

The Novel Fold of Scytovirin Reveals a New Twist For Antiviral Entry Inhibitors

Robert L. McFeeters¹, Changyun Xiong², Barry R. O'Keefe²
Heidi R. Bokesch^{2,3}, James B. McMahon², Daniel M. Ratner⁴
Riccardo Castelli⁵, Peter H. Seeberger⁵ and R. Andrew Byrd^{1*}

¹Structural Biophysics
Laboratory, National Cancer
Institute, Frederick
MD 21702-1201, USA

²Molecular Targets
Development Program
National Cancer Institute
Frederick, MD 21702-1201
USA

³SAIC-Frederick, National
Cancer Institute, Frederick
MD 21702-1201, USA

⁴Department of Chemistry
Massachusetts Institute of
Technology, Cambridge
MA 02139, USA

⁵Laboratory for Organic
Chemistry, Swiss Federal
Institute of Technology (ETH)
Zürich, 8092 Zürich
Switzerland

The solution structure of the potent 95 residue anti-HIV protein scytovirin has been determined and two carbohydrate-binding sites have been identified. This unique protein, containing five structurally important disulfide bonds, demonstrates a novel fold with no elements of extended regular secondary structure. Scytovirin contains two 39 residue sequence repeats, differing in only three amino acid residues, and each repeat has primary sequence similarity to chitin binding proteins. Both sequence repeats form similarly structured domains, with the exception of one region. The result is two carbohydrate-binding sites with substantially different affinities. The unusual fold clusters aromatic residues in both sites, suggesting a binding mechanism similar to other known hevein-like carbohydrate-binding proteins but differing in carbohydrate specificity. Scytovirin, originally isolated from the cyanobacterium *Scytonema varium*, holds potential as an HIV entry inhibitor for both therapeutic and prophylactic anti-HIV applications. The high-resolution structural studies reported are an important initial step in unlocking the therapeutic potential of scytovirin.

© 2007 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: scytovirin; antiviral; HIV; carbohydrate; hevein

Introduction

Protein:carbohydrate interactions play fundamental roles throughout biology. Bacterial infection, cell growth, inflammation, cell mobility, fertilization,

cell to cell adhesion, and viral infection are all influenced by protein:carbohydrate interactions. Therefore protein:carbohydrate interactions are a topic of major interest throughout science. One well-studied group of lectins is the hevein-like family. This family is structurally united by the presence of similar chitin-binding domains (CBDs) and are members of carbohydrate-binding module family 18.¹ Hevein, isolated from the latex of *Hevea brasiliensis*, is composed of 43 residues and its structure has been known since 1991.² Other hevein-like CBDs vary in length, typically from 38–45 residues, and are often found in multiple repeats within a single protein. A dual CBD-like motif is present in the potent antiviral protein scytovirin, originally isolated from the cyanobacterium *Scytonema varium*.

Present addresses: D.M. Ratner, Department of Molecular and Cell Biology, Boston Medical Center, Boston University, Boston, MA, 02215, USA.

Abbreviations used: CBD, chitin-binding domain; RDC, residual dipolar coupling; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; HSQC, heteronuclear single quantum coherence.

E-mail address of the corresponding author:
rabyrd@ncifcrf.gov

Scytovirin is of particular interest due to the fact that it possesses potent and broad spectrum antiviral, including anti-HIV, activity not found in the other hevein-like domain proteins.³ Scytovirin exhibits low nanomolar activity against HIV and binds to the HIV envelope glycoproteins gp160, gp120 and gp41, but it does not bind to the T-cell extracellular CD4 receptor or other common cell surface proteins.⁴ The interaction is carbohydrate-dependent, with scytovirin most readily binding to glycosylated gp120. Scytovirin's ability to completely inactivate laboratory strains and primary isolates of HIV-1⁴ makes it a promising candidate for anti-HIV microbicide development.

Scytovirin contains two 39 residue repeats, which exhibit rather low similarity to a subset of diverse CBD-containing proteins.⁴ The initial sequence analysis offers little indication of the mechanism of antiviral activity beyond similarity to CBDs. The comparison to hevein-like CBDs was not originally examined; however, the sequence similarity between the scytovirin repeats and the hevein domain is approximately 30%. In addition, several features common to hevein family members^{5,6} are present in scytovirin: There are two CBD-like domains separated by a proline-rich linker. Both scytovirin and hevein-like CBDs have a high percentage of cysteine residues, all participating in disulfide bonds. Each domain of scytovirin has three aromatic residues, similar to the conserved triad of aromatic residues found in other hevein-like CBDs. Thus, it seems plausible that scytovirin is another hevein family member.

However, scytovirin shows distinct differences from the hevein-like proteins. First, scytovirin does not bind chitin.⁴ In fact, scytovirin shows a very restrictive carbohydrate binding specificity, having no measurable affinity for monosaccharides or common trisaccharide cores from higher order mannose oligosaccharides^{4,7} (*vide infra*). This observation is in stark contrast to the increasing affinity for longer oligosaccharides reported for other hevein-like CBDs.⁸ Second, the disulfide bonds among the conserved cysteine residues of scytovirin are shuffled compared to the consensus hevein arrangement. Scytovirin has only five cysteine residues per domain. These align exactly with five of the eight cysteine residues per domain in the classical hevein arrangement. The presence of an odd number of cysteine residues in each domain of scytovirin causes the disulfide bonds to rearrange, resulting in two disulfides per domain and one inter-domain disulfide,⁴ unlike the four disulfide bonds typically found in hevein domains. Although scytovirin has three aromatic residues per domain, similar to hevein-like CBDs, the sequence arrangement of the aromatic residues does not match the classic hevein domain sequential arrangement, *vide infra*.

Thus, scytovirin has remained a rather enigmatic, but potent antiviral protein. In order to better understand the antiviral mechanism and enable development of this natural protein, we have

determined the three-dimensional solution structure of recombinant scytovirin and found it to have a novel fold, quite different from hevein. We have also examined the carbohydrate-binding properties with short oligomannose carbohydrates, which were previously identified.⁷ These structural studies are a vital part of the future development for scytovirin and detail the unique characteristics of this protein, both with respect to its antiviral activity and its relationship to other well-studied carbohydrate-binding proteins.

Results

Chemical shift and NOE assignments

Chemical shift assignments of recombinant scytovirin (Figure 1(a)) were complicated by resonance degeneracy. A doubling of most peaks was immediately apparent from the ¹⁵N-heteronuclear single quantum coherence (HSQC) (Figure 1(b)). Duplicity of the amide backbone chemical shifts for a majority of the resonances readily suggested that the sequence repeats of scytovirin have similar backbone folds. There was an indication that each sequence repeat might represent a separate domain of the total protein structure, and we refer to the two sequence repeats as domains, which is borne out in the complete analyses (*vide infra*). Chemical shift degeneracy was exacerbated for side-chains where ¹³C and ¹H resonances were not as well dispersed as ¹⁵N chemical shifts. For illustration, strip plots for a stretch of five consecutive residues in each domain, showing the C^α and C^β resonances and sequential correlations from triple resonance NMR spectra, are shown in Figure 1(c).

The three amino acid differences between sequence repeats of scytovirin aided the sequential assignment process; nevertheless, meticulous manual analysis of standard triple resonance spectra had to be supplemented with additional experiments to unambiguously assign backbone resonances. An HMQC-NOESY-HMQC,⁹ twice taking advantage of ¹⁵N chemical shift dispersion, greatly aided the resolution of inter- versus intra-domain backbone sequential assignments. Pulse sequences to assign through proline resonances¹⁰ were also used to confirm assignments. Information from these experiments was especially useful for the P-G-G-P sequence in each domain and the P-D-P-G-P sequence in the linker. Manual analysis of nuclear Overhauser effect spectroscopy (NOESY) data was also utilized to help resolve chemical shift ambiguity. In summary, nearly complete domain-specific resonance assignments were obtained for the first 94 residues of scytovirin. Residue 95, the C-terminal alanine, was not observable. Over 98% of backbone and 92% of side-chain resonances were accounted for. As expected, each ¹⁵N-HSQC peak pair corresponded to complementary resonances of the 39 residue sequence repeats. Chemical shift

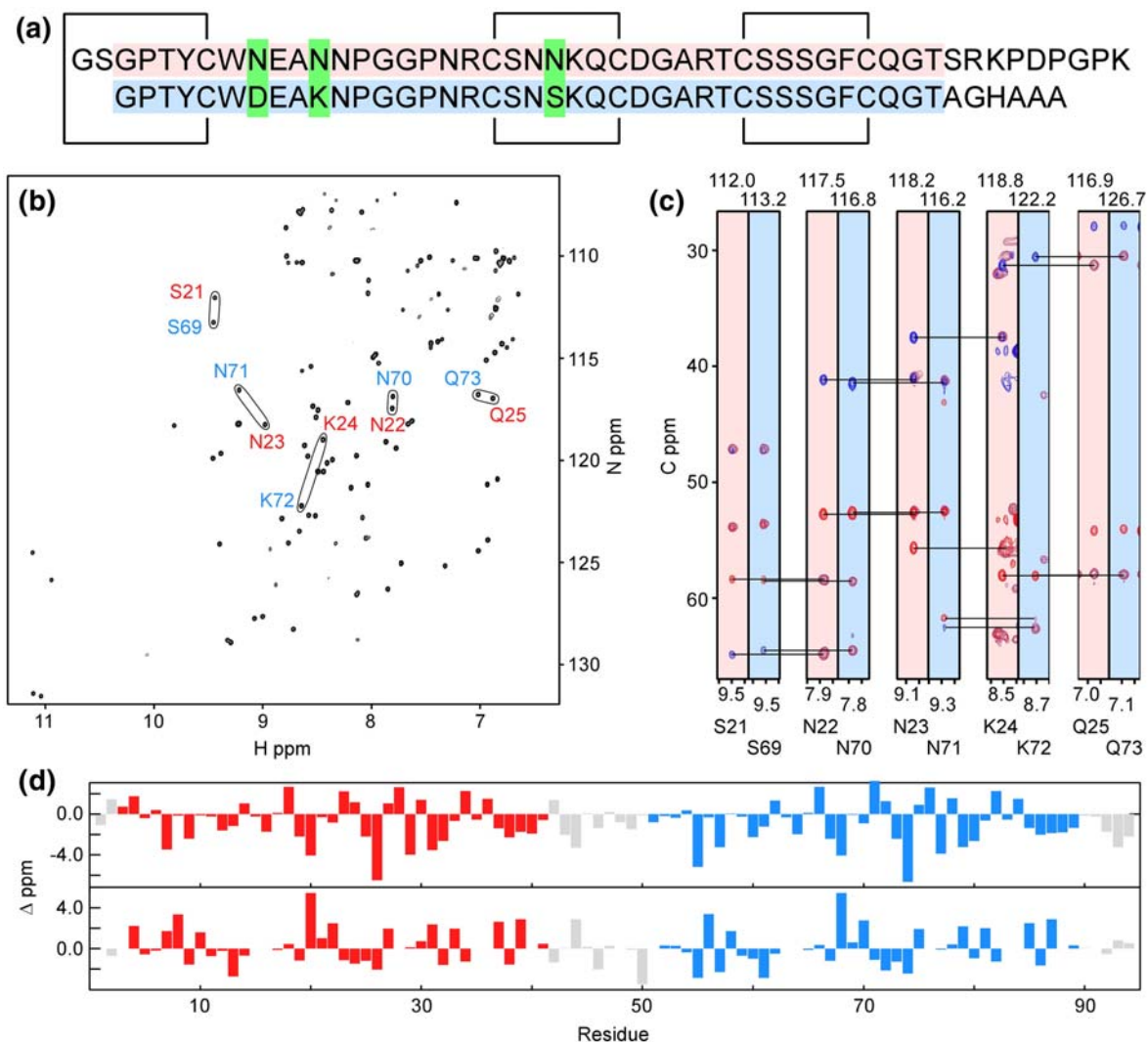


Figure 1. Scytovirin sequence, resonance assignments, and chemical shift indices. (a) Amino acid sequence of scytovirin aligned to show sequence repeats. The background of the first domain (SD1) is shaded in red while the background of the second domain (SD2) is shaded in blue. Black lines indicate disulfide bonds. Green boxes highlight the three natural sequence differences between domains. (b) An ^{15}N -HSQC of scytovirin illustrates peak doubling of complementary resonances. Residues S21, N22, N23, K24, and Q25 from SD1 are labeled in red and complementary residues S69, N70, N71, K72, and Q73 from SD2 are labeled in blue. (c) Slices from HNCACB and CBCA(CO)NH spectra for the same ten residues as (b). HNCACB C^α resonances are shown in red and C^β resonances in blue. CBCA(CO)NH resonances, from previous residue C^α and C^β resonances, are shown in magenta. Background shading is red for SD1 and blue for SD2. (d) The difference in C^α (top) and C^β chemical shifts from random coil values. No extended regions of regular secondary structure are expected, since the characteristic positive difference for α -helices and negative difference for β -sheets is not observed for C^α chemical shifts.

analysis of backbone C^α and C^β resonances^{11,12} (Figure 1(d)) gave the first indication that no extended regions of regular secondary structure were present.

NOESY assignments were also complicated by chemical shift degeneracy, especially for inter-*versus* intra-domain NOEs. No symmetry for the two domains was assumed and an initial structure was calculated from a sparse set of unambiguously assigned NOEs. A moderate resolution structure was obtained and a degree of 2-fold symmetry was immediately apparent. The initial structure allowed for resolution of more inter-*versus* intra-domain NOE ambiguity. After multiple iterations, 1205

intraresidue, 1006 sequential, and 634 long range NOEs were assigned (Table 1).

Structural overview and backbone flexibility

The solution structure of scytovirin was determined using NOEs and residual dipolar coupling data¹³ (RDCs). Initial structures, determined using only NOEs, exhibited a backbone RMSD of $1.4(\pm 0.1)$ Å to the mean structure. The sequence repeats corresponded to structured domains with residues 3–43 forming structural domain 1 (SD1) and 51–89 forming structural domain 2 (SD2). The disulfide bonding pattern, previously determined from pro-

Table 1. NMR and refinement statistics for scytovirin

<i>Distance constraints</i>	
Total NOE	2905
Intra-residue	1247
Inter-residue	1658
Sequential ($ i-j =1$)	679
Medium-range ($ i-j < 4$)	324
Long-range ($ i-j > 5$)	655
Total RDCs	148
NH	82
H ^α C ^α	66
Q (%)	0.25
<i>Structure statistics</i>	
Violations (mean and s.d.)	
Distance constraints > 0.5 Å (Å)	24±2
Average violation > 0.5 Å (Å)	0.69±0.16
Max. distance constraint violation (Å)	1.17
Deviations from idealized geometry	
Bond lengths (Å)	0.01±0.06
Bond angles (°)	1.59±2.55
Impropers (°)	0.08±2.19
Average r.m.s.d. to mean ^a (Å)	
Heavy	0.42±0.11
Backbone	0.51±0.11

^a Calculated from 20 refined structures.

teolysis and peptide mapping,⁵ was readily apparent in these initial structures. Structure calculations with and without disulfide restraints showed only small differences, and subsequent calculations included these restraints. Backbone RMSDs for the domains, aligned individually, were 1.0 Å for SD1 and 1.3 Å for SD2. More striking than the domain similarity was the complete absence of regular secondary structure. In agreement with chemical shift indexing data, no extended α -helices or β -sheets were observed. Circular dichroism (Supplementary Data Figure 1) also indicates an atypical fold. The CD spectrum for scytovirin does not exhibit an α -helical, β -sheet, mixed helix/sheet, or unfolded profile. Refinement against J -modulated H^αC^α and NH RDCs confirmed the unusual fold and total lack of regular secondary structure. Two regions of extended structure are superficially similar to β -sheets. In SD1, residues 28–32 lie antiparallel to residues 38–42. Also, residues 6–12 lie antiparallel to the disjoint sections 17–20 and 26–28. Although reminiscent of β -sheets, neither Phi/Psi angles nor chemical shift indexing strictly indicate the presence of classical β -sheet folds. Furthermore, temperature coefficients of the amide resonances¹⁴ are not consistent with hydrogen bonding patterns expected for antiparallel β -sheets. The same holds for complementary residues in SD2. Additionally, there are no Ile, Leu, or Val residues in the sequence, which results in a very minimal hydrophobic core. The limited hydrophobic core within each domain is consistent with the unusual fold. Thus, scytovirin has no recognizable α -helix or β -sheet elements of secondary structure.

The lowest energy 20 refined scytovirin structures are shown in Figure 2 and coordinates are deposited under PDB accession number 2JMV. In the final

structures, the backbone RMSD of the entire protein improved to 0.42(±0.11) Å (Table 1). Backbone RMSDs of the individual domains improved to 0.25 Å for SD1 and 0.40 Å for SD2. In addition to improved precision, RDC refinement helped define the linker and last loop-like structure of each domain. Calculated *versus* observed RDC plots and cross-validation are included as Supplementary Data Figure 2. A DALI¹⁵ structure search found no similar structures for either SD1, SD2 or full-length scytovirin. Thus scytovirin represents a truly novel fold.

Despite the lack of regular secondary structure, scytovirin appears to be quite ordered on NMR measurable timescales. Transverse and longitudinal ¹⁵N relaxation rates showed no regions of significant variation except in the C terminus. The correlation time derived from these data was 5.1 ns, which corresponds to a monomeric 9.7 kDa protein. Heteronuclear NOEs¹⁶ at multiple field strengths resulted in average values (excluding the two C-terminal residues) of 0.80±0.05 and 0.80±0.07 at 600 and 800 MHz, respectively. Data are included as Supplementary Figure 3. No significant change in average NOE is observed when calculated separately for the individual domains. It is apparent that the backbone of scytovirin does not exhibit flexibility, except the last two C-terminal residues. The N terminus is well ordered, beginning at residue 1, and participates in structuring of SD1. The interactions of residues G1 and S2 with C26, D27, and R30 contribute to the structural difference between domains. Hydrogen/deuterium exchange experiments show that 17 residues in each domain exhibit protected amide protons. All of the protected amides were on the interior of the protein. Some of

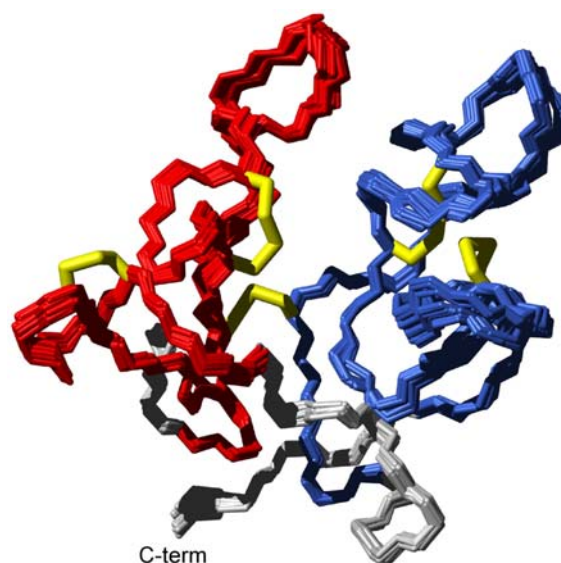


Figure 2. Solution structure of scytovirin. Backbone traces of lowest energy 20 scytovirin structures of 200 calculated from refinement against H^αC^α and NH RDCs. SD1 is colored red, SD2 blue, disulfide bonds in yellow, and the linker and termini grey.

these amides must participate in hydrogen bonds; however, due to the lack of regular secondary structure it was not possible to unambiguously assign hydrogen bonding partners. Therefore, the structures were calculated without such constraints.

Domain interface

The interface between the SD1 and SD2 lies in the middle of the protein and helps stabilize the fold of each domain. The structure of each domain was calculated separately, using intra-domain NOEs. The backbone RMSDs worsened due to the loss of inter-domain NOEs, but the same overall folds were recognized. One end of the interface is cross-linked by the disulfide bond between C7 and C55. Multiple unambiguous NOEs, most notable T5 HG to T53 HG, substantiate the presence of the disulfide bond. The remainder of the interface is predominantly hydrophobic. A large number of unambiguous NOEs are observed between residue C20 and residue Q73. The interface is $\sim 580 \text{ \AA}^2$ and may help stabilize the domain structures, in agreement with initial studies of the individual domains (unpublished results). The linker does not appear to contribute significantly to the interface.

Conserved structural elements

As expected from the similarity of chemical shifts in each domain, the amino acid sequence repeats have similar overall backbone folds; however, the RMSD between SD1 and SD2 is 5 \AA . Close in-

spection reveals a conformational change that explains the difference between the domains. Before discussion of the differences, it is beneficial to understand the likenesses.

Three discreet, common structural elements are apparent in both domains of scytovirin (Figure 3(a)). The first conserved structural element, the "top loop," is formed by residues 12–17(60–65) of SD1 (SD2) (Figure 3(b)). The top loop circumnavigates approximately 240° of a complete circle. The first position of the top loop is the site of one of the three sequence differences (N12/K60) between SD1 and SD2, and the rest of the loop is formed by a P-G-G-P sequence. Both top loops exhibit a similar pattern of heteronuclear NOEs at 600 and 800 MHz. The first glycine (G15/G63) in the loop has an average NOE value of 0.77 ± 0.01 , whereas the second glycine (G16/G64) has a value of 0.84 ± 0.02 . The first glycine is closer to the solvent-exposed apex of the loop and therefore it is reasonable that it is slightly less conformationally restricted. The second conserved structural element, denoted the "middle turn," is formed by residues 22–25(70–73) (Figure 3(c)). A disulfide bond between cysteine residues 20(68) and 26(74) helps demarcate the boundaries of this conserved structural element. In SD1, a probable hydrogen bond between the backbone carbonyl oxygen of residue 22 and the backbone amide proton of residue 25 produces a likeness to a classical turn. However, unlike a classical turn, the last residue of both middle turns introduces an abrupt change in direction. The change in direction is followed by a

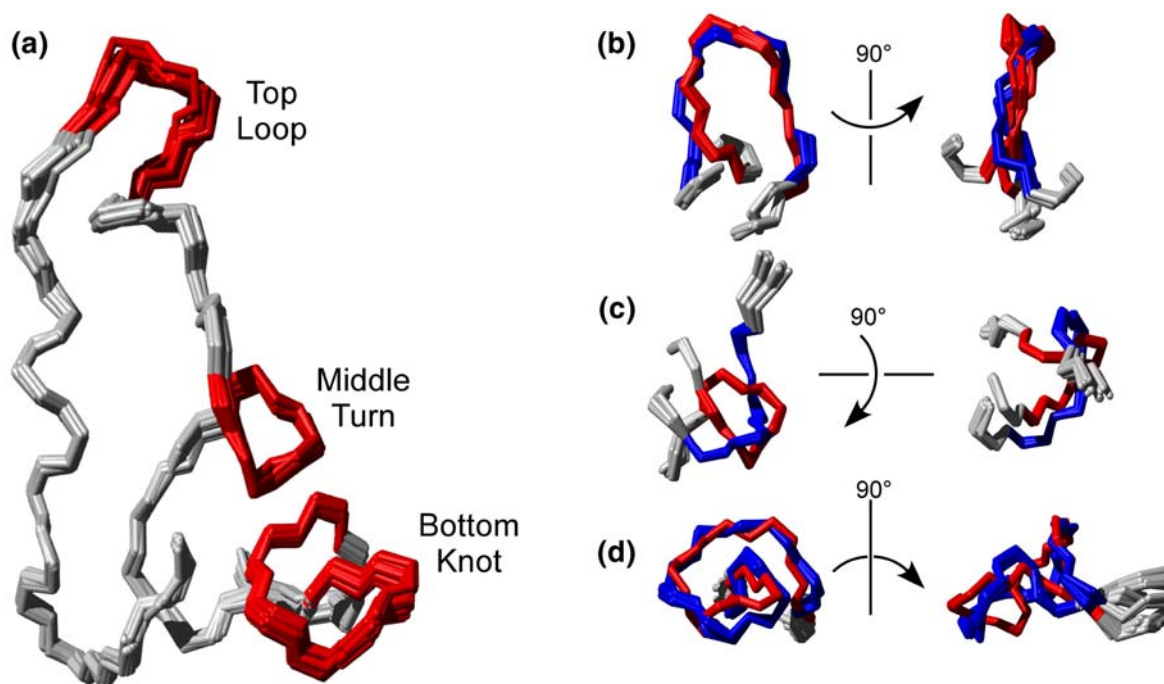


Figure 3. Three major sub-structural elements in scytovirin. (a) Twenty member ensemble of SD1 colored to show the three major loop-like elements conserved in both domains in red. See the text for discussion. (b) The top loop, (c) middle turn and (d) bottom knot of SD1 (red) and SD2 (blue) are aligned to illustrate the structural similarities and differences.

kinked extended region leading to the third conserved structural element and most distinguishing feature of scytovirin, denoted the "bottom knot," which is formed from residues 31–39 (79–87) (Figure 3(d)). As with the middle turn, a disulfide bond between residues 32–38 (80–86) helps define the boundaries of this conserved structural element. Refinement against RDCs greatly improves resolution of this element in each domain, revealing a quite unusual fold.

Domain differences

Both SD1 and SD2 are composed of the same structural elements; however, the relative orientations of the elements within each domain are not the same, resulting in the 5 Å RMSD between SD1 and SD2. The major differences are centered around the middle turn. In SD2, the middle turn is flipped 180° with respect to the rest of the domain when compared to the middle turn of SD1 (Figure 3(c)). Backbone chemical shift differences between SD1 and SD2 are greatest for residues in the middle turn and the extended region directly following (Figure 4). Interactions between residues G1 and C26, D27, R30 of scytovirin are responsible for the difference, since the side-chains intimately pack in this region of SD1 and help determine the orientation of the turn. The same does not hold for SD2, since residues P49 and K50 in the linker, equivalent in position to G1 and S2 for SD1, pack and interact differently with the residues near the middle turn of SD2. Residues preceding P49 in the linker, which are absent at the N terminus, may also contribute to the differences. The final result is a flipped orientation of the middle turn. The structural significance of the two N-terminal residues, G1 and S2, is supported by the decrease in

anti-HIV EC₅₀ by greater than a factor of 500 when these two residues are removed.¹⁷

As a consequence of the reversed middle turn, the top loop is also reversed between domains. The right-handedness of SD1 is mirrored by the left-handedness of SD2. Changes in side-chain packing caused by the natural residue differences of N9 to D57 before the top loop and N12 to K60 at the beginning of the top loop contribute to the reversed handedness. Differences also occur in hydrophobic contributions at position A11/A59. For SD1, the domain interface provides some protection for the A11. For A59 of SD2, protection by the interface is not available so the hydrophobic side-chain buries in the domain, positioning the residue much differently. These relatively small effects combine to cause the regions between the top loop and middle turn to be different between SD1 and SD2. These conformations are confirmed by the observed RDCs for these regions, and the difference leads to the large backbone RMSD between the two domains.

The backbone folds of SD1 and SD2 are quite similar, except for the region between the top loop and middle turn. The concerted change in handedness of the top loop and 180° reversal of the middle turn effectively cancel each other, preserving the remainder of the overall fold. The backbone RMSD of the two domains is 2.5 Å, when the top loop through the middle turn are left out of the calculation. The major structural differences are thereby relegated to only part of the protein, and do not dramatically impact residues near the carbohydrate-binding site (*vide infra*). This allows scytovirin to position residues critical for carbohydrate binding similarly between domains. Thus, both domains of scytovirin retain the ability to bind oligosaccharide, albeit with different affinities.

Oligosaccharide binding

Scytovirin has been shown to bind specific oligosaccharides found on gp41 and gp120.⁴ In particular, scytovirin was shown to bind to a specific tetrasaccharide substructure of the high mannose oligosaccharide normally found decorating HIV-1 envelope glycoproteins.⁷ The Man α (1→2)Man α (1→6)Man α (1→6)Man tetrasaccharide is pictured in Figure 5(a). Chemical shift perturbation mapping of scytovirin was used to monitor titration of this tetrasaccharide (see Figure 5(b)). Different exchange kinetics were observed for residues in SD1 *versus* residues in SD2. At 500 MHz, residues in SD1 exhibit intermediate exchange, whereas residues in SD2 exhibit fast exchange. These exchange regimes can be shifted by conducting the experiments at 800 MHz, where the intermediate exchange becomes slow and the fast exchange becomes intermediate (data not shown). The results suggest significantly different affinities for binding of the tetrasaccharide to each domain. Chemical shifts as a function of ligand concentration were fit to a two-site model (Matlab

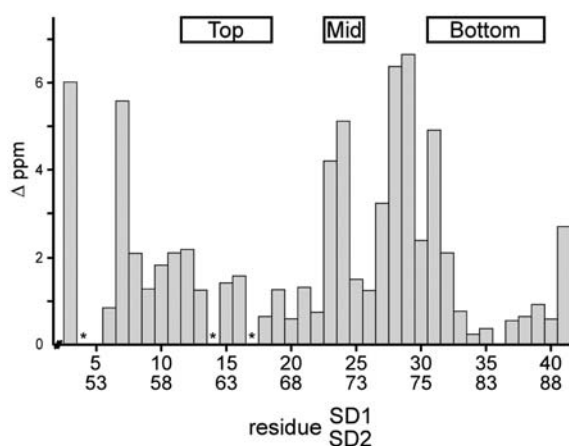


Figure 4. Amide backbone chemical shift difference between domains. The absolute difference of amide backbone chemical shifts for equivalent residues in SD1 and SD2 are plotted. ¹H contributions have been scaled to ¹⁵N ppm. Asterisks mark the position of proline residues. At the top, boxes demark residues composing the top loop, middle turn, and bottom knot.

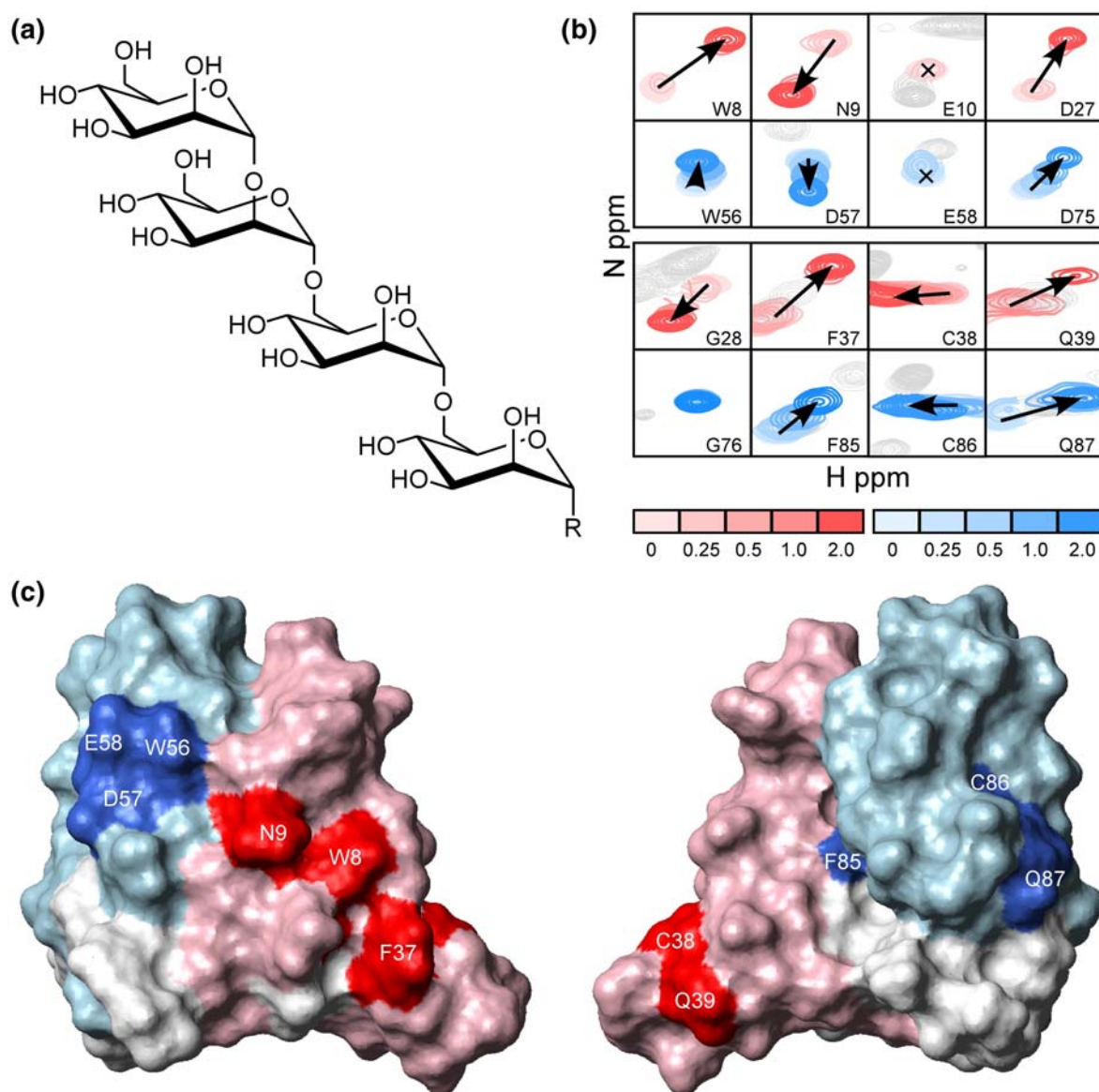


Figure 5. Oligosaccharide titrations reveal binding site. (a) Structure of the tetrasaccharide used for NMR binding studies, where R is the same as in Adams *et al.*⁷ (b) Overlay of sections from ¹⁵N-HSQC spectra collected with increasing amounts of tetrasaccharide. Complementary residues showing the largest chemical shift perturbation are shown. Residues from SD1 are in red and residues from SD2 are in blue. Color gradients from light to dark correspond to tetrasaccharide:scytovirin ratios of 0:1, 0.25:0, 0.5:1, 1:1, 2:1. (c) Surface representation of scytovirin (SD1 pink, SD2 light blue) colored to show residues having largest change in chemical shift upon tetrasaccharide binding (SD1 red, SD2 blue).

algorithm kindly provided by Dr David Fushman, University of Maryland) providing dissociation constant estimates of 30 μ M for SD1 and 160 μ M for SD2 (Supplementary Data Figure 4). These measurements agree with data measured from isothermal titration calorimetry (S. Shenoy & B. R. O’Keefe, personal communication).

Residues that demonstrate backbone amide chemical shift changes upon addition of tetrasaccharide are almost identical for SD1 and SD2. Mapping those residues onto the scytovirin structure is shown in Figure 5(c). It appears that each domain of scytovirin binds oligosaccharide in a similar fashion. This agrees with biochemical

studies of the individually expressed domains.¹⁷ There are three aromatic residues in each domain (SD1, Y6, W8, F37; and SD2, Y54, W56, F85), and, with the exception of F85 in SD2, all the aromatic residues of scytovirin are clustered near the binding sites (Figure 5(c)). Both N^εH^ε1 of tryptophan W8(SD1) and W56(SD2) demonstrate chemical shift perturbations, thus suggesting involvement in tetrasaccharide binding. The clustering of aromatics agrees well with an existing model in which two aromatic residues, separated by one residue in the amino acid sequence, play a critical role in one class of protein–oligosaccharide interactions.¹⁸

Oligosaccharide-binding site comparison

The differences between the two oligosaccharide-binding sites can partially be explained from the apo structure of scytovirin. The N9 residue in SD1, which is the site of side-chain substitution for D57 in SD2, shows large backbone amide chemical shift changes in the presence of tetrasaccharide, implicating involvement in binding. The negative charge of D57 may contribute to decreased oligosaccharide affinity in SD2. Also, the residues between the top loop and middle turn show minor chemical shift perturbations upon addition of tetrasaccharide, and some structural differences between the two domains may contribute to differences in binding. The proximity of G1 and S2 to surface-exposed residues with the largest chemical shift perturbations in SD1 indicate the significance of the ordered N terminus and local structure that contribute to the different oligosaccharide-binding affinity. The lack of chemical shift perturbation for G76 in SD2 differs from the significant change observed for the equivalent G28 in SD1. Thus, even though the sites are quite similar, enough difference exists to cause an appreciable difference in oligosaccharide affinity. Furthermore, both sites exhibit considerable specificity. We have confirmed the previous observation that a related $\text{Man}\alpha(1\rightarrow6)\text{Man}\alpha(1\rightarrow6)\text{Man}$ trisaccharide⁷ does not bind to scytovirin based on the absence of observed chemical shift perturbations in NMR titration experiments.

Discussion

At first glance, scytovirin appears to be another member of the hevein-like family, since it binds oligosaccharides, has a majority of the signature CBD cysteine residues² (Figure 6(a)), and has a similar aromatic triad involved in carbohydrate binding. The likeness of scytovirin to CBDs, however, does not extend to the backbone fold. The scytovirin structure is quite distinct from that of the hevein domain (Figure 6(b)). The difference in structure is, in part, due to the different disulfide bond formation. The absence of a cysteine at position 17 in scytovirin, strictly conserved in all hevein family members, causes a dramatic rearrangement of disulfide bonding (see Figure 6(a)). In scytovirin, the disulfide bonding pattern is C20–C26 and C32–C38 in SD1, C68–C74 and C80–C86 in SD2 with an inter-domain C7–C55 disulfide bond.⁴ This corresponds to a sequential disulfide bonding pattern between the 2nd–3rd and 4th–5th (7th–8th, 9th–10th in SD2) cysteine residues with one inter-domain disulfide bond, between the 1st–6th cysteine residues. The disulfide bonds are very important to the structural integrity of the two domains, and the inter-domain disulfide bond assists in bringing the two domains together to form the overall globular structure. In all hevein family members, the pattern of disulfide bonds is not sequential, but intertwined as evidenced from disulfide bonding between the

conserved 1st–4th, 2nd–5th, 3rd–6th, and 7th–8th cysteine residues (Figure 6(a)). The conserved disulfide pairing makes it reasonable to assume that the hevein domain structure is conserved across all family members, as represented by the *B. argeuini* hevein domain (Figure 6(b)). The hevein domain exhibits short α -helical and β -strand secondary structures, contrary to the absence of such structural elements in scytovirin.

The sequence repeat, or dual hevein motif, of scytovirin is also seen in other CBDs. For example, a recently described tomato lectin shows the same duplication of hevein domains joined by proline-rich linkers.⁵ The tomato lectin has no inter-domain disulfide bond, hence there is no *a priori* expectation that the two domains would contact or interact with one another. The dual motif would be anticipated to have a fold comprised of two typical hevein domains connected by a linker, as suggested by the model in Figure 6(b). Conversely, the inter-domain disulfide bond of scytovirin is reminiscent of mistletoe lectin where two CBDs are held together by a lone, inter-domain disulfide bond.⁶ The conserved hevein disulfide pairing of the tomato lectin (Figure 6(a)) would suggest that the individual domains have the hevein fold and that they are simply tethered. It is not known if there is any contact or interaction between the two domains, as seen in scytovirin. This variation in domain structure and interconnection might provide further differentiation in carbohydrate recognition specificity; however, further investigations are required to examine this possibility.

Each domain of scytovirin contains a triad of aromatic residues that is similar to the triad in hevein domains and is known to be responsible for carbohydrate binding in hevein.¹⁸ Despite the sequence similarity to CBDs, scytovirin lacks measurable binding affinity for chitin, monosaccharides, or common trisaccharides.^{4,7} Scytovirin does not bind the $\text{GlcNAc}\beta(1\rightarrow4)\text{GlcNAc}$ structure of chitin but instead shows high specificity for a tetrasaccharide structure of $\text{Man}\alpha(1\rightarrow2)\text{Man}\alpha(1\rightarrow6)\text{Man}\alpha(1\rightarrow6)\text{Man}$. One plausible explanation for these differences is the altered arrangement of key aromatic residues in the binding sites of scytovirin compared to hevein domains. Interestingly, SD1 of scytovirin preserves the structural arrangement of the aromatic triad (Y6, W8, and F37) and has a nearly identical affinity for its tetrasaccharide-binding partner as hevein does for $(\text{GlcNAc})_4$.⁸ SD2 exhibits a lower apparent binding affinity for the tetrasaccharide, and the aromatic triad (Y54, W56, and F85) is not as tightly clustered. The side-chain of F85 is not readily able to participate in binding due to the reversed arrangement of the middle turn. Thus, we predict that scytovirin and hevein utilize similar protein:carbohydrate interactions; however a different specificity is achieved as a direct result of the altered disulfide bonding pattern and rearrangement of the aromatic triads in each domain. In binding experiments with the tetrasaccharide, the two binding sites are independent with

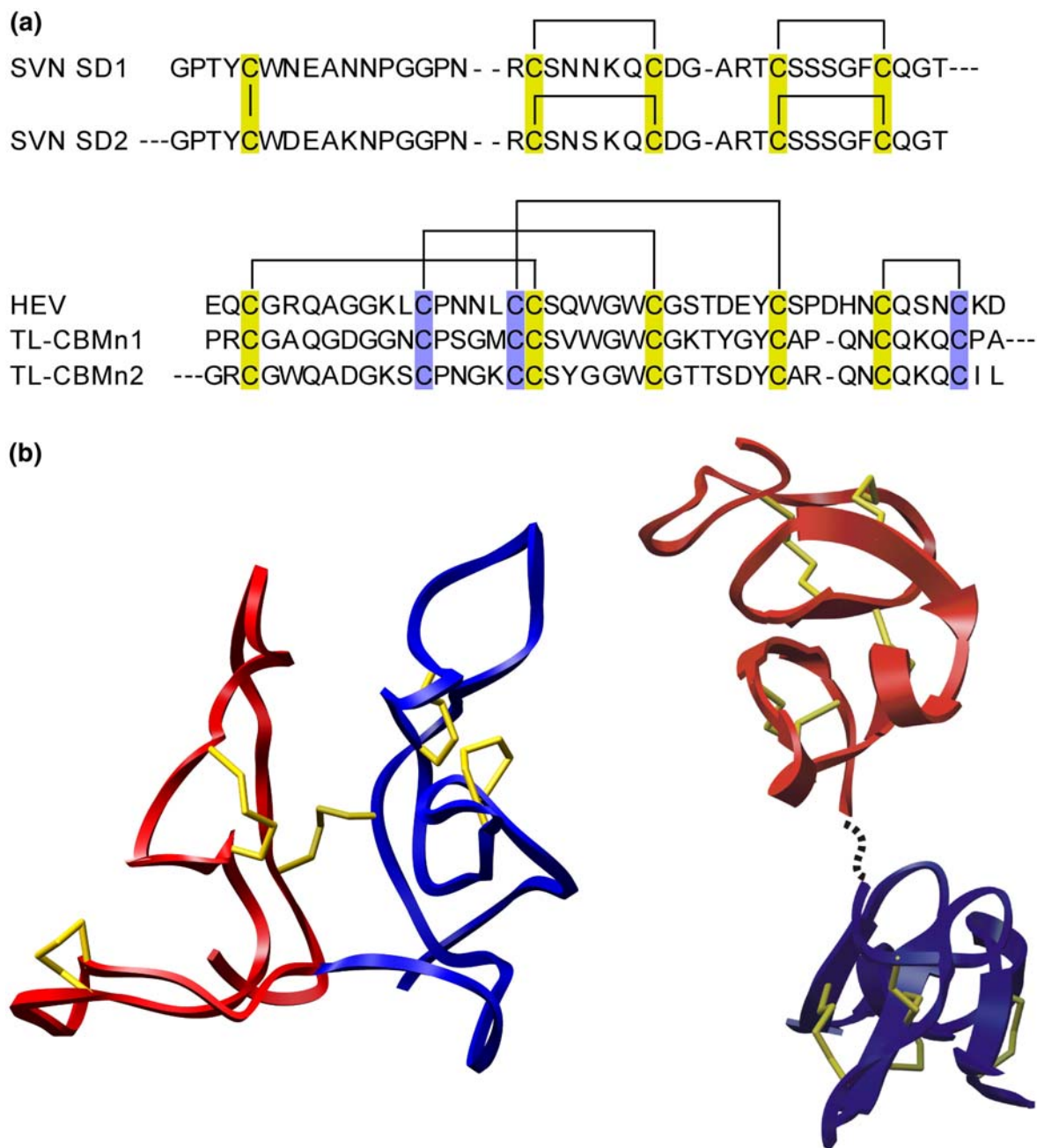


Figure 6. Sequence alignment, disulfide bonding pattern, and domain comparison of scytovirin and hevein-domains. (a) An amino acid sequence alignment of both scytovirin domains (SVN SD1 and SD2), hevein (HEV), and the hevein-like tomato lectin N-terminal chitin-binding modules⁵ (TL-CBMn1 and TL-CBMn2) illustrates the drastically different disulfide bond formation of scytovirin. Disulfide bonds are indicated by lines. Gold boxes indicate cysteine residues conserved in scytovirin and hevein-like proteins, whereas blue boxes indicate cysteine residues only found in hevein-like CBDs. (b) Comparison of the backbone folds of the dual CBDs of scytovirin (left) to a model of two arbitrarily oriented hevein domains (PDB ID: 1HEV) (right). All cysteine residues side-chains are shown, colored in yellow. SVN SD1 is shown in red and SVN SD2 is shown in blue. The model of a dual hevein-domain protein without any inter-domain disulfide bond (right) is formed from two copies of 1HEV and the linker is represented as a broken line.

relatively weak affinity. As noted for other CBDs,⁸ increased affinity is achieved by binding to large carbohydrates or glycoproteins that can deliver carbohydrates to both scytovirin domains, and this multivalency is likely responsible for the high anti-HIV activity of scytovirin.

The presence of two weak carbohydrate-binding sites that result in potent anti-viral activity towards

gp41 and gp120 parallels the observations for another cyanobacterial lectin protein with potent antiviral activity, cyanovirin-N.¹⁹ Cyanovirin-N exhibits two binding sites with micromolar affinity, similar to scytovirin. The cyanovirin-N-binding sites, however, do not involve aromatic residues, based on both NMR titrations²⁰ and the crystal structure of the cyanovirin-N:Man-9 complex.²¹

Similar to scytovirin, cyanovirin-N exhibits very high affinity toward viral glycoprotein epitopes. A recent study of cyanovirin-N¹⁹ observed precipitation of gp41 by cyanovirin-N that could be prevented by mutation of one of the binding sites, suggesting that the precipitation was due to cross-linking but concluding that the nanomolar potency was probably not due to cross-linking. Instead, the multivalency may be related to the spatial specificity of glycoprotein epitopes. Our data suggest that scytovirin may act in a parallel fashion, but the nature of the carbohydrate-binding sites are different and lead to a distinct carbohydrate specificity compared to cyanovirin-N.⁷

The structure of scytovirin introduces a novel fold and new twist to carbohydrate-binding proteins. Knowledge of the carbohydrate-binding sites provides important opportunities for structural engineering to improve this molecule as an antiviral, and particularly anti-HIV agent. Further studies are in progress to better understand the protein:carbohydrate complex. A greater understanding of how scytovirin binds to viral glycoprotein carbohydrates will contribute to advanced development of antiviral therapies and potential topical microbicide prophylaxis.

Materials and Methods

Sample preparation and NMR data acquisition

Scytovirin was expressed and purified as described.²² The NMR buffer used for all experiments was 20 mM Mes (pH 5.5), 100 μ M EDTA, 10 μ M NaN₃. The concentration of scytovirin was 1 mM for the unbound sample and 0.5 mM for the sample used for carbohydrate titration. All data, unless otherwise noted, were collected at 25 °C. The titration, assignment, and ¹³C-edited NOESY data were collected on Varian Inova spectrometers with cryogenically cooled probes at field strengths of 500, 600 and 800 MHz, respectively. ¹⁵N-edited NOESY and proline sequential assignment data were collected at 600 MHz on a Bruker Avance spectrometer. ¹⁵N-edited and ¹³C-edited NOESY mixing times were 100 and 125 ms, respectively. Heteronuclear ¹⁵N-¹H NOEs were collected with a recycle delay of 5.5 s or recycle delay of 2.5 s and a saturation time of 3 s. Hydrogen/deuterium exchange spectra were collected at times of 20 min, 40 min, 1.5 h, 4 h, 16 h, and four days after a lyophilized sample of scytovirin was resuspended in ²H₂O. All data were processed using nmrPipe²³ and visualized using SPARKY.²⁴

Structure calculation and refinement against RDCs

Structure calculations were conducted using XPLOR-NIH.²⁵ Due to the lack of extended secondary structure predicted from C α /C β chemical shifts, database calculated dihedral angle restraints were not used. A total of 1206 intraresidue, 1005 sequential, and 634 long range (defined as NOEs between protons of residues five or greater positions apart in the sequence) were used in the calculation. The lowest energy 20 structures of 200 calculated are reported. The precise and consistent suite of *J*-

modulated pulse sequences¹³ were used to measure NH and H α C α RDCs. Integer multiples of 0.70 and 1.0 ms were used for H α C α and NH coupling delays, respectively. Sixty-seven H α C α and 82 NH couplings were used for structural refinement. Ramachandran statistics included 13 dihedral angles in the most favorable region, 32 in the additionally allowed, 19 generously allowed and five disallowed.

Oligosaccharide titration

Oligosaccharides were synthesized as reported.^{26,27} Titration data were collected at both 500 and 800 MHz in NMR buffer. Ratios of approximately 0:1, 0.5:1, 1:1 and 2:1 were collected for the Man α (1-2)Man α (1-6)Man α (1-6) Man tetrasaccharide. Concentrations of 0:1, 0.5:1, 1:1, 2:1 and 10:1 were collected for the non-binding Man α (1-6)Man α (1-6)Man trisaccharide. At 800 MHz, data were collected at temperatures of 10, 17.5 and 25 °C for temperature dependence of oligosaccharide binding and tentative indication of hydrogen bonding.

Protein Data Bank accession code

Coordinates have been deposited in the RCSB Protein Data Bank under accession number 2JMV.

Acknowledgements

The authors thank Walter Thompson for protein purification, Jess Li for general laboratory support, and Vadim Gaponenko, Hana Křížová, and Amanda Altieri for useful discussions and review of the manuscript. The authors also thank Charles Schwitters for guidance in utilizing XPLOR-NIH and David Fushman for a Matlab algorithm and advice for analyzing kinetic data. This research is supported by the intramural research program of the NIH, National Cancer Institute, Structural Biophysics Laboratory. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.03.030

References

1. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). Carbohydrate-binding modules: fine-

- tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781.
2. Rodriguez-Romero, A., Ravichandran, K. G. & Soriano-Garcia, M. (1991). Crystal structure of hevein at 2.8 Å resolution. *FEBS Letters*, **291**, 307–309.
 3. Charan, R. D., Munro, M. H., O'Keefe, B. R., Sowder, R., McKee, T. C., Currens, M. J. *et al.* (2000). Isolation and characterization of *Myrianthus holstii* lectin, a potent HIV-1 inhibitory protein from the plant *Myrianthus holstii*(1). *J. Nat. Products*, **63**, 1170–1174.
 4. Bokesch, H. R., O'Keefe, B. R., McKee, T. C., Pannell, L. K., Patterson, G. M., Gardella, R. S. *et al.* (2003). A potent novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. *Biochemistry*, **42**, 2578–2584.
 5. Peumans, W. J., Rouge, P. & Van Damme, E. J. (2003). The tomato lectin consists of two homologous chitin-binding modules separated by an extensin-like linker. *Biochem. J.* **376**, 717–724.
 6. Stoeva, S., Franz, M., Wacker, R., Krauspenhaar, R., Guthohrlein, E., Mikhailov, A. *et al.* (2001). Primary structure, isoforms, and molecular modeling of a chitin-binding mistletoe lectin. *Arch. Biochem. Biophys.* **392**, 23–31.
 7. Adams, E. W., Ratner, D. M., Bokesch, H. R., McMahon, J. B., O'Keefe, B. R. & Seeberger, P. H. (2004). Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology; glycan-dependent gp120/protein interactions. [see comment]. *Chem. Biol.* **11**, 875–881.
 8. Asensio, J. L., Canada, F. J., Siebert, H. C., Laynez, J., Poveda, A., Nieto, P. M. *et al.* (2000). Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem. Biol.* **7**, 529–543.
 9. Ikura, M., Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). Detection of nuclear Overhauser effects between degenerate amid proton resonances by heteronuclear three-dimensional NMR spectroscopy. *J. Am. Chem. Soc.* **112**, 9020–9022.
 10. Schubert, M., Ball, L. J., Oschkinat, H. & Schmieder, P. (2000). Bridging the gap: a set of selective ¹H-¹⁵N correlations to link sequential neighbors of prolines. *J. Biomol. NMR*, **17**, 331–335.
 11. Wishart, D. S. & Sykes, B. D. (1994). The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J. Biomol. NMR*, **4**, 171–180.
 12. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31**, 1647–1651.
 13. McFeeters, R. L., Fowler, C. A., Gaponenko, V. V. & Byrd, R. A. (2005). Efficient and precise measurement of H(α)-C(α), C(α)-C', C(α)-C(β) and H(N)-N residual dipolar couplings from 2D H(N)-N correlation spectra. *J. Biomol. NMR*, **31**, 35–47.
 14. Cierpicki, T., Zhukov, I., Byrd, R. A. & Otlewski, J. (2002). Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. *J. Magn. Reson.* **157**, 178–180.
 15. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
 16. Kay, L. E., Torchia, D. A. & Bax, A. (1991). Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
 17. Xiong, C., O'Keefe, B. R., Byrd, R. A. & McMahon, J. B. (2006). Potent anti-HIV activity of scytovirin domain 1 peptide. *Peptides*, **27**, 1668–1675.
 18. Chavez, M. I., Andreu, C., Vidal, P., Aboitiz, N., Freire, F., Canada, F. J. & Jimenez-Barbero, J. (2005). On the importance of carbohydrate-aromatic interaction for the molecular recognition of oligosaccharides by proteins: NMR studies of the structure and binding affinity of AcAMP2-like peptides with non-natural naphthyl and fluoroaromatic residues. *Chem. Eur. J.* **11**, 7060–7074.
 19. Barrientos, L. G., Matei, E., Lasala, F., Delgado, R. & Gronenborn, A. M. (2006). Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Eng. Des. Select.* **19**, 525–535.
 20. Bewley, C. A. & Otero-Quintero, S. (2001). The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *J. Am. Chem. Soc.* **123**, 3892–3902.
 21. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H. *et al.* (2002). Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J. Biol. Chem.* **277**, 34336–34342.
 22. Xiong, C., O'Keefe, B. R., Botos, I., Wlodawer, A. & McMahon, J. B. (2006). Overexpression and purification of scytovirin, a potent, novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. *Protein Expr. Purif.* **46**, 233–239.
 23. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. [see comment]. *J. Biomol. NMR*, **6**, 277–293.
 24. Goddard, T. D. & Kneller, D. G. (0000). SPARKY 3. University of California, San Francisco.
 25. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73.
 26. Plante, O. J., Buchwald, S. L. & Seeberger, P. H. (2000). Halobenzyl ethers as protecting groups for organic synthesis. *J. Am. Chem. Soc.* **122**, 7148–7149.
 27. Ratner, D. M., Plante, O. J. & Seeberger, P. H. (2002). A linear synthesis of branched high-mannose oligosaccharides from the HIV-1 viral surface envelope glycoprotein gp120. *Eur. J. Org. Chem.* **5**, 826–833.

Edited by M. F. Summers

(Received 1 November 2006; received in revised form 6 March 2007; accepted 13 March 2007)