Programming protein self assembly with coiled coils

Hendrik Dietz[1], Thomas Bornschlögl[1], Roland Heym, Frauke König and Matthias Rief

Physik Department E22, Technische Universität München,
James-Franck-Strasse 1, 85748 Garching bei München, Germany
E-mail: mrief@ph.tum.de

Abstract. The controlled assembly of protein domains into supramolecular structures will be an important prerequisite for the use of functional proteins in future nanotechnology applications. Coiled coils are multimerization motifs whose dimerization properties can be programmed by amino acid sequence. Here, we report programmed supramolecular self-assembly of protein molecules using coiled coils and directly demonstrate its potential on the single molecule level by AFM force spectroscopy. We flanked two different model proteins, Ig27 from human cardiac titin and green fluorescent protein (GFP), by coiled coil binding partners and studied the capability of these elementary building blocks to self-assemble into linear chains. Simple sterical constraints are shown to control the assembly process, providing evidence that many proteins can be assembled with this method. An application for this technique is the design of polyproteins for single molecule force spectroscopy with an integrated force-calibration standard.

Proteins provide tailored functionality at the nano-scale. Hence, proteins are likely to play significant roles in future nano-scale technologies. Assembly of nano-scaled protein domains into larger-scale functional assemblies is a difficult task. The formation of self assembled linear filaments in the cytoskeleton, like actin, myosin filaments or the giant muscle protein titin are examples of how nature solves this problem. Can we develop similar assembly strategies to control the assembly of protein subunits? Unlike with DNA, where controlled assembly can already be designed almost at will, folding and structure formation of proteins are yet far from being understood. A major challenge resides in identifying methods that allow for the

---

[1] Contributed equally to this work.

programmed self-assembly of supra-molecular protein complexes. Here, we report a method to assemble linear chains of proteins into supramolecular complexes based on coiled coil formation. A possible application of such protein chains is their mechanical and functional characterization in single molecule force spectroscopy.

Coiled coils are highly specific multimerization motifs frequently found in nature [1]–[4]. Coiled coils are formed from two or more amino acid sequences exhibiting a heptad signature. Upon multimerization, the binding partners assume helical structure and twist around each other to form a coiled coil. An extraordinary property of coiled coils is that the potential binding partners and their relative orientation along the helix axis can be programmed entirely by amino acid sequence [5, 6]. Coiled coils are thus excellent candidates to use for the programmed self-assembly of supramolecular protein complexes.

Figure 1 illustrates how coiled coils can be employed in order to control protein assembly. The multiple cloning site of a conventional DNA protein expression vector is modified such that it is flanked by the coding DNA sequences for two coiled coil binding partners (CCP). The coding DNA sequence of a target protein is then to be inserted between the CCP sequences. Expression of this fusion gene in bacteria leads to a monomeric fusion protein molecule, where the target protein is flanked by two CCP on both of its termini. In the simplest case, the two CCP sequences could be chosen equal and such that they form a homo-dimeric coiled coil. In this case, the fusion monomers will spontaneously assemble to form linear polyprotein chains. Integration of a cysteine residue at a proper site in the CCP sequences will provide a persistent covalent linkage of the coiled coils by disulphide bond formation. By design, it will be straightforward to integrate different CCP sequences that form dimeric, trimeric or higher-order homo- or hetero-coiled coils in order to create branching points for three-dimensional assemblies, inspired by DNA nanotechnology projects [7]–[10].

In order to test this assembly strategy, we first focus on equal CCP sequences that form homo-dimeric coiled coils and study their capability to assemble a single domain of a model protein, Ig27 from human cardiac titin, into linear polyprotein chains. As CCP, we chose a 35 amino acid sequence that is based on a motive known as GCN4-p1 leucine zipper [11] supplemented with a C-terminal cysteine residue. We term this sequence LZ10 [12]. LZ10 forms a parallel, homo-dimeric coiled coil.

We created the corresponding assembly vector by flanking the coding Ig27 DNA sequence by two LZ10 sequences in the multiple cloning site of the standard expression vector pRSET5d (Novagene). Expression in *E. coli* then yielded terminally linked fusion proteins, with one LZ10 sequence on each terminus of Ig27.

Such prepared protein units are able to multimerize in three different geometries that are illustrated in figure 2(a). Dimerization of the C-terminal LZ10 of one Ig27 unit with the N-terminal LZ10 of another unit leads to linkage geometry I. Geometry II arises when two C-terminal LZ10 sequences dimerize, while geometry III arises upon dimerization of two N-terminal LZ10 sequences. The three different modes of assembly are shown in a more schematical drawing in figure 2(d), inset. If such a polyprotein chain is stretched mechanically, geometry I will lead to a coiled coil aligned in parallel to the stretched chain, i.e. in an 'overstretching' geometry, while geometries II and III lead to an orthogonal configuration, i.e. an unzipping geometry. Geometry III differs from geometry II in that the terminal cysteine residue in LZ10 prevents the coiled coil from unzipping by a covalent disulphide bond located on the polyprotein axis. If the assembly of the polyproteins is independent of the target protein's properties and steered by coiled coil formation, the frequency of linkage geometry should be
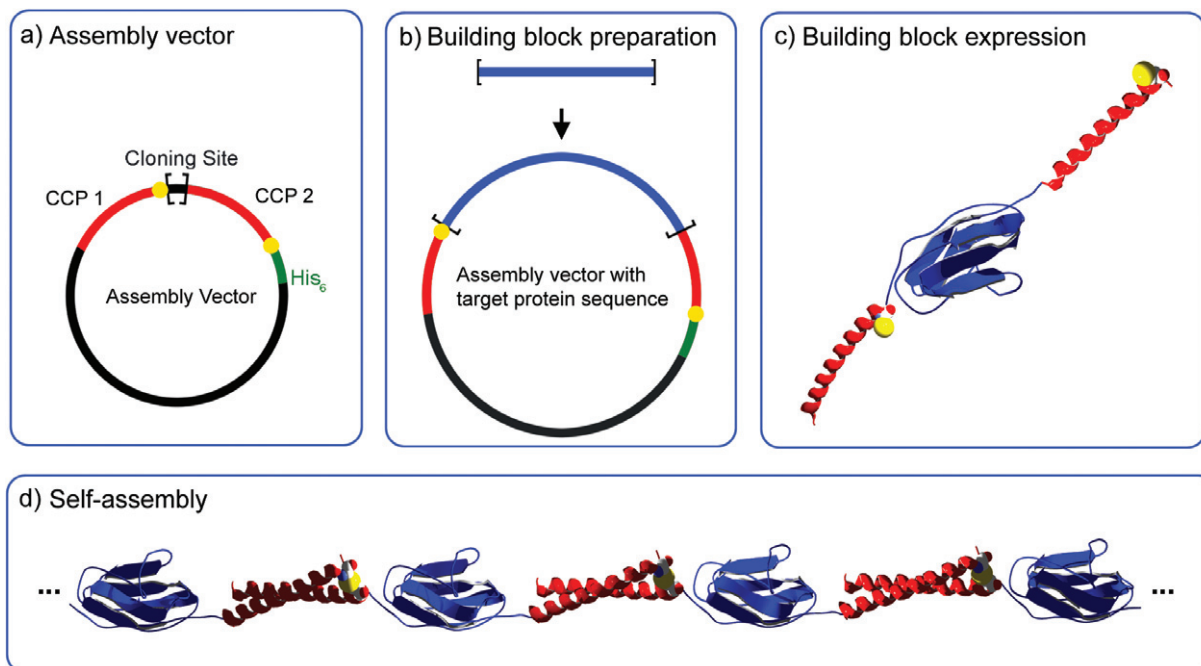
Figure 1. Scheme of coiled coil mediated protein self-assembly. (a) The multiple cloning site of a standard protein expression vector is flanked by the DNA sequences coding for two coiled coil binding partners (CCP1 and CCP2). (b) DNA sequence coding for the protein to be assembled is ligated between the CCP1 and CCP2 sequences. (c) Expression of the fusion gene in a suitable organism, followed by self-assembly via programmed coiled coil formation.

stochastic. Specifically, geometry I should occur in 50% of all cases, while geometries II and III should occur each in 25% of all cases.

Force microscopy using an atomic force microscope (AFM) allows for stretching single polyprotein molecules. The mechanical signature of forced Ig27 domain unfolding has been extensively studied [13]–[17]. The mechanical signature of coiled coil overstretching [18] has been studied as well as their behaviour in unzipping geometry [12]. AFM force spectroscopy allows thus to investigate the desired coiled coil mediated self-assembly of Ig27 molecules into linear polyprotein chains and to directly study the linkage geometries.

Figure 2(b) shows a typical force extension trace obtained with a coiled coil assembled Ig27 polyprotein as schematically shown in figure 2(a). The well known sawtooth-shaped pattern arising from the sequential force induced unfolding of single Ig27 domains [13] is visible at extensions above 280 nm (blue section). In the displayed case, a total of fifteen individual Ig27 domains have been unfolded. The mechanical signature of unzipping and overstretching coiled coils is expected to occur at forces lower than 30 pN. In order to increase resolution in the low force regime, the trace has been recorded by cycling extension and relaxation of the molecule multiple times at extensions between 50 and 180 nm, followed by averaging of the cycles. Figure 1(c) displays a zoom into the force extension response of the coiled coil assembled Ig27 polyprotein. A force plateau becomes clearly visible at forces around 10 pN, followed by a second plateau at 25 pN. The gain in length upon unzipping of a single coiled coil is much larger than upon overstretching a single coiled coil. Unzipping and overstretching of coiled coils
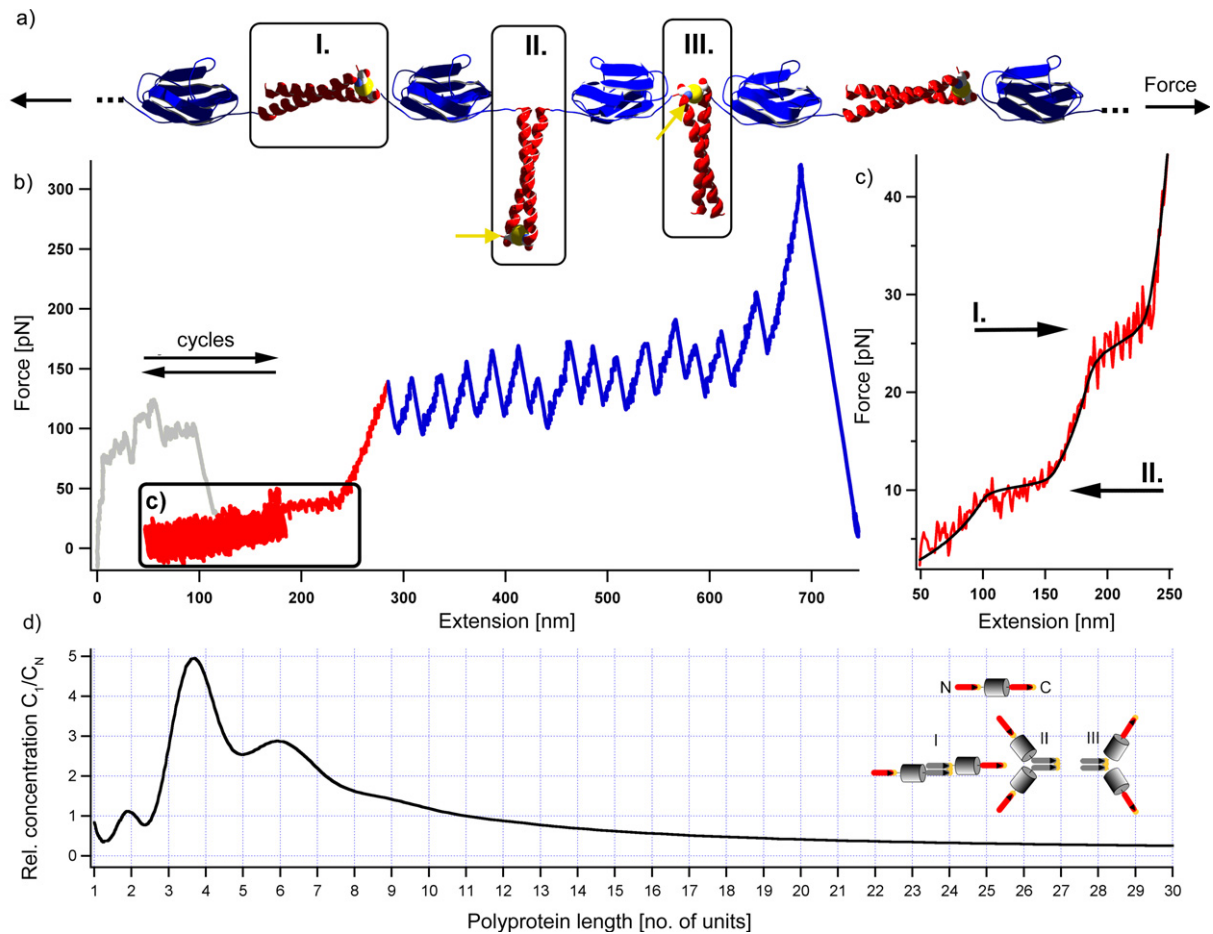
Figure 2. (a) Scheme of a coiled-coil assembled Ig27 polyprotein. Three linkage geometries are possible: (I) overstretching, (II) unzipping, and (III) blocked unzipping. Arrows indicate position of reinforcing disulphide bonds. (b) Force extension response of a coiled coil assembled Ig27 polyprotein, containing 15 subunits. Blue section: sequential unfolding of individual Ig27 domains, grey section: unspecific interaction at low distances to surface; red section: cycling phase. (c) High-resolution force extension response of the linking coiled coils, obtained by averaging the cycles in (b). Overstretching coiled coils results in a force plateau at 25 pN, while unzipping in a plateau at 10 pN. (d) Size-distribution in a coiled-coil assembled Ig27 polyprotein solution, normalized to monomer concentration, obtained by size exclusion chromatography.

occurs close to thermal equilibrium. The work needed to dissociate a single coiled coil can be estimated by the exerted force times the gain in length and equals its free folding energy. The free folding energy of a LZ10 coiled coil is $24\,k_B T$, regardless of linkage geometry [12]. Hence, the lower force plateau at 10 pN must be due to unzipping events with a greater gain in length per unzipped coiled coil, while the higher force plateau at 25 pN must be due to overstretching with the smaller gain in length per overstretched coiled coil. A simple two state equilibrium model (see methods section) allowed us to fit the force extension data and determine the number of coiled coils involved in the different linkage geometries.

In the force trace in figure 2, we determined three unzipping events and seven overstretching events. Hence, seven of the 15 Ig27 domains in the stretched coiled coil linked Ig27 polyprotein have been assembled by coiled coils in overstretching geometry, while three have been assembled in unzipping geometry II. The remaining five Ig27 domains have thus been added to the polyprotein in geometry III, where the linking coiled coils are shielded from unzipping by the blocking disulphide bonds. The trace in figure 2 thus closely resembles the expected ratio of linkage geometry formation of $50:25:25\%$ for each of the three possible linkage geometries. This behaviour is consistently found throughout the investigated coiled coil assembled Ig27 polyproteins and provides evidence that the coiled coil based assembly process is independent of the target proteins chemical properties and should hence depend only on sterical constraints. Size exclusion chromatography shows that Ig27 polyproteins containing more than 30 individual coiled coil linked Ig27 units (see figure 3(d)) can be easily assembled using coiled coil tags.

What is the potential advantage of coiled coil mediated polymerization over purely chemical polymerization exploiting the reactivity of cysteine residues as has been shown previously [19, 20]? First, the degree of polymerization using coiled coils is significantly higher than for chemical polymerization. Another clear disadvantage of chemical polymerization becomes evident when the termini of the protein subunits are located close to each other. Circular dimerization will prevent polymerization in these cases. Green fluorescent protein (GFP) is an example of such a protein, where N and C termini lie close to each other (see figure 3). Indeed, previous attempts to polymerize GFP via reactive cysteines located at the N and C terminus have failed (unpublished results). In a first attempt, we therefore flanked individual GFP molecules by LZ10 sequences and investigated their capability to form linear polyproteins. As is the case for chemical polymerization the assembly of coiled coil flanked GFP molecules ceases at the level of dimers. The three modes of assembly are illustrated in figure 3(a). While in the case of Ig27, all different modes of assembly are possible (see figure 2), for GFP the sterical repulsion between the GFP subunits prevent coiled coil formation in the cases illustrated in figure 3(a) middle and right. In the third mode of assembly where the two GFP subunits lie opposite of each other assembly will terminate at the level of dimers as illustrated in figure 3(a) left. Indeed, chromatography shows that in fact only monomers and dimers are present in such a sample (see figure 3(e)). However, by introducing a longer coiled coil structure, the sterical constraints preventing assembly of GFP can be overcome. To this end, we flanked GFP N-terminally with a single LZ10 sequence and C-terminally with a much longer, triple LZ10 sequence. The resulting fusion protein is shown schematically in figure 3(b). Now, the assembly modes where at least one long zipper is involved (figure 3(b) left and middle) are possible again. Again, chromatography confirms the expected assembly into polyproteins (see figure 3(e)). As expected, AFM force spectroscopy allows the observation of unfolding events from long GFP polyproteins (see figure 3(c)).

In summary, we have demonstrated the potential of a coiled coil based protein multimerization. We collected promising evidence that this technique may form the basis for a programmable protein self-assembly technique. Usage of higher-order than dimeric coiled coils should allow for the construction of three-dimensional structures. The presented technique for the self-assembly of a target protein into linear polyprotein chains will find immediate application in single molecule force spectroscopy. Polyproteins are an essential prerequisite for force spectroscopy experiments [20]–[22]. Complicated recombinant protocols
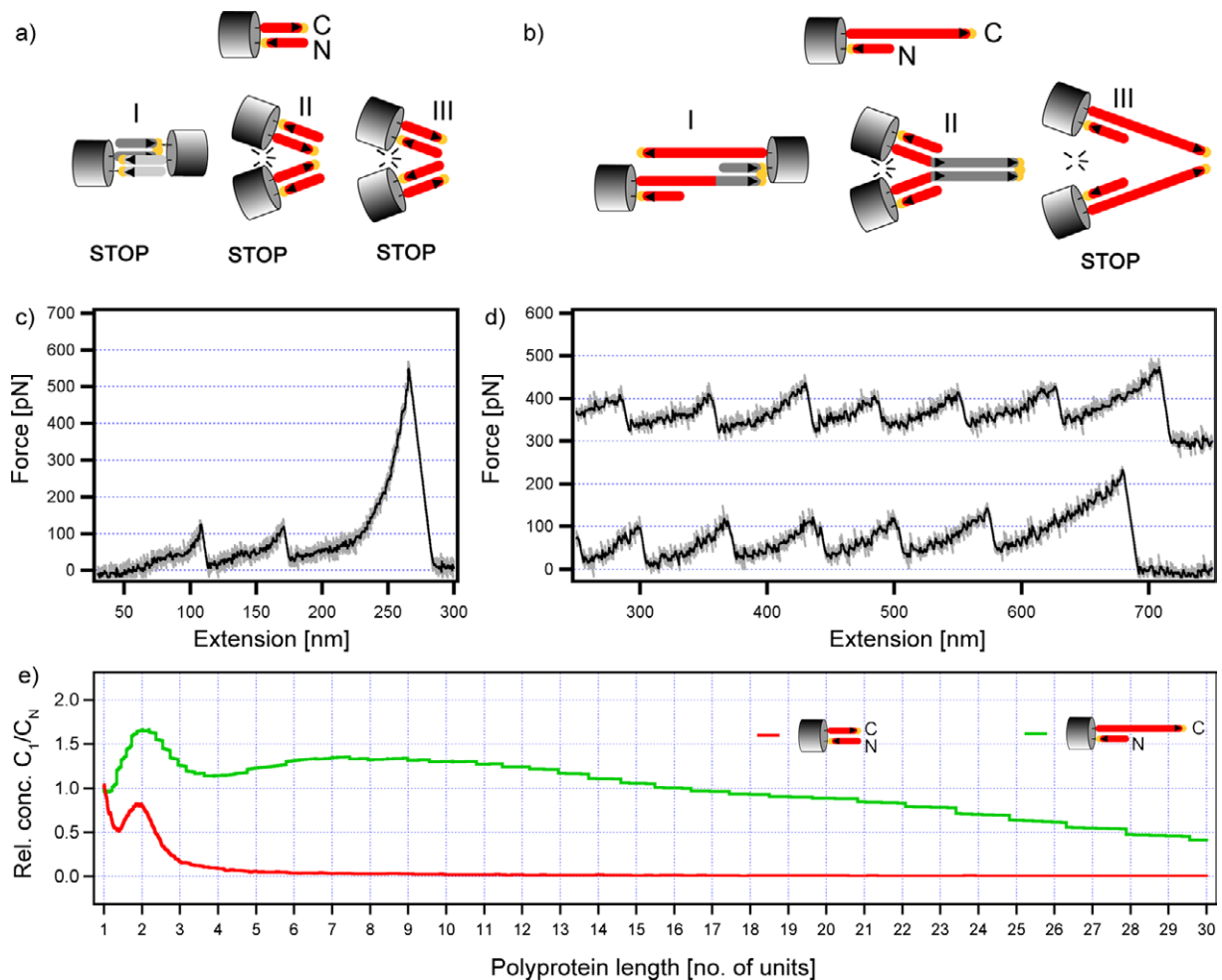
Figure 3. (a) Schematic representation of a GFP flanked terminally by two equal coiled coil binding partners, by geometry of GFP located spatially on the same side of the protein. The linkage geometry I will have a strong tendency for cyclization, since a simple 180° flip around one of the coiled coils will suffice to bring the two remaining coiled coil binding partners in correct dimerization orientation. In geometries II and III, dimerization will be sterically strongly hindered. (b) Scheme of GFP flanked terminally by coiled coil binding partners of different length. Geometries I and II are now open for further polymerization while dimerization along geometry III is still hindered. (c) Force-extension response of a coiled coil mediated GFP dimer as in (a) showing two GFP unfolding events. (d) Typical force extension response of coiled coil assembled GFP polyproteins as in (b), displaying 5 and 6 GFP unfolding events, respectively. (e) Distribution of polyprotein lengths in protein solutions containing GFP flanked with equal coiled coil binding partners as in (a) (red line) and containing GFP flanked with coiled coil binding partners of different length as in (b) (green line) as obtained by size exclusion chromatography.

have been developed for the artificial construction of polyproteins. Coiled coil based protein polymerization is an easy and fast method to polymerize proteins. Specifically, the loading rate independent signature of coiled coil unzipping and overstretching may serve as an integral force calibration standard in single molecule force spectroscopy.

Methods

Expression vectors have been prepared by restriction and ligation of coding sequences into pRSET5d vector (Novagene). Protein expression was done with BL21-CodonPlus (DE3)-RIL *E. coli* cells (Stratagene). Proteins were purified with Ni-NTA based affinity chromatography. Assembly of polyproteins occurs rapidly at protein concentrations above $20\,\mu$M already in the expression cells and during purification steps. Further treatment is not necessary. All force spectroscopy measurements were performed on a custom-built AFM. Gold-coated cantilevers (BioLevers, Olympus, Tokyo) with spring constant of $6\,\text{pN}\,\text{nm}^{-1}$ were used. For the measurements, the protein solutions were applied to a clean glass surface.

In order to determine the number of coiled coils participating in the overstretching and unzipping-like linkage geometries we fitted a two state model to the force extension data. The model considers every single coiled coil to fold two-state like in thermodynamic equilibrium. The ratio between unfolded and folded coiled coils is Boltzmann-distributed where the energetic difference $\Delta G(F)$ between the two states depends on the applied force F [23]:

$$\Delta G(F) = \Delta G_0 + \Delta G_{\text{stretch}}(F, L_{\text{folded}}, L_{\text{unfolded}}) - F \cdot \Delta L \cdot x_{\text{frac}}(F), \tag{1}$$

where ΔG_0 denotes the folding free energy at zero force ($\Delta G_0 = 24\,k_{\text{B}}T$ for LZ10). $\Delta G_{\text{stretch}}$ accounts for the difference in the work required to stretch an unfolded coiled coil to a given force F compared to the work required to stretch a folded coiled coil to the same force. The last term in equation (1) denotes the reversible work upon unfolding of a coiled coil at a force F. x_{frac} denotes the fractional extension of the coiled coil relative to its contour length. We calculated x_{frac} by inverting the WLC-interpolation formula [24] with a persistence length of $p = 0.7\,$nm. The overstretching and unzipping linkage geometries differ by the length increase ΔL upon unfolding (unzipping: $L_{\text{zip-folded}} = 0.9\,$nm, $L_{\text{zip-unfolded}} = 25.6\,$nm, $\Delta L_{\text{zip}} = 24.7\,$nm; overstretching: $L_{\text{str-folded}} = 5.5\,$nm; $L_{\text{str-unfolded}} = 12.8\,$nm; and $\Delta L_{\text{overstretch}} = 7.3\,$nm). The contribution of all coiled coils to the extension of the polyprotein at given force F is then:

$$X(F) = N_{\text{unzipping}} x_{\text{frac}}(F) \cdot \left\{ \frac{L_{\text{zip-unfolded}}}{1 + e^{\Delta G_{\text{zip}}(F)}} + \frac{L_{\text{zip-folded}}}{1 + e^{-\Delta G_{\text{zip}}(F)}} \right\}$$

$$+ N_{\text{overstretching}} x_{\text{frac}}(F) \cdot \left\{ \frac{L_{\text{str-unfolded}}}{1 + e^{\Delta G_{\text{str}}(F)}} + \frac{L_{\text{str-folded}}}{1 + e^{-\Delta G_{\text{str}}(F)}} \right\}.$$

Fitting this expression to the force extension data allowed us to determine the number of unzipping linkage geometries ($N_{\text{unzipping}}$) and the number of overstretching geometries ($N_{\text{overstretching}}$).

Acknowledgment

IOP Institute of Physics **Φ** DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

References

[1] Jahn R and Scheller R H 2006 SNAREs—engines for membrane fusion *Nat. Rev. Mol. Cell Biol.* **7** 631–43

[2] Lupas A N and Gruber M 2005 The structure of alpha-helical coiled coils *Adv. Protein Chem.* **70** 37–78

[3] Gruber M and Lupas A N 2003 Historical review: another 50th anniversary—new periodicities in coiled coils *Trends Biochem. Sci.* **28** 679–85

[4] Woolfson D N and Ryadnov M G 2006 Peptide-based fibrous biomaterials: some things old, new and borrowed *Curr. Opin. Chem. Biol.* **10** 559–67

[5] Charest G and Lavigne P 2006 Simple and versatile restraints for the accurate modeling of alpha-helical coiled-coil structures of multiple strandedness, orientation and composition *Biopolymers* **81** 202–14

[6] Kim B M and Oakley M G 2002 A general method for selection and screening of coiled coils on the basis of relative helix orientation *J. Am. Chem. Soc.* **124** 8237–44

[7] Shih W M, Quispe J D and Joyce G F 2004 A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron *Nature* **427** 618–21

[8] Seeman N C 2005 DNA enables nanoscale control of the structure of matter *Q. Rev. Biophys.* **38** 363–71

[9] Mathieu F, Liao S, Kopatsch J, Wang T, Mao C and Seeman N C 2005 Six-helix bundles designed from DNA *Nano Lett.* **5** 661–5

[10] Liao S and Seeman N C 2004 Translation of DNA signals into polymer assembly instructions *Science* **306** 2072–4

[11] O'Shea E K, Klemm J D, Kim P S and Alber T 1991 X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil *Science* **254** 539–44

[12] Bornschlögl T and Rief M 2006 Single molecule unzipping of coiled coils: sequence resolved stability profiles *Phys. Rev. Lett.* **96** 118102

[13] Rief M, Gautel M, Oesterhelt F, Fernandez J M and Gaub H E 1997 Reversible unfolding of individual titin immunoglobulin domains by AFM *Science* **276** 1109–12

[14] Rief M, Gautel M and Gaub H E 2000 Unfolding forces of titin and fibronectin domains directly measured by AFM *Adv. Exp. Med. Biol.* **481** 129–36

[15] Marszalek P E *et al* 1999 Mechanical unfolding intermediates in titin modules *Nature* **402** 100–3

[16] Li H *et al* 2002 Reverse engineering of the giant muscle protein titin *Nature* **418** 998–1002

[17] Sarkar A, Caamano S and Fernandez J M 2007 The mechanical fingerprint of a parallel polyprotein dimer *Biophys. J.* **92** L36–8

[18] Schwaiger I, Sattler C, Hostetter D R and Rief M 2002 The myosin coiled-coil is a truly elastic protein structure *Nat. Mater.* **1** 232–5

[19] Dietz H, Berkemeier F, Bertz M and Rief M 2006 Anisotropic deformation response of single protein molecules *Proc. Natl. Acad. Sci. USA* **103** 12724–8

[20] Dietz H *et al* 2006 Cysteine engineering of poly proteins for single molecule force spectroscopy *Nat. Protoc.* **1** 80–4

[21] Fisher T E, Marszalek P E and Fernandez J M 2000 Stretching single molecules into novel conformations using the atomic force microscope *Nat. Struct. Biol.* **7** 719–24

[22] Carrion-Vazquez M, Oberhauser A F, Fisher T E, Marszalek P E, Li H and Fernandez J M 2000 Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering *Prog. Biophys. Mol. Biol.* **74** 63–91

[23] Tinoco I and Bustamante C 2002 The effect of force on thermodynamics and kinetics of single molecule reactions *Biophys. Chem.* **101–102** 513–33

[24] Bustamante C, Marko J F, Siggia E D and Smith S 1994 Entropic elasticity of lambda-phage DNA *Science* **265** 1599–600