Native-state fingerprint on the ubiquitin translocation across a nanopore

Fabio Cecconi
Istituto dei Sistemi Complessi (CNR), Via Taurini 19, I-00185 Roma, Italy

Mauro Chinappi
Dipartimento di Ingegneria Industriale, Università di Roma Tor Vergata, Roma I-00133, Italy

(Received 2 August 2019; accepted 11 February 2020; published 2 March 2020)

We study the translocation of the ubiquitin molecule (Ubq) across a channel with a double section which constitutes a general feature of several transmembrane nanopores such as the α-hemolysin (αHL). Our purpose is to establish the structure-dependent character of the Ubq translocation pathway. This implies to find the correspondence, if any, between the translocational unfolding steps and the Ubq native state. For this reason, it is convenient to apply a coarse-grained computational approach, where the protein is described only by the backbone and the force field only exploits the information contained in the native state (in the spirit of Gō-like models, or native-centric models). The αHL-like pore is portrayed as two coaxial confining cylinders: a larger one for the vestibule and a narrower one for the barrel (or stem). Such simplified approach allows a large number of translocation events to be collected by limited computational resources. The co-translocational unfolding of Ubq is described via a few collective variables that characterize the translocation progress. We find two translocation intermediates (stalled conformations) that can be associated with specific unfolding stages. In particular, in the earliest step, the strand S5 unfolds and enters the pore. This step splits the native conformation into two structural clusters packing against each other in the Ubq fold. A second stall occurs when the hairpin of the *N* terminal engages the stem region.

DOI: 10.1103/PhysRevE.101.032401

I. INTRODUCTION

The fast development of nanoscale technology suggests the employment of nanopores as key elements in devices for single biopolymer analysis. Information on the passing molecule is collected by recording how its passage across a nanopore perturbs the free-pore ion current [1–4]. This principle works so remarkably well on DNA that it lays at the core of a well-established high-throughput DNA sequencing technique [5,6]. Unlike DNA, the application of nanopore sensing to proteomics still presents several criticalities to solve [7–9] and, despite recent encouraging progress [10–14], protein sequencing still remains a very challenging task.

Proteins, in fact, have complex structures that can strongly affect the translocation (co-translocation unfolding). Moreover, polypeptides host simultaneously positive and negative charges, and typically their net charge may not be sufficient to generate an easy electrophoretic transport across the nanopore. Also, this often makes it very hard to capture the proteins from the solution to the pore. Since the pioneering work of Oukhaled and co-workers [15], which showed the possibility of translocating proteins in nanopores upon strong denaturation, efforts have multiplied to make protein sequencing by nanopore devices a viable alternative to the standard methods. Obviously, the first step begins from the improvement of the capture process, which is a crucial stage for protein sensing. Protein molecules, indeed, due to their well-known tendency to assume a three-dimensional compact structure, develop resistance against capture in narrow pores. A viable method to enhance the capture events exploits electro-osmotic flows [16–19], while alternative techniques employ charged leaders [20] or molecular motors [21], both to improve the capture rate of proteins and to control their translocation through nanopores. In both cases [20,21], the protein is pulled from one terminal and, interestingly, the resulting co-translocational unfolding pathway turns to be a multistep process. In particular, Ref. [20] interprets the ion-current multilevel as the fingerprint of translocation intermediates, i.e., conformations where proteins remain stuck in the nanopore. This phenomenology is also confirmed by theoretical and computational works [22–28], where the unraveling of proteins during the translocation gives rise to a sequence of stalled and running states.

The multistep (or multistage) character of protein translocation in nanopores has a double implication because, on the one hand, it hinders the sequencing process as the transport does not proceed steadily enough to allow a regular readout of the amino acids. On the other hand, it offers the opportunity to extract information on the translocating molecule by analyzing the sequence of intermediates that it generates along the pathway.

The perspective to map the sequence of stalled states into structural information about the passing protein is, in principle, reasonable, even if it remains an open issue due to the lack of conclusive theoretical, computational, and experimental tests. In principle, filling this gap should allow the development of protocols that are capable, at least, of inferring certain topology properties of protein structures by nanopore analysis.

Demonstration, on a rigorous ground, of to what extent the multistep translocation is actually related to the protein structure is a hard task which can be faced by starting from the simplest cases. In this respect, the ubiquitin (Ubq), for its simple native-state topology, represents the ideal candidate to test this scenario. On the other hand, a brief survey of the literature indicates that Ubq has always been considered a paradigm for experimental and theoretical studies on protein folding and dynamics [29], also in the context of force spectroscopy techniques [30–37]. In addition, Ubq is an excellent model for testing theoretical methods on translocation in nanopores, via all atoms simulations [25,28] as well as by coarse-grained computational approaches [26,38].

In this work, we employ a coarse-grained computational model of Ubq to collect indications on whether the multistep translocation of Ubq reflects some features of its native-state structure. In the same coarse-grained philosophy, the α-hemolysin (αHL) nanopore is also portrayed as a simple confining pipe with two sections emulating its known vestibule-stem architecture. In order to asses the role of the constriction in the αHL-shaped nanopore, we compare the translocation of Ubq in αHL with its translocation across a single section channel, which we termed the nanotubelike (NT) channel.

Our simulations reveal that Ubq translocation is structure dependent as it is practically determined by the resistance that each structural element opposes to its unfolding from the Ubq fold. This resistance turns into translocation intermediates (stalled conformations); in particular, Ubq reveals a major stall when, at the entrance of the pore, it splits into two structural clusters packing against each other in the native state, and identified in Ref. [30].

The paper is outlined as follows. In Sec. II, we show the basic ingredients of our coarse-grained description of the Ubq translocation across a toy-model nanopore mimicking the αHL nanopore. In Sec. III, we discuss the results of the simulations, focusing on the translocation intermediates and their relationship with the geometrical properties of the Ubq native state. In Sec. IV, we attempt the interpretation of the translocation pathway in terms of a simple reaction model. Finally conclusions can be found in Sec. V.

II. COARSE-GRAINED MODEL OF PROTEIN TRANSLOCATION

The Ubq is modeled by a Gō-like force field proposed by Clementi *et al.* [39]. The details about force-field parametrization and implementation can be found in Refs. [39,40]. The chain is represented by only taking into account Cα atom positions since we are mainly interested in the structural rearrangements of the backbone along the translocation pathway. We recall that Gō models [41] are such that the energy function is minimum on the coordinates of the crystallographic structure of the native state. In the present work, such coordinates are extracted from the Protein Data Bank (PDB) entry 1UBI refined at 1.8 Å resolution [42]. A simple way to ensure that the native structure is a minimum of the potential energy is through the notion of native interactions, or contacts [43]. In this work, we consider two residues i and j in native interaction if they share a couple of heavy atoms, i.e., all the

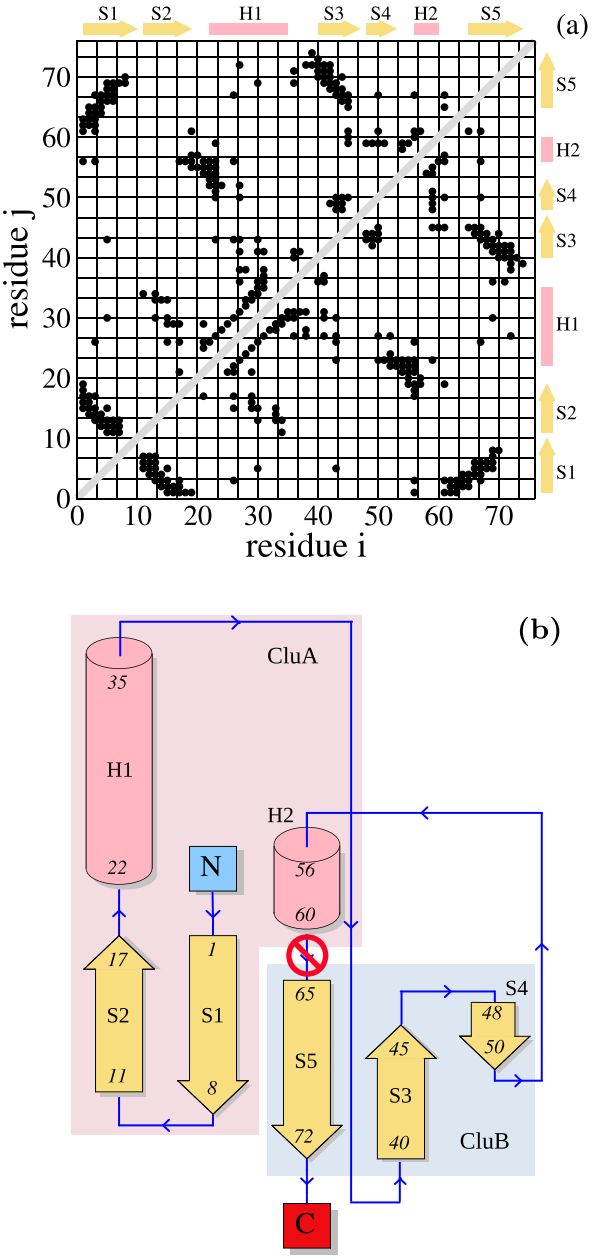

FIG. 1. (a) Heavy-atom contact map of the Protein Data Bank (PDB) structure 1UBI [42]; a cutoff of $R_c = 5$ Å selects $M = 190$ native contacts represented as full dots. (b) The secondary-structure graph displaying five beta strands and two helices. The red signal between H2 and S5 marks the major stalling point encountered in the translocation.

atoms but hydrogens and nitrogens, within a cutoff distance $R_c < 5$ Å. The resulting contact map is reported in Fig. 1, showing the 190 contacts; in Fig. 1(b), we also show the topology graph describing the secondary structure of Ubq. The secondary content of the native Ubq includes two helices, H1 residues (22–35) and H2 (56–60), and short 3_{10} helix 37–39; moreover, five β-strands: S1 (1–8), S2 (11–17), S3 (40–45), S4 (48–50), and S5 (65–72) complete the structure.

The interactions between the beads are associated with peptide bonds, angular bending, torsional deformation, and

native contacts (see [39,40]), leading to the energy function for an N residue protein,

$$\Phi_{G\bar{o}} = \sum_{i=1}^{N-1} V_p(r_{i,i+1}) + \sum_{i=1}^{N-2} V_\theta\left(\theta_i - \theta_i^0\right) + \sum_{i=1}^{N-3} V_\varphi\left(\varphi_i - \varphi_i^0\right) + \sum_{i,j\geqslant i+3} V_{nb}(r_{ij}). \quad (1)$$

The peptide bond term V_p, enforcing the chain connectivity, is a stiff harmonic potential allowing only small oscillations of the bond lengths around their crystallographic values. Likewise, the bending potential V_θ allows only small fluctuations of the bending angles θ_i around their native values θ_i^0. Dihedral potential V_ϕ (associated to torsional deformation) further contributes to the correct formation of the native secondary structure that is also characterized by angles φ_i^0. Finally, the long-range potential V_{nb}, which favors the formation of the correct native tertiary structure, is a collection of two-body 12–10 Lennard-Jones contributions that are attractive between a couple of residues forming native contacts and repelling for non-native couples.

The energy scale of the model is set by a single parameter ϵ [39], which has to be determined by the knowledge of the experimental data; see below.

The confining effect of the αHL nanopore is described, on average, by the following potential with cylindrical symmetry acting in the pore region only, $0 < x < L$:

$$U_p(\mathbf{r}) = 2\epsilon \begin{cases} 0, & y^2+z^2 \leqslant R^2(x) \\ \left[\dfrac{y^2+z^2}{R^2(x)} - 1\right]^m, & y^2+z^2 > R^2(x)\,. \end{cases} \quad (2)$$

Let $r = y^2 + z^2$ be the radial distance from the pore axis (x axis); then the potential (2) provides a vanishing contribution to the interior of the pore [$r \leqslant R(x)$]. On the contrary, it becomes strongly repulsive, as a power of $2m$, when residues try to cross the pore boundary [$r > R(x)$]. The value $m = 4$ is a compromise between the stiffness of the force and the stability of the code. For roughly reproducing the vestibule-barrel shape of the αHL [see Fig. 2(a)], the pore radius is modulated as

$$R(x) = \frac{R_0 + R_1}{2} - \frac{R_0 - R_1}{2} \tanh[\alpha(x - x_c)].$$

$R(x)$ has a sigmoidal shape interpolating between the average radius of the vestibule, $R_0 = 10$ Å, and the radius of the constriction, $R_1 = 4$ Å; moreover, $L \simeq 100$ Å (αHL length). We choose $\alpha = 3$ Å^{-1} to have a steep sigmoid to emulate the almost sudden junction between the vestibule and the barrel in the αHL structure. Again, looking at the αHL geometry, we set $x_c = L/2$ for the position of the constriction, which in the real protein lies approximately on the middle of the channel length. Although we tried to respect the mean sizes of the αHL structure, our channel has been shaped that way to generate a sort of "constriction;" in this respect, it has to be consider only a "toy" version of the true pore. For comparison, we also analyzed the translocation in a nanotubelike pore (NT) where the diameter of the channel is constant and equal to R_0.

In hard coarse-graining approaches, it is customary to represent the membrane, rather than a collection of atoms, with a continuous potential describing the interaction of biopolymers or particles with the membrane walls [45–48]. Following this strategy, we introduce a repulsive short-range force $F_w(x)$, orthogonal to the plane $x = 0$ ($x = L$) [see Fig. 2(a)] and vanishing for $y^2 + z^2 < R^2(x)$. It models the presence of the impenetrable membrane hosting the pore and preventing the protein from entering except through the hole,

$$F_w(x) = \epsilon \begin{cases} -\dfrac{e^{\lambda x}}{x + c}, & x \leqslant 0 \\ 0, & 0 < x < L \\ \dfrac{e^{-\lambda(x-L)}}{x - L + c}, & x \geqslant L. \end{cases} \quad (3)$$

Expression (3) represents a force which is the product of a factor $1/x$ becoming very intense as it approaches the wall, and $\exp(-\lambda|x|)$, which instead decreases very fast away from it to make the repulsion as short ranged as possible. The parameter c is a regularization cutoff preventing numerical overflow. In other terms, Eq. (3) implements, without dramatic singularities, the presence of a repulsive contact force imposed by the steric constraint due to the impenetrability of the membrane. Preliminary numerical tests provided the values $\lambda = 6.0$ Å^{-1} and $c = 10^{-4}$ Å as reasonable.

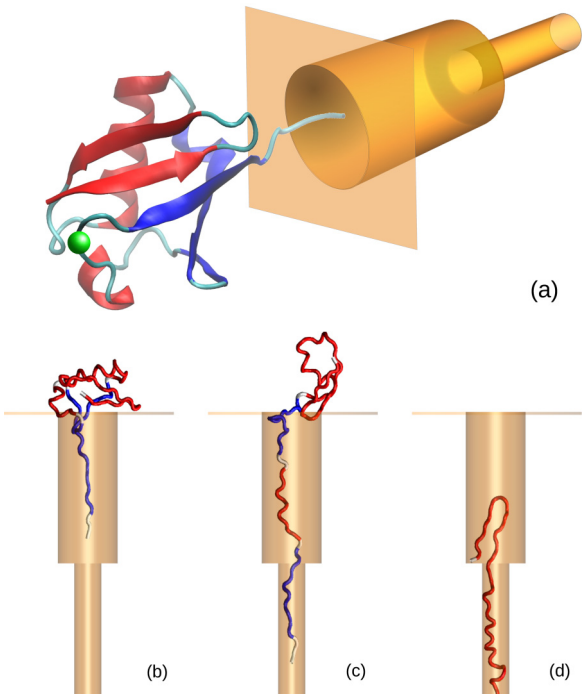

FIG. 2. (a) Ubiquitin native-state cartoon representation with the cylindrical channel shaped in two sections to represent the αHL vestibule-stem structure. Colors identify the two structural clusters of the Ubq fold: CluA (blue) composed of strands S1, S2 and helices H1, H2; CluB (red) formed by strands S3, S4, S5. The green bead marks residue 62, a kind of hinge, separating these two clusters, which is also the site whereby the translocation becomes stalled. (b)–(d) Three stages of Ubq translocation corresponding to (b) the entrance of strand S5, (c) the relevant stalling point characterized by the opening and separation of CluA from CluB, and (d) the translocation of the hairpin S1 and S2, producing a secondary stalling. Images have been drawn using the VMD software [44].

Conversion of the code units into physical units. The knowledge of the experimental $T_f(\mathrm{exp})$ and simulated $T_f(\mathrm{sim})$ unfolding temperatures allows us to determine the energy scale of the Gō model via the simple relation $\epsilon = RT_f(\mathrm{exp})/T_f(\mathrm{sim})$, where $R = 8.31\ \mathrm{J\,K^1\,mol^{-1}}$ is the gas constant. Preliminary thermal unfolding simulations provided $T_f(\mathrm{sim}) = 1.1$ in code units, but, from Ref. [49], we know that for the Ubq $T_f(\mathrm{exp}) \simeq 373$ K, this information set a value $\epsilon \simeq 3.07$ kJ/mol. Once the value of ϵ is known, all the other units can be converted by considering that the Ubq has a mass of 8580 Da, and therefore the resulting average amino acid mass is $M_{aa} = 1.875 \times 10^{-25}$ Kg. This implies that the time unit is $t_u \simeq 3$ ps and the unit of force is $f_u \simeq 10$ pN.

Simulation protocol

The importing mechanism that drives the protein into the pore is simplified to a constant pulling force $(F, 0, 0)$ acting only on the C-terminus bead ($\mathbf{r}_{76}$), which is constrained to slide along the pore axis, $y_{76}(t) = z_{76}(t) = 0$, at every time. To control the temperature, simulations were performed by using a Langevin molecular dynamics,

$$M_{aa}\ddot{\mathbf{r}}_i = -\gamma\dot{\mathbf{r}}_i - \nabla_{r_i}(\Phi_{G\bar{o}} + U_p) + \mathbf{F}\delta_{i,N} + \mathbf{F}_w(x_i) + \mathbf{Z}_i \tag{4}$$

$(i = 1, \ldots, N = 76)$, which is integrated according to a stochastic LEAP-FROG algorithm [50]. In Eq. (4), M_{aa} denotes the average amino acid mass (in the code, $M_{aa} = 1$), $\mathbf{Z}_i$ is a random force with zero average and correlation $\langle Z_{i,\mu}(0)Z_{i,\nu}(t)\rangle = 2\gamma k_B T\delta_{\mu,\nu}\delta(t)$, with $\mu, \nu = x, y, z$ and k_B the Boltzmann's constant ($k_B = 1$). $\mathbf{F}\delta_{i,N}$ is the force of modulus F, pulling only the C-terminus residue ($N = 76$). Finally, $\Phi_{G\bar{o}}$, U_p, and $\mathbf{F}_w$ are given by (1), (2), and (3), respectively.

After several preliminary tests, we used a friction $\gamma = 1.0M_{aa}/t_u$ and a time step $h = 0.002t_u$. The temperature $T = 0.7$ ($T = 336.8$ K) was chosen as a compromise between compactness and floppiness of the simulated conformations to get translocations in reasonable times.

Each translocation run started by positioning the native structure with the C terminus at $\mathbf{x}_{76} = -(10, 0, 0)$ Å with respect to the pore entrance, and thermalizing for $t_{eq} = 10^4$ time steps with the C terminal blocked. Then, the forcing is turned on and the simulation is stopped when all the residues reach the trans side ($x_i > L$).

III. SIMULATION RESULTS

We run a set of $M = 3600$ independent runs, from the native Ubq conformation. The system was restarted from the native state as soon as a full-translocation event had taken place.

The visual inspection of several samples of translocation trajectories reveals that translocation is simply a sorting process of the secondary elements from the C terminus, and therefore the sequence of events

$$C \to S5 \to H2 \to S4 \to S3 \to H1 \to (S1 - S2) \to N$$

reproduces the scanning of the topology graph [Fig. 1(b)] in the reverse order from the C to the N terminus. The notation (S1-S2) indicates that often strands S1 and S2 migrate together. However, three interesting features emerge from the qualitative analysis of the trajectories [see Figs. 2(b)–2(d)]:

(i) The C-pulling translocation is triggered by the entrance of the strand S5 [Fig. 2(b)], but the transport does not begin until the contacts that bound the strands S1 to S5 break down. The separation of S1 and S5 is crucial to allow the splitting of the Ubq into two parts coinciding with two well-defined structural clusters that are packing against each other in the native fold [see Figs. 1(b) and 2(a)]. The first cluster CluA (red) includes the strands S1, S2 and the helices H1, H2, while the second, CluB (blue), includes S3, S4, and S5. The unfolding of S5 in turn implies the sudden unfolding of CluB, which is now free to translocate, while the CluB remains almost intact, offering residual resistance to the translocation [Fig. 2(c)]. In the literature, the split of the Ubq into two halves has already been conjectured in free pulling experiments of poly-Ubq by Schlierf *et al.* [30] and seen by Irback *et al.* [32] in simulations of single Ubq pulling.

(ii) There is an important stalling of the transport between elements S3 and H2 corresponding to the molecule rearrangements described in (i) and sketched in Fig. 2(c).

(iii) The strands S1 and S2 form a very stable hairpin, which often translocates as a single object, violating the single file mode [Fig. 2(d)].

To get a more quantitative characterization of the Ubq translocation, we first compute the dependence of the average translocation time,

$$\tau(F) = \frac{1}{M}\sum_{k=1}^{M} t_k(F),$$

on the strength F of the pulling force, where each t_k is computed as the first time after which all the Ubq residues lie outside the pore in the Trans side. The results are reported in Fig. 3 for a nanotubelike pore (NT) and for the αHL-shaped channel.

An interesting difference emerging from the comparison of the two data sets is that the αHL-like channel seems to be more sensitive to the force variation than the NT-like channel, as can be appreciated from the larger range of τ values.

The $\tau - F$ curves are customarily interpreted in terms of the Bell's formula [51], i.e., actually a simple application of the Kramers' theory [52] to systems under tension, which predicts the exponential behavior

$$\tau(F) \sim \exp\{-\beta[G(x_b) - Fx_b]\} = \tau(0)e^{-\beta F x_b}, \tag{5}$$

where $G(x_b)$ indicates the free-energy barrier separating the Cis and Trans states, x_b is the coordinate of the barrier (transition state) that is supposed to remain practically unchanged with the application of a constant force F, and $\tau(0)$ is the lifetime of the native state in the absence of the force. Equation (5) states that the application of a force exponentially reduces the persistence in the native state. As the inset of Fig. 3 shows, the simulated data deviate from Bell's theory; rather, the data are reasonably fitted by a two-exponential expression,

$$\tau(F) = c_1 e^{-\beta F x_1} + c_2 e^{-\beta F x_2}, \tag{6}$$

where $x_1 = 17.88$ Å and $x_2 = 2.85$ Å are two elongation scales presumably characterizing two different translocation pathways and recalling the scenario found in the mechanical

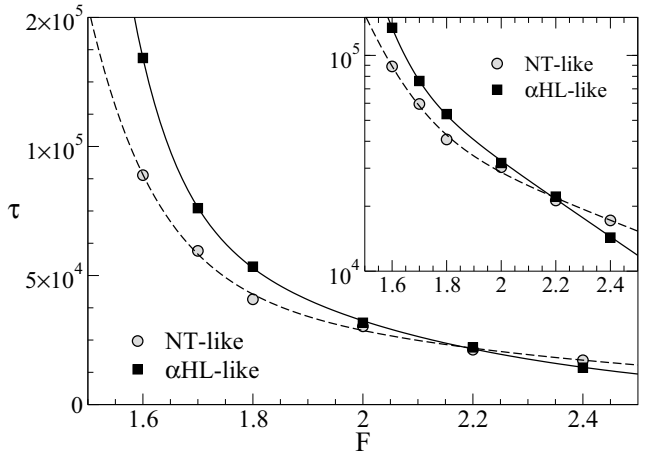

FIG. 3. Average translocation time τ as a function of the intensity F of the pulling force [see Eq. (4)]. Points are the simulation data, for the NT-like (empty circles) and the αHL-like shaped pore (filled squares). The lines are the double-exponential fitting by formula (6). The values of the fitting parameters for the NT-like nanopore are $c_1 = 2.84 \times 10^{10}$, $x_1 = 11.9$ Å, $c_2 = 2.64 \times 10^5$, and $x_2 = 1.62$ Å, while for the αHL-like nanopore, the parameters are $c_1 = 3.16 \times 10^{13}$, $x_1 = 17.9$ Å, $c_2 = 1.72 \times 10^6$, and $x_2 = 2.8$ Å. The inset reports the same plot in linear-logarithmic scale, showing that the data deviate from the simple exponential behavior predicted by Bell's theory [51].

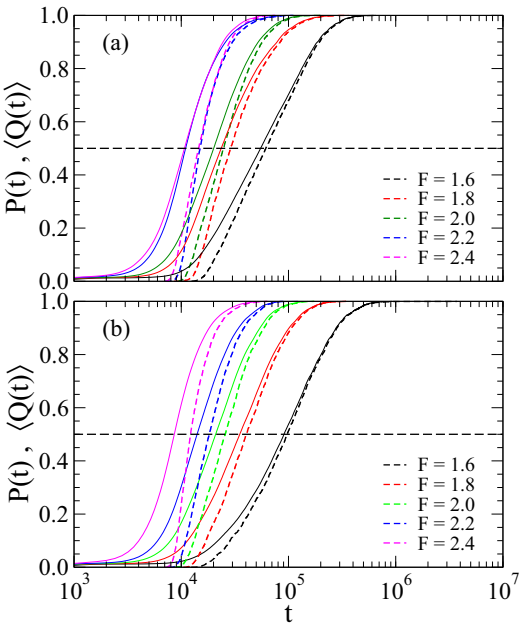

FIG. 4. Translocation probability (dashed line) $P(t)$ vs time at different strengths of the pulling forces, for a set of $M = 3600$ independent translocation events. On the same graph, we also plot the average $\langle Q(t)\rangle$ (solid line), given by Eq. (9). (a) The NT-like pore; (b) the αHL-like pore.

pulling experiments of Ubq [30] where two lengths, $\ell_1 = 80$ Å and $\ell_2 = 200$ Å, were measured. However, a simple correspondence between (x_1, x_2) and (ℓ_1, ℓ_2) is difficult to establish, presumably because there is not a precise mapping between a stretching protocol in the free space and a stretching process into a pore of a membrane. The presence of pores and membranes, indeed, forbids certain rearrangements of the chain, which are instead possible in the free space.

In addition, the approximations introduced by the coarse-grain modeling must also be considered.

The deviation from the Bell's formula suggests that the Ubq translocation cannot be reduced to a simple two-state process from Cis → Trans as required by the Bell's theory [51]. The barrier separating the Cis and Trans state is not dominated by a sharp maximum, but rather is an extended profile, as it often happens in polymer translocation [38,53].

The double-exponential law of $\tau(F)$ can be interpreted as an indication of the presence of at least one intermediate along the translocation pathway. As mentioned in Sec. I, such intermediates are generated by the peculiar way the mechanical unfolding of Ubq couples to the transport across the nanopore. Because the mechanical unfolding is strongly determined by the breaking sequence of those interactions that stabilize the native state of the Ubq, it is reasonable to expect that the translocation pathways are strongly influenced, if not even dictated, by the structure of the Ubq native structure.

We then determined the probability $P(t)$ that Ubq translocates in the time interval $[0, t]$. By definition, $P(t)$ turns to be the cumulative distribution functions (CDFs) of translocation times. For a sample $S = \{t_1 < t_2 < \cdots < t_M\}$ of translocation times that are supposed to be sorted in an ascending way, $P(t)$ is given by plotting the normalized rank k/M against the corresponding t_k. This method has the obvious advantage of being binning independent. The data of $P(t)$ are reported in Fig. 4. To get a more detailed picture of the unfolding process of the Ubq native backbone during its transport into the nanopore, we monitor three indicators (reaction coordinates) that characterize the collective dynamics of the Ubq and describe the progress of the translocation. The first one is

$$N_{\rm cis}(t) = N - \sum_{i=1}^{N} \Theta[x_i(t)], \tag{7}$$

while the second one is

$$N_v(t) = N - \sum_{i=1}^{N} \Theta[x_i(t) - L/2], \tag{8}$$

where $\Theta[s]$ is the unitary step function. $N_{\rm cis}$ and N_v count the number of residues that have not yet entered the pore vestibule ($N_{\rm cis}$) and the barrel (N_v), respectively. In essence, at the initial condition, $N_{\rm cis} = N_v = 76$, i.e., all of the 76 amino acids lie outside the pore at the Cis side. As long as the translocation proceeds, both variables decrease to zero in a way that should depend on the bottlenecks encountered along the pathway, and the highest peaks in the histogram of these variables locate the putative translocation bottlenecks.

The last collective variable that we monitor is

$$Q(t) = \frac{1}{N} \sum_{i=1}^{N} q[x_i(t)], \tag{9}$$

defined by the piecewise function [54],

$$q(x) = \begin{cases} 0, & \text{if } x < 0 \\ x/L, & \text{if } 0 \leqslant x < L \\ 1, & \text{if } x \geqslant 1. \end{cases} \quad (10)$$

The value $Q = 0$ corresponds to the whole protein on the cis side, while $Q = 1$ indicates that a translocation is completed.

The time course of both $N_{\text{cis}}(t)$ and $Q(t)$ on a single realization is a random fluctuating quantity that, however, already reveals the presence of stalled conformations due to intermediates. To avoid strong fluctuations, both indicators have been averaged,

$$\langle Q(t)\rangle = \frac{1}{M}\sum_{r=1}^{M} Q_r(t), \quad \langle N_{\text{cis}}(t)\rangle = \frac{1}{M}\sum_{r=1}^{M} N_{\text{cis},r}(t), \quad (11)$$

over $M = 3600$ independent successful translocation runs.

The average $\langle Q(t)\rangle$ is reported in Fig. 4 as full lines together with the translocation probability $P(t)$ (dashed lines), at different values of the force and for both αHL-like and NT-like pores. It is apparent that these two quantities convey more or less the same information: they both indicate how the Ubq translocation occurs in time, however, their clean sigmoidal shape suggests that $P(t)$ and $\langle Q(t)\rangle$ are not sensitive to the finer details of the dynamics, such as the presence of stalls. In this respect, the average of N_{cis} is a better indicator, as shown by Fig. 5, where $\langle N_{\text{cis}}(t)\rangle$ versus time is plotted at different values of the force and for both pore shapes, i.e., NT-like and αHL-like.

$\langle N_{\text{cis}}(t)\rangle$ starts from the obvious plateau $N_{\text{cis}} = 76$ (Ubq residues) and decreases to zero when translocation is successful. In the middle of the process, $\langle N_{\text{cis}}(t)\rangle$ exhibits an abrupt slope change, which means a sudden slow down of the dynamics. Figure 5 clearly shows the presence of a plateau, at $N_{\text{cis}} = 62$, which is shrinking with the growth of F. This plateau is associated with the presence of a major stall and it is important to remark that the picture is robust because it emerges from the average over a large sample of independent translocation events.

Interestingly, with reference to Figs. 1 and 2, the value $N_{\text{cis}} = 62$ exactly identifies the residue separating the two clusters CluA and CluB of the Ubq fold. Therefore, this result is a robust validation of the most probable pathway scenario: the C-pulling translocation of Ubq is triggered by the entrance of the strand S5, which, in turn, causes the split of Ubq into two parts coinciding with the two structural clusters A and B of the Ubq fold; see Fig. 2.

Figure 5 highlights a certain performance difference between the αHL-like and NT-like pores, where the former seems able to better discriminate the translocation events, since the curves in Fig. 5(b) are well separated at different fields. This does not happen for the NT-like pore where a certain overlap occurs. This feature has already been noted in the data of Fig. 3.

The histograms of N_{cis} collected over the M runs are plotted in Fig. 6 for different values of the pulling force F. The vertical dashed lines mark three important stages of the translocation. The pronounced peak around $N_{\text{cis}} = 62$ confirms that the most important stall occurs when the pulling mechanism

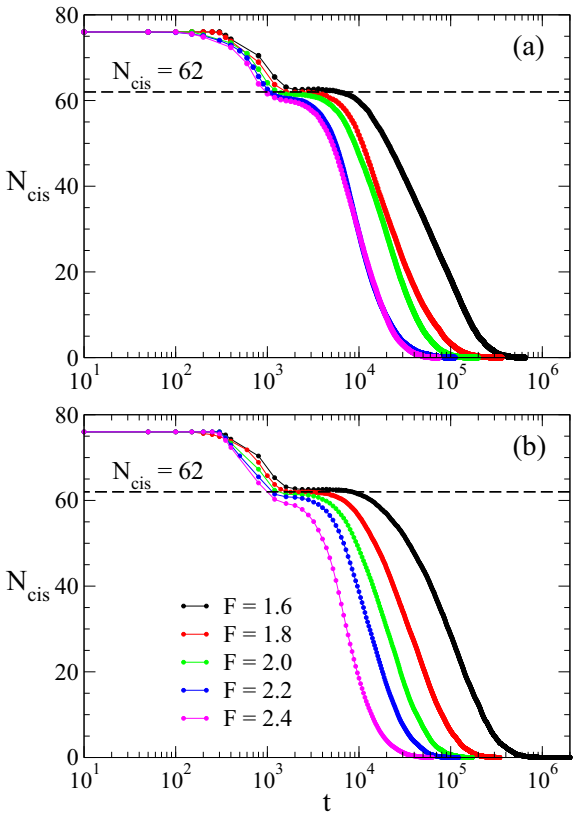

FIG. 5. Time evolution, in logarithmic-linear scale, of the average of N_{cis} over $M = 3600$ independent successful translocation runs. $\langle N_{\text{cis}}(t)\rangle$ develops a variation of the slope occurring around $N_{\text{cis}} = 62$. This identifies the abrupt slowing down of the dynamics that is systematic and robust because it emerges from native structure features. (a) NT-like pore; (b) αHL-like pore.

is able to split the clusters A and B of the Ubq. The histograms display a smaller peak around $N_{\text{cis}} = 73$, corresponding to the entrance of the strand S5 into the nanopore. Finally, the histograms at smaller forces suggest that something also happens around $N_{\text{cis}} = 17$, corresponding to the migration of the strand S2. It should be noted that the peak decreases as F grows, suggesting that to gain sensitivity, it is necessary to work at the lowest pulling rates, in a reasonable compromise between the sensitivity and the simulation or experiment duration length.

To analyze in detail the region $N_{\text{cis}} = 17$, we considered the histograms of N_v [Eq. (8)], counting the residues that have not yet entered the stem region. The data are reported in Fig. 7, showing indeed the peak around $N_v = 17$, indicating that another stalling point involves the simultaneous arrival at the constriction $x_c \sim L/2$, of strands S1 (2–8) and S2 (11–17) which form a resistant block. It can cross the stem either in double or single file conformations. Just to check that the last stalling point of the Ubq translocation is caused by the hairpin S1-S2 arriving at the constriction, we performed dedicated simulations starting from the Ubq conformations with the hairpin placed just before the stem, $x_c \sim L/2$. Then we measured N_v, and its histograms are reported in the lower panel of Fig. 7. The pronounced peak at $N_v = 17$ is strong evidence that a blocking occurs in correspondence of the

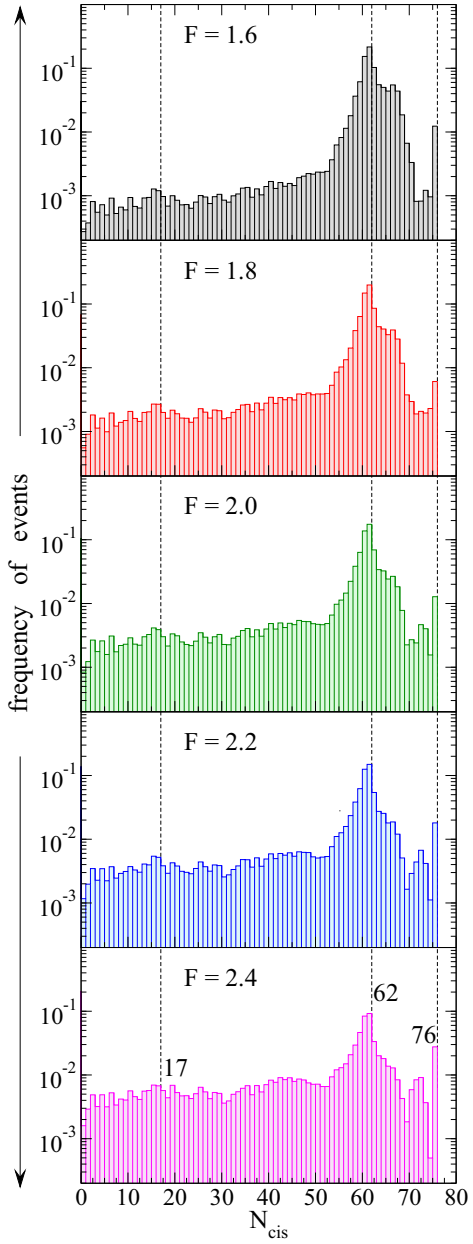

FIG. 6. αHL-like nanopore: Histogram of the indicator $N_{\rm cis}$ collected over 3600 C-pulling runs from the native conformation at different pulling force. The highest peak, $N_{\rm cis} = 62$, is associated to the unfolding of the first structural cluster that includes S1, S2, and H1. The other peak, $N_{\rm cis} = 76$, corresponds to the entrance of the strand S5. The dashed line at $N_{\rm cis} = 17$ marks some other feature involving the strand S2.

strand S2 involved in the hairpin. Notice that 17 is also the length in the sequence of the hairpin S1-S2.

Therefore, due to its stability, this hairpin is able to leave a signal at the entrance of both the vestibule and the constriction.

IV. INTERPRETATION AS A SIMPLE REACTION EQUATION

Simulation results suggest that the translocation dynamics involves at least one relevant intermediate, corresponding to

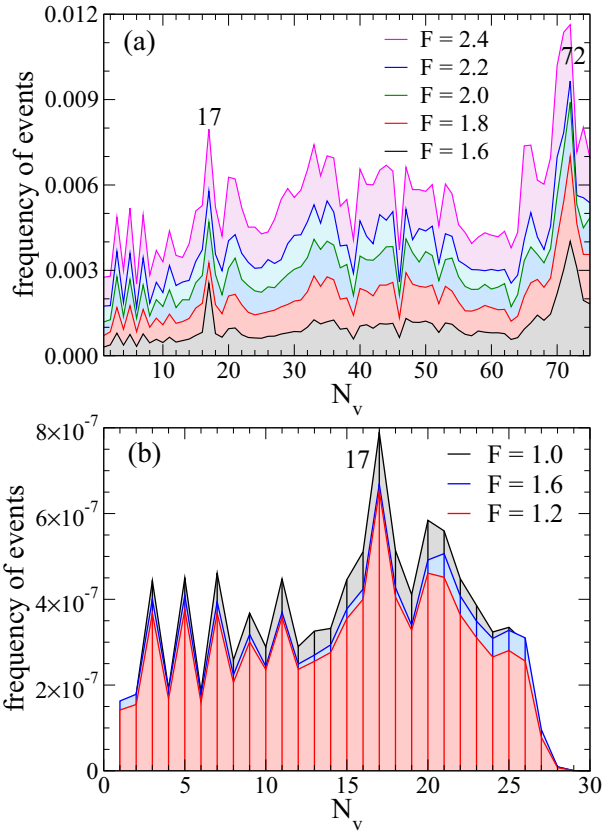

FIG. 7. αHL-like pore: (a) Histogram of the indicator N_v [Eq. (8)], collected over 3600 C-pulling runs starting from the native conformation. The largest peak, $N_v = 72$, is associated with the unfolding of the second structural cluster CluB, in particular with the passage across the stem of the strand S5 (65–72). The second peak, $N_v = 17$, corresponds to the arrival at the stem constriction of the hairpin S1-S2, forming a stable hairpin that determines a stalling point; see the conformation in Fig. 2(d). (b) N_v computed in C-pulling simulations starting from the hairpin (S1-S2) conformations, like the one reported in Fig. 2(d). The plot confirms that the Ubq spends time in the $N_v = 17$, i.e., the stalling state corresponding to the hairpin in the pore constriction.

the Ubq conformations with $N_{\rm cis} = 62$. We thus formulate a toy model to describe such a translocation pathway. We consider that the protein can be in only three different states: (i) C, the entire Ubq is untranslocated, and all the residues are on the Cis side; (ii) I, the Ubq is stuck in the intermediate conformation described in Sec. III and shown in Fig. 2; (iii) T, the Ubq has translocated so all its residues lie in the Trans side. We assume that each state is characterized by a constant number of residues in the Cis side, specifically, $N_{\rm cis}^C = 76$, $N_{\rm cis}^I = 62$, and $N_{\rm cis}^T = 0$. In this simplified scenario, a possible reaction pathway from the Cis to Trans state occurs as follows:

$$C \underset{k_{\rm off}}{\overset{k_c}{\rightleftarrows}} I \xrightarrow{k_T} T, \tag{12}$$

through only one relevant intermediate I. We neglected the improbable case that Ubq, once in the T state, comes back to I. Accordingly, the rate equations for the probabilities P_C, P_I,

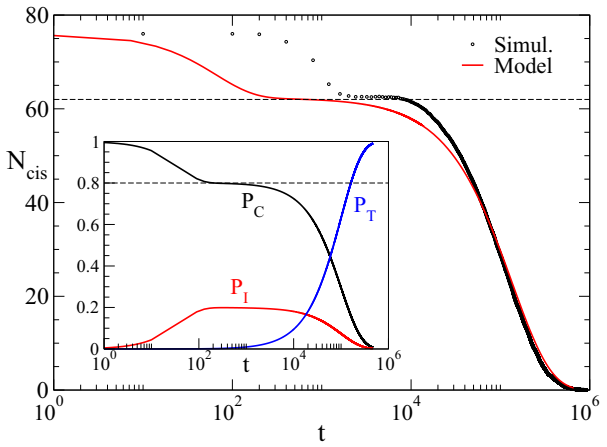

FIG. 8. Kinetics of translocation according to the model in Eq. (12). Main panel: time evolution of $\langle N_{\rm cis}\rangle$ from simulations (dots) compared with Eq. (16) (red line). The horizontal dashed line marks the common stalling behavior. Inset: time evolution of the population of the Cis state (black line). The red curve represents the population of the intermediate I and the blue one is the population of the Trans state. The rates are adjusted to the values $k_c = 5 \times 10^{-3}$, $k_{\rm off} = 2 \times 10^{-2}$, and $k_T = 5 \times 10^{-5}$.

and P_T to be, respectively, in the Cis, I, and Trans states are

$$\begin{aligned}\frac{dP_C}{dt} &= -k_c P_C + k_{\rm off} P_I,\\ \frac{dP_I}{dt} &= k_c P_C - (k_{\rm off} + k_T) P_I,\\ \frac{dP_T}{dt} &= k_T P_I,\end{aligned} \tag{13}$$

with the obvious initial conditions $P_C(0) = 1$, $P_I(0) = P_T(0) = 0$, since all the translocations of Ubq start from the Cis side.

The linearity of Eqs. (13) allows the evolution of $P_C(t)$, $P_I(t)$, and $P_T(t)$ to be found analytically,

$$\begin{aligned}P_C(t) &= \frac{(k_T+\lambda_1)\lambda_2 e^{\lambda_1 t} - (k_T+\lambda_2)\lambda_1 e^{\lambda_2 t}}{k_T(\lambda_2-\lambda_1)},\\ P_I(t) &= \frac{\lambda_1\lambda_2(e^{\lambda_2 t} - e^{\lambda_1 t})}{k_T(\lambda_2-\lambda_1)},\\ P_T(t) &= 1 - \frac{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}{\lambda_2-\lambda_1},\end{aligned} \tag{14}$$

where

$$\lambda_{1,2} = \frac{-(k_c + k_{\rm off} + k_T) \pm \sqrt{(k_c + k_{\rm off} + k_T)^2 - 4k_c k_T}}{2}$$

are the two nonvanishing eigenvalues of the matrix of coefficients in Eq. (13). In this toy model, the time evolution of the ensemble average of the observable $N_{\rm cis}$ reads

$$\langle N_{\rm cis}\rangle = P_C(t) N_{\rm cis}^C + P_I(t) N_{\rm cis}^I + P_T(t) N_{\rm cis}^T, \tag{15}$$

which, since $N_{\rm cis}^T = 0$, reduces to

$$\langle N_{\rm cis}\rangle = P_C(t) N_{\rm cis}^C + P_I(t) N_{\rm cis}^I. \tag{16}$$

The solutions of the above equations are plotted in the inset of Fig. 8, while the comparison of Eq. (16) with the result of the simulations for $N_{\rm cis}$ is shown in the main panel. This simple rate model reproduces the qualitative behavior of the $N_{\rm cis}$, suggesting that an on-pathway intermediate scenario described by Eq. (12) is plausible.

V. CONCLUSIONS

In this work, we analyzed the translocation pathways of ubiquitin (Ubq) pulled from the C terminal across a narrow pore shaped to mimic the vestibule-stem architecture of the α-Hemolysin channel. The purpose was to test the influence of Ubq native-state topology on the translocation dynamics, just to verify if Ubq undergoes a structure-dependent translocation. Ubq, for its simple and clear native-state topology, constitutes an excellent paradigm for this kind of study. Since we were primarily interested in the way the structural unfolding of the Ubq backbone couples to the transport in the nanopore, we employed the Gō model, a simplified coarse-grained protein model [39,41]. By construction, Gō models are built upon the knowledge of the protein native structures, and therefore the structure-dependent proteinlike properties turn to be faithfully described.

The C-pulling translocation runs revealed the presence of stalled dynamics, where the protein remains stuck in a relevant intermediate conformation. By looking at the topology graph in Fig. 1, this intermediate corresponds to the separation of strands S5 and S1 followed by a split in the structural clusters CluA and CluB packing against each other in the Ubq fold. In the literature, this split into two halves has already been conjectured in free pulling experiments of poly-ubiquitin by Schlierf *et al.* [30] and seen by Irback *et al.* [32] in simulations of single Ubq stretching.

The fact that the translocation and the stretching process in the free space produce a similar pathway is quite surprising because the two processes are not directly comparable due to the steric constraints imposed by both the pore and the membrane.

Finally, we found a cooperative translocation of beta-strand S1-S2 that, being very stable, often migrates as a single unit, violating the single file mode. This can be considered a second intermediate, although its signal in the translocation observables is less evident.

It is important to remark that Gō-model force fields do not explicitly take into account the directionality of the hydrogen bonding (HB), so one may wonder if the above results are spoiled by this lack. However, one also has to consider that a pulling process into a pore is expected to be less affected by the HB directionality because it occurs under confinement along a preferential axis. A coarse-grained variant of the Gō model that also takes this into account can be found in Ref. [55]. Our model without HB directionality has been able to reproduce the main features observed in the work of Ref. [32] in which the pulling simulations explicitly treated the HBs. We are confident that the main features of the Ubq translocation pathway seem to be captured by a simple Gō model.

In conclusion, we can claim that the Ubq native-state topology determines certain bottlenecks in the translocation; in particular, the presence of the two clusters CluA and CluB leaves a marked fingerprint in the translocation observables.

[1] W. Shi, A. K. Friedman, and L. A. Baker, Nanopore sensing, Anal. Chem. **89**, 157 (2017).

[2] S. Howorka and Z. Siwy, Nanopore analytics: Sensing of single molecules, Chem. Soc. Rev. **38**, 2360 (2009).

[3] T. Luchian, Y. Park, A. Asandei, I. Schiopu, L. Mereuta, and A. Apetrei, Nanoscale probing of informational polymers with nanopores. Applications to amyloidogenic fragments, peptides, and DNA–PNA hybrids, Acc. Chem. Res. **52**, 267 (2019).

[4] D. Coglitore, N. Giamblanco, A. Kizalaité, P. E. Coulon, B. Charlot, J.-M. Janot, and S. Balme, Unexpected hard protein behavior of BSA on gold nanoparticle caused by resveratrol, Langmuir **34**, 8866 (2018).

[5] M. Jain, H. E. Olsen, B. Paten, and M. Akeson, The oxford nanopore minion: Delivery of nanopore sequencing to the genomics community, Genome Biol. **17**, 239 (2016).

[6] R. D. Maitra, J. Kim, and W. B. Dunbar, Recent advances in nanopore sequencing, Electrophoresis **33**, 3418 (2012).

[7] B. Cressiot, A. Oukhaled, L. Bacri, and J. Pelta, Focus on protein unfolding through nanopores, BioNanoScience **4**, 111 (2014).

[8] M. Chinappi and F. Cecconi, Protein sequencing via nanopore based devices: A nanofluidics perspective, J. Phys.: Cond. Matter **30**, 204002 (2018).

[9] N. Varongchayakul, J. Song, A. Meller, and M. W. Grinstaff, Single-molecule protein sensing in a nanopore: A tutorial, Chem. Soc. Rev. **47**, 8512 (2018).

[10] E. Kennedy, Z. Dong, C. Tennant, and G. Timp, Reading the primary structure of a protein with 0.07 nm 3 resolution using a subnanometre-diameter pore, Nat. Nanotechnol. **11**, 968 (2016).

[11] W. Si, Y. Zhang, G. Wu, Y. Kan, Y. Zhang, J. Sha, and Y. Chen, Discrimination of protein amino acid or its protonated state at single-residue resolution by graphene nanopores, Small **15**, 1900036 (2019).

[12] G. Di Muccio, A. E. Rossini, D. Di Marino, G. Zollo, and M. Chinappi, Insights into protein sequencing with an α-Hemolysin nanopore by atomistic simulations, Sci. Rep. **9**, 6440 (2019).

[13] F. Piguet, H. Ouldali, M. Pastoriza-Gallego, P. Manivet, J. Pelta, and A. Oukhaled, Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore, Nat. Commun. **9**, 966 (2018).

[14] J. Wilson, L. Sloman, Z. He, and A. Aksimentiev, Graphene nanopores for protein sequencing, Adv. Funct. Mater. **26**, 4830 (2016).

[15] G. Oukhaled, J. Mathé, A.-L. Biance, L. Bacri, J.-M. Betton, D. Lairez, J. Pelta, and L. Auvray, Unfolding of Proteins and Long Transient Conformations Detected by Single Nanopore Recording, Phys. Rev. Lett. **98**, 158101 (2007).

[16] M. Boukhet, F. Piguet, H. Ouldali, M. Pastoriza-Gallego, J. Pelta, and A. Oukhaled, Probing driving forces in aerolysin and α-hemolysin biological nanopores: Electrophoresis versus electroosmosis, Nanoscale **8**, 18352 (2016).

[17] A. Asandei, I. Schiopu, M. Chinappi, C. H. Seo, Y. Park, and T. Luchian, Electro-osmotic trap against the electrophoretic force near a protein nanopore reveals peptide dynamics during capture and translocation, ACS Appl. Mater. Interfaces **8**, 13166 (2016).

[18] G. Huang, K. Willems, M. Soskine, C. Wloka, and G. Maglia, Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with frac nanopores, Nat. Commun. **8**, 935 (2017).

[19] E. L. Bonome, F. Cecconi, and M. Chinappi, Electro-osmotic flow through an α-hemolysin nanopore, Microfluid. Nanofluid. **21**, 96 (2017).

[20] D. Rodriguez-Larrea and H. Bayley, Multistep protein unfolding during nanopore translocation, Nat. Nanotechnol. **8**, 288 (2013).

[21] J. Nivala, D. B. Marks, and M. Akeson, Unfoldase-mediated protein translocation through an α-hemolysin nanopore, Nat. Biotechnol. **31**, 247 (2013).

[22] M. Bacci, M. Chinappi, C. M. Casciola, and F. Cecconi, Role of denaturation in maltose binding protein translocation dynamics, J. Phys. Chem. B **116**, 4255 (2012).

[23] D. D. Marino, E. L. Bonome, A. Tramontano, and M. Chinappi, All-atom molecular dynamics simulation of protein translocation through an α-hemolysin nanopore, J. Phys. Chem. Lett. **6**, 2963 (2015).

[24] E. L. Bonome, R. Lepore, D. Raimondo, F. Cecconi, A. Tramontano, and M. Chinappi, Multistep current signal in protein translocation through graphene nanopores, J. Phys. Chem. B **119**, 5815 (2015).

[25] Z. Xu, S. Zhang, J. K. Weber, B. Luan, R. Zhou, and J. Li, Sequential protein unfolding through a carbon nanotube pore, Nanoscale **8**, 12143 (2016).

[26] L. Huang, S. Kirmizialtin, and D. E. Makarov, Computer simulations of the translocation and unfolding of a protein pulled mechanically through a pore, J. Chem. Phys. **123**, 124903 (2005).

[27] P. Tian and I. Andricioaei, Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mitochondrial pore, J. Mol. Biol. **350**, 1017 (2005).

[28] E. L. Bonome, F. Cecconi, and M. Chinappi, Translocation intermediates of ubiquitin through an α-hemolysin nanopore: implications for detection of post-translational modifications, Nanoscale **11**, 9920 (2019).

[29] S. E. Jackson, Ubiquitin: A small protein folding paradigm, Org. Biomol. Chem. **4**, 1845 (2006).

[30] M. Schlierf, H. Li, and J. M. Fernandez, The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques, Proc. Natl. Acad. Sci. USA **101**, 7299 (2004).

[31] J. M. Fernandez and H. Li, Force-clamp spectroscopy monitors the folding trajectory of a single protein, Science **303**, 1674 (2004).

[32] A. Irbäck, S. Mitternacht, and S. Mohanty, Dissecting the mechanical unfolding of ubiquitin, Proc. Natl. Acad. Sci. USA **102**, 13427 (2005).

[33] J. Brujić, K. A. Walther, J. M. Fernandez *et al.*, Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin, Nat. Phys. **2**, 282 (2006).

[34] M. S. Li, M. Kouza, and C.-K. Hu, Refolding upon force quench and pathways of mechanical and thermal unfolding of ubiquitin, Biophys. J. **92**, 547 (2007).

[35] A. Kleiner and E. Shakhnovich, The mechanical unfolding of ubiquitin through all-atom monte carlo simulation with a Gō-type potential, Biophys. J. **92**, 2054 (2007).

[36] P. Szymczak and M. Cieplak, Stretching of proteins in a force-clamp, J. Phys.: Condens. Matter **18**, L21 (2005).
[37] A. Imparato and A. Pelizzola, Mechanical Unfolding and Refolding Pathways of Ubiquitin, Phys. Rev. Lett. **100**, 158104 (2008).
[38] A. Ammenti, F. Cecconi, U. M. B. Marconi, and A. Vulpiani, A statistical model for translocation of structured polypeptide chains through nanopores, J. Phys. Chem. B **113**, 10348 (2009).
[39] C. Clementi, H. Nymeyer, and J. N. Onuchic, Topological and energetic factors: What determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins, J. Mol. Biol. **298**, 937 (2000).
[40] F. Cecconi, C. Guardiani, and R. Livi, Testing simplified proteins models of the hPin1 WW domain, Biophys. J. **91**, 694 (2006).
[41] H. Taketomi, Y. Ueda, and N. Gō, Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions, Int. J. Pept. Protein Res. **7**, 445 (1975).
[42] D. Alexeev, S. M. Bury, M. A. Turner, O. M. Ogunjobi, T. W. Muir, R. Ramage, and L. Sawyer, Synthetic, structural and biological studies of the ubiquitin system: Chemically synthesized and native ubiquitin fold into identical three-dimensional structures, Biochem. J. **299**, 159 (1994).
[43] R. B. Best, G. Hummer, and W. A. Eaton, Native contacts determine protein folding mechanisms in atomistic simulations, Proc. Natl. Acad. Sci. USA **110**, 17874 (2013).
[44] W. Humphrey, A. Dalke, and K. Schulten, VMD–Visual molecular dynamics, J. Mol. Graph. **14**, 33 (1996).
[45] C. Forrey and M. Muthukumar, Langevin dynamics simulations of ds-DNA translocation through synthetic nanopores, J. Chem. Phys. **127**, 015102 (2007).
[46] P. Szymczak, Periodic forces trigger knot untying during translocation of knotted proteins, Sci. Rep. **6**, 21702 (2016).
[47] M. Muthukumar and C. Y. Kong, Simulation of polymer translocation through protein channels, Proc. Natl. Acad. Sci. USA **103**, 5273 (2006).
[48] A. Nadtochiy, D. Melnikov, and M. Gracheva, Filtering of nanoparticles with tunable semiconductor membranes, ACS Nano **7**, 7053 (2013).
[49] G. I. Makhatadze, M. M. Lopez, J. M. Richardson, III, and S. T. Thmos, Anion binding to the ubiquitin molecule, Prot. Sci. **7**, 689 (1998).
[50] K. Burrage, I. Lenane, and G. Lythe, Numerical methods for second-order stochastic differential equations, SIAM J. Sci. Comput. **29**, 245 (2007).
[51] G. I. Bell, Models for the specific adhesion of cells to cells, Science **200**, 618 (1978).
[52] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: fifty years after Kramers, Rev. Mod. Phys. **62**, 251 (1990).
[53] M. Muthukumar, *Polymer Translocation* (CRC Press, Boca Raton, FL, 2016).
[54] J. M. Polson and A. C. M. McCaffrey, Polymer translocation dynamics in the quasi-static limit, J. Chem. Phys. **138**, 174902 (2013).
[55] E.-H. Yap, N. L. Fawzi, and T. Head-Gordon, A coarse-grained α-carbon protein model with anisotropic hydrogen-bonding, Proteins: Struct. Func. Bioinf. **70**, 626 (2008).