Cysteine engineering of polyproteins for single-molecule force spectroscopy

Hendrik Dietz, Morten Bertz, Michael Schlierf, Felix Berkemeier, Thomas Bornschlögl, Jan Philipp Junker & Matthias Rief

Physik Department, Technische Universität München, James-Franck-Strasse, D-85748 Garching bei München, Germany. Correspondence should be addressed to H.D. (dietz@ph.tum.de).

Published online 27 June 2006; doi:10.1038/nprot.2006.12

Single-molecule methods such as force spectroscopy give experimental access to the mechanical properties of protein molecules. So far, owing to the limitations of recombinant construction of polyproteins, experimental access has been limited to mostly the N-to-C terminal direction of force application. This protocol gives a fast and simple alternative to current recombinant strategies for preparing polyproteins. We describe in detail the method to construct polyproteins with precisely controlled linkage topologies, based on the pairwise introduction of cysteines into protein structure and subsequent polymerization in solution. Stretching such constructed polyproteins in an atomic force microscope allows mechanical force application to a single protein structure via two precisely controlled amino acid positions in the functional three-dimensional protein structure. The capability for site-directed force application can provide valuable information about both protein structure and directional protein mechanics. This protocol should be applicable to almost any protein that can be point mutated. Given correct setup of all necessary reagents, this protocol can be accomplished in fewer than 10 d.

INTRODUCTION

The mechanics of proteins are important in many processes in living organisms. A large fraction of the proteins in our body have structural and thus also mechanical function. Examples include muscle proteins, cytoskeletal proteins and proteins of the extracellular matrix. But mechanical processes are also important for our understanding of enzyme activity and molecular motors[1]. Only recently, single-molecule methods like atomic force microscope–based force spectroscopy have made the mechanical properties of protein molecules experimentally accessible[2]. In force spectroscopy experiments, it is crucial to distinguish true single-molecule events from an abundant nonspecific background. Typically, researchers use polyproteins in which many identical subunits are linked covalently in series via their respective N- and C termini. Stretching of such polyproteins in an atomic force microscope will typically result in a characteristic equidistant sawtooth pattern that facilitates detection of single-molecule events. However, only a minority of proteins occur naturally as polyproteins. Therefore, recombinant protocols for the construction of artificially linked polyproteins have been developed[3,4]. There are several limitations to such recombinant strategies: the molecular weight of the polyproteins (and thus the number of individual units in the chain) is generally restricted to less than 150 kDa, and the application of mechanical forces is limited to the termini of the proteins. In this protocol, we describe the step-by-step procedure to construct polyproteins by cysteine engineering. The methodology is much simpler and at the same time overcomes the limitations of recombinant polyprotein construction. The basic concept of cysteine engineering of polyproteins is outlined in **Figure 1**. Cysteine-based polymerization of proteins has been successfully accomplished by Yang *et al.*[5]. Their approach, however, explicitly called for crystallization of the sample in order to obtain polyproteins containing more than three units. The current protocol yields long polymers (up to 30 units) in solution.

There are several new applications that can make use of protein polymerization via cysteine engineering. Because cysteines can be selectively positioned almost anywhere in the protein structure, mechanical force can be applied at predefined sites. Such site-directed force application to a folded protein structure allows for direct measurement of intramolecular distances with Angstrom-level precision. This method has just recently been applied to determine the absolute position of selected amino acids by mechanical triangulation[6] and can thus be used to obtain precise structural information from single protein molecules. In addition, access to protein mechanics has so far been limited to mostly the N- to C-terminal spatial direction. Cysteine engineering can now be applied to probe mechanical protein stability in many more directions. We anticipate that cysteine engineering of polyproteins will find numerous applications answering questions relating to the conformational mechanics of proteins.

Possible limitations

Cysteine engineering cannot control the direction of inclusion of individual modules in the polymerized protein chain. Because of inversion symmetry in the geometry of force application, this uncertainty in the molecular construction will not compromise force extension data in any way. However, interface stabilization (i.e., domain-domain interaction) of neighboring protein modules may, in fact, be a problem arising in multimeric proteins. Yet this is a general problem to all polymerization strategies. Domain-domain interactions can be investigated in detail by following the procedures of Rounsevell *et al.*[7].

In some cases, it may be advantageous to minimize sterical hindrance of polymerization or surface interaction effects as mentioned above by creating copolymers. In addition, the incorporation of protein modules of known force extension signature may provide reference peaks in the atomic force microscope traces as

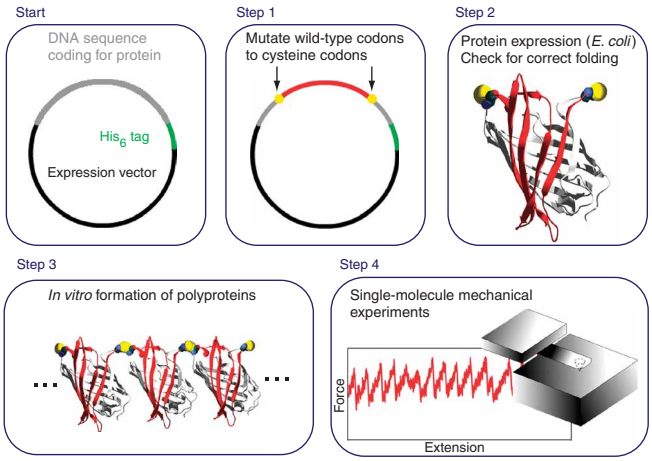

Figure 1 | Basic concept of cysteine engineering of polyproteins for single-molecule force spectroscopy.

previously shown[8,9]. For that purpose, we suggest following the copolymerization steps described below (see Step 17).

Linking cysteines selection criteria

For the present cysteine engineering protocol it is necessary to exchange two wild type residues by cysteines. The two introduced cysteines define the linkage geometry in the polyproteins (and thus, also the geometry of force application in a force spectroscopy experiment). The selection of a specific geometry is therefore mainly determined by the question one is interested in. We therefore only add here some basic criteria. In case of a known protein structure, select two residues with side chains that are clearly accessible to solvent. Preferably, these residues should be located on opposite sides of the protein to prevent steric hindrance of polymerization. In addition, it may be advantageous to select residues that are located in basic regions of the protein. It also may be necessary to mutate solvent-accessible wild-type cysteines to alanines or serines, for example, to prevent possible interfering linkage geometry formation.

For unknown protein structures, we recommend starting by changing one residue at the N terminus and the other at the C terminus to cysteines, since most proteins have solvent-accessible termini. In addition, Ellman's reagent (5,5′-dithio-*bis*(2-nitrobenzoic acid)), available from Sigma, can be used to check for solvent-accessible cysteines in advance (see also Troubleshooting).

MATERIALS

REAGENTS

- T7 promoter–based expression vector containing the DNA sequence encoding the protein of interest (see REAGENT SETUP and **Fig. 1**)
- QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene) plus material called for in the kit manual
- XL10 Gold Ultracompetent *Escherichia coli* (Stratagene) plus material called for in the kit manual
- BL21-CodonPlus(DE3)-RIL *E. coli* (Stratagene) plus material called for in the corresponding manual
- Lysis buffer: 50 mM $NaH_2PO_4 \cdot 2H_2O$, 300 mM NaCl, adjust pH to 8.0 with NaOH
- Series of five elution buffer solutions with increasing concentrations of imidazole: lysis buffer plus 20 mM, 70 mM, 100 mM, 200 mM and 500 mM imidazole, respectively
- Sufficient cultured cells to yield approximately 50 mg of the protein of interest
- Reagents for determining protein concentration (e.g., Bradford assay reagents)

EQUIPMENT

- ▲ CRITICAL Use only sterile containers and pipet tips for all procedure steps.
- French press (preferred instrument for cell rupture) or sonicator
- HisTrap HP columns (GE Healthcare)
- Centricon centrifugal filters (Millipore)
- HiTrap columns (GE Healthcare)
- Superose 10/300 GL column (GE Healthcare)
- Equipment to perform single molecule force spectroscopy experiments (atomic force microscope (e.g., Asylum Research MFP 1D); cantilevers (e.g., Olympus BioLevers), etc.)
- Equipment to confirm correct protein folding (such as circular dichroism equipment)

REAGENT SETUP

Preparation of expression vector coding for double-cysteine mutants Obtain coding DNA sequence of protein of interest and include in an *E. coli* cytosolic expression vector (e.g., the Novagen pRSET family of vectors). For convenience, add a sequence coding for a His_6 tag to facilitate protein purification, if not already included in the expression vector. Other expression organisms and systems may be used; however, the reducing cytosol of *E. coli* prevents premature cysteine linkage formation that may induce protein misfolding. Point-mutate pairwise the selected wild-type residues to cysteines using, for example, the QuikChange Multi Site-Directed Mutagenesis Kit (a thermal cycler is necessary) and mutagenesis primers (e.g., from Metabion) designed according to the kit instructions. Replace wild-type DNA codons of the residues of interest with cysteine codons with high expression yield (TGC or TGT for *E. coli*). Confirm mutations by sequencing.

Expression of mutated protein using *E. coli* expression strain Transform mutated expression vectors into an *E. coli* expression strain. We used BL21-CodonPlus(DE3)-RIL for increased protein yield. Perform protein expression following the manufacturer's protocol. As a guideline, a 500-ml cell culture yields approximately 50 mg of the previously described GFP mutants[6]. Cell culture volume may have to be adjusted to account for less efficient protein expression.

PROCEDURE

Protein preparation

1| Pellet sufficient cells to yield approximately 50 mg of your protein of interest and resuspend in an appropriate volume of lysis buffer 1 (e.g., 30 ml for a 500 ml cell culture).

■ PAUSE POINT Store overnight at –20 °C.

2| Thaw and lyse the cell suspension using a French press or sonicator.

3| Centrifuge for 45 min at 30,000*g* to pellet cell fragments.

4| Separate the supernatant cell extract by decanting or pipetting and filter the cell extract using 0.2 μm filters.

5| Purify protein from cell extract by Ni-NTA affinity chromatography using, e.g., HisTrap HP columns (GE Healthcare) or equivalent. Wash column with three column volumes of each of the elution buffer solutions described in reagents (with increasing concentrations of imidazole).

6| Perform analytical gel electrophoresis to identify the elution fractions that contain the expressed protein.

7| Confirm correct protein folding (using, e.g., circular dichroism, UV spectroscopy, etc.) to check for impairment of the native structure after double-Cys mutations.

8| Determine the protein concentration using, e.g., the Bradford assay.

Polymerization

9| Concentrate protein to at least 0.2 mM, using, e.g., Centricon centrifugal filter units (Millipore).
▲ **CRITICAL STEP** Protein concentration is crucial for speeding up polymerization kinetics. Do not add any reducing agents to the protein solution. Polymerization can be performed in the elution buffer (from Step 8) or in phosphate buffered saline (PBS) pH 7.4. If buffer exchange is necessary, use, for convenience, HiTrap desalting columns (GE Healthcare) instead of dialysis.

10| Shake the protein solution to increase amount of dissolved oxygen. This may help to oxidize the cysteines.

11| Incubate for approximately 80 h—if higher temperatures do not impair native protein structure—at 37 °C to speed up reaction kinetics or until the polymers have a suitable length for force spectroscopy experiments (a minimum of tetramers; an average of octamers is convenient).

12| During incubation, analyze the degree of polymerization from time to time using either low-concentration (e.g., 6% (wt/vol)) polyacrylamide gel electrophoresis (with non-reducing SDS buffer) or size-exclusion analytical chromatography (e.g., Superose 10/300 GL column, GE Healthcare) (see ANTICIPATED RESULTS and **Fig. 2**) or simply perform force spectroscopy experiments. The appearance of equidistant sawtooth patterns in the force extension data (see, for example, **Fig. 1**, Step 4) confirms on the single-molecule level the expected cysteine-based polymerization into linear, tightly linked polyprotein chains as previously shown in solution[6] and in a crystallization assay[5]. The ANTICIPATED RESULTS section may serve as a guideline here.

13| During incubation, aggregation effects like sedimentation may be observed that are absent in non-mutated protein samples. For force spectroscopy experiments, thoroughly spin down aggregates at 10,000*g* (at least for 10 min) and use the supernatant protein solution.

14| After having obtained the desired degree of polymerization, store samples at 4 °C to slow down further polymerization. It is important to note that polymerization still proceeds even at 4 °C. To efficiently quench further polymerization, simply dilute the protein solution (e.g., from 0.2 mM down to 0.02 mM). Polymeric sample solution may also be reduced to monomers and again polymerized multiple times by following the steps described in copolymerization (see Step 17 below).

15| Check for absence of interdomain interactions. We suggest following the procedures by Rounsevell *et al.*[7] or comparing melting curves of polymerized and monomeric samples using differential scanning calorimetry. See also the ANTICIPATED RESULTS section.
■ **PAUSE POINT** The properties of the protein of interest determine how long polymerized samples can be stored and used for force spectroscopy experiments. In the case of GFP polyproteins and Ig27 polyproteins, we did not observe any degradation affecting our experiments even after several months of storage at 4 °C.

Single-molecule force spectroscopy

16| Apply ~10 μl of the centrifuged sample solutions (see Step 14) without further treatment to either glass or freshly gold-evaporated glass surfaces and incubate for 1 h. Perform the single-molecule mechanical experiments. See refs. 10,11 for further information on single-molecule force spectroscopy.
▲ **CRITICAL STEP** Plain glass surfaces generally reduce overall adhesion and unspecific interactions but may impair yield of force extension traces for mutants that unfold at higher forces (>400 pN).

Copolymerization

17| Reduce the linking disulfide bonds of the polymeric samples by applying 20 mM DTT (Cleland's reagent) to the sample solutions and incubate for at least 1 h at room temperature (20° C).
▲ **CRITICAL STEP** Choose already polymerized cysteine mutants with comparable polymerization kinetics.

18| Mix the resulting monomeric sample solutions in the desired stoichiometric ratio.

19| Exchange buffer to remove all traces of DTT by using HiTrap desalting columns (GE Healthcare) twice (consecutively).

20| Repeat polymerization procedure (Steps 9–15) and perform force spectroscopy experiments using atomic force microscope and cantilevers.

● **TIMING**

Mutagenesis and protein expression can in principle be done within 1 week. Polymerization time may vary strongly (from overnight to a few weeks) depending on factors such as protein concentration and cysteine reactivity.

? TROUBLESHOOTING

Polymerization problems

Currently, the kinetics of cysteine-based protein polymerization are not fully understood. Protein concentration is a crucial factor for speeding up reaction kinetics. The polymerization method described here requires two reactive cysteines per monomer. Overall reaction speed will be limited by the cysteine with lowest reactivity. The local environment in the folded protein structure will influence cysteine reactivity, and it is not straightforward to predict reactive cysteine mutations. We recommend, therefore, creating a set of cysteine mutants: e.g., all three possible pairs of cysteine mutations for three residues (e.g., n, i and j). This strategy enables identification of nonreactive cysteines. For instance, if the (n,i) mutant polymerizes as confirmed either by Step 12 or by single-molecule force extension data, but (n,j) and (i,j) mutants do not, the cysteine at position j must be responsible for the low reactivity of the latter two mutants. In turn, if mutants (n,i) and (n,j) do polymerize, the mutant (i,j) should polymerize as well.

ANTICIPATED RESULTS

In this section, we describe our results when performing the above protocol with three different double cysteine–mutated GFP molecules and with a double cysteine–mutated Ig27 domain from human cardiac titin as previously described[6]. These results may serve as a guideline also for other proteins; however, adaptation may prove necessary. We chose to mutate the following native residues to cysteines: 3 and 132, 3 and 212, and 132 and 212 in GFP and residues 3 and 88 in Ig27. After having successfully performed mutagenesis, we went on to protein preparation (Steps 1–5). When purifying the His_6-tagged protein mutants, we observed almost complete elution of all of the four different protein mutants when washing the column with the elution buffer containing 200 mM imidazole. We then concentrated the samples to protein concentrations above 0.2 mM and observed polymers suitable for force spectroscopy experiments after ~80 h of incubation for each of the GFP mutants and the Ig27 mutants as well. All of the four investigated samples did show sedimentation effects when concentrated above 0.2 mM, possibly owing to aggregation effects. For all further analysis, we spun down thoroughly all aggregates and used the supernatant protein solution. We did not observe any impairment of supernatant polyprotein integrity.

As an example, we analyzed (as in Step 12) the double-cysteine mutant GFP(3, 212) protein solution (at 0.2 mM) after 4 d of incubation at 37 °C using gel electrophoresis (6% polyacrylamide, non-reducing SDS buffer plus 2 M urea) and using size-exclusion analytical chromatography (Superose 10/300 GL column), monitoring GFP absorption at 396 nm (GFP excitation maximum) and absorption at 280 nm in a pH 7.5 buffer solution containing 50 mM sodium phosphate plus 200 mM sodium chloride buffer. We observed a smooth distribution (**Fig. 2a**,**b**) of polymer lengths containing at least three, and up to 30, fully functional monomers. Using gel electrophoresis and size-exclusion chromatography, one quickly encounters resolution problems when analyzing cysteine-engineered polyproteins of high molecular weight.

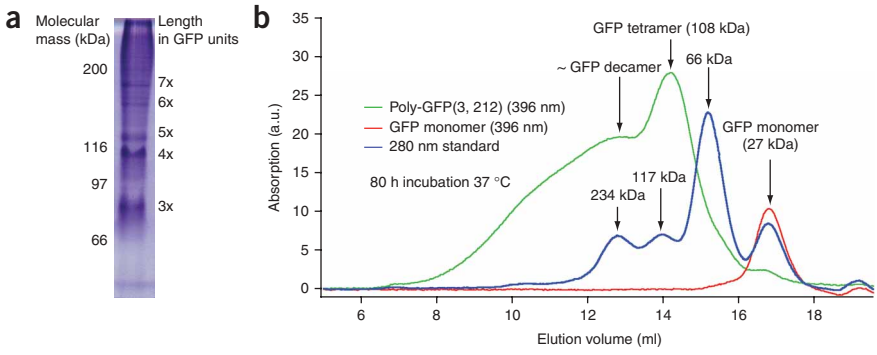

Figure 2 | Dispersion of polymeric lengths in an 0.2-mM GFP(3, 212) polyprotein solution after ~80 h of incubation at 37 °C. (**a**) Analysis using 6% polyacrylamide gel electrophoresis in a non-reducing SDS buffer. (**b**) Analysis by size-exclusion analytical chromatography (Superose 10/300 GL columns) in a pH 7.5 buffer solution containing 50 mM sodium phosphate plus 200 mM sodium chloride buffer. The full flow profile obtained on a poly-GFP(3,212) sample (green solid line) can be attributed to fully functional and, hence, correctly folded GFP polyproteins by a cross-check with the absorption profile of the same sample at 280 nm (data not shown).

We also performed single-molecule force spectroscopy experiments as described in Step 15 with the same GFP(3, 212) sample solution as used for **Figure 2** and observed a very high yield (~20% of all attempts) of force extension traces exhibiting several GFP(3, 212) unfolding events as previously described[6]. In addition, the force extension traces that have been classified as true single-molecule events clearly show a consistent distribution (data not shown) of polymeric length dispersion as in **Figure 2**. We also checked for interdomain interactions using our single-molecule force extension data. A good indication for the absence of domain-domain interaction (besides the methods suggested in Step 15) is also provided by the observation of increasing unfolding forces in sawtooth patterns in force extension traces. This is due to the so-called '*n*-effect'[12]: if the individual modules in the protein chain are statistically independent (that is, interdomain interactions are absent or very weak), then the probability for observing an unfolding event of a single module is higher the more modules are still folded in the stretched chain. This fact then results in the observation of (on average) greater unfolding forces as more domains become unfolded in the stretched chain. We observed the *n*-effect in all of the four investigated polyprotein samples, indicating the absence of domain-domain interactions.

ACKNOWLEDGMENTS This work has been supported by an SFB413 grant of the Deutsche Forschungsgemeinschaft.

AUTHOR CONTRIBUTIONS H.D. developed the protocol, performed and designed research, analyzed the data and wrote the manuscript. M.B. performed size-exclusion chromatography and gel electrophoresis experiments and analyzed the data from these experiments. F.B. performed supporting experiments on GFP polyproteins. M.S., T.B., J.P.J. helped to refine the protocol. M.R. designed research and wrote the manuscript. All authors commented on the manuscript.

COMPETING INTERESTS STATEMENT The authors declare that they have no competing financial interests.

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Bustamante, C., Chemla, Y.R., Forde, N.R. & Izhaky, D. Mechanical processes in biochemistry. *Annu. Rev. Biochem.* **73**, 705–748 (2004).
2. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. & Gaub, H.E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
3. Li, H., Oberhauser, A.F., Fowler, S.B., Clarke, J. & Fernandez, J.M. Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl. Acad. Sci. USA* **97**, 6527–6531 (2000).
4. Carrion-Vazquez, M. *et al.* Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* **74**, 63–91 (2000).
5. Yang, G. *et al.* Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci. USA* **97**, 139–144 (2000).
6. Dietz, H. & Rief, M. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. USA* **103**, 1244–1247 (2006).
7. Rounsevell, R.W., Steward, A. & Clarke, J. Biophysical investigations of engineered polyproteins: implications for force data. *Biophys. J.* **88**, 2022–2029 (2005).
8. Li, L., Huang, H.H., Badilla, C.L. & Fernandez, J.M. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* **345**, 817–826 (2005).
9. Dietz, H. & Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. USA* **101**, 16192–16197 (2004).
10. Fisher, T.E., Marszalek, P.E. & Fernandez, J.M. Stretching single molecules into novel conformations using the atomic force microscope. *Nat. Struct. Biol.* **7**, 719–724 (2000).
11. Clausen-Schaumann, H., Seitz, M., Krautbauer, R. & Gaub, H.E. Force spectroscopy with single bio-molecules. *Curr. Opin. Chem. Biol.* **4**, 524–530 (2000).
12. Zinober, R.C. *et al.* Mechanically unfolding proteins: the effect of unfolding history and the supramolecular scaffold. *Protein Sci.* **11**, 2759–2765 (2002).