
The authors note that, due to a printer’s error, the affiliation for Abbas Mgharbel, Benjamin Audren, Jean-Paul Rieu, and Hélène Delanoë-Ayari at Université de Lyon appeared incorrectly in part. The name of the authors’ laboratory, Laboratoire Matériaux et Phénomènes Quantiques, should instead have appeared as Laboratoire de Physique de la Matière Condensée et Nanostructures. The corrected affiliation line appears below.

Philippe Marmottanta, Abbas Mgharbelb, Jos Käfera, Benjamin Audrenb, Jean-Paul Rieub, Jean-Claude Viala, Boudewijn van der Sandenc, Athanasius F. M. Maréea, François Granera,e, and Hélène Delanoë-Ayarib

aLaboratoire de Spectrométrie Physique, Unité Mixte de Recherche 5588, Université Grenoble I and Centre National de la Recherche Scientifique, 140 Avenue de la Physique, F-38402 Martin d’Hères Cedex, France; bLaboratoire de Physique de la Matière Condensée et Nanostructures, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5586, Université de Lyon, Lyon Université Lyon 1, 43 Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France; cInstitut des Neurosciences Grenoble, Institute National de la Santé et de la Recherche Médicale U836, Chemin Fortune Ferrini, BP 170, F-38042 Grenoble Cedex 9, France; dTheoretical Biology/Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands; and eBiology du Développement, Unité Mixte de Recherche 3215 Institut Curie and Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale U934, 26 Rue d’Ulm, F-75248 Paris Cedex 05, France

www.pnas.org/cgi/doi/10.1073/pnas.0911892106

Correction for “Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins,” by Yalda Javadi and Ewan R. G. Main, which appeared in issue 41, October 13, 2009, of Proc Natl Acad Sci USA (106:17383–17388; first published October 1, 2009; 10.1073/pnas.0907455106).

The authors note that, due to a printer’s error, some text appeared incorrectly on page 17384. In the left column, the second sentence in the third full paragraph, “This yielded [D]50% (midpoint of unfolding) and mD-N (change in solvent-accessible surface area upon protein unfolding) from which ∆GD–N/H2O/∆G0 was calculated (Table S1),” should instead have appeared as “This yielded [D]50% (midpoint of unfolding) and mD-N (change in solvent-accessible surface area upon protein unfolding) from which ∆GD–N/H2O/H2O was calculated (Table S1).” In the right column, the sentence beginning on line 17, “All of the values of ∆GD–N/H2O/H2O calculated are within error for urea and GuHCl denaturations,” should instead have appeared as “All of the values of ∆GD–N/H2O calculated are within error for urea and GuHCl denaturations.” These errors do not affect the conclusions of the article.

www.pnas.org/cgi/doi/10.1073/pnas.0911552106
Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins

Yalda Javadi and Ewan R. G. Main

Centre for Structural and Chemical Biology, Department of Chemistry and Biochemistry, University of Sussex, Falmer near Brighton BN1 9QJ, United Kingdom

Edited by Peter G. Wolynes, University of California, San Diego, La Jolla, CA, and approved August 18, 2009 (received for review July 8, 2009)

Repeat proteins contain short, tandem arrays of simple structural motifs (20–40 aa). These stack together to form nonglobular structures that are stabilized by short-range interactions from residues close in primary sequence. Unlike globular proteins, they have few, if any, long-range nonlinear stabilizing interactions. One ubiquitous repeat is the tetratricopeptide motif (TPR), a 34-aa helix-turn-helix motif. In this article we describe the folding kinetics of a series of 7 designed TPR proteins that are assembled from a combination of identical designed consensus repeats (CTPRans). These range from the smallest 2-repeat protein to a large 10-repeat protein (>350 aa). In particular, we describe how the energy landscape changes with the addition of repeat units. The data reveal that although the CTPRa proteins have low local frustration, their highly symmetric, modular native structure is reflected in their multistate kinetics of unfolding and folding. Moreover, although the initial folding of all CTPRa proteins involves a nucleus with similar solvent accessibility, their subsequent folding to the native structure depends directly on repeat number. This corresponds to an increasingly complex landscape that culminates in CTPRa10 populating a misfolded, off-pathway intermediate. These results extend our current understanding of the malleable folding pathways of repeat proteins and highlight the consequences of adding identical repeats to the energy landscape.

design protein | kinetic traps | misfolding | protein folding

Repeat proteins consist of tandem arrays of simple structural motifs that consist of 20–40 residues. These modules stack together to form elongated, nonglobular protein folds. They are highly abundant, widespread in nature and, second only to immunoglobulins, are the most common protein class to specialize in protein–protein interactions (1). Examples of repeat proteins include tetratricopeptide (TPR), ankyrin, and leucine-rich (LRR) repeats (2, 3). Unusually, unlike globular proteins, these repeat proteins do not rely on complex long-range stabilizing interactions. Instead their recurring modular architectures are dominated by regularized short-range interactions (both inter- and intrarepeat). This distinctive feature, which results in a quasi-1-dimensional structure, has made them extremely attractive targets as models for protein folding and design studies.

The keen interest in repeat proteins has led to successful protein design of consensus TPR (4), ankyrin (5, 6), and LRR proteins (7) and extensive stability and folding studies on natural ankyrin repeat proteins (a 33-residue repeat that forms a β-turn followed by 2 antiparallel α-helices) (8–14). These studies have begun to dissect the equilibrium cooperativity and kinetic folding pathways of repeat proteins. They show that folding is initiated in the most thermodynamically stable unit, with the subsequent route through the energy landscape to the native state being governed by a competition between the stability of individual repeat units and the interactions between repeats. For example, if repeats within a protein have similar stabilities and weak inter-repeat interactions the protein is more likely to have equilibrium intermediates and to fold via a very malleable pathway (more than 1 folding pathway can be accessed) (8, 9, 15, 16). Recently, these studies have been complemented by folding simulations that show and predict that the cooperativity of repeat protein folding decouples on increasing repeat number or where high local frustration occurs (17–19).

Designed consensus repeat proteins provide an excellent system for investigating the fundamental properties of repeat proteins, because each repeat has identical intra- and inter-repeat interactions. Thus, designed repeat proteins are more structurally symmetric than natural repeat proteins and can easily be extended or shortened by adding or removing whole repeats. This ability provides a unique and exquisitely tunable perturbation that differs radically from normal amino acid mutations and enables a wider exploration of repeat protein folding energy landscapes. For example, dependence of stability, folding rate, cooperativity, and thus folding pathway on repeat number can be explored by engineering a series of proteins that increase in size through addition of identical consensus motifs. In particular, we and the Regan laboratory have used 2 series of designed consensus TPRs (a 34-aa, helix-turn-helix motif), called CTPR and CTPRa proteins, to investigate the dependence of thermodynamic characteristics and denatured state on increasing repeat number (Fig. 1; the 2 series only differ by a 2-aa substitution per repeat) (20–23). These studies have highlighted the 1-dimensionality and changing equilibrium properties of increasing repeat number by showing that the thermodynamic unfolding transition can be described and, importantly, predicted by an Ising model that uses nearest-neighbor interactions within a 1-dimensional lattice (22). Such a description implies that thermodynamically, individual repeats unfold independently of each other and thus can populate partially folded configurations. This observation was confirmed when Cortajarena et al. used equilibrium, NMR-detected hydrogen/deuterium exchange to observe sequential unfolding transitions for 2 CTPR proteins containing 2 and 3 repeats, respectively (20). However, we have also previously shown that kinetic folding of these 2 proteins at 20 °C was 2-state over the limited denaturant concentration range measured (23). These studies have been further complemented by folding simulations showing that although the CTPRs possess very low inter- and intrarepeat frustration, their cooperativity of folding as interpreted by correlation length is roughly 3 repeats (18).

In this study we investigated the kinetic folding cooperativity and folding energy landscapes on increasing repeat protein length by comparing the folding kinetics of the largest series of highly symmetric designed repeat proteins to date. This consists of 7 designed CTPRa proteins that range from 2.5 to 10.5 repeats (86–358 aa). Our results show that as repeat number increases the landscape becomes more complex, with intermediates in-
creasingly populated, and interestingly, suggest that off-pathway intermediates become populated for the largest protein. Thus, although the CTPRa constructs have low intra- and inter-repeat frustration, on increasing repeat number their highly symmetric and modular native topology causes their cooperative kinetic folding to uncouple and eventually misfold to form kinetic traps. Finally, the folding pathway of this unique TPR system is compared with previously published work on the folding of other types of repeat protein (8, 17, 18).

Results

Structure of CTPRa Proteins. The consensus TPR proteins (CTPRa)n were built from arraying multiple copies (n) of a 34-aa idealized sequence with a C-terminal single “solvating” helix (22) [Fig. 1 and supporting information (SI) Fig. S1A]. All proteins are stable (Fig. 2) and adopt the distinctive TPR fold with the unique feature of possessing identical modular structures.

Equilibrium Stability of CTPRa Proteins at 10 °C. Equilibrium chemical denaturations using guanidine hydrochloride (GuHCl) and urea were performed at 10 °C and at pH 7.0. Ellipticity at 222 nm and fluorescence at 340 nm was monitored as a function of denaturant concentration to follow each structural transition (Fig. 2 and Fig. S1D). All of the CTPRa proteins underwent a single reversible transition that corresponded to the concurrent loss of native secondary [far-UV circular dichroism (CD)–α-helical] and tertiary (fluorescence) structure, to a denatured state that lacked both. This was consistent with previously reported data at 20 °C and 25 °C (22, 23).

Data from these denaturations were initially analyzed by individually fitting each protein to a 2-state model (24). This yielded [D]j_{0.5} (midpoint of unfolding) and m_D,N (change in solvent-accessible surface area upon protein unfolding) from which ΔG^{H_2O}_{D,N}/ΔG^{H_2O}_{D-N} (free energy of unfolding in water) was calculated (Table S1). However, previous NMR and equilibrium denaturation studies on such designed TPRs have shown that, despite the apparent cooperative equilibrium unfolding, intermediate states are populated through the denaturation transition (21–23). Our data support this because our m_D,N values do not increase in direct proportion to chain length of the CTPRa proteins. Moreover, differential scanning calorimetry was performed that showed the larger proteins to be no longer thermally unfolded in a 2-state manner (25).

Recently a number of studies have analyzed multistate equilibrium repeat unfolding using an Ising model (22, 26). This has been used, as the repetitive and modular repeat structure of repeat proteins coupled with the systematic variation of transition midpoint and m_D,N mimic the nearest-neighbor coupling of the Ising model. Therefore the data presented here were globally fit using the same Ising model as used when the data were fit at 25 °C (22). This gave x_c (midpoint of unfolding of a single α-helix in the protein), m_1 (denaturant dependence of a single α-helix in the protein), J (the coupling energy between α-helices), from which H (half of the difference in free energy between the folded and denatured states of a single helix in the absence of coupling to its neighbors) (Table S2). From these data ΔG^{H_2O}_{D-N} (the free energy of folding in water for a protein with j α-helices) was calculated for a helix, a repeat (2 helices), 1.5 repeats (3 helices) (–2.4 kcal mol\(^{-1}\), –1.0 kcal mol\(^{-1}\), and 0.4 kcal mol\(^{-1}\), respectively), and each CTPRa protein (Table 1 and Table S2). All of the values of ΔG^{H_2O}_{D-N}/ΔG^{H_2O}_{D-N} calculated are within error for urea and GuHCl denaturations. As can be seen, the Ising analysis correctly shows that individual helices or repeats are not stable/unfolded at 10 °C. To obtain a folded and stable protein at least 1.5 consecutive repeats are required [as shown previously (23, 27)]. It is interesting to note that for CTPRa2 and CTPRa3 the ΔG^{H_2O}_{D-N}/ΔG^{H_2O}_{D-N} from 2-state and Ising models are within error, but as the proteins increase in size there is a disparity between results, as one might expect when intermediates are present.
Nonlinearity of both ln \( \Delta G \) and trend of the CTPRa chevrons is the changing that fit well to single exponential equation (Fig. S2). Rates of refolding and unfolding kinetics were best described by a monophasic process within these ranges all of the CTPRa proteins were observed to be very rapid, therefore experiments were performed at 10 °C to expand the range of GuHCl concentration. Although the smallest protein in the series, the tryptophan and tyrosine residues found in each repeat. The unfolding and refolding kinetics of the repeat protein constructs were measured before any rollover, increase slightly before becoming equally fast. Effects of adding of ankyrin repeats to ankyrin-containing TPR motifs have on the rates of refolding and unfolding and trends are immediately apparent from studying the chevrons. These can be broken down into the effects that the addition of TPR motifs have on the rates of refolding and unfolding and the relationship of ln \( \Delta G \) as a function of GuHCl concentration. Rates of refolding and unfolding. It is obvious that the main effect of adding TPR motifs is a considerable decrease in the unfolding rate. Thus, the increase in stability of each protein on increasing repeat number can be attributed to a decrease in the rate of unfolding. This was consistent with previously reported data for 2 designed TPRs at 20 °C (23) and agrees well with studies on the effects of adding of ankyrin repeats to ankyrin-containing proteins (8, 28). In comparison, the refolding rates, when measured before any rollover, increase slightly before becoming equally fast.

Nonlinear relationship of ln \( k_{obs} \) as a function of [GuHCl]. The most striking feature and trend of the CTPRa chevrons is the changing nonlinearity of both ln \( k_f \) and ln \( k_u \) as a function of GuHCl concentration. Although the smallest protein in the series, CTPRa2, exhibits a linear dependence for both refolding and unfolding over the denatured concentration range measured, which is indicative of 2-state behavior, all of the other proteins in the series do not.

When rollover in the refolding arms of the chevron plots are compared, the nonlinearity observed for each progressively bigger CTPRa construct becomes more pronounced until the rollover is no longer a smooth curve but is kinked (Fig. 3 and Fig. S3). The slope of the refolding arm after the kink for the smaller proteins is negative (as one would expect), yet as the repeats become larger the slope becomes flat and for the largest protein (CTPRa10) positive (Fig. 3, Fig. S3, and Table S3). It is interesting to note that the point at which the refolding limb for each protein kinks is essentially the same and coincides with the linear refolding arm of CTPRa2. In a similar manner to the refolding kinetics, the unfolding arms of CTPRa3 to CTPRa10 chevrons have a downward curvature that becomes more curved with increasing repeat number. However, the scale of the rollover and the kinks is less severe.

Unfolding/Refolding Kinetics of CTPRa Proteins. The unfolding and refolding kinetics of CTPRa2 to CTPRa10 were measured using stopped-flow fluorescence spectroscopy as a function of GuHCl concentration. This follows the changes in the fluorescence of the tryptophan and tyrosine residues found in each repeat. The unfolding and refolding kinetics of the repeat protein constructs were observed to be very rapid, therefore experiments were performed at 10 °C to expand the range of GuHCl concentration over which data could be collected. This permitted the refolding of CTPRa4 to CTPRa10 to be followed down to 0 M GuHCl, with CTPRa2 and CTPRa3 measured down to 0.7 M and 0.54 M GuHCl, respectively. Within these ranges all of the CTPRa protein kinetics were best described by a monophasic process that fit well to single exponential equation (Fig. S2). Typical kinetic traces for unfolding and refolding, respectively, overlaid with the unfolded or folded baselines). The natural logarithms of the observed rate constants measured as a function of GuHCl concentration for each CTPRa protein are shown as chevrons plots in Fig. 3 and Fig. S3. A number of striking features and trends are immediately apparent from studying the chevrons. These can be broken down into the effects that the addition of TPR motifs have on the rates of refolding and unfolding and the relationship of ln \( k_{obs} \) as a function of GuHCl concentration.
Whereas previous studies have shown that such nonlinearities can be caused by transient aggregation, ionic effects, and instrumental dead-time (29–31), here they are not. The refolding rate constants for each CTPRa protein were found to be independent of protein concentration over a 100-fold range (measured at 0 M and 0.54 M GuHCl by both pH and [GuHCl] jump experiments, 0.1 μM to 10 μM, and shown in Fig. S3), independent of ionic effects (urea denaturations gave similar rollover to GuHCl denaturations; Fig. S1E) and not limited by instrumental dead-time (folding rates recorded for CTPRa2 are faster than for the curved chevrons of larger CTPRa proteins). Thus, all observed changes in slope of the chevrons can be related to denaturant-dependent changes in either structure or population of differing states in the folding energy landscape.

Confirmation of Significantly Populated Intermediate States During Refolding. To confirm that the curvature observed in the refolding plots of the larger CTPRa proteins is due to the population of intermediate states, refolding experiments were performed with the hydrophobic dye 1-anilinonaphthalene-8-sulfonate (ANS). ANS is known to bind to exposed clusters of hydrophobic residues frequently found in transient folding intermediates (32). On binding to such a hydrophobic surface, ANS undergoes a large change in fluorescence and so is a sensitive probe of the formation of transient species during folding (32). Therefore, refolding experiments in the presence of ANS at a final concentration of 0.54 M GuHCl were performed. Under these conditions, if the curvature observed in Fig. 3 is due to the population of intermediate states, the intermediates should bind ANS and show a change in fluorescence. This is observed for all of the larger CTPRa proteins with significant curvature (CTPRa4 to CTPRa10; Fig. S2 C and D and Table S4). In contrast, CTPRa2 and CTPRa3 do not. This is expected for CTPRa2, because the refolding arm of its chevron is linear over the denatured concentration range measured. The lack of signal from CTPRa3 might be caused by its rapid folding kinetics at 0.54 M GuHCl (350 s^{-1} ± 40). When a higher final GuHCl concentration was used on each CTPRa protein (corresponding to the linear refolding region of each chevron) no signal was observed (Fig. S2 E and F), as expected. The lack of signal at higher GuHCl concentration reflects the fact that the intermediate state is destabilized relative to the denatured state with increasing concentrations of GuHCl.

Characterization of Folding Energy Landscapes. To analyze the energy landscapes of the CTPRa proteins the kinetic data were fit in 2 ways: (i) CTPRa2: a 2-state folding scheme because no ANS binding was observed; and (ii) CTPRa3 to CTPRa10: a 3-state folding scheme because curved chevron plots and/or ANS binding was observed (thus populating refolding intermediates).

Two-state folding of CTPRa2. The data were fitted to a linear 2-state model of folding (SI Appendix Eq. 6). The extrapolated $k_\text{H}_2\text{O}$ and $k_\text{U}$ from the fit were then used to calculate a kinetic $\Delta G_\text{H}_2\text{O}^{\text{U}}$ (Table 1 and Table S3). This was consistent with $\Delta G_\text{U-N}$ and $\Delta G_\text{H}_2\text{O}^{\text{U}}$ calculated from equilibrium data, showing that in the range measured CTPRa2 folds kinetically in a 2-state manner.

Three-state folding of CTPRa3 to CTPRa10. Because of the kinked nature of the chevron plots, the data could not be fitted accurately to the sum of 2 quadratic equations (SI Appendix Eq. 7). Therefore, each dataset was fitted to a simple sequential, on-pathway 3-state model whereby the same intermediate is populated in both refolding and unfolding experiments (SI Appendix Eqs. 8–10 and Fig. S1B) (33).

This model is consistent with our ANS results, because it assumes an intermediate is populated in the dead-time of the stopped-flow instrument, causing the rate to be limited (i.e., where the chevron deviates from “2-state” linearity, the refold-
Discussion

Multistate Kinetic Folding Pathways. In this study we have performed a comprehensive characterization of the folding pathway of a series of designed CTPRa proteins by investigating their unfolding and equilibrium denaturations. In particular, our system allowed for full denaturation and thus complete kinetic analysis of 7 proteins (CTPRa2 to CTPRa10) that ranged from 86 aa to 358 aa. The results on these constructs, which differ greatly in size and stability, has enabled us to shed light on the systematic changes and similarities in folding pathway on increasing repeat number.

To a first approximation, the data are consistent with a minimal dead-end scheme whereby a compact off-pathway intermediate species equilibrates with the denatured state in the dead-time of the experiment (Fig. 3B, Fig. S1C, and Table S5). Thus, although positive slopes have been observed for carbonic anhydrase in the absence of salt (36) and S6 in the presence of salt (37); here, we have explicitly shown that, without any additives, repeat proteins can also exhibit this behavior.

Comparison of Kinetics with Equilibrium and Computational Studies. Our analysis shows that the modular multistate equilibrium folding of the CTPRa proteins (20, 23, 38) is also observed in their kinetic folding and unfolding. Excitingly, our results are also completely consistent with the 2 computational studies on the folding CTPR proteins. First, they mirror the simulations that predict that the kinetic folding pathway of CTPR proteins should be multistate (17, 18). Second, they are consistent with the sequential process obtained from computational simulations by Hagai and Levy (17). These show that as CTPR proteins increase in repeat number, the occurrence of independently folding intermediates (formed from consecutive folded repeats) increases. Third, our results show that multistate kinetics are only observed when a CTPRa construct contains at least 3 CTPR motifs, and Ferreiro et al. (18) showed that the folding cooperativity of the larger CTPR proteins, defined by correlation length, is roughly 3 repeats.

Comparison with Ankyrin Repeat Folding Studies. Although there have been a number published studies on the folding of ankyrin proteins (11, 13), there are few that have characterized such large repeat proteins (9, 10, 15, 16) or compared such a range of repeat protein sizes (8, 28). However, these studies coupled with studies on smaller ankyrin proteins have elegantly shown that folding is
controlled through the thermodynamic interplay between the stability of individual repeats and the interactions between repeats. In general, each natural ankyrin protein seems to have a stable core of a repeat or a few repeats that initiates folding and around which the rest of the structure condenses. Whether folding proceeds with no intermediates being populated, a populated kinetic intermediate, or an intermediate populated even under equilibrium conditions seems to be determined by the stability and size of the core compared with the rest of the protein. The results for CTPRa presented here agree well with this scheme and highlight how the difference in structure between natural repeats and our designed constructs affects their folding (i.e., the CTPRa proteins are built from identical repeats and therefore have no repeat/unit that is more stable or less stable than the rest of the protein). This difference manifests itself in the ready accumulation of intermediates along the folding and unfolding pathway and, when the protein is large enough, an off-pathway intermediate. It is interesting to speculate that Nature may have evolved away from complete consensus proteins, not only for binding, but also to avoid such kinetic traps.

**Conclusions**

Designed TPR motifs have unique properties: identical structural modularity, sequence simplicity, and modular linear structure. We have exploited these properties to explore their folding landscapes by elegantly using the perturbation of adding and subtracting whole repeat motifs. This facilitates a wider exploration of repeat protein folding energy landscapes. Initially on increasing the repeat number, on-pathway intermediates are kinetically populated. Subsequent addition of modular repeats causes off-pathway kinetic traps to predominate. Thus, although proteins constructed of identical repeats have low local frustration, their modular and symmetric structures produce energy landscapes that are prone to kinetic traps. These results complement the existing notion of repeat proteins having very malleable folding pathways and highlight how the structure of consensus repeats affects their folding pathway.

**Materials and Methods**

**Cloning, Protein Production, and Purification.** The designed CTPRa proteins were cloned, expressed, and purified as previously described (22).

**Equilibrium Experiments.** Fluorescence and far-UV CD equilibrium unfolding measurements were performed and analyzed as described in detail in *SI Appendix*.

**Kinetic Experiments.** All experiments were performed as described in detail in *SI Appendix*. In brief, both unfolding and folding phases fitted well to a single-exponential process. No slow, proline isomerization phases were observed in the refolding experiments over a 200-s time scale. It is quite possible that slower phases exist but are difficult to detect owing to instrumental drift.

**Data Analysis.** The dependence of ln koff on [denaturant] for each CTPRa protein was fitted to either a 2-state model or a sequential 3-state model whereby an intermediate is either on pathway (Fig. 5B) or off pathway (Fig. 5C). For full details of equations used see *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank Dr. Sophie Jackson, Dr. Alan Lowe, and Prof. Lynne Regan for insightful discussion, helpful comments, and suggestions; and members of the Department of Chemistry and Biochemistry (University of Sussex), Dr. Lowe, Dr. Jackson, and Dr. Phillips, for critical reading of the manuscript. Y.J. was supported by an Engineering and Physical Sciences Research Council studentship. This work was partly funded by Biototechnology and Biological Sciences Research Council Grant BB031871.

---