Acta Biochimica Polonica

Vol. 53 No. 4/2006, 617–626

on-line at: www.actabp.pl

This review is dedicated to Professor Lech Wojtczak, Chairman of the Editorial Advisory Board of Acta Biochimica Polonica, on the occasion of his 80th birthday

Review

Structural studies of algal lectins with anti-HIV activity

Natasza E. Ziółkowska and Alexander Wlodawer✉

Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD, USA; ✉e-mail: wlodawer@ncifcrf.gov

Received: 10 October, 2006; revised: 18 November, 2006; accepted: 20 November, 2006
available on-line: 27 November, 2006

A number of antiviral lectins, small proteins that bind carbohydrates found on viral envelopes, are currently in pre-clinical trials as potential drugs for prevention of transmission of human immunodeficiency virus (HIV) and other enveloped viruses, such as the Ebola virus and the coronavirus responsible for severe acute respiratory syndrome (SARS). Lectins of algal origin whose antiviral properties make them candidate agents for prevention of viral transmission through topical applications include cyanovirin-N, *Microcystis viridis* lectin, scytovirin, and griffithsin. Although all these proteins exhibit significant antiviral activity, their structures are unrelated and their mode of binding of carbohydrates differs significantly. This review summarizes the current state of knowledge of the structures of algal lectins, their mode of binding of carbohydrates, and their potential medical applications.

Keywords: lectin, HIV, AIDS, topical antivirals, protein structure

INTRODUCTION

Lectins are defined as proteins that bind carbohydrates without initiating their further modifications through associated enzymatic activity (Weis & Drickamer, 1996). They are found in most organisms, ranging from viruses and bacteria to plants and animals (Gabius, 1997; Lis & Sharon, 1998; De Mejia & Prisecaru, 2005). Proteins that are classified as lectins may have totally dissimilar structures, thus the definition is functional rather than structural. Lectins are involved in many biological processes, among them host–pathogen interactions, cell–cell communication, induction of apoptosis, cancer metastasis and differentiation, targeting of cells, as well as recognizing and binding carbohydrates. Some lectins found in algae, such as cyanovirin-N (CV-N) (Boyd *et al*., 1997; Esser *et al*., 1999; Barrientos *et al*., 2003; O'Keefe *et al*., 2003; Helle *et al*., 2006); scytovirin (SVN) (Bokesch *et al*., 2003), *Microcystis viridis* lectin (MVL) (Bewley *et al*., 2004), and griffithsin (GRFT) (Mori *et al*., 2005; Ziółkowska *et al*., 2006) exhibit significant activity against human immunodeficiency virus (HIV) and other enveloped viruses, which makes them particularly promising targets for the development as novel antiviral drugs (De Clercq, 2005; Reeves & Piefer, 2005).

The surfaces of retroviruses such as human immunodeficiency virus (HIV) and many other enveloped viruses are covered by virally-encoded glycoproteins. Glycoproteins gp120 and gp41 present on the HIV envelope are heavily glycosylated, with glycans estimated to contribute almost 50% of the molecular weight of gp120 (Mizuochi *et al*., 1988; Ji *et al*., 2006). Agents that specifically and strongly interact with the glycans may disturb interactions between the proteins of the viral envelope and the cells of the host (Botos & Wlodawer, 2005; Balzarini, 2006). Sugar-binding proteins can crosslink glycans on the viral surface (Sacchettini *et al*., 2001; Shenoy *et al*., 2002) and prevent further interactions with the co-receptors. Unlike the majority of current antiviral therapeutics that act through inhibition of the viral life cycle, lectins can prevent penetration of the host cells by the viruses. Antiviral lectins are best suited to topical applications and can exhibit lower toxicity

Abbreviations: CV-N, cyanovirus N; GRFT, griffithsin; HCV, hepatitis C virus; HIV, human immunodeficiency virus; MVL, *Microcystis viridis* lectin; SVN, scytovirin; UDA, *Urtica dioica.*

than many currently used antiviral therapeutics. Additionally, these proteins are often resistant to high temperatures and low pH, as well as being odorless, which are favorable properties for potential microbicide drugs.

Antiviral activity of a number of lectins that bind high-mannose carbohydrates has been described in the past. Examples of such lectins include jacalin (O'Keefe *et al*., 1997), concanavalin A (Hansen *et al*., 1989), *Urtica diocia* agglutinin (Balzarini *et al*., 1992), *Myrianthus holstii* lectin (Charan *et al*., 2000), *P. tetragonolobus* lectin (Charan *et al*., 2000), and *Narcissus pseudonarcissus* lectin (Balzarini *et al*., 1991). However, lectins derived from marine organisms, a rich source of natural antiviral products (Tziveleka *et al*., 2003), such as CV-N (Boyd *et al*., 1997), SVN (Bokesch *et al*., 2003), MVL (Bewley *et al*., 2004) and GRFT (Mori *et al*., 2005), exhibit the highest activity among the lectins that have been investigated so far (Table 1). Extensive structural studies on antiviral lectins have been undertaken in the last decade, as the knowledge of the three-dimensional structure of putative therapeutic proteins can be very important in their development. Some of the properties of a variety of antiviral lectins have been reviewed recently (Botos & Wlodawer, 2005; Ji *et al*., 2006). This aim of this review is to provide more details of the structural and carbohydrate-binding properties of only a subset of antiviral lectins isolated from various types of algae, since their potency appears to be much higher than the potency of lectins found in other organisms. Interestingly, the three-dimensional structures of these proteins and their modes of carbohydrate binding differ very significantly among them.

CYANOVIRIN-N

CV-N is a lectin originally isolated from blue-green algae *Nostoc ellipsosporum* (Boyd *et al*., 1997). This protein consists of a single chain containing 101 residues and its amino-acid sequence shows obvious duplication. The primary structure of CV-N can be divided into two very similar parts that consist of residues 1–50 and 50–101, respectively (Fig. 1A). The protein is highly resistant to degradation and shows no loss of structural integrity or antiviral activity after treatment with detergents, denaturants, organic solvents, freezing, and heating up to 100ºC (Boyd *et al*., 1997).

Table 1. Anti-HIV activity of several lectins, as summarized by (Ziółkowska *et al*., 2006).

Lectin	EC_{50} (nM)
Jacalin	>227
Myrianthus holstii lectin	150
Urtica diocia agglutinin	105
Concanavalin A	98
N. pseudonarcissus lectin	96
P. tetragonolobus lectin	52
MVL*	30
SVN	0.3
CV-N	0.1
GRFT	0.04

*IC_{50} rather than EC_{50} was reported.

CV-N exhibits a broad range of antiviral activities (reviewed in Barrientos & Gronenborn, 2005). It has been shown that CV-N binds with high affinity to HIV envelope protein gp120 (Boyd *et al*., 1997; Esser *et al*., 1999), blocking the interactions of gp120 with the CD4 coreceptor (Dey *et al*., 2000). CV-N also interacts with another surface glycoprotein, gp41 (O'Keefe *et al*., 2000; O'Keefe, 2001). The specificity of CV-N, however, is not limited only to different strains of HIV and related retroviruses. CV-N inhibits the development of viral cytopathic effects of Ebola virus, binding to its surface envelope glycoprotein (Barrientos *et al*., 2003), blocks influenza infection by binding to the hemagglutinin surface glycoprotein (O'Keefe *et al*., 2003), and also inhibits fusion and/or infection by human herpesvirus 6 and measles virus (Dey *et al*., 2000; Barrientos *et al*., 2003). That lectin has also been shown to exhibit significant activity against hepatitis C virus (HCV) (Helle *et al*., 2006).

The properties of recombinant CV-N produced in *Escherichia coli* have been shown to be virtually identical to those of the native protein isolated from cyanobacterium (Mori *et al*., 1998). The availability of large amounts of recombinant lectin facilitated extended biochemical, antiviral, and structural studies. The structure of CV-N was initially solved by NMR (PDB codes 2ezn, 2ezm) (Bewley *et al*., 1998), indicating that the protein exists in solution in the form of a compact monomer. The molecule is elongated, and, as predicted by analysis of the primary structure (Fig. 1A), consists of two similar domains. Each domain contains five β-strands and two helical turns, with the two halves of the molecule exhibiting considerable similarity (Fig. 2A). However, the two domains of CV-N do not simply follow the duplication of the amino-acid sequence, but are formed after domain swapping (Bennett *et al*., 1995). Domain A is formed by residues 1–39 and 90–101, whereas domain B consists of residues 39–90. Each domain consists of a triple-stranded antiparallel β-sheet and a single β-hairpin. The two domains are connected by helical turns.

Although it was demonstrated that the monomer is the predominant form of CV-N in solution, all reported crystal structures contained domain swapped dimers of the protein (Yang *et al*., 1999). Each chain of CV-N found in the dimer (PDB code

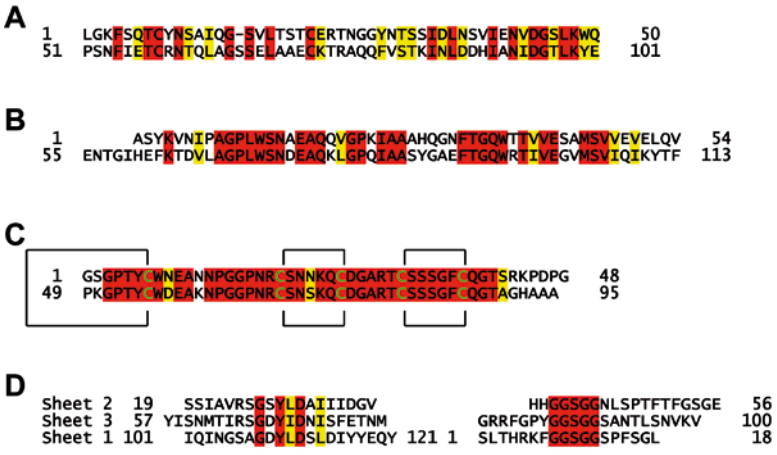

Figure 1. Sequences of the algal lectins discussed in this review, emphasizing the presence of repeats.
Residues that are strictly conserved are shown on red background, and those of a similar character are highlighted in yellow. Cysteines creating disulfide bonds in SVN are marked in green and the disulfide bonds are shown as solid lines. A) CV-N; B) MVL; C) SVN; D) GRFT.

3ezm) is quite extended and cannot exist in solution by itself in that conformation (Fig. 2B). The presence of a flexible linker consisting of residues 49–54 is crucial for enabling the monomer-dimer transition. The two domains move around the linker forming either a monomer observed by NMR (domain A with domain B) or the domain-swapped dimer seen in the crystals. In the latter, domain A from one monomer interacts with domain B′ from the other monomer, and domain B interacts with domain A′; the resulting AB′ and A′B constructs resemble the monomers seen in the NMR structures, but are formed by two separate chains. Transition between the monomeric and dimeric form is accomplished by a change of only two main-chain torsion angles in the linker region (Botos *et al*., 2002a).

CV-N was crystallized at both low (PDB code 3ezm) and high pH (PDB code 1l5b) in different space groups, with the domain-swapped dimer (Fig. 2C) observed under both conditions (Yang *et al*., 1999; Barrientos *et al*., 2002). The only difference between the molecules present in these two crystal forms was in the relative orientation of the two domains. It was shown that in solution the dimer is a metastable form of the protein, although it is stable for months at neutral pH and room temperature (Barrientos *et al*., 2002). However, all attempts to crystallize monomeric CV-N, characterized as such

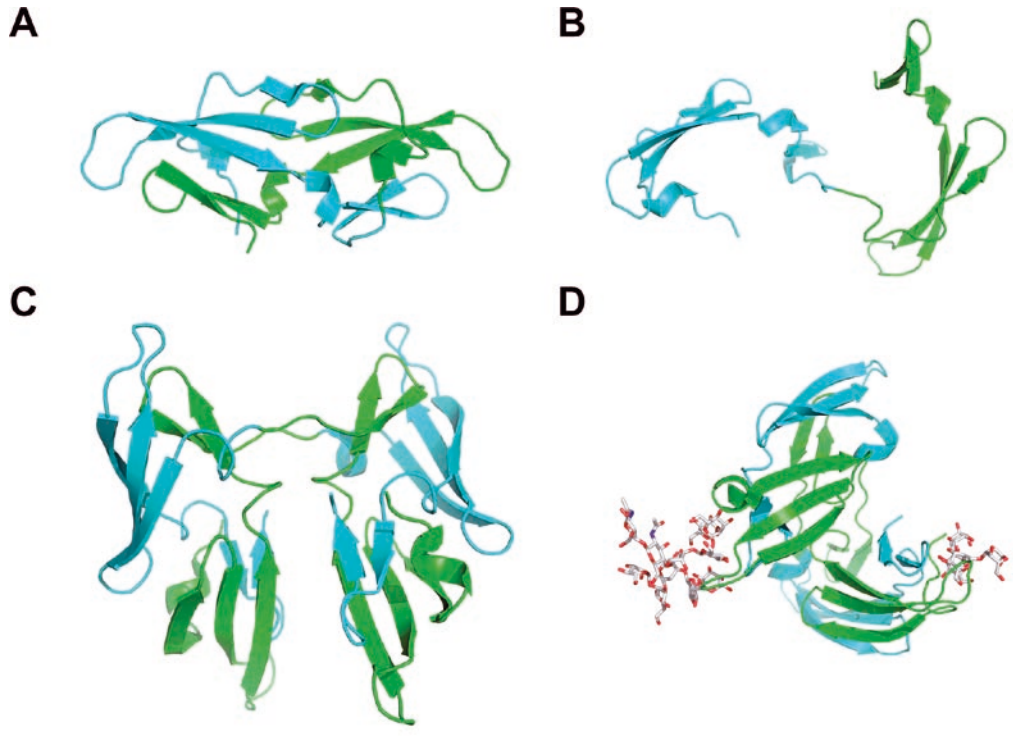

Figure 2. Different structural forms reported for CV-N.
Domain A of each monomer is colored cyan and domain B green. A) Compact monomer observed so far only in the NMR structures; B) Extended monomer that cannot exist by itself, but was seen in all crystal structures and in some NMR structures as part of a dimer; C) Domain-swapped dimer made of two monomers shown in panel B; D) A complex of CV-N with Man9.

by NMR, resulted in crystals containing the dimeric form of the protein.

Residues Pro51 and Ser52 present in the linker region were shown to be important for determining the oligomeric properties of CV-N (Barrientos *et al.*, 2002; Han *et al.*, 2002). A double mutant P51S/S52P was investigated in an attempt to explain the reasons for changes in protein oligomerization state upon crystallization. The mutant exhibits indistinguishable gp120-binding activity compared to the wild type CV-N. The crystal structure of the mutant (PDB code 1lom) demonstrated the difference in orientation of the linker region in comparison to the wild-type domain-swapped dimer (Botos *et al.*, 2002a). Crystals of the mutant could be grown at pH ranging from 4.6 to 10.3, proving that domain swapping is possible under a very wide range of pH values.

A number of structures of the complexes of CV-N with different carbohydrates have been investigated by both NMR and crystallography. The ligands included dimannose Manα1-2Manα (PDB code 1iiy) (Bewley, 2001) and oligosaccharides such as hexamannoside (PDB code 1m5j) or $Man_9GlcNAc_2$ (PDB code 1m5m) (Botos *et al.*, 2002b; Shenoy *et al.*, 2002). These structural studies demonstrated that a CV-N monomer could bind oligosaccharides with different affinity in two sites, described as primary and secondary. The high-affinity primary binding site is located near the linker region of the protein, and thus its shape is influenced by the relative orientation of the domains, as well as by the oligomeric structure of CV-N. The secondary binding site is far from the linker region and has the same conformation in the monomeric and dimeric forms of CV-N. A dimeric form of CV-N shows the presence of four sugar-binding sites. Crystallographic studies of CV-N elucidated structural reasons for the higher affinity for $Man_9GlcNAc_2$ and much lower affinity for the hexamannoside, since the lectin is able to bind directly three mannose rings of $Man_9GlcNAc_2$, but only two mannose rings of the hexamannoside (Botos *et al.*, 2002b), with the remaining parts of the carbohydrates being disordered. The structure of CV-N complexed with $Man_9GlcNAc_2$ is shown in Fig. 2D.

MICROCYSTIS VIRIDIS LECTIN

MVL is 13 kDa protein isolated from cyanobacterium *Microcystis viridis* (Yamaguchi *et al.*, 1999). A molecule of MVL (PDB code 1zhq) consists of a single polypeptide that contains two highly homologous domains, each 54 amino acid long, with about 50% identity (Fig. 1B). The two domains are joined by a linker consisting of 5 amino-acid residues.

MVL was originally shown to bind yeast mannan (Yamaguchi *et al.*, 1999) and its detailed carbohydrate specificity was later determined by NMR titration studies (Bewley *et al.*, 2004). The protein binds oligomannosides such as $Man_6GlcNAc_2$ with sub-micromolar affinity. A number of mannose-containing carbohydrates were tested and it was shown that MVL does not bind α- and β-linked dimannosides, disaccharides Manβ(1→4)GlcNAc and GlcNAcβ(1→4)GlcNAc, or mannotriose. High-affinity carbohydrate binding requires the presence of both the mannose and glucosamine residues and at least tetrasaccharide core structures, such as Man α(→6)Manβ(1→4)GlcNAcβ(1→4)GlcNAc, have to be present. MVL inhibits HIV-1 cell fusion with an IC_{50} about 30 nM (Bewley *et al.*, 2004).

As verified by NMR data, MVL is a monodisperse homodimer in solution (Bewley *et al.*, 2004). The crystal structure of MVL also shows that the protein forms a stable, symmetric homodimer (Williams, Jr. *et al.*, 2005). As expected on the basis of the observed sequence duplication (Fig. 1B), each monomer is formed by two similar domains. Each domain contains a three-stranded antiparallel β-sheet and a single α-helix, the latter located between strands β1

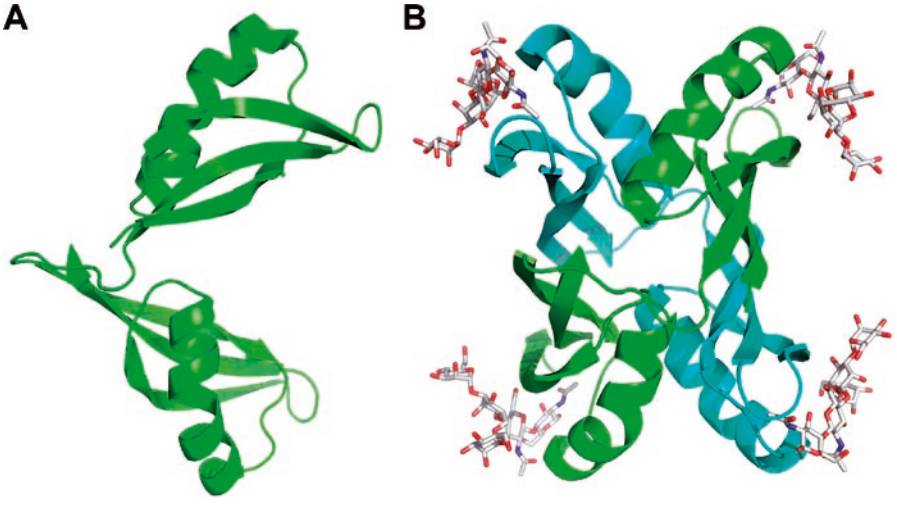

Figure 3. The structure of MVL. Molecule A is colored cyan and molecule B green. A) Boomerang-shaped monomer of MVL; B) The crystal structure of the complex of MVL with $Man_3GlcNAc_2$.

and β2. The interactions between the two domains of the monomer are very limited. Boomerang-shaped monomers (Fig. 3A) interlock and each domain from one monomer contacts both domains from the second monomer.

NMR studies demonstrated that MVL contains two sugar-binding sites per monomer (Bewley *et al*., 2004); this observation was subsequently confirmed by X-ray analysis (Williams, Jr. *et al*., 2005). A crystal structure of the complex of MVL with $Man_3GlcNAc_2$ (PDB code 1zhs) shows four independent carbohydrate binding sites per homodimer, two each within a single polypeptide chain (Williams, Jr. *et al*., 2005) (Fig. 3B). Binding of sugar molecules does not induce any significant conformational changes to MVL. A trisaccharide core interacts with each binding site. The deepest part of the pocket is occupied by the reducing GlcNAc residue of $Man_3GlcNAc_2$ and this sugar unit is essential for defining the specificity of carbohydrate binding.

SCYTOVIRIN

SVN is a lectin isolated from cyanobacterium *Scytonema varium;* surprisingly, a search for expression of similar proteins in other *Scytonema* species using monoclonal antibodies did not yield any positive results (A.W., unpublished). A single chain of SVN contains 95 amino acids; ten of them, which are cysteines, form five intrachain disulfide bonds. Their pattern, elucidated by mass spectrometry of fragments obtained by trypsin digests, was shown to be C7–C55, C20–C26, C32–C38, C68–C74, and C80–C86 (Bokesch *et al*., 2003). Similarly to the two other lectins discussed above, SVN also demonstrates internal sequence duplication, suggesting the presence of two functional domains linked by the C7–C55 disulfide bond (Fig. 1C). The extent of identity of the sequences of the N-terminal part of the molecule (residues 1–48) and the C-terminal part (residues 49–95) is very high (75%).

SVN binds to glycosylated gp160, gp120, and gp41 and interacts with oligosaccharides, specifically α1-2, α1-2, α1-6 linked tetrasaccharide units, but with no reported binding to α1-2, α1-2 linked trisaccharides (Adams *et al*., 2003). Although it does not show significant specificity for mannose or *N*-acetylglucosamine, its binding to gp120 can be inhibited by $Man_8GlcNAc_2$ or $Man_9GlcNAc_2$. The protein displays nanomolar activity against T-tropic strains and primary isolates of HIV-1, appearing to be a good inhibitor of HIV binding and/or fusion (Bokesch *et al*., 2003).

The primary structure of SVN exhibits 55% similarity to the chitin-binding domain of *Volvox carteri* lectin and a slightly lower level of similarity to the sequence of lectin from *Urtica dioica* (UDA) (Harata & Muraki, 2000). However, SVN does not appear to belong to the chitin-binding class of proteins, since it does not bind chitin and has a different disulfide bond pattern (Bokesch *et al*., 2003). Efficient recombinant production of SVN was reported recently (Xiong *et al*., 2006), facilitating further biochemical and structural studies of that protein. A synthetic gene encoding SVN was constructed and expressed in *E. coli*. The recombinant protein was found to have correct disulfide-bonding pattern and exhibit both gp160-binding activity and anti-HIV activity. Although well-diffracting crystals of SVN have been obtained (I. Botos and A.W., unpublished), crystal structure of that protein has not yet been solved. However, NMR studies of SVN have been more successful and its structure should become available in the near future (R. A. Byrd and B. R. O'Keefe, personal communication).

GRIFFITHSIN

GRFT was isolated from the red alga *Griffithsia* sp. collected from the waters off New Zealand. GRFT was shown to display picomolar activity against HIV-1 (Mori *et al*., 2005), moderately interfering with the binding of gp120 to sCD4. The binding of GRFT to soluble gp120 was inhibited by glucose, mannose, and *N*-acetylglucosamine (Mori *et al*., 2005). In addition to inhibiting HIV-1, GRFT was shown to inhibit replication and cytopathy of the coronavirus that causes SARS (Ziółkowska *et al*., 2006).

The gene encoding GRFT has not been isolated, but the amino-acid sequence was obtained directly from protein purified from cyanobacteria. A GRFT molecule consists of a single 121-amino-acid chain (Fig. 1D); residue 31 does not appear to correspond to any standard amino acid and its identity is still unknown (Mori *et al*., 2005). Analysis of the sequence of GRFT (Fig. 1D) shows limited homology (less than 30% identity) to proteins such as jacalin (Aucouturier *et al*., 1987), heltuba (Bourne *et al*., 1999) or artocarpin (Jeyaprakash *et al*., 2004), all members of the β-prism-I family of lectins (Raval *et al*., 2004; Chandra, 2006).

GRFT used for biological and structural studies was prepared as recombinant protein in either *E. coli* (Giomarelli *et al*., 2006) or *Nicothiana benthamiana* (Ziółkowska *et al*., 2006). In both constructs residue 31 of GRFT was replaced by an alanine; this substitution did not seem to affect the carbohydrate-binding properties of the lectin. GRFT expressed in *E. coli* contained an N-terminal 6-His affinity tag followed by a putative thrombin cleavage site, extending the protein sequence by 17 amino acids (Mori *et*

al., 2005; Giomarelli *et al*., 2006); the additional sequence could not be removed and was present in the crystallized protein. The plant-expressed construct did not include any tags, thus resembling more closely the authentic protein, although with an acetylated N terminus and mutated residue 31 (Ziółkowska *et al*., 2006). Although both the His-tagged and the plant-expressed GRFT crystallized easily, crystals grown from the plant-produced material diffracted significantly better, most likely due to the absence of the extension of the polypeptide chain. Crystals of the His-tagged griffithsin contained only a single molecule in the asymmetric unit (PDB code 2gux) whereas all crystal forms grown from the plant-expressed material contained two molecules (PDB codes 2gty, 2gue, 2guc, 2gud, 2hyr, 2hyq) (Ziółkowska *et al*., 2006; 2007).

The fold of GRFT corresponds to the β-prism-I motif (Chothia & Murzin, 1993), observed in a variety of lectins, as well as in some other proteins (Shimizu & Morikawa, 1996). The motif consists of three repeats of an anti-parallel four-stranded β-sheet that form a triangular prism (Fig. 4A). Unlike other members of the family, GRFT forms a domain swapped dimer in which the first two β-strands of one chain are associated with ten strands of the other chain and *vice versa* (Fig. 4B) (Ziółkowska *et al*., 2006).

The crystal structures of GRFT complexes with different monosaccharides and disaccharides, such as mannose (PDB codes 2guc and 2gud), *N*-acetylglucosamine (PDB code 2gue), 1→6α–mannobiose (PDB code 2hyq), and maltose (PDB code 2hyr), were solved and refined at high resolution (up to the atomic resolution of 0.94 Å for the complex with mannose) (Ziółkowska *et al*., 2006; 2007). Very high quality of the diffraction data made it possible to map protein–carbohydrate interactions in considerable detail. In the atomic-resolution structure of the complex with mannose even the presence of both α and β anomers of the sugar bound to the protein was detected (Ziółkowska *et al*., 2006).

Unlike other proteins that belong to the same fold family, a single molecule of GRFT contains three almost identical carbohydrate-binding sites, each capable of binding a monosaccharide through multiple contact points. The six principal sites in the obligatory dimer of GRFT are very similar and are arranged on every monomer in groups of three (Fig. 4C). The carbohydrate-binding sites are formed from the parts of the structure that exhibit extensive sequence conservation, but some of the main chain atoms are involved in specific, but sequence-independent contacts with the carbohydrate molecules; these contacts are very similar in all three sites. GRFT contains three strictly conserved repeats of a sequence GGSGG (Fig. 1D), located in loops that connect the first and fourth strand of each β-sheet. The main chain amide of the last residue of each

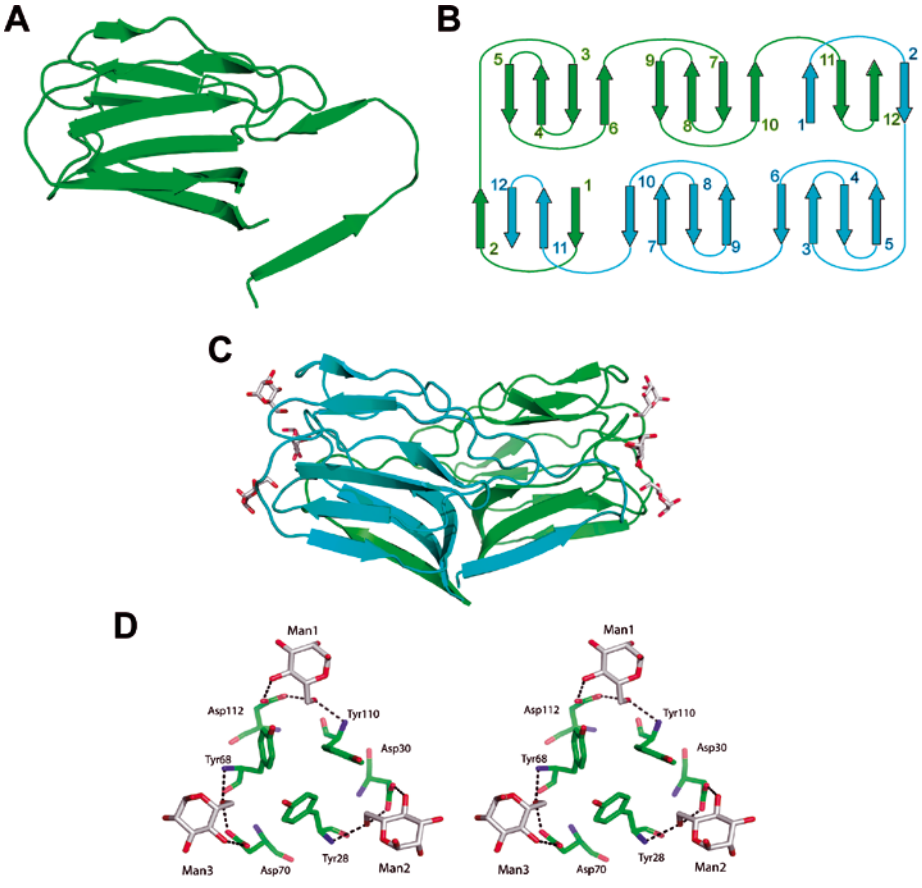

Figure 4. The structure of GRFT.
A) Monomer of GRFT. B) Topology of the domain swapped dimer in which the first two β-strands of one chain are associated with ten strands of the other chain (and *vice versa*). C) A complex of GRFT with mannose. The six principal carbohydrate binding sites are very similar and are arranged in two groups of three on each monomer; only one monomer is shown here. D) Mannose-binding sites 1–3, created principally by molecule A of GRFT. Hydrogen bonds between protein and carbohydrates are marked by dashed lines.

of these sequences participates in creation of a ligand-binding site and the strict conservation of this sequence may be the most important reason for the presence of three monosaccharide binding sites on each molecule of GRFT.

With one known exception, each molecule of the other lectins that are structurally closely related to GRFT contains only a single carbohydrate-binding site. Thus the presence of binding site 1 was reported for all β-prism-I lectins, binding site 2 has only been seen in banana lectin (Meagher *et al*., 2005), whereas binding site 3 is unique to GRFT. Three sugar-binding sites of GRFT form an almost perfect equilateral triangle on the edge of the protein, with the carbohydrate molecules found about 15 Å from each other (Fig. 4D). Very similar interactions are also present in the complexes of GRFT with disaccharides, where the additional sugar units make between zero and two hydrogen bonds with the protein (Ziółkowska *et al*., 2007).

The reported biological activity of GRFT against HIV is >1000-fold higher than the activities reported for several monosaccharide-specific lectins (Table 1**)** (Charan *et al*., 2000; Ziółkowska *et al*., 2006). Since GRFT offers six separate binding sites for mannose in a dimer, the binding potential for the high-mannose oligosaccharides found on the HIV gp120 is significant. Based on crystal structures of GRFT complexed with monosaccharides and disaccharides, it was possible to create a model of a complex of griffithsin with $Man_9GlcNAc_2$ (PDB code 2i43), a branched carbohydrate commonly found on the surface of viral glycoproteins (Fig. 5A) (Ziółkowska *et al*., 2006). The carbohydrate could be docked to the protein quite easily, in a conformation resembling its experimentally-determined structures seen in a complex with a specific antibody (Calarese *et al*., 2003). In the model of the GRFT complex, the core of the carbohydrate, including its reducing end, points away from the protein surface (Fig. 5B). This mode of binding can be easily reconciled with the binding expected for similar N-linked oligosaccharides bound to the viral proteins through their reducing ends. The model suggests that GRFT utilizes multivalent binding modes in order to bind carbohydrates very tightly.

CONCLUSIONS

Although a number of mannose-binding lectins have been shown to interact with the envelope protein of HIV and exhibit antiviral activity (Botos & Wlodawer, 2005), lectins isolated from algae are more potent by several orders of magnitude, and only they were discussed in this re-

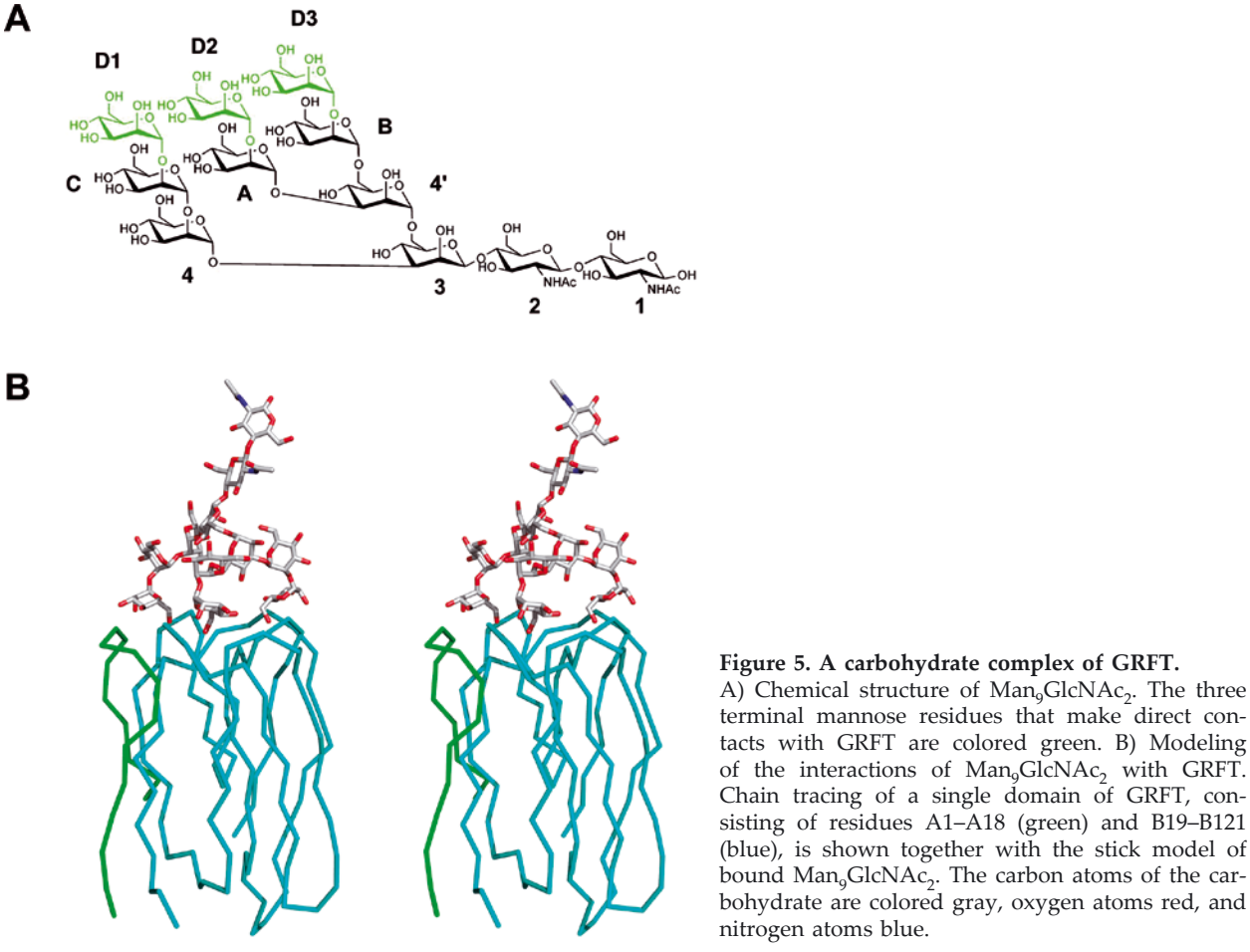

Figure 5. A carbohydrate complex of GRFT.
A) Chemical structure of $Man_9GlcNAc_2$. The three terminal mannose residues that make direct contacts with GRFT are colored green. B) Modeling of the interactions of $Man_9GlcNAc_2$ with GRFT. Chain tracing of a single domain of GRFT, consisting of residues A1–A18 (green) and B19–B121 (blue), is shown together with the stick model of bound $Man_9GlcNAc_2$. The carbon atoms of the carbohydrate are colored gray, oxygen atoms red, and nitrogen atoms blue.

view. However, one should keep in mind that the definition of algae is rather broad. Whereas CV-N, MVL, and SVN were obtained from prokaryotic cyanobacteria, GRFT was isolated from *Griffithsia* sp., belonging to eukaryotic Rhodophyta. The common feature of all four lectins is an internal multiplication of the amino-acid sequence. The lectins isolated from cyanobacteria show duplication in the amino-acid sequence, whereas the sequence of GRFT contains a triple repeat. The duplicated sequences of cyanobacteria-derived lectins lead to folds that include two highly similar domains, observed in the structures of CV-N and MVL, and postulated to be present in SVN as well. On the other hand, the structure of GRFT is described as a prism consisting of three blades, each corresponding to a β-sheet (Fig. 1D). All three lectins with the known structures exhibit domain swapping, which complicates their topological description. The level of structure conservation between the domains of cyanobacterial origin is much higher than in GRFT. The similarity among the three blades of the latter lectin is primarily limited to the amino acids which are engaged in carbohydrate binding. Among the lectins described here, SVN demonstrates the highest identity of the duplicated sequences, and also reveals a very extensive disulfide bond pattern which is not present in the other lectins.

The structures of lectins that have been discussed here consist predominantly of β-strands and sheets, although MVL also contains a long α-helix. Unusual domain swapping is present in the structures of both CV-N and GRFT, but the nature of the phenomenon differs. CV-N dimers swap approximately half of the molecule and additional swapping is present in each monomer as well; in GRFT, only two β-strands out of twelve are swapped. CV-N could be isolated in both monomeric and dimeric form, whereas MVL and GRFT have been observed so far only in their dimeric form. SVN appears to be always monomeric, although the repeated structural motif suggests the presence of two similar domains in each molecule.

The antiviral activity of lectins appears to depend on their ability to bind mannose-containing oligosaccharides present on the surface of viral envelope glycoproteins. GRFT displays picomolar activity against HIV-1, thus being much more potent than CV-N, SVN, and MVL, which exhibit only nanomolar activity (Table 1). Four carbohydrate-binding sites are present in each dimer of CV-N and MVL, whereas the dimer of GRFT provides six potential binding sites. The mode of recognition of the carbohydrates also varies very significantly among these lectins. CV-N recognizes larger oligosaccharide structures of the terminal branches, MVL is specific for four-five units of the oligomannoside core, whereas SVN binds the D3 arm of $Man_9GlcNAc_2$. As postulated by molecular modeling, the unusual geometry of the binding sites of GRFT allows tridentate binding to $Man_9GlcNAc_2$, possibly explaining its much higher antiviral activity in comparison to the activity of the other lectins. It is clear that it is the binding of carbohydrates that is responsible for the antiviral properties of these lectins, but such binding can be accomplished in a variety of ways that will be reflected in the observed antiviral potency.

The observed antiviral activity of CV-N, MVL, SVN and GRFT makes these lectins promising candidates for development as microbicides (Reeves & Piefer, 2005), especially as components of antiviral preparations that could be applied topically with the aim of preventing infection (Turpin, 2002). Preliminary studies on CV-N have been very promising, indicating that this lectin was highly efficient in preventing infection in macaques by chimeric SIV/HIV-1 viruses, delivered in these model systems of HIV-1 infection by either vaginal or rectal routes (Tsai *et al*., 2003; Tsai *et al*., 2004). SVN and GRFT are also in pre-clinical investigations, although the results have not yet been published (B. R. O'Keefe, personal communication). A number of modes of delivery of lectin-based drugs are being contemplated. The most straightforward is to include them as components of microbicides for vaginal or rectal application, or possibly as components of multifunctional contraceptive gels. In particular, the need for development of female-controlled microbicides as protection against infection by HIV is now generally accepted, with considerable interest in the subject expressed recently by the Gates Foundation, among others. Another method of delivery of lectins, potentially even more promising, is their *in situ* expression by modified bacteria similar to those naturally found in vagina. This concept was already proven by studies of the delivery of CV-N in biologically active form using human commensal bacterium *Streptococcus gordonii* (Giomarelli *et al*., 2002). CV-N expressed as both the soluble protein or immobilized on the surface of the bacterium was shown to be an efficient agent for capturing and immobilizing HIV-1, thus preventing infection. Other lectins reviewed here will most likely also be tested in similar applications, thus contributing to the development of completely novel ways of preventing the spread of the disease that has already killed millions of people in the last quarter century.

Acknowledgements

We would like to thank Drs. M. Boyd, I. Botos, and B. R. O'Keefe for their collaboration in structural studies of antiviral lectins and for com-

municating unpublished results. We are grateful to Dr. O'Keefe for critical reading of this manuscript. This project was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

REFERENCES

Adams EW, Ueberfeld J, Ratner DM, O'Keefe BR, Walt DR, Seeberger PH (2003) Encoded fiber-optic microsphere arrays for probing protein-carbohydrate interactions. *Angew Chem Int Ed Engl* **42**: 5317–5320.

Aucouturier P, Mihaesco E, Mihaesco C, Preud'Homme JL (1987) Characterization of jacalin, the human IgA and IgD binding lectin from jackfruit. *Mol Immunol* **24**: 503–511.

Balzarini J (2006) Inhibition of HIV entry by carbohydrate-binding proteins. *Antiviral Res* **71**: 237–247.

Balzarini J, Schols D, Neyts J, Van Damme E, Peumans W, De Clercq E (1991) Alpha-(1-3)- and alpha-(1-6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections *in vitro*. *Antimicrob Agents Chemother* **35**: 410–416.

Balzarini J, Neyts J, Schols D, Hosoya M, Van Damme E, Peumans W, De Clercq E (1992) The mannose-specific plant lectins from *Cymbidium* hybrid and *Epipactis* helleborine and the (N-acetylglucosamine)n-specific plant lectin from *Urtica dioica* are potent and selective inhibitors of human immunodeficiency virus and cytomegalovirus replication *in vitro*. *Antiviral Res* **18**: 191–207.

Barrientos L, Louis JM, Botos I, Mori T, Han Z, O'Keefe BR, Boyd MR, Wlodawer A, Gronenborn A (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state. Reconcilliation of X-ray and NMR structures. *Structure* **10**: 673–686.

Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR (2003) Cyanovirin-N binds to the viral surface glycoprotein, $GP_{1,2}$ and inhibits infectivity of Ebola virus. *Antiviral Res* **58**: 47–56.

Barrientos LG, Gronenborn AM (2005) The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini Rev Med Chem* **5**: 21–31.

Bennett MJ, Schlunegger MP, Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* **4**: 2455–2468.

Bewley CA, Gustafson KR, Boyd MR, Covell DG, Bax A, Clore GM, Gronenborn AM (1998) Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nature Struct Biol* **5**: 571–578.

Bewley CA (2001) Solution structure of a cyanovirin-N: Manα 1-2Manα complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* **9**: 931–940.

Bewley CA, Cai M, Ray S, Ghirlando R, Yamaguchi M, Muramoto K (2004) New carbohydrate specificity and HIV-1 fusion blocking activity of the cyanobacterial protein MVL: NMR, ITC and sedimentation equilibrium studies. *J Mol Biol* **339**: 901–914.

Bokesch HR, O'Keefe BR, McKee TC, Pannell LK, Patterson GM, Gardella RS, Sowder RC, Turpin J, Watson K, Buckheit RW, Jr., Boyd MR (2003) A potent novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. *Biochemistry* **42**: 2578–2584.

Botos I, Wlodawer A (2005) Proteins that bind high-mannose sugars of the HIV envelope. *Prog Biophys Mol Biol* **88**: 233–282.

Botos I, Mori T, Cartner LK, Boyd MR, Wlodawer A (2002a) Domain-swapped structure of a mutant of cyanovirin-N. *Biochem Biophys Res Commun* **294**: 184–190.

Botos I, O'Keefe BR, Shenoy SR, Cartner LK, Ratner DM, Seeberger PH, Boyd MR, Wlodawer A (2002b) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* **277**: 34336–42.

Bourne Y, Zamboni V, Barre A, Peumans WJ, Van Damme EJ, Rouge P (1999) *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. *Structure Fold Des* **7**: 1473–1482.

Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, Gulakowski RJ, Wu L, Rivera MI, Laurencot CM, Currens MJ, Cardellina JH, Buckheit RW, Jr., Nara PL, Pannell LK, Sowder RC, Henderson LE (1997) Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* **41**: 1521–1530.

Calarese DA, Scanlan CN, Zwick MB, Deechongkit S, Mimura Y, Kunert R, Zhu P, Wormald MR, Stanfield RL, Roux KH, Kelly JW, Rudd PM, Dwek RA, Katinger H, Burton DR, Wilson IA (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**: 2065–2071.

Chandra N (2006) Common scaffolds, diverse recognition profiles. *Structure* **14**: 1093–1094.

Charan RD, Munro MH, O'Keefe BR, Sowder RCII, McKee TC, Currens MJ, Pannell LK, Boyd MR (2000) Isolation and characterization of *Myrianthus holstii* lectin, a potent HIV-1 inhibitory protein from the plant *Myrianthus holstii*. *J Nat Prod* **63**: 1170–1174.

Chothia C, Murzin AG (1993) New folds for all-β proteins. *Structure* **1**: 217–22.

De Clercq E (2005) Emerging anti-HIV drugs. *Expert Opin Emerg Drugs* **10**: 241–273.

De Mejia EG, Prisecaru VI (2005) Lectins as bioactive plant proteins: a potential in cancer treatment. *Crit Rev Food Sci Nutr* **45**: 425–445.

Dey B, Lerner DL, Lusso P, Boyd MR, Elder JH, Berger EA (2000) Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *J Virol* **74**: 4562–4569.

Esser MT, Mori T, Mondor I, Sattentau QJ, Dey B, Berger EA, Boyd MR, Lifson JD (1999) Cyanovirin-N binds to gp120 to interfere with CD4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does not affect the CD4 binding site on gp120 or soluble CD4-induced conformational changes in gp120. *J Virol* **73**: 4360–4371.

Gabius HJ (1997) Animal lectins. *Eur J Biochem* **243**: 543–576.

Giomarelli B, Provvedi R, Meacci F, Maggi T, Medaglini D, Mori T, McMahon JB, Gardella R, Boyd MR (2002) The microbicide cyanovirin-N expressed in the surface of commensal bacterium *Streptococcus gordonii* captures HIV-1. *AIDS* **16**: 1351–1360.

Giomarelli B, Schumacher KM, Taylor TE, Sowder RC, Hartley JL, McMahon JB, Mori T (2006) Recombinant production of anti-HIV protein, griffithsin, by auto-induction in a fermentor culture. *Protein Expr Purif* **47**: 194–202.

Han Z, Xiong C, Mori T, Boyd MR (2002) Discovery of a stable dimeric mutant of cyanovirin-N (CV-N) from a T7 phage-displayed CV-N mutant library. *Biochem Biophys Res Commun* **292**: 1036–1043.

Hansen JE, Nielsen CM, Nielsen C, Heegaard P, Mathiesen LR, Nielsen JO (1989) Correlation between carbohydrate structures on the envelope glycoprotein gp120 of HIV-1 and HIV-2 and syncytium inhibition with lectins. *AIDS* **3**: 635–641.

Harata K, Muraki M (2000) Crystal structures of *Urtica dioica* agglutinin and its complex with tri-N-acetylchitotriose. *J Mol Biol* **297**: 673–681.

Helle F, Wychowski C, Vu-Dac N, Gustafson KR, Voisset C, Dubuisson J (2006) Cyanovirin-N inhibits hepatitis C virus entry by binding to envelope protein glycans. *J Biol Chem* **281**: 25177–25183.

Jeyaprakash AA, Srivastav A, Surolia A, Vijayan M (2004) Structural basis for the carbohydrate specificities of artocarpin: variation in the length of a loop as a strategy for generating ligand specificity. *J Mol Biol* **338**: 757–770.

Ji X, Chen Y, Faro J, Gewurz H, Bremer J, Spear GT (2006) Interaction of human immunodeficiency virus (HIV) glycans with lectins of the human immune system. *Curr Protein Pept Sci* **7**: 317–234.

Lis H, Sharon N (1998) Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* **98**: 637–674.

Meagher JL, Winter HC, Ezell P, Goldstein IJ, Stuckey JA (2005) Crystal structure of banana lectin reveals a novel second sugar binding site. *Glycobiology* **15**: 1033–1042.

Mizuochi T, Spellman MW, Larkin M, Solomon J, Basa LJ, Feizi T (1988) Carbohydrate structures of the human-immunodeficiency-virus (HIV) recombinant envelope glycoprotein gp120 produced in Chinese-hamster ovary cells. *Biochem J* **254**: 599–603.

Mori T, Gustafson KR, Pannell LK, Shoemaker RH, Wu L, McMahon JB, Boyd MR (1998) Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein Expr Purif* **12**: 151–158.

Mori T, O'Keefe BR, Sowder RC, Bringans S, Gardella R, Berg S, Cochran P, Turpin JA, Buckheit RW, Jr., McMahon JB, Boyd MR (2005) Isolation and characterization of griffithsin, a novel HIV-inactivating protein, from the red alga *Griffithsia* sp. *J Biol Chem* **280**: 9345–9353.

O'Keefe BR (2001) Biologically active proteins from natural product extracts. *J Nat Prod* **64**: 1373–1381.

O'Keefe BR, Beutler JA, Cardellina JH, Gulakowski RJ, Krepps BL, McMahon JB, Sowder RC, Henderson LE, Pannell LK, Pomponi SA, Boyd MR (1997) Isolation and characterization of niphatevirin, a human-immunodeficiency-virus-inhibitory glycoprotein from the marine sponge *Niphates erecta*. *Eur J Biochem* **245**: 47–53.

O'Keefe BR, Shenoy SR, Xie D, Zhang W, Muschik JM, Currens MJ, Chaiken I, Boyd MR (2000) Analysis of the interaction between the HIV-inactivating protein cyanovirin-N and soluble forms of the envelope glycoproteins gp120 and gp41. *Mol Pharmacol* **58**: 982–992.

O'Keefe BR, Smee DF, Turpin JA, Saucedo CJ, Gustafson KR, Mori T, Blakeslee D, Buckheit R, Boyd MR (2003) Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* **47**: 2518–2525.

Raval S, Gowda SB, Singh DD, Chandra NR (2004) A database analysis of jacalin-like lectins: sequence-structure-function relationships. *Glycobiology* **14**: 1247–1263.

Reeves JD, Piefer AJ (2005) Emerging drug targets for antiretroviral therapy. *Drugs* **65**: 1747–1766.

Sacchettini JC, Baum LG, Brewer CF (2001) Multivalent protein-carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction. *Biochemistry* **40**: 3009–3015.

Shenoy SR, Barrientos LG, Ratner DM, O'Keefe BR, Seeberger PH, Gronenborn AM, Boyd MR (2002) Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. *Chem Biol* **9**: 1109–1118.

Shimizu T, Morikawa K (1996) The β-prism: a new folding motif. *Trends Biochem Sci* **21**: 3–6.

Tsai CC, Emau P, Jiang Y, Tian B, Morton WR, Gustafson KR, Boyd MR (2003) Cyanovirin-N gel as a topical microbicide prevents rectal transmission of SHIV89.6P in macaques. *AIDS Res Hum Retroviruses* **19**: 535–541.

Tsai CC, Emau P, Jiang Y, Agy MB, Shattock RJ, Schmidt A, Morton WR, Gustafson KR, Boyd MR (2004) Cyanovirin-N inhibits AIDS virus infections in vaginal transmission models. *AIDS Res Hum Retroviruses* **20**: 11–18.

Turpin JA (2002) Considerations and development of topical microbicides to inhibit the sexual transmission of HIV. *Expert Opin Investig Drugs* **11**: 1077–1097.

Tziveleka LA, Vagias C, Roussis V (2003) Natural products with anti-HIV activity from marine organisms. *Curr Top Med Chem* **3**: 1512–1535.

Weis WI, Drickamer K (1996) Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* **65**: 441–473.

Williams DC, Jr., Lee JY, Cai M, Bewley CA, Clore GM (2005) Crystal structures of the HIV-1 inhibitory cyanobacterial protein MVL free and bound to $Man_3GlcNAc_2$: structural basis for specificity and high-affinity binding to the core pentasaccharide from n-linked oligomannoside. *J Biol Chem* **280**: 29269–29276.

Xiong C, O'Keefe BR, Botos I, Wlodawer A, McMahon JB (2006) Overexpression and purification of scytovirin, a potent, novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. *Protein Expr Purif* **46**: 233–239.

Yamaguchi M, Ogawa T, Muramoto K, Kamio Y, Jimbo M, Kamiya H (1999) Isolation and characterization of a mannan-binding lectin from the freshwater cyanobacterium (blue-green algae) *Microcystis viridis*. *Biochem Biophys Res Commun* **265**: 703–708.

Yang F, Bewley CA, Louis JM, Gustafson KR, Boyd MR, Gronenborn AM, Clore GM, Wlodawer A (1999) Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol* **288**: 403–142.

Ziółkowska NE, O'Keefe BR, Mori T, Zhu C, Giomarelli B, Vojdani F, Palmer KE, McMahon JB, Wlodawer (2006) Domain-swapped structure of the potent antiviral protein griffithsin and its mode of carbohydrate binding. *Structure* **7**: 1127–1135.

Ziółkowska NE, Shenoy SR, O'Keefe BR, McMahon JB, Palmer KE, Dwek RA, Wormald MR, Wlodawer A (2007) Crystallographic, thermodynamic, and molecular modeling studies of the mode of binding of oligosaccharides to the potent antiviral protein griffithsin. *Proteins: Struct Funct Bioinform*, in press.