

# Slow formation of aggregation-resistant $\beta$ -sheet folding intermediates

Mirco Junker and Patricia L. Clark\*

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556-5670

## ABSTRACT

Protein folding has been studied extensively for decades, yet our ability to predict how proteins reach their native state from a mechanistic perspective is still rudimentary at best, limiting our understanding of folding-related processes *in vivo* and our ability to manipulate proteins *in vitro*. Here, we investigate the *in vitro* refolding mechanism of a large  $\beta$ -helix protein, pertactin, which has an extended, elongated shape. At 55 kDa, this single domain, all- $\beta$ -sheet protein allows detailed analysis of the formation of  $\beta$ -sheet structure in larger proteins. Using a combination of fluorescence and far-UV circular dichroism spectroscopy, we show that the pertactin  $\beta$ -helix refolds remarkably slowly, with multiexponential kinetics. Surprisingly, despite the slow refolding rates, large size, and  $\beta$ -sheet-rich topology, pertactin refolding is reversible and not complicated by off-pathway aggregation. The slow pertactin refolding rate is not limited by proline isomerization, and 30% of secondary structure formation occurs within the rate-limiting step. Furthermore, site-specific labeling experiments indicate that the  $\beta$ -helix refolds in a multistep but concerted process involving the entire protein, rather than via initial formation of the stable core substructure observed in equilibrium titrations. Hence pertactin provides a valuable system for studying the refolding properties of larger,  $\beta$ -sheet-rich proteins, and raises intriguing questions regarding the prevention of aggregation during the prolonged population of partially folded,  $\beta$ -sheet-rich refolding intermediates.

Proteins 2010; 78:812–824.  
© 2009 Wiley-Liss, Inc.

**Key words:** parallel  $\beta$ -sheet; slow kinetics; protein folding; aggregation;  $\beta$ -helix; repeat protein; contact order; topology;  $\beta$ -strand; tryptophan fluorescence.

## INTRODUCTION

Our ability to predict how a protein folds to its native structure, and how this occurs efficiently, avoiding misfolding and aggregation, is still hampered by our incomplete understanding of protein folding mechanisms.<sup>1,2</sup> Many experiments have measured the refolding properties of small, roughly globular proteins.<sup>3,4</sup> Small proteins typically have fast refolding kinetics, fold via two-state kinetic mechanisms, and have minimal propensity for aggregation. These properties facilitate refolding studies, and detailed refolding mechanisms have been determined for several small proteins (see Refs. 5–10 for examples). Likewise, several approaches have been used to predict refolding rates of these small proteins,<sup>11–14</sup> including correlations with either the size of the protein and/or its contact order, that is, the number of amino acids separating residues that interact in the natively folded protein.<sup>15,16</sup> However, while contact order shows some ability to predict refolding rates for globular proteins that fold via two-state mechanisms, it correlates only weakly with refolding rates for proteins that refold via more complicated kinetic mechanisms.<sup>17</sup>

Most proteins, even for simple organisms, are larger than the proteins commonly used for refolding studies: the average protein size is 267 aa in prokaryotes and 361 aa in eukaryotes.<sup>18</sup> Initial global collapse raises a problem for larger proteins: since the number of possible non-native interactions increases with increasing protein size, interactions formed during chain collapse are less likely to represent native contacts than for smaller proteins. A higher ratio of non-native to native interactions in early refolding intermediates could significantly slow the refolding rate, as conformational searching increases.<sup>19</sup> Such searching could also produce a much rougher energy landscape for refolding, and with that a more complex refolding mechanism,<sup>20</sup> which could lead to aggregation.<sup>21</sup> For globular proteins, formation of long-range tertiary structure interactions is typically the rate-limiting step for refolding.<sup>17,22,23</sup> Hence, it is not surprising that attempts to develop truly general models for protein refolding often fail to explain the refolding behavior of

**Abbreviations:** ANS, 8-anilino-1-naphthalenesulfonic acid; a.u., arbitrary units; BCA, bovine carbonic anhydrase; CD, circular dichroism; GdnHCl, guanidinium hydrochloride; IPTG, isopropyl-galactopyranoside; LB, Luria–Bertani broth; PDB, protein data bank; PFS, partly folded state; proK, proteinase K; SDS, sodium dodecyl sulfate.

Grant sponsor: National Science Foundation CAREER Award; Grant number: MCB-0237934; Grant sponsor: American Heart Association Scientist Development; Grant number: 0330151N.

\*Correspondence to: Department of Chemistry and Biochemistry, University of Notre Dame, 251 Nieuwland Science Hall, Notre Dame, IN 46556-5670. E-mail: pclarck1@nd.edu.

Received 31 July 2009; Revised 21 August 2009; Accepted 28 August 2009

Published online 11 September 2009 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.22609

larger proteins<sup>24</sup> or proteins with more complex multi-exponential kinetics.<sup>17</sup>

Moreover, it is quite likely that the fundamental features of protein refolding mechanisms will also vary depending on secondary structure content,<sup>25</sup> and yet model proteins selected for refolding studies are often rich in  $\alpha$ -helical structure. Alpha-helical proteins tend to aggregate less than  $\beta$ -sheet proteins, which can simplify *in vitro* refolding experiments but provides no information on how more aggregation-prone structural topologies might avoid aggregation during refolding. The nature of  $\beta$ -sheet structure means that  $\beta$ -sheet proteins have a higher percentage of residues that form contacts with residues distant in the primary sequence than  $\alpha$ -helical proteins. This higher contact order, and likewise broader landscape for conformational searching during refolding, could also result in slower kinetics and more complex refolding mechanisms. Indeed, a rougher energy landscape for the refolding of all- $\beta$ -sheet proteins, including formation of  $\beta$ -sheet structure in the rate-limiting refolding step, has been observed experimentally.<sup>17,26–32</sup> As a result, refolding studies of larger all- $\beta$ -sheet proteins can provide valuable information regarding the refolding properties of these more aggregation-prone proteins.

An additional complication in the search for fundamental features of protein refolding mechanisms arises from the wide variety of ways that the basic building blocks of protein structural motifs can be assembled into a native protein structure. To reduce this complication, repeat proteins have emerged as a valuable tool to determine the effects of increasing polypeptide chain length on refolding mechanisms. Repeat proteins are comprised of many copies of only a single structural unit, and this linear structure results in relatively low-contact order. Because of the lack of very long-range interactions, the addition of repeats does not increase the structural complexity of the folded protein.<sup>33,34</sup> Interestingly, even though complexity or contact order does not increase with repeat number, larger repeat proteins commonly have slower, more complex refolding kinetics than small repeat proteins.<sup>33</sup> In fact, many repeat proteins refold surprisingly slowly, based on their low-contact order.<sup>35</sup> Clearly, to deduce truly general features of protein refolding mechanisms, we must study the refolding of proteins that are at and above average size.

The  $\beta$ -helix is an all- $\beta$ -sheet repeat structure that permits a systematic approach to the study of parallel  $\beta$ -sheet formation.<sup>36,37</sup> Each repeat of a  $\beta$ -helix consists of a rung with three  $\beta$ -strands connected by loops of various lengths. To date, the refolding of two  $\beta$ -helix proteins, *Salmonella* phage P22 tailspike, and pectate lyase C (PelC) from *Erwinia chrysanthemi*, has been studied in depth. Although these studies have identified specific residues important for refolding,<sup>38–40</sup> and permitted characterization of early folding intermediates,<sup>41–43</sup> the

refolding pathways of these proteins are complicated by additional domains and multimerization (for tailspike),<sup>44</sup> and rate-limiting peptidyl-prolyl isomerization (for PelC).<sup>43</sup>

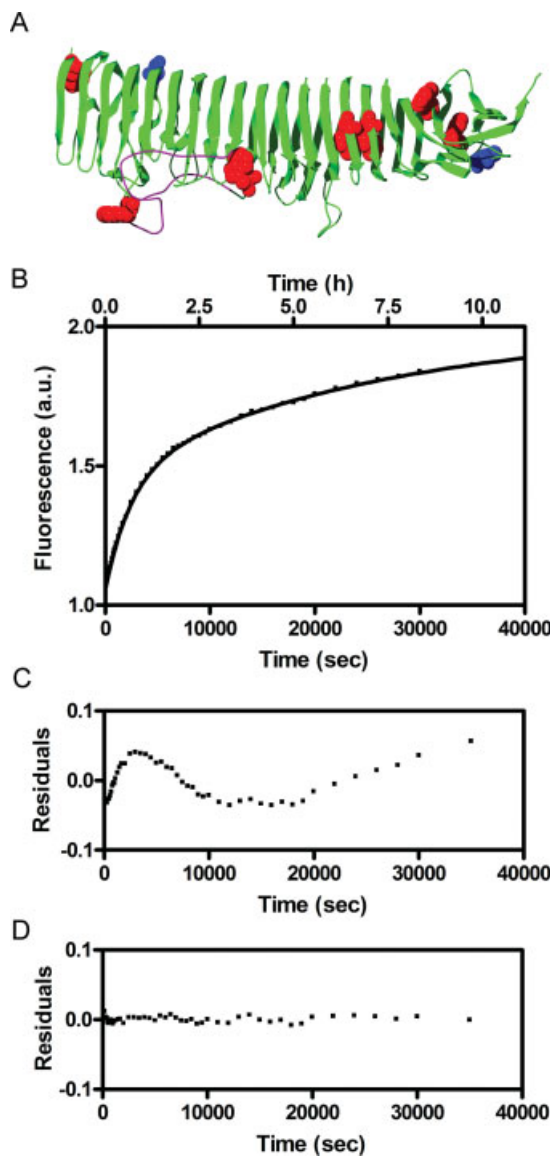
Pertactin is a large monomeric  $\beta$ -helical protein and a virulence factor from *Bordetella pertussis*, the causative agent of whooping cough.<sup>45</sup> The crystal structure of the mature protein<sup>46</sup> reveals pertactin as a 55 kDa, single-domain protein with a 16-rung right-handed parallel  $\beta$ -helix; the  $\beta$ -strands within each rung are connected by short loops or turns [Fig. 1(A)]. This monomeric  $\beta$ -helix is free of cysteines and cofactors that might complicate interpretation of refolding results. Previous studies have shown that pertactin refolds surprisingly slowly, on the hours time scale,<sup>47</sup> orders of magnitude slower than rates predicted from contact order (microseconds to seconds time scale).<sup>16,48</sup> Indeed, the size, topology, and refolding rate of pertactin lie far outside the typical parameters used to develop contact order-based models of refolding kinetics. Here, we report the mechanism of slow refolding of pertactin and determine the rate-limiting step for formation of the  $\beta$ -helix structure. Because of its low propensity to aggregate, pertactin offers the unique possibility to study slow refolding kinetics of a protein with a  $\beta$ -sheet repeat structure. Furthermore, pertactin allows an examination of the formation of  $\beta$ -strands in a larger protein without interfering signals from other structural motifs within the protein.

## RESULTS

### Pertactin refolding is extremely slow and biphasic

We studied the overall kinetics of pertactin refolding, exploiting the seven tryptophan residues that are spread throughout the entire  $\beta$ -helix; these tryptophans vary significantly in their solvent exposure in the native state [Fig. 1(A)]. During refolding, the pertactin fluorescence emission spectrum undergoes changes in both total intensity and maximum emission wavelength.<sup>47</sup> Hence, in addition to the fluorescence emission intensity at 335 nm, we also followed pertactin refolding by measuring the ratio of emission intensities at 335 and 350 nm, particularly for refolding reactions with half-times  $>10$  min. The use of a ratio, rather than absolute intensity, avoids artifacts arising from photobleaching or fluctuations in lamp intensity or photomultiplier tubes, which are more significant for extremely slow refolding reactions.

As shown in Figure 1(B), pertactin refolding is surprisingly slow, requiring  $>10$  h to complete. Kinetic traces required fitting to two exponential rates, as indicated by the residuals in Figure 1. Both rates ( $4.7 \times 10^{-4} \text{ s}^{-1}$  and  $5.9 \times 10^{-5} \text{ s}^{-1}$ ) contributed similar amplitudes to the total observable reaction (Table I). In addition, a burst phase was observed, accounting for about 40% of the



**Figure 1**

Pertactin refolding is very slow and multiexponential. (A) Crystal structure of pertactin (PDB ID: 1dab).<sup>46</sup> Tryptophans are highlighted as red space-filling models and the location of the residues used for site-specific labeling (S117 and T490) are shown in blue. The long loop, containing both *cis*-prolines, is highlighted in purple. (B) Refolding kinetics of pertactin are shown as a function of intrinsic tryptophan fluorescence emission. Shown is the fit to two exponential functions. (C) Fitting to one exponential fit produces nonrandom residuals; (D) after adding a second exponential function, residuals have a random distribution.

total change in signal during refolding. This burst phase could not be fully resolved in stopped flow experiments (dead time  $\sim 20$  ms; data not shown) implying the fast formation of a structured intermediate during the earliest stages of refolding. Approximately 10% of the total amplitude change required more than 12 h to complete (see Fig. 1). This extremely slow kinetic rate could not be

reproducibly quantified, presumably because, over such long-time scales, otherwise insignificant variables such as sample evaporation become increasingly significant.

#### Pertactin does not populate a classic “molten globule” during slow refolding

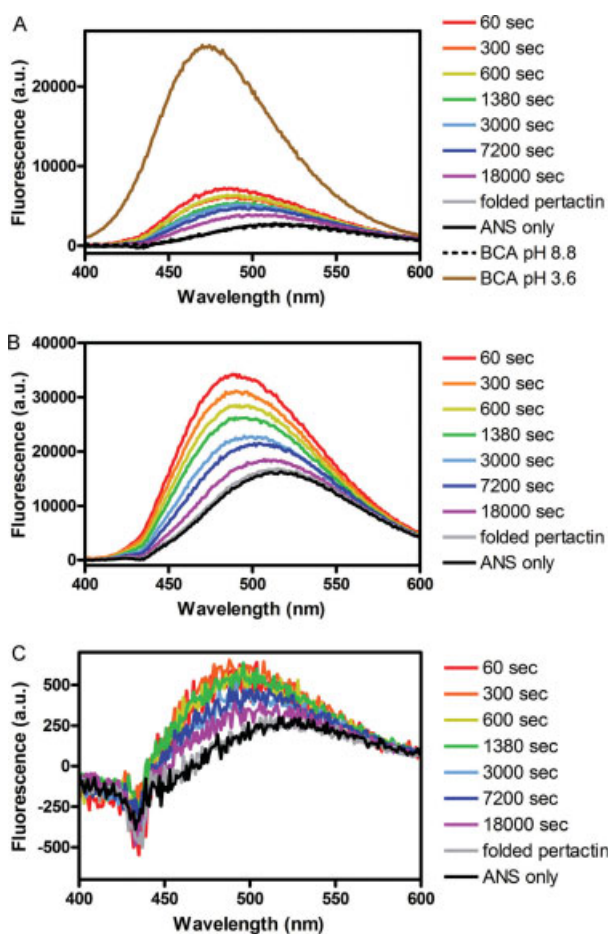
To test for a molten globule-like structure during pertactin refolding, we used 8-anilino-1-naphthalenesulfonic acid (ANS) binding. ANS is known to bind to hydrophobic patches common to early protein refolding intermediates, with binding leading to a significant increase in ANS fluorescence emission.<sup>49,50</sup> For pertactin, however, we observed only a small, 2- to 3-fold increase in ANS fluorescence in the presence of pertactin refolding intermediates [Fig. 2(A)]. In contrast, ANS binding to the known molten globule state of bovine carbonic anhydrase (BCA) produced a more typical  $>10$ -fold increase in ANS fluorescence [Fig. 2(A)].<sup>50</sup> To determine whether differences in the conformations of refolding intermediates and/or the size difference between BCA and pertactin (29 vs. 54 kDa) might alter the number of ANS binding sites on a possible molten globule state, we also tested ANS binding to pertactin refolding intermediates using higher or lower ANS concentrations [Fig. 2(B,C)]. These experiments produced a similar 2 to 3-fold increase in ANS fluorescence, indicating that these pertactin refolding intermediates do not populate a conformation with a large, solvent exposed cluster of hydrophobic residues.

#### Pertactin refolding is not limited by proline isomerization

At room temperature, *cis/trans* isomerization of X-Pro peptide bonds typically occurs on a time scale of  $0.001 \text{ s}^{-1}$ , slower than the refolding of many small, globular proteins,<sup>51</sup> and indeed represents the rate-limiting refolding step for many small proteins.<sup>52</sup> Although the pertactin refolding rates reported earlier are slower than reported rates for X-Pro isomerization, the pertactin sequence includes 28 prolines, and two of the X-Pro peptide bonds are in the *cis*-configuration in the crystal structure (P241 and P256).<sup>46</sup> Both of these *cis*-Pro peptide bonds are located in the long irregular loop that connects rungs 8 and 9 [Fig. 1(A)], and are therefore peripheral to the core of the  $\beta$ -helical structure. Nevertheless, it is still possible that their isomerization could slow the overall pertactin refolding rate. To test the effect of proline isomerization on pertactin refolding kinetics, we used several common techniques. We compared pertactin refolding kinetics with or without the addition of the peptidyl-prolyl isomerases cyclophilin A or FKBP506 (Fig. 3, Table I). In both cases, refolding rates and amplitudes were within experimental error of the uncatalyzed refolding reaction (Table I). Enzymatic catalysis can, however, be limited by the accessibility of the peptide

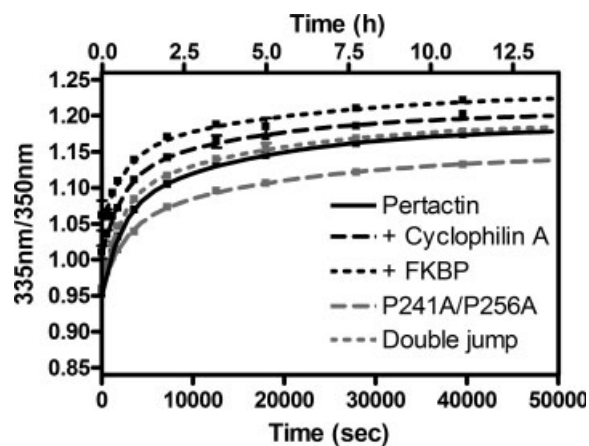
**Table I**Summary of Pertactin Refolding Reactions at 0.4M GdnHCl<sup>a</sup>

	$k_1$ (s <sup>-1</sup> )	$k_2$ (s <sup>-1</sup> )	A <sub>1</sub> (%)	A <sub>2</sub> (%)
Wild type (tryptophans)	$4.7 \times 10^{-4} \pm 0.4 \times 10^{-4}$	$5.9 \times 10^{-5} \pm 0.5 \times 10^{-5}$	51 ± 6	49 ± 6
Wild type (CD)	$3.4 \times 10^{-4} \pm 0.2 \times 10^{-4}$	$5.8 \times 10^{-5} \pm 0.8 \times 10^{-5}$	49 ± 4	51 ± 3
Wild type (double jump)	$5.2 \times 10^{-4} \pm 0.4 \times 10^{-4}$	$5.6 \times 10^{-5} \pm 0.5 \times 10^{-5}$	56 ± 6	44 ± 5
Wild type + cyclophilin A	$4.3 \times 10^{-4} \pm 0.9 \times 10^{-4}$	$5.8 \times 10^{-5} \pm 1.5 \times 10^{-5}$	55 ± 21	45 ± 19
Wild type + FKBP	$3.7 \times 10^{-4} \pm 1.2 \times 10^{-4}$	$4.4 \times 10^{-5} \pm 1.9 \times 10^{-5}$	54 ± 34	46 ± 30
P241A/P256A	$4.0 \times 10^{-4} \pm 0.5 \times 10^{-4}$	$5.1 \times 10^{-5} \pm 0.6 \times 10^{-5}$	46 ± 8	54 ± 8
S117C				
Tryptophan	$4.6 \times 10^{-4} \pm 0.3 \times 10^{-4}$	$6.3 \times 10^{-5} \pm 0.5 \times 10^{-5}$	45 ± 4	55 ± 4
Fluorescein	$2.9 \times 10^{-4} \pm 0.5 \times 10^{-4}$	$5.9 \times 10^{-5} \pm 1.4 \times 10^{-5}$	52 ± 19	48 ± 17
F490C				
Tryptophan	$3.7 \times 10^{-4} \pm 0.5 \times 10^{-4}$	$4.1 \times 10^{-5} \pm 2.0 \times 10^{-5}$	50 ± 13	50 ± 13
Fluorescein	$3.4 \times 10^{-4} \pm 0.2 \times 10^{-4}$	$7.5 \times 10^{-5} \pm 1.1 \times 10^{-5}$	60 ± 9	40 ± 6

<sup>a</sup>The indicated error represents the standard error from three independent experiments.**Figure 2**

The early pertactin refolding intermediate is not a molten globule-like state. Fluorescence emission spectra of ANS added to pertactin at different refolding times. At the indicated refolding times, pertactin samples were mixed with (A) 10  $\mu$ M, (B) 100  $\mu$ M or (C) 1  $\mu$ M ANS, and ANS fluorescence emission was measured. Each spectrum represents an average of three independent experiments. Fluorescence spectra of BCA at pH 3.6 and 8.8 (A) demonstrate the typical 10-fold increase in ANS fluorescence intensity for a classical molten globule. The spectra of ANS only, ANS with folded pertactin, and ANS with BCA pH 8.8 all overlay at the lowest intensity.

bond. Therefore, we also used double jump experiments, in which pertactin was unfolded in 4M GdnHCl for 10 s, followed by immediate dilution to refolding conditions. The double jump refolding rates and amplitudes were within experimental error of the values measured for refolding after full equilibration to the unfolded state (Table I). However, the slow refolding kinetics of pertactin would allow proline isomerization to occur during the folding process itself, making it impossible to conclusively rule out an effect of proline isomerization in these double jump experiments. Hence, we also mutated both *cis*-prolines to alanine. Refolding kinetics and corresponding amplitudes of the single mutants and the double mutant were within experimental error of wild-type

**Figure 3**

Proline isomerization is not a rate-limiting step during refolding. Wild-type pertactin was refolded alone (black, solid line), or with the addition of the peptidyl-prolyl-isomerases cyclophilin A (black, dashed line) or FKBP506 (black, dotted line). Also shown is the refolding of the double mutant in which both *cis*-prolines have been replaced by alanines (grey, dashed line) as well as a double-jump experiment in which wild-type pertactin was unfolded for only 10 s prior to refolding, to prevent isomerization of prolines to a non-native configuration (grey, dotted line).

pertactin (Table I), further demonstrating that proline isomerization is not a rate-limiting step for pertactin refolding.

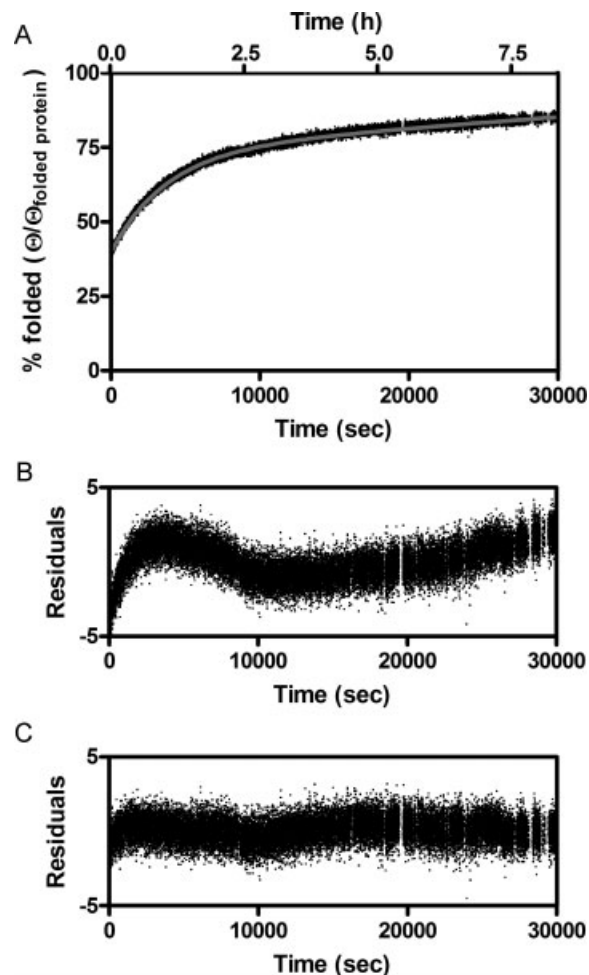
### **$\beta$ -Sheet secondary structure forms during the rate-limiting refolding step**

As the isomerization of peptidyl-prolyl bonds can be excluded as the rate-limiting step, we also examined the rate of secondary structure formation. The pertactin far-UV CD spectrum has a single distinct minimum at 218 nm, typical for  $\beta$ -sheet proteins.<sup>47</sup> We followed the development of this spectral signature during refolding, in a manner analogous to the fluorescence experiments described earlier. A slow biphasic transition from unfolded to folded protein was observed, with kinetics similar to those measured for tryptophan fluorescence during refolding (Fig. 4, Table I). The amplitudes of these two rates are comparable with the amplitudes measured in the fluorescence experiments, and a burst phase with a similar magnitude seen for fluorescence experiments (40%) was also observed for the formation of secondary structure. These data indicate that secondary structure formation occurs during all stages of pertactin refolding, including the rate-limiting step. However, these results do not allow further elucidation of the actual refolding mechanism of pertactin in terms of the structures of refolding intermediates.

### **Equilibrium partly folded state is not a kinetic refolding intermediate**

Pertactin populates a partly folded state (PFS) in equilibrium unfolding/refolding titrations. Analysis of this PFS concluded that the N-terminus is unstructured while the C-terminus is stably folded and resistant to proteolytic digestion.<sup>47</sup> Nevertheless, the origin of the enhanced C-terminal stability is not apparent from the crystal structure of native pertactin.<sup>46</sup> At equilibrium, unfolding from  $F \rightarrow PFS$  and  $PFS \rightarrow U$  each represent  $\sim 50\%$  of the signal change observed by fluorescence and far-UV circular dichroism (CD). Intriguingly, the amplitudes of each of the two observable kinetic phases during refolding also represented  $\sim 50\%$  of the total observable intensity change, raising the possibility that the equilibrium PFS might resemble a kinetic refolding intermediate.

To test the possibility of a kinetic refolding pathway with an intermediate that resembles the PFS, that is, with a folded C-terminus and unfolded N-terminus, we compared the kinetics for refolding starting at the PFS with reactions for complete refolding ( $U \rightarrow F$ ). If the PFS is located along the refolding coordinate between U and F, the maximal activation barrier between PFS and U can be no greater than the highest activation barrier between U and F. Surprisingly, however, pertactin refolding kinetics from the PFS are several orders of magnitude slower than rates measured for the complete  $U \rightarrow F$

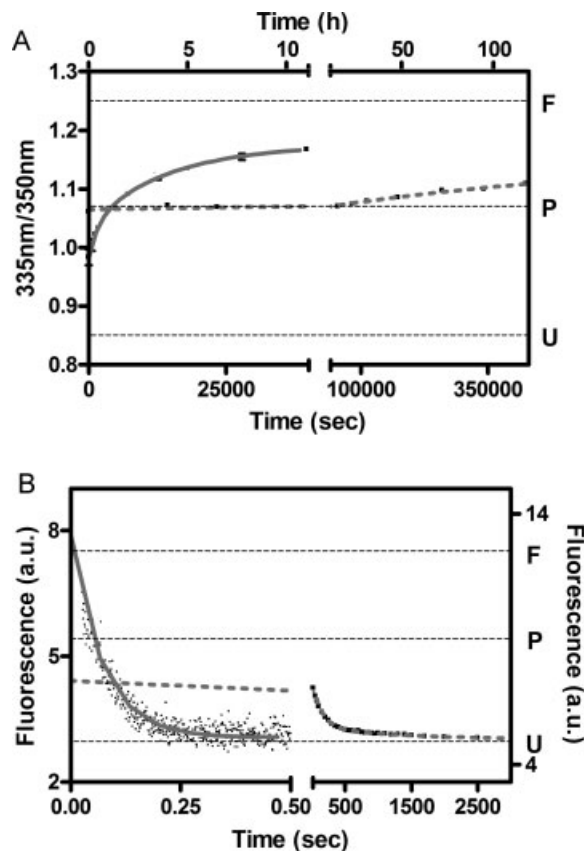


**Figure 4**

Formation of secondary structure occurs during the rate-limiting step of pertactin refolding. Refolding of pertactin from 4 to 0.4M GdnHCl followed by circular dichroism at 218 nm. (A) Shown is the percentage of folded protein based on the CD signals for folded and unfolded protein at 218 nm. Kinetic data were fit to two exponential functions (grey) and the quality of fits is indicated by the residuals for fit with (B) one exponential and (C) two exponential functions.

refolding reaction [Fig. 5(A), Table II]. Hence, the activation barrier between PFS and F is significantly higher than the barrier between U and F. Because  $\sim 10\%$  of the refolding transition occurs slower than 12 h (see earlier), it cannot be ruled out that some pertactin polypeptide chains might populate the PFS en route to the native state, but clearly the PFS does not represent a major on-pathway intermediate for complete refolding of the pertactin  $\beta$ -helix.

We also used diagnostic differences in the proteolytic digestion patterns of the native and PFS as a readout for formation of native pertactin. Native pertactin is largely resistant to proteolytic digestion, while digestion of the PFS leads to a characteristic fragment of 21 kDa.<sup>47</sup> Lim-



**Figure 5**

The pertactin partly folded state is not an on-pathway kinetic refolding intermediate. **A:** Tryptophan fluorescence kinetics for refolding reactions from U to F (solid line) and PFS to F (dotted line). Samples were incubated in 4M and 1.2M GdnHCl, respectively, until equilibrium was reached. Refolding was initiated by diluting samples into refolding buffer leading to a final protein concentration of 0.5  $\mu$ M in 0.4M GdnHCl. Shown is the change in the ratio of tryptophan fluorescence at 335 and 350 nm over time. Samples were incubated at 20°C and each time point represents a fresh aliquot from the master refolding reaction; this strategy minimized interference from photobleaching during the experiment. Results from at least three independent reactions were averaged. Equilibrium fluorescence intensities for the folded (F), partly folded (P), and unfolded state (U) are indicated by the dotted lines. **B:** Kinetic traces for unfolding from F to U (solid line) and PFS to U (dotted line). Folded and partly folded protein (pre-equilibrated in 1.2M GdnHCl) were unfolded in 6M GdnHCl. Kinetics for the reaction from F to U were measured using a stopped-flow apparatus, whereas reactions from P to U were measured using manual mixing techniques. Pertactin concentrations were 5  $\mu$ M for stopped flow and 0.5  $\mu$ M for manual mixing experiments. The expected values for folded, partly folded, and unfolded protein under both conditions are indicated on the right.

ited proteolytic digestion of refolding time points [Fig. 6(A)] confirmed that the PFS can indeed refold to native protein, as indicated by the appearance of protease resistant, full length pertactin, along with minor digestion fragments typically seen for native pertactin.<sup>47</sup> These characteristic digestion patterns can also be used to determine if the PFS is a refolding intermediate during

refolding of pertactin. As shown in Figure 6(B), during refolding, the 21 kDa digestion fragment of the PFS is not detected, confirming that the PFS is not a significant on-pathway intermediate for pertactin refolding.

To further analyze the location of the PFS on the refolding/unfolding reaction coordinate, we also compared unfolding kinetics starting from either 0M (F) or 1.2M GdnHCl (PFS). Although native pertactin is completely unfolded after 1 s of incubation in 6M GdnHCl, the PFS requires more than 30 min to denature under the same conditions [Fig. 5(B); Table II], supporting the idea that the PFS is an off-pathway intermediate. In addition, though complete unfolding (F→U) could be fitted to a single exponential function, unfolding from the PFS required two exponential functions in addition to a burst phase within the dead time of the experiment. The difference in refolding and unfolding kinetics shows that the PFS is not a major on-pathway intermediate for either complete refolding or unfolding of pertactin.

### Chevron plot analysis

Ideally, we would like to determine the relevance of the PFS to unfolding/refolding reactions under physiological conditions, at 0M GdnHCl. Towards this goal, we measured refolding and unfolding rates for the pertactin native (folded), partly folded, and unfolded states, and plotted the kinetic rates as a function of the final GdnHCl concentrations to produce a “chevron plot.” Small proteins typically have unfolding and refolding rates that exhibit a characteristic linear or quasi-linear dependence on denaturant concentration, meaning that these rates can be extrapolated to 0M denaturant to determine the rates and with that the free energy change on unfolding/refolding in buffer. The intersection point of the two arms of the chevron plot typically represents the midpoint of the transition between folded and unfolded structure at equilibrium. In numerous cases, the folding and/or unfolding arm show deviation from strictly linear behavior, in particular at denaturant concentrations far removed from the equilibrium transition midpoint. Although the origin(s) of nonlinear behavior is still controversial, these deviations are often thought to arise due to the presence of folding intermediates and/or broad transition states.<sup>53–55</sup>

The pertactin chevron plot contains three refolding and three unfolding arms, representing the three stable states of pertactin (F, PFS, and U) (see Fig. 7). The presence of two independent transitions at equilibrium restricts our kinetic measurements to narrow ranges of final GdnHCl concentrations: The folding kinetics of all three reactions (U → P, P → F, and U → F) were each determined over a range of final GdnHCl concentrations spanning less than 1M. Moreover, for reactions from F → P, it proved impossible to deconvolve the rates corresponding to unfolding from F → P from rates corre-

**Table II**

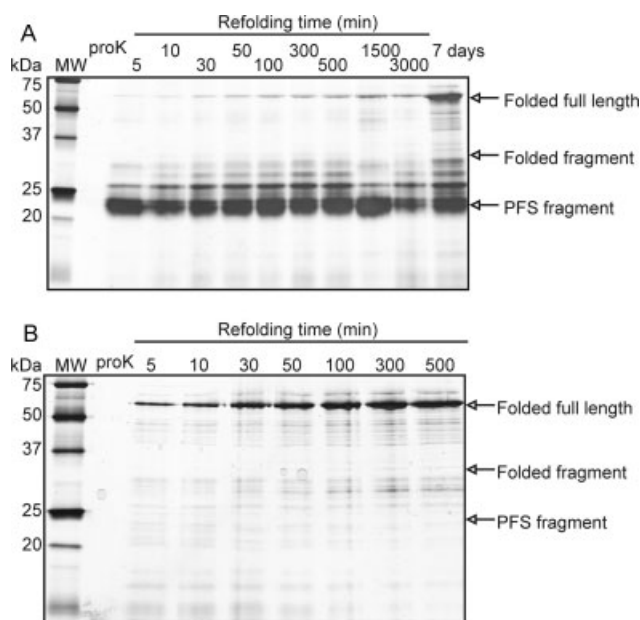
Comparison of Complete Refolding/Unfolding Rates with Refolding (to 0.4M GdnHCl) and Unfolding (to 6M GdnHCl) Starting at the PFS (1.2M GdnHCl)<sup>a</sup>

	$k_1$ (s <sup>-1</sup> )	$k_2$ (s <sup>-1</sup> )	$A_1$ (%)	$A_2$ (%)
<b>Unfolding</b>				
F → U	$1.4 \times 10^1 \pm 0.1 \times 10^1$	NA	NA	NA
P → U	$6.9 \times 10^{-3} \pm 0.3 \times 10^{-3}$	$6.4 \times 10^{-4} \pm 0.9 \times 10^{-4}$	$78 \pm 4$	$22 \pm 2$
<b>Refolding</b>				
U → F	$5.5 \times 10^{-4} \pm 3.3 \times 10^{-4}$	$6.9 \times 10^{-5} \pm 2.8 \times 10^{-5}$	$50 \pm 13$	$50 \pm 13$
P → F	$1.5 \times 10^{-6} \pm 0.9 \times 10^{-6}$	NA	NA	NA

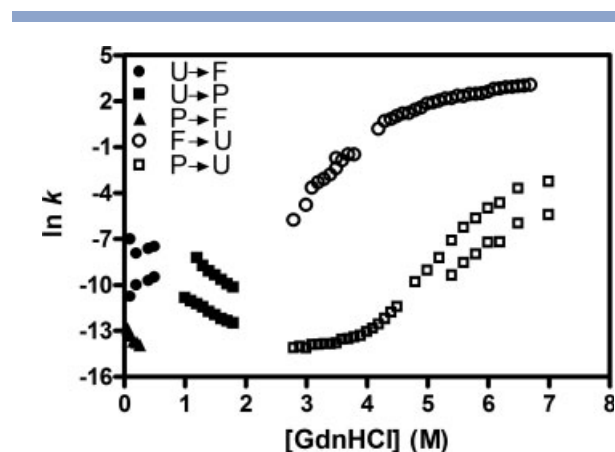
<sup>a</sup>The indicated error represents the standard error from three independent experiments.

sponding to F → U. The overlapping kinetics made a detailed characterization of the unfolding pathway impossible. These limitations unfortunately prohibit robust analysis of the resulting chevron plot with linear fits, and hence an accurate extrapolation to 0M GdnHCl. Notably, however, the majority of pertactin refolding and unfolding reactions have rates below  $1 \times 10^{-3} \text{ s}^{-1}$ , further demonstrating that most pertactin  $\beta$ -helix refolding and unfolding processes are very slow. We measured reaction rates between  $20 \text{ s}^{-1}$  and  $1 \times 10^{-6} \text{ s}^{-1}$ , but in most reactions, even this broad time window was not sufficient to cover the entire amplitude of the corresponding reaction. In

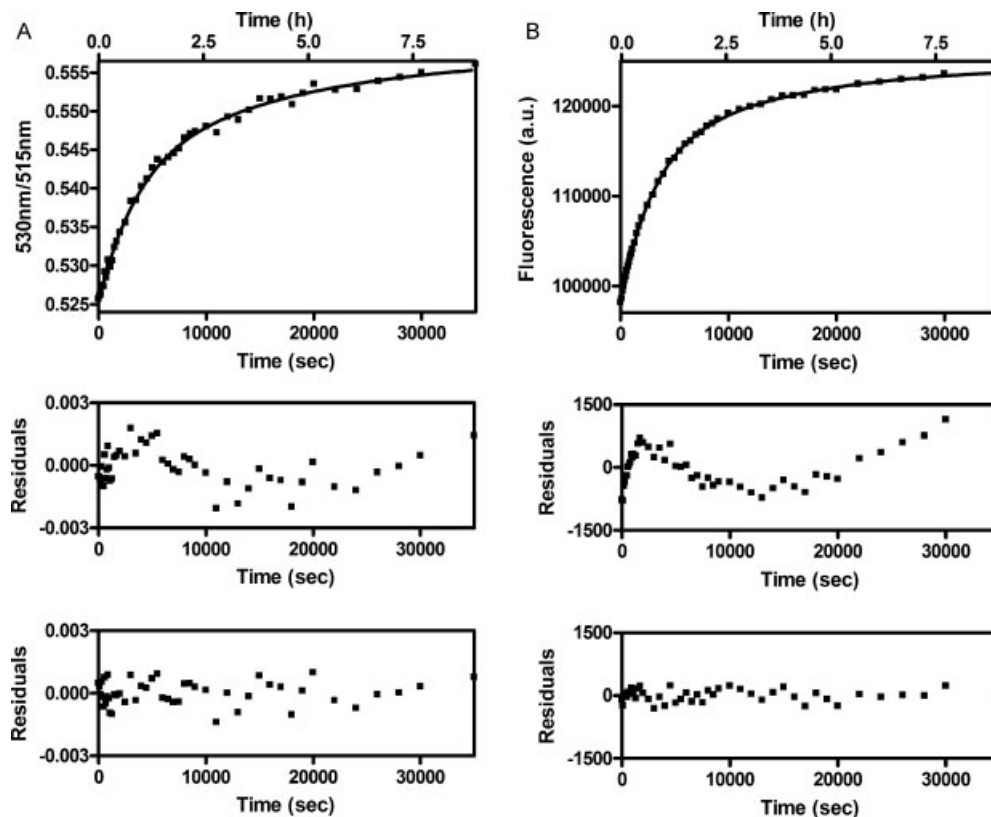
particular, the challenge of resolving very slow reactions with half times significantly longer than 1 day inhibited exhaustive analyses of the entire refolding (or unfolding) reaction. Interestingly, most reactions exhibit at least two independent folding or unfolding exponential rates, and include rollover in the unfolding reactions. These observations support the presence of intermediates and/or broad transition states during the refolding and unfolding reactions. These results inspired us to ask whether refolding kinetics can be used to distinguish between the entire protein folding in a concerted process, or if these multiple refolding rates represent different parts of the protein refolding independently.

**Figure 6**

Limited proteolytic digestion of pertactin during refolding from unfolded and PFS shows pertactin refolding is reversible, but the PFS is not a refolding intermediate. **A:** ProK digestion of pertactin refolding from 1.5 to 0.5M GdnHCl. Samples for each refolding time point were digested with proK for 50 min and analyzed by SDS-PAGE followed by silver staining. **B:** ProK digestion of pertactin refolding from 4 to 0.5M GdnHCl after various refolding incubation times. The characteristic digestion bands for the native protein and PFS<sup>47</sup> are indicated on the right.

**Figure 7**

Chevron plot of pertactin unfolding and refolding kinetics indicates complex pathways. All refolding and unfolding reactions measured the change in pertactin intrinsic tryptophan fluorescence, expressed as either fluorescence intensity at 335 nm (fast reactions) or the ratio of emission between 335 and 350 nm (slow reactions). Indicated GdnHCl concentrations represent the final reaction conditions. Refolding of the unfolded state started at 4M, refolding and unfolding of the PFS started at 1.2M, and unfolding of the folded protein started at 0M GdnHCl. All data was first fit to a single exponential rate; a second rate was introduced only if the single-exponential fit produced nonrandom residuals. Open symbols represent unfolding reactions; closed symbols represent refolding reactions. For all reactions, at least three independent experiments were performed. The data points correspond to the average of these measurements; the standard deviation for each rate is smaller than the size of the data point.



**Figure 8**

Refolding of the pertactin  $\beta$ -helix is a concerted process. **A:** Refolding kinetics of labeled S117C pertactin followed by fluorescein fluorescence emission. Refolding was measured using the ratio in fluorescence intensity between 530 nm and 515 nm. Data required fitting with two exponential functions, as indicated by the residuals for one exponential function (middle) that exhibit an obvious trend. This trend in the residuals was lost with two exponential functions (bottom). **B:** Refolding kinetics of labeled T490C pertactin followed by fluorescein fluorescence emission. Samples were prepared and measured following the same procedure as in (A). Raw change in fluorescence at 515 nm was analyzed and data required fitting to two exponential functions as shown by the residuals for the fit with one (middle) and two (bottom) exponential functions.

### Pertactin $\beta$ -helix folds in a concerted, multistep process

The seven tryptophans in the pertactin sequence are spread throughout the  $\beta$ -helix structure. Therefore, the change in tryptophan fluorescence provides only an overall view of the refolding of the complete  $\beta$ -helix; the intrinsic fluorescence cannot be used to distinguish between refolding at the N-terminus versus C-terminus. A similar limitation applies to the formation of  $\beta$ -sheet structure followed by far-UV CD. Both techniques cannot distinguish between vectorial versus concerted refolding of the  $\beta$ -helix. These two models can, however, be distinguished by kinetic experiments using site-specific fluorescent probes. We therefore introduced a single cysteine residue at a surface-exposed location near either the N- or C-terminus of pertactin [see Fig. 1(A) and Materials and Methods section] and labeled each of these cysteine residues with fluorescein. As shown in Table I, changes in tryptophan fluorescence during refolding of the mutant proteins have similar rates and amplitudes as wild-type

pertactin, indicating that the refolding behaviors of the labeled, single cysteine constructs are comparable with wild-type pertactin. Thermal denaturation also showed no significant destabilization of either the N-terminus or C-terminus for either labeled construct (data not shown).

Changes in the fluorescence properties of the site-specific probes during pertactin refolding allowed us to determine whether one end of the pertactin  $\beta$ -helix folds first, or if the entire  $\beta$ -helix forms in a global, concerted multistep process. For both the N- and C-terminal fluorescein labels, kinetic refolding traces required fitting to two kinetic rates with roughly equal amplitudes (see Fig. 8). These rates and amplitudes are comparable with the values obtained for tryptophan fluorescence and far-UV CD measurements of wild-type pertactin (Table I, Figs. 4 and 8). Therefore, the N-terminus and C-terminus of pertactin fold on similar time scales, indistinguishable from global measurements of secondary or tertiary structure formation. These results indicate pertactin refolds in a concerted process *in vitro*, with significant formation of

secondary structure occurring even in the rate-limiting step.

## DISCUSSION

Extensive studies of the refolding of a variety of small, globular, and often  $\alpha$ -helix-rich proteins have identified some common properties of refolding mechanisms, including global collapse of hydrophobic residues in the initial phase of refolding, typically occurring within microseconds to milliseconds or faster.<sup>56,57</sup> Most proteins with refolding kinetics significantly slower than predicted have refolding mechanisms limited by proline isomerization, multimerization, minor reorientations to a mostly natively folded structure, or other steps not directly related to refolding.<sup>35,58–60</sup> However for pertactin, proline isomerization is not the rate-limiting step for refolding; instead, a significant fraction of secondary structure forms within the final step of pertactin refolding.

It is important to point out that previous equilibrium titration experiments demonstrated that pertactin refolding is fully reversible.<sup>47</sup> Hence, despite a high activation barrier between the unfolded and folded states, and the prolonged population of  $\beta$ -sheet-rich folding intermediates, pertactin is not prone to aggregation during refolding. The slow refolding kinetics implies an extended search around the edge of a refolding funnel, populating conformations that can somehow avoid sliding into an “aggregation funnel.”<sup>21</sup> Importantly, even though other proteins have been identified for which secondary structure formation occurs in the later steps of refolding,<sup>32,61,62</sup> these proteins refold orders of magnitude faster than pertactin.

To further evaluate pertactin refolding properties under native conditions, we analyzed the denaturant dependence of the refolding and unfolding reactions using a chevron plot. However, pertactin is an extremely slow-folding protein with complex kinetic behavior and an equilibrium PFS. Hence, the complexity of the pertactin chevron plot does not permit a detailed analysis of the refolding and unfolding pathways in a manner analogous to analyses of smaller proteins with simpler kinetic behavior. For example, most pertactin unfolding and refolding reactions require fitting to multiple exponential rates, and all include either burst-phase changes or extremely slow steps not measurable by conventional methods. Near the  $F \rightleftharpoons PFS$  and  $PFS \rightleftharpoons U$  transitions, there is overlap of the forward and backward reactions, meaning that these rate constants cannot be attributed to a specific refolding or unfolding pathway. Furthermore, many reactions cannot be measured over a broad range of GdnHCl concentrations, which prohibits extrapolation of the refolding and unfolding rates to 0M GdnHCl.

In light of these complexities, we have used other approaches to characterize the pertactin refolding mechanism. One interesting feature is the lack of a classical

molten globule state before the formation of tertiary structure and significant amounts of secondary structure. The low intensity of ANS fluorescence argues against the formation of large, exposed, hydrophobic surface area during the majority of the refolding reaction. Instead, pertactin refolding intermediates might include only smaller clusters of hydrophobic residues. The lack of a classical molten globule state before formation of tertiary structure is unusual for all- $\beta$ -sheet proteins, and in particular for  $\beta$ -helix proteins. ANS binding to PelC, another  $\beta$ -helix protein, is characterized by a 50-fold increase in ANS fluorescence,<sup>42</sup> even though PelC has a lower percentage of hydrophobic residues than pertactin ( $\sim 50\%$  vs.  $\sim 60\%$ ). In addition, refolding studies on PelC have demonstrated that proline isomerization represents the rate-limiting step for refolding.<sup>43</sup> It should be noted, however, that PelC is significantly shorter (353 aa) than pertactin, resulting in an overall shape that is more similar to globular proteins. In fact, the elongated shape of pertactin raises interesting questions regarding its refolding mechanism. For pertactin, an initial collapsed, molten globule-like state could actually increase the complexity of the refolding mechanism, as it would likely include a large number of non-native, long-range interactions. Therefore, the results for pertactin show that protein size, shape, and structure can significantly alter the protein refolding mechanism.

Site-specific labeling shows that the entire pertactin  $\beta$ -helix refolds in a concerted, multistep process. Such a mechanism might include various kinetic intermediates in which all intermediates produce conformational changes throughout the entire  $\beta$ -helix. For example, an initial collapsed state (but not one that exposes large patches of hydrophobic surface area) might include mispairing of  $\beta$ -strands, followed by sequential rearrangements to form the native  $\beta$ -strand architecture. Alternatively, the protein might refold on different, parallel pathways, with each pathway representing a single, concerted folding step. Regardless, this concerted refolding reaction is surprising when we compare the refolding behavior of pertactin to other repeat proteins. I $\kappa$ B $\alpha$ ,<sup>63</sup> p16<sup>INK4A</sup> (Ref. 64), and other  $\alpha$ -helical ankyrin repeat proteins often have a core domain that folds first, independent of the refolding of other repeat units.<sup>33</sup> Similar behavior was observed for a mixed  $\alpha/\beta$  leucine-rich repeat protein, which forms a folding intermediate with core group of repeats, independent of the number of additional repeats.<sup>65</sup> However, despite the presence of a PFS in pertactin equilibrium titrations, pertactin shows no evidence for a stable core relevant for kinetic refolding. A reason for the lack of a kinetic folding “core” could be the all- $\beta$ -sheet structure of pertactin. Most repeat proteins studied to date have substantial  $\alpha$ -helical structure and hence local interactions that could stabilize a core structure early in the refolding process. In contrast, each rung of the pertactin  $\beta$ -helix has an average size of  $\sim 30$  residues,

and multiple rungs are required to form backbone hydrogen bonds. This lack of short range interactions requires that even a small core structure comprising just two to three rungs would require ordering  $\sim$ 60–90 residues.

Interestingly, even though the PFS does not appear to be a highly populated refolding intermediate *in vitro*, the stability and folding properties of the C-terminal stable core appear relevant to pertactin processing *in vivo*.<sup>66</sup> During secretion through the outer membrane, pertactin passes through a narrow pore that is too narrow to permit secretion of folded pertactin. We have recently determined that pertactin passes through the outer membrane vectorially, from C-terminus to N-terminus.<sup>63</sup> Hence, the C-terminal stable core portion of pertactin reaches the cell surface first, whereas the N-terminus is still located in the periplasm. It has been proposed that extracellular folding of pertactin might serve as a driving force for outer membrane secretion, which occurs in the absence of external energy sources such as ATP or a proton gradient.<sup>66</sup> Crucially, while full length, unfolded pertactin has a complex refolding reaction *in vitro*, vectorial appearance of pertactin on the cell surface could limit the number of possible interactions possible for its C-terminus. The reduced number of possible interactions during secretion through the outer membrane might increase the population of select kinetic intermediates, which could promote folding of the N-terminal portion of pertactin during secretion. Supporting the possible connection between pertactin C-terminal stability and biogenesis *in vivo*, another large  $\beta$ -helix protein from the same family of autotransporter virulence proteins, Pet from *E. coli*, also populates a C-terminal PFS during denaturation titrations.<sup>66,67</sup>

The pertactin refolding kinetics presented here provide insight into the formation of  $\beta$ -sheet structure, but also raise questions regarding the mechanisms used to prevent aggregation of  $\beta$ -sheet rich proteins during refolding. The pertactin  $\beta$ -helix produces no detectable aggregates under the refolding conditions used here. However, the mechanism by which aggregation is prevented is unknown. Aggregates are typically very stable structures, and in most cases, protein aggregation is irreversible.<sup>21,68</sup> Therefore, a fast refolding reaction can favorably shift the ratio of folded versus aggregated protein toward the native structure by forming the kinetically favored product.<sup>69</sup> However, pertactin refolds over the course of several hours. One clue regarding how these long-lived,  $\beta$ -sheet-rich refolding intermediates might avoid aggregation comes from the minimal increase in ANS fluorescence in the presence of pertactin refolding intermediates. This result suggests that hydrophobic residues are buried within these  $\beta$ -sheet-rich folding intermediates, which would help minimize intermolecular interactions that could lead to aggregation.

Examination of the pertactin crystal structure<sup>46</sup> provides additional clues as to how aggregation might be

avoided. While the central rungs of the pertactin  $\beta$ -helix resembles some models of amyloid structure,<sup>70–72</sup> the most C-terminal  $\beta$ -strands form a distorted cap that might inhibit aggregation and/or tail-to-tail dimerization. In addition, in the N-terminus of the  $\beta$ -helix, the hydrophobic core is shielded by a solvent-exposed salt bridge. Both structures have been implicated as mechanisms to prevent intermolecular association of  $\beta$ -sheets.<sup>73</sup> Because these prevention mechanisms focus on the fully folded structure, rather than the refolding pathway, their role in suppressing aggregation during the initial phases of refolding is less clear.<sup>74</sup> Hence, pertactin offers the unique possibility to study how an all  $\beta$ -sheet protein with slow refolding kinetics avoids misfolding into the stable off-pathway PFS,<sup>47</sup> or aggregation, during refolding.

## CONCLUSIONS

Pertactin is an all- $\beta$ -sheet protein with remarkably slow refolding kinetics. We have shown that the  $\beta$ -helix folds in a concerted process with formation of substantial secondary structure included in the rate-limiting step. Furthermore, while pertactin can form a stable PFS at equilibrium, this conformation is not populated during refolding of the native structure from the denatured ensemble. Surprisingly, despite the slow refolding kinetics, aggregation does not occur during refolding to the native or PFS structures. The mechanism of how pertactin avoids aggregation during refolding is unknown, but the similarity of the  $\beta$ -helix fold to the structures of some amyloid fibrils offers the unique possibility to study formation of this structure without the competing, interfering accumulation of macroscopic aggregates.

## MATERIALS AND METHODS

### Molecular biology

Pertactin was expressed using plasmid pPERPLC02, a derivative of pPERPLC01. For pPERPLC02, S298 was replaced with a proline to create a sequence identical to the pertactin sequence used for crystal structure determination. Cys mutants of S117 and F490 as well as Pro  $\rightarrow$  Ala variants of P241 and P256 [Fig. 1(A)] were constructed by site directed mutagenesis of pPERPLC02 using the QuikChange Mutagenesis Kit (Stratagene).

### Protein purification

Wild-type pertactin and Pro  $\rightarrow$  Ala mutants were expressed and purified as described previously.<sup>47</sup> Briefly, 1 L of LB medium was inoculated with 25 mL overnight culture and shaken at 37°C until the OD<sub>600</sub> reached  $\sim$ 0.6. Protein expression was induced by adding IPTG to a final concentration of 250  $\mu$ M. After shaking for addi-

tional 4 h, cells were pelleted, lysed by sonication and incubation with 1 mg/mL lysozyme in sonication buffer (50 mM Tris pH 8.8, 100 mM NaCl, 10 mM benzamide, and 1 mM EDTA). Inclusion bodies were separated by centrifugation, washed with 1% Triton X-100 in sonication buffer, and resuspended in 6M GdnHCl and sonication buffer. Protein was refolded by removal of GdnHCl by dilution to 1M GdnHCl, followed by exhaustive dialysis against 50 mM Tris pH 8.8. Pertactin was purified using an anion exchange column with a final purity of >95% estimated by Coomassie staining. Cysteine mutants were purified using the same procedure, using 50 mM Tris pH 7.0.

### Fluorescence spectroscopy

Fluorescence emission spectra were obtained using a QM-6 T-format fluorimeter (PTI, Birmingham, NJ). Samples were measured at 20°C, using 0.5  $\mu$ M pertactin in 50 mM Tris pH 8.8 in a 1-cm cuvette. Tryptophan emission data were collected under the same conditions as described previously.<sup>47</sup> Fluorescein spectra were collected with an excitation slit width of 3 nm and an emission slit width of 5 nm. Each spectrum represents an average of three spectra collected with a step size of 1 nm and 1 s integration time.

### Kinetic measurements of tryptophan fluorescence

Manually mixed kinetics was also measured using the QM-6 fluorimeter. Tryptophan fluorescence kinetics was collected using an excitation wavelength of 280 nm and emission wavelengths of 335 and 350 nm. The excitation slit width was set at 0.8 nm and emission slit width at 5 nm. For refolding kinetics from the unfolded state, pertactin was first unfolded in 4M GdnHCl in 50 mM Tris pH 8.8 for >1 h; refolding was initiated by diluting samples 10-fold into 50 mM Tris pH 8.8 resulting in 0.5  $\mu$ M pertactin in 0.4M GdnHCl and 50 mM Tris pH 8.8. To minimize photobleaching over long measurement times, a fresh aliquot was used to determine the fluorescence emission intensity for every time point. For fluorescein, an excitation wavelength of 492 nm and emission at 515 and 530 nm was used, with excitation and emission slit widths of 5 nm. Refolding kinetics from the PFS were measured by diluting pertactin pre-equilibrated in 1.0M GdnHCl, 50 mM Tris pH 8.8 (>12 h) into 50 mM Tris pH 8.8 leading to a final pertactin concentration of 0.5  $\mu$ M in 4M GdnHCl. Unfolding kinetics were measured for 0.5  $\mu$ M pertactin in 6M GdnHCl, 50 mM Tris pH 8.8 by mixing pertactin pre-equilibrated in 1.2M GdnHCl (>12 h) with 8M GdnHCl, both in 50 mM Tris pH 8.8.

For fast kinetic experiments, a Biologic SFM-400 stopped flow apparatus was connected to the QM-6 fluorimeter via fiber optic cables. In all cases, 6  $\mu$ M protein

samples in 50 mM Tris pH 8.8 (final concentration) were used, with 5 nm excitation and emission slits. Tryptophans were excited at 280 nm and emission intensity measured at 335 nm. The dead time of the instrument was 20 ms.

### ANS binding assay

Pertactin was unfolded in 4M GdnHCl, 50 mM Tris pH 8.8 for at least 1 h. Refolding was induced by 10-fold dilution into 50 mM Tris pH 8.8, resulting in a final protein concentration of 0.5  $\mu$ M in 0.4M GdnHCl. After various refolding times, 900  $\mu$ L of the refolding protein was added to 100  $\mu$ L of 10  $\mu$ M, 100  $\mu$ M or 1 mM ANS in 50 mM Tris pH 8.8. ANS fluorescence emission spectra were collected immediately after addition to the ANS solution. ANS emission was measured from 400 to 600 nm with an excitation wavelength of 380 nm. Excitation and emission slit width were set at 5 nm and data were collected with an integration time of 0.1 s/nm. Three spectra were averaged for each sample and three independent samples were collected for each time point.

BCA (Sigma) was prepared as a 10 mg/mL stock in 50 mM Tris pH 8.8. Immediately before the experiment, this stock was diluted to a final concentration of 0.5  $\mu$ M in 50 mM Tris pH 8.8 or 3.6, and supplemented with 10  $\mu$ M ANS. ANS fluorescence was measured under conditions identical to those used for pertactin samples.

### Circular dichroism spectroscopy

CD spectra and kinetic traces were collected with an Aviv 62D2 spectropolarimeter. Data were obtained with a bandwidth of 1 nm and an integration time of 1 s. A 2-mm cuvette was used with 5  $\mu$ M protein samples. For kinetic analysis, the change in molar ellipticity was determined at 218 nm. Refolding samples were prepared as described earlier.

### Protein labeling

Pertactin cysteine mutants were labeled with fluorescein following the general guidelines provided by the manufacturer (Molecular Probes). Briefly, each labeling reaction contained 200  $\mu$ M pertactin in 50 mM Tris pH 7.0, 5 mM TCEP, and 1 mM fluorescein. Reactions were incubated at room temperature for at least 12 h, and residual free fluorophores were removed using a S75 size exclusion column. Labeling efficiency was 30% for S117C and 70% for T490C, calculated using the absorbance of the fluorophore at 492 nm compared with the measured protein concentration at 280 nm.

### Proline isomerization experiments

Pertactin samples were unfolded in 4M GdnHCl, 50 mM Tris pH 8.8 for 10 s followed by immediate dilution

to 0.4M GdnHCl and 0.5  $\mu$ M pertactin. For enzymatic analysis, either FKBP 506 or cyclophilin A were added to the refolding buffer in a protein to enzyme ratio of 10:1 (w/w). Refolding kinetics was measured following the change in the ratio of pertactin tryptophan fluorescence emission between 335 and 350 nm. Each point represents the average of at least three experiments. In every case, as explained earlier, each measurement represents a fresh aliquot from a larger stock solution.

### Proteolytic digestion

Pertactin was equilibrated in 1.5 or 4M GdnHCl in 50 mM Tris pH 8.8, followed by dilution to 0.5M GdnHCl in 50 mM Tris pH 8.8 and 7.5 mM CaCl<sub>2</sub>. After various incubation times in 0.5M GdnHCl, pertactin was digested with proteinase K for 50 min. The protease:protein weight ratio was 1:500. Digestion was stopped by boiling for 10 min. Samples were analyzed by SDS-PAGE and visualized with silver staining.

## REFERENCES

- Wetzel R. For protein misassembly, it's the "I" decade. *Cell* 1996;86:699–702.
- Anna LM. How does a knotted protein fold? *FEBS J* 2009;276:365–375.
- Grantcharova V, Alm EJ, Baker D, Horwich AL. Mechanisms of protein folding. *Curr Opin Struct Biol* 2001;11:70–82.
- Fersht AR. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat Rev Mol Cell Biol* 2008;9:650–654.
- Religa TL, Markson JS, Mayor U, Freund SM, Fersht AR. Solution structure of a protein denatured state and folding intermediate. *Nature* 2005;437:1053–1056.
- Maity H, Maity M, Krishna MM, Mayne L, Englander SW. Protein folding: the stepwise assembly of foldon units. *Proc Natl Acad Sci USA* 2005;102:4741–4746.
- Clark PL, Liu ZP, Rizo J, Gierasch LM. Cavity formation before stable hydrogen bonding in the folding of a beta-clam protein. *Nat Struct Biol* 1997;4:883–886.
- Bieri O, Kiefhaber T. Origin of apparent fast and non-exponential kinetics of lysozyme folding measured in pulsed hydrogen exchange experiments. *J Mol Biol* 2001;310:919–935.
- Korzhev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM, Kay LE. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 2004;430:586–590.
- Friel CT, Smith DA, Vendruscolo M, Gsponer J, Radford SE. The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat Struct Mol Biol* 2009;16:318–324.
- Gromiha MM, Thangakani AM, Selvaraj S. FOLD-RATE: prediction of protein folding rates from amino acid sequence. *Nucleic Acids Res* 2006;34 (web server issue):W70–W74.
- Weikl TR, Dill KA. Folding rates and low-entropy-loss routes of two-state proteins. *J Mol Biol* 2003;329:585–598.
- Weikl TR, Palassini M, Dill KA. Cooperativity in two-state protein folding kinetics. *Protein Sci* 2004;13:822–829.
- Prabhu NP, Bhuyan AK. Prediction of folding rates of small proteins: empirical relations based on length, secondary structure content, residue type, and stability. *Biochemistry* 2006;45:3805–3812.
- Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV. Contact order revisited: influence of protein size on the folding rate. *Protein Sci* 2003;12:2057–2062.
- Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 1998;277:985–994.
- Roumestand C, Boyer M, Guignard L, Barthe P, Royer CA. Characterization of the folding and unfolding reactions of a small beta-barrel protein of novel topology, the MTCP1 oncogene product P13. *J Mol Biol* 2001;312:247–259.
- Brocchieri L, Karlin S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res* 2005;33:3390–3400.
- Rutkowska A, Kolinski A. Why do proteins divide into domains? Insights from lattice model simulations. *Biomacromolecules* 2007;8:3519–3524.
- Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Annu Rev Biophys* 2008;37:289–316.
- Clark PL. Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci* 2004;29:527–534.
- Maki K, Cheng H, Dolgikh DA, Roder H. Folding kinetics of staphylococcal nuclease studied by tryptophan engineering and rapid mixing methods. *J Mol Biol* 2007;368:244–255.
- Gosavi S, Whitford PC, Jennings PA, Onuchic JN. Extracting function from a beta-trefoil folding motif. *Proc Natl Acad Sci USA* 2008;105:10384–10389.
- Jones K, Wittung-Stafshede P. The largest protein observed to fold by two-state kinetic mechanism does not obey contact-order correlation. *J Am Chem Soc* 2003;125:9606–9607.
- Brooks CL III. Protein and peptide folding explored with molecular simulations. *Acc Chem Res* 2002;35:447–454.
- Estape D, Rinas U. Folding kinetics of the all-beta-sheet protein human basic fibroblast growth factor, a structural homolog of interleukin-1beta. *J Biol Chem* 1999;274:34083–34088.
- Capraro DT1, Roy M, Onuchic JN, Jennings PA. Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? *Proc Natl Acad Sci USA* 2008;105:14844–14848.
- Varley P, Gronenborn AM, Christensen H, Wingfield PT, Pain RH, Clore GM. Kinetics of folding of the all-beta sheet protein interleukin-1 beta. *Science* 1993;260:1110–1113.
- Samuel D, Kumar TK, Balamurugan K, Lin WY, Chin DH, Yu C. Structural events during the refolding of an all beta-sheet protein. *J Biol Chem* 2001;276:4134–4141.
- Sivaraman T, Kumar TK, Chang DK, Lin WY, Yu C. Events in the kinetic folding pathway of a small, all beta-sheet protein. *J Biol Chem* 1998;273:10181–10189.
- Mizuguchi M, Kroon GJ, Wright PE, Dyson HJ. Folding of a beta-sheet protein monitored by real-time NMR spectroscopy. *J Mol Biol* 2003;328:1161–1171.
- Li H, Frieden C. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and 19F NMR. *Proc Natl Acad Sci USA* 2007;104:11993–11998.
- Barrick D, Ferreira DU, Komives EA. Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Curr Opin Struct Biol* 2008;18:27–34.
- Kloss E, Courtemanche N, Barrick D. Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* 2008;469:83–99.
- Mello CC, Bradley CM, Tripp KW, Barrick D. Experimental characterization of the folding kinetics of the notch ankyrin domain. *J Mol Biol* 2005;352:266–281.
- Jenkins J, Pickersgill R. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol* 2001;77:111–175.
- Yoder MD, Jurnak F. Protein motifs. III. The parallel beta helix and other coiled folds. *FASEB J* 1995;9:335–342.
- Simkovsky R, King J. An elongated spine of buried core residues necessary for in vivo folding of the parallel beta-helix of P22 tail-spike adhesin. *Proc Natl Acad Sci USA* 2006;103:3575–3580.

39. Schuler B, Seckler R. P22 tailspike folding mutants revisited: effects on the thermodynamic stability of the isolated beta-helix domain. *J Mol Biol* 1998;281:227–234.
40. Schuler B, Furst F, Osterroth F, Steinbacher S, Huber R, Seckler R. Plasticity and steric strain in a parallel beta-helix: rational mutations in the P22 tailspike protein. *Proteins* 2000;39:89–101.
41. Benton CB, King J, Clark PL. Characterization of the protrimer intermediate in the folding pathway of the interdigitated beta-helix tailspike protein. *Biochemistry* 2002;41:5093–5103.
42. Kamen DE, Woody RW. A partially folded intermediate conformation is induced in pectate lyase C by the addition of 8-anilino-1-naphthalenesulfonate (ANS). *Protein Sci* 2001;10:2123–2130.
43. Kamen DE, Woody RW. Folding kinetics of the protein pectate lyase C reveal fast-forming intermediates and slow proline isomerization. *Biochemistry* 2002;41:4713–4723.
44. Fuchs A, Seiderer C, Seckler R. In vitro folding pathway of phage P22 tailspike protein. *Biochemistry* 1991;30:6598–6604.
45. Carbonetti NH. Immunomodulation in the pathogenesis of *Bordetella pertussis* infection and disease. *Curr Opin Pharmacol* 2007;7:272–278.
46. Emsley P, Charles IG, Fairweather NF, Isaacs NW. Structure of *Bordetella pertussis* virulence factor P. 69 pertactin. *Nature* 1996;381:90–92.
47. Junker M, Schuster CC, McDonnell AV, Sorg KA, Finn MC, Berger B, Clark PL. Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci USA* 2006;103:4918–4923.
48. Galzitskaya OV, Garbuzynskiy SO, Ivankov DN, Finkelstein AV. Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. *Proteins* 2003;51:162–166.
49. Agashe VR, Shastry MC, Udgaonkar JB. Initial hydrophobic collapse in the folding of barstar. *Nature* 1995;377:754–757.
50. Engelhard M, Evans PA. Kinetics of interaction of partially folded proteins with a hydrophobic dye: evidence that molten globule character is maximal in early folding intermediates. *Protein Sci* 1995;4:1553–1562.
51. Schmid FX. Fast-folding and slow-folding forms of unfolded proteins. *Methods Enzymol* 1986;131:70–82.
52. Schmid FX. Protein folding. Prolyl isomerases join the fold. *Curr Biol* 1995;5:993–994.
53. Kaya H, Chan HS. Origins of chevron rollovers in non-two-state protein folding kinetics. *Phys Rev Lett* 2003;90(25 Pt 1):258104.
54. Khorasanizadeh S, Peters ID, Roder H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* 1996;3:193–205.
55. Sanchez IE, Kiefhaber T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* 2003;325:367–376.
56. Nishimura C, Dyson HJ, Wright PE. The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate. *J Mol Biol* 2002;322:483–489.
57. Lejeune A, Pain RH, Charlier P, Frere JM, Matagne A. TEM-1 beta-lactamase folds in a nonhierarchical manner with transient non-native interactions involving the C-terminal region. *Biochemistry* 2008;47:1186–1193.
58. Fisher KE, Ruan B, Alexander PA, Wang L, Bryan PN. Mechanism of the kinetically-controlled folding reaction of subtilisin. *Biochemistry* 2007;46:640–651.
59. Eckert B, Martin A, Balbach J, Schmid FX. Prolyl isomerization as a molecular timer in phage infection. *Nat Struct Mol Biol* 2005;12:619–623.
60. Kelch BA, Eagen KP, Erciyas FP, Humphris EL, Thomason AR, Mitsui S, Agard DA. Structural and mechanistic exploration of acid resistance: kinetic stability facilitates evolution of extremophilic behavior. *J Mol Biol* 2007;368:870–883.
61. Ratner V, Amir D, Kahana E, Haas E. Fast collapse but slow formation of secondary structure elements in the refolding transition of *E. coli* adenylate kinase. *J Mol Biol* 2005;352:683–699.
62. Kimura T, Uzawa T, Ishimori K, Morishima I, Takahashi S, Konno T, Akiyama S, Fujisawa T. Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. *Proc Natl Acad Sci USA* 2005;102:2748–2753.
63. Ferreira DU, Cho SS, Komives EA, Wolynes PG. The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. *J Mol Biol* 2005;354:679–692.
64. Tang KS, Fersht AR, Itzhaki LS. Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure* 2003;11:67–73.
65. Courtemanche N, Barrick D. The leucine-rich repeat domain of internalin B folds along a polarized N-terminal pathway. *Structure* 2008;16:705–714.
66. Junker M, Besingi RN, Clark PL. Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Mol Microbiol* 2009;71:1323–1332.
67. Renn JP, Clark PL. A conserved stable core structure in the passenger domain beta-helix of autotransporter virulence proteins. *Biopolymers* 2008;89:420–427.
68. Chiti F, Dobson CM. Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 2009;5:15–22.
69. Jahn TR, Radford SE. The yin and yang of protein folding. *FEBS J* 2005;272:5962–5970.
70. Shivaprasad S, Wetzel R. Scanning cysteine mutagenesis analysis of Abeta-(1–40) amyloid fibrils. *J Biol Chem* 2006;281:993–1000.
71. Wasmer C, Lange A, Van Melckebeke H, Siemer AB, Riek R, Meier BH. Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science* 2008;319:1523–1526.
72. Perutz MF, Finch JT, Berriman J, Lesk A. Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci USA* 2002;99:5591–5595.
73. Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci USA* 2002;99:2754–2759.
74. Calloni G, Lendel C, Campioni S, Giannini S, Gliozzi A, Relini A, Vendruscolo M, Dobson CM, Salvatella X, Chiti F. Structure and dynamics of a partially folded protein are decoupled from its mechanism of aggregation. *J Am Chem Soc* 2008;130:13040–13050.