Novel Fold and Carbohydrate Specificity of the Potent Anti-HIV Cyanobacterial Lectin from Oscillatoria agardhii*§
Leonardus M. I. Koharudin‡, William Furey§§, and Angela M. Gronenborn*§
From the Departments of ‡§Structural Biology and §Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260 and the *Biocrystallography Laboratory, Veterans Affairs Medical Center, Pittsburgh, Pennsylvania 15240

Oscillatoria agardhii agglutinin (OAA) is a recently discovered cyanobacterial lectin that exhibits potent anti-HIV activity. Up to now, only its primary structure and carbohydrate binding data have been available. To elucidate the structural basis for the antiviral mechanism of OAA, we determined the structure of this lectin by x-ray crystallography at 1.2 Å resolution and mapped the specific carbohydrate recognition sites of OAA by NMR spectroscopy. The overall architecture of OAA comprises 10 β-strands that fold into a single, compact, β-barrel-like domain, creating a unique topology compared with all known protein structures in the Protein Data Bank. OAA sugar binding was tested against Man-9 and various disaccharide components of Man-9. Two symmetric carbohydrate-binding sites were located on the protein, and a preference for Manα(1–6)Man-linked sugars was found. Altogether, our structural results explain the antiviral activity OAA and add to the growing body of knowledge about antiviral lectins.

HIV infection occurs via virus-cell and cell-cell fusion mediated by the two viral envelope glycoproteins gp120 and gp41 (1–3). gp120 interacts with the CD4 receptor of the host cell, resulting in a conformational change in gp120 that eventually leads to the insertion of the fusion peptide of gp41 into the target membrane, causing membrane fusion (4). The gp120 glycoprotein is remarkably enriched in high mannose N-linked sugars (5), and novel avenues for controlling HIV infection may become available by targeting to the sugars of gp120.

Various lectins, including cyanovirin-N (6), DC-SIGN (7), scytovirin (8), griffithsin (9), MVL (10), and actinohivin (11), are known to bind to the high mannose glycans on gp120, thereby exerting anti-HIV activity. Interestingly, the binding modes and target epitopes on Man-9 are quite distinct for the different lectins. Cyanovirin-N specifically recognizes Manα(1–2)Man-linked mannose substructures, in particular the D1 and D3 arms of Man-9 (12, 13); DC-SIGN preferentially interacts with the Manα(1–3)Manα(1–6)Man trisaccharide (14); griffithsin binds to single mannose units (15); and MVL specifically interacts with the Manα(1–6)Manβ(1–4)GlcnAcβ(1–4)GlcnAc tetrasaccharide (16). Although no crystal structures for protein-carbohydrate complexes are available for scytovirin and actinohivin, sugar binding studies revealed specificities for Manα(1–2)Manα(1–6)Man (17) and Manα(1–2)Man (18), respectively.

Here, we determined the crystal structure of Oscillatoria agardhii agglutinin (OAA), a recently discovered cyanobacterial lectin with potent anti-HIV activity (19); mapped the carbohydrate-binding sites on the protein by NMR spectroscopy; and determined its oligosaccharide specificity. Our results demonstrate that OAA is unique with respect to both its structure and specific carbohydrate binding. Altogether, this work provides the molecular basis for understanding the potent anti-HIV activity of OAA and may aid in its further development as a useful diagnostic and pharmacological reagent in the fight against HIV transmission.

EXPERIMENTAL PROCEDURES

OAA Expression, Purification, and Crystallization—A synthetic OAA gene encoding residues 1–133 (20) was cloned into the pET-26b(+) expression vector (Novagen) using NdeI and XhoI restriction sites at the 5′- and 3′-ends, respectively. For protein expression, Escherichia coli Rosetta 2(DE3) cells (Novagen) were transformed with the pET-26b(+)–OAA vector. Cells were initially grown at 37 °C, induced with 1 mM isopropyl β-D-thiogalactopyranoside at 16 °C, and grown for ~18 h at 16 °C for protein expression. Isotopic labeling of the protein for NMR studies was carried out by growth in modified M9 minimal medium containing 15NH4Cl and/or [13C]glucose as the sole nitrogen and/or carbon source, respectively. For SeMet labeling, cells were grown in modified M9 minimal medium, and SeMet was added to the culture 1 h before induction at 100 mg/liter. For unlabeled samples, cells were cultured in LB medium.

Protein was prepared from the soluble fraction of E. coli after opening the cells by sonication, removal of cell debris by centrifugation, and dialyzing the supernatant overnight.

* This work was supported, in whole or in part, by National Institutes of Health Grant GM080642 (to A. M. G.).
† The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S4.
‡ The atomic coordinates and structure factors (code 3OBL) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).
§ To whom correspondence should be addressed: Dept. of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., BST3/Rm. 1050, Pittsburgh, PA 15260. Tel.: 412-648-9959; Fax: 412-648-9008; E-mail: amg100@pitt.edu.

The abbreviations used are: OAA, O. agardhii agglutinin; CAPS, N-cyclohexyl-3-aminopropanesulfonic acid; MAD, multiwavelength anomalous dispersion; HSO4-, heteronuclear single quantum coherence; QAD, quino-hemoprotein amine dehydrogenase; CHES, 2-(cyclohexylamino)ethanesulfonic acid.

2 The abbreviations used are: OAA, O. agardhii agglutinin; CAPS, N-cyclohexyl-3-aminopropanesulfonic acid; MAD, multiwavelength anomalous dispersion; HSO4-, heteronuclear single quantum coherence; QAD, quino-hemoprotein amine dehydrogenase; CHES, 2-(cyclohexylamino)ethanesulfonic acid.
against 20 mM Tris-HCl buffer (pH 8.5). Further purification involved anion-exchange chromatography on a Q HP column (GE Healthcare) using a linear gradient of NaCl (20–1000 mM) for elution, followed by gel filtration on Superdex 75 (GE Healthcare) in 50 mM sodium acetate, 100 mM NaCl, and 3 mM NaN₃ (pH 5.0). Purified protein fractions were collected and concentrated up to 40 mg/ml using Centriprep devices (Millipore). For crystallization and NMR, the buffer was exchanged to 20 mM Tris-HCl buffer, 100 mM NaCl, and 3 mM NaN₃ (pH 8.0) and with 20 mM sodium acetate, 3 mM NaN₃, and 90:10% H₂O/D₂O (pH 5.0), respectively.

Crystallization trials were carried out by the sitting-drop vapor diffusion method at room temperature using drops consisting of 2 μl of protein and 2 μl of reservoir solutions. Well diffracting crystals were obtained in 1.2 M NaH₂PO₄/0.8 M K₂HPO₄ (pH 5.5), 0.2 M Li₂SO₄, and 0.1 M CAPS (pH 10.5) and in 2.0 M (NH₄)₂SO₄ (pH 5.4), 0.2 M Li₂SO₄, and 0.1 M CAPS (pH 10.5) after 1 day. The SeMet derivative crystals were obtained under the same crystallization conditions as the native crystals. Note that the final pH of the protein in the crystallization drops that yielded crystals for high resolution data collection under the conditions of 1.2 M NaH₂PO₄/0.8 M K₂HPO₄ (pH 5.5), 0.2 M Li₂SO₄, and 0.1 M CAPS (pH 10.5) (see below) was ~6.3.

**Diffraction Data Collection and Structure Determination of OAA—X-ray diffraction data for the SeMet derivative and native crystals were collected at the SER-CAT facility sector 22-BM beam line of the Advance Photon Source at Argonne National Laboratory (Chicago, IL). The multiwavelength anomalous dispersion (MAD) data were collected at wavelengths corresponding to the leading edge, peak, and a high energy remote point of the anomalous scattering plot for selenium (0.9795, 0.9793, and 0.9718 Å, respectively). All diffraction data used for analysis were collected from the crystals grown in 1.2 M NaH₂PO₄/0.8 M K₂HPO₄ (pH 5.5), 0.2 M Li₂SO₄, and 0.1 M CAPS (pH 10.5) (see below) was ~6.3.

**NMR Spectroscopy—**NMR spectra were recorded at 25 °C on Bruker AVANCE 800, AVANCE 700, and AVANCE 600 spectrometers equipped with 5-mm triple-resonance, three-axis gradient probes or x axis gradient cryoprobes. Spectra for backbone resonance assignments were recorded on a 13C/15N-labeled sample in 20 mM sodium acetate, 3 mM NaN₃, and 90:10% H₂O/D₂O (pH 5.0). The protein concentration was similar to the concentration used for crystallization at ~1.0 mg (~40 mg/ml). Three-dimensional HNCACB, CBCA(CO)NH, and 1H-15N NOESY heteronuclear single quantum coherence (HSQC; mixing time of 120 ms) experiments (26, 27) were recorded for complete backbone chemical shift assignment. All spectra were processed with NMRPipe (28) and analyzed using NMRView (29).

**Carbohydrate Binding Studies by NMR Spectroscopy—**Binding of Man-9 (V-Labs) was investigated at 25 °C using 0.020 mM 15N-labeled OAA in 20 mM sodium acetate, 3 mM NaN₃, and 90:10% H₂O/D₂O (pH 5.0) by 1H-15N HSQC spectroscopy at 600 MHz. Because Man-9 is not readily available and very expensive, spectra were recorded for only two titration points at protein/Man-9 molar ratios up to 1:20. Two-dimensional 1H-15N HSQC spectra were recorded after each addition of carbohydrate.

**RESULTS**

**OAA Three-dimensional Structure—**The atomic structure of OAA was determined by x-ray crystallography for the protein comprising Ala-2–Thr-133 (Fig. 1A). Although the protein starts at Ala-2, with Met-1 being completely removed by the E. coli N-terminal methionine aminopeptidase during protein expression (verified by NMR and mass spectrometry), we kept the numbering according to Sato and Hori (20) for consistency.

Initial indexing of diffraction data from the SeMet analog crystal was for the tetragonal system, space group I₄₁, with unit cell dimensions of a = 59.55 and c = 42.66 Å. Although the data processed well in this system with a reasonable merging R value, only half of a protein molecule per asymmetric unit could be accommodated, as a full molecule leads to an extremely low or negative solvent content based on a Matthews probability calculator. Thus, the tetragonal space group was physically impossible. However, the fact that the data merged and scaled well in this space group, permitting only half of a molecule in an asymmetric unit, implied that the protein molecule itself must have a high degree of internal 2-fold symmetry. This was not totally surprising given the high sequence similarity between the two halves of the OAA sequence, with ~77% identity (51/66 residues) and ~86% similarity (57/66 residues) (Fig. 1A).

As expected from the Matthews probability calculation, solution of the structure via the MAD tetragonal data at 1.50 Å resolution revealed a symmetric molecule with half of a chain per asymmetric unit (supplemental Fig. S1A) and disorder at places with sequence differences. To determine the entire molecular structure, diffraction data were collected up
to 1.10 Å resolution from a native crystal and were indexed in lower symmetry space groups, both the monoclinic C2 space group, with unit cell dimensions of $a = 84.21$, $b = 42.77$, and $c = 59.58$ Å and $\beta = 134.88^\circ$, and the triclinic $P1$ space group, with unit cell dimensions of $a = 42.75$, $b = 47.28$, and $c = 47.38$ Å and $\alpha = 78.16^\circ$, $\beta = 62.99^\circ$, and $\gamma = 63.14^\circ$. The half-molecule determined from the MAD data was then utilized as a molecular replacement probe using the program Phaser (30) for the native crystal in both the monoclinic C2 and triclinic $P1$ space groups, and apparent molecular replacement solutions were readily obtained in both cases. Note that two and four half-molecules are expected to be present in the monoclinic C2 and triclinic $P1$ asymmetric units, respectively. Surprisingly, the two half-molecules located in the monoclinic C2 space group asymmetric unit are arranged as two independent units (supplemental Fig. S1B) instead of forming an intact protein chain. The other half-molecules that complete the entire cell are symmetry-related (supplemental Fig. S1C) and cannot be connected to form an intact single chain, even when all half-molecules, symmetry operations, and cell translations are considered. On the other hand, for the molecular replacement solution using the high resolution data in the triclinic cell, four half-molecules were indeed found that properly paired up into two intact molecules, likely constituting the correct solution. Upon electron density map examination, this map also clearly revealed the few expected sequence differences in the two halves of the molecule, thus confirming the structure. This new model was ultimately refined using the REFMAC program (31) in the CCP4 package (22) to a resolution of 1.2 Å, with working and free $R$ factors of 14.0 and 16.9%, respectively (Table 1).

The structure exhibits strong and clear electron density for all residues, with two OAA molecules and four CAPS molecules present in the asymmetric unit. The two polypeptide chains are identical, with average backbone and heavy atom root mean square deviations of 0.04 and 0.24 Å, respectively (supplemental Fig. S2). Two CAPS molecules are in contact with each protein molecule (Fig. 1A). The final model has 98.1 and 100% of all residues in the favored and allowed regions of the Ramachandran plot, respectively, and contains no residues in the disallowed region as evaluated by MolProbity (32). Because of the high resolution, the electron density clearly reveals two conformations for several side chains (Leu-3, Ile-25, Ser-27, Ser-29, Gln-45, Thr-46, Thr-50, Asn-89, Ser-94, Val-105, Thr-113, Thr-117, and Met-118 in monomer A and Leu-3, Asn-5, Gln-22, Glu-24, Ile-25, Gln-45, Thr-46, Thr-50, Asn-68, Asn-89, Gln-112, Thr-113, Thr-117, and Thr-119 in monomer B). Residues with three alternative side chain conformations were observed for Ser-70 in monomer A and Ser-27, Ser-29, Ser-70, and Ser-94 in monomer B.

The overall architecture of OAA is a compact $\beta$-barrel-like domain, comprising a continuous 10-stranded antiparallel $\beta$-sheet (Fig. 1C). Each of the amino acid sequence repeats (Fig. 1A) folds into five $\beta$-strands, denoted as $\beta1$–$\beta5$ (colored blue) and $\beta6$–$\beta10$ (colored light gray) for the first and second repeats, respectively. A very short linker comprising Gly-67–Asn-69 connects these two sequence repeats (colored orange). Within each sequence repeat, the linkers connecting strands $\beta2$ and $\beta3$ and strands $\beta7$ and $\beta8$, respectively, pass across the top or bottom of the barrel. Therefore, the first two $\beta$-strands of each sequence repeat ($\beta1$–$\beta2$ and $\beta6$–$\beta7$) and the next three $\beta$-strands ($\beta3$–$\beta4$–$\beta5$ and $\beta8$–$\beta9$–$\beta10$) are positioned on opposite sides of the barrel (Fig. 1, C and D). In this manner, the two $\beta$-strands from the first sequence repeat ($\beta1$–$\beta2$) are positioned between strands $\beta6$ and $\beta7$ on one side and strands $\beta8$, $\beta9$, and $\beta10$ on the other side of the second se-
sequence repeat (Fig. 1C). Similarly, strands β3, β4, and β5 are located between two and three β-strands, respectively, of the other sequence repeat (Fig. 1C). The swap of β-strands between the two sequence repeats creates an almost perfect C2 symmetric arrangement, with the conformation of the five β-strands in each sequence repeat being extremely similar (Fig. 1D).

The β-barrel structure is stabilized through numerous aliphatic side chain interactions on the inside of the barrel and contributing to the hydrophobic core of the protein (Fig. 2). Tyr-4, Val-6, Trp-23, Ile-25, Val-33, Ile-36, Val-38, Leu-47, Met-51, Tyr-53, Ile-59, Phe-61, Ala-63, and Leu-64 from the first sequence repeat (Fig. 2, left) interact with Tyr-71, Val-73, Trp-90, Leu-92, Val-100, Ile-103, Val-105, Leu-114, Met-118, Tyr-120, Ile-126, Phe-128, and Leu-132 from the second sequence repeat (Fig. 2, right). Apart for two positions (Ile-25/Leu-92 and Ala-63/Gly-130), these residues are identical in the two sequence repeats, and they also exhibit matching rotamer conformations (the same χ1 and χ2 angles). Therefore, residue conservation clearly correlates with structure conservation in OAA. Indeed, this high similarity explains the apparent higher symmetry space group during indexing of the crystallographic data (see above).

Comparison of OAA with Other Protein Structures Reveals Its Novel Fold—An automated search of the Protein Data Bank using the entire structure (the 10-stranded β-barrel) (see Fig. 1C) or the structure of an individual sequence repeat (the chain of five consecutive β-strands) (see Fig. 1D) with the program DALI (33) for related structures did not uncover any proteins with significant overall structural similarity (Z

---

**TABLE 1**

Data collection and phasing and refinement statistics for OAA

<table>
<thead>
<tr>
<th>Data collection</th>
<th>SeMet derivative</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>I4₁</td>
<td>I4₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>59.55, 59.55, 42.66</td>
<td>59.55, 59.55, 42.66</td>
</tr>
<tr>
<td>α, β, γ</td>
<td>90°, 90°, 90°</td>
<td>90°, 90°, 90°</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>59.55, 59.55, 42.66</td>
<td>59.55, 59.55, 42.66</td>
</tr>
<tr>
<td>a, β, γ</td>
<td>90°, 90°, 90°</td>
<td>90°, 90°, 90°</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>22.59-1.50</td>
<td>22.59-1.50</td>
</tr>
<tr>
<td>Rmerge (Å)</td>
<td>0.125</td>
<td>0.119</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.6</td>
<td>99.6</td>
</tr>
<tr>
<td>Redundancy</td>
<td>14.15</td>
<td>14.14</td>
</tr>
</tbody>
</table>

Refinement

| Resolution (Å)                | 42.00-1.20 (1.23-1.20) |
| No. of reflections            | 77,522 |
| Rmerge (Å)                    | 0.140/0.169 (0.210/0.255) |
| No. of atoms                  | 2124 |
| Protein                       | 56 |
| Ligand/ion                    | 344 |
| Water                         | 17.50 |
| B-factors                     | 22.77 |
| r.m.s.d. (Å)                  | 31.97 |
| Bond angle (Å)                | 0.013 |

* Data were obtained from the best diffracting crystal according to crystalization conditions (see “Results” for details).
* The MAD data in the tetragonal I4₁ space group were used as a molecular replacement probe using the program Phaser for the native crystal.
* Refinement was carried out using REFMAC5.
* Values in parentheses are for the highest resolution shell.
* r.m.s.d., root mean square deviation.
score > 4.3). This may not be surprising given the lack of amino acid sequence similarity to any known sequences for which structures are available. The closest match for the entire structure of OAA is the quinohemoprotein amine dehydrogenase (QAD; residues 173–272; Protein Data Bank code 1JJU) (34), for which 67 residues can be superimposed onto the domain of OAA with a Ca atomic root mean square deviation of ∼2.7 Å (Fig. 3A). The other top three structures predicted by DALI are rhizavidin and structures of avidin homologs (supplemental Fig. S3). Their folds are similar to that of QAD but clearly somewhat different from that of OAA.

We also submitted the OAA structure to the PDBeFold server, searching for homologous structures. Again, no significant match was found (Q score < 0.14). Similar to the results of the DALI search, the closest match found by PDBeFold was rhizavidin (Protein Data Bank code 3EW1) and other avidin homologs (supplemental Fig. S3).

Although the overall β-barrel shape of OAA is similar to that of QAD, the secondary structure topology is clearly different. In QAD, the eight antiparallel β-strands are arranged contiguously in space, without any crossover loops between the strands (Fig. 3B). In OAA, the 10 antiparallel β-strands contain two crossovers: between strands β2 and β3 and be-

FIGURE 3. Comparison of the OAA and QAD folds. A, ribbon diagram of QAD (residues 173–272), the closest match to the OAA fold according to the DALI program (33). B, schematic view of the connectivities between the eight β-strands in QAD. C, ribbon diagram of OAA. D, schematic view of the connectivities between the 10 β-strands in OAA. E, top view of the OAA β-strand connectivities. The solid and dashed connecting lines are loops that cross over the top and bottom of the barrel. The topology diagrams (B, D, and E) are colored according to the ribbon representations in A and C.
between strands $\beta 3$ and $\beta 8$. Thus, strand $\beta 3$ is flanked by strands $\beta 4$ and $\beta 8$, and strand $\beta 7$ is flanked by strands $\beta 2$ and $\beta 6$ (Fig. 3D). A top view of the 10 antiparallel $\beta$-strands in OAA is shown in Fig. 3E. Loops that cross over the top and bottom of the $\beta$-barrels are structural hallmarks of the jellyroll fold found in many lectins. Therefore, even if no close match to previously determined lectin structures is found, the general architecture is clearly of the lectin jellyroll type. However, given the different connectivities between the $\beta$-strands in OAA, we believe that the particular secondary structure arrangement seen here is novel and has not been observed for any lectin to date.

Structural Basis for Man-9 Binding and Carbohydrate Specificity of OAA—The two-dimensional $^1$H-$^15$N HSQC spectrum of OAA (Fig. 4) exhibited well dispersed and narrow resonances indicative of a native folded structure. Complete backbone assignments were obtained using three-dimensional HNCA,CB, CBCA(CO)NH, and $^1$H-$^15$N NOESY HSQC spectra. All expected amide backbone resonances were observed, except for Asn-69, which is broad and cannot be detected under the conditions described under “Experimental Procedures.” Note that the two amide resonances of Gly-26 and Gly-93 are unusually high field-shifted proton resonance frequencies (Fig. 4, inset). With the structure in hand and NMR assignments available, it is possible to directly investigate Man-9 binding and the specificity of OAA for carbohydrates using $^1$H-$^15$N HSQC NMR titrations. Titration data sets were recorded for $^{15}$N-labeled OAA at 0.020 mM, and spectra in the absence and presence of $0.016$ mM Man-9 (OAA/Man-9 molar ratio of 1:0.8) and $0.048$ mM Man-9 (OAA/Man-9 molar ratio of 1:2.4) are provided in Fig. 5. As reported previously (19), OAA comprises two carbohydrate-binding sites. Therefore, Man-9/binding site ratios of 0:1 (free), 0:4:1, and 1:2.1 were used.

Chemical shift mapping of $^1$H and $^15$N resonances for free and Man-9-bound OAA allows direct delineation of the ligand-binding sites on the protein as well as determination of apparent binding affinity. As evidenced by the data in Fig. 5...
Man-9-bound OAA at 1:0.8 (Fig. 5A), many more resonances exhibit changes, and those that were already shifted earlier (Gly-26, Gly-60, Arg-62, and Gln-76) now exhibit complete saturation. However, the later titrating resonances have not reached saturation yet, such as Gly-93, Gly-127, Lys-129, and Gln-9. They are the equivalent residues to Gly-26, Gly-60, Arg-62, and Gln-76, respectively. This clearly shows a different behavior induced by Man-9 binding, depending on whether one or both sites are used on the protein. The data also show that the Man-9 affinities for the two binding sites on OAA are distinct: the lower affinity site, site 1 (affected only at high Man-9/protein ratios), comprises the loops connecting strands β1-β2, β7-β8, and β9-β10, and the higher affinity site, site 2, is located at the symmetrically related position and is made up by the loops connecting strands β6-β7, β2-β3, and β4-β5. A qualitative estimation of the binding affinity for the higher affinity site (site 2) yields a $K_d$ value of $\sim 10 \mu M$. Note that residues affected by Man-9 binding at the 1:2.4 molar ratio include Glu-7–Glu-24, Gly-26, Arg-28, Val-33–Ala-35, Thr-50–Leu-65, Glu-72–Ile-91, Gly-93, Arg-95, Val-100–Ala-102, and Thr-117–Lys-129.

It was reported previously that OAA interacts with high mannose glycans comprising a nonreducing terminal mannose and the overall structure Manα(1–3)Manβ(1–4)GlcNAcβ(1–4)GlcNAc (19). As described above, we showed by direct NMR titrations which amino acids of OAA interact with Man-9. To further delineate which of the various saccharide linkages in Man-9 is recognized by OAA, further titrations against the disaccharides Manα(1–2)Man, Manα(1–3)Man, Manα(1–6)Man, and GlcNAcβ(1–4)GlcNAc and the trisaccharide GlcNAcβ(1–4)GlcNAcβ(1–4)GlcNAc were carried out. No binding to Manα(1–2)Man, Manα(1–3)Man, GlcNAcβ(1–4)GlcNAc, and GlcNAcβ(1–4)GlcNAcβ(1–4)GlcNAc was detected (supplemental Fig. S4), even at protein/carbohydrate molar ratios of 1:20. However, OAA interacts with the Manα(1–6)Man disaccharide (Fig. 5C), albeit with low affinity (fast exchange on the chemical shift scale). Residues whose resonances are perturbed by Manα(1–6)Man include Trp-10, Gly-11, Arg-28, Val-34, Met-51, Tyr-53, Gly-60, Trp-77, Gly-78, Ser-86, Met-118, Tyr-120, and Gly-127. All of these are also perturbed upon Man-9 binding, establishing that OAA appears to preferentially recognize the Manα(1–6)Man linkage.

Structural mapping of the Man-9-interacting residues onto the structure of OAA revealed two areas on opposite ends of the molecule (Fig. 6A). One is located at the top and the other at the bottom in the space-filling representation shown in Fig. 6A. All residues involved in Man-9 binding (see above) are colored blue and magenta, with the magenta ones also being affected by the Manα(1–6)Man disaccharide. This subset of residues lies along the short narrow clefts within the two binding surfaces, possibly explaining why only the Manα(1–6)Man disaccharide and not other possible disaccharides of Man-9 can be recognized by OAA.
DISCUSSION

We have determined the 1.2 Å x-ray structure of OAA, an HIV-inactivating cyanobacterial lectin. In contrast to previous suggestions that implied a similar structure to bacterial lectins or to those of cyanobacteria or marine algae, such as Eucheuma serra agglutinin 2 (ESA-2) and Myxobacterium haemagglutinin (MBHA) (19), the three-dimensional structure of OAA is clearly distinct from any known lectins. Indeed, searches for fold similarity using the DALI server failed to uncover any homologs. Therefore, to the best of our knowledge, OAA can be regarded a novel lectin fold. The structural novelty of OAA is intimately related to the unique tandem primary sequence repeat of OAA.

Inspection of the OAA structure reveals that each sequence repeat does not constitute an individual domain. This is similar to what was found for other sequence repeat-containing lectins, such as cyanovirin-N and members of the CVNH family (35–37). However, the exchange of β-strands between the two sequence repeats is quite different in the two cases. In OAA, the interchange of β-strands creates a β-barrel in which all neighboring β-strands are hydrogen-bonded. In cyanovirin-N and other CVNH proteins, the exchange creates a structure in which two hairpins are positioned on top of the two three-stranded β-sheets, thereby creating a bilobal structure.

As described above, two amide resonances of OAA (Gly-26 and Gly-93) exhibit unusual, upfield-shifted proton resonances in the two-dimensional 1H-15N HSQC spectrum (Fig. 4). Inspection of the crystal structure reveals that the corresponding amide hydrogens are located right above the indole rings of Trp-90 and Trp-23 (Fig. 4, inset). Thus, π-hydrogen bonds can be formed between the amide protons of these two Gly residues and the π-electron cloud of Trp, a feature discussed previously (38, 39), but not often observed.

NMR spectroscopy allowed us to determine which of the OAA residues were involved in Man-9 binding. Delineation of two binding sites on the protein for carbohydrates revealed their location in two crevices on the surface of the protein, positioned symmetrically at two ends. In addition, we also determined which disaccharide linkage is recognized by OAA: the protein preferentially binds to Man(1–6)Man compared with all other disaccharide linkages. A schematic model of the OAA/Man-9 recognition for one of the two binding sites is depicted in Fig. 6 (B and C), differing in the two possible orientations of the two GlcNAc moieties. In either case, one or both of the Man(1–6)Man linkages can be accommodated in the binding site cleft formed by the loops connecting strands β1-β2, β7-β8, and β9-β10 in site 1 or by the loops connecting strands β6-β7, β2-β3, and β4-β5 in site 2, respectively.

Because OAA recognizes the Mano(1–3)Mano(1–6)Manβ(1–4)GlcNAcβ(1–4)GlcNAc pentasaccharide, one may speculate that specificity determinants similar to those seen for MVL and Mano(1–6)Mano(1–3)Manβ(1-
Structure and Man-9 Specificity and High Affinity Binding of OAA

4)GlcNAcβ(1–4)GlcNAc pentasaccharide recognition may exist (16). However, inspection of the letter complex structure revealed quite different sugar determinants. Key components in carbohydrate recognition by MVL are the two GlcNAc moieties that sit in a deep pocket of MVL, supplemented by additional interactions between the α(1–6)-linked mannoses. This renders the Manα(1–6)Manβ(1–4)GlcNAcβ(1–4)GlcNAc tetrasaccharide as the central element for recognition (16). In the present OAA structure, on the other hand, the binding cleft is too short to accommodate a tetrasaccharide. This was borne out by our titration data that revealed no interaction between the protein and the GlcNAcβ(1–4)GlcNAc disaccharide. Therefore, OAA recognizes mainly either of the two Manα(1–6)Man disaccharide units, imbedded within the pentasaccharide glycan, resulting in a distinctly different mode of OAA binding to Man-9, compared with MVL.

Interestingly, the Man-9-binding sites in OAA are close to the positions of the bound CAPS molecules in the free protein structure. A similar observation was made for cyanovirin-N, where, in the crystal structure with high mannos e oligosaccharides, a well defined CHEs molecule was bound in one of the sugar-binding sites (13). Both ionic and hydrophobic interactions between CAPS and OAA are observed, with the cyclohexyl ring and the hydrophobic chain of CAPS packing against the aromatic rings of the Trp-10 or Trp-77 side chains. The sulfate moiety forms a hydrogen bond with the backbone amide proton as well as a water-mediated one with the backbone carbonyl oxygens of Gly-124 or Gly-57. Because CHEs and CAPS are structurally related molecules, similar types of interactions are observed in both cases, although binding of CHEs to cyanovirin-N (13) or CAPS to OAA is very weak. Therefore, the presence of these organic molecules in the structures is simply a result of their inclusion in the crystallization buffers. Nevertheless, in certain cases, they may indicate the positions of the actual sugar ligands. Having CAPS molecules bound close to the Man-9-binding sites on OAA may also explain our unsuccessful efforts in obtaining a structure of the OAA-Man-9 complex through either soaking or co-crystallization.

In summary, we have elucidated the tertiary structure of OAA and investigated its carbohydrate specificity. Our results provide insights into the molecular basis and a mechanistic understanding of the protein’s potent anti-HIV activity. In turn, our data may be useful for the development of OAA as a novel reagent in the quest to combat HIV transmission.

Acknowledgments—We thank Dr. Palaniappa Arjunan for help with crystal manipulation and the Advance Photon Source trip preparation, Dr. Andy Howard for Advance Photon Source data collection and processing, and Mike Delk for NMR technical support.

REFERENCES
Structure and Man-9 Specificity and High Affinity Binding of OAA