∗ E-mail: mc@ifpan.edu.pl

Molecular jamming – the cystine slipknot mechanical clamp in all-atom simulations

Łukasz Pepłowski[1], Mateusz Sikora[2], Wiesław Nowak[1], and Marek Cieplak[2]

[1] *Institute of Physics,*

Nicolaus Copernicus University,

Torun, Poland

[2] *Institute of Physics,*

Polish Academy of Sciences,

Warsaw, Poland

Abstract

A recent survey of 17 134 proteins has identified a new class of proteins which are expected to yield stretching induced force-peaks in the range of 1 nN. Such high force peaks should be due to forcing of a slip-loop through a cystine ring, i.e. by generating a cystine slipknot. The survey has been performed in a simple coarse grained model. Here, we perform all-atom steered molecular dynamics simulations on 15 cystine knot proteins and determine their resistance to stretching. In agreement with previous studies within a coarse grained structure based model, the level of resistance is found to be substantially higher than in proteins in which the mechanical clamp operates through shear. The large stretching forces arise through formation of the cystine slipknot mechanical clamp and the resulting steric jamming. We elucidate the workings of such a clamp in an atomic detail. We also study the behavior of five top strength proteins with the shear-based mechanostability in which no jamming is involved. We show that in the atomic model, the jamming state is relieved by moving one amino acid at a time and there is a choice in the selection of the amino acid that advances the first. In contrast, the coarse grained model also allows for a simultaneous passage of two amino acids.

INTRODUCTION

Single-molecule manipulation [1–3] has opened new perspectives on understanding of the mechanical processes taking place in a biological cell and may offer insights into design of nanostructures and nanomachines. Examples of manipulation of a protein include stretching [4], mechanically controlled refolding [5], knot untying [6] and knot tightening [7]. The experimental studies pertained to only a handful of systems and yet demonstrated richness of possible mechanical behaviors. Experiments on stretching generate information on mechanostability. It can be captured by providing the value of F_{max} – the largest force that is needed to unravel the tertiary structure of a protein. F_{max} can be as large as 480 pN – the value measured for scaffoldin c7A [8]. The structural motif that is the core of resistance to stretching is known as the mechanical clamp. In most cases of large mechanostability, the mechanical clamp consists primarily of parallel [9, 10] (and sometimes antiparallel [11]) β-strands that get sheared on pulling at two points of attachment, which are often the terminal amino acids. This mechanism is operational, for instance, in scaffoldin [8] and titin [9, 10].

Recently, we have made stretching simulations for 17 134 proteins within a coarse grained structure-based model [12]. This survey pertained to proteins built of no more than 250 amino acids. Its results are deposited at the BSDB database (Biomolecule Stretching Database, www://info.ifpan.edu.pl/BSDB/) as described in ref. [13]. The survey not only identified about 35 proteins of larger mechanostability than the scaffoldin but it also discovered an entirely new type of the mechanical clamp: the cystine slipknot (CSK). In fact, the 13 top strongest model proteins are endowed with this mechanism. It involves forced dragging of a piece of the backbone, known as the slip-loop, through a cystine ring which is a special case of a knot-loop. Thus CSK does not exist in the native state but it develops through pulling. The cystine ring arises by connecting two backbone segments by two disulfide bonds between two pairs of cysteins. The ring is a part of what is known as the cystine knot (CK) motif [14–16] and has been first discovered in the superfamily of growth factors [17–19] and is responsible for the high mechanical

3

stability of collagen [20]. It typically contains between 8 and 15 residues [14] and it has emerged probably as a result of convergent evolution [21]. The pulling-induced dragging takes place because of the presence of a third disulfide bond that pierces the ring (Figure 1). This CSK mechanism has been predicted [12] to yield F_{max} exceeding even 1000 pN. Such large forces would arise because of jamming resulting from the slip-loop getting stuck in the cystine ring. An identifiable force peak maximum appears if the steric hindrances are overcome and the clinch is released. It should be noted that the acronym CSK has been initially meant to stand for "cysteine slipknot" [12] as the motif involves cysteins. However the term "cystine slipknot" used in this paper seems more apt as it brings the role of the disulfide bonds into fore.

At present, there are three well established families within the cystine knot superfamily: growth factor cystine knots (GFCK, e.g. vascular endothelial growth factor [22, 23]), inhibitor cystine knots (ICK, e.g. scorpion venom proteins [24]) and cyclic cystine knots (CCK) which have no terminal amino acids [25]. The cystine knot proteins studied in this work are members of the first family. The CK proteins are mostly extracellular factors and they are known for their remarkable stability against enzymatic cleavage [15, 16].

These 35 strongest proteins, with the CSK mechanical clamp and without, have not yet been studied in single molecule stretching experiments. Furthermore, our predictions are based on a simple model that accounts for the geometry and size of the side groups in a very approximate manner (primarily through the length parameters in the contact potentials) whereas the success in dragging through the cystine ring crucially depends on the actual size and structure of the moving segments involved. In this paper, we perform all-atom steered molecular dynamics (SMD) simulations and confirm the existence of the CSK mechanical clamps with the values of F_{max} being substantially larger than those associated with the shear-based mechanical clamps. We focus on the top 20 proteins identified in the survey as listed in Table 1. We also consider titin which serves as a benchmark [26–34].

We find that even though the coarse grained model captures the essence of the CSK

clamp and, in particular, that it involves jamming, it misses several interesting atomic-level components. The first component is the different origin of two possible stretching pathways: in the all-atom model what matters is the direction of approach of the pulled backbone to the cystine ring whereas it is the number of amino acids (one or two) that cross the ring simultaneously in the coarse grained model. Another component is that the tendency of developing multiple force peaks on a given pathway is stronger in the all-atom model because size modulations in the side groups may split the force peaks. These modulation are also relevant for thermal stability of the CSK-containing proteins (CSK proteins, for short).

The SMD simulations need to fit in the time scale of several tens of ns and thus have to be performed at pulling speeds which are orders of magnitude greater than what can be achieved in experiments or in coarse grained simulations. The basic pulling velocity used here is 0.04 Å/ps. This necessarily leads to the values of F_{max} which are much larger than experimental, typically an order of magnitude larger [9]. Thus our results on F_{max} are also overblown for this reason and have to be scaled down. We show that the functional form of this scaling down is complicated as it involves a crossover between a logarithmic and linear speed dependence. The latter is expected at high speeds when the drag force dominates. Nevertheless, the values of F_{max} obtained suggest that CSK comes with a much larger mechanostability than the shear-based clamp also at the experimental speeds.

For many CSK proteins there is a considerable variation between trajectories in the values of F_{max}. Pinpointing a specific force-based ranking of the proteins is difficult without generating significant statistics of the trajectories. It need not coincide with the one obtained by using the coarse grained model (and based on the dominant pathways). On the other hand, the ranking of the proteins with the shearing mechanical clamp, and listed in Table 1, is the same. In addition to providing the force-displacement $(F - d)$ trajectories for typically two trajectories for each protein studied, we also characterize the geometry and sequential make-up of the resulting mechanical clamps and study rupturing

5

of the hydrogen bonds.

RESULTS AND DISCUSSION

The force-displacement curves

Table 1 lists 21 proteins that are studied in this paper. The proteins there are arranged according to the value of F_{max} as obtained in the coarse grained model [12]. The energies and forces of that model are determined in terms of an energy parameter ϵ which measures the depth of the potential well in the Lennard-Jones-like native contacts. The forces written in Table 1 are converted to the experimental units using $\epsilon/\text{Å}=110\pm30$ pN – the result obtained by comparing the theoretically determined values of F_{max} to the established experimental data [12]. Notice a considerable error bar in the conversion factor. Another reason for not treating the protein ranking in Table 1 in a verbatim way is that for the CSK proteins at the high end of the force, there is a considerable trajectory dependence in the values of F_{max} both in the coarse grained and all-atom models. Even though the coarse grained values in Table 1 are obtained through averaging over 10 trajectories, a substantial dispersion in the CSK cases remains. The important point is that among the 21 proteins listed in Table 1, 15 have the CSK mechanical clamp as determined in the coarse grained model. They are ranked 1-13 and then 16 and 18. There are more than 100 other CSK proteins in the full set of 17 134 considered in ref. [12] but they are expected to correspond to lower forces, at least within the coarse grained model.

The protocol of our all atom simulations is explained in the Methods section. The resulting $F - d$ plots are displayed in Figs.2 through 4. Full extension is never reached as this would require breaking disulfide bonds. Figures 2 and 3 refer to the CSK proteins whereas Fig.4 to the proteins with the shear mechanical clamp. In the latter case, the values of F_{max} are an order of magnitude smaller and the trajectory dependence is substantially weaker. Previous all-atom studies [9] for titin indicate that F_{max} obtained

6

at such speeds is an order of magnitude higher than the experimental value of 204 pN [30, 31].

Among the 15 CSK proteins, five (1cz8, 2gh0, 1rew, 1m4u, and 3bmp) have both trajectories corresponding to F_{max} of order 15 nN whereas the remaining ten have at least one in the range of 30 nN as in the case of 1bmp (bone morphogenic protein) which is the mechanostability leader in the coarse grained evaluation. Larger statistics of ten trajectories have been obtained for two proteins, 1fzv and 2gyz, as shown in panels (g) and (h) of Fig.3. They indicate that F_{max} takes values in the whole range in which the upper reach is about twice as large as the lower reach. 2gyz is seen to have more trajectories with forces lower than 1fzv whereas in the coarse grained model the two systems appeared to behave similarly.

All of these force values are significantly larger than those associated with the shearing clamp shown in Fig.4. The largest value of F_{max} in this group is ~ 3 nN as observed for the homologous pair 1c4p and 1qqr (panels (a) and (b) in Fig.4). Despite the spread in F_{max} between the trajectories for the CSK proteins, many $F - d$ plots look essentially similar indicating existence of a single pathway. However, some proteins, like 2gyz (panels (a) and (h) of Fig.3) come with two distinctive patterns and thus two pathways: one with two force peaks and F_{max} exceeding 30 nN and another with four peaks and F_{max} of order 15 nN. We shall discuss why these differing patterns arise later on.

Fig.5 addresses the issue of the pulling rate dependence for titin (1tit) and human placenta growth factor (1fzv). The bottom panel shows the $F - d$ curves obtained at several speeds: the bigger the speed the higher the curve and the higher the second force maximum. We focus on the first maximum as its physics is governed by the well studied shear mechanical clamp. The corresponding values of F_{max} are displayed in the top panel. They suggest existence of a gradual crossover from the logarithmic to linear dependence on the speed in the regime of the pulling rates studied. At the slowest speeds, the data points appear to be consistent with a proper logarithmic extrapolation to forces obtained experimentally. Nothing is known about the speed dependence of F_{max} for the

7

CSK proteins. The top panel of Fig.5 displays results for 1fzv based on one (for smaller speeds) or two trajectories. The statistics involved are too poor to guess the functional dependence, however, the titin-like logarithmic dependence on going to the experimental speeds remains a possibility. Thus scaling down of our results for the CSK proteins using the titin-like logarithmic dependence seems to be a reasonable first guess. If so then F_{max} of 30 nN may probably correspond to the experimental 1500 pN. Such forces are comparable to those needed to rupture covalent N-C and C-C bonds: 1500 and 4500 pN respectively [35]. On the other hand, the coarse grained simulations suggested values which would not exceed 1100 pN. At this stage, the coarse grained model is probably more reliable in this respect.

It should be noted that the coarse grained simulations have been performed at speeds much closer to those expected at the experimental conditions. For the CSK proteins, the largest F_{max} obtained is about 5 times larger than the F_{max} for titin. At the speeds considered in the all-atom simulations, the derived forces are typically 10 - 20 times larger than for titin. The extra factor of 2 - 4 compared to the coarse grained model should be attributed to the different resolution of the two models and thus to the different ways in which the cystine ring is penetrated. As will be discussed later, the atomically represented ring distorts to an oblique form significantly and allows for penetration to occur only at an acute angle. This kind of penetration is sensitive to what atoms are facing the plane of the ring. In contrast, the cystine ring in the coarse grained model remains more circular and penetration is more vertical. We do not expect any other changes in the physics of dragging a segment of the backbone through the cystine ring to be relevant.

The mechanical clamp involving shearing of β-strands

In order to set the stage for the discussion of the working of the CSK mechanical clamp, we first discuss the shear clamp. The largest value of F_{max} within this mechanism is predicted to arise in the blood clotting streptokinase β-domain – 1c4p (we consider chain A) and its very close companion hydrolase activator streptokinase domain B – 1qqr. A

8

sequential alignment indicates that 1c4p has two more residues at the N terminus, 1qqr has three more at the C terminus and there are four sites at which the sequences differ. The high sequence identity results in a very close structural similarity.

The $F - d$ curves for the two proteins are shown in the first two panels in Fig.4. The values of F_{max} show only a small spread between two trajectories. They are 3171 and 3315 pN for 1c4p and 3048 and 3037 for 1qqr indicating a bit larger mechanostability of the former. In both cases, the mechanical clamp corresponds to the schematic representation as shown in the top panel of Fig.1. The way this mechanical clamp works is shown in more details in Fig.6 for 1c4p at four stages of unraveling. The major shearing action takes place between the parallel and long β-strands β_1 (11 amino acids) and β_4 (13 amino acids). There is extra shear between strands β_1 and β_2 as well as between β_3 and β_4.

The top left panel of Fig.7 and Fig.8 illustrate the nearly monotonic disappearance of the hydrogen bonds in 1c4p as a result of pulling. At F_{max}, there are no obvious indicators of rupture in the total number of the hydrogen bonds. However, such indicators arise when one looks into the number of the hydrogen bonds in specific pairs of the strands (Fig.7). It is seen that the pairs β_1-β_4 and β_1-β_2 lose their couplings at F_{max}, though some of them reappear temporarily later. On the other hand, the sheet β_3-β_4 loses fewer couplings at F_{max} suggesting their stabilizing role. They are still seen as partly operational in the lower panel of Fig.8 which shows the placement of the hydrogen bonds at d=80 Å. At this stage, the hydrogen bonds in the 15-residue long α-helix remain nearly intact.

The mechanical behavior of the remaining proteins shown in Fig.4 is similar to that found in 1c4p even though the corresponding structures are different and the values of F_{max} are smaller. Here, we discuss the diadenosine tetraphosphate hydrolase 1f3y for which the maximum force (of 2553 – 2624 pN) arises not at the first but at the third force peak. There are five helices in 1f3y (53-65, 85-95, 116-118, 139-145, and 148-164) and seven β-strands ($\beta 3$ is 36-41, $\beta 4$ is 44-46, and the remaining ones are listed in the caption to Fig.7. The two terminal ones unfold at the beginning and generate a small force peak at d around 50 Å. These events are followed by rupturing the sheet formed by

9

the antiparallel strands $\beta2$ and $\beta7$. The second small peak (around $d=160$ Å) comes from unfolding the bonds related to residues in the segment between 113 and 130 which involves loops and the third short helix. Panel (b) of Fig.7 demonstrates that the primary shear mechanical clamp is formed by two parallel strands $\beta1$ and $\beta6$ with some contribution from the antiparallel $\beta5$ and $\beta6$. The bonds between $\beta2$ and $\beta7$ are seen to disappear much earlier.

The cystine slipknot mechanical clamp

The $F-d$ patterns

We can distinguish two characteristic groups of the $F-d$ curves: with one force peak (1fzv, 1qty, 1cz8, 1wq9, 1flt, 1vpf, 1wq8) and with multiple peaks. This feature originates primarily from the size of the slipknot loop. In all of the single maxima cases, the proteins have a short and stiff slipknot loop comprising of order six amino acids. Larger slipknot loops (typically of order 30 amino acids) are more flexible and unfavorable steric clashes with the cytine ring get split into several events in various ways and are protracted in time as seen in Figs.2 and 3. Such long loops arise in 1bmp, 1lxi, 1m4u, 1rew, 2bhk, 2gh0, 2gyz, and 3bmp. However, multiple peaks may sometimes also arise with short loops, as discussed in the next section. This effect relates to sizes of the amino acids in the loop.

We first consider two examples of proteins for which we have observed only single force peaks: placenta growth factor 1fzv and vascular endothelial growth factor 1vpf (both human). In the case of 1fzv, the ring is formed from eight amino acids Cys66–Thr67–Gly68–Cys69–Cys70–Cys113–Glu112–Cys111 that make a ring because of the disulfide bonds Cys66-Cys111 and Cys70-Cys113. The slip-loop involves the segment Leu75–His76–Cys77–Val78–Pro79 (see also Table 2). Dragging is carried out by the disulfide bond Cys35-Cys77. In the case of 1vpf, the ring is created by the disulfide bonds Cys57-Cys102 and Cys61-Cys104 that link the backbone segments Gly58–Gly59–Cys61 and Glu103 into a tight ring. The slip-loop is formed by another segment of the backbone which consists

10

of amino acids Leu65–Glu67–Cys68–Val69–Pro70 (see also Table 2). Dragging is carried out by the Cys36-Cys68 disulfide bond.

The stages of unfolding of 1fzv are illustrated in Fig.9 for the trajectory with F_{max} of 30 390 pN (the other one corresponds to 26 980 pN). Dragging through the cystine ring is seen to take place abruptly – it is accomplished when d varies by less than about 3 Å. In the native state, the slip-loop is stabilized by 12 hydrogen bonds between two long parallel β-strands (amino acids 75–91 and 99–115). Almost all of them get ruptured on dragging the slip-loop through the ring as illustrated in Fig.10. We have demonstrated within the coarse grained model [12] that replacing all of the slip-knot related contacts (inside the slip-loop and between the slip-loop and the ring) by repulsive contacts does not affect F_{max} in a significant manner, indicating that the CSK mechanical clamp operates mostly through overcoming the sterical constraints. Panel (c) in Fig.7 shows that the total number of hydrogen bonds undergoes a sudden drop when the CSK clamp gets ruptured because of a major transformation of the whole structure. Panel (d) in this figure shows that the same observation also applies to the multiple-peak CSK proteins like 1bmp. This has not been the case in the shear mechanical clamps (panels (a) and (b) of Fig.7) because the unraveling at F_{max} is relatively more local and much more gentle.

We now consider the neurotrophic growth factor 2gyz. It is an example of a CSK protein which has (at least) two unfolding pathways, each with several force peaks (another such example is 1m4u). On one pathway F_{max} is nearly twice as large as on the other. The cystine ring consists of eight amino acids: Cys32–Ser33–Gly34–Ser35–Cys36–Cys100–Gly99–Cys98–Cys32. It is indicated by an ellipse in Figs.11 and 12. Fig.11 shows two possible ways of crossing the ring by the pulled cysteine whereas Fig.12, corresponding to the right-hand panels of Fig.11, defines the geometrical parameters of the CSK that will be discussed when describing temporal changes in the structure of the CSK mechanical clamp. The slip-loop is long and is pulled by Cys70 which forms a disulfide bond with Cys5 which is just near the N-terminus. One branch of the slip-loop is 27-residue long (from Arg37 to Cys70) and the other is 33-residue long (from Cys70 to Ala97). The

relevant segment of the slip-loop that has to squeeze through the ring is then Gln67–Pro68–Cys69–Cys70–Arg71–Pro72–Thr73. The steric hindrance arises when this segment of the slip-loop passes near the Cys32–Ser33–Gly34 fragment of the ring. Depending on the history of thermal fluctuations, there are only two ways in which the slip-loop is facing the ring: either the first neighbor Cys32 is closer to the ring or the other first neighbor Arg97 is the one which is closer. The two ways differ by a rotation of 180°. If the approach to the plane of the ring the were vertical, the two amino acids would enter the ring simultaneously. Instead, the symmetry is broken because the approach is at an acute angle since the pulling direction becomes increasingly parallel to the longer axis of the ellipse that represents the narrowing ring. The first way, shown in the left panel of Fig.11, yields the smaller force of 15 nN since cysteine is the smaller sized amino acid of the two. The second way, shown in the right panel, yields F_{max} of 30 nN. The small force case must necessarily come with the striated $F - d$ pattern since the bigger amino acid must follow immediately next which involves further forced expansion of the ring.

The separation into large force and small force pathways, independent of the number of the force peaks, should be observed for each of the CSK proteins studied here. However, for short slip-loops one way is expected to dominate, as determined by its sequential makeup, but both possibilities are seen in stretching of 1vpf and 1wq9. For long slip-loops, there is no particular preference for selection of the initial orientation on approaching the ring. It should be noted that the role of fluctuations depends on the speed of pulling. At small pulling speeds, conformations are expected to get equillibrated better which should reduce choices between various trajectories.

Residue-based description of jamming

We first consider proteins with short slipknot loops such as 1fzv and 1vpf. Figs.13 and 14 illustrate the jamming process through the evolution of the clinched fragment of 1fzv in the immediate vicinity of the force peak. The relaxed, native structure of the fragment is shown at the top for comparison. The fragment comprises the full cystine

ring (in green) and five amino acids of the slip-loop near its dragging head – Cys75 (in yellow). The three-residue long fragments of the slip-loop on the left and on the right of the pulling cysteine are listed in Table 2. The table makes use of the sequential alignments as in refs. [16, 18, 36]. Note that for each of the entries listed in the table, the second sequential neighbor of the pulling cysteine, when counting away from the N-terminus, is always a proline which initiates a β-sheet. In all but three cases, the proline is followed by threonine. Between the pulling cysteine and the proline, the residues can be either small or large.

The cysteine is a relatively small amino acid so it slides into the cystine ring fairly easily. This step, however, bends the slip-loop and raises tension. On further pulling, the neighboring residues (His76 and Val78 shown in brown and pink in the figures respectively) have to penetrate the ring either His76 first or Val78 first with the other of the two following immediately afterwards. The pathway with His75 entering first is dominant because the corresponding conformation is more native-like. It is shown in Fig.13 whereas the other pathway – in Fig.14. The penetration is resisted by the ring and the system is jammed. This stage determines the value of F_{max}. The dominant pathway yields a larger F_{max} since histidine is bigger in size than valine because of its aromatic ring. In this pathway, when His76 stays jammed, its neighbors Leu75 and Asn74 together with Val78 (which is flanking Cys77 from the other side) form a compact complex. This complex has an attached "hook", Pro79 (in blue), which is pivoting on the ring. Immediately after His76 passes through the cystine ring, this complex passes the ring as well. This event determines the largest force and is very quick. The whole process generates a single maximum. Once the bulky Leu75, Asn74 and Val78 complex crosses the loosened ring, the remaining slipknot residues slide through easily in a swinging motion. When the full length of the slip-loop is exhausted and stretching affects only covalent bonds from the protein backbone, the tension grows again. The pathway in which Val78 is the head amino acid which penetrates the cystine ring (Fig.14) is similar with the role of the neighboring amino acids interchanged. For instance, it is Leu75 which forms a hugging hook on the

ring now.

The single peak mechanism involving a passage of a bulky plug is also observed in 1qty, 1wq8, 1flt, 1fzv, 1vpf, and 1wq9. In each of these proteins, the other second neighbor of the pulling cysteine is bulkier than proline and, in the dominant pathway, it is the larger second residue which penetrates the first and paves the way for the smaller companion. It should be pointed out that despite the multitude of the possible values of F_{max}, as seen, e.g. in Fig.3g, there are only two pathways of the ring penetration. The differences in the values arise merely from minute differences in the angle of approach that get emphasized by the very magnitude of the stresses involved.

Multiple peaks arise when the slipknot loop is long and when the amino acids enter in order in which their geometrical sizes keep increasing because the subsequent residues keep forcing the cystine ring ever more open. The resulting sawtooth-like $F - d$ pattern is encountered, for instance, in 1bmp, 2gyz, 2bhk, 2gh0, 1rew, 3bmp. Again, there are basically only two pathways, albeit multipeaked. It is interesting to observe that all of these multiple-peak proteins contain another cysteine. It should be noted that this neighboring cysteine may be involved in forming a dimer state with a similar companion. In the dimeric state, the slip-loop motion through the ring would be prohibited.

We have found that the passage of the slipknot through the cystine ring is generally associated with a rapid rise in the bond and bond-angle contributions to the total energy. All remaining contributions to the total energy, such as electrostatic, van der Waals and those associated with the dihedral and improper angles remain fairly unaffected by crossing of the force peaks.

Curvatures in the backbone

Finally, we discuss the geometry of the CSK mechanical clamp in terms of geometry as characterized by the effective curvatures. Previously [12], we have proposed that the

14

slip-loop can be driven through the cystine ring provided

$$R_{cs} + t_s \;\; < \;\; R_{ck} - t_k \quad , \tag{1}$$

where R_{cs} and R_{ck} are the radii of curvature of the slip-loop and the ring respectively. We have estimated these radii to be of order 7 and 3 Å correspondingly. Symbol t_s denotes the effective thickness of the slip-loop and t_k that of the ring. Both have been assessed to be around 2.5 Å.

Here, we reexamine this condition by determining the local radii of curvature in a way used in the context of the tube picture of proteins [37, 38], i.e. by finding a circle which goes through the corners of a triangle set by three consecutive C$^\alpha$ atoms. Fig.15 shows the thus determined R_{cs} and R_{ck} for 1bmp and 1fzv, together with the semimajor axis, a, of the ellipse assigned to the cystine ring. In the native state, a and R_{ck} are both close to 5 Å for both proteins. R_{cs} is found to be about 3 Å. On pulling, all of these parameters evolve. Until reaching F_{max}, a generally grows whereas R_{ck} and R_{cs} go down. At F_{max}, the radii jump upward, indicating an escape from the constriction, and a jumps downward. At each stage, the condition given by eq. 1 is satisfied if the tube thickness is neglected.

In order to estimate the effect of the thickness in a description which does not invoke an explicit tube, we replace one of the corners of the triangle by another atom. For a new effective R_{cs} we take the backbone oxygen atom associated with the pulled amino acid (the line denoted by O). For a new effective R_{ck} we take the sulphur atom (the line denoted by S) residing on the next coming amino acid (see also Fig.12). These displaced-atom lines generally follow the behavior of the C$^\alpha$ based curvatures, at least until reaching F_{max}. However, they illustrate the magnitude of the effects associated with the tube thickness. The shifts are seen to be substantial and indicate that the condition to go through is borderline or even prohibitive if viewed from the tube-like perspective. We conclude that the atomic-level features and structure are important for the CSK mechanism to work when R_{cs} and R_{ck} are comparable in size.

Concluding remarks

Our all-atom simulations confirm existence of huge mechanostability of the CSK proteins that has been predicted by using the coarse grained model. It should be pointed out that there are differences between the way the slip-knot passes through the cystine ring as described by the two models. The all-atom ring is much tighter because of the bigger excluded volume generated by the atoms in the side groups. Thus the passage process involves significant distortion of the ring so that it becomes a narrow ellipse. As a result, only one amino acid can squeeze through at a time. In contrast, the C$^{\alpha}$-built ring in the coarse-grained model remains fairly circular allowing for the simultaneous passage of up to two amino acids that are the first neighbors of the pulling cysteine. The variations in the magnitude of F_{max} between the trajectories in the coarse grained model, also even up to a factor of 2, arise now not that much from the selection of the amino acid that goes through first but from the possibility of dragging of one or two neighboring amino acids initially.

Our all-atom simulations validate the results obtained through coarse grained modeling in a qualitative way as they support possible existence of the CSK mechanical clamp. The ultimate test, however, should come from experimental pulling studies and it is the experiments that should establish the true values of F_{max}. Our theoretical estimates of F_{max} are not sufficiently precise and should be considered just as indicative of large forces that are expected to be found. The difficulty that such experiments might encounter is the resolvability of the CSK-related force peaks from the background generated by stretching of the peptide bonds.

This stability of the CK proteins appears to be related to the high mechanical stability found in this paper. Specifically, we have shown that the buildup in the tension results from jamming. The sequential neighbors of the pulling cysteine are of a size that prevents them from penetrating the cystine ring without forcing. Thus we expect that thermal fluctuations would not be able to form a slipknot and change the conformation. Similar stability features should also be operational in the CCK and ICK proteins. Thus even

though mechanical stability, in general, need not be related to stability against thermal fluctuations (see, e.g. [39]) there is a relationship in the case of the CK proteins. In particular, it has been shown [40] that a removal of the disulfide bonds in the CK-containing VEGF protein through mutations lowers the melting temperature significantly.

In a recent survey of multidomain proteins [41] still new mechanical clamps have been identified by using the coarse grained model. Among them are two kinds of tensile clamps – one simple, involving contacts between two domains, and another in which a part of the clamp is immobilized by a non-cystine knot-loop. It would be interesting to investigate atomic aspects of the workings of such clamps by using all-atom simulations as done in this paper.

MATERIALS AND METHODS

All simulations have been performed using NAMD 2.6 code [42] with the all-atom CHARMM27 force field [43]. All of the simulations have been carried out using the same computational protocol. The initial structure is downloaded from the PDB database [44] and then placed within a box with rigid water molecules. The layer of the water molecules is at least 8 Å thick. Na^+ and Cl^- ions are introduced into the system at concentration of 0.5 mol/L (between 10 and 20 ions in the box). In order to neutralize the system, additional several ions of one sign are added using a script embedded in the VMD code [45]. The molecules of water are then moved for 100 steps per atom in order to minimize the energy and then equilibrated at the temperature of 300 K during 100 ps of the time evolution. After this stage, the protein atoms are allowed to move and the whole system undergoes 1000 steps of the energy minimization followed by heating from 0 K to 300 K during 50 ps in a stepwise fashion. The last step before stretching consists of 1 ns equilibration at 300 K by using the Langevin dynamics. The time step is set at 1 fs.

In analogy to the coarse-grained approach [46], the SMD simulations [47–51] are implemented by placing the N-terminal C^α atom at a fixed location and by attaching an elastic spring to the C^α atom at the C-terminal amino acid. The elastic constant is equal

17

to 4 kcal/mol/Å^2 (277.9 pN/Å). The other end of the spring is made to move with the pulling velocity v_p of 0.04 Å/ps (4 m/s). The pulling direction is set to go through the line connecting the first and last C^α atoms. In all cases, at least two 10 ns SMD simulations that evolve from the same starting configuration are carried out. This is the configuration obtained at the end of the initial Langevin dynamics. Various trajectories arise due to differences in the time dependence of the Langevin noise applied. Graphical analysis are performed with the VMD 1.8.7 software [45]. Forces and local geometries are analyzed by using homemade scripts.

In order to determine whether two heavy atoms are connected by a hydrogen bond we check whether the distance between them does not exceed 3.3 Å and the angle donor (like O) – H – acceptor (like N) does not exceed 25°. These bonds rupture often and then reform.

The forces and the numbers of the hydrogen bonds are time averaged over the pulling distances of 0.5 Å. In the plots, these data undergo further smoothing out.

ACKNOWLEDGMENTS

[1] A. Galera-Prat, A. Gomez-Sicilia, A. F. Oberhauser, M. Cieplak, and M. Carrion-Vazquez, Curr. Op. Struct. Biol. **20**, 63-69 (2010).

[2] M. Carrion-Vazquez, M. Cieplak, and A. F. Oberhauser, Single molecule biophysics and unfolding forces, in: Encyclopedia of Complexity and Systems Science, Editor-in-chief R. A. Meyers, Springer, New York (2009), ISBN:978-0-387-75888-6.

[3] N. Crampton and D. J. Brockwell, Curr. Opin. Struct. Biol. **20**, 508-517 (2010).

[4] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. B. Li et al., Prog. Biophys. Mol. Biol. **74**, 63-91 (2000).

[5] J. M. Fernandez and H. B. Li, Science **303**, 1674-1678 (2004).

[6] M. T. Alam, T. Yamada, U. Carlsson, and A. Ikai FEBS Lett. **519**, 35-40 (2002).

[7] T. Bornschloegl, D. M. Anstrom, E. Mey, J. Dzubiella, M. T. Rief, and K. Forest, Biophys. J. **96**, 1508-1514 (2009).

[8] A. Valbuena, J. Oroz, R. Hervas, A. M. Vera, D. Rodriguez, et al.. Proc. Natl. Acad. Sci. U S A. **106**, 13791-13796 (2009).

[9] H. Lu and K. Schulten, Chem. Phys. **247**, 141-153 (1999).

[10] D. J. Brockwell, E. Paci, R. C. Zinober, G. Beddard, P. D. Olmsted, D. A. Smith, R. N.Perham, and S. E. Radford, Nat. Struct. Biol. **10**, 731-737 (2003).

[11] J. I. Sułkowska and M. Cieplak, J. Phys. Cond. Mat. **19**, 283201 (2007).

[12] M. Sikora, J. I. Sułkowska, and M. Cieplak, PLoS Comp. Biol., **5**, e1000547 (2009).

[13] M. Sikora, J. I. Sułkowska, B. S. Witkowski, and M. Cieplak, Nucl. Acid. Res. (2010) (published on-line).

[14] D. J. Craik, N. L. Dally, and C. Waine, Toxicon **39**, 43-60 (2001).

[15] K. Berry, M. Thomas, B. Langley, M. Sharma, and R. Kambadur, Am. J. Physiol. Cell. Physiol. **283**, C135-C141 (2002).

[16] K. J. Rosengren, N. L. Daly, M. R. Plan, C. Waine, and D. J. Craik, J. Biol. Chem. **278**, 8606-8616 (2003).

[17] N. Q. McDonald and W. A. Hendrickson, Cell **73**, 421-424 (1993).

[18] P. K. Pallaghy, K. J. Nielsen, and D. J. Craik, Protein Sci, **3**, 1833-1839 (1994).

[19] D. J. Craik, N. L. Daly, T. J. Bond, C. Waine, J. Mol. Biol. **294**, 1327-1336 (1999).

[20] D. Barth, H. J. Musiol, M. Schuett, S. Fiori, A. G. Milbradt, C. Renner, and L. Moroder, Chemistry **9**, 3692-702 (2003).

[21] O. Carugo, S. Lu, J. Luo, X. Gu, S. Liang, S. Strobl, and S. Pongor, Prot. Eng. **14**, 639-646 (2001).

[22] S. Iyer, P. D. Scotney, A. D. Nash, and K. R. Acharya, J. Mol. Biol. **359**, 76-85 (2006).

[23] R. M. Stroud and J. A. Wells, Sci STKE 231: re7 (2004).

[24] C. Combelles, J. Gracy, A. Heitz, D. J. Craik, and L. Chiche, Proteins: Struct. Funct. Bio. **73**, 87-103 (2008).

[25] C. W. Gruber, M. Cemazar, M. A. Anderson, and D. J. Craik, Toxicon **49**, 561-575 (2007).

[26] H. P. Erickson, *Science* **276**, 1090-1092 (1997).

[27] W. A. Linke, M. Ivemeyer, N. Olivieri, B. Kolmerer, J. C. Ruegg, and S. Labeit, *J. Mol. Biol.* **261**, 62-71 (1996).

[28] L. Tskhovrebova, K. Trinick, J. A. Sleep, and M. Simmons, *Nature* **387**, 308-312 (1997).

[29] M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, *Science* **276**, 1112-1116 (1997).

[30] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Science* **276**, 1109-1112 (1997).

[31] M. Carrion-Vasquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez, *Proc. Natl. Acad. Sci. (USA)* **96**, 3694-3699 (1999).

[32] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez, *Proc. Natl. Acad. Sci. (USA)* **98**, 468-472 (2001).

[33] A. F. Oberhauser, P. E. Marszalek, M. Carrion-Vazquez, and J. M. Fernandez, *Nature Struct. Biol.* **6**, 1025-1028 (1999).

[34] S. B. Fowler, R. B. Best, J. L. Toca Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke, *J. Mol. Biol.* **322**, 841-849 (2002).

[35] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H. Gaub, Science **283**, 1727-1730 (1999).

[36] O. Avisian–Kretchmer and A. J. W. Hsueh, Mol. Endocrinology **18**, 1-12 (2004).

[37] J. R. Banavar, T. X. Hoang, A. Maritan, F. Seno, and A. Trovato,Phys. Rev. E **70**, 041905 (2004).

[38] T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. U S A **101**, 7960-7964 (2004).

[39] M. Cieplak and J. I. Sułkowska, J. Chem. Phys. **123**, 194908 (2005).

[40] Y. A. Muller, C. Heiring, R. Misselwitz, K. Welfle, and H. Welfle, J. Biol. Chem. **277**, 43410-43416 (2002).

[41] M. Sikora and M. Cieplak, Nucl. Acid. Res. 39, D443 (2011)

[42] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, C. Chipot, R. D. Skeel, and L. S. K. Kale,J. Comput. Chem. **26**, 1781-1802 (2005).

[43] A. D. MacKerell, D. Bashford, M. Bellot, et al., J. Phys. Chem. B, **102**, 3586-3616 (1998).

[44] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, et al.,Nucl. Acids Res. **28**, 235-242 (2000).

[45] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 27-8, 33-8 (1996).

[46] M. Cieplak, T. X. Hoang, and M. O. Robbins, Proteins: Struct. Funct. Bio. **56**, 285-297 (2004).

[47] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten, J. Mol. Graph. Model **19**, 13-25 (2001).

[48] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Biophys. J. **75**, 662-671 (1998).

[49] E. Paci and M. Karplus, J. Mol. Biol. **288**, 441-459 (1999).

[50] L. Peplowski, K. Kubiak, and W. Nowak, Chem. Phys. Lett. **467**, 144-149 (2008).

[51] W. Nowak and P. Marszalek, Molecular dynamics simulations of single molecule atomic force microscope experiments. in Current Trends in Computational Chemistry, ed. J. Leszczynski, 47-83 (2005), World Scientific, Singapore.
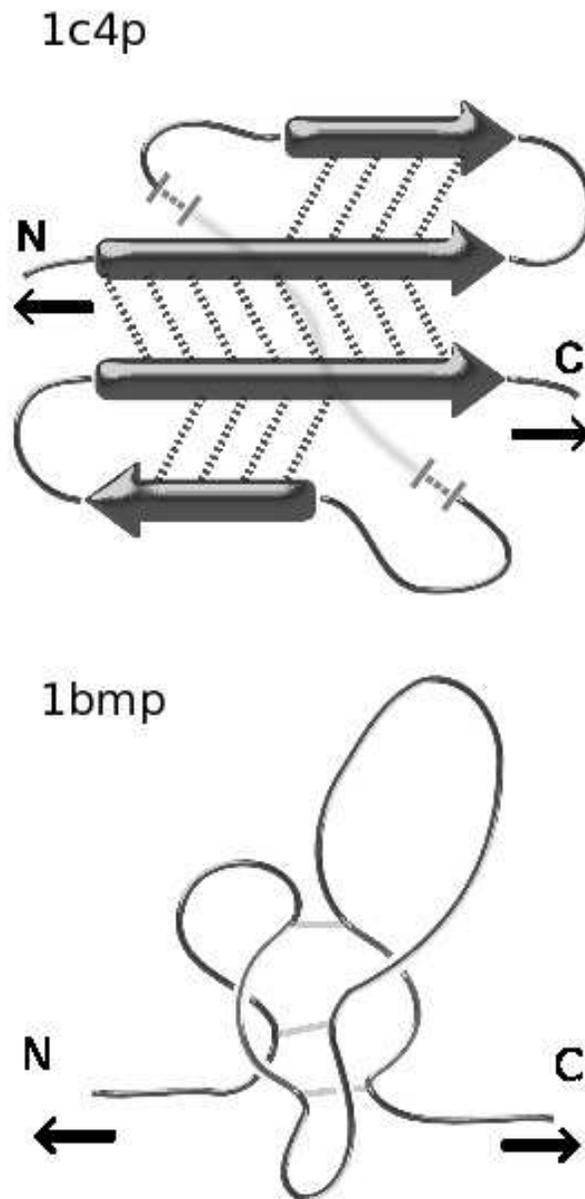
FIG. 1. Two basic classes of the mechanical clamps. The top panel shows the shearing mechanical clamp as illustrated for protein 1c4p. The bottom panel illustrates the cystine slipknot mechanical clamp in the case of protein 1bmp.
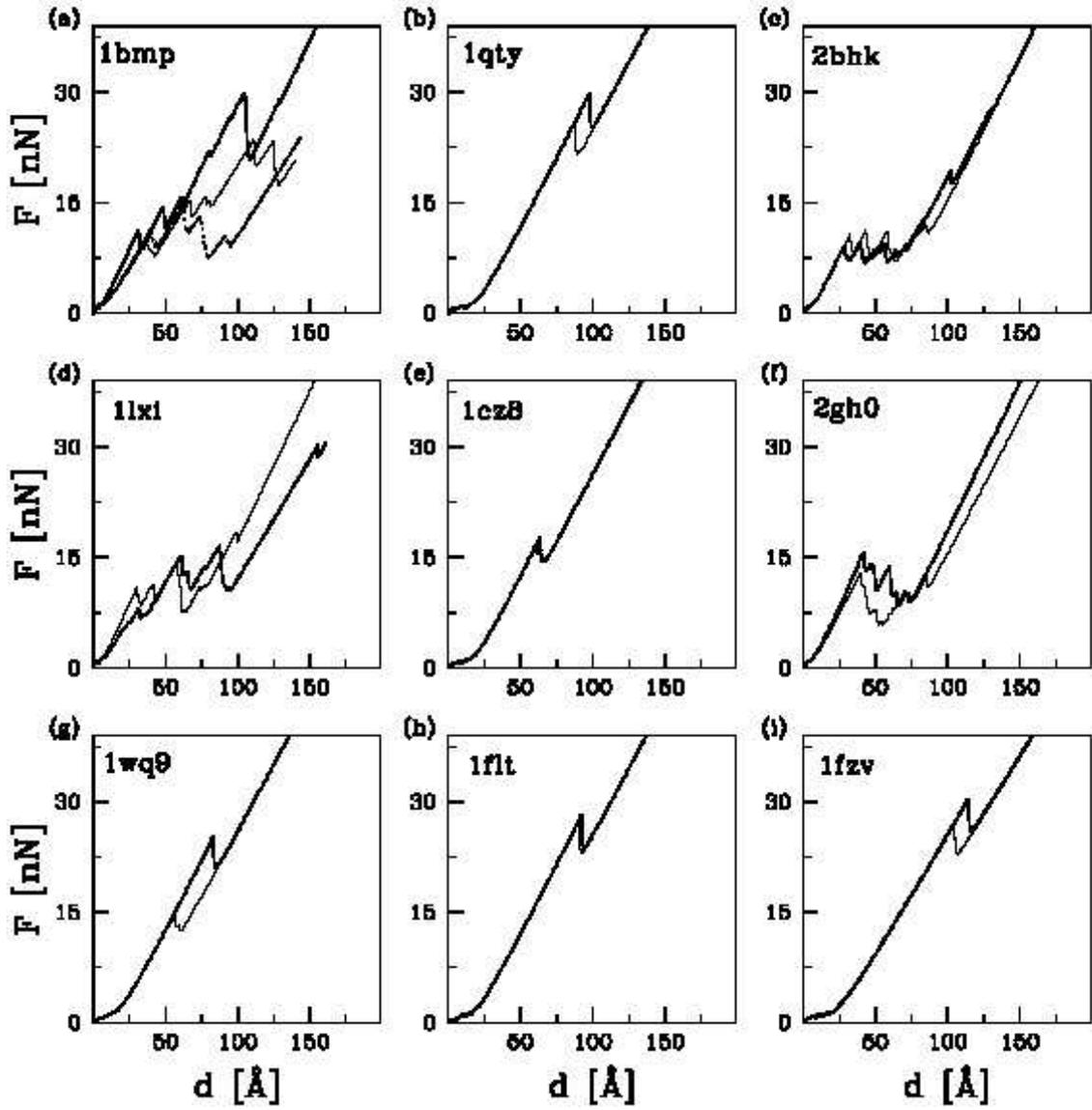
FIG. 2. Examples of the all-atom $F - d$ trajectories for the top 9 proteins listed in Table 1.
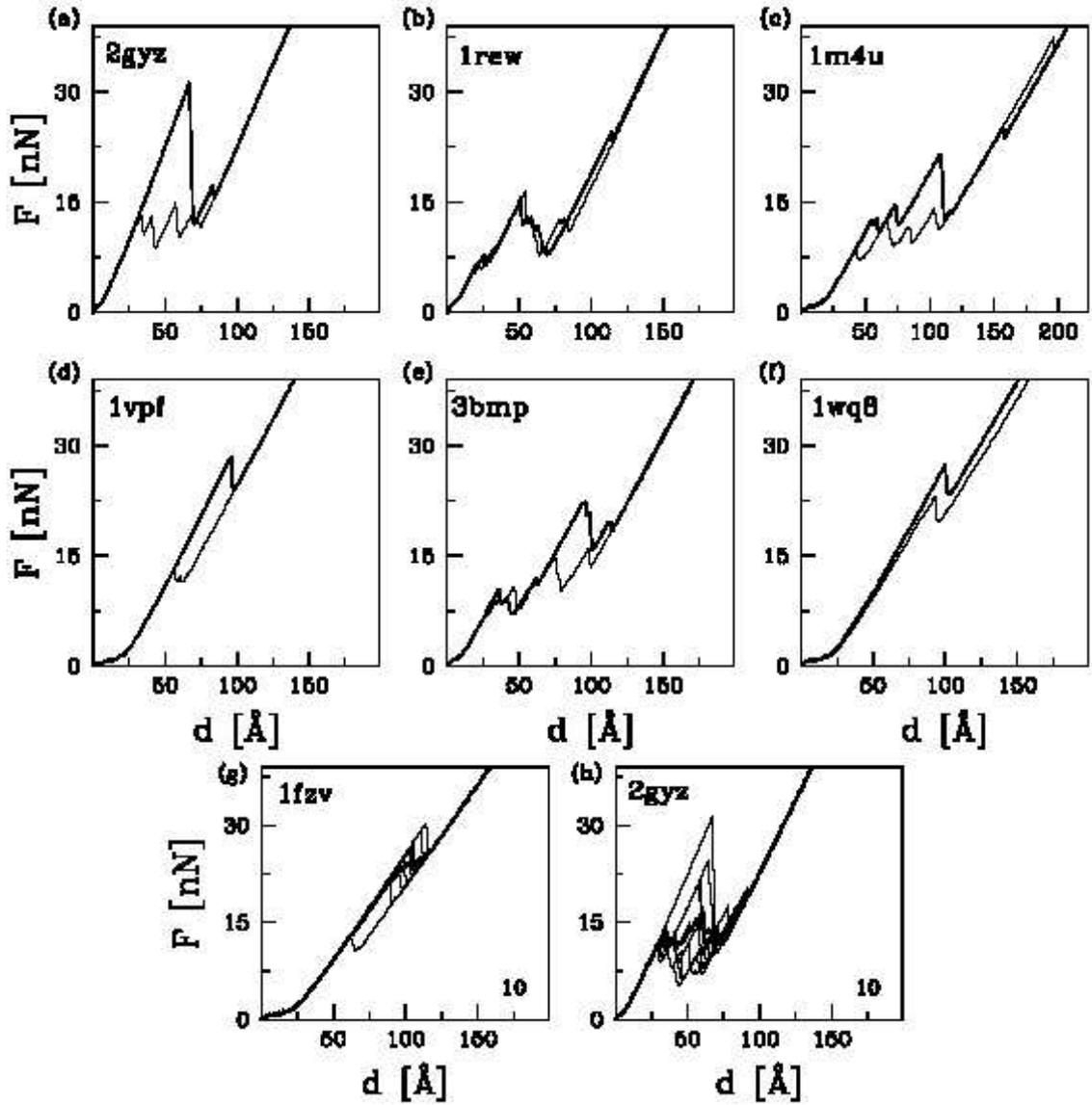
FIG. 3. Examples of the $F - d$ trajectories for the CSK proteins. The top 6 panels show examples of two all-atom $F - d$ trajectories for the proteins ranked as number 10, 11, 12, 13, 16, and 18 in Table 1. Panels (g) and (h) (with 10 written in the lower right corner) show ten $F - d$ trajectories for 1fzv and 2gyz as indicated. The thick solid line corresponds to the average over the ten trajectories.
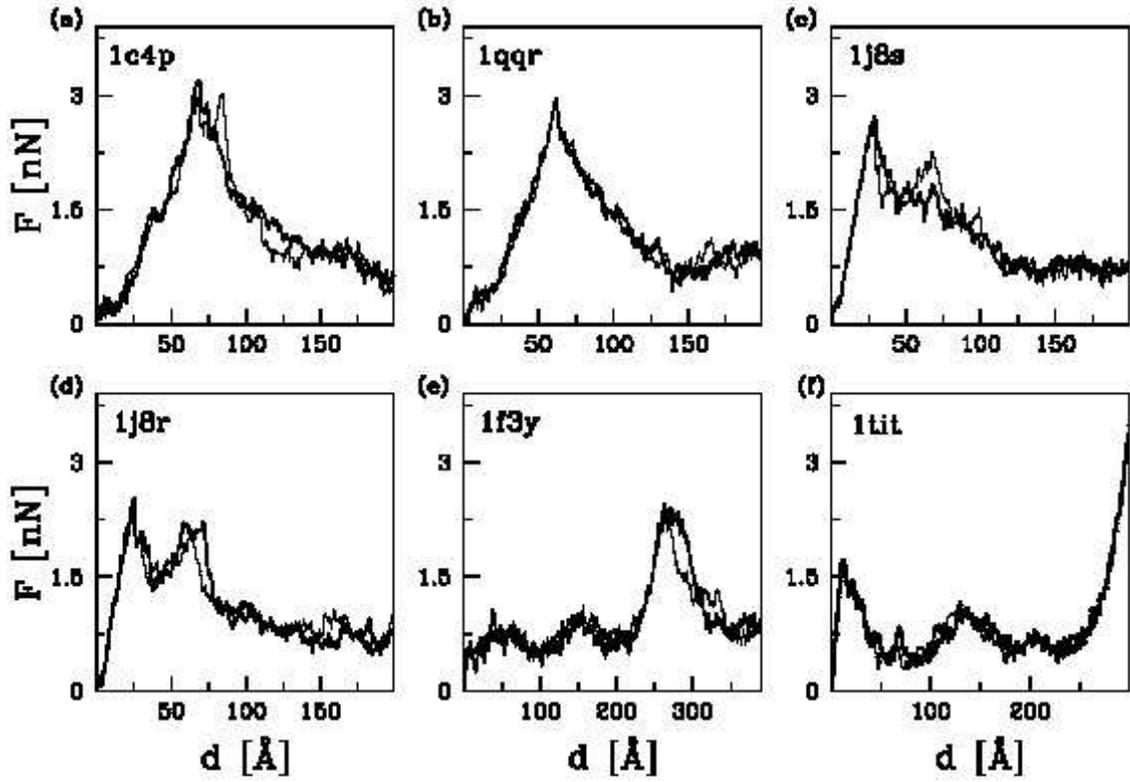
FIG. 4. **Examples of two** $F - d$ **trajectories for proteins are endowed with the shear mechanical clamps.** These proteins are the six remaining entries that are listed in Table 1.
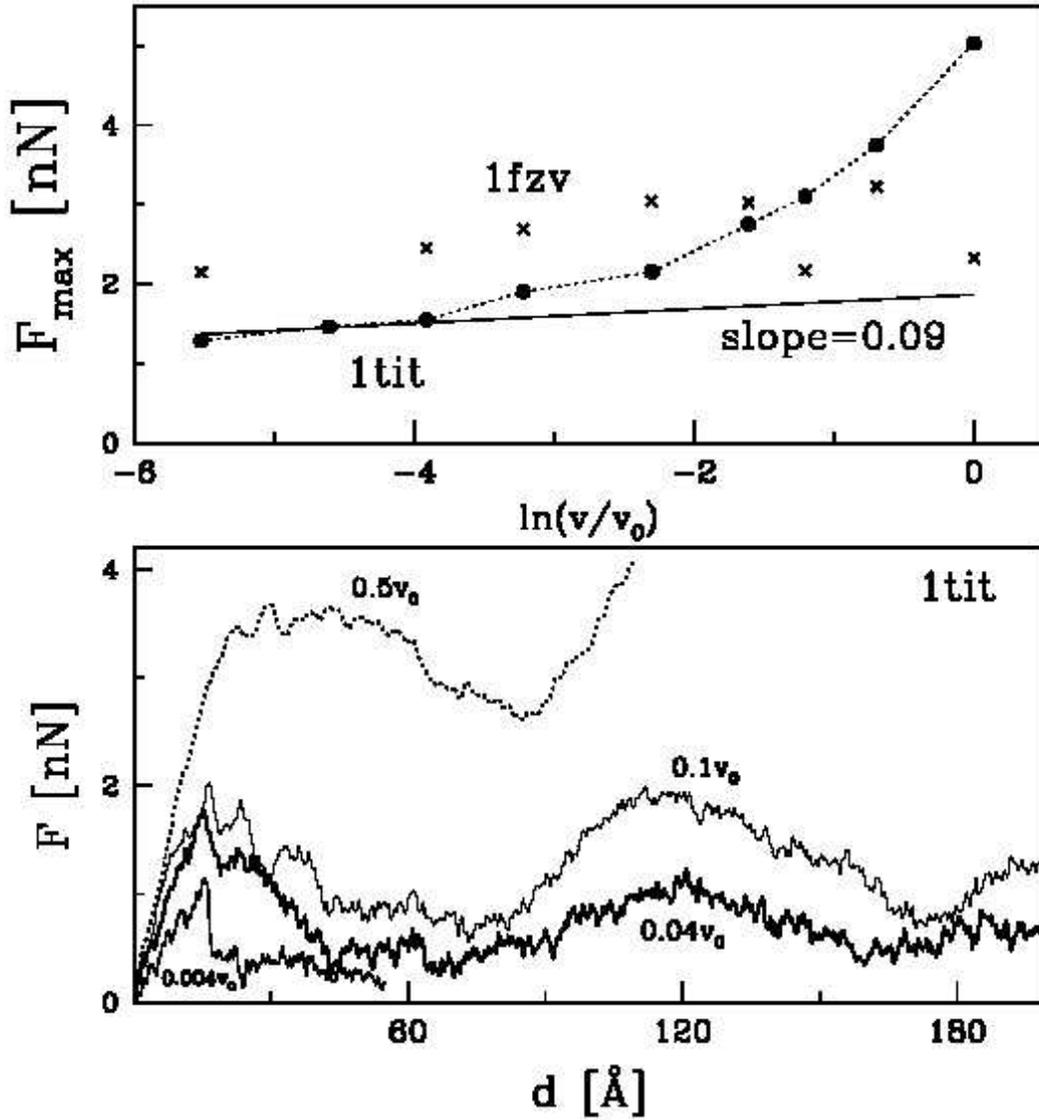
FIG. 5. **The velocity dependence of** F_{max}. The velocity v_0 corresponds to 1 Å/ps. The bottom panel shows the $F - d$ plots for titin at the indicated pulling speeds. The solid data points in the top panel correspond to F_{max} in the first force peak for titin as a function of $ln(v/v_0)$, where v denotes the pulling speed. The data points correspond to single trajectories. The straight line illustrates a perfect logarithmic dependence. The slope indicated would yield the experimental value of 204 pN at 600 nm/s. The crosses correspond to $0.1F_{max}$ in 1fzv. The three smallest speed data correspond to single trajectories and the remaining points are averaged over two trajectories.
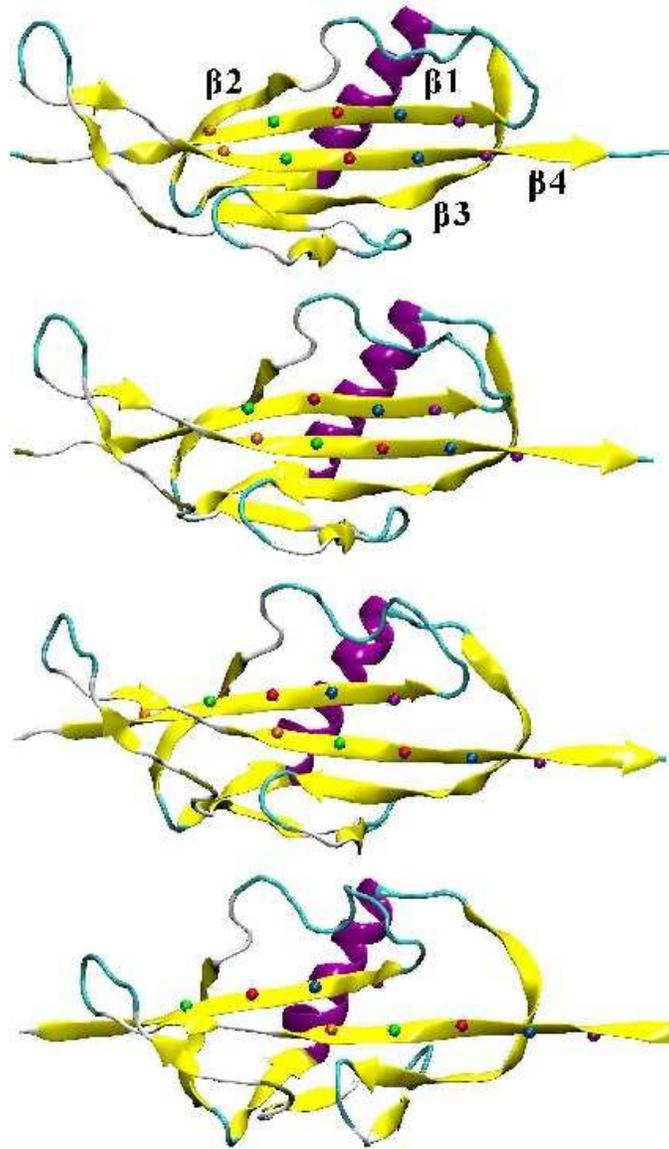
FIG. 6. Rupturing of the mechanical clamp in 1c4p through shear. The trajectory used is one indicated by the thicker line in Fig. 4. The time frames correspond to d of 67, 71, 75, and 83 Å top to bottom respectively. F_{max} arises at d=67 Å. The β-strands 1 through 4 correspond to the segments 158–168, 183-188, 214–226, and 266–278 respectively. The whole structure spans amino acids 149–285.
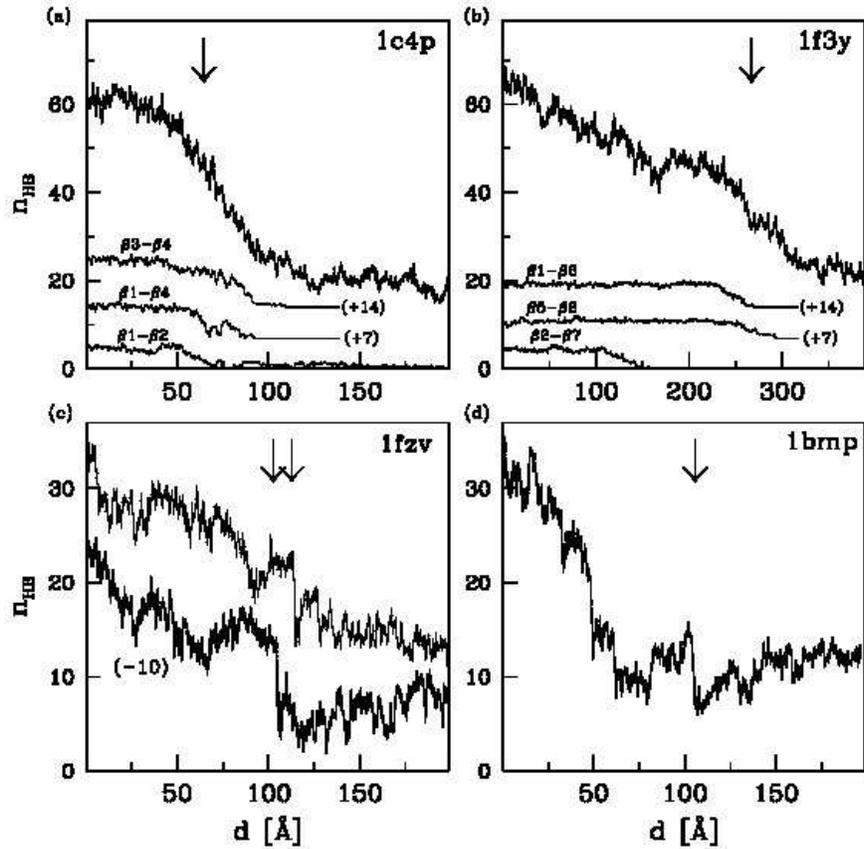
FIG. 7. The number of hydrogen bonds as a function of d in typical trajectories. The arrows indicate locations of the force peaks. The upper two panels correspond to proteins with the shearing mechanical clamps. The top lines show the total number of the hydrogen bonds. The lower lines represent hydrogen bonds between the indicated β-strands. For clarity, they are shifted by the numbers indicated in the brackets. The sequential segments corresponding to the β-strands in 1c4p are listed in the caption of Fig.6. For 1f3y, the definitions of the segments are as follows: $\beta1$ 16–22, $\beta2$ 27–33, $\beta5$ 70–75, $\beta6$ 105–112, $\beta7$ 131–137. The lower panels correspond to the proteins with the cystine slipknot mechanical clamps. The thicker lines show the total number of the hydrogen bonds corresponding to the trajectories indicated by the thicker lines in Fig.2. The thinner line for 1fzv corresponds to the second trajectory generated for this protein.

28
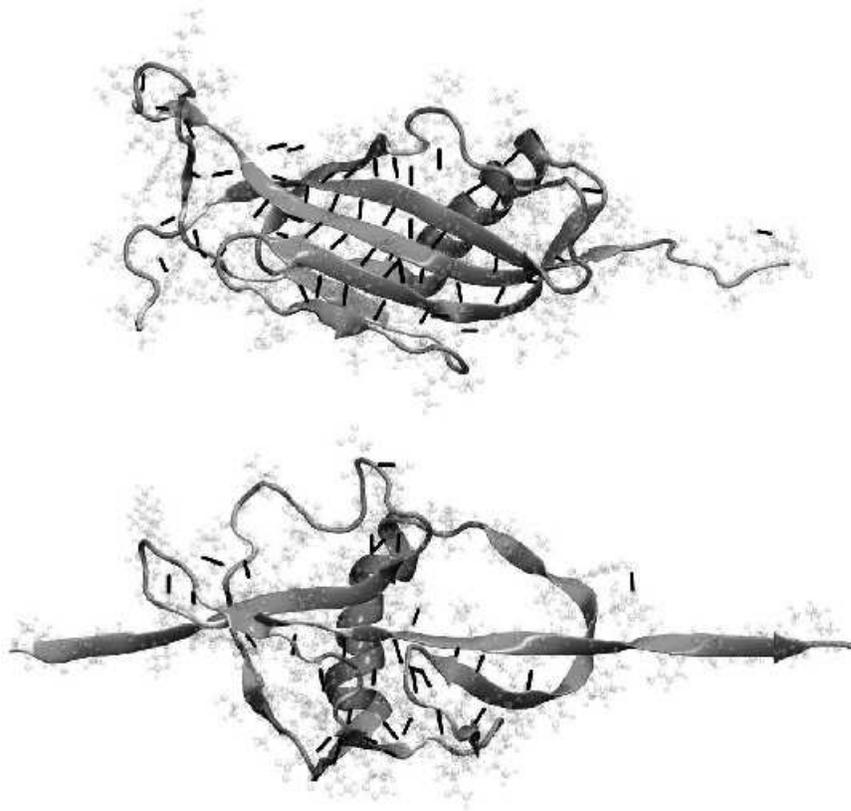
FIG. 8. Hydrogen bonds in 1c4p. They are indicated by solid bars. They may link the backbone or the side chains. The heavy atoms of the side chains are marked in the faint way. The top panel corresponds to d=20 Å and the bottom panel to 80 Å.
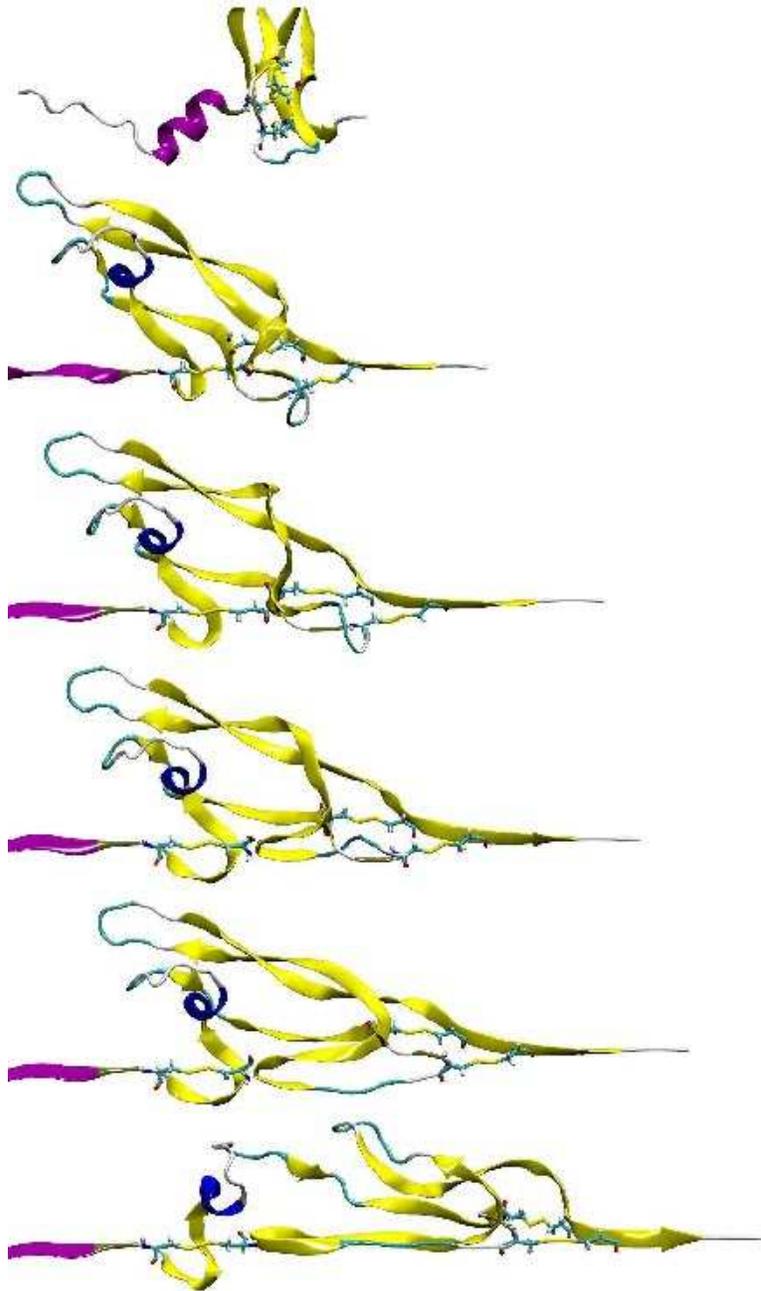
FIG. 9. Stages of rupturing of the cystine slipknot mechanical clamp in protein 1fzv. The time frames correspond to d of 0, 61, 113, 114, 114.5, and 116 Å top to bottom respectively. At the last stage shown here, the tension is decreasing even though the slip-loop is not yet fully pulled through. The biggest jamming is at d=114 Å.

FIG. 10. Hydrogen bonds in 1fzv for d=110 Å (the top panel) and 120 Å (the bottom panel). Like in Fig.8, they are indicated as solid bars. The force peak arises at 114 Å on this trajectory.
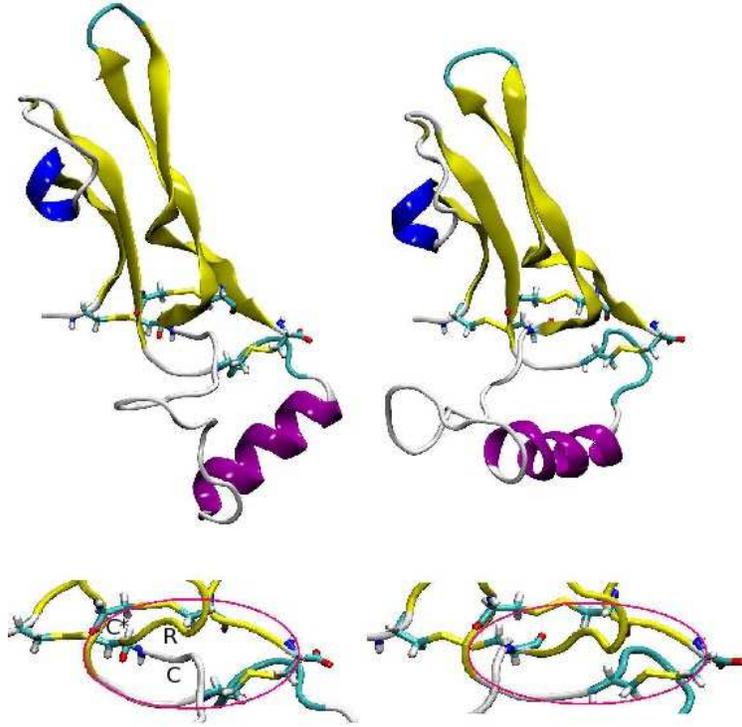
FIG. 11. The action of the cystine slipknot in 2gyz on two pathways. The left panels correspond to the pathway with the smaller value of F_{max}. The right panels – to the pathway with the larger F_{max}. All of the panels correspond to d=22 Å. In the lower panels, C^* indicates the cysteine which is pulled and C the neighboring cysteine. R denotes the neighboring arginine. In the left panels C, is closer to the ring than R. In the right panels, it is the other way around.
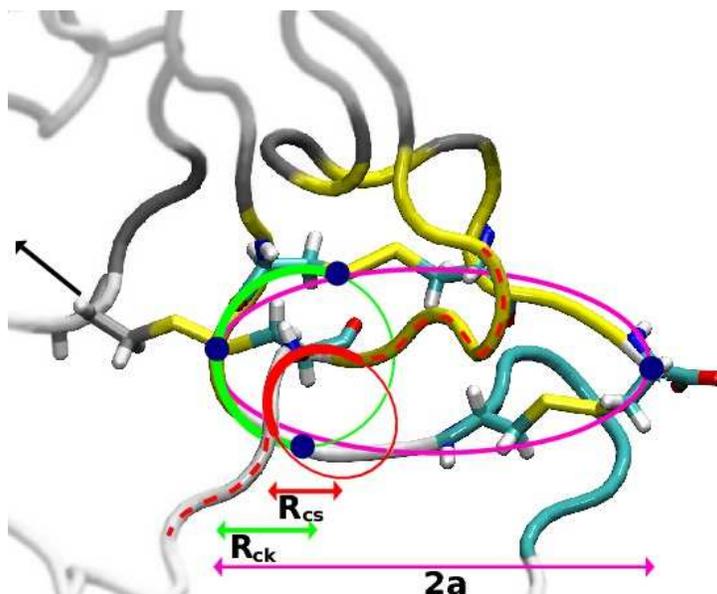
FIG. 12. Definition of the characteristic radii of curvature in the CSK problem as illustrated for 2gyz in its native state. The ellipse in pink encompasses the cystine ring. This loop is closed by two disulfide bonds Cys32-Cys98 and Cys36-Cys100. The major semi-axis of the ellipse is denoted by a and is estimated by taking half of the distance between the C^α atoms on the cystine ring which are furthest away from each other (Cys70 and Cys98). The segment near the "perihelion" is approximated by a circle in green. Its curvature can be determined from a triangle formed by three consecutive amino acids Ser33, Gly34 and Ser35. The corresponding radius of curvature is denoted by R_{ck}. Here, however, the circle shown is set on the triangle based on the two consecutive C^α atoms in Ser33 and Gly34 with the third corner of the triangle placed at the S atom belonging to the fourth amino acid Cys36. In the native state, there is very little difference in the radii generated from these two kinds of the triangles. The circle in red, of radius R_{cs}, corresponds to the curvature obtained by considering the three consecutive C^α: Cys69, Cys70 and Arg71. The middle C^α belongs to a cysteine. This red circle is above the ellipse. On pulling, it gets squeezed and dragged down by Cys70 through the ellipse.
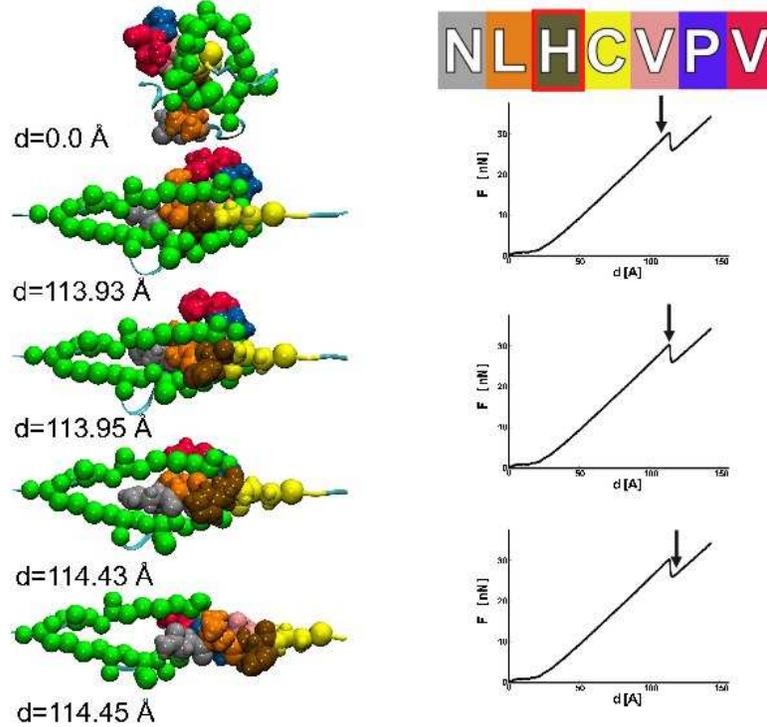
FIG. 13. The action of the cystine slipknot in 1fzv in the strong force pathway. The beads represent van der Waals spheres associated with the heavy amino acids. The different colors correspond to the amino acids indicated at the top. The amino acid that enters the cystine ring right after the cysteine is marked by the rectangular frame in the pictorial representation of the sequence. The snapshots correspond to the instances marked by the arrows on the $F - d$ curves on the right show situation just before sudden drop in tension.
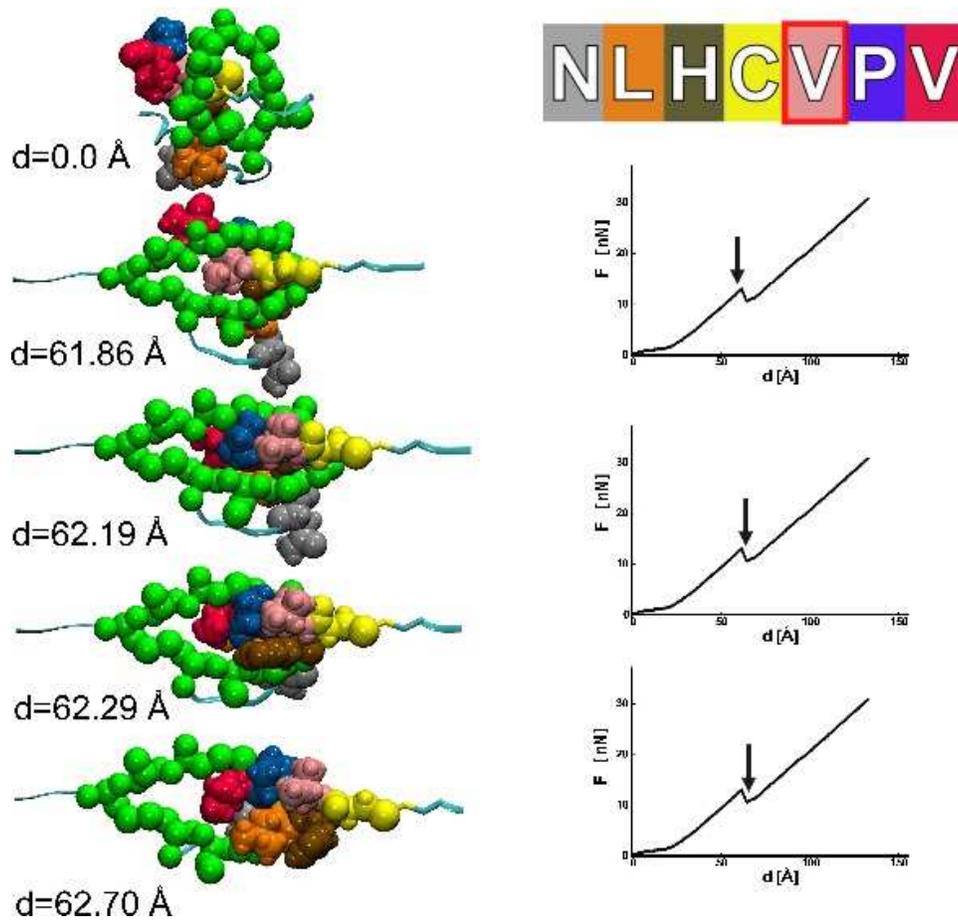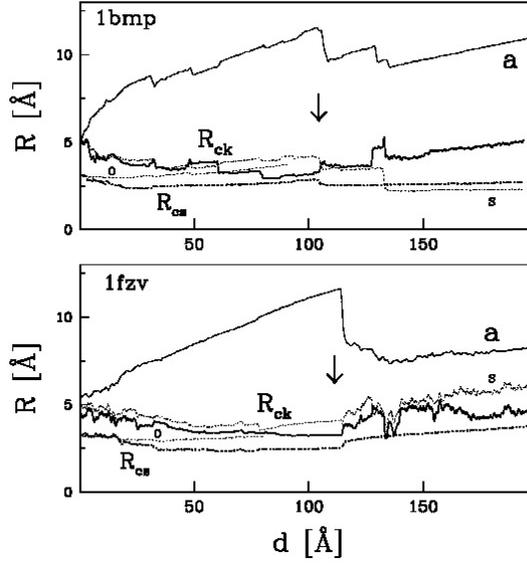
FIG. 14. The action of the cystine slipknot in 1fzv in the "weak" force pathway. The figure is similar to Fig.13 but it corresponds to Val78 entering the first. The trajectory chosen for illustration is the one with the smallest F_{max} in Fig.3g.

FIG. 15. The evolution of the characteristic radii of curvature and of the parameter a in 1bmp (top) and 1fzv (bottom). The arrows indicate the location of the highest force peak on the trajectory analyzed. The curvatures determined based on the three consecutive C^α atoms are marked by the thicker lines. The solid line is for the cystine ring and the dotted line – for the slip-loop. The thin line marked by the symbol S is obtained when the third C^α, belonging to the cysteine in the cystine ring, is replaced by the sulphur atom on the fourth amino acid. This atom forms one of the disulfide bonds defining the cystine ring. The thin line marked by the symbol O is obtained when the forward C^α in the slip-loop (i.e. in the amino acids that is first pulled through the ring) is replaced by the associated backbone oxygen atom.

Tables

TABLE 1: The predicted list of the strongest proteins.

n	PDBid	N	F$_{max}$ [pN]	L$_{max}$ [Å]	d$_{max}$ [Å]	CATH	SCOP
1	**1bmp**	104	**1120**	23.2	176.0	2.10.90.10	g.17.1.2
2	**1qty**	95	**980**	72.1	108.3	2.10.90.10	b.1.1.4
3	**2bhk**	119	**800**	26.5	129.3		
4	**1lxi**	104	**800**	22.5	126.4		g.17.1.2
5	**1cz8**	107	**700**	76.5	149.4	2.10.90.10	b.1.1.1
6	**2gh0**	219	**640**	25.9	104.6		
7	**1wq9**	100	**610**	72.0	131.8	2.10.90.10	g.17.1.1
8	**1flt**	107	**610**	75.6	128.6	2.10.90.10	b.1.1.4
9	**1fzv**	117	**590**	90.4	130.3	2.10.90.10	g.17.1.1
10	**2gyz**	100	**590**	14.4	93.3 &110		
11	**1rew**	103	**580**	21.7	92.7	2.10.90.10	g.7.1.3
12	**1m4u**	139	**580**	52.1	114.5	2.10.90.10	g.17.1.2
13	**1vpf**	94	**580**	68.1	121.6	2.10.90.10	g.17.1.1
14	**1c4p**	137	**560**	106.0	138.8	3.10.20.180	d.15.5.1
15	**1qqr**	138	**550**	110.3	134.8	3.10.20.180	d.15.5.1
16	**3bmp**	114	**550**	33.0	96.1	2.10.90.10	g.17.1.2
17	**1j8s**	193	**540**	77.9	99.3	2.60.40.1370	b.2.3.3
18	**1wq8**	96	**540**	82.6	109.8	2.10.90.10	g.17.1.1

19	**1j8r**	193	**530**	77.7	97.0	2.60.40.1370	b.2.3.3
20	**1f3y**	165	**530**	284.7	336.0	3.90.79.10	d.113.1.1
3580	**1tit**	89	**230**	55.3	43.3	2.60.40.10	b.1.1.4

Table 1. F_{max} is obtained within the structure-based coarse grained model in ref. [12]. The model is defined in terms of the energy parameter ϵ that determines the depth of the potential well in the native contacts. The conversion to pN is by taking the average relationship $\epsilon/\text{Å} \sim 110$ pN. The pulling velocity velocity is $\sim$0.005 Å/ns. The first column indicates the ranking of a model protein, the second – the PDB code, and the third – the number of the amino acids that are present in the structure used. L_{max} denotes the end-to-end distance at which the maximum force arises and d_{max} the corresponding tip displacement. The last two columns give the leading CATH and SCOP codes.

TABLE 2: Sequences of the pulled pieces of the slip-loops.

Protein	PDBid	Slip-loop sequence
BMP-7	1bmp	K P C C A P T
BMP-7	1m4u	K P C C A P T
BMP-7	1lxi	K P C C A P T
BMP-2	3bmp	K A C C V P T
BMP-2	1rew	K A C C V P T
VEGF	1qty	G L E C V P T
VEGF	1cz8	G L E C V P T
VEGF	1flt	G L E C V P T
VEGF	1vpf	G L E C V P T
VEGF	1wq9	S M K C T P V
VEGF toxin	1wq8	S L K C T P V
Human GDF	2bhk	P T C C V P T
GDNF family receptor α-3	2gh0	Q P C C R P T
PlGF	1fzv	N L H C V P V
ARTN isoform 3	2gyz	Q P C C R P T

Table 2. Abbreviations used in the table: BMP - Bone morphogenetic protein, VEGF - Vascular endothelial growth factor, GDF - growth and differentiation factor, GDNF - Glial cell line-derived neurotrophic factor, PlGF - Placenta growth factor, ARTN - Neurotrophic factor artemin.