Single-Molecule Experiments Reveal the Flexibility of a Per-ARNT-Sim Domain and the Kinetic Partitioning in the Unfolding Pathway under Force

Xiang Gao, Meng Qin, Puguang Yin, Junyi Liang, Jun Wang, Yi Cao,* and Wei Wang*

National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing, People’s Republic of China

ABSTRACT Per-ARNT-Sim (PAS) domains serve as versatile binding motifs in many signal-transduction proteins and are able to respond to a wide spectrum of chemical or physical signals. Despite their diverse functions, PAS domains share a conserved structure. It has been suggested that the structure of PAS domains is flexible and thus adaptable to many binding partners. However, direct measurement of the flexibility of PAS domains has not yet been provided. Here, we quantitatively measure the mechanical unfolding of a PAS domain, ARNT PAS-B, using single-molecule atomic force microscopy. Our force spectroscopy results indicate that the structure of ARNT PAS-B can be unraveled under mechanical forces as low as ~30 pN due to its broad potential well for the mechanical unfolding transition of ~2 nm. This allows the PAS-B domain to extend by up to 75% of its resting end-to-end distance without unfolding. Moreover, we found that the ARNT PAS-B domain unfolds in two distinct pathways via a kinetic partitioning mechanism. Sixty-seven percent of ARNT PAS-B unfolds through a simple two-state pathway, whereas the other 33% unfolds with a well-defined intermediate state in which the C-terminal β-hairpin is detached. We propose that the structural flexibility and force-induced partial unfolding of PAS-B domains may provide a unique mechanism for them to recruit diverse binding partners and lower the free-energy barrier for the formation of the binding interface.

INTRODUCTION

Per-ARNT-Sim (PAS) domains are ubiquitous in signal-transduction proteins and serve as versatile interaction hubs (1–4). They sense and respond to various chemical and physical stimuli and regulate the activities of many downstream effector domains. In contrast to their functional diversity, PAS domains have a conserved structure composed of a five-stranded antiparallel β-sheet and several α-helices (5–9) (Fig. S1 in the Supporting Material). The β-sheet is essential for the function of PAS domains, as it can bind to myriad protein targets via the solvent-exposed surface and modulate the binding through conformational changes triggered by environmental signals or cofactor occupancy at the other part of the domain (8,10–16). Because of their universal binding capability, it is suggested that the β-sheet of PAS domains is flexible and adaptable to different binding targets (1). However, a direct and quantitative measurement of the flexibility of the PAS domain is lacking. Here, we applied atomic-force-microscope (AFM)-based single-molecule force spectroscopy to directly probe the conformational change of a PAS domain, ARNT PAS-B, upon elongation force.

Single-molecule force spectroscopy has evolved into a powerful tool to study the conformational dynamics of proteins, especially protein folding/unfolding dynamics (17–22). It allows force to be applied to a protein to trigger the conformational change and quantitatively measure the free-energy landscape along the force direction (23–25). Single-molecule AFM has revealed rich information on the conformational dynamics of proteins, such as the existence of mechanical unfolding intermediate states (26–33), the presence of parallel unfolding pathways (34–38), the structural changes upon ligand binding or protein-protein interactions (39–45), and the capture of rare misfolding events (46,47), most of which are difficult or impossible to study at the ensemble level. Previous single-molecule AFM studies on photoreactive yellow protein (48,49), a representative PAS domain, revealed that the PAS domain is extended by ~3 nm and mechanically destabilized by ~30% in the light-activated state, indicating partial unfolding of photoreactive yellow protein upon photoactivation (48). However, the conformational dynamics of the native PAS domain is not fully addressed. Moreover, in this single-molecule AFM experiment (48), the PAS domain was not stretched along the N- and C-termini but along the axis defined by two engineered cysteine residues. Such force direction may not be physiologically relevant to the signal transduction processes mediated by PAS domains.

Here, we use single-molecule AFM to study conformational dynamics of the ARNT PAS-B domain by applying force along the N- and C-termini, following its native connection in proteins. ARNT PAS-B domain is primarily involved in the binding of HIF-2α for the formation of a dimeric transcriptional regulator complex through an antiparallel β-sheet-β-sheet packing (8). The two terminal β-strands are part of the β-sheet and are proposed to be responsible for the flexibility of the β-sheet (1,7). Our single-molecule AFM results indicate that despite its high thermodynamic stability, the structure of ARNT PAS-B can be disrupted by mechanical forces as low as ~30 pN, which is lower than the force required to disrupt many proteins studied so far. Detailed kinetic studies reveal that...
the low unfolding forces of the ARNT PAS-B domain are due to its long unfolding distance (the distance between the transition state and the native state along the force direction) of ~2 nm (50). Such a long unfolding distance allows ARNT PAS-B to adjust its conformation upon hydrodynamic motion for the binding of different partners without unfolding. Moreover, we find that the ARNT PAS-B domain can unfold through two distinct pathways via a kinetic partitioning mechanism (51). Sixty-seven percent of ARNT PAS-B unfolds in a two-state fashion, whereas 33% of the unfolding events involve a well-defined intermediate state in which the C-terminal β-hairpin is unstructured. We propose that the structural flexibility and the kinetic partitioning in the unfolding pathway may be important for PAS-B domains to interact with multiple binding partners with low sequence identity.

MATERIALS AND METHOD

Protein engineering

Plasmid encoding wild-type ARNT PAS-B was synthesized by Genscript (Nanjing, China). The (PAS-B)₈ polyprotein gene was constructed by iteratively cloning monomer into monomer, dimer into dimer, and tetramer into tetramer using a previously described method based on the identical sticky ends generated by the BamHI and BglII restriction enzymes (New England Biolabs, Ipswich, MA) (52). The polyprotein was expressed in BL21 and purified by Ni-NTA resin (GE Healthcare Bio-Sciences, Little Chalfont, UK). Then the polyprotein was kept at 4°C in solutions of 50 mM Tris (pH 7.5), 15 mM imidazole, 47 mM NaCl, and 5 mM dithiothreitol.

Single-molecule AFM experiments

Single-molecule AFM experiments were carried out on a JPK AFM (ForceRobot 300, JPK Instruments, Berlin, Germany). In all force spectroscopy experiments, we deposited polyprotein onto a mica plate and allowed it to adsorb for 10-20 min. Then, the fluid chamber was filled in ~1 mL buffer containing 50 mM Tris, 17 mM NaCl, and 5 mM dithiothreitol (pH 7.5). The AFM experiments were carried out after allowing the mixture to equilibrate for ~30 min. Cantilevers (Biolever, Olympus, Melville, NY) with a typical spring constant of 6 pN nm⁻¹ were used for all experiments and calibrated using the equipartition theorem before each experiment. The force-extension traces were recorded using JPK software and analyzed using a home-written protocol in Igor 6.0 (WaveMetrics, Lake Oswego, OR).

Monte Carlo simulation

In the Monte Carlo simulation, the mechanical unfolding of the PAS-B domain is modeled as a bifurcation of a three-state and a two-state process. The unfolding rate constants of both native state and intermediate state are dependent on force. The unfolding rate for the native state in both pathways can be described as:

\[ k_{\text{un}} = k_{\text{un}0} \times \exp \left( \frac{F \Delta x_n}{k_B T} \right) \]

where \( k_B \) is the Boltzmann constant, \( T \) is the temperature in Kelvin, \( k_{\text{un}} \) is the unfolding rate constant at a stretching force of \( F \), \( k_{\text{un}0} \) corresponds to the unfolding rate constant at zero force, and \( \Delta x_n \) is the distance between the native state and the transition state (52,53). For the three-state pathway, the \( k_{\text{un}} \) is set at two times that for the two-state pathway to reflect the different probability of these two pathways based on Boltzmann distribution. Thus, when a PAS domain unfolds, it has a 33% probability of unfolding via an intermediate state and a 67% probability of proceeding directly to the unfolded state. In the three-state pathway, the unfolding of the intermediate state follows a similar equation:

\[ k_i = k_{i0} \times \exp \left( \frac{F \Delta x_i}{k_B T} \right) \]

where \( k_i \) is the unfolding rate constant and \( k_{i0} \) is the spontaneous unfolding rate at zero force for the intermediate state. \( \Delta x_i \) is the distance between the intermediate state and the second unfolding transition state.

In the Monte Carlo simulation, we assume that (PAS-B)₈ is elongated in discrete time steps of \( \Delta t \) (typically set to \( 10^{-4} \) s). The contour length of the protein can be calculated from the number of folded states, \( N_f \), the number of intermediate states, \( N_i \), the number of unfolding states, \( N_u \), and the spacer lengths, \( L \).

\[ L_c = L_g + N_f \times L_f + N_i \times L_i + N_u \times L_u \]

where \( L_g \) is the contour length of the PAS domain, and \( L_f, L_i, \) and \( L_u \) are the lengths of a PAS domain in the folded state, the intermediate state, and the unfolding state, respectively. The extension of the polyprotein can be calculated as \( v \) multiplied by \( t \), where \( v \) is the pulling speed and \( t \) is time. The stretching force on poly-PAS can be calculated by the wormlike chain (WLC) model

\[ F = k_BT \frac{1}{p} \left( \frac{1}{4} \times \frac{1}{1 - \frac{v \times t}{L_c} - \frac{1}{4} + \frac{v \times t}{L_c}} \right) \]

where \( p \) is the persistence length and the other parameters are defined above.

The unfolding probability, \( P \), can be calculated as \( k \times \Delta t < < 0 \), where \( k \) is \( k_{\text{un}} \) or \( k_i \), depending on the corresponding unfolding event. Then \( P \) is compared with a random number between 0 and 1 to decide whether the protein unfolds and by which pathway it unfolds. In the two-state pathway, this leads to \( N_i = N_f - 1 \) and \( N_f = N_f + 1 \). In the three-state pathway, this leads to \( N_f = N_f - 1 \) and \( N_i = 1 \). Then, after the unfolding of the intermediate state, \( N_u = N_u + 1 \) and \( N_i = 0 \). The procedure repeats till both \( N_f \) and \( N_i \) reach zero.

RESULTS

Forced unfolding of ARNT PAS-B is characterized by a long unfolding distance to the transition state

To study the mechanical unfolding of PAS-B domain by using AFM-based single-molecule force spectroscopy, we followed a standard polyprotein-based approach (Fig. 1 a) (52). We engineered a polyprotein, (PAS-B)₈, that is made of eight tandem repeats of PAS-B. Stretching (PAS-B)₈ at a constant pulling speed of 400 nm s⁻¹ yields sawtoothlike unfolding patterns (Fig. 1 b) in which each individual peak, except the last one, corresponds to the unfolding of a PAS-B domain in the polyprotein. The last peak corresponds to the detachment of the polyprotein from either the cantilever tip or the substrate. Because the polyprotein is picked up by the cantilever tip from the substrate surface randomly along its
contour, each trace has various numbers of unfolding events from one to eight. Fitting consecutive unfolding events using the WLC model for polymer elasticity yields a contour-length increment ($\Delta L_c$) of ~39 nm and a persistence length ($p$) of 0.5–1.0 nm. Since the PAS-B domain is comprised of 115 amino acids and the distance between the N- and C- termini of a folded PAS-B domain is 2.4 nm (PDB id 1X0O), stretching a PAS-B domain from the folded state to the unfolded state should give rise to a contour-length increment of 39 nm (0.365 nm/aa × 114 aa = 2.4 nm), in good agreement with our experimentally measured value. Therefore, it is certain that the force peaks shown in Fig. 1b are indeed from the unfolding of individual PAS-B domains. The amplitude of these unfolding force peaks is 32.6 ± 6.3 pN (mean ± SD, n = 373), indicating that the unfolding force of PAS-B domains is rather low. This value is much lower than those for many proteins with $\beta$-sheet structures and even some $\alpha$-helical proteins (17,54) and is comparable to the force that can be generated by biological motors (55). The unfolding-force histogram of PAS-B is shown in Fig. 1c. Besides the low unfolding force, another interesting feature for the unfolding force distribution is the narrow distribution. The standard deviation of the unfolding force is as small as 6.3 pN, which is much smaller than that for protein domains containing $\beta$-sheets, such as fibronectin, immunoglobulin, and $\beta$-grasp domains. The distribution of the unfolding force is directly related to the unfolding distance ($\Delta x_u$), the distance between native state and unfolding transition state along the force direction (the length gain between the two stretching points of a protein at its mechanical unfolding transition state (Fig. S2)). The narrower the distribution, the longer is the unfolding distance. Using a standard Monte Carlo simulation procedure, we estimate the unfolding distance of PAS-B to be 2.0 nm and the spontaneous unfolding rate of PAS-B at zero force to be $3.0 \times 10^{-5}$ s$^{-1}$. Generally, the unfolding error using the Monte Carlo method is less than one order of magnitude for the spontaneous unfolding rate constant and <0.3 nm for unfolding distance (Fig. S3). It is worthy of note that such a long unfolding distance of PAS-B is rare for $\beta$-sheet proteins, as the unfolding distance for $\beta$-sheet proteins is typically in the range 0.1–0.5 nm (20,54). Our results indicate that the distance between the two ends of PAS-B can extend by as much as 2.0 nm before it is unfolded. Given that the measured distance between the N- and C-termini based on the NMR structure is only 2.4 nm, such an extension accounts for 75% of its original length, indicating the high flexibility of PAS-B domain upon stretching. Such a long unfolding distance is also confirmed by pulling-speed-dependent experiments. The average unfolding force of PAS-B increases from 29 pN at 50 nm s$^{-1}$ to 35 pN at 3600 nm s$^{-1}$ (Fig. 1d). Using the same unfolding distance of 2.0 nm and unfolding rate constant at zero force of $3.0 \times 10^{-5}$ s$^{-1}$, we are able to reproduce such pulling-speed dependency using Monte Carlo simulations. Such a weak dependency of unfolding force on pulling speed provides additional evidence for the long unfolding distance of PAS-B.

Since the two ends of PAS-B need to extend as long as 2.0 nm to reach the mechanical unfolding transition state,
we would infer that the two terminal $\beta$-strands are partially separated under force before the entire protein is completely unfolded. A schematic drawing of the structure of the native state and transition state of PAS-B domain is shown in Fig. S2. Closely examining the NMR structures of PAS-B reveals that the first two hydrogen bonds between strand-A and strand-I (Thr$^{361.0}$-Asn$^{463.11}$ and Phe$^{363.1}$-Asn$^{461.0}$) are 2.52 $\text{Å}$ and 2.19 $\text{Å}$ in length, respectively, i.e., longer than typical hydrogen bonds in antiparallel $\beta$-strands ($\sim 2$ $\text{Å}$) (7). Such long hydrogen bonds indicate that they are prone to break by stretching. Hence, it is quite possible that the major barrier for mechanical unfolding of PAS-B is located in the middle part of the two terminal $\beta$-strands. However, to pinpoint the exact location of the mechanical unfolding barrier, detailed molecular dynamics simulations are required.

**ARNT PAS-B unfolds through two distinct pathways**

By close examination of the force-extension curves of PAS-B, we found that it unfolds through two distinct pathways (Fig. 2 a). The majority of the unfolding events show two-state unfolding behavior without any detectable mechanical unfolding intermediate states (Fig. 2 b). The contour-length increment for two-state unfolding events is $\sim 39.3 \pm 1.0$ nm (mean $\pm$ SD, $n = 249$), corresponding to full unfolding of PAS-B domains (Fig. 2 c). However, $\sim 33\%$ of the unfolding events show a transient mechanical unfolding intermediate state (Fig. 2 d). It is clear that the positions of the intermediate state are superimposable to each other, indicating that the intermediate state has a well-defined structure. WLC fitting measures the contour-length increment between the native state and the intermediate state ($\Delta L_{cN,I}$) at $10.5 \pm 2.1$ nm (mean $\pm$ SD, $n = 124$) and the contour-length increment between the intermediate state and the unfolded state ($\Delta L_{cI,U}$) at $28.7 \pm 2.5$ nm (mean $\pm$ SD, $n = 124$) (Fig. 2 e). Summing up $\Delta L_{cN,I}$ and $\Delta L_{cI,U}$ yields the total contour-length increment for mechanical unfolding of $39.2 \pm 1.2$ nm for the three-state unfolding pathway, similar to that for the two-state unfolding scenario shown in Fig. 2 c. This indicates that the three-state unfolding events are indeed from the unfolding of the PAS-B domain.

**The mechanical unfolding intermediate state is transient**

Subsequently, we characterized the mechanical unfolding intermediate state. As the position of the intermediate state in the force-extension curves overlaps with the cantilever relaxation phase after the unfolding of the native PAS-B structure, we plotted force against time to more precisely capture the unfolding kinetics of the mechanical unfolding intermediate state and to avoid the cantilever overdamping effect (Fig. 3 a). The unfolding forces for the intermediate states are $26.8 \pm 6.6$ pN (mean $\pm$ SD), which is slightly lower than the unfolding forces of the native state (Fig. 3 b). The average lifetime for the intermediate state is only $\sim 5$ ms. Within such a short timescale, the unfolding of the intermediate state can be considered quasistatic, and the force is approximated as constant in the unfolding process (Fig. 3 a). Therefore, the unfolding kinetics for the intermediate state can be extracted from the dwell-time distribution of the intermediate state in Fig. 3 c. Exponential fitting of the distribution yields an unfolding rate constant of $192 \pm 13$ s$^{-1}$. Using this unfolding rate constant as a constraint, we extracted the unfolding distance of 0.15 nm for the intermediate state and the spontaneous unfolding
Mechanical Unfolding of a PAS Domain

rate constant at zero force of 80 s\(^{-1}\) using the Monte Carlo simulation (Fig. S4). The error for \(k_0\) is less than a half order of magnitude and that for \(\Delta x_u\) is <0.15 nm (Fig. S4). The unfolding forces of the intermediate state at different pulling speeds are also shown in Fig. S4. It is worth noting that the unfolding distance of the intermediate state is much shorter than that of the native state, suggesting that the intermediate state is prone to disruption upon stretching. Moreover, the fast spontaneous unfolding rate of 80 s\(^{-1}\) indicates that the lifetime of the mechanical unfolding intermediate state is only -0.0125 s at zero force. Therefore, this mechanical unfolding intermediate state of PAS-B domain is transient and cannot survive for long even without a stretching force.

The structure of the mechanical unfolding intermediate state

Contour-length increment can be used as a sensitive indicator to map the structure of the mechanical unfolding intermediate state. A \(\Delta L_{C_{N-I}}\) of 10.5 nm corresponds to the detachment of either 34 amino acids from the N-terminus or 37 amino acids from the C-terminus of PAS-B to reach the intermediate state (see Supporting Material). The result-

ing intermediate structures are referred as PAS-B(-N) and PAS-B(-C), respectively (Fig. S5). Another possible structure for the mechanical unfolding intermediate state is one in which both the N-terminal and C-terminal \(\beta\)-strands A and I are disrupted. However, this would lead to a contour-length increment of ~16 nm, longer than the measured \(\Delta L_{C_{N-I}}\) of 10.5 nm. Moreover, this process requires the breaking of interactions between the A-B hairpin and the H-I hairpin, the simultaneous occurrence of which is very unlikely, if not impossible. To unequivocally distinguish the structure of the intermediate state between PAS-B(-N) and PAS-B(-C), we resorted to a more elaborate approach, the glycine-insertion method (29,56). Five glycine residues were inserted at two different positions of the PAS-B domain. This leads to a change in the contour-length increment of the mechanical unfolding events. If the inserted glycine residues are within the mechanical unfolding intermediate state, it will increase \(\Delta L_{C_{N-U}}\) without affecting \(\Delta L_{C_{N-I}}\), and vice versa. We have engineered what to our knowledge are two new proteins with five glycine residues inserted into the turn between A and B \(\beta\)-strands after residue isoleucine 359 (PAS-B-I359/E360-5G) and the turn between H and I \(\beta\)-strands after residue tyrosine 450 (PAS-B-Y450/S451-5G) of PAS-B. However, PAS-B-I359/E360-5G cannot be expressed in Escherichia coli, probably because the glycine insertion affects its proper folding (57). PAS-B-Y450/S451-5G is thermodynamically stable, allowing us to make the polyprotein and use single-molecule AFM to measure the effect of glycine insertion on the contour-length increment (Fig. 4a). Glycine insertion increases the total contour-length increment of the PAS-B domain by ~1.7 nm (Fig. S6) but does not affect the occurrence of the intermediate states. Around 35% of the unfolding events follow the three-state unfolding pathway. Superpositioning the three-state unfolding events of PAS-B-Y450/S451-5G and wild-type PAS-B shows that the \(\Delta L_{C_{N-I}}\) increases by ~1.6 nm, whereas the \(\Delta L_{C_{N-U}}\) remains the same (Fig. 4b). This is further confirmed by the contour-length increment histogram in Fig. 4c. These results indicate that the location at which the five glycine residues are inserted is outside the structure of the intermediate state. Therefore, the structure of the mechanical unfolding intermediate state should be PAS-B(-C), the cyan parts shown in Fig. 4a, in which the H and I \(\beta\) strands are detached.

Is such an intermediate state thermodynamically stable if H and I \(\beta\)-strands are allowed to remain detached for a long time? To address this question, we engineer a truncated mutation of PAS-B corresponding to the PAS-B(-C) structure. However, the circular dichroism spectrum of PAS-B(-C) shows that it mainly adopts a random-coil structure instead of the expected \(\alpha + \beta\) structure, distinct from wild-type PAS-B (Fig. S7). This indicates that PAS-B is not thermodynamically stable in the absence of C-terminal \(\beta\)-strands H and I. Our data are consistent with the chemical denaturation measurement of PAS-B, in which it was found
that the folding and unfolding of PAS-B are both two-state processes without any detectable intermediate state (58).

The free-energy landscape for the mechanical unfolding of PAS-B

Based on our single-molecule AFM data, the mechanical-unfolding free-energy landscape of PAS-B is summarized in Fig. 5. Upon stretching, the N- and C-termini of PAS-B move against each other for ~2.0 nm, leading to the shearing of two terminal β-strands, A and I. The free-energy barrier for this transition is ~24 k_BT, as estimated from the unfolding rate constant (Supporting Material). After surmounting this barrier, unfolding can proceed in two distinct pathways. In the first pathway, the PAS-B domain completely unfolds without navigating any observable barriers. In the second pathway, the PAS-B domain reaches an intermediate state corresponding to a structure with β-strands H and I detached, which is ~10.5 nm away from the native state. Further unfolding requires it to overcome a barrier of 9.4 k_BT to reach the completely unfolded state. Such parallel unfolding pathways have been reported for a few proteins in mechanical unfolding experiments. The detailed molecular mechanism and physiological impact of the kinetic partitioning merit further detailed study.

DISCUSSION

The PAS-B domain is mechanically labile

Mechanical force is involved in many biological processes, including muscle contraction, cell division, cell movement, protein translocation, protein degradation, and signal transduction (55). For many functions, such as muscle contraction, high mechanical stability of proteins is essential (22). However, for signal transduction, high mechanical stability may not be required. PAS domains are important building units of signal-transduction proteins. Here, we show that PAS domains unfold at low mechanical force, distinct from many other proteins studied so far by single-molecule AFM. Such a low unfolding force may be beneficial to its function, because very little external work is required to trigger the conformational change of PAS domains. Such external force may be obtained from external stimuli such as light, chemical signal, or even relative movement of different domains. Therefore, we propose that the low mechanical stability may be common to many PAS domains and other signal-transduction proteins.

The low mechanical stability of the PAS domain originates from its unique three-dimensional structure. Theoretical and experimental efforts have demonstrated that proteins with shear topology, in which two terminal β-strands are antiparallel, are generally mechanically more stable than proteins without this structural motif (59–61). This is because in such a structural arrangement, the hydrogen bonds that bond the two force-bearing termini break concomitantly during the mechanical unfolding process. However, the two termini of PAS domain do not bear the shear topology. Instead, they are arranged in parallel and the hydrogen bonds between them break sequentially upon mechanical stretching. Typically, proteins with such a structural motif are mechanically labile and the ARNT PAS-B reported here is a representative example. Moreover, the major mechanical unfolding barrier resides in the middle of the N- and C-terminal β strands. Such a long unfolding distance not only gives rise to high flexibility of the native structure of PAS-B but also lowers the unfolding force.

Kinetic partitioning in the mechanical unfolding of ARNT PAS-B

It is now widely accepted that proteins can fold to their native structure via multiple pathways by a kinetic partitioning mechanism (51,62), despite the fact that experimental evidence supporting this mechanism is still limited (34–38). The mechanical unfolding of PAS-B provides new, to our
knowledge, evidence for this mechanism. We showed that the mechanical unfolding could proceed through two distinct pathways. In the first path, the C-terminus of the PAS domain unfolds first, leading to a marginally stable unfolding intermediate state in which the C-terminal β hairpin is detached from the rest of the domain. In the second path, the unfolding of the PAS domain does not show any detectable intermediate states. The two pathways account for 33% and 67%, respectively, of the unfolding events. Thus, the free-energy barrier for the three-state unfolding pathway is only ~0.7 k_BT higher than that for the two-state unfolding pathway. Based on the unfolding distance for the native state, such a difference in free-energy barrier will lead to <1 pN difference in the unfolding force, below the detection limit of our AFM setup. There are two possible mechanisms that may explain such kinetic partitioning behavior. For the first one, there are two subpopulations in the native-state ensemble with similar thermodynamic stability. The observed two distinct unfolding pathways result from the unfolding of these two subpopulations. For the second possible mechanism, unfolding can occur through two different pathways, in which the transition states are of similar stability but diverse structure.

Recent single-molecule AFM studies have provided increasing evidence for the kinetic partitioning mechanism of protein unfolding, distinct from the results obtained by other experimental methods (34–38). This is probably because of the unique unfolding/folding pathway probed in single-molecule AFM experiments: the pathway along the direction defined by the stretching force. The folding/unfolding of proteins may be cooperative in the globular free-energy landscape but not in every cross section. Single-molecule AFM provides unique opportunities to study different cross sections of the free-energy landscape of protein folding. Therefore, many surprising folding/unfolding behaviors can be found from these studies, enriching our understanding of the protein folding mechanism and making the free-energy landscape accessible in multiple dimensions experimentally.

Physiological consequences of the flexible native structure

It has been proposed that structural flexibility is important for ligand binding or protein-protein interactions, as it helps proteins to recruit binding partners and lowers the free-energy barrier for the formation of the binding interface. A recent fly-casting mechanism proposed by Wolynes and co-workers indicates that proteins can even partially unfold to facilitate the binding process (63–66). As the major function of PAS-B domains is to bind different ligands and proteins for the signal-transduction process, we suggest that PAS-B domains are able to utilize this unique mechanism for binding in their physiological conditions. Our results show that PAS-B can extend up to 75% of its original end-to-end distance in its native conformation and unfold at a marginal force of <30 pN. Furthermore, the unfolding can occur via a kinetic partitioning mechanism and reach a transient intermediate state in one of the unfolding pathways. Such high structural flexibility makes PAS-B domains ideal candidates for protein-binding motifs. Despite the high flexibility of PAS-B domains
upon mechanical stretching, they are thermodynamically extremely stable (>9 kcal/mol) (58). Thus, the flexibility of PAS-B domains is limited to particular directions. Since the N- and C-termini of PAS-B domains are among the most flexible parts in the structure and are actively involved in the binding process, we believe that such a structural and topological arrangement of PAS-B domains is of biological importance. However, it is still largely unknown whether PAS-B domains are indeed subject to tensile force in their physiological condition. Therefore, further experimental evidence is required to conclude that the flexibility and function of PAS-B domains are indeed regulated by force in vivo. Nonetheless, the studies presented here open the door for the study of anisotropic conformational flexibility of proteins and its biological consequences.

SUPPORTING MATERIAL

Protein sequence, estimation of the free energy barrier, calculation of the number of residues detached from protein at the intermediate state, and seven figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00383-9.

We thank Dr. D. Thirumalai for penetrating discussions and critical comments on the manuscript.

This work is supported by the National Natural Science Foundation of China under grants 11074115, 81221062, 10834002, 10904064, and 31170813, the Natural Science Foundation of Jiangsu Province under grant BK2009008, the program for New Century Excellent Talents in University, and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

REFERENCES


Mechanical Unfolding of a PAS Domain


Please cite this article in press as: Gao et al., Single-Molecule Experiments Reveal the Flexibility of a Per-ARNT-Sim Domain and the Kinetic Partitioning in the Unfolding Pathway under Force, Biophysical Journal (2012), doi:10.1016/j.bpj.2012.03.042.
Single-Molecule Experiments Reveal the Flexibility of a Per-ARNT-Sim Domain and the Kinetic Partitioning in the Unfolding Pathway under Force

Xiang Gao, Meng Qin, Puguang Yin, Junyi Liang, Jun Wang, Yi Cao, and Wei Wang

National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing, People’s Republic of China
Supporting information:

Protein sequence:

The protein sequences of monomer PAS-B (from Ref.(1)), (PAS-B)$_8$, PAS-B(-N) and PAS-B(-C) are shown below. Additional residues at the N and C terminus (in italic) correspond to cloning sites and the purification tags.

PAS-B (from Ref.(1)):

\[
\text{GAMDNVCQPTEFISRHIEGIFTFVDHRCAVGVYQPQELLGKNIVEFCHPEQDLRDSFQQVVLKLGQVLSVMFRSRKQQEWLWMRTSSFTFQNPYSDEIEYIICTNTNKNQQE} \\
\]

(PAS-B)$_8$:

\[
\text{MRGSHHHHHHGS}\text{GAMDNVCQPTEFISRHIEGIFTFVDHRCAVGVYQPQELLGKNIVEFCHPEQDLRDSFQQQVVLKLGQVLSVMFRSRKQQEWLWMRTSSFTFQNPYSDEIEYIICTNTNKNQQE} \\
\]

PAS-B(-N):

\[
\text{MRGSHHHHHHGS}\text{HRCVGVYQPQELLGKNIVEFCHPEQDLRDSFQQVVLKLGQVLSVMFRSRKQQEWLWMRTSSFTFQNPYSDEIEYIICTNTNKNQQE} \\
\]

PAS-B(-C):

\[
\text{MRGSHHHHHHGS}\text{GAMDNVCQPTEFISRHIEGIFTFVDHRCAVGVYQPQELLGKNIVEFCHPEQDLRDSFQQQVVLKLGQVLSVMFRSRKQQE} \\
\]

Estimation of the free energy barrier.

The relationship between the unfolding rate and the free energy barrier can be expressed as the following:

\[
k_0 = A \cdot \exp(-\Delta G / k_B T) \\
\]

(eq.1),

where, A is a pre-exponential factor, $k_B$ is the Boltzmann constant, T is the temperature in Kelvin, $k_0$ is the unfolding rate constant at zero force, and $\Delta G$ is the unfolding free energy barrier. The free energy barrier was estimated by assuming a pre-factor A of $10^6 \text{ s}^{-1}$.

Calculation of the number of residues detached from protein at the intermediate state.
The contour length increment (ΔLc) can be used to estimate the structure of the mechanical unfolding intermediate state. For the two-state unfolding scenario, ΔLc equals the contour length difference of a protein at the transition state, L_t, and the unfold state, L_u. However, because the unfolding distance for most proteins are less than 1 nm, L_t is generally considered to be the same as the contour length at the folded state, L_f. If an intermediate state is involved in the unfolding process, we defined L_{if}, L_{it} and L_{iu} as the contour length of the intermediate state at the folded, transition and unfolded states, respectively. Therefore, the contour length change for a protein to reach the intermediate state is as following:

\[ \Delta Lc_{N-I} = (L_u - L_{iu}) - (L_t - L_{it}) = (L_u - L_{iu}) - (L_f - L_{if}) - (\Delta x_{N-I} - \Delta x_{I-U}) \]  (eq. 2),

where, Δx_{N,I} and Δx_{I,U} are unfolding distance for the native state and the intermediate state, respectively.

Since each amino acid gives rise to a 0.365 nm in the contour length of unfolded peptide,

L_u - L_{iu} = N * 0.365 nm,

Where N is the number of amino acid that is detached in the intermediate state.

Based on the structure of the protein, L_f - L_{if} varies between 0 nm to 2.4 nm depending on N. And, Δx_{N,I} - Δx_{I,U} is 1.9 nm, as estimated from the single molecule AFM results. Therefore, the ΔLc of 10.5 nm from the native state to the intermediate state corresponds to the detachment of either 37 amino acids from the C terminus or 34 amino acids from the N terminus, according to the calculation using eq. 2. These two possible scenarios are shown in Fig.S5. a) and b).
Fig. S1 The topology diagram (a) and secondary structure-sequence alignment (b) of ANRT PAS-B. The diagram is drawn based on the NMR structure (PDB id:1X0O) using PDBsum.
Fig.S2 Schematic illustration of the native state and transition state structure of ARNT PAS-B domain. The end-to-end distance of PAS-B domain increases by 2.0 nm upon stretching at the unfolding transition state. The first two hydrogen bonds (dash lines) between the two terminal beta strands may be broken at the unfolding transition state.
Fig. S3 The Monte Carlo simulation of unfolding force distribution and pulling speed dependent experiment of ANRT PAS-B domain using various set of fitting parameters. Based on these simulations, the error for unfolding distance is less than 15% and the spontaneous unfolding rate constant is less than one order of magnitude.
Fig. S4 Monte Carlo simulation of the mechanical unfolding intermediate state of PAS-B. a) and b) Pulling speed dependent of the unfolding forces for the mechanical unfolding intermediate state. It is of noting that the unfolding force of the intermediate state is not sensitive to the pulling speed. The reason for this is due to the unique unfolding scheme of the intermediate state. In our experiment, the intermediate state always unfolds after the unfolding of the native state. However, after the unfolding of the native state, the force acting on the polyprotein was set to a value of ~30 pN and cannot be fully released before the unfolding of the intermediate state (See Fig. 2). c) and d) Unfolding force histogram of the intermediate state. (The same data set as Fig. 3b) e) and f) The lifetime distribution of the intermediate state. (The same data set as Fig. 3c) The events of lifetime less than 2.5 milliseconds are not included in the histogram because it is comparable with the temporal resolution of our instrument and some of the intermediate states may not be resolved due to the limit of such temporal resolution. The $k_0$ and $\Delta x$ are set as 80 s$^{-1}$ and 0.15 nm respectively in the Monte Carlo simulation, which can adequately reproduce the experimental data (black lines). Lines of other colors refer to the simulated results using different sets of parameters labeled in the figures.
Fig. S5 Two possible unfolding scenarios for the mechanical unfolding intermediate state. Based on the contour length increment from the native state to the intermediate state, the intermediate state corresponds to a structure with either N terminal 34 amino acids detached (highlighted in red) or C terminal 37 amino acids detached (highlighted in green). The N- and C- terminal intermediate states are referred as PAS-B(-N) and PAS-B(-C), respectively.
Fig. S6 The alignment of two representative unfolding traces of wt (PAS-B)$_8$ (red) and (PAS-B-Y450/S451-5G)$_8$ (black) by their first unfolding events. The contour length increment of each unfolding events increases ~1.7 nm ($\Delta x$). The total contour length gain by glycine insertion equals $n \times \Delta x$, where $n$ is the number of unfolding events.
Fig. S7 The intermediate state is thermodynamically unstable. CD spectrum of PAS-B(-C), the truncation mutant of PAS-B lacking the H and I β strands, adopts random coil structure, which is distinct from the CD spectrum of wild type PAS-B.
Reference: