J|A|C|S ARTICLES

Published on Web 10/26/2009

# Unprecedented Glycosidase Activity at a Lectin Carbohydrate-Binding Site Exemplified by the Cyanobacterial Lectin MVL

Syed Shahzad-ul-Hussan,[†] Mengli Cai,[‡] and Carole A. Bewley*[,†]

*Laboratories of Bioorganic Chemistry and Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892*

Received July 16, 2009; E-mail: caroleb@mail.nih.gov

***Abstract:*** Carbohydrate binding proteins, or lectins, are engendered with the ability to bind specific carbohydrate structures, thereby mediating cell−cell and cell−pathogen interactions. Lectins are distinct from carbohydrate modifying enzymes and antibodies, respectively, as they do not carry out glycosidase or glycosyl transferase reactions, and they are of nonimmune origin. Cyanobacterial and algal lectins have become prominent in recent years due to their unique biophysical traits, such as exhibiting novel protein folds and unusually high carbohydrate affinity, and ability to potently inhibit HIV-1 entry through high affinity carbohydrate-mediated interactions with the HIV envelope glycoprotein gp120. The antiviral cyanobacterial lectin *Microcystis viridis* lectin (MVL), which contains two high affinity oligomannose binding sites, is one such example. Here we used glycan microarray profiling, NMR spectroscopy, and mutagenesis to show that one of the two oligomannose binding sites of MVL can catalyze the cleavage of chitin fragments (such as chitotriose) to GlcNAc, to determine the mode of MVL binding to and cleavage of chitotriose, to identify Asp75 as the primary catalytic residue involved in this cleavage, and to solve the solution structure of an inactive mutant of MVL in complex with this unexpected substrate. These studies represent the first demonstration of dual catalytic activity and carbohydrate recognition for discrete oligosaccharides at the same carbohydrate-binding site in a lectin. Sequence comparisons between the N- and C-domains of MVL, together with the sequences of new MVL homologues identified through bioinformatics, provide insight into the evolving roles of carbohydrate recognition.

## Introduction

The complex carbohydrate structures and/or carbohydrate binding proteins that decorate the surfaces of all cells and most viruses and pathogens directly mediate specific adhesion and binding events made possible through precise interactions with cell- and tissue-specific expression of these receptors.[1] Examples include biological processes as fundamental yet diverse as tumor cell migration,[2] fertilization,[3] bacterial and viral infections,[4] and symbiont acquisition.[5] The carbohydrate binding proteins governing these interactions are referred to collectively as lectins and possess a number of defining traits: In general, lectins bind their assigned mono- or disaccharide ligands with weak equilibrium dissociation constants ($K_D$'s) on the order of 0.1−1 mM,[1] and assemble as homodimers, trimers or oligomers to achieve multivalent binding and avidity. Lectins are distinct from antibodies and sugar modifying enzymes in that they are of nonimmune origin and do not catalyze glycosyl transferase nor glycosidase reactions.[6] Though the 3-dimensional structures of the vast majority of lectin structures fall into known protein families or folds, their structures do not reliably predict carbohydrate specificity; thus, lectins are oftentimes described on the basis of their protein family as well as carbohydrate specificity.

Through recent developments in carbohydrate synthesis and screening techniques, the utility of lectins has expanded beyond histology and agglutination assays, where lectin microarrays in particular have been used to reveal temporal changes in bacterial cell surface glycans,[7] to identify carbohydrate antigens during the course of tumor cell progression,[8] and to compare the glycomes of viral envelopes to their host cell.[9] In recent years, lectins that potently inhibit membrane fusion by HIV-1 and other enveloped viruses also have been identified, reminding us of their potential utility as therapeutics. Among those that inhibit HIV-1 envelope-mediated membrane fusion at nanomolar

[†] Laboratory of Bioorganic Chemistry.
[‡] Laboratory of Chemical Physics.

(1) Reviewed in: (a) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357. (b) Lis, H.; Sharon, N. *Annu. Rev. Biochem.* **1986**, *55*, 35.
(2) (a) Kannagi, R.; Izawa, M.; Koiki, T.; Miyazaki, K.; Kimura, N. *Cancer Sci.* **2004**, *95*, 377. (b) Gorelik, E.; Galili, U.; Raz, A. *Cancer Metastasis Rev.* **2001**, *20*, 245.
(3) Reviewed in: Primakoff, P.; Myles, D. G. *Science* **2002**, *296*, 2183. Nixon, B.; Aitken, R. J.; McLaughline, E. A. *Cell. Mol. Life Sci.* **2007**, *64*, 1805.
(4) *Protein-carbohydrate Interactions in Infectious Diseases*, 1st ed.; Bewley, C. A., Ed.; Royal Society of Chemistry: Cambridge, 2006.
(5) (a) Nyholm, S. V.; Stabb, E. V.; Ruby, E. G.; McFall-Ngai, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10231. (b) D'Haeze, W.; Holsters, M. *Glycobiology* **2002**, *12*, 79r for examples.
(6) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321. (a) Goldstein, I. J.; Hughes, R. C.; Monsigny, M.; Osawa, T.; Sharon, N. *Nature* **1980**, *285*, 66. (b) Kocourek, J.; Horejší, V. *Nature* **1981**, *290*, 188. (c) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321.
(7) Hsu, K. L.; Pilobello, K. T.; Mahal, L. K. *Nat. Chem. Biol.* **2006**, *2*, 153.
(8) Hsu, K. L.; Gildersleeve, J. C.; Mahal, L. K. *Mol. Biosyst.* **2008**, *4*, 654.
(9) Krishnamoorthy, L.; Bess, J. W., Jr.; Preston, A. B.; Nagashima, K.; Mahal, L. K. *Nat. Chem. Biol* **2009**, *5*, 244.

 **10.1021/ja905929c** 

concentrations or lower are cyanovirin-N (CVN),[10] scytovirin,[11] *Microcystis viridis* lectin (MVL),[12] *Oscillatoria agardhii* agglutinin (OAA),[13] and griffithsin (GRFT).[14] Remarkably, none of these five lectins share homology at either the primary sequence or structural level. While fine carbohydrate specificity has thus far been determined at an atomic level for CVN[15,16] and MVL[17] only, initial studies employing mono- or disaccharides indicate that each will demonstrate unique fine specificities distinct from the others.[18]

In addition to their potent antiviral activities, these lectins offer elements of novelty at several levels. X-ray and/or NMR structures, determined for all but OAA, reveal that each of these lectins possesses a novel 3-dimensional fold, in at least one case founding a new protein family,[19] and high resolution structures of CVN and MVL in complex with an optimal oligosaccharide presented new carbohydrate binding sites and modes of carbohydrate recognition. In this study we have used glycan array profiling, NMR spectroscopy, isothermal titration calorimetry, and mutagenesis to show that one of the two oligomannose binding sites of MVL can catalyze the cleavage of GlcNAc$\beta$(1−4)GlcNAc$\beta$(1−4)GlcNAc ($GlcNAc_3$, commonly known as chitotriose) to GlcNAc monosaccharides, to establish the mode of binding to $GlcNAc_3$, to identify the primary catalytic residue involved in this cleavage, and to solve by NMR the structure of an inactive mutant of MVL in complex with this unexpected substrate. These studies represent the first demonstration of dual catalytic activity and carbohydrate recognition for discrete oligosaccharides at the same carbohydrate-binding site in a lectin. Sequence comparisons between the N- and C-domains of MVL, together with the sequences of new MVL homologues identified through bioinformatics, provide insight into the evolving roles of carbohydrate recognition.

## Results and Discussion

**Glycan Array Profiling Reveals Unexpected Binding of MVL to Chitotriose and Chitotetraose.** Subsequent to our finding that the cyanobacterial protein MVL potently inhibits HIV-1 entry through carbohydrate-mediated interactions with the HIV-1 surface envelope glycoprotein gp120, we used NMR, X-ray crystallography, and other biophysical techniques to clearly demonstrate the biochemical and structural basis for MVL's fine specificity to $Man_3GlcNAc_2$. A potential shortcoming to the completeness of those studies stemmed from our sole use of commercially available mono-, di-, and trisaccharide fragments of $Man_9GlcNAc_2$ for our initial screening. To gain a more complete recognition profile for MVL, wild type recombinant protein fluorescently labeled with Alexafluor 488 was submitted for profiling on a glycan microarray containing ~200 complex glycans developed by the Consortium for Functional Glycomics.[20] The degree of binding to each of the glycans is shown in Figure 1.[21] Consistent with our NMR and biophysical studies, MVL bound all high mannose oligosaccharides containing at least the high affinity core structures Man$\alpha$(1−6)Man$\beta$(1−4)-GlcNAc$\beta$(1−4)GlcNAc (Man2A) or $Man_3GlcNAc_2$. Surprisingly, MVL also bound *N*,*N*′,*N*″-triacetyl chitotriose ($GlcNAc_3$ or chitotriose) and *N*,*N*′,*N*″,*N*‴-tetraacetyl chitotetraose ($GlcNAc_4$ or chitotetraose). This result was unexpected given the 1.8 Å crystal structure of a 1:2 complex of MVL/$Man_3GlcNAc_2$ that showed MVL to contain two nearly identical Y-shaped carbohydrate binding sites that exhibit a perfect shape complementarity to, and hydrogen bonding network with, the branched $Man_3GlcNAc_2$ structure.

This apparent dual specificity for distinct complex carbohydrate structures is unusual. To confirm this result, binding of MVL to $GlcNAc_3$ and $GlcNAc_4$ was verified by isothermal titration calorimetry, which gave equilibrium dissociation constants ($K_D$'s) of 27 ± 5 and 60 ± 6 $\mu$M, respectively. The binding curves for each of these glycans could only be fit to a 1-site model with a stoichiometry of two. Thus MVL binds $GlcNAc_3$ and $GlcNAc_4$ through two equivalent binding sites with respective $K_D$ values that are 10 and 20 times weaker than those for MVL binding its high mannose ligand $Man_3GlcNAc_2$.

**NMR Determination of Location and Mode of Binding of Chitotriose to MVL.** MVL exists as a symmetrical, obligate homodimer, where each monomer contains two separate binding sites for $Man_3GlcNAc_2$ located in the N- and C-domains of the protein. With complete NMR assignments for MVL in hand, we used chemical shift mapping to identify the binding sites for chitotriose and chitotetraose on MVL. $^1H-^{15}N$ correlation spectra were recorded on samples containing $^{15}N$-labeled MVL free in solution and in the presence of increasing amounts (0.25 equiv) of each saccharide where sample spectra of the titration of $GlcNAc_3$ to MVL are shown in Figure 2. Upon addition of either of $GlcNAc_3$ or $GlcNAc_4$, resonances surrounding the $GlcNAc_2$ core of the carbohydrate binding sites exhibited changes in chemical shifts and/or increased line broadening. In particular, individual cross peaks corresponding to the free and bound form for residues located around the reducing GlcNAc ring could be observed during the course of the titration (Gly10 and Gly69, for example) indicating slow exchange on the NMR time scale. Residues making contact with the nonreducing end of chitotriose, on the other hand, mainly underwent line broadening indicating intermediate exchange at this region of the binding site. Regardless of the carbohydrate used, addition of two equivalents of ligand were required to complete the titration, and resonances located in the N- and C-domains shifted or broadened at the same rate (right panel, Figure 2). These data confirm that like $Man_3GlcNAc_2$, MVL binds chitotriose and chitotetraose through two symmetrically related carbohydrate-binding sites that exhibit similar affinities relative to one another.

(10) Boyd, M. R.; et al. *Antimicrob. Agents Chemother.* **1997**, *41*, 1521.
(11) Bokesch, H. R.; O'Keefe, B. R.; McKee, T. C.; Pannell, L. K.; Patterson, G. M.; Gardella, R. S.; Sowder, R. C., 2nd; Turpin, J.; Watson, K.; Buckheit, R. W., Jr.; Boyd, M. R. *Biochemistry* **2003**, *42*, 2578.
(12) Bewley, C. A.; Cai, M.; Ray, S.; Ghirlando, R.; Yamaguchi, M.; Muramoto, K. *J. Mol. Biol.* **2004**, *338*, 901.
(13) Sato, Y.; Okuyama, S.; Hori, K. *J. Biol. Chem.* **2007**, *282*, 11021.
(14) O'Keefe, B. R.; Vojdani, F.; Buffa, V.; Shattock, R. J.; Montefiori, D. C.; Bakke, J.; Mirsalis, J.; d'Andrea, A. L.; Hume, S. D.; Bratcher, B.; Saucedo, C. J.; McMahon, J. B.; Pogue, G. P.; Palmer, K. E. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6099.
(15) Bewley, C. A. *Structure* **2001**, *10*, 931.
(16) Botos, I.; O'Keefe, B. R.; Shenoy, S. R.; Cartner, L. K.; Ratner, D. M.; Seeberger, P. H.; Boyd, M. R.; Wlodawer, A. *J. Biol. Chem.* **2002**, *37*, 34336.
(17) Williams, D. C., Jr.; Lee, J. Y.; Cai, M.; Bewley, C. A.; Clore, G. M. *J. Biol. Chem.* **2005**, *280*, 29269.
(18) Ziólkowska, N. E.; Shenoy, S. R.; O'Keefe, B. R.; Wlodawer, A. *Protein Sci.* **2007**, *16*, 1485.
(19) Percudani, R.; Montanini, B.; Ottonello, S. *Proteins* **2005**, *60*, 670.
(20) Blixt, O. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17033.
(21) Data sets H:226 (MVL) and H:1170 (MVL mutant II) are publicly available at http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp (accessed Oct 21, 2009) or upon request from the authors.
(22) Mayer, M.; Meyer, B. *J. Am. Chem. Soc.* **2001**, *123*, 6108.

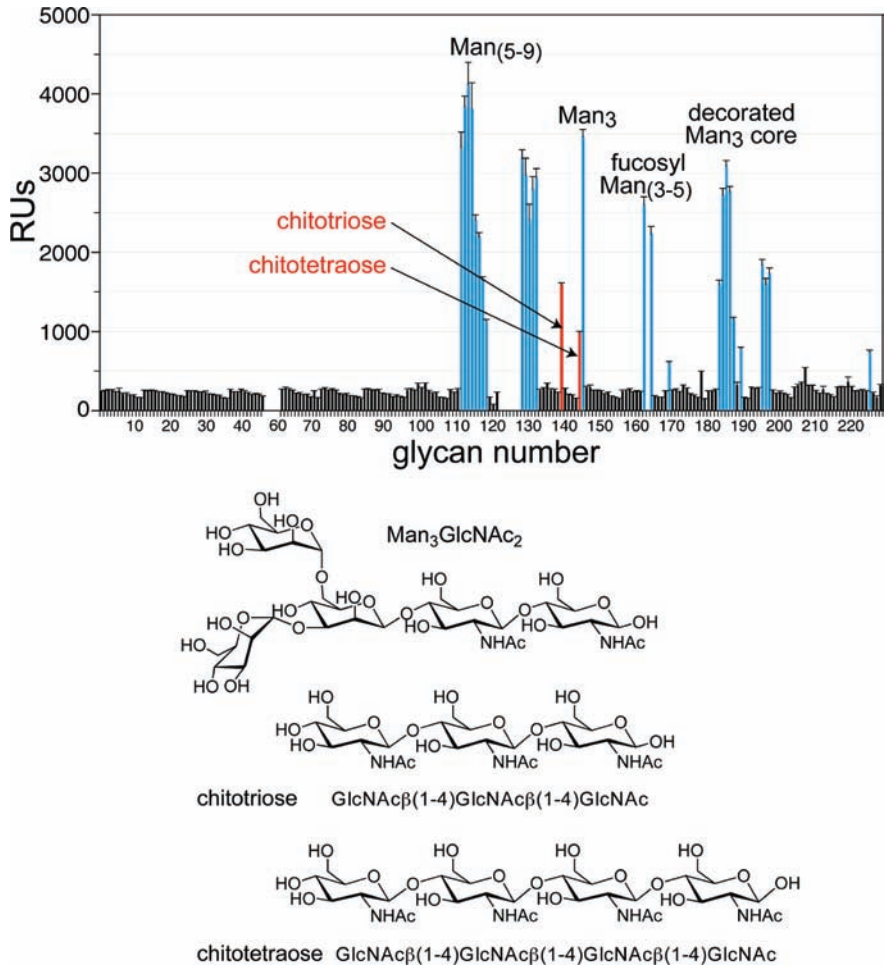


***Figure 1.*** Glycan array profiling of MVL and structures of representative glycans, $Man_3GlcNAc_2$, *N*,*N*′,*N*″ -triacetyl chitotriose ($GlcNAc_3$) and *N*,*N*′,*N*″,*N*‴-tetraacetyl chitotetraose ($GlcNAc_4$), to which MVL binds. Fluorescence response units (RU) correspond to the amount of fluorescently labeled MVL binding to each glycan on the array.

These data are consistent with the ITC experiments whose data were best fit with a stoichiometry of two saccharides per MVL monomer.

To determine the mode and orientation of binding for $GlcNAc_3$ and $GlcNAc_4$, Saturation Transfer Difference (STD) NMR[21] experiments were recorded on samples containing a 60-fold excess of carbohydrate relative to MVL using published experimental conditions.[23] An expansion of the reference and difference spectra showing the *N*-acetyl groups of $GlcNAc_3$ in complex with MVL is shown in Figure 3 where the difference spectrum was normalized to the *N*-acetyl signal of greatest intensity at $\delta$ 2.04. The difference spectrum shows the greatest enhancement (100%) for the $COCH_3$ group of the reducing GlcNAc (ring A). Spectral overlap between $COCH_3$ groups on rings B and C prevented quantification of their individual STD enhancements; however, integration of their overlapping signals yielded a combined enhancement of ~20%. Similar profiles were oberserved in the difference spectra for $Man_3GlcNAc_2$, Man2A and $GlcNAc_4$ in the presence of MVL where the reducing $COCH_3$ group gave the strongest enhancement for all four carbohydrates (spectra supplied in Supporting Information). In contrast, no measurable enhancements were detected for oligomannose fragments Manα(1−3)[Manα (1−6)]Man (mannotriose) and Manα(1−4)GlcNAc in the presence of MVL. Together with the crystal structure of MVL:$Man_3GlcNAc_2$, that revealed a deep hydrophobic pocket surrounding the $COCH_3$ group on the reducing ring, this pattern of enhancements clearly defines the mode and orientation of binding for both types of glycans. Thus, whether present in high mannose oligosaccharides or *N*-acetyl-chitooligosaccharides (chitin fragments), the $GlcNAc_2$ core anchors the carbohydrate to MVL by binding the hydrophobic cleft present in both carbohydrate binding sites. Further, this mode of binding requires that both types of glycans bind MVL in the same orientation.

**Endoglycosidase Activity and Catalysis of MVL toward $GlcNAc_3$ and $GlcNAc_4$.** Dual recognition for distinct complex carbohydrate structures as displayed by MVL binding both the $Man_3GlcNAc_2$ core of high mannose oligosaccharides and chitin fragments is of interest to those studying carbohydrate recognition. To establish the structural basis for this dual specificity, we sought to determine the solution structure of MVL in complex with chitotriose using multidimensional NMR techniques. Samples containing complexes of 1:2 uniformly [$^{13}C$,$^{15}N$]-labeled MVL/$GlcNAc_3$ (1:2 mM) were prepared and HSQC spectra recorded to confirm saturation at both carbohydrate binding sites. On two separate occasions, however, $^{1}H-^{15}N$ correlation spectra recorded 16 h later appeared identical to spectra for free rather than chitotriose-bound MVL, suggesting that MVL was catalyzing the cleavage of $GlcNAc_3$ to GlcNAc.

(23) Lam, S. N.; Acharya, P.; Wyatt, R.; Kwong, P. D.; Bewley, C. A. *Bioorg. Med. Chem.* **2008**, *16*, 10113.

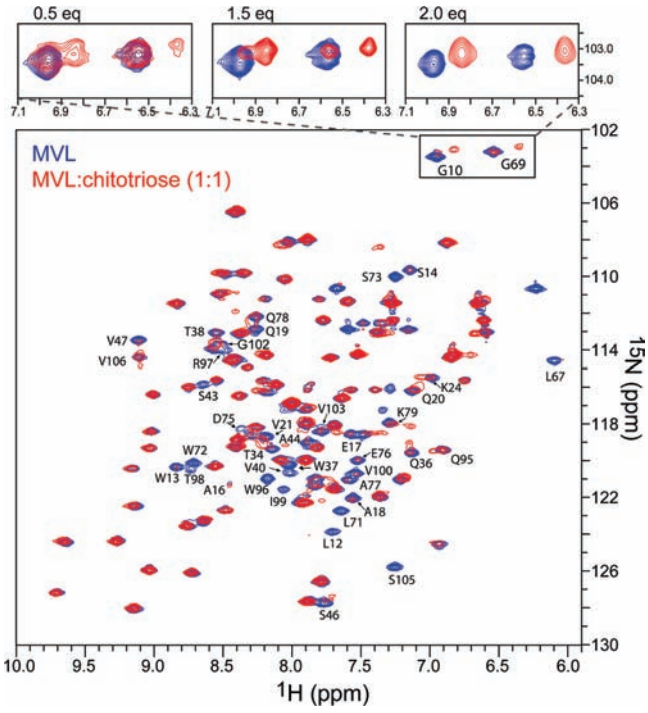


***Figure 2.*** NMR titration of chitotriose to MVL. $^{1}H-^{15}N$ correlation spectra of free MVL (blue) and MVL in the presence of 1 equiv of chitotriose (red), with expansions (above) of an upfield region of the spectrum containing cross peaks for Gly10 and Gly69 of N- and C-domains of MVL, respectively, in the presence of 0.5, 1.5, and 2.0 equiv of chitotriose. Concomitant appearance of cross peaks corresponding to chitotriose-bound resonances indicates two equivalent binding sites on MVL thus requiring 2 equiv of carbohydrate for 1:2 stoichiometric binding.

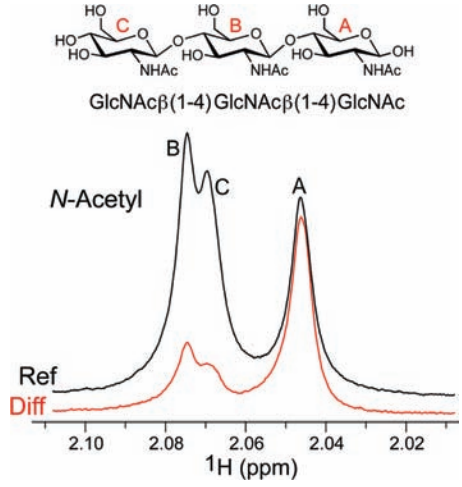


***Figure 3.*** Saturation transfer difference NMR spectra of MVL in the presence of chitotriose. Reference and difference spectra are shown in black and red, respectively, with the difference spectrum normalized to the *N*-acetyl group of the reducing ring A, which gave the strongest enhancement.

To investigate this possibility, the stabilities of $GlcNAc_3$, $GlcNAc_4$, Man2A, and $Man_3GlcNAc_2$ in the presence of MVL (0.2 mM MVL, 2.0 mM ligand) were monitored by $^{1}H$ NMR. The results for $GlcNAc_3$ are shown in Figure 4. At $t = 0$, the $^{1}H$ NMR spectrum for chitotriose in the presence of MVL is similar to that of free chitotriose with the exception of slight line broadening due to binding this 26 kDa protein. After 96 h, signals for chitotriose had disappeared and signals corresponding to GlcNAc had appeared. Further, the line widths of the *N*-acetyl

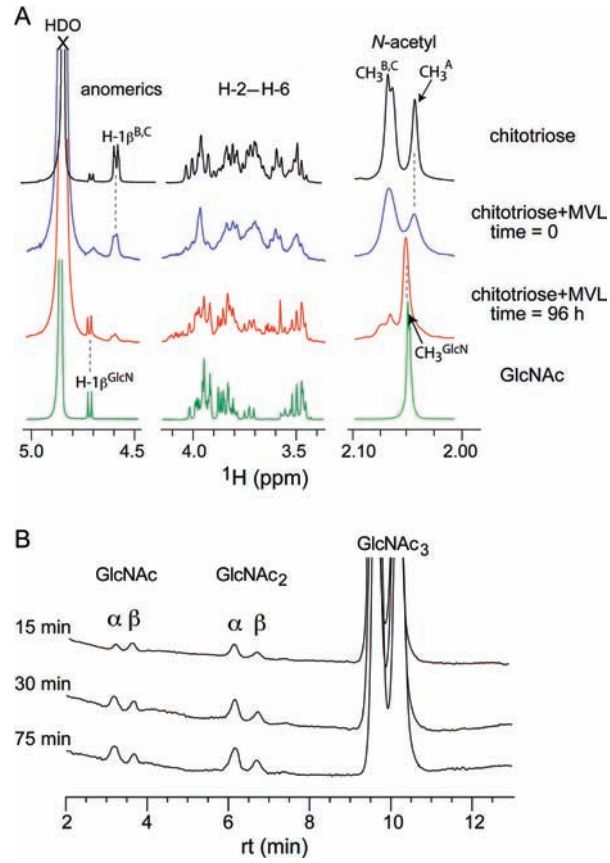


***Figure 4.*** Cleavage of chitotriose ($GlcNAc_3$) by MVL monitored by NMR and HPLC. (A) Expansions of $^{1}H$ NMR spectra (600 MHz) of free chitotriose (black, top), chitotriose in the presence of MVL at $t = 0$ (blue) and $t = 96$ h (red), and free *N*-acetyl glucosamine (green, bottom) as reference. Diagnostic signals used to monitor cleavage of chitotriose are labeled and include its $COCH_3$ group on ring A ($\delta$ 2.04), and anomeric protons of rings A ($\delta$ 4.72), and rings B and C ($\delta$ 4.57−4.59). In a sample containing 1:10 MVL/chitotriose, signals corresponding to the trisaccharide disappear with time, and by 96 h are replaced by signals corresponding to free GlcNAc. (B) HPLC traces of chitotriose with time showing cleavage to GlcNAc and chitobiose ($GlcNAc_2$). The ratio of $\alpha/\beta$ anomers for $GlcNAc_2$ stays constant with time (1.6:1) while the ratio for GlcNAc increases from 0.6:1 to 1.6:1 indicating cleavage between rings B and C with retention of configuration.

signals were much narrower than those of chitotriose indicating absence of binding to MVL. Similar results were obtained when chitotetraose was incubated with MVL where cleavage to monosaccharide was complete within 96 h under the same conditions (200 $\mu$M MVL, 2 mM $GlcNAc_4$). In contrast, $Man_3GlcNAc_2$ and Man2A in the presence of MVL were stable indefinitely, as were solutions of $GlcNAc_3$ and $GlcNAc_4$ in the absence of MVL. Thus, MVL catalyzes the cleavage of chitotriose and chiotote traose to GlcNAc, where cleavage is complete within 16 h for stoichiometric complexes at mM concentrations (spectra provided in Supporting Information), and 96 h for complexes where MVL is present at $\mu$M concentrations and *N*-acetyl-chitooligosaccharides at mM concentrations.

We next investigated the mechanism by which MVL cleaves $GlcNAc_3$ by monitoring cleavage and measuring by HPLC the ratios of $\alpha/\beta$ anomers formed as a function of time. Because chitin polymers are composed of $\beta$-linked GlcNAc residues, and the anomeric center of $\beta$-GlcNAc undergoes mutarotation in 30−60 min to reach its equilibrium

***Table 1.*** Ratios of $\alpha/\beta$ Anomers with Time

| time (min) | GlcNAc | chitobiose |
|---|---|---|
| 15 | 0.6:1 | 1.6:1 |
| 45 | 1.3:1 | 1.6:1 |
| 75 | 1.6:1 | 1.7:1 |

distribution ratio of 1.6:1 $\alpha/\beta$,[24] the mechanism of cleavage can be studied by measuring the size and anomeric ratios of the fragments formed upon enzymatic cleavage. Solutions containing $GlcNAc_3$ and MVL were prepared at 27 °C and aliquots taken for HPLC analysis every 15 min. As seen in Figure 4b, $GlcNAc_2$ was formed with an $\alpha/\beta$ ratio of 1.6:1 at all time points sampled, the value expected at equilibrium. In contrast, GlcNAc appeared in predominantly $\beta$ form with an $\alpha/\beta$ ratio of 0.6:1 (15 min), and underwent mutarotation with time to reach an equilibrium value of 1.6:1 within 75 min (Table 1). Together with our observations that MVL is incapable of cleaving high mannose oligosaccharides or carbohydrates containing a Man$\beta$(1−4)GlcNAc$\beta$(1−4)GlcNAc moiety, these data show that MVL must be cleaving $GlcNAc_3$ at the nonreducing end, between rings B and C and with retention of configuration, thereby exhibiting *N*-acetyl-$\beta$-glucosidase activity.

The mode of cleavage of chitotetraose was also investigated by performing analogous experiments where cleavage of $GlcNAc_4$ by MVL was monitored by HPLC. For this tetrasaccharide, a single product was obtained, namely that of the disaccharide $GlcNAc_2$ (Supporting Information). Thus, like $GlcNAc_3$, MVL cleaves $GlcNAc_4$ at the glycosidic bond separating rings B and C.

**Structure-Guided Identification of a Catalytic Residue.** Further structural studies of MVL in complex with chitotriose required construction of a mutant that lacked enzymatic activity. The cocrystal structure of MVL in complex with $Man_3GlcNAc_2$ and a model of MVL in complex with $GlcNAc_3$ were used to help identify potential catalytic residue/s involved in the cleavage of *N*-acetyl-chitooligosaccharides. Relative to the position of the 2−3 glycosidic bond in $GlcNAc_3$, the model showed Asp75 located in the C-domain of MVL to be the only likely candidate. A single site Asp75Ala mutant was constructed and its binding to and cleavage of $GlcNAc_3$ interrogated by NMR. $^1H-^{15}N$-correlation spectra monitoring titration of D75A with chitotriose appeared identical to those of wild type MVL in all respects except for replacement of the Asp75 signal for that of Ala75, confirming that binding to $GlcNAc_3$ was conserved in both domains of the mutant. Lending further support to the preservation of carbohydrate specificity for this mutant, D75A MVL showed indistinguishable inhibitory activity relative to wild type MVL in an HIV infectivity assay, and glycan array profiling of fluorescently labeled D75A yielded data similar to wild type MVL. Catalytic activity of D75A MVL on $GlcNAc_3$ was assessed by $^1H$ NMR where spectra recorded up to 7 days after preparation of stoichiometric complexes contained signals for chitotriose only, and signals for GlcNAc did not appear until >3 weeks of storage at room temperature. The same results were observed for complexes of D75A MVL and chitotetraose. Thus, Asp75 is a key catalytic residue for the glycosidase activity of MVL. Further, these results established that an Asp75Ala mutant could be used for long-term NMR studies with chitotriose without the carbohydrate undergoing cleavage.

**Solution Structure of D75A MVL in Complex with $GlcNAc_3$.** Previously we solved X-ray crystal structures of free MVL and a 1:2 complex of MVL/$Man_3GlcNAc_2$. At that time, complete $^1H$, $^{13}C$, and $^{15}N$ assignments were made for both protein and carbohydrate using multidimensional NMR techniques. To determine the solution structure of the D75A mutant in complex with chitotriose, complete assignments were made in a similar fashion on a stoichiometric complex using standard multidimensional double and triple resonance NMR experiments. Interproton distance restraints were obtained from 3-D $^{15}N$-separated and $^{13}C$-separated NOE and $^{12}C$-filtered/$^{13}C$-separated intermolecular NOE experiments. NMR titration studies with chitin fragments showed clearly that chemical shift perturbations occurred only for residues in or around the carbohydrate binding sites. Likewise, direct comparison of the 3-D $^{15}N$- and $^{13}C$-separated NOE data recorded on free and bound forms of the protein showed identical NOE patterns for residues not involved in carbohydrate binding. On this basis, we calculated the 3-D structure of 1:2 D75A MVL/$GlcNAc_3$ using conjoined rigid-body/torsion angle dynamics[15,25,26] where the backbone and side chains of residues not involved in carbohydrate binding were held rigid. At the same time interfacial side chains were given rotational degrees of freedom and the carbohydrate was given rotational and translational degrees of freedom during the calculations. Interproton distance restraints between D75A MVL and $GlcNAc_3$ used in the calculations were derived from $^{12}C$-filtered/$^{13}C$-separated NOE spectra, an example of which is shown in Figure 5A. Superimpositions for the 25 lowest energy structures showing close-ups of the N- and C-domains are shown in Figure 5B and C, and structural statistics are presented in Table 2.

$GlcNAc_3$ binds MVL in the carbohydrate binding sites located at the distal ends of each domain (Figure 6). The saccharide binds in an extended conformation with glycosidic $\phi$ (O5−C1−O−C(x)′) and $\psi$ (C1−O−C(x)′−C(x−a)′ angles between rings C and B averaging $-59.7 \pm 0.1°$ and $118.3 \pm 1.6°$, respectively, and between rings B and A averaging $-57.3 \pm 0.4°$ and $104.7 \pm 3.7°$, respectively. At the reducing end, interactions between MVL and $GlcNAc_3$ resemble those observed with $Man_3GlcNAc_2$. For example, the acetate methyl of ring A inserts deep into the hydrophobic pocket formed by and in van der Waals contact with the side chains or C$\alpha$ protons of Pro-11/Pro-70, Trp-13/Trp-72 and Gly 102, and the methyl groups of Leu-12/Leu-71 and Thr-39/Thr-98 (Figure 6, right). This mode of binding is consistent with the strong STD enhancements observed for ring A. In ring B, the carbonyl oxygens of Leu-12/Leu-71 are within hydrogen bonding distance of the acetyl NH, as are the hydroxyl groups of Thr-39/Thr-98 with O-6. Extensive hydrophobic contacts are made between the indole ring of Trp-37/Trp-96 and H-4—H-6 of ring B, as well as H-5 and H-6 of ring C. At the top of the cleft, the side chain O atom of Gln-36/Gln-95 is within hydrogen bonding distance (2.8 Å) to the O-4 atom of ring C. However, relative to $Man_3GlcNAc_2$, the orientation and position of the nonreducing GlcNAc residue of chitotriose, ring C, differs from that of the branching mannose of $Man_3GlcNAc_2$. In particular, the combination of the presence of a C-2 *N*-acetyl group in chitotriose relative to an axial hydroxyl at the same position in $Man_3GlcNAc_2$, and a $\psi$ angle of *ca.* 118° versus 136° for the GlcNAc$\beta$(1−4)GlcNAc linkage in chitotriose versus the Man $\beta$(1−4)GlcNAc linkage in $Man_3GlcNAc_2$ moves the terminal

(24) Kuhn, R.; Haber, F. *Chem. Ber.* **1953**, *86*, 722.
(25) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021.
(26) Bewley, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 1014.

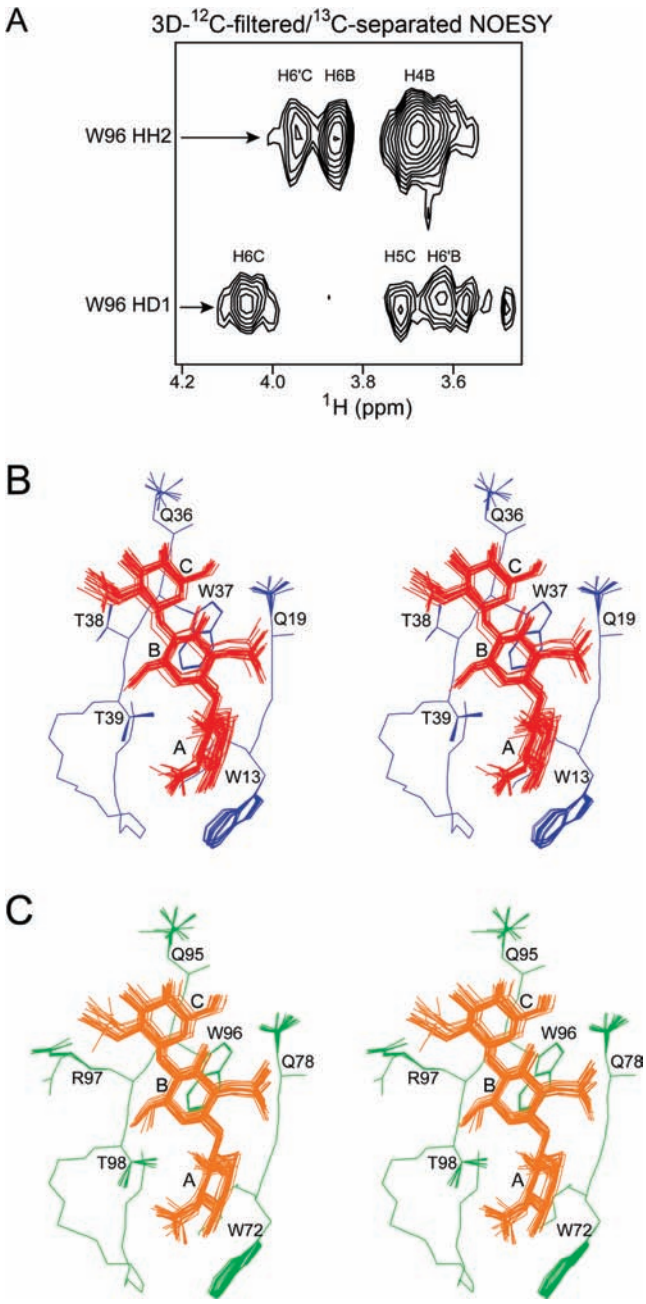


***Figure 5.*** Intermolecular NOE spectrum and superpositions of the NMR ensemble of chitotriose in complex with D75A MVL. (A) Representative $^{13}C$ plane of 3-D $^{12}C$-filtered/$^{13}C$-separated NOE spectrum showing correlations between Trp96 HH2 and HD1 (at $\delta$ 6.06 and 7.17, respectively) and chitotriose. Stereoviews showing 25 structure ensembles of (B) the N-domain and (C) the C-domain of a 1:2 complex of D75A MVL/chitotriose. Protein and carbohydrate are colored blue and red in (B) and green and gold in (C). Side chains displayed correspond to residues for which unambiguous intermolecular NOEs to chitotriose were observed (as in A) and were given rotational degrees of freedom during the structure calculations. Structural statistics appear in Table 2.

GlcNAc residue toward the bottom of the cleft. This subtle movement has the effect of directing additional GlcNAc residues, as would be present in chitotetraose, toward the helix of each domain rather than the $\beta$ strand (Figure 6, right panel). These structural findings are supported by our observation that resonances located in the lower portion of the cleft are shifted upon addition of chitotetraose while those located near the upper sheet go unperturbed (spectra not shown).

***Table 2.*** Structural Statistics for MVL D75A/$GlcNAc_3$ (1:2)

| | N-domain | C-domain |
|---|---|---|
| Distance restraints | | |
| intermolecular NOEs (protein-carbohydrate)[a] | 18 | 19 |
| inter-residue NOEs (carbohydrate) | 4 | 4 |
| Torsion angle restraints | | |
| interfacial side chains of protein[b] | 7 | 8 |
| carbohydrate[c] | 38 | 38 |
| Coordinate precision (Å)[d] | | |
| interfacial side chains of protein | 0.184 (±0.03) | 0.363 (±0.13) |
| carbohydrate | 0.448 (±0.18) | 0.447 (±0.13) |
| Rms deviations from all distance restraints[e] | 0.052 (±0.011) | |
| Rms deviations from all dihedral angle restraints[e] | 0.278 (±0.139) | |

[a] Intermolecular NOE restraints between protein and carbohydrate were derived from 3D $^{12}C$-filtered/$^{13}C$-separated NOE data. [b] Interfacial side chain dihedral angle restraints were determined experimentally from $^{15}N$-separated NOE and ROE experiments. [c] Dihedral angle restraints used in the structure calculations for chitotriose are consistent with a chair conformation for each pyranose unit and were introduced on the basis of intradisaccharide NOEs observed in a 2D $^{12}C$-filtered NOE spectrum. Resonances for the trisaccharide bound through the N-domain and C-domain binding sites are degenerate as observed in the $^{12}C$-filtered NOE spectra of 1:1 and 1:2 MVL D75A/$GlcNAc_3$ complexes; thus, the same dihedral angle restraints were used in calculating the bound structures of each trisaccharide. Dihedral angle restraints for the glycosidic bonds were introduced at the later stages of refinement and were consistent with intersugar NOEs. [d] Precision of the coordinates is defined as the average rms difference between the 25 individual structures and the mean coordinate positions for each site of the complex obtained by best fitting to the protein backbone. (Note the protein backbone and *non*-interfacial side chains are held fixed.) [e] None of the final set of 25 structures exhibited distance violations greater than 0.2 Å or dihedral angle violations greater than 5°.

**Defining the Mechanism of Catalysis.** In terms of catalysis, it was surprising to find that the carboxyl group of the side chain of Asp75 (modeled with PyMol), the residue shown to be critical for glycosidase activity, is located nearly 10 Å away from the glycosidic bond of rings B and C. This is not the first report of its kind, however, as similar observations have been made for other enzymes including serotonin acetyl transferase,[27] an aminoglycoside 2′-*N*-acetyltransferase,[28] and spermine $N^1$-acetyl transferase.[29] Crystal structures of each of these enzymes in complex with their individual substrates revealed indispensable catalytic residues (established by mutagenesis) to be located anywhere from 8−12 Å away from the substrate acceptor. These structural studies suggested catalysis likely occurs through proton relay via a "proton wire".[30,31] Support for this hypothesis came recently with the 2.3 Å crystal structure of a ternary complex comprising spermine $N^1$-acetyl transferase, spermine, and acetyl-CoA[29] that revealed an active site in which none of the demonstrated catalytic residues were within direct or even water mediated hydrogen bonding distance of the substrate. Instead, the structure showed the clear presence of a proton wire comprising two acidic side chains, four water molecules, and

(27) Hickman, A. B.; Namboodiri, M. A. A.; Klein, D. C.; Dyda, F. *Cell* **1999**, *97*, 361.
(28) Vetting, M. W.; Hegde, S. S.; Javid-Majd, F.; Blandhard, J. S.; Roderick, S. L. *Nat. Struct. Biol.* **2002**, *9*, 653.
(29) Montemayor, E. J.; Hoffman, D. W. *Biochemistry* **2008**, *47*, 9145.
(30) Tuckerman, M. E.; Marx, D.; Klein, M. L.; Parrinello, M. *Science* **1997**, *275*, 817.
(31) Kohen, A.; Klinman, J. P. *Acc. Chem. Res.* **1998**, *31*, 397.

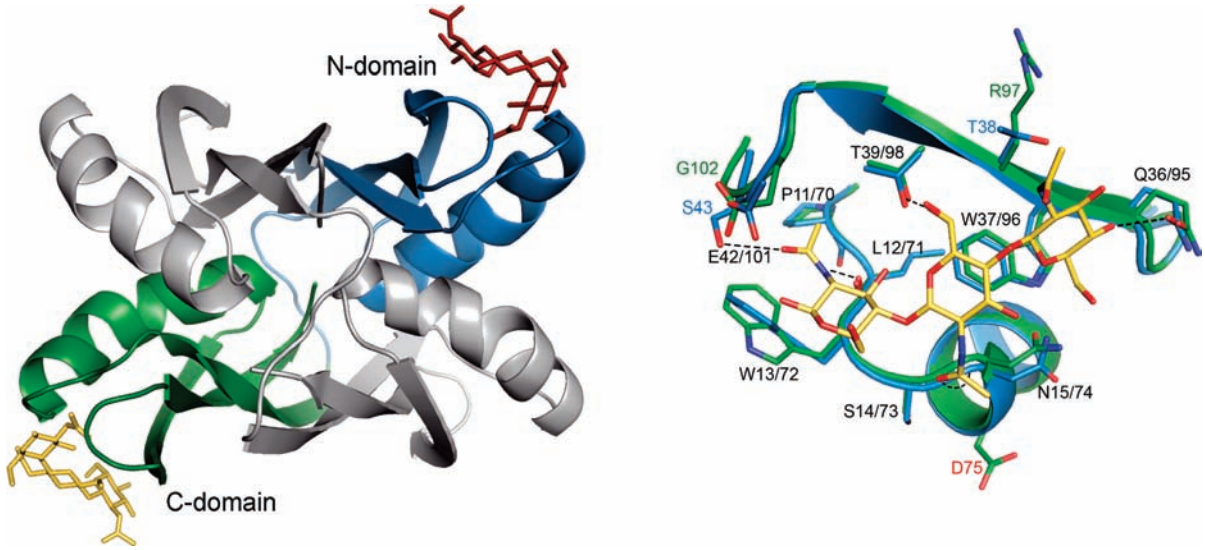


***Figure 6.*** Structure of D75A MVL in complex with chitotriose. (left) Ribbon diagram of homodimeric MVL showing chitotriose bound to a single N-domain (blue and red) and C-domain (green and gold) of one protein molecule. (right) Best-fit superposition of the N- and C-domains of the complex. Side chains containing atoms within 5 Å of chitotriose in either domain are displayed and labeled where each domain contains 11 residues. Of those, only residues in positions 38/97 and 43/102 differ between the N- and C-domains. Dotted lines indicate atoms within direct or water mediated hydrogen bonding distances of chitotriose. Figure generated and distances measured using Pymol.

the amino group of the substrate itself, the combination of which must carry out catalysis.

In our case, numerous attempts to crystallize a complex comprising D75A MVL and chitotriose, including robotic screening of complexes subjected to over a thousand crystallization conditions, as well as seeding with crystalline wild type MVL, have so far been unsuccessful. Nevertheless, in light of the discovery of enzymatic activity at the C-terminal carbohydrate-binding site of MVL, it is instructive to examine the 1.8 Å crystal structure of MVL in complex with $Man_3GlcNAc_2$. In the C-terminal domain, starting from the carboxylate of Asp75, numerous hydrogen bonded water molecules can be traced without interruption to the glycosidic bond between rings B and C (pdb accession number 1zhs). In the N-terminal domain, where the equivalent residue is an alanine (Ala16), no such hydrogen bonded network of water molecules is present and catalytic activity is absent.

## Conclusions

In this study, we have demonstrated by NMR and mutagenesis endoglycosidase activity in the cyanobacterial lectin MVL. This finding is remarkable in that the catalytic activity occurs precisely at a single carbohydrate-binding site that has been characterized biochemically and by X-ray crystallography to be specific for high mannose oligosaccharides. Indeed, this high affinity interaction with oligomannosides provided the basis for defining the protein as a lectin. As discussed earlier, the classification of lectins has remained distinct from enzymes and antibodies due to their lack of glycosidase or glycosyl transferase activity, and their nonimmune origin, respectively. To our knowledge, this is the first example where a carbohydrate-binding site exhibits high affinity for one glycan, here oligomannosides, yet catalyzes the cleavage of a separate family of glycans, namely *N*-acetyl-chitooligosaccharides. By employing complementary NMR techniques, we determined the mode of binding of chitotriose to MVL and measured the time scale and mechanism by which cleavage occurs. Molecular docking and mutagenesis facilitated our identification of Asp75 as a key catalytic residue, allowing us to solve the solution structure of an optimal substrate, chitotriose, in complex with a catalytically deficient mutant, Asp75Ala MVL. Slow rates of cleavage precluded accurate measurement of kinetic rate constants with chitotriose or chitotetraose, as well as a commonly used fluorescent analog that serves as a substrate to most known chitinases. However, our findings that $Man_3GlcNAc_2$, a glycan that contains the same $GlcNAc_2$ core present in chitotriose or chitotetraose, is stable indefinitely in the presence of MVL, while cleavage of chitotriose is complete within 16 h at mM concentrations (Supporting Information), is consistent with MVL possessing glycosidase activity at the nonreducing end of *N*-acetyl-chitooligosaccharides. For chitotriose and chitotetraose, the glycosidic bond of rings B and C is the preferred cleavage site.

Catalytic activity in MVL is restricted to one of the two symmetrically related oligomannose binding sites due to the presence of an aspartic acid residue (Asp75) in the catalytic C-terminal domain versus an alanine residue (Ala16) in the equivalent position of the noncatalytic N-terminal domain. Interestingly, bioinformatics of sequenced microbial genomes uncovered three MVL homologues from other cyanobacteria, and sequence conservation at these positions is observed in all three. In unpublished work, we generated a single site Ala16Asp mutant of MVL where we would predict a second catalytic site would be introduced. In an *Escherichia coli* host however this mutant proved to be toxic to the bacterium and we were unable to produce the protein for enzymatic characterization. Dittman et al. recently showed data suggesting another cyanobacterial lectin, microvirin-N (unrelated to MVL), to be involved in cell−cell recognition in the filamentous cyanobacterium *Microcystis aeuriginosa*.[32] In light of Muramoto and co-worker's original finding that expression of MVL occurs only when the producing organism, *Microcystis viridis*, is grown under low

(32) Kehr, J. C.; Zilliges, Y.; Springer, A.; Disney, M. C.; Ratner, D. D.; Bouchier, C.; Seeberger, P. H.; de Marsac, N. T.; Dittman, E. *Mol. Microbiol.* **2006**, *59*, 893.

nutrient or anoxic conditions,[33] and that this lectin has weak enzymatic activity for chitin fragments together with high affinity for oligomannosides, it is conceivable that MVL represents a snapshot of a lectin gaining enzymatic capability, or an enzyme that has optimized additional carbohydrate selectivity. The role these interesting proteins play in the producing cyanobacteria is a subject of current interest in many laboratories. Ongoing advances in chemistry and glycobiology are likely to uncover dual roles for other proteins originally classified as lectins.

## Experimental Methods

**General Methods.** Equilibrium association constants for chitotriose and chitotetraose binding to MVL were determined by isothermal titration calorimetry (ITC). ITC measurements were performed with a Microcal VP-ITC titration calorimeter, and the data were analyzed using Origin software. In each experiment, 1.484 mL of 20 $\mu$M MVL (monomer) was present in the solution cell, and 40 5-$\mu$L aliquots of ligand (1.6 mM) were added via a 250 $\mu$L rotating stirrer-syringe every 150 s at 25 °C. All solutions were prepared in 10 mM Tris buffer, pH 6.5. Control experiments were performed by titrating each carbohydrate to buffer only. No measurable heats were generated from any of the control experiments.

**Preparation of Recombinant MVL Mutants.** Site-directed mutagenesis of wild type MVL plasmid DNA as template together with primers encoding for the desired mutation was performed using QuikChange protocol. Plasmid DNA from two clones per mutant were sequenced and confirmed to have the desired mutation, and one of each was transformed into *E. coli* BL21(DE3) according to the manufacturer's instructions (Novagen). Unlabeled proteins were overexpressed and purified from *E. coli* grown in Luria−Bertani media, and uniformly $^{15}$N- and $^{15}$N,$^{13}$C-labeled proteins from cells grown in M1 minimal media supplemented respectively with $^{15}$N-$NH_4Cl$ and D-glucose, or $^{15}$N-$NH_4Cl$/$^{13}$C-D-glucose as sole nitrogen and carbon sources. Proteins were purified using a 3-step protocol including $NH_4SO_4$ precipitation (25% w/v), gel filtration (Superdex75, GE Healthcare), and anion exchange (MonoQ, GE Healthcare) chromatography as previously published.[17] When necessary, samples were concentrated in VivaSpin 3500 MWCO concentrators (GE Healthcare) equilibrated in the desired buffer according to the manufacturer's instructions.

**Glycan Array Profiling of MVL.** Fluorescently labeled samples of MVL or MVL D75A were prepared by incubating 1 mg purified protein (500 $\mu$L PBS, pH 8.0) with an ~10-fold excess (0.3 mg) Alexafluor 488 succinamidyl ester (Molecular Probes) at rt for 1 h. After transferring to a Slide-a-lyzer (MWCO 3500), each mixture was dialyzed against PBS, pH 7.4, at 4 °C overnight in the dark. Protein concentrations were determined by $A_{280}$ and $A_{494}$ measurements. Both Alexa 488-labeled proteins were tested in single round HIV-1 infectivity assays as previously described[34] to confirm that labeling did not affect carbohydrate binding nor antiviral activity.

Prior to incubation with the arrays, 1% bovine serum albumin and 0.05% Tween 20 were added to each of the protein solutions. Seventy microliters of each of these samples was incubated on the Consortium for Functional Glycomics printed array (v2.0, 200 oligosaccharides; www.functionalglycomics.org) for 1 h at rt. The slides were washed and then dried under a stream of nitrogen and the arrays read with a Perkin-Elmer Microarray XL4000 scanner and analyzed using Imagene (V.6) image analysis software.

**NMR Spectroscopy.** NMR experiments were performed at 298 K on Bruker Avance 500 or 600 spectrometers equipped with cryogenically cooled, *z*-shielded gradient probes. All NMR samples were prepared with 20 mM $NaPO_4$, 50 mM NaCl, and the pH adjusted to 6.8 with 0.25 M NaOD or DCl (NMR buffer). $D_2O$ samples were prepared by dissolving in deuterated $NaPO_4$ buffer carbohydrate, MVL, or stoichiometric complexes of the two that were lyophilized from 99.9% $D_2O$ three times. NMR samples to be used for Saturation Transfer Difference (STD) NMR experiments contained 20 $\mu$M MVL (monomer), determined by UV absorbance at 280 nm ($\varepsilon$ 26 600 $M^{-1}$ $cm^{-1}$), and 1.2 mM carbohydrate (60 fold excess). 1D STD spectra were acquired with selective irradiation at −1 and +40 ppm (on and off resonance, respectively) using a train of 50 ms Gaussian-shaped radio frequency pulses separated by 1 ms delays and an optimized power level of 58 db. Titration experiments using 250 $\mu$L NMR samples of 0.15 mM $^{15}$N-MVL were performed by recording $^1$H−$^{15}$N correlation spectra of samples in the presence of varying amounts of $GlcNAc_3$ and $GlcNAc_4$. Typically, 0.25−0.5 equiv of ligand were added in 5 $\mu$L aliquots. All solutions were prepared in NMR buffer. 1D and 2D NMR experiments were processed with TopSpin 1.3.

**NMR Structure Determination.** $^1$H, $^{13}$C, and $^{15}$N assignments of U-[$^{13}$C,$^{15}$N] MVL D75A in complex with $GlcNAc_3$ (1:2) were made using standard 3D double and triple resonance experiments for assigning backbone and side chain resonances.[35] $^1$H assignments for $GlcNAc_3$ in the same complex were made from 2D $^{12}$C-filtered HOHAHA and $^{12}$C-filtered NOE data. Interproton distance restraints were derived as follows: intramolecular NOEs within the protein were assigned from 3D $^{15}$N-separated and $^{13}$C-separated NOE experiments (125 ms mixing times), intramolecular NOEs within $GlcNAc_3$ in complex with MVL D75A were assigned from a $^{12}$C-filtered NOE experiment, and intermolecular NOEs between protein and $GlcNAc_3$ came from 3D $^{12}$C-filtered/$^{13}$C-separated NOE experiments (150 and 250 ms mixing times). Interproton distance restraints were classified into the distance ranges 1.8−2.9, 1.8−3.5, 1.8−5, and 1.8−6 Å, corresponding to strong, medium, weak and very weak NOEs. Data sets were processed with NMRPipe[36] and analyzed using the program PIPP.[37]

The structure of a 1:2 complex of MVL D75A/$GlcNAc_3$ was calculated using conjoined rigid body/torsion angle dynamics (using Xplor-NIH[38]) starting from the coordinates of the crystal structure of MVL (pdb accession number 1zhs). While the backbone was fixed, side chains in both the N- and C-domains for which unambiguous intermolecular NOEs could be assigned (Figure 5B,C) were allowed to rotate freely during the calculations, and $GlcNAc_3$ was free to rotate and translate relative to the protein and subject to experimental distance and torsion angle restraints. Nonbonded contacts in the complex were represented by a quartic van der Waals repulsion term and a torsion angle database potential to ensure optimal packing.

**Enzymatic Activity of MVL.** Cleavage of oligosaccharides in the presence of MVL was monitored by NMR and HPLC. First, NMR samples containing 2 mM $GlcNAc_3$, $GlcNAc_4$, Man2A or $Man_3GlcNAc_2$ were prepared in NMR buffer and MVL was added to give a final protein concentration of 200 $\mu$M. (It was not possible to use stoichiometric complexes for these studies because of spectral overlap from the protein.) 1D $^1$H NMR spectra (298 K) were recorded for each sample immediately upon addition of protein, and additional spectra recorded at 12, 24, 48, 72, and 96 h. 1D $^1$H NMR spectra of each of the carbohydrates free in solution were also recorded as reference spectra for direct comparison with spectra of cleavage products.

**Acknowledgment.** We thank Dr. Richard Alvarez for glycan microarray profiling, Mr. Ali Shah for assistance with protein expression and purification, and Drs. Maria-Teresa Gutierrez-Lugo, Marius Clore, Son Lam and Y. C. Lee for contributive discussions

(33) Yamaguchi, M.; Ogawa, T.; Muramoto, K.; Kamio, Y.; Jimbo, M.; Kamiya, H. *Biochem. Biophys. Res. Commun.* **1999**, *265*, 703.
(34) Li, M. *J. Virol.* **2005**, *79*, 10108.
(35) Clore, G. M.; Gronenborn, A. M. *Methods Enzymol.* **1994**, *239*, 349.
(36) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
(37) Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1991**, *95*, 214.
(38) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.

during the course of this work. This research was supported in part by the Intramural Research Program, NIDDK, NIH; the AIDS-Targeted Antiviral Program of the Office of the Director, NIH (C.A.B.); an instrument grant from the Division of AIDS Research, NIAID (C.A.B.); and NIGMS - The Consortium for Functional Glycomics, Core H (GM62116).

**Note Added in Proof.** Tanaka *et al.* recently reported potent anti-HIV activity for the bacterial lectin actinohivin.[39]

(39) Tanaka, H.; Chiba, H.; Inokoshi, J.; Kuno, A.; Sugai, T.; Takahashi, A.; Tsunoda, M.; Suzuki, K.; Takénaka, A.; Sekiguchi, T.; Umeyama, H.; Hirabayashi, J.; Ōmura, S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15633.

**Supporting Information Available:** Complete ref 10; detailed HPLC conditions used to monitor cleavage of chitotriose and chitotetraose; $^1$H NMR assignments for $GlcNAc_3$; STD NMR spectra for $GlcNAc_4$, Man2A and $Man_3GlcNAc_2$ in complex with MVL, and Man2A:MVL competed with $GlcNAc_3$; overlays of $^1$H and $^1$H,$^{15}$N-HSQC spectra recorded during cleavage of $GlcNAc_3$; HPLC traces showing cleavage of $GlcNAc_4$; and full glycan mirroarray data for MVL and Asp75Ala MVL. This material is available free of charge via the Internet at http://pubs.acs.org.

JA905929C