doi:10.1016/S0022-2836(03)00928-8 J. Mol. Biol. (2003) 333, 141–152

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

A Protein Contortionist: Core Mutations of GB1 that Induce Dimerization and Domain Swapping

In-Ja L. Byeon, John M. Louis and Angela M. Gronenborn*

Laboratory of Chemical Physics National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5, Room 130 Bethesda, MD 20892, USA

*Corresponding author

Immunoglobulin-binding domain B1 of streptococcal protein G (GB1), a small (56 residues), stable, single-domain protein, is one of the most extensively used model systems in the area of protein folding and design. Recently, NMR and X-ray structures of a quintuple GB1 core mutant (L5V/A26F/F30V/Y33F/A34F) that showed an unexpected, intertwined tetrameric architecture were determined. Here, we report the NMR structure of another mutant, derived from the tetramer by reverting the single amino acid position F26 back to the wild-type sequence A26. The structure reveals a domain-swapped dimer that involves exchange of the second β-hairpin. The resulting overall structure comprises an eight-stranded β-sheet whose concave side is covered by two α helices. The dimer dissociates into a partially folded, monomeric species with a dissociation constant of 93(±10) μM.

Published by Elsevier Ltd.

Keywords: GB1; core mutants; NMR structure; domain-swapping; oligomerization

Introduction

Over the last decade, several model systems have emerged that have found widespread use for studying protein folding and design. One such protein is GB1, the immunoglobulin-binding domain B1 of streptococcal protein G. GB1 is a small (56 residues), stable, single-domain protein with one α-helix and a four-stranded β-sheet comprising two β-hairpins.[1,2] We previously created an exhaustive set of hydrophobic core mutants of GB1.[3] Nine positions were permuted randomly to comprise the amino acids Leu, Val, Ile, Phe and Met, and from this extensive set of mutants ($\sim 2 \times 10^6$), HS#124, a variant that contained five changes (L5V/A26F/F30V/Y33F/A34F; Figure 1) compared to the wild-type sequence, was selected for structure determination. Interestingly, the quaternary state of the protein was a tetramer and, most surprisingly, the three-dimensional structure of this mutant exhibited a substantial change in backbone geometry, resulting in a drastic global rearrangement of the entire fold into an extensive intertwined architecture.[4]

Here, we report on a further core mutant of GB1 that adopts yet another oligomeric state. This mutant, HS#124^{F26A}, was created by a single amino acid change back from the tetramer mutant sequence at position 26 (Phe) to the wild-type residue (Ala), thus the protein contains four changes (L5V/F30V/Y33F/A34F; Figure 1) compared to the wild-type sequence.

We determined the three-dimensional structure of this quadruple mutant by NMR, and show that for this protein the GB1 fold is restored, albeit rearranged as a domain-swapped dimer. We compare the current structure with those of monomeric wild-type GB1 and the tetramer mutant and discuss critical residues involved in the structural switch. In addition, we characterized the dimer–monomer equilibrium by NMR, gel-filtration and light-scattering.

Results and Discussion

Protein purification and characterization

Each mutation in the quintuple mutant (HS#124) was singly reverted back to the wild-type amino acid and the resulting proteins were characterized. Figure 1 displays the elution profiles for all single

Abbreviations used: GB1, immunoglobulin-binding domain B1 of streptococcal protein G; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy.

E-mail address of the corresponding author: gronenborn@nih.gov

0022-2836/$ - see front matter Published by Elsevier Ltd.

```
          10         20         30         40         50
          |          |          |          |          |
MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE: WT (Monomer)
----V--------------------F---V--FF----------------------: HS#124 (Tetramer)
----V-------------------------V--FF----------------------: HS#124F26A (Dimer)
```

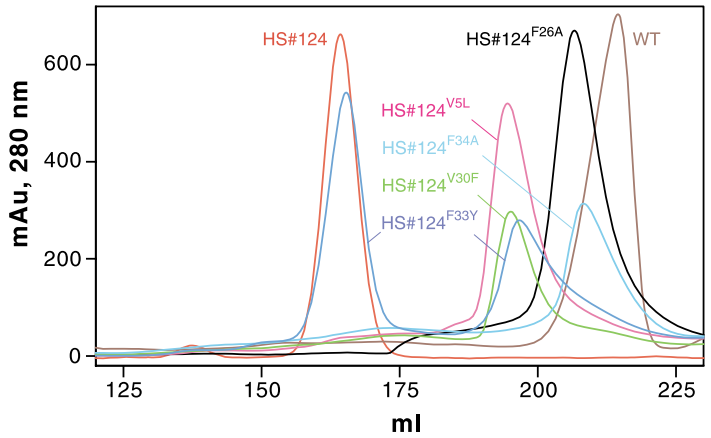

Figure 1. Amino acid sequences and characterization of GB1 mutants by size-exclusion column chromatography. Elution profiles for all single amino acid revertants of the quadruple mutant as well as those of wild-type GB1 and the tetrameric mutant (HS#124) extracted from 200 ml of cell culture (see Materials and Methods) are displayed. The color code is as follows: wild-type GB1; brown, HS#124^{F26A}; black, HS#124; red, HS#124^{F33Y}; blue, HS#124^{V30F}; green, HS#124^{V5L}; magenta, HS#124^{F34A}; cyan.

revertants on a calibrated gel-filtration column in comparison to wild-type, monomeric GB1 and the HS#124 tetramer mutant. As is immediately apparent, none of the single-site revertant proteins elute at the position of wild-type GB1. HS#124^{F33Y} elutes at the tetramer position with a second minor peak eluting at an intermediate position between monomeric wild-type GB1 and the tetramer mutant. Two other revertants, HS#124^{V5L} and HS#124,^{V30F} eluted at this intermediate position as well. HS#124^{F26A} and HS#124^{F34A} eluted last, although somewhat earlier than wild-type GB1. In order to evaluate and characterize these mutant proteins further, $^{1}H-^{15}N$ heteronuclear single quantum coherence (HSQC) spectra were recorded for all eluted peaks. Qualitative inspection revealed that the major earlier eluting peak of HS#124^{F33Y} corresponded to a tetramer structure similar to HS#124, while the minor, later eluting peak exhibited features commonly observed for unfolded/aggregated

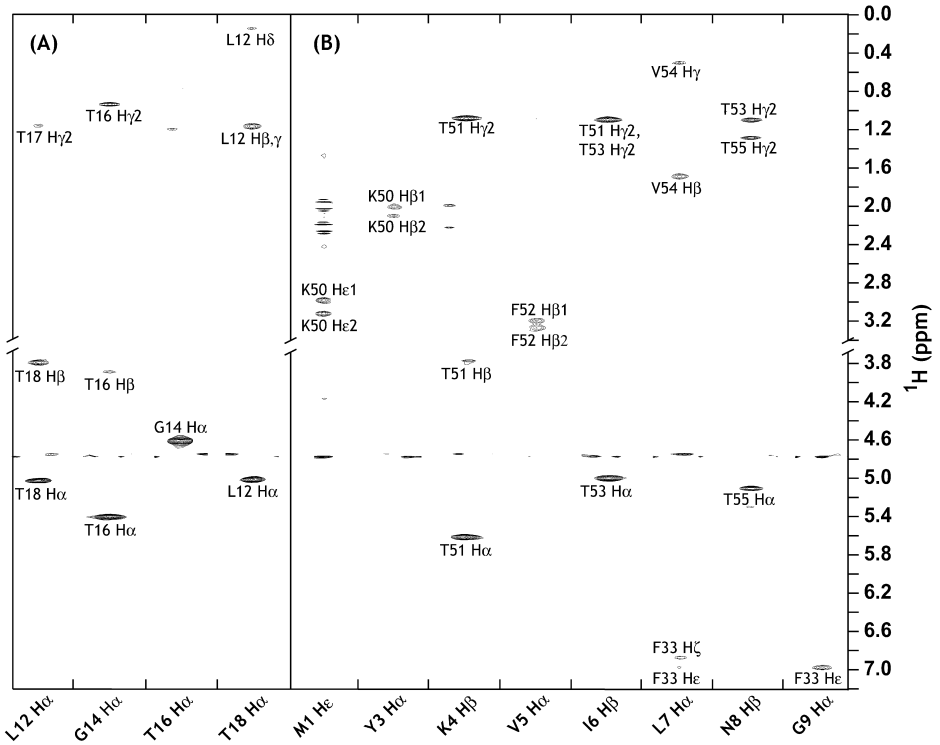

Figure 2. Selected strips from the 3D $^{13}C(f_2)$-edited, $^{12}C(f_3)$-filtered NOE spectrum recorded at 600 MHz on a mixed sample of unlabeled and uniformly $^{15}N,^{13}C$-labeled HS#124^{F26A} GB1 mutant. Intermolecular NOEs between the two β2 strands from each monomer (A) and between β1 from one subunit and β4 from the other subunit (B) are shown.

protein. Similarly, the spectra of HS#124^{V5L}, HS#124^{V30F} and HS#124^{F34A} resembled those of unfolded/aggregated protein and did not contain sharp, resolved resonances that would indicate a well-structured protein. Only HS#124^{F26A} yielded a well-dispersed spectrum, characteristic of a unique and stable architecture for this variant. We therefore selected HS#124^{F26A} (or L5V/F30V/Y33F/A34F-GB1) for structure determination.

Structure determination

Virtually complete ^{1}H, ^{13}C and ^{15}N resonance assignments for the quadruple mutant were obtained using heteronuclear, multidimensional NMR spectroscopy. Initial NMR and nuclear Overhauser effect (NOE) analysis of a uniformly ^{13}C/^{15}N-labeled mutant allowed us to readily identify the α-helix and β-sheet, indicating that the mutant exhibited essentially the same secondary structure as wild-type GB1. Backbone torsion angle prediction using the chemical shifts of the mutant as input for calculating torsion angles with the program TALOS[5] showed good agreement with the angles of the wild-type structure for all but two residues, Val39 and Asp40. This suggested that the regular secondary structure elements (α-helix and β-strands) are conserved in the mutant structure, but that loop structures, especially the loop involving Val39 and Asp40, are different from wild-type GB1. These two residues are located in a linker connecting the α-helix and the second β-hairpin. The linker exhibits an extended conformation in wild-type GB1, whereas in the mutant Val39 and Asp40 are predicted by TALOS to reside in a turn conformation. Furthermore, the β2 strand (Leu12–Thr18) that resides at an edge of the four-stranded β-sheet in the wild-type structure was found to be H-bonded anti-parallel to an additional β-strand. Slowly exchanging amide protons were observed for Lys13, Glu15 and Thr17 in β2, in contrast to the fast exchange behavior for these residues in wild-type GB1. An inter-molecular interface between β2 and the same strand from the other monomer unit was identified and confirmed by NOE analysis of 3D ^{13}C-edited/^{12}C-filtered and ^{13}C-edited/^{12}C,^{14}N-filtered NOESY spectra on a mixed sample of unlabeled and ^{15}N,^{13}C-labeled protein (Figure 2(A)). In addition to this interface, a second dimer interface was found: inter-molecular NOEs were observed between the β1 strand (Gln2-Asn8) of one monomer unit and the β4 strand (Thr51-Thr55) of the other monomer, forming a parallel β-sheet (Figure 2(B)). In wild-type GB1, strands β1 and β4 are the central strands of the four-stranded β-sheet, and their inter-molecular arrangement in the mutant strongly suggested domain swapping.

Structures were calculated with CNS[6] based on 1035 inter-proton distances, 72 hydrogen bond distances and 148 dihedral angles constraints per monomeric subunit. All experimental constraints used in the final structure calculations and pertinent structural statistics are summarized in Table 1. A stereoview of the final 39 conformer ensemble is displayed in Figure 3(A), and ribbon diagrams of a representative structure in two orientations are shown in Figure 3(B). The structures are well defined, satisfy all experimental constraints, display excellent covalent geometry and exhibit atomic r.m.s. deviations of 0.28(±0.05) Å and 0.38(±0.05) Å with respect to the mean coordinate positions for the backbone (N, C^{α}, C') and all

Table 1. Structural statistics

	Ensemble of 39 conformers	Regularized mean structure
r.m.s.d. from experimental constraints		
Distance constraints (Å)[a]		
All (1107)	0.012 ± 0.001	0.010
Intra-monomer (739)	0.012 ± 0.001	0.010
Inter-monomer (296)	0.014 ± 0.001	0.012
Hydrogen bonds (Å)[b]		
Intra-monomer (54)	0.005 ± 0.003	0.007
Inter-monomer (18)	0.011 ± 0.006	0.010
Dihedral angle constraints (deg.) (148)[c]	0.179 ± 0.030	0.220
Residual N–H dipolar couplings (Hz) (56)[d]	1.12 ± 0.11	0.58
r.m.s.d. from idealized covalent geometry		
Bonds (Å)	0.0022 ± 0.0001	0.0020
Angles (deg.)	0.463 ± 0.018	0.463
Impropers (deg.)	0.437 ± 0.029	0.443
Ramachandran plot[e]		
Most favored regions (%)	90.1 ± 1.6	90
Additionally allowed regions (()	9.9 ± 1.6	10
E_{L-J} (kcal/mol)[f]	−377 ± 23	−380
Coordinate precision (Å)[g]		
Backbone atoms	0.28 ± 0.05	–
All heavy atoms	0.67 ± 0.05	–
All heavy atoms excluding disordered side-chains[h]	0.38 ± 0.05	–

The number of constraints per monomeric unit is given in parentheses.

[a] No individual member of the ensemble exhibited distance violations > 0.2 Å.

[b] Hydrogen-bond constraints are for slowly exchanging NH protons only.

[c] Dihedral angle constraints comprise 48 ϕ, 48 ψ, 44 χ_1, and 8 χ_2 angles. None of the structures exhibited dihedral angle violations > 5°.

[d] Residual dipolar coupling constraints involved all the observable backbone NHs (55) and the Trp43 Nϵ^1H.

[e] Calculated by PROCHECK.[33]

[f] E_{L-J} is the Lennard–Jones van der Waals energy calculated with the CHARMM[37] empirical energy function and is not included in the target function for simulated annealing or restrained minimization.

[g] Coordinate precision for the ensemble is defined as the average atomic r.m.s.d. between individual structures and the mean structure.

[h] Excluded disordered surface side-chains are Met1, Asn8 and Glu56 from C^{β} onwards, and Gln2, Lys4, Lys10, Lys13, Glu15, Glu19, Glu27, Lys28, Lys31, Gln32, Glu42 and Lys50 from C^{δ} and onwards.

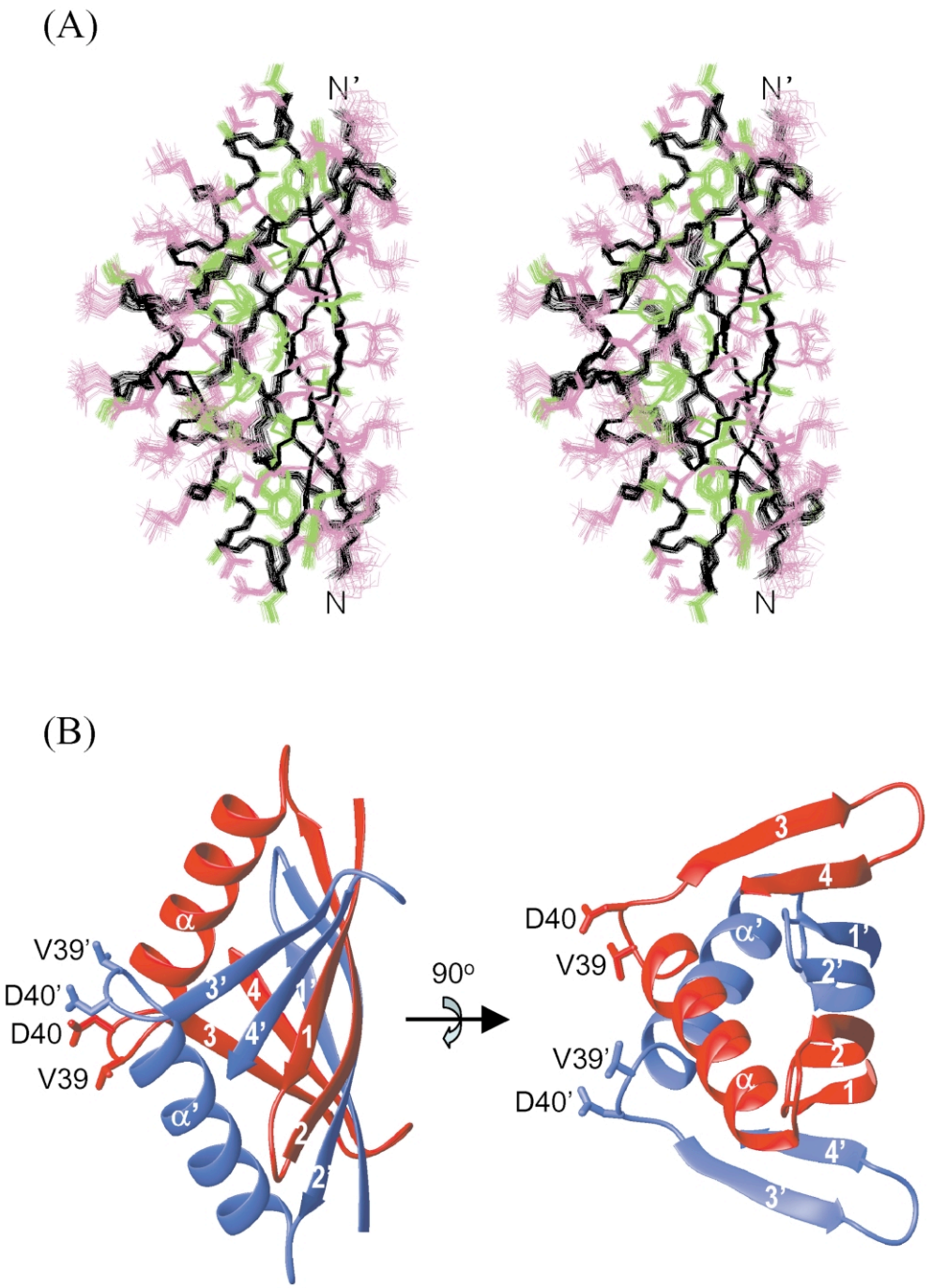

Figure 3. Three-dimensional structure of the HS#124^{F26A} dimer. (A) Stereo view of the best-fit superposition of the final ensemble of 39 NMR structures. Backbone atoms are shown in black and side-chains are colored pink (polar aliphatic) or green (aromatic and non-polar aliphatic). (B) Ribbon diagrams of the regularized mean structure of the domain-swapped dimer in two orientations. Individual monomers are colored red and blue and secondary structural elements are labeled. Side-chains are added for the central hinge–loop residues Val39 and Asp40.

heavy atoms, excluding disordered surface side-chains, respectively. All residues are observed in the favorable region of the Ramachandran plot with >90% of residues in the most favored region. In addition, excellent agreement with the measured residual N–H dipolar couplings was noted with an average violation of 1.14(±0.10) Hz, corresponding to a quality-factor[7] value of 0.061 ± 0.005 and a correlation coefficient of 0.997 ± 0.001. The high quality of the structures reflects the large number of experimental constraints of 25 per residue.

Description of the dimer structure and comparison with monomeric wild-type GB1

A ribbon representation of the structures of the quadruple mutant, with the individual monomer

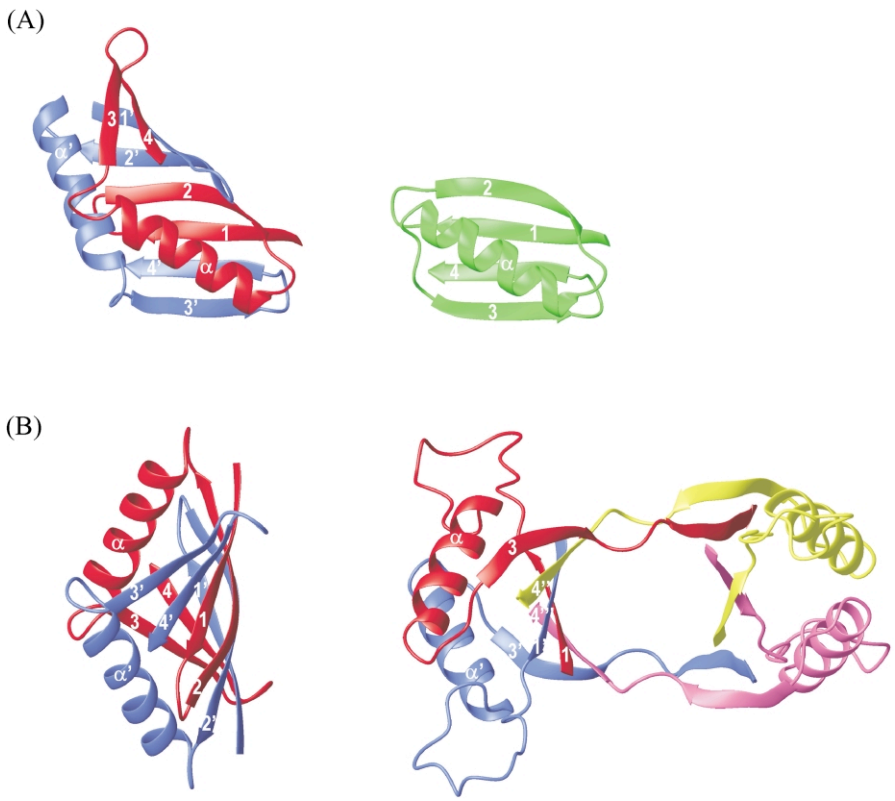

Figure 4. Comparison between the domain-swapped dimeric mutant HS#124^{F26A}, wild-type GB1 (accession code 3GB1) and the HS#124 tetramer (accession code 1MPE). (A) One half of the domain-swapped dimer (blue/red) is depicted in the same orientation as the wild-type GB1 structure (green). (B) Comparison between the domain-swapped dimer and tetramer structures. Each individual polypeptide chain is depicted in a different color and secondary structure elements are labeled.

units colored red and blue, and wild-type GB1, colored green, are displayed in Figure 4(A). The similarity between half of the dimeric mutant and the fold of monomeric wild-type GB1 is apparent, whereas the tetrameric mutant HS#124 exhibits a substantially different fold (Figure 4(B)), with extensive intertwining of all four polypeptide chains.

In the domain-swapped dimeric mutant, exchange of the β3-β4 hairpin between the subunits occurs, resulting in alternation between monomer chains for adjacent hairpins in the eight-stranded β-sheet (Figures 3 and 4). Using the notation introduced by Eisenberg,[8] the "open interface" in the mutant resides between the β2 strands and the "closed interface" is found between β1 and β4. Half of the dimer, composed of the first β-hairpin and the α-helix (residues 1–37) from one polypeptide chain and the second β-hairpin (residues 42–56) from the other chain, is essentially identical to the monomer structure with a pairwise r.m.s.d. value of 1.5 Å for the backbone atoms. In addition to the newly created open interface in the domain-swapped dimer, further inter-molecular contacts exist between the two α-helices and their proceeding loops. The helices are arranged in an anti-parallel manner, crossing at their C termini. This results in a wide V-shape appearance. These V-shaped α helices are buttressing the β-sheet to form a tightly packed hydrophobic core. The only residues for which significant difference in the backbone ϕ and ψ torsion angles between the wild-type and the mutant are observed are those in the linker connecting the α-helix to strand β3. These comprise Gly38, Val39, Asp40 and Gly41, and constitute the hinge–loop in the domain swapping. The ϕ, ψ torsion angles for all four linker residues in both structures are listed in the inset in Figure 5). For all other regions of the polypeptide chain the ϕ,ψ differences lie within the experimental error (±20°). Slightly larger (±50°) deviations are seen for residues 8, 9 and 17, 18 (Figure 5). These residues are located at the end of β1 and β2, respectively, and this region could be influenced structurally by the proximity of the second polypeptide chain in the dimer. It is noteworthy that the central two residues (Val39 and Asp40) in the linker adopt a helical turn structure following the α-helix in the domain-swapped mutant, whereas they exhibit an extended conformation in wild-type GB1 (Figure 5, inset). This change from a somewhat extended conformation in the monomer to a turn in the domain-swapped dimer is unusual for a hinge–loop, since extended conformations in the linker region are observed in many domain-swapped proteins.[8–12] In contrast to most domain-swapped protein structures reported to date, in which the structural elements that are

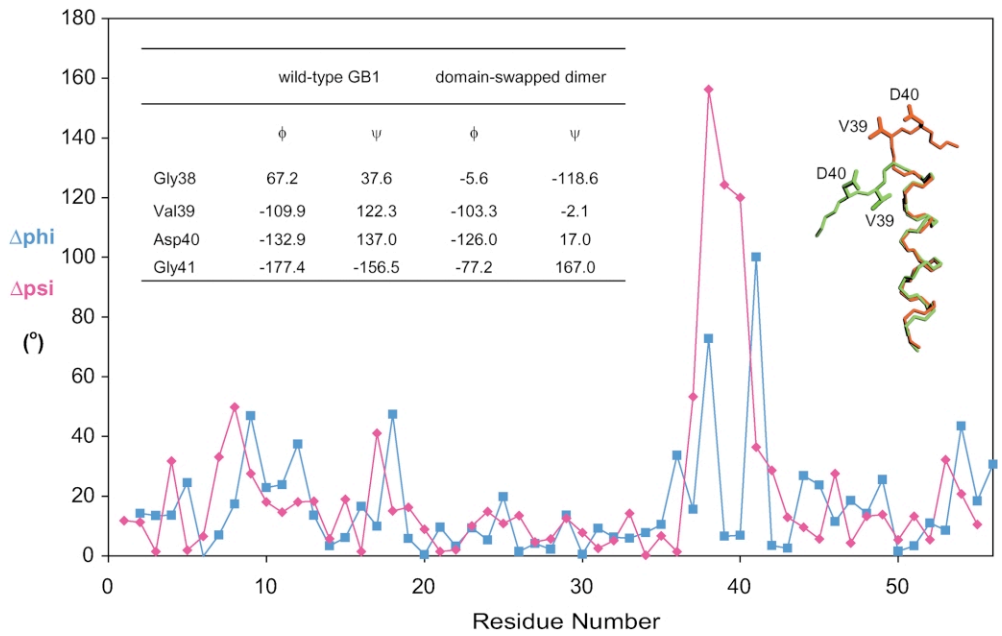

Figure 5. Difference in backbone ϕ (blue) and ψ (pink) torsion angles between the domain-swapped mutant dimer and the wild-type monomer (accession code 3GB1) structures. One insert lists the backbone torsion angles for residues 38–41 in both structures and another displays a backbone comparison between the domain-swapped dimer mutant (red) and monomeric wild-type (green) structures for residues 23–41. For Val39 and Asp40, side-chains are depicted as well.

involved in the swap are single units from either the N or C terminus of the protein,[8] the present structure has essentially half of the molecule swapped. Only very few domain-swapped structures display such extensive inter-change, and for those, the monomers are essentially molecules with internal symmetry. For the present GB1 domain-swapped dimer, the smaller unit is the β3-β4 hairpin, although it may be equally acceptable to define the β1-β2 hairpin plus the helix as the swapped elements.

Comparison of the dimer and tetramer structures

In contrast to the almost identical fold for half of the domain-swapped dimer and the wild-type monomer structures, very limited similarities with the tetramer structure exist. In the tetramer structure, all interactions between β-strands are intermolecular, and each pseudo-dimeric half contains a six-stranded β-sheet. The β-sheet in the domain-swapped dimer is a regular, contiguous eight-stranded sheet, with two hairpin strands alternating between individual polypeptides. Thus, the arrangement of the β-strands in the swapped dimer structure is β3, β4, β1′, β2′, β2, β1, β4′, β3′, while in the tetramer one of the two sheets comprise strands β3, β4″, β1′, β1, β4‴, β3′, with an equivalent arrangement (β3″, β4, β1‴, β1″, β4′, β3‴) in the other half (Figure 4(B)). In addition, a shift in register by two amino acid residues is observed for the β3, β4′ sheet in the tetramer. The pseudo-dimeric half of the tetramer appears less tightly packed than the swapped dimer structure; this is a natural consequence of the loss of two β-stands. The only resemblance that can be gleaned for the two structures pertains to the helix-helix interaction. In both structures, the α-helices interact with their C termini and the subsequent loops, forming a V-shaped arrangement (Figure 4(B)).

Is there a dimer/monomer and dimer/tetramer structure switch?

In order to investigate the structural causes for the switch in quaternary structure from monomer to domain-swapped dimer, a comparison of the wild-type GB1 core with the equivalent portion of the core in the pseudo-monomer half of the dimer was carried out. Figure 6 shows the superposition of these cores and reveals a remarkable difference: in the wild-type monomer structure, the distance between the β-sheet and the α-helix is uniform throughout the entire length of the protein (about 12 Å), whereas in the dimer mutant, a gradual expansion of the core from 12 Å at the N-terminal end to 15 Å at the C-terminal end of the α-helix is observed. Closer inspection points to the importance of the larger side-chain of Phe34, compared to Ala. Phe34 is located at the C-terminal end of the helix and points upwards toward the β-sheet, thereby expanding the core space. This expansion of the core is closely connected to the difference in linker/hinge–loop conformation. Positioning of a bulky side-chain into a monomeric core clearly would impose strain onto the linker. This strain can be released by a local backbone flip, moving

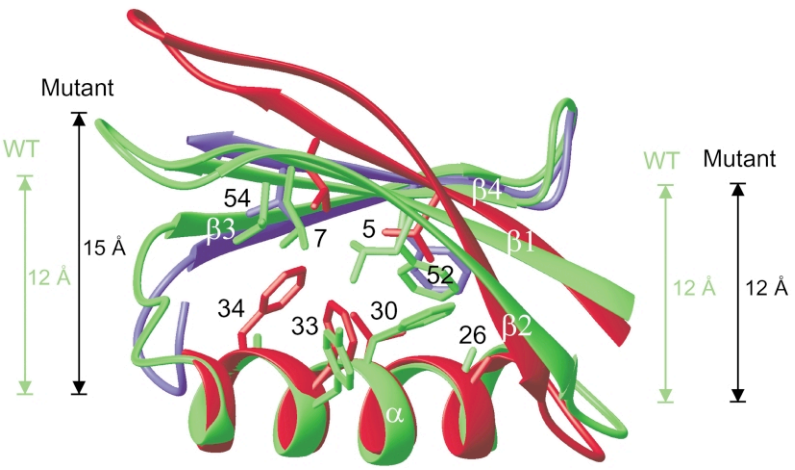

Figure 6. Superposition of the equivalent core structures of the domain-swapped dimer and wild-type GB1 (accession code 3GB1). The superposition involves half of the mutant core structure, composed of the first β-hairpin and the α-helix from one monomer subunit (red) and the second β-hairpin from the other subunit (blue) and the wild-type GB1 core (green). Secondary structure elements and selected side-chains are labeled. The distances between the β-sheet and the α-helix axis at both ends are indicated. The superposition was carried out by best-fitting residues 1–37 from one monomer subunit and residues 43–56 from the other subunit in the mutant and the corresponding residues of wild-type GB1.

the β3-β4 hairpin out of the way, which results in opening of the core, accompanied by a destabilization of the structure. This loss in stability of the individual monomeric units can be compensated for by dimerization. A similar mechanistic explanation was suggested for the tetramerization in HS#124.[4] All other changes appear to have no significant structural effects. Indeed, in a previous set of mutants,[3] the individual L5V and F30V mutations were found to be benign for monomeric GB1. Possible effects of altered side-chains at all four positions in the GB1 wild-type structure were assessed by homology modeling using SCWRL.[13] Three of these resulted in no major energetic penalties. The Ala34 to Phe mutation yielded an energetically unfavorable structure, containing both backbone-backbone and backbone-side-chain clashes, suggesting that A34F may constitute the pivotal change responsible for the change in quaternary structure.

The importance of the individual mutated residues (L5V/F30V/Y33F/A34F) in the integrity of the domain-swapped structure was investigated by modeling and mutagenesis. Inspection of the current dimer structure suggested that the shorter mutant side-chains of the L5V and F30V revertants could be accommodated easily within the core, although possibly causing some destabilization of the structure. The Y33F mutation represents a conservative change and either side-chain should be able to substitute for the other in the respective cores. This is borne out by the fact that the L5V/F30V/A34F triple mutant, containing tyrosine at position 33 as in the wild-type sequence, exists as a domain-swapped dimer. The only difference from the quadruple mutant is a tenfold increase in the dissociation constant, resulting in less dimer at equivalent concentrations (data not shown).

The position of Phe34 in the domain-swapped dimer appeared to be most crucial from a preliminary inspection of the structure. In the Ala34Phe modeled GB1 monomer structure, the phenyl ring clashes with the side-chain of Val39, the central residue in the hinge–loop. This clash, together with the strain in the linker due to the expansion of the core space is predicted to cause severe structural interference. As a consequence, in the domain-swapped dimer structure, the backbone conformation around Val39 is altered, with the side-chain flipped outward and solvent-exposed (see Figures 3 and 5). Calculation of the solvent-accessible surface for Val39 reveals an increase of ~30% from monomer to domain-swapped dimer. In order to test and confirm our hypothesis, we reverted Phe34 in the amino acid sequence of the domain-swapped dimer mutant back to the wild-type Ala residue. As predicted, the resulting mutant protein was monomeric and exhibited a $^{15}N-^{1}H$ HSQC spectrum very similar to that of wild-type GB1. The $^{1}H-^{15}N$ HSQC spectrum for this mutant is shown in Figure 7(B) and, for comparison, the $^{1}H-^{15}N$ HSQC spectrum of HS#124^{F26A} containing both swapped dimer and a monomeric species is displayed alongside. Selected resonances are labeled in both spectra. For the HS#124^{F26AF34A} revertant, the resonances that are particularly indicative of the wild-type GB1 structure are labeled. For instance, note the strikingly different frequencies for E56, L12 and T44 in both spectra.

Taking the above and previous[4] mutational and structural data together, we now clearly have established that a single amino acid change, namely A34F, switches the classical GB1 fold exhibited by the GB1 variant L5V/F30V/Y33F into a domain-swapped dimer, which in turn can be switched into the tetramer structure by an additional single-site mutation, namely A26F.

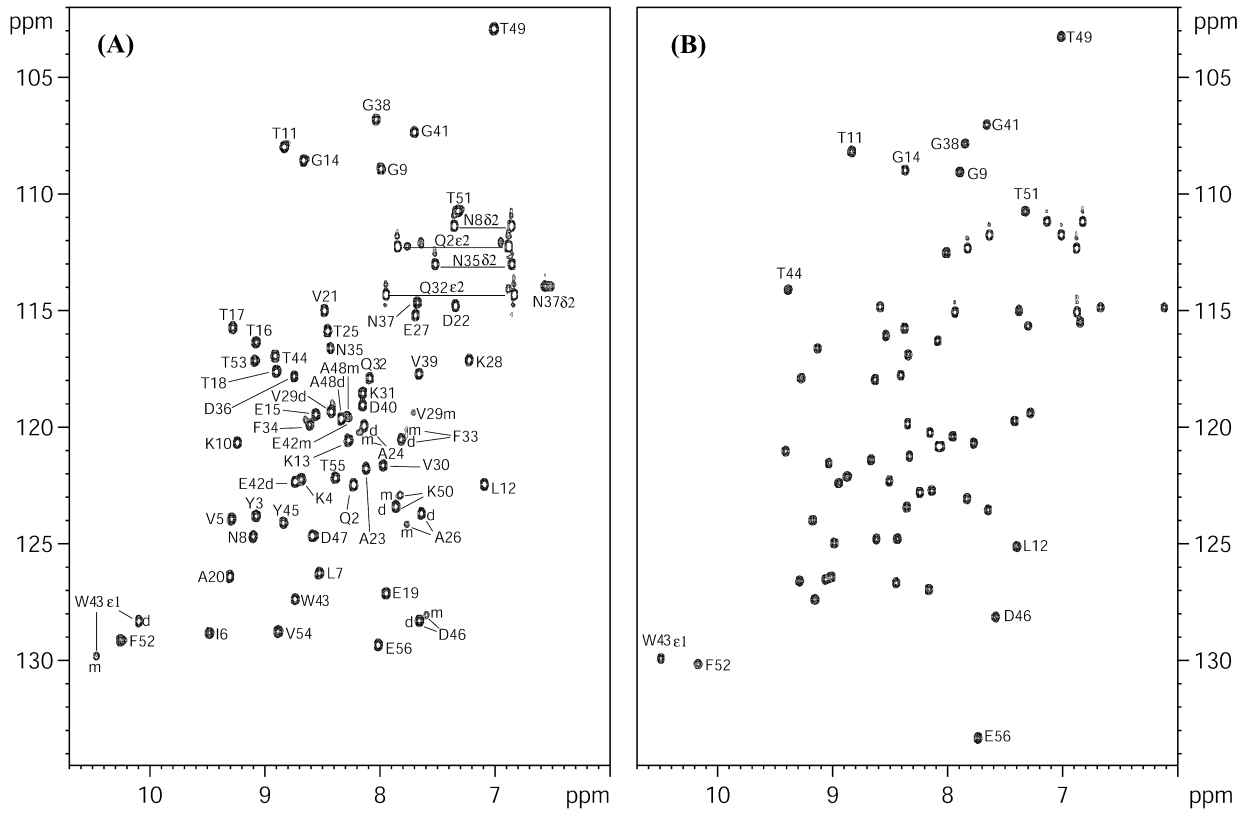

Figure 7. $^{1}H-^{15}N$ HSQC spectra of GB1 variants. (A) Spectrum of HS#124^{F26A} GB1 mutant recorded at 500 MHz and 25 °C. The protein concentration was 1.7 mM in monomer, in 50 mM sodium phosphate buffer (pH 5.5). Assignments are indicated with the residue name and number, and pairs of signals corresponding to side-chain amino groups of asparagines and glutamines are connected by horizontal lines. Cross-peaks arising from the swapped-dimer species are indicated by (d) and those from the monomeric species by (m). (B) Spectrum of the HS#124^{F26AF34A} GB1 mutant recorded at 500 MHz and 25 °C. The spectrum of this mutant is very similar to that of wild-type GB1.

Dimer–monomer equilibrium

In the original $^{15}N-^{1}H$ HSQC spectrum of the domain-swapped dimer mutant, a small number (ca 15% at 1.7 mM; monomer molecular mass; Figure 7(A)) of additional resonances were observed. Upon dilution, these resonances grow in intensity, suggesting that they could arise from a monomeric form. The monomer–dimer equilibrium was confirmed using size-exclusion chromatography in conjunction with in-line multi-angle light-scattering and refractive index detection (Figure 8). Two well-separated peaks are observed, with the first and second peaks corresponding to dimeric and monomeric molecular mass species, respectively. The experimental molecular masses based on the light scattering measurement were 12.0(±0.5) kDa and 6.9(±0.3) kDa, respectively. These values agree well with the predicted molecular mass of 6564 Da (monomer) for a $^{13}C,^{15}N$-labeled mutant. Wild-type GB1 examined in identical fashion as a control, exhibited a single peak with a molecular mass of 6.4(±0.2) kDa, in agreement with the expected molecular mass of 6291 for the ^{15}N-labeled wild-type GB1 (Figure 8). The slightly larger apparent mass for the monomeric species of HS#124^{F26A} compared to wild-type GB1 may reflect the perturbed shape of this species (see below) or residual contributions arising from small amounts of dimer present in the monomeric fraction of this mutant. In addition, it should be noted that the different values are just within the respective errors of the measurements. ^{15}N relaxation studies at 25 °C yielded T_2 values for the Trp43 $N^{\epsilon 1}$ nitrogen atom of ~125 ms and ~170 ms for the major species and the minor species, respectively. These relaxation parameters are consistent with molecular masses of approximately 13 kDa and 6.5 kDa, respectively.

Preliminary NMR analysis of the monomeric species revealed that it is structurally distinct from the compact, wild-type GB1 protein. It appears to exhibit a partially folded structure and most likely corresponds to a stable folding intermediate. The dimerization constant between the partially folded monomer and the domain-swapped dimer was extracted from measurements of peak intensities. The proportions of dimer *versus* monomeric species were determined from the $^{15}N-^{1}H$ HSQC spectra at four different total protein concentrations (16.8 μM, 56 μM, 1.7 mM and 4.3 mM) at 25 °C. The extracted dissociation constants (K_d) and $\Delta G°$ values are essentially identical, with average values of $K_d = 93(\pm 10)$ μM and $\Delta G° =$

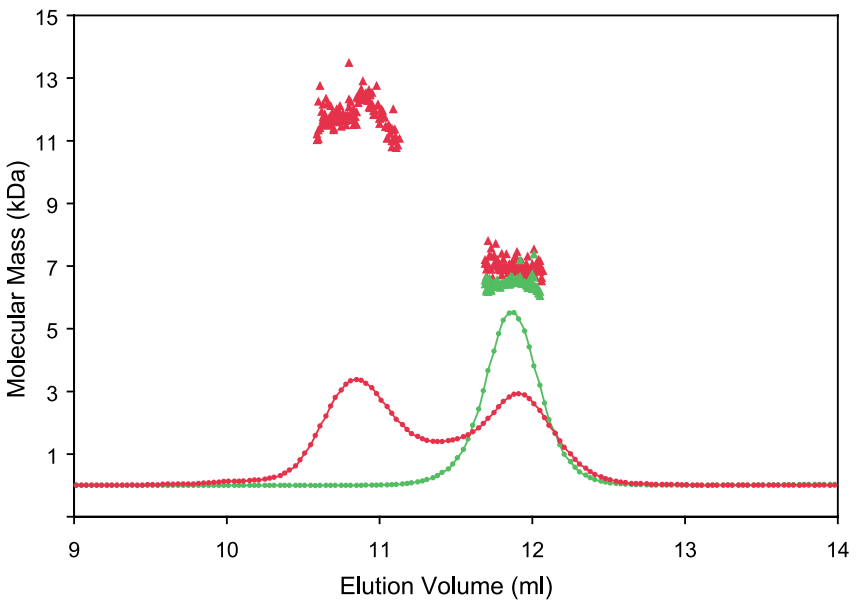

Figure 8. Light-scattering analysis of the HS#124^{F26A} dimer mutant (red) and wild-type GB1 (green). Elution profiles from an analytical S-75 column (1 cm × 30 cm) monitored by refractive index are shown by circles and the predicted molecular masses obtained from the light-scattering measurements are shown by triangles.

5.5(±0.1) kcal mol^{-1}. On the time-scale of the NMR measurements at 25 °C, the equilibrium is established rapidly upon dilution, suggesting that no large kinetic barrier exists between the partially folded monomer and the folded dimer. This is in contrast to the case observed for another domain-swapped dimer investigated in our laboratory, CV-N, which exists as a metastable, kinetically trapped structure at neutral pH, persisting for months at room temperature and below, even upon dilution.[14] Given the fact that the mutant GB1 domain-swapped dimer exists without having a stable, folded monomer, although the analogous, wild-type protein exists as such, the present mutant represents a "quasidomain-swapped" dimer in the notation introduced by Liu & Eisenberg.[8] *Bona fide* domain-swapped dimers exist both, as folded 3D domain-swapped oligomeric structures and in the closed, monomeric conformation, such as cyanovirin-N[14] and cystatin.[12] In these cases, either a kinetically trapped monomer or dimer is observed and it usually requires a long time (up to a couple of months) to reach thermodynamic equilibrium. The fact that we can observe the partially folded, monomeric intermediate by NMR makes its further characterization possible and will allow us to obtain structural data for this otherwise elusive species on the folding landscape.

Concluding Remarks

The current structure represents another example of the amazing versatility exhibited by proteins: the small, highly stable and compact protein GB1 can change topology and quaternary structure simply by a few changes in core residues. Destabilization or opening of the hydrophobic core in monomeric GB1 is compensated for by extending this core *via* dimerization. Interestingly, the present domain-swapped structure is caused by only three (L5V/F30V/A34F) or four (L5V/F30V/Y33F/A34F) core mutations, and a further switch in topology and quaternary structure into a tetramer can be achieved by a single additional amino acid change.[4] Therefore, the single-point mutation A34F can switch the classical monomer GB1 fold into a domain-swapped dimer, and this dimeric structure can be switched further into a tetramer by another single amino acid change of A26F. To our knowledge, this is the only example known for which such drastic structural changes upon single-residue substitutions have been observed. It is hoped that the different structures for these very similar GB1 variants provide interesting test cases for theoretical folding studies.

Materials and Methods

Cloning, expression and purification

Single revertants of the tetrameric HS#124 protein,[4] HS#124^{V5L}, HS#124^{F26A}, HS#124^{V30F}, HS#124^{F33Y} and HS#124^{F34A} were constructed using the HS#124 plasmid DNA, appropriate primers and the QuickChange mutagenesis protocol (Statagene, La Jolla, CA). The HS#124^{F26A} DNA was then used to make the double revertants HS#124$^{F26A/F33Y}$ and HS#124$^{F26A/F34A}$ by the same mutagenesis technique. Purified proteins were verified by mass spectrometry.

Cells were grown at 37 °C either in Luria-Bertani medium or in a modified minimal medium for uniform (>99%) ^{15}N and ^{13}C-labeling with $^{15}NH_4Cl$ and $^{13}C_6$-labeled glucose as the sole nitrogen and carbon sources, respectively. Typically, cells harvested from a 200 ml

culture were suspended in 8-10 ml of PBS (17 mM KH_2PO_4, 50 mM Na_2HPO_4, 1.5 M NaCl, pH 7.4), heated at 80 °C for five minutes, immediately chilled on ice for ten minutes followed by centrifugation at 16,000 rpm (SS-34 rotor) for 30 minutes at 4 °C. The supernatant was passed through a 0.45 μm pore size syringe filter, dialyzed against an excess of deionized water in the cold room, adjusted to 50 mM sodium phosphate (pH 5.5), concentrated using Centriprep YM-3 devices (Millipore corp., Bedford, MA) and loaded onto a Superdex-75 column (2.6 cm × 60 cm; Amersham Biosciences, Piscataway, NJ) equilibrated in 50 mM sodium phosphate buffer (pH 5.5), at a flow-rate of 3 ml/minute at room temperature. Peak fractions were combined and concentrated (Centriprep YM-3), estimated for protein concentration and stored at 4 °C until further use or subjected to an additional reverse-phase HPLC purification step for unlabeled mutant proteins. Unlabeled mutant proteins were bound to POROS R2 resin and eluted using a linear gradient of aqueous 0–60% (v/v) acetonitrile, 0.05% (v/v) trifluoroacetic acid over a period of 16 minutes at a flow-rate of 4 ml/minute. Peak fractions were combined and dialyzed excessively against 50 mM sodium phosphate buffer (pH 5.5), 0.02% (w/v) NaN_3, concentrated and the protein concentration was determined spectrophotometrically using an extinction coefficient of 8250 M^{-1} cm^{-1} at 280 nm.

NMR spectroscopy and structure determination

All spectra for structure determination were recorded at 25 °C on NMR samples containing ca 1.7 mM protein (monomer molecular mass), 50 mM sodium phosphate buffer (pH 5.5) using Bruker DRX800, DMX750, DRX600, DMX600 and DMX500 spectrometers, equipped with 5 mm, triple resonance, three axes gradient probes or *z*-axis gradient cryoprobes. Spectra were processed with NMRPipe[15] and analyzed using NMRDraw and PIPP, CAPP and STAPP.[16] ^{1}H, ^{13}C and ^{15}N sequential assignments and NOE data analysis were carried out using heteronuclear multi-dimensional experiments routinely used in our laboratory.[17–19] 3D HNCACB,[20] CBCA(CO)NH,[21] H(CCO)HN-TOCSY and C(CCO)NH-TOCSY[22,23] experiments were employed for sequential assignments. NOE spectra to derive inter-proton distance constraints included 2D NOESY and ^{12}C,^{14}N-filtered, and 3D ^{13}C-edited, ^{15}N-edited, ^{13}C-edited/^{12}C-filtered, and ^{13}C-edited/^{12}C,^{14}N-filtered NOE experiments,[24,25] recorded using mixing times ranging from 80 to 100 ms. Approximate inter-proton distance constraints were grouped into four distance ranges: 1.8–2.7 Å, 1.8–3.3 Å, 1.8–5.0 Å and 1.8–6.0 Å, corresponding to strong, medium, weak and very weak NOEs, respectively. In addition, 0.5 Å was added to the upper limit of inter-proton distance constraints involving methyl groups. For hydrogen bonds, distance constraints of 1.5–2.5 Å (H–O) and 2.5–3.5 Å (N–O) were employed. Torsion angle constraints were obtained by measuring $^3J_{N-H\beta}$, $^3J_{H\alpha-H\beta}$, $^3J_{N-C\gamma}$, $^3J_{C'-C\gamma}$ and $^3J_{C\alpha-C\delta}$ couplings using quantitative *J* correlation spectroscopy.[26–28] TALOS[5] was used for extracting backbone torsion angle (ϕ,ψ) constraints. Residual HN dipolar couplings were measured using in-phase/anti-phase ^{1}H–^{15}N HSQC experiments[29] on protein samples in magnetically oriented (using 17 mg/ml of pf1 phage solution) and isotropic media (without phage). The measured residual HN dipolar couplings ranged from −16 Hz to 37 Hz. Slowly exchanging amide protons were detected in the ^{1}H–^{15}N HSQC spectrum of lyophilized protein freshly dissolved in ^2H_2O at 15 °C. All experimentally determined distance, torsion angle and residual dipolar coupling constraints (Table 1) were applied in a simulated annealing protocol using the program CNS.[6] The equivalence of the monomeric subunits was ensured using non-crystallographic symmetry constraints by implementing three axes of second order for the dimer *via* a special symmetry NOE class.[30,31] An ensemble of 100 structures was generated and the final family of 39 structures, the major structural family with lowest CNS target function values, was selected. The regularized mean structure was calculated by restrained energy minimization of the geometric mean of the final 39 structures. Quality factor values for the residual N–H dipolar couplings were calculated using PALES.[32] The quality of the structures was analyzed by CNS,[6] PROCHECK,[33] and MOLMOL.[34] All structure Figures were generated with MOLMOL.[34]

Determination of the dissociation constant of the dimeric GB1 mutant

The relative amounts of dimeric and monomeric species were determined from ^{15}N–^{1}H HSQC peak intensities. Volume or peak height measurements yielded essentially the same values (±5% variation). The Trp43 N$^{\epsilon1}$H and Ala 26 backbone NH resonances that exist as well resolved, isolated peaks in the domain-swapped dimer and the partially folded monomer were used for quantification. Final fractions were calculated by averaging values from both resonances using the both determinations (volume and height) for four different protein concentrations (17 μM, 56 μM, 1.70 mM and 4.28 mM). The dissociation constant (K_d) and free energy of dissociation (ΔG°) were calculated using the following equations:

$$D \leftrightarrow 2M$$

$$K_d = [M]^2/[D] = 2P_t f_M^2/f_D = 2P_t f_M^2/(1 - f_M)$$

$$\Delta G^\circ = -RT \ln K_d$$

where [M] and [D] are the molar concentrations of the partially folded monomer and domain-swapped dimer, respectively, P_t is the total protein concentration (monomer molecular mass), f_M and f_D are the measured fractions of monomer and dimer, respectively, R is the gas constant (1.987 cal deg^{-1}·mol^{-1}), and T is the absolute temperature. ΔG° was calculated by normalizing all reactants to the standard concentration of 1 M before applying the logarithm.[35,36]

Size-exclusion chromatography and multi-angle light-scattering

Light-scattering data were obtained using an analytical Superdex-75 column (1.0 cm × 30 cm; Amersham Biosciences, Piscataway, NJ) with in-line multi-angle light-scattering (DAWN EOS, Wyatt Technology, Inc.) and refractive index detectors (OPTILAB DSP, Wyatt Technology, Inc., Santa Barbara, CA). Protein (90 μg at a concentration of ~0.18 mM) was applied to the pre-equilibrated S75 column at a flow-rate of 0.5 ml/minute at room temperature and eluted with 20 mM sodium phosphate buffer (pH 5.5) containing 0.02% (w/v) sodium azide.

Protein Data Bank accession numbers

The atomic coordinates of the 39 conformers of the NMR structure, the corresponding regularized mean structure, together with the experimental distance and angle constrains, have been deposited in the RCSB Protein Data Bank,† accession numbers 1QIO. NMR assignments for the HS#124^{F26A} mutant of GB1 have been deposited at the Biomolecular Magnetic Resonance Databank,‡ access code 5875.

Acknowledgements

We thank Drs D. Garrett and F. Delaglio for software, and J. Baber for technical support. This work was supported, in part, by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the National Institute of Health (to A.M.G.).

References

1. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991). A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science*, **253**, 657–661.
2. Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. (1994). Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry*, **33**, 4721–4729.
3. Gronenborn, A. M., Frank, M. K. & Clore, G. M. (1996). Core mutants of the immunoglobulin binding domain of streptococcal protein G: stability and structural integrity. *FEBS Letters*, **398**, 312–316.
4. Frank, M. K., Dyda, F., Dobrodumov, A. & Gronenborn, A. M. (2002). Core mutations switch monomeric protein GB1 into an intertwined tetramer. *Nature Struct. Biol.* **9**, 877–885.
5. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
6. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
7. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837.
8. Liu, Y. & Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285–1299.
9. 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.
10. Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M. *et al.* (1999). Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J. Mol. Biol.* **288**, 403–412.
11. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M. & Jaskolski, M. (2001). Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nature Struct. Biol.* **8**, 316–320.
12. Staniforth, R. A., Giannini, S., Higgins, L. D., Conroy, M. J., Hounslow, A. M., Jerala, R. *et al.* (2001). Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J.* **20**, 4774–4781.
13. Dunbrack, R. L. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681.
14. Barrientos, L. G., Louis, J. M., Botos, I., Mori, T., Han, Z. Z., O'Keefe, B. R. *et al.* (2002). The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure*, **10**, 673–686.
15. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). Nmrpipe—a multidimensional spectral processing system based on Unix Pipes. *J. Biomol. NMR*, **6**, 277–293.
16. Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A Common-sense approach to peak picking in 2-dimensional, 3-dimensional, and 4-dimensional spectra using automatic computer-analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220.
17. Clore, G. M. & Gronenborn, A. M. (1991). 2-dimensional, 3-dimensional, and 4-dimensional NMR methods for obtaining larger and more precise 3-dimensional structures of proteins in solution. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 29–63.
18. Clore, G. M. & Gronenborn, A. M. (1998). Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.* **16**, 22–34.
19. Bax, A. & Grzesiek, S. (1993). Methodological Advances in Protein NMR. *Accts Chem. Res.* **26**, 131–138.
20. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *J. Magn. Reson. ser. B*, **101**, 201–205.
21. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293.
22. Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1993). A general-method for assigning NMR-spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. Biomol. NMR*, **3**, 225–231.
23. Lin, Y. X. & Wagner, G. (1999). Efficient side-chain and backbone assignment in large proteins: application to tGCN5. *J. Biomol. NMR*, **15**, 227–239.
24. Ikura, M. & Bax, A. (1992). Isotope-filtered 2D NMR of a protein peptide complex-study of a skeletal-muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.* **114**, 2433–2440.
25. Lee, W., Revington, M. J., Arrowsmith, C. & Kay, L. E. (1994). A pulsed-field gradient isotope-filtered 3D C-13 HMQC-NOESY experiment for extracting

† www.rcsb.org
‡ www.bmrb.wisc.edu

intermolecular NOE contacts in molecular-complexes. *FEBS Letters*, **350**, 87–90.
26. Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. & Zhu, G. (1994). Measurement of homonuclear and heteronuclear J-couplings from quantitative J-correlation. *Nucl. Magn. Reson. Pt C*, **239**, 79–105.
27. Grzesiek, S., Kuboniwa, H., Hinck, A. P. & Bax, A. (1995). Multiple-quantum line narrowing for measurement of H-alpha-H-beta J-couplings in isotopically enriched proteins. *J. Am. Chem. Soc.* **117**, 5312–5315.
28. Hu, J. S., Grzesiek, S. & Bax, A. (1997). Two-dimensional NMR methods for determining (chi 1) angles of aromatic residues in proteins from three-bond J(C′C gamma) and J(NC gamma) couplings. *J. Am. Chem. Soc.* **119**, 1803–1804.
29. Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378.
30. ODonoghue, S. I., King, G. F. & Nilges, M. (1996). Calculation of symmetric multimer structures from NMR data using a priori knowledge of the monomer structure, co-monomer restraints, and interface mapping: the case of leucine zippers. *J. Biomol. NMR*, **8**, 193–206.
31. Nilges, M. (1993). A calculation strategy for the structure determination of symmetrical dimers by H-1-NMR. *Proteins: Struct. Funct. Genet.* **17**, 297–309.
32. Zweckstetter, M. & 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.
33. 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.
34. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.
35. Backes, H., Berens, C., Helbl, V., Walter, S., Schmid, F. X. & Hillen, W. (1997). Combinations of the alpha-helix-turn-alpha-helix motif of TetR with respective residues from LacI or 434Cro: DNA recognition, inducer binding, and urea-dependent denaturation. *Biochemistry*, **36**, 5311–5322.
36. Dams, T. & Jaenicke, R. (1999). Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemistry*, **38**, 9169–9178.
37. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Edited by M. F. Summers

(Received 20 May 2003; received in revised form 11 July 2003; accepted 11 July 2003)