Please cite this article in press as: Tsytlonok et al., Complex Energy Landscape of a Giant Repeat Protein, Structure (2013), http://dx.doi.org/10.1016/j.str.2013.08.028

Structure
Article

Complex Energy Landscape of a Giant Repeat Protein

Maksym Tsytlonok,[1,2] Patricio O. Craig,[3,4] Elin Sivertsson,[2] David Serquera,[1] Sarah Perrett,[5] Robert B. Best,[2,7,*] Peter G. Wolynes,[6,*] and Laura S. Itzhaki[2,*]
[1]MRC Cancer Cell Unit, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 0XZ, UK
[2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
[3]Department of Chemistry and Biochemistry, University of California San Diego (UCSD), 9500 Gilman Drive, La Jolla, CA 92093-0374, USA
[4]Center for Theoretical Biological Physics (CTBP), University of California San Diego (UCSD), 9500 Gilman Drive, La Jolla, CA 92093-0374, USA
[5]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China
[6]Center for Theoretical Biological Physics and Department of Chemistry, Rice University, Houston, TX 77251-1892, USA
[7]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA
*Correspondence: robertbe@helix.nih.gov (R.B.B.), pwolynes@rice.edu (P.G.W.), lsi10@cam.ac.uk (L.S.I.)
http://dx.doi.org/10.1016/j.str.2013.08.028

SUMMARY

Here, we reveal a remarkable complexity in the unfolding of giant HEAT-repeat protein PR65/A, a molecular adaptor for the heterotrimeric PP2A phosphatases. The repeat array ruptures at multiple sites, leading to intermediate states with noncontiguous folded subdomains. There is a dominant sequence of unfolding, which reflects a nonuniform stability distribution across the repeat array and can be rationalized by theoretical models accounting for heterogeneous contact density in the folded structure. Unfolding of certain intermediates is, however, competitive, leading to parallel unfolding pathways. The low-stability, central repeats sample unfolded conformations under physiological conditions, suggesting how folding directs function: certain regions present rigid motifs for molecular recognition, whereas others have the flexibility with which to broaden the search area, as in the fly-casting mechanism. Partial unfolding of PR65/A also impacts catalysis by altering the proximity of bound catalytic subunit and substrate. Thus, the repeat array orchestrates the assembly and activity of PP2A.

INTRODUCTION

Proteins repeatedly fold and unfold during their lives in the cell. For example, partly folded states are populated both at the start of a protein's life (cotranslational folding, intracellular trafficking) and at the end (proteasomal degradation); in between, small and large fluctuations of the structure are required to carry out functions such as catalysis, signal transduction, allostery, and mechanical work. The relative stabilities of the different conformations must be tightly controlled to prevent the population of problematic species that lead to malfunction or misfolding. This multitude of states characterizes the so-called "energy landscape" of a protein. Despite many years of research, experimental folding studies are dominated by small proteins, which usually populate only the fully folded and unfolded states. We have turned to a special class of proteins known as tandem-repeat proteins in order to view the whole landscape. They are composed of small motifs (20–40 amino acids), arrays of which pack in a roughly linear fashion to produce elongated, superhelical architectures (Kobe and Kajava, 2000). Their structures comprise only short-range interactions between residues within a repeat or in adjacent repeats, distinguishing them from the topologically complex structures of globular proteins. Herein, we exploit this simplicity and modularity to decipher the folding of a giant repeat protein, revealing the full energy landscape with an exceptionally high level of detail.

The smaller repeat motifs (e.g., ankyrin and tetratricopeptide) have been the main focus of folding studies to date (reviewed in Barrick et al., 2008; Javadi and Itzhaki, 2013). The 39-residue, α-helical *h*untingtin, elongation factor 3, PP2*A* subunit and the lipid kinase *T*OR (HEAT) motif is much larger than these, and it is usually found in giant repeat arrays. The absence of sequence-distant contacts is thought to afford repeat proteins inherent flexibility, and crystallographic data and molecular dynamics simulations hint at how such a property may be employed in the function of large repeat arrays like HEAT-repeat proteins, particularly given their likely exposure to mechanical stress in vivo (Forwood et al., 2010; Grinthal et al., 2010). However, experimental information has been lacking to date. The 590-residue protein PR65/A, comprising 15 HEAT repeats, is the scaffolding subunit of the heterotrimeric serine/threonine protein phosphatase 2A (PP2A) (Figure 1A). PP2A regulates numerous cellular signaling pathways, and this diversity derives from the fact that the enzyme is assembled from different combinations of a catalytic C subunit and one of a large number of regulatory B subunits, connected via the PR65/A subunit (Figure 1A) (Shi, 2009); the B subunit, by binding the substrate, determines the specificity of the enzyme. As well as regulatory subunits, PR65/A also binds a number of viral oncoproteins, which thereby subvert PP2A specificity and activity in some cancers (Mumby, 1995). Despite the importance of PP2A, we still do not fully understand the mechanisms by which its activity is regulated.

Here, we elucidate the unfolding mechanism of PR65/A using a combination of experimental and computational approaches.

Please cite this article in press as: Tsytlonok et al., Complex Energy Landscape of a Giant Repeat Protein, Structure (2013), http://dx.doi.org/10.1016/j.str.2013.08.028

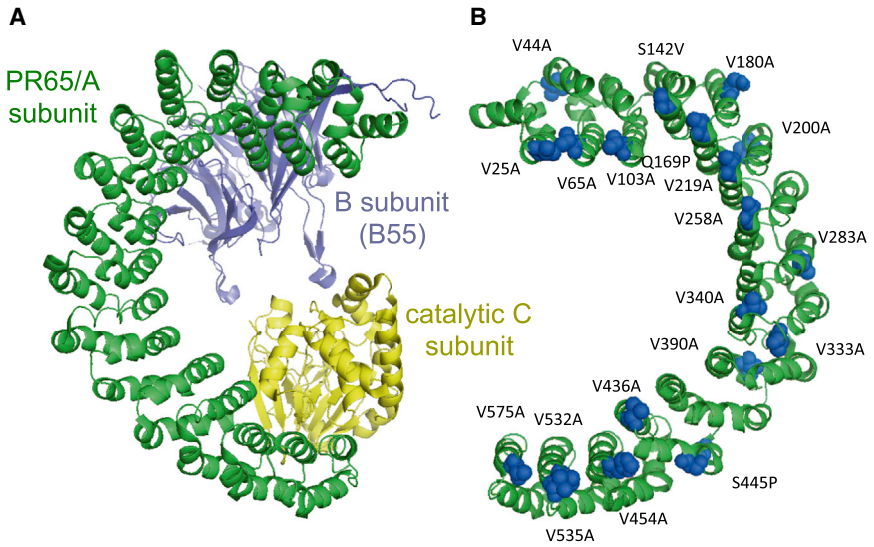

Figure 1. Structure and Stability of the PR65/A Subunit of PP2A

(A) The HEAT-repeat scaffolding subunit PR65/A is in green; the catalytic subunit is in yellow; and the regulatory subunit B55 is in purple (Protein Data Bank [PDB] ID code 3DW8; image generated using PyMOL; http://www.pymol.org).

(B) Structure of PR65/A, with the mutations introduced in this study highlighted in blue, is shown.

We show that the HEAT array ruptures at multiple sites, rather than unfolding propagating from one end of the array as observed for small repeat proteins, leading to an equilibrium intermediate consisting of folded subdomains (HEAT 1–2 and HEAT 11–13) that are noncontiguous in the sequence. Kinetic analysis enables us to delineate a network of partly folded intermediates and to define both sequential and parallel pathways that connect them. The folding/unfolding of the repeat-protein architecture has been likened to driving along a highway, and PR65/A nicely illustrates this analogy: the reaction can only propagate in a one-dimensional manner and is therefore very sensitive to retardation by intermediate species akin to a traffic jam; the longer the highway, the greater the potential for traffic to slow progress (Ferreiro and Wolynes, 2008; Levy et al., 2007). We can destabilize these intermediates by truncation or mutation, smoothing the energy landscape sufficiently in one construct that it unfolds in a single, rapid step. Folding and function may be coupled in the HEAT array of PR65/A, and Wolynes and colleagues have used a fly-casting analogy to describe a similar type of behavior in other proteins (Shoemaker et al., 2000; Trizac et al., 2010); HEAT 1–2 and HEAT 11–13 constitute the rigid core of PR65/A that is important for the recognition of partner proteins, whereas the less-stable HEAT 3–10 and HEAT 14–15 provide the flexibility with which to widen the search area, coordinating the binding of PP2A regulatory B-subunits and catalytic C-subunit at opposite ends of the HEAT array and thereby the timely dephosphorylation of many different substrates. Furthermore, the unfolding of the low-stability central HEAT repeats of PR65/A, which occurs under physiological conditions, will alter the proximity of the catalytic C-subunit to the B-subunit-bound substrate and thereby will directly impact on catalysis. As proposed previously (Grinthal et al., 2010), this dynamic may be particularly important for processive dephosphorylation of multiply phosphorylated and intrinsically disordered substrates, such securin and Tau. Thus, within PP2A, PR65/A is a flexible molecular adaptor, partial unfolding of which will modulate both assembly and catalytic mechanism of the enzyme.

RESULTS

PR65/A Unfolds at Equilibrium via a Hyperfluorescent Intermediate

PR65/A has five tryptophan residues at positions 140 (HEAT 4), 257 (HEAT 7), 417 (HEAT 11), 450 (HEAT 12), and 477 (HEAT 13). The protein unfolds reversibly in a range of different buffers, as monitored by fluorescence. Using a buffer of 50 mM MES at pH 6.5 and a temperature of 25°C, we obtained a denaturation profile in which two transitions are well resolved (Figure 2). Similar profiles were observed regardless of excitation wavelength (280 or 295 nm). The data can be fitted to a three-state model, in which the first transition is from the native state to a hyperfluorescent, red-shifted intermediate species, and the second is from the intermediate to unfolded state (Table S1A available online).

There is no change in the denaturation profile after different incubation times, and the unfolding- and refolding-denaturation curves superimpose yielding the same midpoints and *m*-values within error, indicating reversibility (Figures S1A and S1B). The far-UV circular dichroism (CD)-monitored denaturation curve also shows two transitions (Figure S1D). An intermediate that has higher fluorescence intensity than either of the native or fully denatured states has is populated between ~2 and 4 M urea. The CD spectrum of the intermediate at 2.2 M urea has ~40% of the native ellipticity, and like the native state, it displays the double minima characteristic of α-helical structure. These data are consistent with the approximate five or six HEAT repeats being structured in the intermediate. Further analysis of the intermediate, including its oligomerization behavior, is detailed in the Supplemental Experimental Procedures and Figure S2.

Concomitant Rupture of the HEAT-Repeat Array Occurs at Multiple Distant Sites under Denaturing Conditions

To map out the structure of the equilibrium intermediate, mutations were made of the conserved valine residue at position 24 of the HEAT motifs and of nonconserved valine residues at other positions (Table S1B). The remaining mutations, S142V, Q169P, S445P, and H496P, are mutations to consensus residues, these being natural substitutions that ought to be accommodated by the structure of the repeat. We find that mutations in HEAT repeats 3–7 and HEAT repeats 14–15 shift the midpoint of the first unfolding transition to lower urea concentrations, indicating destabilization of the native state, relative to the intermediate (Figure 2A; Table S1B). Mutations in HEAT repeats 8–10 have little or no effect on the first unfolding transition when monitored by

Structure

Complex Energy Landscape of a Giant Repeat Protein

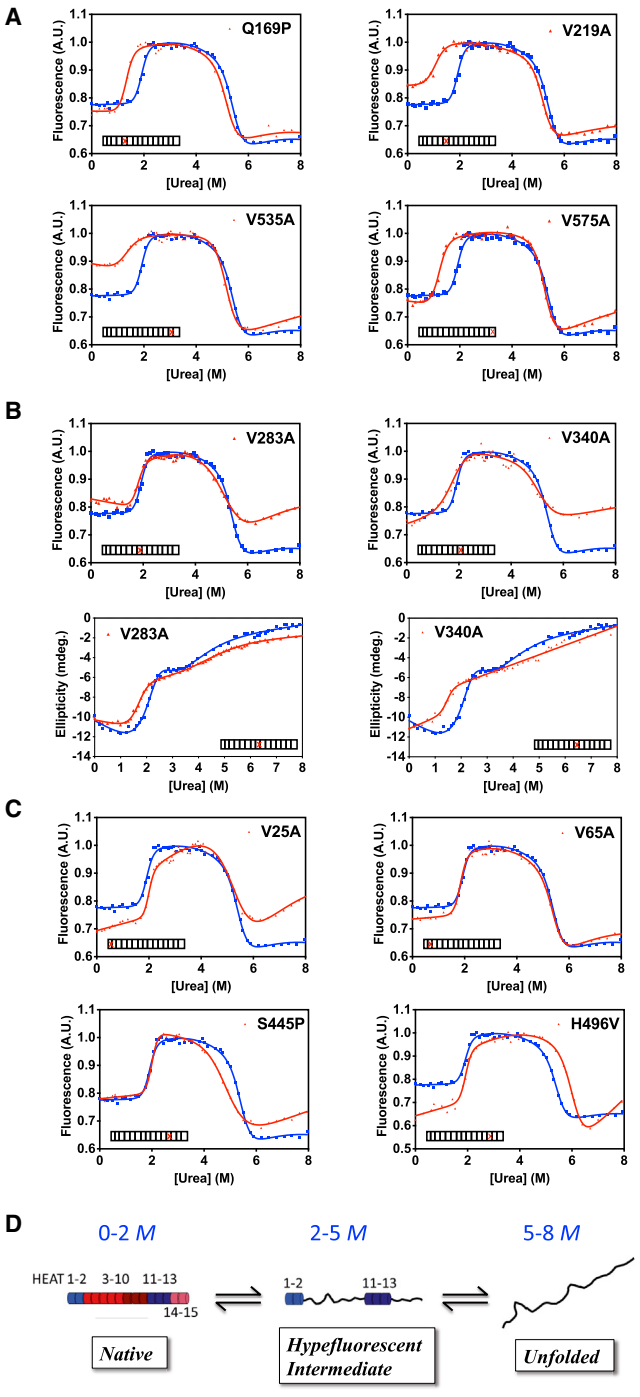

Figure 2. Equilibrium Denaturation Curves of Representative Mutants

(A–C) Denaturation curves of the mutants are shown in red, and for comparison, wild-type is shown in blue.

(A) Effects of mutations in HEAT repeats 3–7 and 14–15, measured by fluorescence are shown.

(B) Effects of mutations in HEAT repeats 8–10, measured by fluorescence and CD; for simplicity, only the first unfolding transition in the CD data is shown.

(C) Effects of mutations in HEAT repeats 1–2 and 11–13, measured by fluorescence are shown. The protein concentration was 0.5 μM for fluorescence and 1 μM for CD.

(D) Schematic of the hyperfluorescent intermediate having structured HEAT repeats 1-2 and 11-13 is shown.

See also Figure S1 and Table S1.

fluorescence; however, destabilization of the native state relative to the intermediate is detected when unfolding of these mutants is monitored by CD (Figure 2B; Table S1C). The lack of fluorescence change associated with unfolding of HEAT repeats 8–10 is likely due to the absence of tryptophan residues in these repeats. Mutations in HEAT repeats 1–2 and 11–13 have little or no effect on the stability of the native state, relative to the intermediate when measured by either fluorescence or CD (Figure 2C; Tables S1B and S1C). The simplest interpretation of the results is that the first transition is associated with the unfolding of HEAT 3–10 and HEAT 14–15, and the second transition is associated with the unfolding of HEAT 1–2 and HEAT 11–13, i.e., HEAT 1–2 and HEAT 11–13 are structured in the hyperfluorescent intermediate, and the other repeats are unstructured (shown schematically in Figure 2D).

Truncations Define the Subdomain Structure and Map the Stability Distribution across the Repeat Array

The one-dimensional-like, modular nature of repeat-protein architectures enables the use of truncations to interrogate the structure and provide a complement to site-directed mutagenesis (Kloss and Barrick, 2009; Mello and Barrick, 2004; Tripp and Barrick, 2004). We truncated the protein at the end of each repeat or in the middle of repeats in those cases in which the truncation at the end of a repeat was not expressed in a soluble form. Several of these truncated variants (subsequently referred to as HEAT 1–13, HEAT 1–10, HEAT 1–9, HEAT 1–8, HEAT 1–5, HEAT 4–15, and HEAT 5–15) were not soluble, and two variants (HEAT 1–6 and HEAT 2–15) were unstable and oligomerized under native conditions. A total of eight truncated variants (HEAT 1–14, HEAT 1–12, HEAT 1–11, HEAT 1–9.5, HEAT 1–8.5 [truncated after the first helix of repeat 8], HEAT 1–7, HEAT 3–15, and HEAT 3–9.5 [truncated after the first helix of repeat 9]) were monomeric and were analyzed further (Tables S1D and S1E).

Although the magnitude of the ellipticity signal at 222 nm could potentially provide information about how much native helical structure is maintained upon truncation, in the case of PR65/A the five tryptophan residues appear to contribute significantly to this ellipticity [as shown by our observation that mutations of the tryptophan residues in full-length PR65/A cause large changes in the 222 nm ellipticity (Figure S3), and the changes are not due to the tryptophan mutants being partially unfolded as gel filtration of these mutants shows that they are fully folded]. Correlating the magnitude of the 222 nm ellipticity signals of the fragments, which contain varying numbers of tryptophan residues, with their helical content, is therefore impossible. Instead, urea-induced unfolding of the variants was used to compare stability and structural integrity across the whole fragment set. A pattern emerges from the CD-monitored denaturation profiles (Figure 3A). Deletion of the most C-terminal repeat (HEAT 1–14) results in a large shift in the first unfolding transition to lower urea concentrations and a large decrease in the apparent *m*-value; upon further truncation (variants HEAT 1–12 and HEAT 1–11), this transition is shifted back to higher urea concentrations; a similar pattern is evident in the fluorescence-monitored

Please cite this article in press as: Tsytlonok et al., Complex Energy Landscape of a Giant Repeat Protein, Structure (2013), http://dx.doi.org/10.1016/j.str.2013.08.028

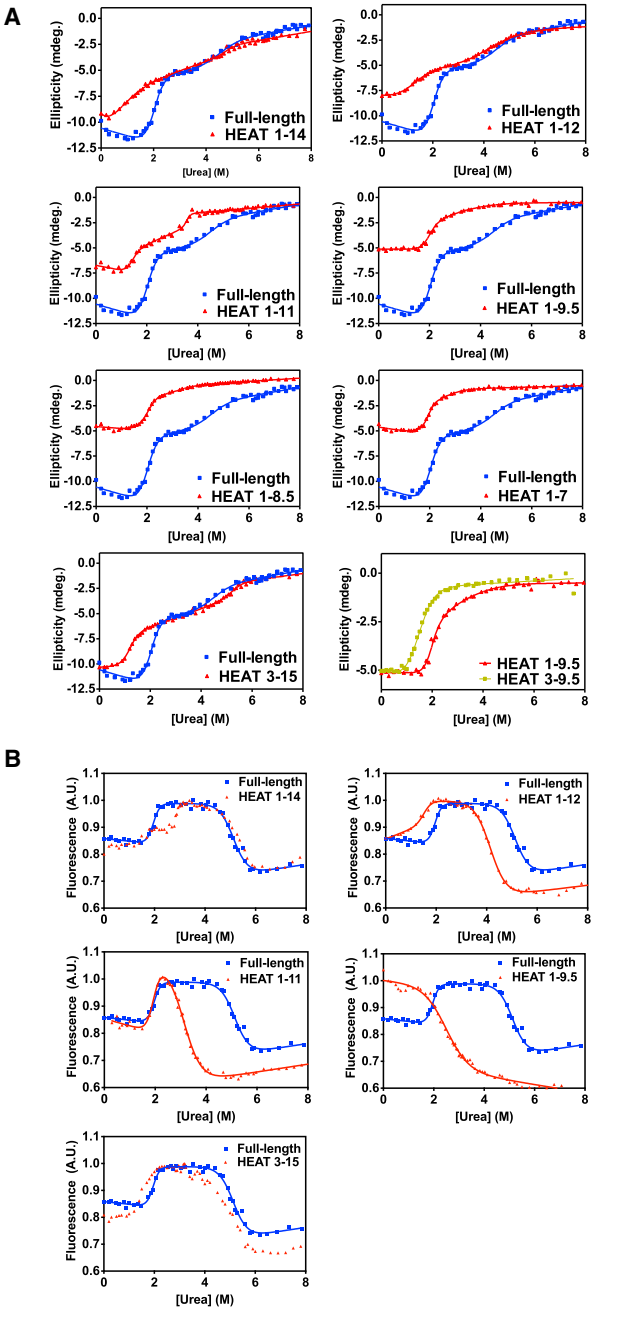

Figure 3. Truncated Variants of PR65/A

(A and B) Denaturation curves of truncated variants measured by CD (A) and fluorescence (B) are shown. The truncations are in red, and the full-length protein is in blue. The variant HEAT 3–9.5, which is both N-terminally and C-terminally truncated, is shown in yellow, with HEAT 1–9.5 in red for comparison. The protein concentration was 1–2 μM, and the data were normalized to 1 μM protein concentration.

(C) Native gels of wild-type and truncated variants incubated in 0–8 M urea are shown. The protein concentration was 2.5 μM.

See also Figures S2 and S3 and Table S1.

denaturation profiles (Figure 3B; Tables S1D and S1E). Deletion of repeat 13 and repeat 12 also results in a large shift to lower urea concentrations in the second unfolding transition monitored by fluorescence and less oligomerization of the intermediate as judged by native PAGE (Figure 3C); moreover, no oligomerization is observed for constructs comprising fewer than ten repeats. These results are consistent with the site-directed mutagenesis data, and they also show that repeats 11–13 are important for the stability of the oligomeric intermediate. A second transition in the CD-monitored denaturation profile is evident for HEAT 1–9.5, HEAT 1–8.5, and HEAT 1–7 and is absent for HEAT 3–9.5, and therefore this transition most likely corresponds to the unfolding of repeats 1–2. The fact that the first transition, corresponding to unfolding of repeats 3–9.5, is also greatly shifted (to a lower urea concentration) upon deletion of repeats 1–2, is consistent with this conclusion: deletion of repeats 1-2 destabilizes the adjacent subdomain 3–9.5, causing it to unfold at lower urea. The CD-monitored denaturation profile does not change greatly between HEAT 1–9.5, HEAT 1–8.5, and HEAT 1–7, suggesting that repeats 8–9.5 are only weakly structured in the absence of any C-terminal repeats.

The truncations are consistent with the model for the unfolding of full-length PR65/A inferred from the point mutants (Figure 2D), in which the hyperfluorescent intermediate has folded repeats 1–2 and 11–13. The truncations help us to further delineate the subdomains within the HEAT-repeat array and provide information about the relative stabilities of these subdomains. Deletion of the N-terminal two repeats (variant HEAT 3–15) shifts the first transition to a much lower urea concentration, indicating that HEAT 1–2 acts as a cap for the HEAT 3–10 moiety. The variant HEAT 1–13 was insoluble, suggesting that HEAT 14–15 acts as a cap for the HEAT 11–13 moiety; however, the mutational data, both at equilibrium and when following the unfolding kinetics (see below), indicate that HEAT 11–13 remains folded when HEAT 14–15 is unfolded, suggesting that HEAT 11–13 is an independently stable moiety and that the HEAT 14–15 cap acts to protect the HEAT 11–13 from oligomerization rather than stabilizing it against unfolding. Deletion of HEAT 15 destabilizes the C-terminal moiety, resulting in this moiety unfolding at much lower urea concentration in the variant HEAT 1–14 than in the full-length protein. Further deletions, of HEAT 13 and HEAT 12, destabilize the HEAT 11–13 moiety, causing this moiety to unfold at very low urea concentrations in variants HEAT 1–12 and HEAT 1–11.

Additional Intermediates Are Revealed in the Unfolding Kinetics

The unfolding kinetic traces are characterized by three phases at urea concentrations above 5 M: a fast phase associated with a

decrease in fluorescence, a slower phase associated with an increase in fluorescence, and a very slow phase associated with a decrease in fluorescence (Figure S4A). The fast and slower phases are of similar amplitudes, and the slowest phase is much smaller in amplitude. At urea concentrations below 5 M, the slowest phase is not present. However, the traces still fit to the sum of three exponential phases because a different slow phase appears, one that is associated with an increase in fluorescence (Figure S4A); this phase is also small in amplitude. The unfolding traces obtained by stopped-flow far-UV CD fit to a single exponential phase at higher urea concentrations and to the sum of two exponential phases at lower urea concentrations (Figure S4A); the rate constants for these phases are similar to those of the fast and slower phases measured by fluorescence (at higher urea concentrations, these two rate constants are very similar, and they can only be distinguished in the fluorescence experiments because they have opposite signs). The fast phase is larger in amplitude than is the slower phase, which is consistent with the assignment that we make below of these phases to the unfolding of specific subdomains of different sizes.

The refolding kinetic traces can be fitted to the sum of three exponential phases (Figure S4A). There is a fast phase associated with an increase in fluorescence and two slower phases associated with decreasing fluorescence. Interestingly, the rate constants increase with increasing urea concentrations between 1.9 and 2.3 M urea, in contrast to the regime below 1.9 M urea and above 2.3 M urea, where they show the usual decrease upon increase in urea concentration (Figure 4A). Nonlinearity can be caused by transient aggregation; however, in the case of PR65/A, the refolding rate constants are independent of protein concentration between 0.2 and 5 μM (Figure S4B), and also there is no aggregation in the initial urea-denatured state. The unfolding kinetics is also independent of protein concentration (Figure S4B). Another explanation is that a misfolded species accumulates that needs to unfold in order to progress to the native state. Because of the complexity of the refolding kinetics and the potential for additional difficulties in interpreting the data arising from the large number of proline residues (26), our subsequent analysis focuses instead on the unfolding kinetics. The plot of the logarithm of the rate constants for unfolding versus urea concentration is shown in Figure 4A.

Unfolding of HEAT 3–10 Occurs Rapidly

Upward curvature is evident in the urea dependence of the rate constants of the fast unfolding phase (Figure 4A), which suggests that there are alternative pathways accessible, as has been observed previously for the large ankyrin-repeat protein D34 (Werbeck et al., 2008). Later, we describe such a model, and we show that we can reproduce both the raw kinetic traces and the upward curvature in the limb using the model. The additional, downward curvature evident in this limb and in the limb of the slower phase can be interpreted as a denaturant-dependent shift between two (or more) sequential transition states separated by a high-energy intermediate (sequential barriers model) (Sánchez and Kiefhaber, 2003) or a gradual movement of the top of a broad energy barrier along the reaction coordinate in accordance with Hammond behavior (Hammond, 1955; Matouschek and Fersht, 1993). Mutations in HEAT repeat 3 (V103A) and in HEAT repeats 8–10 (V283A, V333A, V340A, and V390A) speed up this phase, suggesting that it corresponds to the unfolding of these repeats (Figure 4B, second row; Figure S4D). The mutations V103A and V340A do not show upward curvature, suggesting that only one pathway is now accessible. For V283A, V332A, and V390A, none of which are very destabilizing, both of the alternative pathways are still accessible and therefore there is upward curvature for these mutants as there is for the wild type. Mutations in HEAT 4–7 (S142V, Q169P, V200A, V219A, and V258A) have only small effects on the fast unfolding phase and no significant effect on the other unfolding phases (Figure S4D).

A Second Intermediate Is Formed Subsequently by the Unfolding of HEAT 14–15

The very destabilizing mutations in HEAT repeats 14–15 (e.g., V575A) have the most dramatic effect on the slower unfolding phase, which is associated with an increase in fluorescence, suggesting that this phase arises from unfolding of these elements of structure (Figure 4B, bottom row). Interestingly, however, these mutations also increase the rate of the fast phase described in the previous section. This result can be understood if the mutations accelerate the rate of unfolding of repeats 14–15 such that this process becomes competitive with unfolding of 3–10, and the unfolding kinetics of these two regions can no longer be considered as sequential steps. A simple scheme that can explain the data allows for the unfolding of 3–10, followed by the unfolding of 14–15, or vice versa. The rate for the fast phase is then the sum of the unfolding rates of 14–15 and 3–10, and it can therefore be increased by strongly destabilizing mutations of 14–15. The dominant slow phase is still unfolding of 14–15. For V535A, the slow phase is completely absent (Figure S4D); this is most likely because the mutation is so destabilizing that the initial state observed has repeats 14–15 already unfolded–because they are either already unfolded–because either they are already unfolded at low denaturant, as suggested by the equilibrium fluorescence, or they unfold in the deadtime of the instrument. A similar scenario is most likely responsible for the absence of the slow phase for V575A at higher urea concentrations. For V532A, the slow phase is speeded up to a lesser extent and therefore can still be observed. Later, we describe in detail how such a scheme quantitatively fits the kinetic traces for all the above mutants.

A truncated variant comprising HEAT 13–15 was made in order to simplify the unfolding kinetics; W477 provides a spectroscopic probe in this variant. The unfolding traces of HEAT 13–15 are monophasic and associated with an increase in fluorescence (Figure S4E); this result is consistent with our assignment of the slow phase for full-length PR65/A to the unfolding of HEAT 14–15. These kinetic data are also consistent with the equilibrium data showing that HEAT 14–15 is unfolded in the hyperfluorescent intermediate. Interestingly, V535A has an effect on the slow unfolding kinetics also: whereas for the wild-type there is a slow unfolding phase below 5 M urea associated with an increase in fluorescence (corresponding to unfolding of HEAT 1–2, see below), this phase is absent in the mutant V535A, and instead there is a slow phase associated with a decrease in fluorescence.

HEAT 1–2 Unfolds to Form a Third Intermediate

The biggest effects of mutations V25A and V44A (HEAT 1 and HEAT 2) were on the slowest unfolding phase, observed at

Please cite this article in press as: Tsytlonok et al., Complex Energy Landscape of a Giant Repeat Protein, Structure (2013), http://dx.doi.org/10.1016/j.str.2013.08.028

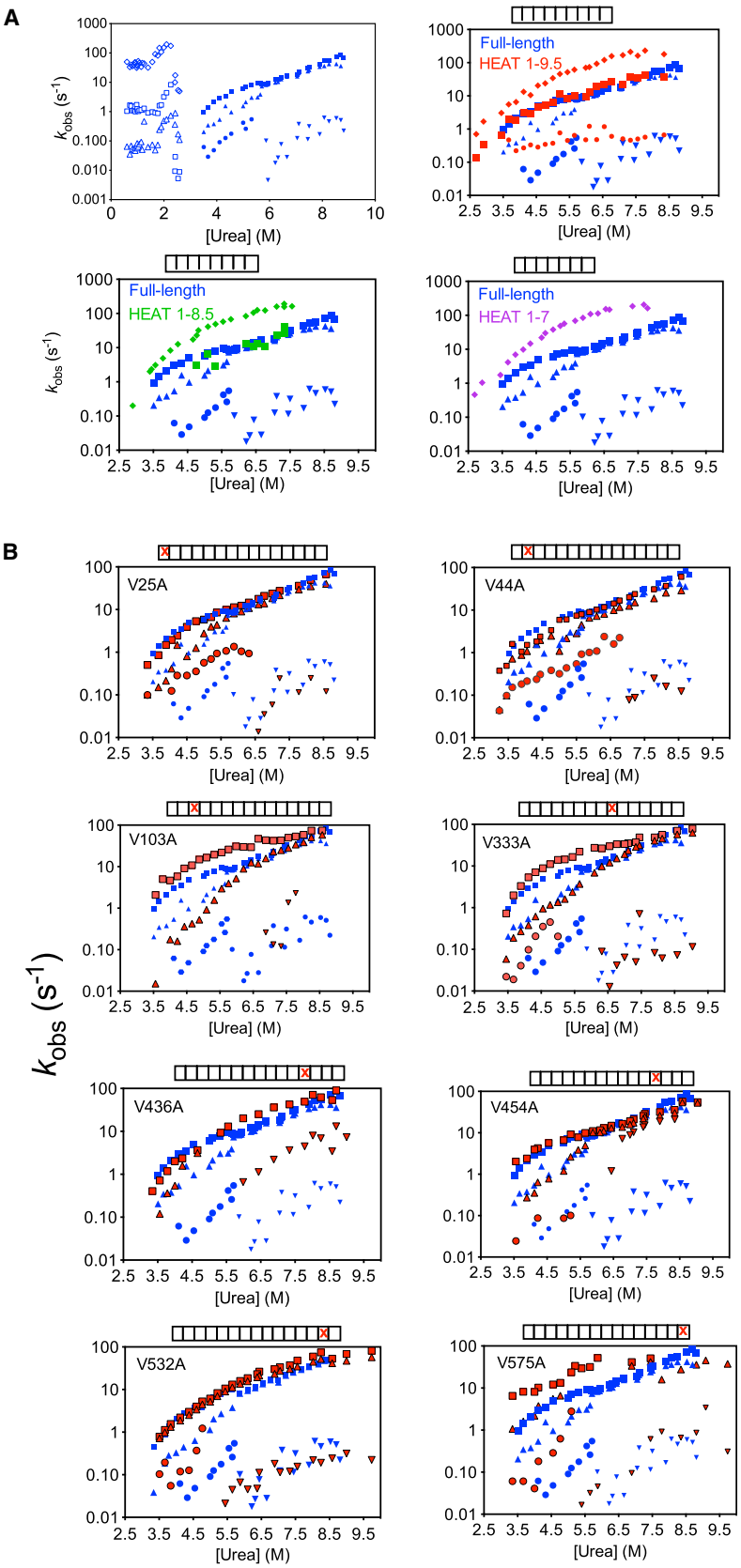

Figure 4. Refolding and Unfolding Kinetics of Full-Length PR65/A and Unfolding Kinetics of the C-Terminally Truncated Variants and of Representative Single-Site Substitutions

(A and B) The urea dependence of the rate constants is shown for the wild-type in blue, and the mutants in red. For wild-type and single-site substitutions, the fast unfolding phase is in squares, the slow unfolding phase is in triangles, the slowest unfolding phase observed at urea concentrations between 3.5 and 5.5 M is in circles, and that observed between 5.5 and 8.5 M urea is in inverted triangles. The location of each mutation is shown schematically. Measurements were made by stopped-flow fluorescence. The protein concentration was 1 μM. The very fast unfolding phase for the truncated variants (which is not observed for the full-length protein) is shown in diamonds.

See also Figure S4 for further mutants.

urea concentrations below 5 M, which is associated with an increase in fluorescence (Figure 4B, first row). The rate constants for this phase are speeded up by these mutations. Moreover, the phase is absent in the N-terminally truncated variant, HEAT 3–15 (Figure S4D). The results suggest that this phase corresponds to the unfolding of HEAT 1–2, which below 5 M urea occurs separately from HEAT 3–10, whereas above 5 M urea, HEAT 1–2 unfolds concomitantly with HEAT 3–10.

Unfolding of HEAT 11–13 Is the Slowest Step

V436A (HEAT 11) and V454A (HEAT 12) speed up the slowest unfolding phase observed for the wild-type above 5.5 M urea (Figure 4B, third row). The absence of this phase at urea concentrations below 5.5 M urea is consistent with it corresponding to unfolding of the HEAT 11–13 moiety: the equilibrium experiments show that this moiety is folded at lower urea concentrations. V436A also has an effect on the other phases; the slow phase, which corresponds to the unfolding of HEAT 1–2 and which is associated with an increase in fluorescence, is absent in its kinetics.

Truncations Simplify the Unfolding Kinetics

The unfolding traces of C-terminally truncated variants HEAT 1–9.5, HEAT 1–8.5, and HEAT 1–7 (all of which have W140 and W257) show a decrease in fluorescence (Figure S4F). The absence of the phase observed for full-length PR65/A, associated with an increase in fluorescence, suggests that the hyperfluorescence of the intermediate arises from the unfolding of repeats in the C-terminal moiety, and this conclusion is consistent with the absence of a hyperfluorescent intermediate for the truncations at equilibrium. The unfolding of HEAT 1–9.5 can be fitted to the sum of three exponential phases (Figures 4A and S4F). There is a very fast phase that is faster than any of the phases observed for the full-length protein, a fast phase that has the same rate constants and urea dependence as the fast phase of the full-length protein, and also a slow phase. The slow phase is very low in amplitude (~5%). The very fast phase is the major phase at high urea concentrations, whereas the fast phase is the major phase at lower urea concentrations. The unfolding of HEAT 1–8.5 can be fitted to the sum of two exponentials and shows the same very fast and fast phases as HEAT 1–9.5 (Figure S4F); the fast phase has very low amplitude, and the slow phase observed for HEAT 1–9.5 is absent (Figure 4A). The unfolding kinetics of HEAT 1–7 shows a single, very fast phase, which is similar to what is observed for HEAT 1–8.5 and HEAT 1–9.5 (Figures 4A and S4F). The presence of a fast phase for variants HEAT 1–8.5 and HEAT 1–9.5 is similar to that observed for full-length PR65/A, and the absence of fast phase in the kinetics of HEAT 1–7 is consistent with the assignment of this phase to the unfolding of HEAT 8–10. We therefore propose the following to explain the appearance of the very fast unfolding phase in these three truncated variants and the presence at low amplitudes in HEAT 1–8.5 and HEAT 1–9.5 of the fast unfolding phase characteristic of full-length PR65/A: in the absence of HEAT 8–10, the HEAT 1–7 variant is able to unfold in a single very rapid step. For HEAT 1–8.5 and HEAT 1–9.5, there is a mixture of "native" states, arising because the HEAT 8–10 moiety is truncated; the truncated moiety is unstructured in the major population, and so unfolding proceeds as it does for HEAT 1–7, whereas the moiety is structured in the minor population, and so unfolding proceeds as it does for the full-length protein.

Unfolding of HEAT 4–7 in a Very Fast Unfolding Step Is Revealed by Double Jump Experiments

Two observations suggest that HEAT 4–7 unfolds after the unfolding of HEAT 8–10 in a very fast step, which, because it is faster than the previous step, is not detected in the kinetic measurements. First, mutations in HEAT repeats 4–7 do not have a clear effect on the any of the four kinetic unfolding phases observed for full-length PR65/A, whereas mutations elsewhere in the protein have clear effects that allow each phase to be assigned to the unfolding of a specific subdomain. Second, when repeats adjacent to HEAT 4–7 are deleted (i.e., the truncated variants HEAT 1–7, HEAT 1–8.5, and HEAT 1–9), a very fast phase is observed that is absent in the unfolding of the full-length protein (Figure 4A). To test whether there is an unfolding additional phase that is missed in the single-jump method, we performed an interrupted refolding experiment in which unfolded protein was allowed to refold for a delay time of between 30 ms and 50 s before unfolding was initiated. The unfolding kinetics of this refolded protein was multiphasic, and all but one of the phases had similar rate constants to those observed by single-jump unfolding of native protein (Figure S4C). The new phase is very fast, and its amplitude first increases and then decreases with increasing refolding delay time, consistent with the accumulation and decay of an intermediate. We therefore propose that this very fast phase is the "missing" phase that corresponds to the unfolding of HEAT 4–7.

Kinetic Model for Unfolding of PR65/A via Four Intermediate States

The above mutation and truncation studies establish qualitatively that the unfolding of PR65/A can be understood in terms of a predominant order of unfolding events. Nevertheless, the fact that some mutants affect more than one of the relaxation rates establishes that a simple sequential unfolding mechanism is inadequate. We have therefore used our data to construct a six-state phenomenological kinetic model that explains both the wild-type data and the data for single point mutants over a wide-range of denaturant concentrations. The scheme is summarized in Figure 5A and includes the native and unfolded proteins as well as four intermediates, labeled I–IV. The initial unfolding of the wild-type protein occurs mainly via unfolding of repeats 3-10 (rate k_1), followed by 14-15 (rate k_2). An alternative possibility is that these two regions could unfold in reverse order, with these parallel pathways explaining several features of the kinetics. The last two steps are unfolding of repeats 1-2 (rate k_3), followed by 11-13 (rates k_4, k_{-4}). Unfolding of 11-13 has been treated as being reversible in the model, because it does not occur at equilibrium below ~5 M urea for the wild-type. The rate k_1 was modeled via a sequential barriers scheme in which unfolding from native to intermediate I occurs via a high-energy intermediate X, i.e., $k_1 = k_{nx}k_{xi}/(k_{xu} + k_{xn})$ (Sánchez and Kiefhaber, 2003): k_1 describes the folding of the largest contiguous "block" of repeats (3-10), and a similar model has been used to describe downward curvature in other long repeat arrays (Lowe and Itzhaki, 2007; Werbeck et al., 2008). The scheme has five kinetic eigenvalues corresponding to the rates k_1, k_2, and k_3

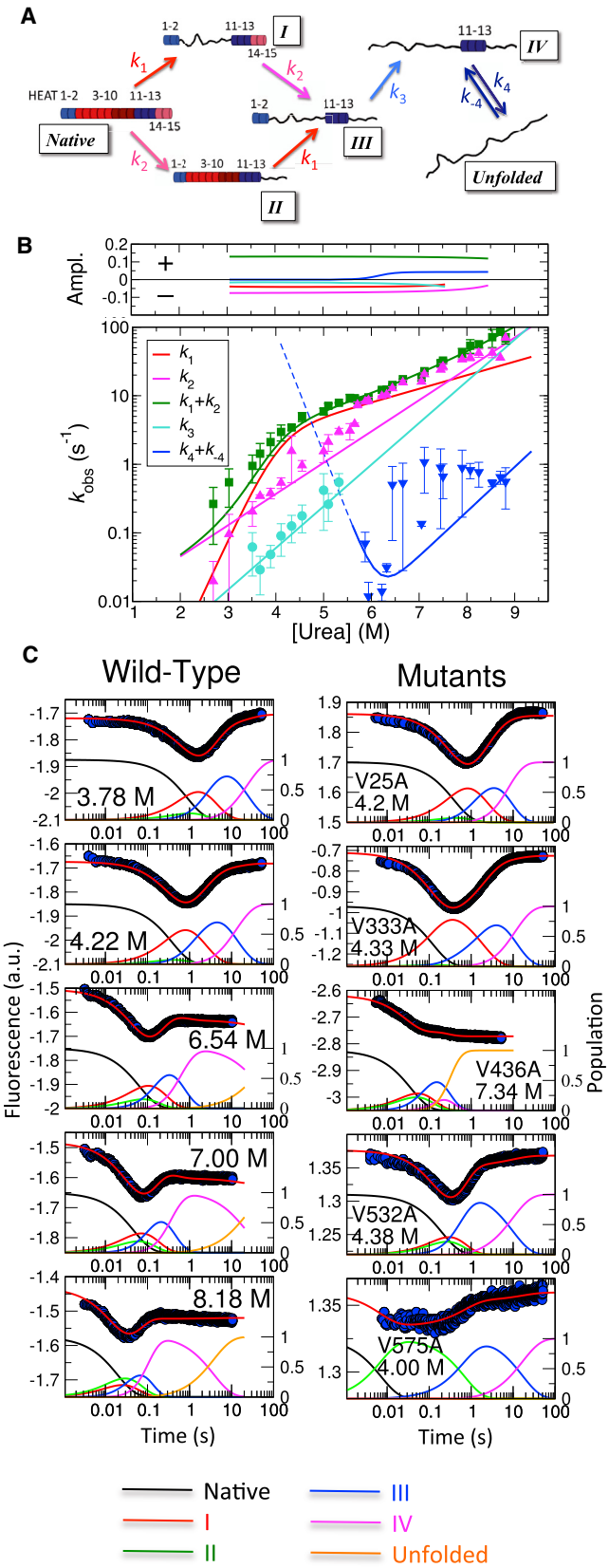

Figure 5. Kinetic Model for PR65/A Unfolding

(A) Kinetic scheme in which the native protein unfolds via a sequence of (parallel) intermediates to the unfolded state is shown. Names of species are given in boxes, and rate coefficients for their interconversion are shown next to the arrows. Color code refers to the region that is unfolding in each step.

(B) Lower: relaxation rates from kinetic model (lines) overlaid on the rates obtained by fitting of exponentials to raw data; color code corresponds to that in (A), as far as possible. Upper: amplitudes corresponding to each relaxation. Positive amplitude corresponds to a decrease of fluorescence with time.

(C) Left axes: raw fluorescence traces (blue symbols) overlaid with the global fit from the kinetic model (red lines). Right axes: time evolution of the populations of the species shown in (A). A range of concentrations are shown for wild-type in the left column, and in the right column, representative traces for five mutants chosen from the different structural regions are shown.

See also Figure S5 for further examples of the kinetic modeling results. See also Table S2.

and the sums k_1+k_2 and k_4+k_{-4} (see the Supplemental Experimental Procedures for details).

The fluorescence signal was modeled as a sum over populations of each species, weighted by their relative fluorescence quantum yield (QY), i.e., the ratio of the fluorescence intensity of each species to that of the fully folded state, at equivalent concentrations. The model was globally fit to the raw stopped-flow fluorescence traces at all urea concentrations for the wild-type and several mutants: V25A (representing repeats 1-2), V333A (3-10), V436A (11-13), and V532A and V575A (14-15). The wild-type fit parameters are rates and *m*-values at 4 M urea and the QY for each species. The mutants shared all wild-type parameters, with only rate coefficient(s) pertaining to the relevant region being fit. For V436A, fitting the QY separately was necessary; this can be rationalized by a direct V436-W477 contact in the structure, causing a change in environment on mutation. The final fit parameters are listed in Table S2.

The five relaxation times are plotted versus urea concentration in Figure 5B (lines). Below 5M urea, the protein unfolds via a series of hypofluorescent intermediates (I, II, and III) to the hyperfluorescent intermediate IV, which is stable up to ~5 M urea, resulting in a "dip" in the fluorescence (Figure 5C). The three exponential phases evident in the raw data (symbols in Figure 5B) are associated with the following: (1) the fastest (green) phase is formation of intermediates I and II with rate k_1 + k_2, with large positive amplitude (i.e., a decrease in fluorescence). Decay of II with very similar rate $k_1 \approx k_1 + k_2$ with small negative amplitude is not resolved because of the small population of intermediate II. (2) The second (magenta) phase is decay of the populated intermediate I with rate k_2, with significant negative amplitude. (3) The third (cyan) phase is conversion of intermediate III to IV with very small negative amplitude.

Above 5 M urea, the fastest phase is still associated with the formation of intermediates I and II with rate k_1+k_2, but upward curvature results from a switch of pathway to that going via intermediate II ($k_2 > k_1$). The second fastest phase now includes decay of both intermediates I and II, because intermediate II is significantly populated. The relaxation with rate k_3 is less evident in the fluorescence, because above 5 M urea, intermediate IV decays further to the unfolded state with rate ~k_4. The amplitude of this final relaxation emerges with significant positive amplitude at urea concentrations above ~6 M (Figure 5B, upper panel).

The fluorescence traces for the mutants are also described by this model, by changing only the rates for the relevant

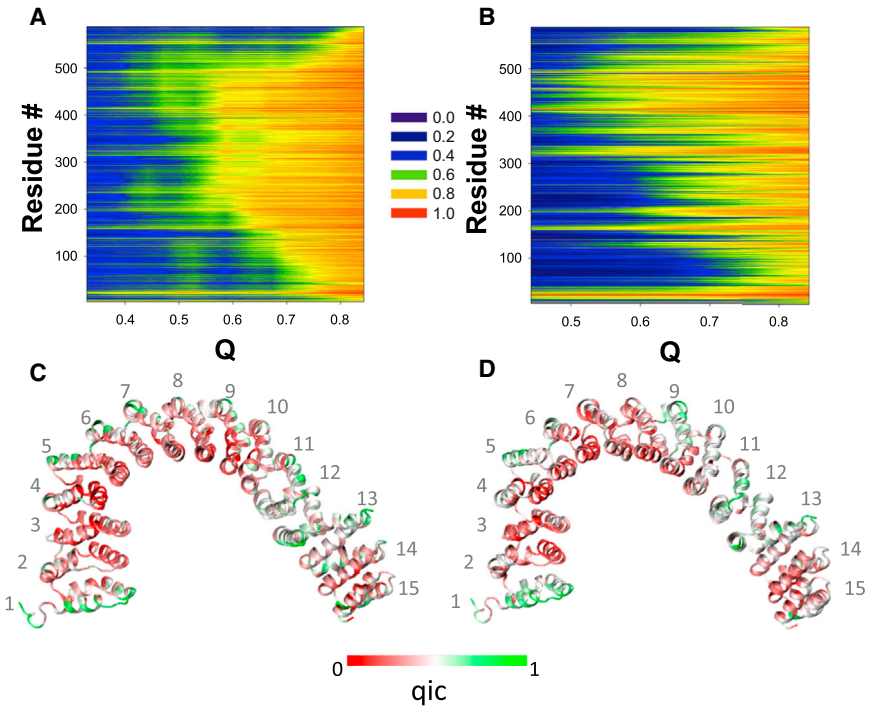

Figure 6. Average Folding Mechanism Derived from Coarse-Grained Simulations

The evolution of the fraction of native contacts for each residue (qic) as a function of the global reaction coordinate Q is shown in (A) and (B) for the Gō and smAMW models, respectively. The structure of the native state of PR65/A (PDB ID code 1B3U chain A), colored by the value of qic of each residue at Q = 0.47, is shown in (C) and (D) for the Gō and smAMW models, respectively. The HEAT-repeat index numbers are indicated.

subdomain. The fits at selected urea concentrations are shown in Figures 5C and S5, and kinetic parameters are given in Table S2. A detailed explanation of the effects of the altered rates on the populations of the intermediates is available in the Supplemental Experimental Procedures. Considering the four separate phases resolved in the fit of the traces to a sum of exponentials (Figure 4A), any kinetic model must include at least five species, implying 20 independent rate coefficients. Our model, with five independent rate coefficients, is thus very conservative and explains the data remarkably well.

Structure-Based Model Simulations of PR65/A Can Rationalize the Variation in Repeat Stability

Structure-based model simulations were next used to analyze the folding energy landscape of PR65/A. This type of model relies on the dominantly funneled nature of protein energy landscapes, which are achieved by evolution through the realization of consistent interresidue interactions that largely favor the native topology compared to other alternative configurations (Bryngelson and Wolynes, 1987). Pure structure-based models in the energy function take into sole account the interactions found in the native structure. This energy function is then used in molecular dynamics simulations to characterize the trade-off between the loss of configurational entropy and the gain of favorable contact interactions along folding trajectories, to characterize the folding mechanism and to identify the dominant pathways in a context in which they can solely be influenced by topology. Many different flavors of structure-based models that incorporate heterogeneity in the strength of interresidue interactions or allow for nonnative interactions have been successfully used in the past to characterize the existence of folding intermediates and the interpretation of folding pathways (Wolynes, 2005).

We used two types of coarse-grained models. One model involves a perfectly funneled energy function (Gō-like model) with a homogeneous contact potential. The other model treats the short- and medium-range interactions with a heterogeneous native structure-based potential but allows for frustration and nonnative contacts between residues distant in sequence ($i - j \geq 12$). These tertiary interactions involve a generalized statistical contact potential that depends on the spatial distance and sequence identity of the interacting residue pairs. It includes water-mediated interactions and a burial energy term. We call this model smAMW for *s*ingle *m*emory *a*ssociative *m*emory Hamiltonian with *w*ater-mediated interactions.

We first ran a series of short simulations at different temperatures to determine the folding temperature (Tf) and then performed a more exhaustive sampling at a temperature near the Tf using umbrella sampling. The Tf observed was 0.9 and 1.0 in reduced temperature units for the pure structure-based and smAMW models, respectively. Figures 6A and 6B show the evolution of the equilibrium average fraction of native contacts (qic) for all residues of PR65/A as a function of the global reaction coordinate, Q, which represents the total fraction of native contacts for the whole protein. ("qic" indicates the reaction coordinate Q applied locally for each residue "i" that reports about the fraction of native contacts "c".) This type of plot gives an indication of which regions become folded at different depths in the funnel, i.e., values of the reaction coordinate. For both models, at low values of Q, the elements that have the highest number of native contacts formed correspond to HEAT 1 (residues 1–42), 11–12 (397–473), and half of HEAT 13 (residues 474–495) (Figures 6C and 6D). These results suggest that these elements fold first in agreement with the experimental observations. The other elements fold at a later stage. The analysis also shows that at low values of Q there is also some partial structure in restricted regions of HEAT 5, HEAT 6, and HEAT 9, although lower than the amount found in the previous repeats. Although the heterogeneity of the contact energy and the influence of nonnative contacts might modulate the order of folding of the different elements, the similarity in the patterns observed for the two models indicates that the native topology plays a dominant role in determining the folding mechanism of this protein.

Please cite this article in press as: Tsytlonok et al., Complex Energy Landscape of a Giant Repeat Protein, Structure (2013), http://dx.doi.org/10.1016/j.str.2013.08.028

DISCUSSION

PR65/A is a giant repeat-protein scaffold that exploits an extended binding interface to regulate the combinatorial assembly of the catalytic C subunit and one of many different regulatory B subunits of the serine/threonine protein phosphatase PP2A (Figure 1). The regular packing of the HEAT repeats of PR65/A is interrupted in two regions, by rotations between HEAT 3 and HEAT 4 and between HEAT 12 and HEAT 13 (Cho and Xu, 2007; Groves et al., 1999; Perry and Kleckner, 2003). Moreover, binding of the C subunit to the C-terminal HEAT repeats and the B subunit to the N-terminal HEAT repeats further modulates the packing, thereby bringing the B-bound substrate into close proximity with the active site of the C subunit (Cho and Xu, 2007; Xu et al., 2006, 2008). This structural variability in PR65/A suggests a dynamic mode of function (Shi, 2009), which is reflected in its energy landscape elucidated here.

Hierarchical Unfolding of PR65/A

The equilibrium unfolding of PR65/A shows two transitions when monitored by fluorescence or CD. The *m*-value for the first transition obtained by fluorescence (4 kcalmol^{-1}M^{-1}) is much smaller than that predicted for the longest contiguous segment that the mutant data show to unfold in this regime (HEAT 3–10, ~320 residues, shown schematically in Figure 2D). Moreover, this *m*-value is larger than that obtained by CD (3 kcalmol^{-1}M^{-1}), and the *m*-values for mutants in these repeats are smaller than that of the wild-type. These results indicate that the unfolding of HEAT 3–10 does not occur in a single step and that the two probes detect these multiple steps somewhat differently. The mutant and fragment data are consistent with a multistep process of PR65/A unfolding involving five subdomains: HEAT 1–2, HEAT 3–7, HEAT 8–10, HEAT 11–13, and HEAT 14–15. We emphasize that we cannot define the boundaries of these subdomains in a very precise manner: for example, a mutation in HEAT 3 has a similar effect on the kinetics to mutations in HEAT 8–10, but we have grouped HEAT 3 together with the adjacent repeats into the HEAT 3–7 subdomain. We are able to delineate the relative stabilities of the subdomains: HEAT 3–7 and HEAT 8–10 are the least stable subdomains; they have similar stabilities to one another, and the interface between them is weak. Consequently, unfolding of these two subdomains is decoupled, and the single transition that is observed at equilibrium has a lower *m*-value than would be expected for a cooperative process; interestingly, the unfolding kinetics of the truncated variants and the double-jump experiments suggest that HEAT 3–7 and HEAT 8–10 unfold in separate steps. The HEAT 14–15 subdomain has low stability, like HEAT 3–7 and HEAT 8–10, and the unfolding of these subdomains may occur in parallel. Subdomains HEAT 1–2 and HEAT 11–13 have the highest stabilities.

We observe that folding of HEAT 1 and HEAT 11–13 is somewhat favored topologically, as shown by the pure structure-based model simulation. This feature is accentuated in the smAMW simulation, indicating that the heterogeneous contact potential and burial energy terms in this model reinforce the different stability of these repeats compared to the others. The simulation results therefore support the experimental evidence of the late unfolding (early folding) of HEAT 1–2 and HEAT 11–13. However, a prediction of the folding order of the other repeats at later stages of the folding funnel was more difficult, most likely because of small differences in their stabilization energy.

Origins of Unfolding Cooperativity in Repeat Proteins

Based on the order of unfolding events inferred from the mutant studies, we have constructed a six-state kinetic model. Using a common set of rate coefficients and relative quantum yields for the different states, we are able to fit the kinetic data quantitatively over a wide-range of denaturant concentrations for both wild-type and mutant proteins. The only parameter that is varied between the wild-type and each mutant is the unfolding rate of the region in which the mutation is located. Although there is a predominant order of unfolding, the model includes parallel unfolding pathways, which explain both the upward curvature in the fastest relaxation rate at high denaturant and the fact that some mutations are able to alter more than one relaxation rate. HEAT 3–10 unfolds most rapidly, followed by HEAT 14–15, and then HEAT 1–2 and HEAT 11–13 unfold in slower steps. Because the distribution of stability between different subdomains of the array is very uneven, this order of unfolding events is strongly preferred, such that only very destabilizing mutations can shift the order; for example, making unfolding of HEAT 14–15 competitive with HEAT 3–10. Similarly, our group and others have shown that alternative kinetic routes are accessible for a number of ankyrin-repeat proteins, such as myotrophin (Lowe and Itzhaki, 2007), D34 (Werbeck et al., 2008), and Notch (Tripp and Barrick, 2008), as was originally predicted based on the symmetric topology of repeat proteins (Ferreiro et al., 2005). When intermediates are destabilized, unfolding of the HEAT repeats can occur in a single step, as occurs, for example, in the truncated variant HEAT 1–7.

The origins and limits of cooperativity in ankyrin-repeat, TPR, and LRR proteins have been explored using both experimental and computational approaches (Ferreiro et al., 2005, 2007, 2008; Hagai and Levy, 2008; Main et al., 2005; Mello and Barrick, 2004; Werbeck and Itzhaki, 2007). These studies indicate that repeats that are distant from one another and therefore not directly in contact can nevertheless unfold in a coupled manner if they and the intervening units have similar intrinsic, low stabilities and the interfaces between them are strong. Cooperativity is thereby maintained over the medium range. If the distribution of stabilities is uneven, an intermediate state will be populated. Moreover, cooperativity in repeat proteins can be easily modulated in a rational way, by changing the stabilities of the individual units; thus, we were able to make variants of the 12-ankyrin repeat protein D34 in which the unfolding of 10 of the 12 repeats was coupled (Werbeck and Itzhaki, 2007). In PR65/A, the stability is unevenly distributed across this large array with the result that multiple intermediates are populated.

Impact of Unfolding on the Function of Giant Repeat Proteins

In small repeat proteins, the unfolding of the low-stability regions may be important for ligand binding, posttranslational modifications, and translocation (Bergqvist et al., 2009; Coleman et al., 2007; Junker et al., 2006; Kloss and Barrick, 2009; Löw et al., 2009; Truhlar et al., 2008; Zweifel et al., 2003). Our results

indicate that for giant repeat proteins, an uneven stability distribution is important for a distinct type of function. Giant repeat arrays bind multiple partners by making use of distantly located sites. The ability to decouple the folding of these sites suggests that they can employ a so-called "fly-casting" mechanism for molecular recognition (Shoemaker et al., 2000; Trizac et al., 2010). PR65/A uses its C-terminal HEAT repeats to bind the catalytic C-subunit of PP2A and the N-terminal HEAT repeats to bind diverse regulatory B-subunits that direct the enzyme to specific substrates (Xu et al., 2008); flexibility within the PR65/A scaffold is apparent even in the crystallized protein, with both B- and C-subunit binding causing significant reconfiguration of the inter- and intrarepeat packing (Shi, 2009). Importantly, we show here that the low-stability, central repeats of PR65/A are partially unfolded under physiological conditions, as seen in the denaturation profiles of wild-type and a representative mutant at 37°C (Figure S1E). Acting as a double-ended fly cast, unfolding fluctuations of the central moiety could function to broaden the search area for binding partners and thereby facilitate molecular recognition at the two ends of the repeat array. The unfolding of the central moiety of PR65/A will also directly impact on the catalytic process, as it will alter the proximity of the catalytic C-subunit and the B-subunit-bound substrate (Figure 1A). Thus, within PP2A, PR65/A is a flexible molecular adaptor, partial unfolding of which modulates both assembly and catalytic mechanism of the enzyme. Our findings show how the absence of sequence-distant contacts affords repeat proteins particular flexibility, consistent with the proposal (Forwood et al., 2010) that they are a distinct structural class midway between globular structured proteins and intrinsically disordered proteins (Wright and Dyson, 2009).

EXPERIMENTAL PROCEDURES

The Experimental Procedures are described in the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.08.028.

ACKNOWLEDGMENTS

We thank the Center of Theoretical Biological Physics for computational resources. We thank Dr. Stephen McLaughlin (Medical Research Council Laboratory of Molecular Biology) for help with analysis of PR65/A oligomers. P.G.W. and P.O.C. acknowledge grants (GM071862 and R01 GM44557) from the National Institutes of Health. L.S.I., M.T., and E.S. are supported by the Medical Research Council of the UK (G1002329) and the Medical Research Foundation. R.B.B. is supported by a Royal Society University Research Fellowship while at Cambridge and by the Intramural Research Programme of the National Institute of Diabetes and Digestive and Kidney Diseases while at the National Institutes of Health. S.P. is supported by the 973 Program (2012CB911000 and 2013CB910700). This work was facilitated by a Royal Society International Joint-Project Grant (to L.S.I. and S.P.).

Received: June 25, 2013
Revised: August 21, 2013
Accepted: August 22, 2013
Published: October 10, 2013

REFERENCES

Barrick, D., Ferreiro, D.U., and Komives, E.A. (2008). Folding landscapes of ankyrin repeat proteins: experiments meet theory. Curr. Opin. Struct. Biol. *18*, 27–34.

Bergqvist, S., Alverdi, V., Mengel, B., Hoffmann, A., Ghosh, G., and Komives, E.A. (2009). Kinetic enhancement of NF-kappaBxDNA dissociation by IkappaBalpha. Proc. Natl. Acad. Sci. USA *106*, 19328–19333.

Bryngelson, J.D., and Wolynes, P.G. (1987). Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. USA *84*, 7524–7528.

Cho, U.S., and Xu, W. (2007). Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. Nature *445*, 53–57.

Coleman, M.L., McDonough, M.A., Hewitson, K.S., Coles, C., Mecinovic, J., Edelmann, M., Cook, K.M., Cockman, M.E., Lancaster, D.E., Kessler, B.M., et al. (2007). Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. J. Biol. Chem. *282*, 24027–24038.

Ferreiro, D.U., and Wolynes, P.G. (2008). The capillarity picture and the kinetics of one-dimensional protein folding. Proc. Natl. Acad. Sci. USA *105*, 9853–9854.

Ferreiro, D.U., Cho, S.S., Komives, E.A., and Wolynes, P.G. (2005). The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. J. Mol. Biol. *354*, 679–692.

Ferreiro, D.U., Cervantes, C.F., Truhlar, S.M., Cho, S.S., Wolynes, P.G., and Komives, E.A. (2007). Stabilizing IkappaBalpha by "consensus" design. J. Mol. Biol. *365*, 1201–1216.

Ferreiro, D.U., Walczak, A.M., Komives, E.A., and Wolynes, P.G. (2008). The energy landscapes of repeat-containing proteins: topology, cooperativity, and the folding funnels of one-dimensional architectures. PLoS Comput. Biol. *4*, e1000070.

Forwood, J.K., Lange, A., Zachariae, U., Marfori, M., Preast, C., Grubmüller, H., Stewart, M., Corbett, A.H., and Kobe, B. (2010). Quantitative structural analysis of importin-β flexibility: paradigm for solenoid protein structures. Structure *18*, 1171–1183.

Grinthal, A., Adamovic, I., Weiner, B., Karplus, M., and Kleckner, N. (2010). PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. Proc. Natl. Acad. Sci. USA *107*, 2467–2472.

Groves, M.R., Hanlon, N., Turowski, P., Hemmings, B.A., and Barford, D. (1999). The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. Cell *96*, 99–110.

Hagai, T., and Levy, Y. (2008). Folding of elongated proteins: conventional or anomalous? J. Am. Chem. Soc. *130*, 14253–14262.

Hammond, G.S.H. (1955). A correlation of reaction rates. J. Am. Chem. Soc. *77*, 334–338.

Javadi, Y., and Itzhaki, L.S. (2013). Tandem-repeat proteins: regularity plus modularity equals design-ability. Curr. Opin. Struct. Biol. *23*, 622–631.

Junker, M., Schuster, C.C., McDonnell, A.V., Sorg, K.A., Finn, M.C., Berger, B., and Clark, P.L. (2006). Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. Proc. Natl. Acad. Sci. USA *103*, 4918–4923.

Kloss, E., and Barrick, D. (2009). C-terminal deletion of leucine-rich repeats from YopM reveals a heterogeneous distribution of stability in a cooperatively folded protein. Protein Sci. *18*, 1948–1960.

Kobe, B., and Kajava, A.V. (2000). When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem. Sci. *25*, 509–515.

Levy, Y., Onuchic, J.N., and Wolynes, P.G. (2007). Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. J. Am. Chem. Soc. *129*, 738–739.

Löw, C., Homeyer, N., Weininger, U., Sticht, H., and Balbach, J. (2009). Conformational switch upon phosphorylation: human CDK inhibitor p19INK4d between the native and partially folded state. ACS Chem. Biol. *4*, 53–63.

Lowe, A.R., and Itzhaki, L.S. (2007). Rational redesign of the folding pathway of a modular protein. Proc. Natl. Acad. Sci. USA *104*, 2679–2684.

Main, E.R., Stott, K., Jackson, S.E., and Regan, L. (2005). Local and long-range stability in tandemly arrayed tetratricopeptide repeats. Proc. Natl. Acad. Sci. USA *102*, 5721–5726.

Matouschek, A., and Fersht, A.R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. Proc. Natl. Acad. Sci. USA *90*, 7814–7818.

Mello, C.C., and Barrick, D. (2004). An experimentally determined protein folding energy landscape. Proc. Natl. Acad. Sci. USA *101*, 14102–14107.

Mumby, M. (1995). Regulation by tumour antigens defines a role for PP2A in signal transduction. Semin. Cancer Biol. *6*, 229–237.

Perry, J., and Kleckner, N. (2003). The ATRs, ATMs, and TORs are giant HEAT repeat proteins. Cell *112*, 151–155.

Sánchez, I.E., and Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. J. Mol. Biol. *325*, 367–376.

Shi, Y.G. (2009). Serine/threonine phosphatases: mechanism through structure. Cell *139*, 468–484.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc. Natl. Acad. Sci. USA *97*, 8868–8873.

Tripp, K.W., and Barrick, D. (2004). The tolerance of a modular protein to duplication and deletion of internal repeats. J. Mol. Biol. *344*, 169–178.

Tripp, K.W., and Barrick, D. (2008). Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape. J. Am. Chem. Soc. *130*, 5681–5688.

Trizac, E., Levy, Y., and Wolynes, P.G. (2010). Capillarity theory for the fly-casting mechanism. Proc. Natl. Acad. Sci. USA *107*, 2746–2750.

Truhlar, S.M., Mathes, E., Cervantes, C.F., Ghosh, G., and Komives, E.A. (2008). Pre-folding IkappaBalpha alters control of NF-kappaB signaling. J. Mol. Biol. *380*, 67–82.

Werbeck, N.D., and Itzhaki, L.S. (2007). Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. Proc. Natl. Acad. Sci. USA *104*, 7863–7868.

Werbeck, N.D., Rowling, P.J., Chellamuthu, V.R., and Itzhaki, L.S. (2008). Shifting transition states in the unfolding of a large ankyrin repeat protein. Proc. Natl. Acad. Sci. USA *105*, 9982–9987.

Wolynes, P.G. (2005). Recent successes of the energy landscape theory of protein folding and function. Q. Rev. Biophys. *38*, 405–410.

Wright, P.E., and Dyson, H.J. (2009). Linking folding and binding. Curr. Opin. Struct. Biol. *19*, 31–38.

Xu, Y., Xing, Y., Chen, Y., Chao, Y., Lin, Z., Fan, E., Yu, J.W., Strack, S., Jeffrey, P.D., and Shi, Y. (2006). Structure of the protein phosphatase 2A holoenzyme. Cell *127*, 1239–1251.

Xu, Y., Chen, Y., Zhang, P., Jeffrey, P.D., and Shi, Y. (2008). Structure of a protein phosphatase 2A holoenzyme: insights into B55-mediated Tau dephosphorylation. Mol. Cell *31*, 873–885.

Zweifel, M.E., Leahy, D.J., Hughson, F.M., and Barrick, D. (2003). Structure and stability of the ankyrin domain of the Drosophila Notch receptor. Protein Sci. *12*, 2622–2632.

