Crystal Structures of Fungal Tectonin in Complex with O-Methylated Glycans Suggest Key Role in Innate Immune Defense

Roman Sommer, Olga N Makshakova, Therese Wohlschlager, Stephanie Hutin, May Marsh, Alexander Titz, Markus Künzler, Annabelle Varrot

Crystal Structures of Fungal Tectonin in Complex with O-Methylated Glycans Suggest Key Role in Innate Immune Defense.

Crystal structures of fungal tectonin in complex with O-methylated glycans suggest key role in innate immune defense

Roman Sommer[1,6], Olga N. Makshakova[2,7], Therese Wohlschlager[3,8], Stephanie Hutin[4], May Marsh[5], Alexander Titz[1,6], Markus Künzler[3], Annabelle Varrot[2,9,*]

1Chemical Biology of Carbohydrates, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), 66123 Saarbr€ucken, Germany

[2]Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France

[3]Institute of Microbiology, Department of Biology, Swiss Federal Institute of Technology (ETH), 8093 Z€urich, Switzerland

[4]Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale, 38000 Grenoble, France

[5]Macromolecular Crystallography, Swiss Light Source, Paul Scherrer Institut, 5232 Villigen PSI, Switzerland

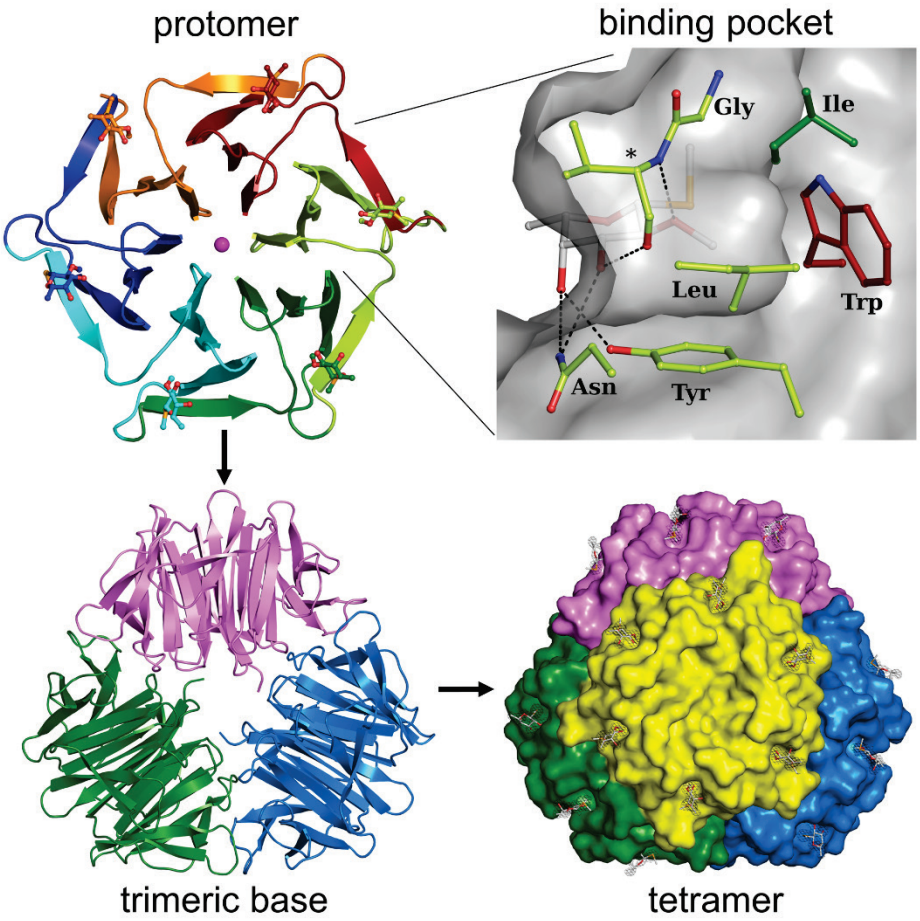
[6]Deutsches Zentrum f€ur Infektionsforschung, Standort Hannover-Braunschweig, Hannover, Germany

[7]Present address: Kazan Institute of Biochemistry and Biophysics, Kazan Scientific Center, Russian Academy of Sciences, Lobachevsky Street 2/31, PO Box 30, 420111 Kazan, Russian Federation

[8]Present address: Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria

[9]Lead Contact

[*]Correspondence: annabelle.varrot@cermav.cnrs.fr



In Brief

Protein-carbohydrate interactions play an important role in innate immunity as they facilitate discrimination between self and non-self. Tectonins are innate defense effectors that specifically recognize O-methylated glycans. Sommer et al. determine the structural basis for tectonin specificity, revealing a conserved recognition motif and a surprising capsid-like tetrameric architecture.

Highlights

-Lb-Tec2 structure solved using O-methylated selenofucoside

-Virus-like tetramers revealed by X-ray structures and confirmed by SAXS

-Lb-Tec2 binding site defines a signature for selectivity for O-methylated glycans

-Our work supports a conserved role of tectonins in innate immunity

SUMMARY

Innate immunity is the first line of defense against pathogens and predators. To initiate a response, it relies on the detection of invaders, where lectin-carbohydrate interactions play a major role. O-methylated glycans were previously identified as non-self epitopes and conserved targets for defense effector proteins belonging to the tectonin superfamily. Here, we present two crystal structures of Tectonin 2 from the mushroom *Laccaria bicolor* (Lb-Tec2) in complex with methylated ligands, unraveling the molecular basis for this original specificity. Furthermore, they revealed the formation of a ball-shaped tetramer with 24 binding sites distributed at its surface, resembling a small virus capsid. Based on the crystal structures, a methylation recognition motif was identified and found in the sequence of many tectonins from bacteria to human. Our results support a key role of tectonins in innate defense based on a distinctive and conserved type of lectin-glycan interaction.

INTRODUCTION

Innate immunity protects organisms against attacks from microbial pathogens or animal predators. In animals, innate immunity is the first line of defense prior to induction of the adaptive immune reaction, whereas in plants and fungi, it is the sole defense strategy (Dambuza and Brown, 2015; Zipfel, 2014). Pattern recognition receptors (PRRs) are produced as part of the innate immune system to recognize foreign and evolutionarily conserved epitopes called pathogen-associated molecular patterns: PAMPS (Brubaker et al., 2015; Zipfel, 2014). In the PRR-producing organism, binding to those "non-self" signatures often induces an immune response, e.g., the production of toxins or the activation of intracellular pathways to fight the invaders. PAMPs are usually easily accessible and shared among groups of pathogens. Classical examples of PAMPs are glycans, which cover the surface of every cell as a dense and complex layer, representing specific markers of a cell's identity. Lipopolysaccharides (LPS) in bacteria, glycolipids in mycobacteria, mannans in yeast, or β-glucans in fungi are the most commonly described microbial glycans (Kumar et al., 2011). PRRs recognizing glycan-based PAMPs are lectins, ubiquitous glycan-binding proteins such as the well-known C-type lectins (Dambuza and Brown, 2015). Fungal lectins have been shown to target glycoepitopes in soil microbes and predators, such as nematodes and insects, thereby inducing toxicity in the target organisms. These toxic lectins therefore act as both PRRs and defense effectors in fungi (Künzler, 2015; Varrot et al., 2013). One example for such a fungal defense lectin is Tectonin 2 from the mushroom *Laccaria bicolor* (Lb-Tec2) with specificity for O-methylated glycans. Lb-Tec2 was previously shown to agglutinate bacteria and to exert nematotoxicity towards *Caenorhabditis elegans*, both activities depending on the presence of O-methylated glycans in the target organisms (Wohlschlager et al., 2014). Despite this knowledge, the molecular basis governing the specificity of Lb-Tec2 remained undetermined.

Lb-Tec2 belongs to the protein family of tectonins found to be conserved throughout evolution and to be implicated in antimicrobial defense (Low et al., 2010). Tectonins comprise tandem repeats of about 35 amino acids of the so-called tectonin domain or TECPR, each of which forms a four stranded antiparallel β-sheet, constituting a blade in a β-propeller protein. The TECPR consensus sequence, first pinpointed in two proteins of the slime mold *Physarum polycephalum,* has previously been identified in proteins from actinomycetes, fungi, marine invertebrates, fish, and human (Capaldi et al., 2015; Chen et al., 2001; Huh et al., 1998; Low et al., 2009; Low et al., 2010; Saito et al., 1995; Schröder et al., 2003; Wohlschlager et al., 2014). Tectonins have been isolated from the hemolymph plasma and/ or the granular hemocytes of horseshoe crabs, fish eggs, the surface of sponge cells as well as the plasma membrane of slime molds. They were shown to bind to components present in bacterial LPS and to exhibit antibacterial activity (Capaldi et al., 2015; Chen et al., 2001; Huh et al., 1998; Low et al., 2010; Saito et al., 1995; Schröder et al., 2003). Many tectonins, including Lb-Tec2, consist of six TECPRs, whereas some others contain additional domains, e.g., the ricin B-type lectin domain found at the *N*-terminus of Tectonin 2 from *P. polycephalum* (Huh et al., 1998). Sequence analyses revealed distant orthologs in human (TECPR1 and TECPR2) as well as in *C. elegans* and *Drosophila melanogaster,* with much bigger size and a more complex architecture, comprising more than six TECPRs and several additional domains (Low et al., 2009). TECPR1 was shown to be involved in a pathway for selective autophagy of bacterial pathogens supporting the role of tectonins in immune defense (Ogawa et al., 2011). The recently published crystal structure of the carp fish-egg lectin FEL confirmed folding as a six-bladed β-propeller (Capaldi et al., 2015).

Methylation is a rare secondary modification of glycans whose function is still unclear (Staudacher, 2012). Methylated glycans are found in plants, fungi, bacteria, worms, and mollusks, but are absent from arthropods and mammals. In bacteria, methylation has been

described both at internal and terminal positions of LPS (Staudacher, 2012), whilst in the nematode *C. elegans*, this modification occurs mainly at the termini of N- and O-glycans (Yan et al., 2015). Although glycan methylation has been described for several organisms, the biosynthesis as well as the basis for molecular recognition by potential binding partners is largely unknown. Given their occurrence and the specific recognition by defense lectins, O-methylated glycans may represent conserved non-self epitopes recognized by the innate immune system in order to identify and combat a large variety of antagonists (Wohlschlager et al., 2014).

Herein, we report on the specific recognition of O-methylated glycans by tectonins through the structural and biophysical study of Lb-Tec2 in complex with two different ligands. As revealed by X-ray crystallography, Lb-Tec2 selectively recognizes methylation on different glycans with a binding site that is highly conserved throughout the tectonin family. The structural data was correlated with experimental binding affinities for the respective ligands and further supported by *in silico* docking studies. A distinctive tetrameric assembly was uncovered in the crystal and was confirmed in solution by small angle X-ray scattering experiments. The quaternary structure confers a very high multivalency, resulting in high avidity for glycan surfaces as determined by surface plasmon resonance experiments. Our results highlight the conservation of a selective recognition motif for a secondary glycan modification within the tectonin family. This strategy allows an organism to combat many invaders that share glycan methylation with a limited set of effector molecules.

RESULTS

In the absence of an appropriate model structure, methyl 2-*O*-methyl-seleno-L-fucopyranoside (2MeSeFuc) was synthesized for phasing purpose (Sommer et al., 2016a). Using the signal of selenium, the Lb-Tec2 structure in complex with 2MeSeFuc was solved by multiwavelength anomalous dispersion (MAD) at 1.65 Å resolution in R32:h (also named H32) space group. We faced an interesting case of static disorder with three different orientations of the protein chains throughout the crystal (detailed in methods). The final refined model, with values of 13.3% for R_{cryst} and 16.7% for R_{free}, consists of protein chain A fully occupied and of protein chain B placed on the symmetry three-fold axis with an occupancy of 0.33 providing the three different protein orientations by crystal symmetry. Data and refinement statistics are summarized in **Table 1**.

Tertiary structure of Lb-Tec2

Lb-Tec2 folds as a six-bladed β-propeller, where each blade consists of a four-stranded antiparallel β-sheet with strands denoted A to D from the center to the outside of the barrel (Chaudhuri et al., 2008). According to this nomenclature, each Lb-Tec2 TECPR repeat consists of strands D-A-B-C with a circular permutation of the β-strand D between blades. It allows the β-propeller closure through a molecular Velcro that presumably grants extra structural stability (**Figure 1A**). The β-propeller exhibits short surface loops and a narrow funnel shaped central channel with a diameter of around 6 Å at the so-called "bottom or exit" opening and 15 Å at the "top or entry" opening. The top displays by convention the DA and BC loops, and is commonly the narrower of the two channel openings (Chen et al., 2011). However, in Lb-Tec2, it is the larger one (**Figure 1A**). The central channel is filled with solvent while its "top" is obstructed by the side chains of Arg12, Arg87, Asp50, Asp125, and Asp203 (**Figure 1B**). A magnesium ion is also found in the channel coordinated to six water

molecules, where three interact with the two carboxylate oxygens of the aspartate residues mentioned above and the remaining three with the main chain oxygen of Ile13, Ile51, Ile88, Ile126, Ile66 and Ile204 (**Figure S1A**).

Quaternary structure of Lb-Tec2

The crystal structure of Lb-Tec2 revealed a distinctive quaternary structure with a tetramer created from three-fold symmetry in contrast to the two- or four-fold symmetry usually observed in other tetrameric structures deposited in the Protein Data Bank (Rose et al., 2017). Three protein chains, related by the three-fold axis, form a symmetric base (**Figure 1C**), which is capped by the fourth protein chain with the symmetry axis in its central channel. This leads to the formation of a pseudo-sphere with dimensions of 64 x 71 x 73 Å (**Figure 1D**). Using on-line size exclusion chromatography combined with small angle X-ray scattering (SAXS) measurements, we determined a ball shaped envelope in agreement with the crystal structure (**Figures 1E and S1B-D, Table S1**). Direct comparison of the predicted scattering curve of the crystal structure to the experimental SAXS data also confirmed the tetrameric arrangement of Lb-Tec2 in solution (**Figure S1D**). Furthermore, application of the Porod-Debye law shows that the protein density calculated for a tetramer (v = 1.13 mg cm^{-3}) is more in line with the empirical density for a protein (1.35-1.37 mg cm^{-3}) than the one calculated for a trimer (v = 0.85 mg cm^{-3}) (Rambo and Tainer, 2011).

Analyses of the tetramer with the PDBePISA (Krissinel and Henrick, 2007) and Protein Interaction Calculator (PIC) (Tina et al., 2007) servers permitted to identify three oligomerization interfaces, created mainly by surface loops at the bottom of the β-propeller (**Figure 1F)**. Interface I comprises residues Lys4, Arg32, surface loops β2-β3 (18-19), β4-β5 (35-40), β8-β9 and residues of the corresponding strands (Arg70-His77). Interface II comprises residues from surface loops β18-β19 (171-172), β20-β21 (188-192) and strand β24

(223-227) and interface III involves residues from surface loop β10-β11 (93-94), β12-β13 (110-114) and β16-β17 (148-154) and from strands β12 (Arg107) and β16 (Arg145-Tyr146). The whole tetramer has a total surface area of 27740 $Å^2$, with 10700 $Å^2$ buried which is equivalent to 39% with on average 10% of each protein chain involved in each interface (**Table S2**). Formation of the trimeric base is mediated by numerous hydrogen bonds and ionic interactions between interfaces I and II and one cation-π interaction between Trp73 and Lys171. The tetramer capping involves hydrogen bonds and ionic interactions of interface III of the three protein chains from trimeric base with either interface I, II and III of the capping protein chain, depending on its orientation. Only one hydrophobic interaction is observed between Val42, Val117 and Leu195 from the capping chain and Val117 of the base protein chains.

Recognition of methylated glycans by Lb-Tec2

2-*O*-Methyl-α-L-fucose (2MeFuc) and 4-*O*-methyl-α-D-mannose (4MeMan) represent potential *in vivo* targets of Lb-Tec2 (Wohlschlager et al., 2014; Yan et al., 2015). Hence, we aimed to dissect the atomic basis of their recognition by Lb-Tec2. The natural 2MeFuc was derivatized by the incorporation of a selenomethyl at the anomeric position into 2MeSeFuc for phasing purpose (Sommer et al., 2016a) and the allyl glycoside of 4MeMan was also prepared (All4MeMan) (Beshr et al., 2016). Both synthetic ligands were located unambiguously in the electron density maps of the crystal structures in complex with Lb-Tec2. Six binding sites, one on each blade, could be identified (**Figure 1A**). In both structures, site 6 in protein chain A was not occupied due to steric hindrance generated by the crystal packing. In the 2MeSeFuc/Lb-Tec2 structure, all other binding sites were occupied, while binding sites 1 and 2 were also found unoccupied in protein chain A in the All4MeMan/Lb-Tec2 structure. Surprisingly, although the crystal was soaked with an

anomeric mixture of α/β 2MeSeFuc, only the β-anomer could be modelled in the electron density.

The binding sites were found on the concave face of each blade in a shallow depression with a pocket accommodating the methyl group. They are created by residues of strands B-C-D and of the surface loop connecting strands D to A (**Figure 1A**). The six subsites are almost identical with all amino acids involved in the carbohydrate recognition conserved apart of one, and all binding sites exhibit the following interactions (**Figure S2A and Table S3**). The ND1 atom of an asparagine forms hydrogen bonds with both O3 and O4 hydroxyl groups of 2MeSeFuc or with both O2 and O3 hydroxyls of All4MeMan. The hydroxyl of a tyrosine is hydrogen bonded to the O4 or O2 hydroxyl of 2MeSeFuc or All4MeMan, respectively. A non-conserved residue of the GXL motif in loop DA forms hydrogen bonds between its main chain nitrogen and the methylated oxygen of the ligand and between its main chain oxygen and the O3 hydroxyl (**Figure 2A**). The β-face of the ligand ring shows CH-π interaction with the phenyl group of the tyrosine mentioned above. The pocket accommodating the methyl group is formed by conserved hydrophobic and aromatic residues: one leucine in loop DA and one isoleucine in strand C of the preceding TECPR build the bottom of the pocket, whilst the side walls are formed by the side chains of one tyrosine in strand C, one tryptophan in strand D of the following TECPR, and the backbone of one glycine in loop DA (**Figure 2A**). Thus, three consecutive TECPRs are necessary to form the pocket involved in methylation recognition. L-fucosides and D-mannosides exhibit a similar relative stereochemistry with two equatorial hydroxyl groups located next to an axial one where O4 is axial, and O3 and O2 are equatorial or O2 is axial, and O3 and O4 are equatorial, respectively. This steric resemblance provides the structural basis for recognition of both types of sugars by Lb-Tec2, a phenomenon observed also in other Man/Fuc-specific lectins, e.g., *Pseudomonas aeruginosa* LecB (Sommer et al., 2016b).

Determination of carbohydrate affinity and avidity of Lb-Tec2

The affinity of Lb-Tec2 for 2MeSeFuc, All4MeMan and GlcNAc in solution was determined by isothermal titration microcalorimetry (**Figure S2B-D**). To determine accurately the affinity, the number of binding sites was fixed for the curve fitting to six as deduced from the crystal structure as described for low affinity systems (Turnbull and Daranas, 2003). The dissociation constant of 6 mM, obtained for 2MeSeFuc, is comparable to the 4.4 mM determined previously for All2MeFuc (Wohlschlager et al., 2014). The affinity for All4MeMan of 16.6 mM is similar to that determined previously for All3MeMan (20.6 mM) (Wohlschlager et al., 2014). A weak interaction (K_d = 140 mM) was measured for GlcNAc (**Table 2**), which has been described as a potential ligand for other tectonins, and used for elution during the purification procedure (Capaldi et al., 2015; Wohlschlager et al., 2014). The order of affinity is therefore as follows α2MeFuc > α4MeMan, α3MeMan > αFuc > α/βGlcNAc.

In order to study Lb-Tec2 avidity, biotinylated 4MeMan (**3)** was synthesized from allyl mannoside **1** after cysteamine addition to the double bond and subsequent coupling of **2** to a biotin moiety (**Figures 2B, S3** and **S4**). Compound **3** was then fixed on a streptavidin-coated sensorchip for surface plasmon resonance (SPR) experiments and different concentrations of Lb-Tec2 were injected. Very fast association and dissociation between protein and the ligand-modified surface were observed precluding kinetic analysis. An apparent K_d of 286 μM was obtained in steady state analysis indicating a 60-fold increase in avidity for biotinylated 4MeMan fixed on a surface compared to its allylated precursor in solution (**Figure 2B**).

***In silico* docking of methylated glycans to Lb-Tec2**

In order to complement the structural information obtained by protein crystallography and to rationalize the binding of different methylated monosaccharides, *in silico* docking studies

were performed with four monosaccharides carrying an allyl aglycon as α-anomer: AllFuc, All2MeFuc, All3MeMan, All4MeMan. The predicted favored orientations of All2MeFuc or All4MeMan in the binding site are in agreement with the crystal structures (**Figure 2C**). AllFuc binds with the same orientation as All2MeFuc to the protein, but lacks the hydrophobic contact usually provided by the methyl group in the hydrophobic pocket (**Figure 2C**). No interaction is observed for the anomeric carbon, which is exposed to the solvent, and explains why both anomers can bind. It is unclear why Lb-Tec2 had a preference for the β-L-fucoside in the crystal since a 5:1 α/β ratio was obtained after synthesis and the β–anomer is usually not naturally found in fucosides (Sommer et al., 2016a). Binding of All3MeMan occurs in a different orientation of the ligand, but in analogy with the other ligands, the methyl group is recognized by the same hydrophobic pocket. In contrast, All3MeMan does not establish a hydrogen bond with the conserved tyrosine, but a weak hydrophobic contact between its hydroxymethyl and the aromatic ring (**Figure 2C**). Docking of GlcNAc led to weak interactions and the methyl group of the N-acetyl pointing towards the binding pocket. The pose is similar to the one observed in the FEL-GlcNAc X-ray structure as described below. The calculated binding energies, using the Generalized Born Solvent Accessible surface area (GBSA) procedure, are in agreement with experimentally determined affinities, predicting higher affinity for All2MeFuc, intermediate affinity for All4MeMan and All3MeMan, lower affinity for AllFuc, and very weak affinity for GlcNAc (**Table S4**).

Docking studies were also performed with the trisaccharide 2-*O*-methyl-L-fucose-α-1,2-D-galactose-β-1,4-D-mannose, a common motif in *C. elegans* N-glycans (Yan et al., 2015). Two clusters of poses were observed, differing mainly by ψ torsion of 2MeFuc-Gal glycosidic linkage in two different orientations which leads to two spatial presentations of the attached mannose residue (**Figure 2C**). Binding energies suggest that pose I is more favorable than pose II but the gain in affinity remains rather small compared to the monosaccharides (**Table**

S4). The galactose moiety does not show any contact with the protein and few hydrogen bonds are observed for the mannose one in each pose. This theoretical consideration is still to be experimentally verified by binding measurements with methylated oligosaccharides, but those are not yet available.

Structural homologs of Lb-Tec2

A search for structural homologs of Lb-Tec2 with DaliLite (Holm and Rosenstrom, 2010) revealed a variety of β-propeller proteins of different functions such as the lectin FEL from carp (Capaldi et al., 2015), followed by β-propellers computationally designed (Voet et al., 2014), the serine/threonine kinase PNKD from *Mycobacterium tuberculosis* (Good et al., 2004), several fucose specific lectins (Arnaud et al., 2014; Houser et al., 2013), enzymes such as a human sialidase (Chavas et al., 2005), and TolB proteins (Abergel et al., 1999). These proteins differ in size and orientation of their surface loops, in presence or absence of small α-helices and in blade orientations. The high plasticity observed for this fold enables its rich functional diversity (Chen et al., 2011).

The crystal structure of FEL was published during the refinement of Lb-Tec2 structures presented here. It displays strong structural similarity to Lb-Tec2 with an rmsd of 2 Å (Capaldi et al., 2015). Differences are mainly found for blade 6, which is shifted by about 2 Å towards the center in Lb-Tec2, and for the "bottom" surface loops, especially the one connecting strands C and D. They differ in size and conformation preventing the formation of a tetrameric structure in FEL (**Figure 3A**). FEL forms dimers via three surface loops at the bottom of the first and second blades. Both FEL and Lb-Tec2 bind a divalent cation in the central cavity, but in a distinct location with different contacts (**Figure 3A**). FEL was shown to recognize GlcNAc in two binding sites in the cocrystal structure (Capaldi et al., 2015). Compared to glycan complexes of Lb-Tec2, the pyranose ring of GlcNAc is shifted towards the solvent leading to fewer interactions. Apart from a direct hydrogen bond between the

oxygen of the N-acetyl group and a main chain nitrogen, all other interactions in binding sites of FEL are water mediated. One conserved water is found at the position where the O3 hydroxyl of 2MeSeFuc or All4MeMan is located. Interestingly, the methyl moiety of the N-acetyl group is found in the same location as the O-methyl groups described above (**Figure 3B**). Thus, FEL may display higher affinity for methylated glycans than for GlcNAc even though this has not been examined experimentally.

Interestingly, Lb-Tec2 is also structurally related to a family of fungal and bacterial lectins specific for fucosylated oligosaccharides found in histo-blood group antigens (rmsd of around 3 Å). The fungal members AFL from *Aspergillus fumigatus* and AOL from *A. oryzae* are dimeric and possess 12 binding sites on two opposite faces (Houser et al., 2013; Makyio et al., 2016). The bacterial members RSL from *Ralstonia solanacearum* and BambL from *Burkholderia ambifaria* are smaller, form β-propellers through trimerization, and exhibit six binding sites (Audfray et al., 2012; Kostlanova et al., 2005). In contrast to tectonins, their binding sites are situated near the propeller "bottom" at the interface between blades, and not on each blade towards the propeller "top" (**Figure 3C**). They specifically recognize fucose via a conserved Arg, Trp/Tyr, and Glu/Gln triad absent in tectonins. This again corroborates the high versatility of this fold.

Identification of Lb-Tec2 orthologs

Our observations suggest that methylation is a major recognition determinant for lectins of the tectonin family. Structural and functional orthologs of Lb-Tec2 were investigated and a Blast search on the NCBI website only provided hits of tectonins containing six TECPRs. Fungal and bacterial orthologs display high sequence identities (85-60%) whilst lower ones are observed for amoebozoan and metazoan orthologs, e.g., from horseshoe crab, lizard, frog, turtles, fishes, and human (40-32%, **Figures 3D** and **S5**). For Lb-Tec2, recognition of glycan

O-methylation required three consecutive TECPRs with the presence of six conserved amino acids in the TECPR repeat: Ile (-1), Gly, Leu, Asn, Tyr, and Trp (+1). Except for glycine, some mutations, e.g., Asn to Asp, or Trp/Tyr to Phe, may be tolerated as long as they do not modify the shape and the size of the binding pocket. Some substitutions, like Asn to Ser/Gly, are expected to influence the specificity and the affinity for carbohydrates due to the loss of important hydrogen bonds.

As shown in **Figure 3E**, a consensus sequence for the recognition of O-methylated glycans can thus be determined and served as signature for tectonins. It resembles the sequence motif described for TECPR repeat identification (Huh et al., 1998). Most orthologs are predicted to comprise six binding sites that are able to specifically recognize O-methylated glycans, apart from those of vertebrate origin, which may have a maximum of five functional binding sites (**Figures 3D** and **S5**). Their binding site 6 is indeed occluded by an insertion and has a very different sequence, as observed in FEL (Capaldi et al., 2015). Some non-vertebrate members may also comprise one or two non-functional sites with variation from one to the other. As a result of many non-conservative mutations, only site 5 is expected to be functional for the predicted fish egg lectin from *Oplegnathus fasciatus* (Uniprot H3JU40). The specificity for O-methylated glycans was demonstrated experimentally for the limulus L6 lectin, which can now be explained by the presence of six functional binding sites (Wohlschlager et al., 2014). In human tectonins, only leukolectin (Walther and Miftari, 2010) may recognize O-methylated glycans in binding sites 2, 3, 4 and 5 since TECPR1 (Ogawa et al., 2011) and TECPR2 (Oz-Levi et al., 2012) do not contain the required sequence in any of their TECPR repeat (data not shown). It seems that complex TECPR proteins, e.g., of human origin and their homologs from *D. melanogaster* and *C. elegans,* which are very distant orthologs of Lb-Tec2 (<8% sequence identity), may recognize a variety of other ligands that remain to be identified.

Discussion

Lectins exhibit very diverse folds and binding site architectures to mirror the diversity of glycans. Since the lectin repertoire of an organism is limited, some lectins evolved to recognize epitopes that are present on several glycans. The present study describes the molecular basis for the recognition of O-methylated glycan by tectonins, a protein family involved in innate immunity, such as Lb-Tec2 and limulus L6 (Wohlschlager et al., 2014). Similar to terminal exocyclic 1,2-diols on microbial glycans being recognized by intelectins (Wesener et al., 2015), methylation seems to be the main determinant for the recognition of glycans by Lb-Tec2 and several others tectonins. As demonstrated here, this modification can be recognized at very different positions of the pyranose ring in L- and D-glycans (**Figure 4A**). Recognition is achieved by a hydrophobic pocket formed by highly conserved amino acids, defining a signature for this specificity (**Figure 3E**). Lb-Tec2 crystal structures revealed the requirement for three consecutive TECPR repeats to form the pocket partly as a result of the exchange of strand D between the blades of the β-propeller. One conserved asparagine is responsible for glycan specificity since it is the only residue involved in hydrogen bonding with the glycan. An equatorial hydroxyl group neighboring the methylated position is necessary for recognition and defines the range of possible ligands. Upon binding, almost half of the glycan is exposed to the solvent, which explains the low affinity determined for monosaccharides. Our results confirm previous findings and show conservation of the specificity for methylated glycans in orthologs with six TECPR repeats from bacteria to human (with leukolectin) (**Figures 3D and S5**). We propose that these glycoepitopes are an important target for innate immune systems since they are widely distributed PAMPs and may broaden the spectrum of targeted organisms (Staudacher, 2012; Wohlschlager et al., 2014). On the other hand, complex tectonins have evolved to recognize other motifs and to

participate to more complex systems such as autophagy. It may reflect the fact that their organism is confronted with a higher repertoire of invaders.

The unexpected tetrameric quaternary arrangement unveiled for Lb-Tec2 results in the distribution of 24 binding sites around its entire surface (**Figure 4B**). Extensive enhancement of binding upon multivalent interactions, *i.e.,* avidity, is a well-known phenomenon for both lectin-glycan and antibody-antigen interactions (Fasting et al., 2012). With 24 binding sites, Lb-Tec2 achieves one of the highest valency ever described for a lectin: only the heptameric hemolytic lectin CEL-III from sea cucumber, depicted in Figure 4B, displays higher valency with 35 carbohydrates binding sites (five per monomer) sitting on the flat pore forming face of the heptamer (Unno et al., 2014). Other toxic lectins, exhibiting high valency, include the dodecavalent mannose specific lectin from snowdrop GNA (Hester et al., 1995) and decameric galactose specific C-type lectin from rattlesnake (Walker et al., 2004). These lectins are also proposed to play a role in innate immunity but differ in the distribution of their binding sites: only part of their surface appears to be dedicated to carbohydrate recognition, whereas in case of Lb-Tec2 the entire protein surface is occupied by binding sites (**Figure 4B**). Based on the conservation of amino acid residues involved in oligomerization, most of the fungal and bacterial orthologs of Lb-Tec2 are also expected to adopt analogous tetrameric structures but actual glycan specificities and the biological functions of these proteins remain to be uncovered.

Multivalency is often essential to the biological function of lectins and can be achieved by oligomerization, the presence of multiple tandem repeats, or by a mix of the two, as observed here for Lb-Tec2 (Taylor et al., 2015). For example, multivalency is critical for the toxicity of the lectins CCL2 from the mushroom *Coprinopsis cinerea* and Griffithsin from red algae (Bleuler-Martinez et al., 2017; Lusvarghi and Bewley, 2016). Their toxicity is proposed to result from carbohydrate mediated crosslinking and hence, aggregation of molecules as well

as spatial rearrangement of glycoconjugates, disturbing underlying cell membranes (Lusvarghi and Bewley, 2016). Indeed, differences in binding site distribution, specificity and valency directly impacts lectin biotoxicity (Bleuler-Martinez et al., 2017). Intriguingly, the ball shaped tetramer of Lb-Tec2, with its binding sites symmetrically spread on each monomer, resembles a small viral capsid. This architecture makes the protein a perfect surveillance molecule, turning it into a formidable weapon. On the other hand, Lb-Tec2 has been proposed to play a role in mycorrhiza formation as well as in mutualistic interactions with mycorrhizae helper bacteria (Deveau et al., 2007). In this context, investigation of the methylated glycan structures recognized in these processes would be of great interest.

Molecular interactions between lectins and their target glycans play a fundamental role in many biological recognition processes, e.g. organismic interactions, and thus, specificity of lectins and diversity of glycans evolve in concert (Springer and Gagneux, 2013). Our study provides molecular insights into an unusual type of lectin-glycan interaction. The binding site architecture described here is dedicated to the exclusive recognition of methylation on glycans and its signature can be used to classify carbohydrate binding proteins into the tectonin family.

ACKNOWLEDGMENTS

The authors thank Emilie Gillon for performing affinity measurements and Dr. Anne Imberty for insightful comments and discussions. A.V. is grateful to Dr. J. Turkenburg and Prof. E. Dodson for helpful advices for space group assignment. The authors are grateful to European Synchrotron Radiation Facility, Grenoble, France for technical support and access to beamlines BM30A-FIP, BM14, ID30-A and BM29 and to Martha Brennich for helpful discussions. Dr. Michael Hoffmann and Dr. Kirsten Harmrolfs are acknowledged for HRMS and NMR measurement of compound **3**. We further thank the Helmholtz Association (to AT, grant no VH-NG-934), the European Commission Marie Curie Action ‘‘Networks for Initial

Training'' (to T.W., grant FP7/2007-2013-EuroGlycoArrays 215536), and the Swiss National Science Foundation (to MK, grant no 31003A-130671) for financial support.

AUTHOR CONTRIBUTIONS

R.S. performed the synthesis and analysis of the ligands and O.N.M the modelling studies. T.W. and A.V. produced and purified the recombinant protein, S.H. treated the SAXS data and calculated the resulting envelope, M.M. performed the initial crystallization trials and crystal testing. A.V. performed the crystallographic experiments and analyzed data. A.T., M.K. and A.V. conceived and supervised the project. A.V. wrote the manuscript with contributions from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Abergel, C., Bouveret, E., Claverie, J.-M., Brown, K., Rigal, A., Lazdunski, C., and Bénédetti, H. (1999). Structure of the *Escherichia coli* TolB protein determined by MAD methods at 1.95 Å resolution. Structure *7*, 1291-1300.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

Arnaud, J., Tröndle, K., Claudinon, J., Audfray, A., Varrot, A., Römer, W., and Imberty, A. (2014). Membrane Deformation by Neolectins with Engineered Glycolipid Binding Sites. Angew. Chem. Int. Ed. Engl**.** *53*, 9267-9270.

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

Beshr, G., Sommer, R., Hauck, D., Siebert, D.C.B., Hofmann, A., Imberty, A., and Titz, A. (2016). Development of a competitive binding assay for the *Burkholderia cenocepacia* lectin

BC2L-A and structure activity relationship of natural and synthetic inhibitors. MedChemComm *7*, 519-530.

Bleuler-Martinez, S., Stutz, K., Sieber, R., Collot, M., Mallet, J.-M., Hengartner, M., Schubert, M., Varrot, A., and Künzler, M. (2017). Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against nematodes. Glycobiology *27*, 486-500.

Brubaker, S.W., Bonham, K.S., Zanoni, I., and Kagan, J.C. (2015). Innate immune pattern recognition: a cell biological perspective. Annu. Rev. Immunol. *33*, 257-290.

Capaldi, S., Faggion, B., Carrizo, M.E., Destefanis, L., Gonzalez, M.C., Perduca, M., Bovi, M., Galliano, M., and Monaco, H.L. (2015). Three-dimensional structure and ligand-binding site of carp fishelectin (FEL). Acta Crystallogr. D Biol. Crystallogr. *71*, 1123-1135.

Chaudhuri, I., Soding, J., and Lupas, A.N. (2008). Evolution of the beta-propeller fold. Proteins *71*, 795-803.

Chavas, L.M., Tringali, C., Fusi, P., Venerando, B., Tettamanti, G., Kato, R., Monti, E., and Wakatsuki, S. (2005). Crystal structure of the human cytosolic sialidase Neu2. Evidence for the dynamic nature of substrate recognition. J. Biol. Chem. *280*, 469-475.

Chen, C.K., Chan, N.L., and Wang, A.H. (2011). The many blades of the beta-propeller proteins: conserved but versatile. Trends Biochem. Sci. *36*, 553-561.

Chen, S.C., Yen, C.H., Yeh, M.S., Huang, C.J., and Liu, T.Y. (2001). Biochemical properties and cDNa cloning of two new lectins from the plasma of *Tachypleus tridentatus*: Tachypleus plasma lectin 1 and 2+. J. Biol. Chem. *276*, 9631-9639.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12-21.

Dambuza, I.M., and Brown, G.D. (2015). C-type lectins in immunity: recent developments. Curr. Opin. Immunol. *32*, 21-27.

Deveau, A., Palin, B., Delaruelle, C., Peter, M., Kohler, A., Pierrat, J.C., Sarniguet, A., Garbaye, J., Martin, F., and Frey-Klett, P. (2007). The mycorrhiza helper *Pseudomonas fluorescens* BBc6R8 has a specific priming effect on the growth, morphology and gene

expression of the ectomycorrhizal fungus *Laccaria bicolor* S238N. New Phytol. *175*, 743-755.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486-501.

Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D Biol. Crystallogr. *67*, 282-292.

Fasting, C., Schalley, C.A., Weber, M., Seitz, O., Hecht, S., Koksch, B., Dernedde, J., Graf, C., Knapp, E.W., and Haag, R. (2012). Multivalency as a chemical organization and action principle. Angew. Chem. Int. Ed. Engl**.** *51*, 10472-10498.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. *42*, 342-346.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K.*, et al.* (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J. Med. Chem. *47*, 1739-1749.

Good, M.C., Greenstein, A.E., Young, T.A., Ng, H.-L., and Alber, T. (2004). Sensor Domain of the *Mycobacterium tuberculosis* Receptor Ser/Thr Protein Kinase, PknD, forms a Highly Symmetric β Propeller. J. Mol. Biol. *339*, 459-469.

Gottlieb, H.E., Kotlyar, V., and Nudelman, A. (1997). NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. J. Org. Chem. *62*, 7512-7515.

Hester, G., Kaku, H., Goldstein, I.J., and Wright, C.S. (1995). Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. Nat. Struct. Mol. Biol. *2*, 472-479.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Carbohydr. Res. *38*, W545-549.

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

Huh, C.G., Aldrich, J., Mottahedeh, J., Kwon, H., Johnson, C., and Marsh, R. (1998). Cloning and characterization of *Physarum polycephalum* tectonins. Homologues of Limulus lectin L-6. J. Biol. Chem. *273*, 6565-6574.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125-132.

Knight, C.J., and Hub, J.S. (2015). WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. Nucleic Acids Res. *43*, W225-230.

Kostlanova, N., Mitchell, E.P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerova, M., and Imberty, A. (2005). The fucose-binding lectin from *Ralstonia solanacearum*. A new type of beta-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. J. Biol. Chem. *280*, 27839-27849.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774-797.

Kumar, H., Kawai, T., and Akira, S. (2011). Pathogen recognition by the innate immune system. Int. Rev. Immunol. *30*, 16-34.

Künzler, M. (2015). Hitting the sweet spot-glycans as targets of fungal defense effector proteins. Molecules *20*, 8144-8167.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3*, 1171-1179.

Low, D.H., Ang, Z., Yuan, Q., Frecer, V., Ho, B., Chen, J., and Ding, J.L. (2009). A novel human tectonin protein with multivalent beta-propeller folds interacts with ficolin and binds bacterial LPS. PLoS One *4*, e6260.

Low, D.H., Frecer, V., Le Saux, A., Srinivasan, G.A., Ho, B., Chen, J., and Ding, J.L. (2010). Molecular interfaces of the galactose-binding protein Tectonin domains in host-pathogen interaction. J. Biol. Chem. *285*, 9898-9907.

Lusvarghi, S., and Bewley, C.A. (2016). Griffithsin: An Antiviral Lectin with Outstanding Therapeutic Potential. Viruses *8*, 296.

Makyio, H., Shimabukuro, J., Suzuki, T., Imamura, A., Ishida, H., Kiso, M., Ando, H., and Kato, R. (2016). Six independent fucose-binding sites in the crystal structure of *Aspergillus oryzae* lectin. Biochem. Biophys. Res. Commun. *477*, 477-482.

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. *63*, 32-41.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355-367.

Nurisso, A., Kozmon, S., and Imberty, A. (2008). Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III. Mol. Simul. *34*, 469-479.

Ogawa, M., Yoshikawa, Y., Kobayashi, T., Mimuro, H., Fukumatsu, M., Kiga, K., Piao, Z., Ashida, H., Yoshida, M., Kakuta, S.*, et al.* (2011). A Tecpr1-dependent selective autophagy pathway targets bacterial pathogens. Cell Host Microbe *9*, 376-389.

Olsen, L.R., Kudahl, U.J., Simon, C., Sun, J., Schönbach, C., Reinherz, E.L., Zhang, G.L., and Brusic, V. (2013). BlockLogo: visualization of peptide and sequence motif conservation. J. Immunol. Methods *400-401*, 37-44.

Oz-Levi, D., Ben-Zeev, B., Ruzzo, E.K., Hitomi, Y., Gelman, A., Pelak, K., Anikster, Y., Reznik-Wolf, H., Bar-Joseph, I., Olender, T.*, et al.* (2012). Mutation in TECPR2 Reveals a Role for Autophagy in Hereditary Spastic Paraparesis. J. Hum. Genet. *91*, 1065-1072.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M.*, et al.* (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J. Synchrotron Radiat. *20*, 660-664.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Crystallogr. *45*, 342-350.

Rambo, R.P., and Tainer, J.A. (2011). Characterizing Flexible and Instrinsically Unstructured Biological Macromolecules by SAS using the Porod-Debye Law. Biopolymers *95*, 559-571.

Repasky, M.P., Shelley, M., and Friesner, R.A. (2002). Flexible Ligand Docking with Glide. In Current Protocols in Bioinformatics (John Wiley & Sons, Inc.).

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320-324.

Rose, P.W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Costanzo, L.D., Duarte, J.M., Dutta, S., Feng, Z.*, et al.* (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res. *45*, D271-D281.

Round, A., Brown, E., Marcellin, R., Kapp, U., Westfall, C.S., Jez, J.M., and Zubieta, C. (2013). Determination of the GH3.12 protein conformation through HPLC-integrated SAXS measurements combined with X-ray crystallography. Acta Crystallogr. D Biol. Crystallogr. *69*, 2072-2080.

Saito, T., Kawabata, S., Hirata, M., and Iwanaga, S. (1995). A novel type of limulus lectin-L6. Purification, primary structure, and antibacterial activity. J. Biol. Chem. *270*, 14493-14499.

Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. *58*, 1772-1779.

Schröder, H.C., Ushijima, H., Krasko, A., Gamulin, V., Thakur, N.L., Diehl-Seifert, B., Muller, I.M., and Müller, W.E. (2003). Emergence and disappearance of an immune molecule, an antimicrobial lectin, in basal metazoa. A tachylectin-related protein in the sponge *Suberites domuncula*. J. Biol. Chem. *278*, 32810-32817.

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 479-485.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Sommer, R., Hauck, D., Varrot, A., Imberty, A., Künzler, M., and Titz, A. (2016a). O-Alkylated heavy atom carbohydrate probes for protein X-ray crystallography: Studies towards the synthesis of methyl 2-O-methyl-L-selenofucopyranoside. Beilstein J. Org. Chem. *12*, 2828-2833.

Sommer, R., Wagner, S., Varrot, A., Nycholat, C.M., Khaledi, A., Haussler, S., Paulson, J.C., Imberty, A., and Titz, A. (2016b). The virulence factor LecB varies in clinical isolates: consequences for ligand binding and drug discovery. Chem. Sci. *7*, 4990-5001.

Springer, S.A., and Gagneux, P. (2013). Glycan evolution in response to collaboration, conflict, and constraint. J. Biol. Chem. *288*, 6904-6911.

Staudacher, E. (2012). Methylation - an uncommon modification of glycans. Biol. Chem. *393*, 675-685.

Taylor, M.E., Drickamer, K., Schnaar, R.L., Etzler, M.E., and Varki, A. (2015). Discovery and Classification of Glycan-Binding Proteins. In Essentials of Glycobiology, A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, and P.H. Seeberger, eds. (Cold Spring Harbor (NY)).

Tina, K.G., Bhadra, R., and Srinivasan, N. (2007). PIC: Protein Interactions Calculator. Nucleic Acids Res. *35*, W473-476.

Turnbull, W.B., and Daranas, A.H. (2003). On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? J. Am. Chem. Soc. *125*, 14859-14866.

Unno, H., Goda, S., and Hatakeyama, T. (2014). Hemolytic Lectin CEL-III Heptamerizes via a Large Structural Transition from α-Helices to a β-Barrel during the Transmembrane Pore Formation Process. J. Biol. Chem. *289*, 12805-12812.

Varrot, A., Basheer, S.M., and Imberty, A. (2013). Fungal lectins: structure, function and potential applications. Curr. Opin. Struct. Biol. *23*, 678-685.

Voet, A.R.D., Noguchi, H., Addy, C., Simoncini, D., Terada, D., Unzai, S., Park, S.-Y., Zhang, K.Y.J., and Tame, J.R.H. (2014). Computational design of a self-assembling symmetrical β-propeller protein. Proc. Natl. Acad. Sci. USA *111*, 15102-15107.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Crystallogr. *36*, 860-864.

Walker, J.R., Nagar, B., Young, N.M., Hirama, T., and Rini, J.M. (2004). X-ray Crystal Structure of a Galactose-Specific C-Type Lectin Possessing a Novel Decameric Quaternary Structure. Biochemistry *43*, 3783-3792.

Walther, B.T., and Miftari, M. (2010). Leukolectins and uses thereof. PCT/GB2009/002569

Wesener, D.A., Wangkanont, K., McBride, R., Song, X., Kraft, M.B., Hodges, H.L., Zarling, L.C., Splain, R.A., Smith, D.F., Cummings, R.D.*, et al.* (2015). Recognition of microbial glycans by human intelectin-1. Nat. Struct. Mol. Biol. *22*, 603-610.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235-242.

Wohlschlager, T., Butschi, A., Grassi, P., Sutov, G., Gauss, R., Hauck, D., Schmieder, S.S., Knobel, M., Titz, A., Dell, A.*, et al.* (2014). Methylated glycans as conserved targets of animal and fungal innate defense. Proc. Natl. Acad. Sci. USA *111*, E2787-2796.

Yan, S., Brecker, L., Jin, C., Titz, A., Dragosits, M., Karlsson, N.G., Jantsch, V., Wilson, I.B., and Paschinger, K. (2015). Bisecting Galactose as a Feature of N-Glycans of Wild-type and Mutant *Caenorhabditis elegans*. Mol. Cell Proteomics *14*, 2111-2125.

Zipfel, C. (2014). Plant pattern-recognition receptors. Trends Immunol. *35*, 345-351.

Main figure titles and legends

Figure 1. Tertiary and quaternary structure of Lb-Tec2

(A) Top and side views of Lb-Tec2 β-propeller colored by TECPR motifs. 2MeSeFuc is depicted in balls and sticks. Numbers correspond to the different blades from *N*- to *C*-terminus and their respective binding sites. Letters describe the nomenclature for β-strand in blade.

(B) Zoom on the central channel of Lb-Tec2 β-propeller. Surface representation of the solvent filling the channel. The residues blocking its exit are depicted as balls and sticks with the carbon atoms colored in blue for Arg12, green for Arg67, pale green for Asp125, cyan for Asp50, and orange for Asp203. Magnesium ion is shown as a sphere (see Figure S1A).

(C) Representation of the trimeric base of Lb-Tec2 tetramer.

(D) Cartoon representations of the Lb-Tec2 tetramer in its three statistical orientations perpendicular or across the three-fold axis with description of rotation axis (left) and its size (right).

(E) Overlay of the SAXS envelope with the tetrameric crystallographic model of Lb-Tec2 (see also Figure S1 and Table S1).

(F) Topology diagram of Lb-Tec2 with numbered β-strands from *N* to *C*-terminus with the same color code as in A. Shaded box represent oligomerization interfaces with interface I, II and III colored in blue, green and beige respectively (see Table S2). Dotted areas localize the carbohydrate binding sites.

Figure 2. Lb-Tec2 binding to O-methylated glycans

(A) Zoom on the interactions in binding site 4. $2F_{obs}–F_{calc}$ electron density around 2MeSeFuc (left) and All4MeMan (right) displayed at 1 σ (0.51 and 0.49 eA^3, respectively). Surface view of the methylation recognition pocket for 2MeSeFuc. Hydrogen bonds are represented by dotted lines. Residues are colored by TECPR motifs (see Figure S2A and Table S3).

(B) Ligand synthesis and SPR experiments. Synthesis scheme of biotinylated 4-O-methyl-α-D-mannopyranoside (**3**) from All4MeMan (**1**). Sensorgrams representing binding of Lb-Tec2 to 4MeMan for a protein concentration range of 0 to 300 μM (middle) and steady state affinity fit (bottom). See also Figures S3 and S4.

(C) In silico docking to Lb-Tec2 (see Table S3). Surface representation of the docking orientation of monosaccharides in Lb-Tec2 binding site: All2MeFuc, AllFuc, All4MeMan, All3MeMan and βGlcNAc as annotated (full name given in Table S5). Overlay of the lowest energy representatives of the two clusters of poses obtained after docking of a common

methylated trisaccharide in nematodes: *O*-methyl-L-fucose-α-1,2-D-galactose-β-1,4-D-mannose (bottom right). Pose I is colored in blue.

Figure 3. Structural and sequence homologies of Lb-Tec2

(A) Views of the overlay of Lb-Tec2 (5FCB, green) and FEL (4RUS, purple) β-propellers from the propeller top (left) and side (right). Cations are depicted as spheres.

(B) Stereo view of the conserved interacting residues in the overlay of binding site 3 of Lb-Tec2 (PDB ID: 5FCB, green) and FEL (PDB ID: 4RUS, purple) structures.

(C) Overlay of blades 4 and 5 from the β-propellers of Lb-Tec2 (PDB ID: 5FCB, green and yellow) and AFL (PDB ID: 4AGI, grey and black) demonstrating conservation of the overall fold but not of the location of the binding sites.

(D) Sequence alignment to identify methylated glycan-binder candidate in the tectonin family (see Figure S5). The proteins are named according to the Uniprot number (Lb-Tec2: B0CZL6_LACBS, L6: LEC6_TACTR, FEL: FEL_CYPCA and Leukolectin: C0KWF0_human) and are colored by origin (fungi: black, bacteria: red, sponge and horseshoe crab: green, fish: orange and human: brown). Residues involved in methylated glycans recognition are represented by colored symbol according to the binding site number (1: magenta, 2: blue, 3: green, 4: red, 5: cyan and 6: orange). Figure drawn with ESPript 3.0 (Robert and Gouet, 2014).

(E) Conservation of the amino acids involved in the recognition of O-methylated glycans in the TECPR repeat depicted with Blocklogo (Olsen et al., 2013).

Figure 4. Recognition of methylated glycans by Lb-Tec2 binding pocket and highly multivalent lectins

(A) Stereo view of the overlay of the models of Lb-Tec2 in complex with methylated monosaccharides obtained by X-ray crystallography and docking studies. Zoom in the binding site 3 with ligands depicted in balls and sticks and colored on carbon atoms: 2MeSeFuc (green), 4MeMan (blue) 3MeMan (orange) and GlcNAc (purple). Allyl aglycon has been removed from mannosides for clarity.

(B) Top and side surface representations of highly multivalent lectins with their binding site distribution. From left to right: Lb-Tec2 (PDB ID 5FCB), CEL-III from sea cucumber with lactulose (PDB ID 3W9T) (Unno et al., 2014), GNA from *Galanthus nivalis* with α-methyl-mannoside (PDB ID 1MSA) (Hester et al., 1995) and the rattlesnake venom galactose-specific C-type lectin (RSL) with lactose (PDB ID 1JZN) (Walker et al., 2004).

Table 1. Data Collection and Refinement Statistics for Lb-Tec2 structures

	2MeSeFuc			All4MeMan	
PDBcode	5FSB			5FSC	
Data collection					
Beamline ESRF	BM30A			BM14	
Space group	R32:h (H32)			R32:h (H32)	
Cell dimensions					
a, *b*, *c* (Å)	69.06 69.06 349.21			69.00 69.00 348.95	
α, β, γ (°)	90 90 120			90 90 120	
	Peak	*Inflection*	*Remote*		
Wavelength	0.98021	0.98041	0.97914	0.97625	
Resolution (Å)	34.53-1.85 (1.89-1.85)	34.53-1.85 (1.89-1.85)	34.53-1.65 (1.68-1.65)	38.77-1.95 (2.00-1.95)	
Measured reflections	300555	300740	419929	183666	
Unique reflections	28130	28131	39383	24046	
R_{merge}[a]	0.124 (0.465)	0.130 (0.488)	0.093 (0.432)	0.143 (0.412)	
Mean I / σI[a]	16.8 (5.4)	16.4 (5.3)	19.5 (5.4)	12.6 (4.8)	
Completeness (%)[a]	100 (100)	100 (100)	100 (100)	100 (100)	
Redundancy[a]	10.7 (10.9)	10.7 (10.9)	10.7 (10.8)	7.6 (7.7)	
$CC_{1/2}$[a]	0.997 (0.934)	0.997 (0.920)	0.998 (0.939)	0.994 (0.923)	
Mid-Slope of Anom Normal Probability	1.349	1.062	1.198	0.97625	
Refinement					
Resolution (Å)	34.53-1.65			38.77-1.95	
No. of work/ free Reflections	37397 / 1983			22904 / 1141	
R_{work} / R_{free}[b] (%)	13.3 / 16.7			14.3 / 18.3	
RMSD					
Bond lengths (Å)	0.015			0.017	
Bond angles (°)	1.70			1.74	
No. of atoms	Chain A	Chain B		Chain A	Chain B
Protein	1747	1651		1747	1649
Ligand	65	78		48	96
Water	287	238		329	188
Average *B*-factors	Chain A	Chain B		Chain A	Chain B
Protein	7.4	8.4		6.0	7.5
Ligand	15.6	14.0		13.6	13.7
Water	20.5	21.8		18.5	18.2
Ramachandran plot (%)					
Favored	97.6			94.6	
Outliers	0			0	

[a]Values in parentheses are for highest-resolution shell. [b] R_{free} was calculated with 5% of the reflections.

Table 2: Affinity measurements by isothermal microcalorimetry.

Ligand	K_d (mM)
AllFuc* [1]	74.9
All2MeFuc* [1]	4.4 ± 1.5
2MeSeFuc*	6.0 ± 0.3
All4MeMan*	16.6 ± 0.7
All3MeMan* [1]	20.6 ± 0.8
GlcNAc*	140.0

Thermograms and corresponding integrated titration curves are available in Figures S2B-D.
Errors indicate deviation of two independent titrations.
***** Full name of the ligand is given in table S5.
[1]: value taken from (Wohlschlager et al., 2014).

Figure 1

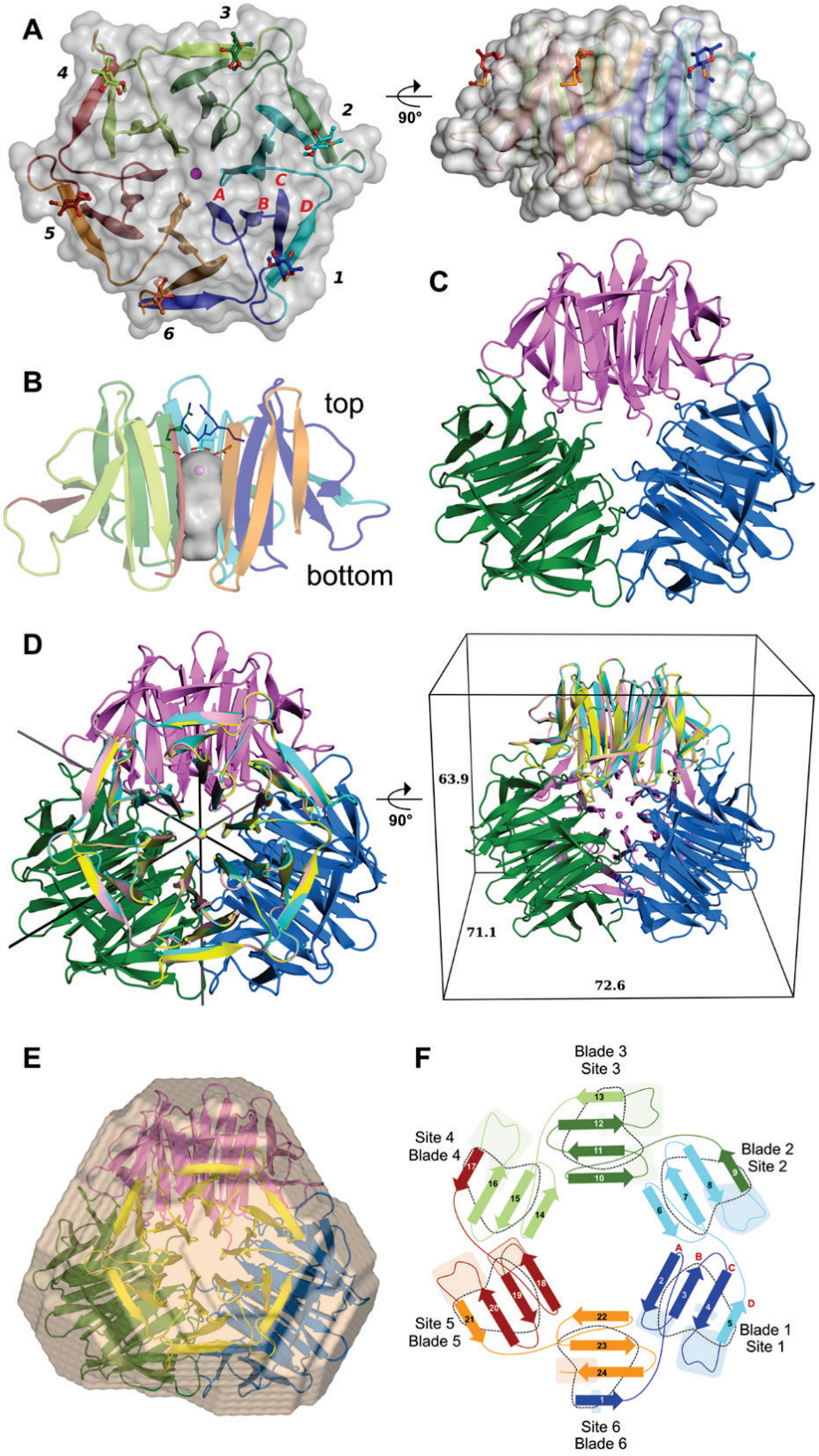

A
3
4
2
C
A
B
D
5
1
6
90°
B
top
bottom
C
D
63.9
90°
71.1
72.6
E
F
Blade 3
Site 3
Site 4
Blade 4
Blade 2
Site 2
Site 5
Blade 5
Blade 1
Site 1
Site 6
Blade 6

Figure 2

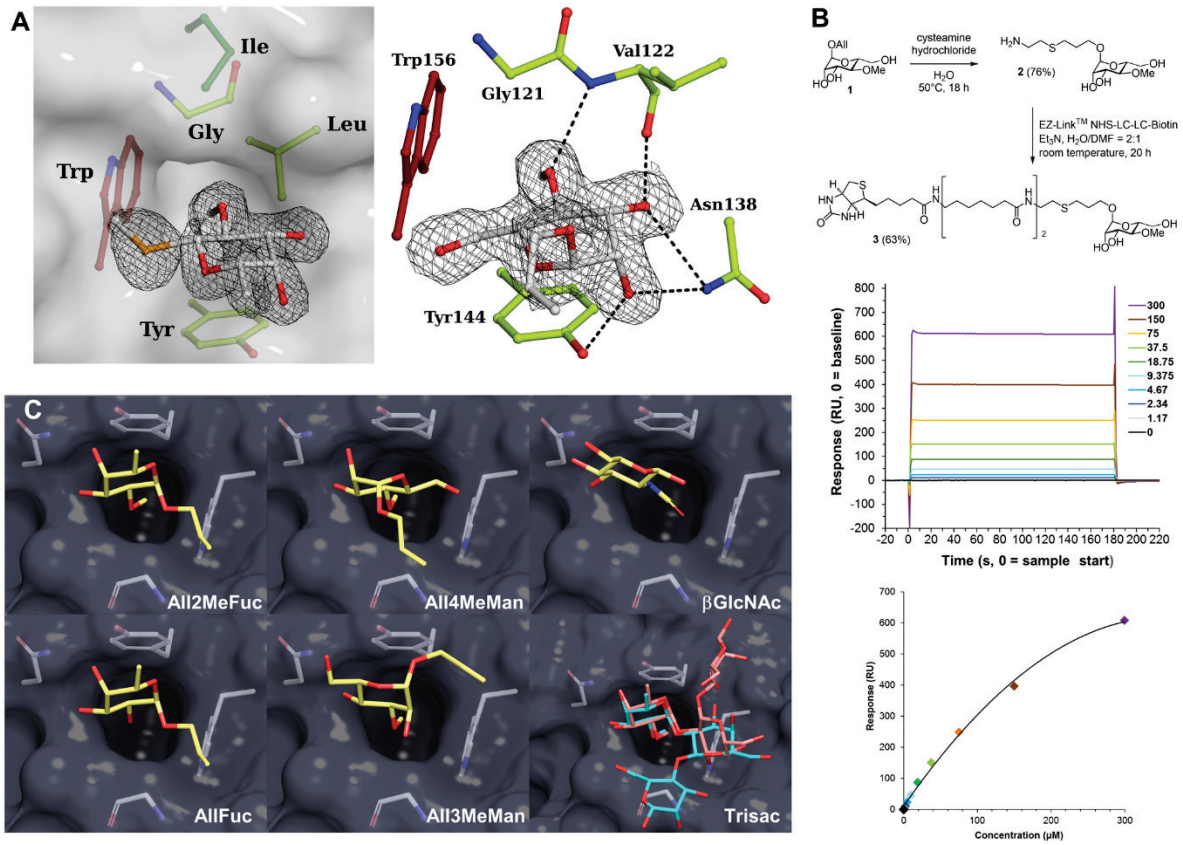

A
Ile
Gly
Leu
Trp
Tyr
Trp156
Gly121
Val122
Asn138
Tyr144
B
cysteamine hydrochloride
H2O
50°C, 18 h
2 (76%)
EZ-Link™ NHS-LC-LC-Biotin
Et3N, H2O/DMF = 2:1
room temperature, 20 h
3 (63%)
Response (RU, 0 = baseline)
Time (s, 0 = sample start)
300
150
75
37.5
18.75
9.375
4.67
2.34
1.17
0
Response (RU)
Concentration (μM)
C
All2MeFuc
All4MeMan
βGlcNAc
AllFuc
All3MeMan
Trisac

Figure 3

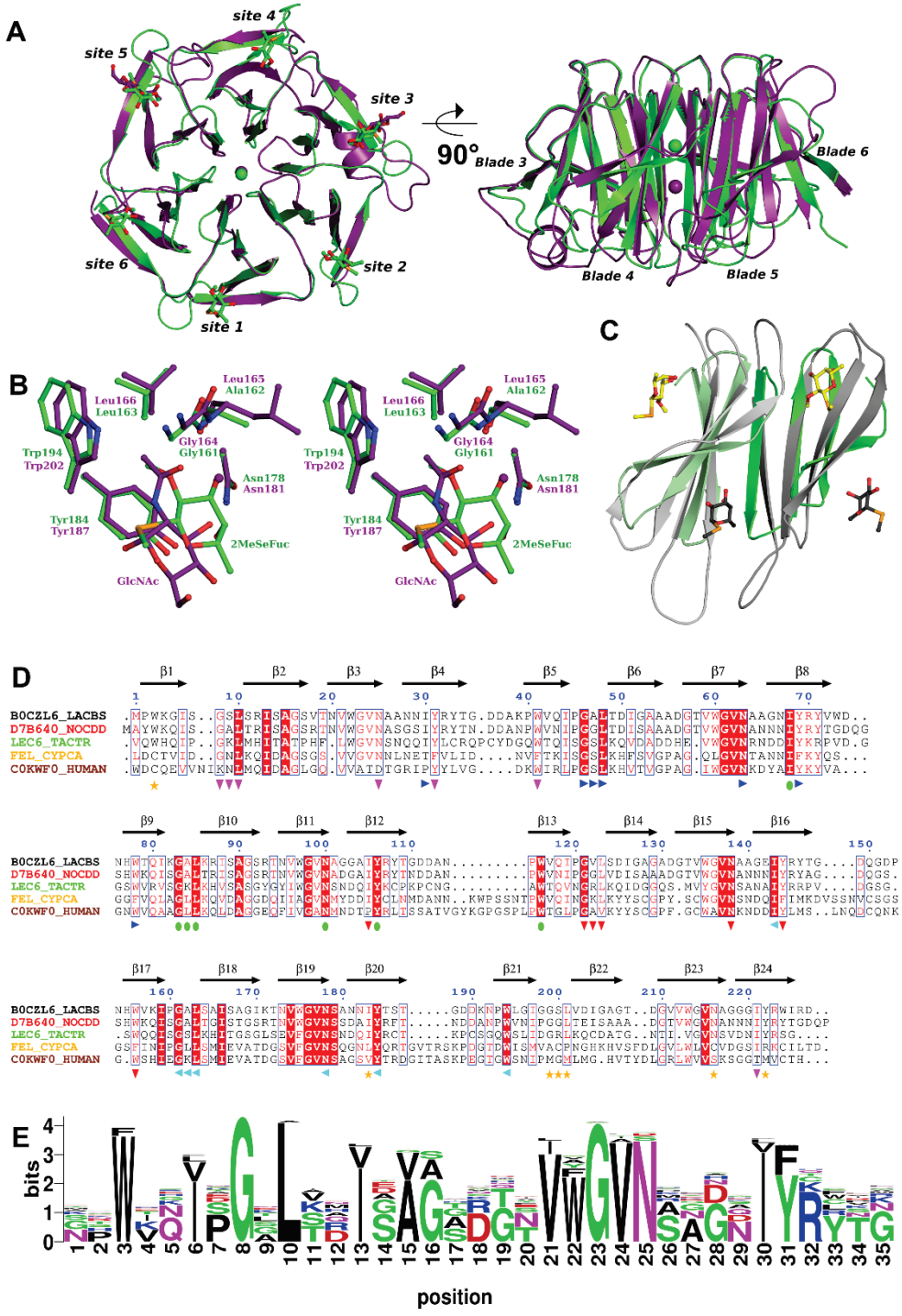
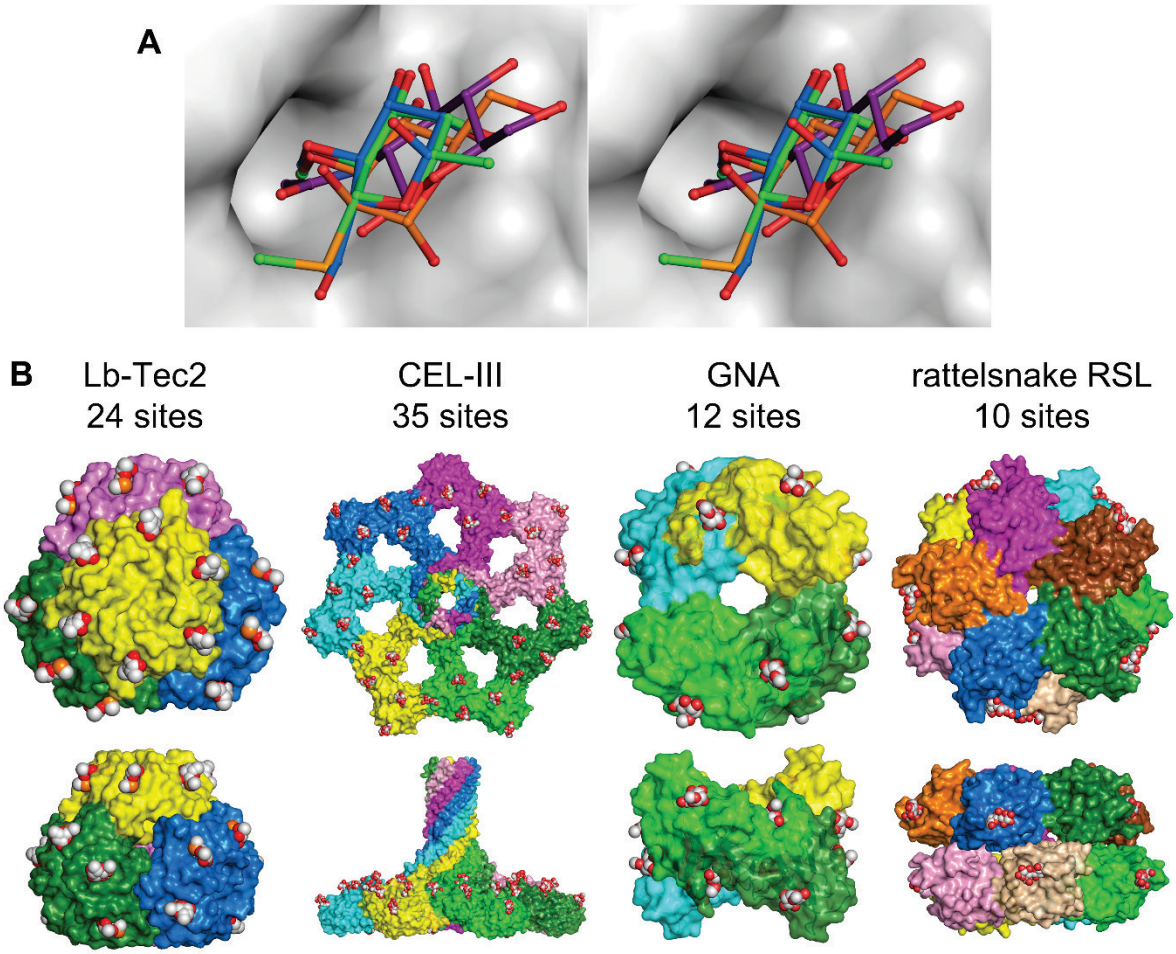

A
site 4
site 5
site 3
site 6
site 2
site 1
90°
Blade 3
Blade 6
Blade 4
Blade 5
B
Leu166
Leu163
Leu165
Ala162
Gly164
Gly161
Trp194
Trp202
Asn178
Asn181
Tyr184
Tyr187
2MeSeFuc
GlcNAc
C
D
β1
β2
β3
β4
β5
β6
β7
β8
B0CZL6_LACBS
D7B640_NOCDD
LEC6_TACTR
FEL_CYPCA
C0KWF0_HUMAN
β9
β10
β11
β12
β13
β14
β15
β16
β17
β18
β19
β20
β21
β22
β23
β24
E
bits
position

Figure 4


A
B
Lb-Tec2
24 sites
CEL-III
35 sites
GNA
12 sites
rattelsnake RSL
10 sites

STAR METHODS

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Bacterial Strains		
BL21(DE3)	Merk-Novagen	69450
Chemicals and Recombinant Protein		
Biacore SPR CM5 sensor chips	GE Healthcare	BR-1000-12
Enrich^{TM} SEC 70	Bio-Rad Ltd	7801070
Cryoloops	Molecular Dimensions Ltd	MD7-133 / MD7-134
GlcNAc	Carbosynth	MA00834
Streptavidin	Sigma Aldrich	85878
EZ-Link NHS-$(LC)_2$-Biotin	Thermo Scientific	21343
Structure Screen 2	Molecular Dimension Ltd	MD1-02
Lb-Tec2	(Wohlschlager et al., 2014)	N/A
Deposited Data		
PDB accession code	This paper	5FSB, 5FSC
Recombinant DNA		
Lb-Tec2 in pET24 expression vector	(Wohlschlager et al., 2014)	N/A
Software and Algorithms		
XDS package	(Kabsch, 2010)	http://xds.mpimf-heidelberg.mpg.de/
CCP4i	(Winn et al., 2011)	http://www.ccp4.ac.uk/
ShelXC/D/E	(Sheldrick, 2010)	http://shelx.uni-goettingen.de/
COOT	(Emsley et al., 2010)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
ATSAS software package	(Petoukhov et al., 2012)	https://www.embl-hamburg.de/biosaxs/software.html
Molprobity	(Chen et al., 2010)	http://molprobity.biochem.duke.edu/
wwPDB Validation server		http://wwpdb-validation.wwpdb.org
Glide and Prime	Schrödinger	https://www.schrodinger.com/
Dalilite	(Holm and Rosenstrom, 2010)	http://ekhidna.biocenter.helsinki.fi/dali_server/
Blast	(Altschul et al., 1990)	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Clustal omega	(Sievers et al., 2011)	https://www.ebi.ac.uk/Tools/msa/clustalo/
ESPript 3.0	(Robert and Gouet, 2014)	http://espript.ibcp.fr/ESPript/ESPript/
PyMOL	Schrödinger	https://pymol.org

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Annabelle Varrot (annabelle.varrot@cermav.cnrs.fr).

METHODS DETAILS

Protein crystallization and data collection.

Lb-Tec2 was expressed and purified as described previously (Wohlschlager et al., 2014). The protein, in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5 and 150 mM NaCl, was concentrated to about 5 mg mL^{-1} by centrifugation on Vivaspin500 column with a 5kDa cut off (Sartorius). The buffer was exchanged to 20 mM Tris pH 8.5 and 150 mM NaCl by dilution and concentration on Vivaspin 500 to 2 or 4 mg mL^{-1}. Crystallization experiments for Lb-Tec2 were performed by the hanging drop vapor diffusion method using 1 + 1 µL drop at 19 ºC. A big cluster of thin plates was obtained after several months in solutions optimized from condition 20 of the Structure Screen II (Molecular Dimensions Ltd) containing 1.3 or 1.6 M $MgSO_4$ and 100 mM sodium 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5. The crystals were soaked with 10 mM 2MeSeFuc (Sommer et al., 2016a) for 2 min and then placed in 80% saturated Li_2SO_4 (~2.5 M) for cryoprotection prior mounting in a cryoloop and flash freezing in liquid nitrogen. For the All4MeMan (Beshr et al., 2016) complex, the crystals were directly transferred for 20 seconds in a solution containing 2.25 M Li_2SO_4 and 10 mM ligand prior mounting and freezing. Data were collected at 100 K at ESRF, Grenoble, France on beamline BM30A-FIP for 2MeSeFuc/Lb-Tec2 complex and BM14 for All4MeMan/Lb-Tec2 complex using an ADSC Q315 CCD and a MARCCD detector respectively. Native crystals were tested on ID30-A using a Pilatus detector but the data were not of sufficient quality for further use. The data were processed using the XDS package (Kabsch, 2010) and the scaling was performed with AIMLESS (Evans, 2011) in the CCP4 Suite (Winn et al., 2011).

Structure solution and refinement.

Using ShelXC/D/E (Schneider and Sheldrick, 2002; Sheldrick, 2010) in R32:h (also named H32) space group, initial phases for Lb-Tec2 were obtained to 1.65 Å from the anomalous

scattering of selenium present in 2MeSeFuc in a three-wavelength MAD diffraction data set comprising peak, inflection and high remote data over 180 degrees. A search for 6 selenium atoms led to an estimated mean FOM of 0.7 and a Pseudo-Free CC of 72.9%. 93% of the expected number of amino acids were observed in the polyalanine model constructed by ShelXE but they corresponded to 1.33 protein chains indicating a possible incorrect spacegroup assignment. Thus, the data were reduced to C2 space group which enabled the modelling of four protein chains. The electron density map was of excellent quality to assign the protein sequence with ArpWarp (Langer et al., 2008). During refinement in C2, it was revealed, that there was statistical static disorder in the crystal resulting from three different orientations of the protein chains related by a threefold axis. One chain was sitting on this axis and presented three superposed orientations obtained by the interchange of three stretches of about 70 amino acids (2-72, 76-147, 155-227). Those stretches have 59% sequence identity leading to a pseudo internal threefold symmetry of the Lb-Tec2 monomer. At non-conserved positions, all different amino acids could be modelled. Since it was very difficult to model the three superposed orientations of that protein chain in C2 and because the threefold axis corresponded to the crystallographic one in H32, final building and refinement was conducted in H32. Only amino acids 74-76 could not be modelled in protein chain B and water or ligand molecules were assigned to the different protein chains with the appropriate occupancy. The All4MeMan complex structure was solved by molecular replacement using PHASER (McCoy, 2007) and the 2MeSeFuc coordinates as search model. Both structures were refined using maximum likelihood restrained refinement with REFMAC 5.8 (Murshudov et al., 2011) iterated with manual model adjustment and electron density inspection in COOT (Emsley et al., 2010). The incorporation of the ligand was performed after inspection of the mFo−DFc weighted maps. Water molecules were introduced automatically using Coot and inspected manually. Ligand libraries were created using Sketcher in CCP4. The stereochemical quality

of the refined models was validated with MolProbity (Chen et al., 2010) and the wwPDB Validation server: http://wwpdb-validation.wwpdb.org prior deposition in the Protein Data Bank.

SAXS measurements.

Small-angle X-ray scattering (SAXS) data were collected at the BioSAXS beamline (BM29) of the ESRF using the HPLC mode (Pernot et al., 2013; Round et al., 2013). 50 µL of Lb-Tec2 at 8.5 mg mL^{-1} was loaded on a size-exclusion column Enrich^{TM} SEC 70 (Bio-Rad Ltd) equilibrated in 20 mM Tris pH 7.5. The data collected from the SEC run were combined into an idealized scattering curve using Primus from the ATSAS package (Petoukhov et al., 2012). Rg was estimated at low *q* values by the Guinier approximation. 40 *ab-initio* models were calculated using DAMMIF which were then averaged, aligned and compared using DAMAVER (Franke and Svergun, 2009; Volkov and Svergun, 2003). The most representative *ab-initio* model selected by DAMAVER was aligned to the crystal structure of Lb-Tec2 using SUPCOMB and the PyMOL Molecular Graphics System, Schrödinger, LLC. The experimental curve was fitted to the theoretical scattering curve using WAXSIS (Knight and Hub, 2015).

Chemical synthesis.

Thin layer chromatography (TLC) was performed using silica gel 60 coated aluminum sheets containing fluorescence indicator (Merck KGaA, Darmstadt, Germany) by charring either in aqueous $KMnO_4$ solution or in a molybdate solution (a 0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in aqueous 10% H_2SO_4) with heating. Medium pressure liquid chromatography (MPLC) was performed on a Teledyne Isco Combiflash Rf200 system using pre-packed silica gel 60 columns from Macherey-Nagel (Düren, Germany). Preparative HPLC was performed on a Ultimate3000 (Dionex, Thermo Scientific) HPLC system using a VP250/10 Nucleodur C-18 Gravity SB 5 µm column from

Macherey-Nagel. Commercial chemicals and solvents were used without further purification. EZ-Link NHS-LC-LC-Biotin was purchased from Thermo Scientific. Deuterated solvents were purchased from Eurisotop (Saarbrücken, Germany). Nuclear magnetic resonance (NMR) spectroscopy was performed on a Bruker Avance III 500 UltraShield spectrometer or on a on a Bruker 700 MHz Avance III (Ascend) spectrometer, equipped with a 5 mm TXI cryoprobe, at 298 K. Low resolution mass spectra were obtained on an amaZon SL spectrometer and high resolution mass spectra were obtained on a maxis 4G hr-qToF mass spectrometer (Bruker Daltonics, Bremen, Germany).

Allyl 4-*O*-methyl-α-D-mannopyranoside (**1**) was synthesized following the protocol of Beshr et al. (Beshr et al., 2016) and methyl 2-O-methyl-seleno-L-fucopyranoside was synthetized following the protocol of Sommer et al. (Sommer et al., 2016a).

3-cysteamine-1-propyl 4-*O*-methyl-α-D-mannopyranoside (**2**). Allyl mannoside **1** (12 mg, 0.051 mmol) and cysteamine hydrochloride (57 mg, 0.51 mmol) were stirred at 50 °C in H_2O (1 mL) for 16 h. Further cysteamine hydrochloride (28.5 mg) was added and the reaction mixture was stirred for additional 2 h prior to lyophilisation. The purification of the residue by MPLC (CH_2Cl_2 to CH_2Cl_2/MeOH = 8:1) yielded 3-cysteamine-1-propyl 4-O-methyl-α-D-mannopyranoside (**2**) as a colorless oil (12 mg, 0.038 mmol, 76%). ESI-MS calcd. $C_{12}H_{26}NO_6S^+$: 312.1; found: 312.1 (Figure S3). ^{13}C NMR (126 MHz, MeOD) δ 101.5, 78.4, 73.8, 72.8, 72.5, 66.8, 62.5, 61.1, 39.9, 30.3, 29.8, 29.2.

6-[biotinamido]-6-hexanamidohexanamido-(3-cysteamine-1-propyl) 4-*O*-methyl-α-D-mannopyranoside (**3**). EZ-Link NHS-LC-LC-Biotin (10.5 mg, 0.018 mmol) in DMF (500 µL) was added dropwise to amino mannoside **2** (5.8 mg, 0.018 mmol) in H_2O (1 mL). The reaction mixture was stirred in presence of Et_3N (1 µL) at r.t. and conversion was monitored by LCMS analysis. The solvent was removed after 20 h by lyophilisation and the residue was purified by semipreparative HPLC (RP18, 10 x 250 mm, 9 mL/min, Eluent A: H_2O, Eluent B:

MeCN, 0-5 min 10% B, 5-40 min 10%-30% B). Biotinylated mannoside **3** was obtained as a white solid (8.7 mg, 0.012 mmol, 63%). ESI-MS calcd. $C_{34}H_{62}N_5O_{10}S_2^+$: 764.4; found: 764.4. HR-MS calcd. $C_{34}H_{62}N_5O_{10}S_2^+$:764.3933; found: 764.3943 (Figure S4**)**. ^{1}H NMR (700 MHz, D_2O) δ 4.86 (d, J = 1.7 Hz, 1H, H-1), 4.63 (dd, J = 7.9, 4.9 Hz, 1H), 4.45 (dd, J = 7.9, 4.5 Hz, 1H), 3.94 (dd, J = 3.4, 1.8 Hz, 1H, H-2), 3.91 – 3.86 (m, 2H), 3.85 – 3.81 (m, 1H), 3.79 (dd, J = 12.2, 5.3 Hz, 1H), 3.64 (ddd, J = 10.0, 5.4, 2.2 Hz, 1H), 3.61 (dt, J = 10.0, 5.8 Hz, 1H), 3.56 (s, 3H, CH_3), 3.45 (t, J = 9.8 Hz, 1H), 3.42 (t, J = 6.6 Hz, 2H), 3.36 (dt, J = 9.8, 5.1 Hz, 1H), 3.25 – 3.16 (m, 4H), 3.02 (dd, J = 13.0, 5.0 Hz, 1H), 2.81 (d, J = 13.1 Hz, 1H), 2.72 (t, J = 6.6 Hz, 2H), 2.68 (td, J = 7.1, 4.9 Hz, 2H), 2.31 – 2.23 (m, 7H), 2.71 – 2.62 (m, 2H), 1.79 – 1.57 (m, 7H), 1.57 – 1.50 (m, 5H), 1.48 – 1.40 (m, 2H), 1.38 – 1.30 (m, 4H).

ITC measurements.

Isothermal titration calorimetry (ITC) experiments were performed using an ITC200 calorimeter (Microcal-Malvern Instruments). The experiments were carried out at 25 °C with the protein and ligands dissolved in 20 mM HEPES pH 7.5 and degassed. The protein concentration was checked by measuring A280 using a theoretical molar extinction coefficient of 91900 M^{-1} cm^{-1} and a molecular weight of 23795 daltons. The protein solution was placed in a 200 µL sample cell and the titration was performed with 20 injections of 2 µL carbohydrate ligands whilst stirring at 1000 rev.min^{-1}. Due to the low affinity, experiments were done in excess of ligand using 50 mM of methylated D-mannoside or L-fucoside and 100 mM of GlcNAc with protein at 150 µM.

Surface plasmon resonance.

All SPR experiments were performed on a Biacore X100 instrument (GE Healthcare) at 25 °C in HBS-ET buffer (10 mM HEPES /NaOH, pH 7.5, 150 mM NaCl, 3 mM EDTA and 0.05% Tween 20) at a flow rate of 30 µL min^{-1}. A research grade CM5 chip (GE Healthcare) was coated with streptavidin (Sigma Aldrich) using standard amine coupling procedures prior

immobilization of 128 resonance units of biotinylated 4MeMan **3** (2 mg mL^{-1} in HBS-ET buffer) on flow channel 2 and activation and deactivation on flow channel 1 as control. Experiments consisted of injection (association 180 s, dissociation 180 s) of various concentrations of protein (2-fold cascade dilutions, from 7000 to 0 μM) on both channels. Regeneration was performed by applying 1 M NaOH on both channels for 30 s at 10 $\mu L min^{-1}$.

Molecular Docking.

Coordinates of the crystal structure (PDB ID: 5FSB) were taken for calculations after deletion of the water molecules. Glide (Schrödinger program package) was previously shown to provide accurate saccharide molecule poses and thus chosen for calculations (Friesner et al., 2004; Nurisso et al., 2008; Repasky et al., 2002). Preparation of protein was fulfilled using program modules implemented in the Schrödinger program suite. Hydrogen atoms were added assessing neutral pH. Protein structure containing ligand molecules was minimized to provide an optimal hydrogen bond pattern. All calculations were performed in OPLS-2005 force field with corresponding partial atomic charges for both protein and sugar molecules. The docking procedure was performed on a rigid receptor while the ligand molecule was allowed to be flexible. Standard precision mode was used for docking. A grid of 21×21×21 Å centered on the ligand was generated without any constraints. Scaling of van der Waals radius to 1 was applied for atoms possessing a partial charge less than 0.25 e. Ligand poses were clustered with rmsd tolerance of 0.5 Å and maximum atomic displacement less than 1.3 Å. A maximum of 10 000 poses could be retained for docking procedure and were sorted according to GlideScore (GScore) function. Emodel being a combination of GScore, molecular mechanics interaction energy and the ligand strain energy was also taken into account (Friesner et al., 2004).

Several best poses were recalculated with Molecular Mechanics (MM) Generalized Born Solvent Accessible Surface Area (GBSA) procedure realized in Prime (Schrodinger) using

implicit waters. The binding free energy was calculated as a difference between the energy of the complex and of free protein and of free ligand. The ligand and residues within 4 Å of the ligand molecule were allowed to move during minimization to improve hydrogen bonds between the ligand and the receptor on one hand and to avoid large structural distortions on the another hand. One pose for each monosaccharide possessing minimal energy after the MMGBSA calculation was taken for further analysis as the best pose. Overall four monosaccharides were taken into account: allyl α-glycosides of fucose, 2-O-methyl-fucose, 3-O-methyl-mannose and 4-O-methyl-mannose. Docking of monosaccharide molecules was performed to all of six binding sites in monomer to prove their equivalence.

All figures were drawn with PyMOL Molecular Graphic System program (Version 2.0, Schrödinger, LLC) unless otherwise stated.

QUANTIFICATION AND STATISTICAL ANALYSIS

ITC measurements.

The experimental data were fitted with a stoichiometry (n) fixed to 6 to a theoretical titration curve using a single site-binding model supplied by Origin 7 software. ΔH (enthalpy change) and K_a (association constant) were obtained from the fitted data. Free energy change (ΔG) and entropy contributions (TΔS) were derived from the equation $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$ (with T the absolute temperature and R = 8.314 $Jmol^{-1}K^{-1}$). Two independent titrations were performed for each sugar apart from GlcNAc where only one was done.

NMR and mass spectrometry

Chemical shifts are given in ppm and were calibrated on residual solvent peaks as internal standard (Gottlieb et al., 1997). Multiplicities were specified as s (singlet), d (doublet), t (triplet) or m (multiplet). The signals were assigned with the help of ^{1}H, ^{1}H - COSY, DEPT-135-edited ^{1}H, ^{13}C -HSQC and ^{1}H, ^{13}C–HMBC experiments. Mass spectrometry data were analyzed using DataAnalysis from Bruker.

Surface plasmon resonance.

Binding was measured as resonance units over time after blank subtraction and the binding sensorgrams were then evaluated using the Biacore X100 evaluation software, version 2.0. Dissociation constants were determined by plotting response at equilibrium (Req, 10 s before the end of injection) against analyte concentration.

DATA AND AVAILABILITY

Accession Numbers

Mass spectrometry and NMR data for the synthetized compounds are described in Figures S3, S4. Coordinates and structure factors for the structures of Lb-Tec2 in complex with 2MeSeFuc and 4MeMan are available at the Protein Data Bank under accession codes 5FSB and 5FSC, respectively.