

Experimental and Theoretical Studies of Mechanical Unfolding of Different Proteins

A. V. Glyakina¹, N. K. Balabaev¹, and O. V. Galzitskaya^{2*}

¹*Institute of Mathematical Problems of Biology, Russian Academy of Sciences,
ul. Institutskaya 4, 142290 Pushchino, Moscow Region, Russia*

²*Institute of Protein Research, Russian Academy of Sciences, ul. Institutskaya 4,
142290 Pushchino, Moscow Region, Russia; fax: 7 (496) 731-8435; E-mail: ogalzit@vega.protres.ru*

Received May 22, 2013

Revision received July 17, 2013

Abstract—Mechanical properties of proteins are important for a wide range of biological processes including cell adhesion, muscle contraction, and protein translocation across biological membranes. It is necessary to reveal how proteins achieve their required mechanical stability under natural conditions in order to understand the biological processes and also to use the knowledge for constructing new biomaterials for medical and industrial purposes. In this connection, it is important to know how a protein will behave in response to various impacts. Theoretical and experimental works on mechanical unfolding of globular proteins will be considered in detail in this review.

DOI: 10.1134/S0006297913110023

Key words: stability, protein, atomic force microscope, denaturant

Experimentally, protein unfolding by extension of their *N*- and *C*-termini has become possible due to progress in atomic force microscopy. At the end of 1986, Binning et al. constructed an atomic force microscope (AFM) that, unlike the scanning tunneling microscope, allowed studying objects with no electric current conductivity. AFM is based on registration of force interactions between the surface of the studied sample and a probe. A nano-size needle that is localized at the terminus of an elastic console called a cantilever is used for this purpose. Force that acts on the probe from the surface side bends the cantilever. The appearance of heights or cavities under the cutting edge changes the force acting on the probe and consequently changes the magnitude of bends of the cantilever. Thus, it is possible to explore the surface landscape by registration of the magnitude of the bend. First of all, Van der Waals forces are envisaged as the forces that act between the probe and the sample, attraction forces dominating at large distances and repulsion forces begin prevailing at further rapprochement of probe and sample [1-3]. AFM has become the essential instrument in biological studies because it theoretically combines two important properties that are necessary to

investigate the relationships between structure and function of biological subjects: high image resolution in the presence of large noise and the ability to operate in aqueous medium that allows observing real-time dynamics of molecular events under physiological conditions.

The main limitation of the approach that restricts the wide use of AFM is that in reality its cross resolution depends on the final sample size and the shape of the needle. In practice, particular working regimes of AFM are used for each biological object. However, the most serious problem is insufficient understanding of interactions arising between the cantilever needle and the sample [3].

The sensitivity of AFM makes this method suitable for studying mechanical properties of proteins. Usually such experiments are conducted on multidomain proteins, or a few similar (or different) proteins are “sewed together” into a chain, and then this chain is stretched out.

Before the beginning of an experiment, the chain made of proteins is allowed to be absorbed on a matrix. Then the chain is attached by one terminus and is stretched. With increasing distance between the cantilever needle and the matrix, the stretched chain of proteins generates reverse force that causes bending of the cantilever, and this, in turn, causes the deviation of a laser beam that is registered with a photodetector (Fig. 1a).

* To whom correspondence should be addressed.

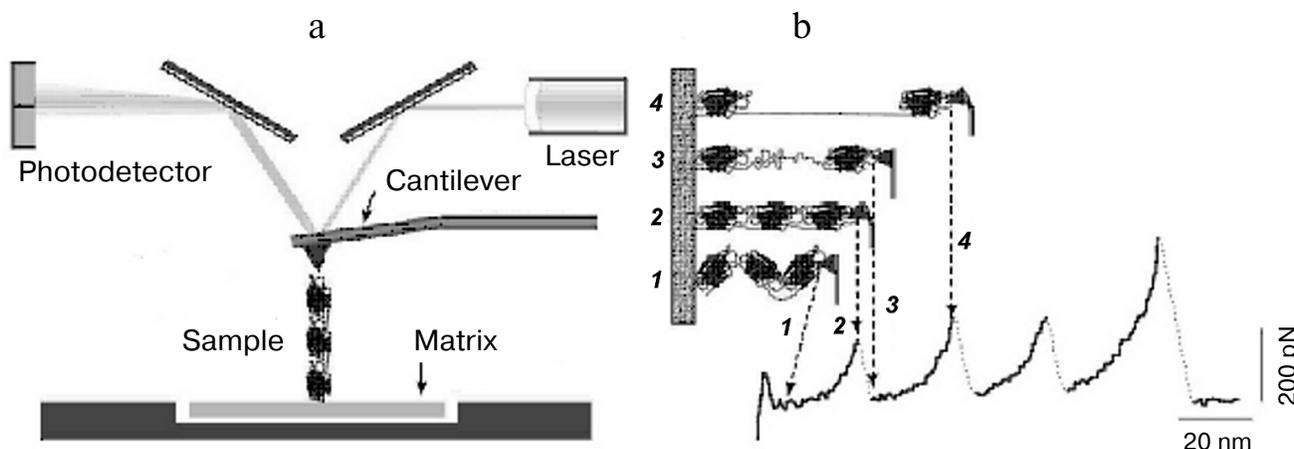


Fig. 1. a) Schematic presentation of the main elements of the atomic force microscope. b) A typical curve of dependence of force on magnitude of stretching obtained in experiments using atomic force microscopy. The figure with some changes was adapted from Fisher et al. [4].

This system can be used to perform spatial manipulations of less than one nanometer, and it can measure forces of a few piconewtons [4].

On the basis of experimental data, a profile of dependence of force on the magnitude of stretching is obtained (Fig. 1b). However, it is hard to interpret this curve: the presence of regular peaks in the curve undoubtedly indicates the multidomain structure of a protein. These peaks correspond to sequential unfolding of single domains in the studied chain. Moreover, false peaks may also arise because a protein may be fully or partly denatured due to the interaction with the matrix, or separate protein domains may accidentally interlace with each other. Such interactions are expressed in the curve in the form of a single peak or set of peaks with different non-periodic amplitudes.

To obtain a structural interpretation of such “sawtooth” dependence of a force on the magnitude of stretching, various model approaches are used in which a full atomic (or coarse-grained) model of a protein [5, 6] is stretched by its termini, and different methods to account for solvent are used (implicitly or explicitly) [7, 8].

For some proteins this dependence is relatively simple, and it shows that unfolding of a protein proceeds according to a simple two-state mechanism, when only two states of the protein are stable during protein unfolding: native and unfolded. In some cases a more complicated behavior is observed with possible existence of intermediate states in the process of mechanical unfolding [9-15].

Two regimes are used in AFM to unfold a protein: the regime of constant velocity and the regime of constant force. In the first case, the matrix is moved aside from the cantilever with constant velocity of 10^2 - 10^5 Å/s. In the second case, force measured is compared with the given value. If these two forces are different, then signal of

“error” is generated, and it is applied to the amplifier linked with the matrix. The matrix changes its position by a feedback mechanism (response time is approximately 20 ms), and thereby the “error” signal is eliminated and the constant force value is maintained [1].

INFLUENCE OF FORCE ON FREE ENERGY BARRIER FOR PROTEIN UNFOLDING

Unfolding of a protein is described by a transition state theory [16, 17]. From this theory, it follows that native and unfolded states of a protein are divided by a free energy barrier. Rates of the transition through this barrier look as:

$$k_u = v \cdot \kappa \cdot \exp(\Delta G_{TS-N}/(k_B \cdot T)), \quad (1)$$

$$k_f = v \cdot \kappa \cdot \exp(\Delta G_{TS-D}/(k_B \cdot T)), \quad (2)$$

where k_u and k_f are velocity constants for unfolding and folding of a protein, correspondingly, v is frequency of oscillations, κ is transition coefficient, ΔG_{TS-N} and ΔG_{TS-D} are activation energy of folding and unfolding, correspondingly, k_B is the Boltzmann constant, and T is absolute temperature. The value of the pre-exponential multiplying factor $v \cdot \kappa$ is unknown for protein folding, but it is sufficiently lower than those for small molecules [18].

Independently from the mode of protein extension (under regime of constant force or constant velocity), the influence of a force on a protein is interpreted by the use of the same theory (Fig. 2). In the absence of a force, the lowest energetic state of a protein is its native state. To reach an unfolded state, a protein must overcome the free energy barrier. The height of the barrier determines the

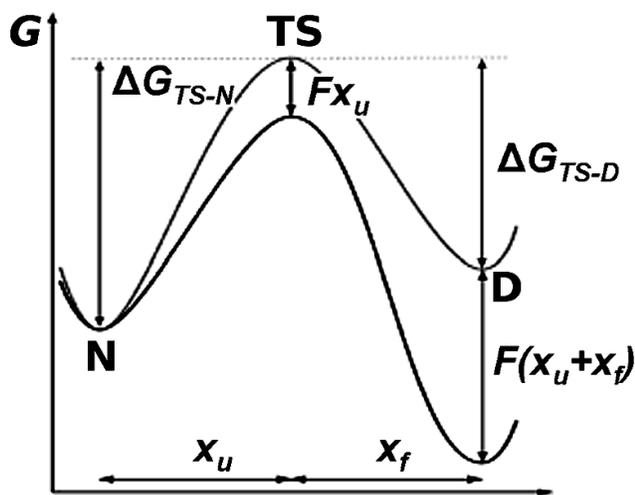


Fig. 2. Free energy barriers for protein unfolding (G) in the absence (gray curve) and in the presence (black curve) of additional external force. Under the action of the force F , the free energy barrier decreases by $F \cdot x_u$. This stabilizes the transition (TS) and unfolded (D) states, but does not influence the native (N) state. The parameters x_u and x_f are distances before the transition state along the reaction ordinate for unfolding and folding pathways, correspondingly. The figure was adapted from a work by Rounsevell et al. [1].

rate with which a protein will be unfolded due to thermal fluctuations. The free energy barrier is decreased under the action of an additional external force F . This increases the probability of protein transition to the unfolded state. Exponential growth of unfolding velocity constant under the action of external force $k_u(F)$ is described by the equation:

$$k_u(F) = v \cdot \kappa \cdot \exp((\Delta G_{TS-N} - F \cdot x_u) / (k_B \cdot T)) = k_u^0 \cdot \exp(F \cdot x_u / (k_B \cdot T)), \quad (3)$$

where k_u^0 is unfolding velocity constant under zero force, F is additional external force, and x_u is the distance to the transition state along the reaction coordinate for the unfolding pathways.

Under protein stretching in the regime of constant velocity, the force (for mechanically stable proteins) acting on the system is a function of the velocity of stretching. In case of higher rates of stretching, there is less time to study the energetic barrier of the protein. Consequently, fewer attempts are necessary to overcome the barrier, and this means that more power of folding is observed in case of high rates of stretching. Such a correlation between probable force of unfolding F and extension velocity v is described by Eq. (4) [19]:

$$F = (k_B \cdot T / x_u) \cdot \ln(x_u \cdot k_c \cdot v / (k_u^0 \cdot k_B \cdot T)), \quad (4)$$

where k_c is the rigidity constant.

Saw-tooth curves obtained experimentally can be analyzed qualitatively using models that describe elasticity of a polymer. There are two such models: a model of freely jointed chain (FJC) and a model of worm-like chain (WLC). In the FJC model a polypeptide is represented in a form of a linear row of rigid rods that can freely rotate in points of joining. In the WLC model a polypeptide is described as a semi-rigid rod (Eq. (5)). It has been shown that the WLC model approximates experimental data better than the FJC model [20].

$$F(x) = (k_B \cdot T / p) / (1 / (4 \cdot (1 - x / L_c)^2) - 1/4 + x / L_c) \quad (5)$$

Two parameters are accounted for in the WLC model: L_c – contour length (full length of unfolded polypeptide), and p – persistent length (length of the shortest rigid component of a chain). Persistent length describes flexibility of a chain. Entropy reductive force F of a polypeptide chain can be calculated at a particular extension x of a given chain. Mainly, in the context of accumulation of data on forced unfolding, the WLC model is only used as checking method. It allows comparing theoretically predicted contour lengths of a protein, which are measured in reality, as well as distances between peaks of the force. The WLC model is necessary for matching of force peaks with domain types in unfolding of chains composed of different protein domains. Additionally, the WLC model is used in Monte Carlo simulation for analysis of data, where calculation of a force under any extension is required.

ANALYSIS OF DATA ON MECHANICAL UNFOLDING OF PROTEINS

The highest peak is taken into account for determination of the unfolding force in experiments on mechanical stretching of proteins. However, there is some distribution of values of unfolding forces at each rate of stretching. So, the mean most probable unfolding force is considered as the unfolding force at any rate of stretching. Although this force is simply determined, the shape of a curve depends on the number of domains in the stretched chain, chain flexibility, and the matrix material and rigidity of the cantilever.

Analysis of data on extension of protein domains connected into one chain [21] showed that the first and the last peaks are higher than the intermediate ones. This result is based on the combination of two competitive actions. First, the probability of revealing an unfolded protein domain at a certain force decreases if a few folded domains exist. In this case, the observing force of unfolding increases. Second, the more domains become unfolded the higher is the elasticity of the system and the lower is the rate of stretching. Then the domain is able to sustain to certain force for a longer time, and the proba-

bility of overcoming the barrier is increased and a lower force of unfolding is observed.

Usually, data on all peaks of the force are taken into account to determine the force of unfolding at a particular rate of stretching. This is allowed in comparative analysis of chains that are composed of the same protein domains. However, there are some obstacles in analysis of chains composed of different protein domains, as well as in comparing experimental results for chains of different length.

The Monte Carlo method is a relatively simple computational model for interpretation of experiments on protein extension [22]. The force applied decreases the barrier for unfolding, and this is described by Eq. (3). The force can be determined at any extension using Eq. (5). Modeling starts at zero force, and initially all domains of the protein are considered to be folded, and their number is equal to the number of domains in the experimental chain.

Equation (6) describes the probability of revealing one unfolded domain from n possible domains during a given time period Δt :

$$p_u^n(F) = n \cdot k_u(F) \cdot \Delta t. \quad (6)$$

The value of Δt should be relatively low to provide unfolding probability less than 1. A generated random number r in the interval from 0 to 1 will determine the direction of the process. If $r < p_u^n(F)$, then the domain unfolds and the length of the construction is increased by the difference in length between unfolded and folded domains. Modeling is made for a set of stretching rates with various combinations of x_u and k_u^0 until the experimental dependence of the average maximum force that is necessary for protein unfolding on extension velocity will be obtained. It is necessary to note that modeling for a few combinations of x_u and k_u^0 will accurately approximate experimental data and, although the difference between x_u values may be low, the corresponding k_u^0 may differ significantly [23, 24].

RELATIONSHIPS BETWEEN PROTEIN TOPOLOGY AND MECHANICAL STABILITY

Mechanical properties of proteins are usually characterized by three parameters [25]:

1) a force under which protein unfolding takes place (in the case of stretching with constant rate), or time when the protein will unfold (in the case of stretching with constant force);

2) dependence of force value on the rate of stretching (in the case of stretching with constant rate), dependence of time when the protein will start to unfold on the value of the applied force (in the case of stretching with constant force);

3) length of a polypeptide chain.

Mechanical unfolding of proteins has been considered in numerous experimental [12, 13, 15, 26-56] and theoretical works [57-75].

Analysis of experimental and theoretical profiles of dependence of force on the distance between termini of a protein revealed that all proteins studied to the date can be divided into three large groups: the first group includes proteins that are able to sustain high force, the second middle force, and the third low force. The first and second groups mainly consist of β -structural proteins possessing parallel terminal β -strands, and the third group includes α -helical proteins. From this it follows that β -structured proteins are more mechanically stable than α -helical ones.

Using molecular dynamics method, Lu and Schulten [60] studied the mechanical unfolding of ten proteins that possess different topology and secondary structure. It has been suggested that hydrogen bonds between neighboring β -strands in β -sheet better stabilize the structure than hydrophobic contacts between helices in α -helical proteins.

The low force (less than 70 pN [26, 27, 31, 33, 47-49]) that usually is required for unfolding of α -helical proteins makes it hard to study them. So, there is still a problem to be elucidated how the mechanical stability of α -helical proteins depends on their helical content, packing of the hydrophobic core, and also the direction of the force applied.

If the protein is of α -helical type, then it is possible to say definitely that it will be mechanically unstable, but the mechanical stability of β -structural proteins varies greatly in experiments [26] as well as in modeling [60]. For example, titin (I27), a protein that demonstrates immunoglobulin-like fold, is unfolded at the force of 200 pN at the rate of stretching of 700 nm/s [40], while domains of fibronectin (FNIII), possessing a similar fold, usually are unfolded at lower forces (for example, TNfn3 – the third domain of fibronectin – unfolded at ~ 120 pN at the rate of stretching of 600 nm/s [36]) (see table). Finally, the first domain of the C2-protein synaptotagmin (C2A) that demonstrates β -sandwich is unfolded at ~ 60 pN [26]. The difference in mechanical stability may be partly explained by relative orientation of N - and C -terminal β -strands. When proteins with parallel terminal β -strands are stretched, the points of force application (N - and C -terminal amino acid residues) are localized at opposite termini of these β -strands. To stretch these proteins, it is necessary initially to destroy simultaneously all interactions (see Fig. 3a) between terminal β -strands, and after that the stable core of the protein will be decomposed. Rather great force will be required for this process. In the case of proteins having anti-parallel terminal β -strands (see Fig. 3b), the N - and C -terminal amino acid residues are localized at the same terminus of these β -strands. In this case, there is consecutive destruction of interactions that stabilize the protein and, as a consequence, there is no high force. These hypotheses are in

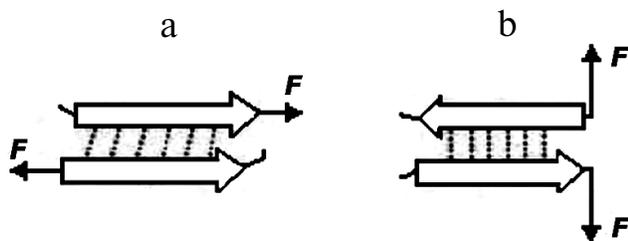


Fig. 3. Hydrogen bonding configurations between terminal β -strands: a) shift-type; b) "zip"-type configuration.

agreement with theoretical predictions and molecular models of simplified systems [76].

The large difference between mechanical stability of I27 and C2A proteins might be explained by the orientation of the terminal β -strands. However, it has been shown theoretically and experimentally that not all proteins possessing parallel terminal β -strands demonstrate the same mechanical stability. Some of them are mechanically more stable (I27, V-CAMI, cadherins), and others are less stable (FNIII-domains and V-CAMII). The difference in mechanical stability of these proteins is due to hydrogen bonding between terminal parallel β -strands in the first group. So, to unfold these proteins it is necessary to destroy all hydrogen bonds to allow structural deformation. In the second case, the proteins possess terminal parallel β -strands that are not linked by hydrogen bonds and, hence, may undergo great deformation with consequent destruction of native interactions.

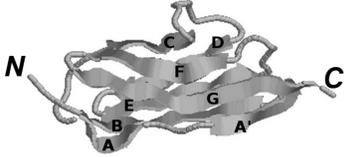
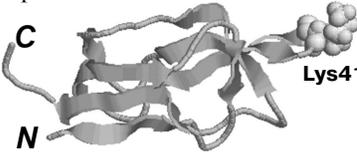
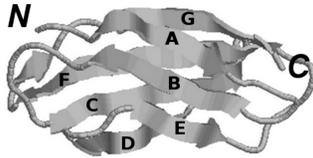
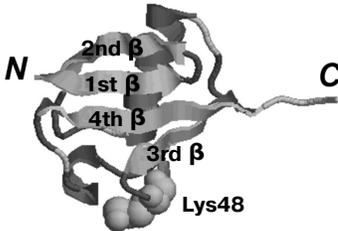
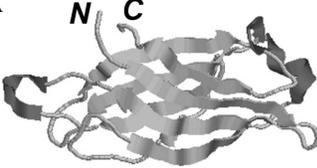
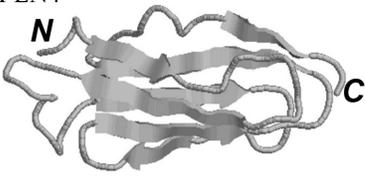
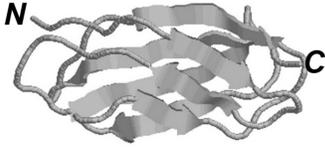
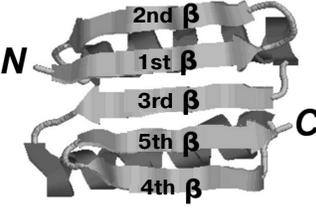
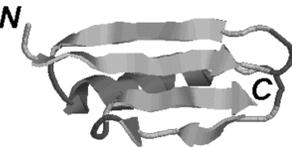
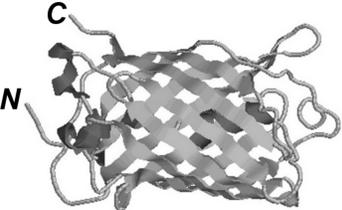
The polydomain protein titin is a protein that has been very thoroughly studied with using AFM. Experimentally, it was shown [22] that increase in contour length of approximately 10 Å in each immunoglobulin domain of titin proceeds before appearance of the main peak of force. Modeling [61, 62] helps to explain this behavior in the following way. The force is applied along the vector connecting the *N*- and *C*-termini of immunoglobulin domains of titin. Hydrogen bonds between β -strands A and B, and also A' and G are the bonds that should be disrupted first to initiate further unfolding. The topology of the immunoglobulin domains in titin determines the impossibility of force transduction along the main chain to unfold the residual part of the protein before disruption of bonds between β -strands A' and G. Differences in maximum unfolding forces for various immunoglobulin domains may appear due to the difference in hydrogen bonds between A' and G and A and B β -strands, or differences in angles between these β -strands and the *N*- and *C*-terminal β -strands that may lead to differences in components of external force necessary for disruption of hydrogen bonding. As β -strands A' and G and A and B are separated, the hydrophobic core becomes exposed to the solvent, and a far lower force is required to unfold the residual part of the domain.

The process of stretching of three proteins was modeled in a work by West et al. [63]: a topologically simple immunoglobulin-binding domain of protein L (PDB code 1hz6), a topologically complex but well studied 27-immunoglobulin-repeat of titin (I27, PDB code 1tit), and a completely α -helical protein of immunity to colicin E7 (Im7, PDB code 1ayi). These proteins were studied during modeling of their stretching on their termini under constant force as well as under constant velocity. It was shown that the trajectories cover different regions of the phase space during the two types of modeling. A number of native contacts between protein termini were determined. Under low rates of stretching, unfolding pathways resembled temperature-induced pathways. The authors wrote that the contacts between terminal β -strands were to be first broken, and only after that the mechanical unfolding of the overall protein molecule occurred. At temperatures close to the melting point of proteins, unfolding proceeds at significantly weaker forces. Linear dependence between applied force and logarithm of unfolding rate is observed in the case of mechanically stable proteins (proteins L and I27). This indicates that the same barrier is overcome at all stretching forces, i.e. the unfolding pathways are similar. For the unfolding pathways in the case of α -helical protein Im7, no dependence was observed.

Modeling of proteins L and I27 showed that there is switching of mechanisms of unfolding depending on the magnitude of the applied force. At zero force, the protein unfolds according to a mechanism through a wide pathway that is not linked with stretching on its termini. Under increasing external force, the mechanism of unfolding is switched to a different one. In the case of low forces (<50 pN), the two mechanisms coexist. Deviation of unfolding pathways becomes lower under higher forces.

Sharma and coauthors [64] constructed the Top7 protein (the table) that is composed of 92 amino acid residues and has a topology not found in nature. This protein is composed of two α -helices and β -sheet consisting of five β -strands. The protein was studied by molecular dynamics, applying to its termini a constant stretching rate. The topology of this protein is that two terminal β -strands (first and fifth β -strands), on which constant stretching rate is applied, are positioned paralleled and not linked with each other by hydrogen bonds. They are linked with each other through one more β -strand (third β -strand). In other words, this protein can be divided into three subunits. The first and third subunits are similar to each other and consist of α -helix and two β -strands. The first subunit contains the first and second β -strands, and the third subunit contains the fourth and fifth β -strands. The second subunit represents simply the third β -strand that connects the first and third subunits (see the table). There are two possible methods of breaking the β -sheet: by moving the first subunit from the second one, or by moving the second subunit from the third subunit. From this it follows: 1) the first subunit slides relative to the sec-

Average maximum force that is required for protein unfolding at given stretching velocity (experimental data)

Protein	Force, pN (Velocity, nm/s)	Protein	Force, pN (Velocity, nm/s)
127 	200 (1000) [44] 180 (600) [44]	E2lip3 	15 ± 10 (700) <i>N</i> - and <i>C</i> -termini [41] 177 ± 3 (700) <i>N</i> -terminus and Lys41 [41]
TNfn3 	120 (600) [36]	Ubiquitin 	203 (300) <i>N</i> - and <i>C</i> -termini [42] 85 ± 20 (300) Lys48 and <i>C</i> -terminus [28, 42]
C2A 	60 (600) [26]	ddFLN4 	63 (250-350) [15]
10FNIII 	74 ± 20 (600) [37]	Spectrin 	70 (1000) [27]
Top7 	155 ± 36 (400) [64]	Protein L 	136 (400) [43]
GFP 	104 (1000) [13]	Protein G 	180 (400) [55]

ond and third subunits; 2) the third subunit slides relative to the first and second subunits. Molecular dynamics simulation of this protein unfolding by its stretching on their termini with constant velocity showed that the peak force correlates with destruction of hydrogen bonds between the first and second subunits: it is easier to break hydrogen bonding between first and third β -strands than that

between the third and fifth β -strands. Comparison of hydrogen bonding between these β -strands shows that hydrogen bonding between the first and third β -strands is not weaker than those between third and fifth β -strands. From this it follows that not only hydrogen bonding, but also other interactions contribute to the mechanical stability of the protein. The authors suggest that there is

G the second β -turn links the third and fourth β -strands (*C*-terminal β -hairpin) [28, 29]. Due to analysis of mechanical unfolding pathways obtained in proteins stretching at constant velocities and application of constant forces on terminal β -strands, their similarities and differences have been revealed [69-72]. We showed that protein G is mechanically more stable than protein L, because it takes more time and higher force to unfold it (Fig. 5). However, the differences in time and force of unfolding disappear on increasing the applied force or stretching velocity. We established the existence of three pathways for unfolding of proteins L and G. In all cases, the native structures of the proteins are decomposed first into two structural units. In the first case, these structural units are represented by [*N*-hairpin + α -helix] and [*C*-hairpin], in the second case by [*N*-hairpin] and [*C*-hairpin] (α -helix is not associated with any from these blocks), and in the third case by [*N*-hairpin] and [α -helix + *C*-hairpin]. In the first case, initially the *C*-terminal β -hairpin is degraded, in the second case the α -helix, and in the third the *N*-terminal β -hairpin. The first unfolding scenario has the greatest probability (it is found in 50 and 48 cases out of 74 cases for proteins L and G, correspondingly) than the two others for both proteins.

INFLUENCE OF ORIENTATION OF SECONDARY STRUCTURE ELEMENTS IN RELATION TO DIRECTION OF APPLIED FORCE ON MECHANICAL STABILITY OF A PROTEIN

To reveal the influence of orientation of secondary structure elements in relation of direction of force application on the mechanical stability of a protein, it is necessary to unfold the same protein by its stretching in different directions. Such experiments and modeling have been performed for proteins E2lip3 and ubiquitin [41, 42, 66]. On attempting to stretch E2lip3 by the *N*-terminus and Lys41, the protein was relatively resistant to stretching. To unfold it, a force of 177 pN at extension velocity of 700 nm/s was required. However, this protein was mechanically unstable (unfolding force during stretching of the protein was less than 15 pN) when the stretching was performed by the *N*- and *C*-termini. Molecular dynamics simulations qualitatively confirmed the experimental data and also showed that unfolding of the protein proceeds through different pathways for its stretching in different directions [41].

Stretching of a protein by applying force on different termini was also performed in the case of ubiquitin. A force of 203 pN at extension velocity 300 nm/s is required to stretch ubiquitin by its *N*- and *C*-termini, while a force of 85 pN at the same extension velocity is necessary to unfold this protein by its Lys48 and *C*-terminus. Also, it was revealed that the stretching of this pro-

tein by its *N*- and *C*-termini is more sensitive to extension velocity, and this means that transition state, which appears at stretching in this direction, looks more native-like than at stretching in another direction (by Lys48 and the *C*-terminus). Distance value x_u at which unfolding starts is approximately 0.25 and 0.63 nm in the case of stretching by the *N*- and *C*-termini and by Lys48 and the *C*-terminus, correspondingly. The experimental data are in agreement with molecular dynamics simulations [42, 66].

The above-mentioned observations indicate that the free energy landscape of proteins is very anisotropic.

INFLUENCE OF INTERACTIONS OF SIDE CHAINS OF AMINO ACID RESIDUES ON MECHANICAL STABILITY OF A PROTEIN

Despite the importance of the type and arrangement of secondary structure elements, it is obvious that there are other factors influencing the mechanical stability of a protein because proteins with similar topologies have different mechanical properties. For example, the immunoglobulin-binding domains of proteins L and G and also ubiquitin have similar topology, and they unfold at forces of 145, 180, and 210 pN at stretching velocity 400 nm/s, correspondingly [28, 42, 43, 50]. Comparison of the mechanical properties and topological characteristics of proteins L and ubiquitin was performed in a work by Brockwell et al. [43]. Distances x_u , at which these proteins begin to unfold, are quite similar (0.22 nm for protein L [43] and 0.24 nm for ubiquitin [28, 42]), and this means that the proteins unfold following similar pathways. However, ubiquitin has fewer hydrogen bonds between the terminal β -strands in comparison with protein L, but it is more mechanically stable. Molecular dynamics simulations showed that these two proteins unfold by shifting of two separate "blocks". One "block" consists of α -helix and the first and second β -strands, and another consists of the third and fourth β -strands. Comparison of native contact maps for each protein showed that although hydrophobic cores of the two proteins are approximately the same size, the number of long distance influencing interactions linking these two "blocks" with each other is significantly lower in protein L. Although it is necessary to stretch these proteins by the same distance to unfold the two proteins, a greater force is required to achieve this distance in the case of ubiquitin because it is necessary to disrupt, first, the interactions of side chains of amino acid residues that link these separate "blocks" in order to provide further transduction of the tension. From this it follows that mechanical stability of a protein depends not only on orientation of secondary structure elements in relation to direction of force application, but also on native interactions in the protein as a whole.

INFLUENCE OF VARIOUS FACTORS ON MECHANICAL STABILITY OF A PROTEIN

The influence of a denaturant on mechanical stability of a protein was studied by Cao and Li [51]. It was shown by experiment that for protein G under increasing denaturant concentration a systematic reduction of its mechanical stability was observed. This means that the free energy barrier between folded and unfolded states of the protein was reduced. Decreasing mechanical stability depends linearly on increasing denaturant concentration. However, the denaturant does not alter the mechanical unfolding pathways and does not shift the transition state position. Experiments using AFM in combination with protein engineering showed that chemical (or thermal) and forced unfolding proceeds through different pathways, and in the case of forced unfolding the overcome free energy barrier is not unique and may change depending on the applied force or by introduction of specific mutations [36, 52].

Solvent molecules also play an important role in conformational dynamics of proteins. It was shown experimentally that glycerol stabilizes the native state of ubiquitin, making it resistant to mechanical unfolding [56].

Calculated equilibrium profiles of free energy as a function of distance for protein CI2 and protein G showed that at different pH values intermediate states appear on the way of forced unfolding of protein CI2. The position of two transition states is shifted to the unstable state under increasing force value. Unfolding of G protein proceeds according to a simple two-state mechanism. The position of the transition state in relation to the folded state does not depend on the applied force [74].

Modeling of protein I27 stretching with constant velocity and constant force at various temperatures showed that the average maximum force and average time of unfolding linearly decrease with increase in temperature. It was shown that the energetic landscape of I27 unfolding does not depend on temperature until the transition state is reached. Competitive pathways in unfolding of this protein appear after passing the transition state [75].

INTERMEDIATE STATES ON MECHANICAL UNFOLDING PATHWAY OF A PROTEIN

It was shown experimentally that topologically complex proteins (10FNIII – 10th domain of fibronectin [11], and GFP – green fluorescent protein [12, 13]) as well as some topologically simple proteins (spectrin [9] and ubiquitin [10]) are unfolded via multi-state mechanisms. The existence of intermediate states on the pathway of mechanical unfolding of proteins is important not only for understanding their functions *in vivo*, but also

because the properties of these experimentally revealed states may be used for more detailed investigation of unfolding pathways than would be possible in the case of two-state proteins. Many of these intermediate states are metastable: their detection became possible only due to development of new techniques. For example, it was found that ubiquitin unfolds according to a two-state mechanism during its stretching under the regime of constant force [10], but not under the regime of constant velocity [28, 42], and from this it follows that ubiquitin may undergo mechanical unfolding through a high energy transition state that is weakly occupied. The ability of AFM to detect metastable states depends on their lifetimes. This is seen in comparison of two intermediate states that arise during mechanical unfolding of GFP [13]. One intermediate state is seemingly formed by detachment of seven amino acid residues at the *N*-terminus of the α -helix, and it is detected as an additional peak on the curve of force dependence on distance between the protein termini. The second intermediate state that is formed after removal of the *N*- or *C*-terminal β -strand from the β -barrel is very unstable, and its presence can be revealed only by the use of special analysis.

In contrast to GFP protein, titin (I27) is unfolded through a stable intermediate state that is a native-like state (only the β -strand A is moved from the core) [44].

It was shown that two other proteins having the same fold as I27 (10FNIII [11] and ddFLN4 [15]) also unfold through intermediate states. However, these domains are mechanically less stable in comparison with I27 (their native and intermediate states unfold at forces of 63 and 53 pN and 90 and 50 pN for ddFLN4 and 10FNIII, correspondingly), and their intermediate states are more unfolded. In any case, these proteins have two mechanically stable states that may facilitate the process of reverse folding, when proteins were subjected to stress during normal functioning of a cell.

The folding of protein G (PDB code 1igd) was modeled by the Monte Carlo method using G \ddot{o} -potentials in a work by Shimada and Shakhnovich [77]. It was shown that unfolding proceeds through an intermediate state. They revealed three possible pathways for unfolding of this protein. The first pathway (the most frequent) passes through intermediate state α -helix–*N*-hairpin, the second one through α -helix–*C*-hairpin, and the third through β -sheet formation. The second and third pathways are seen less often than the first pathway. Although all three pathways pass different intermediate states, they eventually meet in one key moment – the formation of a specific core that includes amino acid residues participating in all secondary structure elements (α -helix and *N*- and *C*-terminal β -hairpins).

In works by Ng et al. and Best et al. [36, 52] experimental and theoretical Φ values obtained during unfolding of proteins by a denaturant and forced unfolding and also from molecular dynamics simulations of forced

unfolding of protein were compared for proteins TNfn3 and I27. They investigated structures of transition states (quick unfolding of the protein proceeds after passage of these structures). Both proteins unfold through the transition state, and unfolding proceeds in a similar manner. Theoretical Φ values are calculated as the ratio of native contact number for amino acid residues in the transition state to the number of contacts for the amino acid residues in the initial structure. Native structure of the protein taken as the initial structure in the work by Ng et al. [36] and the structure of the intermediate state taken from the work by Best et al. [52] were used to obtain theoretical and experimental Φ values.

It was shown experimentally that unfolding of I27 requires greater force than unfolding of TNfn3. A key moment in the forced unfolding of protein I27 is breakage of hydrogen bonds and hydrophobic interactions between β -strands A' and G, but the hydrophobic core of the protein is untouched. Reconstruction of the overall hydrophobic core is observed in protein TNfn3 in addition to loss of interactions between β -strands A and G.

It was revealed that Φ values obtained experimentally and theoretically on mechanical unfolding coincide within the error limits. In both proteins the transition states that are observed during mechanical unfolding differ from transition states observed under unfolding by denaturant.

Geierhaas et al. [78] compared the structures of the transition states of TNfn3 and I27. In this work, they obtained structures that compile an ensemble of transition states using molecular dynamics simulations and experimental Φ values as restrictions, in contrast to the works by Ng et al. and Best et al. [36, 52]. The method uses molecular dynamics calculations that are restricted by a pseudo-energetic function including a series of experimentally obtained Φ values. With these limitations, the ensemble of transition states becomes the most stable state on the surface of the potential energy of a protein. It is more stable than the unfolded state, as occurs in the case of real energetic functions. That is, an additional member is introduced into potentials in molecular dynamics simulations:

$$U = \alpha \cdot \rho^2 / 2, \quad (7)$$

where α is force constant, ρ is root mean square deviation between experimental and theoretical Φ values, and $\rho(t) = (1/N_\Phi) \cdot \sum (\Phi_i^{\text{th}}(t) - \Phi_i^{\text{exp}})^2$ summation is made for all amino acid residues for which experimental Φ values are available; N_Φ is the number of experimental Φ values.

All possible structures that satisfy these limitations are simulated. They do not necessarily represent structures in the transition state under protein unfolding. However, experimentally obtained Φ values are not sufficient to prepare these structures, especially when the number of tested mutations in the proteins is low, i.e. a

small number of Φ values is known. In this case, another experimental value m that is linked with the solvent accessible surface area is commonly used to restrict molecular dynamics or during subsequent selection of the most valuable structures. Proteins TNfn3 and I27 have similar structures (immunoglobulin-like fold), but they differ greatly in their amino acid sequence. Geierhaas et al. [78] consider formation of interactions in which amino acid residues from the folding nucleus participate to be very important for maintenance of topology of immunoglobulin-like fold. Despite the fact that folding nuclei in these proteins are quite similar, the folding nucleus in protein I27 is significantly denser than in TNfn3, and this is expressed at much higher average Φ values for I27 in comparison with TNfn3. They compared structures of native states of these proteins with the structures comprising the ensemble of transition states for these proteins using such parameters as root mean square deviation (RMSD), solvent accessible surface area (SASA), and radius of gyration. It was shown that structures comprising the ensemble of transition states of protein TNfn3 are more unfolded than in the case of I27. The RMSD in the case of native structure for TNfn3 was 7 Å, and for I27 it was 4 Å. The increase in SASA for TNfn3 was 3-fold greater than for I27 (30 versus 9%). The radius of gyration of I27 is not significantly changed, but radius of gyration of TNfn3 increases by 8%. On the basis on these results, the authors concluded that folding nuclei can be sufficiently deformed for folding of different sequences into the same structure. Because the transition state of protein TNfn3 is more unfolded than of protein I27, it was suggested that the first protein may be mechanically less stable than the second. This suggestion is confirmed by the experimental data (the table).

It has been shown in our laboratory that short-lived intermediate states appear on mechanical unfolding pathways of proteins L and G [69-72]. Moreover, it was demonstrated that in protein L and in protein G the most mechanically stable element of the secondary structure is *N*-terminal β -hairpin. It was shown in experiments on unfolding of these two proteins by a denaturant that the folding nucleus of protein L contains *N*-hairpin and the folding nucleus of protein G contains *C*-hairpin. Theoretically, calculated Φ values of proteins L and G unfolding under the impact of a force suggest the presence of the most stable element of the secondary structure, namely *N*-terminal β -hairpin, in folding nuclei of both proteins. The experimental data from the unfolding of these proteins under the influence of denaturant and the theoretical data obtained from modeling of these proteins subjected to force coincide in the case of protein L and do not coincide in the case of protein G.

Thus, it has been shown experimentally using AFM during stretching of proteins that β -structural proteins sustain more significant loads than α -helical or α/β

structural proteins. It is considered that the arrangement of protein secondary structure elements represents the critical factor in determining its mechanical and thermal stability. Thus, a few hydrogen bonds between neighboring β -strands in β -sheet stabilize the structure significantly more than the same number of hydrophobic contacts between helices in α -helical proteins. However, there is no direct correlation between the protein structure and its mechanical and thermal stability. Even small changes in amino acid sequence can significantly influence the mechanical properties of a protein and its thermal stability. The direction of application of force is also a critical factor in determination of the mechanical stability. Detailed investigation of pathways of forced unfolding of a protein is necessary for understanding the molecular nature of mechanical stability of proteins. Computer methods of molecular dynamics simulation allow studying processes of mechanical unfolding of globular proteins at the atomic level. New knowledge may be used in future for development of macromolecules with desired mechanical properties and their further usage as structural blocks for constructing new materials for industrial and medical purposes.

This work was conducted with financial support of the Russian Foundation for Basic Research (grant 11-04-00763), the Presidium of the Russian Academy of Sciences (programs "Molecular and Cellular Biology" (01201353567) and "Fundamental Sciences to Medicine"), and the Ministry of Education and Science of the Republic of Kazakhstan.

REFERENCES

- Rounsevell, R., Forman, J. R., and Clarke, J. (2004) *Methods*, **34**, 100-111.
- Bornschiogl, T., and Rief, M. (2011) *Methods Mol. Biol.*, **783**, 233-250.
- Serdyuk, I. N., Zaccai, N. R., and Zaccai, J. (2007) *Methods in Molecular Biophysics: Structure, Dynamics, Function*, Cambridge University Press, New York.
- Fisher, T. E., Oberhauser, A. F., Carrion-Vazquez, M., Marszalek, P. E., and Fernandez, J. M. (1999) *TIBS*, **24**, 379-384.
- Tozzini, V. (2005) *Curr. Opin. Struct. Biol.*, **15**, 144-150.
- Cavalli, A., Vendruscolo, M., and Paci, E. (2005) *Biophys. J.*, **88**, 3158-3166.
- Lazaridis, T., and Karplus, M. (1999) *Proteins*, **35**, 133-152.
- Roux, B., and Simonson, T. (1999) *Biophys. Chem.*, **78**, 1-20.
- Altmann, S. M., Grunberg, R. G., Lenne, P. F., Ylanne, J., Raae, A., Herbert, K., Saraste, M., Nilges, M., and Horber, J. K. (2002) *Structure*, **10**, 1085-1096.
- Schlierf, M., Li, H., and Fernandez, J. M. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 7299-7304.
- Li, L., Huang, H. H., Badilla, C. L., and Fernandez, J. M. (2005) *J. Mol. Biol.*, **345**, 817-826.
- Abu-Lail, N. I., Ohashi, T., Clark, R. L., Erickson, H. P., and Zauscher, S. (2006) *Matrix Biol.*, **25**, 175-184.
- Dietz, H., and Rief, M. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 16192-16197.
- Hertadi, R., Gruswitz, F., Silver, L., Koide, A., Koide, S., Arakawa, H., and Ikai, A. (2003) *J. Mol. Biol.*, **333**, 993-1002.
- Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A., and Rief, M. (2004) *Nature Struct. Mol. Biol.*, **11**, 81-85.
- Eyring, H. (1935) *J. Chem. Phys.*, **3**, 107-115.
- Laidler, K., and King, C. (1983) *J. Phys. Chem.*, **87**, 2657-2664.
- Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 11615-11617.
- Evans, E., and Ritchie, K. (1997) *Biophys. J.*, **72**, 1541-1555.
- Bustamante, C., Erie, D. A., and Keller, D. (1994) *Curr. Opin. Struct. Biol.*, **4**, 750-760.
- Zinober, R. C., Brockwell, D. J., Beddard, G. S., Blake, A. W., Olmsted, P. D., Radford, S. E., and Smith, D. A. (2002) *Protein Sci.*, **11**, 2759-2765.
- Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. (1997) *Science*, **276**, 1109-1112.
- Best, R. B., Brockwell, D. J., Toca-Herrera, J. L., Blake, A. W., Smith, D. A., Radford, S. E., and Clarke, J. (2003) *Anal. Chim. Acta*, **479**, 87-105.
- Best, R. B., Fowler, S. B., Toca-Herrera, J. L., and Clarke, J. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 12143-12148.
- Brockwell, D. J. (2007) *Curr. Nanosci.*, **3**, 3-15.
- Carrion-Vazquez, M., Oberhauser, A. F., Fisher, T. E., Marszalek, P. E., Li, H., and Fernandez, J. M. (2000) *Prog. Biophys. Mol. Biol.*, **74**, 63-91.
- Lenne, P. F., Raae, A. J., Altmann, S. M., Saraste, M., and Horber, J. K. H. (2000) *FEBS Lett.*, **476**, 124-128.
- Chyan, C. L., Lin, F. C., Peng, H., Yuan, J. M., Chang, C. H., Lin, S. H., and Yang, G. (2004) *Biophys. J.*, **87**, 3995-4006.
- Bertoncini, P., Schoenauer, R., Agarkova, I., Hegner, M., Perriard, J. C., and Guntherodt, H. J. (2005) *J. Mol. Biol.*, **34**, 1127-1137.
- Schoenauer, R., Bertoncini, P., Machaidze, G., Aebi, U., Perriard, J. C., Hegner, M., and Agarkova, I. (2005) *J. Mol. Biol.*, **34**, 367-379.
- Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V., and Marszalek, P. E. (2006) *Nature*, **440**, 246-249.
- Forman, J. R., Qamar, S., Paci, E., Sandford, R. N., and Clarke, J. (2005) *J. Mol. Biol.*, **349**, 861-871.
- Li, L., Wetzel, S., Pluckthun, A., and Fernandez, J. M. (2006) *Biophys. J.*, **90**, L30-L32.
- Junker, J. P., Hell, K., Schlierf, M., Neupert, W., and Rief, M. (2005) *Biophys. J.*, **89**, L46-L48.
- Ainavarapu, S. R., Li, L., Badilla, C. L., and Fernandez, J. M. (2005) *Biophys. J.*, **89**, 3337-3344.
- Ng, S. P., Rounsevell, R. W. S., Steward, A., Geierhaas, Ch. D., Williams, Ph. M., Paci, E., and Clarke, J. (2005) *J. Mol. Biol.*, **350**, 776-789.
- Oberhauser, A. F., Badilla-Fernandez, C., Carrion-Vazquez, M., and Fernandez, J. M. (2002) *J. Mol. Biol.*, **319**, 433-447.
- Li, H., Linke, W. A., Oberhauser, A. F., Carrion-Vazquez, M., Kerkvliet, J. G., Lu, H., Marszalek, P. E., and Fernandez, J. M. (2002) *Nature*, **418**, 998-1002.

39. Wilcox, A. J., Choy, J., Bustamante, C., and Matouschek, A. (2005) *Proc. Natl. Acad. Sci. USA*, **102**, 15435-15440.
40. Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., and Fernandez, J. M. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 3694-3699.
41. Brockwell, D. J., Paci, E., Zinober, R. C., Beddard, G. S., Olmsted, P. D., Smith, D. A., Perham, R. N., and Radford, S. E. (2003) *Nature Struct. Biol.*, **10**, 731-737.
42. Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F., and Fernandez, J. M. (2003) *Nature Struct. Biol.*, **10**, 738-743.
43. Brockwell, D. J., Beddard, G. S., Paci, E., West, D. K., Olmsted, P. D., Smith, D. A., and Radford, S. E. (2005) *Biophys. J.*, **89**, 506-519.
44. Fowler, S. B., Best, R. B., Toca-Herrera, J. L., Rutherford, T. J., Steward, A., Paci, E., Karplus, M., and Clarke, J. (2002) *J. Mol. Biol.*, **322**, 841-849.
45. Best, R. B., Li, B., Steward, A., Daggett, V., and Clarke, J. (2001) *Biophys. J.*, **81**, 2344-2356.
46. Li, H., and Fernandez, J. M. (2003) *J. Mol. Biol.*, **334**, 75-86.
47. Law, R., Carl, P., Harper, S., Dalhaimer, P., Speicher, D. W., and Discher, D. E. (2003) *Biophys. J.*, **84**, 533-544.
48. Law, R., Liao, G., Harper, S., Yang, G., Speicher, D. W., and Discher, D. E. (2003) *Biophys. J.*, **85**, 3286-3293.
49. Bhasin, N., Law, R., Liao, G., Safer, D., Ellmer, J., Discher, B. M., Sweeney, H. L., and Discher, D. E. (2005) *J. Mol. Biol.*, **352**, 795-806.
50. Cao, Y., Lam, C., Wang, M., and Li, H. (2006) *Angew. Chem. Int. Ed. Engl.*, **45**, 642-645.
51. Cao, Y., and Li, H. (2008) *J. Mol. Biol.*, **375**, 316-324.
52. Best, R. B., Fowler, S. B., Herrera, J., Steward, A., Paci, E., and Clarke, J. (2003) *J. Mol. Biol.*, **330**, 867-877.
53. Jagannathana, B., Elmsa, P. J., Bustamante, C., and Marqusee, S. (2012) *Proc. Natl. Acad. Sci. USA*, **109**, 17820-17825.
54. Aggarwal, V., Kulothungan, S. R., Balamurali, M. M., Saranya, S. R., Varadarajan, R., and Ainaravapu, S. R. (2011) *J. Biol. Chem.*, **286**, 28056-28065.
55. Cao, Y., and Li, H. (2007) *Nature Mater.*, **6**, 109-114.
56. Garcia-Manyes, S., Dougan, L., and Fernandez, J. M. (2009) *Proc. Natl. Acad. Sci. USA*, **106**, 10540-10545.
57. Paci, E., and Karplus, M. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 6521-6526.
58. West, D. K., Brockwell, D. J., Olmsted, P. D., Radford, S. E., and Paci, E. (2006) *Biophys. J.*, **90**, 287-297.
59. West, D. K., Olmsted, P. D., and Paci, E. (2006) *J. Chem. Phys.*, **125**, 204910-204917.
60. Lu, H., and Schulten, K. (1999) *Proteins*, **35**, 453-463.
61. Lu, H., Isralewitz, B., Krammer, A., Vogel, V., and Schulten, K. (1998) *Biophys. J.*, **75**, 662-671.
62. Lu, H., and Schulten, K. (2000) *Biophys. J.*, **79**, 51-65.
63. West, D. K., Olmsted, P. D., and Paci, E. (2006) *J. Chem. Phys.*, **124**, 154909.
64. Sharma, D., Perisic, O., Peng, Q., Cao, Y., Lam, C., Lu, H., and Li, H. (2007) *Proc. Natl. Acad. Sci. USA*, **104**, 9278-9283.
65. Sułkowska, J. I., and Cieplak, M. (2008) *Biophys. J.*, **94**, 6-13.
66. Li, P., and Makarov, D. E. (2004) *J. Chem. Phys.*, **121**, 4826-4832.
67. Mitternacht, S., Luccioli, S., Torcini, A., Imparato, A., and Irback, A. (2009) *Biophys. J.*, **96**, 429-441.
68. Caraglio, M., Imparato, A., and Pelizzola, A. (2011) *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.*, **84**, 021918.
69. Glyakina, A. V., Balabaev, N. K., and Galzitskaya, O. V. (2009) *Biochemistry (Moscow)*, **74**, 316-328.
70. Glyakina, A. V., Balabaev, N. K., and Galzitskaya, O. V. (2009) *J. Chem. Phys.*, **131**, 045102.
71. Glyakina, A. V., Balabaev, N. K., and Galzitskaya, O. V. (2009) *Open Biochem. J.*, **3**, 66-77.
72. Glyakina, A. V., Balabaev, N. K., and Galzitskaya, O. V. (2010) *Protein Pept. Lett.*, **17**, 92-103.
73. Kumar, S., and Li, M. S. (2010) *Physics Reports*, **486**, 1-74.
74. O'Brien, E. P., Brooks, B. R., and Thirumalai, D. (2012) *J. Am. Chem. Soc.*, **134**, 979-987.
75. Bung, N., and Priyakumar, U. D. (2012) *J. Mol. Model.*, **18**, 2823-2829.
76. Rohs, R., Etchebest, C., and Lavery, R. (1999) *Biophys. J.*, **76**, 2760-2768.
77. Shimada, J., and Shakhnovich, E. I. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 11175-11180.
78. Geierhaas, C. D., Paci, E., Vendruscolo, M., and Clarke, J. (2004) *J. Mol. Biol.*, **343**, 1111-1123.