

# Thermal versus Mechanical Unfolding in a Model Protein

Rafael Tapia-Rojo,<sup>1, a)</sup> Juan J. Mazo,<sup>2</sup> and Fernando Falo<sup>3, b)</sup>

<sup>1)</sup>*Instituto de Biocomputación y Física de Sistemas Complejos and Departamento de Física de la Materia Condensada, Universidad de Zaragoza, 50009 Zaragoza, Spain.*

<sup>2)</sup>*Instituto de Ciencia de Materiales de Aragón and Departamento de Física de la Materia Condensada, CSIC-Universidad de Zaragoza, 50009 Zaragoza, Spain*

<sup>3)</sup>*Instituto de Biocomputación y Física de Sistemas Complejos and Departamento de Física de la Materia Condensada, Universidad de Zaragoza, 50009 Zaragoza, Spain*

(Dated: 23 October 2019)

Force spectroscopy techniques are often used to learn about the free energy landscape of single biomolecules, typically by recovering free energy quantities that, extrapolated to zero force, are compared to those measured in bulk experiments. However, it is not always clear how the information obtained from a mechanically perturbed system can be related to that obtained using other denaturants, since tensioned molecules unfold and refold along a reaction coordinate imposed by the force, which is unlikely meaningful in its absence. Here, we explore this dichotomy by investigating the unfolding landscape of a model protein, which is first unfolded mechanically through typical force spectroscopy-like protocols, and next thermally. When unfolded by non-equilibrium force extension and constant force protocols, we recover a simple two-barrier landscape, as the protein reaches the extended conformation through a metastable intermediate. Interestingly, folding-unfolding equilibrium simulations at low forces suggested a totally different scenario, where this metastable state plays little role in the unfolding mechanism, and the protein unfolds through two competing pathways<sup>27</sup>. Finally, we use Markov state models to describe the configurational space of the unperturbed protein close to the critical temperature. The thermal dynamics is well understood by a one-dimensional landscape along an appropriate reaction coordinate, however very different from the mechanical picture. In this sense, in our protein model the mechanical and thermal descriptions provide incompatible views of the folding/unfolding landscape of the system, and the estimated quantities to zero force result hard to interpret.

## I. INTRODUCTION

Force spectroscopy techniques have provided a wealth of high resolution data, which resulted in a great step forward towards the understanding of how proteins fold<sup>1,2</sup>. Following the classical meaning of spectroscopy, a biomolecule is perturbed by an external mechanical bias that induces the unfolding transition, and from which information from the biomolecule folding landscape can be inferred. For example, in a force extension protocol, a probe such as an AFM cantilever is retracted at constant velocity, applying an increasing force that eventually unfolds the protein<sup>3,4</sup>. From the relationship between the unfolding force and the pulling velocity, we can learn about the mechanical stability, or the unfolding pathway of the subject molecule, including how it is modulated by physiological factors, such as disulfide bonds<sup>5</sup>, or chaperones<sup>6,7</sup>. In addition, the implementation of force-clamp techniques allowed to apply constant forces to individual proteins and record the unfolding time of the perturbed system<sup>8</sup>, but also to carry out equilibrium folding/unfolding experiments and explore in great detail the folding pathways<sup>9,10</sup>, or rare events which might appear over long timescales<sup>11–13</sup>.

In this context, many theoretical models have been developed to provide tools to analyze the experimental data and obtain physical quantities about the system. Roughly, we can divide such theoretical efforts into those devoted to the kinetics, and those to the equilibrium properties. Bell-Evans model provided a first phenomenological framework which allowed to recover the distance to the transition state and the unfolding rate at zero force from force-extension and constant force experiments<sup>14,15</sup>. Later refinements of this theory allowed to recover also the height of the free energy barrier at zero force<sup>16–18</sup>. On the other hand, the famed Jarzynski equality provides the estimation of equilibrium free energy quantities<sup>19</sup>, and even allows to reconstruct the full free energy profile from such non-equilibrium work measurements<sup>20</sup>. Importantly, in both cases we obtain the estimation of free energy magnitudes extrapolated to zero force, and thus it appears tempting to compare them directly with those obtained from biochemical experiments, such as thermal or chemical denaturation<sup>3,21,22</sup>.

However, an evident question arises in this context; how meaningful is such comparison? When folding and unfolding mechanically, the pulling force imposes a very specific reaction coordinate, which is the pulling vector. As protein folding is known to be a highly complex and multidimensional process, it is unlikely that the distance between the N and C-terminal of a protein provides a relevant order parameter to describe protein folding in the absence of force. Even more, while at very high forces a protein will unfold along the pulling directions, this is not clear to be the case at low forces, where orthogonal de-

<sup>a)</sup>Electronic mail: [Rafa.T.Rojo@gmail.com](mailto:Rafa.T.Rojo@gmail.com). Current Address: Department of Biological Sciences, Columbia University, New York, NY 10027, USA

<sup>b)</sup>Electronic mail: [fff@unizar.es](mailto:fff@unizar.es)

degrees of freedom might be of relevance, compromising the use of the aforementioned theories to explain the data.

Here, we explore these questions by analyzing in high detail the unfolding landscape of the BPN<sub>46</sub> model protein when denatured by temperature and under a mechanical bias. The BPN<sub>46</sub> protein is a highly studied non-Go protein model, due to its rich and complex landscape. It folds spontaneously into a  $\beta$  barrel-like structure, that shows frustration in the specific arrangement of its  $\beta$ -sheets, having also intermediate conformations and a non-trivial folding mechanism when subject to force<sup>23–27</sup>. Inspired by typical force spectroscopy protocols, we first unfold the protein mechanically using non-equilibrium force extension and constant force modes. These protocols measure the unfolding forces and the unfolding rates as a function of the pulling velocity and the pulling force, respectively. By analyzing this dependence, the extrapolated to zero force distance to the transition state and the height of the free energy barrier can be estimated, assuming a simple one-dimensional pathway<sup>16–18</sup>. Furthermore, the extended Jarzynski equality can be used to reconstruct the zero-force free energy profile by sampling the individual force extension trajectories<sup>20</sup>. Our data shows that the protein unfolds by surmounting two distinct free energy barriers, first reaching a mechanical intermediate conformation, previously identified<sup>26,27</sup>. Using the lowest pulling velocity to reconstruct the free energy profile, we obtain a similar unfolding picture, but with lower free energy barriers, likely to the multiple unfolding pathways that compete at very low forces, as previously reported<sup>27</sup>.

In order to explore the folding landscape in the absence of force, we carry out equilibrium simulations in the vicinity of the critical temperature. By using Markov state models and transition state theory, we obtain a detailed description of the conformation landscape of the system, and the unfolding pathway. While we can relate structurally those states excited mechanically and thermally, their role in the protein dynamics seems to be completely different. The thermal unfolding mechanism can be fairly well described as a one-directional pathway with two well-defined intermediate states, different from those observed in the mechanical picture. While our conclusions are specific to the BPN<sub>46</sub> protein model, which has some non-standard features for a simple protein folding model, our data exemplifies some of the problems that might arise when directly comparing data obtained from different denaturants. Protein folding in presence and absence of force are very different processes; hence, it is unclear how extrapolations to zero force inform about a folding transition in the absence of force.

## II. MODEL AND SIMULATION METHODS

The BPN<sub>46</sub> model is a coarsed-grained non-Go protein model<sup>23–26,57</sup>. It has a 46-residue sequence where each residue is represented as a “colored” particle, being either

hydrophobic, hydrophilic, or neutral. The Hamiltonian of the model is defined by four potential terms: a stiff nearest-neighbor harmonic potential, a three-body bending interaction, a four-body dihedral interaction and a sequence-dependent Lennard-Jones potential. This latter term contains the sequence dependency, since hydrophobic residues attract to each other while every other pair see a short-range repulsive potential.

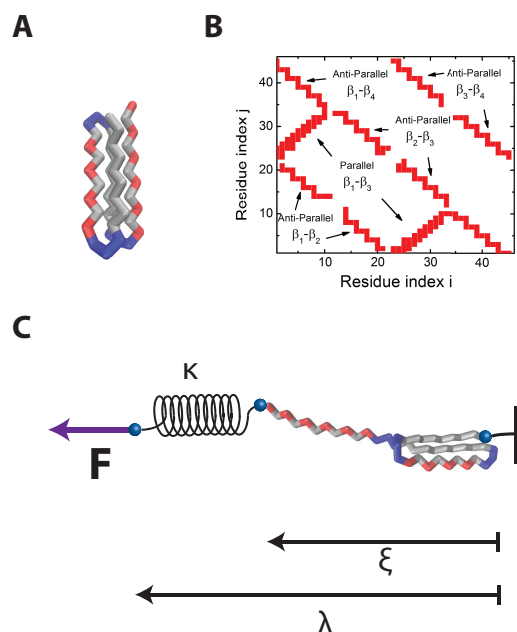


FIG. 1. The BPN<sub>46</sub> model protein. (A) Representation of the native structure of the BPN<sub>46</sub> model protein. Hydrophobic residues (B) are shown in grey, neutral (N) in blue and hydrophilic (P) in red. (B) Map of native contacts of the BPN<sub>46</sub> model protein. It has a  $\beta$  barrel structure arranged in 3 antiparallel and one parallel  $\beta$  sheets. (C) Configuration for the pulling simulations. The protein is fixed from one end while from the other end is attached to a linear spring of constant  $\kappa$ . Two different pulling protocols are carried out. In the force extension one, the  $\lambda$  coordinate is linearly increased  $\lambda = vt$ , where  $v$  is the pulling velocity. In the constant force, the spring is displaced so that a constant force is applied between the ends of the protein. In contrast with  $\lambda$  which is a control parameter,  $\xi$  represents the end to end distance of the protein and is a stochastic magnitude.

The BPN<sub>46</sub> folds spontaneously into a stable native structure as a  $\beta$ -barrel constituted by four  $\beta$ -strands and three neutral turns (see Fig. 1A), mimicking the fold of other well-known domains. However, potential energy analysis of the protein conformational landscape have revealed a multiplicity of structurally similar ground states, separated by large barriers<sup>25</sup>. The structure is held together by the interaction between strands  $\beta_1$  and  $\beta_3$ , formed by just hydrophobic residues. Figure 1B shows the map of native contacts. The hydrophobic strands  $\beta_1$  and  $\beta_3$  run in a parallel disposition, while strands  $\beta_1$ - $\beta_2$ ,  $\beta_2$ - $\beta_3$  and  $\beta_3$ - $\beta_4$  run antiparallel between them.

The behavior of this model protein as a function of temperature and force was reported before<sup>24,26–28</sup>. It ex-

hibits a unfolding transition, with a well-defined peak on its heat capacity at a temperature  $T_c = 0.9kT$ .

In all simulations, we generate stochastic trajectories by integrating numerically the Langevin equation of motion for the 46 residues,

$$m\ddot{\mathbf{r}}_i = -\gamma\dot{\mathbf{r}}_i - \nabla_i \mathbf{V}_{\text{BPN}} + \eta_i, \quad (1)$$

where  $m$  is the unitary mass of each residue,  $\gamma$  the friction coefficient,  $V_{\text{BPN}}$  the potential defined for the BPN<sub>46</sub> model (see App. A), and  $\eta_i$  a gaussian white noise term of zero average, which fulfills the fluctuation-dissipation theorem  $\langle \eta_i \eta_j \rangle = 2kT\gamma\delta(t - t')\delta_{ij}$ .

In the mechanical simulations, we carry out non-equilibrium unfolding by fixing the N-terminal of the protein and attaching a linear spring of constant  $\kappa$  to the C-terminal, used to apply force (see Fig. 1C). Force extension trajectories are generated by displacing the spring at a constant velocity  $v$ , so that the control parameter  $\lambda$  increases linearly  $\lambda = vt$ . Constant force trajectories are generated by suddenly displacing the spring by an amount  $\Delta x$  so that it generates a constant force  $F = \kappa\Delta x$ . In all mechanical simulations, we work at a temperature of  $0.55T_c$ . When working at this temperature the protein unfolds mechanically at a force of  $F_U \approx 20$  pN. Here, we will study nonequilibrium unfolding trajectories at forces above the unfolding force. A conceptually different case is that of the folding-unfolding equilibrium dynamics at a force below the unfolding force, studied in detail in<sup>27</sup> (there,  $T = 0.55T_c$  and  $F = 0.8F_U$ ).

Equilibrium thermal folding-unfolding simulations are carried out at the vicinity of the transition temperature ( $1.1 T_c$ ), to ensure the system visits the maximum number of configurations.

### III. MECHANICAL UNFOLDING

#### A. Analysis Methods

*a. Force Extension:* In force extension trajectories, the experimental output is the force as a function of the pulling coordinate  $\lambda = \xi + F/\kappa$ . Molecular transitions are identified as rupture peaks characterized by the rupture force, which increases with the loading rate  $r_f = v\kappa$ . Bells-Evans phenomenological model predicts a logarithmic dependence of the average rupture force with the loading rate  $\langle f^* \rangle \sim \log(r_f)^{14,15}$ . Extensions of this model predict  $\langle f^* \rangle \sim \log(r_f)^\nu$ , where  $\nu$  is an exponent that depends on the potential shape chosen to model the irreversible rupture process:  $\nu = 2/3$  for a linear-cubic,  $\nu = 1/2$  for a parabolic-cusp and  $\nu = 1$  for a linear potential, which recovers Bell-Evans model<sup>18</sup>. Generally, it can be argued that the linear-cubic landscape is more appropriate, as close to the transition any analytic potential tilted by a pulling force can be approximated by a linear-cubic potential. By fitting the average rupture force  $\langle f^* \rangle$

to the loading rate, the free energy barrier height  $\Delta G^\ddagger$  and position  $x^\ddagger$ , and the intrinsic rate  $k_0$  can be obtained as:

$$\langle f^* \rangle = \frac{\Delta G^\ddagger}{\nu x^\ddagger} \left[ 1 - \left( -\frac{kT}{\Delta G^\ddagger} \log \frac{r_f x^\ddagger}{1.78k_0 e^{\Delta G^\ddagger/kT}} \right)^\nu \right]. \quad (2)$$

*b. Constant Force:* In constant force trajectories, the experimental output is the end-to-end distance  $\xi$  as a function of the simulation time. Molecular transitions are identified as discrete increases in the extension, characterized by the rate of rupture  $k$ , which increases with the pulling force. In a similar way to what described above, Bell-Evans model predicts an exponential dependence of the rate with the pulling force. Extensions of the model predict a more complex behavior, and fitting  $k(F)$  allows recovering the free energy barrier height  $\Delta G_i^\ddagger$  and position  $x_i^\ddagger$ , and the intrinsic rate  $k_0$ :

$$k(F) = k_0 \left( 1 - \frac{\nu F x^\ddagger}{\Delta G^\ddagger} \right)^{1/\nu-1} e^{\frac{\Delta G^\ddagger}{kT} [1 - (1 - \frac{\nu F x^\ddagger}{\Delta G^\ddagger})^{1/\nu}]}. \quad (3)$$

The rate of rupture can be estimated by averaging all rupture traces and fitting to a single exponential or multiple exponentials, if more than a kinetic process is involved.

*c. Reconstruction of the free energy profile:* From non-equilibrium pulling experiments, the free energy profile can be reconstructed by using the extended Jarzynski expression<sup>20</sup>:

$$G_0(\xi_t) = -kT \log \langle \delta(\xi - \xi_t) e^{-\Delta w_t/kT} \rangle \quad (4)$$

where  $G_0(\xi)$  is the free energy profile at zero force along the pulling coordinate  $\xi$ , and  $\Delta w_t$  the difference between the external work done on the system, and the biasing potential,  $\Delta w_t = -\kappa v \int_0^t dt' (\xi(\mathbf{r}_{t'}) - \xi_0 - vt')$ . In this sense, we take time slices and average over different trajectories the work necessary to take the system there, and estimate the free energy at that time slice.

#### B. Results

*a. Force extension:* We carry out force extension simulations at pulling velocities between 0.002 and 0.2 nm/ $\tau$ <sup>29</sup>, recording the dependence of the force with the pulling coordinate  $\lambda$ . Figure 2A shows typical force extension trajectories for  $v = 0.004$  nm/ $\tau$ . Light grey lines show three individual realizations, while the black curve represents the average over a total of 100 trajectories. The protein extends through two different transitions, as every trajectory shows two peaks. We identify in the average a first broad low force transition which leads to  $\lambda \sim 11$  nm, and a second high force one leading a state with  $\lambda \sim 22$  nm. We can relate the first state to the half-stretched conformation (HS), while the second to the

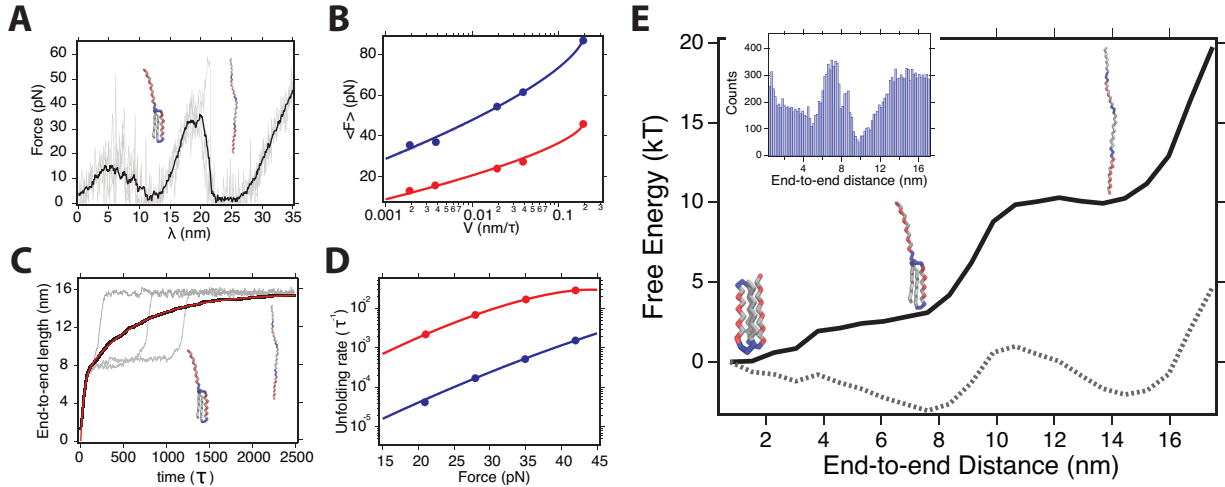


FIG. 2. Non equilibrium mechanical unfolding simulation results (in all cases  $T = 0.55T_c$ ). (A) Typical force extension trajectories at  $v = 0.004 \text{ nm}/\tau$ . Grey lines show individual realizations, and black line the average unfolding trajectory over 100 realizations. The mechanical unfolding in the force extension protocol occurs through two well distinguished transitions, a first low force one at  $\lambda \sim 5 \text{ nm}$  and a second high force one at  $\lambda \sim 17 \text{ nm}$ . (B) Average unfolding force as a function of the pulling velocity for the first transition (red) and second transition (blue). Data points are fit to the Eq. 2 with  $\nu = 2/3$ , obtaining  $\Delta G_1^\ddagger = 7.14 \pm 0.51kT$ ,  $x_1^\ddagger = 0.96 \pm 0.11\text{nm}$ ; and  $\Delta G_2^\ddagger = 10.00 \pm 1.00kT$ ,  $x_2^\ddagger = 0.70 \pm 0.12\text{nm}$ . (C) Constant force trajectories at 42 pN. Grey lines correspond to individual realizations, and black line is the average over 100 realizations. Analogously to force extension experiments, unfolding occurs through two subsequent transitions, one yielding to an extension of  $\sim 8\text{nm}$  and the second to  $\sim 16\text{nm}$ . Red curve is a fit to a two-exponential model, yielding the microscopic rates for the two kinetic transitions. (D) Unfolding rates as a function of the pulling force for the first (red) and second (blue) transitions. Solid lines are fits to Eq. 3 with  $\nu = 2/3$ , obtaining  $\Delta G_1^\ddagger = 8.54 \pm 0.02kT$ ,  $x_1^\ddagger = 1.05 \pm 0.10\text{nm}$ ; and  $\Delta G_2^\ddagger = 12.02 \pm 0.01kT$ ,  $x_2^\ddagger = 0.89 \pm 0.10\text{nm}$ . (E) Free energy profile along the end-to-end distance reconstructed from force-extension trajectories at  $v = 0.002 \text{ nm}/\tau$ , using the extended Jarzynski equality. The profile yields a similar picture, where the unfolding process takes place through a mechanical intermediate with an end-to-end extension of  $\sim 7 \text{ nm}$ —HS conformation. However, the free-energy barrier estimated to reach the intermediate is lower than that using the complete pulling rate range, likely since, due to the low forces involved, multiple pathways are involved, as previously reported. (Dotted line) Tilted landscape under a force of 18 pN. (Inset) Residence time histogram along the end-to-end distance, from which the profile is built.

fully extended conformation, following the notation previously established<sup>27</sup>.

Figure 2B shows the rupture forces fitted to Eq. 2, using  $\nu = 2/3$  where red corresponds to the first event and blue to the second one. Thus, according to this description, there is a first barrier characterized by  $\Delta G_1^\ddagger = 7.14 \pm 0.51kT$  and  $x_1^\ddagger = 0.96 \pm 0.11 \text{ nm}$ , and a second one with  $\Delta G_2^\ddagger = 10.00 \pm 1.00kT$  and  $x_2^\ddagger = 0.70 \pm 0.12 \text{ nm}$ .

*b. Constant force:* We carry out constant force simulations at forces ranging from 20 to 42 pN. Figure 2C shows constant force trajectories where the end-to-end distance  $\xi$  is plotted against the time for a pulling force of 42 pN. Light grey traces correspond to three individual realizations, while the black trace is the average trajectory obtained from averaging 100 realizations. Similarly to the force extension protocol, unfolding occurs sequentially through a mechanical intermediate, with an end-to-end distance of,  $\xi \sim 8 \text{ nm}$ , in accord to the extension of two  $\beta$  strands. As unfolding occurs through two kinetic processes, we represent the average extension as a double exponential  $\langle \xi \rangle(t) = \xi_1(1 - \exp(-k_1 t)) + \xi_2(1 - \exp(-k_2 t))$ , where  $\xi_1$  is the extension of the mechanical intermediate and  $\xi_2$  the length increment to the unfolded states, being  $k_1$  and  $k_2$  the kinetic rates of the two pro-

cesses, respectively.

Unfolding rates obtained at different pulling forces are fitted to Eq. 3, with  $\nu = 2/3$ , where red corresponds to the first barrier and blue to the second one. We obtain  $\Delta G_1^\ddagger = 8.54 \pm 0.02kT$ ,  $x_1^\ddagger = 1.05 \pm 0.10 \text{ nm}$  for the first barrier while  $\Delta G_2^\ddagger = 12.02 \pm 0.01kT$ , and  $x_2^\ddagger = 0.89 \pm 0.10 \text{ nm}$  for the second one. The values are in agreement with those obtained for the force extension protocol.

*c. Reconstruction of the free energy profile:* Finally, we use the force-extensions trajectories to reconstruct the free energy profile along  $\xi$  by using the extended Jarzynski equation (Eq. 4)<sup>20</sup>. Here, taking  $\xi = 0$  as the reference distance, the free energy at different values of  $\xi$  is estimated from the exponential average of the non-equilibrium work required to reach that distance. Jarzynski equality has well-known convergence problems, in particular when the data used is very far from equilibrium—the number of trajectories needed for convergence increase with the exponential of the dissipated work. Hence, we use the slowest pulling speed ( $v = 0.002 \text{ nm}/\tau$ ) for the reconstruction. This landscape is an estimation of the free energy for the unperturbed system, and thus should be directly comparable to the free energy barriers calculated through Eqs. 2 and 3.



Figure 2E shows the reconstructed profile (solid line). The unfolding mechanism suggested here agrees with that depicted by the force-extension and constant force simulations, since the unfolded state is reached through a mechanical intermediate—HS configuration. These three energy minima are more evident when tilting the landscape with a pulling force (dotted line). The extensions of the three involved states match the plateaus at constant force (Fig. 2C), with the native state with a  $\sim 0$  nm end-to-end length, the HS with  $\sim 7$  nm, and the extended state at  $\sim 14$  nm. The unfolded state is reached from the HS state after surmounting a  $\sim 8$  kT, similar to that estimated in the fits of Figs. 2 B and D. By contrast, the native and HS state appear separated by an almost negligible barrier, while the fits to Eq. 2 yielded a barrier of  $\sim 7$  kT. This discrepancy arises likely due to the very low forces at which the transitions occur at this low pulling velocity. As reported at low constant forces<sup>27</sup>, the conformational landscape of the BPN<sub>46</sub> protein under force is complex, and rich dynamics appear between the native and HS state, with several other metastable states separated by very low barriers. Hence, it is likely that at very low pulling speeds, these quasi-equilibrium transitions take place and average out the free energy profile in the Jarzynski reconstruction, yielding to a lower free energy barrier. Indeed, while the average force-extension trajectory at  $v = 0.002$  nm/ $\tau$  shows two clear peaks at forces of  $\sim 15$  and  $\sim 35$  pN, inspection of individual trajectories reveal complex patterns of unfolding peaks between 0 and 8 nm, suggesting complex dynamics, compatible with those that appear at a constant low force<sup>27</sup>.

## IV. THERMAL UNFOLDING

### A. Thermal unfolding simulation details.

We carry out five simulations with a duration of  $t = 10^9 \tau$  each, close to the critical temperature  $T = 1.1T_c$ . Each simulation is previously thermalized during  $t = 10^4 \tau$ , to randomize the initial conditions of each trajectory. The simulation time is sufficient for the system to adequately explore its conformational landscape in equilibrium, given the high temperature, which leads to very fast dynamics between all visited states. Equilibrium is later checked on the Markov network, showing that detailed balance holds.

### B. Analysis Methods: Dimension reduction, Markov state model and transition-path theory

In order to analyze the equilibrium thermal folding-unfolding process, we represent the configurational space of the BPN<sub>46</sub> protein at  $T_c$  as a Markov state network<sup>30</sup>. In this representation, the free energy landscape is shown as a complex network, where the nodes correspond to macrostates (free energy basins) obtained by some clus-

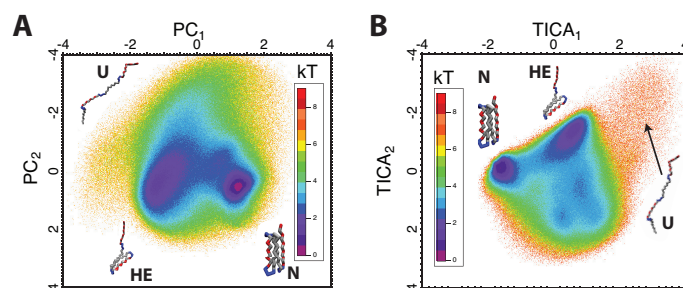


FIG. 3. Comparison between PCA and TICA as dimension-reduction methods. (A) Two-dimensional landscape projected along the first two principal components. (B) Two-dimensional landscape projected along the first two TICA components. Compared to PCA, TICA allows to obtain better defined states that facilitate the discretization and clustering of the trajectories, and the calculation of the Markov State network.

tering method, while the edges—weighted and directed—relate to the kinetic transitions between such states. In a nutshell, building a Markov state network starts by a geometrical discretization of the configurational landscape to build a first microstate network, typically with thousands of nodes, and hence, hard to interpret. Next, this network is coarse-grained by some lumping algorithm, that clusters those states kinetically related, to end up with a smaller, more significant representation of the system's energy landscape. In our case, we use the SSD algorithm as a clustering method<sup>27,31,32</sup>.

The full configurational landscape can have typically several hundreds of dimensions, and a direct discretization of it is a futile effort; hence, it is useful to reduce first the dimensionality of the system to a few, meaningful coordinates. These coordinates should yet be able to identify the large and slow conformational transitions, which would allow to rule out the abundant and meaningless fast fluctuations. Principal Component Analysis (PCA) is often used as a dimension reduction method<sup>27,31–33</sup>, since it identifies the coordinates that contain the largest structural fluctuations about the average conformation. However, it has been recently demonstrated that Time-structure Independent Component Analysis (TICA) is the optimal method for identifying the ‘slow’ order parameters, since it takes into account, not only the spatial variation between structures, but also the timescales over which they occur<sup>34,35</sup>. In our previous work<sup>27</sup>, PCA was sufficient to separate the main macrostates occupied by the system along its dynamics. However, those simulations were carried out at low temperature, and under force; hence, the dynamics were intrinsically slow, given the large free energy barriers that separated the free energy basins of the system. Here, we carry out simulations close to the critical temperature in order to populate the unfolded state without the need of an external bias. Therefore, the dynamics and transitions that define the system dynamics are much abundant and fast, and PCA is a method likely to fail in separating the main

conformational changes of the system.

Figure 3 compares the projection of the simulated dynamics along the first two PCs (A), and the first two TICA components (B). These two components capture the three main conformational states of the system, the native (N), the half-extended (HE) and the unfolded state (U). However, it is very evident by comparing both projections, that TICA does a much better job in defining the energy basins and free energy barriers that separate them. The N and HE states are separated by a  $\sim 2.5$  kT barrier in the PCA projection, while this barrier is of  $\sim 4$  kT in the case of TICA. The unfolded conformation—with a low population—only appears as a spatially spread state (given its large entropy) in the TICA projection, while its representation in the PCA projection is much dimmer. Finally, along the second TICA coordinate, the HE shows multiple states associated to similar conformations separated by small energy barriers, that are not separated with PCA. Hence, this demonstrates the need to use TICA to find meaningful order parameters to build the Markov network.

We use the first three TICA components as our configurational space, and discretize them into 30 bins, obtaining a microstate network with 6657 nodes connected through 228682 links. We then lump the microstates onto kinetically significant macrostates by applying the SSD algorithm. We obtain a first ‘raw’ macrostate network with 45 basins of attraction of macrostates. We refine such network, eliminating basins with an occupation  $\pi_i < 10^{-4}$ , to avoid extremely rare or pathological states. The final macrostate network is made up of 21 nodes, and we take it as the representation of the free energy landscape of the system.

From the equilibrium Markov state network, we can calculate precisely the unfolding pathways by applying transition-path theory<sup>36–38</sup>. Briefly, we start defining the set of native conformations (N) and the set of unfolded conformation (U). Next, we will rank all other states, as intermediates I between N and U, depending on how close they are to the unfolded conformation in terms of unfolding pathways. To this aim, we calculate the committor probability (or unfolding probability)  $q_i^+$  of each state (being  $q_N^+ = 0$  and  $q_U^+ = 1$  by definition) as the solution for the system of equations:

$$-q_i^+ + \sum_{k \in I} T_{ik} q_k^+ = - \sum_{k \in B} T_{ik}, \quad (5)$$

where  $T_{ik}$  is the rate matrix of the Markov network. Intuitively,  $q_i^+$  represents the probability that, state  $i$  reaches the unfolded conformation U before reaching back the native state N. Then, we compute the unidirectional flux to the unfolded state as  $f_{ij}^+ = \max[0, f_{ij} - f_{ji}]$ , where  $f_{ij} = \pi_i(1 - q_i^+)T_{ij}q_j^+$ .  $f_{ij}^+$  defines a network of fluxes from the native N to the unfolded U configurations, from which we extract the individual unfolding pathways.

## C. Results

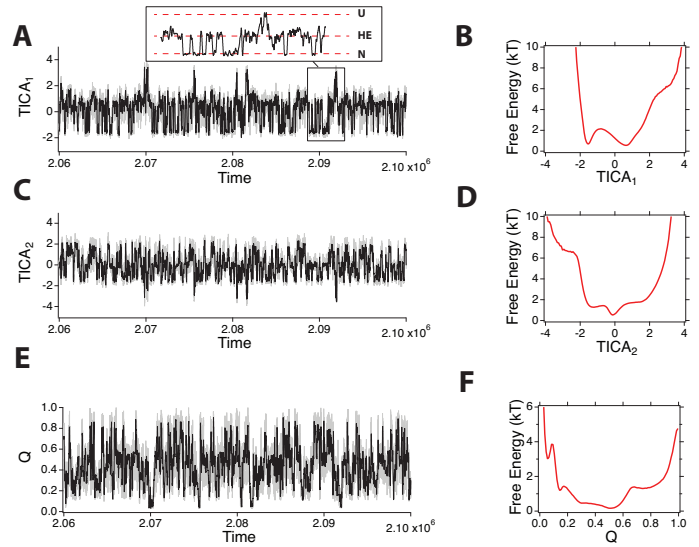


FIG. 4. Fragment of the simulated trajectories along different order parameters, with their projected one-dimensional landscape. (A) Trajectory along the first TICA coordinate. (Inset) detail where the three main states are identified. (B) One-dimensional landscape along the first TICA coordinate. (C) Trajectory along the second TICA coordinate. (D) One-dimensional landscape along the second TICA coordinate. (E) Trajectory along the fraction of native contacts. (F) One-dimensional landscape along the fraction of native contacts.

*a. One dimensional free energy landscapes.* Figure 4 shows a fragment of the simulated trajectories projected along three different order parameters, the first and second TICA components (Figs. 4A and C), and the fraction of native contacts (Fig. 4E). These fragments represent only 1% of the total simulated time, which demonstrates that the system explores its landscape in a much faster timescale than the simulated time window. Right panels are the one-dimensional landscape as obtained from each coordinate. The first TICA component separates the transitions between the three major states, the native (N), half-extended (HE), and unfolded (U) states. Most of the dynamics correspond to very fast transitions between the (N and (HE) states, with occasional visits to the unfolded state. The (N) state has a narrow free energy basin (Fig. 4B), while the (HE) state appears as shallower basin, likely due to the presence of multiple conformations, that will be resolved in higher order TICA components. This suggests that, even though the native state of the system is frustrated, such frustration is removed at the simulation conditions, likely due to the high temperature that sheds the free energy barriers between the multiple ground states. The second TICA coordinate maintains the transitions to the unfolded state, while it shows a much richer conformational landscape, resolving more subtle structures around the (HE) state. Finally, the fraction of native contacts (Q),

is useful to identify the states populated by the dynamics, since it indicates intuitively the degree of nativeness of the states. However, it does not provide a good reaction coordinate to represent the dynamics, since most minima are shallow basins, separated by small barriers, suggesting that multiple states might be averaged out when projecting onto this coordinate.

*b. The equilibrium Markov state network:* Figure 5A shows the macrostate network, calculated as explained above. The size of the nodes is proportional to their population with a cutoff of  $\pi_i < 10^{-2}$ , below which beads have the same size, for visualization reasons. Links with  $T_{ij} > 10^{-5}\tau^{-1}$  are represented as arrows connecting states. We calculate the fraction of native contacts  $Q$  of each node (see Table 1), which allows us to cluster them into six major regions, represented in different colors: The native state (N) ( $Q \approx 0.8$ ; blue); the half-extended configuration (HE) ( $Q \approx 0.5$ ; red); the collapsed ensemble (C), ( $Q \approx 0.3$ ; green); the first intermediate state ( $I_1$ ) ( $Q \approx 0.2$ ; yellow) the unstructured ensemble (UE) ( $Q \approx 0.2$ ; purple), the second intermediate state ( $I_2$ ) ( $Q \approx 0.1$ ; magenta), and the unfolded ensemble (U) ( $Q \approx 0$ ; black).

When comparing with the unfolding mechanism described for the non-equilibrium pulling simulations, besides the native and unfolded state, we find the HE configuration, which resembles the mechanical intermediate described in the landscape of Fig. 2E. Both the HE (thermal) and HS (mechanical) states maintain the core structure and show strand  $\beta 4$  extended. In the thermal network, the HE conformation has  $\pi_{HE} = 0.53$ , and is heavily connected to the native conformation, with fast transitions between them  $T_{HE,N} \approx 3 \times 10^{-5}\tau^{-1}$ . This is similar to the almost barrierless landscape we encountered in the mechanical description in Fig. 2. Additionally, we identify two states that play a central role in the folding/unfolding dynamics; state  $I_1$ —which maintains the hydrophobic interactions present in the native interactions between  $\beta$  strands 1 and 3, providing 30% of the native contacts—and the state  $I_2$ —which maintains little structural similarity with the native state, since the hairpin formed by the weak interactions between the hydrophobic residues in  $\beta$  strands 1 and 2 is the only motif that survives from the native conformation. These two states are bottlenecks in the pathways connecting the native and unfolded states, and removal of any of them will disconnect both ensembles.

*c. Free energy profile along the committor probability:* The committor probability  $q_i^+$ , or unfolding probability  $p_{\text{unfold}}$ , is often argued to be the appropriate reaction coordinate for describing (un)folding transitions<sup>37</sup>. In this sense, we can project the simulated trajectories onto  $p_{\text{unfold}}$ , and calculate a free energy profile along it.

Figure 5B shows the free energy profile along  $p_{\text{unfold}}$ , where the points correspond to actual free energies estimated by binning  $p_{\text{unfold}}$  with a bin size of 0.05, and the solid line is an interpolation to visualize a smooth landscape. All states belonging to the native (N),

TABLE I. Magnitudes characterizing the states detected in the equilibrium network. We characterize the states with their occupation  $\pi_i$ , fraction of native contacts  $Q$ , unfolding probability, and mean escape time  $t_e$  (in  $\tau$  dimensions).

State	$\pi_i$	$Q$	$p_{\text{unfold}}$	$t_e$ ( $\tau \times 10^3$ )
N	0.291	0.82	0.00	14.30
HE	0.372	0.51	0.00	13.70
C	0.264	0.33	0.00	6.25
$I_1$	0.068	0.23	0.14	3.50
UE	0.002	0.19	0.27	1.53
$I_2$	0.036	0.14	0.46	3.92
U	0.002	0.06	0.95	1.80

half-extended (HE) and collapsed (C) ensembles appear lumped onto the same free energy minimum, with  $p_{\text{unfold}} \sim 0$ . This indicates that the majority of the system dynamics involves internal transitions among these three ensembles, that do not lead to unfolding. The intermediate state ( $I_1$ ) is the first relevant minimum, with  $p_{\text{unfold}} \sim 0.15$ . This procedure allows us to identify the transition state as that with 50 % probability of unfolding, which corresponds to the state labeled as  $I_2$  in the landscape network.

*d. Unfolding pathway:* By applying TST, we can determine the unfolding pathways by converting the Markov state network shown in Fig. 5A into a one-directional flux network that connects the native state (N) with the unfolded ensemble (U). Figure 5C shows the unfolding network, depicted from left to right in terms of increasing  $q^+$ , where arrows connecting states have a thickness which is proportional to their flux  $f_{ij}^+$ , being only those connections with  $f_{ij}^+ > 10^{-7}$  represented for visualization reasons.

As suggested by the landscape network, thermal unfolding occurs in a one-dimensional way, where states  $I_1$  and  $I_2$  are bottleneck nodes, that concentrate the unfolding flux. Most of the dynamics between states N, HE and C correspond to conformational changes that do not drive unfolding. However, we can identify states C2 and C3 (belonging to the collapsed ensemble) as precursors of the intermediate  $I_1$ , that eventually leads to state  $I_2$ , and the unfolded state. In this sense, the flux network suggests that the folding/unfolding equilibrium transitions of the protein follow a one-dimensional dynamics, where  $p_{\text{unfold}}$  works indeed as a good reaction coordinate. This unfolding dynamics, although one dimensional as that described in Fig. 2 when applying large forces, follows a very different pathway from the nonequilibrium mechanical unfolding. The half-stretched conformation—to which HE can be ascribed as the non-tensioned homologous—played a very clear role as a mechanical intermediate, as shown in the free-energy profile reconstruction. However, the one-dimensional landscape here has a very different structure, and the state HE, although kinetically relevant, does not play a significant role in the unfolding

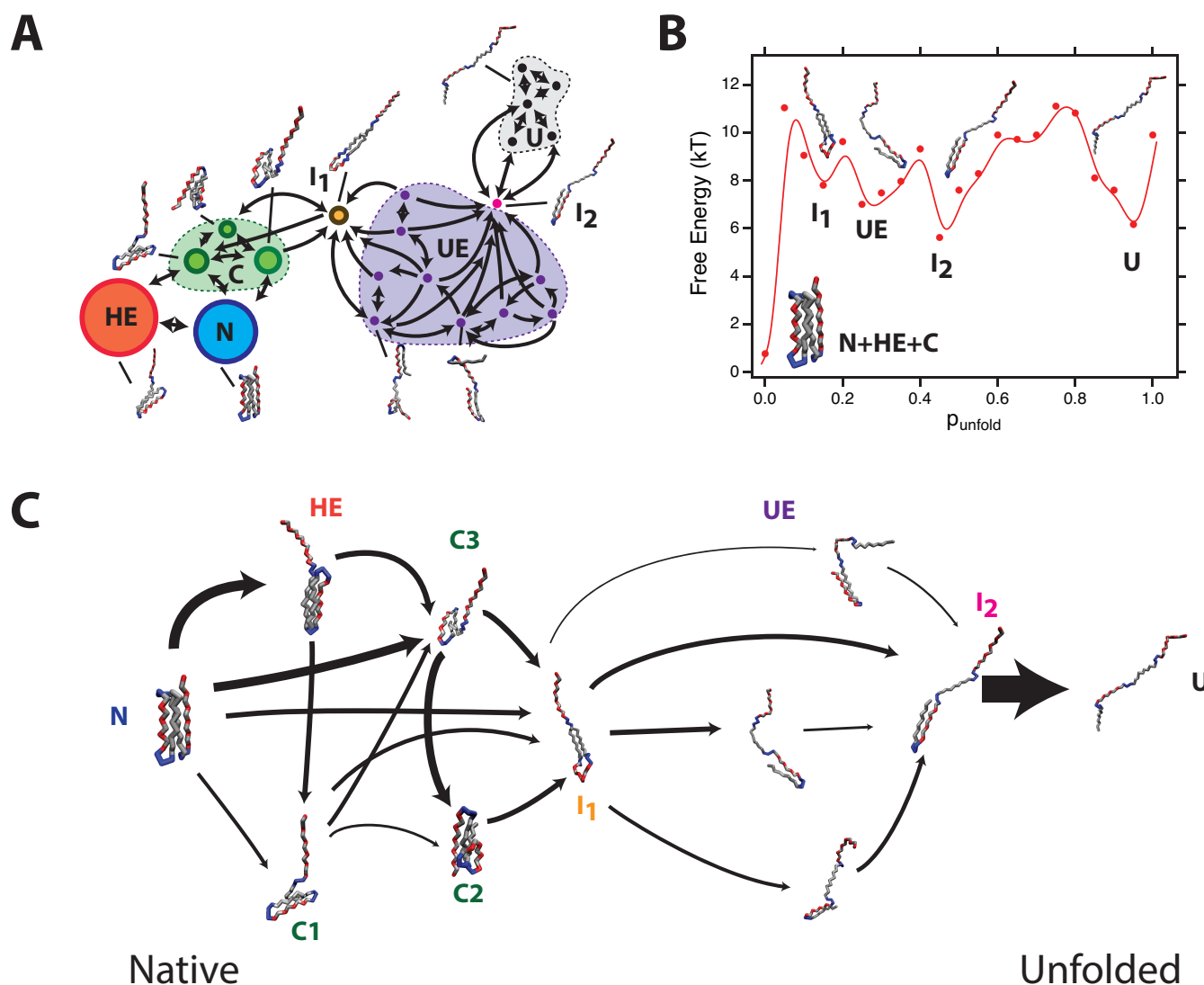


FIG. 5. Markov state model of the BPN<sub>46</sub> model protein close to the critical temperature. (A) Conformational macrostate network at  $T = 1.1T_c$ . Each macrostate is represented as a bead which size is proportional to its population (although states with population  $\pi < 10^{-2}$  are represented as beads of same size). Arrows represent observed transitions between states. Most representative structures are shown next to the associated state. The free energy landscape can be clustered in six major conformational regions, represented by the colors in the beads. (B) Free energy profile along the  $p_{\text{unfold}}$  reaction coordinate. Data points are the clustered histogram from projecting the Markov network onto the reaction coordinate, and the solid line is an interpolation to depict a smooth landscape. (C) Folding pathways as obtained by applying TPT to the macrostate network. Folding/unfolding occurs in a very one-dimensional fashion, being bottleneck states of the unfolding pathway.

pathway. Interestingly, this simple pathway contrasts greatly to the dynamics exhibited by the protein when a small mechanical force is applied. Mechanical forces seemed to excite conformations that do not appear in the absence of force, and thus created a much more complex unfolding dynamics, that could not be described through a one-dimensional landscape<sup>27</sup>.

## V. DISCUSSION

The results presented in this work, together with those previously published<sup>27</sup>, present a threefold vision of the unfolding mechanism of the BPN<sub>46</sub> model protein; while we previously described the equilibrium landscape at low force, here we present the landscape extrapolated to zero force from non-equilibrium pulling simulations—analogous to what often done in force spectroscopy experiments—and an equilibrium landscape obtained from thermal simulations. When comparing these three landscapes, some similarities appear between the



unfolding processes, and, more notably, irreconcilably differences.

The BPN<sub>46</sub> is a simple protein model, that, however, exhibits a complex and rich conformational landscape, including a frustrated ground state, intermediates, or a multidimensional folding mechanism<sup>26,27,39,40</sup>. Therefore, our results must be understood as an exemplification of the dichotomy that might arise when comparing folding mechanisms through different denaturants, such as force and temperature. Interestingly, in every of the unfolding protocols we have used, we recover a similar metastable state, which we dubbed the HS configuration in the mechanical simulations, and HE configuration in the thermal landscape. This state maintains the hydrophobic core between  $\beta$ -strands1-3, with the  $\beta$ -strand4 dislodge from the structure. The prevalence of this conformation in every landscape illustrates a similarity between the landscapes recovered through different techniques. However, the role of this state in the folding/unfolding mechanisms is very different, depending on the protocol we use. In non-equilibrium unfolding simulations, the HS conformation appears as a clear mechanical intermediate, and the pulling trajectories reflect the need to surmount two free energy barriers to reach the stretched state (Fig. 2). These barriers can be estimated using the Dudko-Hummer-Szabo theory<sup>18</sup>, resulting in a  $\sim 7$ -8 kT barrier separating the native and HS conformation. However, in the absence of force, this analogous state appears as a relevant metastable with little role in the unfolding mechanism, similar to what found in equilibrium simulations at very low forces<sup>27</sup>. This indicates that the mechanical regime used to trigger unfolding can also alter the pathway the protein explores, preventing the use of extrapolations to zero-force. Indeed, the reconstruction of the one-dimensional profile using the extended Jarzynski equality reveals a small first barrier, likely due to the competence of several transitions that occur at very low force.

In this sense, the role of mechanical forces in driving protein unfolding must be extrapolated with caution, not just when establishing parallelisms or comparison with unfolding driven by other denaturants (either thermal or chemical), but also when comparing mechanical unfolding under different protocols. Here, we have shown how, surprisingly, the thermal unfolding mechanism of the BPN<sub>46</sub> protein is rather simple; while it exhibits rich dynamics between different metastable states such as the HE or the C ensemble, the protein unfolds in a one dimensional way, with two well defined intermediate states ( $I_1$  and  $I_2$ ). However, when low forces are applied such that we allow equilibrium transitions between the folded and unfolded states—similar to what it is possible to do with force-clamp spectroscopy techniques—the unfolding mechanism is rather complicated. Contrary to what could be intuitively predicted—since the force should impose a reaction coordinate and thus the unfolding mechanism could be easily described by this geometrical coordinate—the unfolding mechanism gets more

complicated, and roughly two major unfolding routes can be identified<sup>27</sup>. When the magnitude of the force is increased so that the equilibrium is shifted to the unfolded conformation, the intuitive behavior arises, and the protein unfolds through the end-to-end reaction coordinate. This implies that the effect of force on the landscape of this protein cannot be modeled over the whole range as a simple linear perturbation with a  $-fx$  term, as commonly done.

Since the advent of force spectroscopy techniques, free energy quantities extracted from pulling experiments have been usually contrasted with those extracted from bulk experiments, such as chemical or thermal denaturation<sup>3,22,41–43</sup>. This comparison implicitly assumes a simple two-state one-dimensional folding landscape, which, while can be reasonable for some systems over a certain scale, should not be generalized. Force is a unique denaturant, since it imposes a topological constraint on the tethered molecule, and defines a very specific reaction coordinate, which might not be relevant in the absence of force. Indeed, mechanical pulling experiments and simulations have been shown to reproduce very different dynamics compared to those exhibited in the absence of force<sup>21,44–48</sup>. For instance, downhill proteins are relevant for folding in solution over a marginal free energy barrier, contradicting the classic two-state picture<sup>49–51</sup>. However, recent force spectroscopy experiments on the gpW downhill protein showed that, under force, this proteins transitions between the folded and unfolded state with two-state dynamics over a barrier of about 2.5 kT<sup>52</sup>. In this sense, recent instrumental development have increased the resolution at which we can interrogate a folding protein, revealing a much more complex scenario than what typically assumed. Protein L, a very classic two-state folder, folds through a ephemeral molten globule-like state that lasts few milliseconds, and therefore is only captured through over very short force quenches<sup>9</sup>. Additionally, the new access to very long timescales, allowed exploring unfolding at low forces, and showing non-monotonic dependences of the unfolding rates with force, which can be interpreted as a multipathway unfolding mechanism<sup>53,54</sup>. Therefore, while our studies here deal with a specific protein model with very particular features, our data exemplifies the dichotomy that can arise, not only when unfolding mechanically and thermally a protein, but also when understanding this mechanism over different force ranges.

In summary, our works provides a practical case of a simple protein that, however, exhibits complex unfolding dynamics that depend strongly on the protocol used to unfold it. Non-equilibrium mechanical protocols as those typically used in single molecule pulling experiments reveal one-dimensional unfolding through a mechanical intermediate, that, however, does not play any role in the unfolding pathway, neither at low or in the absence of force.

## ACKNOWLEDGMENTS

This work was supported by Spanish Ministerio de Economía, Industria y Competitividad (MINECO) Projects No. FIS2014-55867-P and FIS2017-87519-P, co-financed by the Fondo Europeo de Desarrollo Regional (FEDER), and by the Gobierno de Aragón, Grant to the FENOL group, E36\_17R.

## Appendix A: Model parameters and simulation procedures

We use the same model protein employed in<sup>27</sup>. The interaction between residues account for four terms, an stiff harmonic nearest neighbor interaction, a three-body bending interaction, a dihedral four-body interaction, and a sequence-dependent Lennard Jones potential, this is:

$$V_{BLN} = \frac{1}{2}K \sum_{i=1}^{N-1} (r_{i,i+1} - r_0)^2 + \sum_{i=1}^{N-2} [A \cos \theta_i + B \cos 2\theta_i - V_0] + \sum_{i=1}^{N-3} [C_i (1 + \cos \phi_i) + D_i (1 + \cos 3\phi_i)] + \sum_{ij} \epsilon_{ij} \left( \frac{1}{r_{ij}^{12}} - \frac{c_{ij}}{r_{ij}^6} \right),$$

where  $r_{ij}$  is the distance between residues  $i$  and  $j$ ,  $\theta_i$  is the angle between three contiguous residues, and  $\phi$  the dihedral angle. The sequence-independent parameters for the model are  $K = 50$ ,  $r_0 = 1$ ,  $A = 5.118$ ,  $B = 5.308$ ,  $V_0 = -5.295$ , while the sequence-dependent:  $C_i = 0$  and  $D_i = 0.2$  if two or more residues are neutral, or  $C_i = D_i = 1.2$  otherwise; while the Lennard-Jones:  $c_{ij} = 0$ ,  $\epsilon_{ij} = 4$  if either  $i$  or  $j$  are neutral,  $c_{ij} = 1$ , and  $\epsilon_{ij} = 4$  if  $i$  and  $j$  are hydrophobic, and  $c_{ij} = -1$ , and  $\epsilon = 8/3$  otherwise, all adimensional units.

Physical units are estimated in the following way. Length units are recovered assuming that the distance between  $\alpha$  carbons in a protein is 0.38 nm. The energy units are defined assuming the energy of a hydrogen bond  $\epsilon \approx 1.7kT$ , which defines force units as  $\tilde{F} \approx 17.3pN$ . Mass units assume the average mass of an aminoacid  $m_a \approx 3 \times 10^{-22} \text{kg}$ . Thus, time units can be defined, as:  $\tau = \sqrt{m_a r_0^2 / \epsilon_H} \approx 3 \text{ ps}$ .

We carry out all simulations using a self-written code, integrating the overdamped Langevin equations using a stochastic second order Runge-Kutta algorithm<sup>55</sup>. The integration step is  $\Delta t = 0.005\tau$ .

<sup>1</sup>K. C. Neuman, and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods* **5**, 491-505 (2008)

- <sup>2</sup>C. Bustamante, Y. R. Chemla, N. R. Force, and D. Izhaky. Mechanical Processes in Biochemistry. *Ann. Rev. Biochem.* **73**, 705-748 (2004)
- <sup>3</sup>M. Carrión-Vázquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernández. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3694 (1999).
- <sup>4</sup>H. Li, A. F. Oberhauser, S. B. Fowler, J. Clarke and J. M. Fernández. Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6527 (2000).
- <sup>5</sup>P. Kosuri, J. Alegre-Cebollada, J. Feng, A. Kaplan, A. Ingles-Prieto, C. L. Badilla, B. R. Stockwell, J. M. Sanchez-Ruiz, A. Holmgren, and J. M. Fernandez. Protein folding drives disulfide formation. *Cell*. **151**, 794-806 (2012)
- <sup>6</sup>P. Bechtluft, R. G. van Leeuwen, M. Tyreman, D. Tomkiewicz, N. Nouwen, H. L. Tepper, A. J. Driessen, and S. J. Tans. Direct observation of chaperone-induced changes in a protein folding pathway. **30**: 1458-1461 (2007)
- <sup>7</sup>S. Haldar, R. Tapia-Rojo, E. C. Eckels, J. Valle-Orero, and J. M. Fernandez. Trigger factor chaperone acts as a mechanical foldase. *Nature Communications*. **8** 668 (2018)
- <sup>8</sup>J. M. Fernandez, and H. Li. Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein. *Science* **303**, 1674-1678 (2004)
- <sup>9</sup>R. Tapia-Rojo, E. C. Eckels, and J. M. Fernandez. Ephemeral states in protein folding under force captured with a magnetic tweezers design. *Proc. Natl. Acad. Sci. USA* **116**, 7873-7878 (2019)
- <sup>10</sup>K. Neupane, D.A.N. Foster, D. R. Dee, H. Yu, F. Wang, and M. T. Woodside. Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **352** 239-242 (2016)
- <sup>11</sup>I. Popa, J. A. Rivas-Pardo, E. C. Eckels, D. J. Echelman, C. L. Badilla, J. Valle-Orero, and J. M. Fernandez. A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J. Am. Chem. Soc.* **138**, 10546 (2016)
- <sup>12</sup>J. Valle-Orero, J. A. Rivas-Pardo, R. Tapia-Rojo, I. Popa, D. J. Echelman, S. Haldar, and J. M. Fernandez. Mechanical deformation accelerates protein ageing. *Angew. Chem. Int. Ed. Engl.* **56**: 9741-9746 (2017)
- <sup>13</sup>J. Valle-Orero, R. Tapia-Rojo, E. C. Eckels, J. A. Rivas-Pardo, I. Popa, and J. M. Fernández. Proteins Breaking Bad: A Free Energy Perspective. *J. Phys. Chem. Lett.* **8**: 3642-3647 (2017)
- <sup>14</sup>G. I. Bell. Models for the specific adhesion of cells to cells. *Science* **200**, 618 (1978)
- <sup>15</sup>E. Evans, and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72**, 1541. (1997)
- <sup>16</sup>G. Hummer, and A. Szabo. Kinetics from Nonequilibrium Single-Molecule Pulling Experiments. *Biophys. J.* **85**, 5. (2003)
- <sup>17</sup>O. K. Dudko, A. E. Filippoc, J. Klafter, and M. Urbakh. Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proc. Natl. Acad. Sci. USA*, **100**, 11378. (2003)
- <sup>18</sup>O. K. Dudko, G. Hummer, and A. Szabo. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Phys. Rev. Lett.* **96**, 108101- (2006)
- <sup>19</sup>C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78**, 2690. (1997)
- <sup>20</sup>G. Hummer, and A. Szabo, *Proc. Natl. Acad. Sci. USA* **98**, 3658. (2000)
- <sup>21</sup>G. Stirnemann, S. Kang, R. Zhou, and B. Berne. How force unfolding differs from chemical denaturation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3413-3418 (2014)
- <sup>22</sup>L. Sun, J. K. Noel, J. I. Sulkowska, H. Levine, and J. N. Onuchic. Connecting Thermal and Mechanical Protein (Un)folding Landscapes. *Biophys. J.* **107**, 2950-2961 (2014).
- <sup>23</sup>M. A. Miller, and D. J. Wales. Energy landscape of a model protein. *J. Chem. Phys.* **111**, 6610 (1999).
- <sup>24</sup>S. Brown, N. J. Fawzi, T. Head-Gordon. Coarse-grained sequences for protein folding and design. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10712 (2003).

- <sup>25</sup>M. T. Oakley, D. J. Wales, and R. L. Johnston. Energy Landscape and Global Optimization for a Frustrated Model Protein. *J. Phys. Chem. B*. **115**, 11525-29 (2011).
- <sup>26</sup>A. Imparato, S. Luccioli and A. Torcini. Reconstructing the Free-Energy Landscape of a Mechanically Unfolded Model Protein. *Phys. Rev. Lett.* **99**, 168101 (2007).
- <sup>27</sup>R. Tapia-Rojo, S. Arregui, J. J. Mazo, and F. Falo. Mechanical unfolding of a simple model protein goes beyond the reach of one-dimensional descriptions. *J. Chem. Phys.* **141**, 135102. (2014)
- <sup>28</sup>J. D. Honeycutt, and D. Thirumalai. Metastability of the folded states of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3526 (1990)
- <sup>29</sup>The time unit  $\tau$  is usually defined in terms of the mass, distance and energy unit of the system,  $\tau = \sqrt{mr^2/\epsilon}$ . Using the  $C_\alpha$ - $C_\alpha$  distance, the average amino-acid mass and the energy of a hydrogen bond we get  $\tau \sim 3ps$ <sup>27</sup>.
- <sup>30</sup>G. R. Bowman, V. S. Pande and F. Noé (Eds.) *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. Advances in Experimental Medicine and Biology. (2014).
- <sup>31</sup>D. Prada-Gracia, J. Gómez-Gardeñes, P. Echenique, and F. Falo. Exploring the Free Energy Landscape: From Dynamics to Networks and Back. *PLoS Comput. Biol.* **5**, e1000415 (2009).
- <sup>32</sup>R. Tapia-Rojo, J. J. Mazo, J. A. Hernández, M. L. Peleato, M. F. Fillat, and F. Falo. Mesoscopic Model and Free Energy Landscape for Protein-DNA Binding Sites: Analysis of Cyanobacterial Promoters. *PLOS Comput. Biol.* **10**, e1003835 (2014)
- <sup>33</sup>R. Tapia-Rojo, D. Prada-Gracia, J. J. Mazo, and F. Falo. Mesoscopic model for free-energy-landscape analysis of DNA sequences *Phys. Rev. E*. **86**, 021908 (2012).
- <sup>34</sup>G. Pérez-Hernández, F. Paul, T. Giorgino, G. de Fabritiis, and F. Noé. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013).
- <sup>35</sup>C. R. Schwantes, and V. S. Pande, Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **9**, 2000-2009 (2013).
- <sup>36</sup>W. E, and E. Vanden-Eijnden, Towards a Theory of Transition Paths. *J. Stat. Phys.* **123**, 503 (2006)
- <sup>37</sup>F. Noe, C. Schutte, E. Vanden-Eijnden, L. Reich and T.R. Weikl, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19011-19016 (2009)
- <sup>38</sup>P. Metzner, C. Shutte, and E. Vanden-Eijnden. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* **7**, 1192 (2009)
- <sup>39</sup>D. J. Wales, and P. E. J. Dewsbury. Effect of salt bridges on the energy landscape of a model protein. *J. Chem. Phys.* **121**, 10284 (2004)
- <sup>40</sup>D. J. Wales and T. Head-Gordon, Evolution of the Potential Energy Landscape with Static Pulling Force for Two Model Proteins. *J. Phys. Chem. B* **116**, 8394-8411 (2012)
- <sup>41</sup>C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky. Mechanical Processes in Biology. *Annu. Rev. Biochem.* **73**: 705-48 (2004)
- <sup>42</sup>E. J. Guinn, B. Jagannathan and S. Marqusee. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein *Nat. Comm.* **6**, 6861,(2015)
- <sup>43</sup>R. Tapia-Rojo, C. Marcuello, A. Lostao, C. Gomez-Moreno, J. J. Mazo, and F. Falo. A physical picture for mechanical dissociation of biological complexes: from forces to free energies. *Phys. Chem. Chem. Phys.* **19**, 4567 (2017).
- <sup>44</sup>D. de Sancho and R. B. Best. Reconciling Intermediates in Mechanical Unfolding Experiments with Two-State Protein Folding in Bulk *J. Phys. Chem. Lett.* **7**, 3798-3803 (2016)
- <sup>45</sup>G. Stirnemann, and F. Sterpone. Mechanics of Protein Adaptation to High Temperatures *J. Phys. Chem. Lett.* **8**, 5884-5890. (2017)
- <sup>46</sup>O. Languin-Cattoen, S. Melchionna, P. Derreumaux, G. Stirnemann, and F. Sterpone. Three Weaknesses for Three Perturbations: Comparing Protein Unfolding Under Shear, Force, and Thermal Stresses *J. Phys. Chem. B*. **122**, 11922-11930 (2018)
- <sup>47</sup>J. Schonfelder, R Pérez-Jiménez, and V. Muñoz. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution *Nat. Comm.* **7**, 11777 (2015).
- <sup>48</sup>R. Berkovich, S. Garcia-Manyes, K. Klafter, M. Urbakh, J. M. Fernandez. Hopping around an entropic barrier created by force. *Biochemical and Biophysical Research Communications*. **403**, 133-7 (2010)
- <sup>49</sup>W. A. Eaton. Searching for "downhill scenarios" in protein folding. *Proc. Natl. Acad. Sci.* **96**, 5897-99 (1999)
- <sup>50</sup>P. Li, F. Y. Oliva, A. N. Naganathan, and V. Muñoz. Dynamics of one-state downhill protein folding. *Proc. Natl. Acad. Sci.* **106**, 103 (2009)
- <sup>51</sup>M. M. Garcia-Mira, M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz, and V. Muñoz. Experimental Identification of Downhill Protein Folding. *Science*. **298**, 2191-95 (2002)
- <sup>52</sup>J. Schonfelder, D. De Sancho, R. Berkovich, R. B. Best, V. Muñoz, and R. Perez-Jimenez. Reversible two-state folding of the ultrafast protein gpW under mechanical force. *Communications Chemistry*. **1**, 59 (2018)
- <sup>53</sup>G. Yuanm, S. Le, M. Yao, H. Qian, X. Zhou, J. Yan, and H. Chen. Elasticity of the Transition State Leading to an Unexpected Mechanical Stabilization of Titin Immunoglobulin Domains. *Angewandte Chem.* **129**, 5582 (2017)
- <sup>54</sup>C. A. Pierse, and O. K. Dudko. Distinguishing Signatures of Multipathway Conformational Transitions. *Phys. Rev. Lett.* **118**, 088101 (2017)
- <sup>55</sup>H. S. Greenside, and E. Helfand. Numerical integration of stochastic differential equations *Bell Syst. Tech. J.* **60**, 1927 (1981)
- <sup>56</sup>C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco Jr., and C. Bustamante. The ribosome modulates nascent protein folding. *Science* **334** 1723-1727 (2011)
- <sup>57</sup>S. Luccioli, A. Imparato, S. Mitternacht, A. Irbäck and A. Torcini. Unfolding times for proteins in a force clamp. *Phys. Rev. E* **81**, 010902(R) (2010).