Article No. jmbi.1999.2693 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 288, 403–412

JMB

Crystal Structure of Cyanovirin-N, a Potent HIV-inactivating Protein, Shows Unexpected Domain Swapping

Fan Yang[1], Carole A. Bewley[2], John M. Louis[2], Kirk R. Gustafson[3] Michael R. Boyd[3], Angela M. Gronenborn[2], G. Marius Clore[2] and Alexander Wlodawer[1]*

[1]*Macromolecular Structure Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702-1201, USA*

[2]*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda MD 20892-0520, USA*

[3]*Laboratory of Drug Discovery Research and Development DTP, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702-1201, USA*

**Corresponding author*

The crystal structure of cyanovirin-N (CV-N), a protein with potent antiviral activity, was solved at 1.5 Å resolution by molecular replacement using as the search model the solution structure previously determined by NMR. The crystals belong to the space group $P3_221$ with one monomer of CV-N in each asymmetric unit. The primary structure of CV-N contains 101 residues organized in two domains, A (residues 1 to 50) and B (residues 51 to 101), with a high degree of internal sequence and structural similarity. We found that under the conditions of the crystallographic experiments (low pH and 26 % isopropanol), two symmetrically related monomers form a dimer by domain swapping, such that domain A of one monomer interacts with domain B′ of its crystallographic symmetry mate and *vice versa*. Because the two swapped domains are distant from each other, domain swapping does not result in additional intramolecular interactions. Even though one of the protein sample solutions that was used for crystallization clearly contained 100 % monomeric CV-N molecules, as judged by various methods, we were only able to obtain crystals containing domain-swapped dimers. With the exception of the unexpected phenomenon of domain swapping, the crystal structure of CV-N is very similar to the NMR structure, with a root-mean-square deviation of 0.55 Å for the main-chain atoms, the best agreement reported to date for structures solved using both techniques.

Keywords: human immunodeficiency virus; virucides; cyanovirin-N; domain swapping; protein structure

Introduction

Virucides interact directly with virions to reduce or inhibit viral transmission. The highly potent virucidal protein cyanovirin-N (CV-N) irreversibly inactivates diverse strains of human immunodeficiency virus (HIV). CV-N was originally isolated from an extract of the cyanobacterium *Nostoc ellipsosporum* (blue-green algae), after its activity was detected in an anti-HIV screen. However, the physiological function of CV-N in *N. ellipsosporum* remains unknown. Biochemical assays have shown that CV-N binds with a high degree of affinity to the viral surface envelope glycoprotein gp120 (Boyd *et al*., 1997; Mori *et al*., 1997a,b), and that such binding is essential for the antiviral activity of CV-N. Immunoblots did not show significant binding of CV-N to HIV proteins other than gp120 (Boyd *et al*., 1997). The binding of CV-N to soluble gp120 did not interfere with the binding of either soluble CD4 or antibodies directed to the V3 loop or the CD4 binding site, suggesting that CV-N interferes with the interactions of gp120 and other cellular receptors and/or with the conformational changes of gp120 required for viral fusion and entry. Recent studies employing cell-associated gp120 indicated that CV-N could block both CD4-dependent and CD4-independent binding and fusion of HIV-1 with cells (Esser *et al*., 1999).

Abbreviations used: CV-N, cyanovirin N; HIV, human immunodeficiency virus; PBS, phosphate-buffered saline; IL-10, interleukin-10.

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

0022-2836/99/180403–10 $30.00/0

CV-N offers considerable promise as a focus for novel prevention, and possibly also treatment strategies for AIDS, in part due to its broad spectrum of activity against immunodeficiency retroviruses. CV-N can irreversibly inactivate diverse laboratory strains and primary isolates of HIV-1, HIV-2, and simian immunodeficiency virus (SIV; Boyd *et al.*, 1997). Due to the high mutation rate of HIV, drug-resistant strains can evolve rapidly following administration of most anti-HIV drugs. Thus, a drug that could effectively prevent initial viral infection would be more likely to prevent the emergence of drug-resistant virus. Although cell-to-cell fusion and transmission of HIV infection can be irreversibly inhibited *in vitro* at low nanomolar concentrations of CV-N, this protein is not toxic to CEM-SS cells or peripheral blood lymphocytes, even at concentrations as high as 9 μM (Boyd *et al.*, 1997). Finally, CV-N is highly stable and resistant to irreversible denaturation caused by elevated temperature or other denaturation agents (Boyd *et al.*, 1997).

In addition to direct virucidal applications, another potential application of CV-N in AIDS therapy is its successful use in creating a recombinant chimeric toxin molecule in which the translocation and cytotoxic domains of *Pseudomonas* exotoxin A are linked to CV-N (Mori *et al.*, 1997b). In this chimeric molecule, CV-N serves as the targeting moiety that searches for HIV-infected cells expressing gp120. The chimeric molecule shows enhanced cytotoxicity for HIV-infected H9 cells compared with uninfected H9 cells, indicating the feasibility of using CV-N to selectively target and destroy HIV-infected host cells.

CV-N is a single polypeptide chain consisting of 101 amino acid residues, with a calculated molecular mass of 11,013 Da. Based on its primary structure, CV-N was expected to contain two domains with a high degree of similarity in both the primary and tertiary structures (Gustafson *et al.*, 1997). These similarities were indeed observed in the NMR solution structure of CV-N (Bewley *et al.*, 1998). With the exception of this internal sequence similarity, sequence comparison searches showed no significant similarity between CV-N and any other protein. CV-N contains four cysteine residues that form two intramolecular disulfide bonds (Cys8 - Cys22 and Cys58 - Cys73), which are well aligned with each other within their domains except for a single amino acid gap. These disulfide bonds are important for the structural stability and anti-HIV activity of CV-N. Reduction of the bonds and alkylation of the resulting free cysteine residues abolished the anti-HIV activity (Gustafson *et al.*, 1997).

The solution structure of CV-N was solved by NMR using a number of novel restraints, including ^{1}H and ^{13}C shifts, ^{3}J couplings, and residual dipolar couplings (Clore & Gronenborn, 1998). The latter, in contrast to other NMR observations that are dependent on close spatial proximity of atoms (<5 Å), provide *a priori* long-range structural information (Tjandra *et al.*, 1997). Although the resulting ensemble of structures has a high degree of internal consistency (with a backbone precision of 0.2 Å), the impact of refinement against residual dipolar couplings on the resulting coordinate accuracy has not been previously validated by comparison to a subsequently solved high-resolution X-ray structure. Moreover, comparison of the high-resolution X-ray and NMR structures of CV-N provides an opportunity to compare the results of these two principal methods of determining atomic protein structures, this time for high-resolution/high-quality structures, as opposed to lower resolution/quality structures which were the subject of most comparisons in the past (Baldwin *et al.*, 1991; Rozwarski *et al.*, 1994).

Results

Crystals of CV-N belong to the space group $P3_221$, with unit cell parameters of $a = b = 48.07$ Å, $c = 79.08$ Å (Table 1) and with one monomer in each asymmetric unit. The Matthews coefficient and the solvent content are 2.39 Å^3/dalton and 48.2 %, respectively. Using the NMR structure of CV-N as the search model, we readily found a solution with the correct space group enantiomorph, using molecular replacement methods as implemented in the program AMoRe (Navaza, 1994). Several cycles of refinement, followed by manual adjustments of a number of side-chains and by the addition of water molecules, reduced the crystallographic R_{cryst} and R_{free} (Brünger, 1992a) to 19.2 % and 21.1 %, respectively, at 1.5 Å resolution.

The refined crystal structure based directly on the NMR model was well determined, except for residues 50 to 53, which are part of the linker region between the two domains that form a compact molecule in the NMR-determined structure. In that structure, the linker did not show any apparent disorder, as indicated by similar positional deviations of the main-chain atoms of the linker when compared with the other parts of the protein. In the initial crystallographic refinement of CV-N in which a monomeric model was assumed, however, the electron density around the linker was practically non-existent (Figure 1). This result was

Table 1. Summary of data collection and refinement

Space group	$P3_221$
Unit cell (Å)	$a = b = 48.07$, $c = 79.08$
Resolution (Å)	10-1.5
Number of reflections (total/unique)	107,021/17,537
R_{sym} (%) (overall/last shell 1.54-1.5 Å)	5.8/36.8
Completeness (%)	99.9
R_{cryst} (%)	18.4
R_{free} (%)	20.6
RMS deviation of bonds (Å)	0.004
RMS deviation of angles (deg.)	1.24
RMS deviation of dihedrals (deg.)	25.9
RMS deviation of impropers (deg.)	0.563
Average *B*-factor for main-chain atoms (Å^2)	13.56
Number of solvent molecules	207

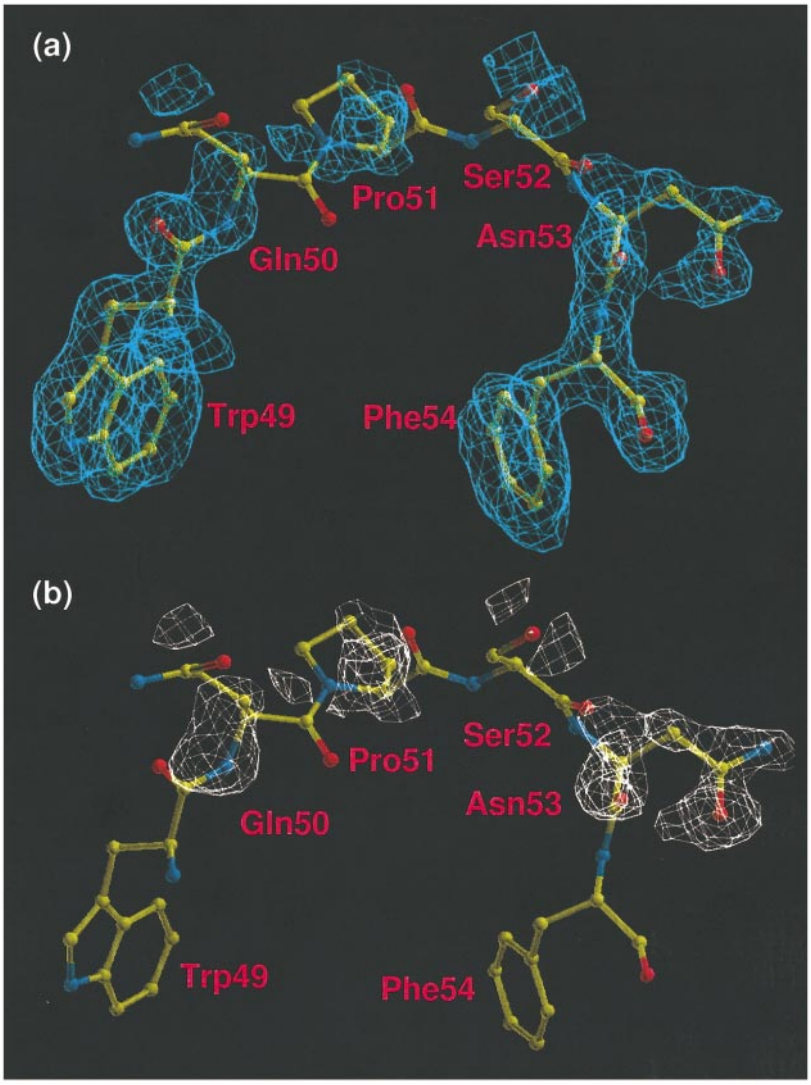

Figure 1. Electron density maps around the linker region of the monomeric crystal structure of CV-N. The maps were calculated using the monomeric crystal structure based directly on the NMR structure. (a) The $2F_o - F_c$ map. (b) The omit $F_o - F_c$ map in which the occupancies of residues Gln50 to Asn53 were set to zero.

also reflected by the average temperature factors of the main-chain atoms of the linker residues and the real-space R_{real} factors (Table 2). While the average temperature factors of Trp49 and Phe54 were 13.5 Å^2 and 16.1 Å^2, respectively, the average temperature factors of residues 50 to 53 were much higher. In addition, in the $2F_o - F_c$ map contoured at the 1 σ level, we observed continuous electron densities that connected residue 49 of one molecule to residue 53 of its crystallographic symmetry mate (Figure 2).

To confirm this observation, we built and refined a model of CV-N that consisted of two domains from two monomers connected by a new linker. Hereinafter, we refer to this new model as the domain-swapped (dimeric) structure (Figure 3). Refinement of the domain-swapped structure reduced the average temperature factors of resi-

Table 2. Comparison of monomeric and domain-swapped dimeric CV-N crystal models

	Monomeric model	Domain-swapped dimeric model
R_{cryst} (%)/R_{free} (%)	19.2/21.1	18.4/20.6
Average R_{real}[a] (%)/B_{ave}[b] (Å^2) of residues ex. 50 to 53	3.9/14.5	3.8/13.5
R_{real} (%)/B_{ave} (Å^2) of residue 49	2.7/13.5	2.4/12.2
R_{real} (%)/B_{ave} (Å^2) of residue 50	57.8/38.5	19.0/21.0
R_{real} (%)/B_{ave} (Å^2) of residue 51	49.7/49.9	23.6/24.7
R_{real} (%)/B_{ave} (Å^2) of residue 52	90.9/51.5	9.5/23.0
R_{real} (%)/B_{ave} (Å^2) of residue 53	38.8/38.0	14.3/20.7
R_{real} (%)/B_{ave} (Å^2) of residue 54	3.9/16.1	4.2/14.5

[a] The real-space R factor of each residue calculated by CNS (Brünger *et al.*, 1998), which shows the disagreement between the structural model and the electron density around the model.
[b] The average temperature factor of the main-chain atoms (C^α, amide nitrogen atom, carbonyl carbon and oxygen atoms).

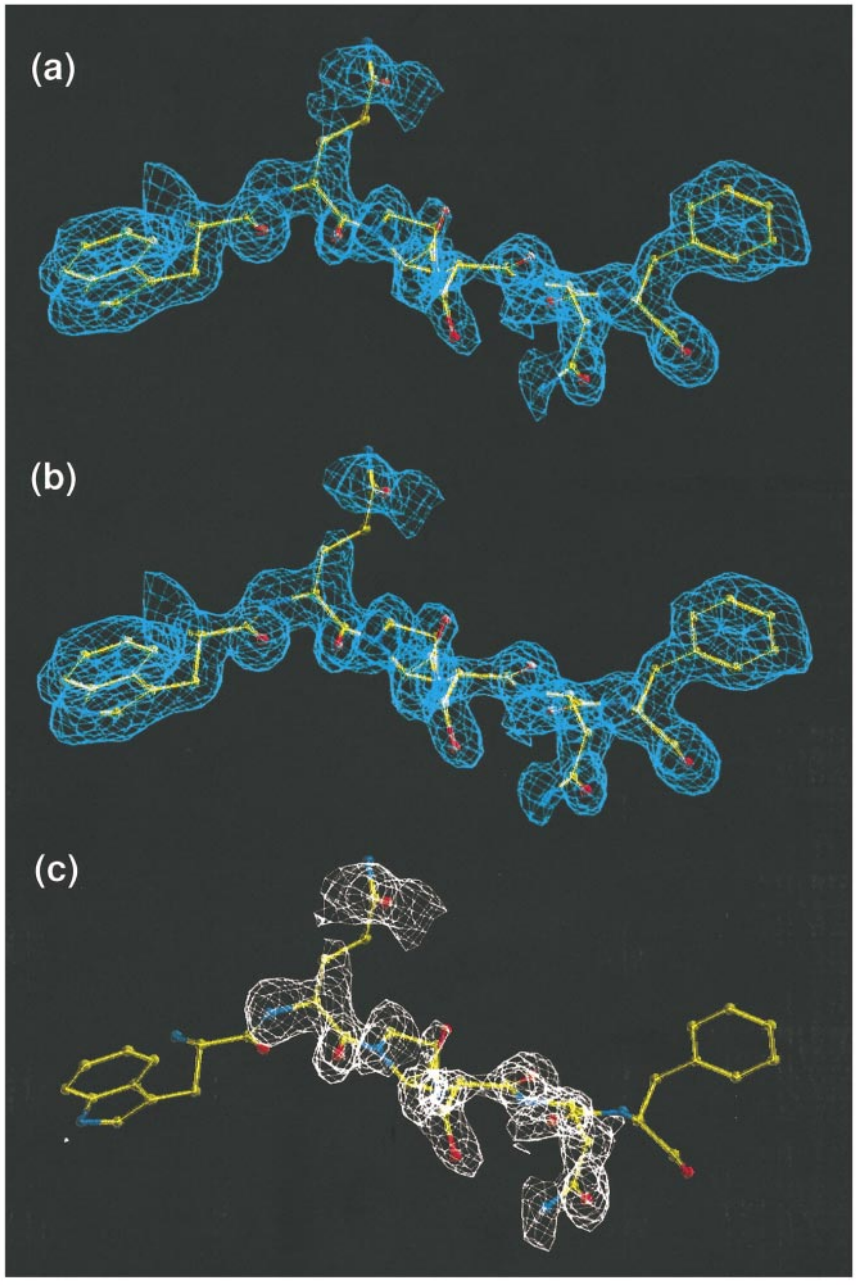

Figure 2. Electron density maps around the region where the chain crosses to the second molecule in the monomeric crystal structure. (a) The $2F_o - F_c$ map calculated from the monomeric crystal structure. The linker in the domain-swapped structure is superimposed on the map. (b) The $2F_o - F_c$ map calculated from the domain-swapped structure. (c) The omit $F_o - F_c$ map calculated by setting the occupancies of residues Gln50 to Asn53 to zero in the domain-swapped structure.

dues 50 to 53 to slightly more than 20 Å^2 (Table 2). The electron density around the new linker is of acceptable quality and, in addition, significant positive electron-density peaks appeared around this region in the omit $F_o - F_c$ map after the occupancies of the four residues were set to zero (Figure 2(c)).

The result that CV-N exists in crystals only as a domain-swapped dimer was unexpected, in view of the predominantly monomeric nature of the protein in solution. Therefore, we also grew crystals from a sample that was independently determined to be monomeric by gel filtration and NMR. A data set extending to 1.75 Å was collected from one of these crystals. In a procedure identical with the one described above, the NMR solution structure was used as the starting model for an independent refinement; nevertheless, the domain-swapped structure was again observed.

Compared with its conformation in the NMR structure, the linker region is extended in the domain-swapped structure. The distance between the C$^\alpha$ atoms of residues 49 and 54 is 10.2 Å in the NMR structure, whereas it is 13.8 Å in the domain-swapped crystal structure. The program PROCHECK showed that all four linker residues in the domain-swapped structure are in the most favored Ramachandran region for random coil, whereas for the crystallographic monomeric structure, one is in the most favored and three are in the generously allowed region. The good geometry of the linker in the domain-swapped structure is further evidence of its existence in the crystals. Although only the residues from Gln50 to Asn53 needed large adjust-

ments, comparison of the ϕ and ψ torsion angles showed that the linker extends from Trp49 to Phe54, since the differences in these angles exceed 20° for these residues (Table 3). Except for the linker region and some flexible side-chains, the domain-swapped structure shares all other structural features of the NMR structure (Bewley *et al.*, 1998), with a high degree of structural similarity between the two domains (Figure 4).

The shape of the monomer in the domain-swapped CV-N structure resembles a bent dumbbell, rather than the football-like monomeric structure. Because there are no direct interactions between the two domains, this structure is probably stable only in the dimeric form. The individual domains in the domain-swapped structure are very similar to each other and to their counterparts in the monomeric crystal structure. Comparing the NMR and crystal structures, we found that the positional RMS deviations of all main-chain atoms between domains A and B are 1.32 Å and 1.31 Å, respectively (Figure 4). Domains A and B in the crystal structure can be fitted separately with their counterparts in the NMR structure, and the positional RMS deviations of all main-chain atoms except for the linker region are 0.52 Å and 0.54 Å, respectively. If domains A and B′ in the dimer are fitted to the whole monomer of the NMR structure, the positional RMS deviation of all main-chain atoms except for the linker region is 0.55 Å.

Discussion

CV-N crystal structures indicated the presence of domain swapping, regardless of whether the crystals were grown from a mixture of dimers and monomers or from purely monomeric samples. However, there is no question that the monomer is the predominant form of CV-N in solution (Figure 5). In particular, analytical ultracentrifugation of the major fraction eluted from a preparative gel filtration column (>95 %) yielded a measured molecular mass of 10,850 Da compared with a predicted mass of 11,013 Da. Furthermore, the sample behaves well, with no evidence of intermolecular association or the existence of a mixture of non-interconverting monomeric and dimeric species. At 35°C, the samples of CV-N used for NMR had $^{1}H_{N}$ and ^{15}N T_2 values of ~40 ms and 150 (±2) ms, respectively, with a rotational correlation time of 4.5 ns and a diffusion anisotropy of 1.5. The values of these relaxation and hydrodynamic parameters are consistent only with a monomer of 10 to 11 kDa. In addition, the measured dipolar couplings are not consistent with the relative orientations of the two halves (A/B′ and A′/B) observed in the dimeric crystal form.

Because the form that crystallized is a domain-swapped dimer, we investigated the origin of the dimeric form. It appears that the domain-swapped dimer can be formed under certain suitable conditions such as those of reverse-phase HPLC used in the purification procedure, or those of the crystallization procedure. A mixture of monomeric and dimeric CV-N is readily obtained by subjecting pure monomeric CV-N to reverse-phase HPLC under acidic conditions (0.05 % trifluoroacetic acid). A ^{1}H-^{15}N correlation spectrum of a lyophilized sample of the HPLC-eluted material dissolved in 90 % H_2O/10 % $^{2}H_2O$ with a measured pH of 2.1 exhibits two sets of cross-peaks in a ratio of about 7:3. The average ^{15}N T_2 for the minor cross-peaks was ~20 ms, corresponding to dimeric CV-N, whereas that of the major species was 40 ms, corresponding to monomeric CV-N. Following direct injection of this low pH sample onto an analytical gel filtration column equilibrated in 50 mM phosphate-buffered saline (PBS; pH 7), 20 % of dimeric CV-N was observed and eluted before the remaining 80 % monomeric CV-N. When the low pH sample was adjusted to pH 6.2 and then analyzed by both analytical gel filtration (on a column equilibrated in 50 mM PBS, pH 7) and ^{1}H-^{15}N correlation spectroscopy, less than 5 % of the dimeric form of CV-N was observed. Moreover, when the pH was lowered again from 6.2 to 3.0 with HCl, the ^{1}H-^{15}N correlation spectrum showed the presence of only a single set of cross-peaks corresponding to monomeric CV-N. Thus, the monomeric form is stable and not easily converted to the dimeric form except by repeating the HPLC purification procedure or by subjecting it to crystallization conditions (low pH and 26 % isopropanol).

Two suggestions have been offered for the ambiguity in defining the two domains of CV-N (Bewley *et al.*, 1998). Based on the high degree of internal sequence similarity, the first and second halves of the molecule can be defined as domains A and B, which we refer to as sequence-based domain definition. By this definition, the interactions between the two domains bury a total of about 3085 Å^2 of accessible surface in the NMR structure (Bewley *et al.*, 1998). To eliminate these interactions, domain A can also be defined as consisting of residues 1 to 39 and 90 to 101 and domain B as consisting of residues 40 to 89 (Bewley *et al.*, 1998), which we refer to as structure-based domain definition. With the observation of domain swapping, the ambiguity in defining the domains may be resolved. The domain-swapped crystal structure shows a flexible linker, or a hinge, extending between residues 49 and 54. Thus, the two sequence-based domains move around the hinge as rigid bodies, forming either a monomer by intramolecular domain interactions (domain A with domain B) or a dimer by intermolecular domain swapping (domain A with domain B′, and domain B with domain A′). There are no interactions between domains A and A′, or B and B′, except for Glu41. Thus the sequence-based domain definition may be more accurate in the case of CV-N.

Domain swapping is a phenomenon that has been discussed for a long time, although its importance has been emphasized only recently

(Schlunegger *et al.*, 1997). It was first reported for RNase A more than 35 years ago (Crestfield *et al.*, 1962), and its first description in a high-resolution structure was provided for a related BS-RNase (Mazzarella *et al.*, 1993; Piccoli *et al.*, 1992). Since then, a number of other proteins, such as diphtheria toxin (DT; Bennett & Eisenberg, 1994; Bennett *et al.*, 1995) and CksHs2 (Parge *et al.*, 1993), have been found to share this feature. The mechanism and evolution of domain swapping have been analyzed (Xu *et al.*, 1998) and reviewed (Bennett *et al.*, 1995). With both monomeric and dimeric states discovered, CV-N provides another example of such a phenomenon. In this particular case, it is clear that the monomeric state is predominant under physiological conditions, as shown by ultracentrifugation, analytical gel filtration, and NMR. In the experiments described here, it seems that the dimeric form of CV-N can be generated either during purification by reverse-phase HPLC or during crystallization (low pH and 26% isopropanol). Under these conditions, the pH of the solution can be quite acidic and the local protein concentration can be much higher. Interestingly, these are also the conditions required for the dimerization of DT (Bennett & Eisenberg, 1994). The decrease in pH was proposed to convert monomeric DT to an open form, which then dimerizes by domain swap-

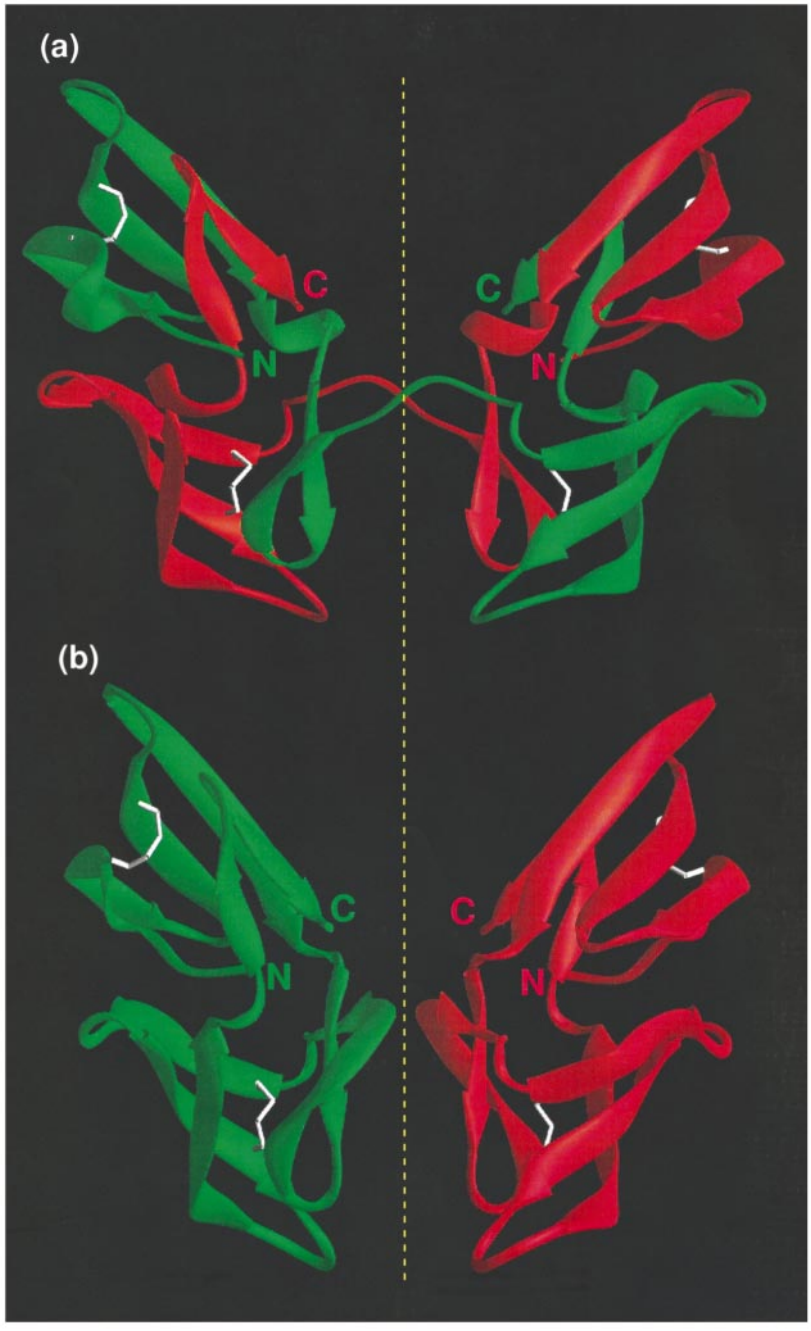

Figure 3 *(legend opposite)*

Figure 3. Crystal structure of CV-N. (a) The domain-swapped CV-N structure. The crystallographic 2-fold symmetry is shown as a yellow broken line. (b) The NMR structure of CV-N is oriented in the same direction as in (a). (c) Another view of the domain-swapped CV-N structure, showing parts of the dimer that are almost perpendicular to each other. (d) The monomeric crystal structure (red), constructed by combining domains A and B′ in the domain-swapped structure, is superimposed on the NMR solution structure (green). The backbone atomic RMS deviation is 0.55 Å.

ping at the high concentrations of the eutectic mixture as the pH returns to neutral during thawing (Bennett & Eisenberg, 1994). Similarly, RNase A is also believed to form an open monomer at low pH and during lyophilization (Crestfield *et al*., 1962).

Domain swapping has been proposed as one of the possible pathways for the evolution of protein oligomerization (Bennett *et al*., 1995; Xu *et al*., 1998). Instead of the unlikely event of mutating pairs of residues on the dimer interface simultaneously, domain swapping provides a stable prototype of a dimer for further evolution (Bennett *et al*., 1995). The crystal structure of interleukin-10 (IL-10) may provide a good example of such a phenomenon (Zdanov *et al*., 1995). IL-10 forms a tight dimer made of two interpenetrating subunits that are stabilized by one intramolecular disulfide bond. The normal monomer of IL-10 cannot exist in solution, since it has many exposed hydrophobic residues. However, rotating a segment of one monomer to the position of its counterpart in the other monomer is prohibited by the intramolecular disulfide bond. It seems that the evolutionary process results in the elimination of the monomeric IL-10 and only the dimeric form is allowed to exist. CV-N has some unique features compared with IL-

Table 3. ϕ and ψ torsion angles of the linker residues

	Domain-swapped crystal structure		NMR structures	
	ϕ	ψ	ϕ	ψ
Lys48	−130	155	−124±1	163±3
Trp49	−74	135	−77±1	82±1
Gln50	−87	132	−166±1	139±1
Pro51	−65	137	−65±2	145±2
Ser52	−81	−3	173±7	13±2
Asn53	−169	170	59±2	60±1
Phe54	−127	−9	−60±1	−42±3
Ile55	−74	−10	−56±3	−21±3

10, DT, and other domain-swapped proteins. The sequence of CV-N indicates that a single molecule of CV-N is probably the result of gene duplication, so that the two domains have a high degree of similarity, whereas the domains of other domain-swapped proteins are usually unrelated. Thus, the interactions between domains A and B in monomeric CV-N molecules may also represent domain swapping or segment swapping.

From the solution data, it is clear that CV-N is predominantly monomeric at neutral pH and the conversion of the minor dimeric form seen after HPLC purification at low pH to the monomeric form upon raising the pH is largely irreversible. We observed, however, that HPLC may not be the only condition leading to dimerization, since crystals containing domain-swapped dimers were also grown from purely monomeric protein, so that the presence of the precipitating agents might also lead to domain swapping. Crystallographic data verifying these phenomena are quite unambiguous. Although the absence of the electron density connecting residues 50 and 53 in the monomeric model would not by itself be sufficient to postulate such molecular organization, the presence of a clear electron density connecting the two crystallographically related molecules cannot be disregarded in the structure refined at a resolution as high as 1.5 Å. Although the dimeric species was the minor components or not present in the samples used for crystallization, only that form of protein could be crystallized. Crystallization of other minor components of protein samples has been reported previously (Duquerroy *et al.*, 1994). The exclusive crystallization of the domain-swapped form of CV-N would have led to incorrect identification of the normal state of the mol-

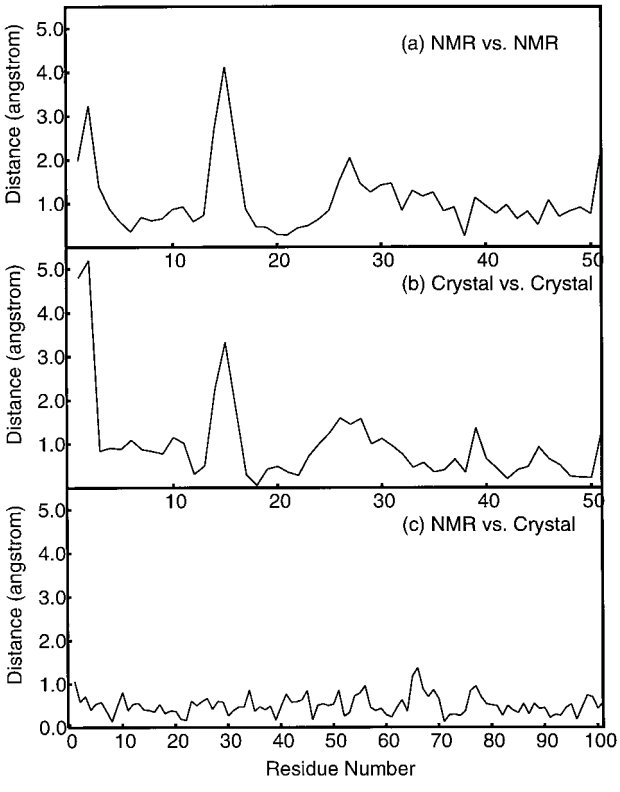

Figure 4. Structural similarity between the domains and molecules of CV-N. Distances between corresponding C^{α} atoms are plotted as a function of residue number after fitting domain A with domain B in the NMR structure (a) and in the domain-swapped crystal structure (b), and after fitting the two domains in the NMR structure with those in the crystal structure (c).

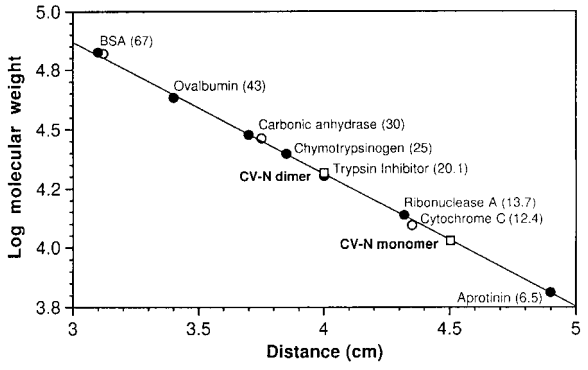

Figure 5. Molecular mass estimation of CV-N by size-exclusion chromatography. Proteins were fractionated on a Superdex-75 column using 50 mM sodium acetate buffer (pH 5.5), 0.02% sodium azide at a flow rate of 0.7 ml/min at ambient temperature. Several proteins were injected individually, as shown by open circles, to rule out anomalous retention of them when injected together. Proteins were also fractionated individually on a Superdex peptide HR column, shown by filled circles, in 20 mM sodium phosphate buffer (pH 6.7), 20 μM EDTA, and 0.02% sodium azide. CV-N monomer (open square) was excluded in both columns with an apparent molecular mass of 10,500 Da. The CV-N dimer (open square) which represented less than 5% of the monomer peak in the Superdex-75 column, eluted with an apparent molecular mass of 19,500 Da. The molecular mass of known proteins are indicated in kDa.

ecule in the absence of the NMR structure, reminding us that crystal structures, even if solved without any errors, may not always correspond to the biologically significant forms of the protein. At this time we do not know if the domain-swapped form of CV-N has any relevance to its antiviral activity, although the possibility that such structures exist should be kept in mind when analyzing the interactions with its viral target.

Materials and Methods

Recombinant CV-N was expressed in *Escherichia coli*; the cloning and initial purification methods are described elsewhere (Mori *et al.*, 1998). Final purification of the sample employed reverse-phase HPLC eluting with an acetonitrile-water gradient containing 0.05% trifluoroacetic acid. Acetonitrile was removed from the CV-N-containing HPLC eluant by rotary evaporation; the residual water and trifluoroacetic acid were removed by lyophilization. Lyophilized CV-N was dissolved in deionized water, and crystallization conditions were established using the Crystal Screen I kit from Hampton Research. After several days of hanging-drop vapor diffusion experiments (Wlodawer & Hodgson, 1975) many small crystals were observed in a droplet containing 20% (v/v) isopropanol, 0.1 M sodium acetate, 0.2 M calcium chloride (pH 4.6). The mother liquor conditions were further refined to 26% (v/v) isopropanol, 0.1 M sodium acetate, 0.15 M calcium chloride, 5% (v/v) polyethylene glycol 400, pH 4.4. Crystals grew to the size of 50 μm × 50 μm × 50 μm after two or three days. A single CV-N crystal was flash-frozen at 100 K and a 1.5 Å resolution X-ray diffraction data set was collected at the wavelength of 0.9795 Å, with a four-element ADSC CCD detector on the synchrotron beamline F2 at CHESS (Cornell University). Diffraction data were processed and scaled with the HKL suite (Otwinowski & Minor, 1997). The structure was solved by molecular replacement methods using the NMR structure as the search model by the program AMoRe (Navaza, 1994). The early stages of refinement utilized X-PLOR (Brünger, 1992b) and SHELXL-97 (Sheldrick & Schneider, 1997), and finally, CNS using maximum-likelihood methods (Brünger *et al.*, 1998). The molecular graphics program O (Jones & Kieldgaard, 1997) was used to rebuild the model, and PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of the structure.

Several experiments were conducted to investigate the transformation between the dimer and the monomer. Equilibrium sedimentation studies were carried out using a Beckman XL-I analytical ultracentrifuge at 20 °C and analyzed using the Beckman XL-A analysis software. Analytical gel filtration studies were performed using both a Superdex-75 column (10 mm × 30 cm, Pharmacia Biotech, NJ) and a Superdex peptide column (10 cm × 30 cm, Pharmacia Biotech, NJ) with 50 mM PBS (pH 7), and a flow rate of 0.5 ml/minute. ^{15}N T_1 and $T_{1\rho}$ relaxation data were acquired and analyzed (Tjandra *et al.*, 1996). The rotational correlation time and diffusion anisotropy were determined by least-square minimization on the basis of the measured relaxation times and the orientations of the internuclear vectors in the NMR structure (Tjandra *et al.*, 1996) and by analysis of the distribution of ^{15}N $T_1/T_{1\rho}$ values (Clore *et al.*, 1998).

Crystallization trials were also set up using a nearly 100% monomeric CV-N sample under the crystallization conditions described above. A single crystal diffraction data set was collected on a Mar345 image plate detector using CuKα radiation generated by a Rigaku rotating-anode X-ray generator. A crystal was flash-frozen at 100 K using an Oxford Cryosystems device. The data set was processed at 1.75 Å using the HKL suite (Otwinowski & Minor, 1997), and the structure was refined using CNS as described above.

Protein Data Bank accession numbers

Coordinates and structure factors have been deposited at the Brookhaven Protein Data Bank under accession numbers 3ezm and r3ezmsf, respectively.

Acknowledgments

We are grateful to Jerry Alexandratos for help with the dynamic light-scattering measurements, and Paul Wingfield for analytical ultracentrifugation. We thank Anne Arthur and Maritta Grau for their excellent editorial assistance and CHESS (Cornell University) for access to their data collection facility. A.W. thanks the Master and Fellows of the Sidney Sussex College and the Department of Biochemistry, University of Cambridge, for a Visiting Fellowship, during which tenure this paper was written. It is numbered 55 in the NCI LDDRD series ''HIV-Inhibitory Natural Products'' (for the past 54 papers, see Esser *et al.*, 1998). This research is sponsored in part by the National Cancer Institute, DHHS, under contract with ABL, as well as by the Intramural AIDS Targeted Antiviral Program of the office of the Director of the National Institutes of Health to G.M.C and A.M.G. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

Baldwin, E. T., Weber, I. T., St. Charles, R., Xuan, J.-C., Appella, E., Yamada, M., Matsushima, K., Edwards, B. F. P., Clore, G. M., Gronenborn, A. M. & Wlodawer, A. (1991). Crystal structure of interleukin-8: symbiosis of NMR and crystallography. *Proc. Natl Acad. Sci. USA*, **88**, 502-506.

Bennett, M. J. & Eisenberg, D. (1994). Refined structure of monomeric diphtheria toxin at 2.3 Å resolution. *Protein Sci.* **3**, 1464-1475.

Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455-2468.

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

Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., II, Buckheit, R. W., Jr., Nara, P. L. & Pannell, L. K., *et al.* (1997).

Discovery of cyanovirin-N, a novel human immunodeficiency virus- inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob. Agents Chemother.* **41**, 1521-1530.

Brünger, A. T. (1992a). The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-474.

Brünger, A. T. (1992b). *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Brünger, A. T., Adams, P. D., Core, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect D*, **54**, 905-921.

Clore, G. M. & Gronenborn, A. M. (1998). New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl Acad. Sci. USA*, **95**, 5891-5898.

Clore, G. M., Gronenborn, A. M., Szabo, A. & Tjandra, N. (1998). Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information. *J. Am. Chem. Soc.* **120**, 4889-4890.

Crestfield, A. M., Stein, W. H. & Moore, S. (1962). On the aggregation of bovine pancreatic ribonuclease. *Arch. Biochem. Biophys.* Suppl. 1, 217-222.

Duquerroy, S., Le, Bras G. & Janin, J. (1994). Lobster enolase crystallized by serendipity. *Proteins: Struct. Funct. Genet.* **18**, 390-393.

Esser, M. T., Mori, T., Mondor, I., Sattentau, Q. J., Boyd, M. R. & Lifson, J. D. (1999). Cyanovirin-N blocks HIV-1 binding and fusion but does not affect sCD4 induced conformational changes in gp120. *J. Virol.* In the press.

Gustafson, K. R., Sowder, R. C., II, Henderson, L. E., Cardellina, J. H., II, McMahon, J. B., Rajamani, U., Pannell, L. K. & Boyd, M. R. (1997). 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.* **238**, 223-228.

Jones, T. A. & Kieldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* **277**, 173-208.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Mazzarella, L., Capasso, S., Demasi, D., Di Lorenzo, G., Mattia, C. A. & Zagari, A. (1993). Bovine seminal ribonuclease: structure at 1.9 Å resolution. *Acta Crystallog. sect. D*, **49**, 389-402.

Mori, T., Shoemaker, R. H., Gulakowski, R. J., Krepps, B. L., McMahon, J. B., Gustafson, K. R., Pannell, L. K. & Boyd, M. R. (1997a). Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. *Biochem. Biophys. Res. Commun.* **238**, 218-222.

Mori, T., Shoemaker, R. H., McMahon, J. B., Gulakowski, R. J., Gustafson, K. R. & Boyd, M. R. (1997b). Construction and enhanced cytotoxicity of a [cyanovirin-N]- [*Pseudomonas* exotoxin] conjugate against human immunodeficiency virus-infected cells. *Biochem. Biophys. Res. Commun.* **239**, 884-888.

Mori, T., Gustafson, K. R., Pannell, L. K., Shoemaker, R. H., Wu, L., McMahon, J. B. & Boyd, M. R. (1998). Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein Expr. Purif.* **12**, 151-158.

Navaza, J. (1994). An automated package for molecular replacement. *Acta Crystallog. sect A*, **50**, 157-163.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Parge, H. E., Arvai, A. S., Murtari, D. J., Reed, S. I. & Tainer, J. A. (1993). Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science*, **262**, 387-395.

Piccoli, R., Tamburrini, M., Piccialli, G., Di Donato, A., Parente, A. & D'Alessio, G. (1992). The dual-mode quarternary structure of seminal RNase. *Proc. Natl Acad. Sci. USA*, **89**, 1870-1874.

Rozwarski, D. A., Gronenborn, A. M., Clore, G. M., Bazan, J. F., Bohm, A., Wlodawer, A., Hatada, M. & Karplus, P. A. (1994). Structural comparisons among the short-chain helical cytokines. *Structure*, **2**, 159-173.

Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. (1997). Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Advan. Protein Chem.* **50**, 61-122.

Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319-343.

Tjandra, N., Wingfield, P., Stahl, S. & Bax, A. (1996). Anisotropic rotational diffusion of perdeuterated HIV protease from ^{15}N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR*, **8**, 273-284.

Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. (1997). Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct. Biol.* **4**, 732-738.

Wlodawer, A. & Hodgson, K. O. (1975). Crystallization and crystal data of monellin. *Proc. Natl Acad. Sci. USA*, **72**, 398-399.

Xu, D., Tsai, C.-J. & Nussinov, R. (1998). Mechanism and evolution of protein dimerization. *Protein Sci.* **7**, 533-544.

Zdanov, A., Schalk-Hihi, C., Gustchina, A., Tsang, M., Weatherbee, J. & Wlodawer, A. (1995). Crystal structure of interleukin-10 reveals the functional dimer with an unexpected topological similarity to interferon gamma. *Structure*, **3**, 591-601.

Edited by R. Huber

(Received 15 December 1998; received in revised form 4 March 1999; accepted 4 March 1999)