The Domain-Swapped Dimer of Cyanovirin-N Is in a Metastable Folded State: Reconciliation of X-Ray and NMR Structures

Laura G. Barrientos,[1] John M. Louis,[1]
Istvan Botos,[2] Toshiyuki Mori,[3] Zhaozhong Han,[3]
Barry R. O'Keefe,[3] Michael R. Boyd,[3]
Alexander Wlodawer,[2]
and Angela M. Gronenborn[1,4]
[1]Laboratory of Chemical Physics
National Institute of Diabetes and Digestive
    and Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20892
[2]Macromolecular Crystallography Laboratory
National Cancer Institute
NCI-Frederick
Frederick, Maryland 21702
[3]Molecular Targets Drug Discovery Program
NCI Center for Cancer Research
National Cancer Institute
NCI-Frederick
Frederick, Maryland 21702

Summary

The structure of the potent HIV-inactivating protein cyanovirin-N was previously found by NMR to be a monomer in solution and a domain-swapped dimer by X-ray crystallography. Here we demonstrate that, in solution, CV-N can exist both in monomeric and in domain-swapped dimeric form. The dimer is a metastable, kinetically trapped structure at neutral pH and room temperature. Based on orientational NMR constraints, we show that the domain-swapped solution dimer is similar to structures in two different crystal forms, exhibiting solely a small reorientation around the hinge region. Mutation of the single proline residue in the hinge to glycine significantly stabilizes the protein in both its monomeric and dimeric forms. By contrast, mutation of the neighboring serine to proline results in an exclusively dimeric protein, caused by a drastic destabilization of the monomer.

Introduction

Three-dimensional (3D) domain swapping is a term coined by Eisenberg [1] for an oligomerization mechanism in which two or more polypeptide chains exchange identical domains. The exchanged portion may consist of a single secondary structure element or an entire globular domain. If exchange is reciprocal between two monomers, dimers are formed, or, if not, oligomers will occur. To date, a large number of domain-swapped dimers have been reported, interestingly, mostly by X-ray crystallography [2]. Indeed, several cases have been reported in which the X-ray structure of the identical protein exhibits a domain-swapped dimer, whereas, in solution, monomeric structures are observed [3–11]. The

classical example for a domain-swapped dimer is bovine pancreatic ribonuclease A (RNase A), discovered already in 1962 by Crestfield, Stein, and Moore under acidic conditions [12], although its X-ray structure was solved only recently [13]. 3D domain swapping has been evoked as a mechanism for amyloid fibril formation [14, 15], and a model for such fiber formation via "polar zippers" has been proposed [16, 17]. Although the detailed pathways that lead to domain-swapped dimers have not been elucidated for most proteins, a general mechanism has emerged; if the folded monomer is subjected transiently to destabilizing conditions, such as low pH, organic solvents, or chaotropic agents, interchange between different segments of the polypeptide chain can occur, especially at high protein concentrations. In this manner, a metastable dimer, which is amenable to structure determination, is produced.

Here we report on studies concerning the domain-swapped dimer formed by the potent human immunodeficiency virus (HIV)-inactivating protein cyanovirin-N (CV-N). We previously solved the solution structure of this protein in its monomeric form by nuclear magnetic resonance (NMR) [10] and the solid-state structure of a domain-swapped dimer by X-ray crystallography [11]. Ribbon diagrams of both the monomeric NMR structure and the swapped dimer X-ray structure are provided in Figure 1. The structure of the monomer in solution can be divided into two pseudosymmetric halves, each of which comprises a triple-stranded β sheet with a β hairpin on top. At the pseudo 2-fold axis, two helical turns connect the halves. As can be appreciated from the view provided in Figure 1, this monomer structure is essentially identical to domains AB′ or A′B of the domain-swapped dimer, with only the hinge residues, Trp49–Phe54, exhibiting significantly altered ϕ, ψ torsion angles [11]. We investigated the folding and the thermodynamic stability of the monomeric and dimeric forms of the wild-type protein, as well as that of a stabilized variant. We demonstrate that the wild-type dimer in solution is in a metastable conformation, which slowly converts into the more stable monomeric form. Characterization of the three-dimensional structure by NMR and crystallography reveals that the CV-N dimer exhibits varying conformations about the hinge region, depending on experimental conditions. Furthermore, we provide evidence that it is possible to stabilize the protein by a mutation in the hinge region, yielding both a stable monomer and dimer of CV-N. In addition, selective destabilization of the monomeric form is also possible by a single mutation.

Results

Folding and Isolation of Dimeric CV-N
During purification of recombinant CV-N, we consistently observed a small fraction of dimeric species, re-

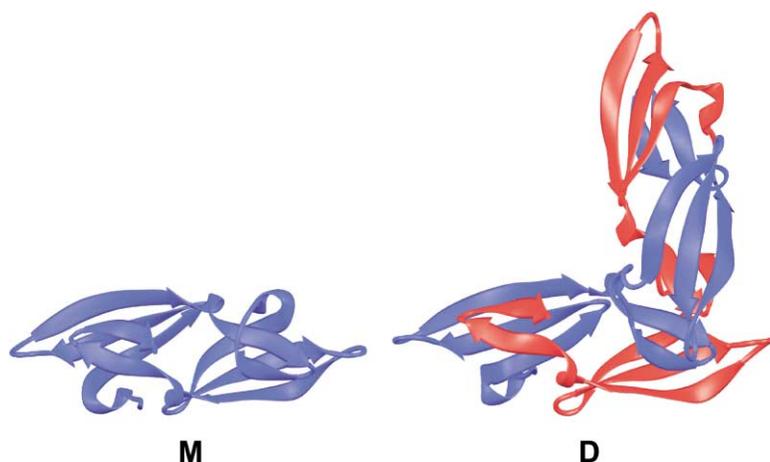[4]Correspondence: gronenborn@nih.gov

Figure 1. Ribbon Representation of the Solution Structure of Monomeric Wild-Type CV-N and the Trigonal X-Ray Structure of the Domain-Swapped Dimer

The different monomer units in the dimer are colored blue and red. M, monomeric wild-type CV-N; D, dimeric domain-swapped wild-type CV-N.

gardless of whether the protein was isolated from the insoluble cellular fraction under denaturing and reducing conditions using refolding or from the soluble fraction in a folded form. Figure 2 shows the result of the purification/folding procedure for CV-N. The profile of the size exclusion chromatography for CV-N, after refolding and removal of the N-terminal His-tag (Figure 2A), clearly shows the presence of a minor fraction of dimeric species and higher-order aggregates. Using the initial purification and refolding procedure detailed in Experimental Procedures, we could not prepare sufficient amounts of dimer for structural and thermodynamic studies. A specific protocol was devised in order to increase the yield of dimeric species. In particular, it was necessary to unfold and refold purified monomeric protein at high concentrations ($\sim$0.3–1 mM) from 8 M urea. In this manner, up to $\sim$40% of the protein can be converted into dimer (Figure 2B) and subsequently isolated by gel filtration. Furthermore, dimeric species can also be obtained by incubation of purified monomer at very high concentration ($>$2 mM) at 38°C (see below). Upon reinjection of the pure dimer fraction, no further partitioning into monomer and dimer is observed.

Solution Characterization

The oligomeric state of monomeric and dimeric CV-N was examined using equilibrium analytical ultracentrifugation, NMR relaxation measurements, and size exclusion chromatography. The ultracentrifugation data for both monomer and dimer samples demonstrated that all were monodisperse and, assuming single, noninteracting species, yielded apparent molecular masses of 11,820 $\pm$ 320 Da and 23,410 $\pm$ 210 Da, respectively. These values are in excellent agreement with those expected for a monomeric (11,621.8 Da) and dimeric (23,243.6 Da) protein. Figure 3 shows a superposition of the two ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of monomeric and dimeric CV-N under identical conditions. Crosspeaks are labeled with residue type and number. Those resonances that are missing or exhibit severe line broadening in the dimer spectrum are underlined. They arise from amino acids surrounding Pro51 in the hinge region (Asn50–Glu56) as well as from residues in a helical loop (Leu36–Glu41)

that is located in close proximity to the proline. It is easily appreciated that the positions of most crosspeaks are essentially identical for the monomer and dimer species. However, discernable differences are observed for the backbone amides of residues close to the tryptophan side chain, namely Leu47, Lys48, and Trp49, as well as Ser38 and Val39. The latter amino acids are located on a helical loop close to the domain interface in the domain-swapped dimer structure previously solved by X-ray crystallography [11] (see also Figures 1 and 7). In addition, the Nϵ1H of the tryptophan side chain resonates in a different location for the monomer and dimer proteins, and these resonances can be used as diagnostic crosspeaks to identify the presence and relative amounts of monomer and dimer. ^{1}H and ^{15}N relaxation measurements at 20°C show that the average T_2 values for the amide nitrogens and for the amide protons for the monomer are $\sim$125 ms and $\sim$50 ms, respectively. The equivalent values for the dimeric protein under identical conditions are $\sim$60 ms and $\sim$23 ms, respectively. The values of these relaxation parameters are consistent with the monomeric and dimeric species exhibiting molecular masses of approximately 11 kDa and 22 kDa, respectively. Likewise, molecular mass estimation by size exclusion chromatography on a calibrated gel filtration column yielded values corresponding to monomer and dimer as described previously [11].

Dimer to Monomer Conversion

A general feature of domain swapping is that monomeric and dimeric forms are separated by a high energy barrier [2]. In the case of CV-N, conversion of dimer to momomer is extremely slow at room temperature or below. Indeed, samples containing 70–300 μM purified dimer did not yield any appreciable amounts of monomer after several months, as assessed by NMR spectroscopy or gel filtration. However, elevating the temperature to 38°C and above speeds up the reaction sufficiently to allow one to follow the conversion. Starting with a sample of 75 μM dimer in 20 mM phosphate buffer (pH 6.0), we determined the ratio of dimer/monomer as a function of incubation time at 38°C by NMR, using the Trp49 Nϵ1H resonance intensities as markers. The corresponding time course for the disappearance of the dimer is illustrated
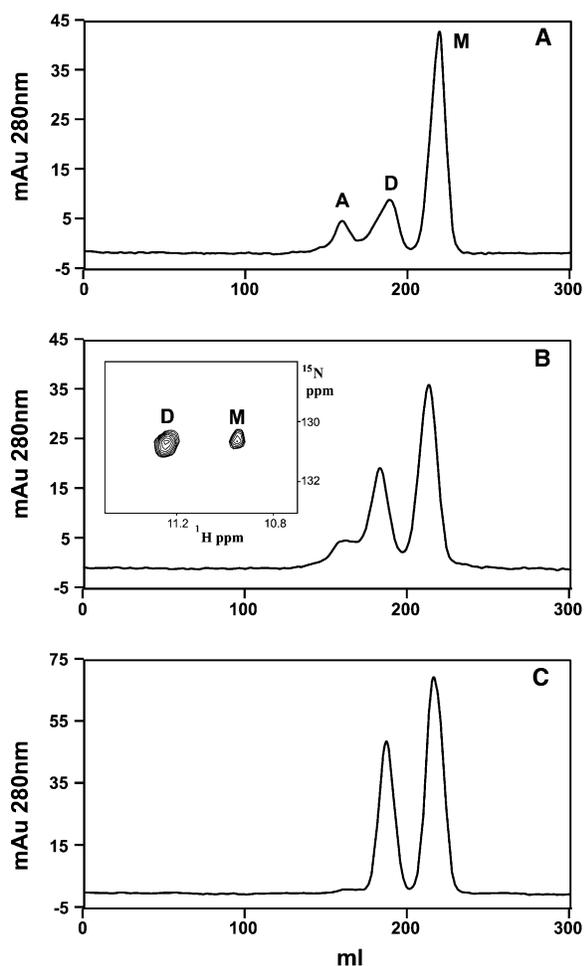
Figure 2. Characterization of the Monomeric, Dimeric, and Higher-Order Aggregates of CV-N by Size Exclusion Column Chromatography

Shown are the elution profiles of CV-N on a Superdex-75 column (see Experimental Procedures).
(A) Profile after extraction from inclusion bodies and refolding from reduced, denatured protein.
(B) After urea unfolding and refolding at high concentration (0.3 mM).
(C) After equilibration of a pure monomeric sample at high concentration (>2 mM) at 38°C for 4 days. The inset in panel B shows the Trp Nε1H region of the ^{1}H-^{15}H HSQC spectrum of a sample refolded at a concentration of 1 mM before column fractionation. The relative intensities of the monomer and dimer peaks in gel filtration agree to within 5% with the intensities of the monomer and dimer signals in the spectrum.

in Figure 4. As is easily appreciated, more than 50% of dimer was converted to monomer after 12 hr. Similarly, incubation overnight at 45°C results in >85% monomer.

The reverse reaction was also monitored. A sample of purified CV-N monomer at a concentration of 2.2 mM was kept at 38°C. After 21 hr, ca. 13% dimer was observed, and an apparent equilibrium was reached after 4 days, yielding monomer and dimer in a 3:1 ratio. In this manner, domain-swapped dimer can also be prepared efficiently (see Figure 2C and Experimental Procedures).

The dissociation constant for wild-type CV-N dimer

was measured quantitatively at 50°C (see Experimental Procedures) and yielded a value of 2.5 mM. The elevated temperature was necessary to establish a fast equilibrium between monomer and dimer. For comparison, the T_m of CV-N is 60°C; therefore, less than 10% of the protein is unfolded.

Equilibrium Denaturation Studies
The stability of monomeric and dimeric CV-N as well as several mutant variants was investigated using urea- and guanidinium hydrochloride (GdnHCl)-induced unfolding. Figure 5A shows the urea unfolding curves, monitored using fluorescence spectroscopy, for monomeric and dimeric wild-type CV-N as well as for the dimeric S52P mutant. The latter was discovered in a screen of a phage-displayed mutant CV-N library [47]. Figure 5B displays the GdnHCl unfolding curves obtained for monomeric wild-type CV-N and the monomer and dimer of the P51G mutant. (Note that complete unfolding of this latter variant could not be achieved with urea.) Inspection of the individual curves yield denaturant concentrations at the midpoint of the transitions, C_m, of 4.7, 3.9, and 3.2 M urea for the wild-type monomer, wild-type dimer, and S52P mutant proteins, respectively. Likewise, midpoint denaturant concentrations of 4.4, 3.2, and 1.6 M GdnHCl were found for the P51G monomer, the P51G dimer, and the wild-type monomer, respectively. Since the fluorescence signal of the monomer and dimer protein is essentially indistinguishable and very different from that of unfolded protein, folded monomer and folded dimer cannot be distinguished, and equilibrium folding/unfolding conditions are only strictly true for the monomeric species of wild-type and P51G mutant CV-N under the conditions employed, therefore precluding the extraction of thermodynamic parameters for the dimeric wild-type and dimeric P52G proteins. For the dilute monomeric proteins, only folding between unfolded and folded monomer is observed. Likewise, the two-state assumption is true for the dimeric S52P mutant CV-N, since no folded monomer can be observed under any conditions. We therefore only extracted thermodynamic parameters for CV-N (M), P51G (M), and S52P (D) from the Gibbs energy function, with ΔG as the free energy of unfolding in the absence of denaturant and with the m value representing the dependence of free energy on denaturant concentration (urea or GdnHCl). Values obtained for these proteins are $\Delta G_{CV-N}(M) = 4.1 \pm 0.2$ kcal · mol^{-1}, $\Delta G_{P51G}(M) = 9.8 \pm 0.5$ kcal · mol^{-1}, and $\Delta G_{S52P}(D) = 4.2 \pm 0.2$ kcal · mol^{-1}. A complete list of parameters is provided in Table 1.

Structure Determination by NMR and Crystallography
Residual NH dipolar couplings ($^1D_{NH}$) were measured for a pure 130 μM dimeric CV-N sample in 25 mM phosphate buffer (pH 8.0) at 20°C using a colloidal suspension of Pf1 for partial alignment. A smaller set obtained for a minor fraction of dimer present in the mixture of monomer/dimer in 4% DMPC/DHPC (3/1) bicelles at 38°C was obtained previously [10]. Residual dipolar couplings were measured from an IPAP spectrum [18], and only those NH resonances that exhibited no overlap were
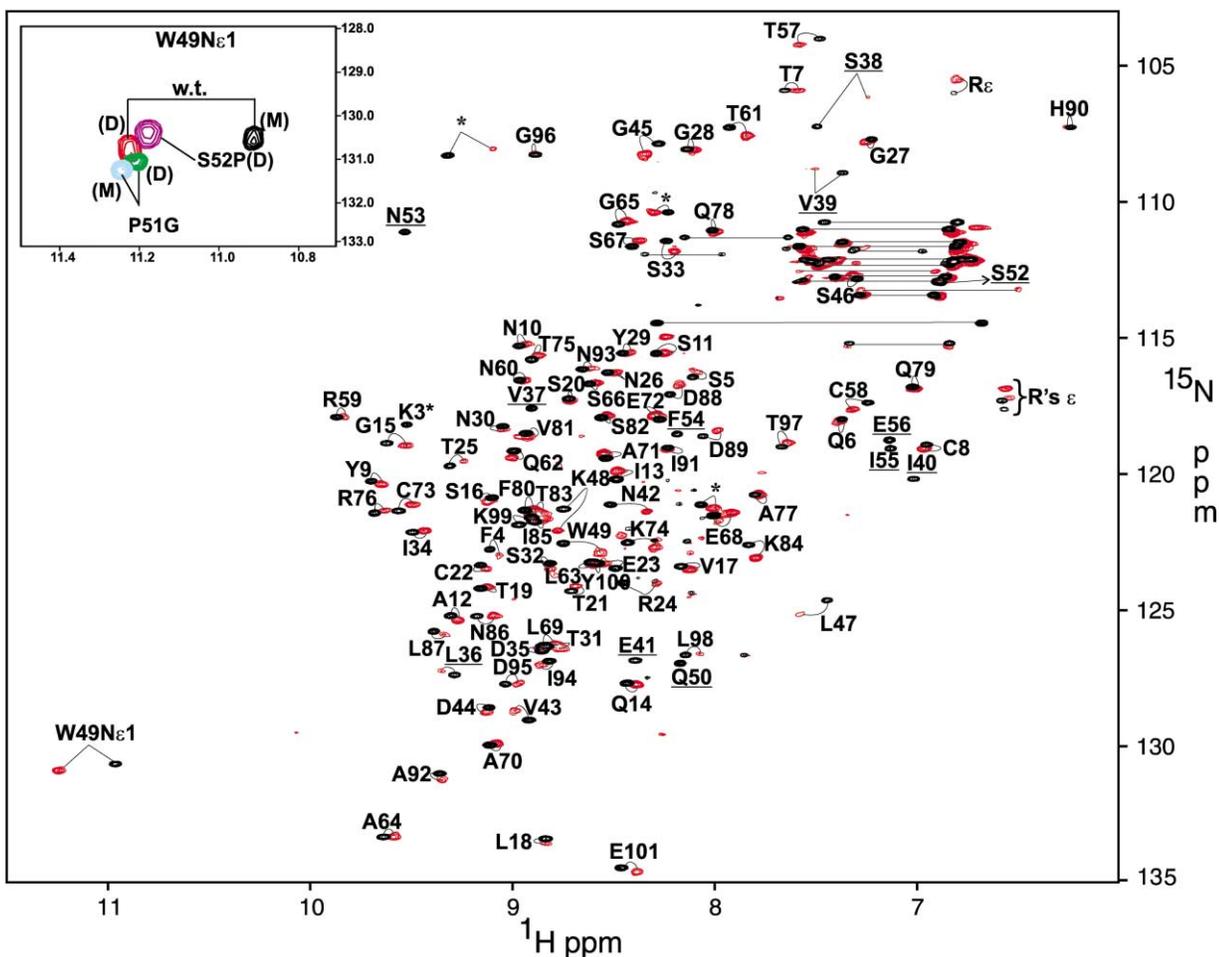
Figure 3. Superposition of the 600 MHz ^{1}H-^{15}N HSQC Spectra of Monmeric CV-N and Dimeric CV-N in 20 mM Sodium Phosphate Buffer, pH 6.0, at 20°C

Monomeric CV-N, black contours; dimeric CV-N, red contours. All crosspeaks are labeled with the amino acid type and position in the mature CV-N sequence. Crosspeaks marked by asterisks arise from the additional amino acids at the N terminus. Underlined residues in the spectrum of the monomer exhibit extreme line brodening in the spectrum of the dimer. The inset shows a close-up of the superposition of the characteristic resonances for the Trp49 Nε1H for both monomeric and dimeric wild-type CV-N (black and red contours) as well as those of the P51G (cyan and green contours) and S52P (magenta contours) mutants.

used. In total, 40 residual dipolar couplings were obtained for the dimer aligned in the phage, and 17 were obtained from the bicelle solutions. Using a monomeric unit extracted from the previously determined trigonal dimeric X-ray structure [11], either AB or A′B′, we calculated the magnitude of the axial component (D_a^{NH}) and the rhombicity (R) of the alignment tensor by singular value decomposition (SVD) [19, 20], yielding values of 14.6 Hz and 0.4 for the Pf1 aligned sample. A ribbon representation of the crystal structure is provided in Figure 6A, and the comparison between the experimentally measured residual dipolar couplings and those calculated based on the X-ray coordinates is provided in Figure 6B. The rmsd value for ^{1}D$_{NH}$ (Pf1) is 5.73 Hz, with a linear correlation coefficient of 0.89.

Our previous crystallographic investigation of CV-N [11] utilized trigonal crystals. A new tetragonal crystal form of CV-N belonging to the space group P4$_1$2$_1$2, with unit cell parameters of a = b = 61.97 Å and c = 148.4 Å, was grown at high pH. A summary of the parameters

used for data collection and refinement is provided in Table 2. The tetragonal structure is also domain swapped, with a full dimer in the asymmetric unit, whereas, in the trigonal structure, there is only one monomer in the asymmetric unit. The tetragonal domains (AB′ and A′B) superimpose with an rms difference of 0.38 Å with each other (backbone of 97 residues) and an rms difference of 0.56 Å (backbone of 96 residues) with the trigonal domain (AB′). While the domains are virtually identical, there is a clear difference in the relative orientation of the two domains between the trigonal and tetragonal X-ray structures (see below). The tetragonal domain-swapped dimer is more "open," with a van der Waals surface area of 9026.4 Å^2 compared to the 8642.2 Å^2 surface area of the trigonal dimer. Using this new X-ray structure (Figure 6C), we again carried out a best fit of the alignment tensor to the coordinates of the structure by SVD and compared the observed residual dipolar couplings with those predicted. The magnitude of the axial component (D_a^{NH}) and the rhombicity (R) of
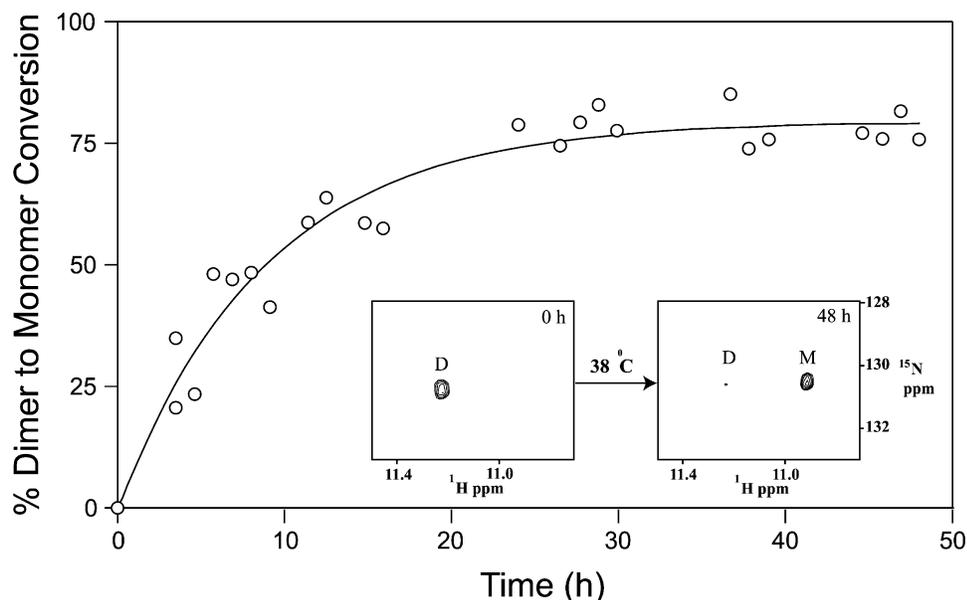
Figure 4. Time Course for the Conversion of Dimer to Monomer for Wild-Type, Domain-Swapped CV-N at 38°C

The relative amount of dimer was quantitated by integration of the Trp49 Nϵ1H resonance intensity. The NMR measurements were carried out at 38°C on samples containing 75 μM protein in 20 mM sodium phosphate buffer (pH 6.0). The inset shows a close-up of the Trp Nϵ1H region prior to incubation at 38°C and after 48 hr.

the alignment tensor are −8.3 Hz and 0.5. As can be appreciated from the correlation plotted in Figure 6D, the agreement between experimental and predicted couplings is considerably worse, with rms difference values of 10.3 Hz for the couplings and linear correlation coefficients 0.56 for the Pf1 data.

Determination of the domain orientation for the solution structure of the dimer was carried out using a procedure analogous to the one described for determining the relative domain orientation in a two-domain protein fragment of a lectin [21]. The starting coordinates were those of two pseudomonomer units (AB′ and A′B) extracted from the refined trigonal 1.5 Å X-ray structure [11], in which proline 51 at the junction between A and B was removed, allowing for free rotation around this junction. We then treated AB′ and A′B as two independent subdomains. Assuming that the orientation of the two subdomains is fixed in solution (at least to a first approximation), the principal axis systems, or alignment frames, of subdomains AB′ and A′B should be equivalent to the alignment system of the entire molecule and, vide infra, to each other. Using the residual dipolar couplings, we calculated the order tensor principal axis systems for each domain. Rotation of pseudosubdomain A′B around the hinge at amino acid position 51 until a superposition of the individual coordinate frames was obtained yielded the final model of the solution dimer. Alternatively, we calculated the order tensor principal axis system for the monomeric unit AB extracted from the dimeric X-ray structure and used the residual dipolar couplings for residues residing in subdomain A and those for amino acids located in subdomain B separately. Rotation of subdomain B around the hinge at residue 51, in an analogous fashion as described above, yielded the relative orientation of A versus B for a mono-

meric unit. Given the C_2 symmetry of the dimer, one of the principal axes of the alignment tensor must be parallel, and the other two orthogonal, to the 2-fold; thus, the dimer model is easily constructed from two AB monomers. The equivalent procedure was carried out for the new tetragonal X-ray structure, yielding identical results. A ribbon representation of the final dimeric solution model is provided in Figure 6E. A comparison of the experimental residual dipolar couplings measured in phage solution (filled circles) and a smaller set obtained using bicelles for alignment (open circles) with those calculated based on this model is shown in Figure 6F. The rmsd values for $^1D_{NH}$ (Pf1) and $^1D_{NH}$ (bicelles) are 1.5 Hz and 1.1 Hz, respectively, and the linear correlation coefficients are 0.99 in both cases. Pertinent parameters for the alignment tensors are 15.1 Hz and 7.6 Hz for the magnitude of the axial component and 0.6 and 0.2 for the rhombicity in Pf1 and bicelle media, respectively. The axis of both alignment tensors relative to the molecular frame of the structure are shown in Figure 6E. Their orientation is clearly different, with differences in angle between the z and x axes of $\geq$45°.

Comparison between Solution and Crystal Dimer

Several of the bipyramidal tetragonal crystals grown at pH 10.3 were dissolved in water, dialyzed against 20 mM phosphate buffer (pH 6) at 4°C, and analyzed by ^{1}H NMR spectroscopy. Only a single resonance was observed in the 1D spectrum for the tryptophan indole region (10.5–11.5 ppm), located at exactly the position of the Trp49 Nϵ1H proton chemical shift identified for the solution CV-N dimer (see Supplemental Figure S1 at http://images.cellpress.com/supmat/supmatin.htm). Therefore, the dimer observed in the crystal structure
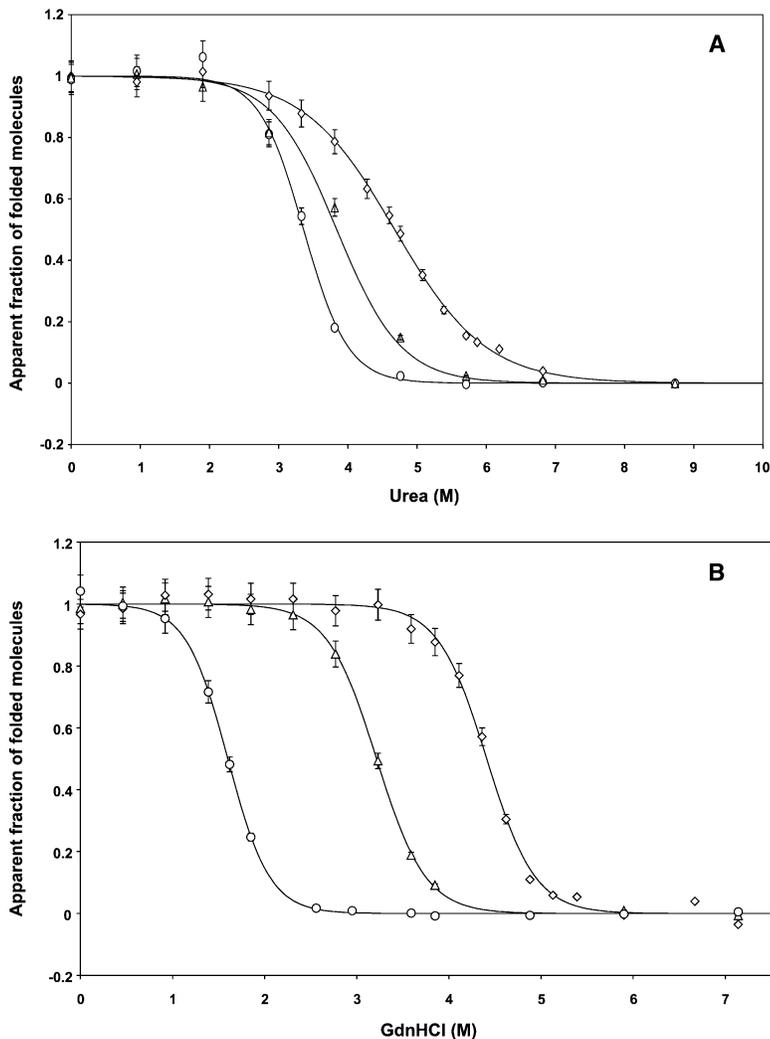
Figure 5. Equilibrium Urea- and GdnHCl-Induced Denaturation Curves for Wild-Type and Mutant CV-N Proteins in Both Monomeric and Dimeric Forms

(A) Urea unfolding curve for monomeric wild-type CV-N (diamonds), dimeric wild-type CV-N (triangles), and the dimeric S52P mutant (circles).

(B) GdnHCl unfolding curve for monomeric P51G mutant CV-N (diamonds), dimeric P51G mutant CV-N (triangles), and monomeric wild-type CV-N (circles). The (un)folding transitions were followed by the intrinsic fluorescence excited at 280 nm. Raw data were converted to the fraction of unfolded molecules and plotted against denaturant concentration. The recovered thermodynamic parameters are summarized in Table 1. All measurements were carried out at 20°C on samples containing 10–20 μM protein in 20 mM sodium phosphate buffer (pH 6.0).

and the dimer in solution have to be very similar in packing and overall structure.

Discussion

Domain-swapped dimers have recently attracted considerable attention [22], since protein polymerization via 3D domain swapping has been suggested to promote fibril formation characteristic of amyloid plaques [15, 17, 23]. Although no atomic structures of such higher-order aggregates have been determined to date, models for oligomerization have been proposed [15, 17, 24]. In particular, side chain hydrogen bonds between glutamine and asparagine residues have been invoked as polar zippers [16, 25], stabilizing antiparallel β sheets, such as those found in amyloid fibrils. Initially thought of as an

Table 1. Thermodynamic Parameters

Protein	[Urea]/2 (M)	[GdnHCl]/2 (M)	ΔG (kcal · mol^{-1})	m (kcal · mol^{-1} · M^{-1})
CV-N (M)[a]	4.7 ± 0.2	—	4.1 ± 0.2	0.88 ± 0.04
CV-N (D)[b]	3.9 ± 0.2	—	—	—
S52P (D)[c,d]	3.2 ± 0.2	—	4.2 ± 0.4	1.6 ± 0.2
CV-N (M)[a]	—	1.6 ± 0.1	4.2 ± 0.2	2.6 ± 0.1
P51G (M)[c]	—	4.4 ± 0.2	9.8 ± 0.5	2.2 ± 0.1
P51G (D)[b,c]	—	3.2 ± 0.2	—	—

[a] Note, identical results were obtained with a C-terminal His-tagged CV-N, demonstrating that the additional residues had no influence on the thermodynamic parameters.
[b] Qualitative assessment; midpoint of unfolding curve.
[c] (His-tag) at the C terminus.
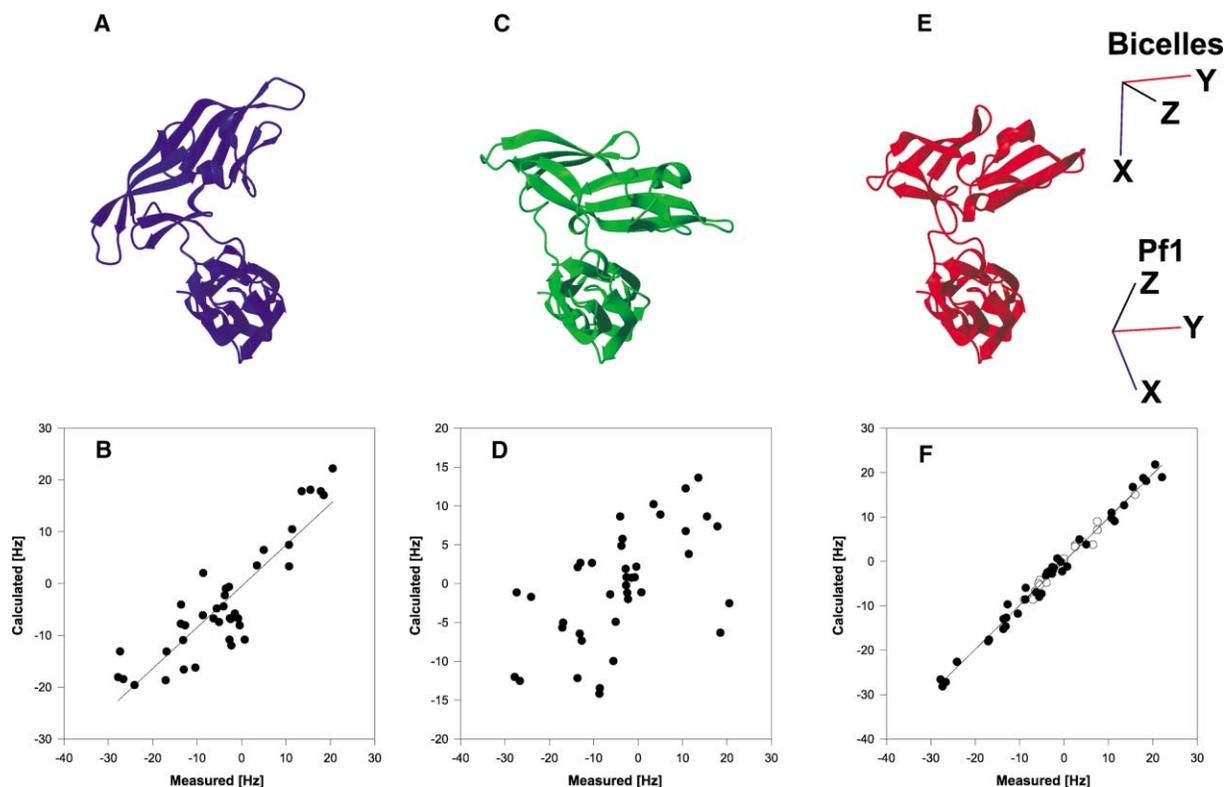[d] The data was fit to the equation for a dimer as described in Experimental Procedures.

Figure 6. Ribbon Representations of the Structures, Top, and Correlations, Bottom, between the Experimentally Measured Dipolar Couplings, $^1D_{obs}NH$, for the Domain-Swapped Dimer of CV-N and Those, $^1D_{calc}NH$, Calculated on the Basis of the Coordinates of the Different Structural Models

Data for the trigonal X-ray structure (A and B), tetragonal X-ray structure (C and D), and the NMR solution structure (E and F). All structures are displayed in the same orientation with respect to the AB′ unit of the dimer. The principal axis frames for the alignment tensors relative to the coordinate frame of the solution structure are also shown in (E).

in vitro curiosity, there is growing evidence that domain-swapped dimers may well be discrete intermediates prior to the formation of well-defined protein aggregates that manifest themselves as fibrous gels for a wide range of different proteins [26]. The protein CV-N is another example in the growing list of proteins for which 3D domain swapping has been observed. As can be ap-

preciated from the data presented in Figure 2, CV-N can fold into different, alternative species exhibiting different molecular masses. As evidenced by a variety of analytical techniques, such as gel filtration, ultracentrifugation, and relaxation measurements by NMR, the two major forms are the monomer and dimer, the latter a domain-swapped species characterized here. Clearly, higher-order aggregates are present as well, but they can be heterogeneous in structure (unpublished data).

For both monomeric and domain-swapped dimeric CV-N to coexist under identical conditions, the free energies for both conformational states must be comparable. The thermodynamic data show that, for wild-type CV-N, the free energy of unfolding is 4.1 kcal/mol for the monomer. Since dimer unfolding for the wild-type is not strictly a two-state system, owing to refolding to also the monomer under equilibrium conditions, it is impossible to obtain an accurate value for ΔG (dimer). We therefore used the S52P mutant as a prototypic example for the dimeric species, given that the midpoint in the urea unfolding curves for wild-type CV-N and the S52P mutant was similar. Analysis of the unfolding curve for the S52P mutant, taking the dimeric nature of the protein into account, yielded a $\Delta G_{S52P}(D) = 4.2 \pm 0.2$ kcal/mol. Substantial stabilization of both monomer and dimer is achieved by a single point mutation of the pivotal proline, Pro51, located in the hinge between the

Table 2. Crystallographic Data Collection and Processing Statistics

Unit cell parameters	a = b = 61.97, c = 148.4, α = β = γ = 90.0
Space group	P4₁2₁2
Molecules/Asymmetric unit	1
Resolution (Å)	2.0
Total reflections	392,421
Unique reflections	20,437
Completeness (%)	
20.0–2.07 Å	99.4
2.07–2.0 Å	99.5
R_{merge} (%)	4.9
Average I/σ	19.4
R factor (%)	23.7
R_{free} (%)	24.5
Rmsd	
Bonds (Å)	0.029
Angles (°)	4.5
Dihedrals (°)	27.9

swapped subdomains in the dimeric structure. For this mutant, unfolding can only be achieved in GdnHCl at room temperature, with a midpoint of denaturation of 4.4 M and 3.2 M GdnHCl for the monomer and dimer species, respectively. For comparison, under identical conditions, the midpoint of denaturation for monomeric wild-type CV-N is observed already at 1.6 M GdnHCl. The free energy of unfolding for the monomer is $\Delta G_{P51G}(M) = 9.8 \pm 0.5$ kcal/mol. The slightly reduced m value observed for the P51G mutant (m = 2.2 kcal/mol M) when compared to the wild-type protein (m = 2.6 kcal/mol M) can easily be explained by the smaller solvent-exposed surface area upon unfolding of the mutant. The energetic difference between the monomer and a domain-swapped dimer can be attributed to a variety of causes. Based simply on entropic considerations (rotational and translational entropy), one would expect that the monomer should be favored. Energetically, on the other hand, a more favorable conformation of the polypeptide chain within the dimer could lower its free energy. Since no new intermolecular interface is formed in the domain-swapped dimer, only the hinge conformation comes into play. In order to offset the intrinsic entropic effect, a difference in hinge conformation could cause an energy difference between the monomer and dimer, favoring the dimer conformation. Therefore, in the present case, the hinge conformation in the loop should be more favorable in the dimeric than in the monomeric structure. Inspection of the solution NMR structure of the monomer reveals that indeed the ϕ, ψ torsion angles of three of the four linker residues (Gln50–Asn53) lie outside the most favored Ramachandran region for a random coil conformation, whereas, in the X-ray structures, all lie within the most favored region. Indeed, considering the secondary structure element originating at the hinge, namely the β hairpin comprising β strands β4 and β5, in the solution structure of the monomer, a crossover of the polypeptide chain is observed for packing the hairpin on top of the triple-stranded β sheet in the C-terminal half of the structure. Such a crossover is generally considered energetically unfavorable in terms of protein architecture. By contrast, the linkers in both X-ray structures exhibit good geometry, and no crossover or strained polypeptide chain is apparent. As noted previously, this difference in local conformation of the linker has very little effect on the overall structure and internal packing of the monomer solution structure or on the pseudomonomeric unit AB′ in the X-ray structure, respectively [11]. Essentially, all interactions present in the monomer are also present in the AB′ unit of the dimer, as can be easily appreciated from Figure 7. Even the local interactions around Pro51 are very similar (Figures 7B and 7C), apart from the fact that Tyr100, His90, and Asp89 reside on the other monomer in the dimer structure. This structural similarity is reflected in almost superimposable backbone amide chemical shifts for the monomer and domain-swapped dimer, immediately apparent by qualitative inspection of the spectra displayed in Figure 3. A noteworthy difference between the ^{1}H-^{15}N HSQC spectra of the monomer and the domain-swapped dimer consists of missing and broadened resonances in the latter. All residues associated with these features reside around Pro51, either

sequentially adjacent in the polypeptide chain or close in space. Their location in the dimer structure is illustrated in Figure 7A, with sequential residues colored green, and those located close in space (albeit on the other monomeric unit) colored yellow. The most likely cause for severe line broadening, even to a degree beyond detection, is motional averaging on the intermediate chemical shift scale, most likely on the milli/microsecond timescale. An obvious candidate responsible for such motion is the proline itself, which clearly could be involved in cis-trans isomerization. Proline residues in folded proteins have been shown by NMR to isomerize, with measured interconversion rates of 0.2/sec for cases where two distinct sets of signals for the cis and trans forms were observed [27]. In the present case we do not observe separate resonances. Instead, severe line broadening leads to loss of these signals, precluding an accurate determination of the lifetimes for the different forms. Nevertheless, it is safe to assume that the two states associated with line broadening via exchange will have faster interconversion rates than states for which separate signals can be observed. Inspection of the two X-ray structures establishes that, in both of them, Pro51 is found exclusively in the trans conformation, with ϕ ranging from $-75°$ to $-85°$ and ψ from 137° to 167°, making cis-trans isomerization an unlikely cause. Clearly, other reasons for line broadening can exist, originating from motions involving either other hinge residues or motions involving the tryptophan ring, which would exhibit large ring current effects on the surrounding residues.

Recent developments in NMR spectroscopy have had a major impact on structural studies of multidomain systems based on exploiting orientational constraints derived from residual dipolar couplings [28, 29]. In particular, the relative domain orientations within multidomain proteins can be determined with relative ease [21, 30]. The domain-swapped dimer of CV-N discussed here constitutes an ideal example for employing this methodology. In solution, the most stable form of the protein is a monomer, and the monomeric structure was previously solved by NMR [10]. Surprisingly, the subsequently solved X-ray structure revealed the protein as a domain-swapped dimer [11]. We previously demonstrated that the difference in quaternary structure (solution monomer versus domain-swapped dimer in the crystal) is not caused by the low pH conditions used for crystal growth [31] and, indeed, both the high- and low-pH crystal forms contain domain-swapped dimers. Our current work establishes that a dimer can also exist in solution, albeit in a kinetically metastable state. However, its lifetime at room temperature or below is sufficiently long for structural studies by NMR. This allowed us to measure a large set of residual dipolar couplings on partially aligned protein in Pf1 phage at 20°C. At this temperature we do not observe any appreciable conversion of dimer to monomer.

The structure of the tetragonal crystal form of CV-N, grown at high pH, again yielded a domain-swapped dimer structure, but with some reorientation of the two domains. In order to assess how the present solution dimer relates to the crystallographic dimer, we dissolved several crystals and recorded a one-dimensional NMR
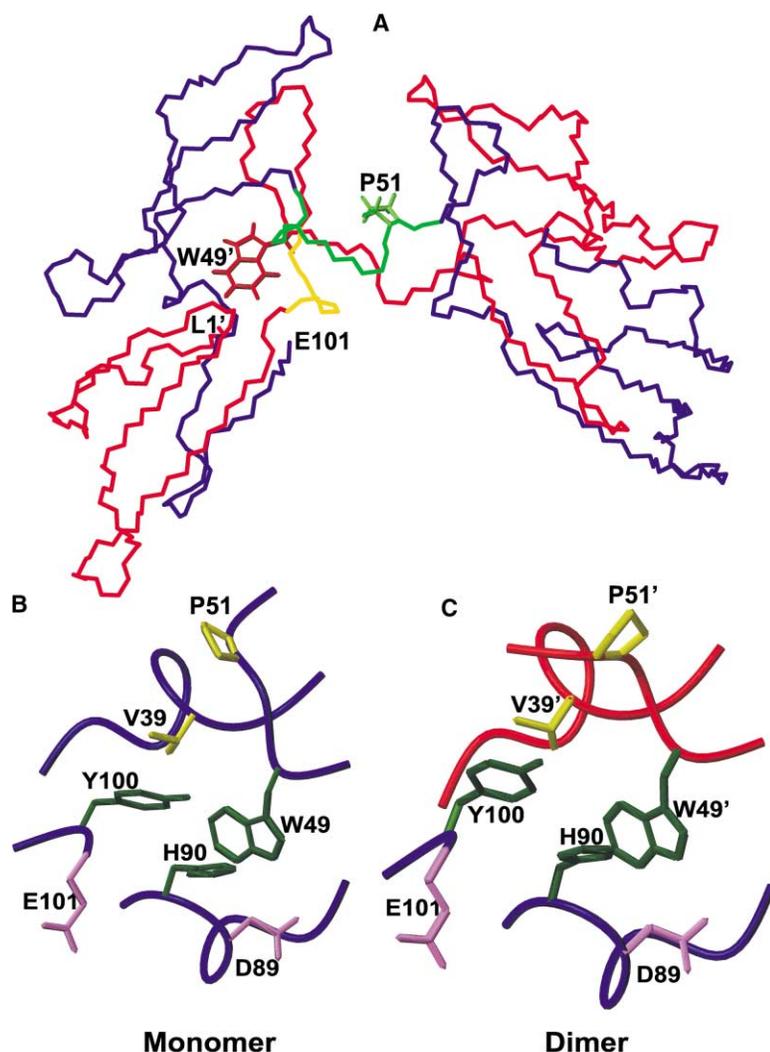
Figure 7. Overall Structure of the Domain-Swapped Dimeric CV-N Structure

(A) The backbone (N, C$_\alpha$, and CO) atoms for monomer AB are shown in blue, and those for monomer A′B′ are shown in red. Hinge and loop residues of AB and A′B′ are colored green and yellow, respectively. Side chains are displayed for the pivotal proline 51 (AB) and tryptophan 49 (A′B′). Environment around the tryptophan side chain in the monomeric NMR structure [10] (B) and the domain-swapped dimeric X-ray structure [11] (C). The polypeptide backbone is displayed in a worm representation; atoms in the same monomer are color coded blue and red. In the monomer structure, this region represents the area around the 2-fold axis of pseudosymmetry at the interface of the two subdomains. In the dimer structure, this region contains residues from both monomeric units (blue and red backbone), with the Trp49 side chain undergoing essentially identical interactions with Asp89, His90, and Tyr100 as observed in the monomer, albeit in an intermolecular mode.

spectrum of the solution. Using the Trp49 Nϵ1H indol resonance as a marker, we only observed a single peak in the appropriate region of the spectrum, exhibiting the proton frequency associated with the dimeric species characterized here (data provided in Supplemental Figure S1 at http://images.cellpress.com/supmat/supmatin. htm). This demonstrates that indeed the domain-swapped dimer observed in the crystal structure also exists in solution. Based on this result, it was deemed appropriate to use the crystallographic structure as the initial model for solving the solution structure of the dimer. The molecular alignment tensors for the individual domains of the domain-swapped CV-N dimer were determined, and the relative domain orientation was established by examination of the pricipal axes of the tensors from these domains as described by Fischer et al. [21].

Irrespective of which of the two X-ray structures was used to determine the domain-domain orientation in solution, identical results were obtained, demonstrating that the experimental dipolar constraints measured in colloidal phage solution are responsible for the relative positioning of the domains. In addition, a smaller number of residual dipolar couplings measured in liquid crystal

bicelle medium is consistent with this orientation. A best fit superposition of the backbone trace for all three swapped dimer structures is shown in Figure 8A. As is easily appreciated, all the structures resemble each other closely, differing solely by reorientation around the hinge region. Indeed, evaluating the relative orientation of the long axes of the two domains (AB′ and A′B) that are almost perpendicular to each other in terms of a torsion angle using the Sγ atoms of the cysteines in the two disulfide bonds (i.e., Cys 8/Cys58′/Cys58/Cys8′), values of −102.2°, −69.9°, and 127.3° are obtained for the trigonal, tetragonal, and solution orientation, respectively. This angle, however, does not take into account the differences in tilt. A more quantitative description of the differences around the hinge between the three structures is afforded by comparison in two orientations as depicted in Figures 8B and 8C. A comparison of the two relevant angles θ and ϕ reveals a difference of 8° and 95° between the trigonal and tetragonal X-ray structures. The NMR structure exhibits a θ difference of 39° and 47° and a ϕ difference of 60° and 35° versus the trigonal or tetragonal X-ray structures, respectively. The variability in domain orientations between the X-ray and solution

A

B

C

X-ray (trigonal)   : θ = 21°; ϕ =-14°
X-ray (tetragonal): θ = 29°; ϕ = 81°
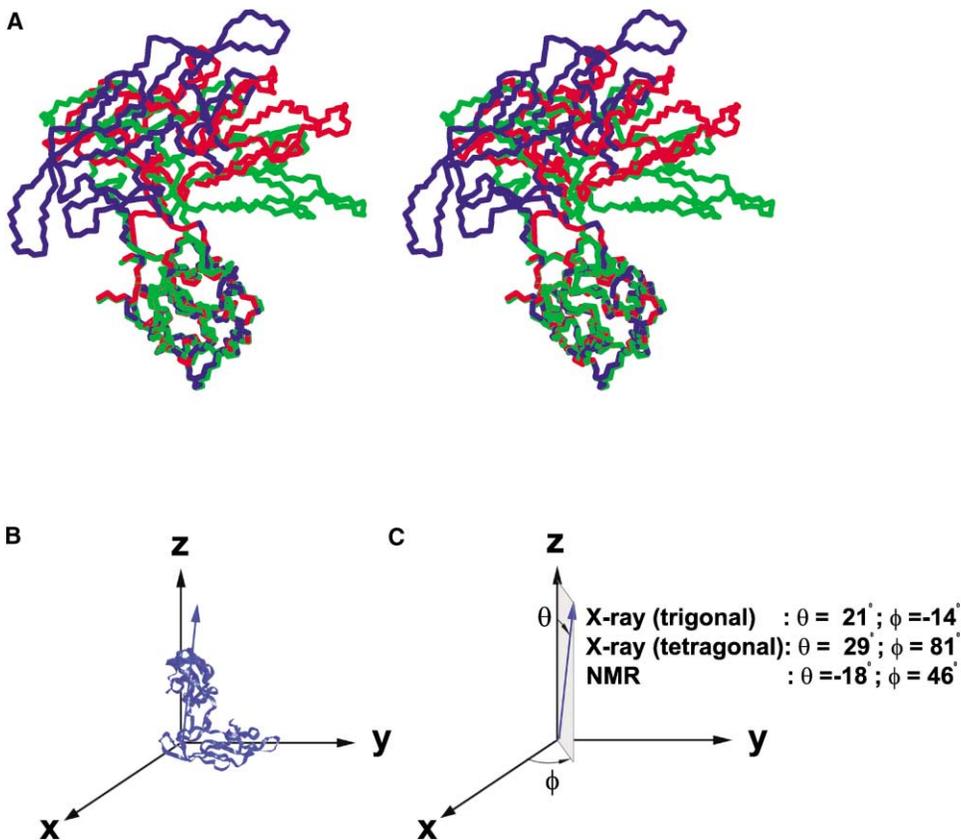NMR   : θ =-18°; ϕ = 46°

Figure 8. Comparison between X-Ray and NMR Structures

(A) Stereoview of a superposition for the three domain-swapped dimer structures. The trigonal and tetragonal X-ray and the solution structures are shown in blue, green and red, respectively. A best fit superposition for the backbone coordinates to the AB′ half of the original X-ray structure was carried out.

(B and C) Schematic illustration of the relative orientation between the two domains. In (B) the trigonal structure is oriented in a cartesian coordinate system such that the long axis of the AB′ domain lies along the y axis. Orientation of the long axis of the A′B domain is defined by the angle θ, with respect to the z axis, and the angle ϕ, which represents the projection of this vector onto the x, y plane. Values for θ and ϕ for all three structures are given in (C).

structures for the domain-swapped dimer of CV-N clearly demonstrates that the protein is inherently capable of structural rearrangements around the hinge region, and, indeed, some flexibility is observed in solution for residues associated with the hinge around Pro51. It should be stressed at this point that the solution structure represents an average structure, and motion of the two domains around the hinge region may well be present. This motion, however, cannot be extensive, and, clearly, both domains do not tumble independently, as if connected by a completely flexible tether. In the latter case we would have not been able to fit all the residual dipolar couplings with a single order tensor principal axis system, and the ranges of observed couplings for the two domains would have been different [21].

Proline residues are well known to have profound influences on protein conformation [32] and stability. Indeed, proline mutations were proposed as a means for enhancing stability by reducing the configurational entropy of unfolding [33]. This concept has not been borne out experimentally, since both stabilization and destabilization of proteins upon mutation to prolines have been observed, with the latter seemingly more common. The

context and, in particular, sequence conservation within protein families seems to be the best predictor for the importance of a proline in terms of stability [34]. Prolines have also been proposed to promote protein-protein interactions by lowering the entropic cost of association, and, interestingly, a preponderance of proline residues has been noted in X-ray structures of oligomers exhibiting arm exchange or domain swaps [35, 36]. Our results with respect to the stability of the monomer and domain-swapped dimer mutants address these questions directly and quantitatively for the CV-N protein. For both the wild-type and P51G mutant CV-N dimer, the domain-swapped dimers are metastable structures, and the mutation of proline to glycine has most likely not changed the relative order of stability between monomer and dimer. The energy difference between monomer and dimer is increased by changing the pivotal proline in the hinge region to glycine, rendering the P51G mutant predominantly monomeric. The most dramatic change in terms of thermodynamic stability is a net increase in stability for the P51G mutant, both for the monomeric and dimeric species. A net stabilization of 5.7 kcal/mol is observed for the monomer, while an exact value for

the dimer cannot be extracted, since unfolding is not strictly two state. In this context it may be interesting to point out that, despite considerable effort, we have not been able to crystallize the P51G protein, most likely because it exists preferentially in a very stable, monomeric form.

Although we have no absolute values with respect to the activation barrier between the monomeric and dimeric states, we have some indication about the relative heights from our experimental results on converting dimer to monomer. Increasing the temperature from 20°C to 38°C and above speeds up the conversion from dimer to monomer considerably, similar to observations reported recently for p13suc1 [36]. We also consistently observed that it takes longer to convert the P51G mutant dimer to monomer under identical experimental conditions, suggesting that the activation barrier for the mutant has to be higher than that for the wild-type CV-N.

Mutation of Ser52 to proline results in an exclusively dimeric protein under native conditions, rendering the polypeptide chain incapable of folding into a monomeric form. The thermodynamic stability of the S52P dimer is essentially identical to that of monomeric wild-type CV-N (see Figure 4A and Table 1). The fact that the S52P mutant protein is only capable of existing in its dimeric folded form can be explained by the additional conformational restriction imposed by the second proline ring on the polypeptide backbone. Indeed, theoretical calculations indicated that the "strain effect" of proline residues is additive along the peptide chain [37], suggesting that the mutant hinge with two adjacent prolines is less flexible than the wild-type hinge. Consequently, the tight backbone conformation around residues 49–53 that allows chain crossing and folding back of the hairpin onto the triple-stranded β sheet in the wild-type CV-N monomer cannot be achieved in this mutant.

Does our data allow us to characterize the state from which strand exchange occurs, or, in other words, does strand exchange occur from the completely unfolded state, as proposed for p13suc1 [36], or from an intermediate state? We believe that strand exchange occurs from a partially folded intermediate based on the evidence presented below. (1) Folding of the dimer from the reduced, monomeric unfolded state occurs by an independent pathway and not via initial monomer formation. Under our folding conditions (see Experimental Procedures), at room temperature, up to 40% of dimer yield can be obtained. However, we never observed more than 10% dimer with time, starting with a pure monomeric sample, unless elevated temperature was used. Indeed, we can manipulate the solution conditions such that we can shift the proportions of monomer and dimer. The experimentally determined dissociation constant for the dimer is 2.5 mM at 50°C. At that temperature less than 10% of the protein is unfolded. (2) Upon unfolding/refolding of pure monomeric CV-N at high concentrations, we observe the formation of monomer, dimer, and aggregated species. Indeed, we purified a fraction of protein eluting at the position of pure domain-swapped dimer on a gel filtration column, which exhibited an ^{1}H-^{15}N HSQC spectrum containing resonances characteristic of random coil structure superimposed on those arising from folded protein. (3) Conversion of

dimer to monomer at 38°C appears to be faster for a sample at high protein concentration than at low protein concentration. If strand exchange would occur from the completely denatured state, one would expect the conversion rate to increase with decreasing protein concentration. In order to fully understand this phenomenon, further experiments will be necessary. However, a possible mechanistic explanation for a faster conversion rate at higher protein concentration could involve reptation, a phenomenon known and studied in depth for polymer chains.

As to the antiviral activity of the different proteins, we note that, for wild-type protein with or without His-tag, essentially identical activities were observed when pure monomer or dimeric protein was used as the reagent. These results are consistent with the fact that, under assay conditions (38°C), essentially all dimer will have converted to monomer within the time course of the assay. Thus, we cannot determine whether the domain-swapped dimeric wild-type CV-N exhibits anti-HIV activity. Under similar assay conditions, we observed wild-type-like activity for the P51G mutant [48] and moderate anti-HIV activity for the pure dimeric S52P mutant protein [47]. It should be pointed out, however, that, given a dimerization constant of 2.5 mM at 50°C, it is highly unlikely that a dimeric species will ever exist under physiological conditions.

Biological Implications

We have carefully characterized the detailed energetic and structural balance between the monomeric and domain-swapped dimeric forms of CV-N. Depending on the experimental conditions, either form predominates and can be isolated. This allowed us to analyze the mechanistic and thermodynamic contributions to domain swapping for this protein in hitherto unprecedented detail. The dimer is a metastable, kinetically trapped intermediate at room temperature and below over a wide pH range, exhibiting comparable thermodynamic stability to the monomer in solution. Under conditions such as those found in the crystal, the dimeric form is stabilized, probably by protein-protein contacts within the crystal lattice, tipping the balance in free energy toward the domain-swapped dimer. The free energy difference between monomer and dimer is changed by mutating the pivotal proline in the hinge region to glycine, rendering the P51G mutant predominantly monomeric. In contrast, mutation of the neighboring hinge loop serine residue to proline (S52P) promotes exclusively swapped dimer formation, most likely due to significantly lowering the entropic cost of intermolecular association. Our results demonstrate that the balance between monomeric and domain-swapped dimeric CV-N can be drastically influenced by experimental conditions and single amino acid changes, supporting the notion that random, transient intramolecular configurations within proteins [38], promoted by small changes in environment, can result in protein oligomerization or aggregation, thereby allowing for the evolution of novel proteins by ordered association or the formation of abnormal aggregates and disease.

Experimental Procedures

Cloning and Expression

The genes for recombinant CV-N and mutants were constructed in pET15b and pET26b vectors (Novagen, Wisconsin), respectively, as described previously [39, 40]. The wild-type CV-N contains GSHMG at the N terminus of the mature 101-amino acid protein [41], and the mutants contain LEHHHHHH at their C termini. All proteins were expressed in *E.coli* BL-21 (DE3) and were uniformly ^{15}N labeled, except for the samples used in the X-ray work. Nucleotide sequences of the cloned DNAs were confirmed by sequencing.

Purification and Folding

The His-tagged wild-type CV-N was purified under denaturing conditions, folded (80 μM), and subjected to thrombin cleavage as described previously [39]. Refolded fractions were concentrated and chromatographed on a Superdex-75 column (HiLoad 2.6 cm $\times$ 60 cm; Amersham Pharmacia Biotech, Piscataway, NJ). Peak fractions (monomer or dimer) were pooled and dialyzed against 5 mM sodium acetate buffer (pH 5.75) and 0.01% NaN$_3$. Mutant CV-N proteins (P51G and S52P) were purified under native conditions using affinity chromatography on Ni-NTA agarose (Qiagen, CA). For these two mutant proteins, the His-tag was not removed. Peak fractions were concentrated and dialyzed against 25 mM sodium acetate buffer (pH 6.0) and 0.05% NaN$_3$,concentrated and chromatographed on a Superdex-75 column (HiLoad 2.6 cm $\times$ 60 cm) in 20 mM sodium phosphate buffer (pH 6.0). For P51G a single peak eluting at the position of monomer was present, while, for S52P, the single peak exhibited a retention time only consistent with a dimeric protein. All purified proteins were concentrated and stored at 4°C.

Preparation of Solution Dimer

A solution containing 0.3–1 mM wild-type CV-N was first unfolded by dialysis overnight at 4°C against 8 M urea and then refolded by extensive dialysis against 20 mM sodium phosphate buffer (pH 6.0). Insoluble aggregates were removed by centrifugation, and the clarified solution was chromatographed on a Superdex-75 column as described above. For the P51G mutant, unfolding was achieved by dialysis at high concentration (>1 mM) against 8 M GdnHCl. For this mutant, essentially no insoluble aggregates were observed. The S52P mutant existed exclusively as a dimeric species, and even refolding at very low concentration ($\leq$ 50 μM) did not yield any folded monomeric protein.

Determination of the Dimerization Constant
for Wild-Type CV-N

Monomeric wild-type CV-N was purified, and samples were made at varying concentrations ranging from 0.5 mM to 4 mM. These samples were incubated at 50°C for 10–30 min and transferred to 4°C, and the amounts of monomer and dimer were quantified by size exclusion chromatography on a Superdex-75 16 $\times$ 60 column (Amersham Biosciences, Piscataway, NJ) in 20 mM sodium phosphate buffer (pH 6.0) at room temperature. A plot of the square of the monomer concentration versus the dimer concentration yielded a straight line with the slope of the fit equal to the dissociation constant for the dimer. The experimentally determined slope was 2.5 $\pm$ 0.7 mM.

Fluorescence Spectroscopy

Fluorescence spectra were obtained on a Perkin Elmer Model LS-50B spectrofluorometer equipped with a temperature-controlled water bath (NESLAB). Measurements were performed with a scan speed of 120 nm/min and data intervals of 0.5 nm. An excitation wavelength of 280 nm was used, and the intrinsic fluorescence (I_{330}/I_{360}) was recorded. All experiments were performed at 20°C in 20 mM sodium phosphate buffer (pH 6.0).

Equilibrium (un)folding induced by GdnHCl or urea was monitored by steady-state tryptophan fluorescence. Each sample was dissolved in 20 mM sodium phosphate buffer (pH 6.0) containing the desired amount of denaturant. Protein concentration was estimated by absorbance using a molar extinction coefficient of 10,220 M^{-1} cm^{-1}. The concentrations of GdnHCl and urea in the stock solutions

were determined by refractive index measurements using a Bausch & Lomb refractometer. The equilibrium constant, K, and the free energy change, ΔG, for the (un)folding reaction were calculated using the following basic equations.

$$\Delta G = -RT\ln K = -RT\ln\left(\frac{f_u}{1 - f_u}\right)$$

$$\Delta G = -RT\ln K = -RT\ln\left(2P_t \frac{f_u^2}{1 - f_u}\right) \text{ (dimer)}$$

where R is the gas constant (1.987 cal $\cdot$ deg^{-1} $\cdot$ mol^{-1}), T is the absolute temperature (K), f_u is the apparent fraction of unfolded molecules, and P_t is the total protein concentration. All unfolding experiments were carried out for protein concentrations of 10–20 μM.

NMR

All NMR spectra were recorded on Bruker DRX-600 and DMX-750 spectrometers equipped with (x, y, z)-shielded gradient triple resonance probes at 20°C. ^{1}H-^{15}N HSQC spectra contained 128 $\times$ 512 complex points in the indirect (^{15}N) and acquisition (^{1}H) dimensions, respectively, with total acquisition times of 61 and 62 ms in both dimensions. Samples contained $\sim$0.075–0.3 mM protein in 20 mM sodium phosphate buffer (pH 6.0) and 0.02% NaN$_3$. A 1D ^{1}H spectrum of the dissolved crystals was recorded using presaturation during the relaxation delay for water suppression by collecting 5000 scans with 800 complex points in the ^{1}H dimension. The equivalent 1D spectrum of a monomer/dimer mixture employed ^{15}N decoupling during acquisition. ^{1}D$_{NH}$ residual dipolar couplings were measured from 2D IPAP [^{15}N-^{1}H]-HSQC experiments [18] using the NMRPipe/ NMRDraw suite of programs [41]. The sample contained 150 μM CV-N dimer in a colloidal phage solution of 11.5 mg/ml Pf1 in 25 mM sodium phosphate buffer (pH 8.0).

X-Ray Crystallography

Crystallization conditions were established using the Wizard II screen (Emerald Biostructures, WA) and the hanging drop, vapor diffusion method [42]. After several days, small crystals were observed in a droplet containing 1 M Na citrate and 0.1 M CHES buffer (pH 9.5). This condition was further refined to 1 M Na citrate and 0.1 M CHES buffer (pH 10.3); silica hydrogel (Hampton Research, CA) was used to prevent excessive nucleation. Equivalent amounts of 30 mg/ml CV-N solution were mixed with 1 M Na citrate and 0.1 M CHES buffer (pH 10.3) on the cured silica hydrogel droplet and equilibrated against a reservoir with the same composition at room temperature. Crystals appeared after 1 week and grew to their full size after 2–3 weeks. Crystals were cryoprotected in a mother liquor solution containing an additional 10% glycerol.

X-ray data were collected at 100 K on the beamline X9B, National Synchrotron Light Source, Brookhaven National Laboratory, with the ADSC Quantum4 CCD detector, using the wavelength of 1.07 Å. Data were processed using the HKL2000 suite [43]; the statistics are summarized in Table 2. The structure was solved by molecular replacement with AmoRe [44], using the available crystal structure (3ezm) transformed into a compact molecule as the search model. Refinement was carried out with CNS [45], and the model was rebuilt with the molecular graphics program O [46].

Conversion of Dimer to Monomer

Purified domain-swapped dimeric CV-N was incubated in 20 mM sodium phosphate buffer (pH 6.0) at 38°C for an extended period of time at concentrations of 75 and 200 μM. The time course for the disappearance of dimer and appearance of monomer was followed by NMR. In particular, the intensities of the Trp49 NϵH resonances in ^{1}H-^{15}N HSQC spectra recorded at 38°C were determined by integration. The relative intensities for the Trp resonance of the dimer were plotted as a function of incubation time for the 75 μM sample. Similar results were obtained for the 200 μM sample. Incubation at 45°C increases the conversion rate and leads to completion overnight. Conversion of the mutant P51G dimeric protein occurs more slowly than for wild-type CV-N dimer under identical experimental conditions. The S52P dimeric protein exists exclusively as dimer and unfolds at higher temperature into a random coil, unstructured polypeptide.

Biological Assays
The anti-HIV activity was characterized using the XTT-terazolium assay [22]. Briefly, 100μl of serial dilutions of either wild-type or mutant CV-N were added to designated wells of a 96-well plate. Subsequently, CEM-SS cells at 1×10^{-5} cells/ml and HIV-1$_{RF}$ virus were added in 50 μl quantities. Plates were incubated at 37°C for seven days and stained with XTT. Four replicates for protection and two for toxicity were carried out.

References

1. Bennett, M.J., and Eisenberg, D. (1994). Refined structure of monomeric diphtheria-toxin at 2.3-angstrom resolution protein. Protein Sci. *3*, 1464–1475.

2. Bennett, M.J., Choe, S., and Eisenberg, D. (1994). Domain swapping—entangling alliances between proteins. Proc. Natl. Acad. Sci. USA *91*, 3127–3131.

3. Kishan, K.V.R., Scita, G., Wong, W.T., DiFiore, P.P., and Newcomer, M.E. (1997). The SH3 domain of Eps8 exists as a novel intertwined dimer. Nat. Struct. Biol. *4*, 739–743.

4. Driscoll, P.C., Cyster, J.G., Campbell, I.D., and Williams, A.F. (1991). Structure of domain-1 of rat lymphocyte-T CD2 antigen. Nature *353*, 762–765.

5. Murray, A.J., Lewis, S.J., Barclay, A.N., and Brady, R.L. (1995). One sequence, 2 folds—a metastable structure of CD2. Proc. Natl. Acad. Sci. USA *92*, 7337–7341.

6. Arvai, A.S., Bourne, Y., Hickey, M.J., and Tainer, J.A. (1995). Crystal-structure of the human cell-cycle protein CKSHS1: single-domain fold with similarity to kinase N-lobe domain. J. Mol. Biol. *249*, 835–842.

7. Vijayalakshmi, J., Mukhergee, M.K., Graumann, J., Jakob, U., and Saper, M.A. (2001). The 2.2 angstrom crystal structure of Hsp33: a heat shock protein with redox-regulated chaperone activity. Structure *9*, 367–375.

8. Mazzarella, L., Capasso, S., Demasi, D., Dilorenzo, G., Mattia, C.A., and Zagari, A. (1993). Bovine seminal ribonuclease—structure at 1.9 Å resolution. Acta Crystallogr. D Biol. Crystallogr. *49*, 389–402.

9. Dursi, A., Oschkinat, H., Cieslar, C., Picone, D., Dalesio, G., Amodeo, P., and Temussi, P.A. (1995). Assignment and secondary-structure determination of monomeric bovine seminal ribonuclease employing computer-assisted evaluation of homonuclear 3-dimensional H-1-NMR spectra. Eur. J. Biochem. *229*, 494–502.

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

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

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

13. Liu, Y.S., Hart, P.J., Schlunegger, M.P., and Eisenberg, D. (1998). The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-angstrom resolution. Proc. Natl. Acad. Sci. USA *95*, 3437–3442.

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

15. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. (2001). Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. Nat. Struct. Biol. *8*, 316–320.

16. Perutz, M.F. (1999). Glutamine repeats and neurodegenerative diseases: molecular aspects. Trends Biochem. Sci. *24*, 58–63.

17. Liu, Y.S., Gotte, G., Libonati, M., and Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. Nat. Struct. Biol. *8*, 211–214.

18. Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J. Magn. Reson. *131*, 373–378.

19. Losonczi, J.A., Andrec, M., Fischer, M.W.F., and Prestegard, J.H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. J. Magn. Reson. *138*, 334–342.

20. Zweckstetter, M., and Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J. Am. Chem. Soc. *122*, 3791–3792.

21. Fischer, M.W.F., Losonczi, J.A., Weaver, J.L., and Prestegard, J.H. (1999). Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. Biochemistry *38*, 9013–9022.

22. Newcomer, M.E. (2001). Trading places. Nat. Struct. Biol. *8*, 282–284.

23. Ogihara, N.L., Ghirlanda, G., Bryson, J.W., Gingery, M., DeGrado, W.F., and Eisenberg, D. (2001). Design of three-dimensional domain-swapped dimers and fibrous oligomers. Proc. Natl. Acad. Sci. USA *98*, 1404–1409.

24. Sinha, N., Tsai, C.J., and Nussinov, R. (2001). A proposed structural model for amyloid fibril elongation: domain swapping forms an interdigitating β-structure polymer. Protein Eng. *14*, 93–103.

25. Perutz, M.F., Johnson, T., Suzuki, M., and Finch, J.T. (1994). Glutamine repeats as polar zippers—their possible role in inherited neurodegenerative diseases. Proc. Natl. Acad. Sci. USA *91*, 5355–5358.

26. Dobson, C.M. (1999). Protein misfolding, evolution and disease. Trends Biochem. Sci. *24*, 329–332.

27. Kordel, J., Forsen, S., Drakenberg, T., and Chazin, W.J. (1990). The rate and structural consequences of proline cis-trans isomerization in calbindin-D9K. NMR studies of the minor (cis-Pro43) isoform and the Pro43Gly mutant. Biochemistry *29*, 4400–4409.

28. Prestegard, J.H., Al-Hashimi, H.M., and Tolman, J.R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q. Rev. Biophys. *33*, 371–424.

29. Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. Methods Enzymol. *339*, 127–174.

30. Skrynnikov, N.R., Goto, N.K., Yang, D.W., Choy, W.Y., Tolman, J.R., Mueller, G.A., and Kay, L.E. (2000). Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. J. Mol. Biol. *295*, 1265–1273.

31. Barrientos, L.G., Gawrisch, K., Cheng, N., Steven, A., and Gronenborn, A.M. (2002). Structural characterization of the dilute aqueous surfactant solution of cetylpyridinium bromide/hexanol/sodium bromide. Langmuir, in press.

32. MacArthur, M.W., and Thornton, J.M. (1991). Influence of proline residues on protein conformation. J. Mol. Biol. *218*, 397–412.

33. Matthews, B.W., Nicholson, H., and Becktel, W.J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc. Natl. Acad. Sci. USA *84*, 6663–6667.

34. Schymkowitz, J.W.H., Rousseau, F., and Itzhaki, L.S. (2000). Sequence conservation provides the best prediction of the role of proline residues in p13suc1. J. Mol. Biol. *301*, 199–204.

35. Bergdoll, M., Remy, M.H., Cagnon, C., Masson, J.M., and Dumas, P. (1997). Proline-dependent oligomerization with arm exchange. Structure *5*, 391–401.

36. Rousseau, F., Schymkowitz, J.W.H., Wilkinson, H.R., and Itz-

haki, L.S. (2001). Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. Proc. Natl. Acad. Sci. USA *98*, 5596–5601.

37. Sak, K., Karelson, M., and Jarv, J. (1998). Quantum chemical modelling of the effect of proline residues on peptide conformation. Int. J. Quantum Chem. *66*, 391–396.

38. Perutz, M.F., and Windle, A.H. (2001). Cause of neural death in neurodegenerative diseases attributable to expansion of glutamine repeats. Nature *412*, 143–144.

39. Barrientos, L.G., Louis, J.M., Hung, J., Smith, T.H., O'Keefe, B.R., Gardella, R.S., Mori, T., Boyd, M.R., and Gronenborn, A.M. (2002). Design and initial characterization of a circular permuted variant of the potent HIV-inactivating protein Cyanovirin-N. Proteins *46*, 153–160.

40. Gustafson, K.R., Sowder, R.C., Henderson, L.E., Cardellina, J.H., McMahon, J.B., Rajamani, U., Pannell, L.K., and 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.

41. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

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

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

44. Navaza, J. (1994). An automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.

45. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

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

47. Han, Z.Z., Xiong, C.Y., Mori, T., and Boyd, M.R. (2002). Construction and screening of a T7 phage-displayed mutant cyanovirin N (CV-N) library; discovery of a stable anti-HIV CV-N dimer. Biochem. Biophys. Res. Commun., in press.

48. Mori, T., Barrientos, L.G., Han, Z.Z., Gronenborn, A.M., Turpin, J.A., and Boyd, M.R. (2002). Functional homologs of Cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. Protein Expr. Purif., in press.

Accession Numbers

The coordinates and structure factors were deposited with the Protein Data Bank under accession codes 1L5B and 1L5E.