Mechanical stretching of proteins: calmodulin and titin

Marek Cieplak

Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

Abstract

Mechanical unfolding of several domains of calmodulin and titin is studied using a Go-like model with a realistic contact map and Lennard–Jones contact interactions. It is shown that this simple model captures the experimentally observed difference between the two proteins: titin is a spring that is tough and strong whereas calmodulin acts like a weak spring with featureless force–displacement curves. The difference is related to the dominance of the α secondary structures in the native structure of calmodulin. The tandem arrangements of calmodulin unwind simultaneously in each domain whereas the domains in titin unravel in a serial fashion. The sequences of contact events during unravelling are correlated with the contact order, i.e., with the separation between contact making amino acids along the backbone in the native state. Temperature is found to affect stretching in a profound way.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Recent developments in technology have enabled studies of large biomolecules through mechanical manipulation. The most common techniques of such a manipulation involve atomic force microscopy and optical tweezers. These

E-mail address: mc@ifpan.edu.pl (M. Cieplak).

techniques target the hydrophobic and hydrogen bond interactions that are an order of magnitude weaker than those corresponding to the covalent bonds. The simplest protocol used in the studies is to anchor one segment of a molecule to a substrate and pull by another segment at a constant speed, v_p. By monitoring the force of resistance to the pull, F, and plotting it versus the tip displacement, d, one obtains an elastic characterization of a molecule that requires theoretical interpretation in terms of the sequence of the rupturing events. The resulting $F-d$ plots have a fine structure consisting of peaks, minima, and plateaus that depend on the molecule and conditions of its environment. The first systems that were studied through stretching were the streptavidin–biotin complex [1,2], DNA [3–5], and the multidomained titin [6–8] that is found in a class of sarcomers in muscles.

Pulling apart two strands of the DNA involves breaking one hydrogen bond at a time and this makes the force ondulate around the value of 13 pN and with an amplitude not exceeding 1 pN [5]. The typical pulling speed in these experiments is 40 nm/s and a five-fold increase in v_p affects the F–d pattern only very weakly. Separating biotin from streptavidin involves stretching many bonds simultaneously and, within the pulling distance of about 10 Å, the peak force is close to 300 pN [2]. Stretching of titin results in a sawtooth-like pattern [6,9–11] where each tooth is attributed to the unwinding of a single domain and has a peak value of order 200 pN. Titin has been found to act like a non-Hookean spring which is strong, tough and nearly reversible [9]. Its properties have inspired biomimetic design of polymers [12]. The calcium binding C2A protein, on the other hand, has been found to be much weaker: the peak force is only of order 60 pN and its sawtooth-like pattern has more structure within each tooth [13]. For poly-calmodulin, another calcium binding protein, there are no significant force peaks since no cluster of hydrogen bonds undergoes breaking [13] until the molecule is fully stretched. The patterns obtained for modular proteins generally depend on whether the modules are connected end-to-end or away from the terminals, like in the case of ubiquitin [14] or lyzozyme [15,16].

The F–d patterns appear to act like finger prints of biomolecules but they also depend on the temperature, T, as evidenced experimentally [17] and theoretically [18,16]. In this paper, we demonstrate that simple geometry-based theoretical models can capture substantial differences in the F–d curves that exist between proteins and show that these differences diminish on increasing the temperature and disappear in the entropic limit. Our presentation is focused on two proteins: the I27 domain of titin and calmodulin corresponding to the protein data bank [19] codes 1tit and 1cfc, respectively. The former has the architecture of the β sandwich [20] with no α helices and with the β content of 32.65%, whereas calmodulin is mostly an α protein that is shown in Fig. 1: the helical content is 54.05% and the β content is 8.11%.

There are obvious advantages to all-atom modelling compared to simplified coarse-grained models: it offers a more realistic description and its nature is more fundamental. All-atom modelling of 1tit [21] leads to the identification of the hydrogen bonds linking the so-called A′ and G strands as being responsible for the maximum force in this case. There are also equally clear drawbacks that are related to the necessity of dealing only with very short time scales, typically of order nano-
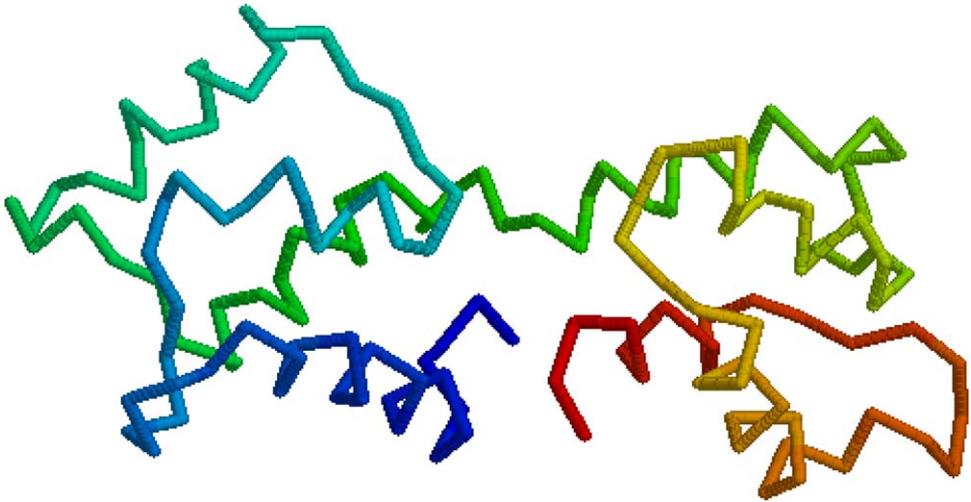
Fig. 1. The backbone representation of the 1cfc structure of calmodulin.

seconds. This results in considering the pulling speeds which are 6–7 orders of magnitude too rapid and which may be responsible for an order of magnitude for big peak forces calculated for titin [21] (another reason for the discrepancy with the experiment could be the surface tension effects due to the droplet of water that surrounds the protein). The models presented here allow for studies that (a) involve more realistic v_p, (b) incorporate tandem connection of several domains, (c) enable comparison to the kinetics of folding, (d) deal with variations of parameters, such as T, and (e) easily compare various proteins.

2. Model

Protein folding is thought to be governed by the geometry of the protein [22–25] and especially by the geometry of its native state [26–28]. One way to incorporate geometry into the model is to follow the prescription of Go [29,30]: construct a Hamiltonian that incorporates the chain-like connectivity and which has a ground state that agrees with the experimentally determined native conformation. Our realization of this prescription within a coarse-grained model that is studied through the techniques of molecular dynamics is outlined in Refs. [31–33].

Briefly, the amino acids are represented by point particles of mass m located at the positions of the C^α atoms. They are tethered by a strong harmonic potential with a minimum at $3.8\,\text{Å}$. The interactions between the amino acids are grouped into contacts of the native and non-native kinds. The distinction is based on taking the atomic representation of the amino acids in the native state and then checking for their possible overlaps. The occurrence of an overlap is determined assuming that the atoms take a spherical space corresponding to the van der Waals radii of the atoms,

enlarged by the factor of 1.24 [34,35] to account for the soft part of the interaction potential. The amino acids (i and j) that are found to overlap in this sense are considered to be forming contacts. These pairs are endowed with the Lennard–Jones potential,

$$V_{ij} = 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{1}$$

such that its minimum agrees with the experimental value of the distance between the C^{α} atoms in the native state. This condition selects a pair by pair value of the length parameter σ_{ij}, whereas the energy parameter ε is kept uniform. ε could be made specific if an understanding regarding the values was reached. It corresponds to many effective non-covalent interactions, such as hydrophobicity and hydrogen bonds, so it should range between 800 and 2300 K. It appears [18] that $\tilde{T} = k_B T / \varepsilon$ of about 0.3, where k_B is the Boltzmann constant, qualitatively reproduces the room temperature elastic behavior of titin.

The properties of the native contacts in the two proteins studied here are illustrated in Fig. 2. There are 209 contacts (89 amino acids) in 1tit and 426 contacts (148 amino acids) in 1cfc. For 1cfc, 71% of the contacts are local—their sequence
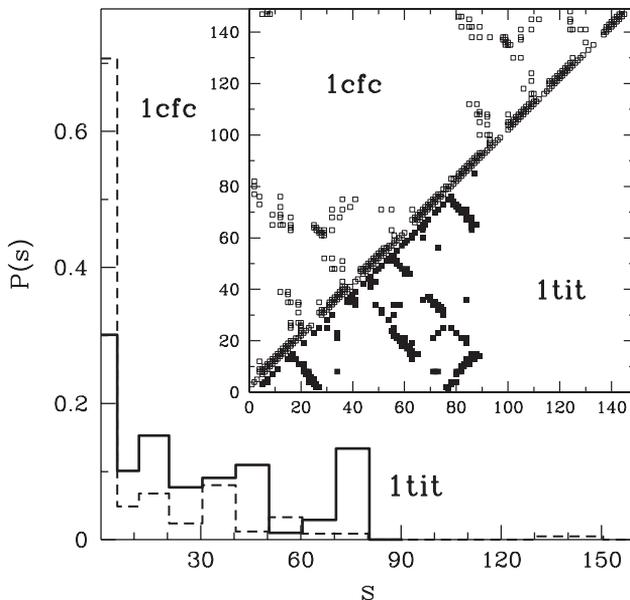
Fig. 2. The distribution of sequential distances in the native contacts. The solid line is for 1tit and the broken line is 1cfc. The bin sizes generally correspond to the distance of 10 except for the very first bin which counts contacts of length smaller or equal to 4 and the second bin which counts contacts with the distance bigger than 4 but not exceeding 10. The inset shows the corresponding contact maps. The contact maps are symmetric and only half is shown for each protein.

distance does not exceed 4. This reflects the high α-content. In contrast, only 30 % of the contacts in 1tit are local. The contact map for 1tit, also shown in Fig. 2, is organised in stripes corresponding to interactions between distinct β-strands. Such patterns are not present in the case of 1cfc suggesting a qualitatively different network of the couplings.

The thermal fluctuations away from the native state are mimicked in the molecular dynamics simulation by introducing the Langevin noise with the damping constant γ of $2m/\tau$, where τ is $\sqrt{m\sigma^2/\varepsilon}$. This corresponds to the situation in which the inertial effects are negligible [33] but a more realistic account of the water environment requires γ to be about 25 times larger [36]. Thus the time scales obtained for $\gamma = 2m/\tau$ need to be multiplied by 25 since a linear dependence on γ has been found [31,32].

Stretching is implemented by attaching both ends of the protein to harmonic springs of spring constant $k = 0.12\varepsilon/\text{Å}^2$, i.e., of order 0.4 N/m, which is typical for atomic force microscopy. The outer end of one spring is held constant, whereas the outer end of the other is pulled along the initial end-to-end vector. Our results are shown for the pulling speed of 0.005 Å/τ, which corresponds to 7×10^6 nm/s. Even though this speed is 3 orders of magnitude faster than in experiments, our previous studies [37,18,38] indicated only small logarithmic corrections on going to still smaller values of v_p.
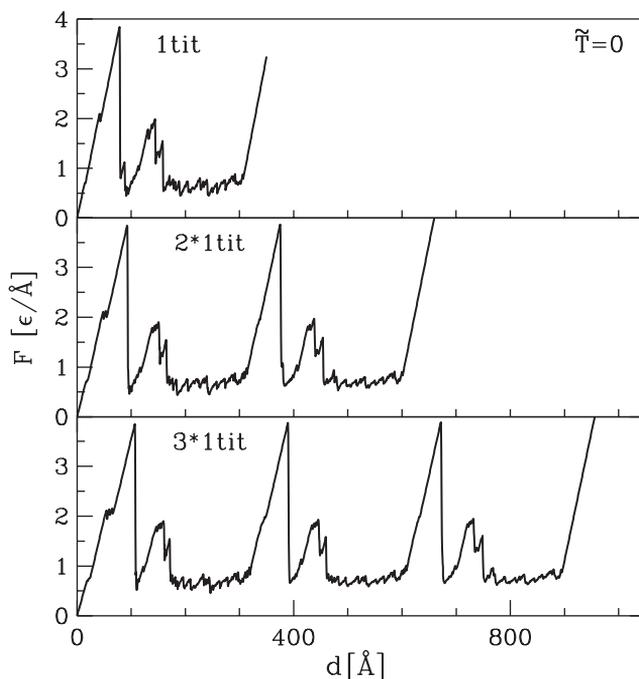
Fig. 3. The force–displacement curves for one, two, and three domains of 1tit, top to bottom respectively, for $\tilde{T} = 0$.
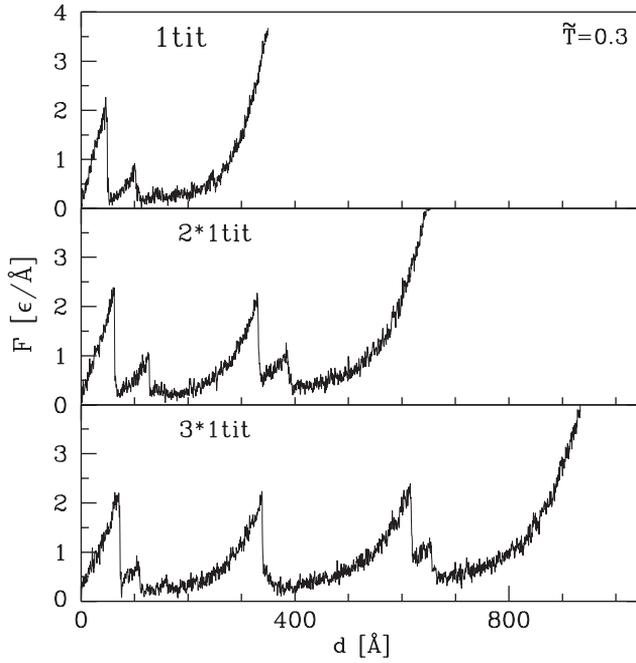
Fig. 4. Same as in Fig. 3 but for $\tilde{T} = 0.3$.
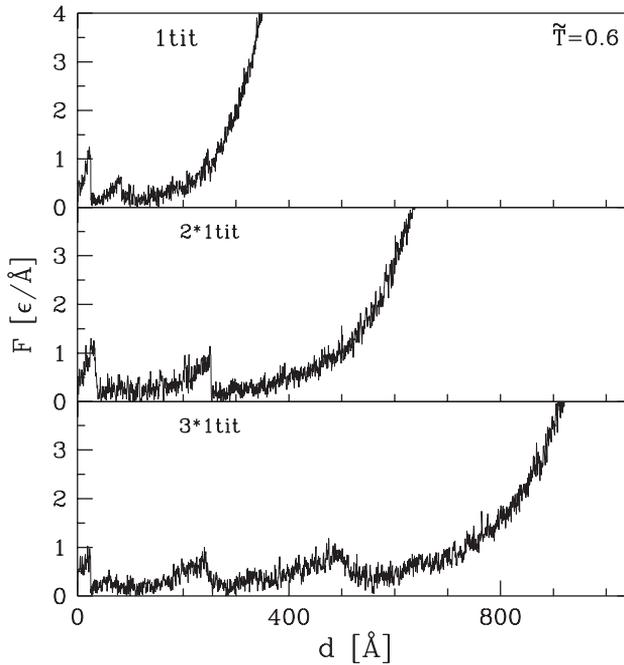
Fig. 5. Same as in Fig. 3 but for $\tilde{T} = 0.6$.

3. The force–displacement curves

Stretching of up to five domains of titin within the Go model has been analyzed in detail in [18]. Here, we present the *F–d* patterns for titin to provide a reference for calmodulin. Fig. 3 shows the *F–d* patterns obtained for one, two, and three domains of 1tit, linked in tandem, at $\tilde{T} = 0$, i.e., when no thermal fluctuations are taken into account. The multidomain patterns are essentially a serial repeat of the single domain curve. The single domain curve has two major force peaks. The first of these has a height of nearly $4\varepsilon/\text{Å}$ and it occurs due to the unravelling of the links between the β-strands that exist at the opposite terminals of the protein. These links are primarily between the strands A′ (amino acids 11–15), A (amino acids 4–7) and G (amino acids 78–88). The second major peak is due to breaking the C–F and B–E links, where B, C, E, and F strands correspond to segments 18–25, 32–36, 55–61, and 69–75, respectively. The small hump on the rising side of the first major peak is due to a rupture in the A–B region [39] and it corresponds to the intermediate state that was identified by Marszalek et al. [10].

Figs. 4 and 5 show the corresponding patterns for $\tilde{T}$ of 0.3 and 0.6. The increase in $\tilde{T}$ results in lowering of the force peaks and making them to occur earlier during the stretching. This is because the thermal fluctuations provide additional unravelling forces. In the entropic limit, reached around $\tilde{T}$ of 0.8, there are no identifiable force
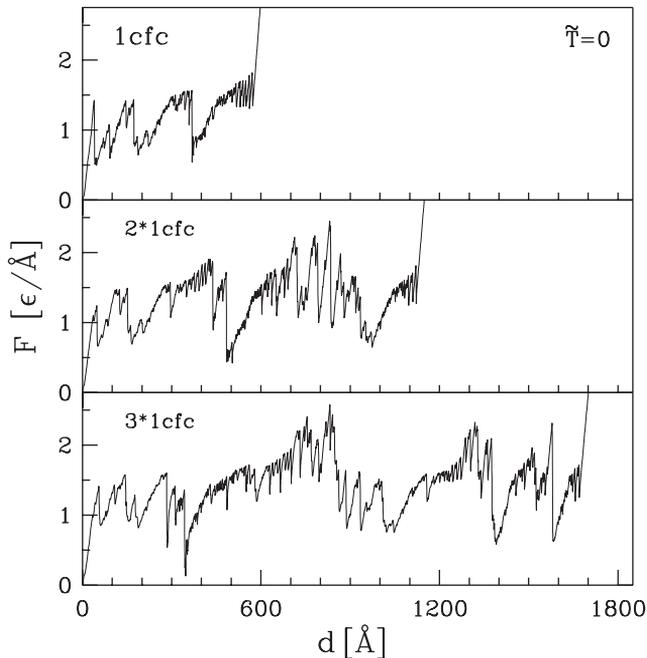
Fig. 6. The force–displacement curves for one, two, and three domains of 1cfc, top to bottom respectively, for $\tilde{T} = 0$.
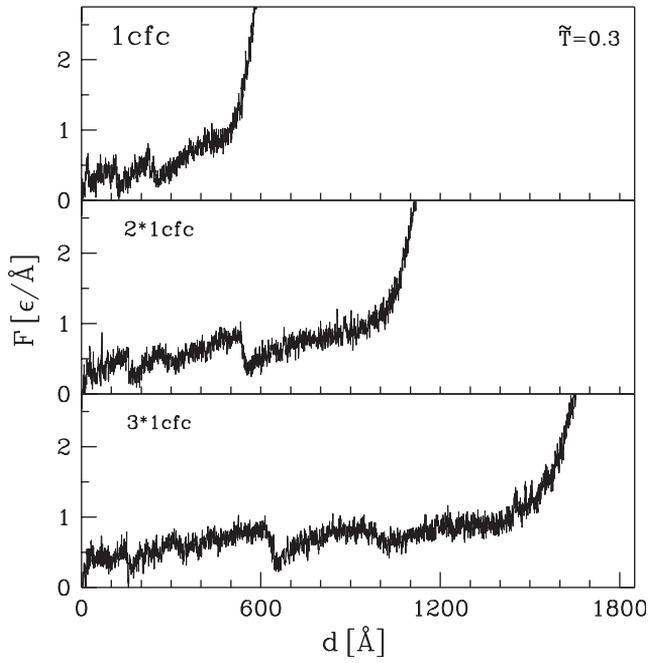
Fig. 7. Same as in Fig. 6 but for $\tilde{T} = 0.3$.
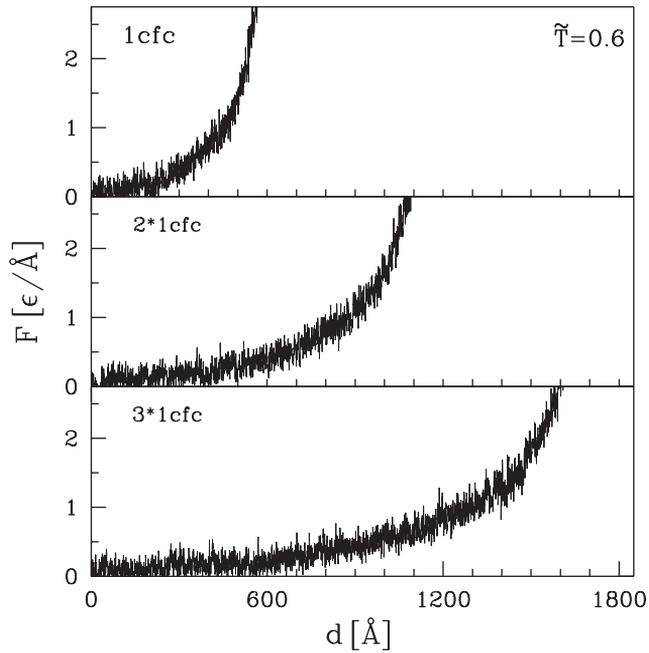
Fig. 8. Same as in Fig. 6 but for $\tilde{T} = 0.6$.

peaks and the *F–d* curves are described by the featureless worm-like-chain model [18,16]. In this limit, the domains unravel simultaneously. At intermediate temperatures, the unravelling is part serial and part parallel. At $\tilde{T} = 0.3$, the stretching of several domains is predominantly serial in character and the $F - d$ curves seem to be qualitatively similar to the saw-tooth patterns obtained experimentally [6,9]. Note that the second major peak force that has been clearly identified in the $\tilde{T} = 0$ trace disappears at $\tilde{T} = 0.3$ except for a weak shadow of it in the first period of the serial pattern.

The multidomain tandem arrangement is constructed so that the C-terminal of one domain is connected to the N-terminal of another by an extra C^{α}–C^{α} bond along the end-to-end direction in a single domain.

The *F–d* curves for calmodulin shown in Figs. 6–8 are the analogs of Figs. 3–5 for titin and they demonstrate an entirely different behavior. The single-domain curve at
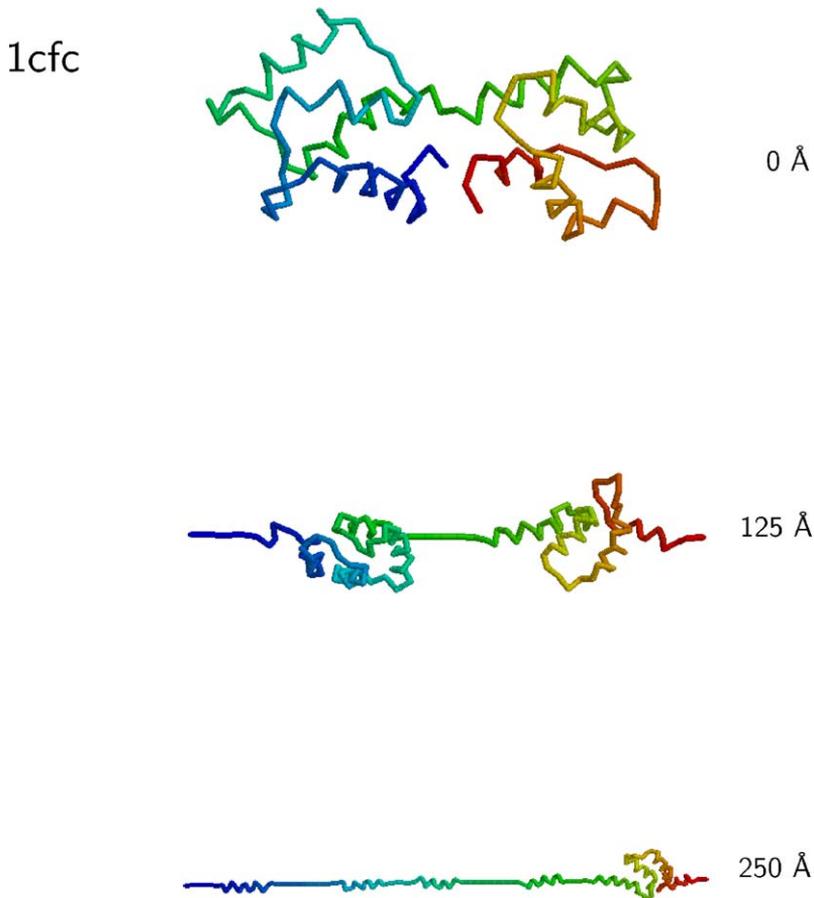
Fig. 9. The conformations of 1cfc during unfolding at $\tilde{T} = 0$. The labels on the right-hand side indicate the corresponding values of *d*.

$\tilde{T} = 0$ displays only minor force peaks. The biggest force (before the stage of fully stretched conformation is reached) is only about $1.5\varepsilon/\text{Å}$—less than 40% of the maximum force found in titin. Furthermore, the multidomain curves are not serial repetitions of the single domain result. Instead, the particular features in the plot get effectively multiplied in segments, indicating a large degree of parallelism in the unwinding. Our studies [40] of two helices connected in series have indicated that they unwind in parallel. Thus the behavior found for calmodulin echos this finding.
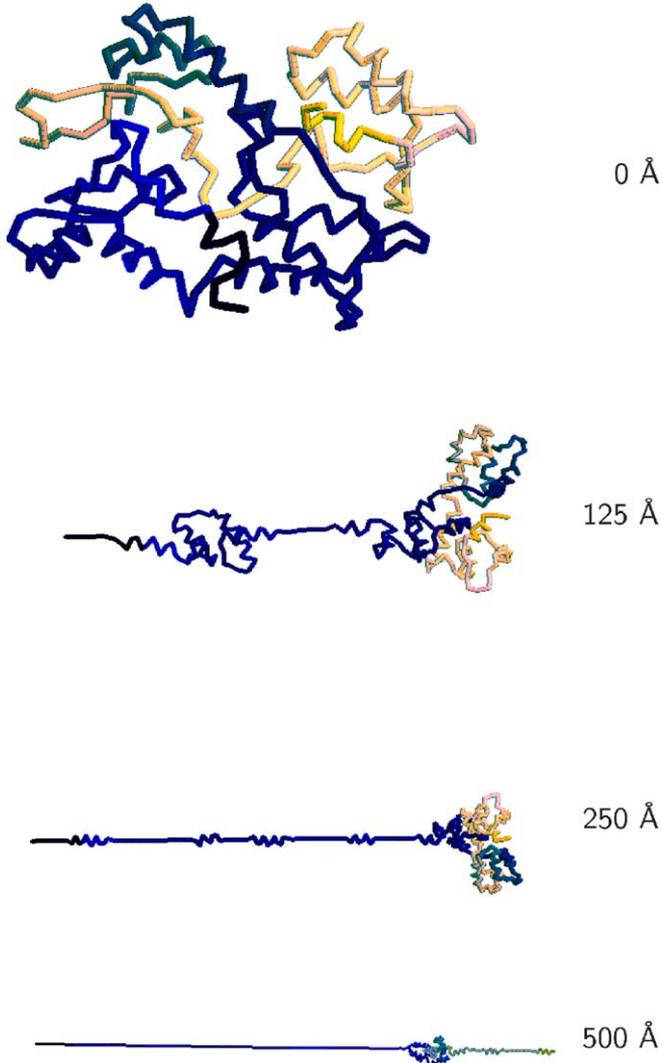
Fig. 10. Same as in Fig. 9 but for two 1cfc domains connected in tandem.

An increase in the temperature causes effects which are consistent with the general scenario [16]—the peaks get lower and become less resolved. The interesting part is that, for calmodulin, the peaks at $\tilde{T} = 0.3$ are so inconspicuous that the curves look as though the system was almost in the entropic/worm-like-chain limit. This limit is fully achieved at $\tilde{T} = 0.6$, as demonstrated in Fig. 8. The nearly featureless nature of the experimental F–d curves taken at room temperature [13] is consistent with the $\tilde{T} = 0.3$ finding based on the Go model.

4. Scenarios of unfolding for calmodulin

Insights into unfolding can be obtained by looking at the snapshots of the process. Fig. 9 shows the snapshots for 1cfc at $\tilde{T} = 0$. Similar to titin, the system starts unwinding by breaking bonds that connect the terminal segments. However, the breakage involves a much smaller force. The next stage involves separation of helices which are parallel to each other, and finally unwinding of the helices themselves. Fig. 10 shows that the stretching of two domains engages both of them from the early stages.

A convenient way to describe the unfolding process is by providing the 'scenario diagrams' in which distances, d_u at which specific contacts break are plotted against
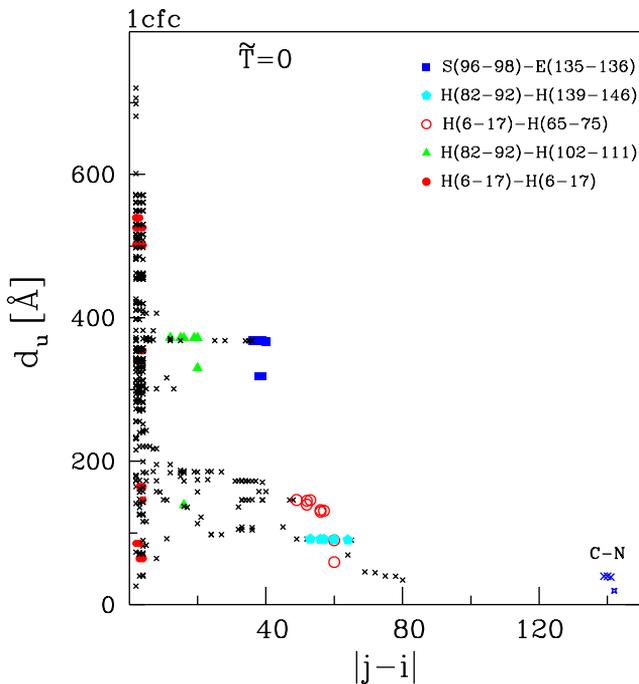
Fig. 11. The scenario diagram for stretching of 1cfc at $\tilde{T} = 0$. The symbols are explained in the main text.

the contact order, i.e., against the sequential distance $s = |j - i|$ between the contact making amino acids i and j. At $\tilde{T} = 0$, there is a unique distance at which a contact breaks. At finite temperatures, contacts may reform a meaningful definition of d_u through a distance at which the contact exists for the last time. The technical criterion for the existence of a contact is that the distance between the amino acids involved is less than 1.5 σ_{ij} [31,32,18].

Figs. 11 and 12 show the stretching scenarios for 1cfc at $\tilde{T} = 0$ and 0.3, respectively. Six sets of interactions are represented by polygonal or circular symbols as indicated in the figures whereas the remaining interactions are marked by the crosses. For instance, the interactions between the 6–17 and 65–75 amino acids in two α-helices are shown as open circles and marked as H(6–17)–H(65–75). The other character symbols follows the conventions of the protein data bank: E is an extended β-strand and S is a bend. The two scenarios are rather similar to each other. For instance, in both cases stretching is initiated in the terminal region indicated by CN at the high end of the contact order. However, the ordering of certain events is switched in time. For instance, the rupturing of the H(82–92)–H(102–111) contacts (the triangles) at $\tilde{T} = 0$ takes place later than that of the H(6–17)–H(65–75) (the
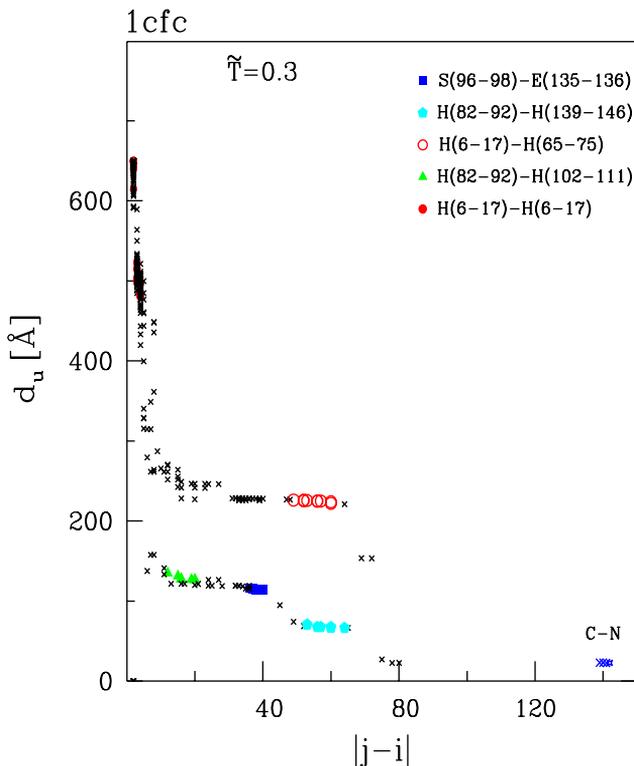
Fig. 12. Same as in Fig. 11 but for $\tilde{T} = 0.3$. The data points are for a single trajectory.

open circles), whereas at $\tilde{T} = 0.3$ the opposite ordering takes place. Furthermore, the $\tilde{T} = 0.3$ events are more noticeably accumulated into well-defined stripes compared to the bigger scatter and single-event resolution seen in the $\tilde{T} = 0$ plot.

We now focus on the room temperature-like case of $\tilde{T} = 0.3$ and consider the stretching scenarios for several domains of calmodulin. Fig. 13 refers to the case of two domains. The star symbols denote the contacts in the first domain, whereas the squares to the second domain. There is a clear intermixing of the symbols which indicates a non-serial character of unfolding. In particular, the contacts in the near-terminal (CN) regions unwind nearly simultaneously in both domains—in marked contrast to what happens in two titin domains [18]. An even heavier mixing takes place in the case of three domains, as shown in Fig. 14.

In summary, the Go-like models can capture the experimentally found difference in the elastic behavior between calmodulin and titin and can elucidate the microscopic picture of the events in stretching. The force–displacement patterns are sensitive to temperature and acquire the worm-like-chain behavior in the entropic limit. Unwinding of modular proteins does not need be serial in nature and calmodulin provides a clear example of such a non-seriality. Its source is mostly α character of the protein.
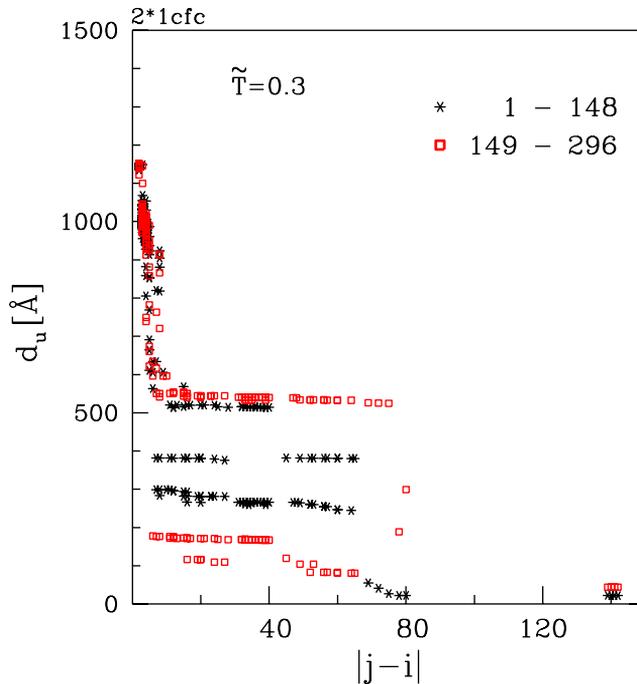
Fig. 13. The scenario diagram for two domains of 1cfc at $\tilde{T} = 0.3$.
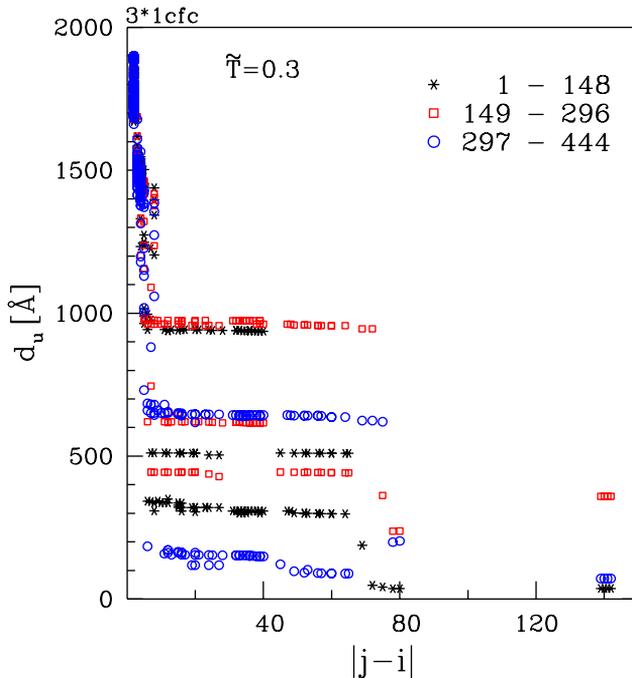
Fig. 14. The scenario diagram for three domains of 1cfc at $\tilde{T} = 0.3$.

Acknowledgements

References

[1] E.L. Florin, V.T. Moy, H.E. Gaub, Science 264 (1994) 415–417.
[2] H. Grubmuller, B. Heymann, P. Tavan, Science 271 (1996) 997–999.
[3] S.B. Smith, Y. Cui, C. Bustamante, Science 271 (1996) 795–799.
[4] A. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.L. Viovy, D. Chatenay, F. Caron, Science 271 (1996) 792–794.
[5] U. Bockelmann, B. Essevaz-Roulet, F. Heslot, Phys. Rev. Lett. 79 (1997) 4489–4492.
[6] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, H.E. Gaub, Science 276 (1997) 1109–1112.
[7] L. Tskhovrebova, K. Trinick, J.A. Sleep, M. Simmons, Nature 387 (1997) 308–312.
[8] M.S.Z. Kellermayer, S.B. Smith, H.L. Granzier, C. Bustamante, Science 276 (1997) 1112–1116.
[9] S.B. Fowler, R.B. Best, J.L. Toca Herrera, T.J. Rutherford, A. Steward, E. Paci, M. Karplus, J. Clarke, J. Mol. Biol. 322 (2002) 841–849.

[10] P.E. Marszalek, H. Lu, H.B. Li, M. Carrion-Vazquez, A.F. Oberhauser, K. Schulten, J.M. Fernandez, Nature 402 (1999) 100–103.

[11] 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 96 (1999) 3694–3699.

[12] Z.B. Guan, J.T. Roland, J.Z. Bai, S.X. Ma, T.M. McIntire, M. Nguyen, J. Am. Chem. Soc. 126 (2004) 2058–2065.

[13] M. Carrion-Vazquez, A.F. Oberhauser, T.E. Fisher, P.E. Marszalek, H. LI, J.M. Fernandez, Prog. Biophys. Mol. Biol. 74 (2000) 63–91.

[14] M. Carrion-Vazquez, H. Li, H. Lu, P.E. Marszalek, A.F. Oberhauser, J.M. Fernandez, Nat. Struct. Biol. 10 (2003) 738–743.

[15] G. Yang, C. Cecconi, W.A. Baase, I.R. Vetter, W.A. Breyer, J.A. Haack, B.W. Matthews, F.W. Dahlquist, C. Bustamante, Proc. Nat. Acad. Sci. 97 (2000) 139–144.

[16] M. Cieplak, T.X. Hoang, M.O. Robbins, Phys. Rev. E 69 (2004) 011912.

[17] H. Janovjak, M. Kessler, D. Oesterhelt, H. Gaub, D.J. Mueller, EMBO J. 22 (2003) 5220–5229.

[18] M. Cieplak, T.X. Hoang, M.O. Robbins, Thermal effects in stretching of Go-like models of titin and secondary structures, Proteins: Struct. Funct. Bio. 56 (2004) 285–297.

[19] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, J. Mol. Biol. 112 (1977) 535–542.

[20] S. Improta, A.S. Politou, A. Pastore, Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity, Structure 15 (1996) 323–327.

[21] H. Lu, K. Schulten, Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interprete atomic force microscopy observations, Chem. Phys. 247 (1999) 141–153.

[22] C. Micheletti, J.R. Banavar, A. Maritan, F. Seno, Phys. Rev. Lett. 82 (1999) 3372–3375.

[23] A. Maritan, C. Micheletti, A. Trovato, J.R. Banavar, Nature 406 (2000) 287.

[24] J.R. Banavar, A. Maritan, Rev. Mod. Phys. 75 (2003) 23–34.

[25] R. Du, V.S. Pande, A.Y. Grosberg, T. Tanaka, E.I. Shakhnovich, J. Chem. Phys. 111 (1999) 10375–10380.

[26] R. Unger, J. Moult, J. Mol. Biol. 259 (1996) 988–994.

[27] K.W. Plaxco, K.T. Simons, D. Baker, J. Mol. Biol. 277 (1998) 985–994.

[28] K.W. Plaxco, K.T. Simons, I. Ruczinski, D. Baker, Biochemistry 39 (2000) 11177–11183.

[29] H. Abe, N. Go, Biopolymers 20 (1981) 1013–1031.

[30] S. Takada, Proc. Natl. Acad. Sci. USA 96 (1999) 11698–11700.

[31] T.X. Hoang, M. Cieplak, J. Chem. Phys. 112 (2000) 6851–6862.

[32] T.X. Hoang, M. Cieplak, J. Chem. Phys. 113 (2001) 8319–8328.

[33] M. Cieplak, T.X. Hoang, Biophys. J. 84 (2003) 475–488.

[34] J. Tsai, R. Taylor, C. Chothia, M. Gerstein, J. Mol. Biol. 290 (1999) 253–266.

[35] G. Settanni, T.X. Hoang, C. Micheletti, A. Maritan, Biophys. J. 83 (2002) 3533–3541.

[36] T. Veitshans, D. Klimov, D. Thirumalai, Folding Des. 2 (1997) 1–22.

[37] M. Cieplak, T.X. Hoang, M.O. Robbins, Proteins: Funct. Struct. Genet. 49 (2002) 114–124.

[38] M. Cieplak, T.X. Hoang, M.O. Robbins, Phys. Rev. E 70 (2004) 011917.

[39] M. Cieplak, A. Pastore, T.X. Hoang, J. Chem. Phys., in press.

[40] M. Cieplak, T.X. Hoang, M.O. Robbins, Proteins: Funct. Struct. Genet. 49 (2002) 104–113.