

Protein mechanical unfolding: a model with binary variables

A. Imparato,^{1,2,*} A. Pelizzola,^{1,2,†} and M. Zamparo^{1,‡}

¹*Dipartimento di Fisica and CNISM, Politecnico di Torino, c. Duca degli Abruzzi 24, Torino, Italy*

²*INFN, Sezione di Torino, Torino, Italy*

A simple lattice model, recently introduced as a generalization of the Wako–Saitô model of protein folding, is used to investigate the properties of widely studied molecules under external forces. The equilibrium properties of the model proteins, together with their energy landscape, are studied on the basis of the exact solution of the model. Afterwards, the kinetic response of the molecules to a force is considered, discussing both force clamp and dynamic loading protocols and showing that theoretical expectations are verified. The kinetic parameters characterizing the protein unfolding are evaluated by using computer simulations and agree nicely with experimental results, when these are available. Finally, the extended Jarzynski equality is exploited to investigate the possibility of reconstructing the free energy landscape of proteins with pulling experiments.

PACS numbers: 87.15.Aa, 87.15.He, 87.15.-v

I. INTRODUCTION

The three-dimensional structure of proteins is strictly connected to the biological functions these molecules perform in living cells [1]. Among various experimental techniques, an increasingly important role in the study of protein structures is being played by single-molecule force spectroscopy, where proteins [2, 3, 4, 5, 6, 7, 8] (but also nucleic acid fragments [9]) are pulled by applying controlled forces to their ends through an atomic force microscope (AFM) or optical tweezers.

By studying the dynamical response of proteins to constant or varying loading, much information on their structure has been gathered [2, 3, 4, 5, 6]. In particular, the possibility of controlling the applied force with high precision has allowed to trace the molecule folding and unfolding pathways [5, 7, 8]. Nevertheless, as the size of the molecules increases, force spectroscopy outcomes cannot be easily related to molecular properties. Thus theoretical models of biomolecules subject to an external force have been developed [10, 11, 12, 13], which represent important tools to study the interplay between protein response to external force and molecular structure.

In most cases, these models are based on a coarse-grained description of the biomolecule, their dynamical degrees of freedom being related to the coordinates and velocities of a suitable set of reference beads, typically one or a few per amino acid [10, 11, 12, 13]. Equilibrium and nonequilibrium results are then obtained by means of (time expensive) computer simulations.

In a recent paper [14] we have approached the problem of protein unzipping from a different point of view, introducing a simple lattice model with binary degrees of freedom, based on (and generalizing) the Wako–Saitô (WS) model of protein folding [15, 16]. Despite its simplicity, the model turned out to exhibit the typical response of real proteins to pulling. In particular, the mechanical unfolding of the 27th immunoglobulin

domain of titin was investigated, considering both force clamp and dynamic loading protocols. Theoretically expected laws were verified and an excellent agreement with the experimental values of the characteristic kinetic parameter was found. The model was also used to investigate the possibility of reconstructing the protein free energy landscape by exploiting an extended version of the Jarzynski equality (JE) [17, 18, 19].

The present paper has several purposes. On the methodological side, we give a full derivation of the relation between our model and the original WS model, and show in detail how to obtain free energy landscapes. On the application side, we consider three other molecules (including BBL, whose thermal behaviour has recently been the subject of some debate, see ref. [20] and references therein, and ref. [21]) in addition to titin and, together with the properties already investigated in [14], we discuss also the probability distribution of unfolding times and forces.

The article is organized as follows. In Section II, we describe the details of our model, its connection with the WS model, and the simulation method. In Section III we present the equilibrium properties of the model proteins, as functions of force and temperature. We then present the results of mechanical unfolding obtained by numerical simulations, first, in Section IV, for the force clamp manipulation and then, in Section V, for the dynamic loading. In Section VI, the free energy landscape of the model proteins is reconstructed from unfolding manipulations, by using an extended Jarzynski Equality. Conclusions are drawn, and future developments are sketched, in Section VII.

II. THE MODEL

In the present section, we define our model and show that in the absence of an external force it reduces to the WS model of protein folding. The latter was introduced in 1978 by Wako and Saitô, in two papers [15, 16] that appear to have been forgotten until recent years. The same model was independently reintroduced by Muñoz, Eaton and coworkers at the end of the '90s [22, 23, 24]. These authors used the model to describe and interpret experimental results, and soon the model became quite popular [25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35]. This

*Electronic address: alberto.imparato@polito.it

†Electronic address: alessandro.pelizzola@polito.it

‡Electronic address: marco.zamparo@polito.it

is why it is often referred to as the Muñoz–Eaton, or Wako–Saitō–Muñoz–Eaton, model. The first recent reference to the original work by Wako and Saitō appeared, as far as we know, in [27].

Similarly to the WS model, the state of a protein is defined according to the conformation of its peptide bond backbone, and in order to reduce the degrees of freedom, we assume that the peptide bonds can exist only in two conformations: native and non-native. Thus a $N + 1$ amino acid protein is represented as a chain of N peptide bonds, and a binary variable m_k is associated to each peptide bond. Bonds are numbered from 1 to N and amino acids from 1 to $N + 1$, bond k connecting amino acids k and $k + 1$. The variable m_k takes the value 0, 1 corresponding to a peptide bond in unfolded or native state respectively.

In order to couple the molecule to an external force we regard stretches of consecutive native bonds (delimited by unfolded bonds) as rigid portions of the protein with their own (native) end-to-end length: in the following l_{ij} will indicate the end-to-end length of the stretch of consecutive amino acids connected by native bonds and delimited by unfolded bonds in positions i and $j > i$: $m_i = m_j = 0$, $m_k = 1$ for $i < k < j$; if we define the quantity

$$S_{ij}(m) \equiv (1 - m_i)(1 - m_j) \prod_{k=i+1}^{j-1} m_k, \quad (1)$$

this condition can be expressed in a more compact form: $S_{ij}(m) = 1$. In the limiting case $j - i = 1$ the stretch reduces to amino acid $j = i + 1$, which is characterized by its native length $l_{i,i+1}$. In the following, “stretch” will refer to stretches of any length, including single amino acids. Boundary conditions $m_0 = m_{N+1} = 0$ are introduced to define the stretches at the protein ends. It is easy to verify that the number of stretches is equal to 1 plus the number of 0’s in the set $\{m_k, k = 1, \dots, N\}$.

We want to keep our model as simple as possible, therefore only two orientations of rigid stretches are considered: given the direction of the external force, a stretch can only be oriented parallel or antiparallel to the force direction. Thus, following [14], we introduce the variables $\sigma_{ij} = \pm 1$ which describe the orientation of the stretch with respect to the external force f . Therefore, the configuration of the molecule is fixed by the set of $(\{m_k\}, \{\sigma_{ij}\})$ values, and for each configuration the protein end-to-end length is given by

$$L(\{m_k\}, \{\sigma_{ij}\}) = \sum_{0 \leq i < j \leq N+1} l_{ij} \sigma_{ij} S_{ij}(m). \quad (2)$$

A cartoon of the model protein is plotted in fig.1. It is worth noting that a given bond configuration $\{m_k\}$ dynamically fixes the set of variables $\{\sigma_{ij}\}$: for each configuration $\{m_k\}$ only those variables $\{\sigma_{ij}\}$ such that $S_{ij}(m) = 1$ are considered.

As in the original WS model [15, 16, 22, 23, 24], in the present model two amino acids i and $j + 1 > i$ interact only if *all* the peptide bonds connecting them along the chain (that is, bonds from i to j) are in the native state, and if they are close enough in the native configuration. The Hamiltonian of

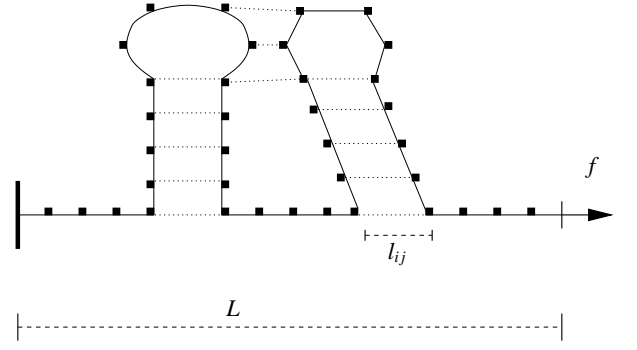


FIG. 1: Cartoon of the model protein, with a force applied to one of the free ends. Dots denote amino acids and dashed lines denote contacts.

the model reads thus

$$\mathcal{H}(\{m_k\}, \{\sigma_{ij}\}, f) = - \sum_{i=1}^{N-1} \sum_{j=i+1}^N \epsilon_{ij} \Delta_{ij} \prod_{k=i}^j m_k - f L(\{m_k\}, \{\sigma_{ij}\}) \quad (3)$$

The quantity $\epsilon_{ij} > 0$ represents the interaction energy between the amino acids i and $j + 1$, (defined as in [24, 30, 34], see below for details), while $\Delta_{ij} = 0, 1$ is the corresponding element of the contact matrix, taking the value 1 if the distance between any two atoms of the two amino acids is smaller than some threshold distance, or the value 0 if none of all the possible atom pairs satisfies this condition, see discussion below.

The microscopic degrees of freedom σ_{ij} do not interact among themselves, hence the partial partition sum over them can be performed analytically. We obtain the effective Hamiltonian \mathcal{H}_{eff} , defined by

$$\sum_{\{\sigma_{ij}\}} \exp[-\beta \mathcal{H}(\{m_k\}, \{\sigma_{ij}\}, f)] = \exp[-\beta \mathcal{H}_{\text{eff}}(\{m_k\}, f)],$$

that reads

$$\mathcal{H}_{\text{eff}}(\{m_k\}, f) = - \sum_{i=1}^{N-1} \sum_{j=i+1}^N \epsilon_{ij} \Delta_{ij} \prod_{k=i}^j m_k - k_B T \sum_{0 \leq i < j \leq N+1} \ln [2 \cosh(\beta f l_{ij})] S_{ij}(m). \quad (4)$$

This effective Hamiltonian is a linear combination of products of consecutive m_k ’s (including the single peptide bond) and therefore it has the same mathematical structure of the WS model, whose Hamiltonian reads

$$\mathcal{H}_0(\{m_k\}) = - \sum_{i=1}^{N-1} \sum_{j=i+1}^N \epsilon_{ij} \Delta_{ij} \prod_{k=i}^j m_k - k_B T \sum_{i=1}^N q_i (1 - m_i), \quad (5)$$

where q_i is the entropic cost of ordering bond i . Given this similarity, the partition function associated to the Hamiltonian (4) can be summed exactly, and the thermodynamic quantities can be obtained, as discussed in ref. [30, 31]. In the case $f = 0$, \mathcal{H}_{eff} reduces to (up to an additive constant)

$$\mathcal{H}_{\text{eff}}(\{m_k\}, 0) = - \sum_{i=1}^{N-1} \sum_{j=i+1}^N \epsilon_{ij} \Delta_{ij} \prod_{k=i}^j m_k - k_B T \ln 2 \sum_{i=1}^N (1 - m_i),$$

which corresponds to the WS model, eq.(5), with $q_i = \ln 2$. In the absence of force, our new orientational degrees of freedom correspond to an entropic gain q_i associated to the unfolded peptide bonds. This quantity takes the value $\ln 2$ because we have made the simplest choice of two possible orientations per stretch. Of course this is a very rough discretization of the actual orientational degrees of freedom of the main chain, namely the dihedral angles. In principle, one could think to orientational variables taking values in a larger set, and this would yield different, even non-uniform, q_i values.

As discussed in ref. [14], in the present model the quantities Δ_{ij} , ϵ_{ij} and l_{ij} are chosen by analyzing the native structure of a given molecule taken by the Protein Data Bank (pdb in the following, <http://www.pdb.org/>). We briefly review here the choice criteria for readability's sake. Two amino acids i and $j + 1$ (with $j + 1 > i + 2$) are in contact ($\Delta_{ij} = 1$) if, in the native state of the protein, at least two atoms from these amino acids are closer than 4 \AA . In this case ϵ_{ij} is taken to be equal to $k\epsilon$, where k is an integer such that $5(k - 1) < n_{at} \leq 5k$, and n_{at} is the number of atoms of the two amino acids whose distance is not larger than the threshold distance. The quantity ϵ is the protein energy scale, and is determined by imposing that, at zero force and at the experimental denaturation temperature T_m , the fraction of folded molecules is $p(T_m) = 1/2$. In order to estimate p we introduce the number of native peptide bonds $M = \sum_{k=1}^N m_k$ and its average density $m = \langle M \rangle / N$, which in the following will be used as an order parameter. The fraction of folded molecules is then estimated, assuming a two-state picture, as $p = (m - m_\infty) / (m_0 - m_\infty)$, where $m_\infty = 1/3$ is the value of m at infinite temperature, while m_0 is a good representative of the folded state [34]. For most molecules we can take $m_0 = m(T = 0) = 1$, an exception being the WW domain of PIN1 (pdb code 1I6C), which orders perfectly only at zero temperature and exhibits a wide plateau in $m(T)$ in the range 200–300 K [34]. In this case we choose $m_0 = m(T = 273 \text{ K}) < 1$.

As far as the parameters l_{ij} are concerned, the generic amino acid i is represented by its $N_i - C_{\alpha,i} - C_i$ sequence. Taking the native state as the reference configuration, l_{ij} is chosen as the distance between the midpoint of the C_i and N_{i+1} atoms and the midpoint of the C_j and N_{j+1} atoms.

In the present paper we shall consider four different molecules of increasing size: 1BBL (37 amino acids, see [34] for details), 1I6C (39 amino acids), 1COA (64 amino acids), 1TIT (89 amino acids). Here and in the following the proteins are indicated with their pdb code. Some results on the mechanical unfolding of 1TIT and a few about 1I6C have already been reported in [14] and will be recalled here for comparison with the other molecules. Results on the thermal unfolding of the other molecules, based on the WS model, have been reported in [30] (1COA) and [34] (1BBL and 1I6C). In particular, in [34] it was shown that 1BBL differs from a clear two-state behaviour, though not being a true downhill folder as some authors have claimed. In the following we shall see that signals of the peculiar behaviour of 1BBL appear also in the mechanical unfolding.

While the equilibrium properties of the present model can be calculated exactly (see next section), results on the unfold-

ing kinetics will be instead obtained by performing Monte Carlo (MC) simulations with Metropolis algorithm, using the effective Hamiltonian (3). In the following t_0 will indicate the system time scale, corresponding to a single MC step.

III. EQUILIBRIUM PROPERTIES

Once the temperature and the external force have been fixed, the macroscopic state of the system is defined by the order parameter m and by the molecule length L .

In figure 2 the equilibrium values of such quantities are plotted as functions of the force, at room temperature, for the different molecules here considered. Inspection of these figures suggests that the two larger molecules exhibit a sharp transition to the unfolded state, while for the two smaller molecules the unfolding is more gradual as the force is increased. The small force plateaux correspond to a global re-

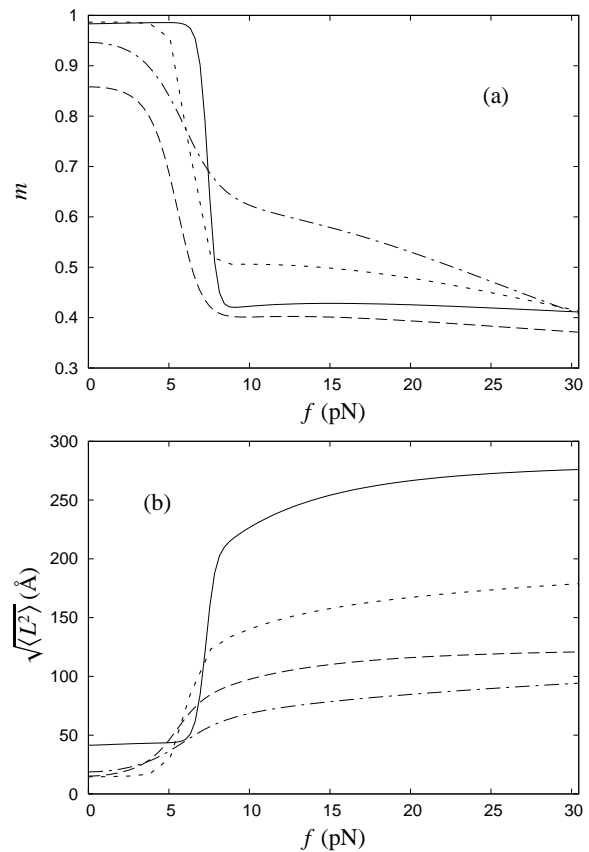


FIG. 2: Panel (a): average order parameter m as a function of the external force f , for different molecules: 1TIT (full line), 1COA (dotted line), 1I6C (dashed line), 1BBL (dashed-dotted line). The temperature value is taken to be $T = 300 \text{ K}$. Panel (b): Root mean square length of the same molecules as a function of the external force f , with $T = 300 \text{ K}$.

orientation of the molecule, which remains in its native state, under the external force. When the force increases a more or less sharp transition to an elongated state occurs. As in the thermal unfolding study [34] we see that 1BBL exhibits

a different, more gradual transition with respect to the other molecules, which exhibit a clearer two-state behaviour.

For each molecule, we can obtain a phase diagram by computing the locus of points in the temperature–force plane where the fraction of folded molecules $p = 1/2$. Such curves are plotted in Fig. 3: inspection of this figure indicates, again, that 1BBL differs qualitatively from the other molecules.

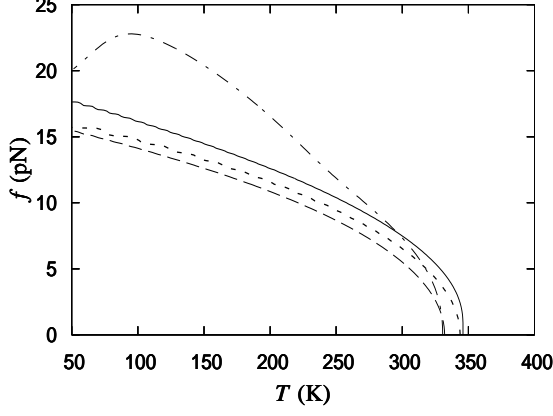


FIG. 3: Phase diagram of 1TIT (full line), 1COA (dotted line), 1I6C (dashed line), 1BBL (dashed-dotted line). For each molecule, the curves are defined by $p(f, T) = 1/2$. The lower-left region of the phase diagram corresponds to folded molecules, the upper-right region corresponds to unfolded molecules.

In order to characterize the macroscopic state of the molecules with the observable quantity L , let us define the constrained zero-force partition function $Z_0(L)$ as follows:

$$Z_0(L) = \sum_{\{m_k\}, \{\sigma_{ij}\}} \delta(L - L(\{m_k\}, \{\sigma_{ij}\})) e^{-\beta H(\{m_k\}, \{\sigma_{ij}\}, f=0)}, \quad (6)$$

and the corresponding free energy by

$$F_0(L) = -k_B T \ln Z_0(L). \quad (7)$$

In appendix A we show how $Z_0(L)$, as defined by eq. (6), can be calculated exactly for our model. In figure 4(a), the equilibrium free energy landscape of the four molecules here considered is plotted as a function of the molecule length, for $T = 300$ K. Inspection of this figure indicates that all the molecules have a minimum of the free energy at small L , the value of the minimum being compatible with the value of $\sqrt{\langle L^2 \rangle}$ at zero force, as plotted in fig. 2(b). At larger force, the free energy function (7) is tilted and reads $F(L, f) = F_0(L) - f \cdot L$: this function exhibits new minima for each molecule, which are compatible with the values of $\sqrt{\langle L^2 \rangle}$ in the large force regime, see fig. 4(b). As an example, let us consider the 1TIT molecule. In fig. 4(b), the function $F(L, f)$, with $f = 10$ pN, exhibits a global minimum at $L \approx 225$ Å, which corresponds to the equilibrium value of $\sqrt{\langle L^2 \rangle}$ for the same molecule at $f = 10$ pN, as shown in fig. 2(b).

The function $F(L, f)$ exhibits an energy barrier at small L , whose height ΔF and width ΔL depend on the value of f . In order to give an example of the typical values of these quantities, let us define $f_{1/2}$ as the force where the molecular length is half of its maximum value.

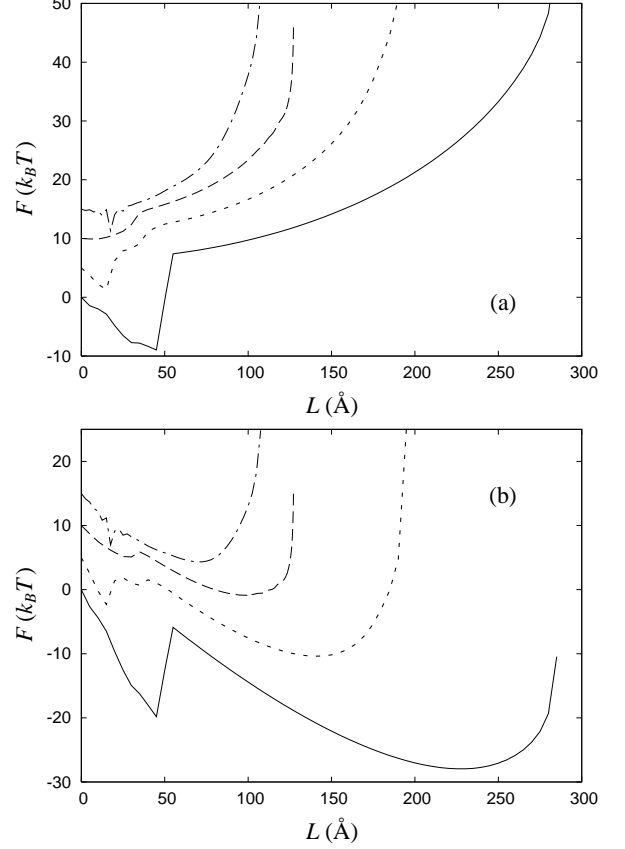


FIG. 4: Panel (a): Free energy landscape as a function of the molecular elongation L , at $T = 300$ K, for different molecules: 1TIT (full line), 1COA (dotted line), 1I6C (dashed line), 1BBL (dashed-dotted line). Panel (b): Tilted free energy landscapes $F(L) - f \cdot L$, with $f = 10$ pN.

In table I, we list the width ΔL and the height ΔF of the energy barrier separating the two minima of the function $F(L) - f \cdot L$, for $f = f_{1/2}$ and $T = 300$ K.

Molecule	ΔL (Å)	$\Delta F(k_B T)$
1BBL	5	2.7
1I6C	10	1.6
1COA	25	5.7
1TIT	10	12.7

TABLE I: Width ΔL and height ΔF of the energy barrier separating the two minima of the free energy $F(L) = F_0(L) - f \cdot L$, with $f = f_{1/2}$, $T = 300$ K, for the molecules here considered.

IV. FORCE CLAMP

In this section, we investigate the unfolding of the model proteins by application of a constant external force: such a manipulation scheme is usually called “force clamp”. The force on the molecule is suddenly increased from 0 to its final value f , and its length is measured [7, 8]. Usually the unfolding of a small molecule or of a portion of a large molecule is viewed as the overcoming of a kinetic barrier in the molecular energy landscape [36]. Such a barrier is characterized by a width x_u along the reaction coordinate, and by a height ΔE_u over the corresponding minimum. Thus, the mean unfolding time is expected to follow the Arrhenius law

$$\tau_u = \omega_0^{-1} e^{\beta(\Delta E_u - f x_u)} = \tau_0 e^{-\beta f x_u}, \quad (8)$$

where ω_0 is a microscopic attempt rate and $\tau_0 = \omega_0^{-1} \exp(\beta \Delta E_u)$ is the mean unfolding time at zero force. Note that Eq. (8) is well defined as long as $f \leq \Delta E_u / x_u$, i.e. as long as the process can be actually considered as a jump process over an energy barrier. For $f > \Delta E_u / x_u$ one expects that the mean escape time τ_u is independent of the external force but rather depends on the microscopic details of the system. On the other hand for too small forces the system will not unfold, if the barrier ΔE_u is sufficiently high ($\beta \Delta E_u \gg 1$).

In order to check whether the unfolding time under constant force obeys such a law, and to extract the kinetic parameter x_u , we run MC simulations to mimic the unfolding of molecules subject to a force clamp with force f , at time $t = 0$. We consider a molecule as unfolded as soon as its length takes the value $L_u = L_{max}/2$, where L_{max} is the molecule length in the completely unfolded state. For each molecule 1000 independent unfolding trajectories are considered. In fig. 5, the mean unfolding time of the four molecules is plotted as a function of the force. In the case of 1TIT, the force does not extend to the small force range, since we find that the unfolding time τ_u goes to infinity in the very small force regime, i.e., the molecules do not unfold at all. Inspection of this fig-

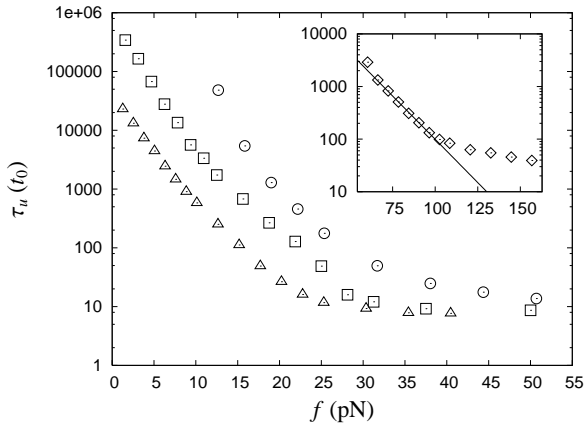


FIG. 5: Mean unfolding time as a function of the force, at $T = 300$ K, for four different molecules: boxes 1I6C, triangles 1BBL, circles 1COA, inset: 1TIT.

ure suggests that the mean unfolding time as a function of the

force follows eq. (8) in a wide range of force values, and then saturates in the large force regime, as found in other works [11, 12]. The force value separating the two regimes depends on the particular molecule. From fits to eq. (8) the values of the kinetic parameter x_u can be obtained for each molecule. In table II the values of the unfolding length are listed for the molecules considered in this paper. In principle, from such a fit procedure one is also able to obtain an estimate of τ_0 . However the quantity τ_0 cannot be expressed in seconds, since this would require to evaluate the molecular time scale t_0 . This, in turn, requires an experimental estimate of τ_0 with the force clamp technique. It is worth noting that, despite the simplicity of our model, the unfolding length for 1TIT is in remarkable agreement with the experimental value $x_u = 2.5 \text{ \AA}$ [3].

Molecule	x_u (Å)
1BBL	14.8 ± 0.4
1I6C	20.3 ± 0.5
1COA	20 ± 2
1TIT	3.13 ± 0.1

TABLE II: Unfolding length x_u as given from linear fits to eq. (8), for the molecules considered in this paper.

The unfolding of the molecule is a stochastic process, and thus the unfolding time varies for each realization of the process. It is thus interesting to study the distribution of the unfolding time. In figure 6, we plot histograms of the unfolding time of the 1TIT molecule at $T = 300$ K, for small and large force. In the large force regime a lognormal distribution was proposed in [12], without a theoretical justification, and shown to fit quite well the results of molecular dynamics simulations. Our results are also well fitted by a lognormal distribution. Nevertheless, we find it interesting to look for a theoretical distribution for the large force regime.

In the large force regime, one can imagine the chain as made up of a number $M < N$ of stretches which can be easily turned by the force from the antiparallel to the parallel (with respect to the force itself) configuration. At the beginning of the pulling process, these stretches will be randomly oriented, half of them antiparallel and half parallel to the force. With a frequency τ_1^{-1} , a stretch will be selected at random by the kinetics and, if antiparallel to the force, turned parallel with probability 1. Therefore, after a time $\tau_u = k\tau_1$, the probability that the chain will be completely elongated in the force direction will be $p(M, k) = [1 - (1 - 1/M)^k]^{M/2}$. The probability distribution of the unfolding time will therefore be approximated by $f(\tau_u = k\tau_1) = p(M, k) - p(M, k-1)$, which for large M can be approximated as

$$f(\tau_u = k\tau_1) = 1/2 \exp[-k/M - M/2 \exp(-k/M)]. \quad (9)$$

As can be seen in Fig. 6, this formula fits reasonably well our data, though not as well as a lognormal. Similar results are obtained for the other molecules (data not shown).

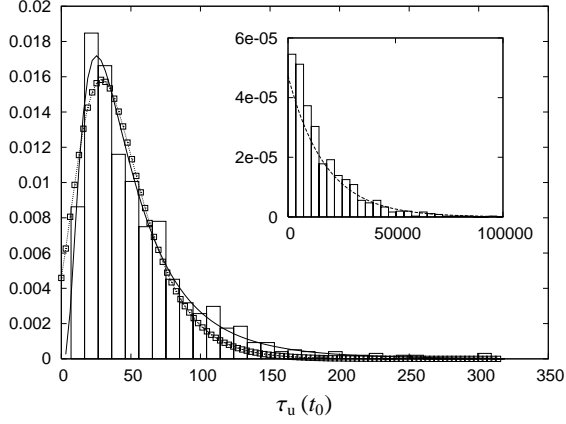


FIG. 6: Histograms of the unfolding time τ_u for the 1TIT molecule in a force clamp, at $T = 300$ K with force $f = 130$ pN (main figure) and $f = 48$ pN (inset). The lines in the main figure are fits of the data to a log-normal function (full line) and to the function (9) (line-points). The line in the inset corresponds to a fit to a negative exponential function.

A. The Kramers problem

We now ask the following question: is the free energy landscape $F(L)$ the “kinetic” potential associated to the diffusive motion of the system along the coordinate L ? In other words, in the zero force regime, is the thermal motion of the molecule a diffusion process across the potential $F_0(L)$? A positive answer to this question would imply that the mean first passage time of the system to a given value L^* of the elongation should be given by the solution of the Kramers problem. The Kramers problem amounts to evaluating the mean first passage time (mfpt) of a Brownian particle, moving in an energy potential $U(x)$, to a given point x_f of the potential. If we let x_0 be the initial position of the particle, the mfpt from x_0 to x_f reads [37]

$$\tau(x_f) = \frac{1}{D} \int_{x_0}^{x_f} dy e^{\beta U(y)} \int_{-\infty}^y e^{-\beta U(z)}, \quad (10)$$

where D is the diffusion coefficient, which basically sets the time scale of the process.

In figure 7(a), we plot the mean first passage time of the 116C molecule at $L^* = L_{max}/2$, as a function of f , for different temperatures, as obtained by eq. (10), where $U(x)$ has been replaced by the free energy $F_0(L) - f \cdot L$, as given by eq. (7). Inspection of this figure suggests that in the small force range, the unfolding kinetics of the protein is actually described by the equilibrium free energy landscape $F(L, f)$. The agreement improves as the temperature decreases. The mean first passage time of the molecule is not recovered by eq. (10) in the large force regime, because the energy difference $F(L^*, f) - F(L = 0, f)$ becomes negative, and thus the motion towards the new minimum at large L becomes purely diffusive, see fig. 7(b).

The same behaviour is found for the 1BBL molecule (data not shown). Since the unfolding time grows exponentially with the size of the molecule, for the two larger molecules

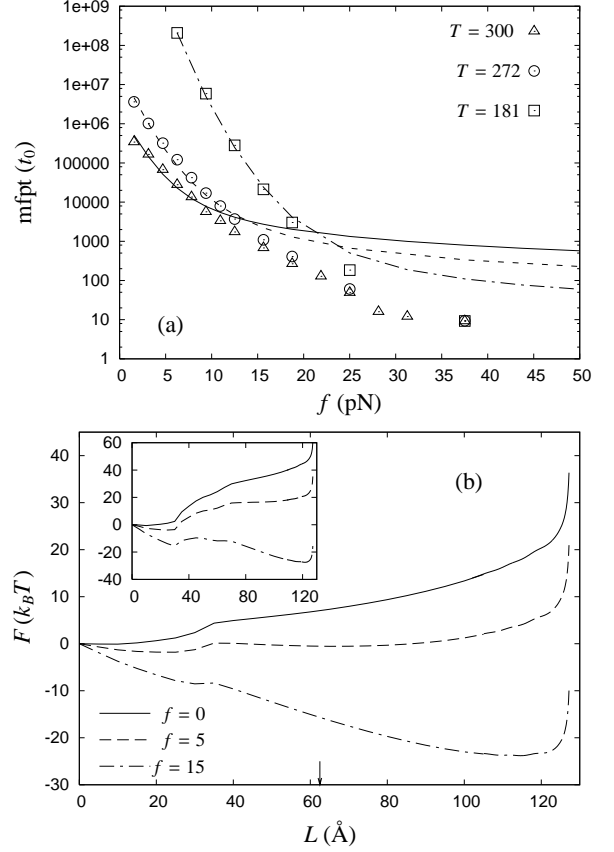


FIG. 7: Panel (a): Mean first passage time at $L^* = L_{max}/2$ for the 116C molecule and for different temperatures (in K), as obtained by direct computer simulations (points) and by eq. (10) (lines). Panel (b): Free energy landscape F as a function of the molecule length L , at $T = 300$ K, for different values of the force (in pN). Inset: Free energy landscape F at $T = 181$ K.

we were not able to observe the unfolding in the small force regime within reasonable computation times.

V. DYNAMIC LOADING

In this section we consider the following manipulation strategy, which is often used in experiments [2, 3, 4, 5, 6]: starting from equilibrium configurations, a time-dependent force is applied to our model molecule and the unfolding time is sampled. As in the case of the force clamp, we define the unfolding time as the first passage time of the molecule length across the threshold value L_u . Here the force is taken to increase linearly with time, with a rate r : this manipulation scheme corresponds to the force-ramp experimental setup discussed in [6]. Thus, the rupture force f_u is given by $f_u = r\tau_u$.

As discussed above, the breaking of a molecular linkage is typically described as a thermally activated escape process from a bound state over a barrier which dominates the kinetics. It can be shown that, if the energy barrier ΔE_u is large (compared to the thermal energy $k_B T$) and rebinding is negli-

gible, the typical unbinding force of a single molecular bond reads [36, 38]

$$f^* = \frac{k_B T}{x_u} \ln \left[\frac{r x_u \tau_0}{k_B T} \right]. \quad (11)$$

In fig. 8(a) the typical unbinding force f^* (the most probable value of f_u) is plotted as a function of the pulling velocity for the 1TIT molecule, for three values of the temperature. Inspection of this figure suggests that the range of values of r , where f^* is a linear function of $\ln r$, depends on the temperature: the smaller is T , the wider is this range. We find that, for this molecule, the value of the unfolding length, $x_u \approx 3 \text{ \AA}$, obtained by fitting the data in the linear regime to eq. (11), is independent of the temperature, as expected (see caption of the figure for the numerical values). Note that the value of the unfolding length, x_u , in the linear regime defined by eq. (11), agrees with that found with the force clamp manipulation, and with the experimental value $x_u = 2.5 \text{ \AA}$ found in ref. [3].

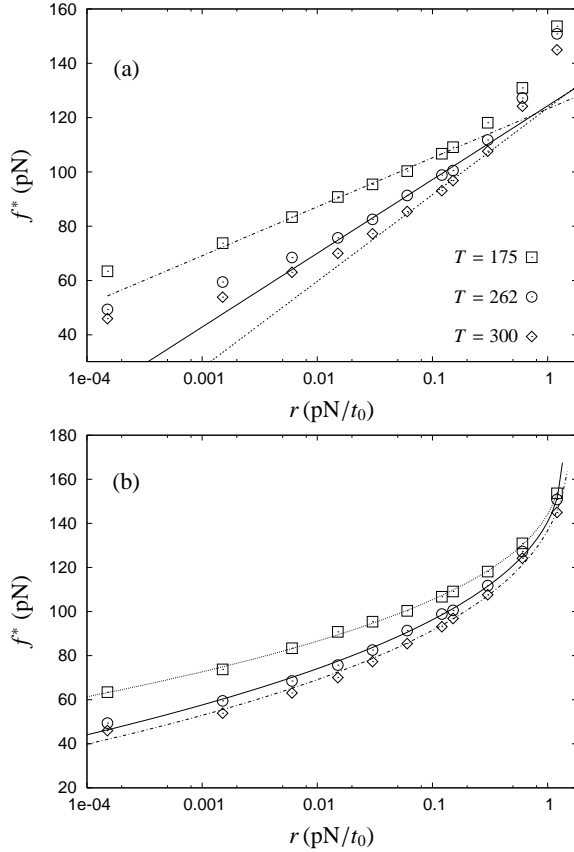


FIG. 8: Panel (a): plot of the typical unbinding force f^* of the 1TIT molecule as a function of the pulling velocity r , for the three values of the temperature. The lines are fits to the data in the linear regime defined by eq. (11). From such fits one can obtain the characteristic unfolding length x_u . We find $x_u = 3.1 \pm 0.1 \text{ \AA}$ for $T = 175 \text{ K}$, $x_u = 3.0 \pm 0.1 \text{ \AA}$ for $T = 262 \text{ K}$, and $x_u = 3.05 \pm 0.2 \text{ \AA}$ for $T = 300 \text{ K}$. Panel (b): the lines are fits of the data to eq. (12).

By repeating the same procedure for the other molecules, we estimate their unfolding lengths, the results are listed in

Molecule	x_u (Å)
1BBL	14.0 ± 0.3
1I6C	17.7 ± 0.5
1COA	22.4 ± 0.5
1TIT	3.05 ± 0.20

TABLE III: Unfolding length x_u , as given from linear fits to eq. (11), for the molecules considered in this paper.

table III. Again, τ_0 cannot be directly compared with experimental values, but we see that the unfolding lengths are comparable to the ones obtained by the force clamp protocol.

However, as discussed above the rupture force is not a linear function of r in the whole range of the pulling rate: figure 8 rather suggests that the slope of f^* increases as r increases. In refs. [38, 39, 40] it has been argued that the appearance of different slopes in the f^* vs. $\ln r$ plot could be the signature of the presence of different escape paths from the folded state. Each of these different paths would be selected by pulling the molecule with a given rate r . Thus, the different slopes in the f^* vs. $\ln r$ plot correspond to different characteristic lengths x_u of the paths. On the other hand, in a recent work [42], considering particular choices of the energy landscape which make exact computations feasible, it has been argued that, the typical unbinding force f^* has a more complex expression

$$f^* = \frac{\Delta E_u}{\nu x_u} \left\{ 1 - \left[\frac{k_B T}{\Delta E_u} \ln \left(\frac{\omega_0 e^\gamma}{\beta x_u r} \right) \right]^\nu \right\}, \quad (12)$$

where the exponent ν depends on the microscopic details of the energy landscape, and γ is the Euler-Mascheroni constant $\gamma \approx 0.577$. Equation (12) reduces to eq. (11) in the limit $\Delta E_u \rightarrow \infty$, or when the exponent ν takes the value 1 [42]. A similar expression for the rupture force was previously proposed in [43], with $\nu = 2/3$. In fig. 8(b), the fits of the typical unbinding force data to eq. (12) are plotted for the 1TIT molecule. The fits turn out to be rather good, but statistical errors are quite large. In order to reduce them, for each molecule, we considered sets of (r, f^*) data at different temperatures, and we made joint fits according to Eq. (12). The values of the unfolding lengths, energy barriers and exponents obtained by these fits are listed in table IV. Comparison of

Molecule	x_u (Å)	ΔE_u ($k_B T$, $T = 300 \text{ K}$)	ν
1BBL	22 ± 1	10.0 ± 0.2	0.61 ± 0.03
1I6C	18 ± 1	13 ± 2	0.7 ± 1
1COA	31.5 ± 4	16 ± 1	0.625 ± 0.06
1TIT	11.5 ± 2	18 ± 1	0.42 ± 0.04

TABLE IV: Unfolding length x_u , barrier height ΔE_u , and characteristic exponent ν as given from fits of the unfolding data with dynamic loading technique to eq. (12).

the unfolding lengths listed in table IV and in table III indicates that the values of the x_u obtained by fitting the typical unbinding force to eq. (12) are rather different from the values obtained by fitting the same data to eq. (11), although, as

discussed before, the fit to eq. (11) is restricted to the intermediate range of values of r . This is in agreement with refs. [42, 43]. In those references it was argued that eq. (12) describes the rupture force in the whole regime of r rather than just in the linear regime, as eq. (11) does. It is worth noting that the values of the exponent ν found for the 1BBL, 1I6C and 1COA molecules (table IV) are compatible with the value $\nu = 2/3$ found in refs. [42, 43] for a particular choice of the energy landscape. Furthermore, the value of the kinetic barrier $\Delta E_u = 18 \pm 1 k_B T$ (at $T = 300$ K) found for the 1TIT molecule, is of the same order of magnitude of the experimental value $\sim 37.3 k_B T$ found in [3].

Similarly to the case of the force clamp, the distribution of the unbinding force exhibits a nontrivial dependence on the system kinetic parameters [41]:

$$P(f) = \frac{1}{\tau_0 r} e^{\beta f x_u} \exp \left[-\frac{k_B T}{r x_u \tau_0} (e^{\beta f x_u} - 1) \right]. \quad (13)$$

The maximum of this distribution corresponds to the typical unbinding force, eq. (11).

In figure 9, the distribution of the unfolding force is plotted for the 1TIT molecule and different rate values. Inspection of fig. 9 suggests that the observed distribution of unfolding force agrees nicely with the expected one. Similar results are obtained for the other molecules.

In order to characterize the different unfolding paths occurring for different values of r , one can assume that the unfolding length x_u is a function of the pulling rate r , and exploit eq. (13) to extract the value $x_u(r)$, so as to estimate the unfolding length at any value of r . In figure 10 the unfolding length x_u of the 1COA and 1TIT molecules are plotted as a function of r , as obtained from this fitting procedure. As expected we find that x_u is a decreasing function of r . Thus, the values $x_u = 22.4 \text{ \AA}$ (1COA) and $x_u = 3.05 \text{ \AA}$ (1TIT), obtained by fitting the typical unbinding force to eq. (11) (see, table III), turn out to be weighted averages of the values $x_u(r)$ as plotted in fig. 10.

VI. EVALUATING THE FREE ENERGY LANDSCAPE FROM PULLING EXPERIMENTS

Given that we can compute exactly the free energy landscape $F_0(L)$ of our model, let us now address the problem of evaluating this landscape through the force manipulation experiments. As discussed above, applying a dynamic loading is equivalent to drive the system out of equilibrium by coupling it to the external potential $U(L, t) = -f(t)L$.

The free energy landscape can be evaluated by using a fluctuation relation, which is an extended form of the Jarzynski equality [18, 19].

$$\left\langle \delta(L - L(\{m_k\}, \{\sigma_{ij}\})) e^{-\beta W} \right\rangle_t = e^{-\beta(F_0(L) - f(t)L)} / Z_0, \quad (14)$$

where $Z_0 = \int dL Z_0(L)$, with $Z_0(L)$ as given by eq. (6). Combining the Jarzynski equality [17], and the weighted histogram method [44], it can be shown that, if the molecule is driven out of equilibrium by a time-dependent external potential coupled

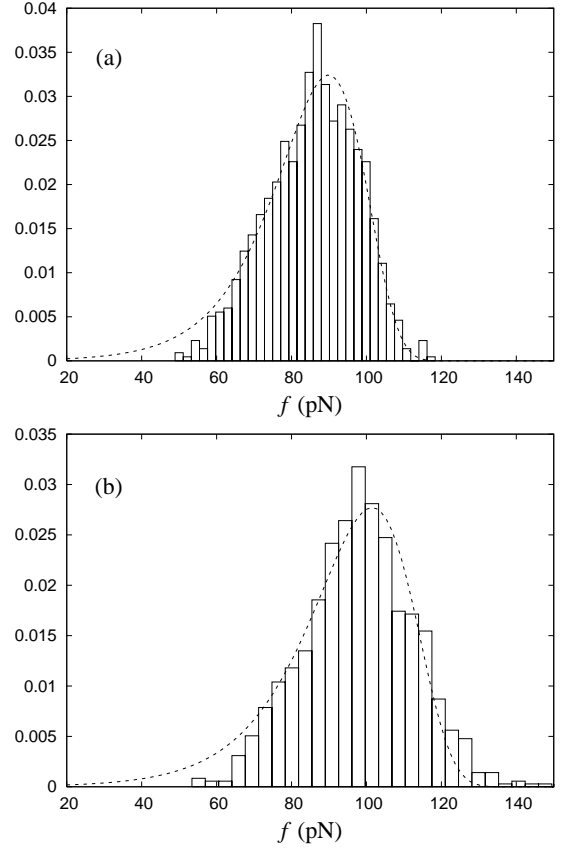


FIG. 9: Distribution of unfolding force under dynamical loading for the 1TIT molecule for (a) $r = 0.06 \text{ pN}/t_0$ and (b) $r = 0.15 \text{ pN}/t_0$, with $T = 300$ K. The dotted line is a fit to eq. (13).

to its length $U(L, t)$, the free energy F as a function of L is given by [18, 19]

$$F(L) = -k_B T \ln \left[\frac{\sum_t \frac{\langle \delta(L - L(\{m_k\}, \{\sigma_{ij}\})) \exp(-\beta W_t) \rangle_t}{\langle \exp(-\beta W_t) \rangle_t}}{\sum_t \frac{\exp(-U(L, t))}{\langle \exp(-\beta W_t) \rangle_t}} \right], \quad (15)$$

where W_t is the thermodynamic work *done* on the system by the external potential, up to the time t , defined as $W_t = \int_0^t dt' \partial \mathcal{H} / \partial t'$, and the average $\langle \dots \rangle_t$ is over all the trajectories of fixed duration t . In an experimental situation, the work W_t is not sampled continuously, but at successive discrete times $0, \Delta t, 2\Delta t, \dots, M\Delta t$. Therefore the sum over t in Eq. (15) runs over these discrete values.

The estimated free energy for the smaller molecules are plotted in fig. 11, as obtained by 10000 independent pulling trajectories, together with the exact ones. As expected [14, 19], the curves obtained by numerical “experiments” collapse onto the expected one as the pulling rate r decreases.

Within the present scheme, it was not possible to evaluate the free energy landscape of the two larger molecules: indeed the work needed to completely unfold the two molecules (1COA, 1TIT) amounts to some hundreds of $k_B T$, and thus the numerical precision in evaluating the average value of $\exp(-\beta W)$ is rather scanty.

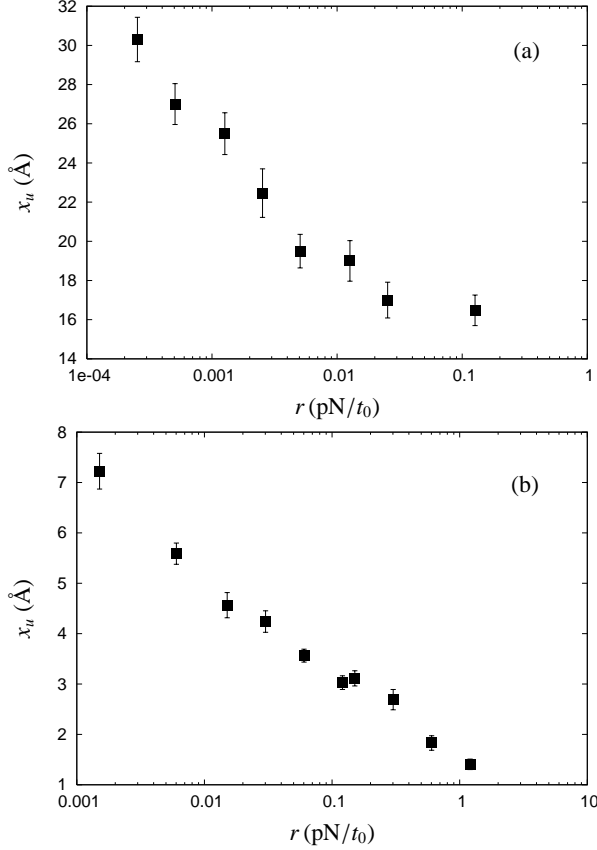


FIG. 10: Unfolding length x_u as a function of r , as obtained from fits of the unfolding force probability distribution to eq. (13) with $T = 300$ K, for the 1COA (a) and 1TIT (b) molecules.

Our results suggest a practical procedure to estimate the free energy landscape of real proteins with dynamic loading experiments. The work done on the molecule has to be sampled for different pulling rates: for each rate eq. (15) provides an estimate of the target function $F_r(L)$. As the rate r decreases, the curves are expected to superimpose more and more: when the difference between the curves is of the order of few $k_B T$, for the whole range of L , the estimate of $F(L)$ can be considered reliable within this small uncertainty.

VII. CONCLUSIONS

In this paper we presented a comparative study of four widely studied proteins, by exploiting a simple model recently introduced. Such a model allows one to obtain analytically the equilibrium properties of the molecules considered. By using MC simulations we also study the unfolding kinetics of the molecules as they are pulled through an external force applied to their free ends. As already discussed in [14], the model turns out to exhibit the typical behaviour of a protein undergoing a mechanical manipulation, both with the force-clamp scheme and with the dynamic-loading scheme.

We systematically study the *equilibrium* free energy landscape of the model molecule, as a function of an experimen-

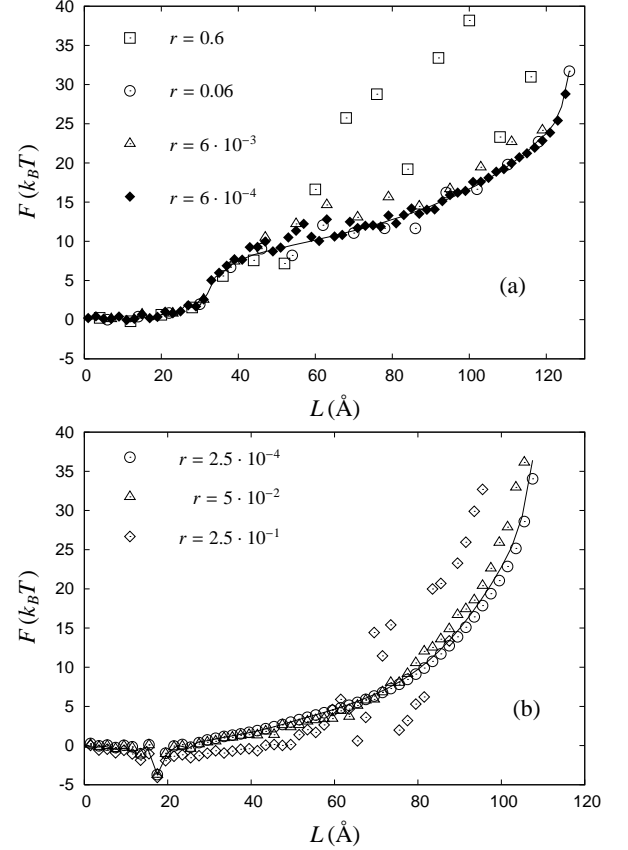


FIG. 11: Reconstructed free energy landscape F of the PIN1 (a), and the 1BBL (b) for $T = 300$ K. The lines correspond to the exact result.

tally accessible coordinate, namely the molecular elongation. By comparing the unfolding time with the expected values, as given by the Kramers' formula, we find that, in the limit of small forces, the free energy landscape (7) represents the *kinetic* energy landscape which governs the unfolding kinetics of the molecules, as discussed in ref. [41]. Furthermore, by comparing the typical width of the energy barrier, as found from eq. (7) and listed in table I, with the unfolding length obtained from out-of-equilibrium unfolding experiments, listed in tables II–III, we conclude that the latter parameters are not related to any typical length in the molecules, but are rather effective parameters.

By considering pulling at different velocities, we found that the unfolding length varies with the pulling velocity: thus our results support the point of view that different escape paths exist for a given molecule, each path being selected by the features of the manipulation.

Finally, the possibility of computing the free energy landscape as an exact result, make our model an excellent test bed to check the application of the fluctuation relations to the study of the equilibrium properties of biomolecules. Our results indicates that the free energy landscape can be recovered by out-of-equilibrium manipulations, and the collapse of the reconstructed curve represents an effective criterion to evaluate the reliability of the results.

In conclusion, we believe that our model is an useful tool to investigate the mechanical unfolding of proteins. We plan to further extend our work by studying the single folding and unfolding paths of proteins, by analyzing the unfolding kinetics of single substructures, so as to compare the results with experimental results.

Acknowledgments

We thank A. Szabo for interesting discussions and J. Klafter for his interest in our work.

APPENDIX A: EVALUATING THE FREE ENERGY LANDSCAPE FROM THE EQUILIBRIUM EXACT SOLUTION

In this appendix we discuss how the summation in the partition function (6) can be performed exactly, so as to obtain the free energy landscape as a function of the molecule length (7). A key point in our approach is to specify a (finite) discrete set of values for the length of a molecule. This is of course not a dangerous assumption, since atomic coordinates in the pdb are given with a finite resolution and it is therefore fair to round the distances l_{ij} (measured in Å) to rational numbers with a finite number of digits. In the applications reported in the present paper we use the resolution 10^{-3} Å. For the sake of simplicity, in the following we will adopt a length unit such that the distances l_{ij} are integer numbers. Notice also that their absolute value is not greater than $L_{\max} = \sum_{i=0}^N l_{ii+1}$, which corresponds to the length of the molecule in the completely unfolded, fully stretched configuration.

We compute the partition function (6) within a recursive scheme, considering sequences of sub-chains made of peptide bonds from 1 to $n \leq N$. For the sub-chain with n bonds we define the interaction energy

$$E_n(m) = - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \epsilon_{ij} \Delta_{ij} \prod_{k=i}^j m_k \quad (\text{A1})$$

and the length

$$L_n(m, \sigma) = \sum_{i=0}^n \sum_{j=i+1}^{n+1} l_{ij} \sigma_{ij} S_{ij}(m). \quad (\text{A2})$$

Notice that $E_n(m) - fL_n(m, \sigma)$ corresponds to the Hamiltonian (3). We will also need the reduced partition function

$$\Xi_n(f) = \sum_m \sum_{\sigma} \exp[-\beta E_n(m) + \beta f L_n(m, \sigma)], \quad (\text{A3})$$

where the first summation is over the first n peptide bond variables and the second one is over the orientations of the corresponding stretches.

Since the length values are integer by choice, we can expand $\Xi_n(f)$ in powers of $e^{\beta f}$ as

$$\Xi_n(f) = \sum_{L=-L_{\max}}^{+L_{\max}} \mathcal{Z}_n(L) e^{\beta f L}. \quad (\text{A4})$$

Our goal is thus the evaluation of $\mathcal{Z}_N(L)$, which corresponds to the partition function $Z_0(L)$ (6) in the main text.

Using the identity

$$1 = 1 - m_n + \sum_{i=1}^n (1 - m_{i-1}) \prod_{k=i}^n m_k, \quad (\text{A5})$$

one can verify the recursive relations

$$\Xi_n(f) = 2 \sum_{i=1}^{n+1} \cosh(\beta f l_{i-1n+1}) A_n^i(f), \quad (\text{A6})$$

and

$$A_n^i(f) = \begin{cases} \exp\left[\beta \sum_{k=i}^n \epsilon_{kn} \Delta_{kn}\right] A_{n-1}^i(f) & \text{if } i \leq n; \\ \Xi_{n-1}(f) & \text{if } i = n+1. \end{cases} \quad (\text{A7})$$

where, for $i \in \{1, \dots, n+1\}$,

$$A_n^i(f) \doteq \sum_m \sum_{\sigma} (1 - m_{i-1}) \prod_{k=i}^n m_k \cdot \exp[-\beta E_n(m) + \beta f L_n(m, \sigma)]. \quad (\text{A8})$$

where the sums over m and σ run as in eq. (A3). In the case $n = 0$ we set $\Xi_0(f) = 2 \cosh(\beta f l_{11})$ and $A_0^1(f) = 1$.

Expanding A_n^i in powers of $e^{\beta f}$ as

$$A_n^i(f) = \sum_{L=-L_{\max}}^{+L_{\max}} a_n^i(L) e^{\beta f L}, \quad (\text{A9})$$

we see that the coefficients of the above expansion satisfy

$$a_n^i(L) = \begin{cases} \exp\left[\beta \sum_{k=i}^n \epsilon_{kn} \Delta_{kn}\right] a_{n-1}^i(L) & \text{if } i \leq n; \\ \mathcal{Z}_{n-1}(L) & \text{if } i = n+1, \end{cases} \quad (\text{A10})$$

with the initial condition $a_0^1(L) = \delta(L)$. In addition we have

$$\mathcal{Z}_n(L) = \sum_{i=1}^{n+1} \left[a_n^i(L - l_{i-1n+1}) + a_n^i(L + l_{i-1n+1}) \right]. \quad (\text{A11})$$

Notice that the parity with respect to L can be exploited to reduce the above scheme to positive L values.

Finally, we want to stress the importance of the integer character of the microscopic lengths l_{ij} . We do not know *a priori* the values that the length of the molecule can assume, so our algorithm needs to span the whole interval $[0, L_{\max}]$, that has thus to be a finite set. On the other hand, the above scheme can be viewed as a polynomial algorithm to find the values that L can assume: there are no configurations (m, σ) with length L when $\mathcal{Z}_N(L) = 0$.

-
- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, (Garland, New York, 2002).
- [2] M. S. Kellermayer, S. B. Smith, H. L. Granzier C. Bustamante *Science* **276**,1112 (1997).
- [3] M. Carrion-Vasquez *et al*, PNAS **96**, 3694 (1999).
- [4] M. Carrion-Vasquez *et al*, *Nat. Struct. Biol.* **10**, 738 (2003); H. Dietz, M. Rief PNAS **101**, 16192 (2004).
- [5] H. Dietz, F. Berkemeier, M. Bertz, M. Rief, PNAS **103**, 12724 (2006).
- [6] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, J. M. Fernandez, PNAS **98**, 468 (2001).
- [7] J.M. Fernandez, H. Li, *Science* **303**, 1674 (2004).
- [8] M. Schlierf, H. Li, J. M. Fernandez PNAS **101**, 7299 (2004).
- [9] C. Danilowicz *et al* PNAS **100**, 1694 (2003); J. Liphardt *et al*, *Science* **292**, 733 (2001) ; B. Onoa *et al* *Science* **299**, 1892 (2003).
- [10] D. K. Klimov, D. Thirumalai *Proc. Natl. Acad. Sci. U.S.A* **97** 7254 (2000).
- [11] M. S. Li, M. Kouza, C.-K. Hu, *Biophysical Journal* **92**, 1 (2007).
- [12] P. Szymczak, M. Cieplak, *J. Phys.: Condens. Matter* **18**, L21 (2006).
- [13] A. Imparato, S. Luccioli, A. Torcini, e-print arXiv:0705.3256.
- [14] A. Imparato, A. Pelizzola, M. Zamparo, *Phys. Rev. Lett.* **98**, 148102 (2007).
- [15] H. Wako, N. Saitô, *J. Phys. Soc. Jpn* **44**, 1931 (1978).
- [16] H. Wako, N. Saitô, *J. Phys. Soc. Jpn* **44**, 1939 (1978).
- [17] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997); C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997); G. E. Crooks, *J. Stat. Phys.* **90**, 1481 (1998); G. E. Crooks, *Phys. Rev. E* **61**, 2361 (1999).
- [18] G. Hummer, A. Szabo, *Proc. Natl. Acad. Sci. USA* **98**, 3658 (2001).
- [19] A. Imparato, L. Peliti, *J. Stat. Mech.-Theory Exp.* P03005 (2006).
- [20] M. Sadqi, D. Fushman, V. Muñoz, *Nature* (London) **442**, 317 (2006); F. Huang, S. Sato, T. D. Sharpe, L. Ying, A. Fersht, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 123 (2007).
- [21] N. Ferguson, T. D. Sharpe, C. M. Johnson, P. J. Schartau, A. R. Fersht, *Nature* (London) **445**, E14 (2007); Z. Zhou and Y. Bai, *Nature* (London) **445**, E16 (2007); M. Sadqi, D. Fushman, V. Muñoz, *Nature* (London) **445**, E17 (2007).
- [22] V. Muñoz, P.A. Thompson, J. Hofrichter, W.A. Eaton, *Nature* **390**, 196 (1997).
- [23] V. Muñoz, E.R. Henry, J. Hofrichter, W.A. Eaton, *Proc. Natl. Acad. Sci. USA* **95**, 5872 (1998).
- [24] V. Muñoz, W.A. Eaton, *Proc. Natl. Acad. Sci. USA* **96**, 11311 (1999).
- [25] A. Flammini, J.R. Banavar, A. Maritan, *Europhys. Lett.* **58**, 623 (2002).
- [26] I. Chang, M. Cieplak, J.R. Banavar, A. Maritan A. *Protein Sci.* **13**, 2446 (2004).
- [27] K. Itoh K. M. Sasai, *Proc. Natl. Acad. Sci. U.S.A* **101**, 14736 (2004); K. Itoh, M. Sasai, *Proc. Natl. Acad. Sci. USA* **103**, 7298 (2006).
- [28] H. Abe, H. Wako, *Phys. Rev. E* **74**, 011913 (2006).
- [29] V.I. Tokar, H. Dreyssé *Phys. Rev. E* **68**, 011601 (2003); V. I. Tokar, H. Dreyssé, *J. Phys. Cond. Matter* **16**, S2203 (2004); V.I. Tokar, H. Dreyssé, *Phys. Rev. E* **71**, 031604 (2005).
- [30] P. Bruscolini, A. Pelizzola *Phys. Rev. Lett.* **88**:258101 (2002).
- [31] A. Pelizzola, *J. Stat. Mech.-Theory Exp.* P11010 (2005).
- [32] M. Zamparo, A. Pelizzola, *Phys. Rev. Lett.* **97**, 068106 (2006).
- [33] M. Zamparo, A. Pelizzola, *J. Stat. Mech.-Theory Exp.* P12009 (2006).
- [34] P. Bruscolini , A. Pelizzola, M. Zamparo, *J. Chem. Phys.* **126**, 215103 (2007).
- [35] P. Bruscolini , A. Pelizzola, M. Zamparo to appear in *Phys. Rev. Lett.*.
- [36] E. Evans, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105 (2001).
- [37] R. Zwanzig, *Nonequilibrium Statistical mechanics*, Oxford University Press, 2001.
- [38] I. Derényi, D. Bartolo, A. Ajdari, *Biophys. J.* **86**, 1263 (2004).
- [39] D. Bartolo, I. Derényi, A. Ajdari, *Phys. Rev. E* **65**, 051910 (2002).
- [40] A. Imparato, L. Peliti, *Eur. Phys. J. B* **39**, 357 (2004).
- [41] E. Evans, K. Ritchie *Biophys. J.* **72**, 1541 (1997).
- [42] O. K. Dudko, G. Hummer, A. Szabo *Phys. Rev. Lett.* **96**, 108101 (2006).
- [43] O. K. Dudko *et al.*, *Proc. Natl. Acad. Sci. U.S.A* **100**, 11378 (2003).
- [44] A. M. Ferrenberg, R. H. Swendsen, *Phys. Rev. Lett* **63**, 1195 (1989).