Simulation of the Mechanical Unfolding of the Ubiquitin by Pulling in Different Directions with Constant Speed

Stanislav Falkovich,[1] *Anatolii Darinskii,*[*1] *Nikolai Balabaev,*[2] *Igor Neelov*[1]

Mechanical extension of the ubiquitin with constant speed in five different directions is simulated on coarse-grained Go-like and all-atom models. The anisotropy of the mechanical resistance of the protein is observed in agreement with experimental data. Differences and similarities between the results obtained for two models are discussed. It is shown that the unfolding begins from the rupture of contacts between residues located in the vicinity of points of the external load application.

Keywords: mechanical unfolding of proteins; molecular dynamics; ubiquitin

Introduction

Many proteins are exposed to mechanical influence *in vivo*,[1–3] therefore mechanical properties of proteins have been actively studied in last decades. One of the most informative methods of the study of mechanical properties of proteins is the method of the Atomic Force Microscopy (AFM) (and a method of optical tweezers which has much in common with AFM).[4,5] In this method load is applied to the ends of concatemer consisting of the several (usually identical) copies of a protein molecule connected in a linear chain by short spacers and the response of the concatemer to the applied load is studied. A pair of amino acid residues located on a surface of protein is usually taken as points of the application of the load. There are two main types of single molecule AFM experiments: a stretching of protein with constant force or with constant speed.

In constant force experiments[6–9] the dependence of a stretching of a molecule on time is determined. Usually in such experiments the values of force in the range 10^{-12} N -10^{-9} N are used. During some time after the beginning of experiment the distance between points of application of force remains constant, and then increases suddenly, that corresponds to the beginning of unfolding of a protein. A time during which the protein molecule remains native under the same constant force, differs for different proteins and is a measure of their resistance to mechanical unfolding: if the protein remains native under constant force for a longer time, then its resistibility is higher.

In experiments of the stretching of a concatemer with a constant speed[10–17] a distance between the points to which the load is applied depends linearly on time and a dependence of the reaction force on time (or extension) is measured. At the initial stages of such experiments only spacers are extended and the protein remains in a native state and the reaction force of concatemer weakly grows with an extension. Further, when the spacers are completely stretched, the load is transferred to protein globule and the reaction force increases. At some extension the value of this force occurs to be enough to unfold one of the protein molecules, incoming into the

[1] Institute of Macromolecular Compounds of Russian Academy of Sciences, Bolshoi Pr., 31, 199004, Saint-Petersburg, Russia
E-mail: adar@imc.macro.ru
[2] Institute of Mathematical Problems of Biology of Russian Academy of Sciences, 4, Institutskaja str., 142290, Pushchino, Moscow Region, Russia
Fax: (812) 328-68-69

concatemer. At this time some intramolecular contacts in the protein break, and the reaction force sharply falls down. The dependence of a reaction force on an extension looks like a ''saw tooth''. Successive unfolding of protein monomers in the concatemer leads to the sequence of such ''saw teeth''. The force and extension necessary for mechanical unfolding of a single protein are determined by averaging over all ''saw teeth''. Usually in experiments the values of pulling speed of 10^{-11}–10^{-8} nm/ps are used. The maximum of the dependence of the reaction force on extension (so-called ''unfolding force'') is used as a measure of resistibility of a protein to external mechanical effect. The value of the unfolding force varies for different proteins, usually from several piconewtons (pN) up to several hundreds of piconewtons.

The results received in AFM-experiments of both types correlate with each other: unfolding of more mechanically resistant protein requires higher values of unfolding force (in a case of loading with a constant speed) and more time (in a case of loading with a constant force). The values of extension corresponding to the beginning of unfolding of a protein have an order of 0.1–1 nm in AFM-experiments of both types.

Each protein molecule has a unique structure and the possibilities of the analytical theory to describe results of AFM-experiments are very restricted. Therefore computer simulation appears to be the most adequate theoretical method, which allows investigate the mechanism of unfolding of proteins under loading. A large number of the papers devoted to simulation of a loading the proteins both with constant force[18–22] and with constant speed[23–27] is published by now. Computer simulation occurred to be able to predict which proteins are more mechanically resistant than others.[20]

Experiments on AFM carried out in last years have shown that alongside with distinctions in mechanical properties of various proteins, the resistibility of the same protein to a load depends on a pulling direction, i.e. anisotropy of mechanical properties of protein is observed. This effect has been shown in experiments on a pulling with constant speed for ubiquitin,[9] protein E2Lip3[14] and Green Fluorescent Protein (Green Fluorescent Protein, GFP[17]). For these proteins the load was applied to various pairs of amino acid residues and the value of an unfolding force (or of the time required for unfolding) depended on a choice of a pair.

For the first two proteins computer simulation of a pulling in various directions with constant force has been carried out with coarse-grained and all-atom models.[14,20,27] The qualitative agreement with experimental data has been shown: the same pulling directions demonstrated enhanced resistibility both in experiment and in simulation. Besides that, the mechanism of spontaneous (without mechanical extension) unfolding of the ubiquitin has been investigated.[20] It was shown, that in latter case the distance between residues with numbers 24 and 52 in a transition state at the top of a barrier of the free energy varies stronger than the distances between all other pairs of residues. For mechanical extension the resistibility of the ubiquitin in the pulling direction 24–52 occurred to be less than in other directions. Authors suggested that the mechanical unfolding occurs easier in that direction in which protein undergoes the greatest deformation during spontaneous unfolding as it is a ''native'' coordinate of reaction of the protein unfolding passing throughout the lowest barrier of free energy.

There are also direct simulations of the pulling with constant speed of the ubiquitin on the base of a coarse-grained Go-like model.[22] In the agreement with the experiment, the unfolding force for a pulling direction 1–76 was higher, than for 48–76. The similar result was obtained in the simulation of pulling of the ubiquitin with a constant speed with use of all-atom model.[9] It was assumed that the mechanical unfolding of the ubiquitin in various

directions is related with the rupture of certain hydrogen bonds.

However several questions were not answered in the works mentioned above. The first one: do all-atom and coarse-grained models predict the similar behavior of a protein molecule under loading with a constant speed? Simulation of unfolding of various proteins under constant force[20] leads to nearly the same ranking of protein resistance for both models. The second question: does a protein start to unfold in the points of application of the load or somewhere inside the protein molecule? In the present paper we try to answer these questions by simulation of the mechanical extension of the ubiquitin with a constant speed in different directions using Go-like and all-atom models.

Ubiquitin was chosen as an object of the study because of a large number of experimental and theoretical works devoted to research of its mechanical properties.[20,22]

Methods

Ubiquitin molecule consists of 76 amino acid residues and has a typical $\alpha+\beta$ topology (see Figure 1).[28]

In the present work are the following pulling directions were chosen: 1–76 (i.e. the load was applied to residues with the numbers 1 and 76), 11–76, 29–76, 48–76 and 24–52. These directions are the same as in the work of Paci et al.[22] Simulation was carried out by a method of molecular dynamics. A pulling speed of $5 \cdot 10^{-3}$ nm/ps typical for simulation of this kind[10–17] was used. It should be noted that this speed is essentially less than speed of a sound in protein.

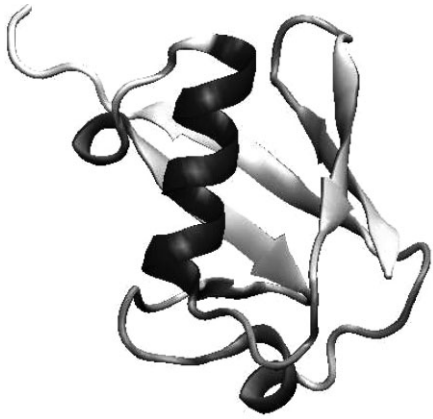

Figure 1.
Schematic structure of the ubiquitin. The image is produced with use of VMD software.[29]

Simulation on the base of a coarse-grained Go-like model was performed with CHARMM package.[30] The Go-like potential[22] was used where the attraction exists only between amino acid residues which have spatial contact in the native structure. Interactions between all other residues are assumed to be repulsive. Thus, the native structure for Go-like model corresponds to the global energy minimum of the system. As an initial structure for simulation the native structure of the ubiquitin molecule (a code in a RCSB Protein Data Bank is 1ubq[31]) was used. Every amino acid residue was substituted by a sphere of the corresponding mass with the center on C_α-atom, and thus the polypeptide chain was considered as a sequence of spheres connected by inextensible bonds. The potential[32] of native interactions between spheres is described with a formula: [33] $E_{ij} = \varepsilon_{ij} \cdot (13\times^{12} - 18\times^{10} + 4\times^{6})$, where $x = \sigma_{ij}/r_{ij}$, r_{ij} is a distance between C_α-atoms of the contacting residues with numbers i and j, σ_{ij} is a distance between C_α-atoms of these residues in the native state of protein. ε_{ij} is the depth of a potential hole dependent on energy of interaction of side radicals and from presence of hydrogen bonds between the residues with the numbers i and j. Actually, it is a Lennard-Jones potential modified to take into account the barrier of free energy describing (de)hydration occurring during formation or break of contact between residues. The view of the potential of native interactions is shown in Figure 2.

Besides that valence and torsion angles potentials are taken into account with the native values of angles corresponding to energy minimums.

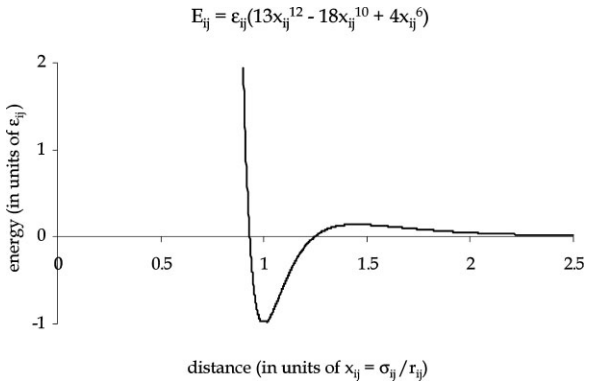

Figure 2.
Dependence of the energy (in units of ε_{ij}) on the distance (in units of x_{ij}) for native interactions.

The temperature 300 K was kept by a Langevin thermostat,[33,34] the step of integration was equal to 15 fs. Two particles (residues i and j for direction i-j) to which loading was applied were attached to two virtual particles by two virtual springs with rigidity 1 N/m, that is much higher than rigidity of van der Waals interactions in protein. The initial positions of 1 and 2 virtual particles coincide with positions of atoms i and j, correspondingly. Virtual particles were displaced with constant speed in opposite directions, and reaction force due to a stretching of virtual springs between virtual particles was measured.

For each of five pulling directions ten molecular-dynamic trajectories were calculated. The values of reaction force and distances between points to which loading was applied, were kept with step 0.1 ps and coordinates of all residues - with step 0.5 ps.

Simulation for all-atom model in explicit water was performed with software package PUMA[35,36] and AMBER-99 forcefield.[37] Collisional thermostat[35,36] was applied to maintain of constant temperature of 300 K. The step of integration was equaled 1 fs. Before the beginning of simulation the protein molecule (PDB ID 1ubq) was located in a parallelepiped filled with molecules of water TIP3P[38] with standard density of 10^3 kg/m^3. The initial number of water molecules in the parallelepiped was 1400 ($10 \times 10 \times 14$) in order to cover a surface of protein with several layers of water. The water molecules overlapping with protein atoms were removed, and 1165 water molecules remained in simulation system. Periodic boundary conditions were not used, the system was placed in sphere-cylinder with repulsive walls.

Atoms to which the load was applied were connected with virtual atoms by the virtual springs having rigidity, comparable with rigidity bonds. These virtual atoms were fixed on the parallel planes moving from each other with constant speed (displacement of the virtual atoms on the plane was allowed).

Five molecular-dynamic trajectories were obtained for each of pulling directions, values of the reaction force of virtual springs have been collected each 0.5 ps, coordinates of all atoms – each 2 ps. Before start of pulling each system was equilibrated for 200 ps.

The dependences of reaction force of the ubiquitin on extension were obtained for each molecular-dynamic trajectory. These dependencies were averaged for each pulling direction (the dependences, obtained with all-atom model and with Go-like model were averaged separately). The ''unfolding point'', i.e. the point of the dependence of reaction force on extension, corresponding to beginning of unfolding, was determined as the top of the first peak,

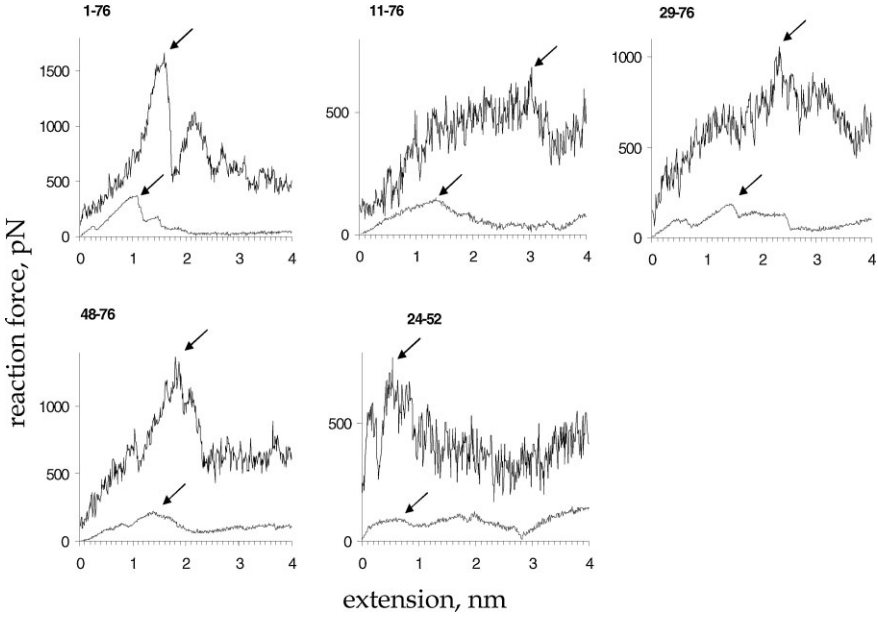

Figure 3.
Dependence of the reaction force on extension for different pulling directions, obtained with all-atom model (upper lines) and Go-like model (lower lines). The "unfolding points" are shown with arrows. All curves were averaged over the all trajectories for the same pulling direction.

followed by a sharp decrease of a reaction force. The unfolding point gives the value of unfolding force and the value of corresponding extension.

Results and Discussion

Dependence of the Reaction Force on the Extension

Dependence of the reaction force on the extension, obtained in simulation is shown as five plots for five different pulling directions at Figure 3. Each plot contains two force-extension curves for full atomic model (upper curve) and Go-model (lower curve), correspondingly. It is easy to see that for each plot (i.e. for each pulling directions) these two curves have much in common.

The curves for pulling direction 1-76 demonstrate a number of peaks with decreasing height, those for 29–76 have a small peak before the highest one and a plateau after it. Dependences for directions 11–76 and 48–76 are qualitatively similar. Curves for pulling direction 24–52, have one peak at low extension and two consequent peaks at larger extensions. For the Go-like model they are of the comparable height while for the all-atom model the second peak is much lower than the first one.

For all pulling directions the reaction force, obtained from simulation of all-atom model is an order of the magnitude higher than that obtained from Go-like model. There are two reasons of the difference between values of the reaction force obtained with different models. The first one is that force fields used in these models are different. The second one is that in the Go-like model only native contacts are taken into account and only they are to be broken when the protein unfolds. But in the all-atom model all interactions (both native and non-native) are taken into account. That is why the broken native contacts can be replaced by non-native ones which can be energetically favorable. Thus, in all-

Table 1.
Values of the unfolding force and corresponding extension

	unfolding force, 10^{-10} N			extension, nm		
	Go	all-atom	experiment[9]	Go	all-atom	experiment[9]
1–76	3.7 ± 0.1	17 ± 1	2	1.1 ± 0.1	1.6 ± 0.1	0.3
48–76	2.3 ± 0.1	14 ± 1	0.9	1.4 ± 0.1	1.8 ± 0.4	0.6
29–76	1.9 ± 0.1	11 ± 1		1.5 ± 0.2	2.3 ± 0.8	
11–76	1.5 ± 0.1	7 ± 1		1.4 ± 0.1	3.0 ± 0.6	
24–52	1.0 ± 0.1	8 ± 1		0.6 ± 0.1	0.5 ± 0.3	

atom model one has to break both native and non-native contacts in order to unfold the protein.

The values of the unfolding force and corresponding extension obtained for each pulling direction with both all-atom model and Go-like model are shown in the Table 1. The pulling directions are arranged in the decreasing order of the value of an unfolding force. It is seen that both models predict near the same order of pulling directions. The only exception is an order of pulling directions 11–76 and 24–52, but the difference for all-atom model is within a statistical error.

In accordance with the experimental data the value of the unfolding force is higher for pulling direction 1–76 than for 48–76, and the value of corresponding extension is higher for pulling direction 48–76 than for 1–76 for both models. However the agreement between the experimental and simulation data is only qualitative. For the Go-like model simulated values of the unfolding force are nearly twice higher than experimental values. Simulated values of the extension corresponding to the start of the unfolding are also higher than experimental ones. We think that this difference is connected with the high pulling speed used in simulation ($5 \cdot 10^{-3}$ nm/ps) in comparison with the experimental speeds (10^{-9}–10^{-5} nm/ps).

Qualitative agreement between the values of extension corresponding to the unfolding point obtained with both simulation models is also observed. The lowest extensions are for the pulling direction 24–52 in both cases. The second lowest extension is for direction 1–76 and for other pulling directions the extensions are higher (see Table 1). The values of extension, corresponding to the unfolding point, are lower for Go-like model than for all-atom model for all pulling directions besides a case of pulling direction 24–52. However the values of extension corresponding to the unfolding point for pulling

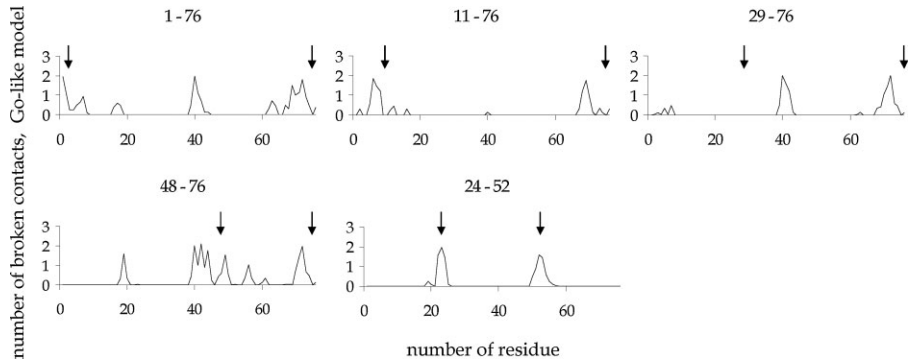

Figure 4.
Profiles number of native interaction broken before the unfolding point for various pulling directions, obtained with Go-like model. Points of the load application are marked with arrows.

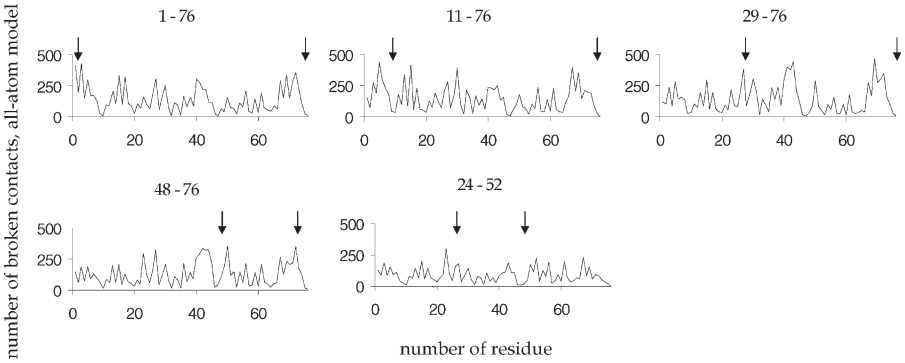

Figure 5.
Profiles of number of native atom-atom contacts broken before the unfolding point for various pulling directions, obtained with all-atom model. Points of the load application are marked with arrows.

direction 24–52 obtained with Go-like and all-atom models are within statistical error.

Breakage of Contacts

It is clear that the process of the mechanical unfolding is connected with the breakage of native contacts. The question arises does this breakage begin from contacts which are located in the close vicinity of the load application or these contacts are distributed over the whole protein? To answer this question it is necessary to calculate the numbers of residues which lose their native contacts in the first stage of unfolding.

We assumed that the native contact in the Go-like model is broken if the distance between contacting residues becomes 1.25 times longer than that in the native state. At such a distance the native interaction is repulsive (Figure 2). Figure 4 shows how many native contacts are lost by a residue with sequence number "i" before the unfolding point was achieved. It is seen that the maximum number of broken contacts are usually observed for residues neighboring along the chain to pulled residues. The only exclusion is the pulling direction 29–76, where the contacts are not lost in a region of residue 29 but are lost in a region of the C-termini. The reason is that the residue 29 belongs to the helix which remains undisturbed up to this moment.

Besides the neighbors along the chain the peaks are observed also for some

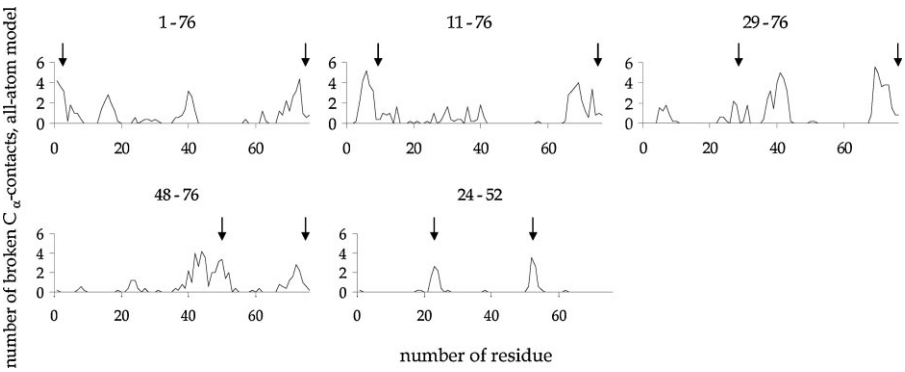

Figure 6.
Profiles of number of native contacts between C_α-atoms broken before the unfolding point for various pulling directions, obtained with all-atom model. Points of the load application are marked with arrows.

neighbors in the spatial structure. In particular the peak near the 40-th residue is observed for pulling directions 1–76, 29–76 and 48–76, where the load is applied to the C-termini. The reason of the appearance of this peak is that the residues near the 40-th one are neighbors of C-terminal residues in a native structure. However we don't see the peak near the residue 40 for pulling direction 11–76, where the load is also applied to the C-termini. What is the difference between this case and the previous ones? This difference is connected with different orientation of the pulling directions relative to contacts between residues belonging to C-terminal and to the region around the residue 40. The first three pulling directions are oriented nearly along these contacts. Thus they break one by one that is known to be an easy-going process. In contrast, the pulling direction 11–76 is oriented transversely to these contacts, and they should be broken simultaneously to separate the C-terminal region from the protein globule.

For the all-atom model we assumed that two atoms lose a native contact if the distance between them exceeds the distance corresponding to the energy minimum of van der Waals interactions. For example, for methyl groups this distance equals to 0.45 nm in AMBER force-field. We choose 0.5 nm as a critical distance for all atom-atom interactions. Results are shown in Figure 5.

It is seen that for all atom model there are essentially more peaks for the number of broken contacts than for the Go-like model. It is because the all-atom model is more detailed than the Go-like one and takes into account not only a change of the distance between centers of residues but also a change of distance between atoms of side groups belonging to these residues. And such changes are not always connected with change of the distances between centers of residues.

In order to compare results for two models we have calculated for the all-atom model a number of broken contacts between C_α-atoms because in the Go-like model amino acid residues were represented by spheres with centers in the C_α-atoms (see Methods). As for Go-like model we assumed that two C_α-atoms loose a native contact if the distance between them becomes 1.25 times longer than that in the native state. The results are shown in Figure 6. The comparison with Figure 4 shows that in both models the same residues lose contacts before the unfolding.

Therefore additional peaks for other residues observed for all-atom model are connected with relative displacements of side groups belonging to contacting residues without an increase of the distance between them.

On the base of above results we can conclude that the critical breakages of intramolecular contacts occur in the vicinity of the points to which the load is applied. The difference in the mechanical resistance of the protein molecule observed by pulling in different directions is connected mainly with the loss of different native contacts between residues as a whole. That is why Go-like and all-atom models predict the same sequence of pulling directions if they are arranged in the order of the mechanical resistance.

Acknowledgements: Work was supported by European program INTAS (No 05-1000004-7747) and the RFBR grant 08-03-00565. We are thankful to Dr. Emanuele Paci for initiation of this research and for helpful discussions.

[1] M. G. L. van den Heuvel, C. Dekker, *Science* **2007**, *317*, 333.
[2] E. A. Evans, D. A. Calderwood, *Science* **2007**, *316*, 1148.
[3] W. A. Linke, *Cardiovasc Res* **2008**, *77*, 637.
[4] M. Sotomayor, K. Schulten, *Science* **2008**, *316*, 1144.
[5] A. F. Oberhauser, M. Carrion-Vazquez, *J. Biol. Chem.* **2008**, *283*, 6617.
[6] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, J. M. Fernandez, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 468.
[7] M. Schlierf, H. Li, J. M. Fernandez, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7299.
[8] J. M. Fernandez, H. Li, *Science* **2004**, *303*, 1674.

[9] M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, J. M. Fernandez, *Nat. Struct. Biol.* **2003**, *10*, 738.
[10] C. Chyan, F. Lin, H. Peng, J. Yuan, C. C. S. Lin, G. Yangy, *Biophys. J.* **2004**, *87*, 3995.
[11] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, H. E. Gaub, *Science* **1997**, *276*, 1109.
[12] M. Carrion-Vasquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, J. M. Fernandez, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3694.
[13] D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, S. E. Radford, *Biophys. J.* **2005**, *89*, 506.
[14] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, S. E. Radford, *Nat. Struct. Biol.* **2003**, *10*, 731.
[15] R. C. Zinober, D. J. Brockwell, G. S. Beddard, A. W. Blake, P. D. Olmsted, S. E. Radford, D. A. Smith, *Protein Sci.* **2002**, *11*, 2759.
[16] D. J. Brockwell, G. S. Beddard, J. Clarkson, R. C. Zinober, A. W. Blake, J. Trinick, P. D. Olmsted, D. A. Smith, S. E. Radford, *Biophys. J.* **2002**, *83*, 458.
[17] H. Dietz, F. Berkemeier, M. Bertz, M. Rief, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12724.
[18] P. Li, D. E. Makarov, *J. Phys. Chem.* **2004**, *108*, 745.
[19] P. Li, D. E. Makarov, *J. Phys. Chem.* **2004**, *121*, 4826.
[20] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, E. Paci, *Biophys. J.* **2006**, *90*, 287.
[21] D. K. Klimov, D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7254.
[22] R. B. Best, E. Paci, G. Hummer, O. K. Dudko, *J. Phys. Chem.* **2008**, *112*, 5968.
[23] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, K. Schulten, *Biophys. J.* **1998**, *75*, 662.
[24] E. Paci, M. Karplus, *J. Mol. Biol.* **1999**, *288*, 441.
[25] E. Paci, M. Karplus, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6521.
[26] J. I. Sulkowska, M. Cieplak, *Biophys. J.* **2008**, *94*, 6.
[27] M. Cieplak, P. E. Marszalec, *J. Chem. Phys.* **2005**, *123*, 194903.
[28] A. Murzin, S. E. Brenner, T. J. P. Hubbard, C. Chothia, *J. Mol. Biol.* **1995**, *247*, 536.
[29] W. Humphrey, A. Dalke, K. Schulten, *J. Molec. Graphics* **1996**, *14*, 33.
[30] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187.
[31] S. Vijay-Kumar, C. E. Bugg, W. J. Cook, Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **1987**, *194*, 531.
[32] J. Karanicolas, C. L. I. Brooks, *J. Mol. Biol.* **2003**, *334*, 309.
[33] H. J. C. Berendsen, *J. of Chem. Phys.* **1984**, *81*, 3684.
[34] A. S. Lemak, N. K. Balabaev, *Molecular Simulation*, **1994**, *13*, 177.
[35] A. S. Lemak, N. K. Balabaev, *Molecular Simulation* **1995**, *15*, 223.
[36] A. S. Lemak, N. K. Balabaev, *J. Comput. Chem.* **1996**, *17*, 1685.
[37] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. SOC.* **1995**, *117*, 5179.
[38] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.

