Glycan-Cyanovirin-N Interactions and Designed WW Domains: Combining Experimental and Computational Studies

by

Brian William Woodrum

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2014 by the
Graduate Supervisory Committee

Giovanna Ghirlanda, Chair
Kevin Redding
Xu Wang

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

Cyanovirin-N (CVN) is a cyanobacterial lectin with potent anti-HIV activity, mediated by binding to the N-linked oligosaccharide moiety of the envelope protein gp120. CVN offers a scaffold to develop multivalent carbohydrate-binding proteins with tunable specificities and affinities. I present here biophysical calculations completed on a monomeric-stabilized mutant of cyanovirin-N, P51G-m4-CVN, in which domain A binding activity is abolished by four mutations; with comparisons made to CVN^{mutDB}, in which domain B binding activity is abolished. Using Monte Carlo calculations and docking simulations, mutations in CVN^{mutDB} were considered singularly, and the mutations E41A/G and T57A were found to impact the affinity towards dimannose the greatest. ^{15}N-labeled proteins were titrated with Manα(1-2)Manα, while following chemical shift perturbations in NMR spectra. The mutants, E41A/G and T57A, had a larger K_d than P51G-m4-CVN, matching the trends predicted by the calculations. We also observed that the N42A mutation affects the local fold of the binding pocket, thus removing all binding to dimannose. Characterization of the mutant N53S showed similar binding affinity to P51G-m4-CVN. Using biophysical calculations allows us to study future iterations of models to explore affinities and specificities. In order to further elucidate the role of multivalency, I report here a designed covalent dimer of CVN, Nested cyanovirin-N (Nested CVN), which has four binding sites. Nested CVN was found to have comparable binding affinity to gp120 and antiviral activity to wt CVN. These results demonstrate the ability to create a multivalent, covalent dimer that has comparable results to that of wt CVN.

WW domains are small modules consisting of 32-40 amino acids that recognize proline-rich peptides and are found in many signaling pathways. We use WW domain sequences to explore protein folding by simulations using Zipping and Assembly Method. We identified five crucial contacts that enabled us to predict the folding of WW domain sequences based on those contacts. We then designed a folded WW domain peptide from an unfolded WW domain sequence by introducing native contacts at those critical positions.

ACKNOWLEDGEMENTS

I would like to foremost thank my advisor Giovanna Ghirlanda, for without whose support this dissertation would not be possible. Secondly, I would like to thank my collaborators, Dr. S. Banu Ozkan, Dr. Taisong Zou, and Ashini Bolia; whose computational work was invaluable to the work presented here. I would also like to thank my committee members for their support: Dr. Kevin Redding for insightful conversations and guidance in this work and Dr. Xu Wang for his incredible guidance in NMR.

Many of the Ghirlanda group members have helped me in various aspects of this research, and I wish to thank them for their contributions. Dr. Haiyan Sun and Dr. Angelo Cereda were instrumental in everything I've learned about molecular biology. A special thanks to Dr. Haiyan Sun for her work on the periplasmic expression of CVN and her passing that knowledge onto me. I would like to thank Jason Maxwell for all his help, both with benchwork and insightful conversations. Dr. Sandip Shinde, who was kind enough to let me work on some projects with him early on, deserves my praise for many of the scientific and career discussions we've had. Dr. Anindya Roy has been a remarkable amount of help, as we entered Dr. Ghirlanda's group at the same time and went through the trials and tribulations of graduate school together. I would also like to thank Dayn Sommer for continually helping with and his willingness to take over all the instrumentation I was responsible for. I would also like to thank James Jursich and, although not a member of my group, Dr. Sara Bowen, for their continued support throughout my graduate experience. I'd also like to thank and acknowledge Dr. Christine

Pruis for all her support during my teaching assistantships and her influence on my teaching philosophy.

I would finally like to thank all the facilities and centers that have been vital to my research. The Proteomics and Protein Chemistry Facility at ASU has been instrumental in providing research support for this project. I'd specifically like to thank John Lopez and Zachary Laughrey for their help with circular dichroism and MALDI-TOF. I'd like to thank Brian Cherry of the ASU MRRC for always being willing to help with the NMR. Andrey A. Bobkov from Sanford-Burnham Medical Research Institute has been very helpful over the years and has run all the ITC presented here. Barry O'Keefe has also been very generous, performing the antiviral activity assays in his laboratory at the National Cancer Institute. The following reagent was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 gp120 CM, Cat # 2968, from DAIDS, NIAID. This work was supported by NSF-MCB award 11212762 (GG), by NIH-NIGMS U54GM094599 (P. Fromme), and by an ASU CLAS Seed Grant.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

1. Cyanovirin-N

1.1. Background on HIV

Human immunodeficiency virus (HIV) is a retrovirus that can lead to the development of acquired immunodeficiency syndrome (AIDS). Bodily fluids transmit HIV during sexual contact or blood contact. HIV can also be transmitted from an infected mother to her baby at birth. According to the Center for Disease Control (CDC), an estimated 1,106,400 people were living in the United States with HIV at the end of 2006.[1] Worldwide, UNAIDS estimated that over 33.4 million people were living with HIV.[2] In 2008, the CDC estimated that nearly 56,300 people in the United States were newly infected with HIV.[3] With so many people being newly infected with AIDS each year, and the number is quite higher for worldwide new infections, it is vitally important to find new drug therapies that will inhibit the transmission of HIV. As discussed below, the treatment of HIV is extremely difficult because of the virus' ability to become drug resistance very rapidly.

It is important to understand the life cycle of HIV in order to find targets for drug therapy. HIV attaches itself to a host cell, infects the host cell with its RNA, and then uses reverse transcriptase to transcribe the RNA into DNA; and then integrates the DNA into the host cell's DNA so that the host cell will translate the viral parts needed for viral reproduction. Through the life cycle, there are many targets at which antiviral therapy can take place to inhibit HIV (Figure 1).

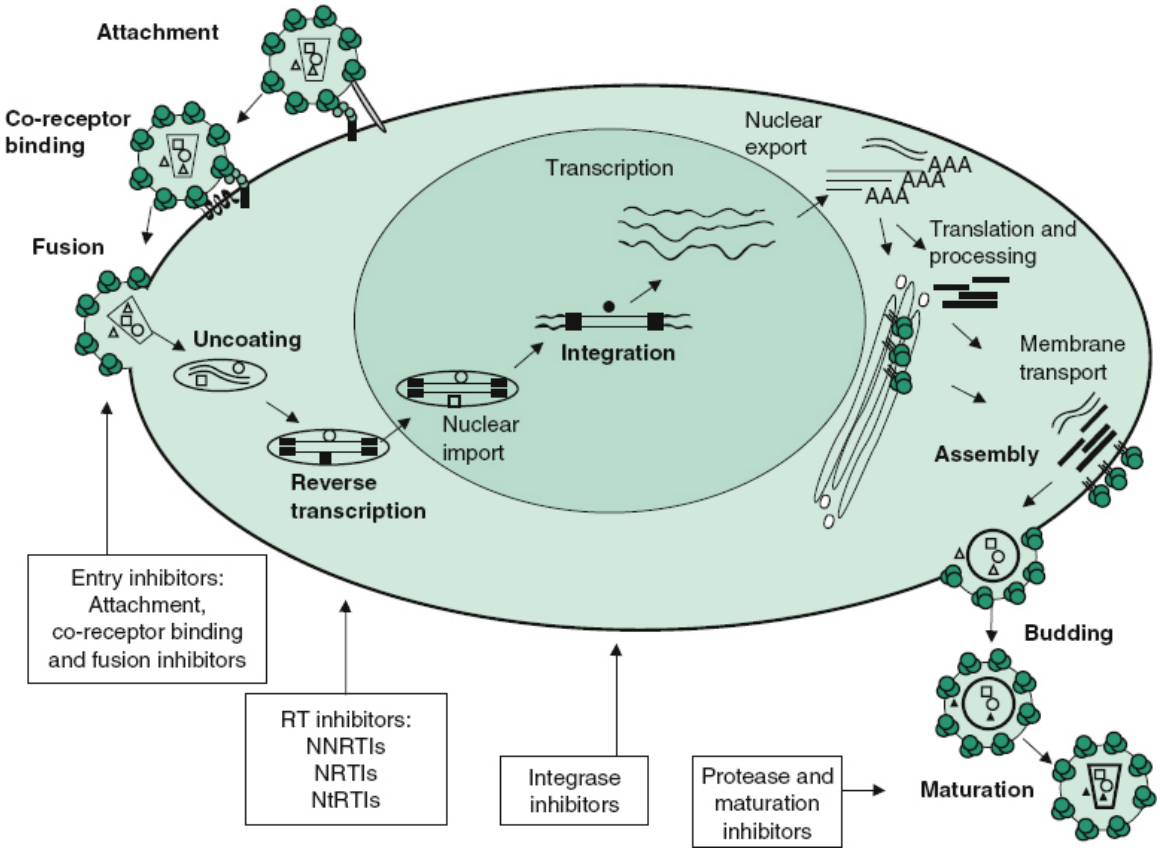


Figure 1: The life cycle of HIV. Current targets for approved therapies are cell fusion, reverse transcriptase, and maturation. New targets include attachment, receptor binding, and integration.[4]

Current therapies for treatment of HIV are classified as anti-retroviral therapy (ART) and highly active anti-retroviral therapy (HAART). A popular, early therapy that was pursued for the treatment of HIV was reverse transcriptase (RT) inhibitors. Many derivations of RT inhibitors have been identified to date, including nucleoside RT inhibitors (NRTIs), nonnucleoside RT inhibitors (NNRTIs), and nucleotide RT inhibitors (NtRTI).[5] However, error-prone reverse transcriptase and sheer number of replication cycles have led to the selection of drug resistant mutants of HIV.[5] Because of this, many

other HIV therapy options have been pursued. These include integrase inhibitors and protease/maturation inhibitors. Protease inhibitors also have problems overcoming resistance, and the genetic basis for this has been well documented.[6] A major problem with both ART and HAART are that these therapies do not cure the disease and mutants with resistance will arise.[7]

A recent target for therapy has been disrupting the cell fusion of the HIV virus and the host cell. Two such drugs being used are enfuvirtide and tifuvirtide.[4] There are many steps in attachment and fusion with the host cell to target, however, shown in Figure 2.

The envelope protein (Env) of HIV mediates recognition of the host cell and then fusion to the host cell. Env is comprised of gp120 and a transmembrane protein, gp41; which are formed after the cleavage of gp160 and then form a trimer of heterodimers. As shown in Figure 2, gp120 forms nonspecific interactions with the target cell membrane. This allows specific interactions to occur with the host cell receptor, CD4, which then is believed to induce conformational changes in gp120. Depending on the tropism of the virus, a co-receptor, either CCR5 or CXCR4, interacts with Env, most likely causing further conformational changes in Env. This last conformational change is believed to result in the insertion of the gp41 helix bundle into the target cell membrane. Most likely, more than one Env trimer is involved in mediating the membrane fusion, as depicted in Figure 2.[4]

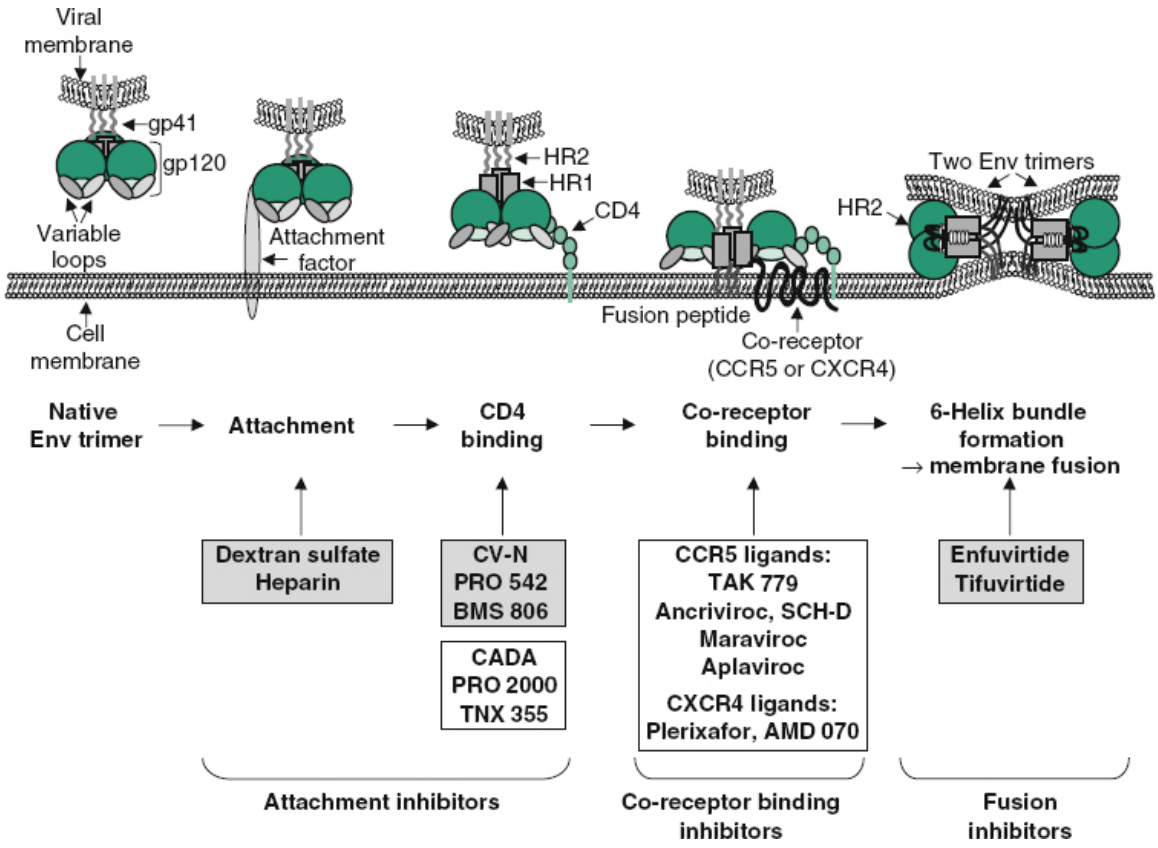


Figure 2. The process in which HIV recognizes and fuses with a host cell presents many opportunities to inhibit infection.[4]

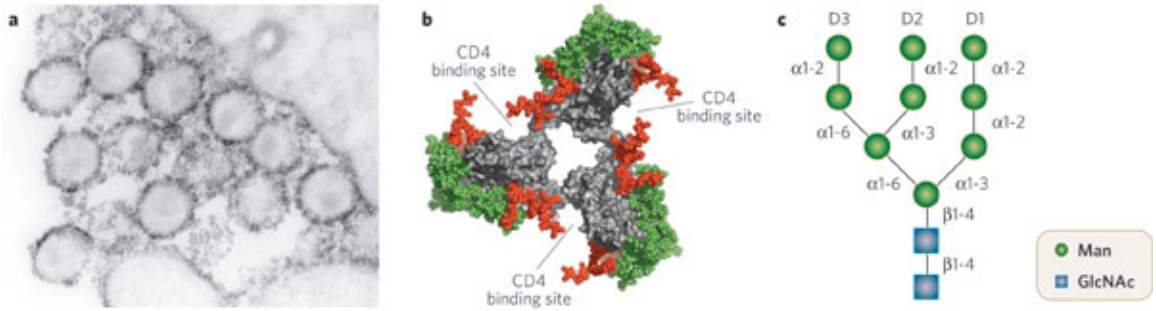


Figure 3. Electron micrograph of HIV-1 (A). The carbohydrates are stained (dark), showing extensive glycosylation concealing the antigenic surface. Molecular model view of gp120 trimer, with glycosylated outer domain (green), and complex sugars (red) that line the CD4 binding sites (B). Linkages of the oligomannose glycan array shown (C).[8]

The heavily glycoslyated protein gp120 plays a large role in HIV's ability to escape the human immune system and in recognition of chemokine co-receptors for host cell binding. HIV is able to escape detection by the immune system because the surface antigens of Env are glycosylated by host enzymes using "immunologically 'self' glycans," or glycans synthesized by the host. Therefore, there is not enough available antigenic surface, and the host's immune system will normally not form antibodies against HIV. Figure 3b shows a molecular model of a trimer of gp120, with Figure 3a showing a schematic of the oligomannose glycan array on gp120. Targeting the glycan array of gp120 removes the chance for HIV to mutate to become resistant to other drugs, because the target is now the glycosylation that HIV must use to escape detection from the immune system.

1.2. Introduction to Cyanovirin-N

1.2.1. Discovery of cyanovirin-N and antiviral properties

Cyanovirin-N (CVN) is an 11-kDa protein that was isolated from the cyanobacterium *Nostoc ellipsosporum*.[9] It is comprised of 10 1 amino acids and is able to be expressed recombinantly in *Escherichia coli*. CVN was discovered on the basis it was found in an extract of cyanobacteria that it exhibited potent inhibition of HIV cytopathy and it was thought to be proteinaceous, and there had not been any evidence of antiviral proteins from cyanobacteria to date. Further studies had shown that CVN interacted with gp120, part of the viral surface envelope glycoprotein.[9,10] CVN was found to have antiviral activity in the sub-nanomolar range. This is extremely potent and it was also encouraging to find that CVN conveyed low cytoxicity to the host cells.

In order to elucidate the structural determinants of CVN that are responsible for the antiviral activity, sequence and numerous structural studies have been performed. Sequence studies have shown that CVN is a linear sequence of 101 amino acids that has no post-translational modifications and 4 cysteine residues that form intramolecular disulfide bonds.[11] CVN was also determined to have two domains in the sequence, that when compared, showed 32% identity ad 26% conservative changes for an overall homology of 58%.[12] It was also found that degradation of the N-terminal and the C-terminal, separately, resulted in excess of 9-fold decrease in potency.[12]

The solution structure of CVN showed that the protein consisted of two symmetric units: domain A that is comprised of amino acids 1-39 and 91-101, and domain B that is comprised of amino acids 40-90.[13] Each of these domains has a binding site for the

Manα(1-2)Manα linkages on the end of gp120. This quaternary structure of CVN also shows a novel fold in which the two domains are pseudosymmetric to each other. Each domain has a triple-stranded anti-parallel β-sheet and a β-hairpin, and are joined by two helical turns.[13]

Unexpectedly, when CVN was crystallized, the structure was found to be a domain-swapped dimer. When crystallized, the protein had domains that were swapped between each symmetrically related monomer. That is such that domain A of one monomer was associated with domain B of the other monomer, and vice versa, giving overall dimerization as A-B' and A'-B.[14] Figure 4 shows the solution structure of the monomer and the x-ray crystal structure of the dimer. As evident from Figure 4, the dimer is extremely similar in structure to the monomer; and only the hinge residues, Trp49-Phe54 that allow the domain-swapping to occur, significantly differ in ϕ and ψ torsional angles.[15] As it turns out, a small fraction of the dimeric species is present in the expression of recombinant CVN, and can be identified through size exclusion chromatography.[15] After analysis of the mechanistic and thermodynamic contributions to domain-swapping, it was determined that the dimer is a "metastable, kinetically trapped intermediate" at many pH's and at room temperature and below.[15] It is believed that as crystallization occurs, the protein-protein interactions in the formation of the crystal lattice structure facilitate the formation of the domain-swapped dimer.[15]

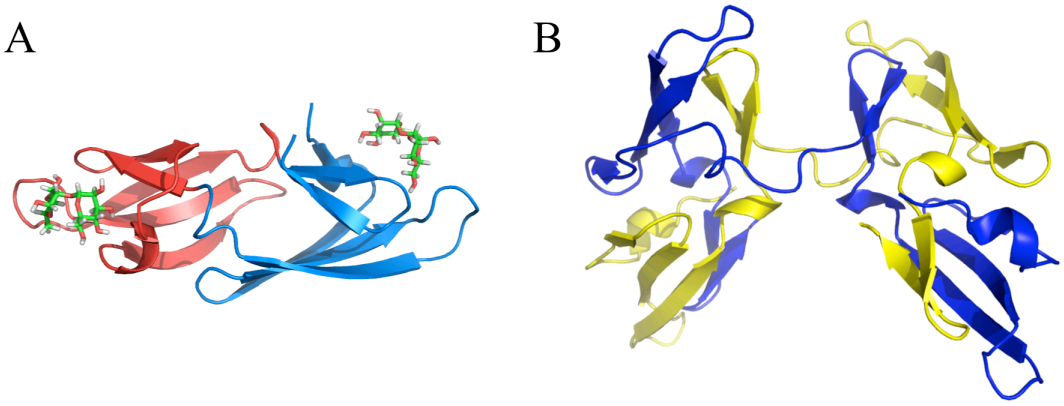


Figure 4. Solution structure of wt CVN as a monomer (A), PDB code 1IIY. Residues 1-39, 91-101 are colored blue, residues 40-90 are colored red, and dimannose is shown in each binding pocket. Domain-swapped dimer of wt CVN (B), PDB code 3EZM, shown with one monomer colored blue, the other yellow. The two parts of the dimer are nearly orthogonal to each other.

1.2.2. Mutants of Cyanovirin-N

(This section in part was originally published in Biochemical Society Transactions. Brian W. Woodrum, Jason D. Maxwell, Ashini Bolia, S. Banu Ozkan, and Giovanna Ghirlanda, The antiviral lectin cyanovirin-N: probing multivalency and glycan recognition through experimental and computational approaches. Biochemical Society Transactions. 2013; 41: 1170-1176. © the Biochemical Society)

In cyanovirin-N, multiple binding interactions with the oligomannoside target could be achieved by simultaneously engaging both domains A and B in the monomeric form, and/or by engaging two or more binding sites in the domain-swapped dimer form. This hypothesis is supported by the difference observed in binding linear oligomannosides and

glycosylated gp120. Whereas each domain, A and B, binds di- and tri-mannose independently with dissociation constants in the micromolar range, monomeric CVN binds gp120 at nanomolar concentrations, suggesting an avidity effect. Further evidence of multivalency is the agglutination resulting from the interaction of CVN at micromolar concentrations with multidentate mannosides such as Man9 and Man8.[16]

To address these issues unequivocally, our lab had previously designed a novel mutant, P51G-m4-CVN (Figure 5), which contains a single high-affinity carbohydrate-binding site (domain B), folds exclusively as a monomer under physiological conditions, and is more stable than wt (wild-type) CVN. The protein was designed by merging two sets of mutations, four in order to abolish glycan binding to domain A[17] and a hinge region mutation, P51G, which had been shown to stabilize the monomer and prevent domain swapping.[19] P51G-m4-CVN was designed to clarify whether the presence of one mannose-binding site is sufficient for antiviral activity or whether multivalent interactions, either via the two sites on each cyanovirin molecule or via the four possible sites on the domain-swapped form, are necessary. This issue had been controversial, owing to the difficulties in clearly discriminating between the possible models, and complicated further by the slight selectivity of each domain for a different linear oligomannoside.[16,17,19-22] As designed, P51G-m4-CVN contains a single functional carbohydrate-binding site (domain B) and folds exclusively as a monomer under physiological conditions; it is possible to obtain the domain-swapped form by unfolding and refolding at millimolar protein concentrations.[18] The high-resolution X-ray structure of the free and dimannose-bound complex revealed important details on the binding

mode.[18,23] Our lab has previously investigated the interaction of P51G-m4-CVN with gp120 by ELISA, revealing a decrease in apparent affinity of ~100-fold compared with wt CVN; high-affinity binding can be rescued by forming the domain-swapped dimer. This was the first evidence linking multivalency and high-affinity binding to gp120; we further tested the antiviral activity of P51G-m4-CVN, revealing loss of activity at concentrations as high as 5 μM. In contrast, wt CVN inhibits HIV with an EC_{50} of approximately 0.1 nM.

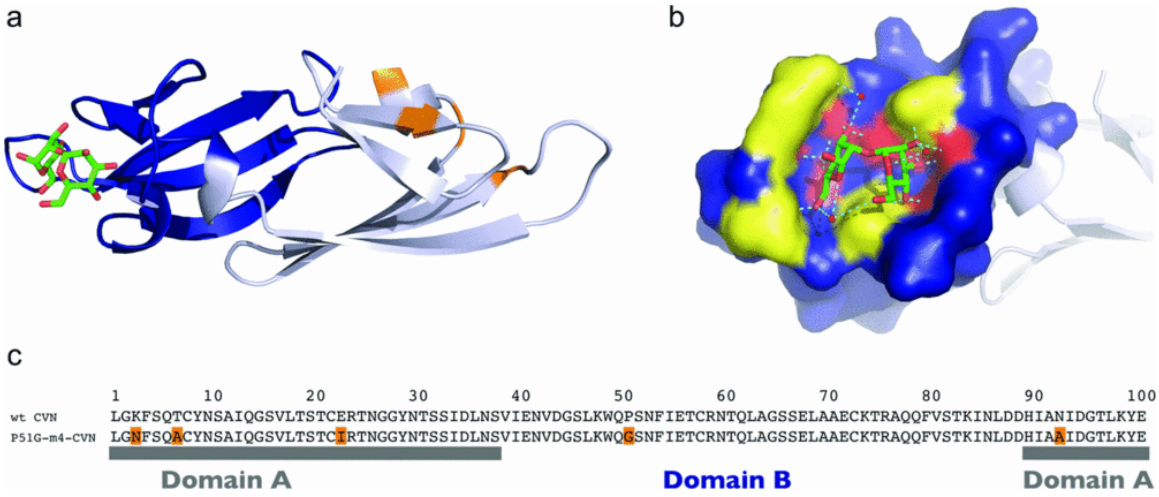


Figure 5. (a) Crystal structure of P51G-m4-CVN (PDB code 2RDK)[18] with domain B in blue and domain A in grey with m4 mutations shown in orange. Only domain B binds dimannose, shown with carbons in green and oxygens in red. (b) The hydrogen-bonding network to the sugar in domain B (blue) in 2RDK. Hydrogen bonds are cyan, with water molecules shown as red dots. The residues that participate in hydrogen-bonding through backbone atoms are highlighted in red, and the residues that participate in hydrogen-bonding through side chains have their side chains highlighted in yellow. (c) Sequence

alignment of wt CVN and P51G-m4-CVN with the mutated residues highlighted in orange.

The complete loss of antiviral activity in P51G-m4-CVN demonstrated the importance of multivalent interactions with the viral gp120, confirming one of the mechanisms that had been hypothesized for antiviral lectins, as illustrated in Figure 6.[24-27] Our conclusions were corroborated by independent studies of a complementary mutant, CVN^{mutDB}, in which the binding site on domain B had been abolished on the background of P51G: this mutant is also inactive against HIV.[21]

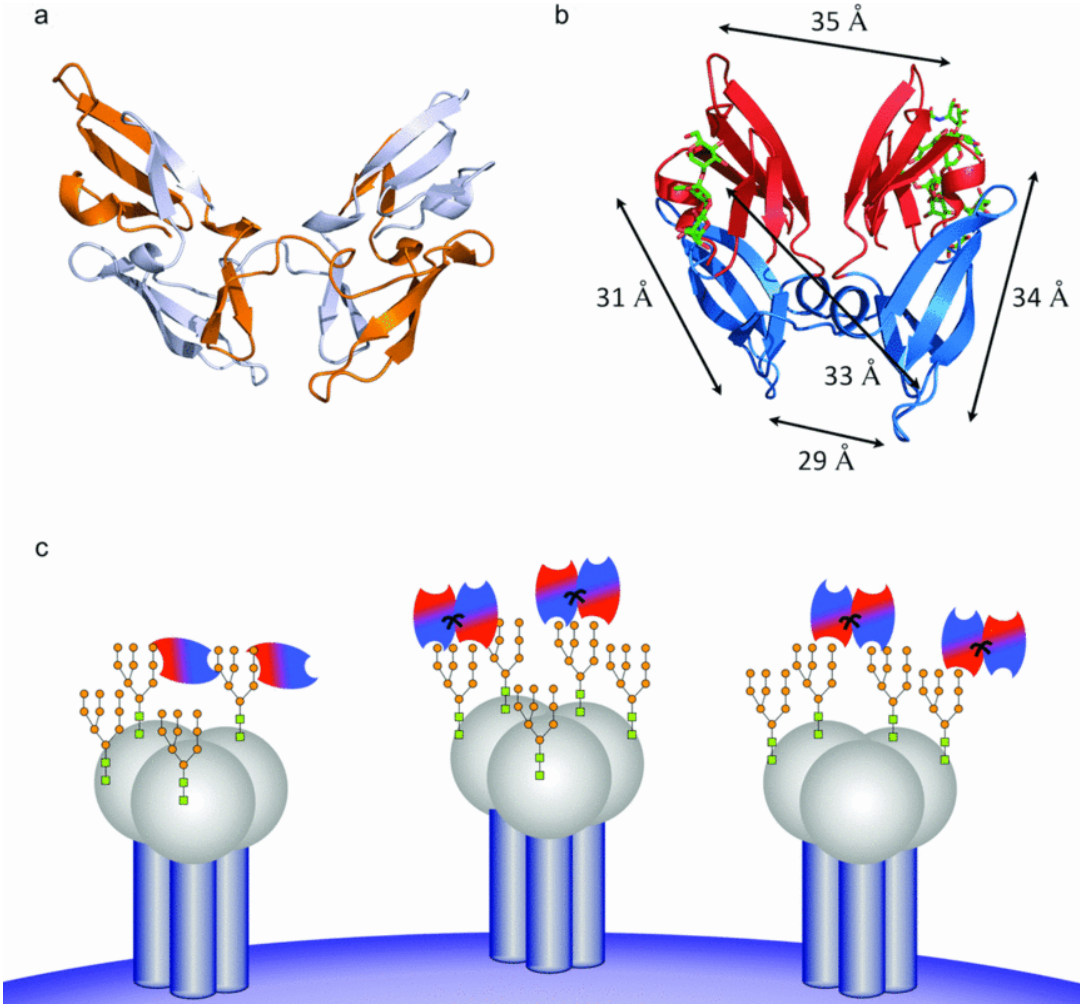


Figure 6. (a) CVN can form a domain-swapped dimer, one monomer is highlighted in orange, the other in grey (PDB code 3EZM). (b) The symmetry of binding sites in the domain-swapped dimer of wt CVN bound with oligosaccharides colored with green carbons, red oxygens and blue nitrogens (PDB code 3GXZ). The binding sites are 30–35 Å apart. (c) Cartoon representation of the possible binding modes of CVN. CVN is depicted as a monomer and a dimer (dimer shown with black hinge region), with domain A in red and domain B in blue. Man9 is represented on trimers of gp120 with GlcNAc represented as green squares and mannose depicted as orange circles.

We then explored whether constructs containing two copies of domain B could recover biological function. Mutants S52P-m4-CVN and Δ50Q-m4-CVN, which form obligate domain-swapped dimers, restore multivalency because they encompass two functional B domains, while containing four mutations that abolish binding in each domain A.[28] We found that these mutants bind gp120 at nanomolar concentrations, and display significant anti-HIV activity. Unfortunately, the hinge mutations resulted in reduced thermostability, which affected their performance in antiviral assays performed at 37˚C; ΔQ50-m4-CVN displayed an EC_{50} of 320 nM, compared with 0.9 nM observed for wt CVN. More complete restoration of activity to levels similar to that of CVN was observed in covalent dimers of monomeric constructs containing inactive domain A and linked by an intermolecular disulfide bond. In these constructs, multivalency is restored by the presence of two intact copies of domainB (BB′) in the dimer, and results in potent antiviral activity (IC_{50} of 6 nM compared with 0.7 nM for wt CVN). A related construct containing two copies of intact domain A (AA′) dimerized via domain-swapping has a somewhat diminished IC_{50} of 60 nM; it is unclear whether these differences arise from the identity of the intact domains in the dimer (BB′ compared with AA′) or from the stability of the dimers (covalent compared with non-covalent dimers).[29]

The ability of intact CVN to form domain-swapped dimers[30] and higher-order oligomers[31] suggests that these species may contribute to antiviral activity through glycan cross-linking.[27] In that case, it may be possible to improve further on the potent antiviral activity of CVN by stabilizing dimeric species, and thus augmenting its multivalency. Mayo and co-workers designed tandem repeats of CVN to create covalently linked

dimers, which form structures highly reminiscent of the domain-swapped dimers observed for wt CVN.[32] The tandem repeat dimer CVN_2L0, which contains two copies of wt CVN and thus has all four binding sites intact (A, A′, B, B′), showed improved neutralization activity compared with wt CVN; the dimer also showed extensive cross-clade activity. Obliteration of binding domains within CVN_2 demonstrated a linear correlation between number of domains and activity. Remarkably, however, no improvements in activity were observed for constructs containing three or four copies of linked CVN, suggesting that the interdomain distances in the domain-swapped dimer of CVN are optimal for binding to gp120.

1.3. Domain B Binding Pocket Study

1.3.1. Abstract

Cyanovirin-N (CVN), a cyanobacterial lectin, exemplifies a class of antiviral agents that inhibit HIV by binding to the highly glycosylated envelope protein gp120. Here, we investigate the energetics of glycan recognition using a computationally inexpensive flexible docking approach, backbone perturbation docking (BP-Dock). We benchmarked our method using two mutants of CVN: P51G-m4-CVN, which binds dimannose with high affinity through domain B, and CVN^{mutDB}, in which binding to domain B has been

abolished through mutation of five polar residues to small nonpolar side chains. We investigated the energetic contribution of these polar residues along with the additional position 53 by docking dimannose to single-point CVN mutant models. Analysis of the docking simulations indicated that the E41A/G and T57A mutations led to a significant decrease in binding energy scores due to rearrangements of the hydrogen-bond network that reverberated throughout the binding cavity. N42A decreased the binding score to a level comparable to that of CVN^{mutDB} by affecting the integrity of the local protein structure. In contrast, N53S resulted in a high binding energy score, similar to P51G-m4-CVN. Experimental characterization of the five mutants by NMR spectroscopy confirmed the binding affinity pattern predicted by BP-Dock. Despite their mostly conserved fold and stability, E41A, E41G, and T57A displayed dissociation constants in the millimolar range. N53S showed a binding constant in the low micromolar range, similar to that observed for P51G-m4-CVN. No binding was observed for N42A. Our results show that BP-Dock is a useful tool for rapidly screening the relative binding affinity pattern of *in silico*-designed mutants compared with wild-type, supporting its use to design novel mutants with enhanced binding properties.

1.3.2. Introduction

Cyanovirin-N (CVN), a 101 aa lectin isolated from *Nostoc ellipsosporum*, exemplifies a novel class of therapeutic agents that target the surface glycans of HIV and other enveloped viruses.[8,10,24,33-36] Its potent activity against HIV is mediated by high-affinity binding to the mannose-rich glycans that decorate the surface of gp120, and requires multivalent interactions through two binding domains within the

protein.[18,23,28,29,32] CVN is unique among lectins in recognizing a glycan unit, Manα(1,2)Manα, with high affinity and specificity: glycan array analysis reveals that only mannose-rich structures containing the specific glycosidic linkage α-Man-(1,2)-α-Man are recognized.[37,38] For these reasons, the protein is emerging as a convenient model system for investigating glycan-protein interactions and testing computational approaches to glycan recognition.

The structure of CVN reveals a novel (to our knowledge) β-sheet fold that comprises two quasi-symmetric domains, defined as A (residues 1-38/90-101) and B (residues 39-89), connected on each side by a short 3_{10} helix. Despite a backbone root mean-square standard deviation (RMSD) of 0.76 Å, the two domains are ~25% identical in sequence, resulting in slightly different glycan selectivity and affinity. Experimental and computational simulation data show a preference for linear trimannose in domain A, and a preference for dimannose (Manα(1,2)Manα) in domain B.[10,20,23,24,27,33-36,39-43] Domain B is generally considered the high-affinity binding site: the reported binding constants are in good agreement and show low-micromolar affinity for Manα(1,2)Manα.[16,18,20,23,28,29,32,39,44] Structural data for CVN and mutants containing an intact domain B in the apo and dimannose-bound forms[16,18,20,23,27,37,38,41-44], complemented by binding studies using oligomannose derivatives[45], contributed to mapping the key interactions in the glycan recognition event on both the protein and the glycan. The origins of the unique specificity and high affinity of wt CVN for dimannose are poorly understood.[46] In particular, the energetic contribution of backbone versus side-

chain interactions, and the role of water-mediated hydrogen bonds in determining binding affinity are still under debate.

Here, we analyzed glycan binding by CVN and its mutants by applying a flexible docking approach called back-bone perturbation docking (BP-Dock), which is based on perturbation response scanning (PRS), a method that combines elastic network models with linear response theory.[47-50] This approach accurately predicted binding affinity patterns of the PDZ domain of PICK1 with various PDZ motifs.[50] BP-Dock enabled us to integrate backbone and side-chain conformational changes of the protein, as well as the flexibility of the glycan ligand, in the docking analysis, which is crucial for accurate prediction of binding affinity.[50] Compared with computationally expensive traditional molecular-dynamics (MD) calculations performed on CVN[42,51-53], this method allows for rapid *in silico* screening of a series of mutants.

For simplicity, we limited our analysis to the high-affinity site, domain B. We benchmarked our method by comparing two well-characterized mutants of CVN (Figure 7): one that comprises an intact domain B (Figure 7 B) and one in which binding to domain B has been abolished (Figure 7 C). The computed binding affinities are consistent with high-affinity binding for one, and with no binding for the other. Extending our docking analysis to single mutants of P51G-m4-CVN, we observed that some mutations led to less favorable binding-energy scores, whereas others did not. Based on these predictions, we experimentally characterized five mutants and obtained binding affinities in agreement with the predictions.

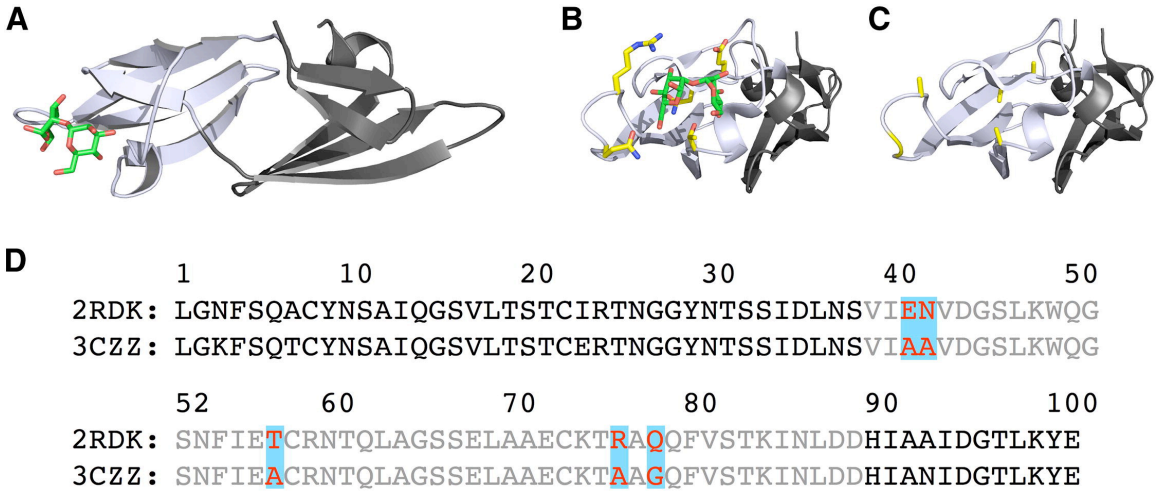


Figure 7. (A) Crystal structure of P51G-m4-CVN (PDB ID: 2RDK) with dimannose bound. (B) Domain B of P51G-m4-CVN, with residues focused on in this study highlighted in yellow (blue, nitrogen atoms; red, oxygen atoms). (C) Domain B of CVN^{mutDB} (PDB ID: 3CZZ), with residues that differ from those of P51G-m4-CVN highlighted in yellow. Domain B of CVN^{mutDB} does not bind dimannose. (D) The sequences of the two mutants are shown aligned with domain A (black) and domain B (gray). The differences between P51G-m4-CVN and CVN^{mutDB} in domain B are highlighted.

Our work demonstrates that rapid docking can be an efficient tool for predicting binding-affinity patterns of various mutants in glycan recognition. However, the method relies on accurate prediction of the initial unbound mutant structure. With such tools, it should be possible to predict mutations that change affinity and specificity for given glycan targets.

1.3.3. Materials and Methods

Benchmark: We analyzed bound mutant structures of P51G-m4-CVN[18,23] and CVN^{mutDB} [39] in this study. The crystal structures for P51G-m4-CVN (PDB code: 2RDK) and CVN^{mutDB} (PDB code: 3CZZ) were retrieved from the Protein Data Bank (http://www.rcsb.org/pdb).[54] These structures were used as templates for binding affinities to dimannose sugar in our flexible docking simulations.

Modeling of single point mutant structures: We first checked the possible impact of single point mutations on the local structure, or the overall stability of the protein. By analyzing the hydrogen-bond network of each individual site within the binding pocket, we found that the side chains of some residues, such as Asn-42 and Thr-57, made critical hydrogen bonds with the backbone or side chain of other positions of CVN (Table 1). Thus, mutating these positions to Ala could alter the stability or lead to local structural change. To capture those effects, we introduced the point mutations into the x-ray structure via Swiss PDB Viewer[55], and performed short (3 ns) replica-exchange MD (REMD) simulations to obtain well-equilibrated mutant models. The initial point mutant structure was seeded into the REMD[56] simulation, and the temperature of the replicas ranged from 270 K to 450 K. Replica spacing was set such that the swap likelihood between replicas approached 0.45. The AMBER force field (ff99SB)[57] with a generalized Born implicit solvent[58] model and a solvent-accessible penalty term of 5 cal/mol Å was used. An α-carbon covariance matrix was calculated from a 1 ns window of the trajectory, and the convergence was checked by the observed high correlation between the slowest fluctuation profiles of the two successive windows. After clustering the snapshots of the

lowest replica, we used the most dominant structure (i.e., the lowest free-energy state) as the predicted mutant model structure.

Table 1. Inter residue hydrogen bond network in P51G-m4-CVN for the specific Domain B residues considered in the study.

Residue #	Donor Residue	Donor Atom	Acceptor Residue	Acceptor Atom	H-bond distance
41	GLU41	N	LYS48	O	2.853
	LYS48	N	GLU41	O	2.902
42	ASN42	ND2	THR57	OG1	2.908
	ASN42	ND2	LYS74	O	2.902
	GLY45	N	ASN42	OD1	2.831
57	THR57	N	PHE54	O	3.062
	ASN42	ND2	THR57	OG1	2.908
	LYS74	N	THR57	O	2.973
76	ARG76	N	ASP44	OD1	2.901
78	GLN78	N	THR75	O	2.889
53	ASN53	N	ASN37	O	2.926
	ASN53	ND2	ASN53	OD1	3.081
	GLU56	N	ASN53	O	3.136

PRS: When a ligand (i.e., dimannose sugar) approaches a receptor (CVN mutants), it exerts force on the binding site, inducing conformational changes. In this work, as a first-order approximation, we mimicked this by applying small random forces on the protein and then computing the displacement of each residue through PRS.[47,59] As a matter of fact, the most intriguing aspect of the BP-Dock approach is that it generates a wide range of binding-induced conformations realized in nature by exerting random unit forces on each residue in the sequence without the presence of ligand. This enabled us to integrate the conformational changes of the receptor (CVN mutants) in the docking analysis, which in turn increased the accuracy of the binding predictions. Thus, ensembles for P51G-m4-CVN, CVN^{mutDB}, and the individual domain B mutant models were obtained by PRS. PRS is a coarse-grained approach that couples elastic network models with linear response theory.[47,49,60] Previously, we have shown that PRS can 1) capture conformational changes upon binding[47], 2) identify key residues that mediate long-range communication and find allosteric pathways[49], and 3) discriminate disease-associated mutants from neutral ones.[61] We compared the dynamics of PRS, which is based on a coarse-grained approach, with those obtained from all-atom MD simulations, and these two estimates showed high correlations.[49,61] However, PRS requires four orders of magnitude less CPU time than MD simulations. Furthermore, although convergence is a bottleneck for MD simulations of larger proteins, it is not an issue in the case of PRS. Therefore, PRS is a fast and efficient approach to incorporate backbone flexibility[48,49,62], which most of the conventional docking protocols cannot do because it is a computationally expensive and difficult procedure.

In the model, the protein structure is viewed as a 3D elastic network in which the Cα atom acts as the nodes and identical springs connect the interacting α-carbons. We applied a random unit force, F, in a random direction to the α-carbon atom of each residue one at a time (i.e., perturbation of a single residue in the chain with a random Brownian kick) sequentially along the chain. The perturbation cascaded through the residue interaction network of the protein and caused other amino acid positions to respond. By using the PRS method, we computed this fluctuation response as a profile, in both direction and magnitude. Thus, the response vector, ΔR, gave the deviation of each residue from its mean position in the *x*, *y*, and *z* directions upon perturbation. This procedure indeed mimicked the natural process of protein binding interactions in a cell as a first approximation, since an approaching ligand applies forces on the receptor protein, inducing conformational change. To minimize the effects of randomness, we performed this procedure 10 times to ensure that the force applied was isotropic and the response vector was averaged. These response fluctuation profiles were used to generate low-resolution multiple receptor conformations (MRCs) based on Cα atoms. If ΔF is a 3N × 1 vector that contains the components of the externally applied force on the selected residues, and H is the Hessian, a 3N × 3N matrix composed of the second derivatives of the potential with respect to the components of the position vectors, we can calculate the response of the elastic network of residues using Eq. 1:

$$\left[\Delta R\right]_{3N\times 1} = \left(\left[H\right]_{3N\times 3N}\right)^{-1}\left[\Delta F\right]_{3N\times 1} \tag{1}$$

where ΔR is the 3N-dimensional vector of the response fluctuations of all the residues.

To derive the perturbed structure, we calculated the final perturbed coordinates for each residue as follows:

$$\left[R_{per}\right]_{3N\times 3N} = \left[R_0\right]_{3N\times 1} + \alpha\left[\Delta R\right]_{3N\times 1} \quad (2)$$

where R_0 is the vector showing initial coordinates of the residues before perturbation, and a is a scaling factor with a value of 50. The arbitrary value of 50 was chosen such that it could yield structures in a range of 0.5-2.0 Å RMSD from the original (unperturbed) crystal structure.[62,63]

The perturbed structures were then clustered to discard similar conformations generated from the perturbations on different residues in the protein. The perturbed structures generated from PRS varied from each other and had a backbone RMSD of 0.58-2.15 Å relative to the initial structure. Once the backbone was perturbed, coarse-grained conformations were obtained by integrating the response fluctuation profile into the original Cartesian coordinates; the side chains were derived from the original unperturbed structure. An all-atom minimization of these structures was done using the AMBER package[64,65] to account for the fluctuation in the side chains upon perturbation on the backbone of the protein and to relieve strain in the system. Another way to couple the side-chains to the perturbed backbone conformations is to update the side-chain positions using the internal coordinates of the initial structure and then follow up by energy minimization. We tested this approach for 2RDK and the E41A and E41G mutants. However, the RMSD between the conformations obtained after minimization with and without updated internal coordinates were in the range of 0.26-0.49 Å for

2RDK, 0.03-0.2 Å for E41A, and 0.27-0.43 Å for E41G. We also observed that the docking results did not show any significant difference compared with results from the minimization approaches discussed above (Table 2). This could be because side-chain conformations are again repacked using either rotamer trials or a full combinatorial search in RosettaLigand.[66,67]

Table 2. Docking results for two methods of all-atom minimization to include side chain position to the perturbed backbone conformations for 2RDK, E41A and E41G mutant.

	Using updated Internal Coordinates		Without using updated Internal Coordinates	
Protein	Rosetta Score (kcal/mol)	X-score (kcal/mol)	Rosetta Score (kcal/mol)	X-Score (kcal/mol)
2RDK	-17.25	-7.58	-17.27	-7.51
E41A	-13.3	-6.55	-13.41	-6.23
E41G	-14.49	-6.9	-14.12	-6.64

Docking with RosettaLigand: An ensemble docking using PRS-generated conformations was performed with RosettaLigand.[66,67] The docking simulations for each structure in the ensemble were performed using the RosettaLigand protocol in the ROSETTA package. RosettaLigand is a method that was specifically developed for docking ligands into protein-binding sites. This method incorporates ligand flexibility by

changing torsional angles and the backbone of the ligand, whereas the whole protein is held fixed throughout the docking simulation. In each case, the coordinates of the dimannose sugar were taken from the crystallographic complex of 2RDK and were treated as a single residue. The ligand was represented as a set of discrete conformations in the docking procedure, ensuring that the ensembles spun a maximal range of the conformational space.[67] In this study, we perturbed the ligand position and orientation randomly with the translation of mean 0.1 Å and rotations of mean 3˚, respectively. The rigid-body orientation and side-chain χ angles of the ligand were optimized using the gradient-based Davidson-Fletcher-Powell algorithm.[66,67] Each docking trajectory consisted of 50 of the Monte Carlo minimization cycles, and we computed 10,000 trajectories to generate a comprehensive ensemble of conformations of the receptor-glycan complex for each protein. The formation of a distinct binding funnel in energy score/RMSD plots was considered as an indication of successful docking, and the final docked conformations were selected based on the lowest free-energy pose in the protein-binding site. The lowest free-energy pose had the lowest Ligsum score among all other docking poses (Ligsum is a RosettaLigand scoring function composed of five energy terms: attractive, repulsive, solvation, hydrogen bonding, and coulombic).[66,67] The overall flexible docking approach is outlined in Figure 8. The binding energies of the bound complexes with the lowest Ligsum score were evaluated with X-Score. X-Score is an empirical scoring function that was developed to rerank the protein-ligand complex and give a more accurate estimate of the binding free energies, which have been found to correlate with the experimental binding constants.[68]

Protein cloning, expression, and purification: Mutagenic primers were used to introduce the E41A, E41G, T57A, N42A, and N53S single mutations into a P51G-m4-CVN sequence in a pET26b(+) vector and then transformed into BL21(DE3) cells (Novagen, Madison, WI). Proteins were expressed as previously described.[18,23] Briefly, 100 mL cell cultures were incubated overnight at 37˚C, and then 10 mL of this culture was added to a 1 L of lysogeny broth (LB). For NMR experiments, a modified M9 minimal media containing $^{15}NH_4Cl$ (Cambridge Isotope Laboratories, Andover, MA) as the sole nitrogen source was used in place of LB. Cells were induced with addition of 1 mM isopropyl β-D-1-thiogalactopyranoside when the OD_{600} reached 0.8 and were grown overnight at 37˚C, and harvested via centrifugation. The cell pellet was lysed by sonication and the insoluble fraction was solubilized using an 8 M urea buffer. The 6× histidine-tagged proteins were purified on a GE HisTrap HP column (GE Healthcare Bio-Sciences, Piscataway, NJ) using a Bio-Rad EconoPump (Bio-Rad, Richmond, CA) under denaturing conditions. Proteins were refolded by dialysis into a 2 M urea buffer, pH 8.0, at room temperature for 12 h, and then overnight into 10 mM Tris, 100 mM NaCl, pH = 8.0, replacing the Tris buffer once. The protein secondary structure and stability were verified by circular dichroism (CD) on a JASCO J-815 CD spectrometer (JASCO, Easton, MD). Temperature denaturation experiments were performed by following the molar ellipticity at 202 nm of samples containing 20 μM protein in 15 mM potassium phosphate; samples were heated over a range of 4–90˚C at 1˚C/min with a data pitch of 0.5˚C.

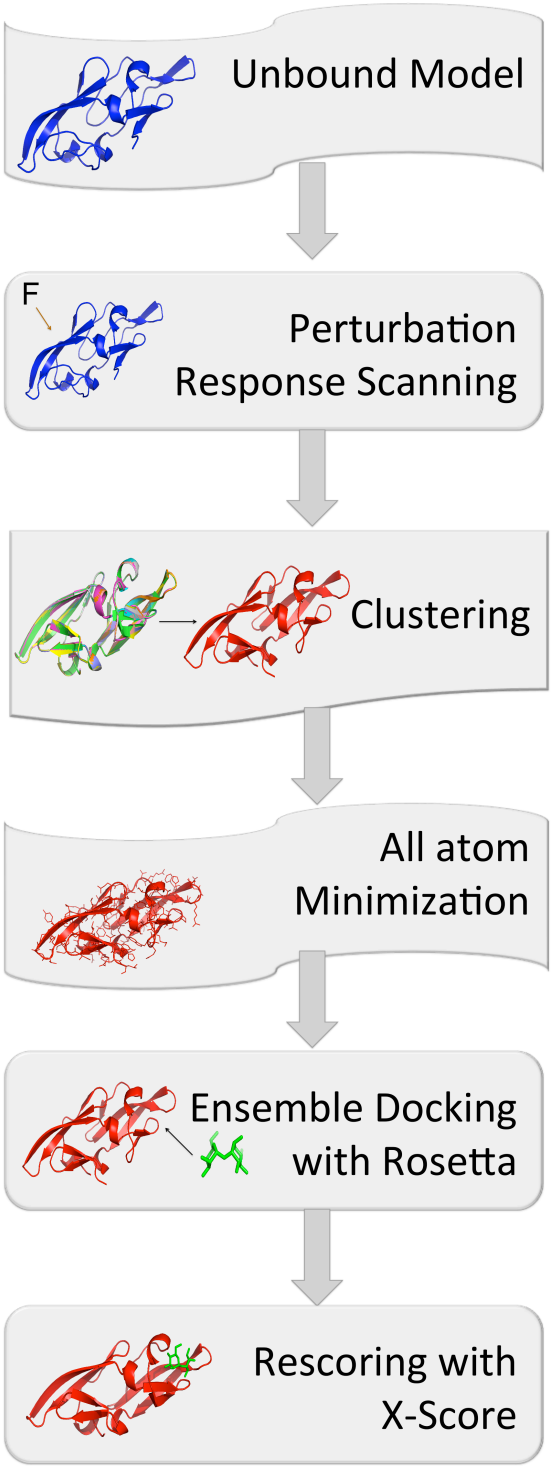


Figure 8. Flow chart of flexible docking method. In our flexible docking approach using PRS model, we generated an ensemble of receptor conformations through several steps:

(i) sequentially exerting random external force on each single-residue of the CVN mutants, (ii) calculating the response fluctuation vector using Perturbation Response Scanning (PRS) method, (iii) constructing the low resolution deformed structures (i.e. backbone) using the response vectors after each single residue perturbation, (iv) clustering the perturbed conformations using k- clustering method, and (v) all-atom minimization of each clustered conformation. Once the multiple receptor conformations (MRCs) ensemble was completed, we performed a docking simulation using the RosettaLigand option in the ROSETTA package for each minimized structure in the ensemble. (vi) Lastly, the binding energies of the CVN-dimannose complexes with the lowest Ligsum score were evaluated with X-Score.

NMR titrations: Purified ^{15}N-labeled protein was buffer exchanged by gel filtration with a GE HiLoad 16/600 Superdex 75 pg column (GE Healthcare Bio-Sciences) on an Agilent 1260 BIO-inert HPLC system (Agilent Technologies, Santa Clara, CA) into 20 mM sodium phosphate buffer, pH 6.0. Fractions containing the monomer of the protein were pooled and concentrated using Amicon Ultracel-3K centrifugal filters (Millipore, Billerica, MA) to 300 μM protein.

Aliquots from a stock solution of 2α-mannobiose (Sigma Aldrich, St. Louis, MO) were titrated into the protein. ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra were obtained on a Varian Innova 500 MHz spectrometer equipped with a room-temperature triple-resonance probe at gradually increasing molar ratios of ligand to

protein. Chemical shifts were observed and the Δδ were determined according to Eq. 3[69,70]:

$$\Delta\delta = \left[\left(\Delta\delta_{HN} \right)^2 + \left(0.17 \times \Delta\delta_N \right)^2 \right]^{1/2} \quad (3)$$

The chemical shifts were plotted against the molar ratio ([ligand]/[protein]) and fitted to Eq. 4:

$$\Delta\delta = 0.5 + \Delta\delta_{\max} \left(M + 1 + \frac{K_d}{[P]} - \sqrt{\left(M + 1 + \frac{K_d}{[P]} \right)^2 - 4M} \right) \quad (4)$$

1.3.4. Results

Our analysis focused on domain B, which has higher affinity and specificity for dimannose (Manα(1,2)Manα) in wt CVN, and for which extensive mutational analysis is available. We used two engineered forms of CVN, P51G-m4-CVN and CVN^{mutDB}, which present the two extremes of the binding affinity range: P51G-m4-CVN contains a wild-type (wt) domain B and thus binds dimannose with high affinity[18,23], whereas CVN^{mutDB} contains an impaired domain B and thus does not bind dimannose at concentrations as high as millimolar[21]. P51G-m4-CVN and CVN^{mutDB} each contain a P51G mutation in the hinge region that stabilizes the monomeric form relative to the domain-swapped form of CVN.[30] The two mutants have been thoroughly characterized, with crystal structures available for CVN^{mutDB} [39] and for P51G-m4-CVN in its apo and bound forms.[18,23]

Using the crystallographic coordinates as a starting point, we generated ensembles of binding-induced conformations by subjecting each protein to PRS. Specifically, we first obtained backbone conformational changes upon perturbing different positions with a

random force. We generated the full-atom models by adding the side chains and performing an all-atom minimization for each conformation in the ensemble. We used RosettaLigand to dock the dimannose ligand into the cavity or binding pocket for each structure in the ensemble, and scored the lowest-energy poses as described in Materials and Methods. The binding energy of dimannose to domain B of P51G-m4-CVN calculated by X-Score is -7.58 kcal/mol and the RMSD of the ligand obtained from the lowest-energy docked pose is only 0.19 Å. In contrast, analysis of CVN^{mutDB} shows a drastic change in the binding energy making it less favorable (-4.68 kcal/mol), corresponding to a change in dissociation constant of four orders of magnitude, and a ligand RMSD of 0.38 Å obtained from the lowest-energy docked pose. Figure 9 compares the binding profiles of P51G-m4-CVN and CVN^{mutDB} for dimannose, and the docking profile observed for P51G-m4-CVN indicates a well-converged distinct binding funnel. In contrast, CVN^{mutDB} shows a poor docking profile in which energetically degenerate structures have high RMSDs.

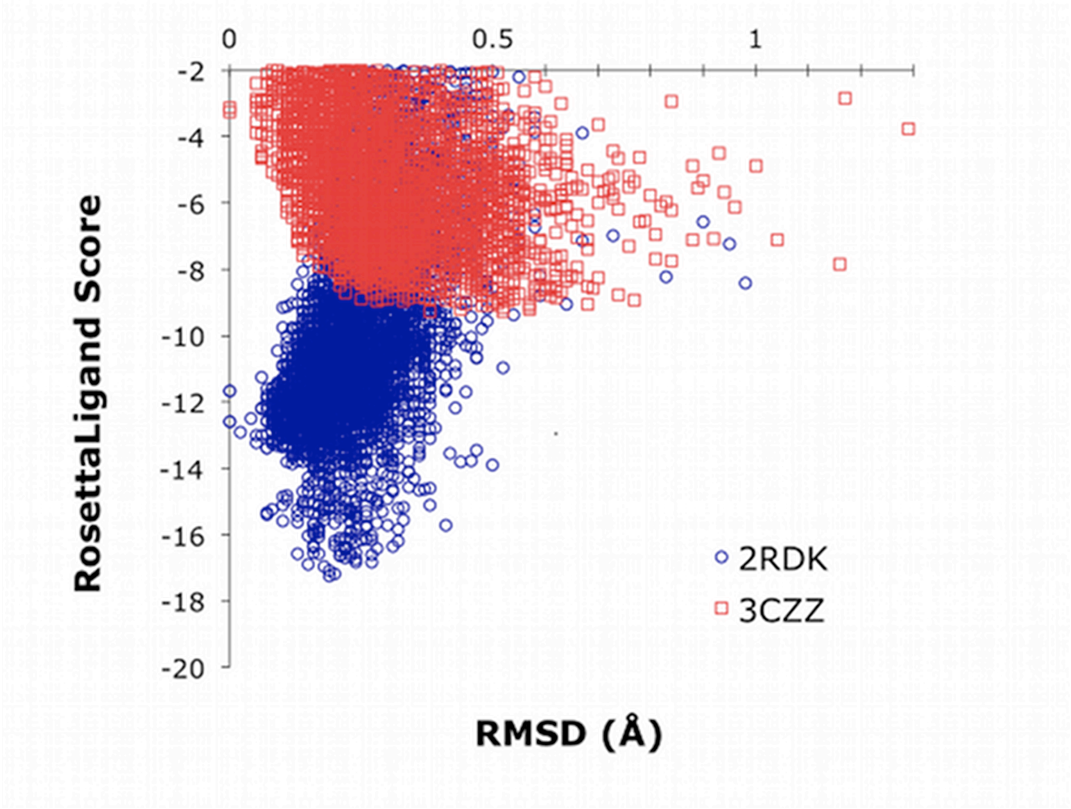


Figure 9. The RosettaLigand energy scores (in kcal/mol) versus RMSD of the docked complex for CVN mutants and dimannose sugar. P51G-m4-CVN is shown in blue circles and CVN^{mutDB} in red squares. The funnel shape of the graph indicates good docking for P51G-m4-CVN. In agreement with experiments P51G-m4-CVN-dimannose has significantly low binding energy score (-17.25 kcal/mol) while the complex of CVN^{mutDB} (-9.28 kcal/mol) does not.

Next, we analyzed each residue's energetic contribution to binding by systematically mutating each position in P51G-m4-CVN to the corresponding amino acid in CVN^{mutDB}: Glu-41, Asn-42, Thr-57, Arg-76 were changed to Ala, and Gln-78 was changed to Gly. We also mutated residue 41 to glycine, which is the most common substitution at that

position in the CVN-homologous family of proteins, CVNH.[37,71] In addition to the positions mutated in CVN^{mutDB}, we mutated N53, which lines the binding pocket and interacts with the glycan, to serine, because serine is the most common substitution at this position in CVNH.[37,71]

Table 3. Binding energy scores (E_{bind}) evaluated from X-Score along with corresponding RMSDs of the ligand ($RMSD_{ligand}$) and protein side chain positions in the binding site ($RMSD_{sidechain}$) for various CVN mutants are shown. Higher negative values indicate a higher binding affinity prediction. Proteins are ordered by decreasing experimental K_d values.

Protein	$RMSD_{ligand}$ (Å)	$RMSD_{sidechain}$ (Å)	Xscore E_{bind} (kcal/mol)	Experimental K_d
P51Gm4CVN	0.19	0.189	-7.58	low μM
N53S	0.88	1.365	-7.06	low μM
R76A	1.06	1.036	-6.9	NA
Q78G	0.98	0.845	-6.85	NA
E41G	0.82	1.261	-6.82	389 ± 73 μM
E41A	0.74	1.155	-6.7	541 ± 118 μM
T57A	0.99	1.08	-6.46	low mM
N42A	1.08	1.808	-5.37	no binding
$CVN^{(mutDB)}$	0.38	1.866	-4.68	no binding

In the absence of crystal structures for the mutants, we generated starting models by introducing the individual point mutations into the coordinates of P51G-m4-CVN and

performing a short REMD run (3 ns). Starting from the most dominant structure of the lowest-temperature REMD run (i.e., the structure with the lowest free energy), we applied our flexible docking approach to each mutant (see Materials and Methods for details). The mutations at positions 41, 42, and 57 resulted in decreased affinity compared with that of P51G-m4-CVN (Table 3). Position 42 accounts for most of the loss of binding energy between the intact domain B in P51G-m4-CVN and the defective domain B in CVN^{mutDB}: the X-Score binding energy of N42A (-5.37 kcal/mol) is in close proximity to the non-binder CVN^{mutDB} (i.e., -4.68 kcal/mol). This drastic loss in binding energy score may indicate the critical role of N42 in stabilizing the structure of the binding pocket. The side chain of N42 forms hydrogen bonds to the backbone of neighboring residues (Table 1); further, this position is very well conserved in the CVNH family.[37,68] T57A, E41A, and E41G also showed a significant decrease in binding affinity, albeit not as drastic as that observed for N42A. In contrast, the binding energy score of N53S mutant (-7.06 kcal/mol) is comparable to that of P51G-m4-CVN (-7.58 kcal/mol). To verify the accuracy of our predictions, we recombinantly expressed E41A, E41G, T57A, N42A, and N53S on the background of P51G-m4-CVN[18,23], and subjected them to experimental characterization.

Effect of single point mutations on fold and stability of CVN: We confirmed that the mutants are monomers, similar to the parent construct, using size exclusion chromatography. The elution profiles of P51G-m4-CVN and the mutants each exhibit a main peak at 89 mL, which accounts for over 80% of the total protein; small peaks at 70 mL correspond to dimer (Figure 10). We assigned monomer and dimer peaks by

comparing the elution profiles with those of recombinant wt CVN, in which the ratio of domain-swapped dimer to monomer is ~40:60.[23,30]

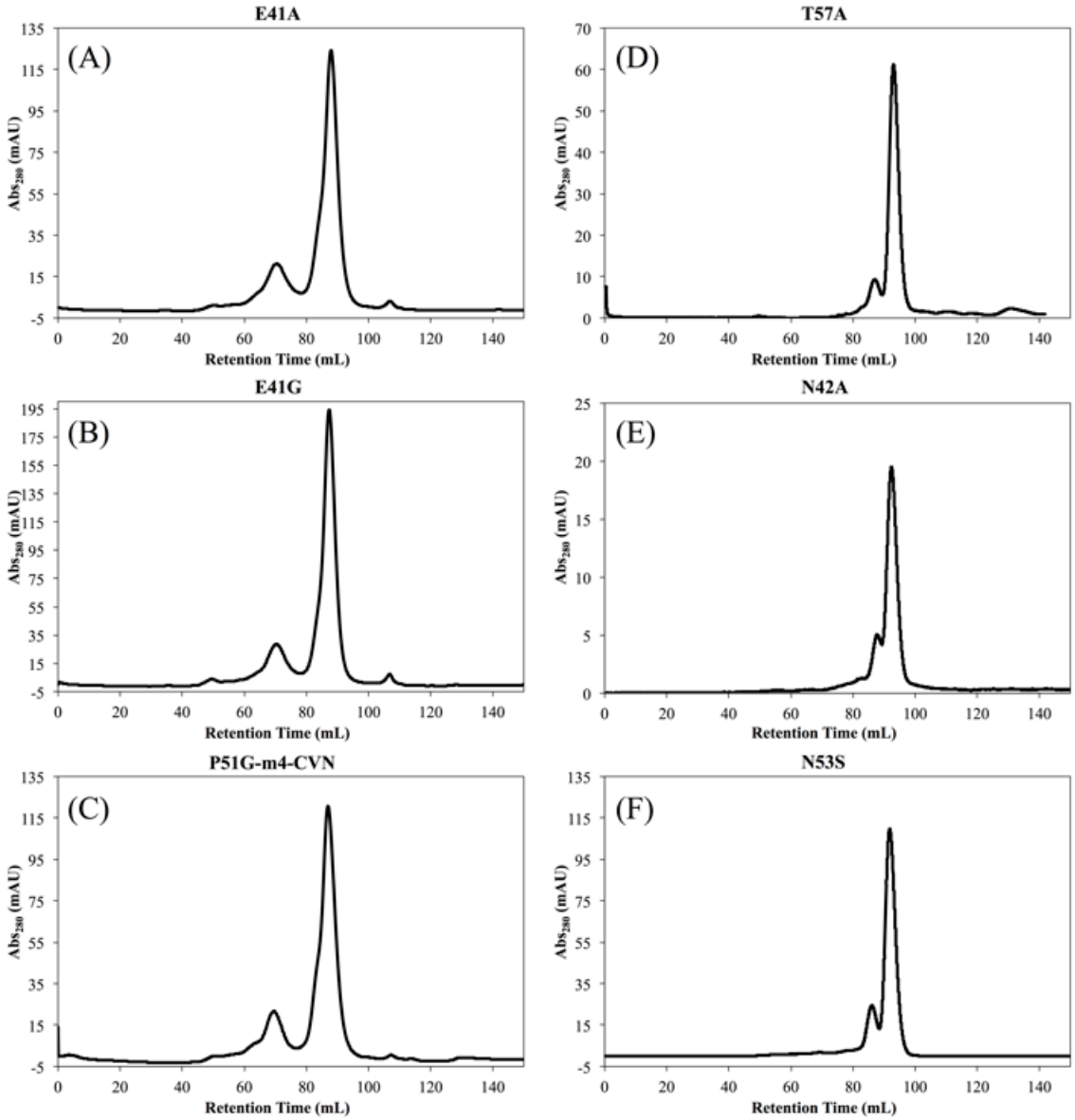


Figure 10. Elution profile of gel filtration of the mutants used in this study: (A) E41A, (B) E41G, (C) P51G-m4-CVN (D) T57A, (E) N42A, and (F) N53S. Monomer fractions were collected and used for circular dichroism studies and ^{1}H-^{15}N HSQC experiments.

The mutations do not affect the overall secondary and tertiary structure of P51G-m4-CVN, as assessed by CD spectroscopy and 2D-NMR. The CD spectra of the mutants are identical to that of P51G-m4-CVN, indicating mainly β structure, with a characteristic minimum at 212 nm and maximum at 190 nm (Figure 11). The ^{1}H-^{15}N HSQC correlation spectra of isotopically labeled mutants show that the backbone amide resonances are well dispersed in both dimensions and resemble the spectrum of P51G-m4-CVN (Figure 13, P51G-m4-CVN; Figure 14, E41A and E41G, red spectra; Figures 15-17 for N42A, T57A, and N53S, respectively). Taken together, the spectroscopic results demonstrate that the mutants display the hallmarks of native proteins, and are consistent with the conservation of the CVN fold.

We assessed the thermodynamic stability of each mutant by thermal denaturation, monitoring the loss of secondary structure at increasing temperatures from 4˚C to 90˚C by CD at 202 nm, the wavelength at which the largest change is observed between folded and unfolded spectra. The thermal denaturation of wt CVN and its mutants is not reversible, which precludes a rigorous analysis of the folding thermodynamics. It is possible, however, to compare the thermal denaturation profiles of a series of mutants and calculate the apparent midpoint of the melting transitions (T_m). The melting curves of the mutants (Figure 12) show that, with the exception of T57A and N42A, the single point mutations do not affect protein stability significantly. The apparent melting points T_m are shown in Table 4.

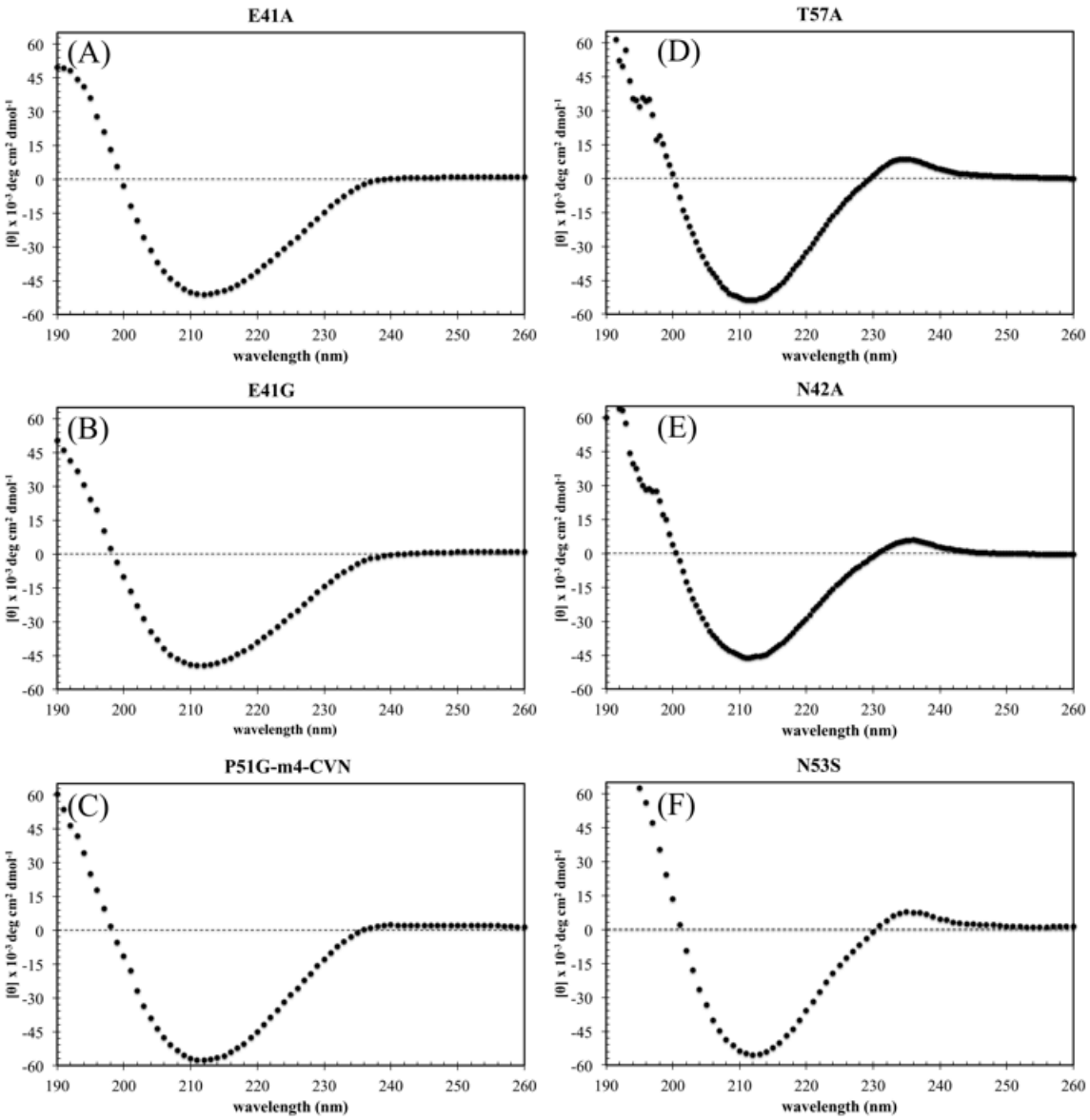


Figure 11. The near UV circular dichroism spectra of the mutants (A) E41A, (B) E41G, (C) P51G-m4-CVN, (D) T57A, (E) N42A, and (F) N53S are shown. All the proteins show the characteristic minimum at 212 nm and maximum at 190 nm expected of cyanovirin-N mutants.

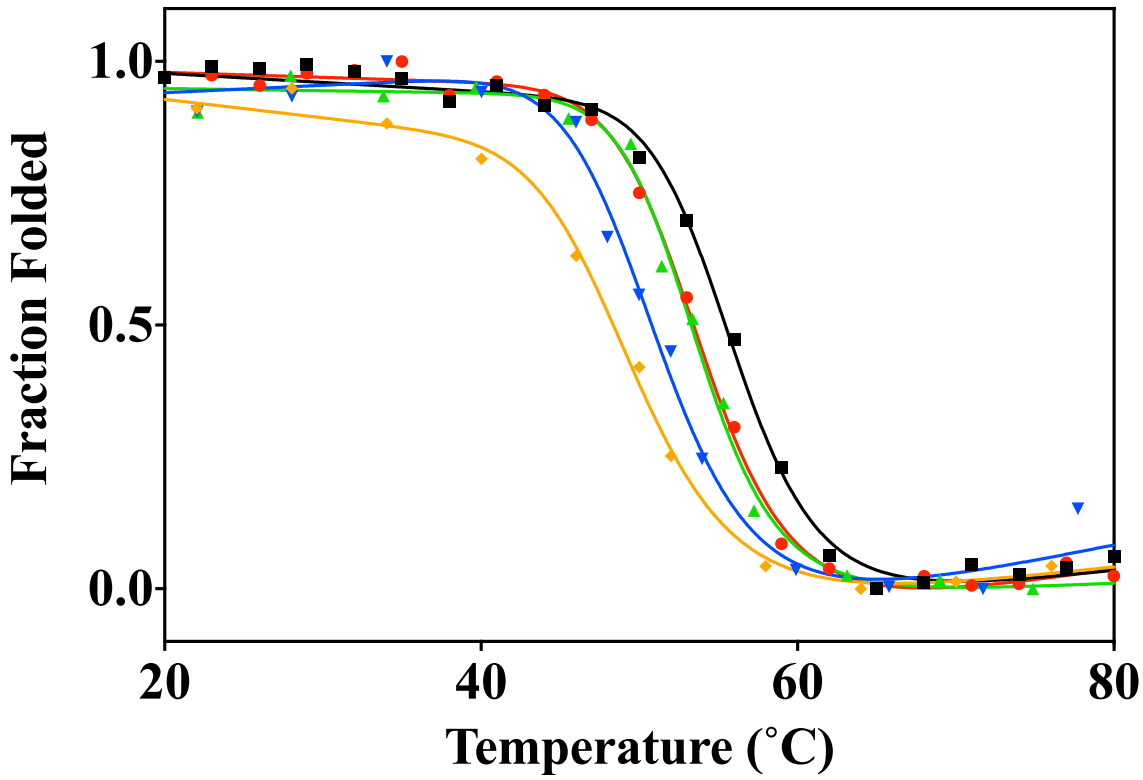


Figure 12. Thermal denaturation profiles of E41A (red circles), E41G (black squares), N53S (green triangles), T57A (blue upside down triangles), and N42A (orange diamonds). The unfolding transitions were monitored at 202 nm from 4 to 90˚C using 20 μM protein in 15 mM potassium phosphate. The melting points are 53˚C, 55˚C, 53˚C, 50.6˚C, and 48˚C, respectively.

Table 4. Melting temperatures of the proteins used in this study are outlined. The proteins unfold irreversibly, preventing a rigorous analysis of folding thermodynamics.

Protein	T_m (˚C)
P51G-m4-CVN	59
E41G	55.3
E41A	53.4
N53S	53.6
T57A	50.6
N42A	48.2

Binding to dimannose (Man2): We used 2D-NMR spectroscopy to quantify the binding of dimannose to P51G-m4-CVN and each of the mutants, by monitoring the changes in chemical shift of the backbone amide signals in uniformly labeled proteins upon addition of ligand through ^{1}H-^{15}N-correlation spectroscopy (HSQC). This technique is particularly well suited for measuring dissociation constants in the mid- to high-micromolar range[72]. Titrations were carried out by adding dimannose from a stock solution up to at least a 15 molar equivalent in 300 μM protein. A comparison of the ^{1}H-^{15}N HSQC spectra of the P51G-m4-CVN before (black) and after (red) addition of 12.5 molar equivalents of dimannose is reported in Figure J. The ^{1}H-^{15}N HSQC spectra resulting from the titration of increasing amounts of dimannose (red to black color) into E41A and E41G are shown in Figure JJ, A and B, respectively. The binding isotherms derived by plotting changes in normalized chemical shifts (Δδ, ppm) of selected

resonances as a function of added dimannose are shown in Figure 14, C and D. As expected, P51G-m4-CVN bound tightly to dimannose, and the changes in chemical shift signals saturated upon addition of slightly more than one equivalent. In contrast, addition of dimannose to E41A and E41G resulted in much weaker binding. The monitored chemical shift changes reached saturation only upon addition of several excess equivalents of dimannose (5.5 molar equivalents). We calculated the equilibrium dissociation constants (K_d) for E41A and E41G by analyzing the changes in chemical shifts versus ligand concentration for several resonances, as described in Materials and Methods. This analysis yielded K_d values of 541 ± 118 μM for E41A, and 389 ± 74 μM for E41G. For T57A, the chemical-shift changes were initially small (<2 molar equivalents) and did not reach saturation even after the addition of 15 molar equivalents (Figure 15). This behavior suggests a K_d in the low millimolar range. N42A did not show any binding to dimannose, as no change in chemical shift was observed in molar equivalents up to 20 (Figure 16). In contrast, binding of dimannose to P51G-m4-CVN displayed the hallmark of tight binding and occurred in fast-to-intermediate exchange mode in the NMR timescale. This behavior prevented an accurate determination of K_d by NMR. Consistent with the computational predictions, N53S behaved similarly to P51G-m4-CVN, and binding was in intermediate exchange (Figure 17). In both cases, analysis of the spectra revealed that the protein was saturated upon addition of one equivalent of dimannose, and further addition of ligand did not result in changes in the signals.

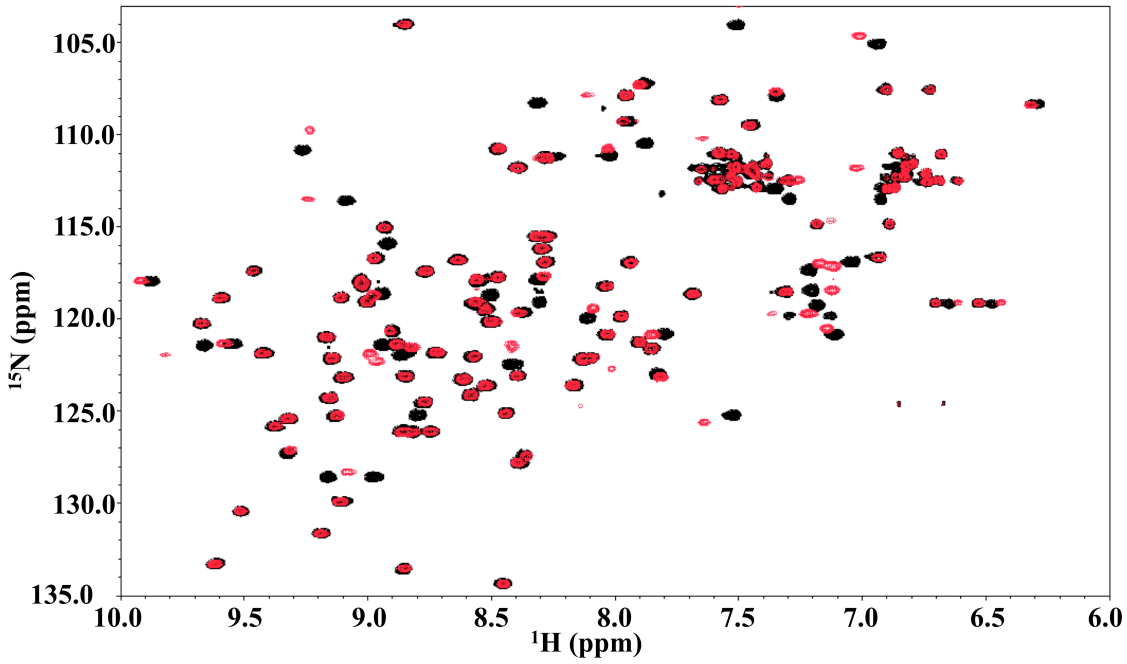


Figure 13. ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of P51G-m4-CVN without dimannose (black) and with 12.5 molar equivalents of dimannose (red).

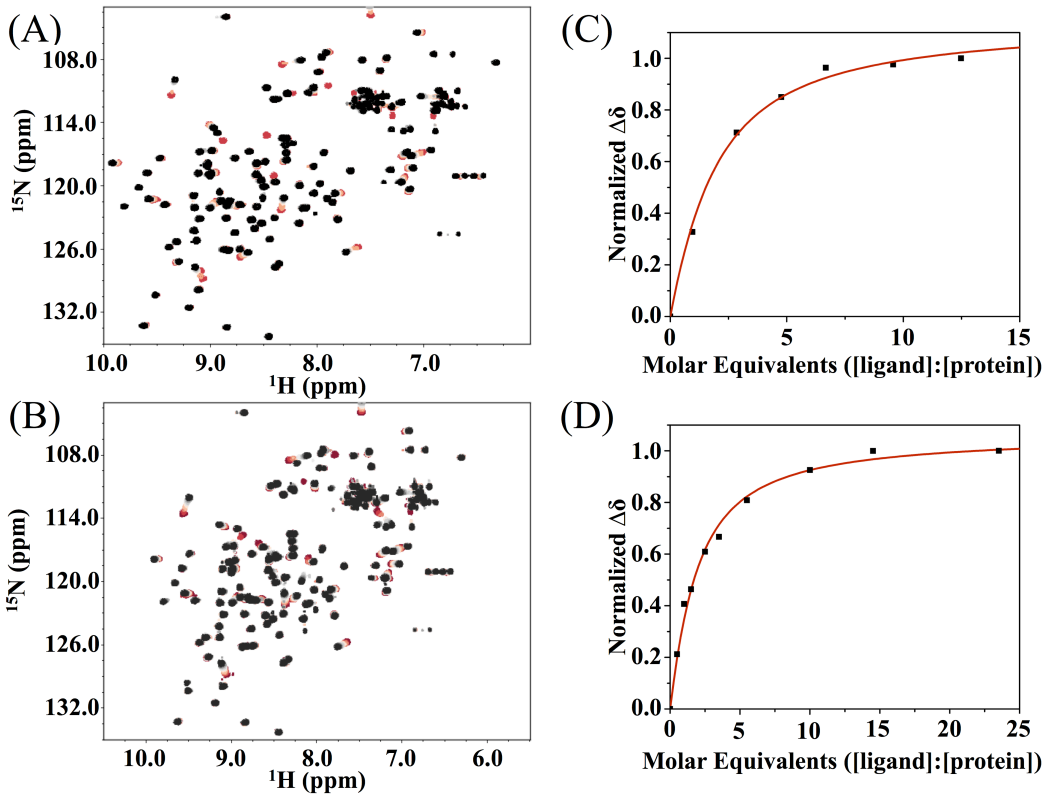


Figure 14. ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of E41A (A) and E41G (B) from titrations with increasing amounts of dimannose (red to black). Resonances are well dispersed and resemble that of the P51G-m4-CVN spectrum, indicating the two mutants are well folded. Example plots and fitting of molar equivalents versus normalized peak shift are shown for a single resonance (E41A in C; and E41G in D). Average K_d from the fits for E41A and E41G are 541 μM and 389 μM, respectively.

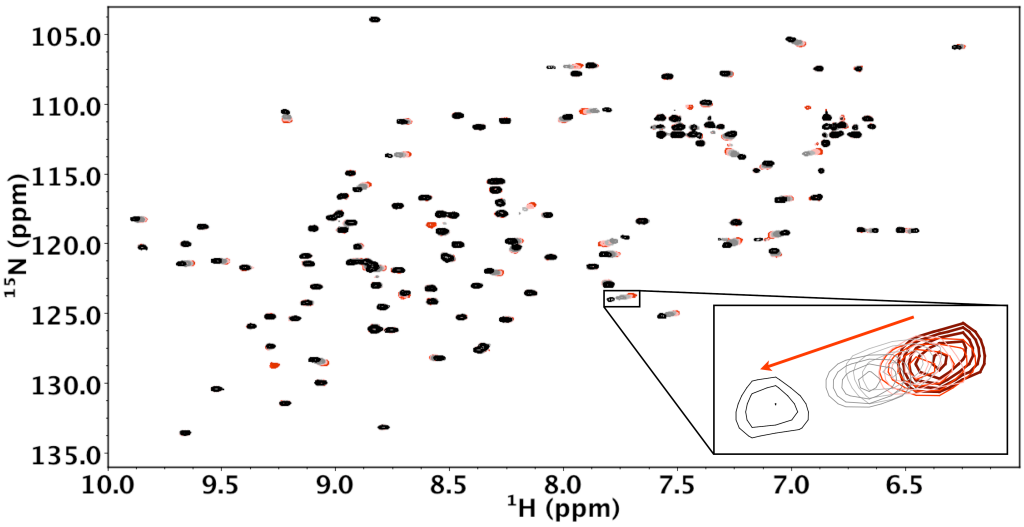


Figure 15. T57A ^{1}H-^{15}N HSQC spectra with increasing molar equivalents of dimannose are shown overlaid. Inset shows blown up view of the change in chemical shift for one signal, with increasing molar equivalents shown with red arrow. T57A binds dimannose at an appreciably larger K_d than P51G-m4-CVN.

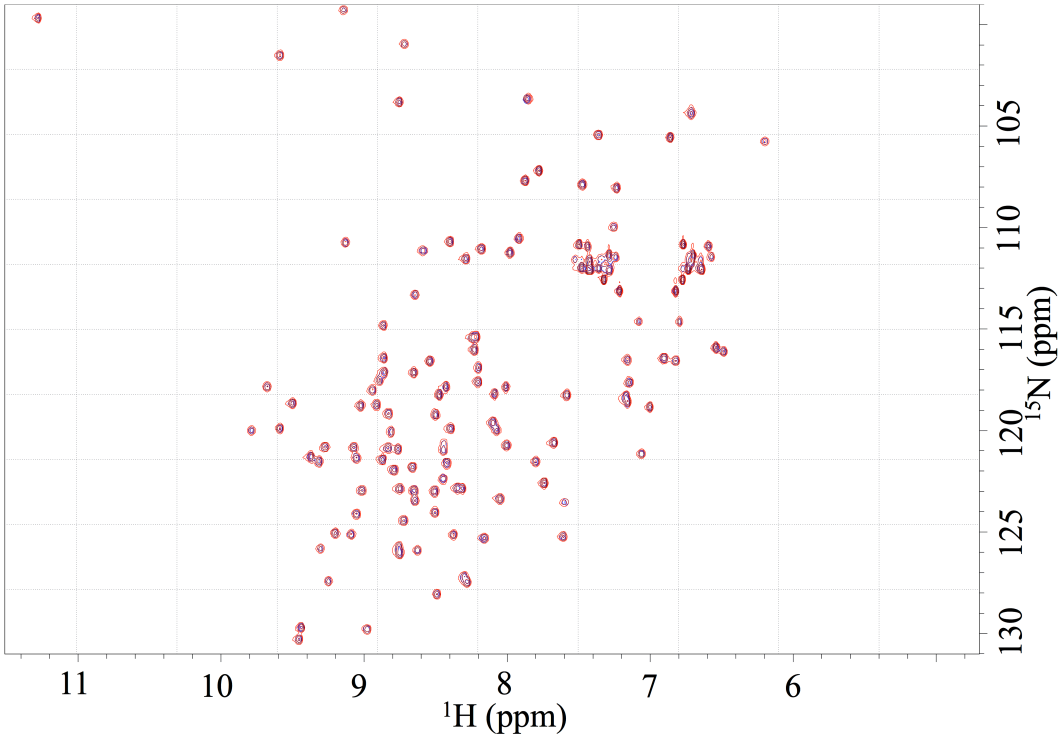


Figure 16. N42A ^{1}H-^{15}N HSQC spectra at 0 molar equivalents dimannose and 20 molar equivalents dimannose are overlaid. The two spectra are identical, signifying that N42A does not bind dimannose.

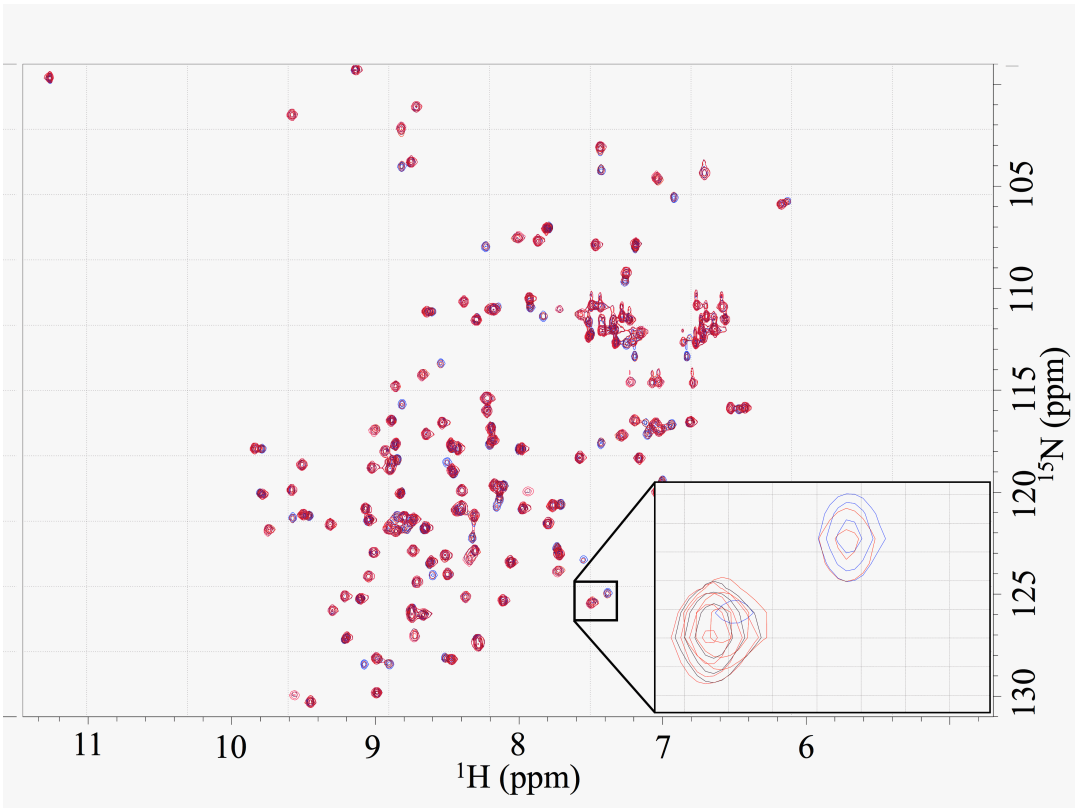


Figure 17. N53S ^{1}H-^{15}N HSQC spectra with increasing molar equivalents of dimannose are shown overlaid. Inset shows blown up view of the change in chemical shift for one signal. N53S was found to be in intermediate exchange, suggesting tight binding.

1.3.6. Discussion

Computationally intensive explicit-solvent MD simulations along with Poisson-Boltzmann-based calculations have been used to evaluate binding free energies and analyze binding interaction networks, using dimannose and trimannose as ligands. Most computational studies have focused on wt CVN[40,42,73] and P51G-m4-CVN.[43,74] The calculated binding energies, in general, reproduced experimental values reported in the literature. Explicit-solvent MD simulations indicated the critical role of water and the

flexibility of the binding pocket in the binding free energy calculations[42], which were later confirmed experimentally.[18,23] These computational studies showed that several hydrogen-bond interactions originate from the backbone amides and carbonyl moieties[43,74]; however, some side-chain hydrogen-bond interactions were also observed, especially in α-Man-(1,2)-α-Man-(1,2)-α-Man (trimannose) recognition.[73]

Despite the wealth of structural and computational data available for wt CVN and its mutants, the role of main-chain interactions versus side-chain interactions in stabilizing CVN-dimannose complexes remained unclear. Here, we extended our flexible docking protocol to the analysis of dimannose recognition in CVN, using two engineered variants (P51G-m4-CVN and CVN^{mutDB}) as benchmarks. We generated backbone-perturbed conformations of CVN mutants by applying random unit forces to individual residues in a sequential manner. We then performed an all-atom minimization after adding the side chains to the backbone-perturbed structures to account for any rotamer changes of the side chains and to relieve any strain in the structure, as perturbations on the backbone can lead to significant changes in the orientation of the side chains. To understand the drastic decrease in binding affinity of CVN^{mutDB} relative to P51G-m4-CVN, we analyzed the hydrogen-bond network in the lowest-energy docked poses of P51G-m4-CVN-dimannose and CVN^{mutDB}-dimannose using Chimera with a cutoff of 4.0 Å. This approach is commonly used in docking studies to roughly estimate critical ligand-receptor interactions.[75] The network for the P51G-m4-CVN-dimannose complex comprises of main-chain nitrogen atoms of Asn-42 and Asp-44 as donors; main-chain oxygen atoms of Asn-42, Ser-52, Asn-53, and Lys-74 as acceptors; and side-chain Oγ of Thr-57 as both

acceptor and donor (Figure 18 A). This hydrogen-bond network is similar to the hydrogen-bond network identified in previous studies.[74,76] In addition, we observed side-chain hydrogen-bonding interactions of OE2 of Glu-41 as acceptor, and NE of Arg-76, NH2 of Arg-76, and NE2 of Gln-78 as donor with the dimannose (Figure 18 A), which was previously seen only in the case of trimannose binding.[73] CVN^{mutDB} retained all of the main-chain hydrogen bonds at Ala-42, Asp-44, Ser-52, Asn-53, and Lys-74; in contrast, all side-chain hydrogen-bonding interactions observed in P51G-m4-CVN were lost. Only one side-chain hydrogen bond, with OE1 of Glu-56 as acceptor, was observed for CVN^{mutDB}. These results emphasize the critical role played by side-chain interactions of residues 41, 57, 76, and 78 in dimannose binding.

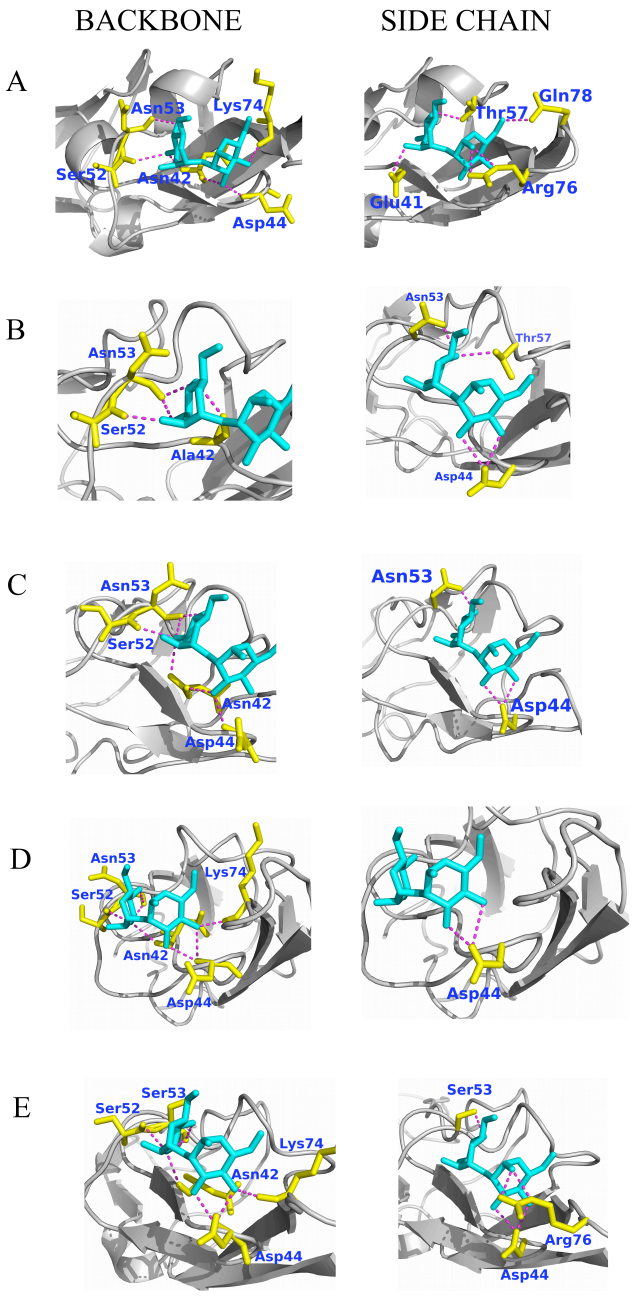


Figure 18. Main chain hydrogen bond network and side chain hydrogen bond network for (A) P51G-m4-CVN (B) N42A (C) E41A (D) T57A and (E) N53S. The dimannose sugar is shown in cyan. The residues that form hydrogen bond with the sugar are shown in yellow and the hydrogen bond in magenta.

A comparison of the hydrogen-bond networks of the E41A, N42A, and T57A mutants with that of the P51G-m4-CVN-dimannose complex showed rearrangements in main-chain and side-chain networks. Specifically, in N42A-dimannose, the main-chain hydrogen bond interactions of the nitrogen of Asn-42 and Asp-44 as donors, and oxygen of Lys-74 as acceptor were lost. The side-chain interactions of OE2 Glu-41, NE and NH2 of Arg-76, and NE2 Gln-78 were lost; however, the side-chain interactions of OD1 Asp-44 and OD1 Asn-53 were gained, preserving the interaction with OG1 of Thr-57 (Figure 18 B). In E41A-dimannose, the main-chain carbonyl oxygen of Lys-74 as acceptor was lost and another oxygen of Ser-52 and Asn- 53 were gained as acceptors. The side-chain interactions of OE2 Glu-41, NE and NH2 of Arg-76, NE2 Gln-78, and OG1 of Thr-57 were lost, and OD1 Asp-44 and OD1 Asn- 53 were gained (Figure 18 C). The T57A mutant retained all of the main-chain hydrogen-bond interactions with dimannose; however, it lost all of the side-chain interactions except OD1 of Asp-44, which explains the slightly higher decrease in binding affinity as compared with the E41A mutant (Figure 18 D). E41G displayed a similar hydrogen-bond network (Figure 19). The analysis shows that compared with the hydrogen-bond network of the P51G-m4-CVN-dimannose complex, N42A lost the highest number of hydrogen bonds (most of which were backbone hydrogen bonds) to dimannose. The loss of hydrogen bonds in T57A and E41A corresponded mostly to side-chain hydrogen bonds. These findings suggest that the identity of the side chains at positions 41 and 57 plays a critical role in dimannose binding. Conversely, our results suggest that N42 defines the integrity (i.e.,

the local stability) of the binding pocket by making critical side-chain-to-backbone hydrogen bonds to the neighboring β-sheet. Although polar-to-Ala mutations have disruptive effects on binding, we observe that conservative Asn to Ser mutation at position 53 preserves a hydrogen-bond pattern conducive to binding. The interactions of main-chain N of Asn-42, OE2 of Glu-41, OG1 of Thr-57, and NE2 of Gln-78 were lost, but additional side-chain interactions of N Asp-44, O of Ser-53, OD1 of Asp-44, and OG of Ser-53 were gained compared with P51G-m4-CVN-dimannose (Figure 18 E). Details of the hydrogen-bond analysis are provided in Table 5.

One common approach in ensemble docking is to use snapshots of the MD trajectories.[77,78] However, the success of this approach depends on whether the simulation covers the wide range of conformations sampled during binding. To gain a better idea of whether BP-Dock can help improve the binding affinity prediction beyond MD, we also performed ensemble docking using clustered snapshots of the REMD trajectories, and compared the results with those obtained by our flexible BP-Dock approach, which is based on a single mutant model. The binding affinities obtained from ensemble MD docking (Table 6) were all in the same range and could not really differentiate between a binder (i.e., N53S) and a nonbinder (N42A). Moreover, it did not capture the relative order of binding affinity (with N53S being the strongest, the binding score decreased in order from E41G to E41A and finally T57A, the weakest binder). However, BP-Dock approach captured the relative order of binding affinities observed experimentally and was capable of differentiating binders from nonbinders.

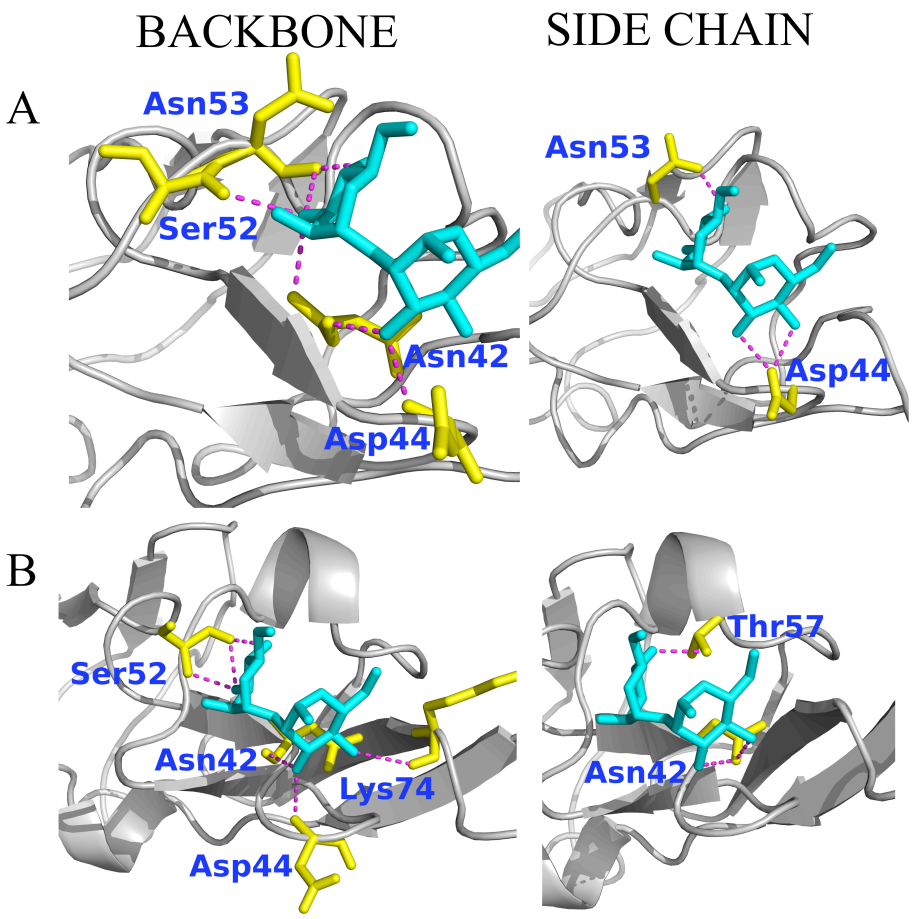


Figure 19. The hydrogen bond network for (A) E41A and (B) E41G:dimannose complex. The dimannose sugar is shown in cyan. The residues that form hydrogen bond with the sugar are shown in yellow and the backbone hydrogen bond interactions are shown in magenta and side chain interactions in green.

Table 5. Complete analysis of hydrogen bond network for P51G-m4-CVN and single point mutants for P51G-m4-CVN.

	Backbone				Side Chain			
	Residue	**Atom**	**H-Bond Distance**	**Role**	**Residue**	**Atom**	**H-Bond Distance**	**Role**
P51G-m4-CVN	Asn42	N	3.097	DONOR	Glu41	OE2	2.874	ACCEPTOR
	Asn42	O	2.839	ACCEPTOR	Thr57	OG1	2.827	ACCEPTOR
	Asp44	N	2.922	DONOR	Thr57	OG1	2.827	DONOR
	Ser52	O	3.139	ACCEPTOR	Arg76	NE	2.988	DONOR
	Asn53	O	2.967	ACCEPTOR	Arg76	NH2	2.871	DONOR
	Lys74	O	2.984	ACCEPTOR	Gln78	NE2	2.915	DONOR
CVN^{mutDB}	Ala42	O	2.948	ACCEPTOR	Glu56	OE1	3.269	ACCEPTOR
	Ala42	N	3.146	DONOR				
	Asp44	N	2.872	DONOR				
	Ser52	O	2.998	ACCEPTOR				
	Asn53	O	2.792	ACCEPTOR				
	Lys74	O	3.043	ACCEPTOR				
N42A	Ala42	O	3.083	ACCEPTOR	Asp44	OD1	2.965	ACCEPTOR
	Ser52	O	3.3	ACCEPTOR	Asp44	OD1	2.729	ACCEPTOR
	Asn53	O	2.883	ACCEPTOR	Asn53	OD1	2.881	ACCEPTOR
	ASN53	O	3.117	ACCEPTOR	Thr57	OG1	2.959	ACCEPTOR
T57A	Asn42	O	2.976	ACCEPTOR	Asp44	OD1	2.826	ACCEPTOR
	Asp44	N	2.997	DONOR	Asp44	OD1	3.476	ACCEPTOR
	Asp44	N	2.997	DONOR				
	Ser52	O	2.983	ACCEPTOR				
	Asn53	O	2.804	ACCEPTOR				
	ASN53	O	2.805	ACCEPTOR				
	Lys74	O	2.758	ACCEPTOR				
E41A	Asn42	N	3.201	DONOR	Asp44	OD1	2.811	ACCEPTOR
	Asn42	O	2.669	ACCEPTOR	Asp44	OD1	2.97	ACCEPTOR
	Asp44	N	3.216	DONOR	Asn53	OD1	2.855	ACCEPTOR
	Ser52	O	3.523	ACCEPTOR				
	Ser52	O	2.99	ACCEPTOR				
	Asn53	O	2.891	ACCEPTOR				
	Asn53	O	2.913	ACCEPTOR				
E41G	Asn42	O	2.833	ACCEPTOR	Asn42	OD1	2.931	ACCEPTOR
	Asp44	N	3.227	DONOR	Asn42	OD1	2.793	ACCEPTOR
	Ser52	N	2.777	DONOR	Thr57	OG1	2.915	ACCEPTOR
	Ser52	O	2.828	ACCEPTOR	Thr57	OG1	2.915	DONOR
	Ser52	O	2.915	ACCEPTOR				
	Lys74	O	3.504	ACCEPTOR				
N53S	Asn42	O	2.966	ACCEPTOR	Asp44	OD1	2.774	ACCEPTOR
	Asp44	N	3.115	DONOR	Asp44	OD1	2.877	ACCEPTOR
	Asp44	N	3.147	DONOR	Ser53	OG	3.274	ACCEPTOR
	Ser52	O	3.204	ACCEPTOR	Ser53	OG	3.274	DONOR
	Ser53	O	2.733	ACCEPTOR	Arg76	NE	2.78	DONOR
	Ser53	O	2.811	ACCEPTOR	Arg76	NH2	2.689	DONOR
	Lys74	O	2.857	ACCEPTOR				

Table 6. Comparison of Ensemble docking using MD snapshot vs BP-Dock cluster docking. The binding energy scores (E_{bind}) by X-score and corresponding RMSDs of the ligand ($RMSD_{ligand}$) and protein side chain positions in the binding site ($RMSD_{sidechain}$) for various CVN mutants are shown. Higher negative values indicate a higher binding affinity prediction.

	MD Ensemble Docking			BP-Dock Cluster docking		
Protein	$RMSD_{ligand}$ (Å)	$RMSD_{sidechain}$ (Å)	X-Score (kcal/mol)	$RMSD_{ligand}$ (Å)	$RMSD_{sidechain}$ (Å)	X-Score (kcal/mol)
N53S	0.14	1.238	-6.92	0.88	1.365	-7.06
E41A	0.22	1.511	-6.19	0.74	1.155	-6.7
E41G	1.51	1.059	-6.03	0.82	1.261	-6.82
N42A	1.16	1.835	-5.93	1.08	1.808	-5.37
T57A	1.59	1.124	-6.4	0.99	1.08	-6.46

Our biophysical characterization by CD, thermal denaturation experiments, and 2D-NMR showed that the N42A and T57A mutations slightly affected the stability of the protein. As observed in P51G-m4-CVN, N53S bound dimannose with dissociation constants in the low micromolar range and in intermediate exchange conditions, indicative of tight binding. On the other hand, a single point mutation of E41A, E41G, and T57A resulted in much weaker dissociation constants (in the low millimolar range)

and in fast exchange on the NMR timescale. However, mutation N42A abolished binding to dimannose.

Overall, our experimental characterizations of these five mutants show that BP-Dock can capture the binding-affinity pattern. The X-Score binding energy scores suggest that the mutations of E41A, E41G, and T57A resulted in a lower binding affinity for dimannose compared with P51G-m4-CVN, indicating the critical role of these side chains in glycan recognition. Mutation N42A was more disruptive than the other mutations, with a binding score closer to that of the nonbinder CVN^{mutDB}.

1.3.6. Conclusions

This work highlights the role of side-chain interactions in stabilizing the complex of dimannose and P51G-m4-CVN, with N42, T57, and E41 emerging as key factors for strong binding to dimannose. As shown here, N42 also appears to stabilize the binding pocket; its absence alters the hydrogen bond network within the protein and the ligand, thus abolishing binding. These effects add to other previously identified key interactions in CVN, such as the hydrogen bond networks with the side chain of Arg-76[42,73] and the main-chain[74], suggesting that a complex interplay between hydrogen bonds contributes to the remarkable affinity and specificity of CVN for α(1,2)-linked mannose.

The experimental verification of flexible BP-Dock suggests that our approach could be utilized as an *in silico* screening tool to assess the effect of mutations on the binding affinity and specificity of CVN-based lectins as compared with the wt CVN.

1.4. Nested Cyanovirin-N, A Covalent Dimer Construct

1.4.1. Introduction

The previous studies in our laboratory and others have shown that single-site binding mutants do not confer anti-viral activity. Having only either domain A (CVN^{mutDB})[39] or domain B (P51G-m4-CVN)[18] causes a significant decrease in binding affinity for gp120 by ELISA and has no antiviral activity. Further studies have shown binding site mutants can have antiviral activity restored by dimerization. Two mutants used a hinge-region deletion or mutation to stabilize the domain-swapped dimer which had two domain B binding sites with domain A binding eliminated, ΔQ50-m4-CVN and S52P-m4-CVN, respectively.[28] These mutants showed increased antiviral activity over the monomer counterpart, P51G-m4-CVN, suggesting multivalent interactions are required for antiviral activity. These results were further corroborated by disulfide-linked monomers of a domain A binding site mutant, $[CVN^{\Delta A}]_{ssd}$, and a domain-swapped dimer domain B binding site mutant stabilized by the ΔQ50 deletion, $[CVN^{\Delta B}]_{dsd}$, that restored the antiviral activity of their defective monomer counterparts.[29] Through these studies, it was concluded that the multivalency of CVN is what confers antiviral activity.

Expanding on the idea of multivalency leads to antiviral activity, other laboratories have attempted to control dimerization in order to increase the number of binding sites in hopes of increasing antiviral activity. Instead of using defective binding site mutants with dimer-stabilizing mutations, such as ΔQ50 of S52P, intact binding sites were combined with the dimer-stabilizing mutations to produce domain-swapped dimers with four binding sites.[19] These domain-swapped dimers, [ΔQ50]CVN and [S52P]CVN, were

found to have antiviral activity only comparable to that of wt CVN.[19] It is very probable, though, that these mutants had similar antiviral activity to wt CVN due to their decreased stability compared to that of wt CVN. These results, along with the inability to control each binding site individually in a domain-swapped dimer, lead to these dimer mutants being insufficient to study the multivalency of CVN further.

In order to better explore multivalency, the Mayo Laboratory designed covalent dimers of CVN by combining, in tandem, two CVN sequences with varying length polypeptide linkers.[32] These CVN_2L_x (where x is the number of amino acids in the linker) covalent dimers were shown to fold in a similar manner to the domain-swapped dimer by using partial carbohydrate binding site mutants and testing the antiviral activity. By including partial binding site mutants on one monomer sequence and the corresponding binding site mutations on the second monomer sequence, if the protein folded in a domain-swapped manner, the protein would not confer antiviral activity; and indeed, the antiviral assays showed this to be true.[32] Interestingly, the fully intact covalent, tandem dimers, CVN_2L_x, showed 3- to 6-fold enhancement of antiviral activity over the monomer of wt CVN.[32] However, it is not clear whether or not the covalent, tandem dimer sequences would form higher order oligomers (domain swapping with other sequences for example) which could have a greater effect on the antiviral activity; although, increasing the number of CVN sequences linked in tandem did not confer higher antiviral activity.[32]

1.4.2. Background On Cyanovirin-N Homologous (CVNH) Family

When CVN was discovered, it displayed a unique sequence and fold not yet found in other genes or proteins. Using sequence alignment and fold-prediction algorithms, a family of CVN homologous proteins, designated CVNH, was found in ferns and fungi.[71] Within this family, there were three types of proteins with similar predicted fold to that of CVN, as depicted in Figure 20. Type I CVNH are comprised of two tandem repeats of CVNH domains, while type II has one CVNH domain that is preceded by a portion of a mold specific gene of unknown function, MS8-h.[71] Type III CVNH also have a MS8-h encoding region at the N-terminus, but are followed by a CVNH domain that is split by a LysM encoding region.[71] The MS8 gene, of which part is found in Type II and III CVNH, is a sequence that encodes a predicted 21 kDa protein that is essential for proper hypha formation in pathogenic fungus.[79] The LysM sequence is a Lysin Motif for a domain that recognizes glycans, peptidoglycans, and chitins; and is typically a domain within lysins, peptidoglycan hydrolases, peptidases, chitinases, esterases, reductases, and nucleotidases.[80]

The Gronenborn laboratory has solved structures and study the function of a few CVNH proteins.[38,76,81] These proteins, as predicted, display remarkable similarity to CVN overall, with differences occurring in loop regions and local, fine structure.[38] All CVNHs with solved structures have also had at least one glycan binding site, albeit not as specific as the binding site of wt CVN.[38,76] More interestingly, the type III CVNH, in which a LysM domain splits two 50 amino acid repeats of CVN, has also had its structure solved by NMR and still retains the binding function of both the CVN domain and the LysM

domain.[81] As seen in Figure 21, the LysM domain (green) folds and enables the separated 50 amino acid repeats of CVN (first repeat is blue, second repeat is yellow) to come back together and fold into a structure closely matching that of wt CVN. Interesting to note, the two domains behaved as two independent domains and the conserved glycine-rich linkers that link the N- and C-terminus of the LysM domain to the CVN domain were shown to be quite flexible.[81]

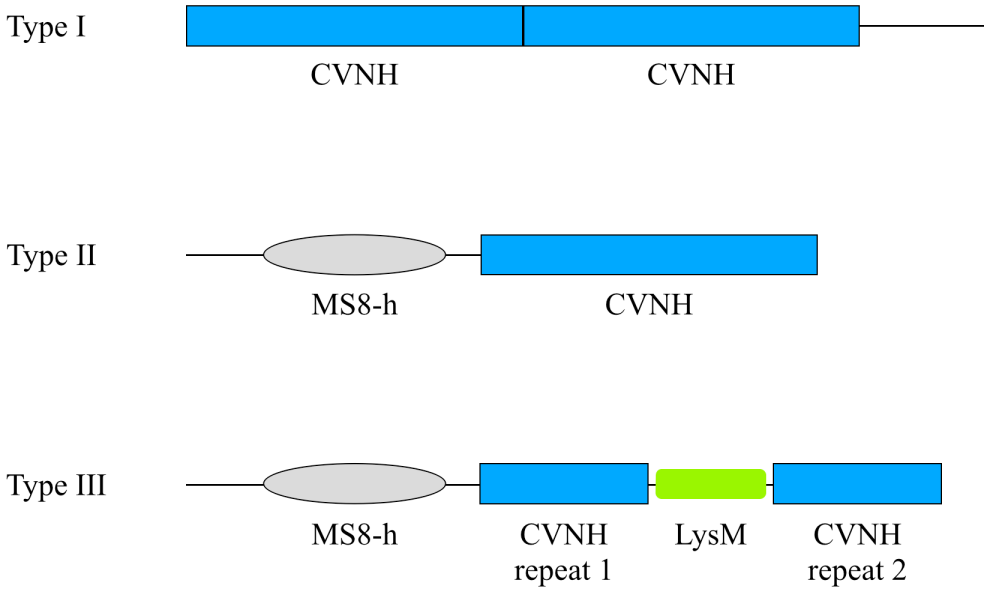


Figure 20. Schematic representation of the three types of the family of cyanovirin-N homologous proteins. Domains are labeled beneath the pictorial representation of the organization in each type. Type I consists of tandem repeats of CVNH (blue), whereas type II and type III are preceded by a partial MS8 gene (gray oval). Type III CNVHs have a LysM inserted between 50 amino acid repeats of CVNH (green).

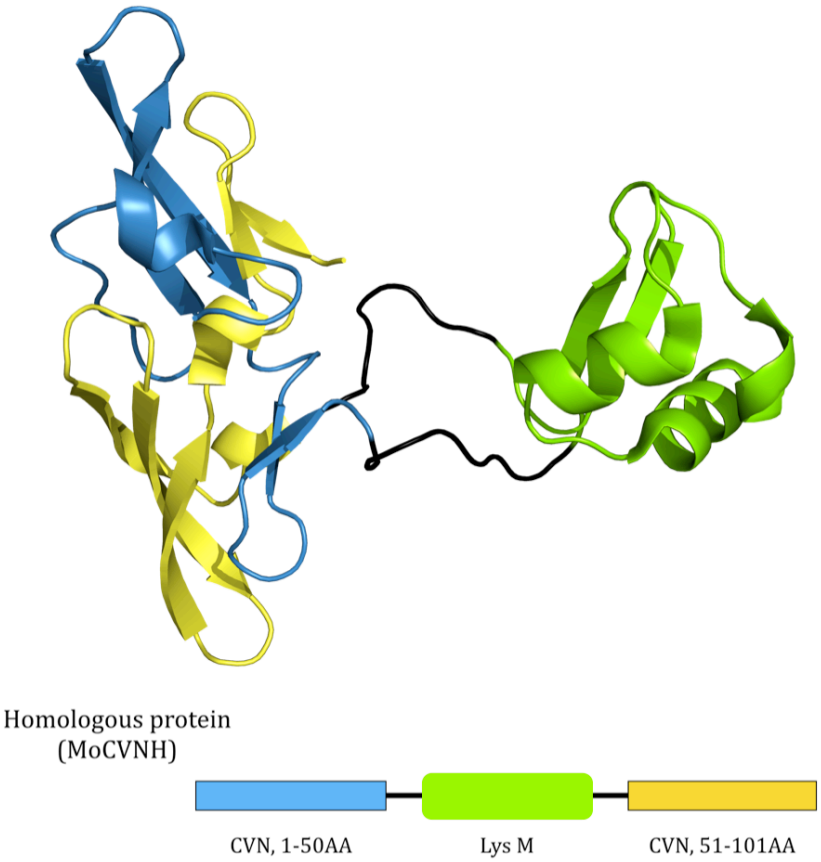


Figure 21. Structure of MoCVNH, a type III CVNH, from PDB ID 2I9Y.[81] The first 50 amino acid repeat of CVN is shown in blue, the LysM in green, the second 50 amino acid repeat of CVN in yellow, and the linkers in black, both in the structure and the domain representation beneath.

1.4.3. Design of Nested Cyanovirin-N

As previously suggested, dimerization and the geometry of the binding sites may be important in conferring antiviral activity, see Figure 6, panel b.[32] Disulfide-linked monomers may not increase the antiviral activity of CVN due to the lack of meeting

probable geometric requirements. The end-to-end, tandom-linked dimers designed by the Mayo group most likely meet these geometric requirements, but domain-swapping with other units of CVN to create higher order oligomers is very possible and likely at higher concentrations.[31,32] In order to further study multivalency and to design a dimer with higher antiviral activity, we aim to design a covalent dimer in which we can control each binding site individually.

Nested CVN was designed on the basis of type III of the CVN homologous family.[71] Using the idea that a full-length CVN sequence could be interrupted, by LysM in the case of type III CVNH, and still fold in a manner that still binds glycans, Nested CVN replaces the LysM domain with another full sequence of CVN, utilizing the same glycine-rich linkers found in type III CVNH, as shown in Figure 22. Constructs were originally designed to include the "P51G" corresponding hinge region mutation in the CVN insert; however, there is recent evidence suggesting the P51G mutation causes a decrease in binding affinity for dimannose (unpublished data). For this reason, the wild hinge region (Nested-P CVN) was introduced back into the insert to test both constructs.

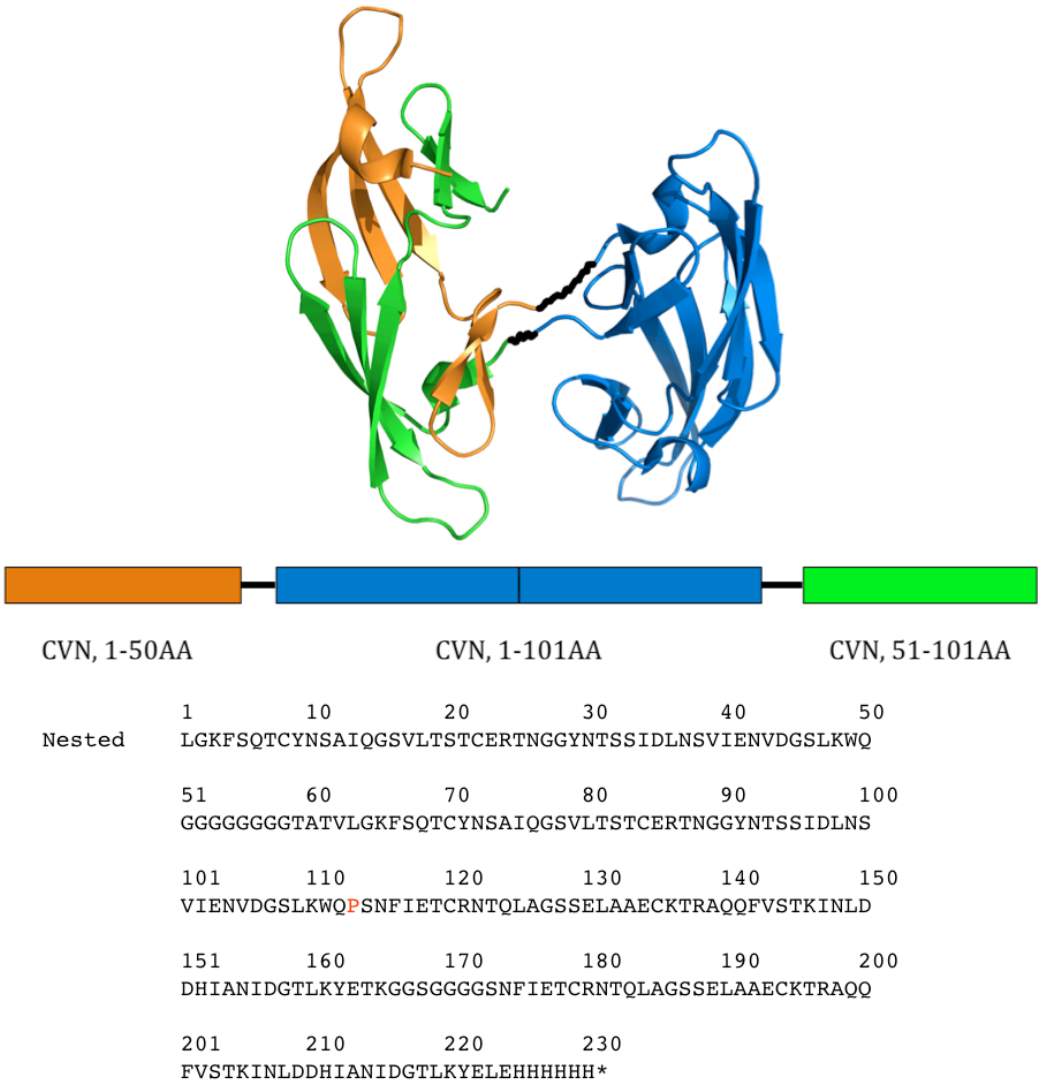


Figure 22. Nested CVN model and sequence are shown. The CVN insert is shown in blue, following the C-terminus of the first 50 amino acid repeat (orange) and preceding the N-terminus of the second 50 amino acid repeat (blue). Each region is connected by a glycine-rich linker, show in black. The sequence of Nested CVN is shown with the corresponding "P51" position in the hinge region of the insert highlighted in red.

1.4.4. Materials and Methods

Genes with the corresponding sequences with pelB leader sequences preceding were ordered from GenScript (Piscataway, NJ) and double digested with NdeI and XhoI. Inserts were ligated into pET26b(+) vectors following the Novagen pET system manual. The vector was transformed into NEB5α cells and BL21 cells from New England Biolabs, Inc. (Ipswich, MA). The protein was expressed from BL21 cells by inoculating 10 mL Luria-Bertani (LB) medium with incubation at 37˚C and shaking at 250 rpm. The cells were grown overnight and then added to 1 L of fresh LB medium. The culture was incubated at 37˚C and 250 rpm shaking until an OD_{600} of approximately 0.8. Protein expression was induced with 0.4 mM final concentration of isopropyl β-D-thiogalactoside (IPTG) and 20 mM benzyl alcohol. Incubation and shaking of the cell culture was continued overnight at 20˚C. When the culture reached an OD_{600} of 1.3-1.5, cells were harvested by centrifugation at 6,000 rpm for 10 minutes.

The periplasmic fraction was collected by resuspending the cell pellet into 30 mM Tris, 20% sucrose, 0.5 mM EDTA, pH =7.5 with stirring at room temperature for 10 minutes. The cells were pelleted by centrifuging at 10,000 rpm at 4˚C. The cells were resuspended in ice cold 5 mM $MgSO_4$ by stirring on ice for 20 minutes. The cells were once again pelleted by centrifugation at maximum speed for 10 minutes, and the supernatant was saved as the periplasmic fraction. The cytoplasmic fraction was collected by sonicating the cells on ice for a total of 20 minutes. This suspension of broken cells was then centrifuged at 15,000 rpm for 20 minutes, saving the supernatant to check for target protein in SDS-PAGE Tricine gels. The pellet of broken cells was resuspended in

100 mM Tris pH=7.5, 2 M urea to wash away any unwanted proteins. The suspension was centrifuged at 15,000rpm for 20 minutes. The 2 M urea wash was repeated until the supernatant was clear after centrifugation. The broken cells were rinsed by suspension in 100 mM Tris and centrifuged at 15,000 rpm for 20 minutes. To collect the any inclusion bodies, the broken cells were dissolved in denaturing conditions, 8 M urea, 20 mM Tris, 20 mM imidazole, and 0.5 M NaCl, pH = 7.5, and centrifuged at 16,500 rpm for 45 minutes. The supernatant was collected, and all fractions were analyzed by Tricine SDS-PAGE for target protein.

Once the target protein was located, the supernatant was filtered and loaded onto a nickel sepharose column (Ni Sepharose High Performance medium from GE Healthcare). The protein was eluted off the column using 400 mM imidazole under native conditions. A Tricine SDS-PAGE gel was performed to determine the fractions in which the target protein was located. To confirm the correct molecular weight of the protein, the protein was run on MALDI-TOF mass spectrometry, using sinapinic acid matrix. Protein concentrations were determined by measuring tryptophan absorbance at 280 nm using an extinction coefficient of 19,540 M^{-1} cm^{-1}. The yield of protein was approximately 10 mg protein per liter of LB medium.

Circular dichroism was utilized to estimate secondary structure and as a preliminary measure to check protein folding. All measurements were performed using a Jasco J-815 spectropolarimeter equipped with a thermostatic cell holder, PTC 424S. Spectra were measured from 240 nm to 190 nm, using a scanning speed of 50 nm/min and a data pitch of 1.0 nm at 15˚C.[82] A 0.1 cm quartz cell was used, with protein concentrations at

approximately 15 μM in 10 mM HEPES, pH = 8.0 buffer. Thermal denaturation studies were also performed by taking full spectra using the previous settings at approximately every 2˚C from 4˚C to 90˚C. Detailed thermodynamic analysis was precluded due to the irreversible unfolding of Nested CVN, which matches that of wt CVN. Apparent T_m was determined by fitting the data to equation 5:

$$\theta = \frac{\left(m_f x + b_f\right) + \left(m_u x + b_u\right) \times e^{-\left(\frac{\Delta H}{R \times \left(\frac{1}{x} - \frac{1}{T_m}\right)}\right)}}{1 + e^{-\left(\frac{\Delta H}{R \times \left(\frac{1}{x} - \frac{1}{T_m}\right)}\right)}} \tag{5}$$

Enzyme-linked immunosorbent assays (ELISA) were utilized to probe how tightly the Nested CVN constructs bound to soluble gp120 (sgp120) compared to wt CVN. The experiment was performed as previously described.[9] In summary, each well except for the last column of 8 wells, of a 96-well plate (Nunc; Maxiscorp, Gaithersburg, MD) was incubated with 100 ng of gp120 in phosphate buffered saline (PBS) for two hours at room temperature. The plates were then rinsed with PBS containing 0.1% Tween 20 (PBST) three times. The plates were then blocked with 3% bovine serum albumin (BSA) at 4˚C overnight. The plates were then washed three times with PBST. Half log serial dilutions of Nested CVN, Nested-P CVN, and wt CVN were then added to the wells in triplicate, for final concentrations ranging from 0.005 nM to 0.05 μM. The last column was left as a blank, and only contained the BSA. The plates were incubated for one hour, and then

washed three times with PBST. Each well was then incubated anti-CVN rabbit polyclonal antibody (final concentration of 10 μg/mL) for 1 hour and then washed with PBST. The plates were then incubated with a secondary anti-rabbit IgG antibodies conjugated to horseradish peroxidase (Sigma, St. Louis, MO) for 1 hour. The plates were then washed again three times with PBST. 100 μL of tetramethylbenzidine peroxidase substrate was then added to each well and allowed to react for approximately 2 minutes until a blue color was observed. The reaction was then quenched with 50 μL of 2 M H_2SO_4, and the absorbance at 450 nm was recorded.

Anti-HIV activity assay of Nested CVN and Nested-P CVN were performed at NCI-Fredericksburg. In short, a modified XTT assay was utilized to determine the anti-HIV activity of Nested CVNand Nested-P CVN, alongside a recombinant CVN (rCVN) that was provided by NCI-Fredericksburg. Experimental procedure involved plating uninfected CEM-SS cells at a density of 1 x 10-5 cells/mL in 50 μL of complete medium in a 96-well plate. Serial dilutions of 100 μL of rCVN, Nested CVN, and Nested-P CVN were added to the wells. The cells were then infected with 50 μL of HIV-1RF virus. The plates were incubated at 37˚C in an atmosphere containing 5% CO2 for 6 days. To estimate cellular viability, 50 μL XTT solution was added to each well and incubated for 4 hours at 37˚C. Cells that had not been infected with the HIV virus could reduce XTT to a soluble, colored formazan while HIV-infected cells could not. Metabolic reduction of XTT to formazan was recorded by absorbance at 450 nm to determine cellular viability.[83]

Isothermal titration calorimetry (ITC) was utilized to ensure that binding sites were still intact and had high affinity for dimannose. To this end, samples were sent to

Sanford-Burnham Medical Research Institute. In short, 3.6 mM dimannose was titrated into 100 μM Nested CVN samples in 20 mM HEPES for high affinity binding site determination, while the low affinity binding site utilized 20 mM dimmanose. Data was fit using common procedures, allowing number of binding sites as a fitting parameter. In some cases, the number of binding sites had to be fixed at four (or two in the case where the high affinity set and low affinity set binding K_d were determined in separate experiments) in order to obtain a proper fitting.

1.4.5. Biophysical Characterization

Using the pelB/pET expression system, soluble protein yields were routinely around 10 mg per liter of broth. To show that target protein was over-expressed, SDS-PAGE was run on pre-induction and post-induction samples (lanes 2 and 3 in Figure 23). SDS-PAGE was also utilized to make sure the protein was localized in the perisplasmic fraction. To ensure the proper fractions were collected for further characterization, and as a preliminary check of purity, SDS-PAGE was utilized again. Figure 23 shows a typical gel, and only fractions that were deemed to have high amounts of target protein with no or little impurities were used for dialysis and further characterization.

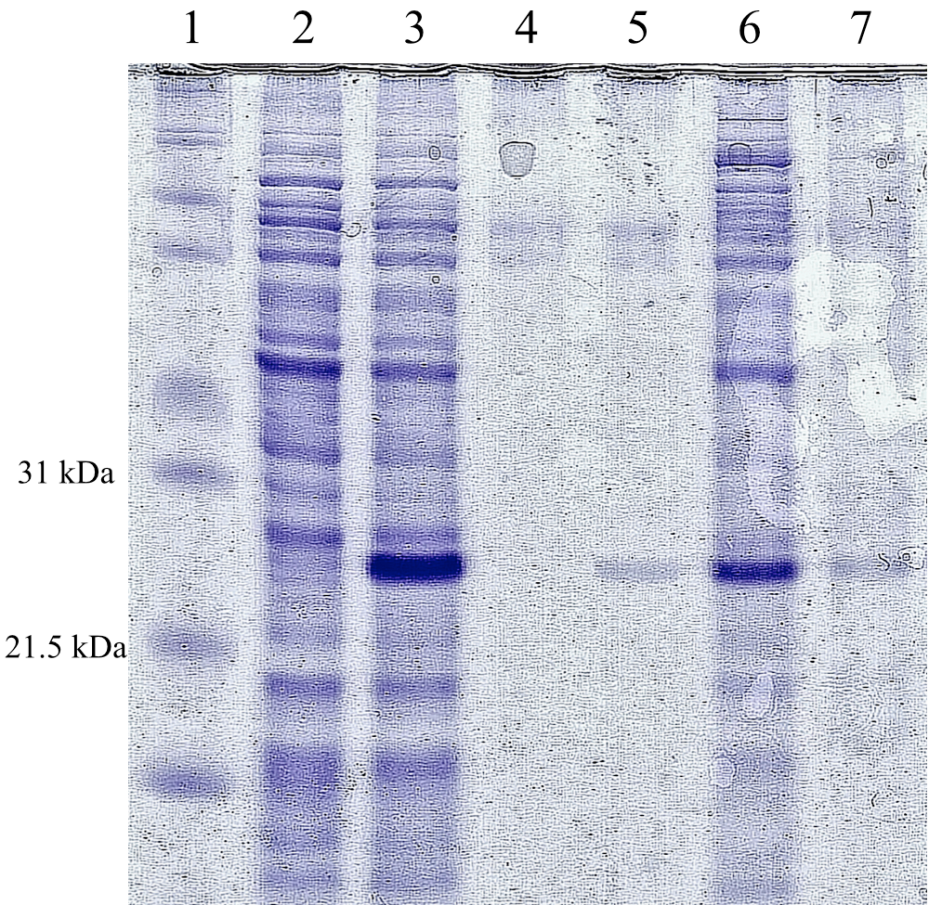


Figure 23: Tricine SDS-PAGE gel of Nested CVN. Important molecular weight markers are labeled in which the calculated weight of Nested CVN appears between. Lane 1 - Mark 12 Unstained standard (Life Technologies); Lane 2 - pre-induction; Lane 3 - post-induction; Lane 4 - sucrose wash; Lane 5 - cytoplasmic fraction; Lane 6 - periplasmic fraction; Lane 7 - inclusion bodies.

To verify that the Nested CVN and Nested-P CVN were expressed correctly, a more accurate molecular weight determination needed to be performed. MALDI-TOF was performed, using sinapinic acid as the matrix, for both Nested CVN and Nested-P CVN.

Nested CVN, was verified by a calculated molecular weight of 24,423 Daltons and Nested-P CVN, was verified by a calculated molecular weight of 24,463 Daltons. Every expression of both proteins' molecular weights was always verified by MALDI-TOF before any further characterization.

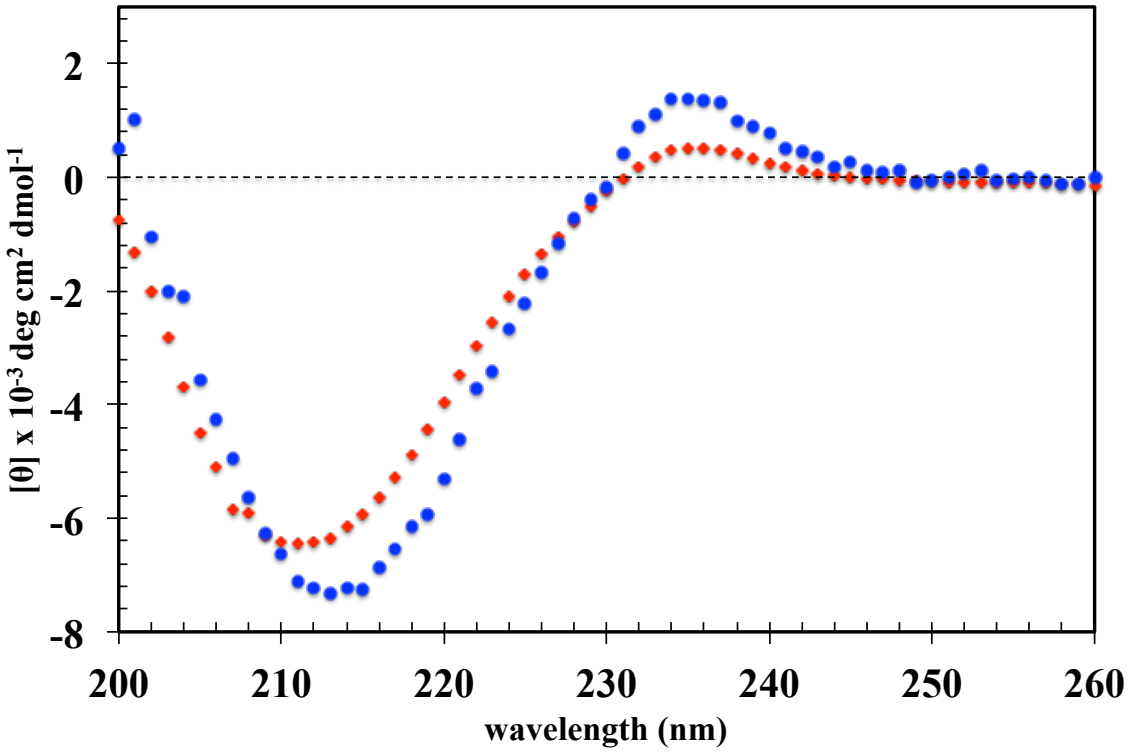


Figure 24: Overall CD spectra of Nested CVN (blue) and Nested-P CVN (red) measured at 4˚C.

To ensure that the Nested constructs were properly folded, circular dichroism was utilized to make inferences about the secondary structure and the resulting spectra were compared to previous spectra of wt CVN. Figure 24 shows the overall CD spectra of Nested CVN and Nested-P CVN. The spectrum for the protein has minimum at

approximately 212 nm and a maximum around 192 nm. The spectrum mostly resembles a β-sheet protein, which is to be expected since the majority of Nested CVN is β-sheet.

To show that the Nested constructs were stable proteins, a thermal denaturation experiment was performed. Figure 25 shows the fraction folded versus temperature at a wavelength of 235 nm for both constructs. Commercially available software (Prism 6, GraphPad Software) was used to fit the sigmoidal curves and determine the midpoint for the melting points of Nested CVN and Nested-P CVN. The average melting point from the two wavelengths used was 56.3˚C, which is stable compared to that of wt CVN at 59.0˚C.[28]

Nested CVN's and Nested-P CVN's interaction with gp120 were measured using ELISA. Wild type CVN was also used, for comparison and as a control as binding to gp120 has been previously described.[9] The results are shown in Figure 26.

The data was fit using commercially available software as described earlier, using the Hill-Slope model. EC_{50} for Nested CVN was determined to be 3.47×10^{-10} M ± 0.33×10^{-10} M. Nested CVN without histag gave similar results (data not shown). EC_{50} for wt CVN was determined to be 3.33×10^{-9} M ± 0.23×10^{-9} M. This is a worse EC_{50} for wt CVN than is typically observed, but highlights some of the shortcomings in reproducibility of ELISA experiments. The data for the EC_{50} of these constructs of CVN, along with other mutants, is summarized in Table 7.

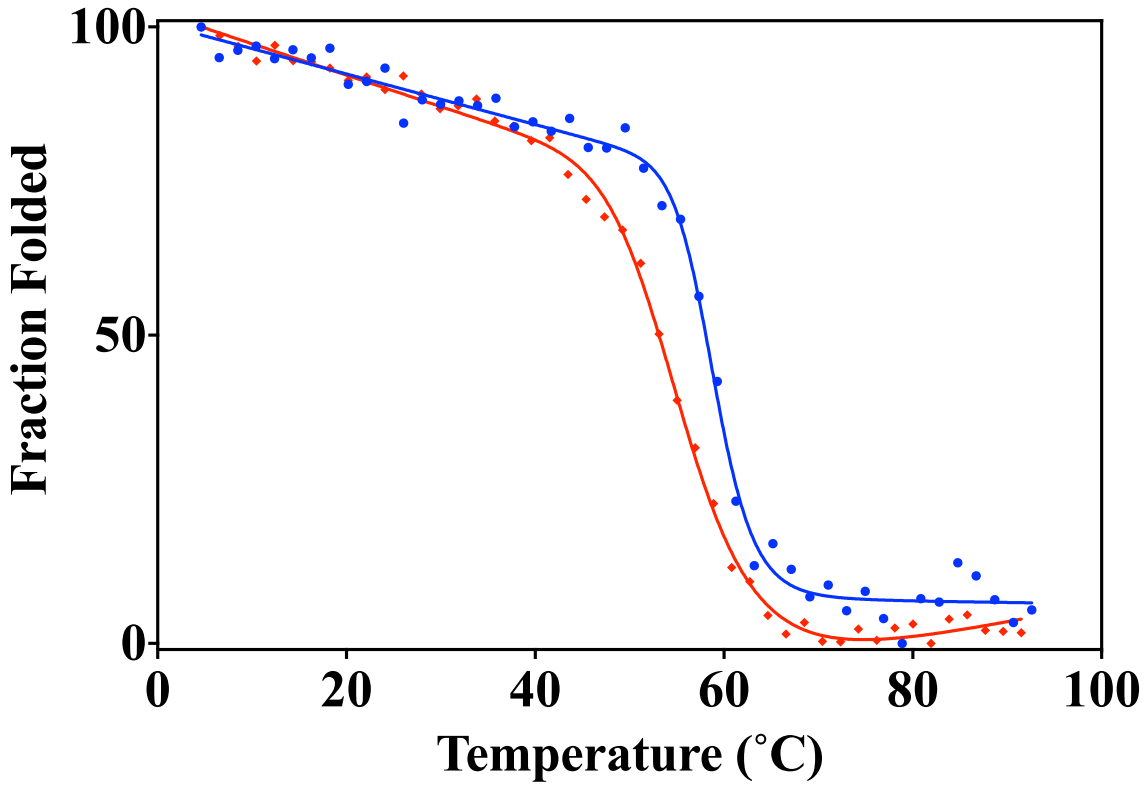


Figure 25: Thermal denaturation curves for Nested CVN (blue) and Nested-P CVN (red). The plots follow the wavelength 235 nm and the data was fit using Prism 6 from GraphPad Software. The T_m for Nested CVN and Nested-P CVN were 59.0˚C and 56.3˚C, respectively. These variants of CVN are stable, similar to wt CVN.

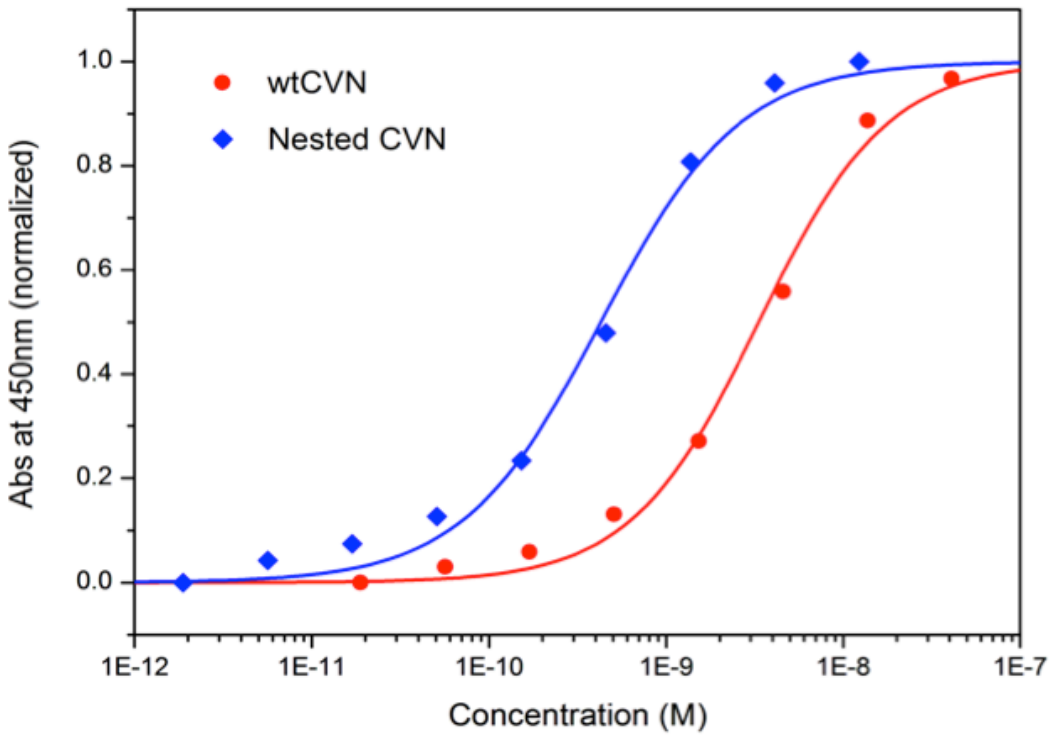


Figure 26: ELISA results of the binding of wtCVN (red circles) and Nested CVN (blue diamonds) to gp120. Serial dilutions of proteins were added to plates coated with gp120 (100 ng/well). For both proteins, polyclonal rabbit anti-CVN antibodies were used to detect bound proteins followed by incubation with anti-rabbit IgG-Peroxidase antibodies; the peroxidase catalyzed oxidation of TMB substrate was monitored at 450 nm.

Table 7: Binding of CVN mutants to HIV gp120

Mutant	EC_{50}, nM
Nested CVN	0.347
wt CVN [A]	0.1
P51G-m4-CVN [B]	173
ΔQ50-m4-CVN [B]	23
S52P-m4-CVN [B]	9.6
m4-CVN [A]	0.6

[A] From reference [18].

[B] From reference [28].

To evaluate the anti-HIV activity of the Nested constructs, a modified XTT-tetrazolium assay was performed. Nested CVN and Nested-P CVN were tested against rCVN. The results of the assay and the fits used to determine the EC_{50} are shown in Figure 27.

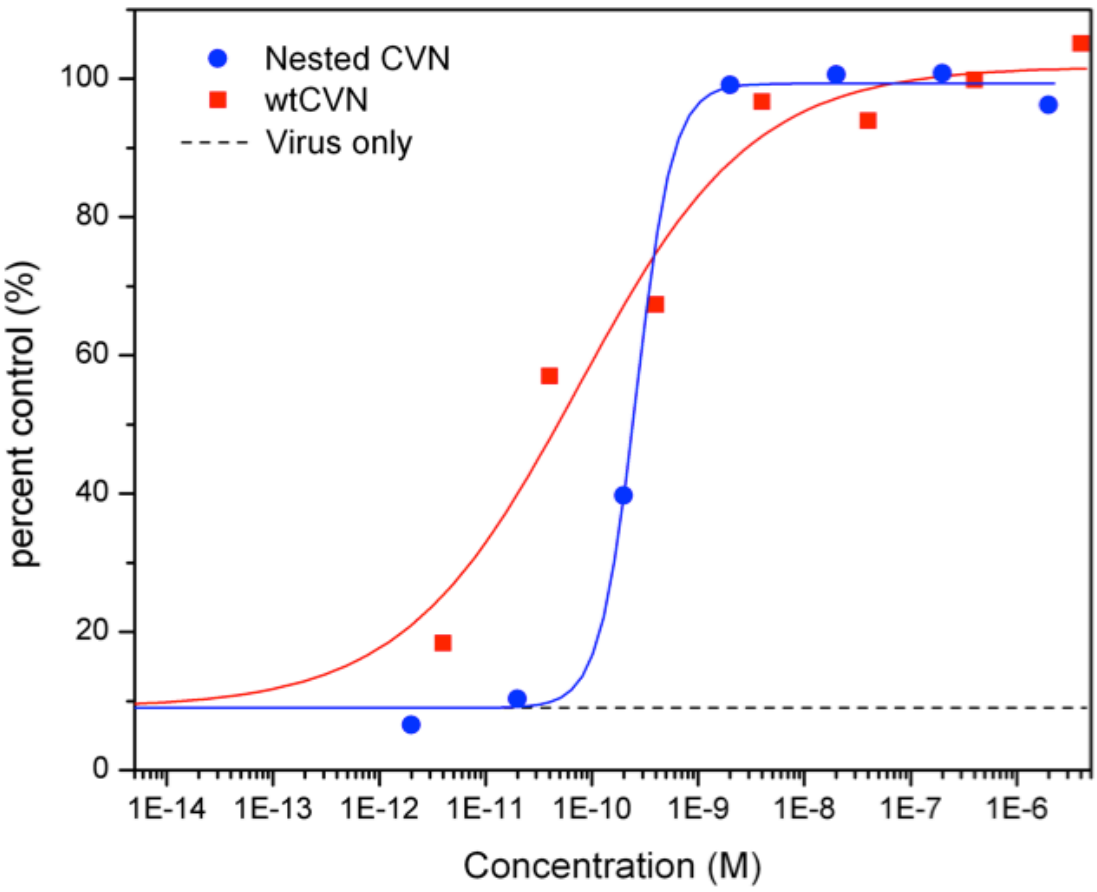


Figure 27: Anti-HIV activity of Nested CVN and wt CVN. Logistic fitting in OriginPro to provide the fit and determine the EC_{50} was used. The bottom boundary of the logistic fit was fixed at the percent control of cells that had only virus added, without any inhibitors. The value for Nested CVN EC_{50} was determined to be $2.60 \times 10^{-10} \pm 0.59 \times 10^{-10}$ M and for CVN was determined to be $7.41 \times 10^{-11} \pm 3.99 \times 10^{-11}$ M.

The values determined for the EC_{50} in the anti-HIV activity of Nested CVN and wt CVN are summarized in Table 8, along with other mutants that have been tested for anti-HIV activity. It is clear that Nested CVN has comparable anti-HIV activity to rCVN, 0.26

nM versus 0.07 nM, respectively. The rCVN anti-HIV activity is a little stronger than typically reported for wt CVN, but is still comparable.

Table 8: Anti-HIV activity of Nested CVN and other mutants

Mutant	EC_{50}, nM
Nested CVN	0.26
wt CVN [A]	0.9
P51G-m4-CVN [B]	>>5,000
ΔQ50-m4-CVN [B]	320
S52P-m4-CVN [B]	≈3000
m4-CVN [A]	240

[A] From reference [18].

[B] From reference [28].

To gain further information about the binding sites and the affinity for dimannose of Nested CVN and Nested-P CVN, isothermal titration calorimetry was performed at Sanford-Burnham Medical Research Institute. The data and fits are shown in Figure 28. In short, the data could be fit with by using a 2 x 2 binding site model; which is expected. Nested CVN had two high affinity binding sites with a K_d of 24.5 μM ± 1.8 μM and two low affinity binding sites with a K_d of 952 μM ± 7.9 μM; while Nested-P CVN had two high affinity binding sites with a K_d of 15.7 μM ± 1.2 μM and two low affinity binding sites with a K_d of 1420 μM ± 16 μM.

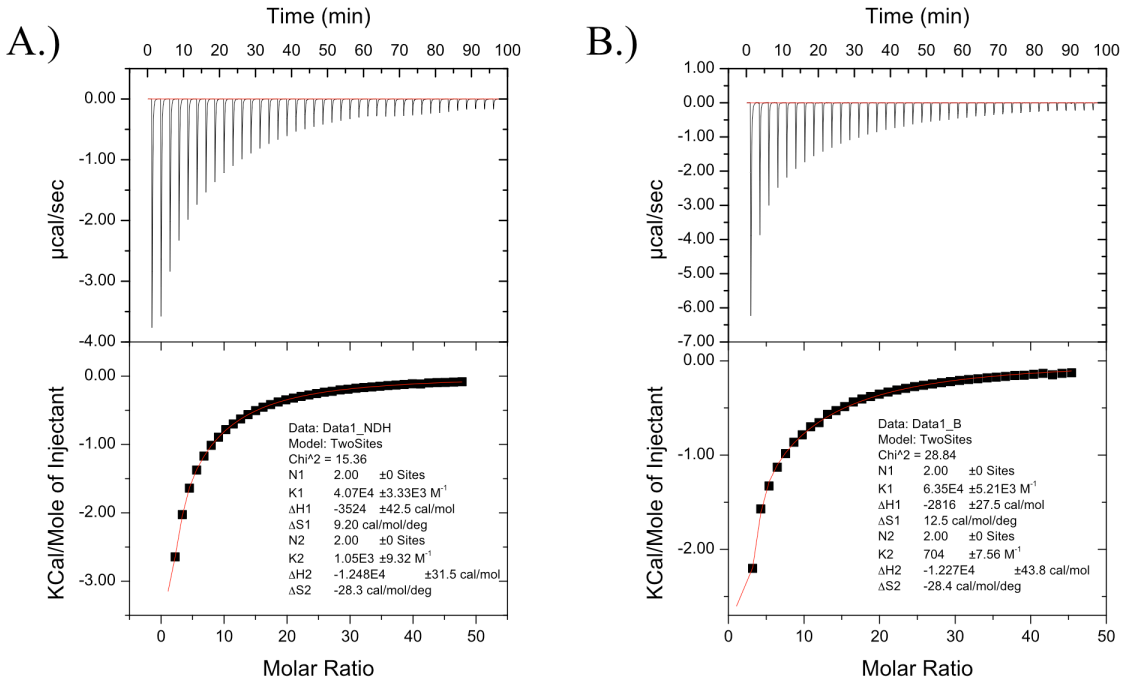


Figure 28. ITC data from Nested CVN (A.) and Nested-P CVN (B.) titrated with dimmanose.

1.4.6. Discussion

Nested CVN and Nested-P CVN has provided insight into the interaction of Nested CVN with gp120 and the role in anti-HIV activity. As a designed covalent dimer, it offers insight into the role of dimerization and multivalency. All data is summarized in Table 9. With the procedure described above, we have routinely been able to produce 30 mg of protein per 500 mL broth. The protein was shown to be fairly stable, in that the melting temperature was found to be 55˚C.

Table 9: Summary of all mutant data

Mutant	state	T_m (˚C)	EC_{50} binding to gp120 (nM)	EC_{50} antiviral activity (nM)
wtCVN	wild type	58	0.1	0.9
m4-CVN	domain A binding mutant	46	0.6	240
P51G-m4-CVN	monomer, domain A binding mutant	64	173	>>5000
ΔQ50 -m4-CVN	dimer, domain A binding mutant	32.5	23	320
S52P-m4-CVN	dimer, domain A binding mutant	39.5	9.6	~3000
Nested CVN	covalent dimer	~58	0.423	423

With the results presented here, it is evident that we designed a covalent dimer of cyanovirin-N, which has comparable potency the wild type protein in binding to gp120 and in antiviral activity assays. In our ELISA study, our findings for Nested CVN binding to gp120 were equivalent to that of wt CVN. Previous reports of wt CVN have an EC_{50} of approximately 0.1 nM, while our study showed an EC_{50} approximately one-fold higher.[9] However, we do believe that this shows that Nested CVN binds gp120 at least as well as wt CVN.

While previous studies have shown that only two binding sites are required for antiviral activity, we show here that four binding sites may change the way in how the

protein binds to gp120 of the Env of HIV.[28] In the case of antiviral activity, it is evident that Nested CVN inhibits HIV fusion differently than wt CVN based on the Hill coefficient of the plot. Nested CVN displays a Hill coefficient of 2.52, while wt CVN has a Hill coefficient of 0.53. Nested CVN then shows cooperative binding in that when one binding site binds, it increases the binding of the next binding site. This is expected, since the molecule is a covalent dimer with four binding sites, when one site binds, the next binding site is already in close proximity to another Manα(1-2)Manα moiety.

The ITC results suggest that all the binding sites are intact and bind dimannose for both Nested CVN and Nested-P CVN. The low micromolar value for the high affinity binding site matches that of wt CVN ITC values obtained by our group (data not shown). The low affinity binding site for the two Nested constructs are approximately 1 mM which is about two times higher than observed by our group for wt CVN (data not shown). Unfortunately, ITC data cannot unequivocally show that all four binding sites are intact as the data for Nested-P CVN can also be fit with one low affinity binding site.

In conclusion, previous mutants of Cyanovirin-N have been used to show that multivalency is needed to confer anti-HIV activity.[18,28] Here, we have shown that covalent dimer of CVN is as stable and as effective at binding to gp120 and in antiviral activity assays. It is also evident that it changes the interactions with the Env protein in antiviral activities, as reflected by the Hill coefficient. Multivalency is required for antiviral activity, and in this study, we show that a covalent dimer that is multivalent confers comparable results to that of wild-type cyanovirin-N.

2. WW Domain

2.1. Introduction

WW domains are a conserved module of 32-40 amino acids found in many different proteins and provide signaling, regulatory, and cytoskeletal function.[84-87] Many studies show that the WW domain recognizes proline-rich peptide regions of larger proteins.[88-91] Due to the computational and experimentally facile nature of the WW domain, the Ranganathan group used WW domains as a model system to study and create artificial sequences using statistical coupling analysis (SCA) and multiple sequence alignment (MSA).[92] By randomly selecting amino acids at each site from natural WW MSA, the group created a series of sequences they termed IC sequences (site-independent conservation).[92] These IC sequences are built randomly, ignoring any statistical coupling. Using Monte Carlo simulation, the group also created a series of sequences which took into account statistical coupling they observed from the SCA; they termed these CC sequences (couple conservation).[92] The CC sequences were built taking into account that some positions coevolved with other positions. Using the natural, IC, and CC sequences, the Ranganathan group determined which sequences would fold by expression tests in *E. coli* and then further experiments using ^{1}H-NMR and fluorescent thermal denaturation studies.[92] Using these methods, the Ranganathan group showed that coupled conservation (pairwise coevolution) was important as some of the designed CC sequences would fold, while none of the IC sequences folded into WW domains.[92] Indeed, the structure of one of their designed sequences, CC45, was solved and showed to have the similar fold to that of two natural WW domains.[92] In a concurrent paper, the Ranganathan group also

showed that some of their designed sequences maintained function in that they would bind proline-rich peptides.[93]

Table 10. Organization of the WW domain sequences. N, native; CC, coupled conservation; and IC, site-independent conservation, sequences were grouped into foldable sequences if they were expressed solubly and a T_m or unfoldable sequences if they were expressed solubly but did not have a T_m.[92]

	N	CC	IC	Total
Foldable	28	12	0	40
Unfoldable	0	19	30	49

Using the sequences generated in the Ranganathan group paper, we collaborated with Dr. Ozkan's group to try to further determine the folding requirements encoded in natural sequences.[92] While the Ranganathan paper only considered statistical information in the sequence space, we aim to use tertiary structural information as well to explore the sequence-structure relationship in WW domain proteins. To this end, we used simulation data provided by zipping and assembly method (ZAM) and a contact probability metric (CPROB) to determine the difference between folders (those sequences that folded) and nonfolders (those sequences which were expressed and soluble, but did not fold into WW domains).[94,95] Our collaborators excluded the sequences from the Ranganathan study that were insoluble or poorly expressed because the reasons for poor expression and insolubility were beyond the scope of our study. From the remaining sequences, our

collaborators grouped the sequences into foldable and unfoldable sequences, those sequences deemed folded showed thermal denaturation profiles and ^{1}H-NMR spectra evident of the native state. Table 10 summarizes the sequences used in this study.

All the WW domain sequences in Table 10 were simulated using ZAM. ZAM is a method which simulates folding with small fragments of the sequences *in silico* and growing the fragments into the full sequence.[94] For the purpose of our study, 8-mer fragments were used. From these simulations, the contact probability (CPROB) within an 8-mer fragment was determined as the calculated fraction of sampled configurations with inter-residue distance less than 8.0 Å.[96] Using this data, a classification model was developed by our collaborators, and maximum likelihood contact probability (MLCPROB) maps were created, shown in Figure 29. From this data, the difference in contact probability between folded and unfolded sequences shows that there are five crucial contacts that exist in the simulated folding of WW domains. To show these contacts are crucial, unfolded sequences were simulated with these contacts artificially forced *in silico*; and through the simulation the unfolded sequences would fold into WW domain-like structures with these forced contacts (data not shown). To further verify these results, we checked for the crucial contacts in the ZAM simulated structures of unfolded and folded sequences. Table 11 shows that with relatively high efficiency, we can predict folded and unfolded sequences based on the five crucial contacts.

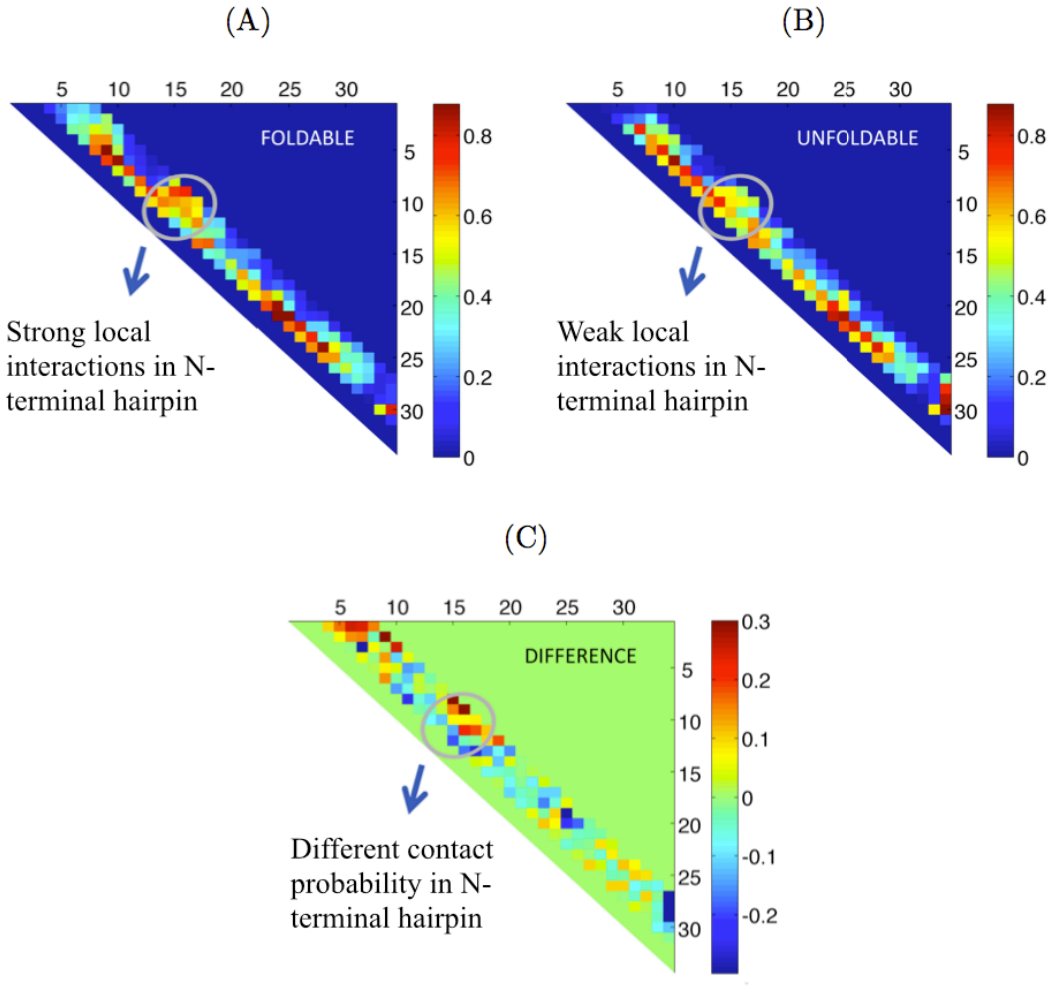
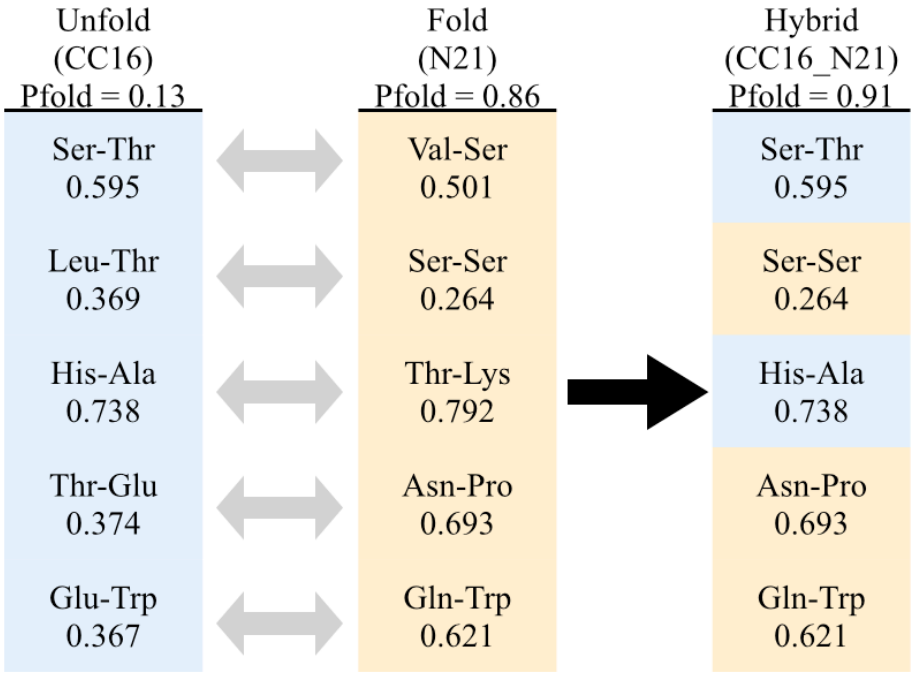


Figure 29. MLCPROB maps for 8-mer fragments for foldable sequences (A), unfoldable sequences (B), and the difference map of unfolded sequence MLCPROB from folded sequence MLCPROB. Figure courtesy of Taisong Zou.

Table 11. Predicted versus observed folded and unfolded WW domain sequences.

	Predicted		
Observed	*unfolded*	*folded*	percent correct
unfolded	41	8	83.7
folded	9	31	77.5

In order to show these crucial contacts are enough to lend tendency of WW domains to fold, we aim to introduce these contacts into an unfolded sequence. To this end, our collaborators tested introducing native contacts *in silico* to unfolded sequences. In this way, hybrid sequences were generated and tested for their Pfold, their probability of folding. Shown in Figure 30 is the method in which we generated a hybrid sequence CC16_N21 that has a high probability of folding. Sequences with the highest probability of folding and showing a WW domain-like fold from the *in silico* simulations were chosen to experimentally test whether these sequences would fold into WW domains and retain the function of WW domains, that is their binding affinity for proline-rich peptides.




```
CC16      GSKLGWTEYHTDAGTEYYYDRQTGESTWEKPEDF
CC16_N21  GSKSGWSEYHNDAGTPYYYDRQTGQSTWEKPEDF
N21       SVESDWSVHTNEKGTPYYHNRVTKQTSWIKPDVL
```

Figure 30. Method for generation of the hybrid sequence CC16_N21. Those combinations of contacts at the crucial points that gave the highest Pfold were chosen for insertion into the CC16 sequence. Below the scheme are the sequences of the unfolded, folded, and hybrid sequence. Pairs of contacts that were changed are highlighted in corresponding colors.

2.2. Materials and Methods

WW domain peptides and group I peptides were synthesized using standard solid phase peptide synthesis (SPPS) protocols.[97] In short, peptides were synthesized using Rink Amide-ChemMatrix® (Matrix Innovations) on a CEM Liberty automated peptide synthesizer with 9-fluorenylmethoxycarbonyl(FMOC)-protected amino acids. Dimethylformamide (DMF) was used as the solvent, 20% piperdine with 1% HBTU (O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate) was used for deprotection, 0.40 M HOBt (N-hydroxy benzotriazole) in DMF was used for activation, and 35% N-Methyl-2-pyrrolidone (NMP)/65% N,N-Diisopropylethylamine, or Hünig's base (DIPEA) used for coupling. Peptides were deprotected and cleaved in one step with 95% trifluoroacetic acid (TFA), 2.5% water, 2.5% triisopropylsilane (TIS). After reducing volume by evaporation under nitrogen, crude peptide was precipitated with cold ether and purified by reverse phase HPLC on an Agilent C18 column. MALDI-TOF analysis was used to verify the molecular weight of the peptides.

^{15}N-labeled peptides were expressed as fusion proteins with maltose binding protein using the pMAL-c5x vector (New England Biolabs). In short, the WW domain peptide sequence preceded by DNA encoding a six histidine tag and a TEV cleavage site was ordered from GenScript with BamHI and NdeI restriction sites for insertion into the pMAL-c5x vector. After double digestion with BamHI and NdeI and purification of the insert and vector with a 1% agarose gel, the insert was ligated into the vector with a ration of 3:1 insert:vector and transformed into NEB 5-alpha competent cells. The transformed cells were cultured and plasmids isolated and sequenced at the DNA

Laboratory at ASU using the forward primer 5'-d(GGTCGTCAGACTGTCGATGAAGCC)-3' and reverse primer 5'-d(TGTCCTACTCAGGAGAGCGTTCAC)-3'. Correct plasmids were transformed into BL21 cells (New England Biolabs) for expression. Fusion proteins were expressed by inoculating 10 mL Luria-Bertani (LB) medium with incubation at 37˚C and shaking at 250 rpm. The cells were grown overnight and then added to 1 L of fresh LB medium. The culture was incubated at 37˚C and 250 rpm shaking until an OD600 of approximately 0.8. Protein expression was induced with 0.4 mM final concentration of isopropyl β-D-thiogalactoside (IPTG) and 20 mM benzyl alcohol. Incubation and shaking of the cell culture was continued overnight at 37˚C. When the culture reached an OD600 of 1.3-1.5, cells were harvested by centrifugation at 6,000 rpm for 10 minutes. The cytoplasmic fraction was isolated by resuspension of the cell pellet in 100 mM Tris, pH=7.5 and sonicating the cells on ice for a total of 20 minutes. This suspension of broken cells was then centrifuged at 15,000 rpm for 20 minutes, saving the supernatant as the cytoplasmic fraction. The fusion proteins were purified by nickel affinity chromatography. Fractions containing the fusion protein were treated with histidine-tagged TEV protease and ran over a nickel column again, this time collecting the fraction that did not bind. MALDI-TOF analysis verified the molecular weight of the WW domain peptide.

Circular dichroism measurements were utilized to determine the fold and stability of the proteins. Full scans were acquired on a JASCO J-815 CD spectrometer (JASCO , Easton, MD) from 260-190 nm at 4˚C in a 1 cm cuvette with protein concentrations at 1 μM (or 1 mM cuvette and 20 μM protein concentration) after overnight equilibration at

4˚C in 20 mM sodium phosphate, pH 7. Thermal denaturation experiments were performed by following molar ellipticity at 227 nm between 4-90˚C with heating at 1˚C/ minute. In the cases where a folded and unfolded baseline were clearly established, data were fitted using equation 6:

$$\theta = \frac{\left(m_f x + b_f\right) + \left(m_u x + b_u\right) \times e^{-\left(\frac{\Delta H}{R \times \left(\frac{1}{x} - \frac{1}{T_m}\right)}\right)}}{1 + e^{-\left(\frac{\Delta H}{R \times \left(\frac{1}{x} - \frac{1}{T_m}\right)}\right)}} \tag{6}$$

where m_f and b_f are the slope and intercept of the folded baseline, respectively, and m_u and b_u are the slope and intercept of the unfolded baseline, respectively. In the cases where a folded or unfolded baseline could not be established, T_m were estimated by fitting the differential melting curve using the van't Hoff relationship.[98] In short, the algebraic derivative was taken of the CD signal, plotted against temperature in Kelvin, and fit to equation 7:

$$\frac{\partial \theta}{\partial T} = Af(1 - f)(T^2) \tag{7}$$

where A is scaling factor and f is substituted with equation 8:

$$f = \frac{K}{K + 1} \tag{8}$$

and finally, T_m is determined by substituting equation 9 for K:

$$K = e^{\left(\frac{\Delta H_{vH}}{R}\right)\times\left(\frac{1}{T_m}-\frac{1}{T}\right)} \quad (9)$$

where ΔH_{vH} is the transition enthalpy and R is the gas constant. The data were fit using Prism 6 from GraphPad Software.

For binding studies, samples were sent to Sanford-Burnham Medical Research Insititude (La Jolla, CA). In short, aliquots of group I peptide from a5 mM stock was titrated into 100 μM WW domain peptide, in 20 mM phosphate buffer, pH 7. Data were fit using standard fitting procedures with a one binding site model.

2.3. Results

Through the biophysical modeling, five local contacts were predicted to be enough predict foldable sequences. Using this finding, foldable sequences were designed from unfolded sequences by introducing native contacts at those positions. The most promising of these designed sequences was CC16_N21, and this sequence was chosen for biophysical characterization.

Thermal denaturation studies were performed on the native WW domain sequence, N21, as a positive control, and of the designed sequence, CC16_N21. After verifying the full circular dichroism spectrum of each peptide had the stereotypical maximum at approximately 227 nm, Figure 31, proteins were melted from 4˚C to 90˚C following the loss of CD signal at 227 nm, Figure 32. Fitting the denaturation of N21 to a two-state model, Equation 6, gives a T_m of 55˚C. Since the folded baseline of N21 has a strong negative slope, the two-state unfolding model may be inappropriate.

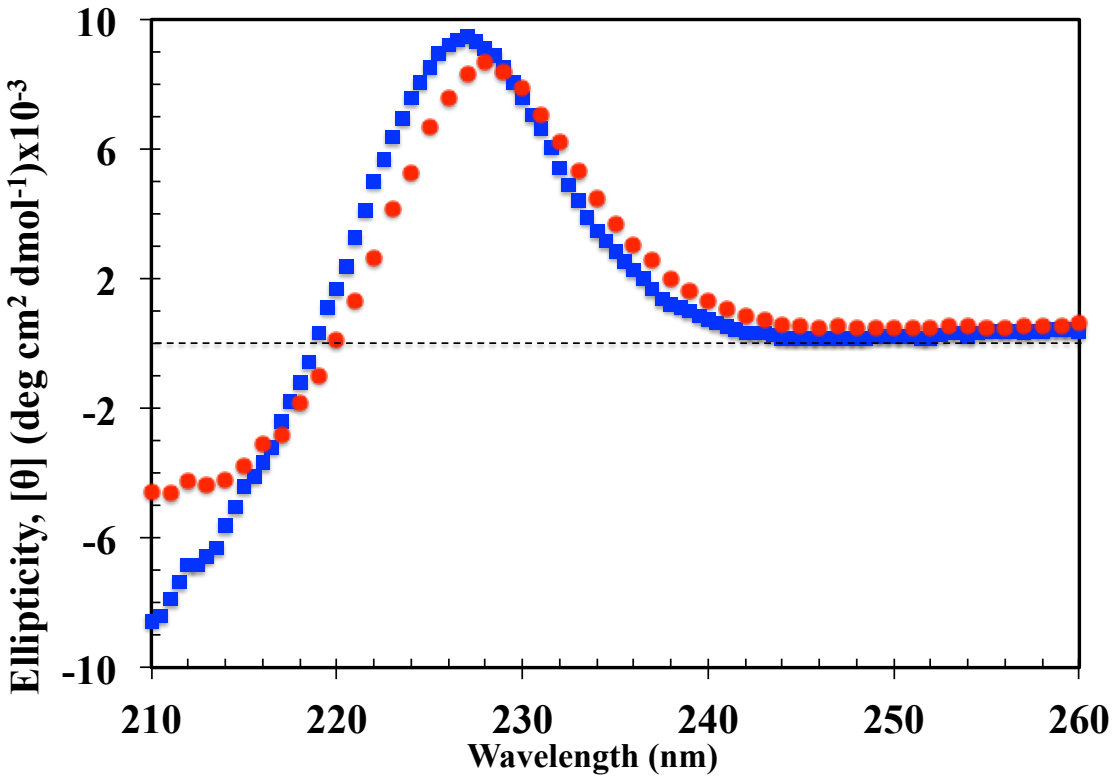


Figure 31. CD spectra of N21 (blue squares) and CC16_N21 (red circles) collected at 4˚C. Both spectra indicate folded WW domains at 4˚C due to the peak at 227 nm.

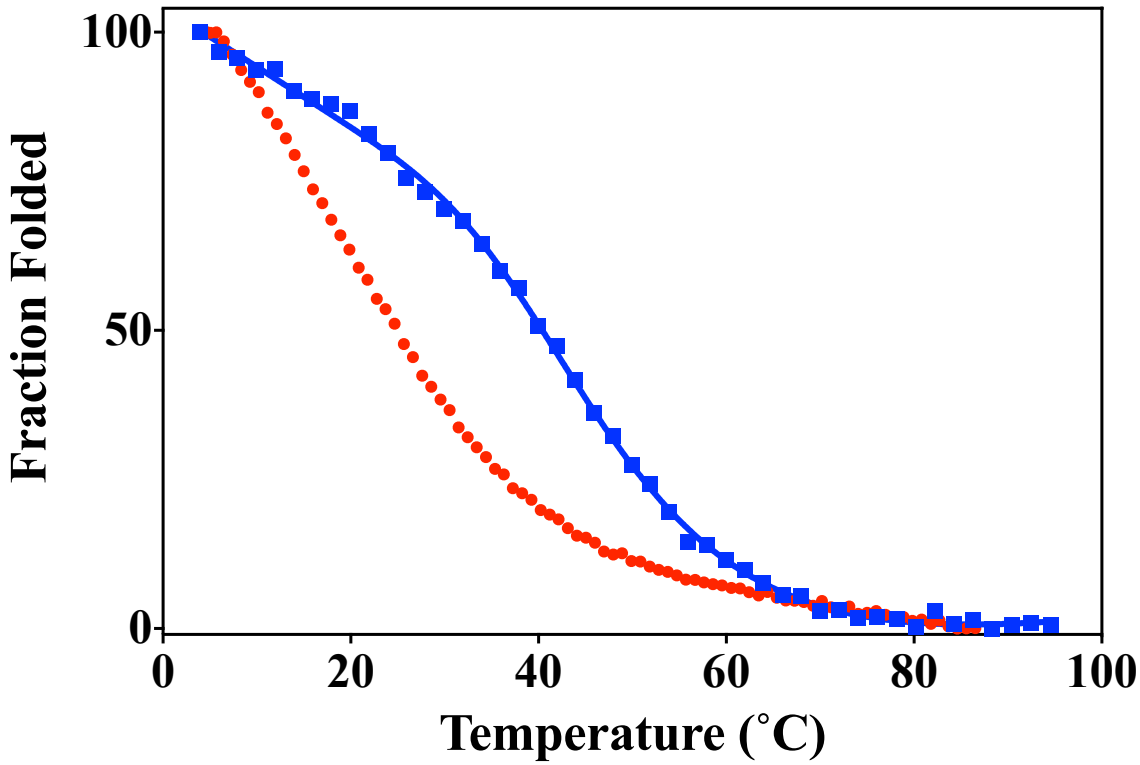


Figure 32. Thermal denaturation curves of N21, blue squares, and CC16_N21, red circles. Due to the lack of a stable, folded baseline, the data was unable to be fit using a two-state unfolding model.

Using the van't Hoff approximations outlined in the materials and methods of this chapter, the differential melting curve (smoothed by averaging 4 neighbor points for clarity sake) of N21 was fitted to Equations 7-9, as shown in Figure 33 to gain a better estimation of the T_m of each peptide.[98] The differential melting curve suggests two separate melting transitions, due to the two minimums in the plot. However, determining the unfolding mechanism of the native peptide is beyond the scope of this study. In order to get a better estimate of the T_m, only the large, final transition was fit to the differential

melting curve/van't Hoff model, Figure 34. This gives an apparent T_m of 42.5˚C for N21. CC16_N21 thermal denaturation data was fit using the differential melting curve/van't Hoff model, as well, since the protein began to melt immediately, as seen in Figure 32. The differential melting curve/van't Hoff fitting for CC16_N21 is shown in Figure 35, and gives the an apparent T_m of 18.6˚C.

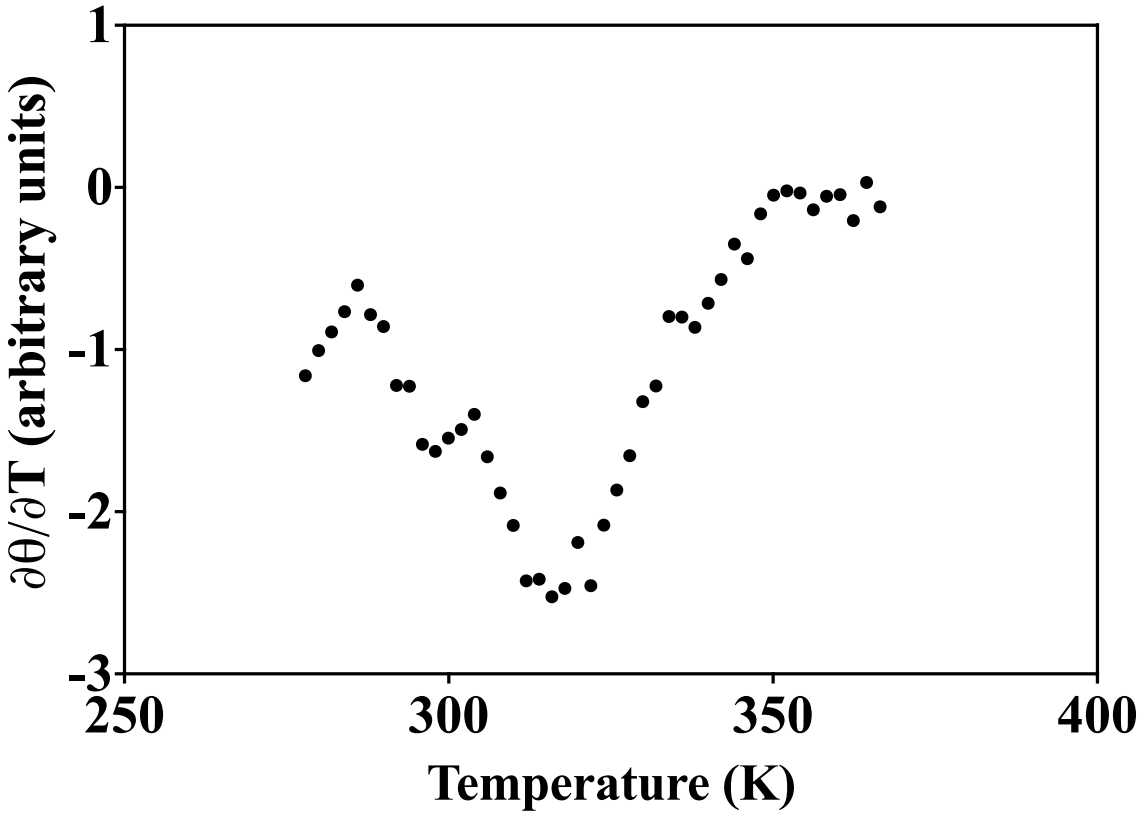


Figure 33. Differential melting curve of N21.There are two minima in the spectra, suggesting that N21 does not follow a two-state model.

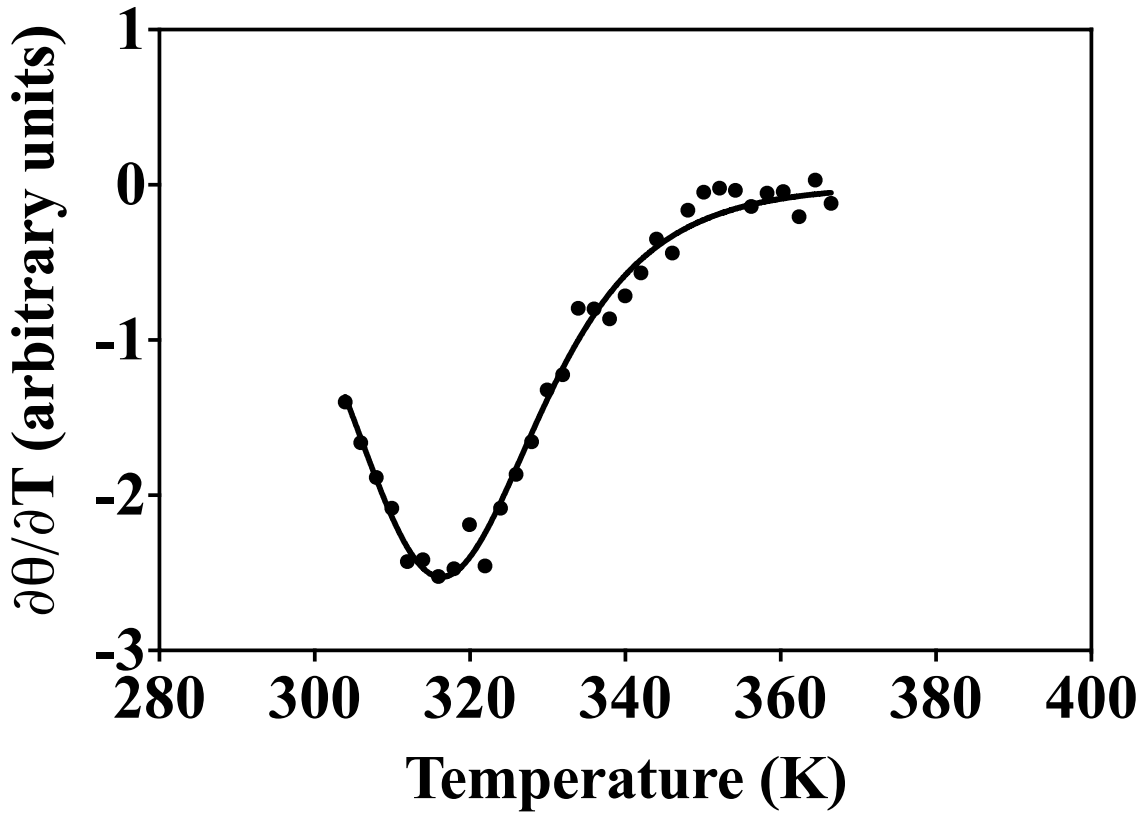


Figure 34. Differential melting curve of N21 showing the large transition with fitted line. The fit gives an apparent T_m of 315.5 K (42.5˚C).

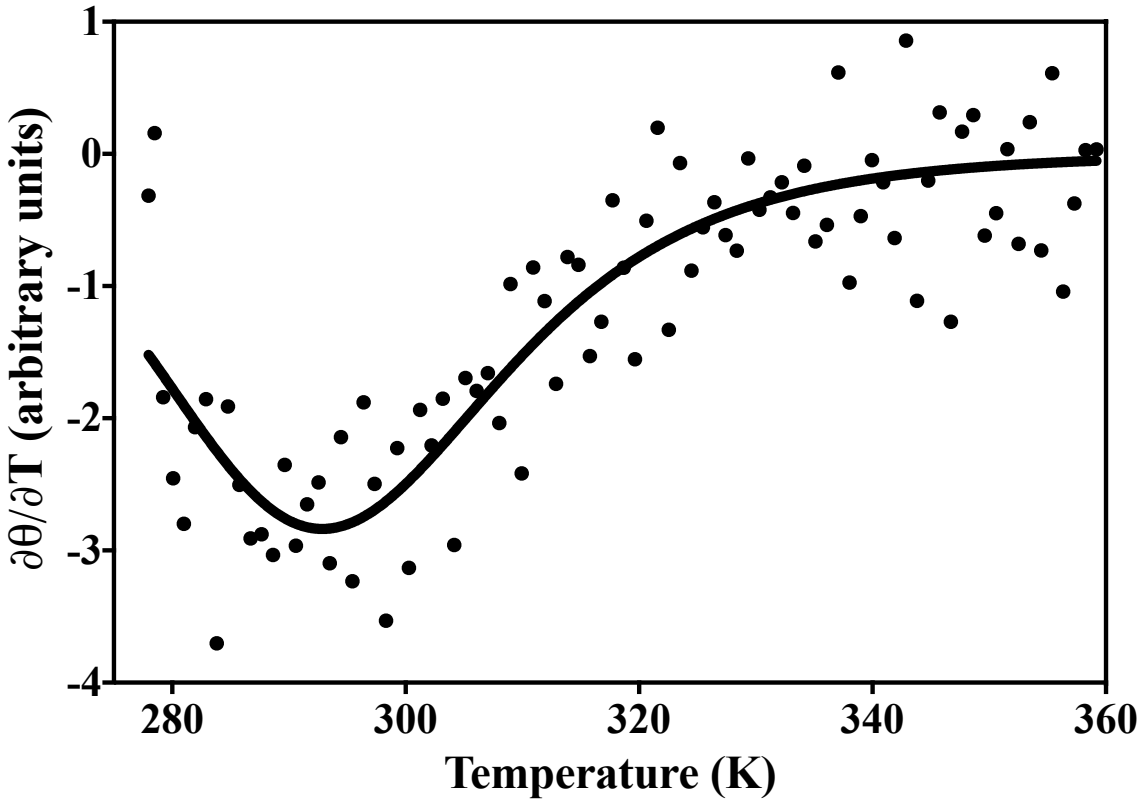


Figure 35. Differential melting curve of CC16_N21 and fit is shown. The apparent T_m for CC16_N21 is 291.6 K (18.6˚C).

The binding affinity for group I peptides was determined by ITC, by titrating in a model group I peptide, EYPPYPPPPYPSG, into CC16_N21 at 4˚C. The titration was completed in duplicate, and the fitting allowed the number of binding sites to be a fitting parameter. The number of sites was given was found to be 1.23 for both titrations, suggesting only one binding site, as expected. The average K_d was found to be 71.0 μM ± 4.7 μM. The fits are shown in Figure 36.

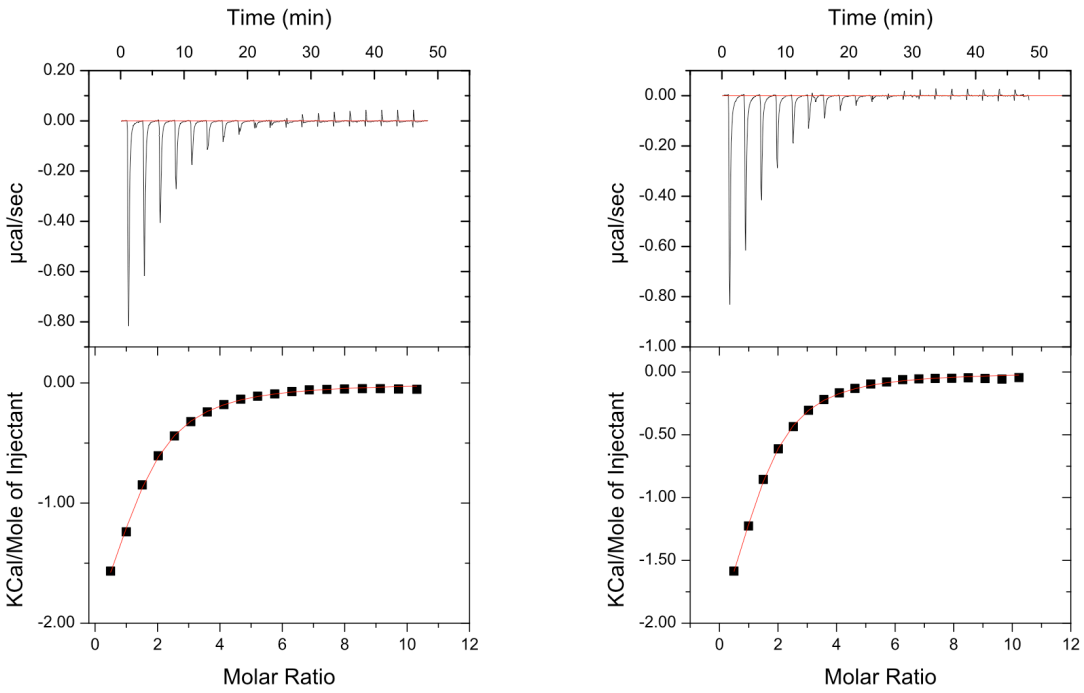


Figure 36. ITC of the group I peptide, EYPPYPPPPYPSG, into CC16_N21. Fitting of the data suggests one binding site and duplicate titrations give 73.5 µM ± 4.4 µM and 68.5 µM ± 5.0 µM, suggesting the designed CC16_N21 retains the function of WW domains.

2.4. Discussion and Conclusion

After characterizing the WW domains as foldable and unfoldable through ZAM, and comparing the contact probability of the folded sequences and unfolded sequences, five crucial contacts were determined.[96] By forcing the contacts in unfolded sequences artificially *in silico* and folding through ZAM, unfolded sequences were found to fold artificially. Utilizing this idea, a foldable sequence was designed from an unfolded sequence by introducing contacts from a native sequence into the unfolded sequence. The most promising sequences were synthesized by solid phase peptide synthesis. The most

promising sequence, CC16_N21, showed a folded CD spectra at 4˚C and matched closely to that of the native sequence (Figure 31). Compared to unfolded sequences (data not shown), the design of CC16_N21 was able to rescue the fold and some of the stability of that of the native sequence, N21 (Figure 31 and Figure 32).

Due to the lack of a folded baseline in the thermal denaturation, the differential melting curve was obtained by taking the algebraic derivative of the CD signal.[98] The differential melting curve was then fit to equations 7, 8, and 9 to estimate the melting point of the native sequence peptide, N21 in Figure 34, and the designed sequence peptide, CC16_N21 in Figure 35.[92,98] While N21 was found to have a moderately stable T_m of 42.5˚C, the lack of a true folded baseline suggests this native sequence is not truly that stable. Also, in the differential melting curve of N21 (Figure 33), there are two minimum suggesting not a true two state melting pathway. While determining the melting pathway of N21 is outside the scope of this study, the ambiguity of the melting point gives credence to our designed peptide, CC16_N21, rescuing the fold and a majority of the stability of the native sequence.

In order to further show that CC16_N21 was a proper WW domain peptide, functional studies were performed. To this end, ITC titrations with a group I peptide, EYPPYPPPPYPSG, were performed. With the low stability of CC16_N21, an apparent T_m of 18.6˚C, all experiments were performed at 4˚C. Fitting the titration showed good agreement with one binding site, and gave an average K_d of 71 μM. This K_d is on the same order of magnitude as other native WW domains and designed WW domains.[93] Given the functional aspect of CC16_N21, and similar fold and stability to native

sequences, we have shown here the design of a foldable WW domain sequence from an unfoldable sequence by introducing native or stabilizing interactions at five crucial contacts.

REFERENCES

(1) Centers for Disease Control and Prevention (CDC). (2008) HIV prevalence estimates--United States, 2006. *MMWR Morb. Mortal. Wkly. Rep. 57*, 1073–1076.

(2) UNAIDS. (2009) AIDS epidemic update. *WHO Library Cataloguing-in-Publication Data* 1–100.

(3) Hall, H., Song, R., Rhodes, P., and Prejean, J. (2008) Estimation of HIV Incidence in the United States. *JAMA: the journal of ….*

(4) Reeves, J. D., and Piefer, A. J. (2005) Emerging drug targets for antiretroviral therapy. *Drugs 65*, 1747–1766.

(5) Cihlar, T., and Ray, A. S. (2010) Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Research 85*, 39–58.

(6) Wensing, A. M. J., van Maarseveen, N. M., and Nijhuis, M. (2010) Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Research 85*, 59–74.

(7) Howett, M. K., and Kuhl, J. P. (2005) Microbicides for prevention of transmission of sexually transmitted diseases. *Curr. Pharm. Des. 11*, 3731–3746.

(8) Scanlan, C. N., Offer, J., Zitzmann, N., and Dwek, R. A. (2007) Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature 446*, 1038–1045.

(9) Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., and Laurencot, C. M. (1997) Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob. Agents Chemother. 41*, 1521.

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

(11) Gustafson, K., Sowder, R., Henderson, L., Cardellina, J., McMahon, J., Rajamani, U., Pannell, L., and Boyd, M. (1997) Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-N, an anti-HIV (human immunodeficiency virus) protein from the cyanobacterium Nostoc ellipsosporum. *Biochemical and Biophysical Research Communications 238*, 223–228.

(12) Mori, T., Shoemaker, R., Gulakowski, R., Krepps, B., McMahon, J., Gustafson, K., Pannell, L., and Boyd, M. (1997) Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. *Biochemical and Biophysical Research Communications 238*, 218–222.

(13) Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M., and Gronenborn, A. M. (1998) Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat Struct Mol Biol 5*, 571–578.

(14) Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., Clore, G. M., and Wlodawer, A. (1999) Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *Journal of Molecular Biology 288*, 403–412.

(15) Barrientos, L. G., and Gronenborn, A. M. (2002) The domain-swapped dimer of cyanovirin-N contains two sets of oligosaccharide binding sites in solution. *Biochemical and Biophysical Research Communications 298*, 598–602.

(16) Shenoy, S. R., Barrientos, L. G., Ratner, D. M., O'Keefe, B. R., Seeberger, P. H., Gronenborn, A. M., and Boyd, M. R. (2002) Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. *Chemistry & Biology 9*, 1109–1118.

(17) Chang, L. C., and Bewley, C. A. (2002) Potent inhibition of HIV-1 fusion by cyanovirin-N requires only a single high affinity carbohydrate binding site: characterization of low affinity carbohydrate binding site knockout mutants. *Journal of Molecular Biology 318*, 1–8.

(18) Fromme, R., Katiliene, Z., Giomarelli, B., Bogani, F., Mc Mahon, J., Mori, T., Fromme, P., and Ghirlanda, G. (2007) A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry 46*, 9199–9207.

(19) Barrientos, L. G., Lasala, F., Delgado, R., Sanchez, A., and Gronenborn, A. M. (2004) Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity. *Structure 12*, 1799–1807.

(20) Bewley, C. A., and Otero-Quintero, S. (2001) The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man8 D1D3 and Man9 with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *J. Am. Chem. Soc. 123*, 3892–3902.

(21) Barrientos, L. G., Matei, E., Lasala, F., Delgado, R., and Gronenborn, A. M. (2006) Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Eng. Des. Sel. 19*, 525–535.

(22) Kelley, B. S., Chang, L. C., and Bewley, C. A. (2002) Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. *J. Am. Chem. Soc. 124*, 3210–3211.

(23) Fromme, R., Katiliene, Z., Fromme, P., and Ghirlanda, G. (2008) Conformational gating of dimannose binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 Å resolution. *Protein Sci. 17*, 939–944.

(24) Botos, I., and Wlodawer, A. (2003) Cyanovirin-N: a sugar-binding antiviral protein with a new twist. *Cell. Mol. Life Sci. 60*, 277–287.

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

(26) Ziółkowska, N. E., and Wlodawer, A. (2006) Structural studies of algal lectins with anti-HIV activity. *Acta Biochim. Pol. 53*, 617–626.

(27) Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R., and Wlodawer, A. (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J. Biol. Chem. 277*, 34336–34342.

(28) Liu, Y., Carroll, J. R., Holt, L. A., McMahon, J., Giomarelli, B., and Ghirlanda, G. (2009) Multivalent Interactions with gp120 Are Required for the Anti-HIV Activity of Cyanovirin. *Biopolymers 92*, 194–200.

(29) Matei, E., Zheng, A., Furey, W., Rose, J., Aiken, C., and Gronenborn, A. M. (2010) Anti-HIV activity of defective cyanovirin-N mutants is restored by dimerization. *Journal of Biological Chemistry 285*, 13057–13065.

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

(31) Koharudin, L. M. I., Liu, L., and Gronenborn, A. M. (2013) Different 3D domain-swapped oligomeric cyanovirin-N structures suggest trapped folding intermediates. *Proceedings of the National Academy of Sciences*.

(32) Keeffe, J. R., Gnanapragasam, P. N. P., Gillespie, S. K., Yong, J., Bjorkman, P. J., and Mayo, S. L. (2011) Designed oligomers of cyanovirin-N show enhanced HIV neutralization. *Proceedings of the National Academy of Sciences 108*, 14079–14084.

(33) Mori, T., and Boyd, M. R. (2001) Cyanovirin-N, a potent human immunodeficiency virus-inactivating protein, blocks both CD4-dependent and CD4-independent binding of soluble gp120 (sgp120) to target cells, inhibits sCD4-induced binding of sgp120 to cell-associated CXCR4, and dissociates bound sgp120 from target cells. *Antimicrob. Agents Chemother. 45*, 664–672.

(34) Barrientos, L. G., O'Keefe, B. R., Bray, M., Sanchez, A., Gronenborn, A. M., and Boyd, M. R. (2003) Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Research 58*, 47–56.

(35) Bolmstedt, A. J., O'Keefe, B. R., Shenoy, S. R., McMahon, J. B., and Boyd, M. R. (2001) Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. *Mol. Pharmacol. 59*, 949–954.

(36) Balzarini, J. (2007) Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nat Rev Micro 5*, 583–597.

(37) Patsalo, V., Raleigh, D. P., and Green, D. F. (2011) Rational and computational design of stabilized variants of cyanovirin-N that retain affinity and specificity for glycan ligands. *Biochemistry 50*, 10698–10712.

(38) Koharudin, L. M. I., Viscomi, A. R., Jee, J.-G., Ottonello, S., and Gronenborn, A. M. (2008) The Evolutionarily Conserved Family of Cyanovirin-N Homologs: Structures and Carbohydrate Specificity. *Structure 16*, 570–584.

(39) Matei, E., Furey, W., and Gronenborn, A. M. (2008) Solution and crystal structures of a sugar binding site mutant of cyanovirin-N: no evidence of domain swapping. *Structure 16*, 1183–1194.

(40) Fujimoto, Y. K., TerBush, R. N., Patsalo, V., and Green, D. F. (2008) Computational models explain the oligosaccharide specificity of cyanovirin-N. *Protein Sci. 17*, 2008–2014.

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

(42) Margulis, C. J. (2005) Computational study of the dynamics of mannose disaccharides free in solution and bound to the potent anti-HIV virucidal protein cyanovirin. *J. Phys. Chem. B 109*, 3639–3647.

(43) Vorontsov, I. I., and Miyashita, O. (2009) Solution and crystal molecular dynamics simulation study of m4-cyanovirin-N mutants complexed with di-mannose. *Biophys. J. 97*, 2532–2540.

(44) Bewley, C. A., Kiyonaka, S., and Hamachi, I. (2002) Site-specific discrimination by cyanovirin-N for alpha-linked trisaccharides comprising the three arms of Man(8) and Man(9). *Journal of Molecular Biology 322*, 881–889.

(45) Sandström, C., Berteau, O., Gemma, E., Oscarson, S., Kenne, L., and Gronenborn, A. M. (2004) Atomic Mapping of the interactions between the antiviral agent cyanovirin-N and oligomannosides by saturation-transfer difference NMR †. *Biochemistry 43*, 13926–13931.

(46) Woodrum, B. W., Maxwell, J. D., Bolia, A., Ozkan, S. B., and Ghirlanda, G. (2013) The antiviral lectin cyanovirin-N: probing multivalency and glycan recognition through experimental and computational approaches. *Biochem. Soc. Trans. 41*, 1170–1176.

(47) Atilgan, C., Gerek, Z. N., Ozkan, S. B., and Atilgan, A. R. (2010) Manipulation of conformational change in proteins by single-residue perturbations. *Biophys. J. 99*, 933–943.

(48) Gerek, Z. N., and Ozkan, S. B. (2010) A flexible docking scheme to explore the binding selectivity of PDZ domains. *Protein Sci. 19*, 914–928.

(49) Gerek, Z. N., and Ozkan, S. B. (2011) Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS Comput. Biol. 7*, e1002154.

(50) Bolia, A., Gerek, Z. N., Keskin, O., Banu Ozkan, S., and Dev, K. K. (2012) The binding affinities of proteins interacting with the PDZ domain of PICK1. *Proteins 80*, 1393–1408.

(51) Xia, J., and Margulis, C. J. (2009) Computational study of the conformational structures of saccharides in solution based on J couplings and the "fast sugar structure prediction software". *Biomacromolecules 10*, 3081–3088.

(52) Xia, J., Margulis, C. J., and Case, D. A. (2011) Searching and optimizing structure ensembles for complex flexible sugars. *J. Am. Chem. Soc. 133*, 15252–15255.

(53) Ramadugu, S. K., Chung, Y.-H., Xia, J., and Margulis, C. J. (2009) When sugars get wet. A comprehensive study of the behavior of water on the surface of oligosaccharides. *J. Phys. Chem. B 113*, 11003–11015.

(54) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank.

(55) Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis 18*, 2714–2723.

(56) Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters 314*, 141–151.

(57) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins 65*, 712–725.

(58) Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers 56*, 275–291.

(59) Atilgan, C., and Atilgan, A. R. (2009) Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol. 5*, e1000544.

(60) Ikeguchi, M., Ueno, J., Sato, M., and Kidera, A. (2005) Protein structural change upon ligand binding: linear response theory. *Phys. Rev. Lett. 94*, 078102.

(61) Nevin Gerek, Z., Kumar, S., and Banu Ozkan, S. (2013) Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol Appl 6*, 423–433.

(62) Bolia, A., Woodrum, B. W., Cereda, A., Ruben, M. A., Wang, X., Ozkan, S. B., and Ghirlanda, G. (2014) A flexible docking scheme efficiently captures the energetics of glycan-cyanovirin binding. *Biophys. J. 106*, 1142–1151.

(63) Bahar, I., Lezon, T. R., Yang, L.-W., and Eyal, E. (2010) Global dynamics of proteins: bridging between structure and function. *Annu Rev Biophys 39*, 23–42.

(64) Kollman, P. A., Dixon, R., Cornell, W., Fox, T., Chipot, C., and Pohorille, A. (1997) The development/application of a “minimalist”organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. *Computer simulation of biomolecular systems 3*, 83–96.

(65) Tsui, V., and Case, D. A. (2000) Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc. 122*, 2489–2498.

(66) Davis, I. W., and Baker, D. (2009) RosettaLigand docking with full ligand and receptor flexibility. *Journal of Molecular Biology 385*, 381–392.

(67) Meiler, J., and Baker, D. (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins 65*, 538–548.

(68) Wang, R., Lai, L., and Wang, S. (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des. 16*, 11–26.

(69) Farmer, B. T., Constantine, K. L., Goldfarb, V., Friedrichs, M. S., Wittekind, M., Yanchunas, J., Robertson, J. G., and Mueller, L. (1996) Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nat Struct Mol Biol 3*, 995–997.

(70) Williamson, M. P. (2013) Progress in Nuclear Magnetic Resonance Spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy 73*, 1–16.

(71) Percudani, R., Montanini, B., and Ottonello, S. (2005) The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins 60*, 670–678.

(72) Bewley, C. A., and Shahzad-ul-Hussan, S. (2013) Characterizing carbohydrate-protein interactions by nuclear magnetic resonance spectroscopy. *Biopolymers 99*, 796–806.

(73) Fujimoto, Y. K., and Green, D. F. (2012) Carbohydrate recognition by the antiviral lectin cyanovirin-N. *J. Am. Chem. Soc. 134*, 19639–19651.

(74) Vorontsov, I. I., and Miyashita, O. (2011) Crystal molecular dynamics simulations to speed up MM/PB(GB)SA evaluation of binding free energies of di-mannose deoxy analogs with P51G-m4-Cyanovirin-N. *J Comput Chem 32*, 1043–1053.

(75) Gray, J. J., Moughon, S. E., Kortemme, T., Schueler-Furman, O., Misura, K. M. S., Morozov, A. V., and Baker, D. (2003) Protein-protein docking predictions for the CAPRI experiment. *Proteins 52*, 118–122.

(76) Matei, E., Louis, J. M., Jee, J., and Gronenborn, A. M. (2011) NMR solution structure of a cyanovirin homolog from wheat head blight fungus. *Proteins 79*, 1538–1549.

(77) Totrov, M., and Abagyan, R. (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology 18*, 178–184.

(78) Andrusier, N., Mashiach, E., Nussinov, R., and Wolfson, H. J. (2008) Principles of flexible protein-protein docking. *Proteins 73*, 271–289.

(79) Tian, X., and Shearer, G. (2002) The mold-specific MS8 gene is required for normal hypha formation in the dimorphic pathogenic fungus Histoplasma capsulatum. *Eukaryotic Cell 1*, 249–256.

(80) Buist, G., Steen, A., Kok, J., and Kuipers, O. P. (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol. Microbiol. 68*, 838–847.

(81) Koharudin, L. M. I., Viscomi, A. R., Montanini, B., Kershaw, M. J., Talbot, N. J., Ottonello, S., and Gronenborn, A. M. (2011) Structure-Function Analysis of a CVNH-LysM Lectin Expressed during Plant Infection by the Rice Blast Fungus Magnaporthe oryzae. *Structure 19*, 662–674.

(82) Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc 1*, 2876–2890.

(83) Gulakowski, R. J., McMahon, J. B., Staley, P. G., Moran, R. A., and Boyd, M. R. (1991) A semiautomated multiparameter approach for anti-HIV drug screening. *J. Virol. Methods 33*, 87–100.

(84) Ilsley, J. L., Sudol, M., and Winder, S. J. (2002) The WW domain: linking cell signalling to the membrane cytoskeleton. *Cell. Signal. 14*, 183–189.

(85) Chen, H. I., and Sudol, M. (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl. Acad. Sci. U.S.A. 92*, 7819–7823.

(86) Rotin, D. (1998) WW (WWP) Domains: From Structure to Function, in *Protein Modules in Signal Transduction*, pp 115–133. Springer Berlin Heidelberg, Berlin, Heidelberg.

(87) Sudol, M. (1996) Structure and function of the WW domain. *Prog. Biophys. Mol. Biol. 65*, 113–132.

(88) Gao, Y.-G., Yang, H., Zhao, J., Jiang, Y.-J., and Hu, H.-Y. (2014) Autoinhibitory structure of the WW domain of HYPB/SETD2 regulates its interaction with the proline-rich region of huntingtin. *Structure 22*, 378–386.

(89) Hu, H., Columbus, J., Zhang, Y., Wu, D., Lian, L., Yang, S., Goodwin, J., Luczak, C., Carter, M., Chen, L., James, M., Davis, R., Sudol, M., Rodwell, J., and Herrero, J. J. (2004) A map of WW domain family interactions. *Proteomics 4*, 643–655.

(90) Macias, M. J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature 382*, 646–649.

(91) Sudol, M., Chen, H. I., Bougeret, C., Einbond, A., and Bork, P. (1995) Characterization of a novel protein-binding module--the WW domain. *FEBS Lett. 369*, 67–71.

(92) Socolich, M., Lockless, S. W., Russ, W. P., Lee, H., Gardner, K. H., and Ranganathan, R. (2005) Evolutionary information for specifying a protein fold. *Nature 437*, 512–518.

(93) Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B., and Ranganathan, R. (2005) Natural-like function in artificial WW domains. *Nature 437*, 579–583.

(94) Ozkan, S. B., Wu, G. A., Chodera, J. D., and Dill, K. A. (2007) Protein folding by zipping and assembly. *Proc. Natl. Acad. Sci. U.S.A. 104*, 11987–11992.

(95) Voelz, V. A., Shell, M. S., and Dill, K. A. (2009) Predicting peptide structures in native proteins from physical simulations of fragments. *PLoS Comput. Biol. 5*, e1000281.

(96) Zou, T. (2014, May 1) Protein Folding & Dynamics Using Multi-Scale Computational Methods (Ozkan, S. B., Ed.). Arizona State Univeristy, Tempe, AZ.

(97) Shinde, S., Cordova, J. M., Woodrum, B. W., and Ghirlanda, G. (2012) Modulation of function in a minimalist heme-binding membrane protein. *J. Biol. Inorg. Chem. 17*, 557–564.

(98) John, D. M., and Weeks, K. M. (2000) van't Hoff enthalpies without baselines. *Protein Sci. 9*, 1416–1419.

APPENDIX A

PERMISSIONS FOR FIGURE 1 AND FIGURE 2.

Home | Account Info | Help

Title: Exploiting the defensive sugars of HIV-1 for drug and vaccine design

Author: Christopher N. Scanlan, John Offer, Nicole Zitzmann and Raymond A. Dwek

Publication: Nature

Publisher: Nature Publishing Group

Date: Apr 25, 2007

Logged in as:
Brian Woodrum
Account #:
3000802559

LOGOUT

Order Completed

Thank you very much for your order.

This is a License Agreement between Brian W Woodrum ("You") and Nature Publishing Group ("Nature Publishing Group"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

Get the printable license.

License Number	3422660549564
License date	Jul 05, 2014
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature
Licensed content title	Exploiting the defensive sugars of HIV-1 for drug and vaccine design
Licensed content author	Christopher N. Scanlan, John Offer, Nicole Zitzmann and Raymond A. Dwek
Licensed content date	Apr 25, 2007
Type of Use	reuse in a dissertation / thesis
Volume number	446
Issue number	7139
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 2
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Glycan-Cyanovirin-N Interactions and Designed WW Domains: Combining Experimental and Computational Studies
Expected completion date	Aug 2014
Estimated size (number of pages)	100
Total	0.00 USD

APPENDIX B

PERMISSION FOR FIGURE 3.

Home | Account Info | Help

Title: Emerging Drug Targets for Antiretroviral Therapy
Author: Dr Jacqueline D. Reeves
Publication: Drugs
Publisher: Springer
Date: Jan 1, 2005

Logged in as:
Brian Woodrum
Account #:
3000802559

LOGOUT

Order Completed

Thank you very much for your order.

This is a License Agreement between Brian W Woodrum ("You") and Springer ("Springer"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

Get the printable license.

License Number	3422651153748
License date	Jul 05, 2014
Licensed content publisher	Springer
Licensed content publication	Drugs
Licensed content title	Emerging Drug Targets for Antiretroviral Therapy
Licensed content author	Dr Jacqueline D. Reeves
Licensed content date	Jan 1, 2005
Volume number	65
Issue number	13
Type of Use	Thesis/Dissertation
Portion	Figures
Author of this Springer article	No
Original figure numbers	figures 1 and 2
Title of your thesis / dissertation	Glycan-Cyanovirin-N Interactions and Designed WW Domains: Combining Experimental and Computational Studies
Expected completion date	Aug 2014
Estimated size(pages)	100
Total	0.00 USD

CLOSE WINDOW

APPENDIX C

PERMISSION TO REPRODUCE BOLIA, A., WOODRUM, B. W., CEREDA, A., RUBEN, M. A., WANG, X., OZKAN, S. B., AND GHIRLANDA, G. (2014) A FLEXIBLE DOCKING SCHEME EFFICIENTLY CAPTURES THE ENERGETICS OF GLYCAN-CYANOVIRIN BINDING. BIOPHYS. J. 106, 1142–1151.

Copyright Clearance Center RightsLink®

My Orders My Library My Profile

Welcome brianwoodrum@gmail.com Log out | Help

- Edit Your Customer Information
- Change Your Password
- Change Your Preferred Language
- Edit Your Billing Address
- Edit Your Shipping Address
- Manage Credit Card Information
- Link Your CCC Account

My Orders > Orders > All Orders

Review Order

Thank you very much for your order.

This is a License Agreement between Brian W \ ·vier"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

Get the printable license.

License Number	3411570689764
License date	Jun 17, 2014
Order Content Publisher	Elsevier
Order Content Publication	Biophysical Journal
Order Content Title	A Flexible Docking Scheme Efficiently Captures the Energetics of Glycan-Cyanovirin Binding
Order Content Author	Ashini Bolia,Brian W. Woodrum,Angelo Cereda,Melissa A. Ruben,Xu Wang,S. Banu Ozkan,Giovanna Ghirlanda
Order Content Date	4 March 2014
Licensed content volume number	106
Licensed content issue number	5
Number of pages	10
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Glycan-Cyanovirin-N Interactions and Designed WW Domains: Combining Experimental and Computational Studies
Expected completion date	Aug 2014
Estimated size (number of pages)	100
Elsevier VAT number	GB 494 6272 12
Price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	**0.00 USD**

Back