Structural analysis and unique molecular recognition properties of a *Bauhinia forficata* lectin that inhibits cancer cell growth

Jacek Lubkowski[1,*], **Sarah V. Durbin**[2], **Mariana C.C. Silva**[3], **David Farnsworth**[2], **Jeffrey C. Gildersleeve**[2], **Maria Luiza V. Oliva**[3], and **Alexander Wlodawer**[1]

[1]Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

[2]Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

[3]Universidade Federal de São Paulo-Escola Paulista de Medicina, Rua Três de Maio, 100, 04044-020 São Paulo, SP, Brazil

Abstract

Lectins have been used at length for basic research and clinical applications. New insights into the molecular recognition properties enhance our basic understanding of carbohydrate-protein interactions and aid in the design/development of new lectins. In this study, we used a combination of cell based assays, glycan microarrays, and X-ray crystallography to evaluate the structure and function of the recombinant *Bauhinia forficata* lectin (BfL). The lectin was shown to be cytostatic for several cancer cell lines included in the NCI-60 panel; in particular, it inhibited growth of melanoma cancer cells (LOX IMVI) by over 95%. BfL is dimeric in solution and highly specific for binding of oligosaccharides and glycopeptides with terminal N-acetylgalactosamine (GalNAc). BfL was found to have especially strong binding (apparent $K_d = 0.5\text{-}1.0$ nM) to the tumor associated Tn antigen. High-resolution crystal structures were determined for the ligand-free lectin, as well as for its complexes with three Tn glycopeptides, globotetraose, and the blood group A antigen. Extensive analysis of the eight crystal structures and comparison to structures of related lectins revealed several unique features of GalNAc recognition. Of special note, the carboxylate group of Glu126, lining the glycan-binding pocket, forms H-bonds with both the N-acetyl of GalNAc and the peptide amido group of Tn antigens. Stabilization provided by Glu126 is described here for the first time for any GalNAc-specific lectin. Taken together, the results provide new insights into the molecular recognition of carbohydrates and provide a structural understanding that will enable rational engineering of BfL for a variety of applications.

Graphical abstract

A combination of cell based assays, glycan microarrays, and X-ray crystallography was used to evaluate the structure and function of the recombinant *Bauhinia forficata* lectin (BfL). The lectin was found to have especially strong binding to the tumor-associated Tn antigen and was shown to be cytostatic for several cancer cell lines included in the NCI-60 panel.
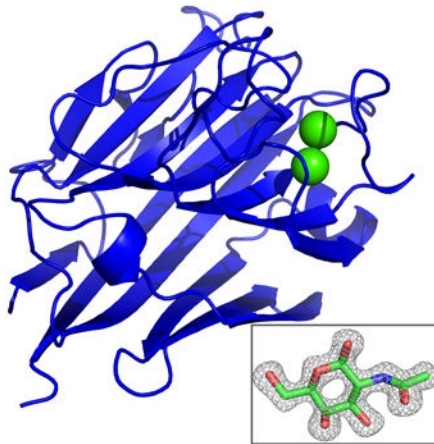
Keywords

Lectin; crystal structure; carbohydrate binding; Tn antigen; cancer cell growth Inhibition

Introduction

Lectins, non-enzymatic proteins that bind carbohydrates, have been the subject of extensive biochemical and structural studies due to the important roles they have in a broad range of biological processes, as well as their potential usage in therapeutic, diagnostic, and technological applications. In addition to their antiviral, antifungal, and antibacterial properties, there is considerable interest in their ability to selectively bind cancer cells and their capacity to prevent growth, proliferation, and migration of cancer cells [1-3]. Many types of cancers are characterized by a modified pattern of carbohydrates present on their cellular surfaces when compared to normal cells [4]. For example, the *O*-linked carbohydrate moieties observed in cancer cells represent products of incomplete glycosylation, terminated at or shortly after GalNAc, the core unit. Among the glycans most often associated with cancer cells are the Tn and T antigens, as well as their sialylated forms [4, 5]. Therefore, there have been major efforts to use lectins to bind the abnormally glycosylated cancer cells for diagnostic and therapeutic applications [1, 6]. Several lectins have been already shown as successful tools in diagnosis of specific cancers, i.e. galectin-4 targeting cancer antigen in breast cancer, *Maackia amurensis* agglutinin recognizing prostate-specific antigen, etc. [7-9]. Lectins have also been reported to display cancer-inhibitory properties in vitro [3, 10, 11]. For these reasons, identification and characterization of novel lectins isolated from a variety of sources has been a highly active area of research.

One lectin of particular interest is BfL, originally isolated from the seeds of *Bauhinia forficata*, a widely distributed tropical leguminous plant called "cow's foot" [12]. The polypeptide chain of BfL, consists of 233 amino acids and is N-glycosylated on asparagines 26 and 108, with carbohydrate content accounting for 8.4% of the mass of the lectin [12]. BfL was found to have a range of interesting biological activities. At concentrations down to 1 μg/ml, BfL agglutinates rabbit and human erythrocytes originating from all blood types [12]. The same authors also showed that, while non-cytotoxic to normal and several types of cancer cells, BfL inhibits the viability of the HER2-dependent MCF7 human breast cancer cell line (35% after 72 hours at 2.5 μM). It was suggested that the effect of BfL on MCF7 cells is linked to its interference with the expression of different subunits of integrin [13].

Although BfL has significant potential for diagnostic, biotechnology, and therapeutic applications, very little is known about the structure, function, and carbohydrate recognition properties of this lectin. For example, no monosaccharides were capable of inhibiting its agglutination activity and the carbohydrate specificity of BfL was unknown at the outset of this study. Structural and biological studies of BfL isolated from seeds of *B. forficata* have been hampered by the difficulty of obtaining highly purified, homogeneous material. Isolation from natural sources is complicated by limited access, seasonal availability, and laborious and expensive purification processes. Additionally, the final preparation is often a non-homogenous mixture of proteins, i.e. due to a heterologous glycosylation, and may not be suitable for certain studies (i.e. structural) or applications (therapeutic). Improved access to BfL would enable more extensive studies on its structure, function, and applications.

Expression of recombinant variants of lectins can facilitate structural and biological studies [14, 15]. In addition, recombinant methods provide rapid access to mutants and/or fusions. In studies presented here, we have developed a protocol for the expression and purification of a recombinant variant of BfL in sufficient quantities to determine its carbohydrate preference, three-dimensional structure, and its influence on the growth of cells in the NCI-60 panel of cancer cell lines [16]. We found BfL to be very specific for binding of N-acetylgalactosamine (GalNAc), both in a free form and as a terminal sugar in oligosaccharides (i.e. blood group A antigens) and glycopeptides (i.e. Tn antigens) [5]. Eight structures of BfL complexed with carbohydrates and glycopeptides elucidated the structural basis of its specificity. We have also found that the glycan-binding pocket of BfL is unique among related lectins and provided its detailed description.

Results

Production of recombinant BfL and formation of the complexes

A synthetic gene encoding the sequence of BfL was expressed in *E. coli*, resulting in insoluble protein that was subsequently subjected to refolding. In addition to the successful protocols described in Materials and Methods, we attempted to express BfL in various other bacterial systems, utilizing different expression vectors (including fusions), bacterial strains, or expression conditions. All attempts led invariably to insoluble products, raising a possibility that the lack of N-glycosylation in bacteria precludes productive folding. Therefore, we also attempted expressing BfL in the drosophila S2 cells, as a secreted protein. Although we identified small quantities of BfL in media, the expression yield was

very low, most likely due to the lectin inhibiting insect cell growth. Therefore, we moved forward with a bacterial expression system.

In bacteria, the insoluble recombinant BfL was expressed in large amounts, becoming the major component of inclusion bodies. In contrast to the ease of expression, the folding procedures proved challenging. Before refolding, inclusion bodies were thoroughly washed in a series of buffers and importantly, most of the divalent metal cations carried from the expression culture were removed at this stage. In our hands, only rapid dilution led to folding of BfL with a modest yield of 5-10%. This process was performed in a 5 liter fermentor vessel (BIOFLO 2000), equipped with an electronically-controller impeller (New Brunswick Sci.) and a P-1 peristaltic pump (GE Healthcare Life Sci). The refolding buffer contained a low concentration of Ca^{2+} ions, shown in the past to be important for retaining a structure stability and carbohydrate affinity of related lectins [17]. This buffer, however, did not contain Mn^{2+}, a cation often found near the carbohydrate-binding pocket in related lectins. After removing the misfolded fraction by filtering, the solution was concentrated and dialyzed against arginine-free buffer (arginine was found critical for successful folding of BfL). It is worth noting that our estimate of the folding yield is most likely too low, as the recombinant BfL readily binds to all media containing cellulose, such as concentration/filtering membranes or chromatography resins. We noticed that while the solubility of non-glycosylated BfL at room temperature is 8-10 mg/ml, it decreases at 5 °C. Because of the latter factor, the final concentration steps were performed at room temperature. Since we did not make particular effort to optimize the folding conditions, it is quite likely that yield of folding can be substantially improved. Nevertheless, the final preparation of recombinant BfL proved to be very pure and homogenous, and the protocol used by us was reproducible.

BfL forms dimers in solution

A majority of the naturally occurring legume lectins form oligomers, usually dimers or tetramers [18]. Oligomerization is intimately related to their agglutinating properties. As mentioned above, the recombinant BfL migrated anomalously on a gel filtration column, eluting in much larger than predicted volume. For instance, a BfL-containing peak eluted from the column concurrently with myoglobin (17,000 Da), suggesting that either BfL does not oligomerize or that it interacts with the chromatographic resin, a phenomenon previously described for other lectins [19]. In order to determine the oligomeric state of the recombinant BfL in solution we utilized analytical centrifugation. Three solutions of this lectin at different concentrations were subjected to sedimentation at 15 °C with simultaneous measurement of the UV absorbance at 280 nm. Using the program Sedfit [20], the experimental UV scans were modeled according to several scenarios that assumed the presence of different species in solution, i.e. monomers, dimers, tetramers, a monomer-dimer equilibrium, etc. An analysis clearly indicated that the best fit to the ultracentrifugation data was for a model consisting of BfL dimers. The values of root-mean-square deviations (r.m.s.d.) between the measured and predicted absorption values were several times lower for the model with BfL dimers as compared to either monomers or tetramers. Since these were sedimentation velocity experiments, we did not measure the dimerization constant for BfL. However, there was no indication of the presence of BfL monomers when the

experimental data were modeled according to the scenario of monomer-dimer equilibrium, suggesting that BfL has a high propensity for dimerization.

Growth inhibition of human cancer cells

Previously, cytotoxic properties of natural BfL were evaluated for three cancer cell lines, MDA-MB-231, MCF 10A, and MCF7 [13]. In those studies, natural BfL was reported to have no effect on the viability of the MDA-MB-231 and MCF 10A cells at the concentration range of 2.5-15 μM; however, the viability of the MCF7 cells was inhibited by 35-40%. To confirm that recombinant BfL was biologically active and to simultaneously evaluate potential activities against a wide range of other cell lines, we utilized the NCI-60 Human Tumor Cell Lines Screen. This screen uses the sulforhodamine B (SRB) colorimetric assay to assess the toxicity of a tested agent on a panel of 60 different human cancer cells lines. Since the SRB assay determines the total protein content in non-lysed cells present in a culture, it allows one to estimate both the cytostatic and cytotoxic effects resulting from exposure to a particular agent. Two of the previously studied cell lines (MCF7, MDA-MB-231) were present in the NCI-60 screen.

Recombinant BfL was evaluated in the NCI-60 screen at a concentration of 1.85 μM. An analysis of data presented in **Figure 1** shows that the inhibitory values for MCF7 and MDA-MB-231 were very similar to the results reported previously for natural BfL, confirming that recombinant BfL is biologically active. In addition, the NCI-60 provided a broad and varied profile of activity for many other cell lines. About 50% of the cell lines were not inhibited by the presence of this lectin (**Figure 1**). A modest effect, representing growth inhibition between 10 and 50%, was observed for 22 cell lines, and a strong effect (>50% inhibition) was observed for five cell lines of different cancers. The strongest inhibition, ~95%, was recorded for the melanoma LOX IMVI cell line. No cytotoxic effects were observed.

Evaluation of carbohydrate binding properties

The carbohydrate binding properties of BfL were evaluated on a glyco-antigen microarray containing approximately 500 array components, including a diverse assortment of *N*-linked glycans, *O*-linked glycans, glycolipid glycans, glycopeptides, and glycoproteins. Binding was assessed at a series of concentrations and apparent dissociation constants (K_d) were determined as described previously [21]. Apparent K_d values signify overall avidity of the multivalent interaction between the lectin and the carbohydrates. Since the lectin can form multivalent complexes with glycans on the surface of a cell, multivalent avidities are more relevant to cell binding than monovalent affinities. Representative apparent K_d values can be found in **Table 1**. The full array data and complete list of apparent K_d values can be found in the Supplementary Material (see **Supplementary Table 1**).

Overall, BfL selectively bound glycans with a GalNAc residue at the non-reducing termini (**Figure 2**). The strongest binding was observed to the Tn antigen. The Tn antigen is a tumor-associated carbohydrate antigen composed of a GalNAc α-linked to either serine or threonine of a polypeptide chain. The array contains a large variety of Tn glycopeptides (55 array components), including serine and threonine forms, variations in peptide sequence around the Tn residue, variations in peptide length, variations in epitope density, attachment

at either the N- or C-terminal end of the peptide, and glycopeptides containing clusters of 2 or 3 Tn residues. BfL bound all forms of the Tn antigen very tightly with apparent K_d values ranging from 0.4 – 1.2 nM. No binding was observed to the corresponding non-glycosylated peptides or the analogous GlcNAc-α, Fuc-α, Man-α, or Gal-α containing glycopeptides even at the highest concentration tested. In most cases, modification of the GalNAc eliminated binding. For example, no binding was observed to core 2, core 3, core 4, or TF/ core 1-containing glycopeptides. One exception was F1α (Galβ1-4GlcNAcβ1-6GalNAcα1-Thr), which is modified at the 6 position. Core 5 (GalNAcα1-3GalNAcα1-Ser) was also a good binder. Since the modification in this case is another GalNAc residue, the key recognition motif is still present. Strong binding to the Tn antigen and preference for an unmodified GalNAc was confirmed with two natural glycoproteins. BfL bound tightly to asialo ovine submaxillary mucin (aOSM), a glycoprotein with >95% of its glycan composed of the Tn antigen. In contrast to aOSM, binding to OSM (~90% sialyl Tn (STn) and 5% Tn), was approximately 15 fold worse than aOSM. Preferential binding to aOSM over OSM indicates little or no recognition of STn. Similarly, BfL bound tightly to asialo-bovine submaxillary mucin (aBSM; ~65% Tn) but had significantly reduced binding to BSM (55% STn, 9% Tn).

Beyond the glycopeptides, BfL demonstrated strong binding to a variety of other GalNAc-α and GalNAc-β terminal oligosaccharides, including the blood group A trisaccharide [BG-A; GalNAcα1-3(Fucα1-2)Gal], GalNAcα1-3Gal, GalNAcα1-6Gal, Forssman di- and tetrasaccharides (GalNAcα1-3GalNAcβ1-3Galα1-4Galβ), Lac-di-NAc (LDN; GalNAcβ1-4GlcNAcβ), Gb4 (GalNAcβ1-3Galα1-4Galβ1-4GlcNAcβ), iGb4 (GalNAcβ1-3Galα1-3Galβ), and GA2 (GalNAcβ1-4Galβ). The terminal GalNAc residue appears to be essential for binding, as structurally-related glycans showed no binding (for examples, compare BG-A, BG-B, and BG-H; also compare Gb3, Gb4, and Gb5; see **Figure 2**). Interestingly, BfL bound well to the BG-A trisaccharide when attached to type 1 and type 2 carrier glycans (producing BG-A1 and BG-A2); however, certain other carrier chains, such as type 3 (BG-A3) and Lewis B (ALeB), resulted in >100 fold reduced binding. Therefore, structural features of the glycan beyond the GalNAc residue can influence recognition by BfL. In contrast to the Tn antigen, binding to BG-A was highly dependent on glycan density; little or no binding was observed at low density (estimated app K_d >1 μM).

Crystal structures of BfL and of its complexes with ligands

Eight crystal structures of BfL, in the carbohydrate-free form and bound to a number of ligands, were determined as part of this work. The ligands included three carbohydrates [α/β-GalNAc, P-antigen (globotetraose, Gb4), and the terminal oligosaccharide from the blood group A antigen, here denoted BG-A], as well as three Tn peptides in which GalNAc was linked to either a serine or a threonine (**Figure 3**). Of the seven high-resolution structures, four were determined in the isomorphous orthorhombic space group $P2_12_12_1$, two were isomorphous in another cell in the same space group, and one belonged to the monoclinic space group $P2_1$ (**Table 2**). All these crystals contained a BfL dimer, representing a biological unit, in the asymmetric unit and diffracted at high-to-atomic resolution (1.71 to 1.17 Å). Crystals of the complex between BfL and Tn glycopeptide Tn-200 also grew in the

space group $P2_12_12_1$ but with different unit cell parameters, contained five dimers in the asymmetric unit, and diffracted to only 2.7 Å resolution.

In common with a large number of related lectins, such as concanavalin A, the prototypical member of this family [22], a monomer of BfL contains two large antiparallel β-sheets, one six-stranded and the other one seven-stranded (**Figure 4**). The dimer is formed by the interactions of the six-stranded sheets from two monomers, arranged in an approximately perpendicular orientation to each other. There are no α-helices present, although a number of helical turns are found in the chain. Two metal-binding sites are present in each monomer (see below). The atomic coordinates of BfL in the eight structures presented here are very similar. For example, the complex with Tn-168 was determined in the orthorhombic crystals at 1.17 Å resolution, and in the monoclinic crystals at 1.35 Å. A superposition of monomers A from both structures, performed with the program SSM [23], yields an r.m.s.d. of 0.26 Å for all 229 pairs of the modeled Cα-atoms. The maximum shift of ~1 Å, observed for Cα of Gln11 is due to the differences in the crystal contacts of this part of the chain. The conformation of the ligand is also almost identical for the visible parts that could be modeled, although the terminal chains of the ligand were visible only in the monoclinic crystals, despite slightly lower resolution of that structure. Superposition of the BfL dimers in the same two complexes yields an r.m.s.d. of 0.50 Å, again attributed primarily to different crystal contacts in the two unit cells. Superposition of molecules A and B in the orthorhombic crystals of the Tn-168 complex yields an r.m.s.d. of 0.28 Å, showing that the agreement between two molecules in the same structure is very comparable to that of the molecules in two different crystal forms.

Superposition of the structure of the Tn-168 complex with the complex of BfL with Gb4 (1.43 Å resolution), both determined using isomorphous orthorhombic crystals, yields an r.m.s.d. of 0.11 Å for the superposition of either monomers A or the complete dimers. This result further supports the notion that the small flexibility of the dimeric structure observed when comparing the structures from non-isomorphous crystals is most likely due to effects of crystal packing. Superposition of other high-resolution structures yielded the r.m.s.d. values ranging from 0.09 Å to 0.22 Å. Not surprisingly, the largest r.m.s.d. value (0.34 Å) was found when the monomers A from Tn-168 and Tn-200 complexes were compared, and it may be related to the considerably lower resolution of the latter structure (2.7 Å). Nevertheless, the two sets of coordinates can be considered to be the same within the errors in their determination.

Metal-binding sites in BfL

In common with all other related lectins, each BfL molecule contains two very highly-conserved metal ions binding sites (**Figure 5**). The site here denoted Ca1 is occupied by a Mn^{2+} ion in all structurally related lectins, whereas Ca^{2+} is found in the second site, Ca2. However, in all BfL structures included in this study, both Ca1 and Ca2 sites are occupied by Ca^{2+} ions.

Since, as discussed later, the coordination environment of the Ca1 site in the recombinant BfL is non-standard for the Ca^{2+} ion, we made an extensive effort to verify the identities of metals found in structures described here. It is worth noting that none of the solutions used

throughout preparation of BfL samples contained Mn^{2+}. Due to thorough washing of inclusion bodies with buffers containing detergents, chaotropes, and chelating agents, a likelihood of carrying this cation from the cell culture media is quite low. To obtain further confirmation, we performed scans of several BfL crystals with the X-rays at the wavelengths spanning the Mn Kα absorption edge (~1.75 Å). These scans revealed no significant absorption that would have indicated the presence of Mn^{2+}. Subsequently, we collected complete X-ray diffraction data sets, tuning the wavelength to values nearing both sides of the Mn absorption edge and calculated anomalous electron density maps. Again, no presence of Mn^{2+} was detected and, except for very weak peaks present at both metal-binding sites, anomalous maps in those areas were quite featureless. Since the calcium absorption edge is far beyond the limit of the wavelengths accessible on the synchrotron beamline used for these experiments, weak anomalous peaks, similar at both sites, suggest the presence of Ca^{2+} in both places. Finally, refining the ion in site Ca1 as Mn^{2+} resulted in a large negative peak in the F_o-F_c electron density map and a higher B-factor than for the surrounding atoms, again indicating that no manganese was present. The average, maximum, and minimum coordinate bond lengths to both ions, together with their standard deviations, are listed in **Table 3**.

Distances between oxygen atoms from the carboxylate group of Asp120 and the Ca^{2+} ion found in site Ca2 are near-perfect for a bidentate ligand (2.44 Å and 2.48 Å, when averaged over all structures). Other coordinating atoms include the main-chain carbonyl oxygen of Trp122, Asn124(OD1), as well as two water molecules, Wat12 and Wat13. The distances and angles for the coordination bonds (**Table 3**) are typical for the Ca^{2+} ions [24] and the atomic displacement factors of the calcium ion are very similar to those of the atoms in its coordination sphere, indicating full occupancy of the site.

The Ca^{2+} ion occupying the Ca1 site is coordinated by three carboxylate oxygens (Glu118(OE2), Glu129(OE1), and Asp120(OD2)), two water molecules (Wat10 and Wat11), and by the NE2 atom of His134. Since in the structures of all other related lectins this site is occupied by Mn^{2+}, it is very likely that natural BfL also contains manganese in this site. The coordination distances in this site are shorter than in the case of site Ca2, most likely due to incomplete adjustment of the neighboring residues in BfL to the presence of a larger cation, Ca^{2+}, rather than Mn^{2+}. However, all distances are still longer than 2.15 Å, usually observed for the O…Mn^{2+} coordination. We postulate that the presence of only calcium in the structures discussed here is a result of the purification and refolding procedures. Furthermore, replacement of Mn^{2+} in natural BfL by Ca^{2+} in the recombinant variant does not introduce any meaningful structural/conformations changes both in the metal-binding site(s) and in the carbohydrate-binding pocket. An example of a similar "forced placement" of Mn^{2+} into Ca^{2+} sites was reported in the past for calmodulin [25], and in that case the replacement also did not induce significant coordinate changes of residues surrounding the metal-binding site.

Whereas the metal ions and their surroundings are well-defined in all high-resolution structures, it is worth mentioning that two different rotamers of Glu129 can be seen in the highest-resolution structure (orthorhombic crystals of the complex with Tn-168). However,

the carboxylate oxygen atoms of both rotamers are found in almost exactly the same locations.

The carbohydrate-binding pocket in BfL and the structural basis of specificity

The carbohydrate-binding pocket in BfL is formed by four discontinuous segments: Ala77-Asp78, Tyr94-Gly95-Gly96-Tyr97, Trp122-Val123-Asn124-Thr125-Glu126-Trp127, and Thr210-Gly211-Gln212. Interactions between BfL and carbohydrate ligands can be divided into primary and accessory. The primary interactions, invariable across different ligands, consist of a set of hydrogen bonds between the lectin and the GalNAc residue, which can be either an isolated molecule or the terminal residue, attached to other carbohydrates (i.e. blood group antigens), or to amino acids as in Tn peptides. These interactions provide gross contribution to the affinity towards a ligand as well as determine the specificity of BfL. Accessory interactions are contributed by parts of the ligand molecule other than the GalNAc residue.

The set of seven primary interactions is defined by invariant hydrogen bonds depicted in **Figure 6**. Two H-bonds, Glu126(O2) … GalNAc(N2) and Gly96(N) … GalNAc(O7), involve the N-acetyl group of the ligand, and they define the specificity of BfL for GalNAc, as opposed to galactose or other non-N-acetylated carbohydrate residues. The side of the binding pocket adjacent to the N-acetyl group of a ligand is formed by the side chains of Tyr97 and Trp127, a hydrophobic section of the Glu126 side chain, and two glycine residues, Gly95 and Gly96. It is worth noting that aromatic residues are commonly found in the carbohydrate-binding pockets of lectins [26]. Collectively, these residues limit the size of the binding pocket, creating a site optimized for accommodation of the carbohydrate residue placed at the non-reducing termini. Additionally, the side chains of Tyr97, Trp127, and Glu126 provide a hydrophobic environment for the methyl group of the N-acetyl and contributing to its additional stabilization. For one of the primary interactions, the H-bond between Gln212(OE1) and GalNAc(O6), the distance and geometry are slightly less conserved across the structures determined in this work; therefore, its contribution to the binding may be somewhat less, compared to other primary interactions. Overall, very small variation in the lengths of primary interactions across different complexes studied here indicates a high structural conservation of the ligand-binding pocket.

The complex of BfL with GalNAc was obtained by soaking a commercial sample of the carbohydrate, containing a mixture of GalNAc anomers, into crystals of BfL. Unexpectedly, each of the two lectin molecules in the asymmetric unit associated with a different anomer. The monomer A of BfL binds the β-GalNAc, whereas the monomer B binds the α anomer. The quality of the electron density does not leave any doubt regarding the identity of a specific GalNAc anomer associated with each monomer of BfL (**Figure 7**). Two water molecules are found within the H-bonding distance of the α anomeric O1 in the molecule B, whereas the anomeric O1 of GalNAc bound to the molecule A is stabilized in its β configuration by a hydrogen bond to the main-chain carbonyl oxygen of Ile16 of a symmetry-related molecule. Whereas the β anomer of GalNAc is more stable than its α counterpart in solution [27], a single H-bond defines the preference towards the α isomer in complex with BfL. The energy difference between both complexes is quite small as a single

crystal contact is able to change the binding preference of BfL from α-GalNAc to β-GalNAc.

The complete blood group A antigen (BG-A) used in this study was defined well by its electron density in the regions of the carbohydrate-binding pockets of both molecules of BfL present in the asymmetric unit. In addition to the set of primary interactions, five accessory H-bonds are contributed by the fucose unit of the ligand. These interactions are made by the O2 oxygen with NE2 of Trp122 and both OE1 and OE2 of Glu126, by O3 with the carbonyl oxygen of Val123, and by the O4 with OG1 of Thr125. The atoms from the central α-D-galactose unit of BG-A do not interact with any BfL residues, except through indirect contacts, via water molecules.

Whereas only two terminal carbohydrate units, β-GalNAc and α-Gal, could be traced bound to monomer A in the complex with Gb4, the complete ligand (four carbohydrate units, β-GalNAc–α-Gal–β-Gal–Glc) could be modeled for a complex formed by monomer B. The accessory interactions include a direct hydrogen bond between the Gb4 and BfL is provided by the O2 atom of the glucose residue and the OG1 atom of Thr125, as well as water-mediated H-bonds between O3 of Glc and OE2 of Glu126, as well as an intramolecular H-bond to O2 of α-Gal.

In the case of Tn peptides, one accessory H-bond is common to all ligands studied here. This bond is formed between Glu126(OE2) and the peptide amide group of the residue (Ser or Thr) linked to the Tn group. In the case of the Tn-(Thr)-peptides, Tn-200 and Tn-207, this is the only direct interaction between the lectin and the peptidic part of the ligand. Two additional accessory H-bonds are present in the structures of the complexes with Tn-(Ser)-peptide, Tn-168. One of them is formed by Glu126(OE2) mentioned above and the hydroxyl group of a Ser adjacent to the glycosylated residue, while the other connects Thr125(OG1) and the peptide oxygen of the hexyl moiety of the ligand. These bonds are visible in the structures determined from both the orthorhombic and monoclinic crystals, although the hexyl moieties are better defined in the latter. To conclude, the accessory interactions in complexes with Tn peptides depend on the specific amino acid sequence of the ligand.

Comparisons with other related lectins

To identify both conserved and unique recognition motifs, we compared the various BfL structures with other related lectins. Whereas structural descriptions are currently available for a wide variety of lectins, only several examples represent proteins that share with BfL both common fold and carbohydrate specificity. Therefore, our structural comparisons focused on this subset of lectins.

Since the coordinates of BfL are nearly invariant in the eight structures presented is this work, in principle any model of BfL is suitable for comparisons with related lectins. We selected the structure of the complex with Tn-168, refined at the highest resolution, 1.17 Å. Using the Dali program [28], we searched for proteins that were structurally-related to BfL and got almost a thousand hits characterized by the Z-score larger than 7.0, indicative of a high structural similarity. The most similar to BfL (Z-score > 38.0) are two related lectins from *Griffonia simplicifolia*, the lectin IV (GS4 - PDB codes 1GSL, 1LED, 1LEC, and

1GNZ - [29]), and the lectin I, isoform GS-1-B$_4$ (GS1 - PDB code 1HQL - [30]). BfL shares 55.3% of amino acid sequence identity with GS4 and a superposition of monomers of these lectins resulted in the r.m.s.d. of 0.97 Å for 226 equivalent Cα-atoms. A similar comparison between BfL and GS1, sharing 55.3% sequence identity, resulted in the r.m.s.d. value 0.92 Å (224 equivalent Cα-atoms). Both BfL and GS4 are dimeric in solution. However, in contrast to structures of BfL, for which dimers (or their multiples) are present in the asymmetric units, in the case of GS4 a dimer is created by a crystallographic symmetry operation. Published definitions of biological assemblies for both *G. simplicifolia* lectins are rather confusing, with GS1 reported as tetramers [30] and GS4 reported as dimers [29], although in crystals both lectins form equivalent tetramers. No such tetramers are present in crystals of BfL. The most compact dimers, present in all three lectins, are created through the interaction of the large 6-stranded β sheets of each molecule, arranged almost perpendicularly to each other (**Figure 4**). A superposition of 455 equivalent Cα-atoms in the dimers of BfL and GS4 resulted in r.m.s.d. of 1.27 Å, indicating that the relative orientation of the molecules forming the dimer is virtually identical in these two lectins. Similar results are obtained when the BfL and GS1 dimers are compared.

Each of these three lectins contains several peptides in *cis*-conformation. The Ala77-Asp78 *cis* peptide, present in all three structures, stabilizes the orientation of the carboxylate group in Asp78 in a way that enables strong interaction with Wat13. This water molecule contributes to the coordination sphere of the metal cation occupying the Ca2 site. The other *cis* peptide present in BfL, Thr210-Gly211, has its equivalent in GS4 (although in that case it is located between valine and glycine), but the insertion into a loop found after glycine in GS1 results in the corresponding *trans* peptide. Whereas the C$_\alpha$-atom of Gly211 is located in close vicinity of Asp78(OD2), a specific role of this unusual conformation is unclear, especially since it is not shared by all related lectins.

Despite close structural relationship of BfL to GS1 or GS4, the substrate specificity of BfL is quite different. Both *G. simplicifolia* lectins show a strong preference for binding to the D-galactose. As discussed earlier, BfL displays high specificity for the GalNAc residue at the non-reducing end with a slight preference for the α-anomer. One H-bonded interaction, between the nitrogen of the N-acetyl group and the OE2 carboxylate atom of Glu126 may be crucial for differentiating between Gal and GalNAc in these lectins. Neither GS1 nor GS4 contains a sufficiently long side chain in the position equivalent to Glu126 in BfL (Asn126 in GS1 and Asp137 in GS4). Furthermore, the side chain of Glu106 in GS1 extends over the site occupied by the N-acetyl group in GalNAc residue present in the complexes of BfL.

Comparisons of the Tn antigen binding sites

Several structures of lectins related to BfL and complexed with Tn antigens were described previously [31-35]. To the best of our knowledge, all structures of complexes between lectins and Tn peptides reported in the past, with a single exception of the soybean lectin SBA [33], describe ligands containing α-D-GalNAc bound to only a serine residue. The earliest such structure was determined at the resolution of 1.97 Å for the complex with the homotetrameric lectin VVLB4, directly purified from *Vicia villosa* (PDB code 1N47; [31]). VVLB4 shares 38.4% amino acid identity with BfL and the superposition of monomers A of

the two structures yields an r.m.s.d. of 1.34 Å for 219 C_α pairs. Another structure of a tetrameric lectin complexed with the same antigen was determined at 2.35 Å resolution using the protein WBA I isolated from *Psophocarpus tetragonolobus* (winged bean) (PDB code 2D3S - [32]). Superposition of molecules A of WBA I and BfL yields an r.m.s.d. of 1.36 Å for 210 equivalent C_α-atoms. Two structures of the *Vateirea macrocarpa* lectin (VML) complexed with the GalNAcα1-O-Ser have been previously determined. The PDB coordinate set 4U36 [34] represents the 1.4 Å structure of a complex with the natural lectin, whereas the coordinates 4XTP [35] were obtained with recombinant protein at 1.97 Å resolution. These two structures are virtually identical (r.m.s.d. 0.30 Å for 229 C_α-pairs). A superposition of the coordinates of BfL complexed with Tn-168 on the VML-Tn yields an r.m.s.d. of 1.34 Å for 216 equivalent C_α-atoms.

A more detailed discussion of the interactions between a Tn peptide and lectin is presented for the complex VVLB4-Tn (**Figure 8**); however, it is generally valid for the other complexes mentioned above. The conformation of the GalNAc molecule is almost identical in both complexes and the hydrogen bonds between the O3 and O4 oxygens of the ligand with the carboxylate of Asp78 in BfL have nearly the same lengths as their counterparts in VVLB4 (with Asp85). In both lectins, the O3 oxygen atom accepts a proton from the side chain amido group of an equivalent asparagine residue (124 in BfL and 129 in VVLB4), forming nearly identical hydrogen bonds. Similarly, O7 of the N-acetyl group in GalNAc is the acceptor in a H-bond with the main chain nitrogen atom of Gly96 (Gly103 in VVLB4). The same asparagine contributes its OD1 atom to the coordination sphere of the Ca2 binding site in both lectins. In contrast to BfL, there are interactions between the serine residue from the Tn peptide and VVLB4. In the complex with Tn-168, described here, the OE2 atom from Glu126 makes H-bonds with both the main chain amide nitrogen of the serine residue linked to the Tn group and the OG atom from adjacent serine residue.

The model of the complex between the soybean lectin SBA and Ala-Pro-Asp-Thr(GalNAc)-Arg, the latter representing an epitope derived from mucin MUC1, is currently the only available structure with the Tn group linked to a threonine residue and the polypeptide component represented by more than one residue (PDB code 4D69 - [33]). The structure of SBA was determined at the resolution of 2.7 Å and it consists of three tetramers of the lectin complexes present in the asymmetric unit. For a comparison with BfL we selected molecule F, which in the structure is associated with the most complete model of Tn peptide, Pro-Ala-Thr-Arg. SBA and BfL share 40.2% amino acid sequence identity, and a superposition of monomers yields a r.m.s.d. value of 1.42 Å for 214 equivalent C_α-atoms. The placement of GalNAc residue in the binding pockets of SBA and BfL, as well as networks of H-bonded interactions with lectin, are very similar. The only exception is the interaction between the O6 atom of the ligand, which forms a very short H-bond with Asp215 in SBA. This aspartate residue does not have an equivalent in BfL; in the latter, the O6 atom interacts with a water molecule. In contrast to BfL, where the amide nitrogen of the serine from Tn-168 forms H-bond with Glu126(OE2) of lectin, the Tn peptide in SBA makes no direct interactions with the protein. Interestingly, since Glu126 in BfL resides within a loop unique to BfL, the position of the carboxylate of Glu126 in the SBA complex is occupied by the aspartate side chain of the Tn peptide.

Additional stabilization of the complexes with Tn peptides in the structures described above is provided by hydrophobic interactions between the rings of GalNAc and an aromatic residue contributed by lectin, oriented in an almost parallel fashion. In the case of BfL, this contribution is most extensive and originates from Trp122, a residue which is equivalent to Tyr in VVLB4, and to Phe in the remaining lectins compared in this section.

Despite all similarities to the lectins described above, the carbohydrate-binding pocket in BfL is distinct. This uniqueness is due to two loops located on the opposite sides of the GalNAc residue ring. The first loop, between Thr121 and Arg133, is longer by two residues in BfL compared to the other lectins. At the tip of that loop is Glu126, which contributes two hydrogen bonds with a molecule of Tn antigen. The second loop, flanked by Thr210 and Glu214, is shorter in BfL by at least four residues than in any homologous lectin discussed here, and its interactions with the GalNAc residue are limited to just one weak hydrogen bond. However, reduction in the size of this loop results in better access to the glycan-binding cavity for larger ligands, such as Tn peptides.

A structural description of complexes with Tn antigen, which is in all cases represented by the GalNAc residue attached to a single serine, is currently also available for three lectins that are structurally distinct from BfL. While there are no meaningful similarities between the topologies and the modes of ligand binding by BfL and these lectin, for completeness we provide their brief description below. One of these proteins is a second lectin isolated from elderberry bark, SNA-II, which belongs to the ricin B family with a β-trefoil topology. In the structure of the SNA-II complex (PDB ID 3CA6; [36]) the atoms O3 and O4 of GalNAc form canonical hydrogen bonds with the side chain of Asp16, with O3 also hydrogen bonded to OE1 of Gln36. Additional hydrogen bonds between the carbohydrate and the protein involve O6 interacting with NE2 of Gln29 and N of Asn19. Extensive hydrophobic interactions are provided by Trp31, the side chain of which is largely parallel to the sugar plane. Neither the serine residue of the Tn antigen nor the N-acetyl group of GalNAc are involved in direct interactions with the protein. Interactions with the *Helix pomatia* (edible snail) lectin (HPA), which belongs to the H-type family, have been extensively used in histopathology to identify cancer cells with aberrant glycosylation, thus establishing the specificity of HPA for binding of the Tn antigen. HPA in its native forms exists as a homohexamer and its fold (PDB ID 2CGZ; [37]) bears no similarity to that of BfL. The interactions with the Tn antigen are also completely different in both lectins. Oxygens O4 and O5 are hydrogen bonded to NE and NH2 of Arg63, and O3 to O of Gly24. The oxygen atom of the N-acetyl group is hydrogen bonded to N of Asp26, whereas the serine residue of the antigen makes no contacts with the protein. CBM32-4 is a carbohydrate-binding module of a large multimodular α-N-acetylglucosaminidase, CpGH89, from the bacterium *Clostridium perfringens* and it does not exhibit any sequence or structure resemblance to BfL. The crystal structure of CBM32-4 in complex with Ser(Tn) (PDB ID 4A44; [38]) shows that this lectin binds the GalNAc utilizing hydrogen bonds between NH1 and NH2 atoms of Arg1423, and O3 and O4 atoms of the carbohydrate, with the former also donating a hydrogen bond to OE2 of Glu1376, and the latter accepting a hydrogen bond from NE2 of His1392. The oxygen of the N-acetyl group makes a water-mediated interaction with two main-chain amide groups, and the aromatic ring of Tyr1395 provides hydrophobic

interactions with the sugar ring. Serine residue of the Tn antigen is not involved in any discernible interactions with the protein.

Discussion

Lectins are used in a wide variety of basic research and clinical applications. For example, lectins are used extensively to monitor changes in carbohydrate expression in cells/tissues. In addition, several lectins and lectin conjugates have progressed into clinical trials for the treatment of cancer. While useful, there are a limited number of lectins with a narrow set of binding specificities, and many carbohydrates cannot be targeted with the currently utilized lectins. Furthermore, many available lectins can only be obtained in small quantities and/or are not accessible in pure, homogeneous form. To overcome these limitations, new lectins and expression methods are needed. Moreover, access to mutants and/or protein fusions could also expand the scope of applications. For example, lectin-GST fusions can be displayed with defined orientation on surfaces for more consistent glycan detection, and mutations of carbohydrate binding domains can provide new lectins with altered selectivity and/or affinity.

BfL is a newly discovered lectin with potential for use in a variety of technological and medical applications. To better understand its binding and biological properties and to fully exploit this protein for various applications, we used a combination of cell binding assays, glycan microarray studies, and X-ray crystallography to gain new insight into the molecular recognition properties and functional attributes of BfL.

First, we developed a successful protocol for the expression and purification of recombinant BfL. The procedures used during production of the recombinant BfL described here were repeated multiple times and the results obtained were consistent in terms of the quality and properties of the final preparations, demonstrating that the expression and purification procedures are robust. Although recombinant BfL is not identical with natural BfL (i.e. it lacks N-glycosylation and contains two Ca^{2+} ions rather than one Ca^{2+} and one Mn^{2+}), it displayed similar biological activity against MCF7 and MDA-MB-231 cells. In addition, this preparation was subsequently found to be active against other cell lines and to bind tightly to a variety of structurally-similar carbohydrates on a microarray. Taken together, the results demonstrate the successful production of pure, well-characterized recombinant BfL in functional form.

Access to recombinant BfL enabled evaluation of its activity against a large panel of cancer cell lines in the NCI-60 screen [16]. BfL displayed modest activity against 22 cell lines and strong activity against five cell lines. The most potent activity was observed for the melanoma cell line LOX IMVI (>95% inhibition). It should be noted that the presence of serum in cell culture media can significantly decrease the cytotoxic effects of lectins, as reported previously [39], and studies on lectins are often carried out with less than 1% serum. The activity assays used in this work were performed under more stringent conditions (5% serum), showing that inhibitory properties of BfL are robust. Furthermore, the consistency of results obtained for the recombinant and natural BfL suggests that the cytotoxic properties of both variants are very similar.

Although BfL was expected to recognize glycans, the specificity of BfL was not known. Access to recombinant BfL enabled in-depth analysis of its carbohydrate-binding properties. Using a glycan microarray containing approximately 500 array components, we evaluated binding of BfL to a diverse collection of potential ligands. BfL was found to bind tightly to glycans carrying a GalNAc residue at the non-reducing termini, such as Tn peptides, blood group A antigens, Forssman di- and tetrasaccharides, and core 5 glycopeptides. Tn peptides were the best binders with apparent K_d values in the range of 0.4 to 1.2 nM. The preference of BfL for a specific GalNAc anomer appears to be of a secondary significance, as shown by the very comparable values of the apparent K_d (0.52 and 0.54 nM for the α- and β-anomer, respectively). We also observed differences in binding due to density of glycans attached to the array. Whereas Tn antigens bound tightly to both the low- and high-density arrays, a strong dependence was observed for blood group A antigens, which displayed very little binding at a low glycan density.

Production of pure, homogenous recombinant BfL also enabled an in depth study of protein structure and revealed the atomic basis of a ligand-specificity. High-resolution crystal structures were determined for the ligand-free lectin, as well as for its complexes with three Tn glycopeptides, globotetraose, and the blood group A antigen. An analysis of the interactions between the lectin and a collection of different ligands identified the subset of seven hydrogen bonds common to all the studied complexes. These structures were then compared to other GalNAc binding lectins for which crystal structures are available. Analysis of the eight crystal structures and comparisons with related lectins revealed several unique features of GalNAc recognition. Of special note, the carboxylate group of Glu126 forms H-bonds with both the N-acetyl of GalNAc and the peptide amido group of Tn antigens. This type of stabilization has not been observed previously for a GalNAc-specific lectin.

Many structures of plant lectins, including members of the Concanavalin A family, have been described previously [40], but it is important to note that the extent of structural and specificity analysis reported in this study is rare. To the best of our knowledge, there is no report presenting structures of a single lectin with such a broad spectrum of ligands. Furthermore, only one rather low-resolution structure of a complex between a lectin (unrelated to BfL) and a Tn peptide in which the ligand molecule extended beyond a single amino acid residue linked to the Tn group was described previously [31]. Therefore, the set of BfL structures determined in this study, that includes a ligand-free lectin, complexes with a simple sugar and small oligosaccharides, as well as with three distinct Tn peptides, is quite unique.

In addition to providing new insights into carbohydrate recognition, the results presented in this study provide a foundation for engineering new variants of BfL with unique properties. The procedure for preparation and purification of recombinant BfL is amenable to preparation of mutants. Additionally, the analysis of carbohydrate binding properties and the structural understanding of carbohydrate recognition create an excellent basis for engineering modified lectins. In particular, the presence of a glutamate residue in position 126 appears to be the major determinant of the specificity for GalNAc, and we predict that its replacement by other amino acids will lead to significant changes in affinity and/or

specificity. We also anticipate that mutations of Gly95 and/or Gly96 could substantially modify the properties of BfL. Whereas the main chain conformations of the two glycine residues are compatible with larger amino acids, addition of side chains will modify the carbohydrate-binding pocket in a way that most likely would prevent an accommodation of the N-acetyl group of GalNAc. These and similar changes to the sequence of BfL could lead to the development of engineered lectins with rationally-designed properties.

Materials and methods

Cloning and expression of BfL-His$_6$

The synthetic gene encoding the sequence of BfL was purchased from Bio Basic Inc. Specifically, the DNA sequence was derived by a back-translation of the previously published amino acid sequence of this lectin [12] with a simultaneous codon optimization, tuned for an expression in *E. coli*. The sequence of natural BfL, followed by the three-amino acids long linker (GlyAlaArg), His$_6$- affinity tag and a termination codon, was cloned into the expression vector pRSET A (Thermo Fisher Sci.) using standard protocols. After purification from monoclonal cultures using standard kits (Qiagen), the expression plasmid was transformed into BL21(DE3) pLysS competent cells (EMD Millipore). For protein production, cells were cultured in a Luria-Bertani (LB) medium, supplemented with 100 μg/ml ampicillin and 30 μg/ml chloramphenicol at 37°C until they reached a mid-log phase (A$_{600}$=0.5), at which stage the expression was induced by the addition of isopropyl β-D-thiogalactopyranoside, IPTG, (American Bioanalytical) to a final concentration of 0.5 mM. After culturing for additional four hours with intermittent extraction of samples for analysis, cells were harvested by centrifugation and stored at −80°C for further processing.

Refolding and purification of BfL-His$_6$

The following buffers and stock solutions have been used in subsequent procedures. Buffer A, 25 mM Tris-HCl (pH 7.5); buffer B, 50 mM Tris-HCl (pH 7.5), 0.15 M NaCl, 5 mM EDTA; buffer C, 50 mM Tris-HCl (pH 7.5), 1 M NaCl, 2% (w/v) Triton X-100, 5 mM EDTA; buffer D, 2 M urea, 0.1 M Tris-HCl (pH 7.5), 5 mM EDTA; buffer E, 25 mM Tris-HCl (pH 7.5), 1 M NaCl, 5 mM EDTA; lysozyme (stock), 50 mg/ml in water; DNAse I (stock), 1 mg/ml in 50% glycerol and 75 mM NaCl; MgCl$_2$ (stock), 0.5 M in water; EDTA (stock), 0.5 M in 50 mM Tris (pH 8.0). All quantities, listed below, are scaled to 1 liter of the cell culture. Bacterial cells, separated from the media by centrifugation, were re-suspended in 13 ml of buffer A and passed two times through homogenizer. After adding 100 μl lysozyme (stock), 250 μl DNAse I (stock), 50 μl MgCl$_2$, (stock), and 12.5 ml of buffer A, suspension was incubated for 30 min at room temperature. Following addition of 350 μl of EDTA (stock), cell suspension was frozen in liquid nitrogen and then quickly thawed at a water bath set to 37 °C. After addition of 200 μl of MgCl$_2$ (stock) suspension was incubated for 30 min at room temperature then supplied with 350 μl EDTA (stock). All following procedures were conducted at 5 °C. After isolating the pellet by centrifugation (20 min. at 20,000×*g*) it was suspended in 10 ml of buffer B (using a glass dounce). The solubilized components were separated by centrifugation (20 min. at 20,000×*g*). Pellet was then suspended in 10 mL of buffer C and sonicated. This step was repeated, each time separating soluble fraction by a centrifugation (20 min. at 20,000×*g*). Subsequently, pellet was

suspended in 10 ml of buffer D and centrifuged (repeated 2×). Finally, buffer E was used for washing inclusion bodies with intermittent centrifugation. Washed inclusion bodies were suspended in a few ml of buffer A and quantified. Aliquots containing 50 mg of inclusion bodies were centrifuged, and after discarding a supernatant, samples were frozen at −80 °C.

All steps of a refolding process were conducted at 5 °C. For refolding, 100 mg of BfL-His6-containing inclusion bodies were dissolved in 100 ml of 6 M guanidinium chloride, buffered with 0.1 M Tris-HCl (pH 7.5) and cleared by centrifugation (30 min., 30,000×g). Solution of denatured lectin was then applied dropwise (0.1 ml/min.) to 2.3 liter of the refolding solution containing 0.1 M Tris-HCl (pH 7.5), 0.5 M NaCl, 0.5 M arginine-hydrochloride (independently buffered with 0.1 M Tris-HCl, pH 7.5), and 1 mM $CaCl_2$. The refolding mixture was stirred continuously at 225 rpm, and the process continued overnight. Next day, solution was cleared by centrifugation followed by filtration through 0.22 μM membrane. Cleared solution was then concentrated to a final volume of about 50-70 mL using a 2.5 l pressure stir cell concentrator with 10 kDa regenerated cellulose membrane. Following the concentration, solution containing refolded recombinant BfL was dialyzed overnight against 3.5 l of buffer containing 50 mM Tris-HCl (pH 7.5), 0.2 M NaCl, and 1 mM $CaCl_2$. On the following day, the protein was concentrated to 8-10 mg/ml at room temperature (solubility of BfL is very temperature-dependent), aliquoted and frozen at −80 °C. The mass spectroscopy analysis of a BfL sample showed the molecular weight of 27,244.8 Da, nearly identical to the theoretical value of 27,245.1 Da.

Prior to any experiments, assays, or measurements, a suitable sample of a recombinant BfL was subjected to gel filtration chromatography using Superdex 75 HR 10/30 column (GE Healthcare Life Sci.) equilibrated with the buffer containing 50 mM Hepes, pH 7.5 and 150 mM NaCl. After merging fractions containing pure recombinant BfL, protein was concentrated to 8-10 mg/ml and used for subsequent studies without freezing.

BfL ligands

2-(acetylamino)-2-deoxy-α/β-D-galactose (α/β-GalNAc, MW=221.21) was purchased from Sigma-Aldrich (SKU A2795). GalNAcβ1-3Galα1-4Galβ1-4Glc (P antigen, globotetrose, Gb4, MW 707.62) was purchased from Elicityl (SKU GLY121, www.elicityl-oligotech.com). GalNAcα1-3(Fucα1-2)Gal (Blood Group A, MW=529.49) was purchased from V-Labs Inc. (SKU L305, www.v-labs.com).

Preparation of Tn-peptides

Three Tn peptides, Ace-Gly-Val-(Tn)Thr-Ser-Ala-Gly-Hexyl-COOH (Tn-200, MW 848.91), Ace-Ser-(Tn)Thr-Val-Gly-Hexyl-COOH (Tn-207, MW 720.71), and N_3-Hexyl-Ser-(Tn)Ser-Val-Gly-Hexyl-COOH (Tn-168, MW 803.87) have been synthesized and purified as described earlier [41] and their detailed characteristics are provided below. Structural formulas representing all these ligands are shown in **Figure 3**.

Ac-Gly-Val-(Tn)Thr-Ser-Ala-Gly-Hexyl-COOH (Tn-200)—The glycopeptide was synthesized using published protocols [41, 42]. Product was 98.0% pure after purification as

shown by analytical reverse phase HPLC. Yield after resin split calculated for last three couplings including coupling of the glycoamino acid: 26.4 mg (89.5%).

^{1}H NMR (400 MHz, D_2O) δ 4.95 (d, J = 3.9 Hz, 1H), 4.65 (d, J = 2.3 Hz, 1H), 4.51 (t, J = 5.7 Hz, 1H), 4.32 (td, J = 10.9, 10.3, 6.8 Hz, 3H), 4.11 (dd, J = 11.0, 3.8 Hz, 1H), 4.06 – 3.82 (m, 9H), 3.76 (d, J = 6.2 Hz, 2H), 3.23 (t, J = 6.8 Hz, 2H), 2.41 (t, J = 7.4 Hz, 2H), 2.12 (dt, J = 14.2, 7.0 Hz, 1H), 2.06 (d, J = 4.5 Hz, 6H), 1.63 (p, J = 7.5 Hz, 2H), 1.54 (p, J = 7.1 Hz, 2H), 1.43 (d, J = 7.2 Hz, 3H), 1.39 – 1.31 (m, 2H), 1.29 (d, J = 6.4 Hz, 3H), 0.99 (dd, J = 6.8, 4.4 Hz, 6H). ^{1}H NMR excludes 13 exchangeable H's. ^{13}C NMR (400 MHz, D_2O) δ 179.59, 176.05, 175.37, 174.55, 174.54, 172.17, 171.85, 171.66, 171.57, 99.52, 77.06, 71.91, 69.16, 68.72, 62.13, 61.85, 60.06, 57.79, 55.54, 50.62, 50.37, 43.22, 43.07, 39.78, 34.30, 30.73, 28.54, 26.05, 24.52, 22.88, 22.27, 19.12, 18.84, 18.31, 16.97. FIA-MS calculated for $C_{35}H_{60}N_8O_{16}$: 849.4 [M-H][+]; found 849.3, 871.4 [M-Na][+]; found 871.3.

Ac-Ser-(Tn)Thr-Val-Gly-Hexyl-COOH (Tn-207)—The glycopeptide was synthesized using published protocols [41, 42]. Product was 97.2% pure after purification as shown by analytical reverse phase HPLC. Yield after resin split calculated for last three couplings including coupling of the glycoamino acid: 24.1 mg (87.0%).

^{1}H NMR (400 MHz, D_2O) δ 4.89 (d, J = 3.9 Hz, 1H), 4.63 (d, J = 2.4 Hz, 1H), 4.59 (t, J = 6.1 Hz, 1H), 4.37 (qd, J = 6.3, 2.5 Hz, 1H), 4.16 – 4.09 (m, 2H), 4.02 (t, J = 6.2 Hz, 1H), 3.98 (d, J = 3.1 Hz, 1H), 3.93 – 3.72 (m, 7H), 3.21 (t, J = 6.9 Hz, 2H), 2.40 (t, J = 7.4 Hz, 2H), 2.12 – 2.01 (m, 1H), 2.06 (d, J = 5.8 Hz, 6H), 1.62 (p, J = 7.5 Hz, 2H), 1.53 (p, J = 7.0 Hz, 2H), 1.39 – 1.31 (m, 2H), 1.28 (d, J = 6.4 Hz, 3H), 0.96 (d, J = 6.8 Hz, 6H). ^{1}H NMR excludes 11 exchangeable H's. ^{13}C NMR (400 MHz, D_2O) δ 179.59, 174.98, 174.75, 174.18, 173.16, 171.94, 171.32, 99.22, 76.06, 72.05, 69.16, 68.74, 61.89, 61.72, 60.37, 57.94, 55.97, 50.40, 43.14, 39.80, 34.28, 30.80, 28.53, 26.06, 24.52, 22.88, 22.25, 18.92, 18.69, 18.48. FIA-MS calculated for $C_{30}H_{52}N_6O_{14}$: 721.4 [M-H][+]; found 721.3, 743.3 [M-Na][+]; found 743.2.

N₃-Hexyl-Ser-(Tn)Ser-Val-Gly-Hexyl-COOH (Tn-168)—The glycopeptide was synthesized using published protocols [41, 42]. Product was 98.9% pure after purification as shown by analytical reverse phase HPLC. Yield after resin split calculated for last four couplings including coupling of the glycoamino acid: 6.5 mg (47.9%).

^{1}H NMR (400 MHz, D_2O) δ 4.90 (d, J = 3.6 Hz, 1H), 4.71 (t, J = 5.9 Hz, 1H), 4.49 (t, J = 5.9 Hz, 1H), 4.23 – 4.10 (m, 2H), 4.02 – 3.92 (m, 2H), 3.86 (dd, J = 12.7, 7.6 Hz, 7H), 3.78 – 3.69 (m, 2H), 3.34 (t, J = 6.9 Hz, 2H), 3.22 (t, J = 7.0 Hz, 2H), 2.37 (q, J = 8.0 Hz, 4H), 2.16 – 2.07 (m, 1H), 2.06 (s, 3H), 1.64 (dq, J = 7.32, 6.76 Hz, 6H), 1.54 (p, J = 7.32 Hz, 2H), 1.46 – 1.29 (m, 4H), 0.97 (d, J = 5.4 Hz, 6H). ^{1}H NMR excludes 11 exchangeable H's. ^{13}C NMR (400 MHz, D_2O) δ 180.01, 177.80, 175.09, 174.22, 172.72, 171.86, 171.39, 98.43, 71.95, 69.06, 68.37, 67.34, 61.79, 61.70, 60.44, 55.98, 54.04, 51.60, 50.32, 43.08, 39.82, 35.80, 34.63, 30.76, 28.56, 28.32, 26.11, 26.07, 25.35, 24.66, 22.70, 18.94, 18.39. FIA-MS calculated for $C_{33}H_{57}N_9O_{14}$: 804.4 [M-H][+]; found 804.4, 826.4 [M-Na][+]; found 826.4.

Sedimentation velocity analysis

The sedimentation analysis was completed for solutions of the recombinant BfL at three protein concentrations, 0.33, 0.42, and 1.0 mg/ml, respectively, in the buffer containing 150 mM NaCl and 50 mM Hepes pH 7.5. The measurements were performed on ProteomeLab XL-I analytical ultracentrifuge (Beckman-Coulter Co.). Samples were equilibrated at 15 °C for one hour and then sedimented at 40,000 rpm. Throughout the experiment, the UV absorbance at 280 nm was measured for more than 350 scans; however, only the first 200 scans were included in the analysis due to complete protein sedimentation. Analysis was performed with the program Sedfit [20].

Glycan microarray fabrication

Glycans used for microarrays were conjugated to either bovine serum albumin (BSA) or human serum albumin (HSA) to produce neoglycoproteins that mimicked the density of spacing of natural glycoproteins. Some glycans were conjugated at high and low densities, as denoted by the number following the array component name. Array printing was done on a MicroGrid II arrayer (Biorobotics) by 946MP1.5 pins (ArrayIt) as previously described [43]. Five hundred and three array components were printed at 125–200 μg/mL in duplicate on epoxy-coated slides (ArrayIt). Atto555 azide dye (Thermo Scientific) at 0.7 ug/mL was mixed in the print buffer to evaluate print quality prior to performing experimental assays. Each slide contained 16 identical subarrays with an average spot size of 80 μm. Prior to being used for experimental assays, the representative slides from the print batch were assessed for quality and reproducibility using a variety of lectins and monoclonal antibodies known to bind certain components. The slides were stored in vacuum-sealed boxes at –20 °C until use.

High-throughput profiling of BfL binding to glycans

Prior to each experiment, each microarray slide was scanned in an InnoScan 1100 AL fluorescence scanner to check for any defects and missing print spots (i.e. presence or absence of soluble Atto555 azide dye). The slides were placed in 16-well modules using metal clips (Grace Bio-Labs) to separate the 16 identical arrays and allow for different experimental conditions. The array slides were blocked overnight at 4 °C in 3% BSA TBS (200 μl/well) and then washed six times with TBS buffer (20 mM Tris-HCl, 150mM NaCl, pH 7.4) containing 0.05% Tween-20 (TBS-T). The BfL samples were prepared in 1% BSA TBS containing 2mM $CaCl_2$ and 0.05% Tween-20 (100 μl/well). All BfL dilutions were measured in duplicate wells printed by different pins on different slides in order to minimize pin and slide variations. After incubating for 2 h at 37 °C and gentle agitation (100 rpm), the slide was washed six times in TBS-T (200 μl/well). To detect bound BfL, Penta-His mouse anti-IgG$_1$ antibody (Qiagen) was diluted 1:200 in 1% BSA TBS and added to the slides (100 μl/well). After incubating for an hour at 37 °C and gentle agitation, the slides were washed six times in TBS-T. Finally, a 1:500 dilution of DyLight549 anti-mouse IgG (Jackson ImmunoResearch) was prepared in TBS containing 1% BSA and 100 μl were added to each well. Slides were covered with aluminum foil to prevent photobleaching and then incubated for 1 hour at 37 °C with gentle agitation. Following six washes with TBS-T, slides were

removed from the module and soaked in TBS-T for five minutes prior to being dried in a centrifuge at 1000 rpm for 5 minutes.

Image analysis and data processing of arrays

The arrays were scanned at 5 μm pixel resolution with an InnoScan 1100 AL Fluorescence Scanner. The photomultiplier tube settings were the same for all experiments to limit unintentional signal variation and slides were scanned at one setting to limit spot signal saturation. Spots were defined as circular features with diameter 60–80 μm, and fluorescence intensity for each component was quantified with GenePix 7.0 software (Molecular Probes). The GenePix array file was manually adjusted for each spot to reflect their actual sizes and placement. Any features marked as missing or defective in the prescan were excluded from further analysis. The background corrected median was used for data analysis, and spots with intensity lower than 150 RFU (1/2 the typical IgM background) were set to 150. The signals for replicate spots on duplicate wells were averaged and log-transformed (base 2) for future analysis. GraphPadPrism6 was used to calculate the apparent K_d curves for the array components.

NCI-60 Human Tumor Cell Lines Screen

Cytotoxic properties of the recombinant BfL were tested against the human cancer cell lines included within the NCI-60 panel [16]. The screen was performed by the Developmental Therapeutics Program (DTP) in the National Cancer Institute according to the published protocols (https://dtp.cancer.gov/discovery_development/nci-60/). The recombinant BfL, at the concentration of 1.85 μM, was subjected to the One Dose Assay which utilizes standard sulforhodamine B colorimetric assay [44]. Each experiment was performed in duplicate and the averaged results, representing the growth of cells in the presence of the lectin relative to their growth in drug-free medium are shown in **Figure 1**. Two melanoma cell lines which are components of the NCI-60 panel, MALME-3M and UACC-257, were excluded due to the failure of cell growth during the assay.

Crystallization

Initial crystallization trials were performed at 20 °C using the Phenix crystallization robot (Art Robbins Instruments) and a wide range of commercial crystallization screens. The sitting droplets were prepared by mixing buffered protein solutions (5 - 8 mg/ml in 50 mM HEPES pH 7.5, and 150 mM NaCl) with the crystal screen solutions. On average, the volume of each droplet was about 0.25 μl with different volume ratios of protein to reservoir solutions. While crystals grew under many different conditions, a common feature seemed to be the presence of medium-to-large molecular weight polyethylene glycol (PEG) as a precipitant and the pH value within a range of 7.5-8.5. In most cases the crystals were shaped as blocks with similar dimensions along all edges. After optimization, the best crystals of the ligand-free BfL, BfL(apo) were grown from hanging drops obtained by mixing a protein solution with the precipitant composed of 12% (w/v) PEG8000, 8% (v/v) ethylene glycol, and 0.1 M Hepes (pH 7.5), in a volume ratio 1:1. Crystals of the complex with GalNAc were obtained by adding a 20-fold molar excess of the ligand to droplets harboring crystals of ligand-free BfL and incubating for several hours to a few days. Crystals

of all other complexes described in this study were obtained by co-crystallization, always using 5-fold molar excess of the ligand over BfL. In each case a solution of the complex was mixed with precipitant solution in the volume ratio 1:1 and the lectin concentration was 5-6 mg/ml. The crystallization conditions used for each complex are presented in **Table 4**. For diffraction experiments, individual crystals were harvested using Litholoops (Molecular Dimensions), briefly flashed in the cryoprotectant solution, in all cases composed of 12% (w/v) PEG8000, 30% (v/v) ethylene glycol, and 0.1 M Hepes (pH 7.5), and frozen in a liquid nitrogen (−200 °C) for subsequent diffraction studies.

Collection of X-ray data

X-ray diffraction data were collected at the Advanced Photon Source in Argonne National Laboratory, (Argonne, IL, USA). All diffraction experiments were conducted at −200 °C. The experimental images processed with subsequent scaling of reflection intensities using the program HKL3000 (HKL Research Inc.) [45]. Details of experimental data collection and processing statistics are presented in **Table 2**.

Structure solution and refinement

The structure of BfL was solved by the molecular replacement method using the program PHASER [46] and the previously published structure of the *Griffonia simplicifolia* lectin 1-B(4) (PDB entry 1HQL, monomer A) as a model [30]. The initial solution was subjected to automatic rebuilding with the program MR Rosetta [47] and structural refinement with the program Refmac5 [48]. The automatic procedures were assisted by manual model modifications aided by the program Coot [49]. At the final stages of refinement, performed at resolution of 1.43 Å, B-factors were modeled anisotropically. The resulting statistics are shown in **Table 5**. Structure of BfL(LF) served as the template for solving all remaining structures reported here. Because of high isomorphism of BfL(LF) and BfL(GalNAc) crystals, a simple substitution of experimental data, followed byrigid-body refinement assured high-quality initial phases associated with well-defined difference electron density outlining the positions of GalNAc ligands. In the case of the remaining complexes, a model of the BfL monomer was used during molecular replacement aided by the program PHASER [46] to define the initial phases. As in the case of BfL(GalNAc), a high quality of initial phases and associated difference electron densities allowed unambiguous identification and placement of the ligands. Further refinement with Refmac5, interspersed with visual inspection and manual corrections to the models with Coot, led to high quality structures. An exception was the lower-resolution structure of BfL(Tn-200) which is less complete and accurate. However, even in that case the location of the Thr-GalNAc fragment of Tn peptide could be unambiguously modeled in the electron density. The statistics for all complexes of BfL included in this study are shown in **Table 5**.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Abbreviations

BfL	*Bauhinia forficata* lectin
GalNAc	N-acetylgalactosamine
NCI	National Cancer Institute
aOSM	asialo ovine submaxillary mucin

References

1. Xu Y, Sette A, Sidney J, Gendler SJ, Franco A. Tumor-associated carbohydrate antigens: a possible avenue for cancer prevention. Immunol Cell Biol. 2005; 83:440–8. [PubMed: 16033540]

2. Nakahara S, Raz A. Biological modulation by lectins and their ligands in tumor progression and metastasis. Anticancer Agents Med Chem. 2008; 8:22–36. [PubMed: 18220503]

3. Yau T, Dan X, Ng CC, Ng TB. Lectins with potential for anti-cancer therapy. Molecules. 2015; 20:3791–810. [PubMed: 25730388]

4. Ju T, Wang Y, Aryal RP, Lehoux SD, Ding X, Kudelka MR, Cutler C, Zeng J, Wang J, Sun X, Heimburg-Molinaro J, Smith DF, Cummings RD. Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers. Proteomics Clin Appl. 2013; 7:618–31. [PubMed: 23857728]

5. Ju TZ, Otto VI, Cummings RD. The Tn antigen-structural simplicity and biological complexity. Angew Chem Int Edit. 2011; 50:1770–1791.

6. Poiroux G, Pitie M, Culerrier R, Lafont E, Segui B, Van Damme EJ, Peumans WJ, Bernadou J, Levade T, Rouge P, Barre A, Benoist H. Targeting of T/Tn antigens with a plant lectin to kill human leukemia cells by photochemotherapy. PLoS One. 2011; 6:e23315. [PubMed: 21858067]

7. Badr HA, Alsadek DM, Darwish AA, Elsayed AI, Bekmanov BO, Khussainova EM, Zhang X, Cho WC, Djansugurova LB, Li CZ. Lectin approaches for glycoproteomics in FDA-approved cancer biomarkers. Expert Rev Proteomics. 2014; 11:227–36. [PubMed: 24611567]

8. Ohyama C, Hosono M, Nitta K, Oh-eda M, Yoshikawa K, Habuchi T, Arai Y, Fukuda M. Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy. Glycobiology. 2004; 14:671–9. [PubMed: 15044396]

9. Syed P, Gidwani K, Kekki H, Leivo J, Pettersson K, Lamminmaki U. Role of lectin microarrays in cancer diagnosis. Proteomics. 2016; 16:1257–65. [PubMed: 26841254]

10. De Mejia EG, Prisecaru VI. Lectins as bioactive plant proteins: a potential in cancer treatment. Crit Rev Food Sci Nutr. 2005; 45:425–445. [PubMed: 16183566]

11. Huang LH, Yan QJ, Kopparapu NK, Jiang ZQ, Sun Y. Astragalus membranaceus lectin (AML) induces caspase-dependent apoptosis in human leukemia cells. Cell Prolif. 2012; 45:15–21. [PubMed: 22172162]

12. Silva MCC, Santana LA, Mentele R, Ferreira RS, de Miranda A, Silva-Lucca RA, Sampaio MU, Correia MTS, Oliva MLV. Purification, primary structure and potential functions of a novel lectin from Bauhinia forficata seeds. Process Biochem. 2012; 47:1049–1059.

13. Silva MCC, de Paula CAA, Ferreira JG, Paredes-Gamero EJ, Vaz AMSF, Sampaio MU, Correia MTS, Oliva MLV. Bauhinia forficata lectin (BfL) induces cell death and inhibits integrin-mediated adhesion on MCF7 human breast cancer cells. BBA-Gen Subjects. 2014; 1840:2262–2271.

14. Streicher H, Sharon N. Recombinant plant lectins and their mutants. Methods Enzymol. 2003; 363:47–77. [PubMed: 14579567]

15. Oliveira C, Teixeira JA, Domingues L. Recombinant lectins: an array of tailor-made glycan-interaction biosynthetic tools. Crit Rev Biotechnol. 2013; 33:66–80. [PubMed: 22530774]

16. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer. 2006; 6:813–23. [PubMed: 16990858]

17. Sharon N, Lis H. Legume lectins--a large family of homologous proteins. FASEB J. 1990; 4:3198–208. [PubMed: 2227211]

18. Brinda KV, Surolia A, Vishveshwara S. Insights into the quaternary association of proteins through structure graphs: a case study of lectins. Biochem J. 2005; 391:1–15. [PubMed: 16173917]

19. Pohleven J, Strukelj B, Kos J. Affinity Chromatography of Lectins. In: Magdeldin S, editorAffinity Chromatography. InTech Rijeka; Croatia: 2012. 49–74.

20. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. 2000; 78:1606–1619. [PubMed: 10692345]

21. Gordus A, MacBeath G. Circumventing the problems caused by protein diversity in microarrays: implications for protein interaction networks. J Am Chem Soc. 2006; 128:13668–9. [PubMed: 17044677]

22. Hardman KD, Ainsworth CF. Structure of concanavalin A at 2.4-A resolution. Biochemistry. 1972; 11:4910–4919. [PubMed: 4638345]

23. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. 2004; D60:2256–2268.

24. Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W. Data mining of metal ion environments present in protein structures. J Inorg Biochem. 2008; 102:1765–76. [PubMed: 18614239]

25. Senguen FT, Grabarek Z. X-ray structures of magnesium and manganese complexes with the N-terminal domain of calmodulin: insights into the mechanism and specificity of metal ion binding to an EF-hand. Biochemistry. 2012; 51:6182–94. [PubMed: 22803592]

26. Hudson KL, Bartlett GJ, Diehl RC, Agirre J, Gallagher T, Kiessling LL, Woolfson DN. Carbohydrate-aromatic interactions in proteins. J Am Chem Soc. 2015; 137:15152–60. [PubMed: 26561965]

27. Takahashi K, Ono S. Calorimetric studies on the mutarotation of D-galactose and D-mannose. J Biochem. 1973; 73:763–70. [PubMed: 4720061]

28. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol. 1993; 233:123–38. [PubMed: 8377180]

29. Delbaere LT, Vandonselaar M, Prasad L, Quail JW, Wilson KS, Dauter Z. Structures of the lectin IV of griffonia simplicifolia and its complex with the Lewis b human blood group determinant at 2.0 A resolution. J Mol Biol. 1993; 230:950–965. [PubMed: 8478943]

30. Tempel W, Tschampel S, Woods RJ. The xenograft antigen bound to Griffonia simplicifolia lectin 1-B(4). X-ray crystal structure of the complex and molecular dynamics characterization of the binding site. J Biol Chem. 2002; 277:6615–21. [PubMed: 11714721]

31. Babino A, Tello D, Rojas A, Bay S, Osinaga E, Alzari PM. The crystal structure of a plant lectin in complex with the Tn antigen. FEBS Lett. 2003; 536:106–10. [PubMed: 12586347]

32. Kulkarni KA, Sinha S, Katiyar S, Surolia A, Vijayan M, Suguna K. Structural basis for the specificity of basic winged bean lectin for the Tn-antigen: A crystallographic, thermodynamic and modelling study. FEBS Lett. 2005; 579:6775–6780. [PubMed: 16310781]

33. Madariaga D, Martinez-Saez N, Somovilla VJ, Coelho H, Valero-Gonzalez J, Castro-Lopez J, Asensio JL, Jimenez-Barbero J, Busto JH, Avenoza A, Marcelo F, Hurtado-Guerrero R, Corzana F,

Peregrina JM. Detection of tumor-associated glycopeptides by lectins: the peptide context modulates carbohydrate recognition. ACS Chem Biol. 2015; 10:747–56. [PubMed: 25457745]

34. Sousa BL, Silva Filho JC, Kumar P, Pereira RI, Lyskowski A, Rocha BA, Delatorre P, Bezerra GA, Nagano CS, Gruber K, Cavada BS. High-resolution structure of a new Tn antigen-binding lectin from Vatairea macrocarpa and a comparative analysis of Tn-binding legume lectins. Int J Biochem Cell Biol. 2015; 59:103–10. [PubMed: 25499445]

35. Sousa BL, Silva-Filho JC, Kumar P, Graewert MA, Pereira RI, Cunha RM, Nascimento KS, Bezerra GA, Delatorre P, Djinovic-Carugo K, Nagano CS, Gruber K, Cavada BS. Structural characterization of a Vatairea macrocarpa lectin in complex with a tumor-associated antigen: A new tool for cancer research. Int J Biochem Cell Biol. 2016; 72:27–39. [PubMed: 26751394]

36. Maveyraud L, Niwa H, Guillet V, Svergun DI, Konarev PV, Palmer RA, Peumans WJ, Rouge P, Van Damme EJ, Reynolds CD, Mourey L. Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from Sambucus nigra. Proteins. 2009; 75:89–103. [PubMed: 18798567]

37. Lescar J, Sanchez JF, Audfray A, Coll JL, Breton C, Mitchell EP, Imberty A. Structural basis for recognition of breast and colon cancer epitopes Tn antigen and Forssman disaccharide by Helix pomatia lectin. Glycobiology. 2007; 17:1077–83. [PubMed: 17652409]

38. Ficko-Blean E, Stuart CP, Suits MD, Cid M, Tessier M, Woods RJ, Boraston AB. Carbohydrate recognition by an architecturally complex alpha-N-acetylglucosaminidase from clostridium perfringens. PLoS One. 2012; 7:e33524. [PubMed: 22479408]

39. Faheina-Martins GV, da Silveira AL, Ramos MV, Marques-Santos LF, Araujo DA. Influence of fetal bovine serum on cytotoxic and genotoxic effects of lectins in MCF-7 cells. J Biochem Mol Toxicol. 2011; 25:290–6. [PubMed: 21523858]

40. Sinha S, Gupta G, Vijayan M, Surolia A. Subunit assembly of plant lectins. Curr Opin Struct Biol. 2007; 17:498–505. [PubMed: 17709241]

41. Zhang Y, Muthana SM, Farnsworth D, Ludek O, Adams K, Barchi JJ Jr. Gildersleeve JC. Enhanced epimerization of glycosylated amino acids during solid-phase peptide synthesis. J Am Chem Soc. 2012; 134:6316–25. [PubMed: 22390544]

42. Zhang Y, Muthana SM, Barchi JJ Jr. Gildersleeve JC. Divergent behavior of glycosylated threonine and serine derivatives in solid phase peptide synthesis. Org Lett. 2012; 14:3958–61. [PubMed: 22817697]

43. Campbell CT, Zhang Y, Gildersleeve JC. Construction and use of glycan microarrays. Curr Protoc Chem Biol. 2010; 2:37–53. [PubMed: 23836542]

44. Skehan P, Storeng R, Scudiero D, Monks A, McMahon J, Vistica D, Warren JT, Bokesch H, Kenney S, Boyd MR. New colorimetric cytotoxicity assay for anticancer-drug screening. J Natl Cancer Inst. 1990; 82:1107–12. [PubMed: 2359136]

45. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: The integration of data reduction and structure solution - from diffraction images to an initial model in minutes. Acta Crystallogr. 2006; D62:859–866.

46. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallograhic software. J Appl Cryst. 2007; 40:658–674. [PubMed: 19461840]

47. DiMaio F, Terwilliger TC, Read RJ, Wlodawer A, Oberdorfer G, Wagner U, Valkov E, Alon A, Fass D, Axelrod HL, Debanu D, Vorobiev SM, Iwai H, Pokkuluri PR, Baker D. Improved molecular replacement by density and energy guided protein structure optimization. Nature. 2011; 473:540–543. [PubMed: 21532589]

48. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. 2011; D67:355–367.

49. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr. 2004; D60:2126–2132.

50. Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. 2010; D66:12–21.
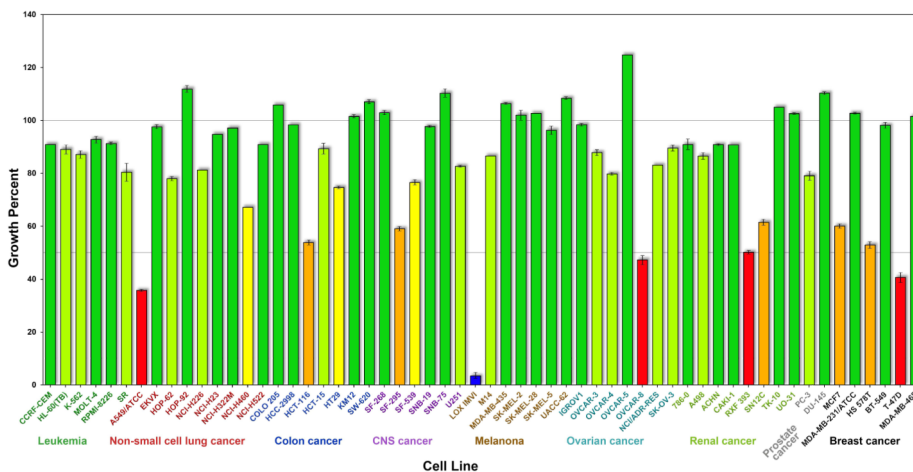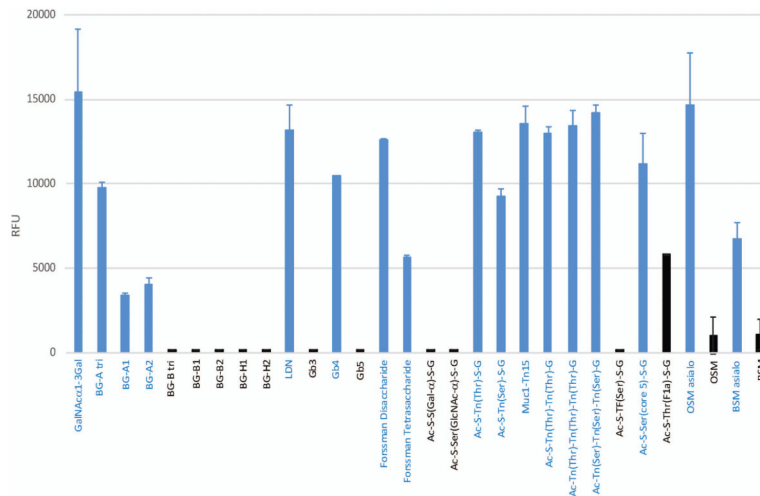
Figure 1.
An effect of the recombinant BfL used at a single concentration (1.85 μM) on the growth of
human cancer cells from the NCI-60 panel. For each cell line, the growth percent (GP)
represents a value relative to the growth of this cell line in an absence of the lectin and is an
average of two experiments (individual values are indicated by error bars). For ease of
analysis, cell lines listed along the horizontal axis are grouped by the type of cancer and the
growth bars are colored according to following scheme: green for GP > 90%, olive for 70%
< GP < 90%, yellow for 50% < GP < 70%, and red for GP < 50%. A single cell line, LOX-
IMVI, the growth of which was inhibited by ~97%, is indicated by a blue bar.

Figure 2.
Binding of *Bauhinia forficata lectin* (BfL) to a subset of glycan microarray components
(complete results are tabulated in the **Supplementary Table 1**). The subset used in this
figure includes all top binders, as well as a selection of non-binders. Binding was measured
as relative fluorescence units (RFU), given on the y-axis. Bars and glycan names in blue
correspond to array components that have a terminal-GalNAc residue. Error bars represent
the standard error (SEM) based on 4 replicate spots per array component (n=4).
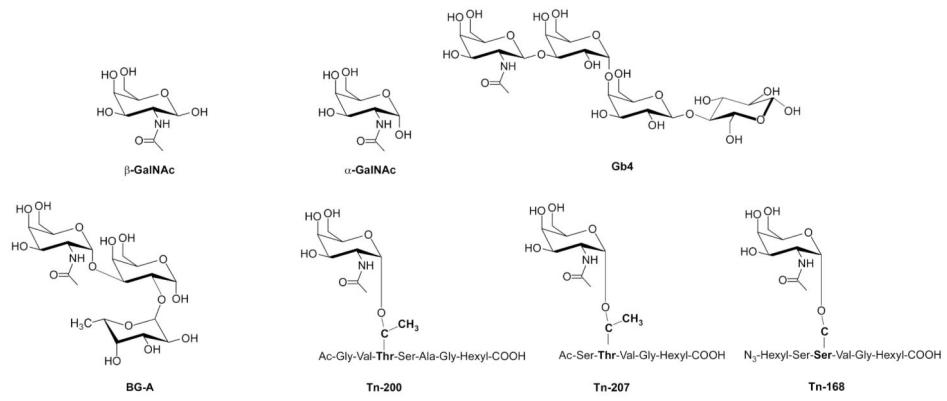
Figure 3.
Chemical structures of the BfL ligands used in the crystallization studies.
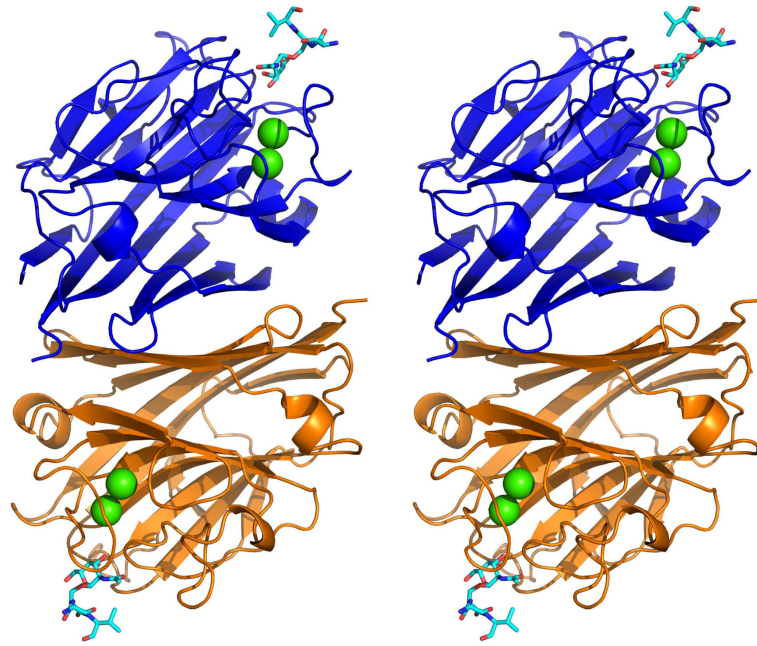
Figure 4.
A stereoview of a BfL dimer. The two molecules forming a dimer are colored orange and blue, calcium ions are green, and the bound Tn-168 glycopeptide is shown in sticks, with the carbons cyan, oxygens red, and nitrogens blue.
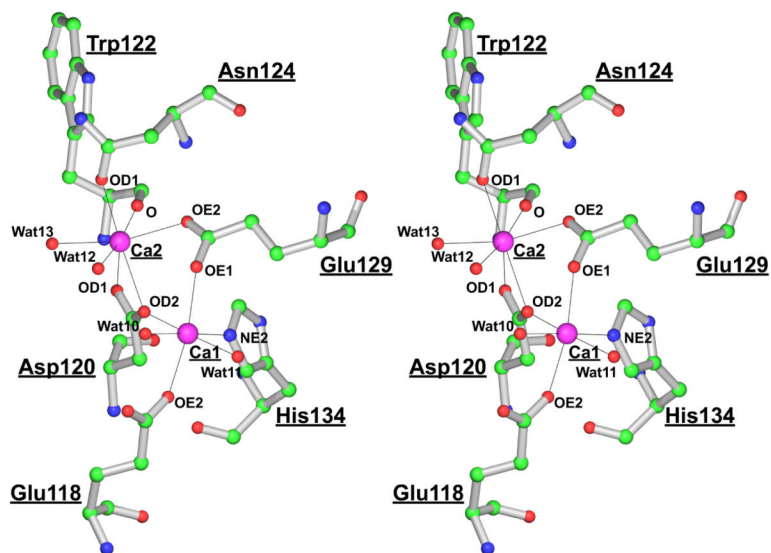
Figure 5.
Metal-binding sites of BfL. Two sites, Ca1 and Ca2, are indicated by larger spheres (magenta). In contrast to the majority of natural legume lectins, where the Ca1 site is occupied by the Mn^{2+} ion, in the recombinant preparation of BfL used in this study both sites are occupied by the Ca^{2+} ions. Each of two cations is coordinated by six ligands, two of which are water molecules. Furthermore, two side chains carboxylates, Asp120 and Glu129, each interacts with both metal ions.
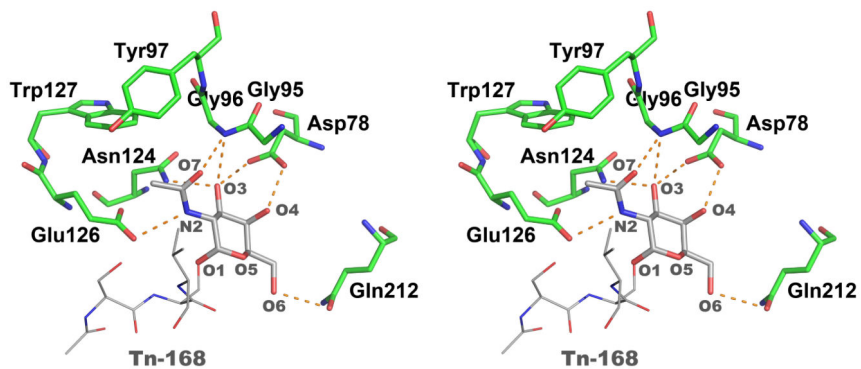
Figure 6.
A stereoview of primary interactions between BfL and the GalNAc residue (show here as part of the complex with Tn-168). The seven hydrogen bonds (dashed lines), common to all complexes of this lectin, provide the main contribution to the binding energy and define the ligand specificity of BfL. The peptide component of the ligand is shown as thin sticks.

β-GalNAc in pocket A

α-GalNAc in pocket B
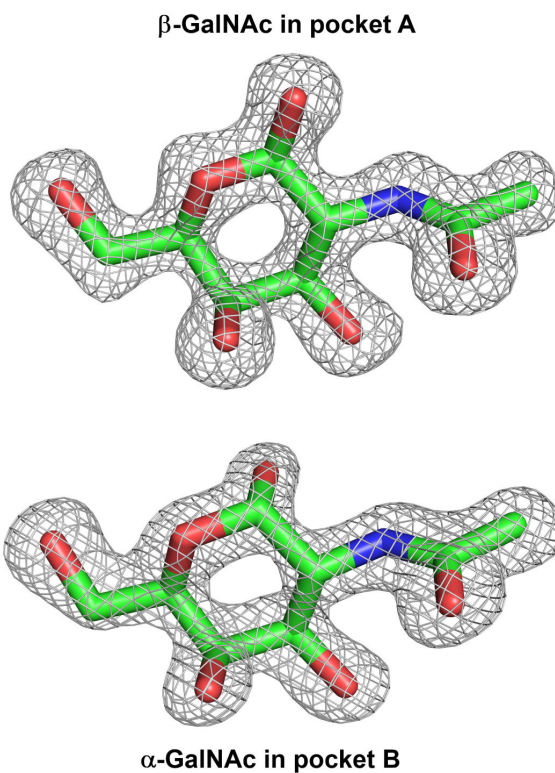
Figure 7.
Structures and associated difference electron densities, 2Fo-Fc, of the GalNAc molecules
identified in carbohydrate-binding pockets of two BfL monomers for the refined model of
BfL-GalNAc. The electron density maps are contoured at 1.5σ level and allow for an
unambiguous identification of the specific GalNAc anomer.
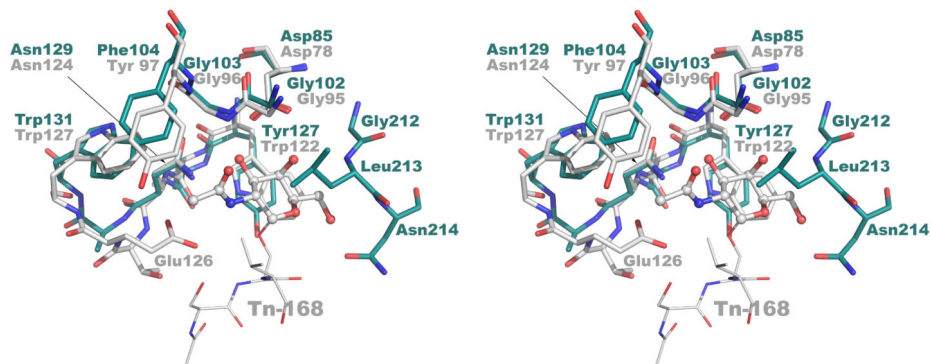
Figure 8.
A stereoview of the carbohydrate-binding sites in BfL (colored gray) and in a homologous lectin VVLB4 (green). The comparison is based on the coordinates of the complex of BfL with Tn-168 (monomer A, orthorhombic crystals) and the structure of VVLB4 with Tn-Ser ligand (PDB code 1N47, monomer A). Both pockets are superimposed based on positions of all atoms in the GalNAc residue; however, only the BfL ligand, Tn-168, is shown in this figure. The protein residues in both lectins are depicted as thick sticks. The α-D-GalNAc residue, common to all Tn antigens, is drawn in ball-and-stick representation, while the peptidic part of the ligand is shown as thin sticks. The loop contributing residues Gly212, Leu213, Asn214 in VVLB4, shown at the right hand side of the figure, is significantly shorter in BfL (not visible here).

Table 1

Representative apparent K_d values for binding to carbohydrate arrays.

ID[&]	Family	Name	RFU$_{max}$	App K_d [nM]	±SE %
5	GalNAc-a	GalNAc-α - 22	17695	0.52	0.08
120	Tn peptide-low	Ac-S-Tn(Ser)-S-G - 04	13903	0.75	0.25
146	Tn peptide-low	Ac-S-Tn(Thr)-S-G - 04	16067	0.71	0.14
147	Tn peptide-low	Ac-A-Tn(Thr)-S-G - 05	16892	0.63	0.15
415	Tn peptide-low	Ac-A-P-G-S-Tn(Thr)-A-P-P-A-G-03	16984	0.65	0.16
79	Tn peptide	Ac-S-Tn(Ser)-S-G - 22	18227	0.65	0.11
499	Tn peptide	Ac-S-Tn(Ser)-V-G-13	17255	0.73	0.21
149	Tn peptide	Ac-A-Tn(Thr)-S-G - 23	17699	0.47	0.08
173	Tn peptide	Ac-P-Tn(Thr)-T-G - 22	17616	0.59	0.09
183	Tn peptide	Ac-S-Tn(Thr)-V-G - 22	17378	0.71	0.14
155	Tn peptide	Ac-T-Tn(Thr)-P-G - 21	18371	0.69	0.12
146	Tn peptide	Ac-V-Tn(Thr)-S-G - 19	18298	0.52	0.15
440	Tn peptide	AzHex-P-D-Tn(Thr)-R-P-NH$_2$-07	15559	1.25	0.52
397	Tn peptide	Muc1-Tn15 [-G-V-T-S-A-P-D-T-R-P-A-P-G-S- Tn(Thr)-A-P-P-A]	17917	0.64	0.14
74	Tn peptide	Ac-Tn(Ser)-Tn(Ser)-Tn(Ser)-G - 16	17650	0.52	0.10
176	Tn peptide	Ac-Tn(Thr)-Tn(Thr)-Tn(Thr)-G - 20	16853	0.51	0.09
372	glycoprotein	OSM (asialo)	22744	0.72	0.47
385	glycoprotein	OSM	17375	10.65	1.13
382	Core-5 peptide	Ac-S-Ser(core 5)-S-G-18	18580	1.02	0.20
341	F1a peptide	Ac-S-Thr(F1α)-S-G - 18	17214	2.51	0.18
44	Blood Group A	BG-A -19	15700	0.87	0.10
68	Blood Group A	BG-A1-12	10151	3.23	1.05
70	Blood Group A	BG-A2-16	10548	2.68	0.91
18	GalNAc-a	GalNAcα1-3Gal - 17	17826	0.50	0.15
17	GalNAc-a	GalNAcα1-6Galβ - 22	19055	0.65	0.13
108	GalNAc-a	Forssman Di - 21	18023	0.92	0.15
289	GalNAc-a	Forssman Tetra-BSA - 13	18518	2.10	0.42
9	GalNAc-b	GalNAc-β - 21	18102	0.54	0.13

ID[&]	Family	Name	RFU$_{max}$	App K$_d$ [nM]	±SE%
234	GalNAc-b	LDN-Sp - 14	18291	0.70	0.13
124	GalNAc-b	GA2di - 37	17241	0.56	0.12
81	GalNAc-b	Gb4 - 09	14003	0.68	0.18

[&] ID refers to the position of this entry in the **Supplementary Table 1**.

% SE denotes the the standard error

Table 2

Diffraction data collection and processing.

Diffraction source Wavelength (Å) Temperature (K) Detector	Beamline 22-ID, SER-CAT, APS, ANL, IL, USA 1.000 100 Rayonix 300HS high speed CCD detector			
	BfL(LF) [#]	**BfL(GalNAc)**	**BfL(Gb4)**	**BfL(BG-A)**
Crystal-to-detector dist. (mm)	175	230	180	220
Rotation range per image (°)	0.25	0.25	0.5	0.5
Total rotation range (°)	125	125	360	200
Exposure time per image (s)	0.25	0.25	0.75	0.75
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
a, b, c (Å)	45.99 58.00 190.6	45.89 58.95 187.3	44.68 88.23 110.8	78.83 44.82 65.77
α, β, γ (°)	90 90 90	90 90 90	90 90 90	90 100.05 90
Mosaicity (°)	0.19	0.87	0.25	0.39
Resolution range (Å)	50 - 1.43 (1.45 - 1.43) [*]	50 - 1.71 (1.74 - 1.71)	50 - 1.43 (1.45 - 1.43)	50 - 1.65 (1.68 - 1.65)
Total No. of reflections	426,074	232,093	1,022,165	178,921
No. of unique reflections	90,961 (3002)	55,964 (2385)	79,081 (2287)	52,931 (2035)
Completeness (%)	95.5 (63.9)	99.0 (86.3)	96.5 (56.3)	96.9 (74.3)
Multiplicity	4.7 (2.5)	4.1 (2.4)	12.9 (4.7)	3.4 (2.2)
$<I/\sigma(I)>$	23.2 (2.4)	11.9 (3.1)	36.8 (2.7)	15.7 (2.4)
R_{merge} [†]	0.040 (0.443)	0.111 (0.367)	0.046 (0.436)	0.0543 (0.300)
$R_{p.i.m}$ [‡]	0.029 (0.327)	0.053 (0.237)	0.019 (0.213)	0.039 (0.230)
Overall B, Wilson plot (Å²)	14.4	13.0	11.0	10.6
$CC_{1/2}$ [&]	0.940 (0.759)	0.938 (0.815)	0.980 (0.862)	0.964 (0.885)

	BfL(Tn-168)	**BfL(Tn-168)**	**BfL(Tn-207)**	**BfL(Tn-200)**
Crystal-to-detector dist. (mm)	128	160	220	260
Rotation range per image (°)	0.25	1.0	0.5	0.5
Total rotation range (°)	180	400	180	250
Exposure time per image (s)	0.25	1.05	0.5	0.5

	Beamline 22-ID, SER-CAT, APS, ANL, IL, USA			
Diffraction source Wavelength (Å) Temperature (K) Detector	1.000 100 Rayonix 300HS high speed CCD detector			
Space group	P2₁2₁2₁	P2₁	P2₁2₁2₁	P2₁2₁2₁
a, b, c (Å)	44.86 88.50 110.8	54.38 46.16 92.11	44.69 88.46 111.	65.08 187.3 258.4
α, β, γ (°)	90 90 90	90 92.56 90	90 90 90	90 90 90
Mosaicity (°)	0.38	0.56	0.80	0.44
Resolution range (Å)	50 - 1.17 (1.19 - 1.17)	40 - 1.35 (1.37 - 1.35)	50 - 1.66 (1.69 - 1.66)	50 - 2.70 (2.80 - 2.70)
Total No. of reflections	1,008,306	793,282	366,225	489,032
No. of unique reflections	147,131 (5701)	100,009 (4951)	52,813 (2432)	85,387 (3781)
Completeness (%)	98.5 (77.1)	99.9 (99.6)	99.7 (94.2)	98.4 (88.8)
Multiplicity	6.9 (4.2)	7.9 (5.7)	6.9 (4.0)	5.7 (5.2)
$\langle I/\sigma(I)\rangle$	17.3 (2.9)	27.6 (2.4)	28.5 (2.0)	14.5 (1.8)
R_{merge} [†]	0.038 (0.426)	0.066 (0.751)	0.051 (0.568)	0.075 (0.660)
$R_{\text{p.i.m}}$ [‡]	0.032 (0.237)	0.027 (0.322)	0.026 (0.292)	0.050 (0.397)
Overall B, Wilson plot (Å²)	13.6	14.4	14.2	32.3
$CC_{1/2}$ [&]	0.966 (0.810)	0.961 (0.761)	0.962 (0.741)	0.966 (0.814)

[#] BfL(LF) refers to the 'ligand-free' structure of BfL.

[*] Values in parentheses are for the highest resolution shell.

[†] $R_{\text{merge}} = \Sigma I(|I - \langle I\rangle|) / \Sigma I(I)$.

[‡] Estimated $R_{\text{p.i.m.}} = R_{\text{merge}}[N/(N-1)]^{1/2}$, where N is the multiplicity of data.

[&] $CC_{1/2} = \Sigma(x - \langle x\rangle)(y - \langle y\rangle) / [\Sigma(x - \langle x\rangle)^2 \Sigma (y - \langle y\rangle)^2]^{1/2}$.

[#] For more extensive definitions of these indicators see http://shelx.uni-ac.gwdg.de/~athorn/pdf/thorn_cshl2014_quality_indicators.pdf and the references cited therein.

Table 3

Coordination distances for two divalent ion binding sites in BfL. Each distance is an average of 14 observations, two in each of the seven high-resolution structures of BfL.

	Mean ± $\sigma^{\&}$, range [Å]
Ca1 … His 134(NE2)	2.37 ± 0.079, 2.25 − 2.50
Ca1 … Asp 120(OD2)	2.20 ± 0.055, 2.10 − 2.30
Ca1 … Glu 129(OE1)	2.24 ± 0.073, 2.15 − 2.34
Ca1 … Glu 118(OE2)	2.24 ± 0.029, 2.19 − 2.29
Ca1 … Wat 10(O)	2.28 ± 0.053, 2.24 − 2.40
Ca1 … Wat 11(O)	2.24 ± 0.082, 2.11 − 2.44
Ca2 … Asn 124(OD1)	2.33 ± 0.026, 2.27 − 2.37
Ca2 … Trp 122(O)	2.33 ± 0.031, 2.25 − 2.38
Ca2 … Glu 129(OE2)	2.44 ± 0.027, 2.39 − 2.47
Ca2 … Asp 120(OD1)	2.48 ± 0.037, 2.44 − 2.56
Ca2 … Asp 120(OD2)	2.44 ± 0.035, 2.35 − 2.49
Ca2 … Wat 12(O)	2.41 ± 0.045, 2.27 − 2.47
Ca2 … Wat 13(O)	2.39 ± 0.031, 2.34 − 2.43

$\&$ σ – standard deviation

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Crystallization conditions for BfL complexes.

Ligand	Composition of the precipitant solution
L305 Tn-168(o) * Tn-207	25% (w/v) PEG1500, 0.1 M Bis–Tris Propane pH 9.0, 0.1 M NaCl
Gb4	28% (w/v) PEG6000, 0.1 M Tris pH 8.5, 0.5 M lithium chloride
Tn-168(m) *	11% (w/v) PEG8000, 0.1 M Hepes pH 7.5, 8% (v/v) ethylene glycol
Tn-200	24% (w/v) PEG1500, 0.1 M Hepes pH 7.5, 0.2 M L-proline

*
Letters 'o' and 'm', in parenthesis, refer to the orthorhombic and monoclinic crystals, respectively.

Table 5

Refinement statistics for BfL and its complexes.

	BfL(LF)	BfL(GalNAc)	BfL(Gb4)	BfL(BG-A)
Resolution range (Å)	30.0 - 1.43 (1.47 -1.43) &	50.0 - 1.70 (1.75 - 1.70)	50.0 - 1.43 (1.47 - 1.43)	50.0 - 1.65 (1.69 - 1.65)
Completeness (%)	89.8 (44.7)	97.2 (77.9)	93.3 (57.5)	93.9 (60.3)
σ-cutoff	none	none	none	none
Reflections, working set	81,670 (2977)	53,241 (3072)	71,808 (3237)	48,757 (2297)
Reflections, test set	4257 (136)	1702 (100)	3713 (163)	2601 (124)
Final R_{cryst}	0.116 (0.252)	0.153 (0.177)	0.105 (0.147)	0.150 (0.230)
Final R_{free}	0.161 (0.309)	0.186 (0.243)	0.108 (0.185)	0.177 (0.258)
No. of non-H atoms				
Protein	3732	3678	3681	3650
Ion	4	4	5	4
Ligand	32	54	121	88
Water	395	359	406	373
Total	4163	4095	4213	4115
R.m.s. deviations				
Bonds (Å)	0.021	0.023	0.020	0.022
Angles (°)	1.986	2.102	1.986	2.016
Average B-factors (Å²)				
Protein	16.3	13.6	16.0	15.4
Ion	12.0	9.5	12.8	9.8
Ligand	28.9	18.9	27.8	21.5
Water	30.3	23.5	33.6	25.7
Ramachandran plot [*]				
Favored regions (%)	98.0	97.0	96.0	98.0
Additionally allowed (%)	2.0	3.0	4.0	2.0
PDB code	5T50	5T52	5T55	5T54
	BfL(Tn-168)		**BfL(Tn207)**	**BfL(Tn200)**

	BfL(LF)	BfL(GalNAc)	BfL(Gb4)	BfL(BG-A)
Resolution range (Å)	50.0 - 1.17 (1.20 -1.17)	40.0 - 1.35 (1.38 -1.35)	50.0 - 1.66 (1.70 - 1.66)	50.0 - 2.70 (2.77 - 2.70)
Completeness (%)	97.2 (71.4)	96.2 (64.4)	97.7 (73.9)	81.2 (38.6)
σ-cutoff	none	none	none	none
Reflections, working set	139,485 (7508)	93,776 (4617)	49,321 (2675)	67,604 (2338)
Reflections, test set	4270 (208)	2927 (142)	2498 (148)	3493 (101)
Final R_{cryst}	0.107 (0.195)	0.105 (0.137)	0.129 (0.177)	0.201 (0.322)
Final R_{free}	0.130 (0.228)	0.141 (0.296)	0.163 (0.212)	0.230 (0.376)
No. of non-H atoms				
Protein	3726	3698	3664	18110
Ion	5	4	5	20
Ligand	120	106	99	460
Water	486	487	420	213
Total	4337	4295	4187	18803
R.m.s. deviations				
Bonds (Å)	0.019	0.021	0.022	0.019
Angles (°)	1.985	2.135	2.054	1.854
Average B-factors (Å²)				
Protein	18.9	17.0	14.8	59.3
Ion	13.5	11.9	10.6	47.1
Ligand	27.6	26.7	24.0	56.2
Water	35.9	35.6	26.9	39.7
Ramachandran plot *				
Favored regions (%)	98.0	98.0	98.0	98.0
Additionally allowed (%)	2.0	2.0	2.0	2.0
PDB code	5T5L	5T5J	5T5P	5T5O

a Values in parentheses are for the highest resolution shell.

* Values calculated with the program MolProbity (http://molprobity.biochem.duke.edu/) [50].