Crystal Structures of the HIV-1 Inhibitory Cyanobacterial Protein MVL Free and Bound to Man₃GlcNAc₂

STRUCTURAL BASIS FOR SPECIFICITY AND HIGH-AFFINITY BINDING TO THE CORE PENTASACCHARIDE FROM *N*-LINKED OLIGOMANNOSIDE*

David C. Williams, Jr.‡, Jae Young Lee§, Mengli Cai‡, Carole A. Bewley¶‖, and G. Marius Clore‡**

From the ‡Laboratory of Chemical Physics, the §Laboratory of Molecular Biology, and the ¶Laboratory of Bioorganic Chemistry, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

The cyanobacterial protein MVL inhibits HIV-1 envelope-mediated cell fusion at nanomolar concentrations by binding to high mannose *N*-linked carbohydrate on the surface of the envelope glycoprotein gp120. Although a number of other carbohydrate-binding proteins have been shown to inhibit HIV-1 envelope-mediated cell fusion, the specificity of MVL is unique in that its minimal target comprises the Manα(1→6)-Manβ(1→4)GlcNAcβ(1→4)GlcNAc tetrasaccharide core of oligomannosides. We have solved the crystal structures of MVL free and bound to the pentasaccharide Man₃GlcNAc₂ at 1.9- and 1.8-Å resolution, respectively. MVL is a homodimer stabilized by an extensive intermolecular interface between monomers. Each monomer contains two structurally homologous domains with high sequence similarity connected by a short five-amino acid residue linker. Intriguingly, a water-filled channel is observed between the two monomers. Residual dipolar coupling measurements indicate that the structure of the MVL dimer in solution is identical to that in the crystal. Man₃GlcNAc₂ binds to a preformed cleft at the distal end of each domain such that a total of four independent carbohydrate molecules associate with each homodimer. The binding cleft provides shape complementarity, including the presence of a deep hydrophobic hole that accommodates the *N*-acetyl methyl at the reducing end of the carbohydrate, and specificity arises from 7–8 intermolecular hydrogen bonds. The structures of MVL and the MVL-Man₃GlcNAc₂ complex further our understanding of the molecular basis of high affinity and specificity in protein-carbohydrate recognition.

HIV[1] infection occurs via virus-cell and cell-cell fusion mediated by the two envelope glycoproteins, gp120 and gp41 (1–3). Gp120 binds to CD4 and chemokine receptors, triggering a series of events that leads to the insertion of the fusion peptide of gp41 into the target membrane and subsequent membrane fusion. The surface of gp120 is characterized by an abundance of high mannose *N*-linked glycosylation sites (4). Post-translational glycosylation of gp120 covers conserved portions of the protein with host-generated carbohydrates that constitute almost half of the molecular mass of gp120 and limit recognition by the immune system, thereby allowing HIV to evade detection (4, 5). The neutralizing antibody 2G12 (6) and the anti-HIV protein cyanovirin (7–10) specifically target high mannose sugars on the surface of gp120. Cyanovirin is a cyanobacterial protein isolated from *Nostoc ellipsosporum* (11, 12) that inhibits HIV-1 fusion at nanomolar concentrations (7, 11) and is currently undergoing preclinical development as a microbicide, having shown 100% efficacy in a primate model (13). Very recently, another cyanobacterial protein known as MVL, originally isolated from a laboratory culture of *Microcystis viridis* NIES-102 (14), has also been shown to inhibit HIV-1 fusion at nanomolar concentrations (15). Cyanovirin and MVL are unusual in that they bind oligosaccharides with very high affinity in the absence of multivalent interactions. Cyanovirin specifically recognizes the disaccharide Manα(1→2)Manα located at the terminal branches of mammalian high mannose oligosaccharides (7–10). In contrast MVL is distinct among oligomannose-binding proteins in that it exhibits fine specificity for the Manα(1→6)Manβ(1→4)GlcNAcβ(1→4)GlcNAc tetrasaccharide core found in *N*-linked oligomannosides (15).

MVL is a monodisperse homodimer in solution with four independent carbohydrate binding sites (15). Each subunit comprises two homologous domains, each consisting of 54 residues with 50% sequence identity between them, connected by a five-residue linker. MVL exhibits no significant sequence similarity to any other known protein family. To elucidate the molecular basis of carbohydrate recognition by MVL we have solved the crystal structures of MVL both free and bound to the pentasaccharide Man₃GlcNAc₂.

EXPERIMENTAL PROCEDURES

Protein Purification and Crystallization—The expression vector for MVL (15) was transformed into strain BL21(DE3) of *Escherichia coli* and grown at 37 °C in Luria-Bertani medium. The culture was induced at an $A_{600\ nm}$ of ~1 with 1.0 mM isopropyl-β-D-thiogalactoside for 2–3 h and harvested by centrifugation at 7800 × *g* for 15 min. The pellet from 1 liter of culture was resuspended in 100 ml of 20 mM NaPO₄ (pH 7.0),

‖ To whom correspondence may be addressed: Laboratory of Bioorganic Chemistry, Building 8, Room 1A-02, NIDDK, National Institutes of Health, Bethesda, MD 20892-0820. Tel.: 301-594-5187; E-mail: caroleb@intra.niddk.nih.gov.

** To whom correspondence may be addressed: Laboratory of Chemical Physics, Building 5, Rm. B1-30I, NIDDK, National Institutes of Health, Bethesda, MD 20892-0520. Tel.: 301-496-0782; Fax: 301-496-0825; E-mail: mariusc@intra.niddk.nih.gov.

---

[1] The abbreviations used are: HIV, human immunodeficiency virus; r.m.s., root mean square; SeMet, selenomethionine.

TABLE I

Data collection and refinement statistics

	Free		Complex
	Native	SeMet	
Data collection			
Space group	P2$_1$	P2$_1$	P1
Wavelength (Å)	1.5418	1.5418	1.5418
Unit cell *a, b, c* (°)	$a = 76.2, b = 74.9, c = 83.9$		$a = 47.8, b = 73.7, c = 78.9$
α, β, γ (degrees)	$\alpha = \gamma = 90, \beta = 100.8$		$\alpha = 90.4, \beta = 98.4, \gamma = 108.9$
Resolution (Å)[a]	50–1.9 (1.97–1.90)	50–1.85 (1.92–1.85)	50–1.8 (1.86–1.80)
Unique reflections	71,428	70,123	85,237
Completeness (%)[a]	97.4 (95.3)	88.7 (46.8)	91.4 (68.2)
R_{merge} (%)[a,b]	9.2 (34.7)	6.9 (45.0)	5.3 (22.3)
$I/\sigma(I)$	14.5 (2.25)	23.6 (2.37)	20.4 (3.30)
Refinement			
Resolution (Å)	20–1.9		20–1.8
Reflections work set	61,777		78,890
Reflections test set	6949		4222
Atoms refined	7867 (865 waters)		8669 (643 waters)
R (R_{free}) (%)[c]	17.7 (21.6)		19.2 (22.1)
R.m.s. deviation in bonds (Å)	0.008		0.005
R.m.s. deviation in angles (degrees)	1.54		1.24
Mean *B*-factor values (Å)	20.2		19.9
% residues in most favorable region of Ramachandran map[d]	92.1		89.1

[a] Data in the highest resolution shell are given in parentheses.

[b] $R_{\text{merge}} = \Sigma_h \Sigma_I |I_{hi} - \langle I_h \rangle| / \Sigma \langle I_h \rangle$, where I_{hi} is the intensity of the *i*th observation of reflection *h,* and $\langle I_h \rangle$ is the average intensity of redundant measurements of the *h* reflections.

[c] *R*-value = $\Sigma |F_o - F_c| / \Sigma |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is calculated from the reflections in the test set.

[d] Calculated with PROCHECK (Ref. 36).

50 mM NaCl, and 2 mM EDTA, lysed by three passages through a microfluidizer, and clarified by centrifugation at $10,000 \times g$ for 15 min. The supernatant was subjected to fractionation by ammonium sulfate precipitation with the serial addition of 10 and 15 g of NH_4SO_4/100 ml of supernatant, incubation on ice for at least 4 h, and centrifugation at $10,000 \times g$ for 15 min. The final pellet (25% w/v NH_4SO_4) was resuspended in 8 ml of 20 mM $NaHPO_4$ (pH 7.0), 200 mM NaCl, and 2 mM EDTA and loaded onto a Superdex 75 26/60 gel filtration column (Amersham Biosciences) equilibrated with the same buffer. The fractions containing MVL were identified by SDS-PAGE, pooled, diluted 4-fold with 20 mM Tris (pH 7.5), loaded onto a HiLoad™ 10/10 MonoQ anion exchange column (Amersham Biosciences) equilibrated with the same buffer, and eluted with a 0–1 M NaCl gradient. The largest peak containing MVL was pooled and dialyzed against 10 mM Tris (pH 7.5).

Selenomethionine (SeMet)-substituted protein was expressed by transforming the same vector into the B834(DE3) methionine auxotrophic strain and growing the bacteria in M9 minimal medium supplemented with 50 mg/liter SeMet (Sigma-Aldrich). The culture was grown at 37 °C up to an $A_{600 \text{ nm}} \sim 0.4$ and induced with 0.3 mM isopropyl-β-D-thiogalactoside at 30 °C for 24 h. The purification was identical to that described above for wild-type MVL.

Crystals were grown in 0.4 M $NH_4H_2PO_4$ at 20 °C. For flash-freezing, 40% ethylene glycol was added to the mother liquor. Co-crystallization of MVL and Man$_3$GlcNAc$_2$ (obtained from Glycotec, Rockville, MD) at 1:2.2–3.0 molar ratios yielded suitable crystals under identical conditions.

Structure Determination and Refinement—MVL and SeMet-MVL crystals of free MVL diffracted to better than 1.9 Å, whereas the MVL-Man$_3$GlcNAc$_2$ co-crystals diffracted to better than 1.8 Å. Data were collected with an in-house x-ray source using a RAXIS IV image plate detector and processed using the HKL2000 software package (16). The structure of free MVL was solved by the single isomorphous replacement and anomalous scattering method (17). 16 of 16 selenium atoms were found, and over 90% of the free MVL molecule was traced using the SOLVE/RESOLVE software (18). The initial model was refined iteratively using the programs O (19) and CNS (crystallography NMR software) (20). The final model of free MVL contains eight MVL monomers (residues 1–113), six (PO$_4$)$^{2-}$, 13 ethylene glycol, and 886 water molecules. The initial model of the co-crystal was generated by molecular replacement with free MVL using the program AMORE (21, 22). The carbohydrate molecules were traced and the model iteratively refined using O (19) and CNS (20). The final co-crystal model contains eight MVL monomers (residues 1–113), 16 Man$_3$GlcNAc$_2$, 2 (PO$_4$)$^{2-}$, 26 ethylene glycol, and 643 water molecules. The data collection and refinement statistics are given in Table I. Structure figures were generated with the programs VMD-XPLOR (23), RIBBONS (24), and GRASP (25).

Residual Dipolar Couplings—Backbone NH residual dipolar couplings (26), $^1D_{NH}$, were measured in solution on a sample of 0.4 mM MVL (dimer) saturated with Man$_3$GlcNAc$_2$ (1.6 mM) in 10 mM sodium phosphate, pH 6.5, by taking the difference in $^1J_{NH}$ couplings between liquid crystalline (5% polyethylene glycol/hexanol with a surfactant to alcohol ratio of 0.96 (27)) and isotropic (water) media. $^1J_{NH}$ couplings were measured using two-dimensional in-phase/antiphase {^{15}N,^{1}H} heteronuclear single quantum coherence experiments at a ^{1}H frequency of 600 MHz on a Bruker DMX600 NMR spectrometer. Best-fitting of the measured $^1D_{NH}$ couplings to the crystal structures of free and complexed MVL was carried out by singular value decomposition using the program SSIA (28). The agreement between observed and calculated $^1D_{NH}$ couplings is expressed as a dipolar coupling *R*-factor, which scales between 0 and 100%, and is defined as the ratio of the r.m.s. deviation between observed and calculated values to the expected r.m.s. deviation if all of the N-H vectors were randomly distributed (29). The latter is given by $[2D_a^2(4 + 3\eta^2)/5]^{1/2}$, where D_a is the magnitude of the axial component of the alignment tensor and η the rhombicity (29). The values of D_a^{NH} and η obtained by singular value decomposition are 7.4 Hz and 0.63, respectively, when best-fitting the dipolar couplings to the coordinates of the free monomer; 7.5 Hz and 0.61, respectively, for the free dimer; 7.2 Hz and 0.63, respectively, for the complexed monomer; and 7.5 Hz and 0.62, respectively, for the complexed dimer.

RESULTS AND DISCUSSION

Crystal Structure of MVL—Native MVL crystallized in space group P2$_1$ and diffracted to 1.9 Å. A SeMet derivative of MVL crystallized under identical conditions in the same space group and diffracted to 1.85 Å. The initial phases were determined by single isomorphous replacement and anomalous scattering, and the structure was refined with data to a final resolution of 1.9 Å. The final structure was well defined with an *R*-factor of 17.7% and a free *R* of 21.6% (Table I). A total of four homodimers of MVL were present in the asymmetric unit, leading to eight unique copies of the MVL monomer in the final structure.

MVL is a 13-kDa polypeptide of 113 amino acids that contains two highly homologous 54-residue domains with 50% sequence identity between them (15). In the following descriptions the secondary structure elements and residue numbers

FIG. 1. **Overview of the crystal structures of MVL free and bound to Man₃GlcNAc₂.** *A, ribbon diagram* of a single monomer of MVL from the crystal structure of the free protein. In the N-domain, strands β1, β2, and β3 and helix α1 comprise residues 2–10, 32–41, 45–53, and 14–29, respectively; in the C-domain, strands β4, β5, and β6 and helix α2 comprise residues 61–69, 91–100, 104–112, and 73–88, respectively. *B, ribbon diagram* of the free MVL dimer with one subunit in *blue* and the other in *red*, together with water molecules (*green spheres*) and associated side chains that participate in water-bridged hydrogen bonds between the two monomers. *C, ribbon diagram* of the MVL-Man₃GlcNAc₂ complex with the bound carbohydrate depicted as *cylinders* and *spheres* (carbon, *yellow*; nitrogen, *blue*; and oxygen, *red*). *D*, combined surface, tube, stick, and space-filling representation of the MVL dimer illustrating the water-filled channel formed between monomers in the dimer. A *surface diagram* of one monomer (*red*) and *tube diagram* of the second monomer (*blue*) are shown with space-filling rendering of water (*gray*) and ethylene glycol (*green*) molecules that fill the channel between the monomers. A *surface diagram* of carbohydrate bound to the surface model of one monomer of MVL and a *stick diagram* of carbohydrate bound to the tube model of the second monomer are shown in *yellow*. *E*, mesh depiction of the electron density plotted at 1.3σ for a *cylinder* and *sphere* representation of Man₃GlcNAc₂ bound to the N-domain of MVL together with the contacting residues (carbon, *yellow* for carbohydrate and *white* for protein; nitrogen, *blue*; and oxygen, *red*).
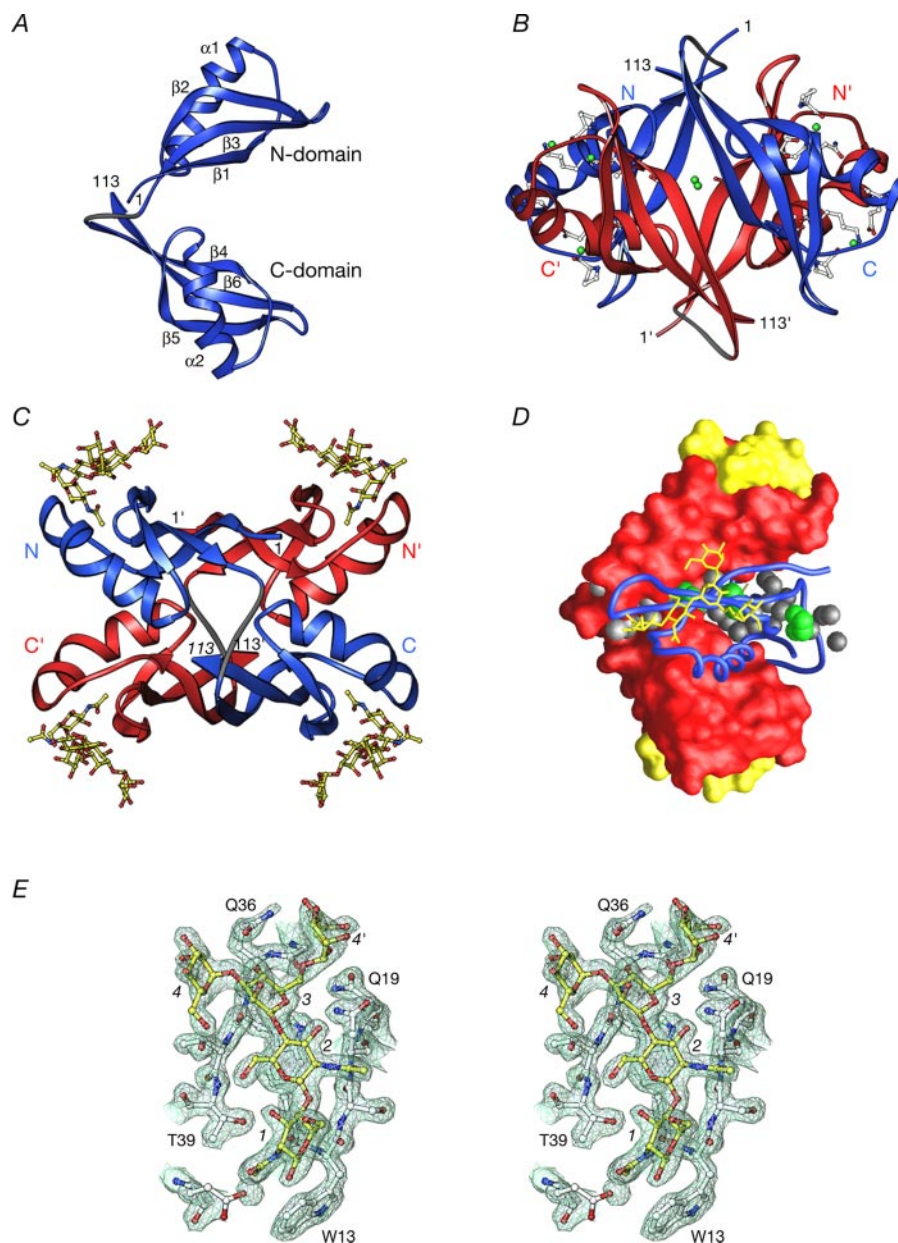
are given for the N-terminal domain (residues 1–54), whereas the equivalent elements and residues in the C-terminal domain (residues 60–113) are provided in parentheses. Each domain contains a three-stranded anti-parallel β-sheet in a −2*X*,1 topology with strand β3 (β6) hydrogen-bonded to strands β1 (β4) and β2 (β5) (Fig. 1*A*). The single α-helix α1 (α2) is located between strands β1 (β4) and β2 (β5) and is packed against one face of the sheet (Fig. 1*A*). Three bulges in the β-strands, between residues 34–35 (93–94), 40–41 (99–100), and 44–45 (103–104), distort the normal twist of the β-sheet such that it wraps around the α-helix and forms a cleft at the distal end of each domain that constitutes the carbohydrate binding pocket (Fig. 1*C*). The Cα backbone r.m.s. difference between the N- and C-domains is ~0.8 Å. The N- and C-domains form a boomerang-shaped structure connected by residues 55–59. The long axes of the two domains are oriented approximately orthogonal to each other; and the C-domain can be superimposed on the N-domain by an ~180° rotation about an axis that bisects the angle between the two domains (Fig. 1, *A* and *D*). There are very limited contacts between the N- and C-domains of the monomer; the interface buries only 156 Å² of surface-

accessible area, and there is a single interdomain hydrogen bond between the side-chain amide group of Gln-30 and the backbone carbonyl oxygen of Thr-112.

MVL forms a homodimer such that the boomerang-shaped monomers interlock, and each domain from one monomer contacts both domains from the second monomer (Fig. 1, *B* and *C*). The α1 and α2 helices of the first monomer contact and lie approximately orthogonal to the α2 and α1 helices, respectively, of the second monomer, whereas the first strand of each β-sheet (β1 or β4) contacts the same strand in the second monomer. The two monomers interact extensively resulting in 3930 Å² of buried accessible surface area with 70% of the buried surface involving nonpolar atoms. The gap volume index (defined as the ratio of gap volume to interface buried accessible surface area) for the interacting surfaces is 1.28, indicative of a tight, complementary interface (30). Three intermolecular hydrogen bonds stabilize the dimer including those between the N-ε2 atom of His-29 of one monomer and the hydroxyl group of Ser-87′ of the other monomer, the side-chain hydroxyls of Thr-64 and Thr-64′, and the hydroxyl group of Thr-64 and the backbone carbonyl oxygen of Asp-65′. At least

eight (4×2) water molecules form hydrogen bonds that bridge the interface between the two monomers (Fig. 1*B*); a single water molecule bridges the backbone carbonyl oxygens of Ala-9 and Pro-11 and the side chain of Gln-83′. A similar network is formed between the backbone carbonyl oxygens of Pro-70 and Lys-24′ and the side chain of Glu-76, between the backbone carbonyl and amide groups of Asn-6 and the same atoms of Asn-6′, and between the backbone carbonyl oxygens of Lys-24 and Ala-68′.

Residual dipolar couplings provide long range orientational information in solution, thereby providing a simple and accurate means of ascertaining whether the structure in the crystal reflects that in solution (26). The relative orientations of the N- and C-domains within the monomer and of the two monomers within the dimer are the same in solution and in the crystal state, as evidenced by the excellent agreement between the measured backbone N-H residual dipolar couplings recorded on the MVL-Man$_3$GlcNAc$_2$ complex in a diluted liquid crystalline medium (of polyethylene glycol/hexanol) and those calculated from the crystal coordinates of both free and carbohydrate-bound MVL. The values obtained for the dipolar coupling *R*-factors (29) are 16.6% for the monomer and 17.3% for the dimer using the coordinates of free MVL, and 16.9% for the monomer and 17.9% for the dimer using the coordinates of MVL complexed to Man$_3$GlcNAc$_2$. These values for the dipolar coupling *R*-factors are within the range expected for 1.5–2-Å resolution crystal structures (26, 31), indicating that the structures in solution and in the crystal are the same.

Intriguingly, despite the tight intermonomer interaction, water molecules fill a channel that extends from the surface adjacent to the linker region of one monomer to the surface adjacent to the linker region of the second monomer (Fig. 1*D*). This channel also contains ethylene glycol molecules from the cryopreservative solution such that the N-δ1 atom of His-29 (from each monomer) is hydrogen-bonded to an ethylene glycol molecule near the center of the channel. The presence of a reactive group such as histidine near the center of a solvent-accessible channel may suggest a potential catalytic or ion-pumping function; however, no such activity has been identified for MVL to date (14, 15).

A search of the Protein Data Bank using the program DALI (32) reveals that the topology of the individual domains of MVL (a triple-stranded β-sheet on top of which lies an α-helix) represents a fairly common motif with several proteins having stretches of around 50 residues (with no significant sequence identity) that can be superimposed onto the individual domains of MVL with Cα atomic r.m.s. differences of 2.5–3.5 Å. However, the combination of the sequential N- and C-domains found in the monomer of MVL and their configuration in the MVL dimer is not observed in any other known protein structure. Moreover, this structural motif has not been reported to be involved in carbohydrate recognition.

Crystal Structure of MVL-Man$_3$GlcNAc$_2$—MVL saturated with Man$_3$GlcNAc$_2$ (1:2.2–3.0 molar ratio) co-crystallized under the same conditions as free MVL. The complex formed crystals in space group P1 that diffracted to 1.8 Å. The initial model was generated by molecular replacement with the structure of free MVL and refined using data to a final resolution of 1.8 Å. Similar to the free MVL crystals, the asymmetric unit contains four MVL dimers. Clear electron density (Fig. 1*E*) demonstrates the presence of four Man$_3$GlcNAc$_2$ ligands bound to each MVL dimer (for a total of 16 ligands in the asymmetric unit). The final model was well defined by the data with an *R*-factor of 19.2% and a free *R* of 22.1% (Table I).

The presence of eight unique MVL monomers per asymmetric unit that were refined independently (without non-crystallographic symmetry restraints) provides an opportunity to as-

sess the conformational variability of MVL within the crystal environment (Fig. 2). Interestingly, the variation between monomers is smaller for MVL in the complex than for free MVL; the average r.m.s. deviation to the mean coordinate positions for free MVL is 0.19 Å for backbone atoms and 0.41 Å for all heavy atoms, and the corresponding values for the MVL-Man$_3$GlcNAc$_2$ complex are 0.12 and 0.31 Å, respectively (and 0.32 Å for the carbohydrate heavy atoms). This observation suggests that binding of Man$_3$GlcNAc$_2$ restricts the conformational variability of the MVL protein even though the free structure is already well ordered. In terms of side chains, this phenomenon is particularly evident for Trp-72 in the C-domain and to a lesser extent for the equivalent residue Trp-13 in the N-domain, which both exhibit a sizeable spread among the eight molecules in free MVL but are fixed in a narrow region of conformational space in the complex (Fig. 2). With one exception, MVL undergoes no significant conformational change upon binding carbohydrate, as evidenced by the observation that the r.m.s. difference between the average coordinates of free and complexed MVL (0.09 Å for backbone atoms and 0.18 Å for all atoms) is smaller than the values for the average r.m.s. deviations to the mean coordinates among the dimers from either structure given above. One residue located in the carbohydrate binding pocket, Thr-38, however, does change its rotameric state from g^- in the free state (for seven of the eight molecules) to t in the complex (Fig. 2*A*), thereby permitting the hydroxyl group of Thr-38 to participate in an intermolecular hydrogen bond with the bound carbohydrate (see below).

Structural Basis for Carbohydrate Binding and Specificity—The bend in the β-sheet forms a surface cleft between the β-sheet and the α-helix at the distal end of each domain (Fig. 3, *A* and *B*). This cleft constitutes the carbohydrate binding pocket that binds with low to submicromolar affinity high mannose oligosaccharides that contain at a minimum the Manα(1→6)Manβ(1→4)GlcNAcβ(1→4)GlcNAc tetrasaccharide core structure (15). The distances between the carbohydrate binding pockets are as follows: ~45 Å between the N- and C-domains of the monomer, ~36 Å between the N- and N′ (and C- and C′)-domains, and ~25 Å between the N- and C′-domains. As depicted in Fig. 3, *A* and *B*, the deepest portion of the cleft is closed on one end by the β-turn between strands β2 and β3 (β5 and β6) and the aromatic ring of Trp-13 (Trp-72). The base of the cleft is formed by Trp-37 (Trp-86), and the sides are lined by residues from the N terminus of helix α1 (α2) and strand β2 (β5). This cleft provides shape complementarity and a hydrogen-bonding network that provides the molecular basis for selectivity and specificity.

The reducing GlcNAc residue of Man$_3$GlcNAc$_2$ fits into the deepest portion of the binding cleft, and the carbohydrate chain extends along and molds to the surface of MVL, such that the pentasaccharide buries ~360 Å² of accessible protein surface area (Fig. 3*B*). Man$_3$GlcNAc$_2$ is involved in extensive hydrogen-bonding and/or van der Waals interactions with 13 residues of MVL, including residues 11–15, 19, 35–39, and 42–43 in the N-domain and residues 70–74, 78, 94–98, and 101–102 in the C-domain. 11 of the 13 interacting residues are identical in the N- and C-domains. Thr-38 and Ser-43 in the N-domain are substituted by Arg-97 and Gly-102 in the C-domain. With only one exception (Thr-38 *versus* Arg-97), the side chains of all residues involved in protein-carbohydrate hydrogen bonds are preserved (Fig. 3, *B* and *C*). The backbone and heavy atom r.m.s. differences between the mean coordinates of the N- and C-domain carbohydrate binding pockets are only 0.10 and 0.18 Å, respectively (Fig. 3*D*). A diagrammatic representation of the hydrogen-bonding and van der Waals interactions between MVL and Man$_3$GlcNAc$_2$ is shown in Fig. 4.
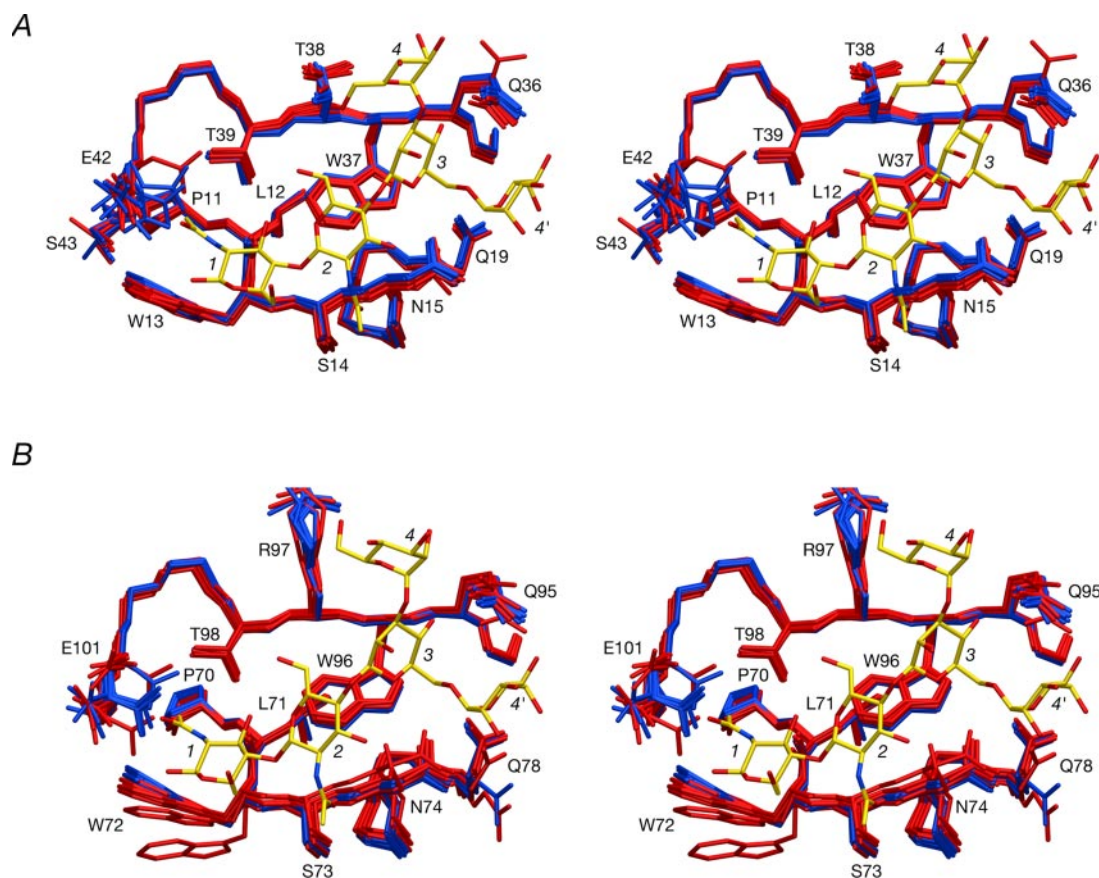
FIG. 2. **Comparison of the carbohydrate binding pockets in MVL free and complexed to Man₃GlcNAc₂.** The backbone of the binding pocket and side chains involved in protein-carbohydrate hydrogen-bonding interactions is displayed, and each *panel* shows a best-fit superposition of the eight MVL monomers in the asymmetric unit in the free and bound states. *A*, N-domain; *B*, C-domain. Free MVL, *red*; complexed MVL, *blue*; Man₃GlcNAc₂ (one of the eight molecules bound to each domain), *yellow*, with oxygen atoms in *red* and nitrogen atoms in *blue*. Note that the O-6 hydroxyl group of the Man4 unit is hydrogen-bonded to the hydroxyl group of Thr-38 in the N-domain pocket (*A*); in the C-domain pocket (*B*) the equivalent residue is Arg-97, and hence no such hydrogen bond can be formed. This difference leads to a 180° rotation about the O-5–C-5–C-6–O-6 dihedral angle of Man4 such that the C-6–O-6 bond points in the opposite direction in the N- and C-domain pockets.

The glycosidic bond torsion angles of Man₃GlcNAc₂ bound to the N- and C-domains of MVL are all located within the preferred conformations observed in crystal structures of isolated oligosaccharides, glycoproteins, and glycan-binding proteins (33). The average values for the glycosidic ϕ (O5-C1-O-C(x)′) and ψ (C1-O-C(x)′-C(x-1)′) torsion angles (taken over all 16 carbohydrate molecules in the asymmetric unit) are $-71.7 \pm 1.4°$ and $101.4 \pm 1.4°$, respectively, for the GlcNAcβ(1→4)GlcNAc linkage; $-76.4 \pm 1.3°$ and $135.6 \pm 1.6°$, respectively, for the Manβ(1→4)GlcNAc linkage; $71.1 \pm 6.1°$ and $-115.0 \pm 7.5°$, respectively, for the Manα(1→3)Man linkage; and $61.5 \pm 2.3°$ and $161.7 \pm 4.3°$, respectively, for the Manα(1→6)Man linkage. The average value of the ω (O-C(x)′-C(x-1)′-C(x-2)′) torsion angle for the Manα(1→6)Man linkage is $51.2 \pm 2.6°$.

A key component of carbohydrate specificity can be attributed to interactions with the reducing GlcNAc1 unit. The acetyl methyl group from GlcNAc1 fits into a deep hole on the surface such that it wedges between and is in van der Waals contact with the side chains of Pro-11 (Pro-70) and Trp-13 (Trp-72) and the methyl groups of Leu-12 (Leu-71) and Thr-39 (Thr-98) (Fig. 3B), whereas the acetyl oxygen atom is hydrogen-bonded to the backbone amide of Ser-43 (Gly-102) (Fig. 3C). The acetyl NH group of GlcNAc1 is hydrogen-bonded to the backbone oxygen atom of Leu-12 (Leu-71), and the O-3 atom of the pyranose ring is hydrogen-bonded to the side-chain hydroxyl group of Thr-39 (Thr-98). The reducing hydroxyl group (O-1) from GlcNAc1 protrudes from the binding site and remains solvent-accessible such that an *N*-linked Asn would not disrupt binding. The

acetyl oxygen and O-6 atoms of the subsequent GlcNAc2 unit are hydrogen-bonded to the backbone amide groups of Asn-15 (Asn-74) and Thr-39 (Thr-98), respectively. In addition the pyranose ring of GlcNAc2 is stacked on top of the benzyl ring of Trp-37 (Trp-96), which lines the base of the binding site. Formation of these hydrogen bonds together with structural complementarity for the *N*-acetyl groups explains the selectivity for β(1→4)-linked GlcNAc present in the Man₃GlcNAc₂ pentasaccharide.

The binding site also selects for a β(1→4)-linked mannose in the third position and additional α(1→3)- and α(1→6)-linked mannose units (Man4 and Man4′, respectively) in the branching positions. The pyranose ring of Man3 is stacked on top of the pyrrole ring of Trp-37 (Trp-96), the side-chain amide group of Gln-36 (Gln-95) forms a hydrogen bond with the O-4 atom of Man3, and the alkyl side chain of Gln-36 (Gln-95) protrudes into the space between the branching mannopyranose units, Man4 and Man4′ (Fig. 3B), thereby discriminating against an α(1→4) linkage at this position. The interatomic distance between the O-ϵ1 atom of Gln-36 and the O-4 atom of Man3 is particularly short (<2.6 Å) suggesting that this is a well formed hydrogen bond that stabilizes the complex with the Manβ(1→4)GlcNAc2 reducing end trisaccharide core. The branched mannose residues, Man4 and Man4′, extend up and away from the binding cleft yet still establish hydrogen bonds with MVL. The α(1→3)-linked Man4 hydrogen-bonds to the side chain of Thr-38 via the O-6 hydroxyl group; however, this is unique to the N-domain because the equivalent residue in the C-domain is Arg-97, which does not form a hydrogen bond
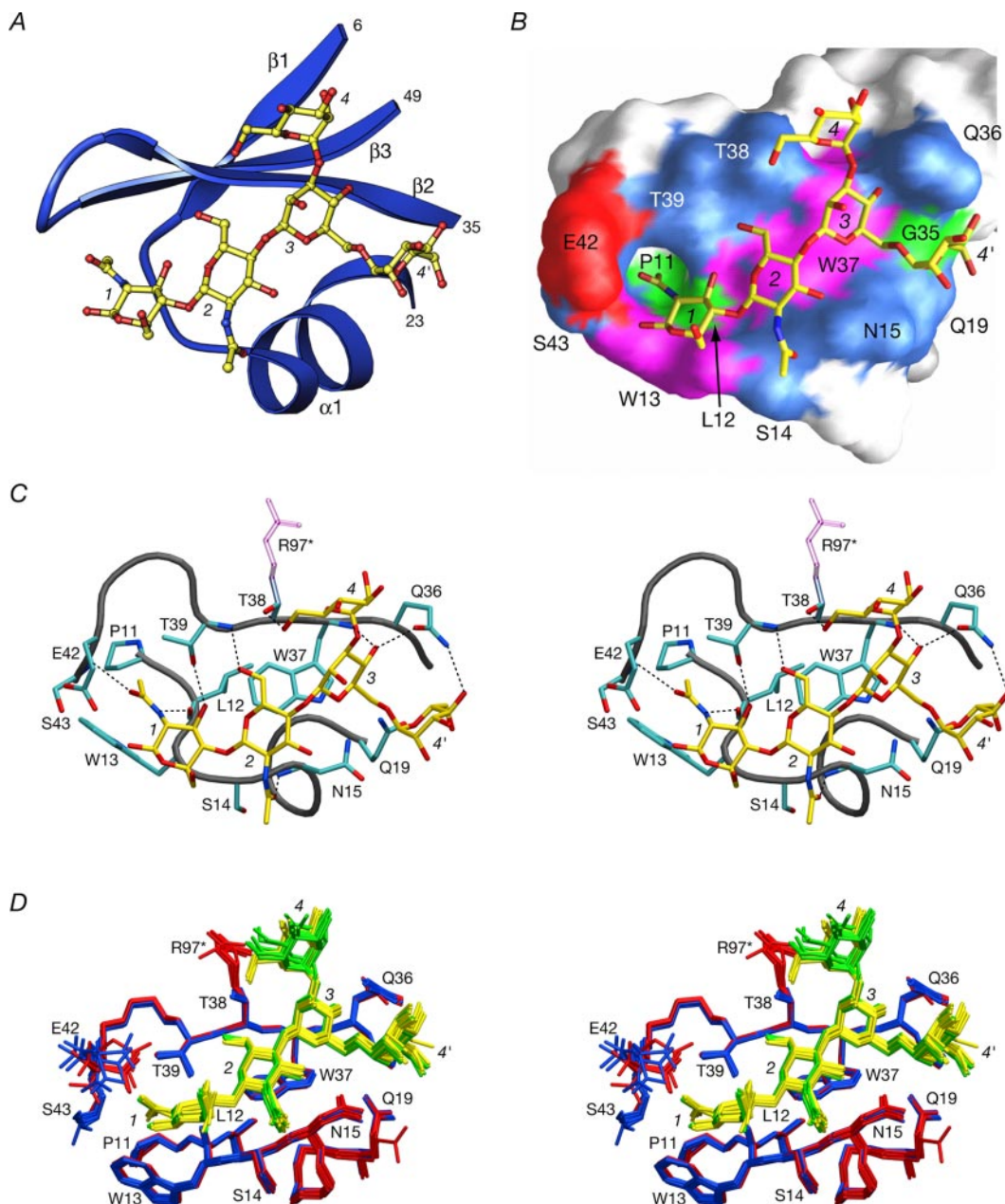
FIG. 3. **Interaction between MVL and Man$_3$GlcNAc$_2$.** *A*, *ribbon diagram* of the carbohydrate binding site from the N-domain of MVL with bound Man$_3$GlcNAc$_2$. Bound carbohydrate is shown in *spheres* and *cylinders* (carbon, *yellow*; nitrogen, *blue*; oxygen, *red*). *B*, surface representation of the carbohydrate binding site from the N-domain with bound Man$_3$GlcNAc$_2$. The carbohydrate is displayed as in *A*, and interacting residues on the protein are colored as follows: hydrophobic, *green*; tryptophan, *magenta*; polar, *cyan*; negatively charged, *red*. *C*, stereoview of the binding site from the N-domain of MVL with bound Man$_3$GlcNAc$_2$ depicting intermolecular hydrogen bonds (*dashed lines*). Side-chain and select main-chain atoms for the interacting residues of MVL and carbohydrate heavy atoms are represented by a *stick model* (carbon, *yellow* for carbohydrate and *cyan* for protein; oxygen, *red*; nitrogen, *blue*), and the backbone of MVL is depicted by a *tube diagram* (*blue*, residues 11–19 and 35–43). Protein residues are labeled with the single-letter code, and the sugar rings of the carbohydrate are numbered in *italics*. The position of Arg-97 (*R97*, denoted with an *asterisk*) from the C-domain binding site is depicted in *transparent lilac* to demonstrate the only contacting side chain that differs between the two domains. *D*, stereoview of a superposition of the 16 carbohydrate binding sites in the asymmetric unit of the crystal structure of the MVL-Man$_3$GlcNAc$_2$ complex with the N-domain shown in *blue*, the C-domain in *red*, and carbohydrate bound to the N- and C-domains in *yellow* and *green*, respectively. The residue numbering is that of the N-domain, and the residue numbers for the C-domain can be obtained by adding 59 to the labeled residues; Arg-97 from the C-domain is denoted by an *asterisk*.

with Man4 (Fig. 3, *C* and *D*). This represents the only difference between the two carbohydrate binding sites and explains why the average heavy atom r.m.s. deviation to the mean coordinates is higher for the carbohydrate bound to the C-domain than for that bound to the N-domain (0.35 *versus* 0.27 Å; Fig. 3*D*). It also accounts for the observation that the only significant differences in glycosidic bond torsion angles for the carbohydrate bound to the N- and C-domains are observed at the Manα(1→3)Man linkage, where ϕ has a somewhat smaller value (67.4 ± 2.4°) in the N-domain than in the C-domain

(74.9 ± 6.3°) and where ψ has a narrower distribution in the N-domain (−114.8 ± 2.7°) than in the C-domain (−115.2 ± 10.3°). Finally, α(1→6)-linked Man4′ forms a hydrogen bond between the O-4 atom and the side chain of Gln-36 (Gln-95) and a water-bridged hydrogen bond between the O-5 atom and the side-chain amide of Asn-15 (Asn-74).

Correlation to Binding Affinities—As previously noted (15), binding of oligomannosides to MVL is associated with large negative enthalpies that increase with the size of the carbohydrate. An increasingly negative entropy change opposes the
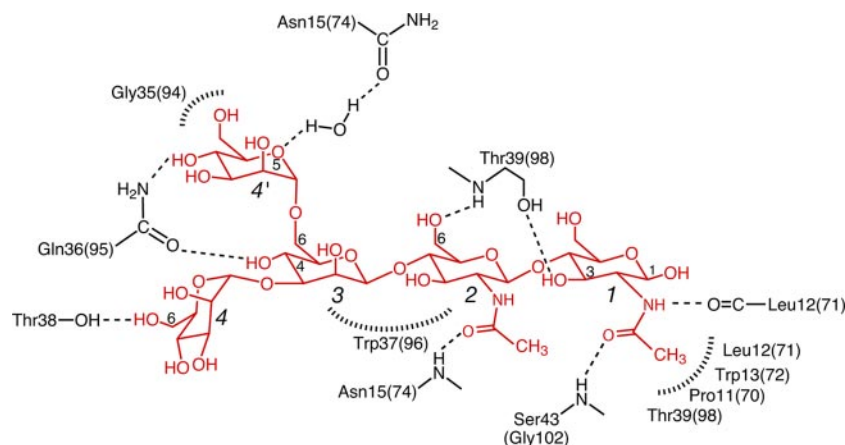
FIG. 4. **Schematic diagram of hydrogen-bonding interactions between MVL and Man₃GlcNAc₂.** The carbohydrate structure is shown in *red* and the contacting protein residues in *black*. *Dashed lines* indicate hydrogen bonds, and the *dashed arcs* represent hydrophobic contacts.

favorable enthalpy change, yet does not fully compensate, so that longer oligomannosides bind with increasing affinity (K_D of 0.2 μM for Man₆GlcNAc₂ *versus* 2.9 μM for Man₃GlcNAc₂). These observations are consistent with the extensive hydrogen-bonding network discussed above. Binding of the mannose chain creates enthalpically favorable hydrogen bonds but restricts conformational freedom of the carbohydrate residues. Modeling the appropriate mannose residues on the structure of the MVL-Man₃GlcNAc₂ complex to form Man₆GlcNAc₂ using glycosidic bond torsion angles in the preferred regions of conformational space (33) and avoiding steric clash suggests the presence of the following additional interactions involving ManB, $\alpha(1{\to}6)$-linked to Man4′, and ManC, $\alpha(1{\to}2)$-linked to Man4. ManB (with the ϕ, ψ, and ω glycosidic torsion angles in the g^+, t, and g^+ conformations, respectively) could form an additional hydrogen bond with the side chain of Gln-19 (Gln-78); ManC (with the ϕ and ψ glycosidic torsion angles in the g^+ and t conformations, respectively) could form hydrogen bonds with the side chains of Glu-49 (Gln-108) and possibly Asn-6 (Asp-65). These hydrogen bonds would account in part for the -7.7 kcal/mol additional enthalpy change associated with the binding of Man₆GlcNAc₂ compared with Man₃GlcNAc₂ (15). Further extension of the branches to Man₉GlcNAc₂ yields no additional protein-carbohydrate contacts accounting for the observation that the affinities of Man₆GlcNAc₂ and Man₉GlcNAc₂ are very similar (15). Hence, the structure reported here accounts for the affinity and selectivity of MVL for oligomannosides.

Inhibition of HIV-1—HIV infects cells through specific recognition of CD4 and chemokine (CCR5 or CXCR4) cell surface receptors by the viral glycoprotein gp120 (1–3). Gp120 is extensively glycosylated by high mannose *N*-linked carbohydrates, which subvert immune recognition by covering the more highly conserved regions of gp120 with oligosaccharides generated by the host (4, 5); hence current research efforts are being directed toward identifying molecules that can bind high mannose oligosaccharides and prevent cellular infection by HIV.

A number of carbohydrate-binding proteins known to bind high mannose oligosaccharides and therefore HIV-1 envelope have been identified and characterized. As recently reviewed by Botos and Wlodawer (34), these include the Ca^{2+}-dependent C-type lectins involved in innate immunity such as mannose-binding protein and the dendritic cell surface receptor DC-SIGN, a variety of plant lectins, as well as several metal-independent carbohydrate-binding proteins that do not fall into established lectin families. Most of these proteins bind their targets with relatively low affinity and typically show specific-

ity for either mannosides (such as mannotriose) or GlcNAc₂, structures that are present in *N*-linked oligosaccharides abundant on gp120 (35). MVL is unique in this regard in that it specifically binds to the Man$\alpha(1{\to}6)$Man$\beta(1{\to}4)$-GlcNAc$\beta(1{\to}4)$GlcNAc tetrasaccharide core structure with low micromolar affinity and requires the presence of both the mannose and glucosamine residues for carbohydrate recognition (15). The reducing end of this tetrasaccharide maintains solvent accessibility in complex with MVL such that an *N*-linked Asn residue would not preclude binding. Hence, MVL can readily bind to the core tetrasaccharide that is *N*-linked to surface residues on gp120. MVL inhibits HIV-1 envelope-mediated cell fusion with an IC_{50} of 30–40 nM, depending on HIV-1 strains, and the process involves binding of at least two molecules of MVL to two independent sites on gp120 (15). The K_D values for binding the individual sites on gp120 are 70–100 nM, comparable with the value of 200 nM for the binding of Man₆GlcNAc₂ and Man₉GlcNAc₂ (15).

Concluding Remarks—In this study we have presented the crystal structures of the cyanobacterial protein MVL, free and in complex with the core pentasaccharide Man₃GlcNAc₂ at 1.9- and 1.8-Å resolution, respectively. The details of this structure show extensive interaction between two monomers of MVL forming a highly stable, symmetric homodimer with four carbohydrate binding sites. Each binding site establishes a complementary pocket that is essentially preformed to bind the Man$\beta(1{\to}4)$GlcNAc$\beta(1{\to}4)$GlcNAc reducing trisaccharide core element with six direct intermolecular hydrogen bonds. Additional contacts and hydrogen bonds form with the branching mannose residues that are entirely consistent with and further explain the thermodynamics of oligomannoside binding by MVL. MVL recognizes a core element in *N*-linked carbohydrates such that MVL binds tightly to gp120 from HIV and inhibits HIV envelope-mediated cell fusion.

The crystal structure of MVL described here is only the second structure of a cyanobacterial carbohydrate-binding protein solved to date. Moreover, it is the second cyanobacterial protein structure shown to contain two carbohydrate binding sites within a single polypeptide chain, the first being cyanovirin-N, which recognizes terminal Man$\alpha(1{\to}2)$Manα disaccharides present on the branching arms of oligomannose-8 and oligomannose-9 (7, 8). MVL and cyanovirin-N bear no sequence or structural similarities to one another and exhibit distinct high-affinity carbohydrate specificities. The high-resolution structure of MVL complexed to Man₃GlcNAc₂ furthers our understanding of specific protein-carbohydrate recognition and provides insight for future design of highly selective, carbohydrate-binding therapeutic agents.

REFERENCES

1. Freed, E. O., and Martin, M. A. (1995) *J. Biol. Chem.* **270,** 23883–23886
2. Moore, J. P., Trkola, A., and Dragic, T. (1997) *Curr. Opin. Immunol.* **9,** 551–562
3. Eckert, D. M., and Kim, P. S. (2001) *Annu. Rev. Biochem.* **70,** 777–810
4. Kwong, P. D., Wyatt, R., Robinson, J. Sweet, R. W., Sodroski, J. A., and Hendrickson, W. A. (1998) *Nature* **393,** 648–659
5. Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu. X., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M., Saag, M. S., Komarova, N. L., Nowak, M. A., Hahn, B. H., Kwong, P. D., and Shaw, G. M. (2003) *Nature* **422,** 307–312
6. Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M. R., Stanfield, R. L., Roux, K. H., Kelly, J. W., Rudd, P. M., Dwek, R. A., Kattinger, H., Burton, D. R., and Wilson, I. A. (2003) *Science* **300,** 2065–2071
7. Bewley, C. A., and Otero-Quintero, S. (2001) *J. Am. Chem. Soc.* **123,** 3892–3902
8. Bewley, C. A. (2001) *Structure (Camb.)* **10,** 931–940
9. Bewley, C. A., Kiyonaka, S., and Hamachi, I. (2002) *J. Mol. Biol.* **322,** 881–889
10. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R., and Wlodawer, A. (2002) *J. Biol. Chem.* **277,** 34336–34342
11. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T. Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., Buckeit, R. W., Nara, P. L., Pannel, L. K., Sowder, R. C., and Henderson, L. E. (1997) *Antimicrob. Agents Chemother.* **41,** 1521–1530
12. Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M., and Gronenborn, A. M. (1998) *Nat. Struct. Biol.* **5,** 571–578
13. Tsai, C. C., Emau, P., Jiang, Y., Tian, B., Morton, W. R., Gustafson, K. R., and Boyd, M. R. (2003) *AIDS Res. Hum. Retroviruses* **19,** 535–541
14. Yamaguchi, M., Ogawa, T., Muramoto, K., Kamio, Y., Jimbo, M., and Kamiya, H. (1999) *Biochem. Biophys. Res. Commun.* **265,** 703–708
15. Bewley, C. A., Cai, M., Ray, S., Ghirlando, R., Yamaguchi, M., and Muramoto, K. (2004) *J. Mol. Biol.* **338,** 901–914
16. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
17. Taylor, G. (2003) *Acta Crystallogr. Sect. D* **59,** 1881–1990
18. Terwilinger, T. C. (2003) *Methods Enzymol.* **374,** 22–37
19. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
20. Brünger, A. T., Adams. P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
21. CCP4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
22. Navaza, J. (1990) *Acta Crystallogr. Sect. A* **46,** 619–620
23. Schwieters, C. D., and Clore, G. M. (2001) *J. Magn. Reson.* **149,** 239–244
24. Carson, M. (1991) *J. Appl. Crystallogr.* **24,** 958–961
25. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11,** 281–296
26. Bax, A, Kontaxis, G., and Tjandra, N. (2001) *Methods Enzymol.* **339,** 127–174
27. Ruckert, M., and Otting, G. (2000) *J. Am. Chem. Soc.* **122,** 7793–7797
28. Zweckstetter, M., and Bax, A. (2000) *J. Am. Chem. Soc.* **122,** 3791–3792
29. Clore, G. M., and Garrett, D. S. (1999) *J. Am. Chem. Soc.* **126,** 2923–2938
30. Jones, S., and Thornton, J. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13–20
31. Williams, D. C., Cai, M., and Clore, G. M. (2004) *J. Biol. Chem.* **279,** 1449–1457
32. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
33. Wormald, M. R., Petrescu, A. J., Pao, Y.-L., Glithero, A., Elliott, T., and Dwek, R. A. (2002) *Chem. Rev.* **102,** 371–386
34. Botos, I., and Wlodawer, A. (2005) *Prog. Biophys. Mol. Biol.* **88,** 233–282
35. Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N., and Gregory, T. J. (1990) *J. Biol. Chem.* **265,** 10373–10382
36. Laskowski, R. A., MacArhur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291