Published in final edited form as:
Biochemistry. 2011 December 13; 50(49): 10698–10712. doi:10.1021/bi201411c.

Rational and Computational Design of Stabilized Variants of Cyanovirin-N which Retain Affinity and Specificity for Glycan Ligands

Vadim Patsalo[a,b], **Daniel P. Raleigh**[c,d], and **David F. Green**[a,b,c,d,*]

[a]Department of Applied Mathematics and Statistics Stony Brook University Stony Brook, New York 11794 USA

[b]Laufer Center for Physical and Quantitative Biology Stony Brook University Stony Brook, New York 11794 USA

[c]Department of Chemistry Stony Brook University Stony Brook, New York 11794 USA

[d]Graduate Program in Biochemistry and Structural Biology Stony Brook University Stony Brook, New York 11794 USA

Abstract

Cyanovirin-N (CVN) is an 11-kDa pseudo-symmetric cyanobacterial lectin that has been shown to inhibit infection by the Human Immunodeficiency Virus (HIV) by binding to high-mannose oligosaccharides on the surface of the viral envelope glycoprotein gp120. In this work we describe rationally-designed CVN variants that stabilize the protein fold while maintaining high affinity and selectivity for their glycan targets. Poisson–Boltzmann calculations and protein repacking algorithms were used to select stabilizing mutations in the protein core. By substituting the buried polar side chains of Ser11, Ser20, and Thr61 with aliphatic groups, we stabilized CVN by nearly 12 °C against thermal denaturation, and by 1 M of GuaHCl against chemical denaturation, relative to a previously-characterized stabilized mutant. Glycan microarray binding experiments confirmed that the specificity profile of carbohydrate binding is unperturbed by the mutations, and is identical for all variants. In particular, the variants selectively bound glycans containing the Manα(1→2)Man linkage, which is the known minimal binding unit of CVN. We also report the slow denaturation kinetics of CVN and show that they can complicate thermodynamic analysis; in particular, the unfolding of CVN cannot be described as a fixed two-state transition. Accurate thermodynamic parameters are needed to describe the complicated free energy landscape of CVN, and we provide updated values for CVN unfolding.

Keywords

protein design; continuum electrostatics; lectin; HIV; microbicide

The 11-kDa lectin Cyanovirin-N (CVN), originally isolated from the freshwater cyanobacterium *Nostoc ellipsosporum*, has generated interest as a potential anti-viral agent.

*Contributing author: Dr. David F. Green Stony Brook University, Math Tower P-137 Stony Brook, New York 11794-3600 phone: (631) 632-9344 fax: (631) 632-8490 david.green@stonybrook.edu.

Supporting Information Available Full amino acid sequences of the proteins in this study, analytical gel filtration results, HSQC NMR spectra, singular value decomposition results, fractional solvent accessibility of CVN sidechains, temperature-dependent CD spectra, and identities of microarray glycans bound by CVN variants. This material is available free of charge via the Internet at http://pubs.acs.org.

Therapeutic interest in CVN stems from its ability to potently and irreversibly inactivate both laboratory-adapted and naturally-occurring strains of the Human Immunodeficiency Virus (HIV) (1). This irreversible inhibition involves tight binding of CVN to the high-mannose oligosaccharides on the viral envelope glycoprotein gp120 (2). The binding event affects the flexibility of gp120, and is thought to hinder the conformational changes which are required for proper interactions with the cell-membrane receptor CD4 and co-receptors CCR5/CXCR4, which are essential for subsequent gp41-mediated membrane fusion (3, 4).

The presence of glycans on gp120 is crucial to the ability of HIV to evade detection by the immune system. The dense glycan coating of gp120 gives rise to what has been termed a "glycan shield" due to its masking of the underlying immunogenic protein epitopes (5). In fact, most of the surface of the proteins exposed on the extraviral side of the envelope is covered by carbohydrates. CVN, along with a number of other lectins, represents an example of a novel class of therapeutic carbohydrate-binding agents against enveloped viruses. The antiviral activity of such molecules is dual in mechanism; first, they are able to bind to the glycans of the viral envelope and block virus entry, and second, long-term exposure to such agents leads to a progressive deletion of the glycosylation sites on the viral surface as an adaptive response to antiviral pressure. In the case of HIV, the deletion of the "glycan shield" is thought to reveal previously-obscured epitopes and allow enhanced detection and neutralization of the virus by the immune system (3, 6). CVN has also been shown to inhibit transmission by several other enveloped viruses; in addition to HIV, CVN inhibits Ebola and herpes simplex viral infection by binding to their respective envelope glycoproteins (7, 8), and influenza infection by binding to hemagglutinin (9).

Wild-type CVN possesses modest stability and has a tendency to form domain-swapped dimers (10). These factors complicate its clinical use and can make biophysical and biochemical studies of the protein difficult. The most promising clinical application of CVN is as a topical microbicide, with hope of reducing HIV transmission in sub-Saharan Africa. It must thus be stable under a wide range of conditions and possess a long shelf life under harsh conditions. The interest in developing CVN as a microbicide has led to investigations of large-scale production of the protein using bacterial, yeast, and plant expression systems (11–13). Stabilization of CVN has implications for recombinant production of the protein as a therapeutic, and can aid in recombinant protein purification. In addition, availability of stabilized CVN variants facilitates studies of mutations which significantly disturb the stability of the protein fold, such as previously-characterized binding-site knockouts (14–18) or designed protein oligomers (19).

We set out to design variants of CVN which are more stable to chemical and thermal unfolding, but retain the full biological activity of the wild-type protein. Here, we present CVN variants which are more stable than the wild-type protein, yet retain the native fold and carbohydrate specificity. The designed homologues represent a new background which is amenable to binding-site redesign or engineering, due to their increased tolerance of higher temperatures and denaturant concentrations compared to wild-type CVN.

Protein stabilization has previously been achieved using a variety of approaches: improving core packing (20), removal of buried polar side chains or unsatisfied salt bridges (21, 22), homology-based strategies (23), mutation of charged surface residues (24), introduction of new disulfide bonds (25), turn redesign (26–28), modulation of unfolded state entropy (29), and rational considerations of unfolded state interactions (30). Often, the substitutions increase stability via a mixture of effects, such as optimized core packing and increased burial of hydrophobic surface area, which may be difficult to deconvolve (31). In this work, we employed a rational design strategy which removes buried polar groups and improves the packing within the protein core to yield increased stability.

We found that kinetic denaturation by guanidine hydrochloride (GuaHCl) is very slow, and that CVN folding is not two-state. Previous studies of CVN folding thermodynamics were misled by the slow unfolding, and we provide updated thermodynamic parameters for CVN. Such data is essential for a description of CVN's complicated free energy landscape.

Materials & Methods

Continuum Electrostatic Calculations

Electrostatic contributions of CVN side chains to protein stability were obtained using standard methods (32) by solving the linearized Poisson–Boltzmann equation (33, 34) using a multigrid finite-difference solver (M.D. Altman and B. Tidor, unpublished) distributed with the Integrated Continuum Electrostatics (ICE) software package (DFG, E. Kangas, Z.S. Hendsch, and B. Tidor, available for licensing through the MIT Technology Licensing Office). Dielectric constants of 2 and 80 were used for the solute and solvent, respectively. The dielectric boundary was defined by the molecular surface using a 1.4 Å radius probe, with radii optimized for this purpose (35). The ionic strength was set to 145 m_M, with a 2.0 Å ion-exclusion layer. A 129-unit grid was used with overfocusing boundary conditions (the longest dimension of the molecule occupying 23%, then 92%, and finally 184% of one edge of the grid).

The electrostatic contribution of a side chain at position *i* to the unfolded state was modeled by its interactions with the carbonyl($i - 1$)-(*i*)-amino($i + 1$) “tripeptide” in the absence of any other protein groups. The conformation of the “tripeptide” was unchanged from that of the folded protein. Electrostatic calculations were performed on 201 snapshots extracted from a 200 ns explicit-solvent molecular dynamics simulation of CVN (Y.K. Fujimoto and DFG, in preparation).

Protein Design Calculations

We used a hierarchical design procedure (36) based on the Dead-End Elimination and A* algorithms (37–42) to find low-energy sequences compatible with the CVN fold. The protein backbone was kept fixed, and the “penultimate” rotamer library of Richardson and colleagues (43) was used for side-chain rotamers. Energies were calculated by the CHARMM potential (44) with the PARAM22 force field (45).

All CHARMM energy terms (bond, angle, dihedral, improper, Lennard-Jones, and electrostatic) were used in the search. Changes to protein stability upon mutation or side-chain rearrangement were approximated from the energetic difference between the folded and the unfolded states based on isolated model side chains. The isolated side chains were acetylated at the N-terminus and N-methylamidated at the C-terminus. Rotamers which clashed with the backbone or with neighboring side chains were eliminated. Sequences greater than 15 kcal/mol above the global minimum energy configuration were discarded. Ten top-scoring configurations for the remaining sequences were re-ranked using Analytical Continuum Electrostatics (46), as implemented in CHARMM. Software written in collaboration with Tidor and colleagues (36, 57) (available for licensing through the MIT Technology Licensing Office) was used for the search.

Cloning, Protein Expression & Purification

Mutants were constructed starting with the synthetic gene coding for Cyanovirin-N (DNA 2.0, Menlo Park, CA) cloned into the pET-26b(+) vector (Novagen) using the NdeI and XhoI restriction sites. Round-the-horn site-directed mutagenesis (S. Moore, unpublished) was performed using the Phusion High-Fidelity PCR Kit. Briefly, non-overlapping primers, with one encoding the desired mutation, were phosphorylated at the 5' end using T4

Polynucleotide Kinase. The phosphorylated primer mix was then added to the PCR reaction, and extension followed for 30–35 cycles. The purified PCR product was then ligated at 16 °C overnight with T4 DNA Ligase, and transformed into XL1-Blue Competent Cells (Novagen). A typical transformation yielded 50–200 colonies. Restriction enzymes, Phusion polymerase, polynucleotide kinase, and ligase were purchased from New England Biolabs. The identity of the mutants was confirmed by DNA sequencing.

Plasmids bearing the appropriate mutations were transformed into BL21(DE3) *E. coli* strain (Novagen). Cells were grown at 37 °C until OD_{600} reached 0.8, and protein expression was induced by addition of 1 mM isopropyl-D-thiogalactoside for 4–5 h. The cells were pelleted by centrifugation at 7500*g*, and frozen at −80 °C until purification.

With the exception of ΔM, the proteins were purified under denaturing conditions. The cell pellet was resuspended in Buffer A (6 M GuaHCl, 20 mM imidazole, 20 mM Tris-HCl, pH 8.0) using 10 mL per gram of cell paste. The cells were disrupted by four passes through a French Press high-pressure homogenizer. The insoluble fraction was immediately pelleted by ultracentrifugation at 100,000 *g*. The supernatant was loaded onto a 5ml His-Trap FF column (GE Healthcare, Piscataway, NJ), connected to an AKTA Explorer 10 FPLC (GE Healthcare), and eluted over 10 column volumes using a gradient of Buffer B (6 M GuaHCl, 300 mM imidazole, 20 mM Tris-HCl, pH 8.0). Dithiothreatol was added to a final concentration of 5 mM.

The proteins were refolded by overnight dialysis against Buffer C (10 mM Tris-HCl, pH 8.0) at room temperature, changing the buffer once. The precipitate was removed by ultracentrifugation, and the soluble fraction was incubated at 37 °C for 24–48 h to increase interconversion of domain-swapped dimer to monomeric protein. The proteins were concentrated by centrifugation, and loaded onto a Superdex 75 26/60 gel filtration column (GE Healthcare) pre-equilibrated with Buffer D (20 mM sodium phosphate, pH 6.0). The monomeric proteins were stored at 4 °C until characterization.

The identity and purity of the recombinant proteins was confirmed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry; all of the overexpressed proteins were found to have an N-terminal methionine residue. Protein concentrations were determined by measuring A_{280} in 6 M GuaHCl using a calculated extinction coefficient of 10 220 M^{-1} cm^{-1}.

Oligomerization State Determination

Analytical gel filtration was performed on a Superdex 75 (10/300 GL) column (GE Healthcare). 100 μL samples were injected onto the column preequilibrated in 20 mM sodium phosphate, 200 mM sodium chloride (pH 7.5) and were eluted in the same buffer at 0.5 mL min^{-1}.

Equilibrium Denaturation Studies

Solutions for equilibrium denaturation studies were prepared as follows. Typically, thirty-two 2 mL samples containing 10 μm protein in buffer D with the appropriate concentration of GuaHCl were incubated at room temperature for 72 h. The concentration of GuaHCl in each sample was determined by refractometry. Samples were analyzed by fluorescence and circular dichroism spectroscopies, described below.

Spectroscopic Analysis

Intrinsic tryptophan fluorescence spectra were collected at 25 °C on a PTI spectrofluorometer (Birmingham, NJ) equipped with a Peltier temperature controller.

Protein concentrations were 10 μM in Buffer D. The excitation wavelength was set at 290 nm and fluorescence emission was collected from 300 nm to 400 nm. To facilitate comparison with earlier work, the intrinsic tryptophan fluorescence ratio I_{330}/I_{360} was also recorded for each sample for 60 s, and averaged to yield the final value.

Circular dichroism measurements were carried out on a Chiroscan spectrometer (Applied Photo-physics, Surrey, UK) equipped with a Peltier temperature controller. Far-UV spectra were recorded from 90 nm to 260 nm in 0.5 nm steps using a path length of 1 mm. Spectral measurements were performed with a bandwidth of 2 nm, with a data acquisition time of 1 s. For thermal denaturation studies, protein concentrations were 15 μM in Buffer D. The sample was heated at 2 °C min^{-1} in steps of 1 °C from 5 °C to 90 °C. Thermal unfolding was measured by following the transition at 200 nm, with a data acquisition time of 5 s at each temperature. θ_{200} was chosen for consistency with previous work and because it shows near-maximal signal difference upon heat denaturation (Supporting Figure S??.)

Thermal denaturation curves were fit to a two-state model as previously described (47, 48) using a custom nonlinear least-squares regression routine, implemented in the statistical environment R (49). Both the pre- and post-transition regions were allowed a linear baseline.

Equilibrium denaturation far-UV spectra in the presence of GuaHCl were collected from 215–245 nm in 0.5 nm increments using 10 μM protein and a path length of 5 mm. The data analysis is discussed below.

NMR spectra were collected at 293 K on a Varian Inova 600 MHz spectrometer. All samples contained 150 μM protein in 20 mM sodium phosphate pH 6.0, 10% D_2O. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was included as an internal reference. Spectra were processed with NMRPipe (50), and analyzed with Sparky (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco)

Glycan Array Screening

Fluorescein-labeled samples of CVN variants were prepared by incubating 0.5 mg purified protein in 50 mM borate, pH 8.5 with a 15-fold molar excess of NHS-Fluorescein (0.3 mg) at room temperature for 1 h. Unreacted dye was removed by desalting on a HiTrap 5 mL column (GE Healthcare), and samples were further dialyzed against PBS overnight in the dark at 4 °C. Protein concentrations were determined by measuring A_{280} and A_{494}, correcting for fluorophore absorbance.

Samples were submitted to the Consortium for Functional Glycomics, where 70 μL of 200 μg mL^{-1} solution were spotted onto Mammalian Printed Array Version 4.1 containing 465 carbohydrates.

Data Analysis

GuaHCl-induced denaturation data were fit as follows. The decomposition of spectroscopic data into basis states and their fractional populations can be represented as a matrix multiplication. CD spectroscopy obeys Beer's law (*i.e.* the observed data is a linear combination of the basis spectra weighted by their fractional populations). If one records a spectrum over *m* wavelengths over *n* experimental conditions (denaturant concentrations), the data populate an $m \times n$ matrix. Decomposition of this data into *k* thermodynamic states (in our case, $k = 3$) is equivalent to the following matrix product:

$$A_{m\times n}=S_{m\times k}F^{T}_{n\times k} \quad (1)$$

The columns of matrix *S* contain the spectra of the thermodynamic states, while the columns of F^T are the fractional populations of the thermodynamic states at each experimental condition. The data are obscured by an $E_{n\times m}$ matrix of experimental noise. Assuming the linear extrapolation (LEM) model, in which the free energy of denaturation is linear with denaturant concentration, the relative free energies of the thermodynamic states at any denaturant concentration are given by four thermodynamic parameters ($\Delta G^{\circ}_{N\rightarrow I}$, $m_{N\rightarrow I}$, $\Delta G^{\circ}_{N\rightarrow D}$, $m_{N\rightarrow D}$), which are defined in Table 4. If the relative free energies of the three states at a particular denaturant concentration are known, their fractional populations are given by the partition function, and the matrix *F* is fully specified. The matrix *S* is then obtained by taking the pseudoinverse:

$$S = A\left(F^T\right)^{+} \tag{2}$$

The goal is to minimize the error between the experimental data *A* and the model SF^T. The objective function in our model is as follows:

$$\|A - SF^T\|_2 + \alpha\|S\|_2 \tag{3}$$

The operation $\|\cdot\|_2$ is the matrix 2-norm. Our objective function is a combination of an error term and a weighted ($\alpha = 0.01$) regularization term. The inclusion of a regularization term was necessary in order to obtain physically meaningful solutions. Without it, the optimization returns spectra which can be nearly infinite in magnitude, but low in fractional population. The regularization weight was set at 0.01, as this value allowed the accurate recovery of thermodynamic parameters in a synthetic data set. The objective function was minimized using the Nelder-Mead simplex algorithm, as implemented in the optim() function of the statistical environment R (49).

Results

Cyanovirin-N is composed of two tandem repeats of 50 and 51 amino acids which share 32% sequence identity (Figure 1A) (51). These repeats give rise to two pseudo-symmetric domains, termed Domain A (residues 1–39 and 90–101) and Domain B (residues 40–89). Each domain contains a carbohydrate-binding site (52) and is stabilized by an internal disulfide bond. The two domains make extensive contacts and share a hydrophobic core. Despite the structural pseudo-symmetry of the two domains, Domain A is not contiguous in sequence, and contains both amino- and carboxy-termini, while the structurally-equivalent residues within Domain B are connected by a short linker (Lys48–Trp49–Gln50–Pro51–Ser52–Asn53).

We designed mutants using the Pro51Gly (P51G) mutation as background. This mutation, located in the linker region of Domain B, has previously been shown both to stabilize monomeric CVN and to increase the yield of monomeric protein relative to the domain-swapped dimer, a metastable state which is formed when the protein is refolded at high concentrations (10). Wildtype CVN lacking this mutation forms a domain-swapped dimer under crystallographic conditions. The crystal structure of domain-swapped CVN shows significant rearrangement in the linker region, and exhibits changes in the ϕ and ψ torsion angles, which are most profound for Ser52 and Asn53 (53). The P51G substitution is thought to stabilize CVN by alleviating backbone strain induced by Pro51, which imparts a positive ϕ angle on Ser52 and places it into a disfavored portion of the Ramachandran plot. In molecular dynamics simulations of wild-type CVN, we observed significant

rearrangements in this linker region, including formation of a *cis*-peptide bond between Pro51 and Ser52 (Y.K. Fujimoto and DFG, in preparation).

Ser11, Ser20, and Thr61: Buried Polar Residues Provide Targets for Rational Design

Initial insight into stabilizing CVN came from the observation that Domain A side chains of Ser11 and Ser20 are buried within the hydrophobic core of the protein, with solvent accessibility of 3.3% and 12.2%, respectively, relative to the Ser side chain in an Ala–Ser–Ala tripeptide (Figure 1B). While the average NMR structure (PDB ID 1IIY) does not show a hydrogen bond between these groups, the interaction is nearly always formed in molecular dynamics simulations of CVN.

In proteins, polar groups are more frequently found on the surface than in the core and often pay significant desolvation penalties upon burial (54). Thr61 and Ala71, which are the Domain B symmetric counterparts of Ser11 and Ser20, pack against one another, are within van der Waals contact of Phe54, and contribute to the hydrophobic core of the protein. These packing interactions are absent in Domain A, despite the proximity of Phe4 (the symmetric equivalent of Phe54) to Ser11 and Ser20 (55).

Ser11 and Ser20 are among the least solvent-exposed side chains in the protein (Table S??). The side chains of these two residues appear to make a direct hydrogen bond in the solution structure of CVN, and this favorable interaction may compensate for the desolvation penalty they pay for burial within a hydrophobic environment (22, 56). Thr61, like Ser11, is also excluded from solvent, and only exposes 2.5% of the side-chain surface area, when compared to a reference tripeptide. Unlike its symmetry-related counterpart, Thr61 does not form a buried hydrogen bond. We hypothesized that replacement of Ser11, Ser20 or Thr61 with an appropriate hydrophobic isostere would stabilize the protein, as replacement of Ser with Ala, and of Thr with Val, was previously shown to be favorable in buried positions within T4 lysozyme (21).

Continuum Electrostatic Calculations Reveal a Destabilizing Buried Polar Bridge

We performed Poisson–Boltzmann (PB) continuum electrostatic calculations in order to estimate the contribution of these groups to protein stability. The calculations afforded the desolvation penalty and the strength of the pairwise electrostatic interactions between the side chains of our target positions (Ser11, Ser20, and Thr61) and other groups in the protein. Ala71, the symmetry-related partner of Ser20, was also included as a control. The calculations evaluate the effect of substituting a side chain with a hydrophobic isostere having the same size and shape, but devoid of charge (22, 56).

Domain A side chains of Ser11 and Ser20 pay a +2.92 and +2.96 kcal/mol desolvation penalty ($\Delta\Delta G°_{solv}$), respectively, upon burial within the core of CVN (Table 1). For each side chain, the desolvation penalty is opposed by the favorable electrostatic interactions with the other groups in the protein. These interaction free energies ($\Delta\Delta G°_{inter}$) sum to −4.25 and −3.39 kcal/mol for Ser11 and Ser20, respectively. The major portion of these interactions comes from the direct interaction between the two groups of −3.08 kcal/mol.

In order to evaluate whether a polar group or its hydrophobic isostere is more favorable to the stability of a protein fold, we calculate the mutation free energy $\Delta\Delta G°_{mut}$, which is tabulated as the sum of $\Delta\Delta G°_{solv}$ and $\Delta\Delta G°_{inter}$. In our model, $\Delta\Delta G°_{mut}$ corresponds to "turning on" the partial charges on the molecular group of interest. A negative value of $\Delta\Delta G°_{mut}$ thus indicates that a group contributes more favorably in the charged state than in the hydrophobic state. For Ser11 and Ser20, an isosteric substitution is unfavorable by +1.33 and +0.43 kcal/mol, respectively (Table 1). A single isosteric substitution is unfavorable at either position 11 or 20 because the remaining group has lost a significant electrostatic

interaction partner, yet still pays a desolvation penalty for burial. However, the *simultaneous* replacement of Ser11 and Ser20 with hydrophobic isosteres is predicted as favorable with an estimated effect of −1.32 kcal/mol. Figure 2A shows the complete *in silico* thermodynamic cycle which follows the replacement of either or both serine side chains with hydrophobic isosteres (denoted S^0).

We set out to compare the magnitude of the electrostatic interactions to that of van der Waals interactions experienced by Ser11 and Ser20 using the coordinates taken from a molecular dynamics simulation of CVN. While the electrostatic interactions of the two side chains with the rest of the protein are favorable, the hydrogen bond between these two groups results in overlap of their van der Waals radii, and an unfavorable direct interaction of +0.85 kcal/mol. These results suggest that the direct interaction between the side chains of Ser11 and Ser20 is mainly electrostatic in nature.

In Domain B, we found that Thr61 pays a desolvation penalty of +2.30 kcal/mol. This penalty is computed to be nearly perfectly offset by favorable electrostatic interactions with the rest of the protein of −2.51 kcal/mol, and the overall effect of replacing Thr61 with a hydrophobic isostere is only slightly unfavorable by +0.21 kcal/mol (Table 1). As expected, the nonpolar Ala71 does not participate in significant electrostatic interactions with Thr61.

The incorporation of an exact hydrophobic isostere for Ser into CVN is not possible. Among naturally-occurring amino acids, Ala is the most conservative nonpolar substitution for Ser and is the closest to an isosteric substitution. We thus designed a mutant termed AATA. In our nomenclature, a protein variant is identified by a four-letter name which designates its amino-acid identity at positions 11, 20, 61, and 71. For example, P51G variant with no other changes would be denoted SSTA, and AATA is P51G with Ser11Ala and Ser20Ala mutations (Table 2). We next employed protein design calculations to determine if other combinations of naturally-occurring amino acids could lead to increased stabilization of CVN.

Design of a Stabilizing Network of Mutations

The PB analysis provides information about isosteric substitutions, but does not consider changes in packing which may result from non-isosteric mutations. Thus we went on to examine potential effects of varying the size and shape of the side chain at positions 11 and 20. In order to determine if the structure of CVN can accommodate other amino-acid side chains at positions 11 and 20, we used a computational protein design algorithm where the design positions were allowed to simultaneously vary in sequence to {Ala, Ser, Thr, Val, Phe, Leu, Ile.} These substitutions were selected due their nonpolar nature or their occurrence within wild-type CVN at the chosen positions. Neighboring residues Phe4, Leu18, Ile34, Leu36, Leu87, Ile91, and Leu98 were allowed to, if necessary, adopt a different rotamer in order to accommodate larger side chains at the design positions. The protein backbone, as well as the remaining side chains, was kept fixed. The design algorithm computed the stability of a conformational arrangement of amino-acid side chains relative to an unfolded state approximation in which the free energy of the unfolded polypeptide is taken as a sum of energies of the individual amino-acid residues in its primary structure (36, 57).

The computational design calculations suggested that a number of mutants would be more stable than the wild-type protein (Figure 2B). In particular, S11V.S20A, S11I.S20A, and S11A.S20A were calculated to be more stable than wildtype by −12.5, −9.0, and −8.6 kcal/mol, respectively. The results suggest a clear preference of alanine to serine at position 20, with X11.S20 always worse energetically than X11.A20 for every substitution X considered. The insights from the protein design calculations led to mutants VATA and IATA (Table 2).

It is worth noting that in interpreting these results the energetic trends are more meaningful than magnitudes. The computational procedure uses a discrete representation of the configurational space and employs an approximate energy function, which ignores important contributions to protein stability (such as side chain configurational entropy). Thus, we do not expect the magnitudes of stabilization from computational design to be quantitatively predictive, but the calculations do reveal potentially stabilizing replacements and complement the PB analysis.

Lessons from CVN Homologues

When CVN was originally discovered, it showed no sequence similarity to any other protein. The structure of CVN revealed a novel fold, which possessed only distant (domain-level) topological similarity to known protein folds. More recently, CVN homologues have been discovered in other prokaryotes (58) and in eukaryotes (59). Pfam release 24.0 lists 116 putative CVN domain sequences across 22 species (60). These genomic data suggest that the CVN domain is a module which often exists within larger multidomain proteins, the function of which is presently unknown (59). There is evidence, however, that these domains adopt structures similar to wild-type CVN and are functional lectins. The solution structures of CVN homologues from *Tuber borchii*, *Ceratopteris richardii*, *Neurospora crassa*, *Microcystis aeruginosa PCC7806*, and *Magnaporthe oryzae* have recently been characterized, and all adopt the same fold (61 –63).

A SeqLogo (64) representation of a multiple sequence alignment of a subset of CVN homologues is shown in Figure 1B. This sequence alignment revealed that across the CVN family, the position corresponding to Ser11 is frequently Val or Ile, while the equivalent of Ser20 is most commonly Ala. Thr61, the Domain B symmetric equivalent of Ser11, is often substituted with Val or Ile. Unfortunately, the size of the CVN family is not yet large enough to determine whether substitutions at these or other positions within the CVN domain are independent or correlated in their conservation (65, 66).

The sequence conservation data are consistent with both computational and intuition-based insights, and reveal a strong preference for nonpolar side chains at the core positions 11, 20, 61, and 71. We thus designed additional symmetrizing mutants, VAVA and IAIA, in order to incorporate the hydrophobic side chains of Val and Ile at position 61 (Table 2) in the background of putative stabilizing mutations in Domain A. The distance between the design positions in the two domains of CVN led us to hypothesize that the effect of substitutions at position 61 in Domain B would be additive with the effect of substituting positions 11 and 20 in Domain A.

Expression Construct: Consistency with Previous Work

The biophysical, structural, and inhibitory properties of CVN have been characterized by a number of different laboratories, with some variation in the exact amino-acid sequence of the protein being studied, in particular at the N-and C-terminal regions. The first source of variation is the identity of the N-terminal amino acid. When the protein is expressed cytoplasmically using the BL21 *E. coli* strain, it accumulates in inclusion bodies. Consequently, the N-terminal Met residue cannot be processed by the *E. coli* methionine aminopeptidase (confirmed by mass spectrometry) and is retained on the polypeptide chain. In order to facilitate disulfide bond formation, CVN has also been successfully expressed as a folded protein in the periplasmic space, and the removal of the localization tag leaves Leu as the N-terminal residue (67, 68). In addition, other biophysical studies used CVN variants with an additional N-terminal Gly–Ser–His–Met–Gly sequence which remained after thrombin cleavage. CVN containing these additional five animo acids at the N-terminus

exhibited anti-HIV activity which is indistinguishable from wild-type protein (10, 69, 70), yet may have thermodynamic properties different from wild type.

Due to the proximity of the Glu101 side chain to the N-terminus in the solution NMR structure of CVN, it is likely that a salt bridge is formed between these two groups at the experimental pH of 6.0. We hypothesized that the addition of a methionine or a longer linker at the N-terminus could thus have an effect on protein stability by altering the salt bridge geometry, or by interfering with its formation. To evaluate this, we expressed the P51G variant in the *E. coli* periplasm using the PelB localization tag. This variant (termed ΔM) is identical in sequence to P51G, except it is one residue shorter due to the removal of the leader tag.

A second source of variation in the sequence of CVN is the presence of a C-terminal His-tag (complete sequence Leu–Glu–His_6). The thermodynamic parameters of non-His-tagged and His-tagged wild-type CVN are identical (12), and a number of additional studies have used this construct (71). For ease of purification, our variants contain this C-terminal His-tag; the amino-acid sequences of the proteins in this study are given in Supporting Figure S??.

Designed Variants Adopt the Wild-type Fold and are More Stable

Designed CVN homologues were expressed recombinantly in *E. coli* and (with the exception of ΔM) purified from inclusion bodies under denaturing conditions. The proteins possess identical secondary and tertiary structure, as judged by far- and near-ultraviolet CD spectroscopy (Figure 3). The far-UV CD spectra of the variants were similar to previously-published spectra for wild-type CVN (70). In addition, the intrinsic tryptophan fluorescence emission for each variant was blue-shifted 20 nm compared to the fluorescence of GuaHCl-denatured protein (not shown), indicating the burial of the unique Trp in CVN (Trp49). All variants were judged to be monomeric by analytical gel filtration chromatography (Supporting Figure S??) at a range of concentrations up to 141 μm. In order to further confirm conservation of the CVN fold upon mutation, we collected HSQC spectra of P51G and of IAIA. The spectra (Supporting Figure S??) show comparable resonance dispersion in both proton and nitrogen dimensions for both proteins, and are consistent with a well-folded structure.

Figure 4 shows the thermal denaturation of CVN homologues monitored by CD spectroscopy, and the extracted thermodynamic parameters are given in Table 3. The parameters determined from thermal denaturation represent apparent values, since thermal denaturation of CVN is not fully reversible, as judged by recovery of CD signal upon heating and re-cooling. For some variants, such as P51G, the signal recovery was close to 90%. However, the denaturation of the more hydrophobic variants was less reversible, and visible aggregation within the cuvette was observed.

For P51G, the midpoint of thermal denaturation was 64.3 °C, consistent with previous studies of His-tagged CVN variants (12). Each of the designed variants were more thermostable than P51G, with thermal stabilization ranging from 5.2 °C for AATA to 11.5 °C for IAIA. In general, greater stabilization was achieved by larger amino-acid substitutions at the design positions. In particular, AATA was less thermostable than VATA, which in turn was less thermostable than IATA. The ΔM variant was more thermostable than P51G by 2.3 °C.

In order to further characterize the stability of the designed variants, we undertook unfolding studies using the chaotrope guanidine hydrochloride (GuaHCl). To facilitate comparison with earlier work, we followed denaturation using the ratio of the intensity of fluorescence emission at 330 nm and at 360 nm, I_{330}/I_{360}. We observed that P51G and other CVN

variants look longer than 48 h to fully unfold when exposed to moderate concentrations of GuaHCl at room temperature. For example, incubation of P51G overnight with increasing concentrations of GuaHCl yielded a biphasic denaturation profile. In contrast, complete equilibration could be achieved over the course of 72 h, and at equilibrium, a single sigmoidal transition was observed (Figure 5A). This is an important observation, since it directly demonstrates that a lengthy equilibration time is required to obtain accurate unfolding curves. The data collected with the shorter overnight equilibration could be forcibly fit to a single unfolding transition, but would yield an apparent stability significantly higher than the actual value. We believe that these effects may have complicated some prior analyses of CVN thermodynamics.

In order to determine the incubation time needed for complete equilibration, we investigated the denaturation kinetics of P51G by following intrinsic tryptophan fluorescence after addition of 3 M GuaHCl (Figure 5B). At this denaturant concentration, the protein appears less than 10% folded at equilibrium, but is 60–70% folded after an overnight incubation at 25 °C, as judged by the fluorescence emission intensity. The recorded denaturation kinetics were complex and could not be modeled by a single exponential decay, or by a sum of two decays, as judged by the non-randomness of the residuals. This implies that P51G denaturation proceeds through at least one kinetic intermediate state, the decay of which is slow. These experiments allowed us to determine the apparent half-life of denaturation (13 h) at 3 M GuaHCl and the time for complete equilibration of CVN variants; 72 h was adequate to completely equilibrate all proteins discussed in this work.

To investigate the equilibrium stability of CVN variants to chemical denaturation, we monitored their intrinsic protein fluorescence and circular dichroism signal. The two techniques allowed us to monitor two probes, with fluorescence monitoring the burial of the unique Trp fluorophore of CVN and CD monitoring the state of the polypeptide backbone. Figure 6 shows far-UV circular dichroism spectra of P51G taken at equilibrium as a function of increasing concentrations of GuaHCl. These spectra lack an isodichroic (isosbestic) point, indicating that the equilibrium folding is also not two-state (72), but rather that the equilibrium unfolding of CVN by GuaHCl is complicated, and populates intermediate states. While both kinetic and equilibrium experiments revealed the presence of multiple states, it is unknown whether intermediate states present in the kinetic and equilibrium denaturation of CVN are structurally related.

In order to extract the thermodynamic parameters from the far-UV CD spectra of CVN variants we employed a thermodynamic model which decomposes spectroscopic data into the spectra of three thermodynamic states and their fractional populations, as given by their free energy differences (73 – 75). The data were fit globally at all wavelengths in order to obtain the most reliable parameters. We note that no single far-UV wavelength showed a clear biphasic transition. When the data were fit to a simple two-state model, the derived thermodynamic parameters were dependent on choice of wavelength — for instance, a plot of apparent fraction unfolded at θ_{220} was non-coincident with fraction unfolded at θ_{235} (data not shown). Fluorescence spectroscopy was not sensitive to the presence of an equilibrium intermediate, and thus tryptophan emission spectra were not used in the thermodynamic model.

The parameters derived from the spectral decomposition are summarized in Table 4. We found that substitutions at positions 11, 20, and 61 stabilized *both* the native and intermediate states. Substitutions of Ser11 and Ser20 with Ala stabilized the native state by 0.94 kcal/mol relative to P51G, in agreement with the computational prediction. Further substitutions at position 11 to Val and Ile stabilized the native state by 1.03 and 2.56 kcal/mol, respectively, relative to P51G.

Substitutions at position 61 (VAVA and IAIA) were accompanied by significant *m*-value changes. The *m*-value for the native to intermediate transition for VAVA is 1.65 while it is 2.32 for the P51G background and 2.1 for ΔM. For VAVA, the free energy difference between the native and denatured states ($\Delta G°_{N\rightarrow D}$) is 6.75 kcal/mol, which is comparable to the stability of the ΔM variant (6.78 kcal/mol). However, the denaturant concentration at which the VAVA unfolding transition is 50% complete is 0.4 M GuaHCl higher than that of ΔM. This is illustrated in Figure 7, which shows the equilibrium denaturation profiles for the designed CVN homologues. The profiles are generated from the equilibrium thermodynamic parameters, and give the fractional populations of the native and intermediate states for all CVN variants as a function of GuaHCl. We found that the equilibrium intermediate state remains significantly populated for all variants at the highest denaturant concentrations tested (up to 6 M GuaHCl), and that all variants populate the intermediate state up to 25% under some experimental conditions.

Carbohydrate Microarray Binding Analysis Shows that Variants Retain Native Binding Specificity

In order to assess the function of our designed variants we carried out fluorescence-detected glycan microarray binding experiments (76) with P51G included as a control. The variants (ΔM not included) were heterogeneously labeled with NHS-Fluorescein at a surface lysine side chain, with the amount of labeling estimated to be approximately 0.5 dye molecules per protein molecule.

Microarray binding experiments confirmed that the designed variants retain the exquisite carbohydrate-binding specificity of wild type CVN. The Consortium for Functional Glycomics (CFG) mammalian printed array (version 4.1) contains 465 diverse carbohydrates, both linear and branched. Consistent with previous microarray studies, we observed that the designed mutants bound exclusively to carbohydrates containing the Manα(1→2)Man disaccharide, which has been previously demonstrated to be the minimal carbohydrate recognized by the two independent binding sites of CVN (61, 76). Figure 8B shows the raw fluorescence signal for our CVN variants. For all variants, the fluorescence emission followed a decaying profile when ranked by the average intensity across all variants, and we observed binding to 10 distinct carbohydrates, shown in Figure 8A. Each of the bound carbohydrates, with the exception of 212, corresponds to a substructure of $Man_9GlcNAc_2$, the largest (tri-antennary) *N*-linked high-mannose oligosaccharide present on the viral surface. Compound 212 is equivalent to the D3 arm of $Man_9GlcNAc_2$ extended by an additional Manα(1→2)Man at the non-reducing end.

As expected, the designed homologues bound to oligosaccharides corresponding to the D1, D2, and D3 arms of $Man_9GlcNAc_2$; the highest fluorescence was observed for glycan 314, which contains all three arms. The second-highest fluorescence was for glycan 212 and the five largest signals were seen for carbohydrates containing the Manα(1→2)Manα(1→2)Man trisaccharide. This is consistent with the oligosaccharide specificity previously observed in a computational study of CVN–carbohydrate binding (77). We must note that the fluorescence intensity cannot be interpreted as affinity for a particular carbohydrate, due to possibilities of a nonuniform degree of labeling and potential differences in density and presentation of various carbohydrates on the microarray slides.

Our stabilized CVN variants showed specificity identical to that of P51G. The proteins did not bind to the large majority of the carbohydrates present on the microarray. Two such compounds (49 and 214) represent the core of $Man_9GlcNAc_2$ and are shown in Figure 8. Consistent with previous studies (61, 76), our mutants specifically bound to the terminal arms of Man_9, and not the glycan core. The full identities of the glycans bound by our CVN variants are given in Table S??.

Discussion

CVN contains a number of polar side chains which are sequestered from solvent. Of the 27 side chains which have a calculated average solvent accessibility below 15%, 10 have polar character (Table S??). These include the disulfide-bonded Cys8–Cys22 and Cys58–Cys73, Trp49, and Asn42 and Asn93, which expose 6% of the Asn surface area. The latter two residues are highly conserved (Figure 1B), and their side chains make key hydrogen bonding interactions which stabilize the two domains of CVN (55). Poisson–Boltzmann calculations suggest that both of these residues are able to overcome their desolvation penalties of +4.30 kcal/mol through favorable electrostatic interactions, and that replacing either Asn42 or Asn93 with a hydrophobic isostere is disfavored by +3.30 and +3.01 kcal/mol, respectively. The remaining three buried polar groups are Ser11, Ser20, and Thr61. Using continuum electrostatic calculations we quantified the desolvation penalty paid by Ser11, Ser20, and Thr61 upon burial within the core of CVN. The single replacement of any of the three groups with a hydrophobic isostere was predicted to be electrostatically unfavorable. However, the Ser11.Ser20 pair was found to contribute unfavorably to CVN stability, and the calculations suggested that the simultaneous replacement of the two serine side chains with hydrophobic isosteres could lead to increased protein stabilization. This analysis highlights the importance of considering multiple mutations during protein design and is similar to observations made with buried salt bridges. The replacement of a single residue in a salt bridge or a polar unit is often highly destabilizing. For instance, the substitution of any charged residue with Ala in a complex “Arg–Glu-Arg” salt bridge triad within Arc repressor has been demonstrated to have a detrimental effect on protein stability and function. However, the simultaneous replacement of the triad with “Met–Tyr–Leu” yielded a variant which significantly more stable that wild type (78). Pairs of buried interacting polar residues can be thought of as analogous to buried salt bridges, albeit with less overall charge density.

In contrast to Ser11 and Ser20, Poisson–Boltzmann calculations suggest that the electrostatic contribution of Thr61 to the stability of CVN is neutral (Table 1). However, we found that the protein could be stabilized by substitutions at position 61. For example, VAVA was 4.7 °C more thermostable than VATA, and IAIA was more thermostable than IATA by the same amount. The observed stabilization may arise due to enhanced van der Waals packing interactions, or may indicate an overestimation of intermolecular interactions with respect to the desolvation penalty in the calculations.

In agreement with the sequence analysis of the CVN family, the computational redesign of Domain A successfully predicted that the protein is able to accommodate Ile, Val, or Ala at position 11. The computational redesign of Domain B, where Thr61 and Ala71 were allowed to vary in sequence, predicted that only a single sequence (T61A.A71) to be more favorable than wild-type (by −2.4 kcal/mol). A number of additional sequences, such as T61V.A71 were predicted to be significantly worse (+10.7 kcal/mol). This could indicate that Domain B is less tolerant of repacking or could highlight limitations of the fixed-backbone model. In this case, since experimental characterization demonstrated stabilization with substitutions at Thr61, the result is most likely an artifact of the fixed-backbone approximation, an effect which has been observed previously (79, 80).

The thermostabilization of our designed CVN variants relative to P51G most likely includes contributions from several factors. The initial 5.3 °C stabilization of AATA relative to P51G likely arises primarily from the removal of the unsatisfied polar groups at positions 11 and 20. Alanine is smaller than serine in volume, so the substitutions are unlikely to add van der Waals packing interactions. Intriguingly, VATA does not have significantly improved thermostability relative to AATA, despite the increased hydrophobicity and size of the Val residue compared to Ala. The IATA mutant has the largest side chain that we expected

could be successfully incorporated at position 11, and it showed improved thermostability relative to AATA by 1.5 °C.

We found that the stabilization of the designed variants can be qualitatively explained by considering the additional nonpolar surface area introduced by the mutations (81). Table 5 gives nonpolar surface areas calculated for all the mutants at positions 11, 20, 61, and 71, as well as the free energy difference ($\Delta\Delta G^\circ_\phi$) expected from the surface area change; the $\Delta\Delta G^\circ_\phi$ is estimated using the "atomic solvation parameter" of 16 cal $Å^{-2}$ mol^{-1} (81).

Our kinetic denaturation studies of P51G reveal that at room temperature the protein requires nearly three days to fully reach equilibrium at intermediate GuaHCl concentrations and suggest that at least one kinetic unfolding intermediate is populated. Equilibrium circular dichroism spectra taken at increasing concentrations of GuaHCl showed an absence of an isodichroic point, and are also incompatible with a two-state model of folding. Singular value decomposition (73) of GuaHCl-dependent CD spectra revealed the presence of three significant spectral components above the experimental noise (Figure S??), consistent with multistate equilibrium denaturation.

A number of previous studies (10, 12, 82) have examined the equilibrium chemical denaturation of wild-type CVN, P51G, and several functional homologues using intrinsic protein fluorescence. Those studies reached a different conclusion about the stability of CVN, and in particular, the denaturation of CVN was deemed two-state. We believe that the present work provides more precise thermodynamic parameters and apparent stabilities than were previously obtained, since earlier studies measured protein stability after overnight incubation with GuaHCl. Overnight incubation is insufficient to allow the system to equilibrate, and this leads to estimated stabilities which are larger than the actual thermodynamic stability (10, 12).

The structural reason for the slow denaturation of the P51G variant of CVN is unknown. While this variant does not contain any proline residues, its disulfide bonds may contribute to the slow denaturation kinetics, as cystine residues have strong preferences for specific C–S–S–C dihedral angles and high energetic penalties for deviation from them (83). For example, both CHARMM (84) and AMBER (85) molecular mechanics force fields implement this dihedral angle as two isoenergetic minima ($\chi_{ss} = \pm 84°$) separated by a barrier height of 6.4 kcal mol^{-1}, comparable to that of the rotation about the peptide bond. Similarly slow unfolding kinetics have been observed for disulfide-containing hen egg white lysozyme (86). However, the slow unfolding kinetics may also be due to other structural factors, having also been observed for the cystine-less four-helix bundle Rop (87).

We found that the presence of an N-terminal methionine residue has a measurable effect on CVN stability. The ΔM variant possessed increased stability when compared to P51G (Table 3). Biophysical studies of the B1 domain of staphylococcal protein G (β1) reported a similar effect; the methionine-containing and methionine-lacking forms of the protein differ by 1.7 kcal/mol in stability (88). Similar effects have been observed for α-lactalbumin (89, 90). For CVN, the effect is not as drastic, yet is significant; we hypothesize that the difference between the stability of P51G and ΔM stems from the fact that the *native* N-terminus, which has the methionine removed, participates in favorable electrostatic interactions in the native state.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Carole Bewley for providing wtCVN NMR assignments, Dr. Hümeyra Taşkent-Sezgin for help with initial work on protein expression, and Dr. Bruce Tidor for access to the ICE software package.

Supported by NIH grant GM086199 to DFG and NSF grant MCB 0919860 to DPR.

Abbreviations used

CVN	cyanovirin-N
GuaHCl	guanidine hydrochloride
HIV	human immunodeficiency virus
HSQC	heteronuclear single quantum coherence
PB	Poisson–Boltzmann

References

1. Gustafson KR, Sowder RC, Henderson LE, Cardellina JH, McMahon JB, Rajamani U, Pannell LK, Boyd MR. Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-N, an anti-HIV (human immunodeficiency virus) protein from the cyanobacterium Nostoc ellipsosporum. Biochem Biophys Res Commun. 1997; 238:223–228. [PubMed: 9299483]
2. Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, Gulakowski RJ, Wu L, Rivera MI, Laurencot CM, Currens MJ, Cardellina JH, Buckheit RW, Nara PL, Pannell LK, Sowder RC, Henderson LE. 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. 1997; 41:1521–1530. [PubMed: 9210678]
3. Balzarini J. Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. Nat Rev Microbiol. 2007; 5:583–597. [PubMed: 17632570]
4. Dey B, Lerner DL, Lusso P, Boyd MR, Elder JH, Berger EA. Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. J Virology. 2000; 74:4562–4569. [PubMed: 10775592]
5. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM. Antibody neutralization and escape by HIV-1. Nature. 2003; 422:307–312. [PubMed: 12646921]
6. Scanlan CN, Offer J, Zitzmann N, Dwek RA. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. Nature. 2007; 446:1038–1045. [PubMed: 17460665]
7. Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR. Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. Antiviral Res. 2003; 58:47–56. [PubMed: 12719006]
8. Tiwari V, Shukla SY, Shukla D. A sugar binding protein cyanovirin-N blocks herpes simplex virus type-1 entry and cell fusion. Antiviral Res. 2009; 84:67–75. [PubMed: 19665490]
9. O'Keefe BR, Smee DF, Turpin JA, Saucedo CJ, Gustafson KR, Mori T, Blakeslee D, Buckheit R, Boyd MR. Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. Antimicrob Agents Chemother. 2003; 47:2518–2525. [PubMed: 12878514]
10. Barrientos LG, Louis JM, Botos I, Mori T, Han Z, O'Keefe BR, Boyd MR, Wlodawer A, Gronenborn AM. The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of x-ray and NMR structures. Structure. 2002; 10:673–686. [PubMed: 12015150]
11. Colleluori DM, Tien D, Kang F, Pagliei T, Kuss R, McCormick T, Watson K, McFadden K, Chaiken I, Buckheit RW, Romano JW. Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. Protein Expr Purif. 2005; 39:229–236. [PubMed: 15642474]

12. Mori T, Barrientos LG, Han Z, Gronenborn AM, Turpin JA, Boyd MR. Functional homologs of cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. Protein Expr Purif. 2002; 26:42–49. [PubMed: 12356469]
13. Sexton A, Drake PM, Mahmood N, Harman SJ, Shattock RJ, Ma JK-C. Transgenic plant production of cyanovirin-N, an HIV microbicide. FASEB J. 2006; 20:356–388. [PubMed: 16354721]
14. Matei E, Zheng A, Furey W, Rose J, Aiken C, Gronenborn AM. Anti-HIV activity of defective cyanovirin-N mutants is restored by dimerization. J Biol Chem. 2010; 285:13057–13065. [PubMed: 20147291]
15. Matei E, Furey W, Gronenborn AM. Solution and crystal structures of a sugar binding site mutant of cyanovirin-N: no evidence of domain swapping. Structure. 2008; 16:1183–1194. [PubMed: 18682220]
16. Fromme R, Katiliene Z, Giomarelli B, Bogani F, Mahon JM, Mori T, Fromme P, Ghirlanda G. A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. Biochemistry. 2007; 46:9199–9207. [PubMed: 17636873]
17. Barrientos LG, Matei E, Lasala F, Delgado R, Gronenborn AM. Dissecting carbohydrate-cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. Protein Eng Des Sel. 2006; 19:525–535. [PubMed: 17012344]
18. Chang LC, Bewley CA. 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. J Mol Biol. 2002; 318:1–8. [PubMed: 12054763]
19. Keeffe JR, Gnanapragasam PNP, Gillespie SK, Yong J, Bjorkman PJ, Mayo SL. Designed oligomers of cyanovirin-N show enhanced HIV neutralization. Proc Natl Acad Sci USA. 2011
20. Dahiyat BI, Mayo SL. Probing the role of packing specificity in protein design. Proc Natl Acad Sci USA. 1997; 94:10172–10177. [PubMed: 9294182]
21. Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, Matthews BW. Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala→Ser and Val→Thr substitutions in T4 lysozyme. Biochemistry. 1993; 32:11363–11373. [PubMed: 8218201]
22. Hendsch ZS, Jonsson T, Sauer RT, Tidor B. Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. Biochemistry. 1996; 35:7621–7625. [PubMed: 8672461]
23. Mok YK, Elisseeva EL, Davidson AR, Forman-Kay JD. Dramatic stabilization of an SH3 domain by a single substitution: roles of the folded and unfolded states. J Mol Biol. 2001; 307:913–928. [PubMed: 11273710]
24. Spector S, Wang M, Carp SA, Robblee J, Hendsch ZS, Fairman R, Tidor B, Raleigh DP. Rational modification of protein stability by the mutation of charged surface residues. Biochemistry. 2000; 39:872–879. [PubMed: 10653630]
25. Robinson CR, Sauer RT. Striking stabilization of Arc repressor by an engineered disulfide bond. Biochemistry. 2000; 39:12494–12502. [PubMed: 11015231]
26. Nauli S, Kuhlman B, Baker D. Computer-based redesign of a protein folding pathway. Nat Struct Biol. 2001; 8:602–605. [PubMed: 11427890]
27. Zhou HX, Hoess RH, DeGrado WF. In vitro evolution of thermodynamically stable turns. Nat Struct Biol. 1996; 3:446–451. [PubMed: 8612075]
28. Predki PF, Agrawal V, Brünger AT, Regan L. Amino-acid substitutions in a surface turn modulate protein stability. Nat Struct Biol. 1996; 3:54–48. [PubMed: 8548455]
29. Anil B, Song B, Tang Y, Raleigh DP. Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability. J Am Chem Soc. 2004; 126:13194–13195. [PubMed: 15479052]
30. Cho J-H, Sato S, Raleigh DP. Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. J Mol Biol. 2004; 338:827–837. [PubMed: 15099748]

31. Malakauskas SM, Mayo SL. Design, structure and stability of a hyperthermophilic protein variant. Nat Struct Biol. 1998; 5:470–475. [PubMed: 9628485]
32. Green DF, Tidor B. Evaluation of electrostatic interactions. Curr Protoc Bioinformatics. 2002
33. Gilson MK, Rashin A, Fine R, Honig B. On the calculation of electrostatic interactions in proteins. J Mol Biol. 1985; 184:503–516. [PubMed: 4046024]
34. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science. 1995; 268:1144–1149. [PubMed: 7761829]
35. Nina M, Beglov D, Roux B. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. J. Phys. Chem. B. 1997; 101:5239–48.
36. Green DF, Dennis AT, Fam PS, Tidor B, Jasanoff A. Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. Biochemistry. 2006; 45:12547–12559. [PubMed: 17029410]
37. Looger LL, Hellinga HW. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. J Mol Biol. 2001; 307:429–445. [PubMed: 11243829]
38. Gordon DB, Mayo SL. Radical performance enhancements for combinatorial optimization algorithms based on the dead-end elimination theorem. J Comput Chem. 1998; 19:1505–1514.
39. Leach AR, Lemon AP. Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. Proteins. 1998; 33:227–239. [PubMed: 9779790]
40. Maeyer MD, Desmet J, Lasters I. The dead-end elimination theorem: mathematical aspects, implementation, optimizations, evaluation, and performance. Methods Mol Biol. 2000; 143:265–304. [PubMed: 11084910]
41. Lasters I, Maeyer MD, Desmet J. Enhanced dead-end elimination in the search for the global minimum energy conformation of a collection of protein side chains. Protein Eng. 1995; 8:815–822. [PubMed: 8637851]
42. Desmet J, Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature. 1992; 356:539–542. [PubMed: 21488406]
43. Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. Proteins. 2000; 40:389–408. [PubMed: 10861930]
44. Brooks BR, Brooks CL III, Mackerell AD Jr. Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The Biomolecular Simulation Program. J Comput Chem. 2009; 30:1545–1614. [PubMed: 19444816]
45. MacKerell A, Bashford D, Bellott M, Dunbrack RL, Evanseck J, Field M, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F, Mattos C, Michnick S, Ngo T, Nguyen D, Prodhom B, Reiher W, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998; 102:3586–3616.
46. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. The Journal of Physical Chemistry. 1996; 100:1578–1599.
47. Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Prot. 2006; 1:2527–2535.
48. Greenfield NJ. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. Nat Prot. 2006; 1:2733–2741.
49. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2011. ISBN 3-900051-07-0
50. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]
51. Bewley CA. Solution structure of a cyanovirin-N:Manα1–2Manα complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. Structure. 2001; 9:931–940. [PubMed: 11591348]

52. Bewley CA, Otero-Quintero S. The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man_8 D1D3 and Man_9 with nanomolar affinity: implications for binding to the HIV envelope protein gp120. J Am Chem Soc. 2001; 123:3892–3902. [PubMed: 11457139]
53. Yang F, Bewley CA, Louis JM, Gustafson KR, Boyd MR, Gronenborn AM, Clore GM, Wlodawer A. Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. J Mol Biol. 1999; 288:403–412. [PubMed: 10329150]
54. Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT. Deciphering the message in protein sequences: tolerance to amino acid substitutions. Science. 1990; 247:1306–1310. [PubMed: 2315699]
55. Bewley CA, Gustafson KR, Boyd MR, Covell DG, Bax A, Clore GM, Gronenborn AM. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. Nat Struct Biol. 1998; 5:571–578. [PubMed: 9665171]
56. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci. 1994; 3:211–226. [PubMed: 8003958]
57. Lippow SM, Wittrup KD, Tidor B. Computational design of antibody-affinity improvement beyond in vivo maturation. Nat Biotechnol. 2007; 25:1171–1176. [PubMed: 17891135]
58. Kehr J-C, Zilliges Y, Springer A, Disney MD, Ratner DD, Bouchier C, Seeberger PH, de Marsac NT, Dittmann E. A mannan binding lectin is involved in cell-cell attachment in a toxic strain of Microcystis aeruginosa. Mol Microbiol. 2006; 59:893–906. [PubMed: 16420359]
59. Percudani R, Montanini B, Ottonello S. The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. Proteins. 2005; 60:670–678. [PubMed: 16003744]
60. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A. The Pfam protein families database. Nucleic Acids Res. 2010; 38:D211–222. [PubMed: 19920124]
61. Koharudin LMI, Viscomi AR, Jee J-G, Ottonello S, Gronenborn AM. The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. Structure. 2008; 16:570–584. [PubMed: 18400178]
62. Shahzad-ul-Hussan S, Gustchina E, Ghirlando R, Clore GM, Bewley CA. Solution structure of the monovalent lectin microvirin in complex with Manα(1–2)Man provides a basis for anti-HIV activity with low toxicity. J Biol Chem. 2011; 286:20788–20796. [PubMed: 21471192]
63. Koharudin LMI, Viscomi AR, Montanini B, Kershaw MJ, Talbot NJ, Ottonello S, Gronenborn AM. Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus Magnaporthe oryzae. Structure. 2011; 19:662–674. [PubMed: 21565701]
64. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990; 18:6097–6100. [PubMed: 2172928]
65. Neher E. How frequent are correlated changes in families of protein sequences? Proc Natl Acad Sci USA. 1994; 91:98–102. [PubMed: 8278414]
66. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science. 1999; 286:295–299. [PubMed: 10514373]
67. Mori T, Shoemaker RH, Gulakowski RJ, Krepps BL, McMahon JB, Gustafson KR, Pannell LK, Boyd MR. Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. Biochem Biophys Res Commun. 1997; 238:218–222. [PubMed: 9299482]
68. Mori T, Gustafson KR, Pannell LK, Shoemaker RH, Wu L, McMahon JB, Boyd MR. Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. Protein Expr Purif. 1998; 12:151–158. [PubMed: 9518455]
69. Shenoy SR, Barrientos LG, Ratner DM, O'Keefe BR, Seeberger PH, Gronenborn AM, Boyd MR. Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. Chem Biol. 2002; 9:1109–1118. [PubMed: 12401495]
70. Barrientos LG, Louis JM, Hung J, Smith TH, O'Keefe BR, Gardella RS, Mori T, Boyd MR, Gronenborn AM. Design and initial characterization of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N. Proteins. 2002; 46:153–160. [PubMed: 11807943]

71. Han Z, Xiong C, Mori T, Boyd MR. Discovery of a stable dimeric mutant of cyanovirin-N (CV-N) from a T7 phage-displayed CV-N mutant library. Biochem Biophys Res Commun. 2002; 292:1036–1043. [PubMed: 11944919]
72. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem. 1991; 60:795–825. [PubMed: 1883209]
73. Henry E, Hofrichter J. Singular value decomposition – application to analysis of experimental data. Meth Enzymol. 1992; 210:129–192.
74. Segel DJ, Fink AL, Hodgson KO, Doniach S. Protein denaturation: a small-angle x-ray scattering study of the ensemble of unfolded states of cytochrome *c*. Biochemistry. 1998; 37:12443–12451. [PubMed: 9730816]
75. Latypov RF, Cheng H, Roder NA, Zhang J, Roder H. Structural characterization of an equilibrium unfolding intermediate in cytochrome *c*. J Mol Biol. 2006; 357:1009–1025. [PubMed: 16473367]
76. Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, Razi N, Stevens DJ, Skehel JJ, van Die I, Burton DR, Wilson IA, Cummings R, Bovin N, Wong C-H, Paulson JC. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc Natl Acad Sci USA. 2004; 101:17033–17038. [PubMed: 15563589]
77. Fujimoto YK, Terbush RN, Patsalo V, Green DF. Computational models explain the oligosaccharide specificity of cyanovirin-N. Protein Sci. 2008; 17:2008–2014. [PubMed: 18809850]
78. Waldburger CD, Schildbach JF, Sauer RT. Are buried salt bridges important for protein stability and conformational specificity? Nat Struct Biol. 1995; 2:122–128. [PubMed: 7749916]
79. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. Science. 1998; 282:1462–1467. [PubMed: 9822371]
80. Mandell DJ, Kortemme T. Computer-aided design of functional protein interactions. Nat Chem Biol. 2009; 5:797–807. [PubMed: 19841629]
81. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature. 1986; 319:199–203. [PubMed: 3945310]
82. Barrientos LG, Lasala F, Delgado R, Sanchez A, Gronenborn AM. Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity. Structure. 2004; 12:1799–1807. [PubMed: 15458629]
83. Hazes B, Dijkstra BW. Model building of disulfide bonds in proteins with known three-dimensional structure. Protein Eng. 1988; 2:119–125. [PubMed: 3244694]
84. MacKerel, A., Jr.; Brooks, C., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The energy function and its parameterization with an overview of the program. In: Schleyer, v. R., et al., editors. The Encyclopedia of Computational Chemistry. Vol. Vol. 1. John Wiley & Sons; Chichester: 1998. p. 271-277.P.
85. Cornell W, Cieplak P, Bayly C, Gould I, Merz K, Ferguson D, Spellmeyer D, Fox T, Caldwell J, Kollman P. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 1996; 118:5179–5197.
86. Laurents DV, Baldwin RL. Characterization of the unfolding pathway of hen egg white lysozyme. Biochemistry. 1997; 36:1496–1504. [PubMed: 9063898]
87. Munson M, O'Brien R, Sturtevant JM, Regan L. Redesigning the hydrophobic core of a four-helix-bundle protein. Protein Sci. 1994; 3:2015–2022. [PubMed: 7535612]
88. Smith CK, Withka JM, Regan L. A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. Biochemistry. 1994; 33:5510–5517. [PubMed: 8180173]
89. Chaudhuri TK, Horii K, Yoda T, Arai M, Nagata S, Terada TP, Uchiyama H, Ikura T, Tsumoto K, Kataoka H, Matsushima M, Kuwajima K, Kumagai I. E ect of the extra N-terminal methionine residue on the stability and folding of recombinant α-lactalbumin expressed in Escherichia coli. J Mol Biol. 1999; 285:1179–1194. [PubMed: 9887272]
90. Chaudhuri TK, Arai M, Terada TP, Ikura T, Kuwajima K. Equilibrium and kinetic studies on folding of the authentic and recombinant forms of human α-lactalbumin by circular dichroism spectroscopy. Biochemistry. 2000; 39:15643–15651. [PubMed: 11112553]
91. Hubbard SJ, Thornton JM. NACCESS. 1993

92. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14:1188–1190. [PubMed: 15173120]

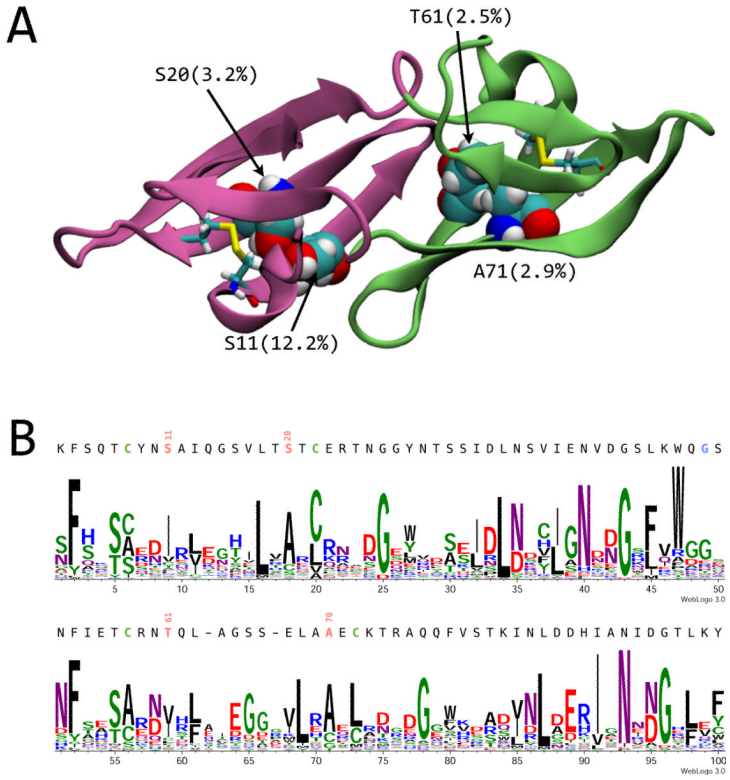

Figure 1.

A. Ribbon representation of Cyanovirin-N based on PDB 1IIY (51). Domain A is colored purple, and Domain B green. Disulfide bonds are shown in licorice representation, and Ser11, Ser20, Thr61, Ala71 are shown in space-filling representation. Fractional side-chain solvent accessibility of these residues is indicated in parentheses. Solvent accessibility was computed with Naccess (91) **B**. Sequence logo representation of CVN family rendered with WebLogo 3.0 (92); the height of the letter stack at each position is proportional to sequence conservation, while the height of the letters within a stack is scaled to relative amino-acid frequency. The numbers below the logo are position indices in the sequence alignment. The first and second repeats are aligned to emphasize the tandemly-repeated nature of CVN. The sequence of P51G mutant of CVN is shown above.

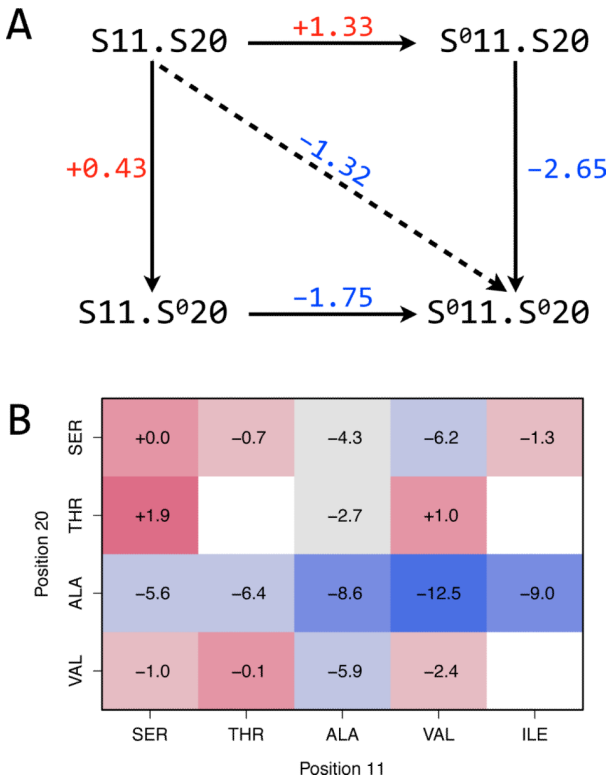

Figure 2.

A. Calculated effect of replacing the side chains of Ser11 and Ser20 with hydrophobic isosteres. All values shown are $\Delta\Delta G°_{mut}$, in kcal/mol, with details provided in Table 1. **B**. Heatmap summary of protein design calculations on Domain A of CVN. Amino acid substitutions as positions 11 and 20, as well as ΔΔG° values for each sequence are shown. The ΔΔG° values are referenced to wild type (Ser11.Ser20). Sequences corresponding to the entries colored in white were determined as “dead ends” in the design procedure.

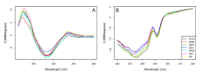

Figure 3.
Far-(**A**) and near-UV (**B**) circular dichroism spectra of CVN mutants confirm mutants possess identical secondary and tertiary structure. Far-UV spectra were recorded at 10 μm protein concentration using a 1 mm path length cuvette. Near-UV spectra were recorded at 100 μm in a 10 μm path length cuvette. Any discrepancies of signal intensity are attributed to errors in determining protein concentration.

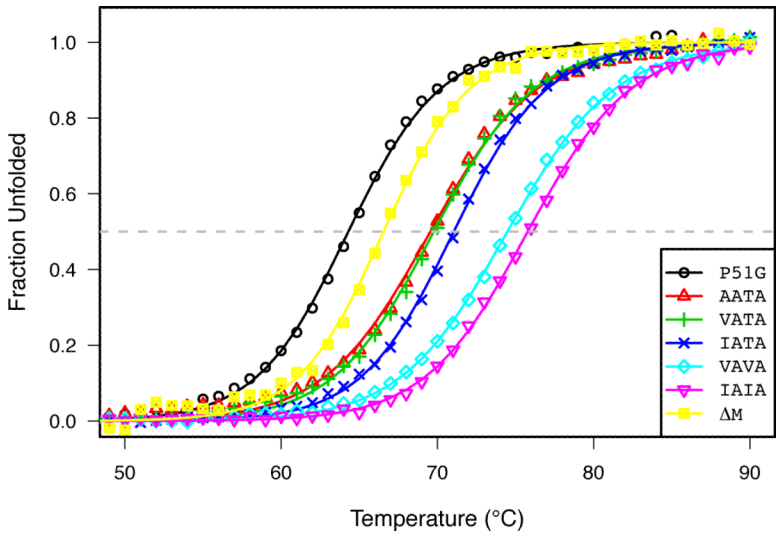

Figure 4.

Thermal denaturation of CVN mutants followed by CD spectroscopy at 200 nm. The raw data were converted to fraction unfolded; only the transition region is shown for clarity. The solid lines are a nonlinear least-squares fit to the data.

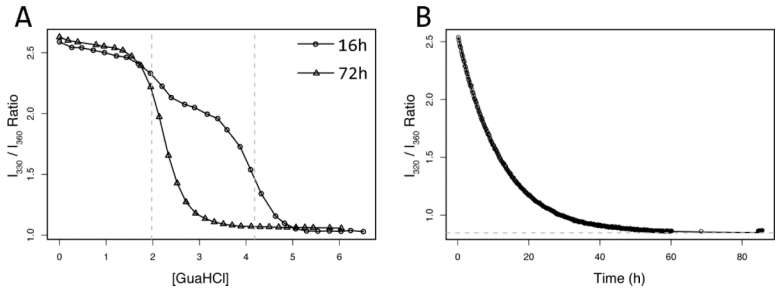

Figure 5.

A. The denaturation profile of P51G monitored by Trp fluorescence emission, after 16- and 72-hour incubations at room temperature in the presence of GuaHCl. The 16-hour profile is biphasic, with dashed grey lines showing the apparent denaturation midpoints of the two phases when they are fitted separately. The numerical derivative of the 16-hour denaturation profile showed two inflection points, at the apparent midpoints of the two phases (not shown). In contrast, 72-hour denaturation profile shows a single transition. These data suggests the presence of a kinetic intermediate in CVN denaturation by GuaHCl. **B**. Fluorescence-monitored denaturation kinetics of P51G after addition of 3 M GuaHCl. A solid line is the fit of the data to a biexponential decay. The apparent half-life of unfolding is 14 h. Note that the data could not be adequately fit by a single or double exponential, as judged by the randomness of the residuals.

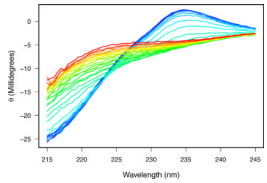

Figure 6.

Far-UV CD wavelength scans of P51G taken at increasing concentrations of GuaHCl after 72 h equilibration at 25 ° C, pH 6.0. The spectra are from 0 M (blue) to 6 M (red) of denaturant. The absence of an isodichroic point at equilibrium demonstrates that the denaturation of CVN by GuaHCl is not strictly two-state.

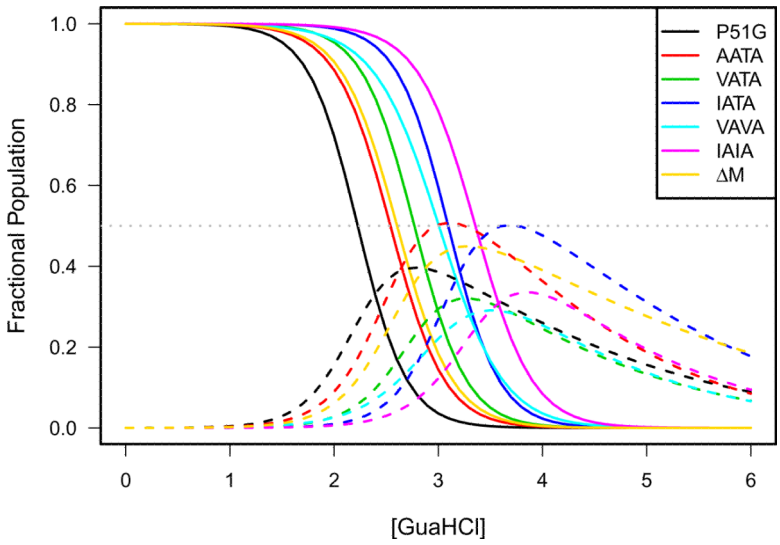

Figure 7.

Fractional populations of CVN variants derived from decomposition of circular dichroism spectra acquired at increasing concentrations of GuaHCl. Solid lines show the fractional populations of the native state; matching dashed lines show the fractional populations of the equilibrium intermediate state for each variant. The dotted grey line indicates 50% population, and is provided for visual reference. The details of the thermodynamic models used to generate these populations are described in the text.

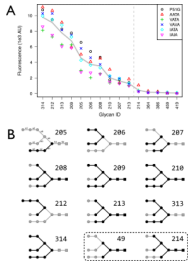

Figure 8.

A. Glycans bound by CVN mutants, ranked by decreasing average fluorescence amongst designed variants. The raw fluorescence intensity of binding to microarray is shown in points. Average fluorescence of all mutants is shown as a thick grey line. Glycan ID corresponds to glycans found on Version 4.1 Glycan Microarray at the Consortium for Functional Glycomics. The dashed vertical line indicates the cutoff for the 10 carbohydrates determined to bind CVN variants. No additional carbohydrates gave rise to signal distinguishable from the experimental noise. **B**. Glycans bound by CVN variants on carbohydrate microarray (shaded in black). For each glycan, the ID is given, as well as the substructure of Man_9 that the glycan corresponds to; Man_9 is shown as background in light gray. Mannose residues are represented by • and GlcNAc residues by ■. The dashed box contains two examples of Man_9 core carbohydrates which were *not* bound by any of the CVN variants discussed here; these are included to illustrate the specificity of binding to Manα(1→2)Man disaccharides.

Table 1

Calculated electrostatic contribution of CVN amino-acid side chains to protein stability.

Molecular group	$\Delta\Delta G^{\cdot}_{solv}$	$\Delta\Delta G^{\cdot}_{inter}$	$\Delta\Delta G^{\cdot}_{mut}$	$\Delta\Delta G^{\cdot}_{inter.partner}$
Ser11	+2.92 ± 0.02	−4.25 ± 0.09	−1.33 ± 0.08	−3.08 ± 0.06
Ser20	+2.96 ± 0.02	−3.39 ± 0.08	−0.43 ± 0.08	−3.08 ± 0.06
Ser11.Ser20	+5.88 ± 0.03	−4.56 ± 0.09	+1.32 ± 0.09	
Thr61	+2.30 ± 0.02	−2.51 ± 0.08	−0.21 ± 0.08	+0.03 ± 0.01
Ala71	+0.03 ± 0.01	+0.20 ± 0.01	+0.23 ± 0.01	+0.03 ± 0.01
Thr61.Ala71	+2.33 ± 0.02	−2.34 ± 0.08	−0.01 ± 0.08	

Values were calculated for 201 molecular dynamics snapshots and averaged. The standard errors of the mean are provided as a measure of uncertainty. All free energy values are in kcal/mol.

Table 2

CVN variants discussed in this work.

	51	11	20	61	71
wild-type	Pro	Ser	Ser	Thr	Ala
P51G (SSTA)	Gly	Ser	Ser	Thr	Ala
AATA	Gly	Ala	Ala	Thr	Ala
VATA	Gly	Val	Ala	Thr	Ala
IATA	Gly	le	Ala	Thr	Ala
VAVA	Gly	Val	Ala	Val	Ala
IAIA	Gly	le	Ala	le	Ala

Numbered columns show the amino acid identity at that position.

Table 3

Apparent equilibrium denaturation parameters.

Variant	T_m (°C)	C_m^{FI}(M)	m^{FI}	C_m^{CD}(M)	m^{CD}
P51G	64.3 ± 0.1	2.3 ± 0.02	2.6 ± 0.05	2.2 ± 0.02	2.5 ± 0.09
AATA	69.6 ± 0.3	2.6 ± 0.01	2.5 ± 0.05	2.5 ± 0.02	2.4 ± 0.14
VATA	69.8 ± 0.2	2.8 ± 0.01	2.6 ± 0.06	2.8 ± 0.02	2.6 ± 0.21
IATA	71.1 ± 0.1	3.1 ± 0.01	2.3 ± 0.05	3.1 ± 0.02	2.0 ± 0.11
VAVA	74.5 ± 0.2	3.0 ± 0.01	2.1 ± 0.06	2.9 ± 0.03	1.7 ± 0.12
IAIA	75.8 ± 0.3	3.5 ± 0.01	2.1 ± 0.06	3.3 ± 0.02	1.7 ± 0.09
ΔM	66.6 ± 0.2	2.6 ± 0.02	1.9 ± 0.11	2.5 ± 0.02	2.0 ± 0.10

The parameters are derived from direct fitting of circular dichroism and fluorescence data. The data were fit to a simple two-state model, as discussed in the text. The provided measures of uncertainty are standard errors of the fit.

Biochemistry. Author manuscript; available in PMC 2012 December 13.

Table 4

Equilibrium denaturation parameters for CVN variants.

Variant	$\Delta G^{\circ}_{N\rightarrow I}$	$m_{N\rightarrow I}$	$\Delta G^{\circ}_{N\rightarrow D}$	$m_{N\rightarrow D}$	$\Delta\Delta G^{\circ}_{N\rightarrow I}$(P51G)	$\Delta\Delta G^{\circ}_{I\rightarrow D}$(P51G)
P51G	5.55	2.32	6.45	2.70		
AATA	5.56	2.10	7.39	2.64	0.01	0.94
VATA	6.34	2.12	7.48	2.57	0.79	1.03
IATA	7.28	2.27	9.01	2.71	1.73	2.56
VAVA	5.44	1.65	6.75	2.13	−0.11	0.30
IAIA	6.83	1.91	8.49	2.41	1.28	2.04
ΔM	5.80	2.10	6.78	2.41	0.25	0.33

The parameters derived from global decomposition of circular dichroism spectra taken as a function of denaturant concentration. The data were fit to a three-state model, as discussed in the text. All free energy values are in kcal/mol.

Table 5

Predicted thermodynamic stabilization based on buried nonpolar surface area.

Variant	$SASA_{np}$ (Å^2)	SASA (Å^2)	$\Delta\Delta G^{\circ}_{\blacklozenge}$ (kcal mol^{-1})
P51G	238	—	0.00
AATA	283	45	0.72
VATA	328	90	1.44
IATA	351	113	1.82
VAVA	367	129	2.07
IAIA	415	177	2.83

Nonpolar surface area was calculated for the side chains at positions 11, 20, 61, and 71 using standard parameters provided with NACCESS. ΔSASA is the estimated change in nonpolar surface area upon mutation, relative to P51G. The expected change in the transfer free energy ($\Delta\Delta G^{\circ}$) was calculated according to Eisenberg (81).