Published in final edited form as:
Proteins. 2008 April ; 71(1): 45–60. doi:10.1002/prot.21652.

Predicting the order in which contacts are broken during single molecule protein stretching experiments

Joanna I. Sułkowska[1], **Andrzej Kloczkowski**[2,*], **Taner Z. Sen**[2], **Marek Cieplak**[1], and **Robert L. Jernigan**[2]

[1] *Institute of Physics, Polish Academy of Sciences, Al. Lotników 46, 02-668 Warszawa, Poland* [2] *L. H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, Iowa 50011*

Abstract

We combine two methods to enable the prediction of the order in which contacts are broken under external stretching forces in single molecule experiments. These two methods are Gô-like models and elastic network models. The Gô-like models have shown remarkable success in representing many aspects of protein behavior, including the reproduction of experimental data obtained from atomic force microscopy. The simple elastic network models are often used successfully to predict the fluctuations of residues around their mean positions, comparing favorably with the experimentally measured crystallographic B-factors. The behavior of biomolecules under external forces has been demonstrated to depend principally on their elastic properties and the overall shape of their structure. We have studied in detail the muscle protein titin and green fluorescent protein and tested for ten other proteins. First, we stretch the proteins computationally by performing stochastic dynamics simulations with the Gô-like model. We obtain the force–displacement curves and unfolding scenarios of possible mechanical unfolding. We then use the elastic network model to calculate temperature factors (*B*-factors) and compare the slowest modes of motion for the stretched proteins and compare them with the predicted order of breaking contacts between residues in the Gô-like model. Our results show that a simple Gaussian network model is able to predict contacts that break in the next time stage of stretching. Additionally, we have found that the contact disruption is strictly correlated with the highest force exerted by the backbone on these residues. Our prediction of bond-breaking agrees well with the unfolding scenario obtained with the Gô-like model. We anticipate that this method will be a useful new tool for interpreting stretching experiments.

Keywords

protein stretching; mechanical unfolding; Gô model; elastic network model; Gaussian network model; titin; green fluorescent protein

INTRODUCTION

Recently, it has been shown that large-scale structural transformations in proteins, such as allostery, can involve large-scale domain motions and possible partial unfolding. The recent work of Miyashita *et al*. and Margakis and Karplus[1,2] has shown that both these processes explain the conformational transition of adenylate kinase, which demonstrates their importance for macroscopic changes of protein conformations. Unfolding can be an important aspect of

*Correspondence to: Andrzej Kloczkowski, L. H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, IA 50011. E-mail: kloczkow@gmail.com.

many biological processes. Protein unfolding can occur for many reasons. The most common is due to mechanical stretching as in muscle proteins. Recently, it has been shown that partial unfolding and refolding proteinquakes[3] can describe conformational changes in myosin. On the other hand, large-scale motions of proteins explain many biological mechanisms. Unfortunately, macroscopic motions of large systems are not yet fully accessible for computational studies using conventional molecular dynamics simulations, due to the extremely long times required for such computations. Recently, it has been shown that experimentally observed large-scale functional motions of large biological structures, such as chaperonins,[4] ribosomes,[5,6] and other biomolecular structures, can be described quite well with elastic network models.

In this article, we apply the elastic network models to explain structural changes in proteins during mechanical unfolding. More precisely, our goal is to apply the Gaussian network model (GNM) to transient protein structures arising during mechanical unfolding simulations in Go-like models. The reason why we use the GNM model is the fact that stochastic simulation can provide a certain number of trajectories but not full statistical ensemble that is needed to calculate the *B*-factors. The GNM model effectively computes thermal fluctuations of residues in the stretched protein even from a single trajectory.

There has been a long controversy over whether some observations such as hydrogen exchange are occurring because of denaturation or through normal structural fluctuations. Recent studies of large conformational transitions,[7,8] where multiple forms of the same protein have been determined, show that there is a strong tendency for the direction of the fluctuations to coincide with the directions of motion required for the different elements of the structure to effect the transition. This does not in itself disprove denaturation as an intermediate stage of this transition; but at least it is suggestive that the protein structure is so cooperative and cohesive that its transition is fully built into the structure and the protein can do little else in making changes to its structure.

The GNM is based on application to a structure in its energetic minimum, which is presumed to be its native state, determined by crystallography. The quasi-static harmonic picture, on which the GNM is based, could be expected to hold only close to the energy minimum and might not remain valid for large distortions. However, there is accumulating evidence that the elastic models do yield the directions of even large scale transitions[7,9] (Jernigan, Song, and Yang, unpublished). To apply GNM to the analysis of structural changes in proteins during stretching, we should first examine mechanisms of mechanical stretching of proteins. During unfolding, the protein moves from one local minimum to another by crossing the energy barrier (with some native contacts occasionally being broken) between these two conformations. We assume that a protein responds to the external mechanical force by moving according to a low frequency motion up to a certain critical displacement, and that the size of this displacement may be considered to be within the regime of linear elasticity. However, as the amplitude of the displacement grows, the nonlinear elastic regime eventually is reached and contacts between residues are broken. Proteins locally/partially unfold if the elastic stress in the structure becomes too high. Following this, the protein might relax to a new conformation where the elastic network theory can be applied again.

Our hypothesis is that, when tension between two residues present in the native contact crosses a certain critical threshold, their mean square fluctuations in a particular slow soft mode should drop almost to zero. Moreover, this high tension should correlate with the highest force acting at these particular residues in the Gô-like model. To test this hypothesis, we have performed the following computational experiment.

We stretch proteins at a constant speed using stochastic molecular dynamics simulations of a coarse-grained Gô-like model. From these simulations, we take a set of several transient conformations (snapshots) at different time points (every 0.5 Å of chain end-to-end distance extension) during the process of mechanical unfolding of the protein. The most interesting of these temporary conformations are analyzed with the GNM model to predict contacts that are most likely to break first upon further stretching. At each of these time steps we compare the order of contact breaking predicted by the GNM with that observed during the simulation experiment. The mechanical unfolding studies are effectively performed at zero temperature to neglect fluctuations, and exclude the possibility of the formation of new non-native contacts during simulations.[10] This approach simplified an analysis of which native contacts should break next. However, it is well known that thermal fluctuations aid the process of unfolding proteins, by decreasing both the unfolding force and a required extension. This can lead sometimes to significant changes in the unfolding scenario that depend strongly on the geometry of the protein.[11,12]

Here, we focus on the 27th immunoglobulin domain of the giant muscle protein titin[13] and the green fluorescent protein GFP.[14] Titin has been one of the prime objects of mechanical studies on single biological molecules[15–19] because of its role in controlling the degree of extension of smooth skeletal and cardiac muscles. Titin contains between 240 and 300 immunoglobulin- (Ig) and fibronectin-like (Fn3) domains that are common building blocks of intracellular proteins involved in ligand recognition and cell adhesion. Ig and Fn3 domains are important for titin interactions, where levels of interdomain mobility and structural stability relate directly to mechanical functions.

I27 unfolds in a unique way through one intermediate state corresponding to a high peak force. GFP is a large size protein and its unfolding pattern is not fully explored. It is known[20–22] that the peak force sensitively depends on the choice of the pulling amino acids. GFP has been studied theoretically in Ref. 23 by Heyon *et al*. Here, we study it by stretching it within the Gô model and then analyzing the phononic modes using the GNM model.

We show that the Gô model allows one to predict the difference in elasticity between I27 and GFP. We show that GFP pulled by the amino acids 3 and 132 exhibits a unique unfolding pathway with a few intermediate states as opposite to the I27. For GFP(3-132) and I27 of titin, we found a few conformations during stretching where the mean square fluctuations in one of the slowest modes have only two very deep minima, which indicates residues between which the contact will break in the next time-step. Moreover, we found that this is strictly correlated with the highest force exerted by the backbone on these residues. Prediction of the order of bond-breaking agrees with the unfolding scenario obtained from the Gô-like model simulations. However, in most cases, especially for GFP, we identify clusters of contacts that will break at the next time-step, instead of single contacts. These clusters of potentially breaking contacts need further verification by comparison with the unfolding scenarios and the behaviors of the forces as a function of distance.

This article is organized as follows. The Materials and Methods section presents briefly the Gô-like model, the parameters used for the stochastic molecular dynamic simulations of protein stretching, and a brief description of the GNM theory. The following Results section first summarizes information about the stretching of titin. This is followed by an analysis of the scenarios of contact-breaking based on the slowest modes in GNM model and the comparison with actual results on unfolding simulations. Next, we calculate forces acting on each residue during stretching. The same analysis is then performed for GFP. Finally, we present conclusions and a discussion of possible future work.

In this article, we consider stretching at constant speed. It should be noticed that another way to accomplish unraveling is through an application of constant force, both experimentally[24] and theoretically by Szymczak and by Cieplak.[25] This process is usually of a two-state kind (i.e., the protein is either folded or fully stretched), what is not interesting in the context of novel application of the GNM-like analysis.

MODELS AND METHODS

We use two different coarse-grained models of proteins: a Gô-like model for the stochastic molecular dynamics simulation of mechanical unfolding and the Gaussian network model (GNM) to predict, from the protein structure, the order in which contacts are most likely to break next upon further mechanical stretching.

Gô-like model

In our Gô-like model,[26] a protein is represented by C^{α} atoms connected along the main chain by massless harmonic springs, each of equilibrium length 3.8 Å , and having a spring constant $k = 100$ [ε/Å ^{2}]. Native amino acid contacts in the protein are assumed to interact by the Lennard–Jones potential

$$E_{ij}=4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (1)$$

where the parameters σ_{ij} are chosen so that the potential minima correspond, pair-by-pair, to the native distances between the C^{α} atoms i and j. To account for natural right-hand nature of helices, the chirality potential is added.[27] To recognize which amino acids comprise the *contact map* of native contacts, we follow the procedure of Tsai *et al*.[28] based on the criterion of overlaps of enlarged atoms. Atoms are represented by spheres with radii a factor of 1.24 larger than the atomic van der Waals radii, to account for the softness of the potential. The residues that do not form native contacts interact through the Lennard–Jones potential shifted and truncated at 4 Å . The energy parameter ε in Eq. (1) is assumed to be uniform, and its effective value divided by the Boltzmann constant k_B is of the order of 900 [K], at least for titin.

In our stretching simulation, both ends of the protein are attached to identical harmonic springs with an elastic constant $k = 0.12$ [ε/Å ^{2}], which corresponds closely to the elasticity of cantilevers in single molecule experiments. The free end of one of the two springs is fixed, whereas the free end of the other spring is pulled with a constant speed, v_p, along the direction of the initial end-to-end vector of the chain. We use a pulling velocity $v_p = 0.005$ [Å /τ], where $\tau=\sqrt{m\sigma^2/\varepsilon}\sim 3$ ps is the characteristic relaxation time for the Lennard–Jones potential.[29] This value of τ is derived for a mass of 118 the average value of the mass of amino acids. It can be argued, however, that, because of the high friction inside a protein, the characteristic time should perhaps be larger.[30] We assume that the protein undergoes stochastic Brownian dynamics, and the equation of motion for the position vector $\boldsymbol{r}$ of each C^{α} is

$$m\ddot{r}=-\gamma\,\dot{r}+F_c+F_{rand} \quad (2)$$

All residues are taken to have the same mass m that is a crude average of masses of all amino acids in a protein. Our previous study[31] showed that the use of a uniform average mass instead of the individual masses of residues leads to almost exactly the same results. The force $\boldsymbol{F}_c$ in Eq. (2) is a net force on the given residue resulting from all interactions in a Gô-like model, whereas $\boldsymbol{F}_{rand}$ is the stochastic random force (Langevin noise) that mimics random hits by the implicit solvent and provides for the thermal equilibrium of the system. The damping constant (friction coefficient) γ in Eq. (2) is taken to be equal to $2m/\tau$, and the dispersion of random

forces $\boldsymbol{F}_{\text{rand}}$ is equal to $2\gamma k_B T\delta(t)$. Such a choice for γ corresponds to a situation in which inertial effects are negligible,[31] but the damping action is not as strong as in water. Increasing γ by ten-fold would result in a ten-fold increase in time scales, thereby yielding a value of v_p that is two orders of magnitude higher than the experimental pulling speeds[10] and correspondingly longer folding times.[31] The equations of motion [Eq. (2)] are solved with a fifth order Gear predictor-corrector algorithm[32] using time-steps of 0.005 τ.

Gaussian network model

The GNM was originally developed for the theory of rubber-like elasticity of random polymer networks[33,34] to calculate fluctuations of junctions and chains inside the network. The model was adapted to proteins in 1997 by Bahar and Erman[35,36] following the earlier result of Tirion[37] who used a single harmonic force parameter to successfully describe atomic motions in proteins. The GNM is usually based, like the Gô model, upon coarse-grained modeling of proteins with a single site per residue. Positions of these sites are usually identified with the coordinates of the C^α atoms in proteins, and it is assumed that both bonded and nonbonded contacts in protein structures are connected by identical uniform massless harmonic springs.

To define which sites are in contact, a uniform cutoff distance R_c is used.[7,35,36,38,39] Residues separated by a distance less than or equal to R_c (including neighbors along the sequence) are assumed to be in contact, and are connected with identical springs. This leads to an elastic network representation of a protein structure in the folded state that resembles a random polymer network, except that a polymer network would be more sparsely interconnected. The dynamics of such an interconnected bead-and-spring model can be described using potentials of the form

$$\begin{aligned} V(R_{ij}) &= \tfrac{1}{2}\gamma(R_{ij} - R_{ij}^0)^2 H(R_c - R_{ij}) \\ &= \tfrac{1}{2}\gamma(\Delta R_{ij})^2 H(R_c - R_{ij}) \end{aligned} \tag{3}$$

where γ is a uniform universal spring constant, and $H(x)$ is the Heaviside step function that equals 1 if $x > 0$, and is zero otherwise. Here, $\boldsymbol{R}_{ij}$ is the distance vector between the ith and jth sites, R_{ij} is its magnitude with the super-script 0 referring to the mean value observed in the reference (native) structure, $\Delta\boldsymbol{R}_{ij}$ is the instantaneous displacement of $\boldsymbol{R}_{ij}$ from the mean value R_{ij}^0, and $(\Delta\boldsymbol{R}_{ij})^2$ is given by the scalar product $(\Delta R_{ij}^T \cdot \Delta R_{ij})$. The reference structure is the crystal structure taken from the Protein Data Bank (PDB). The total energy of the network composed of N nodes is the sum of energies given by Eq. (3) over all interacting pairs of beads, and this increases quadratically with the distortions from the native form and is expressed as

$$V_{\text{tot}} = (\gamma/2)\{\Delta R\}^{\text{T}}\Gamma\{\Delta R\} \tag{4}$$

where Γ is the Kirchhoff (contact) matrix of size $N \times N$, defined on the basis of the cutoff distance R_c, with off-diagonal elements ij being either −1 if nodes i and j are in contact or zero otherwise, and the diagonal elements are defined as the sum of the off-diagonal elements in the ith row (or column) taken with a negative sign. Here $\{\Delta\boldsymbol{R}\}$ is the N-dimensional fluctuation vector $\Delta\boldsymbol{R}$ = col $(\Delta\boldsymbol{R}_1, \Delta\boldsymbol{R}_2, \ldots, \Delta\boldsymbol{R}_N)$ of $\Delta\boldsymbol{R}_i$ for all N nodes, and the superscript T denotes the transpose.

Then the average changes in positions, given either as the correlation $\langle\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j\rangle$ between the displacements of pairs of residues i and j or as the mean-square fluctuations $\langle(\Delta R_i)^2\rangle = \langle\Delta R_i \cdot \Delta R_i\rangle$ for a single residue i, are

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{\int (\Delta R_i \cdot \Delta R_j) \exp(-V_{\text{tot}}/k_{\text{B}}T) d\{\Delta R\}}{\int \exp(-V_{\text{tot}}/k_{\text{B}}T) d\{\Delta R\}} \tag{5}$$

for all *i; j*

This is rigorously re-expressed[33,34] in a simple form as

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{3k_{\text{B}}T}{2\gamma} (\Gamma^{-1})_{ij} \tag{6}$$

where $(\Gamma^{-1})_{ij}$ is the *ij*th element of the inverse of the connectivity matrix Γ. Mean-square fluctuations $\langle (\Delta R\text{i})^2 \rangle$ of the *i*th residue in a protein are given by the *i*th diagonal element of Γ^{-1} in Eq. 6. The mean-square fluctuations in the distance $r_{ij} = |\boldsymbol{R}_i - \boldsymbol{R}_j|$ between two sites *i* and *j* are

$$\begin{aligned} \langle (\Delta r_{ij})^2 \rangle &= \langle (\Delta R_i - \Delta R_j)^2 \rangle \\ &= \frac{3k_{\text{B}}T}{2\gamma} \left[(\Gamma^{-1})_{ii} + (\Gamma^{-1})_{jj} - 2(\Gamma^{-1})_{ij} \right] \end{aligned} \tag{7}$$

One should note that the connectivity matrix Γ has been defined so that the sum of all elements in every row (or column) add to zero. Because of this, det Γ = 0, and the matrix is singular, so that only the pseudoinverse of Γ can be computed through the use of the singular value decomposition method. The pseudoinverse of Γ may be written as $\Gamma^{-1} = \boldsymbol{U} (\Lambda^{-1}) \boldsymbol{U}^{\text{T}}$, where $\boldsymbol{U}$ is the matrix composed of eigenvectors $\boldsymbol{u}_i$ ($1 \leq i \leq N$) of Γ, and Λ is the diagonal matrix of the eigenvalues of Γ. Additionally, it can be proved that all eigenvalues & λ_i of Γ are non-negative.

Mean-square fluctuations of the position of each C^{α} computed from Eq. (6) can be compared with the Debye–Waller thermal factors, which are measured by X-ray crystallography and deposited in the Protein Data Bank as thermal *B*-factors. The relationship between the *B*-factor and the mean square fluctuations for the *i*th residue is given by

$$B_i = 8\pi^2 \langle (\Delta R_i)^2 \rangle / 3 \tag{8}$$

Usually, the spring constant γ in Eq. (6) is treated as a single adjustable parameter chosen to obtain the best fit with experimental *B*-factors for all protein residues. The absolute value of γ does not however affect the relative distribution of fluctuations of residues, but only uniformly rescales them. For native structures, we can calculate the spring constant γ from experimental *B*-factors, and for deformed protein conformations obtained during mechanical stretching, we must rely upon the γ derived from the native state. The *B*-factors computed by GNM usually are in excellent agreement with experimental data.[40] Several earlier studies on elastic network models of biological structures[41,42] suggest that the global dynamics of proteins is insensitive to the details of the model and its parameters, and that the most important slowest modes are determined mainly by the protein's shape.[43] This view of protein structures as rubbery bodies has proved itself to be particularly useful in characterizing a structure's functional motions, which are typically not readily accessed with atomic molecular dynamics. There is now ample evidence supporting the validity of the GNM from a number of other types of experiments—H/D exchange data,[44] NMR order parameters,[45,46] NMR results for the directional motions of individual peptide bonds, and the distortions manifested over multiple structures.[7,8] The matrix Γ^{-1} can be written as the sum of contributions over all individual modes[36]

$$\boldsymbol{\Gamma}^{-1} = \sum_k \lambda_k^{-1} u_k u_k^T \tag{9}$$

where zero eigenvalues (those that physically correspond to rigid body motions of the center of mass of the system) are excluded from the summation. The *i*th component of the eigenvector $\boldsymbol{u}_k$ (corresponding to the *k*th normal mode) specifies the magnitude of fluctuation motions of the *i*th residue in the protein exerted in the *k*th mode. If we order the eigenvalues according to their ascending values starting from zero, then the most important contributions in Eq. (9) are given by the smallest nonzero eigenvalues & λ_k that correspond to the large-scale slow modes, usually corresponding to the large domain motions. The slowest modes play a dominant role in the fluctuation dynamics of protein structures, because their contributions to the mean-square fluctuations scale with λ_k^{-1}. It has been shown that the most important motions of proteins[47–49] or large biological structures (such as the ribosome[5,6]) that are associated with their biological function can clearly be identified with only a few of the slowest modes from the GNM. The large-scale changes of protein conformations between ''open'' and ''closed'' forms, or domain swapping in proteins can also be well explained by elastic network models.[9] In the this article, we will show that analysis of the slow modes in the GNM can aid in predicting possible scenarios for the order of contact breaking during mechanically induced protein unfolding. The GNM is the simplest type of elastic network theory, and the model has been extended to allow anisotropic fluctuations,[38] and hierarchical[50] or mixed[51] levels of coarse graining.

To summarize, both GNM and Gô-like models are based on a coarse-grained description of proteins. In the Gô-like model, we use more realistic atomic-level definitions of the contact map; native contacts interact with Lennard–Jones potentials, sequentially neighboring C^{α}s with a harmonic spring-like potential, and non-native contacts with the repulsive part of Lennard–Jones potentials. Within the GNM, the contact map is created by using a uniform cutoff value and all C^{α}s (both bonded and nonbonded) that are close to one another interact with the same harmonic potential.

RESULTS AND DISCUSSION

Stretching of titin in the Gô-like model

The backbone representation of the 27th immunoglobulin domain of the I band of titin (Ig-I27 with the PDB code 1tit) is shown in Figure 1(a). I27 contains 89 residues that form eight β-strands that are commonly indicated by letters A(4–7), A'(11–15), B(18–25), C(32–36), D(47–52), E(5–61), F(69–75), and G(78–88). The numbers in parentheses indicate the residue indices of the first and the last residue in the given β-strand of the protein sequence. The stretching of titin using the Gô-like model has been analyzed in detail in Refs. 12 and [52]. Mechanical unfolding of titin was studied also by Makarov using kinetic Monte Carlo simulations[53–55] and others.[56–59] Here, we summarize the force–displacement (F/d_u) pattern and the order of contact-breaking for I27 obtained from the Gô-like model dynamics simulations,[11,12] when titin is pulled by the N and C termini. The inset in Figure 2 shows the *F*/*d* curve for one domain of titin I27 at zero temperature ($T = 0$), when no thermal fluctuations are taken into account. Two major peaks and a small hump in the first peak are the most important events during protein stretching. Unfolding starts from the small hump on the left side of the highest peak. This hump is created by the rupture of the A–B region,[57] and it corresponds to the intermediate state that was identified by Marszalek *et al.*[17] From this point, the tension inside the protein increases up to the highest peak in the inset of Figure 2. This peak has a height near 4 (in ε/Å units) and occurs due to unraveling of contacts mostly of the highest contact order, primarily between β-strands A + G and A' + G, and inside β-hairpins A + B and A' + B, a few between β-strands B + G, and half of the contacts inside β-hairpin F + G. The second peak is due to breaking contacts between β-strands C + F and B + E. We can also see the breaking of contacts inside the last remaining β-hairpin D + E at $d_u = 239$ Å before the final rise of the *F*/*d* curve, which eventually leads to rupture of the protein backbone. A convenient way to describe the

order of contact-breaking during stretching is to plot the displacement d_u (in Å) at which the given contact breaks against its contact order, where contact order is the sequential distance $|i-j|$ between amino acids i and j, which are in native contact. This is shown in Figure 2, where different colored symbols and letters mark each pair of β-strands in contact. Asterisks indicate contacts, which do not correspond to those between two β-strands.

Analysis of protein stretching with the GNM model

Cutoff distance and atomic overlap methods for defining the contact map—Our studies are based on coarse-grained models of proteins. This simplifies a complex structure and makes it treatable, but on the other hand introduces limitations for describing the interactions that stabilize protein structures. A description of a protein structure in terms of its amino acid contacts in the native state reflects the most basic and natural properties of proteins. For simple models that use the structural information encoded in contact maps, defining these contacts is the most basic property that influences the agreement found here between theoretical results and experiments. There are a variety of ways in which the amino acids could be defined as forming contacts in the native state. With the GNM, it is common to use a uniform cutoff value of the distances between C^{α}s, but the Gô-like model is based on a more detailed contact map defined at the atomic level. It was shown in a detailed study based on a set of 113 proteins[40] that the optimal cutoff value for the GNM depends on the shape and size of the protein, and the best results for the whole set were obtained for $R_c = 7.3$ Å . However, this study was based on averaging over different protein classes. In our case, for this specific medium-sized β-sheet protein, the best results are obtained for $R_c = 7$ Å .

We calculate mean-square fluctuations for each C^{α} in a protein with the GNM theory based on this contact map (for $R_c = 7$ Å) and compare them with the crystallographic Debye–Waller factors (temperature *B*-factors) for I27. From Eqs. (6) and (8), we find that the universal spring constant $\gamma = 3.8$ kcal mol^{-1} Å $^{-2}$ gives the best agreement with experimental *B*-factors for I27. In the case of the Gô model, beads along the main chain are connected by springs with a spring constant $k = 100\ \varepsilon/$Å ^{2}.

Figure 3(a) shows a comparison of the contact maps used in the Gô-like model based on overlaps of the van der Waals spheres of all heavy atom according to Ref. 28 (part of the plot below the diagonal) with the contact map used in the GNM (plotted above the diagonal) obtained with a uniform cutoff of 7 Å for only C^{α} atoms. The correlation between these two maps is 0.8. The structure of I27 contains nine pairs of β-strands. Using the cutoff $R_c = 7$ Å leads usually to the loss of none or one contact inside each pair of β-strands. However, in the case of the pair of β-strands B + G, we lose all five contacts. These contacts correspond to those between side groups of a small amino acid (Ala, Leu, Ile) and a large one (Phe, Ser). The cutoff value of 7 Å is not large enough to detect such contacts that are easily identified with the atomic map. If we increase the cutoff value, similarity between these maps increases rapidly but we also obtain many new contacts, which do not exist in our atomic version. For $R_c = 8.2$ Å , contacts for B + G β-strands start to appear. This is seen in Figure 3(b) obtained with the cutoff value $R_c = 8.5$ Å . Fortunately, these contacts are rather unimportant for the mechanical stretching of titin, as is shown later.

We have applied both types of contact maps to calculate *B*-factors resulting from GNM theory. We have found that *B*-factors computed from the atomic map correlate with experimental *B*-factors at a value of 0.71, while those calculated using the original GNM contact map with $R_c = 7$ Å correlate with experimental data at 0.77.

Detailed results are plotted in Figure 3(c) that compare experimental *B*-factors (dotted curve) with those computed using the cutoff $R_c = 7$ Å (solid curve) and in Figure 3(d) are displayed the same experimental data (dotted curve) and *B*-factors resulting from the atomic map (solid

curve). Additionally, residues that are part of the secondary structure (β-strands) are marked in Figure 3(c,d) with solid dots, and those in the less structure-regular parts of the protein (coil) are shown as asterisks. This shows that residues in β-strands are less mobile and better modeled by GNM theory than those belonging to the coil. Additionally, we see that GNM identifies well those residues that are more exposed to solvent, like the loop between β-strands B and E, and B and C. These residues have significantly larger fluctuations than the rest of the structure, which agrees with the experimental *B*-factors.

By increasing the cutoff value R_c, we artificially obtain better agreement between the contact maps of the two models as discussed earlier and shown in Figure 3(a,b). However, at the same time, correlations between theoretical *B*-factors and experimental data slowly decline. Additional contacts that arise due to the increase in R_c diminish the relative stiffness of protein secondary structure elements. For example, β-strands B and E create superfluous contacts with other parts of the structure that do not exist in a native state, and are absent in the atomic map. It should be noted that our recent[39] work has shown the occurrence of two maxima: one around 7 Å reported in Ref. 40 and a second one near 11 Å .

Our studies demonstrate that less detailed contact maps based on the universal cutoff actually lead to better predictions of *B*-factors, and the large-scale fluctuation motions inside proteins. They imply that the best cutoff value for the GNM model of titin is $R_c = 7$ Å , because it gives the best agreement with experimental *B*-factors for the whole protein and for all β-strands alone similarly to what is shown in Figure 3(c). We have also examined ten other protein structures having diverse sizes, shapes, and belonging to different protein classes: 1crn, 1aqb, 5pti, 1ubq, 1dz3, 1pga, 1ido, 5rsa, 1ula, and 1aba. Only for the single case of 1aba is the correlation of the computed *B*-factors with the experimental data better for an atomic-based contact map, than for the coarse-grained uniform cutoff contact map. Overall, this reflects the fact that the coarse-grained models are able to capture the cohesiveness of the protein better than do the atomic models.

For the atomic[28] definition of contacts used in our Gô-like model, the separation between C^{α} atoms in native contacts varies from 4.3 to 12.8 Å . Thus the cutoff $R_c = 7$ Å is intermediate between these values. We will retain their definition of contacts in the atomic Gô-like model because the more accurate potential has strong effects on the folding kinetics[31] and small but noticeable effects on mechanical stretching. Thus, different properties may require different details.

Prediction of contact rupture in few slowest modes—No studies of the mechanical unfolding of proteins with the elastic network theories have previously been reported. The elastic network models require a detailed knowledge of the structure that is available only for crystallized, undeformed proteins, and the fluctuation dynamics derived from these models describes only the protein motions around equilibrium conformations. To overcome this limitation, we use stochastic molecular dynamics with the Gô-like model to obtain several transient conformations of the protein during mechanical stretching. These mechanically induced partially unfolded conformations are analyzed in detail by their first slowest modes in the GNM for possible scenarios of breaking contacts upon further stretching.

Our hypothesis is as follows. According to the assumptions of the GNM model, mean square fluctuations describe oscillations of residues in the elastic network around energy minima. However, during mechanical unfolding, we introduce additional strain into the system. First, the system is moved out of its equilibrium conformation, and second, at some critical moments is lead to the breaking of native contacts. The contact-breaking will take place at the position inside the protein where the elastic strain between a pair of residues is larger than a critical value. Such a pair of residues, for which the strain reaches its critical value, is characterized

by vanishing mean square fluctuations, because oscillations of these residues become less and less free. Nonetheless, the initial direction and the amplitude of fluctuations should be well-described by the slowest elastic modes, which have the largest contribution to the fluctuations of residues $\langle(\Delta R_i)^2\rangle$.

Here, we focus on a few different conformations of I27 during stretching corresponding to different values of the mechanical displacement d_u of the chain end-to-end vector (see Fig. 2): an intermediate state (d_u = 42 Å), the highest force state (d_u = 80.5 Å), conformation between the two largest peaks (d_u = 82 Å), the second peak (d_u = 136 Å), after the second peak (d_u = 159 Å), and the stage of stretching of the last hairpin D + E (d_u = 239 Å). Examples of conformations of titin during the various stages of stretching are shown in Figure 1. The results reported below are scaled by the unit value of the spring constant γ, which equals 1 kcal mol^{-1} $Å^{-2}$. The absolute amplitude of motion in each normal mode is arbitrary.

Intermediate state (d_u = 42 Å)—Figure 4(a) shows mean-square fluctuations as a function of the position of residues along the sequence computed from Eqs. (6) and (9) for the first slowest mode in GNM. This mode is the single slowest mode of motion in the I27. We find that amplitudes of fluctuations for a few residues are much larger than that for other parts of the structure. The highest fluctuations correspond to the pairs of residues 1–77(G) and 14(A')–89(G) that were in contact in the native state, but these contacts have already been broken in an early phase of stretching. These contacts appear to break first (at the smallest d_u) in the unfolding scenario shown in Figure 2. Furthermore, we see large fluctuations of residues 25, 26, 64, 65, 66, 67, 75, and 76. They arise because these amino acids have no contacts with the rest of the structure or experience a lot of fluctuational freedom; they lie in the unstructured part of protein. Most importantly, the slowest mode shows several characteristic minima in Figure 4(a) where tension in the structure is the highest. One of the smallest fluctuations corresponds to residue 23 in β-strand B, which has a hydrogen bond with residue 4 in β-strand A. This contact will be broken in the next step of the molecular dynamics simulation of the intermediate state slightly above d_u = 42 Å . Nonetheless, residue 4 has relatively large fluctuations because it is located close to amino acids 1 and 2, which broke contacts with the rest of the structure and fluctuate a lot. Minima inside β-strands G + F and B + E that were discussed earlier correspond to the shear force between two planes stretching in opposite directions. The conformation of I27 in the intermediate state and the position of the residue that experiences the highest tension in the protein are shown in Figure 1(b).

The highest force (d_u = 80.5 Å)—Now we take a closer look at the structure exactly at the maximum of the force (d_u = 80.5 Å in Fig. 2) that leads to the disruption of contacts between β-strands A + G, A' + G, A + B, B + G, and F + G in the Gô-like model. Figure 4(b) shows the fluctuation of atoms in the slowest mode. We see that the β-strand A' is the only one where all residues have noticeable fluctuations. All red crosses belonging to A' are located significantly above the abscissa, differently from other β-strands that have some parts nearer to the abscissa. Amino acids from the A'-strand have no contacts with the rest of the structure, and therefore exhibit noticeable fluctuations. If we take a closer look at the pair of β-strands A + G, all contacts between these two β-strands have been disrupted at earlier stages of stretching. Residues belonging to the β-strand G show two different types of behavior that can be categorized as fluctuating residues and nonfluctuating ones forming a ''hinge point.'' This hinge point is the result of (still unbroken) hydrogen bonds between amino acids 78, 79, 80, and 81 of β-strand G and residues 73, 74, and 75 of β-strand F [indicated by arrows in Fig. 4 (b)]. Residues connected by these hydrogen bonds form the hinge point, associated with the minimum of mean-square fluctuations; the remaining part of β-strand F outside the hinge point fluctuates significantly. Our observations agree well with an unfolding scenario (see Fig. 2) where only half the contacts of the hairpin F + G are broken at d_u = 80.5 Å and the other part will be disrupted during the next stage of unfolding.

Similar behavior is observed for residues of the β-hairpin A + B. All contacts between β-strands A and B have been disrupted earlier. Strand A is almost completely flexible, but residues from B18 to B25 are still immobile because of remaining contacts with strand E. The second mode at the highest force stage of unfolding (not shown) is difficult to interpret. The third mode (not shown) demonstrates that whole strands A and A', and parts of G and B that have no contacts are flexible.

Conformation of titin between the highest peaks on the force-displacement plot—After the stage of the highest force in the curve *F*/*d* (see Fig. 2) is completed, more than half of all native contacts have been broken. The disruption of contacts causes relaxation of the remaining part of titin to a new equilibrium position, which will act strongest against mechanical stretching. This we will observe with the rotation of remaining part of titin. Firstly, the structure will attend to conformation in which hydrogen bonds between the β-strands C + F, B + E, and D + E will be in perpendicular direction to the end-to-end vector (this is a conformation, which enforces the biggest force to unravel structure, unzipping costs less force). Secondly, this leads to the shearing β-strands what we will see finally as the second peak. For displacements from $d_u = 82$ Å up to $d_u = 100$ Å , we observe constant breakage of remaining contacts in the β-hairpin F + G. For further extensions from $d_{u=}$ 100 Å up to $d_u = 120$ Å , only three contacts outside the secondary structure are broken. Figure 4(c) illustrates the contribution of the first (inset) and the second mode, respectively, to fluctuations of residues at $d_u = 82$ Å . We see that the nature of these curves for the slowest modes changes entirely. The plot of the first mode has only one clear and sharp minimum corresponding to the hinge point, whereas other parts of the structure exhibit fluctuations at much larger levels than before. The minimum is observed for residue 13, which is in contact with residue 17. This is one of the contacts outside the secondary structure, which will be broken upon further extension, before reaching the second peak on the *F*/*d* plot. The second mode [Fig. 4(c) down] shows which part of the secondary structure will be disrupted next—these are the remaining hydrogen bonds of β-hairpin F + G (indicated by arrow).

We think that the geometry of contacts is responsible for the significance of the order of modes. Hydrogen bonds within the β-hairpin F + G are perpendicular to the applied external force, while the contact 13–17 has a substantial directional component along the pulling force. As a result, all contacts within the β-hairpin F + G are seen in the second mode, but they will be broken before the 13–17 contact. Up to the moment when all contacts between F + G are disrupted, the shapes of *B*-factor plots in the first two modes are similar, regardless of the extension d_u. In the first mode, we still have a sharp minimum at residue 13, but with different depths. The second mode plots have one or two minima, which correspond to different residues in the β-strands F or G.

The second peak—The second peak on the force–displacement plot (see Fig. 2) during stretching is created mostly by breaking contacts between β-strands C + F and B + E, and contacts outside the secondary structures. For extensions from $d_u \approx 120$ Å up to $d_u = 142$ Å in the first mode, we observe a hinge point created by amino acids 72 and 73 belonging to the F strand and residues 31, 32, and 33 of the C strand (data not shown). The second mode has two sharp minima with significant fluctuations between them that correspond to the unstretched part of titin. These minima are usually located at the first or the second amino acid before and after the unstretched part of titin. There is also a small minimum, less deep than the two mentioned previously, for amino acid 32 in the C strand (data not shown). When more than half of titin is unraveled, stretched part looks like a stationary wave. Thus, the higher modes exhibit such a stationary wave behavior close to chain ends. Odd modes 3, 5, and 7 have quite large harmonic-like oscillations at the end of the titin sequence, while even modes 4, 6, and 8 contain such oscillations at the beginning of the sequence (data not shown). For higher modes, we observe damping in the amplitudes of these oscillations.

Conformation after the second peak—The second peak on the force–displacement plot (see Fig. 2) during stretching has two maxima. The second maximum is created by breaking contacts outside the secondary structure, which are shown as stars around $d_u \approx 158$ Å and contacts between B + E indicated by triangles around $d_u \approx 160$ Å in Figure 2. Fluctuations in the slowest mode exactly at $d_u = 159$ Å are shown in Figure 5(a). In this case, we see two clear minima, which correspond to Leu36 and Glu70. In the next time-step exactly this contact will break.

The last step of stretching—After the second peak in the plot F/d_u, almost the whole structure is stretched. At this point, titin contains only the single secondary structure element —the β-hairpin D + E and a few contacts of short contact order along the chain. In the Gô-like model, contacts within β-hairpin D + E start to break at the displacement of chain ends $d_u = 235$ Å . We have chosen the conformation of titin at $d_u = 239$ Å for more detailed analysis. Figure 4(d) shows the first slowest mode for 1tit at $d_u = 239$ Å . The first mode has one deep broad global minimum for amino acids from 33 up to 40 and two very shallow broad local minima for amino acids inside the β-hairpin D + E. The global minimum is much deeper than those two local ones, because it corresponds to contacts of short order along the chain. For such short-order contacts, the external force acts directly on contacts and keeps this part of the structure under a constant strain. However, we see clearly tension exerted on several residues within the β-hairpin D + E. According to the unfolding scenario shown in Figure 2, we see that residues 49, 50, 51, and 52 in β-strand D and 55, 56, 57, and 58 in β-strand E are under tension. The remaining amino acids 60, 61, and 63 in β-strand E, which previously had contacts with other amino acids outside the β-hairpin, are completely contact-free and exhibit higher fluctuations. Higher order modes show string-like oscillations like those discussed previously for the case of the second peak.

We see that only for the one case of $d_u = 159$ Å do the mean square fluctuations show two clear minima where $\langle (\Delta R_i)^2 \rangle$ are almost zero. In other cases, we can identify instead clusters of potential pairs of residues forming contacts that will break in the next step during the mechanical unfolding. This lack of specificity of order may be explained by the fact that all spring constants in the GNM model are the same. We have modified the GNM model with a random assignment of three different spring constants for the protein elastic network and checked the results for a few transient conformations studied earlier for stretching. For the intermediate state $d_u = 42$ Å , two of the additional four degenerate minima have disappeared. For other values of d_u, we obtained similar results. The use of the uniform spring constant affects the precise identification of the time, at which the pair of residues whose contact to be broken crosses over from the elastic regime to the inelastic one. One should note that normal modes resulting from the ideal harmonic description of the protein cannot correctly explain conformational changes at the moment of breaking the native contacts. Also, we should notice that for all stages of stretching discussed here the GNM provides us only with the magnitude of fluctuations and not with their directions. Directions of motion can be obtained from anisotropic network model (ANM).[36] However, even a small stretching of titin leads to a large number of zero eigenvalues (−30), many more than the six zero eigenvalues for a single rigid body, indicating a decomposition into several independent regions of structure. Additionally, we obtain extremely large fluctuations in the amplitude, especially for the first and the second amino acid before the unstretched part of the structure.

The localization of the strain in the structure before contacts break

The tension exerted at each residue in the protein in the absence of the external force is the same. When we start to pull the protein chain with constant speed, the tension is maximal in some parts of structure. Using the Gô model, we calculate the superposition of the forces from Lennard–Jones interactions and the force transmitted through the protein backbone during the

stretching at each residue. We found that in the case of the forces originating from the backbone, the strongest forces are exerted at the residues for which the fluctuations $\langle(\Delta R_i)^2\rangle$ are the smallest. Figure 5 shows the force from the backbone acting at each residue of I27 at d_u = 159 Å . The strongest force inside the unstretched part of I27 (residues between two dotted lines) acts on residues 36 and 70; however, we also see high tension on C^α = 17. This atom is still in short range contact with its neighbors, which are in an unstretched turn, which is also seen on the right hand side in Figure 1(d). We recall that in the case of d_u = 159 Å the mean square fluctuations obtained from the GNM model indicate these same two residues for which $\langle(\Delta R_i)^2\rangle$ is zero or much lower than other fluctuations. In the cases of other transient structures studied here, the strongest force is usually correlated with the smallest mean square fluctuations in a few slowest modes, but not necessarily always for the first slowest mode. The observation that the smallest fluctuations $\langle(\Delta R_i)^2\rangle$ for some particular residues are correlated with the highest forces confirms our hypothesis that these particular contacts are broken in the next step during protein stretching.

Mean-square fluctuations of the distance between residues

Figure 6 shows the mean-square fluctuations of the distance $\langle(\Delta r_{ij})^2\rangle$ between residues i and j as a function of the displacement d_u during stretching for four sample pairs of contacts. Breaking a contact is seen as a sudden increase of $\langle(\Delta r_{ij})^2\rangle$. We find that fluctuations of distance $\langle(\Delta r_{ij})^2\rangle$ between residues, which are in native contact, initially do not change much but then increase suddenly after the rupture of the contact. If two residues do not have other contacts with the rest of the structure, we observe a rapid increase in fluctuations of the distance following the contact-breaking, as for one of A + G contacts at d_u = 81 Å in Figure 6. Fluctuations of $\langle(\Delta r_{ij})^2\rangle$ for contacts that break at the first peak in Figure 2 are especially large. This is observed for the case of several contacts inside β-hairpin F + G and all contacts between β-strands C + F that break after the first peak. We see these huge fluctuations in the distance in Figure 6 for d_u from 100 to 125 Å . These fluctuations exceed the upper limit of the plot and arise from the conformational changes of 1tit between the two peaks (of Fig. 2). However, these contacts are still unbroken according to the Gô-like model, since the distance is less than 2σ, where σ corresponds to the native distance between atoms. Fluctuations of the distance between two sequentially neighboring residues that interact via a strong harmonic potential in the Gô-like model are almost constant, and their magnitudes are much smaller than the fluctuations in the distance of residues interacting through Lennard–Jones potentials.

Stretching of the GFP in the Gô-like model

As a second protein to test the GNM model, we choose green fluorescent protein (GFP) shown in Figure 7. GFP (PDB ID code: 1emb) consists of 238 amino acids. Its structure corresponds to the β-barrel architecture consisting of 11 β-strands, two long loops, and 3 α-helices. The notation describing the secondary structure elements is the same as in Refs. 20, 22, and 23. AFM experiments with the constant velocity show that the mechanical resistance of GFP critically depends on the linkage.[20–22] For six different ways of pulling, GFP (117–182), GFP (3–132), GFP(181, 212), GFP(132–212), GFP(3–212), and GFP(N-C), the maximal unwinding force is around 600, 407, 356, 127, 117, and 110 pN, respectively, at speed around 3.6 μm/s. The numbers in brackets indicate places in the protein where the force was applied. Moreover, for GFP(N-C) and GFP(3-212) several unfolding routes with a different mechanical resistance were found.

Figure 8 shows the unfolding scenario and the force versus displacement plot for GFP(3-132). The maximal resistance of GFP(3-132) is in good proportion to experimental value obtained for different ways of pulling. Unfolding begins with the rupture of H1 helix leading to an intermediate GFPΔα state as in GFP(N-C). Second intermediate state is created by only partial unfolding of β1-strand. The mechanical resistance appears at 172 Å and arises due to the

simultaneous breaking of contacts between β1–β2, β2–β3, and β1–β6 strands. The latter step involves unfolding β6–β5. This causes breaking of contacts between β4 and β9 in the last step of stretching. Moreover, even though we simulated unfolding at zero temperature in the Gô-like model, we observed temporary breaking of contacts between other parts of GFP. This arises from the rotation of the protein relative to the direction of pulling due to breaking of contacts inside two long loops and helices. In the consequence, GFP(3-132) shows much more mechanical resistance to stretching than GFP(N-C) or GFP(3-212).

Prediction of contacts rupture in the slowest modes for GFP

Figure 9(a) shows that temperature *B*-factors computed from fluctuations of residues $\langle(\Delta R_i)^2\rangle$ in the GNM fit quite well with the experimental data (for residues 60–64 temperature *B*-factors are missing in the PDB file). The barrel-like shape of the protein causes that the center of each β-strand is immobile and stable while all other residues, especially those in loops and helices, are much more flexible. The first intermediate state is created by breaking of the contacts between highly flexible loops and the helix. This prevents us from application of the GNM to the prediction of the order of breaking contacts for this state. The second intermediate state is around 64 Å and is created by a partial unfolding of β1, which breaks contacts with β2. In the GNM model for the conformation at 64 Å , we see the smallest fluctuations for amino acids 11, 12, and 34–36, which belong exactly to the β1 and β2 [Fig. 9(b)]. The next step of unfolding is due to breaking of 26 amino acid contacts that corresponds to the force around 4.5 [ε/Å]. In this case to see the order of breaking the contacts we have to average over at least the first three modes, as it was done earlier for the ribo-some.[5,6] Another possibility is to average unfolded structures over smaller time-steps. Prediction of contact breaking in this case is much more complicated, as within a range of 0.5 Å we could see a cluster of contacts, which are going to break, rather than a single contact. Averaging over the slowest three modes at d_u = 172 Å clearly shows that amino acids in the β-strands β1, β2, and β6 are under the highest tension. These contacts are indicated by arrows in Figure 9(c) and correspond to the structure shown in the same panel. After reaching the extension of 173 Å , the unfolding continues due to the breaking of contacts between β2 and β3 strands and temporary breaking of contacts between β3 and β11, and β5 and β6, as shown in Figure 9(d).

DISCUSSION AND CONCLUSION

We have studied the prediction of the contact-breaking order using the GNM model based on the structures that we obtained during stretching of titin(N-C) and GFP(3-132) in the Gô-like model. We compared the two different contact maps that are used in the GNM and Gô-like models. We found that the simple uniform cutoff of R_c = 7 Å gives the best description of native contacts for the GNM and agrees best with the contact map used in the Gô-like model, at least for titin and GFP. Next, we analyzed the order of contact-breaking between the secondary structure elements. We focused on a few of the most important events during stretching and performed analyses of several slowest modes for these partially unfolded structures for titin and GFP.

Our results show that a simple GNM is able to detect a few conformations where the mean square fluctuations in one of the slowest modes have only two very deep minima indicating two residues between which the native contact is going to break at the next time-step of stretching. Moreover, we have found that this is strictly correlated with the highest force exerted by the backbone on these residues. Our prediction of the order of bond-breaking agrees well with the unfolding scenario obtained for the Gô-like model. However, in most cases and especially for GFP, we can identify only clusters of contacts, which will break in the next time-step during protein stretching, rather than particular contacts. These clusters of contacts that are prone to break need to be compared with the unfolding scenarios and the plots of the forces

as a function of time. Identifying only clusters appears to be a direct result of using uniform spring constants for all interactions. Usually, we have to verify these minima by comparison with the unfolding scenario obtained in Gô-like model simulations and the magnitudes of the force acting at these residues during the mechanical stretching.

To show that our results are reproducible, in addition to titin and GFP, we have performed computational stretching on the Gô-like model of G protein, (PDB ID code: 1pga, 56 residues), crambin (1crn, 46 residues), and ubiquitin (1ubq, 89 residues) and examined a series of their low frequency modes. In the case of 1ubq, to predict the order of contact-breaking, it was sufficient to consider the first slowest mode. However, for 1pga and 1crn, we had to average over the lowest three modes (data are not shown).

Results obtained for the stretching of 1tit with the simple Gô-like model lead to the correct prediction, by using the GNM, of which clusters of contacts break in the next step during stretching. We should, however, notice that in the real stretching experiment of I27 only a single peak in the force–displacement curve is observed. We managed to resolve this disagreement by building a refined version of the Gô-like model, which includes side groups represented by additional C^{β} atoms (work in preparation). Taking into account the side groups allows two amino acids to interact by more than one contact. This can be treated as introducing specificity between interacting atoms. We found that these additional atoms make the process of unfolding more cooperative, producing in effect only a single peak during stretching as shown in Figure 10 and still keeping good proportion between the magnitudes of the mechanical resistance for different directions of pulling for GFP.

The results obtained in the present work are based on a standard Gô-like model (without C^{β} atoms), because it has the same number of nodes as the GNM. This version of GNM has been successfully applied in the past to many biological structures.[5,6,49,60] However, we think that by using generalized versions of both GNM and Gô-like models that include additionally side chain representations and more realistic contact maps in both cases, we would be able to predict even more exactly the order of breaking contacts for stretched proteins.

Understanding of the phononic modes in stretched proteins within the GNM would benefit greatly from a precise experimental determination of structures arising at various stages of the process.

We are already witnessing progress in this matter as low resolution information about transient protein shapes are becoming available. At this stage, however, these shapes do not come with sufficiently precise coordinates of the residues. Before the detailed structures become available, we have generated them through molecular dynamics simulations in the coarse-grained model. These structures appear to be adequate to test the workings of the GNM away from the native state.

Acknowledgements

NIH; Grant numbers: R01GM072014, R33GM066387; Grant sponsor: State Committee for Scientific Research in Poland; Grant number: p203B 03225.

References

1. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci USA 2003;100:12570–12575. [PubMed: 14566052]
2. Maragakis P, Karplus M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. J Mol Biol 2005;352:807–822. [PubMed: 16139299]

3. Itoh K, Sasai M. Dynamical transition and proteinquake in photo-active yellow protein. Proc Natl Acad Sci USA 2004;101:14736–14741. [PubMed: 15466708]
4. Ranson NA, Farr GW, Roseman AM, Gowen B, Fenton WA, Hor-wich AL, Saibil HR. ATP-bound states of GroEL captured by cryo-electron microscopy. Cell 2001;107:869–879. [PubMed: 11779463]
5. Wang YM, Rader AJ, Bahar I, Jernigan RL. Global ribosome motions revealed with elastic network model. J Struct Biol 2004;147:302–314. [PubMed: 15450299]
6. Wang YM, Jernigan RL. Comparison of tRNA motions in the free and ribosomal bound structures. Biophys J 2005;89:3399–3409. [PubMed: 16113113]
7. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng 2001;14:1–6. [PubMed: 11287673]
8. Song G, Jernigan RL. An enhanced elastic network model to represent the motions of domain-swapped proteins. Proteins Struct Funct Bioinformat 2006;63:197–209.
9. Kundu S, Jernigan RL. Molecular mechanism of domain swapping in proteins: an analysis of slower motions. Biophys J 2004;86:3846–3854. [PubMed: 15189881]
10. Cieplak M, Hoang TX, Robbins MO. Stretching of proteins in the entropic limit. Phys Rev e 2004;69:011912.
11. Cieplak M, Hoang TX, Robbins MO. Folding and stretching in a Gô-like model of titin. Proteins Struct Funct Genet 2002;49:114–124. [PubMed: 12211021]
12. Cieplak M, Hoang TX, Robbins MO. Thermal effects in stretching of Gô-like models of titin and secondary structures. Proteins Struct Funct Bioinformat 2004;56:285–297.
13. Tskhovrebova L, Trinick J. Properties of titin immunoglobulin and fibronectin-3 domains. J Biol Chem 2004;279:46351–46354. [PubMed: 15322090]
14. Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormo M, Remington SJ. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. Proc Natl Acad Sci USA 1997;94:2306–2311. [PubMed: 9122190]
15. Erickson HP. Protein biophysics—stretching single protein molecules: titin is a weird spring. Science 1997;276:1090–1092. [PubMed: 9173540]
16. Kellermayer MSZ, Smith SB, Granzier HL, Bustamante C. Folding–unfolding transitions in single titin molecules characterized with laser tweezers. Science 1997;276:1112–1116. [PubMed: 9148805]
17. Marszalek PE, Lu H, Li HB, Carrion-Vazquez M, Oberhauser AF, Schulten K, Fernandez JM. Mechanical unfolding intermediates in titin modules. Nature 1999;402:100–103. [PubMed: 10573426]
18. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 1997;276:1109–1112. [PubMed: 9148804]
19. Tskhovrebova L, Trinick J, Sleep JA, Simmons RM. Elasticity and unfolding of single molecules of the giant muscle protein titin. Nature 1997;387:308–312. [PubMed: 9153398]
20. Dietz H, Rief M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. Proc Natl Acad Sci USA 2004;101:16192–16197. [PubMed: 15531635]
21. Dietz H, Rief M. Protein structure by mechanical triangulation. Proc Natl Acad Sci USA 2006;103:1244–1247. [PubMed: 16432239]
22. Dietz H, Berkemeier F, Bertz M, Rief M. Anisotropic deformation response of single protein molecules. Proc Natl Acad Sci USA 2006;103:12724–12728. [PubMed: 16908850]
23. Hyeon C, Dima RI, Thirumalai D. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. Structure 2006;14:1633–1645. [PubMed: 17098189]
24. Fernandez JM, Li HB. Force-clamp spectroscopy monitors the folding trajectory of a single protein. Science 2004;303:1674–1678. [PubMed: 15017000]
25. Szymczak P, Cieplak M. Stretching of proteins in a force-clamp. J Phys Condens Matter 2006;18:L21–L28.
26. Abe H, Gô N. Non-interacting local-structure model of folding and unfolding transition in globular-proteins. 2 Application to two-dimensional lattice proteins. Biopolymers 1981;20:1013–1031. [PubMed: 7225529]
27. Kwiecinska JI, Cieplak M. Chirality and protein folding. J Phys Condens Matter 2005;17:S1565–S1580.

28. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. J Mol Biol 1999;290:253–266. [PubMed: 10388571]
29. Szymczak P, Cieplak M. Stretching of proteins in a uniform flow. J Chem Phys 2006;125:164903. [PubMed: 17092135]
30. Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. Fold Des 1997;2:1–22. [PubMed: 9080195]
31. Cieplak M, Hoang TX. Universality classes in folding times of proteins. Biophys J 2003;84:475–488. [PubMed: 12524300]
32. Gear CW. Numerical integration of ordinary differential equations. Math Comput 1967;21:146–156.
33. Flory PJ. Statistical thermodynamics of random networks. Proc R Soc London Ser A Math Phys Eng Sci 1976;351:351–380.
34. Kloczkowski A, Mark JE, Erman B. Chain dimensions and fluctuations in random elastomeric networks. 1 Phantom Gaussian networks in the undeformed state. Macromolecules 1989;22:1423–1432.
35. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181. [PubMed: 9218955]
36. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79:3090–3093.
37. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908. [PubMed: 10063201]
38. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515. [PubMed: 11159421]
39. Sen TZ, Feng Y, Garcia J, Kloczkowski A, Jernigan RL. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. J Chem Theor Comput 2006;2:696–704.
40. Kundu S, Melton JS, Sorensen DC, Phillips GN. Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys J 2002;83:723–732. [PubMed: 12124259]
41. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins Struct Funct Genet 1998;33:417–429. [PubMed: 9829700]
42. Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins Struct Funct Genet 2000;41:1–7. [PubMed: 10944387]
43. Lu MY, Ma JP. The role of shape in determining molecular motions. Biophys J 2005;89:2395–2401. [PubMed: 16055547]
44. Bahar I, Wallqvist A, Covell DG, Jernigan RL. Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. Biochemistry 1998;37:1067–1075. [PubMed: 9454598]
45. Haliloglu T, Bahar I. Structure-based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data. Proteins Struct Funct Genet 1999;37:654–667. [PubMed: 10651280]
46. Temiz NA, Meirovitch E, Bahar I. *Escherichia coli* adenylate kinase dynamics: comparison of elastic network model modes with mode-coupling N-15-NMR relaxation data. Proteins Struct Funct Bioinformat 2004;57:468–480.
47. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL. Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry 2002;41:491–501. [PubMed: 11781087]
48. Keskin O, Durell SR, Bahar I, Jernigan RL, Covell DG. Relating molecular flexibility to function: a case study of tubulin. Biophys J 2002;83:663–680. [PubMed: 12124255]
49. Navizet I, Lavery R, Jernigan RL. Myosin flexibility: structural domains and collective vibrations. Proteins Struct Funct Genet 2004;54:384–393. [PubMed: 14747987]
50. Doruker P, Jernigan RL, Bahar I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. J Comput Chem 2002;23:119–127. [PubMed: 11913377]

51. Kurkcuoglu O, Jernigan RL, Doruker P. Collective dynamics of large proteins from mixed coarse-grained elastic network model. QSAR Combinat Sci 2005;24:443–448.

52. Cieplak M. Mechanical stretching of proteins: calmodulin and titin. Phys A Stat Mech Appl 2005;352:28–42.

53. Eom K, Li PC, Makarov DE, Rodin GJ. Relationship between the mechanical properties and topology of cross-linked polymer molecules: parallel strands maximize the strength of model polymers and protein domains. J Phys Chem B 2003;107:8730–8733.

54. Li PC, Makarov DE. Theoretical studies of the mechanical unfolding of the muscle protein titin: bridging the time-scale gap between simulation and experiment. J Chem Phys 2003;119:9260–9268.

55. Makarov DE, Hansma PK, Metiu H. Kinetic Monte Carlo simulation of titin unfolding. J Chem Phys 2001;114:9663–9673.

56. Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, Smith DA, Perham RN, Radford SE. Pulling geometry defines the mechanical resistance of a β-sheet protein. Nat Struct Biol 2003;10:731–737. [PubMed: 12923573]

57. Cieplak M, Pastore A, Hoang TX. Mechanical properties of the domains of titin in a Gô-like model. J Chem Phys 2005;122:054906.

58. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. Biophys J 1998;75:662–671. [PubMed: 9675168]

59. Paci E, Karplus M. Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci USA 2000;97:6521–6526. [PubMed: 10823892]

60. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592. [PubMed: 16143512]

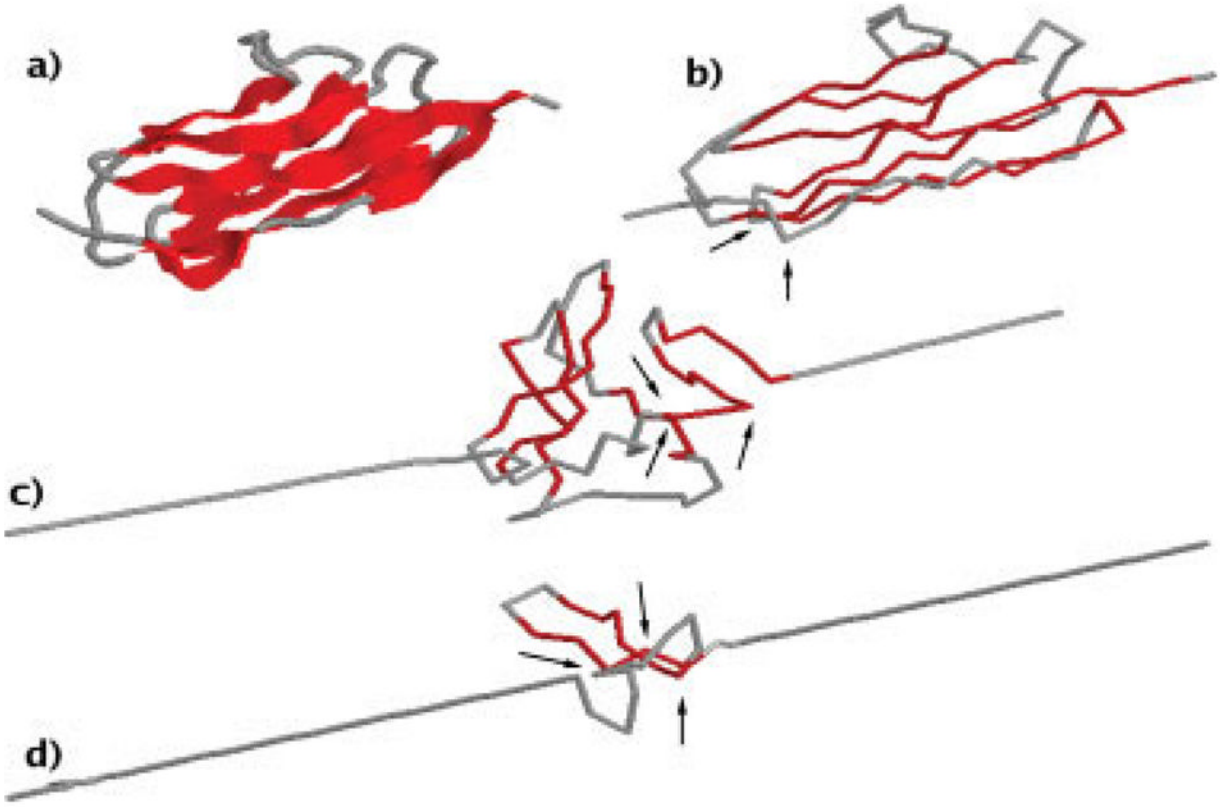

Figure 1.

(a) *The ribbon representation of the I27 domain of titin (PDB code: 1tit) in the undeformed state;* **(b)** *Backbone representation of the stretched 1tit in the intermediate state corresponding to the mechanical displacement of the end-to-*end vector (d_u = 42 Å); **(c)** Structure after the highest force (d_u = 80.5 Å); and **(d)** in the state corresponding to the stretching of the last β hairpin (d_u = 239 Å). See inset in Figure 2 for the location of these points on the force–displacement curve. The arrows indicate the locations interpreted as hinge points for the slowest mode in the GNM. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

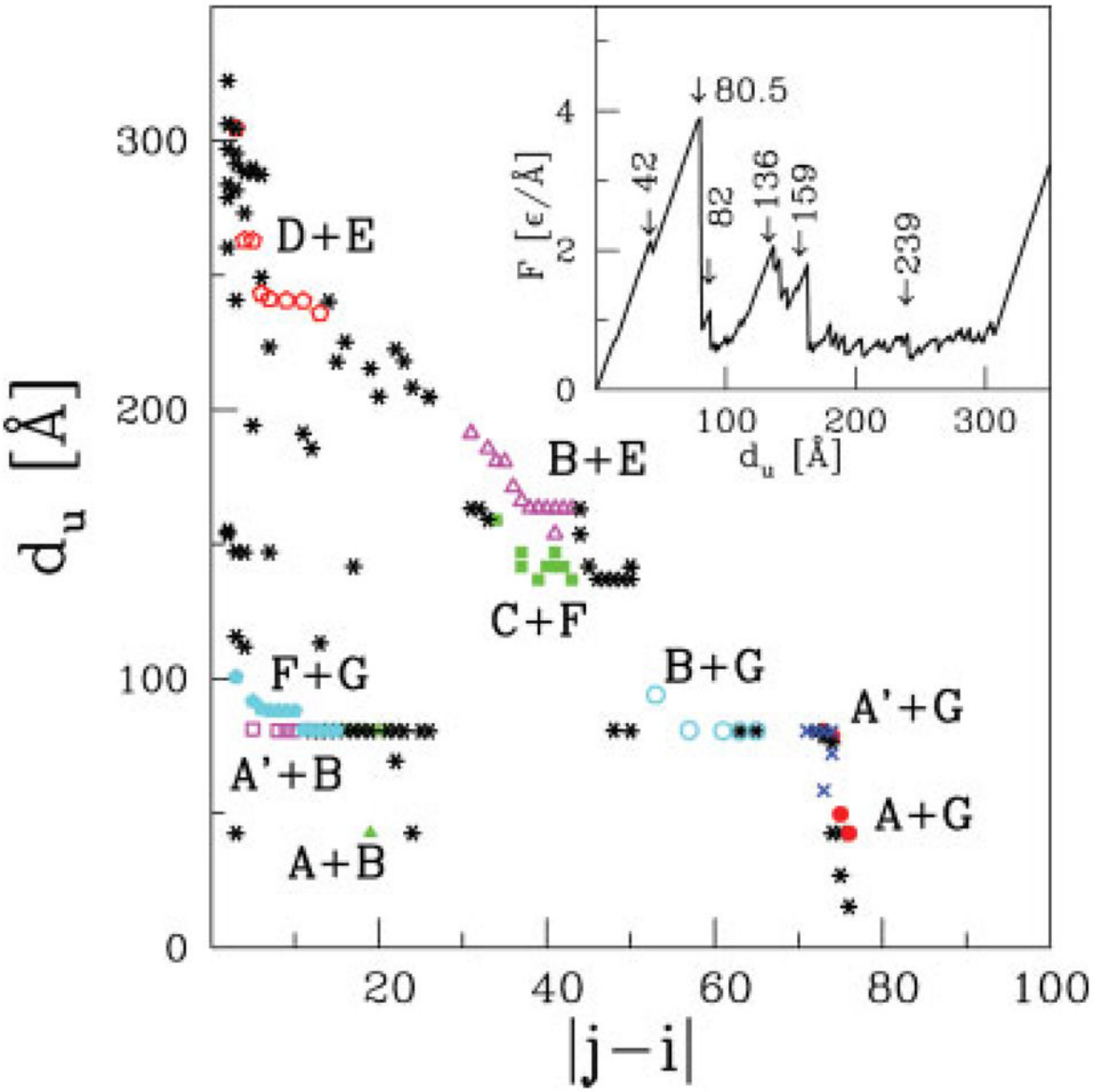

Figure 2.

The inset shows the force–displacement curve of I27 obtained during the computer simulation of stretching in the Gô-like model. Numbers indicate particular displacements (in Å) during stretching that have been chosen for subsequent detailed analysis. The main figure illustrates the unfolding scenario by showing the mechanical displacement d_u [Å] versus the contact order |i – j|. The colored symbols indicate contacts between various β-strands, which are described in detail in the text. Contacts where at least one of the amino acids is not within a β-strand are indicated by asterisks. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

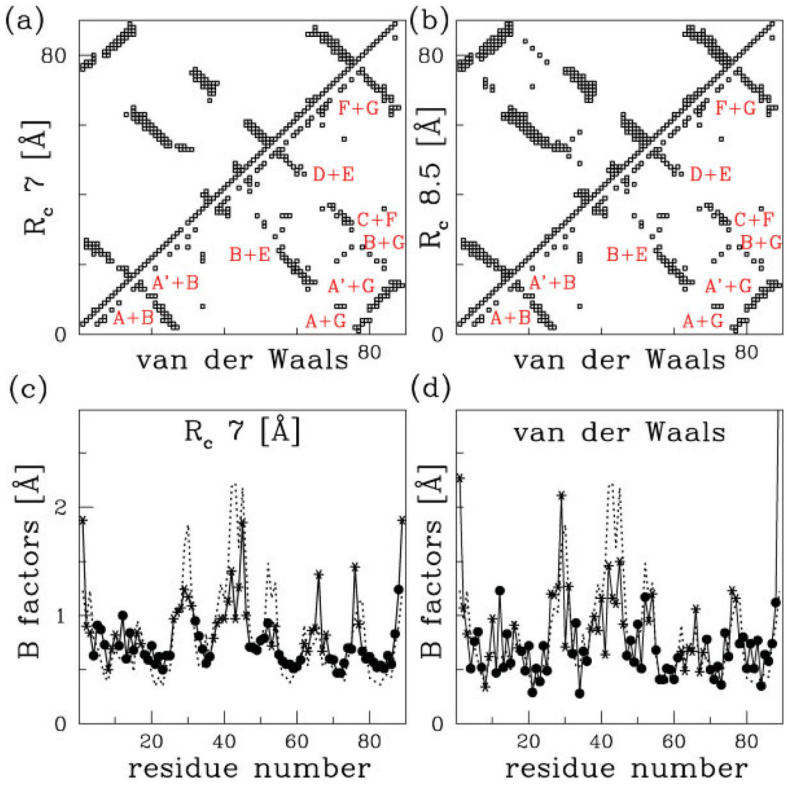

Figure 3.

Top: contact maps obtained for the GNM by using a uniform Cα–Cα cutoff (above the diagonal) compared with more detailed atomic contact maps used in the Gô-like model (below the diagonal), for the native state of 1tit. **(a)** and **(b)** correspond respectively to the GNM cutoffs R_c = 7 Å and R_c = 8.5 Å . Bottom panels (c and d) show temperature factors (B-factors) for 1tit. Dotted and solid curves represent experimental data and GNM predictions, respectively. In **(c)**, the solid curve corresponds to the uniform cutoff (R_c = 7 Å) criterion used to define the contact map. The solid curve in **(d)** corresponds to the overlap criterion of van der Waals spheres for all heavy atoms in the definition of the contact map. The solid dots indicate residues within β-strands, and asterisks all other residues. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

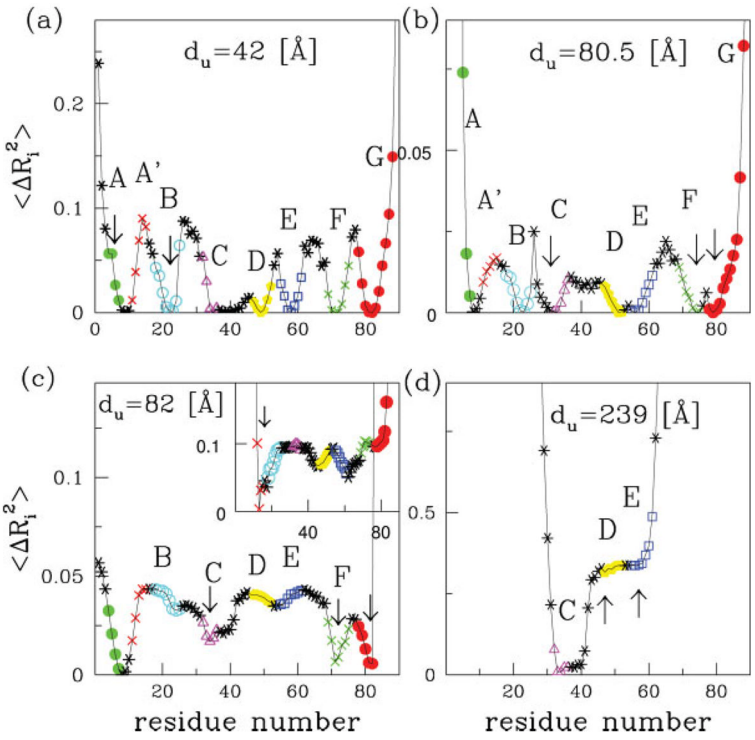

Figure 4.

Mean square fluctuations $\langle(\Delta R_i)^2\rangle$ of amino acids in the slowest modes for particular displacements d_u during stretching. Color symbols and letters indicate amino acids inside specific β-strands, while asterisks denote amino acids belonging to coil parts. The coloring scheme of β-strands is the same as in Figure 2. Arrows indicate hinge points—positions in the protein structure where the force is the highest. Panels correspond respectively to **(a)** intermediate state, first mode; **(b)** the highest force, first mode; **(c)** structure after applying the highest force in first the mode (inset) and in the second mode (main figure); **(d)** structure when the last hairpin starts to break.

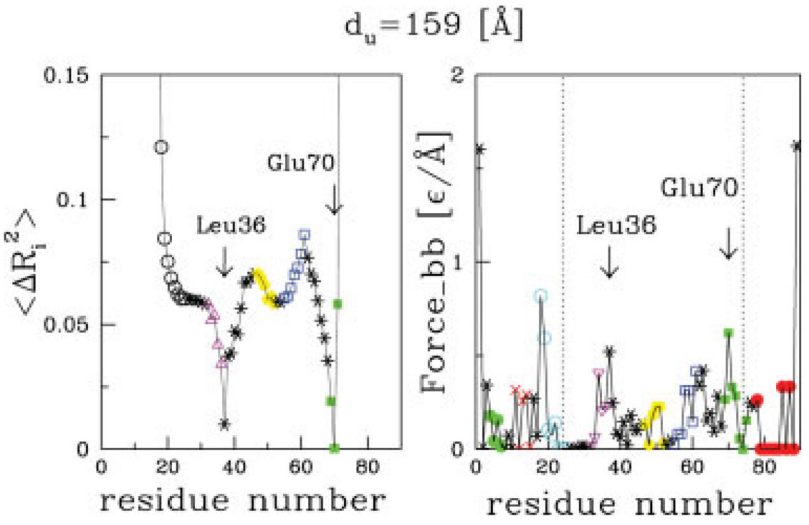

Figure 5.

Left panel: mean square fluctuations $\langle(\Delta R_i)^2\rangle$ of residues at the displacement d_u = 159 Å. Right panel: values of the force exerted by the protein backbone on each amino acid. Arrows in the left panel indicate the smallest fluctuations $\langle (\Delta R_i)^2\rangle$ where the highest forces act on a pair of the amino acids in the right panel. Residues between dotted lines indicate unstretched part of the protein. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

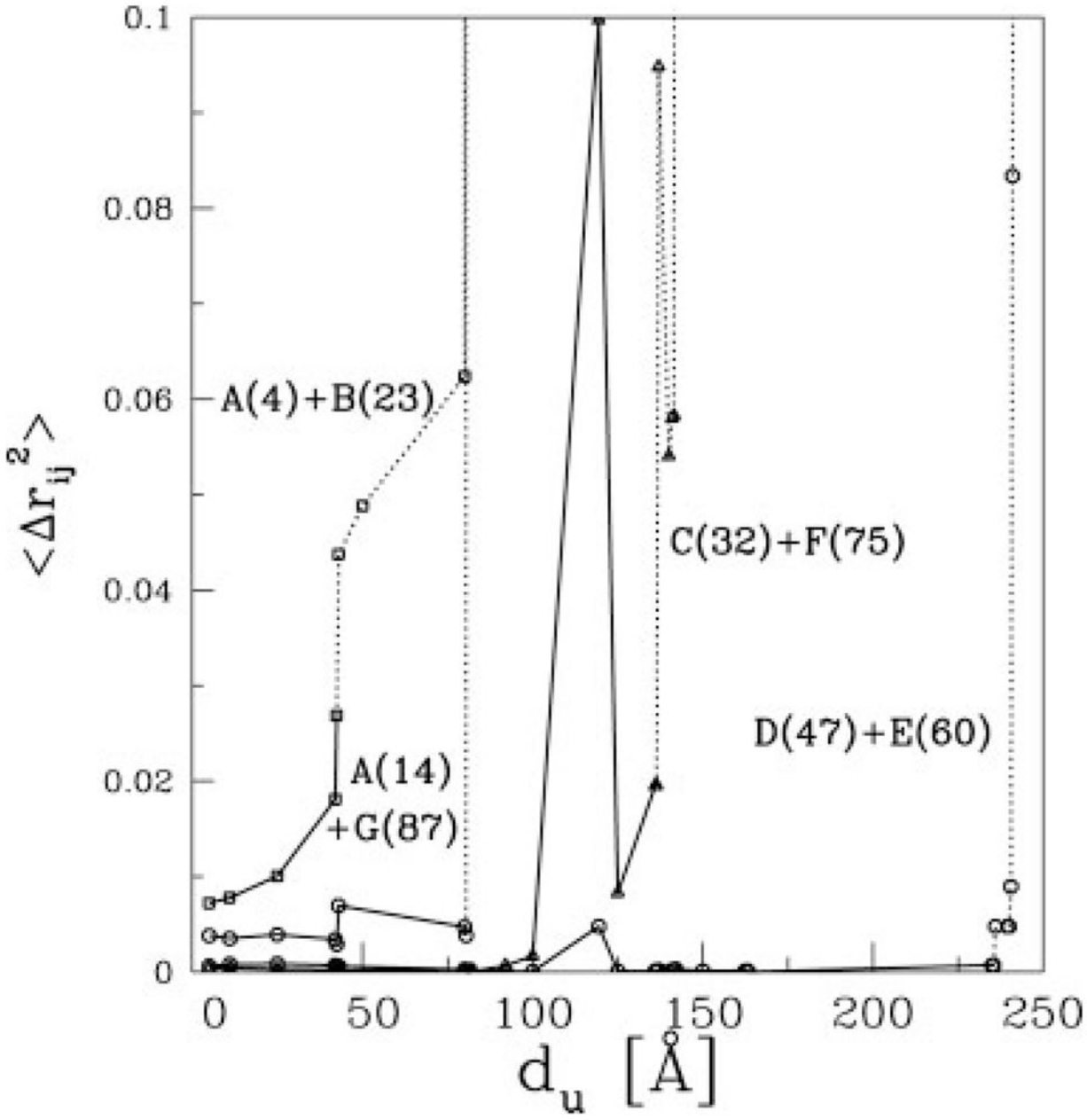

Figure 6.

Mean square fluctuations of the distance $\langle (\Delta r_{ij})^2 \rangle$ between residues i and j, as a function of the displacement d_u during stretching in the Gô-like model. Dotted lines indicate broken contacts, while solid lines show contacts that still exist. Letters with numbers specify β-strands, while numbers indicate particular amino-acids.

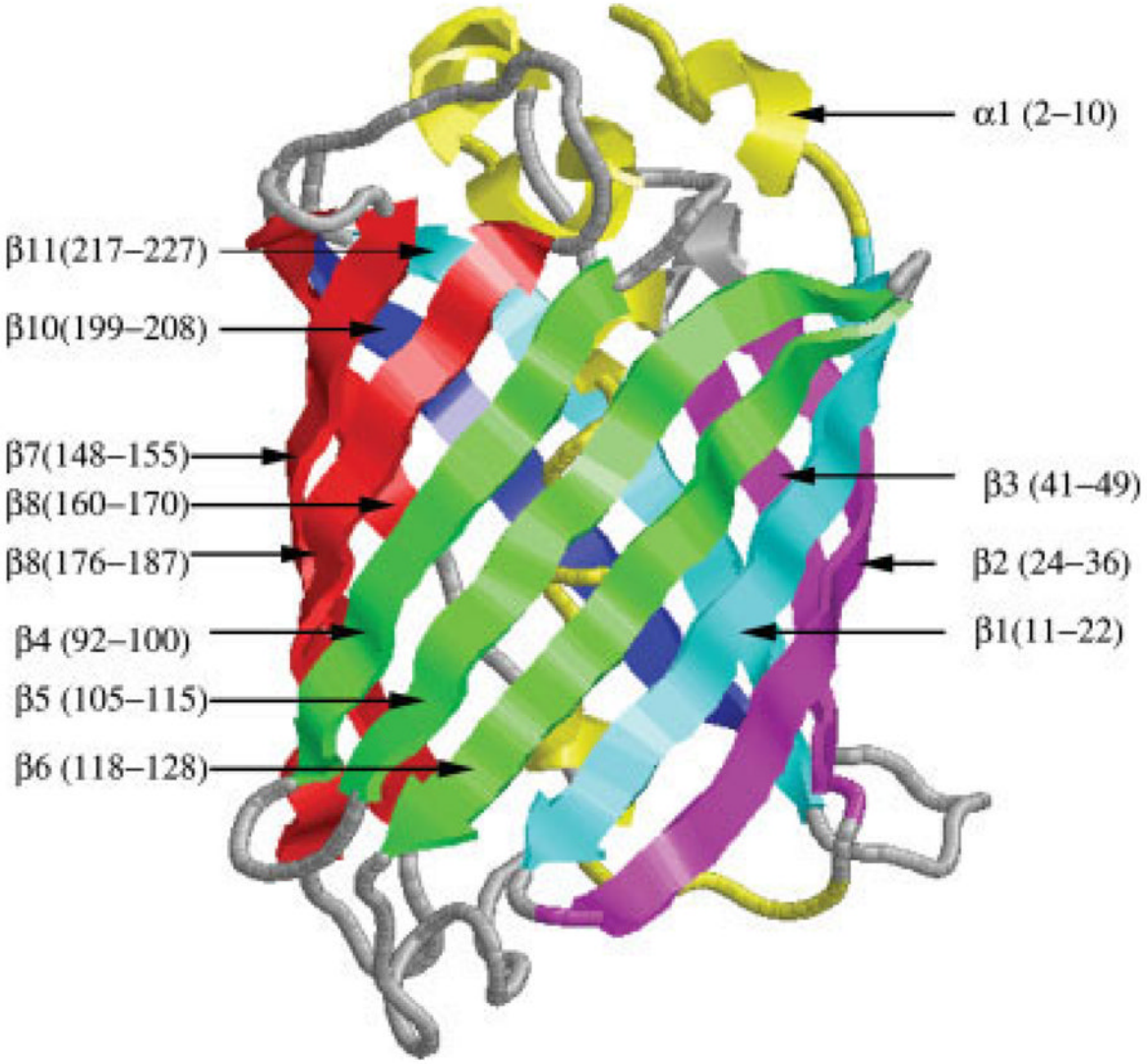

Figure 7.
The ribbon representation of the green fluorescent protein in the undeformed state. The symbols indicate particular β-strands and α-helices, together with the sequence position of the amino acids in each. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

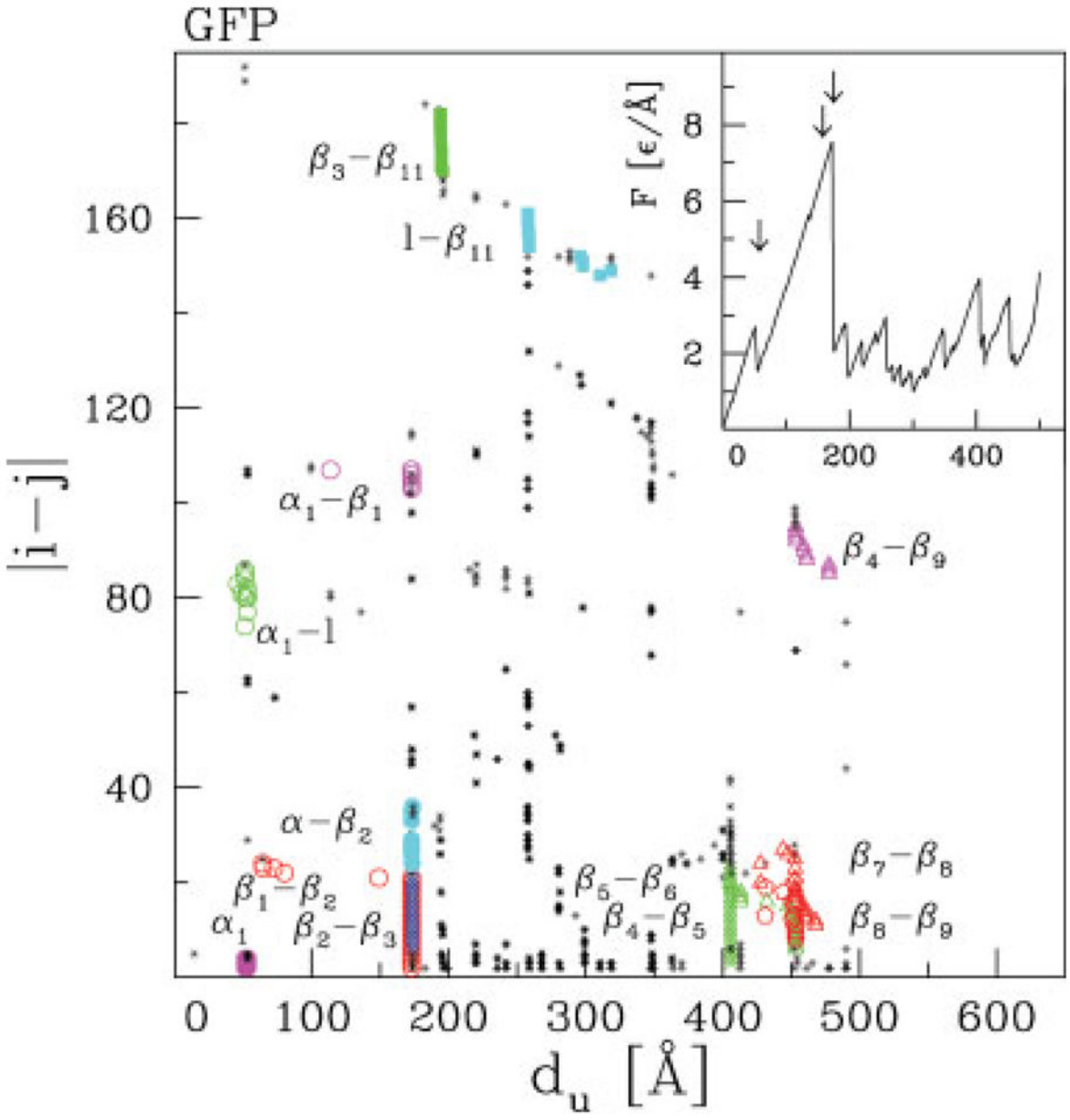

Figure 8.

The inset shows the force–displacement curve of GFP pulled by amino acids 3 and 132. Arrows indicate specific displacements during stretching that have been chosen for the subsequent detailed analysis with the GNM model. The main figure illustrates the unfolding scenario by showing the mechanical displacement d_u versus the contact order |i – j|. Different symbols indicate contacts between various secondary structure elements. Contacts where at least one of the amino acids belongs to a coil are indicated by asterisks. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

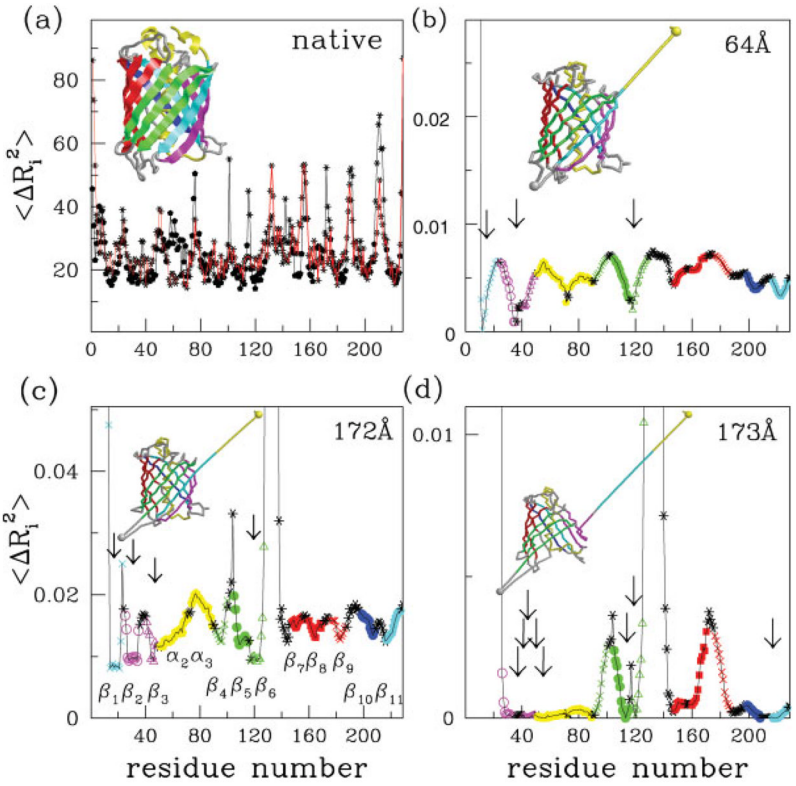

Figure 9.

Mean square fluctuations $\langle(\Delta R_i)^2\rangle$ of amino acids in the slowest modes for particular displacements d_u during stretching. Color symbols and letters indicate amino acids inside specific β-strands, while asterisks denote amino acids belonging to the coil regions. Arrows indicate hinge points—positions in the protein structure where the force is highest. Panels correspond respectively to **(a)** experimental temperature factors (B-factors) from the PBD file and the corresponding values computed from the GNM model; **(b)** intermediate state, first mode; **(c)** the highest force state, averaged over the three lowest modes; **(d)** structure at the highest force, in the second mode. Insets inside each panel show snapshot of the structure corresponding to a given state, where beads indicate amino acids 3 and 132.

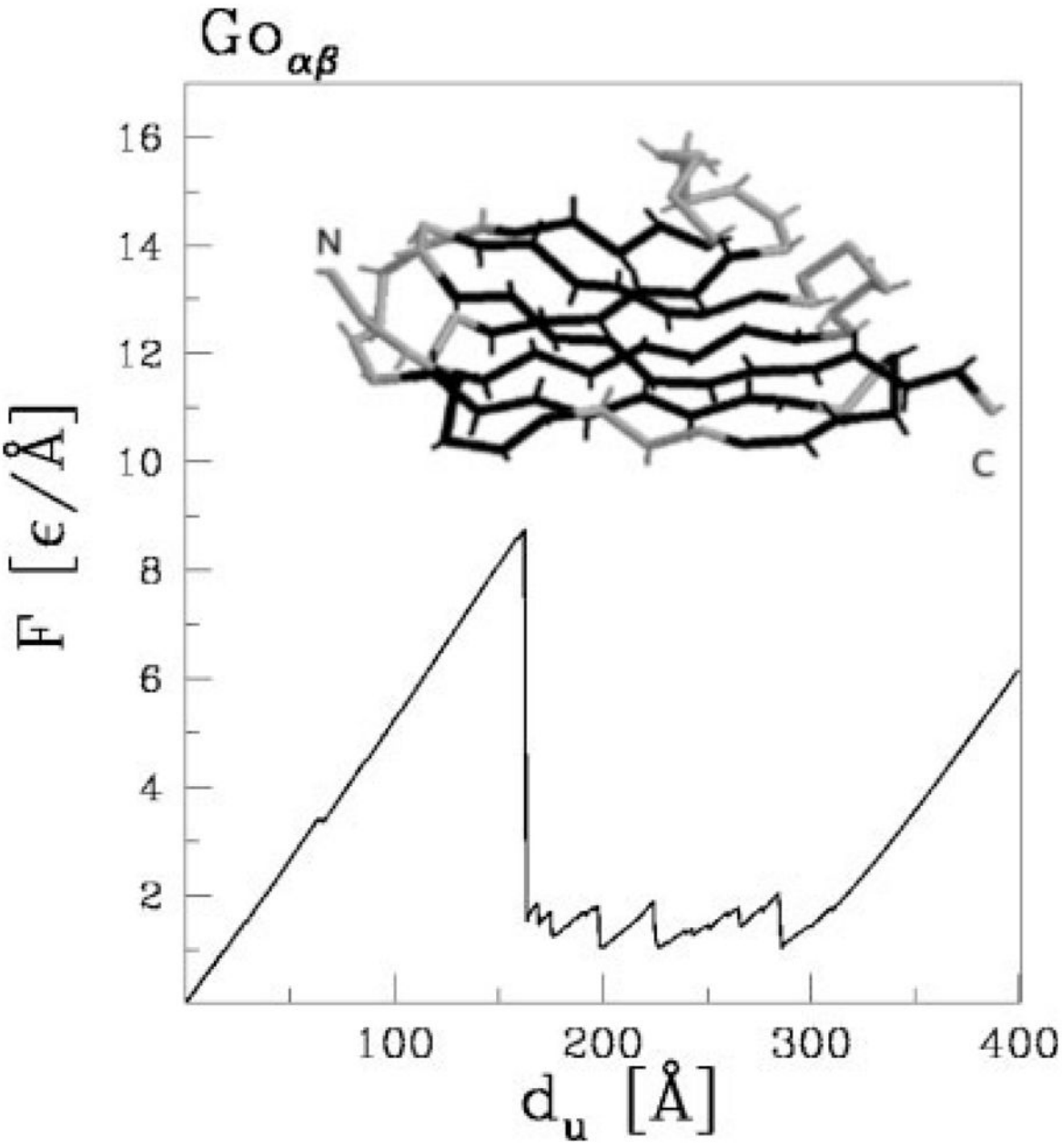

Figure 10.
The stick representation of the I27 and its F–d_u curve obtained during stretching in the Gô αβ-like model based on both the C_α and C_β atoms.