

Non-random-coil Behavior as a Consequence of Extensive PPII Structure in the Denatured State

Aitziber L. Cortajarena¹, Gregg Lois², Eilon Sherman³,
Corey S. O'Hern², Lynne Regan^{1,4*} and Gilad Haran^{3*}

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

²Departments of Mechanical Engineering and Physics, Yale University, New Haven, CT 06520, USA

³Department of Chemical Physics, Weizmann Institute of Science, PO Box 26, Rehovot 76100, Israel

⁴Department of Chemistry, Yale University, New Haven, CT 06520, USA

Received 30 January 2008;
received in revised form
21 June 2008;
accepted 3 July 2008
Available online
11 July 2008

Unfolded proteins may contain a native or nonnative residual structure, which has important implications for the thermodynamics and kinetics of folding, as well as for misfolding and aggregation diseases. However, it has been universally accepted that residual structure should not affect the global size scaling of the denatured chain, which obeys the statistics of random coil polymers. Here we use a single-molecule optical technique—fluorescence correlation spectroscopy—to probe the denatured state of a set of repeat proteins containing an increasing number of identical domains, from 2 to 20. The availability of this set allows us to obtain the scaling law for the unfolded state of these proteins, which turns out to be unusually compact, strongly deviating from random coil statistics. The origin of this unexpected behavior is traced to the presence of an extensive nonnative polyproline II helical structure, which we localize to specific segments of the polypeptide chain. We show that the experimentally observed effects of polyproline II on the size scaling of the denatured state can be well-described by simple polymer models. Our findings suggest a hitherto unforeseen potential of nonnative structure to induce significant compaction of denatured proteins, significantly affecting folding pathways and kinetics.

© 2008 Elsevier Ltd. All rights reserved.

Edited by C. R. Matthews

Keywords: protein folding; fluorescence correlation spectroscopy; PPII helix; denatured state; self-avoiding random walk

Introduction

During the folding process, proteins reach a defined, unique native structure from a poorly defined ensemble of unfolded conformations. Although we have an atomistic description of the structures of the native states of a multitude of proteins, much less is known about the unfolded ensemble. The unfolded state is often treated as some version of a random coil, a realistic model of which is the self-avoiding

random walk (SARW), in which different chain segments cannot occupy the same volume element. For a SARW, the relationship between the radius of gyration R_g (or the hydrodynamic radius R_h) and the length of the chain is given by the expression $R_g \propto N^\nu$, where $\nu = 0.59$ and N is the number of units in the chain.^{1,2} For many different denatured proteins, the experimentally measured relationship between R_g or R_h and N is consistent with the behavior of a SARW.^{3–6} This observation has been interpreted as support for an unstructured unfolded state. Conversely, there have been several reports describing the existence of ‘residual structure’ in the unfolded state,^{7–10} which, in general, does not change the global size scaling of the chain.^{11,12} A detailed understanding of the nature of the unfolded state is essential for a complete understanding of protein folding because any residual structure in the ‘unfolded state’ will clearly play a role in modulating both the thermodynamics and the kinetics of

*Corresponding authors. E-mail addresses:
lynne.regan@yale.edu; Gilad.Haran@weizmann.ac.il.
Abbreviations used: SARW, self-avoiding random walk; CTPR, consensus tetratricopeptide repeat; FCS, fluorescence correlation spectroscopy; GuHCl, guanidinium hydrochloride; Rh6G, rhodamine 6G; PPII, polyproline II; SE, Stokes–Einstein.

folding, and might affect the formation of alternative structures such as amyloid.

Repeat proteins are composed of tandem arrays of a small structural motif.¹³ They are widespread in nature and, in many cases, function as mediators of protein-protein interactions. Their simplified modular structures make them ideal systems for understanding basic principles that drive protein folding.^{14–16} The tetratricopeptide repeat is a 34-

amino-acid motif in which two anti-parallel α -helices stack together.¹⁷ We have previously reported the design and characterization of a consensus tetratricopeptide repeat (CTPR) sequence and have generated a series of proteins with different numbers of tandem repeats of this sequence.^{17–19} Tandem arrays of CTPR units form a regular superhelical extended structure (Fig. 1a).^{18,19} It has been suggested that the folding mechanism of these proteins

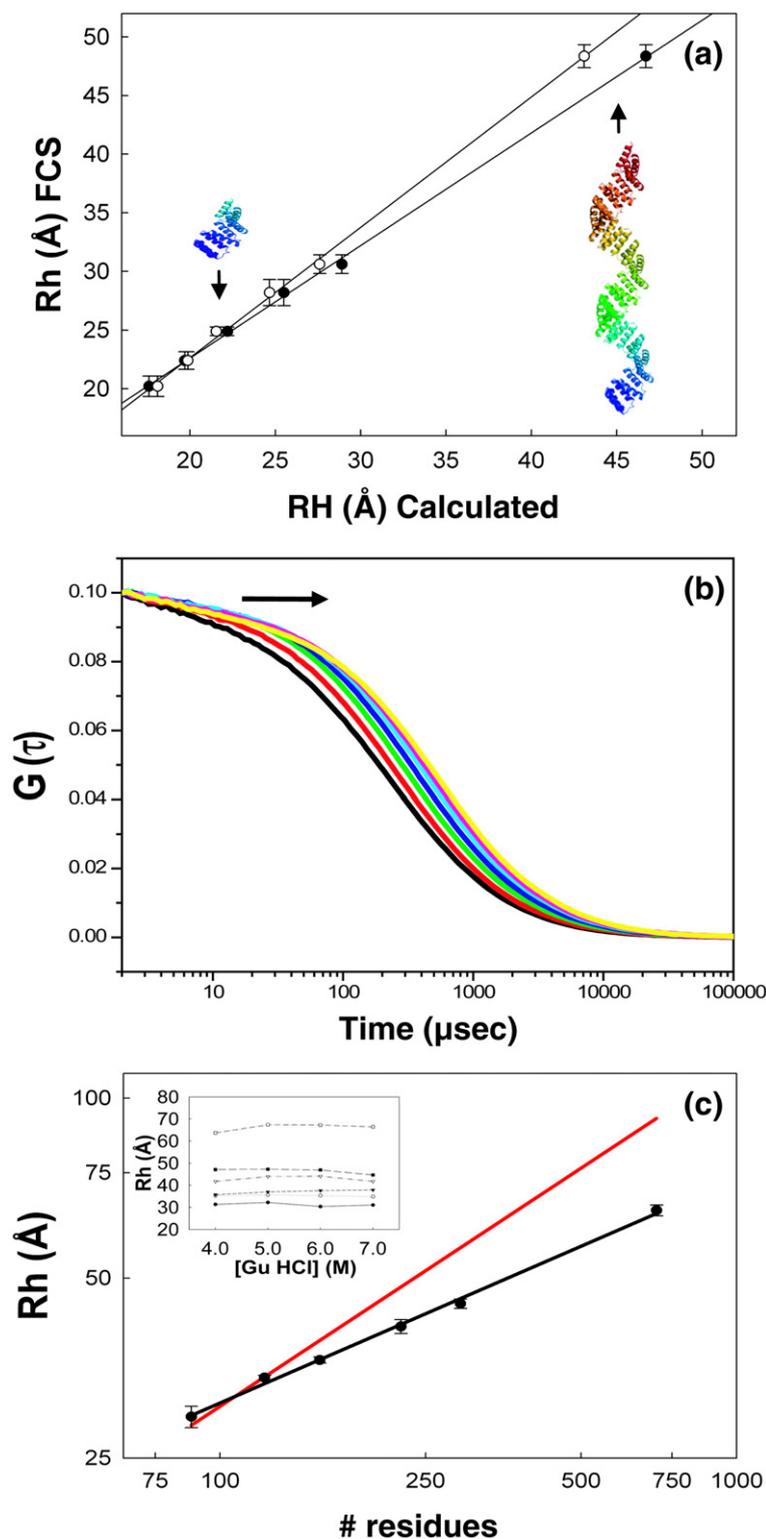


Fig. 1. R_h values of native and unfolded CTPR proteins measured by FCS. (a) Correlation between the R_h values for CTPRs of different lengths (CTPR2, CTPR3, CTPR4, CTPR6, CTPR8, and CTPR20) measured by FCS and the R_h values calculated from the crystal structures of the CTPRs using the program Hydropro,²⁰ (filled circles) or the R_h values calculated using a model that describes the hydrodynamic properties of cylindrical particles (empty circles)²¹ (Eq. (5) in Materials and Methods). The figure shows the ribbon representations of the structures of two CTPR proteins (CTPR4 and CTPR20). (b) FCS cross-correlation curves for the CTPR3 protein at various GuHCl concentrations (from 0 M to 6 M). Arrow indicates the direction of increase in denaturant concentration, which shifts the curves to the right. (c) A log-log plot of R_h values of the different CTPRs under strongly denaturing conditions at 6 M GuHCl (filled dots) and the least-squares fit to the data (solid line). The red line represents the empirical relation between the measured R_h and protein length for a set of highly denatured globular proteins.⁴ The inset shows the dependence of the R_h value for CTPR proteins of different lengths in the unfolded state, CTPR2 (●), CTPR3 (○), CTPR4 (▼), CTPR6 (▽), CTPR8 (■), and CTPR20 (□), indicating that above ~4 M GuHCl, the size of these proteins does not change anymore.

may be different from that of globular proteins.¹⁸ Here we combine spectroscopy and modeling to shed light on the structure of the denatured state of the CTPR proteins. Surprisingly, we find that extensive a nonnative structure leads to a compact denatured state, with size scaling strongly deviating from that expected for a random coil.

Results and Discussion

The denatured state of CTPR proteins is not a random coil

Fluorescence correlation spectroscopy (FCS) probes the hydrodynamic properties of macromolecules by recording fluctuations in the fluorescence they emit.²² We employed FCS to measure the hydrodynamic radii of CTPR proteins in dilute solutions, in both their folded and their unfolded states. For this purpose, we labeled CTPR proteins of 2–20 repeats (CTPR2–CTPR20) with the fluorescent probe Alexa Fluor 488 at a unique C-terminal cysteine residue. We first measured the values of R_h for all the different-length CTPR proteins in the native state. The measured values of R_h (Table 1) are in very good agreement with values calculated either from crystal structures of CTPR proteins using the program Hydropro²⁰ or by treating the folded structures as rods and using an expression for R_h derived for rigid cylinders²¹ (Fig. 1a).

The hydrodynamic properties of CTPR2–CTPR20 were then studied under strongly denaturing conditions at 6 M guanidinium hydrochloride (GuHCl) (see Table 1 and Fig. 1b and c). We calibrated the FCS instrument and corrected the results for viscosity and refractive index effects using rhodamine 6G (Rh6G) as standard (see Materials and Methods). The availability of the full series of CTPR proteins allowed us to obtain a scaling relation for R_h within the series. Unexpectedly, we observed that, for a given length of protein, the measured R_h is *different* from that anticipated if the polypeptide behaves as a SARW. The denatured state of many proteins of different lengths has been studied using either hydrodynamic methods^{4,5} or small-angle X-ray scattering⁶ to measure R_h or R_g , respectively. These data can be fitted to the empirical scaling relation $R_h \propto N^{0.57}$ and $R_g \propto N^{0.59}$, respectively,^{4,6} where N is the number of amino acids. Clearly, these data are in good agreement with the scaling law for a

SARW.^{1,2} For the CTPR proteins, however, the relationship between the measured hydrodynamic radii of the unfolded proteins and N is best described by $R_h \propto N^{0.37}$ [$R_h = (5.49 \pm 0.25)N^{0.37 \pm 0.03}$] (Fig. 1c)†. Similar results were obtained with urea as denaturant (not shown). Thus, under strongly denaturing conditions, the unfolded state ensembles of the CTPR proteins do not behave as a SARW, but are rather more compact, suggesting the presence of a significant structure in the unfolded state‡.

Abundance of polyproline II structure in denatured CTPRs

The properties of proteins are determined by their amino acid composition and sequence. We therefore analyzed the sequence of the CTPR proteins and found that the abundance of several amino acids in the CTPR proteins deviates significantly from that of a reference state, the codon usage in all yeast proteins. Notably, amino acids with a high propensity to form polyproline II (PPII) helical structures, such as P, A, G, and Q,²³ appear in the CTPR sequence, on average, 1.8 times more frequently than in the yeast proteome.

Several recent studies have proposed the existence of a significant population of PPII helical conformations in unfolded proteins, leading to the suggestion that the occurrence of PPII structure may be a general feature of the 'unfolded' state.^{24–27} We used circular dichroism spectroscopy as a sensitive method by which to detect the presence of PPII helix: it is well established that this helix has a distinctive CD spectrum, with a maximum at 229 nm.^{24,28} Figure 2a shows CD spectra of CTPR8 at different concentrations of GuHCl, from 3.5 M to 6 M. Importantly, at 3.5 M GuHCl, the CD signal characteristic of α -helical structure has completely disappeared (data not shown; see also Kajander *et al.*¹⁸) The spectra clearly show a positive peak around 229 nm. The peak intensity increases with increasing GuHCl concentration, reaching a maximum at a concentration of about 4.5–5 M. This concentration is similar to that at which the unfolding titration curves monitored either by CD signal at 222 nm or by R_h also reach a maximum value. Similar spectra were obtained with urea as denaturant (data not shown). CD spectra of PPII helices are also characterized by a negative peak at 210 nm,²⁷ which is very difficult to observe in the presence of GuHCl. We were able to observe the negative peak by taking a CD spectrum with a short

Table 1. R_h values of native and unfolded CTPR proteins

Number of repeats	Number of amino acids	R_h (Å)	
		Native	Unfolded
2	88	20.21±0.87	29.31±1.21
3	122	22.40±0.75	34.07±0.31
4	156	24.89±0.37	36.48±0.42
6	224	28.19±1.11	41.49±1.12
8	292	30.62±0.79	45.30±0.81
20	700	48.36±0.98	64.89±1.32

† The series of R_h values of the native proteins could also be fitted to a power law, giving the relation $R_h = 3.12N^{0.41}$. Note that the preexponent is much smaller than that obtained for the denatured protein series. The rather high exponent is due to the elongated structure of the native CTPRs.

‡ Interestingly, the m -values for the folding of the CTPRs, especially the larger ones, are smaller than expected based on calculations of accessible surface areas of random-coiled denatured proteins of corresponding sizes (data not shown), also indicating compact structures.

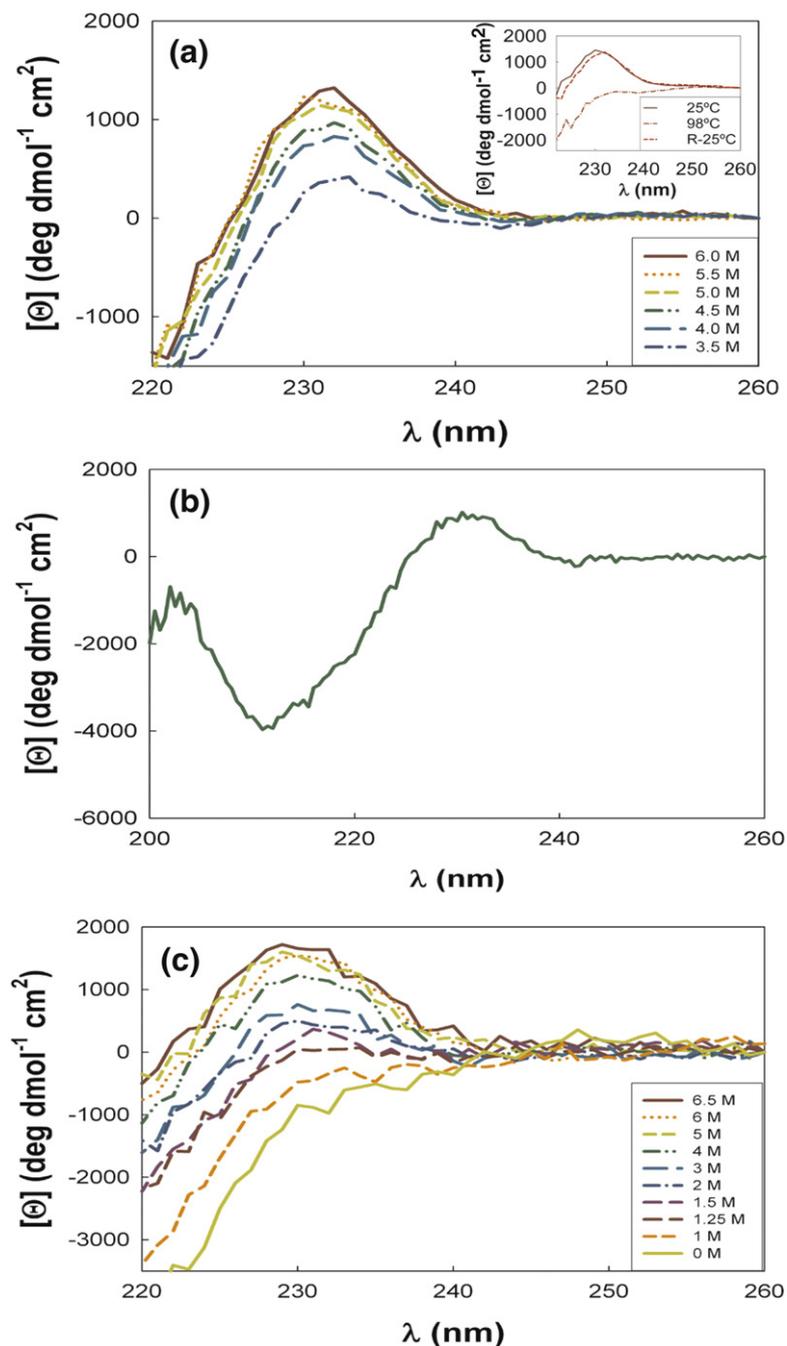


Fig. 2. CD spectra of CTPR8 and B-A peptide showing PPII helical structure. (a) CD spectra of the unfolded state of CTPR8 at increasing GuHCl concentration. The inset illustrates the temperature dependence of the CD signal, indicating a noncooperative but reversible melting of the PPII structure. It shows the CD spectra of CTPR8 at 6 M GuHCl at 25°C (solid line), at 98°C (dashed-dotted line), and at 25°C, after the sample was heated and cooled (dashed line). (b) CD spectrum of the B-A peptide at 2.5 M GuHCl in a 0.1-mm path-length cuvette at 25°C. (c) CD spectra of the B-A peptide in a GuHCl concentration range of 0–6.5 M.

path-length cell and a high protein concentration (not shown).

The modular repeated character of the CTPR series of proteins greatly facilitates our investigation of the nature of PPII structure in the unfolded state. An important question is whether the PPII structure extends through the entire sequence, or whether it is present in discrete segments. We estimated the fraction of PPII in CTPR proteins by comparing the observed mean residue ellipticity at 229 nm ($[\Theta]_{229}$) with $[\Theta]_{229}$ values reported for PPII-containing peptides.²⁷ All the CTPR proteins, from CTPR2 to CTPR20, show a substantial PPII content, which we roughly estimate to be as much as $\sim 50\%$ at 6 M GuHCl. The observed $[\Theta]_{229}$ for PPII is the same,

regardless of the number of repeats in each protein, indicating that there is an equal amount of PPII structure formed per repeat, *and* that there is no cooperativity between repeats in the formation of PPII structure.

CTPR1, however, does *not* exhibit significant PPII structure under denaturing conditions. CTPR1 has the sequence A-B-A_{cap}, where A and B are the two α -helices forming the basic repeat unit, and A_{cap} is a solvating helix (for sequence information, see [Materials and Methods](#)),¹⁷ whereas all the longer CTPR proteins have the sequence (A-B)_n-A_{cap}.^{17,19} CTPR1 thus lacks the B-A interrepeat sequence. Based on the method of Ref. 29, this sequence (but not the B-A_{cap} sequence) can, in fact, be predicted to have

a high propensity for PPII formation. To test this hypothesis, we synthesized a peptide corresponding to the B–A sequence. This peptide does indeed adopt PPII structure, and the CD spectrum (Fig. 2b) shows a positive peak at 229 nm and a negative minimum at 210 nm, characteristic of PPII structure.²⁷ The $[\Theta]_{229}$ value is comparable to that of the CTPR proteins. Moreover, PPII formation by the peptide shows the same denaturant concentration dependence as that observed for the full-length proteins (Fig. 2c). This result confirms that the B–A sequence in the CTPR proteins is responsible for the PPII structure, and that the PPII segments form independently. Because the B–A peptide is unstructured in the absence of denaturant (not shown), our results clearly indicate that GuHCl not only promotes formation of PPII structure by destabilizing the native state but also directly stabilizes this structure.

Finally, the PPII content of peptides has been shown to decrease when the temperature increases.³⁰ We therefore tested the effect of temperature on the PPII content of highly denatured CTPR proteins. We observed a reversible loss of PPII structure with increasing temperature, such that the signal at 98 °C indicates essentially zero PPII content (inset to Fig. 2a).

Lattice simulations of denatured CTPRs

To better understand the effect of PPII structure on the dimensions of the denatured CTPRs, we performed simulations using simple polymer models. First, we studied the behavior of a freely jointed self-avoiding random walking polymer on a three-dimensional lattice for polymers with N equal-sized links of length L , where N varies between 10 and 1000, and L is the lattice constant (see Materials and Methods). To allow comparison with the experimental system, we considered any polymer of length N as containing $N/34$ repeats. For each of these chains, we calculated the R_g , and, as expected for a random walk with excluded volume, a log–log plot of R_g versus N yielded a straight line with a slope of 0.59 (Fig. 3a). To mimic the effect of different lengths of PPII structure within a repeat, we then included in the model links with two different sizes, L (the original) and $N_P L$ (to represent the number of links in the PPII segments), and allowed N_L steps of size L and one step with size $N_P L$ for each repeat. We set the repeat length to 34 ($N_L + N_P = 34$), as described above. This is a reasonable way to model the PPII segments. It is accepted that, at least up to a length of 12 residues, PPII peptides may behave as rigid rods,^{31,32}

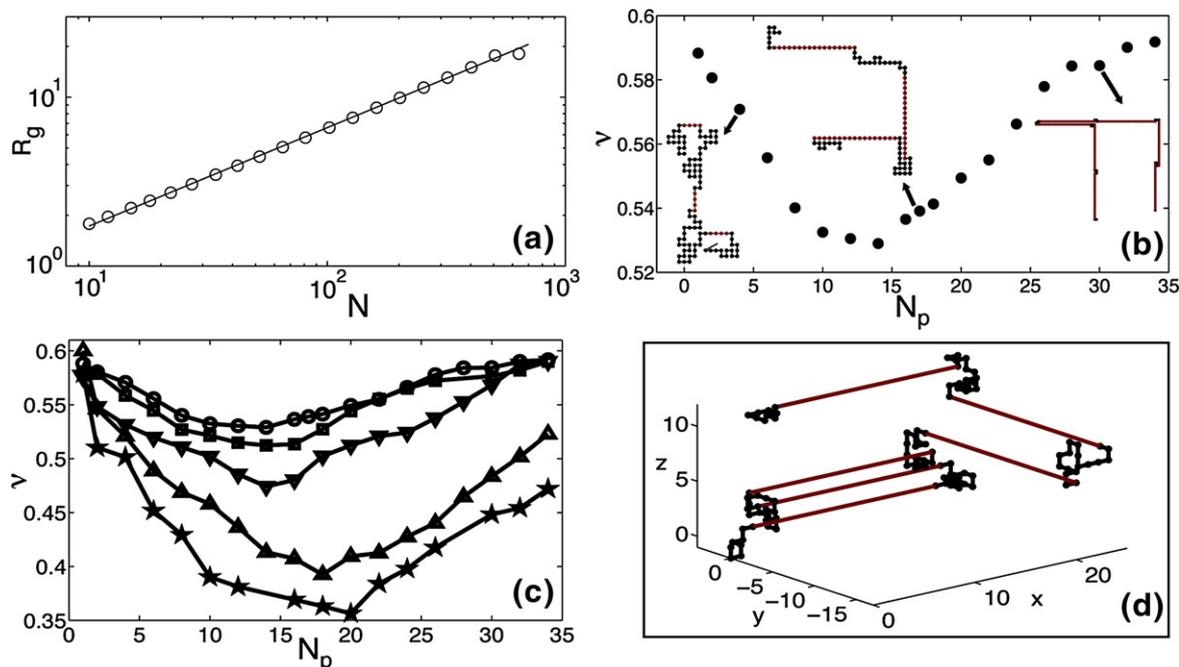


Fig. 3. A polymer model for the effect of PPII on the hydrodynamic properties of the denatured state. (a) A \log_{10} – \log_{10} plot of R_g versus N for the SARW simulation. The ensemble-averaged radius of gyration R_g is plotted against the number of monomers N in a classic excluded volume random walking chain (SARW). The data are fitted to a power law relation $R_g \propto N^\nu$, and we calculate $\nu = 0.587 \pm 0.003$. The measured exponent is consistent with the accepted value of 0.59. (b) The scaling exponent ν as a function of rigid segment (polyproline) length N_P . Two-dimensional representations of the polymer chains are pictured for $N_P = 4, 17,$ and 30 . Note that the ratio between large (red) and small segments (black) is drawn to scale for each value of N_P , but the overall size of the schematic is made smaller for larger N_P values to fit into the figure. The value of ν exhibits a minimum at $N_P = 14$, roughly half of a repeat unit. Error bars on the values of ν are ± 0.003 . (c) The scaling exponent ν as a function of polyproline length N_P and several fractions f of polyproline segments that experience an effective attraction. The data for $f = 0$ (no attraction) is identical with (b). As f increases, the value of ν decreases, and the location of the minimum moves to larger N_P values. Circles: $f = 0$; squares: $f = 0.2$; diamonds: $f = 0.4$; triangles: $f = 0.6$; stars: $f = 0.8$. Error bars on the values of ν are ± 0.003 for $f = 0$ and ± 0.011 all other f values. (d) Visualization of the polymer chain with $N = 204$, $N_P = 18$, and $f = 0.6$ for a representative configuration from the SARW ensemble.

and while isolated nonproline PPII peptides have been shown to be more flexible,³³ there is no such information on nonproline PPII peptides buried within proteins. Note that this scheme creates an *ordered sequence* of rigid and flexible chain stretches. We calculated R_g as a function of N for all N_P values from 1 to 34, and we calculated the scaling exponent ν from the slope of log–log plots of R_g versus N . In Fig. 3b, we show ν versus N_P , where $\nu=0.59$ for $N_P=1$ (i.e., no PPII and the expected random walk with excluded volume behavior). As N_P increases, ν decreases and reaches a minimum when $N_P/34$ is approximately 1/2. The long segments bring the chain to relatively empty regions of the lattice, so that the subsequent steps with small segments are less affected by the excluded volume constraint. This mechanism causes ν to decrease toward the Gaussian coil result of 0.5. As N_P increases further, ν returns to 0.59, and we again have a SARW, but with a larger link size. Note that the scaling exponent ν is smallest when the fraction of PPII segments is 50%, which agrees with experimental estimates of the fraction of PPII residues in each CTPR repeat unit. Although we cannot expect to make fully quantitative comparisons between the results of our simplified simulations and the experimentally observed data for a real polypeptide chain, these results provide a rationale for why the scaling exponent observed for denatured states containing PPII is not the same as that for a random coil, based simply on the intrinsic behavior of mixed polymers.

Interactions between PPII segments are likely to also contribute to the small scaling exponent $\nu \sim 0.37$ found in our experiments.³⁴ We therefore introduced effective attractive interactions between PPII segments in the simulations by minimizing the distance between a fraction (f ; for $0 \leq f \leq 0.8$) of the segments and all other segments in the chain. Incorporating this additional feature into the model has a dramatic effect on still further reducing the value of ν (Fig. 3c and d). Our results suggest that rod-like PPII segments and the interactions between these segments can significantly influence the structure of the unfolded state.

Conclusion

In conclusion, we used FCS to measure the value of R_h under denaturing conditions for a set of repeat proteins with 2–20 tandem copies of the same sequence. Our results revealed that the scaling behavior of R_h with N was *not* as expected for a SARW ($R_h \propto N^{0.37}$ versus the expected $R_h \propto N^{0.59}$), indicating a significantly more compact structure. Further investigations showed that there was considerable PPII structure in the denatured proteins. Moreover, we identified the segment of the repeat that formed a PPII helical structure and showed that PPII was present as regularly spaced, discrete segments throughout the protein. Using simple lattice models of polymers, we showed that chains with a regularly ordered sequence of short and long links (with the latter representing PPII segments) may attain a

structure more compact than that of the standard SARW. The incorporation of interactions between PPII segments into the model leads to further compaction of the chains, with size scaling similar to that observed in the experiment. These results present the most comprehensive picture, to date, of the effect of nonnative PPII structure on the nature of the unfolded state ensemble. The formation of nonnative structure in the denatured state may have implications for the folding mechanism of CTPR proteins, which we are now exploring.³⁵ Finally, since proline-rich segments have been recently shown to be effective in counteracting amyloid fiber formation,³⁶ we hypothesize that PPII structure may, in general, serve to modulate the aggregation of unfolded proteins.

Materials and Methods

Proteins

CTPR proteins are constructed as a tandem array of different numbers of a 34-amino-acid consensus CTPR sequence (sequence of the A–B helices: AEAWYNLG-NAYYKQGDYDEAIEYYQKALELDRS). An additional solvating helix is added to the C-terminus of each protein. This solvating helix (A_{cap} ; AKQNLGNKQKQG) is similar to the A helix, but with some hydrophobic residues substituted by hydrophilic residues. The B–A peptide sequence corresponds to a 34-amino-acid consensus TPR in which the B helix precedes the A helix (Ac-YDEAIEYYQ-KALELDRSAEAWYNLGNAYYKQGD-NH₂).

Preparation of labeled proteins

The cysteine-reactive maleimide form of Alexa Fluor 488 (Molecular Probes, Inc.) was used to label a unique cysteine inserted at the C-terminus of the proteins CTPR2, CTPR3, CTPR4, CTPR6, CTPR8, and CTPR20. The proteins (100 μ M) were reduced by incubation with 10 mM fresh 1,4-dithio-DL-treitol for 15 min. DDT was removed by buffer exchange over a PD-10 column (Amersham Biosciences, Uppsala, Sweden) against 500 mM NaCl and 50 mM phosphate (pH 7.0). A solution of Alexa Fluor 488 maleimide in water was added dropwise into the freshly reduced proteins in the presence of 1 mM tris(2-Carboxyethyl) phosphine and with constant stirring up to a final concentration of dye equal to 0.5 mM. The reactions were allowed to proceed for 4 h at room temperature under continuous stirring, under a nonreducing atmosphere of N₂ in the dark. After 4 h, an excess of β -mercaptoethanol (5% vol/vol) was added to consume the remaining thiol-reactive reagent for 15 min at 4 °C.

The free dye was removed from the reaction mixtures by two consecutive buffer exchange steps over PD-10 columns (Amersham Biosciences), followed by concentration and extensive washing on a Centriprep YM-3 filtration unit (Millipore), using a pH 6.8 buffer with 150 mM NaCl and 50 mM phosphate (phosphate-buffered saline).

The extent of labeling was estimated from the absorption spectra of the proteins using the extinction coefficient for Alexa Fluor 488 $\epsilon_{493} = 72,000 \text{ M}^{-1} \text{ cm}^{-1}$ from the Molecular Probes Handbook. The protein concentration was determined using an extinction coefficient calculated from the amino acid composition.³⁷ The absorbance at 280 nm was corrected for the contribution of the dye.

The protein samples were analyzed by mass spectrometry to confirm the labeling of the protein. Matrix-assisted laser desorption ionization time-of-flight analysis was performed on a Voyager-DE PRO Biospectrometry workstation (AB Applied Biosystems, Foster City, CA). The protein samples were mixed with 9 vol of the matrix solution (10 mg/ml sinapinic acid, 50% acetonitrile, and 0.1% trifluoroacetic acid) and spotted onto the matrix-assisted laser desorption ionization plate. The labeled proteins were frozen and stored in the dark at -80°C .

FCS measurements

FCS measurements were carried out using as excitation source a laser beam at 488 nm from an argon ion laser (Spectra Physics Lasers, Mountain View, CA). The collimated laser beam was focused into the sample, which was mounted on a home-built microscope (60 \times , 1.2 NA water immersion objective; Olympus, Tokyo, Japan). The laser power intensity was adjusted to 70 μW before the microscope. The beam was directed into the objective by a dichroic mirror (500 DCLP; Chroma, Rockingham, VT) and focused 20 μm into the sample. The fluorescence signal was collected through the same objective and transmitted by the dichroic mirror onto a long-pass filter to eliminate any reflected excitation light. It was then focused onto a 50- μm pin hole to eliminate out-of-focus photons and divided onto two single-photon avalanche photodiode units (Perkin-Elmer Photoelectronics, Fremont, CA) by a 50/50 nonpolarizing beam splitter (Unice E-O Services, Chung Li, Taiwan). The signals of the avalanche photodiode units were recorded in the cross-correlation mode for 1 h using a Flex02-12D digital real-time correlator (Correlator.com, Bridgewater, NJ).

All measurements were carried out on protein samples diluted in phosphate-buffered saline to a final concentration of 10 nM, which were transferred into cells constructed from two cover slides and sealed.

Cross-correlation functions were calculated from the signals of the two detectors. The correlation functions presented a fast exponential decay component in the nanoseconds-to-microseconds time scale. In order to calculate the diffusion time of the proteins, the correlation curves were fitted using the data from 20 μs onward, where there is no contribution of photophysical processes, using the following equation:²²

$$G(\tau) = \frac{1}{N} \left(1 + \frac{t}{\tau_D}\right)^{-1} \left(1 + \frac{t}{\tau_D(z_0/r_0)^2}\right)^{1/2} \quad (1)$$

where N is the number of molecules in the focal volume, τ_D is the diffusion time in the radial direction, and r_0 and z_0 are the radial and axial dimensions of the Gaussian observation volume, respectively. The r_0/z_0 ratio was obtained for each experimental condition from a measurement of the dye Rh6G, whose diffusion coefficient is known ($D_{\text{Rh6G}} = 280 \mu\text{m}^2 \text{s}^{-1}$).³⁸ Thus, FCS curves were

§ A recent measurement of the diffusion coefficient of R6G³⁹ gave a value that is larger than the above mentioned number by 30%. However, a survey of literature values of protein hydrodynamic radii (e.g., in Krouglova *et al.*⁴⁰) and our own measurements⁴¹ both suggest that using the new value will lead to hydrodynamic radii that are consistently smaller not only from calculations directly based on Protein Data Bank structures but also from measurements with other methods such as dynamic light scattering.

fitted with two free parameters for the diffusion term (τ_D and N). The diffusion coefficient for each sample was calculated from τ_D using the following equation:

$$\tau_D = \frac{r_0^2}{4D_t} \quad (2)$$

using r_0 values obtained from the Rh6G measurement. The Stokes–Einstein (SE) relation was used to calculate the hydrodynamic radius R_h of the molecules from their diffusion coefficient D_t at a given viscosity η :

$$D_t = \frac{k_B T}{6\pi\eta R_h} \quad (3)$$

where k_B is the Boltzmann constant and T is the absolute temperature.

We also fitted some of the cross-correlation functions over the full time scale, including two independent exponential decays in the equation, to consider the triplet state of the dye and the fast process observed in the nanoseconds-to-microseconds time scale:

$$G(\tau) = \frac{1}{N} \frac{(1 + A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2))}{\left(1 + \frac{t}{\tau_D}\right) \sqrt{1 + \frac{t}{\tau_D(z_0/r_0)^2}}} \quad (4)$$

where τ_1 and τ_2 are the relaxation times of the two processes, and A_1 and A_2 are their amplitudes. In this case, the FCS curves were fitted with two free parameters for the diffusion term (τ_D and N) and with four parameters for the two exponential decays (A_1 , A_2 , τ_1 , and τ_2). The diffusion time results obtained from the fit of the complete correlation function were in agreement with the values obtained from the fit of only the diffusion component of the function.

In order to obtain a well-behaved standard for the assessment of the effect of refractive index changes on apparent diffusion coefficients, we studied the diffusion of Rh6G molecules in all GuHCl solutions. In a simple liquid, the diffusion coefficient is inversely proportional to the viscosity of the medium, as described by SE relation (Eq. (3)). The SE relation implies that the diffusion time relative to that in water τ/τ_w should be proportional to the viscosity relative to water viscosity η/η_w , with a proportionality constant of 1. Figure 4 shows experimental results from a series of measurements in GuHCl solutions.

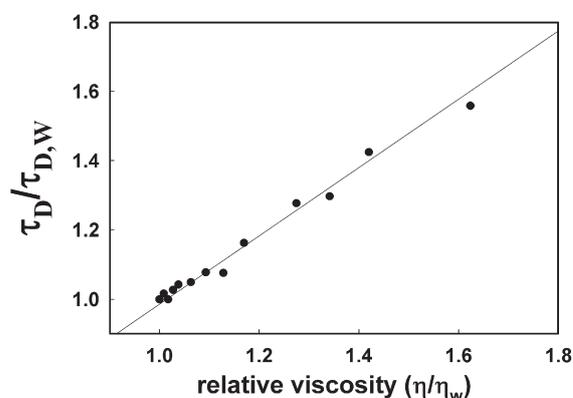


Fig. 4. Dependence of the relative translational correlation time of Rh6G molecules on relative viscosity in 0–6 M GuHCl solutions (filled circles). The line corresponds to the fit of the data to the equation $y = ax$, which gives a slope of 0.98, in agreement with the SE prediction.

The full line in this figure is a fit to $y=ax$, which gives a slope of 0.98, in close agreement with the theoretical SE prediction with a slope of 1. Clearly, the SE relation holds rather well in all GuHCl solutions, showing that refractive index changes or other optical effects do not significantly affect the measured diffusion times. We also controlled for possible effects of laser power on measured diffusion times^{42,43} by repeating some of our measurements at a series of lower powers down to 5 μW . We found slight changes in diffusion time values, which are reflected in error values reported in the study (see Results and Discussion). Such changes did not affect to any significance the scaling of R_h with size, which is the main observation in this study.

Calculation of R_h from crystal structures

The theoretical diffusion coefficients and the hydrodynamic radius for the different-length CTPR proteins were calculated from the X-ray crystal structure coordinates using the program Hydropro version 7c11.²⁰ We used the structures of CTPR8 (Protein Data Bank ID 2AVP)¹⁸ and CTPR20 (Kajander *et al.*, submitted) previously solved by our group. From the CTPR8 structure, we built atomic models for the smaller CTPRs.

Hydropro calculations were performed using the following values in the input file: 3.1 \AA for the atomic element radius and a minibead radius from 1 \AA to 5 \AA (Sigmin and Sigmax).

Calculation of R_h with rod-like model

Because the crystal structures of long CTPR proteins have an extended rod-like shape, we also calculated their hydrodynamic radii using an expression derived by Ortega and Garcia de la Torre for cylindrical particles:²¹

$$R_h = R_h^0(1.009 + 1.395 \times 10^{-2}(\ln p) + 7.8880 \times 10^{-2}(\ln p)^2 + 6.040 \times 10^{-3}(\ln p)^3) \quad (5)$$

In this expression, $R_h^0 = L(3/16p^2)^{1/3}$ is the hydrodynamic radius of a sphere having the same volume as the cylinder, and $p = L/d$ is the ratio between the cylinder length (L) and diameter (d). Using the dimensions of the TPR superhelix obtained from the crystal structure of CTPR8 ($L = 72 \text{\AA}/8$ repeats and $d = 38 \text{\AA}$), we calculated the hydrodynamic properties of the CTPRs of different lengths.

CD spectroscopy

All CD experiments were performed using an AVIV Model 215 CD spectrophotometer (AVIV Instruments, Lakewood, NJ) in 150 mM NaCl and 50 mM phosphate buffer (pH 6.8) at 25 $^\circ\text{C}$. CD spectra of CTPR proteins were acquired at 12.5 μM protein concentration and at 30 μM for the B-A peptide in a 0.2-cm path-length cuvette. To enable us to record the CD data at wavelengths below 215 nm shown in Fig. 2b, we used a higher peptide concentration (760 μM) and shorter path-length cuvettes (0.1 mm). The dynode voltage was monitored at the different wavelengths. From 260 nm to 203 nm, the dynode voltage values were in the optimal range to obtain

reliable data (180–500 V). Only at wavelengths below 203 nm did the dynode voltage exceed 500 V and did the signal-to-noise ratio become poor. The CD spectra were recorded with a bandwidth of 1 nm at 1-nm increments and with an average time of 10 s. Molar ellipticity ($[\Theta]$) was calculated using the following equation:

$$[\Theta] = \frac{\theta}{10cnl} \quad (6)$$

where θ is the ellipticity (in mdeg), c is the molar concentration of the sample (in M), n is the number of amino acids in the protein, and l is the path length (in cm).

The estimation of the PPII helix content from the ellipticity at 228 nm was based on proposed ellipticity values for peptides presenting 100% and 0% PPII helix content.²⁷

Thermal denaturation was monitored at a protein concentration of 12.5 μM by following the ellipticity signal of PPII at 228 nm from 10 $^\circ\text{C}$ to 98 $^\circ\text{C}$ and in the reverse direction from 98 $^\circ\text{C}$ to 10 $^\circ\text{C}$ in a 0.2-cm path-length cuvette. The temperature ramp was performed in 1 $^\circ\text{C}$ steps, with an equilibration time of 1 min at each temperature.

Simulations of self-avoiding random walking chains on a three-dimensional lattice

Simulations of a random walking chain with excluded volume (SARW) were conducted on a three-dimensional cubic lattice with N monomers. To construct a SARW polymer, each monomer was positioned on a lattice site adjacent to the previous monomer, with the constraint that no two monomers can occupy the same lattice site. A polymer chain is complete when N monomers have been placed on the lattice in this self-avoiding fashion. The radius of gyration R_g was calculated as an ensemble average over 10^6 SARW realizations for each N ,⁴⁴ and these data were used to compute the scaling exponent ν defined through the relation $R_g \propto N^\nu$.

The effect of internal structure on the scaling exponent ν was investigated by including the statistical effects of rigid polyproline segments. Motivated by experimental findings, N_P consecutive monomers in the middle of each repeat unit were assigned to represent the polyproline segment, with $1 \leq N_P \leq 34$ (where 34 is the number of monomers in each repeat unit). Monomers that are not in the polyproline segment were placed randomly, as in the SARW. The N_P monomers in the polyproline segment were placed such that they all lie in a straight line in one of the six lattice directions and occupy open lattice positions. The orientation of the segment was chosen randomly from the set of all self-avoiding orientations. This model is an extension of the SARW, and each repeat unit starts with $(34 - N_P)/2$ small steps, followed by one large step of size N_P , and ending with $(34 - N_P)/2$ small steps. This pattern was continued in the simulations over the entire experimental range of 2–20 repeat units. For each value of N_P , the radius of gyration R_g was averaged over an ensemble of 10^6 realizations, and a power law relation was observed between R_g and the number of repeat units. The scaling exponent ν was calculated for each N_P from these data.

The effect of interactions between polyproline segments on the exponent ν was modeled by randomly choosing a fraction f of the polyproline segments to have an effective attraction with other polyproline segments. In our model, if a segment is chosen, its orientation is not assigned randomly. Rather, for each possible self-avoiding orientation, the distance r_i between the center of the chosen

segment and the center of all other existing polyproline segments is calculated. The orientation of the chosen segment is then assigned to be the one that minimizes the sum $\sum_i (r_i^2)$. The radius of gyration R_g was measured and averaged over an ensemble of 10^5 realizations. A power law relation between R_g and the number of repeat units was observed, and ν was calculated for $0 \leq f \leq 0.8$.

Acknowledgements

We thank Professors A. Horovitz, A. D. Miranker, and A. Minsky, as well as members of the Regan and Haran Laboratories, for comments and suggestions on the manuscript. A.L.C. was a recipient of an European Molecular Biology Organization short-term fellowship. This research was made possible, in part, by the historic generosity of the Harold Perlman Family, as well as by partial financial support of the United States–Israel binational science foundation (grant no. 2002371 to G.H.), the National Institutes of Health (grant no. 1R01GM080515-01 to G.H.), and the Human Frontier Science Program (to G.H. and L.R.). Financial support to G.L. and C.S.O. from National Science Foundation grant no. DMR-0448838 and funding from Yale's Institute for Nanoscience and Quantum Engineering to G.L. are gratefully acknowledged.

References

- Grosberg, A. Y. & Kikhlov, A. R. (1994). *Statistical Physics of Macromolecules*. AIP Press, New York.
- Flory, P. J. (1953). *Principles of Polymer Chemistry*. Cornell University Press, New York.
- Tanford, C., Kawahara, K. & Lapanje, S. (1966). Proteins in 6 M guanidine hydrochloride demonstration of random coil behavior. *J. Biol. Chem.* **241**, 1921–1923.
- Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, **38**, 16424–16431.
- Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739–756.
- Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N. *et al.* (2005). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12491–12496.
- Kazmirski, S. L., Wong, K.-B., Freund, S. M. V., Tan, Y.-J., Fersht, A. R. & Daggett, V. (2001). Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4349–4354.
- Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, **293**, 487–489.
- Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T. *et al.* (2002). Long-range interactions within a nonnative protein. *Science*, **295**, 1719–1722.
- Religa, L., Markson, J. S., Mayor, U., Freund, S. M. V. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.
- Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12497–12502.
- Wang, Z., Plaxco, K. W. & Makarov, D. E. (2007). Influence of local, residual structure on the scaling behavior and dimensions of unfolded proteins. *Biopolymers*, **86**, 321–328.
- D'Andrea, L. & Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem. Sci.* **28**, 655–662.
- Lowe, A. R. & Itzhaki, L. S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2679–2684.
- Mello, C. C. & Barrick, D. (2004). An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14102–14107.
- Main, E. R., Stott, K., Jackson, S. E. & Regan, L. (2005). Local and long-range stability in tandemly arrayed tetratricopeptide repeats. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5721–5726.
- Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. (2003). Design of stable alpha-helical arrays from an idealized TPR motif. *Structure*, **11**, 497–508.
- Kajander, T., Cortajarena, A. L., Main, E. R., Mochrie, S. G. & Regan, L. (2005). A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* **127**, 10188–10190.
- Kajander, T., Cortajarena, A. L., Mochrie, S. G. & Regan, L. (2007). Structure and stability of a consensus TPR superhelix. *Acta Crystallogr. Sect. D*, **63**, 800–811.
- Garcia de la Torre, M. L., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719–730.
- Ortega, A. & Garcia de la Torre, J. (2003). Hydrodynamic properties of rodlike and dislike particles in dilute solution. *J. Chem. Phys.* **119**, 9914–9919.
- Krichevsky, O. & Bonnet, G. (2002). Fluorescence correlation spectroscopy: the technique and its applications. *Rep. Prog. Phys.* **65**, 251–297.
- Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E. & Creamer, T. P. (2003). Host–guest scale of left-handed polyproline II helix formation. *Proteins Struct. Funct. Bioinform.* **53**, 68–75.
- Woody, R. W. (1992). Circular dichroism and conformation of unordered polypeptides. *Adv. Biophys. Chem.* **2**, 37–79.
- Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L. & Kallenbach, N. R. (2002). Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9190–9195.
- Kentsis, A., Mezei, M., Gindin, T. & Osman, R. (2004). Unfolded state of polyalanine is a segmented polyproline II helix. *Proteins Struct. Funct. Bioinform.* **55**, 493–501.
- Kelly, M. A., Chellgren, B. W., Rucker, A. L., Trotman, J. M., Fried, M. G., Miller, A.-F. & Creamer, T. P. (2001). Host–guest study of left-handed polyproline II helix formation. *Biochemistry*, **40**, 14376–14383.
- Tiffany, M. L. & Krimm, S. (1973). Extended conformations of polypeptides and proteins in urea and guanidine hydrochloride. *Biopolymers*, **12**, 575–587.
- Vlasov, P. K., Vlasova, A. V., Tumanyan, V. G. & Esipova, N. G. (2005). A tetrapeptide-based method for polyproline II-type secondary structure prediction. *Proteins*, **61**, 763–768.
- Tiffany, M. L. & Krimm, S. (1972). Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. *Biopolymers*, **11**, 2309–2316.

31. Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M. & Eaton, W. A. (2005). Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2754–2759.
32. Stryer, L. & Haugland, R. P. (1967). Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 719–726.
33. Makowska, J., Rodziewicz-Motowidlo, S., Baginska, K., Vila, J. A., Liwo, A., Chmurzynski, L. & Scheraga, H. A. (2006). Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1744–1749.
34. Camacho, C. J. & Schanke, T. (1997). From collapse to freezing in random heteropolymers. *Europhys. Lett.* **37**, 603–608.
35. Cortajarena, A. L., Mochrie, S. G. & Regan, L. (2008). Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange. *J. Mol. Biol.* **379**, 617–626.
36. Faber, P. W., Voisine, C., King, D. C., Bates, E. A. & Hart, A. C. (2002). Glutamine/proline-rich PQE-1 proteins protect *Caenorhabditis elegans* neurons from huntingtin polyglutamine neurotoxicity. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 17131–17136.
37. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423.
38. Magde, D., Elson, E. & Webb, W. W. (1974). Fluorescence correlation spectroscopy: II. An experimental realization. *Biopolymers*, **13**, 29–61.
39. Culbertson, C. T., Jacobson, S. C. & Ramsey, M. J. (2002). Diffusion coefficient measurements in microfluidic devices. *Talanta*, **56**, 365–379.
40. Krouglova, T., Vercammen, J. & Engelborghs, Y. (2004). Correct diffusion coefficients of proteins in fluorescence correlation spectroscopy. Application to tubulin oligomers induced by Mg²⁺ and paclitaxel. *Biophys. J.* **87**, 2635–2646.
41. Sherman, E., Itkin, A., Kuttner, Y. Y., Rhoades, E., Amir, D., Haas, E. & Haran, G. (2008). Using fluorescence correlation spectroscopy to study conformational changes in denatured proteins. *Biophys. J.* **98**, 4819–4827.
42. Eggeling, C., Widengren, J., Rigler, R. & Seidel, C. A. M. (1998). Photobleaching of fluorescent dyes under conditions used for single-molecule detection: evidence of two-step photolysis. *Anal. Chem.* **70**, 2651–2659.
43. Gregor, I., Patra, D. & Enderlein, J. (2005). Optical saturation in fluorescence correlation spectroscopy under continuous-wave and pulsed excitation. *ChemPhysChem*, **6**, 164–170.
44. Wall, F. T. & Erpenbeck, J. J. (1959). New method for the statistical computation of polymer dimensions. *J. Chem. Phys.* **30**, 634.