Structural insights into Aspergillus fumigatus lectin specificity: AFL binding sites are functionally non-equivalent

Josef Houser, Jan Komarek, Gianluca Cioci, Annabelle Varrot, Anne Imberty, Michaela Wimmerova

Structural insights into *Aspergillus fumigatus* lectin specificity - AFL binding sites are functionally non-equivalent.

Authors

Josef Houser[ab], Jan Komarek[ab], Gianluca Cioci[c], Annabelle Varrot[d], Anne Imberty[d] and Michaela Wimmerova[abe]*

[a]Central European Institute of Technology, Masaryk University, Kamenice 5, Brno, 62500, Czech Republic
[b]National Centre for Biomolecular Research, Masaryk University, Kamenice 5, Brno, 62500, Czech Republic
[c]Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, Institut National des Sciences Appliquées, Toulouse Cedex, 31077, France
[d]CERMAV-CNRS, UPR5301, affiliated with Université de Grenoble and Belonging to ICMG, BP53, Grenoble Cedex 9, 38041, France
[e]Department of Biochemistry, Faculty of Science, Masaryk University, Kotlarska 2, Brno, 61137, Czech Republic

Correspondence email: michaw@chemi.muni.cz

Synopsis The interaction between *Aspergillus fumigatus* lectin AFL and biologically relevant oligosaccharides was examined. The functional data were combined with detailed structures of the protein-saccharide complexes to understand the ligand binding at the atomic level.

Abstract *Aspergillus fumigatus* lectin AFL was recently described as a new member of the AAL lectin family. As a lectin from an opportunistic pathogen, it might play an important role in the pathogen's interaction with a human host. A detailed study of the structure of AFL complexed to several mono- and oligosaccharides, including blood group epitopes, was combined with the affinity data from SPR and discussed with previous findings. Its six binding sites are non-equivalent and due to minor differences in amino acid composition they exhibit a marked difference in specific ligand recognition. AFL displays high affinity in micro molar range towards oligosaccharides which were detected in plants but also to those bound on the human epithelia. All these results indicate AFL to be a complex member of the lectin family and a challenging target for future medical research, and due to its binding properties a potentially useful tool in specific biotechnological applications.

Keywords: *Aspergillus fumigatus*, lectin, SPR, protein-saccharide complex, pathogen

Introduction

Saccharides play crucial roles not only in energy metabolism and cell structure, but also in signalling, protein modification or host-pathogen interactions. Various techniques for research in glycosciences have been developed so far (Mechref and Novotny, 2002; Meisen et al., 2011; Tissot et al., 2009), including mass spectrometry, immunological methods and the application of carbohydrate-specific hydrolases, among others. Lectins, ubiquitous carbohydrate-binding proteins, are also widely used for detecting and identifying glycans and glycosylation modifications. They are essential for enzyme-linked lectin assay (ELLA), lectin blotting or sugar-specific labelling (Tissot et al., 2009). Several structural lectin families have been described (http://www.glyco3d.cermav.cnrs.fr). Most of them share the phenomenon of multivalency, achieved by oligomerization and/or the presence of tandem repeats. One example that combines both approaches is the well known lectin from *Aleuria aurantia* (AAL). The protein was described in 1980 (Kochibe and Furukawa, 1980) and its structure was solved twenty years later (Fujihashi et al., 2003; Wimmerova et al., 2003). This lectin forms homodimers, each monomer consisting of six tandem repeats folded in a six-bladed β-propeller. Each monomer bears five L-fucose specific binding sites, sharing the overall binding pattern with the occurrence of hydrogen bonds to Arg and Glu/Gln residues, whereas a neighboring Trp/Tyr residue mediates a CH-π interaction with the apolar surface of fucose. The five sites are similar but non-identical. The average affinity for fucose is in the micromolar range (Wimmerova et al., 2003) but it has been proposed that some of the individual sites display much higher affinities (Olausson et al., 2008). Several efforts have been made to understand the structure/function relationship at the binding site level, in order to correlate each binding site composition with its specificity and affinity (Amano et al., 2004; Olausson et al., 2011; Romano et al., 2011). The studies of bacterial structural homologues (Audfray et al., 2011; Kostlanova et al., 2005) are beneficial but, as these lectins are much smaller and bear only two types of binding sites, they cannot completely explain the observed variability in AAL.

Due to a rapid increase in the number of gene sequences available over the last decade, several homologues of the AAL protein have been identified, including lectins from *Aspergillus fumigatus* and *A. oryzae* (Ishida et al., 2002) . The crystal structure of AFL from *A. fumigatus* was recently solved, as the first from a pathogenic mold (Houser et al., 2013). AFL has also a dimeric arrangement of six-bladed β-propellers as AAL, but it differs in the number of active binding sites, with six of them per monomer (Fig. S1). The differences in the amino acid composition of individual binding sites make this protein an ideal target for detailed studies focused on the relationship between the structure and function of the lectin binding site. In addition, while *A. aurantia* is a harmless fungus, *A. fumigatus* acts as a worldwide allergen and is also a huge threat to immunocompromised patients, with a high mortality rate (Galimberti et al., 2012). Recently, the presence of AFL in *Aspergillus* conidia was demonstrated (Houser et al., 2013; Kuboi et al., 2013), with a high probability to be localized at their surface. AFL recognizes fucose, which is widely present on oligosaccharide epitopes on human tissues (Becker & Lowe, 2003) and was proofed to be used by many pathogens for attachment (Imberty & Varrot, 2008). Furthermore, an immunostimulating effect of AFL on human lung epithelial cells was observed. These findings make AFL a promising target for aspergillosis diagnosis and/or treatment. Better understanding of AFL binding properties can thus be useful in both biotechnology and medicine.

In this paper we describe the crystal structures of several complexes of AFL with various human oligosaccharide receptors. Together with surface plasmon resonance measurements, the link between structure and binding properties among AAL family lectins is described.

Material and Methods

Materials

Methyl-α,L-fucopyranoside was purchased from Interchim, Montluçon, France, L-Galactose and blood group H 2 type trisaccharide were purchased from Dextra Laboratories Ltd., Reading, UK. Other oligosaccharides were purchased from Carbohydrate Synthesis Ltd., Oxford, UK. Basic chemicals were purchased from Sigma-Aldrich, St.Louis, USA, Duchefa, Haarlem, Netherlands and Applichem, Darmstadt, Germany.

Protein expression and purification

The protein AFL was produced as described previously (Houser et al., 2013). Briefly, *Escherichia coli* BL21(DE3)Gold (Stratagene) cells bearing the pET29-*afl* vector were cultivated in standard low-salt LB medium with 50 μg ml^{-1} kanamycin. The induction by 1 mM isopropyl β-D-1-thiogalactopyranoside at 30°C for 3 hrs led to overproduction of AFL. Harvested cells were disintegrated in 20 mM Tris/HCl, pH 7.3 by sonication. AFL was isolated from the protein extract by affinity chromatography on a mannose-agarose resin (Sigma-Aldrich) using isocratic elution. Fractions containing pure AFL protein were pooled, desalted by dialysis against ultrapure water and used for further studies or lyophilized.

Surface plasmon resonance

The SPR measurements were performed using a Biacore 3000 (GE Healthcare) instrument at 25°C with running buffer (10 mM HEPES, 150 mM NaCl, 0.005% (v/v) Tween 20, pH 7.4) at a flow rate of 20 μl min^{-1}. To determine the binding preferences, the direct binding set-up was chosen. The carboxymethyldextran surface of the CM5 chip was activated with EDC (N-ethyl-N-[3-dimethylaminopropyl]carbodiimide)/NHS (N-hydroxysuccinimide) solution using the standard manufacturer's protocol. Three different concentrations of AFL (1, 5 and 50 μg ml^{-1}) in 5 mM MES/NaOH, pH 6.0 buffer were injected at a flow rate of 5 μl min^{-1} for 10 minutes into three flow channels. The relative responses of immobilized AFL were 800, 4500 and 11900 RU, respectively. Finally, the sensor surface was blocked with 1 M ethanolamine. The blank channel was treated similarly, except for the lectin injection. 80 μl of carbohydrate solutions with increasing concentrations in the running buffer were injected into the flow cells using Inject mode. The equilibrium response (after subtracting the response of the reference surface) of each experiment was used to create analyte binding curves. Using the software Origin 7.0 (OriginLab Corp.), the data obtained for a particular ligand (all applicable runs on all channels) were simultaneously fitted applying common K_d for all datasets. The one-site and two-site steady-state affinity models were used, based on equations: $R = R_{max}\, c(ligand) / [K_d + c(ligand)]$ and $R = R_{max1}\, c(ligand) / [K_{d1} + c(ligand)] + R_{max2}\, c(ligand) / [K_{d2} + c(ligand)]$, respectively, where R is the detector response, R_{max} is the maximal response, K_d is the dissociation constant and c(ligand) is the actual concentration of injected ligand. The R_{max1} to R_{max2} ratio in two-site model was not restricted in order to not influence the values of K_{d1} and K_{d2}. Depending on the channel, the experimental R_{max} values typically reached 10 - 40 % of the theoretical R_{max}.

Crystallization and data collection

Lyophilized protein was dissolved in ultrapure water and subsequently used for crystallization experiments using the hanging drop method. The initial crystallization conditions were the same as described previously (Houser et al., 2013), the conditions were optimized for each ligand used. The final crystals were obtained from a set of drops under the following conditions: 2 μl of precipitant (200 mM $CaCl_2$, 25% PEG 4K and 100 mM Tris, pH 8.5) was mixed with 2 μl of protein solution containing 7-15 mg ml^{-1} AFL and a dissolved ligand (Fig. S2). The ligand concentrations used were as follows: blood group A trisaccharide (BGA, 2 mM), Lewis Y tetrasaccharide (Le^Y, 1 mM), α,L-fucosyl(1-6)-N-acetyl-D-glucosamine (1 mM), L-galactose (L-Gal, 1 mM), methyl-β,L-fucoside (1 mM) or an equimolar mixture (1 mM each) of αFuc(1-2)Gal, αFuc(1-3)GlcNAc, αFuc(1-4)GlcNAc and αFuc(1-6)GlcNAc, respectively. Ligand-free AFL was treated the same way omitting the addition of

the ligand. Plates were incubated at 17°C until crystals were formed.

Crystals were cryo-cooled at 100K after soaking for the shortest possible time in reservoir solution supplemented with 15% (*v/v*) glycerol (AFL/BGA), 20% (*v/v*) glycerol (AFL/Le^Y, AFL/Fucα1-6GlcNAc, AFL/βMeFuc) or 40% (*v/v*) PEG400 (ligand-free AFL, AFL/L-Gal, AFL/disaccharide mixture), respectively. The X-ray diffraction experiments were performed at BESSY II in Berlin, Germany on the 14.1 beamline (ligand-free AFL) and 14.2 beamline (AFL/βMeFuc) (Mueller et al., 2012) or at ESRF (European Synchrotron Radiation Facility) in Grenoble, France on the ID23-1 or ID23-2 beamlines (all other complexes) (Nurizzo et al., 2006).

Structure determination

Collected diffraction images were processed using iMosflm 7 (Leslie & Powell, 2007, Battye et al., 2011) or XDS (Kabsch, 2010), respectively, and converted to structure factors using the program package CCP4 version 6.1 (Winn et al., 2011) with 5% of data reserved for R_{free} calculation. The structures of AFL complexes were determined using the molecular replacement method with Molrep 10.2 (Vagin and Teplyakov, 2010) or Phaser 2.5 (McCoy et al., 2007), respectively, with the structure of AFL/MeSeFuc (4agi) (Houser et al., 2013) without the ligands as the starting model. Refinement of the molecule was performed using Refmac5 (Murshudov et al., 1997) alternated with manual model building in Coot 0.7 (Emsley et al., 2010). Sugar residues and other compounds present were placed manually using Coot. Water molecules were added by Coot and checked manually. The addition of alternative conformations where necessary resulted in final structures that were validated by the ADIT validation server (http://rcsb.pdb.org) and deposited in the PDB Database with accession numbers 4ah5 (AFL/Le^Y), 4ah4 (AFL/BGA), 4agt (AFL/αFuc(1-6)GlcNAc), 4uq7 (AFL/L-Gal), 4c1y (AFL/βMeFuc) and 4aha (AFL with equimolar mixture of αFuc(1-2)Gal, αFuc(1-3)GlcNAc, αFuc(1-4)GlcNAc and αFuc(1-6)GlcNAc).

Analytical ultracentrifugation

Sedimentation analysis was performed using a ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter) equipped with an An-60 Ti rotor. Before analysis, purified AFL was brought into the experimental buffer by dialysis and the dialysate was used as an optical reference. Sedimentation velocity experiments were performed with 0.2 mg ml^{-1} AFL. Three different buffers were tested for pH dependence: 20 mM sodium citrate buffer, 150 mM NaCl, pH 4.0; 20 mM Tris/HCl, 150 mM NaCl, pH 7.3 and 20 mM sodium carbonate buffer, 150 mM NaCl, pH 10.0. The influence of ionic strength was analyzed by varying the concentration of NaCl (0 – 1000 mM) in the 20mM Tris/HCl buffer, pH 7.3. Sedimentation velocity experiments were conducted in a standard double-sector centerpiece cell loaded with 420 µl of protein sample and 430 µl of reference solution. Data were collected using absorbance optics at 25° C and a rotor speed of 40,000 rpm. Scans were performed at 280 nm at 8 min intervals and 0.003 cm spatial resolution in continuous scan mode. The partial specific volumes of the protein together with solvent density and viscosity were calculated from the amino acid sequence and buffer composition, respectively, using the software Sednterp (http://bitcwiki.sr.unh.edu). The sedimentation profiles were analyzed with the program Sedfit 13.0 (Schuck, 2000). A continuous size-distribution model for non-interacting discrete species was used to provide a distribution of apparent sedimentation coefficients.

Results

AFL-saccharide interaction

Surface plasmon resonance

The SPR technique was employed in order to determine the binding affinities of AFL towards various biologically relevant saccharides. Immobilization of the protein to the chip surface was chosen to avoid complex interpretation of interactions with a multivalent saccharide-modified surface. The data were processed using

the software Origin 7.0. Using the one-site model, dissociation constants towards fucose and all fucosylated compounds (Table 1) were rather similar and close to 0.1 mM. Affinity for L-galactose was lower (K_d=0.78 mM), indicating the importance of the C6 methyl group of L-fucose. However, as displayed in Fig. S3, the fit between the experimental data and one-site model of binding is not always good, and a more complex model should been envisaged. In theory, a six-binding sites model could be used, but this would result in a very large number of variables that could lead to over-fitting problems. Therefore, the use of a two-sites model is rationalized by the occurrence of two sets of binding sites, depending on the presence of a Trp or a Tyr residue (Houser et al., 2013) as already described for the *Aleuria aurantia* lectin (Amano et al., 2004). This approach usually results in a better fit between the experimental data and model as displayed in Fig. S3, therefore validating the proposition of a second set of binding sites with higher affinities (K_d of few μM) for fucosylated compounds. The ratio between higher and lower affinity sites was usually close to 1:5 or 2:4, what suggests even more complex behaviour of the system. It was only with L-Gal and βMeFuc that the one-site model fitted the data better to the two-sites model.

AFL displays a slight preference towards the α anomeric fucoside, but the β fucoside is also recognized. L-Gal does not bind so efficiently, confirming the hydrophobic interaction due to the fucose methyl group close to the semi-conserved Leu/Ile/Met residues. The highest affinity is reached for the Lewis X and Lewis Y epitopes, but it could be concluded that AFL has a general fucose specificity, since all tested human fucosylated oligosaccharides are efficiently bound by the lectin.

Structural studies

Structure determination

The structures of sugar-free AFL and AFL co-crystallized with various ligands were solved by molecular replacement using the protein coordinate of chain A of the AFL/selenofucoside structure (4agi) as the search model. Ligand-free AFL crystallized in $P2_1$ space group, sugar complexes crystallized in the P1, $P2_1$ or $P2_12_12_1$ space groups (see Table 2 for details).

In all structures, the AFL protein adopts the 6-bladed β-propeller fold previously determined for the MeSeFuc complex. No significant variations of the backbone conformation were observed between the chains within a particular structure or between the different structures, with rmsd varying from 0.10 to 0.33 Å when compared to the previously published structure (Houser et al., 2013). Two single point mutations L20S and R111C reported previously to be natural variation were clearly detected in all of the present complexes. Some cysteine residues in BGA, Le^Y and L-Gal complexes were found to be oxidized, since no DTT or other reducing agents were used during crystallization. Neither cysteines nor their oxidized forms participate in ligand binding. In some complexes additional ions (sodium, chloride, zinc) and organic molecules (PEG) originating from crystallization conditions were detected. They also do not affect the ligand binding or oligomeric state of the protein.

Oligomeric state

AFL is a dimeric protein in solution and crystallizes as a dimer in all complexes, similarly to previously analyzed complex with methyl seleno-α,L-fucopyranoside (Houser et al., 2013). In all structures, the protein associates with the same dimeric interface that involves mainly side chains of Gln7, Tyr109, Asn134, Asn238 and Gln262 residues and a backbone of Asn235, Ser236 and Gly263 residues. One or two dimers are observed in the asymmetric unit depending on the space group (Table 2). Comparison of the AB dimers in the different crystal structures results in rmsd values of 0.20 to 0.24 Å. Comparison with the dimer formed by its structural homologue AAL, the AFL lectin exhibits significant differences. The amino acids involved in the dimeric interface either differ in their sequence or in their position, and that leads to changes in the respective orientation of the two monomers in both dimers (Fig. S4, S5).

The existence of dimeric AFL in solution at neutral pH was originally determined by analytical

ultracentrifugation (Houser et al., 2013). Additional AUC experiments at various pH values were performed to determine the stability of the AFL dimer in solution (Fig. S6). Acidic pH 4, which is reported to affect *A. fumigatus*' conidia binding (Tiralongo et al., 2009), does not affect the oligomeric form at all, hence there is probably no link between AFL oligomeric state and the inhibition of conidia binding. However, a shift to pH 10, which is beyond the typical pH for *A. fumigatus* growth (pH 2.1 – 8.8) (Kwon-Chung & Sugui, 2013), causes a dissociation of 7% of AFL into monomers. The protein dimer was stable at neutral pH with various concentrations of NaCl (150 – 1000 mM; data not shown), which corresponds to the results published previously for lower ionic strength bellow 300 mM NaCl (Houser et al., 2013).

AFL binding sites

AFL contains six binding sites per monomer, each located between two adjacent blades (Houser et al., 2013). According to the nomenclature used previously, site 1 is located between blades I and II and so on, finishing with site 6 between blade VI and blade I. Clear electron density corresponding to AFL ligands was observed in all structures, sometimes in all binding sites, and sometimes in only some of them (Fig. S7). The binding sites in all protein subunits of particular complex display similar behaviour unless stated otherwise.

The molecular structure of AFL co-crystallized with different oligosaccharides revealed only minor changes in side chain positions. The binding mode for the fucosyl part of the oligosaccharides is the same as that observed with MeSeFuc (Fig. S1). Only a minor shift of the fucose position is observed in some complexes with larger oligosaccharides. A summary of the ligands detected in the binding sites of each complex is given in Table 3. In several cases, only the fucose moiety could be identified in the electron density, which indicates a probable flexibility of the remaining part of the oligosaccharide. The AFL structure without any ligand reveals higher flexibility of particular amino acid side chains (especially Arg25 in site 1, Trp141 in site 2 and Trp299 in site5), while the overall shape of binding site is retained (Fig. S8).

All six binding sites share a typical binding motif. Conserved Arg and Glu/Gln residues establish polar interactions with the hydroxyl O-3, O-4 and O-5 of the bound fucose, a Trp/Tyr residue mediates a CH-π interaction with the apolar surface of fucose (C-3, C-4 and C-5). Additional hydrogen bonds may be formed between the saccharide and the other residues. These interactions are different for each binding site and for each ligand. Additional hydrophobic interactions are mainly present between C-6 of the fucose and non-polar amino acids in the binding site, including semi-conserved Leu/Ile/Met in the pocket at the bottom of the site.

AFL/L-Gal complex

AFL was co-crystallized with L-Gal in order to determine the role of the L-Fuc methyl group in sugar binding. The structure revealed electron density corresponding to L-Gal in four out of six binding sites for each monomer (Table 3). In sites 3, 6 and for three out of four chains in site 1, both α and β anomers were detected. Site 2 seems to be more specific for α, L-Gal, since the β anomer could only be seen in chain C.

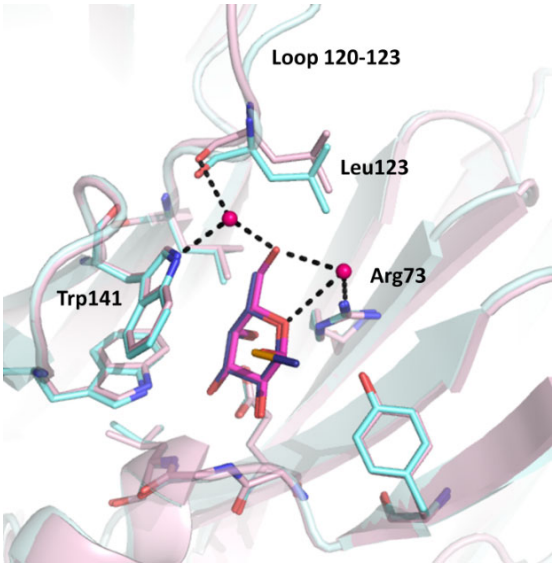


Figure 1 *Comparison of AFL site 2 in chain A of L-Gal complex (pink) with MeSeFuc complex (cyan). L-Gal (carbons in magenta) adopts the same orientation as MeSeFuc (carbons in blue) and is stabilized by two conserved water molecules (shown as spheres). The minor shift in the position of the Leu123 side chain and different orientation of loop 120-123 can be seen at the top of the figure.*

The orientation of L-Gal is identical to that observed for L-Fuc (Fig. 1). The additional OH group points outside the hydrophobic pocket and is stabilized by two bridging water molecules, mainly towards the side chains of conserved amino acids Arg and Tyr/Trp. In site 1, it is further stabilized by a hydrogen bond with a Leu69 main-chain O atom. The position of the amino acid side chains in the vicinity of the ligand are the same as in the complex with MeSeFuc. Only a Leu123 side chain changes its conformation slightly, so that it does not clash with the hydroxyl O6 of L-Gal. However, the displacement of the 120-123 loop observed in chains A and D is probably due to intermonomeric interactions in the crystal. Since no other structural changes were observed, the unfavorable accommodation of O6-L-Gal into the hydrophobic pocket in the binding site together with the restriction of O6 rotational movement (entropic penalty) are the probable causes of its lower affinity towards L-Gal, as determined by SPR.

AFL/βMeFuc complex

AFL is able to recognize βMeFuc, as was proved by glycan array (Houser et al., 2013) and SPR. To more deeply understand the observed difference between the binding of α- and β- linked fucoside, the structure of the AFL/βMeFuc complex was solved. Electron density corresponding to the ligand was detected in all four monomers in the asymmetric unit for sites 3 and 6 and in three out of four monomers in site 1 (Table 3). βMeFuc was also localized in site 5 of chain C, where it is stabilized by intermonomeric interactions enabled by crystal packing. All other sites remain unoccupied or the density for glycerol (cryo-protectant) was detected.

The sugar ligand is stabilized by a Tyr-mediated stacking interaction in sites 1, 3 and 6. This corresponds to a lower affinity as determined by SPR, where the one-site model also fit better than the two-site. The position of βMeFuc in the binding sites is identical to the αMeSeFuc complex, with no additional interaction between the methoxy-group and the protein. Since there is no sterical conflict between βMeFuc and the stacking Trp residue, it is possible that the side chain of Tyr95 and Tyr199 on the opposite side of the binding pocket causes the absence of βMeFuc in site 2 and 4, respectively (Fig. 2). Residues in the equivalent positions of occupied binding sites - Leu39 (site 1), Met140 (site 3) and Val298 (site 6) - do not interfere with βMeFuc, thus enabling its binding.

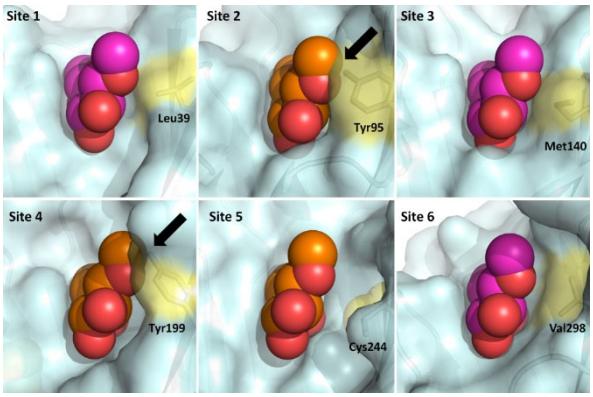


Figure 2 *AFL binding sites with βMeFuc (magenta). In sites 1, 3 and 6, βMeFuc molecule crystallographically assigned in the chain A of the AFL/βMeFuc complex structure is shown (magenta). In sites 2, 4 and 5, where no βMeFuc was detected, shown βMeFuc (orange) was superposed according to the MeSeFuc molecule from the AFL/MeSeFuc complex (4agi). Possible sterical conflicts indicated by black arrow. Tyr95 and Tyr199 that would probably collide with βMeFuc in sites 2 and 4, respectively, and corresponding amino acids in other sites are highlighted in yellow.*

AFL/disaccharide complexes

The disaccharide αFuc(1-6)GlcNAc, which corresponds to core fucosylation on N-glycans, was co-crystallized with the lectin and the electron density of the fucosyl part was clearly seen in all six binding sites. Due to the inherent flexibility of the 1-6 linkage, the GlcNAc moiety was only visible in site 2 of chain A and site 5 of chain B, when stabilized by interactions with a neighboring molecule due to crystal packing. Only one direct interaction between GlcNAc and protein was observed, linking the O5 atom of GlcNAc and the Tyr95-OH in site 2 of chain A (Fig. 3A). Additional interactions are mediated by water bridges connecting O5 to the side chains of Arg73, Asp89 and Tyr95 and O7 to Asp89 and Thr92. In site 5 of chain B, the only interaction with the protein chain is mediated by a water molecule connecting O1 of αGlcNAc and Trp299-NE1. Some other interactions, either direct or mediated by the solvent, are observed with symmetric

molecules, especially at the level of the N-acetyl group. Hence each AFL binding site is able to recognize fucose in the 1-6 linkage, but probably due to the flexibility of this linkage, the GlcNAc residue is mostly moving freely in the solvent and does not bring additional stability to the interaction.

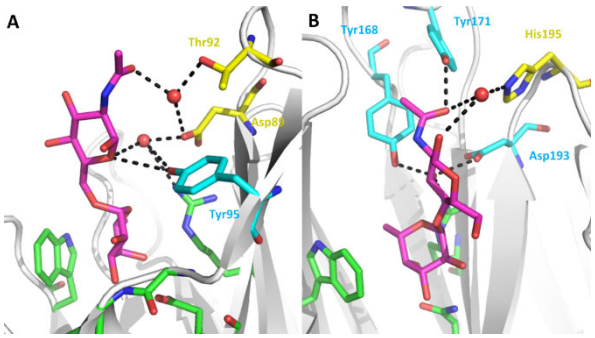


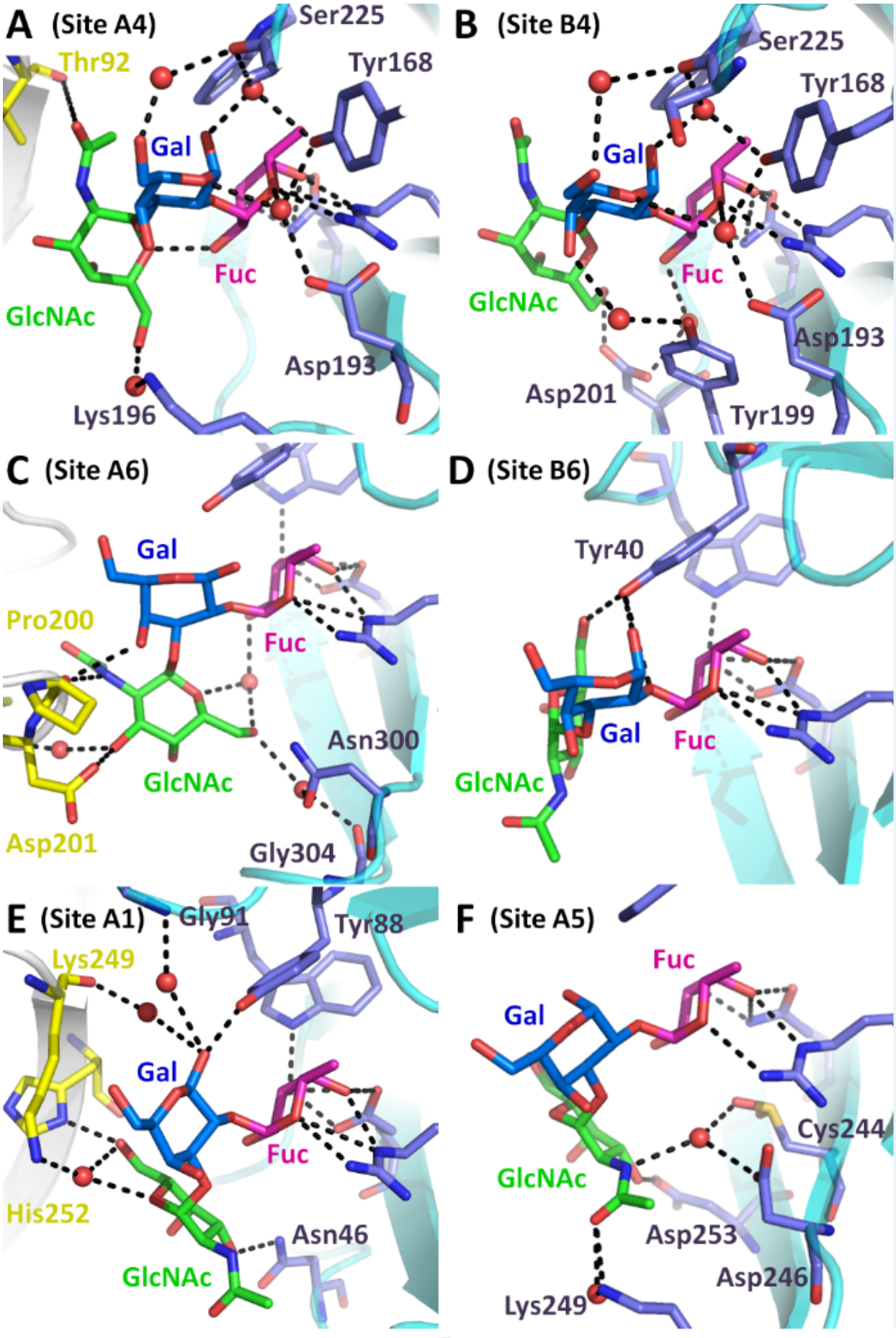
Figure 3 *(A) Binding site 2 of chain A of AFL/Fucα1-6GlcNAc complex. (B) Fucα1-4GlcNAc in binding site 4, chain B of AFL complex crystallized with mixture of fucosylated disaccharides. Hydrogen bonds linking a GlcNAc moiety to the protein side chains are highlighted in both panels. Color code for sugar-binding amino acids: green - fucose binding residues, cyan - direct interaction with GlcNAc, yellow - water mediated interaction with GlcNAc.*

In order to determine the preferences of particular binding sites, AFL was also co-crystallized with an equimolar mixture of four fucosylated disaccharides: αFuc(1-2)Gal, αFuc(1-3)GlcNAc, αFuc(1-4)GlcNAc and αFuc(1-6)GlcNAc. In all six binding sites of both A and B chains, the electron density for the fucose moiety was clearly detected. In site 2 of chain A and sites 1, 2 and 4 of chain B, the density was clear enough to assign the second saccharide unit (Table 3). In all these cases, only αFuc(1-4)GlcNAc was observed. In addition to the conserved interactions in the fucose binding site, GlcNAc is stabilized by direct hydrogen bonds to Tyr88-OH (site 1) or Tyr168-OH, Tyr171-OH and Asp193-OD2 (site 4) and by water-mediated bridges (Fig. 3B). In the other sites the electron density for the second saccharide is not clear, indicating either a large flexibility of the disaccharide linkage (probable for αFuc(1-6)GlcNAc) or a statistical disorder (different oligosaccharides present in one binding site).

AFL/BGA complex

Clear electron density for the whole blood group A trisaccharide was observed in all sites except for site 3 of chain A and sites 2 and 3 of chain B. In addition to the conserved interaction with fucose, Tyr40 (chain B, Fig. 4D) and Tyr88 (both chains A and B, Fig. 4E) mediate hydrogen bonds to the Gal moiety in sites 6 and 1, respectively. The missing interaction of Gal moiety with Tyr40 in site 6 of chain A (Fig. 4C) is affected by intermolecular interaction as described below. Additional interactions occurred between GalNAc and Asp201 side chain (site 4) in chain B (Fig. 4B) or Asp253 side chain (site 5) in chain A (Fig. 4F). The ligands in sites 1, 2, 4 and 6 of chain A are further stabilized by intermolecular interactions in the crystal, however in sites 1 and 2 this does not seem to markedly influence the oligosaccharide conformation comparing it with the computed theoretical energy minimum for the free ligand (Imberty et al., 1995) (Fig. 5). The greatest difference is exhibited by the GalNAcα1-3Gal linkage for the ligand in site 4 of both chains (Fig. 5B), which could be due to a stabilizing interaction between GalNAc-OH6 and Asp201-OD2 (chain B, Fig. 4B), a water bridge to Lys196-NZ or crystal interactions (chain A, Fig. 4A). For site 6, the interactions of the ligand with the neighboring protein monomer in the crystal are of importance. In chain B, the Gal and GlcNAc moiety is linked to the Tyr40-OH. In chain A, the ligand prefers coordination to Pro200-O and Asp201-OD2 of the neighboring protein molecule in the crystal. This leads to a conformational change of the trisaccharide (Fig. 4C, 4D and 5B). Additional stabilizing interactions result in small differences between the observed conformations of the ligand in each particular binding site of the protein, however they generally correspond to the conformations of the particular sugar linkages in PDB accessible structures.

Figure 4 *BGA trisaccharide complex with AFL. Sites are labeled according to chain and site number in each panel. Interacting residues are labeled when stabilizing a Gal or GlcNAc moiety directly or via water bridge. Interacting residues from a neighboring monomer in the crystal (shown in yellow) cause different oligosaccharide conformations in site 6 of chain A and chain B, respectively. Color code: fucose - magenta, galactose - blue, GlcNAc - green. Bridging water molecules are represented as red spheres, hydrogen bonds between saccharide and protein or water molecules, respectively, as black dashed lines*


A (Site A4)
Ser225
Thr92
Tyr168
Gal
Fuc
GlcNAc
Asp193
Lys196
B (Site B4)
Ser225
Tyr168
Gal
Fuc
GlcNAc
Asp193
Asp201
Tyr199
C (Site A6)
Gal
Pro200
Fuc
Asn300
GlcNAc
Asp201
Gly304
D (Site B6)
Tyr40
Gal
Fuc
GlcNAc
E (Site A1)
Gly91
Tyr88
Lys249
Gal
Fuc
His252
Asn46
GlcNAc
F (Site A5)
Fuc
Gal
Cys244
GlcNAc
Asp253
Asp246
Lys249

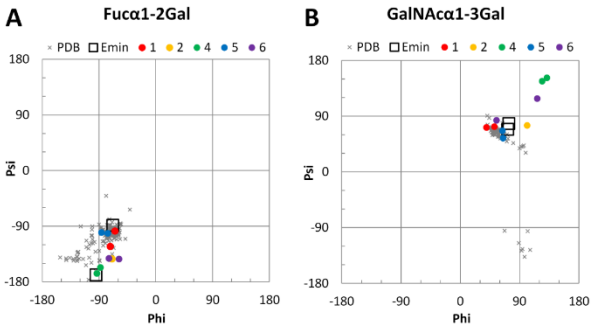


Figure 5 *Sugar Ramachandran plots for BGA epitope in AFL complex. Ligands in each binding site are shown in different colors. Conformation of theoretical energetic minima (Imberty et al., 1995) shown as empty squares. Values for the particular sugar linkage derived from PDB-deposited structures using the program GlyTorsion are shown as grey crosses.*

AFL/Le^Y complex

Lewis Y tetrasaccharide (Le^Y), containing one L-fucose bound by an α-1,2 linkage and one bound by an α-1,3 linkage, was determined to be the epitope best recognized by glycan array experiment of the biologically relevant oligosaccharides (Houser et al., 2013). Clear electron density for the complete Le^Y was observed in binding site 3 of both chains and in site 6 of chain A. For sites 1 and 2, electron densities for three out of four saccharide units were seen. Site 4 did not recognize this saccharide and in site 5, only the density for L-fucose was detected. Both protein subunits show the same binding mode considering the identity of the ligand bound in particular binding site (Table 3). In chain B some saccharide units in sites 1, 2 and 6 are stabilized by crystal contacts, as well as in site 3 of chain A.

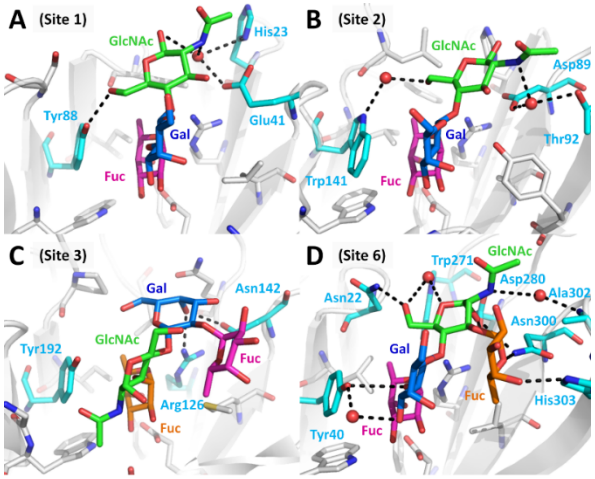


Figure 4 *LeY oligosaccharides bound in AFL chain A sites 1, 2, 3 and 6. Color code: α1-2 linked Fuc - magenta, α1-3 linked Fuc - orange, Gal - blue, GlcNAc - green, amino acids stabilizing saccharide units other than terminal fucose - cyan; bridging water shown as red spheres, hydrogen bonds shown as black dashed lines.*

For sites 1, 2 and 6 in both chains, α-1,2 linked L-fucose is preferentially recognized by the binding site (Fig. 6). The GlcNAc unit is stabilized by Tyr88 and through water bridges by His23 and Glu41 in site 1, while only water bridges are present in site 2, one linking GlcNAc to Asp89 and Thr92 and the second to Trp141. In the case of site 6 of chain A, the O6 of the GlcNAc unit interacts with Asn22-ND2, two additional water bridges link GlcNAc to Trp271 and Ala302 and one water bridge link O3 of the Gal unit to Tyr40 (Fig. 6D). The α anomeric form of the O1 of GlcNAc establishes hydrogen bonds with Asp280-OD2 and Asn300-ND2. However, the β anomer is usually present in biological oligosaccharides, so these interactions probably do not occur in nature.

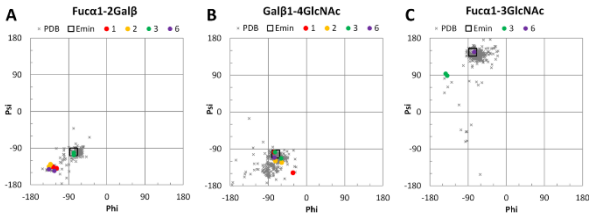


Figure 7 *Sugar Ramachandran plots for Lewis Y epitope in AFL complex. Ligands in each binding site shown in different colors. Conformation of theoretical energetic minimum (Imberty et al., 1995) shown as empty square. Values for the particular sugar linkage derived from PDB-deposited structures using the program GlyTorsion shown as grey crosses*

In site 3 of both chains, the α1-3 linked fucose is bound (Fig. 6C). The Gal unit is coordinated by Arg126 and Asn142 residues, while GlcNAc forms one weak hydrogen bond with Tyr192 in chain B. In site 5 only L-fucose could be assigned, while in site 4 no saccharide was found and the site is occupied by a glycerol molecule, which was used as a cryo-protecting agent.

Looking closer at the conformation of the Lewis Y tetrasaccharide upon binding, distortion of the angles between the bound fucose and the second saccharide unit can be observed when compared to the calculated ideal values (Imberty et al., 1995) (Fig. 7). The rest of the saccharide tends to adopt the conformation of the minimal energy state.

Discussion

The biological role of fungal lectins is the subject of debate for many years. It was proposed, that one of the key roles is the interaction with the environment such as the saccharides in organic matter. The pathogenic fungi may use them to interact with the host tissues, that is the first step of infection.

In this paper we investigated the relationship between the structure and the binding properties of the recently described fucose-specific lectin AFL from *Aspergillus fumigatus*. Several fungal or bacterial homologues of this protein have been described so far (Audfray et al., 2011; Kostlanova et al., 2005; Matsumura et al., 2007), including the closely-related AOL from *Aspergillus oryzae* (Ishida et al., 2002) and the well-known AAL from *Aleuria aurantia* (Wimmerova et al., 2003). AFL has been chosen mainly for its probable role in the pathogenicity of *A. fumigatus*, which is an important opportunistic pathogen causing allergies, bronchopulmonary aspergillosis or aspergilloma (Dagenais and Keller, 2009; Latge, 1999).

We have demonstrated that despite the similarities between AFL and AAL, there are some interesting differences between them. Although both proteins contain tandem repeats, enabling the formation of six bladed β-barrels where the binding sites are located in between adjacent blades, clear differences appear in the dimer interface. Molecules of AFL are shifted with respect to the dimer of AAL, such that the central tunnels are almost aligned in AFL. This results in a larger monomer-monomer interface area (988 $Å^2$ AAL, 1104 $Å^2$ AFL) which implies the greater dimer stability of AFL. The residues forming the intermonomer hydrogen bonds are also different, where the number of hydrogen bonds is higher in AFL dimer.

In contrast to AAL with five binding sites (Wimmerova et al., 2003), AFL possesses six binding sites, all able to bind L-fucose and related saccharides. The amino acids that play a role in ligand binding are not strictly conserved in these lectins. Considering the four amino acids directly involved in Fuc binding (through H-bonds or CH-π interaction), AAL sites vary only in the stacking amino acid, which can be either Trp or Tyr. Additional differences occur in AFL, including the replacement of Glu with Gln, involvement of the main chain in the binding and employment of water molecules to stabilize the interaction. These binding modes have not been observed in other lectins from AAL family.

The analysis of the complexes between AFL and various ligands further confirmed the non-equivalency of its binding sites. While α-methyl-selenofucoside was present in the six binding sites (Houser et al., 2013), L-Gal was only detected in four of them. The binding of L-Gal requires the presence of a glutamate residue for interacting with the O-3 and O-4 of the saccharide. Sites 4 and 5, where L-Gal could not bind, have Gln at this position, but they also present an unfavorable orientation of Tyr168 and Phe274, respectively, which would lead to sterical conflict with O6-L-Gal. This may correlate with the inability to fit SPR data with a two-site model. Therefore, the sites may be more specific for the α,L-fucosylated epitope than the others. Similarly, there are only some binding sites occupied in the complex with βMeFuc. Even though the absence of β-linked fucose can be explained by sterical hindrances, it is noteworthy that only sites with a stacking Tyr residue (sites 1, 3 and 6) are able to accommodate this ligand (Table 3). The only exception is site 5 in chain C, where the ligand is stabilized due to crystal packing. This nicely corresponds to the SPR data, where the one-binding-site model was more suitable for the AFL-βMeFuc interaction. The lower affinity of the interaction is probably due to the fact that stacking with Tyr residues is not as efficient as with Trp residues, as reported previously for homologous lectin RSL from *Ralstonia solanacearum* (Wimmerova et al., 2012).

The structure of the AFL/Le^Y complex revealed an interesting binding pattern, where sites 1, 2 and 6 predominantly bind the α1-2 linked fucose, while site 3 recognizes the α1-3 linked fucose. Site 5 probably binds both Fuc and site 4 is not able to accommodate the whole tetrasaccharide. This clearly demonstrates a high variability among AFL binding sites. To further decipher the

preference of each binding site, we crystallized AFL with a mixture of disaccharides, where Fuc is linked to the other saccharide through various linkages. All binding sites were occupied by a ligand, but only three of them preferred one particular disaccharide from the mixture, i.e. αFuc(1-4)GlcNAc, which is a natural epitope in plants (Fitchette-Laine et al., 1997). This preference could be explained by the small differences in the binding of the ligand, where the fucoside part is crucial, while the identity of the second saccharide only marginally affects the binding. Since the complexes were co-crystallized, the observed differences are probably not caused by packing but due to the amino acid composition of the binding sites. In the complex with the human blood group A trisaccharide, the ligand is recognized by each binding site, but a more detailed view reveals the flexibility of the epitope. The creation of additional hydrogen bonds between the protein and the Gal or GalNAc part of the saccharide can change the torsion angles of the oligosaccharide, as can intermonomeric interactions within the crystal. Generally, AFL-sugar binding in particular site is limited only for large epitopes or for ligands with directly modified fucose, while alpha-fucosylated di- and trisaccharides are recognized by all of the sites.

Several methods are commonly used to determine lectin binding affinities. Here we demonstrated that the surface plasmon resonance method is suitable for AFL, but still has some disadvantages. AFL remains active upon immobilization and the response to binding partners is clear and dose dependent. For low-affinity ligands the change in SPR signal is too weak, but Fuc and its derivatives were suitable analytes. The data were much more accurately fitted by the two-site binding model, showing the presence of non-equivalent binding sites. It was reported for the structural homologous protein RSL, that the stacking interaction between the saccharide unit (L-fucose) and the aromatic side chain (Trp) is crucial for high-affinity binding (Wimmerova et al., 2012). With AFL, two sets of binding sites can be distinguished, one containing Tyr and one containing Trp. This correlates well with the observed two-site binding model suggested by SPR and also with the structure of βMeFuc complex. The lower binding affinity and one-site model suitability nicely correspond to the occupancy of Tyr-containing binding sites observed in this case. The values of K_d are in the micromolar range for all ligands tested. This is also true for homologous AAL (Wimmerova et al., 2003) and AOL (Matsumura et al., 2009). However, no significance was observed for the existence of the high affinity binding site in AFL. The marginal difference in K_d's with respect to various fucosylated disaccharides may be caused by only infrequent interactions of the second saccharide and AFL side chains. Comparing the previous hemagglutination data (Houser et al., 2013) with SPR results, minimal inhibitory concentrations nicely correlates with the low affinity binding sites K_d's, as expected. Three binding sites of both types seem to be more specific towards αFuc(1-4)GlcNAc than towards the other disaccharides tested, based on the crystal structure. As SPR detected αFuc(1-3)GlcNAc being bound with similar strength, both ligands are probably bound in the sites where only fucose was assigned. Also the binding might slightly differ under SPR and crystallization conditions. Of all the oligosaccharides tested, the Lewis Y tetrasaccharide was the highest affinity ligand overall. This corresponds to our previous finding by glycan array (Houser et al., 2013).

It can be concluded, that AFL evolves into a multispecific lectin being able to recognize Fuc in all possible linkages. These could be found not only in the decomposed plant matter in soil, which is the natural environment for *A. fumigatus*, but also in various epitopes on the human tissues. Hence AFL may help *Aspergillus* to invade human hosts through various tissues, especially lungs which are the most prevalent one. Taking all of our findings together, AFL displays a high variability in binding specificity while using a common binding motif. The similarity to the AAL lectin from *Aleuria aurantia* is obvious, yet the differences in the number of binding sites, their specificity and affinity are interesting. Combining our results with previous work on AFL structural homologues (Audfray et al., 2011; Kostlanova et al., 2005; Matsumura et al., 2007; Romano et al., 2011; Wimmerova et al., 2003) can help to understand the rules used by AAL family proteins

in ligand recognition. In addition, AFL being a potential virulence factor of *A. fumigatus*, it might be an interesting drug target. Therefore the finding of binding site non-equivalency is crucial for meaningful inhibitor development and treatment strategies.

Acknowledgements We wish to acknowledge the Consortium for Functional Glycomics, Grant number GM62116, for performing the glycan array studies. We also wish to thank to the European Synchrotron Radiation Facility in Grenoble for access to beamline ID23-1 and Helmholtz-Zentrum Berlin for access to BESSY II beamline 14.2. We would like to thank to Ben Watson-Jones for the language correction. The research leading to these results received financial support from the European Union under the Seventh Framework Programme by CEITEC (CZ.1.05/1.1.00/02.0068) project from European Regional Development Fund, SYLICA (Contract No. 286154 under ‘‘Capacities’’ specific programme), Czech Ministry of Education (LH13055) and the Czech Science Foundation (GA13-25401S). This work was further supported by CNRS, France, the FINOVI foundation and the French cystic fibrosis Association Vaincre la Mucoviscidose.

References

Amano, K., Fujihashi, M., Ando, A., Miki, K. & Nagata, Y. (2004). *Biosci. Biotechnol. Biochem.* **68**, 841-847.

Audfray, A., Claudinon, J., Abounit, S., Ruvoen-Clouet, N., Larson, G., Smith, D. F., Wimmerova, M., Le Pendu, J., Romer, W., Varrot, A. & Imberty, A. (2011). *J. Biol. Chem.* **287**, 4335-4347.

Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. (2011). *Acta crystallographica. Section D, Biological crystallography* **67**, 271-281.

Becker, D. J. & Lowe, J. B. (2003). *Glycobiology* **13**, 41R-53R.

Dagenais, T. R. T. & Keller, N. P. (2009). *Clin. Microbiol. Rev.* **22**, 447-465.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486-501.

Fujihashi, M., Peapus, D. H., Kamiya, N., Nagata, Y. & Miki, K. (2003). *Biochemistry* **42**, 11093-11099.

Galimberti, R., Torre, A. C., Baztan, M. C. & Rodriguez-Chiappetta, F. (2012). *Clin. Dermatol.* **30**, 633-650.

Houser, J., Komarek, J., Kostlanova, N., Cioci, G., Varrot, A., Kerr, S. C., Lahmann, M., Balloy, V., Fahy, J. V., Chignard, M., Imberty, A. & Wimmerova, M. (2013). *Plos. One* **8**, doi: 10.1371/journal.pone.0083077.

Fitchette-Laine, A. C., Gomord, V., Cabanes, M., Michalski, J. C., Saint Macary, M., Foucher, B., Cavelier, B., Hawes, C., Lerouge, P. & Faye, L. (1997). *Plant. J.* **12**, 1411-1417.

Imberty, A., Mikros, E., Koca, J., Mollicone, R., Oriol, R. & Perez, S. (1995). *Glycoconj. J.* **12**, 331-349.

Imberty, A. & Varrot, A. (2008). *Curr. Opin. Struct. Biol.***18**, 567-576.

Ishida, H., Moritani, T., Hata, Y., Kawato, A., Suginami, K., Abe, Y. & Imayasu, S. (2002). *Biosci. Biotechnol. Biochem.* **66**, 1002-1008.

Kabsch, W. (2010). *Acta cryst.* **D66**, 125-132.

Kochibe, N. & Furukawa, K. (1980). *Biochemistry* **19**, 2841-2846.

Kostlanova, N., Mitchell, E. P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerova, M. & Imberty, A. (2005). *J. Biol. Chem.* **280**, 27839-27849.

Kuboi, S., Ishimaru, T., Tamada, S., Bernard, E. M., Perlin, D. S. & Armstrong, D. (2013). *J. Infect. Chemother.* **19**, 1021-1028.

Kwon-Chung, K. J. & Sugui, J. A. (2013). *Plos. Pathog.* **9**, e1003743.

Latge, J. P. (1999). *Clin. Microbiol. Rev.* **12**, 310-350.

Leslie, A. G. W. & Powell, H. R. (2007). *Nato. Sci. Ser. Ii. Math.* **245**, 41-51.

Matsumura, K., Higashida, K., Hata, Y., Kominami, J., Nakamura-Tsuruta, S. & Hirabayashi, J. (2009). *Anal. Biochem.* **386**, 217-221.
Matsumura, K., Higashida, K., Ishida, H., Hata, Y., Yamamoto, K., Shigeta, M., Mizuno-Horikawa, Y., Wang, X., Miyoshi, E., Gu, J. & Taniguchi, N. (2007). *J. Biol. Chem.* **282**, 15700-15708.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl.Cryst.* **40**, 658-674.
Mechref, Y. & Novotny, M. V. (2002). *Chem. Rev.* **102**, 321-369.
Mueller, U., Darowski, N., Fuchs, M. R., Forster, R., Hellmig, M., Paithankar, K. S., Puhringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012). *J. Synchrotron Radiat.* **19**, 442-449.
Meisen, I., Mormann, M. & Muthing, J. (2011). *Biochim. Biophys. Acta* **1811**, 875-896.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240-255.
Nurizzo, D., Mairs, T., Guijarro, M., Rey, V., Meyer, J., Fajardo, P., Chavanne, J., Biasci, J. C., McSweeney, S. & Mitchell, E. (2006). *J. Synchrotron Radiat.* **13**, 227-238.
Olausson, J., Astrom, E., Jonsson, B. H., Tibell, L. A. & Pahlsson, P. (2011). *Glycobiology* **21**, 34-44.
Olausson, J., Tibell, L., Jonsson, B. H. & Pahlsson, P. (2008). *Glycoconj. J.* **25**, 753-762.
Romano, P. R., Mackay, A., Vong, M., DeSa, J., Lamontagne, A., Comunale, M. A., Hafner, J., Block, T., Lec, R. & Mehta, A. (2011). *Biochem. Biophys. Res. Commun.* **414**, 84-89.
Schuck, P. (2000). *Biophys. J.* **78**, 1606-1619.
Tissot, B., North, S. J., Ceroni, A., Pang, P. C., Panico, M., Rosati, F., Capone, A., Haslam, S. M., Dell, A. & Morris, H. R. (2009). *FEBS letters* **583**, 1728-1735.
Vagin, A. & Teplyakov, A. (2010). *Acta cryst.* **D66**, 22-25.
Wimmerova, M., Kozmon, S., Necasova, I., Mishra, S. K., Komarek, J. & Koca, J. (2012). *Plos One* **7**, doi: 10.1371/journal.pone.0046032.
Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. & Imberty, A. (2003). *J. Biol. Chem.* **278**, 27059-27067.
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., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235-242.

Table 1 Apparent K_d values for interaction of biologically important saccharides with AFL using SPR measurement.

Values are given with standard deviation based on measurement on three independent channels with different amounts of AFL immobilized.

Ligand	K_d (one site) [μM]	K_{d1} (two sites) [μM]	K_{d2} (two sites) [μM]
Fuc	209 ± 15	408 ± 40	91.9 ± 32.8
αMeFuc	120 ± 8	958 ± 189	83.0 ± 4.9
βMeFuc	251 ± 29	#	#
L-Gal	775 ± 71	#	#
αFuc(1-2)Gal	82.3 ± 5.2	220 ± 39	14.2 ± 4.3
αFuc(1-3)GlcNAc	71.7 ± 5.7	278 ± 19	7.28 ± 0.91
αFuc(1-4)GlcNAc	63.3 ± 4.9	202 ± 8	4.10 ± 0.42
αFuc(1-6)GlcNAc	84.0 ± 4.7	206 ± 18	15.0 ± 2.1
αFuc(1-2)[αGalNAc(1-3)]Gal Blood group A trisaccharide	68.0 ± 5.9	183 ± 47	11.5 ± 4.5
αFuc(1-2)[αGal(1-3)]Gal Blood group B trisaccharide	61.7 ± 4.7	181 ± 28	10.3 ± 2.3
αFuc(1-2)βGal(1-4)GlcNAc Blood group H type 2 trisaccharide	68.3 ± 7.8	192 ± 34	4.38 ± 1.68
αFuc(1-4)[βGal(1-3)]GlcNAc Lewis a trisaccharide	101 ± 13	442 ± 79	8.24 ± 1.90
αFuc(1-3)[βGal(1-4)]GlcNAc Lewis X trisaccharide	172 ± 11	246 ± 9	1.80 ± 0.53
αFuc(1-2)βGal(1-3)[αFuc(1-4)]GlcNAc Lewis b tetrasaccharide	131 ± 7	249 ± 59	26.6 ± 13.3
αFuc(1-2)βGal(1-4)[αFuc(1-3)]GlcNAc Lewis Y tetrasaccharide	75.8 ± 3.1	113 ± 6	3.85 ± 1.07

Values not determined. Two-site models for βMeFuc and L-Gal were unable to reach meaningful fitting.

Table 2 Data collection and refinement statistics.

Values for the outer shell are given in parentheses.

Ligand	L-Gal	βMeFuc	Fucα1-6GlcNAc	Disaccharide mixture#	Blood group A	Le^Y	None
PDB ID	4UQ7	4C1Y	4AGT	4AHA	4AH4	4AH5	4UOU
Data collection statistics							
Diffraction source	ESRF beamline ID23–1	BESSY II beamline 14.2	ESRF beamline ID23–2	ESRF beamline ID23–2	ESRF beamline ID23–1	ESRF beamline ID23–2	BESSY II beamline 14.1
Wavelength (Å)	0.9763	0.9184	0.8726	0.8726	0.9795	0.8726	0.9184
Space group	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_1$	P1	P1	$P2_1$
a, *b*, *c* (Å)	80.03, 70.44, 117.80	71.21, 90.34, 189.17	47.41, 88.35, 79.78	45.73, 88.38, 78.58	46.42, 47.44, 80.09	47.32, 47.53, 77.36	47.63, 140.22, 78.79
α, β, γ (°)	90.00, 108.34, 90.00	90.00, 90.00, 90.00	90.00, 102.95, 90.00	90.00, 99.63, 90.00	103.61, 91.96, 113.08	96.81, 100.24, 113.61	90.00, 92.15, 90.00
No. of monomers in asymmetric unit	4	4	2	2	2	2	4
Resolution range (Å)	34.44–1.75 (1.85–1.75)	48.14–2.23 (2.35–2.23)	38.87–2.00 (2.11–2.00)	36.41–2.20 (2.32–2.20)	42.03–1.75 (1.84–1.75)	42.01–1.99 (2.10–1.99)	78.74–2.40 (2.53–2.40)
Total no. of reflections	448327 (63804)	168544 (16725)	102700 (11494)	104127 (15841)	138182 (18326)	122864 (15105)	163605 (24505)
No. of unique reflections	121067 (8051)	56964 (5173)	38241 (4925)	27431 (1286)	57394 (7620)	39131 (5255)	40256 (5854)
Completeness (%)	97.3 (95.5)	99.0 (97.8)	88.2 (80.0)	87.5 (71.8)	94.1 (85.3)	96.1 (88.9)	99.7 (99.5)
Redundancy	3.7 (3.7)	8.2 (3.2)	2.4 (2.4)	2.1 (2.1)	2.4 (2.4)	3.1 (2.9)	4.1 (4.2)
$\langle I/\sigma(I)\rangle$	12.0 (3.2)	6.1 (1.5)	7.1 (2.4)	7.1 (2.6)	14.3 (2.5)	11.6 (2.5)	12.9 (8.7)
R_{merge} (%)	0.082 (0.390)	0.130 (0.350)	0.110 (0.440)	0.120 (0.380)	0.050 (0.353)	0.077 (0.379)	0.064 (0.108)
CC(1/2)	0.996 (0.879)	0.996 (0.563)	0.985 (0.567)	0.983 (0.750)	0.999 (0.887)	0.997 (0.892)	0.996 (0.988)
Wilson *B* ($Å^2$)	13.9	30.8	15.9	16.7	15.8	20.8	8.2
Refinement statistics							

Amino acids	4x314	4x314	2x314	2x314	2x314	2x314	4x314
Protein atoms	2449 A; 2455 B; 2449 C; 2448 D	2439 A; 2443 B; 2434 C; 2438 D	2443 A; 2440 B	2444 A; 2451 B	2454 A; 2445 B	2438 A; 2455 B	2437 A; 2439 B; 2437 C; 2442 D
Solvent atoms	1070	137	423	443	556	412	526
Ligand atoms	342	201	175	203	357	344	14
Resolution limits (Å)	34.44–1.76 (1.80–1.76)	47.29–2.23 (2.29–2.23)	38.87–2.00 (2.05–2.00)	36.41–2.20 (2.26–2.20)	42.03–1.75 (1.80–1.75)	42.01–1.99 (2.04–1.99)	78.74–2.40 (2.46–2.40)
No. of reflexions in working set	114970 (8319)	56864 (4095)	36310 (2342)	26054 (3722)	54494 (3283)	37163 (2343)	38050 (2828)
No. of reflexions in test set	6078 (436)	2993 (215)	2172 (110)	2472 (198)	2897 (174)	1967 (119)	2008 (136)
Final R_{cryst} (%)	15.2 (25.4)	24.2 (35.1)	17.2 (28.8)	16.5 (36.4)	16.8 (28.1)	16.8 (24.1)	21.7 (19.7)
Final R_{free} (%)	17.7 (28.4)	31.3 (42.4)	21.2 (29.2)	21.2 (41.7)	21.1 (30.0)	22.5 (31.0)	27.9 (29.8)
Mean B value (Å^2)	17.42	26.96	16.07	15.28	18.95	24.64	11.30
R.m.s. deviations							
Bonds (Å)	0.012	0.012	0.011	0.016	0.012	0.014	0.014
Angles (°)	1.590	1.495	1.447	1.727	1.419	1.648	1.610
Planar groups (Å)	0.008	0.007	0.007	0.008	0.006	0.007	0.007
Chiral volumes (Å^3)	0.096	0.085	0.090	0.101	0.080	0.098	0.088
Ramachandran plot							
Most favoured (%)	97.1	94.1	97.5	97.0	97.9	97.1	95.3
Allowed (%)	2.5	5.4	2.2	2.9	1.8	2.6	4.3
Outliers (%)	0.4	0.5	0.3	0.1	0.3	0.3	0.4

Equimolar mixture of αFuc(1-2)Gal, αFuc(1-3)GlcNAc, αFuc(1-4)GlcNAc and αFuc(1-6)GlcNAc disaccharides used for crystallization.

Table 3 Ligands coordinated in individual binding sites of AFL crystal structures.

Only the clearly defined parts of a saccharide are listed. Saccharide units bound to protein directly are in **bold**, via water molecule underlined and ligands stabilized by crystal packing interactions are in *italics*. In all cases the interaction cutoff was set to 3.6 Å.

Chain	Ligand	L-Gal	L-Gal†	βMeFuc	βMeFuc†	Fucα1-6GlcNAc	Fuc mix#	Blood group A	Le^Y
A	Site 1	**L-Gal**	**L-Gal**	**βMeFuc**	**βMeFuc**	**Fuc**	**Fuc**	α**Fuc**1-2(α***GalNAc***1-3)**Gal**	α**Fuc**1-2βGal1-4**GlcNAc**
	Site 2	**L-Gal**	**L-Gal**	**Glycerol**	-	α**Fuc**1-6***GlcNAc***	α**Fuc**1-4GlcNAc	α**Fuc**1-2(α*GalNAc*1-3)Gal	α**Fuc**1-2Galβ1-4GlcNAc
	Site 3	***L-Gal***	***L-Gal***	**βMeFuc**	**βMeFuc**	**Fuc**	**Fuc**	**Fuc**	αFuc1-2β**Gal**1-4(α**Fuc**1-3)*GlcNAc*
	Site 4	**PEG**	**PEG**	**Glycerol**	-	**Fuc**	**Fuc**	α**Fuc**1-2(α*GalNAc*1-3)Gal	**Glycerol**
	Site 5	**PEG**	**PEG**	**Glycerol**	***βMeFuc***	**Fuc**	**Fuc**	α**Fuc**1-2(α**GalNAc**1-3)Gal	**Fuc**
	Site 6	**L-Gal**	**L-Gal**	**βMeFuc**	**βMeFuc**	**Fuc**	**Fuc**	α**Fuc**1-2(α*GalNAc*1-3)Gal	α**Fuc**1-2βGal1-4(α**Fuc**1-3)**GlcNAc**
B	Site 1	**L-Gal**	**L-Gal**	**βMeFuc**	-	**Fuc**	α**Fuc**1-4***GlcNAc***	α**Fuc**1-2(α**GalNAc**1-3)**Gal**	α**Fuc**1-2β*Gal*1-4**GlcNAc**
	Site 2	**L-Gal**	**L-Gal**	***Glycerol***	**Glycerol**	**Fuc**	α**Fuc**1-4*GlcNAc*	**Fuc**	α**Fuc**1-2β*Gal*1-4*GlcNAc*
	Site 3	**L-Gal**	**L-Gal**	**βMeFuc**	***βMeFuc***	**Fuc**	**Fuc**	**Fuc**	αFuc1-2β**Gal**1-4(α**Fuc**1-3)**GlcNAc**
	Site 4	**PEG**	**PEG**	-	-	**Fuc**	α**Fuc**1-4**GlcNAc**	α**Fuc**1-2(α**GalNAc**1-3)Gal	**Glycerol**
	Site 5	**PEG**	**PEG**	**Glycerol**	-	α**Fuc**1-6*GlcNAc*	**Fuc**	α**Fuc**1-2(α**GalNAc**1-3)Gal	**Fuc**
	Site 6	**L-Gal**	**L-Gal**	***βMeFuc***	**βMeFuc**	**Fuc**	**Fuc**	α**Fuc**1-2(α**GalNAc**1-3)**Gal**	α**Fuc**1-2*Gal*

† Ligands from chains C and D shown instead of A and B, respectively.

Equimolar mixture of αFuc(1-2)Gal, αFuc(1-3)GlcNAc, αFuc(1-4)GlcNAc and αFuc(1-6)GlcNAc disaccharides used for crystallization.