Accepted Manuscript

Crystal structure of the GalNAc/Gal-specific agglutinin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* reveals novel adaptation of a β-trefoil domain

Gerlind Sulzenbacher, Véronique Roig-Zamboni, Willy J. Peumans, Pierre Rougé, Els J.M. Van Damme, Yves Bourne

Please cite this article as: Sulzenbacher, G., Roig-Zamboni, V., Peumans, W.J., Rougé, P., Van Damme, E.J.M. & Bourne, Y., Crystal structure of the GalNAc/Gal-specific agglutinin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* reveals novel adaptation of a β-trefoil domain, *Journal of Molecular Biology* (2010), doi:10.1016/j.jmb.2010.05.038

Crystal structure of the GalNAc/Gal-specific agglutinin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* reveals novel adaptation of a β-trefoil domain

Gerlind Sulzenbacher[1], Véronique Roig-Zamboni[1], Willy J. Peumans[2],

Pierre Rougé[3], Els J.M. Van Damme[2] and Yves Bourne[1]*

[1]Architecture et Fonction des Macromolécules Biologiques (AFMB, UMR-6098) CNRS,

Université Aix-Marseille, Campus Luminy, Case 932, F-13288 Marseille cedex 09, France

[2] Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology,

Ghent University, Coupure Links 653, Ghent, Belgium

[3]Surfaces Cellulaires et Signalisation chez les Végétaux, UMR-CNRS 5546, Pôle de

Biotechnologie Végétale, Toulouse, France

*To whom correspondence should be addressed:

Phone: +33-491825566

Fax: +33-491266720

email: yves.bourne@afmb.univ-mrs.fr

Abstract

A lectin from the phytopathogenic ascomycete *Sclerotina sclerotiorum* that shares only weak sequence similarity with characterized fungal lectins has recently been identified. *Sclerotina sclerotiorum* agglutinin (SSA) is a homodimeric protein consisting of two identical subunits of ~17 kDa and displays specificity primarily towards Gal/GalNAc. Glycan array screening indicates that SSA readily interacts with Gal/GalNAc-bearing glycan chains. The crystal structures of SSA in the ligand-free form and in complex with the Gal-β1,3-GalNAc (T-antigen) disaccharide have been determined at 1.6 and 1.97 Å resolution, respectively. SSA adopts a β-trefoil domain as previously identified for other carbohydrate-binding proteins of the ricin B-like lectin superfamily and accommodates terminal non-reducing galactosyl and N-acetylgalactosaminyl glycans. Unlike other structurally related lectins, SSA contains a single carbohydrate-binding site at site α. SSA reveals a novel dimeric assembly markedly dissimilar to those described earlier for ricin-type lectins. The present structure exemplifies the adaptability of the β-trefoil domain in the evolution of fungal lectins.

Introduction

Numerous mushrooms and other fungi contain carbohydrate-binding proteins commonly known as lectins [1-3]. Though not all previously identified lectins have been purified and characterized, the available biochemical, sequence and structural data provide ample evidence that fungi, and more specifically the largest and diverse phylum *Ascomycota*, express a heterogeneous mixture of carbohydrate-binding proteins.

Fungal lectins can be assigned as storage proteins [4,5] and are involved in morphogenesis and development of the fungi [6,7]. Contrary to established roles of bacterial lectins in host–parasite interactions, the functional roles of lectins expressed by phytopathogenic fungi is poorly understood. Although many believe that fungal lectins mediate host–parasite interactions [8] similar to bacterial adhesins, they have been implicated in the process of specific recognition in mycoparasitism and mediate interactions between parasitic and pathogenic fungi and their hosts [5,9]. However, none of these assigned roles are clearly established.

To further corroborate the diversity of fungal lectins, a cytoplasmic lectin has recently been characterized from the phytopathogenic ascomycete *Sclerotinia sclerotiorum*, a fungal pathogen that has the broadest host range of all known fungi [10]. The so-called *Sclerotinia sclerotiorum* agglutinin (SSA) can be considered the prototype of a small lectin family, which hitherto has been documented exclusively in the *Sclerotiniaceae* family. Although orthologs of SSA have been isolated from sclerotes of several *Sclerotina* species [11], complete sequences are available only for SSA and the orthologs from *Botryotinia fuckeliana, Pyrenophora tritici-repentis* and *Cochliobolus heterostrophus*. Sequence comparison reveals that the SSA orthologs represent a novel family of fungal lectins that shares limited sequence similarity with other fungal lectins. Preliminary biochemical and structural characterization has revealed that SSA readily interacts with glycolipid glycans harboring terminal non-reducing Gal or

3

GalNAc residues and galactosylated N-glycans with highest affinity for α1-3 branched mono- and multi-antennary chains [12]. Since SSA appears to exhibit sugar-binding properties distinct from other previously characterized fungal lectins, it might identify a new lectin subfamily, more widespread within the phylum *Ascomycota*.

A molecular modeling study using threading algorithms has shown that SSA is remotely related to the β-trefoil domain of the non-toxin haemagglutinin HA33/A from *Clostridium botulinum* [13] that is classified in the ricin B-like lectin superfamily. The ricin B chain, which belongs to the ricin-like or R-type lectin family, is characterized by the presence of a triple $(QXW)_3$ motif [14] and is widespread in bacteria, animals and plants, but has not been studied extensively in fungi. Only a few ricin-B type lectins have been isolated from fungi, the best known examples being the multi-modular lectins from the mushrooms *Marasmius oreades* [15] and *Polyporus squamosus* [16], and the pore-forming lectin from the mushroom *Laetiporus sulphureus* [17]. However, we may anticipate that the ricin-B domain is also fairly widespread in the newly available fungal genomes .

In a further step towards understanding the carbohydrate specificity and the oligomeric assembly of this novel fungal lectin, we report the 1.60 Å and 1.97 Å resolution crystal structure of SSA in absence or presence of bound Gal-β1,3-GalNAc (T-antigen). The monomer structure of SSA closely resembles that of the ricin B lectin domain, a trefoil-based fold observed in many lectins and carbohydrate-binding domains, but the shape and hydrophobic character of the unique carbohydrate binding site are markedly modified compared to that of other members of the β-trefoil fold family, consistent with the carbohydrate-binding specificity of the lectin. SSA shares highest structural similarity with the β-trefoil domain of the multi-modular Galα1,3Gal specific lectin from the mushroom *Marasmius oreades* [18], the hemagglutinin HA1 and HA-17 subcomponents from the *Clostridium botulinum* type C progenitor toxin [19] and a serotype D toxin complex from

4

Clostridium botulinum [20]. SSA exhibits a novel dimeric assembly that is rarely observed within members of the ricin-B type lectin family and may participate to multivalent lectin-carbohydrate cross-linking interactions. Altogether these results suggest that SSA identifies a new lectin subfamily with specific sequence and carbohydrate binding properties and might give indications towards the possible involvement of this subfamily of fungal lectins in the regulation of morphogenesis or pathogenesis.

Results and discussion

Quality and overall view of the structure

The structures of SSA in the apo form and in complex with the T-antigen disaccharide were solved from crystals grown in two distinct space groups. They show well-defined electron densities for most of the protein regions and bound sugar and exhibit excellent stereochemistry (Materials & Methods, **Table 1**).

The SSA monomer with overall dimensions of 30 x 30 x 25 Å belongs to the β-trefoil fold family and adopts a typical three-lobed organization, that consists of three four-stranded β sheets (β1-β4, β5-β8 and β9-β12), referred to as subdomains α, β and γ and displaying characteristic pseudo-three fold symmetry [21] (**Fig. 1**). Structure superposition of the three subdomains reveals that the overall structure of these domains is quite similar to each other with rmsd between subdomains in the 1.55 to 1.66 Å range for ~40 Cα atoms, with subdomain β being the smallest. In contrast to most of the extracellular R-type lectins that have been extensively studied, SSA lacks the two disulfide bridges that stabilize the β-trefoil fold, a feature that is also observed in most cytoplasmic ricin-B lectin domains. Moreover, the key/characteristic signature QXW motif reminiscent of the R-type lectins is present only in the third subdomain of SSA [14].

A DALI search for close structural homologs of SSA revealed top-ranked hits (Z-score values > 20) for the N-terminal β-trefoil domain of the *Marasmius oreades* agglutinin (MOA) [18], the C-terminal β-trefoil domain of the HA1 subcomponent of botulinum type C toxin [19] and the 17 kDa HA-17 subcomponent of a serotype D toxin complex from *Clostridium botulinum* [20]. Surprisingly, SSA appears to possess higher structural homology with multimodular proteins harboring a β-trefoil domain rather than with individual/isolated R-type lectins. The rmsd between the SSA structure and those of the three top-ranked homologs above are 1.24 Å, 0.94 Å and 1.21 Å for 125, 124 and 118 Cα atoms, respectively. The major structural

differences between SSA and the three lectin homologs are confined to the shape and dimensions of loop regions connecting β-strands and forging the architecture of the carbohydrate-binding sites.

Glycan array screening

Analysis of the carbohydrate binding specificity of SSA by glycan array screening revealed strong interaction of the lectin with GalNAc-containing oligosaccharides (**Table 2**), consistent with previously reported binding data using various pNP-oligosaccharides as potential ligands [12]. Although there is a clear preference for terminal non-reducing Gal/GalNAc residues, the specificity of SSA is rather broad and glycans with internal Gal/GalNAc residues are also recognized. SSA also reacts with branched oligosaccharides, but addition of extra sugar moieties can lead to low affinity glycans. For instance, glycans with α1,2-fucosyl residues linked to the Gal moiety of the GalNAcα1,4Gal core show a significant response, while those with fucose linked to the terminal Gal, as found in the core H (type 3) blood group antigen, show only a moderate response. Such an interaction with fucosylated glycans was revealed in a previous binding analysis using version 2 of the printed glycan array [22]. While the profile of binding specificity is similar for the three (0.25, 0.5 and 1.0 µg/ml) lectin concentrations tested, complex N-glycans were recognized only at the two higher concentrations (**Supp. Fig. S1**).

The carbohydrate-binding site

The structure of SSA in complex with the Galβ1,3GalNAc disaccharide reveals a well-ordered carbohydrate at only one (site α) out of the three possible sites (**Fig. 2**) and provides detailed lectin-carbohydrate interactions (**Table 3**). The nonreducing Gal moiety is tightly bound at the primary binding site within the α site, consistent with the requirement of

7

unsubstituted non-reducing terminal Gal/GalNAc residues for binding activity [12]. The aromatic ring of Tyr37 establishes a stacking interaction against the B face of Gal while the perpendicularly oriented Trp24 indole ring makes additional Van der Waals contacts with the C6-O6 atom pair. The axial O4 hydroxyl group is tightly anchored to the nitrogen backbone atom of Glu25 and the oxygen atom of the Asn22 side chain. The neighboring O3 hydroxyl is hydrogen bound to the side chains of Asn22 and Asn46 and to Ser44 via a water molecule. The O2 hydroxyl is bound to Asn46 via a water-mediated interaction. The GalNAc moiety is exposed to the solvent and interacts only weakly with residues in the carbohydrate-binding site (**Fig. 2**), consistent with the diversity of sugar linkages and moieties allowed beyond the requirement of terminal Gal or GalNAc residues (**Table 2**). Indeed, only the O4 hydroxyl of GalNAc acts as hydrogen-bond donor to the carbonyl oxygen of Glu25, while the O6 hydroxyl weakly interacts with the Glu25 carboxyl group. The overall conformation of the bound Galβ1,3GalNAc ($\Phi/\Psi = 44°/-10°$) falls within the minimum energy area calculated for the disaccharide [23]. Structural comparison of the apo and sugar-bound form reveals no major conformational changes upon binding of the disaccharide. In the apo form, a glycerol molecule is observed in site α and partly mimics a bound Gal moiety (**Fig. 2**).

Examination of the amino-acid sequence of SSA reveals that the key aromatic side chains (Tyr37 and Trp24) for sugar binding in site α are absent in sites β and γ, and may explain the lack of sugar binding activity in these two sites (**Fig. 1**). The carbohydrate binding site of SSA is shallow and solvent-exposed and may permit accommodation of both the α and β conformation of the 1,3 or 1,4 glycosidic linkage. The carbohydrate-binding site of SSA is filled with 14 well-ordered water molecules that are mostly clustered in the vicinity of the O2 and O3 positions of Gal, suggesting that SSA could accommodate the larger acetamido group of GalNAc, other sugar substituents or bisecting/multi-antennary glycan chains. Indeed, models of the α-GalNAc complex generated using Autodock Vina [24] showed that the docked

orientation of α-GalNac is nearly identical to that of Gal in the T-antigen complex observed in the X-ray crystal structure. The N-acetyl group lies in the vicinity of the Glu25-Gly26 backbone atoms and the side chains of Ser44 and Asn46 (**Fig. 2**), indicating that terminal N-acetylgalactosaminyl moieties could also be easily accommodated into the carbohydrate-binding site of SSA, consistent with the glycan array data (**Table 2**).

Despite low sequence similarity, the carbohydrate-binding site of SSA reveals structural similarities to other previously reported ricin-like or R-type lectin-sugar complexes. Although SSA displays only a single carbohydrate-binding site at site α, the mode of binding to Gal resembles that observed in other galactose-binding lectins. Structural comparison of SSA and the galactose-binding lectin EW29 [25] (from the earthworm *Lumbricus terrestris*), which contains two carbohydrate binding sites at sites α and γ (**Fig. 2**), reveals a similar orientation of the terminal Gal moiety in EW29, with an outward 1.6 Å displacement of the sugar ring compared to its position in SSA along with a shift of the β1-β2 loop. Despite the close structural similarities between SSA and the β-trefoil domain of (MOA) (rmsd of 1.24 Å for 125 Cα atoms), the Gal moiety adopts a distinct position and orientation in the two structures. In MOA, the non-reducing terminal Gal moiety of the blood group B trisaccharide is shifted by ~1.5 Å and rotated by ~30° compared to its position in SSA to favor interactions between the O4 and O6 hydroxyl groups of Gal and MOA (**Fig. 2**).

A novel dimeric interface

The SSA crystal structure identifies a novel dimer interface for a R-type fungal lectin, consistent with data from gel filtration experiments (**Supp. Fig. S2**). The overall architecture of the SSA dimer consists of a compact assembly with dimensions of 60 Å x 30 Å x 25 Å that buries a flat surface on each β-trefoil monomer. The dimer interface is formed by the tight association of the β4 and β5 strands from each monomer with the participation of two

9

peripheral loop regions (β5- β6 and β7- β8) (**Fig. 2**). In turn, the dimeric assembly of SSA differs from the heterodimeric assembly of the lectin domain of ricins [26]. In fact, a structural overlay between SSA and the ricin B chain from the castor bean plant [27] reveals that a nine residues insertion in the β4-β5 loop of SSA prevents formation of a dimeric assembly as seen in the ricin R-type lectin domain. In SSA, the calculated buried surface area to a 1.4 Å probe radius encompasses ~950 A^2 on each subunit and involves 16 residues that are dominantly apolar. The topology of the β-strands recruited in this novel dimer interface is conserved in sequence family members of SSA (see below) and residues involved in this assembly are well conserved, suggesting that this type of dimeric association should be preserved in SSA homologs.

Diversity in the oligomeric assembly of various lectins that share a similar fold has been already described and may account for their multivalent carbohydrate-binding specificities [28]. In the SSA dimer, the two carbohydrate binding sites, which are located on the same face of the dimer, are separated by 38 Å, a distance preventing the formation of cross-linking interactions between different antennae of a single N-glycan. Although the nature and biological signification of cross-linking interactions remains to be investigated, the remote location of the two carbohydrate-binding sites in the SSA dimer may favor interaction with distinct glycan chains harboring terminal galactosyl residues.

BLAST searches using the SSA sequence as a template identified close homologs in the genome of the three plant pathogenic fungi, *Botryotinia fuckeliana* (*Botrytis cinerea*), *Pyrenophora tritici-repentis* and *Cochliobolus heterostrophus,* the former two belonging to the same *Pleosporaceae* family. These three putative SSA homologs share approximately 77%, 27% and 28% sequence identity, respectively, and may belong to the same family of fungal lectins, as key residues at the carbohydrate binding site are conserved in site α along with most of side chains involved in the dimeric assembly (**Fig. 3**).

In summary, the crystal structure of SSA reveals that the specific binding to terminal Gal/GalNAc sugars with 1,3 linkages is achieved by both direct and water-mediated hydrogen bonds to the O4 and O3 hydroxyls of Gal at the non-reducing end. The novel dimeric assembly demonstrates that different quaternary structures exist within the R-type lectin family, resulting from a different association of structurally similar domains. Whereas other types of quaternary arrangements may occur within the extended group of fungal lectins, the restricted number of functional binding sites per β-trefoil domain in SSA could dictate the type of quaternary structures within this new family of fungal lectins and affect the ability to establish multiple interactions and cross-link glycan-receptors. By analogy with the dual storage-defense role attributed to a large group of abundant plant lectins, the *Sclerotiniaceae* lectins might also combine a likely function as storage protein in the development and morphogenesis of the fungus, with a possible defense role against predating organisms. Further studies to elucidate/investigate the biological role of this emerging class of fungal lectins are required.

Materials and Methods

Materials

SSA was isolated from mature *Sclerotina sclerotiorum* sclerotes and purified as described previously [11].

Crystallization and data collection

For crystallization, ammonium sulfate precipitated SSA was extensively dialyzed against 10 mM Hepes pH 7.5 and 150 mM NaCl, and concentrated to 7.3 mg/ml. Crystallization experiments were performed by vapour diffusion with Greiner plates (Greiner-BioOne) using a Freedom (Tecan) and Honeybee (Cartesian) robot. Initial crystals of apo SSA were obtained

at 20° C from a condition containing 2.0 M sodium chloride and 0.1 M sodium acetate pH 4.6 of the Structure Screen 2 (Molecular Dimension Limited). After manual optimisation in Limbro plates, larger crystals appeared within 10 days from typically 1 µl of protein solution and 0,5 µl of reservoir containing 2.1 M sodium chloride and 0.1 M sodium acetate pH 4.6. The SSA-Galβ1,3GalNAc complex was made by mixing the protein solution with 2.5 mM of Galβ1,3GalNAc followed by preincubation overnight at 4°C prior to crystallization experiments. Crystals of the complex appeared at 20°C from a condition containing 1.26 M ammonium sulfate, 0.2 M lithium sulfate, 0.1 M Tris pH 8.5. Crystals selected for data collection were rapidly transferred in the reservoir solution supplemented with 25% glycerol (apo) or 30% PEG 400 (complex), flash cooled at 100 K in a nitrogen gas stream and stored in liquid nitrogen. Data were collected on the ESRF beamlines ID23-1 and ID23-2 (Grenoble, France). Oscillation images were integrated with Mosflm [29] and scaled and merged with SCALA [30].

Structure determination and refinement

Attempts to solve the structure using a Se-labelled D-Gal-β-SePh derivative were unsuccessful. Initial phases were obtained by molecular replacement with the PHASER program [31] using as search model a structural template generated with the protein homology recognition engine Phyre [32] which uses profile-profile matching algorithms for the detection of homologs of known three-dimensional structure, the so-called template-based homology modelling or fold-recognition. The model generated by the Phyre server was based on the crystal structure of *Clostridium botulinum* non-toxin haemagglutinin HA33/A [13] (pdb accession number 1YBI) and led to an unambiguous solution in PHASER with a Z-score value of 51 in the 15 Å to 2.8 Å resolution range. The resulting model, comprising four molecules of SSA in the asymmetric unit, was rebuilt in an automatic fashion with the

program ARP/wARP [33]. Final refinement was then performed with REFMAC [34] using data up to 1.6 Å resolution and the resulting SigmaA-weighted 2mFo-DFc and mFo-DFc electron density maps were used to correct the model with the graphics Emsley 2004}. The structure of the SSA disaccharide complex was solved by molecular replacement with Phaser [31], using apo SSA as a search model. The structure of the complex was subsequently refined and manually adjusted with the programs REFMAC [34] and COOT [35], respectively.

The final apo model comprises residues Gly2 to Lys 153 for the four SSA molecules and two glycerol molecules arising form the cryo buffer; that of the disaccharide complex comprises residues Gly2 to Lys153 for a single SSA molecule, a Galβ1,3GalNAc disaccharide, four sulfate ions arising from the crystallization buffer and two small PEG molecules arising form the cryo buffer. In the two structures, the backbone region of residues Gly58 to Ser60 adopts two alternate conformations while, in the apo structure, the surface loop region Thr103-Glu108 is disordered. The r.m.s.d. between the apo and sugar-bound form is 0.64 Å for 144 Cα atoms. In the apo form, the β2-β3 surface loop region containing the residue pair Trp24-Glu25 at site α undergoes large conformational changes in one out of four molecules, but these movements are mediated by the crystal packing environment. The stereochemistry of the models was analyzed with MolProbity [36]; no residues were found in the disallowed regions of the Ramachandran plot. Data collection and refinement statistics are reported in Table 1. Structural sequence alignment was performed with the RAPIDO server [37]. Figures 1 to 2 were generated with PyMol [38]. Figure 3 was prepared with ESPript [39].

Glycan array screening

The glycan binding specificity of SSA was analyzed by the Consortium for Functional Glycomics using microarrays v3.1 [22]. Full details of protocols used and the ~380 glycans included in the array are presented on the Consortium's web site

http://www.functionalglycomics.org/static/consortium/resources/resourcecoreh8.shtml) and

described by Paulson and co-workers [22]. Lyophilized SSA was dissolved in PBS at 1 mg/ml

and labeled with tetrafluorophenyl (TFP)-Alexa Fluor 488 using the Invitrogen protein

labeling kit following the manufacturer's instructions. Labeled SSA was diluted to 0.25 µg/ml

in Tris buffered saline (20 mM Tris, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, pH 7.4)

containing 1% BSA and 0.05 % Tween 20. An aliquot (70 µl at 0.25 µg/ml) of the labeled

lectin solution was applied to separate microarray slides and incubated under a cover slip for

60 min in a dark, humidified chamber at room temperature. After incubation, the cover slips

were gently transferred to a solution of Tris-buffered saline containing 0.05% Tween 20 and

successively washed 4 times by gently dipping the slides in Tris-buffered saline containing

0.05% Tween 20 and deionized water. After the last wash the slides were spun in a slide

centrifuge for approximately 15 sec to dry and immediately scanned in a ProScanArray

MicroArray Scanner (PerkinElmer) using an excitation wavelength of 488 nm and ImaGene

software (BioDiscovery, Inc., El Segundo, CA) to quantify fluorescence. Data are expressed

as average RFU for binding each glycan calculated by averaging four values after removing

the highest and lowest values; STDEV is the standard deviation, SEM the standard error

measurement and CV is the coefficient of variation (S.D./mean) calculated as %. Analysis of

the carbohydrate specificity of SSA on the glycan array was repeated at lectin concentrations

of 0.50 and 1.0 µg/ml.

Molecular docking

α-GalNAc was selected as a candidate ligand for automated docking to SSA using Autodock

Vina [24]. SSA and the carbohydratre ligands were treated as a rigid protein and flexible

molecules, respectively. The grid of the docking simulation was defined by a 20 Å x 20 Å x

20 Å cube centered on the SSA α site. The docking simulation was performed using the

default parameters. For each ligand, the 9 top-ranked generations based on the predicted

binding affinity (kcal/mol) were analyzed.

Acknowledgements

Accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 2X2S and 2X2T.

References:

1. Wang H., Ng T. B. & Ooi V. E. C. (1998). Lectins from mushrooms. *Mycological Research* **102**, 897-906.

2. Imberty A., Mitchell E. P. & Wimmerová M. (2005). Structural basis of high-affinity glycan recognition by bacterial and fungal lectins. *Curr Opin Struct Biol* **15**, 525-534.

3. Goldstein, I J & Winter (2007). Mushroom lectins. In *Comprehensive glycoscience: From Chemistry to Systems Biology*, Elsevier Ltd, UK.

4. Kellens, JTC & Peumans, WJ (1990). Developmental accumulation of lectins in *Rhizoctonia solanii*: potential role as a storage protein. *J Gen Microbiol* **136**, 2489-2495.

5. Rosen S., Sjollema K., Veenhuis M. & Tunlid A. (1997). A cytoplasmic lectin produced by the fungus *Arthrobotrys oligospora* functions as a storage protein during saprophytic and parasitic growth. *Microbiology* **143**, 2593.

6. Cooper D. N. W., Boulianne R. P., Charlton S., Farrell E. M., Sucher A. & Lu B. C. (1997). Fungal galectins, sequence and specificity of two isolectins from *Coprinus cinereus*. *Journal of Biological Chemistry* **272**, 1514.

7. Yatohgo, T, Nakata, M, Tsumuraya, Y, Hashimoto, Y & Yamamoto, S (1988). Purification and properties of a lectin from the fruiting bodies of *Flammulina velutipes*. *Agric. Biol. Chem* **52**, 1485-1494.

8. Singh R. S., Bhari R. & Kaur H. P. (2010). Mushroom lectins: Current status and future perspectives. *Crit Rev Biotechnol* .

9. Fukazawa Y. & Kagaya K. (1997). Molecular bases of adhesion of *Candida albicans*. *J Med Vet Mycol* **35**, 87-99.

10. Candy L., Van Damme E. J., Peumans W. J., Menu-Bouaouiche L., Erard M. & Rougé P. (2003). Structural and functional characterization of the GalNAc/Gal-specific lectin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* (Lib.) de Bary. *Biochem Biophys Res Commun* **308**, 396-402.

11. Kellens, JTC & Peumans, WT (1992). Lectins in different members of the Sclerotiniaceae. *Mycol. Res.* **96**, 495-502.

12. Van Damme E. J., Nakamura-Tsuruta S., Hirabayashi J., Rougé P. & Peumans W. J. (2007). The Sclerotinia sclerotiorum agglutinin represents a novel family of fungal lectins remotely related to the *Clostridium botulinum* non-toxin haemagglutinin HA33/A. *Glycoconj J* **24**, 143-156.

13. Arndt J. W., Gu J., Jaroszewski L., Schwarzenbacher R., Hanson M. A., Lebeda F. J. & Stevens R. C. (2005). The structure of the neurotoxin-associated protein HA33/A from *Clostridium botulinum* suggests a reoccurring beta-trefoil fold in the progenitor toxin complex. *J Mol Biol* **346**, 1083-1093.

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

15. Winter H. C., Mostafapour K. & Goldstein I. J. (2002). The mushroom *Marasmius oreades* lectin is a blood group type B agglutinin that recognizes the Galalpha 1,3Gal and Galalpha 1,3Galbeta 1,4GlcNAc porcine xenotransplantation epitopes with high affinity. *J Biol Chem* **277**, 14996-15001.

16. Mo H., Winter H. C. & Goldstein I. J. (2000). Purification and characterization of a Neu5Acalpha2-6Galbeta1-4Glc/GlcNAc-specific lectin from the fruiting body of the polypore mushroom *Polyporus squamosus*. *J Biol Chem* **275**, 10623-10629.

17. Tateno H. & Goldstein I. J. (2003). Molecular cloning, expression, and characterization of novel hemolytic lectins from the mushroom *Laetiporus sulphureus*, which show homology to bacterial toxins. *J Biol Chem* **278**, 40455-40463.

18. Grahn E., Askarieh G., Holmner A., Tateno H., Winter H. C., Goldstein I. J. & Krengel U.

(2007). Crystal structure of the *Marasmius oreades* mushroom lectin in complex with a xenotransplantation epitope. *J Mol Biol* **369**, 710-721.

19. Nakamura T., Tonozuka T., Ide A., Yuzawa T., Oguma K. & Nishikawa A. (2008). Sugar-binding sites of the HA1 subcomponent of *Clostridium botulinum* type C progenitor toxin. *J Mol Biol* **376**, 854-867.

20. Hasegawa K., Watanabe T., Suzuki T., Yamano A., Oikawa T., Sato Y., Kouguchi H., Yoneyama T., Niwa K., Ikeda T. & Ohyama T. (2007). A novel subunit structure of *Clostridium botulinum* serotype D toxin complex with three extended arms. *J Biol Chem* **282**, 24777-24783.

21. Murzin A. G., Lesk A. M. & Chothia C. (1992). beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors. *J Mol Biol* **223**, 531-543.

22. Blixt O., Head S., Mondala T., Scanlan C., Huflejt M. E., Alvarez R., Bryan M. C., Fazio F., Calarese D., Stevens J., Razi N., Stevens D. J., Skehel J. J., van Die I., Burton D. R., Wilson I. A., Cummings R., Bovin N., Wong C. H. & Paulson J. C. (2004). Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A* **101**, 17033-17038.

23. Weimar T., Bukowski R. & Young N. M. (2000). The conformation of the T-antigen disaccharide bound to *Maclura pomifera* agglutinin in aqueous solution. *J Biol Chem* **275**, 37006-37010.

24. Trott O. & Olson A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455-461.

25. Suzuki R., Kuno A., Hasegawa T., Hirabayashi J., Kasai K. I., Momma M. & Fujimoto Z. (2009). Sugar-complex structures of the C-half domain of the galactose-binding lectin EW29 from the earthworm *Lumbricus terrestris*. *Acta Crystallogr D Biol Crystallogr* **65**, 49-57.

26. Montfort W., Villafranca J. E., Monzingo A. F., Ernst S. R., Katzin B., Rutenber E., Xuong N. H., Hamlin R. & Robertus J. D. (1987). The three-dimensional structure of ricin at 2.8 A. *J Biol Chem* **262**, 5398-5403.

27. Rutenber E., Katzin B. J., Ernst S., Collins E. J., Mlsna D., Ready M. P. & Robertus J. D. (1991). Crystallographic refinement of ricin to 2.5 A. *Proteins* **10**, 240-250.

28. Sinha S., Gupta G., Vijayan M. & Surolia A. (2007). Subunit assembly of plant lectins. *Curr Opin Struct Biol* **17**, 498-505.

29. Leslie (1992). Recent changes to the MOSFLM package for processing film and image plate data . *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**.

30. CCP4 (1994). The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.

31. McCoy A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D: Biological Crystallography* **63**, 32-41.

32. Kelley L. A. & Sternberg M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371.

33. Perrakis A., Morris R. & Lamzin V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, 458-463.

34. Murshudov, G N, Vagin, A A & Dodson, E J (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255.

35. Emsley P. & Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

36. Davis I. W., Leaver-Fay A., Chen V. B., Block J. N., Kapral G. J., Wang X., Murray L. W., Arendall W. B., Snoeyink J., Richardson J. S. & Richardson D. C. (2007). MolProbity:

all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-W383.

37. Mosca R. & Schneider T. R. (2008). RAPIDO: a web server for the alignment of protein structures in the presence of conformational changes. *Nucleic Acids Research* **36**, W42.

38. Delano, W L (2002). The PyMOL molecular graphics system. In *The PyMOL Molecular Graphics System* (DeLano Scientific LLC, eds), Palo Alto, CA USA.

39. Gouet P., Courcelle E., Stuart D. I. & Métoz F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.

40. Grahn E. M., Winter H. C., Tateno H., Goldstein I. J. & Krengel U. (2009). Structural characterization of a lectin from the mushroom *Marasmius oreades* in complex with the blood group B trisaccharide and calcium. *J Mol Biol* **390**, 457-466.

Legends to figures

Figure 1. Overall view of the structure of SSA. (**A**) Overall structure of the SSA monomer showing the three subdomains, denoted α, β and γ, viewed in two orientations rotated by 90°, and colored green, cyan and orange, respectively. β-strands of a subunit are labeled β1 to β12. Dashed lines indicate the disordered β8-β9 loop region in the apo SSA structure. (**B**) Structure-based sequence alignment of the three individual subdomains of SSA; residues involved in carbohydrate binding at site α are shown by (★).

Figure 2. Carbohydrate binding site at site α. (**A**) Overall view of the dimer with bound Galβ1,3GalNAc in each subunit. (**B**) Close-up views of the carbohydrate binding site in (left) the apo form with bound glycerol (orange) and (middle) the complex with bound Galβ1,3GalNAc (orange), viewed in a similar orientation. The 1.95 Å resolution 2mFo-DFc electron density map (cyan) is contoured at 1 σ. Molecular surface (right) showing the shallow binding site. The position of the docked α-GalNAc (green) is displayed and overlaid with bound Galβ1,3GalNAc. Side chains that directly interact with the disaccharide are labeled. (**C**) Overall view (left) of the *Marasmius oreades* agglutinin (MOA) dimer with bound blood group B Galα1,3(Fucα1,2)Gal trisaccharide (green) [40] (accession code 3EF2). The N- and C-terminal domains from each subunit are colored in yellow and green, respectively. Close-up stereo view (right) of the overlay of the carbohydrate binding sites in SSA and MOA.. (D) Overall view (left) of the EW29 lectin with bound GalNAc (blue) [25] (accession code 2ZQO). Close-up stereo view (right) of the overlay of the carbohydrate binding sites, oriented as in C, in SSA and EW29. Side chains that directly interact with the bound carbohydrate in EW29 and MOA are labeled

19

Figure 3: Sequence alignment of SSA with homologs from *Botryotinia fuckeliana* (Bf-SSA, accession EDN28997.1), *Pyrenophora tritici-repentis* (Ptr-SSA, accession EDU42915.1) *Cochliobolus heterostrophus* (Ch-SSA, accession FK679805.1), as putative representative members of this new class of fungal lectins. Identical residues are indicated with a black background while similar residues are boxed. Amino acids involved in carbohydrate binding and dimer assembly of SSA are indicated by (★) and (▲), respectively. Secondary structure elements are indicated above the sequence.

Table 1. Data collection and refinement

	apo	Galβ1,3GalNAc
Data collection		
Beamline	ID23-1	ID23-2
Resolution (Å) [a]	55-1.60 (1.69-1.60)	60-1.97 (2.08-1.97)
Space group	$P6_1$	$P6_122$
Cell dimensions, a,b,c (Å)	127.32, 127.32, 86.65	51.46, 51.46, 295.99
No. of observations	460339	97652
No. of unique reflections	100976	17040
R_{sym} (%) [b]	0.072 (0.450)	0.140 (0.499)
$<((I)/\sigma(I))>$	16.8 (2.6)	9.0 (3.3)
Redundancy	4.6 (4.0)	5.7 (5.7)
Completeness (%)	96.2 (80.2)	97.88 (98.9)
B factor from Wilson plot ($Å^2$)	16.7	15.7
Refinement		
Resolution (Å)	110-1.60 (1.64 – 1.60)	49 – 1.97 (2.02 – 1.97)
R_{cryst} [c]	14.00 (21.2)	16.30 (21.40)
R_{free} (%) [d]	16.20 (21.0)	19.51 (24.80)
Number of reflections used in refinement	96905 (5289)	15811 (1127)
Number of reflections for R_{free}	4061 (237)	1228 (98)
Number of water molecules	710	169
Rms deviations [e]		
bond length (Å)	0.014	0.014
bond angles (°)	1.51	1.45
chiral volume ($Å^3$)	0.099	0.092
Mean B factors ($Å^2$)		
main / side chain	19.8 / 22.9	27.0 / 29. 4
bound disaccharide	-	27.9
solvent / other ligands	29.3 / 30.8	30.1 / 43.7
Rms deviations on B factors ($Å^2$)		
main chain	1.11	0.81
side chain	1.63	1.64
Ramachandran plot statistics [f]		
% of residues in favored regions	98.5	98.1
% of residues in allowed regions	100	99.4
% of outliers	0	0.6

[a] Values in parentheses are for the highest resolution shell

[b] $R_{sym}=\Sigma_{hkl}(\Sigma_i|I_{hkl}-\langle I_{hkl}\rangle|)/\Sigma_{hkl}|\langle I_{hkl}\rangle|$

[c] $R_{cryst}=\Sigma_{hkl}||Fo|- |Fc||/\Sigma_{hkl}|Fo|$

[d] R_{free} is calculated for randomly selected reflections excluded from refinement

[e] Root mean square deviation from ideal geometry.

[f] Ramachandran plot statistics have been calculated with the MolProbity server [36]

Table 2: Glycan array binding of 0.25 µg/ml SSA: *The 15 glycan structures are ranked from highest relative fluorescence units (RFU) to lowest.*

Glycan number	Glycan	Mean RFU[a]	STDEV[b]	SEM[c]	%CV[d]
9	GalNAcα–Sp8	23843	2458	1229	10
84	GalNAcα1-3Galβ–Sp8	23589	221	111	1
90	GalNAcβ1-4GlcNAcβ–Sp0	21092	348	174	2
85	GalNAcα1-4(Fucα1-2)Galβ1-4GlcNAcβ-Sp8	21025	1327	664	6
329	GalNAcβ1-3Galα1-4Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ-Sp0	19851	810	405	4
19	GalNAcβ–Sp8	19428	1347	673	7
88	GalNAcβ1-3Galα1-4Galβ1-4GlcNAcβ-Sp0	16809	1706	853	10
298	GalNAcα-Sp15	16684	634	317	4
91	GalNAcβ1-4GlcNAcβ–Sp8	14934	582	291	4
300	GalNAcβ1-3Galβ-Sp8	14658	2082	1041	14
346	Galβ1-4GlcNAcβ1-2Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12	11530	555	278	5
4	Galβ1-3GlcNAcβ1-2Manα1-3(Galβ1-3GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp19	7656	571	286	7
83	GalNAcα1-3GalNAcβ–Sp8	7575	54	27	1
55	Fucα1-2Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Sp9	7122	1485	742	21
59	Fucα1-2Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glcβ-Sp9	7074	913	456	13

[a] Relative fluorescence unit

[b] Standard deviation

[c] Standard error measurement

[d] Variation coefficient

Table 3: SSA-Galβ1,3GalNAc interactions[a]

Ligand atom[b]	Direct contact (Å)	Interaction partner
Gal O4	2.8	Asn22 OD1
	2.8	Glu25 N
Gal O3	2.9	Asn22 ND2
	3.0	Asn46 ND2
	2.7	Wat25
Gal 02	2.8	Wat45
Gal C6-O6	**3.3-3.5**	**Trp24 indole ring**
Gal C3-C4	**3.6-3.8**	**Tyr37 phenol ring**
GalNAC O4	2.7	Glu25 O
	2.8	Wat98
GalNAc O7	3.0	Wat46
GalNAc O1	2.7	Wat130
	2.7	Wat118

[a] Contact observed in a single subunit within a 3.2 Å and 4.0 Å distance for hydrogen bonds and hydrophobic carbon-carbon interactions, respectively
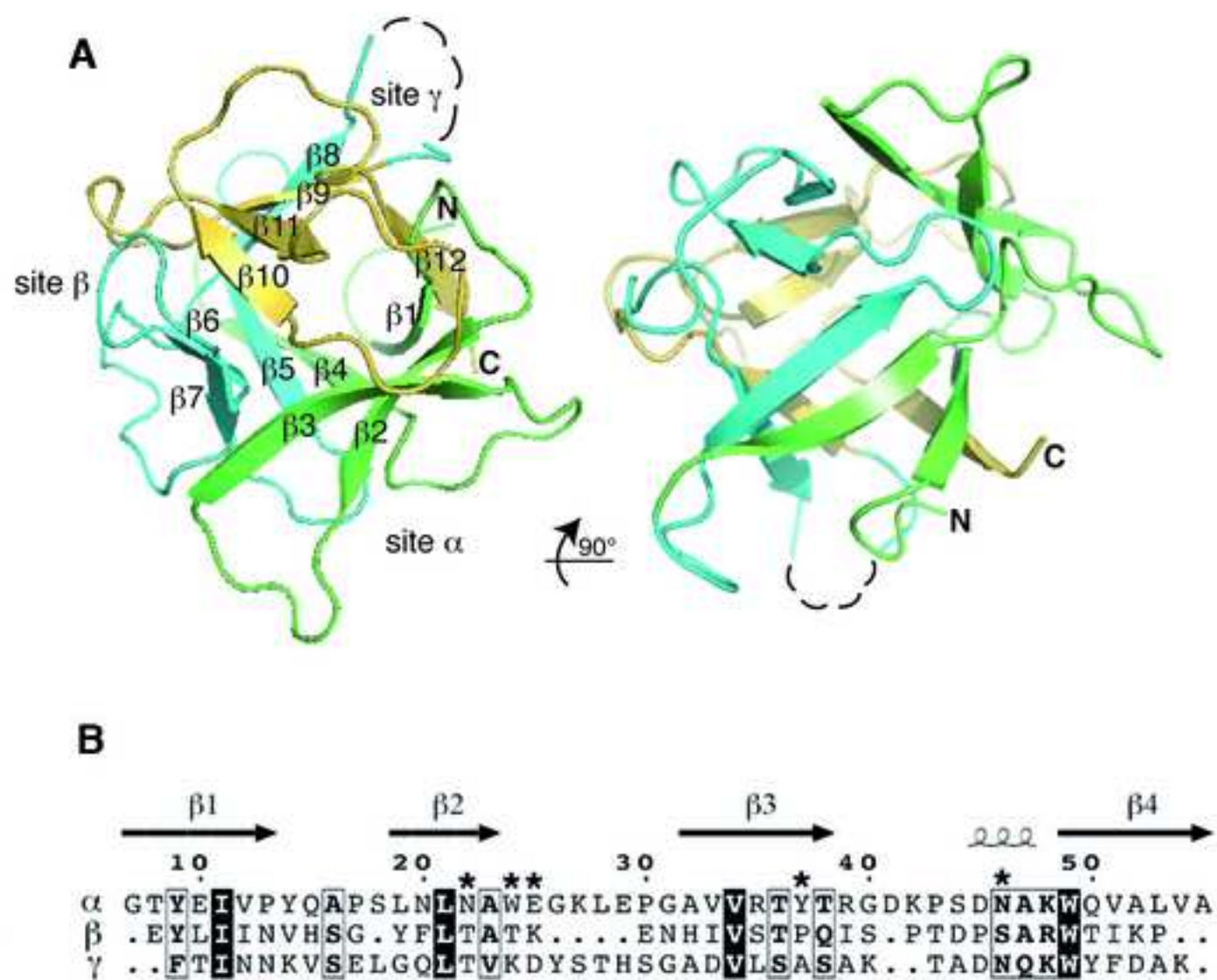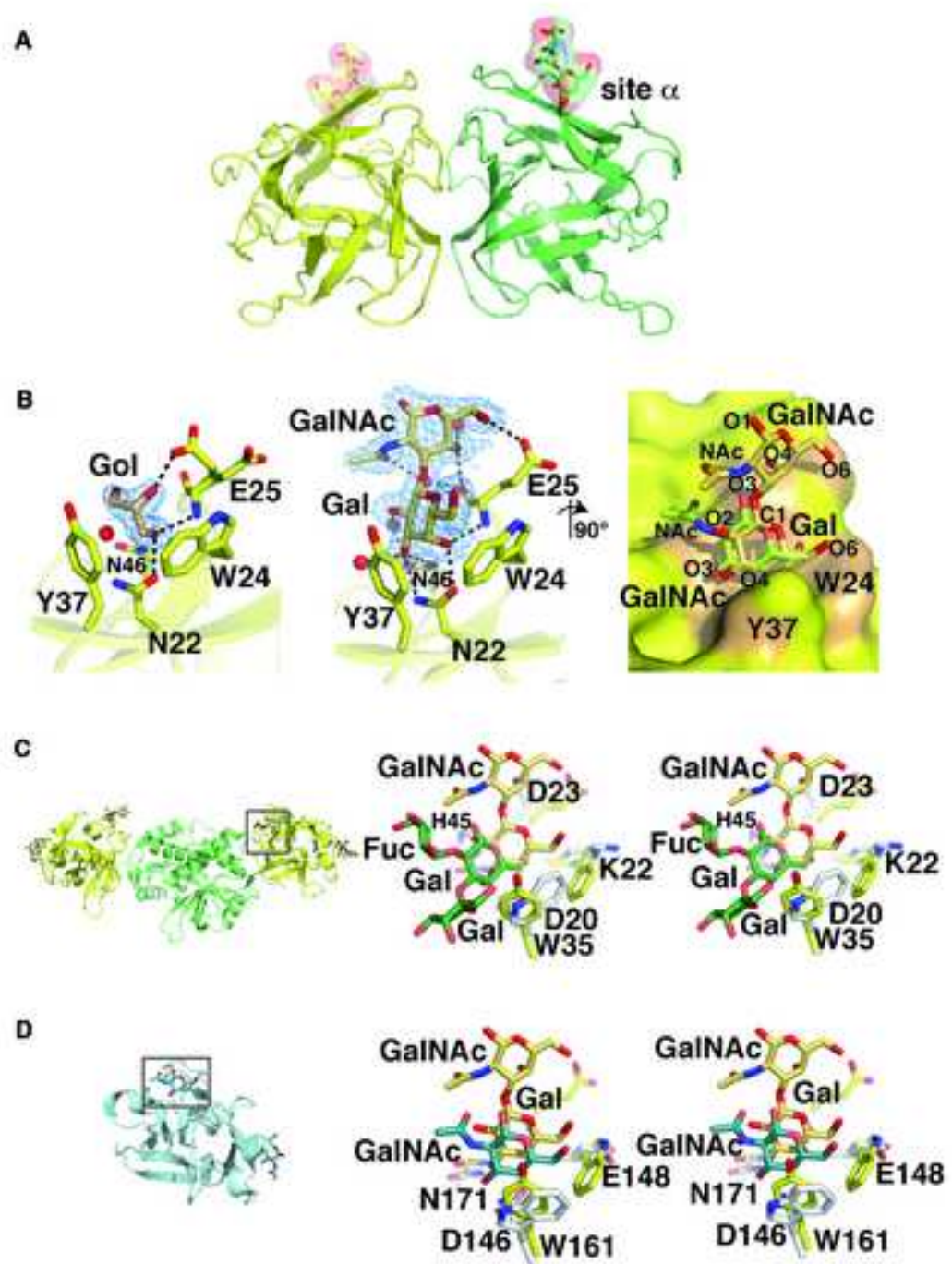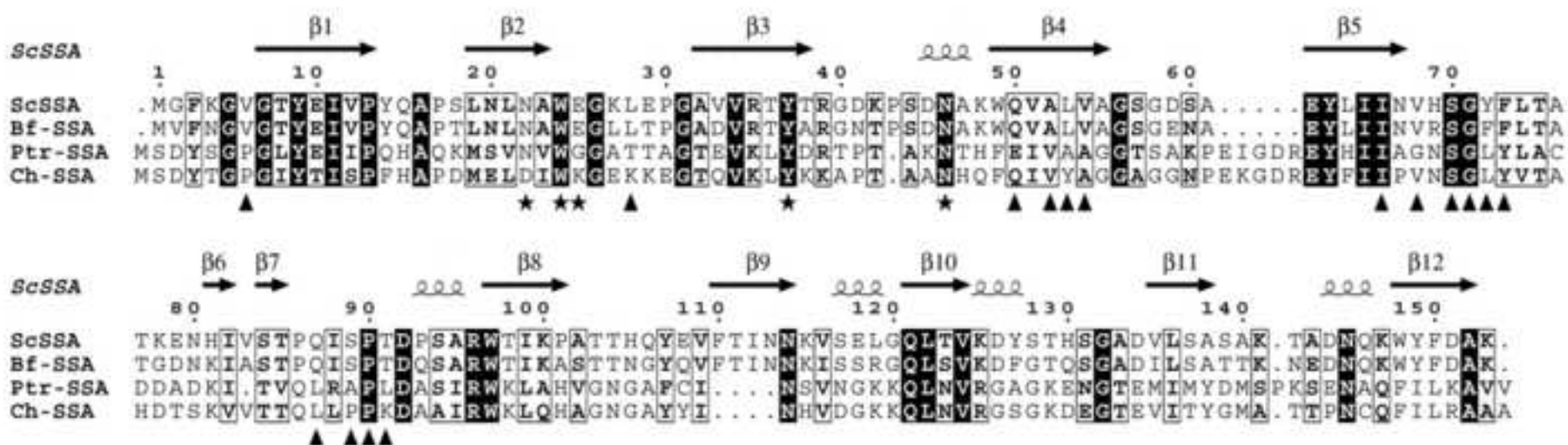[b] Hydrophobic interactions are indicated in bold.

Figure1

A

site γ

β8
β9
β11
β10
β12
β1
site β
β6
β5 β4
β7 C
β3 β2
site α
90°

N
C
N

B

β1   β2   β3   β4

```
          10              20              30              40              50
           *   **          *           *         *
α  GTYEIVPYQAPSLNLNAWEGKLEPGAVVRTYTRGDKPSDNAKWQVALVA
β  .EYLIINVHSG.YFLTATK....ENHIVSTPQIS.PTDPSARWTIKP..
γ  ..FTINNKVSELGQLTVKDYSTHSGADVLSASAK..TADNQKWYFDAK.
```

Fig. 1

Figure2

Fig. 2

Figure3

Fig. 3