JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Exploring the Unfolding Pathway of Maltose Binding Proteins: An Integrated Computational Approach

Carlo Guardiani,[†] Daniele Di Marino,[†] Anna Tramontano,[†] Mauro Chinappi,[‡] and Fabio Cecconi*[,§]

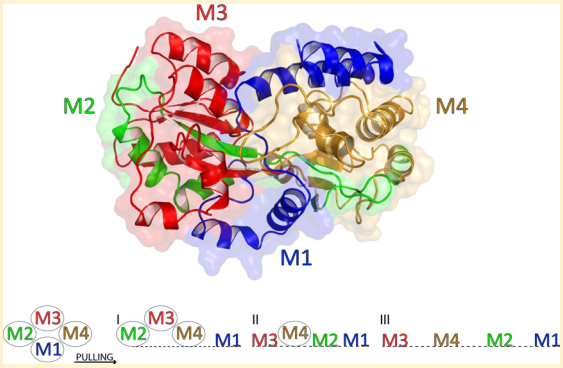
[†]Dipartimento di Fisica, Università di Roma "Sapienza", I-00185, Rome, Italy

[‡]Center for Life Nano Science, Istituto Italiano di Tecnologia (IIT), I-00185, Rome, Italy

[§]CNR—Istituto dei Sistemi Complessi (ISC), Via dei Taurini 19, I-00185, Rome, Italy

Supporting Information

ABSTRACT: Recent single-molecule force spectroscopy experiments on the Maltose Binding Proteins (MBPs) identified four stable structural units, termed *unfoldons*, that resist mechanical stress and determine the intermediates of the unfolding pathway. In this work, we analyze the topological origin and the dynamical role of the unfoldons using an integrated approach which combines a graph-theoretical analysis of the interaction network of the MBP native-state with steered molecular dynamics simulations. The topological analysis of the native state, while revealing the structural nature of the unfoldons, provides a framework to interpret the MBP mechanical unfolding pathway. Indeed, the experimental pathway can be effectively predicted by means of molecular dynamics simulations with a simple topology-based and low-resolution model of the MBP. The results obtained from the coarse-grained approach are confirmed and further refined by all-atom molecular dynamics.



INTRODUCTION

A number of experimental findings[1−5] and theoretical studies[6−10] support the view that the folding dynamics of several proteins is strongly influenced by native state topology, defined as the network of interactions (contacts) between neighboring amino acids in protein native structures. In particular, proteins with similar native states and similar transition states exhibit similar folding pathways despite a distant homology.[3,4] Moreover, simple structural parameters, such as the contact order, have been found to correlate well with the folding rates of small globular proteins.[1,2] While the role of topology in the folding of small, single-domain proteins is widely acknowledged,[11] its impact on the folding mechanism of multidomain proteins is less clear. An interesting benchmark in this regard, is represented by the Maltose Binding Protein (MBP)[12] a monomeric protein comprising two globular domains each formed by two discontinuous fragments (Figure 1), involved in the uptake and transport of maltose by the *Escherichia coli* metabolic system.

Even if MBP is known to exhibit reversible, two-state thermal and chemical denaturation,[13] single molecule force spectroscopy in combination with cysteine cross-link mutations revealed the existence of four stable structural subunits termed *unfoldons* M1, M2, M3, and M4 that, upon force application, unfold in a specific sequence: early breakdown of M1 followed by a simultaneous unfolding of M2 and M3 with a final disruption of M4.[14] The significant overlap between the unfoldons and the discontinuous protein fragments forming the globular domains of MBP (Figure 1C) suggests that the unfoldons might have a topological nature. Therefore, one can address the issue of how to identify unfoldons through the analysis of the interaction network directly using the native state of the MBP. We resort to three different strategies of topological analysis. The first one is a static analysis of the contact network of MBP through a multilevel graph-partitioning algorithm,[15] while the second one is the rigid-unit decomposition proposed by Potestio et al.[16] that, in some sense, takes into account the protein mobility. Both methods identify a number of discontinuous segments with significant overlap with the unfoldons. Instead, a third method, ProFlex,[17,18] performing an unbiased rigidity analysis of the contact network, provides a protein partitioning that only partially corresponds to the unfoldon structure.

The identification of the unfoldons through simple topological methods suggests the applicability of Gō-model simulations[19] to study the MBP mechanical unfolding. When structural properties become crucial, Gō-like approaches are successful in characterizing the mechanical unfolding[20,21] of proteins. They are able to easily identify the structural basis of mechanical stability[22−25] and to unveil the structural influence on protein transport across nanopores.[26−28] Moreover Gō modeling has been successfully applied to a wide range of other subjects such as, sampling of transition states,[29−31] macromolecular crowding and confinement,[32,33] action mechanisms of some enzymes,[34] and biomolecular machines.[35]

Received: April 2, 2014
Published: July 30, 2014

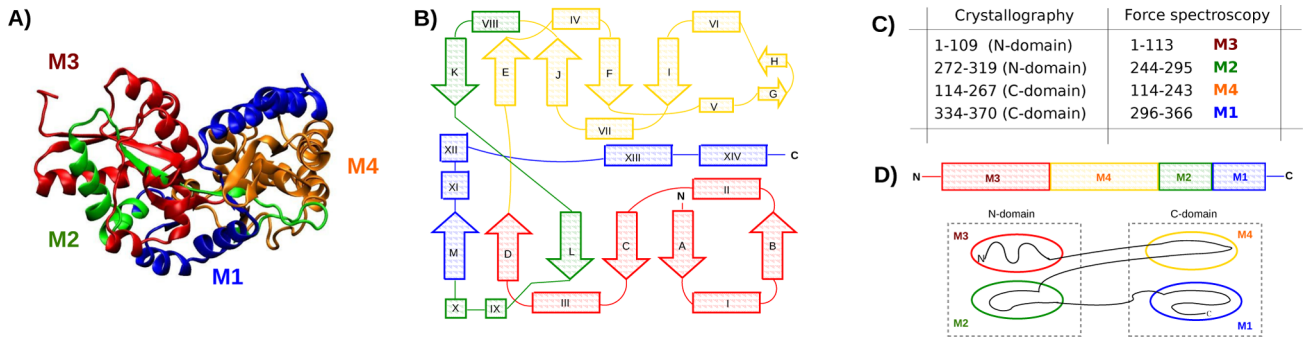


Figure 1. Structure of the Maltose Binding Protein (PDB ID: 4MBP). The color code identifies the four unfoldons: M1 blue; M2 green; M3 red; M4 gold. (A) Crystallographic structure. (B) Topological diagram (Adapted with permission from the work of Spurlino et al.[12] Copyright 1991 American Society for Biochemistry and Molecular Biology). (C) Crystallographic segments and unfoldons as defined in the works of Spurlino et al.[12] and Bertz and Rief[14]. (D) Cartoon representation of the domain and unfoldon organization.

Our simulations show that the steered molecular dynamics (SMD[36,37]) from the C terminus (C-pulling) using the Gō model agree with Bertz and Rief results.[14] Interestingly, the pulling SMD from N terminus (N-pulling) displays the same pathway of the C-pulling at low rates but a different pathway at fast rates. As discussed in the section Results and Discussion, the difference arises because likely the slow pulling first breaks the weakest unit, while at high pulling rates, the first unit to be stretched is the one directly subjected to the force.

Finally since the literature reports several cases of disagreement between the mechanical unfolding pathways predicted by Gō simulations (where non-native contacts are not allowed) and experimental findings[21,38] and non-native interactions are known to play a crucial role in some mechanically induced unfolding,[39] we repeated the mechanical unfolding through all-atom molecular dynamics. The atomistic simulations reproduced the same scenario observed by the Gō model, thus supporting the hypothesis of an important role of the native state topology as a driving force in MBP folding and unfolding.

MODEL AND METHODS

Topological Methods for Protein Partitioning. We applied three strategies for the topological analysis of a protein 3D reference structure, the *graph partitioning method*, *rigid body partitioning*, and *rigid cluster decomposition* with the purpose of obtaining a segmentation of the MBP into fragments naturally leading to the definition of four unfoldons.

(a). Graph Partitioning Method (SCOTCH). The partitioning of a protein can be addressed as a *graph partitioning* problem, where a graph is split in subgraphs each containing the same number of nodes and connected by a minimal number of edges. We found convenient to use the multilevel partitioning algorithm included in SCOTCH software package.[40] The contact map of the α carbons of MBP (PDB ID: 4MBP) computed using a cutoff distance of 7.5 Å was provided as input to SCOTCH. The program was set to partition the contact network in two subgraphs. In order to allow for a nonuniform partition of the graph, we set a maximum load unbalance ratio $\delta = 0.5$ where

$$\delta = \frac{1}{N_{\text{node}}} \sum_k \left| n_k - \frac{N_{\text{node}}}{N_{\text{part}}} \right| \tag{1}$$

with N_{node} being the number of nodes of the graph to partition ($N_{\text{node}} = 370$ for the MBP), N_{part}, the number of subgraphs the cluster must be split in (in our case $N_{\text{part}} = 2$), and n_k being the actual number of nodes assigned to the subgraph k.

(b). Rigid Body Partitioning (PiSQRD). Micheletti and coworkers[16] recently introduced a method for partitioning a protein in nearly rigid blocks exploiting the essential dynamical modes. In this approach a search is performed in the space of possible amino acid groupings to identify the partition in Q blocks that maximizes the fraction f of the mean square displacement of the protein coordinates captured by the rigid unit decomposition

$$f = \frac{\sum_{i=1}^{n} \lambda_i |\mathbf{r}_i^{\text{rb}}|^2}{\sum_{i=1}^{n} \lambda_i} \tag{2}$$

where the sum runs over the n essential modes with largest eigenvalues λ_i and $\mathbf{r}_i^{\text{rb}}$ is the best rigid body fit of the ith essential eigenvector. This quantity can be computed by solving the linear system $\mathbf{r}^{\text{rb}} = \mathbf{AU}$ where $\mathbf{U}$ is the vector of the $6Q$ roto-translation parameters and A is a $3N \times 6Q$ matrix whose elements only depend on the specific rigid-block partitioning that is to be evaluated. In principle, the method requires the solution of a linear system for each of the tens of thousands of possible amino acid groupings which makes it prohibitively expensive. A more convenient approach is thus to perform a preliminary exploration on the space of groupings through the minimization of a phenomenological energy function which ensures that the fluctuations of pairs of residues into the same semirigid block are small compared to the fluctuations of pairs of residues belonging to different units. The energy minimization is performed through a simulated annealing protocol with elementary moves consisting in changes of semirigid block assignment of individual residues. The algorithm is implemented in the PiSQRD web server[41] that we used for our calculations. The algorithm can either import the eigenvectors of a covariance matrix computed during an atomistic MD simulation, or it can employ the Hessian matrix of the β-Gaussian network model[42] calculated from the crystal structure of the protein. We took advantage of both options.

(c). Rigid Cluster Decomposition (ProFlex). For comparison we also employed ProFlex,[18] a computational tool for identifying rigid and flexible regions in protein structures based on a *pebble game* algorithm.[43] Since hydrogen bonds are an important component of the bond-network, it is crucial to input a structure that contains all the polar hydrogens. A parameter E_{cut}, hydrogen-bond energy cutoff, controls the number of hydrogen bonds included in the protein interaction

dx.doi.org/10.1021/ct500283s | J. Chem. Theory Comput. 2014, 10, 3589−3597

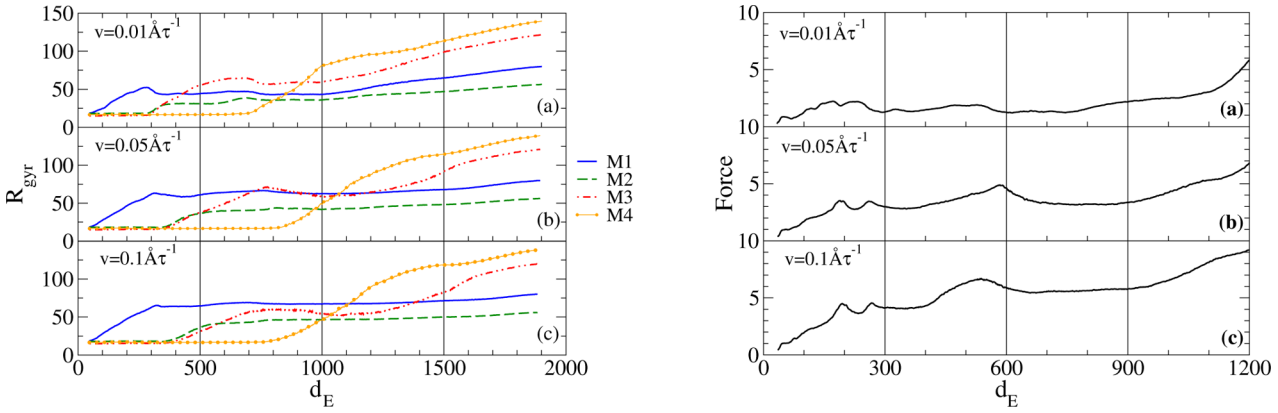


Figure 2. Gō model C-pulling. Evolution of the gyration radius for unfoldons M1, M2, M3, and M4 (left) and of the pulling force (right) as functions of the end-to-end distance (Å) at different pulling speeds. The curves represent average over 50 simulations.

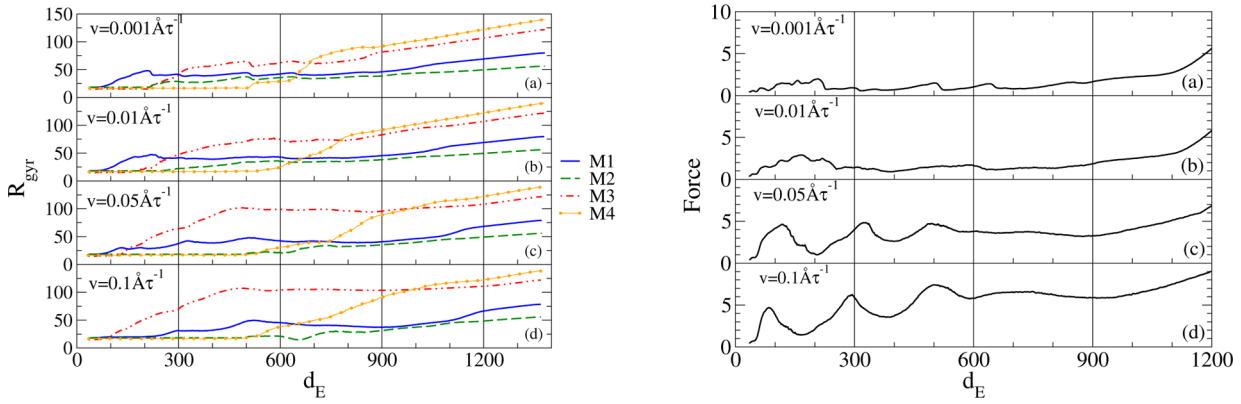


Figure 3. Gō model N-pulling. Evolution of the gyration radius for unfoldons M1, M2, M3, and M4 (left) and of the pulling force (right) as functions of the end-to-end distance (Å) at different pulling speeds. The curves represent average over 50 simulations.

network. In this study we used (i) the equilibrium configuration from an equilibrium MD simulation (see section All-Atom Simulations) and (ii) a conformation obtained from the MBP crystal (PDB ID: 4MBP) hydrogenated using the WHATIF program.[44] Details on the input parameters employed are reported in the Supporting Information.

Gō-Model Simulations. The Gō model is a minimalist off-lattice native-centric model portraying the protein as a chain of beads centered on the positions of the C_α atoms. The Gō energy function reproduces a minimally frustrated funnel landscape whose bottom is occupied by the native state. Here we employ the force field proposed by Clementi et al.[31] supplemented with a harmonic stretching term to perform steered molecular dynamics simulations. All the quantities involved in the coarse-grained Gō-simulations are expressed in reduced units (see the Supporting Information, where also details on the interaction potential are reported). The reduced units of the Gō model can be converted to physical units using appropriate experimental data to set the energy scale. For instance, the match between theoretical and experimental thermal-unfolding temperatures can be used.[45] However, the conversion of the code time unit into picoseconds does not provide deep insights into the physics of the process, as the coarse-grained nature of the model makes the simulation and experimental pulling rates inherently different. We refer the interested reader to the work of Padding and Louis[46] for a general and thorough discussion about coarse-grained approaches and time-scale issues.

Steered molecular dynamics simulations were performed fixing one of the two terminal beads of the protein and connecting the other one to an harmonic spring with force constant $k_s = 0.1$. The free end of the spring was attached to a dummy atom moving along the direction of the native end-to-end vector at a constant velocity. The pulling potential thus reads

$$V_{\text{pull}} = \frac{k_s}{2}|\mathbf{r}_d - \mathbf{r}_p|^2 \tag{3}$$

where $\mathbf{r}_d$ and $\mathbf{r}_p$ are the position vectors of the dummy and pulling beads, respectively. The equations of motion were integrated using a stochastic position Verlet algorithm.[47] The temperature, set to $T = 0.5$, was kept constant through a Langevin thermostat with friction coefficient $\gamma = 0.25$. The time-step was set to $\delta t = 5 \times 10^{-3}$, and the system was equilibrated for a time $t_{eq} = 10^3$ before applying the pulling force. The duration of the SMD run was modulated according to the pulling speed to attain a final elongation of 1500 Å. Physical observables were sampled and averaged over 50 independent runs, specifically, frames were collected every 15 time units in each mechanical unfolding run. In each frame, we measured the pulling force and the gyration radius of every unfoldon and, then, estimated the average conditioned to a given value of the end-to-end distance d_E (see Figures 2 and 3).

All-Atom Simulations. Constant velocity steered molecular dynamics (SMD) simulations[36,37] were performed using the following (quite-standard) protocol. The crystal structure

dx.doi.org/10.1021/ct500283s | J. Chem. Theory Comput. 2014, 10, 3589−3597

for MBP (PDB code: 4MBP[48]) was rotated in order to have the end-to-end vector parallel to x-axis, then the structure was completed with missing atoms and solvated using VMD[49] and 9 Na^+ ions were added to neutralize the system total charge. The total number of atoms is 971 626, the initial box dimension is (1000, 100, 100 Å). After an energy minimization of 10^4 steps, six 0.2 ns NPT equilibration steps (P = 1 atm and T = 50, 100, ..., 300 K) were performed. The final size of the box is (986, 98.6, 98.6 Å). This equilibrated configuration was employed as starting point for both N-pulling (C terminal fixed) and C-pulling (N terminal fixed) NVT runs. The C_α of the pulling terminal is harmonically constrained via the SMD potential

$$U(\mathbf{r}_{N/C}) = \frac{k}{2}[vt - (\mathbf{r}_{N/C}(t) - \mathbf{r}_{N/C}(0))\cdot\hat{\mathbf{n}}]^2$$

with $\mathbf{r}_{N/C}$ as the position of the C_α of the pulling terminal, $\hat{\mathbf{n}}$ as the direction parallel (or antiparallel) to the x-axis, v = 0.05 Å/ps, and k = 7 kcal/mol·Å^2. This pulling speed is several orders of magnitude higher than the experimental pulling speed[14] (v = 1 μ/s). This large discrepancy between experimental and simulation pulling rate is typical of non equilibrium all-atoms MD, and it is needed to explore the phenomenon with the available computational resources. The CHARMM22[50] force field topology parameters and the TIP3P model were employed for the protein and the water, respectively. All simulations were performed using NAMD.[51]

RESULTS AND DISCUSSION

Unfoldons and Topology. The MBP protein has an ellipsoidal shape and is divided in two globular domains (named C- and N-domain) each of them composed by two discontinuous segments; see Figure 1. To unravel the topological properties of the MBP native state, we input the contact map of the α-carbons of the PDB structure ID: 4MBP computed using a cutoff distance of 7.5 Å, into the graph-partitioner of the SCOTCH software. The result is reported in Table 1. The two subgraphs identified, dubbed SG_1 and SG_2, comprise a number of discontinuous segments. Following a common approach in protein partitioning,[16] fragments smaller than one-twentieth of the protein length (i.e., 18 residues for our case) were reassigned to the nearest flanking unit. The merging starts from the smallest fragments. For instance, segment 227−233, belonging to group 1, comprises only seven residues and is located between two stretches belonging to group 2. It is hence merged in a single segment that was assigned to group 2 (115−256).

Table 1. Bipartition of the MBP through the SCOTCH Software Package[40a]

segment	group	group (final)	overlap ω
1−114	1	SG_1 (1−114)	0.99 (M3)
115−226	2		
227−233	1	SG_2 (115−256)	0.95 (M4)
234−256	2		
257−334	1		
335−335	2	SG_1 (257−336)	0.59 (M2)
336−336	1		
337−370	2	SG_2 (337−370)	0.57 (M1)

[a]The algorithm divides the protein into eight segments, whose residues are indicated in column 1. Fragments smaller than one-twentieth of the protein length were reassigned to the nearest flanking unit, starting from the smaller one. The final result of this protocol is listed in column 3. The last column reports the overlap of the four segments with the most similar unfoldon (in brackets).

The merging and reassignment procedure yielded four segments that were compared with the crystallographic domains and the unfoldons using the overlap $\omega(L_i, L_j) = 2n_c/(N_i + N_j)$ where N_i and N_j are the total numbers of residues of the two segments L_i, L_j to be compared, and n_c is the number of their common residues. The two discontinuous fragments making up SG_1 display a remarkable overlap with unfoldons M3 and M2, while the two segments of SG_2 feature high overlap with unfoldons M4 and M1 respectively.

The above MBP splitting is based on a static picture of the native state and one may ask whether an analysis that takes into account the vibrational dynamics would lead to the same scenario. In this respect, the rigid-body partitioning server PiSQRD[16] was applied in two modes. In the first, we submitted to the server the crystal structure of the MPB (PDB ID: 4MBP) to obtain the decomposition via the β-Gaussian network model. In the second, we input into the server the eigenvectors of the covariance matrix corresponding to the 10 largest eigenvalues calculated from the last 0.1 ns of the equilibration at 300 K performed before the SMD run (see section Model and Methods). Since MBP is composed by two globular domains, the most natural choice is requiring the partitioning into two rigid units. As shown in the Supporting Information, this partition captures about 55% of the protein mobility (see Supporting Information Figure S3). The output of PiSQRD was postprocessed removing the shortest fragments and reassigning their amino-acid residues to the neighboring fragments as already done for the SCOTCH outcome. The resulting partition amounts to four fragments that are listed in Table 2.

Table 2. Partitioning of the MBP through the PiSQRD Server[41a]

segment	block	block (final)	overlap ω
1−111	2	RB_2 (1−111)	0.99 (M3)
112−228	1		
229−230	2	RB_1 (112−259)	0.93 (M4)
231−259	1		
260−313	2	RB_2 (260−313)	0.70 (M2)
314−330	1		
331−334	2	RB_1 (314−370)	0.83 (M1)
335−370	1		

[a]As in Table 1, the first and third columns show the boundaries of the discontinuous fragments before and after the merging procedure, respectively. For each fragment, the overlap with the most similar unfoldon is reported in column four.

Both algorithms yielded very similar fragments showing that the partition into unfoldons is robust.

For comparison we also analyzed the MBP structure via ProFlex[18] software. The pre-equilibrated conformation was processed by using four different hydrogen-bond energy cutoffs, E_{cut} = 3.0, −0.1, −1.0, and −1.2 kcal/mol, which amounts to taking into account fewer and fewer hydrogen bonds in the interaction network. As reported in Table 3 and shown in Supporting Information Figure S1, for a large number of H-bonds, E_{cut} = 3.0 kcal/mol, ProFlex identifies two main rigid clusters (RC1 and RC2). RC1 basically includes the whole N terminal subdomain while RC2 comprises the most stable part

Table 3. (Upper) Overlap of the Two Largest ProFlex Clusters RC1 and RC2 for E_{cut} = 3 kcal/mol and the Four Unfoldons (Lower) Same Comparison for E_{cut}= −1.0 kcal/mol[a]

	E_{cut} = 3 kcal/mol	
RC1 frag	unfoldons	overlap ω
5−97		
105−112	M3 1−113	0.94
259−286	M2 244−295	0.70
301−305	M1 296−370	0.12
RC2 frag	unfoldons	overlap ω
113−119		
209−228	M4 114−243	0.35
241−247		
314−325	M1 296−370	0.28
	E_{cut} = −1 kcal/mol	
RC1 frag	unfoldons	overlap ω
8−10		
42−79	M3 1−113	0.57
105−108		
262−286	M2 244−295	0.65

[a]RC2 is not reported since, for this value of E_{cut}, it is no longer recognized as a rigid cluster but as a flexible region, see Supporting Information Figure S1.

of the C terminal subdomain: strands β_E, β_J, β_K, and helices α_{VII} and α_{XII}. Some fragments of RC1 exhibit a good overlap with unfoldons M2 and M3, while a modest overlap is observed between RC2 and M1 and M4 (Table 3). When E_{cut} decreases, RC2 is lost and RC1 becomes smaller and smaller. For instance, at E_{cut}= −1.0 kcal/mol, RC1 still includes strands β_A, β_C, and β_L (which represent the most stable part of the β-sheet of the N terminal subdomain) as well as helices α_{II}, α_{III}, α_{IX}, and α_X. As shown in Table 3, the rigid block, RC1, still retains a good overlap with unfoldons M2 and M3 while the other clusters are so small that no information on the unfoldons can be extracted. A similar analysis repeated on the crystal structure also gave no significant overlap with unfoldons (see Supporting Information Figure S2 and Table S1). The discussion is reported in the Supporting Information.

In summary, at variance with SCOTCH and PiSQRD methods, the ProFlex analysis fails to identify the four unfoldons due to a different working strategy. The SCOTCH and PiSQRD calculations both allow exploiting the a priori information on the existence of two subdomains, while this crucial constraint cannot be entered into ProFlex. Despite these differences, however, ProFlex was capable to identify, without a bias, the N terminal subdomain as the major rigid block composed by M2 and M3.

Gō Simulations. The discussion in the above section has highlighted the topological nature of the unfoldons that can be further confirmed by steered molecular dynamics simulations using the Gō model where the dynamics is completely driven by native interactions and thus by the topology of the native state.

As the mechanical unfolding pathways are known to be affected by the loading rate,[52] we ran simulations at different pulling velocity. Moreover to investigate the dependence of the mechanical unfolding on the force application point, we stretched the MBP from both N and C terminals while keeping the other fixed.

(a) C-pulling. We performed C-pulling simulations at speed 0.01, 0.05, and 0.1 Å/τ, with τ being the code time unit. In Figure 2, we plot the gyration radius R_{gyr} of the unfoldons as a function of the end-to-end distance of the whole protein d_E. At each pulling speed, three different unfolding events can be identified. The stretching of MBP begins with the unraveling of unfoldon M1 until the protein reaches a length of about 250 Å. Immediately after M1, unfoldons M2 and M3 simultaneously start unfolding. Since M2 is approximately only half the length of M3, the curve corresponding to M2 attains a quasi-plateau before the complete unfolding of M3. Unfoldon M4 is the last structural element to be lost. This sequence of breakdown events agrees with the results of Bertz and Rief[14]. The crystallographic structure of MBP reveals that unfoldon M2 is closely intertwined with M3 since they form together the N terminal domain, hence, the unfolding of M2 and M3 is expected to be coupled. In particular, according to the topological diagram of 1, M2 contributes a long β-strand (strand β_L) to the six-stranded β-sheet of the N terminal subdomain. The extraction of this strand when M2 is pulled away destabilizes M3. Differently, although M4 is associated with M1 in the C terminal domain, it interacts with M1 only through weak bonds between α-helices (Figure 1), so that the unfolding of M1 can nearly leave the structure of M4 unaltered. To summarize, the unfolding of M2 and M3 is necessarily coordinated, while M1 and M4 can break down independently even if they belong to the same globular domain.

The force−extension curves, Figure 2 right, display two modest humps at elongations of 200−300 Å and d_E = 500−600 Å. The comparison with the R_{gyr} plots reveals that the first hump corresponds to the breakdown of unfoldon M1 while the latter to the concerted disruption of M2 and M3.

(b) N-pulling. For the N-pulling simulations, we used the same velocities 0.01, 0.05, 0.1 Å/τ of the C-pulling with an additional run at v = 0.001 Å/τ. Figure 3 illustrates the behavior of R_{gyr} as a function of distance d_E of the whole protein. At variance with the C-pulling, the N-pulling is characterized by two different sequences of events depending on fast or slow pulling.

At low pulling velocity (v = 0.01 and 0.001 Å/τ) the unfolding mechanism is similar to the one observed in the C-pulling simulations: unfoldon M1 breaks down first, followed by units M3 and M2 and finally by unfoldon M4. The only difference with the C-pulling simulations, is that M3 begins to unfold slightly earlier and proceeds faster (higher slope of the R_{gyr} curve) than the unraveling of M2.

The scenario instead is very different at high pulling rates (v = 0.05 and 0.1 Å/τ), where we observe an early simultaneous unfolding of M1 and M3 that comes to an end when the protein attains an elongation of about 500 Å. After this event, R_{gyr} of M1 and M3 maintains an approximately stable plateau corresponding to their size (M3 is twice longer than M1). The opening of M1 and M3 is then rapidly followed by the simultaneous breakdown of M2 and M4 that begins when the protein has reached an extension of 500−600 Å.

A further characterization of the unfolding process of M1 and M3 can be achieved through the inspection of the force−extension plots in Figure 3. At high pulling speeds (v = 0.05 and 0.1 Å/τ), the force−extension plots display three peaks at elongations d_E = 100, 300, 500 Å, before the major events in the unfolding pathway. The force peak at d_E = 100 Å precedes the destruction of the helix−loop−helix motif formed by helices α_{XIII} and α_{XIV} which is the first step in the unfolding of M1. The

peak of the force at d_E = 300 Å triggers the stretching of the N terminal part of unfoldon M3, from strand β_A to strand β_C, destroying the β-sheet of the N-domain. Finally, when the force reaches the peak at d_E = 500 Å, M3 becomes completely unfolded and strand β_E is dragged away from strand β_J opening the way to the unraveling of M4.

In this mechanism thus, the unfolding of the N terminal (M3) and C terminal (M1) tails of the MPB is followed by the unraveling of the central part (M4−M2) of the chain. The reason for the early disruption of the two terminal tails however is likely different, and such difference can be rationalized through the following argument: at low rates the first unfoldon to open is the "weakest" one, while at larger rates, there is a higher probability that the breakage involves the unit to which the force is directly applied. This scenario is easily illustrated by the behavior of a one-dimensional chain of four coupled nonlinear oscillators that may be seen as a cartoon of the unfoldons, Figure 4. The system is characterized by the bead coordinates q_0, ..., q_4 and by the global energy function

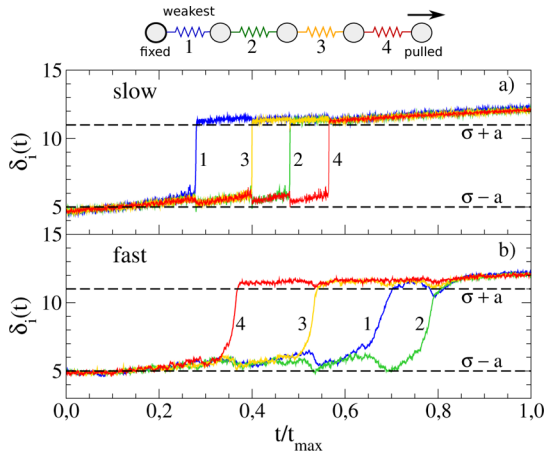


Figure 4. Instances of the stretching dynamics for a chain with four nonlinear springs representing a cartoon of the unfoldons. The parameters of the model (4,5) are ε_1 = 0.8, $\varepsilon_2 = \varepsilon_3 = \varepsilon_4$ = 1.0, σ = 8.0, a = 3.0, k_p = 5.0. The horizontal dashed lines indicate the two well minima. The chain is initialized with all the coordinates in the first well of potential 5, yielding a length $L(0) = (\sigma - a)n$. Slow pulling v_p = 0.34 (a), the weakest bond (i = 1, blue) is the first to be stretched. Fast pulling v_p = 6.8 (b), the pulled bond (i = 4, red) is the first to be broken.

$$E = \sum_{i=1}^{4} V_i(q_i - q_{i-1}) + \frac{k_p}{2}(L(t) - q_4)^2 \quad (4)$$

Bead 0 is grafted to a substrate and cannot move, and bead 4 is pulled through an harmonic spring whose free terminal moves with the law $L(t) = L(0) + v_p t$ as in the SMD. Each nonlinear spring is described by a double-well interaction

$$V_i(u) = \frac{\varepsilon_i}{4}[(u - \sigma)^2 - a^2]^2 \quad (5)$$

which mimics the transition between two states (minima): compact $u_- = \sigma - a$ and broken $u_+ = \sigma + a$, σ is the average interbead distance and a tunes the position of the minima with respect to σ. Each ε_i defines the stiffness of each bond as it determines the barrier height $a^4\varepsilon_i/4$ between the minima. We assume that the weakest bond is the first, thus, ε_1 = 0.8 and $\varepsilon_2 = \varepsilon_3 = \varepsilon_4$ = 1.0.

The Brownian dynamics at temperature T = 1 of each bond length, $\delta_i(t) = q_i(t) - q_{i-1}(t)$ for i = 1, ..., 4 is plotted in Figure 4a and b at slow and fast pulling, respectively. At low rates, we see that the first link to be broken is the weakest one (ε_1 = 0.8, blue). Indeed, when the pulling is so slow that the chain relaxation dynamics takes place, the force applied to the terminal bead has enough time to evenly distribute among the bonds. In this case, the weakest bond is the most probable candidate to the rupture. On the contrary, if the pulling is so fast that no stress relaxation occurs, with the highest probability the pulled bond (i = 4, red) is the first to be stretched. This simple scenario is consistent with the unfolding mechanism observed in the N-pulling simulations, whereby unfoldon M3, the closest to the force application point, is the first to be broken in the case of fast pulling, whereas unfoldon M1, the weakest one, is the first to unravel in the case of slow pulling. In the C-pulling, unfoldon M1 is always the first to be stretched regardless of the pulling speed. Indeed, at low speed, M1 is the first to be broken being the weakest one, while at fast pulling, it breaks ahead of the others being the pulled one. This also explains why, the low-forcing unfolding sequence is the same regardless of the force application point (C and N).

Atomistic Simulations. The Gō-model SMD simulations yield an unfolding mechanism in good agreement with force spectroscopy experiments of the MBP. However, there are details of the protein dynamics, such as the possible formation of non-native contacts and the steric hindrance of side chains, that cannot be captured by the topology-driven models. These limitations suggest the opportunity to repeat the SMD simulations with a more realistic atomistic force-field. As discussed in the section Models and Methods, we used the CHARMM force field[53] in explicit water.

(a) C-pulling. The results of the C-pulling simulation are illustrated in Figure 5a. The evolution of gyration radius of each unfoldon confirms the unfolding mechanism predicted by the Gō simulations: i.e., early breakdown of unfoldon M1 followed by simultaneous disruption of M2 and M3. The force−extension plot (Figure 5b) exhibits two force peaks at the extensions of about 250 and 400 Å. As mentioned in the section Models and Methods, the simulations are run at pulling rates that are orders of magnitude larger than the experimental ones. As a consequence, also the applied forces (several hundreds of piconewtons) are larger than the ones exerted in experiments (decades of piconewtons). In general one could only expect a qualitative agreement between the MD and experimental unfolding pathway,[36] although interesting techniques have been proposed in the literature[54] to extract low-rate data, from an ensemble of high forcing MD simulations.

The inspection of the trajectory reveals that when MBP attains an elongation of about 240 Å, the two helices closest to the C-terminus, α_{XIII} and α_{XIV}, are completely unraveled starting the process of disruption of unfoldon M1. Conversely, when $320 \lesssim d_E \lesssim 360$ Å, the strand β_L under the action of the pulling force drags the portion of unfoldon M3 between helix α_{III} and strand β_D with itself. Finally strand β_L is extracted from the β-sheet of the C domain that is destabilized leaving only a marginal structured element comprising strand β_A, helix α_I and strand β_B. Representative conformations are reported in lower panels of Figure 5. The agreement between the unfolding pathway predicted by atomistic and Gō-model simulations suggests that the folding/unfolding process of MBP is mainly

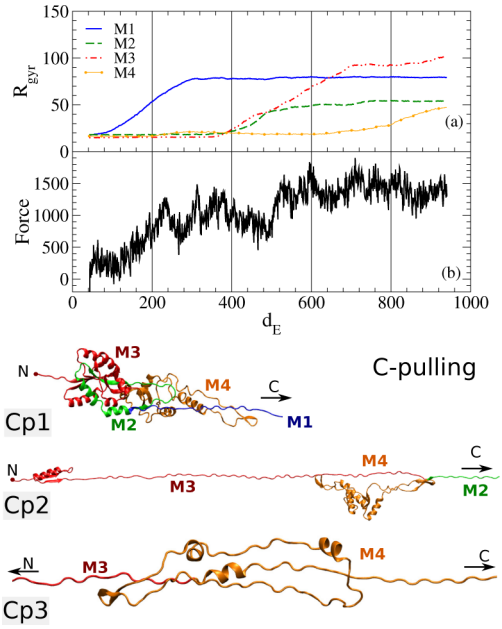


Figure 5. Evolution of the gyration radius (Å, a) for the four unfoldons M1, M2, M3, and M4, and pulling force (pN, b) as a function of the end-to-end distance d_E (Å) in the C-pulling all-atom steered molecular dynamics simulation. The lower panels report three conformations corresponding to d_E = 318 Å (Cp1), d_E = 762 Å (Cp2), and d_E = 940 Å (Cp3).

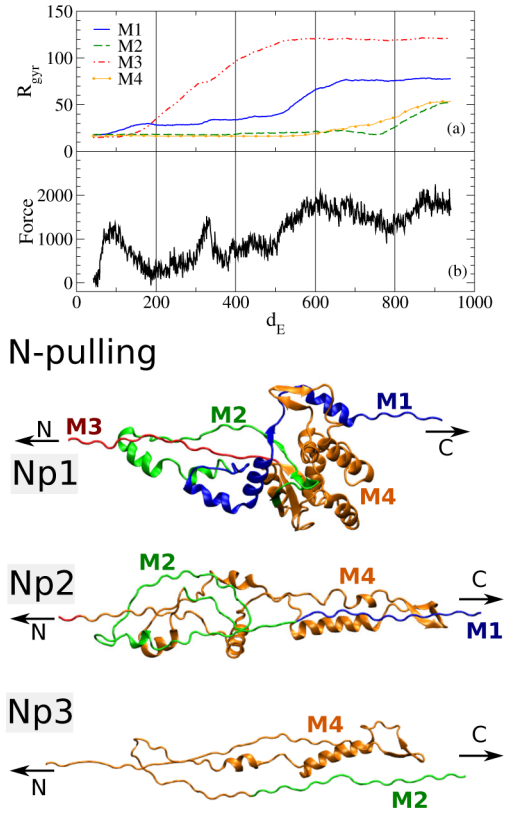


Figure 6. Evolution of the gyration radius (Å, a) for the four unfoldons M1, M2, M3, and M4 and pulling force (pN, b) as a function of the end-to-end distance d_E (Å) in the N-pulling all-atom steered molecular dynamics simulation. The lower panels report three conformations corresponding to d_E = 511 Å (Np1), d_E = 740 Å (Np2), and d_E = 940 Å (Np3).

driven by the topology of the native state while the fine details of the amino-acid sequence apparently only play a minor role.

(b) N-pulling. Figure 6 displays the same profiles for the N-pulling case. The unfolding mechanism appears to be very different from the one observed in the C-pulling run. In fact, unfoldon M3, the closest to the force application point, is unraveled first, followed by the disruption of M1. Finally, unfoldons M2 and M4, corresponding to the central region of the protein sequence, break down almost simultaneously. This mechanism is thus consistent with the sequence of the unfolding events exhibited by the fast N-pulling Gō-model simulations (see Figure 3). The force−extension plot shows two peaks at elongations of 50 and 300 Å respectively and attains a quasi-plateau at $d_E \simeq 550$ Å. The inspection of the trajectory reveals that the first peak corresponds to the unfolding of the $\beta_A-\alpha_I-\beta_B$ motif in unfoldon M3. The second force peak conversely appears immediately before the unfolding of the $\beta_C-\alpha_{III}-\beta_D$ element, again in unfoldon M3. Finally, the first event occurring at the onset of the force plateau is the unraveling of the helical elements of unfoldon M1. Representative configurations are reported in the lower panels of Figure 6.

Role of Non-native Interactions. The agreement between the Gō model and all-atom MD suggests that non-native interactions (not allowed in the Gō model) do not influence the overall unfolding pathway. To support this statement, we computed the occurrence of non-native contacts in the all-atom MD trajectories. We defined two residues to be in non-native contact if the distance d between their α-carbons in the current conformation is lower than R_c = 7.5 Å provided they are not in contact in the native state. The total amount of possible non-native contacts is 66548 (see the Supporting Information for details) but we found that only less than 0.5% of them occurred in more than 10% of conformations along the all-atom unfolding trajectories (see Supporting Information Table S2). However, one could argue that few frequently occurring non-native contacts established in critical positions along the chain may stabilize non-native unfolding intermediates. From the contact maps shown in Figure S4 of the Supporting Information, it is apparent that non-native contacts mainly occur in paired β-strands and that they are established when the pulling force causes the sliding of one strand along the other. Such a displacement yields residue j to interact preferably with either $i + 1$ or $i - 1$, rather than with its native partner i. Due to this peculiar location, the non-native contacts established during the mechanical unfolding of MBP are not able to stabilize non-native intermediates but they, to some extent, enhance the stability of the native conformation. This is in agreement with the results of ref 39 showing that native contacts do play a key role during cotranslocational unfolding but they do not significantly affect the mechanical unfolding in the absence of the pore.

CONCLUSIONS

The present work has been inspired and motivated by a recent force spectroscopy experiment by Rief et al. which identified four structural units, termed unfoldons, in the mechanical unfolding pathway of the Maltose Binding Protein (MBP) and its mutants.[14] We performed a computational study to assess possible topological origins of these unfoldons. Our analysis started from the observation that the crystal structure of the MBP native state consists of two globular domains each

composed by two discontinuous chain fragments.[12] The significant overlap between the unfoldons and the discontinuous fragments of the MBP domains prompted us to apply algorithms similar to those employed for the identification of structural domains.[55] Specifically, we used a multilevel graph partitioning algorithm[15,40] to bisect the protein in two subgraphs linked by a minimal number of edges. The two partitions turned out to be formed by two discontinuous segments showing extensive overlap with the unfoldons. The analysis was repeated using a dynamic partitioning method[16,41] which decomposes the chain into rigid-blocks trying to capture most of the protein mobility. This technique also yielded four fragments that well overlap with the unfoldons. Both protocols used as input the structural information that MBP is made by two domains. A comparison with a completely unbiased approach as ProFlex,[17,18] performing a rigidity cluster analysis, indicates that such information is crucial for a correct identification of unfoldons.

The structural nature of the unfoldons revealed by the partitioning analysis suggested the possibility to predict the MBP mechanical unfolding pathway by means of Gō-model simulations,[19,31] since a Gō-model approach represents the straightforward implementation of the basic topology properties of protein structures into a force-field without resorting to an heavy design. Our C-pulling SMD simulations correctly reproduced the experimental unfolding mechanism[14] at all pulling speeds. In particular, the first event to occur is the detachment and destabilization of unfoldon M1 that is very loosely bound to M4. The subsequent unraveling of unfoldon M2 implies the removal of the long β_L-strand from the β-sheet of the N domain causing the destruction of unfoldon M3, that is followed by the unfolding of M4, the most mechanically stable element of MBP. The same unfolding mechanism was observed in low-speed N-pulling SMD simulations suggesting that the sequence M1−(M2, M3)−M4 must be the minimal free energy pathway. Conversely, when the N-pulling was performed at high speed, the first disrupted unfoldon was the one closest to the point of application of the pulling force i.e. M3. This difference is consistent with the scenario that high-rate pulling protocols more likely break at first the pulled units, while the slow pulling protocols tend to break the weakest units.

Similar results were obtained when the simulations were repeated at the atomistic level using the detailed CHARMM force field.[53]

Our study thus shows that topology is a relevant driving force of the MBP unfolding and confirms the effectiveness of the topology-based low-resolution modeling. Obviously our results do not necessarily imply that topology-based approaches could be used in an acritical way in any problem concerning proteins, but suggest that topology can be a proper conceptual framework even for the analysis of large multidomain proteins. Therefore, the Gō-like models can constitute a viable computational technique for the study of proteins that, for sizes and complex structures, still represent a prohibitive challenge to full atomistic simulations.

■ ASSOCIATED CONTENT

Supporting Information

Details on the Gō model implemented. Fraction of the mobility captured as a function of the number of partitions from PiSQRT server. ProFlex additional results. Non-native contact additional results. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: fabio.cecconi@roma1.infn.it. Phone: (+39) 06-4993-7452. Fax: (+39) 06-4993-7440.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the CINECA (ISCRA project NAPS) for the availability of high performance computing resources and support. This research used the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia.

■ REFERENCES

(1) Simons, K.; Plaxco, K.; Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **1998**, *277*, 985−994.

(2) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I.; Baker, D. Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Mol. Biol.* **1999**, *6*, 1016−1024.

(3) Chiti, F.; Taddei, N.; White, P.; Bucciantini, M.; Magherini, F.; Stefani, M.; Dobson, C. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol.* **1999**, *6*, 1005−1009.

(4) Martínez, J. C.; Serrano, L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Mol. Biol.* **1999**, *6*, 1010−1016.

(5) Fersht, A. R. Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (4), 1525−1529.

(6) Dinner, A. R.; Sali, A.; Karplus, M. The folding mechanism of larger model proteins: role of native structure. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8356−8361.

(7) Alm, E.; Baker, D. Prediction of proteinfolding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11305−11310.

(8) Clementi, C.; Jennings, P. A.; Onuchic, J. N. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1β. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5871−5876.

(9) Koga, N.; Takada, S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.* **2001**, *313*, 171−180.

(10) Plotkin, S.; Onuchic, J. Understanding protein folding with energy landscape theory. Part I: basic concepts. *Q. Rev. Biophys.* **2002**, *35*, 111−167.

(11) Baker, D. A surprising simplicity to protein folding. *Nature* **2000**, *405*, 39−42.

(12) Spurlino, J.; Lu, G.; Quiocho, F. The 2.3 Angstrom resolution structure of the Maltose- or Maltodextrin-binding Protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* **1991**, *266*, 5202−5219.

(13) Ganesh, C.; Shah, A.; Swaminathan, C.; Surolia, A.; Varadarajan, R. Thermodynamic characterization of the reversible, two-state unfolding of Maltose Binding Protein, a large two-domain protein. *Biochemistry* **1997**, *36*, 5020−5028.

(14) Bertz, M.; Rief, M. Mechanical unfoldons as building blocks of Maltose-binding Protein. *J. Mol. Biol.* **2008**, *378*, 447−458.

(15) Hendrickson, B.; Leland, R. A multilevel algorithm for partitioning graphs. *Proceedings of Supercomputing '95*, San Diego, CA, Dec 8, 1995; Technical Report SAND93-1301.

(16) Potestio, R.; Pontiggia, F.; Micheletti, C. Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. *Biophys. J.* **2009**, *96*, 4993−5002.
(17) Jacobs, D.; Rader, A.; Kuhn, L.; Thorpe, M. Protein flexibility predictions using graph theory. *Proteins: Sruct., Funct., Genet.* **2001**, *44*, 150−165.
(18) Jacobs, D.; Hespenheide, B.; Kuhn, L.; Thorpe, M. Protein unfolding: rigidity lost. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3540−3545.
(19) Gō, N.; Abe, H. Noninteracting localstructure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* **1981**, *20*, 991−1011.
(20) Kleiner, A.; Shakhnovich, E. The mechanical unfolding of ubiquitin through all-atom Monte Carlo simulation with a Gō-type potential. *Biophys. J.* **2007**, *92*, 2054−2061.
(21) Li, M.; Kouza, M.; Hu, C. Refolding upon force quench and pathways of mechanical and thermal unfolding of ubiquitin. *Biophys. J.* **2007**, *92*, 547−561.
(22) Cieplak, M.; Marszalek, P. Mechanical unfolding of ubiquitin molecules. *J. Chem. Phys.* **2009**, *123*, 194903.
(23) West, D.; Olmsted, P.; Paci, E. Mechanical unfolding revisited through a simple but realistic model. *J. Chem. Phys.* **2006**, *124*, 154909.
(24) Sulkowska, J.; Cieplak, M. Mechanical stretching of proteins - a theoretical survey of the Protein Data Bank. *J. Phys.: Condens. Matter* **2007**, *19*, 283201.
(25) Kouza, M.; Hu, C.-K.; Li, M. New force replica exchange method and protein folding pathways probed by force-clamp technique. *J. Chem. Phys.* **2008**, *128*, 045103.
(26) Bacci, M.; Chinappi, M.; Casciola, C.; Cecconi, F. Role of Denaturation in Maltose Binding Protein Translocation Dynamics. *J. Phys. Chem. B* **2012**, *116*, 4255−4262.
(27) Bacci, M.; Chinappi, M.; Casciola, C. M.; Cecconi, F. Protein translocation in narrow pores: Inferring bottlenecks from native structure topology. *Phys. Rev. E* **2013**, *88*, 022712.
(28) Cecconi, F.; Bacci, M.; Chinappi, M. Protein Transport Across Nanopores: A Statistical Mechanical Perspective From Coarse-Grained Modeling and Approaches. *Protein Pept. Lett.* **2014**, *21*, 227−234.
(29) Cecconi, F.; Guardiani, C.; Livi, R. Testing simplified protein models of the hPin1 WW domain. *Biophys. J.* **2006**, *91*, 694−704.
(30) Cecconi, F.; Guardiani, C.; Livi, R. Stability and kinetic properties of C5-domain from Myosin Binding Protein C and its mutants. *Biophys. J.* **2008**, *94*, 1403−1411.
(31) Clementi, C.; Nymeyer, H.; Onuchic, J. Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding ? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(32) Homouz, D.; Perham, M.; Samiotakis, A.; Cheung, M.; Wittung-Stafshede, P. Crowded, cell-like environment induces shape changes in aspherical protein. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11754−11759.
(33) Griffin, M.; Friedel, M.; Shea, J. Effects of frustration, confinement and surface interactions on the dimerization of an off-lattice beta-barrel protein. *J. Chem. Phys.* **2005**, *123*, 174707.
(34) Neri, M.; Baaden, M.; Carnevale, V.; Anselmi, C.; Maritan, A.; Carloni, P. Microseconds dynamics simulations of the outer membrane Protease T. *Biophys. J.* **2008**, *94*, 71−78.
(35) Hyeon, C.; Onuchic, J. Internal strain regulates the nucleotide binding site of the kinesin leading head. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2175−2180.
(36) Isralewitz, B.; Gao, M.; Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224−230.
(37) Jin, M.; Andricioaei, I.; Springer, T. A. Conversion between three conformational states of integrin I domains with a C-terminal pull spring studied with molecular dynamics. *Structure* **2004**, *12*, 2137−2147.
(38) Li, M.; Kouza, M. Dependence of protein mechanical unfolding pathways on pulling speeds. *J. Chem. Phys.* **2009**, *130*, 145102.
(39) Tian, P.; Andricioaei, I. Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mithochondrial pore. *J. Mol. Biol.* **2005**, *350*, 1017−1034.
(40) Pellegrini, F.; Roman, J. SCOTCH: a software package for static mapping by dual recursive bipartitioning of process and architecture graphs. *High-Performance Computing and Networking*; Lecture notes in Computer Science; Springer, 1996; Vol. *1067*, pp 493−498.
(41) Aleksiev, T.; Potestio, R.; Cozzini, S.; Micheletti, C. PiSQRD: a web server for decomposing proteins into quasi-rigid dynamical domains. *Bioinformatics* **2009**, *25*, 2743−2744.
(42) Micheletti, C.; Carloni, P.; Maritan, A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. *Proteins: Struc., Funct., Bioinfo.* **2004**, *55*, 635−645.
(43) Jacobs, D.; Hendrickson, B. An algorithm for two-dimensional rigidity percolation: the pebble game. *J. Comput. Phys.* **1997**, *137*, 346−365.
(44) Vriend, G. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **1990**, *8*, 52−56.
(45) Chinappi, M.; Cecconi, F.; Casciola, C. M. Computational analysis of maltose binding protein translocation. *Philos. Mag.* **2011**, *91*, 2034−2048.
(46) Padding, J.; Louis, A. Hydrodynamic interactions and Brownian forces in colloidal suspensions: Coarse-graining over time and length scales. *Phys. Rev. E* **2006**, *74*, 031402.
(47) Honeycutt, R. L. Stochastic runge-kutta algorithms. I. White noise. *Phys. Rev. A* **1992**, *45*, 600−603.
(48) Quiocho, F. A.; Spurlino, J. C.; Rodseth, L. E. Extensive features of tight oligosaccharide binding revealed in highresolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **1997**, *5*, 997−1015.
(49) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33−38.
(50) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(51) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(52) Mitternacht, S.; Luccioli, S.; Torcini, A.; Imparato, A.; Irbäck, A. Changing the mechanical unfolding pathway of FnIII10 by tuning the pulling strength. *Biophys. J.* **2009**, *96*, 429−441.
(53) MacKerell, A., Jr; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(54) Nummela, J.; Andricioaei, I. Exact low-force kinetics from high-force single-molecule unfolding events. *Biophysical Journal* **2007**, *93*, 3373−3381.
(55) Veretnik, S.; Gu, J.; Wodak, S. Identifying structural domains in proteins. In *Structural Bioinformatics*, second ed.; Gu, J., Bourne, P., Eds.; Wiley & Sons Inc, 2009.