Topological features in stretching of proteins

Marek Cieplak[1] and Mateusz Sikora
Laboratory of Biological Physics, Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/48, Warsaw, Poland

Abstract
In the present article, we highlight the diversity of mechanical clamps, some of them topological in nature, that have been found by making surveys of mechanostability of approximately 18 000 proteins within structure-based models. The existence of superstable proteins (with the characteristic unfolding force in the region of 1000 pN) is predicted.

Introduction

Mechanically induced conformational changes of proteins take place in living organisms. Examples include muscle extension, protein translocation, activation of mechano-sensory pathways or of catalytic functions of a protein, and bioadhesion [1–4]. Proteins can also be manipulated mechanically *in vitro* using AFM (atomic force microscopy) or optical tweezers. The most common manipulation is stretching: either at constant force or, usually, at constant speed, v_p. Such *in vitro* studies allow for a detailed characterization of the process. In particular, one can determine F–d patterns at a constant speed, where F denotes tension in the protein and d denotes displacement of the pulling device. Such patterns may come with one or more force peaks. The height, F_{max}, of the largest of them determines the scale of the pattern. Within the experimental range of v_p between 300 and 12 000 nm/s, F_{max} varies with v_p merely logarithmically [5], so the approximate values of F_{max} are meaningful. For constant force experiments, there is also a characteristic force, of the order of F_{max}, above which mechanical unfolding is fast.

The values of F_{max} are typically between 10 and 300 pN [3,5,6], but several proteins have substantially larger values of F_{max}. They include a scaffoldin [7] with F_{max} of 480 pN, green fluorescent protein [8] with F_{max} of up to 548 pN and the protein molecules in the spider capture-silk [9] with F_{max} up to 900 pN. Measurements have been accomplished for only about a hundred proteins so there is little understanding of the scope of possible behaviours. There is thus a need for making predictive theoretical surveys. Owing to the prohibitive computational costs of all-atom models, even for v_p exceeding the experimental speeds by orders of magnitudes, such surveys require coarse-grained models that reduce the number of degrees of freedom and introduce effective couplings between the remaining degrees of freedom. We have made surveys of altogether more than 18 000 proteins by using structure-based coarse-grained models.

We have first considered single-chain proteins comprising up to 250 amino acids [5,10], then multidomain proteins with up to 1021 residues [11] and finally dimeric proteins containing the cystine knot motif [12].

In the present article, we review possible kinds of mechanical clamps, i.e. those structural regions in proteins which are responsible for the enhanced resistance to stretching and thus the emergence of the force peaks. The most common mechanical clamp involves shear between two or more β-strands [1]. In the case of the I27 domain of titin, the shear is between two near-terminal parallel β-strands (linked by six hydrogen bonds) [13] and the corresponding F_{max} is close to 200 pN [6]. However, our surveys have led to the discovery of other kinds of the mechanical clamps. Some of these new clamps do not involve shear, but instead are topological in nature. Furthermore, the corresponding values of F_{max} may be substantially larger than that associated with titin.

The coarse-grained model

Our model [14] describes a monomeric protein as a chain of impenetrable beads located on the C_α atoms that are tethered by harmonic interactions representing covalent interactions: peptide bonds and disulfide bonds. The chain is in an implicit solvent providing thermostatically controlling random forces and damping. Attractive interactions are governed by the native contact map. The map is obtained by representing heavy atoms by (enlarged) van der Waals spheres [15] and checking for their overlaps. If overlaps exist, a contact is present (contacts of the i, $i+2$ type are discarded [5]). A phenomenological way to construct contact potentials has been proposed by Abe and Go [16]; they should be minimized in the native structure. There are many such potentials and the term 'Go-like model' (or 'structure-based model') has no unique meaning. We have analysed 62 variants of possible models [17]. They differ by the functional form of the potential and other attributes. The selection of optimal variants can be based on making comparisons with the experimental values of F_{max} (stretching starts from the native state that defines the model). This was first performed for 29 [17] and then 38 [5] systems.

In the stretching protocol, one terminus is anchored and the other is pulled by an elastic spring mimicking the elasticity of the AFM cantilevers [18]. Values of F_{max} are not sensitive to the choice of the spring constant. The speed v_p in typical simulations is 0.005 Å/τ (1 Å = 0.1 nm), or approximately 500 000 nm/s, since the time scale τ is of order 1 ns due to overdamping. When all native contacts are ruptured, F_{max} diverges because only covalent bonds are then stretched. Among the variants considered, four models were found to be optimal, including one with the Lennard–Jones potentials in the contacts and with the local backbone stiffness defined through a chirality potential [19] favouring the native values of the dihedral angles. In this optimal model, the amplitudes of the potentials are all equal to ε, whereas the length parameters σ_{ij} are determined so that the minimum is located at the native distance between the two C_α residues in contact. We have calibrated ε to be 110 ± 30 pN/Å [5]. The accuracy of the calibration seems sufficient for comparisons of mechanostability between proteins. The room temperature is then within the range 0.3–0.4 ε/k_B. In this range, our model proteins are good folders. Our simulations are carried out either at 0.3 or at 0.35 ε/k_B; the choice affects F_{max} only a little. Mechanical clamps were identified through monitoring of rupture events around F_{max}.

Surveys of mechanostability of proteins and mechanical clamps

We have determined theoretical values of F_{max}, which are available online in the BSDB database described in [20]. We have found that the average F_{max} in the set of 17 134 single-domain chains is 130 pN [5]. The distribution of the forces has a tail reaching above 1000 pN. The types of the mechanical clamps found in monomeric proteins are listed and described in [20]. The most common clamps are due to shear. Their stability depends on the length of the β-sheets involved and on whether there is a stabilization on the sides. However, the top values of F_{max} have been predicted to be associated with extracellular proteins containing a cystine knot motif [21,22] and to have a mechanical clamp that is topological in nature.

The cystine knot motif is an interlaced structural arrangement involving three cystines, i.e. three pairs of cysteine residues connected by disulfide bonds. Two of these cystines connect two short segments of the backbone and transform them into an effective ring (of approximately eight residues). The third one joins two other parts of the backbone through the ring. This tight structure provides substantial thermodynamic stability. The mechanostability of such proteins has not yet been measured, but we expect F_{max} to be at least 1000 pN. The force peaks in such proteins arise due to formation of a cystine slipknot when the ring-piercing cystine drags the backbone through the ring. We have studied the 13 top-strength proteins with the cystine slipknot mechanism also by using all-atom simulations [23] and confirmed the existence of this mechanism. Furthermore, F_{max} has been found to be significantly larger than for titin at the large v_p used.

Figure 1 | **Dynamics of the knot during stretching**

Upper panel: an example of force–displacement (F–d) pattern for YibK methylotransferase (PDB code 1J85). Lower panel: corresponding sequential locations of the knot ends. Adapted from Figure 1 in Sułkowska, J.I., Sułkowski, P., Szymczak, P. and Cieplak, M. (2008) Tightening of knots in proteins. Phys. Rev. Lett. **100**, 058106 with permission.
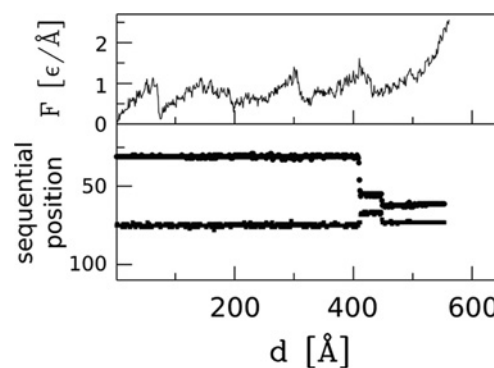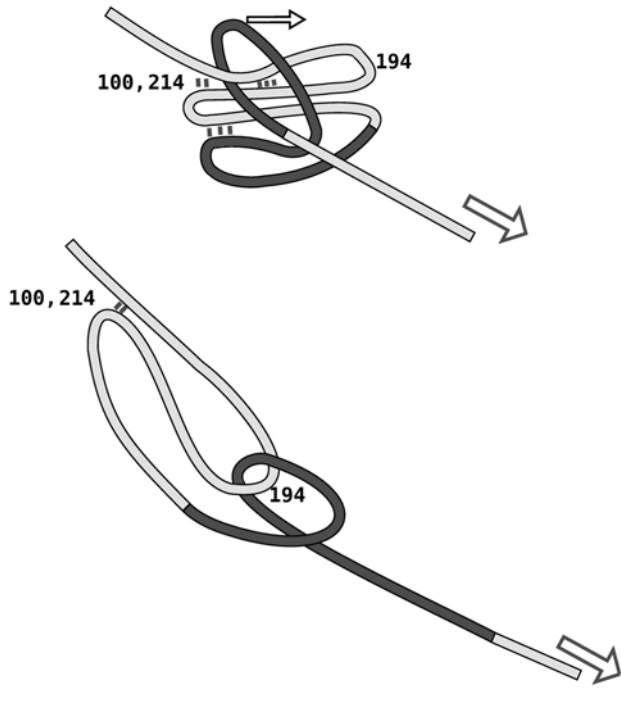

Proteins with native non-cystine knots [24] do not generate novel mechanical clamps. Instead, the knots are tightened in sudden jumps along the sequence [25] instead of diffusing around as found in single DNA molecules [26]. This jumping behaviour is illustrated in Figure 1 for the model methyltransferase YibK comprising 156 residues. When stretching this protein, several force peaks arise, indicating unravelling of the tertiary structure that takes place in stages. Despite the significant unravelling of the structure, the native trefoil knot stays sequentially localized in the same place until the fourth force peak is generated. At this point, the knot ends (defined through a procedure in which the length of the backbone is shortened in steps from both termini until the knot motif disintegrates), instead of being separated by 44 residues, switch to being only 15 residues apart. Later on, there is one more jump to a separation of seven residues and the knot is tightened maximally. Studies of 700 trajectories indicate that there are preferred locations where the knot ends land. These are sharp turns and end points of helices. Since the proteinic backbones do not form closed loops, stretching may result either in knot tightening or in its untying. The outcome of stretching depends on the selection of amino acids to pull by and on the temperature [27], i.e. it depends on the specific trajectory. The presence of a knot appears to enhance thermodynamic stability [28] of a protein, but its other roles remain to be elucidated.

Our survey of the multidomain proteins [11] has led to the discovery of several novel mechanisms such as the tensile mechanical clamp in which two domains, when stretched apart, generate a tensile strain within contacts that couple the domains. Two of these novel clamps are illustrated in Figure 2 for glucosidase. Both images correspond to late stages of unfolding, after multiple shear-based force peaks have been generated and a pseudoknot-loop (shown in black) is formed. In the upper image, the loop cuts through a sequence of

Figure 2 | **Non-shear-based mechanisms involved in stretching of glucosidase with the PDB code 1BHG and 613 residues**

Upper panel: the knot-loop indicated in the darker shade slides to the right and tears the contacts in front of it. Lower panel: the slipknot indicated in the lighter shade generates tensile strain and opens the loop. The first mechanism comes with F_{max} of 3.9 $\varepsilon/\text{Å}$ and the second with 1.8 $\varepsilon/\text{Å}$. The numbers indicate sequential positions of selected amino acids.

Figure 3 | **Mechanisms involved in the C–C′ stretching of proteins 2B1Y (left-hand panel) and 2GH0 (right-hand panel)**

The former involves shear and the latter involves the formation of two cystine slipknots.
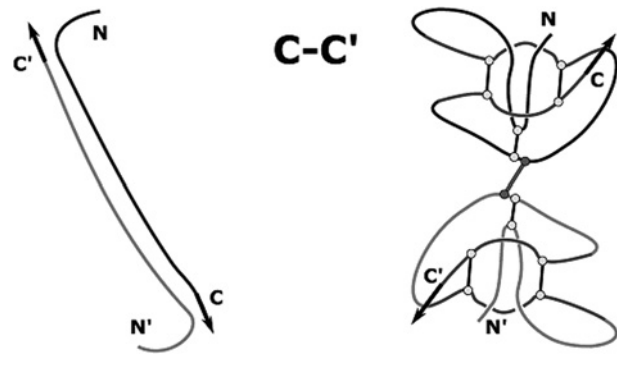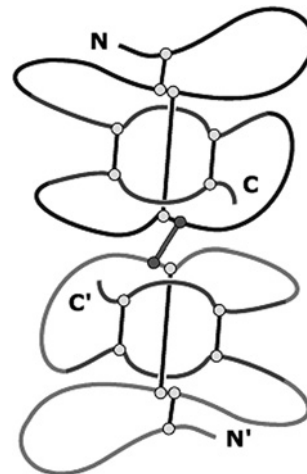

Figure 4 | **Schematic representation of the native state of 1TFG to illustrate the nature of connectivity**

The lighter circles correspond to cysteine residues from the cystine rings. The darker circles show cysteine residues that link the monomers. The termini are indicated. The unprimed symbols refer to one monomer and the primed symbols refer to the other. The straight segments represent disulfide bonds. The central bond links the two monomers together.

contacts in a sliding motion and produces an F_{max} which is nearly twice as large as in titin. Some of these contacts reform after the loop went by. Eventually, this loop starts pushing on the blocking loop which generates a tensile stress on the contacts either left behind or reconstituted during the previous sliding motion and builds another force peak. The protein unravels once the slipknot generated in this process is released.

Dimeric proteins enrich possible mechanical behaviour further still. If the termini of one monomer are denoted by N and C and of another by N′ and C′ then F_{max} depends on which pair of the termini is chosen to implement stretching. For instance, a protein with PDB code 2B1Y shown schematically in the left-hand panel of Figure 3 has the two monomers intertwined in such a way that C–C′ pulling generates strong shear and F_{max} of the order 1000 pN in our model. The N–C′ pulling causes unzipping and F_{max} of approximately 170 pN [12]. Cystatin C is intertwined in a different way and the N–N′ pulling is predicted to result in F_{max} of 770 pN and C–N′ pulling in F_{max} of ∼190 pN [11].

Many of the proteins with the cystine knot are dimers linked by one or two disulfide bonds. Proteins with PDB codes 2GH0 (right-hand panel of Figure 3) and 1TFG (Figure 4) are examples of the former. The C–C′ pulling in

2GH0 results in the generation of two slipknots, in each of the cysteine knots, and hence of two force peaks, with the larger F_{max} of the order 1300 pN, and no force peaks for the N–N′ and N–C pulling as the cystine slipknots cannot form in such geometries [12]. Protein 1TFG is similar to 2GH0, but contains extra disulfide bonds near the N- and N′-termini. These bonds close the backbone into rings (of ten residues) that act as plugs. Stretching by the N–C′ or C–C′ termini results in dragging the plugs through the cystine rings with F_{max} ∼ 1550 pN [12]. This cystine plug mechanism would lead to the largest proteinic mechanostability known, but an experimental verification is necessary.

Funding

References

1 Crampton, N. and Brockwell, D.J. (2010) Unravelling the design principles for single protein mechanical strength. Curr. Opin. Struct. Biol. **20**, 508–517

2 Del Rio, A., Perez-Jimenez, R., Liu, R.C., Roca-Cusachs, P., Fernandez, J.M. and Sheetz, M.P. (2009) Stretching single talin rod molecules activates vinculin binding. Science **323**, 638–641

3 Carrion-Vazquez, M., Oberhauser, A.F., Fisher, T.E., Marszalek, P.E., Li, H. and Fernandez, J.M. (2000) Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. Prog. Biophys. Mol. Biol. **74**, 63–91

4 Vogel, V. (2006) Mechanotransduction involving multimodular proteins: converting force into biochemical signals. Annu. Rev. Biophys. Biomol. Struct. **35**, 459–488

5 Sikora, M., Sułkowska, J.I. and Cieplak, M. (2009) Mechanical strength of 17134 model proteins and cysteine slipknots. PLoS Comp. Biol. **5**, e1000547

6 Fowler, S.B., Best, R.B., Toca Herrera, J.L., Rutherford, T.J., Steward, A., Paci, E., Karplus, M. and Clarke, J. (2002) Mechanical unfolding of a titin Ig domain: structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. J. Mol. Biol. **322**, 841–849

7 Valbuena, A., Oroz, J., Hervas, R., Vera, A.M., Rodriguez, D., Menendez, M., Sułkowska, J.I., Cieplak, M. and Carrion-Vazquez, M. (2009) On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc. Natl. Acad. Sci. U.S.A. **106**, 13791–13796

8 Dietz, H., Berkemeier, F., Bertz, M. and Rief, M. (2006) Anisotropic deformation response of single protein molecules. Proc. Natl. Acad. Sci. U.S.A. **103**, 12724–12728

9 Becker, N., Oroudjev, E., Mutz, S., Cleveland, J.P., Hansma, P.K., Hayashi, C.Y., Makarov, D.E. and Hansma, H.G. (2003) Molecular nanosprings in spider capture-silk threads. Nat. Mat. **2**, 278–283

10 Sułkowska, J.I. and Cieplak, M. (2007) Mechanical stretching of proteins: a theoretical survey of the Protein Data Bank. J. Phys.: Cond. Mat. **19**, 283201

11 Sikora, M. and Cieplak, M. (2011) Mechanical stability of multi-domain proteins and novel mechanical clamps. Proteins: Struct., Funct., Bioinf. **79**, 1786–1799

12 Sikora, M. and Cieplak, M. (2012) Cystine plug and other novel mechanisms of large mechanical stability in dimeric proteins. Phys. Rev. Lett. **109**, 208101

13 Lu, H. and Schulten, K. (1999) Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interpret atomic force microscopy observations. Chem. Phys. **247**, 141–153

14 Cieplak, M. and Hoang, T.X. (2003) Universality classes in folding times of proteins. Biophys. J. **84**, 475–488

15 Tsai, J., Taylor, R., Chothia, C. and Gerstein, M. (1999) The packing density in proteins: standard radii and volumes. J. Mol. Biol. **290**, 253–266

16 Abe, H. and Go, N. (1981) Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. Biopolymers **20**, 1013–1031

17 Sułkowska, J.I. and Cieplak, M. (2008) Selection of optimal variants of Go-like models of proteins through studies of stretching. Biophys. J. **95**, 3174–3191

18 Cieplak, M., Hoang, T.X. and Robbins, M.O. (2004) Thermal effects in stretching of Go-like models of titin and secondary structures. Proteins: Struct., Funct., Bioinf. **56**, 285–297

19 Kwiecinska, J.I. and Cieplak, M. (2005) Chirality and protein folding. J. Phys.: Cond. Mat. **17**, S1565–S1580

20 Sikora, M., Sułkowska, J.I., Witkowski, B. and Cieplak, M. (2011) BSDB: the Biomolecule Stretching Database. Nucleic Acids Res. **39**, D443–D458

21 Sun, P.D. and Davies, D.R. (1995) The cystine-knot growth-factor superfamily. Annu. Rev. Biophys. Biomol. Struct **24**, 269–291

22 Isaacs, N.W. (1995) Cystine knots. Curr. Opin. Struct. Biol. **5**, 391–395

23 Peplowski, L., Sikora, M., Nowak, W. and Cieplak, M. (2011) Molecular jamming: the cysteine slipknot mechanical clamp in all-atom simulations. J. Chem. Phys. **134**, 085102

24 Taylor, W.R. and Lin, K. (2003) Protein knots: a tangled problem. Nature **421**, 25

25 Sułkowska, J.I., Sułkowski, P., Szymczak, P. and Cieplak, M. (2008) Tightening of knots in proteins. Phys. Rev. Lett. **100**, 058106

26 Bao, X.R., Lee, H.J. and Quake, S.R. (2003) Behavior of complex knots in single DNA molecule. Phys. Rev. Lett. **91**, 265506

27 Sułkowska, J.I., Sułkowski, P., Szymczak, P. and Cieplak, M. (2010) Untying knots in proteins. J. Am. Chem. Soc. **132**, 13954–13956

28 Sułkowska, J.I., Sułkowski, P., Szymczak, P. and Cieplak, M. (2008) Stabilizing effect of knots on proteins. Proc. Natl. Acad. Sci. U.S.A. **105**, 19714–19719