Structure and Carbohydrate Recognition by the Nonmitogenic Lectin Horcolin

Vaishali Narayanan,‡ Kishore Babu Bobbili,‡ Nukathoti Sivaji, Nisha G. Jayaprakash, Kaza Suguna,⁎ Avadhesha Surolia,⁎ and Ashok Sekhar*  

ABSTRACT: Lectins are sugar-binding proteins that have shown considerable promise as antiviral agents because of their ability to interact with envelope glycoproteins present on the surface of viruses such as HIV-1. However, their therapeutic potential has been compromised by their mitogenicity that stimulates uncontrolled division of T-lymphocytes. Horcolin, a member of the jacalin family of lectins, tightly binds the HIV-1 envelope glycoprotein gp120 and neutralizes HIV-1 particles but is non-mitogenic. In this report, we combine X-ray crystallography and NMR spectroscopy to obtain atomic-resolution insights into the structure of horcolin and the molecular basis for its carbohydrate recognition. Each protomer of the horcolin dimer adopts a canonical β-prism I fold with three Greek key motifs and carries two carbohydrate-binding sites. The carbohydrate molecule binds in a negatively charged pocket and is stabilized by backbone and side chain hydrogen bonds to conserved residues in the ligand-binding loop. NMR titrations reveal a two-site binding mode and equilibrium dissociation constants for the two binding sites determined from two-dimensional (2D) lineshape modeling are 4-fold different. Single-binding-site variants of horcolin confirm the dichotomy in binding sites and suggest that there is allosteric communication between the two sites. An analysis of the horcolin structure shows a network of hydrogen bonds linking the two carbohydrate-binding sites directly and through a secondary binding site, and this coupling between the two sites is expected to assume importance in the interaction of horcolin with high-mannose glycans found on viral envelope glycoproteins.

INTRODUCTION

Lectins are a ubiquitous class of proteins found in bacteria, viruses, plants, and animals that bind to carbohydrates without enzymatically modifying them.1 Lectins interact with mono and oligosaccharides found on viral and cell surface glycoproteins through their carbohydrate recognition domains, where the lectin–glycan complex is stabilized via hydrogen bonds, van der Waals contacts, hydrophobic interactions, and metal coordination linkages.1−4 Considering the importance of protein–carbohydrate interactions in several biological processes such as receptor-mediated endocytosis, immune response, signal transduction, viral replication, and cell–cell adhesion, lectins have been extensively explored for their therapeutic use over the past few decades.5−13

The mannose-specific jacalin-related lectin (mJRL) family recognizes and binds terminal as well as internal mannose residues in glycoproteins. They have a typical β-prism topology containing four-stranded Greek key motifs that form each face of the prism and harbor up to three sugar-binding sites per protomer.14−18 Members of the mJRL family such as jacalin and banana lectin (BanLec) associate with high-mannose N-glycans on the HIV envelope glycoprotein and block viral entry.19−22 However, their therapeutic potential has been marred by their potent lymphocyte mitogenic activity15,24 that is believed to result from a direct interaction between the lectin and the lymphocyte cell surface glycoprotein CD4.25 Such uncontrolled stimulation of mitosis in T cells not only causes systemic inflammation24 but may actually result in increased viral transmission.

BanLec has two carbohydrate-binding sites per protomer and binds high-mannose glycans in a bidentate mode,19,26 where both sites engage with the same glycan molecule. The mitogenicity and antiviral activity of BanLec are decoupled by the H84T mutation, which disrupts π−π stacking between Y83 and H84, making the two glycan-binding sites independent and incapable of bidentate ligand binding.26,27 Though H84T BanLec is not mitogenic, there are conflicting reports on its ability to neutralize HIV particles, with IC50 values varying between 1 and 46 nM [wild type (wt) IC50 = 0.9 nM].21,26 It is likely that high-mannose chains on the HIV gp120 envelope glycoprotein also need to bind in bidentate fashion for effective inhibition of the virus.28 In support of this

Received: December 6, 2021  
Revised: February 3, 2022  
Published: February 28, 2022
hypothesis, tetrameric BanLec variants, which have eight glycan-binding sites, have 5–77-fold lower IC$_{50}$ values against HIV strains compared to dimeric mutants of BanLec, which have only four.\textsuperscript{21}

Horcolin is a dimeric lectin from *Hordeum vulgare* (barley) belonging to the mJRL family. It is made up of 146 amino acids in each of its protomers. Isothermal titration calorimetry studies have shown that it exhibits monovalent binding to mannose, mannobiose, and mannotriose, while higher mannoooligosaccharides with five, seven, and nine mannose units interact with horcolin in a bivalent mode.\textsuperscript{22} The binding between horcolin and recombinant gp120 from the HIV-1\textsubscript{YU2} strain is very tight, with an equilibrium dissociation constant ($K_a$) of 14.4 nM, and horcolin also has the ability to neutralize HIV-1. Crucially, horcolin does not cause cell cycling in splenocytes, indicating that it does not have detectable mitogenic activity in contrast to some of its other family members such as jacalin and BanLec.\textsuperscript{29}

In this report, we provide high-resolution crystal structures of apo- and methyl-α,D-mannopyranoside-bound horcolin. These structures demonstrate horcolin to be a dimeric protein with a subunit structure resembling other β-prism I fold lectins and carrying two carbohydrate-binding sites (CBS1 and CBS2). We then use multidimensional NMR spectroscopy in conjunction with isotope labeling and perdeuteration to probe the thermodynamics and kinetics of mannose recognition. NMR two-dimensional (2D) lineshape analysis reveals that the two binding sites in horcolin have 4-fold different affinities for mannose. The dichotomy in the ligand-binding sites is confirmed by the use of single-binding-site variants of horcolin, which also suggest an element of allosterity between CBS1 and CBS2.

## EXPERIMENTAL PROCEDURES

**Protein Expression and Purification.** The codon-optimized DNA sequence for the horcolin gene (Uniprot ID: Q5U9T2), either carrying or lacking a C-terminal hexa-His purification tag, was subcloned into the pET-22b(+) vector between the *Nde*I and *Xho*I restriction sites. To generate D39A and D138A horcolin mutants, site-directed mutagenesis was carried out using the Quikchange method with the help of overlapping primers. Additionally, a noncleavable C-terminal 6x-His tag was inserted in the mutant horcolin for purification purposes.

Horcolin was overexpressed in BL21(DE3) cells transformed with the pET-22b(+) plasmid containing the horcolin gene insert. Cells were grown either in Luria-Bertani (LB) broth (unlabeled horcolin) or in M9 minimal media (15N-, 13C-, or 2H-labeled horcolin); 1 g/L 15NH$_4$Cl and 3 g/L [1H,13C] glucose were added to the M9 media as the sole carbon sources, respectively. When the cell culture reached an OD$_{600}$ of 0.8, overexpression of horcolin was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was incubated at 18 °C for 18–20 h. Subsequently, the cells were harvested by centrifugation at 7000 rpm for 20 min at 4 °C and washed with 1X phosphate buffer saline (PBS). The cells were then resuspended in lysis buffer (pH 7.4) containing 1× PBS and 1 mM phenylmethylsulfonyl fluoride (PMFS), sonicated on ice for 30 min (2 s on and 4 s off cycles) at 50% amplitude, and centrifuged twice at 14 000 rpm for 45 min at 4 °C.

For wt horcolin without a hexa-His tag, the supernatant was incubated with 5 mL of mannose–sepharose beads and allowed to bind on an end-rocker for 2 h at 4 °C. The beads were then loaded on a column and washed with 1 L of 1X PBS. Horcolin was eluted in fractions of 2 mL with a 100–300 mM gradient of D-mannose spanning 10 column volumes. Pure protein fractions were dialyzed against 1X PBS and stored at 4 °C. For His-tagged wt horcolin or the single-binding-site mutants, the supernatant after lysis and centrifugation was loaded onto a Ni-NTA column pre-equilibrated with the lysis buffer. After washing the column with 1X PBS buffer containing 20 mM imidazole, the pure protein was eluted with buffer (pH 7.4) containing 50 mM Tris, 200 mM NaCl, and 250 mM imidazole. Pure protein fractions were dialyzed against buffer (pH 7.4) containing 25 mM Tris and 50 mM NaCl to remove the imidazole. Back-exchange of deuterated amides in U-2H,13C,15N-labeled horcolin was carried out by denaturing perdeuterated horcolin in 1X PBS containing 6M guanidine hydrochloride (GdmCl). The protein was then refolded by step-wise dialysis using successive GdmCl concentrations of 6, 3, 1.5, 0.75, and 0 M.

The purity of the protein was checked using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on a 15% Tris-glycine gel. NMR samples were made in 10% (vol/vol) D$_2$O buffer (pH 7.4) containing 25 mM Tris, 50 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 0.03% NaN$_3$. The protein yield was typically 7 mg/L cell culture in both LB and M9 media.

All experiments involving a carbohydrate reported in this manuscript have been carried out using methyl-α-D-mannopyranoside, which has been referred to as mannose throughout the manuscript.

Wt horcolin lacking the C-terminal hexa-His tag was used for all crystallization and NMR studies. Since the single-binding-site variants did not bind mannose strongly enough for efficient purification, a hexa-His tag was used for D39A and D138A horcolin. NMR spectra of wt horcolin without and with the hexa-His tag show only minor chemical shift perturbations (CSP), confirming that the C-terminal hexa-His tag does not induce structural perturbations in horcolin (Figure S10A). Moreover, titration data acquired with His-tagged and tagless horcolin constructs agree very well, demonstrating that the purification tag does not alter mannose-binding characteristics of horcolin (Figure S10B).

**Crystallization and Data Collection.** Crystallization experiments involving apo-horcolin were carried out employing the microbatch under the oil method using a 10 mg/mL protein solution in PBS buffer pH 7.4 (20 mM PBS and 150 mM NaCl). Crystals were obtained in the Crystal Screen I crystallization condition No. 10 from Hampton Research consisting of 0.2 M ammonium acetate, 0.1 M sodium acetate trihydrate pH 4.6, and 30% (w/v) polyethylene glycol 4000. The crystals of apo-horcolin grew to an average size of 0.3 × 0.3 × 0.3 mm$^3$ over a course of 10 days at 22 °C in the orthorhombic space group P2$_2$1$_2$1 with 12 molecules in the asymmetric unit (ASU). Crystals of methyl-α,D-mannopyranoside-bound horcolin were obtained by cocrystallization using the microbatch under the oil method; 300 μL of the protein solution at a concentration of 10 mg/mL in Tris buffer (pH
maps. Ligand molecules were modeled into the electron map and their positions were con

NMRPipe and NMRFAM-Sparky software packages, and 2.18 Å /Da and 44% solvent content with a single subunit/ASU for apo-horcolin and 2.18 Å /Da and 35% solvent content with 12 subunits/six dimers in the ASU. Apo-horcolin subunit coordinates were used as a search model for the complex involving methyl-α-D-mannopyranoside. Refinement cycles were performed with REFMACS. A few rounds of rigid body refinement were performed first, followed by positional and B-factor refinement. Model building was done using Coot, Tight noncrystallographic symmetry (NCS) was employed during the initial stages of refinement and progressively relaxed during the final stages. Initially, ligand molecules were located/identified on the basis of the difference map and their positions were confirmed using simulated omit maps. Ligand molecules were modeled into the electron density where appropriate when the R-factors were approxi-mately 0.23. Oxygen atoms of water molecules were identified on the basis of peaks with heights >3σ in Fo−Fc maps and greater than 1σ in 2Fo−Fc maps. The structures were validated using PROCHECK and MolProbity. A summary of the relevant statistics of the data collection and refinement is given in Table 1.

### Analysis of Structures

**PyMol** was used for structure analysis and generating figures. Interatomic and hydrogen bond distances were calculated using CONTACT in the CCP4 program suite or **PyMol**. Structural superpositions were made using PyMol or ALIGN software. The buried surface area was calculated by employing the PISA software package or PyMol.

**PDB References**

Apo-horcolin (7V4Z)

Horcolin in complex with methyl-α-D-mannopyranoside (7V4S)

### NMR Spectroscopy

NMR spectra were acquired on 14.1 T Agilent (600 MHz 1H Larmor frequency) or 16.5 T Bruker (700 MHz) spectrometers equipped with a 1H/13C/15N triple resonance single-axis gradient cryoprobe (600 MHz spectrometer) or a room-temperature TXI probe (700 MHz Bruker spectrometer). All NMR measurements were carried out at 25 °C. NMR data sets were processed and visualized using NMRPipe and NMRFAM-Sparky software packages, respectively.

**Backbone Assignments**

Backbone 1H, 15N, 13Ca, 13Cβ, 13CO, and 15N resonances of horcolin molecules were assigned experiments carried out on samples of 2.2 mM U, 15N, 0.4 mM U, 13C, 15N or 0.6 mM U, 2H, 13C, 15N horcolin (after back-exchanging amide deuterons to protons). A combination of the following data sets was used for assignment purposes: (U,2H,13C,15N horcolin): HNCACB, HN(CO)CACB, HNCO, HN(CA)CO, HNN. (U,13C,15N horcolin): HNCA, HN(CO)CA, HBHA(CO)NH, HACA(CO)NH. (U,15N horcolin): HNN, 15N-edited nuclear Overhauser enhancement spectroscopy-heteronuclear single quantum coherence spectroscopy (NOESY-HSQC) (140 ms mixing time), and 1H−1H NOESY (150 ms mixing time). Backbone 1H, 15N assignments of D39A and D138A horcolin were transferred from the assignments of wt horcolin using a combination of 15N-edited NOESY-HSQC (150 ms mixing time) and 15N-edited total correlation spectroscopy (TOCSY)-HSQC (53.5 ms mixing time) data sets.

<table>
<thead>
<tr>
<th>Table 1. Data Collection and Refinement Statistics*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>wavelength (Å)</td>
</tr>
<tr>
<td>resolution range (Å)</td>
</tr>
<tr>
<td>space group</td>
</tr>
<tr>
<td>a (Å)</td>
</tr>
<tr>
<td>b (Å)</td>
</tr>
<tr>
<td>c (Å)</td>
</tr>
<tr>
<td>α, β, γ (deg)</td>
</tr>
<tr>
<td>total reflections</td>
</tr>
<tr>
<td>unique reflections</td>
</tr>
<tr>
<td>multiplicity</td>
</tr>
<tr>
<td>completeness (%)</td>
</tr>
<tr>
<td>&quot;I/σ(i)&quot;</td>
</tr>
<tr>
<td>Wilson B-factor (Å)</td>
</tr>
<tr>
<td>reflections used in refinement</td>
</tr>
<tr>
<td>reflections used for Rmerge</td>
</tr>
<tr>
<td>Nوك</td>
</tr>
<tr>
<td>Herr</td>
</tr>
<tr>
<td>CC/1/2</td>
</tr>
<tr>
<td>CC/3</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
</tr>
<tr>
<td>average B-factor (Å)</td>
</tr>
</tbody>
</table>

*Statistics for the highest-resolution shell is shown in parentheses.
To determine the combined and weighted backbone amide chemical shift perturbation (CSP) upon addition of mannose, we used the following equation

\[ \Delta \delta = \sqrt{\Delta H^2 + (0.159 \Delta N)^2} \]

where \( \Delta H \) and \( \Delta N \) are the residue-specific differences in \(^1\)H and \(^{15}\)N chemical shifts extracted from \(^1\)H-\(^{15}\)N HSQC spectra at the beginning and the end of the titration, respectively.

**Fitting and Simulating NMR Titration Profiles.** Two-dimensional lineshape modeling of the titration data was performed within the software package TITAN,\(^{16}\) which numerically simulates, in Liouville space, the evolution of magnetization during a pulse sequence in the presence of chemical exchange. All \(^1\)H-\(^{15}\)N-HSQC spectra of a titration series were processed identically using an exponential window function in both dimensions (4 and 8 Hz line broadening in \(^1\)H and \(^{15}\)N, respectively). Residues that showed a chemical shift perturbation larger than 0.1 ppm upon mannose binding, and whose peaks were not severely overlapped with neighboring resonances, were chosen for lineshape analysis. For the titration of D138A and D39A horcolin, lineshapes of 14 and 16 residues, respectively, were globally fit to a two-state single-site binding model. The following 14 residues were used

(0.526), 19 (0.523), 26 (0.520), 36 (0.515), 51 (0.507), 64 (0.5), 100 (0.481), 150 (0.454), 300 (0.374), 400 (0.321), and 500 (0.267) mM (22 points). In the D138A horcolin titration series, the initial protein concentration was 0.79 mM and the total ligand (protein) concentrations were 0 (0.79), 0.2 (0.784), 0.4 (0.783), 0.8 (0.781), 2 (0.78), 3 (0.778), 6 (0.776), 9 (0.773), 12 (0.77), 16 (0.768), 19 (0.766), 22 (0.763), 28 (0.758), 39 (0.75), 55 (0.737), 78 (0.719), 98 (0.704), and 167 (0.658) mM (18 points). In the D39A mutant titration series, the initial protein concentration was 0.63 mM and the total ligand (protein) concentrations were 0 (0.627), 0.2 (0.623), 0.3 (0.622), 0.6 (0.62), 1 (0.619), 2 (0.618), 5 (0.616), 7 (0.615), 10 (0.613), 12 (0.612), 15 (0.61), 17 (0.609), 22 (0.605), 31 (0.6), 44 (0.593), 63 (0.581), 78 (0.571), 128 (0.540), 274 (0.450), 456 (0.337), and 654 (0.214) mM (21 points).

Figure 1. Crystal structure of apo-horcolin. (A) The domain organization of horcolin, showing the three Greek key motifs in green, blue, and purple colors and the residues at the dimer interface in orange. (B) The subunit structure perpendicular to the pseudo-three-fold axis. The three Greek keys are depicted in the same colors as in panel (A). (C) The dimeric structure of one of the six crystallographically independent dimers of apo-horcolin (left) and rotated by approximately 90° about a horizontal axis (right). The Greek keys have the same colors as in panels (A) and (B). The C-termini of the protomers for the dimer orientation on the right-hand side are hidden below the magenta Greek keys and their approximate locations are indicated by cyan lines. (D) Cartoon representation of the horcolin dimer (gray) with residues burying more than 15 Å² surface area shown as colored spheres (green: carbon, red: oxygen, blue: nitrogen). Amino acid labels in black and orange refer to residues in different protomers of the horcolin dimer.
in the fitting of D138A titration series: V5, M30, G61, E66, E79, V84, F87, D90, V93, R106, G113, G134, F136, and one unassigned residue. The following 16 residues were used in the fitting of D39A titration series: V5, G15, M30, F43, N50, G60, G61, G64, E66, F87, T100, G131, G134, F136, and two unassigned residues. The wt titration data set was fit to a three-state sequential model. The chemical shifts of all three states were fit first and then fixed for the rest of the fitting procedure. Seventeen residues (G14, G15, A40, I41, S55, G56, G60, G61, G64, E66, E79, F87, D90, T107, T112, A133, and one unassigned residue) were globally fit in the wt titration series. The errors in the fit parameters were determined by bootstrapping and are obtained as the standard deviation from the mean of 100 bootstrap replicas. Simulations of peak lineshapes were carried out within TITAN software using parameters derived from the lineshape fitting routine described above.

**RESULTS**

The overall structure of apo-horcolin. As the first step toward understanding the nonmitogenic horcolin at the molecular level, we solved its X-ray crystallographic structure in the ligand-free (apo) state at a resolution of 1.16 Å (Figure 1 and Table 1). Apo-horcolin has previously been shown to be a dimer in solution. It crystallizes in the P2_1_2_1 space group with 12 molecules (or six biological dimers) in the asymmetric unit (ASU). The dimeric arrangement is robust and the six crystallographically independent dimers in each ASU superpose with a pairwise root-mean-square deviation (RMSD) in Cα positions ranging from 0.19 to 0.60 Å. Each molecule contains 12 β-strands arranged in the form of three antiparallel β-sheets (Figure 1A,B). These three Greek key motifs come together to form a β-prism I fold with pseudo-three-fold symmetry characteristic of several highly conserved carbohydrate-binding proteins. The core of this fold is stabilized by interactions between conserved hydrophobic residues such as L96, F116, I126, F129, and I140.

The intersubunit interactions at the dimer interface originate primarily from residues located in β1 (K3–W10) and β10 (G113–V120) and result in the burial of 1600 Å² surface area (800 Å² per protomer, Figure 1C,D). The two β-strands β1 and β10 from different subunits of the dimer pack in an antiparallel manner (Figure 1C). Backbone hydrogen bonds between K6 (CO) and L121 (N), and G8 (N) and P119 (CO), as well as an interesting side chain–backbone hydrogen bond between W10 (Ne1) and R117 (CO), stabilize the dimer interface. The packing of hydrophobic residues such as W10, I114, and L121 at the intersubunit interface also contributes to the dimerization free energy (Figure 1D).

The crystal structure of methyl-α-D-mannopyranoside-bound horcolin. (A) Overlay of the cartoon representations of apo (yellow) and mannose-bound (blue) horcolin. Mannose molecules are shown in sphere representation and colored according to atom type. (B) Electron density for mannose at CBS1 (top) and CBS2 (bottom) in the simulated annealing F̅F̅ omit map. The map is contoured at the 5σ level. (C) A zoomed-in view of the sugar-binding site, showing the elements of the structure important for mannose recognition: LBL1 (red), LBL2 (blue), GG loops (green), and the secondary binding site (cyan). The conserved D39 and D138 residues whose side chains form hydrogen bonds with the bound mannose are shown as sticks.

The errors in the fit parameters were determined by bootstrapping and are obtained as the standard deviation from the mean of 100 bootstrap replicas. Simulations of peak lineshapes were carried out within TITAN software using parameters derived from the lineshape fitting routine described above.

**RESULTS**

The overall structure of apo-horcolin. As the first step toward understanding the nonmitogenic horcolin at the molecular level, we solved its X-ray crystallographic structure in the ligand-free (apo) state at a resolution of 1.16 Å (Figure 1 and Table 1). Apo-horcolin has previously been shown to be a dimer in solution. It crystallizes in the P2_1_2_1 space group with 12 molecules (or six biological dimers) in the asymmetric unit (ASU). The dimeric arrangement is robust and the six crystallographically independent dimers in each ASU superpose with a pairwise root-mean-square deviation (RMSD) in Cα positions ranging from 0.19 to 0.60 Å. Each molecule contains 12 β-strands arranged in the form of three antiparallel β-sheets (Figure 1A,B). These three Greek key motifs come together to form a β-prism I fold with pseudo-three-fold symmetry characteristic of several highly conserved carbohydrate-binding proteins. The core of this fold is stabilized by interactions between conserved hydrophobic residues such as L96, F116, I126, F129, and I140.

The intersubunit interactions at the dimer interface originate primarily from residues located in β1 (K3–W10) and β10 (G113–V120) and result in the burial of 1600 Å² surface area (800 Å² per protomer, Figure 1C,D). The two β-strands β1 and β10 from different subunits of the dimer pack in an antiparallel manner (Figure 1C). Backbone hydrogen bonds between K6 (CO) and L121 (N), and G8 (N) and P119 (CO), as well as an interesting side chain–backbone hydrogen bond between W10 (Ne1) and R117 (CO), stabilize the dimer interface. The packing of hydrophobic residues such as W10, I114, and L121 at the intersubunit interface also contributes to the dimerization free energy (Figure 1D).

The crystal structure of methyl-α-D-mannopyranoside-bound horcolin. (A) Overlay of the cartoon representations of apo (yellow) and mannose-bound (blue) horcolin. Mannose molecules are shown in sphere representation and colored according to atom type. (B) Electron density for mannose at CBS1 (top) and CBS2 (bottom) in the simulated annealing F̅F̅ omit map. The map is contoured at the 5σ level. (C) A zoomed-in view of the sugar-binding site, showing the elements of the structure important for mannose recognition: LBL1 (red), LBL2 (blue), GG loops (green), and the secondary binding site (cyan). The conserved D39 and D138 residues whose side chains form hydrogen bonds with the bound mannose are shown as sticks.

The errors in the fit parameters were determined by bootstrapping and are obtained as the standard deviation from the mean of 100 bootstrap replicas. Simulations of peak lineshapes were carried out within TITAN software using parameters derived from the lineshape fitting routine described above.

The overall structure of apo-horcolin. As the first step toward understanding the nonmitogenic horcolin at the molecular level, we solved its X-ray crystallographic structure in the ligand-free (apo) state at a resolution of 1.16 Å (Figure 1 and Table 1). Apo-horcolin has previously been shown to be a dimer in solution. It crystallizes in the P2_1_2_1 space group with 12 molecules (or six biological dimers) in the asymmetric unit (ASU). The dimeric arrangement is robust and the six crystallographically independent dimers in each ASU superpose with a pairwise root-mean-square deviation (RMSD) in Cα positions ranging from 0.19 to 0.60 Å. Each molecule contains 12 β-strands arranged in the form of three antiparallel β-sheets (Figure 1A,B). These three Greek key motifs come together to form a β-prism I fold with pseudo-three-fold symmetry characteristic of several highly conserved carbohydrate-binding proteins. The core of this fold is stabilized by interactions between conserved hydrophobic residues such as L96, F116, I126, F129, and I140.

The intersubunit interactions at the dimer interface originate primarily from residues located in β1 (K3–W10) and β10 (G113–V120) and result in the burial of 1600 Å² surface area (800 Å² per protomer, Figure 1C,D). The two β-strands β1 and β10 from different subunits of the dimer pack in an antiparallel manner (Figure 1C). Backbone hydrogen bonds between K6 (CO) and L121 (N), and G8 (N) and P119 (CO), as well as an interesting side chain–backbone hydrogen bond between W10 (Ne1) and R117 (CO), stabilize the dimer interface. The packing of hydrophobic residues such as W10, I114, and L121 at the intersubunit interface also contributes to the dimerization free energy (Figure 1D).

The crystal structure of methyl-α-D-mannopyranoside-bound horcolin. (A) Overlay of the cartoon representations of apo (yellow) and mannose-bound (blue) horcolin. Mannose molecules are shown in sphere representation and colored according to atom type. (B) Electron density for mannose at CBS1 (top) and CBS2 (bottom) in the simulated annealing F̅F̅ omit map. The map is contoured at the 5σ level. (C) A zoomed-in view of the sugar-binding site, showing the elements of the structure important for mannose recognition: LBL1 (red), LBL2 (blue), GG loops (green), and the secondary binding site (cyan). The conserved D39 and D138 residues whose side chains form hydrogen bonds with the bound mannose are shown as sticks.

The errors in the fit parameters were determined by bootstrapping and are obtained as the standard deviation from the mean of 100 bootstrap replicas. Simulations of peak lineshapes were carried out within TITAN software using parameters derived from the lineshape fitting routine described above.
As the next step, we assessed the atomic-resolution structural
CBS1 with D90 OD1 that is absent in CBS2 (Figure 3B)
structure of the complex between the O4 atom of mannose at
a water-mediated hydrogen bond observed in the crystal
backbone atoms (Table 1 and Figure 3B,C). However, there is
primarily because most of the hydrogen bonds involve protein
(GXXXD motifs (GAFLD in CBS1 vs GAIVD in CBS2),
(Table 2).

The residues lining the binding pockets in CBS1 and CBS2
hydrogen bonds between the lectin and mannose. These
hydrogen bonds remain similar at both sites despite the
hydrogen bonds involving the O atom of Wat976, the side chain of D90, and a hydroxyl group of mannose (O4), which exists only in CBS1.

The residues lining the binding pockets in CBS1 and CBS2
generate negatively charged depressions that house the
mannose residues, with CBS1 having a larger negative
electrostatic potential than CBS2 (Figure 3A). The predom-
nant noncovalent interactions at both binding sites are
manifested in the Ramachandran map expected for
the Ramachandran map expected for
Figure 4 (A), TALOS-derived backbone dihedral angles ϕ and
ψ (Figure 4C, red circles) predominantly fall in regions of
the Ramachandran map expected for β-sheets (ϕ, ψ: −90,
120°), while a few instances of type I (ϕ, ψ (i + 1): −60, −30°;
ϕ, ψ (i + 2): −90, 0°) and type II (ϕ, ψ (i + 1): −60, 0°; ϕ, ψ
(i + 2): 80, 0°) are also found. The overall distribution of
dihedral angles in solution (Figure 4C, red circles) matches
well with ϕ/ψ angles seen in the crystal structure of apo-
horcolin (Figure 4C, yellow circles), confirming that the
backbone secondary structure is very similar in the crystalline
and solution states.

To evaluate whether the overall arrangement of β-strands in
solution agrees with the three-Greek key β-prism 1 fold of the
crystal structure, we measured 1H-15N NOEs using 15N-
edited NOEYS and 1H-15N NOEYS pulse sequences on
both protonated and perdeuterated (and back-exchanged) 15N-
labeled horcolin. Figure 4D shows a contact map where the
experimentally obtained 1H-15N NOEs are plotted as
features of apo-horcolin in solution using multidimensional NMR spectroscopy. Figures 4A and S1 show the 1H–15N HSQC spectrum of 15N-labeled apo-horcolin. The chemical shift dispersion of the lectin resonances in both the 15N and 1H dimensions is large, confirming that horcolin is folded in solution. Since horcolin is a 30 kDa dimer, perdeuteration was necessary to obtain backbone resonance assignments using a combination of HNCA,CB, HN(CO)CAB, HNCO, and HN(CA)COC three-dimensional (3D) data sets acquired on a
600 MHz NMR spectrometer (see Experimental Procedures). Chemical shift assignments were further verified using the 3D HNN experiment, as well as using i/i + 1 sequential Hα–Hα correlations in 3D 15N-edited NOEYS spectra; 130 out of the expected 138 1H–15N correlations were assigned and the overall backbone assignment (including N, Hα, Cα, Cδ, CO, and Hα atoms) is 91% complete (Table S1).

Backbone chemical shifts are excellent probes of the protein
secondary structure.49 We used the chemical shifts of horcolin
as inputs to the TALOS software package50 to determine the
backbone secondary structure of the protein in solution. The
residue-specific locations of β-strands (orange bars) and loops
cyan line) predicted by TALOS match very well with the
crystal structure, and there is no significant α-helical
conformation detected anywhere (gray line) in horcolin
(Figure 4B). TALOS-derived backbone dihedral angles ϕ and
ψ (Figure 4C, red circles) predominantly fall in regions of
the Ramachandran map expected for β-sheets (ϕ, ψ: −90,
120°), while a few instances of type I (ϕ, ψ (i + 1): −60, −30°;
ϕ, ψ (i + 2): −90, 0°) and type II (ϕ, ψ (i + 1): −60, 0°; ϕ, ψ
(i + 2): 80, 0°) are also found. The overall distribution of
dihedral angles in solution (Figure 4C, red circles) matches
well with ϕ/ψ angles seen in the crystal structure of apo-
horcolin (Figure 4C, yellow circles), confirming that the
backbone secondary structure is very similar in the crystalline
and solution states.

To evaluate whether the overall arrangement of β-strands in
solution agrees with the three-Greek key β-prism 1 fold of the
crystal structure, we measured 1H-15N NOEs using 15N-
edited NOEYS and 1H-15N NOEYS pulse sequences on
both protonated and perdeuterated (and back-exchanged) 15N-
labeled horcolin. Figure 4D shows a contact map where the
experimentally obtained 1H-15N NOEs are plotted as

Table 2. Details of the Interactions at CBS1 and CBS2 of Horcolin with Methyl-α-D-mannopyranoside

<table>
<thead>
<tr>
<th>mannose-bound sugar/water heteroatom</th>
<th>site I</th>
<th>site II</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein/water heteroatom distance (Å)</td>
<td>protein/heteroatom distance (Å)</td>
<td></td>
</tr>
<tr>
<td>O1</td>
<td>15 G N 2.8</td>
<td>64 N 3.0</td>
</tr>
<tr>
<td>O2</td>
<td>15 G N 3.2</td>
<td>64 N 3.5</td>
</tr>
<tr>
<td>O3</td>
<td>13 D138 OD2 2.7</td>
<td>39 OD2 2.6</td>
</tr>
<tr>
<td>O4</td>
<td>15 G N 3.0</td>
<td>36 N 3.0</td>
</tr>
<tr>
<td>O5</td>
<td>13 A135 N 3.4</td>
<td>36 N 3.2</td>
</tr>
<tr>
<td>O6</td>
<td>15 G N 3.1</td>
<td>36 N 3.0</td>
</tr>
<tr>
<td>F136</td>
<td>17 G135 N 3.2</td>
<td>37 N 3.2</td>
</tr>
<tr>
<td>F136</td>
<td>17 G135 O 3.2</td>
<td>37 O 3.2</td>
</tr>
<tr>
<td>D138</td>
<td>22 D138 OD1 2.7</td>
<td>39 OD1 2.7</td>
</tr>
<tr>
<td>Wat976</td>
<td>2.9</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Interactions mediating sugar recognition by horcolin. (A) A zoomed-in view of the electrostatic potential surface of mannose-bound horcolin, showing the negatively charged clefts that house the mannose ligands. Protein–sugar interactions in the complex of horcolin with mannose in CBS1 (B) and CBS2 (C), identified using a distance cutoff of 3.5 Å between the two electronegative atoms and an angular constraint of 65° using the find_pairs command within Pymol. The main difference between the interactions at the two sites is the water-mediated interaction involving the O atom of Wat976, the side chain of D90, and a hydroxyl group of mannose (O4), which exists only in CBS1.
contacts in the upper triangle. The $^{1}$HN−$^{1}$HN contacts shorter than 5 Å observed in the crystal structure of apo-horcolin are indicated in the lower triangle for comparison. β-strands appear as antidiagonal streaks in these plots. The pattern of contacts in the upper and lower triangles agrees very well, showing that the topology of the β-strands, as well as the tertiary structure of horcolin, is retained in solution.

The crystal structure of apo-horcolin shows well-defined electron density for all residues from 3 to 146, including regions of the loop connecting the 12 β-strands. The average B-factors of apo-horcolin are low (16 Å$^2$), indicating the amplitude of dynamics is low throughout horcolin in its high-resolution structure. To characterize the dynamics in the solution state, we calculated squared order parameters ($S^2$), which are measures of the amplitude of fast timescale dynamics, using chemical shift information within the TALOS program. $S^2$ values (Figure 4E) are uniformly high across the entire sequence of horcolin and range only between 0.68 and 0.95, demonstrating that apo-horcolin adopts a rigid 12-stranded β-prism I fold comprised of three Greek key motifs.

**Mannose Binding to Horcolin.** The utility of horcolin as a potential antiviral agent relies crucially on its ability to recognize mannose residues present in high-mannose glycans of glycoproteins attached to the envelopes of viruses such as HIV-1. We next used $^{1}$H−$^{15}$N HSQC-detected NMR titrations to characterize the binding of mannose to horcolin. Mannose was titrated into a solution containing 0.54 mM horcolin (initial concentration), with final mannose concentrations ranging from 0 to 500 mM. $^{1}$H−$^{15}$N HSQC spectra acquired at each mannose concentration showed significant chemical shift changes in a number of resonances including G15, G60, G61, and G64 (Figure 5A). Interestingly, peaks undergoing large chemical shift perturbations also displayed pronounced exchange broadening, with the intensity decreasing at intermediate mannose concentrations and subsequently increasing again. The HSQC correlation from G64, which shows the largest chemical shift perturbation (0.58 ppm) upon
mannose binding, disappears entirely in the sixth HSQC spectrum corresponding to a mannose concentration of 2 mM, while exhibiting intense correlations in spectra at the beginning and the end of the titration. In contrast, resonances from residues such as D38 and D139, which show small chemical shift perturbations (D39: 0.078 ppm, D138: 0.048 ppm), have uniform intensities during the titration. These observations suggest that mannose binding to horcolin occurs at the boundary of the fast-intermediate chemical shift exchange timescale, with the overall exchange broadening determined by the chemical shift difference between the exchanging states, as well as their relative populations.

To quantify the effect of mannose binding, we calculated residue-specific chemical shift perturbations using \( ^1H - ^{15}N \) HSQC spectra corresponding to the beginning (0 mM mannose) and the end (500 mM mannose) of the titration (Figure 5B,C). The overall magnitude of chemical shift changes is small, with an average value of 0.061 ± 0.090 ppm over 130 assigned residues. The largest chemical shift differences between the free and bound forms in the \( ^{15}N \) and \( ^1H \) dimensions are 2.24 and 0.545 ppm, respectively. Among residues in \( \beta \)-strands, the average CSP is 0.04 ppm, indicating that horcolin undergoes a minimal change in secondary structure upon mannose binding, consistent with the close similarity between the crystal structures of apo- and mannose-bound horcolin.

Figure 5B,C shows the chemical shift perturbations occurring in horcolin after binding mannose. All of the perturbations larger than 0.1 ppm are localized to residues present in key mannose recognition motifs identified from the crystal structures of apo- and mannose-bound horcolin, namely, the two ligand-binding loops, the two GG loops, and the secondary binding site. In addition, a number of residues, such as G15 (Figure 5A) and F136 (Figure S2A), show pronounced curvatures in their chemical shift trajectories through the titration, unequivocally demonstrating the presence of two binding sites for mannose on every horcolin protomer.

**Modeling HSQC Lineshapes Using the Bloch-McConnell Equations.** NMR lineshapes in the intermediate-fast
Table 3. Equilibrium Dissociation Constants ($K_d$), As Well As On- ($k_{on}$) and Off-Rate Constants ($k_{off}$) for the Binding of Mannose to Wt, D138A, and D39A Horcolin

<table>
<thead>
<tr>
<th></th>
<th>wt</th>
<th>D138A</th>
<th>D39A</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_d$</td>
<td>19.9 ± 0.1 M</td>
<td>54.00 ± 0.08 M</td>
<td></td>
</tr>
<tr>
<td>$k_{off}$</td>
<td>(2.4 ± 0.2) x 10^3</td>
<td>(15.1 ± 0.2) x 10^3</td>
<td></td>
</tr>
<tr>
<td>$k_{on}$</td>
<td>(1.19 ± 0.08) x 10^3</td>
<td>(2.80 ± 0.05) x 10^3</td>
<td></td>
</tr>
<tr>
<td>$K_d$</td>
<td>4.71 ± 0.02 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{off}$</td>
<td>(2.5 ± 0.1) x 10^3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{on}$</td>
<td>(5.3 ± 0.2) x 10^3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exchange timescale is exquisitely sensitive to exchange rate constants and chemical shift differences between the exchanging states. Lineshape analysis has been extensively used to determine rate constants for conformational interconversion in small molecules as well as to study folding transitions and enzyme-catalyzed Pro cis–trans isomerization. Since mannose binding to horcolin occurs in the intermediate-fast regime on the chemical shift timescale, simple fast exchange equations cannot be used to determine accurate dissociation constants. To quantitatively model the titration data, we used the recently developed TITAN software, which numerically integrates the Bloch-McConnell equations at each titration point (i.e., for each protein and ligand concentration) to directly fit the peak lineshape observed in the corresponding HSQC spectrum.

Regions of the HSQC spectra of horcolin acquired at mannose concentrations ranging from 0 to 500 mM were chosen to include at least one peak with a chemical shift perturbation larger than 0.1 ppm, as well as other resonances with intensities overlapping with the peak of interest. A total of 17 residues, including peaks necessary for accounting for overlapped intensity, were selected for data fitting.

Since the crystal structure of mannose-bound horcolin, as well as the curved NMR chemical shift perturbation trajectories, confirms the existence of at least two mannose-binding sites, we first chose the two-binding-site noncooperative model (NC), given by the following equations for modeling the titration. In the above equations, $P$ is horcolin, $L$ is mannose, and the dissociation constants for binding at the two sites are given as follows:

$$ P + L \overset{k_{on1}}{\underset{k_{off1}}{\rightleftharpoons}} PL1 $$

$$ P + L \overset{k_{on2}}{\underset{k_{off2}}{\rightleftharpoons}} PL2 $$

$$ PL1 + L \overset{k_{on1}}{\underset{k_{off1}}{\rightleftharpoons}} PL12 $$

$$ PL2 + L \overset{k_{on2}}{\underset{k_{off2}}{\rightleftharpoons}} PL12 $$

for the two-site NC mechanism. When the amount of PL2 is small throughout the titration, the dissociation constants for binding at the two sites are given as follows:

$$ K_{d1} = \frac{k_{off1}}{k_{on1}} $$

$$ K_{d2} = \frac{k_{off2}}{k_{on2}} $$

Global fits of the titration data for 17 residues and 22 mannose concentrations to the NC model did not converge, likely because the chemical shifts of PL1 and PL2 cannot simultaneously be unambiguously located from the NMR spectra. To reduce the number of residue-specific parameters, we chose a cluster of four residues that includes G15 and G64 for model analysis. Residues in this cluster respond appreciably to both binding events; G64 changes markedly in chemical shift upon binding of the first mannose, while G15 moves significantly when the second mannose binds, making this cluster an ideal choice for data fitting. Global modeling of this cluster gives $K_d$ values of $59 \pm 2$ mM ($K_{d1}$) and $7.7 \pm 0.1$ mM ($K_{d2}$) for CBS1 and CBS2, respectively (Table S2). However, there are small systematic deviations between the spectra and the fits of G15 (Figure S3). Since G15 is maximally responsive to mannose binding at CBS1, this discrepancy suggests that $K_{d1}$ may be somewhat different from the fit value of 59 mM.

We next evaluated whether the two-site sequential binding mechanism was sufficient to model our titration data. This model is defined by the following equations:

$$ P + L \overset{k_{off1}}{\underset{k_{on1}}{\rightleftharpoons}} PL1 $$

$$ PL1 + L \overset{k_{off1}}{\underset{k_{on1}}{\rightleftharpoons}} PL12 $$

While the number of global kinetic parameters in both the sequential and the NC models is four, there are six fewer residue-specific parameters ($R_{iN}, R_{iHH}, R_{iNH}, R_{iHN}, \sigma_i, \sigma_{iH}$) per residue because PL2 does not feature in the equations. Here, $R_{ij}, R_{ijy}$, and $\sigma_i$ are the longitudinal and transverse relaxation rate constants and the chemical shift of nucleus $j$ ($i \in NH$) in PL2. Since PL2 is not directly visible in the HSQC spectra during the mannose titration, these parameters are poorly defined and removing them from the fitting routine should improve the convergence characteristics.

However, the sequential binding model is an approximation to the two-site NC mechanism and is a good approximation only when the amount of PL2 is small throughout the titration. To determine whether the sequential model was appropriate for our data, we simulated the populations of $P$, PL1, PL2, and PL12 through the titration using the $K_d$ values obtained from fits of the data to the NC model (Figure S4). The simulations show that the maximum fraction of PL2 during the titration is 0.070, with an average over the entire titration of 0.039, clearly indicating that the contribution of the $P \leftrightarrow PL1 \leftrightarrow PL12$ pathway to the formation of doubly bound horcolin is minimal and justifying the use of a $P \leftrightarrow PL1 \leftrightarrow PL12$ sequential model for analyzing the titration data.

HSQC-detected titration data for 17 residues and 22 mannose concentrations were then globally fit to the sequential two-site-binding model to extract the values of $K_{d1}$, $k_{off1}$, $K_{d2}$, and $k_{off2}$ and the chemical shifts of $P$, PL1, and PL12, as detailed in the Experimental Procedures. A representative overlay of the experimental and fit HSQC data (Figures 5D and S5) for the region including G15 and G64 demonstrates
that the chosen model faithfully reproduces the measured intensities and confirms the robustness of the fitting procedure.

Table 3 lists the $K_d$ and $k_{off}$ values obtained from the TITAN fitting routine. The robustness of parameters estimated from a ligand-binding titration depends crucially on the extent to which the binding sites are saturated with the ligand. Accordingly, we calculated the amount of doubly bound horcolin (PL12) and the occupancy of carbohydrate-binding sites at the end of the titration from the $K_d1$ and $K_d2$ of wt horcolin. The percentage of PL12 is 95% and the occupancy factor is 98%, confirming that horcolin is virtually saturated with mannose at the end of the titration. To further validate the results from TITAN modeling, we chose F136, which shows significant chemical shift differences upon binding of mannose to both CBS1 and CBS2, but was left out of the global fitting routine because of extensive peak overlap in adjoining regions of the HSQC spectrum. Using the thermodynamic and kinetic parameters for the horcolin–mannose interaction (Table 3), we simulated the F136 lineshape as a function of mannose concentration (Figure S2). The simulated data agree well with the experimentally observed chemical shift trajectory, showing the reliability of the global parameters recovered from the fit.

The binding affinities of the two ligand-binding sites for horcolin are both weak and fall in the millimolar range ($K_d1 = 19.9$ mM and $K_d2 = 4.71$ mM, Table 3), consistent with previous observations on mannose binding by most lectins from the mJRL family. TITAN fits also provide an estimate of the dissociation rate constant ($k_{off}$) values at each binding site, which was used in conjunction with $K_d$ to determine the on-rate constants ($k_{on}$). $k_{on}$ values fall between 1 and $6 \times 10^5$ M$^{-1}$ s$^{-1}$ for the two sites and are 2 orders of magnitude smaller than the diffusion-limited rate constant in water ($\sim 3 \times 10^8$ M$^{-1}$ s$^{-1}$). This indicates that every collision of mannose with horcolin is not productive and that considerable reorientation may be required before native contacts are established between the hydroxyl moieties in mannose and the binding site residues of horcolin. There may also be transient occlusion of the binding site by the motion of neighboring loop residues, though this seems less likely given the high-order parameter values of the loop regions observed in solution (Figure 4E).

Interestingly, the binding affinity of horcolin for mannose is different between the two sites, with mannose binding 4-fold more tightly to CBS2 ($K_d2 = 4.71$ mM) than CBS1 ($K_d1 = 19.9$ mM). This 4-fold difference can be extracted reliably from NMR titrations and their modeling using TITAN, as seen from the bootstrapped parameter distribution and $\chi^2$ analysis (Figure S6). The difference in $K_d$ values stems almost entirely from the $k_{on}$ value between the two sites [(1.19 ± 0.08) × 105 M$^{-1}$ s$^{-1}$ for CBS1 versus (5.3 ± 0.2) × 105 M$^{-1}$ s$^{-1}$ for CBS2], while the dissociation off-rate constant is virtually identical. This suggests that the interactions that hold mannose in the two binding sites are very similar and need comparable activation free energies to be disrupted, but mannose has more difficulty (slower on-rate constant) in entering and engaging with CBS1 than with CBS2.

To dissect the contribution of the two mannose-binding events to the overall CSPs seen in wt horcolin (Figure 5C), we...
plotted the TITAN-derived chemical shift changes for the first ($P \rightarrow PL1$) and second ($PL1 \rightarrow PL12$) binding events (Figure S7). The largest changes in the chemical shift during $P \rightarrow PL1$ are seen in the second GG loop (G60−T65), as expected from mannose binding to CBS2, while the $PL1 \rightarrow PL12$ transition causes significant changes in the first GG loop (G11−G15), which makes contacts with mannose at CBS1.

Finally, a comparison of the parameters obtained from the sequential and NC mechanisms is provided in Table S2. Both models consistently reveal a measurable difference in binding affinities between the two sites on horcolin, with a 4.2-fold higher affinity for CBS2 from the sequential model and a 7.7-fold higher affinity from the NC model. We have chosen to use the results from the sequential model in all subsequent sections because we obtain good quality fits with a global analysis of all 17 residues, as well as statistics on all affinities and rate constants. The NC model, on the other hand, does not yield fits that are of as good quality, the chemical shifts of $PL1$ and $PL2$ vary arbitrarily, and global fits of all 17 residues do not converge because of ambiguities in residue-specific chemical shift and relaxation parameters.

Structure-Guided Mutagenesis for Validating the Difference in Binding Affinities. The NMR titrations of horcolin with mannose reveal a dichotomy in the molecular recognition of mannose by the two ligand-binding sites in horcolin, with CBS2 displaying a 4-fold higher affinity than CBS1. To the best of our knowledge, this is the first time that a difference in dissociation constants in the two binding sites has been characterized within the jacalin family of lectins. While statistical measures including $\chi^2$ surfaces and bootstrap distributions confirm that the 4-fold difference in $K_d$ values is significant, we desired to further validate this result using site-directed mutagenesis. The structure of mannose-bound horcolin shows that the pattern of protein−ligand interactions is very similar at both CBS1 and CBS2, with most of the hydrogen bonds involving backbone donor or acceptor atoms. However, there is a conserved Asp residue in each site (D138 in LBL1 and D39 in LBL2) whose side chain hydrogen bonds with mannose at that particular site. Accordingly, we chose to mutate D39 and D138 individually to Ala to disrupt these side chain−ligand hydrogen bonds and create single-binding-site variants of horcolin.

NMR chemical shift perturbations observed upon mutagenesis are small with an average value of 0.08 ppm (D138A) and 0.13 ppm (D39A) across the entire protein (Figure S8). The largest changes are localized to the site of the mutation itself and confirm that D39A and D138A horcolin share the same structure as the wt protein. Figures 6A and S9 show $^{1}H−^{15}N$ HSQC-detected NMR titration data of D138A and D39A horcolin, where the mannose concentration is varied from 0 to 167 mM over 18 points for D138A and 0 to 654 mM over 21 points for D39A horcolin. Unlike in wt horcolin, all residues that show chemical shift perturbations upon mannose addition in both mutants trace a linear trajectory from the initial to the final resonance position through the titration, confirming that the stoichiometry of binding is 1:1 in both cases (Figures 6A and S9). Figure 6B,C shows the perturbations in D39A and D138A horcolin when they each bind mannose. As expected from the single-binding-site nature of D39A and D138A horcolin, mannose binding to CBS2 in D138A horcolin does not cause CSPs in LBL1 (Figure 6B), which directly contacts the mannose bound to CBS1 in wt horcolin. Similarly, CSPs for a number of residues in LBL2 are reduced significantly when mannose binds to CBS1 in D39A horcolin (Figure 6C).

To determine the binding thermodynamics and kinetics of glycan recognition, we fit selected resonances with CSP values larger than 0.1 ppm to a two-state model

$$P + L \rightleftharpoons PL$$

using the TITAN software package. A total of 16 residues for D39A and 14 residues for D138A horcolin were chosen for data fitting. Representative G61 resonances of wt, D39A, and D138A in the mannose titration are shown along with the fits from TITAN, and the fits agree very well with the data (Figure 6A). Both mutants show narrow $K_d$ and $k_{off}$ bootstrap distributions and sharp $\chi^2$ surfaces (Figure S6), confirming that the equilibrium and dissociation rate constants can be obtained reliably from the mannose titrations. The $K_d$ for D138A horcolin, where binding occurs only at CBS2, is 3.2 ± 0.3 mM (corresponding to 98% mannose-bound D138A horcolin at the end of the titration), while the $K_d$ of 54.00 ± 0.08 mM for D39A horcolin (94% bound form at the end of the titration), which retains binding only at CBS1, is almost 20-fold smaller. The $k_{on}$ values of both single-binding-site mutants differ by 5.7-fold, and this ratio is similar to the $k_{on}$ values for CBS1 and CBS2 in wt horcolin. However, the $k_{off}$ values, which are comparable for the two binding sites in the wt protein, vary by a factor of 3 in the mutants, increasing the difference in $K_d$ values between CBS1 and CBS2. Taken together, the single-binding-site mutant data thus unequivocally demonstrate that mannose binding at CBS2 occurs more tightly than at CBS1.

**DISCUSSION**

Horcolin is a mannose-specific lectin from barley that binds HIV glycoproteins with high affinity and shows therapeutic potential as an antivirus agent, as it lacks the mitogenic activity that is a common attribute of most other lectins. In this report, we use a combination of X-ray crystallography and solution NMR spectroscopy to characterize the structure of horcolin and its binding to mannose at atomic resolution.

Horcolin is a dimer in both the apo- and mannose-bound forms and adopts a $\beta$-prism I fold composed of three Greek key motifs. Perhaps the best studied mannose-specific $\beta$-prism I fold lectin is BanLec,15,16,47 (Figure 7A), which is evolutionarily close to horcolin and shares 39% sequence identity with it. A superposition of BanLec and horcolin subunit structures gives an RMSD of 1.3 Å, whereas the dimeric structures superpose with an RMSD of 2.2 Å. While most of the reported $\beta$-prism I fold lectins are dimeric or tetrameric,62 there is one example of unusual heptameric quaternary association63 and another instance where octameric assembly has been observed.64 The octamer in the latter case is made up of BanLec-type dimers. The dimer interface in horcolin is composed of two $\beta$-strands ($\beta1$ and $\beta10$) from adjacent subunits that pack in an antiparallel fashion. This dimeric arrangement is similar to several other $\beta$-prism I fold lectins, emphasizing the importance of these interactions in the quaternary organization. Horcolin shows the highest sequence identity with Oryzata lectin (44%), followed by BanLec (39%), AcmJRL (lectin from pineaple, 38%), and HeltubA (37%). All these lectins are evolutionarily close to horcolin and all of them are of plant origin. The subunit structure of horcolin is similar
to those of BanLec, AcmJRL, and other β-prism I fold lectins and lectin domains of known structures, with RMSD values between the Ca atoms of horcolin and Orysata, BanLec, and AcmJRL of 1.60, 1.30, and 1.81 Å, respectively.

In β-prism I fold lectins, a conserved GXXXD motif is involved in carbohydrate recognition. This motif was originally characterized in artocarpin (Artocarpus integrifolia agglutinin), which has a single carbohydrate-binding motif per subunit located in the C-terminal Greek key I. In BanLec, on the other hand, shows two primary binding sites located in Greek keys I and II (Figure 7A), while the residues at the corresponding positions in Greek key III serve as a common secondary site for recognition of mannoooligosaccharides. In BanLec, the two LBLs comprise 129GDFID133 and 34GDVD38 on Greek key I and II,15 while in horcolin, LBL1 and LBL2 are 134GAFLD138 on Greek key I and 35GAI139 on Greek key II.29 A recently characterized mJRL from pineapple, AcmJRL, also has two mannose-binding loops—115GSLVD33 on Greek key I and 35AHAID39 on Greek key II—where the second mannose recognition motif contains Ala in place of Gly, which is found in all other members of the family.17 In all three lectins with two LBLs, horcolin, BanLec, and AcmJRL, the backbone NH of the first three residues in each LBL participate in hydrogen bonding interactions with O6 of mannose. The fourth residue is a hydrophobic amino acid that is not involved in any interaction, while the last residue is Asp, which forms two hydrogen bonds through its side chains with O4 and O6 of mannose. Thus, mannose shows a similar mode of binding with the LBLs of these lectins. Cooperativity between a pair of carbohydrate-binding sites has been observed in a very few instances, that too typically when cross-linked complexes form between lectins and multivalent oligosaccharides.56,67 However, there is no evidence for differences in affinities and rates of monosaccharide binding to a lectin with two very similar binding sites.

Our results illustrate the potential of HSQC-based titrations, in conjunction with NMR lineshape modeling, for elucidating the thermodynamics, kinetics, and mechanism of protein–ligand interactions. Here, the analysis of lineshapes in the 1H−15N HSQC-detected mannose titration clearly demonstrates the presence of two mannose-binding sites in horcolin with $K_d$ values of 4.71 and 19.9 mM. Previous ITC measurements of horcolin with mannose could be fit well using an identical binding sites’ model with a $K_d$ of 5 mM and a stoichiometry of 2. However, ITC lacked the sensitivity to discriminate between the affinities of the two sites because of the weak millimolar range of the dissociation constants and also did not provide insights into potential allosteric communication between CBS1 and CBS2.29 Since the exchange between free and ligand-bound forms of horcolin lies in the intermediate-fast timescale, NMR lineshapes are also sensitive to the kinetics of conformational exchange and enabled us to characterize the $k_on$ and $k_off$ values for mannose binding to horcolin. Moreover, the high solubility of mannose made it possible to saturate the binding of wt and mutant horcolin, despite the weak millimolar ligand-binding affinities of these proteins. While high concentrations of cosolutes such as mannose can elevate solution viscosity and also cause molecular crowding, we have verified that the mannose concentrations ≤0.65 M used in the titrations cause less than 30% increase in solution viscosity and do not measurably affect binding affinities and rate constants (Supporting Information and Figure S11).

The thermodynamics and kinetics of mannose recognition by the single-binding-site variants D138A and D39A horcolin suggest that there is allosteric coupling between the two CBSs. The binding affinities of CBS2 and CBS1 for mannose in wt horcolin are 4.71 and 19.9 mM, corresponding to a binding free energy difference $\Delta \Delta G^o = 0.85$ kcal/mol. While the affinity for CBS2 is similar in D138A horcolin, the affinity of CBS1 decreases by 2.5-fold ($K_d = 54$ mM) in the D39A variant, corresponding to a $\Delta \Delta G^o$ between the two single-site binding events of 1.7 kcal/mol. There are at least two mechanisms for this decrease in the binding affinity of CBS1 between wt and mutant horcolin. First, it is possible that mannose binding by wt horcolin at CBS2 allosterically alters the CBS1 to increase its capacity for recognizing mannose. This mechanism will introduce cooperativity in the binding of mannose to horcolin, and while we do not need a cooperative two-site model to fit the titration data, it is likely that our NMR data sets are not sufficiently sensitive to pick up this cooperative behavior superimposed on a set of already weak
binding affinities. Second, the perturbation of CBS2 consequent to the D39A mutation may be transmitted allosterically to CBS1, spatially rearranging CBS1 and reducing its affinity for mannose. Both these mechanisms suggest the existence of molecular communication between CBS1 and CBS2, and this is supported by the observation that mutation at one site affects sugar-binding rate constants at the other. For example, $k_{\text{on}}$ and $k_{\text{off}}$ at CBS2 increase by 3- and 2-fold, respectively, when D138 at CBS1 is mutated to Ala, while $k_{\text{on}}$ and $k_{\text{off}}$ at CBS1 increase by 2.4- and 6-fold when D39 at CBS2 is mutated to Ala, respectively. Interestingly, CSPs occurring at the secondary binding site upon mannose binding in wt horcolin disappear in both single-site-binding mutants, suggesting that the simultaneous presence of sugar residues at both sites might be necessary to engage the secondary binding site.

To probe the molecular determinants for the communication between CBS1 and CBS2, we examined the interactions established by residues at the two sugar-binding sites. Figure 7B clearly shows the presence of a network of hydrogen bonds linking LBL1 and LBL2 directly as well as through the GG loops and the secondary binding site, providing a molecular pathway for the allosteric transmission of signals between the two sites. This cross-talk between the two CBSs of horcolin is expected to have implications for the binding of higher-order mannose-rich glycans such as those present in the envelopes of viral glycoproteins.

**CONCLUSIONS**

The absence of detectable mitogenicity in horcolin, coupled with its strong and specific recognition of gp120, as well as its ability to neutralize HIV infection, highlights the potential of horcolin as a therapeutic anti-HIV agent. Here, we use X-ray crystallography and multidimensional NMR spectroscopy to elucidate the structural and glycan recognition features of horcolin. Horcolin adopts a $\beta$-prism I fold with three Greek key motifs that carry two carbohydrate-binding sites. Using a combination of NMR 2D lineshape modeling and site-directed mutagenesis, we show that the two sites exhibit at least 4-fold difference in binding affinities and communicate with each other through a network of hydrogen bonds linking the two sites. Our results pave the way for a molecular-level understanding of the recognition of high-mannose glycans and glycoproteins by horcolin and possibly for a dissection of the molecular determinants of mitogenicity in lectins.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.1c00778.

Backbone assignments of horcolin shown on the $^1$H–$^15$N HSQC spectrum of wt apo-horcolin (Figure S1); simulation of $^1$H–$^15$N HSQC spectra of overlapped regions in the titration of wt horcolin with mannose (Figure S2); fits of the titration data to a two-site non-cooperative binding model (Figure S3); justification for the use of a sequential binding model for the mannose-wt horcolin interaction (Figure S4); fits of the titration data to the sequential binding model (Figure S5); TITAN error analysis for $K_d$ and $k_{\text{off}}$ for the mannose binding NMR titration data of D138A, D39A, and wt horcolin (Figure S6); changes in peak position in wt horcolin resulting from mannose binding at CBS1 and CBS2 (Figure S7); $^1$H–$^15$N HSQC overlays of wt and binding site mutants of horcolin (Figure S8); $^1$H–$^15$N HSQC spectra demonstrate that D138A and D39A horcolin are single-binding-site variants (Figure S9); comparison of tagless and tagged wt Horcolin binding with methyl-$\alpha$-d-mannopyranoside (Figure S10); effect of mannose concentration on solution viscosity (Figure S11); NMR resonance assignment statistics for wt apo-horcolin (Table S1); comparison of parameters between the two-site NC and sequential models (Table S2); effect of high mannose concentrations on horcolin-mannose interactions (Text) (PDF)

**Accession Codes**

The coordinates of apo- and mannose-bound horcolin (UniProt ID: Q5U9T2) have been deposited in the Protein Data Bank with accession IDs 7V4Z and 7V4S, respectively. NMR resonance assignments of horcolin have been deposited in the Biological Magnetic Resonance Data Bank (BMRB) with an accession number of 51159.

**AUTHOR INFORMATION**

**Corresponding Authors**

Kaza Suguna — Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India; Email: suguna@iisc.ac.in

Avadhesh Surolia — Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India; orcid.org/0000-0002-2466-2514; Email: surolia@iisc.ac.in

Ashok Sekhar — Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India; orcid.org/0000-0002-8628-7799; Email: ashoksekar@iisc.ac.in

Authors

Vaishali Narayanan — Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Kishore Babu Bobbili — Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Nukathoti Sivaji — Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Nisha G. Jayaprakash — Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.1c00778

**Author Contributions**

V.N. and K.B.B. contributed equally to this work.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The initial X-ray data sets were collected at the X-ray facility for Protein Crystal Structure Determination and Protein Design at the Indian Institute of Science, supported by the Science and Engineering Research Board (SERB) of the Department of Science and Technology (DST). High-resolution data were collected at the XRD2 beamline, Elettra Synchrotron, Trieste, Italy. This work was supported by the DBT/Wellcome Trust India Alliance Fellowship (Grant No. IA/1/18/1/503614) and a DST/SERB Core Research Grant (No. CRG/2019/003457), as well as a start-up grant from IISc awarded to Ashok Sekhar, and a grant (BT/PR27659/BID/7/829/2018) from the Department of Biotechnology (DBT),...
Government of India, to Avadhesha Surolia (A.S.). K.B.B. and N.G.J. are Senior Research Fellows in the DBT grant to A.S. A.S. is a SERB Distinguished Fellow. The authors acknowledge funding for infrastructure support from the following programs of the Government of India: DST-FIST,UGC-CAS, and the DBT-IISc partnership program. V.N. thanks IISc programs of the Government of India: DST-FIST, UGC-funding for infrastructural support from the following.

A.S. is a SERB Distinguished Fellow. The authors acknowledge N.G.J. are Senior Research Fellows in the DBT grant to A.S.

Biochemistry pubs.acs.org/biochemistry 32 1310.


---

Recommended by ACS

**Glycan Cluster Shielding and Antibody Epitopes on Lassa Virus Envelop Protein**
Su Yong Re and Kenji Mizuguchi
FEBRUARY 19, 2023
THE JOURNAL OF CHEMICAL PHYSICS

**Binding of the Bacterial Adhesive FimH to Its Natural, Multivalent High-Mannose Type Glycan Targets**
Maximilian M. Sauer, Rudi Glockshuber, et al.
DECEMBER 13, 2018
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

**The Impact of Sustained Immunization Regimens on the Antibody Response to Oligomannose Glycans**
FEBRUARY 28, 2020
ACS CHEMICAL BIOLOGY

**Impact of Glycan Linkage to Staphylococcus aureus Wall Teichoic Acid on Langerin Recognition and Langerhans Cell Activation**
Astrid Hendriks, Nina M. van Sorge, et al.
FEBRUARY 16, 2021
ACS INFECTIOUS DISEASES