PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Theoretical tests of the mechanical protection strategy in protein nanomechanics

Mateusz Chwastyk,[1] Albert Galera–Prat,[2] Mateusz Sikora,[1,3] Àngel Gómez-Sicilia,[2] Mariano Carrión-Vázquez,[2] and Marek Cieplak[1]*

[1] Laboratory of Biological Physics, Institute of Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warsaw, Poland

[2] Instituto Cajal, Consejo Superior de Investigaciones Cientificas (CSIC), IMDEA Nanociencias and CIBERNED, Av. Doctor Arce, 37, 28002 Madrid, Spain

[3] Institute of Science and Technology Austria, Klosterneuburg, Austria

ABSTRACT

We provide theoretical tests of a novel experimental technique to determine mechanostability of proteins based on stretching a mechanically protected protein by single-molecule force spectroscopy. This technique involves stretching a homogeneous or heterogeneous chain of reference proteins (single-molecule markers) in which one of them acts as host to the guest protein under study. The guest protein is grafted into the host through genetic engineering. It is expected that unraveling of the host precedes the unraveling of the guest removing ambiguities in the reading of the force-extension patterns of the guest protein. We study examples of such systems within a coarse-grained structure-based model. We consider systems with various ratios of mechanostability for the host and guest molecules and compare them to experimental results involving cohesin I as the guest molecule. For a comparison, we also study the force-displacement patterns in proteins that are linked in a serial fashion. We find that the mechanostability of the guest is similar to that of the isolated or serially linked protein. We also demonstrate that the ideal configuration of this strategy would be one in which the host is much more mechanostable than the single-molecule markers. We finally show that it is troublesome to use the highly stable cystine knot proteins as a host to graft a guest in stretching studies because this would involve a cleaving procedure.

Proteins 2014; 82:717–726.

Key words: single molecule force spectroscopy; pFS vectors; stretching of proteins; coarse grained models; cystine slipknots.

INTRODUCTION

A common way to study mechanical stability of proteins is through single-molecule force spectroscopy (SMFS) by stretching the protein under study serially connected to either repeats of itself or repeats of another protein, which acts as a single-molecule marker (see, e.g., Ref. 1–7). The interpretation of the resultant pattern usually involves an assumption about the serial (as opposed to simultaneous) character of unraveling, which is not obvious since the degree of seriality may depend on the temperature (T) (on raising the T, thermal fluctuations become more and more comparable to the effects of forces due to the potentials and make the modules unravel in an increasingly simultaneous manner)[8]. In addition, an arrangement of domains that are repeated in tandem undergoes partial unraveling in all domains before showing single-domain events even if the domains are identical. This may result in $F - d$ patterns that are not simple superpositions of the isolated single-domain patterns.[9]

Recently, a novel approach in SMFS has been established.[10,11] It still involves the stretching of a number of (like or unlike) serially connected proteins; but the

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: European Regional Development Fund under the Operational Programme Innovative Economy NanoFun POIG.02.02.00-00-025/09 (Computer Resources); Grant sponsor: Polish National Science Centre; Grant number: 2011/01/B/ST3/02190 (MC) and 2011/01/N/ST3/02475 (MS); Grant sponsor: MICINN; Grant number: BIO2010-22275; Grant sponsor: Comunidad de Madrid; Grant number: S-2009MAT-1507 (MCV); Grant sponsor: ERA-IB FIBERFUEL project.

*Correspondence to: Marek Cieplak, Institute of Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warsaw, Poland. E-mail: mc@ifpan.edu.pl

Received 6 June 2013; Revised 27 August 2013; Accepted 26 September 2013
Published online 19 February 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24436

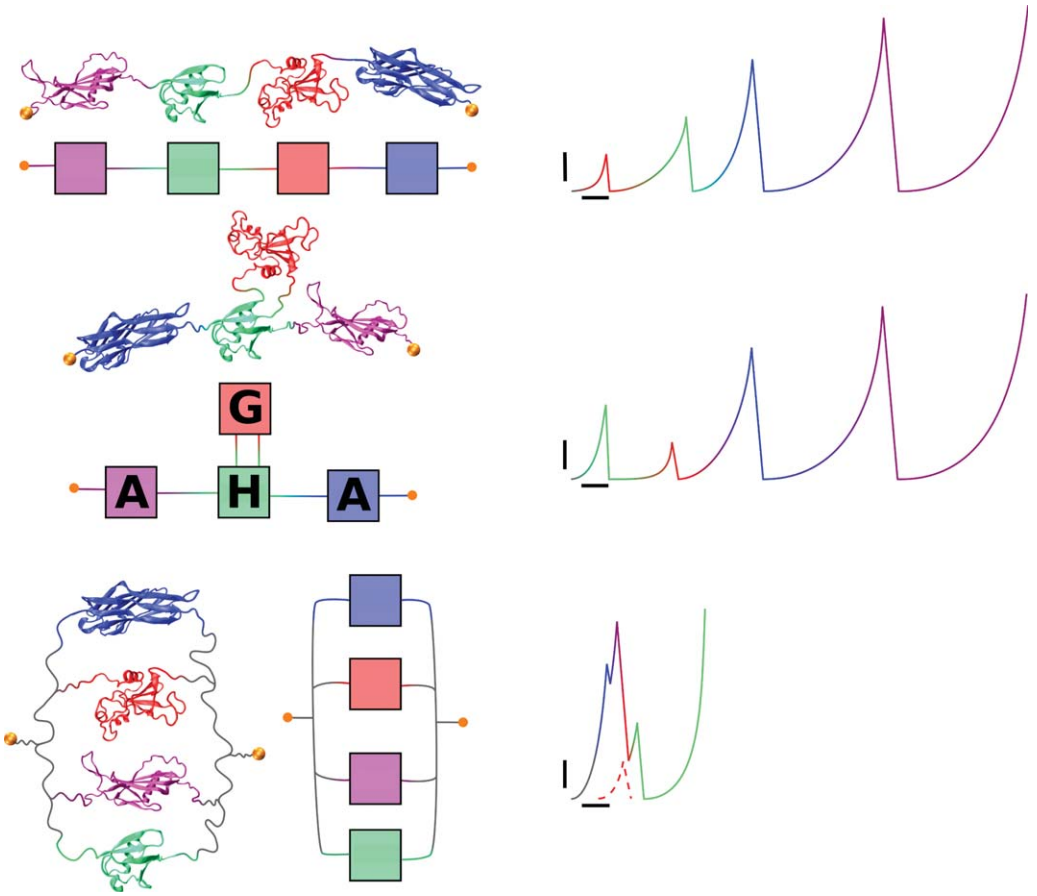

Figure 1

Polyprotein configurations in SMFS: the standard series (the top left panel) and the "host-guest" configurations (the middle left panel) are both different from a hypothetical "parallel" configuration (the bottom left panel). The corresponding force-distance curves are shown in the right panels. The four proteins used and their contributions to the forces are shown in different colors: 1BRN, red; 1UBQ, green; 1G1K, blue; and 1C4P, purple. The vertical and horizontal scale bars in the right hand panels are 100 pN and 10 nm, respectively. In the host-guest configuration, H, G, and A mark the host, guest, and single-molecule markers, respectively. The figure was drawn using the VMD software.[12]

protein under study, referred here as guest (G), is grafted and mechanically protected into a host (H) protein instead of being linked to it in series. For short, we shall call this as an HG connectivity. The grafting is accomplished through cloning of a guest DNA sequence into the host coding sequence and subsequent expression of the fusion protein. When modeling, the grafting is equivalent to cutting one of the peptide bonds in the known host protein and reconnecting the ends through the guest protein. In this article, we provide studies, mostly theoretical, which explore whether the proper workings of this new method involve some conditions that need to be met and whether single-domain patterns are indeed reproduced even though the system studied involves a complex topology. Thus, our question is whether the $F-d$ pattern of an isolated protein is the same as that inside a carrier protein?

The differences between various possibilities to connect proteins are illustrated in Figure 1 (made using VMD software)[12] for four proteins with the Protein Data Bank (PDB)[13] structure codes: 1BNR, 1UBQ, 1G1K, and 1C4P, which correspond to barnase, ubiquitin, cohesin I module of scaffoldin, and the β domain of streptokinase, respectively. In the top panel, they are connected in series. In the bottom panel, all the four proteins are connected in parallel—an arrangement that is purely hypothetical as it has not been carried out experimentally (its engineering would involve disulphide bonding). In the middle panel, 1C4P, 1UBQ, and 1G1K are connected in series but 1UBQ acts like a host to 1BNR—a guest protein. Prima facie one may think that G is in parallel to H, but it is actually grafted inside the host (see Fig. 1). The expected force–extension curves are shown in the right hand panels of Figure 1. In the serial connection, the proteins unfold stochastically and the process tends to start from the module of the lowest mechanical stability and finish on the most mechanostable one. In the HG connection, G is isolated from the tension arising in

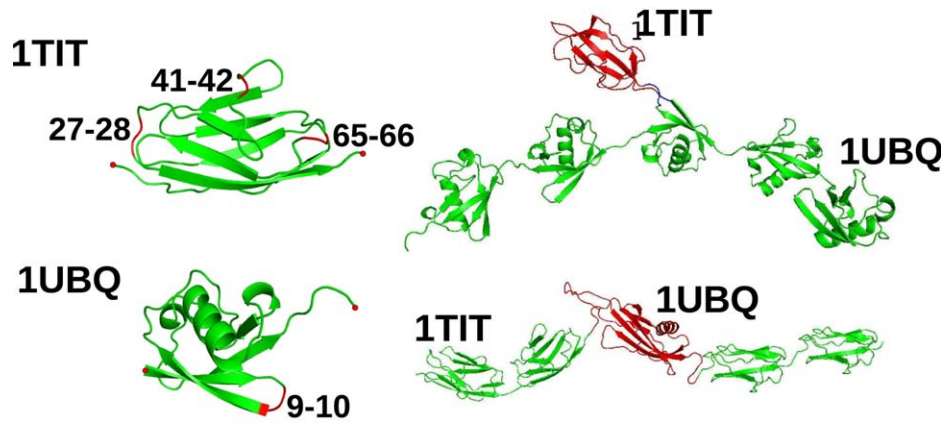

Figure 2

Top left, A cartoon representation of one domain of 1TIT with the indicated places of insertion of a guest protein. The places that are considered in this article are between sites 27 and 28, 41 and 42, and 65 and 66. Bottom left, A similar illustration for ubiquitin. In this case, we consider just one possibility: between sites 9 and 10. Top right: Schematic representation of 1TIT grafted into the third (central) module of a chain of five 1UBQ domains. The cut in this module is between sites 9 and 10. Bottom right: A serial connection of the central ubiquitin to two domains of 1TIT on each side. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

H (i.e., is mechanically protected) until H unfolds. In the parallel connection, nanomechanics is very difficult to study, because unfolding does not depend so much on the force as it does on the size of the molecule: the module with a shorter linker will be the first to unravel. However, the significant interferences between various modules would make the assignment of the force patterns to the structural events complex. These interferences may hide a peak originating from a low-force module, as illustrated for 1BNR in the right bottom panel (represented by a red dashed line). Eventually, when all the modules are unraveled, the further extension will be governed by the shortest module.

A more detailed view of an HG construct is shown in Figure 2 (top right panel) where the middle domain in a chain of five ubiquitins is fused to an I27 domain of titin (1TIT). The bottom right panel in this figure shows the traditional serial linkage of one ubiquitin and four I27s. The expected advantage of the HG connectivity is that unraveling of the host protein should signal the onset of events related to unfolding of the guest molecule and thus make the guest pattern identification process unequivocal.[10]

In this article, we test the HG approach theoretically by grafting several guest proteins of known $F - d$ patterns into linkages of proteins whose $F - d$ patterns are also known, at least theoretically. We stretch the constructed arrangements at constant speed by using a coarse-grained structure-based model described in Ref. 8,14–16 and analyze the resulting unfolding trajectories in terms of how they agree with the pattern of the isolated guest molecule. We consider homogeneous and heterogeneous strings with a grafted module and discuss situations in which the guest and host molecules have mechanostabilities at various ratios. The mechanostability is defined here as the height of the maximum isolated force peak, F_{max}.

Our main finding is that the new method works well when F_{max} of the host molecule is substantially higher than that of the marker molecule as, in this case, the signal from the guest follows immediately after the signal from the host instead of being preceded by a signal from the unraveling linkers mixed with that due to the host. However, there is one more condition that had better be satisfied: The results are more reliable if the host molecule is endowed with a single-force peak. If several force peaks are involved, then peaks coming from G and H partial unfoldings can be interspersed, which may make the reading of the guest $F - d$ signals uncertain. Still the ambiguities are significantly reduced when compared with a serial linking of the guest molecule. Independent of the ability to identify the signal properly, there is still an issue of whether the measured value of F_{max} of the grafted host protein is the same as when in isolation. In many cases, we find that it is about the same within the range of thermal fluctuations. Finally, there could be differences between various ways of selecting sites in the host molecule for the introduction of the guest. Three choices on the accessible surface of I27 and one on the surface of ubiquitin are indicated in the left panels of Figure 2. These cuts do not affect any secondary structures (i.e., are located in loops). Our tests (shown in Supporting Information) do not indicate any noticeable change of the results related to making the choice in the specific cut in I27, but it is likely that some wrong cuts can be found. Overall, the HG approach to measurement of mechanostability is a big improvement over the serial approach. Nevertheless, it is still not free of some ambiguities, which should be removed if the host is much more mechanostable than the single-molecule markers.

THE MODEL AND THE EXPERIMENTAL METHODS

The coarse-grained model

Our starting point is the generation of the native structures. The basic protein structures are obtained from the PDB (the first chain was taken when there were several available). The constructs with grafted proteins are obtained by using the "sculpting tool" of PyMol software.[17] This is a rather crude method of reorganizing the backbone as no energy minimization is involved. However, it assures stereochemical correctness and reconstructs proper topologies. Its usage can be justified by the coarse-grained character of the model. Once the cut in the host protein is made (e.g., between sites 9 and 10 in ubiquitin and 27 and 28 in I27), its conformation near the cut sites is adjusted carefully to allow for making a connection to the termini of the guest protein. In the case of ubiquitin acting as host, its termini were extended by polyalanine linkers; for otherwise, a strong distortion in the original ubiquitin structure would be produced. Examples of the resulting host–guest constructs are shown in Figure 3. The flanking proteins are attached, also using PyMol, to the ready host–guest constructs. It should be noted that, in our model, polyalanine is just a string of beads. In reality, however, it may form secondary structures: either helical or coiled and β structures, depending on the precise properties of the solvent and the length of the peptide.[18–20] For our purposes, these structures are of no consequence as their rupture involves small forces.

The dynamics is implemented within the coarse-grained approach[8,14–16] in which the degrees of freedom are associated with the C^{α} atoms. Their interactions are governed by the contact map as determined through atomic overlaps.[21] The native contacts between C^{α} atoms i and j at distance r_{ij} are described by the Lennard–Jones potential $V(r)=4\epsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$, where σ_{ij} is determined for each pair ij, so that the potential minimum coincides with the native distance. Our estimate of the binding energy parameter $\epsilon \sim 110\,\mathrm{pN/A}$ is based on comparison to experiments on protein stretching.[22] This energy corresponds to about 800 K multiplied by the Boltzmann constant k_B, so the room temperature is closer to $0.35\,\epsilon/k_B$ than $0.3\,\epsilon/k_B$ at which most of our simulations have been performed. The slightly lower temperature usually leads to best folding. Thermostating is implemented by using Langevin noise and damping terms.[23] Nonnative contacts are purely repulsive and are given by the truncated Lennard–Jones potential that is cut at the minimum of 4 Å. Covalent couplings such as those along the backbones are described with a harmonic potential with spring constant $50\,\epsilon/\mathrm{A}^2$. Stretching is implemented by attaching termini of the whole structures to two springs. One of them is anchored and another moving with a speed of 0.005 Å/τ, where τ is of order 1 ns. This time scale takes into account diffusion through an implicit solvent. A native contact is considered broken if r_{ij} exceeds $1.5\,\sigma_{ij}$. Identification of the domain that is being unraveled at a displacement d goes through monitoring of the contacts that get ruptured and plotting the so-called scenario diagrams[8,15] (data not shown here). When making the HG constructs, one has to check that no contact interactions between H and G are introduced. If they are, one can remedy the situation by introducing poly-alanine linkers as illustrated in Supporting Information (see Supporting Information Fig.19).

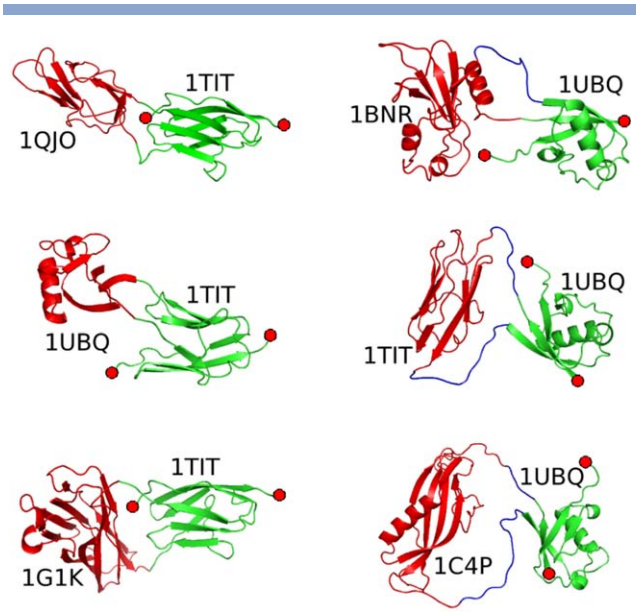

Figure 3

Illustration of how guest proteins are connected to the central host domains. In the left panels, the host molecule is 1TIT and the insertion is between sites 27 and 28. In the right panels, the host molecule is 1UBQ and the insertion is between sites 9 and 10. Polyalanine linkers (shown in blue in the right panels) are used to facilitate the connection. In the case of barnase (1BRN), one linker is sufficient. The names of the guest proteins are indicated on the left of each panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Experimental methodology

We focus on protein 1G1K that has not yet been studied by using the novel HG linkage. The coding sequence for 1G1K was cloned between AgeI/SmaI sites in pFS-2 vector as described[10] using c1C plasmid as a template.[24] Sequence was verified by sequencing and *Escherichia coli* C41 cells were used for protein expression. The protein was purified by Ni^{2+}-affinity, followed by size-exclusion and ionic exchange chromatographies. Finally, clean fractions were concentrated in phosphate-buffered saline/0.2 m*M* ethylenediaminetetraacetic acid/5 m*M* dithiothreitol buffer, aliquoted and stored at −80°C. AFM experiments

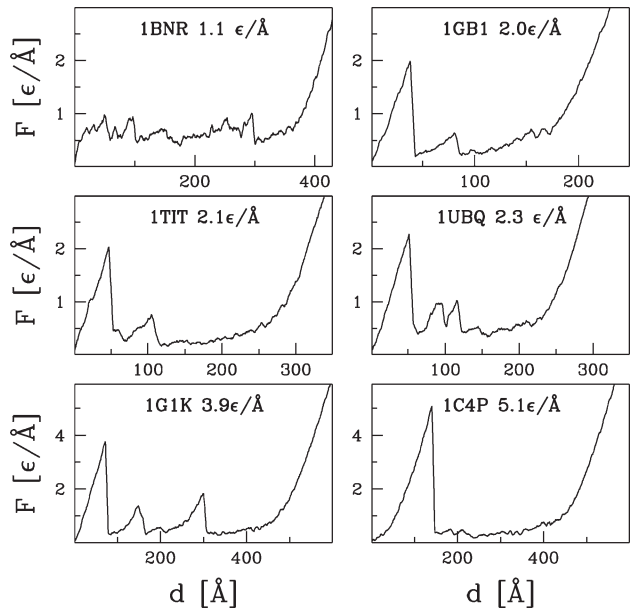

Figure 4
Isolated molecule $F - d$ curves of the proteins studied in this article.

were carried out in a home-made AFM described elsewhere.[24] A drop of 10–20 μL of protein sample was incubated on NTA-Ni^{2+} cover slips for 15 min before SMFS analysis using Biolever (Olympus) cantilevers at a pulling speed of 400 nm/s. Data were collected and analyzed with IGOR PRO (Wavemetrics) and WSxM (Nanotec)[25]. In addition to standard criteria, only those traces showing a peak corresponding to the host unfolding were included in the analysis.

RESULTS AND DISCUSSION

Single proteins

We find that the results of our stretching simulations are governed primarily by the mechanostability properties of individual proteins involved, such as the values of F_{max} and the number of force peaks that are present in the $F - d$ curves. We start by presenting the $F - d$ curves for six single proteins (Fig. 4): barnase – 1BNR,[26] binding domain of streptococcal protein G – 1GB1,[27] I27 domain of titin – 1TIT,[28] ubiquitin – 1UBQ,[29] module of cohesin I – 1G1K,[30] and streptokinase β-domain – 1C4P.[31] For brevity, the PDB structure codes will be used to refer to these proteins from now on.

The proteins are arranged according to their values of F_{max}, as listed in the panels: 1BNR has the smallest mechanostability of the six (1.1 ϵ/Å, i.e., ∼120 pN), while 1C4P has the largest (5.1 ϵ/Å, i.e., ∼560 pN). Under physiological conditions, 1C4P is actually a tetramer, but for simplicity, we envision dealing with an isolated single chain. This protein has been predicted[22] to have the highest value of F_{max} among the 17,134 single chain proteins considered if one puts aside systems in which resistance to stretching is provided by the cystine slipknot mechanism instead of the common mechanism that involves shearing between β-strands. Unlike 1C4P, the second strongest protein, 1G1K, is endowed with a multipeaked $F - d$ pattern indicating that several different regions of the proteins get ruptured independently at various stages of pulling. The predicted value[22] of F_{max} of ∼430 pN is close to 425 pN obtained experimentally.[24] There is some discrepancy for 1BNR—the reported experimental value of F_{max} is about 70 pN[32] and the predicted one is ∼120 pN. However, the agreement with experiment for the remaining proteins 1GB1, 1TIT, and 1UBQ is significantly better with the experimental values of F_{max} being 190 pN,[33] 204 pN,[34] and 203 pN,[35] respectively. It should be noted that any discrepancies between the theoretical and experimental values of F_{max} are not relevant for our purpose because our approach involves combining proteins of various strengths within the same theoretical model and comparing the resulting $F - d$ curves. To make the comparisons, we should take into account error bars that are thermal in nature. They are of order 0.1 ϵ/Å, which is derived from comparing several trajectories.

Homogeneous host chain

We first consider chains of five identical domains. The central domain is connected to two domains on each side; they are numbered 1, 2, 4, and 5. We have found that the number of the flanking proteins affects the overall $F - d$ pattern, but the value of F_{max} associated with the G proteins does not depend on the number of the flanking modules (within the thermal error bars). The central (third) domain is also "cut open" to connect to the guest protein. This central domain then acts as a host to G. Such structures will be referred to by using the shorthand notation:

$$2 * A - H(i - j, G) - 2 * A \ , \qquad (1)$$

where i and j indicate the sites in H to which the termini of G are connected (with or without linkers) and j is just equal to $i + 1$ (though more general schemes can be envisioned). For homogeneous chains, H is identical to A. Otherwise the chain is referred to as being heterogeneous. We consider A to be either 1UBQ or 1TIT. For a comparison, we also stretch proteins connected in series, as described by

$$2 * A - H - G - 2 * A \qquad (2)$$

if H is distinct from G, or $2 * A - G - 2 * A$ otherwise. In each case, we generate five distinct trajectories. We illustrate the findings by showing four $F - d$ curves for the novel HG connectivity and two for the serial

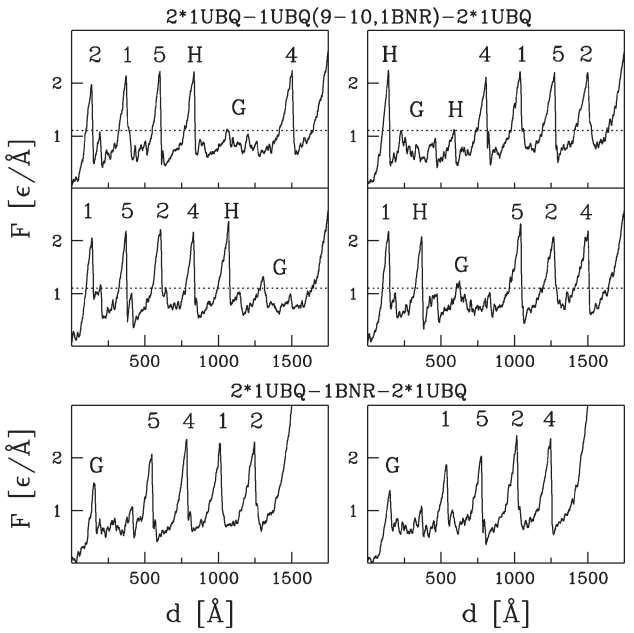

Figure 5

The case of G weaker than H=A. The top four panels show examples of the $F - d$ trajectories for the chain of 1UBQ in which the central (third) module, denoted as H, acts as host to 1BRN, denoted as G. Thus, G=1BRN and H=A=1UBQ. The dotted line indicates the value of F_{max} for an isolated 1BRN. G is inserted between sites 9 and 10 in H.

connectivity. In the main text, we illustrate our findings by discussing four systems, while results for all other systems analyzed are shown in the Supporting Information. In the following, the terms "weak," "strong," and so on refer to the values of F_{max} in isolated proteins.

G of the same strength as A=H

We first consider the situation in which A=H=G (the sign of equality means identity of the protein). The $F - d$ curves are shown in Supporting Information: Supporting Information Figure 1 for 1UBQ and Supporting Information Figure 2 for 1TIT. We observe that the modules unfold in a trajectory-dependent order as the symmetry between individual domains is broken by thermal fluctuations. The random character of unfolding also applies to the serial connectivity. In the HG connectivity, G need not to unravel just past unraveling of H, though in most cases it does. In the serial case, all of the force peaks are essentially the same in height, and the heights agree with the single module value. In the HG case, the values of F_{max} associated with G may depart from the single module case in a stronger way.

G weaker than A=H

Here, we consider A=H=1UBQ and G=1BNR. In the serial connection, G is expected to unravel first as it constitutes the weakest module. This is indeed so as evidenced by the two bottom panels of Figure 5. The remaining domains unfold in an arbitrary order. For the HG connection, G always follows H which means that a rupture of H signals the onset of the unfolding of G. However, we observe one case (the top right panel) in which G unravels after only a partial unfolding of H. Once G gets unraveled, unfolding of H is resumed. The value of F_{max} remains equal, within the error bars related to thermal fluctuations, to the single module case as evidenced by Table 1 in Supporting Information for 40 situations with two strong host molecules 1C4P and 1G1K. On analysing results presented Table 1, we also conclude the structural modifications in the H and G molecules that are necessary to implement the grafting do not affect the values of F_{max} when compared to the values for the unmodified proteins—all differences are smaller than the standard deviations between five trajectories. This indicates that the PyMol-based sculpting tool we use is adequate.

G stronger than A=H and single-peaked

Figure 6 is for A=H=1UBQ and G=1C4P. In contrast to the situation discussed just above, unraveling of G comes last when the modules are in series. Furthermore, G also unravels the last when the connectivity is of the HG type. We have not observed a trajectory in which G would unravel immediately after H. In spite of this, F_{max} for G in the HG strategy is essentially the same as when G is in isolation.

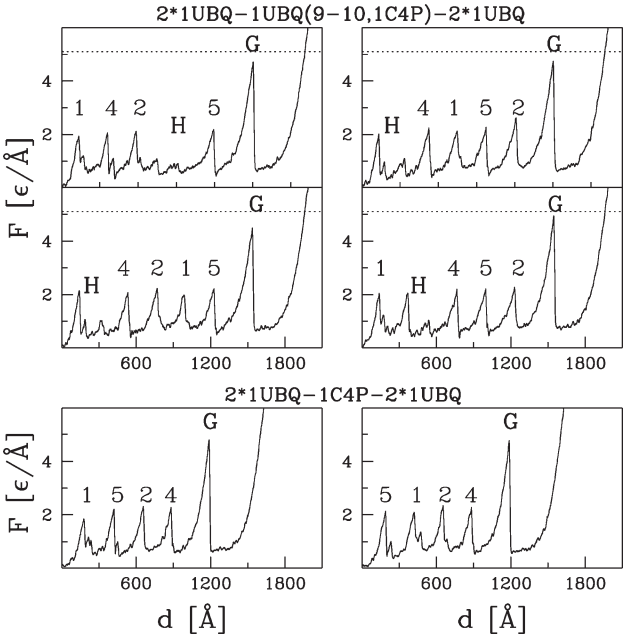

Figure 6

The case of G stronger than H=A. Similar to Figure 5 but with 1C4P replacing 1UBQ as a guest molecule, that is, G=1C4P and H=A=1UBQ.

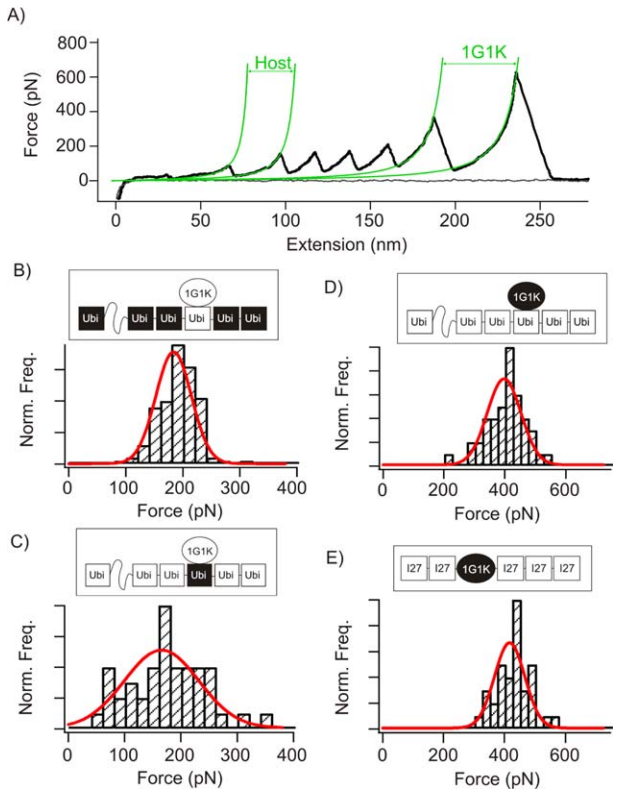

Figure 7

Experimental comparison of the unfolding force distribution of a highly mechanostable protein in series or as a guest of a lower stability host. The comparison involves protein 1G1K as the guest molecule. This protein has not yet been studied experimentally in this context. **A:** An $F-d$ trace of the unfolding of pFS-2/1UBQ-1G1K (equivalent to 2*1UBQ-1UBQ(9-10,1G1K)-2*1UBQ). The host unravels before the guest, and 1G1K unfolding peak typically appears at the end since it is the most mechanostable one. **B–E:** Unfolding force histograms of the single-molecule marker ubiquitin (B), host ubiquitin (C), guest 1G1K (D), and series 1G1K (E) (this last histogram comes from Ref. 24). Insets show a cartoon representation of the proteins used where the module being analyzed is highlighted. Marker ubiquitin shows a similar unfolding force (190 ± 32 pN, $n = 141$) to that reported for other pFS-2 proteins,[10] while host ubiquitin, as expected, shows a more disperse distribution but with a similar average (175 ± 66 pN, $n = 43$). The broader distribution reflects additional degrees of freedom in the system due to the presence of the guest, as expected.[10] Furthermore, grafted and serially arranged 1G1K show similar distributions (403 ± 64 pN, $n = 43$ and 425 ± 55 pN, $n = 35$, respectively). It should be noted that, for 1G1K, the additional peaks observed *in silico* are not resolved experimentally. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Experimental data on guests of various ratios of mechanostability

Recently, the unfolding force of guest proteins with similar[10] or lower[36] mechanical stability than their host has been experimentally studied and found that the grafted protein had a similar mechanostability that the serially arranged one. Here, we also stretch an HG protein where the guest presents higher mechanical stability compared to the host. We have engineered a protein containing cohesin I of *Clostridium cellulolyticum* scaffoldin (1G1K) as a guest in a ubiquitin host of the pFS-2 vector[10] (Fig. 7). In this HG configuration, 1G1K shows an unfolding force of 403 ± 64 pN) (mean ± standard deviation; based on 43 molecules) that is slightly lower than that reported previously (425 ± 55 pN) for the serially arranged protein[24] although this difference is not statistically significant ($P = 0.1$ according to Welch 2 sample t-test). This result supports the predictions of our simulations. Thus, the observed unfolding force of a HG-grafted protein is similar to that measured in serial constructions. It is interesting to note that most $F-d$ recordings obtained had to be discarded due to the absence of the unfolding peak of the host ubiquitin even though 1G1K peak was observed, which was probably due to a destabilization or even the prevention of fold formation of the host module. Similar destabilizing effects of a grafted protein have been described in the literature, and specifically, when two C-cadherin modules were expressed in the pFS-2 vector, no carrier was observed (see Ref. 10 and references therein). This effect might be reduced if longer linkers were added between the host and the guest modules as in the structural models performed in this study.

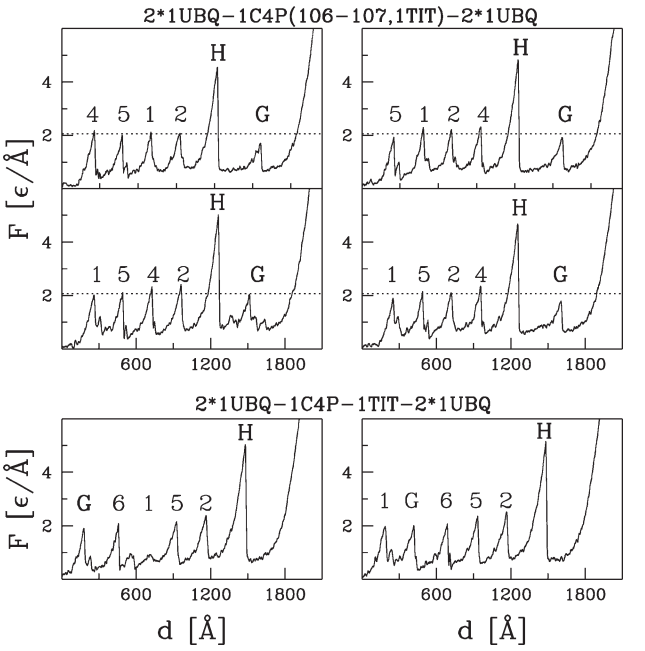

Figure 8

The case of G weaker than both A and H (H > A > G). The top four panels show examples of the $F-d$ trajectories for the chain 1UBQ–1UBQ–1C4P–1UBQ–1UBQ in which 1C4P hosts 1TIT. Thus, here G=1TIT, H=1C4P, and A=1UBQ. The bottom two panels show examples of trajectories when all proteins are connected in series. The symbols above the $F-d$ traces indicate the identity of the domain that unravels around the value of d. The dotted line shows the value of F_{max} for an isolated 1TIT. G is inserted between sites 106 and 107 in H.

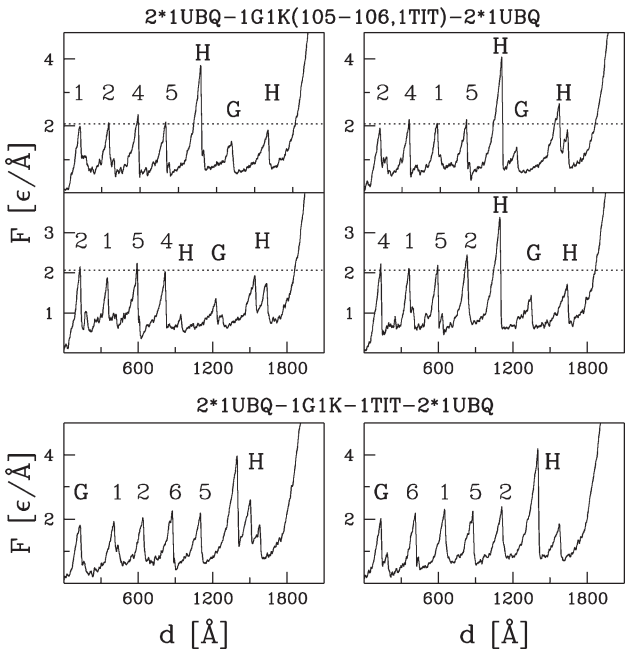

Figure 9

The case of G weaker than both A and H (H > A > G). Similar to Figure 8 but the central host molecule 1C4P is replaced by 1G1K, that is, G=1TIT, H=1G1K, and A=1UBQ.

Heterogeneous host chain

Our analysis so far indicates that the most reliable experimental scenario involving homogeneous chains is one in which the investigated protein G is weaker than A (which coincides with H) since then unraveling of H indeed acts like a marker for the $F - d$ trace associated with G. One can enhance the disparity between H and A by considering heterogeneous chains in which H is stronger than A. Figure 8 gives an example of a strong single-peaked H—it is for A=1UBQ, H=1C4P, and G=1TIT. For the serial connectivity, G may unravel at various stages, but H will always come last. For the HG connectivity, however, the last stage involves unraveling of H that is followed by unraveling of G. The remaining modules unravel earlier and in an arbitrary order. In Figure 8, the values of F_{max} associated with G are either the same or weaker than for the isolated molecule G. In two top panels (the first and the last), the reduction is about 10%. However, this reduction is within the range associated with the size of thermal fluctuations (typically of order $0.1\,\epsilon/k_B$, but smaller than $0.2\,\epsilon/k_B$). For other examples of G listed in Table 1 in Supporting Information (and based on five trajectories in each case), one observes, on average, either a reduction or an increase in the value of F_{max} for G. The change, however, is again within the scale of the thermal fluctuations.

The situation gets more complicated when the strong central H module generates, a multipeaked pattern, as illustrated in Figure 9. Here, 1C4P is replaced by 1G1K. We observe that now unraveling of G is still coupled to unraveling of H. However, the latter proceeds in stages and G unfolds after the first force peak of H is built. The process is also illustrated in Figure 10, which shows the related conformational changes. Using multipeaked strong proteins as host for HG connection is thus not very convenient as the identification of the studied F

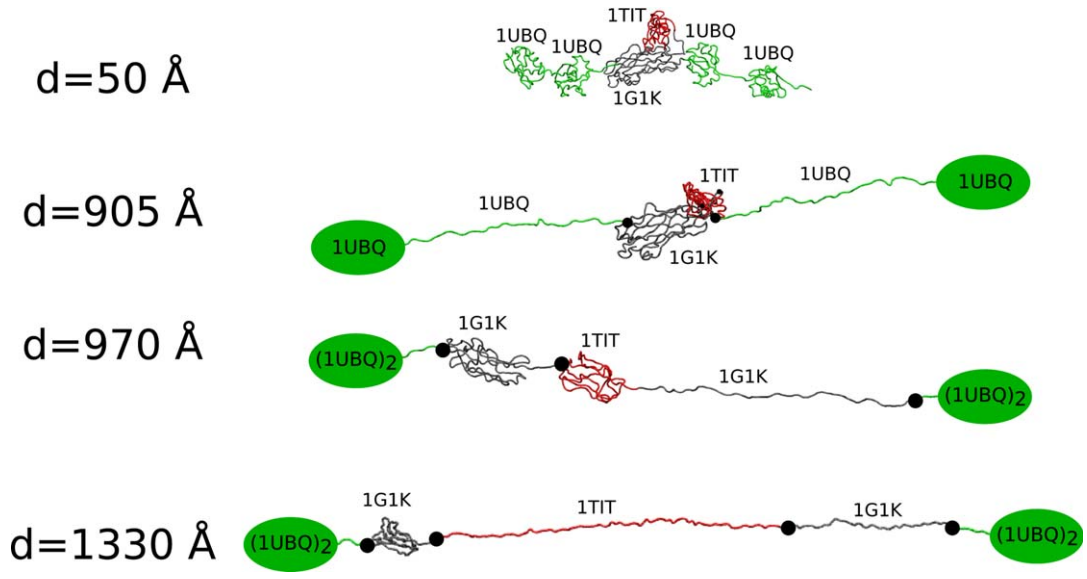

Figure 10

Conformational changes corresponding to the first trajectory (the top left panel) shown in Figure 9. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

– d trace gets difficult. In addition, the values of F_{max} for G are noticeably smaller.

Connecting proteins with cystine knots

The ultimate goal of the new method to study the mechanostability of proteins is to provide unequivocal assignment and characterization of the force peaks associated with a guest protein, which is not trivial particularly in the case of certain intrinsically disordered proteins. As we have shown, this task is helped by using a host that is sufficiently sturdy so that its signature comes later than that of the single-molecule markers and immediately before the features of the guest protein. We have found theoretically[22,37–39] that exceptionally high values of F_{max} are associated with certain proteins, mostly growth factors and super-stable small peptides known as knottins, which incorporate the cystine knot motif.

In Supporting Information, we explore the possibility of using the proteins with the cystine knots as hosts in the novel way of measuring mechanostability. We find that the use would be difficult to implement because cleavage procedures would be required to generate situations in which the relevant mechanical clamp (the cystine slipknot) could be harnessed. Furthermore these procedures themselves would reduce the mechanostability of such hosts.

CONCLUSIONS

We have compared two basic schemes to connect proteins for the purpose of measuring their mechanostability: a serial one and another in which a guest protein is grafted into a host. We have demonstrated, through theoretical and experimental studies involving several different proteins, that the host–guest connectivity works fine for an unambiguous identification of the force–displacement pattern associated with the guest protein independent of whether the main chain of proteins is homogeneous or heterogeneous. One may think that F_{max} of the grafted protein need not, in general, be the same as in its "free" or ungrafted state. However, we have shown that it is, within limits set by thermal fluctuations, provided the host protein comes with a single-force peak.

In general, unraveling of the guest molecule may be separated from that of the host by force peaks originating from the marker molecules. For the guest to unravel just after the host, one of the two conditions has to be met: either the host has to be stronger than the markers or the markers have to be stronger than the guest. It is important to notice that in the HG strategy discussed here, the best configuration, at least experimentally, is to choose host stronger than marker and thus ensuring that guest will be the last to unfold.

ACKNOWLEDGMENTS

We thank R. Hervás for comments on the manuscript. AG-P and AG-S are recipients of FPU (MECD) and JAE-Pre (CSIC) fellowships, respectively.

REFERENCES

1. Ritort F. Single-molecule experiments in biological physics: methods and applications. J Phys Cond Matter 2006;18:R531–R583.
2. Kellermayer MSZ, Smith SB, Granzier HL, Bustamante C. Folding-unfolding in single titin molecules characterized with laser tweezers. Science 1997;276:11126.
3. Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, Fernandez JM. Mechanical and chemical unfolding of a single protein: a comparison. Proc Natl Acad Sci USA 1999;96:36949.
4. Galera-Prat A, Gomez-Sicilia A, Oberhauser AF, Cieplak N, Carrion-Vazquez M. Understanding biology by stretching proteins: recent progress. Curr Op Struct Biol 2010;20:63–69.
5. Brockwell DJ, Paci E, Zinober RC, Beddard G, Olmsted PD, Smith DA, Perham RN, Radford SE. Pulling geometry defines mechanical resistance of β-sheet protein. Nat Struct Biol 2003;10:731737.
6. Crampton N, Brockwell DJ. Unravelling the design principles for single protein mechanical strength. Curr Opin Struct Biol 2010;20: 508–517.
7. Oberhauser AF, Badilla-Fernandez C, Carrion-Vazquez M, Fernandez JM. The mechanical hierarchies of fibronectin observed with single-molecule AFM. J Mol Biol 2002;31:43347.
8. Cieplak M, Hoang TX, Robbins MO. Thermal effects in stretching of Go-like models of titin and secondary structures. Proteins Struct Funct Bio 2004;56:285–297.
9. Cieplak M, Pastore A, Hoang TX. Mechanical properties of the domains of titin in a Go-like model. J Chem Phys 2005;122:054906.
10. Oroz J, Hervas R, Carrion-Vazquez M. Unequivocal single-molecule force spectroscopy of proteins by AFM using pFS vectors. Biophys J 2012;102:682–690.
11. Hervas R, Oroz J, Galera-Prat A, Goñi O, Valbuena A, Vera AM, Gomez-Sicila A, Losad-Urzaiz F, Uversky VN, Menendez M, Laurents DV, Bruix M, Carrion-Vazquez M. Common features at the start of the nuerodegeneration cascade. PLoS Biol 2012;10: e1001335.
12. Humphery W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38.
13. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucl Acids Res 2000;28:235–242.
14. Cieplak M, Hoang TX. Universality classes in folding times of proteins Biophys J 2003;84:475–488.
15. Sułkowska JI, Cieplak M. Mechanical stretching of proteins—a theoretical survey of the Protein Data Bank. J Phys Cond Mat 2007;19: 283201.
16. Sułkowska JI, Cieplak M. Selection of optimal variants of Go-like models of proteins through studies of stretching. Biophys J 2008;95: 3174–3191.
17. The PyMol Molecular Graphics System, Version 1.5.0.4 Schrodinger, LLC; http://www.pymol.org
18. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Chem Phys 2003;119:3559–3566.
19. Vila J, Williams RL, Grant JA, Wojcik J, Scheraga HA. The intrinsic helix-forming tendency of L-alanine. Proc Natl Acad Sci USA 1992; 89:7821–7825.
20. Levy Y, Jortner J, Becker OM. Solvent effects on the energy landscapes and folding kinetics. Proc Natl Acad Sci USA 2001;98:2188–2193.

21. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. J Mol Biol 1999;290:253–266.
22. Sikora M, Sułkowska JI, Cieplak M. Mechanical strength of 17 132 model proteins and cysteine slipknots. PloS Comp Biol 2008;5: e1000547.
23. Grest GS, Kremer K. Molecular dynamics simulation for polymers in the presence of a heat bath. Phys Rev A 1986;33:3628–3631.
24. Valbuena A, Oroz J, Hervas R, Vera AM, Rodrigues D, Menendez M, Sułkowska JI, Cieplak M, Carrion-Vazquez M. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc Natl Acad Sci USA 2009;106:13791–13796.
25. Horcas I, Fernández R, Gómez-Rodríguez JM, Colchero J, Gomez-Herrero J, Baró AM. WSXM: a software for scanning probe microscopy and a tool for nanotechnology. Rev. Scsi. Intrum. 2007; 78:013705.
26. Bycroft M, Ludvigsen S, Fersht AR, Poulsen FM. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. Biochemistry 1991;30:8697–8701.
27. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. Science 1991;253:657–661.
28. Improta S, Politou AS, Pastore A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure 1996;4:323–337.
29. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 1987;194:531–544.
30. Spinelli S, Fierobe HP, Belaich A, Belaich JP, Henrissat B, Cambillau C. Crystal structure of a cohesin module from *Clostridium cellulolyticum*: implications for dockerin recognition. J Mol Biol 2000;304: 189–200.
31. Wang X, Tang J, Hunter B, Zhang XC. Crystal structure of streptokinase beta-domain. FEBS Lett 1999;459:85–89.
32. Best RB, Li B, Steward A, Daggett V, Clarke J. Can non-mechanical proteins withstand force?Stretching barnase by atomic force microscopy and molecular dynamics simulation. Biophys J 2001;81:2344–2356.
33. Cao Y, Li H. Polyprotein of GB1 is an ideal articial elastomeric protein. Nat Mater 2007;6:109114.
34. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 1997;276:1109–1112.
35. Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. Nat Struct Biol 2003;10:73843.
36. Peng Q, Li H. Domain insertion effectively regulates the mechanical unfolding hierarchy of elastomeric proteins: toward engineering multifunctional elastomeric proteins. J Am Chem Soc 2009;131: 14050–14056.
37. Pepłowski L, Sikora M, Nowak W, Cieplak M. Molecular jamming—the cysteine slipknot mechanical clamp in all-atom simulations. J Chem Phys 2011;134:085102.
38. Sikora M, Cieplak M. Formation of cystine slipknots in dimeric proteins. PLoS One 2013;8:e57443.
39. Sikora M, Cieplak M. Cystine plug and other novel mechanisms of large mechanical stability in dimeric proteins. Phys Rev Lett 2012; 109:208101.