ISSN 0006-2979, Biochemistry (Moscow), 2009, Vol. 74, No. 3, pp. 316-328. © Pleiades Publishing, Ltd., 2009.
Original Russian Text © A. V. Glyakina, N. K. Balabaev, O. V. Galzitskaya, 2009, published in Biokhimiya, 2009, Vol. 74, No. 3, pp. 389-403.

Comparison of Transition States Obtained upon Modeling of Unfolding of Immunoglobulin-Binding Domains of Proteins L and G Caused by External Action with Transition States Obtained in the Absence of Force Probed by Experiments

A. V. Glyakina[1], N. K. Balabaev[1], 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: (495) 632-7871; E-mail: ogalzit@vega.protres.ru*

Received March 31, 2008
Revision received June 2, 2008

Abstract—We have studied the extent of coincidence of the pathway of unfolding of protein globules upon experimental modeling of protein unfolding caused by external actions and denaturants. To this end, we compared experimental Φ-values reported in the literature and Φ-values obtained by us upon modeling of unfolding of immunoglobulin-binding domains of proteins L and G caused by external actions at a constant rate. A comparison of the results of calculation with the experimental data shows that the folding pathways for protein L coincide, while those for protein G do not coincide despite structural similarity of these proteins.

DOI: 10.1134/S0006297909030110

Key words: protein structure, molecular dynamics, folding pathway, mechanical unfolding, atom–atomic contacts, folding nucleus, transition state

Current experimental approaches to the determination of folding nuclei in proteins include Φ-analysis [1] and Ψ-analysis [2]. The folding nucleus of a protein is a structural part of the transition state (TS) of the molecule which appears during the folding, i.e. the most unstable state of the protein chain upon its folding. Both approaches are based on the determination of amino acid residues (a.a.) whose changes significantly affect the protein folding rate, altering the stability of the transition state as much as the stability of the native protein. The main difference between the approaches is that the first considers the involvement of amino acid residues in the nucleus, while the second considers the involvement of individual contacts between them.

According to the model of native folding nucleus [1], $\Phi = 1$ shows that this residue is included in the folding nucleus, whereas $\Phi = 0$ shows that it is not included in the folding nucleus. Interpretation of $\Phi \approx 0.5$ is ambiguous: in this case, the residue either is situated at the surface of the folding nucleus, or there are several pathways (and hence several nuclei) of unfolding and the residue is included in one of such alternative nuclei.

As experimental determination of folding nuclei is rather laborious, it may prove expedient to try to predict the TS structure. Prediction of amino acid residues is essential for the formation of folding nuclei would make it possible to outline the most probable folding pathway and, most importantly, to detect structural elements that constrain the process of folding of a protein molecule.

Several approaches for predicting folding nuclei in proteins have been proposed. One of them involves calculation of the all-atom molecular dynamic simulations of protein unfolding [3-6], though this calculation is applicable only to very small proteins. Moreover, it requires simulation of extreme denaturation conditions (500 K, etc.), so that it could be completed within a reasonable amount of time. Therefore, a TS found by such an extreme unfolding may differ greatly from that actually characteristic of a protein during folding [7]: according to the "detailed equilibrium" principle, the pathways of

Abbreviations: a.a., amino acid residues; TS, transition state.
* To whom correspondence should be addressed.

direct and reverse reactions must be the same only when both reactions occur under identical conditions [8]. It should be noted, however, that recent experiments on the molecular dynamic simulations of unfolding of some proteins (consisting of less than 40 a.a.) have been performed at 350 K with the use of state-of-the-art supercomputers [9].

Other methods for predicting folding nucleus structures are based on consideration of a simplified network of protein folding/unfolding pathways [10-12]. These simplified methods (the results of which are, nevertheless, consistent with empirical data [13, 14]) can be applied to the search for folding nuclei in single-domain proteins of various sizes.

The experimental data on the folding nucleus structure show that proteins similar in three-dimensional (3D) structure have, as a rule, similar folding nuclei [15-17]. However there are several exceptions, indicating that folding pathways are sensitive to some features of the amino acid sequence [15, 16]. It was shown that the location of folding nuclei is different in proteins with the same topology, which points to the sensitivity of the folding pathway to the details of the amino acid sequence. Immunoglobulin-binding domains of proteins L and G are structural homologs, but have little sequence similarity. The helix is packed across a four-stranded β-sheet in these proteins. Of interest is the experimental fact that the symmetry of the given topology fully breaks under folding of these proteins (Fig. 1). Thus, the folding nuclei of proteins L and G include *N*- and *C*-hairpins, respectively [18, 19]. Such a result can be explained by the existence of more favorable contacts in the second β-hairpin of protein G. Indeed, the isolated fragment corresponding to the second β-hairpin is stable in aqueous solution [20]. Therefore, the protein region, which was highly capable of forming local structures in the unfolded state, can play an important role in the stabilization of an ensemble of TS. Experimental data for other proteins [21-24] also show that local characteristics of the sequence may be important for choosing the specific pathway of folding.

On the basis of these data it was concluded that topology determines possible folding pathways, and details of packing and orientation of structural elements play an important role in selection of a certain folding pathway. The experimental data on the structure of TS do not permit formulating accurate topological rules determining the position and the structure of the folding nucleus [25]. Specific mutations and global variations in the amino acid sequence of a protein can change its folding pathways without modification of its 3D structure.

Another important question, which arises from these observations, is whether the unfolding pathways of proteins under the action of the external force and denaturant are the same. Simulations of stretching proteins under the external forces allow observing changes of the protein structure in response to mechanical deformation. Mechanical unfolding of two proteins, which have similar structures, is considered in this work. And the question arises how much the unfolding pathways are changed under the action of mechanical force. Some proteins in the cell experience a mechanical effect; therefore, studies of this question are of practical value.

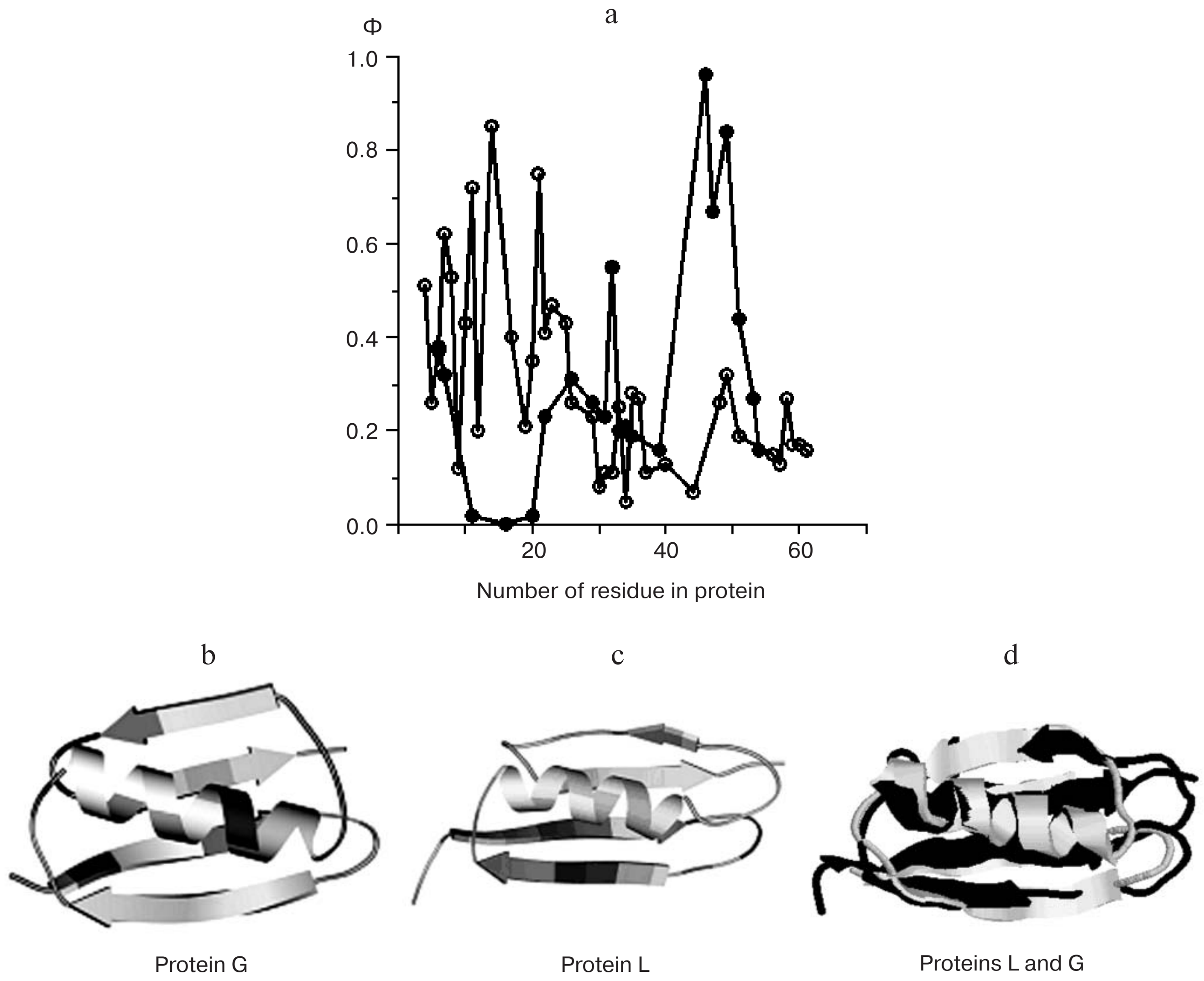
At present, experimental and theoretical Φ-values, obtained from the unfolding caused by force and denaturant and also from simulations of the force unfolding using the method of molecular dynamics have been compared only for two proteins—TI I27 (the 27th immunoglobulin repeat of titin [26]) and TNfn3 (the third domain of type III fibronectin from human tenascin [27]). The authors investigated structures representing a transition state ensemble. These structures were considered to precede the fast unfolding of protein. Both these proteins unfold via an intermediate, and their unfolding occurs in the same way. At that, theoretical Φ-values were calculated as the ratio of the number of native contacts of the amino acid residue in TS to the number of contacts of this residue in the initial structure. To obtain both theoretical and experimental Φ-values, the authors of paper [27] considered the protein native structure as the initial one, whereas the authors of paper [26] analyzed the intermediate structure in the same role.

Experimentally, it was shown that the unfolding of TI I27 requires a larger force applied than for the unfolding of TNfn3. The key event under the force unfolding of TI I27 is the disruption of hydrogen bonds and hydrophobic interactions between β-strands A′ and G. The hydrophobic core of the protein remains intact. As for TNfn3 protein, in addition to the loss of interactions between β-strands A and G, the hydrophobic core of the protein is also rearranged. It turns out that the Φ-values obtained from the theoretical and experimental studies of force unfolding of these proteins (TI I27 and TNfn3) are the same within the error. For both proteins, the transition states observed at force unfolding and denaturation are different [26, 27].

In this work, Φ-values were obtained from the extension of immunoglobulin-binding domains of proteins L and G with constant extension velocities, and compared with experimental Φ-values obtained from the unfolding of these proteins by denaturant [18, 19].

METHODS OF INVESTIGATION

Model and method of modeling. The objects of the study are two immunoglobulin-binding domains of proteins L and G (below, for short, protein L and protein G, respectively). Protein L (the name of the file in the Protein Data Bank is 2ptl) consists of 63 a.a. (residues from 16 to 78 in accordance with the sequence) and includes 951 atoms. Protein G (the name of the file in the Protein Data Bank is 1pgb) consists of 56 a.a. (853



e

Protein L	E	E	**V**	**T**	**I**	**K**	**A**	**N**	**L**	**I**	**F**	A	N	G	S				**T**	**Q**	**T**	**A**
Protein G			**M**	**T**	**Y**	**K**	**L**	**I**	**L**	**N**	**G**					K	T	L	**K**	**G**	**E**	**T**

Protein L	**E**	**F**	**K**	**G**	**T**	**F**	**E**		**K**	**A**	**T**	**S**	**E**	**A**	**Y**	**A**	**Y**	**A**	**D**	**T**	**L**	**K**
Protein G	**T**	**T**	**E**	**A**	**V**	**D**	**A**	A	**T**	**A**	**E**	**K**	**V**	**F**	**K**	**Q**	**Y**	**A**	N	**D**	**N**	**G**

Protein L	K	D	N	G		**E**	**Y**	**T**	**V**	**D**	**V**	**A**	**D**	K	G	Y				**T**	**L**	**N**
Protein G					V	**D**	**G**	**E**	**W**	**T**	**Y**	**D**	**D**				A	T	K	**T**	**F**	**T**

Protein L	**I**	**K**	**F**	A	G
Protein G	**V**	**T**	**E**		

Fig. 1. a) Profiles of Φ-values obtained experimentally for proteins G and L. Closed and open circles are investigated residues of proteins G and L, correspondingly. b, c) Schemes of spatial structures of these proteins shaded according to Φ-values of their amino acid residues from white (Φ = 0) to black (Φ = 1). d) Spatial alignment of proteins L (black) and G (gray). Root-mean-square deviation for aligned Cα atoms of these proteins is 1.38 Å. e) Alignment of primary structures of proteins L and G (aligned regions are shown in bold type).

atoms). Both these proteins consist of two β-hairpins located at the termini (*N*- and *C*-hairpins) and an α-helix between them (Fig. 1). These proteins have similar three-dimensional structures (the root-mean-square deviation of aligned Cα atoms is 1.38 Å), but they differ in amino acid sequences (the identity is about 15%).

The study has been carried out with the help of the method of molecular dynamics using the program PUMA developed at the Institute of Mathematical Problems of Biology, Russian Academy of Sciences. The solution of the system of classical motion equations of atoms has been made in the all-atom force field AMBER-99 [28].

Newton's equations are as follows:

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i\left(\vec{r}_1, ..., \vec{r}_n\right), \tag{1}$$

$$\vec{F}_i = -\frac{\partial U\left(\vec{r}_1, ..., \vec{r}_n\right)}{\partial \vec{r}_i} + F_i^{\text{external}}, \tag{2}$$

where $i = 1, ..., n$; m_i is the mass of the particle; $\vec{r}_i$ is the radius-vector of the particle; $\vec{F}_i$ is the total force acting on the atom from the direction of other particles; $U(\vec{r}_1, ..., \vec{r}_n)$ is the potential energy depending on the positional relationship of all atoms.

The potential energy is represented as the sum of terms:

$$U = U_{\text{v.b.}} + U_{\text{v.a.}} + U_{\text{t.a.}} + U_{\text{p.g.}} + U_{\text{VdW}} + U_{\text{q}}. \tag{3}$$

These terms have different functional view:

$$U_{\text{v.b.}} = \sum_i K_{li}\left(l_i - l_0\right)^2, \tag{4}$$

$$U_{\text{v.a.}} = \sum_i K_{\theta i}\left(\theta_i - \theta_0\right)^2, \tag{5}$$

$$U_{\text{t.a.}} = \sum_i K_{\varphi i}\left[1 \pm \cos(n_i \varphi)\right], \tag{6}$$

$$U_{\text{p.g.}} = \sum_i K_{\varphi i}\left[1 - \cos(2\varphi)\right], \tag{7}$$

where l_0, θ_0 are corresponding equilibrium values of the length of the valence bond (v.b.) and valence angle (v.a.); l_i, θ_i are their current values; t.a. are torsion angles; p.g. are plane groups. K_l, K_θ, K_φ are force constants, which are taken from quantum-chemical calculations. Constants K_l, K_θ, K_φ and l_0, θ_0, n depend on the type of atoms.

Interactions between atoms without valence bonds are described by Lennard–Jones potential:

$$U_{\text{VdW}} = \sum_i \sum_j 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] W_{\text{VdW}}\left(r_{ij}\right), \tag{8}$$

$$W_{\text{VdW}}\left(r_{ij}\right) = \begin{cases} 1, r_{ij} < R_{\text{on}}, \\ \dfrac{\left(R_{\text{off}}^2 - r_{ij}^2\right)^2\left(R_{\text{off}}^2 - 3R_{\text{on}}^2 + 2r_{ij}^2\right)}{\left(R_{\text{off}}^2 - r_{ij}^2\right)^3}, R_{\text{on}} < r_{ij} < R_{\text{off}}, \\ 1, r_{ij} \geq R_{\text{off}}, \end{cases} \tag{9}$$

where U_{VdW} is the energy of Van-der-Waals interactions (Lennard–Jones potential); $W_{\text{VdW}}(r_{ij})$ is the switching function with parameters R_{on} and R_{off}; σ_{ij}, ε_{ij} are parameters of Lennard–Jones potential for the pair of particles *i* and *j*. They are calculated using the combinational rules:

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \text{ and } \sigma_{ij} = (\sigma_i + \sigma_j)/2.$$

These parameters determine the depth of the potential well and the location of its minimum.

Interactions between charged atoms are described by the electrostatic potential:

$$U_{\text{q}} = \sum_i \sum_j \frac{q_i q_j}{\varepsilon r_{ij}} W_{\text{q}}\left(r_{ij}\right), \tag{10}$$

$$W_{\text{q}}\left(r_{ij}\right) = \begin{cases} \left(1 - \dfrac{r_{ij}}{R_{\text{q}}}\right)^2, r_{ij} < R_{\text{q}}, \\ 0, r_{ij} > R_{\text{q}}, \end{cases} \tag{11}$$

where U_{q} is the electrostatic energy; ε is the dielectric constant of medium; $W_{\text{q}}(r_{ij})$ is the screening function with radius R_{q}; r_{ij} is the distance between particles *i* and *j*, which are not valence bonded. In this case, $R_{\text{off}} = R_{\text{q}} = 10.5$ Å.

A model of TIP3P was used for the water, bonds and angles being not fixed, but set by appropriate potential functions. To maintain constant temperature, virtual collision environment was used ("collision thermostat") [29, 30]. The mean collision frequency of atoms with virtual particles was 10 psec^{-1}, the mass of each virtual particle was 1 atomic mass unit. Equations of motion were integrated numerically by using the velocity version of the Verlet algorithm [31] with a time step 0.001 psec.

Initial coordinates of protein atoms were taken from the Protein Data Bank. The proteins were enclosed into a parallelepiped filled with water molecules. The water molecules which sterically overlapped with the protein atoms were removed. Thus, we obtained 1772 water molecules for protein L and 1844 water molecules for protein

G. The whole system (protein + water) was enclosed into a sufficiently large sphere-cylinder with impervious repulsive walls. This sphere-cylinder did not affect the dynamics of protein and water, and at the same time, did not allow water molecules go to infinity, returning them into the modeling region. During the preparation of initial data for the first time, random velocities were assigned to all atoms and relaxation was carried out with fixed terminal atoms of the proteins. For each protein, a series of 10 such calculations in the independent random collision environments were made. These relaxed during 50 psec systems served as initial systems for the following simulations, in which an additional external action increasing the distance between terminal atoms of proteins (in protein L terminal atoms are 1N and 945Cα, and in protein G – atoms 1N and 839Cα) was used.

Totally, 40 independent simulations, which differ in initial data (coordinates and velocities), were performed. Two extension velocity values were taken: $\upsilon = 0.125$ and 0.0625 Å/psec. Simulations were carried out at the temperature of 350 K. On the whole, 20 trajectories of unfolding for each protein with constant velocity were obtained.

The maximal length of the trajectories did not exceed 2000 and 3000 psec under the extension velocities $\upsilon = 0.125$ and 0.0625 Å/psec, respectively. By that time, in the proteins there were no amino acid residues, which had α-helical or β-structural conformations.

The number of contacts between elements of secondary structure and their changes during the force unfolding were analyzed. We calculated atom–atom and residue–residue contacts. Two residues have a contact if the nearest pair of their heavy atoms is at distance < 5 Å. The calculation of the number of atom–atom contacts per residue in protein was carried out in the following way: two atoms were considered in contact with each other if their centers were at a distance of < 5 Å. The residue–residue and atom–atom contacts between two adjacent residues as well as within one residue were not taken into account. The destruction of secondary structure in the proteins was analyzed as well. The secondary structure was defined in sequential moments of time using the program DSSP (Definition Secondary Structure of Proteins) [32].

Calculation of folding nuclei. The experimental Φ-value for residue *r* was determined as in [1]:

$$\Phi = \frac{\Delta_r[F(\mathrm{TS}) - F(\mathrm{D})]}{\Delta_r[F(\mathrm{N}) - F(\mathrm{D})]}\ , \qquad (12)$$

where $\Delta_r[F(\mathrm{N}) - F(\mathrm{D})]$ is the change in the difference of free energies between native (N) and denatured states (D) induced by mutation of residue *r*, and $\Delta_r[F(\mathrm{TS}) - F(\mathrm{D})]$ is the change in the difference of free energies between transition (TS) and denatured states (D) induced by the same mutation.

Experimental Φ-values indicate to what extent the residue and its environment in TS are "native" [1, 33-35]. $\Phi = 1$ indicates that both the residue structure and environment are native in the transition state. $\Phi = 0$ indicates that the residue in the transition state has neither its own native structure nor its native environment. Intermediate Φ-values are usually interpreted as evidence that the environment of the residue is native only partially.

To calculate Φ-values, structures, which compose the ensemble of transition states, were selected from the trajectories obtained. The structures from the region of the maximal force [27] and not lower than half of the force peak height were selected. Thus, different number of points falls into this region on different trajectories. For protein L, 45 structures were selected (on average four structures from each trajectory) at $\upsilon = 0.125$ Å/psec and 47 structures at $\upsilon = 0.0625$ Å/psec (on average four structures from each trajectory). For protein G, 36 structures were selected (on average three structures from each trajectory) at $\upsilon = 0.125$ Å/psec and 65 structures at $\upsilon = 0.0625$ Å/psec (on average six structures from each trajectory). For every amino acid residue from the selected structures, theoretical Φ-values were calculated. The Φ-values were calculated in two ways with consideration of all contacts and only native contacts:

$$\Phi_{\mathrm{th,i}} = \frac{N_i^{\#,\mathrm{nat}}}{N_i^{\mathrm{nat}}}\ ,\quad \widetilde{\Phi}_{\mathrm{th,i}} = \frac{N_i^{\#,\mathrm{all}}}{N_i^{\mathrm{nat}}}\ , \qquad (13)$$

where $N_i^{\#,\mathrm{all}}$ is the number of all atomic contacts the *i*-th amino acid residue has in TS; $N_i^{\#,\mathrm{nat}}$ is the number of native atomic contacts the *i*-th amino acid residue has in TS; N_i^{nat} is the number of contacts the *i*-th amino acid residue has in the native (initial) structure.

The calculated theoretical $\Phi_{\mathrm{th,\ i}}$ and $\widetilde{\Phi}_{\mathrm{th,\ i}}$ values can be compared with experimental $\Phi_{\mathrm{exp,\ i}}$ values and their correlation can be obtained. Since, rarely occurring values $\Phi_{\mathrm{exp,\ r}} < 0$ and $\Phi_{\mathrm{exp,\ r}} > 1$ have no structural interpretation [35], there is a low number of mutations for which experimental values $\Phi_{\mathrm{exp,\ r}} < 0$ and $\Phi_{\mathrm{exp,\ r}} > 1$ are excluded from the comparison with theoretical calculations.

Lists of examined mutations for proteins L and G are given in Table 1.

RESULTS AND DISCUSSION

A set of transition states for unfolding of proteins under the action of external extension velocity. It was shown experimentally that protein L, ubiquitin, and TI I27 are mechanically stable proteins, because their extension by the termini requires force exceeding 100 pN at extension velocity more than 100 nm/sec [36-38]. The data obtained from simulations of extension of different proteins show that the location of β-strands and their ori-

Table 1. Experimental Φ-values (Φ_{exp}) obtained at unfolding of proteins L and G by denaturant and theoretical Φ-values (Φ_{th}) calculated from the modeling of unfolding of proteins L and G under external forces

Mutation	Φ_{exp}	Φ_{th} (averaged over 10 trajectories)			
		only native contacts, υ (Å/psec)		all contacts, υ (Å/psec)	
		0.125	0.0625	0.125	0.0625
1	2	3	4	5	6
		Protein L			
V4A	0.51	0.77	0.70	0.99	0.95
T5A	0.26	0.67	0.59	0.77	0.70
I6A	0.37	0.72	0.69	0.81	0.78
K7A	0.62	0.84	0.83	0.97	0.96
A8G	0.53	0.77	0.75	0.91	0.89
N9A	0.12	0.73	0.73	0.90	0.92
L10A	0.43	0.78	0.79	0.99	1.00
I11A	0.72	1.00	1.00	1.04	1.09
F12A	0.20	0.80	0.79	0.86	0.85
N14A	0.85	1.00	1.00	1.64	1.64
T17A	0.40	1.00	1.00	1.20	1.28
T19A	0.21	0.95	0.91	0.95	0.93
A20G	0.35	0.72	0.66	0.72	0.66
E21A	0.75	0.75	0.72	0.78	0.73
F22A	0.41	0.56	0.56	0.62	0.62
K23A	0.47	0.63	0.61	0.69	0.71
T25A	0.43	0.63	0.62	0.88	0.82
F26G	0.26	0.52	0.44	0.68	0.53
A29G	0.23	0.83	0.80	0.88	0.88
T30A	0.08	0.62	0.61	0.80	0.77
S31G	0.11	0.81	0.78	0.91	0.84
E32G	0.11	0.80	0.72	0.93	0.81
A33G	0.25	0.79	0.71	0.93	0.80
Y34A	0.05	0.52	0.45	0.79	0.77
A35G	0.28	0.70	0.67	0.72	0.73
Y36A	0.27	0.77	0.69	1.02	0.92
A37G	0.11	0.70	0.62	0.83	0.79
L40A	0.13	0.68	0.68	0.73	0.74
N44A	0.07	0.71	0.73	0.81	0.86
T48A	0.26	0.80	0.80	0.80	0.83
V49A	0.32	0.71	0.67	0.80	0.77
V51A	0.19	0.87	0.87	1.04	1.05
Y56A	0.15	0.58	0.53	0.69	0.63
T57A	0.13	0.86	0.84	0.96	0.94
L58A	0.27	0.75	0.73	0.97	0.98

Table 1. (Contd.)

1	2	3	4	5	6
N59A	0.17	0.69	0.70	0.84	0.87
I60A	0.17	0.68	0.65	0.82	0.79
K61A	0.16	0.72	0.71	0.72	0.71
Protein G					
I6A	0.38	0.99	0.99	1.12	1.07
L7A	0.32	0.97	0.97	1.25	1.25
T11A	0.02	0.00	0.00	0.56	1.70
T16A	0.00	0.96	0.97	1.00	1.01
A20G	0.02	0.86	0.88	0.94	1.01
D22A	0.23	0.71	0.72	0.72	0.72
A26G	0.31	1.00	1.00	1.11	1.10
V29A	0.26	0.91	0.93	1.04	1.12
K31G	0.23	0.98	0.98	1.23	1.23
Q32G	0.55	0.99	0.98	0.99	1.98
Y33A	0.20	0.93	0.95	1.08	1.14
A34G	0.21	0.94	0.97	0.97	1.00
N35G	0.19	0.92	0.94	0.99	1.08
V39A	0.16	0.85	0.89	0.91	0.95
D46A	0.96	1.00	1.00	1.08	1.04
D47A	0.67	0.67	0.66	0.67	0.66
T49A	0.84	0.97	0.96	1.01	1.05
T51A	0.44	0.94	0.93	1.09	1.06
T53A	0.27	0.98	1.00	1.14	1.13
V54A	0.16	0.91	0.92	0.95	0.96

entation relative to the force vector are important for the mechanical stability of proteins [36, 39, 40]. The experimental studies of force unfolding of protein L [36] have demonstrated that this protein unfolds by a two-state mechanism without an intermediate.

In our simulations for extension of proteins by the termini with constant velocity, we traced the dependence of force on extension (the force applied to the 1N and 945Cα atoms in protein L, and to the 1N and 839Cα atoms in protein G). The curves are shown in Fig. 2. These forces are averaged over a short time interval (in this case 5 psec) and over the values for two termini. From Fig. 2 we can see that the extension velocity does not influence the profile of the unfolding force from extension. In Table 2, there are characteristics of both proteins—maximal force and extension. Results in Table 2 are averaged over the set of trajectories related to each variant.

From Table 2 we can see that the differences in the average maximal forces for proteins L and G are not significant. At that time, there is a tendency to lowering of the force barrier due to decreasing the extension velocity (for both proteins) and a displacement of the barrier position towards lager extensions (for protein L). It is known that at low extension velocities the pathways of force and temperature unfolding are the same [41].

Figure 3 shows the dependence of the force and number of contacts between the first and fourth β-strands on the time. In both proteins (L and G), the first force peak (it is observed for all trajectories) correlates with a dramatically decreasing number of contacts between the first and fourth β-strands. A general analysis of the trajectories for proteins L and G revealed that the order of disappearance of contacts between β-strands is the same for both proteins: first between the first and fourth β-strands, then between the third and fourth β-strands (*C*-hairpin), and finally between the first and second β-strands (*N*-hairpin).

In protein L in most cases, the disruption of contacts between the first and fourth β-strands coincides with the detachment of the *C*-hairpin from the rest structure; and

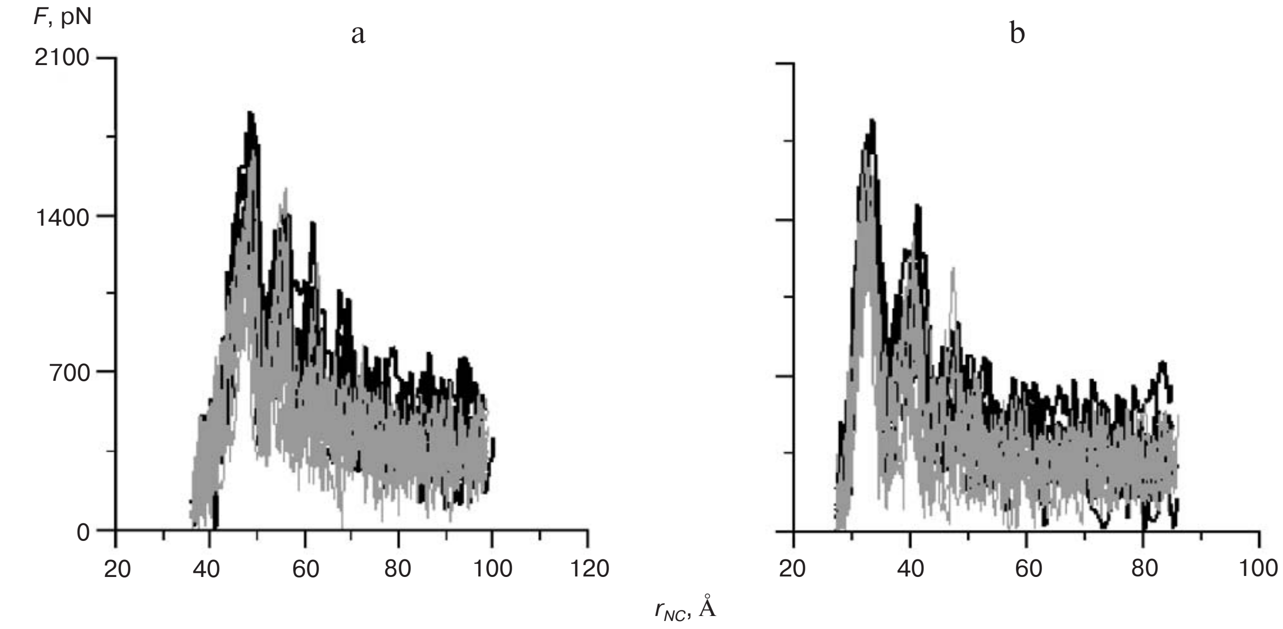


Fig. 2. Change of force F acting on the termini of the protein in points of its fixation, depending on distance r_{NC} between them: a, b) proteins L and G, correspondingly. Black and gray curves, $\upsilon = 0.125$ and 0.0625 Å/psec, correspondingly. Ten curves for each extension velocity are represented.

at first the *C*-hairpin completely separates from the structure and then begins to be destroyed. In this protein, in all except two (in 18 of 20) cases at first the *C*-hairpin is completely destroyed, and then the *N*-hairpin. In protein L in most cases the order of destruction of secondary structure elements is as follows: first the α-helix breaks, then the *C*-hairpin, and finally the *N*-hairpin. A similar situation is observed in the work [41]: at first the contacts between terminal β-strands break, and only then the rest of the protein unfolds mechanically. In TS, there is disruption between two structural units: *N*-hairpin + α-helix and *C*-hairpin. In protein G in most cases, at first there is disruption of contacts between the first and fourth β-strands, then the *C*-hairpin separates from the rest of the structure and is destroyed concurrently. In protein G in all 20 cases, at first the *C*-hairpin breaks and then the *N*-hairpin, the same as in protein L. But in protein G another order of secondary structure destruction is observed: at first the *C*-hairpin breaks, then the α-helix, and finally the *N*-hairpin.

In addition to study of the changing of contacts between the different elements of secondary structure, the destruction of secondary structure in the proteins was analyzed. The secondary structure was defined in sequential moments of time using the program DSSP [32]. It was found that the order of destruction of secondary structure elements in proteins L and G is different. As a rule in protein L, first the α-helix breaks, then the C-hairpin, and finally the N-hairpin. In protein G another order is observed: first the C-hairpin breaks, then the α-helix, and finally the N-hairpin.

Calculation of unfolding nuclei and their comparison with known experimental data on Φ-values. From the analysis of protein structures it can be suggested that the

Table 2. Maximal force $\langle F_{max} \rangle$ and corresponding increase of distance $\langle r^t_{NC} - r^0_{NC} \rangle$ between the *N*- and *C*-termini for proteins at different extension velocities (averaged over sets of trajectories)

Characteristic	Protein L, υ (Å/psec)		Protein G, υ (Å/psec)	
	0.125	0.0625	0.125	0.0625
$\langle F_{max} \rangle$	1614 ± 58	1476 ± 43	1647 ± 37	1503 ± 36
$\langle r^t_{NC} - r^0_{NC} \rangle$*	11.8 ± 0.3	14.2 ± 1.1	5.5 ± 0.3	5.0 ± 0.4

* r^0_{NC} and r^t_{NC} are distances between the *N*- and *C*-termini at the initial moment of time and at the time when the force is maximal, respectively.

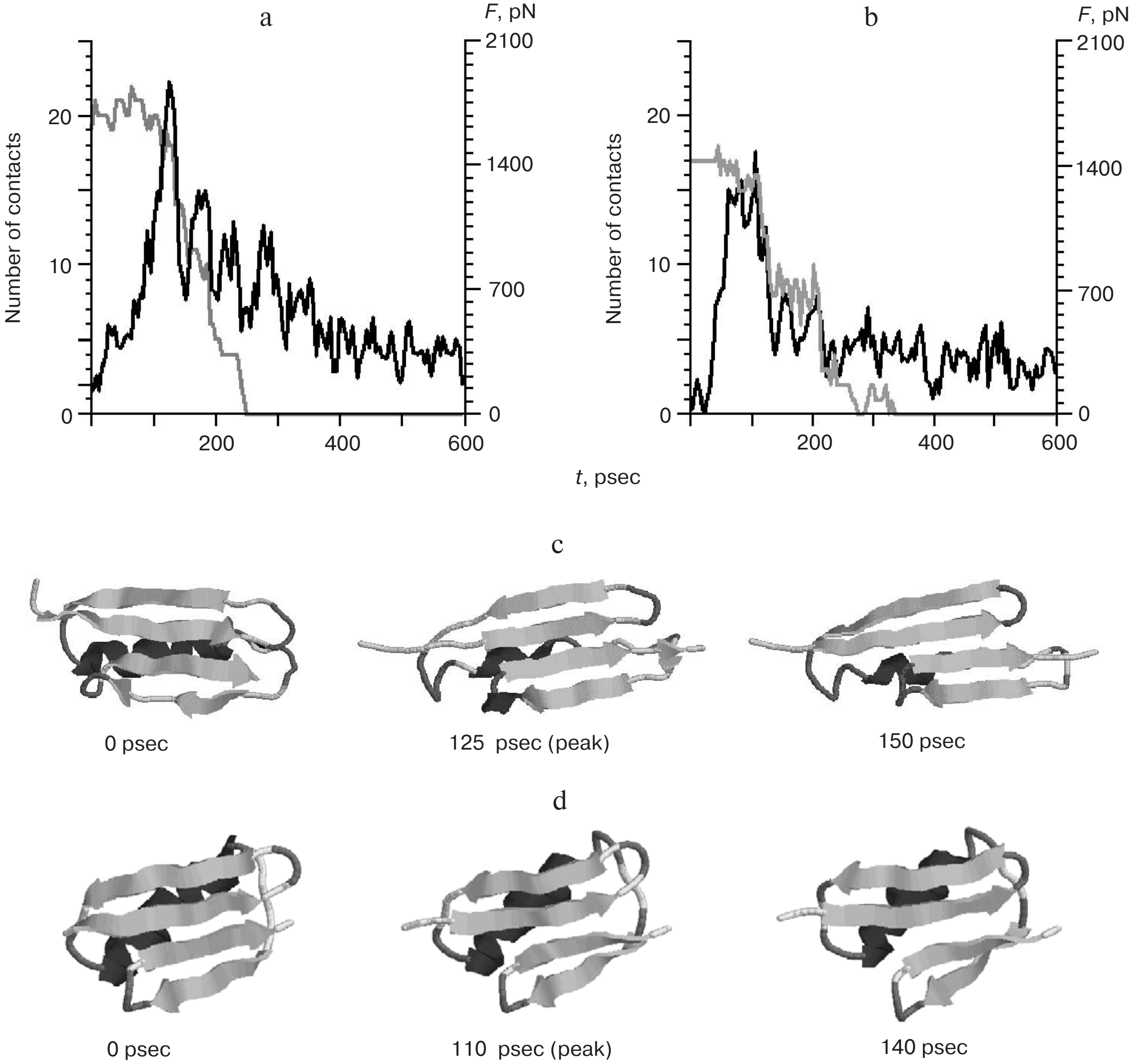


Fig. 3. Dependences of force (black curve) and number of contacts between the first and fourth β-strands (gray curve) on time: a, b) proteins L (υ = 0.125 Å/psec) and G (υ = 0.0625 Å/psec), correspondingly; c, d) structures of proteins L and G corresponding to initial moment of time, maximal peak of force, and minimum of force following by peak for one of the trajectories.

more contacts between secondary structure elements, the more mechanically stable the element. A comparison of the number of contacts (Table 3) and the order of unfolding of two proteins shows that at first disruption between the first and fourth β-strands takes place (22 contacts in protein L and 16 in protein G), then between the third and fourth β-strands, the *C*-hairpin (22 contacts in protein L and 16 in protein G), and finally between the first and second β-strands, the *N*-hairpin (26 contacts in protein L and 24 in protein G).

Ensembles of structures in transition state for proteins L and G were obtained and Φ-values for these structures were calculated. In Figs. 4 and 5 and Table 4, there is comparison of the experimental Φ-values from the literature [18, 19] and the Φ-values obtained in simulations.

From Figs. 4 and 5 one can see that both for protein L and protein G the Φ-values from simulations are higher than the experimental Φ-values. From the theoretical curves (Figs. 4 and 5) one can see that in protein G more non-native contacts appear during the unfolding than in protein L. For protein L, profiles of theoretical Φ-values obtained from different extension velocities are the same, but they somewhat differ for protein G.

Table 3. Structural and physical characteristics of proteins

Characteristic		Protein L	Protein G
Number of residue–residue contacts in protein/number of amino acid residues in protein		234/63 = 3.71	190/56 = 3.39
Number of atom–atom contacts in protein/number of atoms in protein		7170/951 = 7.54	6298/853 = 7.38
Number of hydrogen bonds in backbone		39	36
ln(folding rate, $\sec^{-1}$)		4.1	6.0
ln(unfolding rate, $\sec^{-1}$), 4 M GuHCl		–0.5	0.2
Number of residue–residue (atom–atom) contacts between different elements of secondary structure (all contacts calculated at contact distance 5 Å)	1st and 2nd β-strands	26 (1003)	24 (887)
	3rd and 4th β-strands	22 (838)	16 (662)
	1st and 4th β-strands	22 (885)	16 (759)
	1st and 3rd β-strands	0 (0)	2 (74)
	2nd and 3rd β-strands	0 (0)	0 (0)
	2nd and 4th β-strands	2 (10)	0 (0)
	α-helix and 1st β-strand	12 (346)	9 (415)
	α-helix and 2nd β-strand	13 (308)	6 (217)
	α-helix and 3rd β-strand	7 (238)	7 (228)
	α-helix and 4th β-strand	10 (283)	6 (216)

Table 4. Correlation coefficients between theoretical Φ_{th} and $\tilde{\Phi}_{th}$ values calculated upon modeling of unfolding of proteins L and G under the action of external forces and experimental Φ_{exp} values obtained on unfolding of these proteins by denaturant

Correlation coefficient	Protein L (38 experimental points), υ (Å/psec)		Protein G (20 experimental points), υ (Å/psec)	
	0.125	0.0625	0.125	0.0625
For Φ_{th} and Φ_{exp}	0.39	0.40	0.28	0.25
For $\tilde{\Phi}_{th}$ and Φ_{exp}	0.45	0.45	0.14	–0.05

Note: Φ_{th} was calculated with consideration of only native contacts and $\tilde{\Phi}_{th}$ with consideration of all contacts.

Table 4 shows that correlation coefficients between experimental and theoretical Φ-values for protein L are higher than those for protein G, especially when theoretical Φ-values are calculated with account of all contacts. These coefficients are practically the same under different extension velocities for protein L and differ for protein G. The data obtained for protein L do not contradict the data reported in [41] where it has been shown that there is a linear dependence between the force applied and the logarithm of unfolding rate. All these data indicate that under all stretching forces, protein L overcomes the same barrier, i.e. unfolding pathways are homogeneous. The same concerns protein G for which unfolding pathways are the same for different extension velocities.

In protein L (Fig. 5) theoretical Φ-values averaged over amino acid residues, included in the elements of secondary structure and related to the *N*-hairpin, are higher than the Φ-values related to the α-helix and *C*-hairpin.

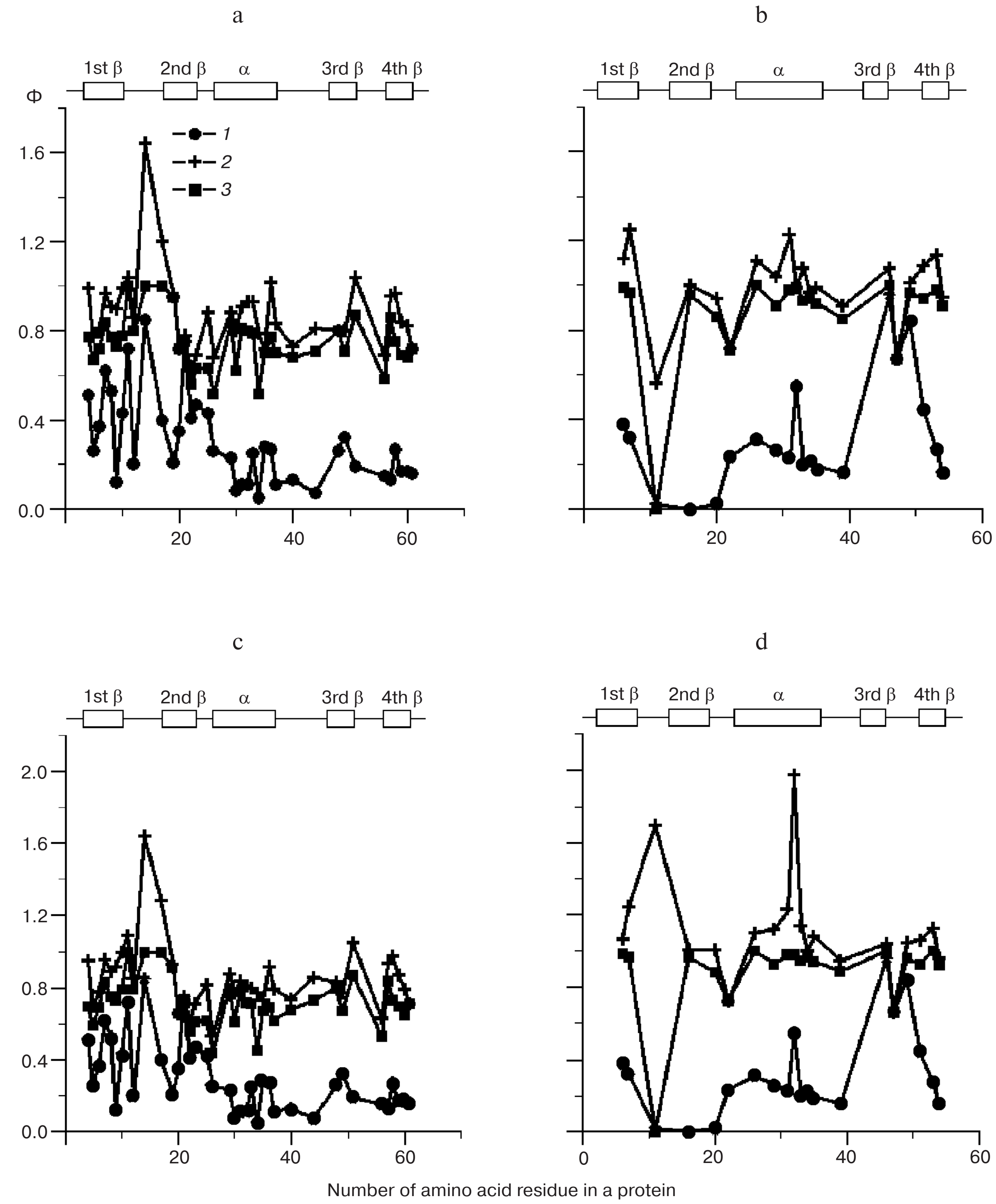


Fig. 4. Profiles of experimental (*1*) and theoretical (*2*, *3*) Φ-values for proteins L (a, c) and G (b, d) under υ = 0.125 Å/psec (a, b) and 0.0625 Å/psec (c, d); *2*, *3*) theoretical Φ-values calculated with consideration of all contacts and only native contacts, respectively.

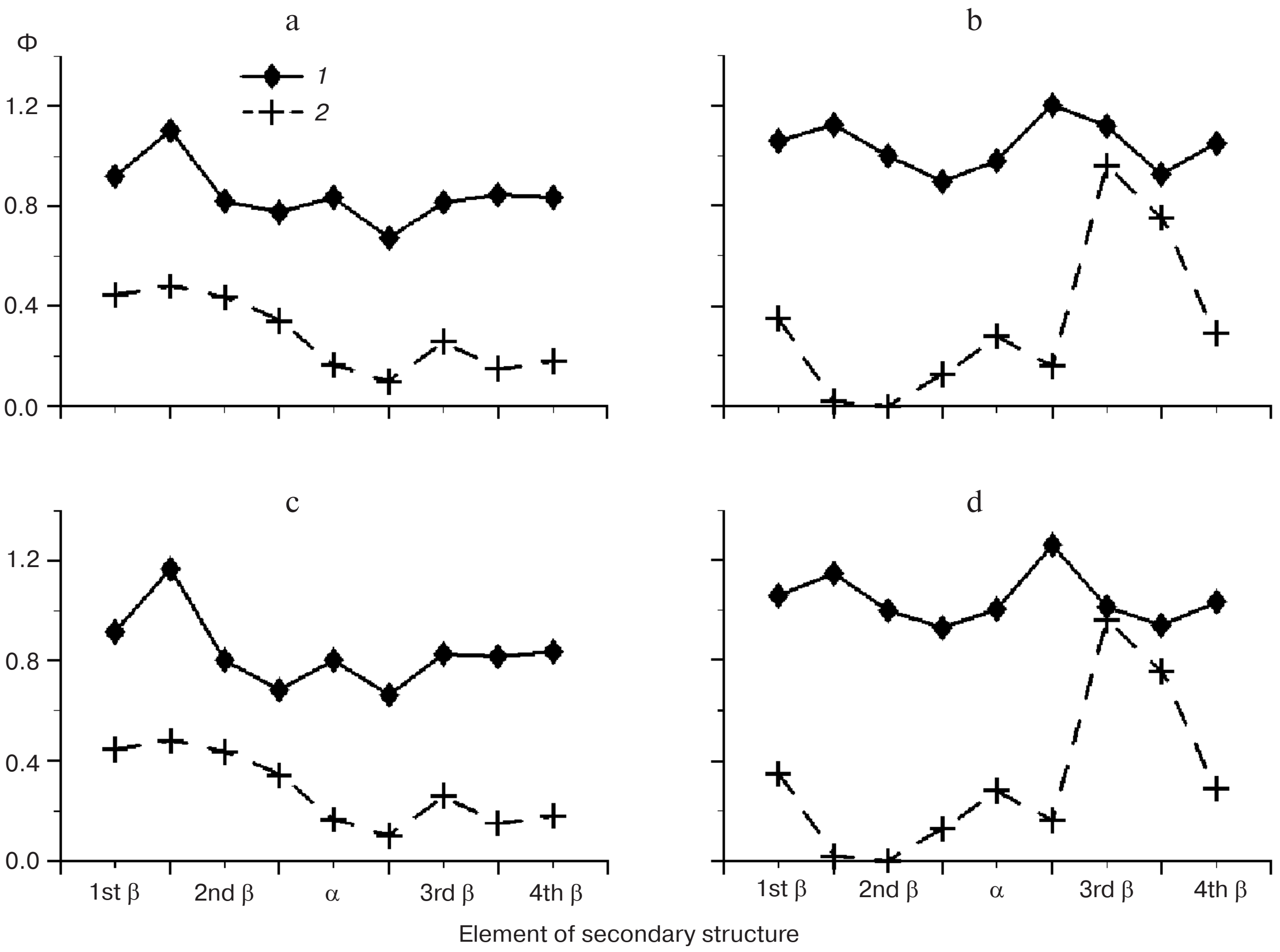


Fig. 5. Profiles of experimental (*1*) and theoretical (*2*) Φ-values for proteins L (a, c) and G (b, d), averaged over amino acid residues belonging to elements of secondary structure under υ = 0.125 Å/psec (a, b) and 0.0625 Å/psec (c, d). Theoretical Φ-values calculated with consideration of all contacts.

This is in agreement with the experimental data from which it follows that the *N*-hairpin is included in the unfolding nucleus of protein L.

It should be noted that at present the correlation coefficient between theoretical and experimental Φ-values equal to 0.5 is considered to be a good one [35]. The discrepancy between theoretical and experimental Φ-values is explained both by errors in the theoretical calculations (reductive models) and errors in the accuracy of experimental measurements (Φ ± 0.1). So, the correlation for these two proteins is 0.3 (protein L, 38 experimental points) and 0.76 (protein G, 20 experimental points) with the landscape approach [35].

It can be suggested that the most mechanically stable element of secondary structure should be included in the folding nucleus of a protein. As it has been shown in this study, both in protein L and in protein G the most mechanically stable element is the *N*-hairpin. It has been demonstrated in the experiments of unfolding of these two proteins by denaturant that the *N*-hairpin is included in the folding nucleus of protein L, and the *C*-hairpin of protein G. The theoretical Φ-values calculated from unfolding of proteins L and G under the action of forces show that in both proteins the most mechanically stable element of secondary structure—the *N*-hairpin is included in the folding nucleus. Thus, the experimental data obtained upon unfolding of these proteins by denaturant and the theoretical data obtained upon modeling of unfolding of these proteins under the action of external forces coincide for protein L and do not coincide for protein G. Although the structures of these proteins are similar, the order of destruction of secondary structure elements is different. In protein L in most cases, the order of destruction of secondary structure elements is as follows: at first the α-helix breaks, then the *C*-hairpin, and finally the *N*-hairpin. But, in protein G at first the *C*-hairpin breaks, then the α-helix, and finally the *N*-hairpin. Therefore, the *N*-hairpin is included in folding nuclei on mechanical unfolding of proteins L and G, but is not included in the folding nucleus of protein G in experiments on unfolding of this protein by denaturant.

This work was supported by the Russian Academy of Sciences ("Molecular and Cell Biology" program), by the Russian Foundation for Basic Research (grants No. 06-03-32814 and No. 08-04-00561), by INTAS grant No. 05-1000004-7747, by the Russian Science Support Foundation, and by Howard Hughes Medical Institute (grant 55005607). Calculations were done at the Joint Supercomputing Center of the Russian Academy of Sciences.

REFERENCES

1. Matouscheck, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature*, **346**, 440-445.
2. Krantz, B. A., and Sosnick, T. R. (2001) *Nature Struct. Biol.*, **8**, 1042-1047.
3. Li, A., and Daggett, V. (1996) *J. Mol. Biol.*, **257**, 412-429.
4. Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1996) *J. Mol. Biol.*, **257**, 430-440.
5. Caflisch, A., and Karplus, M. (1995) *J. Mol. Biol.*, **252**, 672-708.
6. Brooks, C. L., Gruebele, M., Onuchic, J. N., and Wolynes, P. G. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 11037-11038.
7. Finkelstein, A. V. (1997) *Protein Eng.*, **10**, 843-845.
8. Landsberg, P. (1971) *Problems on Thermodynamics and Statistical Physics* [Russian translation], Mir, Moscow.
9. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossman, J. G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V., and Fersht, A. R. (2003) *Nature*, **421**, 863-867.
10. Galzitskaya, O. V., and Finkelstein, A. V. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 11299-11304.
11. Alm, E., and Baker, D. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 11305-11310.
12. Munoz, V., and Eaton, W. A. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 11311-11316.
13. Takada, S. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 11698-11700.
14. Baker, D. (2000) *Nature*, **405**, 39-42.
15. Guerois, R., and Serrano, L. (2001) *Curr. Opin. Struct. Biol.*, **11**, 101-106.
16. Gunasekaran, K., Eyles, S. J., Hagler, A. T., and Gierasch, L. M. (2001) *Curr. Opin. Struct. Biol.*, **11**, 83-93.
17. Perl, D., Welker, Ch., Schindler, T., Schroder, K., Marahiel, M. A., Jaenicke, R., and Schmid, F. X. (1998) *Nature Struct. Biol.*, **5**, 229-235.
18. McCallister, E. L., Alm, E., and Baker, D. (2000) *Nature Struct. Biol.*, **7**, 669-673.
19. Kim, D. E., Fisher, C., and Baker, D. (2000) *J. Mol. Biol.*, **298**, 971-984.
20. Blanco, F. J., Rivas, G., and Serrano, L. (1994) *Nature Struct. Biol.*, **1**, 584-590.
21. Yi, Q., Scalley-Kim, M. L., Alm, E. J., and Baker, D. (2000) *J. Mol. Biol.*, **299**, 1341-1351.
22. Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J., and Serrano, L. (2000) *J. Mol. Biol.*, **297**, 1217-1229.
23. Gillespie, J. R., and Shortle, D. (1997) *J. Mol. Biol.*, **268**, 170-184.
24. Cordier-Ochsenbein, F., Guerois, R., Baleux, F., Huynh-Dinh, T., Lirsac, P. N., Russo-Marie, F., Neumann, J. M., and Sanson, A. (1998) *J. Mol. Biol.*, **279**, 1163-1175.
25. Galzitskaya, O. V. (2002) *Mol. Biol.* (Moscow), **36**, 386-390.
26. Best, R. B., Fowler, S. B., Herrera, J., Steward, A., Paci, E., and Clarke, J. (2003) *J. Mol. Biol.*, **330**, 867-877.
27. 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.
28. Wang, J., Cieplak, A., and Kollman, P. A. (2000) *J. Comp. Chem.*, **21**, 1049-1074.
29. Lemak, A. S., and Balabaev, N. K. (1995) *Mol. Simul.*, **15**, 223-231.
30. Lemak, A. S., and Balabaev, N. K. (1996) *J. Comp. Chem.*, **17**, 1685-1695.
31. Allen, M. P., and Tildesley, D. J. (1987) *Computer Simulation of Liquids*, Clarendon, Oxford.
32. Kabsch, W., and Sander, Ch. (1983) *Biopolymers*, **22**, 2577-2637.
33. Fersht, A. R. (1997) *Curr. Opin. Struct. Biol.*, **7**, 3-9.
34. Privalov, P. L. (1979) *Adv. Protein Chem.*, **33**, 167-241.
35. Garbuzynskiy, S. O., Finkelstein, A. V., and Galzitskaya, O. V. (2004) *J. Mol. Biol.*, **336**, 509-525.
36. 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.
37. Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F., and Fernandez, J. M. (2003) *Nature Struct. Biol.*, **10**, 738-743.
38. Carrion-Vasquez, 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.
39. West, D. K., Olmsted, P. D., and Paci, E. (2006) *J. Chem. Phys.*, **125**, 204910-204917.
40. West, D. K., Brockwell, D. J., Olmsted, P. D., Radford, Sh. E., and Paci, E. (2006) *Biophys. J.*, **90**, 287-297.
41. West, D. K., Olmsted, P. D., and Paci, E. (2006) *J. Chem. Phys.*, **124**, 154909-154918.