Single Molecule Protein Unfolding and Refolding using Atomic Force Microscopy

Thomas Bornschlögl and Matthias Rief

Physikdepartment der TU München, 85748 Garching, Germany

Key Words:

Atomic force microscopy, single molecule technique, mechanical protein unfolding and refolding, poly-protein design, cystein-engineering, coiled-coil-engineering

Abstract

Over the past few years atomic force microscopy (AFM) became a prominent tool to study the mechanical properties of proteins and protein interactions on a single molecule level. Atomic force microscopy together with other mechanical, single molecule manipulating techniques (1) made it possible to probe biological molecules in a way which is complementary to single molecule methods using chemicals or temperature as denaturant (2). For example atomic force microscopy offered new insights into the process of protein folding and unfolding by probing single proteins with mechanical forces. Since many proteins fulfill mechanical function or are exerted to mechanical forces in their natural environment, atomic force microscopy allows to target physiologically relevant questions. Although the number of proteins unfolded with AFM continually increases (3) (4) (5) (6) the total number of proteins studied so far is still relatively small (7).

This chapter aims at giving protocol like instructions for people which are actually getting started using AFM to study mechanical protein unfolding or refolding. The instruction includes different approaches to produce poly-proteins or modular protein chains which are commonly used to screen for true single molecule AFM-data traces. Also the basic principles for data collection with AFM and the basic data analysis methods are explained. For people who want to study proteins that unfold at small forces or for people who want to study protein folding which also occurs typically at small forces (< 30 pN) an averaging technique is explained allowing to increase the force resolution in this regime. For topics which would go beyond the scope of this chapter - as for example the details about different cantilever calibration methods - references are provided.

1. Introduction

In order to anchor proteins between AFM-cantilever tip and substrate one can take advantage of an inherent property proteins have, which is that they stick to various surfaces by non-specific adsorption. The downside of this property is, that the exact anchoring points are not known and that by chance more than one single protein might bind between cantilever tip and surface. A elegant way to circumvent this disadvantages is to use a modular protein chain consisting of many similar or identical protein sub units. If such a modular protein chain binds to tip and surface the exact anchoring points are getting irrelevant because the force onto a protein inside the chain is transduced via its neighbors and their connection can be exactly determined. Also the unfolding of such a highly repetitive structure leads to a repetitive force vs. extension signal when probed with AFM. By taking advantage of this repetitive fingerprint it is possible to select traces where only one single molecule was bound between cantilever tip and surface. In order to study proteins that do not occur naturally as modular protein chains, different protein engineering approaches evolved. For example you can insert your protein of interest into an naturally occurring modular protein or you can directly construct poly-protein chains using coiled coils or “cysteine engineering”. These methods will be described in the first part of the Methods section. Using this modular protein chains the basic principles of an AFM measurement and how to select and analyze the data are explained

in the second part of the Methods chapter. An averaging technique, which allows to increase the force resolution in the low-force regime is explained in the third part of the Methods section.

2. Materials

2.1 Protein Design

1. T7 promotor based E. coli cytosolic expression vector (e.g. pET28 family (Novagen) or pRSET family (Invitrogene))
2. BL21 CodonPlus(DE3)-RIL E.coli Competent Cells (Stratagene)
3. XL10 Gold Ultracompetent E.coli Cells (Stratagene)
4. QiaPrep Spin Miniprep Colums (Qiagen)
5. QiaQuick Gel Extraction Kit (Qiagen)
6. HisTrap HP or FF affinity chromatographie columns (GE Healthcare)
7. Mutagenesis primers (Metabion)
8. Phusion Site Directed Mutagenesis Kit (New England BioLabs)
9. Quickchange Multi Site-Directed mutagenesis Kit (Stratagene)
10. T4 DNA ligase (NewEngland BioLabs)
11. Restriction enzymes for the restiction sites of your choise (NewEngland BioLabs)
12. Lysis buffer: 50 mM NaH_2PO_4,300 mM NaCl adjust pH to 8.0 using NaOH
13. Size-exclusion analytical chromatography columns such as superose 10/300 GL columns (GE Healthcare)
14. French press or Sonicator
15. Centricon centrifugal filter units (Millipore)

2.2 Atomic Force Microscopy

1. Atomic force microscope (e.g. Asylum research)
2. Phosphate buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 (adjust to pH 7.4 with HCl)
3. clean glass coverslips or freshly gold evaporated glass coverslips
4. soft AFM cantilever (e.g. Bio Lever Type A or B, Olympus)

3 Methods

3.1 Protein Design and Purification

To study mechanical protein unfolding, typically researchers stretch modular proteins that have been adsorbed un-specifically between AFM-cantilever tip and surface. Because such a protein consists of similar sub-units one gets a highly repetitive force-extension pattern from one molecule which can be used to choose true single molecule unfolding events for data analysis. To study any protein of interest, different methods based on molecular biology can be used. One of these protein engineering methods is to insert the protein of interest into an already known, naturally occurring modular protein background such as the giant muscle protein titin or the *Dictyostelium discoideum* filamin (ddFLN) (Fig 1. a) - c)). Using this approach the protein of interest should differ in size (amount of amino acids forming the protein) or in average unfolding force from the used modular protein background. Only then the identification of the unfolding event associated with the inserted protein is easily doable. (See for example black unfolding event in Fig 1 c) for GFP inserted in an ddFLN background shown in gray)

Other protein engineering approaches take advantage of artificially designed poly-proteins consisting of identical protein subunits. By stretching such a construct with AFM one gets directly from one single molecule measurement many unfolding events associated with the

protein of interest, which facilitates the generation of datasets with statistical relevance. The design and production of such poly-proteins can be done in different ways:

One approach is to repeat the protein decoding DNA-sequence multiple times within the plasmid by using molecular biology methods (8). This leads after translation to a poly-protein chain where the individual proteins are connected via their N and C termini. This approach is very robust but the molecular weight of the protein construct and therefore the total number of protein subunits on the chain is limited.

A second approach is to add coiled coils onto the N and C termini of the protein of interest. In solution the coiled coil dimerizes thereby leading to a repetitive structure where the protein subunits of interest are connected via coiled coils (Fig 1 d) – f)). This approach leads to long poly-proteins containing a high number of subunits but their connection is still limited to the N and C-terminal ends of the protein (9).

A third way to produce poly-proteins is to introduce two solvent accessible cysteins into the protein structure of interest. In solution the protein then polymerizes via the formation of covalent disulfide bridges. This approach leads to long poly-proteins where the force onto a single molecule can be applied in an desired arbitrary three dimensional direction (10).

In the following we give protocols involving standard molecular biology methods to produce modular proteins which can be used for AFM studies, the advantages and drawbacks are summarized in the notes section.

3.1.1 Insert your protein of interest into a naturally occurring modular protein chain (such as ddFLN1-5)

1. Use a T7 promotor based E. coli cytosolic expression vector (e.g. pET28a+ or pRSET family)

2. Insert the DNA coding for the domains 1 to 5 of the *Dyctiostelium discoideum* actin binding protein (ddFLN1-5 or abp120; PubMed location: P13466) into the multiple cloning site of the expression vector (11). Use codons with high expression yield for E.coli (GeneArt). For cutting use restriction enzymes (New England BioLabs) of your choice. Take care that the restriction site occurs only once in your vector. For re-ligation use the T4 DNA ligase. Alternatively you can use different modular proteins as background protein. (See Note 2).

3 If not already present in the expression vector add an DNA sequence coding for 6 histidine residues (His_6 tag) to the sequence coding for ddFLN1-5 domains using for example the Phusion Site Directed Mutagenesis Kit (New England BioLabs). This allows protein purification using the HisTrap columns afterwards.

4 Insert two restriction sites of your choice between domains 2 and 3 (Aminoacid (AA) position 449 (11)) or between domains 3 and 4 (AA position 549) using the Quickchange Multi Site-Directed mutagenesis Kit or the Phusion Site Directed Mutagenesis Kit

5 Insert the sequence coding for your protein of interest (black sequence in Fig. 1 a) between these restriction sites

6 Confirm your final vector by sequencing (GATC)

7 For protein expression transform the plasmid into an E. coli expression strain such as the BL21-CodonPlus(DE3)-RIL and follow the manufacturers protocol.

8. Centrifuge the gained cell culture and re-suspend the pelleted cells in lysis buffer (e.g. re-suspend cells out of 500 ml cell culture in 30 ml lysis buffer)

9. Lyse the cells using a french press or a sonicator. Keep the cell suspension at 4°.

10. Pellet the cell fragments by centrifuging for 40 min at 30,000 g and 4° and keep the supernatant cell extract.

11. Apply the supernatant cell extract onto Ni-NTA affinity chromatography columns (using e.g. HisTrap FF columns). Wash the columns using the lysis buffer with increasing concentrations of imidazole (use for example 50, 100, 200, and 500 mM imidazol suspensions).

12. Use SDS-Page gel electrophoresis to determine the elution fraction that contains your protein and the degree of purity.

13. If needed you can increase the degree of purification by using size-exclusion analytical chromatography.

14. Determine protein concentration using e.g the Bradford assay (ca. 1 mg/ml is convenient)

15. Continue with the AFM measurement

3.1.2 Construct poly-proteins using coiled coils

1. Use a T7 promotor based E. coli cytosolic expression vector (e.g. pET28 family or pRSET family) containing a His_6 tag sequence (See point 3.1.1.3).

2. Insert two times in a row the DNA sequence coding for a homo-dimeric coiled coil as for example the LZ10 coiled coil (RMKQLEQKVEELLQKNYHLEQEVARLKQLVGECEG (12, 13).) using codons with high expression yield for E.coli (GeneArt)

3. If you prefer to use other homo-dimeric coiled coil sequences determine the a and d positions of the heptad repeat (underlined in the above LZ10 seuqence) using free programs such as coils (http://www.ch.embnet.org/software/COILS_form.html (14)) or paircoils (http://groups.csail.mit.edu/cb/paircoil/cgi-bin/paircoil.cgi) (15)) and change the last a or d position of your coiled coil to a cysteine. (Black circles in Fig 1 d))

4. Insert two restriction sites of your choice between both coiled coil coding sequences using the Phusion Site Directed Mutagenesis Kit

5. Insert your protein of interest between these restriction sites (black in Fig 1 d)).

6. Check your vector by sequencing (GATC)

7. Continue with protein expression and purification (Steps 7-14 of 3.1.1) and the AFM measurement. When you use the LZ10 coiled coil as polymerization motif the protein polymerization will take place already for protein concentrations > 25 μM. Thus polymerization will already occur during the steps for expression and purification so that no further incubation is needed. If you observe sedimentation in your protein solution due to aggregation spin down aggregates at 10,000g for > 10 min and use the supernatant for the AFM experiments.

3.1.3 Construct poly-proteins using “cysteine engineering”

1. Use your T7 promotor based E. coli expression vector containing a His_6 tag sequence.

2. Insert your protein of interest into the multiple cloning site of the vector (See Fig 1 g)

3. Choose two residues within the amino acid sequence of your protein of interest and change them into cysteins using the Phusion Site Directed Mutagenesis Kit. These cysteins will later on define the linkage geometry. Therefore it is critical that they are solvent accessible. Use the protein structure if it is available to determine these linkage points, if not available try the N and C termini as linkage points because they are often solvent accessible. Choosing residues which are located at opposite sites of the protein structure might decrease the hindrance for polymerisation due to sterical restrictions. Also check if your protein already contains solvent accessible cysteins and either exchange them with e.g. alanine or serine or use them as linkage points. For troubleshooting and further details see the protocol (10)

4. Continue with protein expression and purification (Steps 7-14 of 3.1.1).

5. Concentrate the purified protein in the lysis buffer solution to more than 0.2 mM using e.g. Centricon centrifugal filter units.

6. For polymerization incubate the protein solution for ca. 80h at 37° if this temperature does not affect your protein. Else incubate at lower temperatures for longer times. Check progress of the polymerization using e.g. SDS-Page with non-reducing SDS buffer. When the wanted degree of polymerization is reached (average of octamers is convenient for AFM measurements) dilute the protein solution to decrease further polymerization (e.g. dilute to 0.02 mM).

7. Continue with the AFM experiments. If you observe sedimentation in your protein solution due to aggregation spin down aggregates at 10,000g for > 10 min and use the supernatant for the AFM experiments.

3.2 Principle data collection and analysis using AFM

The development and availability of relatively easy to use commercial atomic force microscopes allows the application of this technique by researchers who are interested in single molecule protein folding and unfolding and who don't necessarily have to be AFM-specialists. Most of the commercial available instruments also include automated procedures e.g. in order to calibrate the force constants of the used cantilevers. For details on how to use your specific AFM, the relevant manual will provide much more useful information than this chapter is able to. Therefore we will only shortly explain the basic principles of AFM measurements and provide relevant references for further reading. Due to the unspecific binding of proteins between cantilever tip and surface the collected data will consist of a huge fraction of non-interpretable traces as for example traces where many molecules have bound in parallel. The second part of this chapter therefore provides a protocol how to select relevant and interpretable single molecule data when the modular poly-protein chains from section 3.1 have been used for the measurements. This section also explains the principle steps to determine the length increase due to protein unfolding.

3.2.1 Experimental Setup and data collection

1. Apply ca. 10 μl of the protein suspension to a clean glass cover slip or a freshly gold-evaporated glass surface and incubate for 20 min. To avoid air bubbles when you apply a drop

(ca. 30 µl) of PBS onto the cantilever (Bio Lever Type A or B, Olympus) use degassed PBS. Join both drops and align the AFM as shown in Fig 2 a), where a beam of light which is focused on the tip of the cantilevers back side is reflected into the middle of a 2 segment photo diode. To determine the bending of the cantilever read out the deflection D(t) by reading the intensities I measured on the two photo diodes A and B. ($D(t) = (I_A-I_B)/(I_A+I_B)$)

2. If not already available install an piezo control (Physik Instrumente) that allows e.g. to move the surface back and forth over distances of several µm with constant velocity. Applicable velocities for protein unfolding experiments range between 1 nm/s and 10000 nm/s (some of the limiting factors are instrument drift for slower pulling speeds and hydrodynamic effects for the upper border)

3. Gently approach the cantilver tip towards the surface. You can do so by observing the hysteresis in the deflection signal that originates from the hydrodynamic drag due to a repetitive, forward and backward movement at constant velocities (~ 5 µm/s) of the surface. This hysteresis will increase with decreasing cantilever-surface distance. When close to the surface, apply a triangular voltage signal to the piezo actuator leading to surface movement as shown in Fig 2 b). If you are close enough to the surface you should gain a deflection trace as the one shown in Fig 2 c) which defines the position of the surface (slightly tilted vertical line that also defines the proportionality constant p) and the zero line for the force acting on the cantilever (horizontal line).

4. Check your instrument drift. The deflection should stay constant when the surface is not moved. Particles swimming through the optical path of the light beam might lead to slow variations in the signal while cantilever drift caused e.g. by the bimetal effect might lead to an constantly creeping drift. Check also if the absolute distance of cantilever to surface stays constant over time. Without drift the curve shown in Fig 2 c) should be exactly reproducible after some time and forward and backward traces at slow velocities < 500 nm/s should not significantly differ

5. Start the measurement by repeatedly pressing the cantilever against the surface and by retracting it afterwards with constant velocity. To increase protein adsorption between cantilever tip and surface you can increase the time of cantilever-surface contact.

6. After your measurement calibrate the spring constant of the used cantilever far away from the surface. Commonly the method of thermal equilibration is used and already implemented in most commercial available AFMs. For further details see (16) (17) (18) (19).

7. Calculate from the deflection signal D(t) and from the absolute surface position signal x_0 the force as well as the tip to surface distance z(t) (extension) using:

$$F(t) = k_C/p\ D(t) \qquad \text{and} \qquad z(t) = x_0(t) - 1/p\ D(t)$$

where p can be calculated as shown in Fig 2 c).

3.2.2 Data selection

1. Select those force traces where only one single molecule has been attached between cantilever tip and surface while pulling. Use the following selection criteria:

- Only select sawtooth like pattern with more than 2 sawteeth. The sawteeth should appear equidistantly as in Fig 1 f) for the poly-proteins made with methods 3.1.2 and 3.1.3. (See

3.2.3 for how to measure contour length increases). When ddFLN was used as protein background the sawteeth should look like the grey ones in Fig 1 c).

- Preferentially select curves where the last sawtooth rises to higher force than the forgoing ones. This corresponds to a clear detachment or rupture event of the molecule. After this detachment you should not see any other rising force in the relaxation event of the cantilever.

- Preferentially select force traces where the effect of other bound molecules in the beginning of the force curve is negligible. (See fig 1 f) black event before 100 nm)

As an more objective method to select typical force traces associated with the measured protein you can use a pattern recognition algorithm (20).

2. Compare the force traces from at least 4 different experiments e.g by overlaying them. The contour length gains due to unfolding should be exactly reproducible for different measurements. A total amount of 50 - 100 different unfolding events is convenient. When you used a naturally protein as background (Method 3.1.1), the force traces should contain an additional unfolding event with reproducible length associated with the inserted protein of interest.

3. If you used method 3.1.1 and you observe unfolding events of the matrix protein without a new effect for your protein of interest check if your protein unfolds at low forces using method 3.3.2. Also it might be that your protein unfolds at forces which are much higher than the unfolding forces of the matrix protein. Check this by using method 3.1.2 or 3.1.3.

3.2.3 Determining contour length gain, number of unfolded amino acids and unfolding forces

1. The end to end distance of a polymer in solution increases in a non linear fashion with increasing force. As a quick way (e.g for the first data selection) to compare the length increases of the unfolding proteins you can take the distances between two consecutive sawteeth. Only compare distances that you measured at the same force.

2. Determine the contour length of the polymer by fitting the interpolation formula based on the worm like chain (WLC) model to your data (21):

$$F(z) = \frac{k_B T}{p}\left(\frac{1}{4(1-z/L)^2} - \frac{1}{4} + \frac{z}{L}\right)$$

where k_B is the Boltzmann's constant T is Temperature in Kelvin, L is the contour length (in this formula the total length of the polymer when an infinite force is applied) and p is the persistence length describing the flexibility of the polymer. In the force regime between 50-150 pN the formula reproduces experimentally measured data very well using a persistence length of 0.5 nm (22). (See Notes for other force regimes)

3. Determine the contour length increase due to protein unfolding by measuring the difference in contour length between to consecutive force peaks using step 2. The contour length increase can be used to calculate the amount of unfolding amino acids if the folded structure of the protein is known. The contour length increase ΔL is given by

$$\Delta L = n d_{AA} - d_{folded}$$

where n is the number of unfolding amino acids, d_{AA} is the contour length of a single amino acid and d_{folded} is the initial distance of the amino acids onto which force is applied in the native structure. Using ddFLN and titin as "gauge proteins" we find a value of $d_{AA} = 0.365 \pm 0.002$ nm. Since this value might slightly differ for your AFM you should determine your d_{AA} e.g by using method 3.1.1 and by measuring the contour length increases due to ddFLN unfolding (For ddFLN is n = 100 and $d_{folded} = 4.0 \pm 0.1$ nm).

4. Calculate the amount of unfolding amino acids for your protein of interest using step 3. If you find a value which is deviating from the expected one check if your protein does not show any pre-unfolded parts. When using method 3.1.1 or 3.1.2 you can check this by deleting the presumably unstable N or C-terminal parts of your protein and by redoing the AFM measurements. When using method 3.1.3 insert the Cysteins further away from the termini. To check if parts of your protein unfold already at lower forces use the Method explained in 3.3. Be aware that also stable intermediate states might be observed, leading to one ore more (23) force peaks occurring during protein unfolding (see e.g. the intermediate state of ddFLN4 in Fig 1 c) at ~ 50 nm (24) or the intermediate states occurring in Fig 1 i))

5. Determine the distribution of unfolding forces (the height of the individual sawteeth peaks relative to the 0 force line) for datasets collected at different pulling velocities respectively. From this data you can for example interpolate to the rate of unfolding at 0 force and reconstruct parts of the underlying energy landscape. There are different approaches including monte carlo simulations which have been applied to reconstruct the underlying energy landscape. For more details see (25) (6)

3.3 Protein folding and unfolding close to thermodynamic equilibrium

Atomic force microscopy can not only be used to observe the mechanical unfolding of proteins it also allows to study their refolding. Commonly the unfolding of globular proteins occurs at relatively high forces (> 50 pN) while they refold at forces in the lower pN regime where the resolution of commercially available cantilevers can become the limiting factor. It is demanding but possible to directly observe the refolding of such proteins (26) (27). Another approach which has been already used to deduce the folding properties of a protein is to retract the cantilever towards the surface after the protein has been successfully unfolded. After a certain time during which the protein was allowed to refold one can check if folding had occurred by stretching the whole protein again (28) (24). However AFM is perfectly suited to study protein unfolding and refolding even at very small forces if this process occurs close to thermodynamic equilibrium. In this case either many folding/unfolding transitions can be observed during one stretching cycle of one molecule (29) or many stretching as well as relaxing cycles with highly reproducible force extension pattern can be obtained on the same molecule (30) (12). This allows to easily separate the signal from unspecific background such as drift. It should be noted that the use of an optical trap comprising of a higher force resolution is an alternative for the study of proteins in this force regime, especially if constant force experiments are wanted (1). But the AFM with its higher force constants can provide insights in this low force regime that might be complementary to those measured with optical trapping. The following subchapter wants to provide a protocol for how the force resolution of an AFM in the low force regime can be enhanced by recording many folding/unfolding cycles and by averaging them.

3.3.1 Protein design and first steps

1. Use a naturally occurring modular protein or an artificially designed modular protein as matrix protein and insert only once your protein of interest (See 3.1.1). The used matrix protein should have been already probed with AFM and should be known to be stable enough to not interfere with the unfolding of your protein of interest. The method 3.1.1 facilitates the screening for true single molecule events compared to an poly-protein approach (3.1.2 and 3.1.3) because you expect that the unfolding forces of your protein of interest are small. Moreover the unfolding/folding behavior of one single protein might be already rather complex so that using the poly-protein approach might complicate the data analysis.

2. Start with the standard measurement and data analysis (3.2.1 and 3.2.2) at slow velocities (~ 100 nm/s). This will prove successful protein design and purification and it may give you a first idea of the unfolding behavior of your protein. Since your protein will unfold in the beginning of the force curve where often unspecific interactions occur (e.g grey effect in Fig 3 before ~ 75 nm) data collection using method 3.2.2 might be demanding.

3.3.2 High resolution measurements

1. Apply a consecutively increasing and decreasing voltage signal to the piezo control after the cantilever has touched the surface. The anticipated extension signal (Fig 3 b)) leads to consecutive unfolding/refolding cycles in the beginning of the force curve and thus in the low force regime and is followed by a total retraction from the surface after the last cycle. The calculation of the turning points (a and b in Fig 3 b)) is crucial:

- First approach the cantilever to the surface with constant velocity (leading to a force trace as shown in Fig. 2 c)).

- Use this signal to determine the point of contact of the cantilver tip with the surface (piezoposition 0 in Fig 2c)).

- The turning point a should not be to close to the surface to avoid unspecific attachment of additional proteins during the cycles. On the other hand it should not be to far from the surface to allow for protein refolding at low forces. A value of ~ 20 nm might be convenient to start with.

- The turning point b depends on the contour length increase which you expect from your protein. As a rough guess take the sum of the expected contour length increase of your protein of interest and the contour length of the folded matrix proteins and add 75% of this sum to the position of the turning point a. If you find very often an already unfolded matrix protein in the beginning of the measurement, add 75% of its contour length too. Note that your resolution diminishes with the amount of unfolded polypeptide because the stiffness of the system comprising of cantilever stiffness, stiffness of unfolded polypeptides and stiffness of folded proteins decreases. Therefore it is desirable to avoid pre-unfolded matrix proteins and to measure as close as possible to the surface.

2. To start only collect a few (~ 3) unfolding/refolding cycles per molecule at slow velocities (~ 100 nm/s). The average time after which the protein detaches from cantilever tip or surface is the limiting factor. Try to increase the number of cycles. You can do more cycles if the cycle length is small or if you are measuring at higher velocities. Very often you have to discard the first cycles because they are overlaid by some additionally bound proteins which detach during the first cycles.

3. Check if you see the same unfolding/refolding event in the cycles using box smoothing at low velocities (~ 100 nm/s). Note that this becomes more and more difficult with increasing pulling velocity. Already at velocities of ~ 800 nm/s the unfolding effect might be hard to identify (See Fig 3 c)). Check if you find the expected contour length increase (See 3.2.3). If you don't observe an refolding event and no unfolding events in the following cycles, decrease the turning point position a to facilitate protein folding.

4. Check your instrument drift. You can do this by measuring the contour lengths of the polypeptide after your protein of interest has unfolded (L_0 in Fig 3 a)). This length should not change during the cycles. For illustration are in Fig 3 c) different forward cycles plotted. After protein unfolding the forward cycles follow the grey WLC traces (grey shaded area) that are stacked in 20 pN steps for better comparison.

5. Average the forward and backward traces respectively. Average the deflection vs x_0 signal rather than the force vs. extension signal to avoid averaging effects on the turning point positions. Alternatively you can tilt the force vs. extension traces so that the noise gets vertical again for averaging. This should lead to traces as shown in Fig 3 d) which was measured on one single molecule. In Fig 3 d) the averaged forward trace is shown in black and the backward trace in grey.

6. Determine the area between unfolding and refolding traces which tells you how far the system is away from thermodynamic equilibrium. If you are close to equilibrium (if the hysteresis is < 10 k_BT) the equilibrium averaging method is rectified and you can use a model based on thermodynamic equilibrium (31) to describe your data. If the hysteresis is higher try to decrease it by decreasing the pulling velocity.

7. Do the experiment with different pulling velocities (between 100 and 2000 nm/s)

8. To further increase resolution you can average different pre-averaged traces gained from step 5 if they are comparable. Comparable means that they must have been measured under the same buffer conditions at the same pulling velocities and they must show the same unfolded polypeptide spacer (same absolute value for L_0).

9. You can still use the averaging method for systems which are further away from equilibrium. However it might be easier to determine e.g the unfolding force histograms directly from slowly pulled data than reconstructing it from the averaged traces (32).

10. To measure at very slow velocities (~ 1 nm/s) you might use Ni-NTA coated surfaces in order to increase the time before detachment of protein from the surface occurs (29). Also in this regime measurement drift gets the limiting factor, therefore try to keep drift effects low e.g. by using short light-pathways of the detection system through the sample solution.

Notes

1. The different approaches to construct poly-proteins have different assets and drawbacks which are summarized in table 1. Since some of the them might not immediately work for yor protein of interest you should try different approaches in parallel.

Approach	Using naturally occurring, already probed modular proteins as background (3.1.1)	Constructing poly proteins using coiled coils (3.1.2)	Constructing poly-proteins using cysteine engineering (3.1.3)
Advantages	- Straight forward method because the matrix protein is already known to be measurable with AFM	- Produces long chains of poly-proteins allowing for a quick collection of statistically relevant data	- Produces long chains of poly-proteins allowing for a quick collection of statistically relevant data

	- The matrix protein can be used as "gauge protein" to determine parameters such as d_{aa} (See 3.2.3) - Allows to easily identify complex unfolding pattern e.g. for proteins showing many unfolding intermediate states (23)	- The coiled coil fingerprint can be used to determine the 0 force very accurately.	- Force can be applied in any wanted three dimensional direction - Can be used to determine structural information by mechanical triangulation (22)
Disadvantages	- Force application only in N-C terminal direction - The collection of statistical relevant data may be demanding because you get only one unfolding event per measured molecule	- Force application only in N-C terminal direction - Does not necessarily work for all proteins of interest because of possible sterical hindrances (9)	- Determination of linkage points may be demanding when the protein structure is not known - Does not necessarily work for all proteins because the reactivity of cysteins to form a disulfide bond may be low for some proteins

2. Alternatively to the modular background protein ddFLN used in method 3.1.1 you can take other naturally occurring modular proteins. For examples you can insert your protein of interest in the middle of the domains Ig27-Ig34 from the giant muscle protein titin. The domain boarders as are given in (33) the corresponding AFM experiments in (28). Because all proteins in the chain are in series and therefore simultaneously exerted to the mechanical force, the order in which the proteins unfold depends on their stability. If your protein is the weakest in the chain it will likely unfold at the beginning of the force curve and this event might then often be masked by additional proteins that have bound in parallel. Therefore it might be convenient to insert your protein of interest in a mechanically stable protein background (e.g. titin) as well as in a less stable background (e.g ddFLN). You might also construct your own matrix protein with e.g. the following order: 2 x titin – 2 x ddFLN – your protein of interest- 2 x ddFLN – 2 x-titin

3. When the persistence length p is held constant, the interpolation formula given in 3.2.3 only reproduces experimental data within a certain force regime very well. As mentioned above a persistence length of 0.5 nm reproduces experimental data in the force regime between 50 and 150 pN. To fit data collected in the force regime between 150-300 pN a persistence length of 0.35 nm can be used. For data measured in the low force regime (e.g. with method 3.3.2) between 0-30 pN a persistence length of 0.7 nm leads to good agreement. If you want to compare with very high precision (e.g. to get structural information) different contour length increases measured in different force regimes you have to correct for this effect (22).

Acknowledgements

We thank Hendrik Dietz for providing the data traces in Fig 1 c) and i) and for his major contribution to the protein engineering approaches 3.1.2 and 3.1.3. We also thank Morten Bertz, Michael Schlierf, Felix Berkemeier, Jan Philipp Junker, Roland Heym, Frauke König and Ingo Schwaiger for their important help developing the protein engineering approaches and the data analyzation methods.

1. Bustamante, C., Macosko, J.C., and Wuite, G.J., (2000) Grabbing the cat by the tail: manipulating molecules one by one. *Nat Rev Mol Cell Biol* **1,** 130-136.
2. Borgia, A., Williams, P.M., and Clarke, J., (2008) Single-molecule studies of protein folding. *Annu Rev Biochem* **77,** 101-125.
3. Linke, W.A., and Grutzner, A., (2008) Pulling single molecules of titin by AFM--recent advances and physiological implications. *Pflugers Arch* **456,** 101-115.
4. Zhuang, X., and Rief, M., (2003) Single-molecule folding. *Curr Opin Struct Biol* **13,** 88-97.
5. Clausen-Schaumann, H., Seitz, M., Krautbauer, R., and Gaub, H.E., (2000) Force spectroscopy with single bio-molecules. *Curr Opin Chem Biol* **4,** 524-530.
6. Rounsevell, R., Forman, J.R., and Clarke, J., (2004) Atomic force microscopy: mechanical unfolding of proteins. *Methods* **34,** 100-111.
7. Oberhauser, A.F., and Carrion-Vazquez, M., (2008) Mechanical biochemistry of proteins one molecule at a time. *J Biol Chem* **283,** 6617-6621.
8. 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.
9. Dietz, H., Bornschlögl, T., Heym, R., König, F., and Rief, M., (2007) Programming protein self assembly with coiled coils. *New J. Phys* **9,** 424.

10. Dietz, H., Bertz, M., Schlierf, M., Berkemeier, F., Bornschlögl, T., Junker, J.P., and Rief, M., (2006) Cysteine engineering of polyproteins for single-molecule force spectroscopy. *Nat Protoc* **1,** 80-84.
11. Fucini, P., Koppel, B., Schleicher, M., Lustig, A., Holak, T.A., Muller, R., Stewart, M., and Noegel, A.A., (1999) Molecular architecture of the rod domain of the Dictyostelium gelation factor (ABP120). *J Mol Biol* **291,** 1017-1023.
12. Bornschlogl, T., and Rief, M., (2006) Single molecule unzipping of coiled coils: sequence resolved stability profiles. *Phys Rev Lett* **96,** 118102.
13. Zitzewitz, J.A., Ibarra-Molero, B., Fishel, D.R., Terry, K.L., and Matthews, C.R., (2000) Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J Mol Biol* **296,** 1105-1116.
14. Lupas, A., Van Dyke, M., and Stock, J., (1991) Predicting coiled coils from protein sequences. *Science* **252,** 1162-1164.
15. Berger, B., Wilson, D.B., Wolf, E., Tonchev, T., Milla, M., and Kim, P.S., (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci U S A* **92,** 8259-8263.
16. Hutter, J.L., and Bechhoefer, J., (1993) Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64,** 1868.
17. Butt, H.J., and Jaschke, M., (1995) Calculation of thermal noise in atomic force microscopy. *Nanotechnology* **6,** 1.
18. Burnham, N.A., Chen, X., Hodges, C.S., Matei, G.A., Thoreson, E.J., Roberts, C.J., Davies, M.C., and Tendler, S.J.B., (2003) Comparison of calibration methods for atomic-force microscopy cantilevers. *Nanotechnology* **14,** 1.
19. Walters, D.A., Cleveland, J.P., Thomson, N.H., Hansma, P.K., Wendman, M.A., Gurley, G., and Elings. V., (1996) Short cantilevers for atomic force microscopy. *Rev. Sci. Instrum.* **67,** 3583.
20. Dietz, H., and Rief, M., (2007) Detecting molecular fingerprints in single molecule force spectroscopy using pattern recognition. *Jap J Appl Phys* **46-8B,** 5540.
21. Bustamante, C., Marko, J.F., Siggia, E.D., and Smith, S., (1994) Entropic elasticity of lambda-phage DNA. *Science* **265,** 1599-1600.
22. Dietz, H., and Rief, M., (2006) Protein structure by mechanical triangulation. *Proc Natl Acad Sci U S A* **103,** 1244-1247.
23. Bertz, M., and Rief, M., (2008) Mechanical unfoldons as building blocks of maltose-binding protein. *J Mol Biol* **378,** 447-458.
24. Schwaiger, I., Schleicher, M., Noegel, A.A., and Rief, M., (2005) The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep* **6,** 46-51.
25. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., and Schulten, K., (1997) Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J* **72,** 1568-1581.
26. Fernandez, J.M., and Li, H., (2004) Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303,** 1674-1678.
27. Schlierf, M., Berkemeier, F., and Rief, M., (2007) Direct observation of active protein folding using lock-in force spectroscopy. *Biophys J* **93,** 3989-3998.
28. 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,** 1109-1112.
29. Junker, J.P., Ziegler, F., and Rief, M., (2009) Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science* **323,** 633-637.
30. Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V., and Marszalek, P.E., (2006) Nanospring behaviour of ankyrin repeats. *Nature* **440,** 246-249.
31. Tinoco, I., Jr., and Bustamante, C., (2002) The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophys Chem* **101-102,** 513-533.
32. Bornschlogl, T., and Rief, M., (2008) Single-molecule dynamics of mechanical coiled-coil unzipping. *Langmuir* **24,** 1338-1342.
33. Politou, A.S., Gautel, M., Improta, S., Vangelista, L., and Pastore, A., (1996) The elastic I-band region of titin is assembled in a "modular" fashion by weakly interacting Ig-like domains. *J Mol Biol* **255,** 604-616.

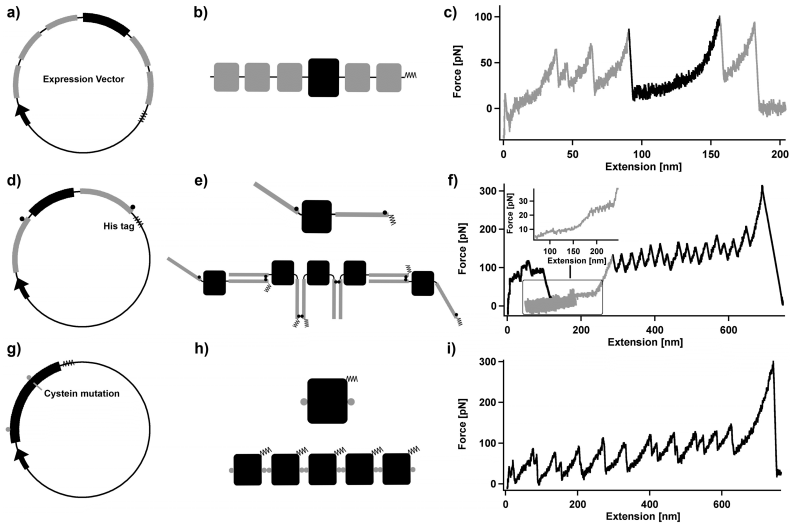

FIGURE 1: Different protein engineering approaches to construct modular protein chains for mechanical protein unfolding experiments using AFM. a)-c) Insertion of the sequence coding for the protein of interest into an naturally occurring modular protein background (such as ddFLN1-5) leads after translation to the modular protein schematically shown in b). AFM measurements reveal the unfolding events of the background protein (grey) as well as the unfolding event of the inserted protein (black in c)). The example curve was measured on GFP inserted between domains 3 and 4 of ddFLN1-5. Data was kindly provided by H. Dietz. d)-f) Flanking the sequence coding for your protein of interest by coiled coil sequences leads after translation to the protein construct shown on top in e). Due to coiled coil dimerization

the single proteins will self assemble into elongated modular protein chains. f) AFM measurements lead to repetitive sawtooth like traces where every sawteeth corresponds to an unfolding event of the protein of interest (black). In the low force regime (here before 200 nm) the unfolding of the connecting coiled coils can be observed using method 3.3.2. A force plateau at ~ 10 pN gets visible which is associated with the unzipping of several coiled coils in series while the force plateau at ~ 25 pN is connected to the overstretching of the coiled coils. (Inset). g) –i) Mutation of two surface exposed amino acid residues in your protein of interest to cysteins leads to the structure schematically shown on top in h). At high concentrations the cysteins will build covalent disulfide bridges which leads to elongated modular protein chains. i) AFM Measurement on such chains reveals sawtooth like force pattern where the sawteeth are associated with the unfolding event of your protein of interest. The trace was measured on GFP connected via the amino acid positions 3 and 212 where also intermediate states can be observed. Data provided by H. Dietz. The traces in c) and i) are measured with the Type A Biolever while f) was measured with the Type B cantilever (Olympus).

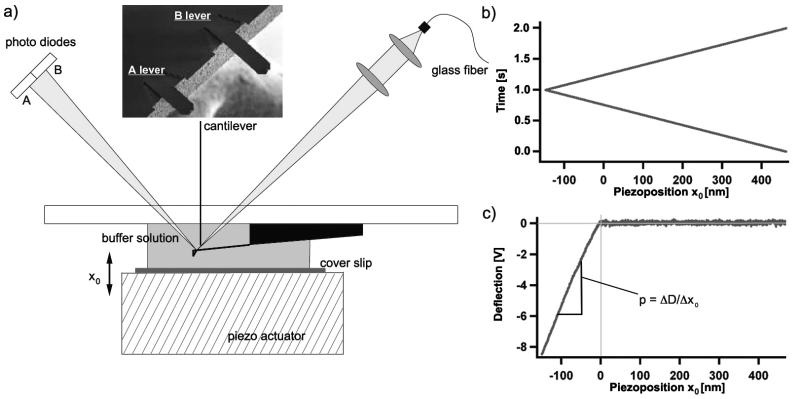

FIGURE 2: Experimental setup. a) schematic drawing of the AFM setup. b) By applying a triangular voltage signal to the piezo actuator the surface should move towards the cantilever and then away from the cantilever with constant velocity. c) Deflection signal when no protein has been attached between cantilever tip and surface after contact

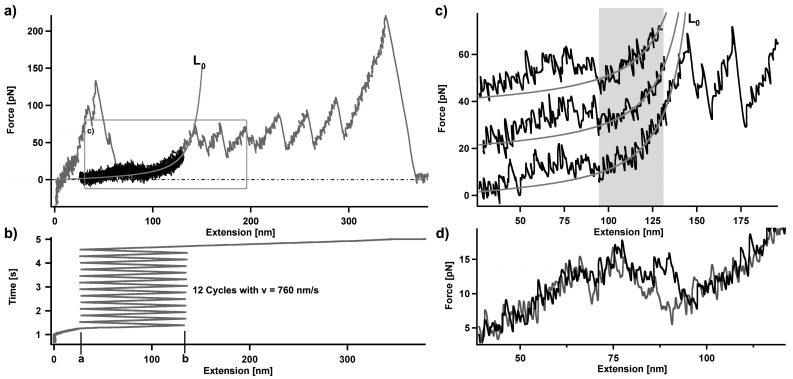

FIGURE 3. High resolution experiments for proteins which unfold and refold close to thermodynamic equilibrium. a) Example curve to study coiled coil unzipping (black at forces < 20 pN) within a ddFLN background protein (unfolding shown in grey at forces of ~ 60 pN) gained with an approach similar to method 3.1.1. b) The surface is consecutively approached and retracted from the cantilever tip which leads to multiple unfolding/refolding cycles. c) Three single unfolding traces stacked by 20 pN steps for better comparison. d) Outcome of averaging the unfolding (black) and refolding (grey) traces respectively.