Published in final edited form as:

J Mol Biol. 2009 November 27; 394(2): 343–350. doi:10.1016/j.jmb.2009.09.005.

Non-sequence-specific interactions can account for the compaction of proteins unfolded under “native” conditions

Jonathan E. Kohn¹, Blake Gillespie², and Kevin W. Plaxco^{3,*}

¹Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720

²Department of Chemistry, California State University, Channel Islands, One University Drive, Camarillo, CA 93012

³Department of Chemistry and Biochemistry, University of California, Santa Barbara, Santa Barbara, CA 93106

Abstract

Proteins unfolded by high concentrations of chemical denaturants adopt expanded, largely structure-free ensembles of conformations well approximated as random coils. In contrast, globular proteins unfolded under less denaturing conditions (via mutations, or transiently unfolded after a rapid jump to native conditions) and molten globules (arising due to mutations or co-solvents) are often compact. Here we explore the origins of this compaction using a truncated, equilibrium-unfolded variant of the 57-residue FynSH3 domain. As monitored by far-UV circular dichroism, NMR spectroscopy and hydrogen exchange kinetics, a four-residue carboxy-terminal deletion of FynSH3 (CΔ4) appears to be largely unfolded even in the absence of denaturant. Nevertheless, CΔ4 is quite compact under these conditions, with a hydrodynamic radius only slightly larger than that of the native protein. In order to understand the origins of this molten globule-like compaction we have characterized a random sequence polypeptide of identical amino acid composition to CΔ4. Notably, we find that the hydrodynamic radius of this random sequence polypeptide also approaches that of the native protein. Thus, while native-like interactions may contribute to the formation of compact “unfolded” states, it appears that non-sequence-specific monomer-monomer interactions can also account for the dramatic compaction observed for molten globules and the “physiological” unfolded state.

Keywords

Natively unfolded; Guanidine Hydrochloride; Folding kinetics; Residual Structure

Introduction

Due to the relative ease with which they are studied, chemically denatured proteins have remained the “gold standard” unfolded state for more than 60 years (1–3). Spectroscopic and small angle scattering studies of such proteins have generally upheld the longstanding view that the chemically-denatured state occupies an ensemble of highly expanded, effectively

© 2009 Elsevier Ltd. All rights reserved.

* Author to whom correspondence should be addressed, Phone: (805) 893-5558, Fax: (805) 893-4120, kwp@chem.ucsb.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

random coil conformations (4, 5). Consequently, the random-coil model remains the starting point for many theoretical models of the folding process (*e.g.*, 6, 7).

In vivo, however, proteins fold in the absence of denaturant and thus the behavior of proteins unfolded under non-denaturing conditions is pertinent to our understanding of folding in the cell. Previous studies of globular proteins unfolded by mutation indicate that the unfolded states populated under more physiological conditions are generally compact (8, 9). For example, unfolded forms of λ -repressor, staphylococcal nuclease and drkSH3 populated in the absence of denaturant are expanded only 8%, 30% and 40% relative to their respective folded forms, markedly less than the ~2-fold expansions observed for proteins of similar size under more highly denaturing conditions (8, 10, 11). Consistent with this, some (12–14), though not all (15, 16), proteins undergo transient compaction prior to folding. This compaction, and other compelling spectroscopic evidence for native-like order in proteins unfolded in the absence of denaturant (5), suggests that the physiological unfolded state may contain significant residual structure (17–20). If present such structure could restrict a newly synthesized protein to a native-like topology from which it is preconfigured to fold rapidly (17). The extent to which the collapse of unfolded proteins under physiological conditions is driven by native-like interactions, however, has seen only limited investigation (21–23) and remains controversial (24–26).

Here we explore the relationship between the chemically denatured state and the more compact unfolded state populated in the absence of denaturants. We do so by comparing the spectroscopic features and molecular dimensions of guanidine hydrochloride-(GuHCl)-unfolded FynSH3 to a variant unfolded by mutation. We also compare these states with the properties of a random sequence polypeptide of identical amino acid composition as our unfolded FynSH3 variant. These comparisons allow us to investigate the extent to which generic heteropolymer behavior, that is composition-driven interactions, is sufficient to account for the compaction of proteins unfolded under near-physiological conditions.

Results

In order to investigate the origins of the observed compaction of proteins under physiological conditions it was first necessary to produce a variant protein that is unfolded at room temperature in the absence of denaturants. Scattered but abundant literature reports suggest that terminal deletion mutations of single domain proteins produce unfolded states after removal of several residues from either terminus (27–30). Fyn-SH3, a physically and biologically well characterized, single domain protein of known structure was selected as our test system (31) and, based on the aforementioned studies, subjected to sequential carboxy-terminal deletions. Folding free-energies were determined for each variant (data not shown) in order to identify the least truncated variant, C Δ 4 (Fig. 1), that lacks a clear denaturant unfolding transition.

The spectral features of C Δ 4 suggest this variant is significantly unfolded at room temperature in the absence of denaturant. For example, while distinct from the circular dichroism (CD) spectrum of the chemically denatured protein, the CD spectrum of C Δ 4 differs significantly from that of the native SH3 fold (Fig. 2). While the unusual CD signature of the SH3 fold renders it difficult to interpret these differences in terms of secondary structure content, the diminished ellipticity in the far-UV clearly indicates a significant loss in structure; in particular, the large, positive peak at 220 nm that presumably arises due to tertiary interactions is almost entirely abolished. Similarly, the 1D NMR spectrum of C Δ 4 exhibits a significant loss of peak dispersion, including the effectively complete loss of a highly shifted methyl resonance (Fig. 3) at -0.55 ppm that is characteristic of native SH3 domains (32). The ^1H , ^{15}N HSQC spectrum of C Δ 4 exhibit

similarly limited spectral dispersion, further confirming the largely unfolded character of this variant (Fig. 4). Finally, the amide-protons of CΔ4 fully exchange within the 7 minute dead time of a manual solvent exchange experiment, which is at least an order of magnitude more rapid than the exchange rate of the native protein (Fig. 5) and suggests that, even in the absence of denaturant, the fixed secondary structure content of CΔ4 is significantly diminished. This said, readily apparent line broadening in its 1D NMR spectra, non-trivial dispersion in its ^1H , ^{15}N HSQC spectra and experimentally-significant ellipticity differences from the chemically denatured wild-type protein suggest that, rather than being fully unfolded, CΔ4 resides somewhere towards the unfolded end of the molten globule spectrum.

Despite the above evidence that CΔ4 is largely unfolded in the absence of denaturant, the variant protein is quite compact under these conditions. Specifically, the hydrodynamic radius (R_h) of CΔ4 is $19.4 \pm 0.3 \text{ \AA}$ (95% confidence intervals), which is within error of the $18.4 \pm 1.0 \text{ \AA}$ R_h measured for the full-length protein (Fig. 6; Table 1). We note, however, that both the full-length FynSH3 and CΔ4 constructs contain highly hydrophilic, unstructured 21-residue amino-terminal tails (retained to facilitate purification and handling) and, because of this, our measurements presumably underestimate the true relative expansion of CΔ4. (Given that this tail is extremely hydrophilic, it appears a reasonable assumption that it does not interact with the collapsed protein and remains a random coil.) The observed compaction of CΔ4 thus appears consistent with the 10 – 30% expansion previously reported for molten globules (9). In the presence of 6M GuHCl the R_h of CΔ4 increases to $23.0 \pm 1.0 \text{ \AA}$ (Fig. 6), which is slightly less than the $25.2 \pm 0.3 \text{ \AA}$ observed for the (four-residue longer) full-length protein under the same conditions (Table 1). These values correspond closely to the R_h predicted for chemically denatured proteins of the same lengths (33).

The observed compaction of CΔ4 and other “physiologically unfolded” and molten globular proteins could arise as a consequence of native-like interactions. Alternatively, the observed collapse could reflect a less specific, Flory-like coil-to-globule transition driven by non-sequence-specific interactions (34, 35). In order to discriminate between these possibilities we have characterized a random-sequence polypeptide (RS) of identical amino acid composition to CΔ4 under the assumption that, whereas native-like interactions are sequence specific a Flory-type collapse will depend only on sequence composition. The circular dichroism spectrum of the randomized RS protein in aqueous buffer is closely similar to that of FynSH3 in 6M GuHCl (Fig. 2) and its 1D NMR spectrum shows little chemical shift dispersion (Fig. 3). However, despite the strong spectroscopic evidence that RS is unfolded under these conditions its R_h also approaches that of the native protein (Table 1). In 6M GuHCl, in contrast, its R_h expands to $24.4 \pm 1.8 \text{ \AA}$, which is effectively indistinguishable from that of the naturally-occurring sequence under the same conditions (Table 1).

Discussion

CΔ4 appears effectively unfolded as monitored by CD, NMR and hydrogen exchange kinetics and yet its dimensions are close to those of the native protein. Similar, albeit somewhat less dramatic compaction has been observed in the unfolded states of λ -repressor, staphylococcal nuclease and drkSH3 that are populated in the absence of denaturant (8, 10, 11), suggesting that the effect is relatively common. Such compaction could be the result of native-like interactions leading to a contraction of CΔ4 relative to the random-coil configuration. A random sequence polypeptide of identical amino acid composition, however, is similarly compact despite exhibiting even less spectroscopic evidence of structure.

CΔ4 might best be thought of as a molten globule. For example, while CΔ4's NMR spectra and hydrogen exchange kinetics strongly argue that it lacks fixed tertiary structure, the protein's NMR spectra exhibit line-broadening characteristic of molten globules (9). Consistent with this, the observed compaction of CΔ4 is similar to that of other molten globules, which typically exhibit molecular dimensions quite close to those of the native state (36). Examples include the molten globule states of α-lactalbumin, for which the radius of gyration, R_g , is expanded only 10% relative to the native protein (36), and apomyoblobin, with an R_g expanded 30% over that of the native protein (37). Nevertheless, it is perhaps surprising that CΔ4 should populate a molten globule given that the "structured" regions of this protein span only 53 residues. We are not aware of any other protein this small for which a stable molten-globule state has been reported.

In contrast to CΔ4, the random sequence variant RS does not exhibit the characteristic spectral fingerprints of a molten globule and appears to be fully unfolded even in the absence of denaturant. Under these same conditions, however, the R_h of RS closely approaches that of both CΔ4 and the native protein. This observation is consistent with the work of Urabe and coworkers, who report that a largely random, 141-residue sequence (selected for solubility rather than any attribute more specifically associated with "foldedness" and not corresponding to any naturally occurring protein) also exhibits molten globule-like compaction (38). It is also consistent with the work of Schuler and co-workers, who have measured pairwise distances across the protein cspB unfolded at very low denaturant and find the monotonic relationship between through-space distance and sequence separation expected for a randomly collapsed state (23). Finally, all of these observations are consistent with classic heteropolymer theory (35), which predicts that a polypeptide chain will undergo a coil-to-globule transition whenever monomer-monomer interactions are more favorable than those between monomers and solvent (as is presumably the case under physiological conditions) (24, 35, 39). Such non-sequence-specific, composition-dependant interactions have been invoked to explain the rapid compaction observed during the early stages of folding for RNase A and cytochrome c (24, 25, 40). The observed compaction of RS is consistent with these claims, as are simulations that predict similar sequence-independent compaction for randomized variants of ubiquitin and λ-repressor (41). Because meaningful native-like interactions are presumably sequence-specific (rather than simply composition-dependant), the compaction observed for random sequences such as these demonstrates that molten globule dimensions can be achieved without the formation of sequence-specific, native-like interactions.

The results presented here do not contradict claims that native-like interactions drive the compaction of non-native states. That is, while we have shown that sequence-independent interactions are sufficient to account for the observed compaction, native-like interactions may also be playing a role in the compaction of CΔ4 and other compact non-native states. Moreover, even if compaction is dominated by sequence-independent interactions it is important to note that such interactions may nevertheless accelerate folding by reducing the conformational search to more compact states (42, 43). Thus the formation of the even non-specific compact states may play an important role in protein folding.

Methods

Truncation variants

The coding sequence for full-length Fyn-SH3 was directionally cloned into a pET-15b (Novagen, Inc.) expression vector introducing an amino-terminal 6xHis tag to facilitate purification. QuikChange PCR (Stratagene, Inc.) was used to generate single amino acid truncations from the carboxy-terminus through successive introductions of a stop-codon. The identity of each truncation clone was then confirmed by DNA sequencing (data not

shown). Mutant plasmids were transformed into BL21(DE3) pLysS (Stratagene, Inc.) and expressed by induction with isopropyl- β -D-thiogalactopyranoside. The protein was purified by Ni-NTA agarose (Qiagen, Inc.) affinity chromatography in 50mM potassium phosphate buffer, pH 7 (termed “aqueous buffer” below) plus 2M GuHCl, which was added in order to minimize loss from aggregation. The final yield of C Δ 4 was approximately 1mg/l.

Design and preparation of the random sequence protein

The random sequence was designed by sequentially numbering each codon of the C Δ 4 gene sequence and then using a random number generator to create five unique, randomized sequences. Each of the five sequences was analyzed using the ProtParam Tool (Swiss Institute of Bioinformatics) to assess expressability in *E. coli* (44). The sequence determined to be the most “expressible,” termed RS (Fig. 1), was then optimized for codon usage (45) and the corresponding synthetic gene was synthesized (Genescript, Inc.) and subcloned into pET-15b (Novagen, Inc.). Expression, purification and approximate yield were as for C Δ 4.

Circular dichroism

CD spectra from 200–270nm were collected on an AVIV model 202 CD-spectrometer (AVIV, Inc.) for full-length FynSH3 in aqueous buffer plus 6M GuHCl, and for C Δ 4 and RS in aqueous buffer (Fig. 2). Sample concentrations were 10 μ M as determined by UV-absorbance (assuming $\epsilon = 16500 \text{ M}^{-1} \text{ cm}^{-1}$).

1D NMR

^1H NMR spectra were collected for wild-type Fyn-SH3, C Δ 4 and RS in aqueous buffer and for wild-type Fyn-SH3 in aqueous buffer plus 6M GuHCl (Fig. 3). Spectra were collected in 10% D $_2$ O at 25°C on a Varian 600 MHz spectrometer (Varian, Inc.) over 128 scans of 1024 points.

HSQC

^{15}N HSQC spectra were collected for C Δ 4 in aqueous buffer, and wild-type FynSH3 in aqueous buffer and in aqueous buffer plus 6M GuHCl (Fig. 4). ^{15}N labeled protein was produced by growing FynSH3 and C Δ 4 expressing BL21(DE3) pLysS (Stratagene, Inc.) *E. coli* in M9 minimal media enriched with 1g/l ^{15}N ammonium chloride (Cambridge Isotopes Laboratory, Inc.). The labeled proteins were then purified as described above. Spectra were collected in 10% D $_2$ O at 25°C on a Varian 600 MHz spectrometer (Varian, Inc.) over 8 scans, 128 increments.

Hydrogen exchange

^1H NMR spectra were collected for C Δ 4 and wild-type Fyn-SH3 in 10% D $_2$ O, 50mM potassium phosphate (pH 7). Protein samples were then rapidly exchanged into 100% D $_2$ O, 50mM potassium phosphate (pD 7) by sephadex-G25 (Sigma, Inc.) spin-column gel-filtration (7 minute dead-time between exchange and collection of first spectrum). Spectra were collected at 25°C on a Varian 600 MHz spectrometer (Varian, Inc.) over 64 scans of 9600 points.

Pulsed-field Gradient NMR

Hydrodynamic radii (R_h) for wild-type Fyn-SH3, C Δ 4 and RS were measured in aqueous buffer and in aqueous buffer plus GuHCl (Fig. 6, Table 1) using pulsed-field gradient NMR using the watersLED pulse sequence (46). Experiments were conducted in 100% D $_2$ O at 25°C on a Varian 600 MHz spectrometer (Varian, Inc.) over 32 scans of 9600 points. (Of note, these experiments were conducted in 100 D $_2$ O in order to minimize interference from

signals arising due to water. Control experiments conducted in protonated buffer in the absence of added denaturant, however, produce an R_h for CΔ4 that is effectively indistinguishable from that observed in 100% D₂O, indicating that deuteration of the solvent is not responsible for the observed compaction –data not shown.) Signal attenuation of the protein and the reference molecule (dioxane at 1mM) were measured as a function of field gradient strength. Hydrodynamic radii were calculated by fitting the relative decay rates of the sample molecule and the dioxane reference to the relationship (33):

$$I = \exp[-(\gamma g \delta)^2 (\Delta - \delta/3) D] \quad (1)$$

Where I is the peak height, γ is the proton gyromagnetic ratio, g is gradient strength, δ and Δ are delays and D is the diffusion constant of the molecule, which is the only fitted parameter. Using the latter parameter we obtain the hydrodynamic radius, $R_{h,}$ of the molecule under investigation using the relationship:

$$R_h = D_{ref} / D_{prot} (R_{h,ref}) \quad (2)$$

where D_{ref} and D_{prot} are the diffusion constants of the reference molecule and the protein respectively, and $R_{h,ref}$ is the known hydrodynamic radius of dioxane.

Acknowledgments

The authors wish to thank Hongjun Zhou for assistance with NMR spectroscopy. This work was supported by NIH grant R01GM062958-01.

Abbreviations

CΔ4	4-residue carboxy-terminal deletion variants of FynSH3
R_h	hydrodynamic radius
GuHCl	guanidine hydrochloride
RS	random sequence polypeptide of identical amino acid composition to CΔ4

References

1. Mirsky AE, Pauling L. On the Structure of Native, Denatured, and Coagulated Proteins. *Proc Natl Acad Sci USA*. 1936; 22:439–447. [PubMed: 16577722]
2. Tanford C, Kawahara K, Lapanje S. Proteins in 6M Guanidine Hydrochloride- Demonstration of Random Coil Behavior. *J Biol Chem*. 1966; 241:1921–1923. [PubMed: 5947952]
3. Millett, I.; Doniach, S.; Plaxco, KW. Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. In: Rose, GD., editor. *Adv Prot Chem*. Vol. 62. Academic Press; 2002. p. 241-262.
4. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiagarajan P, Sosnick TR, et al. Random-coil behavior and the dimensions of chemically unfolded proteins. *Biophys J*. 2004; 88:2472, 2493.101, 12491–12496.
5. McCarney ER, Kohn JE, Plaxco KW. Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit Rev Biochem Mol Biol*. 2005; 40:1–9. [PubMed: 15804623]
6. Snow CD, Qiu LL, Du DG, Gai F, Hagen SJ, Pande VS. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc Natl Acad Sci USA*. 2004; 101:4077–4082. [PubMed: 15020773]

7. Sorin EJ, Pande VS. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J*. 2005; 88:2472–2493. [PubMed: 15665128]
8. Flanagan JM, Kataoka M, Shortle D, Engelman DM. Truncated Staphylococcal Nuclease is Compact but Disordered. *Proc Natl Acad Sci USA*. 1992; 89:748–752. [PubMed: 1731350]
9. Arai, M.; Kuwajima, K. Role of the molten globule state in protein folding. In: Matthews, CR., editor. *Adv Prot Chem*. Vol. 53. Academic Press; 2000. p. 209-282.
10. Chugha P, Sage HJ, Oas TG. Methionine oxidation of monomeric lambda repressor: The denatured state ensemble under nondenaturing conditions. *Prot Sci*. 2006; 15:533–542.
11. Choy WY, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J Mol Biol*. 2002; 316:101–112. [PubMed: 11829506]
12. Roder H, Shastry MCR. Methods for exploring early events in protein folding. *Curr Op Struc Biol*. 1999; 9:620–626.
13. Chen LL, Wildegger G, Kiefhaber T, Hodgson KO, Doniach S. Kinetics of lysozyme refolding: Structural characterization of a non-specifically collapsed state using time-resolved X-ray scattering. *J Mol Biol*. 1998; 276:225–237. [PubMed: 9514723]
14. Pollack L, Tate MW, Darnton NC, Knight JB, Gruner SM, Eaton WA, Austin RH. Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering. *Proc Natl Acad Sci USA*. 1999; 96:10115–10117. [PubMed: 10468571]
15. Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat Struc Biol*. 1999; 6:554–556.
16. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR. Early collapse is not an obligate step in protein folding. *J Mol Biol*. 2004; 338:369–382. [PubMed: 15066438]
17. Shortle, D. The expanded denatured state: An ensemble of conformations trapped in a locally encoded topological space. In: Rose, GD., editor. *Adv Prot Chem*. Vol. 62. Academic Press; 2002. p. 1-23.
18. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease .2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J Mol Biol*. 1997; 268:170–184. [PubMed: 9149150]
19. Crowhurst KA, Tollinger M, Forman-Kay JD. Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. *J Mol Biol*. 2002; 322:163–178. [PubMed: 12215422]
20. Crowhurst KA, Forman-Kay JD. Aromatic and methyl NOES highlight hydrophobic clustering in the unfolded state of an SH3 domain. *Biochemistry*. 2003; 42:8687–8695. [PubMed: 12873128]
21. Nishimura C, Riley R, Eastman P, Fink AL. Fluorescence energy transfer indicates similar transient and equilibrium intermediates in staphylococcal nuclease folding. *J Mol Biol*. 2000; 299:1133–1146. [PubMed: 10843864]
22. Kimura T, Akiyama S, Uzawa T, Ishimori K, Morishima I, Fujisawa T, Takahashi S. Specifically collapsed intermediate in the early stage of the folding of ribonuclease A. *J Mol Biol*. 2005; 350:349–362. [PubMed: 15935376]
23. Hoffman A, Kane A, Nettels D, Hertzog D, Baumgärtel P, Lengefeld J, Reichardt G, Horsley DA, Seckler R, Bakajin O, Schuler B. Mapping protein collapse with single molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA*. 2007; 104:105–110. [PubMed: 17185422]
24. Sosnick TR, Shtilerman MD, Mayne L, Englander SW. Ultrafast signals in protein folding and the polypeptide contracted state. *Proc Natl Acad Sci USA*. 1997; 94:8545–8550. [PubMed: 9238013]
25. Qi PX, Sosnick TR, Englander SW. The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nat Struct Biol*. 1998; 5:882–884. [PubMed: 9783747]
26. Englander SW. Protein folding intermediates and pathways studied by hydrogen exchange. *Ann Rev Bioph Biomol Struc*. 2000; 29:213–238.
27. Alexandrescu AT, Abeygunawardana C, Shortle D. Structure and dynamics of a denatured 131-residue fragment of Staphylococcal Nuclease - A heteronuclear NMR-study. *Biochemistry*. 1994; 33:1063–1072. [PubMed: 8110737]

28. de Prat Gay G, Ruiz-Sanz J, Neira JL, Corrales FJ, Otzen DE, Ladurner AG, Fersht AR. Conformational pathway of the polypeptide-chain of chymotrypsin inhibitor-2 growing from its N-terminus *in-vitro* - parallels with the protein-folding pathway. *J Mol Biol.* 1995; 254:968–979. [PubMed: 7500364]
29. Neira JL, Fersht AR. Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. *J Mol Biol.* 1999; 285:1309–1333. [PubMed: 9887278]
30. Neira JL, Fersht AR. Acquisition of native-like interactions in C-terminal fragments of barnase. *J Mol Biol.* 1999; 287:421–432. [PubMed: 10080903]
31. Viguera AR, Arrondo JLR, Musacchio A, Saraste M, Serrano L. Characterization of the interaction of natural proline-rich peptide with 5 different SH3 domains. *Biochemistry.* 1994; 33:10925–10933. [PubMed: 8086409]
32. Plaxco KW, Gujjarro JI, Morton CJ, Pitkeathly M, Campbell ID, Dobson CM. The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry.* 1998; 37:2529–2537. [PubMed: 9485402]
33. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry.* 1999; 38:16424–16431. [PubMed: 10600103]
34. Flory, PJ. Principles of Polymer Chemistry. Cornell University Press; Ithica, NY: 1953.
35. Nozaki Y, Tanford C. Solubility of Amino Acids, Diglycine, and Triglycine in Aqueous Guanidine Hydrochloride Solutions. *J Biol Chem.* 1970; 245:1648–1652. [PubMed: 5438355]
36. Kataoka M, Kuwajima K, Tokunaga F, Goto Y. Structural characterization of the molten globule of alpha-lactalbumin by solution X-ray scattering. *Prot Sci.* 1997; 6:422–430.
37. Gast K, Damaschun H, Misselwitz R, Mueller-Frohne M, Zirwer D, Damaschun G. Compactness of Protein Molten Globules - Temperature-Induced Structural-Changes of the Apomyoglobin Folding Intermediate. *Eur Biophys J.* 1994; 23:297–305. [PubMed: 7805629]
38. Yamauchi A, Yomo T, Tanaka F, Prijambada ID, Ohhashi S, Yamamoto K, Shima Y, Ogasahara K, Yutani K, Kataoka M, Urabe I. Characterization of soluble artificial proteins with random sequences. *FEBS Lett.* 1998; 421:147–151. [PubMed: 9468296]
39. Gutin AM, Abkevich VI, Shakhovich EI. Is Burst Hydrophobic Collapse Necessary for Protein-Folding. *Biochemistry.* 1995; 34:3066–3076. [PubMed: 7893719]
40. Chan CK, Hu Y, Takahashi S, Rousseau DL, Eaton WA, Hofrichter J. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc Natl Acad Sci USA.* 1997; 94:1779–1784. [PubMed: 9050855]
41. Sosnick TR, Berry RS, Colubri A, Fernandez A. Distinguishing foldable proteins from nonfolders: When and how do they differ? *Proteins.* 2002; 49:15–23. [PubMed: 12211012]
42. Clementi C, Plotkin SS. The effects of nonnative interactions on protein folding rates: Theory and simulation. *Prot Sci.* 2004; 13:1750–1766.
43. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, Davidson AR, Chan HS. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc Natl Acad Sci USA.* 2008; 105:9999–10004. [PubMed: 18626019]
44. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, MR.; Appel, RD.; Bairoch, A. The Proteomics Protocols Handbook. Walker, John M., editor. 2005. p. 571-607.
45. Grosjean H, Fiers W. Preferential codon usage in prokaryotic genes - the optimal codon anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene.* 1982; 18:199–209. [PubMed: 6751939]
46. Altieri AS, Hinton DP, Byrd RA. Association of biomolecular systems *via* pulsed-field gradient NMR self-diffusion measurements. *J Am Chem Soc.* 1995; 117:7566–7561.

WT: TLFEALYDYEARTEDDL⁰SFHKGEKFQILNSSEGDWWEARSLTTGETGYIPSNYVAPV**stop**

CD4: TLFEALYDYEARTEDDL⁰SFHKGEKFQILNSSEGDWWEARSLTTGETGYIPSNY**stop**

RS: EHKEEEPWALGDGYASTE⁰FESIFTDWRASELRYSDIDFTYNLGNTSQKLYGL**Tstop**

Figure 1.

The sequence of full-length FynSH3 (WT), a four-residue carboxy-terminal truncation (CΔ4) and a random-sequence variant (RS) of identical sequence composition to CΔ4. CΔ4 and RS share less than 4% sequence identity, and BLAST searches indicate that the sequence of RS is not statistically significantly related to that of any other protein. Not shown are unstructured, 21-residue his-tag/linkers at the amino-terminus of all three constructs.

Figure 2.

While the far-UV CD spectrum of CΔ4 exhibits slightly greater ellipticity than the chemically denatured protein, it exhibits significantly less ellipticity than the native protein under the same conditions, suggesting a significant reduction in structure. In contrast the spectrum of the RS protein collected under non-denaturing conditions is nearly indistinguishable from that of the chemically denatured, wild type protein, which is thought to adopt a fully unfolded, random-coil configuration (4). The spectrum of GuHCl-denatured protein is truncated at 220 nm due to the strong absorbance of the denaturant below this wavelength.

Figure 3.

CΔ4 exhibits little chemical shift dispersion, suggesting that it lacks fixed tertiary structure. In particular, a strongly shifted methyl resonance at -0.6 that is a signature of the SH3 fold is effectively entirely lost. However, CΔ4 exhibits moderate line broadening (relative to both the native and GuHCl-unfolded protein), suggesting that it may adopt a molten globule-like conformation. RS, in contrast, lacks both significant spectral dispersion and line broadening, suggesting that it is nearly completely unfolded even in the absence of denaturant. A large peak arising from TSP (used as an internal standard) was deleted from the RS spectrum, accounting for the data gap around 0 ppm.

Figure 4.

The ^{15}N HSQC of native FynSH3 exhibits significantly greater chemical shift dispersion than CΔ4, indicating that the latter is relatively unfolded under these non-denaturing conditions. The chemical shift dispersion of FynSH3 in 6M GuHCl is, however, still further reduced. This suggests that CΔ4 resides somewhere along the molten-globule spectrum.

Figure 5.

Hydrogen exchange protection-studies suggest that CΔ4 is largely unfolded. Whereas the native protein has not fully exchanged after 1 hour at pH 7, CΔ4 appears to exchange completely within the 7minute dead time of this manual mixing experiment. Shown are the ratios of total amide-to-aromatic ^1H intensities (normalized to unity for the unexchanged proteins) at intervals after transfer into fully deuterated buffer at pD 7.

Figure 6.

CΔ4 is relatively compact in the absence of denaturant. Shown is the Pulsed Field Gradient-NMR signal attenuation of CΔ4 and dioxane, an internal standard, measured as functions of field gradient strength in the absence and presence of GuHCl. The small discrepancy between the dioxane measurements arises due to the increased viscosity of GuHCl solutions. The larger discrepancy observed between CΔ4 measurements reflects an additional increase in diffusion constant due to expansion of the chain. The fitted curves (equation 1, 2) predict radii of hydration of $19.4 \pm 0.3 \text{ \AA}$ and $23.0 \pm 1.0 \text{ \AA}$ for CΔ4 in the absence and presence of GuHCl respectively.

Table 1

Hydrodynamic radii of FynSH3 variants in the presence and absence of denaturant.

FynSH3 variant	Non-denaturing Conditions ^a	Denaturing Conditions ^a ,
WT	18.4±1.0	25.2±0.3
CΔ4	19.4±0.3	23.0±1.0
RS	20.5±1.0	24.4±1.8

^aMean and 95% confidence intervals.