

A New Folding Paradigm for Repeat Proteins

Tommi Kajander,[†] Aitziber L. Cortajarena,[†] Ewan R. G. Main,^{†,‡} Simon G. J. Mochrie,^{*,†,§} and Lynne Regan^{†,||}

Departments of Molecular Biophysics and Biochemistry, Physics and Applied Physics, and Chemistry, Yale University, New Haven, Connecticut 06520, Department of Chemistry, University of Sussex, Falmer, BN1 9QG, U.K.

Received April 14, 2005; E-mail: simon.mochrie@yale.edu

How a protein's amino acid sequence specifies its structure and properties stands as a grand challenge of the post-genomic era. Repeat proteins,^{1–5} which are composed of tandem arrays of a basic structural motif, account for more than 5% of the proteins in multicellular organisms in the Swiss-Prot database. In addition, leucine-rich repeats, zinc finger repeats, ankyrin repeats, and tetratricopeptide repeats (TPRs)² all rank among the 20 most common protein folds in the Pfam database. It is therefore surprising that the folding of repeat proteins has been little studied,³ especially because their modular, repetitive structures promise a more tractable folding problem than for globular proteins. Here, we demonstrate that the folding of TPR proteins can be quantitatively described by the classical one-dimensional Ising model,^{6,7} which thus represents a new folding paradigm for repeat proteins. Moreover, for the first time, a theoretical model predicts protein stability in detail.

Our approach has been to synthesize and then examine the structure and behavior of a series of designed proteins containing different numbers of an identical repeated unit, which is a consensus sequence based on the natural prevalence of each amino acid at each position in the TPR motif.⁴ We have determined the crystal structure of such a protein, CTPRa8*, which contains eight identical consensus TPR repeats and which is shown in Figure 1A,B. As may be seen from the figure, each repeat is composed of two helices, which are arrayed to form a superhelix. A key feature of this structure, and those of repeat proteins in general,⁵ is that, in contrast to globular proteins, there are no sequentially distant amino acid contacts.² This is illustrated in Figure 1C, which shows a contact map for CTPRa8*, making it clear that CTPRa8* exhibits extensive amino acid contacts only within a helix and between nearest-neighbor helices.

This observation suggests that it may be possible to understand the stability of TPRs on the basis of the collective behavior of the individual helices, interacting with each other via nearest-neighbor interactions. Indeed, as we show in the present communication, the folding/unfolding transitions within a series of consensus TPRs are quantitatively well described by the classical one-dimensional Ising model.^{6,7} According to this description, the TPRs' constituent helices correspond to Ising spins ($s_i = \pm 1$) and interact via a nearest-neighbor coupling. Thus, spin up ($s_i = +1$) in the Ising model corresponds to the folded state of a TPR helix, while spin down ($s_i = -1$) corresponds to the unfolded state. It follows that folding/unfolding of TPRs, and likely of all repeat proteins,⁸ does not conform to the all-or-nothing, folded-or-unfolded, two-state transition that is generally assumed for small globular proteins.⁹ Instead,

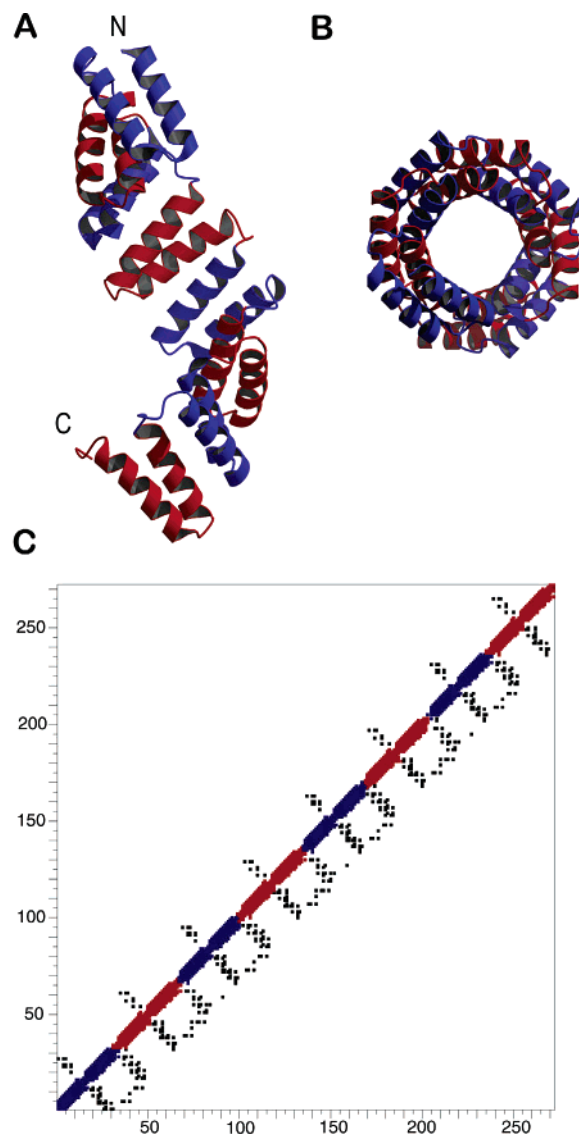


Figure 1. Crystal structure of CTPRa8*. (A) View perpendicular to the superhelical axis. Each TPR repeat is colored either red or blue. (B) View along the superhelical axis. (C) Contact map for CTPRa8*. The axes correspond to the residue numbers in the protein sequence. A square is placed at each position where two residues lie within 3–5 Å of each other in the structure. Therefore, points near the diagonal represent local contacts, while points far from the diagonal correspond to sequentially distant contacts. Contacts between backbone atoms are given above the diagonal, and contacts between all atoms are given below the diagonal. The diagonal is color-coded according to (A) and (B).

the Ising description prescribes the existence of partially folded configurations with significant statistical weight.

[†] Department of Molecular Biophysics and Biochemistry, Yale University.

[‡] University of Sussex.

[§] Departments of Physics and Applied Physics, Yale University.

^{||} Department of Chemistry, Yale University.

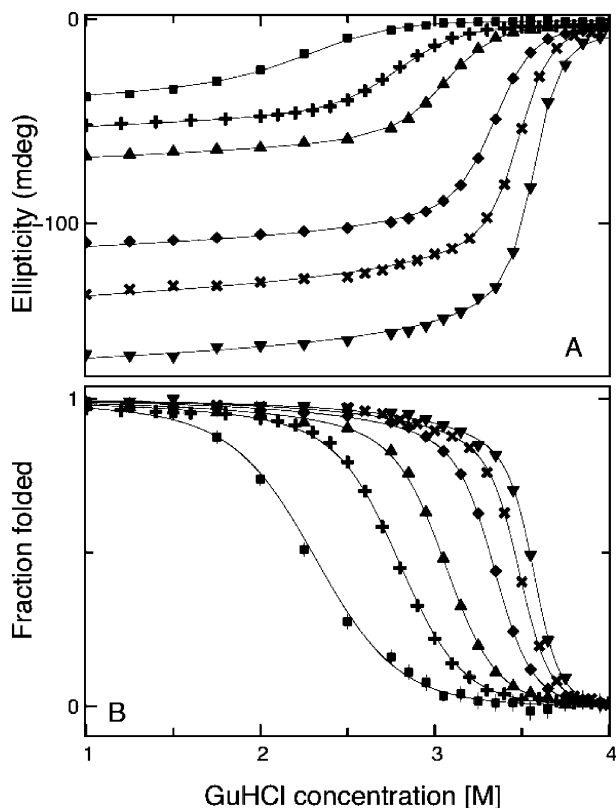


Figure 2. Thermodynamics of CTPRa unfolding. (A) Measured ellipticity and (B) fraction folded vs [GuHCl] for CTPRa2 (■), CTPRa3 (+), CTPRa4 (▲), CTPRa6 (◆), CTPRa8 (×), CTPRa10 (▼) in 50 mM sodium phosphate, pH 6.5, 150 mM sodium chloride, at 25 °C. Solid lines correspond to the best fits to the theoretical description, described in the text, based on the one-dimensional N-spin Ising model. Analogously to the Ising–Zimm–Bragg treatment of the polypeptide helix–coil transition,^{6,7} in which individual amino acids are cast as Ising spins, here our model ascribes a reduced stability to the endmost helices, relative to those not at the end, by incorporating interactions with fictitious down spins at positions 0 and $N + 1$.

The Ising model free energy is

$$G = k_B T \sum_i (-J s_i s_{i+1} - H s_i) \quad (1)$$

The parameters of the model are: H , which is $1/(k_B T)$ multiplied by one-half of the difference in free energy between the folded and unfolded states of a single helix in the absence of coupling to its neighbors; J , which specifies the coupling between neighboring helices; and N , which is the number of helices. We take $H = m_1(x - x_c)/2$,^{10,11} where x is the concentration of guanidium hydrochloride ([GuHCl]), x_c is the value of [GuHCl] at which $H = 0$, and m_1 is assumed to be independent of x .

How can we test this hypothesis? Because the variation of thermodynamic behavior for different numbers of identical, coupled subunits is a textbook signature of collective effects,⁷ the regularity and simplicity of CTPRs and the possibility of creating them in various multiplicities render them an ideal system in which to test this model of repeat protein stability. Therefore, we constructed a series of CTPRs, named CTPRa n , where n is the number of TPR repeats, for which the number of helices ($N = 2n + 1$) varies between 5 and 21. We have characterized the stability of these proteins as a function of GuHCl concentration and temperature in 50 mM sodium phosphate, pH 6.5, 150 mM sodium chloride, at 25 °C, via circular dichroism (CD) and fluorescence measurements. Figure 2A depicts the measured ellipticity as a function of guanidinium hydrochloride concentration ([GuHCl]) for CTPRa2,

CTPRa3, CTPRa4, CTPRa6, CTPRa8, and CTPRa10. The rapid change in each profile with increasing GuHCl concentration corresponds to the transition from the folded to the unfolded state.

Each transition in Figure 2 can be individually quite well described by the usual two-state model (fits not shown). However, it is immediately apparent from these data that, as the number of repeats increases, there is a systematic variation of both the transition midpoint, which occurs at increasing GuHCl concentrations for increasing numbers of helices, and the slope of the unfolding curve, which also increases for increasing numbers of repeats. Qualitatively, these observations mimic what is expected from an Ising description.^{6,7} However, to quantitatively investigate the applicability of the Ising model, it is necessary to directly fit its predictions to the experimental data of Figure 2. To this end, we have taken the ellipticity for CTPRa n ($\Delta\theta_n$) to be given by

$$\Delta\theta_n = (a_n + b_n x)f + (c_n + d_n x)(1 - f) \quad (2)$$

where $x = [\text{GuHCl}]$, a_n , b_n , c_n , and d_n are parameters that describe the [GuHCl] dependence of the ellipticity (assumed to vary linearly in the folded and unfolded states) and $f = f(H, J)$ is the fraction of the protein that is folded, according to the Ising model.^{6,7}

Within the Ising description, the thermodynamic behavior of all CTPRs must be given by the same values of J , x_c , and m_1 . Therefore, we have fitted all of the data of Figure 2A simultaneously, using a common value of each of J , x_c , and m_1 as fitting parameters, in addition to the quantities a_n , b_n , c_n , and d_n for each profile. The best fits obtained this way are shown as the solid lines in Figure 2A. Clearly, the model provides an excellent description of the experimental ellipticity for all of the proteins studied. This is highlighted in Figure 2B, which plots the corresponding fraction folded. The best-fit parameter values are $x_c = 3.82 \pm 0.01$ M, $J = 1.90 \pm 0.02$, and $m_1 = 0.96 \pm 0.01$ M⁻¹ with goodness-of-fit parameter $\chi^2 = 1.6$. The economy of just three fitting parameters may be contrasted with the 12 parameters (a transition midpoint and a midpoint slope for each of the six unfolding curves) that would be needed to describe the behavior of each of the six proteins seen in Figure 2B as an independent two-state transition.

Because CTPRs are composed of alternating helices that differ in sequence, one might expect that a more elaborate model, which incorporates different stabilities for the different helices, would be more appropriate. However, we estimate that the two helices of CTPRa are of similar stability,¹² and it can be shown¹³ that, for alternating helices with similar stabilities, the predicted thermodynamic behavior is difficult to distinguish from that of the simpler model employed here. Because the simpler model evidently describes the data well, it is the one we have chosen to use.

How general is the behavior of Figure 2? The Ising model also accurately reproduces the systematics of CTPRa folding/unfolding as a function of temperature ($T_c = 87 \pm 0.2$ °C, $J = 1.64 \pm 0.02$, and $m_1 = 0.049 \pm 0.001$ °C⁻¹ with $\chi^2 = 1.5$). In addition, we have characterized the folding behavior of several members of a different series of designed TPRs (CTPR1, CTPR2, and CTPR3^{4,14}) in 50 mM sodium phosphate, pH 6.8, 150 mM NaCl. The Ising model provides an excellent description in this case too, with best-fit parameters $x_c = 5.44 \pm 0.03$ M, $J = 3.27 \pm 0.05$, and $m_1 = 0.91 \pm 0.01$ M⁻¹ with $\chi^2 = 1.4$. Whether characterized via CD or fluorescence (data not shown), the folding/unfolding behavior of CTPRs is well fit by the Ising model.

Evidently, the Ising model describes the thermodynamics of TPR folding/unfolding under a variety of circumstances. But what really has been gained? First, the Ising model is predictive. Because thermodynamic data from just two different CTPRs in a series are sufficient to determine J , x_c , and m_1 , the Ising model quantitatively

predicts the behavior of all additional CTPRs in that series. Thus, we may view four of the model profiles shown in Figure 2 as predictions. Remarkably, these stand as the first examples where one can successfully predict not only the stabilities of proteins but also the shapes of their unfolding curves.

Second, the Ising model requires a new microscopic picture: In the usual two-state transition, a protein is essentially always either completely folded or completely unfolded with only brief transient behavior. By contrast, the Ising description implies that near the transition midpoint, partially folded configurations occur with significant probability. For example, for CTPRa3 at a GuHCl concentration just below the midpoint of the unfolding transition, about one-quarter of the molecules are partially unfolded. Because of the overall structural similarities among different repeat proteins (repetition of a basic structural motif and sequentially localized contacts) we believe that the Ising model treatment will be widely applicable to repeat proteins in general⁸ and thus represents a new folding paradigm. Of course, for natural repeats, we may expect that each corresponding Ising spin will have a different intrinsic free energy (H_i) and a different coupling (J_{ij}) with its neighbors.

Beyond repeat proteins, does the behavior of CTPRa tell us anything about the folding of globular proteins? The folding of several small, single-domain globular proteins has been well-studied. As in the case of individual repeat proteins, the thermodynamic behavior of small globular proteins can usually be well-described by the traditional two-state model. However, in many cases, elegant hydrogen-exchange measurements clearly reveal a richer free-energy landscape than simply two free energy minima.^{15,16} For example, T4 lysozyme clearly shows two main subdomains with different stabilities.¹⁶ Beyond this division into two parts, the authors of ref 16 note that, within each smaller element of secondary structure, the residue-specific stability is clustered, indicating that it may be sensible to envision each element of secondary structure itself as a degree of freedom that can be either folded or unfolded (i.e., as an Ising spin). In light of such observations, which suggest that there are important similarities between the thermodynamics of repeat proteins, elucidated here, and of globular proteins, we may hope that, in future studies, the simplicity and regularity of repeat proteins

will facilitate insights that are applicable not only to repeat proteins but also to the folding of globular proteins.

Acknowledgment. We are grateful to Joana Tala for skilled technical assistance in the early stages of this work and to Jeff Knight for his advice and help with the mass spectroscopy analyses. We thank Max Paoli, Nick Read, and Jimin Wang for insightful discussions. We are also grateful to Vivian Stojanof and staff at X6A at the National Synchrotron Light Source (NSLS) for their invaluable assistance. The NSLS is supported by the U.S. Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, under contract No. DE-AC02-98CH10886. T.K. acknowledges support from the Helsingin Sanomat Centennial Foundation (Finland). A.L.C. is a recipient of a postdoctoral fellowship from the Spanish Ministry of Education, Culture and Sports.

Supporting Information Available: Detailed materials and methods. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Andrade, M. A.; Perez-Iraxeta, C.; Ponting, C. P. *J. Struct. Biol.* **2001**, *134*, 117–131.
- (2) D'Andrea, L.; Regan, L. *Trends Biochem. Sci.* **2003**, *28*, 655–662.
- (3) Tang, K. S.; Fersht, A. R.; Itzhaki, L. *Structure* **2003**, *11*, 67–73.
- (4) Main, E. R. G.; Xiong, Y.; Cocco, M. J.; D'Andrea, L.; Regan, L. *Structure* **2003**, *11*, 497–508.
- (5) Das, A. K.; Cohen, P. W.; Barford, D. *EMBO J.* **1998**, *17*, 1192–1199.
- (6) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526–535.
- (7) Nelson, P. *Biological Physics: Energy, Information, Life*; W. H. Freeman: New York, 2003.
- (8) Mello, C. C.; Barrick, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *99*, 16029–16034.
- (9) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428–10435.
- (10) Ahmad, F.; Bigelow, C. C. *J. Biol. Chem.* **1982**, *257*, 12935–12938.
- (11) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266–280.
- (12) Munoz, V.; Serano, L. *Biopolymers* **1997**, *41*, 495–509.
- (13) Montroll, E. W.; Goel, N. S. *Biopolymers* **1966**, *4*, 855–878.
- (14) Main, E. R. G.; Stott, K.; Jackson, S. E.; Regan, L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5721–5726.
- (15) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192–197.
- (16) Llinas, M.; Gillespie, B.; Dahlquist, F. W.; Marqusee, S. *Nat. Struct. Biol.* **1999**, *6*, 1072–1078.

JA0524494