

Solution structure of cyanovirin-N, a potent HIV-inactivating protein

Carole A. Bewley¹, Kirk R. Gustafson², Michael R. Boyd², David G. Covell³, Ad Bax¹, G. Marius Clore¹ and Angela M. Gronenborn¹

The solution structure of cyanovirin-N, a potent 11,000 *M_r* HIV-inactivating protein that binds with high affinity and specificity to the HIV surface envelope protein gp120, has been solved by nuclear magnetic resonance spectroscopy, including extensive use of dipolar couplings which provide *a priori* long range structural information. Cyanovirin-N is an elongated, largely β -sheet protein that displays internal two-fold pseudosymmetry. The two sequence repeats (residues 1–50 and 51–101) share 32% sequence identity and superimpose with a backbone atomic root-mean-square difference of 1.3 Å. The two repeats, however, do not form separate domains since the overall fold is dependent on numerous contacts between them. Rather, two symmetrically related domains are formed by strand exchange between the two repeats. Analysis of surface hydrophobic clusters suggests the location of potential binding sites for protein–protein interactions.

The initial events that lead to HIV infection include binding of the virus to the host cell, activation of the virus, and ultimately virus–cell or cell–cell fusion¹. During the first step of HIV infection, the viral surface envelope glycoprotein gp120 interacts with the CD4 receptor of the host cell, upon which gp120 undergoes a conformational change² sufficient to accommodate a subsequent interaction between gp120 and a member of the α and β chemokine receptor families, now commonly referred to as coreceptors^{2,3}. Concurrently, gp41 dissociates from gp120, associates with the target membrane and mediates fusion. In this paper we describe the three-dimensional solution structure of a newly discovered cyanobacterial protein, named cyanovirin-N, which is a highly potent inhibitor of diverse laboratory adapted strains and clinical isolates of HIV-1, as well as HIV-2 and SIV⁴. The antiviral activity of cyanovirin-N is mediated, at least in part, through high affinity binding to gp120^{4,5}. Cyanovirin-N is currently under joint NCI/NIH investigation as a broad-spectrum virucidal and therapeutic agent against HIV.

Cyanovirin-N was originally isolated from an aqueous extract of a cultured cyanobacterium, *Nostoc ellipsosporum*⁴, and was identified in a screening effort aimed at the discovery of new sources of HIV inhibitors⁶. The primary sequence and disulfide bonding pattern were determined by conventional biochemical techniques^{4,7}, and a synthetic gene was constructed for over-expression of the protein^{4,8}. Analysis of the primary sequence of cyanovirin-N revealed the presence of two internal repeats of 50 and 51 amino acids that show strong sequence similarity to one another, and equivalent positions of the disulfide bonds (Fig. 1)⁷. Cyanovirin-N is extremely resistant to physico-chemical degradation and can withstand treatment with denaturants, detergents, organic solvents such as acetonitrile or methanol, multiple freeze-thaw cycles, and heat (up to 100 °C) with no subsequent loss of antiviral activity⁴. The primary sequence of cyanovirin-N shares no similarity with other proteins thus far deposited in public protein data bases.

Structure determination

The solution structure of cyanovirin-N was determined using double and triple resonance multidimensional heteronuclear NMR spectroscopy, making use of uniformly ¹⁵N- and ¹⁵N/¹³C-labeled protein^{9–11}. The final ensemble of structures was calculated by simulated annealing¹² on the basis of 2,509 experimental NMR restraints, including 334 residual dipolar couplings (¹D_{NH}, ¹D_{CH}, ¹D_{C α C'}}, ¹D_{NC'}} and ²D_{HNC'}}). The latter are a function of the orientation of interatomic vectors relative to the molecular alignment tensor, and hence provide qualitatively different structural information from that afforded by other NMR observables, such as NOEs, coupling constants and chemical shifts, which are reliant on close spatial proximity of atoms¹³. A summary of the structural statistics is provided in Table 1, and a superposition of the final ensemble of 40 simulated annealing structures is shown in Fig. 2a. It should be noted that the inclusion of the residual dipolar couplings in the structure refinement increases the coordinate precision from 0.3 Å to 0.15 Å for the backbone atoms and from 0.54 Å to 0.45 Å for all heavy atoms. The atomic r.m.s. shift in the mean coordinates resulting from the inclusion of the dipolar couplings is 0.66 Å for the backbone atoms and 0.81 Å for all heavy atoms.

Description of the structure

Cyanovirin-N has the shape of an elongated prolate ellipsoid, ~55 Å in length with a maximum width of ~25 Å. The secondary structure elements comprise 10 β -strands and four short ₃₁₀-helical turns. The overall topology can be described as follows (Figs 1a, 2b). In the first sequence repeat (residues 1–50) a four-residue helical turn (residues 3–6) precedes a three-stranded antiparallel β -sheet (β 1, β 2 and β 3 from residues 7–14, 16–24 and 28–36 respectively). β -strand 1 is characterized by a wide β -bulge¹⁴ at Asn 10 and Ser 11. A three-residue helical turn (residues 37–39) connects β -strand 3 to a β -hairpin formed by β -strands 4 (residues 39–43) and 5

¹Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA. ²Laboratory of Drug Discovery Research and Development, National Cancer Institute, Frederick, Maryland 21702-1201, USA. ³Frederick Cancer Research and Development Center, National Cancer Institute, Frederick, Maryland 21702-1201, USA.

Correspondence should be addressed to G.M.C. or A.M.G. email: clore@speck.niddk.nih.gov or gronenborn@vger.niddk.nih.gov

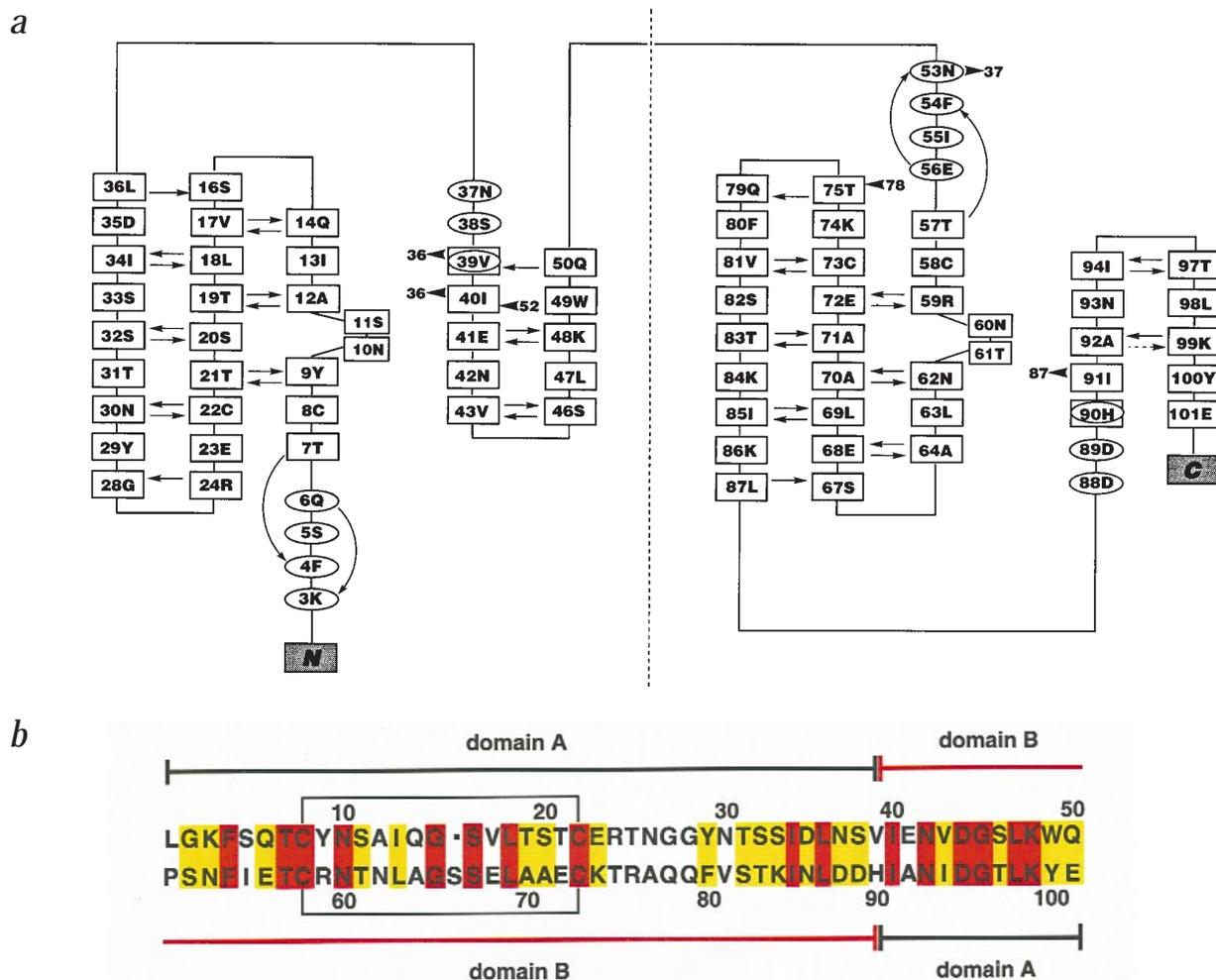


Fig. 1 Secondary structure elements and structure alignment of cyanovirin-N. **a**, Secondary structure and global topology of cyanovirin. Boxed residues denote β -strands, circled residues 3_{10} -helical turns, solid arrows indicate backbone hydrogen bonds from amide protons to carbonyl oxygens, the dotted arrow indicates hydrogen bonds bridged by a water molecule, and offset boxes denote the positions of β -bulges. The dotted line demarcates the first and second sequence repeats. **b**, Structure alignment of the first and second sequence repeats. Yellow boxes denote conserved residues, red boxes identical residues; black and red lines demarcate domains A and B respectively.

(residues 46–50). The β -hairpin is directed away from the three-stranded antiparallel β -sheet at an angle of $\sim 140^\circ$ (Figs 2b, 3a). This topology is exactly repeated in the second sequence repeat (residues 51–101) wherein the second triple-stranded antiparallel β -sheet is formed by β -strands 6 (residues 57–64 with a wide β -bulge at Asn 60 and Thr 61), 7 (residues 67–75) and 8 (residues 79–87), and the second β -hairpin by β -strands 9 (residues 91–94) and 10 (residues 97–100). The first and second sequence repeats are oriented opposite to one another with respect to the pseudosymmetric two-fold axis, which is directed into the plane of the paper in the view shown in Fig. 2b, such that β -strands 1 and 6 lie adjacent and are oriented antiparallel to one another. This arrangement places each of the β -hairpins on top of the triple stranded β -sheet of the other half of the molecule with an angle of $\sim 40^\circ$ between the β -hairpin and the central strand of the underlying β -sheet.

Although cyanovirin-N can be clearly divided into two sequential sequence repeats, the individual repeats do not form separate domains (Fig. 3a,b). Rather, the overall fold of

the molecule depends on numerous contacts between the two repeats. Indeed, the interaction between the two repeats of cyanovirin-N buries 3,085 \AA^2 of accessible surface area (1,551 \AA^2 for the first repeat and 1,534 \AA^2 for the second). In addition, there are several electrostatic interactions between the first and second repeats which include hydrogen bonds between Gly 15(NH) and Thr 61(O γ), Asn 37(N δ H $_2$) and Asn 53(O δ), Asp 44(O δ) and both Arg 76(NH) and Arg 76(N ϵ), and Leu 47(NH) and Thr 83(O γ).

The overall structure of cyanovirin-N, however, can be divided into two symmetrically related domains, A and B, formed by strand exchange between the two sequence repeats (Fig. 3c). Domain A contains the N- and C-termini and comprises residues 1–39 and 90–101; domain B extends from residues 39 to 90. Thus, domains A and B correspond to the 'top' and 'bottom' halves of the molecule in the views shown in Fig. 3a,c. Each domain contains the triple-stranded antiparallel β -sheet of one repeat and the β -hairpin of the other repeat, and the two domains are joined together by helical turns 2 (residues 37–39) and 4 (residues 88–90) (Fig. 3c).

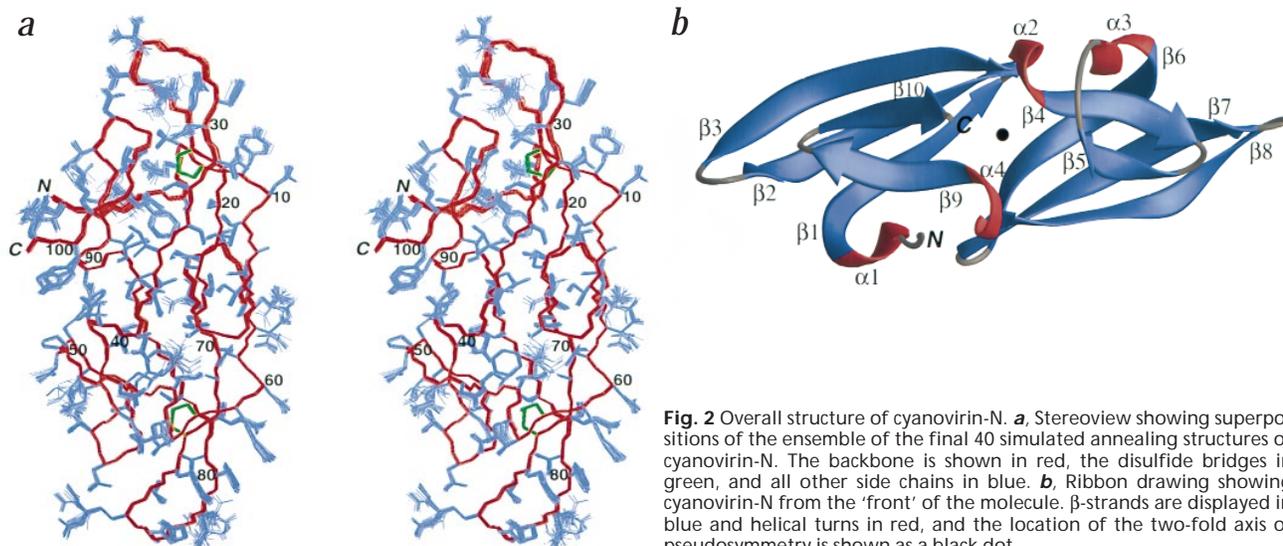


Fig. 2 Overall structure of cyanovirin-N. **a**, Stereoview showing superpositions of the ensemble of the final 40 simulated annealing structures of cyanovirin-N. The backbone is shown in red, the disulfide bridges in green, and all other side chains in blue. **b**, Ribbon drawing showing cyanovirin-N from the 'front' of the molecule. β -strands are displayed in blue and helical turns in red, and the location of the two-fold axis of pseudosymmetry is shown as a black dot.

The positioning of the β -hairpins with respect to the underlying triple-stranded β -sheets is determined by numerous hydrophobic interactions. In domain A, Phe 4, Leu 18, Ile 34 and Leu 36 located in the β -sheet of the first repeat interact with Leu 87, Ile 91 and Leu 98 which are part of the β -hairpin of the second repeat (Fig. 4a). Note that leucines 36 and 87 are located close to the two-fold axis and assume inverted positions in the packing of domain B (Fig. 4b). The disulfide bridge between Cys 8 and Cys 22 has a right-handed hook conformation, and the $S\gamma$ atoms are in van der Waals contact with side chains of Phe 4, Asn 93 and Leu 98. A hydrogen bond between the $N\delta H_2$ group of Asn 93 and the $O\gamma$ atom of Thr 7 bridges β -strands 9 and 1. Finally, a tightly bound water molecule, identified by the observation of ROEs from water to the amide protons of Ala 92, Lys 99 and Glu 101, serves to bridge hydrogen bonds between the backbone carbonyls of His 90 and Lys 99 and the backbone amide of Ala 92.

Fig. 4b shows domain B viewed in the same orientation as domain A displayed in Fig. 4a. A nearly identical set of interactions is apparent: namely, the equivalently positioned side chains of Phe 54, Leu 69, Ile 85 and Leu 87 of the second triple-stranded β -sheet pack against Leu 36, Ile 40 and Leu 47 of the first β -hairpin. Note that the side chain conformation of Ile 85 with χ_1/χ_2 angles in the t/t rotamers differs from the equivalent Ile 34 which has χ_1/χ_2 angles in the g/t rotamers, presumably due to the presence of the tightly bound water wedged between β -strands 9 and 10. The disulfide bridge between Cys 58 and Cys 73 occupies the same position as that between Cys 8 and Cys 22, with the $S\gamma$ atoms in close contact with Phe 54, Asn 42 and Leu 47; and a hydrogen bond occurs between the side chains of Asn 42 (in β -strand 4) and Thr 57 (in β -strand 6). In addition, the side chains of Thr 61 and Ala 71, which are in van der Waals contact with one another, pack against the aromatic ring of Phe 54 providing an auxiliary element to the core not seen in domain A where two serines (at positions 11 and 20) are substituted for Thr 61 and Ala 71.

At the interface of the two domains (Fig. 4c), a cluster of hydrophobic residues, comprising Val 39, His 90, Trp 49 and Tyr 100, brings together helical turns 2 and 4 and the tips of β -strands 5 and 10 respectively (Fig. 4c). Finally, the interactions between Trp 49 and Asp 89, which include both hydrophobic

contacts and a hydrogen bond between the $N\epsilon H$ atom of Trp 49 and the $O\delta$ atom of Asp 89, contribute to the closing of this gap.

Given the strong internal sequence similarity of the two repeats of cyanovirin-N, the nearly identical structures observed for the two repeats is not surprising. A structural alignment between the first and second repeats show 16 residues to be identical and 19 residues to be conservatively replaced giving rise to an overall 70% similarity (Fig. 1b), and the two repeats can be superimposed with a $C\alpha$ atomic r.m.s. difference of 1.3 Å for all 50 residues (Fig. 3b). Likewise, domains A and B can be superimposed with a $C\alpha$ atomic r.m.s. difference of 1.3 Å (Fig. 3d). A noteworthy feature that results from the internal two-fold pseudosymmetry is the adjacent placement of the N- and C-termini, wherein the side chains of Leu 1 and Glu 101 are in van der Waals contact. Thus Leu 1 and Glu 101 form a noncovalent bridge over β -strand 9 that mimics the loop connecting β -strand 5 and helical turn 3, which similarly crosses over β -strand 4 (Fig. 2b). This close arrangement of the N- and C-termini may help to explain results from mutagenesis studies that showed that removal of three consecutive residues from either the N- or C-terminus reduces the antiviral activity by over two orders of magnitude⁵.

Similarities to other proteins

The function of cyanovirin-N in the cyanobacterium from which it was isolated remains unknown. An automated search of protein and gene sequence data banks¹⁵ failed to return any sequences similar to cyanovirin-N. Similarly, a search of the Brookhaven protein structure database using the program DALI¹⁶ did not reveal any proteins with significant overall structural similarity (Z score > 3) to cyanovirin-N. The two individual domains, however, bear a distant resemblance to the SH3 fold¹⁷ (Fig. 5). The closest match is with the hyperthermophile DNA binding protein Sac7d¹⁸ for which 41 residues can be superimposed onto each domain of cyanovirin-N with a $C\alpha$ atomic r.m.s. difference of ~3 Å. The match comprises the triple stranded antiparallel β -sheet and the two helical turns within each domain. The β -hairpin, however, in the domains of cyanovirin-N, is flipped by ~180° about its long axis, relative to that in Sac7d and other SH3

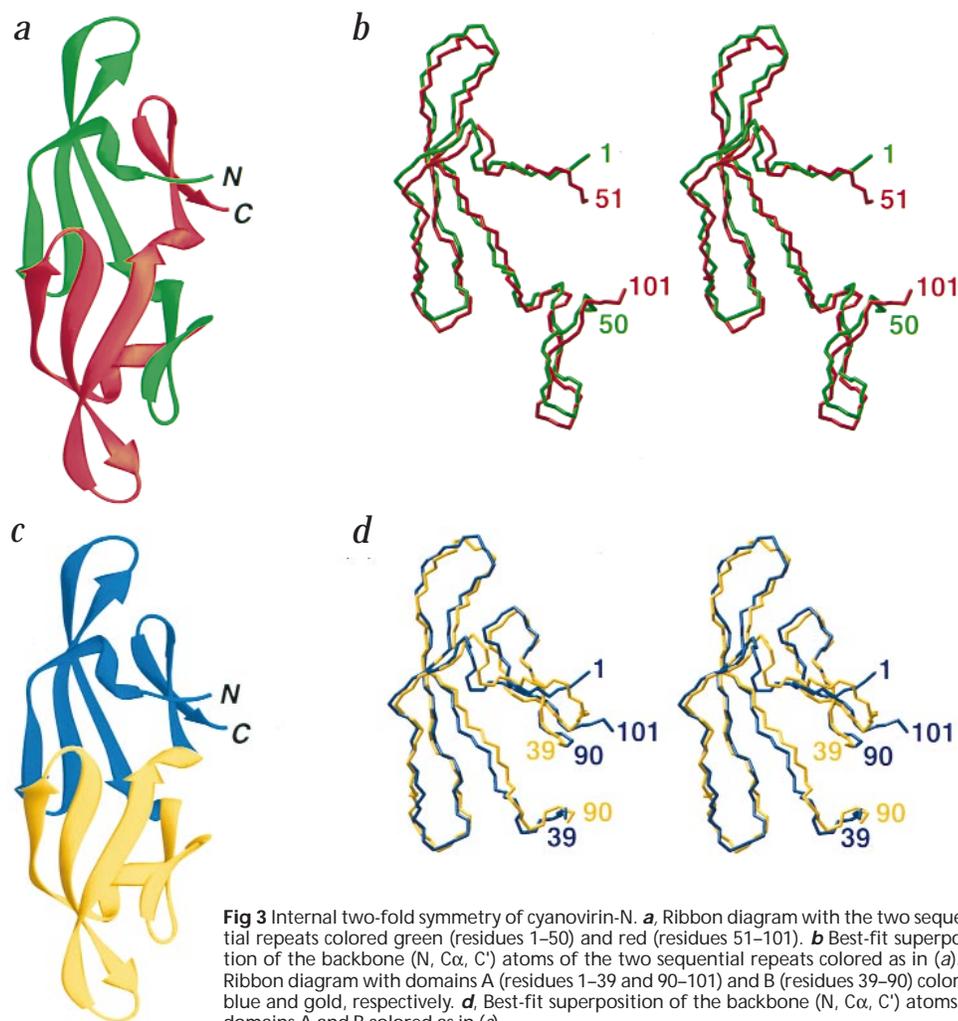


Fig 3 Internal two-fold symmetry of cyanovirin-N. **a**, Ribbon diagram with the two sequential repeats colored green (residues 1–50) and red (residues 51–101). **b** Best-fit superposition of the backbone (N, C α , C') atoms of the two sequential repeats colored as in (a). **c**, Ribbon diagram with domains A (residues 1–39 and 90–101) and B (residues 39–90) colored blue and gold, respectively. **d**, Best-fit superposition of the backbone (N, C α , C') atoms of domains A and B colored as in (c).

domains. It should be noted that in SH3 domains of signaling proteins the equivalent structural element to this hairpin is extended into a long, so-called RT loop which forms part of the binding site for polyproline containing peptides¹⁷.

Potential binding surfaces of cyanovirin-N

Several observations in regard to the antiviral activity of cyanovirin-N suggested early on that its ability to inactivate diverse strains of HIV may be a result of interactions between cyanovirin-N and the HIV envelope glycoprotein gp120^{4,5}. First, pretreatment of virus with cyanovirin-N prevented virus-cell fusion and infection of cells, while pretreatment of cells with cyanovirin-N offered no protection. Second, delayed addition experiments revealed that cyanovirin-N had to be added to cells before or shortly after addition of virus to afford maximum antiviral activity. And third, cyanovirin-N inactivated diverse laboratory strains and clinical isolates of HIV-1, HIV-2 and SIV, all of which share similar surface envelope glycoprotein functions. In the case of HIV-1, the clinical isolates included M-tropic, T-tropic and dual-tropic strains, all of which were inhibited at comparable low nanomolar concentrations⁴. Collectively these results suggested that cyanovirin-N inhibits fusion and viral transmission by direct interactions with the virus as opposed to

the target cell and any of its receptors (such as CD4, CXCR4, CCR5). Through a variety of experimental approaches, cyanovirin-N was shown to bind avidly to gp120, including recombinant non-glycosylated gp120⁴. Further, pretreatment of cyanovirin-N with exogenous, virus-free gp120 resulted in a concentration-dependent decrease in antiviral activity⁴. The recombinant cyanovirin-N used in the NMR structural studies had gp120 binding and anti-HIV properties that were indistinguishable from those of cyanovirin-N isolated from its natural source⁴.

Since surface hydrophobicity plays a key role in protein–protein interactions^{19–22}, we have mapped the most hydrophobic surface clusters on cyanovirin-N using the method of Covell and coworkers^{20–22} to predict which regions of cyanovirin-N may be interacting with the viral envelope. Fig. 6 shows two views of a surface representation of cyanovirin-N onto which have been mapped the electrostatic potential (left-hand panels) and the two highest ranking hydrophobic clusters (center panels).

The top ranking surface hydrophobic cluster (Fig. 6a) is located in domain A, comprises Leu 1, Gly 2, Lys 3, Gln 6, Thr 7, Thr 25, Asn 26, Asn 93, Ile 94 and Asp 95, and is centered around the N-terminus and the turns connecting β -strands 2 and 3 and β -strands 9 and 10. This cluster of amino acids forms an extensive and curved ridge that surrounds the cleft between the first triple-stranded β -sheet and the second β -hairpin (Fig. 6a, right). Comparison of the hydrophobic cluster with the electrostatic surface shows that this ridge, as well as the cleft, is predominantly neutral, with the exception of the positive charge from the N ζ H₃ group of Lys 3 and the negative charge from the carboxylate of Asp 95 located at the top right and bottom left corners respectively, of the hydrophobic cluster. The observation that a large portion of this highest ranking hydrophobic cluster consists of the first three N-terminal residues provides a rationale for the finding that a cyanovirin-N mutant lacking the first three N-terminal residues is virtually inactive⁵. It should be noted that the symmetrically equivalent surface region in domain B of cyanovirin-N, which would consist of helical turn 3 and the turns between β -strands 4 and 5 and β -strands 7 and 8 (surface not shown) carries a formal charge of -2 wherein the neutral residues Gln 6, Thr 26 and Ala 92 (not labeled) are substituted by Glu 56, Arg 76 and Glu 42

respectively, and Lys3 is replaced by Asn53, resulting in a large decrease in surface hydrophobicity and a very different charge distribution.

The second highest ranking surface hydrophobic cluster is located in domain B, comprises Ala 64, Gly 65, Ser 66, Glu 68, Ala 70, Lys 84 and Asn 86, spans regions of β -strands 6, 7 and 8 and includes the turn between β -strands 6 and 7 (Fig. 6b). Ala 70, Ala 64, Gly 65 and Ser 66 form a neutral vertical ridge while Lys 84 and Glu 68 present positive and negative charges respectively, at the left edge of the region (Fig. 6b).

It is known that cyanovirin-N does not bind to gp120 at either of the V3 loop or the CD4 binding site⁴ regions of gp120, which have been well characterized. Thus, future structural studies of cyanovirin-N complexed to the relevant, folded domain of gp120 should shed light on the molecular mechanisms of the conformational changes necessary for initiation of the fusion event, and may also provide more structural clues concerning the interactions between gp120 and gp41.

Methods

Expression, purification and sample preparation. Cloning and expression of a synthetic gene for cyanovirin-N has been described elsewhere⁸. Briefly, the appropriate DNA coding sequence was subcloned into the *E. coli* vector pFLAG, followed by transformation of *E. coli* strain BL21. Uniform (>95%) ¹⁵N and ¹³C labeling was obtained by growing the cells in modified minimal medium containing ¹⁵NH₄Cl and/or ¹³C₆-glucose as the sole nitrogen and carbon sources respectively. Cells were grown at 37 °C, and protein expression was induced for three hours with 1 mM isopropyl-D-thiogalactoside. The cells were harvested, resuspended in 10 mM Tris buffer, pH 7.4, 5 mM EDTA, and 5 mM benzimidazole, lysed by passage through a French press, and cleared by centrifugation. The supernatant was applied directly to a preparative Bakerbond C4 wide pore column and eluted in a stepwise manner with water, 2:1 v/v water:methanol, 1:2 v/v water:methanol, and methanol. Fractions containing cyanovirin-N were combined and further purified by reversed-phase HPLC using a C18 column equilibrated with 0.05% trifluoroacetic acid and eluted with a linear gradient of 20% to 40% acetonitrile. Based on amino acid analysis and UV spectroscopy, the extinction coefficient of cyanovirin-N is 9,400 mol⁻¹ cm⁻¹ at 280nm. Samples for NMR contained ~1.4 mM protein at pH 6.1.

NMR spectroscopy. NMR experiments were carried out at 27 °C on Bruker DMX500, DMX600 and DMX750 spectrometers equipped with x,y,z-shielded gradient triple resonance probes. Spectra were processed with the NMRPipe package²³, and analyzed using the programs PIPP and STAPP²⁴. ¹H, ¹⁵N and ¹³C resonance assignments were made from the following 3D through-bond heteronuclear correlation experiments: CBCA(CO)NH, CBCANH, HNCO, HNHA, DIPSI-H(CCO)NH, DIPSI-C(CCO)NH, HCCH-COSY and HCCH-TOCSY⁹⁻¹¹. ³J_{H_Nα}, ³J_{C_αC_γ} (aromatic, methyl and methylene), ³J_{N_Cγ} (aromatic, methyl and methylene), and ³J_{C_αC_β} couplings were measured by quantitative J correlation spectroscopy²⁵. ϕ and ψ backbone torsion angle restraints were derived from ³J_{H_Nα} and ³J_{C_OCO} coupling constants, the three-bond amide deuterium isotope effect on the ¹³C_α shifts (measured from a 3D HCA(CO)N recorded on a protein sample dissolved in 50% H₂O/50% D₂O)²⁶, and the backbone (¹⁵N, NH, ¹³C_α, ¹³C_β, ¹³C', H_α) secondary chemical shifts using the program TALOS (G. Cornilescu, F. Delaglio and A.B., in preparation). The latter comprises a database of residue triplets correlating ϕ/ψ angles (derived from high resolution structures) and their corresponding backbone secondary chemical shifts, and is based on the premise that when a string of three amino acid shows high similarity in secondary shifts and residue type with a string of amino acids in the database, the central residues of the two strings are likely to have similar backbone torsion angles. Side chain torsion angles were derived from NOE/ROE data and three-bond heteronuclear coupling constants. Interproton distance restraints were derived from 3D ¹⁵N (120 ms mixing time) and ¹³C (45 and 120 ms mixing time) separated NOE experiments, and 4D ¹³C/¹⁵N (140 ms mixing time) and ¹³C/¹³C (140 ms

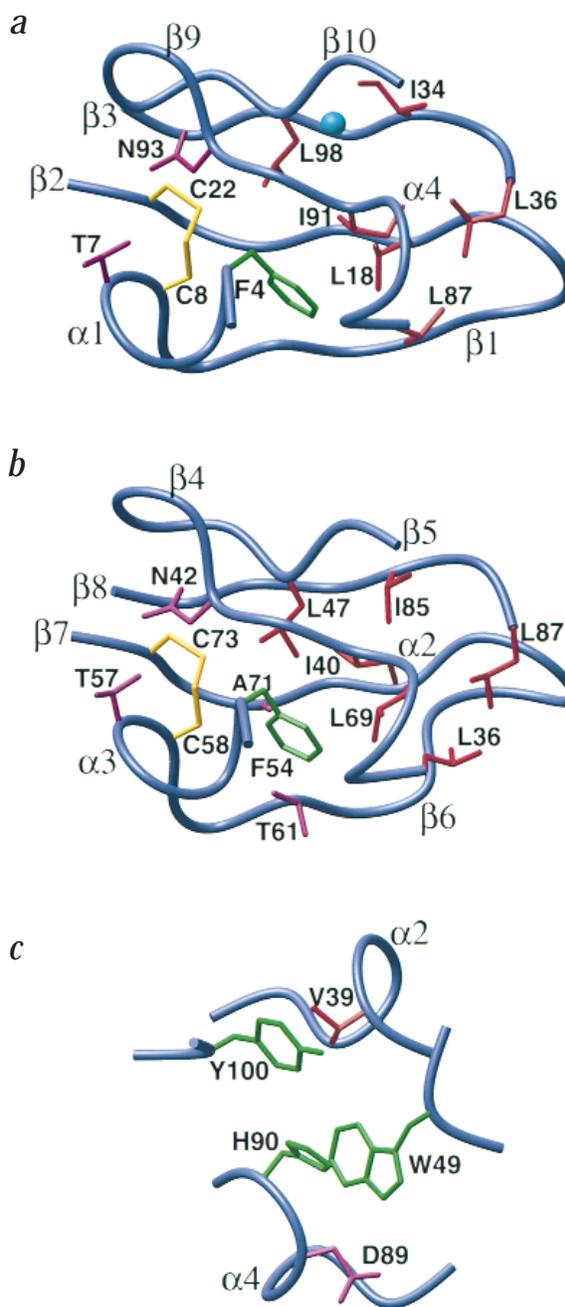


Fig. 4 Side chain contacts forming the core of cyanovirin-N. Views showing the core of **a**, domain A and **b**, domain B. Domain B is shown in the same orientation as domain A. **c**, The interface of the two domains. A α backbone worm is shown in blue, hydrophobic residues in red, aromatic residues in green, disulfide bonds in yellow, and all other residues in magenta. The tightly bound water in domain A which serves to bridge backbone hydrogen bonds between β -strands 9 and 10 is shown as a cyan colored sphere.

mixing time) separated NOE experiments⁹⁻¹¹. Location of bound water was determined by means of a 2D H₂O-ROE-¹H-¹⁵N-HSQC (mixing time 45 ms) spectrum²⁷. ¹D_{NH}, ¹D_{C_αH}, ¹D_{C_αC_γ}, ¹D_{NC}, and ²D_{HNC} residual dipolar couplings were obtained by taking the difference in the corresponding J splittings measured on oriented (in 4% 3:1 DMPC:DHPC at 38 °C) and isotropic (in water) cyanovirin-N²⁸. ¹J_{NH}, ¹J_{C_αH}, ¹J_{C_αC_γ} couplings were obtained from a 2D IPAP {¹⁵N,¹H}-HSQC experiment to generate two spectra containing either the upfield

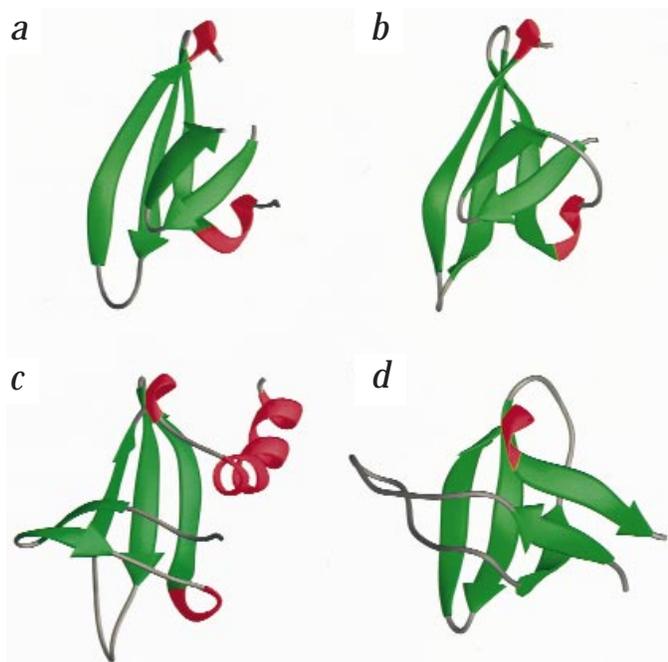


Fig. 5 Comparison of the fold of domains **a**, A and **b**, B of cyanovirin-N with **c**, the hyperthermophile DNA binding protein Sac7d¹⁸ and **d**, the SH3 domain of spectrin⁴⁷

or downfield ¹⁵N doublet component²⁹, a 3D ¹H α -coupled(F_1) HCA(CO)N experiment, and a 2D constant time ¹³C'-coupled/¹H α -decoupled(F_1) ¹H-¹³C HSQC experiment respectively. ¹J_{NC'} and ²J_{HNC'} couplings were obtained from a 2D ¹³C'-coupled/¹³C α -decoupled(F_1) ¹H-¹⁵N HSQC experiment³⁰. The precision of the measured ¹D_{NH}, ¹D_{C α H}, ¹D_{C α C'}}, ¹D_{NC'} and ²D_{HNC'} dipolar couplings was \sim 0.5–1.0 Hz, \sim 1–1.5 Hz, \sim 1.0–1.5 Hz, \sim 0.5–1.0 Hz and \sim 1.0–1.5 Hz respectively.

The measured ¹D_{NH} values ranged from -31 to +20 Hz, and the normalization factors (given by $\gamma_N\gamma_H\langle r_{NH}^{-3}\rangle/\gamma_A\gamma_B\langle r_{AB}^{-3}\rangle$ where γ and r represent gyromagnetic ratios and distances respectively) employed for ¹D_{C α H}, ¹D_{C α C'}}, ¹D_{NC'} and ²D_{HNC'} relative to ¹D_{NH} were 0.48, 5.36, 9.04 and 3.04 respectively. The magnitude of the axial and rhombic components of the alignment tensor **D**^{NH} were obtained by examining the distribution of normalized dipolar couplings³¹ which yielded values of $D_a^{NH} = -17.0$ Hz and $R = 0.17$, where D_a^{NH} is the axial component of the tensor and R is the rhombicity defined as the ratio of the rhombic to axial components of the tensor. This value of D_a^{NH} corresponds to a value of 1.48×10^{-3} for A_a which is the unitless axial component of the molecular alignment tensor **A**. Heteronuclear ¹⁵N-¹H NOEs were measured as described³² and identified only a single residue with a ¹⁵N-¹H NOE less than 0.6, namely Ser 52 which had an NOE value of \sim 0.4.

Structure calculations. Approximate interproton distance restraints, derived from multidimensional NOE spectra, were grouped into four distance ranges, 1.8–2.7 Å (1.8–2.9 Å for NOEs involving NH protons), 1.8–3.3 Å (1.8–3.5 Å for NOEs involving NH protons), 1.8–5.0 and 1.8–6.0 Å, corresponding to strong, medium, weak and very weak NOEs respectively⁹. 0.5 Å was added to the upper bound for distances involving methyl groups to account for the higher apparent intensity of the methyl resonances. Distances involving non-stereospecifically assigned methylene protons, methyl groups, and H δ and H ϵ protons of Tyr and Phe, were represented as a $(\Sigma r^{-6})^{-1/6}$ sum³³. The structures were calculated by simulated annealing^{12,34} using the program CNS³⁵, adapted to incorporate pseudopotentials for three-bond coupling constants³⁶, secondary ¹³C α /¹³C β chemical shifts³⁷, proton chemical shifts^{38,39} and residual dipolar coupling⁴⁰ restraints, and a conformational database potential for the non-bonded contacts derived from very high resolution (1.7 Å or better) X-ray structures^{41,42}. The target function that is minimized during simulated annealing and restrained regularization comprises quadratic harmonic potential terms for covalent geometry, ³J_{HNH α} coupling constant restraints, secondary ¹³C α and ¹³C β chemical shift restraints, ¹H chemical shift restraints, and dipolar coupling restraints; square-well quadratic potentials for the experimental distance and torsion angle restraints; a quartic van der Waals repulsion term and a conformational database potential term for the non-bonded contacts. The latter biases sampling during simulated annealing refinement to conformations that are likely to be energetically possible by effectively limiting the choices of dihedral angles to those that are known to be physically realizable^{41,42}. There were no hydrogen-bonding, electrostatic, or 6–12 Lennard-Jones empirical potential energy terms in the target function.

Structure figures were generated using the programs MOLMOL⁴³, GRASP⁴⁴ and RIBBONS⁴⁵. The secondary struc-

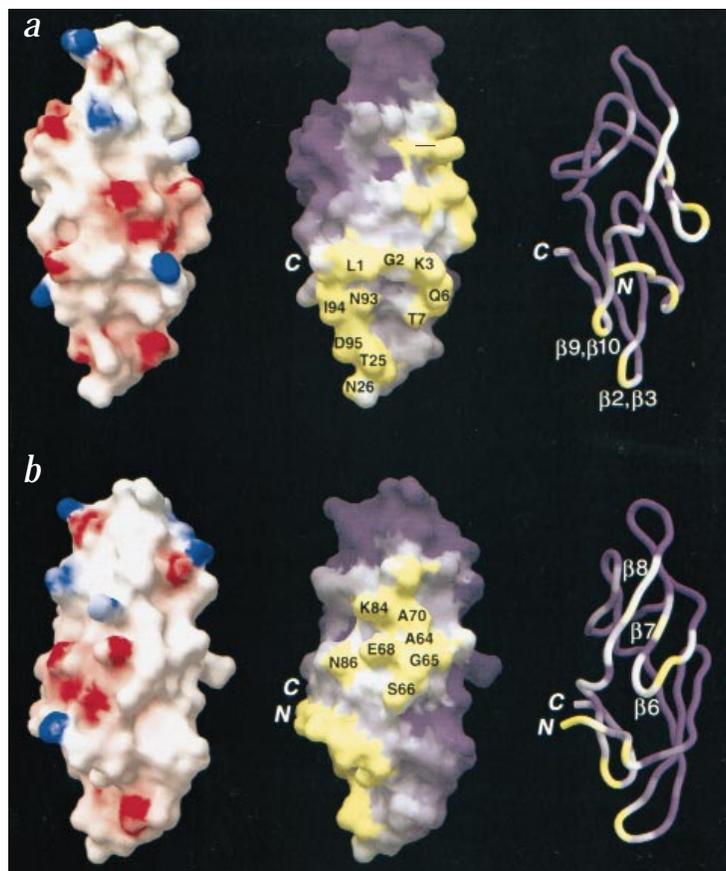


Fig 6 a,b Two views mapping the electrostatic potential (left-hand panels) and the two highest ranking surface hydrophobic clusters (center panels) on the molecular surface of cyanovirin-N. The first (a) and second (b) highest ranking hydrophobic clusters are located in domains A and B respectively. In the left-hand panels, the electrostatic potential is colored from red (negative charge) to blue (positive charge). In the center panels regions of highest hydrophobicity are colored yellow, those of lowest hydrophobicity are colored purple, and the gradient from yellow to white to purple corresponds to decreasing hydrophobicity. Shown in the right-hand panels are the α worm representations in the same orientation as the corresponding surfaces where hydrophobicity has been mapped onto the backbone worm with the same color scheme used for the center panels.

Table 1 Structural statistics¹

	<SA>	($\bar{S}\bar{A}$) _r
R.m.s. deviations from experimental distance restraints (Å) ²		
All (1,241)	0.013 ± 0.001	0.010
interresidue sequential (i - j = 1) (418)	0.008 ± 0.003	0.005
interresidue medium range (1 < i - j ≤ 5) (171)	0.005 ± 0.002	0.007
interresidue long range (1 < i - j ≤ 5) (540)	0.016 ± 0.002	0.014
intraresidue (20)	0.028 ± 0.011	0.016
bound water (8)	0.001 ± 0.004	0.000
H-bonds (84)	0.017 ± 0.006	0.009
R.m.s. deviations from exptl dihedral restraints (°) (334) ²	0.266 ± 0.052	0.175
R.m.s. deviations from ³ J _{H_NC_α} coupling constants (Hz) (81) ²	0.60 ± 0.01	0.61
R.m.s. deviations from secondary ¹³ C shifts (p.p.m.)		
¹³ C _α (82)	0.85 ± 0.01	0.84
¹³ C _β (75)	1.17 ± 0.01	1.16
R.m.s. deviations from ¹ H shifts (p.p.m.) (362)	0.25 ± 0.002	0.25
R.m.s. deviations from residual dipolar couplings (Hz)		
¹ D _{NH} (Hz) (84)	0.50 ± 0.02	0.50
¹ D _{CH} (Hz) (77)	1.12 ± 0.03	1.13
¹ D _{C_αC'} (Hz) (44)	1.26 ± 0.01	1.25
¹ D _{N_C'} (Hz) (66)	0.55 ± 0.01	0.56
² D _{H_NC'} (Hz) (63)	1.25 ± 0.01	1.26
Deviations from idealized covalent geometry		
bonds (Å) (1519)	0.004 ± 0.0007	0.005
angles (°) (2,724)	0.699 ± 0.007	0.757
impropers (°) (775)	0.754 ± 0.019	0.784
Measures of structure quality		
E _L (kcal mol ⁻¹) ³	-434 ± 5	-434
PROCHECK ⁴		
Residues in most favorable region of Ramachandran plot	85.4 ± 0.7	87.0
No. of bad contacts per 100 residues	5.7 ± 1.1	5.0
Coordinate precision (Å) ⁵		
backbone (N, C _α , C', O)	0.15 ± 0.02	
all non-hydrogen atoms	0.45 ± 0.03	

¹The notation of the NMR structures is as follows: <SA> are the final 30 simulated annealing structures; $\bar{S}\bar{A}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best-fitted to each other (residues 1–101); ($\bar{S}\bar{A}$)_r is the restrained regularized mean structure obtained by restrained regularization of the mean structure $\bar{S}\bar{A}$. The number of terms for the various restraints is given in parentheses. The final values for the force constants employed for the various terms in the target function employed for simulated annealing are as follows: 1,000 kcal mol⁻¹ Å⁻² for bond lengths, 500 kcal mol⁻¹ rad⁻² for angles and improper torsions (which serve to maintain planarity and chirality), 4 kcal mol⁻¹ Å⁻⁴ for the quartic van der Waals repulsion term (with the van der Waals radii set to 0.8 times their value used in the CHARMM PARAM19/20 parameters), 30 kcal mol⁻¹ Å⁻² for the experimental distance restraints (interproton distances and hydrogen bonds), 200 kcal mol⁻¹ rad⁻² for the torsion angle restraints, 1 kcal mol⁻¹ Hz⁻² for the ³J_{H_NC_α} coupling constant restraints, 0.5 kcal mol⁻¹ p.p.m.⁻² for the secondary ¹³C chemical shift restraints, 7.5 kcal mol⁻¹ p.p.m.⁻² for the ¹H chemical shift restraints, 1.0 kcal mol⁻¹ Hz⁻² for the ¹D_{NH} dipolar coupling restraints, 1.0, 0.035, 0.050 and 0.108 kcal mol⁻¹ Hz⁻² for the normalized (relative to ¹D_{NH}) ¹D_{C_αH}(NH), ¹D_{C_αC'}(NH), ¹D_{N_C'}(NH) and ¹D_{H_NC'}(NH) dipolar coupling restraints respectively, and 1.0 for the conformational database potential.

²None of the structures exhibited interproton distance violations greater than 0.5 Å, dihedral angle violations greater than 3°, or ³J_{H_NC_α} coupling constant violations greater than 2 Hz. The torsion angle restraints consist of 100 φ, 97 ψ, 80 χ₁, 47 χ₂ and 10 χ₃ angles. Protein backbone hydrogen bonding restraints (two per hydrogen bond, r_{NH-O} = 1.5–2.8 Å, r_{N-O} = 2.4–3.5 Å) were introduced during the final stages of refinement according to standard criteria based on amide H-D exchange experiments, backbone three-bond couplings, ¹³C_α/¹³C_β secondary shifts and NOE data. The distance restraints involving the single bound water comprised three interproton distance restraints between backbone amides and water and five distance restraints for three hydrogen bonds from two backbone carbonyls and one backbone amide to water. Only structurally useful intraresidue NOEs are included in the restraints (that is, involving protons separated by more than three bonds).

³The Lennard-Jones van der Waals energy was calculated with the CHARMM PARAM19/20 parameters and is not included in the target function for simulated annealing or restrained minimization.

⁴The overall quality of the structure was assessed using the program PROCHECK⁴⁴. There were no φ/ψ angles in the disallowed region of the Ramachandran plot. The dihedral angle G-factors for φ/ψ, χ₁/χ₂, χ₁/χ₃/χ₄ are -0.37 ± 0.02, 0.44 ± 0.04, 0.04 ± 0.11 and -0.06 ± 0.13 respectively.

⁵Defined as the average r.m.s. difference (residues 1–101) between the final 40 simulated annealing structures and the mean coordinates.

ture and topology was analyzed using the program PROMOTIF¹⁴. Electrostatic calculations were performed with GRASP⁴⁴. Calculation, ranking and mapping of surface hydrophobic clusters was carried out as described^{20–22}.

Coordinates. The coordinates of the ensemble of 40 simulated annealing structures, the restrained regularized mean structure, and the complete list of experimental NMR restraints and ¹H, ¹⁵N and ¹³C assignments have been deposited in the Brookhaven Protein Data Bank (accession codes 2EZM, 2EZN and 2EZMMR).

Acknowledgments

We thank D. Garrett and F. Delaglio for software support; R. Tschudin and L. Cartner for technical support; L. Pannell for mass spectrometry; M. Caffrey, M. Cai, B. O'Keefe and N. Tjandra for numerous useful discussions. C.A.B. is a recipient of a Cancer Research Institute postdoctoral fellowship. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health to G.M.C., A.M.G and A.B.

Received 1 May, 1998; accepted 5 June, 1998.

1. Freed, E.O. & Martin, M.A. The role of human immunodeficiency virus type I envelope glycoproteins in virus infection. *J. Biol. Chem.* **270**, 23883–23886 (1995).
2. Wilkinson, D. HIV-1: cofactors provide the entry keys. *Curr. Biol.* **6**, 1051–1053 (1996).
3. D'Souza, M. P. & Harden, V. A. Chemokines and HIV-1 second receptors - confluence of two fields generates optimism in AIDS research. *Nature Med.* **2**, 1293–1300 (1996).
4. Boyd, M. R. et al. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob. Agents Chemother.* **41**, 1521–1530 (1997).
5. Mori, T. et al. Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV-inactivating protein. *Biochem. Biophys. Res. Comm.* **238**, 218–222 (1997).
6. Boyd, M.R. In *AIDS, etiology, diagnosis, treatment and prevention*. (DeVita, V. R., Hellman, S. & Rosenberg, S. A., eds) 305–319 (Alan Liss, New York: 1988).
7. Gustafson, K.R. et al. Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-N, an anti-HIV protein from the cyanobacterium *Nostoc ellipsosporum*. *Biochem. Biophys. Res. Comm.* **238**, 223–228 (1997).
8. Mori, T. et al. Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein Exp. Purific.* **12**, 151–158 (1998).
9. Clore, G.M. & Gronenborn, A.M. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390–1399 (1991).
10. Clore, G. M. & Gronenborn, A. M. Determining the structures of larger proteins and protein complexes by NMR. *Trends Biotech.* **16**, 22–34 (1998).
11. Bax, A. & Grzesiek, S. Methodological advances in protein NMR. *Acc. Chem. Res.* **26**, 131–138 (1993).
12. Nilges, M., Gronenborn, A.M., Brünger, A. T. & Clore G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Prot. Engng.* **2**, 27–38 (1988).
13. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. Use of dipolar ^1H - ^{15}N and ^1H - ^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct. Biol.* **4**, 732–738 (1997).
14. Hutchinson, E. G. & Thornton, J. M. PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220 (1996).
15. Altschul, S. F. et al. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
16. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
17. Kuriyan, J. & Cowburn, D. Structures of SH2 and SH3 domains. *Curr. Opin. Struct. Biol.* **3**, 828–837 (1993).
18. Edmondson, S. P., Qiu, L., Schriver, J. W. Solution structure of the DNA-binding domain of Sac7d from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry* **34**, 13289–13304 (1995).
19. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13–20 (1996).
20. Young, L., Jernigan, R. L. & Covell, D. G. A role for surface hydrophobicity in protein-protein recognition. *Prot. Sci.* **3**, 717–729 (1994).
21. Covell, D. G., Smythers, G. W., Gronenborn, A. M. & Clore, G. M. Analysis of hydrophobicity in the α and β chemokine families and its relevance to dimerization. *Prot. Sci.* **3**, 2064–2072 (1994).
22. Villoutreix, B. O., Härdig, Y., Wallqvist, A., Covell, D. G., de Frutos, G. & Dahlback, B. Structural investigation of C4p-binding protein by molecular modeling: localization of putative binding sites. *Prot. Sci.* **in the press** (1998).
23. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
24. Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220 (1991).
25. Bax, A. et al. Measurement of homo- and hetero-nuclear J couplings from quantitative J correlation. *Meth. Enz.* **239**, 79–106 (1994).
26. Ottiger, M. & Bax, A. An empirical correlation between amide deuterium isotope effects on $^{13}\text{C}\alpha$ chemical shifts and protein backbone conformation. *J. Am. Chem. Soc.* **119**, 8070–8075 (1997).
27. Grzesiek, S. & Bax, A. Measurement of amide proton exchange rates and NOE with water in $^{13}\text{C}/^{15}\text{N}$ enriched calcineurin B. *J. Biomol. NMR* **3**, 627–638 (1993).
28. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).
29. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998).
30. Delaglio, F., Torchia, D. A. & Bax, A. Measurement of ^{15}N - ^{13}C J couplings in staphylococcal nuclease. *J. Biomol. NMR* **1**, 439–446 (1991).
31. Clore, G. M., Gronenborn, A. M. & Bax, A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **133**, 216–221 (1998).
32. Grzesiek, S. & Bax, A. The importance of not saturating H_2O in protein NMR: application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594 (1993).
33. Nilges, M. A calculational strategy for the structure determination of symmetric dimers by ^1H -NMR. *Protein Struct. Funct. Genet.* **17**, 297–309 (1993).
34. Clore, G.M. & Gronenborn, A.M. New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5891–5898 (1998).
35. Brünger, A.T. et al. Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr. D* (1998) **In the press**.
36. Garrett, D. S. et al. The impact of direct refinement against three-bond HN-C α H coupling constants on protein structure determination by NMR. *J. Magn. Reson.* **B104**, 99–103 (1994).
37. Kuszewski, J., Qin, J., Gronenborn, A. M. & Clore, G. M. The impact of direct refinement against $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson.* **B106**, 92–96 (1995).
38. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Reson.* **B107**, 293–297 (1995).
39. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. A potential involving multiple proton chemical-shift restraints for non-stereospecifically assigned methyl and methylene protons. *J. Magn. Reson.* **B112**, 79–81 (1996).
40. Clore, G. M., Gronenborn, A. M. & Tjandra, N. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131**, 159–162 (1998).
41. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Prot. Sci.* **5**, 1067–1080 (1996).
42. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Reson.* **125**, 171–177 (1997).
43. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14** 51–5, 29–32 (1996).
44. Nichols, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Protein Struct. Funct. Genet.* **11**, 281–296 (1991).
45. Carson, M. RIBBONS 4.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
47. Musacchio, A., Noble, M., Paupit, R., Wierenga, R. & Saraste, M. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **359**, 851–855 (1992).