Published in final edited form as:
Proteins. 2009 December ; 77(4): 904–915. doi:10.1002/prot.22514.

A Designed Chimeric Cyanovirin-N Homolog Lectin: Structure and Molecular Basis of Sucrose Binding

Leonardus M.I. Koharudin[1], **William Furey**[2,3], and **Angela M. Gronenborn**[1,2,*]

[1]Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260, USA

[2]Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260, USA

[3]Biocrystallography Laboratory, PO Box 12055, VA Medical Center, Pittsburgh, PA 15240

Abstract

The NMR and X-ray structures of a designed chimeric Cyanovirin-N homolog (CVNH) protein were determined. The individual halves of the structure are similar to their counterparts in the parent proteins, with domains A and B resembling the structures of TbCVNH and NcCVNH, respectively. No significant differences between the solution and crystal conformations were observed, although details in loop conformations and distinct crystal packing-induced features are present. Carbohydrate binding studies by NMR revealed affinity and specificity for Glcα(1-2)Frc and Manα(1-2)Man and the parental half that is devoid of any sucrose affinity in NcCVNH was transformed into a genuine sucrose binding site in the context of the chimera. The atomic details of sugar recognition are seen in the crystal structure of the protein with two bound Glcα(1-2)Frc molecules. Both sugars exhibit different conformations around the glycosidic bond and engage in unique hydrogen bonding networks in the two sites. Although the protein is able to bind two Manα(1-2)Man molecules, a property associated with HIV-inactivation, no anti-HIV activity was observed for the hybrid protein. The present results provide the structural basis for sugar recognition in the CVNH family and aid in deciphering the relationship between sugar binding and anti-HIV activity.

Keywords

NMR; X-ray; CVNH; lectin; sucrose

Introduction

Sucrose is ubiquitously used in our daily food for its pleasingly sweet taste. In mammals it functions as an energy source, being channeled through various biochemical pathways into glycolysis and the tricarboxylic acid cycle in the production of ATP and NADH1. Its structure is composed of α-D-glucose (Glc) and β-D-fructose (Frc). The location of the glycosidic linkage between the C1 and C2 carbons renders Glcα(1-2)Frc a true non-reducing sugar, i.e. no hydroxyl group is available for further addition of another carbohydrate unit. This chemical particularity as well as the fact that Glcα(1-2)Frc is only produced in plants and no other organisms, makes sucrose unique in the collection of natural disaccharides. In

*Corresponding author and contact information: Angela M. Gronenborn, Department of Structural Biology, University of Pittsburgh School of Medicine, Biomedical Science Tower 3, 3501 Fifth Avenue, Pittsburgh, PA 15260, amg100@pitt.edu, Phone: (412) 648 9959, Fax: (412) 648 9008.

plants, sucrose is a major product of photosynthesis and serves as a source of carbon skeletons and energy.

Plant lectins constitute a large protein family that collectively recognizes carbohydrates. High-resolution structures of lectin/sugar complexes have contributed considerably to our understanding of specific carbohydrate recognition[2,3]. Recently, we determined solution NMR structures and carbohydrate specificities for three members of the Cyanovirin-N Homolog (CVNH) lectin family[4,5]. Among these proteins, TbCVNH, a protein from the truffle *Tuber borchii*, was found to not only bind Manα(1-2)Man (denoted simply as Man2α) and higher mannose oligosaccharides like most CVNH family members, but also to sucrose. Interestingly, the binding sites for both Man2α and sucrose overlap and reside in domain A of TbCVNH. Our structural work showed that binding to the same protein site is due to very similar conformations of the two disaccharides that match a surface binding cleft in TbCVNH. The other CVNH proteins were NcCVNH, a protein from *Neurospora crassa*, that possesses a sugar binding site on domain B, and CrCVNH, a fern protein, with two sugar binding sites on both domains, A and B. CrCVNH and NcCVNH specifically recognize Man2α and oligomannoses, with no detectable affinity for sucrose.

Recognition of di- and tri-mannosides in the CVNH family has been extensively studied both by NMR and X-ray crystallography, primarily in CV-N[6-9]. However, no information is available for sucrose recognition by any CVNH family protein. Structural work aimed at obtaining either inter-molecular NOEs between TbCVNH and sucrose in solution or crystals of the TbCVNH/sucrose complex has not yielded results, due to weak binding and flexibility of the polypeptide chain in domain B. In order to progress in our efforts to decipher the sucrose-protein interaction at the atomic level, we designed a chimeric protein, keeping domain A of TbCVNH and replacing domain B with domain B of NcCVNH (Figure 1). We hoped that a more stable protein would result from this design and that the protein would still recognize sucrose like the parent protein, TbCVNH.

This designed chimera, LKAMG, provided the opportunity for detailed elucidation of the structural determinants for sucrose interaction. The three dimensional structures of this chimera were determined both by NMR spectroscopy and X-ray crystallography. Using direct NMR titrations we investigated its sugar binding specificities, which, surprisingly, revealed higher affinities than those observed with the parental proteins. We also solved the crystal structure of LKAMG with sucrose. Structural comparisons between sugar-free and bound protein and between free and bound sucrose allowed us to characterize the determinants of protein-sucrose interactions. Despite its sugar binding properties, LKAMG is devoid of anti-HIV- activity.

Altogether, our present data provides the structural basis for understanding the specific interaction between sucrose and a CVNH protein and contribute to our knowledge of protein-carbohydrate recognition in general.

Materials and Methods

Expression constructs

LKAMG was designed to combine domain A of TbCVNH and domain B of NcCVNH. Domain A of TbCVNH comprises residues 1 to 39 and residues 90 to 103 while domain B of NcCVNH spans residues 42 to 95. In order to synthesize a synthetic gene encoding the chimeric protein, individual DNA fragments encoding residues 1 to 39 of TbCVNH (PCR1), residues 42 to 95 of NcCVNH (PCR2), and residues 90 to 103 of TbCVNH (PCR3) were amplified using the parent vectors as templates. PCR1 used oligonucleotides 5′-GCGAAATTAATACGACTCACTATA- GGGG-3′ (T7 promoter) as forward and 5′-

GTTGTTTCCGATGATGGTATCCAGATCCAGTTCGCTGG -3′ (R1) as reverse primers. For PCR1, the first 15 nucleotides (underlined) are nucleotides that code for residues 42 to 46 of NcCVNH. PCR2 (amplification of residues 42 to 95 of NcCVNH) was carried out using the oligonucleotides 5′-GGATCTGGATACC-ATCATCGGAAACAACGATGGTCACTTCC-3′ as forward (F2) and 5′- CTGAAT-ACGGTTCAGGTTAACATCGCGGTCGTGGAACTC as reverse (R2) primers. For PCR2 the first 13 nucleotides (underlined) in the forward primer are nucleotides that encode residues 35 to 39 of TbCVNH and the first 15 nucleotides (underlined) in the reverse primer encode residues 90 to 94 of TbCVNH. PCR3 (amplification of residues 90 to 103 of TbCVNH) was carried out using the oligonucleotides 5′-CGCGATGTTAACCTGAACCGTATTCAGAACGTTAA- CGGTC-3′ as forward (F3) and 5′- CAAAAAACCCCTCAAGACCCGTTTAGAG-3′ as reverse primers (T7 terminator), respectively. For PCR3 the first 12 nucleotides (underlined) in the forward primer encode residues 92 to 95 of NcCVNH and the sequence of the T7 terminator was used as the reverse primer. The second step in generating a continuous coding gene for the chimeric protein required linking the amplified PCR1 and PCR2, and PCR2 and PCR3 DNA fragments to create PCR4 and PCR5, respectively. PCR4 was created by mixing small, equal amounts of PCR1 and PCR2 products and using the forward and reverse primers from PCR1 and PCR2, T7 promoter and R2, respectively. Similarly, PCR5 was constructed, using the forward and reverse primers from PCR2 and PCR3, F2 and R3, respectively. The final step in creating the gene involved linking PCR4 and PCR3 and PCR5 and PCR1. Again, equal amounts of PCR products from the previous step were amplified with T7 promoter and T7 terminator primers, respectively. Once the complete DNA fragment encoding the full LKAMG was generated, the fragment was cloned into the pET-15b(+) expression vector (Novagen), using NdeI and XhoI restriction sites at the 5′ and 3′ ends, respectively. The N-terminal His-tagged LKAMG gene generated a protein with 3 additional residues (Gly-Ser-His) at the N-terminus after thrombin cleavage.

Protein purification

For protein expression, *E.coli* BL21 (DE3) cells (Stratagene) were transformed with pET-15b-LKAMG vector. Cells were grown at 37° C and induced with 1 mM IPTG for 4 h. Isotopic labeling was carried out by growing the cultures in modified M9 minimal media containing $^{15}NH_4Cl$ and/or ^{13}C-glucose as sole nitrogen and/or carbon sources, respectively.

Proteins were purified on a Ni^{2+} affinity column (GE Healthcare), using a linear gradient of imidazole (20-500 mM) for elution, followed by thrombin digestion and gel filtration on Superdex75 (GE Healthcare), either in 20 mM NaPhosphate buffer, 100 mM NaCl, 0.02% NaN_3, (pH 6.0) for NMR studies or in 20 mM Tris-HCl buffer, 100 mM NaCl, 1 mM DTT, 0.02% NaN_3, (pH 8.0) for crystallization. Purified proteins were concentrated using centriprep concentrators (Millipore) up to ~60 mg/ml for crystallization and to ~15.0 mg/ml (~1.2 mM) for NMR. For NMR samples, the buffer was simultaneously exchanged to 20 mM NaPhosphate, 0.02% NaN_3, 90/10% H_2O/D_2O (pH 6.0).

The oligomeric state of the chimeric protein was assessed by size-exclusion chromatography in conjunction with in-line multi-angle light scattering and refractive index detection.

NMR spectroscopy and solution structure determination

All NMR spectra for resonance assignments and NOE identification were recorded at 25° C on either ^{15}N-labeled or $^{13}C/^{15}N$-labeled samples. Protein concentrations were ~1.2 mM. Spectra were recorded on Bruker Avance800, Avance700, and Avance600 spectrometers equipped with 5-mm triple-resonance, three-axes gradient probes or *z*-axis gradient cryoprobes. Assignments were achieved (100% backbone and >90% side chain) using 3D

HNCACB, CBCA(CO)NH, H(CCO)-NH, (H)C(CO)-NH, HCCH-TOCSY, and HBHA(CO)NH experiments10,11. Distance constraints were derived from 3D simultaneous ^{13}C- and ^{15}N-NOESY experiments (mixing time 120 ms)12. Spectra were processed with NMRPipe13 and analyzed using NMRView14. NOE cross-peak assignments and intensity-to-distance calibrations were automatically generated using the program CYANA15. Hydrogen bond constraints were obtained from H/D exchange experiments and NOESY analysis. Distance constraints of 1.8–2.5 Å (H-O) and 2.5–3.5 Å (N-O) were used for each H-bond. Backbone torsion angles (φ and ψ) were predicted using TALOS16. All restraint information was applied in a simulated annealing protocol using CNS17. 800 structures were calculated, from which 30 structures with the lowest CNS target function values were selected for refinement using explicit water implemented in ARIA18. For the final 30 structures of LKAMG, 95.1 % of all residues were in the favored region of the Ramachandran plot, as evaluated by Molprobity19. All structural figures were generated with PyMOL20.

A summary of the experimental constraints as well as pertinent structural statistics are provided in Table 1. The atomic coordinates and NMR constraints have been deposited in the RCSB Protein Data Bank under accession code 2KJL.

Determination of dissociation constants for sucrose and Man2α binding

Sugar binding was followed in NMR titration experiments using ^{15}N-labeled protein in 20 mM Na phosphate buffer, 0.02% NaN_3 (pH 6.0), at 25° C by recording ^{1}H-^{15}N HSQC spectra on a Bruker Avance700 MHz spectrometer. For sucrose titrations, increasing amounts of sucrose were added to the protein and spectra were recorded for each addition up to a final molar ratio of sucrose:protein of 16:1, 16:1, and 28:1 for TbCVNH, NcCVNH, and LKAMG, respectively. For Man2α titrations, increasing amounts of Man2α were added to the protein and spectra were recorded for each addition up to a final molar ration of Man2α:protein of 8:1 for all proteins. Titration curves were plotted for four resonances that exhibited sizable, saturable shifts and no peak overlap in order to extract reliable apparent K_d values. The four resonances used were those of residues S2, Y3, C23, and N25 for domain A of both TbCVNH and LKAMG and those of G51, R81, D82, and C83 for domain B of LKAMG. The corresponding resonances of NcCVNH (G53, R83, D84, and C85) exhibited almost no changes at all or very small non-saturable shifts (e.g. N45 and D84) at the end of the titrations with sucrose. Therefore, K_d values cannot be determined. Similarly, for the Man2α titrations, the resonances used for the binding curves were those of residues Y3/Y3, C23/C23, N25/N25, and V96/V100 for domain A of TbCVNH/LKAMG. For domain B of NcCVNH/LKAMG, amide resonances of N45/N43, N56/N54, C85/C83, and E88/E86 were used. Dissociation constants were obtained by non-linear best fitting of the ^{1}HN titration curves using KaleidaGraph (Synergy Software, Reading, PA), averaging over the four curves.

Crystallization

All crystallization trials were carried out by the sitting drop vapor diffusion method at 4° C. A cluster of needles was seen in 0.2 M Li_2SO_4, 0.1 M Tris-HCl (pH 8.5), and 30% PEG 4000. This condition was further optimized by varying protein concentration (35, 40, 50, and 60 mg/ml), Li_2SO_4 concentration (0.15, 0.20, 0.25, and 0.30 M), and PEG 4000 concentration (20, 25, 30, and 35%), while keeping the ratio of protein to mother liquor constant at 8 to 1 μL.

For co-crystallization of protein-sugar complexes, protein solutions (~40 mg/ml) were incubated overnight with either sucrose or Man2α at a molar ratio of 1:40 (protein:disaccharide). Crystallization conditions with both disaccharides were identical to

those used for the free protein (0.2 M Li_2SO_4, 0.1 M Tris-HCl (pH 8.5), and 30% PEG 4000 and protein to mother liquor ratio of 8 to 1 μL). Crystals were only detected in the sucrose-LKAMG co-crystallization wells, not with Man2α.

X-ray data collection and processing

X-ray diffraction data were collected on flash-cooled crystals (-180°C) using a Rigaku FR-E generator with Saturn 944 CCD or R-AXIS HTC image plate detectors. For the free protein, two different crystal forms grew: a monoclinic form with space group $P2_1$ grew from a protein concentration of 40 mg/ml and an orthorhombic form with $P2_12_12_1$ space group from 50 mg/ml. Data for the $P2_1$ and $P2_12_12_1$ crystals were collected up to 1.56 and 1.36 Å resolution, respectively. Data for the $P2_1$ crystal of the sucrose complex was collected up to 1.88 Å resolution.

All diffraction data were processed, integrated, scaled using d*TREK software21, and eventually converted to mtz format using the CCP4 package22. All pertinent statistics are summarized in Table 2. The unit-cell dimensions of the $P2_1$ crystal were a = 33.88 Å, b = 34.44 Å, and c = 45.63 Å (α and γ = 90°, β = 101.24°) and for the $P2_12_12_1$ crystal a = 45.51 Å, b = 46.20 Å, and c = 50.38 Å (α, β, and γ = 90°), with an estimated solvent content of 41% (V_m = 2.09 $Å^3$/Da) and 42% (V_m = 2.12 $Å^3$/Da) based on Matthews Probability Calculator (http://www.ruppweb.org/Mattprob/). Both crystals contain one molecule per asymmetric unit. The $P2_1$ crystal of sucrose complex contains two molecules per asymmetric unit with unit-cell dimensions of a = 44.74 Å, b = 39.37 Å, and c = 86.52 Å (α and γ = 90°, β = 98.16°) and an estimated solvent content of 59% (V_m = 3.02 $Å^3$/Da).

Phases were determined by molecular replacement using an all Ala model of the pseudo-monomeric structure of the domain-swapped wild-type CV-N structure (PDB:1M5M), comprising residues 1 to 49 of monomer A and residues 54 to 101 of monomer B as structural probe in PHASER23. After generation of the initial model, the chain was re-built using the program Coot24. Iterative refinement was carried out by alternating between manual re-building in Coot24 and automated refinement in REFMAC25. Similar procedures were adopted for the diffraction data collected from the $P2_12_12_1$ crystal of the free protein and the $P2_1$ crystal of the complex. In the later case, the structure obtained from $P2_1$ crystal was used as the structural probe for molecular replacement.

All final models show clear electron density for all residues, except for the last residue in the $P2_1$ crystal of the free protein. For the complex, additional electron density was present in domains A and B of both monomers in the asymmetric unit. The extra density in both domains allowed fitting one sucrose molecule into each binding site. For all three polypeptide models, almost all residues are located in the allowed region of the Ramachandran plot, as evaluated by Molprobity19. 96.4, 95.3, and 93.5 % of residues reside in the favored region in the $P2_1$ free, the $P2_12_12_1$ free, and the $P2_1$ complex forms, respectively. Refinement statistics for the final models are provided in Table 2. All structural figures were generated with PyMOL20. The atomic coordinates and diffraction data for LKAMG have been deposited in the RCSB Protein Data Bank under accession codes 3HNU, 3HNX, and 3HP8.

Anti-HIV assay

Anti-HIV assay for LKAMG were carried out as described in 4.

Results and Discussion

Design and characterization of LKAMG

LKAMG was designed to posses the CVNH fold with two, different sugar binding sites, by combining domain A of TbCVNH with domain B of NcCVNH (Figure 1). Domain A of TbCVNH comprises residues 1 to 39 and residues 90 to 103 that form a helical turn (α1), a three-stranded β-sheet (β1 to β3) with the second helical turn following after strand β1, and a β-hairpin (β9 and β10). Domain B of NcCVNH is composed of residues 42 to 95, forming helix α3, a β-hairpin (β4 and β5), and another three-stranded β-sheet (β6 to β8). The helical turn α4 connects domains A and B, joining strands β8 and β9. The entire protein contains 107 residues.

LKAMG was expressed as a N-terminally His tagged protein and purified as described for the parent proteins[4]. Unlike CV-N that can exist both in monomeric and domain-swapped dimeric forms[26,27], the parental TbCVNH and NcCVNH proteins exist solely as monomeric proteins[4]. Since it is believed that the hinge region influences the partitioning between monomer and dimer states and the hinge sequence in LKAMG is derived in its entirety from the monomeric parental NcCVNH, we anticipated that LKAMG would also be monomeric in solution. As expected, multi-angle light scattering with refractive index detection showed only monomer, both, at NMR concentration and under crystallization conditions (see further results below).

The 2D ^{1}H-^{15}N HSQC spectrum of LKAMG reveals excellent chemical shift dispersion and line-widths, indicating that the protein is stably folded (Figure 2A). Resonances of residues residing in domain A and B are generally similar in amide chemical shifts to their counterparts in TbCVNH (colored in magenta) and NcCVNH (colored in cyan), respectively. Notable differences ($\Delta\delta$ > 1ppm), however, were observed for the following residues; M1, L12, T13, L37, D38, I40, I41, N43, W50, G51, N54, F55, T56, E57, T59, I75, D82, R96, I97, and Q107 (inset of Figure 2A). Residues L37 to N43 and R96 to I97 are located in regions that join the two domains. Therefore, chemical shift differences are expected for this area of the protein. Also, the hinge region, comprising residues W50 to N54, together with M1 and Q107, experience different environments since these regions reside at the interface between the two pseudo-symmetric domains of the protein. Indeed, this region showed large chemical shift differences from the parental proteins. Interestingly, helix α3, comprising residues F55 to T59, exhibited the largest chemical differences (2-7 ppm). In the parental protein (NcCVNH), these residues in α3 interact with amino acids in the long neighboring loop that connects strands β1 and β2[4]. Since this loop in the chimeric protein is short, like in TbCVNH, it does not reach helix α3. Therefore, the large chemical shift differences observed for F55 to T59 are caused by the lack of interaction between these two structural elements. Several other residues also underwent appreciable chemical shift changes, such as L12, T13, I75, and D82. In TbCVNH, residues L12 and T13 reside in strand β1 and L12 is involved in hydrophobic contacts with the conserved L64 on strand β6[4]. In the chimeric protein, strand β6 is derived from NcCVNH and L64 is replaced by an aromatic residue, F66, causing sizable differences in resonance frequencies due to ring current effects. As for I75 and D82, both these residues reside on strand β7 with I75 next to the aromatic F64, while D82 contacts helix α3 in the chimera. Naturally, some structural adjustment will occur in these regions and individual amino acids will experience a different structural and electronic environment compared to the parental protein.

In order to establish and confirm that the chimera binds the same sugars as the parental proteins, Man2α and sucrose were titrated into LKAMG and binding was evaluated by ^{1}H-^{15}N HSQC spectroscopy. Man2α is recognized by both parent proteins and binds on domain A in TbCVNH and on domain B in NcCVNH, while sucrose only interacts with

domain A of TbCVNH[4]. As expected, the chimeric protein interacts with both sugars (Figure 2B and 2C). However, several surprising aspects were noted: first, the interaction of Man2α with LKAMG is ~2 and ~8 fold tighter than with either of the parental proteins, with apparent dissociation constants of 1.8 ± 0.2 and 0.3 ± 0.0 mM for domains A and B of LKAMG compared to 3.7 ± 0.5 mM for TbCVNH (domain A) and 2.5 ± 0.3 mM for NcCVNH (domain B), respectively. Second, domain B of LKAMG, which is equivalent to domain B of NcCVNH, also interacts with sucrose (Figure 2C). This was completely unexpected since no sucrose binding was detected in domain B of NcCVNH[4] and domain A of NcCVNH was found not to interact with any sugar at all. The apparent K_d values for sucrose binding to domains A and B of LKAMG were 15.2 ± 1.9 mM and 7.3 ± 1.0 mM, respectively. The apparent K_d value for sucrose binding to domain A in TbCVNH was 4.7 ± 0.3 mM. For domain B in NcCVNH, no saturable binding of sucrose was observed, thus, no corresponding K_d value can be determined. One possible explanation for this finding may relate to the structural adjustment in the sugar binding regions in the chimeric protein. Indeed the backbone chemical shift differences support such conformational rearrangement in the hinge region and helix α3, both of which contribute to the sugar binding site in domain B of CVNH proteins.

LKAMG Structures –NMR

The NMR solution structure of LKAMG was solved using uniformly ^{15}N- and $^{13}C/^{15}N$-labeled protein and methodology commonly employed10. Resonance assignments for LKAMG are 95.8 % complete. Using the automatic NOE assignment and intensity-to-distance conversion modules in CYANA15 yielded a total of 2650 unique proton distance constraints, extracted from 3- and 4-D ^{15}N- and ^{13}C-edited NOESY spectra. Additionally, 106 hydrogen bond constraints were added based on the presence of slowly exchanging amide protons in hydrogen-deuterium exchange experiments and from analyzing NOESY patterns in the ^{15}N- and ^{13}C-NOESY spectra. Dihedral angle constraints were predicted from chemical shifts using the program TALOS16 and a total of 68 residues exhibited ψ and φ angles that were well predicted. Structures were calculated with CNS28, incorporating a total of 2891 experimental constraints. The final 30 conformer ensemble exhibits backbone and heavy atom r.m.s. distributions of 0.23 ± 0.04 Å and 0.61 ± 0.04 Å, respectively. A best fit superposition of the C_α traces of the 30 LKAMG conformers is displayed in Figure 3A. All individual members of the ensemble exhibit excellent covalent geometry and no violations with respect to the experimental data are present (Table 1).

LKAMG Structures –X ray

In parallel with structural determination by NMR spectroscopy, crystallization trials of LKAMG yielded diffraction quality crystals in two different space groups: $P2_1$ and $P2_12_12_1$ that diffracted up to 1.56 and 1.36 Å, respectively. Phasing information was obtained by molecular replacement using a pseudo-monomeric structure of the crystallographic wild-type CV-N model (PDB:1M5M)7, composed of residues 1 to 49 of monomer A and residues 54 to 101 of monomer B, as a structural probe. A summary of X-ray data and refinement statistics for both crystal structures is presented in Table 2. The final refined models consist of residues 1–106 (Figure 3B) and 1-107 (Figure 3C) for the $P2_1$ and $P2_12_12_1$ crystals, respectively. Only weak electron density was observed for the three tag residues, G-S-H, at the N-terminus in both crystal forms and the last residue (Q107) in the $P2_1$ crystal. Therefore, these amino acids could not be accurately positioned and were excluded from the model. The unique availability of high-resolution structures for the identical protein in solution and two different crystal forms allows for a direct comparison of their global structures as well as for detecting important local differences.

Comparison between solution and crystal structures of LKAMG

The large number of constraints and high-resolution diffraction data yielded well defined models for LKAMG. Its overall fold, both, in solution and the crystal, closely resembles that of the parental molecules (Figures 3 and S1)4. It comprises two triple-stranded β-sheets (one formed by strands β1, β2, and β3 and the other formed by strands β6, β7, and β8), two β-hairpins (one formed by strands β9 and β10 and the other formed by strands β4 and β5) and four 3_{10}-helical turns (α1, α2, α3, and α4). Superpositions of the solution and the two crystal structures yield pairwise r.m.s.d. values of 1.00 ± 0.05 Å and 1.84 ± 0.04 Å for backbone and all heavy atoms, respectively for the $P2_1$ structure (Figure 3D) and 0.97 ± 0.05 Å and 1.81 ± 0.04 Å for the $P2_12_12_1$ structure (Figure 3E).

Noticeable backbone variations between the solution and the two crystal structures can easily be discerned in the three loop regions (connecting strands β1 and β2, connecting strands β6 and β7, and connecting strands β7 and β8) (Figure 3D and 3E). These local differences are most likely caused by the different environments encountered by these loops in solution and the crystal lattice. In solution, the loops are not interacting with any neighboring molecules and the protein moves around freely. In the crystal lattice, on the other hand, the loops are involved in contacts with neighboring molecules, restricting its conformational space. Indeed, crystal-packing effects clearly play the major role in the different loop conformations. A detailed analysis of the equilibrium dynamics for this protein and the effects of crystal packing on dynamics is provided in an accompanying manuscript (*Liu et al.*).

The conformations of the two X-ray models are very similar, exhibiting pairwise r.m.s.d. values of 0.36 ± 0.00 Å and 0.92 ± 0.00 Å for backbone and all atoms, respectively (Figure 3F). A small shift in the backbone conformation is observed for the β1, β2 and β7, β8 loops. Solvent exposed side chain chains are found in different conformations in the two crystal structures, contributing to the higher r.m.s.d. value for all atoms. While all core hydrophobic residues and a few solvent exposed residues exhibit identical conformations (Figure S2A), most of the surface amino acids, in particular Arg, Glu, Asp, Gln, and Asn residues are distinct in the two structures. Examples of residues that exhibit different side-chain conformations like D27, Q53, and E86 are depicted in supplementary Figure S2B. Almost half of all amino acids differ in side chain conformation in the two structures and they are spread throughout the sequence: M1, N9, L12, T13, N14, G15, G16, R17, R20, E22, R24, N25, A26, D27, G28, N29, Q49, N53, I62, R63, E72, L76, R81, D82, C83, N84, E86, F87, D89, R90, D91, N95, R96, Q98, N101, R103, V104, F106, and Q107. The fact that different side chain conformations are observed is not due to phasing bias or error, since their individual removal during refinement does not change their conformations in the electron density map. Therefore, undoubtedly, crystal packing in the two different space groups lies at the heart of this variability in side chain structure and may explain why these side chains exhibit conformational averaging in solution.

Anti-HIV assay

The HIV-inactivating properties of wild type CV-N are connected the protein's ability to bind to mannose sugars on the viral envelope glycoprotein gp12029. Since LKAMG contains two Man2α binding sites, one in each domain, analogous to monomeric wild type CV-N, it seemed prudent to test whether the protein exhibited any effect in HIV infection assays. No HIV inactivation was observed at nanomolar or micromolar concentrations. While the dissociation constant for Man2α for the wild-type CV-N is in the micromolar range, we noticed that the dissociation constant for Man2α for LKAMG lies in the millimolar range (see above). Therefore, the ~1000 fold lower affinity towards Man2α may be the reason for the loss of anti-HIV activity for LKAMG.

Sucrose bound LKAMG

LKAMG was co-crystallized in the presence of sucrose and the structure of the complex was obtained from a monoclinic crystal that diffracted up to 1.88 Å. Two molecules (designated monomer A and B throughout the manuscript) were found in the asymmetric unit, in contrast to the sugar free protein crystals that contained only one molecule per asymmetric unit. Both monomers are very similar with only minor differences in the β1, β2 and β7, β8 loops (Figure S3A). These minor differences are reminiscent of the differences observed between the $P2_1$ and $P2_12_12_1$ crystals for the free proteins (Figure 3F). Therefore, conformational variability clearly indicates that the protein is flexible in these loop areas.

In the final refined structural model, clear electron density for the four bound sucrose molecules is observed and their molecular structures as well as the associated electron density in domain A and B of monomer A are displayed in Figure 4A and 4B. Only very minor differences exist when comparing the bound ligands in equivalent binding sites in molecule A and B, whereas remarkably different structures are seen in domains A and B (Figure S3B). Torsion angles around the glycosidic bond are Φ = 72.3° and 73.4° and Ψ = -93.3° and -84.4° for sucrose in domain A of monomer A and B, respectively, and Φ = 114.3° and 104.2° and Ψ = -49.2° and -42.3° for sucrose in domain B of monomer A and B, respectively. Given their almost identical conformations in equivalent sites in monomers A and B we only consider and discuss bound sucrose in monomer A throughout this manuscript for the complex.

Intriguingly, the bound sucrose molecules in domains A and B are quite different (Figure 4C) illustrating that sucrose can adopt a variety of conformations around its glycosidic bond. Small energetic differences caused by different interacting side chains on the protein will select those conformers that yield the lowest energy for the complex. Interestingly, simple inspection of LKAMG in the presence of sucrose revealed that more resonances were affected in domain A than in domain B (Figure 2C). Sucrose binding to domains A and B of LKAMG yielded apparent K_d values of 15.2 ± 1.9 mM and 7.3 ± 1.0 mM, respectively. The observation of such weak binding by NMR is in complete accord with the data from the X-ray models in which no major conformational changes in the protein were noted for the sucrose complex (Figure S3B). It appears that sucrose simply sits in crevices on the surface of the protein with specific contacts provided by shape complementarity and hydrogen bonds between hydroxyl groups on the sugar and pliable polar protein side chains. In the present complex, the sucrose molecules predominantly utilize the C3 and C4 attached hydroxyl groups on both glucosyl and fructosyl rings for hydrogen bonding in both domains (Figure 4D and 4E). For domain A interactions, the C3 hydroxyl group of glucosyl ring is involved in hydrogen bonding with the amide nitrogen of N99 and carbonyl oxygen of Q98 side chain, while the C4 hydroxyl hydrogen bonds with the backbone carbonyl oxygen of S2 and the hydroxyl group of the S6 side chain. On the fructosyl ring, the C3 hydroxyl group accepts a hydrogen bond from the N101 amide and donates a hydrogen bond to the backbone carbonyl oxygen of N99, while the C4 hydroxyl group hydrogen bonds with the backbone carbonyl oxygen of R24. In domain B, the C3 hydroxyl group of the glucosyl ring is involved in a water mediated hydrogen bond with the backbone amide of Q53 and the side chain carbonyl oxygen of N54, while the C4 hydroxyl group hydrogen bonds with the backbone amide of N43 and the backbone carbonyl oxygen of Q53. On the fructosyl ring, the C3 hydroxyl group forms hydrogen bonds with the backbone amide of D45 and the backbone carbonyl oxygen of N43, while the C4 hydroxyl group hydrogen bonds with the backbone carbonyl oxygen of R81. Interestingly, R24, N99, and N101 in domain A are the equivalent residues to R81, N43 and D45 in domain B. Therefore, these conserved polar residues appear to play important roles in carbohydrate binding.

Although the overall theme of predominantly using C3- and C4-attached hydroxyl groups on either ring for hydrogen bonding with the protein is common to the interaction of sucrose with LKAMG in both domains, several differences in the hydrogen bond networks are present. In domain A, the C2-attached hydroxyl group of the glucosyl ring makes a hydrogen bond with the backbone carbonyl oxygen of N99 and a water mediated hydrogen bond with the side chain carbonyl oxygen of Q107, while the C6-attached hydroxyl groups of the glucosyl and fructosyl rings are hydrogen bonding to each other. In contrast, sucrose that is bound in domain B uses both C6-attached hydroxyls for protein binding. The C6-hydroxyl on the glucosyl ring hydrogen bonds to the amide N43 side chain and, through a water molecule, to the backbone carbonyl oxygen of F55 and the backbone amide of T58. On the fructosyl ring, the C6-hydroxyl group hydrogen bonds with the amide nitrogen of the R81 side chain and, through a water molecule, with the side chain hydroxyl group of T58 and the side chain oxygen of E57. Given the conformational variability of the bound sucrose in our present structures, it seems likely that even more conformations around the glycosidic linkage may be possible upon interacting with proteins, even when the essential C3- and C4-attached hydroxyl groups are engaging in similar interactions. In particular, it may well be the case that multiple conformations are present in solution and that upon crystallization a predominant conformation is selected. This clearly could be the case for the sucrose in domain A.

In order to assess the conformational selection from all possible sucrose structures, we compared the two bound sucrose conformations present in our complexes to other bound or free sucrose structures available in structural databases (Figure 5). For protein-bound sucrose, we selected crystal structures of sugar complexes from the RCSB PDB with resolution better or equal to 1.80 Å31. This yielded a total of 35 structures, 11 of which contained more than one sucrose molecules as ligand. Thus, a total of 48 bound sugar molecules were available. For the free sucrose, 7 different sucrose structures from the Cambridge Structural Database (CSD) were kindly provided by Tracy Allgood (personal communication).

All bound sucrose conformations extracted from the PDB database are similar with Φ and Ψ torsion angles clustered around 103° and -45°, respectively (Figure 5A), although four notable exceptions are seen: 2GNX, 2HHSa and b, and 1IW0 clearly lie outside this clustered region. The 7 free sucrose structures exhibit almost identical conformations with torsion angle values of Φ = 108° and Ψ = -45° (Figure 5B). The bound sucrose in domain B of LKAMG, both in monomer A and B (Figure 5C), is structurally very similar to the major cluster in the protein-bound crystal structures and free sucrose, while the one in domain A clearly is different (Figure 5C). Interestingly, it is similar to the sucrose conformation reported for 1IW0, where sucrose is not a known ligand of the protein, but was introduced in the crystallization mix. Therefore, the conformation observed in domain A may represent a lower affinity structure than the one in domain B, consistent with the lack of hydrogen bonds that are formed by the C6-attached hydroxyls of both glucosyl and fructosyl rings in domain A compared to domain B. The missing hydrogen bond by the C6-attached hydroxyl is substituted by a hydrogen bond involving the C2-attached hydroxyl in the glucosyl ring of sucrose when bound in domain A of LKAMG. The above crystallographic findings agree well with the affinity data derived from NMR titration experiments.

In summary, only very small changes in sucrose and protein conformations are noted upon complex formation, and subtle rearrangements around the glycosidic torsion angle may allow maneuvering most hydroxyl groups on the carbohydrate into place for hydrogen bond formation. Binding to shallow, surface exposed sites on proteins with mobile polar side chains may allow lectins to select conformations of the sugar ligands around their stable

energy minimum, providing versatility without sacrificing specificity, albeit at the expense of affinity. This general principle may apply to most carbohydrate-lectin interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. In-Ja Byeon for expert help with NMR experiments and fruitful discussions, M. Delk for NMR technical support and Dr. Laurel Lagenaur for expertise and help with HIV assays. The plasmid encoding NcCVNH was kindly provided by Simone Ottonello. This work was supported by a National Institutes of Health Grant GM080642 (to A.M.G.).

References

1. Voet, DJ.; Voet, JG. Biochemistry. New York: J. Wiley & Sons; 2003.
2. Sharon N, Lis H. The structural basis for carbohydrate recognition by lectins. Adv Exp Med Biol. 2001; 491:1–16. [PubMed: 14533786]
3. Bouckaert J, Hamelryck T, Wyns L, Loris R. Novel structures of plant lectins and their complexes with carbohydrates. Curr Opin Struct Biol. 1999; 9(5):572–577. [PubMed: 10508764]
4. Koharudin LMI, Viscomi AR, Jee JG, Ottonello S, Gronenborn AM. The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. Structure. 2008; 16(4): 570–584. [PubMed: 18400178]
5. Percudani R, Montanini B, Ottonello S. The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. Proteins. 2005; 60(4):670–678. [PubMed: 16003744]
6. Bewley CA. Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. Structure. 2001; 9(10):931–940. [PubMed: 11591348]
7. Botos I, O'Keefe BR, Shenoy SR, Cartner LK, Ratner DM, Seeberger PH, Boyd MR, Wlodawer A. Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. J Biol Chem. 2002; 277(37):34336–34342. [PubMed: 12110688]
8. Barrientos LG, Matei E, Lasala F, Delgado R, Gronenborn AM. Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. Protein Eng Des Sel. 2006; 19(12):525–535. [PubMed: 17012344]
9. Shenoy SR, Barrientos LG, Ratner DM, O'Keefe BR, Seeberger PH, Gronenborn AM, Boyd MR. Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. Chem Biol. 2002; 9(10):1109–1118. [PubMed: 12401495]
10. Clore GM, Gronenborn AM. Determining the structures of large proteins and protein complexes by NMR. Trends Biotechnol. 1998; 16(1):22–34. [PubMed: 9470228]
11. Bax A, Grzesiek S. Methodological advances in protein NMR. Accounts of Chemical Research. 1993; 26(4):131–138.
12. Sattler M, Maurer M, Schleucher J, Griesinger C. A Simultaneous ^{15}N, ^{1}H-HSQC and ^{13}C, ^{1}H-HSQC with sensitivity enhancement and a heteronuclear gradient-echo. Journal of Biomolecular NMR. 1995; 5(1):97–102.
13. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. Journal of Biomolecular NMR. 1995; 6(3):277–293. [PubMed: 8520220]
14. Johnson BA, Blevins RA. NMRView - a computer-program for the visualization and analysis of NMR data. Journal of Biomolecular NMR. 1994; 4(5):603–614.
15. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. Journal of Molecular Biology. 1997; 273(1):283–298. [PubMed: 9367762]

16. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of Biomolecular NMR. 1999; 13(3):289–302. [PubMed: 10212987]

17. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallographica Section D-Biological Crystallography. 1998; 54:905–921.

18. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M. Refinement of protein structures in explicit solvent. Proteins-Structure Function and Genetics. 2003; 50(3):496–506.

19. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007; 35(Web Server issue):W375–383. [PubMed: 17452350]

20. DeLano, WL. The PyMOL Molecular Graphics System. 2002. http://www.pymol.org

21. Pflugrath JW. The finer things in X-ray diffraction data collection. Acta Crystallogr D Biol Crystallogr. 1999; 55(Pt 10):1718–1725. [PubMed: 10531521]

22. Collaborative computational project N. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50:760–763. [PubMed: 15299374]

23. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr. 2007; 63(Pt 1):32–41. [PubMed: 17164524]

24. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60(Pt 12 Pt 1):2126–2132. [PubMed: 15572765]

25. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997; 53(Pt 3):240–255. [PubMed: 15299926]

26. Bewley CA, Gustafson KR, Boyd MR, Covell DG, Bax A, Clore GM, Gronenborn AM. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. Nat Struct Biol. 1998; 5(7):571–578. [PubMed: 9665171]

27. Yang F, Bewley CA, Louis JM, Gustafson KR, Boyd MR, Gronenborn AM, Clore GM, Wlodawer A. Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. J Mol Biol. 1999; 288(3):403–412. [PubMed: 10329150]

28. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998; 54(Pt 5):905–921. [PubMed: 9757107]

29. Barrientos LG, Gronenborn AM. The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. Mini Rev Med Chem. 2005; 5(1): 21–31. [PubMed: 15638789]

30. Matei E, Furey W, Gronenborn AM. Solution and crystal structures of a sugar binding site mutant of cyanovirin-N: no evidence of domain swapping. Structure. 2008; 16(8):1183–1194. [PubMed: 18682220]

31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1):235–242. [PubMed: 10592235]

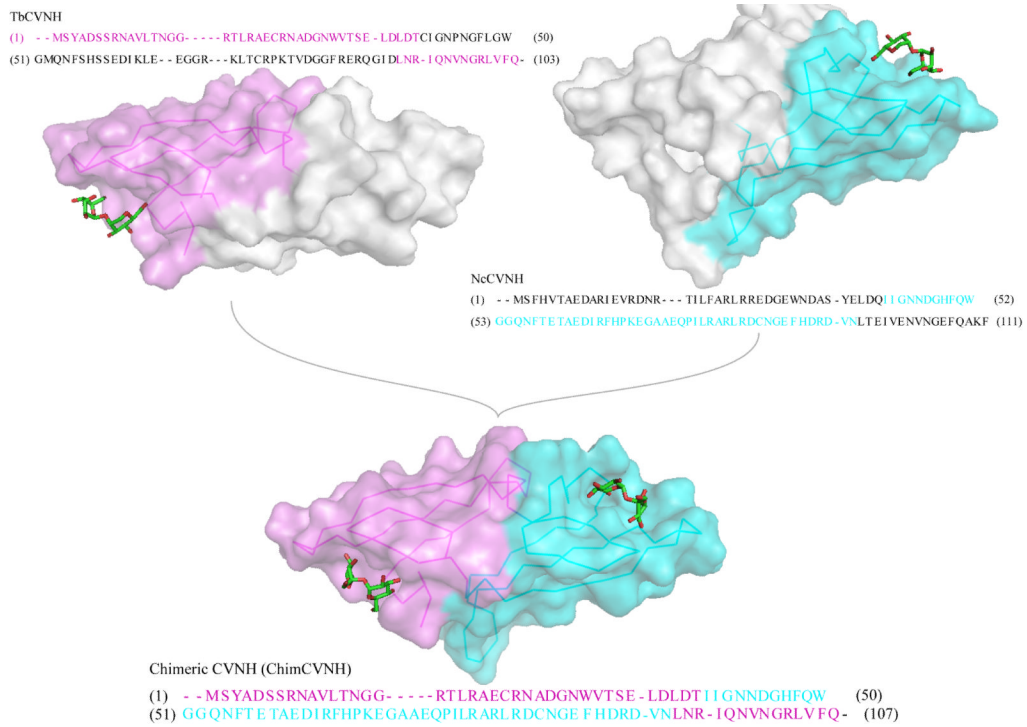

Figure 1. Design of LKAMG

The chimeric protein was created by combining the pseudo-symetric halves of TbCVNH and NcCVNH and possesses the CVNH fold. TbCVNH contains a single sugar-binding site on domain A (magenta) and NcCVNH binds sugar in domain B only (cyan). The amino acid sequence of LKAMG comprises residues 1 to 39 and residues 90 to 103 of domain A of TbCVNH and residues 42 to 95 of domain B of NcCVNH.

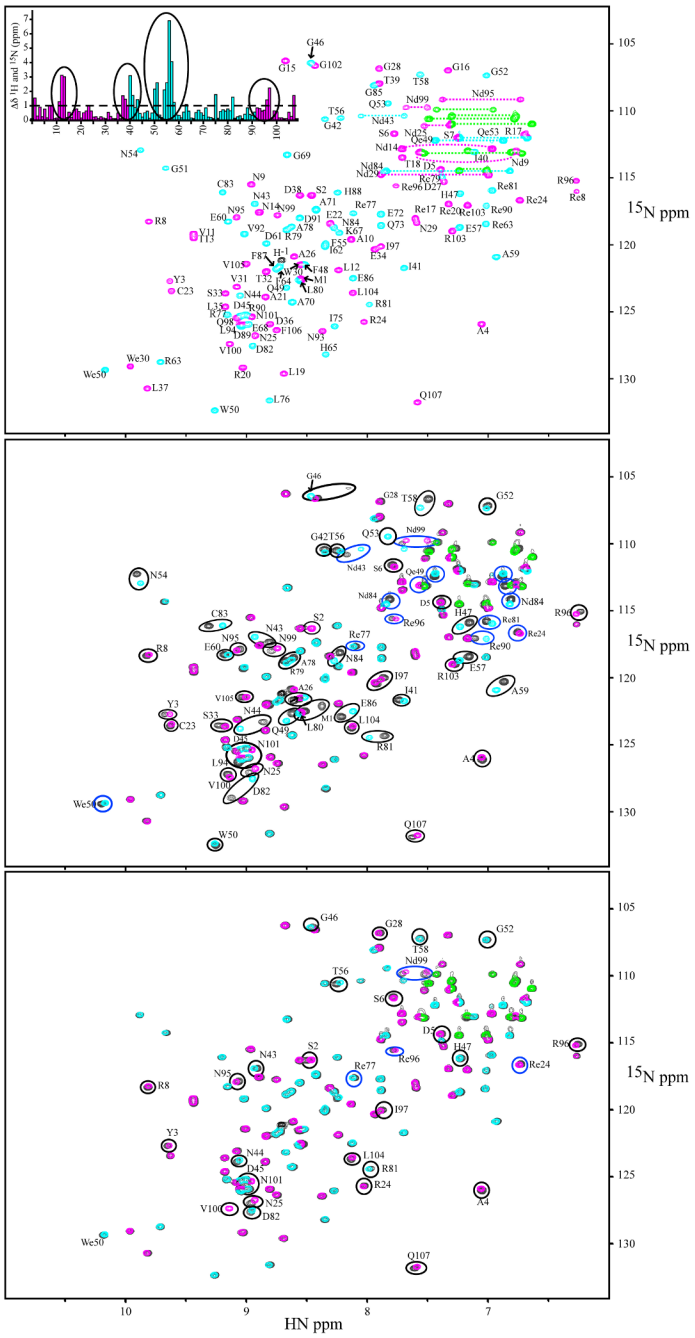

Figure 2. ^{1}H-^{15}N HSQC NMR spectra and carbohydrate binding of LKAMG

(A) ^{1}H-^{15}N 2D HSQC spectrum of LKAMG in 20 mM NaPhosphate, 0.02% NaN_3 (pH 6.00) at 298 K (top panel). Resonances corresponding to residues originating from TbCVNH and NcCVNH are displayed in magenta and cyan, respectively. (B) and (C) Sugar titrations of LKAMG with Man2α (B) and sucrose (C). Affected amino acids, color coded as in (A), are encircled and labeled by residue name and number. Ligand free protein resonances are shown in black and side chain amino groups of Asn and Gln residues are colored green. Amide chemical shift differences between the chimeric and parental proteins, plotted along the polypeptide chain, are provided in the insert (top panel) calculated based on a formula of $\Delta\delta\ (ppm) = \sqrt{[(\Delta\delta(^{1}H_N)^2 + (\Delta\delta(^{15}N)^2/5)]}$.

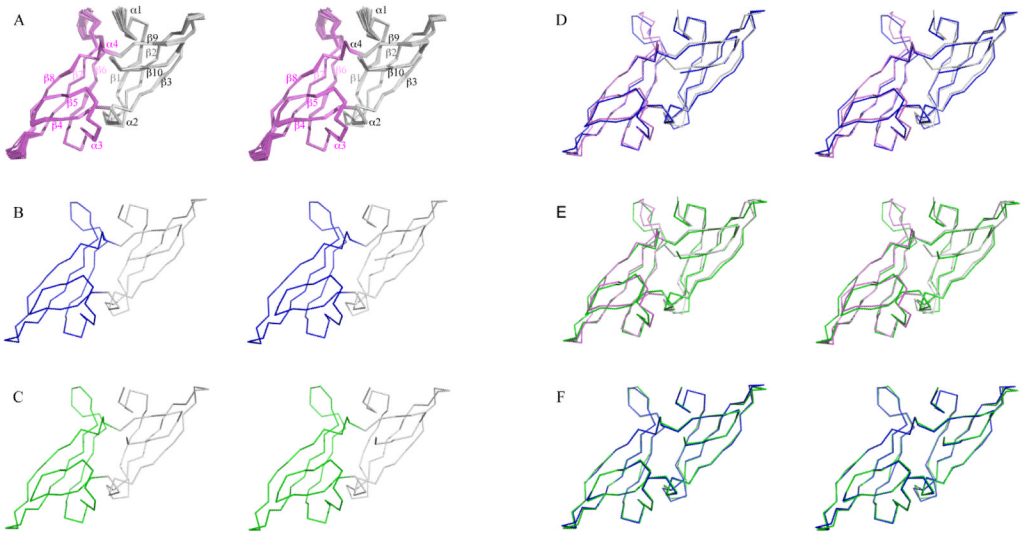

Figure 3. NMR and X-ray structures of LKAMG

(A) Stereo view of the 30 conformer ensemble (Cα representation) determined by NMR. Secondary structure elements are labeled only in (A) but apply throughout the figure. (B) and (C) Stereo views of the $P2_1$ and $P2_12_12_1$ X-ray models, respectively. Domain A is shown in grey throughout and domain B is colored magenta, blue, and green for the NMR ensemble, the $P2_1$ and the $P2_12_12_1$ X-ray models, respectively. (D), (E) and (F) Comparison between the NMR and X-ray models. (D) Stereo views of the superposition between the lowest energy structure in the NMR ensemble (magenta) and the $P2_1$ X-ray model (blue) and (E) the $P2_12_12_1$ X-ray model (green). (F) Stereo views of the best-fit superposition of the two X-ray models. No significant differences are observed between the solution and crystal structures, although local details in loop conformations induced by crystal packing are present, especially for the loops that connect strands β2 and β3, strands β6 and β7, and strands β7 and β8.

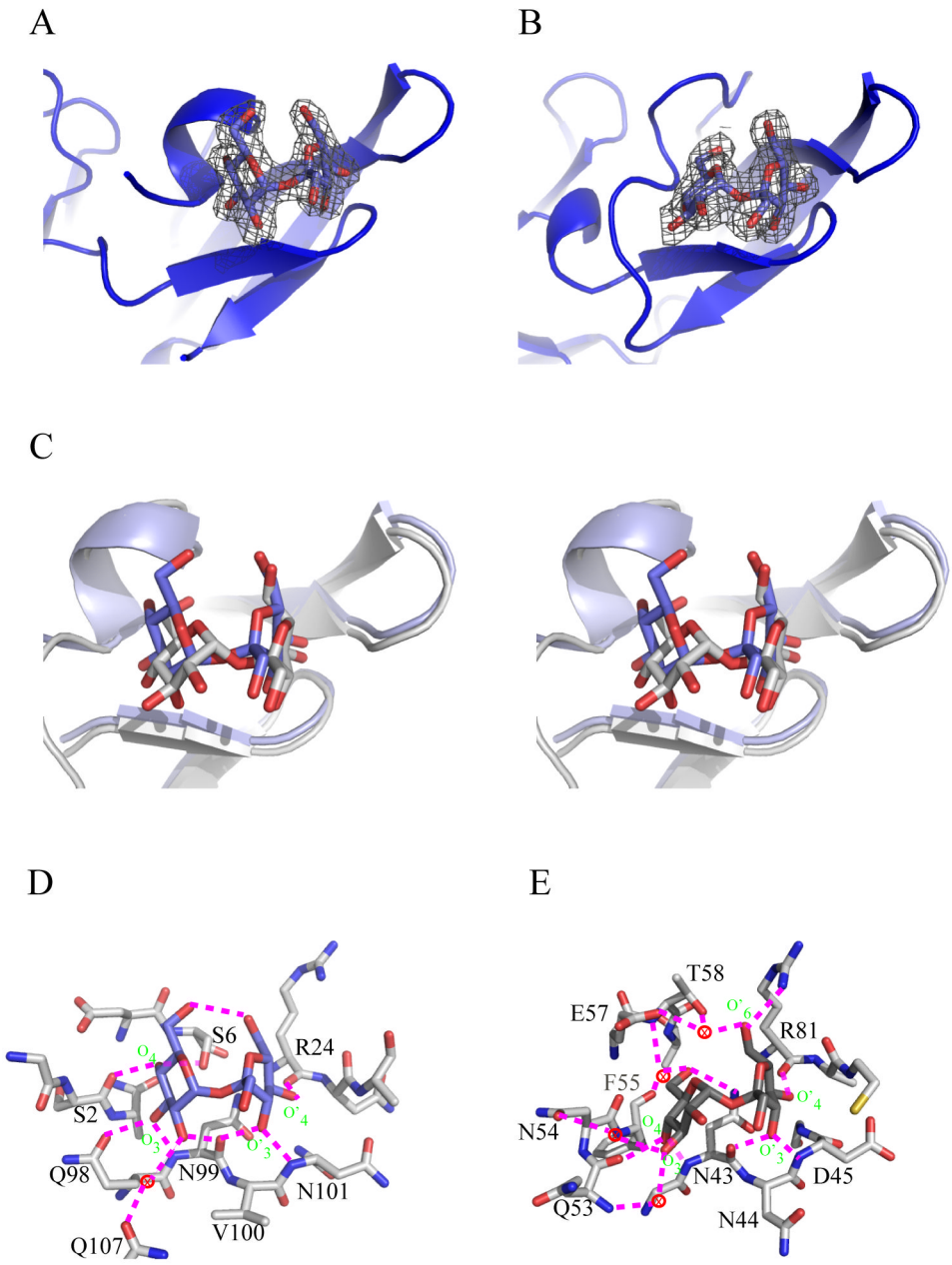

Figure 4. Sucrose conformations in the LKAMG-sucrose complex

(A) and (B) Electron density with the fitted sucrose molecules and (C) stereo view of the superposition of the two different sucrose moieties in domains A and B (in monomer A). The sugar molecules are shown in stick representation and the protein as ribbon. (D) and (E) Hydrogen bond network between sucrose and the protein in domains A and B, respectively.

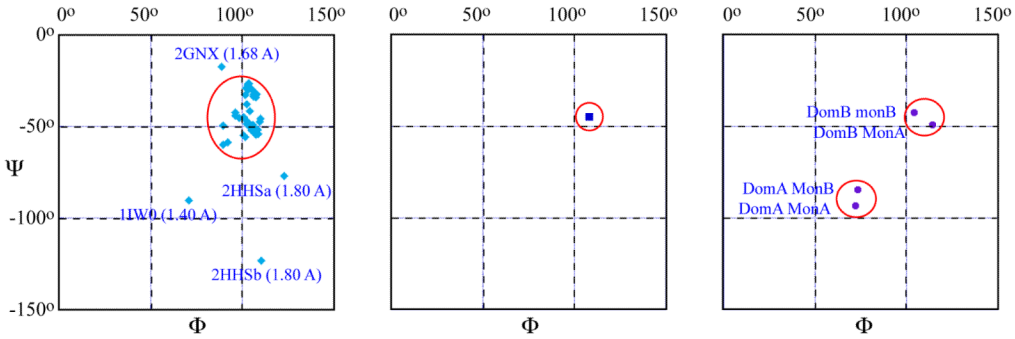

Figure 5. Ψ, Φ distribution of the sucrose glycosidic linkage for lectin-bound and free sucrose

The Ψ- angle is defined by $C1_g$-$O1_g$-$C2_f$-$O5_f$ and the Φ-angle by $O5_g$-$C1_g$-$O1_g$-$C2_f$ around the glycosidic linkage. (A) Values were extracted for protein-bound structures deposited in the RCSB PDB database for models that diffracted to 1.80 Å or better and contained at least one sucrose molecule as ligand. The subscripts g and f refer to atoms in the glucosyl and fructosyl rings, respectively. The four outliers of the cluster are labeled with their PDB accession codes and the resolution of the structures. (B) Conformations for seven free sucrose molecules were obtained from the CSD database. (C) Sucrose conformations in the present complex structure. Conformations for the sugars in domains A and B of both monomers are shown.

Table 1
NMR and refinement statistics for LKAMG

	107aa 30 structures*
NMR distance & dihedral constraints	
Distance constraints	
Total NOE	2650
Intra residue (|i-j| = 0)	467
Inter residue	
Sequential (|i-j| = 1)	650
Medium-range (1 <|i-j| < 5)	411
Long range (|i-j > 5)	1122
Hydrogen bonds	106
Dihedral angle constraints	
phi	67
psi	68
Structure Statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.024 ± 0.001
Dihedral angle constraints (°)	0.662 ± 0.066
Max. dihedral angle violation (°)	5
Max. distance constraint violation (Å)	0.5
Deviations from idealized geometry	
Bond lengths (Å)	0.0045 ± 0.0001
Bond angles (°)	0.518 ± 0.011
Impropers (°)	0.476 ± 0.017
Mean r.m.s.d.** (Å)	
Heavy atoms	0.61 ± 0.04
Backbone atoms	0.23 ± 0.04

* 800 structures were calculated and 30 structures with the lowest CNS energies were refined in explicit water.

** r.m.s.d. values were calculated for residues 2 - 105.

Table 2
Data collection and refinement statistics (molecular replacement) for LKAMG

	$P2_1$ Free[a]	$P2_12_12_1$ Free[a]	$P2_1$ Bound[a]
Data collection			
Space group	$P2_1$	$P2_12_12_1$	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	33.88, 34.44, 45.63	45.51, 46.20, 50.38	44.50, 39.30, 86.13
α, β, γ (°)	90.00, 101.24, 90.00	90.00, 90.00,90.00	90.00, 97.63, 90.00
Resolution (Å)	24.49-1.56 (1.62-1.56) [b]	27.26-1.35 (1.40-1.35) [b]	35.70-1.88 (1.95-1.88) [b]
R_{merge}	0.047 (0.175)	0.039 (0.337)	0.166 (0.413)
$\langle I / \sigma I \rangle$	19.9 (3.2)	40.7 (2.7)	3.9 (1.4)
Completeness (%)	95.5 (71.4)	97.7 (77.3)	96.7 (76.1)
<Redundancy>	3.90 (2.21)	13.32 (2.94)	3.08 (2.92)
Refinement			
Resolution (Å)	1.56	1.37	2.00
No. reflections	12,776	21,673	19,030
R_{work} / R_{free}	0.183 / 0.225 [c]	0.168 / 0.210 [c]	0.224 / 0.277 [c]
No. atoms			
Protein	864	932	1717
Ligand (Sucrose)	0	0	92
Ligand (DTT)	0	0	16
Water	169	154	244
B-factors			
Protein	15.88	16.12	29.26
Ligand (Sucrose)	0	0	29.99
Ligand (DTT)	0	0	53.15
Water	28.08	29.51	37.64
R.m.s. deviations			
Bond lengths (Å)	0.009	0.007	0.012
Bond angles (°)	1.141	1.090	1.596

[a] Each diffraction dataset was obtained from the best-diffracting single crystal.

[b] Values in parentheses are for highest-resolution shell.

[c] Refinements were carried out using the program Refmac5.