Published in final edited form as:

FEBS J. 2013 May ; 280(9): 2056–2067. doi:10.1111/febs.12229.

***Burkholderia oklahomensis* agglutinin is a canonical two-domain OAA-family lectin: structures, carbohydrate binding, and anti-HIV activity**

Matthew J. Whitley¹, William Furey^{2,3}, Sireesha Kollipara⁴, and Angela M. Gronenborn¹

¹Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA

²Department of Pharmacology & Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA

³Biocrystallography Laboratory, Veterans Affairs Medical Center, Pittsburgh, PA 15240, USA

⁴Department of Pathology, Microbiology, & Immunology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

Abstract

Burkholderia oklahomensis EO147 agglutinin (BOA) is a 29 kDa member of the OAA family of lectins. Members of the OAA family recognize high-mannose glycans and, by binding to the HIV envelope glycoprotein 120 (gp120), block the virus from binding to and entering the host cell, thereby inhibiting infection. OAA-family lectins comprise either one or two homologous domains, with a single domain possessing two glycan binding sites. We solved the structure of BOA in the ligand-free form as well as in complex with four molecules of 3 α ,6 α -mannopentaose, the core unit of the N-linked high-mannose structures found on gp120 *in vivo*. This is the first structure of a double-domain OAA lectin in which all four binding sites are occupied by ligand. The structural details of the BOA-glycan interactions presented here, along with the determination of affinity constants and HIV inactivation data, shed further light onto the structure-function relationship in this important class of anti-HIV proteins.

Keywords

lectins; anti-HIV activity; high-mannose glycans; sugar binding; x-ray crystallography; NMR

Introduction

More than 34 million people around the world are infected with the human immunodeficiency virus (HIV) today, and 1.7 million people died from acquired immunodeficiency syndrome (AIDS) in 2011 alone. Enormous progress has been made over the last decade in combating the disease, and powerful drugs are now available that attack the virus at various points in its life cycle. For example, reverse transcriptase inhibitors such as etravirine [1] target the enzymatic process by which the viral RNA genome is retrotranscribed into DNA inside the infected cell, while integrase inhibitors such as

Corresponding author: Angela M. Gronenborn, Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., 1050 BST3, Pittsburgh, PA 15261, USA. Telephone: (412) 648-9959; Fax: (412) 648-9008; amg100@pitt.edu.

Database: The atomic coordinates and structure factors for ligand-bound and ligand-free BOA have been deposited into the Protein Data Bank under the accession numbers **4GK9** and **4GU8**, respectively.

raltegravir [2] block the enzymatic integration of any successfully created retroviral DNA into the host genome. If reverse transcription and integration proceed normally, drugs such as darunavir [3] can still prevent the creation of new infectious virions inside an infected cell by inactivating HIV protease, thus preventing newly synthesized viral polyproteins from being cleaved into their mature forms.

Although combination therapies with these different types of drugs have been very successful in reducing viral load in patients, they combat HIV only *after* the virus has already entered the target cell, rendering the potentially alternative strategy of blocking entry a desirable avenue to explore. Molecules that play a role in viral entry include the envelope glycoproteins 120 (gp120) and 41 (gp41) on the viral surface and CD4 and chemokine receptors [4-8] on the target cell. Existing drugs such as enfuvirtide [9, 10] and maraviroc [11] disrupt the interactions between these viral and host proteins by binding to gp41 or the CCR5 co-receptor, thus preventing the virus from fusing to its host target. Potentially the most promising target for entry inhibitors, however, may be gp120 itself because of its central role in the entry mechanism. However, targeting gp120 has been difficult for several reasons. The functional form of gp120 is trimeric [12, 13], with part of the surface of each monomer hidden and unavailable for generating an immune response. A considerable portion of the remaining exposed surface of trimeric gp120 is covered by a dense layer of N-linked glycans that effectively shields the underlying protein from the immune system [14].

A major step forward in developing strategies for blocking viral entry was the discovery of broadly neutralizing antibodies that recognize the gp120 carbohydrate shield itself; these antibodies include 2G12 [15, 16] and PGT121 [17], among others. They interact with high-mannose and complex glycans on gp120, penetrate the glycan shield, and are thought to act by crosslinking Env trimers on the viral surface. The same mechanism applies to the HIV-inactivating potential of several lectins [18, 19]. Many oligomannose-binding proteins have been discovered over the last two decades, and, at present, numerous lectin-mediated antiviral activities have been described [20, 21]. Examples of antiviral lectins include concanavalin A from jack-beans, jacalin from the jackfruit *A. heterophyllum*, cyanovirin-N from the cyanobacterium *N. ellipsosporum*, scytovirin from the cyanobacterium *S. varium*, and griffithsin from the red alga *Griffithsia* sp., among others. Among these, cyanovirin-N has shown promise as a microbicide for preventing rectal [22] and vaginal transmission [23, 24] of HIV, validating the role of such lectins in developing multifaceted approaches to prevent HIV infection. One of the most recently discovered lectins with anti-HIV activity, *Oscillatoria agardhii* agglutinin (OAA), exhibits no homology to any of the other classes, and its amino acid sequence contains an unusually high number of glycines (~20%) [25, 26]. The crystal structure of OAA revealed that the protein adopts a novel β -barrel-like topology (unique in the Protein Data Bank at the time of structure determination) and that it recognizes Man α (1-6)Man disaccharide units [27] in a manner structurally distinct from cyanovirin-N [28]. At present, a search of GenBank with the OAA sequence yields 19 hits from 13 different organisms. Of these 19 proteins, six possess only a single domain, and 13 contain two domains. In the present work, we structurally and biochemically characterize a double-domain OAA-family agglutinin from *Burkholderia oklahomensis* EO147. Many proteobacteria of the *Burkholderia* genus are plant or human pathogens, causing such diseases as melioidosis and infecting the lungs of people suffering from cystic fibrosis. *B. oklahomensis* itself was first identified in an infected leg wound of a farmer in the U.S. state of Oklahoma [29]. X-ray structures of both the ligand-free and ligand-bound forms of the agglutinin were determined, and details of 3 α ,6 α -mannopentose recognition were uncovered. In addition, glycan binding in solution was assessed by NMR, and BOA's anti-HIV activity was tested in single-round HIV infectivity assays. Our results, in conjunction with previous work on other HIV-inactivating lectins, contribute to the further development of lectins as anti-HIV therapeutics.

Results and Discussion

Analysis of the glycan-bound BOA structure and details of the interaction with 3 α ,6 α -mannopentaose

The protein construct used for crystallization contains three additional N-terminal residues (labeled S₂V₁D₀ in the PDB file) that remain after enzymatic removal of the N-terminal hexahistidine tag, followed by the 276 residues of BOA's primary sequence. BOA also contains a unique, ten residue N-terminal tail that is not present in of any other members of the OAA family. NMR titration analysis with the BOA homologue OAA demonstrated that 3 α ,6 α -mannopentaose, the core unit of high-mannose glycans (Figure 1), is the essential binding epitope [28]. We therefore reasoned that 3 α ,6 α -mannopentaose is the ligand of choice for co-crystallization trials with BOA. We solved the x-ray crystal structure of the BOA-glycan complex containing four bound molecules of 3 α ,6 α -mannopentaose and refined the model to 1.9 Å resolution; this structure is the first structure of a double-domain OAA family member in which all four binding sites are occupied by carbohydrate. The co-crystals exhibit the primitive tetragonal space group $P4_32_12$ and contain a single copy of the protein-ligand complex in each asymmetric unit. The polypeptide chain partitions into two β -barrel-like domains connected to each other by a short three-residue linker. The sequence in each barrel comprises two 66-amino-acid repeats forming five antiparallel β -strands. An amino acid sequence alignment of all members of the OAA family discussed in this paper is provided in Figure 2. Sequences of the single domains of OAA and PFA are aligned with the first domains of BOA and MBHA. Glycan binding site residues are underlined and color-coded differently for each binding site.

The overall structure of ligand-bound BOA is displayed in Figure 3. BOA's two β -barrels share 66% sequence identity over 126 aligned residues, and a structural superposition of the two domains performed using the PDB Protein Comparison Tool [30] yields a C α RMSD of 0.6 Å. The angle between the long axes of the two barrels is approximately 85° (Figure 3B). Six side chains (N45, Q75, D112, N191, N205, and N251) were modeled in two conformations based on clear density for alternate conformers in the F_o-F_c difference electron density contoured at 3 σ . The peptide bond between L274 and P275 was found to be in the *cis*-conformation. The experimental 2F_o-F_c electron density was unambiguous for both the protein and all four carbohydrate ligands, and an example of quality of the density is provided in Figure 4A.

For illustration purposes we have selected binding site III, the first binding site in the C-terminal β -barrel domain, to delineate the interactions between BOA and the glycan, but a complete list of contacts for all binding sites is provided in Table 2. Most interactions involve residues that are located in the loops and turns that connect the β -strands of the barrels. The reducing end of the pentasaccharide (M3) is held in place by a combination of hydrogen bonding and hydrophobic interactions: the guanidinium group of R236, located in the loop connecting β -strands 7 and 8, is engaged in hydrogen bonds (3.4 and 2.7 Å between heavy atoms) with the C3 and C4 hydroxyls of M3, respectively. The M3 C4 hydroxyl is also hydrogen bonded (2.7 Å) to the side-chain carboxylate group of E264 in the β 9- β 10 loop. In addition, the E264 carboxylate group is in hydrogen bonding distance to the guanidinium group of R236, and it appears that a salt bridge between R236 and E264 correctly positions the two side chains to recognize the hydroxyls of the M3 pyranose ring. The opposite face of the M3 ring packs against the indole side chain of W152, located in the β 1- β 2 loop. The tight packing between the tryptophan indole and the M3 ring can be seen clearly in the view provided in Figure 4B. The glutamate, arginine, and tryptophan residues that contact the M3 sugar are absolutely conserved among all OAA-family lectins (Figure 2).

Additional interactions involving mannose units other than M3 consist of hydrogen bonds between the protein backbone and sugar hydroxyl groups. The amide group of G265 is engaged in hydrogen bonds with the O5 and O6 atoms of M4' (2.9 and 3.2 Å, respectively), and the carbonyl oxygen of the following residue, P266, serves as a hydrogen bond acceptor for the C2 hydroxyl of M4'. The backbone amides of another two absolutely conserved glycines, G153 and G154, both interact with the C6 hydroxyl of M5' (3.4 and 3.0 Å, respectively), and G153 makes an additional hydrogen bond to the M5' C5 hydroxyl (2.9 Å). All these protein-ligand hydrogen bonds are associated with the core trisaccharide unit comprised of M5'α(1-3)M4'α(1-6)M3 that is part of the D2 arm of Man-9 (Figure 1), and they are depicted as dashed thick black lines in Figure 4A; no interactions were observed between BOA and mannosyl rings M4 or M5''. Structurally, the binding site for the pentasaccharide forms a shallow surface pocket from which the interacting residues detailed above project towards the glycan. The pyranose rings M3 and M4' are located completely within the pocket, while M5' is located at one end. The mannoses M4 and M5'' are found outside the binding cleft and do not interact with the protein. A surface representation of the protein depicting the ligand in the binding cleft is displayed in Figure 4C.

In light of the conservation of the protein-sugar interactions in all four binding sites, we examined the conformation of each of the bound sugars. A least-squares superposition of all four bound sugars is displayed in Figure 4D and color-coded by binding site as in Figure 2. The superposition reveals that the core trisaccharide M5'α(1-3)M4'α(1-6)M3 is conformationally invariant among the four bound ligands, whereas mannose rings M4 and M5'', which are not located in the binding site and are not restrained by interactions with the protein, clearly display a higher degree of conformational heterogeneity compared to the core trisaccharide.

Comparison of ligand-bound BOA with other OAA-family lectins

Only two other structures of OAA-family lectins in complex with carbohydrates are presently available. The first is OAA, the single domain founding member of the family that possesses two carbohydrate binding sites, and the second is the recent structure of MBHA, a two-domain lectin that comes from *Myxococcus xanthus* and has four sugar binding sites [31]. In the MBHA structure, however, only two of the four ligand binding sites were occupied; the unoccupied sites are occluded by sugars bound to neighboring MBHA molecules in the crystal. The structures of these other OAA-family lectins exhibit the same β-barrel-like fold first seen in OAA, which is not unexpected given their high sequence similarity (Figure 2). Within BOA, the sequence making up the first and second β-barrel domains exhibits 72% and 65% sequence identity with OAA, respectively, and a structural superposition of the individual BOA domains with OAA yields Cα RMSDs of 0.6 and 0.7 Å, respectively. No major differences between either the protein structures or the conformation of the bound 3α,6α-mannopentose ligands were discernible, although some minor conformational differences in the loops can be seen. We also compared ligand-bound BOA to MBHA, in which only two of the four binding sites are occupied by glycan, with the empty binding sites occluded by sugars that are bound to neighboring MBHA molecules. BOA and MBHA share 65% sequence identity over 259 residues. A superposition of the backbone structures for BOA and MBHA (Figure 5) reveals noticeable differences in loop conformations regardless of whether the loops are involved in ligand binding, resulting in overall Cα and all-heavy-atom RMSDs of 1.1 and 1.25 Å, respectively. Although it is tempting to speculate that crystal packing lies at the root of some of these loop conformational differences, this is unlikely to be the sole explanation, because not every individual loop packs against a neighboring protein molecule in the crystal. Thus it is more likely that both crystal packing effects as well as structural disorder contribute to these differences. It is interesting to note, however, that the conformational variation in loop

structures appears not to influence the conformation of bound 3 α ,6 α -mannopentose, as these are essentially identical (gray and red, respectively, in Figure 5). The positions and orientations of each M3 pyranose ring superimpose extremely well, while some slight variability is observed at the ends of the oligosaccharide where no interactions with the protein restrict the position of these rings.

At this juncture it is interesting to ask the question how it is possible that the ligands are positioned identically in the two structures even if the loop conformations differ noticeably in many places. The answer lies in the invariant positioning of the glycan-contacting residues of BOA and MBHA. For example, the critical glycine residues that are engaged in backbone hydrogen bonds with the sugar (see Figure 4A) are located identically in both structures, and the same holds for the key Trp, Arg, and Glu side chains (detailed for each binding site in Table 2). Therefore, although differences in the loop conformations are present, they still allow for optimal positioning of the interacting groups and thus recognition of the glycan.

Analysis of the ligand-free BOA structure

In the absence of ligand, BOA crystallized in the primitive orthorhombic space group $P2_12_12_1$, with each asymmetric unit containing eight copies of the protein related to each other by noncrystallographic symmetry (NCS). The quality of the electron density at the N-termini of the eight copies varies and results in differences in the first few amino acids that could be modeled reliably in each chain. The full N-terminal sequence (starting at position -2 with the SVD sequence remaining after removal of the hexahistidine tag) could be modeled for chains A, B, and D, while chains F and H were modeled from position 1 (Met-1, the first actual residue of the BOA sequence) onward, and chains C, E, and G from residue 2 (Thr-2) onward. Other than these differences at the N-termini, the entire primary sequence of BOA could be modeled into the electron density for each of the eight copies. Overall, the protein adopts the same tandem β -barrel-like fold as the ligand-bound form. Structural superposition of all NCS copies using SuperPose [32] yielded C α and heavy-atom RMSDs from the mean structure of 0.26 and 0.48 Å, respectively, indicating that all the non-crystallographically related copies are nearly identical.

Furthermore, there are only minor differences between the ligand-free copies of the protein and the sugar-bound form (C α RMSD of 0.4 Å). Figure 6A depicts the superposition of all eight backbone traces, with the structurally invariable elements (β -strands and some loops) depicted in light gray and the structurally variable loops colored red. These structurally variable loops are also associated with distinct, alternative conformations for the side chains of Trp-85 and Trp-152 and variable positioning of the Trp-18 and Trp-218 side chains within a single rotameric state (cyan side chains in Figure 6B-C). For four residues following each of these tryptophans, the protein backbone adopts slightly different paths in each of the eight copies. These four loops provide the key recognition elements for sugar binding. As pointed out above, the four tryptophan side chains pack tightly against the reducing end of the oligomannose ligand, with the two glycines following each tryptophan involved in backbone hydrogen bonding with the glycan. Therefore, the structural plasticity of these loops in the absence of ligand may permit recognition and easy adjustment once the ligand is present. The increased loop disorder in the absence of ligand is reflected in a larger ambiguity in the electron density of these regions, as well as by increases of 20-30 Å² in the isotropic temperature factors (B values) compared to the rest of the protein.

NMR analysis of 3 α ,6 α -mannopentose binding

Given the variable binding loop conformations in the glycan-free BOA crystal structure, we investigated BOA's binding affinity for 3 α ,6 α -mannopentose using NMR by recording a

series of ^1H - ^{15}N HSQC spectra in the presence of increasing concentrations of sugar. Binding is in slow exchange on the chemical shift scale, and spectra in the absence (black) and presence (red) of a saturating excess of sugar are displayed in Figure 7A. Because we could not assign the BOA spectrum, we simply analyzed several residues for which the free and bound resonances were easily identifiable by inspection. One such pair of resonances is boxed in Figure 7A, and the varying intensities of these resonances with increasing amounts of sugar is shown in Figure 7B-F. The intensity of the free and bound cross-peaks for any given residue reflects the population of the free and bound states for that residue and permits extraction of the dissociation constant. Note that, for the slow exchange case, each titration point yields an independent estimate of K_D , unlike in the fast exchange regime, for which a complete curve is necessary for K_D determination. By analyzing a total of 13 titration points for five cross-peaks, we obtained K_D estimates ranging from 20-67 μM , with an average value of $47 \pm 20 \mu\text{M}$. Because assignments for residues in the individual binding sites are not available, the above K_D value relates to the binding to all four binding sites and is an average value. Previous data obtained for 3 α ,6 α -mannopentose binding to individual binding sites in other OAA-family lectins revealed essentially identical dissociation constants ($\sim 10 \mu\text{M}$) for the two binding sites in OAA, whereas the individual binding site affinities in PFA and MBHA were found to differ by a factor of 1.5-2 [28, 33]. Even though precise affinities are not available for each individual binding site in BOA, our data show that binding to 3 α ,6 α -mannopentose is slightly weaker than for other OAA-family lectins. The difference, however, is at most a factor of ~ 5 and thus not likely to be important for the overall activity.

BOA is a nanomolar-scale inhibitor of HIV infectivity

BOA's ability to block HIV entry was tested in single-cycle infectivity assays. Four independent samples of BOA at concentrations ranging from 0-1000 nM were each tested in duplicate, and the normalized experimental data are displayed in Figure 8. The data yielded IC_{50} values of 10-14 nM with an average of $12.2 \pm 0.7 \text{ nM}$. This value is nearly identical to published values of 12, 12, and 15 nM for the homologues OAA, PFA, and MBHA, respectively [33]. The experimentally identical IC_{50} values for all tested OAA-family lectins demonstrate that double- and single-domain OAA-family lectins exhibit equivalent anti-HIV activity and support our previous finding that the presence of a second domain plays no appreciable role in the ability of OAA-family lectins to block HIV infection. As to whether the identical IC_{50} values of the single- and double-domain proteins might result from dimerization of the single-domain members leading to a functional mimic of a double-domain protein, we find no evidence to that effect: all our biochemical and biophysical characterizations of the OAA-family lectins have shown that the proteins are monomers in solution, even at concentrations several orders of magnitude higher than those used in the anti-HIV activity assays.

Conclusions

We have determined and analyzed the structures of *Burkholderia oklahomensis* EO147 agglutinin in the ligand-free state and bound to four molecules of the Man-9 core structure 3 α ,6 α -mannopentose, a key component of the high mannose glycans on the HIV envelope protein gp120. The sugars are bound in a shallow surface pocket, positioned using a combination of hydrogen bonding and hydrophobic interactions between the sugar and protein residues that are absolutely conserved among all known members of the OAA family. The ligand-bound BOA structure is the first structure of a double-domain OAA-family lectin in which all four binding sites are occupied by sugar. The two double-domain structures (BOA and MBHA) share the same overall fold, although minor differences in loop backbone conformations are observed. Carbohydrate binding affinities are also very

similar, and anti-HIV assays yielded essentially identical IC₅₀ values for HIV entry inhibition. The functional equivalence among the four binding sites in BOA and MBHA is the result of equivalent positioning of contact residues even if conformational differences in the associated loop backbone structures are present. The structural data described here for BOA contributes to further illuminate the structure-function relationships in the OAA family, an important new class of anti-HIV lectins.

Experimental Procedures

BOA subcloning

The primary sequence of BOA was obtained from the National Center for Biotechnology Information protein database (reference sequence **ZP_02360833.1**), and a synthetic pUC57 plasmid encoding a codon-optimized *E. coli* gene was purchased from GenScript. The coding sequence was amplified from the pUC57 plasmid using the polymerase chain reaction and ligated between the SalI and BamHI restriction sites of a pET15b vector modified to code for an N-terminal hexahistidine tag cleavable by tobacco etch virus (TEV) protease.

BOA expression and purification

For crystallographic studies, a single colony of *E. coli* BL21 Star (DE3) cells (Life Technologies, Grand Island, NY, USA), freshly transformed with the pET15b plasmid encoding the hexahistidine-tagged BOA, was used to inoculate 5 mL of LB medium supplemented with 100 µg/mL ampicillin and grown overnight at 37 °C in a rotary shaker operating at 250 rpm. This starter culture was added to 1 L of LB medium containing 100 µg/mL ampicillin and shaken at 37 °C until the optical density at 600 nm (OD₆₀₀) reached approximately 0.7. Protein expression was induced by the addition of IPTG to a final concentration of 1 mM and allowed to proceed by shaking at 18 °C for another 16-18 h. ¹⁵N-labeled BOA for NMR experiments was produced using the same protocol, except that modified M9 medium containing ¹⁵NH₄Cl as the sole nitrogen source was used as the growth medium. Cells from each liter of culture were harvested by centrifugation, resuspended in 40 mL of buffer A [50 mM KH₂PO₄ (pH 8.0), 300 mM KCl, 10 mM imidazole, 4 mM DTT], and lysed using a microfluidizer. After clarifying the cell lysate by centrifugation, the supernatant was loaded onto a 5 mL Bio-Scale Mini Profinity IMAC Cartridge (Bio-Rad, Hercules, CA, USA) equilibrated in buffer A. His-tagged BOA was eluted from the column using a linear gradient of 0-50% buffer B [50 mM KH₂PO₄ (pH 8.0), 300 mM KCl, 500 mM imidazole, 4 mM DTT] and dialyzed overnight at 4 °C against 4 L of buffer C [50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 375 µM EDTA, 4 mM DTT], followed by hexahistidine tag cleavage by incubation with 1 mg of His-tagged TEV protease for 24 hours at 4 °C. The reaction mixture was again passed over the IMAC cartridge in order to separate newly tag-free BOA from tagged TEV protease and any residual tagged BOA. Finally, tag-free BOA was concentrated and passed over a HiLoad 26/60 Superdex 75 prepgrade gel filtration column (GE Healthcare, Piscataway, NJ, USA) equilibrated in buffer D [25 mM sodium acetate (pH 5), 25 mM NaCl, 4 mM DTT, 0.02% NaN₃]. Protein that was not immediately used was flash frozen in liquid nitrogen in buffer D supplemented with 10% v/v glycerol and stored at -80 °C until needed.

BOA crystallization

BOA was crystallized in CombiClover crystallization plates (Emerald Biosystems, Bainbridge Island, WA, USA) using the sitting drop vapor diffusion method. Diffraction-quality ligand-free BOA crystals were successfully grown at room temperature by mixing 2 µL of 6.6 mg/mL BOA in buffer D with 1 µL of a reservoir solution consisting of 0.2 M NaCl, 0.65 M (NH₄)₂HPO₄, 0.1 M sodium citrate (pH 5.5), and 10% v/v glycerol, followed

by equilibration against 500 μ L of reservoir solution. Diffraction-quality co-crystals of BOA bound to 3 α ,6 α -mannopentaose (Sigma-Aldrich) were grown at room temperature by exchanging BOA from buffer D into phosphate-buffered saline (PBS, pH 7.25), adding the sugar in PBS at a final ligand:protein molar ratio of 6:1, and mixing 2 μ L of the resulting 6.6 mg/mL protein solution with 1 μ L of a reservoir solution consisting of 1.2 M sodium citrate and 0.15 M sodium cacodylate (pH 6.5), followed by equilibration against 500 μ L of reservoir solution. Sugar-free and sugar-bound crystals each grew over a period of approximately 5-7 days.

X-ray diffraction data collection and structure solution

Sugar-free BOA crystals were cryoprotected by serially transferring the crystals into reservoir solution aliquots containing increasing concentrations of glycerol. After soaking for several minutes in the final reservoir solution containing 25% v/v glycerol, the crystals were flash frozen in liquid nitrogen.

Diffraction data (180 images of oscillation width 1° each) for sugar-free BOA were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22-ID at the Advanced Photon Source of Argonne National Laboratory. Sugar-bound BOA co-crystals were cryoprotected by soaking them for 30 s in a 50% saturated solution of sucrose in reservoir solution, followed by flash freezing in liquid nitrogen. Data for a ligand-bound BOA crystal (185 images of oscillation width 1° each) were collected in-house using a Rigaku FR-E SuperBright generator equipped with VariMax HR optics and an R-Axis HTC detector. Data processing and reduction steps for all data sets were carried out with the d*TREK package [34]. Phasing of the ligand-bound BOA data was accomplished by molecular replacement using Phaser [35], with the structure of ligand-free OAA (PDB 3S5V) serving as the search model. Initial automated model building was performed with ARP/wARP 7.2 [36], and subsequent cycles of manual model correction and refinement were carried out using Coot [37] and Phenix 1.7.3 [38], respectively. This step also included manual building and placement of the four bound ligands. The ligand-free BOA data were phased by molecular replacement in Phaser using the ligand-bound BOA protein structure determined in the present work as the phasing model. Initial model building was carried out using Phenix AutoBuild, and cycles of manual model correction and automatic refinement, including the application of noncrystallographic symmetry restraints, were performed using Coot and Buster 2.10.0, respectively. Data collection and refinement statistics for both structures are provided in Table 1. The structure factors and refined molecular models for both ligand-free and ligand-bound BOA were deposited at the PDB under accession codes **4GU8** and **4GK9**, respectively. All structural figures in this paper were created using CCP4mg [39].

Characterization of carbohydrate binding by NMR spectroscopy

To characterize the binding affinity of BOA for 3 α ,6 α -mannopentaose, ^1H - ^{15}N HSQC spectra were recorded at 25 °C with increasing amounts of glycan using a Bruker Avance 900 MHz NMR spectrometer equipped with a triple-resonance TCI cryoprobe and a z-axis pulsed field gradient coil. The concentration of ^{15}N -labeled BOA was 100 μM , and the sugar concentration ranged from 0-2721 μM . The ^1H and ^{15}N carriers were placed at 4.7 and 118 ppm, respectively, and the spectral widths were 15.4 ppm in ^1H and 34 ppm in ^{15}N . All NMR data were processed using NMRPipe and visualized using NMRDraw [40]. Glycan binding was in slow exchange on the chemical shift scale, and distinct ligand-bound and ligand-free resonances were observed for residues involved in ligand binding. The dissociation constant K_D was estimated based on the relative volumes of cross-peaks (populations of species) corresponding to the free and bound states and the known total amount of protein and ligand present.

Anti-HIV activity assays

BOA inhibition of HIV-1 infectivity was measured using a luciferase assay. TZM-bl cells maintaining a Tat-inducible luciferase reporter gene and expressing CD4, CCR5, and CXCR4 were added to a 96-well plate (10^4 cells in 100 μ L of Dulbecco's modified Eagle's medium per well). After 16 hours, the medium was aspirated, and wild type HIV (R9) virus (1 ng) and BOA (final concentrations ranging from 0-1000 nM) were added to each well. After incubation for 48 hours at 37 °C, 100 μ L of Steady-Glo luciferase substrate (Promega) were added to each well, and after 5 minutes luciferase activity was quantitated using a Packard TopCount NXT luminometer. Nonlinear regression analyses of the data were performed in SigmaPlot 10.0 and KaleidaGraph 4.02.

Acknowledgments

This project was supported by NIH Grant GM080642 to A.M.G. M.J.W. thanks Drs. Lisa Charlton, Palaniappa Arjunan, and Krishnamoorthy Chandrasekhar for stimulating discussions regarding crystallography. The authors acknowledge use of the Advanced Photon Source, which is supported under Contract No. W-31-109-Eng-38 by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences.

Abbreviations

BOA	<i>Burkholderia oklahomensis</i> EO147 agglutinin
CCR5	CC chemokine receptor type 5
CD4	cluster of differentiation 4 glycoprotein
CXCR4	CXC chemokine receptor type 4
DTT	dithiothreitol
IPTG	isopropyl β -D-1-thiogalactopyranoside
LB	lysogeny broth
MBHA	myxobacterial hemagglutinin
NCS	noncrystallographic symmetry
OAA	<i>Oscillatoria agardhii</i> agglutinin
PDB	protein data bank
PFA	<i>Pseudomonas fluorescens</i> agglutinin
TEV	tobacco etch virus

References

1. Das K, Clark AD Jr, Lewi PJ, Heeres J, De Jonge MR, Koymans LM, Vinkers HM, Daeyaert F, Ludovici DW, Kukla MJ, De Corte B, Kavash RW, Ho CY, Ye H, Lichtenstein MA, Andries K, Pauwels R, De Bethune MP, Boyer PL, Clark P, Hughes SH, Janssen PA, Arnold E. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *J Med Chem.* 2004; 47:2550–60. [PubMed: 15115397]
2. Markowitz M, Nguyen BY, Gotuzzo E, Mendo F, Ratanasuwan W, Kovacs C, Prada G, Morales-Ramirez JO, Crumpacker CS, Isaacs RD, Gilde LR, Wan H, Miller MD, Wenning LA, Tepller H. Rapid and durable antiretroviral effect of the HIV-1 Integrase inhibitor raltegravir as part of combination therapy in treatment-naive patients with HIV-1 infection: results of a 48-week controlled study. *J Acquir Immune Defic Syndr.* 2007; 46:125–33. [PubMed: 17721395]

3. Ghosh AK, Dawson ZL, Mitsuya H. Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV. *Bioorg Med Chem.* 2007; 15:7576–80. [PubMed: 17900913]
4. Alkhatib G, Berger EA. HIV coreceptors: from discovery and designation to new paradigms and promise. *Eur J Med Res.* 2007; 12:375–84. [PubMed: 17933717]
5. Klasse PJ. The molecular basis of HIV entry. *Cell Microbiol.* 2012; 14:1183–92. [PubMed: 22583677]
6. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature.* 1998; 393:648–59. [PubMed: 9641677]
7. Olshevsky U, Helseth E, Furman C, Li J, Haseltine W, Sodroski J. Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *J Virol.* 1990; 64:5701–7. [PubMed: 2243375]
8. Pollard SR, Rosa MD, Rosa JJ, Wiley DC. Truncated variants of gp120 bind CD4 with high affinity and suggest a minimum CD4 binding region. *EMBO J.* 1992; 11:585–91. [PubMed: 1537337]
9. Wild C, Greenwell T, Matthews T. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. *AIDS Res Hum Retroviruses.* 1993; 9:1051–3. [PubMed: 8312047]
10. Wild CT, Shugars DC, Greenwell TK, McDanal CB, Matthews TJ. Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc Natl Acad Sci U S A.* 1994; 91:9770–4. [PubMed: 7937889]
11. Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, Mori J, Rickett G, Smith-Burchnell C, Napier C, Webster R, Armour D, Price D, Stammen B, Wood A, Perros M. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob Agents Chemother.* 2005; 49:4721–32. [PubMed: 16251317]
12. Center RJ, Leapman RD, Lebowitz J, Arthur LO, Earl PL, Moss B. Oligomeric structure of the human immunodeficiency virus type 1 envelope protein on the virion surface. *J Virol.* 2002; 76:7863–7. [PubMed: 12097599]
13. Zhu P, Chertova E, Bess J Jr, Lifson JD, Arthur LO, Liu J, Taylor KA, Roux KH. Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc Natl Acad Sci USA.* 2003; 100:15812–7. [PubMed: 14668432]
14. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature.* 1998; 393:705–11. [PubMed: 9641684]
15. Sanders RW, Venturi M, Schiffner L, Kalyanaraman R, Katinger H, Lloyd KO, Kwong PD, Moore JP. The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *J Virol.* 2002; 76:7293–305. [PubMed: 12072528]
16. Trkola A, Purtscher M, Muster T, Ballaun C, Buchacher A, Sullivan N, Srinivasan K, Sodroski J, Moore JP, Katinger H. Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol.* 1996; 70:1100–8. [PubMed: 8551569]
17. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Phung P, Fling S, Wong CH, Phogat S, Wrin T, Simek MD, Koff WC, Wilson IA, Burton DR, Poignard P. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature.* 2011; 477:466–70. [PubMed: 21849977]
18. Barrientos LG, Gronenborn AM. The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini Rev Med Chem.* 2005; 5:21–31. [PubMed: 15638789]
19. Ji X, Olinger GG, Aris S, Chen Y, Gewurz H, Spear GT. Mannose-binding lectin binds to Ebola and Marburg envelope glycoproteins, resulting in blocking of virus interaction with DC-SIGN and complement-mediated virus neutralization. *J Gen Virol.* 2005; 86:2535–42. [PubMed: 16099912]
20. Li Y, Zhang X, Chen G, Wei D, Chen F. Algal lectins for potential prevention of HIV transmission. *Curr Med Chem.* 2008; 15:1096–104. [PubMed: 18473805]

21. Ziolkowska NE, Wlodawer A. Structural studies of algal lectins with anti-HIV activity. *Acta Biochim Pol.* 2006; 53:617–626. [PubMed: 17128290]
22. Tsai CC, Emau P, Jiang Y, Tian B, Morton WR, Gustafson KR, Boyd MR. Cyanovirin-N gel as a topical microbicide prevents rectal transmission of SHIV89.6P in macaques. *AIDS Res Hum Retroviruses.* 2003; 19:535–41. [PubMed: 12921090]
23. Tsai CC, Emau P, Jiang Y, Agy MB, Shattock RJ, Schmidt A, Morton WR, Gustafson KR, Boyd MR. Cyanovirin-N inhibits AIDS virus infections in vaginal transmission models. *AIDS Res Hum Retroviruses.* 2004; 20:11–8. [PubMed: 15000694]
24. Buffa V, Stieh D, Mamhood N, Hu Q, Fletcher P, Shattock RJ. Cyanovirin-N potentially inhibits human immunodeficiency virus type 1 infection in cellular and cervical explant models. *J Gen Virol.* 2009; 90:234–43. [PubMed: 19088294]
25. Sato Y, Murakami M, Miyazawa K, Hori K. Purification and characterization of a novel lectin from a freshwater cyanobacterium, *Oscillatoria agardhii*. *Comp Biochem Physiol B Biochem Mol Biol.* 2000; 125:169–77. [PubMed: 10817903]
26. Sato Y, Okuyama S, Hori K. Primary structure and carbohydrate binding specificity of a potent anti-HIV lectin isolated from the filamentous cyanobacterium *Oscillatoria agardhii*. *J Biol Chem.* 2007; 282:11021–9. [PubMed: 17314091]
27. Koharudin LM, Furey W, Gronenborn AM. Novel fold and carbohydrate specificity of the potent anti-HIV cyanobacterial lectin from *Oscillatoria agardhii*. *J Biol Chem.* 2011; 286:1588–97. [PubMed: 20961847]
28. Koharudin LM, Gronenborn AM. Structural basis of the anti-HIV activity of the cyanobacterial *Oscillatoria Agardhii* agglutinin. *Structure.* 2011; 19:1170–81. [PubMed: 21827952]
29. Glass MB, Steigerwalt AG, Jordan JG, Wilkins PP, Gee JE. *Burkholderia oklahomensis* sp. nov., a *Burkholderia pseudomallei*-like species formerly known as the Oklahoma strain of *Pseudomonas pseudomallei*. *Int J Syst Evol Microbiol.* 2006; 56:2171–6. [PubMed: 16957116]
30. Prli A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, Bourne PE. Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics.* 2010; 26:2983–5. [PubMed: 20937596]
31. Koharudin LM, Kollipara S, Aiken C, Gronenborn AM. Structural insights into the anti-HIV activity of the *Oscillatoria agardhii* agglutinin homolog lectin family. *J Biol Chem.* 2012; 287:33796–811. [PubMed: 22865886]
32. Maiti R, Van Domselaar GH, Zhang H, Wishart DS. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.* 2004; 32:W590–4. [PubMed: 15215457]
33. Koharudin LM, Kollipara S, Aiken C, Gronenborn AM. Structural insights into the anti-HIV activity of the *Oscillatoria agardhii* agglutinin homolog lectin family. *J Biol Chem.* 2012
34. Pflugrath JW. The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr.* 1999; 55:1718–1725. [PubMed: 10531521]
35. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007; 40:658–674. [PubMed: 19461840]
36. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* 2008; 3:1171–9. [PubMed: 18600222]
37. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–32. [PubMed: 15572765]
38. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:213–21. [PubMed: 20124702]
39. McNicholas S, Potterton E, Wilson KS, Noble ME. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr.* 2011; 67:386–94. [PubMed: 21460457]

40. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–93. [PubMed: 8520220]

FIGURE 1.

Schematic representation of the structure of Man-9. The glycosidic linkages between each sugar residue are indicated. M and GN stand for mannose and N-acetylglucosamine residues, respectively, while the designations D1-3 denote the individual arms of this triantennary sugar. The five mannosyl residues in solid circles represent 3 α ,6 α -mannopentaose, the core structure of Man-9 that was co-crystallized with BOA in the present work.

FIGURE 2.

Multiple sequence alignment of OAA family members. BOA and MBHA contain two sugar-binding domains, whereas OAA and PFA are single-domain proteins. Residues highlighted in blue are absolutely conserved among these lectins. Amino acids that interact with 3α6α-mannopentose are underlined in red, cyan, magenta, and green, respectively, for binding sites I, II, III, and IV. Note that the binding site residues belong to the absolutely conserved subset in the alignment.

FIGURE 3.

The refined co-crystal structure of monomeric BOA bound to 4 molecules of 3 α ,6 α -mannopentaose. The view in panel B involves rotation around the x -axis by approximately 45° from that in panel A. Note that the long axes of the two barrels are rotated relative to each other by approximately 85°.

FIGURE 4.

Summary of the intermolecular interactions involved in the recognition of 3 α ,6 α -mannopentaose by BOA. The binding site III structure is used as a representative example. Panel A displays the experimental 2F_o-F_c electron density contoured at 1 σ for the pentasaccharide (magenta) and the protein residues that interact with it (light blue). The ligand density was readily interpretable in all four binding sites. Protein residues that form hydrogen bonds with the ligand are labeled, and hydrogen bonds are represented by dashed thick black lines. Panel B displays van der Waals surfaces, illustrating the packing interaction between M3, the reducing end of the ligand, and the indole ring of Trp152. The space-filling model in panel C reveals the surface cleft in which the sugar rests, with the interacting residues colored light blue. Panel D displays a superposition of the four sugars bound to BOA. The sugars are color-coded according to the binding site in which they are located using the same color scheme as in Figure 2. Note that the sugar rings that interact with the protein (M3, M4', and M5') exhibit less conformational variability than those that do not interact with the protein.

FIGURE 5.

Structural comparison of the ligand-bound forms of BOA and MBHA. BOA and its bound glycans are shown in gray, and MBHA and its two bound sugars are colored salmon and red, respectively. Although the secondary structures overlay extremely well, there is some variability in the loops between BOA and MBHA.

FIGURE 6.

Superposition of the eight noncrystallographically related BOA molecules found in the asymmetric unit of the ligand-free crystal. The only noticeable areas of conformational heterogeneity are shaded in red, and these loop regions contain the conserved WGG motif involved in binding the ligand in each binding site (panel A). Panels B and C provide enlarged views of the key tryptophan side chain (colored cyan) in two different binding sites. Trp18 (panel B) exhibits some conformational variability within a single rotameric state in the absence of ligand. This also holds for Trp218. Both Trp85 (panel C) and Trp152 adopt two distinct rotamers.

FIGURE 7.

Representative NMR data used to estimate the dissociation constant K_D for 3 α ,6 α -mannopentaose binding to BOA. Panel A shows the ^1H - ^{15}N HSQC spectrum of BOA. The black spectrum, recorded in the absence of ligand, represents the starting point of the titration, while the red spectrum, recorded in the presence of a saturating excess of ligand, represents the end of the titration. Panels B-F illustrate the titration of the boxed signal located at (8.25, 131) ppm in panel A. In panels B-F, the sugar:protein molar ratios are 0, 1.45, 2.5, 4.0, and 27.2, respectively. With increasing sugar, the signal from free BOA (left cross-peak in each panel) decreases in intensity, while the signal from sugar-bound protein (right cross-peak) grows in intensity. The K_D was estimated from the volumes of the cross-peaks in each pair that were fitted to Gaussian profiles using NMRPipe utilities.

FIGURE 8.

Nonlinear regression analysis of BOA's anti-HIV activity. Data points from four replicate experiments are shown in black, red, green, and yellow, respectively. The fitted IC_{50} value for each curve is given in the figure, with an average of all the experimental values of 12.2 ± 0.7 nM.

Table I

Data Collection and Refinement Statistics

	Ligand-Free BOA	Ligand-Bound BOA
Data Collection		
Space Group	$P2_12_12_1$	$P4_32_12$
Unit Cell Dimensions		
a, b, c (Å)	119.72, 130.58, 149.99	75.62, 75.62, 132.32
α, β, γ (°)	90, 90, 90	90, 90, 90
Data Collection Wavelength (Å)	1.0	1.5418
Unique Reflections	91,789	30,821
Redundancy	7.0	12.1
Resolution (Å)	18.41-2.40	38.10-1.90
Mosicity (°)	0.48	0.44
R_{merge}	0.068	0.080
$I/\sigma(I)$	13.0(4.1) ^a	17.8(3.6) ^b
Completeness (%)	99.5(99.1) ^a	99.4(92.7) ^a
Refinement		
Resolution (Å)	2.40	1.90
R_{work} ^b	0.194(0.262) ^a	0.165(0.268) ^a
R_{free} ^b	0.229(0.329) ^a	0.193(0.331) ^a
No. Atoms in ASU		
Protein	16,248	2,076
Ligands/Ions	38	230
Water	799	403
Average B -factors		
Protein	55.9	25.3
Ligands/Ions	52.2	38.4
Water	54.0	39.9
R.M.S. Deviations		
Bond Lengths (Å)	0.010	0.011
Bond Angles (°)	1.09	1.07
MolProbity Statistics		
Ramachandran Favored (%)	98.3	98.6
Ramachandran Allowed (%)	1.6	1.4
Ramachandran Outlier (%)	0.1	0.0
All-Atom Clashscore	3.8	1.6
PDB Accession Code	4GU8	4GK9

^aValues in parentheses refer to the highest-resolution shell.

^bR-factors were calculated according to $R = \Sigma ||F_{\text{Obs}}| - |F_{\text{Calc}}|| / \Sigma |F_{\text{Obs}}|$. The free sets consisted of 5% of the total number of measured reflections, and the free reflections were not used in any stage of structure refinement.

Table IIDetails of the interactions between BOA and 3 α ,6 α -mannopentaose in all four binding sites

Protein	Ligand	Interaction Type
<i>Binding Site I</i>		
W18 indole	M3 ring	Packing
G19 amide	M5' O5, M5' O6	BB H-bonds
G20 amide	M5' O6	BB H-bond
R103 N η 2	M3 O3, M3 O4	SC H-bonds
E131 Oe1	M3 O4	SC H-bond
G132 amide	M4' O5, M4' O6	BB H-bonds
P133 O	M4' O2	BB H-bond
<i>Binding Site II</i>		
W85 indole	M3 ring	Packing
G86 amide	M5' O5, M5' O6	BB H-bonds
G87 amide	M5' O6	BB H-bond
R36 N η 2	M3 O3, M3 O4	SC H-bonds
E64 Oe1	M3 O4	SC H-bond
G65 amide	M4' O5, M4' O6	BB H-bonds
P66 O	M4' O2	BB H-bond
<i>Binding Site III</i>		
W152 indole	M3 ring	Packing
G153 amide	M5' O5, M5' O6	BB H-bonds
G154 amide	M5' O6	BB H-bond
R236 N η 2	M3 O3, M3 O4	SC H-bonds
E264 Oe1	M3 O4	SC H-bond
G265 amide	M4' O5, M4' O6	BB H-bonds
P266 O	M4' O2	BB H-bond
<i>Binding Site IV</i>		
W218 indole	M3 ring	Packing
G219 amide	M5' O5, M5' O6	BB H-bonds
G220 amide	M5' O6	BB H-bond
R170 N η 2	M3 O3, M3 O4	SC H-bonds
E197 Oe1	M3 O4	SC H-bond
G198 amide	M4' O5, M4' O6	BB H-bonds
P199 O	M4' O2	BB H-bond

BB and SC refer to backbone and side-chain H-bonds, respectively.