Native topology determines force-induced unfolding pathways in globular proteins

D. K. Klimov, and D. Thirumalai

PNAS 2000;97;7254-7259
doi:10.1073/pnas.97.13.7254

This information is current as of March 2007.

Online Information & Services
High-resolution figures, a citation map, links to PubMed and Google Scholar, etc., can be found at:
www.pnas.org/cgi/content/full/97/13/7254

References
This article cites 22 articles, 11 of which you can access for free at:
www.pnas.org/cgi/content/full/97/13/7254#BIBL

This article has been cited by other articles:
www.pnas.org/cgi/content/full/97/13/7254#otherarticles

E-mail Alerts
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

Rights & Permissions
To reproduce this article in part (figures, tables) or in entirety, see:
www.pnas.org/misc/rightperm.shtml

Reprints
To order reprints, see:
www.pnas.org/misc/reprints.shtml

Notes:
Native topology determines force-induced unfolding pathways in globular proteins

D. K. Klimov* and D. Thirumalai*

Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742

Communicated by Bruce J. Berne, Columbia University, New York, NY, April 21, 2000 (received for review February 21, 2000)

Single-molecule manipulation techniques reveal that stretching unravels individually folded domains in the muscle protein titin and the extracellular matrix protein tenascin. These elastic proteins contain tandem repeats of folded domains with β-sandwich architecture. Herein, we propose by stretching two model sequences (S1 and S2) with four-stranded β-barrel topology that unfolding forces and pathways in folded domains can be predicted by using only the structure of the native state. Thermal refolding of S1 and S2 in the absence of force proceeds in an all-or-none fashion. In contrast, phase diagrams in the force-temperature (f,T) plane and steered Langevin dynamic studies of these sequences, which differ in the native registry of the strands, show that S1 unfolds in an all-or-none fashion, whereas unfolding of S2 occurs via an obligatory intermediate. Force-induced unfolding is determined by the native topology. After proving that the simulation results for S1 and S2 can be calculated by using native topology alone, we predict the order of unfolding events in Ig domain (Ig27) and two fibronectin III type domains (9FnIII and 10FnIII). The calculated unfolding pathways for these proteins, the location of the transition states, and the pulling speed dependence of the unfolding forces reflect the differences in the way the strands are arranged in the native states. We also predict the mechanisms of force-induced unfolding of the coiled-coil spectrin (a three-helix bundle protein) for all 20 structures deposited in the Protein Data Bank. Our approach suggests a natural way to measure the phase diagram in the (f,C) plane, where C is the concentration of denaturants.

Native topology determines force-induced unfolding pathways in globular proteins

Single-molecule manipulation techniques reveal that stretching unravels individually folded domains in the muscle protein titin and the extracellular matrix protein tenascin. These elastic proteins contain tandem repeats of folded domains with β-sandwich architecture. Herein, we propose by stretching two model sequences (S1 and S2) with four-stranded β-barrel topology that unfolding forces and pathways in folded domains can be predicted by using only the structure of the native state. Thermal refolding of S1 and S2 in the absence of force proceeds in an all-or-none fashion. In contrast, phase diagrams in the force-temperature (f,T) plane and steered Langevin dynamic studies of these sequences, which differ in the native registry of the strands, show that S1 unfolds in an all-or-none fashion, whereas unfolding of S2 occurs via an obligatory intermediate. Force-induced unfolding is determined by the native topology. After proving that the simulation results for S1 and S2 can be calculated by using native topology alone, we predict the order of unfolding events in Ig domain (Ig27) and two fibronectin III type domains (9FnIII and 10FnIII). The calculated unfolding pathways for these proteins, the location of the transition states, and the pulling speed dependence of the unfolding forces reflect the differences in the way the strands are arranged in the native states. We also predict the mechanisms of force-induced unfolding of the coiled-coil spectrin (a three-helix bundle protein) for all 20 structures deposited in the Protein Data Bank. Our approach suggests a natural way to measure the phase diagram in the (f,C) plane, where C is the concentration of denaturants.

Material and Methods

Off-Lattice Models for β-Barrel Proteins. We use coarse-grained continuum representations for proteins in which only the positions of Cα-carbons are retained (12). Sequences S1 and S2 are modeled as chains of 46 connected beads of three types, hydrophobic B, hydrophilic L, and neutral N. The potential energy includes the contributions from bond-length and bond-angle potentials, the dihedral angle potential VDih(φ), and nonbonded potentials (12). Vnonb(φ) is given by A(1 + cosφ) + B(1 + cos3φ). Along the β-strands trans-states are preferred, and accordingly, we chose A = B = 1.2ϵh, where ϵh = 1.25 kcal/mol is the strength of hydrophobic interactions. In the turn regions (i.e., in the vicinity of a cluster of N residues), A = 0 and B = 0.2ϵh. The interaction between a pair of hydrophobic residues is

\[ V_{\text{nb}}(r) = 4\lambda\epsilon_h \left[ \left( \frac{a}{r} \right)^{12} - \left( \frac{a}{r} \right)^{6} \right], \]

where \( \lambda \) is a random factor unique for each pair of B residues (12) and \( a = 3.8 \text{ Å} \) is the residue size. When the sequence is subject to external force \( f \), the total energy is written as \( E_{\text{tot}} = E_p - (f \cdot r) \), where \( r \) is the end-to-end vector.

Abbreviations: AFM, atomic force microscopy; SMD, steered molecular dynamics; SLD, steered Langevin dynamics; TS, transition state.

*To whom reprint requests should be addressed: E-mail: thirum@glue.umd.edu or klimov@glue.umd.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.
Results

Phase Diagrams for S1 and S2. We first studied the unfolding of off-lattice models of proteins. The phase diagrams for S1 and S2, both with β-sandwich native topology (Fig. 1a and c), are dramatically different (Fig. 1b and d). There is a small green region in Fig. 1b (S1) that represents states with intermediate values of the fraction of native contacts, \(<\xi_o>\). This small region suggests that force-induced unfolding would be an all-or-none process. In contrast, the phase diagram for S2 (Fig. 1d) shows a broad region of intermediate states along the edge of stability of the native state (red region). This broad region indicates that unfolding must occur via an intermediate. The conclusion that S1 must unfold in an all-or-none manner and that S2 must unfold through an intermediate, which naturally emerges from the equilibrium phase diagram, is verified explicitly by SLD simulations (see below).

The threshold force required to unfold the native state (determined from the edges of the red region) for S1 is larger than that required for S2 at all temperatures. At \(T = 300\) K, the equilibrium unfolding force for S1 is about 28 pN, whereas for S2, it is only 14 pN. We deduce that the native state of S1 is more stable than that of S2. Equilibrium folding and unfolding simulations in the absence of force show that the free energy of stability of S1 is larger than that of S2 by a factor of 2.5 (4.3 kcal/mol vs. 1.7 kcal/mol). Thus, the phase diagrams can be used to obtain the free energies of stability in the absence of force.

Stretching Under Constant Force. To probe the structural transitions on stretching, we performed SLD simulations by applying a constant force \(f\) along the z axis. For S1, the stretched state is reached in a single kinetic step, in which all strands unfold nearly simultaneously (Fig. 2a). There is, however, considerable heterogeneity in the time scales in which individual molecules (green curves in Fig. 2a) unfold. The strands unfold by an unzipping process in a cooperative manner without populating intermediates. A qualitatively different unfolding mechanism is found for S2 (Fig. 2b). The normalized extension, \(<\xi_o(t)>/L\) (\(L\) is the contour length), shows a biexponential behavior (blue line in Fig. 2b), indicating that S2 unravels by a two-step process. The unfolding kinetics is further revealed by examining the extension profiles for 20 individual trajectories (green curves in Fig. 2b).

On a short time scale, the polypeptide chain unfolds from the native state \(N\) to a stable on-pathway intermediate \(I\) (the green region in Fig. 1d represents the area of stability of \(I\)). The transition \(N \rightarrow I\) is reversible, i.e., on turning off the force, the chain spontaneously reaches the folded state without any evidence of hysteresis. Individual trajectories reach \(I\) (Fig. 2b Inset) by unzipping strand D. Fig. 2b shows that \(I\) is obligatory, i.e., all of the molecules reach \(S\) (stretched state) through \(I\). The \(I \rightarrow S\) transition also shows a broad distribution of unfolding times. Remarkably, individual transitions \(I \rightarrow S\) occur over a narrower time interval compared with the events leading to \(I\) (compare with Fig. 2b Inset). The unfolding of \(I\), in which three strands are pulled apart, occurs by a shearing-type motion, which is dynamically more cooperative (occurs extremely sharply) compared with the unzipping motion.

Pulling Speed Dependence. On application of force, the barrier between \(N\) and \(S\) is lowered such that the unfolding rates increase (21). Thus, the threshold force and the unfolding rate must depend on the pulling speed \(v\). We performed SLD simulations by fixing one of the termini and pulling the other at a constant \(v\) (Eq. 1). Force-extension profiles for \(S1\) show a single peak at \(f = 100\) pN at an extension of about 0.07. (Fig. 3a). We conclude that the TS for force-induced unfolding occurs close to the folded native state, which is reminiscent of the unfolding seen in Ig27 (1, 6). For S2, there are two peaks in the force-extension profile.
The one at $f = 58$ pN with an extension of about 0.1L represents the breakup of strand D (Fig. 1c), thus creating the stable intermediate I. The dominant peak ($f \approx 130$ pN and $\Delta z \approx 0.44L$) represents the unraveling of I. From the dependence of the fraction of native contacts $\langle Q(t, T) \rangle$ on $\Delta z$ (red curve in Fig. 3b), we conclude that, after strand D unfolds, there are no major structural changes until $\Delta z \approx 0.44L$, when the polypeptide chain dynamically reaches S. The simulations done at varying speeds result in the same mechanism as is suggested by constant force simulations. The unfolding mechanism is consistent with the expectations from the equilibrium phase diagrams (Fig. 1), a fact that underscores the importance of measuring the diagram of states.

As predicted by Evans and Ritchie (21), the average threshold unfolding force depends logarithmically on the pulling speed (Fig. 3c). The unfolding forces, namely, $f_{p,S1}$ (for S1) and $f_{p,S2}$ (for unfolding I in S2), may be written as $f_{p,a} = f_{eq,a} \ln(v/v_{min})$ ($a = S1$ or S2). Here, $f_{eq,a}$ is the equilibrium unfolding force (obtained from the phase diagram) and the minimum pulling speed $v_{min} = k_u \Delta z$ ($k_u$ is the unfolding rate at $f = 0$, and $\Delta z$ is the extension in going from N to S). We also show the scaled results for titin (1) and the recombinant tenascin segment TnFnIII (5). The parameterization proposed here (which is natural when the overall spring constant of the protein is comparable to the stiffness of cantilever) produces a scaling plot in which the results of simulations for model sequences and proteins lie essentially on one line. Thus, scaling the experimental data by $v_{min}$ gives a direct measurement of $f_{eq}$. From the dependence of $f_{eq}$ on denaturant concentration, which is easier to change than $T$ in experiments, the phase diagram (Fig. 1) for the various proteins can be measured.

**Native Topology Determines Force-Induced Unfolding Events.** The dramatic differences in the force-induced unfolding between S1 and S2, which are masked in thermal refolding (see legend to Fig. 1), can be predicted in terms of the topology of the native state. We have devised the following method (see Materials and Methods for technical details) by using the topology and the energetics of stabilization of the native structures to predict the order of unfolding events. (i) Starting with the native state, compute the energy of each strand by adding the various contact energies between the residues in a given strand with the rest of the structure. Here, $f_{eq,a}$ is the equilibrium unfolding force (obtained from the phase diagram) and the minimum pulling speed $v_{min} = k_u \Delta z$ ($k_u$ is the unfolding rate at $f = 0$, and $\Delta z$ is the extension in going from N to S). The one with the highest energy would unfold next and so on. (iv)

---

**Figure 1.** Native structures and phase diagrams for S1 and S2. Schematic sketch of the native structures for S1 (a) and S2 (c). In S1, terminal strands are tightly packed and are stabilized by an almost equal number of tertiary contacts (59 and 58, respectively). Strand D in S2 is exposed and forms only 35 contacts as opposed to the 57 formed by strand A. Thermal refolding ($f = 0$) for S1 and S2 shows two-state behavior. The phase diagrams for S1 (b) and S2 (d) are quantified by the thermal average of the fraction of native contacts $\langle Q(t, T) \rangle$. The values of $\langle Q \rangle$ are color coded according to the scales shown on the right. The green area in b, which corresponds to intermediate $\langle Q \rangle$, is extremely narrow. There is a large green region in the diagram for S2 (d). This large area suggests that S2 unfolds through partially unfolded structure I, in which the strand D is stretched, whereas strands A–C maintain their native conformation.
If there are two strands of nearly equal energies, we predict that they may nearly simultaneously unfold or that there may be parallel unfolding pathways. With this method, we successfully reproduced the simulated force-induced unfolding characteristics of S1 and S2.

Predicting Force-Induced Unfolding Characteristics for β-Sandwich Proteins. We applied this method to predict the unfolding pathways of three β-sandwich proteins, Ig27 (PDB code 1tit), 9FnIII (1fnf), and 10FnIII (1fnf), as well as the α-helical repeats of spectrin (1aj3). For all these modules, unfolding occurs by a shearing-type motion just as in I → S for S2. For Ig27 of the I band of titin, we find that the strand A unfolds first followed by A’ and G (Fig. 4a). Strands C, D, and E retain their structural integrity until late in the unfolding process. These observations are in accord with SMD simulations (10). An estimate of the energy barrier, with Q as a reaction coordinate (ref. 23; Fig. 4e), shows that force-induced unfolding TS is highly structured and occurs close to the native state (1, 2, 6, 17).

From the pulling speed dependence of the forces required to unfold individual domains of titin (1) and the unfolding rate at $f = 0$ (ref. 6; see legend to Fig. 3c), we predict that the equilibrium unfolding force is about 18 pN. A recent AFM experiment has shown that there is a “hump” in the force-extension curves (at $f = 100$ pN) on stretching tandem of repeats of individual Ig domains of human cardiac titin (17). The hump corresponds to an initial event in which strand A starts to shear away from the rest of the module (10, 17). Using various statistical potentials (see Materials and Methods), we find that the equilibrium unfolding force for A to unravel is $f_{eq} \approx 7$ pN. With this value, we predict that the initial intermediate in Ig27 module should occur at $f_s \approx f_{eq} \ln(v/v_{min}) \approx 80$ pN at $v = 0.5\mu m/s$.

Our method predicts that, in 10FnIII, strand G unravels first. This unraveling is followed by A, B, and F. On unfolding of B, the major barrier is overcome (Fig. 4e). The structure containing strands E, C, and D remains intact until late in the unfolding process. The calculations also predict a minor barrier between S1 and S2.
the contact potentials used, thus suggesting parallel unfolding pathways (24). The energy difference is sufficiently small that the potentials cannot discriminate between the two. (ii) The force required to unfold strand F is less than that needed to unfold strand B in 10FnIII. Because the threshold force is proportional to stability (for the conformations with similar topologies; Fig. 1), we conclude that 9FnIII is less stable than 10FnIII (19). (iii) Although the location and the structure of the unfolding TS for 9FnIII is rather similar to those for 10FnIII, the energy barrier is smaller; therefore, the unfolding rate for 9FnIII is greater than for 10FnIII.

**Force-Induced Unfolding of Spectrin.** Finally, we predict the unfolding pathways for the triple-helix spectrin for which AFM measurements of only the unfolding forces are available (13). Here, helix A unfolds first and is followed by a simultaneous unraveling of B and C helices (Fig. 4d). There is only one energy barrier for the unfolding process. The calculation of the energy profile for unfolding (Fig. 4c) indicates that unfolding TS of spectrin occurs close to the native state. Remarkably, on unraveling of helix A, the stability of the remaining structure is greatly compromised, which suggests that there should be no intermediate for spectrin unfolding. These conclusions were unchanged when we applied our method to all 20 PDB structures for spectrin and used different statistical potentials (see Materials and Methods). Using the estimates Ie = 4.4 pN and v = 8.1 × 10^{-2} μm/s (see Materials and Methods), we predict that the force required to unfold spectrin domain at = 0.8 μm/s is 40 pN (experimental value is ~30 pN; ref. 13). Rief et al. (13) have reported the unfolding forces for spectrin at two pulling speeds (0.8 μm/s and 8 μm/s). From these measurements and using our parameterization of the dependence of on (Fig. 4c), we can calculate the unfolding rate when = 0. We find that kₐ(f = 0) = 10^{-4} s⁻¹, which is only a factor of 20 smaller than the value obtained by using scaling arguments. Because f scales only logarithmically with , this difference results in a relatively small error in predictions for f.

**Conclusions**

In this article, we have provided a general framework for predicting the force-induced unfolding characteristics (pathways, pulling speed dependence of the unfolding forces, and location of TSs). Our approach is based on the hypothesis that unfolding by stretching can be understood by using only the topology of the native state. This hypothesis was firmly established by studying in detail the unfolding of two off-lattice model sequences with β-sandwich architecture similar to that of Ig27 and fibronectin domains. The remarkable differences between the stretching behaviors of two model sequences, which are not evident in thermal refolding calculations, are predicted quantitatively with our method. The success of our method in nearly quantitatively predicting the pulling speed dependence and location of TSs in Ig27, 10FnIII, 9FnIII, and spectrin confirms the validity of our theoretical model. Thus, given the native state of a protein, one can predict the force profiles for tandem repeats of identical protein modules.

Our method for predicting the characteristics of force-induced unfolding also suggests a way to measure the phase diagram of the sort shown in Fig. 1. In AFM and optical tweezer experiments, it is easier to alter the concentration of denaturants, C, than temperature. We propose that, by measuring the unfolding forces as a function of C, the equilibrium phase diagram in the plane of (C) can be obtained. Such experiments (especially when combined with fluorescence resonance energy transfer measurements) can provide a detailed map of the energy landscape of proteins.

Although the model devised herein is in nearly quantitative agreement with the experiments conducted thus far, we should...
emphasize its certain limitations. In particular, the location of the unfolding TS and associated barriers can be affected by the entropy of the unfolded regions of the chain. However, under tension, the fluctuations in these regions are expected to be small compared with the case of thermally induced (or denaturant-induced) unfolding.

One of the major issues is the relationship between the sequence of events in thermally induced and force-induced unfolding of proteins. The unfolding pathways may be similar in the limit of very slow pulling speeds. However, it is less clear whether unfolding pathways remain unaltered at pulling speeds normally used in experiments. Recent study of forced unfolding of β-hairpin indeed suggests that unfolding events depend on the pulling speed (25). This difference is probably exaggerated in SMD simulations because of very large pulling speeds. It would be desirable to carry out experimental and computational studies to compare thermal and forced unfolding pathways. Our preliminary results on S1 and S2 sequences suggest that there are differences in the thermal and forced unfolding.

We have focused only on the stretching of proteins in this article. The theory and the computational protocol is general and can be used to describe the stretching of other biomolecules and complexes. In particular, applications to RNA stretching can be made as long as the crystal structure is known. Such applications would be particularly useful given that much less is known about the folding landscape of RNA (26).

Note Added in Proof. We have applied our computational protocol to obtain the unfolding force for T4 lysozyme. Using the statistical potentials derived in ref. 16, we found that f ≈ 73 pN at the pulling speed V = 1 μm/s. The experimental value is 64 ± 16 pN (7).

We are grateful to Prof. G. H. Lorimer and Dr. M. Rief for penetrating discussions. This work was supported in part by National Science Foundation Grant CHE99-75150.