

## Article

## Protein Unfolding by Biological Unfoldases: Insights from Modeling

Michał Wojciechowski,<sup>1</sup> Piotr Szymczak,<sup>2</sup> Mariano Carrión-Vázquez,<sup>3</sup> and Marek Cieplak<sup>1,\*</sup><sup>1</sup>Institute of Physics, Polish Academy of Sciences, Warsaw, Poland; <sup>2</sup>Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland; and <sup>3</sup>Instituto Cajal, Consejo Superior de Investigaciones Científicas and Instituto Madrileño de Estudios Avanzados en Nanociencia, Madrid, Spain

**ABSTRACT** The molecular determinants of the high efficiency of biological machines like unfoldases (e.g., the proteasome) are not well understood. We propose a model to study protein translocation into the chamber of biological unfoldases represented as a funnel. It is argued that translocation is a much faster way of unfolding a protein than end-to-end stretching, especially in a low-force regime, because it allows for a conformational freedom while concentrating local tension on consecutive regions of a protein chain and preventing refolding. This results in a serial unfolding of the protein structures dominated by unzipping. Thus, pulling against the unfoldase pore is an efficient catalyst of the unfolding reaction. We also show that the presence of the funnel makes the tension along the backbone of the substrate protein nonuniform even when the protein gets unfolded. Hence, the stalling force measured by single-molecule force spectroscopy techniques may be smaller than the traction force of the unfoldase motor.

## INTRODUCTION

Intracellular proteins are incessantly synthesized and degraded (1). Such a regulated turnover of proteins is one of the basic mechanisms taking place in a cell. It controls enzyme activity, temporarily limits the action of transcription factors, and removes damaged proteins, preventing their accumulation as a part of the quality control system of the cell. Selective degradation usually occurs in multi-protein complexes known as AAA+ proteases (ATPases associated with various cellular activities). In eukaryotes and archaea, these nanomachines are known as proteasomes and the most common structure found is denoted as 26S (2,3). In bacteria, there are several types of the degradation complexes, for instance ClpXP and Lon (4,5).

The 26S proteasome consists of two subunits: 19S and 20S. The first of these is an ATP-dependent cap that captures an ubiquitinated protein, unfolds it (6), and then directs it to the 20S proteasome. 20S is the actual shredding center where cleavage of peptide bonds takes place. The functions of unfolding and degradation are spatially separated in the ClpXP complex (a model system for these proteases): the ATP-dependent hexameric unfoldase ClpX recognizes, unfolds, and translocates the tagged proteins to the ClpP peptidase chamber for proteolysis (7,8). Similar to other AAA+ ATPases, both 26S and ClpXP have a barrel-like form with a narrow entrance at both tops of the barrel.

The mechanics of the degradation process raises many questions. Some of them have been addressed experimentally through biochemistry and single molecule studies.

For instance, Maillard et al. (7) have considered degradation of the green fluorescent protein (GFP) by ClpXP. The ClpXP was anchored to one polystyrene bead and the GFP to another, and the interactions between ClpXP and GFP were monitored. Translocation velocity was found to depend on the force,  $F_r$ , due to the bead. In the limit of  $F_r = 0$ , the velocity was ~80 amino acids (AA) per second while stalling took place at ~20 pN. Thus, the force with which ClpXP pulls on the protein appears to be equal to 20 pN.

Aubin-Tam et al. (8) have considered a similar arrangement but a different substrate: eight distinct domains of the immunoglobulin-like filamin-A, which was flanked by an *ssrA* tag on the ClpX side and Halo Tag protein on the other side. The forces involved and translocation speed have been found to be similar to those in Maillard et al. (7). Despite all these advances, two main questions remain related to the mechanism of unfolding and translocation by AAA+ proteases: how different is this process from the N-C geometry of pulling used in single-molecule force spectroscopy, and how does a (passive or active) probe affect the natural (untethered) unfolding process of these nanomachines?

Typical AAA+ proteases have a ring-shaped hexamer, which uses the energy of multiple cycles of ATP binding and hydrolysis to produce a conformational change. It generates a power stroke that unfolds substrate proteins by translocation through a central axial pore. ATP binding and hydrolysis, and phosphate release in the region located between the large and small AAA+ domains, can alter their rotation by causing rigid body motions that propagate around the ring and are then transmitted to the substrate (at least partly) through conserved loops that protrude into

---

Submitted March 25, 2014, and accepted for publication July 15, 2014.

\*Correspondence: [cieplak@ifpan.edu.pl](mailto:cieplak@ifpan.edu.pl)

Editor: Michael Feig.

© 2014 by the Biophysical Society  
0006-3495/14/10/1661/8 \$2.00



the central pore. Thus, the ring exerts two kinds of forces: longitudinal and torsional (9).

Here, we consider only the longitudinal forces, and develop an effective model of the proteasome in which the degradation process is mimicked by pulling a protein into a geometrically sculpted potential. For the sake of simplicity, we represent the pore as rigid and purposely disregard the role of allostery and heterogeneity of interactions (10–13). In this way we can dissect the contribution of threading through a narrow pore from other factors (rate-limiting effects of substrate topology, allosteric asymmetry, or specific interactions within the pore). We show that the unfolding and translocation dynamics of the protein in single-molecule force spectroscopy experiments in the presence of proteasome is a result of an interplay of three forces: the motor force, the pore reaction force, and the force due to the external measuring device.

## METHODS

### The model of the protein-proteasome system

The protein-proteasome system is molecularly large, so modeling it by molecular-dynamics methods is helped by introducing several approximations. We describe the protein in the coarse-grained Gō-like fashion with an implicit solvent as done in Cieplak and Hoang (14), Sulkowska and Cieplak (15,16), and Sikora et al. (17). We model the proteasome as an effective external potential in an analogy to studies of protein translocation through cylindrical pores (18–24), but now with a smooth and generic funnel-like pore. The atomic-level time-dependent roughness may or may not affect the ease of degradation. We mimic the motor action of the proteasome through a force that pulls the protein into the structure. As a simplification, we consider this force to be constant even though it is expected to be intermittent or periodic due to the cyclic nature of the ATP hydrolysis. A graphical representation of our model is shown in Fig. 1 A.

A more computer-demanding model would involve a coarse-grained molecular description of the proteasome as used in Kravats et al. (25). The motor action is then introduced by imposing an asymmetric coupling of the substrate to the loops in the central channel. This coupling yields discrete translocation steps. The translocation process is shown to be the rate-limiting element in degradation of an  $\alpha$ -helical protein with weak stability in the C-terminus region (25).

### Modeling the substrate proteins

The proteins studied are listed in Table 1 together with characteristic forces that will be defined later. The choice of the proteins is motivated by several factors. Even though proteasomes can degrade disulfide bonded polypeptides and up to three polypeptide chains (26,27), for simplicity, we start with proteins without such bonds. We select one protein with large mechanostability: cohesin (the structure code PDB:1AOH) (28). We also consider proteins that have been studied experimentally in the context of degradation either by the proteasome or by the Clp. These are: GFP (PDB:1GFL) (29,30), ribonuclease barnase (PDB:1BNR), mouse dihydrofolate reductase (PDB:1U71) (31), ribonuclease H (PDB:1RIL) (30,32), and I27 domain of titin (PDB:1TIT) (30).

In each model protein, the AAs are represented by the C $^{\alpha}$  atoms that are tethered into a chain. Potential wells are assigned to the native contacts by attributing enlarged van der Waals spheres (33) to all heavy atoms in a native conformation of the protein and then checking whether a pair of

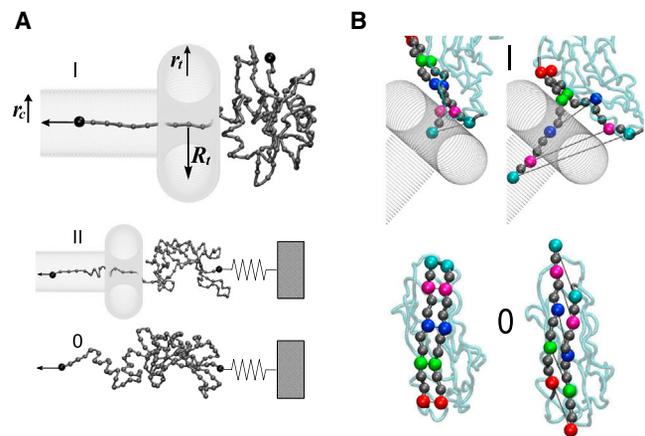


FIGURE 1 (A) Pulling protocols considered: I and II involve the proteasome whereas 0 does not. (Horizontal arrows) Direction of the proteasomal traction. In the constant  $v_p$  case, stretching is implemented by attaching the pulled end to a spring (not shown) whose other end is pulled with speed  $v_p$ . (B) Protocol I explains the geometry of the modeled protein-proteasome system.  $R_r$  and  $r_i$  are the major and minor radii, respectively; other radii are explained in the text. The equation for the torus is  $(x^2 + y^2 + z^2 + R_r^2 - r_i^2)^2 = 4R_r^2(x^2 + y^2)$ . Two consecutive snapshots of the model cohesin during the initial stages of unraveling. Strands  $\beta_1$  and  $\beta_9$  of protein PDB: 1AOH are highlighted by showing the corresponding AAs as spheres. The lower snapshots correspond to protocol 0. The highlighted strand on the left (right) is  $\beta_1$  ( $\beta_9$ ). The colors used (in the on-line version) merely highlight examples of AA pairs that are immediate partners in the  $\beta$ -sheet. The upper snapshots correspond to protocol I-C. Strand  $\beta_9$  is seen to enter the inside of the funnel. To see this figure in color, go online.

amino acids  $i$  and  $j$  ( $j \geq i + 2$ ) has at least one case of the atomic overlap between them. If it does, the pair is said to form a “native contact” (15,16). The potential associated with a contact between C $^{\alpha}$  atoms separated by distance  $r_{ij}$  is given by the potential

$$V_{ij}(r) = 4\epsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right].$$

Its depth,  $\epsilon$ , has been calibrated (17) to be approximately equal to 110 pN  $\times$   $\text{\AA}$ , which correlates well with the experimental data. This calibration was obtained by considering a range of pulling speeds,  $v_p$ , and by extrapolating the magnitudes of the maximal force peaks to those measured at the speeds used in the experiments. Notice that stretching starts in a native state near which Gō-like models work the best. A calibration based on, say, folding properties can yield different estimates. Considering other variants of the Gō-like models with other definitions of the contact map and descriptions of the backbone stiffness (here, it is based on the local chirality (16)) can also affect the calibration. The length parameters  $\sigma_{ij}$  are determined so that the potential minimum agrees with the native distances between  $i$  and  $j$ . Nonnative contacts correspond to softly repulsive spheres of radius 2  $\text{\AA}$ . Bonded interactions are modeled by the harmonic potential with the spring constant of 50  $e/\text{\AA}$ . We have neglected the presence of the covalent bond between Ser-65 and Gly-67 in the chromophoric segment of PDB:1GFL. We have checked that its effect is minor and within the thermal noise when stretching at constant  $v_p$ . The solvent is represented by damping and random fluctuational forces whose amplitude depends on the temperature,  $T$ . Our simulations are performed at  $k_B T = 0.3\epsilon$ , which is close to the optimal folding  $T$  of the model proteins (14) ( $k_B$  is the Boltzmann constant). With our calibration of  $\epsilon$  (equivalent to  $\sim 1.6$  kcal/mol), this choice is also consistent with studies performed in a vicinity of the room temperature.

**TABLE 1** Constant speed and constant force simulations

Code	Protein	II-C	II-N	I-C	I-N	0
Constant speed <sup>a</sup>						
PDB:1AOH	Cohesin	3.7	4.5	<b>2.3</b>	2.5	4.4
PDB:1GFL	Green fluorescent protein	<b>2.3</b>	2.3	2.5	2.5	2.7
PDB:1VCD	Nudix	3.5	4.0	<b>2.0</b>	3.9	4.1
PDB:1O72	Cytolysin	3.1	2.6	2.5	<b>1.9</b>	4.1
PDB:1GWY	Cytolysin	3.1	2.8	2.7	<b>1.9</b>	3.9
PDB:1ODI	Purine nucleoside phosphorylase	3.4	3.2	<b>3.1</b>	4.1	3.8
PDB:1Y2X	Lectin	2.6	2.9	<b>1.7</b>	2.2	3.6
PDB:2PF6	Lutheran glycoprotein	3.7	2.7	2.0	<b>1.8</b>	3.6
PDB:1OTX	Purine nucleosidase phosphorylase	3.1	2.7	<b>2.6</b>	3.8	3.6
PDB:2DSD	ADP-ribose pyrophosphatase	2.6	3.2	<b>2.1</b>	2.9	3.6
PDB:1NW4	Purine nucleoside phosphorylase	3.3	<b>2.9</b>	3.3	3.4	3.5
PDB:1U71	Dihydrofolate reductase	2.3	2.5	<b>2.0</b>	3.0	2.6
PDB:1TIT	I27 domain of titin	2.6	2.4	1.7	<b>1.3</b>	2.1
PDB:1RIL	Ribonuclease H	2.1	3.3	<b>1.6</b>	2.9	1.8
PDB:1BNR	Barnase	1.6	1.9	1.8	<b>1.3</b>	<b>1.2</b>
Constant force <sup>b</sup>						
PDB:1AOH	Cohesin	2.4	2.9	<b>1.2</b>	1.5	2.7
PDB:1GFL	Green fluorescent protein	<b>1.1</b>	<b>1.1</b>	<b>1.1</b>	1.3	2.7
PDB:1BNR	Barnase	0.8	0.6	0.7	0.7	<b>0.4</b>
PDB:1TIT	I27 domain of titin	1.6	1.3	0.9	<b>0.5</b>	1.3
PDB:2DSD	ADP-ribose pyrophosphatase	1.3	1.7	<b>1.1</b>	1.7	2.1

<sup>a</sup>The values of  $F_{\max}$ , in units of  $e/\text{\AA}$ , in the constant speed simulations ( $v_p = 0.005 \text{ \AA}/\tau$ ) for the protocols indicated. The data are based on at least three trajectories in each case. The data point to the existence of single dominant pathways. The lowest forces of resistance to pulling, among the protocols, are highlighted. The error bars related to the thermal noise are  $\sim 0.1 e/\text{\AA}$  (a detailed discussion of this issue is in Chwastyk et al. (37)) so the selection of the lowest force is not always clear-cut. For instance, protocols I-N and 0 yield practically the same  $F_{\max}$  so both are highlighted in bold.

<sup>b</sup>The values of  $F_p$ , in units of  $e/\text{\AA}$ , obtained by extrapolation of constant force simulations to the timescales corresponding to 80 residues/s. The error bars are  $\sim 0.15 e/\text{\AA}$ .

## Modeling the proteasome

The model consists of two connected structures, these being a torus placed atop a cylinder, which creates a funnel-like structure. The torus mimics the entrance to an unfoldase like the proteasome and the cylinder—the part where protein degradation takes place in this structure. The funnel defines a region where the protein is pulled and prevents the elongated protein from refolding. The motivation for the model comes from the fact that the biological proteasome is composed of the core particle 20S and two regulatory caps 19S. Both substructures share an axial channel. The core itself consists of four coaxially placed heptameric rings. The protease active sites are located on the interior surface of the two rings in the middle. Thus the target protein must enter the axial channel before it is degraded. The diameter of the channel has been determined to be  $\approx 15 \text{ \AA}$  (34,35), as defined by the average distance between opposing heavy atoms on the inner side of the proteasome.

To account for some flexibility in the channel, we model it as a pore of radius  $r_c = 8 \text{ \AA}$  (see Fig. 1 A). The length of the pore can be larger than the actual axial size of the whole proteasome because we do not model the very process of degradation. The width of the pore is wide enough to accommodate a hydrated polypeptide chain but not the secondary structures. The outer two rings of 20S form a gate through which proteins enter the core chamber. This is the region that recognizes and binds polyubiquitin tags of the proteins. 19S generates a force that pulls the substrate. We represent the cap by a torus with the major radius  $R_t = 13 \text{ \AA}$  and the minor radius  $r_t = 6 \text{ \AA}$  so that the opening in its narrowest place has a radius of  $7 \text{ \AA}$ , which is smaller than the radius of the cylindrical pore describing the core particle. This disparity accounts for an extra cavitylike space that forms between the

cap and the core particle. Other than the experimentally established condition on the size of the hole in the torus, its geometrical parameters have been chosen arbitrarily. We have been guided by the condition that the torus should be large enough that couplings of the protein with the outside surface do not affect the pulling process. These parameters define the funnel-like chamber into which the substrate is squeezed. The adjustment of the protein to the geometry of the chamber facilitates unfolding.

The interaction between the protein and the surface of the torus as well as the inner surface of the cylindrical pore is assumed to be repulsive, and given by the truncated Lennard-Jones potential

$$V_i^s = \begin{cases} 4\epsilon \left[ \left( \frac{\sigma}{d_i} \right)^{12} - \left( \frac{\sigma}{d_i} \right)^6 \right] & , d_i \leq r_{\min} \\ 0 & , d_i > r_{\min} \end{cases} \quad (1)$$

where  $d_i$  is the closest distance between the  $i$ th AA and the torus/pore surface. The distance  $r_{\min} = 6 \text{ \AA}$  is the coordinate of the minimum of the potential, which takes into account excluded volumes of residues and wall atoms. We take  $\sigma = 0.5^{1/6} r_{\min} = 5.345 \text{ \AA}$ .

## The motor action

The motor action of the proteasome is represented by a force,  $F_p$ , acting on the most forward AA; it can be either C- or N-terminal. The value of  $F_p$  is expected to be small so that the resulting typical unfolding time,  $t_U$ , would be beyond computational possibilities. Instead, we establish the dependence of  $t_U$  on  $F_p$  in the feasible range, and extrapolate it to the experimental time-scales as determined from the degradation speed.

It is instructive to first discuss constant  $v_p$  stretching, because this elucidates apparently novel features in unfolding scenarios brought in by the proteasome and illustrates distinct possible modes of operation of the system. We take  $v_p = 0.005 \text{ \AA}/\tau$ , where  $\tau \sim 1 \text{ ns}$  (17). We consider three pulling protocols (Fig. 1 A). They are defined both for constant  $v_p$  and constant  $F_p$  situations. The basic one is denoted by “I”: the protein is pulled into the funnel by one of the termini, whereas the other is unconstrained. This is how the proteasome is expected to work. The tag for degradation can be placed at the C-terminus (protocol I-C) or at the N-terminus (protocol I-N). The role of the directionality of proteasome pulling has been assessed by Berko et al. (36). However, this is not how the optical tweezers experiments have been performed and hence we consider protocol “II” (and its subcases II-C and II-N). In this protocol, one end is pulled, but another is tethered to a stationary spring, which is meant to represent a measuring device (such as a polystyrene sphere of an optical tweezer apparatus). We take the spring constant of  $0.06 e/\text{\AA}^2$  as in typical atomic-force microscopy (AFM) cantilevers. This arrangement has been designed with the purpose of measuring the pulling force exerted by the proteasome. Finally, the protocol denoted by 0 is one that does not involve a proteasome—it is the standard AFM-based pulling.

The results may depend on the initial orientation of the protein relative to the model proteasome. We have adopted a procedure in which both termini are located on the main axis of the proteasome and away from the torus. The center of mass of the protein is then moved toward the torus in small steps until one of the heavy atoms collides with it. The placement just before this last step is taken as the starting state for most of the simulations and from then on, the protein is represented only by its  $C^\alpha$  atoms. We also consider starting states in which the center of mass is shifted away from the near-touch situation.

## RESULTS

### Stretching at constant speed

At constant  $v_p$ , one monitors the tension,  $F$ , in the backbone and observes patterns with force peaks. A characteristic

force,  $F_{\max}$ , associated with a protein is defined as a height of the maximal peak. It depends on  $v_p$ . The upper part of Table 1 for 15 proteins, including cohesin and GFP, lists values of  $F_{\max}$  obtained in the three protocols for one  $v_p$ . There are several conclusions we draw from it:

1. The force peaks in proteasome-mediated pulling are usually lower than those in protocol 0.
2. Forces are usually lower in protocol I than in II because the free terminus allows the substrate to rotate and adopt a less obstructive orientation. However, holding the other end positions the substrate centrally, which may generate an advantage in some situations.
3. Except for GFL, pulling by different termini results in a different  $F_{\max}$  value.

The explanation for the first of these observations is that interactions with the pore walls either generate an extra force that disrupts the protein or facilitate adoption of an easier pathway. Fig. 1 B illustrates the difference between protocols 0 (top) and I-C (bottom) for PDB:1AOH. Without the proteasome, the first and largest force peak (top-left panel of Fig. 2) is due to shear between strands  $\beta_1$  and  $\beta_9$ . Strand  $\beta_1$  comprises two substrands  $\beta_{1a}$  (residues 6–11) and  $\beta_{1b}$  (12–15). Strand  $\beta_9$  has two substrands  $\beta_{9a}$  (136–140) and  $\beta_{9b}$  (142–147). With the proteasome, however, shearing is replaced by unzipping of these strands (Fig. 2, middle left panel) because strand  $\beta_9$  is pulled into the funnel-pore whereas  $\beta_1$  stays anchored at the opening and  $F_{\max}$  is nearly halved.

We now consider the  $F-d$  traces for PDB:1AOH, where  $d$  denotes displacement of the pulling spring. Results for other proteins are shown in the Supporting Material. Stretching of PDB:1AOH in protocol 0 has been shown to have  $F_{\max}$  close to 480 pN ( $4.3 \text{ } \epsilon/\text{\AA}$ ) (28). This protein consists of 147 AAs that form nine  $\beta$ -strands and one short  $\alpha$ -helix linked (Fig. 2). The remaining panels show the plots of  $F$  versus  $d$  for the specified protocols. Other symbols indicate the crucial contacts that rupture at the corresponding force peak (for instance,  $\beta_1-\beta_9$  means contacts between strands  $\beta_1$  and  $\beta_9$ ). The importance degree of a group of contacts is determined by removing the contacts and monitoring the change in  $F_{\max}$  (15). In the absence of the proteasome, the main force peak is associated with shear between  $\beta_1$  and  $\beta_9$  and the last force peak is between  $\beta_4$  and  $\beta_7$ . The second peak involves rupture of several structures.

Proteasome-mediated pulling is strongly asymmetric, i.e., it depends on the terminus being pulled as well as the protocol. For I-C, II-C, and II-N, the first force peak still involves  $\beta_1$  and  $\beta_9$ . After the first peak, the  $F-d$  curves are all very different from each other and from the nonproteasome case. In the case of I-C, the height of the first peak is reduced, as discussed before. The next, strong force peak at  $\sim 200 \text{ \AA}$  is still due to unzipping, but in this case, the  $\beta_7$  strand is drawn out of two nearby strands,  $\beta_{2,4}$ . Thus each residue from  $\beta_7$  has to break two types of contacts

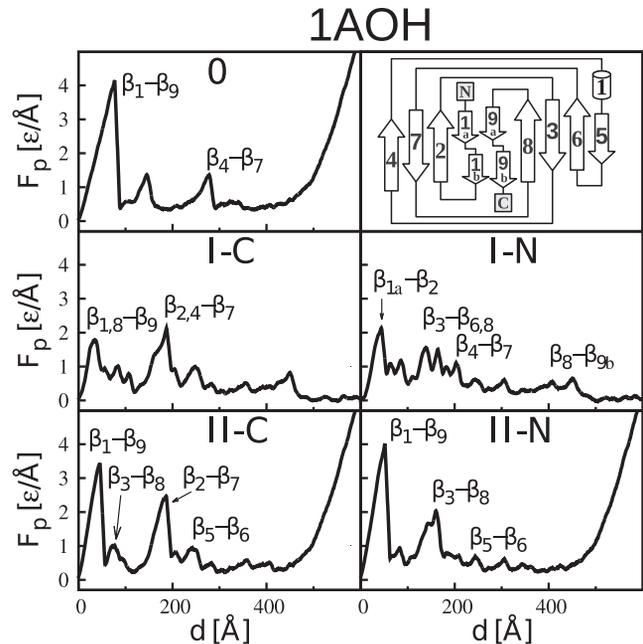


FIGURE 2 (Top right panel) Connections between the secondary structures of PDB:1AOH. (Cylinder)  $\alpha$ -helix ( $\alpha_1$  comprises AAs 64–68). (Arrows)  $\beta$ -strands, as follows:  $\beta_1$ , which consists of two substrands  $\beta_{1a}$  (6–11) and  $\beta_{1b}$  (12–15),  $\beta_2$  (19–28),  $\beta_3$  (36–44),  $\beta_4$  (48–57),  $\beta_5$  (69–74),  $\beta_6$  (78–86),  $\beta_7$  (99–109), and  $\beta_8$  (115–128); and  $\beta_9$ , which consists of two substrands  $\beta_{9a}$  (136–140) and  $\beta_{9b}$  (142–147). The remaining panels show the  $F-d$  curves obtained at constant speed for various indicated pulling protocols. The unit of force,  $\epsilon/\text{\AA}$ , corresponds to  $\sim 110 \text{ pN}$ . The symbols near the force peaks indicate contacts between the secondary structures that contribute to the generation of the peak in a dominant way. The procedure that identifies the origin of the force peaks involves checking the effect of removal of contacts in specific groups. In the case of the second peak, the candidates are  $\beta_3-\beta_8$ ,  $\beta_2-\beta_7$ , the loop between  $\beta_6$  and  $\beta_7$ , and the loop between  $\beta_8$  and  $\beta_9$ . However, removal of these contacts does not lower the second force peak in any noticeable manner but affects the unfolding pathway already near the first peak. It is, therefore, difficult to identify the origin of the second peak in this way.

(one with  $\beta_2$  and the other with  $\beta_4$ ). The presence of the proteasome introduces asymmetry to what would otherwise be a shearing process: a pulling force on one strand in a  $\beta$ -sheet is different in magnitude than the opposite force acting on its partner strand. As a result, the contacts between a set of strands do not rupture simultaneously. The resulting torque rotates the protein away from the shearing direction and leads to unzipping.

The nature of the unfolding events and the values of  $F_{\max}$  depend on the protocol. In particular, the mode of measurement used in protocol II need not probe the action of the proteasome as obtained in protocol I, because protocol II restricts rotations. We have also considered II-N and II-C for PDB:1RG7, with and without its ligand methotrexate, as studied experimentally by Johnston et al. (38). We model the ligand as a fairly stiff object comprising three harmonically coupled beads that also have contact interactions with the protein. We enhance the ligand-protein interactions by a factor of 10, inasmuch as they are much stronger than in

hydrogen bonds (39). We find that in II-N pulling at constant  $v_p$ , the height of the first force peak doubles from  $2.4 \text{ e}/\text{\AA}$  on coupling to the ligand. This agrees with the finding that methotrexate inhibits proteolysis (38) from the N-terminus. In the II-C case, we find that the peaks that appear later grow significantly (see Fig. S5 in the Supporting Material).

### Stretching at constant force

The processes at constant  $F_p$  are described by plots of the end-to-end distance,  $L$ , as a function of time,  $t$ , for individual trajectories and by determining the median characteristic unfolding time,  $t_U$ . In protocols 0 and II,  $t_U$  is defined as  $t$  needed for  $L$  to reach 85% of the full backbone length. At this stage, there are almost no contacts left. In protocol I, it is the translocation time. It should correspond to the first entry of the last AA to the central pore. However, fluctuations may cause temporary backtracking, so the nominal point of entry is shifted  $25 \text{ \AA}$  inward.

Fig. 3 shows examples of the  $t$ -dependence of  $L$  for PDB:1AOH. The top-right panel shows  $t_U$  as a function of  $F_p$  for the five protocols. A transition between the low- and high-force regimes taking place in proteins (40,41) is carried over to situations with the proteasomes, which mirrors the results for a peptide in a simpler pore (42,43). The functions, however, are shifted. In the low-force regime,  $\tau_u(F)$  depends on  $F_p$  exponentially as the unraveling processes are dominated by waiting periods needed to cross

an energy barrier. In the high-force regime, unfolding starts almost immediately. The crossover force,  $F'$ , is interpreted as the threshold force at which the kinetic barrier between the folded and unraveled states vanishes. We observe that for PDB:1AOH, the values of  $F'$  decrease between the protocols in the order: 0, II-N, II-C, I-N, and I-C, indicating that the presence of the proteasome facilitates unraveling, especially in the I-C case.

In Fig. 3, the forces have been selected based on two criteria:

1.  $F_p$  is within the regime of the exponential growth in  $t_U$ , and
2. The timescales are comparable between the panels and are of  $\sim 10^6 \tau$ .

The plots come with several plateaus that correspond to the waiting periods in intermediate states. The symbols next to the  $L$  curves indicate contacts rupturing at the specific increase in  $L$ . The order of the rupturing events is relatively conserved for each protocol used (see Fig. S1). Similar conclusions can be drawn for PDB:1GFL (238 AAs) (Fig. 4)—the proteasome is seen to facilitate unraveling, but  $F'$  decreases with the protocols in a different order: 0, I-N, II-N, II-C, and I-C.

The order of unraveling events for protocol I-C is consistent with the fluorescence spectroscopy studies of unfolding

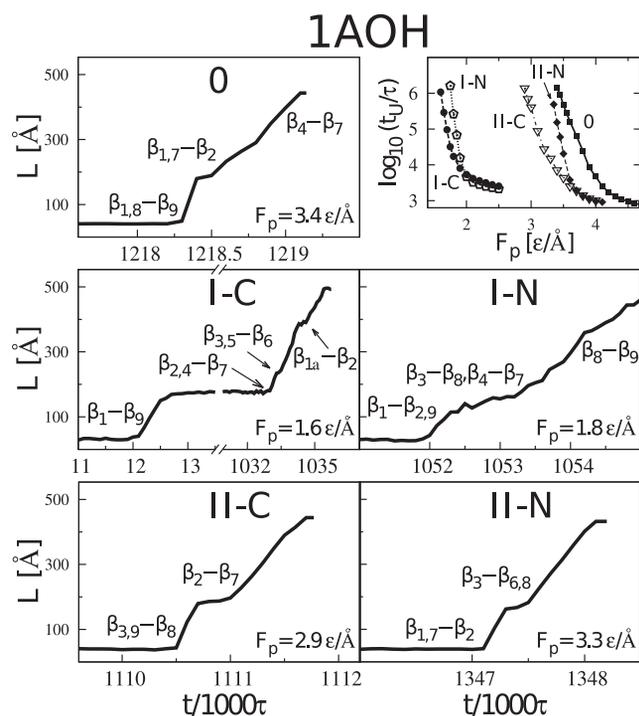


FIGURE 3 Constant force unraveling of PDB:1AOH for the protocols indicated. (Top right) Median unfolding time as a function of  $F_p$ ; remaining panels show examples of the time dependence of  $L$  for the forces shown.

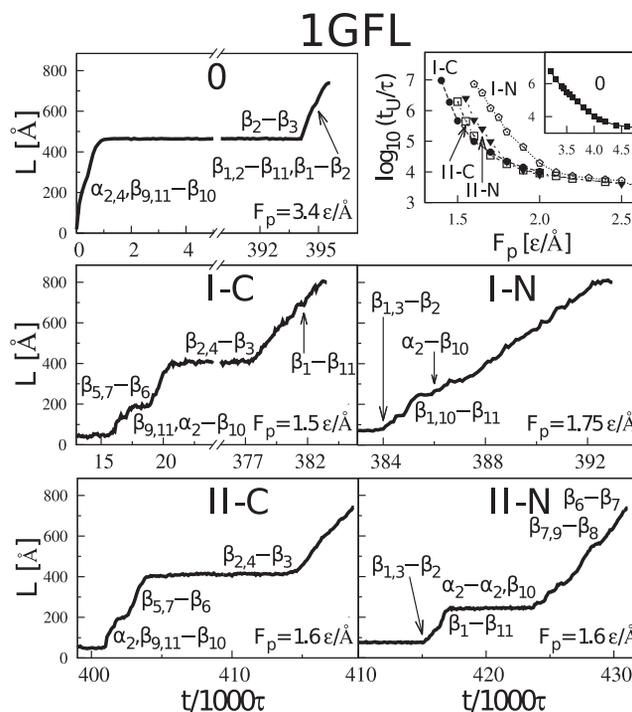


FIGURE 4 Constant force unraveling for PDB:1GFL for the protocols indicated. The assignment of secondary structures is as follows:  $\alpha_1$  (4–9),  $\alpha_2$  (56–65),  $\alpha_3$  (68–72),  $\alpha_4$  (75–82),  $\alpha_5$  (83–87),  $\alpha_6$  (156–159),  $\beta_1$  (12–23),  $\beta_2$  (25–36),  $\beta_3$  (40–49),  $\beta_4$  (91–101),  $\beta_5$  (104–115),  $\beta_6$  (118–128),  $\beta_7$  (148–155),  $\beta_8$  (160–171),  $\beta_9$  (175–188),  $\beta_{10}$  (199–208), and  $\beta_{11}$  (216–227). Symbol  $L_{\beta_3}$  denotes the loop that starts after strand  $\beta_3$ .

of a C-tagged PDB:1GFL by the AAA+ ClpXP protease (44). These studies have demonstrated that an initial step in ClpXP unfolding involves extraction of the C-terminal  $\beta$ -strand (denoted as  $\beta_{11}$  in Nager et al. (44), and as  $\beta_{10}$  in this article). In our model, we observe that the contacts that get broken in the first stage (at  $F_p = 1.5 \text{ } \epsilon/\text{\AA}$ ) are those which involve this very strand. The resulting intermediate has been shown (44) to be populated at low ATP concentrations, stalling further unfolding. Robust further degradation takes place only at high ATP concentration. It should also be noted that protocols I-C and II-C yield the same unfolding scenarios for the model PDB:1GFL. In the first intermediate seen in the left-bottom panel of Fig. 4 (II-C), the first 130 N-terminal residues are still in their near-native state, which is in agreement with Maillard et al. (7). The initial stages of unfolding under protocol 0 are similar, but they also involve severing contacts of  $\beta_{10}$  with  $\alpha_4$ .

The analysis of  $L$  as a function of  $t$  reveals the existence of a number of intermediate states corresponding to multiple kinetic barriers. The associated waiting times,

$$\tau_\kappa \sim e^{(\delta E_\kappa - F \delta x_\kappa) / k_B T},$$

are widely distributed. The index  $\kappa$  labels various energy minima involved. Due to the exponential form of  $\tau_\kappa(\delta E_\kappa)$  in the small force regime, the highest barrier dominates and determines the slope of  $\log \tau_u(F)$  dependence. This allows us to estimate the unfolding forces at experimentally reported timescales of  $\sim 1$  s needed to translocate 80 residues. Note, however, that these estimates should be treated with caution due to the simplified energy landscape adopted in our model and the fact that the position of the transition state,  $\Delta x_\kappa$ , can itself be force-dependent (45,46). The estimates are given in the lower part of Table 1 for five proteins. The lowest force is for PDB:1BNR in protocol II-N:  $0.6 \text{ } \epsilon/\text{\AA}$ , i.e., nearly 70 pN (with our calibration of the  $\epsilon$ ), which is approximately three times larger than the forces measured (7). For PDB:1GFL, the extrapolated force in protocol II-C would be  $\sim 120$  pN. A further discussion of these issues is provided in the Supporting Material.

### Tension at the supported backward terminus

The results discussed in the previous section seem to suggest that the motor force needed to unfold the proteins is considerably larger than the stalling force measured in the experiments of Maillard et al. (7), defined as a force that needs to be applied to the microbead in an optical-tweezers experiment to halt the translocation process. However, as discussed by Alegre-Cebollada et al. (47), these two forces are in general different. As shown in the inset of Fig. 5 (see also Fig. S11), there are three external forces acting on the protein in such a case: the motor force  $F_p$ , the pore reaction force  $F_t$  (21,47) (with which the rim of the pore acts on the neighboring residues), and the force due to the external measuring device ( $F_r$ ). Except during the unfolding

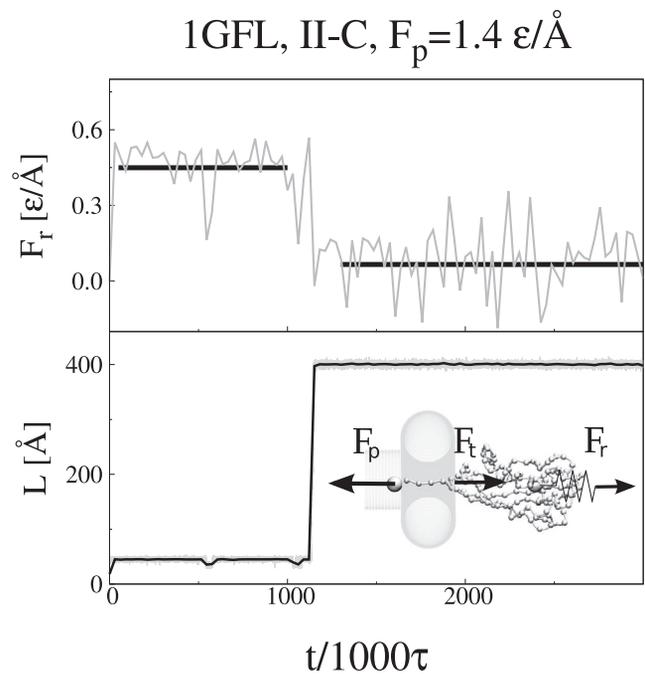


FIGURE 5 An example of a jumping event in model PDB:1GFL in protocol II-C at  $F_p = 1.4 \text{ } \epsilon/\text{\AA}$ . (Bottom panel)  $L$  in one trajectory. (Inset) Forces involved in a schematic way. The force due to the torus is denoted by  $F_t$  so that  $F_r - F_p - F_t$ . (Top panel) Corresponding tension at the backward terminus. (Horizontal lines) Average levels of the force.

events, the total force on the protein should sum up to zero, i.e.,  $F_r + F_t = F_p$ , so that the measured  $F_r$  will always be smaller than the pulling force  $F_p$ . However, because the translocation takes place after the unfolding event, we expect  $F_t$  to have a relatively small contribution to the overall force balance. In the Supporting Material, we show that for a homopolymer (with no attractive contacts),  $F_r$  fluctuates around  $F_p$  if  $F_p$  is  $> \sim 0.8 \text{ } \epsilon/\text{\AA}$  (i.e., larger than  $\sim 80$  pN). However, at smaller pulling forces—of relevance for the proteasome—the homopolymer is not as taut and the two forces differ noticeably. This is because of the entropically formed coiled conformations that wander off the pulling axis and exert a force on the funnel.

It is of interest to study the interplay among  $F_p$ ,  $F_t$ , and  $F_r$  during the entire translocation process. Fig. 5 shows  $F_r$  near a jumping event in PDB:1GFL in protocol II-C. The value  $F_r$  first fluctuates around  $F_{ri} = 0.44 \text{ } \epsilon/\text{\AA}$  and then jumps down to a new fluctuation level of  $\sim 0.05 \text{ } \epsilon/\text{\AA}$ . (We neglect the very beginning of pulling when the monitoring spring is slack.) In this case, the motor force is equal to  $1.4 \text{ } \epsilon/\text{\AA}$ —three times larger than  $F_{ri}$ , independent of the calibration of  $\epsilon$ . Much of the resistance to pulling is taken up by the proteasome itself.

Furthermore, let us note that the motor force  $F_p$  does not generate a uniform tension in the backbone (see Fig. S11). Only the parts of the protein already inside the pore and near the pore entrance are stretched with the force  $F_p$ , whereas the rest of the protein is stretched with force  $F_r$ . In this particular case, approximately two-thirds of  $F_p$  is

spent on overcoming the wall resistance at the entrance and just one-third generates tension that is monitored. We find that  $F_{r,i}$  is essentially constant as a function of  $F_p$ . It does depend, however, on the distance,  $s$ , by which each atom of the protein and the attached spring can be shifted away from the near-touch situation when constructing the initial state of the system. One expects a better adaptation for the translocation when  $s$  is increased.

For  $s$  of 3 Å, we get  $F_{r,i}$  of  $\sim 0.6 \text{ e}/\text{Å}$ , i.e., 0.43 of  $F_p$ . Pilot runs for  $s$  of 10 Å yield an  $F_{r,i}$  value quite distinct from  $F_p$ . The fact that the folded part of the protein is under the inhomogeneous tension needs to be taken into account in any theoretical model of the translocation process. In particular, for small  $F_r$  the largest tension ( $\sim F_p$ ) acts on the protein structures near the pore rim. The unfolding rate of those structures are then accelerated by the factor  $\exp(-F_p \delta x / k_B T)$ , hence they are more prone to break first, which results in a unfolding pathway different from that in AFM-based pulling. However, as  $F_r$  is increased, the tension in the remaining part of the protein ( $\sim F_r$ ) builds up. Finally, near the stalling (where  $F_r$  becomes closer to  $F_p$ ), all of the structures are stretched with a similar force and the unfolding pathway will converge toward that observed in standard AFM unfolding experiments.

It should be noted that even though the translation and unfolding are overlapping in time, the actual measurement of the stalling force takes place after the unfolding event (7). It is thus worthwhile to analyze the situation after this event. In such a case, the pore rim reaction force,  $F_r$ , suddenly drops to a much lower value and there appears a force imbalance between  $F_r$  and  $F_p$ . Because the motion is overdamped, the center of mass of the protein will then begin to move with velocity  $v \sim (F_p - F_r - F_t)$ . Such a linear dependence of the translocation velocity on  $F_r$  has indeed been observed experimentally (see Fig. 2B in Maillard et al. (7)), and is in itself a confirmation of the nonzero value of  $F_r$ .

## DISCUSSION

Our studies show that proteasome-induced translocation is a more efficient way of unfolding the proteins than AFM-induced unfolding unless the protein is small and weakly stable, as then overcoming the constriction has a bigger effect than mechanical help in unraveling. The interplay between the pulling force and the steric forces at the pore entrance create inhomogeneous tension in the chain, which catalyzes unfolding. Such a setup allows for making adjustments and rotations that facilitate unfolding. Furthermore, the region of the top local tension moves along the protein as it unfolds, but stays near the pore entrance. Finally, the steric constraints associated with the pore may prevent refolding (48) in analogy to the situation in mitochondrial-pore translocation (49). As shown in Figs. 3 and 4, such a focused, cotranslocational unfolding is much more efficient than an AFM-mediated force-clamp unfolding, with the unfolding times being

several orders-of-magnitude shorter. In addition, repetitive pulling, invoked in Tian and Andricioaei (24) and Kravats et al. (25), may help reduce generation of transient misfolded conformations at the expense of a longer overall duration of the process (and large consumption of ATP) compared to continuous pulling. These mechanisms should operate independent of the disordering influence of the poly-ubiquitin tags discussed in Hagai and Levy (50) and Hagai et al. (51). It is difficult to assess the relative weights of these various mechanisms, however. The general tendencies to enhance efficiency are subject to variations, depending on a specific protein pulled. Pore-mediated unfolding proceeds along a different pathway (49), which may lead to kinetic traps and dramatic slowing down, as shown for barnase (24).

Finally, we have shown that the force with which the folded part of the protein is stretched in single-molecule force spectroscopy experiments involving proteasome ( $F_r$ ) is in general considerably smaller than the motor force ( $F_p$ ). This fact needs to be taken into account in the theoretical interpretation of the experimental results. Considering that protein unfolding by the AAA+ motor of the proteasome is similar to that of the mitochondrial import (13,31,32,52), our conclusions can likely have a bearing on experiments on mitochondrial and other translocases.

## SUPPORTING MATERIAL

Fourteen figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00780-2](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00780-2).

We appreciate discussions with A. Galera-Prat, À. Gómez-Sicilia, R. Hervás, D. Laurents, Y. Levy, M. Sikora, and G. Stan.

Computer resources were financed by the European Regional Development Fund under the Operational Programme Innovative Economy NanoFun grant No. POIG-02-02-00-00-025/09. In Poland, this research has been supported by the National Science Center grants No. N N202 055440 (to M.C. and P.S.) and No. 2011/01/B/ST3/02190 (to M.W.). In Spain, it has been supported by Ministerio de Ciencia e Innovación grant No. BIO2010-22275 and Comunidad de Madrid grant No. S-2009MAT-1507.

## REFERENCES

1. Alberts, B., A. Johnson, ..., P. Walter. 2002. *Molecular Biology of the Cell*. Garland Science, New York.
2. Coux, O., K. Tanaka, and A. L. Goldberg. 1996. Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* 65:801–847.
3. Bochtler, M., L. Ditzel, ..., R. Huber. 1999. The proteasome. *Annu. Rev. Biophys. Biomol. Struct.* 28:295–317.
4. Goldberg, A. L. 1990. ATP-dependent proteases in prokaryotic and eukaryotic cells. *Semin. Cell Biol.* 1:423–432.
5. Gottesman, S. 1996. Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* 30:465–506.
6. Ruschak, A. M., T. L. Religa, ..., L. E. Kay. 2010. The proteasome antechamber maintains substrates in an unfolded state. *Nature.* 467:868–871.
7. Maillard, R. A., G. Chistol, ..., C. Bustamante. 2011. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell.* 145:459–469.

8. Aubin-Tam, M.-E., A. O. Olivares, ..., M. J. Lang. 2011. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell*. 145:257–267.
9. Sauer, R. T., and T. A. Baker. 2011. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* 80:587–612.
10. Sen, M., R. A. Maillard, ..., C. Bustamante. 2013. The ClpXP protease unfolds substrates using a constant rate of pulling but different gears. *Cell*. 155:636–646.
11. Tonddast-Navaei, S., and G. Stan. 2013. Mechanism of transient binding and release of substrate protein during the allosteric cycle of the p97 nanomachine. *J. Am. Chem. Soc.* 135:14627–14636.
12. Kravats, A. N., S. Tonddast-Navaei, ..., G. Stan. 2013. Asymmetric processing of a substrate protein in sequential allosteric cycles of AAA+ nanomachines. *J. Chem. Phys.* 139:121921.
13. Prakash, S., and A. Matouschek. 2004. Protein unfolding in the cell. *Trends Biochem. Sci.* 29:593–600.
14. Cieplak, M., and T. X. Hoang. 2003. Universality classes in folding times of proteins. *Biophys. J.* 84:475–488.
15. Sułkowska, J. I., and M. Cieplak. 2007. Mechanical stretching of proteins—a theoretical survey of the Protein Data Bank. *J. Phys. Cond. Mat.* 19:283201.
16. Sułkowska, J. I., and M. Cieplak. 2008. Selection of optimal variants of Gö-like models of proteins through studies of stretching. *Biophys. J.* 95:3174–3191.
17. Sikora, M., J. I. Sułkowska, and M. Cieplak. 2008. Mechanical strength of 17,132 model proteins and cysteine slipknots. *PLoS Comp. Biol.* 5:e1000547.
18. Muthukumar, M. 2007. Mechanism of DNA transport through pores. *Annu. Rev. Biophys. Biomol. Struct.* 36:435–450.
19. Kirmizialtin, S., V. Ganesan, and D. E. Makarov. 2004. Translocation of a  $\beta$ -hairpin-forming peptide through a cylindrical tunnel. *J. Chem. Phys.* 121:10268–10277.
20. Huang, L., S. Kirmizialtin, and D. E. Makarov. 2005. Computer simulations of the translocation and unfolding of a protein pulled mechanically through a pore. *J. Chem. Phys.* 123:124903.
21. West, D. K., D. J. Brockwell, and E. Paci. 2006. Prediction of the translocation kinetics of a protein from its mechanical properties. *Biophys. J.* 91:L51–L53.
22. Makarov, D. E. 2009. Computer simulations and theory of protein translocation. *Acc. Chem. Res.* 42:281–289.
23. Szymczak, P. 2013. Tight knots in proteins: can they block the mitochondrial pores? *Biochem. Soc. Trans.* 41:620–624.
24. Tian, P., and I. Andricioaei. 2005. Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mitochondrial pore. *J. Mol. Biol.* 350:1017–1034.
25. Kravats, A., M. Jayasinghe, and G. Stan. 2011. Unfolding and translocation pathway of substrate protein controlled by structure in repetitive allosteric cycles of the ClpY ATPase. *Proc. Natl. Acad. Sci. USA*. 108:2234–2239.
26. Lee, C., S. Prakash, and A. Matouschek. 2002. Concurrent translocation of multiple polypeptide chains through the proteasomal degradation channel. *J. Biol. Chem.* 277:34760–34765.
27. Burton, J. L., Y. Xiong, and M. J. Solomon. 2011. Mechanisms of pseudosubstrate inhibition of the anaphase promoting complex by Acl1. *EMBO J.* 30:1818–1829.
28. Valbuena, A., J. Oroz, ..., M. Carrión-Vázquez. 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl. Acad. Sci. USA*. 106:13791–13796.
29. Weber-Ban, E. U., B. G. Reid, ..., A. L. Horwich. 1999. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature*. 401:90–93.
30. Kenniston, J. A., T. A. Baker, ..., R. T. Sauer. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell*. 114:511–520.
31. Lee, C., M. P. Schwartz, ..., A. Matouschek. 2001. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol. Cell*. 7:627–637.
32. Kenniston, J. A., R. E. Burton, ..., R. T. Sauer. 2004. Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J. Struct. Biol.* 146:130–140.
33. Tsai, J., R. Taylor, ..., M. Gerstein. 1999. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 290:253–266.
34. Zhang, F., M. Hu, ..., Y. Shi. 2009. Structural insights into the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. *Mol. Cell*. 34:473–484.
35. Groll, M., L. Ditzel, ..., R. Huber. 1997. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature*. 386:463–471.
36. Berko, D., S. Tabachnick-Cherny, ..., A. Navon. 2012. The direction of protein entry into the proteasome determines the variety of products and depends on the force needed to unfold its two termini. *Mol. Cell*. 48:601–611.
37. Chwastyk, M., A. Galera, ..., M. Cieplak. 2014. Theoretical tests of the mechanical protection strategy in protein nanomechanics. *Proteins*. 82:717–726.
38. Johnston, J. A., E. S. Johnson, ..., A. Varshavsky. 1995. Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. *J. Biol. Chem.* 270:8172–8178.
39. Schlegel, H. B., M. Poe, and K. Hoogsteen. 1981. Models for binding of methotrexate to *Escherichia coli* dihydrofolate reductase: direct effect of carboxylate of aspartic acid 27 upon ultraviolet spectrum of methotrexate. *Mol. Pharmacol.* 20:154–158.
40. Szymczak, P., and M. Cieplak. 2006. Stretching of proteins in a force-clamp. *J. Phys. Cond. Mat.* 18:L21–L28.
41. Luccioli, S., A. Imparato, ..., A. Torcini. 2010. Unfolding times for proteins in a force clamp. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 81:010902.
42. Kirmizialtin, S., L. Huang, and D. E. Makarov. 2006. Computer simulations of protein translocation. *Phys. Status Solidi*. 243:2038–2047.
43. Matysiak, S., A. Montesi, ..., C. Clementi. 2006. Dynamics of polymer translocation through nanopores: theory meets experiment. *Phys. Rev. Lett.* 96:118103.
44. Nager, A. R., T. A. Baker, and R. T. Sauer. 2011. Stepwise unfolding of a  $\beta$  barrel protein by the AAA+ ClpXP protease. *J. Mol. Biol.* 413:4–16.
45. Dudko, O. K., J. Mathé, ..., G. Hummer. 2007. Extracting kinetics from single-molecule force spectroscopy: nanopore unzipping of DNA hairpins. *Biophys. J.* 92:4188–4195.
46. Huang, L., and D. E. Makarov. 2008. Translocation of a knotted polypeptide through a pore. *J. Chem. Phys.* 129:121107.
47. Alegre-Cebollada, J., P. Kosuri, and J. M. Fernández. 2011. Protease power strokes force proteins to unfold. *Cell*. 145:339–340.
48. Martin, A., T. A. Baker, and R. T. Sauer. 2008. Protein unfolding by AAA+ protease: critical dependence on ATP-hydrolysis rates, energy landscapes, and substrate engagement. *Nat. Struct. Biol.* 15:139–145.
49. Sato, T., M. Esaki, ..., T. Endo. 2005. Comparison of the protein-unfolding pathways between mitochondrial protein import and atomic-force microscopy measurements. *Proc. Natl. Acad. Sci. USA*. 102:17999–18004.
50. Hagai, T., and Y. Levy. 2010. Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. *Proc. Natl. Acad. Sci. USA*. 107:2001–2006.
51. Hagai, T., A. Azia, ..., Y. Levy. 2011. Intrinsic disorder in ubiquitination substrates. *J. Mol. Biol.* 412:319–324.
52. Wilcox, A. J., J. Choy, ..., A. Matouschek. 2005. Effect of protein structure on mitochondrial import. *Proc. Natl. Acad. Sci. USA*. 102:15435–15440.