

# Order Statistics Theory of Unfolding of Multimeric Proteins

A. Zhmurov,<sup>†‡</sup> R. I. Dima,<sup>§</sup> and V. Barsegov<sup>†‡\*</sup>

<sup>†</sup>Department of Chemistry, University of Massachusetts, Lowell, Massachusetts; <sup>‡</sup>Moscow Institute of Physics and Technology, Moscow, Russia; and <sup>§</sup>Department of Chemistry, University of Cincinnati, Cincinnati, Ohio

**ABSTRACT** Dynamic force spectroscopy has become indispensable for the exploration of the mechanical properties of proteins. In force-ramp experiments, performed by utilizing a time-dependent pulling force, the peak forces for unfolding transitions in a multimeric protein ( $(D)_N$ ) are used to map the free energy landscape for unfolding for a protein domain  $D$ . We show that theoretical modeling of unfolding transitions based on combining the observed first ( $f_1$ ), second ( $f_2$ ), ...,  $N^{\text{th}}$  ( $f_N$ ) unfolding forces for a protein tandem of fixed length  $N$ , and pooling the force data for tandems of different length,  $n_1 < n_2 < \dots < N$ , leads to an inaccurate estimation of the distribution of unfolding forces for the protein  $D$ ,  $\psi_D(f)$ . This problem can be overcome by using Order statistics theory, which, in conjunction with analytically tractable models, can be used to resolve the molecular characteristics that determine the unfolding micromechanics. We present a simple method of estimation of the parent distribution,  $\psi_D(f)$ , based on analyzing the force data for a tandem  $(D)_n$  of arbitrary length  $n$ . Order statistics theory is exemplified through a detailed analysis and modeling of the unfolding forces obtained from pulling simulations of the monomer and oligomers of the all- $\beta$ -sheet  $WW$  domain.

## INTRODUCTION

Mechanical functions of intra- and extracellular multimeric proteins play an essential role in diverse biological processes from cytoskeleton support and cell motility (1) to cell adhesion and the formation of the extracellular matrix (2,3), to muscle contraction and relaxation (4), and to blood clotting (5). For example, actin cross-linking filamins play an important role in cellular locomotion (1,6), in which the mechanical stability of filamin domains provides the flexibility necessary for actin cross-links (7). The extracellular matrix (ECM) determines the tensile strength and the elasticity of tissues. Fibronectin type III ( $FnIII$ ) domains of the  $Fn$  tandems of the ECM offer multiple binding sites for the components of the ECM assembly, including the integrin receptors, triggered by mechanical stretching (8). Titin, a giant modular protein, which consists of immunoglobulin ( $Ig$ ) and fibronectin ( $Fn$ ) domains (1,9), plays a crucial role in muscle contraction and relaxation (10). Fibrinogen ( $Fb$ ) polymerizes and aggregates into fibrin oligomers, which assemble into the branched network of fibrin fibers called a blood clot (5,11). To form a temporary plug to seal an injury, fibrin fibers sustain the large shear stress due to arterial blood flow.

Many mechanically active proteins perform their biological functions by adopting a modular structure  $(D)_N$  or  $D_1 - D_2 - \dots - D_N$  formed by connected head-to-tail identical protein domains  $D$  or nonidentical domains  $D_1, D_2, \dots, D_N$ . These structures are referred to as homogeneous tandems of identical protein domains  $D$  or heterogeneous tandems of nonidentical domains  $D_1, D_2, \dots, D_N$ . An example of the homogeneous protein tandems (oligomers)

$(WW)_4$  and  $(WW)_6$ , formed by C-terminal-to-N-terminal connected all- $\beta$ -sheet  $WW$  domains, is presented in Fig. 1. Titin consists of a few hundreds of  $Ig$  and  $Fn$  domains, with the number of these domains varying among the different titin molecules (1). Fibronectin tandems are formed by  $\sim 20$   $Fn$  domains of types  $FnI$ ,  $FnII$ , and  $FnIII$  (8). In addition, engineered polyprotein constructs are used by experimentalists to characterize the mechanical properties of proteins. For example, fibrinogen oligomers  $(Fb)_N$  of varying length  $N = 3-16$  have been used as single-chain models to explore the physical characteristics of fibrin oligomers (12).

Single-molecule techniques, which utilize atomic-force microscopy and optical tweezer-based force spectroscopy, are widely used to access conformational transitions in proteins (12–15). In a force-ramp assay on a multimeric protein  $(D)_N$  composed of identical domains ( $D$ ), the time-dependent mechanical force  $f(t) = r_f t$ , where  $r_f = \kappa v_0$  is the force-loading rate that depends on the cantilever spring constant  $\kappa$  and the pulling speed  $v_0$ , is used to induce the mechanical unfolding reactions (16). These are analyzed by building the histogram of unfolding forces (peak forces) (12),  $f_1, f_2, \dots, f_N$ , extracted from the force-extension curves, to infer the probability density function (pdf) of unfolding forces for a protein domain  $D$ ,  $\psi_D(f)$ . In principle, one can obtain  $N$  force values for a tandem  $(D)_N$  in just one measurement, but in practice, because the cantilever tip can pick up a tandem of any length  $n$  ( $1 \leq n \leq N$ ), the number of peak forces varies from one measurement to another. The force data from many measurements and for several values of  $n$  are then combined into a single set and analyzed.

Combining all the force data for a homogeneous tandem  $(D)_n$  of fixed length  $n$ , i.e., forming a single set

Submitted April 1, 2010, and accepted for publication July 14, 2010.

\*Correspondence: [valeri\\_barsegov@uml.edu](mailto:valeri_barsegov@uml.edu)

Editor: Edward H. Egelman.

© 2010 by the Biophysical Society  
0006-3495/10/09/1959/10 \$2.00

doi: 10.1016/j.bpj.2010.07.012

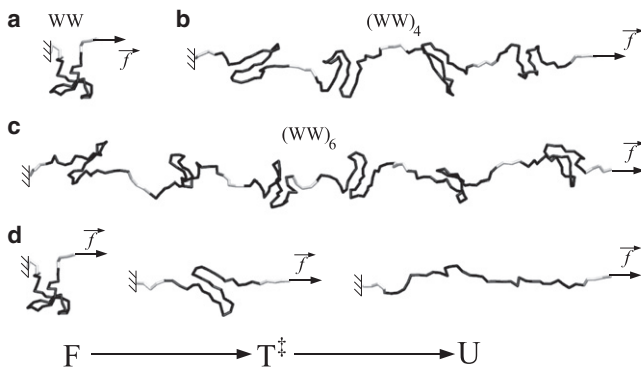


FIGURE 1 (a) Schematic representation of the all- $\beta$ -sheet WW domain (PDB code: 1PIN) in the folded state. (b and c) Schematic representation of the native structures for the tandems of four WW domains (b) and six WW domains (c), formed by the C-terminal-to-N-terminal connected WW domains (shown in *solid representation*) through flexible linkers (shown in *shading*). In pulling simulations for the WW monomer (*single domain*), the time-dependent mechanical force  $f(t)$  (*solid arrows*) is applied to the C-terminus of the molecule in the direction of the end-to-end vector, while the N-terminus is constrained. In pulling simulations for  $(WW)_4$  and  $(WW)_6$ , the force is applied to the C-terminus of the last WW domain, while the N-terminus of the first domain is fixed. (d) The forced unfolding transitions from the native state  $F$  to the globally unfolded state  $U$  in the single WW domain and in the WW domains forming the tandems  $(WW)_4$  and  $(WW)_6$  occur in a single step via formation of the transition state for unfolding  $T^\ddagger$ . This is followed by the simultaneous disruption of almost all the native contacts stabilizing the folded state.

$$f_{(D)_n}^{\text{comb}} = \{f_1, f_2, \dots, f_n\} \quad (1 \leq n \leq N),$$

and pulling together the force data for tandems of different length  $n_1 \leq n_2 \leq \dots \leq N$ ,

$$f_{(D)_{n_1, (D)_{n_2}, \dots, (D)_N}}^{\text{comb}} = \left\{ \left\{ f_{(D)_{n_1}}^{\text{comb}} \right\}, \left\{ f_{(D)_{n_2}}^{\text{comb}} \right\}, \dots, \left\{ f_{(D)_N}^{\text{comb}} \right\} \right\}$$

is based on the hypothesis that the forced unfolding transitions observed in a tandem  $(D)_N$  characterize the mechanical properties of a single protein  $D$ . In statistical terms, this translates into the assumption that the unfolding forces obtained for the tandem and for the single domain are sampled from the same distribution, i.e.,  $\psi_D(f) = \psi_{(D)_N}(f)$ . We tested this assumption theoretically by carrying out pulling simulations for the WW monomer and oligomers  $(WW)_n$ , formed by connected identical all- $\beta$ -sheet WW domains, subject to force-ramp  $f(t)$ . The computer models and the simulation details are presented in the next section. We find that the histograms of unfolding forces for the single WW domain differ from the force histograms obtained for the combined data,  $f_{(WW)_n}^{\text{comb}} = \{f_1, f_2, \dots, f_n\}$ , for the tetramer  $(WW)_4$  ( $n = 4$ ) and for the hexamer  $(WW)_6$  ( $n = 6$ ) in the location of the most probable force  $f^*$  and in the width of the distribution, which quantifies fluctuations around the average force  $\bar{f}$  (Fig. 2, a and b). Indeed, the difference between  $f^*$  for WW and  $(WW)_4$  and for WW and  $(WW)_6$  of  $\sim 10$ – $15$  pN is significant, and is larger than the experimental error. The

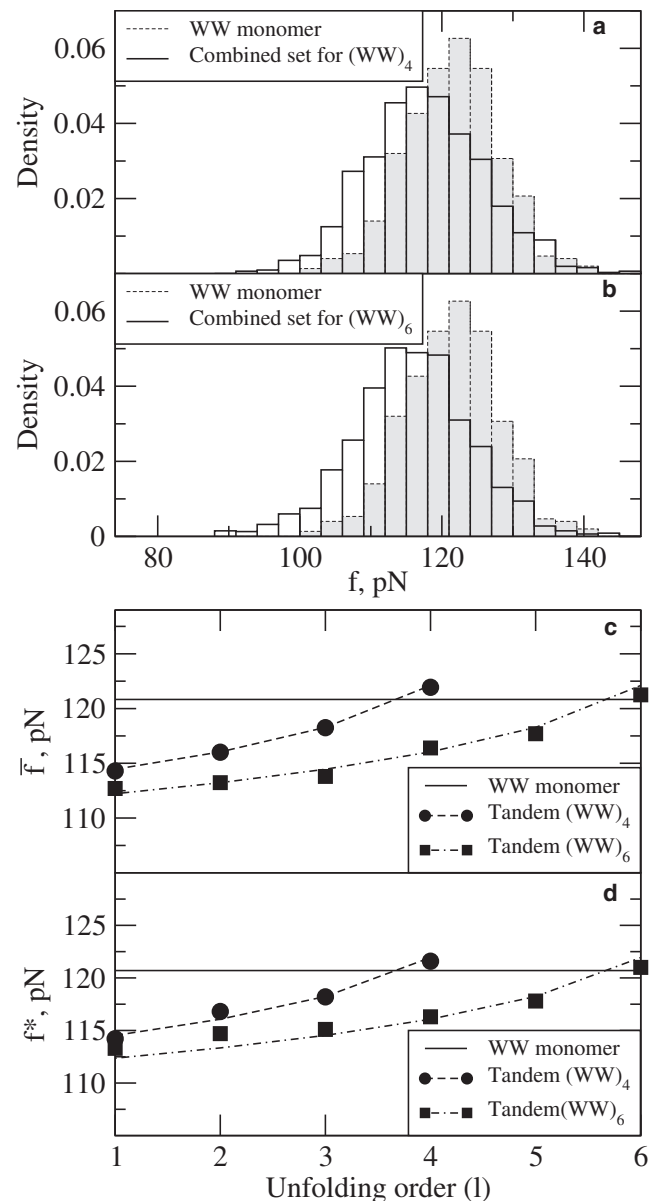


FIGURE 2 (a and b) Comparison of the histograms of the combined unfolding forces,  $f_{(WW)_n}^{\text{comb}} = \{f_1, f_2, \dots, f_n\}$ , for the WW oligomers  $(WW)_4$  (a) and  $(WW)_6$  (b) with the histogram of unfolding forces for the single WW domain (monomer). Each histogram is constructed using  $p = 260$  data points (peak forces) and the Freedman-Diaconis rule for the optimal bin size,  $h_{\text{opt}} = 2IQRp^{-1/3}$  ( $IQR$  is the interquartile range) (33,34). (c and d) Comparison of the average force,  $\bar{f}_{(WW)_n, l}$  (c), and the most probable force,  $f_{(WW)_n, l}^*$  (d), as a function of the unfolding order  $l = 1, 2, \dots, 4$  for  $(WW)_4$  and  $l = 1, 2, \dots, 6$  for  $(WW)_6$  (*data points*), with the same quantities,  $\bar{f}_{WW}$  and  $f_{WW}^*$ , for the WW monomer (*horizontal reference lines*). (*Ascending dashed and dash-dotted lines*) Values of  $\bar{f}_{1:n-l+1}$  and  $f_{1:n-l+1}^*$  obtained using Order statistics theory.

width of the histogram is  $\sim 14$  pN for WW versus 20 pN for  $(WW)_4$  and  $(WW)_6$ . We also compared the average and most probable unfolding forces for the WW monomer,  $\bar{f}_{WW}$  and  $f_{WW}^*$ , with the same quantities,  $\bar{f}_{(WW)_n, l}$  and  $f_{(WW)_n, l}^*$ , calculated for the first through the fourth unfolding

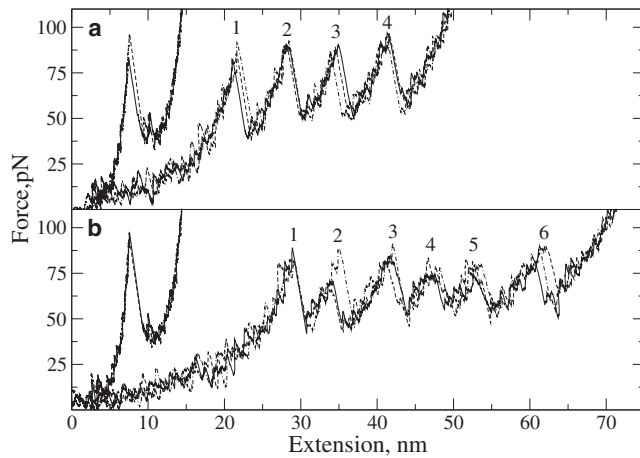


FIGURE 3 (a and b) Force-extension curves obtained for the WW monomer (curves with a single force peak) and oligomers  $(WW)_4$  (a) and  $(WW)_6$  (b). The most typical force-extension profiles are overlaid to demonstrate the stochastic nature of the biomechanical unfolding reactions in multimeric proteins. The force peaks,  $f_{(WW)_n,l}$  ( $l = 1, 2, \dots, n$ ), for consecutive unfolding transitions in the tandems  $(WW)_4$  and  $(WW)_6$  are numbered as 1, 2, ..., 4 (a) and 1, 2, ..., 6 (b), respectively. Each unfolding transition is accompanied by a substantial tension drop in the protein chain, which results in characteristic sawtooth force-extension profiles.

transition ( $l = 1, 2, \dots, 4$ ) for  $(WW)_4$  (Fig. 3 a), and for the first through the sixth unfolding transition ( $l = 1, 2, \dots, 6$ ) for  $(WW)_6$  (Fig. 3 b). We see that the data points for  $\bar{f}_{(WW)_n,l}$  and for  $f_{(WW)_n,l}^*$  are below the reference line for  $\bar{f}_{WW}$  and for  $f_{WW}^*$ , respectively, for all the values of the unfolding order  $l$ , except for the last transition, i.e.,  $l = 4$  for  $(WW)_4$  and  $l = 6$  for  $(WW)_6$  (Fig. 2, c and d).

These results demonstrate that the physical characteristics of protein domains ( $D$ ) combined in a tandem  $(D)_n$  differ from those for the same but single domain  $D$ . Hence, the inference of the parent pdf of unfolding forces for the (monomer)  $D$ ,  $\psi_D(f)$ , and the mapping of the free energy landscape for unfolding require using a more elaborate theoretical framework. Why are the distributions of unfolding forces for the WW monomer and for the oligomers  $(WW)_4$  and  $(WW)_6$  different (Fig. 2, a and b)? Why is every next unfolding transition in  $(WW)_4$  and  $(WW)_6$  characterized by larger values of  $\bar{f}_{(WW)_n,l}$  and  $f_{(WW)_n,l}^*$  compared to every previous transition (Fig. 2, c and d)?

Here, we use Order statistics (17,18) to describe the mechanical unfolding reactions in proteins subject to a time-dependent pulling force (force-ramp). We show that Order statistics offers an intuitive simple explanation for the observed results (Fig. 2), and answers the long-standing question whether the mechanical property of a tandem  $(D)_N$  is the sum of mechanical properties of single domains  $D$ . Order statistics theory allows for accurate interpretation and modeling of the unfolding force data observed in multimeric proteins (protein tandems) and engineered polyproteins. We also present an algorithm for the estimation of the parent distribution of unfolding forces for a protein

domain  $D$ ,  $\psi_D(f)$ , based on the first-order statistic of unfolding forces collected for a tandem  $(D)_n$ .

## THEORETICAL METHODS

### Computer models of the WW monomer and oligomers

We used  $C_\alpha$ -based self-organized polymer models (19) for the single WW domain (monomer), and for the tandems (oligomers)  $(WW)_4$  and  $(WW)_6$  formed by the connected identical all- $\beta$ -sheet WW domains (Fig. 1). The protein tandems  $(WW)_4$  and  $(WW)_6$  serve as a conceptual representation of wild-type multidomain proteins. The WW domain (34 residues, Protein DataBank (PDB) code: 1PIN (20)), which has been extensively studied experimentally (21,22) and computationally (23,24), is involved in the regulation of several cellular processes. The WW domain has been a paradigm for describing folding and unfolding of the  $\beta$ -sheet proteins (21). The mechanical unraveling of WW can be described by the single-step kinetics of unfolding,  $F \rightarrow U$ , from the folded state  $F$  to the unfolded state  $U$ , which makes this protein a suitable test system. The tandems  $(WW)_4$  and  $(WW)_6$  were constructed by connecting the  $N$ - and  $C$ -termini of the adjacent WW domain using linkers of four neutral residues (Fig. 1, b and c). Short flexible linkers do not change the micromechanics of unfolding in multimeric proteins. Each residue was represented by its  $C_\alpha$ -atom with the  $C_\alpha$ - $C_\alpha$  covalent bond distance of  $a = 3.8 \text{ \AA}$ , which corresponds to the length of a peptide bond. Structural analysis of WW,  $(WW)_4$ , and  $(WW)_6$  has revealed that the number of native contacts stabilizing their folded state are, respectively, 65, 260, and 390 ( $R_C = 8.0 \text{ \AA}$ -cutoff distance).

The molecular potential energy function for a protein conformation, specified in terms of the residue coordinates  $\{r_i\}$ ,  $i = 1, 2, \dots, M$ , is given by (16,19,25)

$$\begin{aligned}
 V_{\text{MOL}} &= V_{\text{FENE}} + V_{\text{NB}}^{\text{ATT}} + V_{\text{NB}}^{\text{REP}} \\
 &= -\sum_{i=1}^{M-1} \frac{k}{2} R_0^2 \log \left( 1 - \frac{(r_{i,i+1} - r_{i,i+1}^0)^2}{R_0^2} \right) \\
 &\quad + \sum_{i=1}^{M-3} \sum_{j=i+3}^M h \left[ \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^0}{r_{ij}} \right)^6 \right] \Delta_{ij} \\
 &\quad + \sum_{i=1}^{M-2} l \left( \frac{\sigma}{r_{i,i+2}} \right)^6 + \sum_{i=1}^{M-3} \sum_{j=i+3}^M l \left( \frac{\sigma}{r_{ij}} \right)^6 (1 - \Delta_{ij}),
 \end{aligned} \tag{1}$$

where the distance between any two interacting residues  $i$  and  $i+1$  is  $r_{i,i+1}$ , whereas  $r_{i,i+1}^0$  is its value in the native (PDB) structure. The first term in Eq. 1 is the finite extensible nonlinear elastic (FENE) potential, which describes the backbone chain connectivity. Here,  $R_0 = 2 \text{ \AA}$  is the tolerance in the change of a covalent bond, and  $k = 1.4 \text{ N/m}$  is the force constant. The second term in Eq. 1 is the Lennard-Jones potential ( $V_{\text{NB}}^{\text{ATT}}$ ), which accounts for the native interactions that stabilize the folded state. We assumed that if the non-covalently linked residues  $i$  and  $j$  ( $|i-j| > 2$ ) are within the cutoff distance  $R_C$  in the native state, i.e.,  $r_{ij} < R_C$ , then  $\Delta_{ij} = 1$ , and zero otherwise. We used a uniform value of  $\epsilon_n = 1.5 \text{ kcal/mol}$ , which specifies the strength of the nonbonded interactions, and all nonnative interactions (fourth term in Eq. 1) were treated as repulsive ( $V_{\text{NB}}^{\text{REP}}$ ). Additional constraint was imposed on the bond angle formed by residues  $i, i+1$ , and  $i+2$  by including the repulsive potential with parameters  $\epsilon_l = 1 \text{ kcal/mol}$  and  $\sigma = 3.8 \text{ \AA}$ , which quantify, respectively, the strength and the range of repulsion. To ensure self-avoidance of the protein chain, we set  $\sigma = 3.8 \text{ \AA}$ .

### Simulations of mechanical unfolding

The unfolding dynamics was obtained by integrating the Langevin equations for each particle position  $r_i$  in the overdamped limit,

$$\eta d\mathbf{r}_i/dt = -\partial V/\partial \mathbf{r}_i + \mathbf{g}_i(t).$$

Here,  $V = V_{\text{MOL}} - \mathbf{f}\mathbf{X}$  is the total potential energy,  $\mathbf{g}(t)$  is the Gaussian distributed random force,  $\eta$  is the friction coefficient. To mimic the force-ramp measurement, in each simulation run the  $N$ -terminal  $C_{\alpha}$ -atom of the molecule was constrained and a time-dependent force  $\mathbf{f}(t) = f(t)\mathbf{n}$  with the magnitude  $f(t) = r_f t$  was applied to the C-terminal  $C_{\alpha}$ -atom of molecule in the direction  $\mathbf{n}$  of the end-to-end vector  $\mathbf{X}$  (Fig. 1). The Langevin equations were propagated with the time step  $\Delta t = 0.08\tau_H = 20$  ps, where  $\tau_N = \zeta \varepsilon_r \tau_L / k_B T$ .

Here,  $\tau_L = (ma^2/\varepsilon_h)^{1/2} = 3$  ps,  $\zeta = 50$  is the dimensionless friction constant for a residue in water ( $\eta = \zeta m/\tau_L$ ), and  $m \approx 3 \times 10^{-22}$  g is the residue mass (26). Pulling simulations were carried out at room temperature ( $T = 300$  K) using the bulk water viscosity, which corresponds to the friction coefficient  $\eta = 7.0 \times 10^5$  pN ps/nm. We used experimentally relevant values of  $\kappa = 35$  pN/nm (cantilever spring constant) and  $v_0 = 2.5$   $\mu\text{m/s}$  (pulling speed), which translate to a force-loading rate  $r_f = \kappa v_0 = 87.5$  nN/s. Three sets of 260 unfolding trajectories were generated for  $WW$ ,  $(WW)_4$ , and  $(WW)_6$  molecules, and the unfolding forces (peak forces) were extracted from the force-extension curves (Fig. 3).

Structural analysis of the unfolding trajectories revealed that the single  $WW$  domain and the  $WW$  domains in the tandems  $(WW)_4$  and  $(WW)_6$  unfold in a single step,  $F \rightarrow U$  (Fig. 1 d), which is also evident from sharp force peaks observed in the force-extension profiles for  $WW$ , and for  $(WW)_4$  and  $(WW)_6$  (Fig. 3). The single-step unfolding transition in the  $WW$  domain is accompanied by the simultaneous rupture of a substantial fraction of the native contacts under the maximum force load (peak force). The fraction of native contacts in the folded (unfolded) state  $F(U)$  is  $\approx 0.85$  ( $\approx 0.15$ ), which agrees well with the results of folding simulations (23). The onset of unfolding transition occurs at the critical molecular elongation of  $X^* \approx (0.60 - 0.68)\Delta L \approx (7.5 - 8.5)\text{nm}$ , where  $\Delta L = 33a \approx 12.5$  nm is the maximal extension of the  $WW$  monomer. The molecular spring constant  $k_{\text{sp}}$ , which quantifies the elastic deformation of the native structure, was extracted from the force-extension curves (Fig. 3). We found that  $k_{\text{sp}} \approx 17.5$  pN/nm for the single  $WW$  domain, and that  $k_{\text{sp}} = 13.5$  pN/nm, 9.4 pN/nm, 8.0 pN/nm, and 9.0 pN/nm for the first ( $l = 1$ ), second ( $l = 2$ ), third ( $l = 3$ ), and fourth unfolding transition ( $l = 4$ ), respectively, for the tandem  $(WW)_4$ . We found similar variation in  $k_{\text{sp}}$  with the unfolding order  $l$  for the tandem  $(WW)_6$  (data not presented). The subsequent statistical analysis revealed that there are no domain-domain interactions in  $(WW)_4$  and  $(WW)_6$ , and that the unfolding forces are independent and identically distributed (i.i.d.).

## ORDER STATISTICS THEORY

### Force-clamp assay: statistics of unfolding times

Let us first consider the forced unfolding transitions observed in a force-clamp experiment on a homogeneous tandem  $(D)_N$  carried out using a constant pulling force  $f$ . Order statistics based description of the results of force-clamp measurements is given in our previous articles (27,28). In this type of measurement, the tandem end-to-end distance  $X$  shows characteristic stepwise increases,  $X_1(t_1)$ ,  $X_2(t_2)$ , ...,  $X_N(t_N)$ , observed at times  $t_1$ ,  $t_2$ , ...,  $t_N$ , which mark the unfolding transitions in the protein domains forming the tandem (13,14). Assume that the unfolding times,  $t_1$ ,  $t_2$ , ...,  $t_N$ , are independent and identically distributed (i.i.d.) random variables. The i.i.d. assumption holds, e.g., when the tandem is formed by identical protein domains that unfold independently, i.e., when there are no domain-domain interactions (27,28). Any domain  $D$  can unfold at any given time,  $t_1$ ,  $t_2$ , ...,  $t_N$ , with equal probability.

Then, by the very method of instrumentation used in the force-clamp measurements, i.e., the force applied remains constant and the mechanical fluctuations are minimal, the recorded unfolding times are time-ordered, i.e.,

$$t_{1:N} < t_{2:N} < \dots < t_{N:N},$$

where  $t_{l:N} = t_l$  ( $l = 1, 2, \dots, N$ ). In statistical terms, the unfolding times comprise a set of ordered time variates or order statistic  $t_{l:N}$  (27,28), where the  $l^{\text{th}}$  order statistic is defined as the  $l^{\text{th}}$  observation among  $N$  observations in a tandem of length  $N$ , such that  $l-1$  ( $N-l$ ) values are below (above) it. For example, for a tandem  $(D)_4$  the first, second, third, and fourth unfolding times  $t_1$ ,  $t_2$ ,  $t_3$ , and  $t_4$  are, respectively, order statistics  $t_{1:4}$ ,  $t_{2:4}$ ,  $t_{3:4}$ , and  $t_{4:4}$ . The challenge is to extract information about the unobserved parent time data, which describe the mechanical unfolding of protein domains  $D$  forming the tandem  $(D)_N$ , from the observed ordered time data  $t_{l:N}$ .

The  $l^{\text{th}}$  order statistic is described by the cumulative distribution function (cdf)  $\Phi_{l:N}(t)$  defined to be the probability that the  $l^{\text{th}}$  unfolding time  $t_{l:N}$  does not exceed  $t$ , i.e.,  $\Phi_{l:N}(t) = \text{Prob}(t_l \leq t)$ , and the pdf,  $\phi_{l:N}(t) = d\Phi_{l:N}(t)/dt$ . Under the i.i.d.-assumption, the cdf and pdf of the  $l^{\text{th}}$  unfolding times, observed in a force-clamp assay, are given by (17,18)

$$\begin{aligned} \Phi_{l:N}(t) &= \sum_{j=l}^N \binom{N}{j} \Psi(t)^j (1 - \Psi(t))^{N-j}, \\ \phi_{l:N}(t) &= N \binom{N-1}{l-1} \Psi(t)^{l-1} (1 - \Psi(t))^{N-l} \psi(t), \end{aligned} \quad (2)$$

where  $\binom{n}{j} = n!/j!(n-j)!$ .

In Eq. 2, the combinatorial prefactors take into account the number of possible combinations because there are multiple unfolding scenarios in a protein tandem. The function  $\Psi(t)$  is the parent cdf of unfolding forces for the single protein domain (monomer)  $D$ , defined as the probability that the unfolding transition in  $D$  has occurred before or at time  $t$ , and the parent pdf is  $\psi(t) = d\Psi(t)/dt$ . Hence, the expressions in Eq. 2 provide the necessary connection between the observed ordered time data for a tandem  $(D)_N$ , described by the cdf  $\Phi_{l:N}(t)$  and the pdf  $\phi_{l:N}(t)$ , and the unobserved parent distribution functions  $\Psi(t)$  and  $\psi(t)$ .

### Force-ramp assay: statistics of unfolding forces

In a force-ramp experiment, an applied force  $f(t) = r_f t$  increases linearly in time  $t$  with a force-loading rate  $r_f$ . The forced unraveling of a multimeric protein  $(D)_n$  is reflected in a sawtooth profile of the mechanical tension, experienced by the tandem chain, as a function of the tandem extension  $X$  (force-extension curve), in which consecutive force maxima (peak forces),  $f_1(X_1)$ ,  $f_2(X_2)$ , ...,  $f_N(X_n)$ , mark the unfolding transitions that occur in protein



domains (Fig. 3) (12). As in the force-ramp assay, the cantilever tip can pick up a tandem  $(D)_n$  of any length  $n$  ( $1 \leq n \leq N$ ), and when the unfolding forces are i.i.d., any domain could have unfolded at any given time resulting in any given peak force  $f_1, f_2, \dots, f_N$  with equal probability. The i.i.d.-assumption holds when the dynamics of propagation of the mechanical tension in the tandem chain remains decoupled from the unfolding kinetics (28). At the experimental pulling speeds of  $\nu_0 = 0.5\text{--}10 \mu\text{m/s}$ , the consecutive unfolding transitions can be viewed as independent events, but a significant increase in  $\nu_0$  might result in correlated transitions.

In the force-ramp assay, each unfolding transition is accompanied by a substantial tension drop in the tandem chain, which is reflected in the tension decrease from the maximum value (peak force) to some lower value (base line). As a result, the mechanical tension must be restored first before the next unfolding transition occurs (Fig. 3). In essence, once an unfolding transition in the tandem of length  $n$  has occurred, the pulling experiment is restarted. For example, after the first unfolding transition, the next (second) unfolding event is the first transition, but in the shorter tandem of length  $n - 1$ , the third unfolding event is the first transition in the tandem of length  $n - 2$ , etc. Hence, the first ( $f_1$ ), second ( $f_2$ ), etc., and  $n^{\text{th}}$  ( $f_n$ ) peak force is the first unfolding force in the tandem of length  $n$  ( $f_{1:n}$ ), the first unfolding force in the tandem of length  $n - 1$  ( $f_{1:n-1}$ ), etc., and, finally, the first unfolding force in the tandem of unit length ( $f_{1:1}$ ), respectively. For example, for a tandem  $(D)_4$  the first, second, third, and fourth unfolding forces  $f_1, f_2, f_3$ , and  $f_4$  are, respectively, the first-order statistics  $f_{1:4}, f_{1:3}, f_{1:2}$ , and  $f_{1:1}$ .

The peak forces  $f_l$  observed for a tandem  $(D)_n$  form the first-order statistic of unfolding forces in a sample of decreasing size  $n - l + 1 = n, n - 1, \dots, 1$  (number of still folded protein domains), i.e.,  $f_l = f_{1:n-l+1}$  ( $l = 1, 2, \dots, n$ ). Switching to force variables ( $f = r_f t$ ) and setting  $l = 1$  in Eq. 2, we obtain the cdf and the pdf of the first-order statistic of unfolding forces,

$$\begin{aligned} \Phi_{1:n-l+1}(f) &= 1 - (1 - \Psi(f))^{n-l+1}, \\ \phi_{1:n-l+1}(f) &= (n - l + 1)(1 - \Psi(f))^{n-l} \psi(f). \end{aligned} \quad (3)$$

The expressions in Eq. 3 provide the connection between the observed first-order statistic of unfolding forces, i.e., the cdf  $\Phi_{1:n-l+1}(f)$  and the pdf  $\phi_{1:n-l+1}(f)$ , and the unobserved parent distributions,  $\Psi(f)$  and  $\psi(f)$ . Also, the first expression in Eq. 3 shows that the probability that the first unfolding transition has not occurred at some finite force value  $\leq f$  (or time  $\leq t = f/r_f$ ) in a tandem of length  $n$ ,  $(1 - \Psi(f))^n$ , is exactly  $(1 - \Psi(f))$ -times smaller than the probability for a tandem of length  $n - 1$ ,  $(1 - \Psi(f))^{n-1}$ . Therefore, the first unfolding event ( $l = 1$ ) occurs faster, and hence at a lower force, in a tandem of longer length, which is intuitively clear as there are more domains in a longer tandem that could unfold. This

implies that the average first unfolding force  $\bar{f}_{1:n-l+1}$  should increase with  $l$ , which explains the results of Fig. 2 c.

### Order statistics-based modeling of unfolding forces

To exemplify Order statistics theory, we calculated the pdf of the first-order statistic of unfolding forces,  $\phi_{1:n-l+1}(f)$  (Eq. 3), and the average first unfolding force,  $\bar{f}_{1:n-l+1}$ , for a generic tandem  $(D)_n$ . We assumed the exponential kinetics of unfolding ( $F \rightarrow U$ ) and the Bell model for the force dependence of the unfolding rate constant  $k(f) = k_0 \exp [fx^\ddagger/k_B T]$  (29,30), where  $k_0$  is the force-free rate and  $x^\ddagger$  is the distance from the folded state  $F$  to the transition state  $T^\ddagger$  (Fig. 1 d). Under these assumptions, the parent cdf and the pdf of unfolding forces (for a single protein  $D$ ) are given by (30)

$$\begin{aligned} \Psi(f) &= 1 - \exp \left[ -\frac{k_B T k_0}{r_f x^\ddagger} \left( e^{fx^\ddagger/k_B T} - 1 \right) \right], \\ \psi(f) &= \frac{k_0}{r_f} \exp \left[ \frac{fx^\ddagger}{k_B T} - \frac{k_B T k_0}{r_f x^\ddagger} \left( e^{fx^\ddagger/k_B T} - 1 \right) \right]. \end{aligned} \quad (4)$$

The results of the calculation of  $\phi_{1:n-l+1}(f)$  and  $\bar{f}_{1:n-l+1}$  ( $l = 1, 2, \dots, n$ ) for generic tandems  $(D)_4$  and  $(D)_6$  are presented in Fig. 4 (parameters of the model are summarized in the caption). We see that the curves of  $\phi_{1:n-l+1}(f)$  for  $(D)_4$  shift toward larger forces as  $n - l + 1$  decreases, or  $l$  increases (Fig. 4 a). We obtained similar results for  $(D)_6$  (data not shown). Hence, on average, every next transition is characterized by a larger force compared to every previous transition, which is also evident from the plots of  $\bar{f}_{1:n-l+1}$  (Fig. 4 b). The average unfolding forces,  $\bar{f}_{1:n-l+1} = \bar{f}_l$ , for  $(D)_4$  and  $(D)_6$  underestimate the average unfolding force for the single domain  $D$  for all values of  $l$ , except for the last transition, for which, by construction,  $l = n$  and  $\phi_{1:1}(f) = \psi(f)$  (Eq. 3), and hence,  $\bar{f}_{1:1} = \bar{f}$  (Fig. 4 b).

We also compared the pdf of unfolding forces, obtained by combining the first-order statistic pdfs  $\psi_{(D)_n}^{\text{comb}}(f) = (1/n) \sum_{l=1}^n \phi_{1:n-l+1}(f)$  for tandems  $(D)_4$  and  $(D)_6$ , with the (parent) pdf of unfolding forces for the single domain  $D$ ,  $\psi_D(f)$ . We remind the reader that building the histogram of unfolding forces based on the combined set of force data for a tandem  $(D)_n$ —a procedure used to analyze the experimental data—is statistically equivalent to constructing the estimate of the pdf,  $\psi_{(D)_n}^{\text{comb}}(f)$ . We see that  $\psi_{(D)_n}^{\text{comb}}(f)$  for tandems  $(D)_4$  and  $(D)_6$  is markedly different from  $\psi_D(f)$  for the single domain  $D$  (Fig. 4 c), which implies the inequality of the distributions  $\psi_D(f) \neq \psi_{(D)_n}^{\text{comb}}(f)$ . These results validate the use of Order statistics to describe the mechanical unfolding of multimeric proteins. Order statistics captures all the statistical properties of unfolding forces observed for the  $WW$  monomer and oligomers. The curves of  $\psi_{(D)_n}^{\text{comb}}(f)$  are shifted to lower force values and are characterized by a larger width compared with  $\psi_D(f)$  (Fig. 4 c). This is in agreement with the statistical properties of  $\psi_{(WW)_n}^{\text{comb}}(f)$  (Fig. 2, a and b). The average forces,  $\bar{f}_{1:n-l+1}$ ,

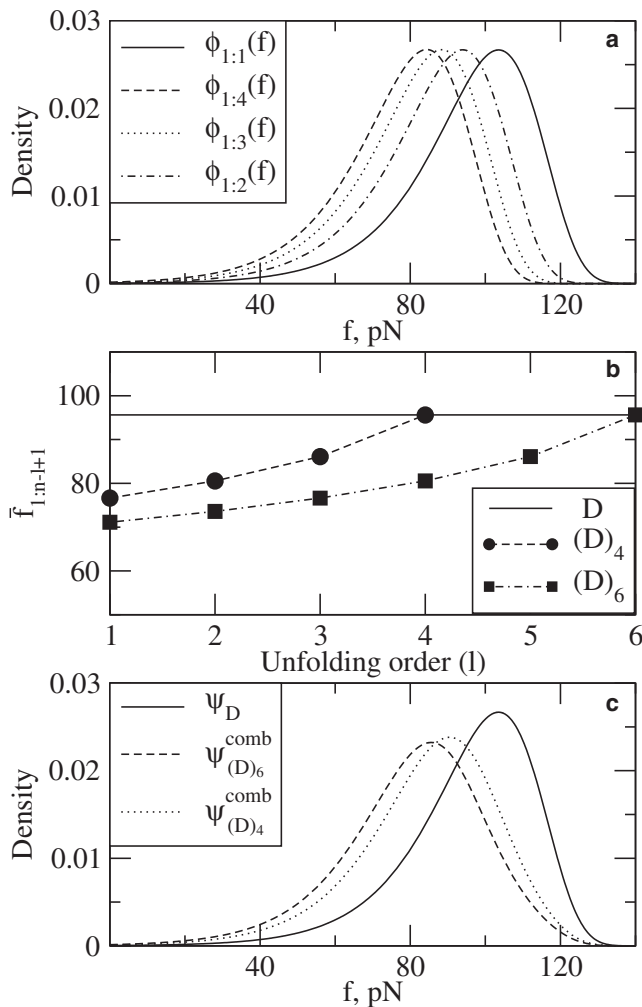


FIGURE 4 (a) Theoretical curves of the first-order statistic pdf of unfolding forces,  $\phi_{1:n-l+1}(f)$  ( $l = 1, 2, \dots, n$ ), describing the forced unfolding transitions in a generic tandem  $(D)_4$ . (b) The average unfolding forces,  $\bar{f}_{1:n-l+1}$ , for generic tandems  $(D)_4$  and  $(D)_6$  (data points) are compared with the average force for a single domain  $D$ ,  $\bar{f}$  (horizontal reference line). (c) Theoretical pdf curves for the combined unfolding forces ( $f_{(D)_n}^{comb} = \{f_1, f_2, \dots, f_N\}$ ),  $\psi_{(D)_n}^{comb}(f)$ , for tandems  $(D)_4$  and  $(D)_6$  are compared with the parent pdf for the monomer  $D$ ,  $\psi_D(f)$ . Model calculations are performed using the force-ramp protocol ( $v_0 = 2.5 \mu\text{m/s}$  and  $\kappa = 10 \text{ pN/nm}$ ), and the Bell model with parameters  $k_0 = 1.0 \text{ s}^{-1}$ ,  $x^\ddagger = 0.3 \text{ nm}$ , and  $k_B T = 4.14 \text{ pN nm}$ .

for  $(D)_4$  and  $(D)_6$  follow a similar dependence on  $l$  (Fig. 3 b) as the average first, second, etc., force,  $\bar{f}_{(WW)_n,l}$ , for  $(WW)_4$  and  $(WW)_6$  (Fig. 2, c and d). We observed a similar agreement between the most probable unfolding forces,  $f_{1:n-l+1}^*$ , for  $(D)_4$  and  $(D)_6$  and  $f_{(WW)_n,l}^*$  for  $(WW)_4$  and  $(WW)_6$  (data not shown).

## RESULTS

### Analysis of unfolding forces for tandems $(WW)_4$ and $(WW)_6$

Calculations of the Order statistics measures for a generic tandem  $(D)_n$  show that using the Bell model results in distri-

butions that are skewed to the left (Fig. 4, a and c). Yet, the force histograms for the  $WW$  monomer and oligomers seem to be more Gaussian-like (Fig. 2, a and b). The force-extension profiles for  $WW$ ,  $(WW)_4$ , and  $(WW)_6$  show that the mechanical tension increases almost linearly with the molecular extension  $X$  (Fig. 3, a and b), which allows us to use a harmonic approximation to model the force data. For this reason, we used the model of a Brownian particle evolving on a harmonic potential (31,32) to describe the time evolution of  $X$  (Appendix 1). In this analytically tractable model, the distribution of  $X$  and, hence, of the unfolding force  $f = k_{sp}X$  are Gaussian ( $k_{sp}$  is the molecular spring constant). In addition, the  $WW$  domain is a small protein, and the entropic effects due to the chain elongation can be safely neglected. In the framework of this model, the (parent) cdf of unfolding forces for a single domain is given by

$$\Psi(f) = \frac{1}{2} \text{Erfc} \left[ \frac{F^* - \bar{F}(f)}{\sqrt{2w(f)}} \right], \quad (5)$$

where  $\text{Erfc}(x)$  is the complementary error function, and  $F^* = k_{sp}X^*$  is the mechanical tension proportional to the chain extension  $X^*$ , which corresponds to the protein in the unfolded state  $U$  (Fig. 1 d). In Eq. 5,

$$\bar{F}(f) = f - \tilde{f} \left( 1 - \exp \left[ -f/\tilde{f} \right] \right)$$

is the average mechanical tension of the chain,

$$w(f) = k_B T k_{sp} \left( 1 - \exp \left[ -2f/\tilde{f} \right] \right)$$

is the width, and  $\tilde{f} = r_f \tau$  is the characteristic force (Appendix 1). The parent pdf of unfolding forces for the single domain is  $\psi(f) = d\Psi(f)/df$ . The Order statistics cdfs and pdfs for the tandems  $(WW)_4$  and  $(WW)_6$ ,  $\Phi_{1:n-l+1}(f)$  and  $\phi_{1:n-l+1}(f)$  ( $l = 1, 2, \dots, n$ ), are then obtained by substituting  $\Psi(f)$  and pdf  $\psi(f)$  for the single domain  $WW$  in the expressions in Eq. 3.

We described the distributions of the first through the fourth unfolding force,  $\psi_{(WW)_4,l}(f)$  ( $l = 1, 2, 3$ , and 4), obtained from the force-extension curves for the tandem  $(WW)_4$  (Fig. 3 a), using Order statistics theory and the Brownian oscillator model. The model has five parameters:  $k_B T$ ,  $D$ ,  $k_{sp}$ ,  $X^*$ , and  $r_f$ . However, in our calculations, we fixed the value of  $k_B T = 4.14 \text{ pN nm}$  (room temperature),  $D = 5.5 \times 10^{-4} \text{ cm}^2/\text{s}$  (diffusion constant for a residue in water at room temperature), and  $r_f = 87.5 \text{ nN/s}$  (force-ramp used in simulations), and we varied the two remaining parameters,  $X^*$  and  $k_{sp}$ . The results of the comparison of  $\psi_{(WW)_4,l}(f)$  with the theoretically derived Order statistics measures,  $\phi_{1:n-l+1}(f)$ , are presented in Fig. 5 a, where we also display the (parent) pdf for the  $WW$  monomer,  $\psi_{WW}(f)$ . We used nonparametric density estimates (33,34) of the pdfs  $\psi_{(WW)_4,l}(f)$  and  $\psi_{WW}(f)$ , which provide a more accurate description of the probability density mass,

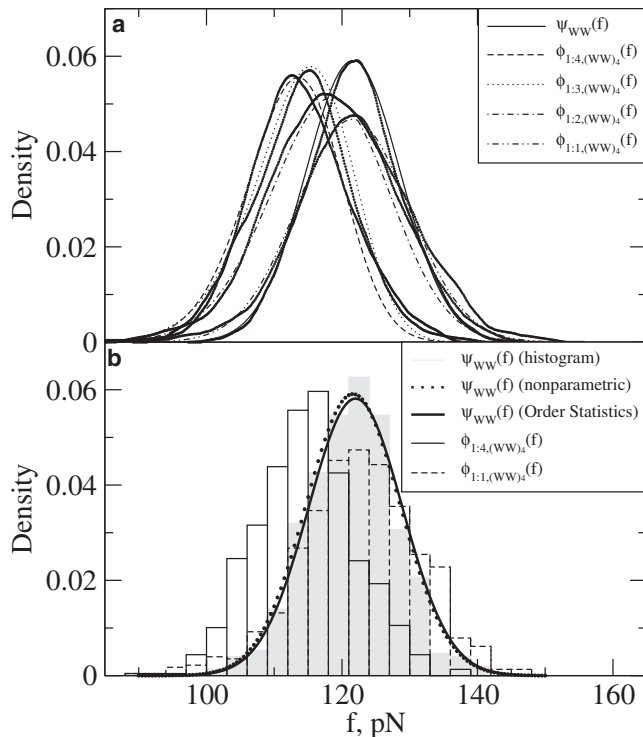


FIGURE 5 (a) Distributions of unfolding forces for the WW monomer and the oligomer  $(WW)_4$ . Nonparametric density estimates of the force pdf for the WW monomer,  $\psi_{WW}(f)$  (data points), and the pdfs for the first through the fourth unfolding forces for  $(WW)_4$ ,  $\psi_{(WW)_4,l}(f)$ ,  $l = 1, 2, 3$ , and 4 (data points), are compared, respectively, with the theoretical pdf curves of unfolding forces for the WW monomer,  $\psi_{WW}(f)$ , and with the theoretical curves of the first-order statistic pdfs of unfolding forces for  $(WW)_4$ ,  $\phi_{1:n-l+1}(f)$ , i.e.,  $\phi_{1:4,(WW)_4}(f)$ ,  $\phi_{1:3,(WW)_4}(f)$ ,  $\phi_{1:2,(WW)_4}(f)$ , and  $\phi_{1:1,(WW)_4}(f)$ . (b) The histogram-based estimate (shaded bars) and the nonparametric density-based estimate (data points) of the pdf of unfolding forces for the WW monomer,  $\psi_{WW}(f)$ , obtained from pulling simulations, are compared with the theoretically derived estimate of the same density, obtained using the Order statistics based inference (solid curve). The histogram-based estimates of the pdf of the first and fourth unfolding forces,  $\phi_{1:4,(WW)_4}(f)$  (open bars) and  $\phi_{1:1,(WW)_4}(f)$  (open bars), are shown for comparison.

especially for the small data samples of 260 points (peak-forces) used in this work. We see that the Order statistics pdfs  $\phi_{1:n-l+1}(f)$  follow closely  $\psi_{(WW)_4,l}(f)$  for all values of  $l = 1, 2, 3$  and 4 (Fig. 4 a). In agreement with  $\psi_{(WW)_4,l}(f)$ , the Order statistics measures  $\phi_{1:n-l+1}(f)$  are characterized by larger values of  $f_{1:n-l+1}^*$  (maxima) for larger  $l$ . We obtained a similar agreement for the tandem  $(WW)_6$  (data not shown). Numerical values of  $X^*/\Delta L$  and  $k_{sp}$ , determined from the fit of  $\phi_{1:n-l+1}(f)$  to  $\psi_{(WW)_4,l}(f)$ , are in the range of  $X^*/\Delta L = 0.57-0.71$  and  $k_{sp} = 14.0-16.3$  pN/nm (Table 1) and, hence, are in good agreement with the values of  $X^*/\Delta L = 0.6 - 0.68$  and  $k_{sp} = 8.0-17.5$  pN/nm, obtained from the structural analysis of the simulation output. We also calculated the average and the most probable forces,  $\bar{f}_{1:n-l+1}$  and  $f_{1:n-l+1}^*$ , as a function of  $l$  for  $(WW)_4$  and  $(WW)_6$ , which describe well  $\bar{f}_{(WW)_4,l}$  and  $f_{(WW)_4,l}^*$  obtained from pulling simulations (Fig. 2, c and d).

TABLE 1 Parameters for the Brownian oscillator model,  $X^*/\Delta L$  and  $k_{sp}$ , for the WW domain

Parameters	$\psi_{WW}(f)$	$\psi_{1,(WW)_4}(f)$	$\psi_{2,(WW)_4}(f)$	$\psi_{3,(WW)_4}(f)$	$\psi_{4,(WW)_4}(f)$
$X^*/\Delta L$	0.71	0.57	0.61	0.58	0.64
$k_{sp}$ , pN/nm	15.5	14.1	14.0	15.1	16.3

These are obtained from the fit of the theoretical pdf curves of unfolding forces for the WW monomer,  $\psi_{WW}(f)$  (parent density), and the first-order statistics pdf curves,  $\phi_{1:n-l+1}(f)$  ( $l = 1, 2, \dots, 4$ ) for the tandem  $(WW)_4$ , to the nonparametric density estimates of the pdfs of unfolding forces for the WW domain,  $\psi_{WW}(f)$  (parent force data), and to the pdfs of unfolding forces for the first through the fourth unfolding transitions for  $(WW)_4$ ,  $\psi_{(WW)_4,l}(f)$  (Fig. 5 a), obtained from pulling simulations (see Fig. 3 a).

## DISCUSSION

### Main results

Dynamic force spectroscopy has evolved over the last two decades into the standard experimental tool for the exploration of the biomechanical unfolding reactions at the single-molecule level. In these experiments, a constant force  $f$  (force-clamp) or a time-dependent force  $f(t) = r_t t$  (force-ramp) is applied to a protein tandem (oligomer)  $(D)_N$  to induce unfolding in protein domains. In our recent articles (27,28), we showed that, because in the force-clamp measurement the applied pulling force (hazard) remains constant, the recorded unfolding times form sets of ordered time variates,  $t_1 = t_{1:N}$ ,  $t_2 = t_{2:N}$ ,  $\dots$ ,  $t_N = t_{N:N}$ , or the Order statistics of unfolding times,  $t_l = t_{l:N}$ , of increasing order  $l = 1, 2, \dots, N$ . We also developed an Order statistics based approach to model the unfolding time data (27,28). In this work, we focused on the unfolding forces analyzed in force-ramp experiments, where each unfolding transition is accompanied by a substantial decrease in the mechanical tension reflected in the sawtoothlike force-extension profiles (Fig. 3). We showed that the peak forces for the first, second,  $\dots$ ,  $N^{\text{th}}$  unfolding transition,  $f_1, f_2, \dots, f_N$ , observed in the force-extension curves, are, indeed, the first-order statistics for decreasing tandem length  $N, N-1, \dots, 1$ , i.e.,  $f_1 = f_{1:N}, f_2 = f_{1:N-1}, \dots, f_N = f_{1:1}$ , or  $f_l = f_{1:N-l+1}$  ( $l = 1, 2, \dots, N$ ).

Excellent agreement between the distributions of unfolding forces, the average and the most probable unfolding forces, obtained from pulling simulations for the WW monomer and oligomers  $(WW)_4$  and  $(WW)_6$ , and the corresponding first-order statistics measures validates the use of Order statistics to describe the biomechanical unfolding reactions in multimeric proteins. This theory captures all the statistical properties of unfolding forces observed for multimeric proteins, and explains why the statistics of forces for a protein tandem differs substantially from the force statistics for a single protein domain. Specifically, in accord with the results of pulling simulations for  $(WW)_4$  and  $(WW)_6$ , the theory predicts correctly that the distributions of unfolding forces,  $\psi_l(f) = \phi_{1:N-l+1}(f)$ , should shift to larger force values (Fig. 4 a), and that the average and the

most probable forces,  $\bar{f}_l = \bar{f}_{1:N-l+1}$  and  $f^*_l = f^*_{1:N-l+1}$ , should increase with increasing unfolding order  $l = 1, 2, \dots, N$  (Fig. 2, *c* and *d*). In agreement with the results of simulations (Fig. 2, *a* and *b*), Order statistics theory predicts that combining all the unfolding forces for a tandem  $(D)_N$  and binning these data into the force histogram results in an imprecise estimate of the distribution of unfolding forces for the protein domain  $D$ ,  $\psi_D(f)$  (Fig. 4 *C*). Hence, Order statistics theory verifies that the combined unfolding forces for a homogeneous tandem  $(D)_N$  and the unfolding forces for a single domain  $D$  are not sampled from the same distribution. Clearly, combining the force data for tandems of different length, does not resolve this problem.

Traditional description of the force-driven unfolding involves pulling together the unfolding time data,  $t_{(D)_N}^{\text{comb}} = \{t_1, t_2, \dots, t_N\}$  (force-clamp), or the force data,  $f_{(D)_N}^{\text{comb}} = \{f_1, f_2, \dots, f_N\}$  (force-ramp), observed for a tandem  $(D)_N$ . These are used to model the parent pdf of unfolding times,  $\psi_D(t)$ , or the pdf of unfolding forces,  $\psi_D(f)$ , for a protein domain  $D$ . In our previous work (27), we showed that the statistical inference of the parent pdf based on analyzing the combined data sample is fully justified in the case of force-clamp measurements of unfolding times. This is because the  $(l + 1)^{\text{st}}$  Order statistic in a tandem of length  $n$  is related to the  $l^{\text{th}}$  Order statistic in a tandem of length  $n - 1$  via a recurrence relation, which for the pdf of unfolding times reads:

$$n\phi_{l:n-1}(t) = (n-1)\phi_{l:n}(t) + l\phi_{l+1,n}(t). \quad (6)$$

Using Eq. 6 recursively, we obtain the parent pdf,  $\psi_D(t)$ , which, by construction, is the pdf of the first unfolding times in the tandem of the unit length,  $\phi_{1:1}(t)$  (17,18),

$$\psi_D(t) \equiv \phi_{1:1}(t) = \frac{1}{n} \sum_{l=1}^n \phi_{l:n}(t). \quad (7)$$

Hence, in the context of force-clamp measurements, Eq. 7 allows one to estimate the parent pdf,  $\psi_D(t)$ , by summing over the Order statistics pdfs,  $\phi_{l:n}(t)$ , of all orders  $1 \leq l \leq n$  (when the unfolding times are i.i.d.). In practice, this is equivalent to combining all the ordered time data  $t_{l:n}$  and binning the data into a single histogram, which is how traditional analyses of unfolding times are carried out (see Fig. 9 in (27)). However, as we showed in this work, in a force-ramp measurement one observes the first-order statistics of unfolding forces,  $f_{1:n-l+1}$ , and, hence,

$$\psi_D(f) \neq \frac{1}{n} \sum_{l=1}^n \phi_{1:n-l+1}(f).$$

With help from Order statistics we now understand why the pdf of unfolding forces for the  $WW$  monomer,  $\psi_{WW}(f)$ , differs from the pdf of combined forces for the oligomers  $(WW)_N$ , i.e.,  $\psi_{(WW)_N}^{\text{comb}}(f) \neq \psi_{WW}(f)$  (Fig. 1, *a* and *b*), which also explains why the procedure of combining the force

data cannot be used. We also understand why the average and the most probable forces,  $\bar{f}_{1:n-l+1} = \bar{f}_l$  and  $f^*_{1:n-l+1} = f^*_l$ , gradually increase with the unfolding order  $l$ .

### Estimation of the parent density $\psi(f)$

The results of application of Order statistics theory to the  $WW$  monomer and oligomer  $(WW)_4$  show that the first-order statistics pdf for the last transition ( $l = 4$ ),  $\phi_{1:n-l+1}(f) = \phi_{1:1}(f)$ , agrees well with the parent density  $\psi_{WW}(f)$  in terms of the average force, i.e.,  $\bar{f}_{1:1} \approx 120.9$  pN versus  $\bar{f}_{WW} \approx 120.6$  pN (Fig. 5 *a*). This is because, by construction,  $\psi_{WW}(f) = \phi_{1:1}(f)$ , and  $\bar{f}_{WW} = \bar{f}_{1:1}$ . Yet,  $\phi_{1:1}(f)$  differs from  $\psi_{WW}(f)$  in the width, quantified by the standard deviation  $\sigma$ , i.e.,  $\sigma_{1:1}(f) \approx 8.5$  pN versus  $\sigma_{WW}(f) \approx 6.8$  pN (Fig. 5 *a*). The width of  $\phi_{1:1}(f)$  is increased because the last unfolding transition in the tandem occurs at higher mechanical tension  $F(f)$  (see Eq. 5), compared to the first transition ( $l = 1$ ), described by  $\phi_{1:n-l+1}(f) = \phi_{1:4}(f)$ . This results in smaller fluctuations in the extension,  $\delta X \sim \sqrt{k_B T / k_{sp}}$ , and, hence, in larger force fluctuations,  $\delta f \sim \sqrt{k_B T k_{sp}}$ , for the last transition ( $l = 4$ ) compared to the first transition ( $l = 1$ ). Because of this, the first-order statistics pdf for the first transition ( $l = 1$ ),  $\phi_{1:4}(f)$ , is closer to the parent density,  $\psi_{WW}(f)$ , in terms of the width ( $\sigma_{1:4}(f) \approx 6.9$  pN), but deviates from  $\psi_{WW}(f)$  in the average value ( $\bar{f}_{1:4} \approx 114.1$  pN), because  $\bar{f}_{WW} \neq \bar{f}_{1:4}$  (Fig. 5 *a*). Hence, in general, the force distributions for the first few (last few) transitions in a tandem are closer to the parent density in terms of the width (average force).

These statistical properties of unfolding forces can be used to resolve the parent density. In the Brownian oscillator model, when  $f \gg \bar{f}$ , the parent pdf is given by the Gaussian density,

$$\psi_D(f) = (2\pi w)^{-1/2} \exp[-(f - F^*)^2 / 2w],$$

where  $w = k_B T k_{sp}$  and  $F^* = k_{sp} X^*$ . Hence, by setting  $w = \sigma_{1:n}^2(f)$  and  $F^* = \bar{f}_{1:1}$  in the above expression for  $\psi_D(f)$ , where  $\sigma_{1:n}(f)$  and  $\bar{f}_{1:1}$  are, respectively, the standard deviation for the first unfolding force ( $f_{1:n} = f_1$ ) and the average  $n^{\text{th}}$  unfolding force ( $f_{1:1} = f_n$ ) for a tandem  $(D)_n$ , one can estimate  $\psi_D(f)$ . We utilized this approach to calculate the parent pdf for the  $WW$  domain,  $\psi_{WW}$ , using  $\sigma_{1:4}(f) = 6.9$  pN and  $\bar{f}_{1:1} = 120.9$  pN obtained for the tandem  $(WW)_4$ . The results obtained show that the Order statistics based estimate, computed theoretically, compares well with the nonparametric-density estimate and with the histogram-based estimates of  $\psi_{WW}(f)$  (Fig. 5 *b*). The model-free inference of the parent cdf,  $\Psi_D(t)$ , from the first-order statistics measures for the tandem  $(D)_n$ ,  $\Phi_{1:n-l+1}(f)$ , can be based on Eq. 3 for the cdf, which can be written as

$$\Psi(f) = 1 - (1 - \Phi_{1:n-l+1}(f))^{1/(n-l+1)}.$$



This equation can be used for each cdf  $\Phi_{1:n-l+1}(f)$  ( $l = 1, 2, \dots, n$ ), and for each tandem length  $n$ .

## CONCLUSIONS

We developed a new, to our knowledge, theoretical framework, inspired by Order statistics, to describe the biomechanical unfolding reactions observed in dynamic force-ramp experiments on wild-type multimeric proteins and engineered polyproteins. We also showed that the Brownian oscillator model can be used to model the forced unfolding transitions in small proteins such as the *WW* domain. There is a long-standing question whether the mechanical property of a multimeric protein can be viewed as the sum of the mechanical properties of individual protein domains. Essentially, this is a restatement of the central limit theorem applied to multimeric proteins. Our results show that, for the i.i.d. unfolding times, observed for a homogeneous tandem of noninteracting identical protein domains under the influence of the constant force  $f$ , the answer is in the affirmative, because  $\psi_D(t) = (1/N)\sum_{l=1}^N \phi_{l:N}(t)$  (Eq. 7) and, hence, the average unfolding time is  $\bar{t}_D = (1/N)\sum_{l=1}^N \bar{t}_{l:N}$  from Bura et al. (27). However, the mechanical property of a multimeric protein is not the sum of the mechanical properties of constituting domains when it is subject to the time-dependent force  $f(t)$ , because the distributions of unfolding forces for the consecutive first, second, ...,  $N^{\text{th}}$  unfolding transitions,  $\phi_{1:N-l+1}(f)$ , do not sum up to the parent density  $\psi_D(f)$ , i.e.,  $\bar{f}_D \neq (1/N)\sum_{l=1}^N \bar{f}_{1:N-l+1}$ . Hence, the statistical properties of the important molecular metrics quantifying the mechanical response of multimeric proteins and polyprotein chains depend, in part, on the nature of mechanical perturbation. The Order statistics based inference enables one to obtain accurate estimates of the parent density using the force data for protein tandems of arbitrary length.

## APPENDIX 1: BROWNIAN OSCILLATOR MODEL

We describe the force-induced elongation of a protein chain  $X(t)$  using the one-dimensional Brownian motion of a particle in a harmonic potential  $U(X) = k_{\text{sp}}X^2/2$  (31,32), where  $k_{\text{sp}}$  is the molecular spring constant. The particle is subject to the time-dependent pulling force  $f(t) = r_f t$  and the Gaussian random force  $g(t)$  with zero mean,  $\langle g(t) \rangle = 0$ , and two-point correlation function,  $\langle g(t)g(0) \rangle = 2D\delta(t)$  ( $D$  is the diffusion constant). Within the framework of the Brownian oscillator model, the conditional probability (Green function) for extending the molecule from  $X'$  to  $X$  over time  $t$  is given by

$$G(X, X', t) = \left( \frac{1}{2\pi w(t)} \right)^{1/2} \exp \left[ -\frac{(X - \bar{X}(t))^2}{2w(t)} \right] \quad (8)$$

where

$$\bar{X}(t) = X' \exp[-t/\tau] + (r_f/k_{\text{sp}})(t - \tau(1 - \exp[-t/\tau]))$$

is the average extension,  $\tau = \eta/k_{\text{sp}}$  is the characteristic timescale, and

$$w(t) = (k_{\text{B}}T/k_{\text{sp}})(1 - \exp[-2t/\tau])$$

is the width. We assume that the unfolding transition occurs when  $X(t)$  reaches the critical extension  $X^*$ , at time  $t$ . The cumulative distribution function (cdf) or the unfolding probability is given by

$$\Psi(t) = \int_{X^*}^{\infty} dX \int_0^{\infty} dX' G(X, X', t) P_{\text{eq}}(X'), \quad (9)$$

where  $P_{\text{eq}}(X)$  is the initial distribution of  $X$  in the folded state. Assuming that  $X$  is sharply peaked around the equilibrium extension  $X = 0$ , i.e.,  $P_{\text{eq}}(X) = \delta(X)$ , which is the case for the *WW* domain, performing the integration over  $X$  and  $X'$ , and switching from the unfolding times  $t$  to the unfolding forces  $f = r_f t$ , we obtain the parent cdf for the single domain,  $\Psi(f)$  (Eq. 5).

We thank Dr. Efstathia Bura (George Washington University) for her help with nonparametric density estimation of the force data.

This work was supported by the startup fund from the University of Massachusetts at Lowell (to V.B.). This project was supported in part by National Science Foundation grant No. MCB-0845002 (to R.I.D.).

## REFERENCES

1. Stossel, T. P., J. Condeelis, ..., S. S. Shapiro. 2001. Filamins as integrators of cell mechanics and signaling. *Nat. Rev. Mol. Cell Biol.* 2: 138–145.
2. McEver, R. P. 2002. Selectins: lectins that initiate cell adhesion under flow. *Curr. Opin. Cell Biol.* 14:581–586.
3. Barsegov, V., and D. Thirumalai. 2005. Dynamics of unbinding of cell adhesion molecules: transition from catch to slip bonds. *Proc. Natl. Acad. Sci. USA.* 102:1835–1839.
4. Linke, W. A., M. Ivemeyer, ..., S. Labeit. 1996. Towards a molecular understanding of the elasticity of titin. *J. Mol. Biol.* 261:62–71.
5. Weisel, J. W. 2008. Biophysics. Enigmas of blood clot elasticity. *Science.* 320:456–457.
6. Feng, Y., and C. A. Walsh. 2004. The many faces of filamin: a versatile molecular scaffold for cell motility and signaling. *Nat. Cell Biol.* 6:1034–1038.
7. Schwaiger, I., A. Kardinal, ..., M. Rief. 2004. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat. Struct. Mol. Biol.* 11:81–85.
8. Zinober, R. C., D. J. Brockwell, ..., D. A. Smith. 2002. Mechanically unfolding proteins: the effect of unfolding history and the supramolecular scaffold. *Protein Sci.* 11:2759–2765.
9. Rief, M., M. Gautel, ..., H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science.* 276:1109–1112.
10. Labeit, S., M. Gautel, ..., J. Trinick. 1992. Towards a molecular understanding of titin. *EMBO J.* 11:1711–1716.
11. Brown, A. E. X., R. I. Litvinov, ..., J. W. Weisel. 2009. Multiscale mechanics of fibrin polymer: gel stretching with protein unfolding and loss of water. *Science.* 325:741–744.
12. Brown, A. E. X., R. I. Litvinov, ..., J. W. Weisel. 2007. Forced unfolding of coiled-coils in fibrinogen by single-molecule AFM. *Biophys. J.* 92:L39–L41.
13. Li, H. B., A. F. Oberhauser, ..., J. M. Fernandez. 2000. Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl. Acad. Sci. USA.* 97:6527–6531.
14. Brujic, J., R. I. Z. Hermans, ..., J. M. Fernandez. 2006. Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nat. Phys.* 2:282–286.

15. Clemen, A. E.-M., M. Vilfan, ..., M. Rief. 2005. Force-dependent stepping kinetics of myosin-V. *Biophys. J.* 88:4402–4410.
16. Mickler, M., R. I. Dima, ..., M. Rief. 2007. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc. Natl. Acad. Sci. USA.* 104:20268–20273.
17. David, H. A., and H. N. Nagaraja. 2003. *Order Statistics*. Wiley Interscience, New York.
18. Gumbel, E. J. 2004. *Statistics of Extremes*. Dover Publications, New York.
19. Hyeon, C., R. I. Dima, and D. Thirumalai. 2006. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure.* 14:1633–1645.
20. Koepf, E. K., H. M. Petrassi, ..., J. W. Kelly. 1999. WW: an isolated three-stranded antiparallel  $\beta$ -sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* 8: 841–853.
21. Jäger, M., H. Nguyen, ..., M. Gruebele. 2001. The folding mechanism of a  $\beta$ -sheet: the WW domain. *J. Mol. Biol.* 311:373–393.
22. Ferguson, N., J. Berriman, ..., A. R. Fersht. 2003. Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc. Natl. Acad. Sci. USA.* 100:9814–9819.
23. Karanicolas, J., and C. L. I. Brooks, 3rd. 2003. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design? *Proc. Natl. Acad. Sci. USA.* 100:3954–3959.
24. Cheung, M. S., D. Klimov, and D. Thirumalai. 2005. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl. Acad. Sci. USA.* 102:4753–4758.
25. Dima, R. I., and H. Joshi. 2008. Probing the origin of tubulin rigidity with molecular simulations. *Proc. Natl. Acad. Sci. USA.* 105:15743–15748.
26. Veitschans, T., D. K. Klimov, and D. Thirumalai. 1997. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* 2:1–22.
27. Bura, E., D. K. Klimov, and V. Barsegov. 2007. Analyzing forced unfolding of protein tandems by ordered variates. 1. Independent unfolding times. *Biophys. J.* 93:1100–1115.
28. Bura, E., D. K. Klimov, and V. Barsegov. 2008. Analyzing forced unfolding of protein tandems by ordered variates. 2. Dependent unfolding times. *Biophys. J.* 94:2516–2528.
29. Bell, G. I. 1978. Models for the specific adhesion of cells to cells. *Science.* 200:618–627.
30. Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. *Biophys. J.* 72:1541–1555.
31. Doi, M., and S. F. Edwards. 1994. *The Theory of Polymer Dynamics*. Oxford University, New York.
32. Barsegov, V., G. Morrison, and D. Thirumalai. 2008. Role of internal chain dynamics on the rupture kinetic of adhesive contacts. *Phys. Rev. Lett.* 100:248102–248105.
33. Scott, D. W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York.
34. Bura, E., A. Zhmurov, and V. Barsegov. 2009. Nonparametric density estimation and optimal bandwidth selection for protein unfolding and unbinding data. *J. Chem. Phys.* 130:015102–015116.