Engineering of neo-lectins and janus lectins

Simona Notova

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant
Spécialité : Biologie Structurale et Nanobiologie
Unité de recherche : CEntre de Recherche sur les MAcromolécules Végétales

Ingénierie de neo-lectines et lectines Janus

Engineering of neo-lectins and janus lectins

Présentée par :

Simona NOTOVA

Direction de thèse :

Anne IMBERTY	Directrice de thèse

Rapporteurs :

Birte SVENSSON
PROFESSEUR, Danmarks Tekniske Universitet
Richard DANIELLOU
PROFESSEUR DES UNIVERSITES, Université d'Orléans

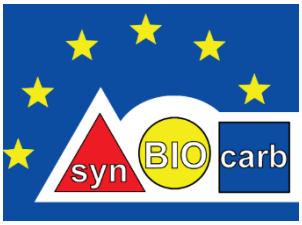
Thèse soutenue publiquement le **29 juin 2022**, devant le jury composé de :

Anne IMBERTY DIRECTRICE DE RECHERCHE, CNRS délégation Alpes	Directrice de thèse
Birte SVENSSON PROFESSEUR, Danmarks Tekniske Universitet	Rapporteure
Richard DANIELLOU PROFESSEUR DES UNIVERSITES, Université d'Orléans	Rapporteur
Marc JAMIN PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Président
Franck FIESCHI PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Examinateur
Winfried ROEMER PROFESSEUR, Albert-Ludwigs-Universität Freiburg	Examinateur

Invités :

The present work was carried out as a part of European Training Network SynBIOcarb. SynBIOcarb bring together a diverse team of biologists, chemists and protein engineers from across the Europe and bridges together synthetic biology and glycobiology for the development and exploitation of diagnostics and targeted drug delivery.

SynBIOcarb has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska Curie grant agreement no. 81402.



This thesis is dedicated to my closest friends.

You have inspired me to change my path.

ACKNOWLEDGEMENT

I would like to thank Prof. Birte Svennson, Prof. Marc Jamin, Prof. Richard Daniellou, Prof. Franck Fieschi, and Prof. Winfried Römer for accepting to be part of my thesis committee and for evaluating of my Ph.D. thesis.

I would also like to thank Prof Franc Fieschi, Dr Ludovic Landemarre, and Dr. Olivier Lerouxel for being part of my CSI committee and for their useful tips and pieces of advice.

Also, I would like to thank all the members of ETN SynBIOcarb, firstly, for the opportunity to be part of it and for being such amazing friends and colleagues. It is a pity that Covid19 affected our pre-planned meetings but even though we kept the spirit and when we met, it was the same as at the beginning. I very much enjoyed all the time we spent together, you taught me so many interesting things, and we have so beautiful memories which I will carry with me. I would also like to thank Dr. Jan Tkac, Dr. Jaroslav Katrlik, Paras and Juvi from Slovak Academy of Sciences, and Dr. Juergen Mairhofer and Natalia from EnGenes Biotech for welcoming me and giving me an opportunity to complete my secondments at your institutions.

I would like to thank all the members of the GBMS group at Cermav CNRS (former and current) for such a warm welcome and friendly environment. Anne, you deserve the biggest thank you from all people and I will come back to it later ☺. Annabelle, thank you for the hours of discussion we spent together. You taught me a lot, especially in structural biology, but not only ☺ Emilie and Valou, the best technicians a person could wish for, thank you for your constant guidance and help. Thank you to Serge, you made every day brighter with your constant good mood ☺ And huge thanks are going to all the students I have a chance to meet. Francois, Lukas, Daniel, Martin, Ons, Nathan, Luca, Sarah, Sue, Kanhaya, Rafael, Dania, Margherita, Dylan, Jalaa, Katie, Josha (I hope that I did not forget anyone) for all the science talks, coffee breaks, lunches and all the amazing time we spent together. I will miss you!

Also, I would like to thank all people in Cermav CNRS with whom we cross our path. Thank you for all the kind of help you offered me.

Special thanks are going to the team of Prof. Winfried Römer, Francesca, and Lina for establishing such a wonderful collaboration. I really appreciate all your help, all the scientific discussion, and all results we have achieved together. I would also like to thank Dr. Frederique Lisacek, Dr. Francois Bonnardel for all the hard work we have completed together. I would

like to thank to Dr. Laurent Heux, Maeva and Josselin, who showed me the world of polymer science and made me to like it a little bit more ☺ I would also like to acknowledge all the scientists who participated in obtaining and summarizing our results in manuscripts.

I would like to thank Dr. Jean-Marie Bourhis for his guidance and help with small-angle x-rays scattering experiments.

I would like to thank my family, especially my mum, dad, and brother who supported me in all my decisions and always believed in me. Mami, oci, ďakujem vám za všetko, čo ste mi v živote dali.

I would like to thank my boyfriend, Pierre-Edouard, who stood by me during this journey and who unconditionally supported me all the time. Merci.

I would like to thank all my amazing, wonderful friends, from Slovakia, Mobka, Tomi and Luky, and also those whom I found here in Grenoble, Sofia, Loli, Dani, Kelly, Lefteris, Lukas, and many others. When I first arrived in Grenoble, I was alone but soon enough you became my family. Since then, we lived so many memories and for all of these experiences, I thank you all! Dakujem! Gracias! Obrigado! Merci! Ευχαριστώ!

Loli, Sofia, you have a special place in my heart and I cannot even say how grateful I am for all the memories we have shared together. Since the moment we spent the first holiday together you became my best friends, my motivation, my support. Thank you! Gracias! Obrigado!

And last but definitely not least, my biggest THANKS are dedicated to my supervisor Dr. Anne Imberty, Queen of Lectins. Anne, I am so grateful for the opportunity you gave me because thanks to you I could change the direction of my path. You are an incredible advisor, always being there for me and supporting me even in the hardest moments. From the moment I started to work in Cermav, I felt comfortable with you and you gave me a lot of courage to believe in myself. You knew exactly how to work with me, how to motivate me, and how to exceed my efforts and achievements over and over. I thank you for every challenge you gave me, you are my biggest idol and motivation, and if one day I myself become an advisor, I wish that one day I will be at least half of the type of supervisor as you are right now. M

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

AA – amino acids

AAL- *Aleuria aurantia* lectin

aaRS – amino-acyl tRNA synthetase

ACG - *Agrocybe cylindracea* lectin

ADCC - antibody-dependent cell-mediated cytotoxicity

AraP – arabinose promoter

AvFc – Avaren lectibody

BC2LC - *Burkholderia cenocepacia* lectin

cAA – canonical amino acid

CBM – carbohydrate-binding module

CBM40 - *Clostridium perfringens* carbohydrate-binding module

$CBM77_{Rf}$ - *Ruminococcus flavefaciens* carbohydrate-binding module

CEA - colorectal cancer biomarker

CEL-III - *Cucumaria echinata* lectin

CERMAV - Centre de Recherches sur les Macromolécules Végétales

CNCs – cellulose nanocrystals

CNRS - Centre national de la recherche scientifique

CN-V – cyanocirin-n

ConA – *Canavalia ensiformis* lectin

CRD – carbohydrate-recognition domain

CTB - cholera toxin B-subunit

DNA – deoxynucleic acid

DC-SIGNR - dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin

DSC – differential scanning calorimetry

EDC – extracellular domain

EEL - *Euonymus europaeus* lectin

EPN – binding motif composed of glutamic acid, proline, and asparagine

ETN – European training network

Fc – crystallization fragment

FXG – fucosylated xyloglucane

GAG – glycosaminoglycan

GalA – galacturonic acid

Gal-1 – galectin 1

GalNAc – N-acetylgalactosamine

GlcNAc – N-acetylglucosamine

gp120 – envelope glycoprotein

GUVs – giant unilamellar vesicles

HA – hemagglutination assay

HG - homogalacturonan

HisTag – 6 x histidine tag

HIV – human immunodeficiency virus

hPSCs – human pluripotent stem cells

IA – inhibition assay

IgG – immunoglobulin G

IMAC - immobilized metal ion affinity

ITC – isothermal titration calorimetry

ITN – innovative training network

LacNAc – N-acetyllactosamine

LacP – lactose promoter

LecA - *Pseudomonas aeruginosa* lectin

LecB - *Pseudomonas aeruginosa* lectin

LSL - *Laetiporus sulphureus* lectin

MAH - *Maackia amurensis* lectin

MBP-A - mannose-binding lectin A

MOA – *Marasmius oreades* lectin

mRNA – messenger ribonucleic acid

ncAA – non-canonical amino acid

NMR – nuclear magnetic resonance

PCR – polymerase chain reaction

PDB – protein data bank

PFT – pore-forming toxin

QPD – binding motif composed of glutamine, proline, and aspartic acid

RE – restriction enzymes

RhaP – rhamnose promoter

RNA – ribonucleic acid

RSL – *Ralstonia solanacearum* lectin

SaroL-1 – *Salpingoeca rosetta* lectin

SAXS – small-angle x-ray scattering

SBI – supplementation-based incorporation

SCS – stop codon suppression

SEC – size exclusion chromatography

SLB – supported lipid bilayer

SPR – surface plasmon resonance

StrA – sortase A

STx1B – Shiga toxin B subunit

tRNA – transfer RNA

SUMMARY

Glycobiology is a rapidly growing field of natural sciences with a focus on glycans, glycoconjugates and glycan binding proteins. Lectins are sugar-binding proteins present in all types of organisms and they display wide range of biological functions. As encoded in their name (from Latin - *legere* – to select), each lectin is specific only to a finite group of glycans and in order to ensure higher affinity, they are generally multivalent. Lectins are powerful glycan-profiling tools and even though some of them already found their applications, the discovery of novel lectins is still desirable. In this regard, involvement of synthetic biology and protein engineering are of high interest for building of lectin architecture and tuning their specificity.

Various approaches for lectin discovery or engineering are presented in this thesis. The thesis is composed of several chapters, where the introduction is dedicated to the general description of lectins and lectin engineering with the respect to their specificity and topology, including a short review on engineering of β-propeller and β-trefoil lectins. The results are presented in three scientific articles (in the format of preprint or manuscripts in preparation). The first publication describes the discovery and characterization of novel pore-forming lectin with specificity toward cancer cells glyco-epitope. In the second manuscript, synthetic biology approach was used to create artificial proteins with the ability to recognize plant cell wall polymers and to be used as glue proteins in the construction of an artificial plant cell wall. The third manuscript generalizes the Janus lectin strategy as a universal tool for creation of bispecific chimeras with increased valency. The last chapter summarizes the achieved results and propose new perspectives and challenges giving the special importance to the continuation of lectin engineering.

RÉSUMÉ

La glycobiologie est un domaine des sciences naturelles en plein essor qui se concentre sur les glycanes, les glycoconjugués et les protéines de liaison aux glycanes. Les lectines sont des protéines de reconnaissance des sucres présentes dans tous les types d'organismes et elles présentent un large éventail de fonctions biologiques. Comme indiqué dans leur nom (du latin - *legere* - sélectionner), chaque lectines est spécifique seulement d'un groupe limité de glycanes et, afin d'assurer une affinité plus élevée, elles sont généralement multivalentes. Les lectines sont de puissants outils de profilage des glycanes et même si certaines d'entre elles ont déjà trouvé leurs applications, la découverte de nouvelles lectines est toujours souhaitable. Dans cette optique, l'implication de la biologie synthétique et de l'ingénierie des protéines est d'un grand intérêt pour la modification de l'architecture ou même de la spécificité des lectines.

Différentes approches pour la découverte ou l'ingénierie des lectines sont présentées dans cette thèse. La thèse est composée de plusieurs chapitres, où l'introduction est dédiée à la description générale des lectines et de leur ingénierie pour ce qui concerne leur spécificité et leur topologie. Elle comprend une brève revue de l'ingénierie des lectines β-propeller et β-trefoil. Les résultats sont présentés dans trois articles scientifiques (préprint ou manuscrit en préparation). La première publication décrit la découverte et la caractérisation d'une nouvelle lectine formant des pores dans les membranes et présentant une spécificité pour des glyco-epitopes de cellules cancéreuses. Dans le deuxième manuscrit, une approche de biologie synthétique a été utilisée pour créer des protéines artificielles ayant la capacité de reconnaître les polysaccharides de la paroi cellulaire végétale et ainsi les utiliser comme glue dans la construction de paroi cellulaire végétale artificielle. Le troisième manuscrit généralise la stratégie des lectines. Janus comme un outil universel pour la création de chimères bispécifiques avec une valence accrue. Le dernier chapitre résume les résultats obtenus et propose de nouvelles perspectives et de nouveaux défis en accordant une importance particulière à la poursuite de l'ingénierie des lectines.

1 INTRODUCTION

1.1 Glycobiology

Saccharides (also called sugars, carbohydrates, or glycans) are together with nucleic acids, proteins, and lipids, the main building blocks in all living organisms. These biomolecules complement each other while having different roles and functions. Proteins and nucleic acids were the first ones to be studied due to their more accessible sequence and structure identification. On the other hand, sugars are far more complex moieties, and only in the late 1980s a new field of natural sciences, called glycobiology, was established (Varki et al., 2017).

Glycobiology is nowadays a rapidly growing field focused on the study of saccharides (free or attached to other biomolecules), including glycomics (identification and analysis of structure and function of glycan sets), enzymology (biosynthesis and degradation of glycans), and glycan interaction with proteins. Unlike proteins, glycans are not directly encoded in the genome. Even though they are synthesized by only narrow spectra of enzymes (glycosyltransferases and glycan-processing enzymes), their composition and structure can be represented by many combinatorial possibilities (type of monosaccharide, branching points, substituents variations, etc.) (Fig. 1). Thus, their characterization is often challenging and requires multiple approaches (physical, chemical, enzymatic, glycan-binding proteins, etc.).

The biological functions of glycans are very broad and diverse. Because they are so widely distributed, their classification can be challenging. For example, Varki (2017) in his extensive review divided the biological roles of glycans into four groups – i) structural and modulatory roles, ii) intraspecies recognition, iii) interspecies recognition, iv) molecular mimicry of host glycans, whereas all these categories might involve glycan-binding proteins (lectins, glycan specific antibodies, and glycosaminoglycan-binding proteins). Additionally, glycans can have more than one function and therefore they can fit into more than one group.

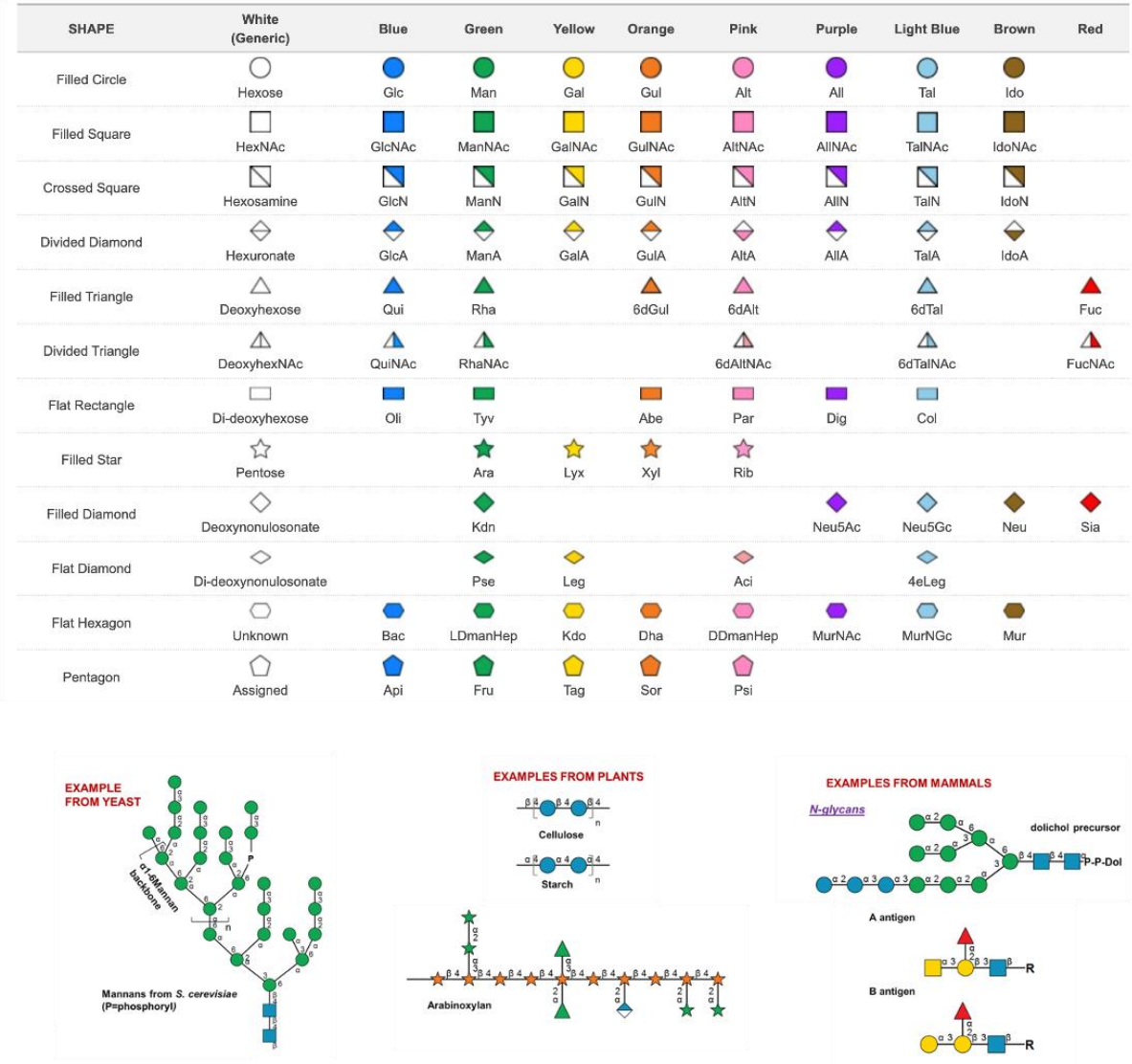


Figure 1: **Drawing of symbol nomenclature for glycans (SNFG)**. This representation is widely accepted by the scientific community and allows illustrating of glycan structures in a clear manner with the respect to their branching points. On the bottom, few examples of glycans present in yeast, plants, and mammals are represented as examples of diversity. Adapted from (Neelamegham et al., 2019; Varki et al., 2015)

The first group (i) involves numerous functions of free or conjugated glycans in the protection, stabilization, organization, and nutrition storage. A typical example is the glycocalyx, the physical barrier composed of various glycan structures, such as mucins, present on the cell surface which serves as a protective "wall" but at the same time operates as a communication interface between the cell and extracellular environment. Intraspecies recognition (ii) refers to glycan functions within one organism. Most of the time the function depends on the interaction between glycan-binding protein and glycan whereas this contact represents a communication

node. Depending on the interaction, there are various outcomes, such as cell death triggering, protein degradation, intercellular signaling or adhesion, cell-matrix interaction, but also induction of fertilization and reproduction processes, or stimulation of the immune system. On a similar base, interspecies recognition (iii) involves also a large spectrum of signaling and communication pathways between glycans and glycan-binding proteins but in this case, the recognition happens between two different organisms. These interactions play a key role in host/pathogen relations and depending on such contacts the pathogens adhere to host cells and initiate the infection. Similarly, the host can recognize the pathogen and induce an immune response leading to pathogen elimination. Protein-glycan interactions are also involved in symbiosis, another type of interspecies recognition where both organisms live together and in some way profit from the existence of the other. Due to the fact that the host organism recognizes the glycans on the surface of pathogens, molecular mimicry of host glycans (iv) appeared during evolution. Indeed, some pathogens found a way how to decorate their surfaces with glycan structures similar or identical to the ones on the host, in order to trick the host immune system and thus avoid the recognition. Glycan mimicry can be used also for host protection, e.g., soluble oligosaccharides present in human milk, which favors symbionts and protect from pathogens in the infant (Varki, 2017).

Glycans are involved in the most extended post-translation modification of proteins called glycosylation, in which a sugar moiety is covalently linked to an amino acid (AA) chain. Such modification provides to proteins the properties already mentioned above (signaling, interaction, etc.) but additionally might improve protein solubility and stability, increase its hydrophilicity and even participate with proper folding. Therefore, correct glycosylation is crucial for the functionality of proteins and cells in general. Disfunction of glycosylation is associated with a very serious congenital handicap, while alteration of glycosylation during life is correlated with all major chronic diseases (cancer, inflammation…) (Varki et al., 2009, 2017).

1.2 Lectins

Lectins are carbohydrate/glycan/sugar-binding proteins present in all types of organisms. They recognize free carbohydrates or the ones present in glycoproteins, glycolipids, and other glycoconjugates. The module responsible for glycan recognition is called the carbohydrate-recognition domain (CRD) and is usually specific for a certain type of monosaccharide or glycan moiety. Lectins can display one or multiple CRDs and they differ from glycosaminoglycan (GAG) binding proteins and carbohydrate-binding modules (CBMs) which are usually associated with enzymes and present as monovalent domains. Even though lectins recognize carbohydrates, they are not enzymes and are not involved in sugar modifications. However, they display a large panel of biological functions, including cell-cell recognition, host-pathogen recognition, signaling, or immune response, but they have a non-immune origin (Sharon & Lis, 2004).

1.2.1 Lectin structures

1.2.1.1 Lectin-glycan recognition

Lectins together with CMBs represent a large diversity of glycan recognition tools. The binding site, or binding pocket, of lectins, is usually a shallow but very well-defined area where the contact between protein and ligand is established (Fig. 2) (Varki et al., 2017). The structural bases for glycan-lectin interaction are generally defined by polar interaction, i.e., direct or water-mediated hydrogen bonds, van der Waals interactions, and π-stacking, with sometimes occurrence of salt bridges and ion-mediated interactions (Fig. 2). The hydrogen bond is very typical for sugar/protein complexes since hydroxyl groups –OH (acceptor or donor) are widely present in glycan residues establishing close contact with the amide, carboxyl, or carbonyl group of the main or side chain of the protein AAs. Additionally, such an interaction can be water-mediated as well. Van der Waals forces are driven by the presence of atoms or molecules in close proximity and even though this type of interaction is considered weak, with an increasing number of atoms involved, the strength of the interaction is increasing as well. Salt bridges are usually established between negatively charged sugars, i.e., N-acetylneuraminic acid, and positively charged AAs, i.e., lysine, arginine, and histidine. Regardless of the fact that carbohydrates are considered highly polar molecules due to their hydroxyl groups, they

contain nonpolar parts as well, formed mainly by -CH on the carbon ring. Thus, a stacking arrangement is frequently observed between the apolar face of the sugar ring and an aromatic ring of AAs, generally termed π-stacking, where π refers to the electron cloud of the aromatic ring of the AA side chain. Some lectins require the presence of divalent cations for establishing a stable binding with their ligand. The typical examples are C-type lectins in which the Ca^{2+} ion is necessary for bond formation between glycan and protein (Weis & Drickamer, 1996).

In general, lectins are specific to a particular type of monosaccharide (galactose, fucose, mannose, sialic acid, etc.), oligosaccharide, or glycoconjugate. While describing them, we refer to fucose-binding lectin, galactose-binding lectin, etc. Moreover, their specificity can be even more exclusive if they are selective to α or β sugar anomers or the position of linkages in oligosaccharides (1-2, 1-3, 1-4). As an example, lectin MOA from fairy ring mushroom *Marasmius oread*es (Kruger et al., 2002) can be mentioned, which is often described as α-galactose-binding lectin, with the preference for the 1-3 linkage.

Because of the diversity, there is a constant effort to determine the lectin specificity by several complement experimental approaches. This requires the combination of biophysical and structural methods of which some are described in chapter 3.

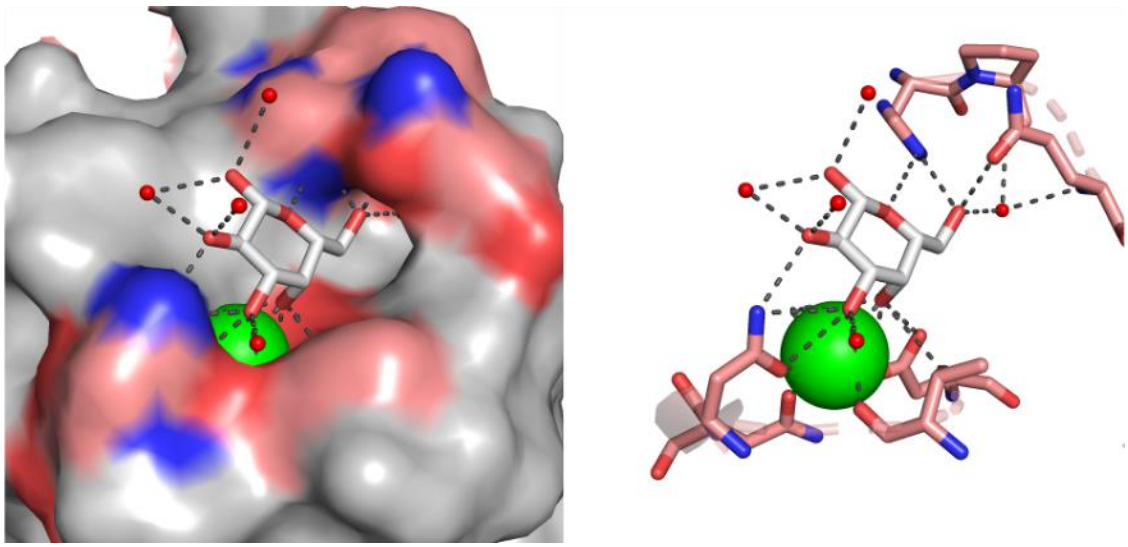


Figure 2: Example of lectin binding site in complex with a ligand. Lectin LecA from *Pseudomonas aeruginosa* in complex with galactose (white sticks) (PDB 1OKO). The interaction between LecA and Gal is established via Ca^{2+} ion (green sphere), and direct (polar/nonpolar) or water-mediated (red spheres) contacts (black dashes). The AAs involved in ligand binding are represented by light orange spheres (left) or sticks (right).

1.2.1.2 Lectin folds

Lectins are present in all organisms, from viruses to mammals, plants, and others. They present a large variety of folds and sequences. Since a large number of 3D structures of lectins are available in the Protein Data Bank (PDB) (Berman et al., 2000), their fold can be used to define a classification. According to UniLectin3D (https://www.unilectin.eu/unilectin3D/) (Bonnardel, Mariethoz, et al., 2019), a manually curated database containing structural information of lectins, there are 75 protein fold classes and the most represented are shown in Fig. 3.

Many lectin folds are rich in β-strand secondary structures, such as β-sandwiches and β-prisms. β-propeller and β-trefoil lectins have the particularity to present tandem repeat, resulting from evolutive replication of simple domains. Their architecture is of interest for protein engineering and they will be more precisely described here.

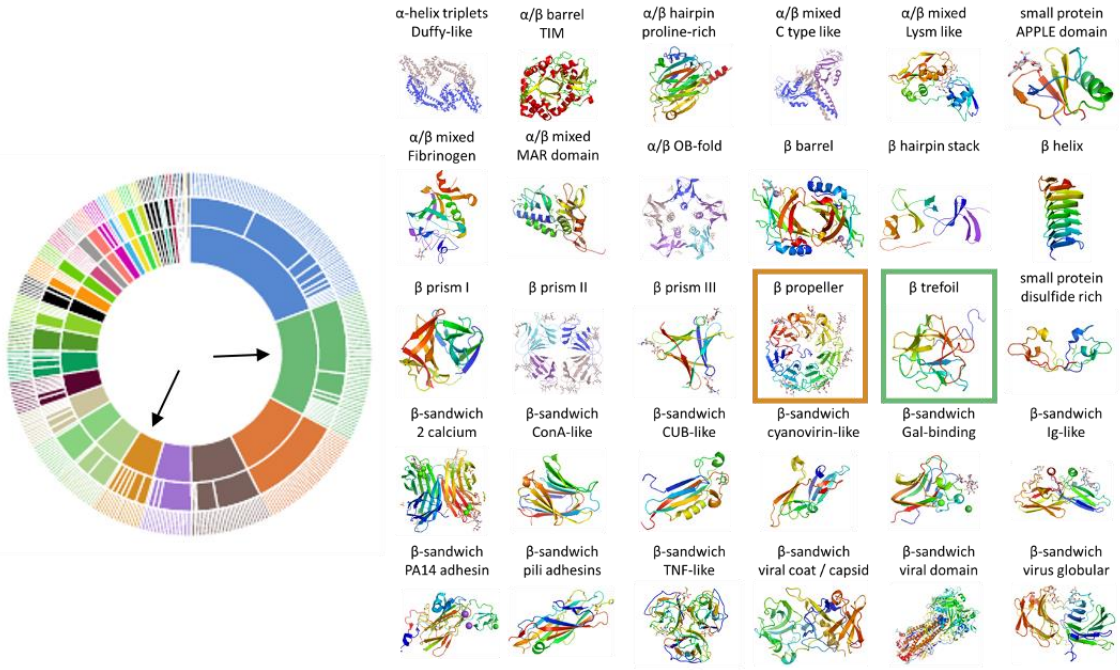


Figure 3: A selection of structural folds observed in lectins. The sun diagram (left) represents a screenshot from the UniLectin3D database of various structural folds found in lectins structures. The arrows indicate β-propellers (orange) and β-trefoils (green) whereas these two classes are depicted in squares of the same color on the right.

The β-propeller fold is widely spread in nature and results in a donut-like shape composed of four to ten repeats (blades), each made of four antiparallel β-strands. This architecture is

commonly generated by the tandem organization (duplication at the gene level), although oligomerization of protomers is also observed (Bonnardel, Kumar, et al., 2019). Due to their architecture, β-propellers have several binding sites on one face of the donut, oriented in the same direction giving them the advantage of multivalent interaction with glycans on cell surfaces. Although they are very well conserved in structure, the AA sequence is variable between β-propeller lectins, which makes them a perfect target for the study of the evolution of proteins and for the protein engineering (Arnaud et al., 2014; Noguchi et al., 2019; Ribeiro et al., 2018; Smock et al., 2016; Terada et al., 2017; Voet et al., 2014; Yadid & Tawfik, 2007). Currently, there are several databases gathering the information about β-propeller lectins, such as UniLectin3D (manually curated database of experimentally solved lectins structures) (Bonnardel, Mariethoz, et al., 2019), and PropLec (prediction and identification of β-propeller lectins) (Bonnardel, Kumar, et al., 2019).

β-trefoil is, according to UniLectin3D, the second most present structural fold found in lectins. Interestingly, the first lectin described, ricin, belongs to the β-trefoil fold family (Rutenber & Robertus, 1991). The β-trefoil architecture comprises three lobes (repeats) forming a barrel structure at one extremity and the triangular arrangement of hairpins at the other. Each lobe accommodates one binding site, resulting in a total of three binding sites per β-trefoil. However, this number is only theoretical, and each β-trefoil can vary the number of binding sites from one to three. Additionally, β-trefoils are often conjugated with an extra domain with diverse functions, such as enzymatic, toxic, pore-forming, etc. Similar to β-propellers, β-trefoils display a very well conserved structure while the sequence similarity is poor. Therefore, they are excellent targets for protein engineering as well, as previously described in numerous publications (Bleuler-Martinez et al., 2017; Hu et al., 2015; Terada et al., 2017; Yabe et al., 2007) and as explored in the present work (see chapter 6).

1.2.1.3 Lectin multivalency

Most of the lectin-monosaccharide interactions are considered as low affinity (micromolar to millimolar range). The affinity is generally stronger for longer ligands such as oligosaccharides or polysaccharides, however, nature developed other strategies how to overcome this issue. The multivalent character of the interaction is a key parameter for the increase of affinity. Due to multivalency, the number of active binding partners increases and so does the strength of interaction which in this case can be described rather as avidity (multivalent interaction) instead

of affinity (monovalent interaction) (Varki et al., 2017). Protein multivalency comes from domains tandem repeats, oligomerization of protomers, or protein presentation of the cell surface. We have discussed the importance of the multivalent organization of lectins, especially in the context of protein engineering, in a review article (see chapter 1.4) (Notova et al., 2020).

Additionally, the topology of binding partners can affect the character of interaction. For example, if lectin binding sites are all oriented in the same direction, usually, an increase in binding strength is expected. On the other hand, if they are arranged in the opposite direction, the outcome might be different and e.g., cross-linking is observed. Likewise, the surface representation of ligand is of high importance, and glycan attachment to other biomolecules, i.e., lipids in the cell membrane, and their topology can enhance the binding or even be absolutely critical for the interaction.

1.2.1.4 Association of lectins with additional domain

Lectins, as protein moieties, can be composed of more than one domain. These domains can be characterized as CRDs or have an additional function, i.e., enzymatic, pore-forming, oligomerization interface, ability to bind to another ligand or others (Varki et al., 2017).

Many glyco-active enzymes present several structural domains that display different function. In addition to the catalytic module, the presence of a CRD, or in this case referred to as a carbohydrate-binding module (CBM) is often observed. Enzyme cleaving one sugar, such as sialidase, may be linked to a sialic acid-binding domain that recognized the substrate, improving the efficiency (Colman & Ward, 1985). An example of one of the most complex systems is represented by cellulosomes, a multi-enzyme machine involved in the decomposition of the plant cell wall, notably cellulose. Such cellulosomes consisting of complex architecture with scaffold domains, CBMs, and enzymes are produced by bacteria living in the rumen and are responsible for the plant cell wall degradation (Israeli-Ruimy et al., 2017; Venditto et al., 2016).

Lectins or CRDs can be also associated with a toxic domain. Two very well-known bacterial examples are cholera toxin from *Vibrio cholerae* (De, 1959) and Shiga toxin from *Shigella dysenteriae* (Keusch, 1998) which are both composed of pentameric lectin domain that binds to cell surface glycans associated with a toxin that is internalized in a host cell (Fig. 4A, B). The association of CRD domain with a pore-forming domain, as previously observed in LSL lectin from mushroom *Laetiporus sulphureus* (Mancheño et al., 2005) or CEL-III from marine

invertebrate *Cucumaria echinata* (Unno et al., 2014) is a less common but noteworthy example (Fig. 4C). The pore-forming activity is triggered by membrane-associated glycan recognition followed by a structural reorganization of protein monomers and their assembly into the oligomer (Boyd & Bubeck, 2018; Mancheño et al., 2005; Savva et al., 2019; Unno et al., 2014). Such activity is probably part of the defensive and/or offensive acts of these organisms. Moreover, the crystal structure of LSL revealed similarities with β pore-forming toxin (PTF) aerolysin (Cirauqui et al., 2017; Degiacomi et al., 2013), an attractive target for biotechnological applications, and thus brings potential possibilities to the field of pore-forming lectins as well.

The additional domain can also serve as an oligomerization interface like previously described in lectin MOA from mushroom *Marasmius oreades* (Grahn et al., 2007) or C-type lectin human receptor DC-SIGNR ((Feinberg et al., 2005). By oligomerization, lectins increase the number of their binding sites, and therefore the stronger or multiple-directed interaction is achieved.

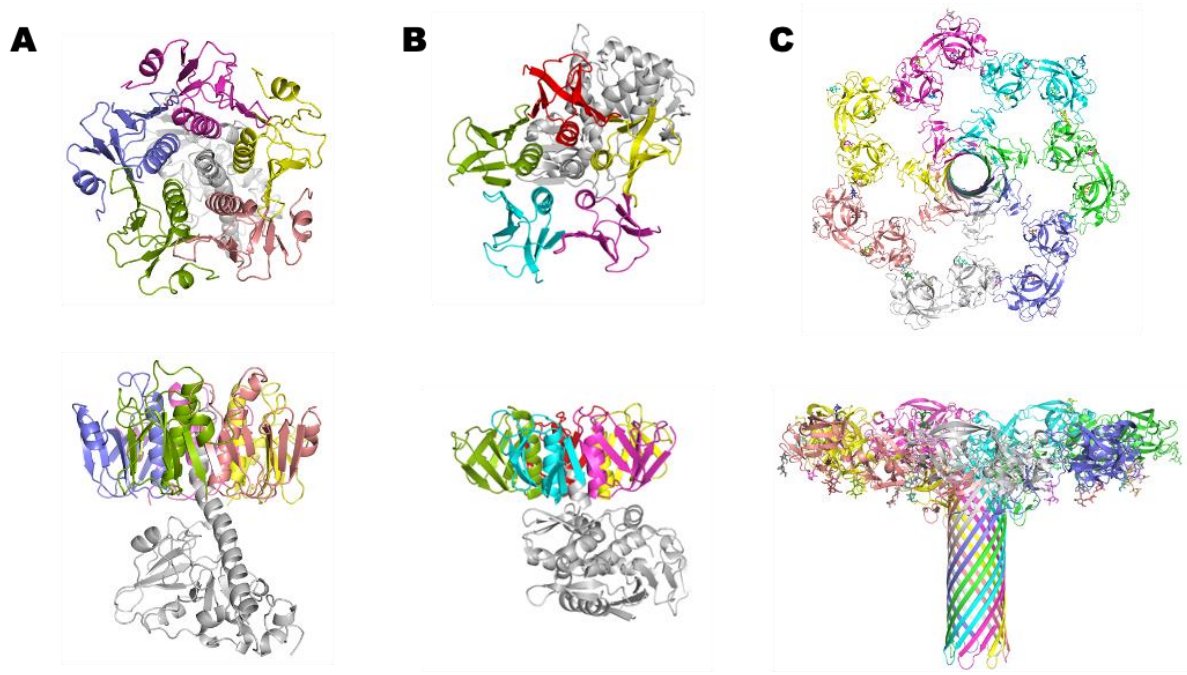


Figure 4: Examples of lectin structures with toxic domain. (A) Cholera toxin (PDB 1XTC) and (B) Shiga toxin (PDB 1R4Q) are composed of pentameric lectin domain (in colors) associated with toxin domain (grey). C) Heptameric pore CEL-III (PDB 3W9T).

1.2.2 Applications of lectins

Glycans and glycoconjugates are widely present in nature. Their composition and conformation are extensively diverse and since lectins can specifically and reversibly recognize and bind glycans, they can be found in versatile applications. Over more than 100 years of lectin history, these proteins started to be used in several fields of research, either as biological agents or as glycan-profiling tools (Sharon & Lis, 2004).

1.2.2.1 Lectins as glycan-profiling tools

Lectins have a great potential to serve as glycan-profiling tools and many are commercially used. As a result of their ability to hemagglutinate erythrocytes by recognizing different oligosaccharides present on the surface of red blood cells, already from the early 20th century, lectins found their first application in the blood group typing (Morgan & Watkins, 2000; Sharon & Lis, 2004). Later on, immobilized lectins were used in affinity chromatography for the isolation and purification of polysaccharides and glycoproteins. Currently, several commercial products can be found on the market, e.g., Agarose Wheat Germ Lectin resin (Merck) with specificity for N-acetyl glucosamine (GlcNAc) or HiTrap Capto Lentil Lectin (Cytiva) for purification of α-D-mannose and α-D-glucose containing biomolecules.

Lectins were also applied in the field of histochemistry and cytochemistry. They can be easily labeled by fluorescent dye or biotin and therefore used for the detection of glycoconjugates by microscopy (fluorescent, optical, or electron) or flow cytometry (Dan et al., 2016). For example, the company Thermo Fischer Scientific offers 18 lectin-fluoro-conjugates such as peanut lectin (terminal β-galactose), concanavalin A (α-D-mannose and α-D-glucose), soybean lectin (terminal N-acetyl galactosamine GalNAc), ricin (galactose, lactose), wheat germ agglutinin (sialic acid, GlcNAc), and others coupled with various fluorescent labels. Additionally, lectins found their application also in lectin blotting, a technology analogous to immunoblotting (western blot). The method is more specific to glycoproteins and together with histo- and cytochemistry, this procedure represents a standard in the glycan-profiling techniques. The recognition of glycoconjugates can be used for simple detection or even trafficking in the cells, tissues, and organelles (Dan et al., 2016).

The glycosylation of proteins can alternate depending on the physiological state of the cell and therefore the glycan profile can change upon disease development, such as cancer. Early detection of cancer or metastasis is always challenging. However, prompt identification of

cancer biomarkers by lectins could lead to the acceleration of cancer recognition and therefore increase the chances of patient survival. Yang et al. (2012) developed a quantum-dot-based lectin biosensor for the detection of colorectal cancer biomarker (CEA). By immobilization of lectins EEL (*Euonymus europaeus* lectin, specific for mannose) and ConA (*Canavalia ensiformis* lectin specific for mannose and glucose) on the aldehyde-modified glass surface, the presence of cancer biomarker CEA was detected even bellow the required diagnostic cut-off value.

The idea of constructing a lectin-based biosensor can be exploited also in a different field than cancer biomarker detection. As summarized in the recent review (Mi et al., 2021) lectin-based biosensors were used for rapid detection of foodborne pathogens. Even though lectin biosensors offer many advantages, such as high sensitivity, fast identification, etc., the absence of a high-throughput format is still limiting.

1.2.2.2 Lectin microarray

A special attention is dedicated to the newest application of lectins, called lectin microarray. The method was first time published in 2005 (Angeloni et al., 2005; Kuno et al., 2005; Pilobello et al., 2005; Zheng et al., 2005) and since then, several companies developed their own setups which are currently available on the market, such as GLYcoDiag, Creative Biolabs, CD Microarrays, Asparia Glycomics, RayBiotech, and others.

The idea behind this technology is to gather and immobilize various lectins on the surface of the chip and use them for the detection of glycans or glycoconjugates present in the sample. In comparison with mass spectrometry-based methods (also used for characterization of glycans), lectin microarray is considered a simpler, highly sensitive, robust, and especially high-throughput technique that requires lower sample purity (crude extract, cells) and can directly analyze glycans attached to their conjugates (avoiding purification step). However, the technique is more suitable for comparative studies than quantity determination (Dang et al., 2020).

There are several ways how to perform lectin microarray. In the direct assay, the lectins with known specificity are directly immobilized on the surface of the chip and subsequently incubated with fluorescently-labeled glycoproteins (Fig. 5A). The same composition is used in the antibody-overlay lectin sandwich array but this time, the bound glycoproteins are probed with labeled antibodies (Fig. 5B). Due to antibodies, the setup is more sensitive in the detection

and already trace amounts of glycoproteins can be identified. The exact opposite is used in the lectin-overlay antibody sandwich array and therefore antibodies are immobilized on the chip while labeled lectins are used for profiling glyco-variants in the proteins (Fig. 5C). All these techniques require labels, mostly fluorescent ones (Dang et al., 2020; Propheter et al., 2011).

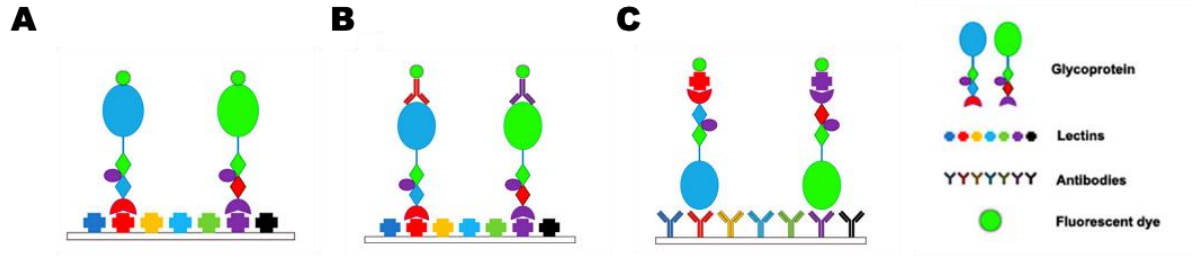


Figure 5: Different types of lectin microarrays. A) Direct assay, B) antibody-overlay lectin sandwich, C) lectin-overlay antibody sandwich. Adapted from Dang et al. (2020).

As previously described, the nature of glycoproteins can vary due to physiological changes, e.g., carcinogenesis. Therefore, lectin microarray represents a suitable tool for the biomarker detection (Dang et al., 2020; Hirabayashi et al., 2013). Additionally, the lectin array can be used as a quality-control tool for monitoring proteins glycosylation and cell differentiation (e.g., induced pluripotent stem cells and embryonal stem cells) (Tateno et al., 2011).

1.3 Discovery of new lectins

The first description of lectins was found in the doctoral thesis of Peter Hermann Stillmark in 1888 where he characterized ricin, a highly toxic lectin isolated from castor beans. Due to their ability to agglutinate cells/erythrocytes, lectins were originally referred to as agglutinins/hemagglutinins. In 1954 William C. Boyd and Elizabeth Shapleigh introduced the term "lectin" from the Latin word *legere* - "to select". Throughout the 20th century, more and more plant lectins were characterized but only in the 1960s, the first animal lectin (asialoglycoprotein receptor) was discovered in the serum (Sharon & Lis, 2004). Later on, the presence of lectins was also confirmed in other organisms and soon they attracted ample attention due to their ability to detect, isolate and characterize glycoconjugates. Since then, many new lectins were discovered, characterized, and found their biotechnological and diagnostic applications. However, the urge for discovering novel lectins and understanding their biological functions still remains.

1.3.1 Classical approach – extraction from natural sources

Originally, lectins were and still are extracted from their natural sources such as plants, fungi, or animals by classical biochemical methods. This usually includes biomaterial isolation, homogenization and crude extract serial purifications whereas additional steps might be applied (protein precipitation, filtration, etc.) (Fig. 6A). The introduction of affinity chromatography with immobilized carbohydrates in 1965 (Agrawal & Goldstein, 1965), rapidly increased the number of obtained lectins. In 1973, the first recombinant protein (insulin) was produced and this technology was also applied to lectin production. However, this was conditioned by the fact that the primary structure of the lectin was already known. Compared to lectin extraction from natural sources, recombinant protein production brings a lot of advantages, such as more controlled expression, easier purification, or higher protein yields. However, *in vivo* heterologous production of some lectins can be very challenging, due to their complex structure (disulfide bonds) or toxicity to production cells, and alternatives, such as cell-free protein production might be a solution (Jaroentomeechai et al., 2020).

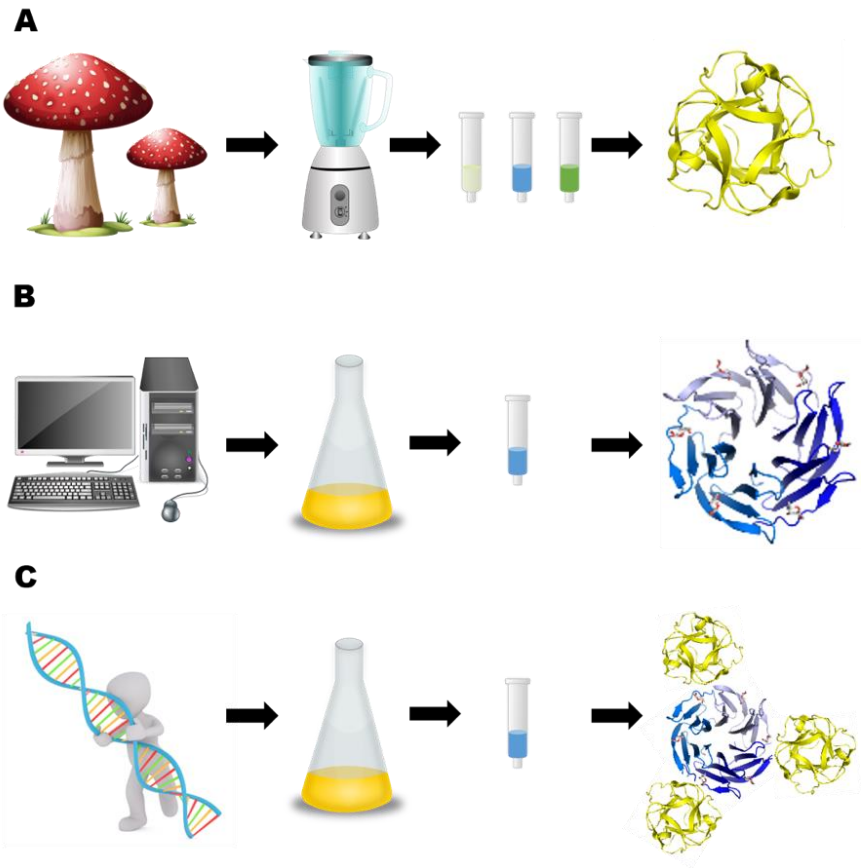


Figure 6: **Three main strategies used in lectin discovery**. A) Classical approach is based on the isolation of lectins from natural sources, B) Data mining or in silico approach use searching tools, such as databases, for prediction and discovery of new lectins, C) Synthetic biology involves numerous tools for protein engineering to create novel protein variants.

1.3.2 Data mining approach

With the exponential volume of data produced by sequencing organism's genomes, bioinformatics is another way to discover new lectins. *In silico* approach can be also used for searching for the best candidate/match which possesses specific requirements. The prediction is mostly based on bioinformatics analysis of either nucleotide or AA sequence of proteins and their comparisons to already known lectins. For this purpose, online databases represent a very useful and powerful tool how to execute an effective virtual screening. In the field of life sciences, there are several well-known databases that gather information about genomes and proteins. A widely recognized example is the UniProt database which provides a free set of protein sequences with their functional annotation (Bateman et al., 2021). Once the suitable candidate is recognized, the protein is produced as recombinant, purified, and subsequently characterized (Fig 6B). On the other hand, the databases with a focus on glycobiology are much

less represented. Nevertheless, a web-based glycoinformatic resource Glyco@Expasy gathers information about various glycan-oriented tools, databases, and searching interfaces (Mariethoz et al., 2018).

UniLectin3D is a manually curated database of experimentally solved lectin structures (Bonnardel, Mariethoz, et al., 2019). Each structural family of lectins is characteristic by its fold (β-propeller, β-trefoil, etc.) but often by its sequence motifs as well. This provides an alternative way how to identify new lectins or CRDs. In the age of high-throughput sequencing, currently, more than 115 000 genomes have been already sequenced and deposited in the databases (RefSeq, January 2022). Therefore, these genomes can be used for sequence screening and detection of CRDs or lectins as performed by LectomeXplore (Bonnardel et al., 2021). The presence of a particular sequence motif allows identifying the structural fold of predicted CRD or lectin.

The database LectomeXplore is a noteworthy example of a database dedicated to the prediction of lectins. Bonnardel et al., (2021) proposed a new lectin classification based on (i) 35 lectin domain folds, (ii) 109 classes of lectin sharing at least 20% sequence similarity, and (iii) 350 families of lectin sharing at least 70% sequence similarity. LectomeXplore offers multiple ways to search for new lectins as shown in Figure 7. If there is an interest in particular organisms, a search by taxonomic tree including archaea, bacteria, eukaryote, or viruses is preferable. On the other hand, exploration by structural fold with the extension to lectin classes is available as well (Fig 7A). Once a potential candidate is chosen, the website displays several panels including NCBI and UniProt reference, if available, species, length of the protein, score, etc. (Fig 7B). The domain is also aligned and compared to the reference consensus sequence and the AAs involved in the glycan-binding site in the reference PDB structure are shown as well. LectomeXplore also offers a gene viewer tool where the position of predicted lectin can be localized on the gene and therefore can be used as verification and confirmation of correct prediction.

An alternative is the PropLec database (Bonnardel, Kumar, et al., 2019) which is focused only on the prediction of β-propeller lectins. For the prediction, the database has defined the conserved motifs at the level of the individual blade, which avoids the difficulties encountered with tandem repeats. PropLec together with LectomeXplore uses the interface of UniLectin3D and therefore filter parameters, such as lectin families, taxonomy, or similarity score, occur as

well. On the other hand, PropLec offers a new feature in which the prediction is done by the selection of the number of blades.

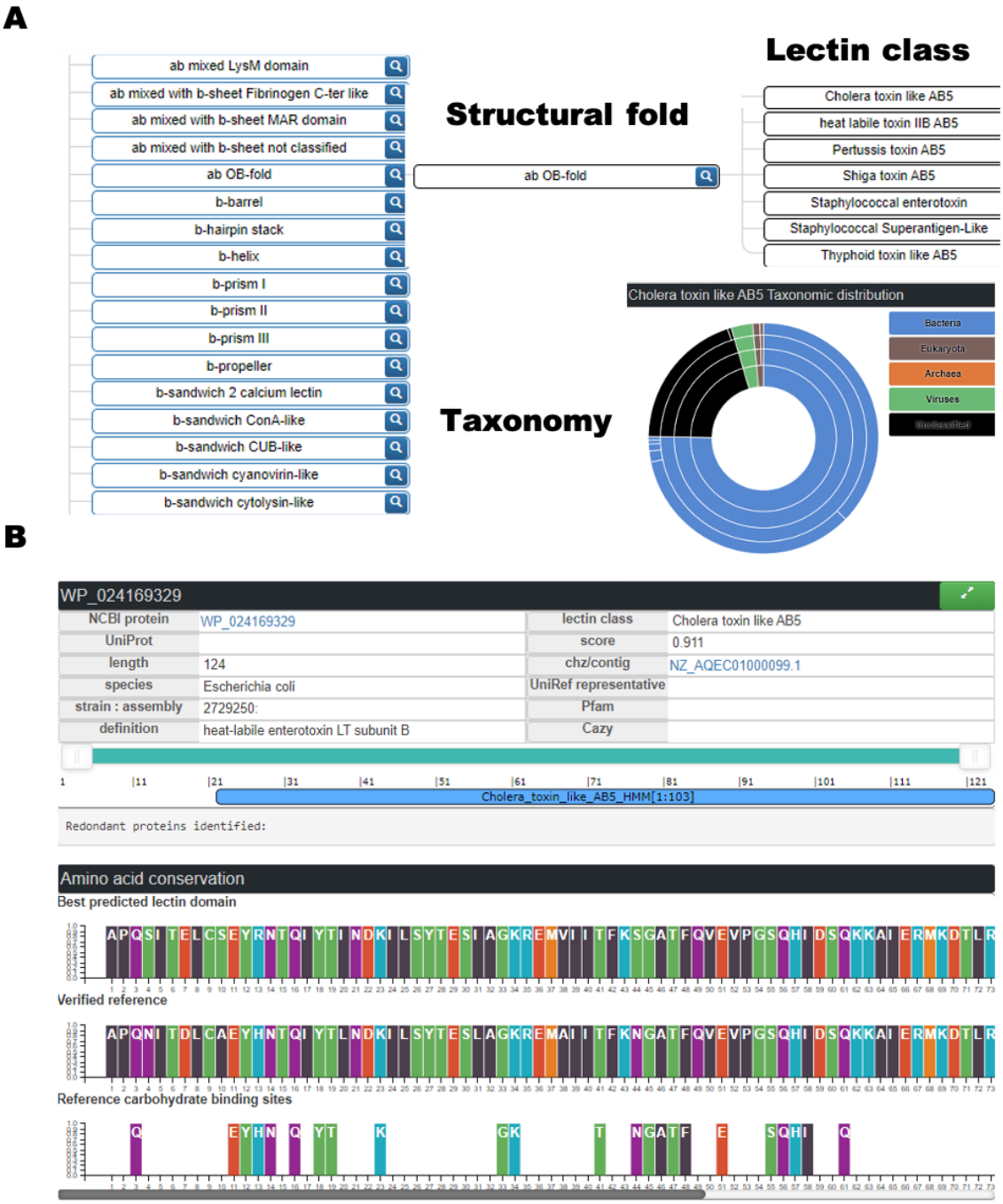


Figure 7: LectomeXplore as a prediction tool for the discovery of new lectins. A) The database offers prediction based on taxonomy, structural fold, and lectin class, B) Example of a screenshot of LectomeXplore information panel and amino acid alignments of predicted and the reference sequence.

The *in silico* approach was also used for the prediction of β-trefoil lectin SaroL-1, closely described in chapter 4. Similar to PropLec, the database dedicated to the prediction of β-trefoil lectins was created, using a sequence motif at the lobe level, and named TrefLec. The database is as well part of the UniLectin3D portal and was currently published in our newest article (Notova et al., 2022).

Prediction of specificity for a putative lectin is much more complicated. Lundstrøm et al., (2022) created LectinOracle, an open-access deep-learning model that predicts lectin/ligand interactions. LectinOracle was compared with the experimentally proved lectin specificities and the results were mostly in agreement with literature-annotated studies. Moreover, LectinOracle has the potential to analyze whole lectomes and thus evaluate their role in host-pathogen interactions. Similarly, Bojar et al., (2022) used deep machine learning for evaluating the specificities of commercially available lectin by glycan array. Even though most of the results corresponded to the ones which were previously described in the literature, for certain lectins they revealed additional insights of lectin binding. Nonetheless, the prediction needs to be always confirmed with experimental observation since information like biological function or protein organization is lacking.

1.3.3 Lectin engineering and synthetic biology

As described previously, numerous lectins have been discovered in nature and many of those found their applications. However, there are certain drawbacks, such as not sufficient selectivity or affinity, that might be disadvantageous for lectins. Therefore, in the last decades, lectins became a target of protein engineering with the clear vision to create suitable candidates with the potential in biotechnology and biomedicine applications.

1.3.3.1 Synthetic biology

Synthetic biology is a rapidly-growing, multidisciplinary field that comprises the design and engineering of biomolecules, biological machines, and even the creation of novel biological systems and life forms, i.e., artificial cells and organisms. The main purpose of synthetic biology is to introduce new functions and properties which does not exist in nature, however, these qualities are often inspired by nature itself. Throughout the history, several important milestones, such as the development of molecular cloning techniques and recombinant protein production, automated genome sequencing, CRISPR/Cas9, bacterial biofuel production, and numerous incorporations of built circuits into the cells, have been achieved and found their applications in the field (Cameron et al., 2014). Synthetic glycobiology can be considered a subgroup of synthetic biology and its main area is the engineering of glycosylation for the production of novel glycoconjugates and alteration of glycocalyx or engineering of carbohydrate-binding proteins and carbohydrate-active enzymes (Hirabayashi & Arai, 2019; Jaroentomeechai et al., 2020; Kightlinger et al., 2020; Turnbull et al., 2019; Ward et al., 2021). In this chapter, we focus on lectin engineering with a special interest in the introduction of additional specificity and multivalency.

1.3.3.2 Engineering of specificity

As stated in previous chapters, lectins display a large panel of biological functions and therefore already found their applications in numerous fields. However, it could be of interest to control their properties and for example narrow their specificity and selectivity. The modified lectin would be in theory more suitable for a particular purpose or application.

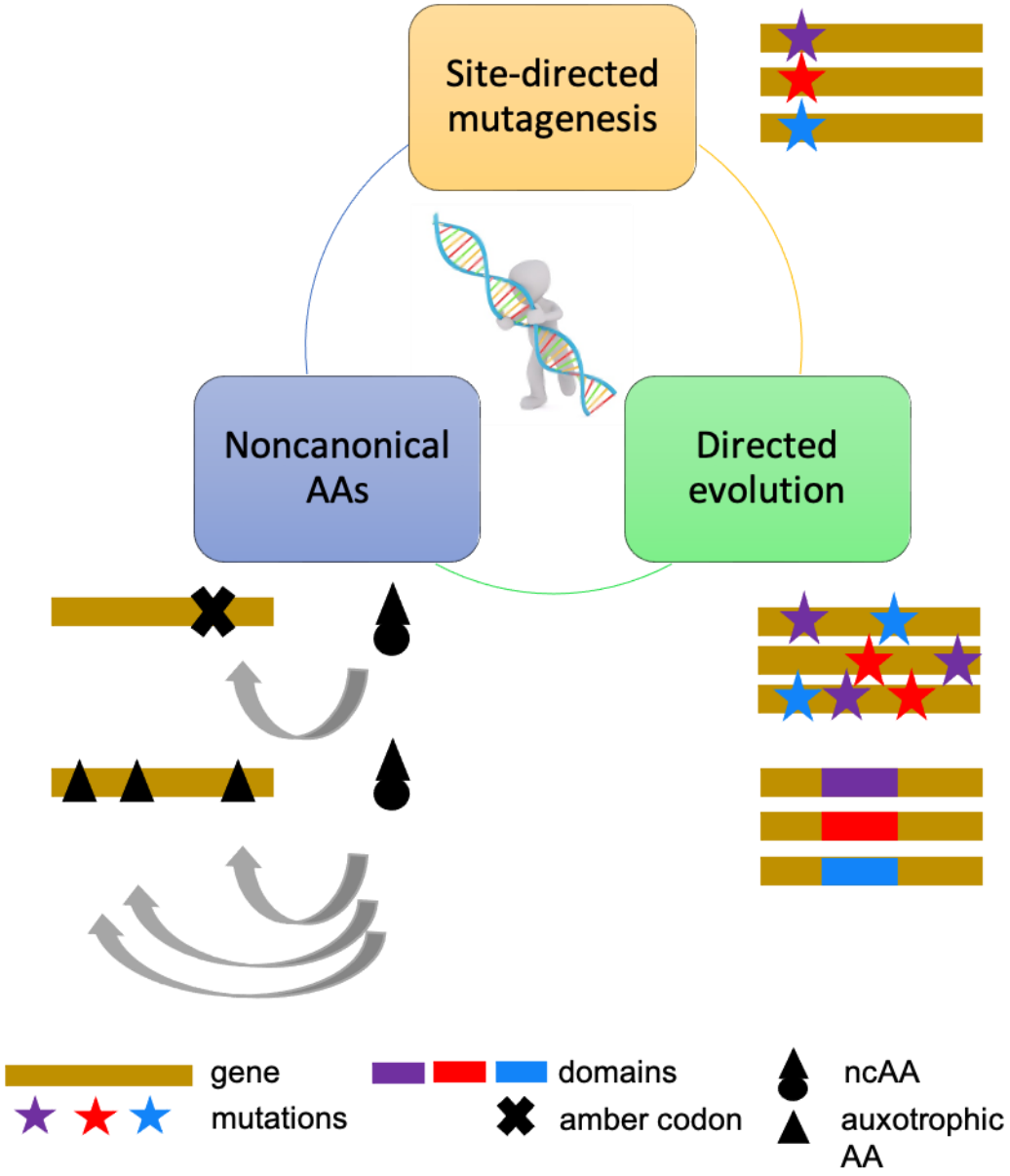


Figure 8: Synthetic biology as a tool for engineering of lectin specificity. Lectin specificity can be altered at the gene level with modifications performed by directed evolution approaches like site-directed mutagenesis, random mutagenesis, and domain shuffling. The second option involves the incorporation of noncanonical amino acids in the position of an amber codon, or by replacement of auxotrophic amino acids.

As shown in Figure 8, various attempts to engineer lectin specificity have been performed, including methods of directed evolution, such as point or site-directed mutagenesis, random mutagenesis, exon or domain shuffling, or even incorporation of non-canonical amino acids (Hu et al., 2015).

Site-directed mutagenesis

Probably one of the very first examples of engineering of lectin specificity was described by Drickamer in 1992. In this study, the specificity of a C-type mannose-binding lectin A (MBP-A) from a rat was modified toward galactose. Firstly, the consensus sequence motif for selective recognition of monosaccharides was determined. C-type lectins specific for mannose are characterized by the presence of a QPD motif whereas the mannose-binding motif is described as EPN. Subsequently, the galactose binding motif was substituted for the mannose, and the new mutant was able to bind galactose, however with some residual activity toward mannose. This study thus might be considered as the beginning of the lectin engineering era and since then few successful attempts have been reported.

One such example was reported by Adam et al., (2007) where site-directed mutagenesis was applied on lectin LecB from *Pseudomonas aeruginosa.* LecB is a fucose binding lectin with unusually high affinity due to the unique mode of binding mediated by two calcium ions. Its ortholog, RS-IIL from *Ralstonia solanacearum* is a mannose-specific lectin and as revealed by structural comparison of AAs involved in carbohydrate and calcium-binding pocket, the specificity is under the control of one amino acid in a variable tripeptide. Therefore, the study was designed to mutate this triad of AAs in order to shift the specificity from fucose to mannose. These mutants were subsequently examined by biophysical and structural studies. Successfully, certain mutants showed improved affinity toward mannose, but this did not exceed the one from RS-IIL. Surprisingly, more promising results were shown with other mutants that showed even higher affinity for the common LecB ligands than parental protein. Thus, rational design and site-directed engineering of lectins have great potential in biotechnological and medical applications.

By site-specific mutation of AAs in the binding site of lectin MAH from *Maackia amurensis*, a library of 35 mutants was created. MAH is specific for sialic acid, especially for T-antigen and disialyl T-antigen and the mutated-lectin library showed preferences for different glycoconjugates. Furthermore, these mutants have unique binding specificity which was confirmed by the recognition of various surface carbohydrates on different cell lines, and thus they can be used for cell line profiling, i.e. in the lectin microarray (Maenuma et al., 2008).

Randomized direct evolution approach

Generally, the strategy of directed evolution is of high interest in protein engineering. As deduced from the name, this approach "directs" the evolution by a series of random mutations that are directed to a target ligand. The process is supposed to follow the real events of evolution, however in a speed-up procedure.

Site-directed mutagenesis may be combined with a randomized directed evolution approach and together might serve as a tool for altering already known specificity (Romano et al., 2011). Lectin AAL from mushroom *Aleuria aurantia* is a β-propeller lectin with five fucose binding sites that do not distinguish between linkages in α-1,2, α-1,3, α-1,4, and α-1,6. The AAL mutants were created by point mutations and domain exchanges to create the binding sites with selectivity for certain linkages. By this approach, certain mutants showed almost 10-times higher preferences for α-1,6 while compared to native AAL. Thus, lectin engineering toward increased selectivity has potential in various applications, i.e., diagnostics assays.

Yabe et al., (2007) continued with directed evolution and applied the strategy of error-prone PCR in β-trefoil lectin EW29 from earthworm *Lumbricus terrestris.* The strategy was to mimic natural evolution which was followed by a reinforced ribosome display system, resulting in including additional specificity in a lectin. The parental lectin EW29 is specific for galactose whereas its mutant showed affinity to α-2,6sialylated glycans. Nonetheless, the mutated lectin maintained the original specificity to galactose as well, however with decreased affinity.

Noncanonical amino acid incorporation

A different approach can be used by incorporating non-canonical amino acids (ncAAs) in order to modulate lectin-carbohydrate interaction. The proteins are composed of 20 canonical AAs (cAAs) however there are hundreds of possible variants, i.e., ncAAs which are not coded in the genome but can be obtained by chemical modifications. These variants, more precisely their side chains, are decorated with azido-, alkyne-, alkene-, tetrazine-, cyclooctyne-, ketone- or other functional groups, and they can be incorporated into the protein chain by harnessing the host ribosomal translation system and thus they can induce structural and functional changes.

There are several ways how to achieve protein incorporation of ncAAs. One of the options involves a method named stop codon suppression (SCS) (Wiltschi, 2012). This method is

preferentially used for the incorporation of ncAA in a single specific position of protein. There are three stop codons, UGA (opal), UAA (ochre), and UAG (amber), however, in bacteria *E. coli*, the amber codon is the rarest one and thus it is unlikely that this codon terminates the sequence of essential genes. Therefore, an amber codon can be reassigned to a ncAA. In order to reach incorporation of ncAA in the side of an amber codon, a selective pair of amino-acyl tRNA synthetase/amber suppressor tRNA (aaRS/tRNA) specific for amber codon and ncAA must be exploited or developed. Such a pair must be orthogonal, i.e., mutually compatible, and recognized by the host ribosomes. At the same time, aaRS/tRNA should not cross-react with the host system, i.e., aaRS does not accept cAA, and amber suppressor tRNA is not charged with cAA (Fig. 9B). Nonetheless, the host termination machinery is in non-stop competition with the SCS system which often results in low yields or truncated proteins (Dumas et al., 2015; Tobola et al., 2019).

On the other hand, several lectins have been modified by the supplementation-based incorporation (SBI) method (Shanina et al., 2021; Tobola et al., 2018, 2022; Wiltschi, 2012). This method is dependent on the engineering of an auxotrophic host system for certain cAA and during recombinant protein production, the host strain is supplied by its analog, i.e., ncAA or its precursor (Fig. 9C). The main advantage of this technique, especially if used in bacteria *E. coli*, is that high yields and qualitative incorporation are achieved. Moreover, *E. coli* is relatively easily engineered to be auxotrophic and such strains are already commercially available. SBI's bottleneck might be the possibility of disruption of the protein structure or function once the ncAA is incorporated, however, as described in the next paragraphs, these properties could be advantageous, too. Furthermore, ncAA might be incorporated also in host proteins which can affect the general fitness of the host (Tobola et al., 2019, Wiltschi, 2012).

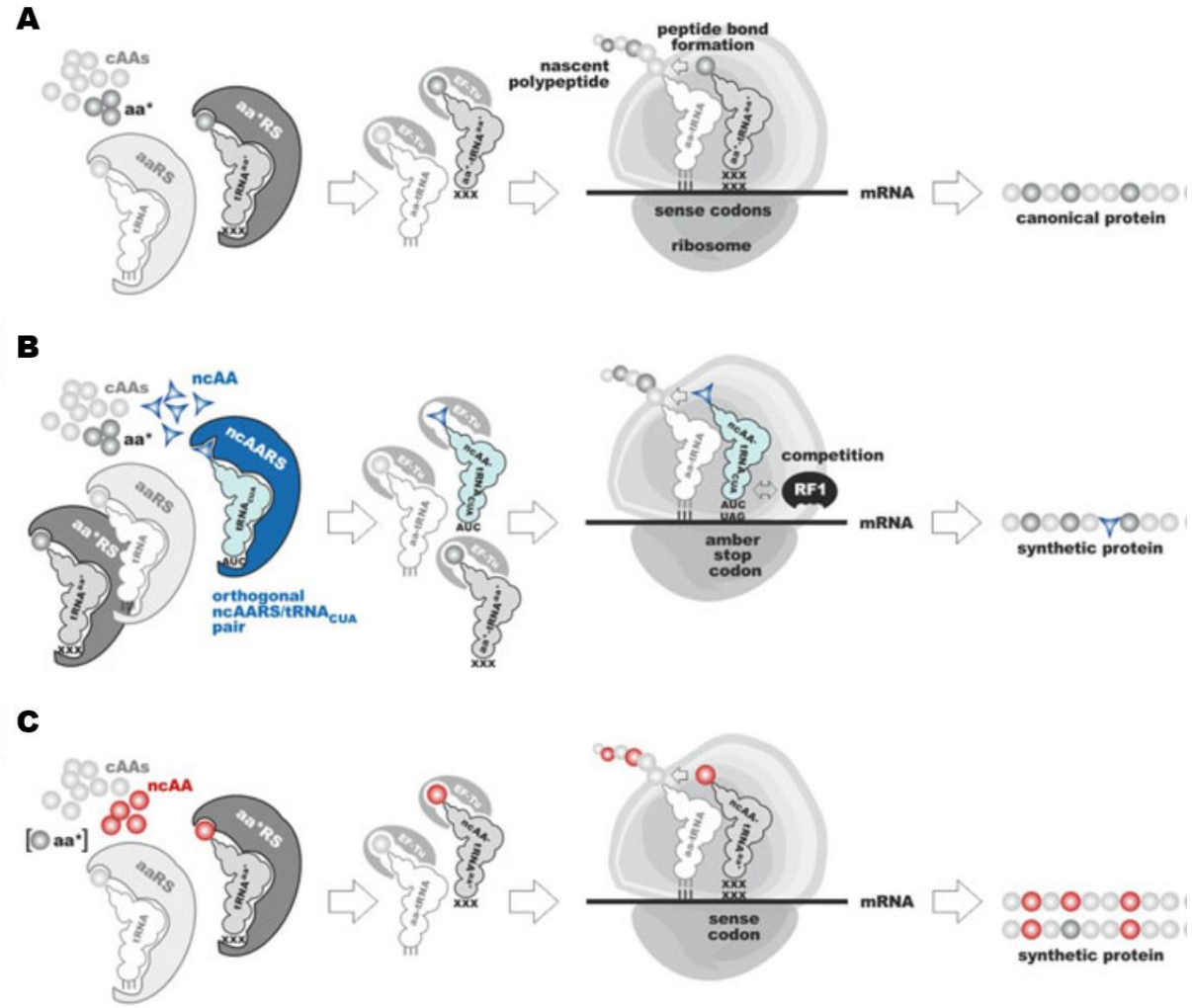


Figure 9: Translation process modified by incorporation of ncAA. A) naturally occurring ribosomal translation, B) Stop codon suppression (SCS) translation, C) Supplementation-based incorporation (SBI) translation. Adapted from Glieder et al. (2015).

By this approach, Tobola et al., (2018) modulated lectin RSL from *Ralstonia solanacearum* with fluorinated tryptophan analogs which resulted in changes in lectin stability and specificity. The incorporation was assured by using tryptophan auxotrophic *E. coli* strain, and bacteria were supplemented with indol compounds which served as tryptophan-analog precursors. RSL is a β-propeller with six binding sites for fucose. Each binding site contains three tryptophans which are establishing hydrogen, hydrophobic, and π stacking interactions with sugar residue. They created together four variants, each with a different position of fluorine on the indole ring of Trp. These variants were compared with the parent protein RSL in terms of stability, specificity, affinity, and structure as well. Their results showed that modification of Trp side chains allowed the binding to oligosaccharide ligands, but modified the thermodynamic

signature and stability of lectin. Moreover, the variants display altered preferences for different blood group oligosaccharides due to extra fluorine. Similarly, the structural analysis showed that incorporation of ncAA might induce various effects, such as loss of stacking or formation of an additional bond between fluorine and sugar.

Tobola et al. (2022) subsequently modified Galectin 1 (Gal-1) with the same approach. Gal-1 is specific for β-galactosylated glycans, i.e., N-acetyllactosamine (LacNAc), and is involved in numerous biological functions in mammalian cells. Over the years, Gal-1 became one of the most studied lectins and thus this makes it a perfect candidate for lectin engineering. Contrary to RSL, Gal-1 has only one Trp located in the binding site, thus incorporation of ncAA was more subtle. Parental protein was substituted in total with 12 different Trp noncanonical variants. Even though some variants showed similar binding properties than parental Gal-1, interestingly, two variants with 7-azatryptophan and 7-fluorotryptophan showed significantly reduced affinity to 3'-sulfated glycans. These findings, therefore, support the idea that incorporation of ncAA might be a suitable strategy for fine-tuning lectin specificity.

1.3.3.3 Engineering of topology

Lectins are multivalent and the spatial position and orientation of their carbohydrate-binding sites, i.e., their topology, directly participate in their function. Nonetheless, the engineering of lectin architecture is currently of interest. Differently from specificity engineering, Figure 10 shows that this field focuses mostly on protein symmetry (neo-lectins), valency (clickable lectins, coiled-coil, fusion proteins), the introduction of additional specificity, or even the creation of chimeras with additional property. These approaches result in the formation of proteins with novel functional and structural properties and with potential for biomedical and biotechnological applications.

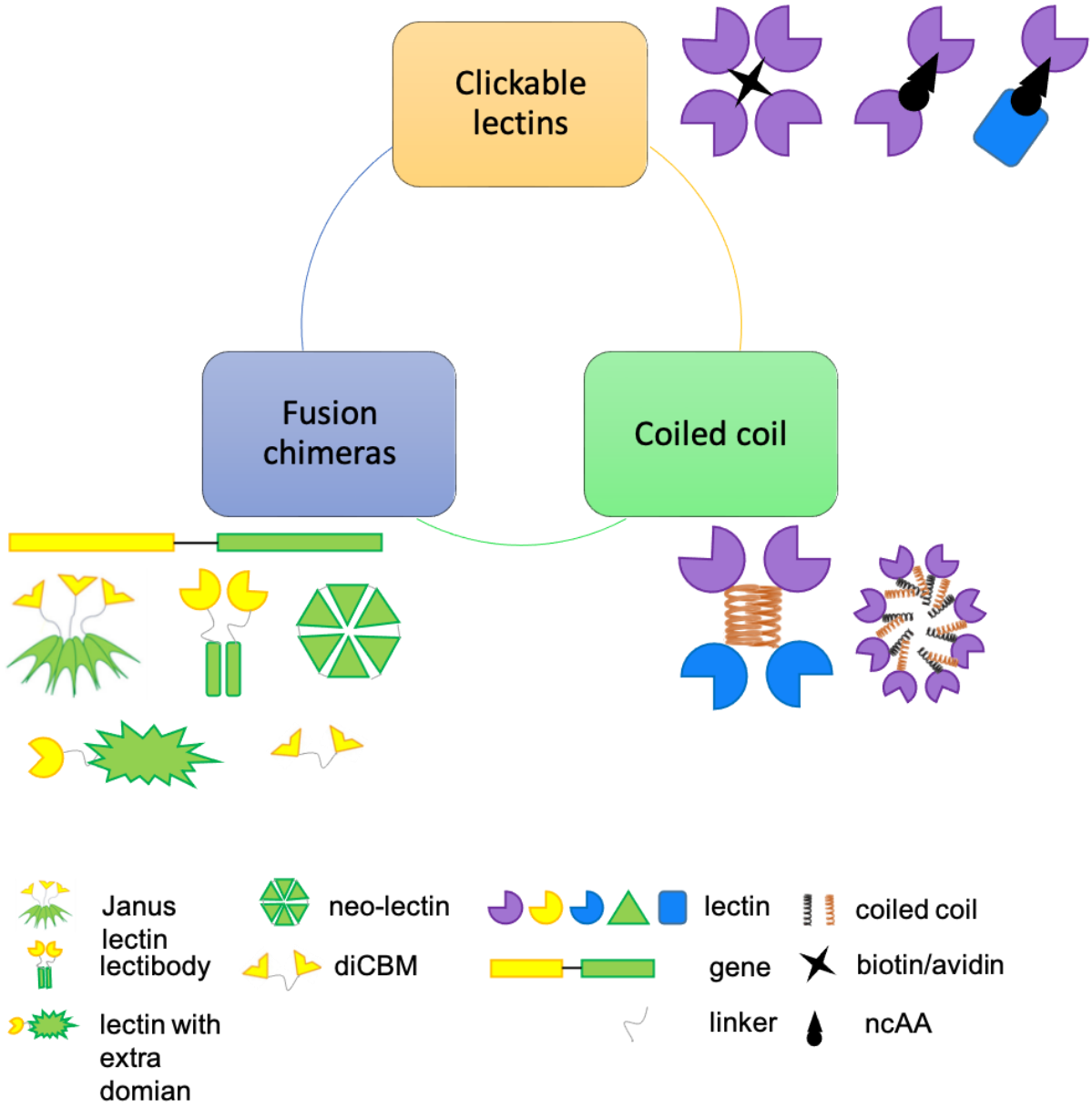


Figure 10: Synthetic biology as a tool for engineering of lectin topology. There are numerous possibilities for how to alter lectin topology. The strategy "clickable lectins" involves different types of clickable handles, with introduction of biotinylated sequences or noncanonical amino acids, which conjugate two or more lectins together. Coiled-coil employs helical structures to organize proteins in a higher oligomerization state. The third option comprises introduction of linker to create fusion chimeras. This group is also probably the most diverse with proteins with modulation of valency and introduction of additional specificity, or functionality.

Conjugation of lectins by a clickable handle

Conjugation of proteins by the addition of a "handle" that can have multiple functions can be obtained by chemical approaches (Hermanson, 2013). However, protein engineering can also be used in order to obtain lectin conjugates, for example, to produce covalent multimer. This can be achieved for example by incorporating ncAA. Tobola et al. (2019) modified in this way three lectins, RSL, STx1B (Shiga toxin 1 B-subunit), and human Galectin-1 (Gal-1) with three different ncAAs (Fig. 11A). The lectins incorporated ncAAs (L-azidohomoalanine (Aha) and L-homopropargylglycine (Hpg) by SBI method, and L-azidolysine (AzK) by SCS method) in different positions, while the binding pocket was kept intact in order to maintain lectin functionality. NcAAs conjugation was first performed with a fluorophore to prove their accessibility. In a second step, direct conjugation between monomers was performed resulting in homo- and hetero-lectin chimeras (Fig. 11B). After optimization, it appeared that the insertion of a linker compatible with "click" groups of ncAAs resulted in a higher yield. Several linker possibilities were tested and they all proved to be capable of creating homo-dimers, i.e., Gal-1/Gal-1, RSL/RSL, STx1B/STx1B, whereas the presence of only one "clickable" handle was sufficient for conjugation. However, with the increasing number of oligomerization, the dimer formation appeared to be more problematic and the presence of monomers was observed as well as in the case of direct conjugation. Nonetheless, engineering of lectins by ncAAs has the potential to create lectins with interesting topology, and additionally, it might introduce extra specificity and therefore create 'superlectin'.

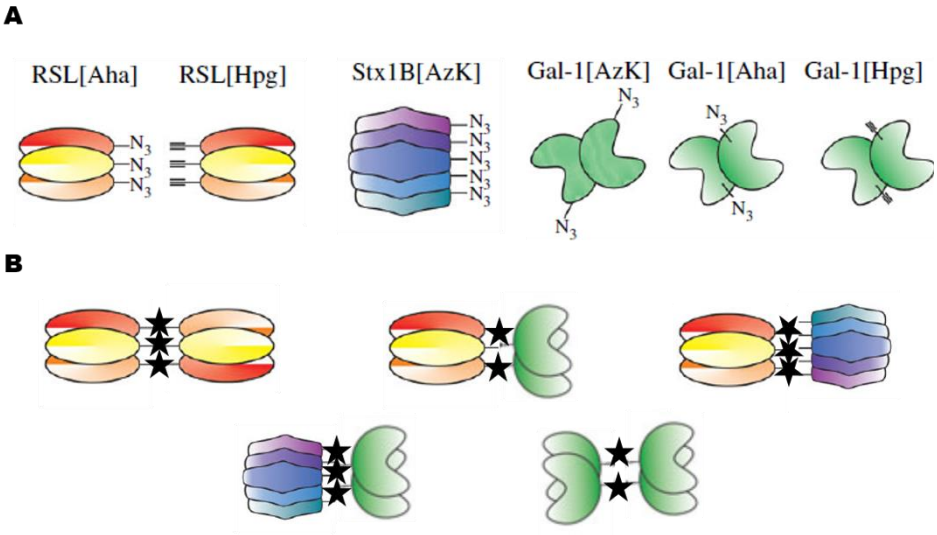


Figure 11: NcAAs as a clickable handle in lectin engineering. A) Schematic representation of lectins RSL, Stx1B, and Gal-1 with incorporated ncAAs. B) Formation of lectin's homo- and heteromers by direct or linker-mediated conjugation. Adapted from Tobola et al. (2019).

Another approach relies on a widely exploited strategy for tetramerization, i.e., conjugation between biotin and avidin or streptavidin. Due to their strong interaction, these binding partners found their applications in many biotechnological applications (Wilchek & Bayer, 1990). A variety of protein sites can be biotinylated by chemical conjugation, however, such modification is not site-specific. Nonetheless, by the addition of a consensus sequence of 13 AA to the protein terminus, site-specific biotinylation can be achieved (Powlesland et al., 2006; Schatz, 1993). Powlesland et al. (2006) by this approach created artificial tetrameric proteins composed of mouse SIGNR1 CRD (a homologue of DC-SIGNR) by conjugating four biotin-anchored lectin monomers with a tetrameric streptavidin. Thus, by precise insertion of biotin in a protein structure, synthetic constructs with increased valency can be prepared.

Continuing with this approach Achilli et al. (2020) created an artificial tetrameric lectin DC-SIGNR, a human C-type lectin receptor present on dendritic cells. DC-SIGNR CRD (carbohydrate recognition domain) naturally assembles as a tetramer through the extracellular domain (EDC) and thus engineering a synthetic construct mimicking the same valency of CRDs is of high interest. The so-called TETRALEC (artificial version of DC-SIGNR) was created in two steps. First, the N-terminus of DC-SIGNR CRD was extended by three glycine residues. This oligopeptide was recognized by the enzyme sortase A, SrtA, which anchored a biotinylated peptide LPRT-OMe at the N-terminus as previously described (Antos et al., 2009; Mao et al., 2004). The resulting biotinylated monomeric CRDs were clicked together into tetramers by NeutrAvidin, a tetrameric protein derived from avidin. The protein was tested for its specificity and binding properties while compared to the natural tetrameric DC-SIGNR and monomeric CRD. TETRALEC selectivity for the ligands is comparable with DC-SIGNR however the spatial organization of binding sites differs and might affect some of the interactions. Despite this fact, this strategy could be generally applied to other lectins, especially those which are monovalent, in order to create their tetrameric form and therefore possibly increase their affinity.

The strategy of enzymatic site-specific conjugation of biotin is of interest also as a tool for lectin functionalization, e.g., attachment of fluorophores (Kurhade et al., 2021).

Coiled-coil conjugations result in the increase of valency

Modification of lectin topology might be achieved by incorporation of peptides able to fold and oligomerize as coiled-coil domains. Coiled-coil is a structural motif, widely present in proteins, usually involving two or multiple alpha-helical organizations for various functional but also structural purposes. Therefore, the anchoring of such structures was used for the creation of higher protein assemblies. The helix is usually composed of the consensus pattern of seven AAs, referred also as heptad repeat, HPPHPPP, where H stands for hydrophobic and P for polar AA. Repetition of the heptad results in alpha-helix of various lengths and coiled-coil adopts parallel or anti-parallel organization (Truebestein & Leonard, 2016).

By this approach Fettis et al. (2019) engineered a chimera composed of Galectin-1 (G1) and Galectin-3 (G3) assembled as a dimer through a two-stranded α-helical coiled-coil. G1 was previously engineered as a fusion tandem repeat lectin G1/G1 (Fettis & Hudalla, 2018) however, a new chimera of G1/G3 was created by a fusion of G1 and G3 *via* peptide linker that forms α-helical coil and results in chimera dimerization. The so-called zipper combines 2 domains of G1 and two domains of G3 and therefore results in the increase of multivalency and dual functionality. The G1/G3 zipper chimera was compared with the G1 monomer, G1/G1 fusion tandem repeat, and G1/G3 monomer. Taken together, the results showed that multimeric organization improved the binding activity of chimera compared to other variants. Additionally, G1/G3 zipper induced T-cell death event at low concentrations and thus with further exploration of its efficacity might be considered an immunomodulatory protein therapeutic.

The introduction of coiled-coil structures might result also in the assembling of proteins into a higher oligomerization state (Ross et al. 2019). B-subunit of pentameric cholera toxin CTB was appended with α-helical coiled-coil (Fig. 12A) and structural studies demonstrated the formations of tubes composed of 12 pentamers where all CTB binding sites aligned at the external face of the tube and the coiled-coil clustered at the internal one (Fig, 12B, C). Each pentamer is connected with the neighboring one *via* three α-helices, from which two belong to one protein molecule and the third one to another. Although CTB naturally assembles as pentamer, the tube-like organization is driven by coiled-coil, and therefore such a strategy might be of interest for the formation of novel supramolecular structures.

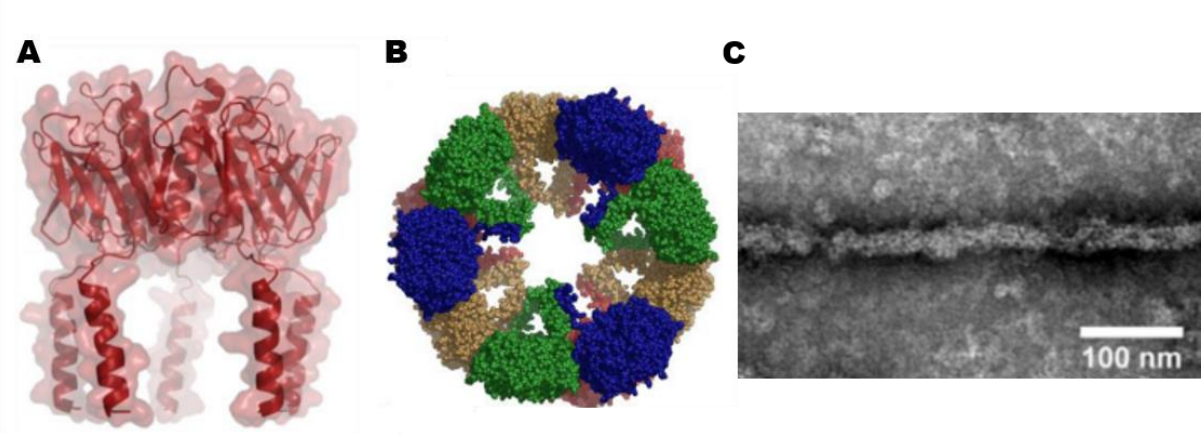


Figure 12: Addition of coiled-coil sequences to cholera toxin induces tube-like formation. A) Model of pentameric CTB appended with coiled-coils. B) Ring assembly observed in the crystal structure by packing of three asymmetric units, each containing four pentamers (green, blue, yellow, red) (PDB 6HSV). C) Tubes formed by stacking of ring assemblies. Adapted from (Ross et al., 2019).

Coiled-coil could be beneficial for engineered biomaterials as well. In the work of Ramberg et al. (2021), lectin RSL was fused with parallel trimeric coiled-coil heptad and subsequently cocrystallized with Q7, cucurbit[7]uril, a synthetic macrocyclic molecule also considered as suitable host molecules, i.e., for protein encapsulation. The newly created crystalline architecture organizes into layer-like structures where the protein part modulates the stiffness and porosity of such assembly. Thus, the incorporation of protein-like molecules into biomaterials might moderate their properties and find their applications as protein cargoes.

Fusion chimeras and the concept of neo-lectins

Lectin topology might be engineered also for obtaining higher symmetry than the one in the natural lectins (Smock et al., 2016; Terada et al., 2017; Voet et al., 2014; Yadid & Tawfik, 2011). β-propeller and β-trefoil structural folds are of high interest and their engineering could bring novel properties to the protein stability, functionality, etc., but also revealed the evolutional signature of these lectins. The importance of this topic is more described in our review (Notova et al., 2020) which is also included in chapter 1.4.

Terada et al. (2017) used computational protein design to create a highly symmetrical β-trefoil lectin. The synthetic protein is called Mitsuba (three-leaf in Japanese). Mytilec, a lectin discovered in mollusks that binds to α-galactose and possesses anti-cancer properties, was used

as a template. Naturally occurring Mytilec assembles as a dimer and its engineered monomeric version was unstable (Terada et al., 2016). Strikingly, the artificial Mitsuba showed a large increase in stability compared to monomeric Mytilec, but the cytotoxic properties were diminished.

On the contrary, a reverse approach was applied by Yadid & Tawfik, (2007) where Tachylectin from horseshoe crab *Tachypleus tridentatus*, originally a monomeric β-propeller composed of five blades, was artificially reconstructed from five individual blades of randomly fragmentized Tachylectin 2, resulting in a homo-oligomeric protein with the same binding properties as parental lectin. Supporting these studies, indeed, the creation of highly symmetrical *de novo* proteins might be a great strategy for the reconstruction of ancestral protein forms.

Almost 10 years ago, the concept of neo-lectins was developed in CERMAV CNRS by the team of Dr. Anne Imberty based on the β-propeller lectin RSL (Arnaud et al., 2014). Neo-lectins are defined as synthetic lectins with controlled valency. The template lectin, RSL from *Ralstonia solanacearum,* naturally occurs as a trimer with six binding sites for fucose (Fig. 13A). Upon binding to giant unilamellar vesicles (GUVs) decorated with fucosylated glycolipids, RSL induces structural and morphological changes on the membrane, i.e., invagination, resulting in tubule formation and as previously observed, by reducing the number of the binding sites from six to three, these properties diminished while the avidity stayed preserved (Arnaud et al., 2013). Therefore Arnaud et al. (2014) prepared a library of different RSL neo-lectins with modified valency ranging from six to zero binding sites. In the first step, the monovalent RSL was engineered on the gene level, i.e., three monomeric RSL connected with a linker (Fig. 13B). The new lectin was named neoRSL_VI and preserved all six binding sites for fucosylated ligands. The biophysical and structural characterization showed almost no differences in its activity and organization in comparison with parental RSL. Monomeric NeoRSL_VI was more accessible for further engineering and therefore the mutations were introduced into the binding sites. As shown in Figure 13C, 13 different variants were prepared. Surprisingly, all mutants with valency in the range from one to six (except the mutant with zero binding sites) maintained similar affinity towards the tested ligands in solution, and avidity was maintained as well (except for the mutants with zero and one binding sites). Neo-lectins were also tested with glycodecorated GUVs in order to compare their membrane behavior to parental RSL. Strikingly, the topology of the binding sites plays a crucial role in neo-lectin internalization. If the two binding sites are neighboring, the invagination is still observed (90 – 70% of efficiency if compared to RSL) despite only 2 binding sites being present. On the

other hand, if the binding sites are more distant, the effect of invagination is diminished even though 3 binding sites are present in neo-lectin (Fig. 13D). Altogether, these findings pointed to the fact that avidity and lectin topology are not always in direct correlation and thus multiple factors should be considered while engineering a neo-lectin with desired characteristics, i.e., cargo proteins for drug delivery.

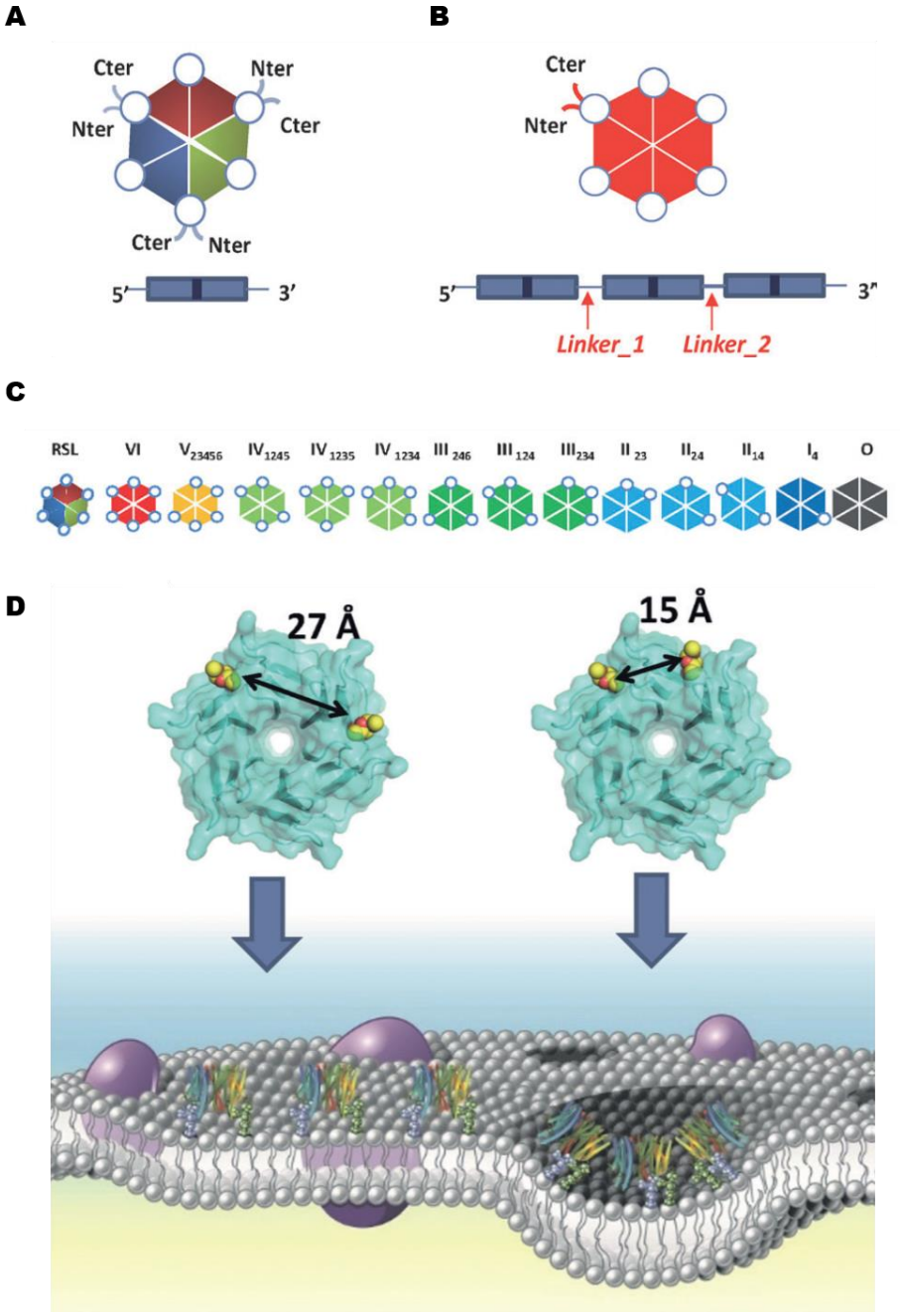


Figure 13: Neo-lectin RSL valency and its effect on avidity and membrane uptake. A) RSL lectin oligomerizes as a trimer and possesses 6 binding sites for fucose, B) NeoRSL_VI designed as monomeric RSL, C) A library of 13 different NeoRSLs with valency controlled from six to zero, D) The location of binding pockets is critical for RSL-induced membrane invaginations. Adapted from (Arnaud et al., 2014).

Fusion chimeras and the multivalency

The fusion chimeras can be created also with the purpose to increase lectin multivalency or even introduce novel specificity. Supplementing the previous approach of clickable lectins, Achilli et al. (2020) in the same study designed also dimeric CRD of DS-SIGN by fusing it with Fc antibody fragment. The fusion protein also showed similar behavior as synthetic and natural tetrameric CRD of DC-SIGNR. The same strategy was as well used with other lectins, e.g., Mincle and SIGNR3 (Lightfoot et al., 2015; Rabes et al., 2015), however, such constructs result only in divalent assembly.

Keeffe et al. (2011) designed oligomers of cyanovirin-n (CN-V) in order to increase the neutralization of viruses. CN-V is a mannose-specific lectin from cyanobacterium *Nostoc ellipsosporum* that displays antiviral properties to many enveloped viruses, including human immunodeficiency virus (HIV). CN-V is composed of two tandem-repeat domains, both containing one binding site for high mannose glycoconjugates (present in the envelope proteins, e.g., gp120), and even though two binding sites have different affinities toward ligand, they are both crucial for virus neutralization. CN-V is monomeric in solution, however, it crystallized as domain swapped dimer. Thus Keeffe et al. (2011) in their study created a set of fusion protein (two, three and four CN-V tandem repeats) with different linkers in order to compare their properties with natural CN-V. The "dimer" composed of two tandem repeats of CN-V crystallized with the same domain-swapping as observed for the wild type. The tandem repeat organization increased lectin multivalency in solution and therefore virus neutralization effect was significantly improved (similar potency to anti-HIV antibodies). Interestingly, the knockout studies showed the importance of the presence of two B-type binding sites in the virus neutralization. While the presence of two tandem repeat of CN-V showed a significant effect in HIV neutralization, the addition of a third or fourth lectin repeat did not increase its potency further. Although it is not confirmed, this could be conditioned by the position of binding sites in domain-swapped dimer which might allow better cross-linking of high-mannose glycoconjugates of gp120 while the addition of extra CN-V does not favor such a special orientation of lectin. Nonetheless, tandem repeat dimer CN-V proved to be a great candidate for HIV neutralization.

Another interesting approach how to engineered lectin was proposed by Irumagawa et al. (2022) where the multivalency was introduced by the creation of lectin nano-blocks, the artificial structures created by linking dimeric lectin ACG from *Agrocybe cylindracea* (specific

for β-galactose derivates, including T-antigen) and dimeric *de novo* protein WA20 (Arai et al., 2012). In their study, the synthetic lectin nano-blocks were constructed by fusion of ACG and WA20 with different linkers resulting in self-assembling into di-, tetra-, hexa-, octa- and decameric structures (Fig. 14). The complexes were characterized in terms of their size and binding properties, i.e., hemagglutination, and cell staining. The lectin nano-blocks have a preference to organize themselves into di- and tetramers, nonetheless other multimeric stages were detected as well. The novel structures showed the same binding properties as ACG alone however they showed higher affinity due to multimerization. Additionally, the representatives of di-, tetra-, and hexameric lectin nano-blocks were structurally characterized by small-angle x-ray scattering (SAXS) and the models of such assemblies were built. Indeed, Irumagawa et al. developed an interesting strategy on how to increase the valency and induce avidity, and thus this method could be applied to a variety of lectins.

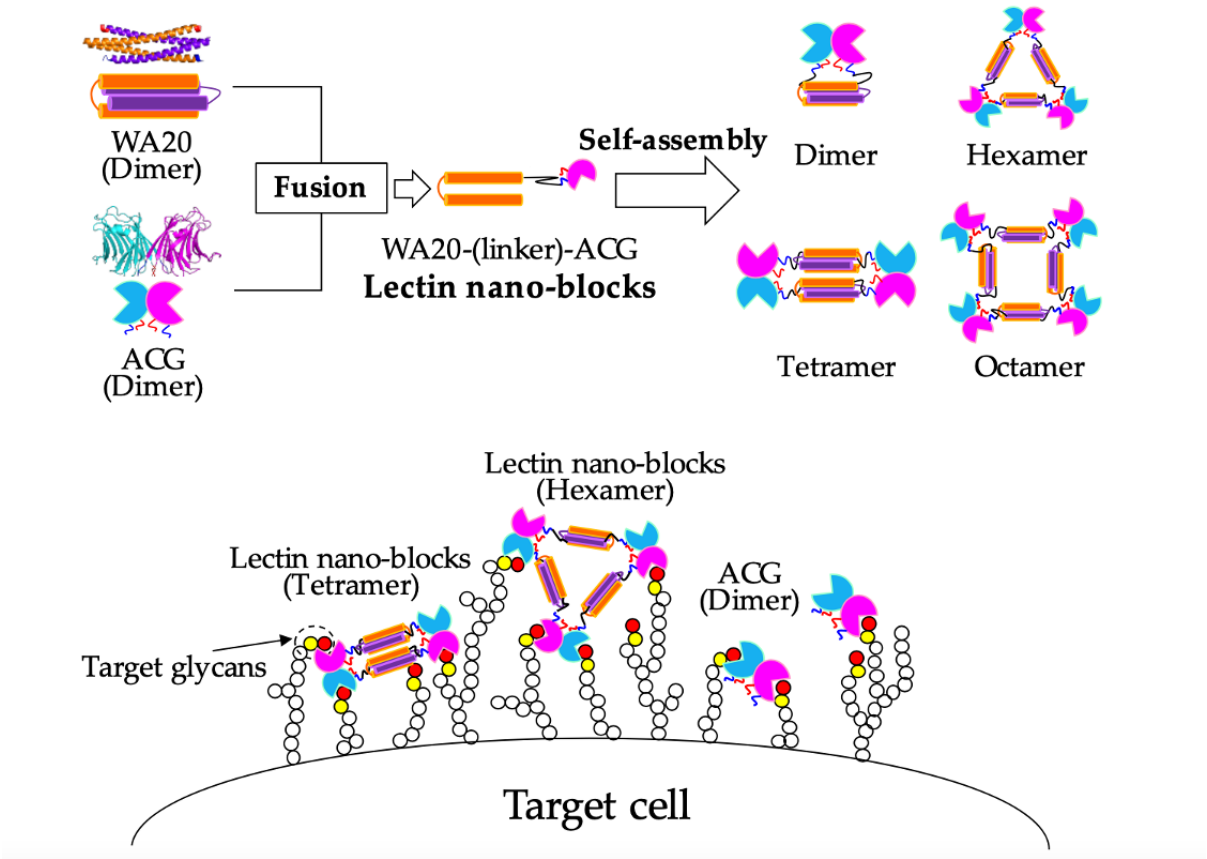


Figure 14: The lectin nano-blocks as candidates for cell labeling. These synthetic proteins were created as a fusion of dimeric lectin ACG and dimeric *de novo* protein WA20. The fusion proteins spontaneously assemble into multimers and due to multivalency the affinity to their binding partners was significantly increased. Adapted from Irumagawa et al. (2022)**.**

Nonetheless, not only lectins but also CBMs are an attractive target for protein engineering. Connaris et al., (2009) enhanced the affinity of CBM40 (NanH) from *Vibrio cholerae* to α-2,3, α-2,6- and α-2,8-linked sialosides by engineering tandem repeats of CBM40, i.e., monoCBM40, diCBM40, triCBM40 and tetraCBM40. TetraCBM40 displayed an increase of affinity 700-1500-fold to α-2,3-siallylactose when compared to native monomeric CBM40. Continuing with this approach Ribeiro et al., (2016) designed a divalent form of CBM40 (NanI) from *Clostridium perfringens* which resulted in a 10-fold increase of binding affinity towards sialylated oligosaccharides. Both examples support the idea that CBMs are suitable candidates for protein engineering and the creation of multimeric modules with higher affinity. Due to the fact that CBMs are often part of enzymes recognizing sugars, their exploration and handling bring novel opportunities in the field of biological probes. We have also followed this approach and there is a detailed description of the design, production, and characterization of a divalent form of $CBM77_{Rf}$ from *Ruminococcus flavefaciens* in chapter 5.

Janus lectins – bispecific fusion chimeras with increased valency

The above mentioned lectin RSL was also used for another type of synthetic biology approach for the creation of so-called Janus lectin (Ribeiro et al., 2018). The name is derived from the Roman god Janus who had two faces and therefore this concept implies the idea that two lectins, their domains or lectin, and CBM are rationally designed as a bispecific chimera. The *de novo* protein is engineered on the gene level and subsequently produced in a suitable expression system. So far, only one Janus lectin has been described. Ribeiro et al. (2018) designed and created RSL-CBM40, a Janus lectin composed of a β-propeller lectin RSL associated with CBM40 from *Clostridium perfringens* (Fig. 15A). Both proteins showed relatively strong affinity to their ligands and since they have been previously engineered (Arnaud et al., 2014; Ribeiro et al., 2016) they proved to be perfect candidates for such an approach. After assembling of gene and production of protein, the protein is oligomerized into a trimeric structure with six binding sites for fucose on the one site/face and three binding sites for sialic acid on the other one (Fig. 15A).

For the RSL part, the expected binding properties were observed. However, the introduction of multivalency of CBM40 resulted in 20 to 200 times stronger affinity to its binding partners if compared to its dimeric and monomeric forms respectively. The ability of RSL-CBM40 to bind its ligands at the same time was tested as well and due to its opposite orientation of binding

interfaces, Janus lectin happens to be a perfect candidate for building materials layer by layer. Additionally, the protein layer appears to be stiffer and therefore introduces novel properties to the material. Its capability of binding two ligands simultaneously was tested as well in the presence of two different types of GUVs decorated with fucosylated and sialylated glycolipids respectively, which resulted in their cross-linking.

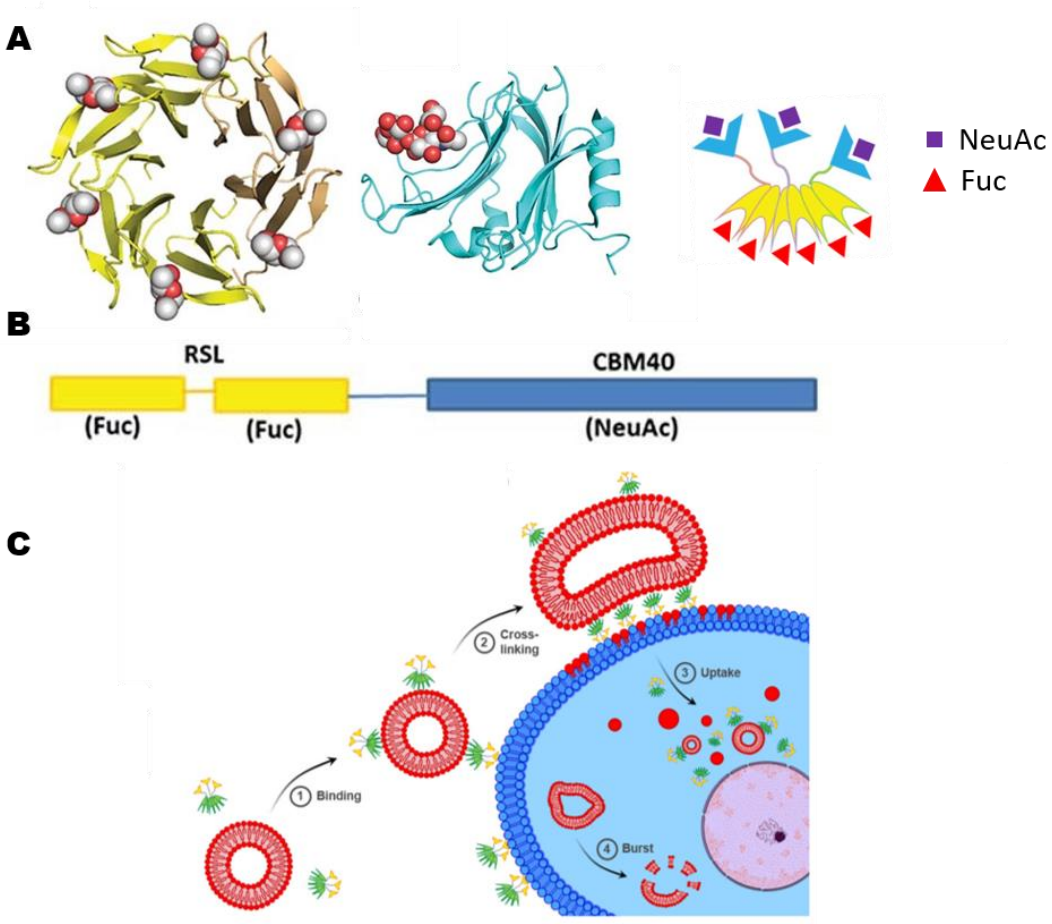


Figure 15: Janus lectin RSL-CBM40. A) Crystal structures of trimeric RSL (yellow ribbon) complex with six fucose (PDB 2BT9) and CBM40 (cyan) complex with 3'-siallylactose (PDB 5FRE) and schematic representation of Janus lectin RSL-CBM40 in complex with six fucose and three sialic acid ligands, B) The gene sequence *rsl-cbm40 was* designed as a fusion of RSL monomer at N-terminus linked together with monomeric CBM40 at C-terminus, C) RSL-CBM40 is able to cross-link between fucose-decorated GUVs and cancer cells overexpressing sialic acid, resulting in internalization of vesicles. Adapted from (Ribeiro et al., 2018; Siukstaite et al., 2021).

The rational design of Janus lectin to bind sialic acid has the potential for the detection of hypersialylation, a physiological defect occurring in the numerous cancer types, including lung, breast, ovarian, pancreatic, and prostate cancer (Dobie & Skropeta, 2021). In the framework of

the SynBIOCarb network, we have published a scientific article where Siukstaite et al. (2021) used the Janus lectin RSL-CBM40 to demonstrate that it can crosslink fucose-decorated GUVs with sialic-rich H1299 cancer cells. Strikingly, the intact liposome uptake into the cell-mediated by lectin was observed suggesting that bispecific lectins could find their application in targeted drug delivery systems (Fig. 15C). This publication is attached to the thesis and can be found in the annex.

By Janus lectin strategy, novel topology but also specificity is introduced with the potential application in plant cell wall engineering, as closely describe in chapter 5 or cell labeling in chapter 6.

Fusion chimeras with additional properties

Fusion of an additional domain to lectins can bring additional properties, such as toxicity or activation of the immune system, that could find application, especially in the medical field.

Such fusion was built based on lectin BC2LC from *Burkholderia cenocepacia* with narrow and exclusive specificity for human pluripotent stem cells (hPSCs). Moreover, once bound to the cell, its internalization is induced. This selectivity is of special interest in the field of organ transplantation especially when the organs are cultured from hPSCs since the residual presence of original cells can lead to cancer development. Tateno et al. (2015) engineered a lectin-toxin fusion protein with the ability to eliminate hPSCs from cell cultures. The lectin was fused with exotoxin A from *Pseudomonas aeruginosa* and showed targeted removal of induced and embryonic hPSCs with a concentration of 10 μg/mL. Additionally, Tateno & Saito (2017) engineered a more potent variant by fusion of previously described protein with two extra domains of exotoxin A. The lectin-toxin chimera was again produced in *E. coli* but this time the protein showed 556-fold higher toxicity against induced hPSCs. Thus, the creation of such chimeras has great potential in cell-cultured organ production or regenerative medicine.

Another noteworthy strategy is focused on the creation of “lectibody” – a chimeric protein composed of lectin and Fc fragment of IgG antibody. The lectin, dimerized through the Fc structure, is selected for its specificity towards glycan present on the target, either cancer cells, or pathogens. Once the lectin is bound, the Fc part activates the immune system *via* cell receptors or complement proteins. Such lectibody was prepared by Hamorsky et al. (2019) as a fusion of Fc fragment with actinohivin lectin variant (Avaren), called AvFc. The lectibody is

specific for high-mannose-type glycans present on the envelope glycoproteins (gp120) of the HIV virus. Actinohivin by itself has already proved anti-HIV properties, making it a suitable drug candidate against the HIV virus. However, the lectin is prone to aggregate, and therefore an engineered version, Avaren, was created and used for lectibody. The fusion protein was produced in the plant expression system and due to dimerization showed higher affinity to its ligands, inhibited several HIV viruses, and even induced Fc-mediated cytotoxicity (Fig. 16). Thus, AvFc might be considered a novel and promising candidate for HIV targeting.

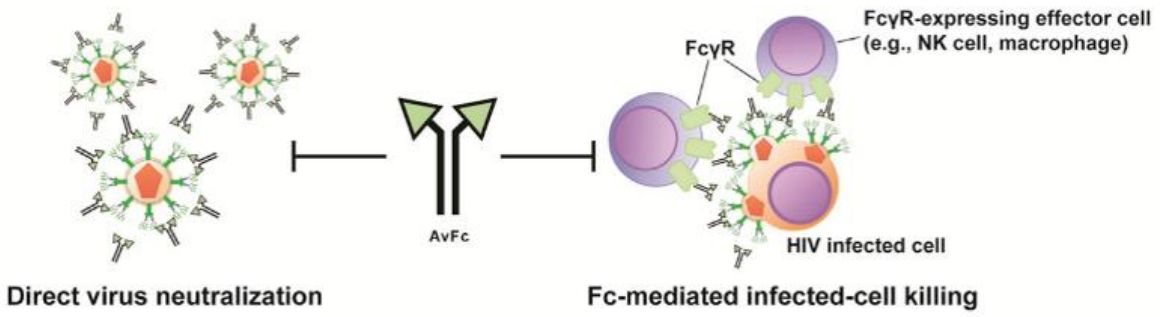


Figure 16: Schematic representation of Lectibody AvFc. The fusion protein has two mode of actions, either virus neutralization through lectin domain or Fc mediated cytotoxicity.

Encouraged by the success, Oh et al. (2022) continued with AvFc molecule but this time tested its activity against hyper-mannosylated cell surfaces. Hyper-mannosylation, as a pathological event, was detected in several cancers, i.e., breast, colorectal, pancreatic, ovarian, prostate, lung, and others. Thus, AvFc was tested with several cancerous cell lines and showed that it is able to bind these cells due to the high-mannose glycans present on the cell surface. Subsequently, AvFc inhibits the activation of receptors involved in cell growth and the Fc part of the lectibody elicits the ADCC (antibody-dependent cell-mediated cytotoxicity) activity. Altogether, AvFc is probably the first antitumor agent targeting cancer-associated high-mannose glycans and the lectibody strategy has the potential to be developed in numerous fields and find its applications in drug development and targeting.

1.4 Review article: Structure and engineering of tandem repeat lectins

During the first months of my Ph.D. studies, we got an invitation to write a short informative review about lectin engineering. The review was published in the journal Current Opinion in Structural Biology, 2020, 62, 39-47 with the title Structure and engineering of tandem repeat lectins. The article focuses on the importance of multivalency in lectin/ligand interactions, explains where the protein multivalency is coming from, and later describes the latest achievements in the field of lectin engineering, especially in β-propeller and β-trefoil fold of lectins. I contributed to this review mostly by β-trefoil fold section, i.e., text and figures, and general manuscript preparation and inclusion of corrections.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Structure and engineering of tandem repeat lectins

Simona Notova[1,†], François Bonnardel[1,2,3,†], Frédérique Lisacek[2,3,4], Annabelle Varrot[1] and Anne Imberty[1]

Through their ability to bind complex glycoconjugates, lectins have unique specificity and potential for biomedical and biotechnological applications. In particular, lectins with short repeated peptides forming carbohydrate-binding domains are not only of high interest for understanding protein evolution but can also be used as scaffold for engineering novel receptors. Synthetic glycobiology now provides the tools for engineering the specificity of lectins as well as their structure, multivalency and topologies. This review focuses on the structure and diversity of two families of tandem-repeat lectins, that is, β-trefoils and β-propellers, demonstrated as the most promising scaffold for engineering novel lectins.

Addresses
[1] Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France
[2] SIB Swiss Institute of Bioinformatics, CH-1227 Geneva, Switzerland
[3] Computer Science Department, UniGe, CH-1227 Geneva, Switzerland
[4] Section of Biology, UniGe, CH-1205 Geneva, Switzerland

Corresponding author: Imberty, Anne (anne.imberty@cermav.cnrs.fr)
[†] Both first authors contributed equally to this work.

Current Opinion in Structural Biology 2020, 62:39–47

This review comes from a themed issue on **Carbohydrates**

Edited by **Sony Malhotra** and **Paul A Ramsland**

For a complete overview see the Issue and the Editorial

Available online 13th December 2019

https://doi.org/10.1016/j.sbi.2019.11.006

Introduction

Lectins are ubiquitous protein receptors able to bind complex oligosaccharides. They display a large panel of biological functions and many lectins are useful in biotechnology and biomedicine applications. Lectins are generally multivalent, that is, they contain several binding sites for the simultaneous attachment to several target carbohydrates. While multivalency is usually associated with strong avidity for glycoconjugates, the topology of the binding sites in space is also of importance for generating high specificity, for example, towards the branches of *N*-glycans and for biological function. Recently, it was demonstrated that specific arrangements of lectin binding sites favor glycolipid clusters and therefore affect the structure and dynamics of the cell membranes [1].

The multivalent properties of proteins relate to their highly symmetrical shapes. Most of the time, the symmetry of a protein structure stems from the oligomerisation of identical protomers. In oligomeric lectins, the carbohydrate-binding sites are generally spread in different directions resulting in structures with dihedral symmetry (D2 to D8) or tetrahedral one (Figure 1). In contrast, the cyclic symmetry Cn generates structures where all binding sites are on the same side, therefore perfectly fitted to bind glycosylated surfaces, such as cell membranes. Few oligomeric cyclic lectins are observed and they adopt C2 to C7 symmetry (Figure 1). Other lectins generate multivalency yet through a different strategy, using tandem repeats of a single motif to create cyclic and dihedral symmetries. In such case, the evolutionary origin of structure symmetry has been demonstrated to occur through gene duplication and fusion events [2].

The Lectin3D database [3] that gathers all known 3D-structures of lectins has been recently updated and remodeled as Unilectin3D [4••]. It contains more than 2000 structures in over 100 lectin families (i.e. unrelated sequences). While most are oligomeric, several different lectin families contain tandem repeats. A two-fold repeat with C2 symmetry is observed in some members of the β-prism I (e.g. banana lectin, griffithsin), cyanovirin and few other families. Tandem repeats with C3 cyclic symmetry are observed in the β-prism II (e.g. piocyn, monocot lectins) and β-trefoil families and is expected to occur in many lectins with the novel β-prism III fold [5]. Higher cyclic symmetry (Cn) is the hallmark of β-propellers lectins. Combination of tandem repeat and oligomerization results in even higher valency (Figure 1).

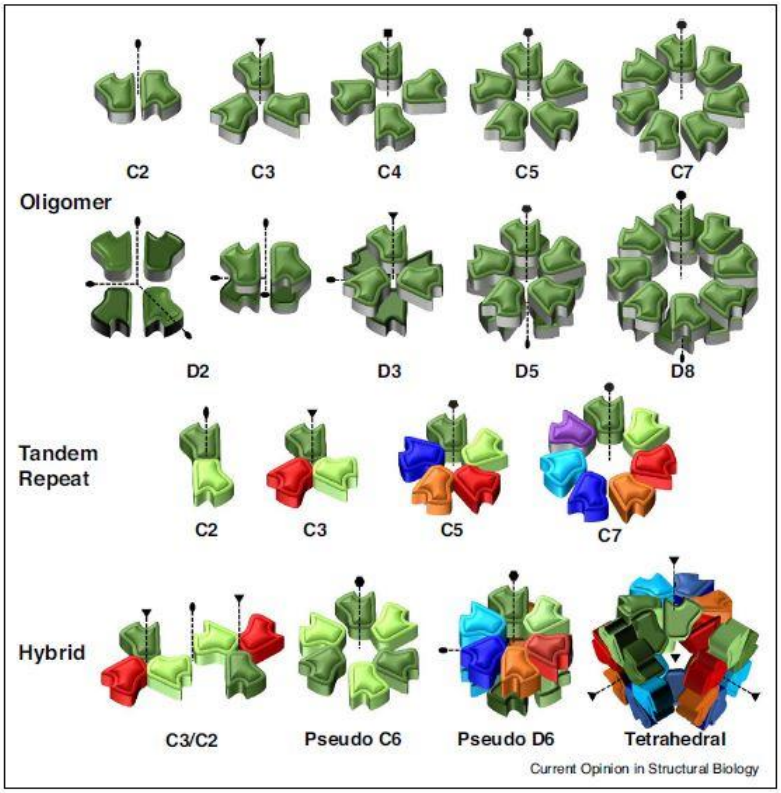
Tandem repeat lectins and their symmetry are of high interest for understanding protein evolution but also for the design of 'neo-lectins' with engineered structures and binding sites. In the emerging domain of synthetic glycobiology [6], engineering lectins provide powerful tools for deciphering the glycocode, but can also be applied in biotechnology for quality control of recombinant therapeutic glycoproteins, or in clinics as biomarkers for histopathology. Engineering strategies have been developed with some success [7,8••]. In this review, we focus on the two families of tandem repeat lectins that have been demonstrated to be the most promising scaffold for engineering novel lectins, that is, β-propellers and β-trefoils.

β-Propeller lectins

β-Propellers lectins: large panel of structures

The β-propeller fold is widely spread and results in proteins with four to ten repeats of four-stranded

Figure 1

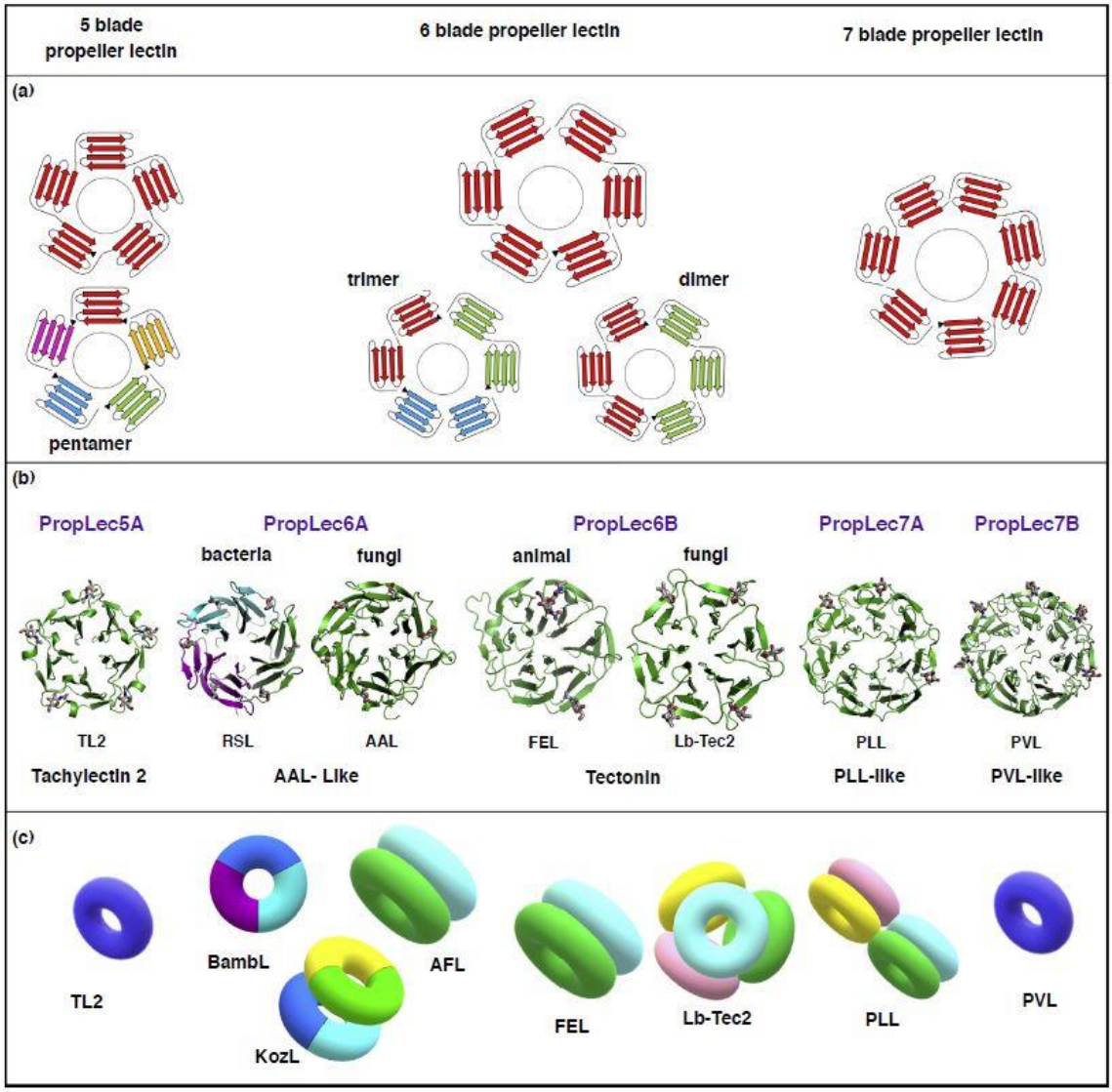


Schematic representation of different types of symmetry observed in oligomeric lectins, tandem-repeat lectins, and hybrid types. Lectin domain with identical peptide sequences are represented by identical color. The symmetry family is indicated below each schematic representation and symmetry axis (or pseudo-symmetry axis) are represented by dotted lines.

antiparallel β-sheets (i.e. blades of the propeller) assembled in a donut shape [9–11]. Members of the β-propeller family have diverse functions, ranging from signalling to enzymatic activity, and they are characterized by their structural rigidity. The known β-propeller lectins have been recently collected, compared and grouped into five classes based on the HMM signature of their blades [12••]. They are characterized by the positioning of five to seven carbohydrate binding sites on the same side of the donut, resulting in very efficient binding to glycoconjugates (Figure 2). β-propeller lectins are present in some bacterial and fungal pathogens and involved in infection by targeting glycans on host tissue, such as AFL/FleA in *Aspergillus fumigatus* that binds fucose and triggers a response from the host [13,14]. They are also involved in the innate immunity of several mushrooms and invertebrates and act by recognizing PAMPs (*pathogen-associated* molecular patterns) such as *N*-acetylated or methylated carbohydrates [15••]. Variations also occur at the supramolecular level: the donut-shaped β-propeller frequently associates back-to-back in solution, therefore presenting the carbohydrate-binding sites on opposite faces. Tetramer assemblies of such toruses can result in box-like shapes as in the lectin of *Photorhabdus luminescens* (PLL) [16] or virus-like shapes as in Tectonin 2 from *Laccaria bicolor* [15••].

The PropLec6A class of β-propellers contains fungal and bacterial proteins and is rather unique with respect to the blade assembly that forms the propeller. While all fungal structures display the classical 6-repeat architecture, the bacterial ones are much shorter and their sequences contain only two or three blades. The β-propeller of lectins of *Ralstonia solanacearum* (RSL) and *Burkholderia ambifaria* (BambL) result from the trimerization of short 90 amino acid-long peptides comprising two repeats [17,18] whereas the one found in *Kordia zhangzhouensis* (KozL) assembles as a dimer of longer peptides comprising three repeats [12••]. Most recent studies have concluded that all β-propellers have a common monophyletic

Figure 2



Representation of selected β-propeller lectins. **(a)** Schematic representation of the different folds and assemblies observed in lectins. Each blade is represented as a bundle of 4 β-strands. **(b)** Structure of representative lectins in each family with carbohydrate ligands. Proteins chains are represented as ribbons and carbohydrate ligands are represented as sticks. **(c)** Different types of functional assemblies of the β-propeller lectins. Each donut shape represents one propeller structure. In all panels, different peptides are represented by different colors.

origin, in spite of the different families stemming from independent amplification and diversification pathways from an ancestral peptide [19]. The β-prism-II proteins, which include several lectin classes, have been proposed to also share this ancestral peptide [20]. Interestingly, the only oligomeric β-propellers, that is, possible fossil forms, identified so far are the microbial PropLec6A lectins (RSL, BambL and KozL). It has been suggested that they represent less stable evolutionary intermediates that have been replaced by fully amplified peptides during evolution [19].

β-Propeller lectins are robust proteins, easy to produce in bacterial fermentation and to handle in the lab, which explains their wide use in biotechnology. Fucose-specific

lectins from *Aleuria aurantia* (AAL) and *Aspergillus orizae* (AAO) are available from companies and included in lectin arrays [21]. The specificity for terminal GlcNAc of the *Psathyrella velutina* lectin (PVL) can be used for labeling some cancer cells and tissues [22] and both PVL and its homolog in *Agrocybe aegerita* (AANL) are very interesting tools for the glycoproteomics of *O*-GlcNAcylation [23,24•].

β-Propellers: scaffolds for engineering symmetry and multivalency

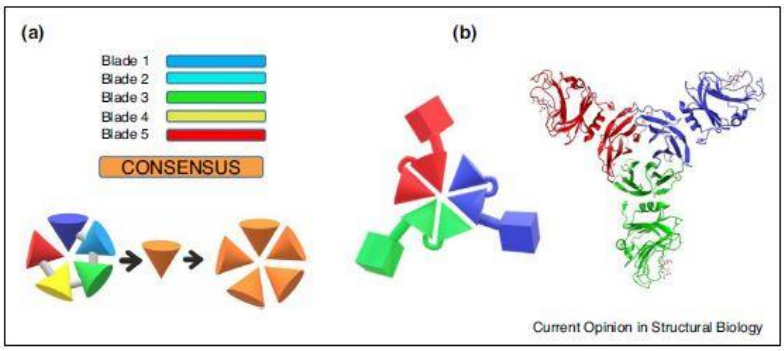
The β-propeller fold has been an excellent scaffold for understanding and reproducing Evolution and several groups demonstrated the possibility to reconstruct symmetrical proteins. The Pizza protein was computationally designed and produced as a perfectly symmetrical sixfold propeller [25], followed by the Tako and Ika peptides that self-assemble as symmetrical eightfold propellers [26]. Tachylectin-2, the only propeller lectin with five blades, has been also used due to the remarkable conservation of amino acids among the blades. Genetic library screening by phage display resulted in the identification of 2-blade fragments that could assemble as new lectins with sugar-binding activity [27]. A novel approach was proposed through the alignment of Tachylectin-2 and *Nematostella vectensis* lectin sequences and the reconstruction of an ancestor blade, a 43 amino acid sequence that self-assembles in a 5-blade lectin able to efficiently bind glycoconjugates (Figure 3) [28••].

Directed mutagenesis has been used for lowering the number of binding sites in β-propeller lectins, particularly in the LecB6 class. An AAL variant produced by site-directed mutagenesis to present only one functional fucose-binding site displayed a decrease in agglutination activity [29]. Furthermore, β-propeller lectins are powerful tools for studying membrane dynamics as they can cluster glycan heads and invaginate the membrane. Decreasing the valency of the RSL lectin from six to three binding sites resulted in the loss of capacity to invaginate the cell membrane of giant vesicles decorated with Lewis^a-pseudolipids [30]. The construction of a library of neolectins with systematic variations of the number and localization of binding sites demonstrated the crucial importance of distances between the binding sites, that is, the effect of the lectin topology on the dynamics of targeted glycolipids [1].

The LecB6 family displays strong affinity for fucose, partly due to the interaction between aromatic amino acids (Trp or Tyr) stacking with the hydrophobic face of the monosaccharides. In the AAL fungal lectin, only five of the six binding sites bind fucose, and different occupancy of oligosaccharides indicates differences among the five sites. Alanine-mutation mapping on Trp and Tyr residues established that sites 1, 2 and 4 are more efficient for binding [31]. Synthetic biology was used for introducing fluorinated indol analogs of the tryptophan residues at different positions in RSL. Introduction of non-canonical amino acids resulted in lectin variants with altered specificity towards fucosylated blood group oligosaccharides [32••].

Mutations of other amino acids in the binding sites were also explored and AAL mutants displaying altered binding affinities to different fucosylated oligosaccharides were obtained [33]. The Q224N mutant has higher affinity for core fucosylated *N*-glycan and is of special interest as biomarker for hepatocellular carcinoma [34•]. Chemical functionalization of RSL generated conjugates with 3, 6, or 12 polyethyleneglycol molecules (PEG) that displayed altered carbohydrate-binding properties [35]. The RSL peptide has also been used for the development of novel thermo-responsive affinity sugar binders by fusion with

Figure 3



Engineering of propellers wit schematic representation of the associated strategy. **(a)** neo-lectin obtained by construction of a consensus peptide that self-assembles in a 5-blade lectin [28••]. **(b)** Fusion of a two-blade fucose binding peptide and a sialic acid binding peptide in a chimera peptide that assembles as Janus 6-blade lectin [37••]. In all panels, different peptides are represented by different colors.

elastin-like polypeptides [36]. The fusion of the RSL peptide with a sialic acid binding domain of another bacterium resulted in the first Janus-lectin, able to bind six fucoses on one face and three sialic acids on the other face. Such supramolecular assembly was demonstrated to form an organized layer with bivalent glyco-compounds and to associate glyco-vesicles in proto-tissues [37••].

β-Trefoil lectins

β-Trefoil proteins are composed of a three-lobed architecture consisting of peptide repeats forming a barrel structure at one extremity and a triangular arrangement of hairpins at the other [38]. β-Trefoil lectins are largely spread in bacteria, fungi, plants and animals and often attached to other proteins with toxic or enzymatic activity. They are classified as family 13 of the carbohydrate-binding modules when attached to carbohydrate active enzymes [39]. While very conserved in shape, the sequence conservation is very weak and only a few hydrophobic residues are conserved in the core. This involves the QxW motif that is repeated in each foil in most classes of β-trefoil lectins [40], albeit with some exceptions. The threefold symmetry can be degenerated and many β-trefoil lectins present only one or two carbohydrate active binding sites (Figure 4). While originally described as specific for galactose, β-trefoil lectins have been evidenced to bind a variety of sugars, including rhamnose in the structure of a recently solved protozoan lectin [41].

β-Trefoil lectin: structural plasticity for multiple function

The function of β-trefoil lectins can vary depending on the organism and ligands but many of them serve as a recognition module for an attached toxin domain, as for ricin-B, the first structure to be determined [42]. The β-trefoil domains often play a role in innate immunity as demonstrated by CCL2 from the *Coprinopsis cinerea* fungus which is toxic towards nematodes. The dimeric lectin acts by binding a fucosylated epitope present in the worm digestive system [43••]. β-Trefoil lectins are also useful in biotechnology. Some have fine specificity that can be used for glycan sorting and labeling, such as the *Marasmius oreades* mushroom MOA lectin recognizing specifically α-galactose in blood group B epitope [44]. Actinohivin in Actinomycetes is a broadly neutralizing lectin against HIV-1 infection that targets the gp120 envelope glycoprotein [45]. Indeed, the crystal structure of the lectin confirms the presence of oligomannose in the three bindings sites of the peptide [46].

Lectins from marine invertebrates are of special interest due to their role in the protection against pathogens. Mytilec from the mussel *Mytilus galloprovincialis* has antibacterial and antifungal functions. Both Mytilec and the related *Crenomytilus grayanus* CGL lectin bind with high avidity the αGal1-4Gal epitope present on globotriaosyl ceramide in cancer cells. Both lectins have anti-cancer properties since this binding downregulates cell growth and leads to apoptosis [47•,48•].

Engineering β-trefoil structures and specificity

The plasticity of β-trefoil lectins makes them a convenient tool to study evolution. The construction of artificial lectins, mimicking their evolutionary pathways, was attempted for both plant and invertebrate proteins. A consensus sequence was designed based on the lactose-binding ricin B. The sequence was repeated to obtain the three-fold globular protein named ThreeFoil [49]. The resulting structure is almost perfectly symmetrical and exhibits high thermal stability. Interestingly, the carbohydrate ligand, that is, lactose, acts as a chaperone and helps in the folding of the artificial lectin [50•].

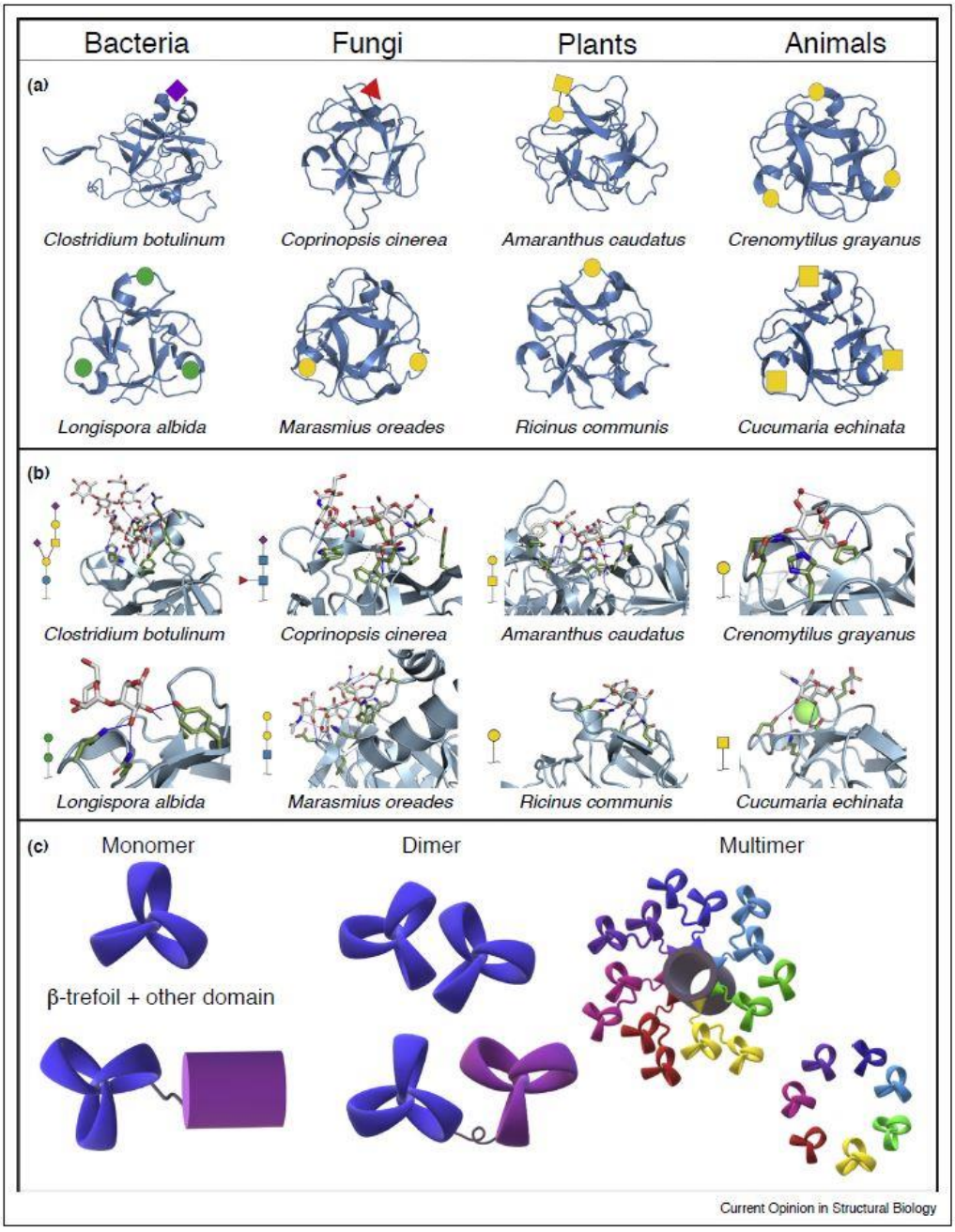
MytiLec is also an excellent candidate for engineering artificial lectins thanks to its high internal similarity and the presence of one active sugar-binding site in each lobe of the trefoil structure. In its native form, Mytilec consists of a tight homodimer and its specificity towards cancer antigen is of interest in biotechnology. The substitution of apolar residues by polar ones at the dimerization interface, resulted in a monomeric mutant, albeit with low stability [48•]. The loss of cytotoxic activity demonstrated that dimerization is crucial for cellular impact, as confirmed for the toxicity of CGL2 [43••]. The ancestral reconstruction method was used to create a threefold perfectly symmetrical protein derived from Mytilec and called Mitsuba (three-leaf in Japanese) [51••]. Mitsuba interacts with cells which express globotriose on their surfaces and shows good stability but only weak anticancer activity.

Yabe *et al.* demonstrated the process of natural evolution mimicry by changing the specificity of earthworm EW29Ch lectin from galactose to sialic acid. The lectin was engineered by error-prone PCR with a reinforced ribosome display system and the original lactose-binding pocket was extended to adapt sialylation on the galactose ligand. With this approach, the authors created the first sialic acid specific R-type lectin [52]. In 2012, this method was further developed by Hu *et al.* [53], who improved the process by introducing a high-throughput screening system with glycoconjugate microarrays. The engineered mutant is the first receptor specific for 6-sulfo-galactose terminated glycans, and is a potential marker for various diseases. Further analysis showed that the change of specificity was reached just by a single E20K point mutation.

Conclusion and discussion

Tandem-repeat lectins originate from an evolution strategy based on gene duplication resulting in the multiplication of binding sites with ideal topology for binding to substrates. The identification of repeated short peptides with carbohydrate-binding activities is difficult and it is

Figure 4



Representation of selected β-trefoil lectins. **(a)** Variations of structure and specificity observed in β-trefoil lectins for bacteria (botulin toxin and actinohivin), fungi (CCL2 and MOA, plants (amaranthus agglutinin and ricin) and animals (mussel CGL and sea cucumber CEL-III). Ligands are represented using the Symbol Nomenclature for Glycans [57]. **(b)** Sugar binding sites of the same lectins as in section A with hydrogen bonds as blue lines, water molecule and calcium ions as red and green balls, respectively. **(c)** Schematic representation of different architecture associated with β-trefoil lectins, including monomer, covalent or non-covalent dimers, and multimer (hemolytic lectin CEL-III).

likely that novel lectin families are yet to be identified in Nature. For the known folds, such as β-propellers, appropriate bioinformatics tools turned out essential for efficient search in genomic databases [12••].

Tandem-repeat lectins such as β-propellers and β-trefoils are excellent tools as bio-markers due to their strong avidity for glycoconjugates presented on cell membranes. Engineering strategies, from single point mutation to incorporation of non-canonical amino acids, or directed evolution, are now available for designing and controlling the specificity and the topology of such lectins. Creating such multivalent receptors through artificially constructed tandem repeats is a route of interest. It generally strongly enhances the receptor affinity for its ligand [54,55]. It has also been demonstrated that expression as tandem-repeat sequences improves the production of recombinant algal lectins in bacterial systems [56]. This is a definite option for the future development of artificial lectins, molecules that will have many applications in health science and biotechnology.

Conflict of interest statement

Nothing declared.

Acknowledgements

The authors acknowledge support by the ANR PIA Glyco@Alps (ANR-15-IDEX-02), the Alliance Campus Rhodanien Co-funds (http://campusrhodanien.unige-cofunds.ch) and the H2020 Marie Skłodowska-Curie Action n° 814029 (synBIOcarb).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Arnaud J, Tröndle K, Claudinon J, Audfray A, Varrot A, Römer W, Imberty A: **Membrane deformation by neolectins with engineered glycolipid binding sites**. *Angew Chem Int Ed Engl* 2014, **53**:9267-9270.

2. Blaber M, Lee J, Longo L: **Emergence of symmetric protein architecture from a simple peptide motif: evolutionary models**. *Cell Mol Life Sci* 2012, **69**:3999-4006.

3. Pérez S, Rivet A, Imberty A: **3D-lectin database**. In *Glycoscience: Biology and Medicine*. Edited by Taniguchi N, Hart GW, Seeberger P, Wong CH. Springer; 2015:283-289.

4. •• Bonnardel F, Mariethoz J, Salentin S, Robin X, Schroeder M, Pérez S, Lisacek F, Imberty A: **UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands**. *Nucleic Acids Res* 2019, **47**: D1236-D1244
Description of the database of 2000 structures of lectins classified on origin and fold, with cross-links to literature, other databases in glycosciences and functional data such as known specificity.

5. Cabanettes A, Perkams L, Spies C, Unverzagt C, Varrot A: **Recognition of complex core-fucosylated N-glycans by a mini lectin**. *Angew Chem Int Ed Engl* 2018, **57**:10178-10181.

6. Turnbull WB, Imberty A, Blixt O: **Synthetic glycobiology**. *Interface Focus* 2019, **9** 20190004.

7. Hu D, Tateno H, Hirabayashi J: **Lectin engineering, a molecular evolutionary approach to expanding the lectin utilities**. *Molecules* 2015, **20**:7637-7656.

8. •• Hirabayashi J, Arai R: **Lectin engineering: the possible and the actual**. *Interface Focus* 2019, **9**:20180068
Up-to-date review on strategies for engineering lectin specificities. It also discusses possible approaches to confer sugar-binding properties on synthetic proteins and peptides.

9. Chen CK, Chan NL, Wang AH: **The many blades of the beta-propeller proteins: conserved but versatile**. *Trends Biochem Sci* 2011, **36**:553-561.

10. Fülöp V, Jones DT: **β Propellers: structural rigidity and functional diversity**. *Curr Opin Struct Biol* 1999, **9**:715-721.

11. Jawad Z, Paoli M: **Novel sequences propel familiar folds**. *Structure* 2002, **10**:447-454.

12. •• Bonnardel F, Kumar A, Wimmerova M, Lahmann M, Perez S, Varrot A, Lisacek F, Imberty A: **Architecture and evolution of blade assembly in β-propeller lectins**. *Structure* 2019, **27**:764-775
Bioinformatics approach for identification of β-propeller lectins in genomes that reveals their widespread distribution of across species.

13. Houser J, Komarek J, Kostlanova N, Cioci G, Varrot A, Kerr SC, Lahmann M, Balloy V, Fahy JV, Chignard M *et al.*: **A soluble fucose-specific lectin from *Aspergillus fumigatus* conidia - structure, specificity and possible role in fungal pathogenicity**. *PLoS One* 2013, **8**:e83077.

14. Richard N, Marti L, Varrot A, Guillot L, Guitard J, Hennequin C, Imberty A, Corvol H, Chignard M, Balloy V: **Human bronchial epithelial cells inhibit *Aspergillus fumigatus* germination of extracellular conidia via FleA recognition**. *Sci Rep* 2018, **8**:15699.

15. •• Sommer R, Makshakova ON, Wohlschlager T, Hutin S, Marsh M, Titz A, Kunzler M, Varrot A: **Crystal structures of fungal tectonin in complex with O-methylated glycans suggest key role in innate immune defense**. *Structure* 2018, **26**:391-402
An unusual capsid-shape assembly of lectins with direct link to the function as defense against pathogens and predator.

16. Kumar A, Sykorova P, Demo G, Dobes P, Hyrsl P, Wimmerova M: **A novel fucose-binding lectin from *Photorhabdus luminescens* (PLL) with an unusual heptabladed beta-propeller tetrameric structure**. *J Biol Chem* 2016, **291**:25032-25049.

17. Audfray A, Claudinon J, Abounit S, Ruvoën-Clouet N, Larson G, Smith DF, Wimmerová M, Le Pendu J, Römer W, Varrot A *et al.*: **Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes**. *J Biol Chem* 2012, **287**:4335-4347.

18. Kostlanová N, Mitchell EP, Lortat-Jacob H, Oscarson S, Lahmann M, Gilboa-Garber N, Chambat G, Wimmerová M, Imberty A: **The fucose-binding lectin from *Ralstonia solanacearum*: a new type of -propeller architecture formed by oligomerisation and interacting with fucoside, fucosyllactose and plant xyloglucan**. *J Biol Chem* 2005, **280**:27839-27849.

19. Chaudhuri I, Soding J, Lupas AN: **Evolution of the beta-propeller fold**. *Proteins* 2008, **71**:795-803.

20. Kopec KO, Lupas AN: **beta-Propeller blades as ancestral peptides in protein evolution**. *PLoS One* 2013, **8**:e77074.

21. Hirabayashi J, Yamada M, Kuno A, Tateno H: **Lectin microarrays: concept, principle and applications**. *Chem Soc Rev* 2013, **42**:4443-4458.

22. Audfray A, Beldjoudi M, Breiman A, Hurbin A, Boos I, Unverzagt C, Bouras M, Lantuejoul S, Coll JL, Varrot A *et al.*: **A recombinant fungal lectin for labeling truncated glycans on human cancer cells**. *PLoS One* 2015, **10**:e0128190.

23. Machon O, Baldini SF, Ribeiro JP, Steenackers A, Varrot A, Lefebvre T, Imberty A: **Recombinant fungal lectin as a new tool to investigate O-GlcNAcylation processes**. *Glycobiology* 2017, **27**:123-128.

24. • Liu W, Han G, Yin Y, Jiang S, Yu G, Yang Q, Yu W, Ye X, Su Y, Yang Y *et al.*: **AANL (*Agrocybe aegerita* lectin 2) is a new facile tool to probe for O-GlcNAcylation**. *Glycobiology* 2018, **28**:363-373

The AANL lectin is proven as a useful tool for enrichment and identification of *O*-GlcNAcylated proteins and peptides; a fine example of the use of lectin in glycoproteomics

25. Voet AR, Noguchi H, Addy C, Simoncini D, Terada D, Unzai S, Park SY, Zhang KY, Tame JR: **Computational design of a self-assembling symmetrical beta-propeller protein.** *Proc Natl Acad Sci U S A* 2014, **111**:15102-15107.

26. Noguchi H, Addy C, Simoncini D, Wouters S, Mylemans B, Van Meervelt L, Schiex T, Zhang KYJ, Tame JRH, Voet ARD: **Computational design of symmetrical eight-bladed β-propeller proteins.** *IUCrJ* 2019, **6**:46-55.

27. Yadid I, Tawfik DS: **Functional beta-propeller lectins by tandem duplications of repetitive units.** *Protein Eng Des Sel* 2011, **24**:185-195.

28. •• Smock RG, Yadid I, Dym O, Clarke J, Tawfik DS: **De novo evolutionary emergence of a symmetrical protein is shaped by folding constraints.** *Cell* 2016, **164**:476-486
First example of reconstruction of artificial functional lectin from 47-residue ancestral motifs that form five-bladed propellers via oligomeric assembly.

29. Olausson J, Astrom E, Jonsson BH, Tibell LA, Pahlsson P: **Production and characterization of a monomeric form and a single-site form of *Aleuria aurantia* lectin.** *Glycobiology* 2011, **21**:34-44.

30. Arnaud J, Claudinon J, Tröndle K, Trovaslet M, Larson G, Thomas A, Varrot A, Römer W, Imberty A, Audfray A: **Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity.** *ACS Chem Biol* 2013, **8**:1918-1924.

31. Amano K, Fujihashi M, Ando A, Miki K, Nagata Y: **Involvement of tyrosines at fucose-binding sites of *Aleuria aurantia* lectin: non-equal response to site-directed mutagenesis among five sites.** *Biosci Biotechnol Biochem* 2004, **68**:841-847.

32. •• Tobola F, Lelimousin M, Varrot A, Gillon E, Darnhofer B, Blixt O, Birner-Gruenberger R, Imberty A, Wiltschi B: **Effect of non-canonical amino acids on protein-carbohydrate interactions: structure, dynamics and carbohydrate affinity of a lectin engineered with fluorinated tryptophan analogs.** *ACS Chem Biol* 2018, **13**:2211-2219
First inclusion of fluoro-tryptophan analogs, that is, non-canonical amino acids, in a lectin binding site, resulting in change of specificity towards histo-blood group oligosaccharides.

33. Romano PR, Mackay A, Vong M, DeSa J, Lamontagne A, Comunale MA, Hafner J, Block T, Lec R, Mehta A: **Development of recombinant *Aleuria aurantia* lectins with altered binding specificities to fucosylated glycans.** *Biochem Biophys Res Commun* 2011, **414**:84-89.

34. • Norton P, Comunale MA, Herrera H, Wang M, Houser J, Wimmerova M, Romano PR, Mehta A: **Development and application of a novel recombinant *Aleuria aurantia* lectin with enhanced core fucose binding for identification of glycoprotein biomarkers of hepatocellular carcinoma.** *Proteomics* 2016, **16**:3126-3136
Development of a candidate oncofetal marker for liver cancer by engineering a fungal lectin and modifying specifically its binding affinity to fucose linkages.

35. Ramberg KO, Antonik PM, Cheung DL, Crowley PB: **Measuring the impact of PEGylation on a protein-polysaccharide interaction.** *Bioconjug Chem* 2019, **30**:1162-1168.

36. Arnold L, Chen R: **Novel thermo-responsive fucose binding ligands for glycoprotein purification by affinity precipitation.** *Biotechnol Bioeng* 2014, **111**:413-417.

37. •• Ribeiro JP, Villringer S, Goyard D, Coche-Guerente L, Höferlin M, Renaudet O, Römer W, Imberty A: **Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells.** *Chem Sci* 2018, **9**:7634-7641
Design of the first chimeric bispecific lectin, with two rationally oriented and distinct recognition surfaces that bind independently to both fucosylated and sialylated glycoconjugates.

38. Murzin AG, Lesk AM, Chothia C: **beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors.** *J Mol Biol* 1992, **223**:531-543.

39. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ: **Carbohydrate-binding modules: fine-tuning polysaccharide recognition.** *Biochem J* 2004, **382**:769-781.

40. Hazes B: **The (QxW)3 domain: a flexible lectin scaffold.** *Protein Sci* 1996, **5**:1490-1501.

41. Khan F, Suguna K: Crystal structures of a β-trefoil lectin from *Entamoeba histolytica* in monomeric and a novel disulphide bond-mediated dimeric forms. *Glycobiology*, in press.

42. Rutenber E, Robertus JD: **Structure of ricin B-chain at 2.5 A resolution.** *Proteins* 1991, **10**:260-269.

43. •• Bleuler-Martinez S, Stutz K, Sieber R, Collot M, Mallet JM, Hengartner M, Schubert M, Varrot A, Kunzler M: **Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against nematodes.** *Glycobiology* 2017, **27**:486-500
Illustration of the protective function of lectin in fungi and of the requirement of multivalency for maintaining its toxicity towards intervertebrates.

44. Grahn E, Askarieh G, Holmner A, Tateno H, Winter HC, Goldstein IJ, Krengel U: **Crystal structure of the *Marasmius oreades* mushroom lectin in complex with a xenotransplantation epitope.** *J Mol Biol* 2007, **369**:710-721.

45. Hoorelbeke B, Huskens D, Ferir G, Francois KO, Takahashi A, Van Laethem K, Schols D, Tanaka H, Balzarini J: **Actinohivin, a broadly neutralizing prokaryotic lectin, inhibits HIV-1 infection by specifically targeting high-mannose-type glycans on the gp120 envelope.** *Antimicrob Agents Chemother* 2010, **54**:3287-3301.

46. Zhang F, Hoque MM, Jiang J, Suzuki K, Tsunoda M, Takeda Y, Ito Y, Kawai G, Tanaka H, Takenaka A: **The characteristic structure of anti-HIV actinohivin in complex with three HMTG D1 chains of HIV-gp120.** *Chembiochem* 2014, **15**:2766-2773.

47. • Liao JH, Chien CT, Wu HY, Huang KF, Wang I, Ho MR, Tu IF, Lee IM, Li W, Shih YL *et al.*: **A multivalent marine lectin from *Crenomytilus grayanus* possesses anti-cancer activity through recognizing globotriose Gb3.** *J Am Chem Soc* 2016, **138**:4787-4795
A promising example for development of lectin-based biosenzor based on the specificity to globotriose (Gb3). Gb3 represents marker for various types of cancer and in this work, CGL reduced viability of breast cancer cell line.

48. • Terada D, Kawai F, Noguchi H, Unzai S, Hasan I, Fujii Y, Park SY, Ozeki Y, Tame JRH: **Crystal structure of MytiLec, a galactose-binding lectin from the mussel *Mytilus galloprovincialis* with cytotoxicity against certain cancer cell types.** *Sci Rep* 2016, **6**:28344
The prototype of mussel lectin, showing dose-dependent cytotoxic activity to certain cancer cell lines, whereas the engineered monomer looses haemagglutination and cytotoxic activity

49. Broom A, Doxey AC, Lobsanov YD, Berthin LG, Rose DR, Howell PL, McConkey BJ, Meiering EM: **Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein.** *Structure* 2012, **20**:161-171.

50. • Broom A, Ma SM, Xia K, Rafalia H, Trainor K, Colon W, Gosavi S, Meiering EM: **Designed protein reveals structural determinants of extreme kinetic stability.** *Proc Natl Acad Sci U S A* 2015, **112**:14605-14610
ThreeFoil, as an example of how rational design can improve solubility and folding kinetics. Moreover, lactose, as a ligand, served as chaperone in the process of Three-Foil folding which supported the advantage of designing the fold and function together.

51. •• Terada D, Voet ARD, Noguchi H, Kamata K, Ohki M, Addy C, Fujii Y, Yamamoto D, Ozeki Y, Tame JRH *et al.*: **Computational design of a symmetrical β-trefoil lectin with cancer cell binding activity.** *Sci Rep* 2017, **7**:5943
Design of Mitsuba (three leaf in Japanese), an artificial protein engineered from MytiLec. This β-trefoil lectin is composed of 3 identical tandem repeats and retains sugar binding activity.

52. Yabe R, Suzuki R, Kuno A, Fujimoto Z, Jigami Y, Hirabayashi J: **Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-binding protein by natural evolution-mimicry.** *J Biochem* 2007, **141**:389-399.

53. Hu D, Tateno H, Kuno A, Yabe R, Hirabayashi J: **Directed evolution of lectins with sugar-binding specificity for 6-sulfo-galactose.** *J Biol Chem* 2012, **287**:20313-20320.

54. Connaris H, Crocker PR, Taylor GL: **Enhancing the receptor affinity of the sialic acid-binding domain of *Vibrio cholerae* sialidase through multivalency.** *J Biol Chem* 2009, **284**:7339-7351.

55. Ribeiro J, Pau W, Pifferi C, Renaudet O, Varrot A, Mahal LK, Imberty A: **Characterization of a high-affinity sialic acid specific CBM40 from *Clostridium perfringens* and engineering of divalent form.** *Biochem J* 2016, **473**:2109-2118.

56. Hwang HJ, Han JW, Jeon H, Han JW: **Induction of recombinant lectin expression by an artificially constructed tandem repeat structure: a case study using *Bryopsis plumosa* mannose-binding lectin.** *Biomolecules* 2018, **8**:E146.

57. Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lutteke T, O'Boyle N, Packer NH, Stanley P, Toukach P *et al.*: **Updates to the symbol nomenclature for glycans guidelines.** *Glycobiology* 2019, **29**:620-624.

2. AIM OF THE THESIS AND OBJECTIVES

Lectin selectivity and architecture determine the character and strength of binding toward its ligands. In respect to these aspects, the main aim of this thesis was to implement synthetic biology and protein engineering in order to create artificial lectins.

The creation of artificial lectins or proteins in general brings also new questions to the protein nomenclature. Often, it is not an easy task to identify these new biomolecules correctly and the opinion of scientific community can differ as well. The concept of protein chimeras is known for quite some time and refers to a synthetic protein composed of at least two different functional domains or proteins where such association comes from gene fusion, click chemistry, etc. On the other hand, the neo-lectins might be considered as a class of synthetic lectins with controlled valency or specificity, however, such lectin can be composed only from one type of lectin or carbohydrate binding module which are organized in tandem repeats.

With this classification, we established our main objectives:

- ***Discovery and characterization of novel lectins with interesting topology and specificity.***
  Novel lectin architectures and well-defined specificity are always of high interest and therefore the development of searching tools, such databases, and subsequent empirical cross-examination represent one of the fundamental approach in biology.
- ***Neo-lectins engineering.***
  Lectins or carbohydrate binding modules can be engineered by various approaches of synthetic biology. Our focus was on the increase of lectin valency with the expectation of higher affinity toward ligands.
- ***Generalization of synthetic biology approach, i.e., Janus lectin, as a universal tool for constructions of bispecific chimeras with improved valency***.
  We intended to create novel lectin chimeras in order to demonstrate that previously developed strategy of Janus lectin can be applicable on various lectin folds and selectivity.

The main three objectives of this Ph.D. thesis are summarized in individual scientific-article manuscripts, from which the first one was already published in the preprint server BioRxiv and in the journal Communications Biology. The other two are presented as manuscripts in preparation and were deposited in preprint server BioRxiv and submitted to journals where there are currently under revisions.

3. METHODS

3.1 Gene design

All proteins characterized in this work have been obtained as recombinant material from the bacterial culture of *Escherichia coli*. The DNA encoding the protein sequence can be obtained from the original organisms by extraction techniques and amplification, but nowadays, it is easier to design the gene of interest and purchase it when the sequence is known. The original DNA sequence is obtained from literature or from databases, and some modifications, i.e., "gene design" are necessary before introducing it in the *E. coli* host. Gene design includes codon optimization, the creation of novel genes by fusion of additional domains and linkers, and the supplement of different tags for production and purification.

Codon optimization is usually the first step. The genetic code is degenerated and therefore each amino acid can be coded by more than one codon, i.e., a triplet of nucleotide. The frequency of codons varies from one species to another, which is reflected in the efficiency of each organism to use codons. For better translation, it is therefore recommended to optimize the DNA sequence for the host organism. Additionally, extra sequences, such as fusion tags and sites for restriction enzymes (REs) can be introduced. Using the synthetic biology approach, we can now create also novel proteins by suitable gene design. Divalent lectins are produced by connecting two copies of the same gene with the insertion of a linker. Bispecific chimeras (Janus lectins) are designed as a fusion of two genes coding for two different lectin domains, again with the insertion of a linker. In all cases, the designed genes were synthesized by the company Eurofins.

3.2 Cloning

Cloning is a process in which the gene of interest is inserted into the suitable vector followed by vector introduction into the expression host system. The whole procedure can be divided into several steps, 1. gene and vector cleavage by REs, 2. ligation, 3. positive clones screening, and 4. insertion of the vector into the expression system. Since my thesis was focused on bacterial expression system, we can refer to vectors as plasmids and the insertion of a vector

into the host system as transformation. The first step of cloning is gene and plasmid cleavage by the same types of REs. Ideally, two different REs are used resulting in 5' and/or 3' noncompatible protruding ends. After, the cleaved plasmid and gene are separated and purified by agarose electrophoresis. The correct bands are snipped and ligated together by DNA ligase. During ligation, gene and plasmid of interest are linked together, in the oriented direction. However, self-ligation of the plasmid can occur and therefore screening for positive clones is necessary. Once confirmed by polymerase chain reaction (PCR) that the ligation was successful, the plasmid is transformed into a bacterial host system. During this process, the bacteria are exposed to stress conditions (e. g. increase of temperature – heat-shock transformation) in which the take-up of foreign genetic material (plasmid) from the environment is achieved.

3.3 Heterologous expression and protein production

The heterologous expression is a process in which the gene is expressed in the host organism whereas this organism does not contain the gene naturally and therefore was inserted inside by various processes, e. g. transformation. There are several ways how to produce the protein recombinantly either in cell-based or cell-free systems. Cell-free production, as indicated in the name, is performed *in vitro.* On contrary, cell-based production is done *in vivo*, generally in bacteria, yeast, insects or mammalian cells. For the need of this thesis, only a bacterial expression system is described. Bacteria, such as *Escherichia coli*, represent one of the most widely used expression host. Its main advantages include cheap manufacturing, easy manipulation, a fast growth rate and a high yield of products. On the other hand, the bacterial system is not suitable for proteins that require complex post-translational modifications, such as glycosylation, although there are already attempts to overcome this issue (Wacker et al., 2002). The overexpression occurs either by increasing the number of plasmid copies or by increasing the strength of the promoter region. In this work, several variants of plasmid pET were used. pET is a low copy, powerful, and widely used vector in bacterial expression systems. It is designed for high-level recombinant expression due to the usage of T7 promoter instead of regular bacterial promoters such as lactose promoter LacP. In this case, the gene for T7 polymerase is usually integrated into the host genome and it is under the control of other commonly used promoters such as LacP, AraP, or RhaP. Therefore, upon induction, the T7

RNA polymerase is expressed and subsequently used for gene overexpression on the pET plasmid. The protein can be expressed in cytoplasm, periplasm or secreted into media. The protein is preferably produced in soluble form and with correct folding. The insoluble production can occur as well while the misfolded protein is entrapped in inclusion bodies and subsequential refolding is necessary. In order to avoid protein misfolding, lower production temperature or optimization of inducer concentration can be applied. Alternatively, the change of expression plasmid, bacterial strain or host system might be considered.

3.4 Protein purification

Protein purification is a method in which the protein is isolated from the complex mixture. In our case, the proteins were purified from the cell lysate and therefore, disruption of bacteria is required prior to purification. In the next step, centrifugation, heavy cellular particles, such as membranes or inclusion bodies, are separated from the supernatant containing all soluble proteins. Afterward, a suitable purification strategy is chosen, in our case, we used affinity purification and size exclusion chromatography (SEC). As indicated in the name, affinity chromatography is based on specific binding or interaction between protein and a ligand. In this thesis, we have used two different modes of binding – immobilized metal ion affinity (IMAC) and immobilized sugar affinity. In the first case, the principle of the method uses the fact that a peptide containing several histidines displays an affinity for certain metal ions, such as nickel or cobalt. A tag composed of at least 6 histidine residues (HisTag) is fused at the C- or N-terminus of the protein. Nickel (Ni), as a part of the stationary phase, is immobilized on agarose beads by chelation using nitrilotriacetic acid (NTA). Once the protein is applied to the Ni-NTA column, the coordination bond between Ni and His is formed. Low affinity bound proteins and contaminants can be removed by low concentration of imidazole, a small molecule that mimics histidine properties. By increasing the concentration of imidazole, the proteins of interest can be eluted. IMAC is a widely used technique applicable for almost all types of proteins. The second affinity method requires more specific interaction, in our case lectin-sugar. In this case, the carbohydrate is immobilized on a matrix, e.g., agarose beads. The bond established between protein and ligand is noncovalent and its strength depends on lectin affinity to sugar. Also, only correctly folded and active lectins can be immobilized. The main advantage of this method over IMAC is the higher specificity because only lectin specific for certain sugar

interacts with the column. However, if the lectin specificity is unknown or its affinity is very weak, other purification strategies must be used. In order to obtain highly pure protein, affinity chromatography is usually followed by size exclusion chromatography (SEC). During this step, the protein separation is based on their size (larger proteins are eluted first) and can remove potential aggregated. Additionally, the protein size can be estimated from the calibration curve assigned to the column.

3.5 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a quantitative, label-free method which is considered as a gold standard technique for biophysical characterization for molecular interactions. ITC provides complete information about affinity, stoichiometry and thermodynamic profile of the interaction. It allows us to characterize the binding properties of proteins and their ligands by measuring the heat exchanges occurring during the interaction at constant pressure and temperature. The system is composed of two identical cells surrounded in adiabatic jacket and a syringe filled with titrant solution which is injected into sample cell while constantly stirring. Sample cell is usually filled with protein solution whereas the reference cell is filled with buffer or water. Once the titrant is injected into sample cell, depending on the thermodynamics of interaction, the heat is either released (exothermic reaction) or absorbed (endothermic reaction). This heat change causes a difference of temperature between sample and reference cell and the system compensates the difference between temperatures by applying an electric power (μcal/s) in order to maintain constant temperature in both cells. The heat exchange (Q) is proportional to the thermal power and is calculated as the integration under each peak:

$$Q = V\,\Delta H[M]\;\; Ka[L]/(1 + Ka[L])$$

where V is the volume of the cell, ΔH is the enthalpy of binding per mole of ligand, M is the total macromolecule concentration including bound and free fractions, Ka is the binding constant, and L is the free ligand concentration.

The thermodynamic principle applied in ITC is characterized by formula:

$$-RT\;\ln Ka = \;\Delta G = \;\Delta H - T\Delta S$$

where the Ka value relates to Gibbs free energy of binding (ΔG) which can be also expressed as enthalpic (ΔH) and entropic (ΔS) contribution to the binding affinity, where R is universal gas constant and T is absolute temperature in Kelvin. The thermodynamical signature of the interaction, primary ΔG, indicates how strong is the binding and therefore two ligands, with the same affinities, can have a different mode of binding if their ΔH and ΔS are altered (Menéndez, 2020).

The binding event is described as an equilibrium of association constant Ka, where M and L represent the concentrations of unbound reactants, macromolecule, and ligand, ML is the concentration of complex and Kd is a dissociation constant.

$$\mathrm{Ka} = ([\mathrm{ML}])/([\mathrm{M}][\mathrm{L}]) = 1/\mathrm{Kd}$$

Prior to the experiment, the suitable concentration of macromolecule and ligand is selected. This step is critical for a successful titration experiment and as an indicator, a dimensionless parameter *c* is calculated:

$$\mathrm{c} = \mathrm{Ka}[\mathrm{M}]\ \mathrm{n}$$

whereas values $1 \leq c \leq 1000$ (ideally $10 \leq c \leq 100$) indirectly guarantee the achievement of sigmoid shaped isotherms, allowing for the fitting of n, Ka and ΔH from the first equation above. Experiments conditions with *c* values below 10 or above 100 do not allow for fitting three parameters and require some assumption on the interactions for obtaining Ka.

3.6 Surface plasmon resonance

Surface plasmon resonance (SPR) is a label-free optical method that is used for the biophysical characterization of biomolecules interactions. The difference compared to ITC is the immobilization of either the protein or its binding partner on the surface of a chip. The immobilized molecule is called the ligand whereas the molecule injected in the aqueous solution is referred to as the analyte. SPR measures the refractive index of the sensor surface and as the analyte binds to the ligand, the increase of mass results in the increase of the refractive index. The change of refractive index is measured in real-time and plotted directly as sensorgram - resonance units (RU) versus time. The analyte is injected through reference (no ligand) and sample channels (with immobilized ligand) which allows to subtract

background response. SPR gives access to the kinetics of the interaction. The affinity (K_A and/or K_D) can be obtained by measuring k_{on} and k_{off}, or by equilibrium binding assays where various concentrations of the analyte are injected in several cycles and the level of binding is measured at equilibrium. A regeneration step is generally necessary between injections to dissociate the analyte. If the rate at which the analyte binds ligand exceeds the rate at which it is delivered to the surface the event referred to as mass transport occurs. Mass transport leads to misinterpretation and wrong determination of the actual kinetics and therefore, the optimization of experimental conditions, such as lower level of immobilized ligand and higher flow rate might be necessary. If the ligand immobilized on the chip shows long-term stability and the surface can be sufficiently regenerated, the SPR chips can be reused (Merwe, 2000). There are different types of chips available in the market, but probably the most used are CM5 chips functionalized with carboxymethylated dextran covalently linked to the gold surface.
When compared to ITC, the SPR technique requires smaller sample consumption and is less sensitive to aggregation. SPR and ITC are complementary methods used for determining biophysical characterization and the usage of both can bring different insights into the understanding of the interaction of biomolecules.

3.7 Differential scanning calorimetry

Differential scanning calorimetry (DSC) is a label-free technique used for determining the thermal stability of proteins or other biomolecules (lipids, nucleic acids). The main principle of DSC is to follow the denaturation of the sample as a function of increase of temperature is with constant pressure. During thermal denaturation, the proteins undergo structural changes commonly referred to as unfolding. These changes result in heat absorption caused by the reorganization of non-covalent bonds and therefore DSC measures ΔH of denaturation event, as well as T_m (transition temperature midpoint) in which unfolded and folded state, are in equilibrium. Generally, higher T_m signifies higher stability of the protein in the solution. Similar to ITC, DSC is composed of two identical cells (sample and reference) which are maintained at the same temperature during the measurement. The reference cell is filled with buffer and the sample cell with a protein solution and both cells are heated at the same scan rate. Once the protein starts to denature, a temperature difference occurs between the cells, and the electric power needed to maintain equal temperature is recorded. The thermogram of the reference cell is used for data subtraction. The thermodynamical profile of the studied sample

provides information about the thermal stability of the sample but also about various events occurring during the denaturation process such as oligomer destabilization followed by denaturation of monomer or unfolding of multiple distinct domains at different temperatures. Additionally, DCS can be used in protein-ligand interaction studies, where ligand usually increase the stability of protein in solution resulting in increased T_m.

3.8 Hemagglutination, hemolysis, inhibition assay

Lectins, as carbohydrate-binding proteins are known for their ability to agglutinate cells, especially human or animal erythrocytes (in this thesis rabbit erythrocytes). Therefore, hemagglutination assay (HA) and inhibition assay (IA) are routinely used methods for characterization of activity and specificity of multivalent lectins. HA is standardly performed in a microtiter plate where the lectin solution is serially diluted, mixed with blood solution, and incubated. If the lectin is active and specific to any sugar present on the surface of erythrocytes, the agglutination is observed as a blurry, unclear solution, preventing blood clotting at the bottom of the wells. Since the test is performed with serial dilution of lectin solution, the lowest concentration needed for the hemagglutination is estimated. In specific cases, the lectins can induce hemolysis instead of hemagglutination, resulting in partial or complete lysis of red blood cells. Visually, this event looks similar to hemagglutination and therefore further centrifugation step is necessary in order to evaluate the results correctly. When the lectin shows activity towards erythrocytes, the inhibition assay (IA) can be used to determine lectin specificity. In such a case, lectin is pre-incubated with a serially diluted carbohydrate. Then, the solution of erythrocytes is added and incubated as in HA. The pre-incubation step with carbohydrates should result in the occupation of lectin binding sites and therefore inhibition of hemagglutination or hemolysis and the crucial concentration of inhibitor can be determined as IC_{50} value, albeit with rather limited quantitative information.

3.9 X-rays crystallography

In order to determine protein structure by X-rays crystallography, several steps need to be followed, i) obtention of protein crystals of good quality, ii) data collection using X-rays source

such as synchrotron, iii) data reduction, solving of phase and first model building and iv) structure refinement and deposition to Protein Data Bank (PDB).

The crystallization of protein is empirical and is dependent on many factors (temperature, protein concentration, pH, additives, presence of ligand, etc.). There is no universal guideline suggesting the best crystallization conditions therefore the strategy “try-success” is the most common approach and usually extensive screening is performed. Commercially available crystallization screens, e.g., Morpheus II, BCS screen, Clear Strategy, etc., provide pre-mixed solutions and therefore allow screening of many different conditions. The crystallization depends also on the concentration, purity and conformational homogeneity of the protein samples. Proteins, especially lectins, can be co-crystallized in the presence of their ligands, however, the crystallization conditions might differ from the apo-protein. Methods such as vapor diffusion, dialysis, or micro-batch are commonly used in protein crystallography, however, only vapor diffusion will be described. In the vapor diffusion, the protein sample is mixed with a reservoir solution in a specific in ratio, e.g., 1:1, and is placed aside from the reservoir (sitting drop) or atop (hanging drop). The chamber is sealed, in order to maintain the most controlled crystallization conditions, and over time, due to the difference in concentrations between drop and reservoir solution, the water from protein solution (drop) vaporizes into the reservoir, resulting in supersaturation followed by crystallization.

The crystallization process is characterized by two events, nucleation and crystal growth. The supersaturation is the driving force of crystallization and is a critical step for crystal nucleation. During this process, the proteins arrange into a certain pattern and this gives the base for the crystal nucleus. In the ideal case, only one or few crystal nuclei are present and are further grown resulting in bigger size crystals. If there are a lot of nuclei (tens or hundreds) crystals are too small, do not grow and usually are not suitable for data collection. In the case that supersaturation is not reached, no crystallization is observed. On the other hand, if we reach over supersaturation, protein precipitation is observed.

Once the crystal is obtained, it is cryoprotected and mounted into a loop prior to flash freezing in liquid nitrogen in order to formed vitreous ice and limit radiation damages. A soaking procedure, for ligand or phasing agent incorporation, may be performed before crystal freezing. The loop mounted on magnetic pin are locked in a puck and constantly maintained in the cryo conditions until they arrive at the synchrotron facility for the data collection (ii). The crystal is

placed in the X-ray beam using a sample changer and the diffraction pattern data are recorded by a detector.

A crystal is a regular and periodical arrangement of a structural pattern (one or several atoms or molecules) in 3D space. This pattern repeats itself in a lattice (array of points) and the minimal repeating unit of a lattice is called the unit cell. The lattice can generate a whole crystal by repetitive translation along its principal axes and is defined by three vector lengths a, b, and c, and angles α, β, and γ. They are only 14 different 3D configurations into which atoms or molecules can be arranged in crystals known as 14 Bravais lattices. Some lattices have inherent symmetry such as rotations, mirror planes, inversions and/or translation leading to 230 space groups. Only 65 are encountered in protein crystallography as proteins are chiral and mirror planes or inversion center cannot appear in their crystals. This brings us to the term asymmetric unit, which is defined as the smallest fraction of unit cell that can help to generate the complete unit cell by applying only symmetry operators space groups.

When X-rays are applied to a periodic object as a crystal, diffraction will be observed only when the beams diffracted from one unit cell are in phase with those from other unit cells. In the 2D diffraction pattern recorded, each reflection, i.e., spot, corresponds to the beam diffracted by a group of lattice planes. Each reflection is defined by its Miller indices (h, k and l) and its structure factor F_{hkl} composed of two parameters, the phase and the amplitude of the wave diffracted from the crystal lattice plane (h, k, l). The amplitude is proportional to the number of electrons of the lattice plane and is directly calculated from the intensity of reflections measured on the diffraction pattern. A diffraction experiment requires to measure a large number of reflections and as only a finite number can be obtained on the diffraction pattern in a crystal orientation (image), several orientations need to be recorded to collect a full data set and the amount of orientation will depend on the crystal symmetry. All the images are analyzed using computer program such as XDS which allows space group determination, indexation (giving Miller indices to each spot), measure intensity for each reflection and merging of equivalent reflections but also scaling.

In X-ray crystallography, the phases information is lost and only the amplitudes can be obtained from the diffraction pattern. Several methods have been developed and are used to solve the so-called phase problem and therefore the protein structure. When there is a known structure for a homologous protein, the most commonly used method is the molecular replacement (MR) in which the phases are calculated from model. If there is no similar structure, the phases are

solved experimentally, e.g., by multiwavelength anomalous diffraction (MAD) in which atoms with special properties (high atomic number, anomalous scattering) are incorporated into protein. Data are collected at different wavelengths and their analysis permits the location of

those specific atoms and thus the determination of the phases. Once the phases and amplitudes are known, the initial electron density map is calculated using Fourier transform and an initial model is build. This contains errors and the differences between the atomic positions of the model (calculated) and the actual structure (observed) must be minimized through a process called refinement usually carried out through iterative cycles and by taking into account geometrical restrains. Program such as Refmac5 interspaced with manual model building with the software COOT used for atomic model and electron density visualization were used. Ligands and water molecules are added to the structure as well. The stereochemical quality of the final model is checked during the validation step prior deposition of the atomic coordinates with the structure factors into the Protein Data Bank.

4. PREDICTION AND CHARACTERIZATION OF NOVEL PORE FORMING LECTIN

4.1 Summary

The first manuscript is dedicated to the discovery and characterization of a novel β-trefoil lectin from primitive marine eukaryote *Salpingoeca rosetta.* The project actually started as a side project of my Ph.D. and with time it developed as of my main projects.

This study is a collaboration with a former Ph.D. student Francois Bonnardel who was supervised by Dr. Anne Imberty (CERMAV CNRS) and Dr. Frederique Lisacek (Swiss Institute of Bioinformatics, Switzerland). He developed a database for the prediction of β-trefoil lectins in the genomes. My contribution to this project was to experimentally confirm the correct functionality of the database, i.e., the identification of suitable candidates (SaroL-1), gene design, cloning, protein production, protein purification, and biophysical (ITC, DSC), and structural characterization (x-rays crystallography) of the lectin. To support the value of our data, we established a collaboration within the SynBIOcarb network with the team of Prof. Winfried Römer and another two Ph.D. students, ESR4 Lina Siukstaite and ESR5 Francesca Rosato. This allowed us a better understanding of lectin behavior on the cells (carried out by Francesca) and on GUVs (carried out by Lina). Once we gathered our data, we concluded our results in the scientific article where I contributed to the manuscript and figures preparation and corrections.

The manuscript presented here was submitted to preprint server BioRxiv, DOI: 10.1101/2022.02.10.479907, however the revised version with the title *The choanoflagellate pore-forming lectin SaroL-1 punches holes in cancer cells by targeting the tumor-related glycosphingolipid Gb3*, with updated data such as electron microscopy for visualization pore structures, was submitted and published by the journal Communications Biology DOI: 10.1038/s42003-022-03869-w.

4.2 Scientific article I

A pore-forming β-trefoil lectin with specificity for the tumor-related glycosphingolipid Gb3

Simona Notova [1†], François Bonnardel [1,2,3,†], Francesca Rosato [4,5], Lina Siukstaite [4,5], Jessica Schwaiger [4,5], Jia Hui Lim [1], Nicolai Bovin [6], Annabelle Varrot [1], yu Ogawa [1], Winfried Römer [4,5,7*], Frédérique Lisacek [2,3,8*] and Anne Imberty [1*]

[1] Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France
[2] Swiss Institute of Bioinformatics, CH-1227 Geneva, Switzerland
[3] Computer Science Department, UniGe, CH-1227 Geneva, Switzerland
[4] Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany
[5] Signalling Research Centers BIOSS and CIBSS, University of Freiburg, 79104 Freiburg, Germany
[6] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow 117997, Russian Federation
[7] Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, 79104 Freiburg, Germany
[8] Section of Biology, UniGe, CH-1205 Geneva, Switzerland.

† Both first authors contributed equally to this work.

* To whom correspondence should be addressed. Anne Imberty (anne.imberty@cermav.cnrs.fr) Frédérique Lisacek (frederique.lisacek@sib.swiss) Winfried Römer (winfried.roemer@bioss.uni-freiburg.de)

Abstract

Lectins are efficient multivalent glycan receptors, deciphering the glyco-code on cell surfaces. The β-trefoil fold, characterized by three lobe-shaped repeats, is adopted by several classes of lectins, often associated with other domains having enzymatic or toxic activity. Based on the UniLectin3D database classification, the sequence signature of trefoil lobes was defined and used to predict 44714 lectins from 4497 species. Among them, SaroL-1 from the lower eukaryote *Salpingoeca rosetta* was predicted to contain both β-trefoil and aerolysin-like pore-forming domain. Recombinant SaroL-1 binds to galactose and derivatives, with a stronger affinity for cancer-related α-galactosylated epitopes such as glycosphingolipid Gb3 embedded in giant unilamellar vesicles or cell membranes. Crystal structures in complex with Gb3 trisaccharide and GalNAc show similarity with pore-forming toxins. Recognition of the αGal epitope on glycolipids was necessary for hemolysis of rabbit erythrocytes and toxicity on cancer cells through carbohydrate-dependent pore-formation.

Keywords

β-trefoil, lectins, oligomerisation, carbohydrate-binding protein, glycosphingolipid Gb3, aerolysin, giant unilamellar vesicles, cancer cells

Introduction

Lectins are protein receptors that bind complex carbohydrates without modifying them, and therefore participating in the signaling function of the glycocode encoded in glycoconjugates such as glycolipids and glycoproteins at the cell surface [1,2]. Lectins participate in multiple biological processes, such as embryonic development, cell growth and immunomodulation, and are crucial for the interactions between microorganisms and host cells (pathogenicity, symbiosis). Lectin domains are often associated with other functional proteins such as enzymes or toxins. Life-threatening examples are ricin [3] or cholera toxin [4], in which the lectin domain is responsible for the specificity and adhesion to cell surface glycans, prior to the cellular uptake of the toxin that interferes with metabolism.

A different mode of action is observed in pore-forming toxins (PFTs) that oligomerise and create holes into membranes of bacteria or host cells [5,6]. The specific β-pore forming toxins (β-PFTs) form pores fully lined by β-strands and include the aerolysin family [7,8]. These proteins contain a conserved aerolysin *C*-terminal domain and *N*-terminal domain that adopts different topologies targeting the cell surface, some of them with a lectin-like fold. At present, lectin-dependent β-PFTs have been described in fishes, i.e., natterin-like protein from zebrafish [9] and lamprey [10], in sea cucumber [11] and fungi [12]. Such modular proteins are of high interest since the lectin specificity can be employed to induce the cytotoxicity, for example, in cancer cells, as tested with the lamprey lectin [10]. Among other strategies, identifying novel pore-forming lectins can be a valuable tool for research and therapy.

A new software has been recently developed for the identification and annotation of lectins in proteomes, thereby allowing the search of β-PFT-containing lectins. This tool takes advantage of a structure-based classification of lectins proposed in UniLectin3D, a database of manually curated and classified lectin 3D structures including their oligomeric status and carbohydrate binding sites [13]. Tandem repeats are common in lectins, such as β-trefoils or β-propellers, and challenge conventional sequence motif finding methods. Nonetheless, detection is improved by considering each repeat independently with precise delineation based on the 3D shape. This approach was validated with β-propellers [14] and extended to β-trefoils.

β-trefoil lectins are small proteins consisting of three repeats of the same conserved binding motif, and this fold is widely distributed in nature [15]. They have been identified in bacteria, fungi, plants and animals, with a broad variety of sequences and local conformations but with conserved aromatic amino acids that create a common central hydrophobic core (Fig. 1a). β-trefoil lectins are popular in protein engineering for designing new scaffolds with high symmetry that can be associated with other domains. This functional and modular versatility is resourceful in synthetic biology, as recently reviewed [16]. Also, β-trefoil lectins bind efficiently to glycosylated surfaces and target specific cell types. For example, the Mytilec family initially identified in mussels binds to globotriaosyl ceramide (Gb3) that is overexpressed in some metastatic cancer [17-21]. Engineered Mytilec with perfectly symmetrical β-trefoil such as Mitsuba (three-leaf in Japanese) successfully demonstrates the potential of these small and symmetrical modules for recognizing cancer cells [22].

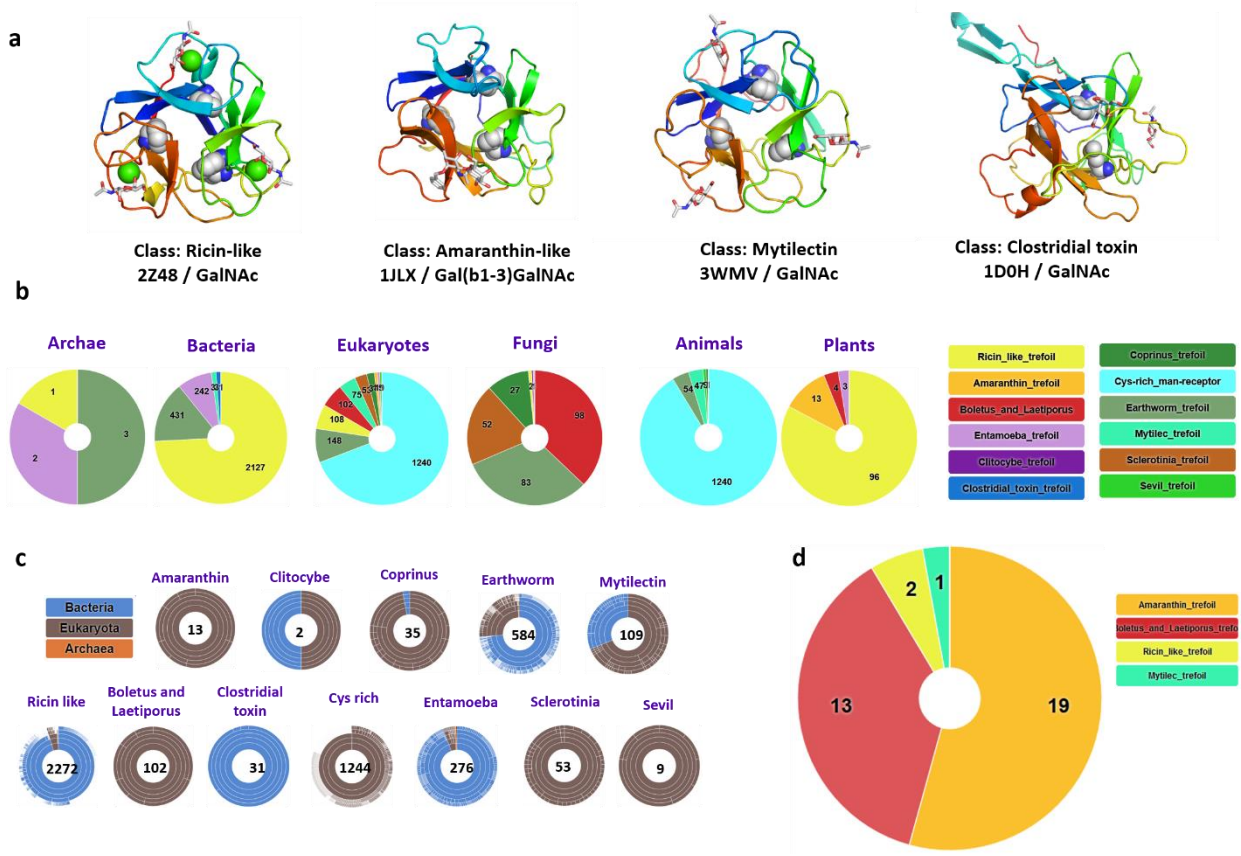


Figure 1. *a) Selected examples of β-trefoil lectins from different classes. The binding peptide is represented by a rainbow-colored ribbon, the bound sugar by sticks, and the conserved hydrophobic core-forming amino acids by spheres. b) Sunburst statistics for predicted β-trefoil lectins in different classes, in selected domains of life. c) Classification of predicted β-trefoil lectins and distribution of sequences in the TrefLec database. d) Prediction of β-trefoil lectins with an aerolysin domain based on the corresponding CATH domain (CATH entry 2.170.15.10).*

In this article, we introduce a new database of the UniLectin3D platform. TrefLec detects and classifies β-trefoil lectins resulting from screening translated genomes. TrefLec was then used to search for the occurrence of β-trefoils with an associated aerolysin domain to discover new β-PFT-lectins. Having identified a candidate of interest, we present here the broad structural and functional characterization of a novel pore-forming lectin.

Results

The TrefLec database for the prediction of β-trefoil lectins in genomes

The UniLectin3D classification spans 109 classes defined upon a 20% amino acid sequence identity cut-off. UniLectin3D contains 212 X-ray structures of β-trefoil lectins (i.e., 9% of the database content) corresponding to 63 distinct proteins. Although all structures share the same fold where hydrophobic amino acids form the β-trefoil core, they are spread across12 different classes. The Ricin-like class is the most populated, with 123 crystal structures presenting a very conserved fold in all kingdoms of life. Other types of β-trefoil are observed in invertebrates and fungi or can be involved in botulism and tetanus infection.

The 12 classes of β-trefoil lectins defined in UniLectin3D were used as the basis for constructing the TrefLec database of predicted β-trefoil lectins. The classes served for the identification of conserved motifs. Sequence alignment was performed at the "lobe" level for each class, i.e., using the three repeats for each sequence. Twelve characteristic Hidden Markov Model (HMM) profiles that capture amino acid variations were built, one for each class (Supplementary Fig. 1). The QxW signature that was observed earlier in Ricin-like lectins [23] is common to most classes, although with some degeneration, confirming an evolutionary link in most β-trefoils. In some classes, such as Amaranthin or Mytilec, the classic QxW signature is absent, but the same topology is preserved for the hydrophobic core.

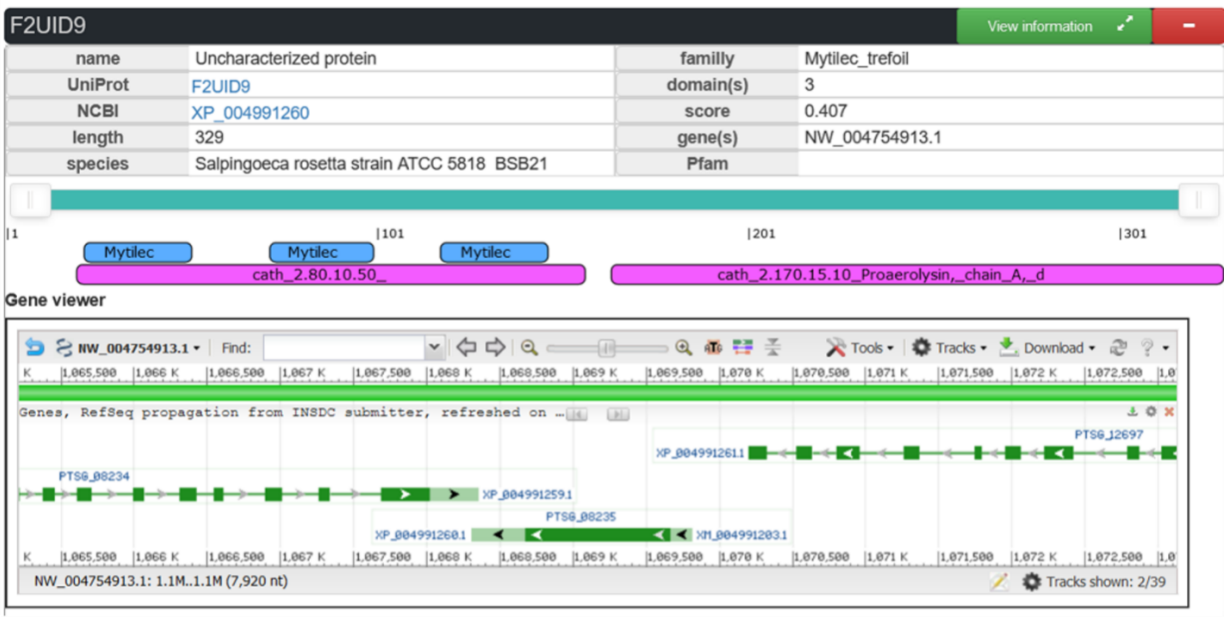
The 12 motifs were used to search in UniProt-trEMBL, UniProt-SwissProt and NCBI-nr protein databases for all kingdoms, including bacteria, viruses, archaea and eukarya with animals, fungi, and plants, comprising 197.232.239 proteins in 108.257 species for the RefSeq release of May 2021. This resulted in the identification of 4830 filtered sequences of putative lectins using a score cut-off of 0.25 (44714 unfiltered) in 1660 species (4497 unfiltered) (Fig. 1b, 1c). Sequence similarity of lectins between different classes created some overlap, i.e. a small number of sequences were predicted to occur in several classes, but the proportion remained low and was calculated, for example, to be lower than 2% across the Ricin-like, Entamoeba and earthworm classes (Supplementary Fig. 2).

The majority of sequences belong to the Ricin-like class (48%), the Cys-rich man receptor (27%) and the Earthworm-lectin (12%). All classes are not represented equally in the different kingdoms. The large Ricin-like class is over-represented in plants and bacteria. Fungi span the most extensive diversity of classes. Some classes are specific to one kingdom with for example Amaranthin-type lectins predicted to occur only in eukaryotes, specifically in plants [24]. Similarly, Coprinus β-trefoils are primarily identified in genomes of basidiomycetes fungi. Cys-rich lectins are sub-domains of membrane receptors occurring in vertebrates [25]. Clostridial toxins are only predicted in clostridial bacteria. In contrast, Mytilecs, known for their binding to cancer cells and cytolytic activity [20,26] have been structurally characterized in mollusks but predicted to occur in several invertebrates and bacterial species.

Search for new β-PFT-lectins in the TrefLec database

The TrefLec database (https://unilectin.eu/trefoil/) provides information on additional domains associated with the β-trefoil domain and predicted various enzymatic or toxic functions (Supplementary Table 1). The β-trefoils of the Ricin-like class are associated with glycosyl hydrolases, lipases or other enzymes. The β-trefoil Cys-rich domain is part of the macrophage mannose receptor that also contains fibronectin and multiple C-type protein domains [25]. The web interface can be used to search for specific domains defined in reference data sources such as CATH, the Protein Structure Classification database [27]. When searching for aerolysin or proaerolysin (CATH domain 2.170.15.10), 35 lectin-containing sequences were identified (Fig. 1d). Twelve sequences from fungi contain a "Boletus_and_Laetiporus_trefoil", all with strong similarity with the well described β-PFT-lectin LSL from *Laetiporus sulphureus* [12]. A related sequence is observed in the genome of the primitive plant *Marchantia polymorpha*. Only one β-PFT-lectin related to Ricin-type trefoil is predicted in bacteria *(Minicystis rosea*). Nineteen sequences from plants are predicted to contain an Amaranthin-type β-trefoil. Such plant toxins have not been structurally characterized, but their phylogeny has been recently reviewed and a role in the stress response was proposed [24]. In the remaining four sequences, a eukaryotic Mytilec domain was identified in the genome of *Salpingoeca rosetta,* a single-cell and colony-forming micro-eukaryotic marine organism belonging to the choanoflagellates group [28]. Lectins of the Mytilec-like class have been demonstrated to bind the αGal1-4Gal epitope on glycosphingolipid Gb3 from cancer cells [21].

a



b

```
N-ter:      MSSTFEPATDSPLPVPGVQ  19

α-lobe:     ---------YFLQHVQSGKYVHPHGGSDMPGNDTALVLHHGFDEKRDALRW  61

β-lobe:     VFVNDAENKHQLKHYSSGKFVHPKGGKV--GKFATLVVHSSP--GRPETMI  108

γ-lobe:     -EMVQEDGRTYLRHTDSDYYVHPHGGSPNPGDNTRLVYYSG---YRPSLAFLAIPA  160

Aerolysin:  ETLFVDRIEIHQAQALESINTITSLSDEHRNDTDQPVQTSISVALEESLQDSAQLSF
            ERCFGLKVGSEFEVGLPLVGKTKVSVQFSGSWKSSTIKGEVRTSAVKVQINEHVTIP
            PGKCVQIRIDTRRCTKTAPATMYLRTASGIEVQRETTVTSTYHYDQEVHVVPVTN   329
```

Figure 2. SaroL-1 sequence information. *a) Excerpt of the TrefLec page of the predicted lectin from Salpingoeca rosetta with information about the protein, the domains and the gene. b) Peptide sequence of SaroL-1 with separation of the domains and alignment of lobes for the β-trefoil domain. Amino acids corresponding to the signature of Mytilec-like class are highlighted in yellow, amino acids predicted to be involved in carbohydrate binding in green, and amino acids involved in the hydrophobic core in blue.*

Fig. 2 depicts the TrefLec entry of the protein identified in *Salpingoeca rosetta* and referred here as SaroL-1. The sequence (F2UID9 in UniProt) consists of 329 amino acids. The 166 amino acid sequence at the *C*-terminus displays 30% identity with aerolysin domains in various organisms. Three repeats are located in the *N*-terminal region with 41% identity with the artificial Mytilec Mitsuba [22] and 33% identity with other members of the Mytilec-like family from the Crenomytilus or Mytilus genera, with apparent conservation of the amino acids involved in carbohydrate binding.

Binding of SaroL-1 lectin domain to αGal-containing epitopes

The SaroL-1 gene was designed and fused with a 6-His-tag sequence and a cleavage site for Tobacco Etch Virus (TEV) protease at the *N*-terminus. The SaroL-1 protein was expressed in a soluble form in the BL21(DE3) strain of *Escherichia coli* and purified using immobilized metal ion affinity and size exclusion chromatography (Supplementary Fig. 3). Data provided by SEC-MALS and SDS PAGE analysis showed that SaroL-1 appears to be monomeric in solution with an estimated molecular weight of 36.86 ± 0.76 kDa.

The binding of SaroL-1 to different galactosyl-ligands was assessed in solution by isothermal titration calorimetry (ITC). The monosaccharides *N*-acetylgalactosamine (GalNAc) and α-methyl galactoside (GalαOMe) displayed a similar millimolar affinity with a Kd of 2.2 and 2.8 mM, respectively. All tested αGal disaccharides and the p-nitrophenyl-α-D-galactopyranoside (PNPG) derivative bound with affinities twice as strong, with a Kd close to 1 mM, except for αGal1-4Gal, the terminal disaccharide of the globoside Gb3, that was characterized as the highest affinity ligand (Kd = 390 ± 0.20 μM). Lactose that contains βGal displayed very weak binding, being 20 times less efficient than αGal1-4Gal, confirming the preference for the αGal epitope (Fig. 3a, b). Affinity values and thermodynamics parameters are listed in Supplementary Table 2. ITC isotherms are shown in Fig. 3a and Supplementary Fig. 4.The affinity of SaroL-1 to oligosaccharides in solution is not very strong but the avidity effect may result in much stronger binding to glycosylated surfaces. The ability of fluorescent-labelled SaroL-1 to bind multivalently to giant unilamellar vesicles (GUVs) with dimensions matching those of human cells [29] was then evaluated by confocal imaging, using a fluorescent lipid as a membrane marker. The GUVs were functionalized with diverse naturally occurring glycolipids, including wild type Gb3 mixture from porcine and lactosylceramide (Lac-cer), and several synthetic analogs consisting of the oligosaccharide attached to the phospholipid DOPE through a spacer molecule (Function-Spacer-Lipid, FSL). Strong binding of SaroL-1 was observed with GUVs containing FSL-Gb3 presenting the αGal1-4βGal1-4Glc trisaccharide (Fig. 3c). The binding of fluorescent SaroL-1 was of the same order as the one observed on GUVs containing natural wild type Gb3 mixture from porcine, indicating the absence of effect of the artificial linker (Fig. 3c). In addition to binding to the surface of GUVs, SaroL-1 formed clusters, probably through multivalent recruitment of glycolipids, and membrane invaginations were observed in association with these clusters. These observations corroborate previous findings in other systems of multivalent lectins and glyco-decorated GUVs [30-33]. In particular, invaginations are consistent with glycolipid dynamics induced by the clustering of sugar heads induced by the binding of multivalent lectins.

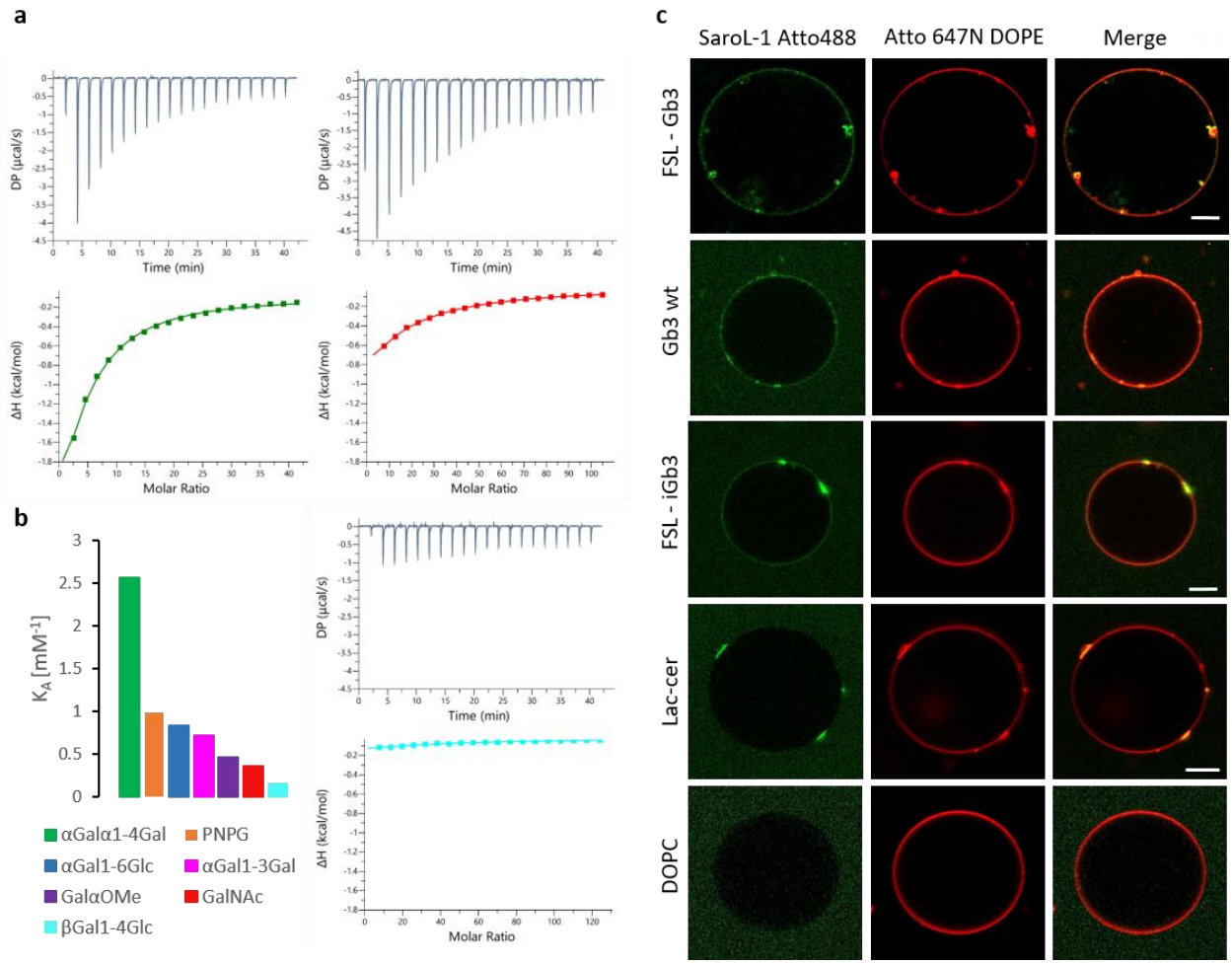


Figure 3. SaroL-1 recognizes αGal epitopes. a) *Representative ITC isotherms of SaroL-1 with αGal1-4Gal (green), GalNAc (red), and lactose (βGal1-4Glc) (cyan),* ***b)*** *Comparison of K_A values of various binding partners for SaroL-1,* ***c)*** *200 nM of SaroL-1 (green) binds to GUVs (red; fluorescent lipid Atto 647N) functionalised with either FSL-Gb3, Gb3 wt, FSL-iGb3, and lactosylceramide (Lac-cer). SaroL-1 induces tubular membrane invaginations in some cases, as visible for FSL-Gb3 GUVs and SaroL-1 clustering on GUVs, as visible for Gb3 wt, FSL-iGb3 and Lac-Cer GUVs. GUVs without functional group (DOPC) serves as negative control and show no binding of SaroL-1. The GUVs were composed of DOPC, cholesterol, glycolipid of choice, and membrane dye to the molar ratio of 64.7:30:5:0.3, respectively. Scale bars are 10 µm.*

SaroL-1 bound to a lesser extent to GUVs containing FSL-iGb3 that presents the αGal1-3βGal1-4Glc epitope. In this case, some clustering of SaroL-1 was also observed at the surface of the GUVs, but no membrane invaginations were formed. Finally, only very weak binding of labelled SaroL-1 was observed on lactosylceramide-containing GUVs (Fig. 3c) that present a βGal terminal sugar. No binding of SaroL-1 was observed to negative controls, namely DOPC GUVs without glycolipids (Fig. 3c) nor to GUVs decorated with the fucosylated oligosaccharides FSL-A and FSL-B, containing the blood group A and B trisaccharide,

respectively (Supplementary Fig. 5). These oligosaccharides do have αGalNAc and αGal, respectively, but the presence of neighboring fucose prevented the binding by SaroL-1.

Binding of SaroL-1 to H1299 cells

The interactions of SaroL-1 with human cells were investigated on the human lung epithelial cell line H1299, a non-small cell lung cancer (NSCLC) characterized by increased cell surface Gb3 expression [34]. We monitored the binding of Cy5-labeled SaroL-1 (SaroL-1-Cy5) to cell surface receptors by flow cytometry (Fig. 4a, b) and its intracellular uptake by confocal imaging (Fig. 4c). In flow cytometry analysis, two different concentrations of SaroL-1 were applied (2 µg/mL and 5 µg/mL). Cells were incubated with SaroL-1-Cy5 for 30 minutes on ice, then the unbound lectin was washed away to reduce unspecific signal, and fluorescence intensity was measured directly with FACS Gallios. The flow cytometry analysis revealed a strong binding of the protein to the cell surface in a dose-dependent manner. Fig. 4a shows a remarkable shift in fluorescence intensity for the samples treated with 2 µg/mL and 5 µg/mL of SaroL-1 (cyan and orange histograms) compared to the negative control (red histogram) after 30 minutes of incubation without reaching signal saturation.

To inhibit the conversion of ceramide to glucosylceramide and accordingly the subsequent biosynthesis of Gb3, H1299 were incubated with PPMP, a chemical inhibitor of glucosylceramide synthase (GCS) activity, used to deplete Gb3 expression. H1299 cells were incubated with 2 µM PPMP for 72 hours before flow cytometry analysis. Subsequently, cells were stimulated with 2 µg/mL or 5 µg/mL of fluorescently labeled SaroL-1 (SaroL-1-Cy5) for 30 minutes on ice (Fig. 4b) and samples were analyzed as described above. Histograms of fluorescence intensities revealed a significant reduction in SaroL-1 binding to the plasma membrane compared to Fig. 4a for both tested concentrations. These results suggest a crucial role of the glycosphingolipid Gb3 as a cell surface receptor for SaroL-1.

Subsequently, H1299 cells were imaged with confocal microscopy to investigate the intracellular uptake of SaroL-1. For these experiments, the concentration of SaroL-1 was set to 10 µg/mL to moderately increase fluorescence signal intensity. After 30 minutes and 1 hour of incubation at 37°C (Fig. 4c), the internalization of SaroL-1-Cy5 into H1299 cells became visible, as shown by the fluorescent signal in red. SaroL-1 seemed widely distributed in the intracellular environment and at the plasma membrane (white arrows) at both time points. In conclusion, our observation made by flow cytometry and fluorescence imaging confirm that SaroL-1 can interact with human cells and induce its intracellular uptake.

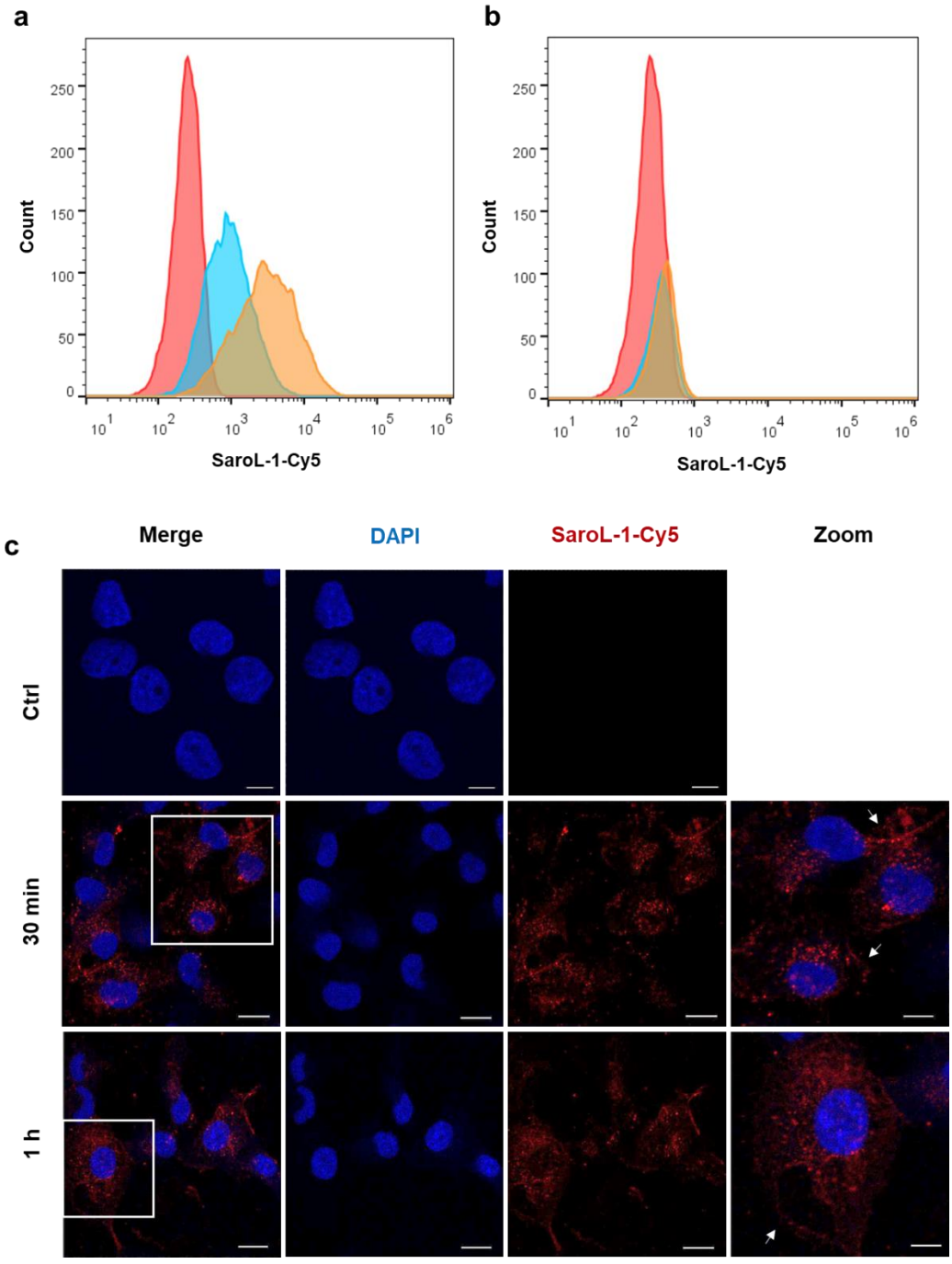


Figure 4. SaroL-1 shows dose-dependent binding and intracellular uptake into H1299 cells. *a) Histogram of fluorescence intensity of H1299 cells incubated for 30 min at 4°C with increasing concentrations of SaroL-1-Cy5 (red: negative control, cyan: 2 µg/mL, orange: 5 µg/mL). The shift in fluorescence intensities indicated that SaroL-1 binds to the H1299 cell surface. b) Histogram of fluorescence intensity of H1299 cells pre-treated for 72 h with the GSL synthesis inhibitor PPMP and incubated for 30 min at 4°C with increasing concentrations of SaroL-1-Cy5 (red: negative control, cyan: 2 µg/mL, orange: 5 µg/mL). In the absence of Gb3, the binding of SaroL-1 to the plasma membrane was remarkably reduced. c) Confocal imaging of H1299 human lung epithelial cells incubated with 10 µg/mL Cy5-conjugated SaroL-1 (red) for different time points at 37°C. The fluorescent signals accumulate at the plasma membrane and in the intracellular space of treated cells. Nuclei were counterstained by DAPI. Scale bars represent 10 µm.*

Crystal structure of SaroL-1 in complex with ligands

Crystallisation experiments were performed with SaroL-1 in complex with different αGal- and GalNAc-containing mono-,di- and tri-saccharides. Several crystals were obtained, those of the complex SaroL-1/GalNAc and SaroL-1/Gb3 trisaccharide showed suitable diffraction and datasets were collected at 1.7 Å and 1.8 Å, respectively (Supplementary Table 3). Attempts to solve the crystal structures by molecular replacement were unsuccessful. Thus the selenium-methionine variant of SaroL-1 was expressed, purified, and crystallised for experimental determination of the phases. A multi-wavelength anomalous diffraction (MAD) dataset was collected at 2.3 Å and used to solve and refine the structure of SeMet SaroL-1 (Supplementary Table 3) (PDB code 7QE3). The latter was subsequently used to solve the structure of the complexes of SaroL-1 with GalNAc (PDB code 7QE4) and the Gb3 trisaccharide (PDB code 7R55) by molecular replacement.

The overall structure of SaroL-1/GalNAc consists of two monomers in the asymmetric unit. They are highly similar (RMSD = 0.52 Å) and do not present extensive contact, confirming the monomeric state of the lectin in solution. SaroL-1 is composed of two domains, a β-trefoil domain at the *N*-terminal region shown in blue and an elongated domain (green) consisting of seven β-strands forming a twisted β-sheet (Fig. 5a). Clear electron density for three GalNAc monosaccharides is observed in the lectin domain, corresponding to the three sites α, β and γ, classically observed in the β-trefoil structure (Supplementary Fig. 6). In all sites, both α- and β-anomers of the GalNAc monosaccharides are present, with stronger occupancy for the α anomer. The three binding sites share strong sequence similarities, albeit with some variations in the amino acids and contacts (Fig. 5c and Supplementary Table 4). The axial O4 of αGalNAc establishes hydrogen bonds with the side chain of conserved His and Arg residues in all sites. This Arg also interacts with the O3 hydroxyl establishing fork-like contacts with two adjacent oxygens of the sugar ring. The O6 hydroxyl is hydrogen-bonded to the main chain of a conserved Gly residue and to the side chain oxygen of a more variable Asn/Asp/Glu residue. Several water molecules are also involved in bridging the protein and the carbohydrate. Moreover, each GalNAc is stabilized in the binding pocket by a C-π-stacking interaction between its hydrophobic face and the aromatic ring of His (site α and β) or Tyr (site γ). The N-acetyl group of GalNAc establishes mostly water-mediated contacts, confirming that it is not crucial for affinity.

The complex of SaroL-1 with the Gb3 trisaccharide presents the same packing as the one with GalNAc. The protein structure is also similar in both complexes albeit with a variation in the fold of the aerolysin β-sheets that present a small kink in its medium region, resulting in an angular deviation of approximately 17° (Supplementary Fig. 7e). The terminal α-galactose of the trisaccharide occupies the exact location and establishes the same contact as GalNAc in the other complex. Electron density was observed in binding sites β and γ only (Fig. 5b and Supplementary Fig. 6). Site α is not occupied, presumably because of close proximity with the neighboring monomer in the crystal. The second galactose residue (βGal) is perpendicular to His98 in site β and Tyr146 in site γ with hydrogen bonds involving ring oxygen O5 of galactose (Fig. 5d). In both cases, an acidic amino acid (Asp43 in site β and Glu92 in site γ) bridges

between the two Gal residues by establishing two strong hydrogen bonds between O6 of αGal and O2 of βGal. Finally, the reducing Glc also participates in the hydrogen bond network through hydroxyl O2 interacting with Asn42 (site β) or Lys91 (site γ). Several bridging water molecules are also involved in the binding network.

The β-trefoil domain of SaroL-1 belongs to the Mytilec-like class of lectins and the above-cited His and Gly belong to the HPXGG sequence motif conserved in this class (Fig. 2b) [21]. The *N*-terminal domain of SaroL-1 demonstrates strong structural similarity with the β-trefoil of the lectins from *Mytilus galloprovincialis* (Mytilec) [21], *Crenomytilus grayanus* (CGL) [20] and synthetic construct Mitsuba [22] with a sequence identity of 34%, 33% and 41%, and RMSD = 0.66 Å, 0.69 Å and 0.81 Å, respectively (Fig. 5e and Supplementary Fig. 7a-d). The CGL structure has been obtained in a complex with Gb3 with the trisaccharide fully visible only in one of the three binding sites. The location of the terminal αGal is similar in CGL and SaroL-1, while the other part of the trisaccharide shows a significant variation as demonstrated in the superimposition of SaroL-1 and CGL complexes (Fig 5e).

Although the sequence of the *C*-terminal elongated β-sheet domain of SaroL-1 has no similarity to those from known structures, its structure is very similar to the pore-forming region of the aerolysin-type β-PFTs, such as LSL a *Laetiporus sulphureus* lectin [12] and the ε-toxin of *Clostridium perfringens* [35,36] (Fig. 5f). However, sequence identities are about 20%. In its pro-aerolysin state, i.e., the solution state, the β-PFT fold is characterized by an extended shape consisting of long and short β-strands, creating two main domains [37].

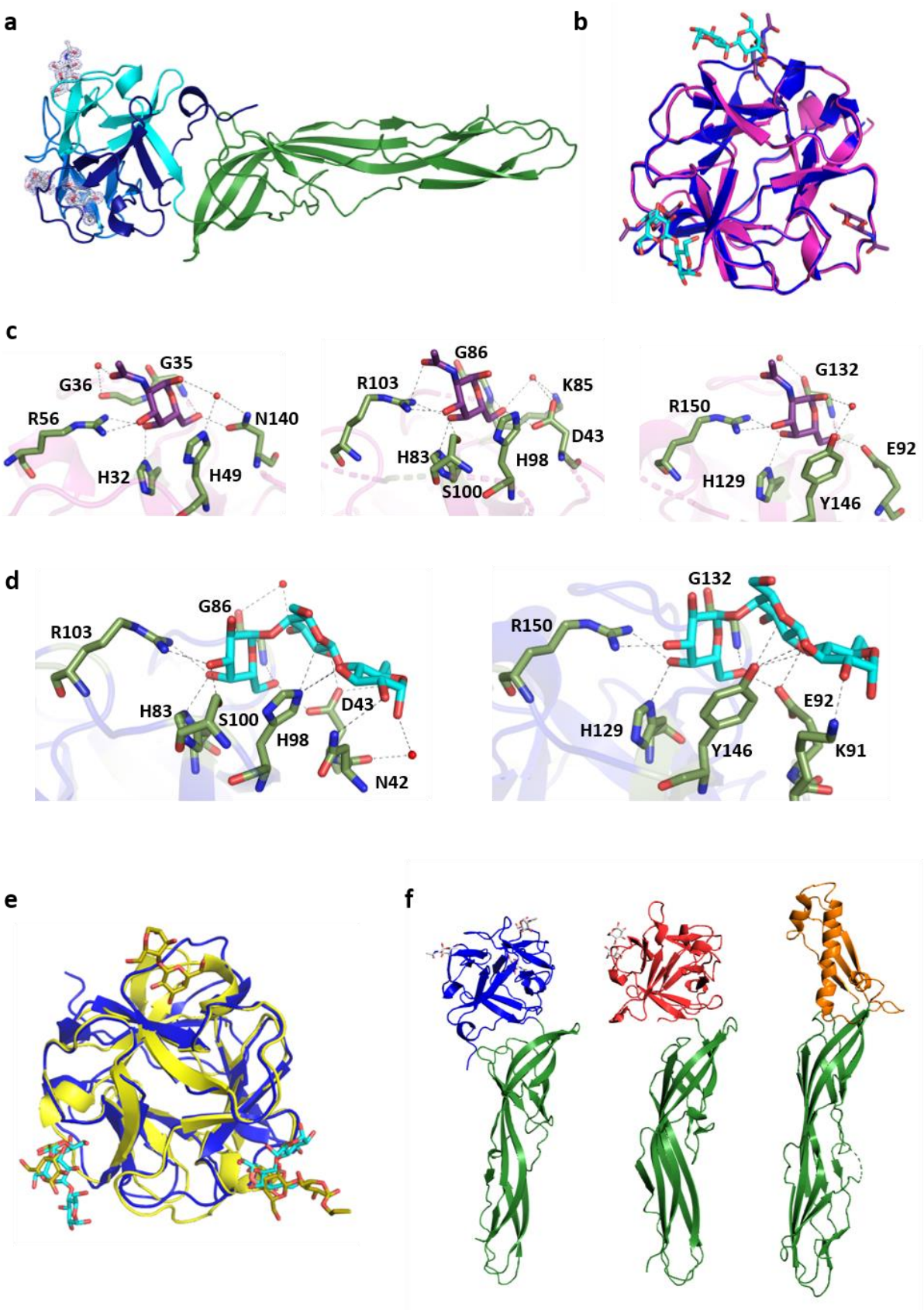


Figure 5: Crystal structure of SaroL-1. *a) Cartoon representation of monomeric SaroL-1 in complex with GalNAc. β-trefoil-domain colored in blue and aerolysin domain in green. The GalNAc ligands are displayed in their electron density map as sticks. b) Superimposition of β-*

trefoil lectin domains, (7QE4, light magenta), (7R55, blue) in complex with 3 molecules of GalNAc (violet) and 2 molecules of Gb3 (cyan). c) Zoom on α, β and γ binding sites with GalNAc (violet) polar contacts are represented as dashed lines and bridging water molecules as red spheres. d) Zoom on the interactions with Gb3 (cyan) in binding β and γ sites, polar contacts are represented as dashed lines and bridging water molecules as red spheres. e) Overlay of β-trefoil domains of SaroL-1 (blue) in complex with Gb3 (cyan) and of monomeric CGL (5F90) (yellow) in complex with Gb3 and αGal1-4Gal (yellow). f) Comparison of the structures of monomeric SaroL-1, pore-forming lectin LSL (1W3A) and ε-toxin (1UYJ) from left to right. The pro-aerolysin domain is colored in green and the membrane-binding domain in blue (SaroL-1), red (LSL) and orange (ε-toxin).

Pore-forming property of SaroL-1

The presence of the hemolytic/pore-forming domain indicates that SaroL-1 could form pore-like structures upon membrane binding, which would fit with our first observation of the alteration of glycolipid dynamics described above. To test our hypothesis of pore-formation, we incubated wt Gb3-containing GUVs (indicated by the fluorescent lipid DOPE-Alexa647N; red color) with 200 nM unlabeled SaroL-1 and 3 kDa dextran labeled by Alexa Fluor ™ 488 (dextran-AF488; green color) and monitored dextran influx into GUVs for two hours by using confocal microscopy (Fig. 6a). After 30 minutes of incubation, 45% of the 185 total observed GUVs in the experiments were filled with dextran-AF488. The number of GUVs filled with dextran-AF488 steadily increased over time and reached 69% out of 178 GUVs after two hours of incubation with SaroL-1 (Fig. 6a). The control group of wt Gb3-containing GUVs incubated together with dextran-AF488 but without SaroL-1 showed less than 1% influx of dextran of total 393 GUVs after 2 hours. The 3 kDa size of fluorescently labeled dextran corresponds to a hydrodynamic radius of 18 Å [38]. The slow equilibration observed for dextran entering in the GUVs is consistent with the diameter range of 1 to 4 nm reported for the pore of aerolysin-like toxins [39].

The role of the lectin domain in pore formation was investigated by adding the soluble galactose analog PNPG that competes with Gb3 in the binding site. We pre-treated 200 nM SaroL-1 with 10 mM PNPG for 15 minutes at RT in PBS buffer and then added the solution to wt Gb3-containing GUVs. In the PNPG-treated group, the influx of dextran was fully inhibited for the first 30 minutes, as none of 435 of total GUVs were filled with dextran. However, after 2 hours, 4% of total 466 GUVs were filled with dextran (Fig. 6b). Therefore, it was demonstrated that the Gb3-binding activity of the lectin domain is necessary for pore formation. Obtained results indicate that SaroL-1 plays a role in dextran-AF488 influx to wt Gb3-containing GUVs compared to the negative control and prefers the glycosphingolipid Gb3 over PNPG.

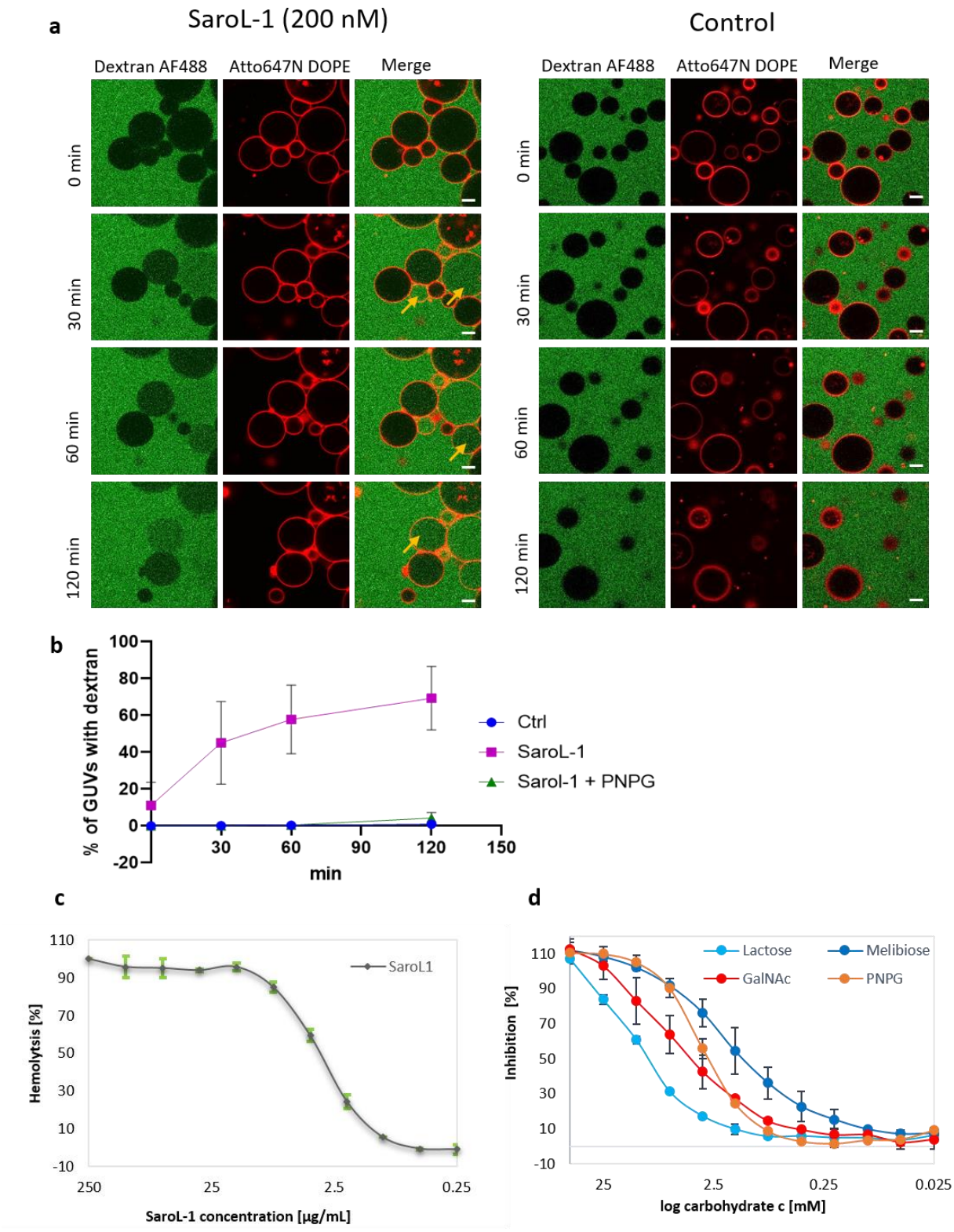


Figure 6. Pore-forming and hemolytic activity of SaroL-1. *a) SaroL-1 (unlabeled, 200 nM) triggers the influx of dextran-AF488 (green) into wt Gb3-containing GUVs (red) via its pore-forming activity. In the control group without SaroL-1, there was no visible influx of dextran-AF488 detected. Yellow arrows indicate events of dextran-AF488 influx to wt Gb3-GUVs. The GUVs were composed of DOPC, cholesterol, wt Gb3, and membrane dye to the molar ratio of 64.7:30:5:0.3, respectively. The scale bars represent 10 µm. b) Kinetics of SaroL-1 driven*

dextran-AF488 influx to wt Gb3-GUVs. Two-way ANOVA followed by Dunnett's. Mean values ± SD are shown. Data represent three independent experiments, n=3. The molecular weight of fluorescently labelled dextran is 3 kDa. The total amount of control GUVs was at 0 min – 225 GUVs, 30 min – 344 GUVs, 60min – 349 GUVs and 120 min – 393 GUVs. For SaroL-1 experiment with wt Gb3-GUVs, total amount of GUVs was 0 min – 230 GUVs, 30 min – 185 GUVs, 60 min – 183 GUVs and 120 min – 178 GUVs. In PNPG treated group there were total amount of GUVs at 0 min – 322 GUVs, at 30 min – 435 GUVs, 60 min – 447 GUVs and at 120 min 446 GUVs. c) Graphical visualization of hemolytic activity of SaroL-1 where IC_{50} was estimated as 6.3 μg/mL. The hemolytic activity was calculated as a percentage compared to the negative control, whereas the highest absorbance observed was considered as 100 % hemolysis. d) Inhibition of the hemolytic activity of SaroL-1 by PNPG, GalNAc, melibiose and lactose. The values were compared to the hemolytic activity of SaroL-1 in the same concentration without the presence of any inhibitor and subsequently recalculated considering 0% inhibition at the lowest value for each inhibitor.

When adding SaroL-1 to rabbit red blood cells, hemolysis was observed instead of hemagglutination classically induced by multivalent lectins [40]. The hemolysis was confirmed by measuring the absorbance at 540 nm (Fig. 6c). SaroL-1 appeared to be a potent hemolytic agent since a concentration of 1 μg/mL is sufficient for damaging erythrocytes, whereas nearly complete hemolysis was observed at a concentration of 8 μg/mL. Analysis of the red blood cells by microscopy demonstrated the occurrence of almond-like shaped erythrocytes after a few minutes of exposition to the lectin (Supplemental Fig. 8). The number of erythrocytes decreases with increasing incubation time as cells are possibly lysed due to morphological changes. The role of the lectin domain in the pore formation was assayed by pre-incubating SaroL-1 with several carbohydrates in different concentrations before measuring the hemolysis. Melibiose and PNPG appeared as the most efficient competitors (Fig. 6d) with complete inhibition of hemolysis at 10 mM (IC_{50} (melibiose) = 1.8 mM, IC_{50} (PNPG) = 3.3 mM). In agreement with ITC and X-rays data, GalNAc was also an efficient inhibitor (IC_{50} = 6.5 mM), while lactose action was weaker (IC_{50} = 14.5 mM).

Toxicity of SaroL-1 towards cancer cells

As the aerolysin domain may cause osmotic lysis and cell death [37], we determined the cytotoxic effect of SaroL-1 on H1299 cells after 24 hours treatment using a cell proliferation assay (MTT). The assay is based on the cleavage of tetrazolium salt MTT to form a formazan dye by metabolic-active cells, suitable for quantifying cell proliferation and viability. SaroL-1 shows cytotoxic activity in fetal calf serum (FCS)-containing medium, and the percentage of viable cells after 24 hours decreased by half at a concentration of 10 μg/mL. Based on the results depicted in Fig. 7a, increasing concentrations of SaroL-1 decreased the proliferation and viability of human epithelial cells *in vitro* in a dose-dependent manner. No significant cytotoxicity was observed upon treatment with PBS as negative control.

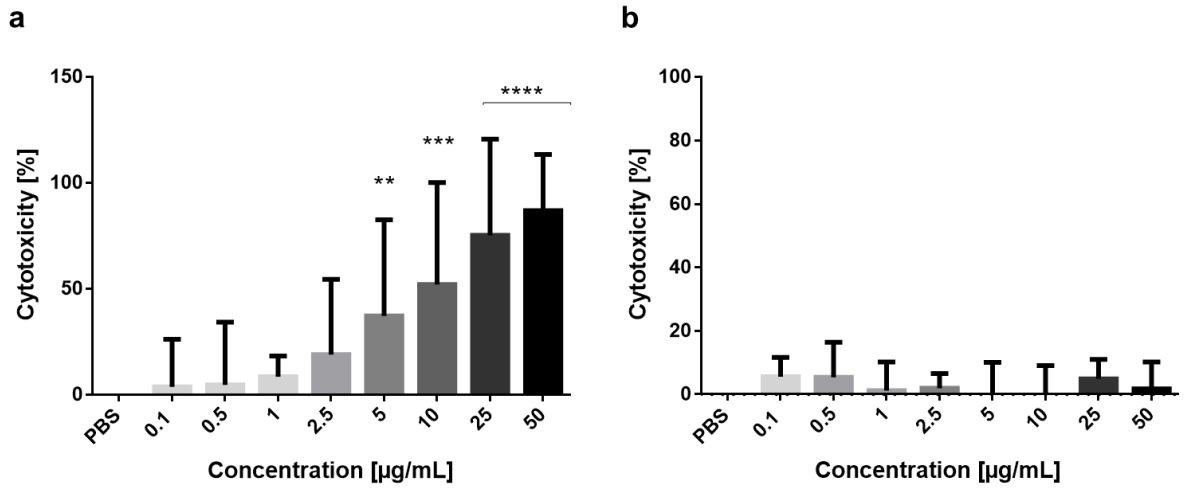


Figure 7. Cytotoxic activity of SaroL-1 against H1299 cells. *a) Dose-dependent increase of cytotoxicity following the addition of purified SaroL-1 in a standard cell proliferation assay (MTT) demonstrating increment of cytotoxicity after 24 hours of incubation compared to treatment with PBS. Cell viability is reduced by approximately 87% after stimulation with SaroL-1. b) The soluble sugar PNPG inhibits SaroL-1 cytotoxicity. H1299 were incubated with increasing concentrations of SaroL-1 pre-treated with 10 mM PNPG. Cell proliferation assay (MTT) was used to assess SaroL-1´s cytotoxic activity after 24 hours in comparison treatment with PBS. The results indicate that cell viability is preserved when SaroL-1 glycan-binding sites are saturated with soluble 10 mM PNPG. Differences to the control were analyzed for significance by using two-tailed unpaired t-test. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.*

Additionally, we saturated the carbohydrate binding pockets of SaroL-1 with PNPG to prove that a glycan-driven binding and internalization of the protein is essential to exert its cytotoxicity on cells. Therefore, increasing concentrations of the lectin (0.1, 0.5, 1, 2.5, 5, 10, 25, 50 µg/mL) were pre-incubated with 10 mM PNPG for 15 minutes at RT, then the solution was added to cells in a standard MTT assay. Strikingly, in the presence of PNPG, cell death was largely reduced by more than 90% after 24 hours of incubation (Fig. 7b). Treated cells preserved viability, even in the presence of high protein concentrations. These results indicate that SaroL-1 activity is efficiently inhibited by 10 mM PNPG, leading to a substantial decrease in cell cytotoxicity. Remarkably, we demonstrate that SaroL-1 can exert cytotoxic activity on H1299 cells only upon binding to glycosylated receptors exposed at the plasma membrane.

Undeniably, cell anchorage not only provides the structural support for a cell, but it mediates crucial survival signals for the cells, providing access to nutrients and growth factors. Furthermore, it has been established that processes that modify cell adhesion leading to the loss of cell anchorage may induce cell death [41]. We therefore investigated the possible role of SaroL-1´s hemolytic domain in the disruption of cell adhesion. We proved that SaroL-1, mainly upon binding to the Gb3 receptor, causes a dose-dependent rounding and detachment of cells compared to PBS-treated cells, ultimately leading to cell death (Supplementary Fig. 9).

Model of SaroL-1 transmembrane pore

β-PFTs of the aerolysin-family occur as a monomer in solution and form pores in membranes according to the following steps: they bind to a cell-surface receptor, oligomerise while generating β-hairpins from each of six to seven individual monomers and then, produce a vertical β-stranded pore which varies in size from 20 to 25 Å. To visualize the relative orientation of the lectin- and pore-forming domain of SaroL-1 in the pore architecture, a 3D model was built, using a template selected from crystal structures of a toxin for which both solution-monomeric and pore-heptameric data were available. The ε-toxin of *C. perfringens* [35,36] matched the requirements. Although the primary sequence homology of the pore-forming domains of SaroL-1 and ε-toxin (PDB 1UYJ) is low (12%), the 3D-structures of the monomeric state are strikingly similar (Fig. 5e) and were used for the template alignment (Supplementary Fig. 10). From this, a heptameric SaroL-1 pore was built with SwissModel [42] based on the membrane pore structure of ε-toxin (PDB 6RB9) (Fig. 8b). The lectin domain was then linked on the *N*-terminal extremity of each hairpin, in a conformation bringing all carbohydrate-binding sites towards the surface of the membrane.

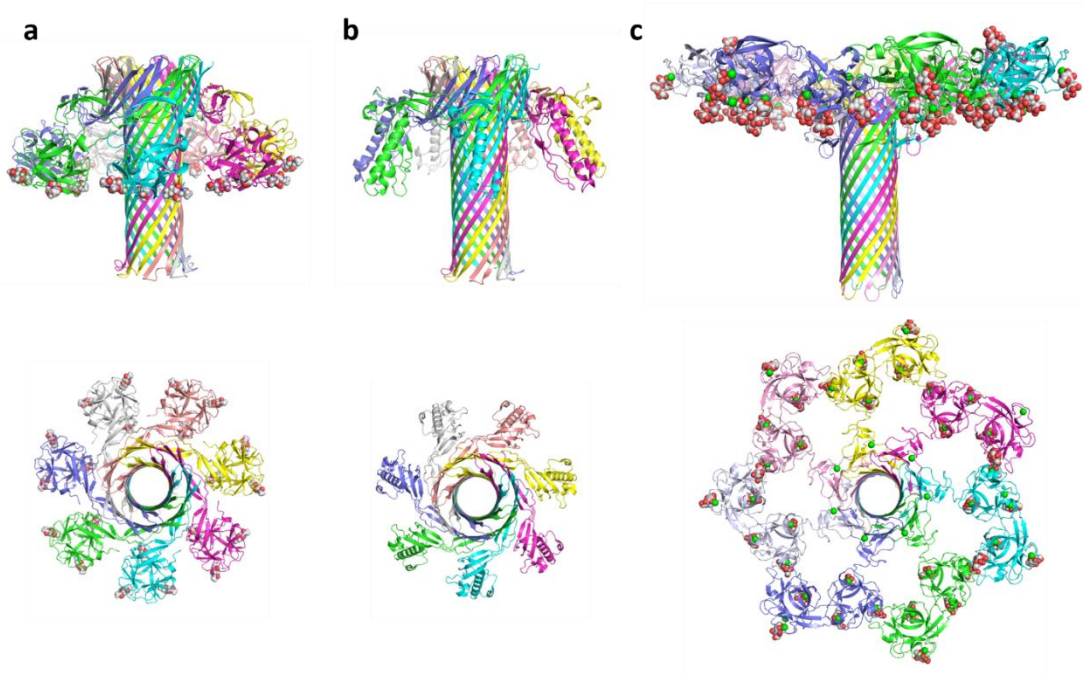


Figure 8. Prediction of SaroL-1 model *a) Preliminary model of membrane-bound heptameric SaroL-1 built by homology modeling using the structure similarity displayed in Fig. 3. b) Crystal structure of heptameric ε-toxin of C. perfringens (PDB 6RB9). c) Crystal structure of heptameric lactose-binding lectin Cel-III from Cucumaria echinata complexed with βGal derivative (PDB 3WT9).*

The resulting preliminary model (Fig. 8a) confirms that the lectin domain can adopt an orientation that locates all 21 αGal binding sites of the heptamer in a plane, which would correspond to the surface of the glycosylated host cell membrane. Extensive molecular

dynamics studies would be necessary to explore all possible orientations of the lectin domain, as well as the molecular mechanisms occurring during pore-formation, as previously performed for aerolysin [37].

Discussion

Lectins have been primarily studied for their ability to bind to complex carbohydrates, with applications in biotechnology, histology, and diagnostics [43,44]. Their role in signaling has been recently emphasized due to their ability to cross-link membrane glycoproteins and glycolipids [33]. The interest is now high for identifying lectins associated with an additional domain carrying a particular function that can be specifically targeted to glycan-bearing cells. Furthermore, the development in protein engineering and synthetic biology highlights lectins and carbohydrate-binding modules as building blocks in designing novel molecules such as artificial and multivalent dimers increasing affinity to sialic acid [45] or Janus lectins that associate two domains with different carbohydrate specificity [46,47]. In this context, β-trefoil lectins are pretty attractive: they are small and compact, very stable due to their hydrophobic core and multivalent through tandem repeats, thereby increasing avidity for their substrates.

Furthermore, they can be engineered for adapting their specificity to a precise glycan target [48] or enhancing their internal symmetry [22]. The TrefLec database contains predicted sequences for each of the 12 β-trefoil lectin classes, resulting from mining translated genomes. This resource offers the option of searching the tens of thousands of β-trefoil candidates for new lectin sequences and possibly associated with the additional functional domain(s). This approach was validated by identifying Mytilec family members previously found only in mollusks and other marine animals, for instance, in the lower eukaryote *Salpingoeca rosetta*. A new Mytilec sequence was selected from TrefLec, consisting of a Mytilec-like and an aerolysin domain despite the poorly informative initial UniProt record mentioning an uncharacterized protein.

The identification of a novel β-PFT with putative membrane-binding domain specific for the αGal epitope was completed by compelling evidence of carbohydrate-mediated interactions. Indeed, β-PFT aerolysin-like toxins have a conserved pore-forming architecture and a large variety of membrane-binding domains. *Aeromonas* aerolysin binds to a glycosylphosphatidylinositol-anchored protein [49], and the *Clostridium* ε-toxin binds to proteins associated with lipid rafts [50]. Several β-PFTs have a lectin-type membrane-binding domain, such as CEL-III in the sea cucumber, including 35 binding sites for galactose in its hemolytic heptameric form [11]. The fungal β-PFT from *L. sulfureus* has a lectin domain specific for lactose that also occurs in many other fungi [51]. Plant β-trefoil lectins of the Amaranthin class were also predicted to be associated with aerolysin in many plants [24]. SaroL-1 is, however, the first identified β-PFT with a lectin-type membrane-binding domain specific for an epitope associated with several cancers. Its specificity and biological activity were established with different biophysical and cell biological approaches. SaroL-1 showed preferences for α-galactosides with much stronger affinity than for β-galactosides, such as lactose. The specific

binding to the glycosphingolipid Gb3, a cancer biomarker existing in Burkitt's lymphoma [52], ovarian [53], colorectal [54], breast [55] and pancreatic cancer [56] is of particular interest. As recently reviewed [57], several lectins have been identified as being specific to Gb3. Therefore, we tested SaroL-1 on the Gb3-positive H1299 lung epithelial cell line and observed binding and intercellular uptake. However, the binding was vastly diminished on Gb3-depleted cells pointing to a Gb3-dependent mode of interaction. Some Mytilec-like lectins, such as CGL [20,26] and Mytilec [58] were shown to be cytotoxic to certain cancer cell lines through Gb3-mediated interactions.

Similarly, SaroL-1 binding induced a dose-dependent cytotoxic effect resulting in the detachment of H1299 cells from the cell culture dish. Despite of the fact that already sub-micromolar concentrations effectively killed 50 % of cells the cytotoxic effect is markedly inhibited by pre-incubation of SaroL-1 with the αGal-containing competitor PNPG, which once more confirms a glycan-dependent mode of binding and action.

The crystal structures of SaroL-1 in complex with GalNAc and Gb3 trisaccharide provided the atomic basis for its specificity. They confirmed the presence of an aerolysin domain with a structure similar to LSL and aerolysin. We could propose a model of the entire pore using ε-toxin from *C. perfringens* as a template. The ability of SaroL-1 to form a pore was experimentally validated through hemolysis assays and visualized by the influx of labelled dextran penetrating into giant unilamellar vesicles. It could also be demonstrated that SaroL-1 is cytotoxic and induces cell detachment. The lectin-binding step is necessary for the pore-formation since the pre-incubation with a high affinity galactose derivative diminishes or abolishes hemolysis, GUVs penetration, and cytotoxicity. Furthermore, SaroL-1 did not bind to H1299 cells lacking Gb3 expression, nor it exerts cytotoxicity when galactose derivatives are present in solution.

Recently, PFTs gained interest for their application as biotechnological sensors and delivery systems. Aerolysin is a well-characterized nanopore, wild-type and mutated forms demonstrated promising results as nanopores for direct sensing of nucleotide acids and proteins [59,60], or even charged polysaccharides [61]. The applications for early diagnosis, such as detecting circulating cancer cells, are promising [62]. Aerolysin was demonstrated to identify most of the twenty proteinogenic amino acids [63]. Together with the characterization of new aerolysins from biodiversity, these findings pave the way for nanopore-based sequencing of proteins. With an increasing number of possible applications for PFTs, SaroL-1 is a suitable candidate for further exploration in the field of nanopore technology and cancer diagnosis and treatment.

Material and methods

Screening of trefoil lectins

Trefoil structures in UniLectin3D are grouped together in a class if they share 20% of sequence similarity with any other structure of the same class. Within each trefoil class, the lectin lobes

have been manually selected based on PDB 3D structures and aligned with the Muscle software [64]. When a protein has several associated 3D structures, only one of them is selected.

Twelve distinct trefoil lectin classes were defined. The sequences defining the trefoil lobes in each class are extracted individually and aligned together. This resulted in collecting 12 multiple alignments of lobe conserved regions. These were used as input to HMMER-hmmbuild [65], using default parameters and symfrac (the minimum residue fraction necessary to set a position as consensus in the alignment) set at 0.8, to generate twelve characteristic profiles for each class. HMMER is an established tool that produces Hidden Markov Models (HMM) [66], that is, a signature/profile for a group of similar protein sequences. Using these 12 trefoil lectin class profiles, potential lectin sequences were predicted in UniProtKB (UniProt April 2019)[67] and NCBI-nr (non-redundant July 2019) [68]. TrefLec was updated in 2021 and, at the time of writing, includes more β-trefoil candidates. Nonetheless, SaroL-1 was predicted with 2019 data. The two large protein sequence datasets (millions of sequences) were screened with the 12 HMM profiles using HMMER-hmmsearch, with default parameters and a p-value below 0.01 [65].

Further filtering was applied to discard highly similar (>98%) proteins of the same species (only one instance was kept), as well as identified domains shorter than 10 amino acids. The HMMER tool generates a statistical score for each predicted lectin. Since each sequence is compared to 12 profiles, a prediction is assigned 12 scores, which, by nature of the score definition (bit score), are not comparable across the 12 classes. That is why, to use a single cut-off value in TrefLec, a normalised prediction/similarity score (between 0 and 1) was defined. It reflects the similarity between the predicted lobe sequence and the matched profile lobe consensus sequence. In this way, for each sequence, the 12 prediction results can be ranked, and the top one is selected following published procedure [69].

Gene design and cloning

The original gene sequence of the uncharacterized protein from *Salpingoeca rosetta* (strain ATCC 50818 / BSB-021) (NW_004754913.1) was obtained from the UniLectin3D database. The gene *sarol-1* was ordered from Eurofins Genomics (Ebersberg, Germany) after codon optimization for the expression in the bacteria *Escherichia coli*. The restriction enzyme sites of NdeI and XhoI were added at 5' and 3' ends, respectively. The synthesized gene was delivered in plasmid pEX-A128-SaroL-1. This plasmid and the pET-TEV vector [70] were digested with the NdeI and XhoI restriction enzymes to ligate *sarol-1* in pET-TEV to fuse a 6-His Tag cleavable with TEV protease at the *N*-terminus of SaroL-1. After transformation by heat shock in *E. coli* DH5α strain, a colony screening was performed, and the positive plasmid was amplified and controlled by sequencing.

Protein expression

E. coli BL21(DE3) cells were transformed by heat shock with pET-TEV-SaroL-1 plasmid prior preculture in Luria Broth (LB) media with 25 μg/mL kanamycin at 37°C under agitation at 180 rpm overnight. The next day, 10 mL of preculture was used to inoculate 1 L LB medium with

25 µg/mL kanamycin at 37°C and agitation at 180 rpm. When the culture reached OD_{600nm} of 0.6 - 0.8, the protein expression was induced by adding 0.1 mM isopropyl β-D-thiogalactoside (IPTG), and the cells were cultured at 16°C for 20 hours. The cells were harvested by centrifugation at 14000 × *g* for 20 min at 4°C and the cell paste was resuspended in 20 mM Tris/HCl pH 7.5, 100 mM NaCl, and lysed by a pressure cell disruptor (Constant Cell Disruption System) with a pressure of 1.9 kBar. The lysate was centrifuged at 24 000 × *g* for 30 min at 4°C and filtered on a 0.45 µm syringe filter prior to loading on an affinity column.

Seleno-Methionine Protein Expression

Substitution of the methionine of SaroL-1 by selenomethionine was performed according to protocol [71]. *E. coli* BL21(DE3) pET TEV-SaroL-1 were precultured in 5 mL LB Broth media with 25 µg/mL kanamycin at 37°C with agitation at 180 rpm until OD_{600nm} reached 0.6. The pre-culture was centrifuged (10 min, 3000 × *g*, 4°C) and washed twice by 1 × M9 medium. 1 × M9 media is composed of salts (Na_2HPO_4-$2H_2O$, KH_2PO_4, NaCl, NH_4Cl) enriched with 0.4% glucose, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1 µg thiamine, and sterile water. The cells were cultured in 50 mL 1 × M9 minimal salts media and 25 µg/mL kanamycin at 37°C with agitation at 180 rpm overnight. Then, the preculture was transferred into 1 L 1 × M9 medium with the addition of 25 µg/mL kanamycin at 37°C and agitation at 180 rpm. When OD_{600nm} reached 0.6-0.8, the mixture of the amino acids (Lys, Phe, Thr, Ile, Leu, Val, SeMet) was added and incubated for 15 min at 37°C with agitation at 180 rpm prior to induction and treatment of the cells as described above for the wild-type protein. The cell pellet was stored at -20°C prior to purification.

Protein purification

The sample was loaded on 1 mL HisTrap column (Cytiva) pre-equilibrated with 20 mM Tris/HCl pH 7.5, 100 mM NaCl (Buffer A). The column was washed with Buffer A to remove all contaminants and unbound proteins. The SaroL-1 was eluted by Buffer A in steps during which the concentration of imidazole was increased from 50 mM to 500 mM. The fractions were analyzed by 12% SDS PAGE and those containing SaroL-1 were collected and deprived of imidazole using a PD10 desalting column (Cytiva). The sample was concentrated by Pall centrifugal device with MWCO 10 kDa prior to loading onto Enrich 70 column (BioRad) previously equilibrated with Buffer A for further purification. After analysis by SDS-PAGE, the pure protein fractions were pooled, concentrated and stored at 4°C.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed with MicroCaliTC200 (Malvern Panalytical). Experiments were carried out at 25°C ± 0.1°C. SaroL-1 and ligands samples were prepared in Buffer A. The ITC cell contained SaroL-1 in a concentration range from 0.05 mM to 0.16 mM. The syringe contained the ligand solutions in a concentration from 10 to 50 mM. 2 µL of ligands solutions were injected into the sample cell at intervals of 120 s while stirring at 750 rpm. Integrated heat effects were analysed by nonlinear regression using one site binding model (MicroCal PEAQ-ITC Analysis software). The experimental data were fitted to a theoretical curve, which gave the dissociation constant (Kd) and the enthalpy of binding (ΔH).

Protein labelling

SaroL-1 was dissolved at 1 mg/mL in Dulbecco's phosphate-buffered saline (PBS) and stored at 4°C prior to usages. For fluorescent labelling, Cy5 (GE Healthcare) mono-reactive NHS ester and NHS-ester conjugated Atto488 (Thermo Fisher) were used. Fluorescent dyes were dissolved at a final concentration of 1 mg/mL in water-free DMSO (Carl RothGmbH & Co), aliquoted, and stored at -20°C before usage according to the manufacturer´s protocol. For the labelling reaction, 100 µL of lectin (1 mg/mL) was supplemented with 10 µL of a 1 M $NaHCO_3$ (pH 9) solution. Hereby, the molar ratio between dye and lectin was 2:1. The labelling mixture was incubated at 4°C for 90 min, and uncoupled dyes were separated using Zeba Spin desalting columns (7 kDa MWCO, 0.5 mL, Thermo Fischer). Labelled SaroL-1 was stored at 4°C, protected from light.

Composition and preparation of giant unilamellar vesicles (GUVs)

GUVs were composed of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), cholesterol (both AvantiPolar Lipids, United States), Atto 647N 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE; Sigma Aldrich), and one of the following glycolipids at a molar ratio of 64.7:30:0.3:5. The glycolipids are globotriaosylceramide (Gb3, Matreya), FSL-Gb3 (Function-Spacer-Lipid with globotriaosyl saccharide), and corresponding FSL-isoGb3 (Function-Spacer-Lipid with iso-globotriaosyl saccharide) synthetized as previously described [72,73], FSL-A(tri) (Function-Spacer-Lipid with blood group A trisaccharide; SigmaAldrich), FSL-B(tri) (Function-Spacer-Lipid with blood group B trisaccharide; Sigma Aldrich) or lactosylceramide (LC, Sigma Aldrich).

GUVs were prepared by the electroformation method as earlier described [74]. Briefly, lipids dissolved in chloroform of a total concentration of 0.5 mg/mL were spread on indium tin oxid-covered (ITO) glass slides and dried in a vacuum for at least one hour or overnight. Two ITO slides were assembled to create a chamber filled with sucrose solution adapted to the osmolarity of the imaging buffer of choice, either HBSS (live-cell imaging) or PBS (GUVs only imaging). Then, an alternating electrical field with a field strength of 1 V/mm was implemented for 2.5 hours at RT. Later we observed the GUVs in chambers manually built as described [74].

Imaging of SaroL-1 binding to GUVs

Samples of GUVs and SaroL-1 were imaged using a confocal fluorescence microscope (Nikon Eclipse Ti-E inverted microscope equipped with Nikon A1R confocal laser scanning system, 60x oil immersion objective, NA = 1.49 and four laser lines: 405 nm, 488 nm, 561 nm, and 640 nm). Image acquisition and processing were made using the software NIS-Elements (version 4.5, Nikon) and open-source Fiji software (https://imagej.net/software/fiji/).

Dextran influx into GUVs

Dextran Alexa Fluor 488 (Dextran-AF488) with a MW of 3 kDa (Thermo Fischer) was added to the observation chamber (0.02 mg/mL) in PBS buffer together with 200 nM SaroL-1. Then wt Gb3-containing GUVs (40 µL) were added for imaging and monitoring of pore-formation. Dextran-AF488 was present in the GUVs' surrounding solution. For PNPG treatment, 200 nM of SaroL-1 was pre-incubated with 10 mM PNPG dissolved in DMSO for 15 min at RT in the observation chamber. Directly after, dextran-AF488 (0.02 mg/mL) and 40 µL wt Gb3-

containing GUVs were added to the observation chamber and monitored using confocal microscopy.

Hemolytic assay (HA) and inhibition of hemolytic activity (IHA)

HA was performed in U-shaped 96-well microtiter plates. SaroL-1 with an initial concentration of 1 mg/ml was diluted by serial 2-fold dilution in Buffer A. Rabbit erythrocytes were purchased from Atlantis France and used without further washing. The erythrocytes were diluted to a 4% solution in 150 mM NaCl. Protein solution and rabbit erythrocytes were mixed together in a 1:1 ratio and incubated for 1 h at 37°C. After the samples were centrifuged, 2600 × *g* for 15 min at RT and the absorbance of the supernatant was measured at 540 nm by a Tecan reader.

IHA was performed in U-shaped 96-well microtiter plates. Various carbohydrates (PNPG, GalNAc, lactose, melibiose) with the initial concentration of 100 mM were serially 2-fold diluted in Buffer A. SaroL-1 with a final concentration in the well 8 µg/mL was added into ligand solution in 1:1 ratio and incubated for 1 hour at RT. 4% rabbit erythrocytes were added to the mixture in a 1:1 ratio and incubated for 1-2 h at 37°C. The samples were centrifuged, 2600 × *g* for 15 min at RT, and the absorbance of the supernatant was measured at 540 nm by a Tecan reader.

Optical microscopy

SaroL-1 (4 µg/mL and 8µg/mL) was added into a 4% solution of rabbit erythrocytes and observed after incubation for 5 and 30 min, respectively, at RT. The samples were observed Zeiss Axioplan 2 microscope with 40x magnification. 4% solution of rabbit erythrocytes was used as a negative control.

Cell culture

The human lung epithelial cell line H1299 (American Type Culture Collection, CRL-5803) was cultured in a complete medium, consisting of Roswell Park Memorial Institute (RPMI) medium supplemented with 10% fetal calf serum (FCS) and 2 mM L-glutamine at 37°C and 5% CO_2. Cells were stimulated with different concentrations of SaroL-1 in a complete medium for indicated time points.

Flow cytometry analysis

H1299 cells were detached with 1.5 mM EDTA in PBS -/-, and 1×10^5 cells were counted and transferred to a U-bottom 96 well plate (Sarstedt AG & Co.). To determine the binding of SaroL-1 to surface receptors, cells were incubated with fluorescently labelled protein for 30 min at 4°C and protected from light, in comparison with PBS-treated cells as a negative control. To deplete cells of glucosylceramide-based glycosphingolipids, they were cultivated 72 hours in the presence of 2 µM DL-threo-1-phenyl-2-palmitoylamino-3-morpholino-1-propanol (PPMP) (Sigma-Aldrich) to inhibit the synthesis of glucosylceramide-based GSLs [75] and incubated with fluorescent SaroL-1 as previously described. Subsequently, cells were centrifuged at 1600 × *g* for 3 min at 4°C to remove unbound lectin. The samples were then washed two times with ice-cold FACS buffer (PBS (-/-) supplemented with 3% FCS (*v/v*). After the last washing step, the cells were re-suspended with FACS buffer and transferred to

FACS tubes (Kisker Biotech GmbH & Co) on ice and protected from light. The fluorescence intensity of treated cells was measured with FACS Gallios from Beckman Coulter. The samples were further analyzed using FlowJo V.10.5.3.

Fluorescence imaging of SaroL-1 binding and cellular uptake by confocal microscopy

Between 4 and 5 × 10^4 cells were seeded on 12-mm glass coverslips in a 4-well plate and cultured for one day before the experiment. Cells were incubated with SaroL-1-Cy5 for indicated time points at 37°C, in comparison with PBS-treated cells as a negative control. Subsequently, cells were fixed with 4% paraformaldehyde for 15 min at RT and quenched with 50 mM ammonium chloride for 5 min. The membrane was permeabilised, and cells were blocked by 0.2% Saponin in 3% BSA in PBS (*w/v*) for 30 min. The cell nuclei were counterstained with DAPI (5 × 10^{-9} g/L), and the samples were mounted on coverslips using Mowiol (containing the anti-bleaching reagent DABCO). Samples were imaged utilising a laser scanning confocal microscope system from Nikon (Eclipse Ti-E, A1R), equipped with a 60x oil immersion objective and a numerical aperture (*NA*) of 1.49. The images were further analyzed using NIS-Element Confocal 4.20 from Nikon and ImageJ 1.52a from Laboratory for Optical and Computational Instrumentation. A minimum of three biological replicates with ≥ 20 cells per condition were analyzed.

Cell proliferation (MTT) assay

To investigate the cytotoxic effect of SaroL-1 on human cells, H1299 cells were treated with increasing concentrations of SaroL-1 for 24 hours in a standard MTT assay. 3.5 × 10^4 cells per well were transferred to a 96-well plate with a U-bottom. The cells were centrifuged at 1600 × *g* for 3 min at RT. The cell pellet was re-suspended in 100 µL of variously concentrated protein solutions (0.1, 0.5, 1, 2.5, 5, 10, 25, 50 µg/mL) and transferred to a 96-well flat-bottomed plate. The cells were incubated for 24 hours at 37°C. Subsequently, 10 µL of MTT labelling solution (MTT Cell Proliferation Kit, Roche) was added to each well, and the cells were incubated for 4 hours at 37°C. Then, 100 µL of the solubilisation reagent was added to each well, and the plate was incubated at 37°C overnight. The next day, the absorbance of the samples was measured at 550 nm using a BioTek microplate reader. To assess SaroL-1 cytotoxicity in the presence of the soluble sugar PNPG, 3.5 × 10^4 cells per well were transferred to a 96-well plate with a U-bottom. Variously concentrated protein solutions (0.1, 0.5, 1, 2.5, 5, 10, 25, 50 µg/mL) were preincubated with 10 mM PNPG dissolved in complete medium, for 15 min at RT. The cells were centrifuged at 1600 × *g* for 3 min at RT. Next, the cell pellet was re-suspended in 100 µL protein solution and transferred to a 96-well flat-bottomed plate. The cells were incubated for 24 hours at 37°C, and the MTT assay was performed as described above. The data was further analyzed using GraphPad Prism software.

Cell detachment assay

To determine the role of SaroL-1 in cell detachment, 3 × 10^5 cells per well were counted and allowed to adhere overnight in a 6-well plate. The next day, cells were incubated for 2, 4, or 8 h with two different concentrations of SaroL-1 (10 µg/mL and 50 µg/mL) diluted in a complete medium compared to a negative control. As a positive control, cells were incubated with trypsin-EDTA for 10 min at 37°C, to induce 100% cell detachment. The number of cells in

suspension at the indicated time points was quantified by analyzing the supernatant with the CytoSmart Corning cell counter.

Statistical analysis

All data in graphs are presented as mean ± standard deviation (SD) and were calculated from the results of independent experiments. Statistical testing was performed with GraphPad Prism software and Microsoft Excel using data of ≥ 3 biological replicates. Statistical differences in independent, identical samples were determined with a two-tailed, unpaired t-test. Tests with a p-value ≤ 0.05 are considered statistically significant and marked with an asterisk (*).p-values ≤ 0.01 are shown as two asterisks (**), ≤ 0.001 are summarized with three asterisks (***) and ≤ 0.0001 are indicated as four asterisks (****).

Crystallisation and structure determination of SaroL-1

The protein dissolved in Buffer A to 9 mg/mL was co-crystallised with 10 mM GalNAc and 10 mM Gb3 trisaccharide after incubation for at least 2 hours at RT. Crystallization screening was performed using the vapor diffusion method with hanging drops of 2 µL drops containing a 1/1 (*v/v*) mix of protein and reservoir solution at 19°C. Crystal clusters were obtained after several days from a solution containing 0.1 M buffer system 4, pH 6.5 (MOPSO/Bis-Tris, Molecular Dimensions), 100 mM AA Morpheus I, 25 or 35% PEG SMEAR MEDIUM for native SaroL-1 GalNAc and GB3 complex crystals, respectively and 0.1 M buffer system 4 pH 6.5, 100 mM AA (Arg, Thr, Lys, His), 45% precipitant mix 6 Morpheus II (Molecular Dimensions) for SeMet protein. Single crystals were directly mounted in a cryoloop and flash-freezed in liquid nitrogen. Diffraction data were collected at 100 K at the Soleil Synchrotron (Paris, France) on Proxima 2 for SeMet and GalNAc structures and Proxima 1 for the GB3 complex using a DECTRIS EIGER X 16M detector. The data were processed using XDS [76] and XDSME (GitHub). All further computing was performed using the CCP4i suite [77]. Data quality statistics are summarized in Supplemental Table 3. The SeMet SaroL-1 served as the initial structure where the phases were solved experimentally, and the model was built using Crank2 on MAD data, including peak and inflection point [78]. The structures of SaroL-1/GalNAc and SaroL-1/Gb3 were solved by molecular replacement using PHASER and the coordinates of SeMet protein as search model [79]. The structures were refined with restrained maximum likelihood refinement using REFMAC 5.8, and local NCS restrains [80] iterated with manual rebuilding in Coot [81]. Five percent of the observations were set aside for cross-validation analysis, and hydrogen atoms were added in their riding positions and used for geometry and structure-factor calculations. Incorporation of the ligand was performed after inspection of the ARP/WARP 2Fo-DFc weighted maps. The library for hexanetriol was constructed with Sketcher and Libcheck in CCP4i. Water molecules were inspected manually. The model was validated with the wwPDB Validation server: http://wwpdb-validation.wwpdb.org. The coordinates were deposited in the Protein Data Bank under code 7QE3 for SeMet SaroL-1, 7QE4 for the native SaroL-1 in complex with GalNAc and 7R55 for the native SaroL-1 in complex with Gb3.

Acknowledgements. This research was funded by the European Union Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement synBIOcarb (No.

814029). AI, FL and FB acknowledge support from Glyco@Alps (ANR-15-IDEX-02), Labex Arcane/CBH- EUR-GS (ANR-17-EURE-0003) and the Alliance Campus Rhodanien. WR acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC-294 and EXC-2189) and for Major Research Instrumentation (project number: 438033605), by the Ministry for Science, Research and Arts of the State of Baden-Württemberg (Az: 33-7532.20) and by the Freiburg Institute for Advanced Studies (FRIAS).

This publication is partially based upon work from COST Action CA18103 (INNOGLY), supported by COST (European Cooperation in Science and Technology). We acknowledge SOLEIL for provision of synchrotron radiation facilities (BAG proposal 20191314) and we would like to thank Martin Savko and William Shepard for assistance in using beamline Proxima 2 and Andrew Thompson for assistance using Proxima 1. The authors would like to thanks the chromatography platform from CERMAV and in particular Eric Bayma for performing SEC-MALS experiments.

Author contributions. F.B. developed the database and the prediction method and built the search and visualisation interface under the guidance of F.L. and A.I., S.N. designed the sequence, produced, analysed, and characterised the lectin. L.S. provided experiments with GUVs, F.R. and J.S. performed cell-based assays, N.B. provided FSL-Gb3 and FSL-iGb3, S.N. and A.V. solved the crystal structure and refined it, A.I. created the model of pore-formation, F.B., S.N., F.R., L.S., A.I., and F.L. wrote the manuscript and prepared figures, A.V., W.R., F.L., and A.I. further revised the manuscript.

Declaration of interest. The authors declare no competing interests.

References

1 Gabius, H. J. *et al.* What is the Sugar Code? *Chembiochem*, doi:10.1002/cbic.202100327 (2022).
2 Lis, H. & Sharon, N. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637-674, doi:10.1021/cr940413g (2002).
3 Rutenber, E. *et al.* Crystallographic refinement of ricin to 2.5 A. *Proteins* **10**, 240-250 (1991).
4 Merritt, E. A. *et al.* Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. *Protein Sci.* **3**, 166-175 (1994).
5 Iacovache, I., van der Goot, F. G. & Pernot, L. Pore formation: an ancient yet complex form of attack. *Biochim. Biophys. Acta* **1778**, 1611-1623, doi:10.1016/j.bbamem.2008.01.026 (2008).
6 Li, Y. *et al.* Structural Basis of the Pore-Forming Toxin/Membrane Interaction. *Toxins* **13**, doi:10.3390/toxins13020128 (2021).
7 Podobnik, M., Kisovec, M. & Anderluh, G. Molecular mechanism of pore formation by aerolysin-like proteins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**, doi:10.1098/rstb.2016.0209 (2017).
8 Szczesny, P. *et al.* Extending the aerolysin family: from bacteria to vertebrates. *PLoS One* **6**, e20349, doi:10.1371/journal.pone.0020349 (2011).
9 Jia, N. *et al.* Structural basis for receptor recognition and pore formation of a zebrafish aerolysin-like protein. *EMBO Rep.* **17**, 235-248, doi:10.15252/embr.201540851 (2016).

10 Pang, Y. *et al.* Crystal structure of a cytocidal protein from lamprey and its mechanism of action in the selective killing of cancer cells. *Cell Commun. Signal.* **17**, 54, doi:10.1186/s12964-019-0358-y (2019).
11 Unno, H., Goda, S. & Hatakeyama, T. Hemolytic lectin CEL-III heptamerizes via a large structural transition from alpha-helices to a beta-barrel during the transmembrane pore formation process. *J. Biol. Chem.* **289**, 12805-12812, doi:10.1074/jbc.M113.541896 (2014).
12 Mancheno, J. M., Tateno, H., Goldstein, I. J., Martinez-Ripoll, M. & Hermoso, J. A. Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *J. Biol. Chem.* **280**, 17251-17259 (2005).
13 Bonnardel, F. *et al.* UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Res.* **47**, D1236–D1244, doi:10.1093/nar/gky832 (2019).
14 Bonnardel, F. *et al.* Architecture and evolution of blade assembly in β-propeller lectins. *Structure* **27**, 764–775, doi:https://doi.org/10.1016/j.str.2019.02.002 (2019).
15 Murzin, A. G., Lesk, A. M. & Chothia, C. beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors. *J. Mol. Biol.* **223**, 531-543, doi:10.1016/0022-2836(92)90668-a (1992).
16 Notova, S., Bonnardel, F., Lisacek, F., Varrot, A. & Imberty, A. Structure and engineering of tandem repeat lectins. *Curr. Opin. Struct. Biol.* **62**, 39-47, doi:10.1016/j.sbi.2019.11.006 (2020).
17 Chernikov, O. V. *et al.* A GalNAc/Gal-specific lectin from the sea mussel *Crenomytilus grayanus* modulates immune response in macrophages and in mice. *Sci. Rep.* **7**, 6315, doi:10.1038/s41598-017-06647-5 (2017).
18 Fujii, Y. *et al.* A GM1b/asialo-GM1 oligosaccharide-binding R-type lectin from purplish bifurcate mussels Mytilisepta virgata and its effect on MAP kinases. *FEBS J.* **287**, 2612-2630, doi:10.1111/febs.15154 (2020).
19 Hasan, I. *et al.* MytiLec, a Mussel R-Type Lectin, Interacts with Surface Glycan Gb3 on Burkitt's Lymphoma Cells to Trigger Apoptosis through Multiple Pathways. *Mar. Drugs* **13**, 7377-7389, doi:10.3390/md13127071 (2015).
20 Liao, J. H. *et al.* A multivalent marine lectin from *Crenomytilus grayanus* possesses anti-cancer activity through recognizing globotriose Gb3. *J. Am. Chem. Soc.* **138**, 4787-4795, doi:10.1021/jacs.6b00111 (2016).
21 Terada, D. *et al.* Crystal structure of MytiLec, a galactose-binding lectin from the mussel *Mytilus galloprovincialis* with cytotoxicity against certain cancer cell types. *Sci. Rep.* **6**, 28344, doi:10.1038/srep28344 (2016).
22 Terada, D. *et al.* Computational design of a symmetrical β-trefoil lectin with cancer cell binding activity. *Sci. Rep.* **7**, 5943, doi:10.1038/s41598-017-06332-7 (2017).
23 Hazes, B. The (QxW)3 domain: a flexible lectin scaffold. *Protein Sci.* **5**, 1490-1501, doi:10.1002/pro.5560050805 (1996).
24 Dang, L., Rouge, P. & Van Damme, E. J. M. Amaranthin-Like Proteins with Aerolysin Domains in Plants. *Front Plant Sci* **8**, 1368, doi:10.3389/fpls.2017.01368 (2017).
25 Taylor, M. E., Conary, J. T., Lennartz, M. R., Stahl, P. D. & Drickamer, K. Primary structure of the mannose receptor contains multiple motifs resembling carbohydrate-recognition domains. *J. Biol. Chem.* **265**, 12156-12162 (1990).
26 Chernikov, O., Kuzmich, A., Chikalovets, I., Molchanova, V. & Hua, K. F. Lectin CGL from the sea mussel Crenomytilus grayanus induces Burkitt's lymphoma cells death via interaction with surface glycan. *Int. J. Biol. Macromol.* **104**, 508-514, doi:10.1016/j.ijbiomac.2017.06.074 (2017).
27 Sillitoe, I. *et al.* CATH: increased structural coverage of functional space. *Nucleic Acids Res.* **49**, D266-D273, doi:10.1093/nar/gkaa1079 (2021).

28 Dayel, M. J. *et al.* Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. *Dev. Biol.* **357**, 73-82, doi:10.1016/j.ydbio.2011.06.003 (2011).
29 Omidvar, R. & Romer, W. Glycan-decorated protocells: novel features for rebuilding cellular processes. *Interface Focus* **9**, 20180084, doi:10.1098/rsfs.2018.0084 (2019).
30 Arnaud, J. *et al.* Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity. *ACS Chem. Biol.* **8**, 1918-1924, doi:10.1021/cb400254b (2013).
31 Arnaud, J. *et al.* Membrane deformation by neolectins with engineered glycolipid binding sites. *Angew. Chem. Int. Ed.* **53**, 9267-9270, doi:10.1002/anie.201404568 (2014).
32 Romer, W. *et al.* Shiga toxin induces tubular membrane invaginations for its uptake into cells. *Nature* **450**, 670-675, doi:10.1038/nature05996 (2007).
33 Thuenauer, R. *et al.* The *Pseudomonas aeruginosa* lectin LecB causes integrin internalization and inhibits epithelial wound healing. *mBio* **11**, e03260-03219 doi:10.1128/mBio.03260-19 (2020).
34 Zhuo, D., Li, X. & Guan, F. Biological Roles of Aberrantly Expressed Glycosphingolipids and Related Enzymes in Human Cancer Development and Progression. *Front. Physiol.* **9**, 466, doi:10.3389/fphys.2018.00466 (2018).
35 Cole, A. R. *et al.* Clostridium perfringens epsilon-toxin shows structural similarity to the pore-forming toxin aerolysin. *Nat. Struct. Mol. Biol.* **11**, 797-798, doi:10.1038/nsmb804 (2004).
36 Savva, C. G. *et al.* The pore structure of Clostridium perfringens epsilon toxin. *Nat. Commun.* **10**, 2641, doi:10.1038/s41467-019-10645-8 (2019).
37 Cirauqui, N., Abriata, L. A., van der Goot, F. G. & Dal Peraro, M. Structural, physicochemical and dynamic features conserved within the aerolysin pore-forming toxin family. *Sci. Rep.* **7**, 13932, doi:10.1038/s41598-017-13714-4 (2017).
38 Bohrer, M. P., Deen, W. M., Robertson, C. R., Troy, J. L. & Brenner, B. M. Influence of molecular configuration on the passage of macromolecules across the glomerular capillary wall. *J. Gen. Physiol.* **74**, 583-593, doi:10.1085/jgp.74.5.583 (1979).
39 Knapp, O., Stiles, B. & Popoff, M. R. The Aerolysin-Like Toxin Family of Cytolytic, Pore-Forming Toxins. *Open Toxinol. J.* **3**, 53-68 (2010).
40 Mrazkova, J., Malinovska, L. & Wimmerova, M. Microscopy examination of red blood and yeast cell agglutination induced by bacterial lectins. *PLoS One* **14**, e0220318, doi:10.1371/journal.pone.0220318 (2019).
41 Grossmann, J. Molecular mechanisms of "detachment-induced apoptosis--Anoikis". *Apoptosis* **7**, 247-260, doi:10.1023/a:1015312119693 (2002).
42 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).
43 Lis, H. & Sharon, N. Lectins as molecules and as tools. *Ann. Rev. Biochem.* **1986**, 35-67 (1986).
44 Manning, J. C. *et al.* Lectins: a primer for histochemists and cell biologists. *Histochem. Cell Biol.* **147**, 199-222, doi:10.1007/s00418-016-1524-6 (2017).
45 Connaris, H., Crocker, P. R. & Taylor, G. L. Enhancing the receptor affinity of the sialic acid-binding domain of Vibrio cholerae sialidase through multivalency. *J. Biol. Chem.* **284**, 7339-7351, doi:10.1074/jbc.M807398200 (2009).
46 Ribeiro, J. P. *et al.* Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells. *Chem. Sci.* **9**, 7634 - 7641 doi:10.1039/C8SC02730G (2018).
47 Siukstaite, L. *et al.* The two sweet sides of Janus lectin drive crosslinking of liposomes to cancer cells and material uptake *Toxins* **13**, 792, doi:10.3390/toxins13110792 (2021).

48 Yabe, R. *et al.* Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-binding protein by natural evolution-mimicry. *J. Biochem.* **141**, 389-399, doi:10.1093/jb/mvm043 (2007).
49 Diep, D. B., Nelson, K. L., Raja, S. M., Pleshak, E. N. & Buckley, J. T. Glycosylphosphatidylinositol anchors of membrane glycoproteins are binding determinants for the channel-forming toxin aerolysin. *J. Biol. Chem.* **273**, 2355-2360, doi:10.1074/jbc.273.4.2355 (1998).
50 Stiles, B. G., Barth, G., Barth, H. & Popoff, M. R. Clostridium perfringens epsilon toxin: a malevolent molecule for animals and man? *Toxins* **5**, 2138-2160, doi:10.3390/toxins5112138 (2013).
51 Lebreton, A. *et al.* Comprehensive phylogenetic and bioinformatics survey of lectins in the fungal kingdom. *J. Fungi* **7**, 453, doi:10.3390/jof7060453 (2021).
52 Taga, S., Mangeney, M., Tursz, T. & Wiels, J. Differential regulation of glycosphingolipid biosynthesis in phenotypically distinct Burkitt's lymphoma cell lines. *Int. J. Cancer* **61**, 261-267, doi:10.1002/ijc.2910610220 (1995).
53 Jacob, F. *et al.* Serum antiglycan antibody detection of nonmucinous ovarian cancers by using a printed glycan array. *Int. J. Cancer* **130**, 138-146, doi:10.1002/ijc.26002 (2012).
54 Kovbasnjuk, O. *et al.* The glycosphingolipid globotriaosylceramide in the metastatic transformation of colon cancer. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 19087-19092, doi:10.1073/pnas.0506474102 (2005).
55 LaCasse, E. C. *et al.* Shiga-like toxin-1 receptor on human breast cancer, lymphoma, and myeloma and absence from CD34(+) hematopoietic stem cells: implications for ex vivo tumor purging and autologous stem cell transplantation. *Blood* **94**, 2901-2910 (1999).
56 Maak, M. *et al.* Tumor-specific targeting of pancreatic cancer with Shiga toxin B-subunit. *Mol Cancer Ther* **10**, 1918-1928, doi:10.1158/1535-7163.MCT-11-0006 (2011).
57 Siukstaite, L., Imberty, A. & Römer, W. Structural diversities of lectins binding to the glycosphingolipid Gb3 *Front. Mol. Biosci.* **8**, 704685, doi:10.3389/fmolb.2021.704685 (2021).
58 Fujii, Y. *et al.* A lectin from the mussel Mytilus galloprovincialis has a highly novel primary structure and induces glycan-mediated cytotoxicity of globotriaosylceramide-expressing lymphoma cells. *J. Biol. Chem.* **287**, 44772-44783, doi:10.1074/jbc.M112.418012 (2012).
59 Cao, C. *et al.* Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores. *Nat. Commun.* **10**, 4918, doi:10.1038/s41467-019-12690-9 (2019).
60 Cressiot, B. *et al.* Aerolysin, a Powerful Protein Sensor for Fundamental Studies and Development of Upcoming Applications. *ACS Sens.* **4**, 530-548, doi:10.1021/acssensors.8b01636 (2019).
61 Fennouri, A., Ramiandrisoa, J., Bacri, L., Mathe, J. & Daniel, R. Comparative biosensing of glycosaminoglycan hyaluronic acid oligo- and polysaccharides using aerolysin and hemolysin nanopores. *Eur. Phys. J. E Soft Matter* **41**, 127, doi:10.1140/epje/i2018-11733-5 (2018).
62 Xi, D., Li, Z., Liu, L., Ai, S. & Zhang, S. Ultrasensitive Detection of Cancer Cells Combining Enzymatic Signal Amplification with an Aerolysin Nanopore. *Anal. Chem.* **90**, 1029-1034, doi:10.1021/acs.analchem.7b04584 (2018).
63 Ouldali, H. *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176-181, doi:10.1038/s41587-019-0345-2 (2020).
64 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
65 Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200-W204, doi:10.1093/nar/gky448 (2018).
66 Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755-763, doi:10.1093/bioinformatics/14.9.755 (1998).
67 UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).

68 Haft, D. H. *et al.* RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* **46**, D851-D860, doi:10.1093/nar/gkx1068 (2018).
69 Bonnardel, F. *et al.* Proteome-wide prediction of bacterial carbohydrate-binding proteins as a tool for understanding commensal and pathogen colonisation of the vaginal microbiome. *Biofilms and Microbiomes* **7**, 49, doi:10.1038/s41522-021-00220-9 (2021).
70 Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W. & Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.* **81**, 6807-6816, doi:10.1128/JVI.00338-07 (2007).
71 Doublie, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523-530 (1997).
72 Korchagina, E. *et al.* Toward creating cell membrane glyco-landscapes with glycan lipid constructs. *Carbohydr. Res.* **356**, 238-246, doi:10.1016/j.carres.2012.03.044 (2012).
73 Volynsky, P. *et al.* Why human anti-Galalpha1-4Galbeta1-4Glc natural antibodies do not recognize the trisaccharide on erythrocyte membrane? Molecular dynamics and immunochemical investigation. *Mol. Immunol.* **90**, 87-97, doi:10.1016/j.molimm.2017.06.247 (2017).
74 Madl, J., Villringer, S. & Römer, W. in *Chemical and Synthetic Approaches in Membrane Biology Springer Protocols Handbooks* (ed A.K. Shukla) 17-36 (Humana Press Inc, 2017).
75 Abe, A. *et al.* Improved inhibitors of glucosylceramide synthase. *J. Biochem. (Tokyo)* **111**, 191-196, doi:10.1093/oxfordjournals.jbchem.a123736 (1992).
76 Kabsch, W. Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).
77 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
78 Pannu, N. S. *et al.* Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 331-337, doi:10.1107/S0907444910052224 (2011).
79 McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D. Biol. Crystallogr.* **63**, 32-41, doi:10.1107/S0907444906045975 (2007).
80 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).
81 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

A pore-forming β-trefoil lectin with specificity for the tumor-related glycosphingolipid Gb3

Supplementary Information

Supplementary Table 1*: Selection of protein particular architectures composed of β-trefoil domain(s) and other functional domains, identified using Pfam motifs.*

Family	PFAM	Species	ProteinAC	PfamAC
Ricin-like trefoil	Melibiase 2	*Catenulispora acidiphila*	C7Q336	PF16499, PF17801
Ricin-like trefoil	Lipase GDSL 2	*Catellatospora citrea*	A0A419XYK	PF13472
Ricin-like trefoil	Arabinase	*Lentzea aerocolonigenes*	A0A0F0GI39	PF09206
Ricin-like trefoil	Glycosyltransferase GT2	*Trachymyrmex septentrionalis*	A0A151JT25	PF00535
Ricin-like trefoil	Glycosylhydrolase GH16	*Streptomyces sp* BK335	A0A4V2ULV4	PF00722
Ricin-like trefoil	CBM 6 (cellulose-binding domain), GSDH (Glucose / Sorbosone dehydrogenase), PKD 4 (Polycystic Kidney Disease)	*Saccharothrix sp* ALI22I	A0A1V2PXN8	PF03422, PF07995, PF18911
Coprinus trefoil	LSM RNA-binding proteins	*Thanatephorus cucumeris*	A0A0B7G0E7	PF01423
Earthworm trefoil	Lipase GDSL 2	*Cellulomonas persica*	A0A510UZ07	PF13472
Cys-rich man-receptor trefoil	Fibronectin type II, C-type lectin-like	*Homo sapiens*, *Rattus norvegicus*	Q9UBG0, Q4TU93	PF00040, PF05473
Fungi and Clostridium trefoil	Aerolysin	*Laetiporus sulphureus*	Q7Z8V1	PF01117

Supplementary Table 2: *ITC results of SaroL-1 binding to different ligands. The stoichiometry N was fixed to value 3 in all cases.*

Ligand	**Kd (mM)**	**Δ H (kcal/mol)**	**Δ G (kcal/mol)**	**-T Δ S (kcal/mol)**
αGal1-4Gal	0.39 ± 0.02	-6.36 ± 0.20	-4.65	1.71
PNPG	1.01 ± 0.02	-7.36 ± 0.11	-4.09	3.28
αGal1-6Glc	1.20 ± 0.06	-12.7 ± 0.39	-3.98	8.76
αGal1-3Gal	1.38 ± 0.28	-7.46 ± 0.11	-3.90	3.56
GalαOMe	2.17 ± 0.27	-5.67 ± 0.46	-3.63	2.03
GalNAc	2.76 ± 0.01	-11.8 ± 0.31	-3.49	8.35
βGal1-4Glc	6.66 ± 1.4	-4.98 ± 1.20	-2.97	2.01

Supplementary Table 3: *Data collection and refinement statistics for SaroL-1/GalNAc, SaroL-1/Gb3 and Se-M SaroL-1/GalNAc*

Data collection						
Protein name	SaroL-1/GalNAc		SaroL-1/Gb3		Se-M-SaroL-1	
Beamline	Soleil Synchrotron Proxima-2		Soleil Synchrotron Proxima-1		Soleil Synchrotron Proxima-2	
Wavelength	0.980107		0.978565		0.979415	
Space group	$P2_12_12_1$		$P2_12_12_1$		$P2_1$	
Unit cell dimensions	57.16 59.25 209.61 90.00 90.00 90.00		57.34 58.99 210.30 90.00 90.00 90.00		44.99 58.58 166.42 90.00 90.68 90.00	
Resolution (Å)	45.19 – 1.70		45.18 – 1.84		39.78 – 2.20	
$\boldsymbol{R_{merge}}$	0.078 (0.616)		0.093 (1.034)		0.095 (0.503)	
$\boldsymbol{R_{pim}}$	0.041 (0.339)		0.050 (0.572)		0.072 (0.383)	
Mean I* / σ*I	14.6 (3.1)		11.8 (1.7)		11.2 (3.0)	
Completeness (%)	100 (100)		99.8 (96.9)		99.5 (98.5)	
Redundancy	8.5 (8.3)		8.2 (7.8)		4.9 (5.0)	
CC1/2	0.999 (0.890)		0.999 (0.792)		0.996 (0.852)	
Nb reflections	678912 (34686)		515983 (28809)		215732 (18811)	
Nb unique reflections	79471 (4185)		62904 (3696)		44177 (3794)	
Refinement						
Resolution (Å)	44.28 – 1.70		45.14 – 1.84		39.77 - 2.20	
No. reflections	75465		59836		42022	
No. free reflections	3914		2982		2095	
$\boldsymbol{R_{work}}$ / $\boldsymbol{R_{free}}$	0.169 / 0.206		0.184 / 0.240		0.189 / 0.231	
R.m.s Bond lengths (Å)	0.008		0.018		0.014	
Rmsd Bond angles (°)	1.433		2.201		1.920	
Rmsd Chiral (Å^3)	0.007		0.109		0.011	
Clashscore	2		2		4	
No. atoms / Bfac (Å^2)	Chain A	Chain B	Chain A	Chain B	Chain A	Chain B
Protein	5095/20.7	5115/19.0	2543/28.7	2527/30.9	2481/38.0	2482/29.5
Sugar	120 /18.9	120 / 21.0	68 / 34.2	68 / 33.5	15 / 42.58	-
Water	486 / 31.3	542 / 30.0	331 / 35.9	249 / 36.3	191 / 40.1	280/34.3
Ramachandran	98.6		98.1		97.9	
Allowed / Favored / Outliers (%)	97.8		97.0		96.4	
	0.00		0.16		0.00	
PDB Code	7QE4		7R55		7QE3	

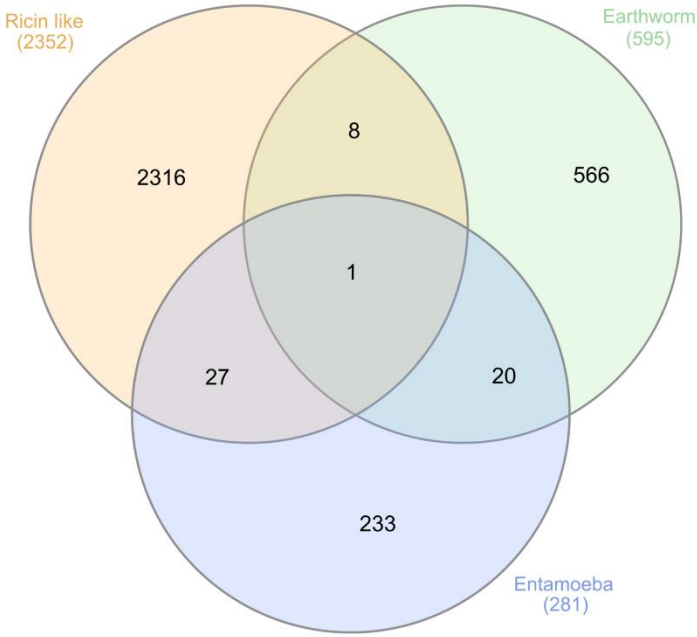
*Values in brackets are for highest-resolution shell.

** Riding hydrogen atoms were added to the coordinate file of 7QE4

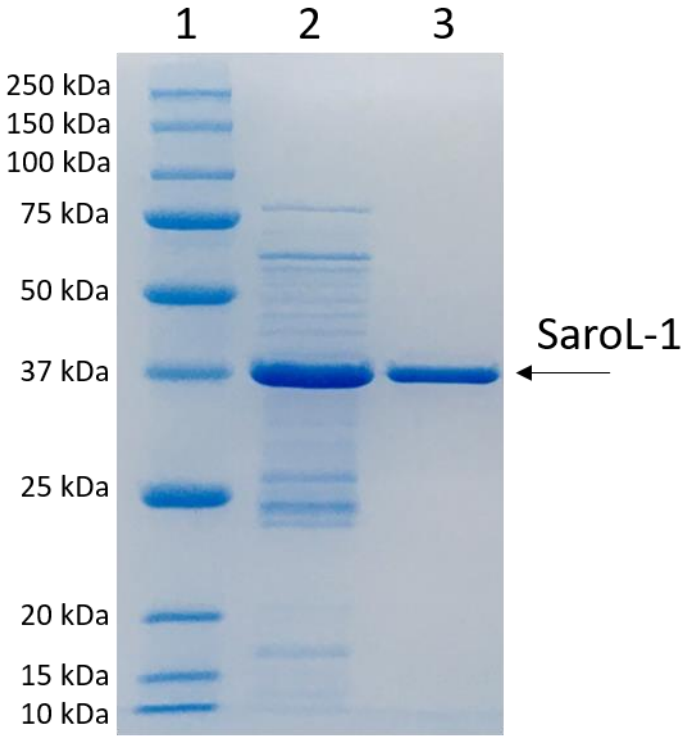
Supplementary Table 4: *Direct hydrogen bonds observed in native SaroL-1 structures between α-GalNAc and Gb3 trisaccharide with amino acids of the three binding sites. Distances (Å) are averaged between the two chains (standard deviation < 0.1 Å).*

		Site α	***Site β***	***Site γ***
αGalNAc	O3	R26.NH2 2.85	R103.NH2 2.85	R150.NH2 2.85
	O4	R26.NH1 2.9	R103.NH1 2.95	R150.NH1 2.90
	N2	H32.ND1 2.60	H83.ND1 2.75	H129.ND1 2.75
	O6	G35.N 2.90 N140.OD1 2.50	G86.N 2.85 D43.OD1 2.65	G132.N 2.75 E92.OE2 2.55
Gb3	O3 (αGal)		S100.OG 2.6 R103.NH1 2.85	R150.NH2 2.8
	O4 (αGal)		R103.NH2 3 H83.ND1 2.7	R150.NH2 2.9 H129.ND1 2.6
	O6 (αGal)		D43.OD1 2.75 G86.N 2.9	E92.OE2 2.5 G132.N 2.75
	O2 (βGal)		D43.OD2 2.5	E92.OE2 2.65
	O5 (βGal)		H98.NE2 3.35	Y146.OH 3.5
	O1 (αGlc)			K91.NZ 3.2
	O2 (αGlc)		N42.ND2 3.35	K91.NZ 2.8
	O3 (αGlc)		H98.NE2 3.1	Y146.OH 3
	O4 (αGlc)		H98.NE2 3.2	Y146.OH 3.45
	O6 (αGlc)		D43.OD2 3.3	

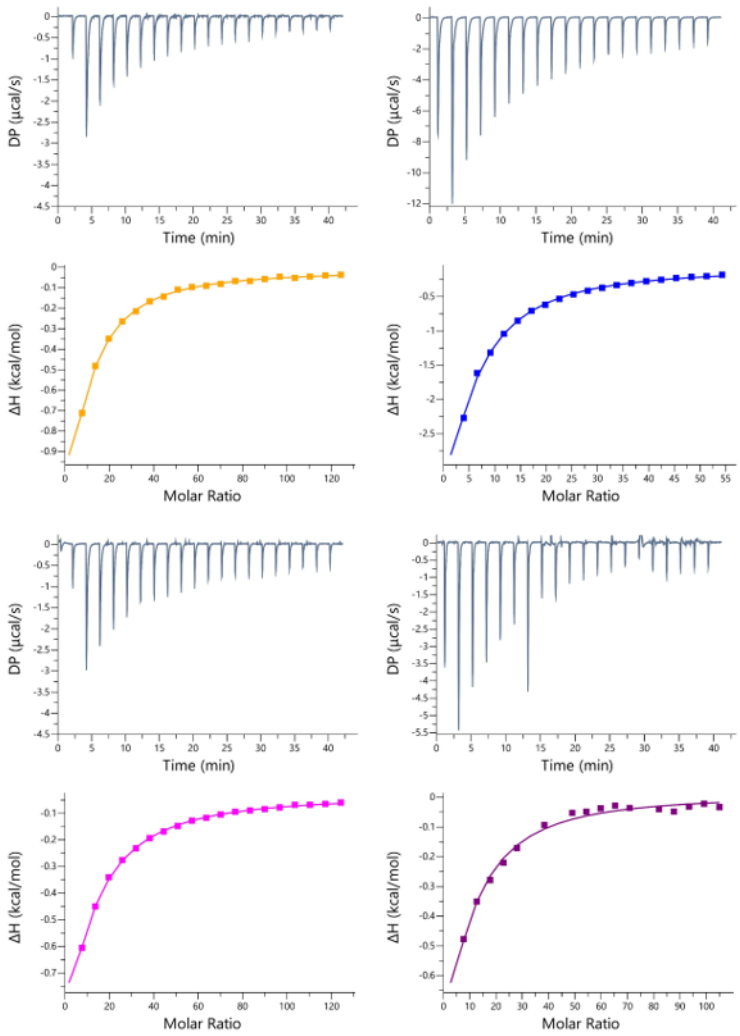
Supplementary Figure 1*: The classification of β-trefoil lectins and their binding motif signature generated by WebLOGO (Crooks et al. 2004, doi: 10.1101/gr.849004.)*



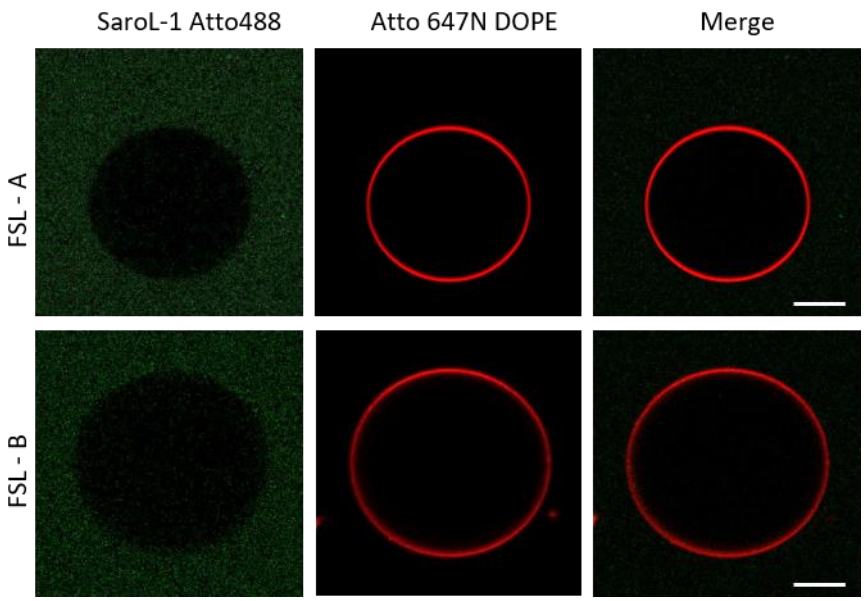
Supplementary Figure 2*: Venn diagram representing the overlap between the β-trefoil classes in predicted lectins, for a score > 0.25. Graphics by InteractiVenn (Heberle et al. 2015, doi: 10.1186/s12859-015-0611-3).*



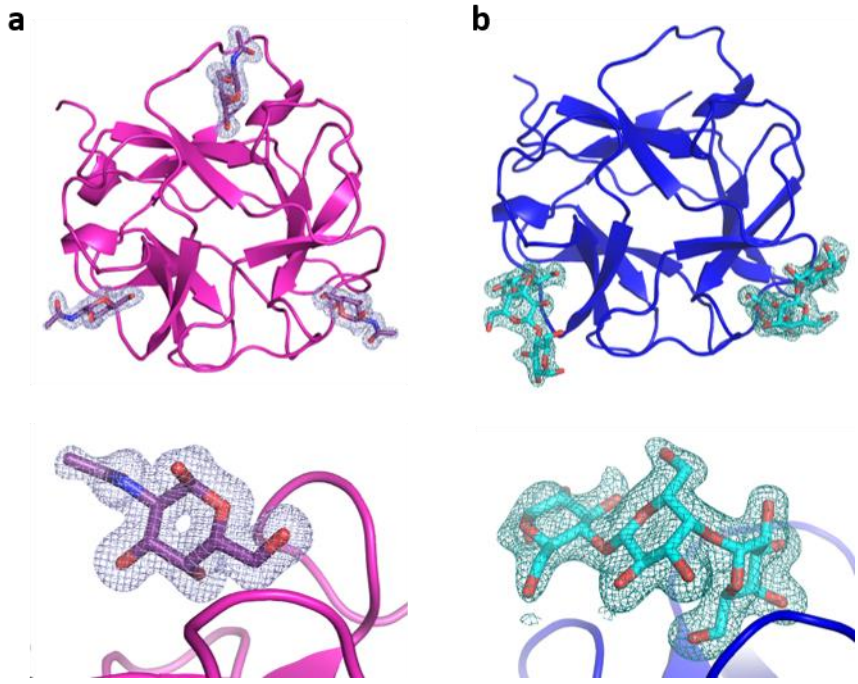
Supplementary Figure 3: *Analysis of SaroL-1 by denaturating SDS page gel electrophoresis. 12 % SDS gel, row 1 – protein marker, row 2 – elution of metal affinity chromatography, row 3 – elution of size exclusion chromatography. The molecular weight of SaroL-1 was estimated as 36.86 ± 0.76 kDa.*



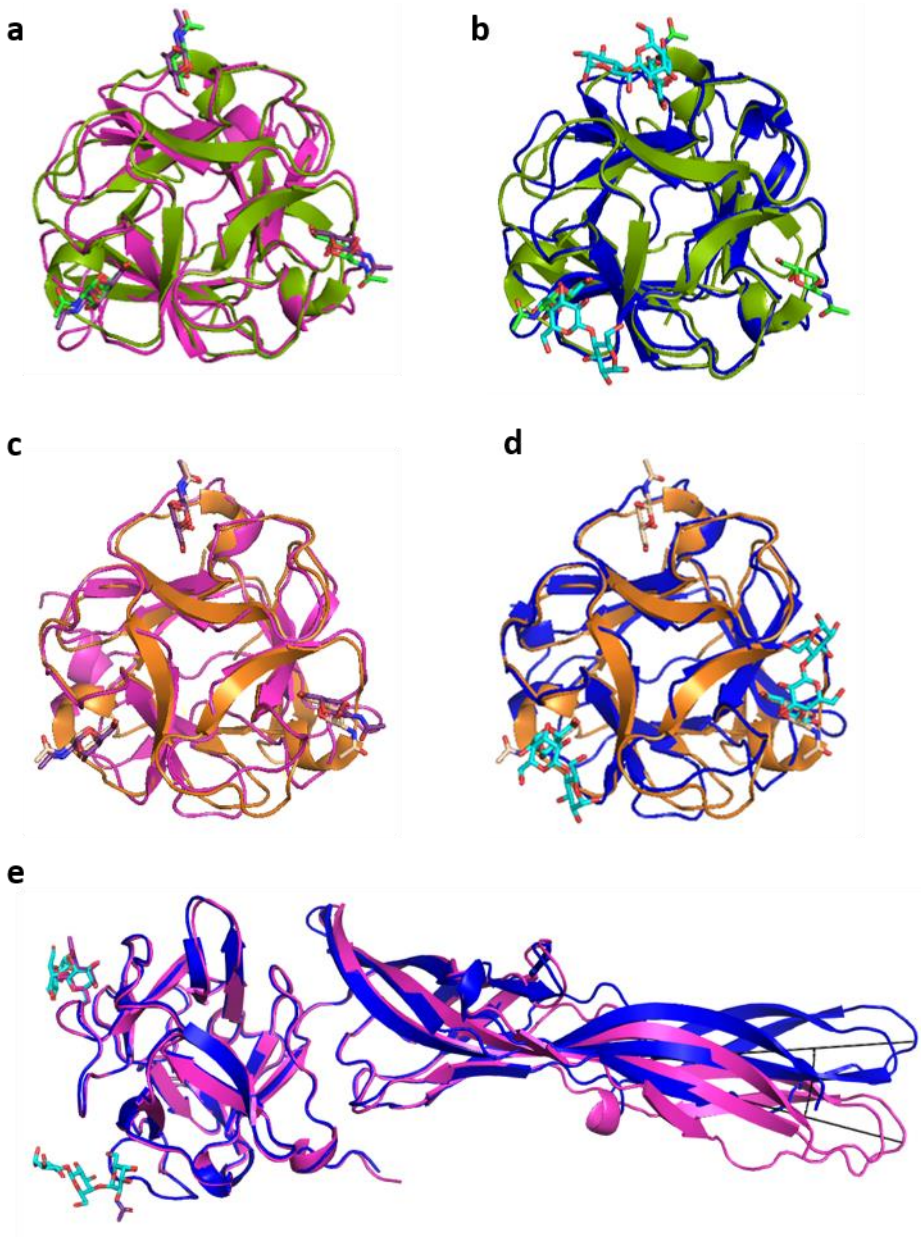
Supplementary Figure 4: *ITC data of SaroL-1 with different carbohydrate, the thermograms (top) and integrated peaks (bottom). The ITC cell contained SaroL-1 in the concentration range of 0.050-0.116 mM. The syringe contained ligands, such as PNPG (orange), αGal1-6Glc (melibiose) (blue), αGal1-3Gal (magenta) and GalαOMe (purple) in the concentration range of 10-50 mM.*



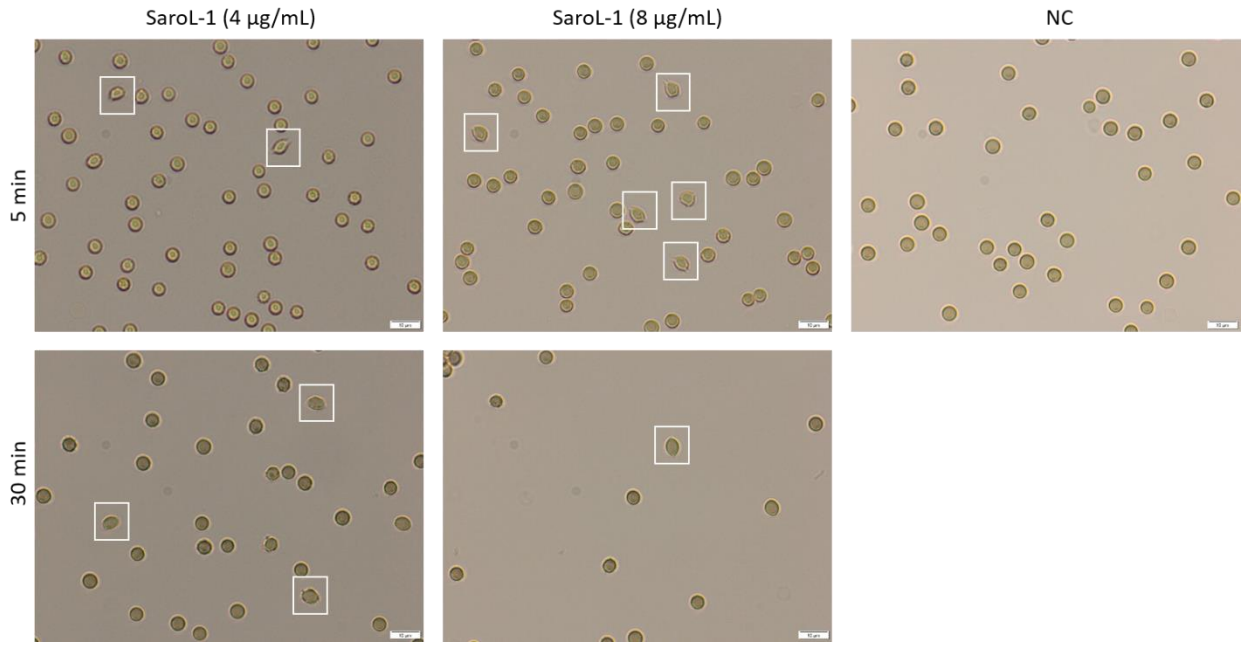
Supplementary Figure 5: *Binding assay of 200 nM of SaroL-1 (green) and GUVs (red). GUVs are functionalised with FSL-A and FSL-B (function-spacer-lipid with either blood group A or blood group B trisaccharide). Scale bars represent 10 µm.*



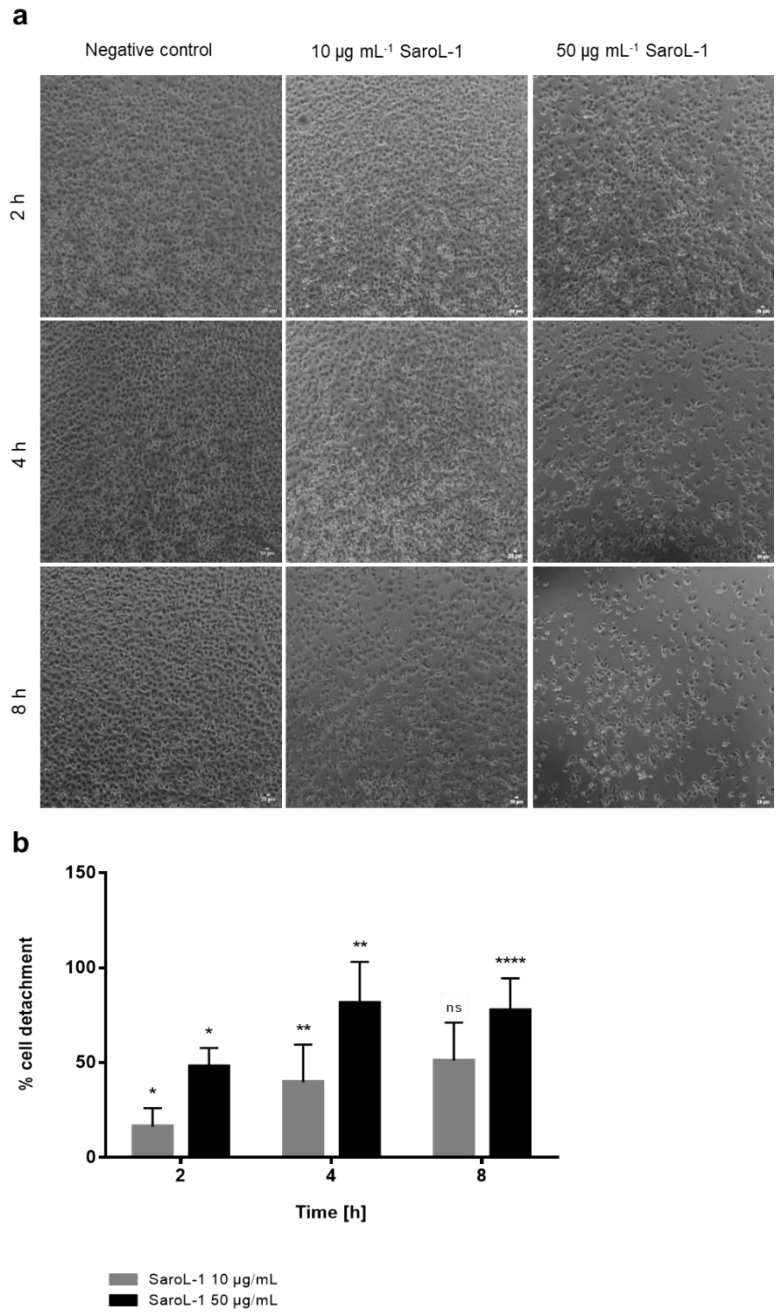
Supplementary Figure 6: *Electron density map of ligands. a) β-trefoil domain of SaroL-1 in complex with αGalNAc, zoom on β-site, b) β-trefoil domain of SaroL-1 in complex with Gb3, zoom on β-site.*



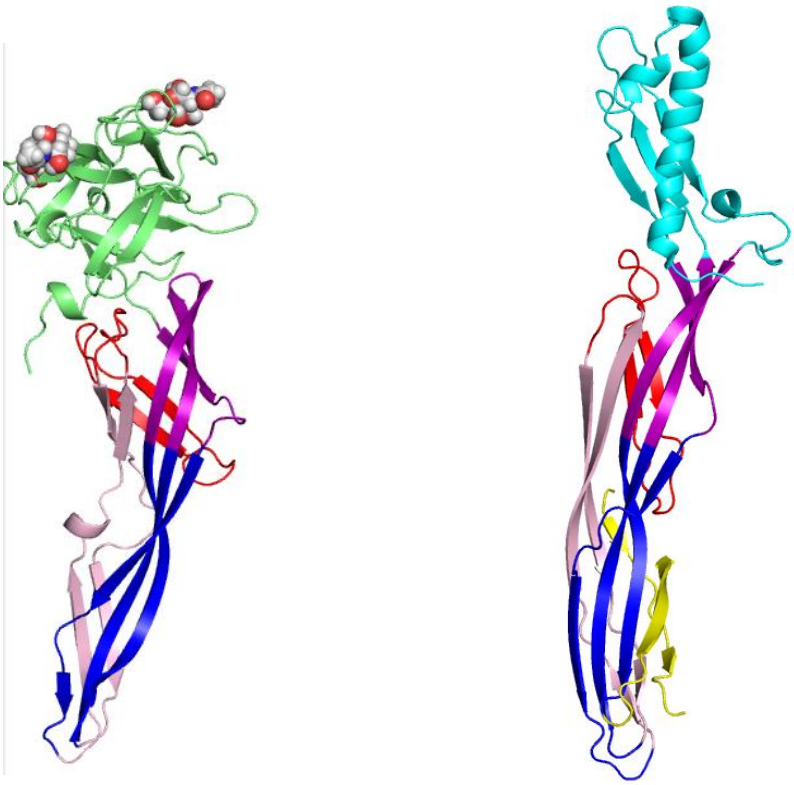
Supplementary Figure 7*: Structures superimposition. a) β-trefoil domain of SaroL-1 in light magenta in complex with GalNAc (violet) (7QE4) with Mytilec (olive green) in complex with GalNAc (green) (3WMV). b) β-trefoil domain of SaroL-1 in blue in complex with Gb3 (cyan) (7R55) with Mytilec (olive green) in complex with GalNAc (green) (3WMV). c) β-trefoil domain of SaroL-1 in light magenta in complex with GalNAc (violet) (7QE4) with Mitsuba (orange) in complex with GalNAc (pale orange) (5XG5). d) β-trefoil domain of SaroL-1 in blue in complex with Gb3 (cyan) (7R55) with Mitsuba (orange) in complex with GalNAc (pale orange) (5XG5). d) Shift of pore-forming domain between chain A of SaroL-1/GalNAc (light magenta) with SaroL-1/Gb3 (blue). The difference angle is measured as SaroL-1/Gb3/D194 - SaroL-1/GalNAc/R287 - SaroL-1/GalNAc/D194 = 17.6°.*



Supplementary Figure 8: *The almond-like shape of the rabbit erythrocytes is observed once incubated with SaroL-1 (4 µg/mL and 8 µg/mL). After 5 and 30 min of incubation, the red blood cells underwent morphological changes (white rectangles) due to SaroL-1 binding. A solution of untreated rabbit erythrocytes (NC) functioned as negative control showing the natural shape of red blood cells.*



Supplementary Figure 9: *SaroL-1 binding to H1299 cells induce cell detachment. a) Adherent H1299 cells were treated with SaroL-1 (10 µg/mL and 50 µg/mL) for 8 hours and observed by light microscopy. SaroL-1 caused a dose-dependent rounding and detachment of cells in comparison to treatment with PBS. At the highest concentration, SaroL-1's induced detachment of cells is visible at early time point (2 h), whereas 10 µg/mL SaroL-1 promoted cell detachment at 8 h. b) Quantification of cell detachment induced by SaroL-1 treatment for different time points (2, 4, 8 hours). The increase in cell detachment was quantified by analyzing the supernatant with CytoSmart Corning cell counter (means ± SD, n = 3). Significant difference compared to the negative control for each time point, *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001.*



Supplementary Figure 10: *Structural similarities between SaroL-1 (left) and ε-toxin of C. perfringens (PDB 1UYJ, right), used for aligning the sequences for further model building of the pore-forming domain.*

5. BUILDING ARTIFICIAL PLANT CELL WALL ON LIPID BILAYER BY ASSEMBLING POLYSACCHARIDES AND ENGINEERED PROTEINS

5.1 Summary

The second project I have been involved in focused on the construction and characterization of Janus lectin RSL-CBM77$_{Rf}$ and the divalent form of CBM77$_{Rf}$ with the purpose to build an artificial plant cell wall. Throughout this project, we established collaboration within CERMAV CNRS with the team of Dr. Laurent Heux and his Ph.D. students Maeva Touzard and Josselin Mante, and Dr. Liliane Coche-Guérente (DCM, Grenoble).

This project has for me a special significance because during it I had the opportunity to co-supervise two master's students, Nathan Cannac (M2) and Luca Rabagliati (M1). With Luca, we worked together on the gene design of divalent CBM77$_{Rf}$, cloning, protein expression, purification, and biophysical characterization by ITC. With Nathan, we designed the gene sequences for Janus lectins and did the cloning, protein expression, purification, and ITC characterization. I was not involved in the QCM-D experiments, nor polymers purification or characterization.

Our findings are summarized in the publication entitled *Building artificial plant cell wall on lipid bilayer by assembling polysaccharides and engineered proteins* which was submitted to preprint server BioRxiv, DOI: 10.1101/2022.07.25.501355 and subsequently submitted to the journal ACS Synthetic Biology where the final revisions have been done.

5.2 Scientific article II

Building artificial plant cell wall on lipid bilayer by assembling polysaccharides and engineered proteins

Simona Notova [a,#], Nathan Cannac [a,#], Luca Rabagliati [a], Maeva Touzard [a], Josselin Mante [a], Yotam Navon [b], Liliane Coche-Guérente [c], Olivier Lerouxel [a], Laurent Heux [a,*] & Anne Imberty[a,*]

[a] Université Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France
[b] The Pulp and Paper Research & Technical Centre, 38044 Grenoble, France
[c] Université Grenoble Alpes, CNRS, DCM, 38000 Grenoble, France

co-premier's authors

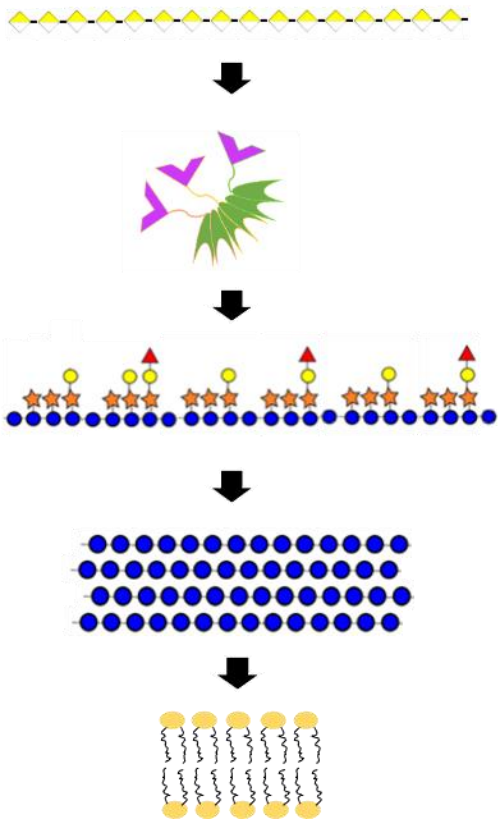
* To whom correspondence should be addressed. Anne Imberty (anne.imberty@cermav.cnrs.fr) and Laurent Heux (laurent.heux@cermav.cnrs.fr)

Keywords: synthetic biology, protein engineering, lectins, carbohydrate-binding protein, plant cell wall

Abstract

The cell wall constitutes a fundamental structural component of plant cells, providing them with mechanical resistance and flexibility. Mimicking that wall is a critical step in the conception of an experimental model of the plant cell. The assembly of cellulose/hemicellulose in the form of cellulose nanocrystals and xyloglucans as a representative model of the plant cell wall has already been mastered, however, those models lacked the pectin component. In this work, we used engineered protein for bridging pectin to hemicellulose, therefore assembling a complete cell wall mimics. We first engineered proteins, i.e. carbohydrate-binding module from *Ruminococcus flavefaciens* able to bind oligo-galactorunan, resulting in high affinity polygalacturonan receptors with Kd in the micromolar range. A Janus protein, with cell wall gluing property, was then designed by assembling this CBM with a *Ralstonia solanacearum* lectin specific for fucosylated xyloglucans. The resulting supramolecular architecture is able to bind fucose-containing xyloglucans and homogalacturonan ensuring high affinity for both. A two-dimension assembly of an artificial plant cell wall was then built first on synthetic polymer and then supported lipid bilayer. Such artificial cell wall can serve as a basis for the development of plant cell mechanical models and thus deepen the understanding of the principles underlying various aspects of plant cells and tissues.

Graphical abstract



Introduction

The composite material nature of the primary plant cell walls brings them their essential functions in controlling cell shapes, providing structural rigidity but also flexibility and establishing communications, exchange and also protecting barriers against pathogens (McNeil, Darvill et al. 1984, Zhang, Gao et al. 2021). The primary cell walls consist mainly of cellulose microfibrils embedded in two other polysaccharides networks, hemicellulose, and pectin, with a low amount of proteins. Cellulose occurs as semi-crystalline microfibrils and forms a mesh-like structure around the cells. Hemicellulose, such as xyloglucan, adsorbs onto cellulose, crosslinking the microfibrils and forming an entangled network that bring further strength and extensibility (Pauly, Albersheim et al. 1999). Pectins are the major charged polysaccharides mainly composed of homogalacturonans, they form a network in close contact but independent of cellulose and through phase separation processes, are a driving force for cell wall expansion and morphogenesis (Haas, Wightman et al. 2020). Pectin's homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells (Haas, Wightman et al. 2021). Both networks gave a complex mechanical behavior to the plant cell wall essential to provide elasticity, plasticity and deformability upon mechanical stress (Zhang, Yu et al. 2021). Although present in lower amounts, cell wall proteins are prone to be involved in cell wall dynamics, expansins, and other structural proteins that cross-link the component of the cell wall and should participate in its mechanical properties (Jamet, Canut et al. 2006).

The remarkable composition of plant cell walls inspired the design of biomimetic materials, that are environmentally friendly and have unique properties in strengths, lightness, and resistance (Teeri, Brumer et al. 2007). Applications range from a medical domain with polysaccharide-based vesicles for encapsulating active compounds to the coating, or even packaging. Recently, the engineering of multilayer microcapsules based on polysaccharides nanomaterials resulted in vesicles with various applications (Lombardo and Villares 2020). For the formation of such vesicles, cellulose nanomaterials are often the main component in a bio-mimetic approach, with the addition of xyloglucan or pectin. The most recent study succeeded in assembling cellulose microfibrils and pectins on lipid-based vesicles and therefore reached one step closer to the construction of artificial plant cells (Paulraj, Wennmalm et al. 2020). Even though the individual components of the plant cell wall found their applications, the construction of synthetic plant cell walls remains a challenge.

Biomimetic cell walls can also be obtained in the form of film, i.e., supported multilayer assembly that can be analyzed by a variety of techniques. Such films were first obtained based on cellulose nanocrystals (CNCs) that can be easily obtained from plant material with strong acid treatment, hydrolyzing the amorphous regions of cellulose and releasing intact crystalline blocks (Rånby 1949, Habibi, Lucia et al. 2010). CNCs have high colloidal stability and represent an excellent model for mimicking the naturally occurring cellulose microfibrils. Multilayers of CNCs and xyloglucan (hemicellulose) can be obtained relatively easily due to the strong interactions, mainly driven by hydrophobic interaction occurring between the nonpolar faces of the glucose chain present in both polysaccharides (Hayashi, Takeda et al. 1994, Hanus and Mazeau 2006, Lopez, Bizot et al. 2010, Mazeau and Charlier 2012). In order

to mimic plant cell wall, a lipidic component is necessary. Supported lipid bilayer (SLB) is a widely used and characterized model of cell membranes (McConnell, Watts et al. 1986, Richter and Brisson 2005). Interaction between CNCs and SLB is the crucial step for an artificial plant cell wall and has also been described after tuning of parameters and the resulting multilayer has been fully characterized (Navon, Jean et al. 2020). Additionally, the addition of xyloglucan to this lipid-supported nanocellulose film has also been reported (Navon 2020).

The next step is therefore to include a layer of pectin on the multilayer. Pectins are a family of complex, charged and functionalized polysaccharides, the simplest model available is homogalacturonan (HG) i.e., a linear polymer of α1-4 linked galacturonic acid (GalA). Unmethylated HG can be obtained from different plant sources after enzymatic treatment. The adsorption of such soluble and charged polysaccharides on CNCs-xyloglucan film is a difficult process, partly due to the competition between xyloglucan and pectin for the binding to cellulose (Zykwinska, Thibault et al. 2008). In the cell wall, a tiny amount of proteins is present, and extensins are known to link the pectin (Sede, Borassi et al. 2018). We, therefore, propose to engineer protein glue that would stabilize the interaction between hemicellulose and pectin for the assembly of an artificial cell wall.

Proteins, able to bind to oligo or polysaccharides without modifying them, belong to two main families. Lectins are generally multivalent proteins with several carbohydrate-binding sites and exist with a variety of sequence and fold (Bonnardel, Mariethoz et al. 2019). On the other hand, carbohydrate-binding modules (CBMs) are monovalent protein domain (Lombard, Golaconda Ramulu et al. 2014), generally associated with a carbohydrate-active enzyme. Many of them play a role in biomass degradation, and are specific for plant polysaccharides (Gilbert, Knox et al. 2013) . While CBMs have low affinity for glycans, it has been demonstrating that associating two of them by flexible linkers results in 2 to 3 order of magnitude enhancement of the affinity (Connaris, Crocker et al. 2009, Ribeiro, Pau et al. 2016). Such synthetic biology approach was also used for building a bi-specific Janus lectin, assembling six fucose binding domains with three sialic acid binding domain (Ribeiro, Pau et al. 2016). This construct was able to crosslink vesicles decorated with the appropriate glycolipids, but also to assemble a multilayer of glycosylated multivalent compounds, forming a multilayer assembly.

We propose here to design a chimeric lectin with one face able to bind hemicellulose and one face specific for homogalacturonan. Such construct should be able to glue the xyloglucan/cellulose cement deposited on supported lipid, with the pectin matrix added above, therefore providing a more complete model of the whole plant cell wall.

Results

Selection of biomolecules for building of artificial cell wall polysaccharides

Construction of artificial cell walls involved the polysaccharides displayed in Figure 1, i.e. cellulose nanocrystals (CNC), fucosylated xyloglucan (FXG) as representative hemicellulose polysaccharide and linear homogalacturonan (HG) consisting of galacturonic acid (GalA) linked by α1-4 linkages which represents the simplest model for pectin. The preparation, purification and characterization of these three polysaccharides components are described in the material and methods section.

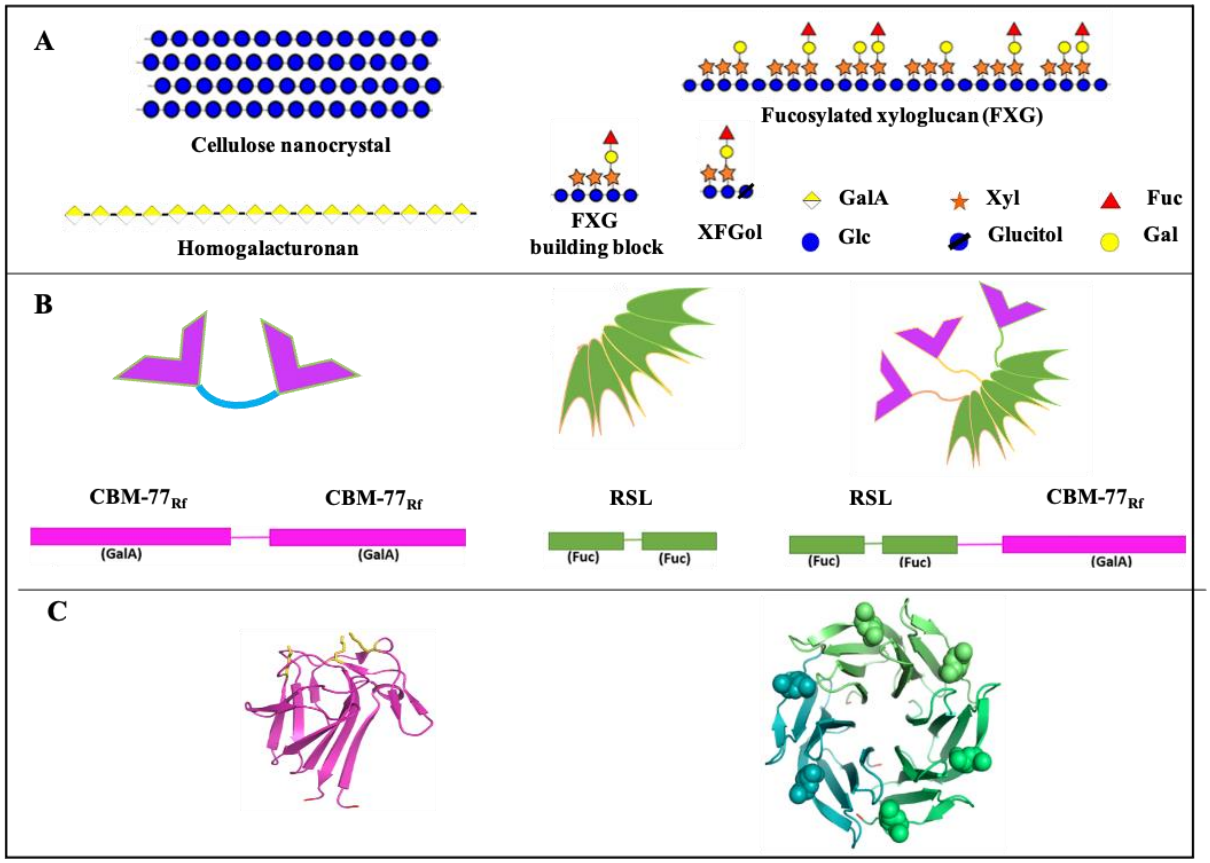


Figure 1: **Biomolecules used in the building of artificial cells**. A: Polysaccharides such as crystalline nanocellulose (CNC), homogalacturonan (HG) from pectin, and fucosylated xyloglucan (FXG) from hemicellulose together with a schematic representation of monosaccharides in agreement with SNFG nomenclature (Neelamegham, Aoki-Kinoshita et al. 2019). B: Schematic representation and peptide organization of engineered diCBM77$_{Rf,}$ RSL and Janus RSL-CBM77$_{Rf}$. C: Three-dimensional structure of GalA-specific monomeric CBM77$_{Rf}$ from *Ruminococcus flavefaciens* (PDB 5FU5) and fucose-specific trimeric RSL from *Ralstonia solanacearum* (PDB 2BT9) with protein represented by ribbon and fucose represented by spheres.

In order to create an interface able to crosslink hemicellulose and pectin, we aim at building a chimeric lectin that binds both polysaccharides. In the first step, we searched for carbohydrate-binding modules (CBMs) with specificity for HG in databases such as CAZY (Drula, Garron et al. 2022) and CBM-Carb (https://cbmdb.glycopedia.eu/). This resulted in the identification of two previously characterized proteins, YeCBM32 from *Yersinia enterolitica* (Abbott, Hrynuik et al. 2007) and $CBM77_{RfPL1/9,}$ here referred as $CBM77_{Rf}$ from *Ruminococcus flavefaciens* (Venditto, Luis et al. 2016). Both of them are modules associated with pectin lyases involved in the degradation of plant cell walls and display a strong affinity for HG and oligogalacturonans with a size greater than six GalA residues. After testing for possibility of recombinant production, $CBM77_{Rf}$ was selected for the project, first for engineering it as a dimer in order to check if sufficient avidity can be obtained for pectin, and then as a multivalent chimeric construct (Figure 1). In order to create a chimeric lectin with two specificities, a hemicellulose specific-binding protein has to be associated with $CBM77_{Rf}$. It was previously demonstrated that the fucose binding lectin from bacteria *Ralstonia solanacearum*, presents strong affinity for oligomers of fucosylated xyloglucans (Kostlanová, Mitchell et al. 2005). This lectin associates as a trimer and possesses six fucose binding sites due to the presence of tandem repeats in each monomer. RSL and $CBM77_{Rf}$ were then associated as chimeric lectin through addition of a linker peptide.

Engineered lectins with specificity for plant cell wall polysaccharides

Design and production of divalent $CBM77_{Rf}$ with high avidity for homogalacturonan

The engineering of CBMs was already proven as an efficient tool for increasing protein valency and affinity (Connaris, Crocker et al. 2009, Ribeiro, Pau et al. 2016). Using similar strategy, a dimeric $CBM77_{Rf}$ ($diCBM77_{Rf}$) was designed in order to confirm that multivalent organization has the potential to increase the affinity for HG. The protein was engineered as a tandem repeat of two $CBM77_{Rf}$ connected through flexible linker ALNGSELGSGSGLSSLGEYKDI which was previously used in the design of another divalent CBM, diCBM40 (Ribeiro, Pau et al. 2016). $diCBM77_{Rf}$ is composed of 270 amino acids and was fused with N-terminal His-tag associated with TEV (tobacco etch virus) protease site. The protein was recombinantly produced in soluble form in bacteria *E. coli* BL21(DE3) and purified by immobilized metal ion chromatography followed by His-tag cleavage by TEV protease. After the TEV cleavage, the protein consisted of 251 AA with an estimated molecular weight of 25.7 kDa (Figure 2A).

Design and production of a Janus lectin with the ability to cross-link pectin and xyloglucan

The strategy to obtain a Janus lectin was to connect the gene coding for RSL monomer (N-terminus) with the one of CMB-77 domain (C-terminus) through additional oligonucleotide sequence coding for flexible linker (GGGGSGGGGS) as previously used for engineering of Janus lectin (Ribeiro, Villringer et al. 2018). The resulting protein consists of 214 amino acids and was recombinantly produced in soluble form in *Escherichia coli* KRX and subsequently purified by affinity as previously described for lectin RSL (Kostlanová, Mitchell et al. 2005).

RSL-CBM77$_{Rf}$ has an estimated molecular weight of 22.1 kDa however, SDS-PAGE analysis showed mostly the presence of protein with the estimated size of 66 kDa suggesting that RSL-CBM77$_{Rf}$ assemble as a stable trimer that does separate even during denaturation (Figure 2B). The oligomerization is under the control of RSL and the protein displays three HG binding sites site on one face and six fucose binding sites on the other face.

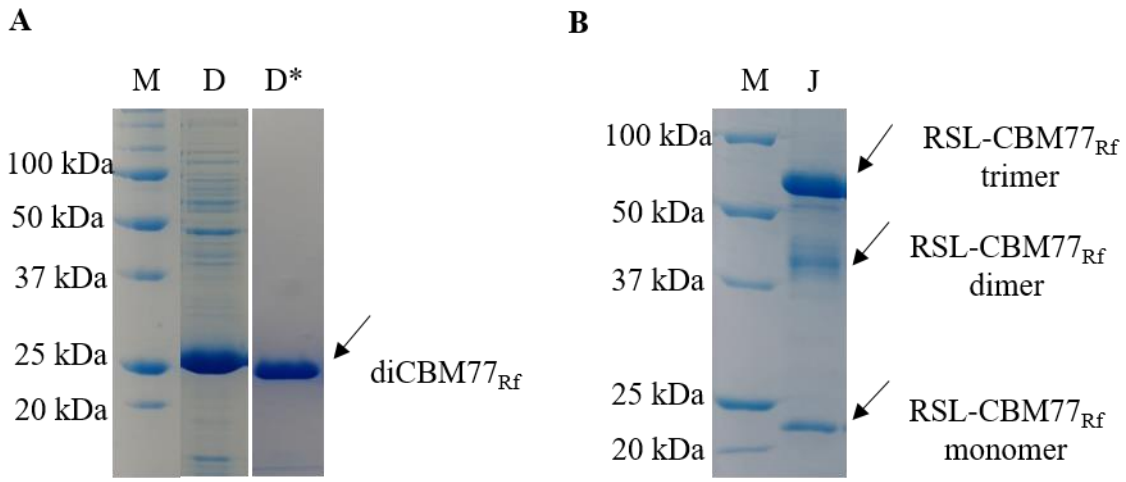


Figure 2: SDS page analysis of diCBM77$_{Rf}$ and Janus lectin RSL-CBM77$_{Rf}$. The protein samples were analyzed under denaturing conditions on 12% polyacrylamide gel. M-protein marker, D-diCBM77$_{Rf}$, D*- diCBM77$_{Rf}$ cleaved by TEV protease, J-Janus lectin RSL-CBM77$_{Rf}$. A) monomeric diCBM77$_{Rf}$ with an estimated size of 27.9 kDa. The impurities were eliminated once the protein was cleaved with TEV protease and the His-tag was removed, B) Janus lectin RSL-CBM77$_{RF}$ appears to be trimeric even in denaturation conditions, the trimerization is a result of RSL stable oligomerization. The molecular weight was estimated as 66 kDa, 44 kDa, and 22 kDa corresponding to trimeric, dimeric, and monomeric RSL-CBM77$_{Rf}$, respectively.

Binding to plant cell wall oligo- and polysaccharides

diCBM77$_{Rf}$ binding to oligo- and polygalacturonate

The efficiency of diCBM77$_{Rf}$ for binding to pectin fragments was assayed by ITC, with the comparison of affinity towards oligomers and polysaccharides. When titrating diCBM77$_{Rf}$ with hepta-galacturonic acid (GalA_DP7) large exothermic peaks were obtained, indicating exothermic binding (Figure 3A). The affinity is in the millimolar range, and the sigmoid curve could not be obtained. On the opposite, binding to homogalacturonan (HG), resulted in strong avidity, with an affinity value of 780 nM (Figure 3B). The stoichiometry (N) indicated the binding of 14 diCBM77$_{Rf}$ units per HG polysaccharide. Since the polysaccharide sample was characterized with molecular weight (MW) of approx. 35 kDa, corresponding to a degree of polymerization (DP) of 180 GalA, each diCBM77$_{Rf}$ occupies 14 GalA residues (either linear or cross-link), i.e., seven GalA per CBM77$_{Rf}$ monomer, therefore GalA_DP7 can be considered as effective binding unit. When calculating affinity and thermodynamic contribution per

CBM77$_{Rf}$ (i.e., per heptamer of GalA), the affinity is 20 µM. The engineering of a diCBM77$_{Rf}$, therefore, resulted in multiplying the avidity by 60-fold when compared to monomeric diCBM77$_{Rf}$ for the equivalent substrate.

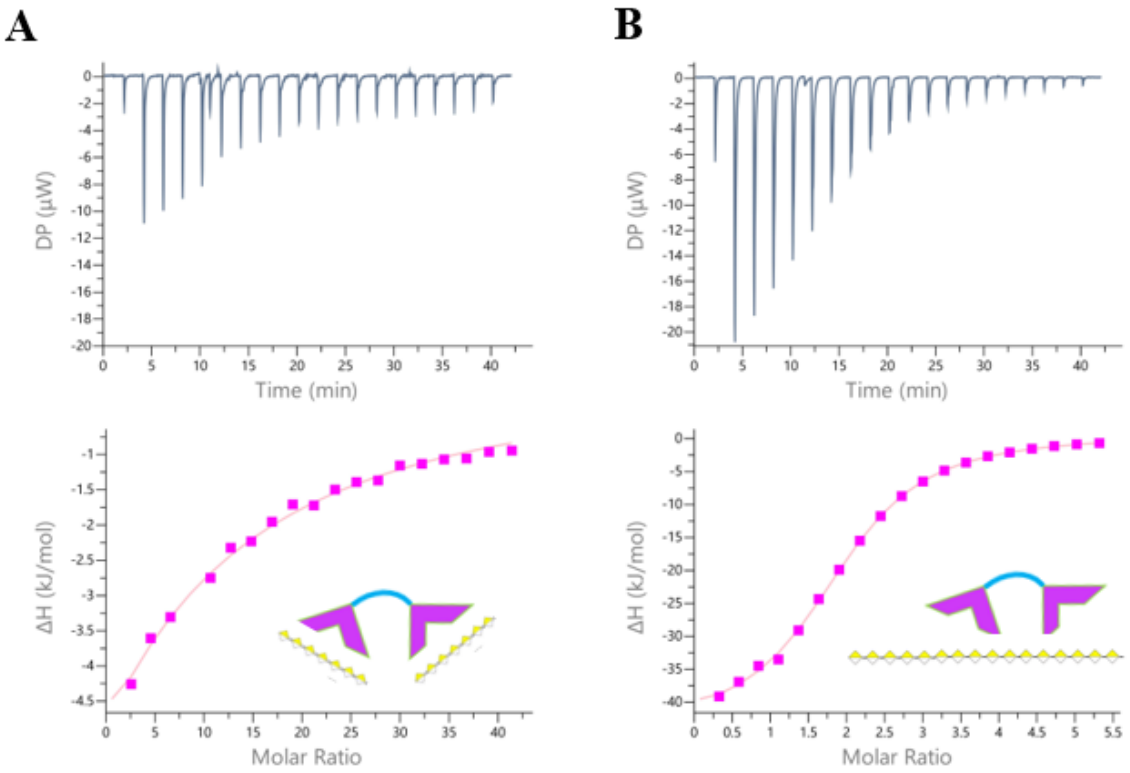


Figure 3: ITC data for the binding of diCBM77$_{Rf}$ on pectin oligosaccharide Gal_DP7 (A) and polysaccharide HG (B). The figures were prepared using ligand concentration per effective binding unit, i.e., GalA_DP7. Top: thermogram obtained by injecting oligo or polysaccharides into the protein solution. Bottom: integration of peaks (rectangles) and obtained fit (line). ITC conditions: A) 50 µM diCBM77$_{Rf}$ in buffer A with 10 mM GalA_DP7 in 20 mM Tris pH 7.5, B) 100 µM diCBM77$_{Rf}$ in buffer A with 100 µM HG in buffer A.

Janus lectin RSL-CBM77$_{Rf}$ binding to plant cell wall oligo and polysaccharides

The Janus lectin, consisting of a trimer of RSL-CBM77$_{Rf}$, was tested by ITC for its binding capacity toward polysaccharides and oligosaccharides. RSL-CBM77$_{Rf}$ binds to GalA_DP7, indicating that the addition of RSL peptide at the N-terminal extremity of CBM77$_{Rf}$ does not alter its activity. In fact, trimerization enhanced the affinity of CBM77$_{Rf}$ towards GalA_DP7, with a Kd of 103 µM, therefore 10 times higher than what was observed for diCBM77$_{Rf}$ (Figure 4A). Similarly binding to HG is more efficient for Janus lectin RSL-CBM77$_{Rf}$ than for diCBM77$_{Rf}$, resulting in a very strong affinity of 211 nM (Figure 4B). When considering GalA_DP7 as the effective binding unit for CBM77$_{Rf}$, which again corresponds closely to the observed stoichiometry (approx. 10 Janus lectin trimer per HG polysaccharides), an affinity of 4.8 µM is obtained, approximately 20 times better than for the oligomer in solution. The affinity for HG is four time better for the Janus lectin than for the diCBM77$_{Rf}$. Analysis of the

thermodynamic contribution indicates that the gain in affinity is related to a lower entropy barrier. It may be that the architecture of Janus lectin brings the three CBMs close to each other, resulting in a much more efficient binding of the substrate, and therefore stronger affinity as described for other multivalent systems (Dam and Brewer 2002). However, decrease in entropy barrier is also observed when comparing Janus lectin RSL-CBM77$_{Rf}$ and diCBM77$_{Rf}$ binding to the oligosaccharide, indicating that other effects also play a role.

Table 1: Thermodynamic data expressed as a function of the carbohydrate ligand concentration, either as the whole molecule or as the "effective binding unit" concentration of the ligand, i.e., seven GalA stretch motif for HG, one fucosylated motif or ten monosaccharide building block for apple FXG. N indicates the measured ligand/protein stoichiometry.

Prot.	Ligand	N	Kd (µM)	ΔG (kJ/mol)	ΔH (kJ/mol)	-TΔS (kJ/mol)
diCBM77$_{Rf}$						
	GalA_DP7	2[a] (1) [b]	1280 ± 100	-16.5	-63.7 ± 1.3	47.1
	HG	0.074 ± 0.00 (0.037) [b]	0.78 ± 0.02	-34.8	-1135 ± 15	1110
	HG (per GalA_DP7)	1.91 ± 0.01 (0.95) [b]	20 ± 0.4	-26.8	-44.2 ± 0.45	17.3
Janus RSL-CBM77$_{Rf}$						
	GalA_DP7	0.63 ± 0.01	103 ± 15	-22.8	-53 ± 4	30.2
	HG	0.031 ± 0.001	0.211 ± 0.01	-38.1	-1125 ± 5	1085
	HG (per GalA_DP7)	0.78 ± 0.01	4.84 ± 0.31	-30.0	-43.7 ± 0.2	13.7
	XFGol	1.88 ± 0.01	3.9 ± 0.4	-30.9	-40.1 ± 0.1	9.2
	FXG	0.168 ±0.01	0.051 ± 0.02	-41.8	-327.5 ± 7.5	285.3
	FXG (per Fuc unit)	2.0 ± 0.1	0.62 ± 0.2	-35.6	-27.3 ± 0.6	8.3
	FXG (per building block)	5.37 ± 0.26	1.65 ± 0.58	-33.2	-10.3 ± 0.3	22.9

[a]: value fixed during fitting procedure; [b]: Stoichiometry ligand/protein calculated per CBM77$_{Rf}$ monomer for comparison

As for binding to fucosylated oligosaccharides, each Janus lectin RSL-CBM77$_{Rf}$ binds with strong affinity to mono-fucosylated octasaccharide XFGol (Kd = 3.9 µM) (Figure 4C). The observed stoichiometry of 1.88 is in agreement with the occurrence of two fucose binding site per protein monomer (i.e. six per Janus lectin trimer). The obtained affinity is very similar to that measured for RSL for control experiments (Figure 5) and also corresponds to previously published results with oligosaccharide XXFG with the affinity of 2.8 µM obtained by surface plasmon resonance (Kostlanová, Mitchell et al. 2005). The Janus lectin displays much stronger avidity for apple fucosylated xyloglucan (FXG) than for oligosaccharides, with a measured Kd of 51 nM (Figure 4D). The measured stoichiometry of 0.168 corresponds to six RSL monomers (each with two fucose binding sites) per polysaccharide FXG, indicating the presence of a

minimum of 6 to 12 fucose residues per polymer. This is in agreement with the known structures of apple pomace xyloglucan (Watt, Brasch et al. 1999) and with the evaluation by gel permeation chromatography (GPC) and decomposition analysis (see material and methods). GPC revealed the estimated molecular weight of apple FXG of 55 kDa which corresponds to approximately 300 monosaccharide residues and decomposition analysis confirmed that 5% of this polymer is fucose, which corresponds to 15 monosaccharide units (Table 4). The affinity of the lectin was therefore recalculated per fucose binding unit with the value of 620 nM. The stoichiometry 2 indicates that both binding sites of the RSL part of monomeric RSL-CBM77$_{Rf}$ are occupied with the fucose. The affinity was also determined per building block (i.e. xyloglucan repeat of XFG or XXG - see Fig1A). Their average composition of ten monosaccharide units would result in 30 such building blocks per FXG polysaccharide. The observed stoichiometry of 5.4 building block for each monomeric RSL-CBM77$_{Rf}$ confirms that fucose is present in every 2 or 3 building blocks of the polymer FXG.

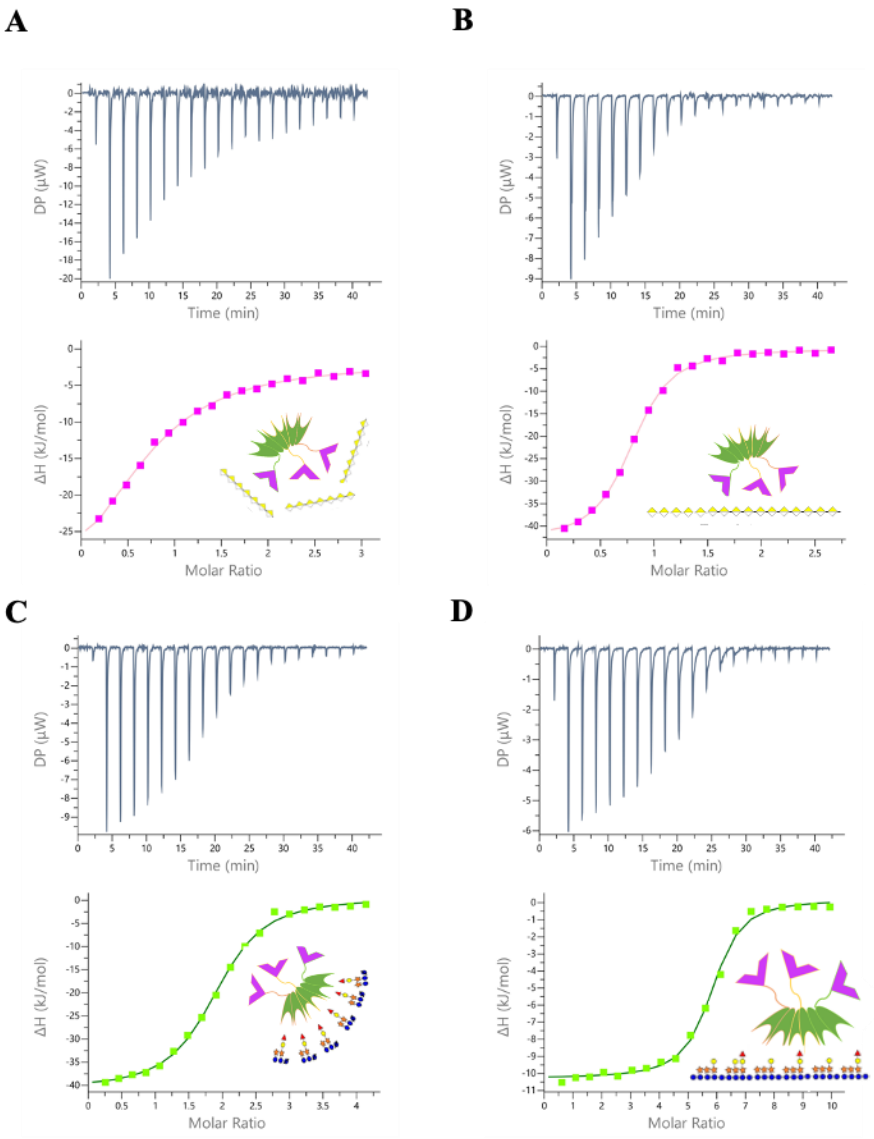


Figure 4: Comparison of ITC isotherms of RSL-CBM77$_{Rf}$ obtained by titration with oligosaccharides and polysaccharides. Top: thermograms are obtained by injecting oligo or polysaccharides into the solution of protein. The figures were prepared using ligand concentration per effective binding unit, i.e., seven GalA stretch motif for HG or ten monosaccharide building block for apple FXG. Bottom: integration of peaks (rectangles) and obtained fit (line). ITC conditions: A) 204 µM RSL-CBM77$_{Rf}$ in buffer A with 3 mM GalA_DP7 oligosaccharide in 20 mM Tris pH 7.5, B) 100 µM RSL-CBM77$_{Rf}$ in buffer A with 100 µM HG in buffer A, C) 50 µM RSL-CBM77$_{Rf}$ in buffer A with 1 mM XFGol oligosaccharide in buffer A, D) 50 µM RSL-CBM77$_{Rf}$ in buffer A with 75 µM apple FXG in buffer A.

Control experiments indicating presence of polysaccharide contaminant

Control experiments were performed with both RSL and diCBM77$_{Rf}$ in order to check the specificity.

RSL alone binds as expected on XFGol and FXG polymer (Figure 5A, B). The binding to the oligomers was identical than for Janus lectin. For binding to FXG polymer we determined the affinity and stoichiometry per fucose unit and FXG building block. As shown in Table 2 in both cases, the affinity of RSL and RSL-CBM77$_{Rf}$ are comparable. The difference is however in stoichiometry per FXG building block, suggesting the binding of 4 building blocks per RSL monomer. Nevertheless, these finding just confirmed our hypothesis and calculation about the presence of fucose 50% of building block repeats of FXG polymer.

In order to prove the specific mode of interaction for each polysaccharide control experiments were performed by ITC as well. RSL did not bind to neither oligosaccharide GalA_DP7 nor the polymer HG (Figure 5C, D). Similarly, diCBM77$_{Rf}$ was tested with oligosaccharide XFGol and apple FXG. Surprisingly, a strong binding was observed with both solutions (Figure 5E, F). More investigations were conducted and diCBM77$_{Rf}$ was tested with 2-fucosyllactose, corresponding to terminal disaccharide of XFGol, and non fucosylated xyloglucan from tamarin seeds. The experimental conditions were identical as those for XFGol and FXG and as shown on Fig. 5G and 5H, in this case no binding was observed. These findings brought us to the conclusion that diCBM77$_{Rf}$ does not interact with FXG since it does not bind to pure fucosylated oligosaccharide, nor to the defucolylated xyloglucan.

It is therefore very likely that the high affinity binding observed for diCBM77$_{Rf}$ for XFGol and FXG polysaccharide is due the presence of an unknown contaminant. Indeed, the stoichiometry of XFGol binding to diCBM77$_{Rf}$ (n=12) indicates that only 8-16 % of the ligands are active. Oligosaccharide XFGol (Elicityl) was characterized with purity 85% (provided by manufacturer) and therefore 15% of the sample composition is unknown. Similarly, apple FXG was characterized by decomposition analysis confirming the presence of additional monosaccharides such arabinose, mannose and rhamnose (Table 4). Since XFGol was produced from the same source as FXG (apple pomace), the same contamination might be present in both samples. Further investigation is needed in order to determine the actual essence of the interaction. Nevertheless, this observation is not a barrier to the construction of artificial cell wall.

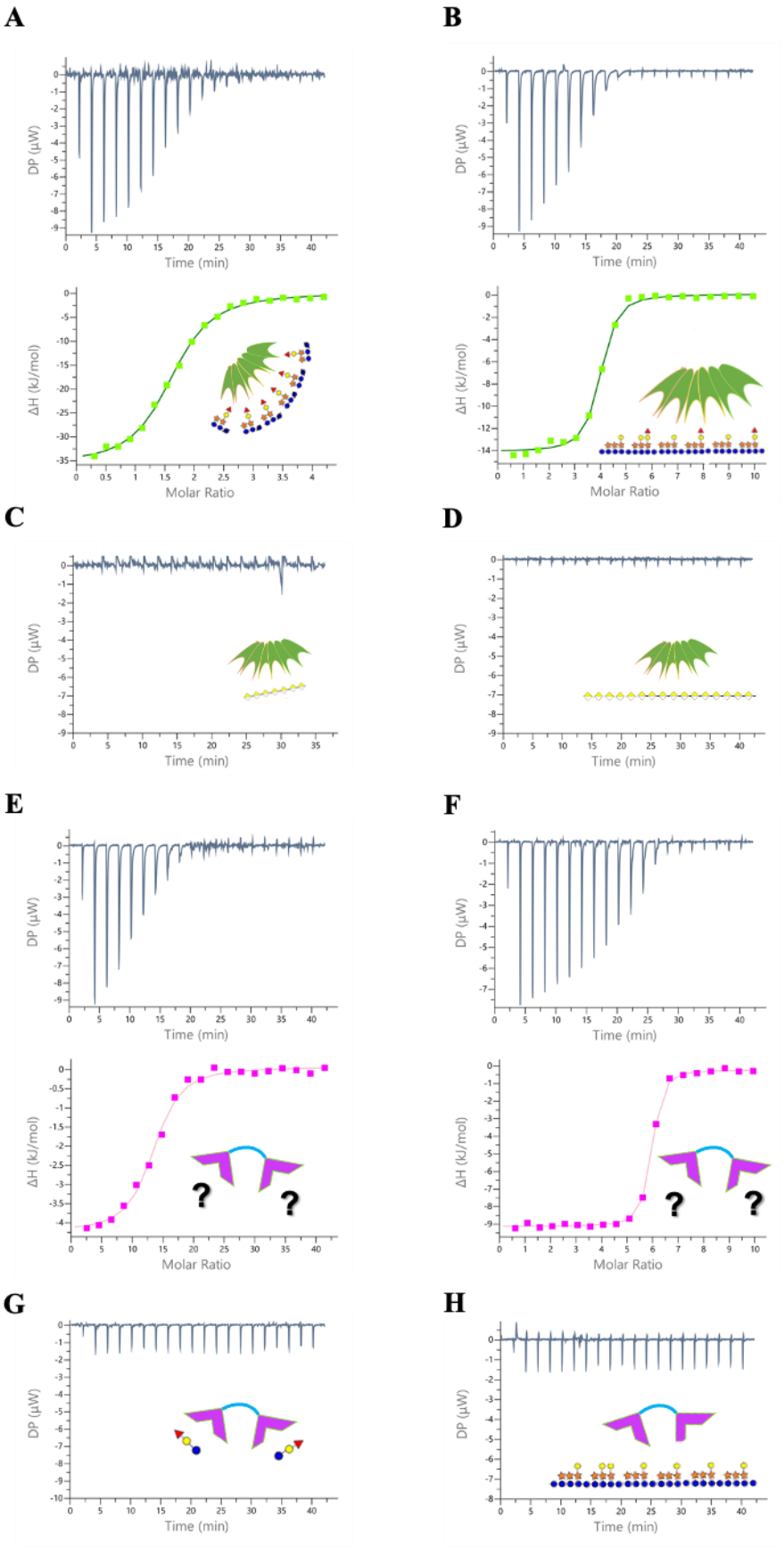


Figure 5: Control ITC experiments. ITC conditions: A) 50 µM RSL in buffer A with 1 mM XFGol in buffer A, B) 50 µM RSL in buffer A with 100 µM apple FXG in buffer A, C) 50 µM RSL in buffer

A with 10 mM GalA_DP7 oligosaccharide in 20 mM Tris pH 7.5, D) 50 µM RSL in buffer A with 50 µM HG in buffer A, E) 50 µM diCBM77$_{Rf}$ in buffer A with 10 mM XFGol in buffer A, F) 50 µM diCBM77$_{Rf}$ in buffer A with 50 µM apple FXG in buffer A, G) 50 µM diCBM77$_{Rf}$ in buffer A with 10 mM 2-fucosyl lactose in buffer A, H) 50 µM diCBM77$_{Rf}$ in buffer A with 4 µM tamarin nonfucosylated XG in buffer A. Top: thermograms obtained by injecting of oligo or polysaccharides into the protein solution. Bottom (except C, D): integration of peaks (rectangles) and obtained fit (line).

Table 2: Thermodynamic data of control experiments. The data were prepared in the same way as previously shown in Table 1.

Prot.	Ligand	N	Kd (µM)	ΔG (kJ/mol)	ΔH (kJ/mol)	-TΔS (kJ/mol)
diCBM77$_{Rf}$						
	XFGol	11.85 ± 0.95	21.6 ± 1.9	-26.6	-3.96 ± 0.37	22.7
	FXG	0.12 ± 0.00	0.013 ± 0.01	-45.5	-427.5 ± 0.5	381.5
	FXG (per building block)	5.72 ± 0.02	0.62 ± 0.3	-35.9	-8.9 ± 0.0	27.1
	2-fucosyl lactose		nb			
	XG		nb			
RSL						
	GalA_DP7		nb			
	HG		nb			
	XFGol	1.69 ± 0.1	5.1 ± 1.1	-30.3	-36.9 ± 0.9	6.6
	FXG	0.163 ± 0.01	0.46 ±0.01	-41.9	-343 ± 3	301.5
	FXG (per Fuc units)	2 ± 0.02	0.55 ± 0.08	-35.7	-28.6 ± 0.3	7.2
	FXG (per building block)	3.9 ± 0.1	1.1 ± 0.2	-34.0	-14.3 ± 0.1	19.8

nb: no binding observed

Building of supported synthetic plant cell wall

A multilayer model of plant cell walls was prepared and analyzed qualitatively through Quartz Crystal Microbalance with energy Dissipation monitoring technique (QCM-D). QCM-D measurements were used to monitor the adsorption of CNCs, polysaccharides, and proteins, on the surface of a silica-coated quartz crystal that was first functionalized with a thin film of polyethylene-imine (PEI), a branched cationic polymer which was previously demonstrated to provide a strong basis for adhesion of CNC (Navon, Jean et al. 2020) or with supported lipid bilayer (SLB).

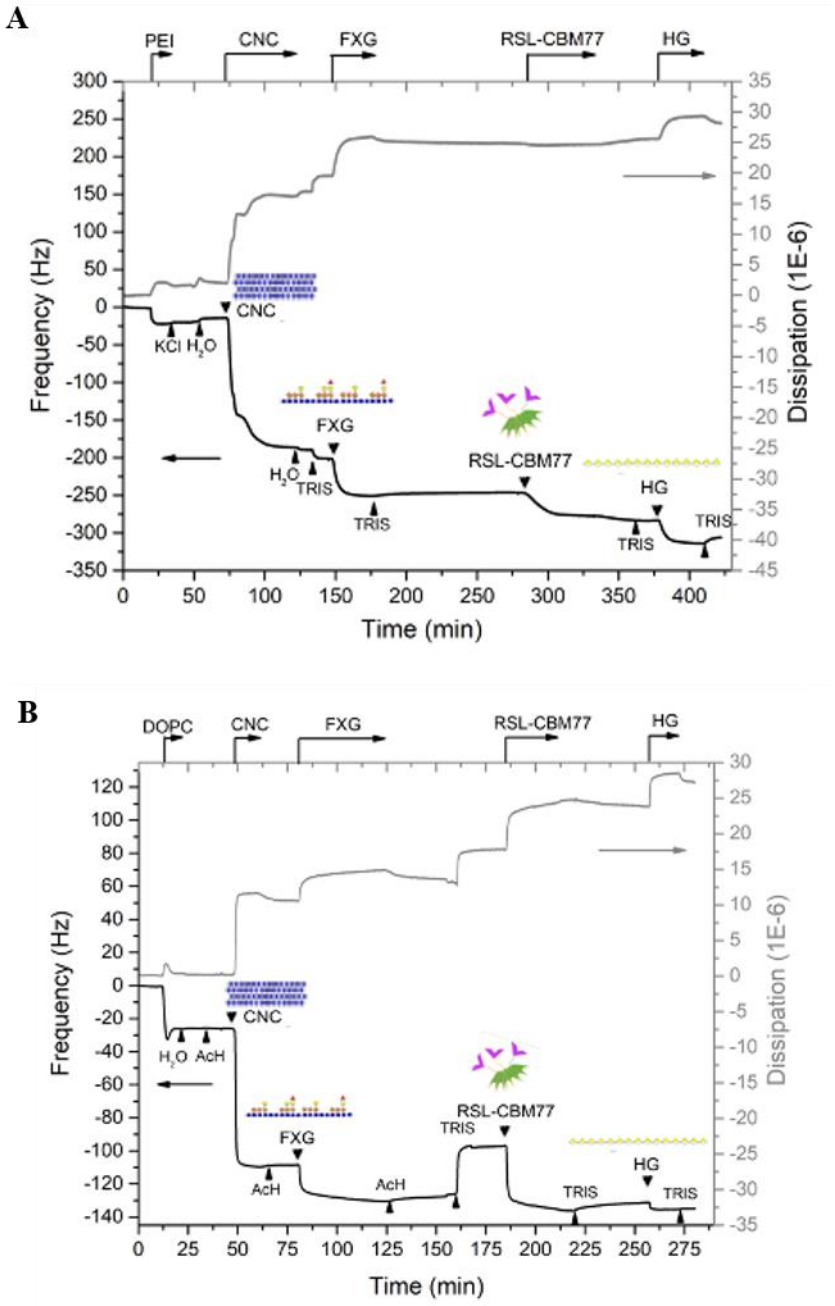


Figure 6: **Frequency and dissipation variation over time of the quartz surface for the 7th overtone with layer-by-layer stacking**. **A:** PEI (1g.L^{-1}) in 500mM KCl , CNC (1g.L^{-1}) and FXG (1g.L^{-1}) in pure water. This was followed by rinsing with 20mM Tris (pH=7.5), 100mM NaCl, an injection of RSL-CBM77$_{RfPL1/9}$ (0.2, 2, and 10µM), and injection of HG in the same buffer. **B:** DOPC SLB (0.1g.L^{-1}) in water, CNC (1g.L^{-1}) and FXG (1g.L^{-1}) in 100mM acetate. This was followed by rinsing with 20mM Tris (pH=7.5), 100mM NaCl, injection of RSL-CBM77$_{Rf.}$ and purified HG in the same buffer

Figure 6A displays the evolution of resonance frequency (related to the change in mass) and energy dissipation of the shear oscillation (related to the viscoelastic properties of the

oscillating mass) for each layer deposited on the quartz crystal covered with PEI. The addition of solution of nanocrystals generates a large shift in frequency, followed by less marked ones for the addition of FXG, Janus RSL-CBM77$_{Rf}$ lectin and HG (Table 3).

Table 3: Frequency and dissipation variation quantified from experiments displayed in Figure 6.

Film supported on PEI			**Film supported on SLB**		
Layer buffer	$\Delta f_7/7$ (Hz)	ΔD_7 (10^{-6})	Layer buffer	$\Delta f_7/7$ (Hz)	ΔD_7 (10^{-6})
PEI(KCl)	-20.5	+1.5	**SLB (water)**	-25.2	+0.2
CNC (water)	-175.5	+15	**CNC (acetate)**	-82.5	+10.4
FXG (Tris)	- 44.5	+5	**FXG (acetate)**	-19.1	+3
RSL-CBM77$_{Rf}$ (Tris)	-38	+0.7	**RSL-CBM77$_{Rf}$ (Tris)**	-34.8	+6.1
HG (Tris)	-23.5	+2.4	**HG (Tris)**	-3.5	+3.4

In a second experiment, the multilayer film was built on a supported lipid bilayer (SLB) of DOPC that more closely mimics natural plant cell wall than PEI. Deposition of DOPC vesicles that open on the glass and form a lipid bilayer. The CNCs were then added as described previously resulting in a strong frequency change. The amplitude was smaller than the one obtained on PEI but in agreement with a previous study on SLB (Navon, Jean et al. 2020). Similarly, the deposition of FXG resulted in a clear frequency shift. Except for a necessary change of buffer (from acetate to Tris) that creates small shifts in the frequency due to the change in solution viscosity, the addition RSL-CBM77$_{Rf}$ did result in a frequency shift equivalent to the one measured with the PEI-functionalized surface. A smaller variation is finally obtained when adding the homogalacturonan solution.

Regarding the changes in energy dissipation for each layer, we observed a noticeable increase in *D* upon adsorption of CNC, FXG, RSL-CBM77, and HG layers. The ratio ($\Delta D/-\Delta f$) measuring the change in energy dissipation per surface-coupled mass unit characterizes the softness of the film, and it was demonstrated that a homogeneous film should be considered as rigid only for ($\Delta D/-\Delta f$) << 4 x 10^{-7} Hz^{-1}. (Reviakine, Johannsmann et al. 2011). In our case, the ratios were found slightly lower or closed to this threshold value. These values are related to the viscoelastic properties of the films indicating the formation of soft and highly hydrated films.

Control experiments were performed with RSL instead of Janus lectin, resulting in smaller step in frequency upon addition, consistent with its lower molecular weight. The absence of the CBM77$_{Rf}$ moiety on the lectin resulted in total absence of binding of homogalacturonan for the last layer (Figure 7). The other control experiments made use of non fucosylated XG instead of FXG. In this case, no change in frequency is observed after RSL-CBM77$_{Rf}$ injection confirming the need of fucose for the addition of this layer (Figure 7).

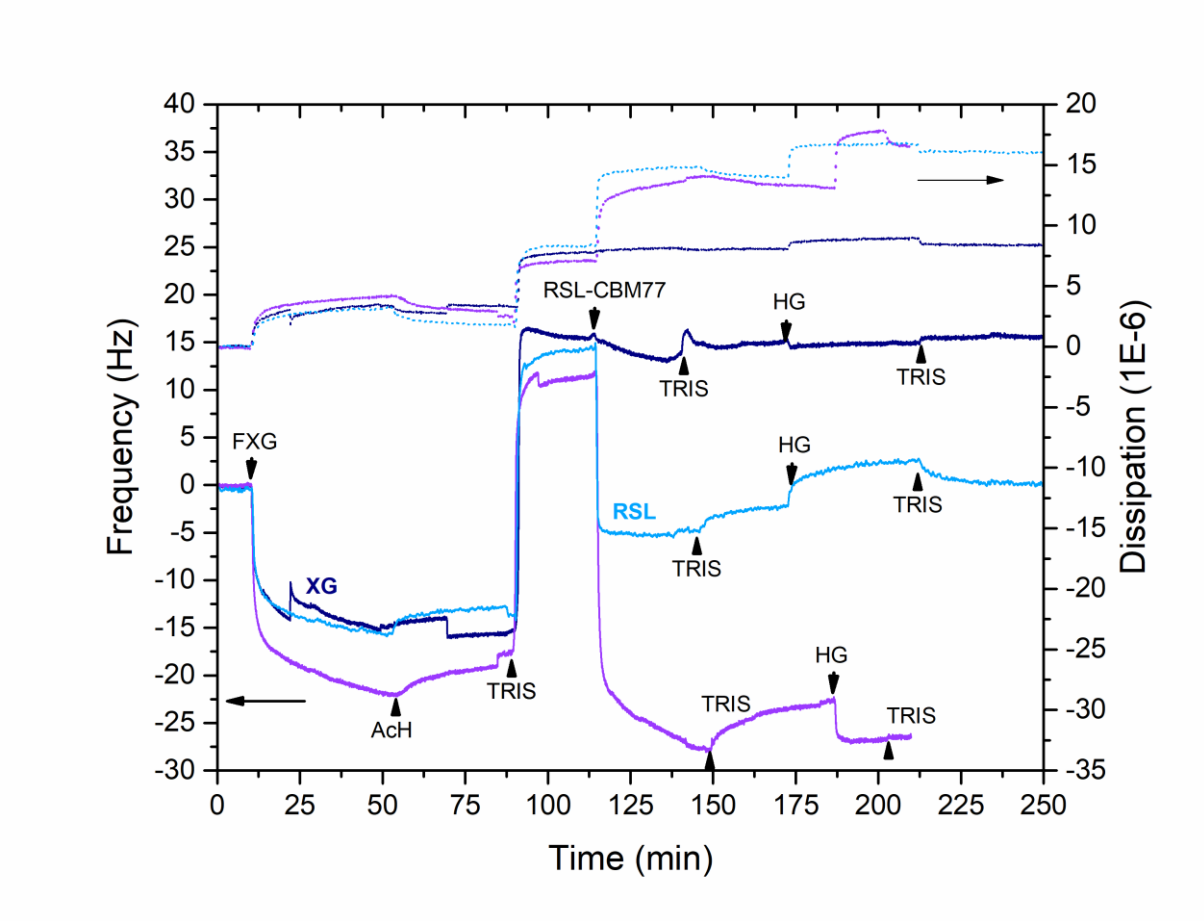


Figure 7: **Overlay of control experiments.** All the experiments were performed on multilayer constructure on SLB whereas either Janus lectin RSL-CBM77$_{Rf}$ was replaced with RSL (light blue) and no binding of HG is observed in the last step or the nonfucosylated XG (dark blue) was used instead of FXG and in this case, no binding of RSL-CBM77$_{Rf}$ nor of HG was observed. The original data with FXG and RSL-CBM77$_{Rf}$ are shown in violet.

Discussion

In the development of synthetic biology, the obtainment of artificial cells is a goal pursued by many teams. General efforts concentrate on the content of such cell, i.e. minimum genetic content to reproduce life, while it is also necessary to re-create surfaces of pseudo-cells with properties that are close enough to the real ones. Mimicking the surface of animal cells necessitates the incorporation of glycolipids, glycoproteins or other glycoconjugates to create the glycocalyx necessary for cell-cell interaction. In the case of plant cells, the complexity is much higher and polysaccharides with very different nature and properties must be assembled for obtaining the mechanical and biological properties obtained in Nature. We demonstrated here that the addition of a small fraction of well-designed protein participates in the building of cell wall mimics.

The diCBM77$_{Rf}$ that has been constructed here has the ability to label strongly galacturonans, which could be of interest as a biomarker. Indeed, a limited number of antibodies are available as markers of polysaccharides in plants, and the molecular structure recognized by these antibodies often remains poorly characterized (Ruprecht, Bartetzko et al. 2017). CBMs are more robust than antibodies but display weak affinity due to their monovalent character. Engineered diCBMs with specificities towards a component of plant cell walls may therefore represent useful tools for analyzing plant polysaccharides. Due to the large catalog of specificities toward plant polysaccharides exhibited by CBMs (Boraston, Bolam et al. 2004, Duffieux, Marcus et al. 2020) , the concept can be extended towards other cell wall components.

This CBM77$_{Rf}$ associated with lectin RSL resulted in a new tool for inducing the association of two soluble plant polysaccharides, i.e. pectin (homogalacturonan) and hemicellulose (fucosylated xyloglucan). The supported artificial plant cell wall that has been constructed is characterized by a single layer of cellulose nanocrystals on a lipid bilayer, with a thickness of 4-5 nm for SLB and 6-8 nm for CNCs in the case of a quasi-monolayer, according to previous work. FXG addition resulted in a more limited thickness of approximately 2 nm (Navon 2020). The introduction of protein in the cell wall model allowed for completion with a thin pectin layer, but also influenced the flexibility, as indicated by analysis of the dissipation.

In conclusion, supported cell wall mimics that have been obtained in this study are of interest for research purposes. This opens the route for obtaining artificial plant cell walls. Cellulose-based containers built on lipid vesicles have already been described and completing such construct with different polysaccharides and proteins is now possible.

Material and Methods

Oligosaccharides and polymers preparation

Oligosaccharide XFGol was purchased from Elicityl (PEL101GV) and the reported 85% purity was checked by NMR analysis. GalA_DP7 was obtained from the CERMAV collection (Grenoble, France) (Gouvion, Mazeau et al. 1994). A purity check was performed by gel permeation chromatography indicating that the sample consists mainly of DP7 with limited contamination by DP8. 2-fucosyl lactose was obtained from CERMAV collection and the purity was checked by NMR analysis. Homogalacturonan (HG) from orange was obtained from Sigma (P3889) and further purified by dialysis in order to remove impurities and small polymers. The absence of the methyl ester functional group was verified by NMR analysis. Gel permeation chromatography on SB 806 M HQ column indicated an MW of 35 kDa. Nonfucosylated xyloglucan from tamarin fruit (Glyloid 3S, Dainippon) was further purified by ethanol in order to remove the contaminants. Fucosylated xyloglucan (FXG) was obtained from apple pomace, inspired by [35]. Gel permeation chromatography on SB 806 M HQ column gave a molecular weight of 55 kDa for FXG and acidic hydrolysis treatment followed by anion-exchange liquid chromatography indicated that the fucose content is 5% (Table 4).

Table 4: Monosaccharide composition of apple FXG presented as mass percentage.

	Fuc	Rha	Ara	Gal	Glc	Man	Xyl	Total
%	5.30	3.71	3.85	10.85	29.49	9.05	21.39	83.63
mass*	± 0.32	± 0.13	± 0.10	± 0.61	± 0.94	± 0.23	± 0.80	± 3.14

* the mass percentage corresponds to monosaccharide mass (in g) in 100 g of sample.

The CNCs suspensions were prepared by acid hydrolysis of cotton linters (Buckeye Cellulose Corporation) according to reported method (Dong, Revol et al. 2998) in the form of a suspension at 2.47% wt obtained by hydrolysis of cotton fiber with sulfuric acid. This treatment has the particularity to graft sulfate groups on the CNC, covering them with negative charges. The CNC suspension was subsequently sonicated for 30 s at 10% and then filtered using a 0.45 µm syringe filter.

Gene design and cloning

The original amino acid sequence of the carbohydrate-binding module $CBM77_{Rf}$ from *Ruminococcus flavefaciens* was obtained from the PDB database. The gene *dicbm77$_{Rf}$* was designed as a tandem repeat of two individual $CBM77_{Rf}$ domains connected *via* ALNGSELGSGSGLSSLGEYKDI linker. The gene *rsl-cbm77$_{Rf}$* was designed as a fusion chimera with RSL at N-terminus and $CBM77_{Rf}$ at C-terminus *via* linker GGGGSGGGGS. The genes were ordered from Eurofins Genomics (Ebersberg, Germany) after codon optimization for the expression in the bacteria *Escherichia coli*. The restriction enzyme sites of NdeI and XhoI were added at 5' and 3' ends, respectively. The synthesized genes were delivered in plasmid pEX-A128-di$CBM77_{Rf}$ and pEX-A128-RSL-$CBM77_{Rf}$ respectively. The plasmid pEX-A128-di$CBM77_{Rf}$ and the pET-TEV vector (Houben, Marion et al. 2007) were digested with the NdeI and XhoI restriction enzymes to ligate *dicbm77$_{Rf}$* in pET-TEV to fuse a 6-His Tag cleavable with TEV protease at the *N*-terminus of di$CBM77_{Rf}$. Similarly, plasmids pEX-A128-RSL-$CBM77_{Rf}$ and the pET-25b+ were digested with the NdeI and XhoI restriction enzymes to ligate *rsl-cbm77$_{Rf}$* into pET-25b+. After transformation by heat shock in *E. coli* DH5α strain, a colony screening was performed, and the positive plasmids were amplified and controlled by sequencing.

Protein expression

E. coli BL21(DE3) cells were transformed by heat shock with pET-TEV-di$CBM77_{Rf}$ plasmid prior pre-culture in Luria Broth (LB) media with 25 µg/mL kanamycin at 37°C under agitation at 180 rpm overnight. The next day, 10 mL of pre-culture was used to inoculate 1 L LB medium with 25 µg/mL kanamycin at 37°C and agitation at 180 rpm. When the culture reached OD_{600nm} of 0.6 - 0.8, the protein expression was induced by adding 0.1 mM isopropyl β-D-thiogalactoside (IPTG), and the cells were cultured at 16°C for 20 hours.

Correspondingly, *E. coli* KRX cells were transformed by heat shock with pET-25b+-RSL-$CBM77_{Rf}$ plasmid and pre-cultured in LB media substituted with 50 µg/mL ampicillin at 37°C under agitation at 180 rpm overnight. The following day, 10 mL of pre-culture was used to inoculate 1 L LB medium with 50 µg/mL ampicillin at 37°C and agitation at 180 rpm. When

reached OD_{600nm} of 0.6 - 0.8, the protein expression was induced by adding 1% L-rhamnose, and the cells were cultured at 16°C for 20 hours.

The cells were harvested by centrifugation at 14000 × *g* for 20 min at 4°C and the cell paste was resuspended in 20 mM Tris/HCl pH 7.5, 100 mM NaCl (Buffer A), and lysed by a pressure cell disruptor (Constant Cell Disruption System) with a pressure of 1.9 kBar. The lysate was centrifuged at 24 000 × *g* for 30 min at 4°C and filtered on a 0.45 µm syringe filter prior to loading on an affinity column.

Protein purification

diCBM77$_{Rf}$

The cell lysate was loaded on 1 mL HisTrap column (Cytiva) pre-equilibrated with Buffer A. The column was washed with Buffer A to remove all contaminants and unbound proteins. The diCBM77$_{Rf}$ was eluted by Buffer A in steps during which the concentration of imidazole was increased from 50 mM to 500 mM. The fractions were analyzed by 12% SDS PAGE and those containing diCBM77$_{Rf}$ were collected and deprived of imidazole by dialysis in Buffer A. N-terminal His-tag was removed by TEV cleavage with the ratio 1:50 mg of TEV:protein in the presence of 0.5 mM EDTA and 1mM TCEP over night at 19°C. After, the protein mixture was purified on 1 mL HisTrap column (Cytiva) and the pure protein was concentrated by Pall centrifugal device with MWCO 10 kDa and stored at 4°C.

RSL- CBM77$_{Rf}$

Likewise, the cell lysate was loaded on 10 mL D-mannose-agarose resin (Merck) pre-equilibrated with Buffer A. The column was washed with Buffer A to remove all contaminants and unbound proteins and the flow-through was collected. RSL-CBM77$_{Rf}$ was eluted by Buffer A with the addition of 100 mM D-mannose in one step. Due to not sufficient binding capacity of the column, the flow-through was reloaded on the column several times and the protein was eluted as described previously. The fractions were analyzed by 12% SDS PAGE and those containing RSL-CBM77$_{Rf}$ were collected and dialyzed against Buffer A. The protein was concentrated by Pall centrifugal device with MWCO 10 kDa and the pure protein fractions were pooled, concentrated, and stored at 4°C.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed with MicroCaliTC200 (Malvern Panalytical). Experiments were carried out at 25°C ± 0.1°C. Proteins and ligands samples were prepared in Buffer A, except of oligosaccharide GalA_DP7 which was prepared in 20 mM Tris pH 7.5. The ITC cell contained proteins in a concentration range from 0.05 mM to 0.2 mM. The syringe contained the ligand solutions in a concentration from 50 µM to 10 mM. 2 µL of ligands solutions were injected into the sample cell at intervals of 120 s while stirring at 750 rpm. Integrated heat effects were analyzed by nonlinear regression using one site binding model (MicroCal PEAQ-ITC Analysis software). The experimental data were fitted to a theoretical curve, which gave the dissociation constant (Kd) and the enthalpy of binding (ΔH).

Decomposition analysis of FXG

Apple FXG 4 mg was placed in a dry bath in the presence of 1 mL of 2N TFA, for 4 h at 110°C. After cooling, the sample was filtered on 0.2 µm then diluted 40 times in reverse osmosis water. Subsequently, the sample was injected on anion exchange liquid chromatography, on a DIONEX ICS6000 system equipped with a CarboPac PA1 pre-column and column (250*2 mm) and a PAD (pulsed amperometry) detector. The injected sample volume was 5 µL and the purification run at 28°C. The elution was carried out at 0.25 mL/min under the following conditions: T0 to T18 min: NaOH 0.016 M, T18 to T20 min: NaOH 0.016 M + linear gradient in NaOAc from 0 to 0.02 M, T20 to T25: linear gradient in NaOH from 0.016 M to 0.2 M and linear gradient in NaOAc from 0.02 M to 0.5 M, T25.1 to T30: NaOH 0.1 M + NaOAc 0.5 M. A calibration curve for each monosaccharide (fucose, rhamnose, arabinose, galactose, glucose, mannose, xylose, ribose, galacturonic acid, glucuronic acid was performed ($2*10^{-3}$ to $3*10^{-2}$ mg/mL) in the same operating conditions. The sample is injected three times.

Quartz crystal microbalance with dissipation monitoring (QCM-D) measurement

QCM-D measurements were performed using Q-Sense E4 instruments (Biolin Scientific) equipped with one to four flow modules. Silica-coated AT-cut quartz crystals (QSX 303) were purchased from Biolin Scientific (Stockholm, Sweden). Besides measurement of bound mass, which is provided from changes in the resonance frequency, f, of the sensor crystal, the QCM-D technique also provides structural information of biomolecular films via changes in the energy dissipation, D, of the sensor crystal. f and D were measured at the fundamental resonance frequency (4.95 MHz) as well as at the third, fifth, seventh, ninth, eleventh, and thirteenth overtones (i = 3, 5, 7, 9, 11 and 13). Normalized frequency shifts $Df = Df_i/i$ and dissipation shifts $\Delta D = \Delta D_i$ corresponding to the seventh overtone only were presented for clarity.

The sensors were rinsed with a 2 wt % SDS aqueous solution followed by abundant water rinsing and dried under a nitrogen stream. The sensors were then exposed to a UV/ozone lamp for 10 min prior to the measurements.

Flow rate was controlled using a peristaltic pump (ISM935C, Ismatec, Switzerland). The temperature of the E4 QCM-D platform and all solutions were stabilized to ensure stable operation at 24 °C. All buffers were previously degassed in order to avoid air bubble formation in the fluidic system. All experiments were performed in three replicates.

Polyethylene-imine (PEI), a branched cationic polymer which can serve as a good primer was injected (1 g. L^{-1}) in water. In all the experiments done with PEI, the tubing diameter was 0.25 mm and the flow rate was set at 10 µL.min^{-1}. Second injection was performed with CNC (1 g. L^{-1}) then FXG (1 g. L^{-1}) in pure water. This was followed by rinsing with 20 mM Tris (pH=7.5), 100 mM NaCl and injection of RSL-$CBM77_{Rf}$ (different concentrations up to 10 µM) and then HG (1 g. L^{-1}) in the same buffer. Nonfucosylated XG from tamarin or RSL were used in control experiments.

To produce an SLB on QCM-D, a small unilamellar vesicles (SUVs) suspension was injected into the measurement chamber. The DOPC (1,2-dioleoyl-sn-glycero-3-phosphocholine) SUVs were prepared as previously described (Navon, Jean et al. 2020) and injected in 100 mM acetate

(pH=3). With SLB surface was done using tubing with a 0.75 mm diameter and a flow of 50 µL.min^{-1}. CNC (1 g. L^{-1}) was then injected in the same buffer followed by FXG (1 g. L^{-1}) then rinsed with 20 mM Tris (pH=7.5). This buffer was also used for injection of RSL-CBM77$_{Rf}$ (10 µM) and HG (1 g. L^{-1}).

Acknowledgements

This research was funded by the European Union Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement synBIOcarb (No. 814029). The authors thank the nano-bio platform at ICMG, DCM, University of Grenoble Alpes (ICMG FR 2607) for the use of QCM-D as well as the Glyco@Alps program (Investissements d'Avenir Grant No. ANR-15-IDEX-02) and Labex Arcane/CBH- EUR-GS (ANR-17-EURE-0003). The authors acknowledge PCANS platform at Cermav CNRS, especially Laurine Buon and Eric Bayma for the oligo and polysaccharide characterization. We would also like to thank Microbiology platform at Cermav CNRS for 2-fucosyl lactose.

Abbreviations

CBM – carbohydrate-binding module, CNCs – crystal nanocellulose, FCX – fucosylated xyloglucan, GalA_DP7 - hepta-galacturonic acid, HG – homogalacturonan, PEI - polyethylene-imine, SLB – supported lipid bilayer, XFGol - mono-fucosylated octasaccharide

Author contributions: S.N., N.C., L.R. and A.I. designed the sequence, produced, analysed, and characterised the lectins by ITC, N.C., L.C. and L.H. performed QCM-D experiments, M.T. provided and characterized HG, J.M. provided and characterized FXG and nonfucosylated XG, S.N., L.C., A.I., and L.H. wrote the manuscript and prepared figures, N.C., L.R., Y.N., L.C., O.L., L.H., and A.I. further revised the manuscript.

Competing interests. The authors declare no competing interests.

References

Abbott, D. W., S. Hrynuik and A. B. Boraston (2007). "Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from Yersinia enterolitica." J Mol Biol **367**(4): 1023-1033.

Bonnardel, F., J. Mariethoz, S. Salentin, X. Robin, M. Schroeder, S. Pérez, F. Lisacek and A. Imberty (2019). "UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands." Nucleic Acids Research **47**: D1236–D1244.

Boraston, A. B., D. N. Bolam, H. J. Gilbert and G. J. Davies (2004). "Carbohydrate-binding modules: fine-tuning polysaccharide recognition." Biochem J **382**(Pt 3): 769-781.

Connaris, H., P. R. Crocker and G. L. Taylor (2009). "Enhancing the receptor affinity of the sialic acid-binding domain of Vibrio cholerae sialidase through multivalency." J Biol Chem **284**(11): 7339-7351.

Dam, T. K. and C. F. Brewer (2002). "Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry." Chem. Rev. **102**(2): 387-429.

Dong, X. M., J.-F. Revol and D. G. Gray (2998). "Effect of microcrystallite preparation conditions on the formation of colloid crystals of cellulose." Cellulose **5**: 19-32.

Drula, E., M. L. Garron, S. Dogan, V. Lombard, B. Henrissat and N. Terrapon (2022). "The carbohydrate-active enzyme database: functions and literature." Nucleic Acids Res **50**(D1): D571-D577.

Duffieux, D., S. E. Marcus, J. P. Knox and C. Herve (2020). "Monoclonal Antibodies, Carbohydrate-Binding Modules, and Detection of Polysaccharides in Cell Walls from Plants and Marine Algae." Methods Mol Biol **2149**: 351-364.

Gilbert, H. J., J. P. Knox and A. B. Boraston (2013). "Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules." Curr Opin Struct Biol **23**(5): 669-677.

Gouvion, C., K. Mazeau, A. Heyraud, F. R. Taravel and I. Tvaroska (1994). "Conformational study of digalacturonic acid and sodium digalacturonate in solution." Carbohydr Res **261**(2): 187-202.

Haas, K. T., R. Wightman, E. M. Meyerowitz and A. Peaucelle (2020). "Pectin homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells." Science **367**(6481): 1003-1007.

Haas, K. T., R. Wightman, A. Peaucelle and H. Hofte (2021). "The role of pectin phase separation in plant cell wall assembly and growth." Cell Surf **7**: 100054.

Habibi, Y., L. A. Lucia and O. J. Rojas (2010). "Cellulose nanocrystals: chemistry, self-assembly, and applications." Chem Rev **110**(6): 3479-3500.

Hanus, J. and K. Mazeau (2006). "The xyloglucan-cellulose assembly at the atomic scale." Biopolymers **82**(1): 59-73.

Hayashi, T., T. Takeda, K. Ogawa and Y. Mitsuishi (1994). "Effects of the degree of polymerization on the binding of xyloglucans to cellulose." Plant Cell Physiol **35**(6): 893-899.

Houben, K., D. Marion, N. Tarbouriech, R. W. Ruigrok and L. Blanchard (2007). "Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family." J Virol **81**(13): 6807-6816.

Jamet, E., H. Canut, G. Boudart and R. F. Pont-Lezica (2006). "Cell wall proteins: a new insight through proteomics." Trends Plant Sci **11**(1): 33-39.

Kostlanová, N., E. P. Mitchell, H. Lortat-Jacob, S. Oscarson, M. Lahmann, N. Gilboa-Garber, G. Chambat, M. Wimmerová and A. Imberty (2005). "The fucose-binding lectin from *Ralstonia solanacearum*: a new type of b -propeller architecture formed by

oligomerisation and interacting with fucoside, fucosyllactose and plant xyloglucan." J. Biol. Chem. **280**: 27839-27849.

Lombard, V., H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat (2014). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic Acids Res **42**(Database issue): D490-495.

Lombardo, S. and A. Villares (2020). "Engineered Multilayer Microcapsules Based on Polysaccharides Nanomaterials." Molecules **25**(19).

Lopez, M., H. Bizot, G. Chambat, M. F. Marais, A. Zykwinska, M. C. Ralet, H. Driguez and A. Buleon (2010). "Enthalpic studies of xyloglucan-cellulose interactions." Biomacromolecules **11**(6): 1417-1428.

Mazeau, K. and L. Charlier (2012). "The molecular basis of the adsorption of xylans on cellulose surface." Cellulose **19**: 337-349.

McConnell, H. M., T. H. Watts, R. M. Weis and A. A. Brian (1986). "Supported planar membranes in studies of cell-cell recognition in the immune system." Biochim Biophys Acta **864**(1): 95-106.

McNeil, M., A. G. Darvill, S. C. Fry and P. Albersheim (1984). "Structure and function of the primary cell walls of plants." Annu Rev Biochem **53**: 625-663.

Navon, Y. (2020). Interaction of plant cell wall building blocks : towards a bioinspired model system, Université Grenoble Alpes & Ben-Gurion university of the Negev.

Navon, Y., B. Jean, L. Coche-Guerente, F. Dahlem, A. Bernheim-Groswasser and L. Heux (2020). "Deposition of Cellulose Nanocrystals onto Supported Lipid Membranes." Langmuir **36**(6): 1474-1483.

Neelamegham, S., K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, T. Lutteke, N. O'Boyle, N. H. Packer, P. Stanley, P. Toukach, A. Varki, R. J. Woods and S. D. Group (2019). "Updates to the Symbol Nomenclature for Glycans guidelines." Glycobiology **29**(9): 620-624.

Paulraj, T., S. Wennmalm, D. C. F. Wieland, A. V. Riazanova, A. Dedinaite, T. Gunther Pomorski, M. Cardenas and A. J. Svagan (2020). "Primary cell wall inspired micro containers as a step towards a synthetic plant cell." Nat Commun **11**(1): 958.

Pauly, M., P. Albersheim, A. Darvill and W. S. York (1999). "Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants." Plant J **20**(6): 629-639.

Rånby, B. G. (1949). "Aqueous colloidal solutions of cellulose micelles." Acta Chemica Scandinavica **3**: 649-650.

Reviakine, I., D. Johannsmann and R. P. Richter (2011). "Hearing what you cannot see and visualizing what you hear: interpreting quartz crystal microbalance data from solvated interfaces." Anal Chem **83**(23): 8838-8848.

Ribeiro, J., W. Pau, C. Pifferi, O. Renaudet, A. Varrot, L. K. Mahal and A. Imberty (2016). "Characterization of a high-affinity sialic acid specific CBM40 from *Clostridium perfringens* and engineering of divalent form." Biochemical Journal **473**: 2109-2118.

Ribeiro, J. P., S. Villringer, D. Goyard, L. Coche-Guerente, M. Höferlin, O. Renaudet, W. Römer and A. Imberty (2018). "Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells." Chemical Science **9**: 7634 - 7641

Richter, R. P. and A. R. Brisson (2005). "Following the formation of supported lipid bilayers on mica: a study combining AFM, QCM-D, and ellipsometry." Biophys J **88**(5): 3422-3433.

Ruprecht, C., M. P. Bartetzko, D. Senf, P. Dallabernadina, I. Boos, M. C. F. Andersen, T. Kotake, J. P. Knox, M. G. Hahn, M. H. Clausen and F. Pfrengle (2017). "A Synthetic Glycan Microarray Enables Epitope Mapping of Plant Cell Wall Glycan-Directed Antibodies." Plant Physiol **175**(3): 1094-1104.

Sede, A. R., C. Borassi, D. L. Wengier, M. A. Mecchia, J. M. Estevez and J. P. Muschietti (2018). "Arabidopsis pollen extensins LRX are required for cell wall integrity during pollen tube growth." FEBS Lett **592**(2): 233-243.

Teeri, T. T., H. Brumer, 3rd, G. Daniel and P. Gatenholm (2007). "Biomimetic engineering of cellulose-based materials." Trends Biotechnol **25**(7): 299-306.

Venditto, I., A. S. Luis, M. Rydahl, J. Schuckel, V. O. Fernandes, S. Vidal-Melgosa, P. Bule, A. Goyal, V. M. Pires, C. G. Dourado, L. M. Ferreira, P. M. Coutinho, B. Henrissat, J. P. Knox, A. Basle, S. Najmudin, H. J. Gilbert, W. G. Willats and C. M. Fontes (2016). "Complexity of the Ruminococcus flavefaciens cellulosome reflects an expansion in glycan recognition." Proc Natl Acad Sci U S A **113**(26): 7136-7141.

Watt, D. K., D. J. Brasch, D. S. Larsen and L. D. Melton (1999). "Isolation, characterisation, and NMR study of xyloglucan from enzymatically depectinised and non-depectinised apple pomace." Carbohydrate Polymers **39**: 165-180.

Zhang, B., Y. Gao, L. Zhang and Y. Zhou (2021). "The plant cell wall: Biosynthesis, construction, and functions." J Integr Plant Biol **63**(1): 251-272.

Zhang, Y., J. Yu, X. Wang, D. M. Durachko, S. Zhang and D. J. Cosgrove (2021). "Molecular insights into the complex mechanics of plant epidermal cell walls." Science **372**(6543): 706-711.

Zykwinska, A., J.-F. Thibault and M.-C. Ralet (2008). "Competitive binding of pectin and xyloglucan with primary cell wall cellulose. ." Carbohydrate Polymers **74**: 957–961.

6. EXTENDING JANUS LECTINS ARCHITECTURE, AND APPLICATION FOR PROTOCELL LABELING

6.1 Summary

In order to extend the concept of Janus lectin, I design a novel architecture, that is still using the fucose-specific trimeric RSL scaffold but this time decorated with α-galactose specific β-trefoil lectin, instead of the NeuAc and GalA specific β-sandwich CBMs described above (Notova et al., 2022, Ribeiro et al., 2018).

The results are described below as a draft for a publication, however, still need to be improved. The project is a collaboration within the SynBIOcarb network with the team of Prof. Winfried Römer and his Ph.D student ESR4 Lina Suikstaite (GUVs experiments) and Dr. Ludovic Landemarre from the company GlycoDiag and his Ph.D. student ESR14 Federica Vena. My personal involvement in this project includes all the experiments described in the manuscript, except the gene design of RSL-MOA (designed and synthesized before I started the project) and GUVs experiments. The production of the isolated β-trefoil domain of MOA (MOAβT) was performed under my mentoring by Federica Vena during her SynBIOCarb secondment in CERMAV. She now continues with the project in the GlycoDiag company for implementing MOAβT in lectin arrays for the detection of Galili epitope on therapeutical glycoproteins. As part of the comparison between Janus lectins, I also aimed for obtaining the solution conformation of proteins by small-angle X-ray scattering (SAXS) but the results are not of sufficient quality for publication, and they have not been included.

The medified version was submitted to the preprint server BioRxiv, with the title *Extending Janus lectins architecture: characterization and application to protocells*, DOI: 10.1101/2022.08.15.503968 and the subsequently submitted to Computational and Structural Biotechnology Journal where the article is currently under the revision.

6.2 Scientific article III

Extending Janus lectins architecture, and application for protocell labeling

Simona Notova [a], Lina Suikstaite [b,c], Francesca Rosato [b,c], Federica Vena [d], Aymeric Audfray [e], Nicolai Bovin [f], Ludovic Landemarre [d], Winfried Römer [b,c,g], Anne Imberty [a*]

[a] Université Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France
[b] Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany
[c] Signalling Research Centers BIOSS and CIBSS, University of Freiburg, 79104 Freiburg, Germany
[d] GLYcoDiag, 2 rue du Cristal, 45100 Orléans
[e] Malvern Panalytical SAS, 30 rue du docteur Levy, 69200 Vénissieux
[f] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow 117997, Russian Federation
[g] Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, 79104 Freiburg, Germany

* To whom correspondence should be addressed. Anne Imberty (anne.imberty@cermav.cnrs.fr)

Abstract

Synthetic biology is a rapidly growing field with applications in biotechnology and biomedicine. Through various approaches, remarkable achievements, such as cell and tissue engineering, have been already accomplished. In synthetic glycobiology, the engineering of glycan binding proteins is being developed for producing tools with precise topology and specificity. Here we developed the concept of chimeric lectins, i.e., Janus lectin, with increased valency, and additional specificity. Novel engineered lectin assembled as a fusion between β-propeller from R*alstonia solanacearum* and β-trefoil domain from fungus *Marasmius oreades* is specific for fucose and α-galactose and the unique protein architecture allows to bind these ligands simultaneously. The protein activity was tested with glycosylated giant vesicles, resulting in proto-tissue-like structures through cross-linking of vesicles. The synthetic protein binds to HT1299 cancer cells by its two domains. Furthermore, the biophysical properties of this Janus lectin were compared with the two already existing Janus lectins. Denaturation profiles of the proteins indicate that the fold of each has a significant role in protein stability and should be considered during protein engineering.

Keywords: synthetic biology, protein engineering, carbohydrate-binding protein, Galili epitope, lectins

Introduction

Lectins, originally known as agglutinins because of their multivalency, selectively interact with glycoconjugates, a property that found many applications in the field of biotechnology and biomedicine as glycan-profiling tools (Sharon & Lis, 2004). Fine tuning the specificity or the valency of lectins is a promising domain to obtain novel tools. Thus, synthetic biology approaches, such as engineering of protein architecture, bring novel possibilities for lectin applications (Fettis et al., 2019; Irumagawa et al., 2022; Oh et al., 2022; Ramberg et al., 2021; Ross et al., 2019). Lectin engineering is a domain of the of synthetic glycobiology (Terada et al., 2017; Ward et al., 2021) and it can be performed at different levels, from glycan specificity to supramolecular architecture.

The trimeric oligomerization of *Ralstonia solanacearum* lectin (RSL) in a β-propeller shape (Fig. 1A) was used as a scaffold for building Janus lectins (Ribeiro et al., 2018, Notova et al., 2022) with two faces with different specificities. Two Janus lectins, RSL-CBM40 and RSL-$CBM77_{Rf}$, were obtained by fusion of RSL with two carbohydrate-binding modules (CBMs) with different specificity but rather similar shape, consisting of a single β-sandwich domain. This resulted in the creation of synthetic bispecific chimeras able to establish the interaction with fucose on one side and sialic acid (RSL-CBM40) or homogalacturonans (and RSL-$CBM77_{Rf)}$, respectively, on the other side.

Other protein domains with different specificity, but also topology diverging from the β-sandwich CBM could be considered for building novel Janus lectins from the trimeric RSL scaffold. β-trefoils are robust domains that contain internal tandem forming a three-lobed architecture with almost no secondary structures (Murzin et al., 1992). They display a large variety of functions, and many show carbohydrate-binding activity. They are considered as lectins due to their multivalence, but also as CBM (CBM13) since they sometimes occur as domains associated with glyco-active enzymes or toxins (Notova et al., 2022). While the shape is well conserved, the sequences do not show strong similarity, except for the presence of hydrophobic residues in the core and a QXW repeat present in most sequences (Hazes, 1996). A large number of crystal structures of β-trefoil lectins are available (212 from 63 different proteins) and based on their differences in sequences have been classified into 12 classes in the UniLectin3D database (Bonnardel, Mariethoz, et al., 2019). Datamining in genomes based on the lobe sequence signature of each class resulted in the prediction of thousands of putative β-trefoil lectins spanning all kingdoms of life (Notova et al., 2022). The threefold symmetry results in the presence of three carbohydrate-binding sites referred to as α, β, and γ although variations in amino acid sequences sometimes result in only one or two active binding sites.

Among the known structures of β-trefoil lectin, the agglutinin from the fairy ring mushroom *Marasmius oreades* (MOA) presents an interesting strict specificity for αGal containing oligosaccharides. The lectin rose attention already in the second half of the 20th century due to its high selectivity and affinity to blood group B oligosaccharide whereas almost no binding was detected with blood group A or H oligosaccharides (Estola & Elo, 1952; Winter et al., 2002). MOA is specific for oligosaccharides with terminal non-reducing α1-3 linked galactose, while galactosides linked in α1-2, α1-4, and α1-6 showed no binding (Winter et al., 2002).

Due to its preference for the αGal1-3Gal disaccharide, MOA binds efficiently to blood group B (Galα1-3 [Fucα1-2]Gal) but also to linear oligosaccharides such as Galα1-3Galβ1-4Glc on glycosphingolipid isoGb3 and Galα1-3Galβ1-4GlcNAc (Galili epitope) on glycoprotein (Macher & Galili, 2008). This latter epitope is present on the cell surface of most mammals, but not in humans and apes due to the deactivation of the α1-3galactosyltranferase during evolution (Galili et al., 1988). Humans possess preformed antibodies directed against Galα1-3Gal, which are responsible for hyperacute rejection of animal (mainly porcine) organs in the xenotransplantation (Cooper et al., 1994). They are also responsible for a strong immune response against biodrugs such as therapeutical antibodies if produced in rodent cell culture with active α1-3galactosyltranferase (Chung et al., 2008). Highly selective lectin is therefore needed for the identification of such epitopes (Kruger et al., 2002; Tateno & Goldstein, 2004; Winter et al., 2002).

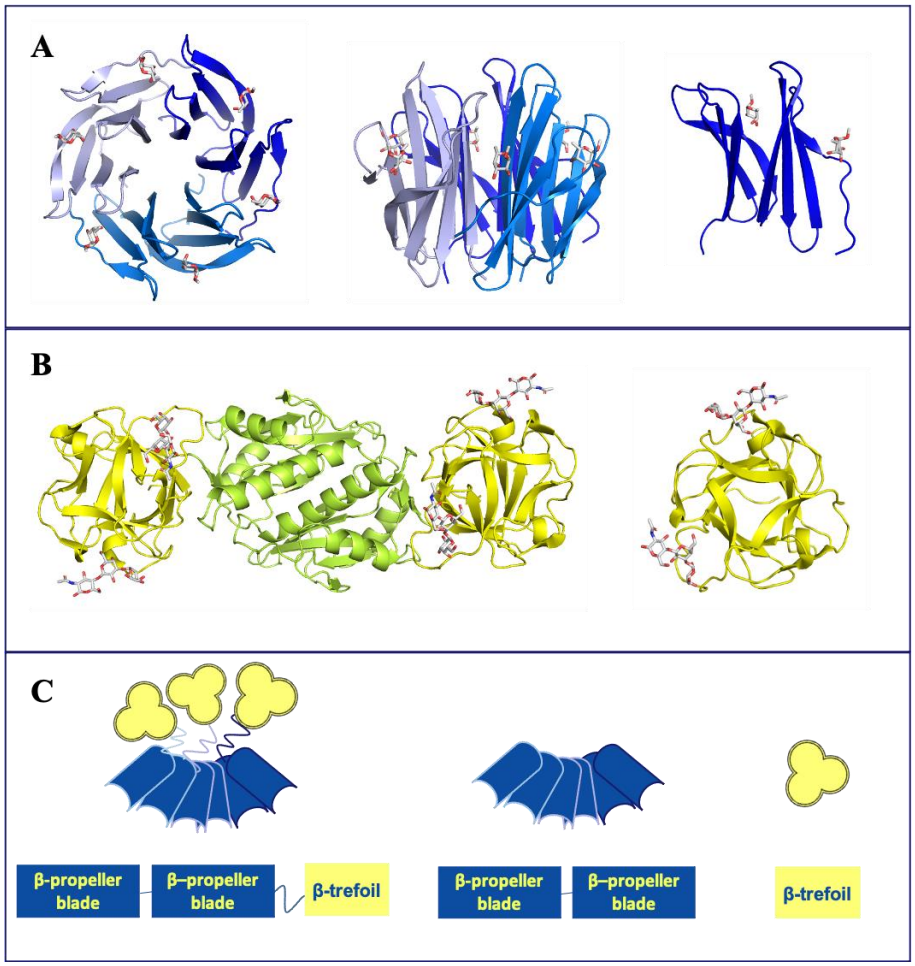


Figure 1: Schematic representations of Janus lectin RSL-MOA and crystal structures of its individual protein components. *A) Left and middle: Top and side view of the crystal structure of trimeric RSL (blue carton) in complex with α-MeFuc (sticks) with six fucose binding sites for fucose. Right: structure of RSL monomer (PDB code 2BT9). B) Left: Cartoon representation of the crystal structure of dimeric MOA, β-trefoil domain (yellow) in complex with Galα1-3Galβ1-4GlcNAc and dimerization domain (green). Right: β-trefoil domain of MOA with different orientation (PDB code 2IHO). C) schematic representation of Janus lectin RSL-MOA, RSL, MOAϐT and peptide domains.*

The crystal structure of MOA revealed that the lectin assembles as a homodimer with monomer composed of two distinct domains, adopting β-trefoil fold at N-terminus and α/β fold at C-terminus (Fig 1B). The C-terminal domain serves as a dimerization interface and has a proteolytic function (Cordara et al., 2011). The lectin β-trefoil domain has three-fold symmetry and the conserved motif $(Gln\text{-}X\text{-}Trp)_3$ is involved in the hydrophobic core of the structure. Co-crystallization studies of MOA with Galα1-3Galβ1-4GlcNAc revealed that each binding site has different ligand occupancy, emphasizing the fact that slight differences in amino acids might affect the binding (Fig 1B) (Grahn et al., 2007).

In the present study, we designed a Janus lectin as a fusion between monomeric RSL at N-terminus and the MOA β-trefoil domain at C-terminus (Fig 1C). We designed, produced, and characterized the Janus lectin RSL-MOA with double specificity toward fucosylated and α-galactosylated glycans. We also compared the biophysical behavior of this novel Janus-lectin with RSL-CBM40 (Ribeiro et al., 2018) and RSL-$CBM77_{Rf}$ (Notova et. al., 2022). Additionally, we have engineered the β-trefoil domain of MOA (MOAβT) and compared its activity with RSL-MOA.

Results

Design and production of β-trefoil domain of MOA (MOAβT)

MOA occurs naturally as a dimer, with a strong association between two C-terminal domains. Monomeric MOA composed only of the β-trefoil domain would be of interest for biotechnology applications. The gene of MOA β-trefoil domain was defined as the 156 N-terminal AAs of MOA full sequence and was amplified by PCR and subcloned into vector pET TEV as a fusion with N-terminal His-tag sequence. The resulting protein was named MOAβT and was recombinantly produced in soluble form in bacteria *E. coli* BL21(DE3) and purified by immobilized metal ion chromatography followed by His-tag cleavage by TEV (tobacco etch virus) protease. The resulting protein has an estimated molecular weight of 17.2 kDa (Fig. 2A). TEV cleavage followed by additional immobilized metal ion chromatography reduced the sample contaminants and the protein purity was verified by 12% SDS PAGE electrophoresis (Fig. 2A).

Design and production of Janus lectin RSL-MOA

Janus lectin RSL-MOA was designed as a gene fusion of monomeric RSL (N-terminus) and β-trefoil domain of lectin MOA (C-terminus). Connection of the two protein domains by the eight AAs linker PNGELLSS results in the protein sequence of 255 AA length. The gene sequence was optimized for bacterial expression and the synthesized gene was subcloned into plasmid pET25b+. The protein was produced in soluble form in bacteria *E. coli* KRX. The protein was subsequently purified by affinity chromatography on an agarose-mannose column due to the interaction between the RSL domain and mannose residues. Protein analysis by 12% SDS PAGE electrophoresis showed that Janus lectin RSL-MOA has an apparent size of 83 kDawhich corresponds to a trimer. RSL oligomerization appears to be resistant to denaturation

conditions Smaller amount of dimer and monomer are also visible, probably because of partial denaturation (Fig. 2B).

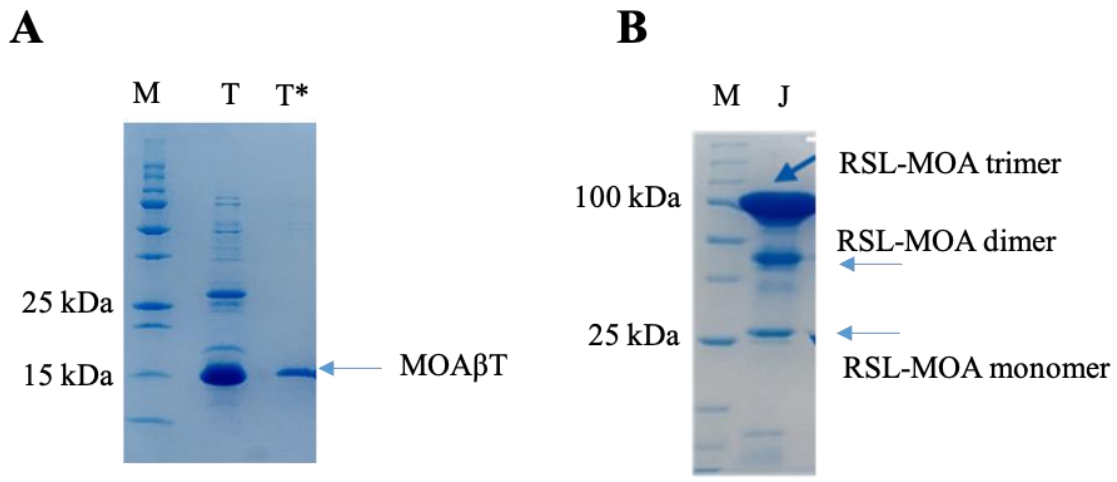


Figure 2: SDS PAGE analysis of MOAβT and RSL-MOA. *The protein samples were analyzed under denaturing conditions on 12% polyacrylamide gel. M-protein marker, J-Janus lectin, T- MOAβT, T*-MOAβT cleaved by TEV protease. A) MOAβT with the estimated size of 17.2 kDa appears as monomeric and the impurities were eliminated during TEV cleavage and subsequent purification of protein. B) Janus lectin RSL-MOA is resistant to denaturation conditions and on the gel appears as a monomer (28 kDa), dimer (56 kDa), and trimer (83 kDa) whereas the trimer is the most abundant.*

Biophysical characterization of MOAβT and RSL-MOA by isothermal titration calorimetry (ITC)

MOAβT affinity towards Galα1-3Gal disaccharide, the terminal disaccharide of Galili epitope, was assayed by titration microcalorimetry. Large exothermic peaks were obtained at the beginning of titration. The affinity was not strong enough to obtain a sigmoidal curve, and therefore, as recommended in such cases (Turnbull & Daranas, 2003), the stoichiometry (N) was fixed, using a value of N = 2 since as previously shown by Grahn et al., (2007) MOA binding sites has different occupancy and we expected at least two of them to be active. A dissociation constant (Kd) of 150 μM was obtained, which confirms the functionality of isolated β-trefoil domain, and is in good agreement with Kd measured for the whole MOA protein, i.e., 182 μM (Winter et al., 2002). (Fig 3A).

RSL-MOA was designed to possess six binding sites for fucose and up to nine binding sites for α-galactose on opposites faces. The functionality of both binding interfaces was tested by various biophysical approaches. In order to evaluate if all sites are functional, we designed an ITC experiment with a consecutive injection of both ligands. First, αGal1-3Gal was titrated into the cell containing RSL-MOA. Subsequently, the cell content from the first experiment (complex RSL-MOA/Galα1-3Gal) was titrated by α-MeFuc (Fig 3B).

The thermogram obtained by titrating Galα1-3Gal in RSL-MOA solution was very similar to the one obtained for MOAβT. The dissociation constant (Kd = 226 μM) was comparable with

the Kd for the isolated β-trefoil domain (above) and for the whole MOA (Winter et al., 2002). The following titration by α-MeFuc resulted in a sigmoid shape, due to stronger affinity, with a measured stoichiometry of N= 2, corresponding to the presence of two fucose binding sites per RSL monomer. The Kd was measured to 0.4 μM, in excellent agreement with previously measured affinity (0.7 μM) for RSL (Kostlánová et al., 2005). This experiment proved that both parts of RSL-MOA are functional and able to bind their ligands at the same time.

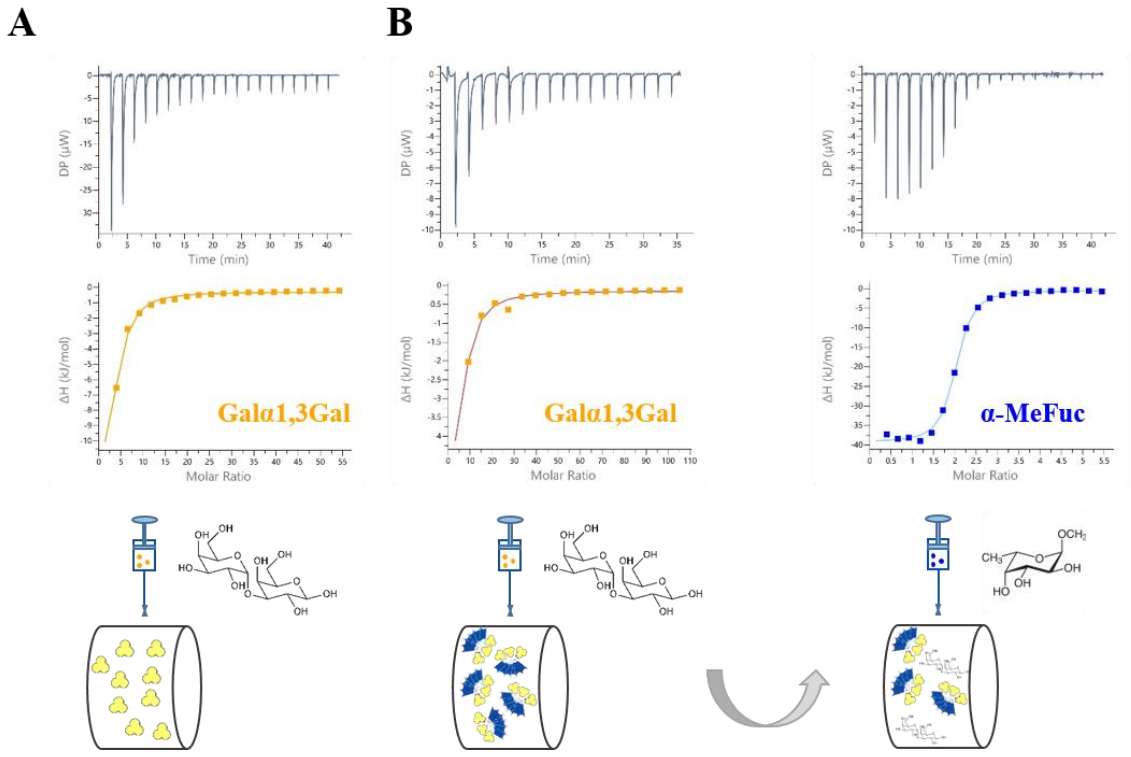


Figure 3: ITC thermographs of MOAβT and Janus lectin RSL-MOA. *A) MOAβT binds Galα1-3Gal with a similar affinity as parental protein MOA. B) RSL-MOA can simultaneously bind two binding partners, i.e., Galα1-3Gal and α-MeFuc confirming the activity of both functional domains.*

Biophysical characterization of RSL-MOA by surface plasmon resonance (SPR)

To evaluate the effect of multivalency and therefore measure avidity instead of affinity, surface plasmon resonance was used. Chips CM5 were primarily coated with streptavidin. Subsequently, different ligands, i.e., biotinylated PAA-α-fucose, PAA-α-galactose, and PAA-β-galactose, were immobilized (see methods). A high-density α-galactose chip was prepared by immobilization of 200 μg/mL biotinylated PAA-α-Gal and used to evaluate the binding of the MOA part of Janus lectin RSL-MOA. However, no binding of RSL-MOA was observed, probably because of its very low affinity for the monosaccharide α-galactose (K_D = 8 mM) (Winter et al. 2002).

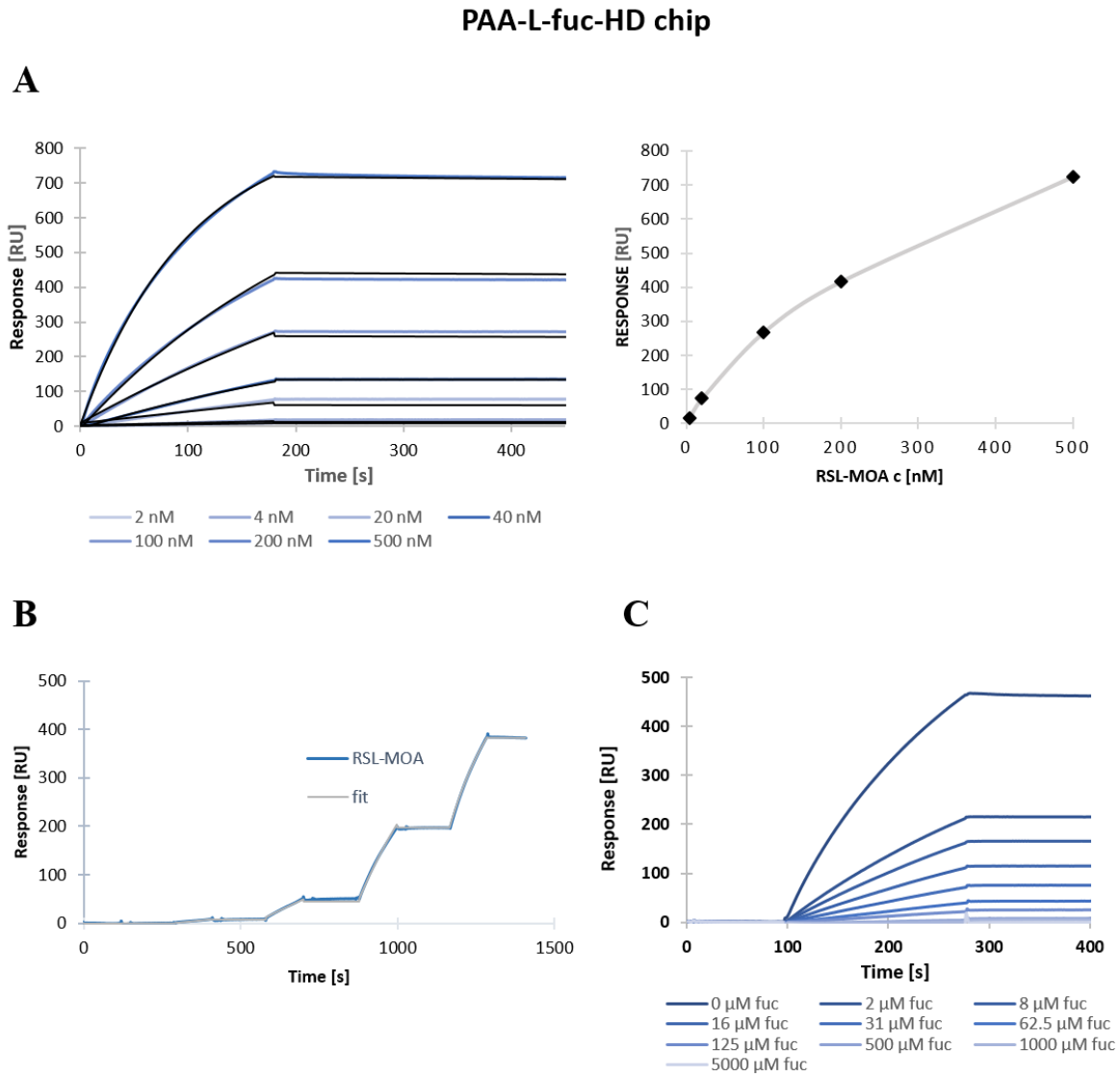


Figure 4: SPR sensorgrams of Janus lectin RSL-MOA on CM5-PAA-L-fuc HD chip. *A) Top: Titration experiment of different concentrations of RSL-MOA (blue) and fitting curves (1:1 binding fit) (black) (left) and steady state analysis of RSL-MOA titration (right). B) Single cycle kinetics of RSL-MOA with concentrations 10, 50, 250, and 500 nM. C) Inhibition of 200 nM RSL-MOA by various concentration of L- fucose.*

In order to evaluate the binding of RSL part, we have prepared two different SPR chips. First, a high-density (HD) chip was prepared by immobilization of 200 µg/mL PAA-L-fucose. Subsequently, RSL-MOA in the concertation range from 2 to 500 nm was injected on the chip. The regeneration step was performed between each injection using 1 M L-fucose solution. As shown in Figure 4A, a dose-dependent response was obtained, however, RSL-MOA bound too strongly to HD fucose chip with a steep association phase and no dissociation event. The steady state analysis showed that even at the highest protein concertation (500 nM) the chip surface was not saturated by RSL-MOA, i.e., the plateau phase was not reached (Fig. 4A). Nonetheless, the fitting of kinetics constant could be performed and values for affinity (K_D) and kinetics (k_{on} and k_{off}) were estimated (Table 1). As an alternative, a single-cycle kinetic experiment was performed. RSL-MOA was injected at four concentrations (10 to 500 nM) (Fig. 4B). The

advantage of this method is that no regeneration step is needed. The estimated K_D was one order lower if compared to titration experiment, however such variation is acceptable in binding events characterized by avidity. In order to evaluate the capacity of monosaccharides to inhibit the multivalent binding, 200 Nm RSL-MOA was pre-incubated with various concentrations of L-fucose (2 to 5000 μM). in order to compete with its binding chip surface. As shown in Figure 4C, the complete inhibition was achieved in the presence of high concentrations of fucose (1000 and 5000 μM), and IC_{50} was estimated to be 2.2 μM.

Table 1: SPR statistics from titration and single cycle kinetics experiments. *The experiments were carried on HD of LD PAA-L-Fuc CM5 chips with the flow 30 μL/min and 10 μL/min, respectively.*

Experimental condition	k_{on} (1/Ms)	k_{off} (1/s)	K_D (M)	R_{MAX} (RU)	Chi^2
Chip CM5 PAA-L-fuc HD Classical kinetics	$1.99*10^4$	$3.44*10^{-5}$	$1.7*10^{-9}$	865	104
Chip CM5 PAA-L-fuc HD Single cycle kinetics	$1.14*10^4$	$6.26*10^{-6}$	$5.5*10^{-10}$	573	4.42
Chip CM5 PAA-L-fuc LD Classical kinetics	$4.42*10^2$	$2.08*10^{-7}$	$4.9*10^{-10}$	3030	8.23
Chip CM5 PAA-L-fuc LD Single cycle kinetics	$1.84*10^4$	$1.09*10^{-7}$	$5.9*10^{-12}$	1960	40.4

In general, HD chips are not suitable for lectins the undesirable effect of mass transfer (de Mol & Fischer, 2010). Therefore we prepared a low density (LD) CM5 chip which was functionalized from a mixture (200 μg/mL) of PAA-L-fucose/PAA-β-galactose in the ratio 1:9. RSL-MOA was injected in increasing concentrations (5 to 1000 nM). The regeneration step was performed between each injection using 1 M L-fucose solution. Again, a dose-dependent response was obtained (Fig 5A), but still with a very slow dissociation event. Even though, the curvature of steady state analysis is closer to plateau phase when compared to HD chip. The saturation of chip surface was not reached even with highest concertation of protein (1000 nM). The fitting is less favorable than in the case of HD chip, nevertheless, the affinity and kinetics of binding were estimated (Table 1) obtaining values of $K_D = 1.7*10^{-9}$ M. K_D correlates with previously observed results, however with some differences with k_{on} and k_{off}, that could be attributed to the difficulty to evaluated precisely k_{off} with a very slow dissociation event. A single-cycle kinetic experiments were performed, too (Fig. 5B). The kinetics of binding is comparable with HD chip, and result in a stronger avidity ($5.9*10^{-12}$) (Table 1). Additionally, RSL-MOA was pre-incubated with various concentrations of L-fucose (0, 5, 10, 25, 50, 100, 250, 500, 1000, 2500, 5000 μM) in order to compete with its binding on LD chip. As shown in Figure 5C, the complete inhibition was achieved in the presence of high concentrations of fucose (2500 and 5000 μM), and IC_{50} was estimated to be 7.8 μM.

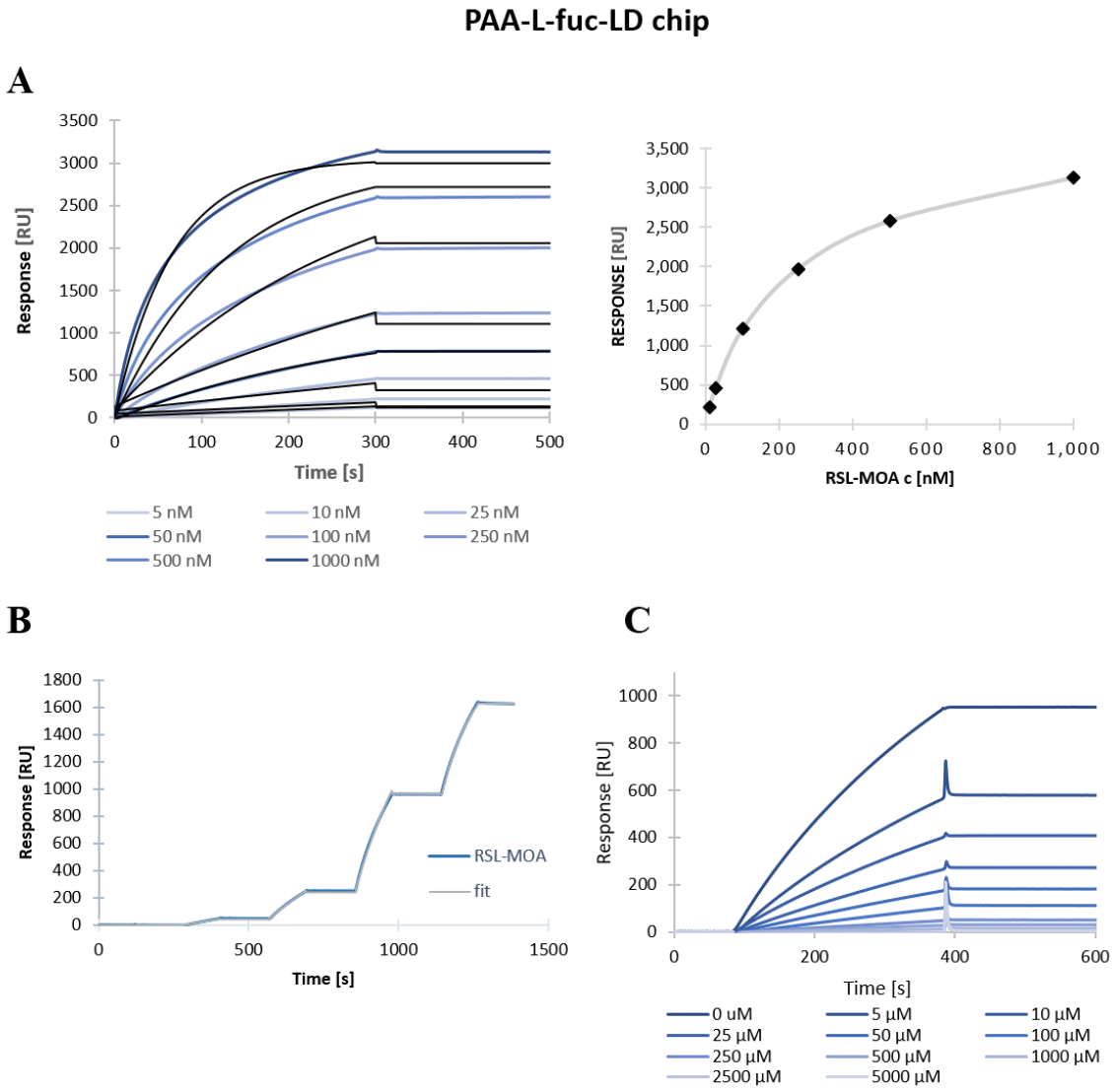


Figure 5: SPR sensorgrams of Janus lectin RSL-MOA on CM5-PAA-L-fuc LD chip. *A) Top: Titration experiment of different concentrations of RSL-MOA (blue) and fitting curves (1:1 binding fit) (black) (left) and steady state analysis of RSL-MOA titration (right). B) Single cycle kinetics of RSL-MOA with concentrations 10, 50, 250, and 500 nM. C) Inhibition of 50 nM RSL-MOA by various concentration of L- fucose.*

The ability of RSL-MOA to bind glycan-decorated surface of SPR chips was confirmed by several experimental setups. The affinity constant varies from nanomolar to picomolar, illustrating the difficulty to quantify binding when avidity is the major event. Surprisingly, the LD chip gave less reliable data than the HD one, which could be due to rather high immobilization rate (higher RU response with the lectin on the LD), and to the slower flow used in the experiment.

These results are however in agreement with previous ones evaluating RSL avidity for fucose chip (Arnaud et al., 2013). This avidity constant is at least 1000-fold higher than the affinity measure for fucose monosaccharide in solution, confirming the very strong cluster effect

resulting from the topology of the RSL β-propeller with the presentation of six binding sites on the same face, binding very efficiently to fucose presented in a multivalent manner on a surface.

MOAβT and RSL-MOA induce crosslinking of glycosylated giant unilamellar vesicles (GUVs)

The binding properties of MOAβT and RSL-MOA were tested against glycolipid-containing giant unilamellar vesicles (GUVs) in order to test their ability to bind to membranes. Fluorescently labeled GUVs (Atto647N-DOPE) were decorated with synthetic glycolipid analogs attached to functional spacer lipid (FSL) molecules, such as FSL-isoglobotriose, FSL-A, and FSL-B carrying Galα1-3Galβ1-4Glc, blood group A and blood group B oligosaccharides, respectively. These GUVs were subsequently incubated with MOAβT and RSL-MOA fluorescently labeled with Atto488.

Surprisingly, MOAβT-Atto488 (500 nM) (green) showed almost no interaction with FSL-iGb3-GUVs (red) (Fig. 6A). On the contrary, MOAβT-Atto488 (500 nM) (green) binds FSL-B-GUVs (red) and even crosslinks the vesicles (Fig. 6B). The lectin clearly concentrates at the interface created between two GUVs resulting in a flattening of the vesicles. The β-trefoil topology allows the binding of blood group B oligosaccharides on two different GUVs bringing them together.

RSL-MOA-Atto488 (green) binds efficiently to the surface of FSL-iGb3-GUVs (red), even at a concentration of 200 nM (Fig. 6C). The difference is striking when compared to the absence of binding of MOAβT-Atto488 (500 nM). The super-multivalency results from the presence of three β-trefoil domains, so in total nine possible αGal binding sites make a very strong difference in terms of ability to bind to the glycosylated surface. In this case, also, cross-linking of vesicles is observed.

The capacity of RSL-MOA to crosslink GUVs carrying different oligosaccharides was tested using two fluorescently labeled vesicles, FSL-iGb3-GUVs (Atto647N-DOPE red) and FSL-A-GUVs (Atto488-DOPE green). When the unlabeled RSL-MOA is added to the GUVs solution, cross-linking is observed between red and green GUVs, with minimal occurrence of cross-linking between GUVs of the same color. Herein, due to the double specificity of lectin RSL-MOA, we confirmed that multivalent and two-site oriented topology has clear potential in vesicles cross-linking (Fig 6D).

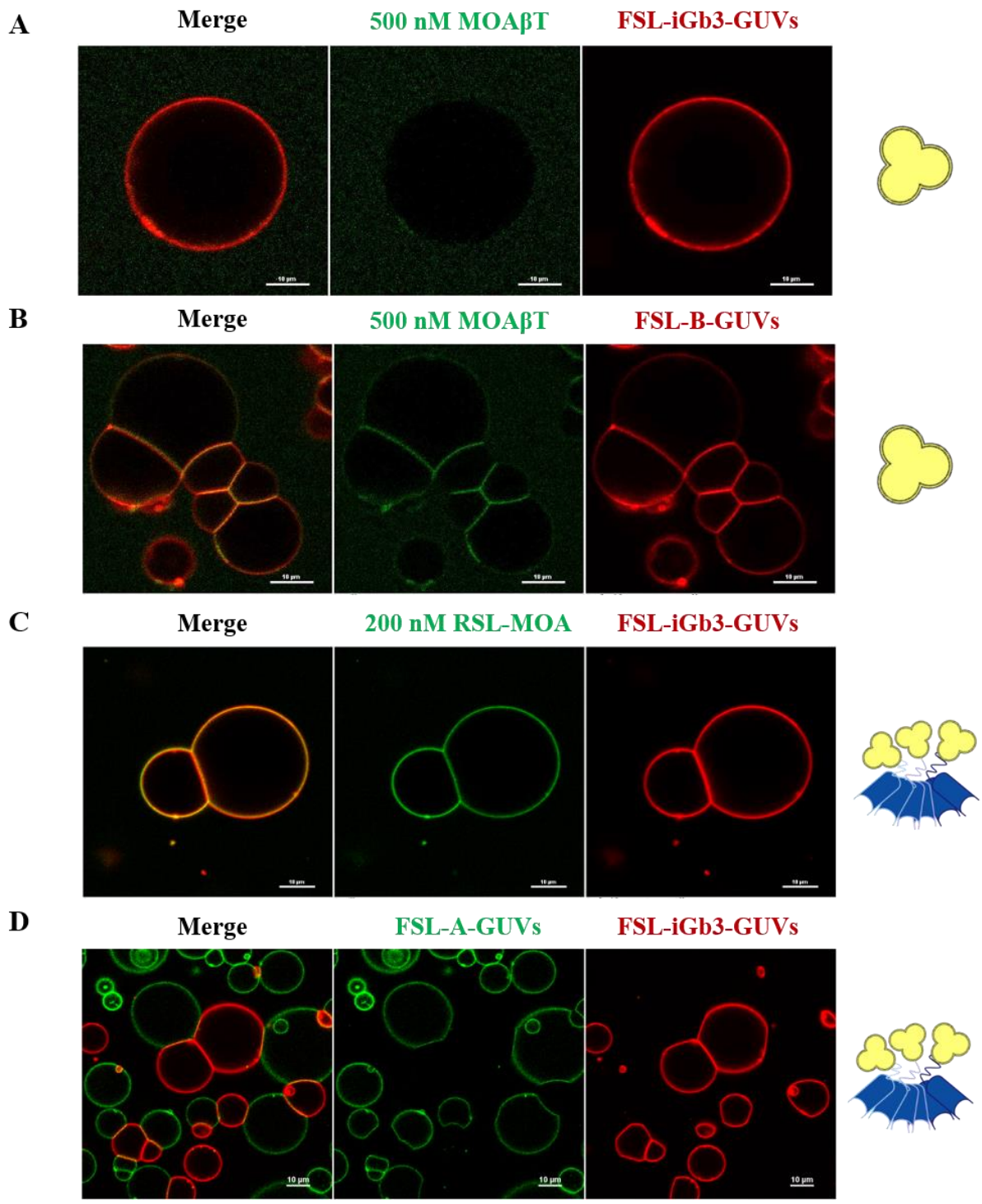


Figure 6: The binding properties of MOAβT and RSL-MOA with glycodecorated GUVs. *A) 500 nM MOAβT-Atto488 (green) shows almost no binding to the FSL-iGb3-GUVs (red, fluorescent lipid Atto647N). B) 500 nM MOAβT-Atto488 (green) binds and crosslinks FSL-B-GUVs red, fluorescent lipid Atto647N). C) 200 nM RSL-MOA-Atto488 (green) binds and crosslinks FSL-iGb3-GUVs (red, fluorescent lipid Atto647N). D) Unlabeled 200 nM RSL-MOA crosslinks two different GUVs, FSL-iGb3-GUVs (red, fluorescent lipid Atto647N) and FSL-A-GUVs (green, fluorescent lipid Atto488). The GUVs were composed of DOPC, cholesterol, glycolipid of choice, and membrane dye to the molar ratio of 64.7:30:5:0.3, respectively. Scale bars are 10 µm.*

RSL-MOA binds to human epithelial cells

The ability of RSL-MOA to recognize and bind to glycans on the surface of human cells was investigated by flow cytometry. RSL-MOA was fluorescently labeled with a Cy5 dye (RSL-MOA-Cy5) and incubated with the non-small cell lung cancer cell line H1299, for 30 minutes at 4°C. As depicted in the histograms of fluorescence intensity (Fig. 7A), cells treated with increasing concentrations of RSL-MOA-Cy5 (0.07 – 0.7 µM) showed binding of the lectin to the cell surface in a dose-dependent manner. In order to confirm lectin specificity and glycan-driven binding to receptors at the plasma membrane of H1299, a series of inhibition assays were performed. RSL-MOA-Cy5 was pre-incubated with 100 mM L-fucose or 100 mM synthetic analogue of α-Gal epitope, p-nitrophenyl-α-D-galactopyranoside (PNPG), respectively. The high concentrations of ligands were chosen due to previous observations with lectin RSL, for which a complete inhibition of binding to fucosylated receptors on the surface of H1299 cells was achieved only in the presence of 100 mM L-fucose.

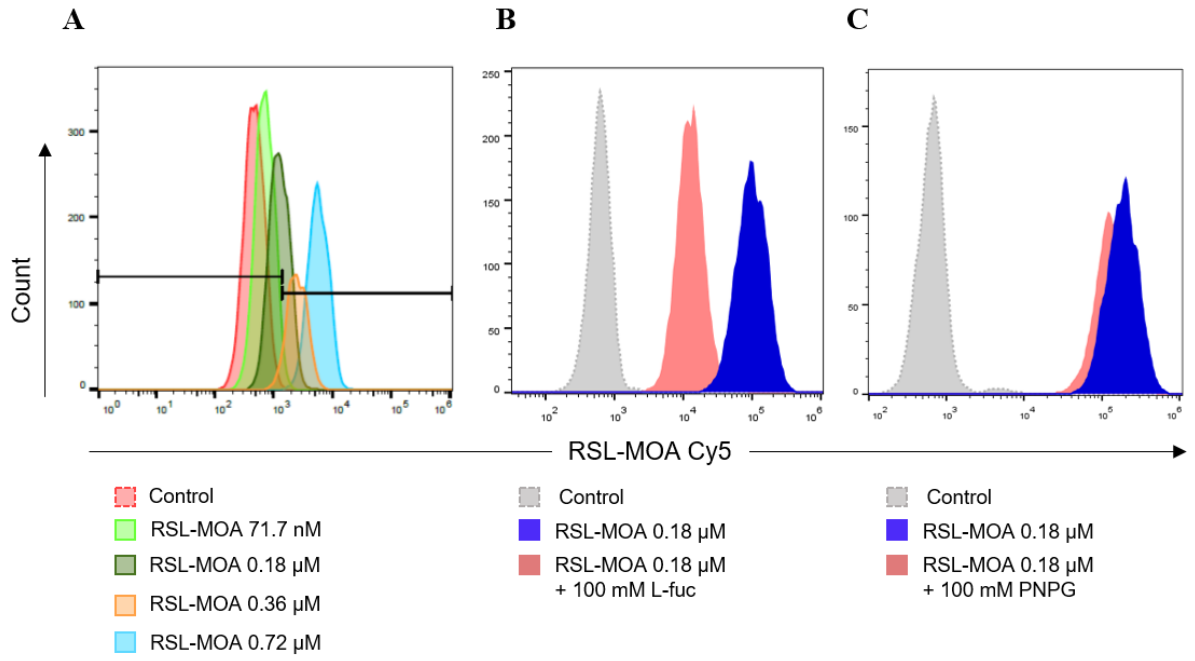


Figure 7: RSL-MOA shows dose-dependent binding to H1299 cells. *A) Histogram plot representing gated living H1299 cells pre-incubated with fluorescently labeled RSL-MOA-Cy5. Histograms of fluorescence intensity show a dose-dependent trend in binding according to used lectin concentrations. B) Histograms of fluorescence intensity of H1299 treated with RSL-MOA-Cy5 pre-incubated with 100 mM L-fucose. Inhibition of RSL domain with high concentrations of L-fucose determine a shift in fluorescence intensity towards lower values (pink histogram), suggesting a reduced binding of lectin to H1299 cells. C) Histograms of fluorescence intensity of H1299 treated with RSL-MOA-Cy5 pre-incubated with 100 mM PNPG shows almost no reduction in lectin binding to cell surface. The control experiments correspond to samples of H1299 cells without any addition of lectin sample.*

The presence of 100 mM L-fucose significantly decrease RSL-MOA-Cy5 binding to treated cells (Fig. 7B) but does not abolish them. Increasing the concentration up to 500 mM of fucose does not result in additional binding (data not show). The interaction of the engineered lectin

with cell surface is therefore not completely inhibited upon saturation of the fucose-binding sites, suggesting that the MOA domain partially compensate the binding to H1299 cells. On the other hand, 100 mM PNPG has almost no effect on RSL-MOA-Cy5 binding to treated cells, probably due to the low affinity of MOA toward monosaccharides (Fig. 7C) (Winter et al., 2002). These result suggest that both RSL and MOA domains of Janus lectin are able to bind to glycans present on H1299 cancer cells.

Comparison of thermal stability of Janus lectins and their distinct domains

Janus lectins are created by assembling protein domains with different biophysical properties. It is therefore of interest to evaluate the effect of the fusion of these modules, and also to compare the biophysical properties of Janus lectins.

The thermal stability of RSL-MOA and its constituting domains was measured with differential scanning calorimetry (DSC), using equipment provided by the company Malvern Panalytical. Thermograms of MOAβT and RSL are displayed in Figure 8. These two modules have very different thermal stability with a denaturation/midpoint temperature (Tm) of 42°C for the MOAβT and 92°C (with two Tm at 90.5°C and 93.8°C) for the β-propeller RSL. It could be that MOAβT is easily denatured because it is a domain separated from the whole original protein. The very strong thermal resistance of RSL is related to the compact and robust architecture of the β-propeller and the two slightly different Tms extracted from the curve may correspond to the dissociation of the trimer followed by the unfolding of each monomer. As expected, adding the ligand to the protein solution increased the Tm of RSL to approximately 100°C and 103°C, for sugar concentrations of 125 and 250 µM, respectively.

The thermal unfolding of RSL-MOA and other Janus lectins was carried out in the presence of a high concentration of L-fucose since the protein was tested after the purification procedure. RSL-MOA displayed two events of denaturation at very different temperatures, i.e., 40°C and 115°C. From the results obtained on the separated domain, it is clear that the lower Tm corresponds to the unfolding of the MOA β-trefoil and the second event, with higher Tm, corresponds to the denaturation of RSL. The Tm of the MOA moiety is not different when isolated or linked to RSL. The Tm of the RSL moiety is higher than for the β-propeller alone, which is probably due to the high concentration of fucose in the medium.

Other Janus lectins were also assayed by DSC for comparison. The proteins, RSL-CBM40, and RSL-CBM77$_{Rf}$ were obtained as described previously (Notova et al., 2022, Ribeiro et al., 2018) and the buffer was supplemented by L-fucose as in the case of RSL-MOA. As already experienced with RSL-MOA, two separate events of denaturation were observed (Fig. 8) (Table 2). The high-temperature Tm, corresponding to the denaturation of RSL, is observed, with a value as high as 121°C for RSL-CBM40. CBM77$_{Rf}$ and CBM40 have a similar β-sandwich fold, and they present rather similar melting temperatures of 67°C and 65°C, respectively. This value is more than 25 °C higher than the one observed for MOA, indicating a clear difference in thermal stability between β-sandwich and β-trefoil.

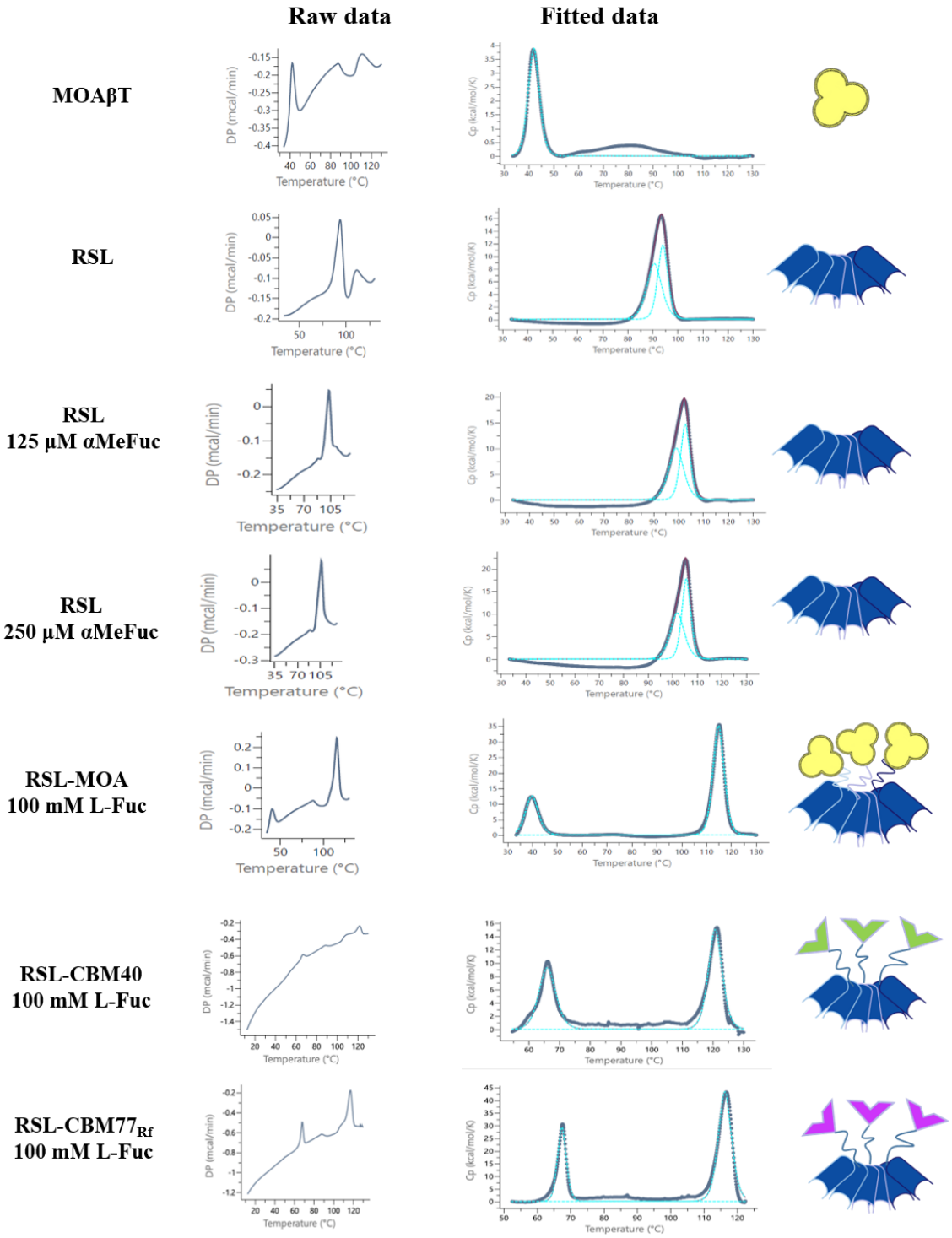


Figure 8: Thermal stability analysis of MOAβT, RSL, and Janus lectins RSL-MOA, RSL-CBM40, and RSL-CBM77$_{Rf}$ with or without ligands. *The figure displays experimental conditions, raw data, fitted data and schematic representation of proteins.*

Table 2: Thermal signatures of denaturation profiles of MOAβT, RSL, and Janus lectins. Scan rate was fixed at 200 °C/min.

Protein	Ligand	Protein (g/L & μM)	T_m1 (°C)	T_m2 (°C)
MOAβT	-	0.5 (29)	41.8	-
RSL	-	0.25 (25)	90.5	93.8
	125 μM αMeFuc	0.25 (25)	98.9	102.5
	250 μM αMeFuc	0.25 (25)	101.6	105.3
RSL-MOA	100 mM L-fuc	0.25 (9)	39.5	114.5
RSL-CBM77$_{Rf}$	100 mM L-fuc	0.5 (18)	67.4	116.6
RSL-CBM40	100 mM L-fuc	0.5 (16)	65.5	121.1

Discussion and Conclusion

In this work, we extended the concept of Janus lectin and therefore developed a universal strategy for increasing lectin valency and introducing an additional specificity (Notova et al., 2022, Ribeiro et al., 2018). Janus lectins are engineered as fusion chimeras and due to the presence of lectin RSL, these synthetic proteins assemble as trimers, resulting in the multiplication of lectin binding sites and thus expected higher affinity toward ligands. The first two Janus lectin, RSL-CBM40 (Ribeiro et al., 2018; Siukstaite et al., 2021) and RSL-CBM77$_{Rf}$ (Notova et al., 2022) were designed as fusion chimeras of lectin and CBM domain with a possible application as drug carriers or protein glue in plant cell wall engineering. However, here we proposed an alternative where Janus lectin RSL-MOA is composed of two individual lectin domains, β-trefoil from lectin MOA and β-propeller of RSL.

The biophysical properties of RSL-MOA were compared with MOAβT, the engineered β-trefoil domain of MOA lectin. Both proteins showed the same affinity toward ligands in solution (ITC) confirming their activity. However, their behavior was different when tested with glycodecorated liposomes (GUVs experiments). We observed that MOAβT showed almost no binding to FSL-iGb3-GUVs while RSL-MOA interacts with such vesicles and even cross-link them. This is probably conditioned by the super-multivalency of engineered RSL-MOA that presents nine binding sites, instead of only three in MOAβT. On the other hand, MOAβT displayed active binding with blood group B GUVs (FSL-B-GUVs) confirming that the lectin domain is able to bind glycoconjugates on the membrane surface. RSL-MOA also has the ability to bind two different GUVs and cross-link them. This confirms that engineered topology does not affect protein binding and such a lectin is of interest for highly selective tool development and labeling of proto-tissues.

The SPR analysis of RSL-MOA revealed the strong dose-dependent binding of β-propeller RSL toward high and low density PAA-L-Fuc CM5 chips. Even though almost no dissociation

event was observed, the fitting procedure made us enough confident to estimate binding constants. The variations between them are probably related to the fact that the binding is considered as avidity, not simple affinity, and therefore the characterization of such event is generally very complex. Nonetheless, RSL-MOA showed at least 1000 higher affinity that in solution supporting the fact that addition of MOA does not affect it binding properties, neither avidity.

The ability of RSL-MOA to recognize glycans exposed at the surface of H1299 cells was analyzed by flow cytometry. RSL-MOA showed a strong binding in dose-dependent manner. H1299 cells are characterized by the presence of highly fucosylated glycoconjugates, among the others, on their surface (Jia et al., 2018). Furthermore, we intended to evaluate if saturation of RSL-MOA sugar-binding sites with 100 mM L-fucose or 100 mM PNPG, diminish the interaction between lectin and cells. The addition of 100 mM L-fucose to the lectin, indeed, reduced the binding of RSL-MOA to H1299 cells, nonetheless, it is not completely prevented. We believe that the presence of MOA domain partially compensates the interaction between the lectin and the cell surface. Our hypothesis is supported by previous observations with lectin RSL, which also showed similar binding properties to H1299 cells, however, this interaction could be completely inhibited by 100 mM L-fucose. On the other hand, 100 mM PNPG does not affect RSL-MOA binding suggesting that synthetic analogue of α-galactose does not bind MOA with sufficient affinity which was already observed previously with α-galactose monosaccharide (Winter et al., 2002) and confirmed by our SPR observations as well.

The thermal stability of three Janus lectins, i.e., RSL-MOA, RSL-CBM40, and RSL-CBM77$_{Rf}$, was compared by DSC. During protein denaturation, two events of unfolding are observed for all Janus lectins implying the fact that each protein domain has different stability. Structurally similar CBMs (β-sandwich fold) share almost the same Tm, however β-trefoil is much more thermally unstable. On the other hand, β-propeller RSL showed extremely high Tm, whereas the addition of the ligand has a tremendous effect on its stability. Our findings, therefore, suggest that the structural fold of the individual domains should be taken into account when designing novel Janus lectins if stability is to be considered.

Until now, three Janus lectin with different specificities and architecture have been engineered. Therefore, we are confident that with this strategy, the novel bispecific lectins with improved valency can be obtained and found their applications in numerous fields of biotechnology and biomedicine.

Methods

Gene design and cloning

MOAβT

Gene for β-trefoil domain of MOA was obtained by polymerase chain reaction where plasmid pET-25b+RSL-MOA was used as a template. The primers, TTACATATGAGCTTACGTCG TGGC (Forward) and ATTACTCGAGTTACATGCGGTTGAAGTACC (Reverse) were designed to align to the full sequence of β-trefoil domain of MOA and were ordered from Eurofins Genomics (Ebersberg, Germany). The restriction enzyme sites of NdeI and XhoI were added at 5' and 3' ends, respectively. Subsequently, gene *moaβt* and plasmid pET-TEV vector (Houben et al., 2007) were digested by RE NdeI and XhoI and ligated resulting in pET-TEV-MOAβT vector. After transformation by heat shock in *E. coli* DH5α strain, a colony screening was performed, and the positive plasmids were amplified and controlled by sequencing.

RSL-MOA

The original amino acid sequence of the lectin MOA from *Marasmius oreades* was obtained from the PDB database. The gene *rsl-moa* was designed as a fusion chimera with RSL at N-terminus and β-trefoil domain of MOA at C-terminus *via* linker PNGELLSS. The gene was ordered from Eurofins Genomics (Ebersberg, Germany) after codon optimization for the expression in the bacteria *Escherichia coli*. The restriction enzyme sites of NdeI and XhoI were added at 5' and 3' ends, respectively. The synthesized gene was delivered in plasmid pEX-A2-RSL-MOA. Subsequently, plasmid pEX-A2-RSL-MOA and the pET-25b+ were digested by NdeI and XhoI restriction enzymes to ligate *rsl-moa* in pET-25b+. After transformation by heat shock in *E. coli* DH5α strain, a colony screening was performed, and the positive plasmids were amplified and controlled by sequencing.

Protein expression

E. coli BL21(DE3) cells were transformed by heat shock with pET-TEV-MOAβT plasmid prior preculture in Luria Broth (LB) media with 25 μg/mL kanamycin at 37°C under agitation at 180 rpm overnight. The next day, 10 mL of preculture was used to inoculate 1 L LB medium with 25 μg/mL kanamycin at 37°C and agitation at 180 rpm. When the culture reached OD_{600nm} of 0.6 - 0.8, the protein expression was induced by adding 0.1 mM isopropyl β-D-thiogalactoside (IPTG), and the cells were cultured at 16°C for 20 hours.

E. coli KRX (Promega) cells were transformed by heat shock with pET-25b+-RSL-MOA plasmid and pre-cultured in LB media substituted with 50 μg/mL ampicillin at 37°C under agitation at 180 rpm overnight. The following day, 10 mL of preculture was used to inoculate 1 L LB medium with 50 μg/mL ampicillin at 37°C and agitation at 180 rpm. When reached OD_{600nm} of 0.6 - 0.8, the protein expression was induced by adding 1% L-rhamnose, and the cells were cultured at 16°C for 20 hours.

The cells were harvested by centrifugation at 14000 × *g* for 20 min at 4°C and the cell paste was resuspended in 20 mM Tris/HCl pH 7.5, 100 mM NaCl (Buffer A), and lysed by a pressure cell disruptor (Constant Cell Disruption System) with a pressure of 1.9 kBar. The lysate was

centrifuged at 24 000 × *g* for 30 min at 4°C and filtered on a 0.45 µm syringe filter prior to loading on an affinity column.

Protein purification

MOAβT

The cell lysate was loaded on 1 mL HisTrap column (Cytiva) pre-equilibrated with Buffer A. The column was washed with Buffer A to remove all contaminants and unbound proteins. The MOA was eluted by Buffer A in steps during which the concentration of imidazole was increased from 25 mM to 500 mM. The fractions were analyzed by 12% SDS PAGE and those containing MOAβT were collected and deprived of imidazole by dialysis in Buffer A. N-terminal His-tag was removed by TEV cleavage with the ratio 1:50 mg of TEV:protein in the presence of 0.5 mM EDTA and 1mM TCEP over night at 19°C. After, the protein mixture was repurified on 1 mL HisTrap column (Cytiva) and the pure protein was concentrated by Pall centrifugal device with MWCO 3 kDa and stored at 4°C.

RSL-MOA

The cell lysate was loaded on 10 mL D-mannose-agarose resin (Merck) pre-equilibrated with Buffer A. The column was washed with Buffer A to remove all contaminants and unbound proteins and the flow-through was collected. RSL-MOA was eluted by Buffer A with the addition of 100 mM D-mannose or 100 mM L-fucose in one step. Due to not sufficient binding capacity of the column, the flow-through was reloaded on the column several times and the protein was eluted as described previously. The fractions were analyzed by 12% SDS PAGE and those containing RSL-MOA were collected and dialyzed against Buffer A. The protein was concentrated by Pall centrifugal device with MWCO 30 kDa and the pure protein fractions were pooled, concentrated, and stored at 4°C.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed with MicroCaliTC200 (Malvern Panalytical). Experiments were carried out at 25°C ± 0.1°C. Proteins and ligands samples were prepared in Buffer A. The ITC cell contained proteins in a concentration range from 0.05 mM to 0.2 mM. The syringe contained the ligand solutions in a concentration from 50 µM to 10 mM. 2 µL of ligands solutions were injected into the sample cell at intervals of 120 s while stirring at 750 rpm. Integrated heat effects were analyzed by nonlinear regression using one site binding model (MicroCal PEAQ-ITC Analysis software). The experimental data were fitted to a theoretical curve, which gave the dissociation constant (Kd) and the enthalpy of binding (ΔH).

Surface plasmon resonance (SPR)

The SPR experiments were performed using a Biacore X100 biosensor instrument (GE Healthcare) at 25°C. Biotinylated polyacrylamide-attached (PAA) sugars, such as L-fucose, α-galactose, and β-galactose (Lectinity) were immobilized on CM5 chips (GE Healthcare) pre-coated with streptavidin, as previously described (Ribeiro et al., 2016). In the sample cell, the PAA-sugars were immobilized either as a mixture for low-density chips (1:9 of sugar of interest: non-interacting sugar) or as an individual PAA-sugar. The solutions were prepared at

a concentration 200 µg/mL in 10 mM HEPES buffer pH 7.5 with 100 mM NaCl and 0.05% Tween 20 (Buffer S). The reference cell was prepared in the same way as described for the sample cell, however, this time only non-interacting sugars were used. All the experiments were carried out in Buffer S. For the titration experiments, different concentrations of protein were injected onto the chip with the flow of 30 µL/min, and the protocol included steps: association time 300 s, dissociation time 300 s, 2x regeneration step for 180 s by 1 M fucose. The singly cycle kinetics experiment was carried out with the flow of 10 µL/min with an association and dissociation time of 120 s, while the regeneration was done only at the end of the run. For the inhibition assay, RSL-MOA with the concentration of 50 nM was pre-incubated with various concentrations of L-fucose for minimum 1 h at 4°C and subsequently injected onto the fucose chip. The used protocol was identical to the one used for titration experiments. The data analysis was performed by BIAevaluation software.

Protein labeling

RSL-MOA and MOAβT were dissolved at 1 mg/mL in Dulbecco's phosphate-buffered saline (PBS) and stored at 4°C prior to usage. For fluorescent labeling, NHS-ester conjugated Atto488 (Thermo Fisher) or Cy5 (GE Healthcare) was used. Fluorescent dye was dissolved at a final concentration of 10 mg/mL in water-free DMSO (Carl RothGmbH & Co), aliquoted, and stored at -20°C before usage according to the manufacturer´s protocol. For the labeling reaction, 100 µL of lectin (1 mg/mL) was supplemented with 10 µL of a 1 M $NaHCO_3$ (pH 9) solution. Hereby, the molar ratio between dye and lectin was 5:1 for RSL-MOA (cell assays) and 2:1 for RSL-MOA and MOAβT for GUVs assays. The labeling mixture was incubated at 4°C for 90 min, and uncoupled dyes were separated using Zeba Spin desalting columns (7 kDa MWCO, 0.5 mL, Thermo Fischer). Labeled lectins were stored at 4°C, protected from light.

Composition and preparation of giant unilamellar vesicles (GUVs)

GUVs were composed of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), cholesterol (both AvantiPolar Lipids, United States), Atto647N 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE; Sigma Aldrich), and one of the following glycolipids at a molar ratio of 64.7:30:0.3:5. The glycolipids are FSL-A (Function-Spacer-Lipid with blood group A trisaccharide) (SigmaAldrich), FSL-B (Function-Spacer-Lipid with blood group B trisaccharide) (Sigma Aldrich) or FSL-isoGb3 (Function-Spacer-Lipid with iso-globotriaosyl saccharide).

GUVs were prepared by the electroformation method as earlier described (Madl et al., 2016). Briefly, lipids dissolved in chloroform of a total concentration of 0.5 mg/mL were spread on indium tin oxid-covered (ITO) glass slides and dried in a vacuum for at least one hour or overnight. Two ITO slides were assembled to create a chamber filled with sucrose solution adapted to the osmolarity of the imaging buffer of choice, either HBSS (live-cell imaging) or PBS (GUVs only imaging). Then, an alternating electrical field with a field strength of 1 V/mm was implemented for 2.5 hours at RT. Later we observed the GUVs in chambers manually built as described (Madl et al., 2016).

Imaging of RSL-MOA and MOAβT binding to GUVs

Samples of GUVs and lectins were imaged using a confocal fluorescence microscope (Nikon Eclipse Ti-E inverted microscope equipped with Nikon A1R confocal laser scanning system, 60x oil immersion objective, NA = 1.49, and four laser lines: 405 nm, 488 nm, 561 nm, and 640 nm). Image acquisition and processing were made using the software NIS-Elements (version 4.5, Nikon) and open-source Fiji software (https://imagej.net/software/fiji/).

Cell culture

The human lung epithelial cell line H1299 (American Type Culture Collection, CRL-5803) was cultured in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% fetal calf serum (FCS) and 4 mM L-glutamine at 37°C and 5% CO2, under sterile conditions. Cells were cultivated in standard TC-dishes 100 (Sarstedt AG & Co. KG, Numbrecht, Germany) until 90% confluence, detached with trypsin (0.05% trypsin-EDTA solution; Sigma-Aldrich Chemie GmbH, Darmstadt, Germany) and re-seeded for a subculture or for experiments. For experiments, cells were incubated with different concentrations of RSL or RSL-MOA for indicated time points.

Flow Cytometry Analysis

H1299 cells were detached with 2 mL of 1.5 mM EDTA in PBS (-/-), and 1×10^5 cells were counted and transferred to a U-bottom 96 well plate (Sarstedt AG & Co. KG, Numbrecht, Germany). To quantify protein binding to cell surface receptors, cells were incubated with different concentrations of fluorescently labeled RSL or RSL-MOA lectin for 30 min at 4°C and protected from light compared to PBS-treated cells as a negative control. For the saturation of glycan-binding sites, 100 nM RSL AF488 or 0.18 µM RSL-MOA Cy5 were preincubated with 25, 50 or 100 mM soluble L-fucose or with 100 mM 4-Nitrophenyl α-D-galactopyranoside (PNPG; Sigma-Aldrich, Chemie GmbH, Darmstadt, Germany), for 30 min at RT and in the absence of light. At the end of pre-incubation, the solution was diluted 100 times and added to cells for 30 min at 4°C, in the dark. Subsequently, cells were centrifuged at 1600 × g for 3 min at 4°C and washed twice with FACS buffer (PBS (-/-) supplemented with 3% FCS v/v). After the last washing step, the cells were resuspended with FACS buffer and transferred to FACS tubes (Kisker Biotech GmbH Co. KG, Steinfurt, Germany) on ice and protected from light. The fluorescence intensity of treated cells was monitored at FACS Gallios (Beckman Coulter Inc., Brea, CA, USA) and further analyzed using FlowJo V.10.5.3.

Differential scanning calorimetry (DSC)

The DCS experiments were carried out in Micro-Cal PEAQ DSC instrument (Malvern Panalytical). The protein samples were prepared in Buffer A, which could be substituted with the addition of a ligand. The sample cell contained the protein solutions in the concentration range of 9-29 mM. The reference cell contained the same buffer as present in the sample but without protein. The increase in temperature was measured from 20-130°C with a scan rate of 200°C/min. The data were analyzed by Micro-Cal PEAQ DSC software with a non-two-state and progress baseline method. fitting model.

Acknowledgment: This research was funded by the European Union Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement synBIOcarb (No. 814029). The authors acknowledge Malvern Panalytical for Micro-Cal PEAQ DSC instrument.

Bibliography

Arnaud, J., Claudinon, J., Tröndle, K., Trovaslet, M., Larson, G., Thomas, A., Varrot, A., Römer, W., Imberty, A., & Audfray, A. (2013). Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity. *ACS Chemical Biology*, *8*(9), 1918–1924. https://doi.org/10.1021/cb400254b

Bonnardel, F., Mariethoz, J., Salentin, S., Robin, X., Schroeder, M., Perez, S., Lisacek, F. D. S., & Imberty, A. (2019). Unilectin3d, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky832

Chung, C. H., Mirakhur, B., Chan, E., Le, Q.-T., Berlin, J., Morse, M., Murphy, B. A., Satinover, S. M., Hosen, J., Mauro, D., Slebos, R. J., Zhou, Q., Gold, D., Hatley, T., Hicklin, D. J., & Platts-Mills, T. A. E. (2008). Cetuximab-Induced Anaphylaxis and IgE Specific for Galactose-α-1,3-Galactose. *New England Journal of Medicine*, *358*(11), 1109–1117. https://doi.org/10.1056/nejmoa074943

Cooper, D. K. C., Koren, E., & Oriol, R. (1994). Oligosaccharides and Discordant Xenotransplantation. *Immunological Reviews*, *141*(1). https://doi.org/10.1111/j.1600-065X.1994.tb00871.x

Cordara, G., Egge-Jacobsen, W., Johansen, H. T., Winter, H. C., Goldstein, I. J., Sandvig, K., & Krengel, U. (2011). Marasmius oreades agglutinin (MOA) is a chimerolectin with proteolytic activity. *Biochemical and Biophysical Research Communications*, *408*(3), 405–410. https://doi.org/10.1016/j.bbrc.2011.04.031

de Mol, N. J., & Fischer, M. J. E. (2010). Surface Plasmon Resonance: Methods and Protocols. In *Life Sciences*. https://doi.org/10.1007/978-1-60761-670-2

ESTOLA, E., & ELO, J. (1952). Occurrence of an exceedingly weak A blood group property in a family. *Annales Medicinae Experimentalis et Biologiae Fenniae*.

Fettis, M. M., Farhadi, S. A., & Hudalla, G. A. (2019). A chimeric, multivalent assembly of galectin-1 and galectin-3 with enhanced extracellular activity. *Biomaterials Science*, *7*(5). https://doi.org/10.1039/c8bm01631c

Galili, U., Shohet, S. B., Kobrin, E., Stults, C. L., & Macher, B. A. (1988). Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. *The Journal of Biological Chemistry*, *263*(33), 17755–17762. https://doi.org/10.1016/s0021-9258(19)77900-9

Grahn, E., Askarieh, G., Holmner, Å., Tateno, H., Winter, H. C., Goldstein, I. J., & Krengel, U. (2007). Crystal Structure of the Marasmius Oreades Mushroom Lectin in Complex with a Xenotransplantation Epitope. *Journal of Molecular Biology*, *369*(3), 710–721. https://doi.org/10.1016/j.jmb.2007.03.016

Hazes, B. (1996). The (QxW)3 domain: A flexible lectin scaffold. *Protein Science*, *5*(8), 1490–1501. https://doi.org/10.1002/pro.5560050805

Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W. H., & Blanchard, L. (2007). Interaction of the C-Terminal Domains of Sendai Virus N and P Proteins: Comparison of Polymerase-Nucleocapsid Interactions within the Paramyxovirus Family. *Journal of Virology*, *81*(13). https://doi.org/10.1128/jvi.00338-07

Irumagawa, S., Hiemori, K., Saito, S., Tateno, H., & Arai, R. (2022). Self-Assembling Lectin Nano-Block Oligomers Enhance Binding Avidity to Glycans. *International Journal of Molecular Sciences*, *23*(2). https://doi.org/10.3390/ijms23020676

Jia, L., Zhang, J., Ma, T., Guo, Y., Yu, Y., & Cui, J. (2018). The function of fucosylation in progression of lung cancer. *Frontiers in Oncology*, *8*(DEC), 1–10. https://doi.org/10.3389/fonc.2018.00565

Kostlánová, N., Mitchell, E. P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerová, M., & Imberty, A. (2005). The fucose-binding Lectin from Ralstonia solanacearum: A new type of β-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M505184200

Kruger, R. P., Winter, H. C., Simonson-Leff, N., Stuckey, J. A., Goldstein, I. J., & Dixon, J. E. (2002). Cloning, expression, and characterization of the Galα1,3Gal high affinity lectin from the mushroom Marasmius oreades. *Journal of Biological Chemistry*, *277*(17), 15002–15005. https://doi.org/10.1074/jbc.M200165200

Macher, B. A., & Galili, U. (2008). The Galα1,3Galβ1,4GlcNAc-R (α-Gal) epitope: A carbohydrate of unique evolution and clinical relevance. *Biochimica et Biophysica Acta - General Subjects*, *1780*(2), 75–88. https://doi.org/10.1016/j.bbagen.2007.11.003

Madl, J., Villringer, S., & Römer, W. (2016). *Delving into Lipid-Driven Endocytic Mechanisms Using Biomimetic Membranes*. https://doi.org/10.1007/8623_2016_7

Murzin, A. G., Lesk, A. M., & Chothia, C. (1992). β-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. *Journal of Molecular Biology*, *223*(2), 531–543. https://doi.org/10.1016/0022-2836(92)90668-A

Notova, S., Bonnardel, F., Rosato, F., Siukstaite, L., Schwaiger, J., Bovin, N., Varrot, A., Römer, W., Lisacek, F., & Imberty, A. (2022). A pore-forming β-trefoil lectin with specificity for the tumor-related glycosphingolipid Gb3. *BioRxiv*. https://doi.org/10.1101/2022.02.10.479907

Oh, Y. J., Dent, M. W., Freels, A. R., , Qingwen Zhou, C. B., Lebrilla, & Matoba, Michael L. Merchant, N. (2022). Antitumor activity of a lectibody targeting cancer-associated high-mannose glycans. *Molecular Therapy*, *30*(4), 1523–1535.

Ramberg, K. O., Guagnini, F., Engilberge, S., Wrońska, M. A., Rennie, M. L., Pérez, J., & Crowley, P. B. (2021). Segregated Protein–Cucurbit[7]uril Crystalline Architectures via Modulatory Peptide Tectons. *Chemistry - A European Journal*, *27*(59), 14619–14627. https://doi.org/10.1002/chem.202103025

Ribeiro, Joao P., Pau, W., Pifferi, C., Renaudet, O., Varrot, A., Mahal, L. K., & Imberty, A. (2016). Characterization of a high-affinity sialic acid-specific CBM40 from Clostridium perfringens and engineering of a divalent form. *Biochemical Journal*, *473*(14), 2109–2118. https://doi.org/10.1042/BCJ20160340

Ribeiro, João P., Villringer, S., Goyard, D., Coche-Guerente, L., Höferlin, M., Renaudet, O.,

Römer, W., & Imberty, A. (2018). Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells. *Chemical Science*, *9*(39), 7634–7641. https://doi.org/10.1039/c8sc02730g

Ross, J. F., Wildsmith, G. C., Johnson, M., Hurdiss, D. L., Hollingsworth, K., Thompson, R. F., Mosayebi, M., Trinh, C. H., Paci, E., Pearson, A. R., Webb, M. E., & Turnbull, W. B. (2019). Directed Assembly of Homopentameric Cholera Toxin B-Subunit Proteins into Higher-Order Structures Using Coiled-Coil Appendages. *Journal of the American Chemical Society*, *141*(13), 5211–5219. https://doi.org/10.1021/jacs.8b11480

Sharon, N., & Lis, H. (2004). History of lectins: From hemagglutinins to biological recognition molecules. In *Glycobiology*. https://doi.org/10.1093/glycob/cwh122

Siukstaite, L., Rosato, F., Mitrovic, A., Müller, P. F., Kraus, K., Notova, S., Imberty, A., & Römer, W. (2021). The two sweet sides of janus lectin drive crosslinking of liposomes to cancer cells and material uptake. *Toxins*, *13*(11), 1–19. https://doi.org/10.3390/toxins13110792

Tateno, H., & Goldstein, I. J. (2004). Partial identification of carbohydrate-binding sites of a Galα1,3Galβ1,4GlcNAc-specific lectin from the mushroom Marasmius oreades by site-directed mutagenesis. *Archives of Biochemistry and Biophysics*, *427*(1), 101–109. https://doi.org/10.1016/j.abb.2004.04.013

Terada, D., Voet, A. R. D., Noguchi, H., Kamata, K., Ohki, M., Addy, C., Fujii, Y., Yamamoto, D., Ozeki, Y., Tame, J. R. H., & Zhang, K. Y. J. (2017). Computational design of a symmetrical β-trefoil lectin with cancer cell binding activity. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-06332-7

Turnbull, W. B., & Daranas, A. H. (2003). On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *Journal of the American Chemical Society*, *125*(48). https://doi.org/10.1021/ja036166s

Ward, E. M., Kizer, M. E., & Imperiali, B. (2021). Strategies and Tactics for the Development of Selective Glycan-Binding Proteins. *ACS Chemical Biology*, *16*(10), 1795–1813. https://doi.org/10.1021/acschembio.0c00880

Winter, H. C., Mostafapour, K., & Goldstein, I. J. (2002). The mushroom Marasmius oreades lectin is a blood group type B agglutinin that recognizes the Galα1,3Gal and Galα1,3Gal,β1,4GlcNAc porcine xenotransplantation epitopes with high affinity. *Journal of Biological Chemistry*, *277*(17), 14996–15001. https://doi.org/10.1074/jbc.M200161200

7. CONCLUSIONS AND PERSPECTIVES

Glycobiology is a rapidly-developing field of natural sciences. Researchers all around the world start to realize how important is the role of glycans and glycoconjugates in intercellular connection and signaling. Lectins, as glycan-recognition proteins, can be considered mediators for such communication. The interaction is reversible and lectins do not modify the glycan structures. Each lectin is selective, which means that it can interact only with limited types of glycoconjugates. In order to assure that the interaction between glycan and lectin is strong enough, lectins are usually multivalent. The increase in the binding sites can enlarge the affinity toward ligand, especially on the surface interfaces, i.e., cell membrane. Due to their properties, lectins found various applications in biotechnology and biomedicine, however, there is a constant urge to discover or engineer new lectins with more controlled and defined properties.

In this thesis, we have introduced several ways how new lectins can be discovered or how their topology or even specificity can be modified by approaches of synthetic biology. The importance of lectin engineering was also discussed in our review article. We have accomplished all our objectives and the results of my Ph.D. were presented and discussed in three separate scientific-article manuscripts (the fourth one in the annex section where I contributed as a co-author).

The first manuscript is focused on the discovery of new lectins and their characterization. We developed a novel lectin database, TrefLec, designed for the prediction of β-trefoil lectins. The functionality of the database was confirmed by an empirical approach where a predicted lectin was evaluated by experimental procedures. The study includes a spectrum of various laboratory methods, from molecular cloning, protein production and purification, biophysical characterization, and cellular and proto-tissues studies supplemented with structural characterization and molecular modeling. We confirmed that the newly discovered lectin, SaroL-1, truly possesses β-trefoil architecture and we evaluated its specificity. As predicted, this lectin is specific for α-galactosylated glycoconjugates including the cancer epitope Gb3. More interestingly, SaroL-1 showed a cytotoxic effect upon binding to Gb3+ cancer cells and induced cell lysis in rabbit erythrocytes. The cytotoxicity is conditioned by lectin/glycoconjugate interaction as confirmed by several experimental procedures. In fact, SaroL-1 is composed of two domains, β-trefoil and pore-forming domain and upon membrane binding, this lectin undergoes extensive structural changes and oligomerizes into a pore. To

our knowledge, SaroL-1 is the first pore-forming lectin specific to cancer epitope Gb3 making him a suitable candidate for further development in pore nanotechnology and cancer therapy.

The second project was very different to the first one. We used synthetic biology approach to create artificial proteins with increased valency and dual specificity. Engineered lectins proved that the multivalency affected their affinity. Additionally, we aimed to expand the idea of synthetic biology even further in order to construct artificial plant cell walls. Plant cell walls and their individual components, i.e., cellulose, hemicellulose, and pectin, are, indeed, extensively studied over decades. There have been several attempts to construct an artificial plant cell wall, however, to our knowledge, there is no evidence of such reconstruction, mostly because of the nonsuccess of pectin incorporation. On the other hand, the interaction between lipid bilayer, cellulose, and hemicellulose or pectin has been already proven. Therefore, we designed a protein glue, bispecific Janus lectin RSL-CBM77$_{Rf}$, that is able to interact at the same time with fucosylated xyloglucan (hemicellulose representative) and homogalacturonan (pectin representative). Furthermore, we verified the activity of protein RSL-CBM77$_{Rf}$ in the multilayer organization and we confirmed that thanks to our engineered lectin, pectin, i.e., homogalacturonan, was successfully attached to the fucosylated xyloglucan/ cellulose nanocrystals/lipid bilayer assembly.

In the last manuscript draft, the concept of Janus lectins was enriched by the creation of a bispecific chimera composed of two distinct lectin domains, MOA β-trefoil and RSL β-propeller. Third Janus lectin, RSL-MOA, possesses the specificity for fucosylated and α-galactosylated glycoconjugates. Its binding activity was compared with the engineered monomeric β-trefoil domain of MOA, MOAβT, and the results clearly proved the importance of lectin multivalency. Furthermore, all three Janus lectins, RSL-CBM40, RSL-CBM77$_{Rf,}$ and RSL-MOA, plus MOAβT and RSL were compared in terms of their thermal stability. All Janus lectins have two events of denaturation indicating that the melting temperature of individual domains is variable. Surprisingly, regarding the melting temperature, β-sandwich fold, present in both CBMs, is significantly more stable than β-trefoil fold which highlights the fact that the selection of structural fold should be taken into account while designing future Janus lectins.

The findings we presented here could be further explored in future research. The pore-forming toxins reveal the structural complexity, and for their determination, the usage of cryo-electron microscopy would be of interest. SaroL-1 as the first described pore-forming lectin specific for cancer epitope Gb3, has a great potential for protein engineering. Firstly, the engineering and

creation of the individual β-trefoil domain would be of interest in order to compare the cytotoxic properties with SaroL-1. Additionally, β-trefoil domain could be fused with another lectin giving creation to Janus lectin, with increased specificity for cancer cells and potential application in diagnostics or even treatment. On the other hand, the idea of joining the pore-forming domain to different proteins is even more intriguing. Is it possible that such an assembly would behave in a similar way as SaroL-1 and thus induce a pore formation? For now, we can only hypothesize the answer, nonetheless, in the field of pore-engineering, various attempts, e.g., directed modification of pH sensibility, optogenetics, nanopore biosensors, or sequencing, have been developed.

The construction of primitive artificial plant cell wall was successful in the multilayer format. The next step would involve the use of the lipid vesicles, as more acute model for cells. This would allow to build close-to-reality 3D proto-cell coated with individual components of the plant cell wall. Although the idea might seem simple, there is a lot of polymer science behind it. We understand how the protein behaves in the presence of individual polymers but it is of interest to provide mechanical studies as well and see the effect of protein on the polymers and vice versa. On the other hand, the engineering of proteins, e.g., CBMs, could have a potential for polymer cross-linking due to multivalency. Proper protein design could for example result in sample jellification or induce a glue-like effect.

Janus lectins have been presented as a universal strategy for the creation of bispecific lectin. Indeed, due to RSL oligomerization, the valency was increased resulting in higher affinity. The concept of various protein scaffolds has been already partially explored. However, I am more intrigued by the idea of additional specificity. Most of the up-to-date engineered lectins are bispecific. But what if we could introduce the third specificity? The strategy of fusion proteins might be considered a choice number one since has been proven to be efficient already several times. Nonetheless, protein co-expression could be another interesting strategy in which two protein variants, e.g., two RSLs with two different extra domains, would be produced in one expression system at the same time and during protein oligomerization would assemble together as a mixture. These would also give a heterogeneous protein population but with the process optimization, maybe, one day we can produce such multi-specific lectins.

Even though the area of protein engineering is relatively young, we demonstrated, that it is certainly a powerful and progressive tool that brings new opportunities and possibilities into protein sciences, and nowadays it is our creativity that starts to rule our knowledge.

8. ANNEX

8.1 Scientific article IV

During my Ph.D. I have been involved in the continuation of the project of the first Janus lectin RSL-CBM40. The project was a collaboration within SynBIOcarb network, with the team of Prof. Winfried Römer (University of Freiburg, Germany) and another two Ph.D. students, ESR4 Lina Suikstaite and ESR5 Francesca Rosato. Lina carried out the experiments on GUVs and Francesca was involved in cell studies. My participation in this project was the production and purification of Janus lectin RSL-MOA. Later on, I was engaged in manuscript preparation and corrections.

The results were currently published in the scientific paper entitled “The Two Sweet Sides of Janus Lectin Drive Crosslinking of Liposomes to Cancer Cells and Material Uptake” and was published in 2021 in the journal Toxins, 13 (11), p. 729. Janus lectin RSL-CBM40 was designed to be specific to fucosylated and sialylated glycoconjugates (João P. Ribeiro et al., 2018). This study continued with previous results and Janus lectin RSL-CBM40 was involved in the crosslinking of fucosylated GUVs with hypersialylated cancer cells H1299. Moreover, RSL-CBM40 induced the cellular uptake of the intact liposome inside of the cancer cell which proves the concept that lectins could be potential drug carriers.

toxins

MDPI

Article

The Two Sweet Sides of Janus Lectin Drive Crosslinking of Liposomes to Cancer Cells and Material Uptake

Lina Siukstaite [1,2,*,†], Francesca Rosato [1,2,*,†], Anna Mitrovic [1,2], Peter Fritz Müller [1,2], Katharina Kraus [1,2], Simona Notova [3], Anne Imberty [3,*] and Winfried Römer [1,2,4,*]

1 Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany; anna.mitrovic@pluto.uni-freiburg.de (A.M.); peter.fritz.mueller@bioss.uni-freiburg.de (P.F.M.); k.kraus97@yahoo.de (K.K.)
2 Signaling Research Centers BIOSS and CIBSS, University of Freiburg, 79104 Freiburg, Germany
3 University of Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France; simona.notova@cermav.cnrs.fr
4 Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, 79104 Freiburg, Germany
* Correspondence: lina.siukstaite@bioss.uni-freiburg.de (L.S.); francesca.rosato@bioss.uni-freiburg.de (F.R.); anne.imberty@cermav.cnrs.fr (A.I.); winfried.roemer@bioss.uni-freiburg.de (W.R.); Tel.: +49-(0)-761-203-67500 (W.R.)
† Contributed equally.

Abstract: A chimeric, bispecific Janus lectin has recently been engineered with different, rationally oriented recognition sites. It can bind simultaneously to sialylated and fucosylated glycoconjugates. Because of its multivalent architecture, this lectin reaches nanomolar avidities for sialic acid and fucose. The lectin was designed to detect hypersialylation—a dysregulation in physiological glycosylation patterns, which promotes the tumor growth and progression of several cancer types. In this study, the characteristic properties of this bispecific Janus lectin were investigated on human cells by flow cytometry and confocal microscopy in order to understand the fundamentals of its interactions. We evaluated its potential in targeted drug delivery, precisely leading to the cellular uptake of liposomal content in human epithelial cancer cells. We successfully demonstrated that Janus lectin mediates crosslinking of glyco-decorated giant unilamellar vesicles (GUVs) and H1299 lung epithelial cells. Strikingly, the Janus lectin induced the internalization of liposomal lipids and also of complete GUVs. Our findings serve as a solid proof of concept for lectin-mediated targeted drug delivery using glyco-decorated liposomes as possible drug carriers to cells of interest. The use of Janus lectin for tumor recognition certainly broadens the possibilities for engineering diverse tailor-made lectin constructs, specifically targeting extracellular structures of high significance in pathological conditions.

Keywords: chimeric carbohydrate; protein engineering; hypersialylation; giant unilamellar vesicles; drug delivery; cancer cell targeting; live-cell imaging

Key Contribution: The present study describes the ability of the bifunctional Janus lectin to target hypersialylation on the cell surface of human epithelial cancer cells and to trigger the internalization of liposomal lipids and complete liposomes.

Citation: Siukstaite, L.; Rosato, F.; Mitrovic, A.; Müller, P.F.; Kraus, K.; Notova, S.; Imberty, A.; Römer, W. The Two Sweet Sides of Janus Lectin Drive Crosslinking of Liposomes to Cancer Cells and Material Uptake. *Toxins* **2021**, *13*, 792. https://doi.org/10.3390/toxins13110792

Received: 28 September 2021
Accepted: 4 November 2021
Published: 9 November 2021

1. Introduction

Lectins are proteins that recognize and bind carbohydrate moieties attached to proteins and lipids [1–4]. They can interact with mono- or oligosaccharides through non-covalent interactions involving hydrogen bonds, van der Waals, and hydrophobic forces. Lectins are present in all organisms: first described in plants, multiple lectins were isolated from microorganisms and animals in subsequent years [5,6]. Their interactions with carbohydrates, which can be highly specific and reversible [7,8], are crucial for numerous inter- and intracellular processes mediated by unique glycosylation patterns at the cell surface [6,9].

Thus, lectins play a significant role in biological processes such as embryonic development, cell growth and differentiation, endocytosis and molecular trafficking, as well as immunomodulation, inflammation, metastasis, and apoptosis [2,6,10–12]. Pathogens, in turn, use lectins to recognize and attach to carbohydrates present on host tissue, an essential process for invasion and infection [13–15].

The carbohydrate specificity of lectins is the underlying basis for their applications in biological and therapeutic research. Indeed, lectins have been investigated for a variety of new medical approaches, including cancer diagnosis, imaging, drug delivery, and treatment [16,17]. In the past few decades, it has been observed that there are precise distinctions in glycan structure and composition between physiological and pathological states. In cancer, cell surface alterations in glycan synthesis and expression have been widely described [18,19]. These aberrant glycosylation patterns are often associated with malignant transformation, development of metastasis, and more aggressive phenotypes [20]. Such carbohydrate structures are known as tumor-associated carbohydrate antigens (TACAs) and constitute promising targets for lectin-based therapy.

Interestingly, several bacterial lectins have found application in tumor detection or treatment. Shiga toxin (Stx), produced by *Shigella dysenteriae* and enterohemorrhagic *Escherichia coli* (EHEC) strains, has been explored to target human tumors in which the glycosphingolipid globotriaosylceramide (also known as Gb3) is overexpressed (Supplementary Figure S1). The selective recognition of Gb3 by the B-subunit of the Shiga toxin (StxB) has raised interest in using this molecule as a tool for tumor imaging or as a carrier for cytotoxic compounds to cancer cells [21,22]. Similarly, the cholera toxin (Ctx) from *Vibrio cholerae* is an effective protein in tumor targeting. The specific affinity for the ganglioside GM1, which is highly expressed in the blood–brain barrier, resulted in the investigation of its B-subunit (CtxB) for new anti-glioma chemotherapy strategies [23] or as a sensitive retrograde neuronal tracer [24–27]. Furthermore, a novel nanoparticle formulation exploited wheat germ agglutinin (WGA) binding to its glycan receptor for local delivery of paclitaxel to the lung [28]. The receptor-mediated cellular uptake of WGA-functionalized PEG nanoparticles has been proven to be more efficient and led to a superior in vitro anticancer activity compared to the clinical paclitaxel formulation [29]. However, lectins may induce some degrees of toxicity in host cells, either indirectly by being associated with a cell-damaging, catalytic domain (e.g., Ctx, Stx, MytiLec), or directly through their ability to agglutinate red blood cells, or by affecting membrane structure (e.g., MOA) [30,31]. The toxicity is mainly related to the multivalency of lectins. This potential toxicity, which can be an asset when targeting (cancer) cells, has to be taken into account when designing lectin-based vectorization strategies.

Among the frequently occurring dysregulations in glycosylation patterns, sialylation represents an established hallmark of several cancer types, including lung, breast, ovarian, pancreatic, and prostate cancer [32]. Hypersialylation, which occurs either via (i) the upregulation of sialyltransferase enzymes; (ii) the downregulation of neuraminidase enzymes; or (iii) both [33,34], results in an accumulation of sialic acid exposed on cell surfaces. Hypersialylation promotes continuous tumor growth metastasis through several mechanisms, including enhancing immune evasion and stimulating tumor invasion and migration [32]. In clinical practice, sialylation-based cancer biomarkers such as pancreatic cancer marker CA19-9 (Sialyl-Lewis a: sLe^a) and the sialylated mucin biomarkers CA125 (MUC16) and CA15-3 (MUC1 analog) are used to detect ovarian and breast cancers [35,36]. Different strategies have been adopted in the past few years to target these alterations in sialic acid expression, such as using chemical inhibitors of sialyltransferases and specific silencing of the gene expression of sialyltransferases [37]. Nevertheless, new treatments targeting different interplays of sialylation are continuously explored.

A tandem construction of two carbohydrate-binding modules (CBM) from the NanI sialidase of *Clostridium perfringens* (di-CBM40) was previously described to have strong avidity for sialylated surfaces, with a Kd of 1.3 μM for 3′-siallylactose chips compared to 14.4 μM for monomeric CBM40 [38]. The high avidity for sialylated surfaces displayed

by engineered multivalent CBM40 is of high interest in designing drug delivery strategies toward sialylated epitopes on cell surfaces. Recently, a chimeric bispecific Janus lectin was engineered with two rationally oriented and distinct carbohydrate recognition surfaces [39]. It was designed by assembling the sequences of the sialic acid-binding CBM40 and the fucose-binding lectin of *Ralstonia solanacearum* (RSL) (Figure 1a) and referred to as FS-Janus lectin. As RSL naturally assembles as a trimeric β-propeller fold [40], FS-Janus lectin accordingly adopts a trimeric structure comprising one RSL trimer (six fucose binding sites) linked to three CBM40 monomers (Figure 1b) [38,40]. The multivalent presentation of binding sites on each face of the FS-Janus lectin allows for simultaneous recognition and high avidity binding toward fucosylated and sialylated surfaces, with a Kd in the range of 50–100 nM [39]. The dual sugar specificity of FS-Janus lectin enabled crosslinking of heterogeneous GUV populations functionalized either with the ganglioside GM3 (exposing a terminal sialic acid) or the function-spacer-lipid construct comprised of the blood group A trisaccharide (FSL-A, with terminal fucose) linked to the phospholipid DOPE, resulting in the organization of proto-cells into modular structures resembling proto-tissues.

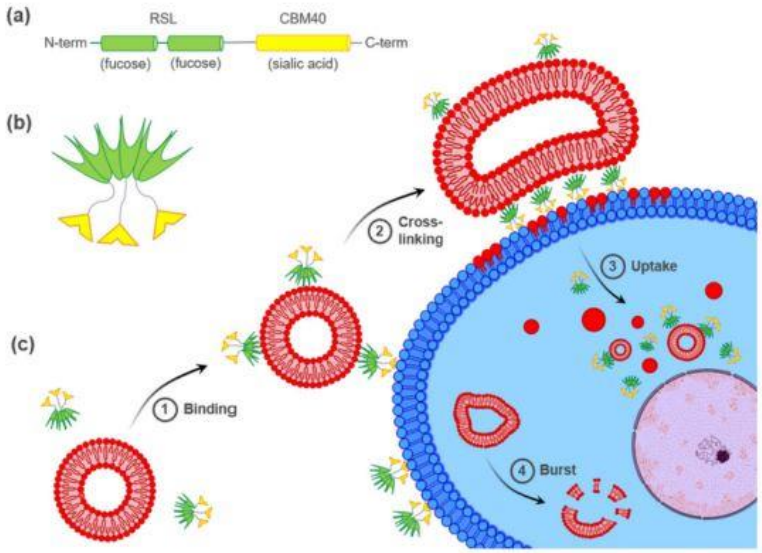


Figure 1. Hypothetical model of FS-Janus lectin-mediated crosslinking of glyco-decorated liposomes and human cancer cells leading to the cellular uptake of complete liposomes and liposomal content. (**a**) The peptide sequence of FS-Janus lectin with two blades of RSL (green) connected to CBM40_NanI (yellow) through a peptide linker. (**b**) Schematic representation of FS-Janus lectin presenting six fucose-binding sites at the upper face and three sialic acid-binding sites at the bottom face. (**c**) FS-Janus lectin binds simultaneously to glyco-decorated liposomes and the cell surface (1) and crosslinks them (2). These tight interactions lead to the cellular uptake of complete liposomes and liposomal content (3). Intracellular liposome transport and burst (4).

In this study, we investigated the ability of FS-Janus lectin to crosslink glyco-decorated GUVs to human hypersialylated epithelial cancer cells and to induce liposomal uptake and transfer of liposomal content. A hypothetical model of FS-Janus lectin-mediated interactions is depicted in Figure 1c. Remarkably, we demonstrated that the FS-Janus lectin is a suitable tool for targeting hypersialylation and delivery of glyco-decorated liposomes to cancer cells, which opens up new possibilities in drug delivery with high specificity.

2. Results

2.1. FS-Janus Lectin Induces Lipid Exchange between GM3- and Blood Group A Trisaccharide-Decorated GUVs

In addition to recent studies that have shown crosslinking of glyco-decorated vesicles (so-called proto-cells) and the organization of proto-cells into proto-tissues (i.e., the assembly of proto-cells into tissue-like structures) mediated by FS-Janus lectin [39], we aimed to further characterize the potential of the FS-Janus lectin. We used giant unilamellar vesicles (GUVs) (i.e., lipid bilayer spheres with variable sizes up to the dimensions of human cells that can be functionalized with diverse glycans) [41]. We incubated FS-Janus lectin (200 nM) with two populations of GUVs, functionalized with either GM3 (monosialodihexosylganglioside, containing sialic acid) or FSL-A (function-spacer-lipid comprised of the blood group A trisaccharide containing fucose, linked to the phospholipid DOPE) and monitored the evolving interactions over time. Vesicles were labeled by fluorescent lipids with an Atto 647N fluorophore (red color) and an Atto 488 fluorophore (green color) for GM3- and FSL-A-containing GUV populations, respectively. We first observed crosslinking and proto-tissue formation of two distinct populations of sialylated and fucosylated liposomes accompanied by clearly visible shape changes of liposomes, particularly at the interfaces of liposomes (Figure 2, Supplementary Video S1). Having a closer look at the zoomed areas of Figure 2, it became apparent that after 55 min in zoom 1 and after 107 min in zoom 2, the lipids of the two liposome populations mixed over time. Membrane areas of liposomes that were not forming interfaces and that were initially stained only in red became partially green, and vice versa. Thus, a lipid exchange triggered by FS-Janus lectin can be observed as indicated by yellow arrows between two differently glyco-decorated populations of GUVs. However, neither crosslinking, shape changes, nor lipid exchange were observed when the two liposome populations were incubated without FS-Janus lectin (Figure 2, control, Supplementary Video S2).

Moreover, when RSL and di-CBM40 were incubated as separate, non-chimeric components of FS-Janus lectin, we did not observe crosslinking between fucosylated and sialylated populations of GUVs (Supplementary Figure S2a,b). The obtained results showed that FS-Janus lectin can crosslink liposomes and mediate lipid exchange between sialylated and fucosylated GUVs. These results inspired us to further test the properties of FS-Janus lectin on living human cancer cells.

2.2. FS-Janus Lectin and di-CBM40 Bind and Internalize into Human Epithelial Cancer Cells

As a first step to monitor the potential of FS-Janus lectin to target cancer cells and trigger the exchange of liposomal content, we evaluated its functionality in binding to human cancer cells via flow cytometry and confocal imaging. Recent studies showed that modifications in α2,6-sialylation are associated with lung cancer progression. The invasiveness and tumorigenicity of non-small cell lung cancer (NSCLC) might be mediated by alterations in the expression of sialyltransferases, as shown in A549 and H1299 NSCLC cells in vitro [42]. Thus, the human NSCLC cell line H1299 was selected as the preferred cell line for these studies. To assess the capacity of FS-Janus lectin to target sialic acid residues specifically, we compared its binding activity to the sialic acid-binding, purified di-CBM40 protein.

The fluorescent FS-Janus lectin was incubated with H1299 cells for binding studies (Figure 3a). Three different concentrations of FS-Janus lectin (16, 32, and 64 nM) were tested to determine the optimal one. Cells were incubated with fluorescent FS-Janus lectin for 30 min on ice. Then, the unbound lectin was washed away to decrease unspecific signals. The flow cytometry analysis revealed a strong binding of the protein, probably to both fucosylated and sialylated membrane structures in a dose-dependent manner (Figure 3a). Stimulated cells were positive to all tested concentrations. The 32 nM concentration was chosen as the preferred working concentration for further binding studies by flow cytometry, as the shift in fluorescence intensity appeared remarkably distinct without reaching signal saturation.

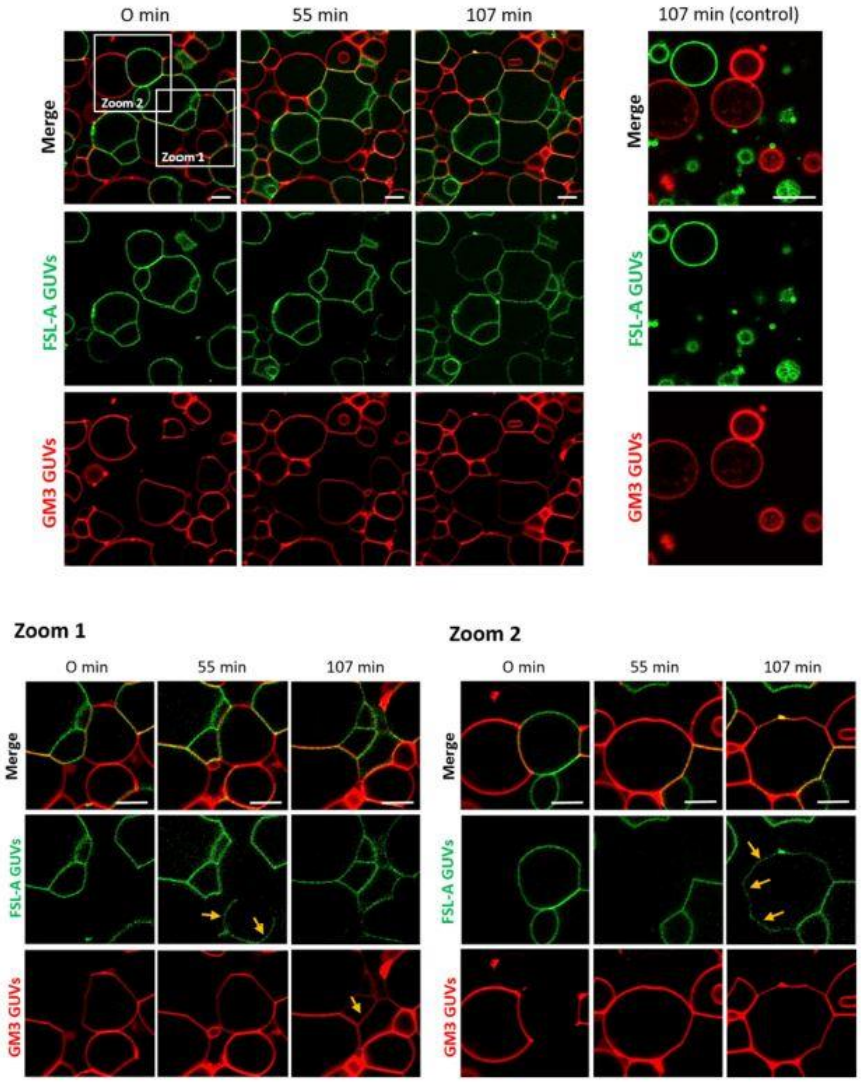


Figure 2. FS-Janus lectin triggers crosslinking and lipid exchange between GM3- and blood group A trisaccharide- (FSL-A) functionalized liposomes. GM3 GUVs (red color; labeled with the fluorescent lipid DOPE-Atto 647N) and FSL-A GUVs (green color; labeled with the fluorescent lipid DOPE-Atto 488) were incubated with 200 nM FS-Janus lectin (unlabeled) at room temperature and were monitored for 120 min using confocal microscopy. The crosslinking between GUVs starts immediately after the addition of FS-Janus lectin, and lipid exchange becomes visible after 55 min of incubation. Some examples are highlighted by yellow arrows. Without the addition of FS-Janus lectin, the liposomes do not crosslink and maintain their round shape over the entire incubation time. The scale bars represent 10 µm.

For comparison, the binding of the sialic acid-specific di-CBM40 to the same cancer cell line was also evaluated by flow cytometry (Figure 3b). When treated with increasing concentrations of di-CBM40, H1299 cells were efficiently stained at the plasma membrane. The binding of CBM40 to cancer cells provided a strong fluorescent signal in a range of protein concentrations from 15 to 214 nM, which also appeared to be dose-dependent (Figure 3b).

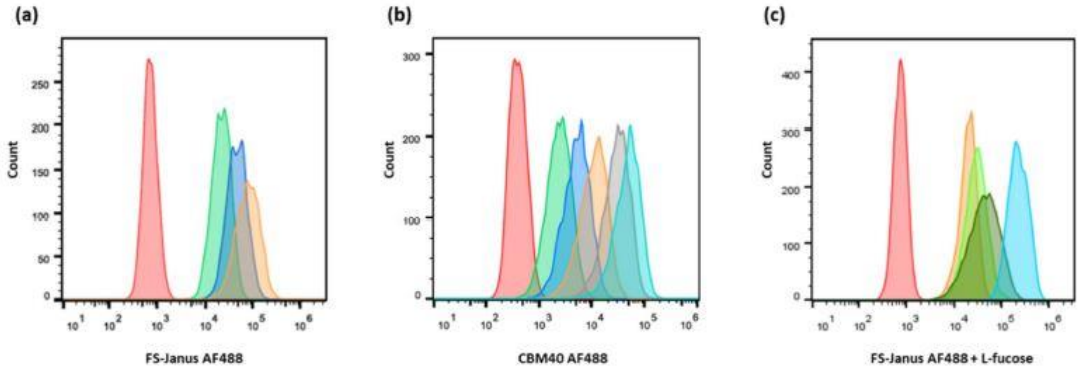


Figure 3. Dose-dependent binding of FS-Janus lectin and di-CBM40 to H1299 lung epithelial cells. Flow cytometry analysis of gated living H1299 cells incubated for 30 min on ice with FS-Janus lectin (**a**), di-CBM40 (**b**), and FS-Janus lectin in the presence of soluble L-fucose (**c**). (**a**) Histogram of fluorescence intensity of H1299 cells stimulated with different concentrations of FS-Janus lectin AF488 (red: negative control, green: 16 nM, blue: 32 nM, orange: 64 nM). (**b**) Histogram of fluorescence intensity of H1299 cells stimulated with different concentrations of di-CBM40 AF488 (red: negative control, green: 15 nM, blue: 21 nM, orange: 42 nM, grey: 107 nM, light blue: 214 nM). (**c**) Histogram of fluorescence intensity of H1299 cells stimulated with FS-Janus lectin AF488 (32 nM) pre-incubated with different concentrations of soluble L-fucose for 30 min at room temperature (RT) to saturate fucose-binding sites (red: negative control, light blue: FS-Janus lectin 32 nM, dark green: FS-Janus lectin 32 nM + 25 mM L-fucose, green: FS-Janus lectin 32 nM + 50 mM L-fucose, orange: FS-Janus lectin 32 nM + 100 mM L-fucose).

The specific inhibition of each of the binding sites by corresponding glycans was checked by pre-incubating FS-Janus lectin with different concentrations of soluble L-fucose or 3′-sialyllactose. The concentration range of L-fucose for the complete inhibition of the RSL domain was tested before with RSL alone by flow cytometry (Supplementary Figure S3a).

H1299 cells were then treated with fluorescent FS-Janus lectin in the presence of 25, 50, and 100 mM of L-fucose for 30 min on ice. After washing away the unbound lectin, a strong signal was recorded by flow cytometry (Figure 3c), confirming the proper functionality and predicted affinity of the CBM40_NanI domain for sialylated receptors.

Next, cells were imaged with confocal microscopy to investigate the cellular uptake of FS-Janus lectin. For these experiments, the concentration of FS-Janus lectin was increased to 150 nM to improve image quality. Cells were pre-incubated with fluorescent FS-Janus lectin for 30 min on ice, and then the unbound lectin was washed away. This condition corresponded to 0 min at 37 °C in the experiment, resulting in plasma membrane staining by the fluorescent lectin. Subsequent incubation at 37 °C revealed uptake and intracellular trafficking of FS-Janus lectin, leading to an accumulation of FS-Janus lectin in the perinuclear area of the cells (white arrows), clearly visible at around 60 min of internalization (Figure 4a).

Similarly, H1299 were stimulated with di-CBM40 for internalization studies. For confocal imaging, a protein concentration of 214 nM was chosen. The intracellular uptake, trafficking, and accumulation of di-CBM40 in the perinuclear area (white arrows, Figure 4b) suggest that this domain could drive the cellular uptake of FS-Janus lectin by specifically targeting hypersialylated cell surface receptors. Furthermore, we investigated the uptake and intracellular accumulation of FS-Janus lectin, which was pre-incubated with 100 mM L-fucose in H1299 cells by confocal imaging (Figure 4c). Incubation at 37 °C with FS-Janus lectin revealed intracellular trafficking and perinuclear accumulation driven by the CBM40_NanI domain of FS-Janus lectin (white arrows). The same experiment was performed with di-CBM40, and soluble 3′-sialyllactose was used to saturate its sialic acid binding sites for flow cytometry analysis (Supplementary Figure S3b). Binding, uptake, and internalization of RSL (Supplementary Figure S4a,b) and FS-Janus lectin in the presence of the inhibitor 3′-sialyllactose (Supplementary Figure S4c,d) were then monitored.

Supplementary Figure S5 shows the co-localization of FS-Janus lectin and the trans-Golgi network.

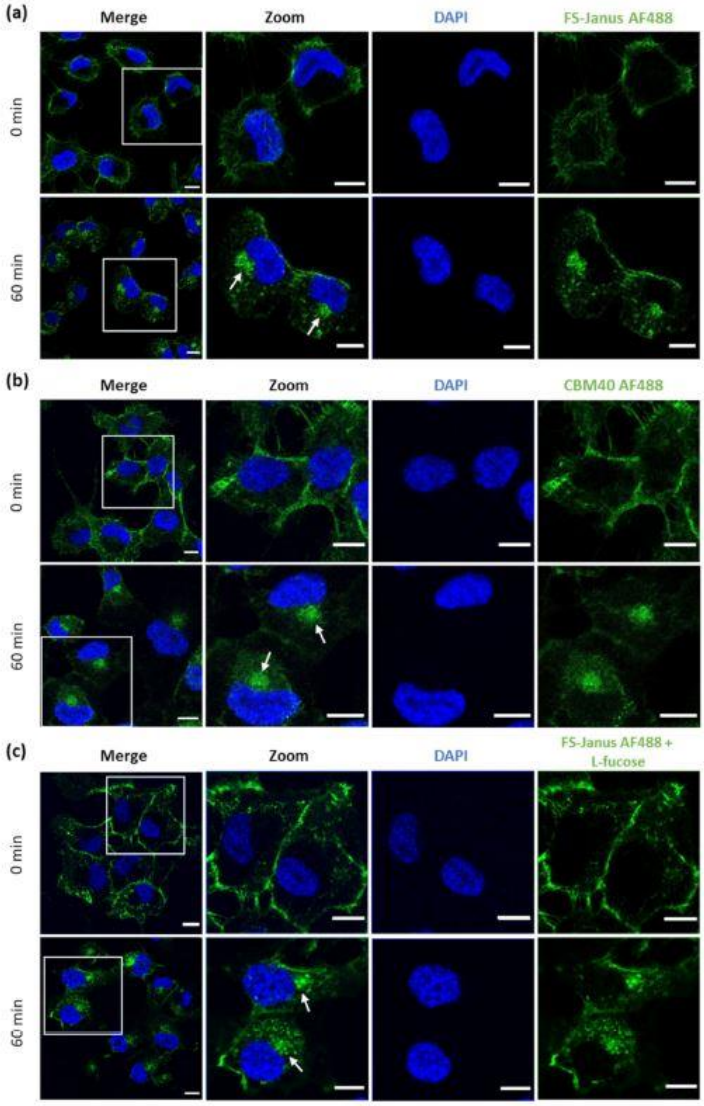


Figure 4. Fluorescence imaging revealed the internalization of FS-Janus lectin and di-CBM40 into H1299 cells. (**a**) Confocal imaging of fluorescently labeled FS-Janus lectin (green color) incubated with H1299 cells at different time points. FS-Janus lectin is internalized within 60 min and accumulates in the perinuclear region of H1299 cells (white arrows). Nuclei were counterstained by DAPI. (**b**) Confocal imaging of di-CBM40 (green color) binding and uptake into H1299 cells at different time points. di-CBM40 was observed intracellularly in H1299 cells after 60 min (as indicated by white arrows). (**c**) Confocal imaging of fluorescently labeled FS-Janus lectin (green color) incubated with H1299 cells at different time points. In the presence of soluble L-fucose, FS-Janus lectin was internalized within 60 min and accumulated in the perinuclear region of H1299 driven by its CBM40 domains. Nuclei were counterstained by DAPI. Scale bars represent 10 µm.

The observations made by flow cytometry and fluorescence imaging confirm that the FS-Janus lectin can selectively target sialylated conjugates on the surface of human epithelial lung cancer cells, and the CBM40 domain is sufficient for intracellular uptake.

2.3. FS-Janus Lectin Mediates Crosslinking of Liposomes with H1299 Cells and the Cellular Uptake of Liposomal Content

To test the ability of FS-Janus lectin to bind to cellular and synthetic membrane systems simultaneously, we incubated blood group A trisaccharide-decorated GUVs and H1299 cells together with FS-Janus lectin and monitored the interactions by live-confocal microscopy. When H1299 cells and blood group A trisaccharide-decorated GUVs were incubated over two hours without FS-Janus lectin, GUVs and H1299 cells did not interact visibly (Supplementary Figure S6 and Supplementary Video S3). Upon treatment with 200 nM FS-Janus lectin, GUVs (red color) in proximity to H1299 cells (partially stained in blue color) started to adhere to cell surfaces. GUVs co-localized at the contact sites to cells over time, where the FS-Janus lectin (green color) was enriched, creating straight, elongated interfaces (Figure 5, white arrows, Supplementary Video S4). Therefore, we concluded that the bispecific FS-Janus lectin mediates crosslinking between liposomes and H1299 cells. The RSL domain binds to the fucosylated surface of blood group A trisaccharide-decorated GUVs and the CBM40 domains to the sialylated receptors of H1299 cells.

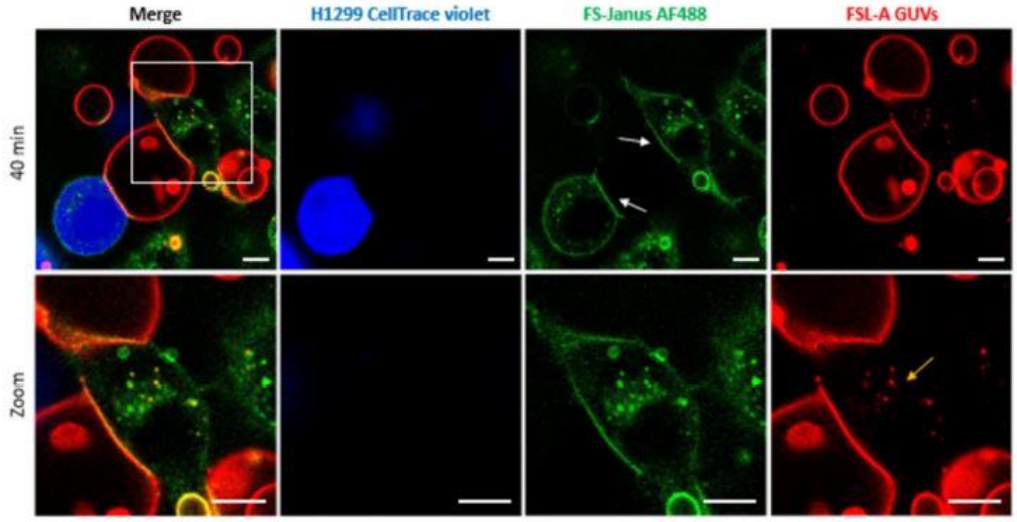


Figure 5. Crosslinking of blood group A trisaccharide-decorated GUVs and H1299 cells is mediated by 200 nM FS-Janus lectin. FS-Janus lectin is enriched at the interfaces (white arrows). The GUVs (red color; labeled with the fluorescent lipid DOPE-Atto 647N) were incubated with the H1299 cells (blue color; partially stained with CellTrace™ Violet) and FS-Janus lectin (green color; labeled with AF488). The yellow arrow indicates perinuclear accumulations of liposomal lipids and FS-Janus lectin. Live-cell imaging experiments were performed at 37 °C by using confocal microscopy and were recorded for 120 min. Scale bars represent 10 µm.

In addition to FS-Janus lectin-mediated crosslinking of liposomes and cells, we observed that FS-Janus lectin was internalized into H1299 cells, accumulating mainly in the perinuclear area over time (Figures 5 and 6). Strikingly, liposomal lipids could also be found in the perinuclear area (yellow arrows in both figures), colocalizing with FS-Janus lectin. At this stage, we can only speculate that liposomal content is transferred to the host cell plasma membrane, from where it may be endocytosed. This could happen along with the uptake of FS-Janus lectin shown in Figure 5. Figure 6 depicts the kinetics of uptake and the enrichment of FS-Janus lectin (red color) and fluorescent lipids (green color) in the perinuclear area over a time window of around two hours (Supplementary Video S5).

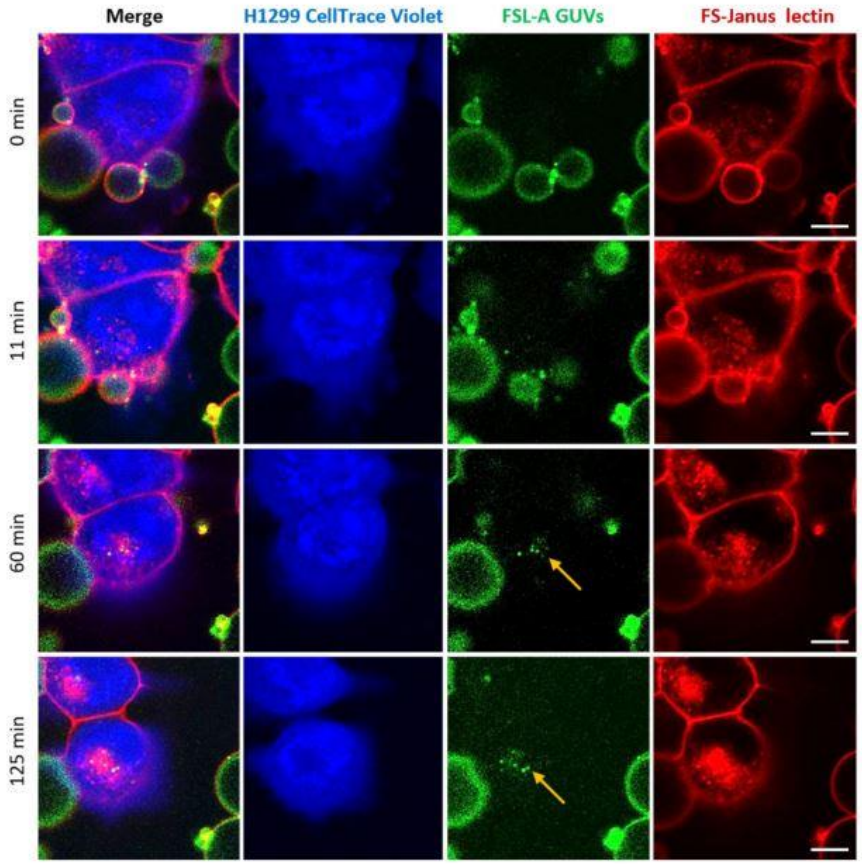


Figure 6. Kinetics of the cellular uptake and perinuclear accumulation of liposomal lipids and FS-Janus lectin. H1299 cells (blue color; partially stained with CellTrace™ Violet) were incubated with 200 nM FS-Janus lectin (red color, labeled with Atto 647) and blood group A trisaccharide-decorated GUVs (green color; labeled with the fluorescent lipid DOPE-Atto 488). Yellow arrows indicate perinuclear accumulation of fluorescent liposomal lipids and FS-Janus lectin. Live-cell imaging experiments were performed at 37 °C by using confocal microscopy, and the kinetics of cellular uptake were recorded for 180 min. Scale bars represent 10 µm.

2.4. FS-Janus Lectin Induces the Uptake of Complete Liposomes into H1299 Cells

One of the most promising observations we made in the framework of these studies marked the complete uptake of intact FSL-A GUVs into H1299 cells, as depicted in Figure 7 and Supplementary Video S6. Images were acquired in a live-cell imaging setting and monitored by confocal microscopy at 37 °C for 120 min. For the selected GUV, the uptake seemed to be mediated by FS-Janus lectin and took place between 25 and 30 min (indicated by yellow arrows). After the first contact of the FS-Janus lectin-bound GUV with the H1299 cell, the plasma membrane engulfed the liposome, and the filopodia-like structure of the cell appeared to drag the liposome rapidly toward the cell body (it took approximately 5 min). Even after internalization, FS-Janus lectin stood bound to the GUV. The cellular uptake of the GUV can also be nicely observed by following the displacement of the CellTrace™ Violet stain (blue color) in the cytosol by the black circular area (the liposome). As observed in Figure 4a, FS-Janus lectin alone also accumulated in the perinuclear area.

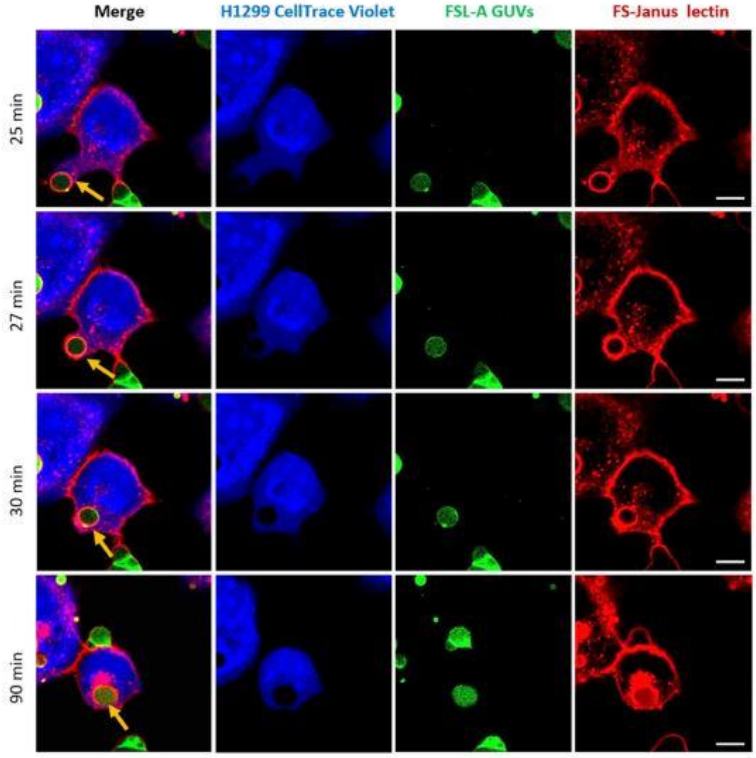


Figure 7. Complete uptake of blood group A-decorated GUVs into H1299 cells. FS-Janus lectin (200 nM, red color; labeled with Atto 647) triggers the internalization of complete liposomes (green color; labeled with fluorescent lipid DOPE-Atto 488) into H1299 cells (blue color; partially stained with CellTrace™ Violet). Yellow arrows point to one GUV that is taken up between 25 and 30 min. Additionally, inside the cell, the liposome remains covered by FS-Janus lectin. Live-cell imaging experiments were performed at 37 °C by using confocal microscopy, and the kinetics of uptake were recorded for 120 min. Scale bars represent 10 µm.

We successfully detected liposomal uptake events multiple times for several biological replicates, using two different concentrations of FS-Janus lectin (either 200 nM or 500 nM). The uptake processes involved liposomes of different sizes and took place at various time points with varying uptake speeds. No uptake events were detected without the addition of FS-Janus lectin to glyco-decorated liposomes and H1299 cells. All these observations allowed us to conclude that the uptake of blood group A-functionalized liposomes into the cytoplasm of H1299 cells was mediated by FS-Janus lectin.

2.5. Uptaken Liposomes Are Deformed and Burst

Strikingly, after the cellular uptake of liposomes triggered by FS-Janus lectin (here: 500 nM), we observed liposome deformations, shrinkages, and bursts in the cytosol of H1299 cells. The experiment presented in Figure 8 and Supplementary Video S7 was recorded over 180 min using a live-cell confocal microscopy setting. The liposome was internalized within the first 25 min of incubation. FS-Janus lectin (red color) remained bound to the GUV (green color), even after internalization. Additionally, in this experiment, the cellular uptake of the GUV can be nicely tracked by the movement of the black circular area in the cytosol of the cell stained by the CellTrace™ Violet marker (blue color). Around 110 min, small deformations of the liposome became apparent. In the following minutes, the liposome further deformed, and several smaller vesicular structures (nicely visible in

green color) formed inside the initially internalized liposome, along with a decrease in size and partial disappearance of bound FS-Janus lectin. Between 127 and 155 min, the liposome as the entity and sub-compartments ruptured and disappeared almost completely. Whether these processes were driven by the host cell, FS-Janus lectin, or the reorganization of membrane lipids remains unclear at this stage, and further research is needed.

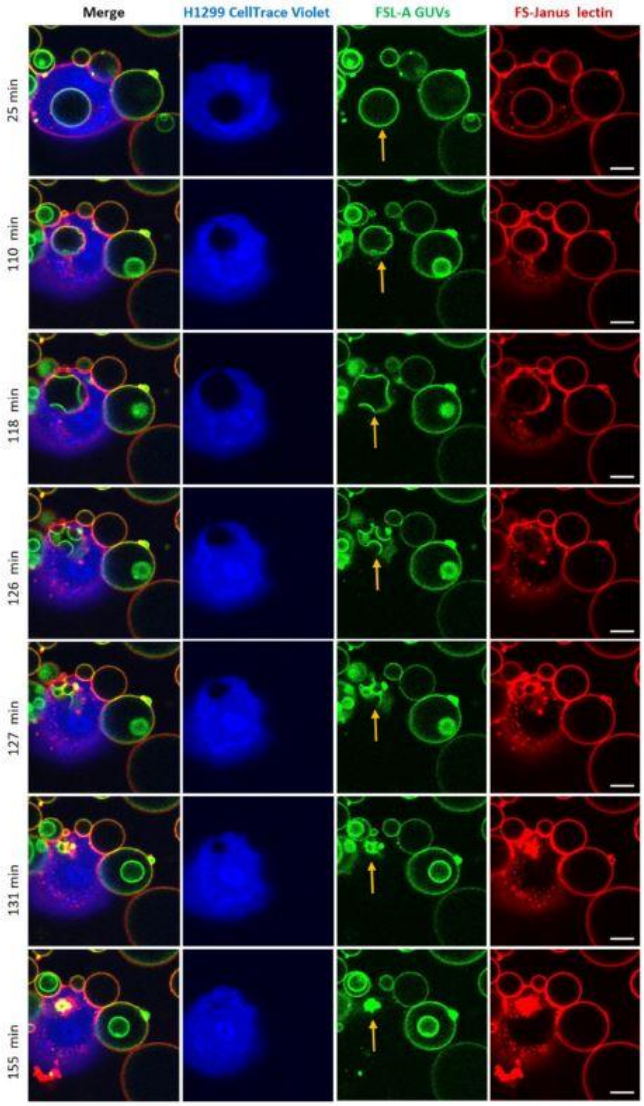


Figure 8. An internalized blood group A decorated GUV undergoes deformations, shrinkage, and bursts inside H1299 cells. FS-Janus lectin (500 nM, red color; labeled with Atto 647) triggered the internalization of complete liposomes (green color; labeled with fluorescent lipid DOPE-Atto 488) into H1299 cells (blue color; partially stained with CellTrace™ Violet). Once internalized, the selected GUV became deformed, reduced its size, and burst. The yellow arrows point to these events. Live-cell imaging experiments were performed at 37 °C by using confocal microscopy and were recorded for 120 min. Scale bars represent 10 µm.

3. Discussion

We successfully demonstrated that FS-Janus lectin induces the intracellular uptake of liposomes via its dual sugar specificity. These findings provide a proof of concept for lectin-mediated targeted drug delivery using liposomes as drug carriers to cells of interest.

Initially, this study showed that FS-Janus lectin induces crosslinking and lipid exchange between fucosylated and sialylated GUVs. In addition, the CBM40 domain, which recognizes sialic acid, drives the binding and internalization of FS-Janus lectin into H1299 cells when the RSL domain, which explicitly recognizes L-fucose, is blocked. These results inspired us to investigate the properties of FS-Janus lectin further.

As described previously, the trivalent assembly of CBM40 domains is very promising as it binds specifically to sialic acid with nanomolar avidity [38,39]. Sialic acid in cancer is an attractive target for therapeutic applications as aberrant sialylation has a significant role in cancer development. Sialoglycans such as sLe^a, sLe^x, STn, and GM2 are present on the surface of cancer cells and contribute extensively to cancer progression by several mechanisms, including immune system evasion and driving tumor growth and metastasis formation [43,44]. Therefore, targeting of sialic acid moieties has grown to be an innovative strategy for cancer therapy as the area for personalized medicine develops rapidly [43,45].

Janus lectin represents the concept of a scaffold with unlimited combinations of lectin modules as it can be engineered to target multiple alterations within the glycome. For instance, a well-documented example of TACA is the globoside Gb3 (also known as CD77 and P^k blood group antigen), overexpressed in breast, ovarian, pancreatic, and colorectal cancer as well as in Burkitt's lymphoma [46]. The glycosphingolipid Gb3 is an attractive therapeutic target, and lectins such as the Shiga toxin B-subunit and LecA from *P. aeruginosa* can serve for lectin-mediated tumor targeting. The replacement of the CBM40 domain in the Janus architecture with StxB or LecA represents another option to test the liposomal delivery of drugs to malignant cells of interest in the future. Furthermore, the Janus lectin approach might open new possibilities in immunotherapy, where the immune system is redirected toward tumor cells. Over the past few decades, researchers have investigated the effects of stimulation of human lymphocytes with lectins. They evaluated the activation and proliferation of B and T cells in response to lectins of different origins [47]. For instance, the Phytohemagglutinin P (PHA-P), a lectin from *Phaseolus vulgaris* (red kidney bean), possesses T cell-activating properties, which have been reported by many authors [47,48]. The idea of a Janus lectin that fuses a T cell-stimulating lectin to a TACA-specific lectin would raise novel opportunities for selective T cell-mediated cytotoxicity.

As our results demonstrate, the FS-Janus lectin can crosslink fucose-decorated vesicles with H1299 cells and becomes enriched at the interfaces. This is encouraging for targeted drug delivery as various drugs can be loaded into vesicles. Certainly, the use of liposomes in medicine over the past few decades has offered significant prospects for effective therapies in a broad range of pathological conditions [49,50]. Lipid-based drug delivery systems have seen a remarkable increase in their application with a plethora of therapeutic compounds and diagnostic agents, described as carriers for chemotherapeutic molecules, gene therapy, and bioactive agents [50].

In addition to crosslinking, incubation of H1299 cells together with fucosylated vesicles and FS-Janus lectin leads to an uptake of liposomal membrane lipids. The uptaken lipids appear to colocalize with FS-Janus lectin inside the cell and accumulate in the perinuclear area over time. The internalization of smaller molecules into cells can be achieved by various endocytic mechanisms such as clathrin- or caveolin-dependent or independent endocytosis. Moreover, during vesicle uptake, we detected an elongation of the cell membrane toward the vesicles (Figure 7). This could indicate the formation of filopodia, which are known to be exhibited in H1299 cells [51]. Filopodia are actin-based membrane protrusions found in various cell types that contribute to cell migration and surface adhesion as well as endocytosis [52,53]. Macrophages, for instance, use filopodia to detect pathogens or apoptotic cells within their environment. By specific phagocytic receptors displayed on the tips of the filopodia, macrophages can bind to the surface of the

pathogen or cell, which subsequently initiates a process of internalization [54]. A recent study demonstrated that filopodia on HeLa cells, a human cervical cancer cell line, serves as a route of endocytic uptake of extracellular vesicles and other nanoparticles [54]. In addition, dynamic clustering and dispersion of lipid rafts at the lamellipodia of myoblasts contributes to both cell-to-cell adhesion and plasma membrane union during the cell-to-cell fusion processes [55]. We, therefore, assumed that FS-Janus lectin bound to the membrane triggered the adhesion and uptake of GUVs.

The processes mentioned above suggest that FS-Janus lectin binding can provoke actin polymerization, leading to protrusions or filopodia formation at the plasma membrane and, subsequently, a pull-in of the liposome into H1299 cells.

Interestingly, when liposomes were functionalized with WGA, a lectin, they were only bound to adenocarcinoma human alveolar basal epithelial (A549) cells, but neither the lipid material nor liposome itself was uptaken. However, when the WGA-liposomes were incubated with oral epithelial cells (healthy), the enhanced uptake of liposomes was observed [56]. Recently, it was shown that multivalent targeting of cancer (K562) cells by core-crosslinked elastin/resilin-like polypeptide micelles led to micelle to cell crosslinking [57]. Additionally, another type of vesicle, the TAT-peptide modified vesicles, was taken up by endocytosis in ovarian carcinoma cells [58].

Strikingly, during experiments conducted with FSL-A-GUVs, we observed a burst and decay of intact vesicles inside the cytosol of cells, accompanied by shape deformation and shrinking of the respective vesicle. This process deserves further investigation, whether the burst was induced due to the nature of FS-Janus lectin alone or combined with the cell cytoskeletal pressure. However, it is known that actin dynamics can also trigger deformation of the liposome membrane [59]. Another hypothesis suggests that during intracellular trafficking, vesicles drive along the microtubules. Furthermore, the cytoskeleton is actively working on organelles, creating tension, and resulting in organelle rupture.

We speculate that the rupture of vesicles can be caused by membrane reorganization and phase separation, responsible for the burst of uptaken liposomes in our experiments. In many cases, but not as a rule, direct cytosolic delivery of drugs loaded in liposomes is favored to prevent the degradation of pharmaceuticals in lysosomes [59].

To summarize, our findings show a prospective proof of concept for lectin-mediated uptake of lipid particles from vesicles and of intact vesicles that even burst inside cancer cells. On the other hand, it must be taken into account that sialylated residues are also exposed on healthy cells. This work does not address the interaction of FS-Janus lectin with non-malignant cells, but possible off-target interactions must be investigated in the future. Furthermore, the precise mechanism of vesicle uptake driven by the FS-Janus lectin has yet to be understood and needs to be addressed in further studies.

Taken together, we show encouraging properties of FS-Janus lectin that promote further investigations regarding future applications, especially in the field of targeted drug delivery to cancer cells.

4. Conclusions

Lectins are arising in many applications in medicine as potential tools for specific targeting and drug delivery. Additionally, to recognize cells of interest, lectin–carbohydrate interactions can trigger vesicular transport into or across epithelial cells. These promising features are applied to the study of the chimeric FS-Janus lectin, which led to the delivery of vesicles to human cancer cells exhibiting hypersialylation. The concept of bifunctional Janus lectin described here is highly versatile since it can be adapted to other glycan epitopes by changing the lectins composing the scaffold. This opens a wide range of applications based on the selective targeting of the glycome in many pathological conditions.

5. Materials and Methods

5.1. Production of Lectins and Labeling

The constructs of FS-Janus lectin and CBM40 were produced in *Escherichia coli* BL21 (DE3). The detailed description of cloning, production, and purification was previously described [38,39]. Briefly, the genes of FS-Janus lectin and CBM40 were subcloned into vectors pET25(+) and pET45(+), respectively. The plasmids were transformed into *Escherichia coli* BL21 (DE3) and cultured in Luria–Bertani (LB) medium supplemented with 100 µg mL^{-1} ampicillin at 37 °C with agitation (180 rpm). Upon reaching OD_{600} of 0.6–0.8, the culture was induced by IPTG and transferred at 16 °C for 20 h with agitation (180 rpm). The cells were harvested by centrifugation and disrupted by a cell disruption system. After the centrifugation of the cell lysate, the supernatant was loaded on affinity columns, i.e., mannose agarose for FS-Janus lectin or nickel sepharose affinity for CBM40. The proteins were dialyzed, and the purity was verified by SDS-PAGE gel electrophoresis.

Lectins were dissolved at 1 mg/mL in Dulbecco's phosphate-buffered saline (PBS) and stored at 4 °C prior usages. For fluorescent labeling, NHS-ester conjugated Alexa Fluor 488, Atto488, Alexa Fluor 647, Atto 488 DOPE, or Atto 647N DOPE (Thermo Fisher Scientific Inc., USA) were used. Fluorescent dyes were dissolved at a final concentration of 1 mg/mL in water-free DMSO (Carl Roth GmbH & Co. KG, Germany), aliquoted, and stored at −20 °C before usage. For the labeling reaction, 200 µL of lectin (1 mg/mL) was supplemented with 20 µL of a 1 M $NaHCO_3$ (pH 8.5) solution. For FS-Janus lectin, the molar ratio between dye and lectin was set to 7:1, and the final ratio was 1.5:1. For di-CBM40, the molar ratio between dye and lectin was set to 3:1, and the final ratio resulted in being 0.5:1. The labeling mixture was incubated at 4 °C for 90 min, and uncoupled dyes were separated using Zeba Spin desalting columns (7k MWCO, 0.5 mL, Thermo Fisher Scientific Inc., USA). Labeled lectins were stored at 4 °C, protected from light.

5.2. Composition and Preparation of GUVs

GUVs were composed of 1,2-dioleoyl-sn-glycero-3- phosphocholine (DOPC), cholesterol (both Avanti Polar Lipids, United States), Atto 647N 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE; Sigma-Aldrich Chemie GmbH, Germany), and either FSL-A(tri) (function-spacer-lipid with blood group A trisaccharide; Sigma-Aldrich Chemie GmbH, Germany) and Ganglioside GM3 from bovine milk (Avanti Polar Lipids, United States) at a molar ratio of 64.7:30:0.3:5 mol%.

GUVs were prepared by the electroformation method as earlier described [58]. Briefly, lipids dissolved in chloroform with a total concentration of 0.5 mg/mL were spread on indium tin oxide (ITO-covered) glass slides and dried in a vacuum for at least one hour or overnight. Two ITO slides were assembled to a chamber filled with sucrose solution adapted to the osmolarity of the imaging buffer of choice, either HBSS (live-cell imaging) or PBS (GUVs only imaging). Then, an alternating electrical field with a field strength of 1 V/mm was implemented for 2.5 h at room temperature. Later, we observed the GUVs in chambers manually built as described [60].

5.3. Cell Culture

The human lung epithelial cell line H1299 (American Type Culture Collection, CRL-5803) was cultured in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% fetal calf serum (FCS) and 4 mM L-glutamine at 37 °C and 5% CO_2, under sterile conditions. Cells were cultivated in standard TC-dishes 100 (Sarstedt AG & Co. KG, Germany) until 90% confluency, detached with trypsin (0.05% trypsin-EDTA solution; Sigma-Aldrich Chemie GmbH, Germany) re-seeded for a subculture or for experiments. For experiments, cells were stimulated with different concentrations of FS-Janus lectin and CBM40 for indicated time points.

5.4. Flow Cytometry Analysis

H1299 cells were detached with 2 mL of 1.5 mM EDTA in PBS (-/-), and 1×10^5 cells were counted and transferred to a U-bottom 96 well plate (Sarstedt AG & Co. KG, Germany). To quantify protein binding to cell surface receptors, cells were incubated with different concentrations of fluorescently labeled FS-Janus lectin and CBM40 solutions for 30 min at 4 °C and protected from light compared to PBS-treated cells as a negative control. For the saturation of fucose-binding sites, 32 nM FS-Janus lectin was pre-incubated with 25, 50, or 100 mM soluble L-fucose, for 30 min at RT, in the absence of light. At the end of pre-incubation, the solution was diluted 100 times and added to cells for 30 min at 4 °C, in the dark. Subsequently, cells were centrifuged at 1600× *g* for 3 min at 4 °C and washed twice with FACS buffer (PBS (-/-) supplemented with 3% FCS *v*/*v*). After the last washing step, the cells were re-suspended with FACS buffer and transferred to FACS tubes (Kisker Biotech GmbH Co. KG, Germany) on ice and protected from light. The fluorescence intensity of treated cells was monitored at FACS Gallios (Beckman Coulter Inc., USA) and further analyzed using FlowJo V.10.5.3.

5.5. Lectin Stimulation and Fluorescence Microscopy

Between 4 and 5×10^4 H1299 cells were seeded on 12-mm glass coverslips in a 4-well plate and allowed to adhere. The next day, cells were stimulated with fluorescently labeled FS-Janus or CBM40 for 30 min at 4 °C, then washed once with PBS and incubated at 37 °C for the indicated time points. Subsequently, cells were fixed with 4% paraformaldehyde for 15 min at RT. The membrane was permeabilized, and cells were blocked by 0.2% Saponin in 3% BSA in PBS (*w*/*v*) for 30 min. Nuclei were counterstained with DAPI (5×10^{-9} g/L), and the samples were mounted on coverslips using Mowiol (containing the anti-bleaching reagent DABCO). Samples were imaged by means of a laser scanning confocal microscope system (Nikon Eclipse Ti-E inverted microscope equipped with Nikon A1R confocal laser scanning system, 60x oil immersion objective, numerical aperture (*NA*) of 1.49, with four lasers: 405 nm, 488 nm, 561 nm, and 640 nm). The images were further analyzed using NIS-Element Confocal 4.20 from Nikon and ImageJ 1.52a from Laboratory for Optical and Computational Instrumentation. A minimum of three biological replicates with ≥20 cells per condition were analyzed.

5.6. Live-Cell Imaging

H1299 cells were stained with a violet CellTrace™ proliferation kit (Thermo Fisher Scientific Inc., USA) one day prior to the experiment. A volume containing a total of 2.5×10^5 cells was transferred to a 15 mL tube and centrifuged (280× *g*, 1 min). After discarding the supernatant, the cell pellet was re-suspended in 1 mL PBS, and 5 µM of violet CellTrace™ proliferation marker in DMSO was added. The solution was incubated for 20 min at 37 °C and 5% CO_2 in the absence of light. The staining solution was diluted with 5 mL of pre-warmed RPMI medium and incubated for 5 min at 37 °C and 5% CO_2, according to the manufacturer's instructions. The cells were pelleted by centrifugation (280× *g*, 1 min), re-suspended in 2 mL of RPMI complete medium, and 0.5 mL of cell solution was transferred to each well of a 4-well glass-bottom dish (CELLview™; Greiner Bio-One International GmbH, Germany). The cells were incubated overnight at 37 °C and 5% CO_2 to ensure adherence. On the day of the experiment, cells were pre-washed with and kept in Hanks' Balanced Salt Solution (HBSS, Thermo Fisher Scientific Inc., USA) supplemented with 1% (*v*/*v*) FCS, 1 mM L-glutamine, 1% (*v*/*v*) NEAA, 10 mM HEPES, 0.55% (*w*/*v*) D-Glucose, while imaging. GUVs were pre-incubated with fluorescently labeled FS-Janus lectin for 5 min and applied to the cells with a final lectin concentration of 200 or 500 nM. GUVs and H1299 cell samples were imaged at 37 °C using an incubator stage (Okolab, Italy) mounted onto a confocal laser scanning microscope (Nikon Eclipse Ti-E, A1R). Image acquisition and processing were conducted using the software NIS-Elements (version 4.5, Nikon).

5.7. Chemical Reagents

The following reagents were obtained from commercial sources: RPMI 1640, PBS (-/-), FCS, and L-glutamine were all purchased from Gibco (Thermo Fisher Scientific Inc., USA). The following chemicals were obtained from Roth: BSA, DABCO, DAPI, EDTA, Mowiol, NH_4Cl, paraformaldehyde.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/toxins13110792/s1, Figure S1: Examples of tumor-associated carbohydrate antigens. Figure S2: Non-chimeric components of FS-Janus lectin—trimeric RSL and dimeric di-CBM40 do not crosslink GUVs containing different glycospecies. Figure S3: Binding of trimeric RSL and dimeric CBM40 to H1299 in the presence of soluble sugars. Figure S4: Dose-dependent binding and intracellular uptake of RSL and FS-Janus lectin treated with 3´-sialyllactose in H1299 lung epithelial cells. Figure S5: The FS-Janus lectin accumulates at the perinuclear area of H1299. Figure S6: Control experiment of FSL-A GUVs and H1299 cells, co-incubated together without FS-Janus lectin. Video S1: FS-Janus lectin triggers crosslinking and lipid exchange between GM3- and blood group A trisaccharide-functionalized liposomes. Video S2: Control group of liposomes without the addition of FS-Janus lectin. Video S3: Control group without FS-Janus lectin for the crosslinking experiment of blood group A trisaccharide-decorated GUVs and H1299 cells. Video S4: Live-cell imaging of crosslinking of blood group A trisaccharide-decorated GUVs and H1299 cells is mediated by FS-Janus lectin, enriched at the interfaces. Video S5: Kinetics of the cellular uptake and perinuclear co-accumulation of liposomal lipids and FS-Janus lectin. Video S6: Full uptake of blood group A-decorated liposome into H1299 cells. Video S7: An internalized blood group, A trisaccharide-decorated GUV, undergoes deformations, shrinkage, and bursts inside H1299 cells.

Author Contributions: Conceptualization of the study, L.S., F.R., and W.R.; Methodology, L.S., F.R., A.M., P.F.M., K.K., and S.N.; Software, L.S. and F.R.; Formal analysis, L.S., F.R., and A.M.; Investigation, L.S., F.R., A.M., and K.K.; Data curation, L.S., F.R., and A.M.; Writing—original draft preparation, L.S. and F.R.; Writing—review and editing, L.S., F.R., P.F.M., A.I., and W.R.; Supervision, W.R.; Project administration, W.R.; Funding acquisition, A.I. and W.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement synBIOcarb (No. 814029). Moreover, WR acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC-294 and EXC-2189) and the German Research Foundation grants RTG 2202 (Transport across and into membranes) and Major Research Instrumentation (project number: 438033605); the Ministry for Science, Research and Arts of the State of Baden-Württemberg (Az: 33-7532.20); and by the Freiburg Institute for Advanced Studies (FRIAS). AI acknowledges support from Glyco@Alps (ANR-15-IDEX-02) and Labex Arcane/CBH- EUR-GS (ANR-17-EURE-0003). This publication is partially based upon work from COST Action CA18103 (INNOGLY), supported by COST (European Cooperation in Science and Technology).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Acknowledgments: The article processing charge was mostly funded by the Baden-Wuerttemberg Ministry of Science, Research and Art and the University of Freiburg in the funding programme Open Access Publishing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mody, R.; Joshi, S.H.; Chaney, W. Use of Lectins as Diagnostic and Therapeutic Tools for Cancer. *J. Pharmacol. Toxicol. Methods* **1995**, *33*, 1–10. [CrossRef]
2. Gorelik, E.; Galili, U.; Raz, A. On the Role of Cell Surface Carbohydrates and Their Binding Proteins (Lectins) in Tumor Metastasis. *Cancer Metastasis Rev.* **2001**, *20*, 245–277. [CrossRef] [PubMed]
3. Bies, C.; Lehr, C.-M.; Woodley, J.F. Lectin-Mediated Drug Targeting: History and Applications. *Adv. Drug Deliv. Rev.* **2004**, *56*, 425–435. [CrossRef]
4. Minko, T. Drug Targeting to the Colon with Lectins and Neoglycoconjugates. *Adv. Drug Deliv. Rev.* **2004**, *56*, 491–509. [CrossRef] [PubMed]
5. Rini, J.M. Lectin Structure. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 551–577. [CrossRef]
6. Lis, H.; Sharon, N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev.* **1998**, *98*, 637–674. [CrossRef]
7. Sharon, N. Lectins: Carbohydrate-Specific Reagents and Biological Recognition Molecules. *J. Biol. Chem.* **2007**, *282*, 2753–2764. [CrossRef]
8. Ambrosi, M.; Cameron, N.R.; Davis, B.G. Lectins: Tools for the Molecular Understanding of the Glycocode. *Org. Biomol. Chem.* **2005**, *3*, 1593–1608. [CrossRef]
9. Weis, W.I.; Drickamer, K. Structural Basis of Lectin-Carbohydrate Recognition. *Annu. Rev. Biochem.* **1996**, *65*, 441–473. [CrossRef]
10. Collins, B.E.; Paulson, J.C. Cell Surface Biology Mediated by Low Affinity Multivalent Protein–Glycan Interactions. *Curr. Opin. Chem. Biol.* **2004**, *8*, 617–625. [CrossRef]
11. Ghazarian, H.; Idoni, B.; Oppenheimer, S.B. A Glycobiology Review: Carbohydrates, Lectins and Implications in Cancer Therapeutics. *Acta Histochem.* **2011**, *113*, 236–247. [CrossRef]
12. Ohtsubo, K.; Marth, J.D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126*, 855–867. [CrossRef]
13. Imberty, A.; Varrot, A. Microbial Recognition of Human Cell Surface Glycoconjugates. *Curr. Opin. Struct. Biol.* **2008**, *18*, 567–576. [CrossRef]
14. Brandley, B.K.; Schnaar, R.L. Cell-Surface Carbohydrates in Cell Recognition and Response. *J. Leukoc. Biol.* **1986**, *40*, 97–111. [CrossRef]
15. Eierhoff, T.; Bastian, B.; Thuenauer, R.; Madl, J.; Audfray, A.; Aigal, S.; Juillot, S.; Rydell, G.E.; Muller, S.; de Bentzmann, S.; et al. A Lipid Zipper Triggers Bacterial Invasion. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 12895–12900. [CrossRef]
16. Youan, B.-B.C.; Coulibaly, F.S. Current Status of Lectin-Based Cancer Diagnosis and Therapy. *AIMS Mol. Sci.* **2017**, *4*, 1–27. [CrossRef]
17. Müller, S.K.; Wilhelm, I.; Schubert, T.; Zittlau, K.; Imberty, A.; Madl, J.; Eierhoff, T.; Thuenauer, R.; Römer, W. Gb3-Binding Lectins as Potential Carriers for Transcellular Drug Delivery. *Expert Opin. Drug Deliv.* **2017**, *14*, 141–153. [CrossRef]
18. Varki, A.; Kannagi, R.; Toole, B.P. Glycosylation Changes in Cancer. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2009.
19. Pinho, S.S.; Reis, C.A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* **2015**, *15*, 540–555. [CrossRef]
20. Rodrigues, J.G.; Balmaña, M.; Macedo, J.A.; Poças, J.; Fernandes, Â.; de-Freitas-Junior, J.C.M.; Pinho, S.S.; Gomes, J.; Magalhães, A.; Gomes, C.; et al. Glycosylation in Cancer: Selected Roles in Tumour Progression, Immune Modulation and Metastasis. *Cell. Immunol.* **2018**, *333*, 46–57. [CrossRef] [PubMed]
21. Odumosu, O.; Nicholas, D.; Yano, H.; Langridge, W. AB Toxins: A Paradigm Switch from Deadly to Desirable. *Toxins* **2010**, *2*, 1612–1645. [CrossRef] [PubMed]
22. Johannes, L.; Römer, W. Shiga Toxins—from Cell Biology to Biomedical Applications. *Nat. Rev. Microbiol.* **2010**, *8*, 105–116. [CrossRef]
23. Guan, J.; Zhang, Z.; Hu, X.; Yang, Y.; Chai, Z.; Liu, X.; Liu, J.; Gao, B.; Lu, W.; Qian, J.; et al. Cholera Toxin Subunit B Enabled Multifunctional Glioma-Targeted Drug Delivery. *Adv. Healthc. Mater.* **2017**, *6*, 1700709. [CrossRef] [PubMed]
24. Fort, P.; Sakai, K.; Luppi, P.-H.; Salvert, D.; Jouvet, M. Monoaminergic, Peptidergic, and Cholinergic Afferents to the Cat Facial Nucleus as Evidenced by a Double Immunostaining Method with Unconjugated Cholera Toxin as a Retrograde Tracer. *J. Comp. Neurol.* **1989**, *283*, 285–302. [CrossRef] [PubMed]
25. Lima, D.; Coimbra, A. Morphological Types of Spinomesencephalic Neurons in the Marginal Zone (Lamina I) of the Rat Spinal Cord, as Shown after Retrograde Labelling with Cholera Toxin Subunit B. *J. Comp. Neurol.* **1989**, *279*, 327–339. [CrossRef] [PubMed]
26. Luppi, P.-H.; Fort, P.; Jouvet, M. Iontophoretic Application of Unconjugated Cholera Toxin B Subunit (CTb) Combined with Immunohistochemistry of Neurochemical Substances: A Method for Transmitter Identification of Retrogradely Labeled Neurons. *Brain Res.* **1990**, *534*, 209–224. [CrossRef]
27. Luppi, P.-H.; Sakai, K.; Salvert, D.; Fort, P.; Jouvet, M. Peptidergic Hypothalamic Afferents to the Cat Nucleus Raphe Pallidus as Revealed by a Double Immunostaining Technique Using Unconjugated Cholera Toxin as a Retrograde Tracer. *Brain Res.* **1987**, *402*, 339–345. [CrossRef]
28. Mo, Y.; Lim, L.-Y. Mechanistic Study of the Uptake of Wheat Germ Agglutinin-Conjugated PLGA Nanoparticles by A549 Cells. *J. Pharm. Sci.* **2004**, *93*, 20–28. [CrossRef]

29. Mo, Y.; Lim, L. Preparation and in Vitro Anticancer Activity of Wheat Germ Agglutinin (WGA)-Conjugated PLGA Nanoparticles Loaded with Paclitaxel and Isopropyl Myristate. *J. Control. Release* **2005**, *107*, 30–42. [CrossRef] [PubMed]
30. Winter, H.C.; Mostafapour, K.; Goldstein, I.J. The Mushroom Marasmius Oreades Lectin Is a Blood Group Type B Agglutinin That Recognizes the Galα1,3Gal and Galα1,3Galβ1,4GlcNAc Porcine Xenotransplantation Epitopes with High Affinity. *J. Biol. Chem.* **2002**, *277*, 14996–15001. [CrossRef]
31. Terada, D.; Kawai, F.; Noguchi, H.; Unzai, S.; Hasan, I.; Fujii, Y.; Park, S.-Y.; Ozeki, Y.; Tame, J.R.H. Crystal Structure of MytiLec, a Galactose-Binding Lectin from the Mussel Mytilus Galloprovincialis with Cytotoxicity against Certain Cancer Cell Types. *Sci. Rep.* **2016**, *6*, 28344. [CrossRef]
32. Dobie, C.; Skropeta, D. Insights into the Role of Sialylation in Cancer Progression and Metastasis. *Br. J. Cancer* **2021**, *124*, 76–90. [CrossRef]
33. Büll, C.; Stoel, M.A.; den Brok, M.H.; Adema, G.J. Sialic Acids Sweeten a Tumor's Life. *Cancer Res.* **2014**, *74*, 3199–3204. [CrossRef]
34. Peixoto, A.; Relvas-Santos, M.; Azevedo, R.; Santos, L.L.; Ferreira, J.A. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. *Front. Oncol.* **2019**, *9*, 380. [CrossRef]
35. Chen, K.; Blixt, O.; Wandall, H.H. Mucins as biomarkers in cancer. In *Mucins and Cancer*; Future Medicine Ltd.: London, UK, 2013.
36. Drake, P.M.; Cho, W.; Li, B.; Prakobphol, A.; Johansen, E.; Anderson, N.L.; Regnier, F.E.; Gibson, B.W.; Fisher, S.J. Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation. *Clin. Chem.* **2010**, *56*, 223–236. [CrossRef] [PubMed]
37. Vajaria, B.N.; Patel, K.R.; Begum, R.; Patel, P.S. Sialylation: An Avenue to Target Cancer Cells. *Pathol. Oncol. Res.* **2016**, *22*, 443–447. [CrossRef]
38. Ribeiro, J.P.; Pau, W.; Pifferi, C.; Renaudet, O.; Varrot, A.; Mahal, L.K.; Imberty, A. Characterization of a High-Affinity Sialic Acid-Specific CBM40 from Clostridium Perfringens and Engineering of a Divalent Form. *Biochem. J.* **2016**, *473*, 2109–2118. [CrossRef] [PubMed]
39. Ribeiro, J.P.; Villringer, S.; Goyard, D.; Coche-Guerente, L.; Höferlin, M.; Renaudet, O.; Römer, W.; Imberty, A. Tailor-Made Janus Lectin with Dual Avidity Assembles Glycoconjugate Multilayers and Crosslinks Protocells. *Chem. Sci.* **2018**, *9*, 7634–7641. [CrossRef] [PubMed]
40. Kostlánová, N.; Mitchell, E.P.; Lortat-Jacob, H.; Oscarson, S.; Lahmann, M.; Gilboa-Garber, N.; Chambat, G.; Wimmerová, M.; Imberty, A. The Fucose-Binding Lectin from Ralstonia Solanacearum. *J. Biol. Chem.* **2005**, *280*, 27839–27849. [CrossRef]
41. Omidvar, R.; Römer, W. Glycan-Decorated Protocells: Novel Features for Rebuilding Cellular Processes. *Interface Focus* **2019**, *9*, 20180084. [CrossRef] [PubMed]
42. Yuan, Q.; Chen, X.; Han, Y.; Lei, T.; Wu, Q.; Yu, X.; Wang, L.; Fan, Z.; Wang, S. Modification of A2,6-Sialylation Mediates the Invasiveness and Tumorigenicity of Non-Small Cell Lung Cancer Cells In Vitro and In Vivo via Notch1/Hes1/MMPs Pathway. *Int. J. Cancer* **2018**, *143*, 2319–2330. [CrossRef]
43. Büll, C.; Boltje, T.J.; Balneger, N.; Weischer, S.M.; Wassink, M.; van Gemst, J.J.; Bloemendal, V.R.L.J.; Boon, L.; van der Vlag, J.; Heise, T.; et al. Sialic Acid Blockade Suppresses Tumor Growth by Enhancing T Cell-Mediated Tumor Immunity. *Cancer Res.* **2018**, *78*, 3574–3588. [CrossRef] [PubMed]
44. Munkley, J.; Scott, E. Targeting Aberrant Sialylation to Treat Cancer. *Medicines* **2019**, *6*, 102. [CrossRef]
45. Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C.A. Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? *Cancer Cell* **2019**, *36*, 6–16. [CrossRef]
46. Đevenica, D.; Čikeš Čulić, V.; Vuica, A.; Markotić, A. Biochemical, Pathological and Oncological Relevance of Gb3Cer Receptor. *Med Oncol.* **2011**, *28*, 675–684. [CrossRef]
47. Miller, K. The Stimulation of Human B and T Lymphocytes by Various Lectins. *Immunobiology* **1983**, *165*, 132–146. [CrossRef]
48. Morgan, D.; Ruscetti, F.; Gallo, R. Selective in Vitro Growth of T Lymphocytes from Normal Human Bone Marrows. *Science* **1976**, *193*, 1007–1008. [CrossRef]
49. Sercombe, L.; Veerati, T.; Moheimani, F.; Wu, S.Y.; Sood, A.K.; Hua, S. Advances and Challenges of Liposome Assisted Drug Delivery. *Front. Pharmacol.* **2015**, *6*, 286. [CrossRef]
50. Hua, S.; Wu, S.Y. The Use of Lipid-Based Nanocarriers for Targeted Pain Therapies. *Front. Pharmacol.* **2013**, *4*, 143. [CrossRef]
51. Yamada, H.; Takeda, T.; Michiue, H.; Abe, T.; Takei, K. Actin Bundling by Dynamin 2 and Cortactin Is Implicated in Cell Migration by Stabilizing Filopodia in Human Non-Small Cell Lung Carcinoma Cells. *Int. J. Oncol.* **2016**, *49*, 877–886. [CrossRef]
52. Gallop, J.L. Filopodia and Their Links with Membrane Traffic and Cell Adhesion. *Semin. Cell Dev. Biol.* **2020**, *102*, 81–89. [CrossRef] [PubMed]
53. Vonna, L.; Wiedemann, A.; Aepfelbacher, M.; Sackmann, E. Micromechanics of Filopodia Mediated Capture of Pathogens by Macrophages. *Eur. Biophys. J.* **2007**, *36*, 145–151. [CrossRef] [PubMed]
54. Aderem, A.; Underhill, D.M. Mechanisms of Phagocytosis in Macrophages. *Annu. Rev. Immunol.* **1999**, *17*, 593–623. [CrossRef] [PubMed]
55. Mukai, A.; Kurisaki, T.; Sato, S.B.; Kobayashi, T.; Kondoh, G.; Hashimoto, N. Dynamic Clustering and Dispersion of Lipid Rafts Contribute to Fusion Competence of Myogenic Cells. *Exp. Cell Res.* **2009**, *315*, 3052–3063. [CrossRef] [PubMed]
56. Brück, A.; Abu-dahab, R.; Borchard, G.; Schäfer, U.F.; Lehr, C.-M. Lectin-Functionalized Liposomes for Pulmonary Drug Delivery: Interaction with Human Alveolar Epithelial Cells. *J. Drug Target.* **2001**, *9*, 241–251. [CrossRef]

57. Weber, P.; Dzuricky, M.; Min, J.; Jenkins, I.; Chilkoti, A. Concentration-Independent Multivalent Targeting of Cancer Cells by Genetically Encoded Core-Crosslinked Elastin/Resilin-like Polypeptide Micelles. *Biomacromolecules* **2021**, *22*, 4347–4356. [CrossRef] [PubMed]
58. Fretz, M.M.; Koning, G.A.; Mastrobattista, E.; Jiskoot, W.; Storm, G. OVCAR-3 Cells Internalize TAT-Peptide Modified Liposomes by Endocytosis. *Biochim. Biophys. Acta Biomembr.* **2004**, *1665*, 48–56. [CrossRef]
59. Simon, C.; Kusters, R.; Caorsi, V.; Allard, A.; Abou-Ghali, M.; Manzi, J.; di Cicco, A.; Lévy, D.; Lenz, M.; Joanny, J.-F.; et al. Actin Dynamics Drive Cell-like Membrane Deformation. *Nat. Phys.* **2019**, *15*, 602–609. [CrossRef]
60. Madl, J.; Villringer, S.; Römer, W. Delving into Lipid-Driven Endocytic Mechanisms Using Biomimetic Membranes. In *Chemical and Synthetic Approaches in Membrane Biology*; Humana Press: New York, NY, USA, 2016.

toxins

Supplementary Materials: The Two Sweet Sides of Janus Lectin Drive Crosslinking of Liposomes to Cancer Cells and Material Uptake

Lina Siukstaite, Francesca Rosato, Anna Mitrovic, Peter Fritz Müller, Katharina Kraus, Simona Notova, Anne Imberty and Winfried Römer

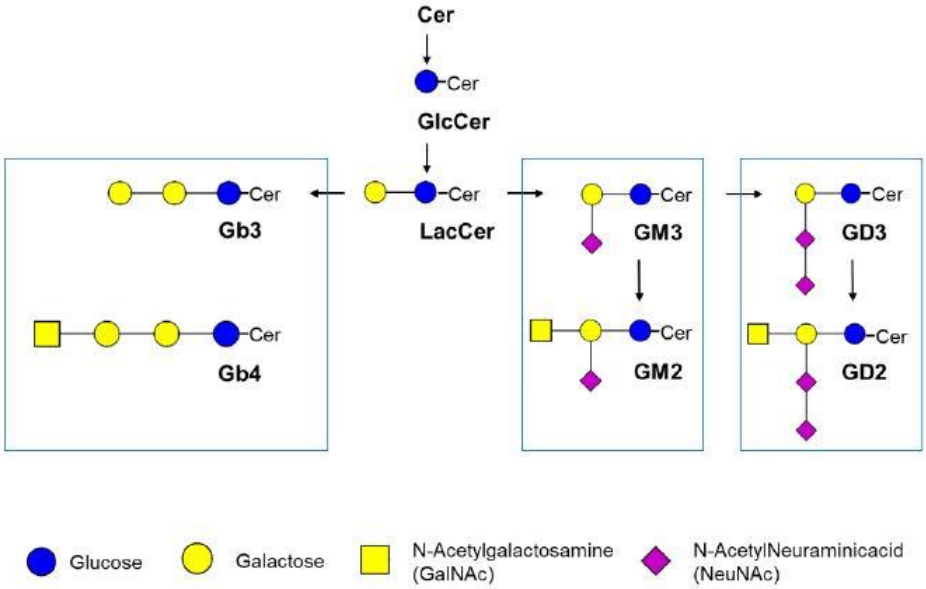


Figure S1. Examples of tumor-associated carbohydrate antigens. Schematic representation of glycolipids found in several malignant tumors.

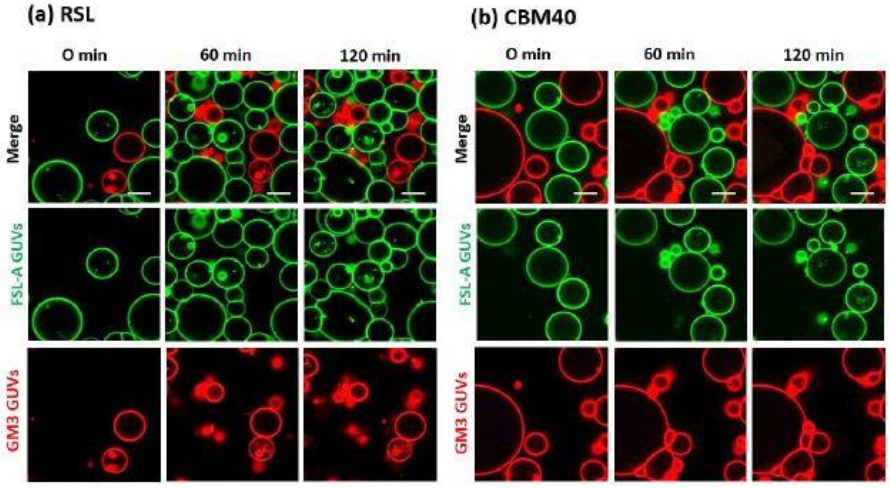


Figure S2. Non-chimeric components of FS-Janus lectin - trimeric RSL and dimeric di-CBM40 do not crosslink GUVs containing different glycospecies. (a) Trimeric RSL (200 nM) does not crosslink FSL-A- and GM3-containing GUVs. (b) Dimeric di-CBM40 (200 nM) does not crosslink GM3-GUVs with FSL-A-containing GUVs. GM3-GUVs (red color; labeled with the fluorescent lipid DOPE-Atto 647N) and FSL-A-GUVs (green color; labeled with the fluorescent lipid DOPE-Atto 488) were incubated with 200 nM of either RSL or CBM40 (unlabeled) at room temperature and were monitored for 120 min using confocal microscopy. The scale bars represent 10 μm.

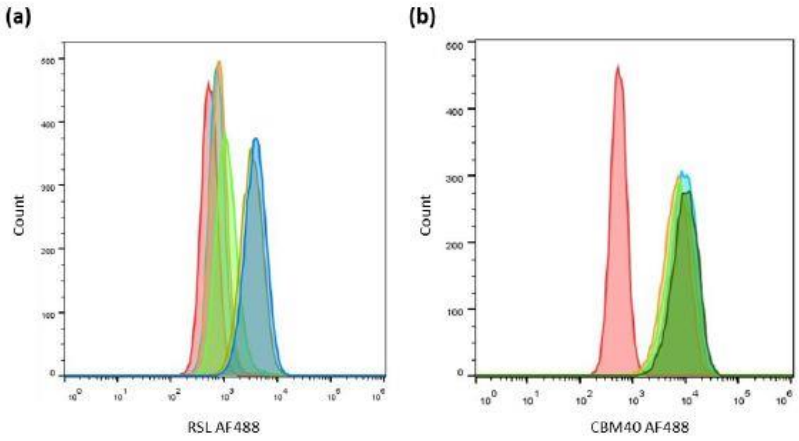


Figure S3. Binding of trimeric RSL and dimeric CBM40 to H1299 in presence of soluble sugars. Flow cytometry analysis of gated living H1299 cells incubated for 30 min at 4 °C with RSL and CBM40 (a, b). (a) Histogram of fluorescence intensity of H1299 cells stimulated with 0.2 μM of RSL AF488 pre-incubated with increasing concentrations of soluble L-fucose for 30 min at RT (red: negative control, light blue: RSL + 100 mM L-fucose, orange: RSL + 50 mM L-fucose, light green: RSL + 25 mM L-fucose, green: RSL + 10 mM L-fucose, blue: RSL, untreated). In presence of 100 mM L-fucose, the binding of RSL to H1299 is remarkably reduced and nearly to zero. The 25, 50 and 100 mM L-fucose concentrations were used to saturate FS-Janus fucose-binding sites. (b) Histogram of fluorescence intensity of H1299 cells stimulated with 42 nM of CBM40 AF488 pre-incubated with increasing concentrations of 3′-sialyllactose (3′-SL) for 30 min at RT (red: negative control, light blue: CBM40 + 50 mM 3′-SL, orange: CBM40 + 25 mM 3′-SL, light green: CBM40 + 10 mM 3′-SL, green: CBM40, untreated). A total blocking of CBM40 binding to H1299 cells was not achieved in the reported range of inhibitor concentrations. Higher concentrations of 3′-sialyllactose were not sufficient to fully block protein binding to cell surface (not shown). The 25 and 50 mM 3′-sialyllactose concentrations were used to saturate FS-Janus sialic acid-binding sites.

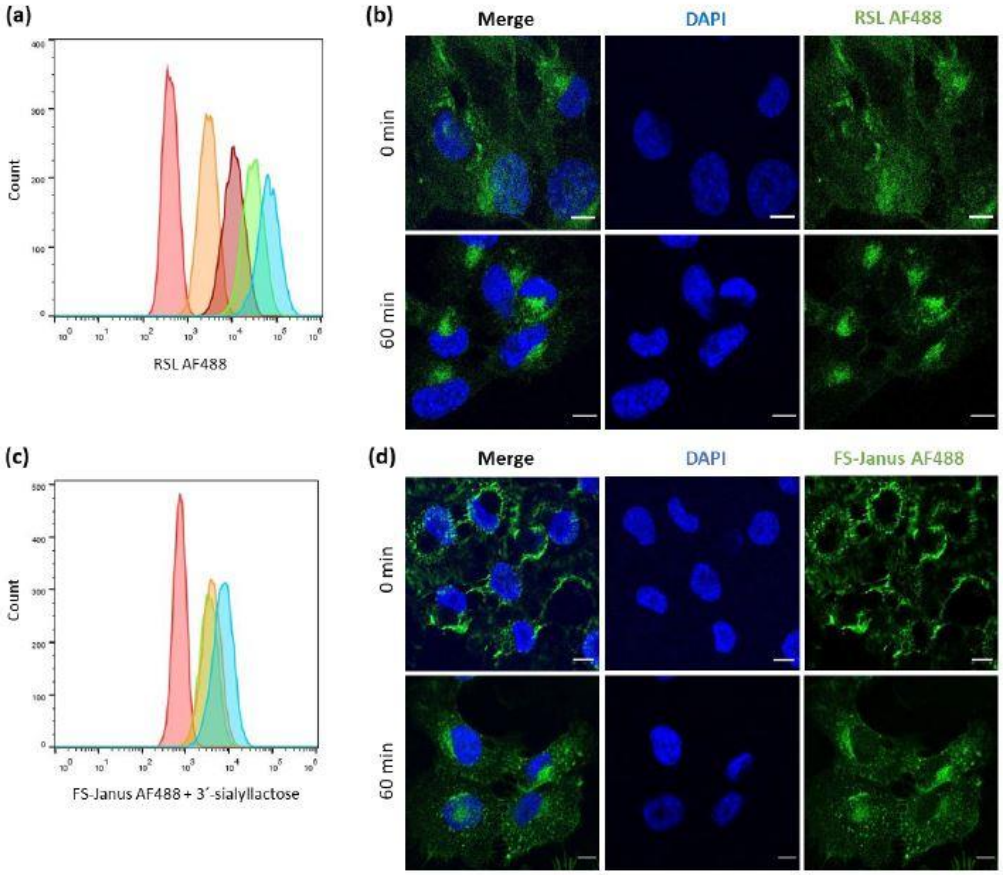


Figure S4. Dose-dependent binding and intracellular uptake of RSL and FS-Janus lectin treated with 3´-sialyllactose in H1299 lung epithelial cells. Flow cytometry analysis of gated living H1299 cells incubated for 30 min at 4°C with RSL (a) and FS-Janus lectin + 3´-SL (c), confocal imaging of RSL (green color) binding (0 min) and cellular uptake (after 60 min of stimulation) (b) and confocal imaging of FS-Janus lectin + 3´-SL at 0 min and after internalization (60 min) (d). (a) Histogram of fluorescence intensity of H1299 cells stimulated with different concentrations of RSL AF488 (red: negative control, orange: 51 nM, dark red: 0.2 µM, green: 0.5 µM, blue: 10 µM). (b) Confocal imaging of fluorescently labelled RSL (green color) incubated with H1299 cells at different time points. Cells were pre-incubated with RSL AF488 (0.62 µM) for 30 min on ice. Subsequently, the lectin solution was washed away and replaced by complete medium, and cells were incubated at 37 °C for 60 min. RSL is internalized within 60 min and accumulates in the perinuclear area. Nuclei were counterstained by DAPI. Scale bars represent 10 µm. (c) Histogram of fluorescence intensity of H1299 cells stimulated with FS-Janus AF488 incubated with different concentrations of soluble 3´-sialyllactose for 30 min at RT, to saturate the CBM40 carbohydrate-binding sites (red: negative control, green: FS-Janus lectin 32 nM + 50 mM 3´-SL, orange: FS-Janus lectin 32 nM + 25 mM 3´-SL, blue: FS-Janus lectin 32 nM).). (d) Confocal imaging of fluorescently labeled FS-Janus lectin (green color) incubated with H1299 cells at different time points. FS-Janus AF488 (150 nM) was pre-incubated with soluble 3´-sialyllactose (100 mM) for 30 min at RT, and the solution was added to cells. Cells were treated as described in (b). In presence of 3´-sialyllactose, FS-Janus lectin is internalized within 60 min and accumulates in the perinuclear region of H1299. Nuclei were counterstained by DAPI. Scale bars represent 10 µm.

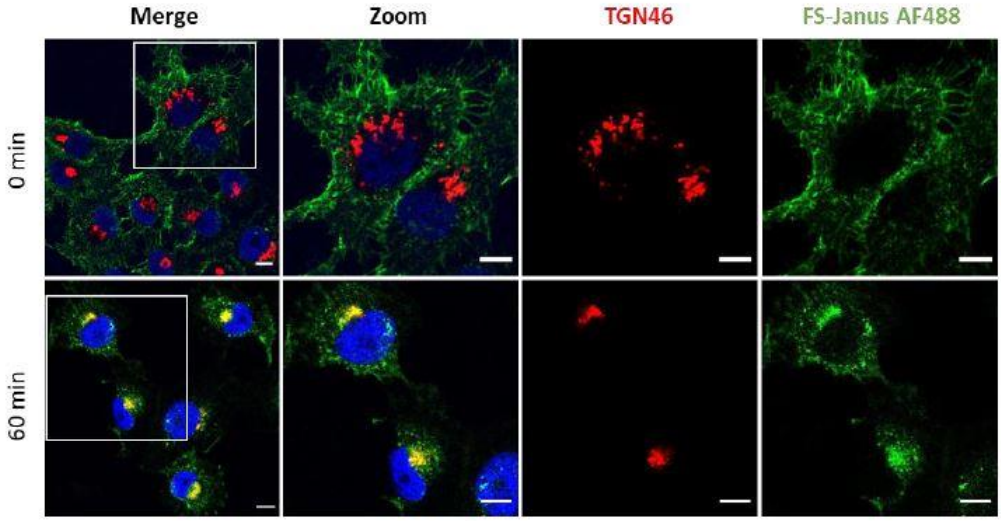


Figure S5. The FS-Janus lectin accumulates at the perinuclear area of H1299. Fluorescence co-localization studies of FS-Janus lectin (green) and *trans*-Golgi network marker TGN46 (red) after 60 min of lectin stimulation. Cells were pre-incubated with FS-Janus AF488 (150 nM) for 30 min on ice. Subsequently, the lectin solution was washed away and replaced by complete medium, and cells were incubated at 37 °C for 60 min. FS-Janus is internalized within 60 min and accumulates at the perinuclear area. Fixed cells were incubated with primary antibody (1:100) for 1 h at RT. After three washes, cells were stained with fluorescently labelled secondary antibody (1:200) for 30 min at RT in the dark. Nuclei were counterstained by DAPI. Scale bars represent 10 µm.

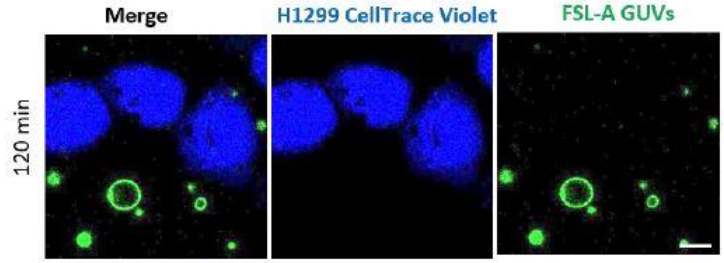


Figure S6. Control experiment of FSL-A GUVs and H1299 cells, co-incubated together without FS-Janus lectin. After 120 minutes of incubation, FSL-A GUVs do not interact visibly with H1299 cells. The GUVs (green color) are labelled with the fluorescent lipid DOPE-Atto488, the H1299 cells (in blue color) are partially stained with CellTrace™ Violet. Scale bars represent 10 µm.

Supplementary Videos

Video S1. FS-Janus lectin triggers crosslinking and lipid exchange between GM3- (red color; labelled with the fluorescent lipid DOPE-Atto 647N) and blood group A trisaccharide- (green color; labelled with the fluorescent lipid DOPE-Atto 488) functionalized liposomes. The crosslinking between GUVs starts immediately after the addition of FS-Janus lectin (200 nM, unlabelled), and lipid exchange becomes apparent during 2 hours of incubation. The scale bar represents 10 µm.

Video S2. Liposome dynamics without the addition of FS-Janus lectin. The GM3- (red color; labelled with the fluorescent lipid DOPE-Atto 647N) and blood group A trisaccharide- (green color; labelled with the fluorescent lipid DOPE-Atto 488) functionalized liposomes do not crosslink and keep their round shape over the entire incubation time. The scale bar represents 10 µm.

Video S3. Crosslinking experiment of blood group A trisaccharide-decorated GUVs and H1299 cells without the addition of FS-Janus lectin. The GUVs (red color) are labelled with the fluorescent lipid DOPE-Atto647N, the H1299 cells (in blue color) are partially stained with CellTrace™ Violet. Live-cell imaging experiments were performed at 37°C by using confocal microscopy and were recorded for 120 min. Scale bar represent 10 μm.

Video S4. Crosslinking of blood group A trisaccharide-decorated GUVs and H1299 cells, which is mediated by FS-Janus lectin (200 nM). The GUVs (red color) are labelled with the fluorescent lipid DOPE-Atto647N, the H1299 cells (in blue color) are partially stained with CellTrace™ Violet and FS-Janus lectin (green color) is labelled with AF488. FS-Janus lectin is enriched at the interfaces. Live-cell imaging experiments were performed at 37°C by using confocal microscopy and were recorded for 120 min. Scale bar represents 10 μm.

Video S5. Kinetics of the cellular uptake and perinuclear co-accumulation of liposomal lipids and FS-Janus lectin. H1299 cells (blue color; partially stained with CellTrace™ Violet) were incubated over in total of three hours in a live-imaging setting with 200 nM FS-Janus lectin (red color, labelled with AF647) and blood group A trisaccharide-decorated GUVs (green color; labeled with DOPE-Atto 488) at 37°C. Scale bar represents 10 μm.

Video S6. Full uptake of blood group A-decorated liposome into H1299 cells. Live-cell imaging experiments were performed at 37°C by using confocal microscopy, and the kinetics of uptake were recorded for 120 minutes. FS-Janus lectin (200 nM, red color; labelled with AF647) triggers the internalization of complete liposomes (green color; labelled with DOPE-Atto 488) into H1299 cells (blue color; partially stained with CellTrace™ Violet). Inside the cell, the liposome remains covered by FS-Janus lectin. Scale bar represents 10 μm.

Video S7. An internalized blood group A trisaccharide-decorated GUV undergoes deformations, shrinkage, and bursts inside H1299 cells. Live-cell imaging experiments were performed at 37°C by using confocal microscopy over 120 minutes. FS-Janus lectin (500 nM, red color; labelled with AF647) triggered the internalization of complete liposomes (green color; labelled with DOPE-Atto 488) into H1299 cells (blue color; partially stained with CellTrace™ Violet). Once internalized, the selected GUV got deformed, reduced its size and bursted. Scale bar represents 10 μm.

9. BIBLIOGRAPHY

Achilli, S., Monteiro, J. T., Serna, S., Mayer-Lambertz, S., Thépaut, M., Roy, A. Le, Ebel, C., Reichardt, N. C., Lepenies, B., Fieschi, F., & Vivès, C. (2020). Tetralec, artificial tetrameric lectins: A tool to screen ligand and pathogen interactions. *International Journal of Molecular Sciences*, *21*(15), 1–20. https://doi.org/10.3390/ijms21155290

Adam, J., Pokorná, M., Sabin, C., Mitchell, E. P., Imberty, A., & Wimmerová, M. (2007). Engineering of PA-IIL lectin from Pseudomonas aeruginosa - Unravelling the role of the specificity loop for sugar preference. *BMC Structural Biology*, *7*, 1–13. https://doi.org/10.1186/1472-6807-7-36

Agrawal, B. B., & Goldstein, I. J. (1965). Specific binding of concanavalin A to cross-linked dextran gels. *The Biochemical Journal*. https://doi.org/10.1042/bj0960023C

Angeloni, S., Ridet, J. L., Kusy, N., Gao, H., Crevoisier, F., Guinchard, S., Kochhar, S., Sigrist, H., & Sprenger, N. (2005). Glycoprofiling with micro-arrays of glycoconjugates and lectins. *Glycobiology*, *15*(1). https://doi.org/10.1093/glycob/cwh143

Antos, J. M., Chew, G. L., Guimaraes, C. P., Yoder, N. C., Grotenbreg, G. M., Popp, M. W. L., & Ploegh, H. L. (2009). Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *Journal of the American Chemical Society*, *131*(31), 10800–10801. https://doi.org/10.1021/ja902681k

Arai, R., Kobayashi, N., Kimura, A., Sato, T., Matsuo, K., Wang, A. F., Platt, J. M., Bradley, L. H., & Hecht, M. H. (2012). Domain-swapped dimeric structure of a stable and functional de Novo Four-Helix bundle protein, WA20. *Journal of Physical Chemistry B*, *116*(23). https://doi.org/10.1021/jp212438h

Arnaud, J., Claudinon, J., Tröndle, K., Trovaslet, M., Larson, G., Thomas, A., Varrot, A., Römer, W., Imberty, A., & Audfray, A. (2013). Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity. *ACS Chemical Biology*, *8*(9), 1918–1924. https://doi.org/10.1021/cb400254b

Arnaud, J., Tröndle, K., Claudinon, J., Audfray, A., Varrot, A., Römer, W., & Imberty, A. (2014). Membrane deformation by neolectins with engineered glycolipid binding sites. *Angewandte Chemie - International Edition*, *53*(35), 9267–9270. https://doi.org/10.1002/anie.201404568

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., … Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1100

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. In *Nucleic Acids Research* (Vol. 28, Issue 1). https://doi.org/10.1093/nar/28.1.235

Bojar, D., Meche, L., Meng, G., Eng, W., Smith, D. F., Cummings, R. D., & Mahal, L. K. (2022). A Useful Guide to Lectin Binding: Machine-Learning Directed Annotation of 57 Unique Lectin Specificities. *ACS Chemical Biology*. https://doi.org/10.1021/acschembio.1c00689

Bonnardel, F., Kumar, A., Wimmerova, M., Lahmann, M., Perez, S., Varrot, A., Lisacek, F., & Imberty, A. (2019). Architecture and Evolution of Blade Assembly in β-propeller Lectins. *Structure*, *27*(5), 764-775.e3. https://doi.org/10.1016/j.str.2019.02.002

Bonnardel, F., Mariethoz, J., Pérez, S., Imberty, A., & Lisacek, F. (2021). LectomeXplore, an update of UniLectin for the discovery of carbohydrate-binding proteins based on a new lectin classification. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1019

Bonnardel, F., Mariethoz, J., Salentin, S., Robin, X., Schroeder, M., Perez, S., Lisacek, F. D. S., & Imberty, A. (2019). Unilectin3d, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky832

Boyd, C. M., & Bubeck, D. (2018). Advances in cryoEM and its impact on β-pore forming proteins. *Current Opinion in Structural Biology*, *52*, 41–49. https://doi.org/10.1016/j.sbi.2018.07.010

Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. In *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro3239

Chung, C. H., Mirakhur, B., Chan, E., Le, Q.-T., Berlin, J., Morse, M., Murphy, B. A., Satinover, S. M., Hosen, J., Mauro, D., Slebos, R. J., Zhou, Q., Gold, D., Hatley, T., Hicklin, D. J., & Platts-Mills, T. A. E. (2008). Cetuximab-Induced Anaphylaxis and IgE Specific for Galactose-α-1,3-Galactose. *New England Journal of Medicine*, *358*(11), 1109–1117. https://doi.org/10.1056/nejmoa074943

Cirauqui, N., Abriata, L. A., Van Der Goot, F. G., & Dal Peraro, M. (2017). Structural, physicochemical and dynamic features conserved within the aerolysin pore-forming toxin family. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-13714-4

Colman, P. M., & Ward, C. W. (1985). Structure and diversity of influenza virus neuraminidase. In *Current topics in microbiology and immunology* (Vol. 114). https://doi.org/10.1007/978-3-642-70227-3_5

Connaris, H., Crocker, P. R., & Taylor, G. L. (2009). Enhancing the receptor affinity of the sialic acid-binding domain of Vibrio cholerae sialidase through multivalency. *Journal of Biological Chemistry*, *284*(11), 7339–7351. https://doi.org/10.1074/jbc.M807398200

Cooper, D. K. C., Koren, E., & Oriol, R. (1994). Oligosaccharides and Discordant Xenotransplantation. *Immunological Reviews*, *141*(1). https://doi.org/10.1111/j.1600-065X.1994.tb00871.x

Cordara, G., Egge-Jacobsen, W., Johansen, H. T., Winter, H. C., Goldstein, I. J., Sandvig, K., & Krengel, U. (2011). Marasmius oreades agglutinin (MOA) is a chimerolectin with proteolytic activity. *Biochemical and Biophysical Research Communications*, *408*(3), 405–410. https://doi.org/10.1016/j.bbrc.2011.04.031

Dan, X., Liu, W., & Ng, T. B. (2016). Development and Applications of Lectins as Biological Tools in Biomedical Research. *Medicinal Research Reviews*. https://doi.org/10.1002/med.21363

Dang, K., Zhang, W., Jiang, S., Lin, X., & Qian, A. (2020). Application of Lectin Microarrays for Biomarker Discovery. *ChemistryOpen*, *9*(3), 285–300. https://doi.org/10.1002/open.201900326

de Mol, N. J., & Fischer, M. J. E. (2010). Surface Plasmon Resonance: Methods and Protocols.

In *Life Sciences*. https://doi.org/10.1007/978-1-60761-670-2

De, S. N. (1959). Enterotoxicity of bacteria-free culture-filtrate of vibrio cholerae. *Nature*, *183*(4674). https://doi.org/10.1038/1831533a0

Degiacomi, M. T., Iacovache, I., Pernot, L., Chami, M., Kudryashev, M., Stahlberg, H., Van Der Goot, F. G., & Dal Peraro, M. (2013). Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1312

Dobie, C., & Skropeta, D. (2021). Insights into the role of sialylation in cancer progression and metastasis. In *British Journal of Cancer* (Vol. 124, Issue 1). https://doi.org/10.1038/s41416-020-01126-7

Drickamer, K. (1992). Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature*, *359*, 710–713.

Dumas, A., Lercher, L., Spicer, C. D., & Davis, B. G. (2015). Designing logical codon reassignment-Expanding the chemistry in biology. *Chemical Science*, *6*(1), 50–69. https://doi.org/10.1039/c4sc01534g

ESTOLA, E., & ELO, J. (1952). Occurrence of an exceedingly weak A blood group property in a family. *Annales Medicinae Experimentalis et Biologiae Fenniae*.

Feinberg, H., Guo, Y., Mitchell, D. A., Drickamer, K., & Weis, W. I. (2005). Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. *Journal of Biological Chemistry*, *280*(2), 1327–1335. https://doi.org/10.1074/jbc.M409925200

Fettis, M. M., Farhadi, S. A., & Hudalla, G. A. (2019). A chimeric, multivalent assembly of galectin-1 and galectin-3 with enhanced extracellular activity. *Biomaterials Science*, *7*(5). https://doi.org/10.1039/c8bm01631c

Fettis, M. M., & Hudalla, G. A. (2018). Engineering Reactive Oxygen Species-Resistant Galectin-1 Dimers with Enhanced Lectin Activity. *Bioconjugate Chemistry*, *29*(7), 2489–2496. https://doi.org/10.1021/acs.bioconjchem.8b00425

Galili, U., Shohet, S. B., Kobrin, E., Stults, C. L., & Macher, B. A. (1988). Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. *The Journal of Biological Chemistry*, *263*(33), 17755–17762. https://doi.org/10.1016/s0021-9258(19)77900-9

Glieder, A., Kubicek, C. P., Mattanovich, D., Wiltschi, B., & Sauer, M. (2015). Synthetic biology. In *Synthetic Biology*. https://doi.org/10.1007/978-3-319-22708-5

Grahn, E., Askarieh, G., Holmner, Å., Tateno, H., Winter, H. C., Goldstein, I. J., & Krengel, U. (2007). Crystal Structure of the Marasmius Oreades Mushroom Lectin in Complex with a Xenotransplantation Epitope. *Journal of Molecular Biology*, *369*(3), 710–721. https://doi.org/10.1016/j.jmb.2007.03.016

Hamorsky, K. T., Kouokam, J. C., Dent, M. W., Grooms, T. N., Husk, A. S., Hume, S. D., Rogers, K. A., Villinger, F., Morris, M. K., Hanson, C. V., & Matoba, N. (2019). Engineering of a Lectibody Targeting High-Mannose-Type Glycans of the HIV Envelope. *Molecular Therapy*, *27*(11), 2038–2052. https://doi.org/10.1016/j.ymthe.2019.07.021

Hazes, B. (1996). The (QxW)3 domain: A flexible lectin scaffold. *Protein Science*, *5*(8), 1490–1501. https://doi.org/10.1002/pro.5560050805

Hermanson, G. T. (2013). Bioconjugate Techniques: Third Edition. In *Bioconjugate Techniques: Third Edition*. https://doi.org/10.1016/C2009-0-64240-9

Hirabayashi, J., & Arai, R. (2019). Lectin engineering: The possible and the actual. In *Interface Focus*. https://doi.org/10.1098/rsfs.2018.0068

Hirabayashi, J., Yamada, M., Kuno, A., & Tateno, H. (2013). Lectin microarrays: Concept, principle and applications. *Chemical Society Reviews*, *42*(10), 4443–4458. https://doi.org/10.1039/c3cs35419a

Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W. H., & Blanchard, L. (2007). Interaction of the C-Terminal Domains of Sendai Virus N and P Proteins: Comparison of Polymerase-Nucleocapsid Interactions within the Paramyxovirus Family. *Journal of Virology*, *81*(13). https://doi.org/10.1128/jvi.00338-07

Hu, D., Tateno, H., & Hirabayashi, J. (2015). Lectin engineering, a molecular evolutionary approach to expanding the lectin utilities. *Molecules*, *20*(5), 7637–7656. https://doi.org/10.3390/molecules20057637

Irumagawa, S., Hiemori, K., Saito, S., Tateno, H., & Arai, R. (2022). Self-Assembling Lectin Nano-Block Oligomers Enhance Binding Avidity to Glycans. *International Journal of Molecular Sciences*, *23*(2). https://doi.org/10.3390/ijms23020676

Israeli-Ruimy, V., Bule, P., Jindou, S., Dassa, B., Moraïs, S., Borovok, I., Barak, Y., Slutzki, M., Hamberg, Y., Cardoso, V., Alves, V. D., Najmudin, S., White, B. A., Flint, H. J., Gilbert, H. J., Lamed, R., Fontes, C. M. G. A., & Bayer, E. A. (2017). Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions. *Scientific Reports*, *7*(February), 1–15. https://doi.org/10.1038/srep42355

Jaroentomeechai, T., Taw, M. N., Li, M., Aquino, A., Agashe, N., Chung, S., Jewett, M. C., & DeLisa, M. P. (2020). Cell-Free Synthetic Glycobiology: Designing and Engineering Glycomolecules Outside of Living Cells. *Frontiers in Chemistry*, *8*(July), 1–22. https://doi.org/10.3389/fchem.2020.00645

Jia, L., Zhang, J., Ma, T., Guo, Y., Yu, Y., & Cui, J. (2018). The function of fucosylation in progression of lung cancer. *Frontiers in Oncology*, *8*(DEC), 1–10. https://doi.org/10.3389/fonc.2018.00565

Keeffe, J. R., Gnanapragasam, P. N. P., Gillespie, S. K., Yong, J., Bjorkman, P. J., & Mayo, S. L. (2011). Designed oligomers of cyanovirin-N show enhanced HIV neutralization. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(34), 14079–14084. https://doi.org/10.1073/pnas.1108777108

Keusch, T. (1998). The Rediscovery of Shiga Toxin and Its Role in Clinical Disease. *Japanese Journal of Medical Science & Biology*, *51*, 5–22.

Kightlinger, W., Warfel, K. F., Delisa, M. P., & Jewett, M. C. (2020). Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synthetic Biology*, *9*(7), 1534–1562. https://doi.org/10.1021/acssynbio.0c00210

Kostlánová, N., Mitchell, E. P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerová, M., & Imberty, A. (2005). The fucose-binding Lectin from Ralstonia solanacearum: A new type of β-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan.

Journal of Biological Chemistry. https://doi.org/10.1074/jbc.M505184200

Kruger, R. P., Winter, H. C., Simonson-Leff, N., Stuckey, J. A., Goldstein, I. J., & Dixon, J. E. (2002). Cloning, expression, and characterization of the Galα1,3Gal high affinity lectin from the mushroom Marasmius oreades. *Journal of Biological Chemistry*, *277*(17), 15002–15005. https://doi.org/10.1074/jbc.M200165200

Kuno, A., Uchiyama, N., Koseki-Kuno, S., Ebe, Y., Takashima, S., Yamada, M., & Hirabayashi, J. (2005). Evanescent-field fluorescence-assisted lectin microarray: A new strategy for glycan profiling. *Nature Methods*, *2*(11). https://doi.org/10.1038/nmeth803

Kurhade, S. E., Weiner, J. D., Gao, F. P., & Farrell, M. P. (2021). Functionalized High Mannose-Specific Lectins for the Discovery of Type I Mannosidase Inhibitors. *Angewandte Chemie - International Edition*, *60*(22), 12313–12318. https://doi.org/10.1002/anie.202101249

Lightfoot, Y. L., Selle, K., Yang, T., Goh, Y. J., Sahay, B., Zadeh, M., Owen, J. L., Colliou, N., Li, E., Johannssen, T., Lepenies, B., Klaenhammer, T. R., & Mohamadzadeh, M. (2015). SIGNR 3-dependent immune regulation by Lactobacillus acidophilus surface layer protein A in colitis . *The EMBO Journal*, *34*(7). https://doi.org/10.15252/embj.201490296

Lundstrøm, J., Korhonen, E., Lisacek, F., & Bojar, D. (2022). LectinOracle: A Generalizable Deep Learning Model for Lectin–Glycan Binding Prediction. *Advanced Science*, *9*(1). https://doi.org/10.1002/advs.202103807

Macher, B. A., & Galili, U. (2008). The Galα1,3Galβ1,4GlcNAc-R (α-Gal) epitope: A carbohydrate of unique evolution and clinical relevance. *Biochimica et Biophysica Acta - General Subjects*, *1780*(2), 75–88. https://doi.org/10.1016/j.bbagen.2007.11.003

Madl, J., Villringer, S., & Römer, W. (2016). *Delving into Lipid-Driven Endocytic Mechanisms Using Biomimetic Membranes*. https://doi.org/10.1007/8623_2016_7

Maenuma, K., Yim, M., Komatsu, K., Hoshino, M., Takahashi, Y., Bovin, N., & Irimura, T. (2008). Use of a library of mutated Maackia amurensis hemagglutinin for profiling the cell lineage and differentiation. *Proteomics*, *8*(16), 3274–3283. https://doi.org/10.1002/pmic.200800037

Mancheño, J. M., Tateno, H., Goldstein, I. J., Martínez-Ripoll, M., & Hermoso, J. A. (2005). Structural analysis of the Laetiporus sulphureus hemolytic pore-forming lectin in complex with sugars. *Journal of Biological Chemistry*, *280*(17), 17251–17259. https://doi.org/10.1074/jbc.M413933200

Mao, H., Hart, S. A., Schink, A., & Pollok, B. A. (2004). Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *Journal of the American Chemical Society*, *126*(9). https://doi.org/10.1021/ja039915e

Mariethoz, J., Alocci, D., Gastaldello, A., Horlacher, O., Gasteiger, E., Rojas-Macias, M., Karlsson, N. G., Packer, N. H., & Lisacek, F. (2018). Glycomics@ExPASy: Bridging the gap. *Molecular and Cellular Proteomics*, *17*(11), 2164–2176. https://doi.org/10.1074/mcp.RA118.000799

Menéndez, M. (2020). Isothermal Titration Calorimetry: Principles and Applications. *ELS*, *1*(1), 113–127. https://doi.org/10.1002/9780470015902.a0028808

Merwe, P. A. Van Der. (2000). Surface plasmon resonance In: Harding S, Chowdhry PZ (eds)

Protein–ligand interactions: a practical approach. *Physics*, *627*, 1–50. http://www.springerlink.com/index/10.1007/978-1-60761-670-2

Mi, F., Guan, M., Hu, C., Peng, F., Sun, S., & Wang, X. (2021). Application of lectin-based biosensor technology in the detection of foodborne pathogenic bacteria: A review. *Analyst*, *146*(2), 429–443. https://doi.org/10.1039/d0an01459a

Morgan, W. T. J., & Watkins, W. M. (2000). Unravelling the biochemical basis of blood group ABO and Lewis antigenic specificity. In *Glycoconjugate Journal* (Vol. 17, Issues 7–9). https://doi.org/10.1023/A:1011014307683

Murzin, A. G., Lesk, A. M., & Chothia, C. (1992). β-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. *Journal of Molecular Biology*, *223*(2), 531–543. https://doi.org/10.1016/0022-2836(92)90668-A

Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütteke, T., O'Boyle, N., Packer, N. H., Stanley, P., Toukach, P., Varki, A., & Woods, R. J. (2019). Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology*, *29*(9). https://doi.org/10.1093/glycob/cwz045

Noguchi, H., Addy, C., Simoncini, D., Wouters, S., Mylemans, B., Van Meervelt, L., Schiex, T., Zhang, K. Y. J., Tame, J. R. H., & Voet, A. R. D. (2019). Computational design of symmetrical eight-bladed β-propeller proteins. *IUCrJ*. https://doi.org/10.1107/S205225251801480X

Notova, S., Bonnardel, F., Lisacek, F., Varrot, A., & Imberty, A. (2020). Structure and engineering of tandem repeat lectins. *Current Opinion in Structural Biology*, *62*, 39–47. https://doi.org/10.1016/j.sbi.2019.11.006

Notova, S., Bonnardel, F., Rosato, F., Siukstaite, L., Schwaiger, J., Bovin, N., Varrot, A., Römer, W., Lisacek, F., & Imberty, A. (2022). A pore-forming β-trefoil lectin with specificity for the tumor-related glycosphingolipid Gb3. *BioRxiv*. https://doi.org/10.1101/2022.02.10.479907

Oh, Y. J., Dent, M. W., Freels, A. R., , Qingwen Zhou, C. B., Lebrilla, & Matoba, Michael L. Merchant, N. (2022). Antitumor activity of a lectibody targeting cancer-associated high-mannose glycans. *Molecular Therapy*, *30*(4), 1523–1535.

Pilobello, K. T., Krishnamoorthy, L., Slawek, D., & Mahal, L. K. (2005). Development of a lectin microarray for the rapid analysis of protein glycopatterns. *ChemBioChem*, *6*(6). https://doi.org/10.1002/cbic.200400403

Powlesland, A. S., Ward, E. M., Sadhu, S. K., Guo, Y., Taylor, M. E., & Drickamer, K. (2006). Widely divergent biochemical properties of the complete set of mouse DC-SIGN-related proteins. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M601925200

Propheter, D. C., Hsu, K. L., & Mahal, L. K. (2011). Recombinant lectin microarrays for glycomic analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *723*. https://doi.org/10.1007/978-1-61779-043-0_6

Rabes, A., Zimmermann, S., Reppe, K., Lang, R., Seeberger, P. H., Suttorp, N., Witzenrath, M., Lepenies, B., & Opitz, B. (2015). The C-type lectin receptor mincle binds to streptococcus pneumoniae but plays a limited role in the anti-pneumococcal innate immune response. *PLoS ONE*, *10*(2). https://doi.org/10.1371/journal.pone.0117022

Ramberg, K. O., Guagnini, F., Engilberge, S., Wrońska, M. A., Rennie, M. L., Pérez, J., & Crowley, P. B. (2021). Segregated Protein–Cucurbit[7]uril Crystalline Architectures via Modulatory Peptide Tectons. *Chemistry - A European Journal*, *27*(59), 14619–14627. https://doi.org/10.1002/chem.202103025

Ribeiro, Joao P., Pau, W., Pifferi, C., Renaudet, O., Varrot, A., Mahal, L. K., & Imberty, A. (2016). Characterization of a high-affinity sialic acid-specific CBM40 from Clostridium perfringens and engineering of a divalent form. *Biochemical Journal*, *473*(14), 2109–2118. https://doi.org/10.1042/BCJ20160340

Ribeiro, João P., Villringer, S., Goyard, D., Coche-Guerente, L., Höferlin, M., Renaudet, O., Römer, W., & Imberty, A. (2018). Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells. *Chemical Science*, *9*(39), 7634–7641. https://doi.org/10.1039/c8sc02730g

Romano, P. R., Mackay, A., Vong, M., deSa, J., Lamontagne, A., Comunale, M. A., Hafner, J., Block, T., Lec, R., & Mehta, A. (2011). Development of recombinant Aleuria aurantia lectins with altered binding specificities to fucosylated glycans. *Biochemical and Biophysical Research Communications*, *414*(1), 84–89. https://doi.org/10.1016/j.bbrc.2011.09.027

Ross, J. F., Wildsmith, G. C., Johnson, M., Hurdiss, D. L., Hollingsworth, K., Thompson, R. F., Mosayebi, M., Trinh, C. H., Paci, E., Pearson, A. R., Webb, M. E., & Turnbull, W. B. (2019). Directed Assembly of Homopentameric Cholera Toxin B-Subunit Proteins into Higher-Order Structures Using Coiled-Coil Appendages. *Journal of the American Chemical Society*, *141*(13), 5211–5219. https://doi.org/10.1021/jacs.8b11480

Rutenber, E., & Robertus, J. D. (1991). Structure of ricin B-chain at 2.5 Å resolution. *Proteins: Structure, Function, and Bioinformatics*, *10*(3). https://doi.org/10.1002/prot.340100310

Savva, C. G., Clark, A. R., Naylor, C. E., Popoff, M. R., Moss, D. S., Basak, A. K., Titball, R. W., & Bokori-Brown, M. (2019). The pore structure of Clostridium perfringens epsilon toxin. *Nature Communications*, *10*(1), 25–28. https://doi.org/10.1038/s41467-019-10645-8

Schatz, P. J. (1993). Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: A 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Bio/Technology*. https://doi.org/10.1038/nbt1093-1138

Shanina, E., Siebs, E., Zhang, H., Varón Silva, D., Joachim, I., Titz, A., & Rademacher, C. (2021). Protein-observed 19F NMR of LecA from Pseudomonas aeruginosa. *Glycobiology*, *31*(2), 159–165. https://doi.org/10.1093/glycob/cwaa057

Sharon, N., & Lis, H. (2004). History of lectins: From hemagglutinins to biological recognition molecules. In *Glycobiology*. https://doi.org/10.1093/glycob/cwh122

Siukstaite, L., Rosato, F., Mitrovic, A., Müller, P. F., Kraus, K., Notova, S., Imberty, A., & Römer, W. (2021). The two sweet sides of janus lectin drive crosslinking of liposomes to cancer cells and material uptake. *Toxins*, *13*(11), 1–19. https://doi.org/10.3390/toxins13110792

Smock, R. G., Yadid, I., Dym, O., Clarke, J., & Tawfik, D. S. (2016). De Novo Evolutionary Emergence of a Symmetrical Protein Is Shaped by Folding Constraints. *Cell*, *164*(3), 476–486. https://doi.org/10.1016/j.cell.2015.12.024

Tateno, H., & Goldstein, I. J. (2004). Partial identification of carbohydrate-binding sites of a Galα1,3Galβ1,4GlcNAc-specific lectin from the mushroom Marasmius oreades by site-directed mutagenesis. *Archives of Biochemistry and Biophysics*, *427*(1), 101–109. https://doi.org/10.1016/j.abb.2004.04.013

Tateno, H., Onuma, Y., Ito, Y., Minoshima, F., Saito, S., Shimizu, M., Aiki, Y., Asashima, M., & Hirabayashi, J. (2015). Elimination of tumorigenic human pluripotent stem cells by a recombinant lectin-toxin fusion protein. *Stem Cell Reports*. https://doi.org/10.1016/j.stemcr.2015.02.016

Tateno, H., & Saito, S. (2017). Engineering of a Potent Recombinant Lectin-Toxin Fusion Protein to Eliminate Human Pluripotent Stem Cells. *Molecules (Basel, Switzerland)*, *22*(7), 1–9. https://doi.org/10.3390/molecules22071151

Tateno, H., Toyota, M., Saito, S., Onuma, Y., Ito, Y., Hiemori, K., Fukumura, M., Matsushima, A., Nakanishi, M., Ohnuma, K., Akutsu, H., Umezawa, A., Horimoto, K., Hirabayashi, J., & Asashima, M. (2011). Glycome diagnosis of human induced pluripotent stem cells using lectin microarray. *Journal of Biological Chemistry*, *286*(23). https://doi.org/10.1074/jbc.M111.231274

Terada, D., Kawai, F., Noguchi, H., Unzai, S., Hasan, I., Fujii, Y., Park, S. Y., Ozeki, Y., & Tame, J. R. H. (2016). Crystal structure of MytiLec, a galactose-binding lectin from the mussel Mytilus galloprovincialis with cytotoxicity against certain cancer cell types. *Scientific Reports*, *6*(June), 1–11. https://doi.org/10.1038/srep28344

Terada, D., Voet, A. R. D., Noguchi, H., Kamata, K., Ohki, M., Addy, C., Fujii, Y., Yamamoto, D., Ozeki, Y., Tame, J. R. H., & Zhang, K. Y. J. (2017). Computational design of a symmetrical β-trefoil lectin with cancer cell binding activity. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-06332-7

Tobola, F., Lelimousin, M., Varrot, A., Gillon, E., Darnhofer, B., Blixt, O., Birner-Gruenberger, R., Imberty, A., & Wiltschi, B. (2018). Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. *ACS Chemical Biology*, *13*(8), 2211–2219. https://doi.org/10.1021/acschembio.8b00377

Tobola, F., Lepšík, M., Zia, S. R., Leffler, H., Nilsson, U. J., Blixt, O., Imberty, A., & Wiltschi, B. (2022). Engineering the Ligand Specificity of the Human Galectin-1 by Incorporation of Tryptophan Analogues. *ChemBioChem*, *23*(5), 1–7. https://doi.org/10.1002/cbic.202100593

Tobola, F., Sylvander, E., Gafko, C., & Wiltschi, B. (2019). 'Clickable lectins': Bioorthogonal reactive handles facilitate the directed conjugation of lectins in a modular fashion. *Interface Focus*, *9*(2). https://doi.org/10.1098/rsfs.2018.0072

Truebestein, L., & Leonard, T. A. (2016). Coiled-coils: The long and short of it. *BioEssays*, *38*(9), 903–916. https://doi.org/10.1002/bies.201600062

Turnbull, W. B., & Daranas, A. H. (2003). On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *Journal of the American Chemical Society*, *125*(48). https://doi.org/10.1021/ja036166s

Turnbull, W. B., Imberty, A., & Blixt, O. (2019). Synthetic glycobiology. In *Interface Focus* (Vol. 9, Issue 2). https://doi.org/10.1098/rsfs.2019.0004

Unno, H., Goda, S., & Hatakeyama, T. (2014a). Hemolytic lectin cel-iii heptamerizes via a large structural transition from α-helices to a β-barrel during the transmembrane pore formation process. *Journal of Biological Chemistry*, *289*(18), 12805–12812. https://doi.org/10.1074/jbc.M113.541896

Unno, H., Goda, S., & Hatakeyama, T. (2014b). Hemolytic lectin cel-iii heptamerizes via a large structural transition from α-helices to a β-barrel during the transmembrane pore formation process. *Journal of Biological Chemistry*, *289*(18), 12805–12812. https://doi.org/10.1074/jbc.M113.541896

Varki, A. (2017). Biological roles of glycans. *Glycobiology*, *27*(1), 3–49. https://doi.org/10.1093/glycob/cww086

Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F. G., Lütteke, T., … Kornfeld, S. (2015). Symbol nomenclature for graphical representations of glycans. *Glycobiology*, *25*(12). https://doi.org/10.1093/glycob/cwv091

Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., & Etzler, M. E. (2009). Essentials of Glycobiology, second ed. In *Essentials of Glycobiology*.

Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., & Seeberger, P. H. (2017). Essentials of glycobiology, third edition. In *Cold Spring Harbor Laboratory Press*.

Venditto, I., Luis, A. S., Rydahl, M., Schückel, J., Fernandes, V. O., Vidal-Melgosa, S., Bule, P., Goyal, A., Pires, V. M. R., Dourado, C. G., Ferreira, L. M. A., Coutinho, P. M., Henrissat, B., Knox, J. P., Baslé, A., Najmudin, S., Gilbert, H. J., Willats, W. G. T., & Fontes, C. M. G. A. (2016). Complexity of the Ruminococcus flavefaciens cellulosome reflects an expansion in glycan recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), 7136–7141. https://doi.org/10.1073/pnas.1601558113

Voet, A. R. D., Noguchi, H., Addy, C., Simoncini, D., Terada, D., Unzai, S., Park, S. Y., Zhang, K. Y. J., & Tame, J. R. H. (2014). Computational design of a self-assembling symmetrical β-propeller protein. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15102–15107. https://doi.org/10.1073/pnas.1412768111

Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., & Aebi, M. (2002). N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (New York, N.Y.)*, *298*(5599), 1790–1793. https://doi.org/10.1126/science.298.5599.1790

Ward, E. M., Kizer, M. E., & Imperiali, B. (2021). Strategies and Tactics for the Development of Selective Glycan-Binding Proteins. *ACS Chemical Biology*, *16*(10), 1795–1813. https://doi.org/10.1021/acschembio.0c00880

Wilchek, M., & Bayer, E. A. (1990). Applications of avidin-biotin technology: Literature survey. *Methods in Enzymology*. https://doi.org/10.1016/0076-6879(90)84257-H

Wiltschi, B. (2012). Chapter 13 Expressed Protein Modifi cations : Making. *Methods*, *813*, 211–225. https://doi.org/10.1007/978-1-61779-412-4

Winter, H. C., Mostafapour, K., & Goldstein, I. J. (2002). The mushroom Marasmius oreades lectin is a blood group type B agglutinin that recognizes the Galα1,3Gal and Galα1,3Gal,β1,4GlcNAc porcine xenotransplantation epitopes with high affinity. *Journal of Biological Chemistry*, *277*(17), 14996–15001. https://doi.org/10.1074/jbc.M200161200

Yabe, R., Suzuki, R., Kuno, A., Fujimoto, Z., Jigami, Y., & Hirabayashi, J. (2007). Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-binding protein by natural evolution-mimicry. *Journal of Biochemistry*, *141*(3), 389–399. https://doi.org/10.1093/jb/mvm043

Yadid, I., & Tawfik, D. S. (2007). Reconstruction of Functional β-Propeller Lectins via Homo-oligomeric Assembly of Shorter Fragments. *Journal of Molecular Biology*, *365*(1), 10–17. https://doi.org/10.1016/j.jmb.2006.09.055

Yadid, I., & Tawfik, D. S. (2011). Functional -propeller lectins by tandem duplications of repetitive units. *Protein Engineering, Design and Selection*, *24*(1–2), 185–195. https://doi.org/10.1093/protein/gzq053

Yang, C., Xu, C., Wang, X., & Hu, X. (2012). Quantum-dot-based biosensor for simultaneous detection of biomarker and therapeutic drug: First steps toward an assay for quantitative pharmacology. *Analyst*. https://doi.org/10.1039/c2an15894a

Zheng, T., Peelen, D., & Smith, L. M. (2005). Lectin arrays for profiling cell surface carbohydrate expression. *Journal of the American Chemical Society*, *127*(28). https://doi.org/10.1021/ja0505550

SUMMARY

Glycobiology is a rapidly growing field of natural sciences with a focus on glycans, glycoconjugates and glycan binding proteins. Lectins are sugar-binding proteins present in all types of organisms and they display wide range of biological functions. As encoded in their name (from Latin - *legere* – to select), each lectin is specific only to a finite group of glycans and in order to ensure higher affinity, they are generally multivalent. Lectins are powerful glycan-profiling tools and even though some of them already found their applications, the discovery of novel lectins is still desirable. In this regard, involvement of synthetic biology and protein engineering are of high interest for building of lectin architecture and tuning their specificity. Various approaches for lectin discovery or engineering are presented in this thesis. The thesis is composed of several chapters, where the introduction is dedicated to the general description of lectins and lectin engineering with the respect to their specificity and topology, including a short review on engineering of β-propeller and β-trefoil lectins. The results are presented in three scientific articles (in the format of preprint or manuscripts in preparation). The first publication describes the discovery and characterization of novel pore-forming lectin with specificity toward cancer cells glyco-epitope. In the second manuscript, synthetic biology approach was used to create artificial proteins with the ability to recognize plant cell wall polymers and to be used as glue proteins in the construction of an artificial plant cell wall. The third manuscript generalizes the Janus lectin strategy as a universal tool for creation of bispecific chimeras with increased valency. The last chapter summarizes the achieved results and propose new perspectives and challenges giving the special importance to the continuation of lectin engineering.

RÉSUMÉ

La glycobiologie est un domaine des sciences naturelles en plein essor qui se concentre sur les glycanes, les glycoconjugués et les protéines de liaison aux glycanes. Les lectines sont des protéines de reconnaissance des sucres présentes dans tous les types d'organismes et elles présentent un large éventail de fonctions biologiques. Comme indiqué dans leur nom (du latin - *legere* - sélectionner), chaque lectines est spécifique seulement d'un groupe limité de glycanes et, afin d'assurer une affinité plus élevée, elles sont généralement multivalentes. Les lectines sont de puissants outils de profilage des glycanes et même si certaines d'entre elles ont déjà trouvé leurs applications, la découverte de nouvelles lectines est toujours souhaitable. Dans cette optique, l'implication de la biologie synthétique et de l'ingénierie des protéines est d'un grand intérêt pour la modification de l'architecture ou même de la spécificité des lectines. Différentes approches pour la découverte ou l'ingénierie des lectines sont présentées dans cette thèse. La thèse est composée de plusieurs chapitres, où l'introduction est dédiée à la description générale des lectines et de leur ingénierie pour ce qui concerne leur spécificité et leur topologie. Elle comprend une brève revue de l'ingénierie des lectines β-propeller et β-trefoil. Les résultats sont présentés dans trois articles scientifiques (préprint ou manuscrit en préparation). La première publication décrit la découverte et la caractérisation d'une nouvelle lectine formant des pores dans les membranes et présentant une spécificité pour des glyco-epitopes de cellules cancéreuses. Dans le deuxième manuscrit, une approche de biologie synthétique a été utilisée pour créer des protéines artificielles ayant la capacité de reconnaître les polysaccharides de la paroi cellulaire végétale et ainsi les utiliser comme glue dans la construction de paroi cellulaire végétale artificielle. Le troisième manuscrit généralise la stratégie des lectines. Janus comme un outil universel pour la création de chimères bispécifiques avec une valence accrue. Le dernier chapitre résume les résultats obtenus et propose de nouvelles perspectives et de nouveaux défis en accordant une importance particulière à la poursuite de l'ingénierie des lectines.