SapL1: A New Target Lectin for the Development of Antiadhesive Therapy Against *Scedosporium apiospermum*

Dania Martínez-Alarcón[1, 2], Jean-Philippe Bouchara[3], Roland J. Pieters[2], Annabelle Varrot[1*].

1 Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France. dania.martinez.alarcon@gmail.com, annabelle.varrot@cermav.cnrs.fr

2 Utrecht University, 3584 CG Utrecht, The Netherlands. R.J.Pieters@uu.nl

3 Host-Pathogen Interaction Study Group, GEIHP, EA 3142, SFR ICAT 4208, UNIV Angers, UNIV Brest, Institut de Biologie en Santé, IRIS, CHU d'Angers, Angers, France. jean-philippe.bouchara@univ-angers.fr

* Corresponding author

Email: annabelle.varrot@cermav.cnrs.fr

Abstract

Scedosporium apiospermum is an emerging opportunistic fungal pathogen responsible for life-threatening infections in immunocompromised patients. This fungus exhibits limited susceptibility to all current antifungals and, due its emerging character, its pathogeny and virulence factors remain largely unknown. Carbohydrate binding proteins such as lectins are involved in host-pathogen interactions and may constitute valuable therapeutic targets to inhibit microbial adhesion to the host cells by using carbohydrate mimics. However, such lectins are still unidentified in *S. apiospermum*. Here, we present the first report of the identification and characterization of a lectin from *S. apiospermum* named SapL1. SapL1 is homologous to the conidial surface lectin FleA from *Aspergillus fumigatus* known to be involved in the adhesion to host glycoconjugates present in human lung epithelium. The present report includes a detailed strategy to achieve SapL1 soluble expression in bacteria, its biochemical characterization, an analysis of its specificity and

affinity by glycan arrays and isothermal titration calorimetry (ITC), as well as the structural characterization of its binding mode by X–ray crystallography. The information gathered here contribute to the understanding of glycosylated surface recognition by *Scedosporium* species and is essential for the design and development of antiadhesive glycodrugs targeting SapL1.

Introduction

During the last decades, an increased incidence of invasive infections, especially in immunosuppressed patients, has been caused by previously rare fungal pathogens[1]. Among those pathogenic fungi, species from the *Scedosporium* genus are particularly dangerous since their mortality rate may be over 80%[2]. *Scedosporium* infections mainly affect immunocompromised patients such as solid organ transplant (SOT) or hematopoietic stem cell transplant (HSCT) recipients, or patients with chronic granulomatous disease. Together with *Fusarium* species and some dematiaceous fungi, *Scedosporium* species are responsible for approximately 10% of the mycosis caused by filamentous fungi in hematopoietic cell transplant recipients and up to 19% in solid organ transplant patients.[3] In addition, *Scedosporium* species rank the second among the filamentous fungi colonizing the respiratory tract of people with cystic fibrosis (CF), after *Aspergillus fumigatus*, and because of the propensity of these fungi to disseminate in case of immune deficiency, this fungal colonization of the airways is considered in some centers as a contraindication to lung transplantation which remains the ultimate treatment in CF[4]. Apart from lung transplantation, this fungal colonization of the airways which usually starts in the adolescence and becomes chronic in up to 19% of the patients, may also lead to a chronic inflammation and sometimes to an allergic bronchopulmonary mycosis[2]. Furthermore, it has been reported that some species of this genus can also affect immunocompetent patients[5].

The genus *Scedosporium* comprises more than ten species *(Scedosporium angustum, S. apiospermum, S. aurantiacum, S. boydii, S. cereisporum, S. dehoogii, S. desertorum, S. ellipsoideum, S. fusoideum, S. minutisporum, S. rarisporum and S. sanyaense)* whose taxonomy has been previously described based on molecular phylogenetic by using the sequences of four genetic loci[2,6]. However only the genome of *S. boydii, S. aurantiacum, S. dehoogii* and *S. apiospermum* have been completely sequenced to date[7-9]. Among *Scedosporium* species, *S. apiospermum* (formerly considered as the asexual form of *Pseudallescheria boydii*) is the most outstanding. This cosmopolitan microfungus is responsible for localized to severe or fatal disseminated infections in humans and is one of the most common *Scedosporium* species (with *S. boydii* and *S. aurantiacum*) capable to colonize chronically the lungs of CF patients[3,10].

The treatment of *Scedosporium* infections is challenging because its efficacy depends on timely diagnosis, which is based on morphological detection by microscopy, histological analysis, and culture on selective media. This process is time consuming and may lead to false negative results, especially from respiratory secretions of patients with CF because of the common co-colonization by other fungi or bacteria [2]. Furthermore, *Scedosporium* species display a primary resistance to classical antifungals such as 5-flucytosine, amphotericin B, and the first-generation triazole drugs, fluconazole or itraconazole and exhibit a limited susceptibility to the newest generations of antifungal drugs, *i.e.* echinocandins, voriconazole and isavuconazole[2,11]. Nowadays, the first-line treatment for *Scedosporium* infections involves a combination therapy that includes the use of voriconazole in conjunction with other antifungals[11,12]. However, due to common recurrences, even without interruption of treatment, the recovery rates are poor and mortality remains over 65% while it is almost 100% when dissemination occurs, reason why it has aroused special interest despite its emerging character[12,13].

Scedosporium infections begin with conidial adherence to tissues, followed by germination and hyphal elongation[14]. This adherence process allows it to avoid cleansing mechanisms

aimed to eradicate the invading pathogens and represents the initial step towards infection[14-17]. Conidial adhesion is mediated by cell surface molecules (CSM), including different types of carbohydrates such as polysaccharides and glycoconjugates. Some of the most important carbohydrates CSM described to date in *Scedosporium* species include peptidorhamnomannans (PRMs)[18-20], α-glucan[21], melanin[22], ceramide monohexosides[20,23], N-acetyl-D-glucosamine-containing molecules[24] and mannose/glucose-rich glycoconjugates[20]. The presence and/or abundance of these molecules on the cell surface varies according to the stage of development and is of great relevance to understand fungal pathobiology[20]. The carbohydrates binding proteins known as lectins also act as CSMs and were shown to have an essential role during pathogenesis. However, the knowledge about lectins in emerging microfungi is very limited to date.

Since some lectins from microorganisms have been demonstrated to be involved in the host recognition and adhesion process, they have been used to develop anti-adhesive therapy through the use of carbohydrates and their mimics[15-17,25-28]. This approach is particularly promising since it does not kill the pathogen nor arrest its cell cycle. Consequently, resistance frequencies are rare[16,25] and expected side effects in non-target tissues are lower than those caused by conventional antifungal compounds. This may allow its implementation as a prophylactic therapy for immunocompromised patients[17].

Due to the emerging character of *Scedosporium apiospermum*, there is very limited information on its mechanisms of recognition and anchoring to the host. Furthermore, lectins from this microorganism have not been characterized to date, what hinders the development of an anti-adhesive therapy. Conversely, closely related filamentous fungi were investigated leading to the identification and characterization of their host binding modes. For example, in *A. fumigatus,* which is a saprophytic mold also responsible for bronchopulmonary infections in receptive hosts, the lectin FleA (or AFL) was identified and revealed a role in host-pathogen interactions[29]. FleA is a six-bladed β-propeller homodimer located on the conidial surface that recognize human blood group antigens

and mediates *A. fumigatus* binding to airway mucins and macrophages glycoproteins in a fucose-dependent manner[30]. In healthy individuals, this anchorage is critical for the mucociliary clearance process and the macrophagy, in fact it has been described that fleA-deficient (ΔfleA) conidia are even more pathogenic than wild type (WT) conidia both, in healthy and chemically immunocompromised mice[30,31]. However, CF patients represent a very particular scenario because the mucus in their lungs is thicker, in relation with mucin overproduction and the high content of calcium ions, which modulates the supramolecular organization of MUC5B by protein cross-link[31,32]. This contributes to the suboptimal transport properties of mucus and compromises the pathogen clearance mechanisms[31]. Furthermore, the aberrant glycosylation of the mucins MUC5B and MUC5B in CF lungs causes, among other things, an increase in the abundance of sialyl-Lewis X and Lewis X determinants in the non-reducing end of mucins carbohydrate chains[33,34], which can also be translated as an increase in the fucose content. Therefore, in this context, the FleA (and homologous proteins) anchoring to the mucus layer plays an essential role in the colonization of the CF lungs by *A. fumigatus*.

Here, we have used the recently sequenced genome of *S. apiospermum*[9] to identify a putative homologue of FleA that we have called SapL1 for *Scedosporium apiospermum* Lectin 1. The present report comprises SapL1 identification, its production in bacteria, an analysis of the fine specificity and affinity of the recombinant protein, as well as its structural characterization by X–ray crystallography.

Results

Production and purification of SapL1

The hypothetical protein XP_016640003.1 (EMBL accession number), encoded by SAPIO_CDS9261 and from now on referred as SapL1, was identified through data mining using the FleA (pdb entry 4D4U[35]) sequence as bait into the genome of the reference strain

S. apiospermum IHEM 14462[9]. During the sequence analysis, we identified that the first 74 amino acids of the putative protein (Uniprot A0A084FYP2) are a very disordered region that is not present in any other protein of the same family. We suspect that this peptide could be due to an error during the genome annotation and propose that SapL1 sequence actually starts at methionine-75 (Fig. S1). The SapL1 coding sequence (from Met-75) was fused to an N-terminal 6xHis tag cleavable by the Tobacco Etch virus protease (TEV) under regulation of *trc* and *T7* promoters into pProNde and pET-TEV vectors, respectively (Fig. 1A). Expression was performed in *Escherichia coli* and purification was carried out using immobilized metal affinity chromatography (IMAC). Unfortunately, the original expression yield with both vectors (~0.35 mg·L^{-1} of culture) was too low to proceed with characterization studies. Therefore, we explored new alternatives to enhance the expression. First, the thioredoxin protein (Trx) as well as 6-His tag and TEV cleavage site were fused at the N-terminus of SapL1 by subcloning into the pET32-TEV vector (Fig. 1A). This strategy substantially increased the SapL1 production yield but most of the protein remained insoluble as part of inclusion bodies (Fig. 1B).

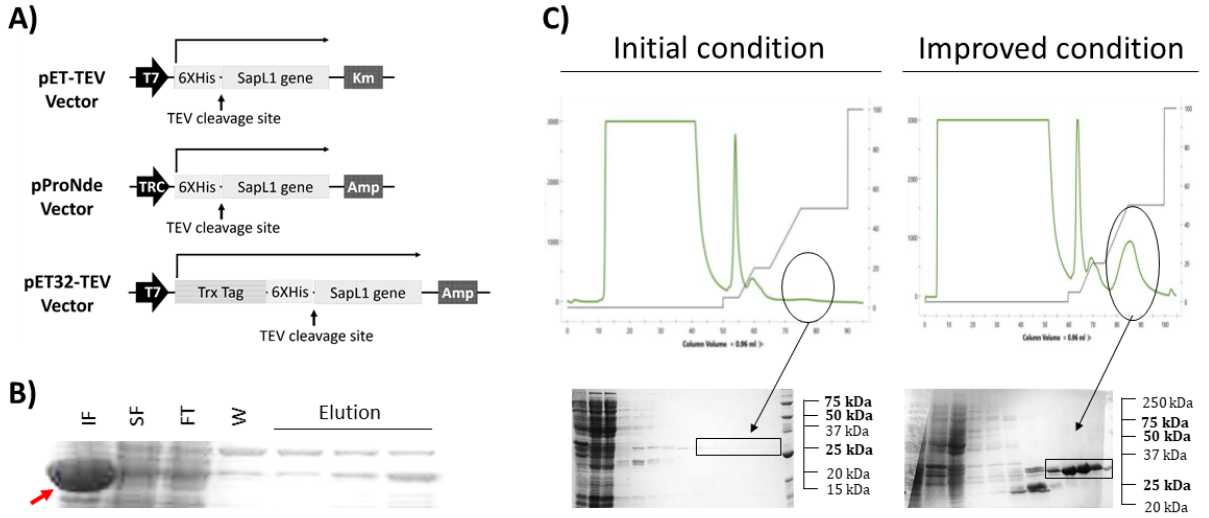

Figure 1

A wide range of expression conditions were subsequently assayed to achieve sufficient soluble expression by modification of various parameters such as growth temperature, host strain, inducer concentration, optical density of the culture at induction, culture

duration, etc. Sixty-nine different sets of parameters were assayed (Table S1) and it was possible to improve the yield up to 4 mg·L^{-1} (Fig. 1C). The best set of conditions for SapL1 expression was using *E. coli* strain TRX, pProNde vector, LB media, growth at 37°C and 160 rpm until OD_{600} = 0.4, before a switch of the temperature to 16°C and overnight induction at OD 0.8 with 0.05 mM IPTG and 1% L-rhamnose (Rh).

Rhamnose influence on SapL1 solubility

Interestingly, during the optimization of the expression conditions, we found that rhamnose (Rh) plays an essential role in the solubility and stability of SapL1. Therefore, we performed a new set of experiments to demonstrate the correlation between the amount of protein recovered after purification and the Rh concentration in the media. For this, we investigated SapL1 expression under the previously described conditions modified by supplementation of the culture medium with different concentrations of Rh (0.1%, 0.15%, 0.2%, 0.5% and 1%). Purification parameters were set to obtain high purity product and kept identical for all experiments. Figure 2 shows the chromatograms obtained for representative concentrations accompanied by their respective SDS-PAGE profile. To quantify SapL1 expression in these experiments, we integrated the area under peaks corresponding to the protein (Fig. 2, black bars). Thus we confirmed that SapL1 recovery is directly proportional to the Rh concentration in the culture medium. Then, the experiment was repeated at three different concentrations (0.1%, 0.15%, 0.2%) using 20°C as the induction temperature, and production behavior was similar to that found at 16°C, but with lower performance.

Additional experiments also showed that SapL1 expression can be induced at high concentration of rhamnose in a dose dependent manner, even without addition of IPTG when the pProNde vector is used (data not shown). For this reason, we initially consider that the higher production rates of SapL1 could be attributed to the dissociation of the *LacI*-DNA complex caused by an increase in cytosolic potassium concentration since the accumulation of this ion is one of the main adaptive responses of *E. coli* to hypertonic

shocks, and its concentration has been correlated to the loss of interactions between DNA and proteins in *in vitro* assays[36-38]. However, previous reports have demonstrated that, *in vivo*, the effects of macromolecular crowding caused by plasmolysis (the reduction of intracellular water content, as an adaptive response to osmotic shock) on the activity of cytoplasmic proteins, is sufficient to buffer the kinetics of association of DNA-protein against changes in [K^+].[38,39]

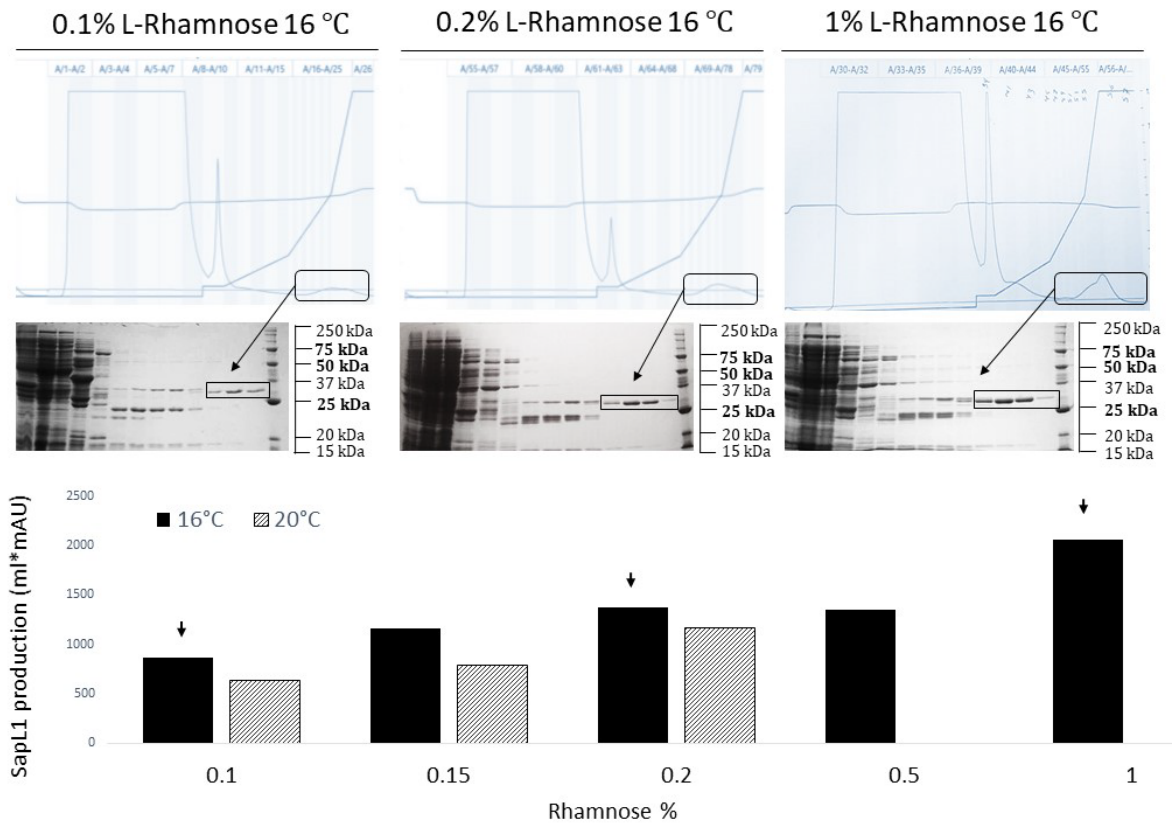

Figure 2

Thus, it seems that the higher recovery rates of SapL1 is actually caused by a combination of two main parameters: 1) the leak of genetic repression by dissociation of the *LacI-DNA* complex (characteristic of *lac* promoters) and 2) a positive effect on the solubility and folding of the protein due to the increased concentration of organic osmoregulators, whose synthesis is induced in *E. coli* as an adaptive response to the external hypertonicity caused, in this case, by the addition of rhamnose to the culture medium[36-44]. This hypothesis was later supported by the finding that the addition of glycerol, instead of

rhamnose, displays the same effect on production and solubility of SapL1 (data not shown). Plasmolysis may also play an essential role in SapL1 stabilization, since it has been shown that macromolecular crowding has positive effects on protein folding and in some cases it can drive self-association of improper folded proteins into functional oligomers[38].

Biochemical characterization

We assayed the thermal stability of SapL1 in 26 different buffers in a pH range of 5 to 10 through Thermal Shift Assay (TSA). The most suitable condition for this protein was MES buffer 100 mM pH 6.5, where a single denaturing event at T_m of 55°C was observed (Fig. S2). Then, to estimate the molecular size of the native protein, we performed size exclusion chromatography using an ENrich™ SEC 70 column (Bio-Rad) and 20 mM MES, 100 mM NaCl, pH 6.5 as mobile phase. However, the protein displayed strong non-specific interactions with the matrix of the column, and SapL1 could not be eluted even using 5 M of NaCl. Interestingly, it could be recovered when the buffer was supplemented with 20 mM α-methyl fucoside or L-rhamnose, evidencing similar effects on SapL1 elution for both sugars. Due to the impossibility to estimate the molecular weight of SapL1 by size exclusion chromatography on this resin, we performed measurements in solution using Dynamic Light Scattering (DLS). We obtained a monodisperse peak corresponding to a protein of 72 ± 29.4 kDa corroborating that SapL1 forms dimers as FleA and other proteins of this family (monomer MW: 40 kDa, data not shown)[29]. The range of the standard deviation also suggested an ellipsoidal shape, which is characteristic for the dimers in this lectin family[45].

Carbohydrate binding properties

A hemagglutination assay showed that recombinant SapL1 agglutinated rabbit red blood cells at 0.97 $\mu g \cdot ml^{-1}$ (Fig. 3A). It confirms that the recombinant lectin is active and that its heterologous production in *E. coli* did not alter its hemagglutinating properties.

To identify the potential ligands of SapL1 on epithelial cell surfaces, we submitted SapL1 to the glycan array version 5.4 of the Consortium for Functional Glycomics (USA) consisting of 585 mammalian glycans. It was labelled with Fluorescein Isothiocyanate (FITC) in a molar ratio of 0.426 and its binding properties were analysed at two different concentrations (5 and 50 $\mu g \cdot mL^{-1}$). As expected from its homology with FleA, SapL1 recognizes fucosylated oligosaccharides independently of the fucose linkage. The α1-2 and α1-3/4 linked fucosides displayed the highest affinity whilst α1-6 linked ones, the lowest. The weakest interactions with fucosylated compounds were reported for branched oligosaccharides (Fig. S3). The affinity of SapL1 for L-fucose and α-methyl-fucoside was determined by Isothermal Titration Calorimetry (ITC) and the K_d was found to be 225 μM and 188 μM, respectively with stoichiometry fixed to 1 since the measurements were done in the presence of an excess of ligand (Fig. 3B). These values are in agreement with the affinity constant (around 110 μM) reported for FleA for α-methyl-fucoside.[46] No binding interaction was observed for SapL1 with rhamnose by ITC (data not shown).

Analysis of the glycans constituting blood group determinants revealed that SapL1 binds to all epitopes with a preference for H type 2 (Fucα1-2Galβ1-4GlcNAcβ) then Lewis^a (Galβ1-3(Fucα1-4)GlcNAcβ) and Lewis^x (Galβ1-4(Fucα1-3)GlcNAcβ). However, most of the recognized branched oligosaccharides contained the core fucose Fucα1-6. Epitopes with two fucose units, such as Lewis^b and fucosylated polylactosamine, were also well recognized. Addition of a galactose or a GalNAc as in blood group B or A antigens did not impair Fucα1-2 recognition (Fig. 3C). Spacers used to join the carbohydrates to the chip also display a strong influence on binding. It is remarkable that 70% of the 90 positive binders contained either the spacer Sp0 ($CH_2CH_2NH_2$) or Sp8 ($CH_2CH_2CH_2NH_2$). Those spacers also display a strong influence on binding, especially for small glycans such as Galα1-3(Fucα1-2)Galβ1-4(Fucα1-3)GlcNAcβ which was recognized when attached to Sp0

but not to Sp8. This may be due to a steric hindrance caused by the modification of carbohydrate presentation on the surface of the chip.

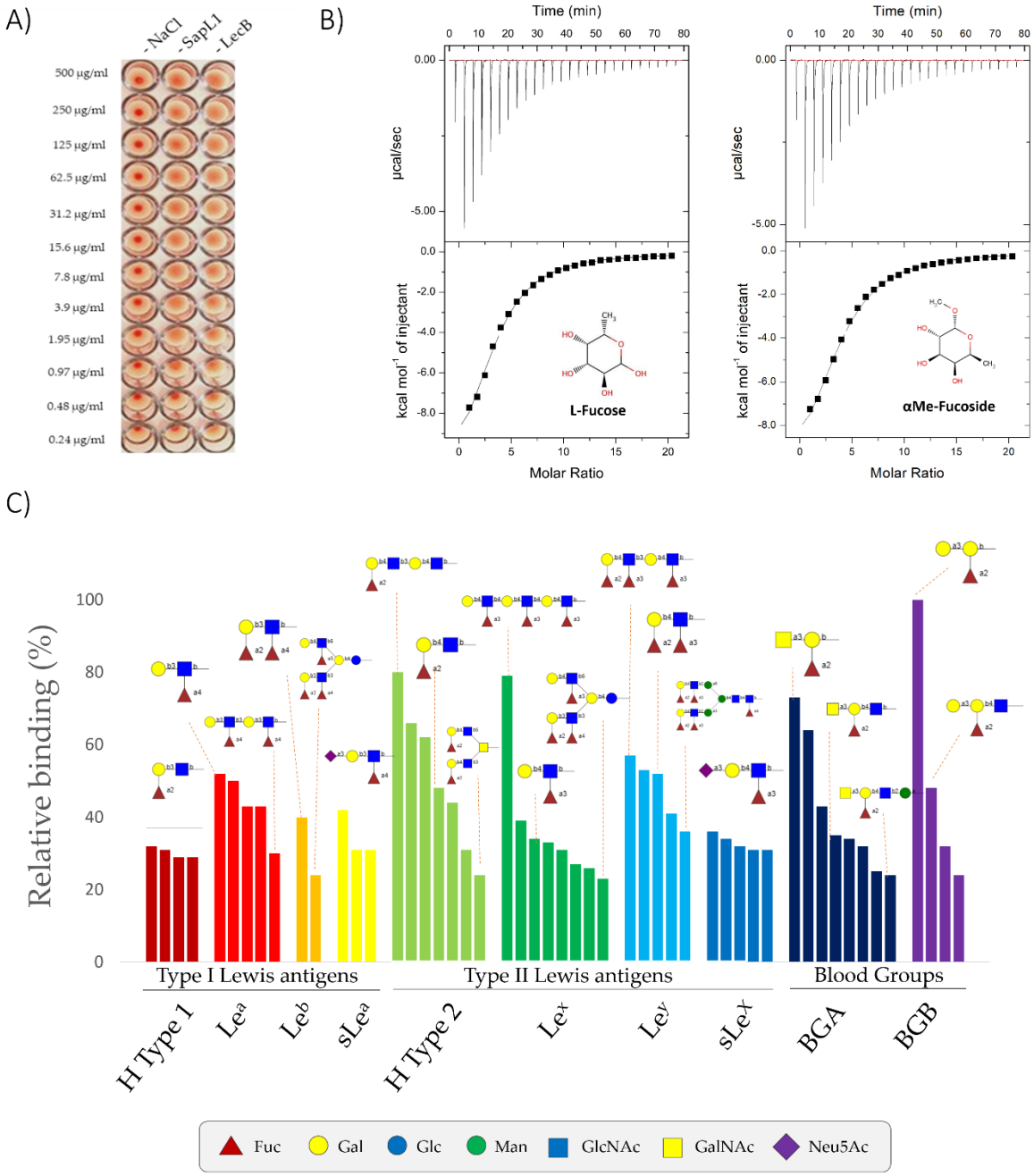

Figure 3

Overall structure of SapL1

SapL1 was co-crystallized with α-methyl-fucoside and the structure of the complex was solved by molecular replacement at 2.3 Å resolution in the $P2_1$ space group using the

coordinates of FleA (PDB code 4D4U[35]) as the search model. The asymmetric unit contained two monomers, assembled as a dimer with all 295 amino acids visible apart of the N-terminal methionine. See data collection and refinement statistics in Table 1. SapL1 folds into the canonical six-bladed β-propeller with six-binding sites at the interface between blades typical for this family of lectin (Fig. 4). A fucose moiety was found in 5 and 4 of the six binding pockets of chains A and B, respectively while glycerol originating from the crystallizing solution was found in the other binding sites.

Table 1. Data-collection and refinement statistics. *Values in parentheses are for the outer shell

Data collection		
Beamline	SOLEIL Proxima 1	
Wavelength (A°)	0.97857	
Space group	$P2_1$	
Unit cell dimensions a, b, c (Å), α, β, γ (°)	76.06, 45.66, 83.48, 90, 105.05, 90	
No. of monomers in ASU	2	
Resolution (Å)	40.0 – 2.4 (2.46-2.4)	
R*merge*	0.085 (0.428)	
R*pim*	0.07 (0.356)	
Mean *I/σ (I)*	5.3 (1.6)	
Completeness (%)	98.2 (98.4)	
Multiplicity	2.7 (2.7)	
CC1/2	0.991(0.7)	
No. reflections /No. Unique reflections	58305/ 21570	
Refinement		
Resolution (Å)	20.0-2.3	
No. of reflections in working set / Free set	21561 / 1089	
R work/ R free	17.5 / 24.0	
R.m.s Bond lengths (Å)	0.017	
Rmsd Bond angles (°)	1.94	
Rmsd Chiral (Å³)	0.087	
No. atoms / Bfac (Å²)	Chain A	Chain B
Protein	2222/32.64	2220/40.83
Ligand and heterogen	89/31.56	66/36.914
Waters	111/31.018	69/33.206
Ramachandran Allowed / Favored/ Outliers (%)	96 / 4 / 0	95 / 4 / 1
PDB Code	6TRV	

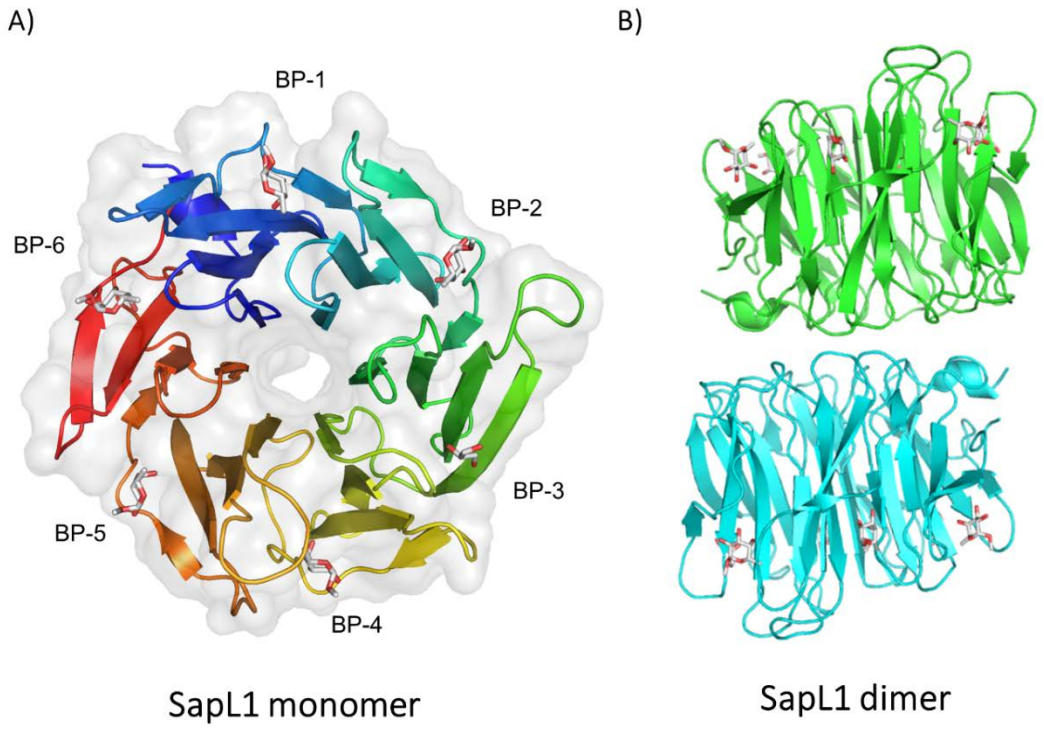

Figure 4

The overall fold of SapL1 and FleA as well as the overall dimer are very similar with a rsmd of 1.2 and 1.26 Å, respectively. They share 43% of sequence identity and both proteins present the same distribution of β-strands except for the lack of the last blade of strand 4 in SapL1, and the external face of blade 5, in which FleA displays an elongated β-strand (β22) that is split in two in SapL1 (β22/23). It is remarkable that the structure of the first 3 blades is highly conserved in both proteins, while the second half (blades 3-6) displays the largest discrepancies. Furthermore, FleA also presents two additional small α-helices (ή3 and ή4), located in the loops between sheets that serve as connectors for blades (2/3 and 5/6, respectively, Fig. 5). Similar conclusions can be drawn when comparing SapL1 to its homologue AOL in *Aspergillus oryzae*[47]. Those proteins have similar dimerization interface and we noticed differences on the other side of the propeller where sugar binding occurs. The lectins present surface loops of different size and sequence leading to change in the architecture of the related binding sites that impacts their affinity as discussed in more details below. The first member structurally characterized in this lectin family was AAL from the mushroom *Aleuria aurantia*[48]. It

presents differences on both sides of the β-propeller with the microfungal members of this family mainly at the level of the surface loops. This leads to a different dimerization interface and the dimer cannot overlay. This also affects fucose binding since AAL only has five functional binding sites. The role of these lectins is still unknown, in particular for the mushroom members.

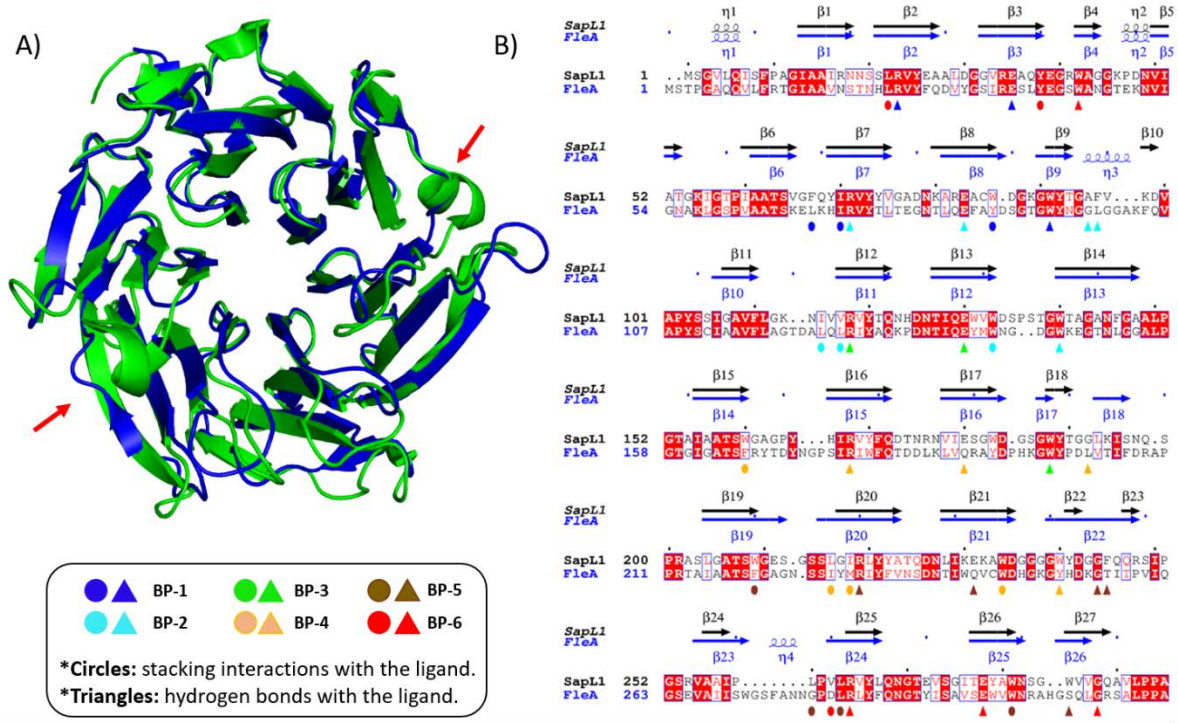

Figure 5

Protein-ligand interactions

Due to divergence in the tandem repeat sequence forming each blade of the propeller, the six binding sites of SapL1 monomer are not equivalent, but they share important conserved features. Hydrophobic interactions are observed between the C6 of the fucose and at least three residues of the protein (mainly isoleucine, tryptophan/tyrosine and leucine). The O2 and O3 hydroxyls make strong hydrogen bonds with a conserved triad of amino acid consisting of an arginine, a glutamic acid and a tryptophan. In the cases where glycerol was found in the binding pocket, its own hydroxyls mimic the interactions of fucose with these same residues (Fig. 6). It is to be noted that SapL1 binding sites are

more conserved than FleA binding site where a glutamine can replace the glutamic acid and a tyrosine the tryptophan in the triad (Fig. S4).

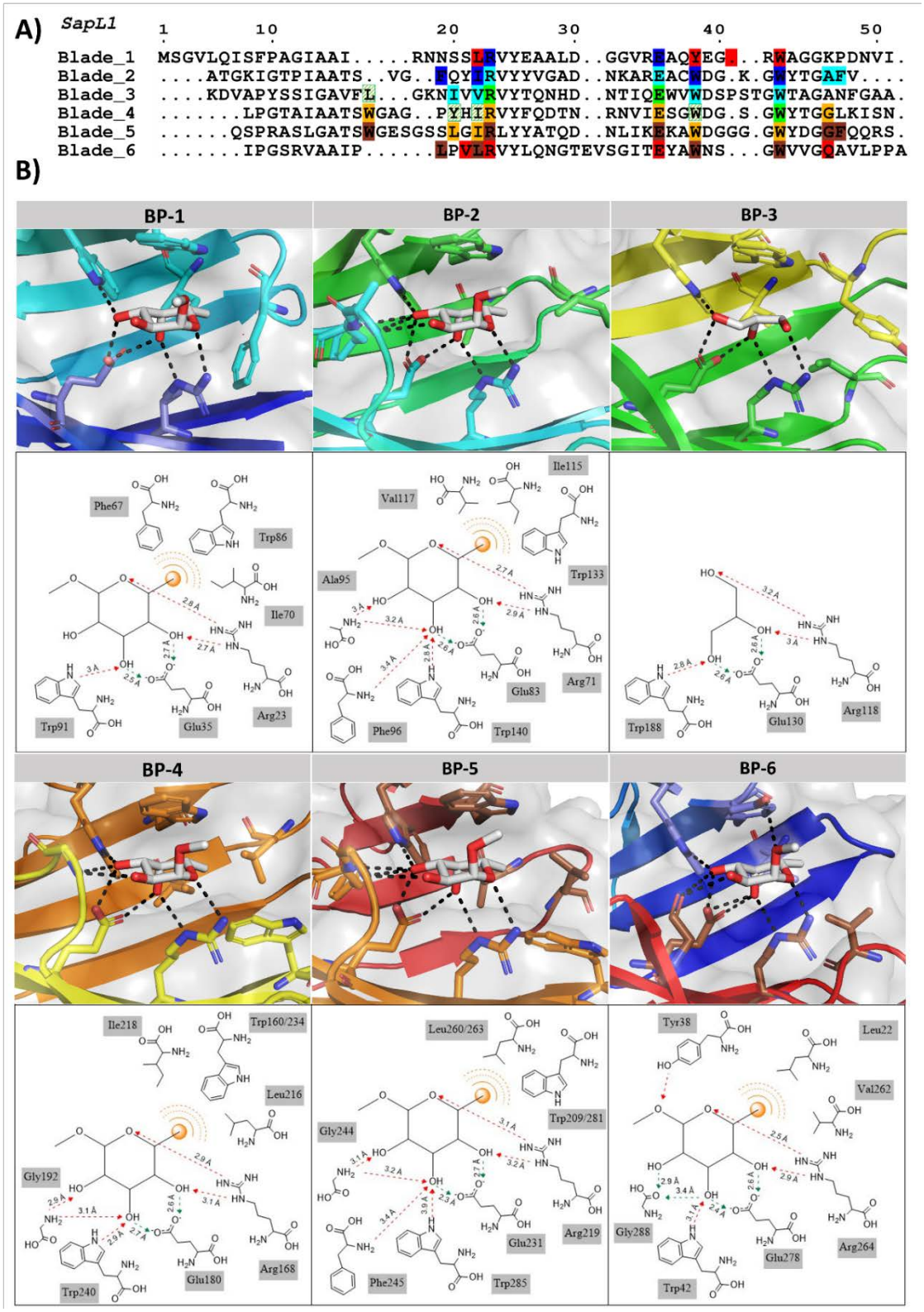

Figure 6

The O2 hydroxyl seemed to be the most versatile position, since it established interactions with a loop adjacent in four of the six binding sites (BPs 2, 4, 5 and 6). This is particularly interesting since the binding pockets that do not contain this loop (BPs 1 and 3) were mostly occupied by glycerol instead of fucose, indicating that those interactions could be responsible for enhancing the affinity and might be explored for development of inhibitors.

Discussion

The development novel techniques for diagnosis has exposed that, until now, the incidence of fungal infections has been underestimated.[1,49] In comparison with pathogenic bacteria, which have been the focus of attention for several years, there is a lack of information about the virulence factors and host recognition systems of pathogenic fungi[50]. This hinders the development of new drugs against this microorganisms, while their resistance to the current ones is progressing quickly[13,24,49]. Therefore, there is a rush in obtaining information that may contribute to the development of new antifungal agents.

In this study, we present the identification, production and characterization of a new carbohydrate binding protein from the emerging microfungus *S. apiospermum*, whose host-anchoring mechanism is completely unknown. To the best of our knowledge, this is the first studied lectin for this opportunistic pathogen. SapL1 is homologous to the conidial surface lectin FleA from *A. fumigatus* known to be involved in adhesion to the host glycoconjugates present in mucins and human lung epithelium[30]. During our analysis, we have identified a possible error on the annotation of the initial methionine of SapL1 and we have relayed this information to corroborate this finding and to update the entry if necessary.

In our initial attempts to produce recombinant SapL1, we observe that the conventional parameters of expression in *E. coli*, result in the production of insoluble protein that was

mainly found in inclusion bodies. Later, experimental data indicated that the presence of rhamnose or glycerol was required for SapL1 solubilization during bacterial expression and that the production yield was directly proportional to their concentration in the culture medium. This suggests that the stabilization of SapL1 derives from an osmomodulator effect that, either directly or indirectly, enhances its proper folding. This osmomodulator effect may be due to the leak of genetic repression by dissociation of the repressor (*LacI-DNA)* complex[36-38] in the first case, or in the second case to the increased concentration of organic osmoregulators, which are known to enhance protein folding and solubility, presumably also reinforced by the macromolecular crowding produced by plasmolysis[36-44]. These findings, and the extensive research carried out to produce SapL1 in its soluble form, provide important insights for the heterologous expression of eukaryotic lectins tending to be produced as insoluble proteins in *E. coli.* Besides, our findings highlight the useful role of organic osmoregulators during heterologous expression of proteins to favour their proper folding and to avoid the formation of inclusion bodies.

We demonstrated that SapL1 is strictly specific for fucosylated carbohydrates and recognized all blood group types present in the glycan array screening. These results are particularly interesting since it has been shown that there is six times more content of fucose α1,3/4 linked to glycoproteins in the CF airways[33]. This phenomenon is mainly due to an increased expression of the α1,3-fucosyltransferase, which is involved in the synthesis of sialyl-Lewis X and Lewis X determinants attached to mucus mucins[34]. Therefore, the fact that SapL1 recognizes equally the α1-2 and α1-/3/4-linked fucosides may explain the high incidence of scedosporiosis in CF patients. Besides, it correlates with the presence of *SapL1* gene in all the pathogenic strains of *Scedosporium* whose genome has been sequenced to date. However, there is no evidence suggesting that the phenotype of blood group could have an influence in this recognition conversely to previous reports on other pathogens like *Pseudomonas aeruginosa* and *Haemophilus influenza*[51-53].

Finally, the specificity and affinity of each one of the six binding pockets of SapL1 has been deeply analyzed to highlight the features that must be explored for the design of efficient inhibitors. Within our analysis, we have found that the binding pockets are non-equivalent but they all share the features necessary for fucose recognition. This example of divergence can also be found in the other lectins of the same family with FleA being the member with the greatest differences between pockets known to date[35]. It is remarkable that although SapL1 has a relatively low sequence identity with other members of this family, its specificity and affinity are very similar to those reported for FleA from *A. fumigatus*[46], AAL from the orange peel mushroom *Aleuria aurantia*[48] and the bacterial lectins BambL and RSL, from *Burkholderia ambifaria* and *Ralstonia solanacearum,* respectively[54,55]. Together with the crystallographic data, this confirms that the structure and function of this lectin family are highly conserved, settling the possibility for development of a broad-spectrum therapy.

Overall, our research has revealed the first insights about the recognition of human glycoconjugates by *S. apiospermum* lectins and contributes to the general understanding of the host-binding process during the early stages of infection. The detailed information exposed here, places SapL1 as a promising target for the diagnosis and/or treatment of *Scedosporium* infections and it will be of great value to guide the development of antiadhesive glycodrugs against this pathogen.

Methods

Production

The coding sequence for SapL1 was optimized for expression in *Escherichia coli* and ordered at Eurofins Genomics (Ebersberg, Germany) before cloning into the expression vectors pET-TEV[56], pET32-TEV[46] and pProNde. pProNde is a homemade vector where the NcoI restriction site of pProEx HTb (EMBL, Heidelberg) has been replaced by a NdeI restriction site by PCR. Then, plasmids were introduced into *E. coli* strains by thermal

shock at 42°C and different expression parameters were assayed (Table S1) until soluble expression was achieved. The best set of conditions for SapL1 expression were: *E. coli* TRX strain, pProNde vector, LB medium, cells grown at 37°C and 160 rpm until 0.4 DO_{600}, then cultures transferred to 16°C until 0.8 OD_{600nm} and protein induction carried out overnight by the addition of isopropyl-β-D-1-thiogalactoside (IPTG) to a final concentration of 0.05 mM and 1% L-rhamnose (Rh).

Purification

Cells were subsequently harvested by centrifugation at 5000 g for 10 min and resuspended in buffer A (50 mM Tris-HCl, 500 mM NaCl, pH 8.5). After addition of 1 µl of Denarase (C-LEcta GmbH, Leipzig, Germany) and moderate agitation during 30 minutes at room temperature, cells were disrupted at 1.9 kbar using a cell disruptor (Constant Ltd Systems). Cell debris were removed by centrifugation at 22000 g for 30 minutes. The supernatant was filtered using 0.45 micrometer membranes (PES, ClearLine) and the protein purification was carried out by IMAC using 1 mL His-Trap FF columns (GE Healthcare Life Sciences) and a NGC™ chromatography system (Bio-Rad). After loading the supernatant, the column was rinsed thoroughly with buffer A until stabilization of the baseline. Bound proteins were eluted through the addition of buffer B (50 mM Tris-HCl, 500 mM NaCl, 500 mM Imidazole, pH 8.5) in an 0-500 mM imidazole gradient over 20 mL. Fractions containing SapL1 were pooled and concentrated by ultrafiltration (Pall, 10kDa cut-off) prior buffer exchange to 50 mM Tris-HCl, 100 mM NaCl, pH 8.5 using PD10 desalting columns (GE Healthcare Life Sciences). Then, the fusion was cleaved off overnight at 19°C using the TEV protease produced in the lab in 1:50 ratio and addition of 0.5 mM EDTA and 0.25 mM TCEP. The sample was loaded on the His-trap column to separate the cleaved protein collected in the flow-through from the TEV protease and potential uncleaved sample retained and eluted with imidazole. Sapl1 containing fractions were concentrated by centrifugation to the desired concentration.

Size exclusion chromatography

Size exclusion chromatography (SEC) was performed using a High-Resolution ENrich™ SEC 70 column (Bio-Rad) on NGC™ chromatography system (Bio-Rad). Column was equilibrated with 50 ml of buffer D (20 mM MES, 100 mM NaCl, pH 6.5) and 200 µl of sample at 10 mg mL^{-1} were injected into the system followed by 40 ml isocratic elution on buffer D supplemented with 20 mM α-methyl fucoside, 0.1% L-rhamnose, 5 M NaCl, 2 M NaCl or 500 mM NaCl, according to the experiment. Fractions were monitored by the absorbance at 280 nm and 0.5 mL fractions were collected in the resolving region of the column.

Dynamic light scattering (DLS)

DLS analyses were performed using a Zetasizer™ NanoS (Malvern Panalytical) with a 40 µl quartz cuvette. Measurements were performed in triplicate on protein sample at 1 $mg{\cdot}ml^{-1}$ in buffer C (50 mM Tris-HCl, 100 mM NaCl, pH 8.5) after centrifugation.

Thermal Shift Assay (TSA)

The thermal stability of SapL1 was analyzed by TSA with the MiniOpticon real-time PCR system (Bio-Rad). Prior assay, buffer stocks at 100 mM and a mixture containing 70 µl of SapL1 at 1 mg ml^{-1}, 7 µl of 500x Sypro Orange (Merk Sigma-Aldrich,) and 63 µl of ultrapure H_2O were prepared. Then, 7.5 µl of H_2O, 12.5 µl of the corresponding buffer and 5 µl of the protein/Sypro mixture were mixed in 96-well PCR microplates. The heat exchange test was then carried out from 20°C to 100°C with a heating rate of $1°C{\cdot}min^{-1}$. Fluorescence intensity was measured with Ex/Em: 490/530 nm and the data processing was performed with the CFX Manager software.

Isothermal Titration Calorimetry (ITC)

Experiments were performed using a Microcal ITC200 calorimeter (Malvern Panalytical) with 40 µl of L-fucose 2.5 mM in the syringe and 200 µl of protein 0.04 mM in the sample cell. 2 µl injections in a range of 120 seconds while stirring at 1000 rpm were performed.

Experimental data were adjusted to a theoretical titration curve by the Origin ITC Analysis software.

Hemagglutination assay

Agglutination test was performed with fresh rabbit erythrocytes (bioMérieux, Lyon) in U 96-well plates (Nalgene). For the test, 150 mM NaCl was used as a negative control and 1 mg ml^{-1} LecB of *P. aeruginosa* as a positive control. 50 µl of sample was prepared at 0.1 mg·ml^{-1} and submitted to serial double dilutions. 50 µl of rabbit erythrocytes 3% were added to each well prior incubation of the plate at room temperature. After 2 hours, the result of the experiment was evaluated and agglutination activity was calculated according to the dilution of the protein.

Glycan arrays

Protein was labeled with Fluorescein Isothiocyanate (FITC) (Merk Sigma-Aldrich) according to the supplier's instructions with slight modifications. Briefly two milligrams of protein were dissolved in 1 ml of buffer E (100 mM Na_2CO_3, 100 mM NaCl, pH 9); then, 40 µl of FITC at 1 mg·ml^{-1}, previously dissolved in DMSO, were gradually added to the protein solution and the mixture was gently stirred at room temperature overnight. Next day, the solution was supplemented with NH_4Cl to a final concentration of 50 mM and free FITC was removed using PD10 column with PBS as mobile phase. Protein concentration was determined at A_{280} and FITC at A_{490} using a NanoDrop™ 200 (Thermo Scientific) and Fluorescein/Protein molar ratio (F/P) was estimated by the following formula:

$$\text{Molar}\frac{\text{F}}{\text{P}} = \frac{\text{MW}}{389} \times \frac{\frac{A_{495}}{195}}{[A_{280} - (0.35 \times A_{495})]E^{0.1\%}}$$

Where MW is the molecular weight of the protein, 389 is the molecular weight of FITC, 195 is the absorption E 0.1% of bound FITC at 490 nm at pH 13.0, (0.35 X *A*495) is the correction factor due to the absorbance of FITC at 280 nm, and E 0.1% is the absorption at 280 nm of a protein at 1.0 mg·ml^{-1}, Being an ideal F/P should be 0.3 > 1.

Labeled lectin was sent to the Consortium for Functional Glycomics (CFG; Boston, MA, USA) and its binding properties were assayed at 5 and 50 $\mu g \cdot ml^{-1}$ on a "Mammalian Glycan Array version 5.4" which contain 585 glycans in replicates of 6. The highest and lowest signal of each set of replicates were eliminated and the average of the remaining data was normalized to the percentages of the highest RFU value for each analysis. Finally, the percentages for each glycan were averaged at different lectin concentrations.

Crystallization and data collection

Crystal screening was performed using the hanging-drop vapour diffusion technique by mixing equal volumes of pure protein at 5 $mg \cdot ml^{-1}$ and precipitant solutions from commercial screenings of Molecular Dimensions (Newmarket, UK). 2 µl drops were incubated at 19°C until crystals appeared. A subsequent optimization of positive conditions for SapL1 crystallization was carried out and crystals suitable for X-ray diffraction analysis were obtained under solution containing 100 mM Bicine pH 8.5, 1.5 M ammonium sulfate $(NH_4)_2SO_4$ and 12% v/v glycerol. Crystals were soaked in mother liquor supplemented with 10% (v/v) glycerol, prior to flash cooling in liquid nitrogen. Data collection was performed on PX1 beamline at SOLEIL Synchrotron (Saint Aubin, FR) using a Pixel detector.

Structure determination

Data were processed using XDS[57] software and were converted to structure factors using the CCP4 program package v.6.1[58], with 5% of the data reserved for Rfree calculation. The structure was determined using the molecular-replacement method with Phaser v.2.5[59], using the structure of FleA dimer (PDB entry 4D4U[35]) as starting model. Model refinement was performed using REFMAC 5.8[60] alternated with manual model building in Coot v.0.7[61]. Sugar residues and other compounds that were present were placed manually using Coot and validated using Privateer[62]. The final model has been validated and deposited in the PDB Database with accession number 6TRV.

Figures

Figures were created using PyMOL Version 1.8.4 (Schrödinger), ChemDraw Version 15, ESPript Version 3.0[63] and PowerPoint 16.

References

1 Tuite, N. L. & Lacey, K. Vol. 968 (eds Louise O'Connor & Barry Glynn) 1-23 (Humana Press, Totowa, NJ, 2013).

2 Ramirez-Garcia, A. *et al. Scedosporium* and *Lomentospora*: An updated overview of underrated opportunists. *Med. Mycol.* **56**, S102-S125, doi:10.1093/mmy/myx113 (2018).

3 Goldman, C., Akiyama, M. J., Torres, J., Louie, E. & Meehan, S. A. *Scedosporium apiospermum* infections and the role of combination antifungal therapy and GM-CSF: a case report and review of the literature. *Medical Mycol. Case Rep.* **11**, 40-43, doi:10.1016/j.mmcr.2016.04.005 (2016).

4 Rammaert, B. *et al.* Perspectives on Scedosporium species and Lomentospora prolificans in lung transplantation: Results of an international practice survey from ESCMID fungal infection study group and study group for infections in compromised hosts, and European Confederation. *Transpl. Infect. Dis.* **21**, 1-8, doi:10.1111/tid.13141 (2019).

5 Luplertlop, N. *Pseudallescheria/Scedosporium* complex species: from saprobic to pathogenic fungus. *J. Mycol. Med.* **28**, 249-256, doi:10.1016/j.mycmed.2018.02.015 (2018).

6 Gilgado, F., Cano, J., Gené, J. & Guarro, J. Molecular phylogeny of the *Pseudallescheria boydii* species complex: proposal of two new species. *J. Clin. Microbiol.* **43**, 4930-4942, doi:10.1128/JCM.43.10.4930-4942.2005 (2005).

7 Duvaux, L. *et al.* Draft genome sequence of the human-pathogenic fungus *Scedosporium boydii*. *Genome Announc.* **5**, e00871-00817, doi:10.1128/genomeA.00871-17 (2017).

8 Pérez-Bercoff, Å. *et al.* Draft genome of Australian environmental strain WM 09.24 of the opportunistic human pathogen *Scedosporium aurantiacum*. *Genome Announc.* **3**, e01526-01514, doi:10.1128/genomeA.01526-14 (2015).

9 Vandeputte, P. *et al.* Draft genome sequence of the pathogenic fungus *Scedosporium apiospermum*. *Genome Announc.* **2**, e00988-00914, doi:10.1128/genomea.00988-14 (2014).

10 Pereira De Mello, T. *et al.* Conidial germination in *Scedosporium apiospermum, S . aurantiacum, S . minutisporum* and *Lomentospora prolificans*; influence of growth conditions and antifungal susceptibility profiles. *Mem. Inst. Oswaldo Cruz* **111**, 484-494, doi:10.1590/0074-02760160200 (2016).

11 Tortorano, A. M. *et al.* ESCMID and ECMM joint guidelines on diagnosis and management of hyalohyphomycosis: *Fusarium spp., Scedosporium spp.* and others. *Clin. Microbiol. Infect.* **20**, 27-46, doi:10.1111/1469-0691.12465 (2014).

12 Troke, P. *et al.* Treatment of scedosporiosis with voriconazole: clinical experience with 107 patients. *Antimicrob. Agents Chemother.* **52**, 1743-1750, doi:10.1128/AAC.01388-07 (2008).

13 Rodriguez-tudela, J. L. *et al. Scedosporium apiospermum:* changing clinical spectrum of a therapy-refractory opportunist. *Med. Mycol.* **44**, 295-327, doi:10.1080/13693780600752507 (2006).

14 Cortez, K. J. *et al.* Infections Caused by Scedosporium spp. *Clin. Microbiol. Rev.* **21**, 157-197, doi:10.1128/CMR.00039-07 (2008).

15 Theuretzbacher, U. & Piddock, L. J. V. Non-traditional antibacterial therapeutic options and challenges. *Cell Host Microbe* **26**, 61-72, doi:10.1016/j.chom.2019.06.004 (2019).

16 Krachler, A. M. & Orth, K. Targeting the bacteria-host interface strategies in anti-adhesion therapy. *Virulence* **4**, 284-294, doi:10.4161/viru.24606 (2013).

17 Ofek, I., Hasty, D. L. & Sharon, N. Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol. Med. Microbiol.* **38**, 181-191, doi:10.1016/S0928-8244(03)00228-1 (2003).

18 Figueiredo, R. T. *et al.* TLR4 recognizes *Pseudallescheria boydii* conidia and purified rhamnomannans. *J. Biol. Chem.* **285**, 40714-40723, doi:10.1074/jbc.M110.181255 (2010).

19 Pinto, M. R., Gorin, P. A. J., Wait, R., Mulloy, B. & Barreto-Bergter, E. Structures of the O-linked oligosaccharides of a complex glycoconjugate from *Pseudallescheria boydii*. *Glycobiology* **15**, 895-904, doi:10.1093/glycob/cwi084 (2005).

20 Pinto, M. R. *et al.* Involvement of peptidorhamnomannan in the interaction of *Pseudallescheria boydii* and HEp2 cells. *Microbes Infect.* **6**, 1259-1267, doi:10.1016/j.micinf.2004.07.006 (2004).

21 Bittencourt, V. C. B. *et al.* An a-glucan of *Pseudallescheria boydii* is involved in fungal phagocytosis and Toll-like receptor activation. *J. Biol. Chem.* **281**, 22614-22623, doi:10.1074/jbc.M511417200 (2006).

22 Ghamrawi, S. *et al.* Cell wall modifications during conidial maturation of the human pathogenic fungus *Pseudallescheria boydii*. *PLoS ONE* **9**, e100290-e100290, doi:10.1371/journal.pone.0100290 (2014).

23 Rollin-Pinheiro, R., Liporagi-lopes, L. C., Meirelles, J. V. D. & Souza, L. M. D. Characterization of *Scedosporium apiospermum* glucosylceramides and their involvement in fungal development and macrophage functions. *PLoS ONE* **9**, e98149-e98149, doi:10.1371/journal.pone.0098149 (2014).

24 De Mello, T. P. *et al. Scedosporium apiospermum, Scedosporium aurantiacum, Scedosporium minutisporum* and *Lomentospora prolificans*: a comparative study of surface molecules produced by conidial and germinated conidial cells. *Mem. Inst. Oswaldo Cruz* **113**, 1-8, doi:10.1590/0074-02760180102 (2018).

25 Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta* **1760**, 527-537, doi:10.1016/j.bbagen.2005.12.008 (2006).

26 Tamburrini, A., Colombo, C. & Bernardi, A. Design and synthesis of glycomimetics: Recent advances. *Med. Res. Rev.* **40**, 495-531, doi:10.1002/med.21625 (2020).

27 Sattin, S. & Bernardi, A. Glycoconjugates and glycomimetics as microbial anti-adhesives. *Trends Biotechnol.* **34**, 483-495, doi:10.1016/j.tibtech.2016.01.004 (2016).

28 Pieters, R. J. Carbohydrate mediated bacterial adhesion. *Adv. Exp. Med. Biol.* **715**, 227-240, doi:10.1007/978-94-007-0940-9_14 (2011).

29 Houser, J. *et al.* A soluble fucose-specific lectin from *Aspergillus fumigatus* conidia - structure, specificity and possible role in fungal pathogenicity. *PLoS ONE* **8**, e83077-e83077, doi:10.1371/journal.pone.0083077 (2013).

30 Kerr, S. C. *et al.* FleA expression in *Aspergillus fumigatus* is recognized by fucosylated structures on mucins and macrophages to prevent lung infection. *PLOS Pathog.*, e100555-e100555, doi:10.1371/journal.ppat.1005555 (2016).

31 Sakai, K., Hiemori, K., Tateno, H., Hirabayashi, J. & Gonoi, T. Fucose-specific lectin of *Aspergillus fumigatus:* binding properties and effects on immune response stimulation. *Med. Mycol.* **57**, 71-83, doi:10.1093/mmy/myx163 (2019).

32 Thornton, D. J., Rousseau, K. & McGuckin, M. A. Structure and function of the polymeric mucins in airways mucus. *Annu. Rev. Physiol.* **70**, 459-486, doi:10.1146/annurev.physiol.70.113006.100702 (2008).

33 Glick, M. C., Kothari, V. A., Liu, A., Stoykova, L. I. & Scanlin, T. F. Activity of fucosyltransferases and altered glycosylation in cystic fibrosis airway epithelial cells. *Biochimie* **83**, 743-747, doi:10.1016/s0300-9084(01)01323-2 (2001).

34 Lamblin, G. *et al.* Human airway mucin glycosylation: a combinatory of carbohydrate determinants which vary in cystic fibrosis. *Glycoconj. J.* **18**, 661-684, doi:10.1023/A:1020867221861 (2001).

35 Houser, J. *et al.* Structural insights into *Aspergillus fumigatus* lectin specificity : AFL binding sites are functionally non-equivalent. *Acta Crystallogr. D Biol. Crystallogr.*, 442-453, doi:10.1107/S1399004714026595 (2015).

36 Frank, D. E. *et al.* Thermodynamics of the interactions of Lac repressor with variants of the symmetric Lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.* **267**, 1186-1206, doi:10.1006/jmbi.1997.0920 (1997).

37 Capp, M. W. *et al.* Compensating effects of opposing changes in putrescine (2+) and K+ concentrations on lac repressor-lac operator binding: in vitro thermodynamic analysis and in vivo relevance. *J. Mol. Biol.* **258**, 25-36, doi:10.1006/jmbi.1996.0231 (1996).

38 Cayley, S., Lewis, B. A., Guttman, H. J. & Record, M. T. Characterization of the cytoplasm of *Escherichia coli K-12* as a function of external osmolarity implications for protein-DNA interactions in vivo. *J. Mol. Biol.* **222**, 281-300, doi:10.1016/0022-2836(91)90212-o (1991).

39 Richeys, B. *et al.* Variability of the intracellular ionic environment of *Escherichia coli.* Differences between *in vitro* and *in vivo* effects of ion concentrations on protein-DNA interactions and gene expression. *J. Biol. Chem.* **262**, 7157-7164 (1987).

40 Harries, D. & Rosgen, J. Vol. 84 679-735 (Elsevier, 2008).

41 Massiah, M. A., Wright, K. M. & Du, H. Obtaining soluble folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS: application to LB and M9 minimal media. *Curr. Protoc. Protein Sci.* **84**, 1-24, doi:10.1002/0471140864.ps0613s84 (2016).

42 Pettitt, B. M. & Bolen, D. W. Protein folding, stability, and solvation structure in osmolyte solutions. *Biophys. J.* **89**, 2988-2997, doi:10.1529/biophysj.105.067330 (2005).

43 Singh, L. R. *et al.* Forty years of research on osmolyte-induced protein folding and stability. *J. Iran. Chem. Soc.* **8**, 1-23 (2011).

44 Ajito, S., Iwase, H., Takata, S.-i. & Hirai, M. Sugar-mediated stabilization of protein against chemical or thermal denaturation. *The Journal of Physical Chemistry B* **122**, 8685-8697, doi:10.1021/acs.jpcb.8b06572 (2018).

45 Bonnardel, F. *et al.* Architecture and evolution of blade assembly in β-propeller lectins. *Structure* **27**, 764-775.e763, doi:10.1016/j.str.2019.02.002 (2019).

46 Lehot, V. *et al.* Multivalent fucosides with nanomolar affinity for the *Aspergillus fumigatus* lectin FleA prevent spore adhesion to pneumocytes. *Chem. Eur. J.* **24**, 19243-19249, doi:10.1002/chem.201803602 (2018).

47 Makyio, H. *et al.* Six independent fucose-binding sites in the crystal structure of *Aspergillus oryzae* lectin. *Biochem. Biophys. Res. Commun.* **477**, 477-482, doi:10.1016/j.bbrc.2016.06.069 (2016).

48 Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. & Imberty, A. Crystal structure of fungal lectin. Six-bladed β-propeller fold and novel fucose recognition mode for *Aleuria aurantia* lectin. *J. Biol. Chem.* **278**, 27059-27067, doi:10.1074/jbc.M302642200 (2003).

49 Debourgogne, A., Dorin, J. & Machouart, M. Emerging infections due to filamentous fungi in humans and animals: only the tip of the iceberg? *Env. Microbiol. Rep.* **8**, 332-342, doi:10.1111/1758-2229.12404 (2016).

50 Hogan, L. H., Klein, B. S. & Levitz, S. M. Virulence factors of medically important fungi. *Clin. Microbiol. Rev.* **9**, 469-488, doi:10.1128/CMR.9.4.469-488.1996 (1996).

51 Kuo, K. C., Kuo, H. C., Huang, L. T., Lin, C. S. & Yang, S. N. The clinical implications of ABO blood groups in *Pseudomonas aeruginosas* sepsis in children. *J. Microbiol. Immunol. Infect.* **46**, 109-114, doi:10.1016/j.jmii.2012.01.003 (2013).

52 Scanlin, T. F. & Glick, M. C. Terminal glycosylation in cystic fibrosis. *Biochim. Biophys. Acta* **1455**, 241-253, doi:10.1016/S0925-4439(99)00059-9 (1999).

53 Taylor-Cousar, J. L. *et al.* Histo-blood group gene polymorphisms as potential genetic modifiers of infection and cystic fibrosis lung disease severity. *PLoS ONE* **4**, e4270-e4270, doi:10.1371/journal.pone.0004270 (2009).

54 Audfray, A. *et al.* Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes. *J. Biol. Chem.* **287**, 4335-4347, doi:10.1074/jbc.M111.314831 (2012).

55 Kostlanova, N. *et al.* The fucose-binding lectin from *Ralstonia solanacearum*. A new type of β-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. *J. Biol. Chem.* **280**, 27839-27849, doi:10.1074/jbc.M505184200 (2005).

56 Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W. H. & Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.* **81**, 6807-6816, doi:10.1128/jvi.00338-07 (2007).

57 Kabsch, W. XDS. *Acta Crystallographica* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

58 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).

59 McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

60 Murshudov, G. N. & Nicholls, R. A. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).

61 Emsley, P. & Lohkamp, B. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

62 Aguirre, J. *et al.* Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833-834, doi:https://doi.org/10.1038/nsmb.3115 (2015).

63 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320-324, doi:10.1093/nar/gku316 (2014).

Acknowledgments: We would like to thank to Valérie Chazalet and Emilie Gillon for their technical assistance. We are grateful to the Consortium for Functional Glycomics for providing the glycan array resource. We also would like to thank for access to the beamlines BM30A-FIP at the European Synchrotron Radiation Facility (ESRF), Grenoble, France where initial tests were performed and Proxima 1 at SOLEIL Synchrotron, Saint Aubin, France (proposal number 20170827), where final data were collected. Thanks to our local contacts Jean-Luc Ferrer, Serena Sirigu, and Pierre Legrand for their assistance and technical support.

Author Contributions: D.M.A. performed cloning, protein expression, protein purification, protein labelling, hemmaglutination, isothermal microcalorimetry measurements. D.M.A. participated in all steps of protein crystallography and data analysis, wrote the original draft of the manuscript and prepared all figures. J-P.B. provided the genome data for the bioinformatics studies and contributed to revise the manuscript. R.J.P. participated in the supervision of D.M.A, funding acquisition and revision of the manuscript. A.V. administered the project, conceived the design of the study, obtained funding, evaluated the results, contributed to data analysis, supervised D.M.A., helps in the preparation of the manuscript and contributed to revise the manuscript and the figures. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant (H2020-

MSCA-ITN-2017-EJD-765581), from the French cystic fibrosis association Vaincre la Mucoviscidose. The project was also supported by Glyco@Alps (ANR-15-IDEX02).

Accession numbers

The structure of SapL1 was deposited as PDB ID: 6TRV.

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

Figure captions

Figure 1. SapL1 production. A) Schematic representation of the genetic constructs used for the expression of SapL1. B) SDS-PAGE of fractions collected from SapL1 purification using pET32-TEV vector. IF: Insoluble fraction, SF: Soluble fraction, FT: Flow through, W: wash and elution, from right to left respectively. Red arrow indicates inclusion bodies from insoluble fractions. C) Representative chromatograms of SapL1 purification before and after the process with their respective profile on 15% SDS-polyacrylamide gels (insoluble fraction, soluble fraction, flow through, washing and elution, from left to right respectively). The fractions containing SapL1 are delimited.

Figure 2. Rhamnose influence on SapL1 solubility. A) SapL1 purification chromatograms at 0.1%, 0.2%, and 1% rhamnose with their corresponding SDS-PAGE profile. IF: Insoluble fraction, SF: Soluble fraction, FT: Flow through, W: wash and elution, from left to right respectively. Black rectangles indicate the elution peaks and elution fractions containing SapL1 on SDS-polyacrylamide gels. B) Numerical values derived from the integration of areas under the peaks corresponding to SapL1 for each experiment. Black and striped bars represent the results for experiments performed at 16°C and 20°C, respectively.

Figure 3. SapL1 carbohydrate binding properties. A) Hemagglutination assay of SapL1 on fresh rabbit erythrocytes. Negative and positive controls consist of 150 mM NaCl and the lectin LecB from *Pseudomonas aeruginosa*, respectively. B) Titration of SapL1 with L-fucose and α-methyl-fucoside with the thermogram and the integration displayed at the top and bottom, respectively. C) Analysis of the interactions of SapL1 with blood group epitopes. The graph shows the relative binding of SapL1 to glycans containing Lewis and ABH blood group antigens from the 90 hits identified as binders.

Figure 4: SapL1 overall structure. A) Surface and cartoon representation of SapL1 monomer colored from blue (*N*-terminal end) to red (*C*-terminal end). BP: Binding

pockets. B) Representation of SapL1 dimer colored by monomer with ligand depicted in sticks.

Figure 5: SapL1/FleA comparison. A) Overlay of SapL1 (blue, PDB: 6TRV)) and FleA (green, PDB: 4D4U) structures. Red arrows indicate the main differences between structures. B) Sequence alignment of SapL1 and FleA with display of their secondary structure elements. The residues involved in ligand interactions of each pocket are indicated by circles for hydrophobic and stacking interactions and by triangles for hydrogen bonds colored according to the binding pocket: BP-1, blue; BP-2 cyan; BP-3 green; BP-4, orange, BP-5, brown; BP-6, red.

Figure 6: SapL1-ligand interactions. A) The sequence alignment of the 6 blades of SapL1. The residues involved in ligand binding interactions are indicated in solid boxes: BP-1, blue; BP-2 cyan; BP-3 green; BP-4, orange, BP-5, brown; BP-6, red. Striped green boxes indicate the four additional residues expected to be involved in fucose binding within BP-3 that were not visible in the structure since only glycerol was attached to this pocket. B) Zoom on the interactions of ligand with each binding site of SapL1 and their schematic representation.