Force induced unfolding of proteins

I. Mechanically active proteins

Native structures of wild-type proteins may unfold as external conditions change. Usually unfolding occurs as a result of raising temperature, increasing the concentration of denaturants such as urea or guanidinium chloride, or changing pH. Only recently it has been recognized that certain wild-type proteins are subject to external force and in the course of performing biological functions they may unravel to adopt highly extended, stretched conformations.

The process of native state unraveling due to external force is referred to as mechanical or force-induced unfolding. Mechanical unfolding is implicated in many cellular processes, such as muscle contraction and relaxation, cell adhesion, transport of proteins through membranes, cytoskeleton elasticity etc. To appreciate the features of mechanically active proteins consider several examples.

Titin: This protein, which apparently is the longest covalently linked molecule (the length of up to 1 μm), plays a crucial role in muscle contraction and relaxation. Titin spans almost half of the muscle sarcomer, including A- and I-band regions, and consists of about 30,000 amino acids and 300 domains.

There are two type of domains, immunoglobulin (Ig) and fibronectin (Fn) modules, as well as PEVK regions, which contain large number of charged residues. Although PEVK regions are unstructured (Fig. 1), both Ig and Fn have well-defined, β-sandwich native fold. Fn and Ig domains and PEVK regions are linked in tandem and form long necklace-like structure (Fig. 1).

Because of modular construction titin behaves as a non-linear entropic spring upon application of external stretching force. The mechanical response of titin includes three stages:

(i) Initial straightening of domains in a protein (small forces, Fig. 1a);

(ii) stretching of PEVK regions (intermediate forces, Fig. 1b);

(iii) one-by-one sequential unfolding of individual domains (large forces, Fig. 1c)

It appears that the extensions of titin in skeletal muscles fall into first two categories. However, titin in cardiac muscles experiences larger stretching forces and there is experimental evidence that individual Ig domains do unfold.

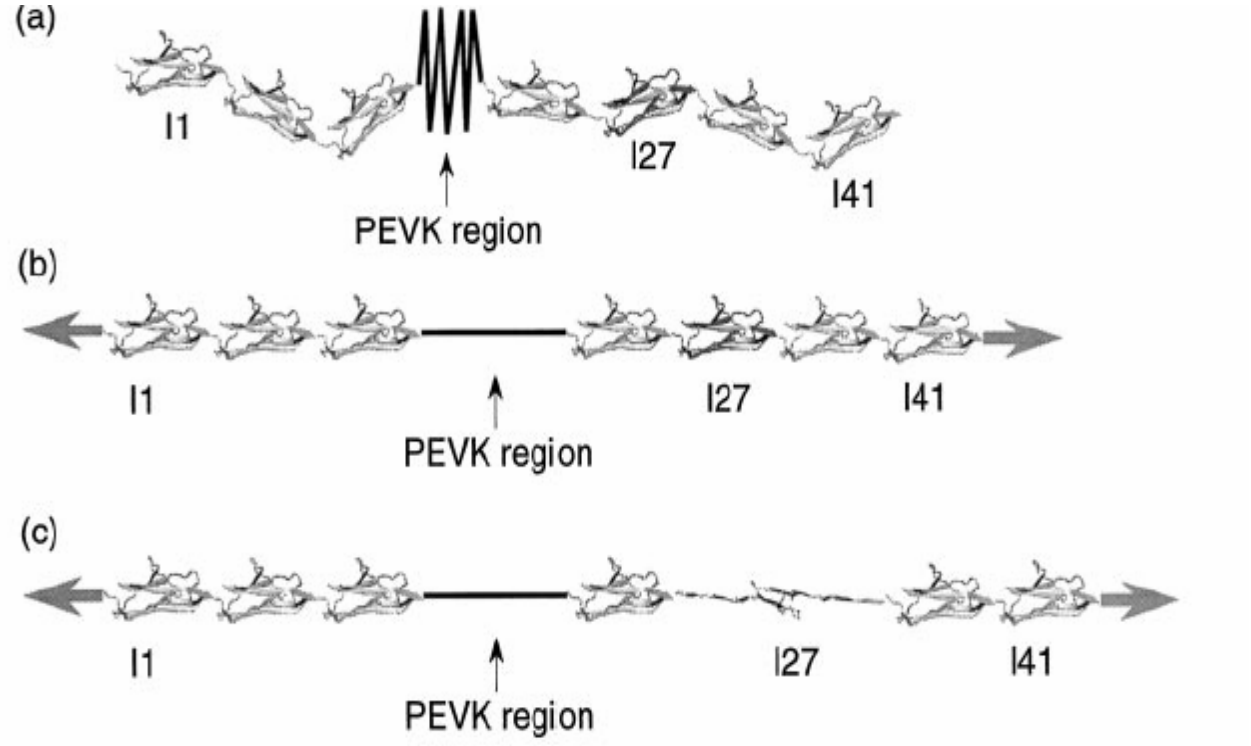

Fig. 1 Modular structure of titin (part of I-band region is shown). Individual Ig domains and PEVK regions are linked in a tandem of structural repeats. Successive stages of titin unfolding are displayed as a function of increasing force ((a) through (c)) (*Biophys. J.*, **75**, 662 (1998)).

I-band region of titin, which is largely responsible to titin stretching, consists of PEVK regions and Ig domains (Fig. 1). The number of Ig domains varies from 37 to 90 in various titin molecules. Although their sequences differ, Ig domains have similar native structure. As an example consider Ig27 domain (Fig. 2). This domain has 89 amino acids and the β-sandwich native structure with two β-sheets (strands A'GFC and ABED).

The native β-sandwich topology is stabilized by extensive hydrophobic core. Note that the domain terminals (the beginning of the strand A and the end of the strand G) are located on the opposite sides of the native structure. Such native architecture of the domain permits modular construction of titin. The native distance between terminals (so called end-to-end distance r_{1N}) is 49 Å, but it increases almost 10 times upon stretching (in the fully extended state $r_{1N}\approx$330 Å). The equilibrium force required to unfold Ig domains ranges from approximately 5 to 40 pN.

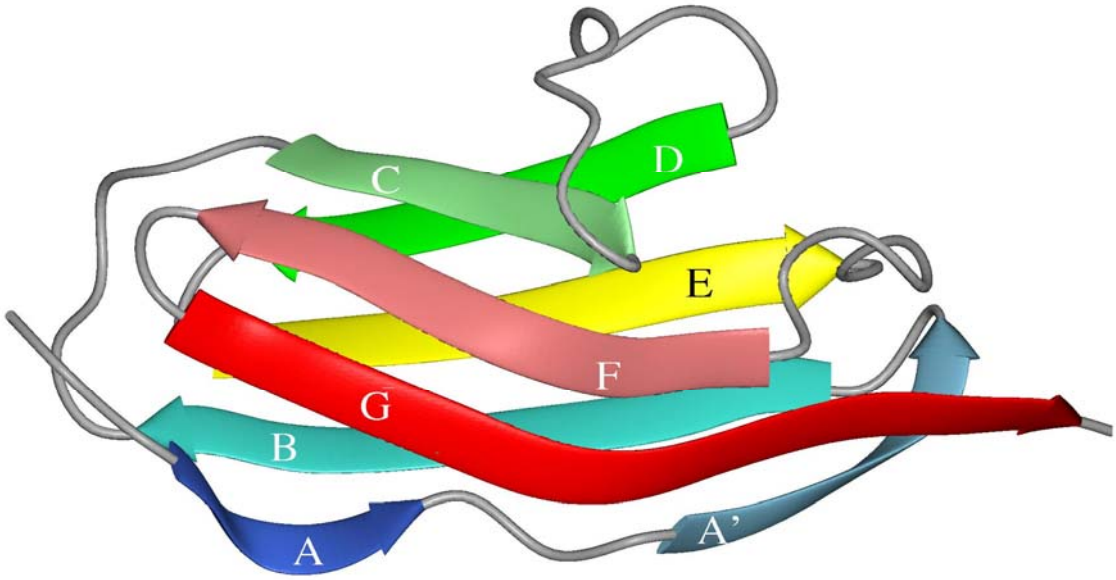

Fig. 2 Native structure of Ig27 domain in titin. Letters indicate β-strands.

Fibronectin: Extracellular matrix (ECM) is a collection of proteins in the extracellular environment, which serves as a template for adhesion of cells. Fibronectin, an ECM protein, is composed of about 20 distinct modules (namely, Fn domains, which are classified into FnI, FnII, and FnIII types). Binding between integrins and Fn domains anchors a cell onto ECM. Experiments show that fibronectin in ECM is subject to mechanical tension (of about 5 pN), which is strong enough to unfold some of Fn domains. Consequently, the length of fibronectin may extend up to four-fold in ECM.

The remarkable feature of FnIII domains is that their mechanical stretching exposes or changes the affinity of cryptic binding sites, which are buried in the FnIII fold in the absence of external tension. Once exposed cryptic binding sites initiate interactions between fibronectin molecules and trigger their aggregation. In such elongated conformations fibronectin forms fibril-like macromolecular assemblies. Therefore, stretching of fibronectin in ECM controls cell adhesion and mobility.

Consider the structure of the 10^{th} FnIII domain (10FnIII, Fig. 3). As Ig27 10FnIII domain is a typical β-sandwich protein composed of two β-strands. The integrin binding site in10FnIII is located in the RGD loop between the strands F and G. Mechanical stretching of RGD loop reduces the affinity of the binding site. As a result the interactions with cell's integrin proteins become broken and cell disengages from ECM. Releasing the external tension restores the affinity of the RGD binding site. Therefore, 10FnIII module functions as mechanical nano-switch, which binds or releases cells depending of applied force. In general, integrin and cryptic binding sites are distributed among different FnIII domains in fibronectin. Interestingly, mechanical unfolding turns off one binding sites and activates new ones, depending on the applied force.

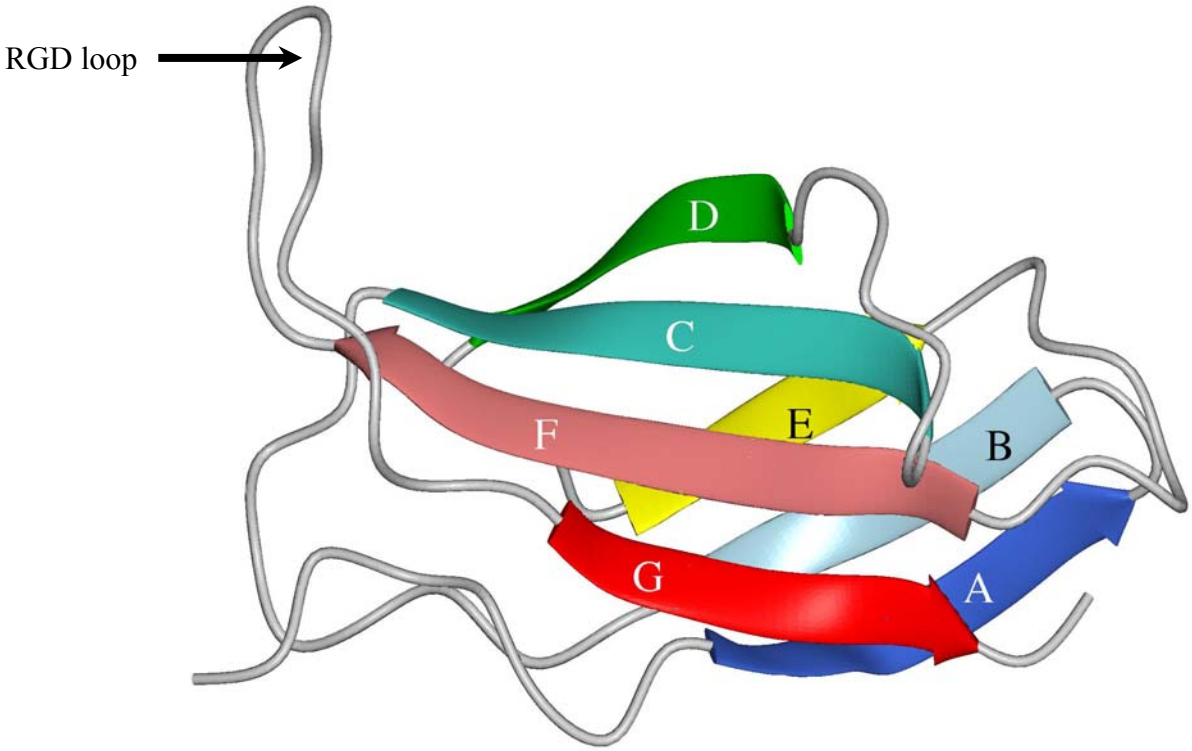

Fig. 3 Native structure of 10FnIII domain of fibronectin. Letters denote β-strands. The integrin binding site is located in the RGD loop.

Most other mechanically active proteins have generally similar native topology as those described above for Ig27 and 10FnIII. The common structural characteristics of these proteins are:

1. β-sandwich native topology, in which two β-sheets are packed against each other;

2. In order to accommodate tandem arrangement of individual domains chain terminals are located on the opposite sides of native conformations;

3. Hydrogen bonds stabilize β-sheets, while hydrophobic interactions maintain packing of β-sheets in the β-sandwich.

Other mechanically active proteins: Although most of mechanically active proteins contain β-structure, there are few examples of proteins with other native state architectures. For instance, α-spectrin, a three helix bundle protein, is found in the cytoskeleton of erythrocytes. Being the component of cell walls, it determines their mechanical properties with respect to deformation, stretching etc. Another example is α/β protein barnase, which is implicated in partial force induced unfolding during its transport through membranes.

II. Experimental studies of mechanical unfolding

Experimental study of force induced unfolding of single protein molecules has become possible with the development of atomic force microscopy (AFM) technique (Fig. 4). A multidomain protein is adsorbed on a golden template. A tip of the cantilever (Fig. 4) is pressed against the adsorbed protein and then is retracted with a constant speed. Cantilever's tip forms interactions with one of the protein molecules and as it is gradually moved away a protein extends as shown in Fig. 4.

Protein extension is triggered by unfolding of protein's domains. Because protein is extended beyond its equilibrium length, the molecular restoring force is generated, which causes the cantilever to bend. Laser registers the bending angle, and therefore, the unfolding force. The unique feature of AFM is the ability to record force-extension curves for individual protein molecules with the accuracy of few pico-Newtons.

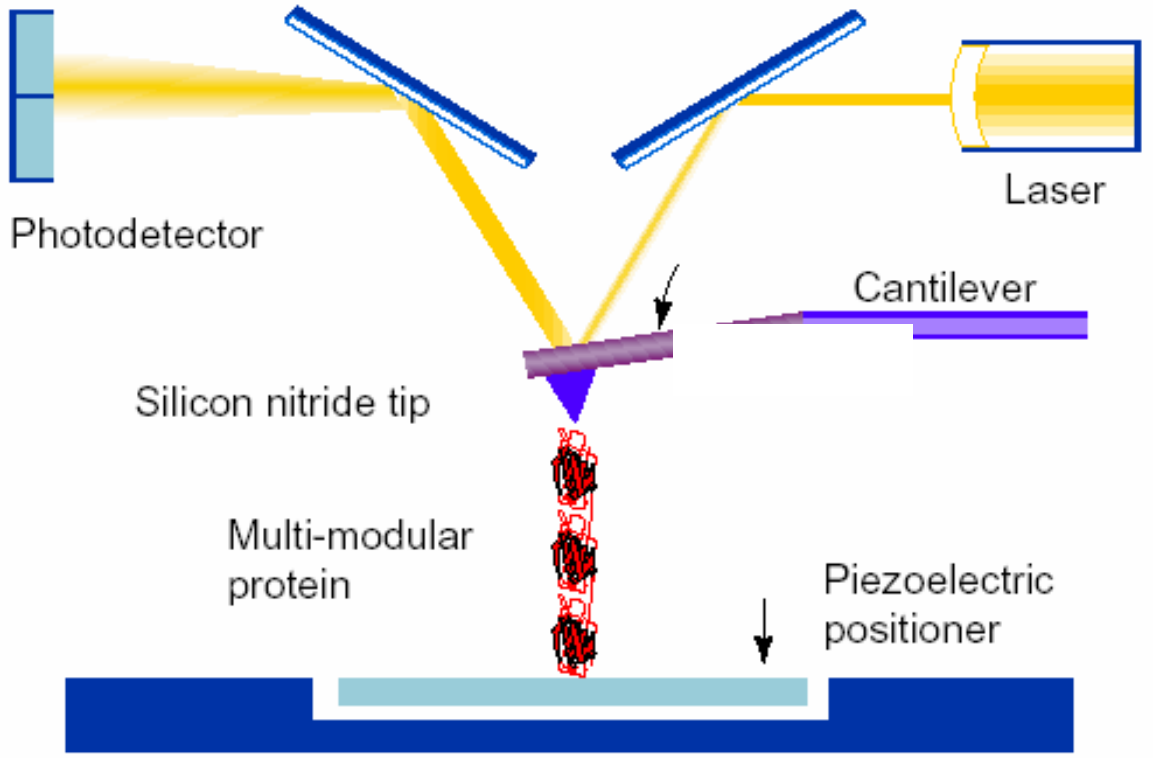

Fig. 4 Diagram of the typical AFM experiment (*Trends in Biochemical Sciences* **24**, 379 (1999)).

The example of three typical force-extension profiles recorded by AFM is shown in Fig. 5. This figure shows how the force generated in the unfolding protein depends on the extension. All such force-extension plots have a characteristic sawtooth-like profile, which is interpreted as a signature of a modular structure of the protein. The profile in Fig. 5 was obtained for the protein construct made of 12 domains. In this experiment the

number of well-defined force peaks never exceeded 12, but often was actually less, because cantilever's tip could pick up the adsorbed protein anywhere along its sequence. Furthermore, the amplitudes of force peaks (excluding the first and the final ones) are generally maintained from one stretch experiment to another. Although it is not possible to prove this directly, all observations suggest that each force peak (apart from initial and final ones) corresponds to the unfolding of an individual protein domain. Therefore, AFM experiments measure the force amplitudes and spacing between peaks. Note that the force amplitude does depend on the speed of cantilever retraction (so called pulling speed).

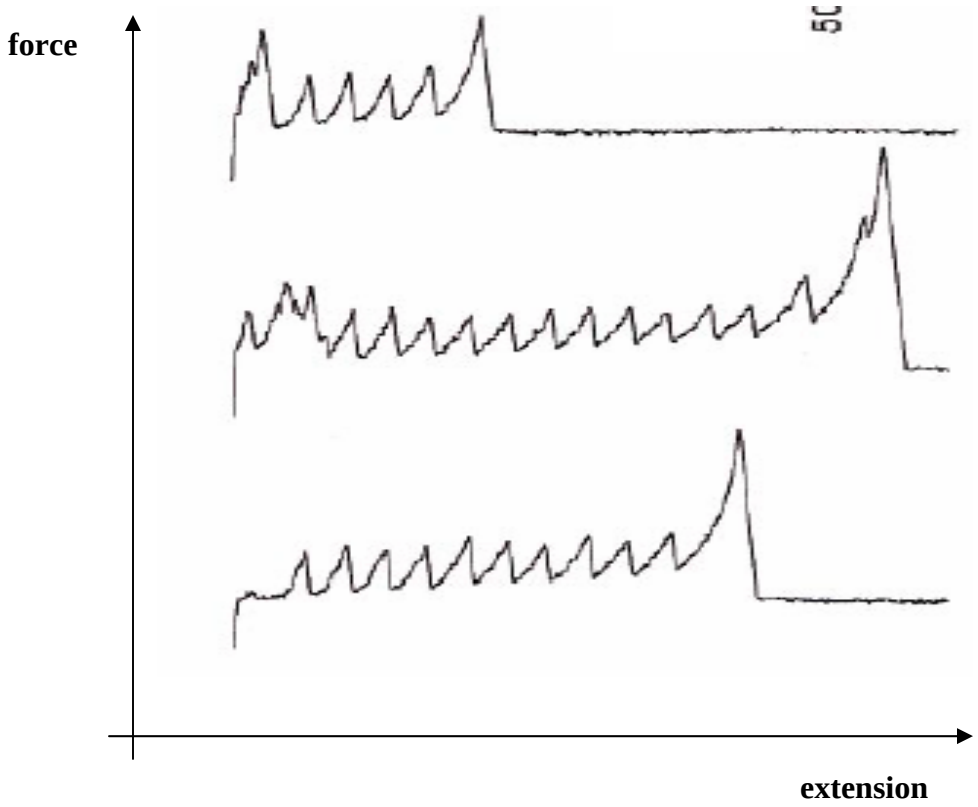

Fig. 5. Examples of the AFM force-extension profiles recorded for 12 domain protein (*Nature Structural Biology* **7**, 719 (2000)).

Force-induced unfolding of titin has been studied by Gaub and coworkers (*Science* **275**, 1295 (1997)). Specifically, the experiments registered the forces generated by unfolding of the tandem of eight titin domains (Ig27 through Ig34) at the pulling speed of 1 μm/s. It was found that unfolding forces were distributed in the range between 150 to 300 pN. Interestingly, the peaks were arranged in the ascending order of their amplitude, suggesting that the weakest domains unfold first. Similar results were reported for wild-type tenascin (the average force amplitude is 137 pN, *Nature* **393**, 181 (1998)), which, as titin, is composed of β-sandwich domains. Surprisingly, the mechanical stability of α-helix proteins is far lower. AFM experiments conducted for spectrin revealed that the

average unfolding force is between 25 and 35 pN (*Journal of Molecular Biology* **286**, 553 (1999)).

Because wild-type proteins have heterogeneous domain composition, the peaks in their force-extension profiles cannot be easily identified with specific domains. The solution is offered by protein engineering methods, which permit to produce polyproteins, containing identical domains. The force-extension profile (upper plot in Fig. 6) is obtained for the wild-type fragment containing Ig domains, whereas the lower plot describes the unfolding of the engineered $(Ig27)_8$ construct, which features nearly constant peak amplitude. The latter one allows for unambiguous determination of the average unfolding force for Ig27 domain (≈204 pN, *Proceedings of the National Academy of Sciences USA* **96**, 3694 (1999)). It is worth noting that the average unfolding force for the polyprotein $(Ig28)_8$ composed of Ig28 domains is 257 pN.

This result is intriguing because the stability of Ig28 domain against chemical denaturation is lower than for Ig27. Therefore, the stability of the native state against thermal or chemical denaturation does not correlate with the mechanical stability. Furthermore, measurements of the force-extension profiles for $(Ig27\text{-}Ig28)_4$ construct established that the average force increases to 300 pN, thus implicating strong interdomain interactions. Accordingly, one may conclude that the mechanical properties of modular proteins are not, as a rule, the sum of the mechanical properties of their components.

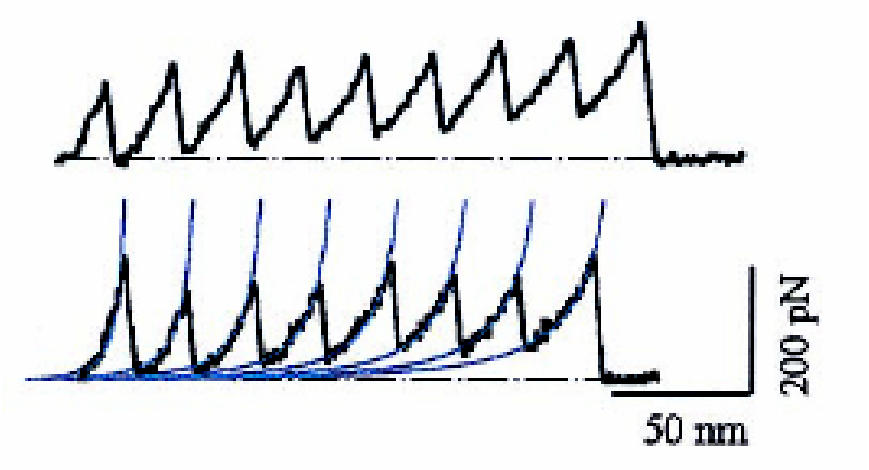

Fig. 6 Force-extension profiles for wild-type Ig domains (upper plot) and engineered polyprotein $(Ig27)_8$ composed of identical Ig27 domains (*Proceedings National Academy of Sciences* **96**, 3694 (1999)).

Worm-like chain (WLC) model: It is possible to extract not only force amplitudes from force-extension profiles, but the number of residues being unfolded in one unfolding event. For this purpose the ascending slopes of the force peaks may be fit with the worm-like chain model (WLC). The WLC model describes the equilibrium stretching of an

entropic chain without excluded volume interactions. The equation that relates entropic force f generated as a result of extension by the length z is given by

$$f(z) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1 - \frac{z}{L_c}\right)^{-2} - \frac{1}{4} + \frac{z}{L_c}\right], \qquad (1)$$

where L_c is the total contour length of the protein, p is the persistence length, T is the temperature and k_B is the Boltzmann constant (*Macromolecules* **28**, 8759 (1995)). Because the adjustable parameters in WLC model are, L_c and p, one can compute the length of the protein ΔL released by unfolding event by fitting successive force peaks. Fits of the force peaks using WLC model are indicated in Fig. 6 by blue curves. Using this methodology the contour length of the fibronectin domain in tenascin was estimated to be 24.7 nm.

WLC model offers the possibility to register even more subtle events in force-induced unfolding. It was observed that AFM force-extension profiles for Ig27 have a "hump" on the ascending slope of force peaks (Fig. 7). WLC fits show that the protein length released between steps 2 and 3 is about 284 Å, which corresponds to unfolding of about 70 amino acids. This estimate is less by about 20 than the total number of residues of Ig27. Therefore, a separate initial unfolding event takes place, which results in the unraveling of about 20 amino acids (described by the fit "1" in Fig.7). More detailed study of the "hump" established that it is related to preliminary unfolding of the strand A, during which three hydrogen bonds between A and B (Fig. 2) are broken. Specific mutation K6P (replacement of lysine with proline at the position 6) disrupts these hydrogen bonds and completely eliminates the "hump".

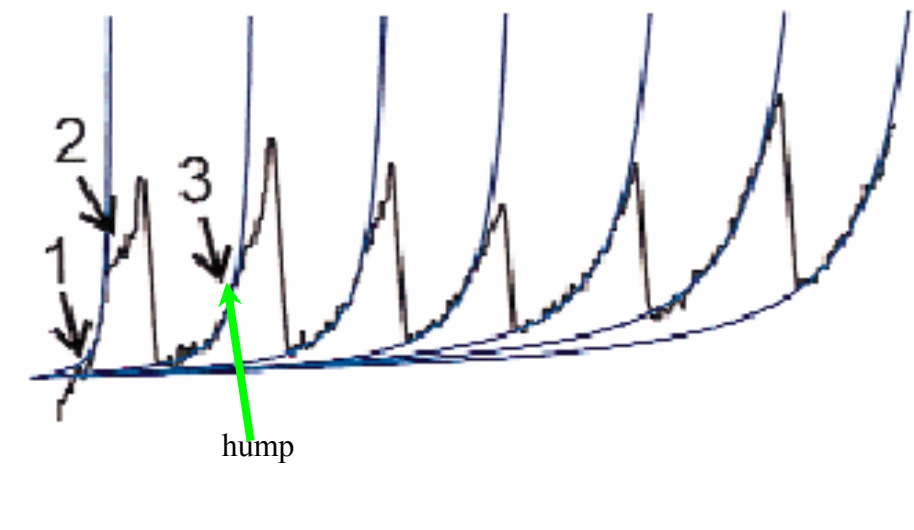

Fig. 7 Force-extension profile for Ig27 polyprotein, which registers the transient unfolding intermediate as indicated by the characteristic "hump" (*Nature Structural Biology* **7**, 719 (2000)).

III. Physics of mechanical unfolding

Simple theory of force-induced structural transitions was proposed by Bell in 1978 (*Science* **200**, 618 (1978)). For proteins the Bell's idea can be formulated as follows. Consider a constant force *f,* which acts upon protein terminal. Application of force has to be taken into account in the free energy of the protein by subtracting the term *fx,* where *x* is the length by which protein is extended. In other words, the potential energy of a protein is reduced by the work *fx* produced by the force. Therefore, the free energy $\Delta F(f) = \Delta F(0) - fx$.

The unfolding rate k_u can be written as

$$k_u = k_0 e^{-\frac{\Delta F(0) - fx}{k_B T}} = k_u\left(f = 0\right) e^{\frac{fx}{k_B T}} \qquad (2)$$

Eq. (2) shows that the unfolding rate increases upon application of force, because the free energy unfolding barrier is reduced by *fx*. Unfolding of proteins is a probabilistic event associated with crossing free energy barrier ΔF. In the absence of force ΔF is too high to allow frequent unfolding of a protein. However, external force makes the barrier smaller and, thus, increases the probability of unfolding during experiment. This is exactly the reason why it is possible to measure force-extension profiles as shown in Figs. 5-7.

Mechanical unfolding vs thermal unfolding: It is important to draw distinction between mechanical unfolding and the unfolding initiated by changing temperature (or, equivalently, denaturant concentration) (Fig. 8). Folding from thermally denatured state (denoted as RC in Fig. 8) is largely determined by the trade-off between large entropy in the unfolded state and low energy in the native state. During thermal folding protein "looses" entropy, but gains energy, i.e. attains the low energy state. The process is reversed during thermal unfolding. In this case, the energy of a protein increases, but so does the entropy. As a result protein enters the free energy minimum associated with thermally unfolded state RC.

Mechanical unfolding is conceptually different. When protein unfolds, external force drastically reduces the entropy of the unfolded state by stretching and extending a polypeptide chain. Therefore, chain entropy plays less important role in mechanical unfolding and a single reaction coordinate, such as end-to-end distance or number of native contacts, should adequately describe the unfolding process.

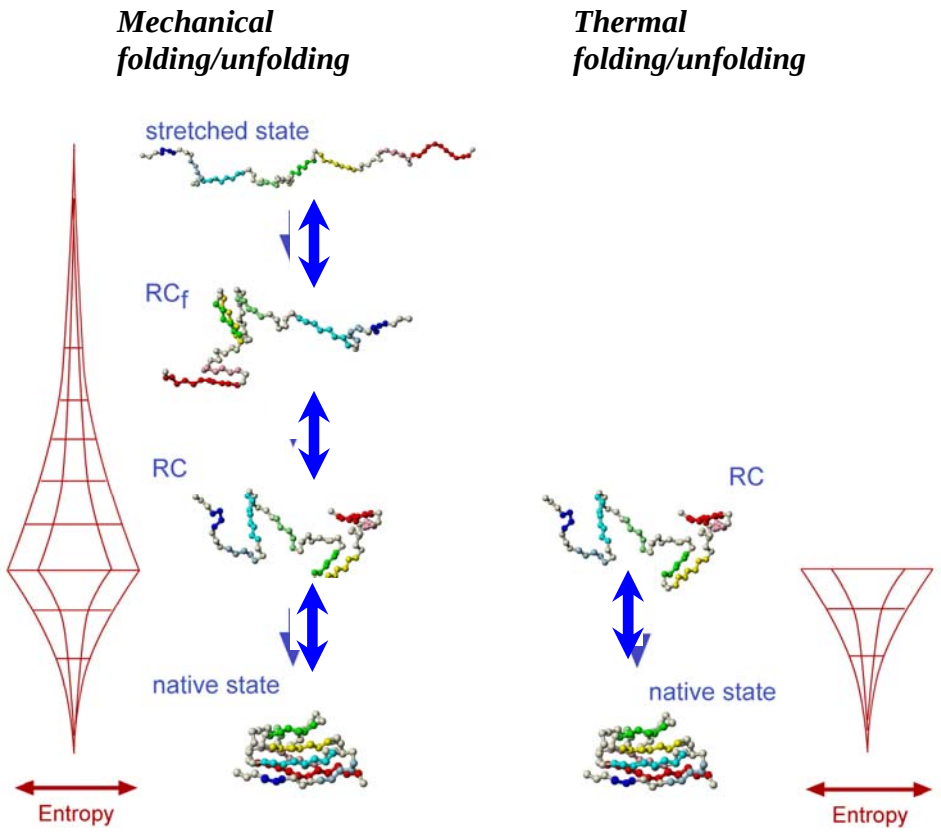

Fig. 8 The conceptual diagram of the differences between mechanical and thermal unfolding. RC_f denotes partially stretched unfolded states, which serve as intermediates in mechanical unfolding.

IV. Lattice model simulations of mechanical unfolding

General principles of mechanical unfolding of proteins may be studied using simple lattice model of proteins. Consider a 3D extension of the lattice model introduced in the Protein Folding lecture. Here we will describe this model more formally. Each amino acid is represented by a single bead confined to the vertices of the 3D lattice. In order to introduce excluded volume interactions we require that the lattice vertices cannot be occupied by the beads more than once. Assume now that the interactions between beads (apart from excluded volume) are only possible, if a pair of beads occupies nearest neighbor lattice sites. Then the potential energy of the chain is

$$E_p = \sum_{i,j} B_{ij}\delta(r_{ij} - a),$$

where a is the lattice spacing, r_{ij} is the distance between residues i and j, B_{ij} is the energy of interaction between i and j, and $\delta(0)=1$ and 0, otherwise. If force f is applied to the terminal of the protein, the total potential energy is modified is

$$E_{tot} = E_p - fr_{1N},$$

where r_{1N} is the end-to-end distance. Let us now consider how the equilibrium properties of the lattice protein change as a function of two variables, temperature T and force f. Corresponding phase diagram is presented in Fig. 9, in which protein states are color coded according to the scale on the right (*Proceedings of the National Academy of Sciences USA* **96**, 6166 (1999)). Red color represents high native content (roughly speaking, native state is formed) and blue color indicates low native content (unfolded state). It is seen that the native state is confined to the region of low temperatures and forces. Increasing the temperature (along T axis, f=0) results in temperature induced unfolding (red→blue transition). However, unfolding transition also takes place if one keeps T fixed and increases f. This is the force-induced unfolding described above. Because the green strip representing intermediate (partially folded) states disappears as f increases, the force-induced unfolding transition is sharper and narrower than the temperature induced one.

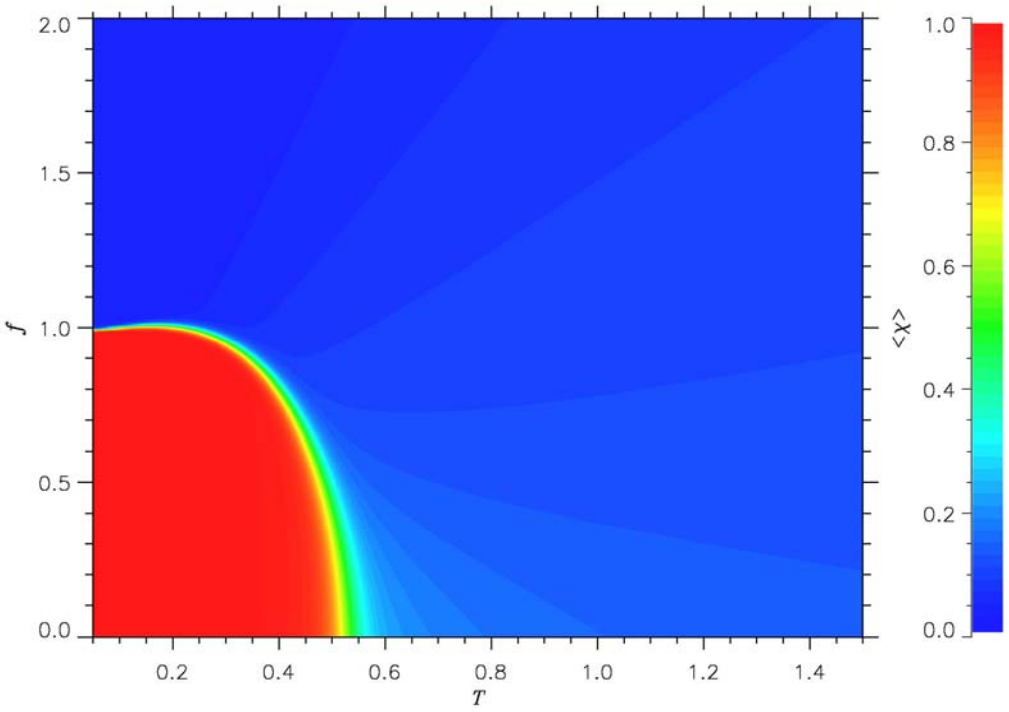

Fig. 9 Phase diagram of force-induced unfolding in the (f,T) plane (see text for details) (*Proceedings of the National Academy of Sciences USA* **96**, 6166 (1999)).

Lattice model of proteins can be used to illustrate Eq. (2). Let us plot the free energy of the chain F as a function of the number of native contacts Q, which can be served as an approximate unfolding reaction coordinate. Fig. 10 demonstrates that the free energy unfolding barrier $\Delta F^{\dagger}$ indeed decreases as f becomes larger in accord with Eq. (2). Accordingly, the unfolding rate k_u is expected to increase with f. Additional simulations directly verified this prediction. Furthermore, one can also predict that at certain force the unfolding rate k_u stops increasing and levels off.

From the unfolding simulations we can estimate the value of force f_c, at which k_u becomes maximal and then compute the extension required to unfold the protein as

$$x = \frac{\Delta F^{\pm}(f=0)}{f_{opt}}$$

The interesting result of this lattice model exercise is that the critical extension needed to unfold a protein is very small ($x \sim 0.02 L_c$, where L_c is the protein contour length). This estimate is consistent with experimental data obtained for the force-induced unfolding of Ig27.

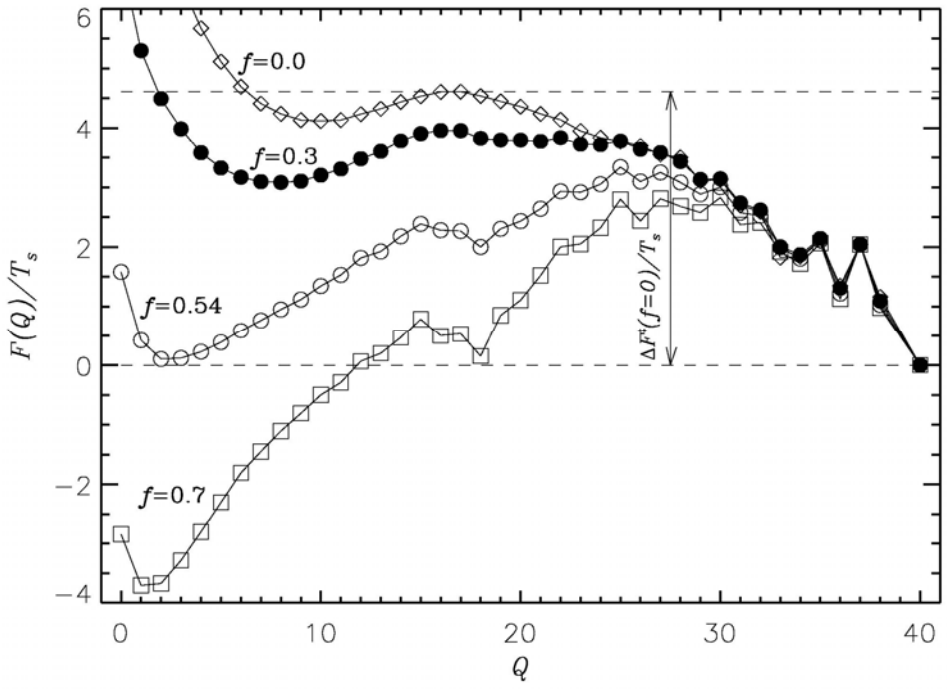

Fig. 10 Free energy as a function of the number of native contacts Q at various values of force f. Vertical double sided arrow marks the unfolding free energy barrier $\Delta F^{\dagger}$ in the absence of force (*Proceedings of the National Academy of Sciences USA* **96**, 6166 (1999)).

V. Native topology determines mechanical unfolding

It is natural to expect that a native topology plays a crucial role in mechanical unfolding. This is because spatial arrangement of the elements of the native structure (e.g., helices and strands) and their interactions should determine how these elements are unraveled and extended from the native state. The importance of the native topology has been most clearly demonstrated using coarse-grained off-lattice protein models (Fig. 11, *Proceedings of the National Academy of Sciences USA* **97**, 7254 (2000)). Similar to lattice model each amino acids is represented by a single bead (or "united atom"). To reflect amino acid diversity three types of residues are considered – hydrophobic (in blue), hydrophilic (in red), and neutral (in gray). The native structure of a β-barrel in Fig. 11 is stabilized by hydrophobic interactions. The main difference between lattice and off-lattice models is that the latter one does not confine amino acids to lattice sites and allows continuous change of residues' coordinates.

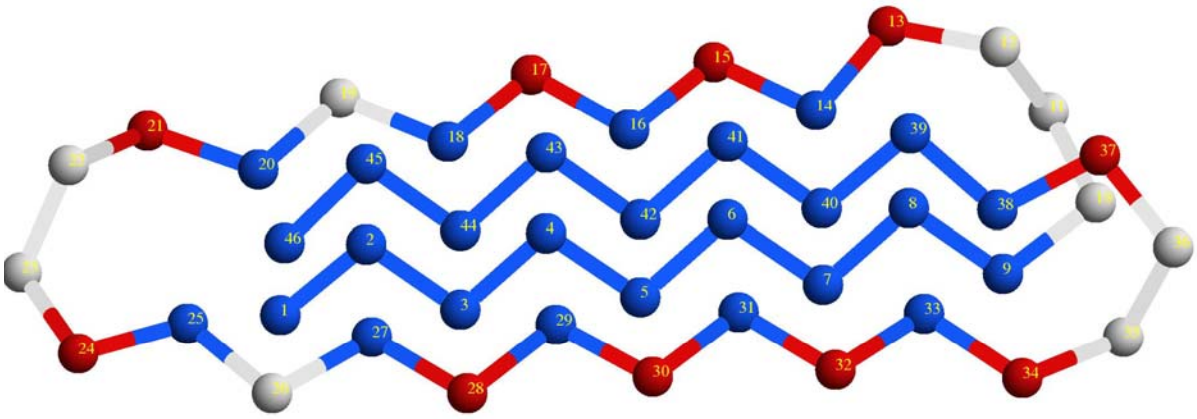

Fig. 11 Native structure of off-lattice model β-barrel.

Two different protein sequences were considered. Sequence S1 has a symmetric native state, in which both terminal β-strands (A and D) make approximately equal number of native interactions (accordingly, their energies in the native state are similar (Fig. 12a)). In the native state of S2 terminal β-strand A is buried, but the strand D is relatively exposed and is weakly coupled with the rest of the native state (Fig. 12a). Therefore, the spatial arrangement of β-strands for S2 is asymmetric.

Consider now the mechanical unfolding simulations, in which the N-terminal is fixed (tethered to the template) and the C-terminal is pulled away with the constant speed. In these simulations we record the resisting force, which is generated by the protein being stretched, as a function of the relative (with respect to equilibrium) extension $\Delta z/L$, where L is the contour length. Fig. 12b demonstrates that force-extension profiles for S1 and S2 are very different. S1 unfolds in a single step, when $\Delta z/L \approx 0.07$, and the corresponding force amplitude is 90 pN. In contrast, S2 unfolds in two steps. The first, small force peak of ≈60 pN occurs at small extensions, whereas the dominant force maximum of 120 pN is located at $\Delta z/L=0.45$.

Careful analysis of mechanical unfolding established that a singe peak for S1 corresponds to simultaneous unfolding of all β-strands. In S2 unfolding is decoupled into two stages. Small force peak is related to unraveling of the strand D. Unfolding of the strands A-C produces the second, dominant force peak at much larger extensions.

Differences in mechanical unfolding of S1 and S2 are linked to the differences in the native topology. Because S1 has a symmetric native state (Fig. 12a), all elements of the native state unfold concurrently. In S2 symmetry in the spatial arrangement of β-strands is broken and unfolding starts with the "weakest" element, which is the strand D. The rest of the S2 structure unfolds cooperatively by a sheer-type motion. Therefore, differences in the native topology find profound manifestation in the force-extension profiles.

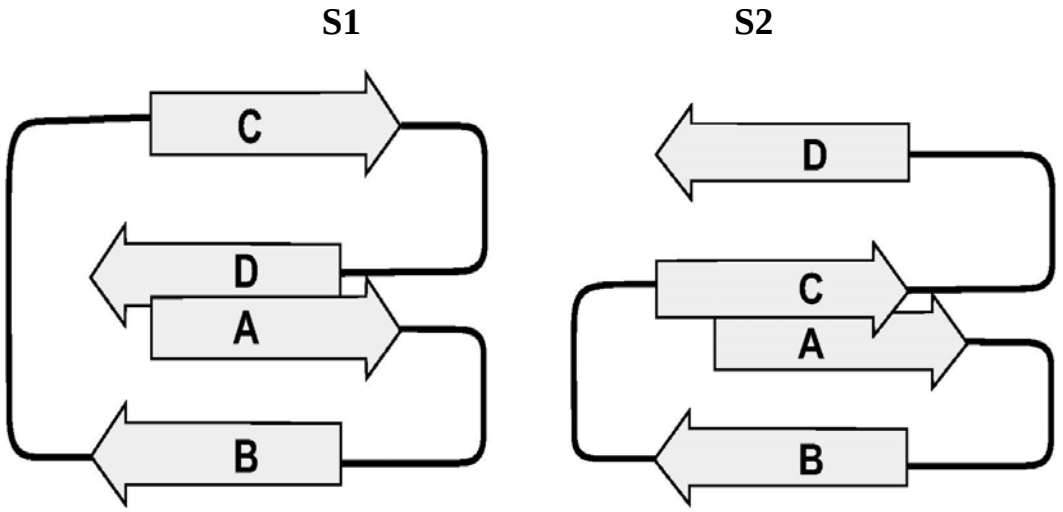

Fig. 12a Diagrams of the native structures for S1and S2.

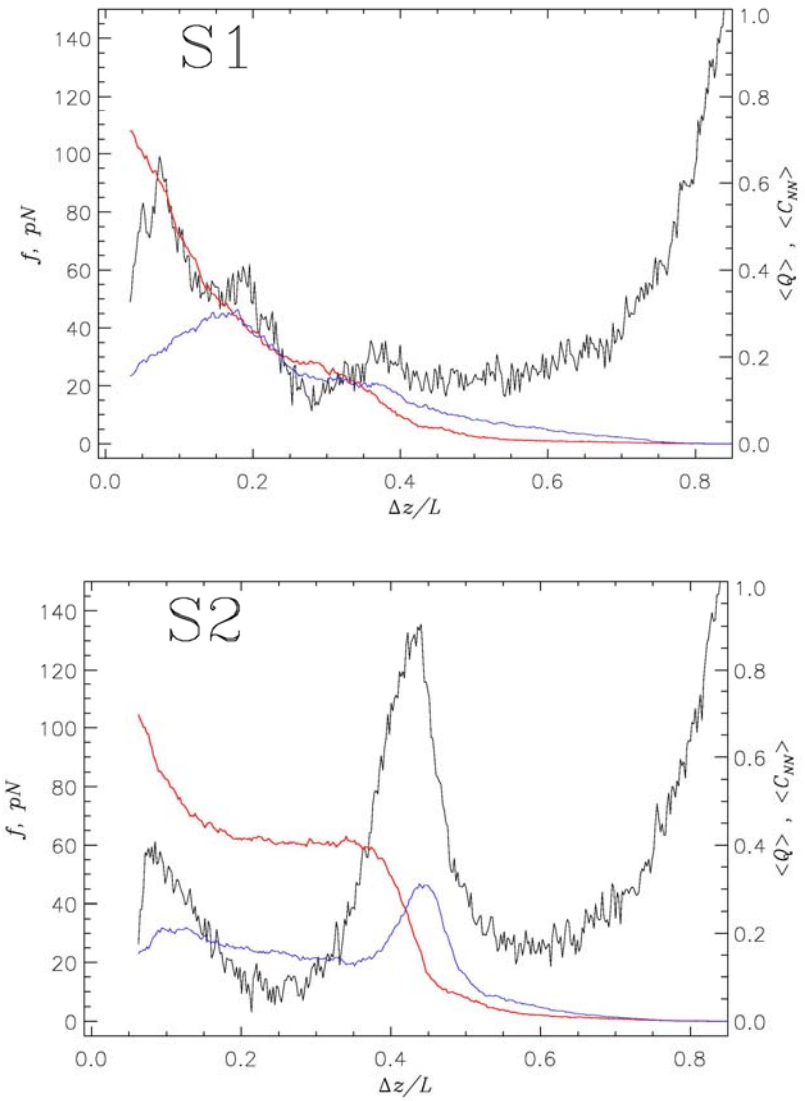

Fig. 12b Force-extension profiles for S1 and S2 (*Proceedings of the National Academy of Sciences USA* **97**, 7254 (2000)).

VI. Conclusions

Mechanical unfolding is the process of unraveling of native states of proteins as a result of application of external force. Biological functions of many proteins involve, at least, partial mechanical unfolding in the course of muscle functions, cell adhesion etc. Mechanically active proteins typically have native β-structure and are designed to withstand larger forces than those which are not typically subject to external tension. Experiments and simulations start to uncover basic characteristics of force-induced unfolding. It follows from these studies that mechanical unfolding is largely determined by the native topology.