Multiplexed Single Molecule Observation and Manipulation of Engineered Biomolecules

Markus A. Jobst

SIGILL. UNIVERSITAT. LUDOVIC. MAXIM.

München 2018

Multiplexed Single Molecule Observation and Manipulation of Engineered Biomolecules

Dissertation

an der

Ludwig-Maximilians Universität München
Fakultät für Physik
Lehrstuhl für angewandte Physik

vorgelegt von

Markus A. Jobst

aus Erding

München, 27.02.2018

Erstgutachter: Prof. Dr. Hermann E. Gaub
Zweitgutachter: Prof. Dr. Ulrich Gerland
Tag der mündlichen Prüfung: 18.04.2018

ZUSAMMENFASSUNG

Molekulare Prozesse in Organismen werden oft von Strukturelementen ermöglicht, die mechanischen Kräften standhalten können. Ein Beispiel hierfür ist das mikrobielle und hierarchisch aufgebaute Proteinsystem des Zellulosoms. Enzyme und die Rezeptor-Liganden Komplexe Cohesin-Dockerin (Coh-Doc) arbeiten hierbei für die effiziente Hydrolyse von pflanzlichen Polysacchariden zusammen. Die Coh-Doc Komplexe können bemerkenswerten Kräften standhalten, um in den extremen Umweltbedingungen, in denen die Mikroorganismen teilweise leben, die Wirtszellen und Enzyme an ihre Substrate binden zu können. Die vorliegende Arbeit untersucht den Einfluss von mechanischer Kraft auf solche Biomoleküle mittels Einzelmolekülmessungen.

Die hohe Symmetrie des Bindeinterfaces des Coh-Doc Typ I Komplexes aus *Clostridium thermocellum* ermöglicht zwei verschiedene Konformationen, die vergleichbare Affinität und Stärke aufweisen. Im Rahmen dieser Arbeit konnte ich beide in den Wildtyp-Molekülen und unter nativen Bedingungen nachweisen.

Eines der stärksten bekannten nicht-kovalenten Rezeptor-Liganden Systeme, Coh-Doc Typ III aus *Ruminococcus flavefaciens* wurde charakterisiert, und die Kernrolle des benachbarten xModuls für die Stabilität des gesamten Komplexes sowie die Rolle der bimodalen Kraftverteilung untersucht.

Solch hohe Kräfte vermindern die Genauigkeit der gemessenen Konturlängeninkremente von Proteinentfaltungen, indem sie Konformationsänderungen der Poly-Ethylenglykol (PEG) Oberflächenanker in wässrigen Puffersystemen verursachen. Mit Elastin-ähnlichen Polypeptiden (ELP) als Anker wurde dieses Problem gelöst: durch die Ähnlichkeit des Peptid-Rückgrates von ELPs mit dem entfaltener Proteine beeinflussen diese die Genauigkeit des Experiments nicht.

Für die Optimierung von Messdurchsatz und Vergleichbarkeit entwickelte ich an einer Mikrofluidik-Plattform zur *in vitro* Proteinsynthese und -immobilisierung. Das Coh-Doc System wurde hierbei als Binde-Molekül für gemultiplexte Messungen integriert. Die dadurch ermöglichte Nutzung einer einzigen AFM Messsonde für die Messung verschiedener Moleküle erlaubt die nötige Kraftpräzision, um molekulare Mechanismen bis auf die Ebene einzelner Aminosäuren aufzuklären.

Des weiteren habe ich den Coh-Doc Komplex in einem rein auf Proteininteraktionen basierten 'Cut and Paste' Assay für den modularen Aufbau molekularer Systeme implementiert. Dieses ermöglicht schnelle Phänotypisierung geometrischer Anordnunungen und die Untersuchung von Wechselwirkung zwischen Enzymen mittels definierter Positionierung auf Einzelmolekülebene.

Um die Kraftantwort komplexer Systeme besser verstehen und letztendlich gestalten zu können, ergänzte ich die Untersuchung von Proteinsystemen mit derer von DNA-Origami Strukturen. Die Ergebnisse der Kraftspektroskopie an DNA wurden mit Computersimulationen verglichen, und trotz des großen Unterschieds ihrer Ladungsraten stimmen beide Methoden gut überein. Dadruch legen sie die Grundlagen für ein besseres Verständnis komplexer molekularer Superstrukturen.

Abstract

Molecular processes in organisms are often enabled by structural elements resilient to mechanical forces. For instance, the microbial and hierarchical cellulosome protein system comprises enzymes and the receptor-ligand complexes Cohesin-Dockerin (Coh-Doc), that act in concert for the efficient hydrolysis of plant polysaccharides. The Coh-Doc complexes can withstand remarkably high forces to keep host cells and enzymes bound to their substrates in the extreme environmental conditions the microorganisms frequently live in. This work focuses on the investigation of mechanical stability of such biomolecules on the single-molecule level.

The highly symmetric binding interface of the Coh-Doc type I complex from *Clostridium thermocellum*, enables two different binding conformations with comparable affinity and similar strength. I was able to show that both conformations exist in the wild-type molecules and are occupied under native conditions.

I further characterized one of the strongest non-covalent protein complexes known, Coh-Doc type III from *Ruminococcus flavefaciens* by elucidating the pivotal role of the adjacent xModule domain for the mechanical stabilization of the whole complex and the role of the bimodal rupture force distribution.

Such large forces impair accuracy of measured contour length increments in unfolding studies by inducing conformational changes in poly-ethylene glycol (PEG) linkers in aqueous buffer systems. This problem was solved by introducing elastin-like polypeptides (ELP) as surface tethers. Having a peptide backbone similar to that of unfolded proteins, ELP linkers do not alter accuracy of the single-molecule force spectroscopy (SMFS) assay.

To provide high throughput and precise comparability, I worked on a microfluidic platform for the *in vitro* protein synthesis and immobilization. The Coh-Doc system was hereby integrated as a binding handle for multiplexed measurements of mechanostability. Employing a single AFM probe to measure multiple different molecules facilitates force precision required to shed light onto molecular mechanisms down to the level of single amino acids.

I also applied the Coh-Doc complex to a purely protein based single-molecule cut and paste assay for the bottom-up assembly of molecular systems for quick phenotyping of spatial arrangements. With this system, interactions in enzymatic synergies can be studied by defined positioning patterns on the single molecule level.

To understand and design force responses of complex systems, I complemented the investigation of protein systems with SMFS studies on DNA Origami structures. The results of SMFS on DNA were compared to a simulation framework. Despite their difference in force loading rates, both methods agree well within their results, enabling better fundamental understanding of complex molecular superstructures.

Contents

Part I

Scientific Context

1
Biomolecules

1.1 Overview

How do the fundamental processes at the foundation of life work? A question of this significance is sought to be answered, when scientists investigate the properties of and the interplay between the multitude of biomolecules at work in living organisms. Biomolecules in general are organic compounds that range across several orders of magnitude of sizes and levels of organization: many molecules like hormones, vitamins, neurotransmitters and metabolites can act on their own, *e.g.,* in metabolic or signaling pathways, but others can also be organized on higher orders to form supramolecular structures. In many cases, only few types of molecules (*i.e.,* monomers) join together to form oligomers by either covalent or non-covalent bonds, thereby constituting larger structures. These processes can be extended to large quantities of monomers, where they form polymers (generally molecular entities similar to oligomers, only with the requirement of a minimum number of subunits).

Almost all functional parts of living systems base on this principle, that only few types of small constituents can combine to larger parts, that not only differ in size and complexity, but surpass their components by enabling all new capabilities: from energy conversion to information storage, from structural functions to enzymatic catalysis, from cell signaling to - ultimately and of course several abstraction layers further - even consciousness. All of these parts, from the smallest molecule or ion as a subunit to multi-domain protein complexes, chromosomes or even cellular networks like a human brain either are or consist of highly versatile biomolecules.

1.2 Polymers for Structural Stability and Energy Storage

Prominent examples of this principle are polymers that primarily serve structural functions across a vast number of plants (*e.g.,* cellulose and lignin), fungi or arthropods (*e.g.,* chitin), or abundant polysaccharides such as starch and glycogen, used by many organisms for energy storage. A simple illustration for how different materials can be generated from the same molecular components is the difference of amylose, amylopectin (both components of starch), and glycogen: despite their chemical identity on the subunit level, all assembled from glucose monomers, solely their differences in branching frequency affect storage density, solubility and availability for enzymatic breakdown. All three are composed of linear chains of glucose, linked by $\alpha(1\rightarrow4)$-glycosidic bonds [1]. While glycogen, the variant found in animals, fungi and bacteria, branching through $\alpha(1\rightarrow6)$-glycosidic bonds in around nine percent of bonds, is a highly soluble and quickly available energy source for example in muscle cells, the other two molecules amylopectin, branching in around five percent and the helically packed amylose with next to no branching are less soluble, allow only slow hydrolysis, but instead higher storage densities in amyloplasts of plant cells [2–4].

1.3 RNA & DNA - Catalysis, Transfer and Storage of Genetic Information

Other immensely important examples for biopolymers include the poly-nucleotides RNA and DNA, made from only four different ribo- or deoxyribo- nucleic acids, respectively, and capable of storing an enormous number of blueprints for RNA and protein machinery. By covalently combining an arbitrary number n of only four different nucleotides in a chain, an enormous amount of information can be stored in such a molecule, growing exponentially with length and quickly reaching uncountable extent $\propto 4^n$ [5, 6]. In nature, this principle not only allows compact storage of information on how to build and regulate expression of functional molecules, but also serves a variety of other purposes like evolutionary selection by providing diversity, epigenetics, or exchange of information between organisms by lateral gene transfer [7–10].

1.4 Proteins - from Structure to Function

For proteins, this principle is taken even further, combining 20+2 proteinogenic amino acids into chains that can fold and combine on different levels of organization to perform distinct functions [11]. Enzymes for example, are highly ordered structures that catalyze conversion of their substrates by lowering activation energies for specific reactions [12]. The molecular arrangement hereby is not only organized by the plain sequence of amino acids (primary structure), but also on higher levels like three dimensional arrangement of nearby residues within the chain forming α-helices or β-sheets (secondary structure), their binding to different parts along the chain to form larger three dimensional tertiary structures ('folding'), or non-covalent binding of such domains to other domains of the same type or others in heterogeneous macromolecules (quaternary structure) [13–15]. These in turn, can again be responsible for building polymers from their subunits to gain functional entities, or breaking them down, and thereby closing the circle.

Such functional macromolecules display a high degree of complexity and are frequently adapted from nature by science and bioengineering to perform similar tasks orchestrated by the subtle and elegant super-organization of uncountable subunits into large and complicated systems. Those approaches include screening for functional molecules in living organisms, but also and more and more often purely synthetic approaches, that design functionality from scratch and invent completely new proteins [16–19]. As these kinds of endeavours prove to be quite difficult due to the complexity, every progress in understanding fundamental principles of how such systems work can potentially also signify progress in promoting those approaches. One way to investigate such mechanisms is the field of single-molecule techniques combined with methods of molecular cloning to pinpoint and understand effects and details of their constituents [20–22].

2

Experimental Strategies

2.1 From Gene Synthesis to the Dynamic Force Spectrum

To answer a scientific question and perform the ensuing experimental design, frequently the acquisition and modification of novel proteins is required. These modifications can for example be the addition of functional peptide tags for immobilization, or whole fusion domains as expression helpers, fingerprint domains or fluorophores for detection, as well as site-directed mutagenesis of the target domain to probe for effects and relevance of single amino acids on specific functions or stability. In force spectroscopy assays, addition and exchange of binding tags and mutations to the domains of interest are frequently performed to adapt to experimental requirements. This is fostered by countless helpful advances to molecular biology that emerged over the last decades of scientific discovery. From *in silico* design, the planning and construction of genes, to advances in cloning techniques and improvements of enzyme capabilities, to optimized laboratory strains of *Escherichia coli* (*E. coli*) for speciality applications, there are many tools that can aid production, extraction, and isolation of biomolecular samples suitable for investigation. For example, it is possible to combine two protein domains of completely different organisms and each with their own function into one construct, covalently linked *via* their peptide backbone and have them recombinantly expressed for use in *in vivo* or *in vitro* studies, to address a broad variety of questions of different scientific background. The following paragraphs focus on the principal methods applied within the course of this thesis for the measurement of recombinantly produced protein receptor-ligand pairs in atomic force microscope-based single-molecule force spectroscopy.

2.1.1 Gene Synthesis and Molecular Cloning

Once a target molecule for analysis in single-molecule force spectroscopy is identified, there are several ways towards the implementation of an experimental design. The DNA coding for the desired proteins can either be amplified from genomic, cDNA or other vector libraries, or freshly synthesized from scratch by solid-phase synthesis [23–25]. One of the most commonly applied methods for synthesis is the four-step phosphoramidite chemistry with repeated cycles of acid-catalyzed deprotection of oligo-nucleotide ends, base addition (condensation), optional capping of unreacted 5′-hydroxyl groups, and finally oxidation of the phosphite triester group to phosphate [26, 27]. With these kinds of reaction strategies, oligomers with tens of base pairs of length can be synthesized. The resulting oligomers can be used as substrates for the assembly of larger fragments (mostly the actual 'gene synthesis') by enzymatic ligation of fragments at their complementary overlaps [28] or by polymerase cycling assembly, where overlapping oligos are amplified and extended into double-stranded genes [29]. By now, these methods have reached remarkably high efficiencies, rendering commercial synthesis of sequences up to the kilo-base pair range a viable option for laboratories, and thereby enabling versatile and flexible design of DNA and protein constructs.

Genomic, plasmidic, or even linear starting material with the coding region for the desired protein can be subcloned into suitable expression vectors using polymerase chain reaction (PCR) based methods for amplification and introduction of modifications. Any desired tags, *e.g.*, for affinity purification, periplasmic export, secretion, or for post-translational modifications, or even whole fusion domains can be introduced, combined or exchanged. In case of custom synthesized linear genes with complementary ends, the desired DNA can readily be inserted into a target vector. With other types of template DNA samples to start from, the target genes have to be PCR-amplified first [30–33], to yield sufficient amounts for the assembly and subsequent transformation of cells, and to attach the correct overlaps for vector insertion at their ends.

After successful amplification, the sample is usually treated with a restriction endonuclease (*e.g.*, *Dpn*I) to digest methylated template DNA. This step helps preventing unwanted background of transformed clones with non-altered plasmids at the later stage of plasmid isolation. To increase efficiencies for the following step, especially if large volumes of the PCR product are used, the sample can then be inactivated by heat and PCR purified to dispose of the enzymes, obsolete and possibly reaction inhibiting components such as ions, dNTPs, primers, and remaining template fragments.

The state of the art method to clone amplified fragments into vectors used mostly during the work on this thesis is a one-pot Gibson assembly [34, 35]. It provides capability of scarless cloning (no unwanted cloning artefacts in the sequence), simultaneous insertion of multiple genes, and high consistency and reliability of rapidly constructing designed plasmids for protein expression, as long as PCR amplifications of inserts and backbone comply as desired. The fundamental strategy behind Gibson assembly is the isothermal recombination of double-stranded DNA with specific and unique overlaps at their ends corresponding to those of their anticipated annealing partner. A 5′ exonuclease to reveal single-stranded overlaps for annealing, a polymerase to repair remaining single-stranded sections after annealing of complementary overlaps, and a ligase to finally seal the nicks and covalently join the fragments together are deployed at 50 °C. The only requirement is that the overlapping sequences at the ends have a melting temperature above 50 °C. If multiple fragments to be cloned into a vector prove problematic due to cross-reactivities, formation of hairpins or other issues, efficiencies can be improved by an intermediate overlap-extension PCR prior to Gibson assembly. Flanking primers of the first and last fragment are hereby used to amplify a single insert that includes all sub-fragments, which are polymerized across their corresponding overlaps within a single reaction. Subsequent Gibson assembly then only requires insertion of a single fragment into the vector.

The many other options for subcloning include Golden Gate assembly, which utilizes Type IIS restriction enzymes to cut DNA beyond their recognition sites for scarless cloning and allows for cloning of highly repetitive genes [36, 37], or traditional restriction enzyme ligation cloning with strong sequence constraints and the drawback of cloning scars in between fragments [38, 39] [1].

The assembled plasmidic product can then be used to transform bacteria made competent for uptake of exogenous DNA by chemical induction or electroporation [41–43]. The transformation of cells is a biotechnological application of the principle of horizontal gene transfer: under harsh environmental conditions, bacteria can take up extracellular genetic material to quickly adapt to challenging conditions through acquisition of foreign genes [10, 44, 45]. Artificial induction of competence can be exploited to introduce plasmidic DNA into the cells and thereby enable reliable transformation of cells for replication of plasmidic DNA or expression of genes, among other applications. Transformed cells are grown overnight on agar plates at a

---

[1] An overview of current techniques in molecular cloning is given by Casini *et al.* (2015) [40]

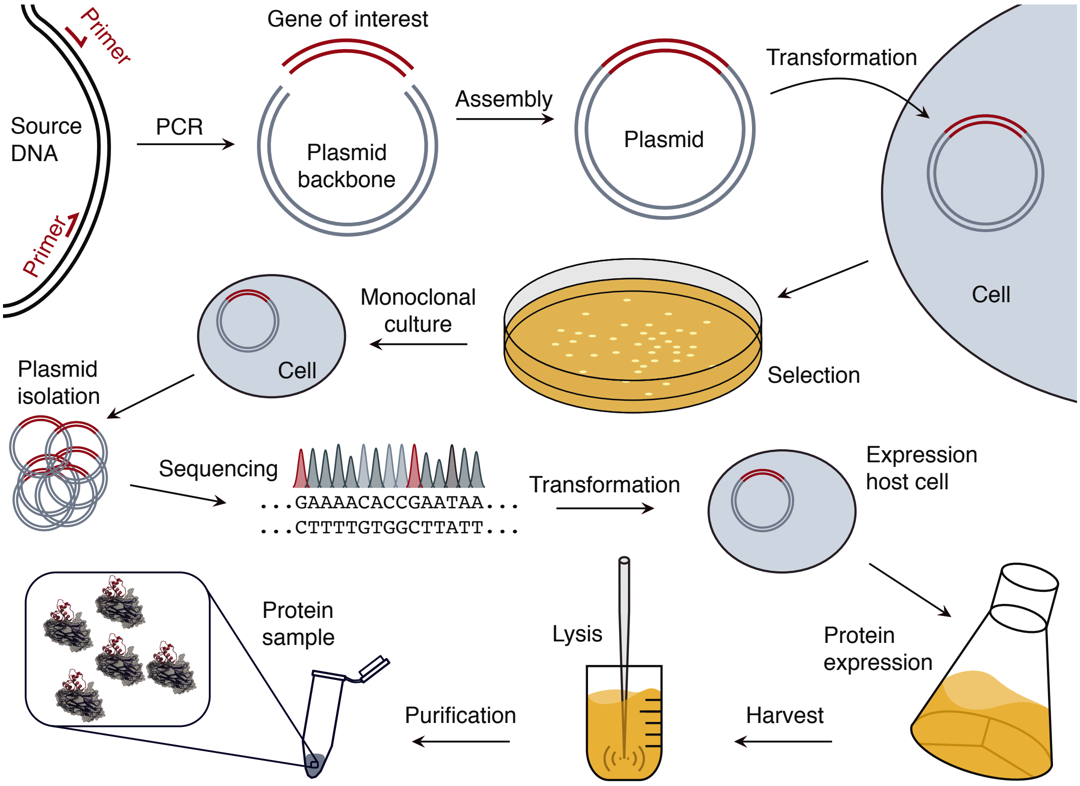


Figure 2.1. The process from genes to a protein sample. Double stranded source DNA (*e.g.*, genomic or synthesized) is amplified *via* polymerase chain reaction (PCR), and assembled into a plasmid with a gene assembly method like Gibson or Golden Gate assembly. Cells are transformed with the resulting plasmid and selected on agar plates after growth in presence of antibiotics. Large amounts of plasmids can be pufiried from monoclonal cultures and confirmed by sequencing. Following transformation of expression hosts, the target proteins of interest are then overexpressed in liquid medium and subsequently harvested and purified.

density that allows for the isolation of single monoclonal colonies ('clones'). A few of them are then cultured in separate vessels with liquid media and their plasmids isolated. Although many DNA polymerases have proofreading capabilities, and therefore only small error rates in their activity, there still can be errors introduced into the target sequence at various stages during the subcloning procedure. To ensure successful cloning, relevant parts of the plasmids can for example be amplified by PCR and then checked for their length by gel-electrophoresis. Promising candidates can then be confirmed by DNA sequencing. The amount of obtained plasmidic material is easily sufficient for subsequent rounds of subcloning, transformation of expression cultures and storage.

2.1.2 Plasmid Components

Important components on plasmids for gene expression in *E. coli* like the pET28a vector include a number of regulatory and other elements to ensure high yield of overexpressed target protein:

An antibiotic resistance gene selects for successfully transformed clones and minimizes background of non-transformed cells, and allows to maintain evolutionary pressure to keep the plasmid in a culture. Both goals are then achieved by growing the cultures in presence of the corresponding antibiotic (*e.g.*, KanR, kanamycin resistance).

An origin of replication (ORI, replicon) ensures the amplification of the plasmid in cells. The replicon ensures replication and determines the copy number of plasmids within a cell, which is regulated by balance of positively and negatively regulating mechanisms of the host cell. The number of plasmids per cell range in the lower tens for 'low' or 'mid copy number' replicons (*e.g.*, pET vectors with pBR322 ORI), and in the mid hundreds for 'high copy number' replicons (*e.g.*, pUC vectors with pMB1 ORI) [46–48].

A lacI repressor is encoded on the plasmid and expressed basally (transcribed with an endogenous polymerase) to bind to the lac operator and thereby prevent the T7 RNA polymerase from mRNA synthesis of the expression cassette. Binding of lactose or derivates to the repressor releases it from the operator, and the T7 RNA polymerase can bind to its promoter and transcribe the corresponding gene. The T7 RNA polymerase is exogenous to derivatives of B strain *E. coli*, which minimizes interference with host cell metabolism. Many strains suitable for T7 promoted gene expression have the λ-phage DE3 gene encoding for the T7 RNA polymerase introduced into their genome ('λDE3 lysogens') [49, 50].

Promoter and terminator regions that allow (*e.g.*, T7) RNA polymerases to bind, transcribe and unbind the coding fragment of the DNA.

A ribosomal binding site (RBS), encoding the prokaryotic mRNA Shine-Dalgarno consensus sequence `AAGAAG` (typically followed by `U` in *E. coli*) upstream of the coding region, to recruit the ribosome to the transcribed mRNA and thereby initiate translation [51, 52].

A start codon (mostly the nucleotide triplet `AUG` on the mRNA level), encoding a methionine signals the start of translation.

A stop codon (`UAG`, `UAA`, or `UGA`) terminates translation of an mRNA sequence by the ribosome [53–56].

After successful translation of mRNA into the corresponding amino acid sequence, the desired protein construct can ideally fold (often with the help of chaperones) into their corresponding functional structure. Post-translational modifications can be necessary for the correct functionality of the protein, like the activation of Asx ligases through proteolytic maturation at low pH [57]. For modifications like N-terminal tagging, it is important to note, that the initial methionine and many small residue amino acids in the subsequent position are cleaved by a methionine peptidase in *E. coli* [58–60].

2.1.3 Gene Expression

A small volume of liquid growth medium with the appropriate antibiotics is inoculated with a suitable expression strain transformed with the plasmid containing the coding region for the desired protein construct ('pre-culture'). In its late exponential growth phase, when cell densities are sufficiently high, the culture is still growing quickly, but the medium is not yet turned acidic, the culture is transferred to a larger volume. A ratio of around 1:100 of pre-culture and the chosen expression medium (again with appropriate antibiotics) then usually represents a good trade-off between density of healthy cells and introduction of acidic reagents, metabolic waste and dead cells. Typical growth media include at least a carbon source, a nitrogen source, a buffer system and a mixture of salts, often complemented by cofactors to enhance

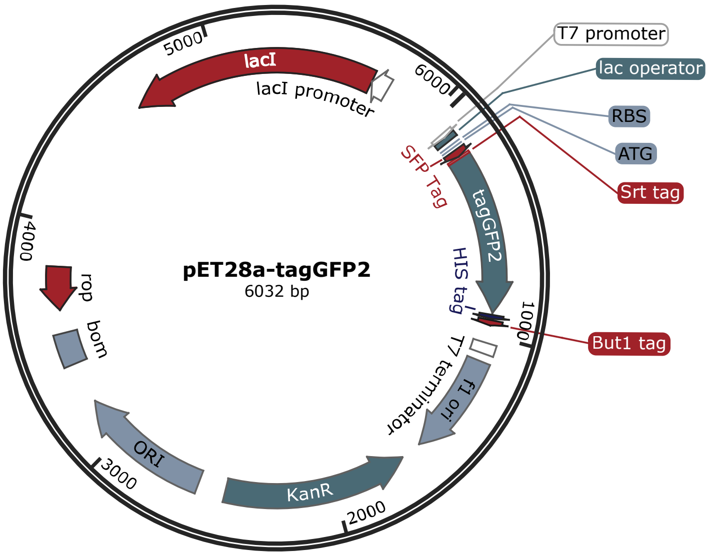


Figure 2.2. pET28a-tagGFP2 vector for subcloning and gene expression. This plasmid includes a gene encoding for a tagGFP2, labelling tags for enzymatic ligations (SFP, Srt, and But1), and a tag for HIS affinity chromatography. Other components described in section 2.1.2 are indicated.

growth. Defined and minimal media consisting of these components are economical and can be required for specific scientific questions, but to maximize over-expression of proteins for isolation, rich media are frequently used. They additionally contain complex formulations from biological origins like algal or yeast extracts, are not exactly determined, and enable significantly higher cell densities and protein yields.

Genes under the control of the T7lac promoter are inhibited from expression in absence of lactose, because the lacI repressor binds its operator preceding the gene. Lactose or derivates like Isopropyl β-D-1-thiogalactopyranoside (IPTG, a lactose derivate inaccessible to the host cell metabolism) release the repressor from the DNA upon binding, although the exact molecular mechanism is not yet completely understood [61]. At high glucose concentrations as the preferred carbon source, expression of lac controlled genes is suppressed, but occurs at low (basal) levels in presence of lactose, because the endogenous RNA polymerase can only bind transiently to its promoter region, and therefore T7 RNA polymerase is only expressed at low levels. A second control mechanism regulates glucose dependence of expression in this case: The catabolite activator protein (CAP) promotes tight binding of the polymerase by binding to the CAP binding site upstream of the promoter on the DNA, mediated by cyclic adenosine monophosphate (cAMP). This assist of tight RNA polymerase binding ultimately allows for high expression levels [62]. Since cAMP levels are regulated by the rate of glucose transport across the cell wall through adenylate cyclase activity, this catabolite repression mechanism inhibits expression of T7lac controlled genes also in presence of lactose for the most part. Apart from that, lactose uptake rates are also inhibited by glucose transport through direct inhibition of lactose permease, resulting in low intracellular lactose levels during glucose metabolism [63].

This mechanism can be utilized to control recombinant protein expression with two conceptually different approaches: One relies on growing the cells to an optimum density (which is measured by optical density at 600 nm, OD_{600}) and manually inducing expression by addition of IPTG to the medium (*e.g.*, LB medium, 'lysogeny broth'). The OD_{600} has to be monitored constantly to prevent missing the right moment during the phase of maximum growth rate. Since IPTG is not a substrate for the β-galactosidase of the lactose metabolism, its concentration remains constant after manual induction, yielding stable expression rates. The other approach involves automatic induction in media with both glucose and lactose as carbon sources (*e.g.*, ZYM-5052), triggered by the depletion of glucose as the preferred carbon source. Once the lactose metabolism has replaced glucose metabolism, T7lac controlled genes are highly overexpressed [64]. This strategy reduces workload and thereby allows for quick and efficient screening. After expression, cells can be harvested by centrifugation, and after decanting the supernatant, the cell pellet can be frozen and stored at −80 °C until purification.

2.1.4 Protein Purification

To obtain the desired overexpressed protein constructs from the harvested cells, they are resuspended either for chemical (often detergent-based) or physical lysis by methods such as freeze-thaw cycling, ultrasonication or liquid homogenization, optionally assisted by enzymatic digestion of the polysaccharide cell wall components through lysozymes. After centrifugation and filtering to get rid of cell wall debris and other large non-soluble components of the lysate, the sample can be further purified by chromatography methods based on various principles like affinity, size exclusion, ion exchange, or free-flow-electrophoresis. Affinity chromatography techniques frequently rely on short peptide tags in the amino acid backbone of the protein. For example, a poly-histidine tag can complex nickel ions chelated in a matrix of nitrilotriacetic acid, which is incorporated into a resin, *e.g.*, made of agarose. After loading the sample onto a column made of this substrate, the bound fraction can be rinsed with buffer to flush out unwanted molecules, and subsequently eluted with another buffer containing imidazole or histidine to compete for binding to the nickel ions. Alternatively, a change in pH or salt concentration that modifies binding strength, can also be used for elution. Because proteinogenic amino acids with an aromatic group adsorb light at 280 nm wavelength, a UV sensor in the flow after the column can be used to indicate the volume containing usable concentrations of target protein. Tyrosines or tryptophanes are required for this method, because phenylalanines adsorb light of this wavelength to a much lesser extent [65, 66]. In case none of these amino acids are present, adsorption of the peptide bonds at 205 nm can alternatively be utilized for determination of concentrations [67]. This inline analysis allows to choose the most promising fractions for purity analysis in gel-electrophoresis. After this, functionality should be tested to ensure recovery and proper folding of the correct construct. Afterwards, the buffer of the samples can be exchanged and proteins concentrated either for storage at −80 °C, or directly applied in an assay.

2.1.5 Bacterial Strains

For gene expression in *E. coli*, typically dedicated expression strains are transformed with the plasmids, although some strains are capable of both, at least moderate yield reproduction of plasmidic DNA for molecular cloning and gene expression for obtaining the proteins encoded on the plasmids. A typical strain for cloning and subcloning purposes would be DH5α, a descendant of the K-12 laboratory strain. It is optimized for transformation efficiency, for plasmid yield and quality by suppression of heterologous recombination, for insert stability by removal of some endonucleases, and capable of blue/white screening. For protein expression on contrary, BL21(DE3), a derivative of the B strain would be a typical strain to use because of its genomic and IPTG inducible expression of T7 RNA polymerase to initiate expression of the target proteins and its deficiency of some proteases to prevent digestion of recombinantly overexpressed target proteins [49, 68] [2]. Other strains have also been optimized for quick growth (NEB Turbo, Mach1 T1R), tunable expression by titration for membrane-bound or insoluble proteins (Lemo21(DE3)), chaperone co-expression, or periplasmic expression and are readily available from various suppliers [70]. Other potentially helpful properties for gene expression are the usage of rare codons by co-expression of their corresponding tRNA (Rosetta, CodonPlus) [71, 72], disulfide formation in the cytoplasm (FA113, Origami or SHuffle strains) [73, 74], expression of toxic proteins (BL21(DE3)NH, NEB Turbo, Lemo21(DE3)) [75, 76], or even prokaryotic protein glycosylation in bacteria [77, 78].

---

[2] A head-to-head comparison of *E. coli* B and K-12 genomes can be found in Studier *et al.* (2009) [69]

2.1.6 Other Systems for Protein Synthesis

Apart from the relatively simple protein expression in bacterial cells, there are several eukaryotic expression systems that can be utilized, if the drawbacks of bacterial systems prevent successful production of functional proteins. These other systems can be mammalian, insect, yeast, algal, or from other plant cells, each with their special requirements in gene structure, culture conditions and with benefits like possible modifications or drawbacks like time scale or cost efficiency. The benefits also include eukaryotic post-processing, target protein solubility or toxicity, special folds or protein sizes like antibodies or complex proteins, or requirement of special control of the experimental environment.

Another option is the expression in cell-free systems (in vitro translation and transcription, IVTT), which are either crude extracts harvested from whole cells, or are separately purified and reconstituted to precise concentrations in 'pure' systems. Cell-free expression systems are also available for all types of prokaryotic and eukaryotic protein synthesis [79–81]. Among other potential speciality applications like amino acid replacement, tRNA replacement or high-throughput assays, these systems have advantages in preparation time, manipulation of reaction conditions up to extremes, availability of proteins with high cytotoxicity or proteins highly prone to aggregation, degradation or misfolding, but are costly and typically have comparatively low yields.

Part of the work on this thesis was the development of a microfluidic platform for the *in vitro* transcription and translation (IVTT) of target proteins for subsequent investigation in a multiplexed SMFS assay (see section 5.4). Cell-free expression of proteins without the requirement of optimization and manual purification enables quick phenotyping and comparison of mechanical properties. The technology developed here enabled characterization and quantitative comparison of a range of homologous molecules in a similar assay with high precision in a later study [82].

2.2 Single-Molecular Force Measurements

2.2.1 Instrumentation

Single-Molecule Force Spectroscopy (SMFS) is a nanometer-scale technique that enables to gain insights into molecular mechanisms. Several different types of instrumentation have emerged over the past few decades, and they are able to determine unfolding and unbinding force distributions of individual molecular structures.

Optical tweezers (OT) for example, can routinely measure forces acting on single beads trapped in laser foci and moving the beads apart from each other, or away from a cover slide. Molecules are either coupled to two beads or to a bead and a glass substrate [83, 84]. Magnetic tweezers (MT) make use of paramagnetic beads on an glass slide to exert magnetic forces by precisely positioning permanent magnets above them. A unique benefit of these instruments is an extension that enables application of torque in addition to translational forces to molecules and measure their responses (magnetic torque tweezers, MTT). A small extra magnet is placed at the side of the main magnet to generate an asymmetric magnetic field. By this method, for example the twisting and bending of double-stranded DNA can be explored [85–87]. Very recent developments include acoustic force spectroscopy (AFS), and centrifugal force microscopes (CFM) [88–91].

Initially developed from the scanning electron microscope for imaging surfaces with atomic resolution, the atomic force microscope (AFM) has been applied to a broad range of imaging and force spectroscopy measurements [92–95]. Since early SMFS experiments like the measurement of individual receptor-ligand pairs, characterization of the elastic response of DNA, or reversible unfolding of Titin Immunoglobulin domains, many developments improved this method [96–98]. Further uses of the AFM based force spectroscopy not only include measurement-driven assays, but also applications focussing on the assembly and manipulation of functional surfaces and nano-scale positioning of enzymes or fluorescent reporters, *e.g.*, single-molecule cut and paste. Hereby, the recorded force data merely serves as a control of delivery and estimate of number of molecules, rather than for determining molecular parameters [99, 100].

Still, new methods for measuring forces applied to molecules are currently emerging, where single molecules themselves act as biosensor force probes and are read out by means of fluorescence. Förster resonance energy transfer (FRET) efficiency or lifetime measurements are utilized to determine the force response of molecules immobilized and probed by biosensors like DNA Origami or even *in vivo* by engineered vinculin derivatives [101, 102].

2.2.2 Fingerprint Domains

To ensure specific pulling on single molecules in protein-SMFS, often so called fingerprint domains are fused to the domains of interest. They serve for correct identification of structures through their individual force response, and rule out artefacts like unspecifically adsorbed molecules in unknown geometries or contaminations of any kind. In the case of poly-protein pulling, these can for example be multiple titin immunoglobulin domains, which have the molecule of interest in between them, and are simultaneously used as unspecific pulling handles by adsorbing to an unpassivated cantilever tip. In specific receptor-ligand SFMS, on the contrary, the fingerprint domains can be any previously characterized domain. Especially the unfolding forces and contour length increments should be known. Other than that, quick refolding capability in the measurement buffer is often necessary or at least helpful. The fingerprint domain can then be covalently fused to the receptor or ligand domain, either on the gene level for expression as a fusion domain, linked post-translationally through specific enzymes, such as SFP, sortase or OaAEP1, or by other means, such as labelling of non-natural amino acids [103–107]. Co-expression as fusion proteins has the advantage, that the domains can also serve multiple purposes as expression, solubility or purification helpers, whereas post-translational modification fosters versatility and quick adaptation of experimental design to new findings and requirements.

2.3 AFM Measurement Protocols

A typical measurement of a single-molecule interaction in a receptor-ligand AFM-SMFS experiment in constant speed mode is schematically shown in figure 2.3. A laser is reflected off the back side of a micrometer sized cantilever and indicates differences in force by changing its position on a quadrant photo diode. The tip of the cantilever features an extrusion on the order of a micrometer, which culminates in an apex in the nanometer range, directed towards the sample surface. On both, tip of the extrusion and sample surface, the complementary molecules of interest are immobilized chemically. In standby, the laser spot is adjusted such that the differential voltage between the segments of the diode is zero. Any acting force on the lever arm, *e.g.*, indentation into a sample on the surface, or a molecule tethered from the surface to the tip of the lever, can be recorded as a change in differential signal of the photo diode segments. To initiate binding of the molecules immobilized on the glass surface and the tip of the cantilever, they are brought into contact under controlled conditions up to a certain force (typically around 100 pN), such that the samples are not compromised by physical means. Upon retraction of the cantilever, the bound molecules are stretched and thereby loaded with force. Depending on the mechanical hierarchy, sub-domains and receptor-ligand binding give in, unfold and unbind sequentially. After each unfolding event, a certain length is added to the free contour length L of the polymer backbone, such that the force acting on the molecules drops recognizably until the retraction distance catches up. At the end of one cycle, once all molecules are detached, a new spot on the sample surface is addressed (some nanometers away from the initial position in x-y direction) to probe new molecules immobilized on the surface. A broad range of measurement protocols is commonly used with the AFM in SFMS mode. Each has their benefits and drawbacks and has to be chosen according to the nature of the molecule under investigation and the scientific question to be answered.

Constant Speed or distance ramp mode is the easiest to implement on the instrument side, but the most complicated to analyse: On the one hand, commercial piezoelectric positioning systems are often readily deployable in these kind of assays, since the positioners already either possess distance sensors and their controllers the respective methods to run closed-loop positioning, or have other means to overcome positioning hysteresis to an acceptable precision. On the other hand, analysis algorithms need to take into account the non-linear entropic response of the polymer under investigation, complicating extraction of usable parameters from the raw data and underlying theoretical foundations did not include the non-linearly increasing forces on the molecules resulting from constant speed measurement mode.

In Force Clamp mode, after reaching a pre-set trigger value in constant speed mode, the measured deflection on the cantilever is fed back into the positioning loop of the z-axis piezoelectric actuator. Thereby, a previously set force can be applied to the molecule throughout one measurement. The cantilever has to be calibrated in advance if accurate force values are to be probed. Depending on the corner frequency of the whole actuated part of the instrument, it takes a certain time in the milliseconds range to readjust the position after a domain rupture, until the set force is reached again. The longer this time is, the less accurate the measured lifetimes of domains can get. The instrument can be tuned primarily by the inertia of the instrument and the integral part of the PID loop. The latter is limited when oscillations due to large error values or due to noise on the deflection signal have to be suppressed. For this purpose, a new type of precise piezoelectric positioners can be employed, that do not

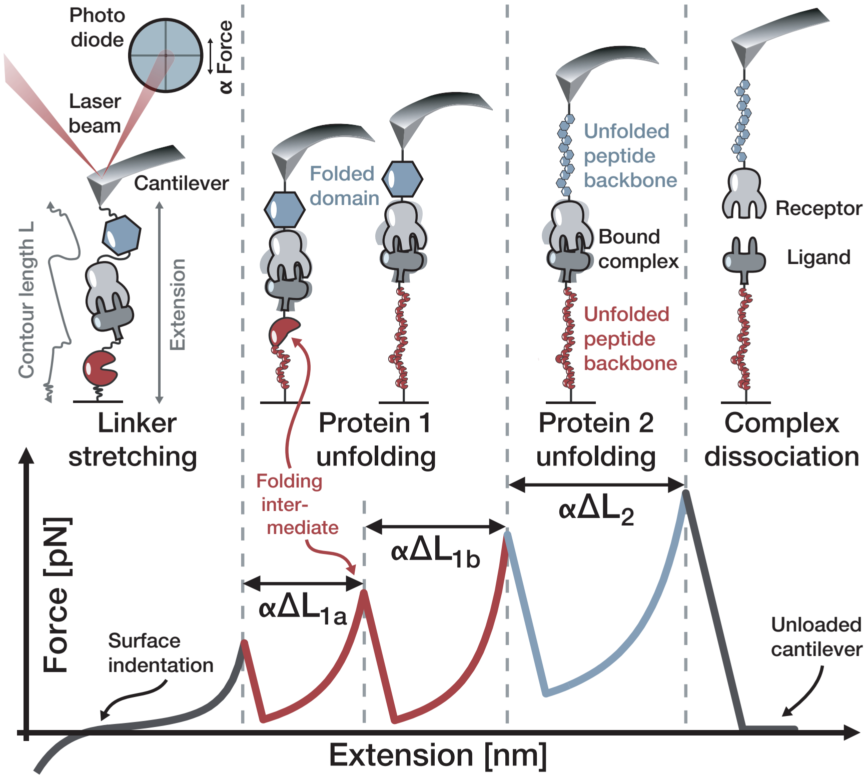


Figure 2.3. Schematic view of an AFM-SMFS experiment in constant speed mode. A focused laser beam is reflected from the backside of a micrometer sized cantilever to indicate its bending on a quadrant photo diode. Movement of the cantilever with nanometer precision by piezoelectric actuators restricts conformational freedom of the molecules tethered to the nanometer sized tip of the cantilever and thereby force is acted on the measurement needle. During retraction, the lever and tethered molecules get increasingly loaded with force, and protein domains give in and unfold in sequence according to their mechanical strengths until finally, the receptor-ligand complex is dissociated. The force-extension behaviour of the polymeric linkers is governed by entropic forces and display the typical non-linear curvature until rupture of a domain. In contrast, the linear decrease in force after each rupture is due to the hookean spring behaviour of the cantilever (adapted from [108] with permission).

rely on closed-loop positioning without introducing hysteresis on the positioning. Analysis in this mode focuses on average lifetimes of the domains of interest at given forces.

Force Ramp mode is similar to force clamp, with the difference, that the force set point is linearly increased over time. Typically, a starting value is chosen, such that the loop does not start to trigger in the thermal noise of the signal, and sufficient that a specific attachment of a molecule can be expected. From there on, the set value is ramped at a constant rate over time. This change of the set value is identical with the force loading rate of the molecule, which makes analysis much more straightforward compared to constant speed modes.

From the instrument perspective, the three modes can also be thought of in reverse order: force clamp simply is a force ramp with a increase rate of zero piconewton per second; and constant speed is a force clamp with an infinite trigger value, that never gets activated. There are even more elaborate modes, for example mapping the attractive interaction of few molecules on the cantilever with molecules on cell surfaces. This mode records a topological image of the surface in the conventional AFM imaging way, but additionally measures specific adhesion interactions and maps them synchronously to the places where they occur. This allows for example for the estimation of lateral diffusion coefficients of receptors within membranes:

THE IDEA OF MAPPING MOLECULAR RECOGNITION[3] by simultaneously measuring surface topography and force-extension data ('force volume mapping' or 'affinity imaging') was introduced early [109, 110], and refined to remarkable temporal and spatial resolution. While these molecular recognition imaging techniques turned out to be a valuable tool for detecting and locating specific binding sites on surfaces, their development into dynamic recognition force imaging [111–113] greatly increased temporal and spatial resolution, while still yielding information about surface elasticity and adhesion, as well as identifying biomolecules at the same time.

THESE MULTIPARAMETRIC IMAGING MODES[4] can simultaneously detect physical properties of the surface and forces exerted on specific biomolecular binding sites. The AFM cantilever oscillates with amplitudes around 100 nm at sub- or low kHz frequencies to measure force-distance data, and simultaneously records image topography and other surface properties at sub- or low Hz line scanning frequencies. The recorded force and topography data is collected orders of magnitude faster compared to force volume mapping methods, yielding imaging speeds comparable to conventional AFM imaging methods [114, 115]. Another benefit of this method is that a large range of loading rates for receptor-ligand dissociation events can be probed in a single experiment, due to the largely varying cantilever tip velocities. Recently, this method was applied to gain nanometer-scale resolution imaging data of a G protein-coupled receptor (PAR1) in proteoliposomes while characterizing their ligand-binding energy landscape [116] from loading rates ranging from 10^3 to 10^6 $\mathrm{pN\,s^{-1}}$, already two orders of magnitude higher than conventional force-distance based SMFS. Another recent study demonstrates the ability of this technique to distinguish two different binding events on opposite sides of engineered PAR1 by their unbinding force, and thereby determine their orientation within the lipid bilayer [117].

LATERAL FORCE SENSORS.[5] A slightly different approach developed a T-shaped cantilever [118, 119] to drive it at its flexural resonance frequency (around 9 kHz) and record force data from cantilever torsion, resulting in a lateral laser deflection signal that was acquired while imaging the sample in conventional tapping mode. Due to the cantilever's high torsional resonance (around 115 kHz), unbinding dynamics could be measured at the microsecond time scale and at extraordinarily high loading rates of up to nearly 10^9 $\mathrm{pN\,s^{-1}}$ [119], about four orders of magnitude faster than conventional SMFS. Force curves and therefore unbinding events and their corresponding force values could be mapped with high spatial and temporal resolution, while providing

---

[3] This paragraph was published by Ott *et al.* (2017) [108] in the Journal of Structural Biology and adapted with permission from Elsevier Inc. Copyright ©2016 Elsevier Inc.

[4] see Footnote 3

[5] see Footnote 3

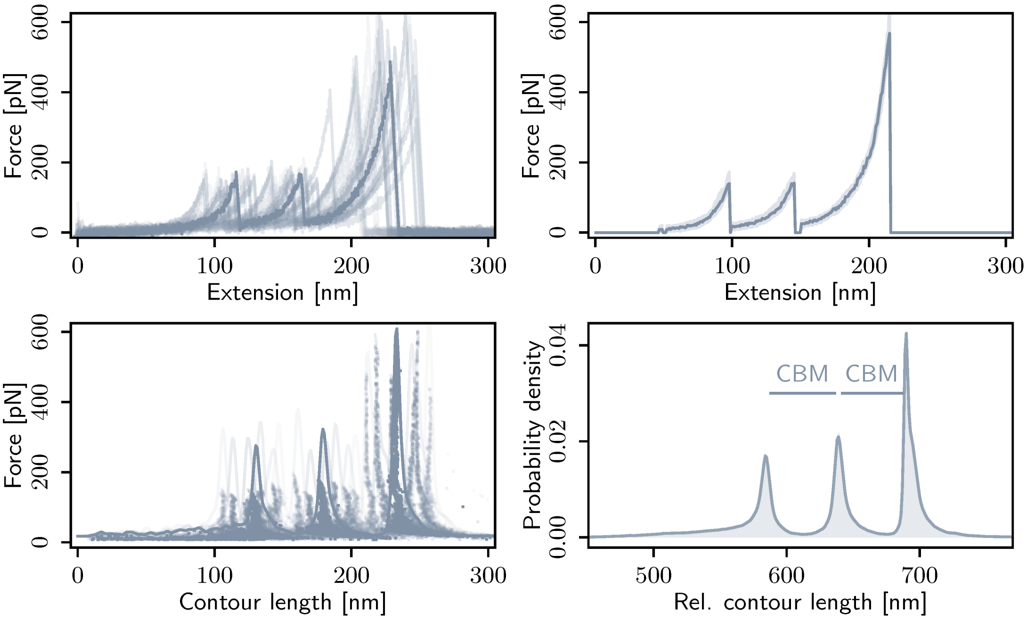


Figure 2.4. Data reduction of constant speed SMFS data. Top left: State-of-the art experiments can yield large numbers of individual unfolding curves, frequently with numbers among several hundreds. All these measurements get sorted by their types of unfolding events (typically, order of contour length increments after ruptures) and then superimposed by different methods. **Top right:** One of these methods include several rounds of cross-correlation in contour length space and offsetting the original force-distance data on the distance axis according to the maximum correlation. Resulting data can be superimposed, and sliced into nanometer bins to find the most probable values and full-width-half-maxima for each extension bin. **Bottom left:** Another strategy is to transform individual curves into contour length space and calculate the probability densities of their energy barrier positions. **Bottom right:** Global contour length transformation from superposition of all individual transformations, distance shift after cross-correlation and estimation of the probability density by a KDE with a silverman estimator. The horizontal lines indicate the gain on free contour length by unfolding two CBM domains subsequently.

AFM images that were simultaneously recorded as surface topography. Mechanical elasticity properties of the substrate were also detected in the phase signal.

2.4 Data Reduction for Analysis

To extract meaningful parameters from the data to understand the investigated molecular mechanisms, sufficient statistics are required, as not only noise is compromising measurements, but the fundamental processes at work are inherently stochastic. On top of this, like in many single-molecule techniques, typically many of the taken data in AFM-SMFS are not usable due to unspecific pick-up of contaminations, no successful target molecule binding, or multiple molecules measured simultaneously. These data are in general not straightforward to interpret and often need to be excluded from analysis. To handle large datasets typically on the order of hundreds of thousands of force-extension curves, sophisticated strategies combined with powerful sorting algorithms are required.

After converting raw voltages into distances and forces with the respective calibration values of the instrument, further corrections are applied, like the compensation of hookean lever bending due to the applied force, or offsets on the force and distance axes to adjust for zero loading and position of surface contact, thereby obtain absolute numbers [120]. From this point on, a number of strategies were developed to classify data by its characteristic patterns and extract unfolding increments after and rupture forces of events in each curve.

Initial steps in sorting the vast number of data usually include filtering out empty traces, that show no adhesive interaction, multiple interactions, that have far too many peaks than expected and those, that show interactions far outside the range of combined tether and unfolded backbone linker lengths. Thresholds should of course be applied carefully to not exclude meaningful data and thereby introduce artificial bias into the measurement. Remaining curves can then either be fitted with polymer elasticity models or transformed into contour length space with the same models to identify increments in free contour lengths ΔL_c^{ij} from peak i to peak j for all subsequent i, j in each curve. In contour length space, the spacing of energy barriers is directly observable [121]. Once all pathways in a molecular system are identified, the corresponding peaks for each individual event type can be plotted against their respective loading rates and the kinetic models can be fitted to either a single distribution or over a whole dynamic force spectrum (see section 3.2).

Data transformed into contour length space can be cross-correlated with all curves in a single data set and then shifted by its maximum value along the x axis. All curves can then be superimposed and processed with a kernel density estimator, to assemble a global probability density for the whole data set (see figure 2.4). This allows to quickly identify most prominent features in contour length increments, or to observe effects of cross-linking or insert studies at a glance [122].

If on the contrary determination of elasticity parameters are of prime interest, a different approach is often more feasible: to gain representative 'master curves' of single unfolding pathways, the maximum value of cross-correlation in contour length space can also be used to shift the original data in force-distance space along the x-axis. By iterative correlation with the assembled global probability density and subtraction of most probable shift in the final assembly, bias introduced by the choice of the initial curves to process can be reduced, and representative absolute values can be obtained. Application of shift and correlation thresholds allows to easily filter out less probable populations and noisy data. Binning of superimposed data in force-distance space along the x-axis into nanometer-sized slices allows calculation of a kernel density estimate (KDE) along the y-axis for each slice. The maxima of the KDEs and their full-width-half-maxima in all slices assemble the master curve. This algorithm does not necessarily reproduce absolute rupture forces to the highest accuracy, but the resulting single trace resembles the elastic behaviour and the contour length increments in between peaks of the most probable pathway well and allows for investigation of the elasticity on a single representative force-extension curve [123].

3

Molecular Response to External Forces

3.1 Polymer Elasticity

The force extension-behaviour of linear polymers in aqueous solutions does in general not follow hookean spring characteristics, but rather has predominantly entropy driven properties. The subunits of a polymer try to explore their accessible space, given by the end-to-end distance of the polymer, and the flexibility of the chain. By pulling apart the ends, the number of accessible conformations for the chain are restricted, and thereby, an entropic force acts upon the ends, trying to increase possible states and therefore configurational entropy [124]. An interesting property of such entropic springs is that forces rise with higher temperatures. As the entropic penalty for stretching the molecule gets larger with temperature according to the relation for the Gibbs free energy $\Delta G = \Delta H - T\Delta S$. A number of theoretical frameworks were developed to describe this entropically governed force-distance dependence adequately across large force ranges. Among those are the freely jointed chain (FJC) or Gaussian chain model, the worm-like chain (WLC) model, the freely rotating chain (FRC) model or combined approaches with transition ranges, such as the model described by Livadaru, Netz, and Kreuzer (2003) [125]. It is important to note, that elasticity parameters like the Kuhn length l_k, or the persistence length l_p do not resemble distances in the molecular structure (*e.g.,* the atomic distance of alpha carbons), since they plainly represent statistical monomer sizes with no physical equivalent.

3.1.1 FJC

The FJC is the most general model, and it treats the polymer as a flexible linear chain of contour length l_c without restrictions on bond angles or any other long range interactions. In this picture, the elastic response of the polymer to external force is purely entropically driven. The orientation of the polymer's discrete segments is described by a statistical three dimensional random walk with independent rigid segments of length l_k ('Kuhn length', parameter for the local stiffness), and therefore, the orientation of the segments as uncorrelated, and they can rotate freely at their interconnections [124, 126–129]. The correlation between tangent vectors $\vec{t}_i$ and $\vec{t}_j$ of any segments i and j is zero: $\langle \vec{t}_i \cdot \vec{t}_j \rangle = \langle \vec{t}_i \rangle \cdot \langle \vec{t}_j \rangle = 0$. *Via* the partition function of the system, the free energy of the whole chain can be found, yielding an expression for the end-to-end distance for a given force. through differentiation:

$$\langle x \rangle = l_c \left(\coth \frac{F l_k}{k_\mathrm{B} T} - \frac{k_\mathrm{B} T}{F l_k} \right) \tag{3.1}$$

$$= l_c \mathcal{L} \left(\frac{F l_k}{k_\mathrm{B} T} \right), \tag{3.2}$$

with the Langevin function $\mathcal{L}(a) = \coth a - \frac{1}{a}$. Equation 3.1 can also be solved for the force F by making use of the inverse Langevin function $\mathcal{L}^{-1}$:

$$F = \frac{k_\mathrm{B} T}{l_k} \mathcal{L}^{-1} \left(\frac{\langle x \rangle}{l_c} \right). \tag{3.3}$$

A recent approximation of the inverse Langevin function with a maximum relative error of less than 0.18 % is given by Petrosyan (2017) [130]:

$$\mathcal{L}^{-1}(a) \simeq 3a + \frac{a^2}{5} \sin \left(\frac{7a}{2} \right) + \frac{a^3}{1-a}. \tag{3.4}$$

3.1.2 WLC

Kratky and Porod (1949) have developed a model that resembles more of a directional continuity by focussing on the limit of a much larger number of elements n compared to their relative extension $n \ll l_n / l_c$ [131]. This has been called 'worm-like chain' and presents an intermediate solution between the description of a rigid rod and a flexible coil. The local stiffness is given by the persistence length l_p, which does not represent a physical distance in the molecular structure, but is rather a descriptive elasticity parameter in the model. The addition of long-range flexibility lets the trajectory vary continuously, resulting in a smoother chain compared to the FJC [124].

The WLC model treats the polymer as a rod of contour length l_c with flexible subunits. The persistence length l_p resembles the distance at which the directional correlation along the rod orientation drops to $\frac{1}{\mathrm{e}}$. In the discrete case, the directional correlation of tangent vectors $\vec{t}$ at position s_0 with distance Δs along the polymer is given by the exponential power law

$$\langle \vec{t_i} \cdot \vec{t_j} \rangle = \exp\left(-\frac{|i-j|}{l_p} \right). \tag{3.5}$$

Molecules with higher persistence lengths, such as double-stranded DNA (around 35 nm are therefore stiffer than those with short persistence lengths, such as polyethylene glycol (sub-nanometer) [132]. Force acting on the ends of the polymer then reduces the conformational space for its subunits and induces an entropic restoring force, dependent on the extension of the molecule. Since no analytical solution to this framework is known, Marko and Siggia came up with the most noted approximate interpolation formula for the WLC force-extension relation [133, 134]:

$$F_\mathrm{WLC}(x) = \frac{k_\mathrm{B} T}{l_p} \left(\frac{1}{4(1 - \frac{x}{l_c})^2} + \frac{x}{l_c} - \frac{1}{4} \right) \tag{3.6}$$

Recently, more accurate approximations for the force-extension relation with relative errors less than 0.9 % were published, representing a significant improvement over the firstly proposed approximation of Marko and Siggia with relative errors of up to 17 %. Also, an approximation for the extension-force dependence with errors less than 0.95 % compared to the numerical solution is given by Petrosyan (2017) [130].

3.1.3 FRC

As a modification of the FJC model, the freely rotating chain model adds the constraint of fixed polar angles of the bonds to the system. The segments are free to rotate around the torsional (azimuthal) angles [131, 135, 136].

$$\langle \vec{t}_i \cdot \vec{t}_j \rangle = l^2 (\cos\theta)^{|i-j|} \tag{3.7}$$

3.1.4 Polymer Backbone Stretching

To account for stretching of the polymer backbone at high forces, 'extensible' model modifications were introduced for both FJC and WLC. Hookean segment spring constants k_s in linear stretching terms compensate for additional enthalpic deformations or extensibility of segments [124, 133, 137]. These models still fail to describe the full range of forces accessible by the AFM, what led to the application of *ab-initio* quantum-chemical calculations to correct for backbone stretching at high forces [138]. Minimization of energy and subsequent derivation delivers a term for the force, dependent on specific material constants for unit cell length at equilibrium a_0, and elastic stretching moduli γ_i:

$$F = \sum_{i=1}^{\infty} \gamma_i \left(\frac{a}{a_0} - 1 \right)^i \tag{3.8}$$

For peptides with a unit cell length of $a_0 = 0.73\,\mathrm{nm}$, the following values for the elastic constants were found: $\gamma_1 = 27.4\,\mathrm{nN}$ and $\gamma_2 = 109.8\,\mathrm{nN}$, with higher orders being negligible.

3.1.5 Force-induced Conformational Change of PEG: a Two-state Model

Poly-ethylene glycol (PEG) linkers are commonly applied to SMFS experiments for tethering molecules of interest to Cantilever tip or cover slide surfaces covalently [104, 107, 139]. While this strategy allows for both, site-specific immobilization of target biomolecules, and passivation against undesired binding of contaminating molecules, PEG molecules display a distinct conformational change upon application of force. Oesterhelt, Rief, and Gaub (1999) found, that the elastic response of PEG molecules in the apolar solvent hexadecane is described well by an extensible Langevin function (FJC model, equation 3.1 with an additional linear term to account for segment elasticity due to bond angle torsion $N_s \cdot F/k_s$, with N_s, the number of segments). In aqueous solutions though, single water molecules can form two hydrogen bonds with the adjacent oxygens of the PEG subunits, effectively stabilizing a shortened trans-trans-gauche (ttg) conformation of the PEG monomers in absence of force. With rising force acting on the polymer, the probability for overcoming enthalpic stabilization by hydrogen bonding rises, shifting an increasingly larger proportion of

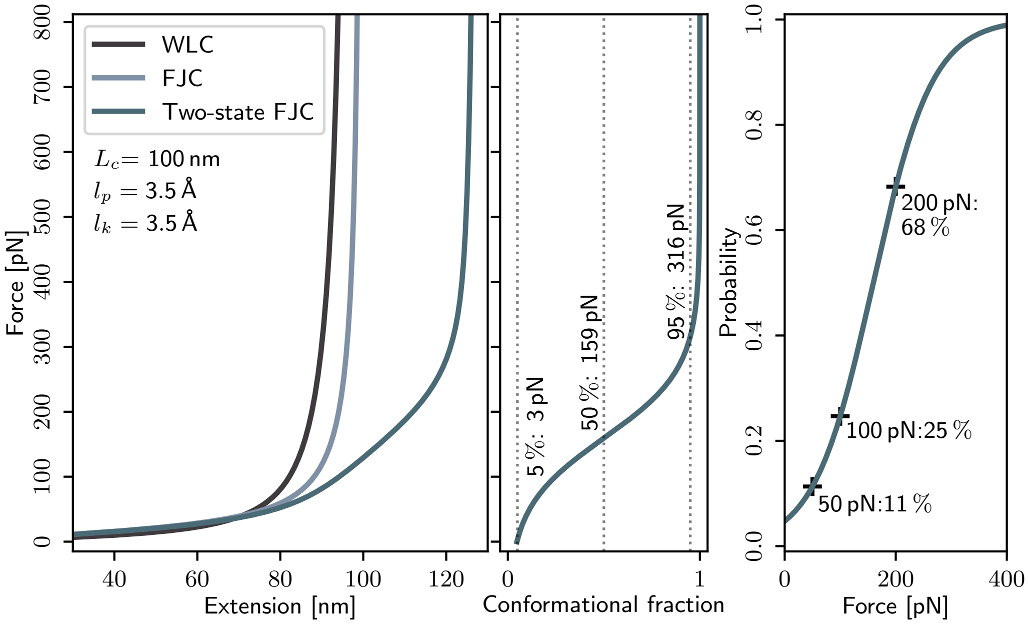


Figure 3.1. Polymer Elasticity Models. The left panel shows force-extension dependencies of a hypothetical polymer of 100 nm contour length and a persistence length (WLC), or Kuhn length (FJC) of 0.35 nm in the models WLC, FJC, and the two-state FJC for PEG linkers in aqueous solutions. The center panel shows the elongated fraction of subunits with the shared force axis of the previous plot. The right panel shows the probability of a subunit in elongated state dependent on the applied force, along with some sample force values. The conformational transition takes place over a broad force range and can therefore compromise measurements of increments in free contour length dramatically. The PEG molecules get elongated by a factor of roughly 1.27 from unloaded state to above 300 piconewton of force loading.

the subunits into the elongated all-trans (ttt) conformation. The overall net penalty in Gibbs free energy from enthalpy difference due to loss of possible hydrogen bonds and entropically favourable stretched state was found to be $\Delta G_0 = 3k_\mathrm{B}T$.

A modification of the FJC model, introducing a Markovian two-level system for the conformational states of the subunits describes the force response of the polymer in water well [140]: Given a Boltzmann distribution for the ratio of both populations $N_{ttg}/N_{ttt} = \mathrm{e}^{\Delta G/k_\mathrm{B}T}$ depending on the free energy altered by the applied force $\Delta G(F) = \Delta G_0 - F(l_{ttt} - l_{ttg})$ is then

$$\langle x \rangle = N_s \left(\frac{l_{ttt}}{\mathrm{e}^{-\Delta G/k_\mathrm{B}T} + 1} + \frac{l_{ttg}}{\mathrm{e}^{\Delta G/k_\mathrm{B}T} + 1} \right) \cdot \left[\coth\left(\frac{F l_k}{k_\mathrm{B}T} \right) - \frac{k_\mathrm{B}T}{F l_k} \right] + \frac{N_s F}{k_s}. \tag{3.9}$$

Solving the Boltzmann distribution for the ratio of states for the force dependent probability of the elongated state shows that the force range of the transition from nearly all subunits in ttg-conformation to all subunits in ttt-state reached from zero to more than 300 pN of, effectively rendering many AFM based SMFS experiments affected by this effect (see figure 3.1, center and right panel for a corresponding plot):

$$p(F) = \frac{1}{\exp\left(F\frac{l_{ttg}-l_{ttt}}{k_{\mathrm{B}}T} + \Delta G\right) + 1} \tag{3.10}$$

This effect can cancel out in case only extensions at comparable forces are considered. Furthermore, the magnitude of this effect scales with the lengths of the linkers. For varying forces, the apparent contour length increments are affected severely if this two-state model is not applied. For typical PEG-linked protein pulling SMFS experiments, the model is not applicable, since the ratio of PEG linker and peptide linkers change with each unfolding event. Extending models further to account for these effects as well introduces a number of additional free parameters, complicating fitting procedure and thereby increasing danger of ending up in local minima in fitting algorithms.

To avoid this problem in the first place, a substitution for PEG linkers in AFM based SMFS experiments was developed during the work on this thesis, eliminating the need for complicated extended models by exclusively utilizing peptide linkers (see section 5.2) [123]. A prior study enabled this method through the development of a molecular cloning technique for highly repetitive protein polymers, thereby allowing for selection of arbitrary lengths or compositions of peptide linkers with any tags for immobilization of target proteins [141].

3.2 Kinetic Theory of Unbinding and Unfolding Dynamics Under Force

In dynamic force spectroscopy of receptor-ligand pairs, kinetic and energetic parameters of the complex are of interest [1]. The methods most prominently used to extract this information from SMFS experiments is to vary the loading rate by measuring the rupture forces at different pulling speeds in constant speed mode [142–144], or with different slopes in force ramp mode [145]. The obtained rupture force data are then assembled into a dynamic force spectrum, a plot of most probable rupture forces against their corresponding loading rates. In their comprehensive guide to analysis of SMFS data sets, Noy and Friddle [146] explain the basic physics of bond stretching. An SMFS measurement corresponds to the stretching of multiple elastic components in series, including the projection of the bond potential onto the pulling axis, the cantilever modelled as a harmonic spring and potential linker molecules with non-linear elasticity deviating from those under investigation. Such a scenario gives rise to bound and unbound states separated by free energy barriers. By pulling on the harmonic spring, this energy landscape is constantly modulated. Since thermal fluctuations are orders of magnitude faster than changes in the external force, the transition from a bound to an unbound state is thermally driven in common loading rate regimes, as described by [147–149]. These models describe a linear dependence of the rupture force on the natural logarithm of the loading rate and give access to the zero-force off rate k_{off}^{0} (exponentially amplified under force) and the distance to the transition state Δx. Theoreticians extended this framework and accounted for modulation of Δx by the applied force [150], and the possibility of rebinding at slow

---

[1] This paragraph was published by Ott *et al.* (2017) [108] in the Journal of Structural Biology and adapted with permission from Elsevier Inc. Copyright ©2016 Elsevier Inc.

loading rates [151]. These newer models predict a non-linear dependence of the most probable rupture force on the loading rate and give the height of the free energy barrier to unbinding ΔG as an additional parameter. Such non-linear trends were observed experimentally, and a comprehensive list of such data sets is given in the work of Friddle, Noy, and De Yoreo (2012). Joint experimental and computational data sets were also analyzed in recent studies [152, 153]. As Noy and Friddle (2013) point out, these models should only be used if the force spectrum of interest indeed exhibits a non-linear trend. If this is not given, fitting non-linear models results in non-meaningful fit parameters and the phenomenological model should be used instead.

Even with the application of force by the AFM cantilever, the unbinding process is still induced by thermal fluctuations in routine SMFS experiments, where the applied force does not completely eliminate the energy barrier [149]. In the view of the model introduced by Bell (1978) and further developed by Izrailev *et al.* (1997) or Evans and Ritchie (1997) [147–149], the unbinding of a molecular complex by application of force with an AFM cantilever can be seen as the superposition of a bond with the interaction potential $U_b(x)$, and the harmonic potential of the moving hookean spring $U_h(x) = \frac{1}{2}k_c(x - vt)^2$, with spring constant k_c. The total one-dimensional potential as the free energy profile along the reaction coordinate x can then be written as $U_{tot}(x) = U_b(x) + U_h(x)$, where v is the (constant) retraction velocity of the harmonic trap and t the time.

3.2.1 From Arrhenius to Kramers' Kinetic Theory of Reaction Rates

Historically, the rate of unbinding k_{off}^0 in absence of force for two molecules A and B interacting in an equilibrium reaction

$$AB \underset{k_{\text{on}}}{\overset{k_{\text{off}}}{\rightleftharpoons}} A + B \tag{3.11}$$

was initially empirically discovered by Hood and then rationalized by van't Hoff and Arrhenius in a law of the form

$$k_{\text{off}}^0 = A \exp\left(-\frac{\Delta G}{k_{\text{B}}T}\right) \tag{3.12}$$

where A is a constant prefactor, ΔG the height of the energy barrier from bound to transition state, and Δx the distance form the energy minimum of the bound state to the maximum of the barrier along the reaction coordinate [154–157]. Later on, the pre-factor A was estimated heuristically for a number of chemical reactions [158]. The detailed theoretical derivation of this formula was done by Kramers (1940), when he related Einstein's theory of Brownian motion (1905) to rate theory [159]: The Markovian Langevin equation for non-linear Brownian motion together with an external force field $U(x)$ in the shape of a one-dimensional asymmetrical double-well potential and an fluctuation force $\xi(t)$ lay the starting point for the Kramers theory of reaction rates [159, 161, 162]

$$m\ddot{x} = -U'(x) - \gamma m\dot{x} + \xi(t). \tag{3.13}$$

Hereby, m denotes the reacting molecule's mass, γ the velocity relaxation rate as a damping term. The fluctuation force $\xi(t)$ is Gaussian white noise with vanishing mean $\langle \xi(t) \rangle = 0$, and therefore obeys the fluctuation-dissipation theorem:

$$\langle \xi(t)\xi(s) \rangle = k_{\mathrm{B}} T m \gamma (t - s). \tag{3.14}$$

Under these prerequisites, Kramers derives the Fokker-Planck equation for the Brownian motion dynamics in phase space under the influence of the non-linear potential. Thereby he describes the time-evolution of the probability density $p(x, t)$ at velocity $v = \dot{x}$ to find the molecule at a given location along the reaction coordinate and at a given time. By evaluating the probability current in phase space, he was able to find an expression for the pre-factor A of equation 3.12 in the Smoluchowski limit of strong friction $\gamma \gg \omega_{\mathrm{T}}$:

$$A \underset{\gamma \gg \omega_{\mathrm{T}}}{\longrightarrow} \frac{\omega_{\mathrm{T}} \omega_{\mathrm{B}}}{\gamma 2\pi} \tag{3.15}$$

with ω_{T} the angular frequency of the transition state and ω_{B} the angular frequency of the bound state, and therefore the unbinding rate yields

$$k_{\mathrm{off}}^{0} = \frac{\omega_{\mathrm{T}} \omega_{\mathrm{B}}}{\gamma 2\pi} \exp\left(-\frac{\Delta G}{k_{\mathrm{B}} T} \right) \tag{3.16}$$

Note that one of the important findings of Kramers condensed in this formula was that the kinetics is governed not only by the thermal oscillation frequency of the bond, but also limited by the coupling of the vibrations to the surrounding molecules, and therefore prone to viscous damping. This coupling is denoted in the pre-factor A. A remarkable review on this topic is given by Hänggi, Talkner, and Borkovec (1990), where they additionally cover a multitude of other interesting contributions to reaction-rate theory [162]. Talkner and Łuczka have given a concise derivation of the kinetic model of time-dependent systems starting from the Fokker-Planck theory [163].

3.2.2 Impact of External Driving Force on the Natural Off-Rate

The first to recognize the negative exponential dependence of lifetimes of solid materials on uni-axial tensile stress σ acting upon it, and come up with a heuristically determined kinetic equation to describe it, was Zhurkov in 1965 [164]. He found the lifetimes of small specimens of materials like strips or fibers of metals, alloys, crystals or polymers all to follow the same exponential relation of the form

$$\tau = \tau_0 \exp\left(\frac{U_0 - x_\beta \sigma}{k_{\mathrm{B}} T} \right), \tag{3.17}$$

where τ_0, U_0, and x_β are characteristic material properties. He further noted, that τ_0 "by both its magnitude and dimensionality (...) coincides with the reciprocal of the natural oscillation frequency of atoms in solids" [164]., and noticed U_0 to be equal to the inter-atomic binding energy in the crystal lattice or the chemical bonds in chain macromolecules for the respective materials. The stress on a bond only decreases the height of the energy barrier by the magnitude of its product with x_β, and the rupture itself is still carried out by thermal fluctuations overcoming the remainder of the barrier.

$$\frac{1}{\tau} = A_0 \exp\left(\frac{x_\beta \sigma - U_0}{k_\mathrm{B} T}\right). \tag{3.18}$$

Bell put this model into the biological context of receptor-ligand bonds on cell surfaces in 1978 and identified the lifetime at zero force $\tau(F = 0)$ with the inverse kinetic rate constant of unbinding $(k_\mathrm{off})^{-1}$ (in this context the tensile stress σ is simply the force F acting on the molecules, x_β the distance to the transition state Δx, and U_0 is interpreted as ΔG) [147]. Therefore the characteristic time constant t_r^0 or inverse of the attempt frequency ν_0 of the diffusive barrier crossing matches the reciprocal pre-factor A:

$$t_\mathrm{r}^0 = \frac{1}{\nu_0} = \frac{l_\mathrm{B} l_\mathrm{T}}{D} = \frac{\gamma 2\pi}{\omega_\mathrm{T}\omega_\mathrm{B}} = \frac{1}{A}, \tag{3.19}$$

with a length scale each for the shape of the bond potential at the bound state and the transition state (l_B, l_T), and a damping term D. t_r^0, also known as diffusive relaxation time, is specific for each individual type of molecular bond, since it is determined by the bond parameters l_B, describing the thermal spread of states in the bound minimum, and l_T the statistically weighted width of the energy barrier at the transition state towards the unbound state, as well as the friction-limited local diffusivity D [148, 165]. Combining the insights from equations 3.16, 3.18 and 3.19 then yields the rate of dissociation under force:

$$k_\mathrm{off}(F) = k_\mathrm{off}^0 \exp\left(\frac{F\Delta x}{k_\mathrm{B}T}\right) = \frac{1}{t_\mathrm{r}^0} \exp\left(\frac{F\Delta x}{k_\mathrm{B}T} - \frac{\Delta G}{k_\mathrm{B}T}\right). \tag{3.20}$$

3.2.3 Derivation of the Probability Distribution and its Maximum

Evans and Ritchie also discuss the effects of different potential shapes like a deep harmonic well or a inverse power law attraction and finally derive the probability for bond dissociation under dynamically increasing external forces: Under the assumption of a quasi-equilibrium at any time, the bound fraction n_b obeys by the first-order linear differential equation $\dot{n}_b(t) = -k_\mathrm{off}(F(t))n_b(t)$. With the fraction of already dissociated (unbound) bonds being $n_u = 1 - n_b$, and replacing t by F(t) as a continuous and reversible function of its argument, integration of this rate equation results in

$$n_u(F) = 1 - \exp\left(-\int_0^F \frac{k_\mathrm{off}(f)}{\dot{f}} \mathrm{d}f\right). \tag{3.21}$$

By differentiation $\frac{\partial}{\partial F} n_u(F)$, the probability of bond failure at force F and given loading rate $\dot{F}$ can be found:

$$p(F) = \frac{k_\mathrm{off}(F)}{\dot{F}} \exp\left(-\int_0^F \frac{k_\mathrm{off}(f)}{\dot{f}} \mathrm{d}f\right) \tag{3.22}$$

$$\underset{(3.20)}{=} \frac{k_\mathrm{off}^0}{\dot{F}} \exp\left(\frac{F\Delta x}{k_\mathrm{B}T} - k_\mathrm{off}^0 \int_0^F \frac{1}{\dot{f}} \exp\left(\frac{f\Delta x}{k_\mathrm{B}T}\right) \mathrm{d}f\right). \tag{3.23}$$

The integral can be evaluated analytically for a constant and positive loading rate $\dot{F}$

$$p(F) = \frac{k_{\mathrm{off}}^0}{\dot{F}} \exp\left(\frac{F\Delta x}{k_{\mathrm{B}}T} - k_{\mathrm{off}}^0 \left[\frac{k_{\mathrm{B}}T}{\dot{F}\Delta x} \exp\left(\frac{f\Delta x}{k_{\mathrm{B}}T}\right)\right]_0^F\right) \tag{3.24}$$

$$= \frac{k_{\mathrm{off}}^0}{\dot{F}} \exp\left(\frac{F\Delta x}{k_{\mathrm{B}}T} - k_{\mathrm{off}}^0 \frac{k_{\mathrm{B}}T}{\dot{F}\Delta x}\left(\exp\left(\frac{F\Delta x}{k_{\mathrm{B}}T}\right) - 1\right)\right), \tag{3.25}$$

and then differentiated to find the maximum of the probability function ($\frac{\partial}{\partial F}p(F) = 0$) and therefore the most probable rupture force $\langle F \rangle$. The model predicts a linear dependence between the most probable rupture force and the logarithm of the loading rate:

$$\langle F \rangle = \frac{k_{\mathrm{B}}T}{\Delta x} \ln\left(\frac{\dot{F}\Delta x}{k_{\mathrm{off}}^0 k_{\mathrm{B}}T}\right) \tag{3.26}$$

To determine the values for the zero-force off rate k_{off}^0 and the distance to the transition state Δx, either equation 3.25 can be fitted to a single force distribution of a single loading rate or equation 3.26 can be fitted to the most probable force values over many orders of magnitudes of loading rates (the dynamic force spectrum).

3.2.4 Applicability and Advanced Models

It is important to understand that this model is only valid within the adiabatic limit, where the molecular relaxation time scale is orders of magnitude shorter (faster processes) compared to the time scale of the external driving. Therefore, the loading rate must be kept sufficiently low. For higher pulling speeds or loading rates in an SMFS experiment, better suited models have been developed and are imperative [152, 166, 167]. On the other end of the loading rate spectrum, at very low loading rates, the combined driving trap and bond system can reach into the equilibrium force, where binding and unbinding happen both at comparable rates [151, 168]. This behaviour is not covered by the model shown here, since for the combination of measurement techniques and protein complexes investigated here, the influence of this regime is reasonably negligible. Both reach and strength of the equilibrium regime, as well as the forces within the kinetic regime are affected by the hookean probe spring constant, so additional care has to be taken apart from plain instrument calibration variance, when results from different experiments are being compared or accurate values for intrinsic molecular properties are of interest [168–170].

Another noteworthy aspect is that this model is best applied to data stemming from a force ramp protocol. In constant speed mode, often the loading rate is estimated to simply comply with the product of the trap or cantilever stiffness and its velocity $\dot{F} = kv$, but for non-negligible polymeric linker lengths this is not a valid assumption, rather a rough approximation at best, since the polymer acts as an entropic spring and has a highly non-linear force response upon increase in extension. In case of sufficiently narrow rupture force distributions, within which the loading rate can be assumed to be nearly constant this may be a good enough approximation. The impact of the elastic properties of the linkers with varying lengths was for example demonstrated by Friedsam *et al.* by Monte Carlo simulations for polymers modelled by the Markovian two-level system (see section 3.1.5), and for continuous distribution of linker lengths [171]. A resolution to this problem is for example the introduction of an FJC or a WLC force-extension relation into the Bell-Evans framework, to predict the loading rate as a function of F in order to respect the force history of the system more accurately as was done by Ray, Brown, and Akhremitchev, or Dudko, Hummer, and Szabo [172–174], respectively. Another, more general and versatile formalism was shown by Freund, taking into account the inherent stiffness of the loading device (see previous paragraph), as well as an arbitrary time dependence of the force exerted on the bond, which among other applications might be of particular interest for multiple bonds rupturing in series (*e.g.*, presence of fingerprint domains, or unfolding intermediates) [175].

The model introduced here also has the drawback of a fixed distance to the transition state Δx, which fails to describe a deformation of the energy barrier upon force application. To address this issue, extensions to this theoretical framework were proposed to include a moving energy barrier [150]. This extended model spins the thought of differently shaped potentials further (see [148]) and introduces a formalism to parametrize the shape of the energy profile into an equation for the most probable rupture force, including a linear-cubic ($\nu = 2/3$) and a cusp-like ($\nu = 1/2$) potentials. For $\nu = 1$, it simply falls back to well-established Bell-Evans equation (3.25).

3.3 Selection of Specific Binders Investigated with the Kinetic Model

Scientific literature has a broad variety of receptor-ligand-like binders, that have been investigated with SMFS. Among these are protein:protein, protein:small-molecule, protein:peptide, Protein:DNA, or DNA-DNA interaction, a selection of which is listed in table 3.1. It lists the molecular weight of each receptor and ligand in kDa, measured rupture forces in pN, the chosen immobilization chemistry in the corresponding publication and the pulling geometry, as well as the reference. It is important to note that rupture forces depend on the loading rate and on the stiffness of the probing device, so the values listed here may only have demonstrative purpose. Especially the dependence of rupture forces on the probe spring stiffness can render comparability between experiments cumbersome (see section 3.2.4) [169, 170]. For more detailed information on the kinetic parameters of some of these systems, table 3.2 additionally shows the distance from the bound state to the transition state in nm, and the off-rate in s^{-1} for selected complexes. These values are used in the plots of figure 3.2, where the dynamic force spectra of all known events of cohesin:dockerin interactions measured in the Gaub Lab are shown. The upper panel shows the overview to provide a scale between the largely varying strengths, and the lower panel zooms into the lower force region, where many events overlap in their values. All data are calculated with the Bell-Evans model (equations 3.25 and 3.26), over a large range of theoretical loading rates, except from the low force CohE:XDoc(Ctta) events, which are calculated with the Dudko-Hummer-Szabo Model, just as how the original data were fitted in the corresponding publication.

Table 3.1. Unbinding Strengths of selected molecular binders, that can be suitable as specific handles for SMFS experiments. Note that rupture forces in almost all cases change significantly with loading rates according to the kinetic theory. If available, the loading rates are given or estimated from the in the original publications. To provide better insight into this behaviour, figure 3.2 shows this dependence for some Cohesin:Dockerin binders. In general, rupture forces also depend on the probe spring constant and on the molecular loading geometry, and therefore on the immobilization sites used for surface conjugation. The given terminus indicates where the respective pulling handle is located in the fusion protein amino acid chain. In case of DNA, the immobilization terminus is given. The terminology X (Y) Z signifies: conjugation of molecule X to Z, cat alyzed by enzyme Y

Binding molecules	**MW [kDa]**	**Force [pN]**	**Ld. Rate [pN s^{-1}]**	**Strategy of Immobilization**	**Termini**	**Ref**
Coh(GH84C X82):Doc (*C.p.*)	-	50-63	$10^3 - 10^4$	NHS-PEG5k-Mal+ CoA(SFP)ybbR	N:C	[176]
Coh(CipA):DocS (*C.t.*)	15.4/8.3	90 – 100	$10^3 - 10^4$	NHS-PEG5k-Mal+Cys	C:C SGL	[143]
		120 – 150	$10^3 - 10^4$		C:C DBL	[143]
CohE:XDoc(Ctta) (*R.f.*)	21.6/26.2	600 – 750	$10^4 - 10^5$	NHS-PEG5k-Mal+Cys	N:C	[142]
		~ 120	10^4	NHS-PEG5k-Mal+ CoA(SFP)ybbR	C:C LF	[153]
		~ 600	10^4		C:C HF	
NiNTA:HIS6	0.2/0.8	153 ± 57	6.6×10^3	Gold-Cys	-	[177]
Avidin:Biotin	66-69/0.2	160 ± 20	NA	Nonspecific adsorption (BSA)	-	[96]
Avidin:Desthiobiotin		125 ± 20	NA		-	
Avidin:Iminobiotin		85 ± 15	NA		-	
Streptavidin:Biotin	52.8/0.2	60 – 90	$10^3 - 10^4$	Biotinylated BSA	-	[178]
		100 – 180	$10^4 - 10^5$		-	
Calmodulin:skMLCK	16.7/1.1	19.3 ± 0.8	EQM	Pulldown via NI-NTA	-	[179]
StrepTagII:mvST	1.1/58.4	90 – 175	$10^3 - 10^5$	NHS-PEG5k-Mal+ [Cys or CoA(SFP)ybbR]	C:C	[144]
		50 – 90	$10^3 - 10^5$		N:C	
Biotin:mvSA	0.2/54	200 – 230	$10^3 - 10^5$	NHS-PEG5k-Mal+Cys	N:C	[180]
scFv:GCN4(7P14P)	26.7/4.0	45 – 65	$10^2 - 10^4$	NHS-PEG5k-Mal+Cys	N:C	[181]
Anti-DIG:DIG	170/0.4	25 – 90	$10^2 - 10^5$	NHS-PEG6k	-	[182]
20bp dsDNA, shear	26.2/9.3	54 ± 14	2×10^3	NHS-PEG5k-Mal+SH	5′/5′	[183]
30bp dsDNA, shear	26.2/12.3	65 ± 14	$\sim 2 \times 10^3$	NHS-PEG5k-Mal+SH	5′/5′	
poly(GC) dsDNA, zipper	-	20±3	EQM	unspecific adsorption to Au and Si_3N_4	-	[184]
poly(AT) dsDNA, zipper	-	9±3	EQM		-	

Abbreviations: Coh: Cohesin; Doc: Dockerin, *C.p.*: *Clostridium perfringens*; DocS: Dockerin from Cels48S; CohE: Cohesin from ScaE; XDoc: XModule-Dockerin; *C.t.*: *Clostridium thermocellum*; *R.f.*: *Ruminococcus flavefaciens*; mvST: monovalent Streptactin; mvSA: monovalent Streptavidin; DIG: digoxigenin; NHS: N-hydroxysuccinimide; PEGnk: poly-ethylene glycol of n kDa molecular weight; Mal: maleimide; Cys: cysteine; CoA: coenzyme A; SFP: 4′-phosphopantetheinyl transferase; ybbR-Tag: peptide sequence DSLEFIASKLA; LF: low force unbinding path; HF: high force unbinding path.; NA: data not available; EQM: equilibrium conditions | In analogy to Ott *et al.* (2017) [108]

Table 3.2. Kinetic Parameters of selected molecular binders, that can be suitable as specific handles for SMFS experiments. The given terminus indicates where the respective pulling handle is located in the fusion protein amino acid chain.

Binding molecules	**Δx [nm]**	**k^0_{off} [s^{-1}]**	**Termini**	**Reference**
Coh(GH84C X82):Doc (*C.p.*)	0.77	1.1×10^{-2}	N:C	Milles *et al.* [176]
Coh(CipA):DocS (*C.t.*)	0.7	3×10^{-5}	C:C SGL	Stahl *et al.* [143]
	0.6	1×10^{-6}	C:C DBL	
CohE:XDoc(Ctta) (*R.f.*)	0.13	7.3×10^{-7}	N:C +XMOD	Schoeler *et al.* [142]
	0.19	4.7×10^{-4}	N:C -XMOD	Schoeler *et al.* [142]
	0.42	5×10^{-3}	C:C DHS/LF	Schoeler *et al.* [153]
StrepTagII:mvST	0.23	3.4×10^{-1}	C:C	Baumann *et al.* [144]
	0.45	6×10^{-1}	N:C	
Biotin:mvSA	0.38 ± 0.02	3×10^{-6}	N:C	Sedlak *et al.* [180]
scFv:GCN4(7P14P)			N:C	Morfill *et al.* [183]
Anti-DIG:DIG			-	Neuert *et al.* [182]
20bp dsDNA, shear geometry	2.89±0.77	$8.1\pm0.8\times10^{-14}$	5′/5′	Morfill *et al.* [183]

Abbreviations: Coh: Cohesin; Doc: Dockerin, *C.p.*: *Clostridium perfringens*; DocS: Dockerin from Cels48S; CohE: Cohesin from ScaE; XDoc: XModule-Dockerin; *C.t.*: *Clostridium thermocellum*; *R.f.*: *Ruminococcus flavefaciens*; mvST: monovalent Streptactin; mvSA: monovalent Streptavidin; +XMOD: xModule is intact; -XMOD: xModule unfolded before; SGL: rupture of the single peak pathway; DBL: final rupture of the double peak pathway; BE: Bell-Evans model; DHS: Dudko-Hummer-Szabo model; LF: low force unbinding path; HF: high force unbinding path.

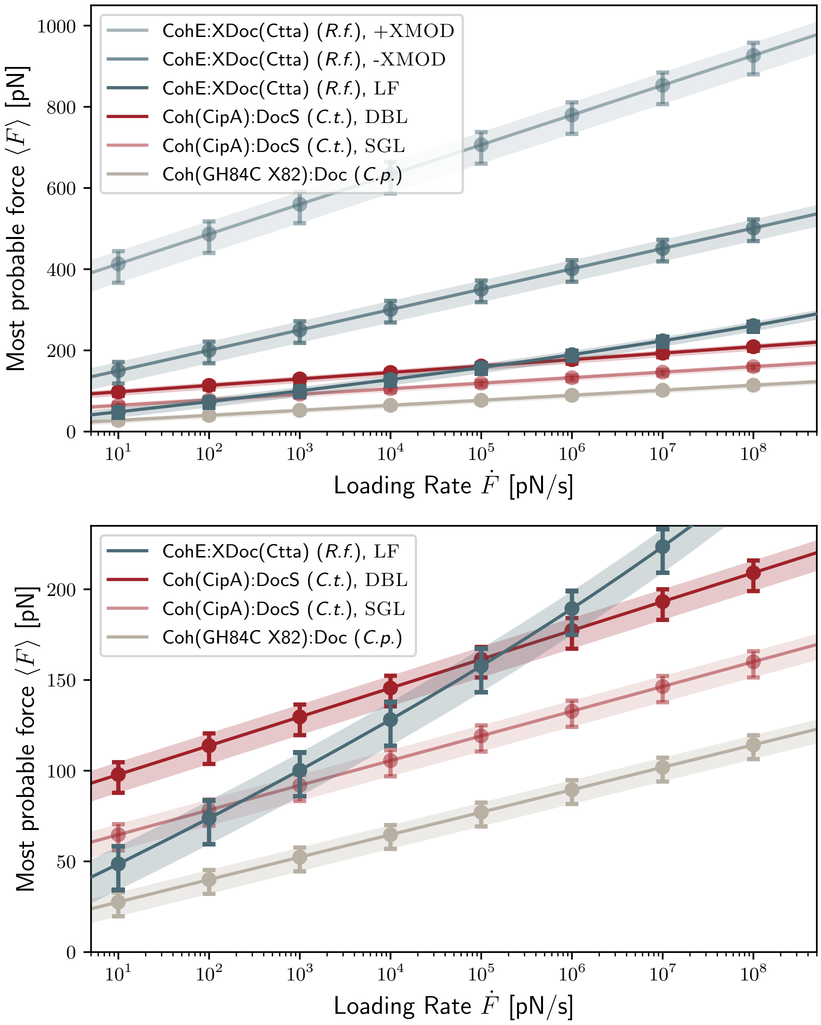


Figure 3.2. Theoretical force spectra for various Cohesin:Dockerin complexes, obtained by inserting experimentally found values (table 3.2) for Δx and k^0_{off} into the Bell-Evans equation 3.25 (most probable forces $\langle F \rangle$, and asymmetric full-width-half-maxima of probability distributions (FWHM) for individual loading rates) and equation 3.26 (global dependencies of $\langle F \rangle$ and FWHMs (shaded areas) over the whole range of loading rates). Upper panel: all Cohesin:Dockerin complexes; Lower panel: zoom into the low-force region for better visibility of weaker rupture events. Complexes from *Clostridium perfringens* (*C.p.*) are shown in grey, complexes from *Clostridium thermocellum* (*C.t.*) in red tones, and complexes from *Ruminococcus flavefaciens* (*R.f.*) in shades of blue. It is notable, that due to it's narrow potential well width, the type 3 CohE:XDoc has a extraordinarily steep loading rate dependence compared to other Coh:Doc interaction, even for the non-native low force population.

4

Application to Biological Systems

4.1 Cellulosomal Components as Model Systems

Cellulosomes are vast and complex enzymatic machineries used by anaerobic fungi and bacteria for the decomposition of cellulosic and hemicellulosic and other plant cell wall substrates into simple sugars, that can be used by their metabolism. Cellulose consists of glucose subunits connected covalently *via* β(1→4)-glycosidic bonds [185]. It is stabilized by intra-and intermolecular hydrogen bonds to from microfibrils and cellulose fibrils on different levels of organization. Additional protection against degradation is provided by interconnection and interweaving as a copolymer with other substrates, like lignin or hemicellulose, rendering enzymatic digestion even more difficult.

The cellulosome solves this problem by organizing a number of different enzymes, each specialized in their own task. They can act together to break down the complex copolymer efficiently [186, 187]. These enzymes include endo-cellulases, that can cut the crystalline polymer strands into smaller cellulose subunits, thereby creating new polymer ends, and exo-cellulases, that can digest those ends processively. Others can further process their products, such as β-glucosidases being specialized for the hydrolysis of di- or tetrasaccharides into monosaccharides.

Furthermore, the cellulosome features cellulose binding modules (CBM), that can anchor the digestion scaffold holding the enzymes in close proximity to their substrate [188–190]. The CBMs are covalently bound to the peptide backbone of the scaffolding protein network as fusion domains. Substructures are attached non-covalently by highly specific and high-affinity receptor-ligand protein domains termed cohesin (Coh) and Dockerin (Doc). These domains in turn are either expressed as fusion domains with single endogenous cellulolytic enzymes (Type I Coh-Doc) or with yet other scaffolds, enabling branching of the whole structure (Type II Coh-Doc). Typically, a scaffold molecule has many Coh domains and often a CBM. In gram-positive bacteria, the initial scaffold is usually attached to the host cell by a Coh-Doc (Type III Coh-Doc) interaction, with the Coh domain being anchored by a sortase enzyme *via* its `LPXTG` peptide motif [191].

Depending on the specific task that the Coh and Doc domains fulfil, and the position where they are located within the scaffold, they display highly differing mechanical strengths and chemical affinities [82]. Table 3.1 shows mechanical strength of selected Coh-Doc pairs (among other specifically binding molecules), Table 3.2

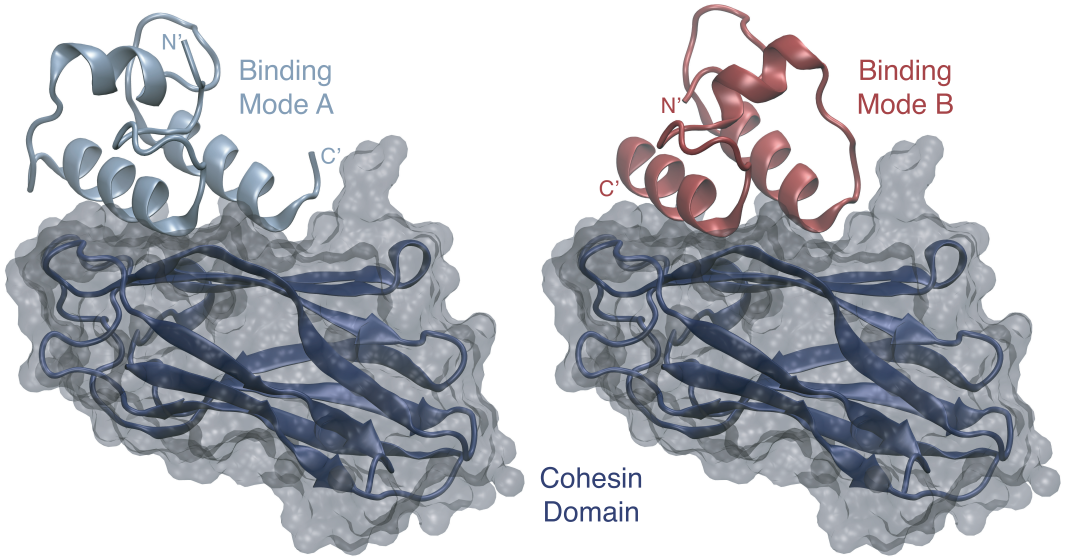


Figure 4.1. Dual binding modes of the Coh-Doc type I (*C.t.*) binding interface. Both conformations are from the exact same viewing angle (light blue and red), with the Coh domains (grey/dark blue) aligned with VMD's multisec tool [192]. In the two modes, the Doc domain is rotated by approximately 180 degrees on the Coh binding surface (Mode A: PDB 2CCL, Mode B: PDB 1OHZ).

shows kinetic parameters accordingly, and figure 3.2 visualizes the dynamic force spectra of Coh-Doc interactions. The non-covalent organization is thought to benefit exchange of enzymes, when they are broken or need to be replaced due to substrate variation without the need for rebuilding the whole scaffold system, thereby saving precious resources.

The orthogonal yet highly homologous high-affinity binding domains, the modularity, and the enzymatic composition render the cellulosome an interesting subject of scientific investigation. During the work on this thesis, the main focus was set to utilize them as part of a high-throughput SMFS assay in a microfluidic chip format (section 5.4), elucidate their extraordinary strength and high symmetry binding interface (sections 4.2, 5.1, 5.3, and 6.2), and apply their modularity to a force hierarchy-driven molecular assembly assay (sections 4.3 and 6.3).

4.2 Dual Binding Modes

The small type I (*Clostridium thermocellum*, *C.t.*) Doc domain has a duplicated calcium induced folding F-hand motif, resulting in twin helices each with an adjacent calcium loop. Both are connected with an intermediate short helix and in anti-parallel orientation relative to each other [193–196]. The structure shows a rotational symmetry not only by the two helices as secondary structural elements, but also crucial amino acids responsible for binding the surface of the Coh counterpart are included in the symmetry. This intriguing geometry lead to the proposal of two possible binding conformations, which could be confirmed by mutational studies in crystal structures for *C.t.*, *Clostridium cellulolyticum* (*C.c.*), and *Acetovibrio cellulolyticus* (*A.c.*) type I Doc domains [197–199]. The larger type III (*Ruminococcus flavefaciens*, *R.f.*) Doc, comprises an adjacent Xmodule structure, supposedly for structural stabilization [142, 153, 200], and shows also some symmetry within the binding interface, but not to the extent of the type I class domains.

The existence of two possible specific binding conformations within a single interface is thought to provide flexibility in cellulosome assembly to overcome steric hindrances by enzymes, allow for different and possibly evolving enzyme combinations to degrade challenging and inflexible solid substrate, or enable enhancement and orientational optimization of degradation by rearranging through rotation during the hydrolysis processes [197, 198]. Based on the conservation throughout many bacterial dockerin domains, this characteristic is thought to be fundamental to the catalysis mechanism, since dual binding possibilities may contribute to the conformational plasticity. Although the linkers between catalytic and binding domains in Doc-enzyme fusion proteins can also provide flexibility, the rigidity of those linkers was shown to increase upon binding to the corresponding Coh domain on the scaffold, which could be then in turn be compensated for by the two modes of binding [198, 201, 202]. Other possible benefits of this principle include robustness against loss-of-function mutations, synergistic communalism within communities of different microbes, or the reduction of proteolysis in long linkers while remaining flexibility through dual modes [198]. A completely different explanation could be the plain chance of probability of the genesis of symmetry in evolutionary variation processes due to algorithmic bias [6]. During the work on this thesis, for the first time evidence was found that this dual mode of binding is not only detectable with mutational studies in the crystallized molecules by blocking one of both modes, but both modes indeed exist in parallel in the wild-type molecule and under native conditions (see section 5.3) [122]. The high analogy and symmetry in the type III (*r.f.*) Coh-Doc binding interface suggest existence of a similar mechanism in these molecules and evidence searched for in this system as well (section 6.2).

4.3 Molecular Cut and Paste Assay

Since its invention in 2008, single-molecule cut and paste (SMCP) aimed to enable the creation of new functions by building signaling cascades, or investigate synergistic enzymatic activities through the spatial arrangement of biomolecular patterns [99]. It complements other bottom-up approaches like dip-pen nano-lithography, or DNA-Origami [203–205] as a directed spatial assembly method. It enables control of position, density, composition, geometry of individual molecules by specificity and force hierarchy of handle molecules.

Since the macromolecular concept of pliers does not have an equivalent in the nanoscopic world, the whole concept relies on chemical affinities and mechanical hierarchies for transferring the molecules. This is can be achieved by the force loading geometry (shear or zipper) and length of DNA strand hybridization, and incorporation of antibodies, or protein receptor-ligand interactions was also shown to be useful [206, 207]. a force hierarchy between three handle molecule systems for each step, the following assembly cycle can be realized: a high density storage area (depot, low unbinding force) allows specific pick-up of single or few molecules with an AFM cantilever (medium force). After transfer to an empty target area, the cargo molecules

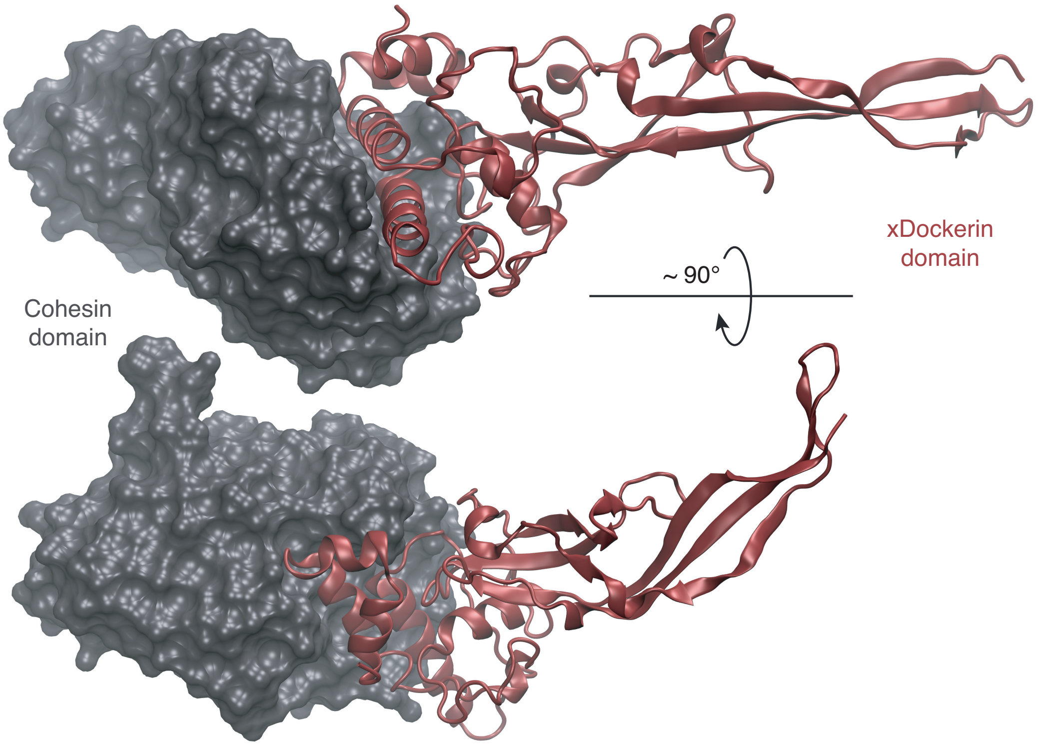


Figure 4.2. Binding interface of type III Cohesin:Dockerin (PDB 4IU3). The parallel double helix interface of the type III Doc domain displays high symmetry not only in their tertiary structure, but also in their amino acids responsible for binding. This is analogous to the the type I Coh:Doc interface.

are bound to their final handles (high force). Repetition of these steps allow assembly of larger structures or combination of different types of molecules in an area.

The combination of AFM SMFS with a total internal fluorescence microscope allows control of the process up to super-resolution [100, 208]. Also hybrid approaches of assembling patterns of binding sites were shown, which then allowed self-assembly of fluorescent nanoparticles or reconstitution of split aptamers [121, 209]. Mostly, DNA handles were used, which allow high precise tuneability due to the selectivity of their nucleotide hybridization. A pure protein based concept, that allows integration *via* IVTT systems into microfluidic devices was part of the work on this thesis (see section 6.3) [210–213].

4.4 Force Spectroscopy on DNA Origami Superstructures

DNA origami nanotechnology exploits sequence-specific hybridization of DNA strands to create arbitrary structures by folding a single-stranded large DNA backbone with short oligo-nucleotides ('staples') into desired two- or three-dimensional shapes [204, 214, 215]. Usually, a backbone strand derived from M13mp18 bacteriophage genomic DNA gets folded by chemically synthesized staple strands in in solutions of excess staple molar ration and under high divalent salt conditions. Staples stabilize the folded structure by crossing over between adjacent backbone strands

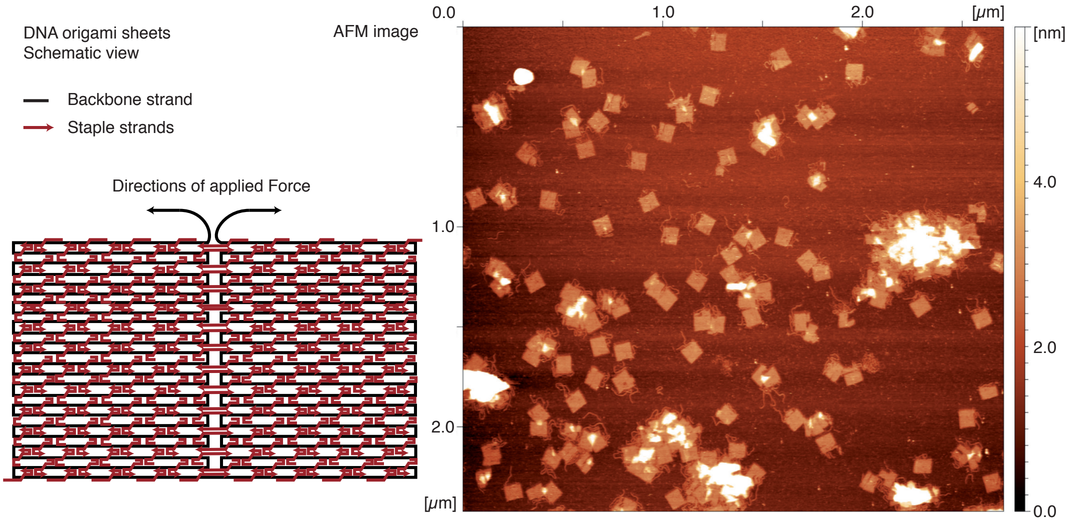


Figure 4.3. DNA origami. Left: schematic view of a single DNA origami sheet with backbone strand in black, staple strands in red. Right: AFM image of the DNA origami sheets measured in section 6.1 (courtesy of Dr. David M. Smith).

and partially hybridize with nucleotides in each strand. The process is based on self-assembly after melting molecules to single-stranded DNA, and promoting specific complementary hybridization of the desired parts by slowly cooling down the solution. Cooperativity effects can correct for misfolding of structures, thereby make the process extraordinarily reliable, and also allow for isothermal folding strategies [216–218]. Recent developments also include DNA origami-graphene hybrids for nanopore sequencing [219], hybrid DNA-protein origami by using sequence specific DNA binding domains of transcription activator like (TAL) effectors as staples [220], the development of single stranded DNA and RNA origami without the need for staple strands [221], or enabling mass production by bacteriophage-based production of staples from self-excising cassettes by auto-catalytic DNAzyme sequenced [222]. Within the work on this thesis, a coarse-grained simulation approach (oxDNA framework) was compared with experimental data to investigate mechanical stability and force-induced unfolding behaviour of DNA origami macromolecular structures (figure 4.3 shows an AFM image of the origami samples measured in AFM SMFS). It was shown, that both approaches agree well within their results and can be applied to further develop mechanical functionality within the DNA origami and understand principles of mechanotransduction within macromolecules (see section 6.1 [223]).

4.5 PUBLICATION P1: State of the Art SMFS

This publication addresses developments in SMFS techniques during the past few years. Applications of biomolecular technologies to the method, innovations in instrument and assay design, and data analysis and theoretical framework are highlighted. My contribution of writing primarily focused on the technical methodology and instrument development. I co-developed the concept for the manuscript, and authored the chapters on time resolution, time-scale gap between experiment and simulation, stability and force precision, multi-parametric imaging modes, end lateral force sensors. I co-designed the figures and co-wrote their captions, and proof-read and finalized the whole manuscript.

Single-molecule Force Spectroscopy on Poly-proteins and Receptor-ligand Complexes: The Current Toolbox

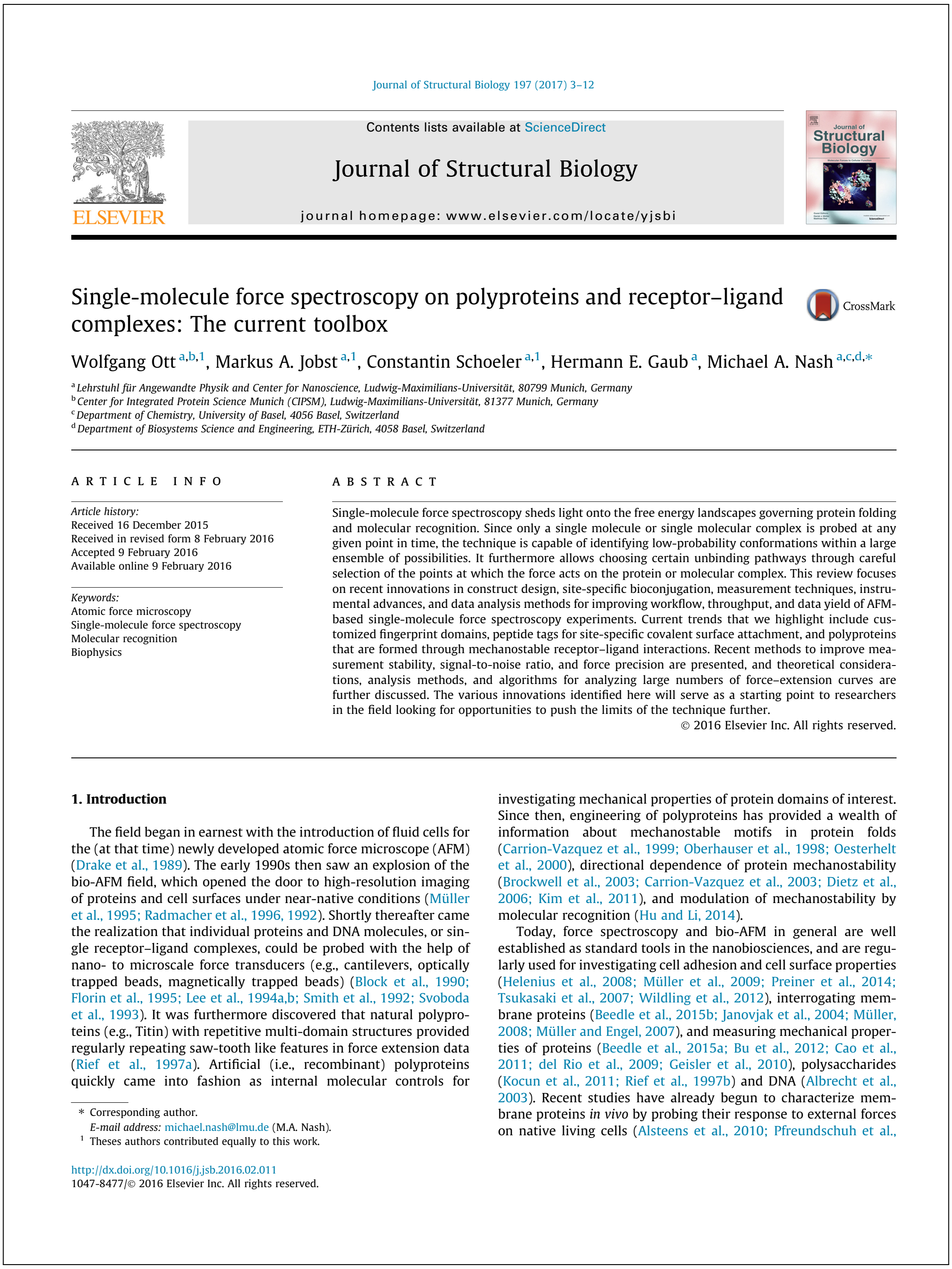
Wolfgang Ott[†], Markus A. Jobst[†], Constantin Schoeler[†],
Hermann E. Gaub and Michael A. Nash

[†]*these authors contributed equally to this publication*

published in

Journal of Structural Biology, 197 (1), 3-12, (2017)

Reprinted from [108], with permission from Elsevier



Journal of Structural Biology 197 (2017) 3–12

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox

Wolfgang Ott [a,b,1], Markus A. Jobst [a,1], Constantin Schoeler [a,1], Hermann E. Gaub [a], Michael A. Nash [a,c,d,*]

[a] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[b] Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, 81377 Munich, Germany
[c] Department of Chemistry, University of Basel, 4056 Basel, Switzerland
[d] Department of Biosystems Science and Engineering, ETH-Zürich, 4058 Basel, Switzerland

ARTICLE INFO

Article history:
Received 16 December 2015
Received in revised form 8 February 2016
Accepted 9 February 2016
Available online 9 February 2016

Keywords:
Atomic force microscopy
Single-molecule force spectroscopy
Molecular recognition
Biophysics

ABSTRACT

Single-molecule force spectroscopy sheds light onto the free energy landscapes governing protein folding and molecular recognition. Since only a single molecule or single molecular complex is probed at any given point in time, the technique is capable of identifying low-probability conformations within a large ensemble of possibilities. It furthermore allows choosing certain unbinding pathways through careful selection of the points at which the force acts on the protein or molecular complex. This review focuses on recent innovations in construct design, site-specific bioconjugation, measurement techniques, instrumental advances, and data analysis methods for improving workflow, throughput, and data yield of AFM-based single-molecule force spectroscopy experiments. Current trends that we highlight include customized fingerprint domains, peptide tags for site-specific covalent surface attachment, and polyproteins that are formed through mechanostable receptor–ligand interactions. Recent methods to improve measurement stability, signal-to-noise ratio, and force precision are presented, and theoretical considerations, analysis methods, and algorithms for analyzing large numbers of force–extension curves are further discussed. The various innovations identified here will serve as a starting point to researchers in the field looking for opportunities to push the limits of the technique further.

1. Introduction

The field began in earnest with the introduction of fluid cells for the (at that time) newly developed atomic force microscope (AFM) (Drake et al., 1989). The early 1990s then saw an explosion of the bio-AFM field, which opened the door to high-resolution imaging of proteins and cell surfaces under near-native conditions (Müller et al., 1995; Radmacher et al., 1996, 1992). Shortly thereafter came the realization that individual proteins and DNA molecules, or single receptor–ligand complexes, could be probed with the help of nano- to microscale force transducers (e.g., cantilevers, optically trapped beads, magnetically trapped beads) (Block et al., 1990; Florin et al., 1995; Lee et al., 1994a,b; Smith et al., 1992; Svoboda et al., 1993). It was furthermore discovered that natural polyproteins (e.g., Titin) with repetitive multi-domain structures provided regularly repeating saw-tooth like features in force extension data (Rief et al., 1997a). Artificial (i.e., recombinant) polyproteins quickly came into fashion as internal molecular controls for investigating mechanical properties of protein domains of interest. Since then, engineering of polyproteins has provided a wealth of information about mechanostable motifs in protein folds (Carrion-Vazquez et al., 1999; Oberhauser et al., 1998; Oesterhelt et al., 2000), directional dependence of protein mechanostability (Brockwell et al., 2003; Carrion-Vazquez et al., 2003; Dietz et al., 2006; Kim et al., 2011), and modulation of mechanostability by molecular recognition (Hu and Li, 2014).

Today, force spectroscopy and bio-AFM in general are well established as standard tools in the nanobiosciences, and are regularly used for investigating cell adhesion and cell surface properties (Helenius et al., 2008; Müller et al., 2009; Preiner et al., 2014; Tsukasaki et al., 2007; Wildling et al., 2012), interrogating membrane proteins (Beedle et al., 2015b; Janovjak et al., 2004; Müller, 2008; Müller and Engel, 2007), and measuring mechanical properties of proteins (Beedle et al., 2015a; Bu et al., 2012; Cao et al., 2011; del Rio et al., 2009; Geisler et al., 2010), polysaccharides (Kocun et al., 2011; Rief et al., 1997b) and DNA (Albrecht et al., 2003). Recent studies have already begun to characterize membrane proteins *in vivo* by probing their response to external forces on native living cells (Alsteens et al., 2010; Pfreundschuh et al.,

---

* Corresponding author.
E-mail address: michael.nash@lmu.de (M.A. Nash).
[1] Theses authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jsb.2016.02.011
1047-8477/© 2016 Elsevier Inc. All rights reserved.

2015). There are a number of review articles that thoroughly cover the field from the early years (Carvalho et al., 2013; Casuso et al., 2011; Hoffmann and Dougan, 2012; Lee et al., 2007; Li and Cao, 2010; Marszalek and Dufrêne, 2012; Müller and Dufrêne, 2008; Neuman and Nagy, 2008; Noy, 2011; Rief and Grubmüller, 2002; Sirbuly et al., 2015; Woodside and Block, 2014).

Despite the high level of interest and well-developed method of AFM-SMFS (Single Molecule Force Spectroscopy), there have remained several limitations to the technique that prevent researchers from fully taking advantage of mechano-phenotyping of molecules and cell surfaces. Specifically, low experimental throughput and low yield of useable single-molecule interaction curves have both hampered the widespread adoption of the method, and its application for studying a large number of proteins. The purpose of this review is to highlight recent developments in bioconjugate chemistry, instrumentation, and data processing/ algorithms which aim at improving the design process, yield, measurement quality and throughput of AFM-SMFS experiments.

2. Unfolding fingerprints

In typical AFM-SMFS experiments, many thousand force–extension curves are recorded, but only a fraction of these curves contain useable data that describe the behavior of a single molecule. Typically, the majority of curves (~80–99%) contain no interaction, a multiplicity of interactions that are difficult to interpret, or unspecific adhesion events as measurement artifacts. The experimenter is left searching for a needle in a haystack, looking for single-molecule interactions among a vast excess of unusable force–extension curves. In order to filter the data efficiently, the SMFS community has identified a broad range of proteins that can be used as specific identifiers in unfolding traces. We refer to these domains as ‘fingerprints’ because they provide a unique unfolding step or ‘contour-length increment’ of defined length that can be used as a filter during data processing. These fingerprint domains are typically globular protein domains with individual unfolding forces and length increments varying across a large range. This ability to choose the length increments and unfolding forces of the fingerprint domains has enabled the design of custom fusion proteins with well-controlled unfolding behaviors. Recent surveys of mechanical properties of different protein domains are provided by Sułkowska and Cieplak (2007), Hoffmann and Dougan (2012).

3. Receptor–ligand SMFS

Protein–protein and protein-small molecule interactions have been widely analyzed with SMFS. Reports of receptor–ligand SMFS include measurements on biotin–avidin (Florin et al., 1994; Lee et al., 1994a,b; Moy et al., 1994; Rico and Moy, 2007; Yuan et al., 2000), antigen–antibody interactions (Hinterdorfer et al., 1996; Morfill et al., 2007; Schwesinger et al., 2000) along with several other protein–protein or small molecule interactions (Lee et al., 2007; Mitchell et al., 2007; Schmidt et al., 2012).

One limitation in the standard method of receptor–ligand SMFS is that the signal lacks single-molecule specificity. Depending on the proteins involved and the experimental conditions (i.e., blocking/passivation steps), and since typically no fingerprint molecules are used, it can be difficult to differentiate non-specific interactions from specific protein-protein recognition. A second limitation of many receptor–ligand SMFS experiments is that pulling geometry is not strictly controlled. While in a standard polyprotein experiment, the force is applied strictly between the N- and C-termini of each domain, coupling of receptors and ligands to AFM tips and substrates is often done through amide linkages formed between amine groups on the proteins and activated NHS-ester groups on the surface or cantilever. This implicates a diversity of pulling geometries which are not strictly controlled, resulting in rupture force distributions that are smeared out or otherwise distorted.

4. Receptor–ligand SMFS with fingerprints

Our group has worked on improving the technique for receptor–ligand SMFS out of sheer necessity (Fig. 1). We were interested in studying a family of receptor–ligand proteins (i.e., cohesin–dockerin, Coh–Doc) involved in carbohydrate recognition and degradation by anaerobic bacteria (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2015, 2014; Stahl et al., 2012). These protein receptor–ligand complexes are responsible for building up large extracellular networks of structural scaffold proteins and enzymes. They are linked into these structural networks in well-defined and known orientations (e.g., N-terminal or C-terminal anchoring points). It is important to note that when pulling apart a receptor–ligand complex consisting of two proteins, there are four possible terminal pulling configurations (i.e., N-N′, N-C′, C-N′, C-C′) (Fig. 1B). Many of the Coh–Doc complexes we are interested in possess a clear ‘physiological’ pulling configuration found in nature, and ‘non-physiological’ or ‘non-native’ configurations. To understand their natural mechanical adaptations giving rise to their remarkable assembly strategy, we sought to characterize the mechanical stability of these receptor–ligand complexes in both their native and non-native loading configurations. We found a way to ensure specific interactions by basically combining two previously separate modes of AFM-SMFS (i.e., on polyproteins and receptor–ligand complexes). We fused the Coh and Doc domains separately to different fingerprint domains, and recombinantly produced each construct as a single fusion protein. The fingerprints serve two purposes: (1) they provide site-specific attachment sites through engineered cysteine residues or peptide ligation tags (see section 5) to strictly control loading geometry; (2) they provide predetermined increments in contour length which allows us to filter the datasets for specific single-molecule interactions (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2015, 2014; Stahl et al., 2012).

This configuration yields several advantages: We now have the ability to study mechanical stability of receptor–ligand pairs and unfolding of individual domains (i.e., the fingerprints) in a single-experiment with high yield and specificity, eliminating measurement artifacts. We also have a systematic and straightforward way to probe effects of pulling geometry on receptor–ligand unbinding, and to compare native and non-native pulling configurations. The gene design (i.e., N- or C-terminal fingerprint domains) directly reflects the conformation to be investigated. Furthermore, a specific protein domain of interest can now easily be fused to a mechanostable Coh–Doc receptor–ligand pair for characterization. Depending on the expected domain unfolding forces, an appropriately fitting protein receptor–ligand pair can be chosen from a wide range of well-characterized molecules (Table 1). We note that this table does not include every receptor–ligand probed by AFM. For an extensive list of receptor–ligands that were explored with AFM, see Lee et al. (2007). Currently, the mechanically most stable receptor–ligand pair is a Coh–Doc type III complex derived from *R. flavefaciens*, with loading-rate dependent rupture forces between 600 and 800 pN (Schoeler et al., 2015, 2014). Another interaction in a similar force range is the trimeric titin–telethonin complex described by Bertz et al. (2009).

5. Site-specific bioconjugation

Many polyprotein experiments rely on non-specific adsorption of polyproteins onto surfaces (e.g., mica, gold). Receptor–ligand

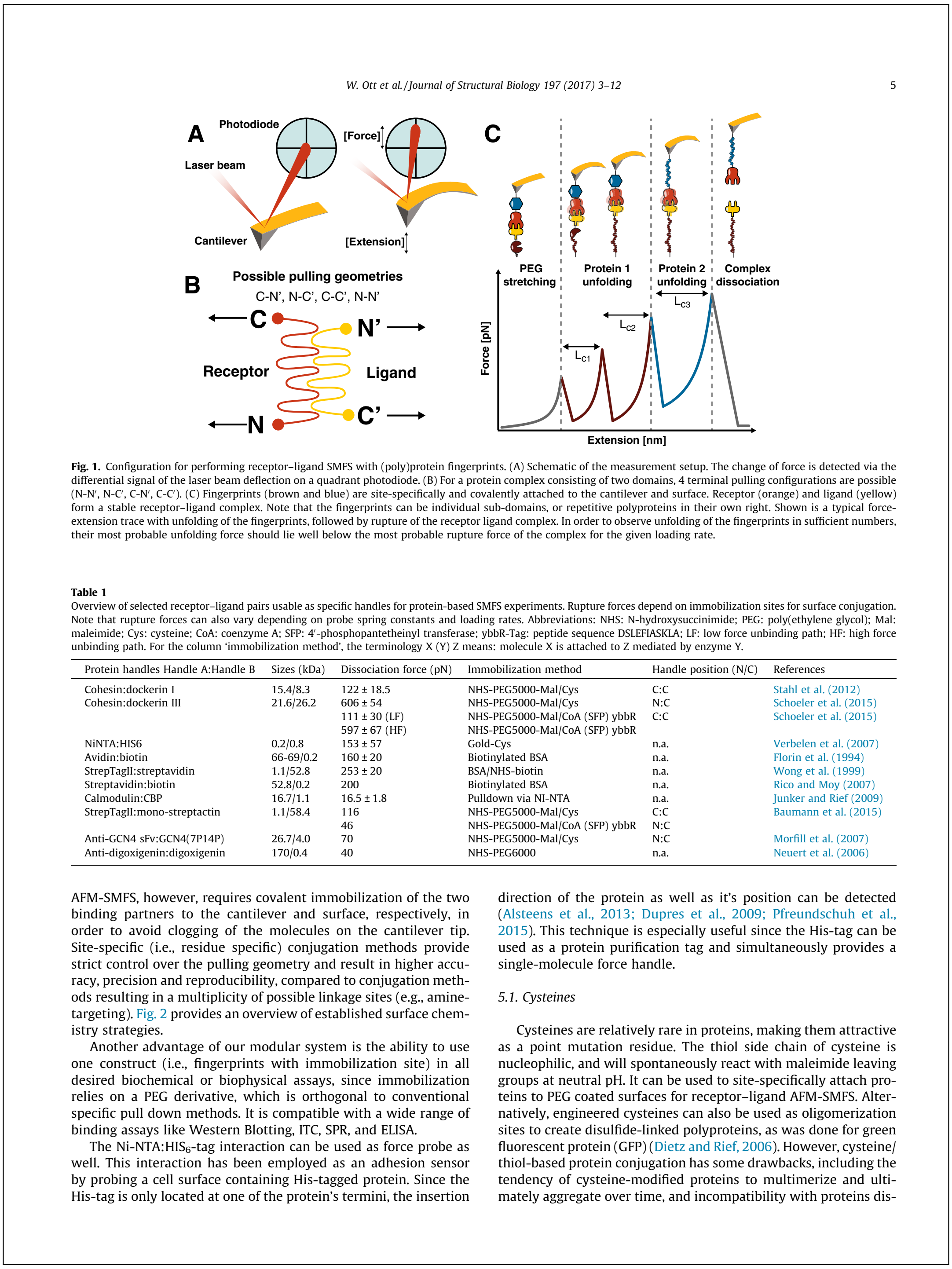


Fig. 1. Configuration for performing receptor–ligand SMFS with (poly)protein fingerprints. (A) Schematic of the measurement setup. The change of force is detected via the differential signal of the laser beam deflection on a quadrant photodiode. (B) For a protein complex consisting of two domains, 4 terminal pulling configurations are possible (N-N′, N-C′, C-N′, C-C′). (C) Fingerprints (brown and blue) are site-specifically and covalently attached to the cantilever and surface. Receptor (orange) and ligand (yellow) form a stable receptor–ligand complex. Note that the fingerprints can be individual sub-domains, or repetitive polyproteins in their own right. Shown is a typical force-extension trace with unfolding of the fingerprints, followed by rupture of the receptor ligand complex. In order to observe unfolding of the fingerprints in sufficient numbers, their most probable unfolding force should lie well below the most probable rupture force of the complex for the given loading rate.

Table 1
Overview of selected receptor–ligand pairs usable as specific handles for protein-based SMFS experiments. Rupture forces depend on immobilization sites for surface conjugation. Note that rupture forces can also vary depending on probe spring constants and loading rates. Abbreviations: NHS: N-hydroxysuccinimide; PEG: poly(ethylene glycol); Mal: maleimide; Cys: cysteine; CoA: coenzyme A; SFP: 4′-phosphopantetheinyl transferase; ybbR-Tag: peptide sequence DSLEFIASKLA; LF: low force unbinding path; HF: high force unbinding path. For the column ‘immobilization method’, the terminology X (Y) Z means: molecule X is attached to Z mediated by enzyme Y.

Protein handles Handle A:Handle B	Sizes (kDa)	Dissociation force (pN)	Immobilization method	Handle position (N/C)	References
Cohesin:dockerin I	15.4/8.3	122 ± 18.5	NHS-PEG5000-Mal/Cys	C:C	Stahl et al. (2012)
Cohesin:dockerin III	21.6/26.2	606 ± 54	NHS-PEG5000-Mal/Cys	N:C	Schoeler et al. (2015)
		111 ± 30 (LF)	NHS-PEG5000-Mal/CoA (SFP) ybbR	C:C	Schoeler et al. (2015)
		597 ± 67 (HF)	NHS-PEG5000-Mal/CoA (SFP) ybbR		
NiNTA:HIS6	0.2/0.8	153 ± 57	Gold-Cys	n.a.	Verbelen et al. (2007)
Avidin:biotin	66-69/0.2	160 ± 20	Biotinylated BSA	n.a.	Florin et al. (1994)
StrepTagII:streptavidin	1.1/52.8	253 ± 20	BSA/NHS-biotin	n.a.	Wong et al. (1999)
Streptavidin:biotin	52.8/0.2	200	Biotinylated BSA	n.a.	Rico and Moy (2007)
Calmodulin:CBP	16.7/1.1	16.5 ± 1.8	Pulldown via NI-NTA	n.a.	Junker and Rief (2009)
StrepTagII:mono-streptactin	1.1/58.4	116	NHS-PEG5000-Mal/Cys	C:C	Baumann et al. (2015)
		46	NHS-PEG5000-Mal/CoA (SFP) ybbR	N:C	
Anti-GCN4 sFv:GCN4(7P14P)	26.7/4.0	70	NHS-PEG5000-Mal/Cys	N:C	Morfill et al. (2007)
Anti-digoxigenin:digoxigenin	170/0.4	40	NHS-PEG6000	n.a.	Neuert et al. (2006)

AFM-SMFS, however, requires covalent immobilization of the two binding partners to the cantilever and surface, respectively, in order to avoid clogging of the molecules on the cantilever tip. Site-specific (i.e., residue specific) conjugation methods provide strict control over the pulling geometry and result in higher accuracy, precision and reproducibility, compared to conjugation methods resulting in a multiplicity of possible linkage sites (e.g., amine-targeting). Fig. 2 provides an overview of established surface chemistry strategies.

Another advantage of our modular system is the ability to use one construct (i.e., fingerprints with immobilization site) in all desired biochemical or biophysical assays, since immobilization relies on a PEG derivative, which is orthogonal to conventional specific pull down methods. It is compatible with a wide range of binding assays like Western Blotting, ITC, SPR, and ELISA.

The Ni-NTA:HIS$_6$-tag interaction can be used as force probe as well. This interaction has been employed as an adhesion sensor by probing a cell surface containing His-tagged protein. Since the His-tag is only located at one of the protein's termini, the insertion direction of the protein as well as it's position can be detected (Alsteens et al., 2013; Dupres et al., 2009; Pfreundschuh et al., 2015). This technique is especially useful since the His-tag can be used as a protein purification tag and simultaneously provides a single-molecule force handle.

5.1. Cysteines

Cysteines are relatively rare in proteins, making them attractive as a point mutation residue. The thiol side chain of cysteine is nucleophilic, and will spontaneously react with maleimide leaving groups at neutral pH. It can be used to site-specifically attach proteins to PEG coated surfaces for receptor–ligand AFM-SMFS. Alternatively, engineered cysteines can also be used as oligomerization sites to create disulfide-linked polyproteins, as was done for green fluorescent protein (GFP) (Dietz and Rief, 2006). However, cysteine/thiol-based protein conjugation has some drawbacks, including the tendency of cysteine-modified proteins to multimerize and ultimately aggregate over time, and incompatibility with proteins dis-

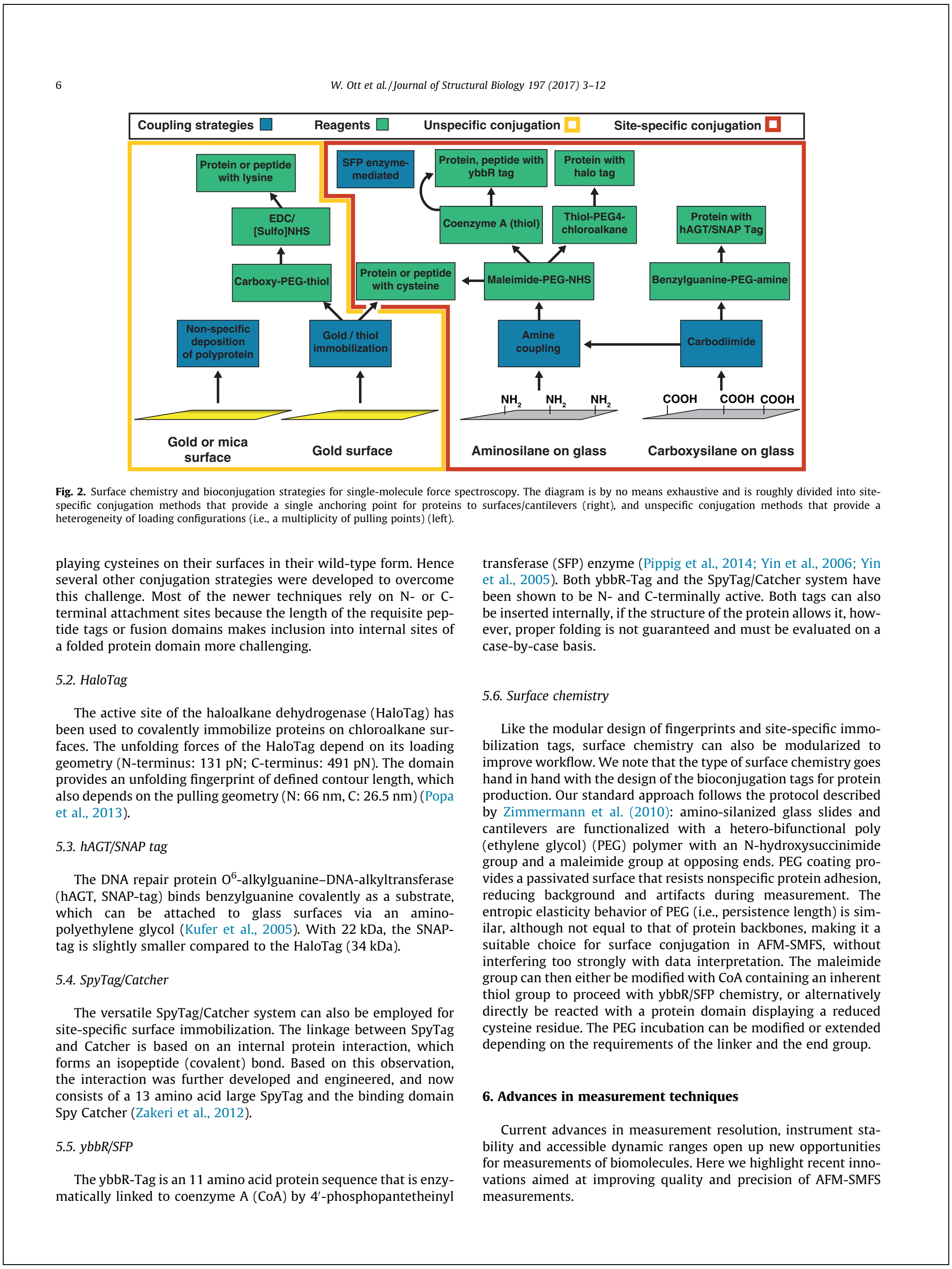


Fig. 2. Surface chemistry and bioconjugation strategies for single-molecule force spectroscopy. The diagram is by no means exhaustive and is roughly divided into site-specific conjugation methods that provide a single anchoring point for proteins to surfaces/cantilevers (right), and unspecific conjugation methods that provide a heterogeneity of loading configurations (i.e., a multiplicity of pulling points) (left).

playing cysteines on their surfaces in their wild-type form. Hence several other conjugation strategies were developed to overcome this challenge. Most of the newer techniques rely on N- or C-terminal attachment sites because the length of the requisite peptide tags or fusion domains makes inclusion into internal sites of a folded protein domain more challenging.

5.2. HaloTag

The active site of the haloalkane dehydrogenase (HaloTag) has been used to covalently immobilize proteins on chloroalkane surfaces. The unfolding forces of the HaloTag depend on its loading geometry (N-terminus: 131 pN; C-terminus: 491 pN). The domain provides an unfolding fingerprint of defined contour length, which also depends on the pulling geometry (N: 66 nm, C: 26.5 nm) (Popa et al., 2013).

5.3. hAGT/SNAP tag

The DNA repair protein O^6-alkylguanine–DNA-alkyltransferase (hAGT, SNAP-tag) binds benzylguanine covalently as a substrate, which can be attached to glass surfaces via an amino-polyethylene glycol (Kufer et al., 2005). With 22 kDa, the SNAP-tag is slightly smaller compared to the HaloTag (34 kDa).

5.4. SpyTag/Catcher

The versatile SpyTag/Catcher system can also be employed for site-specific surface immobilization. The linkage between SpyTag and Catcher is based on an internal protein interaction, which forms an isopeptide (covalent) bond. Based on this observation, the interaction was further developed and engineered, and now consists of a 13 amino acid large SpyTag and the binding domain Spy Catcher (Zakeri et al., 2012).

5.5. ybbR/SFP

The ybbR-Tag is an 11 amino acid protein sequence that is enzymatically linked to coenzyme A (CoA) by 4′-phosphopantetheinyl transferase (SFP) enzyme (Pippig et al., 2014; Yin et al., 2006; Yin et al., 2005). Both ybbR-Tag and the SpyTag/Catcher system have been shown to be N- and C-terminally active. Both tags can also be inserted internally, if the structure of the protein allows it, however, proper folding is not guaranteed and must be evaluated on a case-by-case basis.

5.6. Surface chemistry

Like the modular design of fingerprints and site-specific immobilization tags, surface chemistry can also be modularized to improve workflow. We note that the type of surface chemistry goes hand in hand with the design of the bioconjugation tags for protein production. Our standard approach follows the protocol described by Zimmermann et al. (2010): amino-silanized glass slides and cantilevers are functionalized with a hetero-bifunctional poly (ethylene glycol) (PEG) polymer with an N-hydroxysuccinimide group and a maleimide group at opposing ends. PEG coating provides a passivated surface that resists nonspecific protein adhesion, reducing background and artifacts during measurement. The entropic elasticity behavior of PEG (i.e., persistence length) is similar, although not equal to that of protein backbones, making it a suitable choice for surface conjugation in AFM-SMFS, without interfering too strongly with data interpretation. The maleimide group can then either be modified with CoA containing an inherent thiol group to proceed with ybbR/SFP chemistry, or alternatively directly be reacted with a protein domain displaying a reduced cysteine residue. The PEG incubation can be modified or extended depending on the requirements of the linker and the end group.

6. Advances in measurement techniques

Current advances in measurement resolution, instrument stability and accessible dynamic ranges open up new opportunities for measurements of biomolecules. Here we highlight recent innovations aimed at improving quality and precision of AFM-SMFS measurements.

6.1. Improved time resolution

In general, the timescales relevant for protein un-/folding and the corresponding timescale for thermally induced crossing of energy barriers are not fully detectable by common SMFS techniques, which typically resolve slower than 50 µs. Early on, the importance of developing high-speed AFM imaging and force spectroscopy through miniaturization of cantilevers with high resonance frequencies and low viscous drag coefficients was appreciated (Viani et al., 1999a,b). Nonetheless, only recent studies were able to overcome timescale limitations to observe, for example, extraordinarily slow protein misfolding transitions (~0.5 ms) using optical tweezers (Yu et al., 2015). Furthermore, advanced statistical methods extended optical tweezers SMFS time resolution to the ~10 µs range (Žoldák et al., 2013), and optimization of AFM cantilevers for SMFS has pushed the limit toward resolution on the microsecond timescale (Edwards et al., 2015). These developments allow experimentally accessible ranges to approach the lower limits of fast folding transition dynamics (Chung et al., 2012; Schuler and Hofmann, 2013), resolving short-lived intermediate states and yielding important insights into other fast conformational dynamics.

6.2. Bridging the timescale gap to steered molecular dynamics simulations

Recently, experimental measurements were brought into proximity (Dong and Sahin, 2011; He et al., 2012; Schoeler et al., 2015) or even overlap (Rico et al., 2013) with all atom steered molecular dynamics (SMD) simulations. Depending on the size and thus complexity of the simulated system, it has so far been possible to achieve SMD simulation timescales in the nanosecond to mid-microsecond range (Freddolino et al., 2008; Heymann and Grubmüller, 2001; Lee et al., 2009). Rico *et al.* developed a high speed force spectroscopy AFM based on an Ando-type high speed imaging AFM (Ando et al., 2001), with a high resonance frequency (600 kHz) miniature multilayer piezoelectric actuator (calibrated before each experiment and run in open loop mode), and a short cantilever with a high resonance frequency (550 kHz in liquid), and low viscous damping. This system was used to record protein unfolding data at extremely high speeds. To reduce hydrodynamic drag, the sample surface was tilted against the direction of the movement. With these improvements and data acquisition in the megahertz range, they were able to record meaningful and interpretable data at pulling speeds of up to 4000 µm/s, which is about 2–3 orders of magnitude faster than conventional methods and starts overlapping with the range of SMD simulations (Rico et al., 2013). Despite these successes, care must be taken because underdamped or ‘ringing’ cantilevers like the ones used here are not in agreement with the basic assumptions of the traditional SMFS framework, but can be improved by custom cantilever optimization procedures at the cost of time resolution (Edwards et al., 2015).

6.3. Long-term stability and force precision

Sophisticated measurements of complex biological systems or single molecules often require extraordinarily stable low-drift instruments, capable of continuous long-term data acquisition to gain sufficient and reliable statistics. Active stabilization techniques were developed to enable routine long-term stability and Ångstrom scale precision at room temperature for optical trap setups: differential sample position was measured and regulated with two independently stabilized and MHz modulated lasers, backscattered on sample and probe, and recorded separately on a single photodiode using lock-in amplifiers (Walder et al., 2015). This method is deemed applicable to surface-based and dual-beam optical traps, magnetic tweezers, AFM setups and optical microscopy, including super-resolution techniques.

AFM cantilever long-term stability and force precision can be increased even further by partially removing the reflective gold coating from the cantilever to dramatically reduce cantilever bending caused by the bimetallic effect (Churnside et al., 2012). Stability and precision improvements, which still retain high measurement bandwidths, enable and improve on picoscale force and subnanoscale motion measurements of molecular properties and dynamics in various biological systems. These may include groundbreaking investigations like the observation of single RNA polymerase base pair stepping (Abbondanzieri et al., 2005; Zhou et al., 2013), base pair unwinding of helicases (Cheng et al., 2011) and prion misfolding pathways (Yu et al., 2015, 2012). More details on long-term stability measurements and force precision are covered in the recent review of Edwards and Perkins (2016).

6.4. Mapping molecular recognition events: multiparametric imaging modes

The idea of mapping molecular recognition by simultaneously measuring surface topography and force–extension data (‘force volume mapping’ or ‘affinity imaging’) was introduced early (Hinterdorfer et al., 1996; Ludwig et al., 1997), and refined to remarkable temporal and spatial resolution. While these molecular recognition imaging techniques turned out to be a valuable tool for detecting and locating specific binding sites on surfaces, their development into dynamic recognition force imaging (Hinterdorfer and Dufrêne, 2006; Raab et al., 1999; Zhang et al., 2014) greatly increased temporal and spatial resolution, while still yielding information about surface elasticity and adhesion, as well as identifying biomolecules at the same time.

Multiparametric imaging modes can simultaneously detect physical properties of the surface and forces exerted on specific biomolecular binding sites. The AFM cantilever oscillates with amplitudes around 100 nm at sub- or low kilohertz frequencies to measure force–distance data, and simultaneously records image topography and other surface properties at sub- or low hertz line-scanning frequencies. The recorded force and topography data is collected orders of magnitude faster compared to force volume mapping methods, yielding imaging speeds comparable to conventional AFM imaging methods (Alsteens et al., 2012; Pfreundschuh et al., 2014). Another benefit of this method is that a large range of loading rates for receptor–ligand dissociation events can be probed in a single experiment, due to the largely varying cantilever tip velocities. Recently, this method was applied to gain nm-scale resolution imaging data of a G protein-coupled receptor (PAR1) in proteoliposomes while characterizing their ligand-binding energy landscape (Alsteens et al., 2015) from loading rates ranging between 1e3 and 1e6 pN/s, already two orders of magnitude higher than conventional force–distance based SMFS. Another recent study demonstrates the ability of this technique to distinguish two different binding events on opposite sides of engineered PAR1 by their unbinding force, and thereby determine their orientation within the lipid bilayer (Pfreundschuh et al., 2015).

6.5. Lateral force sensors

A slightly different approach developed a T-shaped cantilever (Dong et al., 2009; Dong and Sahin, 2011) to drive it at its flexural resonance frequency (~9 kHz) and record force data from cantilever torsion, resulting in a lateral laser deflection signal that was acquired while imaging the sample in conventional tapping mode. Due to the cantilever’s high torsional resonance (~115 kHz), unbinding dynamics could be measured at the microsecond timescale and at extraordinarily high loading rates of up to nearly 1e9 pN/s (Dong and Sahin, 2011), about four orders of magnitude faster than conventional SMFS. Force curves and therefore unbinding events and their corresponding force values could be mapped with high spatial and temporal resolution, while providing AFM images that were simultaneously recorded as surface topography. Mechanical elasticity properties of the substrate were also detected in the phase signal.

7. Theory and data analysis

7.1. The data analysis problem

Technical advances greatly increasing the throughput of AFM-SMFS measurements have made automated data analysis protocols an essential requirement. In practice, researchers face the problem of extracting meaningful single molecule signal from large datasets that contain an abundance of unusable data. The use of well-defined fingerprint domains with known unfolding patterns facilitates this procedure greatly. To avoid tedious and time-consuming manual sorting of thousands of data traces, and potential introduction of bias into the data analysis procedure, algorithms which identify the fingerprint unfolding length increments and classify the data correspondingly have been developed and implemented with success (Bosshart et al., 2012; Jobst et al., 2015; Kuhn et al., 2005; Puchner et al., 2008).

7.2. Polymer elasticity models and contour length transformations

Single molecule force measurements generally only gain access to a protein's extension under a given force. The stochastic nature of domain unfolding or complex dissociation under force as well as the non-linear elastic behavior of the polymer backbone chain makes analysis in force-extension space difficult. The same unfolding event is observed over a range of different positions in force–extension curves for multiple measurement cycles as shown in Fig. 3B i.

From a physicist's point of view, mechanical stretching of an unfolded protein domain is described by polymer elasticity models such as the worm-like chain (WLC) (Bustamante et al., 1994), the freely jointed chain (FJC) (Ortiz and Hadziioannou, 1999), or the freely rotating chain (FRC) model (Livadaru et al., 2003). These models contain the free contour length L of the polymer, including surface tethers and unfolded protein backbone, as a parameter. The free contour length is simply the length of the polypeptide along the contour of the biopolymer chain, given a specific folding state (e.g., Fig. 3A). Under a set of physically relevant constraints (L, x, $F > 0$, $x < L$), these elasticity models provide one-to-one mappings from force–extension space into force-contour length space. The models can be solved for the contour length parameter (Jobst et al., 2013; Puchner et al., 2008), yielding an expression for the contour length as a function of force and extension $L(F,x)$. This function can be used to transform force–extension traces from constant speed or force clamp/ramp experiments into contour length space (Fig. 3B ii). The calculated contour length then can be binned (Fig. 3B iii), aligned, and subsequently averaged to precisely locate energy barriers (Fig. 3B iv) along a protein's unfolding pathway, and to classify data sets based on unfolding patterns. This idea was first proposed by Puchner et al. (2008) and has been successfully applied in multiple AFM-SMFS studies (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2014; Stahl et al., 2012; Thoma et al., 2015).

7.3. Worm-like chains, freely-rotating chains and beyond

The WLC model accurately describes a protein's stretching response for forces up to approximately 150 pN. While many protein unfolding or dissociation events take place well within this force regime, some interactions like titin Ig domain unfolding (Rief et al., 1997a), cohesin unfolding (Valbuena et al., 2009), disso-

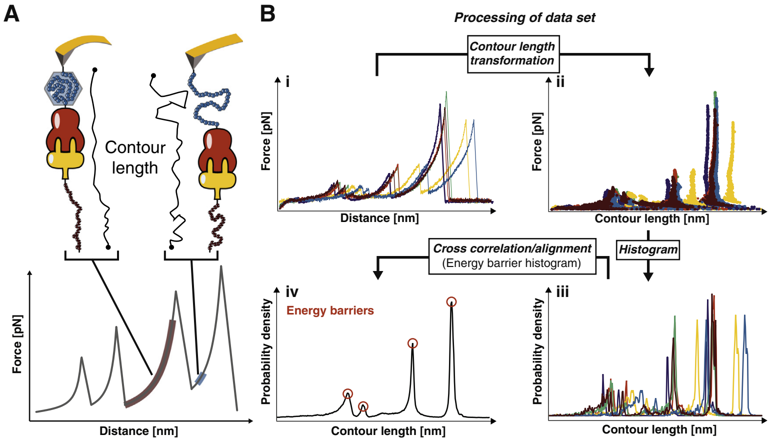


Fig. 3. Assembly of contour length histograms for screening AFM-SMFS datasets. (A) Force–extension traces are transformed into contour length space using an appropriate polymer elasticity model. The choice of the model depends on the force range. (B) Following transformation, the data (i) are plotted in force-contour length space (ii). Force and contour length thresholds are applied and the data are histogrammed (projected onto contour length axis) with an appropriate bin width, i.e., nanometer scale, to obtain the diagram in (iii). Each trace analyzed this way can be searched for a specific contour length increment (distance between two peaks in the probability density vs. contour length plot) corresponding to one of the fingerprints. To obtain a master histogram describing all the observed increments in a dataset, individual histograms reflecting a specific unfolding pathway are aligned by cross-correlation and offsetting along the contour-length axis (iv).

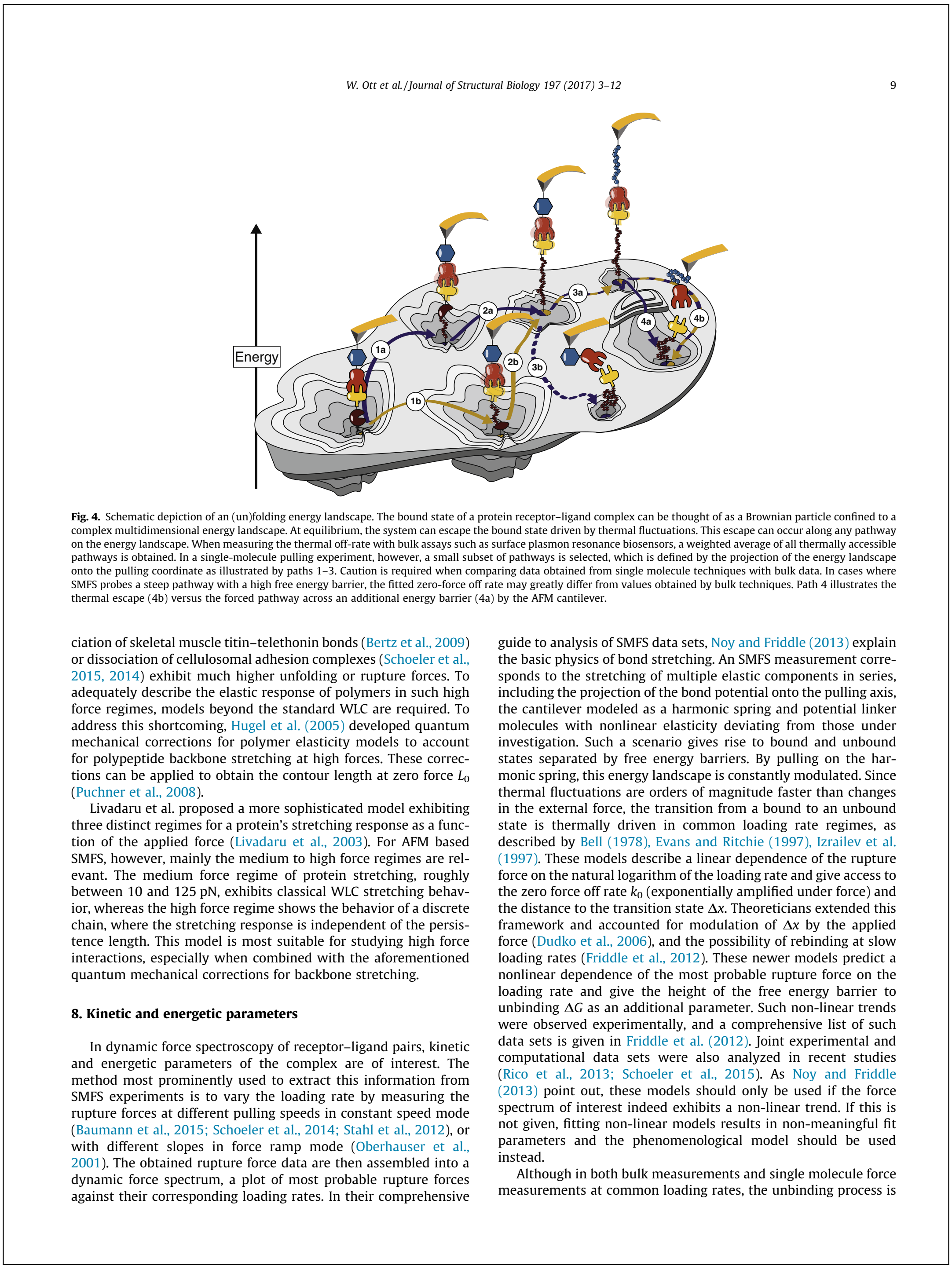


Fig. 4. Schematic depiction of an (un)folding energy landscape. The bound state of a protein receptor–ligand complex can be thought of as a Brownian particle confined to a complex multidimensional energy landscape. At equilibrium, the system can escape the bound state driven by thermal fluctuations. This escape can occur along any pathway on the energy landscape. When measuring the thermal off-rate with bulk assays such as surface plasmon resonance biosensors, a weighted average of all thermally accessible pathways is obtained. In a single-molecule pulling experiment, however, a small subset of pathways is selected, which is defined by the projection of the energy landscape onto the pulling coordinate as illustrated by paths 1–3. Caution is required when comparing data obtained from single molecule techniques with bulk data. In cases where SMFS probes a steep pathway with a high free energy barrier, the fitted zero-force off rate may greatly differ from values obtained by bulk techniques. Path 4 illustrates the thermal escape (4b) versus the forced pathway across an additional energy barrier (4a) by the AFM cantilever.

ciation of skeletal muscle titin–telethonin bonds (Bertz et al., 2009) or dissociation of cellulosomal adhesion complexes (Schoeler et al., 2015, 2014) exhibit much higher unfolding or rupture forces. To adequately describe the elastic response of polymers in such high force regimes, models beyond the standard WLC are required. To address this shortcoming, Hugel et al. (2005) developed quantum mechanical corrections for polymer elasticity models to account for polypeptide backbone stretching at high forces. These corrections can be applied to obtain the contour length at zero force L_0 (Puchner et al., 2008).

Livadaru et al. proposed a more sophisticated model exhibiting three distinct regimes for a protein's stretching response as a function of the applied force (Livadaru et al., 2003). For AFM based SMFS, however, mainly the medium to high force regimes are relevant. The medium force regime of protein stretching, roughly between 10 and 125 pN, exhibits classical WLC stretching behavior, whereas the high force regime shows the behavior of a discrete chain, where the stretching response is independent of the persistence length. This model is most suitable for studying high force interactions, especially when combined with the aforementioned quantum mechanical corrections for backbone stretching.

8. Kinetic and energetic parameters

In dynamic force spectroscopy of receptor–ligand pairs, kinetic and energetic parameters of the complex are of interest. The method most prominently used to extract this information from SMFS experiments is to vary the loading rate by measuring the rupture forces at different pulling speeds in constant speed mode (Baumann et al., 2015; Schoeler et al., 2014; Stahl et al., 2012), or with different slopes in force ramp mode (Oberhauser et al., 2001). The obtained rupture force data are then assembled into a dynamic force spectrum, a plot of most probable rupture forces against their corresponding loading rates. In their comprehensive guide to analysis of SMFS data sets, Noy and Friddle (2013) explain the basic physics of bond stretching. An SMFS measurement corresponds to the stretching of multiple elastic components in series, including the projection of the bond potential onto the pulling axis, the cantilever modeled as a harmonic spring and potential linker molecules with nonlinear elasticity deviating from those under investigation. Such a scenario gives rise to bound and unbound states separated by free energy barriers. By pulling on the harmonic spring, this energy landscape is constantly modulated. Since thermal fluctuations are orders of magnitude faster than changes in the external force, the transition from a bound to an unbound state is thermally driven in common loading rate regimes, as described by Bell (1978), Evans and Ritchie (1997), Izrailev et al. (1997). These models describe a linear dependence of the rupture force on the natural logarithm of the loading rate and give access to the zero force off rate k_0 (exponentially amplified under force) and the distance to the transition state Δx. Theoreticians extended this framework and accounted for modulation of Δx by the applied force (Dudko et al., 2006), and the possibility of rebinding at slow loading rates (Friddle et al., 2012). These newer models predict a nonlinear dependence of the most probable rupture force on the loading rate and give the height of the free energy barrier to unbinding ΔG as an additional parameter. Such non-linear trends were observed experimentally, and a comprehensive list of such data sets is given in Friddle et al. (2012). Joint experimental and computational data sets were also analyzed in recent studies (Rico et al., 2013; Schoeler et al., 2015). As Noy and Friddle (2013) point out, these models should only be used if the force spectrum of interest indeed exhibits a non-linear trend. If this is not given, fitting non-linear models results in non-meaningful fit parameters and the phenomenological model should be used instead.

Although in both bulk measurements and single molecule force measurements at common loading rates, the unbinding process is thermally driven, caution is required when comparing their data. While at unbiased equilibrium, all thermally accessible pathways from the bound state are sampled and the off rate is consequently measured as a weighted average, single molecule force measurements select only a small subset of these pathways due to the defined pulling geometry, as illustrated by paths 1–3 in Fig. 4. In cases where the energy landscape is highly asymmetric and the pulling experiment probes a steep pathway, the off rates obtained from single molecule *vs.* bulk measurements might differ greatly (see Fig. 4, paths 4a vs 4b).

9. Summary and outlook

We highlighted recent advances in experimental design, molecular design, sample preparation, measurement and analysis methods for AFM-SMFS on polyproteins and receptor–ligand complexes. We summarized site-specific bioconjugation strategies to obtain well-defined pulling geometries for improved reliability and reproducibility of experiments. We also highlighted receptor–ligand pairs with high mechanical strength (e.g., cohesin–dockerin), and their application as specific pulling handles in AFM-SMFS for improving experimental throughput and curve yield. Finally, we touched on recent innovations in positional control and cantilever microfabrication for improving time and force resolution and stability of the measurement, on emerging techniques for mapping force responses of surfaces to their topologies, and we discussed theoretical considerations for analyzing large numbers of curves.

In the future, there remain several technical challenges that need to be addressed. One of the limitations of AFM is that it covers a relatively high force range, yet there exist a multitude of biological interactions in the low-force regime that are of interest. Further technical advances in instrument design, cantilever fabrication, and feedback control might further improve force resolution and thereby enable such experiments. A second area for improvement involves sample throughput and parallel screening. With the development of more elaborate, sophisticated and well defined surface immobilization strategies and protein handles, significant gains in throughput can be envisioned. Innovations of the chemistry in combination with efficient data analysis protocols and state of the art instrumentation may pave the way towards in depth study of complex, multi-domain protein systems.

These advances in experimental design and throughput would greatly benefit from refined theoretical frameworks that account for parameters such as cantilever stiffness and ringing whilst maintaining analytical tractability. Consequently, with improved methodology we anticipate the community will be able to address an even wider range of questions about mechanical adaptations of proteins and protein complexes in the future.

Acknowledgments

The authors gratefully acknowledge funding from an advanced Grant to HEG from the European Research Council (Cellufuel Grant 294438), and from a Society in Science – Branco Weiss Fellowship administered by ETH Zürich, Switzerland to MAN. We are also thankful for the financial support of the Deutsche Forschungsgemeinschaft SFB 863.

References

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., Block, S.M., 2005. Direct observation of base-pair stepping by RNA polymerase. Nature 438, 460–465. http://dx.doi.org/10.1038/nature04268.

Albrecht, C., Blank, K., Lalic-Multhaler, M., Hirler, S., Mai, T., Gilbert, I., Schiffmann, S., Bayer, T., Clausen-Schaumann, H., Gaub, H.E., 2003. DNA: a programmable force sensor. Science 301, 367–370. http://dx.doi.org/10.1126/science.1084713.

Alsteens, D., Garcia, M.C., Lipke, P.N., Dufrêne, Y.F., 2010. Force-induced formation and propagation of adhesion nanodomains in living fungal cells. Proc. Natl. Acad. Sci. USA 107, 20744–20749. http://dx.doi.org/10.1073/pnas.1013893107.

Alsteens, D., Dupres, V., Yunus, S., Latgé, J.-P., Heinisch, J.J., Dufrêne, Y.F., 2012. High-resolution imaging of chemical and biological sites on living cells using peak force tapping atomic force microscopy. Langmuir 28, 16738–16744. http://dx.doi.org/10.1021/la303891j.

Alsteens, D., Trabelsi, H., Soumillion, P., Dufrêne, Y.F., 2013. Multiparametric atomic force microscopy imaging of single bacteriophages extruding from living bacteria. Nat. Commun. 4, 1–7. http://dx.doi.org/10.1038/ncomms3926.

Alsteens, D., Pfreundschuh, M., Zhang, C., Spoerri, P.M., Coughlin, S.R., Kobilka, B.K., Müller, D.J., 2015. Imaging G protein-coupled receptors while quantifying their ligand-binding free-energy landscape. Nat. Methods 12, 845–851. http://dx.doi.org/10.1038/nmeth.3479.

Ando, T., Kodera, N., Takai, E., Maruyama, D., Saito, K., Toda, A., 2001. A high-speed atomic force microscope for studying biological macromolecules. Proc. Natl. Acad. Sci. USA 98, 12468–12472. http://dx.doi.org/10.1073/pnas.211400898.

Baumann, F., Bauer, M.S., Milles, L.F., Alexandrovich, A., Gaub, H.E., Pippig, D.A., 2015. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. Nat. Nanotechnol. 1–7. http://dx.doi.org/10.1038/nnano.2015.231.

Beedle, A.E.M., Lezamiz, A., Stirnemann, G., Garcia-Manyes, S., 2015a. The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. Nat. Commun. 6, 1–9. http://dx.doi.org/10.1038/ncomms8894.

Beedle, A.E.M., Williams, A., Relat-Goberna, J., Garcia-Manyes, S., 2015b. Mechanobiology – chemical origin of membrane mechanical resistance and force-dependent signaling. Curr. Opin. Chem. Biol. 29, 87–93. http://dx.doi.org/10.1016/j.cbpa.2015.09.019.

Bell, G.I., 1978. Models for the specific adhesion of cells to cells. Science 200, 618–627. http://dx.doi.org/10.1126/science.347575.

Bertz, M., Wilmanns, M., Rief, M., 2009. The titin–telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. Proc. Natl. Acad. Sci. USA 106, 13307–13310. http://dx.doi.org/10.1073/pnas.0902312106.

Block, S.M., Goldstein, L., Schnapp, B.J., 1990. Bead movement by single kinesin molecules studied with optical tweezers. Nature 348, 348–352. http://dx.doi.org/10.1038/348348a0.

Bosshart, P.D., Frederix, P.L.T.M., Engel, A., 2012. Reference-free alignment and sorting of single-molecule force spectroscopy data. Biophys. J. 102, 2202–2211. http://dx.doi.org/10.1016/j.bpj.2012.03.027.

Brockwell, D.J., Paci, E., Zinober, R.C., Beddard, G.S., Olmsted, P.D., Smith, D.A., Perham, R.N., Radford, S.E., 2003. Pulling geometry defines the mechanical resistance of a beta-sheet protein. Nat. Struct. Biol. 10, 731–737. http://dx.doi.org/10.1038/nsb968.

Bu, T., Wang, H.-C.E., Li, H., 2012. Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1. Langmuir 28, 12319–12325. http://dx.doi.org/10.1021/la301940g.

Bustamante, C., Marko, J.F., Siggia, E.D., Smith, S., 1994. Entropic elasticity of lambda-phage DNA. Science 265, 1599–1600. http://dx.doi.org/10.1126/science.8079175.

Cao, Y., Li, Y.D., Li, H., 2011. Enhancing the mechanical stability of proteins through a cocktail approach. Biophys. J. 100, 1794–1799. http://dx.doi.org/10.1016/j.bpj.2011.02.030.

Carrion-Vazquez, M., Oberhauser, A.F., Fowler, S.B., Marszalek, P.E., Broedel, S.E., Clarke, J., Fernandez, J.M., 1999. Mechanical and chemical unfolding of a single protein: a comparison. Proc. Natl. Acad. Sci. USA 96, 3694–3699. http://dx.doi.org/10.1073/pnas.96.7.3694.

Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P.E., Oberhauser, A.F., Fernandez, J.M., 2003. The mechanical stability of ubiquitin is linkage dependent. Nat. Struct. Biol. 10, 738–743. http://dx.doi.org/10.1038/nsb965.

Carvalho, F.A., Martins, I.C., Santos, N.C., 2013. Atomic force microscopy and force spectroscopy on the assessment of protein folding and functionality. Arch. Biochem. Biophys. 531, 116–127. http://dx.doi.org/10.1016/j.abb.2012.11.007.

Casuso, I., Rico, F., Scheuring, S., 2011. Biological AFM: where we come from – where we are – where we may go. J. Mol. Recogn. 24, 406–413. http://dx.doi.org/10.1002/jmr.1081.

Cheng, W., Arunajadai, S.G., Moffitt, J.R., Tinoco, I.J., Bustamante, C., 2011. Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. Science 333, 1746–1749. http://dx.doi.org/10.1126/science.1206023.

Chung, H.S., McHale, K., Louis, J.M., Eaton, W.A., 2012. Single-molecule fluorescence experiments determine protein folding transition path times. Science 335, 981–984. http://dx.doi.org/10.1126/science.1215768.

Churnside, A.B., Sullan, R.M.A., Nguyen, D.M., Case, S.O., Bull, M.S., King, G.M., Perkins, T.T., 2012. Routine and timely sub-picoNewton force stability and precision for biological applications of atomic force microscopy. Nano Lett. 12, 3557–3561. http://dx.doi.org/10.1021/nl301166w.

del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J.M., Sheetz, M.P., 2009. Stretching single Talin rod molecules activates vinculin binding. Science 323, 638–641. http://dx.doi.org/10.1126/science.1162912.

Dietz, H., Rief, M., 2006. Protein structure by mechanical triangulation. Proc. Natl. Acad. Sci. USA 103, 1244–1247. http://dx.doi.org/10.1073/pnas.0509217103.

Dietz, H., Berkemeier, F., Bertz, M., Rief, M., 2006. Anisotropic deformation response of single protein molecules. Proc. Natl. Acad. Sci. USA 103, 12724–12728. http://dx.doi.org/10.1073/pnas.0602995103.

Dong, M., Sahin, O., 2011. A nanomechanical interface to rapid single-molecule interactions. Nat. Commun. 2, 1–6. http://dx.doi.org/10.1038/ncomms1246.

Dong, M., Husale, S., Sahin, O., 2009. Determination of protein structural flexibility by microsecond force spectroscopy. Nat. Nanotechnol. 4, 514–517. http://dx.doi.org/10.1038/nnano.2009.156.

Drake, B., Prater, C.B., Weisenhorn, A.L., Gould, S.A.C., Albrecht, T.R., Quate, C.F., Cannell, D.S., Hansma, H.G., Hansma, P.K., 1989. Imaging crystals, polymers, and processes in water with the atomic force microscope. Science 243, 1386–1389. http://dx.doi.org/10.1126/science.2928794.

Dudko, O.K., Hummer, G., Szabo, A., 2006. Intrinsic rates and activation free energies from single-molecule pulling experiments. Phys. Rev. Lett. 96, 108101. http://dx.doi.org/10.1103/PhysRevLett. 96.108101.

Dupres, V., Alsteens, D., Wilk, S., Hansen, B., Heinisch, J.J., Dufrêne, Y.F., 2009. The yeast Wsc1 cell surface sensor behaves like a nanospring in vivo. Nat. Chem. Biol. 5, 857–862. http://dx.doi.org/10.1038/nchembio.220.

Edwards, D.T., Perkins, T.T., 2016. Optimizing force spectroscopy by modifying commercial cantilevers: improved stability, precision, and temporal resolution. J. Struct. Biol. 1–13. http://dx.doi.org/10.1016/j.jsb.2016.01.009.

Edwards, D.T., Faulk, J.K., Sanders, A.W., Bull, M.S., Walder, R., LeBlanc, M.-A., Sousa, M.C., Perkins, T.T., 2015. Optimizing 1-µs-resolution single-molecule force spectroscopy on a commercial atomic force microscope. Nano Lett. 15, 7091–7098. http://dx.doi.org/10.1021/acs.nanolett.5b03166.

Evans, E., Ritchie, K., 1997. Dynamic strength of molecular adhesion bonds. Biophys. J. 72, 1541–1555. http://dx.doi.org/10.1016/S0006-3495(97)78802-7.

Florin, E.L., Moy, V.T., Gaub, H.E., 1994. Adhesion forces between individual ligand–receptor pairs. Science 264, 415–417. http://dx.doi.org/10.1126/science.8153628.

Florin, E.L., Rief, M., Lehmann, H., Ludwig, M., Dornmair, C., Moy, V.T., Gaub, H.E., 1995. Sensing specific molecular-interactions with the atomic-force microscope. Biosens. Bioelectron. 10, 895–901. http://dx.doi.org/10.1016/0956-5663(95)99227-C.

Freddolino, P.L., Liu, F., Gruebele, M., Schulten, K., 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys. J. 94, L75–L77. http://dx.doi.org/10.1529/biophysj.108.131565.

Friddle, R.W., Noy, A., De Yoreo, J.J., 2012. Interpreting the widespread nonlinear force spectra of intermolecular bonds. Proc. Natl. Acad. Sci. USA 109, 13573–13578. http://dx.doi.org/10.1073/pnas.1202946109.

Geisler, M., Xiao, S., Puchner, E.M., Gräter, F., Hugel, T., 2010. Controlling the structure of proteins at surfaces. J. Am. Chem. Soc. 132, 17277–17281. http://dx.doi.org/10.1021/ja107212z.

He, C., Genchev, G.Z., Lu, H., Li, H., 2012. Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations. J. Am. Chem. Soc. 134, 10428–10435. http://dx.doi.org/10.1021/ja3003205.

Helenius, J., Heisenberg, C.-P., Gaub, H.E., Müller, D.J., 2008. Single-cell force spectroscopy. J. Cell Sci. 121, 1785–1791. http://dx.doi.org/10.1242/jcs.030999.

Heymann, B., Grubmüller, H., 2001. Molecular dynamics force probe simulations of antibody/antigen unbinding: entropic control and nonadditivity of unbinding forces. Biophys. J. 81, 1295–1313. http://dx.doi.org/10.1016/S0006-3495(01)75787-6.

Hinterdorfer, P., Dufrêne, Y.F., 2006. Detection and localization of single molecular recognition events using atomic force microscopy. Nat. Methods 3, 347–355. http://dx.doi.org/10.1038/nmeth871.

Hinterdorfer, P., Baumgartner, W., Gruber, H.J., Schilcher, K., Schindler, H., 1996. Detection and localization of individual antibody–antigen recognition events by atomic force microscopy. Proc. Natl. Acad. Sci. USA 93, 3477–3481. http://dx.doi.org/10.1073/pnas.93.8.3477.

Hoffmann, T., Dougan, L., 2012. Single molecule force spectroscopy using polyproteins. Chem. Soc. Rev. 41, 4781–4796. http://dx.doi.org/10.1039/c2cs35033e.

Hu, X., Li, H., 2014. Force spectroscopy studies on protein–ligand interactions: a single protein mechanics perspective. FEBS Lett. 588, 3613–3620. http://dx.doi.org/10.1016/j.febslet.2014.04.009.

Hugel, T., Rief, M., Seitz, M., Gaub, H.E., Netz, R.R., 2005. Highly stretched single polymers: atomic-force-microscope experiments versus ab-initio theory. Phys. Rev. Lett. http://dx.doi.org/10.1103/PhysRevLett.94.048301.

Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., Schulten, K., 1997. Molecular dynamics study of unbinding of the avidin–biotin complex. Biophys. J. 72, 1568–1581. http://dx.doi.org/10.1016/S0006-3495(97)78804-0.

Janovjak, H., Struckmeier, J., Hubain, M., Kedrov, A., Kessler, M., Müller, D.J., 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. Structure 12, 871–879. http://dx.doi.org/10.1016/j.str.2004.03.016.

Jobst, M.A., Schoeler, C., Malinowska, K., Nash, M.A., 2013. Investigating receptor–ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. J. Vis. Exp., e50950 http://dx.doi.org/10.3791/50950.

Jobst, M.A., Milles, L.F., Schoeler, C., Ott, W., Fried, D.B., Bayer, E.A., Gaub, H.E., Nash, M.A., 2015. Resolving dual binding conformations of cellulosome cohesin–dockerin complexes using single-molecule force spectroscopy. Elife 4, 1031. http://dx.doi.org/10.7554/eLife.9.

Junker, J.P., Rief, M., 2009. Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. Proc. Natl. Acad. Sci. USA 106, 14361–14366. http://dx.doi.org/10.1073/pnas.0904654106.

Kim, M., Wang, C.-C., Benedetti, F., Rabbi, M., Bennett, V., Marszalek, P.E., 2011. Nanomechanics of streptavidin hubs for molecular materials. Adv. Mater. 23, 5684–5688. http://dx.doi.org/10.1002/adma.201103316.

Kocun, M., Grandbois, M., Cuccia, L.A., 2011. Single molecule atomic force microscopy and force spectroscopy of chitosan. Colloids Surf. B Biointerfaces 82, 470–476. http://dx.doi.org/10.1016/j.colsurfb.2010.10.004.

Kufer, S.K., Dietz, H., Albrecht, C., Blank, K., Kardinal, A., Rief, M., Gaub, H.E., 2005. Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. Eur. Biophys. J. 35, 72–78. http://dx.doi.org/10.1007/s00249-005-0010-1.

Kuhn, M., Janovjak, H., Hubain, M., Müller, D.J., 2005. Automated alignment and pattern recognition of single-molecule force spectroscopy data. J. Microscopy 218, 125–132. http://dx.doi.org/10.1111/j.1365-2818.2005.01478.x.

Lee, C.-K., Chrisey, L.A., Colton, R.J., 1994a. Direct measurement of the forces between complementary strands of DNA. Science. http://dx.doi.org/10.1126/science.7973628.

Lee, G.U., Kidwell, D.A., Colton, R.J., 1994b. Sensing discrete streptavidin–biotin interactions with atomic force microscopy. Langmuir. http://dx.doi.org/10.1021/la00014a003.

Lee, C.-K., Wang, Y.-M., Huang, L.-S., Lin, S., 2007. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction. Micron 38, 446–461. http://dx.doi.org/10.1016/j.micron.2006.06.014.

Lee, E.H., Hsin, J., Sotomayor, M., Comellas, G., Schulten, K., 2009. Discovery through the computational microscope. Structure 17, 1295–1306. http://dx.doi.org/10.1016/j.str.2009.09.001.

Li, H., Cao, Y., 2010. Protein mechanics: from single molecules to functional biomaterials. Acc. Chem. Res. 43, 1331–1341. http://dx.doi.org/10.1021/ar100057a.

Livadaru, L., Netz, R.R., Kreuzer, H.J., 2003. Stretching response of discrete semiflexible polymers. Macromolecules 36, 3732–3744. http://dx.doi.org/10.1021/ma020751g.

Ludwig, M., Dettmann, W., Gaub, H.E., 1997. Atomic force microscope imaging contrast based on molecular recognition. Biophys. J. 72, 445–448. http://dx.doi.org/10.1016/S0006-3495(97)78685-5.

Marszalek, P.E., Dufrêne, Y.F., 2012. Stretching single polysaccharides and proteins using atomic force microscopy. Chem. Soc. Rev. 41, 3523–3534. http://dx.doi.org/10.1039/c2cs15329g.

Mitchell, G., Lamontagne, C.-A., Lebel, R., Grandbois, M., Malouin, F., 2007. Single-molecule dynamic force spectroscopy of the fibronectin–heparin interaction. Biochem. Biophys. Res. Commun. 364, 595–600. http://dx.doi.org/10.1016/j.bbrc.2007.10.034.

Morfill, J., Blank, K., Zahnd, C., Luginbühl, B., Kühner, F., Gottschalk, K.-E., Plückthun, A., Gaub, H.E., 2007. Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy. Biophys. J. 93, 3583–3590. http://dx.doi.org/10.1529/biophysj.107.112532.

Moy, V.T., Florin, E.L., Gaub, H.E., 1994. Intermolecular forces and energies between ligands and receptors. Science 266, 257–259. http://dx.doi.org/10.1126/science.7939660.

Müller, D.J., 2008. AFM: a nanotool in membrane biology. Biochemistry 47, 7986–7998. http://dx.doi.org/10.1021/bi800753x.

Müller, D.J., Dufrêne, Y.F., 2008. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. Nat. Nanotechnol. 3, 261–269. http://dx.doi.org/10.1038/nnano.2008.100.

Müller, D.J., Engel, A., 2007. Atomic force microscopy and spectroscopy of native membrane proteins. Nat. Protocols 2, 2191–2197. http://dx.doi.org/10.1038/nprot.2007.309.

Müller, D.J., Büldt, G., Engel, A., 1995. Force-induced conformational change of bacteriorhodopsin. J. Mol. Biol. 249, 239–243. http://dx.doi.org/10.1006/jmbi.1995.0292.

Müller, D.J., Helenius, J., Alsteens, D., Dufrêne, Y.F., 2009. Force probing surfaces of living cells to molecular resolution. Nat. Chem. Biol. 5, 383–390. http://dx.doi.org/10.1038/nchembio.181.

Neuert, G., Albrecht, C., Pamir, E., Gaub, H.E., 2006. Dynamic force spectroscopy of the digoxigenin–antibody complex. FEBS Lett. 580, 505–509. http://dx.doi.org/10.1016/j.febslet.2005.12.052.

Neuman, K.C., Nagy, A., 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat. Methods 5, 491–505. http://dx.doi.org/10.1038/nmeth.1218.

Noy, A., 2011. Force spectroscopy 101: how to design, perform, and analyze an AFM-based single molecule force spectroscopy experiment. Curr. Opin. Chem. Biol. 15, 710–718. http://dx.doi.org/10.1016/j.cbpa.2011.07.020.

Noy, A., Friddle, R.W., 2013. Practical single molecule force spectroscopy: how to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope. Methods 60, 142–150. http://dx.doi.org/10.1016/j.ymeth.2013.03.014.

Oberhauser, A.F., Marszalek, P.E., Erickson, H.P., Fernandez, J.M., 1998. The molecular elasticity of the extracellular matrix protein tenascin. Nature 393, 181–185. http://dx.doi.org/10.1038/30270.

Oberhauser, A.F., Hansma, P.K., Carrion-Vazquez, M., Fernandez, J.M., 2001. Stepwise unfolding of titin under force-clamp atomic force microscopy. Proc. Natl. Acad. Sci. USA 98, 468–472. http://dx.doi.org/10.1073/pnas.021321798.

Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H.E., Müller, D.J., 2000. Unfolding pathways of individual bacteriorhodopsins. Science 288, 143–146. http://dx.doi.org/10.1126/science.288.5463.143.

Ortiz, C., Hadziioannou, G., 1999. Entropic elasticity of single polymer chains of poly (methacrylic acid) measured by atomic force microscopy. Macromolecules 32, 780–787. http://dx.doi.org/10.1021/ma981245n.

Otten, M., Ott, W., Jobst, M.A., Milles, L.F., Verdorfer, T., Pippig, D.A., Nash, M.A., Gaub, H.E., 2014. From genes to protein mechanics on a chip. Nat. Methods 11, 1127–1130. http://dx.doi.org/10.1038/nmeth.3099.

Pfreundschuh, M., Alsteens, D., Hilbert, M., Steinmetz, M.O., Müller, D.J., 2014. Localizing chemical groups while imaging single native proteins by high-resolution atomic force microscopy. Nano Lett. 14, 2957–2964. http://dx.doi.org/10.1021/nl5012905.

Pfreundschuh, M., Alsteens, D., Wieneke, R., Zhang, C., Coughlin, S.R., Tampé, R., Kobilka, B.K., Müller, D.J., 2015. Identifying and quantifying two ligand-binding sites while imaging native human membrane receptors by AFM. Nat. Commun. 6, 1–7. http://dx.doi.org/10.1038/ncomms9857.

Pippig, D.A., Baumann, F., Strackharn, M., Aschenbrenner, D., Gaub, H.E., 2014. Protein–DNA chimeras for nano assembly. ACS Nano 8, 6551–6555. http://dx.doi.org/10.1021/nn501644w.

Popa, I., Berkovich, R., Alegre-Cebollada, J., Badilla, C.L., Rivas-Pardo, J.A., Taniguchi, Y., Kawakami, M., Fernandez, J.M., 2013. Nanomechanics of HaloTag tethers. J. Am. Chem. Soc. 135, 12762–12771. http://dx.doi.org/10.1021/ja4056382.

Preiner, J., Kodera, N., Tang, J., Ebner, A., Brameshuber, M., Blaas, D., Gelbmann, N., Gruber, H.J., Ando, T., Hinterdorfer, P., 2014. IgGs are made for walking on bacterial and viral surfaces. Nat. Commun. 5, 1–8. http://dx.doi.org/10.1038/ncomms5394.

Puchner, E.M., Franzen, G., Gautel, M., Gaub, H.E., 2008. Comparing proteins by their unfolding pattern. Biophys. J. 95, 426–434. http://dx.doi.org/10.1529/biophysj.108.129999.

Raab, A., Hansma, H.G., Badt, D., Smith-Gill, S.J., 1999. Antibody recognition imaging by force microscopy. Nature. http://dx.doi.org/10.1038/12898.

Radmacher, M., Tillmann, R.W., Fritz, M., Gaub, H.E., 1992. From molecules to cells – imaging soft samples with the atomic force microscope. Science 257, 1900–1905. http://dx.doi.org/10.1126/science.1411505.

Radmacher, M., Fritz, M., Kacher, C.M., Cleveland, J.P., Hansma, P.K., 1996. Measuring the viscoelastic properties of human platelets with the atomic force microscope. Biophys. J. 70, 556–567. http://dx.doi.org/10.1016/S0006-3495(96)79602-9.

Rico, F., Moy, V.T., 2007. Energy landscape roughness of the streptavidin–biotin interaction. J. Mol. Recogn. 20, 495–501. http://dx.doi.org/10.1002/jmr.841.

Rico, F., Gonzalez, L., Casuso, I., Puig-Vidal, M., Scheuring, S., 2013. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. Science 342, 741–743. http://dx.doi.org/10.1126/science.1239764.

Rief, M., Grubmüller, H., 2002. Force spectroscopy of single biomolecules. Chemphyschem 3, 255–261. http://dx.doi.org/10.1002/1439-7641(20020315)3:3<255::AID-CPHC255>3.0.CO;2-M.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M., Gaub, H.E., 1997a. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276, 1109–1112. http://dx.doi.org/10.1126/science.276.5315.1109.

Rief, M., Oesterhelt, F., Heymann, B., Gaub, H.E., 1997b. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. Science 275, 1295–1297. http://dx.doi.org/10.1126/science.275.5304.1295.

Schmidt, S.W., Filippov, P., Kersch, A., Beyer, M.K., Clausen-Schaumann, H., 2012. Single-molecule force-clamp experiments reveal kinetics of mechanically activated silyl ester hydrolysis. ACS Nano 6, 1314–1321. http://dx.doi.org/10.1021/nn204111w.

Schoeler, C., Malinowska, K.H., Bernardi, R.C., Milles, L.F., Jobst, M.A., Durner, E., Ott, W., Fried, D.B., Bayer, E.A., Schulten, K., Gaub, H.E., Nash, M.A., 2014. Ultrastable cellulosome–adhesion complex tightens under load. Nat. Commun. 5, 1–8. http://dx.doi.org/10.1038/ncomms6635.

Schoeler, C., Bernardi, R.C., Malinowska, K.H., Durner, E., Ott, W., Bayer, E.A., Schulten, K., Nash, M.A., Gaub, H.E., 2015. Mapping mechanical force propagation through biomolecular complexes. Nano Lett. 15, 7370–7376. http://dx.doi.org/10.1021/acs.nanolett.5b02727.

Schuler, B., Hofmann, H., 2013. Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. Curr. Opin. Struct. Biol. 23, 36–47. http://dx.doi.org/10.1016/j.sbi.2012.10.008.

Schwesinger, F., Ros, R., Strunz, T., Anselmetti, D., Guntherodt, H.-J., Honegger, A., Jermutus, L., Tiefenauer, L., Plückthun, A., 2000. Unbinding forces of single antibody–antigen complexes correlate with their thermal dissociation rates. Proc. Natl. Acad. Sci. USA 97, 9972–9977. http://dx.doi.org/10.1073/pnas.97.18.9972.

Sirbuly, D.J., Friddle, R.W., Villanueva, J., Huang, Q., 2015. Nanomechanical force transducers for biomolecular and intracellular measurements: is there room to shrink and why do it? Rep. Prog. Phys. 1–22. http://dx.doi.org/10.1088/0034-4885/78/2/024101.

Smith, S.B., Finzi, L., Bustamante, C., 1992. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. Science 258, 1122–1126. http://dx.doi.org/10.1126/science.1439819.

Stahl, S.W., Nash, M.A., Fried, D.B., Slutzki, M., Barak, Y., Bayer, E.A., Gaub, H.E., 2012. Single-molecule dissection of the high-affinity cohesin–dockerin complex. Proc. Natl. Acad. Sci. USA 109, 20431–20436. http://dx.doi.org/10.1073/pnas.1211929109.

Sułkowska, J.I., Cieplak, M., 2007. Mechanical stretching of proteins—a theoretical survey of the protein data bank. J. Phys. Condens. Matter 19, 283201. http://dx.doi.org/10.1088/0953-8984/19/28/283201.

Svoboda, K., Schmidt, C.F., Schnapp, B.J., Block, S.M., 1993. Direct observation of kinesin stepping by optical trapping interferometry. Nature 365, 721–727. http://dx.doi.org/10.1038/365721a0.

Thoma, J., Burmann, B.M., Hiller, S., Müller, D.J., 2015. Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. Nat. Struct. Mol. Biol. 22, 795–802. http://dx.doi.org/10.1038/nsmb.3087.

Tsukasaki, Y., Kitamura, K., Shimizu, K., Iwane, A.H., Takai, Y., Yanagida, T., 2007. Role of multiple bonds between the single cell adhesion molecules, nectin and cadherin, revealed by high sensitive force measurements. J. Mol. Biol. 367, 996–1006. http://dx.doi.org/10.1016/j.jmb.2006.12.022.

Valbuena, A., Oroz, J., Hervas, R., Manuel Vera, A., Rodriguez, D., Menendez, M., Sulkowska, J.I., Cieplak, M., Carrion-Vazquez, M., 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc. Natl. Acad. Sci. USA 106, 13791–13796. http://dx.doi.org/10.1073/pnas.0813093106.

Verbelen, C., Gruber, H.J., Dufrêne, Y.F., 2007. The NTA-His6 bond is strong enough for AFM single-molecular recognition studies. J. Mol. Recogn. 20, 490–494. http://dx.doi.org/10.1002/jmr.833.

Viani, M.B., Schaffer, T.E., Chand, A., Rief, M., Gaub, H.E., Hansma, P.K., 1999a. Small cantilevers for force spectroscopy of single molecules. J. Appl. Phys. 86, 2258–2262. http://dx.doi.org/10.1063/1.371039.

Viani, M.B., Schaffer, T.E., Paloczi, G.T., Pietrasanta, L.I., Smith, B.L., Thompson, J.B., Richter, M., Rief, M., Gaub, H.E., Plaxco, K.W., Cleland, A.N., Hansma, H.G., Hansma, P.K., 1999b. Fast imaging and fast force spectroscopy of single biopolymers with a new atomic force microscope designed for small cantilevers. Rev. Sci. Instrum. 70, 4300–4303. http://dx.doi.org/10.1063/1.1150069.

Walder, R., Paik, D.H., Bull, M.S., Sauer, C., Perkins, T.T., 2015. Ultrastable measurement platform: sub-nm drift over hours in 3D at room temperature. Opt. Express 23, 16554–16564. http://dx.doi.org/10.1364/OE.23.016554.

Wildling, L., Rankl, C., Haselgrübler, T., Gruber, H.J., Holy, M., Newman, A.H., Zou, M.-F., Zhu, R., Freissmuth, M., Sitte, H.H., Hinterdorfer, P., 2012. Probing binding pocket of serotonin transporter by single molecular force spectroscopy on living cells. J. Biol. Chem. 287, 105–113. http://dx.doi.org/10.1074/jbc.M111.304873.

Wong, J., Chilkoti, A., Moy, V.T., 1999. Direct force measurements of the streptavidin–biotin interaction. Biomol. Eng. 16, 45–55. http://dx.doi.org/10.1016/S1050-3862(99)00035-2.

Woodside, M.T., Block, S.M., 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu. Rev. Biophys. 43, 19–39. http://dx.doi.org/10.1146/annurev-biophys-051013-022754.

Yin, J., Straight, P.D., McLoughlin, S.M., Zhou, Z., Lin, A.J., Golan, D.E., Kelleher, N.L., Kolter, R., Walsh, C.T., 2005. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc. Natl. Acad. Sci. USA 102, 15815–15820. http://dx.doi.org/10.1073/pnas.0507705102.

Yin, J., Lin, A.J., Golan, D.E., Walsh, C.T., 2006. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. Nat. Protocols 1, 280–285. http://dx.doi.org/10.1038/nprot.2006.43.

Yu, H., Liu, X., Neupane, K., Gupta, A.N., Brigley, A.M., Solanki, A., Sosova, I., Woodside, M.T., 2012. Direct observation of multiple misfolding pathways in a single prion protein molecule. Proc. Natl. Acad. Sci. USA 109, 5283–5288. http://dx.doi.org/10.1073/pnas.1107736109.

Yu, H., Dee, D.R., Liu, X., Brigley, A.M., Sosova, I., Woodside, M.T., 2015. Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. Proc. Natl. Acad. Sci. USA 112, 8308–8313. http://dx.doi.org/10.1073/pnas.1419197112.

Yuan, C., Chen, A., Kolb, P., Moy, V.T., 2000. Energy landscape of streptavidin–biotin complexes measured by atomic force microscopy. Biochemistry 39, 10219–10223. http://dx.doi.org/10.1021/bi992715o.

Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., Howarth, M., 2012. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc. Natl. Acad. Sci. USA 109, 4347–4348. http://dx.doi.org/10.1073/pnas.1115485109.

Zhang, S., Aslan, H., Besenbacher, F., Dong, M., 2014. Quantitative biomolecular imaging by dynamic nanomechanical mapping. Chem. Soc. Rev. 43, 7412–7429. http://dx.doi.org/10.1039/C4CS00176A.

Zhou, J., Schweikhard, V., Block, S.M., 2013. Single-molecule studies of RNAPII elongation. Biochem. Biophys. Acta 1829, 29–38. http://dx.doi.org/10.1016/j.bbagrm.2012.08.006.

Zimmermann, J.L., Nicolaus, T., Neuert, G., Blank, K., 2010. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. Nat. Protocols 5, 975–985. http://dx.doi.org/10.1038/nprot.2010.49.

Žoldák, G., Stigler, J., Pelz, B., Li, H., 2013. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. Proc. Natl. Acad. Sci. USA 110, 18156–18161. http://dx.doi.org/10.1073/pnas.1311495110.

Part II

Results

5

Peer Reviewed Research Articles

5.1 Publication P2: High Force Protein Complexes

The focus of this publication was the characterization of the mechanically highly stable receptor-ligand protein complex Coh-Doc type 3 (*R.f.*). A combination of *in silico* and *in vitro* experiments highlights the extraordinary force resistance of this complex and finds evidence for a force-activated stabilization mechanism in the interplay of the protein binding partners. My contribution to this project was the co-development of a highly stable, fully automated AFM SMFS instrument and the development of an advanced control software to maintain stability by drift correction and enable high statistical power through long-term measurements. Both taken together enabled experiments with routinely achieving tens to hundreds of thousands of raw data traces. These then typically yielded quantities of interpretable data in the rage from hundreds to thousands of individual curves.

Ultrastable Cellulosome-adhesion Complex Tightens Under Load

Constantin Schoeler[†], Klara H. Malinowska[†], Rafael C. Bernardi,
Lukas F. Milles, Markus A. Jobst, Ellis Durner, Wolfgang Ott,
Daniel B. Fried, Edward A. Bayer, Klaus Schulten,
Hermann E. Gaub and Michael A. Nash

[†]*these authors contributed equally to this publication*

published in

Nature Communications, 5, 1-8, (2014)

nature COMMUNICATIONS

ARTICLE

Received 25 Jun 2014 | Accepted 22 Oct 2014 | Published 8 Dec 2014

DOI: 10.1038/ncomms6635

OPEN

Ultrastable cellulosome-adhesion complex tightens under load

Constantin Schoeler[1,*], Klara H. Malinowska[1,*], Rafael C. Bernardi[2], Lukas F. Milles[1], Markus A. Jobst[1], Ellis Durner[1], Wolfgang Ott[1], Daniel B. Fried[3], Edward A. Bayer[3], Klaus Schulten[2,4], Hermann E. Gaub[1] & Michael A. Nash[1]

Challenging environments have guided nature in the development of ultrastable protein complexes. Specialized bacteria produce discrete multi-component protein networks called cellulosomes to effectively digest lignocellulosic biomass. While network assembly is enabled by protein interactions with commonplace affinities, we show that certain cellulosomal ligand–receptor interactions exhibit extreme resistance to applied force. Here, we characterize the ligand–receptor complex responsible for substrate anchoring in the *Ruminococcus flavefaciens* cellulosome using single-molecule force spectroscopy and steered molecular dynamics simulations. The complex withstands forces of 600–750 pN, making it one of the strongest bimolecular interactions reported, equivalent to half the mechanical strength of a covalent bond. Our findings demonstrate force activation and inter-domain stabilization of the complex, and suggest that certain network components serve as mechanical effectors for maintaining network integrity. This detailed understanding of cellulosomal network components may help in the development of biocatalysts for production of fuels and chemicals from renewable plant-derived biomass.

---

[1] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany. [2] Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3] Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel. [4] Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.A.N. (email: michael.nash@lmu.de).

Cellulosomes are protein networks designed by nature to degrade lignocellulosic biomass[1]. These networks comprise intricate assemblies of conserved subunits including catalytic domains, scaffold proteins, carbohydrate binding modules (CBMs), cohesins (Cohs), dockerins (Docs) and X-modules (XMods) of unknown function. Coh:Doc pairs form complexes with high affinity and specificity[2], and provide connectivity to a myriad of cellulosomal networks with varying Coh:Doc network topology[3–5]. The most intricate cellulosome known to date is produced by *Ruminococcus flavefaciens* (*R.f.*)[6,7] and contains several primary and secondary scaffolds along with over 220 Doc-bearing protein subunits[8].

The importance of cellulolytic enzymes for the production of renewable fuels and chemicals from biomass has highlighted an urgent need for improved fundamental understanding of how cellulosomal networks achieve their impressive catalytic activity[9]. Two of the mechanisms known to increase the catalytic activity of cellulosomes are proximity and targeting effects[10]. Proximity refers to the high local concentration of enzymes afforded by incorporation into nanoscale networks, while targeting refers to specific binding of cellulosomes to substrates. Protein scaffolds and CBM domains are both critical in this context as they mediate interactions between comparatively large bacterial cells and cellulose particles. As many cellulosomal habitats (for example, cow rumen) exhibit strong flow gradients, shear forces will accordingly stress bridging scaffold components mechanically *in vivo*. Protein modules located at stressed positions within these networks should therefore be preselected for high mechanostability. However, thus far very few studies on the mechanics of carbohydrate-active proteins or cellulosomal network components have been reported[11].

In the present study we sought to identify cellulosomal network junctions with maximal mechanical stability. We chose an XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose in the rumen. The complex links the *R. flavefaciens* cell wall to the cellulose substrate via two CBM domains located at the N-terminus of the CttA scaffold, as shown in Fig. 1a. The crystal structure of the complex solved by X-ray crystallography[12] is shown in Fig. 1b. XMod-Doc tandem dyads such as this one are a common feature in cellulosomal networks. Bulk biochemical assays on XMod-Docs have demonstrated that XMods improve Doc solubility and increase biochemical affinity of Doc:Coh complex formation[13]. Crystallographic studies conducted on XMod-Doc:Coh complexes have revealed direct contacts between XMods and their adjacent Docs[12,14]. In addition, many XMods (for example, PDB 2B59, 1EHX, 3PDD) have high β-strand content and fold with N- and C-termini at opposite ends of the molecule, suggestive of robust mechanical clamp motifs at work[15,16]. These observations all suggest a mechanical role for XMods. Here we perform AFM single-molecule force spectroscopy experiments and steered molecular dynamics simulations to understand the mechanostability of the XMod-Doc:Coh cellulosomal ligand–receptor complex. We conclude that the high mechanostability we observe originates from molecular mechanisms, including stabilization of Doc by the adjacent XMod domain and catch bond behaviour that causes the complex to increase in contact area on application of force.

Results and Discussion

Single-molecule experiments. We performed single-molecule force spectroscopy (SMFS) experiments with an atomic force miscroscope (AFM) to probe the mechanical dissociation of XMod-Doc:Coh. Xylanase (Xyn) and CBM fusion domains on the XMod-Doc and Coh modules, respectively, provided identifiable unfolding patterns permitting screening of large data sets of force-distance curves[17–19]. Engineered cysteines and/or peptide tags on the CBM and Xyn marker domains were used to covalently immobilize the binding partners in a site-specific manner to an AFM cantilever or cover glass via poly(ethylene glycol) (PEG) linkers. The pulling configuration with Coh-CBM immobilized on the cantilever is referred to as configuration I, as shown in Fig. 1c. The reverse configuration with Coh-CBM on the cover glass is referred to as configuration II. In a typical

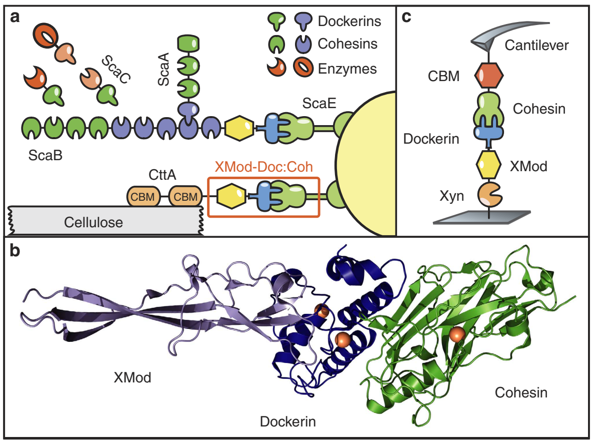


Figure 1 | System overview. (**a**) Schematic of selected components of the *R. flavefaciens* cellulosome. The investigated XMod–Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose is highlighted in orange. (**b**) Crystal structure of the XMod-Doc:Coh complex. Ca^{2+} ions are shown as orange spheres. (**c**) Depiction of experimental pulling configuration I, with Coh-CBM attached to the cantilever tip and Xyn–XMod–Doc attached to the glass surface.

experimental run we collected about 50,000 force extension traces from a single cantilever. We note that the molecules immobilized on the cantilever and glass surfaces were stable over thousands of pulling cycles.

We sorted the data by first searching for contour length increments that matched our specific xylanase and CBM fingerprint domains. After identifying these specific traces (Fig. 2a), we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history. To assign protein subdomains to the observed unfolding patterns, we transformed the data into contour length space using a freely rotating chain model with quantum mechanical corrections for peptide backbone stretching (QM-FRC, Supplementary Note 1, Supplementary Fig. 1)[20,21]. The fit parameter-free QM-FRC model describes protein stretching at forces >200 pN more accurately than the commonly used worm-like chain (WLC) model[20,22]. The resulting contour length histogram is shown in Fig. 2b. Peak-to-peak distances in the histogram represent contour length increments of unfolded protein domains. Assuming a length per stretched amino acid of 0.365 nm and accounting for the folded length of each subdomain, we compared the observed increments to the polypeptide lengths of individual subdomains of the Xyn-XMod-Doc and Coh-CBM fusion proteins. Details on contour length estimates and domain assignments are shown in Supplementary Table 1.

Unfolding patterns in configuration I showed PEG stretching followed by a three-peaked Xyn fingerprint (Fig. 1a, top trace, green), which added 90 nm of contour length to the system. Xyn unfolding was followed by CBM unfolding at ~ 150 pN with 55 nm of contour length added. Finally, the XMod-Doc:Coh complex dissociated at an ultra-high rupture force of ~ 600 pN. The loading rate dependence of the final rupture event for curves of subtype 1 is plotted in Fig. 2c (blue). The measured complex rupture force distributions are shown in Supplementary Fig. 2.

Less frequently (35–40% of traces) we observed a two-step dissociation process wherein the XMod unfolded before Doc:Coh rupture as shown in Fig. 2a (middle trace, orange). In these cases, the final dissociation exhibited a much lower rupture force (~ 300 pN) than the preceding XMod unfolding peak, indicating the strengthening effect of XMod was lost, and XMod was no longer able to protect the complex from dissociation at high force. The loading rate dependency of Doc:Coh rupture occurring immediately following XMod unfolding is shown in Fig. 2c (grey).

In configuration II (Fig. 2a, bottom trace), with the Xyn-XMod-Doc attached to the cantilever, the xylanase fingerprint was lost after the first few force extension traces acquired in the data set. This indicated the Xyn domain did not refold within the timescale of the experiment once unfolded, consistent with prior work[17,18]. CBM and XMod unfolding events were observed repeatedly throughout the series of acquired force traces in both configurations I and II, indicating these domains were able to refold while attached to the cantilever over the course of the experiment.

We employed the Bell-Evans model[23] (Supplementary Note 2) to analyse the final rupture of the complex through the effective distance to the transition state (Δx) and the natural off-rate (k_{off}). The fits to the model yielded values of $\Delta x = 0.13$ nm and $k_{\mathrm{off}} = 7.3 \times 10^{-7}\,\mathrm{s}^{-1}$ for an intact XMod, and $\Delta x = 0.19$ nm and $k_{\mathrm{off}} = 4.7 \times 10^{-4}\,\mathrm{s}^{-1}$ for the 'shielded' rupture following XMod unfolding (Fig. 2c). These values indicate that the distance to the transition state is increased following XMod unfolding, reflecting an overall softening of the binding interface. Distances to the transition state observed for other ligand–receptor pairs are typically on the order of ~ 0.7 nm (ref. 17). The extremely short Δx of 0.13 nm observed here suggests that mechanical unbinding for this complex is highly coordinated. We further analysed the unfolding of XMod in the Bell-Evans picture and found values of $\Delta x = 0.15$ and $k_{\mathrm{off}} = 2.6 \times 10^{-6}\,\mathrm{s}^{-1}$. The loading

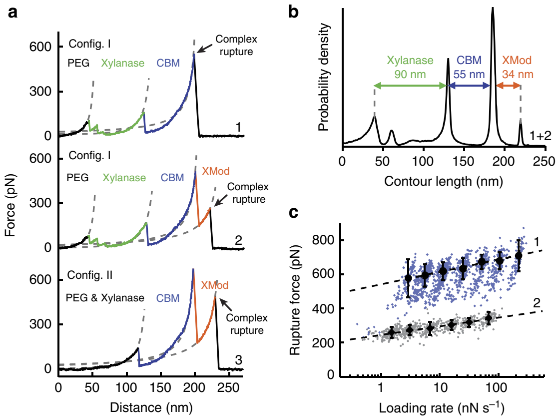


Figure 2 | Experimental SMFS unfolding traces. (**a**) Unfolding fingerprints from pulling configuration I (curves 1 & 2) and configuration II (curve 3). The QM-FRC model (dashed lines) was used to estimate the contour lengths of the unfolded modules. (**b**) Contour length histogram obtained from 127 force extension traces (Config. I). The peak-to-peak increments correspond to Xyn, CBM and XMod amino-acid sequence lengths. (**c**) Dynamic force spectra for the final Doc:Coh complex rupture peaks obtained from 2,122 force-extension traces. The blue points show Doc:Coh ruptures that occurred with an intact XMod, while grey points show ruptures immediately following XMod unfolding. Black circles and diamonds represent the most probable rupture force/loading rate obtained by Gaussian fitting at each pulling speed. Error bars are ±1 s.d. Dashed lines are least square fits to the Bell-Evans model.

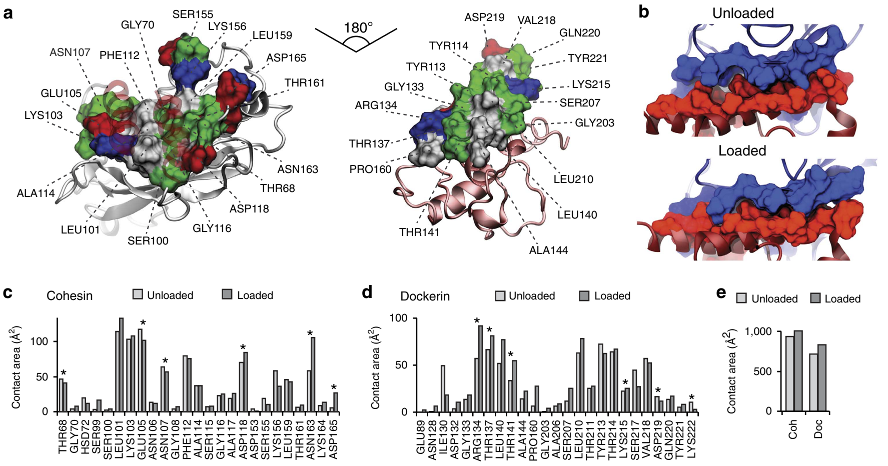


Figure 3 | Analysis of binding interface and catch bond mechanism from SMD. (**a**) Surface plots for the main interacting residues of Coh (left) and Doc (right). Hydrophobic residues are shown in grey, polar residues in green, and negative and positive residues in red and blue, respectively. Both Coh and Doc exhibit a hydrophobic patch in the centre of the binding surface that is surrounded by polar and charged residues. (**b**) Rearrangement of binding residues of Coh (blue) and Doc (red) under force. Following mechanical loading, an interdigitated complex is formed that resembles teeth of a zipper. (**c**,**d**) Surface contact area of interacting residues of Coh (c) and Doc (d) in the absence and presence of force. Residues forming prevalent hydrogen bonds are indicated with stars. (**e**) Total contact surface area of Coh and Doc in unloaded and loaded conformations.

rate dependence for this unfolding event is shown in Supplementary Fig. 3.

The exceptionally high rupture forces measured experimentally (Fig. 2) are hugely disproportionate to the XMod-Doc:Coh biochemical affinity, which at $K_D \sim 20$ nM (ref. 12) is comparable to typical antibody–antigen interactions. Antibody–antigen interactions, however, will rupture at only ~60 pN at similar loading rates[24], while bimolecular complexes found in muscle exposed to mechanical loading *in vivo* will rupture at ~140 pN (ref. 25). Trimeric titin–telethonin complexes also found in muscle exhibit unfolding forces around 700 pN (ref. 26), while Ig domains from cardiac titin will unfold at ~200 pN (ref. 27). The XMod-Doc:Coh ruptures reported here fell in a range from 600 to 750 pN at loading rates ranging from 10 to 100 nN s^{-1}. At around half the rupture force of a covalent gold-thiol bond[28], these bimolecular protein rupture forces are, to the best of our knowledge, among the highest of their kind ever reported. The covalent bonds in this system are primarily peptide bonds in the proteins and C-C and C-O bonds in the PEG linkers. These are significantly more mechanically stable than the quoted gold-thiol bond rupture force (~1.2 nN) (ref. 29) and fall in a rupture force range >2.5 nN at similar loading rates. Therefore, breakage of covalent linkages under our experimental conditions is highly unlikely. We note that the high mechanostability observed here is not the result of fusing the proteins to the CBM or Xyn domains. The covalent linkages and pulling geometry are consistent with the wild-type complex and its dissociation pathway. *In vivo*, the Coh is anchored to the peptidoglycan cell wall through its C-terminal sortase motif. The XMod–Doc is attached to the cellulose substrate through two N-terminal CBM domains. By pulling the XMod–Doc through an N-terminal Xyn fusion domain, and the Coh through a C-terminal CBM, we established an experimental pulling geometry that matches loading of the complex *in vivo*. This pulling geometry was also used in all simulations. The discontinuity between its commonplace biochemical affinity and remarkable resistance to applied force illustrates how this complex is primed for mechanical stability and highlights differences in the unbinding pathway between dissociation at equilibrium and dissociation induced mechanically along a defined pulling coordinate.

Steered molecular dynamics. To elucidate the molecular mechanisms at play that enable this extreme mechanostability, we carried out all-atom steered molecular dynamics (SMD) simulations. The Xyn and CBM domains were not modelled to keep the simulated system small and reduce the usage of computational resources. This approximation was reasonable as we have no indication that these domains significantly affect the XMod–Doc:Coh binding strength[30]. After equilibrating the crystal structure[12], the N-terminus of XMod–Doc was harmonically restrained while the C-terminus of Coh was pulled away at constant speed. The force applied to the harmonic pulling spring was stored at each time step. We tested pulling speeds of 0.25, 0.625 and 1.25 Å ns^{-1}, and note that the slowest simulated pulling speed was ~4,000 times faster than our fastest experimental pulling speed of 6.4 μm s^{-1}. This difference is considered not to affect the force profile, but it is known to account for the scale difference in force measured by SMD and AFM[31,32].

SMD results showed the force increased with distance until the complex ruptured for all simulations. At the slowest pulling speed of 0.25 Å ns^{-1} the rupture occurred at a peak force of ~900 pN, as shown in Supplementary Fig. 4 and Supplementary Movie 1. We analysed the progression and prevalence of hydrogen bonded contacts between the XMod–Doc and Coh domains to identify

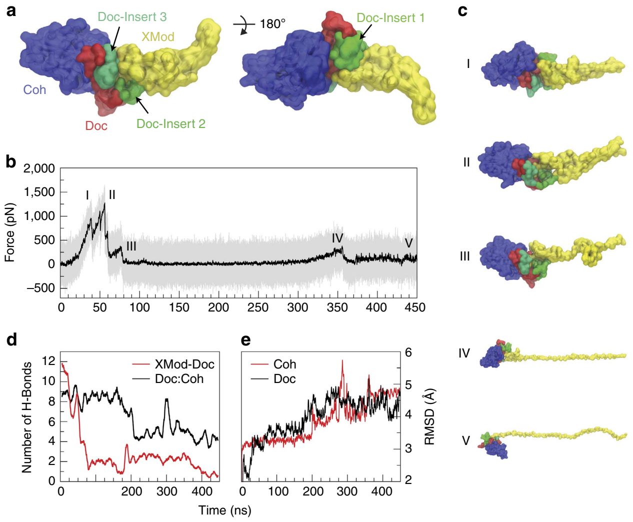


Figure 4 | SMD shows unfolding of XMod destabilizes Doc:Coh binding interface. XMod was unfolded by moving the harmonic restraint to the C terminus of XMod while the N terminus was moved at 0.625 Å ns^{-1}. (**a**) Surface representation of XMod-Doc:Coh complex with Doc insert sequences. Coh is shown in blue, Doc in red and green (inserts), and XMod in yellow. (**b**) Force time trace of XMod unfolding. The domain starts to unfold in several substeps starting at ~400 pN. Snapshots at different time steps are labelled I-V and are shown in (**c**). Steps IV and V are shown at smaller scale. (**d**) Average number of hydrogen bonds between Doc:Coh (black) and XMod-Doc (red). XMod-Doc contact is dominated by the insert sequences 1–3. (**e**) Root mean squared deviation (RMSD) of Doc (black) and Coh (red).

key residues in contact throughout the entire rupture process and particularly immediately before rupture. These residues are presented in Fig. 3a,c,d and Supplementary Figs 5,6. The simulation results clearly reproduced key hydrogen bonding contacts previously identified[12] as important for Doc:Coh recognition (Supplementary Fig. 5).

The main interacting residues are shown in Fig. 3a,b. Both Coh and Doc exhibit a binding interface consisting of a hydrophobic centre (grey) surrounded by a ring of polar (green) and charged residues (blue, positive; red, negative). This residue pattern suggests the hydrophilic side chains protect the interior hydrophobic core from attack by water molecules, compensating for the flat binding interface that lacks a deep pocket. The geometry suggests a penalty to unbinding that stabilizes the bound state. Further, we analysed the contact surface areas of interacting residues (Fig. 3b–e). The total contact area was found to increase due to rearrangement of the interacting residues when the complex is mechanically stressed, as shown in Fig. 3e and Supplementary Movie 2. Doc residues in the simulated binding interface clamped down on Coh residues upon mechanical loading, resulting in increased stability and decreased accessibility of water into the hydrophobic core of the bound complex (Fig. 3b). These results suggest that a catch bond mechanism is responsible for the remarkable stability[33] under force and provide a molecular mechanism which the XMod–Doc:Coh complex uses to summon mechanical strength when needed, while still allowing relatively fast assembly and disassembly of the complex at equilibrium. The residues that increase most in contact area (Fig. 3c,d) present promising candidates for future mutagenesis studies.

Among the 223 Doc sequences from *R. flavefaciens*, six subfamilies have been explicitly identified using bioinformatics approaches[8]. The XMod–Doc investigated here belongs to the 40-member Doc family 4a. A conserved feature of these Doc modules is the presence of three sequence inserts that interrupt the conserved duplicated F-hand motif Doc structure. In our system, these Doc sequence inserts make direct contacts with XMod in the crystallized complex (Fig. 1) and suggest an interaction between XMod and Doc that could potentially propagate to the Doc:Coh binding interface. To test this, an independent simulation was performed to unfold XMod (Fig. 4). The harmonic restraint was moved to the C-terminus of XMod so that force was applied from the N- to C-terminus of XMod only, while leaving Doc and Coh unrestrained. The results (Fig. 4b) showed XMod unfolded at forces slightly higher than but similar to the XMod–Doc:Coh complex rupture force determined from the standard simulation at the same pulling speed. This suggested XMod unfolding before Doc:Coh rupture was not probable, but could be observed on occasion due to the stochastic nature of domain unfolding. This was consistent with experiments where XMod unfolding was observed in ~35–40% of traces. Furthermore, analysis of the H-bonding between Doc and XMod (Fig. 4d, red) indicated loss of contact as XMod unfolded, dominated by contact loss between the three Doc insert sequences and XMod. Interestingly, XMod unfolding clearly led to a decrease in H-bonding between Doc and Coh at a later stage (~200 ns) well after XMod had lost most of its contact with Doc, even though no force was being applied across the Doc:Coh binding interface. This provided evidence for direct stabilization of the Doc:Coh binding interface by XMod.

As shown in Fig. 4e, the root mean squared deviation (RMSD) of Doc increased throughout the simulation as XMod unfolded. Coh RMSD remained stable until it started to lose H-bonds with Doc. Taken together this suggests that, as XMod unfolded, Coh and Doc became more mobile and lost interaction strength, potentially explaining the increase in Δx from 0.13 to 0.19 nm on unfolding of XMod in the experimental data sets. Apparently the XMod is able to directly stabilize the Doc:Coh interface, presumably through contact with Doc insert sequences that then propagate this stabilizing effect to the Doc:Coh binding interface.

In summary, we investigated an ultrastable XMod-Doc:Coh complex involved in bacterial adhesion to cellulose. While previously the role of XMod functioning in tandem XMod-Doc dyads was unclear[12,14], we show that XMod serves as a mechanical stabilizer and force-shielding effector subdomain in the ultrastable ligand–receptor complex. The Doc:Coh complex presented here exhibits one of the most mechanically robust protein–protein interactions reported thus far, and points towards new mechanically stable artificial multi-component biocatalysts for industrial applications, including production of second-generation biofuels.

Methods

Site-directed mutagenesis. Site-directed mutagenesis of *R. flavefaciens* strain FD1 chimeric cellulosomal proteins. A pET28a vector containing the previously cloned *R. flavefaciens* CohE from ScaE fused to cellulose-binding module 3a (CBM3a) from *C. thermocellum*, and a pET28a vector containing the previously cloned *R. flavefaciens* XMod-Doc from the CttA scaffoldin fused to the XynT6 xylanase from *Geobacillus stearothermophilus*[12] were subjected to QuikChange mutagenesis[34] to install the following mutations: A2C in the CBM and T129C in the xylanase, respectively.

For the construction of the native configuration of the CohE-CBM A2C fusion protein Gibson assembly[35] was used. For further analysis CohE-CBM A2C was modified with a QuikChange PCR[36] to replace the two cysteins (C2 and C63) in the protein with alanine and serine (C2A and C63S). All mutagenesis products were confirmed by DNA sequencing analysis.

The XynT6-XDoc T129C was constructed using the following primers:
5′-acaaggaaggtaagccaatggttaatgaatgcgatccagtgaaacgtgaac-3′
5′-gttcacgtttcactggatcgcattcattaaccattggcttaccttccttgt-3′

The CBM-CohE A2C was constructed using the following primers:
5′-ttaactttaagaaggagatataccatgtgcaatacaccggtatcaggcaatttgaag-3′
5′-cttcaaattgcctgataccggtgtattgcacatggtatatctccttcttaaagttaa-3′

The CohE-CBM C2A C63S was constructed using the following phosphorylated primers:
5′-ccgaatgccatggccaatacaccgg-3′
5′-cagaccttctggagtgaccatgctgc-3′

Expression and purification of Xyn-XMod-Doc. The T129C Xyn-XMod-Doc protein was expressed in *E. coli* BL21 cells in kanamycin-containing media that also contained 2 mM calcium chloride, overnight at 16 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with tris-buffered saline (TBS) buffer containing 20 mM imidazole and 2 mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was dialysed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at −20 °C. The concentrations of the protein stock solutions were determined to be ~5 mg ml^{-1} by absorption spectrophotometry.

Expression and purification of Coh-CBM. The Coh-CBM C2A, C63S fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicol containing ZYM-5052 media[37] overnight at 22 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with a polyacrylamide gravity flow column. The protein solution was concentrated with an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at −80 °C. The concentrations of the protein stock solutions were determined to be ~5 mg ml^{-1} by absorption spectrophotometry.

Sample preparation. In sample preparation and single-molecule measurements calcium supplemented TBS buffer (Ca-TBS) was used (25 mM TRIS, 72 mM NaCl, 1 mM $CaCl_2$, pH 7.2). Cantilevers and cover glasses were functionalized according to previously published protocols[18,38]. In brief, cantilevers and cover glasses were cleaned by UV-ozone treatment and piranha solution, respectively. Levers and glasses were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce surface amine groups. Amine groups on the cantilevers and cover glasses were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker in sodium borate buffer. Disulfide-linked dimers of the Xyn-XMod-Doc proteins were reduced for 2 h at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with Ca-TBS measurement buffer, centrifuged at 850 r.c.f. for 3 min, and the supernatant was collected with a micropipette. Reduced proteins were diluted with measurement buffer (1:3 (v/v) for cantilevers, and 1:1 (v/v) for cover glasses), and applied to PEGylated cantilevers and cover glasses for 1 h. Both cantilevers and cover glasses were then rinsed with Ca-TBS to remove unbound proteins and stored under Ca-TBS before force spectroscopy measurements. Site-specific immobilization of the Coh-CBM-ybbR fusion proteins to previously PEGylated cantilevers or coverglasses was carried out according to previously published protocols[39]. In brief, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH 7.2) for 1 h at room temperature. Levers or surfaces were then rinsed with Ca-TBS to remove unbound CoA. Coh-CBM-ybbR fusion proteins were then covalently linked to the CoA surfaces or levers by incubating with Sfp phosphopantetheinyl transferase for 2 h at room 37°. Finally, surfaces or levers were subjected to a final rinse with Ca-TBS and stored under Ca-TBS before measurement.

Single-molecule force spectroscopy measurements. SMFS measurements were performed on a custom built AFM[40] controlled by an MFP-3D controller from Asylum Research running custom written Igor Pro (Wavemetrics) software. Cantilever spring constants were calibrated using the thermal noise/equipartition method[41]. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from 0.2 to 6.4 μm s^{-1}. An x-y stage was actuated after each force-extension trace to expose the molecules on the cantilever to a new molecule at a different surface location with each trace. Typically 20,000–50,000 force-extension curves were obtained with a single cantilever in an experimental run of 18–24 h. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw data sets contained a majority of unusable curves due to lack of interactions or nonspecific adhesion of molecules to the cantilever tip, select curves showed single-molecule interactions. We filtered the data using a combination of automated data processing and manual classification by searching for contour length increments that matched the lengths of our specific protein fingerprint domains: Xyn (~89 nm) and CBM (~56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

Data analysis. Data were analysed using previously published protocols[17,18,22]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11$ nm connected by a fixed angle $\gamma = 41°$ and and assembled into barrier position histograms using cross-correlation. Detailed description of the contour length transformation can be found in Supplementary Note 1 and Supplementary Fig. 1.

For the loading rate analysis, the loading rate at the point of rupture was extracted by applying a line fit to the force vs time trace in the immediate vicinity before the rupture peak. The loading rate was determined from the slope of the fit. The most probable rupture forces and loading rates were determined by applying Gaussian fits to histograms of rupture forces and loading rates at each pulling speed.

Molecular dynamics simulations. The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97 Å resolution and is available at the protein data bank (PDB:4IU3). A protonation analysis performed in VMD[42] did not suggest any extra protonation and all the amino-acid residues were simulated with standard protonation states. The system was then solvated, keeping also the water molecules present in the crystal structure, and the net charge of the protein and the calcium ions was neutralized using sodium atoms as counter ions, which were randomly arranged in the solvent. Two other systems, based on the aforementioned one, were created using a similar salt concentration to the one used in the experiments (75 mM of NaCl). This additional salt caused little or no change in SMD results. The overall number of atoms included in MD simulations varied from 300,000 in the majority of the simulations to 580,000 for the unfolding of the X-Mod.

The MD simulations in the present study were performed employing the NAMD molecular dynamics package[43,44]. The CHARMM36 force field[45,46] along with the TIP3 water model[47] was used to describe all systems. The simulations were done assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 Å was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were

treated using the particle-mesh Ewald (PME)[48] method. The equations of motion were integrated using the r-RESPA multiple time step scheme[44] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. Before the MD simulations all the systems were submitted to an energy minimization protocol for 1,000 steps. The first two nanoseconds of the simulations served to equilibrate systems before the production runs that varied from 40 to 450 ns in the 10 different simulations that were carried out. The equilibration step consisted of 500 ps of simulation where the protein backbone was restrained and 1.5 ns where the system was completely free and no restriction or force was applied. During the equilibration the initial temperature was set to zero and was constantly increased by 1 K every 100 MD steps until the desired temperature (300 K) was reached.

To characterize the coupling between Doc and Coh, we performed SMD simulations[49] of constant velocity stretching (SMD-CV protocol) employing three different pulling speeds: 1.25, 0.625 and 0.25 Å ns^{-1}. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically (center of mass of ASN5), and moving a second restraint point, at the end of the Coh domain (center of mass of GLY210), with constant velocity in the desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the z-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the z-axis. Owing to the flexibility of the linkers, this approach reproduces the experimental set-up. All analyses of MD trajectories were carried out employing VMD[42] and its plug-ins. Secondary structures were assigned using the Timeline plug-in, which employs STRIDE criteria[50]. Hydrogen bonds were assigned based on two geometric criteria for every trajectory frame saved: first, distances between acceptor and hydrogen should be < 3.5 Å; second, the angle between hydrogen-donor-acceptor should be $< 30°$. Surface contact areas of interacting residues were calculated employing Volarea[51] implemented in VMD. The area is calculated using a probe radius defined as an *in silico* rolling spherical probe that is screened around the area of Doc exposed to Coh and also Coh area exposed to Doc.

References

1. Doi, R. H. & Kosugi, A. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* **2,** 541–551 (2004).
2. Carvalho, A. *et al.* Cellulosome assembly revealed by the crystal structure of the cohesin—dockerin complex. *Proc. Natl Acad. Sci. USA* **100,** 13809–13814 (2003).
3. Smith, S. P. & Bayer, E. A. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* **23,** 686–694 (2013).
4. Bayer, E. A., Lamed, R., White, B. A. & Flint, H. J. From cellulosomes to cellulosomics. *Chem. Rec.* **8,** 364–377 (2008).
5. Demain, A. L., Newcomb, M. & Wu, J. H. D. Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* **69,** 124–154 (2005).
6. Jindou, S. *et al.* Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium Ruminococcus flavefaciens. *FEMS Microbiol. Lett.* **285,** 188–194 (2008).
7. Ding, S. Y. *et al.* Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* **183,** 1945–1953 (2001).
8. Rincon, M. T. *et al.* Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS ONE* **5,** e12476 (2010).
9. Himmel, M. E. *et al.* Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315,** 804–807 (2007).
10. Fierobe, H.-P. *et al.* Degradation of cellulose substrates by cellulosome chimeras Substrate targeting versus proximity of enzyme components. *J. Biol. Chem.* **277,** 49621–49630 (2002).
11. Valbuena, A. *et al.* On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl Acad. Sci. USA* **106,** 13791–13796 (2009).
12. Salama-Alber, O. *et al.* Atypical cohesin-dockerin complex responsible for cell-surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.* **288,** 16827–16838 (2013).
13. Adams, J. J., Webb, B. A., Spencer, H. L. & Smith, S. P. Structural characterization of type ii dockerin module from the cellulosome of *Clostridium thermocellum*: calcium-induced effects on conformation and target recognition. *Biochemistry* **44,** 2173–2182 (2005).
14. Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc. Natl Acad. Sci. USA* **103,** 305–310 (2006).
15. Sikora, M. & Cieplak, M. Mechanical stability of multidomain proteins and novel mechanical clamps. *Proteins Struct. Funct. Bioinf.* **79,** 1786–1799 (2011).
16. Brunecky, R. *et al.* Structure and function of the Clostridium thermocellum cellobiohydrolase A X1-module repeat: enhancement through stabilization of the CbhA complex. *Acta. Crystallogr.* **68,** 292–299 (2012).
17. Stahl, S. W. *et al.* Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl Acad. Sci. USA* **109,** 20431–20436 (2012).
18. Jobst, M. A., Schoeler, C., Malinowska, K. & Nash, M. A. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *J. Vis. Exp.* **82,** e50950 (2013).
19. Otten, M. *et al.* From genes to protein mechanics on a chip. *Nat. Methods* **11,** 1127–1130 (2014).
20. Livadaru, L., Netz, R. R. & Kreuzer, H. J. Stretching response of discrete semiflexible polymers. *Macromolecules* **36,** 3732–3744 (2003).
21. Hugel, T., Rief, M., Seitz, M., Gaub, H. & Netz, R. Highly stretched single polymers: atomic-force-microscope experiments versus *ab-initio* theory. *Phys. Rev. Lett.* **94,** 048301 (2005).
22. Puchner, E. M., Franzen, G., Gautel, M. & Gaub, H. E. Comparing proteins by their unfolding pattern. *Biophys. J.* **95,** 426–434 (2008).
23. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor—ligand bonds explored with dynamic force spectroscopy. *Nature* **397,** 50–53 (1999).
24. Morfill, J. *et al.* Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy. *Biophys. J.* **93,** 3583–3590 (2007).
25. Berkemeier, F. *et al.* Fast-folding α-helices as reversible strain absorbers in the muscle protein myomesin. *Proc. Natl Acad. Sci. USA* **108,** 14139–14144 (2011).
26. Bertz, M., Wilmanns, M. & Rief, M. The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. *Proc. Natl Acad. Sci. USA* **106,** 13307–13310 (2009).
27. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402,** 100–103 (1999).
28. Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H. E. How strong is a covalent bond? *Science* **283,** 1727–1730 (1999).
29. Xue, Y., Li, X., Li, H. & Zhang, W. Quantifying thiol-gold interactions towards the efficient strength control. *Nat. Commun.* **5,** 4348 (2014).
30. Bomble, Y. J. *et al.* Modeling the self-assembly of the cellulosome enzyme complex. *J. Biol. Chem.* **286,** 5614–5623 (2011).
31. Sotomayor, M. & Schulten, K. Single-molecule experiments in vitro and in silico. *Science* **316,** 1144–1148 (2007).
32. Grubmüller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **271,** 997–999 (1996).
33. Thomas, W. *et al.* Catch-bond model derived from allostery explains force-activated bacterial adhesion. *Biophys. J.* **90,** 753–764 (2006).
34. Wang, W. & Malcolm, B. A. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *Biotechniques* **26,** 680–682 (1999).
35. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
36. Sawano, A. & Miyawaki, A. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* **28,** e78 (2000).
37. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expres. Purif.* **41,** 207–234 (2005).
38. Zimmermann, J. L., Nicolaus, T., Neuert, G. & Blank, K. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat. Protoc.* **5,** 975–985 (2010).
39. Yin, J., Lin, A. J., Golan, D. E. & Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **1,** 280–285 (2006).
40. Gumpp, H., Stahl, S. W., Strackharn, M., Puchner, E. M. & Gaub, H. E. Ultrastable combined atomic force and total internal fluorescence microscope. *Rev. Sci. Instrum.* **80,** 063704 (2009).
41. Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64,** 1868 (1993).
42. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14,** 33–38 (1996).
43. Kalé, L. *et al.* NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151,** 283–312 (1999).
44. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
45. Best, R. B. *et al.* Optimization of the additive CHARMM All-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ 1and χ 2dihedral Angles. *J. Chem. Theory Comput.* **8,** 3257–3273 (2012).
46. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
47. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–934 (1983).
48. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

49. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72,** 1568–1581 (1997).
50. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinf.* **23,** 566–579 (1995).
51. Ribeiro, J. V., Tamames, J. A. C., Cerqueira, N. M. F. S. A., Fernandes, P. A. & Ramos, M. J. Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems. *Chem. Biol. Drug Des.* **82,** 743–755 (2013).

Acknowledgements

We gratefully acknowledge funding from an advanced grant of the European Research Council (Cellufuel Grant 294438) and from DFG SFB 1032 and the Excellence Cluster Center for Integrated Protein Science Munich. This work was supported by grants from the National Institutes of Health (NIH, 9P41GM104601 to K.S.) and the National Science Foundation (NSF, MCB-1157615 to K.S.). Simulations made use of the Texas Advanced Computing Center (TACC) as part of the Extreme Science and Engineering Discovery Environment (XSEDE, MCA93S028 to K.S.) and the NCSA Blue Waters sustained-petascale supercomputer as part of the general allocations (Simulations of Cellulosomal Subunits: Components of a Molecular Machinery for Depolymerization of Feedstock for Production of Second Generation Biofuels, to K.S.). A grant to E.A.B., H.E.G. and M.A.N. from GIF, the German-Israeli Foundation for Scientific Research and Development is also noted. Additional support was obtained from grants (No. 1349) to E.A.B. from the Israel Science Foundation (ISF) and the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry. M.A.N. acknowledges funding from Society in Science - The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland.

Author contributions

Performed and analysed SMFS experiments: C.S., K.H.M., L.F.M., M.A.J., E.D. and M.A.N.; performed and analysed MD simulations: R.C.B. and K.S.; provided proteins and DNA cloning vectors: W.O., D.B.F. and E.A.B.; wrote and edited the manuscript: C.S., K.H.M., R.C.B., E.A.B., K.S., H.E.G. and M.A.N.; supervised research: E.A.B., K.S., H.E.G. and M.A.N.

Additional information

Accession codes: Plasmids used in this study are available through Addgene (https://www.addgene.org) under the following accession codes: Xylanase-Xmodule-Dockerin: 60865; Cohesin-CBM: 60866.

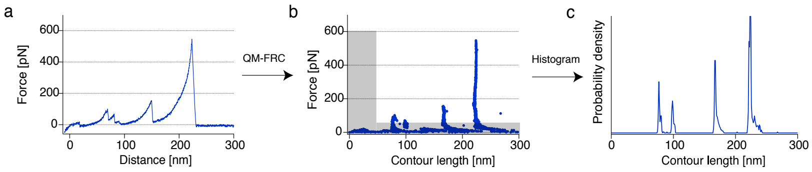
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

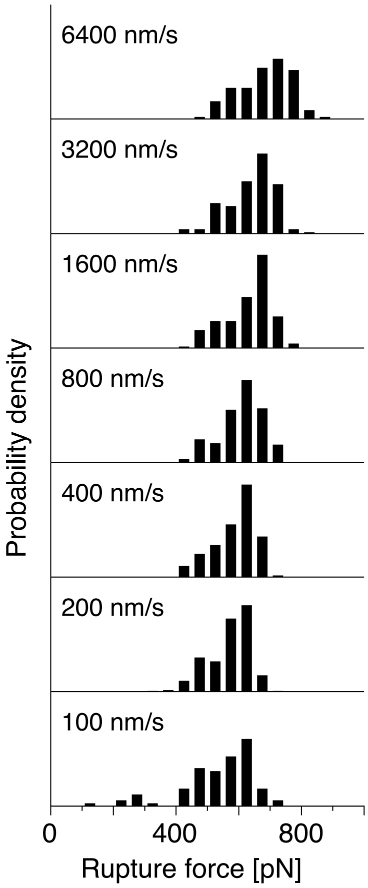
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Schoeler, C. *et al.* Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* 5:5635 doi: 10.1038/ncomms6635 (2014).

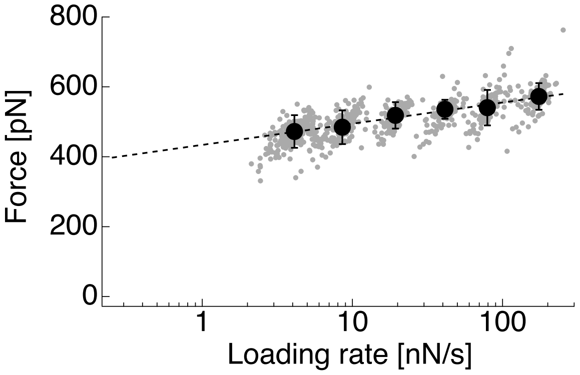
Supplementary Figures



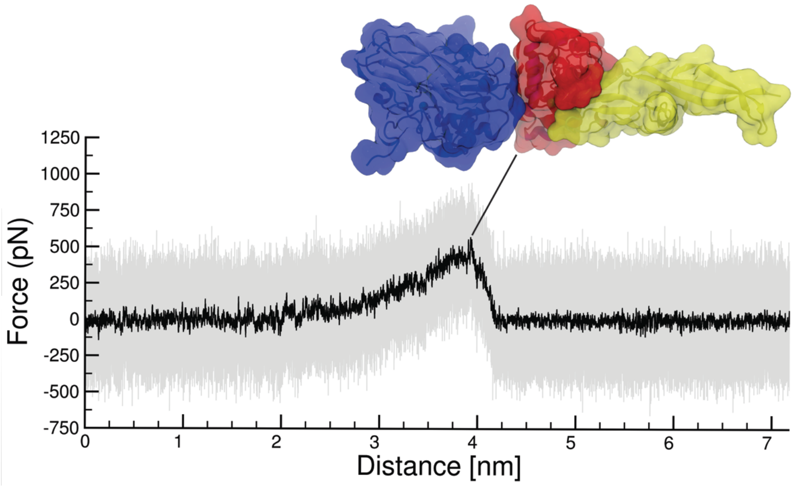
Supplementary Fig. 1: Assembly of contour length histograms. **a** Force-extension traces are transformed into contour length space using a QM-corrected FRC model with parameters $\gamma = 41°$, and $b = 0.11\,\text{nm}$. **b** In force-contour length space, force and contour length thresholds are applied and the data are histogrammed with a bin width of 1 nm to obtain the histogram in **c**. To obtain a master histogram, individual histograms reflecting a specific unfolding pathway are cross-correlated and aligned by offsetting by the maximum correlation value.



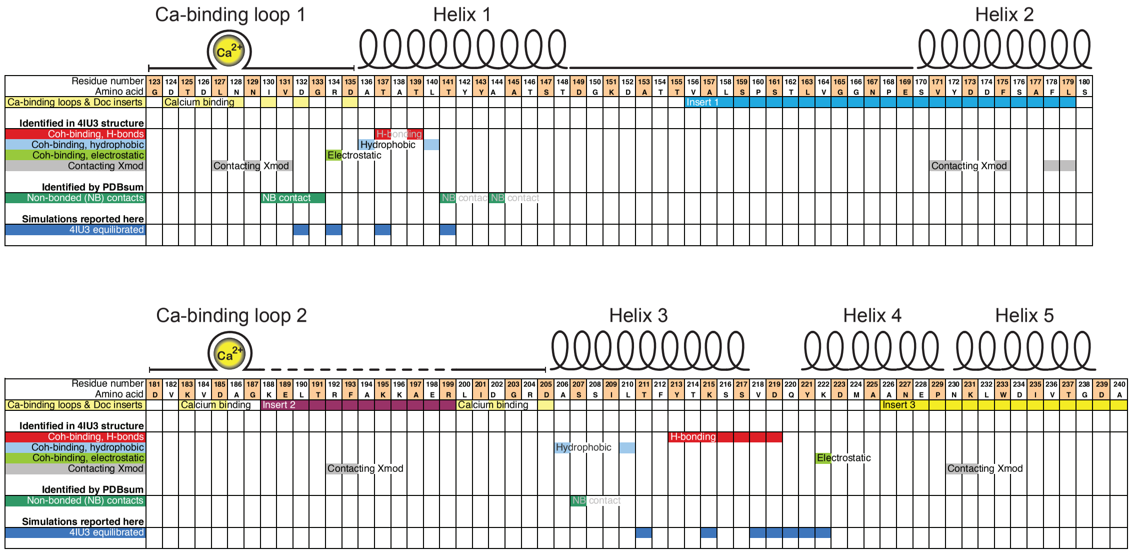
Supplementary Fig. 2: Complex rupture force histograms for pulling speeds ranging from 100 nm s^{-1} to 6400 nm s^{-1}. Pulling speeds are indicated next to the histograms. Only traces with an intact XMod were taken into account (no XMod unfolding observed, corresponding to Fig. 2, trace 1). At the slowest pulling speed data suggest the presence of a lower rupture force population.



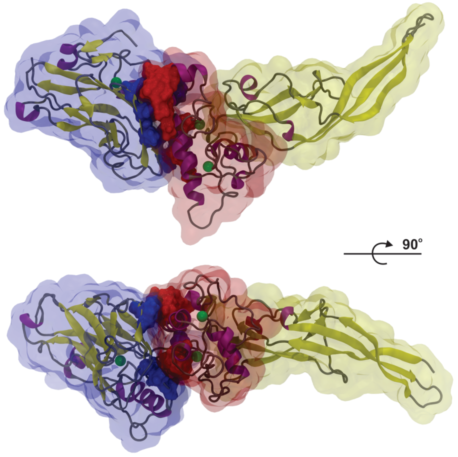
Supplementary Fig. 3: Dynamic force spectrum for XMod unfolding obtained from 654 force-extension traces. The gray points show single XMod unfolding events. Black circles represent the most probable rupture forces and loading rates obtained by Gaussian fitting at each pulling speed. Error bars are ±1 standard deviation. The dashed line is a least squares fit to the Bell-Evans model that yielded $\Delta x = 0.15\,\mathsf{nm}$ and $k_{off} = 2.6 \times 10^{-6}\,\mathsf{s}^{-1}$.



Supplementary Fig. 4: Force distance trace obtained by SMD at a pulling speed of 0.25 Å ns^{-1}. Force values at each time step are shown in gray, with average force calculated every 200 ps in black. The inset is a snapshot of the XMod-Doc:Coh complex immediately prior to rupture. XMod is shown in yellow, Doc in red and Coh in blue.

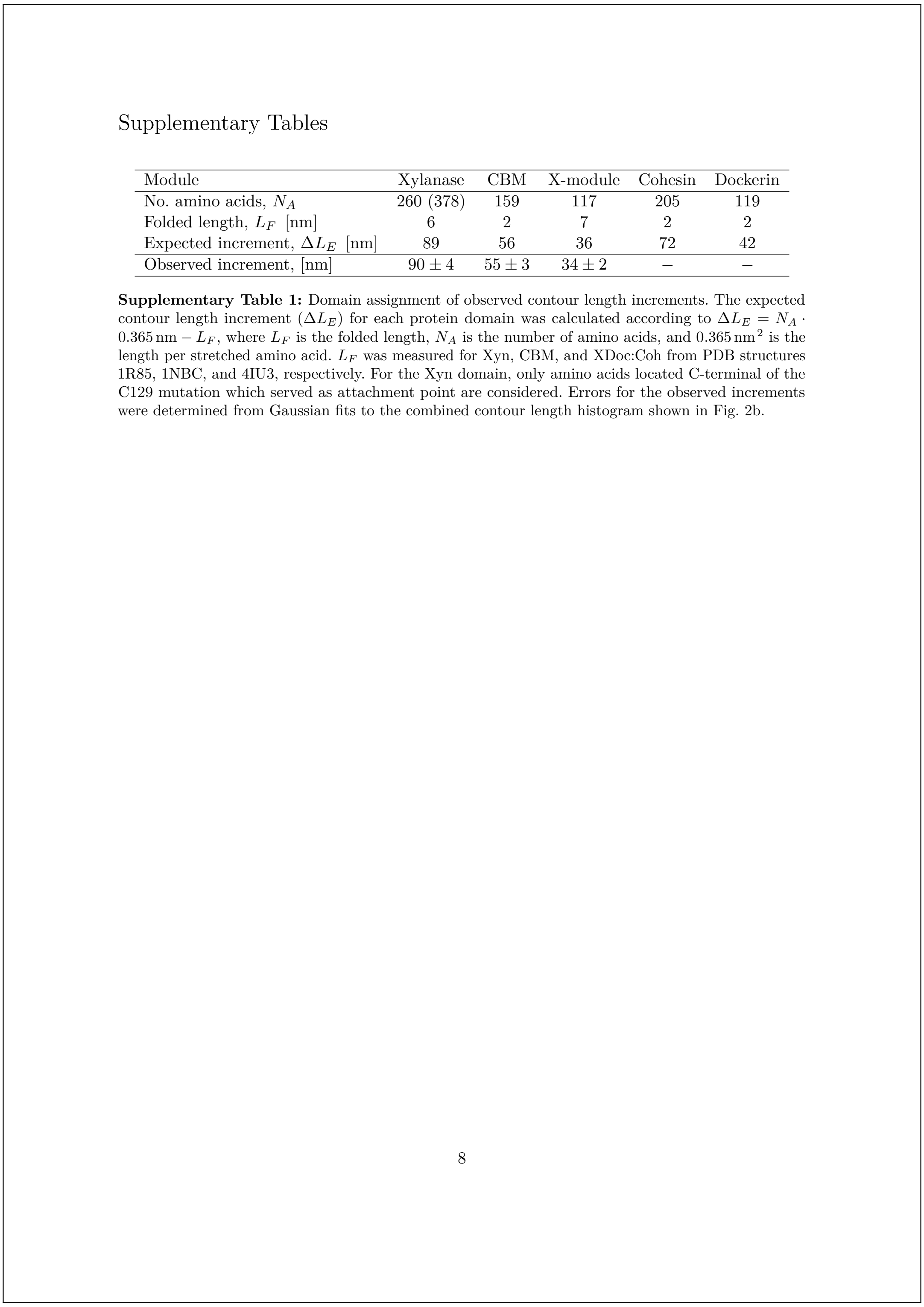


Supplementary Fig. 5: Amino acid sequence and secondary structure elements of the *Ruminococcus flavefaciens* dockerin module. The table includes interactions with the Coh and XMod domains identified in the PDB code 4IU3 crystal structure, PDBsum[1] and MD simulations.



Donor	Acceptor	Prevalence
LYS215-S	GLU105-S	87%
ASN107-S	ASP219-S	48%
ASP219-B	ASN107-S	39%
THR141-S	ASP118-S	31%
LYS156-S	ASP132-S	22%

Supplementary Fig. 6: Hydrogen bond contacts between XMod-Doc (yellow and red surface, respectively) and Coh (blue surface). The residues that have hydrogen bonds lasting for more than 10% of the simulation time are represented in a glossy surface. In the bottom of the figure the five most prevalent hydrogen bond interactions are presented. The letter S or B indicate if the respective interaction is made by the amino acid side chain or backbone.



Supplementary Tables

Module	Xylanase	CBM	X-module	Cohesin	Dockerin
No. amino acids, N_A	260 (378)	159	117	205	119
Folded length, L_F [nm]	6	2	7	2	2
Expected increment, ΔL_E [nm]	89	56	36	72	42
Observed increment, [nm]	90 ± 4	55 ± 3	34 ± 2	–	–

Supplementary Table 1: Domain assignment of observed contour length increments. The expected contour length increment (ΔL_E) for each protein domain was calculated according to $\Delta L_E = N_A \cdot 0.365\,\text{nm} - L_F$, where L_F is the folded length, N_A is the number of amino acids, and 0.365 nm[2] is the length per stretched amino acid. L_F was measured for Xyn, CBM, and XDoc:Coh from PDB structures 1R85, 1NBC, and 4IU3, respectively. For the Xyn domain, only amino acids located C-terminal of the C129 mutation which served as attachment point are considered. Errors for the observed increments were determined from Gaussian fits to the combined contour length histogram shown in Fig. 2b.

Supplementary Notes

Supplementary Note 1: QM-FRC Model for Polymer Elasticity

The freely rotating chain model[3] considers bonds of length b, connected by a fixed angle γ. The torsional angles are not restricted. The stretching behavior in the FRC picture is given by

$$\frac{x}{L} = \begin{cases} \frac{Fa}{3k_BT} & \text{for } \frac{Fb}{k_BT} < \frac{b}{p} \\ 1 - \left(\frac{4Fp}{k_BT}\right)^{-\frac{1}{2}} & \text{for } \frac{b}{p} < \frac{Fb}{k_BT} < \frac{p}{b} \\ 1 - \left(\frac{cFb}{k_BT}\right)^{-1} & \text{for } \frac{p}{b} < \frac{Fb}{k_BT} \end{cases} \tag{1}$$

where $a = b\frac{1+\cos\gamma}{(1-\cos\gamma)\cos\frac{\gamma}{2}}$ is the Kuhn length, and $p = b\frac{\cos\frac{\gamma}{2}}{|\ln(\cos\gamma)|}$ is the effective persistence length in the FRC picture.

To account for backbone elasticity of the polypeptide chain at high force, quantum mechanical *ab-initio* calculations can be used to obtain the unloaded contour length at zero force. A polynomial approximation to these calculations can be used to obtain the unloaded contour length at zero force L_0:

$$F = \gamma_1\left(\frac{L}{L_0} - 1\right) + \gamma_2\left(\frac{L}{L_0} - 1\right)^2 \tag{2}$$

where the $\gamma_1 = 27.4\,\mathrm{nN}$, and $\gamma_2 = 109.8\,\mathrm{nN}$ are the elastic coefficients reported for polypeptides[4].

Supplementary Note 2: Bell-Evans Model for Mechanically Induced Receptor Ligand Dissociation

The Bell-Evans model was used to estimate the distance to the transition state (Δx) and the natural off-rate (k_{off}) of individual rupture events:

$$\langle F\rangle = \frac{k_BT}{\Delta x}\ln\frac{\Delta x \cdot \dot{F}}{k_{off}k_BT} \tag{3}$$

where k_B is Boltzmann's constant, T is the temperature and $\dot{F}$ is the loading rate at the point of rupture.

Supplementary Methods

Materials

Silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation) with a nominal spring constant of 100 pN/nm (25 kHz resonance frequency in water) were used. Circular coverglasses, 2.4 cm in diameter, were obtained from Menzel Gläser (Braunschweig, Germany). 3-Aminopropyl dimethyl ethoxysilane (APDMES) was purchased from ABCR GmbH (Karlsruhe, Germany). NHS-PEG-Maleimide ($5\,\mathrm{kDa}$) was purchased from Rapp Polymer (Tübingen, Germany). Immobilized TCEP Disulfide Reducing Gel was obtained from Thermo Scientific (Pittsburgh, PA). The following standard chemicals were obtained from Carl Roth (Karlsruhe, Germany) and used as received: tris(hydroxymethyl)aminomethane (TRIS, >99% p.a.), $\mathrm{CaCl_2}$ (>99% p.a.), sodium borate (>99.8% p.a), NaCl (>99.5% p.a.), ethanol (>99% p.a.), and toluene (>99.5% p.a.). Borate buffer was 150 mM, pH 8.5. The measurement buffer for force spectroscopy was Tris-buffered saline (TBS, 25 mM TRIS, 75 mM NaCl, pH 7.2) supplemented with $\mathrm{CaCl_2}$ to a final concentration of 1 mM. All buffers were filtered through a sterile $0.2\,\mu\mathrm{m}$ polyethersulfone membrane filter (Nalgene, Rochester, NY, USA) prior to use.

Protein Sequences

Sequences of protein constructs used in this work are listed here. Domains as well as engineered tags and residues are color-coded.

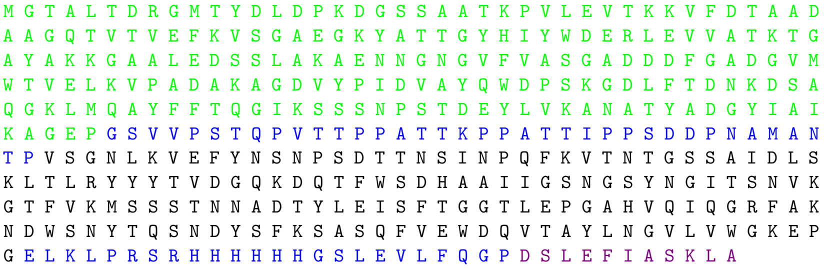
Xyn-XModDoc

Xylanase T129C
Linker or extra residues
X-module
Dockerin type III

```
M S H H H H H H K N A D S Y A K K P H I S A L N A P Q L D Q R Y K N E F T I G A
A V E P Y Q L Q N E K D V Q M L K R H F N S I V A E N V M K P I S I Q P E E G K
F N F E Q A D R I V K F A K A N G M D I R F H T L V W H S Q V P Q W F F L D K E
G K P M V N E C D P V K R E Q N K Q L L L K R L E T H I K T I V E R Y K D D I K
Y W D V V N E V V G D D G K L R N S P W Y Q I A G I D Y I K V A F Q A A R K Y G
G D N I K L Y M N D Y N T E V E P K R T A L Y N L V K Q L K E E G V P I D G I G
H Q S H I Q I G W P S E A E I E K T I N M F A A L G L D N Q I T E L D V S M Y G
W P P R A Y P T Y D A I P K Q K F L D Q A A R Y D R L F K L Y E K L S D K I S N
V T F W G I A D N H T W L D S R A D V Y Y D A N G N V V V D P N A P Y A K V E K
G K G K D A P F V F G P D Y K V K P A Y W A I I D H K V V P N T V T S A V K T Q
Y V E I E S V D G F Y F N T E D K F D T A Q I K K A V L H T V Y N E G Y T G D D
G V A V V L R E Y E S E P V D I T A E L T F G D A T P A N T Y K A V E N K F D Y
E I P V Y Y N N A T L K D A E G N D A T V T V Y I G L K G D T D L N N I V D G R
D A T A T L T Y Y A A T S T D G K D A T T V A L S P S T L V G G N P E S V Y D D
F S A F L S D V K V D A G K E L T R F A K K A E R L I D G R D A S S I L T F Y T
K S S V D Q Y K D M A A N E P N K L W D I V T G D A E E E
```

Coh-CBM C2A, C63S

CBM (C2A, C63S)
Linker or extra residues
CohIII
ybbR-Tag



M G T A L T D R G M T Y D L D P K D G S S A A T K P V L E V T K K V F D T A A D
A A G Q T V T V E F K V S G A E G K Y A T T G Y H I Y W D E R L E V V A T K T G
A Y A K K G A A L E D S S L A K A E N N G N G V F V A S G A D D D F G A D G V M
W T V E L K V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T D N K D S A
Q G K L M Q A Y F F T Q G I K S S S N P S T D E Y L V K A N A T Y A D G Y I A I
K A G E P G S V V P S T Q P V T T P P A T T K P P A T T I P P S D D P N A M A N
T P V S G N L K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I D L S
K L T L R Y Y Y T V D G Q K D Q T F W S D H A A I I G S N G S Y N G I T S N V K
G T F V K M S S S T N N A D T Y L E I S F T G G T L E P G A H V Q I Q G R F A K
N D W S N Y T Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G K E P
G E L K L P R S R H H H H H H G S L E V L F Q G P D S L E F I A S K L A

Supplementary References

[1] Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M. PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences* **22**, 488–490 (1997).

[2] Dietz, H.; Rief, M. Protein structure by mechanical triangulation. *Proceedings of the National Academy of Sciences* **103**, 1244–1247 (2006).

[3] Livadaru, L.; Netz, R. R.; Kreuzer, H. J. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules* **36**, 3732–3744 (2003).

[4] Hugel, T.; Rief, M.; Seitz, M.; Gaub, H.; Netz, R. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-Initio Theory. *Physical Review Letters* **94**, 048301 (2005).

5.2 PUBLICATION P3: A Solution to the Conformational Linker Stretching

This article introduces a new kind of polymer linker for the site-specific and covalent surface immobilization of the proteins under investigation in SMFS assays. The new method solves the problem of conformational transitions in commonly used tethers, and thereby improves accuracy in determination of contour length increments greatly. The main part of my contribution to this publication was the application of polymer elasticity models to the data, and the development of a new kind of data reduction strategy in large SMFS data sets, that condenses a large number of single force-distance traces into a representative 'master curve' while conserving the elastic response of the polymers. I furthermore wrote substantial parts of the manuscript, created the data-driven figures, and contributed to the experiment design and analysis of the data.

Elastin-like Polypeptide Linkers for Single-molecule Force Spectroscopy

Wolfgang Ott†, Markus A. Jobst†, Magnus S. Bauer, Ellis Durner,
Lukas F. Milles, Michael A. Nash, and Hermann E. Gaub

†*these authors contributed equally to this publication*

published in

ACS Nano, 11(6), 6346-6354, (2017)

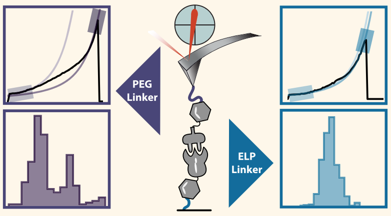
ACSNANO
www.acsnano.org
ARTICLE

Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy

Wolfgang Ott,[†,‡,⊥] Markus A. Jobst,[†,⊥] Magnus S. Bauer,[†] Ellis Durner,[†] Lukas F. Milles,[†]
Michael A. Nash,[§,∥] and Hermann E. Gaub*[,†]

[†]Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, 80799 Munich, Germany
[‡]Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[§]Department of Chemistry, University of Basel, 4056 Basel, Switzerland
[∥]Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland

Supporting Information



ABSTRACT: **Single-molecule force spectroscopy (SMFS) is by now well established as a standard technique in biophysics and mechanobiology. In recent years, the technique has benefitted greatly from new approaches to bioconjugation of proteins to surfaces. Indeed, optimized immobilization strategies for biomolecules and refined purification schemes are being steadily adapted and improved, which in turn has enhanced data quality. In many previously reported SMFS studies, poly(ethylene glycol) (PEG) was used to anchor molecules of interest to surfaces and/or cantilever tips. The limitation, however, is that PEG exhibits a well-known trans–trans–gauche to all-trans transition, which results in marked deviation from standard polymer elasticity models such as the worm-like chain, particularly at elevated forces. As a result, the assignment of unfolding events to protein domains based on their corresponding amino acid chain lengths is significantly obscured. Here, we provide a solution to this problem by implementing unstructured elastin-like polypeptides as linkers to replace PEG. We investigate the suitability of tailored elastin-like polypeptides linkers and perform direct comparisons to PEG, focusing on attributes that are critical for single-molecule force experiments such as linker length, monodispersity, and bioorthogonal conjugation tags. Our results demonstrate that by avoiding the ambiguous elastic response of mixed PEG/peptide systems and instead building the molecular mechanical systems with only a single bond type with uniform elastic properties, we improve data quality and facilitate data analysis and interpretation in force spectroscopy experiments. The use of all-peptide linkers allows alternative approaches for precisely defining elastic properties of proteins linked to surfaces.**

KEYWORDS: *single-molecule force spectroscopy, elastin-like polypeptides, biopolymer spacer, sortase coupling, protein ligation*

Refined Techniques in SMFS. Single-molecule force spectroscopy (SMFS) is a state-of-the-art technique in the rapidly growing field of molecular biomechanics.[1−3] Tools and methods are being steadily developed to improve ease of sample handling, sensitivity, reproducibility, and reliability.[4,5] In parallel, the biochemical toolbox is expanded continuously, enabling analysis of more complex and demanding biological systems. Improvements such as the use of orthogonal binding handles,[6−9] diverse biomolecule immobilization strategies,[10−14] and alternative methods for protein synthesis (*i.e.*, recombinant bulk expression or cell-free *in vitro* expression) are all examples of significant technical advances that have been achieved in recent years.[15]

Requirements for Recording Large Data Sets and Challenges Arising Therefrom. A key requirement to probe multiple different protein domains in a single experiment is the ability to use a single cantilever over extended periods of time to achieve a large number of force−extension traces. For this purpose, two main advances are worth noting, the first of them being the improvement of geometrically defined covalent surface tethering and the second being the discovery and characterization of the type III cohesin−dockerin (Coh:Doc) interaction.[7] Coh:Doc receptor−ligand pairs can withstand remarkably high forces in a SMFS assays and exhibit extremely high long-term functionality. This latter property is particularly important for carrying out multiplexed experiments where many proteins deposited onto the same surface and spatially

Received: April 18, 2017
Accepted: June 7, 2017
Published: June 7, 2017

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

separated are pulled apart using the same receptor-modified cantilever. In such a configuration, Coh:Doc is used as a binding handle to successfully and continuously unfold target proteins for over 24 h of measurement time without significant loss of binding activity. Data sets of typically several tens of thousands of force−extension curves can easily be obtained using type III Coh:Doc, dramatically outperforming other mechanostable interactions (*e.g.*, biotin−avidin).

The ability to measure with a single cantilever over several days allows interrogation of different types or variants of proteins immobilized on different positions of the same substrate (*i.e.*, protein microarrays) and to achieve statistical significance over the course of a single experiment. This leads to large data sets and requires the use of sophisticated algorithms to identify and extract specific single-molecule interactions among a large number of traces with poor signal, such as empty traces, multiple interactions in parallel, or nonspecific interactions. Independent of the size of the data sets though, elasticity models whether applied as part of elaborate algorithms or fitted manually to single curves have in the past been required to account for the different elastic contributions stemming from heterogeneous stretching behavior of mixed poly(ethylene glycol) (PEG)−protein polymer backbone.

Conformational Changes of PEG Linker Molecules Obscure Molecular Characteristics of Interest. When performing SMFS in an elevated force regime using PEG as linker molecules, additional challenges arise. A conformational transition of PEG occurs in a force range of up to *ca.* 300 pN, resulting in an approximately linear force−extension regime.[16−18] In aqueous solutions, PEG exhibits a trans−trans−gauche conformation. With rising force on the polymer, the occupancy of conformations is shifted to all-trans, effectively increasing the net polymer contour length. Analysis methods such as fitting standard elasticity models to the data or detecting contour length increments within said force range are therefore compromised and would, for a quantitative description, require improved heterogeneous elasticity models.

PEG is a highly flexible polymer with a low persistence length, while peptide bonds have restricted degrees of freedom. These restrictions alter the stretching behavior and give rise to marked differences in comparison to PEG. Furthermore, the ratio of PEG linker length to unfolded protein backbone length is not constant over the course of an unfolding trace, which means fitting parameters must be optimized for different sections of the curve as more domains unfold. This issue becomes particularly significant and noticeable when probing protein unfolding and receptor−ligand unbinding in a high force regime and is also problematic when unfolding occurs across a broad range of forces.

Benefits of ELP Linkers in SMFS. In this study we investigate the feasibility of biological peptide polymers to circumvent this problem. We selected well-characterized elastin-like polypeptides (ELPs) as a suitable candidate for this purpose. The progression of cloning techniques of repetitive genes in recent years has set the stage for precisely defined protein polymers and opened up the ability to design, produce, and purify protein spacers of well-defined contour length and chemical composition for single-molecule experiments.[19−22] ELPs exhibit similar elasticity behavior as unfolded protein backbone and are completely monodisperse, a key advantage compared to synthetic polymers such as PEG. Monodisperse ELP linkers fused directly to a protein of interest allow for complete control of the lengths of a nanomechanical system from the surface up to the force transducer, which is not true for the chemically synthesized PEG polymers with non-negligible polydispersity. Since ELPs are expressed recombinantly in *Escherichia coli* (*E. coli*), their production is easily scaled up, resulting in lower costs compared to commercially available heterobifunctional PEGs. Furthermore, ELPs can be produced with N-/C-terminal protein ligation tags, which can be used for specific and bio-orthogonal surface chemistry in SMFS sample preparation.

ELPs are synthetic biopolymers derived from tropoelastin domains. They are composed of a repetitive amino acid heptamer "Val-Pro-Gly-Xaa-Gly",[23] where Xaa is a guest residue that can be any amino acid apart from proline. The guest residue influences the hydrophobicity of the protein and impacts the lower critical solution temperature, the point at which the ELP undergoes a soluble-to-insoluble phase transition. At this environment-dependent cloud point, ELPs change their conformation and precipitate, resulting in clouding of the solution.

ELPs are intrinsically disordered proteins that do not fold into well-defined secondary and tertiary structures, but rather remain unfolded and flexible, a property that is ideally suited to their application as spacer/linker molecules for SMFS.[24] We hypothesized that ELPs would therefore be a suitable choice to achieve both surface passivation and site-specific immobilization in single-molecule nanomechanical experiments. The bulky yet flexible features of ELPs inhibit nonspecific protein binding to the surface, while enabling ligation of other proteins due to the high degree of accessibility of N- or C-terminally fused peptide tags. Post-translational protein ligation methods have made it possible to move from organic chemical conjugation methods toward enzyme-mediated covalent immobilization, for example utilizing sortase A or Sfp.[14,25] Both enzymes catalyze sequence- and site-specific reactions yielding uniform protein orientation at the surface.

ELPs have previously been the subject of atomic force microscopy (AFM) studies. For example, AFM was used to support theoretical predictions about the behavior of ELPs above and below their cloud point, as well as to study ELP elasticity.[26−28] This study was carried out entirely below the cloud point, so that intermolecular interactions between ELPs were negligible. In contrast to prior studies, we employ ELPs as spacer molecules with other protein domains attached. Our results show that ELPs provide several benefits over PEG linkers in SMFS attributable primarily to the features of having uniform elastic properties and monodisperse linkers.

This study offers an attractive substitute for established PEG systems using all-protein ELP linkers. The immobilization strategy provides precise control over the elastic properties of multicomponent protein mechanical systems linked between a glass surface and a force transducer. Our approach transfers advances in smart polymer research to SMFS experiments and describes the improvements achieved through this alternative surface anchoring strategy.

RESULTS AND DISCUSSION

SMFS with Receptor−Ligand Polyproteins Employing Site-Specific Immobilization. Typically PEG linkers with an *N*-hydroxysuccinimide (NHS) group are linked to an amino-silanized surface. The other end of the PEG contains a reactive group for protein immobilization, which in most cases is a thiol-reactive maleimide group. Figure 1A illustrates a Coh:Doc-

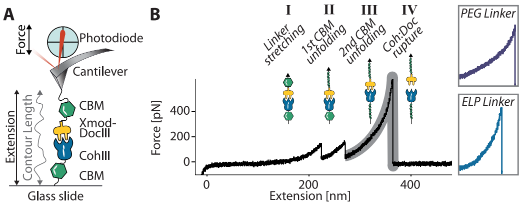


Figure 1. (A) SMFS configuration: Cantilevers are functionalized with CBM-Xmod-DocIII fusion proteins. Glass slides are modified with CohIII-CBM constructs. (B) Coh:Doc-based SMFS unfolding trace. Following Coh:Doc complex formation at zero extension, retraction of the cantilever results in mechanical stretching of the receptor:ligand-linked polyprotein. (I) Spacer molecules are fully extended and stretched. (II, III) The weakest links in the chain, usually the fingerprint domains (here: CBM), are unfolded in series. (IV) Finally, the Coh:Doc complex dissociates under force. The unfolded CBM domains can then refold after the complex rupture. The cantilever is now free to probe a different molecule on the surface. The insets on the right side qualitatively illustrate the differences in linker stretching in the high-force regime as observed in the final peak for constructs immobilized using PEG and ELP linkers. A quasi-linear regime of PEG stretching attributable to the conformational transition from trans-trans-gauche to all-trans is clearly visible for PEG in contrast to ELP.

based SMFS experiment. Proteins anchored to a functionalized glass surface are probed by the corresponding receptor fusion protein covalently linked to the cantilever tip. A characteristic unfolding curve recorded at constant speed is shown in Figure 1B. After the Coh:Doc complex is formed by contacting the cantilever with the surface, force is applied by retracting the base of the cantilever. The signal is detected by a quadrant photodiode with a laser that is reflected off the back side of the cantilever. Bending of the cantilever is translated into a differential voltage output of the photodiode. Upon retraction of the cantilever base at constant speed, the polymer linker is stretched first (Figure 1B, I). Subsequently, the weakest component in the system unfolds. In this case two carbohydrate binding modules (CBMs) are unfolded consecutively (Figure 1B, II and III). Finally, the force increases to a level where the receptor ligand pair dissociates. Following Coh:Doc rupture, the force drops to zero (Figure 1B, IV) and the cantilever is free to probe another molecule at a different location on the surface.

In order to identify data traces that show specific single-molecule interactions, a multilevel sorting algorithm is used to search for characteristic unfolding patterns of the fingerprint domains. This algorithm takes into account the unfolding forces and the measured increases in contour length (*i.e.*, contour length increments) of the peptide backbone upon unfolding of the various fingerprint domains.[29] Independent of the analysis method, however, accurate polymer elasticity models are required to quantify the hidden lengths of the folded proteins that are released by the unfolding events, giving rise to the limitations of PEG systems described above.

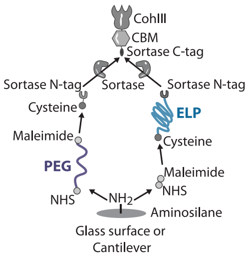


Figure 2. Comparison of immobilization strategies. For standard immobilization with PEG spacers, NHS chemistry was used to link PEG to amino-silanized surfaces. Protein constructs were then coupled *via* cysteine-sortase tag peptides to the maleimide end-groups on the PEG spacers. For immobilization with ELP linkers, a small-molecule NHS-maleimide cross-linker with a negligible contour length of 0.83 nm was used to couple cysteine-ELP spacers with a sortase-tag to the amino-silanized surface. In both cases, a fusion protein of interest, consisting of a CBM fingerprint domain and a mechanostable Coh receptor, was enzymatically coupled to the immobilized molecules on the surface in a subsequent step. Depicted is the functionalization of the glass surface with CohIII. The functionalization of the cantilever tip with DocIII followed a similar scheme.

Adaptation of Surface Chemistry to Tether Protein Domains to ELP Linkers. The comparison of PEG with ELP linkers was carried out by cloning and recombinantly expressing two different ELPs both with 120 nm theoretical contour length (ELP$_{120\ \mathrm{nm}}$, assuming 0.365 nm per amino acid).[30] One ELP linker contained an N-terminal sortase-tag (“GGG”) and a C-terminal cysteine. The other ELP linker had a sortase-tag at its C-terminus (“LPETGG”) and a cysteine at the N-terminus. Two analogous bioconjugation routes were used to attach ELP or PEG linkers to cantilevers and glass surfaces (Figure 2). To achieve a direct comparison, 15 kDa PEG linkers of similar contour lengths (~120 nm) were used. For PEG experiments, 15 kDa NHS-PEG-maleimide was immobilized onto an amino-silanized glass slide (PEG$_{120\ \mathrm{nm}}$). The maleimide groups of the PEG reacted with a GGGGG-Cys peptide, leaving the sortase N-tag available for subsequent derivatization. For ELP experiments, a small-molecule cross-linker (sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate, sulfo-SMCC), which added negligible contour length (0.83 nm) to the system, was first immobilized onto amino-silanized glass, followed by coupling with GGG-ELP$_{120\ \mathrm{nm}}$-Cys. Both strategies resulted in the sortase N-tag being available for conjugation *via* sortase-mediated enzymatic ligation. The protein of interest (CohIII-CBM-LPETGG) was linked by sortase A to ELP or PEG (Figure 2). The same strategy was used for the cantilever, except GGG-Xmod-DocIII was conjugated by sortase A to Cys-ELP$_{120\ \mathrm{nm}}$-LPETGG or to PEG$_{120\ \mathrm{nm}}$-coupled Cys-LPETGG. Our enzyme-mediated protein immobilization approach has the advantage of site-specific linkages and results in a homogeneous

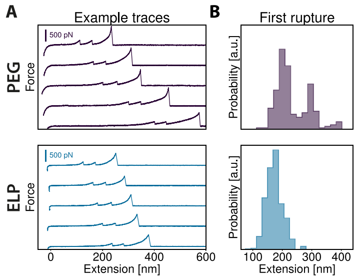


Figure 3. Comparison of dispersity of PEG and ELP linkers. (A) Typical force−extension traces for PEG (purple) and ELPs (blue). In the PEG linker experiment, the unfolding events occur over a wider range of absolute extension values, whereas unfolding events with ELP linkers occur over a narrow range. (B) Histograms showing the distribution of extension values corresponding to the first CBM unfolding event in each curve (PEG: *N* = 219; ELP: *N* = 521). Due to the polydispersity of the PEG linkers, three discrete populations with different extensions are clearly visible, while for ELPs only one population is observed.

orientation of the proteins at the surface. Such uniformly immobilized proteins lead to a well-defined propagation of the applied force through the molecular complex under investigation and to well-defined distributions of the unfolding/rupture events in the force−extension curves. The use of N- and C-terminal tags for surface chemistry also ensured that only full-length (*i.e.*, fully translated) ELPs were measured in the experiment.

AFM experiments performed with ELPs as linkers showed a higher percentage of clearly identifiable single-molecule unfolding traces. We attribute this to the bulky character of the ELPs. They provide a less dense surface immobilization of the biomolecules of interest when compared to PEG-based immobilization. This behavior is advantageous since high surface density frequently causes multiple interactions between surface- and cantilever-bound molecules in SMFS experiments (Supplemental Figure S1). Multiple interactions are generated when more than one receptor−ligand interaction is formed in parallel. The complicated unfolding and unbinding traces that result from multiple bonds pulled in parallel are hardly interpretable and therefore discarded from the analysis (Supplemental Figure S2). Efficient passivation of glass surfaces against nonspecific adhesion of proteins requires a dense PEG surface layer, to prevent proteins from nonspecifically sticking to the glass surface. Approaches such as titrating functional (*i.e.*, maleimide end-groups) with nonfunctional (*i.e.*, CH_3 end-groups) PEG or changing the concentration of binding agents or proteins of interest can improve the process. In our experience, however, surface immobilization with ELP instead of PEG linkers leads to better passivation of the surface and a higher percentage of single-molecule traces without the need for any titration of functional and nonfunctional linkers.

Comparison of Dispersity between PEG and ELP Linkers. All unfolding traces were presorted by an automated analysis routine, selecting for single interactions that display two consecutive CBM unfolding events. Following the automated sorting, deletion of obviously erroneous curves (typically 10%) caused by, for example, baseline drift was performed manually.[7,29] PEG unfolding traces showed wildly varying initial extensions prior to the first CBM unfolding event. This is likely caused by the non-negligible polydispersity of PEG, as we did not observe multiple discrete populations with ELP experiments. The intrinsic monodispersity of ELP molecules is a clear advantage. Since they are produced recombinantly in *E. coli* with functional tags *in vivo*, only full-length protein sequences have the necessary terminal peptide tags that allow for surface immobilization. Additionally, ELPs were purified with inverse transition cycling (ITC), a method developed for ELP purification based on their reversible precipitation behavior. Possibly shorter ELPs are removed during the process, since their cloud point is higher than for $ELP_{120\ nm}$. Although the polydispersity of chemically synthesized PEGs (mass distribution ∼10−20 kDa) is sufficiently low for many applications, it leads to a noticeable impact in SMFS.

The influence of PEG polydispersity on the SMFS data is illustrated in Figure 3A, which shows SMFS traces recorded with both PEG and ELP linkers and also shows example traces of the shortest and largest extensions found in a typical type III Coh:Doc data set. Figure 3B shows a histogram of extension values at which the first CBM unfolding event occurred. For ELPs, the distribution shows one peak centered at an extension value that is expected based on the known ELP linker length. In the case of the PEG experiment, however, three distinct populations are observed. This can be understood by considering that at the level of single molecules a polydisperse distribution results in discrete peaks representing the corresponding lengths of the discrete polymeric linkers on the cantilever tip. We interpret the distributions as being caused by three different PEG molecules with different lengths attached to the tip. Although the discrete distributions could conceivably be caused by different positions of the molecule attachment points to the AFM cantilever tip, this effect should be the same for ELPs. Moreover, varying linker lengths also reflect in varying steepness of the force−extension trace peaks, which would not occur simply because of attachment geometry (Figure 3A, PEG traces). We exclusively observed monomodal distributions for ELPs; therefore an anchor position effect seems not to play a major role. This polydispersity is clearly disadvantageous, since multiple linker lengths render data analysis more difficult. Curves cannot simply be overlaid in force−distance space due to varying loading rates. Furthermore, for constant-speed SMFS experiments, loading rate populations in dynamic force spectra will be broadened due to the probabilistic nature of the thermally driven rupture events.

We note that the PEG-modified surfaces are softer than ELP-modified surfaces during indentation of the tip into the polymer brush, as determined by the curvature at the beginning of each trace. The firmer ELP-modified surfaces require a lower indentation force to reach a linear force−distance regime after the initial soft indentation. For calibrating the inverse optical lever sensitivity, this is advantageous since high indentation forces can damage the molecules attached to the tip through adsorption and denaturation processes.[31]

Uniform ELP Stretching Behavior Minimizes Artifacts. We hypothesized that by replacing synthetic PEG linkers with biological ELP linkers, and thereby having a single type of polymer backbone throughout the mechanical system, better defined elasticity properties for the recording of force curves would be achievable. The persistence lengths of ELP peptide backbones should be comparable to those of unfolded protein

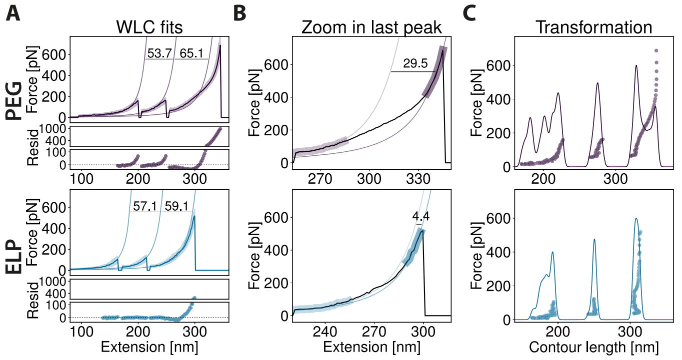


Figure 4. Elasticities of PEG and ELP linkers. (A) Superposition of multiple protein unfolding curves ("master curves") from SMFS experiments with PEG (purple, $N = 73$) and ELP linkers (blue, $N = 151$). The lower plots of each graph in panel A show the residuals of each WLC fit. Note that the residual plots are split into two subranges, shown in two windows from −35 to 120 pN (lower window) and from 120 to 1100 pN (upper window). The applied WLC model was extended by *ab initio* quantum mechanical calculations to correct for the enthalpic stretching of the polymer backbone.[33] Data were fitted with a fixed persistence length of 0.4 nm. The fits show that the stretching behavior of the mixed polymer system with PEG linkers deviates markedly at elevated forces from the predictions of the elasticity model, whereas the ELP curves agree reasonably well. (B) Final stretch and the Coh:Doc rupture event were fitted with the qmWLC model with two different contour lengths in the lower and upper force regime. The PEG molecules undergo a conformational transition,[16] resulting in different measured contour lengths for each force regime. For ELP molecules, a comparable transition was reported,[27,34] which apparently contributes to a much lower extent, so that SMFS experiments are much less affected. The differences in fitted contour length between the two fits are 29.5 nm for PEG linkers and 4.4 nm for ELP linkers. (C) Contour length transformations[29,35] of PEG and ELP master curves (purple and blue points). Ideally, the transformation results in data points aligning on vertical lines, where each line represents an energy barrier position for each stretching regime between two peaks in force−extension space. A KDE (Gaussian kernel, bandwidth: 2.5 nm) was calculated for the transformed data. The ELP data set showed the expected three peaks for the three unfolding and dissociation events, whereas the PEG data exhibit an irregular distribution with additional maxima.

domains, since they both consist of the same type of peptide-bonded polymer chains. This matching of the persistence length should be advantageous compared to PEG, which contains repeats of ethylene oxide groups with lower stiffness. Accurate description of the mechanical system under investigation by elasticity models plays a crucial role in determining characteristic parameters such as persistence lengths and contour length increments.

Previous studies had shown that at forces below 100 pN PEG elasticity may be satisfactorily described by standard elasticity models.[16] In a systematic study in this force range, we compared ELP and PEG linkers and corroborated these earlier results. The data and a thorough discussion thereof are given in the Supporting Information (see particularly Supplemental Figure S3).

At elevated forces, however, stretching of PEG through its conformational transition causes marked deviations from ideal polymer behavior. In aqueous environments, water molecules bridge neighboring ethylene oxide monomers by hydrogen bonding to two adjacent oxygen groups in the PEG backbone. By this means, water stabilizes the trans−trans−gauche configuration with a binding energy of around 3 kT. When PEG is stretched, however, the subunits of the backbone are forced increasingly into a slightly longer all-trans configuration and the bound water molecules are released. This conformational change, which contributes prominently to the polymer elasticity in the force range of up to ca. 300 pN, causes an increase in the measured net contour length of the polymer backbone.[16,17]

Figure 4A shows assemblies of multiple data traces ("master curves") of PEG- and ELP-linked proteins, respectively. The master curves are obtained by first aligning force−extension traces along the extension axis using an algorithm to maximize cross-correlation values in contour length space and then finding most probable force values of aligned traces in force distance space (see the Materials and Methods section). A recently introduced worm-like chain (WLC) approximation model[32] with an *ab initio* quantum mechanical correction for backbone stretching at high forces[33] (qmWLC) was then fitted to the traces with a fixed persistence length of 0.4 nm.

In the case of PEG linkers, a pronounced linear regime between 100 and 300 pN is visible in the last stretch prior to Coh:Doc rupture. As a consequence, the qmWLC cannot model this polymer correctly. ELPs do not show such a conformational change to this extent, and therefore the elasticity model fits satisfyingly. A fitting approach where the persistence length is also a free fit parameter is shown in Supplemental Figure S4. This approach misused the persistence length to compensate for the gauche-to-trans conformational change in the polymer; therefore, it resulted in largely unrealistic values for the contour length increments.

Figure 4B shows details of the last stretch before the Coh:Doc dissociation, highlighting the difference between PEG and ELP linkers. Two separate fits in the respective low- and high-force regimes illustrate the differences in polymer length before and after the conformational transition. We note that ELPs were also reported to have a force-induced conformational change, in this case based on proline cis−trans isomerization that also extends the contour length.[27,34] However, the low number of prolines in the overall sequence (every fifth amino acid) in the ELP motif renders this effect much smaller compared to the conformational change of PEG and will be camouflaged by signal noise in typical experiments with proteins.

Figure 4C shows the transformation into contour length space using the qmWLC model. A kernel density estimate (KDE) was used (Gaussian kernel, bandwidth of 2.5 nm) to generate smooth functions describing the contour length increments observed between unfolding or rupture events, which in this case included 2× CBM unfolding and Coh:Doc dissociation. In the case of PEG linkers, the KDE−contour length distribution shows several peaks. This is because of the failure of the qmWLC model to accurately describe the force response of the polymer. Determining the contour length increments between the peaks of the KDE proves problematic even for this relatively simple exemplary case of two large fingerprint unfolding events and a receptor ligand dissociation. Smaller unfolding steps or even folding intermediates, which appear as substeps, would be even harder to pinpoint with the PEG system. In the case of ELP-immobilized proteins, only three distinct peaks appear, with much more clearly identifiable contour length increments between the peaks.

CONCLUSION

PEG linkers have successfully been employed in numerous studies to anchor biomolecules of interest to surfaces for SMFS. In the low-force regime (below 100 pN) the extended WLC model describes their elastic properties with sufficient accuracy for the majority of applications. For elevated forces, however, the conformational transitions in the PEG backbone would necessitate further development of elasticity models for a convincing description.[16] Moreover, the inherent polydispersity of PEGs, together with their complex elasticity, complicates data analysis and reduces the amount of information that can be deduced from SMFS.

The ELP-based linkers, however, have proven in our studies to be significantly improved linker molecules for surface immobilization and passivation purposes in single-molecule force experiments. ELPs are monodisperse, are highly flexible, and readily allow for direct, site-specific tethering. We showed that these features lead to more accurate measurements of contour length increments in receptor−ligand polyprotein force spectroscopy experiments. A well-established elasticity model suffices for the data analysis.

Even at low forces, the PEG subunits already start to change their conformational state occupancy. At 50 pN, the probability for their elongated state is already above 10%.[16] Therefore, the findings we present here are also relevant for investigations at lower forces or in systems that should be analyzed over a large range of forces. PEG linkers may still deliver satisfying results, as long as data in similar force ranges can be compared. In some cases, elasticity parameters such as the Kuhn length or persistence length can heuristically compensate for effects not explicitly described by the model. As soon as different force ranges of multiple domains need to be compared, though, the varying proportions of elongated (all-trans) *versus* non-elongated (trans−trans−gauche) PEG subunits cannot simply be accounted for by the elasticity parameter, and therefore measured contour length increments get distorted. Different biochemical approaches like those described here are thus necessary to gain meaningful insights. These scenarios include, for example, shielded unfolding events or small substeps, where the force cannot drop sufficiently in between stretching events.

The ELPs investigated here represent only one formulation of the vast variety of smart polymer linkers that could be utilized in SMFS experiments. Further studies are required to evaluate other nonstructured, non-proline-containing protein linkers to determine their suitability for SMFS studies, since the amino acid side chain composition may affect the persistence length[36,37] or give rise to nonentropic behavior. Biotechnological characteristics, *i.e.*, recombinant production yields and ease of purification, are as important as the biophysical requirements, which renders the easily produced ELPs particularly attractive. Other smart polymers should be similarly accessible to perform as suitable alternatives. The reported approach can be applied to enhance SMFS studies with purified proteins on functionalized surfaces as shown here or alternatively to modify cantilevers for chemical recognition imaging and force spectroscopy on artificial membranes or cell surfaces. It can easily be adopted by standard molecular biology equipped laboratories to streamline the procedure and improve data quality for resolving smaller unfolding features with high accuracy. Studies on smart polymers as tethers for SMFS experiments might also help to develop environmentally responsive surfaces, which bear potential for exciting applications in the nanobiosciences.

MATERIALS AND METHODS

All reagents were at least of analytical purity grade and were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Carl Roth GmbH (Karlsruhe, Germany). All buffers were filtered through a 0.2 μm poly(ether sulfone) membrane filter (Nalgene, Rochester, NY, USA) prior to use. The pH of all buffers was adjusted at room temperature.

A 300 amino acid long ELP was the basis for the AFM linker constructs used in this study, and the underlying cloning and protein purification procedure of the ELP is described in detail elsewhere.[19] The ELP sequence was $[(\mathrm{VPGVG})_5\text{-}(\mathrm{VPGAG})_2\text{-}(\mathrm{VPGGG})_3]_6$ and is referred to as $\mathrm{ELP}_{120\ \mathrm{nm}}$.

Standard molecular biology laboratories capable of producing recombinant proteins are equally capable of expressing ELPs, since both rely on the same principles, reagents, and instrumentation. With our plasmids provided at Addgene, cloning can even be avoided and production of ELP linkers for protein immobilization can be performed right away.

Cloning. A detailed description of the cloning procedure of the constructs can be found in the Supporting Information (Figures S5−S11). ELP sequences used in this study, along with 40 nm length variants and binding handles, are deposited at Addgene and available upon request (Addgene accession numbers: 90472: Cys-$\mathrm{ELP}_{120\ \mathrm{nm}}$-LPETGG, 90475: Cys-$\mathrm{ELP}_{40\ \mathrm{nm}}$-LPETGG, 91571: GGG-$\mathrm{ELP}_{40\ \mathrm{nm}}$-Cys, 91572: GGG-$\mathrm{ELP}_{120\ \mathrm{nm}}$-Cys, 91697: CohIII-CBM-HIS-LPETGG, 91698: GGG-HIS-CBM-Xmod-DocIII).

Transformation of Cells. A 2 μL amount of Gibson assembly or ligation reaction transformed *DH5α* cells (Life Technologies GmbH, Frankfurt, Germany; 30 min on ice, 1 min at 42 °C, 1 h at 37 °C in SOC medium) was used. The cells were plated on 50 μg/mL kanamycin-containing LB agar and incubated overnight at 37 °C. Clones were analyzed with Colony PCR, and clones with amplicons of appropriate lengths were sent to sequencing.

Protein Expression. Chemically competent *E. coli NiCo21(DE3)* (New England Biolabs, Ipswich, MA, USA) were transformed with 50 ng of plasmid DNA for the expression of all constructs used in this study. Transformed cells were incubated in autoinduction ZYM-5052 media (for ELP containing constructs supplemented with 5 mg/mL proline, valine, and 10 mg/mL glycine; 100 μg/mL kanamycin) for 24 h (6 h at 37 °C, 18 h at 25 °C).[38] Expression cultures were harvested *via* centrifugation (6500*g*, 15 min, 4 °C), the supernatant was discarded, and the pellets were stored at −80 °C until further lysis.

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

Throughout the whole purification process, for ELPs containing a cysteine, 1 mM tris(2-carboxyethyl)phosphine (TCEP, Thermo Fisher Scientific Inc., Waltham, MA, USA) or 1 mM of dithiothreitol (DTT) was added to the respective buffers. Cell pellets with proteins containing no HIS-tag were solubilized in 50 mM Tris-HCl pH 7.5 (supplemented with cOmplete, EDTA-free protease inhibitor cocktail, Sigma-Aldrich, St. Louis, MO, USA), and all other pellets in lysis buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 10% (w/v) glycerol, 0.1% (v/v) Triton X-100, 5 mM $MgCl_2$, DNase I 10 μg/mL, lysozyme 100 μg/mL).

Cys-ELP$_{120\ nm}$-LPETGG and GGG-ELP$_{120\ nm}$-Cys were purified with the ITC method.[39] After resolubilization, the cells were lysed by sonication (Bandelin Sonoplus GM 70, tip: Bandelin Sonoplus MS 73, Berlin, Germany; 40% power, 30% cycle, 2 × 10 min). The cells were kept on ice during the sonication procedure. The soluble fraction was separated from the insoluble cell debris by centrifugation (15000*g*, 4 °C, 1 h). In a first heating step (60 °C, 30 min) of the supernatant, most of the *E. coli* host proteins precipitated. The fraction of the collapsed ELPs was resolubilized by cooling the suspension for 2 h to 4 °C on a reaction tube roller. The insoluble host proteins were pelleted by centrifugation (15000*g*, 4 °C, 30 min). Further purification steps were necessary to increase the purity of the ELP solution. This was done by repeated thermoprecipitation of the ELP followed by redissolution.

The ELP solution was clouded by adding 1 M acetate buffer (final concentration 50 mM, pH 2.5) and 2 M NaCl. A heating step (60 °C, 30 min) ensured all ELPs were collapsed. A hot centrifugation (3220*g*, 40 °C, 75 min) was necessary to separate the high-salt, low-pH solution from the ELP pellet, which was resolubilized in 50 mM Tris-HCl (pH 7.0) after discarding the supernatant. The solution was incubated for 2 h at 4 °C to resolubilize all ELPs completely. A cold centrifugation step (3220*g*, 4 °C, 60 min) isolated the remaining insoluble fraction of the suspension. After decanting the supernatant, the salt concentration was increased and pH lowered, to precipitate the ELPs again. This cycle was repeated three times or extended if the purity of the solution was not high enough.

The constructs CohIII-CBM-HIS-LPETGG and GGG-HIS-CBM-Xmod-DocIII were expressed and lysed as described above. After the first centrifugation, the supernatant was, however, filtered (0.45 μm) and applied to a HisTrap FF (GE Healthcare Europe GmbH, Freiburg, Germany). Unspecifically bound proteins on the column were removed by washing five column volumes (25 mM Tris-HCl pH 7.8, 500 mM NaCl, 20 mM imidazole, Tween 20 0.25% (v/v), 10% (v/v) glycerol). Finally, the desired HIS-tag containing protein was eluted (25 mM Tris-HCl pH 7.8, 500 mM NaCl, 300 mM imidazole, Tween 20 0.25% (v/v), 10% (v/v) glycerol).

For long-term storage the protein solutions of the different constructs were concentrated (Amicon Ultra-15 centrifugal filter units 10K MWCO, Merck KGaA, Darmstadt, Germany) and reduced with 5 mM TCEP overnight (at 4 °C) for constructs that contained a cysteine. The buffer of the reduced ELP solution was exchanged (Zeba spin desalting columns 7K, Thermo Fisher Scientific Inc.) to 50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, with a pH of 7.2, and 10% (v/v) glycerol and flash frozen in liquid nitrogen in small aliquots to be stored at −80 °C. All other proteins were exchanged with 25 mM Tris-HCl, 75 mM NaCl, and 5 mM $CaCl_2$ with a pH of 7.2 and supplemented with a final glycerol concentration of 20% (v/v). No loss of functionality of the ELPs (cross-linking and passivation capability) could be detected, when stored buffered or lyophilized in small aliquots at −80 °C, over the duration of more than one year.

SDS-PAGE (Any kD Mini-PROTEAN stain-free gels, Bio-Rad Laboratories GmbH, Hercules, CA, USA) was employed to detect any impurities. Since ELPs could not be stained with the stain-free technology, an Alexa Fluor 647-C_2-maleimide dye (Thermo Fisher Scientific Inc.) was incubated for 1 h at room temperature with the ELP solution. An appropriately diluted protein solution was mixed with 5× loading buffer (250 mM Tris-HCl, pH 8.0, 7.5% (w/v) SDS, 25% (v/v) glycerol, 0.25 mg/mL bromophenol blue, 12.5% (v/v) 2-mercaptoethanol) and heated for 5 min at 95 °C.

ELP concentration was photometrically determined at 205 nm (Ultrospec 3100 Pro, Amersham Biosciences, Amersham, England, and TrayCell, Hellma GmbH & Co. KG, Müllheim, Germany). For all other constructs an absorption measurement at 280 nm led to the concentration (NanoDrop UV−vis spectrophotometer, Thermo Fisher Scientific Inc.). The extinction coefficient was determined theoretically for ELPs at 205 nm[40] and 280 nm[41] for all other fusion proteins.

AFM Sample Preparation. Force spectroscopy samples, measurements, and data analysis were prepared and performed according to previously published protocols.[10,35] Silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation, Tokyo, Japan; nominal spring constant: 100 pN/nm; 25 kHz resonance frequency in water) were used as force probes. Surface chemistry for cantilevers was similar to that for coverslips (Menzel Gläser, Braunschweig, Germany; diameter 24 mm). Surfaces were amino-silanized with 3-(aminopropyl)dimethylethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany). α-Maleinimidohexanoic-ω-NHS PEG (NHS-PEG-Mal, Rapp Polymere, Tübingen, Germany; PEG-MW: 15 kDa) was used as a linker for the sortase peptides (GGGGG-C and C-LPETGG, Centic Biotec, Heidelberg, Germany) in PEG-linked experiments. The cysteine-containing ELPs were linked to the surface with a sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate) cross-linker (sulfo-SMCC, Thermo Fisher Scientific Inc.). PEG or cross-linker (10 mM) was dissolved in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5.

Sortase-catalyzed coupling of the fingerprint molecules (GGG-CBM-Xmod-DocIII and CohIII-CBM-LPETGG) was done in 25 mM Tris-HCl, pH 7.2, 5 mM $CaCl_2$, and 75 mM NaCl at 22 °C for 2 h. Typically, 50 μM ELP or sortase peptide was coupled with 25 μM fingerprint molecule and 2 μM sortase enzyme.

In between both of the cross-linking steps (PEG, SMCC, or ELP, peptide reaction) surfaces were rinsed with water and dried with nitrogen. After immobilization of the fingerprint molecules, surfaces were rinsed in measurement buffer (25 mM Tris-HCl, pH 7.2, 5 mM $CaCl_2$, 75 mM NaCl). The reaction of the different surface chemistry was done spatially separated by using silicone masks (CultureWell reusable gaskets, Grace Bio-Laboratories, Bend, OR, USA). The mask was applied after silanization and removed under buffer after the last immobilization step.

AFM-SMFS Measurements. Data were taken on custom-built instruments (MFP-3D AFM controller, Oxford Instruments Asylum Research, Inc., Santa Barbara, CA, USA; piezo nanopositioners: Physik Instrumente GmbH & Co. KG, Karlsruhe, Germany, or Attocube Systems AG, Munich, Germany).

Instrument control software was custom written in Igor Pro 6.3 (Wavemetrics Inc., Portland, OR, USA). Piezo position was controlled with a closed-loop feedback system running internally on the AFM controller field-programmable gate array. A typical AFM measurement took about 12 h and was done fully automated and at room temperature. Retraction velocity for constant-speed force spectroscopy measurements was 0.8 μm/s. Cantilever spring constants were calibrated after completing all measurements on different spots on the surface using the same cantilever. This was done by utilizing the thermal method applying the equipartition theorem to the one dimensionally oscillating lever.[31,42]

Force−Extension Data Analysis. Obtained data were analyzed with custom-written software in Python (Python Software Foundation, Python Language Reference, version 2.7, available at http://www.python.org), utilizing the libraries NumPy, SciPy, and Matplotlib.

Raw voltage data traces were transformed into force distance traces with their respective calibration values after determining the zero force value with the baseline position. A correction of the force-dependent cantilever tip *z*-position was carried out. Force distance traces were filtered for traces showing two CBM unfoldings and a subsequent type III cohesin−dockerin dissociation, without preceding Xmodule unfolding.[7] This screening was carried out by detecting maximum-to-maximum distances of kernel density estimate (Gaussian kernel, bandwidth 1 nm) peaks in contour length space in each single trace, after applying thresholds for force, distance, and number of peaks. For sorting data sets, transformation of force distance data into contour length space was done with a manually fixed persistence length of 0.4 nm, to measure distances of energy barrier positions.[29,43] Sorting was done allowing generous errors to the expected increments to account for the conformational stretching of the spacer molecules. Fits to the force−extension data with the WLC model had the following parameters additionally to the values mentioned in the figure captions, if not stated otherwise: initial guess for persistence length: 0.4 nm; fit precision: 1×10^{-7}. For assessment of transformation quality, the inverse worm-like-chain model was applied for transformation of force distance traces into the contour length space in a force window of 10 to 125 pN and with a persistence length previously fitted to each peak separately: The global mean value of each data set for each peak was used. Final alignments of the whole data sets were assembled by cross-correlation.

Master Curve Assembly. The master curves were assembled by cross-correlation of each force−distance trace of a presorted data set with all previous curves in contour length space, starting with a random curve. Each curve was shifted on its x axis to fit the maximum correlation value and added to the set assembly in contour length space. Subsequently, a second run was performed, cross-correlating each curve with the previously assembled set, to facilitate an equal correlation template for every curve, independent of its occurrence. Finally, the most probable shift was calculated by a KDE and subtracted from each curve to get representative absolute distances with respect to the origin. Distance and correlation value thresholds were applied to filter out less probable PEG populations and otherwise badly fitting data. In a final step, all overlaid raw data points in force−distance space were binned on the x axis into nanometer-sized slices, and their densities on the y axis were estimated by a KDE for each slice. Near the rupture events, where the kernel density estimates cannot unambiguously identify maxima of the data slices, the value was set to zero. Therefore, after each rupture, a small "gap" is visible, which was not included in data points used for fitting. Their most probable value and the corresponding full width at half-maxima then assembled the master curve. Although by this procedure representative absolute rupture forces for the domains are not necessarily reproduced to the highest accuracy, the most probable and most representative pathway of the elastic behavior in between peaks is resembled well.

qmWLC model. For WLC fits and transformations into contour length space, a recently improved approximation, solved for the extension, was used,[32] adding correction terms for quantum mechanical backbone stretching.[33]

With the abbreviations

$$f = FL_p/kT \tag{1}$$

$$b = \exp\left(\sqrt[4]{\frac{900}{f}}\right) \tag{2}$$

WLC fits were done with the model formula

$$x = L_{\mathrm{corr}}\left(\frac{4}{3} - \frac{4}{3\sqrt{f+1}} - \frac{10b}{\sqrt{f}(b-1)^2} + \frac{f^{1.62}}{3.55 + 3.8f^{2.2}}\right) \tag{3}$$

With the quantum mechanical correction,

$$L_{\mathrm{corr}} = \frac{L_{c,0}}{2y_2}\left(\sqrt{4Fy_2 + y_1^2} - y_1 + 2y_2\right) \tag{4}$$

where y_1 and y_2 are the *ab initio* parameters from the original publication.

Transformations were performed with the model contour length:

$$L_c = \frac{x}{\frac{4}{3} - \frac{4}{3\sqrt{f+1}} - \frac{10b}{\sqrt{f}(b-1)^2} + \frac{f^{1.62}}{3.55 + 3.8f^{2.2}}} \tag{5}$$

With the reverse quantum mechanical correction for zero force contour length,

$$L_{c,0} = \frac{L_c}{\frac{1}{2y_2}\left(\sqrt{y_1^2 + 4y_2F} + 2y_2 - y_1\right)} \tag{6}$$

with x being the extension, L_c the model contour length, F the force, L_p the persistence length, k Boltzmann's constant, T the temperature, y_1 and y_2 the quantum mechanical correction parameters, L_{corr} the qm-corrected contour length, and $L_{c,0}$ the reverse qm-corrected contour length at zero force. As a nonlinear fitting algorithm, a Levenberg−Marquardt least-squares minimization method was applied.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsnano.7b02694.

> Further details on experimental methods, supplementary results, and sequence information (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: gaub@lmu.de.

ORCID

Hermann E. Gaub: 0000-0002-4220-6088

Author Contributions

$^{\perp}$W. Ott and M. A. Jobst contributed equally to this work.

Author Contributions

W.O.: experiment design, sample preparation, measurements, data analysis, writing of manuscript; M.A.J.: experiment design, data analysis, writing of manuscript; M.S.B.: data analysis; E.D.: sample preparation; L.F.M.: data analysis; M.A.N.: experiment design, writing of manuscript; H.E.G.: experiment design, writing of manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Advanced Grant "Cellufuel" of the European Research Council and the Deutsche Forschungsgemeinschaft through SFB 1032. M.A.N. acknowledges support from an ERC Starting Grant "Molecular Mechanical Adhesives" number 715207 and from Society in Science−the Branco Weiss Fellowship from ETH Zurich. We thank T. Verdorfer and C. Schoeler for proofreading and helpful discussions.

REFERENCES

(1) Cao, Y.; Li, H. Engineered Elastomeric Proteins with Dual Elasticity Can Be Controlled by a Molecular Regulator. *Nat. Nanotechnol.* **2008**, *3*, 512−516.

(2) Lv, S.; Dudek, D. M.; Cao, Y.; Balamurali, M. M.; Gosline, J.; Li, H. Designed Biomaterials to Mimic the Mechanical Properties of Muscles. *Nature* **2010**, *465*, 69−73.

(3) Rivas-Pardo, J. A.; Eckels, E. C.; Popa, I.; Kosuri, P.; Linke, W. A.; Fernández, J. M. Work Done by Titin Protein Folding Assists Muscle Contraction. *Cell Rep.* **2016**, *14*, 1339−1347.

(4) Ott, W.; Jobst, M. A.; Schoeler, C.; Gaub, H. E.; Nash, M. A. Single-Molecule Force Spectroscopy on Polyproteins and Receptor−ligand Complexes: The Current Toolbox. *J. Struct. Biol.* **2017**, *197*, 3−12.

(5) Bull, M. S.; Sullan, R. M. A.; Li, H.; Perkins, T. T. Improved Single Molecule Force Spectroscopy Using Micromachined Cantilevers. *ACS Nano* **2014**, *8*, 4984−4995.

(6) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. Single-Molecule Dissection of the High-

Affinity Cohesin-Dockerin Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20431−20436.

(7) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; E, G. H.; Nash, M. A. Ultrastable Cellulosome-Adhesion Complex Tightens under Load. *Nat. Commun.* **2014**, *5*, 1−8.

(8) Baumann, F.; Bauer, M. S.; Milles, L. F.; Alexandrovich, A.; Gaub, H. E.; Pippig, D. A. Monovalent Strep-Tactin for Strong and Site-Specific Tethering in Nanospectroscopy. *Nat. Nanotechnol.* **2015**, *11*, 89−94.

(9) Milles, L. F.; Bayer, E. A.; Nash, M. A.; Gaub, H. E. Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium Perfringens. *J. Phys. Chem. B* **2017**, *121*, 3620−3625.

(10) Zimmermann, J. L.; Nicolaus, T.; Neuert, G.; Blank, K. Thiol-Based, Site-Specific and Covalent Immobilization of Biomolecules for Single-Molecule Experiments. *Nat. Protoc.* **2010**, *5*, 975−985.

(11) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E690−E697.

(12) Popa, I.; Rivas-Pardo, J. A.; Eckels, E. C.; Echelman, D.; Valle-Orero, J.; Fernandez, J. M. A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J. Am. Chem. Soc.* **2016**, *138*, 10546−10553.

(13) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M. Nanomechanics of HaloTag Tethers. *J. Am. Chem. Soc.* **2013**, *135*, 12762−12771.

(14) Pippig, D. A.; Baumann, F.; Strackharn, M.; Aschenbrenner, D.; Gaub, H. E. Protein-DNA Chimeras for Nano Assembly. *ACS Nano* **2014**, *8*, 6551−6555.

(15) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. From Genes to Protein Mechanics on a Chip. *Nat. Methods* **2014**, *11*, 1127−1130.

(16) Oesterhelt, F.; Rief, M.; Gaub, H. E. Single Molecule Force Spectroscopy by AFM Indicates Helical Structure of Poly(ethylene-Glycol) in Water. *New J. Phys.* **1999**, *1*, 1−11.

(17) Liese, S.; Gensler, M.; Krysiak, S.; Schwarzl, R.; Achazi, A.; Paulus, B.; Hugel, T.; Rabe, J. P.; Netz, R. R. Hydration Effects Turn a Highly Stretched Polymer from an Entropic into an Energetic Spring. *ACS Nano* **2017**, *11*, 702−712.

(18) Xue, Y.; Li, X.; Li, H.; Zhang, W. Quantifying Thiol-Gold Interactions towards the Efficient Strength Control. *Nat. Commun.* **2014**, *5*, 4348.

(19) Ott, W.; Nicolaus, T.; Gaub, H. E.; Nash, M. A. Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation. *Biomacromolecules* **2016**, *17*, 1330−1338.

(20) Tang, N. C.; Chilkoti, A. Combinatorial Codon Scrambling Enables Scalable Gene Synthesis and Amplification of Repetitive Proteins. *Nat. Mater.* **2016**, *15*, 419−424.

(21) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules* **2010**, *11*, 944−952.

(22) Meyer, D. E.; Chilkoti, A. Genetically Encoded Synthesis of Protein-Based Polymers with Precisely Specified Molecular Weight and Sequence by Recursive Directional Ligation: Examples from the Elastin-like Polypeptide System. *Biomacromolecules* **2002**, *3*, 357−367.

(23) Gray, W. R.; Sandberg, L. B.; Foster, J. A. Molecular Model for Elastin Structure and Function. *Nature* **1973**, *246*, 461−466.

(24) Roberts, S.; Dzuricky, M.; Chilkoti, A. Elastin-like Polypeptides as Models of Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589*, 2477−2486.

(25) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. Reprogramming the Specificity of Sortase Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13343−13348.

(26) Urry, D. W.; Hugel, T.; Seitz, M.; Gaub, H. E.; Sheiba, L.; Dea, J.; Xu, J.; Parker, T. Elastin: A Representative Ideal Protein Elastomer. *Philos. Trans. R. Soc., B* **2002**, *357*, 169−184.

(27) Valiaev, A.; Lim, D. W.; Oas, T. G.; Chilkoti, A.; Zauscher, S. Force-Induced Prolyl Cis-Trans Isomerization in Elastin-like Polypeptides. *J. Am. Chem. Soc.* **2007**, *129*, 6491−6497.

(28) Valiaev, A.; Dong, W. L.; Schmidler, S.; Clark, R. L.; Chilkoti, A.; Zauscher, S. Hydration and Conformational Mechanics of Single, End-Tethered Elastin-like Polypeptides. *J. Am. Chem. Soc.* **2008**, *130*, 10939−10946.

(29) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. Comparing Proteins by Their Unfolding Pattern. *Biophys. J.* **2008**, *95*, 426−434.

(30) Dietz, H.; Rief, M. Exploring the Energy Landscape of GFP by Single-Molecule Mechanical Experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16192−16197.

(31) Proksch, R.; Schäffer, T. E.; Cleveland, J. P.; Callahan, R. C.; Viani, M. B. Finite Optical Spot Size and Position Corrections in Thermal Spring Constant Calibration. *Nanotechnology* **2004**, *15*, 1344−1350.

(32) Petrosyan, R. Improved Approximations for Some Polymer Extension Models. *Rheol. Acta* **2017**, *56*, 21−26.

(33) Hugel, T.; Rief, M.; Seitz, M.; Gaub, H. E.; Netz, R. R. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments versus *Ab-Initio* Theory. *Phys. Rev. Lett.* **2005**, *94*, 048301.

(34) Valiaev, A.; Lim, D. W.; Schmidler, S.; Clark, R. L.; Chilkoti, A.; Zauscher, S. Hydration and Conformational Mechanics of Single, End-Tethered Elastin-like Polypeptides. *J. Am. Chem. Soc.* **2008**, *130*, 10939−10946.

(35) Jobst, M. A.; Schoeler, C.; Malinowska, K.; Nash, M. A. Investigating Receptor-Ligand Systems of the Cellulosome with AFM-Based Single-Molecule Force Spectroscopy. *J. Visualized Exp.* **2013**, e50950.

(36) Stirnemann, G.; Giganti, D.; Fernandez, J. M.; Berne, B. J. Elasticity, Structure, and Relaxation of Extended Proteins under Force. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 3847−3852.

(37) Cheng, S.; Cetinkaya, M.; Gräter, F. How Sequence Determines Elasticity of Disordered Proteins. *Biophys. J.* **2010**, *99*, 3863−3869.

(38) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expression Purif.* **2005**, *41*, 207−234.

(39) MacEwan, S. R.; Hassouneh, W.; Chilkoti, A. Non-Chromatographic Purification of Recombinant Elastin-like Polypeptides and Their Fusions with Peptides and Proteins from Escherichia coli. *J. Visualized Exp.* **2014**, e51583.

(40) Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Sci.* **2013**, *22*, 851−858.

(41) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.; Appel, R.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protocols Handbook* **2005**, 571−607.

(42) Hutter, J. L.; Bechhoefer, J. Calibration of Atomic-Force Microscope Tips. *Rev. Sci. Instrum.* **1993**, *64*, 1868−1873.

(43) Jobst, M. A.; Milles, L. F.; Schoeler, C.; Ott, W.; Fried, D. B.; Bayer, E. A.; Gaub, H. E.; Nash, M. A. Resolving Dual Binding Conformations of Cellulosome Cohesin- Dockerin Complexes Using Single-Molecule Force Spectroscopy. *eLife* **2015**.

Supporting Information: Elastin-Like Polypeptide Linkers for Single Molecule Force Spectroscopy

Wolfgang Ott[a,b,⊥], Markus A. Jobst[a, ⊥], Magnus S. Bauer[a], Ellis Durner[a], Lukas F. Milles[a], Michael A. Nash[c,d], Hermann E. Gaub[a,#]

[a] Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, 80799 Munich, Germany.

[b] Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany.

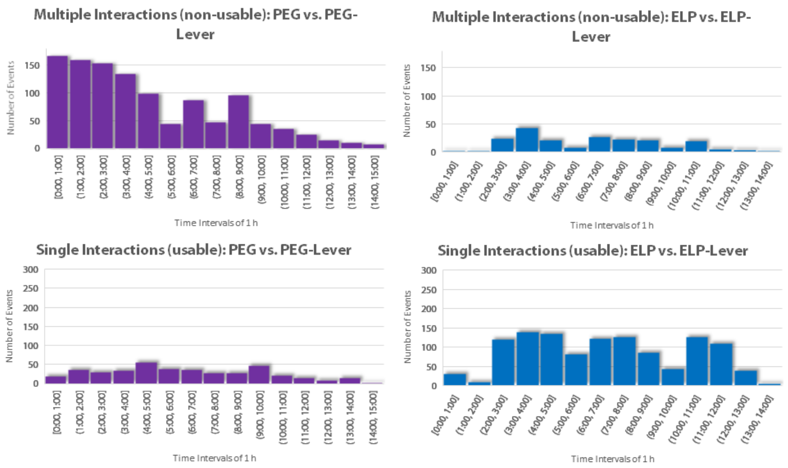
[c] Department of Chemistry, University of Basel, 4056 Basel, Switzerland.

[d] Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland.

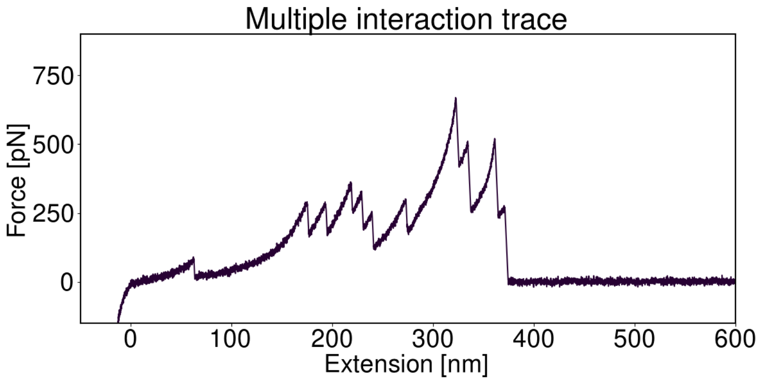
[⊥] These authors contributed equally to this work

[#] Corresponding author: gaub@lmu.de

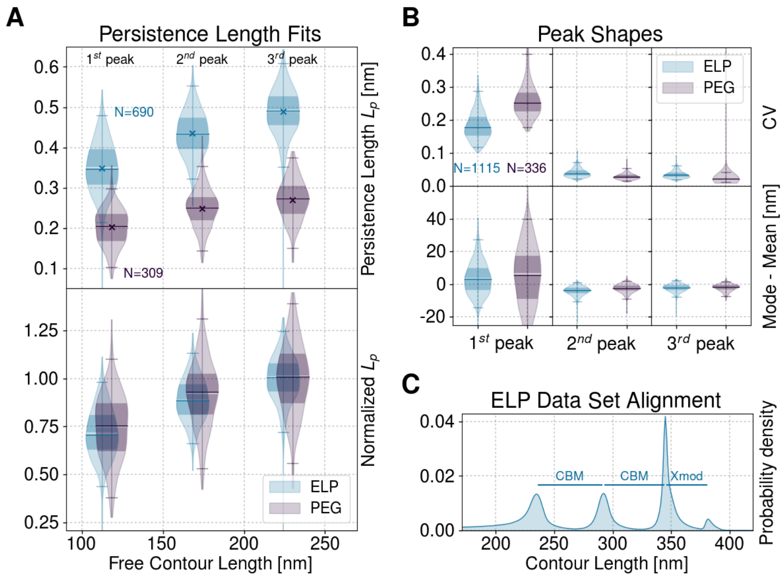
Supporting Information



Supplemental Figure S1. Number of curves within a 1 h timeframe were binned in one histogram bar. Multiple traces were traces with more than 10 peaks (**Supplemental Figure S2** shows an exemplary multiple interaction trace). Left (purple) is the PEG-lever versus the PEG-immobilization and right (blue) ELP-lever versus ELP-immobilization. The two top panels show number of multiple interactions over time. The bottom panels show number of single specific interactions over time.



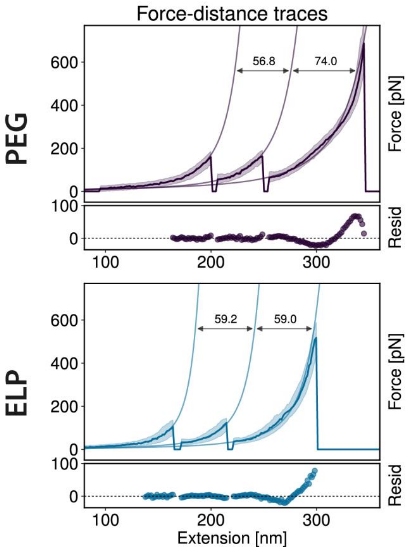
Supplemental Figure S2. A typical example trace displaying multiple interactions.



Supplemental Figure S3: Performance of contour length transformations. (A) Observed persistence lengths. Upper plot: observed persistence lengths preceding each CBM and Coh:DocIII unfolding/rupture peak as measured by WLC fits in the force range of 30 to 125 pN (ELP: 0.35, 0.44, and 0.49 nm; PEG: 0.20, 0.25, and 0.27 nm). Lower plot: same data normalized to the respective last peak means. The qualitative behavior over the unfolding of the peaks is similar for both constructs. **(B)** Assessment of transformation quality. Coefficient of variation (CV) as a measure of distribution broadness and distance of mode to mean as a measure of peak symmetry show better performance for ELP data for the first peaks. Later peaks show better performance of PEG data, although the differences are negligible. Transformations were done with the inverse WLC model only for data points between 10 and 125 pN. Persistence lengths for the transformations were chosen as the mean values of the WLC fits to each peak as shown in panel (A). **(C)** Alignment of transformed ELP curves in contour length space. Two CBM increments and one Xmod unfolding prior to Coh:Doc rupture are clearly detectable.

Low force performance of ELP linkers

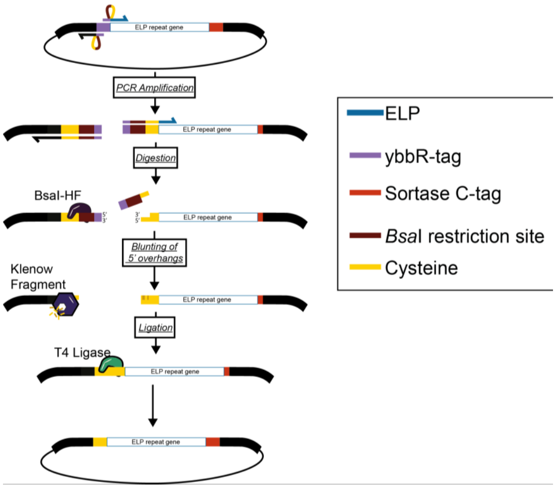
For this analysis, only forces in a range from 10 to 125 pN were taken into account, to minimize the effects of conformational stretching. The elastic properties of the first stretching event of a data trace are dominated by the linker molecules. As more protein domains unfold, the peptide backbone of the unfolded domains contributes increasingly to the overall elastic response. Contour length transformations of force distance data were performed with the mean fitted persistence lengths of each peak, as shown in **Supplemental Figure S3, Panel A** (0.35, 0.44, and 0.49 nm for ELP data peaks; 0.20, 0.25, and 0.27 nm for PEG data peaks), to account for varying persistence lengths over the course of each pulling cycle. The persistence length as a measure for the stiffness of a polymer is lower for PEG than for ELP with bulky side chains and rotational restrictions of the peptide backbone. Comparable changes of persistence lengths over the course of an unfolding experiment were also observed earlier in other studies.[1,2] The distribution width and asymmetry of each peak in contour length space were evaluated separately by the coefficient of variation and the calculated difference of statistical mode and mean. A comparison of all datasets revealed that for the first unfolding peak, ELP datasets display slightly superior properties: the first peak for data with ELP linker tethering is sharper and more symmetric (**Supplemental Figure S3, Panel B**) as indicated by the narrower distribution and lower coefficient of variation. For the subsequent peaks 2 and 3, both PEG and ELP linkers perform similarly and the differences become negligibly small. Although the impact on data quality in this low force regime examined here, was not as severe as expected, ELP linkers seem to exhibit advantageous behavior for the first stretching events of each curve, and might improve accuracy in determining the following contour length increments to identify protein domains.



Supplemental Figure S4: Master curves fits with persistence lengths as an additional free fit parameter. If the persistence length is not kept fix, but also fitted to the data, it is clearly visible, that this parameter is optimized to compensate the conformational stretching effect for PEG datasets. While the qmWLC model fit itself looks better and has lower residuals compared to the fixed persistence length fit, the resulting contour length increment is way off and does not yield any meaningful value, rendering the model useless to extract information from the data. The two CBM domains have the exact same amino acid sequence and therefore should show the same contour length increments upon unfolding.

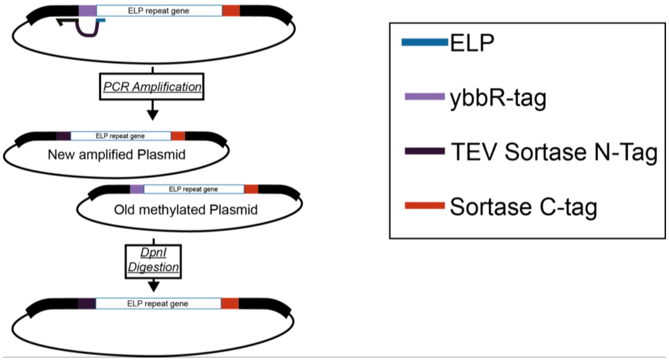
Cloning of ELP linkers. Standard PCR was used for amplification of DNA (Phusion High-Fidelity PCR Master Mix, Thermo Fisher Scientific Inc., Waltham, MA, USA). Melting temperatures were adjusted according to the employed primers (see **Table S1**, below).

A plasmid encoding ybbR-ELP$_{120\ nm}$-LPETGG described earlier[3] was modified to yield the plasmid Cys-ELP$_{120\ nm}$-LPETGG. PCR amplification of the plasmid with primers annealing at and downstream of the ybbR-tag was the first step (**Supplemental Figure S5**). The gene for the ELP is a highly repetitive sequence, hence it was necessary to anneal the forward primer at the ybbR-tag to create a unique attachment site. Since the ybbR-tag had to be removed, a *Bsa*I restriction site was incorporated with a primer downstream of the annealing region of the forward primer. The reverse primer had a cysteine encoded at its 5' end. After successful PCR amplification, the product was digested (*Bsa*I and *Dpn*I) and blunted (1h, 37°C, 5 Min, 80°C). The blunting reaction was performed in parallel with 1 µl of Klenow Fragment enzyme and the addition of 1 mM dNTPs (Thermo Fisher Scientific Inc., Waltham, MA, USA)).
After purification (QIAquick PCR purification kit or gel extraction kit (Qiagen GmbH, Hilden, Germany) the ligation reaction was set up: 1 µl of a T4 Ligase (10U/µl, Thermo Fisher Scientific Inc., Waltham, MA, USA was supplemented with 1 µl ATP (10 mM), 0.5 µl PEG-6000, 1 µl T4 Polynucleotide Kinase (PNK) and buffered in CutSmart buffer (New England Biolabs, Ipswich, MA, USA).



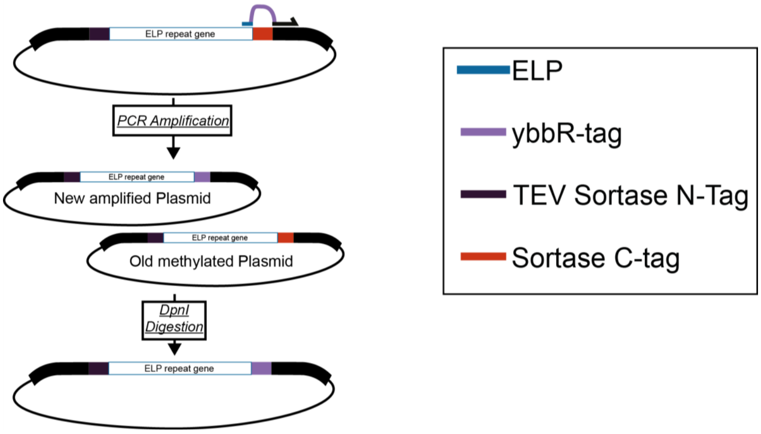
Supplemental Figure S5. Cloning scheme for Cys-ELP$_{120\ nm}$-LPETGG.

For the creation of the TEV-GGG-ELP$_{60\ \mathrm{nm}}$-LPETGG plasmid, a plasmid encoding ybbR-ELP$_{60\ \mathrm{nm}}$-LPETGG[1] was mutated with one QuikChange primer[4], annealing up- and downstream of the ybbR-tag introducing DNA encoding a TEV-site and a triple glycine. The TEV cleavage site was introduced to ensure full cleavage of the N-terminal methionine. This was assumed to be necessary, since Sortase A only works with glycines at the very N-terminal start of a protein. The QuikChange reaction was done with 50 ng DNA template, 1 µl of primer (10 pmol/µl) in 20 µl Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA, see **Supplemental Figure S6**).



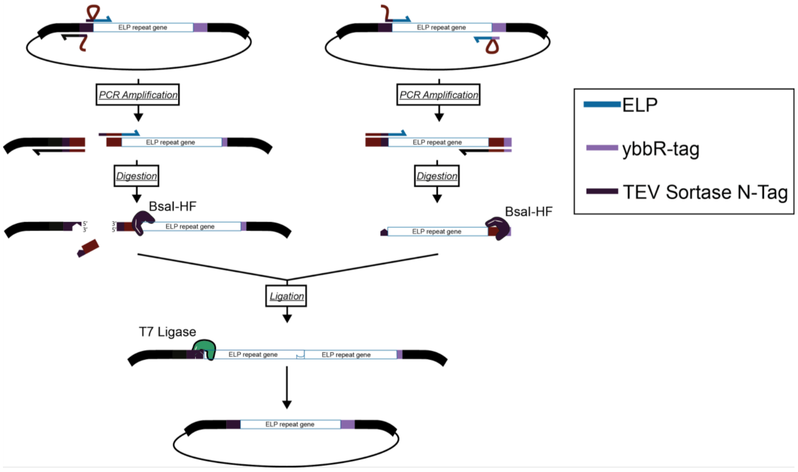
Supplemental Figure S6. Cloning scheme for TEV-GGG-ELP$_{60\ \mathrm{nm}}$-LPETGG.

The newly obtained plasmid was modified again with QuikChange to exchange the C-terminal Sortase-tag with a ybbR-tag (**Supplemental Figure S7**).



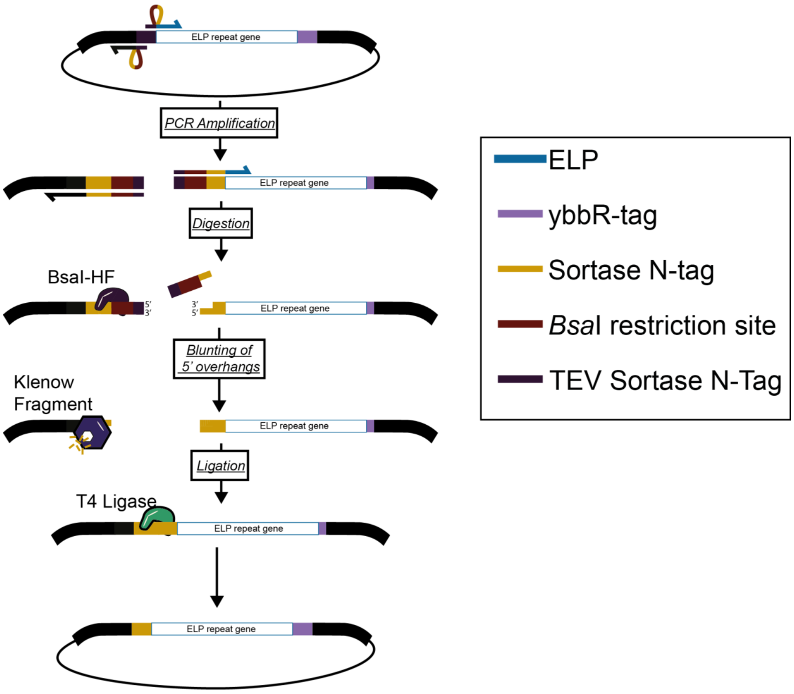
Supplemental Figure S7. Cloning scheme for TEV-GGG-ELP$_{60\ nm}$-ybbR.

The ELP gene cassette was duplicated by insertion of a gene sequence encoding [(VPGVG)$_5$-(VPGAG)$_2$-(VPGGG)$_3$]$_3$ into the linearized vector containing TEV-GGG-ELP$_{60\ \mathrm{nm}}$-ybbR. This was done by GoldenGate cloning.[5] For this purpose, both vector and insert were amplified with primers encoding flanking *Bsa*I restriction sites. The *Bsa*I sites were designed to match the corresponding end of insert and backbone, without leaving any cloning scars. After *Bsa*I digestion and purification of the PCR product *via* gel extraction, both of the parts were ligated with their corresponding sticky ends (2.5 µl CutSmart buffer, 1.25 µl T7 ligase, 2.5 µl ATP (10 mM); New England Biolabs, Ipswich, MA, USA, see **Supplemental Figure S8**).



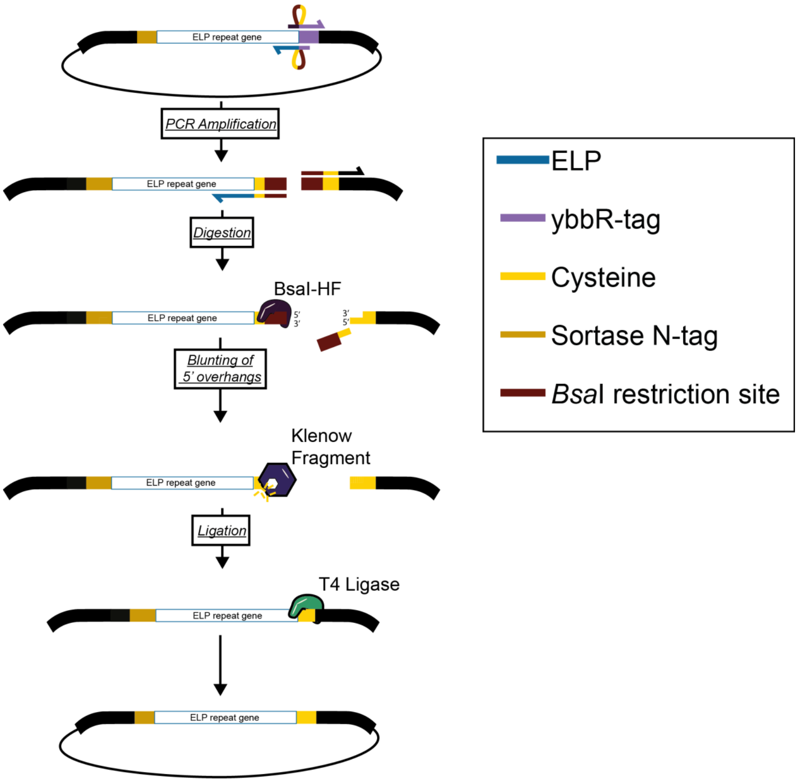
Supplemental Figure S8. Cloning scheme for TEV-GGG-ELP$_{120\ \mathrm{nm}}$-ybbR.

Experiments showed that the *E. coli* methionine aminopeptidases already fully digested the N-terminal methionine proceeding the polyglycine. Hence, removal of the TEV cleavage site was desired to simplify the ELP production process. This was achieved by a linearization reaction, *Bsa*I digestion and religation as described above. Primers were designed to anneal at the TEV-site and encoded a *Bsa*I restriction site upstream of the triple glycine (**Supplemental Figure S9**).


Construction of GGG-ELP$_{120nm}$-ybbR

Supplemental Figure S9. Cloning scheme for GGG-ELP$_{120\ nm}$-ybbR

Finally, the C-terminal ybbR-tag was switched to a cysteine. The reverse primer attached at the codons of the ybbR-tag with a *Bsa*I restriction site. The forward primer encoded a cysteine at its 5' end and annealed downstream of the stop codon. The linear plasmid was processed as described above (**Supplemental Figure S10**).

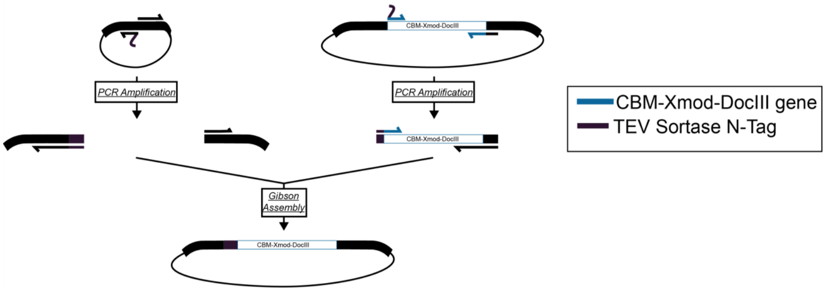


Supplemental Figure S10. Cloning scheme for GGG-ELP$_{120\ \mathrm{nm}}$-Cys

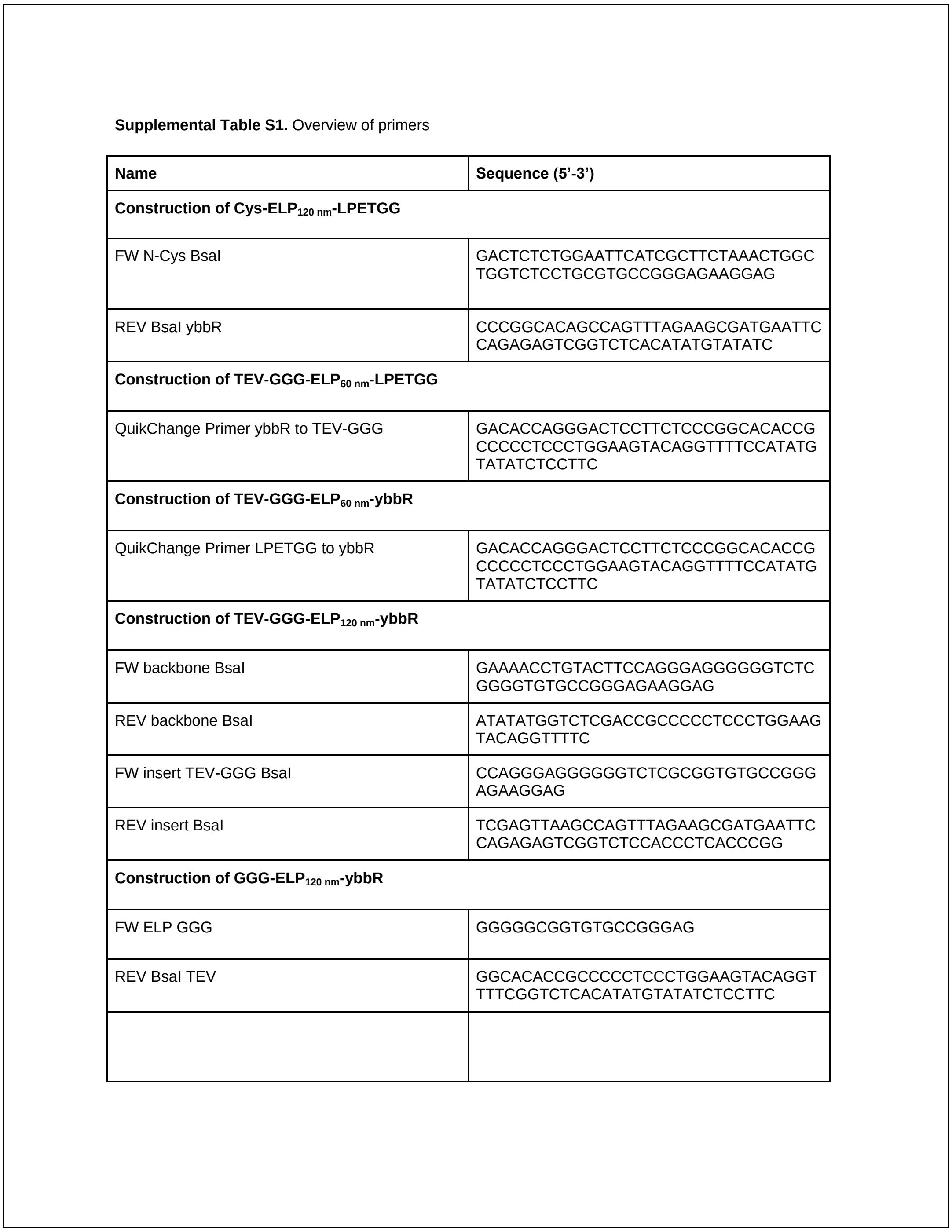
Cloning of GGG-HIS-CBM-Xmod-DocIII and CohIII-CBM-HIS-LPETGG.
Basis for the construction were two plasmids published by Schoeler *et al.*[6] The plasmid encoding the gene for CohIII-CBM was linearized with primers encoding the Sortase C-tag. 4.5 µl of the PCR product was directly digested with 1 µl *Dpn*I (Thermo Fisher Scientific Inc., Waltham, MA, USA), 3’ ends were phosphorylated with 1 µl T4 PNK (New England Biolabs, Ipswich, MA, USA) and the ends were religated with 1 µl T4 Ligase (10U/µl, Thermo Fisher Scientific Inc., Waltham, MA, USA) (15 Min at 37°C, 45 Min 22°C). The 10 µl reaction was supplemented with 1 µl ATP (10 mM), 0.5 µl PEG-6000 and 1 µl CutSmart buffer (10x, New England Biolabs, Ipswich, MA, USA).

The plasmid encoding the CohIII domain had a cloning scar (encoding the amino acids “GT”) at the N-terminus. Glycine and threonine were removed since one single glycine is already reactive with the “LPETGG” in a Sortase A catalyzed reaction. This was done with a sequential linearization and religation reaction (as described above).

The CBM-Xmod-DocIII gene was subcloned with Gibson Assembly into a linearized vector with a TEV site followed by a Sortase N-tag. 10 µl of the HiFi MasterMix (2x, New England Biolabs, Ipswich, MA, USA), were mixed with a 10-fold molar excess of insert to the backbone (reaction volume 20 µl, 1 hr, 50°C; **Supplemental Figure S11**). Similar to the GGG-ELP$_{120\ nm}$-Cys, the unnecessary TEV site was removed, since *E. coli* already digested the N-terminal methionine sufficiently. This was achieved by employing the same procedure as described for CohIII-CBM linearization and religation.

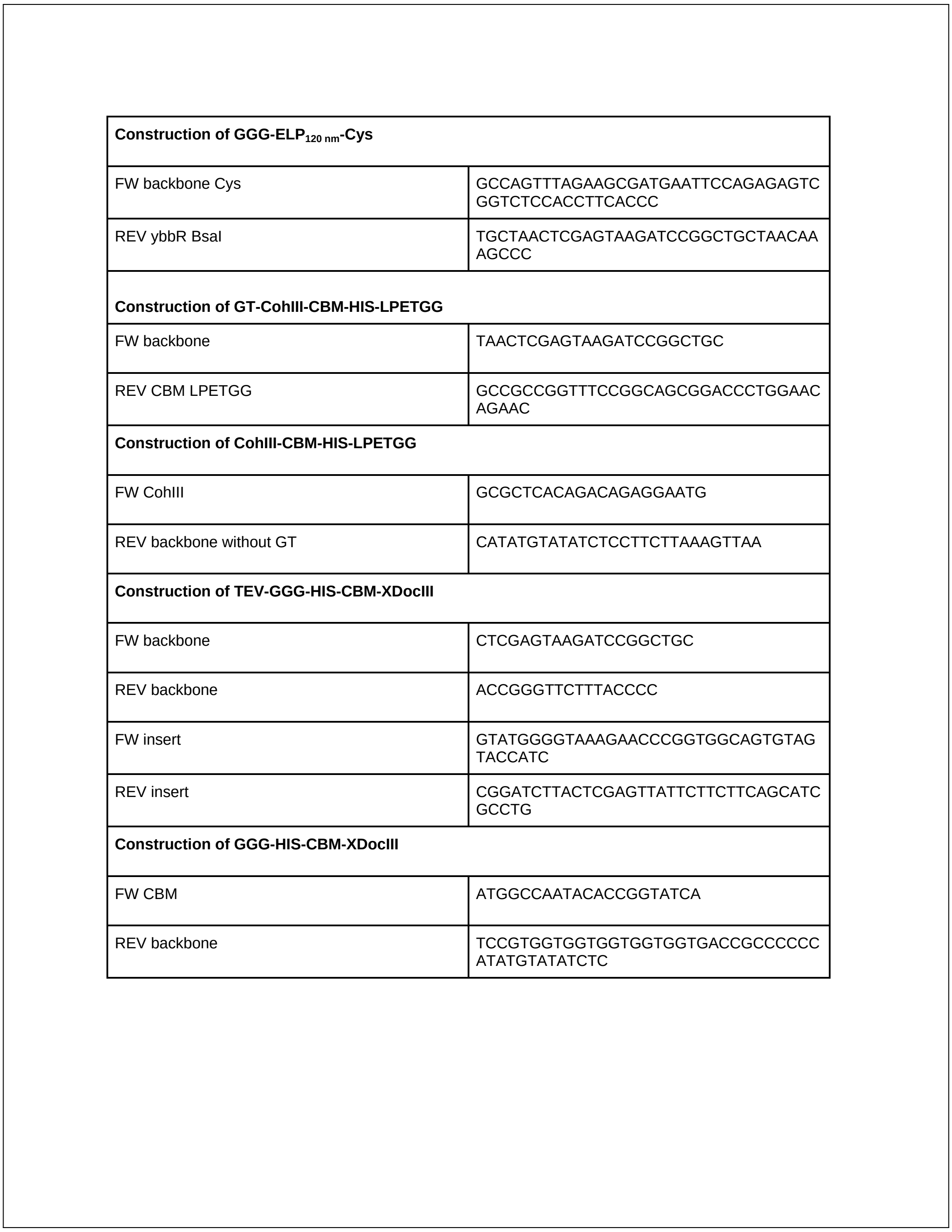


Supplemental Figure S11. Cloning scheme for TEV-GGG-CBM-Xmod-DocIII



Supplemental Table S1. Overview of primers

Name	**Sequence (5’-3’)**
Construction of Cys-ELP$_{120\ nm}$-LPETGG	
FW N-Cys BsaI	GACTCTCTGGAATTCATCGCTTCTAAACTGGCTGGTCTCCTGCGTGCCGGGAGAAGGAG
REV BsaI ybbR	CCCGGCACAGCCAGTTTAGAAGCGATGAATTCCAGAGAGTCGGTCTCACATATGTATATC
Construction of TEV-GGG-ELP$_{60\ nm}$-LPETGG	
QuikChange Primer ybbR to TEV-GGG	GACACCAGGGACTCCTTCTCCCGGCACACCGCCCCCTCCCTGGAAGTACAGGTTTTCCATATGTATATCTCCTTC
Construction of TEV-GGG-ELP$_{60\ nm}$-ybbR	
QuikChange Primer LPETGG to ybbR	GACACCAGGGACTCCTTCTCCCGGCACACCGCCCCCTCCCTGGAAGTACAGGTTTTCCATATGTATATCTCCTTC
Construction of TEV-GGG-ELP$_{120\ nm}$-ybbR	
FW backbone BsaI	GAAAACCTGTACTTCCAGGGAGGGGGGTCTCGGGGTGTGCCGGGAGAAGGAG
REV backbone BsaI	ATATATGGTCTCGACCGCCCCCTCCCTGGAAGTACAGGTTTTC
FW insert TEV-GGG BsaI	CCAGGGAGGGGGGTCTCGCGGTGTGCCGGGAGAAGGAG
REV insert BsaI	TCGAGTTAAGCCAGTTTAGAAGCGATGAATTCCAGAGAGTCGGTCTCCACCCTCACCCGG
Construction of GGG-ELP$_{120\ nm}$-ybbR	
FW ELP GGG	GGGGGCGGTGTGCCGGGAG
REV BsaI TEV	GGCACACCGCCCCCTCCCTGGAAGTACAGGTTTTCGGTCTCACATATGTATATCTCCTTC



Construction of GGG-ELP$_{120\ nm}$-Cys	
FW backbone Cys	GCCAGTTTAGAAGCGATGAATTCCAGAGAGTCGGTCTCCACCTTCACCC
REV ybbR BsaI	TGCTAACTCGAGTAAGATCCGGCTGCTAACAAAGCCC
Construction of GT-CohIII-CBM-HIS-LPETGG	
FW backbone	TAACTCGAGTAAGATCCGGCTGC
REV CBM LPETGG	GCCGCCGGTTTCCGGCAGCGGACCCTGGAACAGAAC
Construction of CohIII-CBM-HIS-LPETGG	
FW CohIII	GCGCTCACAGACAGAGGAATG
REV backbone without GT	CATATGTATATCTCCTTCTTAAAGTTAA
Construction of TEV-GGG-HIS-CBM-XDocIII	
FW backbone	CTCGAGTAAGATCCGGCTGC
REV backbone	ACCGGGTTCTTTACCCC
FW insert	GTATGGGGTAAAGAACCCGGTGGCAGTGTAGTACCATC
REV insert	CGGATCTTACTCGAGTTATTCTTCTTCAGCATCGCCTG
Construction of GGG-HIS-CBM-XDocIII	
FW CBM	ATGGCCAATACACCGGTATCA
REV backbone	TCCGTGGTGGTGGTGGTGGTGACCGCCCCCCATATGTATATCTC

Supplemental Table S2. Biophysical parameters of the employed ELPs.

ELP repeats $(5)_x$	**ε_{205} [1/M cm][7]**	**Molecular weight [Da][8]**	**Isoelectric point**	**Amino acids in ELP repeats (total)[8]**	**Total Length [nm][9]** (.365 nm per aa)
Cys-ELP$_{120\text{ nm}}$-LPETGG	851370	24763.08	3.20	300 (307)	112.06
GGG-ELP$_{120\text{ nm}}$-Cys	843030	24379.63	3.23	300 (304)	110.96

Protein Sequences

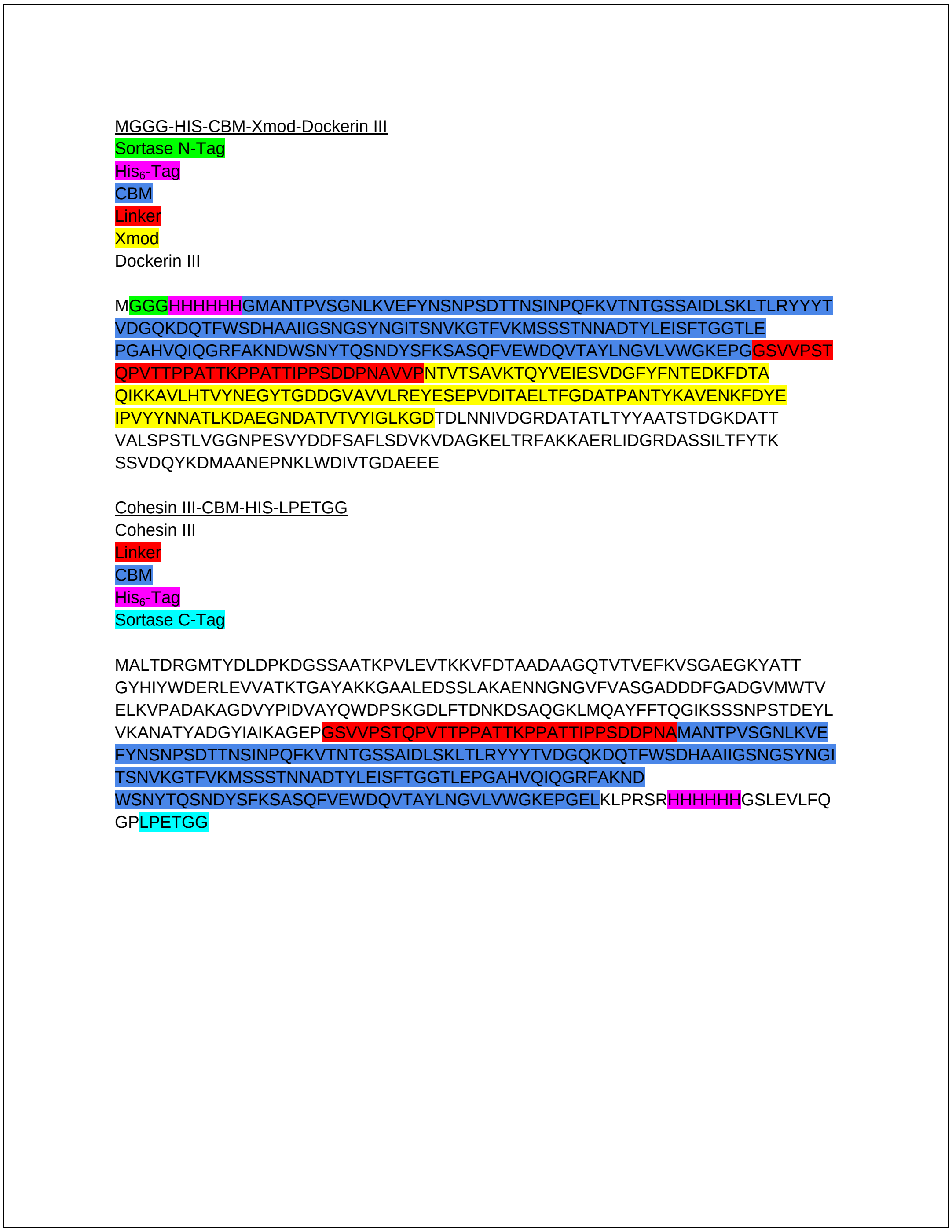
GGG-ELP$_{120\text{ nm}}$-Cys

Sortase N-Tag
ELP
Cysteine

GGGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGV
PGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVG
VPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGV
GVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPG
AGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVP
GGGVPGEGC

Cys-ELP$_{120\text{ nm}}$-LPETGG
Cysteine
ELP
Sortase C-Tag

MCVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVP
GVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGV
PGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVG
VPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGA
GVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPG
GGVPGEGLPETGG



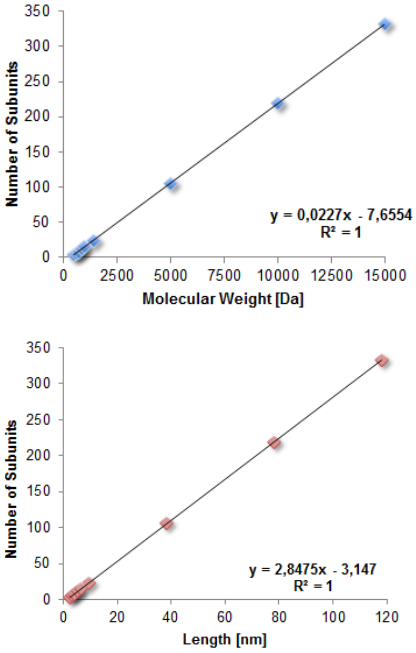
MGGG-HIS-CBM-Xmod-Dockerin III
Sortase N-Tag
His$_6$-Tag
CBM
Linker
Xmod
Dockerin III

MGGGHHHHHHGMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYT
VDGQKDQTFWSDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLE
PGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVPST
QPVTTPPATTKPPATTIPPSDDPNAVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTA
QIKKAVLHTVYNEGYTGDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYE
IPVYYNNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATT
VALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTFYTK
SSVDQYKDMAANEPNKLWDIVTGDAEEE

Cohesin III-CBM-HIS-LPETGG
Cohesin III
Linker
CBM
His$_6$-Tag
Sortase C-Tag

MALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATT
GYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGADDDFGADGVMWTV
ELKVPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPSTDEYL
VKANATYADGYIAIKAGEPGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAMANTPVSGNLKVE
FYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWSDHAAIIGSNGSYNGI
TSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKND
WSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGELKLPRSRHHHHHHGSLEVLFQ
GPLPETGG

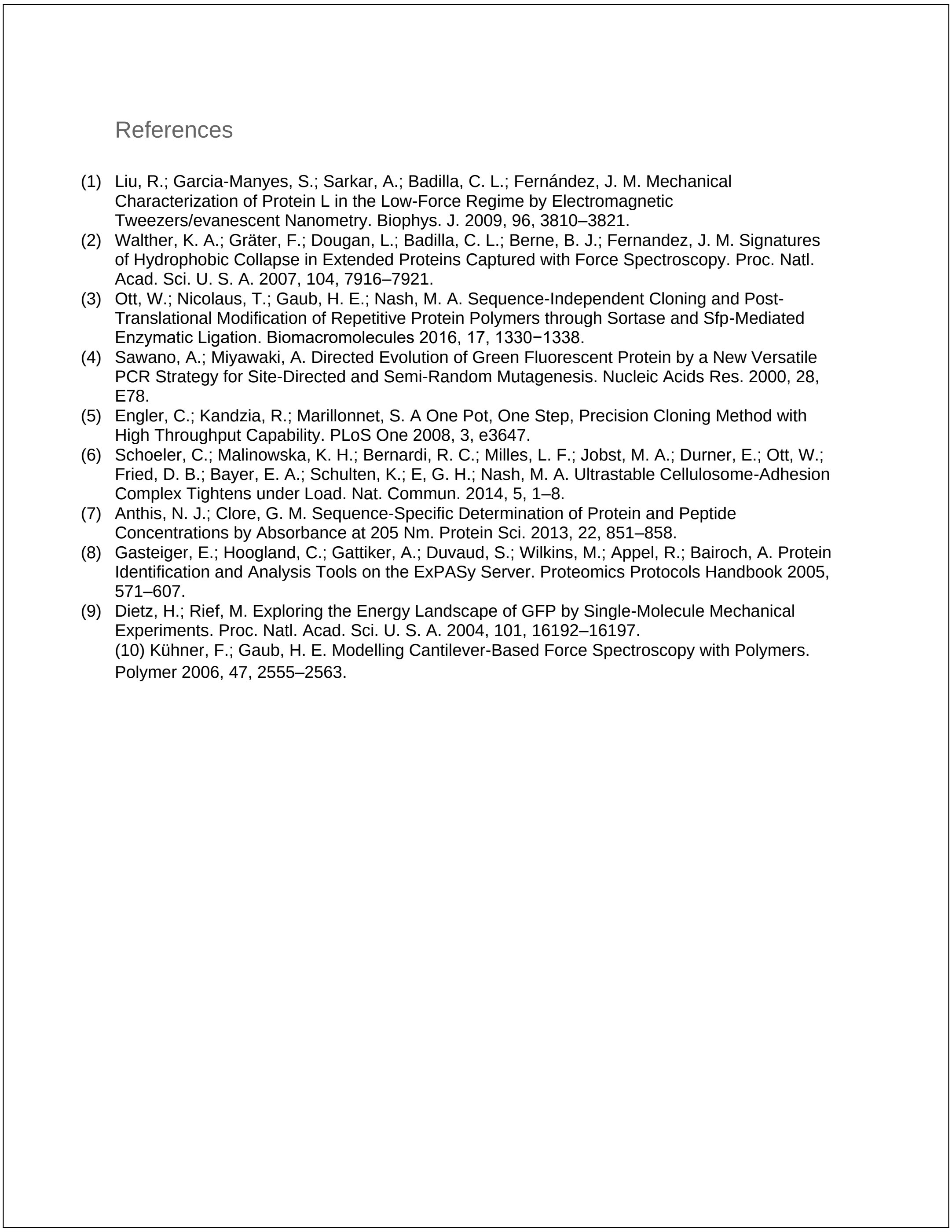
Linker Length. The artefacts generated by PEG linkers at elevated forces can be reduced by shortening the linker molecules. Usually our force spectroscopy experiments employ spacers with 40 nm length. Many SMFS assays utilize these 5 kDa PEG linkers, where the effect is scaled down proportionally with length, however still present. Further truncation would minimize the influence of the conformational change of PEG spacers, but in return raise other concerns: i) reduced mechanical isolation of the molecules under investigation by low pass filtering from transducer oscillations, to ensure purely thermally driven unfolding and dissociation events and defined loading rates[10], ii) reduced passivation of the surfaces against nonspecific adsorption, and iii) influence of surface effects and effects of the linker molecules themselves on the domains of interest. Employing peptide based smart polymers as linkers offer a new solution to this issue, avoiding linker artefacts almost entirely.



Supplemental Figure S12. Conversion of PEG molecular weights with functional end groups into their corresponding lengths. Based on the molecular weight of PEGs with functional groups maleimide and NHS, the number of subunits for various PEGs can be determined. Subsequently, the PEG contour lengths for a given number of subunits can be calculated. The data were obtained from the NHS-PEG-maleimide portfolio of Thermo Scientific and Rapp Biopolymers.

Supplemental Table S3. Overview of average molecular weight and length of PEG-Polymers. In blue are the calculated polymer sizes, in black the data the calculation is based on. Number of subunits were always round to the next integer.

Molecular Weight [Da]	**Number of Subunits**	**Length [nm]**
513.3	4	2.5
601.6	6	3.2
689.71	8	3.9
865.92	12	5.3
1394.55	24	9.5
1000	15	6.4
5000	106	38.3
10000	220	78.1
15000	333	118.0



References

(1) Liu, R.; Garcia-Manyes, S.; Sarkar, A.; Badilla, C. L.; Fernández, J. M. Mechanical Characterization of Protein L in the Low-Force Regime by Electromagnetic Tweezers/evanescent Nanometry. Biophys. J. 2009, 96, 3810–3821.

(2) Walther, K. A.; Gräter, F.; Dougan, L.; Badilla, C. L.; Berne, B. J.; Fernandez, J. M. Signatures of Hydrophobic Collapse in Extended Proteins Captured with Force Spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 2007, 104, 7916–7921.

(3) Ott, W.; Nicolaus, T.; Gaub, H. E.; Nash, M. A. Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation. Biomacromolecules 2016, 17, 1330−1338.

(4) Sawano, A.; Miyawaki, A. Directed Evolution of Green Fluorescent Protein by a New Versatile PCR Strategy for Site-Directed and Semi-Random Mutagenesis. Nucleic Acids Res. 2000, 28, E78.

(5) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS One 2008, 3, e3647.

(6) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; E, G. H.; Nash, M. A. Ultrastable Cellulosome-Adhesion Complex Tightens under Load. Nat. Commun. 2014, 5, 1–8.

(7) Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. Protein Sci. 2013, 22, 851–858.

(8) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.; Appel, R.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protocols Handbook 2005, 571–607.

(9) Dietz, H.; Rief, M. Exploring the Energy Landscape of GFP by Single-Molecule Mechanical Experiments. Proc. Natl. Acad. Sci. U. S. A. 2004, 101, 16192–16197.
(10) Kühner, F.; Gaub, H. E. Modelling Cantilever-Based Force Spectroscopy with Polymers. Polymer 2006, 47, 2555–2563.

5.3 PUBLICATION P4: Dual Binding Modes in Protein Complexes

In this research article it was demonstrated that the dual binding mode of Coh-Doc type 1 (*C.t.*), earlier suggested by mutational crystallography studies, also exists in the wild-type molecules and under native conditions. A bias of the fingerprint unfolding force distribution showed that the difference in mechano-stability of the handle complex originates in a pre-probing decision (*i.e.* binding in two differing modes) as opposed to branching of pathways during probing. My contribution to this work was the main experimental design, substantial writing of the manuscript, cloning and expression of target proteins, and the predominant part of the data acquisition, analysis, and interpretation.

Resolving Dual Binding Conformations of Cellulosome Cohesin-dockerin Complexes Using Single-molecule Force Spectroscopy

Markus A. Jobst, Lukas F. Milles, Constantin Schoeler,
Wolfgang Ott, Daniel B. Fried, Edward A. Bayer,
Hermann E. Gaub and Michael A. Nash

published in

eLIFE, 4, e10319, (2015)

eLIFE
elifesciences.org

RESEARCH ARTICLE

Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy

Markus A Jobst[1,2], Lukas F Milles[1,2], Constantin Schoeler[1,2], Wolfgang Ott[1,2], Daniel B Fried[3], Edward A Bayer[4], Hermann E Gaub[1,2], Michael A Nash[1,2]*

[1]Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-University, Munich, Germany; [2]Center for Nanoscience, Ludwig-Maximilians-University, Munich, Germany; [3]Kean University, New Jersey, United States; [4]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel

Abstract Receptor-ligand pairs are ordinarily thought to interact through a lock and key mechanism, where a unique molecular conformation is formed upon binding. Contrary to this paradigm, cellulosomal cohesin-dockerin (Coh-Doc) pairs are believed to interact through redundant dual binding modes consisting of two distinct conformations. Here, we combined site-directed mutagenesis and single-molecule force spectroscopy (SMFS) to study the unbinding of Coh:Doc complexes under force. We designed Doc mutations to knock out each binding mode, and compared their single-molecule unfolding patterns as they were dissociated from Coh using an atomic force microscope (AFM) cantilever. Although average bulk measurements were unable to resolve the differences in Doc binding modes due to the similarity of the interactions, with a single-molecule method we were able to discriminate the two modes based on distinct differences in their mechanical properties. We conclude that under native conditions wild-type Doc from *Clostridium thermocellum* exocellulase Cel48S populates both binding modes with similar probabilities. Given the vast number of Doc domains with predicteddual binding modes across multiple bacterial species, our approach opens up newpossibilities for understanding assembly and catalytic properties of a broadrange of multi-enzyme complexes.

DOI: 10.7554/eLife.10319.001

***For correspondence:** michael.nash@lmu.de

Competing interests: The authors declare that no competing interests exist.

Funding: See page 17

Received: 23 July 2015
Accepted: 28 October 2015
Published: 31 October 2015

Reviewing editor: Taekjip Ha, Johns Hopkins University School of Medicine, United States

Introduction

Cellulosomes are hierarchically branching protein networks developed by nature for efficient deconstruction of lignocellulosic biomass. These enzyme complexes incorporate catalytic domains, carbohydrate binding modules (CBMs), cohesin:dockerin (Coh:Doc) pairs, and other conserved features (***Demain et al., 2005***; ***Bayer et al., 2004***; ***Schwarz, 2001***; ***Béguin and Aubert, 1994***; ***Smith and Bayer, 2013***; ***Fontes and Gilbert, 2010***). A central attribute of cellulosome assembly is the conserved ~75 amino acid type-I Doc domain typically found at the C-terminus of cellulosomal catalytic domains. The highly conserved consensus Doc sequence from *Clostridium thermocellum (Ct)* is shown in ***Figure 1A***. Dockerins guide attachment of enzymes into the networks by binding strongly to conserved Coh domains organized within non-catalytic poly (Coh) scaffolds. In addition to their nanomolar binding affinities, many archetypal Coh:Doc pairs are thought to exhibit dual binding modes (***Carvalho et al., 2007***; ***Pinheiro et al., 2008***; ***Currie et al., 2012***). The bound Doc domain can adopt two possible orientations that differ by ~180° rotation on the Coh surface, as shown in ***Figure 1B***. The two binding modes originate from duplicated F-hand sequence motifs, a conserved structural feature found among type-I dockerins (***Pagès et al., 1997***). The duplicated F-hand motifs resemble EF-hands found in eukaryotic calcium binding proteins (e.g., calmodulin), and provide

eLife digest Some bacteria use cellulose, the main component of plant cell walls, as a food source. The enzymes that break down cellulose are anchored onto a protein scaffold in a structure called the cellulosome on the bacteria's surface. This anchoring occurs through an interaction between receptor proteins known as 'cohesin' domains on the scaffold proteins and 'dockerin' ligands on the enzymes.

Most receptor-ligand interactions only allow the two proteins to bind in a single, fixed orientation. However, cohesins and dockerins are suspected to bind in two different configurations. It has been difficult to investigate the populations of these different configurations because most experimental techniques investigating protein binding take average measurements from many molecules at once. As the binding modes are extremely similar, these methods have been unable to distinguish between the two cohesin-dockerin binding configurations without introducing mutations, in part because these configurations are very similar to each other.

Jobst et al. used a technique called single-molecule force spectroscopy to investigate cohesin-dockerin interactions between individual molecules. This technique applies a force that separates, or 'unbinds', cohesin and dockerin, by pulling individual complexes of the two binding partners apart with a nanoscale probe. In the experiments, *E. coli* bacteria were made to produce mutant versions of dockerin that can only bind to cohesin in one orientation. This allowed each binding configuration to be studied individually. The results of these experiments revealed the mechanical unbinding patterns of each cohesin-dockerin configuration, and showed that it is possible to use these patterns to distinguish between the two configurations. A complimentary set of experiments revealed that wild-type (non-mutated) cohesin-dockerin complexes occupy both configurations in approximately equal amounts, and do not switch modes once bound.

Further single-molecule experiments together with computer simulations will provide a more detailed picture of how cohesin and dockerin fit together in the two configurations. Such experiments could also reveal how cohesin and dockerin contribute to the break down of cellulose inside living cells and how they could be used for the precise assembly of single proteins.

DOI: 10.7554/eLife.10319.002

internal sequence and structural symmetry to Doc domains. Rotating Doc by ~180° with respect to Coh (***Figure 1B,C***) results in an alternatively bound complex with similarly high affinity involving the same residues on Coh recognizing mirrored residues within Doc. The dual binding mode is thought to increase the conformational space available to densely packed enzymes on protein scaffolds, and to facilitate substrate recognition by catalytic domains within cellulosomal networks (***Bayer et al., 2004***). From an evolutionary perspective, the dual binding mode confers robustness against loss-of-function mutations, while allowing mutations within Doc to explore inter-bacterial species cohesin-binding promiscuity in cellulosome-producing microbial communities. Coh:Doc interactions and dual binding modes are therefore important in the context of cellulose degradation by cellulosome-producing anaerobic bacterial communities.

However, direct experimental observation of the dual binding modes for wild-type Doc has thus far proven challenging. Ensemble average bulk biochemical assays (e.g., surface plasmon resonance, calorimetry, enzyme-linked immunosorbent assays) are of limited use in resolving binding mode populations, particularly when the binding modes are of equal thermodynamic affinity. Crystallography is challenging because the complex does not adopt a unique molecular conformation, but rather exhibits a mixture of two conformations thereby hindering crystal growth. Structural data on the dual binding mode have typically been collected using a mutagenesis approach, where one of the binding modes was destabilized by mutating key recognition elements (***Carvalho et al., 2007***; ***Pinheiro et al., 2008***). This approach, however, while resolving the structures of each bound complex, cannot determine if one binding mode is dominant for wild-type Doc, or if that dominance is species or sequence dependent. Coarse grained molecular dynamics has also predicted dual modes of interaction between Coh and Doc (***Hall and Sansom, 2009***), but direct experimental evidence of both binding modes for wild-type Doc has remained elusive. Improved fundamental understanding of the dual binding mode could shed light onto the molecular mechanisms by which these multi-

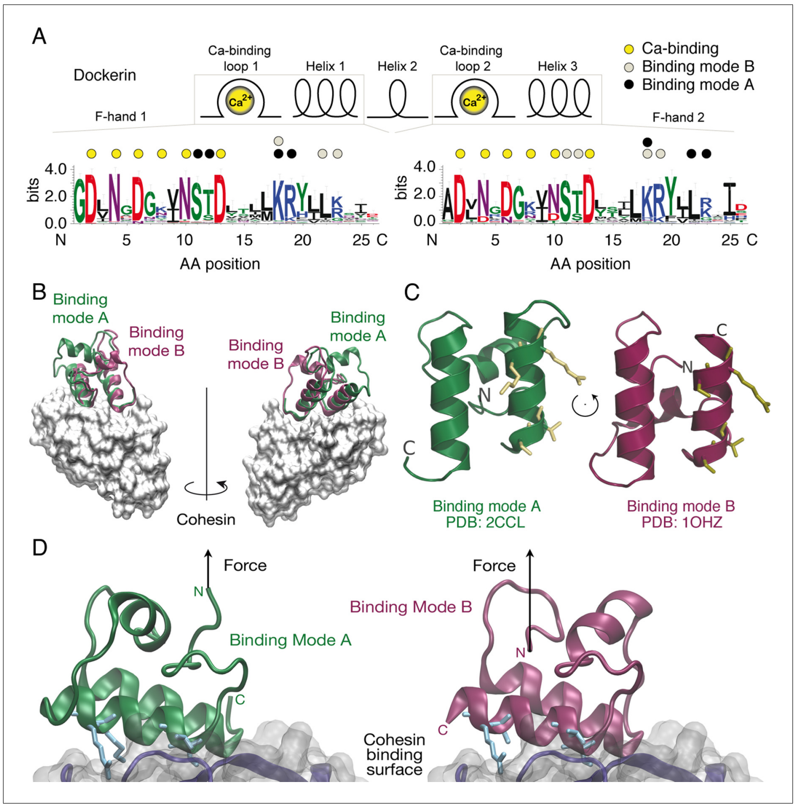


Figure 1. Cohesin:Dockerin dual binding modes. (**A**) Secondary structure and consensus sequence logo (***Crooks, 2004***) assembled from 65 putative *Ct* type-I Doc variants. Dots above the amino acid codes indicate residues involved in: Ca^{2+} coordination (yellow), mode A binding (black), and mode B binding (gray). Letter colors represent chemical properties: Green, polar; purple, neutral; blue, basic; red, acidic; black, hydrophobic. Crucial Coh-binding residues are located at positions 11, 12, 18, 19, 22, and 23 in each F-hand motif. (**B**) Coh:Doc complex crystal structures showing overlaid Doc domains in the two binding modes. Images were generated by aligning the Coh domain (gray) from PDB 2CCL (green, binding mode (**A**) and 1OHZ (red, binding mode (**B**) using the VMD plugin MultiSeq (***Humphrey et al., 1996***; ***Roberts et al., 2006***). (**C**) View of the Doc binding interface for each mode from the perspective of Coh. The conserved binding residues at positions 11, 12, 18, and 19 in the F-hand motif relevant for binding in the corresponding mode are depicted as stick models (yellow). (**D**) Close-up view of the interface for each binding mode with arrows indicating the location and direction of applied force. Binding residues 11, 12, 18, and 19 for binding mode A and 45, 46, 52, and 53 for binding mode B are shown as blue stick models. The Coh domain is oriented the exact same way in both views.

DOI: 10.7554/eLife.10319.003

enzyme complexes self-assemble and achieve synergistic conformations, as well as provide a new approach to designing systems for protein nanoassembly (***Kufer et al., 2009***; ***2008***).

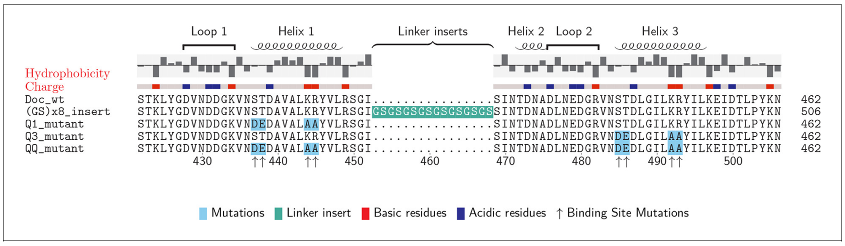
Here, we used SMFS (***Li and Cao, 2010***; ***Engel and Müller, 2000***; ***Woodside and Block, 2014***) to study wild-type and mutant Doc from exocellulase Cel48S of *C. thermocellum* (*Ct*-DocS). We demonstrate that specific unfolding/unbinding trajectories of individually bound Coh:Doc complexes are characteristic of the binding modes. To validate our approach, we produced Doc mutants that exhibited a preferred binding mode. We performed single-molecule pulling experiments on bound Coh:mutant Doc complexes and observed a strong bias in the probability of two clearly distinguishable unfolding patterns, termed 'single' and 'double' rupture types for each binding mode mutant. We further probed the unbinding mechanism of the double rupture events using poly (Gly-Ser) inserts to add amino acid sequence length to specific sections of Doc as a means to identify which portions of Doc unfolded. Finally, we used the inherent differences in mechanical stability of each binding mode, and the effects these differences had on the unfolding force distributions of an adjacent domain, to directly observe and quantify binding mode populations for wild-type Doc.

Results

Protein design

The wild-type and mutant Doc sequences used in this work were aligned (***Beitz, 2000***) and are presented in ***Figure 2***. Among *Ct*-Doc domains, a Ser-Thr pair located at positions 11 and 12 of F-hand motif 1 (N-terminal helix 1) is highly conserved (***Figure 1A***). This Ser-Thr pair is H-bonded to Coh in binding mode A (***Figure 1A***, black dots). Analogously, binding mode B refers to the configuration where the Ser-Thr pair from helix 3 dominates the H-bonding to Coh (***Figure 1A***, gray dots). Binding mode B was previously crystallized for a homologous *Ct*-Doc (***Carvalho et al., 2003***). Mutation of the Ser-Thr pair in helix 3 to Ala-Ala was used to bias binding and thereby crystallize binding mode A for the same Doc (***Carvalho et al., 2007***). A similar targeted mutagenesis approach was also used to obtain crystal structures of a *Clostridium cellulolyticum* Doc in each binding mode (***Pinheiro et al., 2008***).

To preferentially select for a specific binding mode (A or B), we prepared Doc sequences that incorporated 4 amino acid point mutations, referred to as quadruple mutants ('Q'). To design quadruple mutants, we noted that recent structural work reported a set of *Ct*-Doc domains that differ from the canonical duplicated Ser-Thr sequences. These non-canonical Docs were found to exhibit only a single binding mode (***Brás et al., 2012***; ***Pinheiro et al., 2009***). In one of these non-canonical Doc domains, an Asp-Glu pair was found in place of Ser-Thr. Since the Coh surface is negatively charged, we postulated that including Asp-Glu in place of Ser-Thr within one of the F-hands could be used to effectively knock out a given binding mode for our canonical Doc. Additionally, we incorporated double alanine mutations to replace the conserved Lys-18 Arg-19 residues of a given F-hand motif, further destabilizing a targeted binding mode. Q1 refers to a quadruple mutant where helix 1 has been modified at four positions (i.e. S11D-T12E-K18A-R19A). Q3 refers to the quadruple mutant where helix 3 has been modified at four positions (i.e. S43D-T44E-K50A-R51A). As a negative



```
Doc_wt        STKLYGDVNDDGKVNSTDAVALKRYVLRSGI................SINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN   462
(GS)x8_insert STKLYGDVNDDGKVNSTDAVALKRYVLRSGIGSGSGSGSGSGSGSGSSINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN   506
Q1_mutant     STKLYGDVNDDGKVNDEDAVALAAYVLRSGI................SINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN   462
Q3_mutant     STKLYGDVNDDGKVNSTDAVALKRYVLRSGI................SINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN   462
QQ_mutant     STKLYGDVNDDGKVNDEDAVALAAYVLRSGI................SINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN   462
```

Figure 2. Doc sequences used in this study (N- to C-terminus). Doc_wt: wild-type sequence; hydrophobicity and charge graphs are displayed for the wild-type-Doc (red: positively charged, blue: negatively charged); (GS)x8_insert: A (Gly-Ser)$_8$ linker was incorporated between helix 1 and helix 2; Q1_mutant: Quadruple mutant in helix 1. Four point mutations (DE/AA) were incorporated into Doc helix 1 to knock out binding mode A; Q3_mutant: Quadruple mutant in helix 3. Four point mutations (DE/AA) were incorporated into Doc helix 3 to knock out binding mode B; QQ_mutant: Non-binding control with both binding modes knocked out. Numbers below indicate amino acid number of the fusion protein construct starting from the xylanase N-terminus.

DOI: 10.7554/eLife.10319.004

control, we prepared a mutant referred to as 'QQ' that incorporated quadruple mutations into both helices 1 and 3.

Doc domains were expressed as fusion domains attached to the C-terminal end of xylanaseT6 (Xyn) from *Geobacillus stearothermophilus* to improve solubility and expression levels as previously reported (***Stahl et al., 2012***). The Xyn domain also acts as a so-called fingerprint in AFM force extension traces to provide a means for screening datasets and searching for known contour length increments. We use the term 'contour length' to refer to the maximum length of a stretched (unfolded) polypeptide chain. Our screening process identified single-molecule interactions and ensured correct pulling geometry. For the Coh domain, we chose cohesin 2 from *Ct*-CipA expressed as a C-terminal fusion domain with the family 3a carbohydrate binding module (CBM) from *Ct*-CipA. In order to exclude artifacts arising from fingerprint domains, protein immobilization or pulling geometry, a second set of fusion proteins was cloned, expressed and probed in complementary experiments using a flavoprotein domain from the plant blue light receptor phototropin (iLOV) (***Chapman et al., 2008***). All protein sequences are provided in the 'Materials and methods' section.

Single-molecule unfolding patterns

The pulling configuration for single-molecule AFM experiments is shown in ***Figure 3A***. CBM-Coh was site-specifically and covalently attached to an AFM cantilever tip and brought into contact with a glass surface modified with Xyn-Doc. The mechanical strength of protein domains and complexes will strongly depend on the pulling points (i.e. sites at which the molecule is attached to cantilever/surface). The site-specific attachment chemistry used here was precisely defined by the chosen residue of immobilization, ensuring the same loading geometry was used on the complex for each and every data trace. After formation of the Coh:Doc complex, the cantilever was retracted at a constant speed that ranged from 200 to 3200 nm/s while the force was monitored by optical cantilever deflection. The resulting force-distance traces were characteristic of the series of energy barriers crossed by the protein complex along the unfolding/unbinding pathway. A sawtooth pattern was consistently observed when molecular ligand-receptor complexes had formed. Sorting the data using contour length transformation (***Puchner et al., 2008***) and identifying traces that contained a Xyn contour length increment (~89 nm) allowed us to screen for single-molecule interactions (***Stahl et al., 2012***), as described in our prior work on Coh:Doc dissociation under force (***Stahl et al., 2012***; ***Schoeler et al., 2014***; ***Jobst et al., 2013***; ***Otten et al., 2014***; ***Schoeler et al., 2015***).

Typical single-molecule interaction traces from such an experiment are shown in ***Figure 3B, C*** and in ***Figure 3—figure supplement 1***. Following PEG linker stretching, an initial set of peaks

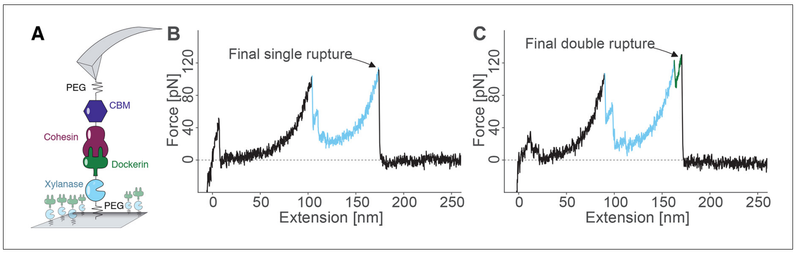


Figure 3. Overview of the experimental configuration and recorded single-molecule unfolding and unbinding traces. (**A**) Schematic depiction showing the pulling geometry with CBM-Coh on the AFM Cantilever and Xyn-Doc on the glass substrate. Each fusion protein is site-specifically and covalently immobilized on a PEG-coated surface. (**B-C**) Each force vs. extension trace shows PEG linker stretching (black), xylanase unfolding and subsequent stretching (blue), and Coh:Doc complex rupture. The Coh:Doc complex rupture occurred in two distinct event types: single (**B**) and double (**C**) ruptures. The 8-nm contour length increment separating the double peaks was assigned to Doc unfolding (C, green).

DOI: 10.7554/eLife.10319.005

The following figure supplement is available for figure 3:

Figure supplement 1. Representative sample of force traces.

DOI: 10.7554/eLife.10319.006

sequentially decreasing in force was assigned to xylanase unfolding and stretching. This domain when unfolded added ~89 nm of free contour length to the system. The final peak (s) corresponded to rupture of the Coh:Doc complex, and occurred as either 'single' or 'double' rupture events. The contour length increment between the two double event peaks was found to be ~8 nm, that is, 8 nm of hidden contour length was added to the biopolymer during a sub-step of Doc unbinding (see 'Discussion'). The 8-nm contour length increment was also observed in complementary experiments employing other fusion domains: xylanase was swapped for an sfGFP domain and CBM was swapped out for an iLOV domain. In these new fusions, the 8 nm Doc increment was still observed, indicating it was not caused by a specific fusion domain. As we show below, double and single rupture events were associated with binding modes A and B, respectively. CBM unfolding length increments (~57 nm) were only rarely observed because the Coh:Doc complex only rarely withstood forces sufficiently high to unfold CBM (***Stahl et al., 2012***).

Ensemble average binding experiments

Binding experiments were carried out in bulk to evaluate the binding affinity of wild-type, Q1, Q3, and QQ Doc sequences to wild-type Coh. Xyn-Doc fusion protein variants were immobilized in a microwell plate and exposed to tag red fluorescent protein (TagRFP) (***Merzlyak et al., 2007***) fused to Coh (TagRFP-Coh) across a range of concentrations, followed by rinsing and subsequent fluorescence readout (***Figure 4A***). The data clearly showed that Q1 and Q3 Doc sequences, each with a mutated binding mode, maintained high-binding affinity with dissociation constants (K_d) in the nM range. These values are in good agreement with previous reports on homologous type-I Doc domains (***Brás et al., 2012***; ***Sakka et al., 2011***). This suggested that mutant Doc domains with one destabilized binding mode were still able to recognize fluorescent protein fused Coh with strong affinity by relying on the alternative binding mode that was preserved. The QQ double knockout mutant, however, showed no appreciable binding over the concentration range tested. This negative control showed that DEAA quadruple mutations were in fact effective at eliminating binding for the targeted modes.

Single-molecule rupture statistics of binding mode mutants

For each Doc tested, we collected tens of thousands of force-extension traces and selected for further analysis only those traces showing the ~89 nm xylanase contour length increments and no other anomalous behavior, resulting in typically 200–3000 usable single-molecule interaction curves per experiment. We determined the number of Coh:Doc unbinding events that occurred as single or double rupture peaks. The results are shown in ***Figure 4B***. The wild-type Doc showed double rupture events in ~57% of the cases, and single rupture events in ~43% of the cases. The mutant designed to knock out binding mode A (Q1), showed a single event probability of ~77%, and a double event probability of ~23%. The mutant designed to knock out binding mode B (Q3) showed a single event probability of ~41%, and a double event probability of ~59%. It is clear from these data that the Q1 mutant has a strong bias toward single peaks that is not observed in the wild-type leading to preliminary assignment of single peaks to binding mode B.

For all double events, we determined the force difference of the second peak relative to the first (***Figure 4C***). Q1 and wild-type on average showed second peaks that were ~15–20% higher in force than the first peak. Q3 meanwhile showed clearly different behavior. Although the ratios of single to double peaks were nearly identical between wild-type and Q3, differences in the relative force between the first and second peaks differentiated wild-type and Q3 (***Figure 4C***). Double peaks for the Q3 mutant were more likely to show a shielded behavior, where the second peak was lower in force than the first peak by ~10%. Although the Q3 mutant showed the same single vs. double event probability as wild-type, the double events for Q3 were distinguishable from those of the wild-type based on this observed decrease in the rupture force of the second peak. The second barrier of the double events was therefore weaker in Q3 than for wild-type. This weaker 2nd double peak for the Q3 mutant combined with similar single/double peak ratios as wild-type leads us to believe that the number of double peaks is being underestimated systematically for the Q3 mutant. Generally, each binding mode still allows for the occurrence of a single event (albeit with different likelihood), in which the whole Doc domain unbinds without an additional unfolding substep. Since the second and final energy barrier for complex dissociation is weaker than the first for the Q3 mutant, the

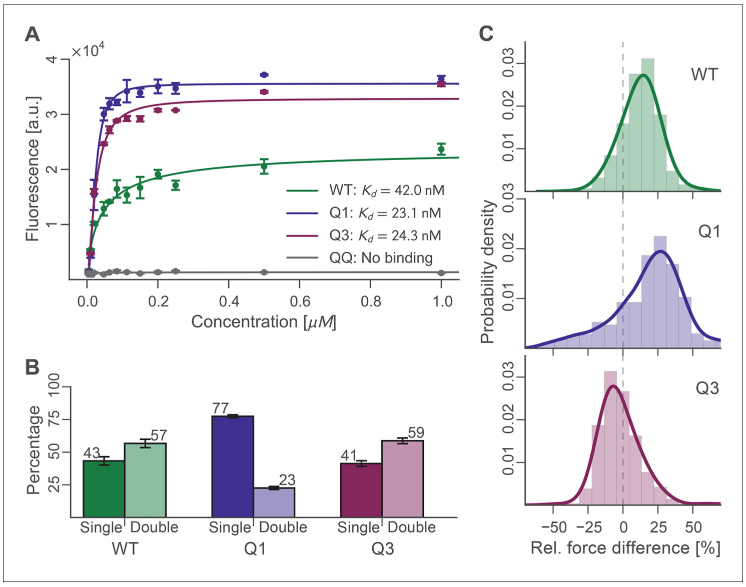


Figure 4. Bulk and single-molecule characterization of Doc mutants. (**A**) Fluorescence binding curve showing binding of TagRFP-labelled Coh to wild-type and mutant Doc nonspecifically immobilized in a 96-well plate. Both Q1 and Q3 mutants bound TagRFP-Coh similarly to wild-type with dissociation constants (K_D) in the low nM range. The negative control QQ mutant showed no binding. Solid lines are 4 parameter logistic nonlinear regression model fits to the data. Error bars represent the standard deviation of three independent samples. (**B**) Event probabilities for single (opaque colors) and double (translucent colors) Coh:Doc rupture peaks determined for Doc wild-type and DE/AA quadruple mutants. Data originate from 947, 4959, and 1998 force-extension traces from wild-type, Q1 and Q3 variants, respectively. Error bars represent 95% Clopper-Pearson confidence intervals based on the beta probability distribution. (**C**) Relative difference in double peak rupture forces for the different Doc variants. Positive values indicate a stronger final peak. Histograms represent concatenated data from various pulling speeds. Drawn lines are kernel density estimates calculated on the raw data.

DOI: 10.7554/eLife.10319.007

The following source data is available for figure 4:

Source data 1. Probability Data.

DOI: 10.7554/eLife.10319.008

probability for the molecule to pass both barriers simultaneously is increased, thus resulting in a higher percentage of single events.

Probing the 8-nm length increment with poly (GS) inserts

We sought to identify the molecular origin of the 8 nm contour length increment separating the double event peaks by engineering additional amino acid sequence length into the Doc domain. Amino acid insert sequences have previously been used to probe length increments in AFM force spectroscopy experiments (***Bertz and Rief, 2009***) (***Carrion-Vazquez et al., 1999***). By adding additional amino acids to the polypeptide chain at a particular location, insert sequences increase the gain in contour length following unfolding of a subdomain in a predictable way. Any change in the observed length increment can be pinpointed to the position in the molecule where the unfolding event occurs. In this case, we engineered flexible $(GS)_8$ insert sequences directly into wild-type Doc between helices 1 and 2, in a flexible loop that was not expected to interfere with either of the two binding modes. Structural homology models (***Figure 5A***) of the wild-type Doc and $(GS)_8$ insert

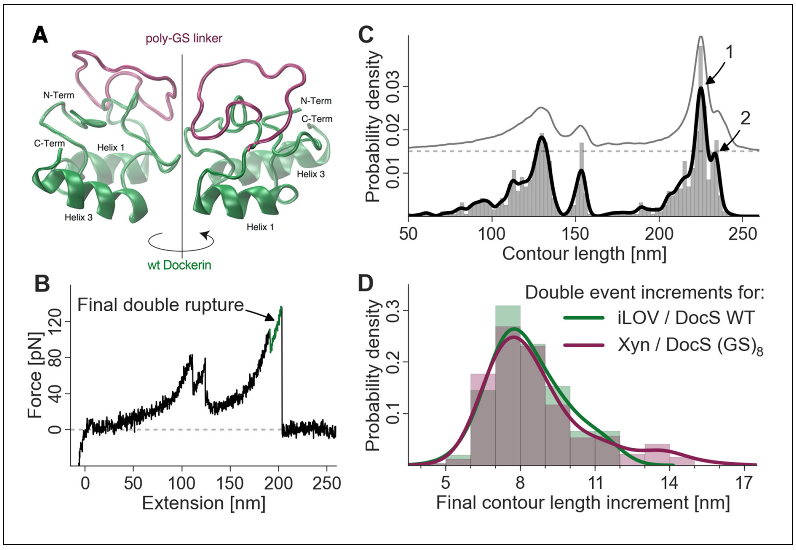


Figure 5. Probing the final contour length increment with Poly (GS) inserts. (**A**) Structural homology model overlay of wild-type and mutant Doc containing a $(GS)_8$-linker between helix 1 and helix 3. The wild-type Doc is shown in green. The 16 amino acid long GS-insert is shown in purple (***Kelley and Sternberg, 2009***) (remaining Doc domain not shown). (**B**) Typical force extension trace with final double rupture event depicted in green (arrow). (**C**) Histogram and kernel density estimate of the transformation of the single force extension trace in panel B into contour length space (black) and kernel density estimate of the whole dataset of single molecule Xyn-Doc:Coh-CBM traces bearing xylanase fingerprint and final double rupture (gray, offset in y-direction for readability) in contour length space. (**D**) Histograms (bars, bin width: 1 nm), kernel density estimates (drawn lines, bandwidth: 0.75 nm, gaussian kernel), and statistical test (Kolmogorov-Smirnov, 'KS test') are each calculated on the raw data of the final increments (peak-to-peak distances) in contour length space (x-distance between arrow 1 and 2 in panel (**C**). Maxima for final double event increments lie at 7.75 nm and 7.73 nm for iLOV-Coh:Doc (wild-type)-sfGFP (N = 255) and Xyn-Doc $(GS)_8$:Coh-CBM (N = 320) final ruptures, respectively (a two-sample KS test on the raw data indicates no significant difference in the data distributions (p-value of 21.7%).

DOI: 10.7554/eLife.10319.009

sequence were calculated using the Phyre server (***Kelley and Sternberg, 2009***). If the 8-nm contour length increment was caused by sequential unbinding of Doc helices 1 and 3 in wild-type Doc, then double peaks for the poly (GS) constructs should show an increase in the double peak contour length increment. As shown in ***Figure 5B,C and D***, the contour length histogram for $(GS)_8$ Doc was indistinguishable from the wild-type Doc. No additional contour length was gained due to additional amino acids inserted between Doc helices 1 and 2. Since the Doc was anchored to the glass slide through an N-terminal xylanase domain, this result indicated that the unfolding event responsible for the 8-nm length increment must be located upstream (i.e. N-terminal) from the site of the $(GS)_8$-insert. This result suggested that unfolding of calcium binding loop 1 and helix 1 in Doc was the source of the 8-nm length increment.

Single-molecule evidence of dual binding mode

To finally confirm the presence of both bound conformations in wild-type Coh:Doc complexes, we replaced xylanase with sfGFPand CBM with iLOV as the contour length marker or fingerprint domains. iLOV was chosen as a superior unfolding fingerprint domain because it does not show multiple unfolding substeps (in contrast to xylanase), which simplified analysis. Also iLOV has an unfolding force distribution that lies in a similar range as the Coh:Doc complex dissociation single and

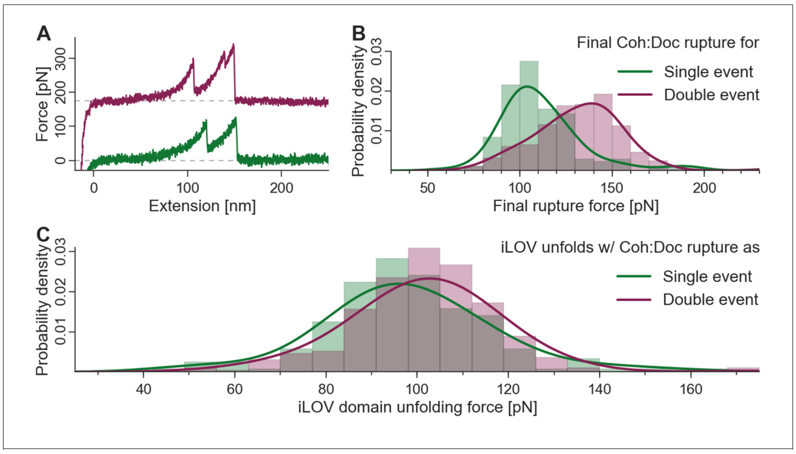


Figure 6. Biasing of unfolding force distributions by dual binding mode. (**A**) Typical force traces showing iLOV unfolding with final single (green) and double (purple) complex ruptures. The curve terminating in a double peak is offset in the y-direction for clarity. (**B**) Final complex rupture force distribution for single and double events. Double events are more mechanically stable. (**C**) iLOV domain unfolding forces for final single (green) and double (red) events at a pulling velocity of 800 nm/s. Histograms (bars), kernel density estimates (lines), and statistical tests are each obtained from the raw data. Maxima for iLOV unfolding lie at 96.0 pN and 102.7 pN for single (N = 172) and double (N = 277) final ruptures, respectively. A two-sample Kolmogorov-Smirnov test showed significant differences in the data distributions (p-value of 0.09%). Since the data were all recorded with a single cantilever and both event types were distributed equally throughout the runtime of the measurement, no systematic biasing is expected. Because of the lower force distribution of final single peaks, the iLOV unfolding force distribution is truncated compared to final double peak force traces, supporting the notion that the binding mode is set prior to mechanical loading of the complex.

DOI: 10.7554/eLife.10319.010

double peaks, allowing for effective biasing of the iLOV unfolding force distributions by the inherent stability difference between single and double event peaks. ***Figure 6A*** shows characteristic single and double event curves containing iLOV unfolding (36-nm contour length increment) followed by Coh:Doc rupture as a single or double event. The rupture force distributions of the single and double event (second peak) ruptures are shown in ***Figure 6B***. The most probable rupture force for single events was ~104 pN, while for double events this value was ~140 pN at a pulling speed of 800 nm/s. We next calculated the unfolding force distributions of the iLOV domain for curves that terminated with single events or double events. If the Coh:Doc complex ruptured before iLOV unfolding was observed, the curve was eliminated from the dataset because it lacked a fingerprint domain length increment. This criterion for inclusion in the dataset results in a biasing of the iLOV unfolding forces, since the maximum of the fingerprint unfolding force distribution that can be observed must lie below that of the Coh:Doc complex. The fact that we observed a downward shift in the iLOV unfolding forces (***Figure 6C***) for curves that terminated in the less mechanically stable single rupture event is confirmation that the single- and double-event peaks arise from separate bound conformations. Each mode has a distinct mechanical stability and energy landscape that is set at the time of receptor-ligand binding, that is once bound, the conformation of the complex does not change. If single- and double-event unbinding patterns were simply two competing pathways out of the same bound state, then the downward shift in rupture force distribution would not be observed for the iLOV unfolding forces. Although this shift in rupture force distributions is comparatively subtle, it can be observed accurately with high statistical significance. We note that the datasets for both binding modes were measured with the same cantilever throughout the runtime of the whole experiment. Calibration and drift issues therefore did not interfere with the required accuracy.

Discussion

The relatively small ~8 kDa Doc domains exhibit an internal sequence and structural symmetry that is believed to give rise to a dual mode of binding to Coh, as shown in ***Figure 1***. In order to study this remarkable plasticity in molecular recognition in greater detail, we prepared a series of mutants (***Figure 2***) designed to either knock out a specific binding mode or add length to the molecule at a specific position. Bulk experiments showed that Doc mutants Q1 and Q3, originally designed to suppress one of the binding modes, were still able to bind Coh with high affinity, while the double knockout did not bind (***Figure 4A***). The equilibrium affinities of Coh binding to Q1, Q3, or wild-type were all similarly high with K_Ds in the low nM range, in good agreement with literature values (***Sakka et al., 2011***), suggesting the two binding modes are thermodynamically equivalent and rendering them indistinguishable with conventional methods such as ELISA or calorimetry. Techniques like surface plasmon resonance could possibly show differing values for on- and off-rates for the mutants, but would still not be able to resolve the binding modes within a wild-type population.

Force spectroscopy with the AFM interrogates individual molecules, and measures their mechanical response to applied force. Since the technique is able to probe individual members of an ensemble, it provided a means to quantify binding mode configurations by assigning unfolding/unbinding patterns to the binding mode adopted by the individual complexes. Site-directed Q1 and Q3 mutations supported the assignment of binding mode A to a characteristic double rupture peak dissociation pathway. Single events were assigned to binding mode B and showed no Doc unfolding substep prior to complex rupture.

We consistently observed 8 nm of added contour length that separated the Doc double peaks. Since force is applied to Doc from the N-terminus, we analyzed the Doc sequence starting at the N-terminus and searched for reasonable portions of Doc that could unfold in a coordinated fashion to provide 8 nm of contour length. The results from the GS-insert experiments (***Figure 5***) indicated no change in the double-event contour length increment, regardless of the added GS-insert length located between helix 1 and 3 in Doc. This result is consistent with the 8 nm length increment being located N-terminally from the GS-insert site, implicating unfolding of Doc calcium binding loop 1 and helix 1 as the source of the 8 nm. This length accurately matches the estimated length increment for unfolding calculated from the crystal structure (***Figure 1D***).

Although this result could also be consistent with the 8 nm increment being located somewhere outside the Doc domain in the polyprotein, we deem this scenario highly unlikely. The 8 nm increment cannot be located in the Xyn or CBM domains because we have accounted for Xyn and CBM lengths in their entirety based on the observed 89 nm and rare 57 nm length increments here and in a previous study (***Stahl et al., 2012***), and for confirmation swapped out those domains for different proteins completely (i.e. iLOV and GFP). The remaining possibility that the 8 nm is located within the Coh domain is also not likely since the barrel-like structure of the Coh is known to be mechanically highly stable (***Valbuena et al., 2009***; ***Hoffmann et al., 2013***). Also, if the 8-nm length increment were due to partial Coh unfolding, the Q1 and Q3 mutants would not be expected to affect the single/double peak ratio or force differences between the double event peaks as was observed (***Figure 4B, C***). The GS-insert data suggest the 8-nm length increment is located within Doc, upstream (N-terminal) from the GS-insert site implicating calcium loop 1 and helix 1 in this unfolding event.

Finally, we observed that an inherent difference in the mechanical stability of single and double event rupture peaks (***Figure 6B***) could be used as a feature by which to discriminate the binding modes. Our analysis algorithm accepted only the force curves that first showed iLOV fingerprint domain unfolding followed by either a single- or double-rupture peak. By observing a small but significant downward shift in the iLOV unfolding force distribution when analyzing curves that terminated in the less stable single-event peak, we confirmed the single-event peaks originate from a unique conformation that is 'set' at the time of complex formation.

Taken together, we propose an unbinding mechanism where the first barrier of the double peaks represents unfolding of the N-terminal calcium binding loop and unraveling of alpha helix 1 up to the Lys-Arg pair at sequence positions 18 and 19 in the wild-type structure in binding mode A. Based on a length per stretched amino acid of 0.4 nm, the expected contour length for unfolding the Doc domain up to this position would be 7.6 nm, in good agreement with the measured value of 8 nm within experimental error. A portion of the N-terminal calcium binding loop (i.e. residues S11-

T12) is involved in binding to D39 in Coh. The first peak of the double events is attributed to breakage of this interaction and simultaneous unfolding of calcium loop 1 and alpha helix 1 up to the Lys-Arg pair at sequence positions 18 and 19. Another contributing factor is the intramolecular clasp that has been identified as a stabilizing mechanism among similar type-I Doc domains (***Slutzki et al., 2013***). A recent NMR structural study (***Chen et al., 2014***) on the same wild-type Doc used in this work confirmed a hydrophobic ring-stacking interaction between Tyr-5 and Pro-66. Confirmation of this clasp motif by NMR means the head and tail of the Doc are bound together, additionally stabilizing the barrier that is overcome in the first of the double event peaks. In this scenario, subsequent to breaking the interactions between the calcium binding loop and Coh, disrupting the intramolecular clasp and unfolding the N-terminal loop-helix motif, the remaining bound residues including Lys-18, Arg-19, Lys-50, Leu-54, and Lys-55 stay bound to Coh and are able to withstand substantial force on their own, eventually breaking in the second and final of the double rupture peaks. Prior work further supports this unbinding mechanism, revealing that a progressive N-terminal truncation of Doc did not affect the interaction largely, unless the truncation reached the Lys-18 and Arg-19 residues (***Karpol et al., 2009***). This corroborates the idea of the C-terminal end of helix 1 being a crucial part of the binding site within the complex. Single rupture peaks were thus observed when the wild-type complex was bound in binding mode B, and no unfolding of Ca-binding loop 1 or helix 1 occurred. Force was propagated directly to bound residues Lys-18, Leu-22, and Arg-23 which when broken resulted in complete complex dissociation.

Given the fingerprint biasing phenomenon (***Figure 6C***), we finally sought to correct the single/double peak counting statistics (***Figure 4B***) in order to correct for undercounting of single peaks due solely to their failure to reach sufficiently high forces to unfold the fingerprint domain. Only traces showing a fingerprint were analyzed to ensure defined unfolding geometry. Using the rupture force distributions of singles, doubles, iLOV, and xylanase domains, we calculated the probability of occurrence of fingerprint unfolding at a force higher than the single-event ruptures. This overlap probability was found to be 0.85 for iLOV and 0.40 for xylanase. When the single/double peak ratios for were corrected for this effect, the final binding mode ratios (binding mode A/binding mode B, i.e., doubles/singles) were found to be 0.95 and 0.87 for xylanase-Doc and iLOV, respectively. These ratios are close to 1 indicating comparable probability of each binding mode after accounting for biasing the single/double peak counting statistics due to fingerprint domain stability. We note that these numbers are also slightly lower than unity due to the exclusion of double peaks that occurred before unfolding of the fingerprint domains. Further details on rupture force distributions and overlap statistics are shown in ***Figure 7***. As the magnitude of biasing changes with the unfolding force distributions of each fingerprint domain, overlaps in the probability distributions allow for normalizing single/double event ratios of experimental data sets with different fingerprinting domains. For the Coh:Doc complex unbinding event, biasing (undercounting) is more pronounced for the mechanically weaker single ruptures. This normalization procedure shows the relative difference of biasing between single and double events, as double events are less biased than single events.

The biological significance of Coh-Doc interactions in the context of cellulosome assembly and catalysis cannot be overstated. Their high affinity and specificity, along with their modularity, thermostability, and their ultrastable mechanical properties all make Coh-Doc unique from a biophysics perspective, and attractive from an engineering standpoint. Dual binding mode Doc domains are broadly predicted among many cellulosome producing bacteria (e.g. *C. thermocellum, C. cellulolyticum, R. flavefaciens*), however relatively few have been confirmed experimentally (***Carvalho et al., 2007***; ***Pinheiro et al., 2008***; ***Brás et al., 2012***). In fact, the direct effect of single vs. dual binding modes on the ability of cellulosomes to convert substrate into sugars is currently unknown. It is therefore unclear whether or not dual binding modes affect, for example, the catalytic properties of native or engineered synthetic cellulosomes.

However, it is important to note that cellulosome producing bacteria invariably live among communities with other microorganisms, which may be producing cellulases and cellulosomes of their own. In such an environment, a dual binding mode could enable organisms to produce enzymes that are able to bind to a neighboring species' scaffoldins, yet still retain high-affinity interactions with host scaffoldins. They would be able to combine resources with neighboring cells in a mixed microbial consortium. The dual binding mode could therefore allow genetic drift to explore interspecies protein binding. Indeed, cross-species reactivity between Coh and Doc has been reported (***Haimovitz et al., 2008***). Cellulosome-producing microbes may therefore be pursuing a middle

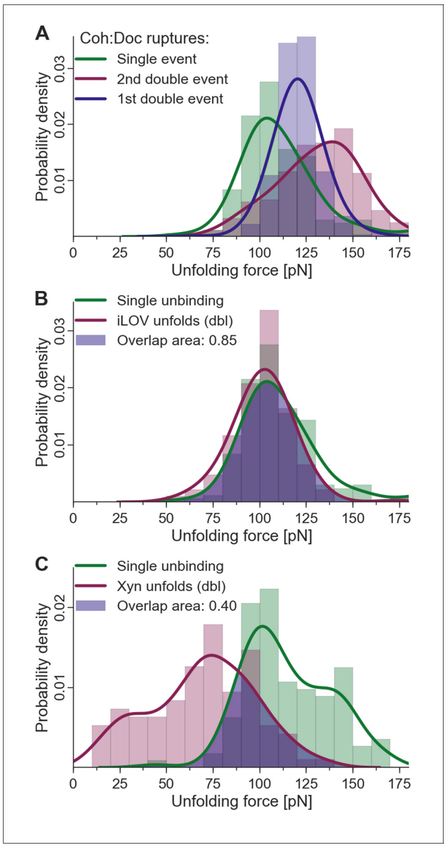


Figure 7. Fingerprint unfolding and complex unbinding forces. (**A**) Rupture force distribution of final complex ruptures for single (green), first (purple) and second (red) double unbinding events. (**B**) Overlap area (purple) of iLOV domain unfolding force distribution (red) (iLOV-doubles curve class) with the rupture force distribution (green) for single-event complex ruptures. (**C**) Overlap area (purple) of Xyn domain unfolding force distribution (red) (Xyn-doubles curve class) with the rupture force distribution (green) for single-event complex ruptures. Overlaps in probability distributions allow normalizing single-event counts to double events to account for different biasing caused by the different unfolding forces of the fingerprint domain. Biasing occurs, because for overlapping force distributions of fingerprint unfolding and complex ruptures, unbinding events are more likely to take place without fingerprint unfolding if the two distributions are closer together. For the Coh:Doc unbinding, this effect is more pronounced for the weaker single ruptures. Because double events are also biased, this still does not give a true quantification, but only compensates for the differences of biasing. The non-bell-evans-like shape of the single rupture peaks, especially in the region of the 1st double event peak (**A**) suggests that this class of curves does not contain a single type of unbinding mechanism, but rather a superposition of different event types.

DOI: 10.7554/eLife.10319.011

ground between protein synthesis strictly for selfish vs. communal usage. By distinguishing the presence of each binding mode for wild-type Doc domains, the single-molecule biophysical approach presented here based on differences in mechanical hierarchies will facilitate further study into the significance of the dual binding mode.

In summary, the dual binding mode of Coh:Doc domains has so far proven resistant to explicit experimental characterization. Crystallography combined with mutagenesis has provided snapshots of the two modes, but resolving each of the modes for wild-type Doc under near native conditions has up until now not been possible. We have demonstrated the advantages of a single-molecule approach in resolving these subtle differences in molecular conformations of bound complexes. Despite having equal thermodynamic binding affinity, when mechanically dissociated by pulling from the N-terminus of Doc, binding mode A was more mechanically stable with an additional energy barrier. This mechanical difference was exploited to probe the two binding modes independently from one another, providing direct observation of this unique mechanism in molecular recognition. In the future, harnessing control over binding modes could offer new approaches to designing molecular assembly systems that achieve defined protein orientations.

Materials and methods

Site-directed mutagenesis of plasmid DNA

A pET28a vector containing the previously cloned xylanaseT6 from *Geobacillus stearothermophilus* (***Salama-Alber et al., 2013***) and DocS dockerin from *Clostridium thermocellum* Cel48S were subjected to QuikChange mutagenesis (***Wang and Malcolm, 1999***) to install the following mutations: Q1, Q3, and QQ in the dockerin and T129C in the xylanase, respectively.

For insertion of the $(GS)_4$ and $(GS)_8$ linkers into the Doc domain, exponential amplification with primers bearing coding sequences for the inserts at their 5'-ends was performed with a Phusion High-Fidelity DNA polymerase (New England Biolabs, MA). PCR products were then blunt end ligated using KLD Enzyme Mix and KLD Reaction Buffer from the Q5 site directed mutagenesis kit (New England Biolabs, MA). The modified DNA constructs were used to transform *Escherichia coli* DH5-alpha cells, grown on kanamycin-containing agar plates and subsequently screened. All mutagenesis products were confirmed by DNA sequencing analysis.

Primers used for inserting the $(GS)_8$ linker into the Doc domain:
Fw 5'-ggttctggctccggttctggctccagcatcaacactgacaat-3'
Rev 5'-agaaccggagccagagccggaacctatacctgatctcaaaacatatct-3'

Protein expression and purification

Fusion proteins HIS-CBM A2C-Coh2 (*C.t.*) were expressed in *E. coli* BL21(DE3)RIPL cells in kanamycin-containing media supplemented with 2mM calcium chloride overnight at 16°C. After harvesting, cells were lysed by sonication, and the lysate was subjected to heat treatment at 60°C for 30 min to precipitate the bulk of the host bacterial proteins, leaving the expressed thermophilic proteins in solution. The lysate was then pelleted, and the supernatant fluids were applied to a beaded cellulose column and incubated at 4°C for 1 hr. The column was then washed with 50 mM Tris buffer (pH 7.4) containing 1.15 M NaCl, and the protein was eluted using a 1% (vl/v) triethylamine aqueous solution. Tris buffer was added to the eluent and the solution was neutralized with HCl.

Fusion proteins HIS-Xyn T129C-DocS (*C.t.*) wild-type, Q1, and Q3 mutants were expressed as described above. Following heat treatment, the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer containing 20mM imidazole and 2mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was then dialyzed to remove the imidazole.

Fusion proteins ybbR-HIS-CBM A2C-Coh2 (*C.t.*), ybbR-HIS-Xyn T129C-DocS (*C.t.*) wild-type and QQ mutants and ybbR-HIS-Xyn T129C-DocS (*C.t.*) $(GS)_4$ insert were expressed in *E. coli* BL21(DE3) RIPL cells; ybbR-HIS-Xyn T129C-DocS (*C.t.*) $(GS)_8$ insert fusion protein variants were expressed in *E. coli* NiCo21(DE3)RIPL cells. Cultivation and expression was done in ZYM-5052 autoinduction media (***Studier, 2005***) containing kanamycin (and chloramphenicol, in case of the NiCo21(DE3)RIPL cells) overnight at 22°C, overall 24 hr. After harvesting, cells were lysed using sonication. The lysate was then pelleted by centrifugation at 39,000 rcf, the supernatant fluids were applied to Ni-NTA columns and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with polyacrylamide gravity flow columns or with polyacrylamide spin desalting columns.

All protein solutions were concentrated with Amicon centrifugal filter devices and stored in 50% (v/v) glycerol at -20°C (ybbR-free constructs) or -80°C (ybbR-bearing constructs). The concentrations of the protein stock solutions were determined to be in the order of 1–15 mg/mL by absorption spectrophotometry at a wavelength of 280 nm.

ELISA-like binding assay

1 µM of Xyn-Doc fusion proteins (wild-type Q1, Q3, QQ Doc fusions) bearing either wild-type or mutant Doc domains were adsorbed onto surfaces of the wells of a 96-well nunc maxi sorp plate (Thermo Scientific, Pittsburgh, PA). After blocking (2% (w/v) BSA, 0.05% Tween 20 in TBS buffer) and several rinsing steps, a red fluorescent protein-cohesin (StrepII-TagRFP-Coh2 (*C.t.*), Addgene ID 58,710 (**Otten et al., 2014**)) fusion construct was incubated to the unspecifically immobilized Doc fusion proteins over a range of concentrations. After further rinsing, the fluorescence of the TagRFP domain was measured with a multi-well fluorescence plate reader (M1000 PRO, Tecan Group Ltd., Männedorf, Switzerland). Fluorescence values were plotted against their corresponding concentration values for each protein variant, and 4 parameter logistic nonlinear regression model functions were fitted to the data to determine the transition point of the curve.

Surface immobilization strategies

The Xyn domain had a cysteine point mutation at position 129 (Xyn T129C) to facilitate covalent attachment to a glass surface *via* Polyethylene glycol (PEG)-maleimide linkers. There were no other cysteines within the Xyn or Doc domains, which ensured site-specific immobilization of the molecule and defined mechanical loading of Doc from the N-terminus for the AFM experiments. The CBM domain likewise contained an A2C cysteine point mutation for covalent attachment to the cantilever tip *via* PEG-maleimide linkers. The second set of fusion proteins sfGFP-Doc and iLOV-Coh was covalently attached to coenzyme A bearing PEG linkers by their terminal ybbR tags.

AFM sample preparation

For AFM measurements, silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation nominal spring constant: 100 pN/nm; 25 kHz resonance frequency in water), and glass coverslips (Menzel Gläser, Braunschweig, Germany; diameter 22mm) were used. 3-Aminopropyl dimethyl ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany), α-Maleinimidohexanoic-ω-NHS PEG (NHS-PEG-Mal, Rapp Polymere, Tübingen, Germany; PEG-MW: 5 kDa), immobilized tris (2-carboxylethyl)phosphine (TCEP) disulfide reducing gel (Thermo Scientific, Pittsburgh, PA), tris (hydroxymethyl) aminomethane (TRIS, >99% p.a., Carl Roth, Karlsruhe, Germany), $CaCl_2$ (>99% p.a., Carl Roth, Karlsruhe, Germany), sodium borate (>99.8% p.a., Carl Roth, Karlsruhe, Germany), NaCl (>99.5% p.a., Carl Roth, Karlsruhe, Germany), ethanol (>99% p.a.), toluene (>99.5% p.a., Carl Roth, Karlsruhe, Germany) were used as received. Sodium borate buffer was 150 mM, pH 8.5. Measurement buffer for AFM-SMFS was tris-buffered saline supplemented with 1 mM $CaCl_2$ (TBS, 25 mM TRIS, 75 mM NaCl, 1 mM $CaCl_2$ pH 7.2). All buffers were filtered through a sterile 0.2 µm polyethersulfone membrane filter (Nalgene, Rochester, NY) prior to use.

Force spectroscopy measurement samples, measurements and data analysis were prepared and performed according to previously published protocols (**Jobst et al., 2013**;**Otten et al., 2014**). In brief, NHS-PEG-Maleimide linkers were covalently attached to cleaned and amino-silanized silicon nitride AFM cantilevers and cover glasses. The respective protein constructs were covalently linked either *via* engineered cysteine residues to the maleimide groups of the surface on the sample directly, or *via* Sfp phosphopantetheinyl transferase-mediated attachment of a terminal ybbR tag to coenzyme A, which was previously attached to the maleimide groups of the surface.

AFM-SMFS measurements

AFM data were recorded in 25 mM TRIS pH 7.2, 75 mM NaCl and 1mM $CaCl_2$ buffer solution (TBS). Retraction velocities for constant speed force spectroscopy measurements varied between 0.2 and 3.2 µm/s. Cantilever spring constants were calibrated utilizing the thermal method applying the equipartition theorem to the one dimensionally oscillating lever (***Hutter and Bechhoefer, 1993***; ***Cook et al., 2006***). Measurements were performed on custom built instruments, deploying an Asylum Research (Santa Barbara, CA, USA) MFP-3D AFM controller and Physik Instrumente (Karlsruhe, Germany) or attocube (Munich, Germany) piezo nanopositioners (***Gumpp et al., 2009***). After each measurement, the xy-stage was actuated by 100 nm to probe a new spot on the surface and measure new individual Xyn-Doc fusion molecules. Instrument control software was programmed in Igor Pro 6.3 (Wavemetrics). The retraction speed was controlled with a closed-loop feedback system running internally on the AFM controller field-programmable gate array (FPGA).

Force-extension data analysis

Data analysis and plotting was performed in Python (Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org) utilizing the libraries NumPy and SciPy (***van der Walt et al., 2011***) and Matplotlib (***Hunter, 2007***).

Measured raw data were analyzed by determining the zero force value with the baseline position and applying a cantilever bending correction to the z-position. The resulting force distance traces were coarsely screened for peaks as sudden drops in force and curves with less than three peaks (such as in ***Figure 3—figure supplement 1***, panel F) were excluded, as they contain no clearly identifiable signal. Force-distance traces were transformed into contour length space with the inverse worm-like-chain model (***Jobst et al., 2013***), assuming a fixed persistence length of 0.4 nm. Screening for the 89 nm xylanase, the 36nm iLOV and the final 8 nm final double rupture increment was performed by finding their corresponding local maxima in a kernel density estimate with bandwidth b = 1 nm. Thresholds in force, distance, and peak counts were applied to sort out nonspecific and multiple interactions. All curves were ultimately selected for the xylanase or iLOV fingerprint and checked manually. For the counting statistics, double peaks were detected as an increment of 8 +- 4 nm in contour length for final rupture peaks in the contour length plot, given that the curve showed one of the fingerprints. If a double peak was detected, the force difference was determined as the percentual difference between the first and the final rupture peak force.

Barrier position diagrams were assembled using optimal alignment through cross-correlation (***Puchner et al., 2008***; ***Otten et al., 2014***). The numbers of points included in fitted histograms are provided in the figure captions, along with the statistical tests and significance values obtained.

Amino acid sequences

pET28a-HIS-XynT129C-DocS (*C.t.*) wild-type

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNKQLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNVVVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNSTDAVALKRYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-ybbR-HIS-XynT129C-DocS (*C.t.*) 16aa GS Insert

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNKQLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNVVVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNSTDAVALKRYVLRSGIGSGSGSGSGSGSGSSINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) Q1 mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-

QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNVVVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNDEDAVALAAYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) Q3 mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNKQLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNVVVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNSTDAVALKRYVLRSGISINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) QQ mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNKQLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNVVVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNDEDAVALAAYVLRSGISINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN

pET28a-ybbR-HIS-sfGFP-DocIS (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGITHGMDELYKKVVPGTPSTKLYGDVNDDGKVNSTDAVALKRYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-ybbR-HIS-CBM A2C-Coh2 (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASMCNTVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWCDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAGSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVNVGNAT

pET28a-ybbR-HIS-iLOV-Coh2 (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASGSPEFIEKNFVITDPRLPDNPIIFASDGFLELTEYSREEILGRNARFLQGPETDQATVQKIRDAIRDQRETTVQLINYTKSGKKFWNLLHLQPVRDQKGELQYFIGVQLDGSDHVGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAGSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVNVGNAT

pET28a-StrepII-TagRFP-Coh2 (*C.t.*)

MWSHPQFEKVSKGEELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILATSFMYGSRTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFPSNGPVMQKKTLGWEANTEMLYPADGGLEGRSDMALKLVGGGHLICNFKTTYRSKKPAKNLKMPGVYYVDHRLERIKEADKETYVEQHEVAVARYCDLPSKLGHKLNGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAGSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVNVGNAT

Acknowledgements

The authors acknowledge Carlos Fontes, Sarah Teichmann, Stefan Stahl, and Ellis Durner for helpful discussions. Support for this work was provided by the ERC Advanced Grant CelluFuel, and the EU 7th Framework Programme NMP4- SL-2013-604530 (CellulosomePlus), and the German-Israeli Foundation (GIF) for Scientific Research and Development. MAN acknowledges support from Society in Science – The Branco Weiss Fellowship from ETH Zurich.

Additional information

Funding

Funder	Grant reference number	Author
European Research Council	294438	Hermann E Gaub
European Commission	NMP4- SL-2013-604530	Daniel B Fried
German-Israeli Foundation for Scientific Research and Development	G-147-207.4-2012	Edward A Bayer Hermann E Gaub Michael A Nash
Society in Science	Branco Weiss Fellowship	Michael A Nash

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

MAJ, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting and revising the article, Contributed reagents; LFM, Conception and design, Analysis and interpretation of data, Drafting and revising the article, Contributed unpublished essential data or reagents; CS, Acquisition of data, Analysis and interpretation of data, Drafting and revising the article; WO, Acquisition of data, Contributed reagents, Drafting and revising the article; DBF, EAB, Conception and design, Contributed reagents, Drafting and revising the article; HEG, MAN, Conception and design, Analysis and interpretation of data, Drafting and revising the article

References

Bayer EA, Belaich JP, Shoham Y, Lamed R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology* **58**:521–554. doi: 10.1146/annurev.micro.57.030502.091022

Beitz E. 2000. TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics* **16**:135–139. doi: 10.1093/bioinformatics/16.2.135

Bertz M, Rief M. 2009. Ligand binding mechanics of maltose binding protein. *Journal of Molecular Biology* **393**:1097–1105. doi: 10.1016/j.jmb.2009.08.066

Brás JL, Alves VD, Carvalho AL, Najmudin S, Prates JA, Ferreira LM, Bolam DN, Romão MJ, Gilbert HJ, Fontes CM. 2012. Novel clostridium thermocellum type i cohesin-dockerin complexes reveal a single binding mode. *The Journal of Biological Chemistry* **287**:44394–44405. doi: 10.1074/jbc.M112.407700

Béguin P, Aubert J-P. 1994. The biological degradation of cellulose. *FEMS Microbiology Reviews* **13**:25–58. doi: 10.1111/j.1574-6976.1994.tb00033.x

Carrion-Vazquez M, Marszalek PE, Oberhauser AF, Fernandez JM. 1999. Atomic force microscopy captures length phenotypes in single proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96**:11288–11292. doi: 10.1073/pnas.96.20.11288

Carvalho AL, Dias FMV, Prates JAM, Nagy T, Gilbert HJ, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA. 2003. Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America* **100**:13809–13814. doi: 10.1073/pnas.1936124100

Carvalho AL, Dias FMV, Nagy T, Prates JAM, Proctor MR, Smith N, Bayer EA, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA, Gilbert HJ. 2007. Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America* **104**:3089–3094. doi: 10.1073/pnas.0611173104

Chapman S, Faulkner C, Kaiserli E, Garcia-Mata C, Savenkov EI, Roberts AG, Oparka KJ, Christie JM. 2008. The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**:20038–20043. doi: 10.1073/pnas.0807551105

Chen C, Cui Z, Xiao Y, Cui Q, Smith SP, Lamed R, Bayer EA, Feng Y, Cui Q, Smith SP, Lamed R, et al. 2014. Revisiting the NMR solution structure of the Cel48S type-I dockerin module from clostridium thermocellum reveals a cohesin-primed conformation. *Journal of Structural Biology* **188**:188–193. doi: 10.1016/j.jsb.2014.09.006

Cook SM, Lang KM, Chynoweth KM, Wigton M, Simmonds RW, Schäffer TE. 2006. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotechnology* **17**:2135–2145. doi: 10.1088/0957-4484/17/9/010

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* **14**:1188–1190. doi: 10.1101/gr.849004

Currie MA, Adams JJ, Faucher F, Bayer EA, Jia Z, Smith SP. 2012. Scaffoldin conformation and dynamics revealed by a ternary complex from the clostridium thermocellum cellulosome. *The Journal of Biological Chemistry* **287**:26953–26961. doi: 10.1074/jbc.M112.343897

Demain AL, Newcomb M, Wu JHD. 2005. Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews* **69**:124–154. doi: 10.1128/MMBR.69.1.124-154.2005

Engel A, Müller DJ. 2000. Observing single biomolecules at work with the atomic force microscope. *Nature Structural Biology* **7**:715–718. doi: 10.1038/78929

Fontes CM, Gilbert HJ. 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annual Review of Biochemistry* **79**:655–681. doi: 10.1146/annurev-biochem-091208-085603

Gumpp H, Stahl SW, Strackharn M, Puchner EM, Gaub HE. 2009. Ultrastable combined atomic force and total internal fluorescence microscope. *Review of Scientific Instruments* **80**:063704. doi: 10.1063/1.3148224

Haimovitz R, Barak Y, Morag E, Voronov-Goldman M, Shoham Y, Lamed R, Bayer EA. 2008. Cohesin-dockerin microarray: diverse specificities between two complementary families of interacting protein modules. *Proteomics* **8**:968–979. doi: 10.1002/pmic.200700486

Hall BA, Sansom MSP. 2009. Coarse-grained MD simulations and protein−protein interactions: the cohesin−dockerin system. *Journal of Chemical Theory and Computation* **5**:2465–2471. doi: 10.1021/ct900140w

Hoffmann T, Tych KM, Hughes ML, Brockwell DJ, Dougan L. 2013. Towards design principles for determining the mechanical stability of proteins. *Physical Chemistry Chemical Physics* **15**:15767–15780. doi: 10.1039/c3cp52142g

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**:33–38. doi: 10.1016/0263-7855(96)00018-5

Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**:90–95. doi: 10.1109/MCSE.2007.55

Hutter JL, Bechhoefer J. 1993. Calibration of atomic-force microscope tips. *Review of Scientific Instruments* **64**:1868 doi: 10.1063/1.1143970

Jobst MA, Schoeler C, Malinowska K, Nash MA. 2013. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *Journal of Visualized Experiments* **82** doi: 10.3791/50950

Karpol A, Kantorovich L, Demishtein A, Barak Y, Morag E, Lamed R, Bayer EA. 2009. Engineering a reversible, high-affinity system for efficient protein purification based on the cohesin-dockerin interaction. *Journal of Molecular Recognition* **22**:91–98. doi: 10.1002/jmr.926

Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**:363–371. doi: 10.1038/nprot.2009.2

Kufer SK, Puchner EM, Gumpp H, Liedl T, Gaub HE. 2008. Single-molecule cut-and-paste surface assembly. *Science* **319**:594–596. doi: 10.1126/science.1151424

Kufer SK, Strackharn M, Stahl SW, Gumpp H, Puchner EM, Gaub HE. 2009. Optically monitoring the mechanical assembly of single molecules. *Nature Nanotechnology* **4**:45–49. doi: 10.1038/nnano.2008.333

Li H, Cao Y. 2010. Protein mechanics: from single molecules to functional biomaterials. *Accounts of Chemical Research* **43**:1331–1341. doi: 10.1021/ar100057a

Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, Gaintzeva A, Lukyanov KA, Lukyanov S, Gadella TW, Chudakov DM. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* **4**:555–557. doi: 10.1038/nmeth1062

Otten M, Ott W, Jobst MA, Milles LF, Verdorfer T, Pippig DA, Nash MA, Gaub HE, Milles LF, Verdorfer T, Pippig DA, et al. 2014. From genes to protein mechanics on a chip. *Nature Methods* **11**:1127–1130. doi: 10.1038/nmeth.3099

Pagès S, Bélaïch A, Bélaïch JP, Morag E, Lamed R, Shoham Y, Bayer EA. 1997. Species-specificity of the cohesin-dockerin interaction between clostridium thermocellum and clostridium cellulolyticum: prediction of specificity determinants of the dockerin domain. *Proteins* **29**:517–527.

Pinheiro BA, Proctor MR, Martinez-Fleites C, Prates JA, Money VA, Davies GJ, Bayer EA, Fontesm CM, Fierobe HP, Gilbert HJ. 2008. The clostridium cellulolyticum dockerin displays a dual binding mode for its cohesin partner. *The Journal of Biological Chemistry* **283**:18422–18430. doi: 10.1074/jbc.M801533200

Pinheiro BA, Gilbert HJ, Sakka K, Sakka K, Fernandes VO, Prates JA, Alves VD, Bolam DN, Ferreira LM, Fontes CM. 2009. Functional insights into the role of novel type I cohesin and dockerin domains from clostridium thermocellum. *The Biochemical Journal* **424**:375–384. doi: 10.1042/BJ20091152

Puchner EM, Franzen G, Gautel M, Gaub HE. 2008. Comparing proteins by their unfolding pattern. *Biophysical Journal* **95**:426–434. doi: 10.1529/biophysj.108.129999

Roberts E, Eargle J, Wright D, Luthey-Schulten Z. 2006. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**:382 doi: 10.1186/1471-2105-7-382

Sakka K, Sugihara Y, Jindou S, Sakka M, Inagaki M, Sakka K, Kimura T. 2011. Analysis of cohesin-dockerin interactions using mutant dockerin proteins. *FEMS Microbiology Letters* **314**:75–80. doi: 10.1111/j.1574-6968.2010.02146.x

Salama-Alber O, Jobby MK, Chitayat S, Smith SP, White BA, Shimon LJ, Lamed R, Frolow F, Bayer EA. 2013. Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *The Journal of Biological Chemistry* **288**:16827–16838. doi: 10.1074/jbc.M113.466672

Schoeler C, Malinowska KH, Bernardi RC, Milles LF, Jobst MA, Durner E, Ott W, Fried DB, Bayer EA, Schulten K, Gaub HE, Nash MA. 2014. Ultrastable cellulosome-adhesion complex tightens under load. *Nature Communications* **5**:5635–35. doi: 10.1038/ncomms6635

Schoeler C, Bernardi RC, Malinowska KH, Durner E, Ott W, Bayer EA, Schulten K, Nash MA, Gaub HE. 2015. Mapping mechanical force propagation through biomolecular complexes. *Nano Letters* **15**:7370–7376. doi: 10.1021/acs.nanolett.5b02727

Schwarz WH. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology* **56**:634–649. doi: 10.1007/s002530100710

Slutzki M, Jobby MK, Chitayat S, Karpol A, Dassa B, Barak Y, Lamed R, Smith SP, Bayer EA, Barak Y, et al. 2013. Intramolecular clasp of the cellulosomal ruminococcus flavefaciens ScaA dockerin module confers structural stability. *FEBS Open Bio* **3**:398–405. doi: 10.1016/j.fob.2013.09.006

Smith SP, Bayer EA. 2013. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Current Opinion in Structural Biology* **23**:686–694. doi: 10.1016/j.sbi.2013.09.002

Stahl SW, Nash MA, Fried DB, Slutzki M, Barak Y, Bayer EA, Gaub HE. 2012. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America* **109**:20431–20436. doi: 10.1073/pnas.1211929109

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**:207–234. doi: 10.1016/j.pep.2005.01.016

Valbuena A, Oroz J, Hervas R, Vera AM, Rodriguez D, Menendez M, Sulkowska JI, Cieplak M, Carrion-Vazquez M. 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proceedings of the National Academy of Sciences of the United States of America* **106**:13791–13796. doi: 10.1073/pnas.0813093106

Wang W, Malcolm BA. 1999. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *BioTechniques* **26**:680–682.

Woodside MT, Block SM. 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annual Review of Biophysics* **43**:19–39. doi: 10.1146/annurev-biophys-051013-022754

van der Walt Stefan, Colbert SC, Varoquaux Gaël. 2011. The NumPy array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **13**:22–30. doi: 10.1109/MCSE.2011.37

5.4 PUBLICATION P5: Microfluidic Chip IVTT SMFS

The following publication presents a multiplexed SMFS assay based on cell-free *in vitro* transcription and translation (IVTT) in a microfluidic platform for the parallelized synthesis, covalent immobilization, and SMFS measurement of target proteins. This lab-on-a-chip approach to AFM-based SMFS makes the tedious classical way of gene expression in *E. coli* cultures, and subsequent affinity purification obsolete. Thereby it enables higher throughput by performing parallel synthesis of different target molecules, and their subsequent covalent surface functionalization in a single microfluidic platform. The resulting array of molecules can readily be used in an SMFS assay. I contributed to this work by developing suited AFM measurement protocols, performing SMFS experiments, and analyzing and interpreting the data. I co-authored the manuscript and created the data-driven parts of the figures.

From Genes to Protein Mechanics on a Chip

Marcus Otten[†], Wolfgang Ott[†], Markus A. Jobst[†],
Lukas F. Milles, Tobias Verdorfer, Diana A. Pippig,
Michael A. Nash and Hermann E. Gaub

[†]*these authors contributed equally to this publication*

published in

Nature Methods, 11(11), 1127-1130, (2014)

Reprinted from [213], with permission from Nature Publishing Group

From genes to protein mechanics on a chip

Marcus Otten[1,2,4], Wolfgang Ott[1,2,4], Markus A Jobst[1,2,4], Lukas F Milles[1,2], Tobias Verdorfer[1,2], Diana A Pippig[1–3], Michael A Nash[1,2] & Hermann E Gaub[1,2]

Single-molecule force spectroscopy enables mechanical testing of individual proteins, but low experimental throughput limits the ability to screen constructs in parallel. We describe a microfluidic platform for on-chip expression, covalent surface attachment and measurement of single-molecule protein mechanical properties. A dockerin tag on each protein molecule allowed us to perform thousands of pulling cycles using a single cohesin-modified cantilever. The ability to synthesize and mechanically probe protein libraries enables high-throughput mechanical phenotyping.

Mechanical forces play a pivotal role in biological systems by performing tasks such as guiding cell adhesion[1], inducing gene expression patterns[2] and directing stem cell differentiation[3]. At the molecular level, mechanosensitive proteins act as sensors and transducers, communicating the presence and direction of applied forces to downstream signaling cascades. Conformational changes in response to mechanical forces[4] and energetic barriers along unfolding pathways can be probed by single-molecule force spectroscopy (SMFS) techniques[4]. Such techniques, including optical tweezers, magnetic tweezers and atomic force microscopy (AFM), have been used to interrogate high-affinity receptor-ligand binding[5], measure unfolding and refolding dynamics of individual protein domains[6–8], observe base-pair stepping of RNA polymerases[9] and identify DNA stretching and twisting moduli[10].

Despite these successes, SMFS experiments have been limited by low throughput. Experimental data sets typically contain a majority of unusable force-distance traces owing to the measurement of multiple molecular interactions in parallel or a lack of specific interactions. Typical yields of interpretable single-molecule interaction traces in SMFS experiments vary between 1% and 25%. The incapacity of SMFS to quickly screen libraries of molecular variants has hindered progress toward understanding sequence-structure-function relationships at the single-molecule level. In particular, the need to prepare each protein sample and cantilever separately increases experimental workload and gives rise to calibration uncertainties. Therefore, methods to interrogate the mechanical behavior of different proteins in a parallel and streamlined format with the same cantilever would offer distinct advantages. Such a screening approach could characterize single-molecule properties such as unfolding forces, interdomain mechanical signatures and mechanically activated catch-bond behavior[1]. Screening of these properties could find applications in biotechnology and human health studies in which mechanical dysregulation or misfolding is suspected to play a role in pathology[11].

Here we developed a platform for parallel characterization of individual protein mechanics in a single experiment (**Fig. 1**). Microspotted gene arrays were used to synthesize fusion proteins *in situ* using cell-free gene expression. Proteins were covalently immobilized inside multilayer microfluidic circuits. A single cantilever was then positioned above the protein array and used to probe the mechanical response of each individual protein via a common C-terminal dockerin (Doc) fusion tag. Genes of interest were chosen such that each gene product exhibited an identifiable unfolding pattern when loaded from the N to the C terminus. Each target protein was expressed with an N-terminal 11-amino-acid ybbR tag, which was used to covalently and site-specifically link the protein to the surface via Sfp synthase–catalyzed reaction with coenzyme A (CoA)[12]. At the C terminus the proteins contained a 75-amino-acid cellulosomal Doc from *Clostridium thermocellum*[13] as a specific handle targeted by the cohesin (Coh)-modified cantilever.

The gene microarray was aligned and reversibly bonded to a microfluidic chip known as MITOMI (mechanically induced trapping of molecular interactions). The chip has been used in the past for screening transcription factors[14,15] and mapping interaction networks[16]. More recently, our group employed MITOMI chips for molecular force assays[17]. In this work, MITOMI chips featured 640 dumbbell-shaped unit cells in a flow layer and 2,004 micromechanical valves in a control layer. Each unit cell was equipped with pneumatic 'neck', 'sandwich' and 'button' valves (**Fig. 1a**) according to design principles of soft lithography[18]. Each neck valve protected the microspotted DNA in the back chamber from exposure to other reagents during surface patterning in the front chamber. The sandwich valves prevented chamber-to-chamber cross contamination, ensuring that only a single protein variant was present in each sample spot. For surface chemistry in the front chamber, the button valves were actuated to shield the sample spots, allowing *n*-dodecyl β-D-maltoside passivation in the surrounding area. Releasing the button valves allowed subsequent functionalization with CoA-poly(ethylene glycol) (CoA-PEG) in the sample area under the buttons serving as the protein immobilization site. We expressed the genes by

---

[1]Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-Universität, Munich, Germany. [2]Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität, Munich, Germany. [3]Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, Munich, Germany. [4]These authors contributed equally to this work. Correspondence should be addressed to M.A.N. (michael.nash@lmu.de).

RECEIVED 3 MARCH; ACCEPTED 29 JULY; PUBLISHED ONLINE 7 SEPTEMBER 2014; CORRECTED AFTER PRINT 5 NOVEMBER 2014; DOI:10.1038/NMETH.3099

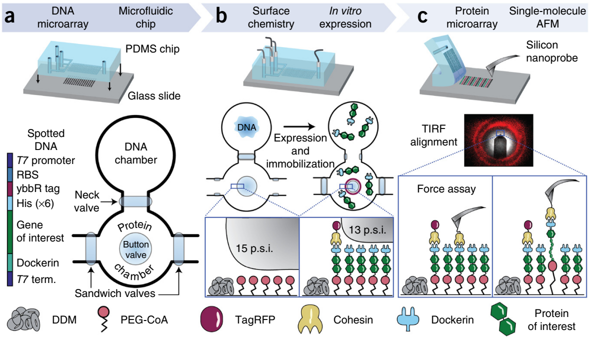


Figure 1 | Method workflow. (**a**) A gene array was spotted onto a glass slide. Genes were designed with a common set of flanking sequences, including a *T7* promoter region, ybbR tag, dockerin tag and *T7* terminator (term.). The multilayer microfluidic chip featuring 640 unit cells was aligned to the DNA microarray and bonded to the glass slide. Each unit cell comprised a DNA chamber, a protein chamber, and superseding elastomeric control valves actuated by pneumatic pressure. PDMS, poly(dimethylsiloxane). (**b**) Control valves were used for spatially selective surface modification of each protein chamber with poly(ethylene glycol)–coenzyme A (PEG-CoA) and for fluidic isolation of each chamber before *in vitro* expression of the microspotted DNA. Fluorescence labeling with TagRFP-cohesin was achieved by partial button-valve pressurization, leaving only an outer concentric ring of immobilized gene products exposed to the labeling solution. DDM, *n*-dodecyl β-D-maltoside. (**c**) After removal of the microfluidic device, the resulting well-defined, covalently attached protein microarray was accessed from above with a cohesin-functionalized atomic force microscope (AFM) cantilever. Single-molecule unfolding traces of each of the protein constructs were thus acquired sequentially at each corresponding array address with a single cantilever in a single experiment.

incubating an *in vitro* transcription and translation cell extract at 37 °C with the spotted DNA in the back chamber. The synthesized proteins then diffused to the front chamber, where they were covalently linked to the surface via an Sfp-catalyzed reaction of surface-bound CoA with solution-phase N-terminal ybbR peptide tags (**Fig. 1b**). Partial pressurization of the button valve[19] was used for tagging an outer concentric portion of the sample area with a fluorescently (TagRFP) tagged Coh that specifically bound to the C-terminal Doc tag of each target protein, thereby confirming successful protein synthesis and surface immobilization (**Supplementary Fig. 1**). Finally, the microfluidic device was removed from the glass slide to provide access to the protein array from above. Using this approach, we generated microarrays of site-specifically and covalently immobilized proteins for subsequent SMFS experiments, starting from a conventional gene array.

An inverted three-channel total-internal-reflection fluorescence/atomic force microscope (TIRF-AFM)[20] was used to position the cantilever in the center of the fluorescent rings in the protein array and perform SMFS measurements (**Fig. 1c**). The Coh-modified cantilever was used to probe the surface for expressed target proteins containing the C-terminal Doc tag. Upon surface contact of the cantilever, formation of a Coh-Doc complex allowed measurement of target-protein unfolding in a well-controlled pulling geometry (N to C terminus). We retracted the probe at constant velocity and recorded force-extension traces that characterized the unfolding fingerprint of the target protein. This approach-retract process could be repeated many times at each array address to characterize each expression construct.

Several unique features of the C-terminal Doc tag make it particularly suitable as a protein handle for SMFS. Its small size of 8 kDa does not notably add to the molecular weight of the gene products, which is advantageous for cell-free expression. Additionally, Doc exhibits a specific and high-affinity interaction with Coh domains from the *C. thermocellum* scaffold protein CipA. Coh was used both for fluorescence detection of the expression constructs and for modification of the cantilever. On the basis of our prior work, the Coh-Doc interaction is characterized to be high affinity, with a dissociation constant K_d in the low nanomolar range and rupture forces >125 pN at a loading rate of 10 nN/s (ref. 21). Our prior work also indicated that upon forced dissociation, Doc exhibited a characteristic double sawtooth rupture peak with a contour length increment of 8 nm separating the two peaks. We used this two-pronged double rupture event at the end of each force-extension trace as a positive indicator

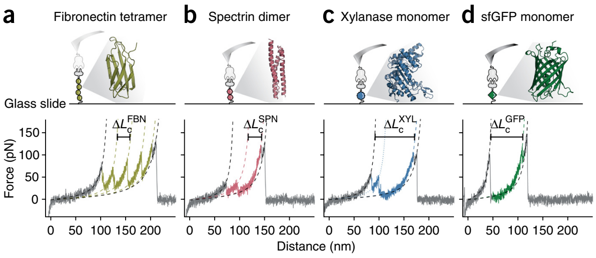


Figure 2 | Representative single-molecule force traces recorded in different protein spots on a single chip with a single cantilever. (**a**–**d**) Four proteins of interest, anchored between the coenzyme A (CoA)-functionalized surface and the cohesin-functionalized cantilever, were probed: fibronectin tetramer (**a**), spectrin dimer (**b**), xylanase monomer (**c**) and sfGFP monomer (**d**). The crystal structure and pulling configuration (top) are shown for each construct. Each single-molecule force-distance trace (bottom) shows the individual unfolding fingerprint of the protein of interest followed by a respective protein of interest followed by a common, final double sawtooth peak (gray) that is characteristic of the cohesin-dockerin rupture. Experimental data were fitted with the worm-like chain model (dashed lines). Unfolding intermediates were also observed (fitted for only xylanase in **c**; dotted colored line).

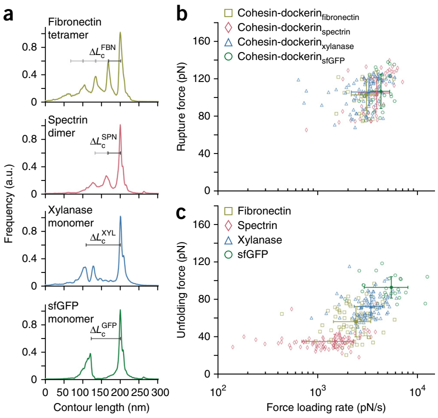


Figure 3 | Unfolding and rupture statistics from multiple force traces. (**a**) Relative frequency of observing given contour lengths determined by transforming and aligning multiple force traces into contour length space via the worm-like chain model. Shown are diagrams for the fibronectin tetramer ($n = 27$, $\Delta L_c^{FBN} = 33$ nm), spectrin dimer ($n = 50$, $\Delta L_c^{SPN} = 34$ nm), xylanase monomer ($n = 91$, $\Delta L_c^{XYL} = 93$ nm) and sfGFP monomer ($n = 25$, $\Delta L_c^{GFP} = 79$ nm). (**b**) Rupture force versus loading rate of the final cohesin-dockerin dissociation event. (**c**) Unfolding force versus loading rate for each protein of interest. The populations in **b** and **c** were fitted with two-dimensional Gaussians. Respective means and s.d. are plotted in the corresponding colors as solid symbols and error bars. a.u., arbitrary units.

that the gene of interest was completely expressed through to the C terminus (**Fig. 2**). Furthermore, this double rupture peak indicated that the interaction with the Coh-modified cantilever was specific and that the pulling geometry was strictly controlled such that force was applied to the molecule of interest from the N to the C terminus.

To validate and demonstrate our approach, we expressed genes of interest comprising well-known fingerprint domains in the SMFS literature. We produced multimeric polyproteins including tetrameric human type-III fibronectin (FBN)[22] and dimeric chicken brain α-spectrin (SPN)[23]. We also synthesized monomers of endo-1,4-xylanase T6 from *Geobacillus stearothermophilus* (XYL)[21], superfolder GFP (GFP)[24] and twitchin kinase[25]. In all cases, the target proteins were fused to N-terminal ybbR and C-terminal Doc tags (**Supplementary Figs. 2**–**6**). Unfolding data for FBN, SPN, XYL and GFP were obtained using a single cantilever to probe a single microarray (**Figs. 2** and **3**). Twitchin kinase was found not to express in sufficient yield to provide reliable unfolding statistics.

We transformed force-extension data (**Fig. 2**) into contour length space[26] using the worm-like chain model and compared the measured contour length increments with the amino acid sequence lengths of each protein and literature values. The observed contour lengths and rupture forces were consistent with our expectations. FBN showed a fourfold-repeated sequence of rupture peaks at contour length increments of 32 nm (ΔL_c^{FBN}; **Fig. 2a**) frequently interrupted by an intermediate peak at 10–12 nm, both features characteristic of FBN[22]. SPN showed two regular sawtooth-like peaks with contour lengths of 33 nm (ΔL_c^{SPN}; **Fig. 2b**)[23]. XYL exhibited a decreasing multipeaked unfolding fingerprint with a contour length increment of 92 nm (ΔL_c^{XYL}; **Fig. 2c**), occasionally showing additional increments corresponding to unfolding of remaining XYL subdomains, a result consistent with the prior study and accounting for N-terminal immobilization of XYL[21]. GFP unfolding showed a contour length increment of 74 nm (ΔL_c^{GFP}; **Fig. 2d**)[24]. As each protein in the array contained the same C-terminal Doc tag, the final two rupture peaks in all force traces represented rupture of the Coh-Doc complex regardless of the protein of interest.

In our system, surface densities of expressed proteins were comparable to those obtained in conventional SMFS experiments. Uninterpretable and nonspecific interactions were excluded from the analysis (**Supplementary Fig. 7**). By collecting multiple unfolding traces, we assembled contour length diagrams for each protein of interest[26,27] (**Fig. 3a**) and confirmed the predicted contour length increments on the basis of the encoded amino acid sequences in each DNA spot. Coh-Doc rupture events for all protein constructs in the array clustered to the same population in the force-loading rate plot, independently of the preceding rupture peaks from the protein of interest (**Fig. 3b**). The Coh-Doc ruptures agreed with previously reported values at similar loading rates[21]. The unfolding events of the proteins of interest produced distinct populations in the force-loading rate plots (**Fig. 3c**). The unfolding events depended on the internal structure and the unfolding pathway of the fingerprint domain when stretched between its N and C termini. SPN, for example, an elongated 3-helix bundle, was previously reported to exhibit a broader energy well ($\Delta x = 1.7$ nm; ref. 23) and showed a flatter distribution of unfolding forces than that of the more compact globular FBN domain with a shorter, steeper potential ($\Delta x = 0.4$ nm; ref. 22).

In summary, our flexible approach efficiently streamlines protein expression, purification and SMFS into a single integrated platform (**Supplementary Discussion**). The approach should be compatible with other *in vitro* expression systems including extracts derived from insects, rabbit reticulocytes and human cell lines, and it is capable of introducing post-translational modifications and non-natural amino acids, allowing, for example, the screening of site-directed mutants. Our method allows for synthesis of cytotoxic proteins or proteins with a tendency to form inclusion bodies during bulk expression. In addition to providing greatly improved throughput, our system has the advantage of measuring multiple constructs with one cantilever, thereby eliminating errors introduced when performing multiple calibrations on different samples with uncertainties of ~10% (ref. 28). Detecting subtle differences in mechanical stability with this high-throughput approach could therefore be used to perform mechanical phenotyping experiments on similarly stable families of mutant proteins. This workflow opens the door to large-scale screening studies of protein nanomechanical properties.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Addgene: pET28a-ybbR-HIS-sfGFP-DocI, 58708; pET28a-ybbR-HIS-CBM-CohI, 58709; pET28a-StrepII-TagRFP-CohI, 58710; pET28a-ybbR-HIS-Xyl-DocI, 58711;

pET28a-ybbR-HIS-10FNIII(x4)-DocI, 58712; pET28a-ybbR-HIS-Spec(x2)-DocI, 58713.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
M.O. is grateful to the Elite Network of Bavaria (IDK-NBT) for a doctoral fellowship. M.A.N. acknowledges support from Society in Science—The Branco Weiss Fellowship administered by the ETH Zürich. The authors acknowledge support from the DFG Sonderforschungsbereich 1032 and the European Research Council Grant Cellufuel (Advanced Grant 294438). The authors thank E. Bayer (Weizmann Institute) for starting genetic materials used for Doc and Coh modules.

AUTHOR CONTRIBUTIONS
M.O., M.A.N. and H.E.G. designed the research; M.O., W.O., M.A.J. and T.V. performed experiments; D.A.P. helped with immobilization strategies; M.O., W.O., M.A.J., L.F.M. and M.A.N. performed data analysis; M.O., W.O., M.A.J., M.A.N. and H.E.G. cowrote the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Thomas, W.E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E.V. *Cell* **109**, 913–923 (2002).
2. Li, C. & Xu, Q. *Cell. Signal.* **12**, 435–445 (2000).
3. Engler, A.J., Sen, S., Sweeney, H.L. & Discher, D.E. *Cell* **126**, 677–689 (2006).
4. Müller, D., Helenius, J., Alsteens, D. & Dufrêne, Y.F. *Nat. Chem. Biol.* **5**, 383–390 (2009).
5. Florin, E.-L., Moy, V.T. & Gaub, H.E. *Science* **264**, 415–417 (1994).
6. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. & Gaub, H. *Science* **276**, 1109–1112 (1997).
7. Fernandez, J.M. & Li, H. *Science* **303**, 1674–1678 (2004).
8. Oesterhelt, F. *et al. Science* **288**, 143–146 (2000).
9. Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R. & Block, S.M. *Nature* **438**, 460–465 (2005).
10. Bryant, Z. *et al. Nature* **424**, 338–341 (2003).
11. Linke, W.A. *Cardiovasc. Res.* **77**, 637–648 (2008).
12. Yin, J. *et al. Proc. Natl. Acad. Sci. USA* **102**, 15815–15820 (2005).
13. Bayer, E.A., Belaich, J.-P., Shoham, Y. & Lamed, R. *Annu. Rev. Microbiol.* **58**, 521–554 (2004).
14. Maerkl, S.J. & Quake, S.R. *Science* **315**, 233–237 (2007).
15. Rockel, S., Geertz, M., Hens, K., Deplancke, B. & Maerkl, S.J. *Nucleic Acids Res.* **41**, e52 (2013).
16. Gerber, D., Maerkl, S.J. & Quake, S.R. *Nat. Methods* **6**, 71–74 (2009).
17. Otten, M., Wolf, P. & Gaub, H.E. *Lab Chip* **13**, 4198–4204 (2013).
18. Thorsen, T., Maerkl, S.J. & Quake, S.R. *Science* **298**, 580–584 (2002).
19. Garcia-Cordero, J.L. & Maerkl, S.J. *Chem. Commun.* **49**, 1264–1266 (2013).
20. Gumpp, H., Stahl, S.W., Strackharn, M., Puchner, E.M. & Gaub, H.E. *Rev. Sci. Instrum.* **80**, 063704 (2009).
21. Stahl, S.W. *et al. Proc. Natl. Acad. Sci. USA* **109**, 20431–20436 (2012).
22. Li, L., Huang, H.H.-L., Badilla, C.L. & Fernandez, J.M. *J. Mol. Biol.* **345**, 817–826 (2005).
23. Rief, M., Pascual, J., Saraste, M. & Gaub, H.E. *J. Mol. Biol.* **286**, 553–561 (1999).
24. Dietz, H. & Rief, M. *Proc. Natl. Acad. Sci. USA* **103**, 1244–1247 (2006).
25. Greene, D.N. *et al. Biophys. J.* **95**, 1360–1370 (2008).
26. Puchner, E.M., Franzen, G., Gautel, M. & Gaub, H.E. *Biophys. J.* **95**, 426–434 (2008).
27. Jobst, M.A., Schoeler, C., Malinowska, K. & Nash, M.A. *J. Vis. Exp.* **82**, e50950 (2013).
28. Gibson, C.T., Smith, D.A. & Roberts, C.J. *Nanotechnology* **16**, 234–238 (2005).

ONLINE METHODS

Chip fabrication. Ready-to-use wafers for flow and control layers of the 640-chamber MITOMI design were obtained from Stanford Microfluidics Foundry (design name DTPAd)[14]. The flow wafer features 15-µm-high features, rounded by photoresist reflow, whereas the control wafer features a rectangular cross-section.

Microfluidic chips were cast in poly(dimethylsiloxane) (PDMS) from these wafers. For the control layer, Sylgard 184 (Dow Corning) base and curing agent were mixed at a ratio of 5:1 by weight, poured onto the wafer, degassed and partially cured for 20 min at 80 °C. For the flow-layer wafer, a 20:1 base–to–curing agent mixture of Sylgard 184 was spin-coated for 75 s at 1,600 r.p.m. and partially cured for 30 min at 80 °C. The control layer chips were cut out, inlet holes were punched and the chips were aligned onto the spin-coated PDMS on the flow-layer wafer. After the two-layer chips were baked for 90 min at 80 °C, they were cut and removed from the wafer, and inlet/outlet holes were punched. Microfluidic chips were stored for up to 6 weeks.

Cloning. For the construction of the fusion proteins, Gibson assembly[29] was used. A ratio of 0.07 pmol vector to 0.3 pmol of insert was used for the fusion reaction. The primer sequences are provided in **Supplementary Table 1**. A pET28a plasmid was linearized with primers 1 and 2. The dockerin type I–encoding gene was isolated from the xylanase-dockerin type I construct[21] with primers 3 and 4. Codon-optimized sequences were purchased from GeneArt/Invitrogen. The genes of interest were designed in such a way that they already contained sequences overlapping those of their neighboring partners (pET28a and dockerin type I). In the case of the spectrin, two domains were linked with a flexible glycine-serine (×6) linker. For fibronectin, four type III domains were fused separated by glycine-serine (×6) linkers. The expression vector in all cases was a pET28a plasmid with a modified multiple cloning site (sequence attached: plasmids are available at Addgene, **Supplementary Table 2**). After construction, clones were verified via sequencing and amplified in NEB 5-alpha *Escherichia coli* cells. Following plasmid preparation, samples were concentrated up to 500 ng/µl before microspotting.

DNA microspotting. A 24 × 60–mm #1 thickness coverslip (Thermo Scientific) was silanized with 3-aminopropyldimethylethoxysilane (ABCR) following literature protocols[30].

The DNA solution containing 1% (w/v) nuclease-free bovine serum albumin (Carl Roth) in nuclease-free water was microspotted under humid atmosphere onto the silanized coverslip using the GIX Microplotter II (Sonoplot) and a glass capillary with a 30-µm tip diameter (World Precision Instruments) according to the manufacturer's instructions in a rectangular 40 × 16 pattern with 320-µm column pitch and 678-µm row pitch. Alignment of the DNA array and the microfluidic chip was done manually using a stereomicroscope. Bonding between the glass cover slip and microfluidic device was achieved by thermal bonding for 5 h at 80 °C on a hot plate.

Protein synthesis on-chip. The microfluidic device was operated at a pressure of 4 p.s.i. in the flow layer and 15 p.s.i. in the control layer. Operation started with the button and neck valves actuated for surface passivation. The flow layer was passivated by flushing through standard buffer (25 mM Tris, 75 mM NaCl, 1 mM CaCl2, pH 7.2) for 5 min and 2% *n*-dodecyl β-D-maltoside (Thermo Scientific) in nuclease-free H_2O for 30 min (ref. 31). Next the button valve was opened, and borate buffer (50 mM sodium borate, pH 8.5) was flushed through for 30 min to deprotonate aminosilane groups on the glass surface.

For maleimide/coenzyme A functionalization, a solution of 5 mM NHS-PEG-maleimide (MW = 513 Da, Thermo Scientific) in borate buffer was flushed through for 45 min. The device was then rinsed with nuclease-free H_2O for 5 min, followed by 30 min of 20 mM coenzyme A (Merck) in coupling buffer (50 mM sodium phosphate, pH 7.2, 50 mM NaCl, 10 mM EDTA). The button valve was then actuated to protect the functionalized surface area followed by 5 min of rinsing with standard buffer.

S30 T7 HY (Promega) *in vitro* transcription and translation mix, supplemented with 1 µL T7 polymerase (Promega) and 0.5 µL RNase inhibitor (Invitrogen), was then flushed into the chip, filling the DNA chambers (neck valve open).

The neck valve was then closed, and the channels were filled with 4′-phosphopantetheinyl transferase (Sfp synthase) in Sfp buffer (50 mM HEPES, 10 mM $MgCl_2$). The chip was then incubated at 37 °C on a hot plate. After 1 h of incubation, the neck and the button valves were opened to allow Sfp synthase–catalyzed linkage of expressed protein to the coenzyme A–functionalized area below the button. At the same time the sandwich valves were actuated to avoid chamber-to-chamber cross-contamination. After another 1.5 h of incubation, the neck and button valves were closed, the sandwich valves were opened and the chip was rinsed with standard buffer for 20 min.

To verify successful protein expression and immobilization on the functionalized surface area, a fluorescent detection construct (TagRFP–cohesin type I (2 µg/ml) in standard buffer) was flushed through the device for 10 min with the button valve actuated. The sandwich valves were then actuated, and the button valve partially released by decreasing the pressure to 11 p.s.i. After 20 min of incubation at room temperature, the sandwich valves were opened, and the chip flushed with standard buffer for 20 min. Fluorescence images of all chambers were recorded on an inverted microscope with a 10× objective (Carl Zeiss), featuring an electron-multiplying charge-coupled device (EMCCD) camera (Andor). Prior to force spectroscopy experiments, the chip was stored in buffer at 4 °C.

Directly before measurement, the PDMS chip was peeled off from the glass substrate under buffer, revealing the microarray while avoiding drying of the functionalized surface. The array surface was then rinsed several times with buffer. We did not encounter any problems with cross-contamination between chambers.

Cantilever functionalization. A silicon-nitride cantilever bearing a silicon tip with a tip radius of ~8 nm (Biolever mini, Olympus) was silanized with ABCR as described previously[30]. Protein functionalization was performed in a similar way as reported previously[27,31]. Briefly, a 50 µM solution of CBM A2C–cohesin from *C. thermocellum* in standard buffer was incubated with 1:2 (v/v) TCEP beads (Tris (2-carboxyethyl) phosphine disulfide reducing gel, Thermo Scientific), previously washed with standard buffer, for 2.5 h. The cantilever was submerged in borate buffer for 45 min to deprotonate primary amine groups on the silanized surface and then incubated with 20 mM NHS-PEG-maleimide (MW = 5 kDa, Rapp Polymere) in borate buffer for 60 min.

The cantilever was rinsed sequentially in three beakers of deionized H_2O. TCEP beads were separated from the protein solution by centrifugation at 1,000g for 1 min. Next the cantilever was incubated for 60 min with reduced protein solution, which was diluted to a concentration of 1 mg/mL with standard buffer. Finally the cantilever was rinsed sequentially in three beakers of standard buffer and stored submerged in standard buffer in humid atmosphere at 4 °C for up to 24 h before use.

Force spectroscopy. A custom-built TIRF (total internal reflection fluorescence)-AFM (atomic force microscope) hybrid[20,30] was used to conduct the force spectroscopy measurements. The TIRF microscope was used to image fluorophores in up to three different color channels simultaneously using an iChrome MLE-S four-color laser (Toptica Photonics), an Optosplit III triple emission image splitter (Cairn Research) and a Xion3 EMCCD camera (Andor). A long-range stick-slip *xy* piezo nanopositioning system (ANC350, Attocube Systems) allowed access to the whole microchip array as well as fine spatial sampling of different surface molecules on the nanometer scale within each protein spot. Cantilever actuation in the *z* direction was performed by a LISA piezo-actuator (Physik Instrumente) driven by an MFP3D AFM controller (Asylum Research).

The following force spectroscopy protocol was performed repeatedly in each functionalized protein target area. The cantilever approach velocity was 3,000 nm/s, dwell time at the surface was 10 ms and retract velocity was 800 nm/s. Data were recorded with 6,250-Hz sampling rate. The cantilever typically had a spring constant in the range of 100 pN/nm and a resonance frequency of 25 kHz in water. Accurate calibration of the system was performed by the nondestructive thermal method[32,33] using corrections to account for discrepancies from the original theory[27,34].

Data and statistical analysis. The raw data were converted from photodiode voltages into force values in newtons, and the following standard corrections were applied. The zero force value for the unloaded cantilever in each curve was determined by averaging over 40-nm extension after the final complex rupture and subtracting this value from each force value in the curve. The position of the surface was determined by finding the force value closest to 0 in a small neighborhood of the first non-negative force value in the force-extension trace. The *z* piezo position was corrected for the true tip-sample separation due to deflection of the lever as a function of the force for a Hookean spring.

A pattern-recognition software based on a package described previously[26] and adapted in-house chose the curves showing worm-like chain force responses of the stretched protein constructs. Example curves showing multiple, unspecific or no interactions are shown in **Supplementary Figure 7**, together with a single xylanase trace for comparison. The expected protein backbone contour length increments for each construct were detected in contour length space: the real part of the following numerically solved inverse worm-like chain (WLC) formula[27] was used to transform force-extension data into force–contour length space for every measured force curve:

$$L(x) = \frac{x}{6u}\left(3 + 4u + \frac{9 - 3u + 4u^2}{g(u)} + g(u)\right)$$

where

$$g(u) = \left(27 - \frac{27}{2}u + 36u^2 - 8u^3 + \frac{3\sqrt{3}}{2}\sqrt{-u^2\left((4u-3)^3 - 108\right)}\right)^{\frac{1}{3}}$$

and

$$u = F\frac{L_p}{kT}$$

with L the contour length, x the extension, F the force, L_p the persistence length, k Boltzmann's constant and T the temperature. Transformed data points were combined in a Gaussian kernel density estimate with a bandwidth of 1 nm and plotted with a resolution of 1 nm. In these resulting energy-barrier position diagrams, the contour length increments could easily be determined. The transformation was performed with the following parameters: persistence length $L_p = 0.4$ nm, thermal energy $kT = 4.1$ pN nm. Force and distance thresholds were applied at 10 pN and 5 nm, respectively. The measurement data sets in each protein spot on the chip typically showed a yield of 0.5–5% specific interactions.

The force peaks corresponding to protein domain unfolding events, as well as those corresponding to final ruptures, were line fitted in force-time space to measure the loading rate of each individual event.

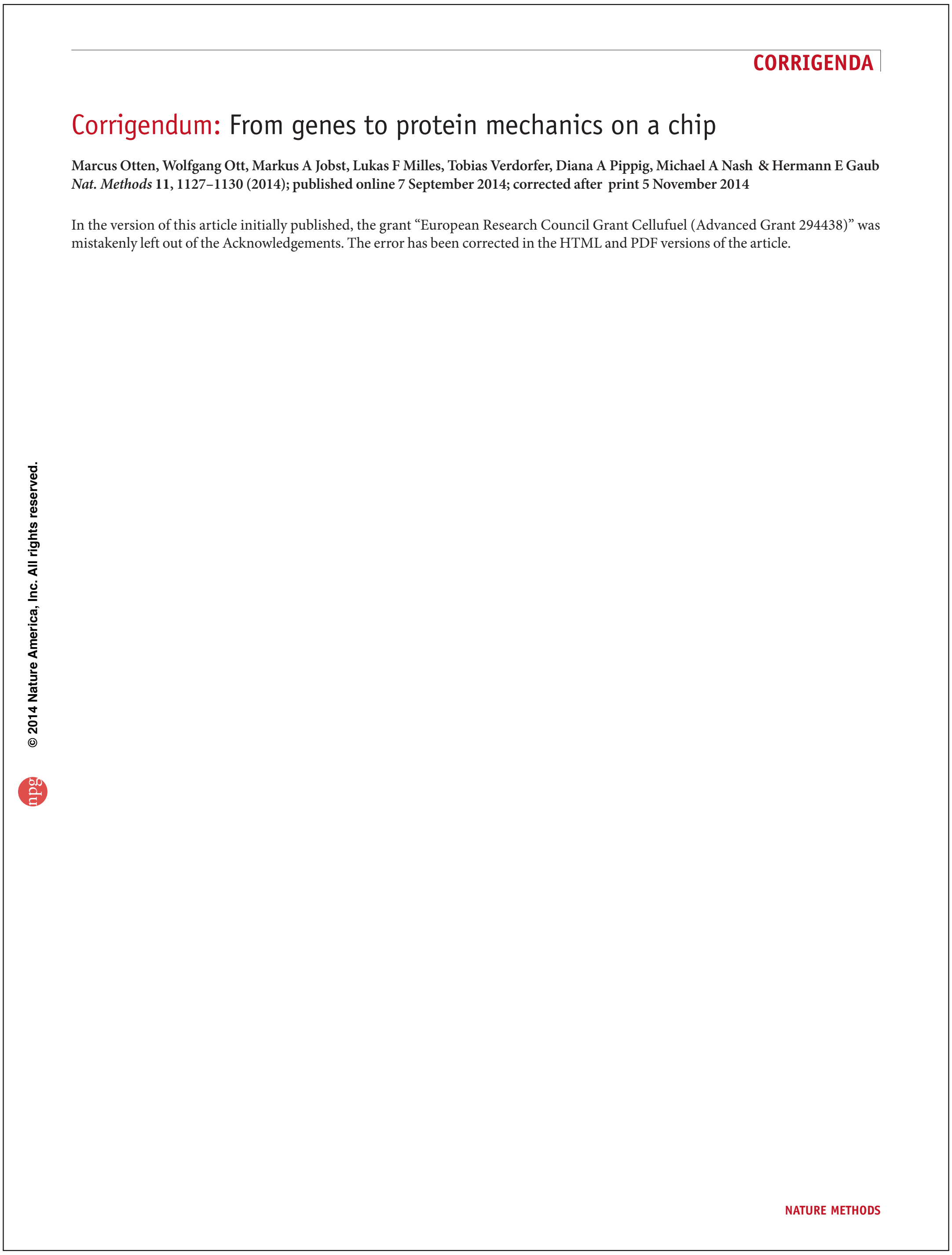
WLC fits for demonstrative purposes in **Figure 2** were done by using the following formula:

$$F(x) = \frac{kT}{L_p}\left(\frac{1}{4(1 - x/L)^2} + \frac{x}{L} - \frac{1}{4}\right)$$

with F the force, k the Boltzmann's constant, T the temperature, L_p the persistence length, x the extension and L the contour length.

Discrepancies between contour length increments in fitted single-molecule traces and aligned contour length diagrams are artifacts caused by the fixed persistence length in the contour length transformation, whereas the WLC fits to single force traces treat both contour length and persistence length of each stretch as free parameters. An overview of the yield of interpretable curves of all constructs is available in **Supplementary Table 3**.

29. Gibson, D.G. *et al. Nat. Methods* **6**, 343–345 (2009).
30. Zimmermann, J.L., Nicolaus, T., Neuert, G. & Blank, K. *Nat. Protoc.* **5**, 975–985 (2010).
31. Huang, B., Wu, H., Kim, S. & Zare, R.N. *Lab Chip* **5**, 1005–1007 (2005).
32. Hutter, J.L. & Bechhoefer, J. *Rev. Sci. Instrum.* **64**, 1868 (1993).
33. Cook, S.M. *et al. Nanotechnology* **17**, 2135–2145 (2006).
34. Proksch, R., Schäffer, T.E., Cleveland, J.P., Callahan, R.C. & Viani, M.B. *Nanotechnology* **15**, 1344–1350 (2004).

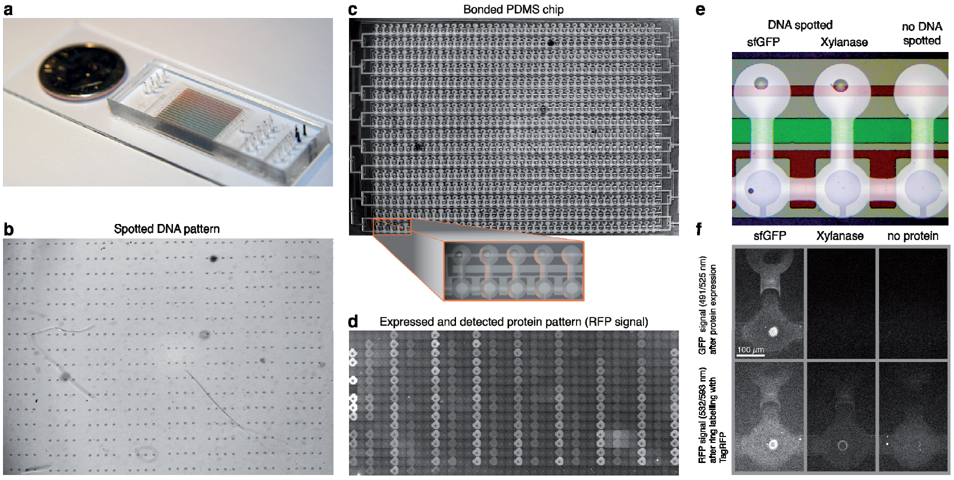


Corrigendum: From genes to protein mechanics on a chip

Marcus Otten, Wolfgang Ott, Markus A Jobst, Lukas F Milles, Tobias Verdorfer, Diana A Pippig, Michael A Nash & Hermann E Gaub

Nat. Methods **11, 1127–1130 (2014); published online 7 September 2014; corrected after print 5 November 2014**

In the version of this article initially published, the grant "European Research Council Grant Cellufuel (Advanced Grant 294438)" was mistakenly left out of the Acknowledgements. The error has been corrected in the HTML and PDF versions of the article.

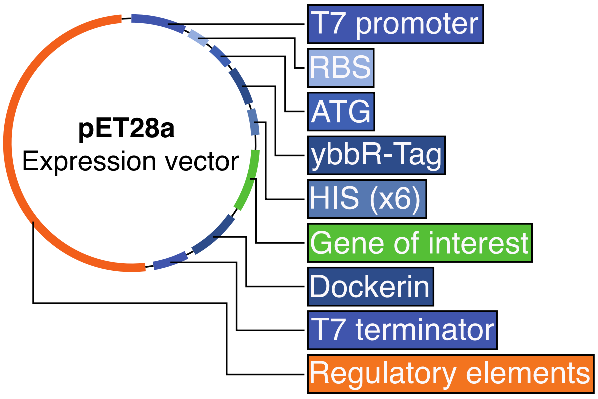


Supplementary Figure 1

Microfluidic chip overview.

(a) Photograph of a microfluidic chip bonded to a glass slide with a US dime for scale. Control channels are filled with food dye for better visualization. (b) Pattern of a typical DNA array, consisting of repeats of rows with four different genes and one row with nothing spotted as negative control. (c) Photograph of a bonded PDMS chip onto the glass slide with DNA spots in the back chamber. The orange highlighted frame shows a zoom in of the bottom left corner. (d) Typical fluorescence collage assembled from 640 single fluorescence micrographs of each protein chamber on one single chip shows pattern of expressed protein (assembly not to scale). Fluorescence signal of TagRFP reveals expression levels and Dockerin specificity. Here, low passivation of the protein chamber facilitates visualization. (e) Three of 640 adjacent dumbbell-shaped chambers, one with sfGFP DNA spotted (left), one with Xylanase DNA (center) and one negative control without DNA (right). Control channels are visualized with food dye: neck valve (green), sandwich valve (red), and button valve (blue). (f) Fluorescence images showing GFP signal (top) from expressed and immobilized ybbR-sfGFP-Dockerin (left), ybbR-Xylanase-Dockerin (center) with negative control lacking the spotted DNA (right). The bottom row shows the signal from the TagRFP detection construct, which specifically bound to the Dockerin tag via the Cohesin domain.

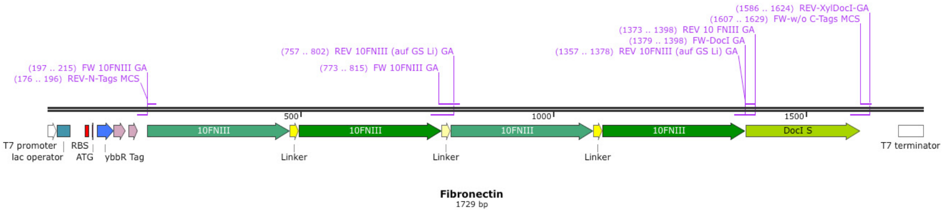
Nature Methods: doi:10.1038/nmeth.3099



Supplementary Figure 2

Diagram of the expression vector pET28a with an individual gene of interest.

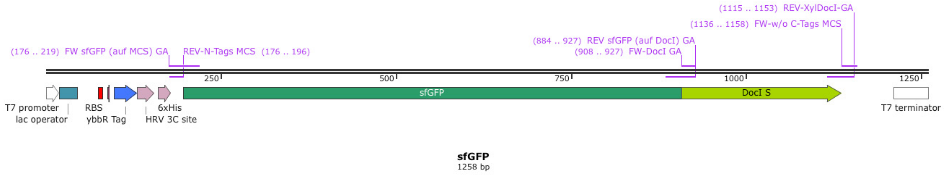
Nature Methods: doi:10.1038/nmeth.3099



Supplementary Figure 3

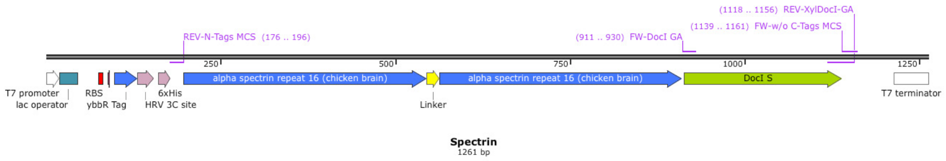
Schematic of the fibronectin tetramer gene cassette.

Nature Methods: doi:10.1038/nmeth.3099



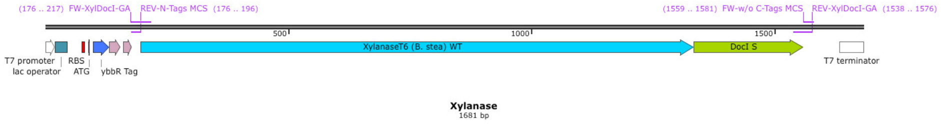
Supplementary Figure 4

Schematic of the sfGFP dimer gene cassette.



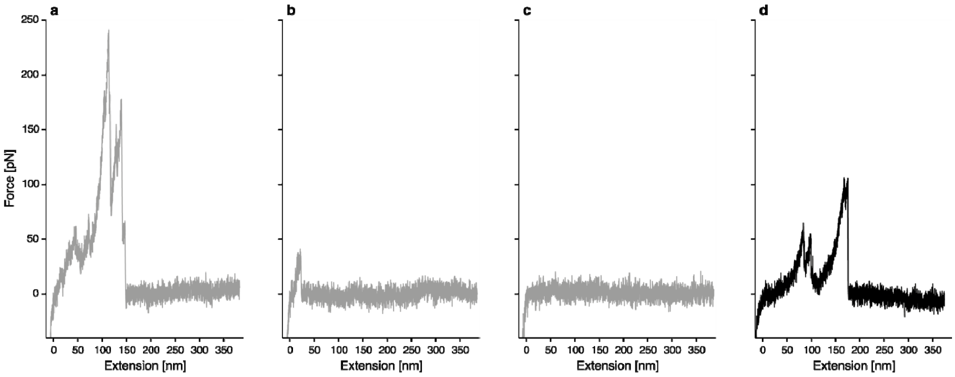
Supplementary Figure 5

Schematic of the spectrin dimer gene cassette.



Supplementary Figure 6

Schematic of the xylanase gene cassette.



Supplementary Figure 7

Exemplary force traces

Example curves showing (a) uninterpretable interaction, (b) non-specific interaction of cantilever with surface, (c) no interaction, and (d) a specific Xylanase-Dockerin unfolding and unbinding trace. Curves similar to those shown in a-c were excluded from the analysis.

Nature Methods: doi:10.1038/nmeth.3099

Supplementary Discussion

Typically in SMFS experiments, rupture force – loading rate plots are used to characterize k_{off} and Δx, the unbinding (or unfolding) probability per time unit and the distance to the transition state along the reaction coordinate, respectively, providing direct information about the energy landscape governing protein folding[1]. SMFS experiments are also complemented by all-atom simulations of such systems *in silico*. Recently, it was shown that high speed SMFS experiments could be performed at speeds achievable in molecular dynamics simulations[2], overcoming a long standing discrepancy between experiment and simulation.

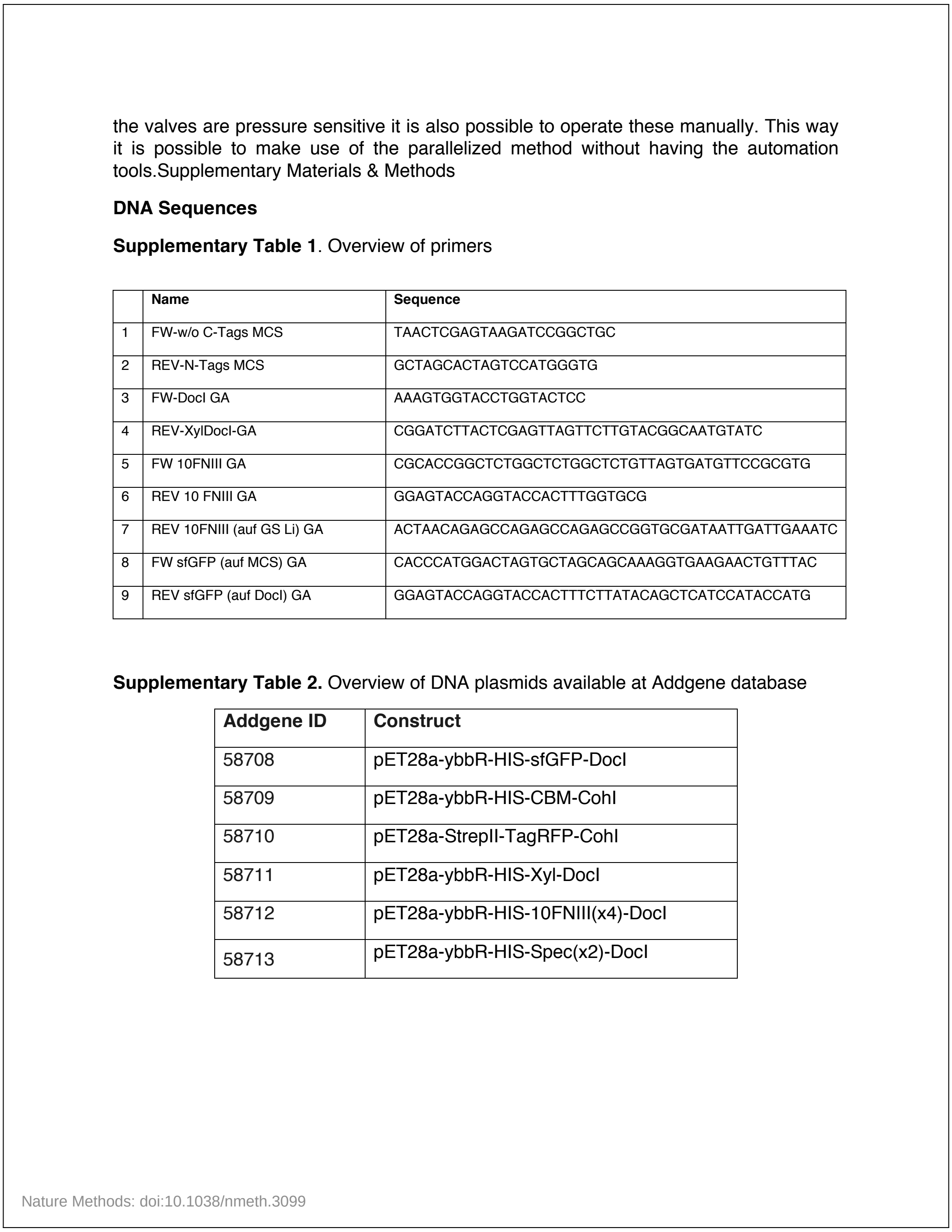
In analyzing single-molecule unfolding curves (i.e., **Fig. 2**), we note that the spotted DNA at the measured array addresses correctly corresponded to the domain of interest encoded by the corresponding spotted DNA at that position. For example, the fibronectin tetramer was measured at array position (237), the spectrin dimer at position (239), the xylanase monomer at position (196), and the sfGFP monomer at position (238), corresponding to the correct genes deposited into the expression chambers at those array positions (**Fig. 2**). Typically 10–15 immobilization chambers per microarray were measured. Typically several thousand force curves were acquired giving rise to dozens of interpretable single-molecule interaction curves.

Upper force limit

Here we extend the discussion regarding the upper force limit for the SMFS-MITOMI system. In all force-distance data traces, the last rupture events represent unbinding of the Coh-Doc complex, not unfolding of a domain. This rupture force of the Coh-Doc complex represents an upper limit in force for the entire construct, since the Doc is used as a handle sequence grabbed by the Coh-modified cantilver. The system we described can therefore interrogate domains with mechanical rupture forces that lie below that of Coh-Doc (~125 pN at 10 nN/s). If proteins with larger unfolding forces should be investigated, other Coh-Doc domains that show even higher complex rupture forces can be used. The Coh-Doc pair from *R. flavefaciens*, for example (PDB 4IU3) exhibits rupture forces over 600 pN at these loading rates (unpublished data). This could alternatively be used as a handle sequence to interrogate mechanically more stable domains of interest.

Computerized image analysis can be used to automate cantilever positioning above the fluorescent rings and subsequent acquisition of unfolding traces at each array address in combination with online force curve analysis to further increase throughput. Additionally, well-characterized reference proteins on the same chip may serve as calibration standards further minimizing uncertainty in absolute force values.

It is possible to operate the MITOMI device in a simplified way without the need for microspotting template DNA and chip alignment. This manual option should encourage the interested community to apply the suggested method to their single molecule force spectroscopy experiments. MITOMI enables the experimenter to prepare up to 16 different constructs in one column with 40 repeats each by flow-loading the DNA. Since



the valves are pressure sensitive it is also possible to operate these manually. This way it is possible to make use of the parallelized method without having the automation tools.Supplementary Materials & Methods

DNA Sequences

Supplementary Table 1. Overview of primers

	Name	**Sequence**
1	FW-w/o C-Tags MCS	TAACTCGAGTAAGATCCGGCTGC
2	REV-N-Tags MCS	GCTAGCACTAGTCCATGGGTG
3	FW-DocI GA	AAAGTGGTACCTGGTACTCC
4	REV-XylDocI-GA	CGGATCTTACTCGAGTTAGTTCTTGTACGGCAATGTATC
5	FW 10FNIII GA	CGCACCGGCTCTGGCTCTGGCTCTGTTAGTGATGTTCCGCGTG
6	REV 10 FNIII GA	GGAGTACCAGGTACCACTTTGGTGCG
7	REV 10FNIII (auf GS Li) GA	ACTAACAGAGCCAGAGCCAGAGCCGGTGCGATAATTGATTGAAATC
8	FW sfGFP (auf MCS) GA	CACCCATGGACTAGTGCTAGCAGCAAAGGTGAAGAACTGTTTAC
9	REV sfGFP (auf DocI) GA	GGAGTACCAGGTACCACTTTCTTATACAGCTCATCCATACCATG

Supplementary Table 2. Overview of DNA plasmids available at Addgene database

Addgene ID	**Construct**
58708	pET28a-ybbR-HIS-sfGFP-DocI
58709	pET28a-ybbR-HIS-CBM-CohI
58710	pET28a-StrepII-TagRFP-CohI
58711	pET28a-ybbR-HIS-Xyl-DocI
58712	pET28a-ybbR-HIS-10FNIII(x4)-DocI
58713	pET28a-ybbR-HIS-Spec(x2)-DocI

Nature Methods: doi:10.1038/nmeth.3099

Multiple cloning site for the protein of interest:

N terminal region

T7 promoter | lac operator | RBS | ATG | ybbr Tag | HRV 3C protease site | HIS Tag (x6)

TAATACGACTCACTATAGG|GGAATTGTGAGCGGATAACAATTCC|CCTGTAGAAATAATTTTGT
TTAACTTTAAG|AAGGA|GATATACAT|ATG|GGTACC|GACTCTCTGGAATTCATCGCTTCTAA
ACTGGCT|CTGGAAGTTCTGTTCCAGGGTCCG|CTGCAG|CACCACCACCACCACCAC|CCATGG
ACTAGTGCTAGC

C terminal region

Dockerin Type I | T7 terminator

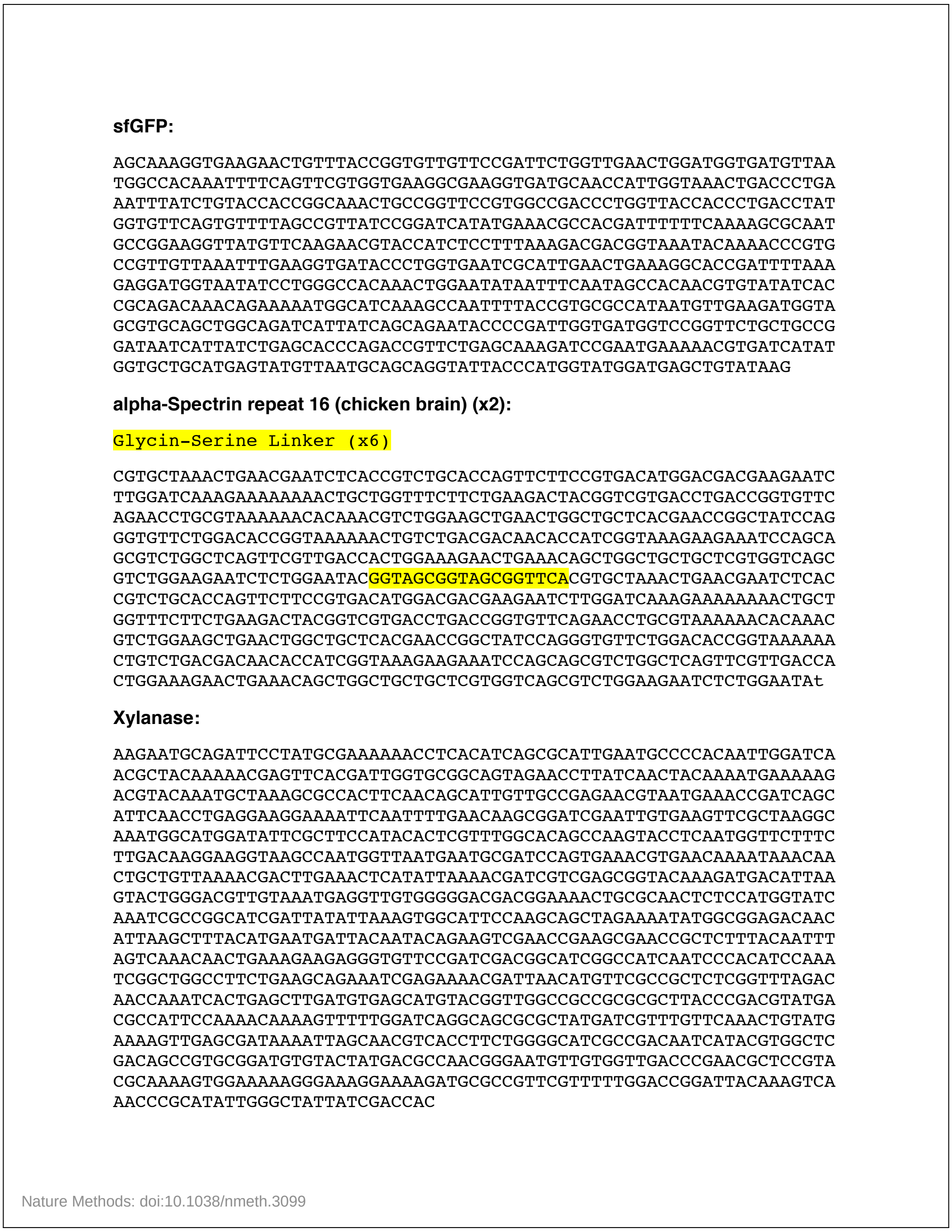
AAAGTGGTACCTGGTACTCCTTCTACTAAATTATACGGCGACGTCAATGATGACGGAAAAGTTAA
CTCAACTGACGCTGTAGCATTGAAGAGATATGTTTTGAGATCAGGTATAAGCATCAACACTGACA
ATGCCGATTTGAATGAAGACGGCAGAGTTAATTCAACTGACTTAGGAATTTTGAAGAGATATATT
CTCAAAGAAATAGATACATTGCCGTACAAGAAC|TAA|CTCGAGTAAGATCCGGCTGCTAACAAA
GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA|CTAGCATAACCCCTTGGGG
CCTCTAAACGGGTCTTGAGGGGTTTTTT

10 FibronectinIII (4x):

Glycin-Serin Linker (x6)

GTTAGTGATGTTCCGCGTGATCTGGAAGTTGTTGCAGCAACCCCGACCAGCCTGCTGATTAGCTG
GGATGCACCGGCAGTTACCGTTCGTTATTATCGTATTACCTATGGTGAAACCGGTGGTAATAGTC
CGGTTCAAGAATTTACCGTTCCGGGTAGCAAAAGCACCGCAACCATTAGCGGTCTGAAACCGGGT
GTTGATTACACCATTACCGTTTATGCCGTTACCGGTCGTGGTGATTCACCGGCAAGCAGCAAACC
GATTAGCATTAACTATCGTACCGGTAGCGGTAGTGGTAGCGTTTCAGATGTGCCTCGCGACCTGG
AAGTGGTGGCTGCCACACCGACCTCACTGCTGATCTCATGGGATGCCCCTGCCGTGACCGTGCGC
TATTATCGCATCACATATGGCGAGACAGGTGGCAATTCACCTGTGCAAGAATTCACAGTTCCTGG
TTCAAAAAGTACCGCCACAATTTCTGGCCTGAAACCTGGCGTGGATTACACAATCACAGTGTATG
CAGTGACAGGTCGCGGTGATAGTCCGGCAAGTTCAAAACCGATTTCAATCAATTATCGCACCGGC
TCTGGCTCTGGCTCTGTTAGTGATGTTCCGCGTGATCTGGAAGTTGTTGCAGCAACCCCGACCAG
CCTGCTGATTAGCTGGGATGCACCGGCAGTTACCGTTCGTTATTATCGTATTACCTATGGTGAAA
CCGGTGGTAATAGTCCGGTTCAAGAATTTACCGTTCCGGGTAGCAAAAGCACCGCAACCATTAGC
GGTCTGAAACCGGGTGTTGATTACACCATTACCGTTTATGCCGTTACCGGTCGTGGTGATTCACC
GGCAAGCAGCAAACCGATTAGCATTAACTATCGTACCGGTAGCGGTAGTGGTAGCGTTTCAGATG
TGCCTCGCGACCTGGAAGTGGTGGCTGCCACACCGACCTCACTGCTGATCTCATGGGATGCCCCT
GCCGTGACCGTGCGCTATTATCGCATCACATATGGCGAGACAGGTGGCAATTCACCTGTGCAAGA
ATTCACAGTTCCTGGTTCAAAAAGTACCGCCACAATTTCTGGCCTGAAACCTGGCGTGGATTACA
CAATCACAGTGTATGCAGTGACAGGTCGCGGTGATAGTCCGGCAAGTTCAAAACCGATTTCAATC
AAttaTCGCACC

Nature Methods: doi:10.1038/nmeth.3099



sfGFP:

AGCAAAGGTGAAGAACTGTTTACCGGTGTTGTTCCGATTCTGGTTGAACTGGATGGTGATGTTAA
TGGCCACAAATTTTCAGTTCGTGGTGAAGGCGAAGGTGATGCAACCATTGGTAAACTGACCCTGA
AATTTATCTGTACCACCGGCAAACTGCCGGTTCCGTGGCCGACCCTGGTTACCACCCTGACCTAT
GGTGTTCAGTGTTTTAGCCGTTATCCGGATCATATGAAACGCCACGATTTTTTCAAAAGCGCAAT
GCCGGAAGGTTATGTTCAAGAACGTACCATCTCCTTTAAAGACGACGGTAAATACAAAACCCGTG
CCGTTGTTAAATTTGAAGGTGATACCCTGGTGAATCGCATTGAACTGAAAGGCACCGATTTTAAA
GAGGATGGTAATATCCTGGGCCACAAACTGGAATATAATTTCAATAGCCACAACGTGTATATCAC
CGCAGACAAACAGAAAAATGGCATCAAAGCCAATTTTACCGTGCGCCATAATGTTGAAGATGGTA
GCGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGTTCTGCTGCCG
GATAATCATTATCTGAGCACCCAGACCGTTCTGAGCAAAGATCCGAATGAAAAACGTGATCATAT
GGTGCTGCATGAGTATGTTAATGCAGCAGGTATTACCCATGGTATGGATGAGCTGTATAAG

alpha-Spectrin repeat 16 (chicken brain) (x2):

Glycin-Serine Linker (x6)

CGTGCTAAACTGAACGAATCTCACCGTCTGCACCAGTTCTTCCGTGACATGGACGACGAAGAATC
TTGGATCAAAGAAAAAAAACTGCTGGTTTCTTCTGAAGACTACGGTCGTGACCTGACCGGTGTTC
AGAACCTGCGTAAAAAACACAAACGTCTGGAAGCTGAACTGGCTGCTCACGAACCGGCTATCCAG
GGTGTTCTGGACACCGGTAAAAAACTGTCTGACGACAACACCATCGGTAAAGAAGAAATCCAGCA
GCGTCTGGCTCAGTTCGTTGACCACTGGAAAGAACTGAAACAGCTGGCTGCTGCTCGTGGTCAGC
GTCTGGAAGAATCTCTGGAATACGGTAGCGGTAGCGGTTCACGTGCTAAACTGAACGAATCTCAC
CGTCTGCACCAGTTCTTCCGTGACATGGACGACGAAGAATCTTGGATCAAAGAAAAAAAACTGCT
GGTTTCTTCTGAAGACTACGGTCGTGACCTGACCGGTGTTCAGAACCTGCGTAAAAAACACAAAC
GTCTGGAAGCTGAACTGGCTGCTCACGAACCGGCTATCCAGGGTGTTCTGGACACCGGTAAAAAA
CTGTCTGACGACAACACCATCGGTAAAGAAGAAATCCAGCAGCGTCTGGCTCAGTTCGTTGACCA
CTGGAAAGAACTGAAACAGCTGGCTGCTGCTCGTGGTCAGCGTCTGGAAGAATCTCTGGAATAt

Xylanase:

AAGAATGCAGATTCCTATGCGAAAAAACCTCACATCAGCGCATTGAATGCCCCACAATTGGATCA
ACGCTACAAAAACGAGTTCACGATTGGTGCGGCAGTAGAACCTTATCAACTACAAAATGAAAAAG
ACGTACAAATGCTAAAGCGCCACTTCAACAGCATTGTTGCCGAGAACGTAATGAAACCGATCAGC
ATTCAACCTGAGGAAGGAAAATTCAATTTTGAACAAGCGGATCGAATTGTGAAGTTCGCTAAGGC
AAATGGCATGGATATTCGCTTCCATACACTCGTTTGGCACAGCCAAGTACCTCAATGGTTCTTTC
TTGACAAGGAAGGTAAGCCAATGGTTAATGAATGCGATCCAGTGAAACGTGAACAAAATAAACAA
CTGCTGTTAAAACGACTTGAAACTCATATTAAAACGATCGTCGAGCGGTACAAAGATGACATTAA
GTACTGGGACGTTGTAAATGAGGTTGTGGGGGACGACGGAAAACTGCGCAACTCTCCATGGTATC
AAATCGCCGGCATCGATTATATTAAAGTGGCATTCCAAGCAGCTAGAAAATATGGCGGAGACAAC
ATTAAGCTTTACATGAATGATTACAATACAGAAGTCGAACCGAAGCGAACCGCTCTTTACAATTT
AGTCAAACAACTGAAAGAAGAGGGTGTTCCGATCGACGGCATCGGCCATCAATCCCACATCCAAA
TCGGCTGGCCTTCTGAAGCAGAAATCGAGAAAACGATTAACATGTTCGCCGCTCTCGGTTTAGAC
AACCAAATCACTGAGCTTGATGTGAGCATGTACGGTTGGCCGCCGCGCGCTTACCCGACGTATGA
CGCCATTCCAAAACAAAAGTTTTTGGATCAGGCAGCGCGCTATGATCGTTTGTTCAAACTGTATG
AAAAGTTGAGCGATAAAATTAGCAACGTCACCTTCTGGGGCATCGCCGACAATCATACGTGGCTC
GACAGCCGTGCGGATGTGTACTATGACGCCAACGGGAATGTTGTGGTTGACCCGAACGCTCCGTA
CGCAAAAGTGGAAAAAGGGAAAGGAAAAGATGCGCCGTTCGTTTTTGGACCGGATTACAAAGTCA
AACCCGCATATTGGGCTATTATCGACCAC

Nature Methods: doi:10.1038/nmeth.3099

Detection construct RFP-Cohesin:

TagRFP-Cohesin:

T7 promoter | *lac operator* | *RBS* | *ATG* | *StrepII Tag* | *TagRFP* |
Linker | *Cohesin* | T7 terminator

```
TAATACGACTCACTATAGG|GGAATTGTGAGCGGATAACAATTCC|CCTGTAGAAATAATTTTGT
TTAACTTTAAG|AAGGA|GATATACAT|ATG|GGTACC|TGGTCTCACCCGCAGTTCGAAAAA|G
TTTCTAAAGGTGAAGAACTGATCAAAGAAAACATGCACATGAAACTGTACATGGAAGGTACTGTT
AACAACCACCACTTCAAATGCACCTCTGAAGGTGAAGGTAAACCGTACGAAGGTACTCAGACCAT
GCGTATCAAAGTTGTTGAAGGTGGTCCGCTGCCGTTCGCTTTCGACATCCTGGCTACCTCTTTCA
TGTACGGTTCTCGTACCTTCATCAACCACACCCAGGGTATCCCGGACTTCTTCAAACAGTCTTTC
CCGGAAGGTTTCACCTGGGAACGTGTTACCACCTACGAAGACGGTGGTGTTCTGACCGCTACCCA
GGACACCTCTCTGCAAGACGGTTGCCTGATCTACAACGTTAAAATCCGTGGTGTTAACTTCCCGT
CTAACGGTCCGGTTATGCAGAAAAAAACCCTGGGTTGGGAAGCTAACACCGAAATGCTGTACCCG
GCTGACGGTGGTCTGGAAGGTCGTTCTGACATGGCTCTGAAACTGGTTGGTGGTGGTCACCTGAT
CTGCAACTTCAAAACCACCTACCGTTCTAAAAAACCGGCTAAAAACCTGAAAATGCCGGGTGTTT
ACTACGTTGACCACCGTCTGGAACGTATCAAAGAAGCTGACAAAGAAACCTACGTTGAACAGCAC
GAAGTTGCTGTTGCTCGTTACTGCGACCTGCCGTCTAAACTGGGTCACAAACTGAAC|GGCAGTG
TAGTACCATCAACACAGCCTGTAACAACACCACCTGCAACAACAAAACCACCTGCAACAACAATA
CCGCCGTCAGATGATCCGAATGCA|GGATCCGACGGTGTGGTAGTAGAAATTGGCAAAGTTACGG
GATCTGTTGGAACTACAGTTGAAATACCTGTATATTTCAGAGGAGTTCCATCCAAAGGAATAGCA
AACTGCGACTTTGTGTTCAGATATGATCCGAATGTATTGGAAATTATAGGGATAGATCCCGGAGA
CATAATAGTTGACCCGAATCCTACCAAGAGCTTTGATACTGCAATATATCCTGACAGAAAGATAA
TAGTATTCCTGTTTGCGGAAGACAGCGGAACAGGAGCGTATGCAATAACTAAAGACGGAGTATTT
GCAAAAATAAGAGCAACTGTAAAATCAAGTGCTCCGGGCTATATTACTTTCGACGAAGTAGGTGG
ATTTGCAGATAATGACCTGGTAGAACAGAAGGTATCATTTATAGACGGTGGTGTTAACGTTGGCA
ATGCAACA|TAA|CTCGAGTAAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTG
CTGCCACCGCTGAGCAATAA|CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTT
TTT
```

Molecular weights of synthesized fusion proteins

ybbR-(Fibronectin)$_4$-Dockerin Type I: 53 kDa
ybbR-(Spectrin)$_2$-Dockerin Type I: 40 kDa
ybbR-Xylanase-Dockerin Type I: 56 kDa
ybbR-sfGFP-Dockerin Type I: 39 kDa
ybbR-Twitchin-Dockerin Type I: 52 kDa

Supplementary Table 3. Yield of interpretable curves

Construct	**Interpretable Curves**
GFP	25 out of 15258 = 0.16 %
Fibronectin	27 out of 26653 = 0.1 %
Xylanase	91 out of 5553 = 1.64 %
Spectrin	50 out of 10344 = 0.48%

References

1. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature* **397,** 50–53 (1999).
2. Rico, F., Gonzalez, L., Casuso, I., Puig-Vidal, M. & Scheuring, S. High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations. *Science* **342,** 741–743 (2013).

5.5 PUBLICATION P6: AFM Single Molecule Force Spectroscopy

This work details the experimental method of site-specific receptor-ligand AFM-based SMFS. Methodological concepts of covalent surface immobilization, SFMS assay, and data analysis in contour length space are explained using the example of cellulosomal molecules. I co-designed the work, and co-authored the manuscript.

Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy

Markus A. Jobst[†], Constantin Schoeler[†],
Klara Malinowska, Michael A. Nash

[†] *these authors contributed equally to this publication*

published in

Journal of Visualized Experiments, (82), (2013)

Reprinted from [120], with permission from MyJoVE Corporation jove Journal of Visualized Experiments www.jove.com

Video Article

Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy

Markus A. Jobst*[1], Constantin Schoeler*[1], Klara Malinowska[1], Michael A. Nash[1]

[1] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität

[*] These authors contributed equally

Correspondence to: Michael A. Nash at michael.nash@physik.uni-muenchen.de

URL: http://www.jove.com/video/50950
DOI: doi:10.3791/50950

Keywords: Bioengineering, Issue 82, biophysics, protein unfolding, atomic force microscopy, surface immobilization

Date Published: 12/20/2013

Citation: Jobst, M.A., Schoeler, C., Malinowska, K., Nash, M.A. Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy. *J. Vis. Exp.* (82), e50950, doi:10.3791/50950 (2013).

Abstract

Cellulosomes are discrete multienzyme complexes used by a subset of anaerobic bacteria and fungi to digest lignocellulosic substrates. Assembly of the enzymes onto the noncatalytic scaffold protein is directed by interactions among a family of related receptor-ligand pairs comprising interacting cohesin and dockerin modules. The extremely strong binding between cohesin and dockerin modules results in dissociation constants in the low picomolar to nanomolar range, which may hamper accurate off-rate measurements with conventional bulk methods. Single-molecule force spectroscopy (SMFS) with the atomic force microscope measures the response of individual biomolecules to force, and in contrast to other single-molecule manipulation methods (*i.e.* optical tweezers), is optimal for studying high-affinity receptor-ligand interactions because of its ability to probe the high-force regime (>120 pN). Here we present our complete protocol for studying cellulosomal protein assemblies at the single-molecule level. Using a protein topology derived from the native cellulosome, we worked with enzyme-dockerin and carbohydrate binding module-cohesin (CBM-cohesin) fusion proteins, each with an accessible free thiol group at an engineered cysteine residue. We present our site-specific surface immobilization protocol, along with our measurement and data analysis procedure for obtaining detailed binding parameters for the high-affinity complex. We demonstrate how to quantify single subdomain unfolding forces, complex rupture forces, kinetic off-rates, and potential widths of the binding well. The successful application of these methods in characterizing the cohesin-dockerin interaction responsible for assembly of multidomain cellulolytic complexes is further described.

Video Link

The video component of this article can be found at http://www.jove.com/video/50950/

Introduction

Cellulosomes are large multienzyme complexes displayed on the surface of anaerobic cellulolytic bacteria (*e.g. C. thermocellum*) that have evolved to efficiently depolymerize plant cell wall lignocellulose into soluble oligosaccharides[1]. A central attribute of cellulosomes is the high-affinity cohesin-dockerin interaction. In the most prominent paradigm, a highly conserved 60-75 amino acid type I dockerin module is displayed at the C-terminal end of the various bacterial enzymes. The dockerin module directs assembly of synergistic combinations of enzymes onto the noncatalytic scaffold protein ('scaffoldin'), which comprises a polyprotein of cohesin domains that are specific for the type I dockerin module. At higher levels, cellulosome architecture can become very complex, incorporating alternative cohesin and dockerin pairs (*e.g.* type II, type III) that anchor the structures to the cell surface and allow for the assembly of branched structures containing multiple scaffoldins[2]. The various cohesin-dockerin types, despite having related structures, exhibit differential binding specificities suppressing cross reactivity with unintended scaffoldins or components from other cellulosome-producing bacterial species. While bioinformatic approaches have successfully identified thousands of unique cellulosomal components at the genetic level, comparatively few protein structures are known, and the mechanisms at work in cohesin-dockerin specificity determination remains an active area of structural biology research.

Since the invention of the atomic force microscope (AFM) by Binnig *et al.*[3], various AFM operational modes have been developed and continuously improved, including noncontact imaging, oscillation mode imaging[4], and single molecule force spectroscopy (SMFS)[5,6]. SMFS has evolved into a widely used technique to directly probe individual proteins[7-11], nucleic acids[12-15], and synthetic polymers[16-19]. In a typical SMFS experiment to investigate receptor-ligand binding[20,21], an AFM cantilever tip is modified with one of the binding partners, while a flat glass surface is modified with the complementary binding partner. The modified cantilever is brought into contact with the surface allowing the partners to bind. The base of the cantilever is then withdrawn at constant speed and the force is measured using the optical lever deflection method. The resultant force-distance data traces exhibit sawtooth-like peaks if binding was established. In cases where the binding partners are fused to multiple protein domains, each peak in the force-distance trace can be correlated to the unfolding of a single protein domain or folded subdomain, while the last peak corresponds to rupture of the protein binding interface. The specific positions of the force-resistant elements can be used as a fingerprint to identify the various protein domains of interest. This method can be used to interrogate important amino acids involved in protein folding and stabilization. Many models have been reported in the literature to treat the characteristic force extension behavior observed in SMFS experiments. The most commonly used models include the freely jointed chain (FJC) model[22], the worm-like chain (WLC) model[18,23-25], and the freely rotating chain (FRC) model[25,26].

In our prior work[11], we used single-molecule force spectroscopy to investigate the interaction of cohesin and dockerin modules. Here, we present an experimental protocol for glass surface and cantilever functionalization with enzyme-dockerin and CBM-cohesin protein constructs. We also present an AFM-based SMFS protocol including data acquisition and analysis procedures. The described protocol can easily be generalized to other molecular systems, and should prove particularly useful to researchers interested in high-affinity receptor ligand pairs.

Protocol

A schematic of the pulling geometry used in this work to probe the cohesin-dockerin interaction is shown in **Figure 1A.** The protein immobilization protocol reported here for cantilever and cover glass functionalization is a modified version of the procedure published previously[27]. The proteins were expressed from plasmid vectors in *E. coli* using conventional methods. The proteins were designed with a solvent-accessible thiol group, which was used in combination with maleimide chemistry to tether the protein via a stable thioether linkage to the cover glass surface and cantilever. The engineered cysteine residues in both the CBM-cohesin and xylanase-dockerin fusion proteins were located towards the N-terminal side of the proteins, away from the cohesin-dockerin binding interface[11]. A detailed overview of the chemical bonding employed in protein immobilization is shown in **Figure 1B.**

1. Sample Preparation

1. **Buffers**
    1. Prepare Tris buffered saline supplemented with calcium (TBS): 25 mM TRIS, 75 mM NaCl, 1 mM $CaCl_2$, pH 7.2
    2. Prepare sodium borate buffer: 50 mM $Na_2B_4O_7$, pH 8.5
    The process flow diagram showing sample preparation steps is shown in **Figure 2.**

    When handling cantilevers and cover glasses, self-locking tweezers are recommended.

2. **Aminosilanization of cover glass** (approximately 1.5 hr)
    1. Place cover glass (24 mm diameter, 0.5 mm thickness) in a PTFE holder.
    2. Sonicate cover glass in 1:1 ethanol : ultrapure water (v/v) for 15 min.
    3. Rinse cover glass with ultrapure water.
    4. Place cover glass in piranha solution (1:1 H_2SO_4 (concentrated) : H_2O_2 (30%) (v/v)) for 30 min, then thoroughly rinse with ultrapure water. Dry cover glass under a gentle stream of N_2. Caution: piranha solution is extremely corrosive. Eye and skin protection are required.
    5. Submerge cover glass in 45:5:1 ethanol : ultrapure water : 3-aminopropyl dimethyl ethoxysilane (v/v). Place on a shaker at RT for 60 min (approximately 50 rpm).
    6. Submerge cover glass sequentially in ethanol and ultrapure water (2x each). Dry cover glass under a gentle stream of N_2.
    7. Bake cover glass in an oven (80 °C for 30 min).
    8. Silanized cover glasses may be stored under argon for up to 6 weeks.

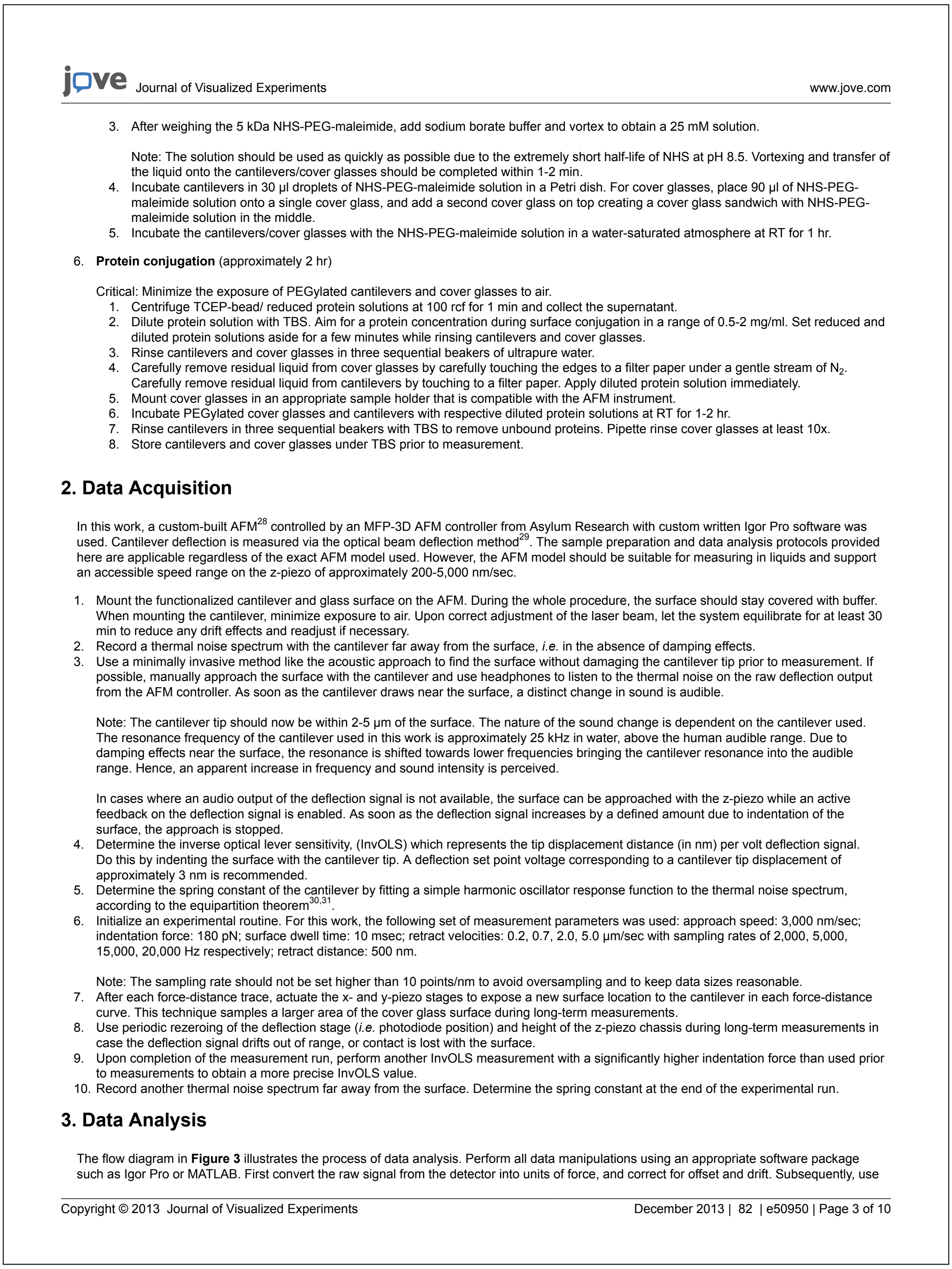
3. **Aminosilanization of cantilevers** (approximately 1.5 hr)

    NOTE: The presented protocol for tip functionalization is appropriate for silicon cantilever tips.
    1. Place cantilevers on a clean glass slide. Treat with UV-ozone for 15 min.
    2. Submerge cantilevers for 3 min in 1:1 ethanol : 3-aminopropyl dimethyl ethoxysilane (v/v) with a catalytic amount (0.25%, (v/v)) of ultrapure water.
    3. Rinse cantilevers with gentle stirring sequentially for 60 sec in beakers of toluene, ethanol, and ultrapure water. Carefully dry cantilevers on filter paper between rinses.
    4. Place levers on a clean glass slide and bake (80 °C for 30 min).

4. **Protein disulfide reduction** (approximately 3 hr)

    All solutions should be prepared to obtain approximately 30 µl of diluted protein per cantilever and 20 µl of diluted protein per cover glass. Protein solutions should be mixed with Tris(2-carboxyethyl)phosphine (TCEP) disulfide reducing gel in a ratio of 1:2 (v/v).
    1. Prepare aliquots of TCEP disulfide reducing beads in micro-tubes. It is recommended to cut micropipette tips with scissors to widen the hole diameter when pipetting the TCEP bead slurry.
    2. Rinse TCEP bead slurry with 1 ml TBS buffer, and centrifuge at 850 rcf for 3 min.
    3. Carefully remove and discard the supernatant with a micropipette.
    4. Repeat steps 1.4.2-1.4.3 2x.
    5. Apply concentrated protein solution (1-10 mg/ml) to the TCEP beads (1:2 protein : TCEP bead slurry (v/v)) and gently mix by stirring with micropipette tip. Avoid introducing air bubbles.
    6. Place protein/TCEP bead slurry mixture on a rotator for 2.5 hr.

5. **PEGylation of cover glasses and cantilevers** (approximately 1.5 hr)
    1. Prior to modification with NHS-PEG-maleimide linkers, soak aminosilanized cantilevers and cover glasses in sodium borate buffer (pH 8.5) for 45 min to deprotonate primary amine groups on the surface.
    2. Ensure that the NHS-PEG-maleimide powder is warmed up to RT before opening the cap and weighing the appropriate amount for a 25 mM solution. Unused NHS-PEG-maleimide should be stored under argon at -20 °C. Approximately 30 µl of polymer solution per cantilever, and 90 µl per 2 cover glasses (sandwiched together) is required.



3. After weighing the 5 kDa NHS-PEG-maleimide, add sodium borate buffer and vortex to obtain a 25 mM solution.

   Note: The solution should be used as quickly as possible due to the extremely short half-life of NHS at pH 8.5. Vortexing and transfer of the liquid onto the cantilevers/cover glasses should be completed within 1-2 min.
4. Incubate cantilevers in 30 µl droplets of NHS-PEG-maleimide solution in a Petri dish. For cover glasses, place 90 µl of NHS-PEG-maleimide solution onto a single cover glass, and add a second cover glass on top creating a cover glass sandwich with NHS-PEG-maleimide solution in the middle.
5. Incubate the cantilevers/cover glasses with the NHS-PEG-maleimide solution in a water-saturated atmosphere at RT for 1 hr.

6. **Protein conjugation** (approximately 2 hr)

   Critical: Minimize the exposure of PEGylated cantilevers and cover glasses to air.
   1. Centrifuge TCEP-bead/ reduced protein solutions at 100 rcf for 1 min and collect the supernatant.
   2. Dilute protein solution with TBS. Aim for a protein concentration during surface conjugation in a range of 0.5-2 mg/ml. Set reduced and diluted protein solutions aside for a few minutes while rinsing cantilevers and cover glasses.
   3. Rinse cantilevers and cover glasses in three sequential beakers of ultrapure water.
   4. Carefully remove residual liquid from cover glasses by carefully touching the edges to a filter paper under a gentle stream of N_2. Carefully remove residual liquid from cantilevers by touching to a filter paper. Apply diluted protein solution immediately.
   5. Mount cover glasses in an appropriate sample holder that is compatible with the AFM instrument.
   6. Incubate PEGylated cover glasses and cantilevers with respective diluted protein solutions at RT for 1-2 hr.
   7. Rinse cantilevers in three sequential beakers with TBS to remove unbound proteins. Pipette rinse cover glasses at least 10x.
   8. Store cantilevers and cover glasses under TBS prior to measurement.

2. Data Acquisition

In this work, a custom-built AFM[28] controlled by an MFP-3D AFM controller from Asylum Research with custom written Igor Pro software was used. Cantilever deflection is measured via the optical beam deflection method[29]. The sample preparation and data analysis protocols provided here are applicable regardless of the exact AFM model used. However, the AFM model should be suitable for measuring in liquids and support an accessible speed range on the z-piezo of approximately 200-5,000 nm/sec.

1. Mount the functionalized cantilever and glass surface on the AFM. During the whole procedure, the surface should stay covered with buffer. When mounting the cantilever, minimize exposure to air. Upon correct adjustment of the laser beam, let the system equilibrate for at least 30 min to reduce any drift effects and readjust if necessary.
2. Record a thermal noise spectrum with the cantilever far away from the surface, *i.e.* in the absence of damping effects.
3. Use a minimally invasive method like the acoustic approach to find the surface without damaging the cantilever tip prior to measurement. If possible, manually approach the surface with the cantilever and use headphones to listen to the thermal noise on the raw deflection output from the AFM controller. As soon as the cantilever draws near the surface, a distinct change in sound is audible.

   Note: The cantilever tip should now be within 2-5 µm of the surface. The nature of the sound change is dependent on the cantilever used. The resonance frequency of the cantilever used in this work is approximately 25 kHz in water, above the human audible range. Due to damping effects near the surface, the resonance is shifted towards lower frequencies bringing the cantilever resonance into the audible range. Hence, an apparent increase in frequency and sound intensity is perceived.

   In cases where an audio output of the deflection signal is not available, the surface can be approached with the z-piezo while an active feedback on the deflection signal is enabled. As soon as the deflection signal increases by a defined amount due to indentation of the surface, the approach is stopped.
4. Determine the inverse optical lever sensitivity, (InvOLS) which represents the tip displacement distance (in nm) per volt deflection signal. Do this by indenting the surface with the cantilever tip. A deflection set point voltage corresponding to a cantilever tip displacement of approximately 3 nm is recommended.
5. Determine the spring constant of the cantilever by fitting a simple harmonic oscillator response function to the thermal noise spectrum, according to the equipartition theorem[30,31].
6. Initialize an experimental routine. For this work, the following set of measurement parameters was used: approach speed: 3,000 nm/sec; indentation force: 180 pN; surface dwell time: 10 msec; retract velocities: 0.2, 0.7, 2.0, 5.0 µm/sec with sampling rates of 2,000, 5,000, 15,000, 20,000 Hz respectively; retract distance: 500 nm.

   Note: The sampling rate should not be set higher than 10 points/nm to avoid oversampling and to keep data sizes reasonable.
7. After each force-distance trace, actuate the x- and y-piezo stages to expose a new surface location to the cantilever in each force-distance curve. This technique samples a larger area of the cover glass surface during long-term measurements.
8. Use periodic rezeroing of the deflection stage (*i.e.* photodiode position) and height of the z-piezo chassis during long-term measurements in case the deflection signal drifts out of range, or contact is lost with the surface.
9. Upon completion of the measurement run, perform another InvOLS measurement with a significantly higher indentation force than used prior to measurements to obtain a more precise InvOLS value.
10. Record another thermal noise spectrum far away from the surface. Determine the spring constant at the end of the experimental run.

3. Data Analysis

The flow diagram in **Figure 3** illustrates the process of data analysis. Perform all data manipulations using an appropriate software package such as Igor Pro or MATLAB. First convert the raw signal from the detector into units of force, and correct for offset and drift. Subsequently, use models of biopolymer elasticity to locate energy barriers in the unfolding pathways, and identify protein subdomains. Finally, kinetic and energetic parameters of the receptor-ligand interaction are obtained.

1. **Unit conversion and data corrections**
    1. Multiply the raw deflection signal (volts) by the InvOLS (nm/volt) and spring constant (pN/nm) to convert the detector voltage into units of force.
    2. Offset the data such that the unloaded cantilever has a force value of zero pN by first averaging the force values from the last 10% of the force-distance trace (acquired farthest away from the surface), and then subtracting the average from all force values in the data trace.
    3. Offset each trace in the x-direction such that the first intercept with the distance axis occurs at a distance of 0 nm.
    4. The InvOLS is dependent on the laser spot position on the cantilever. Even small amounts of drift in the optical readout system may cause noticeable changes in the InvOLS when the footprint of the cantilever is comparable to the laser spot size. Correct for this by analysis of the noise on the deflection signal at zero force. Assuming constant ambient conditions, noise on the deflection signal is directly proportional to the InvOLS.
        1. Measure the route mean square (RMS) deflection value (noise level under zero force) of the last 10% of each force-distance trace.
        2. Plot the noise vs. the curve number and apply a suitable fit. Typically an exponential fit in the form of $N(n_i) = N_0 \cdot e^{-kn_i}$ will work best, where N is the noise, ni is the curve number, and N0 and k are fit parameters. A linear fit may also be appropriate for certain data sets.
        3. Determine a scaling factor (SF) for each curve:

           Equation 1:

           $$(SF) = \frac{N_0 e^{-kn_i} + C}{N_0 e^{-kn_F} + C}$$

           where, ni is the curve number, nf is the final curve number, and C is an offset.
        4. Next divide all the force values in each individual curve by the scaling factor. This procedure scales each curve by the ratio of the RMS noise value of the current curve to the RMS noise value of the final trace that was acquired immediately prior to the InvOLS measurement.
        5. Perform a deflection correction to transform the distance axis (z) to molecular extension (z*). This accounts for bending of the cantilever under force which shortens the distance between cantilever tip and sample from the value reported by the z-piezo sensor position.

           Equation 2:

           $$z^* = z - \frac{F}{k}$$

           Where *z* is the measured z-sensor position, *F* the force acting on the cantilever and *k* the spring constant.

2. **Contour length analysis**

    The contour length of a protein is the maximum stretched length of the polypeptide chain. The folding state of a protein refers to its geometry and end to end distance determined by secondary and tertiary structure. The contour length of a protein is directly related to its folding state[9,25,32]. The position of specific ruptures in force-extension traces varies widely due to polydispersity of PEG linkers, as well as external parameters such as temperature, buffer properties and loading rates. This complicates direct data analysis but can be overcome by transforming force-extension data into contour length space. This technique enables averaging over huge datasets, and allows automatic pattern recognition to be used to identify characteristic unfolding events. It is therefore possible to sort individual force traces depending on the type of interaction exhibited. The following previously described procedure[25] is used to transform force-extension data into contour length space.
    1. Solving the WLC model (Equation 3)[23] for the contour length *L* at a fixed persistence length *p* results in Equation 4, which provides the contour length *L(x,u)*. Here, *x* is the distance and *u=F*p/kBT*, where *kB* is Boltzmann's constant and T is the temperature. Only real solutions can be considered. Additional constraints are *x<L*, *F>0, L>0, x>0*;

       Equation 3:

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

Equation 4:

$$L(x,u) = \frac{x}{6u}\left(3 + 4u + \frac{9 - 3u + 4u^2}{g(u)} + g(u)\right)$$

where,

$$g(u) = \left(27 - \frac{27}{2}u + 36u^2 - 8u^3 + \frac{3\sqrt{3}}{2}\sqrt{-u^2\left[(4u-3)^3 - 108\right]}\right)^{\frac{1}{3}}$$

2. Plot the transformed data points in a force vs. contour length plot. Apply a force threshold of approximately 10 pN to exclude noise. Unspecific interactions can be excluded by applying a long-pass length filter. Assemble a histogram of contour lengths.
3. Cross-correlate[33] the obtained histograms with a template histogram, and offset along the x-axis to correct for PEG polydispersity. Use the resulting correlation values to measure the similarity of individual data traces. Thereby, data traces can be sorted into predefined classes to simplify further analysis.
4. Use a similar technique to find repeating features in a single trace by autocorrelation, *e.g.* for multiple Ig-domain unfolding.
5. Sort traces manually to investigate other unfolding events.

3. **Loading Rate Analysis**

   Extract kinetic and energetic information about receptor-ligand dissociation by applying suitable models to the force spectrum, *i.e.* the rupture-force vs. ln(loading-rate) plot.

   1. For a given pulling speed, determine the rupture force and loading rate for rupture events of interest:
      1. Perform a line fit to a force-time trace in the vicinity of the rupture event of interest. Determine the loading rate from the slope of the line fit to the peak. Repeat this procedure for every trace showing the rupture event of interest.
      2. Determine the most probable rupture force by applying a Gaussian fit to a histogram of the rupture forces. Alternative fit functions are possible.
      3. Determine the most probable loading rate.

   2. Repeat steps 3.3.3.1 - 3.3.3.3 for all pulling speeds.
   3. Plot the most probable rupture forces against the natural logarithm of the most probable loading rates to obtain the force spectrum.
   4. Apply a suitable theoretical model to the force spectrum to extract kinetic and energetic parameters (**Figure 4C**). In many cases, the linear Bell-Evans model[20,34] can be used and will yield good estimates for *koff*, the dissociation rate in the absence of force, and Dx*,* the distance to the transition state along the reaction coordinate, as shown in Equation 5.

Equation 5:

$$\langle F \rangle = \frac{k_B T}{\Delta x}\ln\left(\frac{\dot{F}\cdot \Delta x}{k_B T\cdot k_{off}}\right)$$

Representative Results

We used the described procedure to investigate a type I cohesin-dockerin pair from *C. thermocellum*. Upon successful binding of the cohesin-dockerin pair, the recorded force distance traces showed characteristic peak patterns. A typical trace is shown in **Figure 4a.** Every peak in the trace represents the unfolding of one protein subdomain with the last peak corresponding to the dissociation of the receptor-ligand complex.

For the CBM-cohesin-dockerin-xylanase complex investigated in this work, the initial rise in force corresponds to stretching of the PEG linker molecules. The subsequent series of up to three descending force dips reflects the unfolding of the xylanase domain. The final peak represents the rupture of the cohesin-dockerin binding interface.

All recorded force-distance traces were transformed to force-contour length space. The resulting barrier position histogram is shown in **Figure 4B.** The data show a contour length increment of approximately 89 nm. The xylanase domain consists of 378 amino acids, 260 of which are located C-terminally from the engineered cysteine residue. From the crystal structure, the folded length of the domain is assumed to be 6 nm.

Further assuming a length per stretched amino acid of 0.365 nm[35], the measured 89 nm increment can be unambiguously assigned to the unfolding of the xylanase domain. This is consistent with previously published results[11].

To probe the energy landscape of the cohesin-dockerin interaction, we analyzed a total of 186 data traces obtained with four different pulling speeds (0.2, 0.7, 2.0, and 5.0 µm/sec). The resulting force spectrum is shown in **Figure 4C.** Fitting Equation (5) to the data yields values for *koff* and Dx of 3.13 x10^{-5}/sec and 0.70 nm, respectively. These values are in good agreement with previously published results[11].

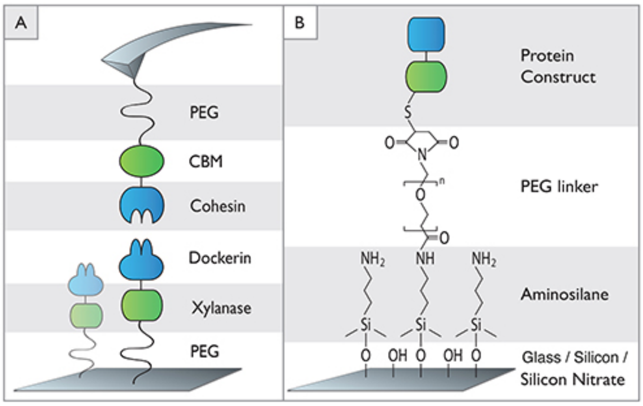


Figure 1. Schematic of biomolecule immobilization. (**A**) Xylanase-dockerin fusion proteins are attached to the glass slide via PEG linkers. The cantilever is similarly modified with a cohesin protein fused to a cellulose binding module (CBM). (**B**) Depiction of chemical bonding employed in cover glass and cantilever functionalization. Click here to view larger image.

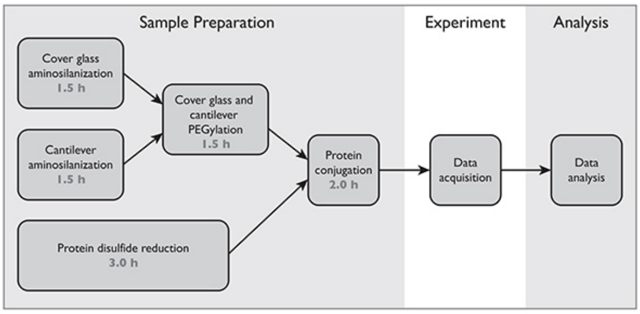


Figure 2. Process flow diagram showing sample preparation steps followed by data acquisition and analysis. Click here to view larger image.

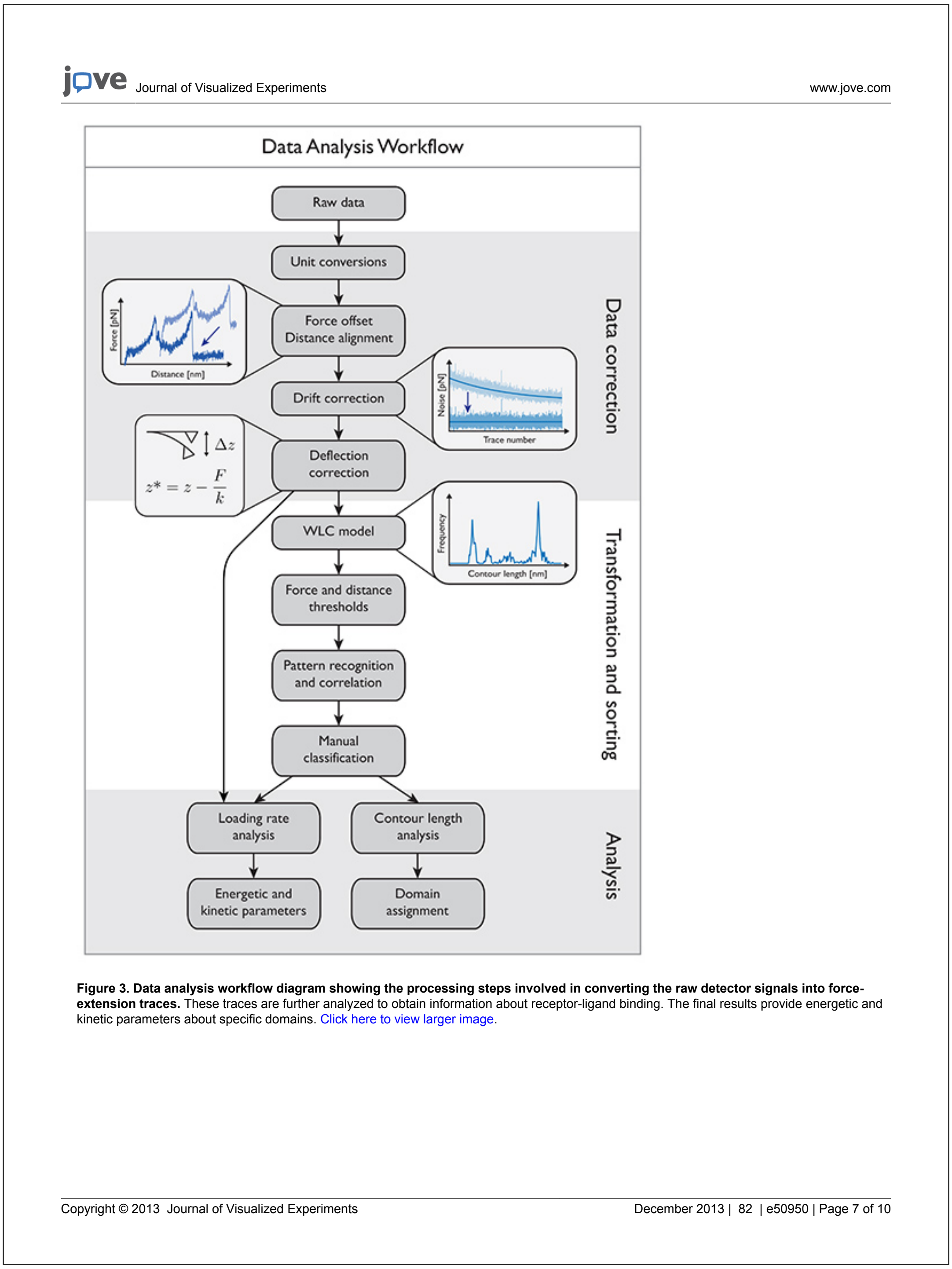


Figure 3. Data analysis workflow diagram showing the processing steps involved in converting the raw detector signals into force-extension traces. These traces are further analyzed to obtain information about receptor-ligand binding. The final results provide energetic and kinetic parameters about specific domains. Click here to view larger image.

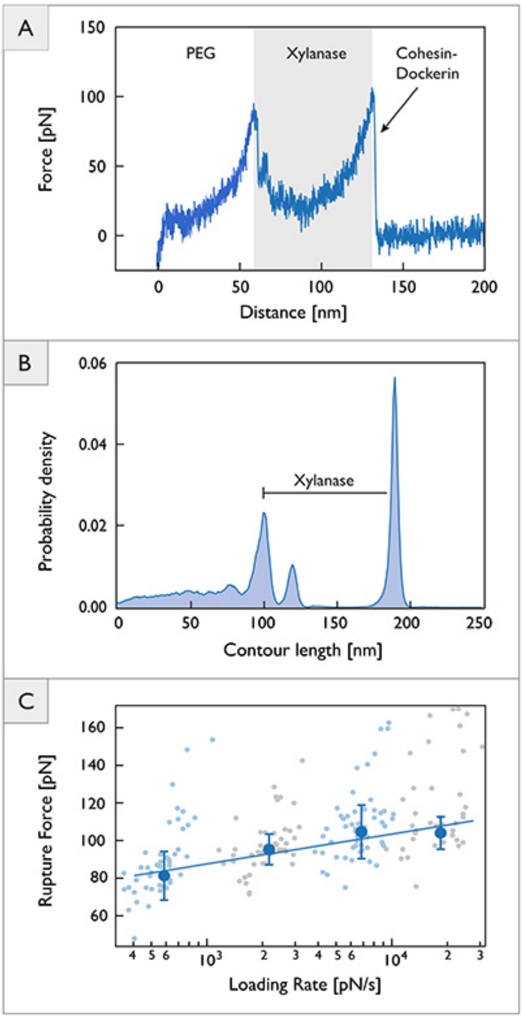


Figure 4. Single molecule force spectroscopy data on cohesin-dockerin. (**A**) Typical unfolding trace showing PEG linker stretching, xylanase unfolding, and rupture of the cohesin-dockerin binding interface. (**B**) Contour length histogram assembled from 314 force distance traces exhibiting energy barrier positions along the contour length. (**C**) Dynamic force spectrum obtained from 186 force-extension traces. Large blue circles represent the most probable rupture force at a given loading rate. The solid line represents a least squares fit to Equation 5. Rupture event populations are shown in the background. Error bars represent standard deviation obtained from Gaussian fits. Click here to view larger image.

Discussion

To obtain meaningful data from single molecule force spectroscopy experiments, it is crucial to achieve well-defined and reproducible pulling geometries. The protocol used here results in site-specific immobilization of protein complexes in a defined pulling geometry.

The cantilevers used in this study were chosen due to their force sensitivity and high resonance frequency in water. Moreover, the small tip curvature of approximately 10 nm is advantageous for single molecule experiments due to reduced likelihood of multiple interactions. However, the small footprint (38x16 μm^2) of the cantilever arm complicates the adjustment of the laser beam when the optical deflection method[29] is used. The diameter of the focused laser beam in the setup used for this study is comparable to the width of the cantilever. As a result, obtaining a steady sum signal can be difficult. The laser drift on the cantilever can be partially compensated for using noise analysis across the data curves to correct the inverse optical lever sensitivity, as we have described. A new atomic force microscope with a shortened optical path and smaller laser spot size is currently under development in our group to improve data quality.

To obtain reliable information about rupture events, analysis of many traces is necessary. Since single molecule force spectroscopy measurements are subject to various fluctuations, averaging in force-extension space is not constructive. Barrier position histograms, however, once aligned in contour length space can be averaged since they are independent of fluctuations. As a result, even tiny features in the unfolding pathway are resolved. Conventional superposition of force extension traces does not achieve this kind of resolution.

In a force regime above 500 pN, a corrected WLC model accounting for electron cloud elasticity (QM-WLC) describes force-extension behavior better than the classical WLC model[18]. The freely rotating chain[26] model (FRC) can also be used in a high force regime. With rupture forces up to 125 pN, the cohesin-dockerin interface shows one of the strongest receptor-ligand interactions reported in the literature. The WLC model was used in this work and in practice there was little difference between WLC, QM-WLC, and FRC models for analysis of cohesin-dockerin unfolding traces.

The conventional Bell-Evans[20,34] model was used to analyze the force-loading rate dependency of the cohesin-dockerin binding interface. Recent works[36,37] have extended the theoretical framework for the interpretation of single molecule experiments. These models treat nonlinear trends in the force spectra. Furthermore, they produce the free energy barrier height DG of the dissociation event. To observe distinct nonlinear trends in the force spectra, loading rates need to be varied over many orders of magnitude. Realizing extremely low loading rates is theoretically achievable using extremely slow z-piezo pulling speeds, however in practice this poses a challenge due to drift in the tip-substrate distance. Extremely high loading rates can also be difficult to obtain since increasing noise might obscure certain features in the recorded force-distance traces. Choice of the theoretical model must be balanced with these practical aspects of data acquisition while considering the specific proteins under investigation. In many cases the linear Bell-Evans model is entirely sufficient.

In conclusion, a complete experimental protocol for the study of receptor-ligand interactions using AFM-based single-molecule force spectroscopy has been presented. The positioning accuracy and force sensitivity of the atomic force microscope in conjunction with versatile biomolecule immobilization strategies provide an excellent toolbox for the investigation of receptor-ligand systems for structural biology studies.

Disclosures

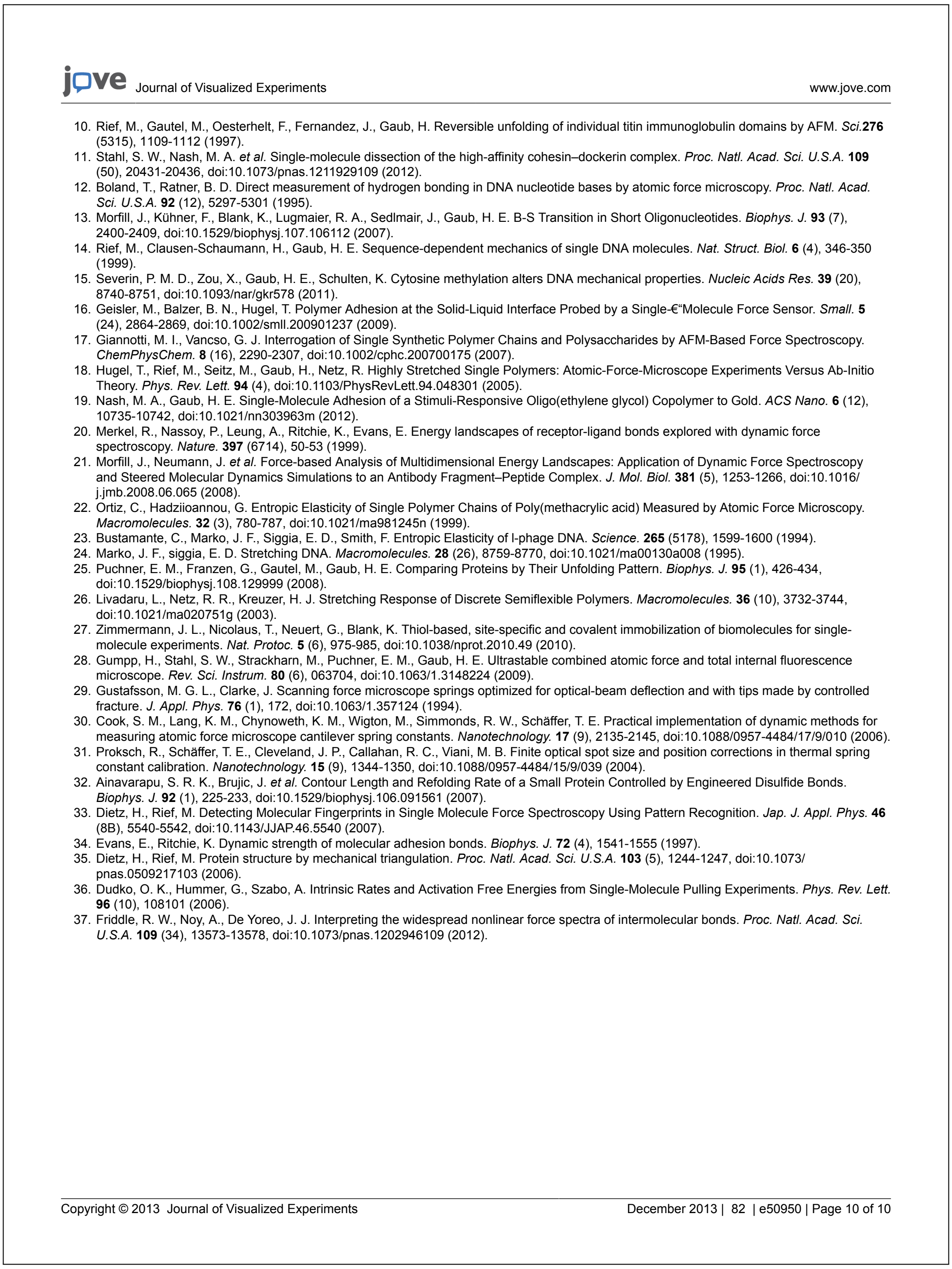
The authors declare that they have no competing financial interests.

Acknowledgements

The authors acknowledge funding from a European Research Council advanced grant to Hermann Gaub. Michael A. Nash gratefully acknowledges funding from Society in Science - The Branco Weiss Fellowship program. The authors thank Edward A. Bayer, Yoav Barak, and Daniel B. Fried at the Weizmann Institute of Science for generously providing the proteins used in this study. The authors thank Hermann E. Gaub, Elias M. Puchner, and Stefan W. Stahl for helpful discussions.

References

1. Bayer, E. A., Belaich, J. P., Shoham, Y., Lamed, R. The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* **58,** 521-554, doi:10.1146/annurev.micro.57.030502.091022 (2004).
2. Bayer, E. A., Lamed, R., White, B. A., Flint, H. J. From Cellulosomes to Cellulosomics. *Chem. Rec.* **8** (6), 364-377, doi:10.1002/tcr.20160 (2008).
3. Binnig, G., Quate, C. F. Atomic Force Microscope. *Phys. Rev. Lett.* **56** (9), 930-933, doi:10.1103/PhysRevLett.56.930 (1986).
4. García, R., Perez, R. Dynamic atomic force microscopy methods. *Surf. Sci. Rep.* **47** (6), 197-301 (2002).
5. Engel, A., Müller, D. J. Observing single biomolecules at work with the atomic force microscope. *Nat. Struct. Biol.* **7** (9), 715-718 (2000).
6. Li, H., Cao, Y. Protein Mechanics: From Single Molecules to Functional Biomaterials. *Acc. Chem. Res.* **43** (10), 1331-1341, doi:10.1021/ar100057a (2010).
7. Florin, E.-L., Moy, V. T., Gaub, H. E. Adhesion forces between individual ligand-receptor pairs. *Science.* **264** (5157), 415-417, doi:10.1126/science.8153628 (1994).
8. Oberhauser, A., Hansma, P., Carrion-Vazquez, M., Fernandez, J. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **98** (2), 468-472 (2001).
9. Puchner, E. M., Gaub, H. E. Force and function: probing proteins with AFM-based force spectroscopy. *Curr. Opin. Struct. Biol.* **19** (5), 605-614, doi:10.1016/j.sbi.2009.09.005 (2009).



10. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J., Gaub, H. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Sci.***276** (5315), 1109-1112 (1997).
11. Stahl, S. W., Nash, M. A. *et al.* Single-molecule dissection of the high-affinity cohesin–dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.* **109** (50), 20431-20436, doi:10.1073/pnas.1211929109 (2012).
12. Boland, T., Ratner, B. D. Direct measurement of hydrogen bonding in DNA nucleotide bases by atomic force microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **92** (12), 5297-5301 (1995).
13. Morfill, J., Kühner, F., Blank, K., Lugmaier, R. A., Sedlmair, J., Gaub, H. E. B-S Transition in Short Oligonucleotides. *Biophys. J.* **93** (7), 2400-2409, doi:10.1529/biophysj.107.106112 (2007).
14. Rief, M., Clausen-Schaumann, H., Gaub, H. E. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* **6** (4), 346-350 (1999).
15. Severin, P. M. D., Zou, X., Gaub, H. E., Schulten, K. Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.* **39** (20), 8740-8751, doi:10.1093/nar/gkr578 (2011).
16. Geisler, M., Balzer, B. N., Hugel, T. Polymer Adhesion at the Solid-Liquid Interface Probed by a Single-€“Molecule Force Sensor. *Small.* **5** (24), 2864-2869, doi:10.1002/smll.200901237 (2009).
17. Giannotti, M. I., Vancso, G. J. Interrogation of Single Synthetic Polymer Chains and Polysaccharides by AFM-Based Force Spectroscopy. *ChemPhysChem.* **8** (16), 2290-2307, doi:10.1002/cphc.200700175 (2007).
18. Hugel, T., Rief, M., Seitz, M., Gaub, H., Netz, R. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-Initio Theory. *Phys. Rev. Lett.* **94** (4), doi:10.1103/PhysRevLett.94.048301 (2005).
19. Nash, M. A., Gaub, H. E. Single-Molecule Adhesion of a Stimuli-Responsive Oligo(ethylene glycol) Copolymer to Gold. *ACS Nano.* **6** (12), 10735-10742, doi:10.1021/nn303963m (2012).
20. Merkel, R., Nassoy, P., Leung, A., Ritchie, K., Evans, E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature.* **397** (6714), 50-53 (1999).
21. Morfill, J., Neumann, J. *et al.* Force-based Analysis of Multidimensional Energy Landscapes: Application of Dynamic Force Spectroscopy and Steered Molecular Dynamics Simulations to an Antibody Fragment–Peptide Complex. *J. Mol. Biol.* **381** (5), 1253-1266, doi:10.1016/j.jmb.2008.06.065 (2008).
22. Ortiz, C., Hadziioannou, G. Entropic Elasticity of Single Polymer Chains of Poly(methacrylic acid) Measured by Atomic Force Microscopy. *Macromolecules.* **32** (3), 780-787, doi:10.1021/ma981245n (1999).
23. Bustamante, C., Marko, J. F., Siggia, E. D., Smith, F. Entropic Elasticity of l-phage DNA. *Science.* **265** (5178), 1599-1600 (1994).
24. Marko, J. F., siggia, E. D. Stretching DNA. *Macromolecules.* **28** (26), 8759-8770, doi:10.1021/ma00130a008 (1995).
25. Puchner, E. M., Franzen, G., Gautel, M., Gaub, H. E. Comparing Proteins by Their Unfolding Pattern. *Biophys. J.* **95** (1), 426-434, doi:10.1529/biophysj.108.129999 (2008).
26. Livadaru, L., Netz, R. R., Kreuzer, H. J. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules.* **36** (10), 3732-3744, doi:10.1021/ma020751g (2003).
27. Zimmermann, J. L., Nicolaus, T., Neuert, G., Blank, K. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat. Protoc.* **5** (6), 975-985, doi:10.1038/nprot.2010.49 (2010).
28. Gumpp, H., Stahl, S. W., Strackharn, M., Puchner, E. M., Gaub, H. E. Ultrastable combined atomic force and total internal fluorescence microscope. *Rev. Sci. Instrum.* **80** (6), 063704, doi:10.1063/1.3148224 (2009).
29. Gustafsson, M. G. L., Clarke, J. Scanning force microscope springs optimized for optical-beam deflection and with tips made by controlled fracture. *J. Appl. Phys.* **76** (1), 172, doi:10.1063/1.357124 (1994).
30. Cook, S. M., Lang, K. M., Chynoweth, K. M., Wigton, M., Simmonds, R. W., Schäffer, T. E. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotechnology.* **17** (9), 2135-2145, doi:10.1088/0957-4484/17/9/010 (2006).
31. Proksch, R., Schäffer, T. E., Cleveland, J. P., Callahan, R. C., Viani, M. B. Finite optical spot size and position corrections in thermal spring constant calibration. *Nanotechnology.* **15** (9), 1344-1350, doi:10.1088/0957-4484/15/9/039 (2004).
32. Ainavarapu, S. R. K., Brujic, J. *et al.* Contour Length and Refolding Rate of a Small Protein Controlled by Engineered Disulfide Bonds. *Biophys. J.* **92** (1), 225-233, doi:10.1529/biophysj.106.091561 (2007).
33. Dietz, H., Rief, M. Detecting Molecular Fingerprints in Single Molecule Force Spectroscopy Using Pattern Recognition. *Jap. J. Appl. Phys.* **46** (8B), 5540-5542, doi:10.1143/JJAP.46.5540 (2007).
34. Evans, E., Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72** (4), 1541-1555 (1997).
35. Dietz, H., Rief, M. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. U.S.A.* **103** (5), 1244-1247, doi:10.1073/pnas.0509217103 (2006).
36. Dudko, O. K., Hummer, G., Szabo, A. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Phys. Rev. Lett.* **96** (10), 108101 (2006).
37. Friddle, R. W., Noy, A., De Yoreo, J. J. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci. U.S.A.* **109** (34), 13573-13578, doi:10.1073/pnas.1202946109 (2012).

6
Manuscripts for Research Articles

6.1 Manuscript M1 (submitted for publication): SMFS on DNA Origami

The goal of the research presented in this manuscript was to validate the oxDNA2 framework for simulating in SMFS data and to elucidate mechanisms of forced unraveling of large DNA Origami structures. Coarse-grained simulations of rothemund sheet DNA Origami were compared to AFM-based SMFS experiments and the Origami unfolding behaviour was assessed in detail. My contribution to this work was the development of a suited surface immobilization strategy, the sample preparation and the subsequent acquisition and analysis of the experimental SMFS data.

Force-induced Unravelling of DNA Origami

Megan C. Engel, David M. Smith, Markus A. Jobst, Martin Sajfutdinow,
Tim Liedl, Flavio Romano, Lorenzo Rovigatti,
Ard A. Louis, and Jonathan P. K. Doye

Reproduced with permission from ACS Nano, submitted for publication

Force-induced unravelling of DNA Origami

Megan C. Engel,[*,†] David M. Smith,[‡] Markus A. Jobst,[¶] Martin Sajfutdinow,[‡]
Tim Liedl,[¶] Flavio Romano,[§] Lorenzo Rovigatti,[†,‖,⊥] Ard A. Louis,[†] and
Jonathan P. K. Doye[#]

†*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, United Kingdom*
‡*Fraunhofer Institute for Cell Therapy and Immunology IZI, Perlickstraße 1, 04103 Leipzig*
¶*Department für Physik, Ludwig-Maximilians-Universität Amalienstr. 54 80799 München*
§*Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia, Via Torino 155, 30172 Venezia Mestre, Italy*
‖*CNR-ISC, Uos Sapienza, Piazzale A. Moro 2, 00185 Roma, Italy*
⊥*Dipartimento di Fisica,* Sapienza *Università di Roma, Piazzale A. Moro 2, 00185 Roma, Italy*
#*Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, United Kingdom*

E-mail: megan.engel@physics.ox.ac.uk

Keywords

DNA nanotechnology, DNA origami, self-assembly, molecular dynamics, coarse-grained modelling, force spectroscopy, AFM

Abstract

The mechanical properties of DNA nanostructures are of widespread interest as applications that exploit their stability under constant or intermittent external forces become increasingly common. We explore the force response of DNA origami in unprecedented detail by combining AFM pulling experiments with simulations using oxDNA, a coarse-grained model of DNA at the nucleotide level, to study the unravelling of an iconic origami system: the Rothemund tile. We contrast the force-induced melting of the tile with simulations of an origami 10-helix bundle. Finally, we simulate a recently-proposed origami biosensor, whose functioning hinges on origami behaviour under tension. We observe characteristic stick-slip unfolding dynamics in our force-extension curves and unravelling from the scaffold ends inwards for both the Rothemund tile and the helix bundle. Our results highlight the effect of design on force response: we observe regular, modular unfolding for the Rothemund tile that contrasts with strain-softening of the 10-helix bundle that leads to catastropic failure under constant force conditions. The detailed visualization of the yielding events provided by simulation allows the preferred pathways through the complex unfolding free-energy landscape to be mapped, a key factor in determining relative barrier heights being the extensional release per base pair broken. We thus shed light on two important questions: how stable DNA nanostructures are under external forces; and what design principles can be applied to enhance stability.

The behavior of DNA under tension has been studied extensively both theoretically and experimentally. Single-molecule force studies have been employed to elucidate the elastic properties of double- and single- stranded DNA;[1–3] to characterize the mechanisms of duplex instability at higher forces;[4–9] and to examine torsional buckling of the helix under a mixture of torque and tension.[10,11] These explorations have shed light on the cellular processes that mediate structural changes in DNA, including gene regulation, transcription, and recombination.[12–14] Furthermore, the now well-understood behavior of duplex DNA under force has been exploited for a wide range of applications, including employing DNA in molecular force sensors[15] to characterize mechanosensitive receptors[16] and cellular traction forces.[17]

Denaturing of the DNA helix is typically modelled as a thermally-activated process, in which the dissociation rate is governed by the height of the free energy barrier between the fully-zippered state – stabilized by hydrogen bonding and base stacking – and a transition state characterized by a critical number of broken base pairs.[18,19] Application of an external biasing force favours more extended configurations, lowering the effective barrier to the transition state until it becomes surmountable by thermal energy.[20]

The force at which duplex DNA yields under tension depends on duplex length,[19,21] force application geometry, whether static or varying tension is applied and for how long,[18–20] and solution conditions including temperature[19,22] and salt concentration.[23,24] Under shear stress, when tension is applied at either end of one strand of the duplex, DNA duplexes typically unbind between 30-65 pN,[8,21] while in an unzipping geometry, where tension is applied to both strands at the same end of the duplex, yielding occurs between 10-20 pN.[18,25,26] The marked difference between the critical forces in shearing and unzipping geometries can be understood by considering the extension gained per base pair broken in each case. Since a force F biases the unfolding landscape by an amount $\sim F\Delta z$, with z the extension of the DNA along the force direction, the breaking of a single base pair in a shearing geometry, which allows an extension gain of $\Delta z/\mathrm{bp} \sim 0.3\,\mathrm{nm}$, will be less favourable than base pair breakage in an unzipping geometry, which allows for an extension gain of $\Delta z/\mathrm{bp} \sim 1.3\,\mathrm{nm}$.[19]

The drive to produce DNA modules sufficiently rigid for use as building blocks for crystalline arrays and nanomechanical devices cultivated more complex arrangements of the well-studied DNA duplex[27,28] – namely, DNA helices linked by multiple crossover junctions. One of the simplest and earliest of the designs is two helices bundled together by a double crossover (DX) to form a modular tile.[29] Such basic tile motifs can be expanded to construct 3D structures, including triangular tensegrity lattices,[30] polyhedra,[31] and nanotubes.[32]

The advent of DNA origami – a construction technique employing kilobase-long scaffold strands joined together by shorter, staple strands to form arbitrary shapes[33] – allowed construction of even larger and more intricate designs, which expanded the reach of DNA nanotechnology to nanosensing,[34,35] nanomedicine,[36] and nanoelectronics.[37] Finally, a similar, but scaffold-free, approach was proposed by Peng Yin and coworkers,[38] in which only short single strands with four distinct binding domains are used. These versatile single-stranded tiles (SSTs) can also be fashioned into a wide array of 2D[39] and 3D[40] shapes.

As many applications depend on the increased structural rigidity of nanostructures as compared to duplex DNA, there has been a significant effort to characterize their elastic properties. Ligation closure experiments on the DX tile revealed a persistence length about twice that of duplex DNA.[41] Even larger persistence lengths, $\sim 10-20\times$ that of duplex DNA, have been achieved with nanotubes comprised of 6 helices;[42,43] for bundles of more helices, persistence lengths $40-200\times$ that of dsDNA have been observed.[44]

More sophisticated applications of DNA nanotechnology require an understanding of the force response beyond that of the basic elastic properties. Such understanding is crucial particularly for applications that exploit behaviour under force, such as biosensors,[35,45,46] springs,[47] force probes[48,49] and tensegrity structures.[50,51] Further, characterizing the long-term stability of nanostructures under constant or intermittent stress is necessary for their use as movable joints and mechanical components like hinges[52] and as hybrid nanopores, which undergo a constant denaturing stress from ionic current flow.[53,54] The recent use of DNA nanotube hydrogels to study theoretical models of semiflexible polymers underscores the importance of knowledge of DNA nanostructural stability under force, as it hinged upon the robustness of the nanotubes under shear force.[55]

Given their substantial contribution to our understanding of the DNA duplex, force spectroscopy experiments are a logical stride in the development of DNA nanotechnology. Experiments have only just begun to explore the more complex force-response behaviour of DNA nanostructures, however; magnetic tweezer experiments have revealed torsional rigidities of origami 6-helix bundles in excess of ∼ 4× those of dsDNA,[42] and recent optical tweezer experiments on origami nanotubes found disassembly forces of 40-50 pN.[56] Here, we build a more comprehensive picture of DNA origami force response by combining coarse-grained simulations with AFM experiments to characterize the unraveling of a standard origami system: the Rothemund tile.[33] We also perform simulations of two previously-published experimental systems for which force-extension data are available: a 10-helix bundle[57] and a biosensor[35] for which the force-extension behaviour signals the presence of aptamers in solution.

By pursuing a combined simulation and experimental approach, we are able to correlate features in force-extension data directly to underlying structural changes, yielding insights inaccessible via experiment alone. To access the timescales relevant to experiments, we use oxDNA,[58,59] a nucleotide-level coarse-grained model, rather than an atomistic description. Because of oxDNA's previous success in reproducing DNA nanostructures,[60–62] capturing DNA mechanical response to tension,[19,63] and describing DNA twist,[64,65] it is well-suited to the present study. The inability of oxDNA to describe S-DNA – the proposed form of the overstretched DNA duplex at high salt concentrations – is not an issue as the origamis unravel at lower forces experimentally than those at which S-DNA forms.

By forcibly unravelling DNA origami, we characterize departures from duplex behaviour and probe the basic physics of origami force compliance, as well as the dynamics of force propagation through the structures. We also probe the mechanical limits of DNA nanostructures under external stress, knowledge of which is crucial for applications that utilize nanostructures to apply or resist force.[47,50,51] Our results provide further validation of oxDNA's robustness in treating DNA origami. More importantly, they illustrate the possible fruits of combining experimental force spectroscopy with simulation to explore unfolding pathways and probe origami mechanical resistance to force in unprecedented detail, which can inform rational design of future nanomechanical structures.

Results and Discussion

Rothemund tile

Our system, shown in Figures 1(a) and (b), is very similar to Rothemund's original design, though we retain a 20-bp stem hairpin that Rothemund excised. The 6909-nuforce-extensioncleotide scaffold is bound by 30-nt and 32-nt staples, most of which have three domains, with one long and two short domains. The design has 24 rows with a central seam dividing the origami into two equivalent halves, and a very regular pattern of staples (Fig. S1).

To mimic the AFM pulling procedure, in which one end of the scaffold is affixed to a glass coverslip and the other to the retractable AFM cantilever, simulations were performed by placing the ends of the tile in harmonic traps (see arrows in Figure 1(b)) and moving one relative to the other at a constant rate. Simulated force loading rates were 3.2×10^8 pN/s and 3.2×10^9 pN/s, and experimental force loading rates were 1.8×10^5 pN/s and 1.4×10^4 pN/s. By 'force loading rate', we mean the average rate of force increase as given by the product of harmonic trap stiffness and speed. This value is thus an instrumental property, and the loading

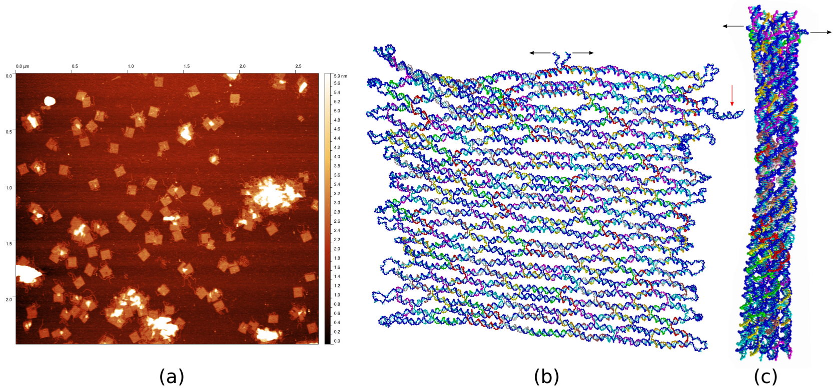


Figure 1: The two main systems studied in this work. (a) An AFM image of Rothemund tiles deposited on a surface. (b) An oxDNA representation of the Rothemund tile. The 20bp-stem hairpin indicated by the red arrow was omitted in Rothemund's original design.[33] (c) An oxDNA representation of the 10-helix bundle presented in Bae *et al.*[57]. For both simulated systems, forces were applied to the nucleotides at the ends of the scaffold, as indicated by the black arrows.

rate the molecule actually 'feels' differs due to the presence of linker strands as well as the particular configuration of the molecule; estimates of the experimental force loading rate from fits to force-extension curves yield values 1-3 orders of magnitude lower than instrumental loading rates. While our simulated loading rates are many times greater than the experimental loading rates, any conversions of coarse-grained timescales to absolute time units must be interpreted with caution. The values above give a "worst-case" disparity; since coarse-graining speeds up diffusive dynamics relative to the microscopic time scales that set the coarse-grained time unit,[66] the resulting telescoping of timescales[67] renders the effective force loading rates for simulations closer to experiment than they would appear from the above values.

Figure 2 contains the experimental (a) and simulated ((b) and (c)) force-extension curves (FECs) for the Rothemund tile. The experimental curves terminate before the full structure is unraveled – which would occur around 4000 nm – because the pulling handle between the molecule and the AFM tip breaks. For domains under mechanical load, there is a non-zero probability that statistically stronger ones break before weaker ones. This probability scales with the magnitude of overlap between the rupture force distributions of both domains.[68,69] In the single molecule force spectroscopy (SMFS) experiments performed here, the pulling handle is loaded throughout the whole retraction and force rarely drops to near zero between rupture events. The state of the pulling handle therefore does not get reset completely, and the probability for rupture increases with extension. It is thus unsurprising to see SMFS curves that do not unfold the whole origami structure in a single pulling cycle.

Nonetheless, several salient features are visible in the data. Immediately striking is the regular, sawtooth pattern present in all of the traces, a signature of the regularity of design in the tile. The tile unravels in units evidenced by these sawtooth rips, indicative of cooperative unfolding within each subunit. Notable, though, is that each 'rip' occurs at more or less the same rupture force: the unfolding of one unit does not render subsequent units easier to unfold.

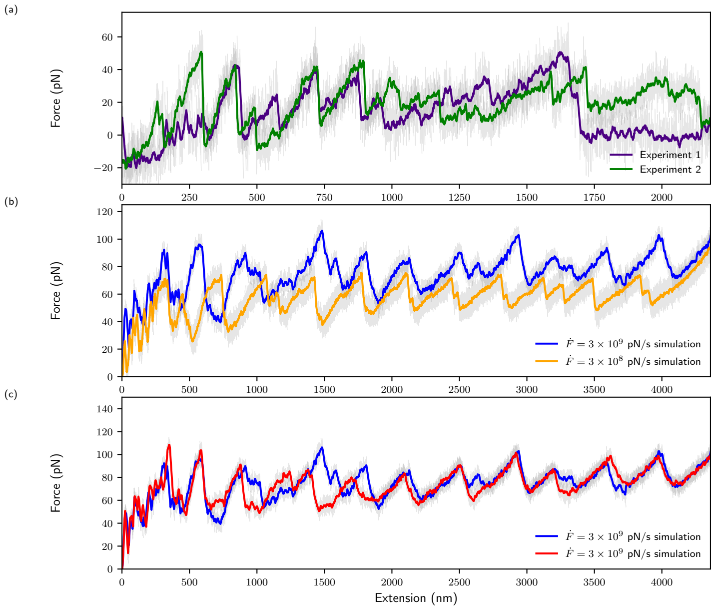


Figure 2: Force-extension curves for the Rothemund tile. (a) Experimental AFM curves at force loading rates of $\dot{F} = 1.8 \times 10^5\,\mathrm{pN/s}$ (purple) and $\dot{F} = 1.4 \times 10^4\,\mathrm{pN/s}$ (green). (b) Simulated pulling curves at $\dot{F} = 3 \times 10^9\,\mathrm{pN/s}$ (blue) and $\dot{F} = 3 \times 10^8\,\mathrm{pN/s}$ (orange). (c) Simulated pulling curves, both collected at $\dot{F} = 3 \times 10^9\,\mathrm{pN/s}$. In all cases, colored traces are averages of gray curves and represent effective sampling rates of $50\,\mathrm{s}^{-1}$ (experiment), $1.3 \times 10^7\,\mathrm{s}^{-1}$ (simulated, slow), and $6.6 \times 10^7\,\mathrm{s}^{-1}$ (simulated, fast).

Also evident in both experimental and simulated curves is the fundamental stochasticity of the process: even under an identical external protocol, the tile unfolds along a slightly different pathway every time. The experimental curves in (a), both collected at the same force loading rate, and the simulated curves in (c), also collected at the same rate, differ nonetheless from one another since barrier crossing is a thermally-activated diffusive process.[70]

Both the experimental and simulated FECs exhibit a decrease in the slope of the force rise preceding each sawtooth 'rip' as more of the tile unravels, indicating a decrease in the apparent force loading rate.[71,72] This is a direct consequence of the effective increase in the linker length as the origami subunits closest to the handles unfold, releasing ssDNA, which behaves as an entropic spring that couples to the harmonic traps to govern the overall force response of the system. The elastic behaviour of ssDNA – typically modelled as a modified freely-jointed chain[1] or a worm-like chain[73] – is highly nonlinear, and is known to reduce the loading rate transmitted to a molecule.[71] Thus, as more ssDNA unravels, the apparent force loading rate decreases.[71] This accounts for the sharper, more jagged features visible near the beginning of the traces being softened, as sharp features are smoothed through convolution with the stiffness of the increasingly long polymeric tethers.[74] The stochasticity in the force-extension traces is more pronounced the farther from equilibrium the protocol is; in the (non-physical) equilibrium limit, the traces would appear identical for every pull. As the apparent force loading rate decreases, therefore, so too does the variation from pull to pull. We can rationalize this by considering that the width of the characteristic rupture force distribution for a given structural element increases with force loading rate, so a smaller spread in observed rupture forces is expected as loading rate decreases.[75,76] This explains the increased regularity of features as the extension grows. The final segment of the simulated curves corresponds to stretching ssDNA.

The simulated FECs also exhibit a dependence of rupture force on probe velocity, with the force required to unravel specific features increasing with pulling speed. This behaviour of rupture force under dynamic loads was first described by Evans and Ritchie,[20] following Bell,[77] who predicted that rupture force is an increasing function of pulling speed. Later, more refined theories have corroborated this,[75,76,78] and experiments are frequently analyzed using this framework.[5,79,80] In this context, we can rationalize the maximum rupture force increasing from $\sim 50\,\mathrm{pN}$ for the experimental loading rates to $\sim 70\,\mathrm{pN}$ for the slower simulated rate, and finally to $\sim 100\,\mathrm{pN}$ for the fastest simulated rate. We reiterate that comparing simulation and experimental timescales is a subtle task, as coarse graining speeds up intrinsic dynamics.[66] To obtain a clearer picture of how the loading rates compare, one could find the computational loading rate for which the average rupture force matched the experimentally observed one.

The added value of simulations lies in the ability to correlate force-extension data with structural changes. Figure 3 tracks the breakage of native base pairs as the tile unravels for two pulling speeds. At $t = 0$, the tile is fully natively bonded. Each yellow block corresponds to two half-rows: the blue 'stripes' punctuating the yellow regions are regions of single-stranded scaffold, which can be seen in Figure 1 (b) at the end of each row. The nucleotide labelled '0' is subjected to the moving harmonic trap, while the nucleotide labelled '6908' sits in a stationary trap. Clearly, the structure unravels from the scaffold ends inwards, corresponding to the location of applied force.

Some asymmetry in the application of force is visible in the plots: 'persistent domains' – regions where staples remain partially bound by a single domain to scaffold regions even after junctions have broken – are more numerous in the lower half of the plots, corresponding to the half of the structure furthest from the moving trap. For example, in the faster simulations, at 5.0 µs (after 15 rows have unfolded), there are 39 persistent domains in this lower half, while the upper half contains only 20 such domains. Similarly, at 8.0 µs, after all rows have unfolded, 44 persistent domains remain in the

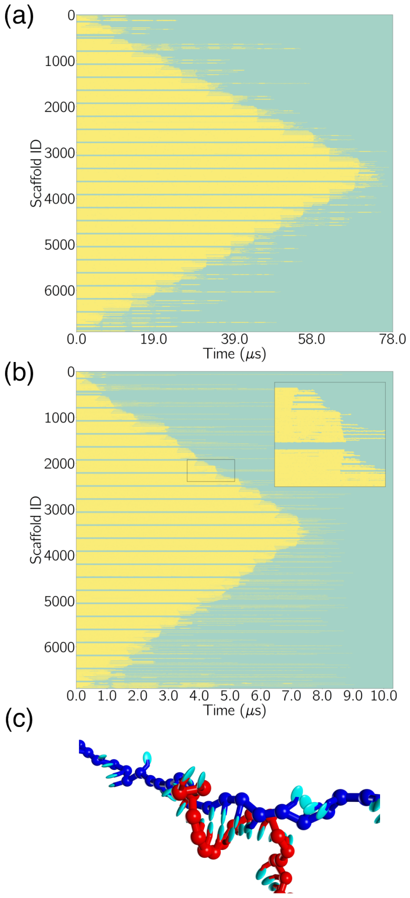


Figure 3: The pattern of native-bond unfolding in pulling simulations of the Rothemund tile. Yellow regions indicate bonds that are still in their native configuration; blue regions indicate an unbonded scaffold nucleotide. (a) Slower pulling rate; (b) faster pulling rate, corresponding to the blue trace in Fig. 2 (b) and (c). The inset shows detail of the unfolding mechanism for a pair of half rows.Nucleotide 0 corresponds to the end associated with the moving trap. (c) An illustration of a persistent staple domain with three remaining base pairs connecting it to the scaffold; that these base pairs do not prevent the scaffold backbone aligning with the force makes these last base pairs more resistant to force-induced melting.

lower half and only 22 in the upper region. This asymmetry is an indication that the timescale over which force is increasing is faster than the time for the force to propagate through the structure; that is, one half of the tile 'feels' a somewhat larger instantaneous force than the other half, making the force-induced melting of these persistent domains more likely. The effect is thus more pronounced for the faster pulling rate. Indeed, in the slower simulations, there is much less asymmetry in the persistent domains; e.g. just after ∼32.0 µs (after 11 rows have unfolded), there are 11 persistent domains in the lower half and 13 in the upper half. The delayed propagation of stress along polymer chains has been noted elsewhere.[81]

A further difference between the two rates is visible in the recurrent breakage and reformation of base pairs in 'persistent' staple domains, which occurs much more frequently at the slowest pulling rate. When the force approaches a maximum, the persistent domains begin to melt and the number of base pairs in these domains diminishes; but once the "rip" has occurred and the force in the system decreases, at the slowest rates there is sufficient time for the domain to reanneal. Note that the last few base pairs between staple and scaffold are significantly more resistant to force-induced melting, because the scaffold back-bone can then approximately align itself with the direction of the force and so the extensional gain from strand melting is much reduced;[19] see Figure 3 (c).

From the plots of base pair breakage versus time, we can glean the basic mechanism by which the Rothemund tile unfolds. First, note that the routing pattern for the tile's rows means that the force is alternately antiparallel and parallel to the duplex; see Figure 4. The gain in free energy per broken base pair is $\sim F\Delta z/\mathrm{bp}$, where $\Delta z/\mathrm{bp}$ is the extension gain per base pair broken. For rows where the scaffold strand is anti-parallel to the force (Fig. 4(a)), each broken base pair contributes the base-base distance along the backbone plus the rise of a base pair in the duplex – in oxDNA, 9.8 Å. By contrast, when the scaffold strand is parallel to the force (Fig. 4(b)) $\Delta z/\mathrm{bp}$ is a more modest ∼3 Å, the difference between the backbone length and the base-pair rise of the duplex. Consequently, the force required to break a base pair in the parallel geometry is much greater than the force required to break a base pair in the antiparallel geometry. This is analogous to the fact that the DNA duplex rupture force for a shear geometry is much higher than that for an unzipping geometry.[19,26,63,82]

Consider the upper half of Figures 3 (a) and (b). The top half of each yellow block corresponds to a half-row with 'shear-like,' or parallel, pulling geometry, and the bottom half corresponds to a half-row with 'unzipping-like,' or antiparallel, pulling geometry.

The force increases roughly linearly during the pulling on the 'parallel' rows until a maximum force is reached that allows the full half-row to be sheared off. At the point when the final seam staple in the row is removed, the force exceeds that which would be necessary to unzip the subsequent half-row; thus this next half-row yields rapidly. Most of the base pairs in the lower half of each yellow half-row break over a short period of time, and the consequent rapid release of scaffold strand leads to the sudden drop in the force. Also, for those staples whose long domain is in the half-row that unzips, yielding generally occurs by the unbinding of the shorter domain from the next half-row, leaving the staple attached to the released scaffold. This is visible as a series of blue lines in the top half of the next yellow block, as well as the thin yellow lines associated with the persistent domains in the bottom halves of the yellow blocks.

An additional feature evident from Fig. 3 is that there is a tendency for the corresponding left-hand and right-hand pairs of half-rows to yield almost simultaneously, giving these plots their rough mirror symmetry. This tendency is more evident both at the slower pulling rate, and at later times. One of the effects of this synchronization is to reduce the variability in the FECs at later times.

This mechanism, wherein two rows unravel co-operatively, leads to the regular features in the FECs. This can be quantified by performing a Lomb-Scargle[83] periodicity analysis of one

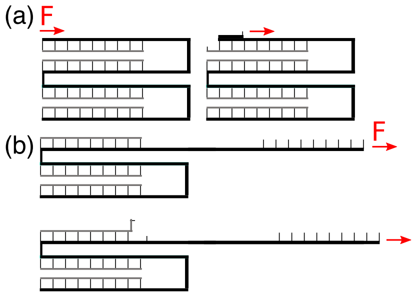


Figure 4: Antiparallel (a) and parallel (b) force application modalities. When force is applied antiparallel to the helix, the geometry is 'unzipping-like' and carries an extension gain per base pair of 9.8 Å. Applying force parallel to the helix, or in a 'shear-like' geometry, by contrast only yields an extension gain per base pair of ∼3 Å. This difference renders the forces observed for rows that are 'sheared' to be much higher than those for rows that are 'unzipped'.

of the FECs in Figure 2. Fig. S4 shows a strong peak that occurs very close to the length of two rows of the tile.

The simulations also allow us to directly visualize the changes in configuration underlying these events. Figure 5 depicts the mechanisms underlying a typical "sawtooth" feature in the FEC. At point A, the top row is in a 'shear-like' pulling geometry. As force increases, the base pairs become increasingly strained, particularly on the left-hand side of the tile, corresponding to the moving harmonic trap. Structure B reveals the tile immediately after the final staple of the topmost left-hand half-row – the central seam – has been broken, but with the seam staple between the lower two half-rows still just in place. The half-row underneath then begins to unzip, and the left-hand half-row has completely unzipped by stage C. The consequent removal of the second bridging seam staple has a significant effect on the unravelling of the right-hand side. It allows the section nearest the seam to rotate about the nearest junction, dramatically changing the geometry of the pulling. At C, unzipping of the second half-row on the right-hand side has begun, even though the staples nearest the seam are still intact in a subunit that is now roughly vertical. The small peak in the FEC at C corresponds to this yielding of the right-hand side. The key role of the seam staples underlies the synchronized unravelling of the pairs of half-rows. Two rows, corresponding to the basic sawtooth unit in the force-extension plots, have been unravelled between A and D – further confirmation of our picture of the basic mechanism of unfolding for the Rothemund tile.

10-helix bundle

The 10-helix bundle was originally designed by Bae *et al.*[57] to explore the concept of mechanical origami assembly as an alternative to annealing—the scaffold was held at constant tension to eliminate transient secondary structure as folding progressed—but the force-extension curve associated with disassembly was also probed. It is both topologically and structurally more complex than the Rothemund tile. Firstly, the staple routings on the 1768-nucleotide scaffold are more non-local, with some staples spanning up to 5 domains. Secondly, the structure is 3D rather than 2D. Despite these differences, the FECs share many common features with those of the Rothemund tile: the basic sawtooth-like, 'stick-slip' dynamics, shown in Figure 6; an overall increase in rupture forces as pulling rate increases; and increased reproducibility with increasing extension in the traces performed under the same

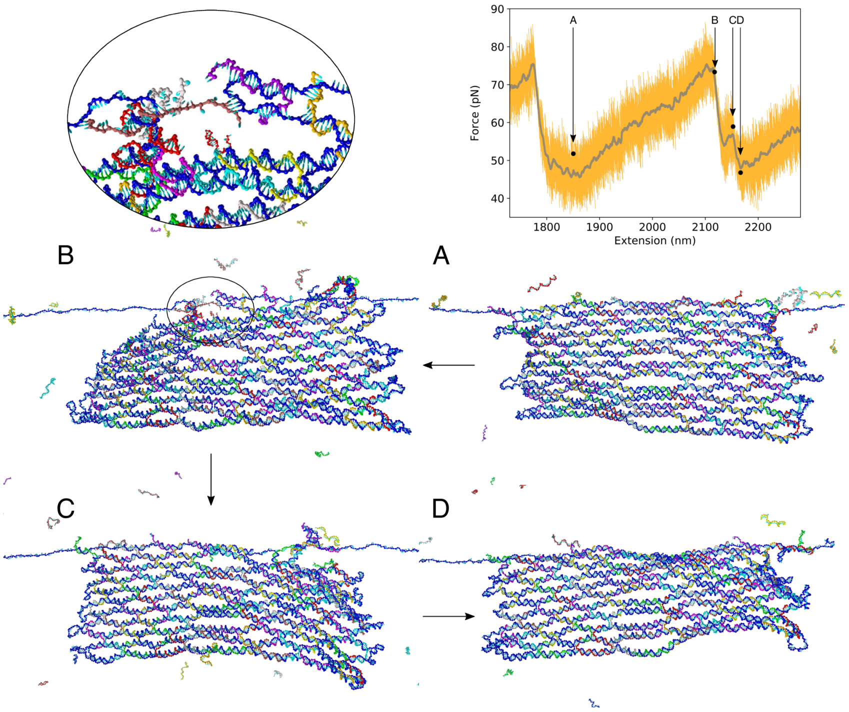


Figure 5: Structural illustration of the unfolding mechanism of the tile. A section of a simulated force-extension curve at $\dot{F} = 3 \times 10^8$ pN/s is shown; the gray curve is an average. The sawtooth feature that peaks at point B corresponds to the top row being sheared open; subsequently, the next row unzips. The small bump at C occurs because the right half of the row unzips at a slightly later time than the left.

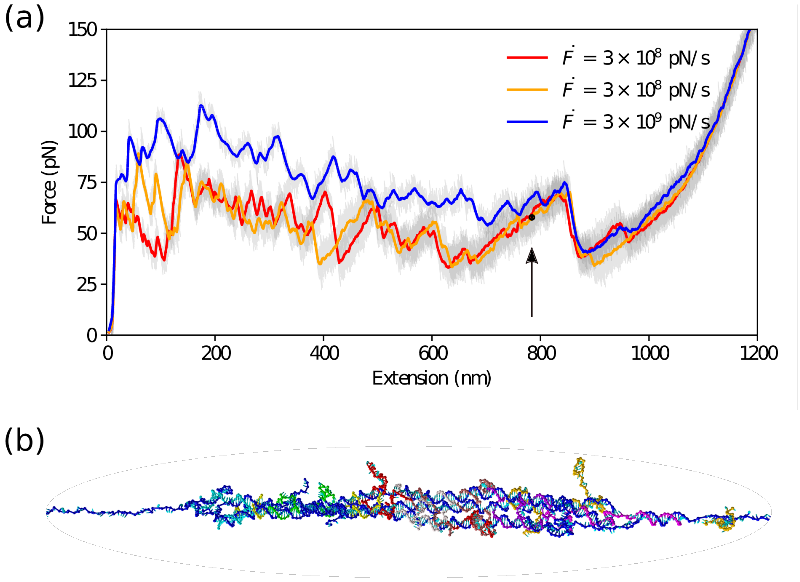


Figure 6: (a) Force-extension curves corresponding to unfolding the 10-helix bundle with moving harmonic traps at loading rates of $\dot{F} = 3 \times 10^9$ pN/s (blue) and $\dot{F} = 3 \times 10^8$ pN/s (red and green). (b) Structural snapshot, corresponding to the black marker and arrow in (a), of the final rupture feature in the force-extension curves: three helices are in a shear geometry.

conditions.

One consequence of the more complex structure, however, is that there is a less obvious preferred folding pathway; a multiplicity of possible unfolding events may yield the same change in extension. That the helix bundle indeed samples multiple unfolding pathways is clear from Figures 7 and 8. In one simulation, the bundle unfolds along a single helix before other helices unravel. In another trace, collected under identical simulation conditions, other helices begin to yield from the opposite scaffold end before the first is finished unravelling. Even along a single pathway, there is no clear uniformity in terms of a repeating unfolding unit; Figure S4 reveals an absence of periodic signatures in Fig. 6(a) associated with any particular length scales.

Also notable is the fact that in many regions, successive rupture events occur at progressively lower forces. For all traces, the highest forces reached occur near the start of the unravelling process. This 'strain-softening' effect arises because, unlike with the Rothemund tile, the gain in extension per base pair broken varies significantly as the helix bundle unravels. Figure 9 reveals the difference: whereas the behaviour of Δz/bp is essentially linear for the tile, the helix bundle Δz/bp contains regions of curvature, wherein breaking of key domains renders the structure increasingly compliant; it 'catastrophically' fails. This unfolding behaviour has been described previously as 'shielded' unfolding, in which structural topology prevents weaker barriers from being compromised before stronger ones.[84] A specific structural example of this shielding is shown in Figure 10: once the first full helix has unravelled at ~ 170 nm, the remainder of the bundle is comparatively easier to unfold. Given the multi-domain nature of the staples, unravelling the first helix weakens many of the staples holding subsequent helices in place; this gives rise to the strain-softening observed. The plots in Fig. 7 (particularly (b)) illustrate the weakening of the interior regions of the bundle. Unravelling does not simply occur from the ends inwards, as with the tile; as the final failure is approached, the structure appears to separate into multiple blocks. Strain-softening is visible up to ~ 650 nm extension. At this point, as illustrated in Figure 6 (b), three helices remain and force application occurs roughly in a shearing geometry, accounting for the ~ 75 pN force required to unfold this final feature – not far below the initial maximum force of ~ 85 pN.

Employing a linearly increasing force protocol rather than harmonic traps makes the bundle's catastrophic failure even more evident: Figure 11 contains FECs for a range of force loading rates; scaffold ends are both held at the same, linearly increasing, tension. Again, we note the increase in rupture force for increasing force loading rate. After a few initial rupture events, the bundle essentially unfolds 'all at once,' evidenced by the long linear regions in the force-extension data. Once the monotonically increasing force surpasses the critical force at which all landscape barriers are removed, unfolding enters a purely downhill regime. Thus, the force application protocol is clearly of central importance in the analysis of molecular mechanical properties; in the case of harmonic traps, the molecule has opportunities to partially relax, allowing us to distinguish more detailed features in the force-extension curves.

Bae *et al.*[57] also noted cooperative unfolding of the helix bundle, which in their magnetic tweezer pulling experiments began to yield at ~ 30 pN – a rupture force half as large as for our systems, which yield at ~ 60 pN under the slowest pulling rate. Their experimental force loading rate was not provided, but it was certainly smaller than our slowest rate, which probably mostly accounts for the discrepancy.

It is noticeable that there is a pronounced asymmetry with respect to scaffold ends in Fig. 7(a), but to a much lesser extent for the faster pulling rate. This pulling-rate dependence is the opposite of the behaviour observed for the tile (Fig. 3(a) and (b)), and is a result of an asymmetry in the helix bundle's design. Specifically, there is an additional nick close to the end of helix 0 (see Fig. S2) that will lead to a somewhat lower barrier for unravelling to be initiated from this end. If initiation occurs from just one end, the origami will then tend to rotate so that the points at which the force

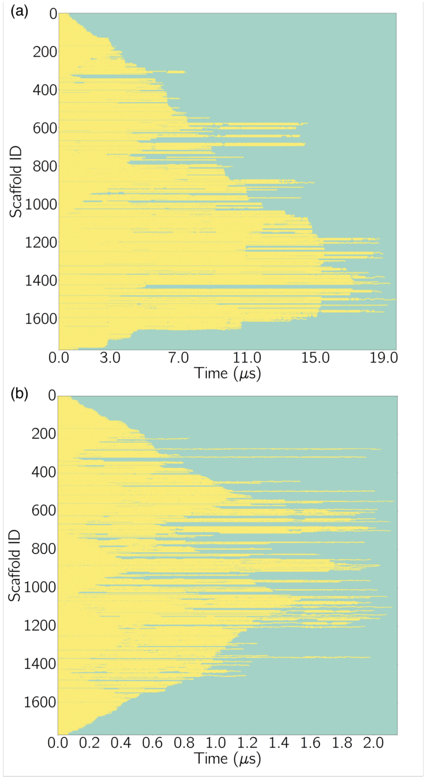


Figure 7: The pattern of native-bond unfolding in pulling simulations of the 10-helix bundle at (a) $\dot{F} = 3 \times 10^8$ pN/s (the dark red curve in Fig. 6(a)) and (b) $\dot{F} = 3 \times 10^9$ pN/s (the blue curve in Fig. 6(a)). Yellow regions indicate bonds that are still in their native configuration; blue regions indicate an unbonded scaffold nucleotide.

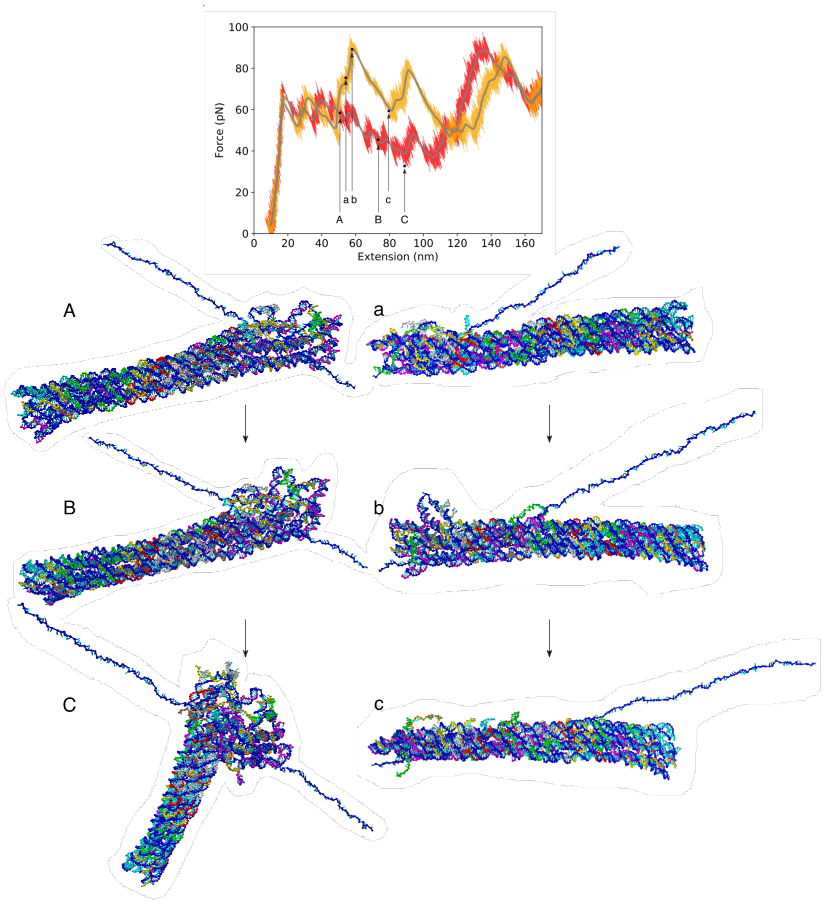


Figure 8: Structures exemplary of the stochasticity manifest in the force-extension unfolding curves for the 10-helix bundle. Curves for each simulation, performed at the force loading rate $\dot{F} = 3 \times 10^9$ pN/s, are shown at top, and the locations of the structural snapshots are indicated. The unfolding pathway of the red force-extension curve, shown on the left, exhibits unravelling from both ends before the first helix has completely yielded. The unfolding pathway of the orange force-extension curve, however, features a complete unfolding of the first helix before others begin to yield.

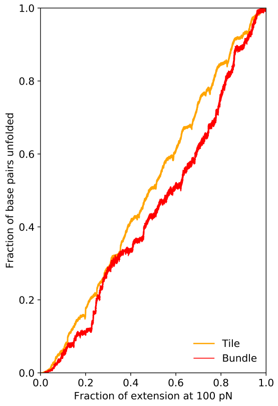


Figure 9: The fraction of base pairs unfolded as a function of the fractional extension (measured with respect to the extension at 100 pN) for the Rothemund tile and 10-helix bundle. The Rothemund tile exhibits a regular, linear pattern of Δz/bp as rows alternatingly shear and unzip. By contrast, the curve for the helix bundle is much more non-linear, and curvature in the plot indicates a strain-softening, or shielding, effect: once certain key basepairs unfold, subsequent basepairs are easier to break.

acts on the origami are co-linear, which in turn leads to a greater extensional release if that row continues to unravel. Furthermore, the resulting localized strain-softening will make it more likely that subsequent unraveling will continue from this end. When the pulling rate is higher, as in Figure 7 (b), the system has less time to find the most favourable pathway, and a more symmetric picture emerges.

7-tile biosensor

Our final case study allows us to further explore the limits and capabilities of coarse-grained origami force pulling simulations, and is of the 7-tile biosensor of Koirala *et al.*[35]. The 14 761-nucleotide origami consists of seven rectangular tiles joined by 44-base pair duplex 'locks', where the breaking of the locks can be observed as a large change in extension due to the connected tiles then being able to re-orient. The slowest pulling rate that we used was insufficient to observe separated opening of the 6 inter-tile locks as in experiment; instead, all the locks were in the process of opening at the same time (Figs. S9 and S10). However, the locks break in the order expected, outermost to innermost, just not individually; Figure 12 illustrates the unlocking mechanism. While the signature of the locks opening can be seen in the force-extension data (Figure S10), the resolution of individual events is not comparable to the experimental force-extension curves[35] given the relatively high force loading rate. Furthermore, at this high rate, the tiles themselves begin to unravel somewhat before all of the locks have broken, a sign that our simulations are far from equilibrium. A detailed correlation of structural changes to force-extension data is certainly feasible for this system, but would require further simulation work at slower pulling rates.

Conclusion

Here, we have used molecular simulations performed using the oxDNA coarse-grained model to characterize the mechanical force response of two archetypal DNA origami systems, the

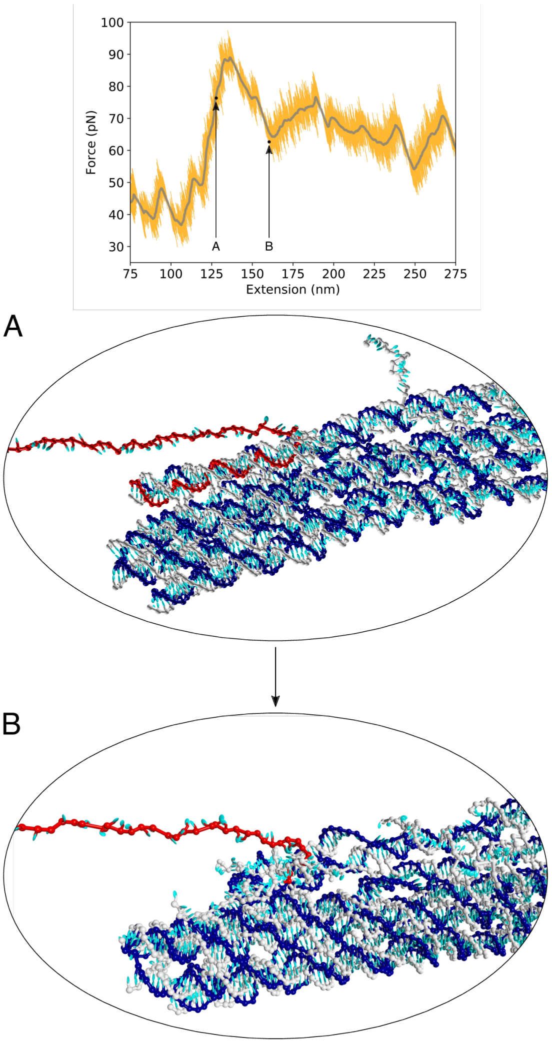


Figure 10: Structural illustration of the first large unfolding rip in the force-extension curve collected at $\dot{F} = 3 \times 10^9$ pN/s. The origin of the 'strain-softening effect', in which subsequent unfolding rips occur at lower forces, can be seen in the images: once the first helix (red) completely unwinds between A and B, many of the multiple-domain staples (white) attached to other helices (blue) have been compromised. Subsequent helices thus require lower forces to yield.

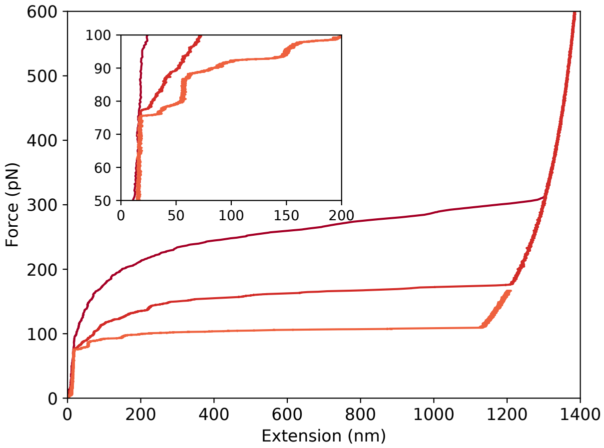


Figure 11: Simulated force-extension curves for the 10-helix bundle collected under a linearly increasing force protocol. Force loading rates decrease from top to bottom: $\dot{F} = 3.2 \times 10^8$ pN/s, 3.2×10^7 pN/s, 3.2×10^6 pN/s.

Rothemund tile and a 10-helix bundle, and to explore the force response of a 7-tile biosensor. We also performed AFM pulling experiments on the tile, enabling us to correlate features in experimental data with structural changes revealed in simulation. Both experiments and simulations exhibited regular, sawtooth unfolding behaviour in the force-extension curves for the Rothemund tile, and captured yielding at similar forces; our simulated yielding force was ∼15 pN higher than the experimental 50 pN due to higher pulling rates.

Our force-extension data yields insight into the complex free-energy landscapes for origami unfolding, underscoring the much richer force-induced melting behaviour of origami compared to dsDNA. We observe stochasticity in the unfolding process for both systems; multiple pathways, cooperative unfolding, and strain-softening in the unfolding of the 10-helix bundle; and geometry-dependent Δz/bp. Since force biases the pathways taken through the unfolding landscape, we expect the assembly landscapes for these systems to exhibit even greater complexity.

Variations in rupture forces with pulling rates imply that any designed structural resistance to force will be timescale dependent.[19] The effect of design on force-response behaviour has been highlighted by the disparate results obtained for the Rothemund tile, which unfolds modularly, and the 10-helix bundle, whose non-local staple routings lead to strain-softening and catastrophic failure.

Our results also reveal a glimpse of the complexities of force propagation through origami structures, manifesting in asymmetric unravelling behaviour. Valuable future investigations could consider in more detail how force propagation dynamics are complicated by the presence of junctions and multi-domain staples.

We have demonstrated the possibilities for improving comprehension of origami mechanical response and informing rational design of nanostructures through joint simulation and experimental studies, and hope future work will continue to profit from the insights offered by DNA origami force spectroscopy.

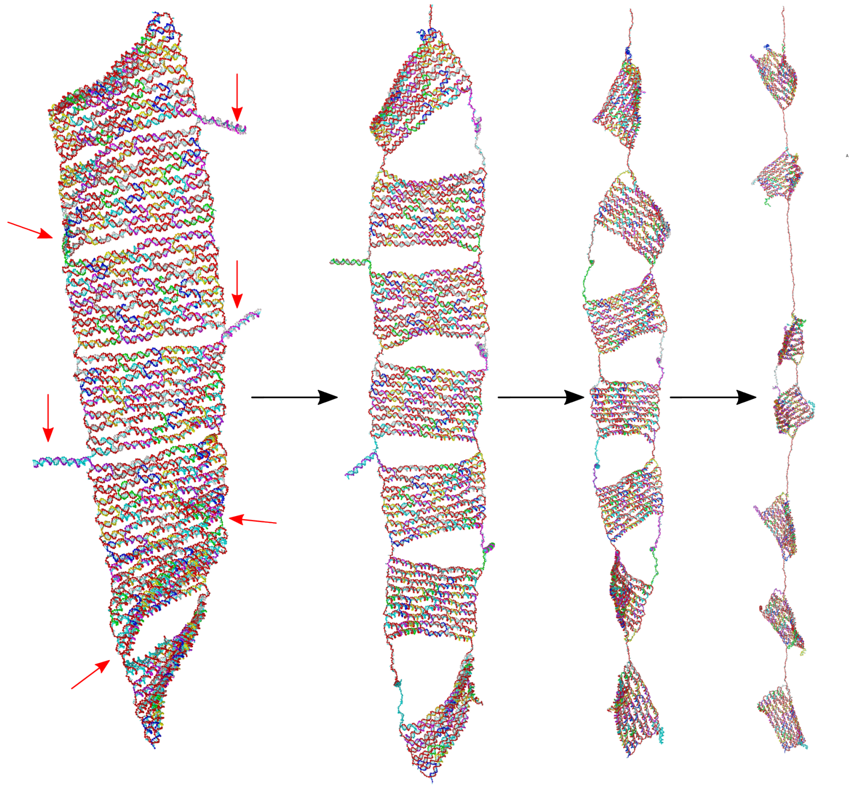


Figure 12: Structural snapshots illustrating the unfolding mechanism of the 7-tile biosensor of Koirala *et al.*[35] at a force-loading rate of $\dot{F} = 3.2 \times 10^8\,\mathrm{pN/s}$. The duplex 'locks' are labelled with red arrows in the first panel. The end-to-end length in the leftmost panel is 181 nm, and in the rightmost panel is 475 nm.

Methods

Experiments

Briefly, scaffold '7560'[85] from bacteriophage M13mp18[33] was double functionalized by ligation with biotin and thiol group to facilitate attachment to either the streptavidin-coated AFM tips or maleimide-functionalized substrate. The circular, single stranded scaffold was digested by BamHI-HF and EcoRI-HF after partial pre-hybridization at restriction sites. The linearized scaffold was purified by solid phase extraction. Afterwards, partially double-stranded scaffold was ligated with 10× excess of double-stranded inserts carrying either biotin or thiol (5'-BIO; 3'-THIOL + 5'-Phosphate) by T4 DNA ligase. This ligated scaffold was either used in its unpurified form (Fig.2a, green curve) to fold DNA origami rectangles,[86] or was purified from excess functionalization strands by size exclusion (Fig.2a, red curve) in spin filtration columns and subsequently folded. The folding procedure was already reported elsewhere.[86] The freshly folded DNA origami was purified by PEG precipitation.[87] Prior to 'unzipping' experiments, the origami pellet was resolved in 1× TE buffer (pH 8) supplemented with 10 mM MgCl2 and 2 mM TCEP. TCEP is required to reduce oxidized thiol-groups and enable maleimide conjugation reaction on surfaces. Afterwards, the resolved origami samples were coupled to AFM sample slides.

Correct DNA origami folding from digested scaffold was confirmed by AFM imaging; see Fig 1. Freshly cleaved mica surfaces were coated by 0.01% poly-L-ornithine to facilitate origami immobilization. Imaging was carried out in intermittent contact mode.

AFM Sample Preparation

Clean glass cover slides and UV-ozone cleaned silicon nitride AFM cantilevers were covalently functionalized first with amino-silane and subsequently with 5kDa NHS-PEG-maleimide polymer linkers. Reduced sulfhydryl groups of DNA origami were coupled to the free maleimide groups on the cover slides by forming permanent thiol-ether bonds.[68,88] Maleimide groups on the cantilevers were bound to coenzyme A, and subsequently to a monovalent variant of StrepTactin (monoST)[89,90] or a monomeric variant of streptavidin via Sfp phosphopantetheinyl transferase (SFP) with ybbR tags (amino acid sequence DSLEFIASKLA) for use as specific AFM pulling handles.[91,92]

SMFS Measurement Protocol

SMFS experiments were performed in TE buffer supplemented with magnesium at room temperature with custom-built instruments. Pulling velocities were set to $1\,\mu\mathrm{m\,s^{-1}}$ or $1.6\,\mu\mathrm{m\,s^{-1}}$. Cantilever and glass slide positioning was controlled by PID feedback loops during measurement. For analysis, curves with peaks at more than 150 nm extension and more than 30 pN force were sorted automatically, after thermal cantilever calibration, interferometric piezo calibration, and data conversion, similar to previous work.[92–95]

Simulations

We ran molecular dynamics (MD) simulations using oxDNA, a coarse-grained model whose basic unit is a rigid nucleotide that interacts with other nucleotides through stacking, hydrogen-bonding, excluded volume, electrostatic and backbone potentials. An Andersen-like thermostat[96] ensured diffusive particle motion in the canonical ensemble. Simulations were performed on GPUs and ranged in duration from 10^8 to 10^9 steps. At the slowest pulling rate, the Rothemund tile simulation took ~73 days, and that of the 7-tile biosensor took 63 days. Temperatures relevant to experiments were used: 20 °C for the tile, 36 °C for the helix bundle,[57] and 30 °C for the 7-tile assembly; and we used a high salt concentration typical of origami studies, $[\mathrm{Na}^+] = 0.5\,\mathrm{M}$.

Sequence dependence did not significantly alter the unfolding behaviour for the tile, and was only incorporated into the tile simulations at 3.2×10^9 pN/s. Simulations of the 10 HB and 7-tile structures did not incorporate sequence dependence.

In all cases of harmonic pulling, the centres-of-mass of the scaffold end nucleotides were subjected to 3D harmonic traps. Scaffold end-to-end separation along the axis separating the traps was measured and subtracted from the separation between trap centres to yield an 'effective' trap displacement, which was multiplied by the stiffness of the traps in series to give the instantaneous force. For linear pulling, a constantly increasing tension was applied to scaffold end nucleotides.

Additional simulation details are provided in the supplementary material.

Conflict of Interest: The authors declare no competing financial interest.

Acknowledgment. M.C.E. is grateful to the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Rhodes Trust for financial support. The authors acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility (http://dx.doi.org/10.5281/zenodo.22558) and the resources provided by the "Cambridge Service for Data Driven Discovery" (CSD3) system operated by the University of Cambridge Research Computing Service funded by EPSRC Tier-2 capital grant EP/P020259/1. Part of this work has been supported through the Fraunhofer Attract project 601 683.

Supporting Information Available: Supporting material is available free of charge at . . .

References

1. Smith, S. B.; Cui, Y.; Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and sinle-stranded DNA molecules. *Science* **1996**, *271*, 795–799.
2. Bosco, A.; Camunas-Soler, J.; Ritort, F. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.* **2014**, *42*, 2064–2074.
3. McIntosh, D. B.; Ribeck, N.; Saleh, O. A. Detailed scaling analysis of low-force polyelectrolyte elasticity. *Phys. Rev. E* **2009**, *80*, 041803.
4. Essevaz-Roulet, B.; Bockelmann, U.; Heslot, F. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 11935–11940.
5. Strunz, T.; Oroszlan, K.; Schäfer, R.; Güntherodt, H.-J. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11277–11282.
6. Khner, F.; Morfill, J.; Neher, R. A.; Blank, K.; Gaub, H. E. Force-Induced DNA Slippage. *Biophys. J.* **2007**, *92*, 2491 – 2497.
7. van Mameren, J.; Gross, P.; Farge, G.; Hooijman, P.; Modesti, M.; Falkenberg, M.; Wuite, G. J. L.; Peterman, E. J. G. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18231–18236.
8. Gross, P.; Laurens, N.; Oddershede, L. B.; Bockelmann, U.; Peterman, E. J. G.; Wuite, G. J. L. Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.* **2011**, *7*, 731–736.
9. Romano, F.; Chakraborty, D.; Doye, J. P. K.; Ouldridge, T. E.; Louis, A. A. Coarse-grained simulations of DNA overstretching. *J. Chem. Phys.* **2013**, *138*.
10. Forth, S.; Deufel, C.; Sheinin, M. Y.; Daniels, B.; Sethna, J. P.; Wang, M. D. Abrupt Buckling Transition Observed during the Plectoneme Formation of Individual DNA Molecules. *Phys. Rev. Lett.* **2008**, *100*, 148301.
11. Nomidis, S. K.; Kriegel, F.; Vanderlinden, W.; Lipfert, J.; Carlon, E. Twist-Bend Coupling and the Torsional Response of Double-Stranded DNA. *Phys. Rev. Lett.* **2017**, *118*, 217801.

12. Zhang, Y.; McEwan, A.; Crothers, D.; Levene, S. Analysis of In-Vivo LacR-Mediated Gene Repression Based on the Mechanics of DNA Looping. *PLoS ONE* **2006**, *1*, e136.

13. Goodman, S. D.; Nash, H. A. Functional replacement of a protein-induced bend in a DNA recombination site. *Nature* **1989**, *341*, 251–254.

14. Bustamante, C.; Bryant, Z.; Smith, S. B. Ten years of tension: single-molecule DNA mechanics. *Nature* **2003**, *421*, 423–427.

15. Goktas, M.; Blank, K. G. Molecular Force Sensors: From Fundamental Concepts toward Applications in Cell Biology. *Adv. Mater. Interf.* **2017**, *4*, 1600441.

16. Wang, X.; Ha, T. Defining Single Molecular Forces Required to Activate Integrin and Notch Signaling. *Science* **2013**, *340*, 991–994.

17. Blakely, B. L.; Dumelin, C. E.; Trappmann, B.; McGregor, L. M.; Choi, C. K.; Anthony, P. C.; Duesterberg, V. K.; Baker, B. M.; Block, S. M.; Liu, D. R. *et al.* A DNA-based molecular probe for optically reporting cellular traction forces. *Nat. Methods* **2014**, *11*, 1229–1232.

18. Cocco, S.; Monasson, R.; Marko, J. F. Force and kinetic barriers to unzipping of the DNA double helix. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8608–8613.

19. Mosayebi, M.; Louis, A. A.; Doye, J. P. K.; Ouldridge, T. E. Force-Induced Rupture of a DNA Duplex: From Fundamentals to Force Sensors. *ACS Nano* **2015**, *9*, 11993–12003.

20. Evans, E.; Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **1997**, *72*, 1541–1555.

21. Hatch, K.; Danilowicz, C.; Coljee, V.; Prentiss, M. Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs. *Phys. Rev. E* **2008**, *78*, 011920.

22. Williams, M. C.; Wenner, J. R.; Rouzina, I.; Bloomfield, V. A. Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching. *Biophys. J.* **2001**, *80*, 1932–1939.

23. Wenner, J. R.; Williams, M. C.; Rouzina, I.; Bloomfield, V. A. Salt Dependence of the Elasticity and Overstretching Transition of Single DNA Molecules. *Biophys. J.* **2002**, *82*, 3160 – 3169.

24. Fu, H.; Chen, H.; Koh, C. G.; Lim, C. T. Effects of magnesium salt concentrations on B-DNA overstretching transition. *Eur. Phys. J. E* **2009**, *29*, 45–49.

25. Huguet, J. M.; Bizarro, C. V.; Forns, N.; Smith, S. B.; Bustamante, C.; Ritort, F. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 15431–15436.

26. Rief, M.; Clausen-Schaumann, H.; Gaub, H. E. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Mol. Biol.* **1999**, *6*, 346–349.

27. Seeman, N. C. Nanomaterials Based on DNA. *Annu. Rev. Biochem.* **2010**, *79*, 65–87.

28. Jones, M. R.; Seeman, N. C.; Mirkin, C. A. Programmable materials and the nature of the DNA bond. *Science* **2015**, *347*, 1260901.

29. Fu, T. J.; Seeman, N. C. DNA double-crossover molecules. *Biochemistry* **1993**, *32*, 3211–3220.

30. Liu, D.; Wang, M.; Deng, Z.; Walulu, R.; Mao, C. Tensegrity:? Construction of Rigid DNA Triangles with Flexible Four-Arm DNA Junctions. *J. Am. Chem. Soc.* **2004**, *126*, 2324–2325.

31. He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **2008**, *452*, 198.
32. Mathieu, F.; Liao, S.; Kopatsch, J.; Wang, T.; Mao, C.; Seeman, N. C. Six-Helix Bundles Designed from DNA. *Nano Lett.* **2005**, *5*, 661–665.
33. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **2006**, *440*, 297–302.
34. Bell, N. A. W.; Engst, C. R.; Ablay, M.; Divitini, G.; Ducati, C.; Liedl, T.; Keyser, U. F. DNA Origami Nanopores. *Nano Lett.* **2012**, *12*, 512–517.
35. Koirala, D.; Shrestha, P.; Emura, T.; Hidaka, K.; Mandal, S.; Endo, M.; Sugiyama, H.; Mao, H. Single-Molecule Mechanochemical Sensing Using DNA Origami Nanostructures. *Angew. Chem. Int. Ed.* **2014**, *126*, 8275–8279.
36. Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Oliveira, C. L. P. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **2009**, *459*, 73–76.
37. Pinheiro, A. V.; Han, D.; Shih, W. M.; Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.* **2011**, *6*, 763–772.
38. Yin, P.; Hariadi, R. F.; Sahu, S.; Choi, H. M. T.; Park, S. H.; LaBean, T. H.; Reif, J. H. Programming DNA Tube Circumferences. *Science* **2008**, *321*, 824–826.
39. Wei, B.; Dai, M.; Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **2012**, *485*, 623.
40. Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **2012**, *338*, 1177–1183.
41. Sa-Ardyen, P.; Vologodskii, A. V.; Seeman, N. C. The Flexibility of DNA Double Crossover Molecules. *Biophys. J.* **2003**, *84*, 3829–3837.
42. Kauert, D. J.; Kurth, T.; Liedl, T.; Seidel, R. Direct Mechanical Measurements Reveal the Material Properties of Three-Dimensional DNA Origami. *Nano Lett.* **2011**, *11*, 5558–5563.
43. Wang, T.; Schiffels, D.; Martinez Cuesta, S.; Kuchnir Fygenson, D.; Seeman, N. C. Design and Characterization of 1D Nanotubes and 2D Periodic Arrays Self-Assembled from DNA Multi-Helix Bundles. *J. Am. Chem. Soc.* **2012**, *134*, 1606–1616.
44. Schiffels, D.; Liedl, T.; Fygenson, D. K. Nanoscale Structure and Microscale Stiffness of DNA Nanotubes. *ACS Nano* **2013**, *7*, 6700–6710.
45. Ke, Y.; Lindsay, S.; Chang, Y.; Liu, Y.; Yan, H. Self-Assembled Water-Soluble Nucleic Acid Probe Tiles for Label-Free RNA Hybridization Assays. *Science* **2008**, *319*, 180–183.
46. Gu, H.; Yang, W.; Seeman, N. C. DNA Scissors Device Used to Measure MutS Binding to DNA Mis-pairs. *J. Am. Chem. Soc.* **2010**, *132*, 4352–4357.
47. Iwaki, M.; Wickham, S. F.; Ikezaki, K.; Yanagida, T.; Shih, W. M. A programmable DNA origami nanospring that reveals force-induced adjacent binding of myosin VI heads. *Nat. Commun.* **2016**, *7*, 13715.
48. Funke, J. J.; Ketterer, P.; Lieleg, C.; Schunter, S.; Korber, P.; Dietz, H. Uncovering the forces between nucleosomes using DNA origami. *Sci. Adv.* **2016**, *2*, 1600974.
49. Nickels, P. C.; Wünsch, B.; Holzmeister, P.; Bae, W.; Kneer, L. M.; Grohmann, D.; Tinnefeld, P.; Liedl, T. Molecular force spectroscopy with a DNA origami–based

nanoscopic force clamp. *Science* **2016**, *354*, 305–307.

50. Liedl, T.; Hogberg, B.; Tytell, J.; Ingber, D. E.; Shih, W. M. Self-assembly of three-dimensional prestressed tensegrity structures from DNA. *Nat. Nanotechnol.* **2010**, *5*, 520–524.
51. Simmel, S. S.; Nickels, P. C.; Liedl, T. Wireframe and Tensegrity DNA Nanostructures. *Acc. Chem. Res.* **2014**, *47*, 1691–1699.
52. Marras, A. E.; Zhou, L.; Su, H.-J.; Castro, C. E. Programmable motion of DNA origami mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 713–718.
53. Bell, N. A. W.; Engst, C. R.; Ablay, M.; Divitini, G.; Ducati, C.; Liedl, T.; Keyser, U. F. DNA Origami Nanopores. *Nano Lett.* **2012**, *12*, 512–517.
54. Wei, R.; Martin, T. G.; Rant, U.; Dietz, H. DNA Origami Gatekeepers for Solid-State Nanopores. *Angew. Chem. Int. Ed.* **2012**, *51*, 4864–4867.
55. Schuldt, C.; Schnauß, J.; Händler, T.; Glaser, M.; Lorenz, J.; Golde, T.; Käs, J. A.; Smith, D. M. Tuning Synthetic Semiflexible Networks by Bending Stiffness. *Phys. Rev. Lett.* **2016**, *117*, 197801.
56. Shrestha, P.; Emura, T.; Koirala, D.; Cui, Y.; Hidaka, K.; Maximuck, W. J.; Endo, M.; Sugiyama, H.; Mao, H. Mechanical properties of DNA origami nanoassemblies are determined by Holliday junction mechanophores. *Nucleic Acids Res.* **2016**, *44*, 6574–6582.
57. Bae, W.; Kim, K.; Min, D.; Ryu, J.-K.; Hyeon, C.; Yoon, T.-Y. Programmed folding of DNA origami structures through single-molecule force control. *Nat. Commun.* **2014**, *5*, 5654.
58. Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, 085101.
59. Šulc, P.; Romano, F.; Ouldridge, T. E.; Rovigatti, L.; Doye, J. P. K.; Louis, A. A. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.* **2012**, *137*, 135101.
60. Snodin, B. E. K.; Randisi, F.; Mosayebi, M.; Šulc, P.; Schreck, J. S.; Romano, F.; Ouldridge, T. E.; Tsukanov, R.; Nir, E.; Louis, A. A. *et al.* Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.* **2015**, *142*, 234901.
61. Schreck, J. S.; Romano, F.; Zimmer, M. H.; Louis, A. A.; Doye, J. P. K. Characterizing DNA Star-Tile-Based Nanostructures Using a Coarse-Grained Model. *ACS Nano* **2016**, *10*, 4236–4247.
62. Sharma, R.; Schreck, J. S.; Romano, F.; Louis, A. A.; Doye, J. P. K. Characterizing the Motion of Jointed DNA Nanostructures Using a Coarse-Grained Model. *ACS Nano* **2017**, *11*, 12426–12435.
63. Romano, F.; Chakraborty, D.; Doye, J. P. K.; Ouldridge, T. E.; Louis, A. A. Coarse-grained simulations of DNA overstretching. *J. Chem. Phys.* **2013**, *138*, 085101.
64. Matek, C.; Ouldridge, T. E.; Doye, J. P. K.; Louis, A. A. Plectoneme tip bubbles: Coupled denaturation and writhing in supercoiled DNA. *Sci. Rep.* **2015**, *5*, 7655.
65. Matek, C.; Ouldridge, T.; Levy, A.; Doye, J.; Louis, A. DNA cruciform arms nucleate through a correlated but asynchronous cooperative mechanism. *J. Phys. Chem. B* **2012**, *116*, 11616–11625.
66. Depa, P.; Chen, C.; Maranas, J. K. Why are coarse-grained force fields too fast? A look at dynamics of four coarse-grained polymers. *J. Chem. Phys.* **2011**, *134*, 014903.

67. Padding, J. T.; Louis, A. A. Hydrodynamic interactions and Brownian forces in colloidal suspensions: Coarse-graining over time and length scales. *Phys. Rev. E* **2006**, *74*, 031402.

68. Jobst, M. A.; Milles, L. F.; Schoeler, C.; Ott, W.; Fried, D. B.; Bayer, E. A.; Gaub, H. E.; Nash, M. A. Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *eLIFE* **2015**, *4*, 180–183.

69. Schoeler, C.; Verdorfer, T.; Gaub, H. E.; Nash, M. A. Biasing effects of receptor-ligand complexes on protein-unfolding statistics. *Phys. Rev. E* **2016**, *94*, 042412.

70. Kramers, H. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7*, 284 – 304.

71. Ray, C.; Brown, J. R.; Akhremitchev, B. B. Correction of Systematic Errors in Single-Molecule Force Spectroscopy with Polymeric Tethers by Atomic Force Microscopy. *J. Phys. Chem. B* **2007**, *111*, 1963–1974.

72. Neuert, G.; Albrecht, C.; Pamir, E.; Gaub, H. Dynamic force spectroscopy of the digoxigenin antibody complex. *FEBS Lett.* **2006**, *580*, 505–509.

73. Alemany, A.; Ritort, F. Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins. *Biopolymers* **2014**, *101*, 1193–1199.

74. Woodside, M. T.; Block, S. M. Reconstructing Folding Energy Landscapes by Single-Molecule Force Spectroscopy. *Annu. Rev. Biophys.* **2014**, *43*, 19–39.

75. Dudko, O. K.; Hummer, G.; Szabo, A. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Phys. Rev. Lett.* **2006**, *96*, 108101.

76. Bullerjahn, J. T.; Sturm, S.; Kroy, K. Theory of rapid force spectroscopy. *Nat. Commun.* **2014**, *5*, 4463.

77. Bell, G. Models for the specific adhesion of cells to cells. *Science* **1978**, *200*, 618–627.

78. Hummer, G.; Szabo, A. Kinetics from Nonequilibrium Single-Molecule Pulling Experiments. *Biophys. J.* **2003**, *85*, 5–15.

79. Getfert, S.; Reimann, P. Hidden Multiple Bond Effects in Dynamic Force Spectroscopy. *Biophys. J.* **2012**, *102*, 1184–1193.

80. Ritchie, D. B.; Foster, D. A. N.; Woodside, M. T. Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16167–16172.

81. Fugmann, S.; Sokolov, I. M. Scaling of the rupture dynamics of polymer chains pulled at one end at a constant rate. *Phys. Rev. E* **2009**, *79*, 021803.

82. Strackharn, M.; Stahl, S. W.; Puchner, E. M.; Gaub, H. E. Functional Assembly of Aptamer Binding Sites by Single-Molecule Cut-and-Paste. *Nano Lett.* **2012**, *12*, 2425–2428.

83. Scargle, J. D. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **1982**, *263*, 835–853.

84. Puchner, E. M.; Gaub, H. E. Single-Molecule Mechanoenzymatics. *Annu. Rev. Biophys.* **2012**, *41*, 497–518.

85. Douglas, S. M.; Marblestone, A. H.; Teerapittayanon, S.; Vazquez, A.; Church, G. M.; Shih, W. M. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **2009**, *37*, 5001–5006.

86. Sajfutdinow, M.; Uhlig, K.; Prager, A.; Schneider, C.; Abel, B.; Smith, D. M. Nanoscale patterning of self-assembled monolayer (SAM)-functionalised substrates with single molecule contact printing. *Nanoscale* **2017**, *9*, 15098–15106.

87. Stahl, E.; Martin, T. G.; Praetorius, F.; Dietz, H. Facile and Scalable Preparation of Pure and Dense DNA Origami Solutions. *Angew. Chem. Int. Ed.* **2014**, *53*, 12735–12740.

88. Jobst, M. A.; Schoeler, C.; Malinowska, K.; Nash, M. A. Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy. *J. Vis. Exp.* **2013**, 50950.

89. Baumann, F.; Bauer, M. S.; Milles, L. F.; Alexandrovich, A.; Gaub, H. E.; Pippig, D. A. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat. Nanotechnol.* **2015**, *11*, 89.

90. Erlich, K. R.; Baumann, F.; Pippig, D. A.; Gaub, H. E. Strep-Tag II and Monovalent Strep-Tactin as Novel Handles in Single-Molecule Cut-and-Paste. *Small Methods* **2017**, *1*, 1700169.

91. Pippig, D. A.; Baumann, F.; Strackharn, M.; Aschenbrenner, D.; Gaub, H. E. Protein-DNA Chimeras for Nano Assembly. *ACS Nano* **2014**, *8*, 6551–6555.

92. Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. From genes to protein mechanics on a chip. *Nat. Meth.* **2014**, *11*, 1127.

93. Ott, W.; Jobst, M. A.; Bauer, M. S.; Durner, E.; Milles, L. F.; Nash, M. A.; Gaub, H. E. Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano* **2017**, *11*, 6346–6354.

94. Hutter, J. L.; Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **1993**, *64*, 1868–1873.

95. Cook, S. M.; Scäffer, T. E.; Chynoweth, K. M.; Wigton, M.; Simmonds, R. W.; Lang, K. M. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotechnol.* **2006**, *17*, 2135.

96. Russo, J.; Tartaglia, P.; Sciortino, F. Reversible gels of patchy particles: Role of the valence. *J. Chem. Phys.* **2009**, *131*, 014504.

Supporting Information: Force-induced unravelling of DNA Origami

Megan C. Engel,* David M. Smith, Markus A. Jobst, Martin Sajfutdinow, Tim Liedl, Flavio Romano, Lorenzo Rovigatti, Ard A. Louis, and Jonathan P. K. Doye

E-mail: megan.engel@physics.ox.ac.uk

Additional Structural Details

Rothemund tile

In his seminal DNA origami paper, Rothemund presented a rectangular tile folded from M13mp18 DNA.[1] Here, we use a nearly identical structure: a 284-nt wide (Rothemund's original system was 288-nt wide), 24-helix long rectangle with two extended ends, corresponding to elongating sequences inserted into the multiple cloning site (MCS), protruding at a central seam. We also retain a 73-nt region Rothemund excised due to its containing a 20-bp stem hairpin. In the AFM experiments, the MCS sites were 433-nt and 217-nt long and were incorporated into double-stranded handles for attachment to the AFM tip and cover slide; the total handle lengths were 433-nt and 295-nt. For the simulations, since the unfolding of the origami rectangle itself was of interest, these MCS regions were replaced by 10-nucleotide poly-A segments. The folded origami rectangle in our case comprises a 6909-nucleotide scaffold bound by 192 staple strands: 48 30-nt staples and 144 32-nt staples. Most of the staples bind to 3 domains and a few are 1- or 2-domain staples; see Figure S1.

10-helix bundle

In contrast to the Rothemund tile, our second system is based on a hexagonal lattice and is comprised of 10 helices: four 171-nt long, two 175-nt long, two 181-nt long, and two 185-nt long. The 1768-nt scaffold has 51 staples, many of which are 4 or 5 domain (see Figure S2). A key difference between the helix bundle (HB) and the Rothemund tile lies in the relative non-locality of the bundle's staple domains.

7-tile biosensor

The 7-tile biosensor of Koirala *et al.*[3] comprises 14 761 nucleotides arranged as rectangular tiles joined by 44-base pair duplex 'locks'. In Koirala's work, the ends of the system are held in harmonic traps. The sequences of the 'locks' can be specifically designed to bind with aptamers of interest such that their presence in solution is revealed by the opening of one or more locks, manifesting as a sudden increase in extension and decrease in force.

Additional Experimental Methods

Scaffold linearization and functionalization and DNA origami folding

Enzymes and CutSmart-Buffer was purchased from New England BioLabs GmbH (Germany, Frankfurt), modified DNA oligos from Biomers.net GmbH (Germany, Ulm), unmodified oligos from Eurofins Genomics (Germany, Ebersberg) and all other chemicals from Sigma-Aldrich Chemie GmbH (Germany, Munich).

Scaffold '7560'[4] was purified from bacterio-

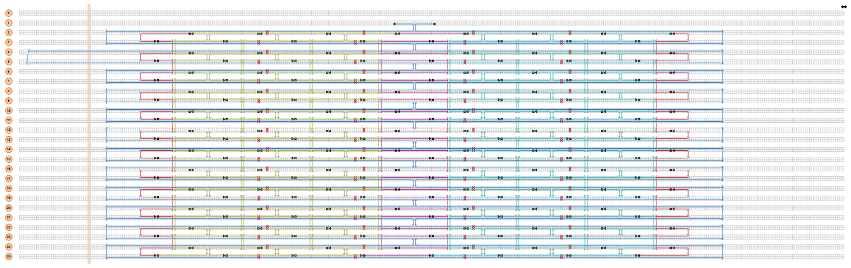


Figure S1: Cadnano structure for the Rothemund tile.

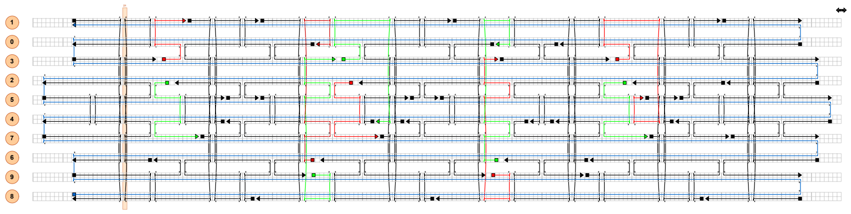


Figure S2: Cadnano structure for the 10-helix bundle of Bae *et al.*[2].

phage M13mp18 according to the protocols in Ref. [1]. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constant Initially, 15 nM of the circular, single-stranded scaffold was hybridized to 50 nM of 'CUT' sequences in 1× CutSmart Buffer to form dsDNA around restriction sites. After mixing the 7560 and CUT strands, the DNA was initially denatured at 80 °C for 5 min, and subsequently cooled to room temperature by a stepwise reduction of temperature of 1 K every 5 seconds. Afterwards, 20 units of BamHI-HF and EcoRI-HF were added to the DNA mix, and the double-digest was carried out for 45 min at 37 °C. The linearized scaffold was then purified from the excised segment by solid phase extraction (NucleoSpin Kit, Macherey & Nagel, Germany, Düren).

The linearized scaffold was ligated to modified oligos carrying either biotin- or thiol- (BIO; THIO) group, to facilitate attachment to either the streptavidin-coated AFM tips or maleimide-functionalized substrate. For this, both ends of the linearized scaffold were hybridized to a 5× excess of entanglement prevention (ENT) sequences, which are expected to enhance the availability of the ends for ligation reactions due to the increased stiffness and exposure of dsDNA from the otherwise highly entangled ssDNA scaffold. A 5× excess of the restriction sites' complement (COMP) strand was also added to the mix in order to create a sticky end for hybridizing to the partially double stranded inserts (INS - added at 10× excess). It should be noted that 3'-THIOL-EcoRI additionally functionalized with 5'-phosphate, as necessary for the ligation reaction. Ligation was performed in CutSmart-Buffer using T4 DNA Ligase and 1 mM ATP for 4 hours at 16 °C, before being heat inactivated at 65 °C for 10 minutes.

This ligated scaffold was either used in its unpurified form (Fig. 2a, green curve) to fold DNA origami rectangle, or was purified from excess functionalization strands by size exclusion (Fig. 2a, red curve) in spin filtration columns (Amicon Ultra-0.5 ml 100kDa MWCO, Merck KGaA, Darmstadt, Germany). The structure and staple sequences were previously reported,[5] and contain a nearly identical staple pattern to

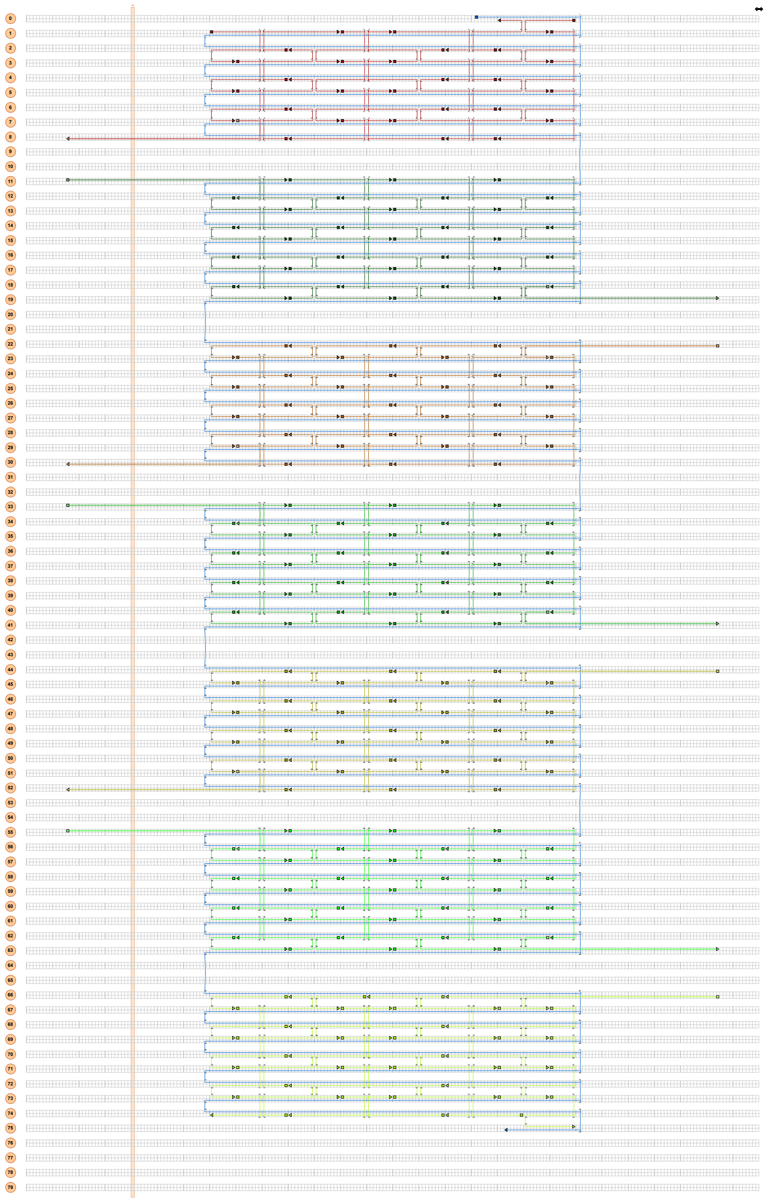


Figure S3: Cadnano structure for the 7-tile assembly of Koirala *et al.*[3].

the original 'Rothemund Rectangle',[1] however with a permutation of the scaffold sequence, and the addition of single-base 'skips' throughout the structure to eliminate global twist of the flexible sheet. For the folding reaction, 10 nM of scaffold was mixed with a 10× excess of staples and initially denatured for 5 minutes at 90 °C, then subsequently hybridized by successively reducing the temperature in steps of 1 K / minute to 20 °C.

The freshly folded DNA origami was separated from excess staple strands and functionalization strands via three successive rounds of PEG precipitation.[6]

Prior to 'unzipping' experiments, the origami pellet was resolved in buffer containing 10 mM Tris, 1 mM EDTA, 10 mM MgCl2 and 2 mM TCEP buffer with pH 8 for 2 hours at ambient temperature to achieve a concentration of ~70 nM. TCEP is required to reduce oxidized thiol-groups and enable maleimide conjugation reaction on surfaces. Afterwards, the resolved origami samples were coupled to AFM sample slides.

AFM Sample Preparation

Glass cover slides were cleaned (piranha etched with sulfuric acid and hydrogen peroxide, 1:1 v/v), then rinsed with deionized water and submerged for one hour in a mixture of 2% (v/v) 3-Aminopropyl dimethyl ethoxysilane (APDMES, abcr GmbH, Karlsruhe, Germany), 88% ethanol (>99% p.a., Carl Roth, Karlsruhe, Germany), and 10% deionized water while shaking. After drying at 80 °C for 30 minutes, the slides display amine groups for subsequent amide bond coupling to the NHS ester groups of alphamaleinimidohexanoic-ω succinimidyl ester poly(ethylene glycol) (NHS-PEG-Mal, Rapp Polymere, Tübingen, Germany) heterobifunctional polymers with a molecular weight of 5000 Da (10 mM, 30 minutes at room temperature in 100 mM HEPES buffer, pH 7.5). After rinsing with water, reduced sulfhydryl groups of the origami were coupled to the free maleimide groups (10 mM Tris,1 mM EDTA, 10 mM $MgCl_2$ and pH 8 for 2 hours at room temperature and then 4 °C overnight) to form permanent thiol-ether bonds.[7,8]

Silicon nitride cantilevers (MLCT, Bruker, Billerica, United States, or Biolever mini, BL-AC40TS-C2, Olympus Corporation, Tokyo, Japan) were cleaned with UV-ozone treatment (UVOH 150 LAB, FHR Anlagenbau GmbH, Ottendorf-Okrilla, Germany), incubated with a mixture of 5uL deionized water, 1 ml ethanol and 1 ml APDMES for 5 minutes at room temperature, and subsequently rinsed with toluene (>99.5% p.a., Carl Roth, Karlsruhe, Germany), ethanol and deionized water, then dried for 30 minutes at 80 °C). After incubation with hetero-bifunctional PEG linkers (10 mM, NHS-PEG-Mal, 30 minutes at room temperature in HEPES buffer, pH 7.5) and thorough rinsing with deionized water, they were coupled to coenzyme A (Merck KGaA, Darmstadt, Germany) at 1 mM in coupling buffer (50 mM sodium phosphate, pH 7.2, 50 mM NaCl, 10 mM EDTA). A monovalent variant of Strep-Tactin (monoST)[9,10] or a monomeric variant of streptavidin as AFM pulling handles were coupled via Sfp phosphopantetheinyl transferase (SFP) with ybbR tags (amino acid sequence DSLEFIASKLA) to the immobilized coenzyme A molecules on the cantilevers (2 mM SFP, 30 mM monoST or mcSA2 for 2 hours at room temperature.[11,12]

Single Molecule Force Spectroscopy Measurement Protocol

Single molecule force spectroscopy (SMFS) experiments were performed in TE buffer supplemented with magnesium (10 mM Tris, 1 mM EDTA, 10 mM $MgCl_2$ and pH 8) at room temperature with custom built instruments (MFP-3D AFM controller, Oxford Instruments Asylum Research, Inc., Santa Barbara, CA, USA; piezo nanopositioners: Physik Instrumente GmbH & Co. KG, Karlsruhe, Germany, or Attocube Systems AG, Munich, Germany). Pulling velocities were set to $1\,\mu m\,s^{-1}$ or $1.6\,\mu m\,s^{-1}$, sampling rate to 12.5 kHz. Cantilever and glass slide positioning was controlled by PID feedback loops on the field-programmable gate arrays during measurement. For analysis, curves with peaks at more than 150 nm extension and more than 30 pN force were sorted automatically, after transformation of raw sensor voltages and correction of zero-force offset and position due to lever bending, similar to previous work.[12,13] Cantilever stiffness calibration was done after the measurement with the thermal method, and piezo calibration was done interferometrically.[14,15]

Correct DNA origami folding from digested scaffold was confirmed by AFM imaging; see Fig. 1 in the main text. Freshly cleaved mica surfaces (Plano GmbH, Wetzlar, Germany) were coated with 100 µl of 0.01% poly-L-ornithine for 5 min and subsequently rinsed with buffer containing 10 mM Tris, 1 mM EDTA, 10 mM MgCl2 and pH 8. Origami samples were immobilized on the surfaces for 10 min. Imaging was carried out with Nanowizard 3 (JPK Instruments AG, Berlin, Germany) in intermittent contact mode in origami buffer using BL-AC40TS cantilevers (Olympus Corporation, Tokyo, Japan).

The stiffness of the AFM cantilever was $\sim$ 110 pN nm^{-1} and the rate of tip retraction 1.6 μm/s for one experiment, and $\sim$ 14.3 pN nm^{-1} and the rate of tip retraction 1.0 μm/s for a second. The experimental force loading rates were 1.8×10^5 pN/s and 1.4×10^4 pN/s.

Additional Simulation Details

Coarse-grained model for DNA
In oxDNA, DNA is treated as a chain of rigid nucleotides, with sites corresponding to the backbone and the base. Hydrogen bonding, stacking, coaxial stacking, cross-stacking, backbone chain connectivity, and excluded volume interactions are captured by potential functions, the details of which can be found elsewhere.[16,17] The model allows for Watson-Crick specific base-pairing, but cannot capture non-canonical base-pairing interactions. Two versions of the model exist: oxDNA2 incorporates major-minor helix grooves and specificity in stacking interactions, which are neglected by the earlier iteration; both versions were parametrized to reproduce experimental results.[18] In this work, we use oxDNA2 for the Rothemund tile and the 7-tile biosensor, and the original oxDNA version for the 10 HB, as oxDNA2 was not released at the time these simulations were undertaken. We expect there to be very little difference between the results for the two oxDNA versions for the current systems, as both capture the thermodynamics of hybridization and the mechanical properties of double-stranded and single-stranded DNA equally well.

Simulations
We ran long molecular dynamics (MD) simulations on GPUs, which offer a speed-up of 25× over CPUs. An Andersen-like thermostat[19] ensured diffusive particle motion in the canonical ensemble, and served to dissipate the energy imparted to the system by external force application. The number of steps between velocity refresh attempts (`newtonian_steps`) was set to 103 and the overall monomer diffusion coefficient resulting from the thermostat (`diff_coeff`) was 2.5. Simulations were performed at the temperatures relevant to experiments: 20 °C for the tile and 36 °C for the helix bundle,[2] and at a high salt concentration typical of origami studies, $[\mathrm{Na}^+] = 0.5\,\mathrm{M}$. The 7-tile assembly was simulated at 30 °C and $[\mathrm{Na}^+] = 0.5\,\mathrm{M}$.

For all systems, harmonic traps of stiffness 5.7 pN/nm were placed on either end of the scaffold, and one was moved relative to the other at a fixed rate. Two rates of trap movement are explored in this work: 5.6×10^8 nm s^{-1} and 5.6×10^7 nm s^{-1}. These parameters yield our reported force loading rates: 3.2×10^9 pN/s and 3.2×10^8 pN/s. Our simulated trap stiffness is the same order of magnitude as stiffnesses typical of optical tweezer SMFS experiments, but is significantly lower than stiffnesses typical of AFM experiments. This lowers our simulated force loading rates relative to the AFM experiments and allows us to approach experimental rates more closely. Further, our parameters were chosen based on optimizing the joint requirements of sufficient resolution of features of interest and feasible simulation time given our computational resources.

The Rothemund tile required roughly 6×10^8 MD steps to unfold at 3.2×10^9 pN/s, or 225 GPU hours, and 6×10^9 MD steps to unfold at 3.2×10^8 pN/s, or 74 days on GPU. Unfolding the 10 HB required roughly 1.4×10^8 MD steps, or 140 GPU hours, at 3.2×10^9 pN/s and 1.3×10^9 MD steps, or 60 days, at 3.2×10^8 pN/s. Breaking all locks in the 7-tile system required $\sim 3.3 \times 10^8$ MD steps at 3.2×10^8 pN/s, or 750 GPU hours; the simulations are significantly slower than for the tile due to the ~ 2000 additional nucleotides.

Sequence dependence did not significantly alter the unfolding behaviour for the tile, and was only incorporated into the tile simulations at 3.2×10^9 pN/s. Simulations of the 10 HB and 7-tile structures did not incorporate sequence dependence.

Forces in oxDNA are applied to the centre of mass of the nucleotide. In the case of harmonic traps, forces are extracted by monitoring the displacement of the scaffold end nucleotides' centres of mass from the harmonic trap centres, then multiplying this by the trap stiffnesses (5.7 pN/nm in all cases). In practice, an effective trap stiffness is computed from the two individual trap stiffnesses k: $\frac{1}{k_{eff}} = \frac{1}{k} + \frac{1}{k}$ or $k_{eff} = k/2$, as the traps are in series. The simulations monitor the projection of the end-to-end distance of the scaffold onto the force axis, which is then subtracted from the distance between the harmonic trap centres. The difference is multiplied by k_{eff} to produce the force-extension curves in the main text.

For a native base pair to be classified as 'present' in generating the blue-yellow plots, the nucleotides must have a hydrogen bonding energy less than -0.1 in simulation units (-4.142×10^{-21} J).

References

1. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **2006**, *440*, 297–302.
2. Bae, W.; Kim, K.; Min, D.; Ryu, J.-K.; Hyeon, C.; Yoon, T.-Y. Programmed folding of DNA origami structures through single-molecule force control. *Nat. Commun.* **2014**, *5*, 5654.
3. Koirala, D.; Shrestha, P.; Emura, T.; Hidaka, K.; Mandal, S.; Endo, M.; Sugiyama, H.; Mao, H. Single-Molecule Mechanochemical Sensing Using DNA Origami Nanostructures. *Angew. Chem. Int. Ed.* **2014**, *126*, 8275–8279.
4. Douglas, S. M.; Marblestone, A. H.; Teerapittayanon, S.; Vazquez, A.; Church, G. M.; Shih, W. M. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **2009**, *37*, 5001–5006.
5. Sajfutdinow, M.; Uhlig, K.; Prager, A.; Schneider, C.; Abel, B.; Smith, D. M. Nanoscale patterning of self-assembled monolayer (SAM)-functionalised substrates with single molecule contact printing. *Nanoscale* **2017**, *9*, 15098–15106.
6. Stahl, E.; Martin, T. G.; Praetorius, F.; Dietz, H. Facile and Scalable Preparation of Pure and Dense DNA Origami Solutions. *Angew. Chem. Int. Ed.* **2014**, *53*, 12735–12740.
7. Jobst, M. A.; Schoeler, C.; Malinowska, K.; Nash, M. A. Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy. *J. Vis. Exp.* **2013**, 50950.
8. Jobst, M. A.; Milles, L. F.; Schoeler, C.; Ott, W.; Fried, D. B.; Bayer, E. A.; Gaub, H. E.; Nash, M. A. Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *eLIFE* **2015**, *4*, 180–183.
9. Baumann, F.; Bauer, M. S.; Milles, L. F.; Alexandrovich, A.; Gaub, H. E.; Pippig, D. A. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat. Nanotechnol.* **2015**, *11*, 89.

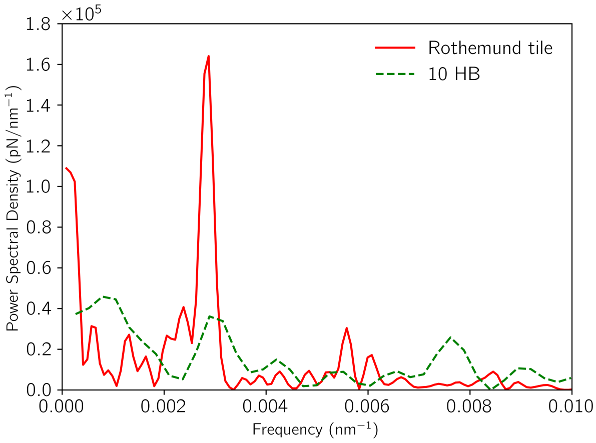


Figure S4: Lomb-Scargle periodogram indicating periodicity in the force-extension curves of the Rothemund tile (red) and the helix bundle (green), both pulled at $\dot{F} = 3 \times 10^8$ pN/s. The red peak near 0 is an artefact of the truncated curve – it would correspond to a very long timescale oscillation where the force returned to 0 at ∼8000nm. The highest peak, at 0.0029 nm^{-1}, corresponds to a length of 345 nm, very close to the length of two rows of the tile: $2\times$ 284 nt $\times$ 0.64 nm = 363 nm, confirming the picture of the tile unfolding two rows at a time. The helix bundle's force-extension curve lacks any clear periodicity.

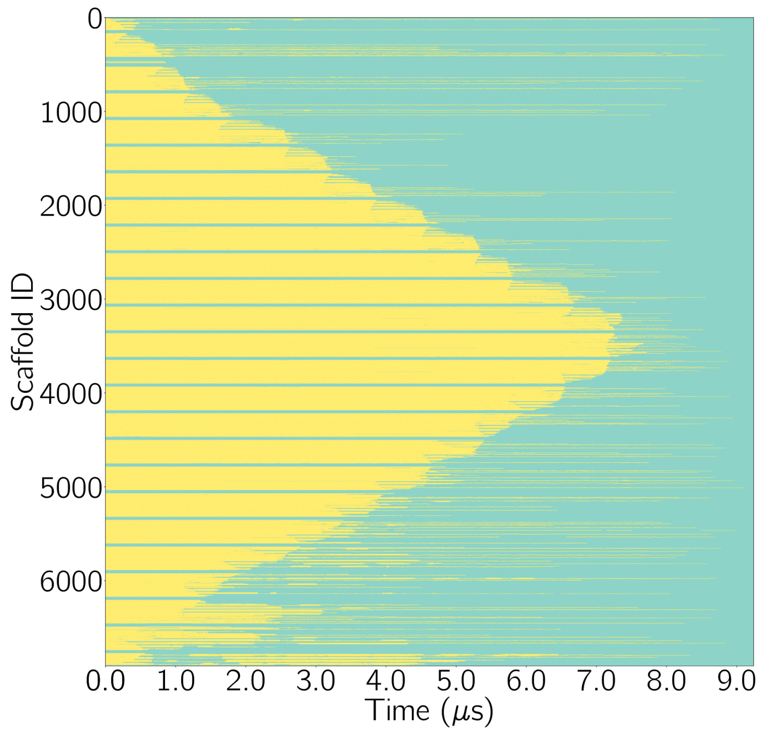


Figure S5: Pattern of native bond unfolding in pulling simulation of the Rothemund tile performed at force loading rate $\dot{F} = 3 \times 10^9$ pN/s, and represented by the red curve in Fig. 2(c) of the main text; the plot for the other run under these conditions is included in the main text.

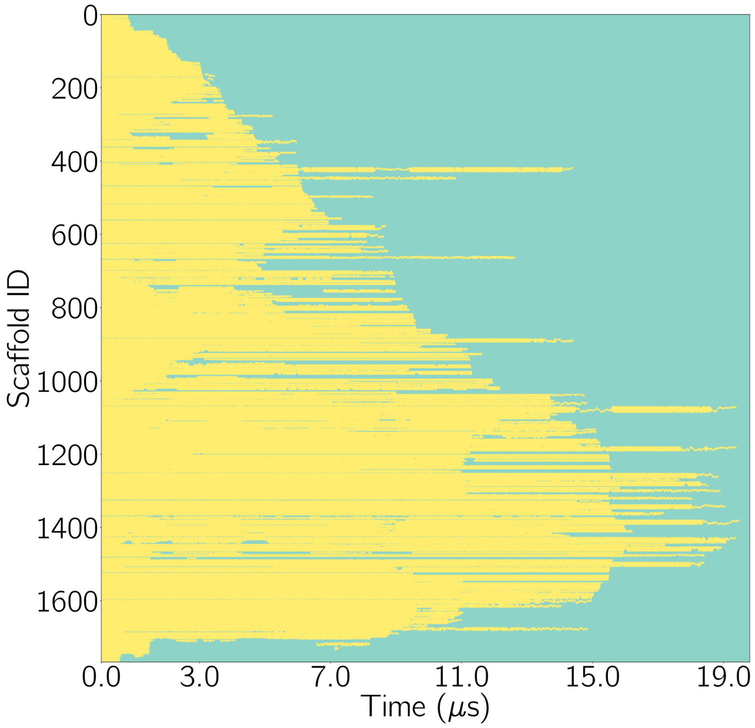


Figure S6: Pattern of native bond unfolding in pulling simulation of the 10-helix bundle performed at force loading rate $\dot{F} = 3 \times 10^8$ pN/s, and represented by the orange curve in Fig. 6(a) of the main text; the plot for the other run under these conditions is included in the main text.

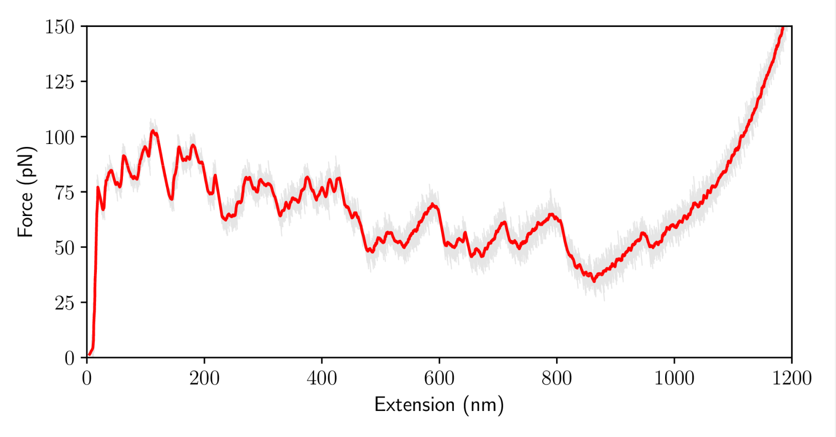


Figure S7: Force-extension curve for pulling simulation of the 10-helix bundle performed at force loading rate $\dot{F} = 6 \times 10^9$ pN/s.

10. Erlich, K. R.; Baumann, F.; Pippig, D. A.; Gaub, H. E. Strep-Tag II and Monovalent Strep-Tactin as Novel Handles in Single-Molecule Cut-and-Paste. *Small Methods* **2017**, *1*, 1700169.
11. Pippig, D. A.; Baumann, F.; Strackharn, M.; Aschenbrenner, D.; Gaub, H. E. Protein-DNA Chimeras for Nano Assembly. *ACS Nano* **2014**, *8*, 6551–6555.
12. Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. From genes to protein mechanics on a chip. *Nat. Meth.* **2014**, *11*, 1127.
13. Ott, W.; Jobst, M. A.; Bauer, M. S.; Durner, E.; Milles, L. F.; Nash, M. A.; Gaub, H. E. Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano* **2017**, *11*, 6346–6354.
14. Hutter, J. L.; Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **1993**, *64*, 1868–1873.
15. Cook, S. M.; Scäffer, T. E.; Chynoweth, K. M.; Wigton, M.; Simmonds, R. W.; Lang, K. M. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotechnol.* **2006**, *17*, 2135.
16. Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, 085101.
17. Snodin, B. E. K.; Randisi, F.; Mosayebi, M.; Šulc, P.; Schreck, J. S.; Romano, F.; Ouldridge, T. E.; Tsukanov, R.; Nir, E.; Louis, A. A. *et al.* Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.* **2015**, *142*, 234901.
18. John SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys.* **2004**, *33*, 415–440.
19. Russo, J.; Tartaglia, P.; Sciortino, F. Reversible gels of patchy particles: Role of the valence. *J. Chem. Phys.* **2009**, *131*, 014504.

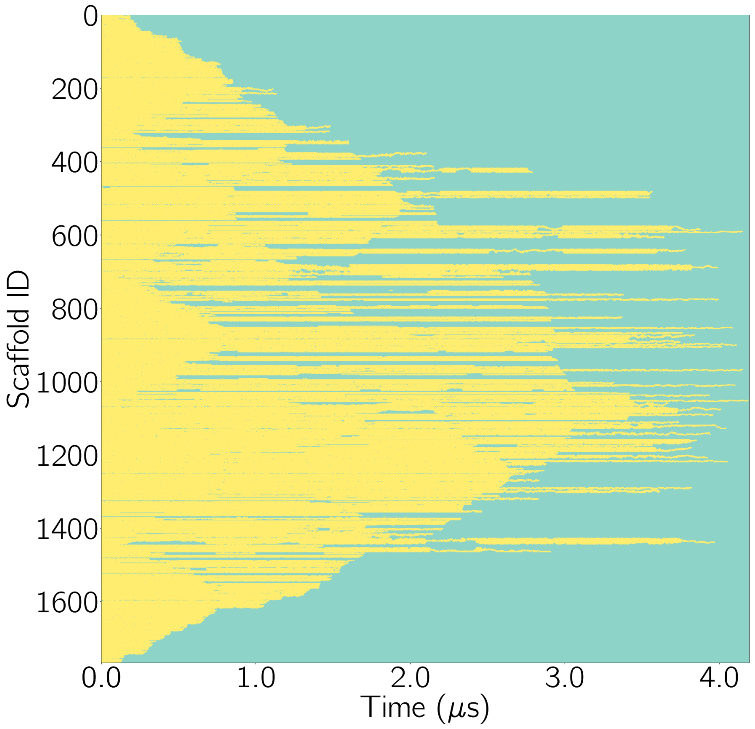


Figure S8: Pattern of native bond unfolding in pulling simulation of the 10-helix bundle performed at force loading rate $\dot{F} = 6 \times 10^9$ pN/s.

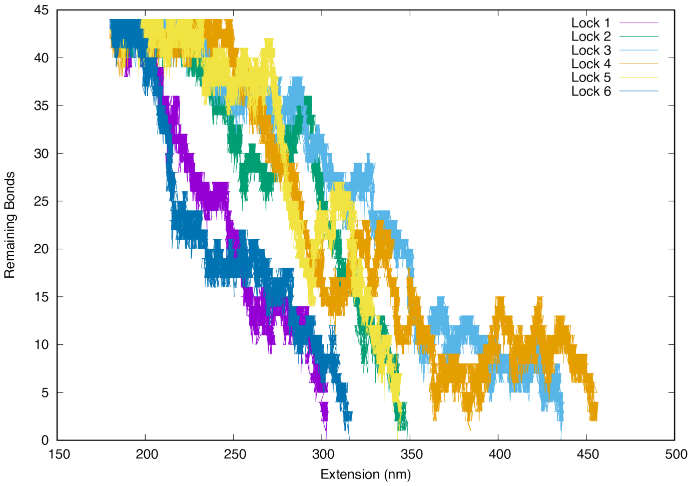


Figure S9: Bond breakage as a function of end-to-end extension for the six 44-base pair aptamer locks in the 7-tile assembly. Lock 1 is located nearest the moving harmonic trap (the uppermost in Fig. 12 of the main text), lock 2 is the next furthest away, and so on to lock 6, which is closest to the stationary harmonic trap. The pulling rate of $\dot{F} = 3 \times 10^8$ pN/s is too rapid to differentiate the breaking of individual locks; sets of two locks yield approximately simultaneously, from the 'outside' (near the traps) to the 'inside'.

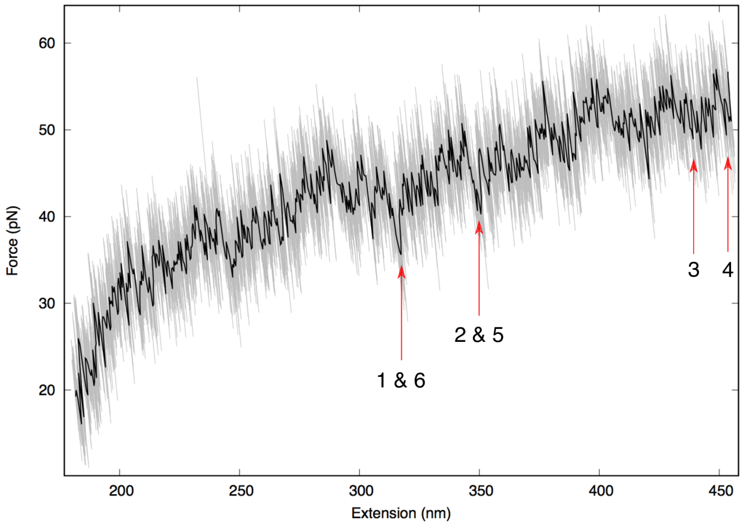


Figure S10: Force-extension curve for the 7-tile assembly for loading rate $\dot{F} = 3 \times 10^8$ pN/s. Light gray curve represents data at a sampling rate of $4 \times 10^8\,\mathrm{s}^{-1}$, and the black curve is an average, representing a sampling rate of $4 \times 10^6\,\mathrm{s}^{-1}$. Aptamer lock breakage events are indicated by the red arrows. Slower pulling rates would enable individual lock-openings to be seen more clearly.

6.2 MANUSCRIPT M2 (in Preparation): Dual Binding Modes and Dynamic Force Spectroscopy on Coh-Doc type 3 (*R.f.*)

This manuscript focuses on origin of the bimodal rupture distribution in the Coh-Doc type 3 (*R.f.*) complex and the evaluation of a potential existence of a dual binding mode. A new instrument was developed to determine a broad dynamic force spectrum of the complex dissociation in force ramp mode. My contribution was partially the instrument development, programming and data analysis.

Wide-Range Dynamic Force Spectroscopy on Cohesin Dockerin

Ellis Durner, Markus A. Jobst, Wolfgang Ott, Hermann E. Gaub

Manuscript in preparation

6.3 MANUSCRIPT M3 (in Preparation): Protein-based Molecular Cut and Paste Assembly

During the work on this project, a protein-only system for the AFM-directed single-molecule cut and paste assembly was developed. Post-translational protein modification, hierarchical fusion domains, and high affinity immobilization allow a DNA-free system to enable investigation of DNA polymerase assays or microfluidic lab-on-a-chip approaches for multiplexed assays. I designed, cloned and expressed the protein system, developed the instrument control routines, and conducted the experiments, and analyzed the data.

Protein-based Molecular Cut and Paste Assembly

Markus A. Jobst, Wolfgang Ott, Ellis Durner, Hermann E. Gaub

Manuscript in preparation

Part III

Appendix

List of Publications

[1] W. Ott*, **M. A. Jobst***, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, and H. E. Gaub, “Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy”, ACS Nano, 2017, DOI: 10.1021/acsnano.7b02694

[2] W. Ott*, **M. A. Jobst***, C. Schoeler*, H. E. Gaub, and M. A. Nash, “Single-molecule force spectroscopy on polyproteins and receptor-ligand complexes: The current toolbox”, Journal of Structural Biology, 2016, DOI: 10.1016/j.jsb.2016.02.011

[3] **M. A. Jobst**, L. F. Milles, C. Schoeler, W. Ott , D. B. Fried, E. A. Bayer, H. E. Gaub, and M. A. Nash, “Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy”, eLife, 2015, DOI: 10.7554/eLife.10319

[4] M. Otten*, W. Ott*, **M. A. Jobst***, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, and H. E. Gaub,“From genes to protein mechanics on a chip”, Nature Methods, 2014, DOI: 10.1038/nmeth.3099

[5] C. Schoeler*, K. H. Malinowska*, R. C. Bernardi, L. F. Milles, **M. A. Jobst**, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, and M. A. Nash, “Ultrastable cellulosome-adhesion complex tightens under load”, Nature Communications, 2014, DOI: 10.1038/ncomms6635

[6] **M. A. Jobst***, C. Schoeler*, K. Malinowska, and M. A. Nash, “Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy”, Journal of Visualized Experiments, 2013, DOI: 10.3791/50950

[7] M. Engel, D. M. Smith, **M. A. Jobst**, M. Sajfutdinow, T. Liedl, F. Romano, L. Rovigatti, A. A. Louis, and J. P. K. Doye, “Force-induced unravelling of DNA Origami”, *submitted for publication*

[8] E. Durner*, **M. A. Jobst***, W. Ott, and H. E. Gaub, “Wide-Range Dynamic Force Spectroscopy on Cohesin Dockerin”, *in preparation*

[9] **M. A. Jobst**, W.Ott, E. Durner, and H. E. Gaub, “Protein-Based Molecular Cut and Paste Assembly”, *in preparation*

---

* These authors contributed equally to this work

List of Figures

Bibliography

[1] J. Berg, J. Tymoczko, and L. Stryer, "Glycogen metabolism", in *Biochemistry, 5th edition*, New York: W H Freeman, 2002, ISBN: 0-7167-3051-0 (cit. on p. 3).

[2] S. G. Ball, "The intricate pathway of starch biosynthesis and degradation in the monocellular alga Chlamydomonas reinhardtii", *Australian Journal of Chemistry*, vol. 55, no. 2, pp. 49–59, 2002. DOI: `10.1071/CH02052` (cit. on p. 4).

[3] J. Fettke, M. Hejazi, J. Smirnova, E. Höchel, M. Stage, and M. Steup, "Eukaryotic starch degradation: Integration of plastidial and cytosolic pathways", *Journal of Experimental Botany*, vol. 60, no. 10, pp. 2907–2922, Jul. 2009. DOI: `10.1093/jxb/erp054` (cit. on p. 4).

[4] H. Kajiura, H. Takata, T. Akiyama, R. Kakutani, T. Furuyashiki, I. Kojima, T. Harui, and T. Kuriki, "In vitro synthesis of glycogen: The structure, properties, and physiological function of enzymatically-synthesized glycogen", *Biologia*, vol. 66, no. 3, pp. 387–394, Jan. 2011. DOI: `10.2478/s11756-011-0053-y` (cit. on p. 4).

[5] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid", *Nature*, vol. 171, no. 4356, pp. 737–738, Apr. 1953. DOI: `10.1038/171737a0`. arXiv: `arXiv:1011.1669v3` (cit. on p. 4).

[6] A. A. Louis, "Contingency, convergence and hyper-astronomical numbers in biological evolution", *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 58, pp. 107–116, Aug. 2016. DOI: `10.1016/j.shpsc.2015.12.014` (cit. on pp. 4, 35).

[7] M. Ptashne, "Gene regulation by proteins acting nearby and at a distance", *Nature*, vol. 322, no. 6081, pp. 697–701, Aug. 1986. DOI: `10.1038/322697a0` (cit. on p. 4).

[8] M. Gutierrez-Arcelus, H. Ongen, T. Lappalainen, S. B. Montgomery, A. Buil, A. Yurovsky, J. Bryois, I. Padioleau, L. Romano, A. Planchon, E. Falconnet, D. Bielser, M. Gagnebin, T. Giger, C. Borel, A. Letourneau, P. Makrythanasis, M. Guipponi, C. Gehrig, S. E. Antonarakis, and E. T. Dermitzakis, "Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing", *PLOS Genetics*, vol. 11, no. 1, C. D. Brown, Ed., e1004958, Jan. 2015. DOI: `10.1371/journal.pgen.1004958` (cit. on p. 4).

[9] C. D. Allis and T. Jenuwein, "The molecular hallmarks of epigenetic control", *Nature Reviews Genetics*, vol. 17, no. 8, pp. 487–500, Aug. 2016. DOI: `10.1038/nrg.2016.59` (cit. on p. 4).

[10] S. M. Soucy, J. Huang, and J. P. Gogarten, "Horizontal gene transfer: Building the web of life", *Nature Reviews Genetics*, vol. 16, no. 8, pp. 472–482, Aug. 2015. DOI: `10.1038/nrg3962` (cit. on pp. 4, 7).

[11] A. Ambrogelly, S. Palioura, and D. Söll, "Natural expansion of the genetic code", *Nature Chemical Biology*, vol. 3, no. 1, pp. 29–35, Jan. 2007. DOI: `10.1038/nchembio847` (cit. on p. 4).

[12] A. Gutteridge and J. M. Thornton, “Understanding nature’s catalytic toolkit”, *Trends in Biochemical Sciences*, vol. 30, no. 11, pp. 622–629, Nov. 2005. DOI: `10.1016/j.tibs.2005.09.006` (cit. on p. 4).

[13] C. M. Dobson, A. Šali, and M. Karplus, “Protein folding: A perspective from theory and experiment”, *Angewandte Chemie International Edition*, vol. 37, no. 7, pp. 868–893, Apr. 1998. DOI: `{10.1002/(SICI)1521-3773(19980420)37:7<868::AID-ANIE868>3.0.CO;2-H}` (cit. on p. 4).

[14] S. Radford, “Protein folding: Progress made and promises ahead”, *Trends in Biochemical Sciences*, vol. 25, no. 12, pp. 611–618, Dec. 2000. DOI: `10.1016/S0968-0004(00)01707-2` (cit. on p. 4).

[15] S. W. Englander and L. Mayne, “The nature of protein folding pathways”, *Proceedings of the National Academy of Sciences*, vol. 111, no. 45, pp. 15 873–15 880, Nov. 2014. DOI: `10.1073/pnas.1411798111` (cit. on p. 4).

[16] F. H. Arnold and A. A. Volkov, “Directed evolution of biocatalysts”, *Current Opinion in Chemical Biology*, vol. 3, no. 1, pp. 54–59, Feb. 1999, ISSN: 13675931. DOI: `10.1016/S1367-5931(99)80010-6` (cit. on p. 4).

[17] C. Jäckel, P. Kast, and D. Hilvert, “Protein design by directed evolution”, *Annual Review of Biophysics*, vol. 37, no. 1, pp. 153–173, Jun. 2008. DOI: `10.1146/annurev.biophys.37.032807.125832` (cit. on p. 4).

[18] M. S. Packer and D. R. Liu, “Methods for the directed evolution of proteins”, *Nature Reviews Genetics*, vol. 16, no. 7, pp. 379–394, Jul. 2015. DOI: `10.1038/nrg3927`. arXiv: `arXiv:1011.1669v3` (cit. on p. 4).

[19] P.-S. Huang, S. E. Boyken, and D. Baker, “The coming of age of de novo protein design”, *Nature*, vol. 537, no. 7620, pp. 320–327, Sep. 2016. DOI: `10.1038/nature19946` (cit. on p. 4).

[20] K. Bavishi and N. S. Hatzakis, “Shedding light on protein folding, structural and functional dynamics by single molecule studies”, *Molecules*, vol. 19, no. 12, pp. 19 407–19 434, 2014. DOI: `10.3390/molecules191219407` (cit. on p. 4).

[21] H. Miller, Z. Zhou, J. Shepherd, A. J. M. Wollman, and M. C. Leake, “Single-molecule techniques in biophysics: A review of the progress in methods and applications”, *Reports on Progress in Physics*, vol. 81, no. 2, p. 024 601, Feb. 2018. DOI: `10.1088/1361-6633/aa8a02` (cit. on p. 4).

[22] M. Ashwini, S. B. Murugan, S. Balamurugan, and R. Sathishkumar, “Advances in molecular cloning”, *Molecular Biology*, vol. 50, no. 1, pp. 1–6, Jan. 2016. DOI: `10.1134/S0026893316010131` (cit. on p. 4).

[23] H. Khorana, K. Agarwal, H. Büchi, M. Caruthers, N. Gupta, K. Klbppe, A. Kumar, E. Ohtsuka, U. RajBhandary, J. van de Sande, V. Sgaramella, T. Tebao, H. Weber, and T. Yamada, “Ciii. total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast”, *Journal of Molecular Biology*, vol. 72, no. 2, pp. 209–217, Dec. 1972. DOI: `https://doi.org/10.1016/0022-2836(72)90146-5` (cit. on p. 5).

[24] K. Itakura, T. Hirose, R. Crea, A. Riggs, H. Heyneker, F. Bolivar, and H. Boyer, "Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin", *Science*, vol. 198, no. 4321, pp. 1056–1063, Dec. 1977. DOI: `10.1126/science.412251` (cit. on p. 5).

[25] M. D. Edge, A. R. Greene, G. R. Heathcliffe, P. A. Meacock, W. Schuch, D. B. Scanlon, T. C. Atkinson, C. R. Newton, and A. F. Markham, "Total synthesis of a human leukocyte interferon gene", *Nature*, vol. 292, no. 5825, pp. 756–762, Aug. 1981. DOI: `10.1038/292756a0` (cit. on p. 5).

[26] S. Roy and M. Caruthers, "Synthesis of DNA/RNA and their analogs via phosphoramidite and h-phosphonate chemistries", *Molecules*, vol. 18, no. 11, pp. 14 268–14 284, Nov. 2013. DOI: `10.3390/molecules181114268` (cit. on p. 5).

[27] S. Kosuri and G. M. Church, "Large-scale de novo DNA synthesis: technologies and applications", *Nature Methods*, vol. 11, no. 5, pp. 499–507, May 2014. DOI: `10.1038/nmeth.2918` (cit. on p. 5).

[28] L.-C. Au, F.-Y. Yang, W.-J. Yang, S.-H. Lo, and C.-F. Kao, "Gene synthesis by a LCR-based approach: High-level production of Leptin-L54 using synthetic gene in Escherichia coli", *Biochemical and Biophysical Research Communications*, vol. 248, no. 1, pp. 200–203, Jul. 1998. DOI: `10.1006/bbrc.1998.8929` (cit. on p. 6).

[29] W. P. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, and H. L. Heyneker, "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides", *Gene*, vol. 164, no. 1, pp. 49–53, Oct. 1995. DOI: `10.1016/0378-1119(95)00511-4` (cit. on p. 6).

[30] R. Saiki, S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich, and N. Arnheim, "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia", *Science*, vol. 230, no. 4732, pp. 1350–1354, Dec. 1985. DOI: `10.1126/science.2999980` (cit. on p. 6).

[31] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich, "Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction", *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 51, no. Table 1, pp. 263–273, Jan. 1986. DOI: `10.1101/SQB.1986.051.01.032` (cit. on p. 6).

[32] K. B. Mullis and F. A. Faloona, "Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction", in *Methods in Enzymology*, C, vol. 155, 1987, pp. 335–350. DOI: `10.1016/0076-6879(87)55023-6` (cit. on p. 6).

[33] R. Saiki, D. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. Horn, K. Mullis, and H. Erlich, "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase", *Science*, vol. 239, no. 4839, pp. 487–491, Jan. 1988. DOI: `10.1126/science.2448875` (cit. on p. 6).

[34] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison III, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases", *Nature Methods*, vol. 6, no. 5, pp. 343–345, May 2009. DOI: `10.1038/nmeth.1318` (cit. on p. 6).

[35] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith, and J. C. Venter, "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome", *Science*, vol. 329, no. 5987, pp. 52–56, Jul. 2010. DOI: `10.1126/science.1190719` (cit. on p. 6).

[36] C. Engler, R. Kandzia, and S. Marillonnet, "A one pot, one step, precision cloning method with high throughput capability", *PLoS One*, vol. 3, no. 11, e3647, 2008. DOI: `10.1371/journal.pone.0003647` (cit. on p. 7).

[37] C. Engler, R. Gruetzner, R. Kandzia, and S. Marillonnet, "Golden gate shuffling: A one-pot DNA shuffling method based on type IIs restriction enzymes", *PLoS ONE*, vol. 4, no. 5, J. Peccoud, Ed., e5553, May 2009. DOI: `10.1371/journal.pone.0005553` (cit. on p. 7).

[38] S. N. Cohen, A. C. Y. Chang, H. W. Boyer, and R. B. Helling, "Construction of biologically functional bacterial plasmids in vitro", *Proceedings of the National Academy of Sciences*, vol. 70, no. 11, pp. 3240–3244, Nov. 1973. DOI: `10.1073/pnas.70.11.3240` (cit. on p. 7).

[39] D. Nathans and H. O. Smith, "Restriction endonucleases in the analysis and restructuring of DNA molecules", *Annual Review of Biochemistry*, vol. 44, no. 1, pp. 273–293, Jun. 1975. DOI: `10.1146/annurev.bi.44.070175.001421` (cit. on p. 7).

[40] A. Casini, M. Storch, G. S. Baldwin, and T. Ellis, "Bricks and blueprints: Methods and standards for DNA assembly", *Nature Reviews Molecular Cell Biology*, vol. 16, no. 9, pp. 568–576, Sep. 2015. DOI: `10.1038/nrm4014` (cit. on p. 7).

[41] D. Hanahan, "Studies on transformation of Escherichia coli with plasmids", *Journal of Molecular Biology*, vol. 166, no. 4, pp. 557–580, Jun. 1983. DOI: `10.1016/S0022-2836(83)80284-8` (cit. on p. 7).

[42] A. Taketo, "DNA transfection of Escherichia coli by electroporation", *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, vol. 949, no. 3, pp. 318–324, Mar. 1988. DOI: `10.1016/0167-4781(88)90158-3` (cit. on p. 7).

[43] W. J. Dower, J. F. Miller, and C. W. Ragsdale, "High efficiency transformation of E.coli by high voltage electroporation", *Nucleic Acids Research*, vol. 16, no. 13, Sep. 1988 (cit. on p. 7).

[44] F. Griffith, "The significance of pneumococcal types", *Journal of Hygiene*, vol. 27, no. 02, pp. 113–159, Jan. 1928. DOI: `10.1017/S0022172400031879` (cit. on p. 7).

[45] O. T. Avery, C. M. MacLeod, and M. McCarty, "Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III", *Journal of Experimental Medicine*, vol. 79, no. 2, pp. 137–158, Feb. 1944. DOI: `10.1084/jem.79.2.137` (cit. on p. 7).

[46] S. Lin-Chao and H. Bremer, “Effect of the bacterial growth rate on replication control of plasmid pBR322 in Escherichia coli”, *MGG Molecular & General Genetics*, vol. 203, no. 1, pp. 143–149, Apr. 1986. DOI: `10.1007/BF00330395` (cit. on p. 8).

[47] C. Yanisch-Perron, J. Vieira, and J. Messing, “Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mpl8 and pUC19 vectors”, *Gene*, vol. 33, no. 1, pp. 103–119, Jan. 1985. DOI: `10.1016/0378-1119(85)90120-9` (cit. on p. 8).

[48] M. Rossi, P. Brigidi, A. Gonzalez Vara y Rodriguez, and D. Matteuzzi, “Characterization of the plasmid pMB1 from Bifidobacterium longum and its use for shuttle vector construction”, *Research in Microbiology*, vol. 147, no. 3, pp. 133–143, Jan. 1996. DOI: `10.1016/0923-2508(96)80213-0` (cit. on p. 8).

[49] F. W. Studier and B. A. Moffatt, “Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes”, *Journal of Molecular Biology*, vol. 189, no. 1, pp. 113–130, May 1986. DOI: `10.1016/0022-2836(86)90385-2` (cit. on pp. 8, 11).

[50] P. Daegelen, F. W. Studier, R. E. Lenski, S. Cure, and J. F. Kim, “Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3)”, *Journal of Molecular Biology*, vol. 394, no. 4, pp. 634–643, Dec. 2009. DOI: `10.1016/j.jmb.2009.09.022` (cit. on p. 8).

[51] J. Shine and L. Dalgarno, “Determinant of cistron specificity in bacterial ribosomes”, *Nature*, vol. 254, no. 5495, pp. 34–38, Mar. 1975. DOI: `10.1038/254034a0` (cit. on p. 8).

[52] N. Malys, “Shine-dalgarno sequence of bacteriophage T4: GAGG prevails in early genes”, *Molecular Biology Reports*, vol. 39, no. 1, pp. 33–39, Jan. 2012. DOI: `10.1007/s11033-011-0707-4` (cit. on p. 8).

[53] F. R. Blattner, “The complete genome sequence of Escherichia coli K-12”, *Science*, vol. 277, no. 5331, pp. 1453–1462, Sep. 1997. DOI: `10.1126/science.277.5331.1453` (cit. on p. 9).

[54] R. H. Epstein, A. Bolle, and C. M. Steinberg, “Amber mutants of bacteriophage T4D: Their isolation and genetic characterization”, *Genetics*, vol. 190, no. 3, pp. 831–840, Mar. 2012. DOI: `10.1534/genetics.112.138438` (cit. on p. 9).

[55] S. Brenner, A. O. W. Stretton, and S. Kaplan, “Genetic code: The ‘nonsense’ triplets for chain termination and their suppression”, *Nature*, vol. 206, no. 4988, pp. 994–998, Jun. 1965. DOI: `10.1038/206994a0` (cit. on p. 9).

[56] S. Brenner, L. Barnett, E. R. Katz, and F. H. C. Crick, “UGA: A third nonsense triplet in the genetic code”, *Nature*, vol. 213, no. 5075, pp. 449–450, Feb. 1967. DOI: `10.1038/213449a0` (cit. on p. 9).

[57] R. Yang, Y. H. Wong, G. K. T. Nguyen, J. P. Tam, J. Lescar, and B. Wu, “Engineering a catalytically efficient recombinant protein ligase”, *Journal of the American Chemical Society*, vol. 139, no. 15, pp. 5351–5358, Mar. 2017. DOI: `10.1021/jacs.6b12637` (cit. on p. 9).

[58] P. H. Hirel, M. J. Schmitter, P. Dessen, G. Fayat, and S. Blanquet, "Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid.", *Proceedings of the National Academy of Sciences*, vol. 86, no. 21, pp. 8247–8251, Nov. 1989. DOI: `10.1073/pnas.86.21.8247` (cit. on p. 9).

[59] Y.-D. Liao, J.-C. Jeng, C.-F. Wang, S.-C. Wang, and S.-T. Chang, "Removal of N-terminal methionine from recombinant proteins by engineered E. coli methionine aminopeptidase", *Protein Science*, vol. 13, no. 7, pp. 1802–1810, Jul. 2004. DOI: `10.1110/ps.04679104` (cit. on p. 9).

[60] F. Frottin, A. Martinez, P. Peynot, S. Mitra, R. C. Holz, C. Giglione, and T. Meinnel, "The proteomics of N-terminal methionine cleavage", *Molecular & Cellular Proteomics*, vol. 5, no. 12, pp. 2336–2349, Dec. 2006. DOI: `10.1074/mcp.M600225-MCP200` (cit. on p. 9).

[61] A. J. Griffiths, W. M. Gelbart, J. H. Miller, and R. C. Lewontin, "Regulation of the Lactose System", in *Modern Genetic Analysis*, New York: W. H. Freeman, 1999, ch. Regulation of Gene Transcription (cit. on p. 10).

[62] T. Malan, A. Kolb, H. Buc, and W. R. McClure, "Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter", *Journal of Molecular Biology*, vol. 180, no. 4, pp. 881–909, Dec. 1984. DOI: `10.1016/0022-2836(84)90262-6` (cit. on p. 10).

[63] A. Vazquez, Q. K. Beg, M. A. DeMenezes, J. Ernst, Z. Bar-Joseph, A.-L. Barabási, L. G. Boros, and Z. N. Oltvai, "Impact of the solvent capacity constraint on E. coli metabolism", *BMC Systems Biology*, vol. 2, no. 1, p. 7, 2008. DOI: `10.1186/1752-0509-2-7` (cit. on p. 10).

[64] F. W. Studier, "Protein production by auto-induction in high-density shaking cultures", *Protein Expression and Purification*, vol. 41, no. 1, pp. 207–234, Mar. 2005. DOI: `10.1016/j.pep.2005.01.016` (cit. on p. 10).

[65] G. Beaven and E. Holiday, "Ultraviolet absorption spectra of proteins and amino acids", in *Adv Protein Chem*, London, 1952, pp. 319–386. DOI: `10.1016/S0065-3233(08)60022-4` (cit. on p. 11).

[66] S. C. Gill and P. H. von Hippel, "Calculation of protein extinction coefficients from amino acid sequence data", *Analytical Biochemistry*, vol. 182, no. 2, pp. 319–326, Nov. 1989. DOI: `10.1016/0003-2697(89)90602-7` (cit. on p. 11).

[67] R. Scopes, "Measurement of protein by spectrophotometry at 205 nm", *Analytical Biochemistry*, vol. 59, no. 1, pp. 277–282, May 1974. DOI: `10.1016/0003-2697(74)90034-7` (cit. on p. 11).

[68] A. H. Rosenberg, B. N. Lade, C. Dao-shan, S.-W. Lin, J. J. Dunn, and F. Studier, "Vectors for selective expression of cloned DNAs by T7 RNA polymerase", *Gene*, vol. 56, no. 1, pp. 125–135, Jan. 1987. DOI: `10.1016/0378-1119(87)90165-X` (cit. on p. 11).

[69] F. W. Studier, P. Daegelen, R. E. Lenski, S. Maslov, and J. F. Kim, “Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes”, *Journal of Molecular Biology*, vol. 394, no. 4, pp. 653–680, Dec. 2009. DOI: `10.1016/j.jmb.2009.09.021` (cit. on p. 11).

[70] T. Makino, G. Skretas, and G. Georgiou, “Strain engineering for improved expression of recombinant proteins in bacteria”, *Microbial Cell Factories*, vol. 10, no. 1, p. 32, 2011. DOI: `10.1186/1475-2859-10-32` (cit. on p. 11).

[71] A. M. Baca and W. G. Hol, “Overcoming codon bias: A method for high-level overexpression of plasmodium and other AT-rich parasite genes in Escherichia coli”, *International Journal for Parasitology*, vol. 30, no. 2, pp. 113–118, Feb. 2000. DOI: `10.1016/S0020-7519(00)00019-9` (cit. on p. 11).

[72] H. P. Sørensen, H. U. Sperling-Petersen, and K. K. Mortensen, “Production of recombinant thermostable proteins expressed in Escherichia coli: Completion of protein synthesis is the bottleneck”, *Journal of Chromatography B*, vol. 786, no. 1-2, pp. 207–214, Mar. 2003. DOI: `10.1016/S1570-0232(02)00689-X` (cit. on p. 11).

[73] P. H. Bessette, F. Aslund, J. Beckwith, and G. Georgiou, “Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm”, *Proceedings of the National Academy of Sciences*, vol. 96, no. 24, pp. 13 703–13 708, 1999. DOI: `10.1073/pnas.96.24.13703` (cit. on p. 11).

[74] J. Lobstein, C. A. Emrich, C. Jeans, M. Faulkner, P. Riggs, and M. Berkmen, “SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm”, *Microbial Cell Factories*, vol. 11, no. 1, p. 753, Dec. 2012. DOI: `10.1186/1475-2859-11-56` (cit. on p. 11).

[75] S. S. Poletto, I. O. da Fonseca, L. P. de Carvalho, L. A. Basso, and D. S. Santos, “Selection of an Escherichia coli host that expresses mutant forms of mycobacterium tuberculosis 2-trans enoyl-ACP(CoA) reductase and 3-ketoacyl-ACP(CoA) reductase enzymes”, *Protein Expression and Purification*, vol. 34, no. 1, pp. 118–125, Mar. 2004. DOI: `10.1016/j.pep.2003.10.009` (cit. on p. 11).

[76] F. Saida, M. Uzan, B. Odaert, and F. Bontems, “Expression of highly toxic genes in E. coli: Special strategies and genetic tools”, *Current Protein and Peptide Science*, vol. 7, no. 1, pp. 47–56, Feb. 2006. DOI: `10.2174/138920306775474095` (cit. on p. 11).

[77] M. Wacker, “N-linked glycosylation in campylobacter jejuni and its functional transfer into E. coli”, *Science*, vol. 298, no. 5599, pp. 1790–1793, Nov. 2002. DOI: `10.1126/science.298.5599.1790` (cit. on p. 11).

[78] J. Pandhal, S. Y. Ow, J. Noirel, and P. C. Wright, “Improving N-glycosylation efficiency in Escherichia coli using shotgun proteomics, metabolic network analysis, and selective reaction monitoring”, *Biotechnology and Bioengineering*, vol. 108, no. 4, pp. 902–912, Apr. 2011. DOI: `10.1002/bit.23011` (cit. on p. 11).

[79] F. Katzen, G. Chang, and W. Kudlicki, “The past, present and future of cell-free protein synthesis”, *Trends in Biotechnology*, vol. 23, no. 3, pp. 150–156, Mar. 2005. DOI: `10.1016/j.tibtech.2005.01.003` (cit. on p. 12).

[80] G. Rosenblum and B. S. Cooperman, “Engine out of the chassis: Cell-free protein synthesis and its uses”, *FEBS Letters*, vol. 588, no. 2, pp. 261–268, Jan. 2014. DOI: `10.1016/j.febslet.2013.10.016` (cit. on p. 12).

[81] A. Zemella, L. Thoring, C. Hoffmeister, and S. Kubick, “Cell-free protein synthesis: Pros and cons of prokaryotic and eukaryotic systems”, *ChemBioChem*, vol. 16, no. 17, pp. 2420–2431, Nov. 2015. DOI: `10.1002/cbic.201500340` (cit. on p. 12).

[82] T. Verdorfer, R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash, and H. E. Gaub, “Combining in vitro and in silico Single-Molecule force spectroscopy to characterize and tune cellulosomal scaffoldin mechanics”, *J. Am. Chem. Soc.*, 2017. DOI: `10.1021/jacs.7b07574` (cit. on pp. 12, 33).

[83] K. Svoboda and S. M. Block, “Biological applications of optical forces”, *Annual Review of Biophysics and Biomolecular Structure*, vol. 23, no. 1, pp. 247–285, Jun. 1994. DOI: `10.1146/annurev.bb.23.060194.001335` (cit. on p. 12).

[84] K. C. Neuman and S. M. Block, “Optical trapping”, *Review of Scientific Instruments*, vol. 75, no. 9, pp. 2787–2809, Sep. 2004. DOI: `10.1063/1.1785844` (cit. on p. 12).

[85] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, “Magnetic torque tweezers: Measuring torsional stiffness in DNA and RecA-DNA filaments”, *Nature Methods*, vol. 7, no. 12, pp. 977–980, Oct. 2010. DOI: `10.1038/nmeth.1520` (cit. on p. 12).

[86] X. J. A. Janssen, J. Lipfert, T. Jager, R. Daudey, J. Beekman, and N. H. Dekker, “Electromagnetic torque tweezers: A versatile approach for measurement of single-molecule twist and torque”, *Nano Letters*, vol. 12, no. 7, pp. 3634–3639, Jun. 2012. DOI: `10.1021/nl301330h` (cit. on p. 12).

[87] S. K. Nomidis, F. Kriegel, W. Vanderlinden, J. Lipfert, and E. Carlon, “Twist-bend coupling and the torsional response of double-stranded dna”, *Phys. Rev. Lett*, vol. 118, p. 217 801, 21 May 2017. DOI: `10.1103/PhysRevLett.118.217801` (cit. on p. 12).

[88] G. Sitters, D. Kamsma, G. Thalhammer, M. Ritsch-Marte, E. J. G. Peterman, and G. J. L. Wuite, “Acoustic force spectroscopy”, *Nature Methods*, vol. 12, no. 1, pp. 47–50, Nov. 2014. DOI: `10.1038/nmeth.3183` (cit. on p. 13).

[89] K. Halvorsen and W. P. Wong, “Massively parallel single-molecule manipulation using centrifugal force”, *Biophysical Journal*, vol. 98, no. 11, pp. L53–L55, Jun. 2010. DOI: `10.1016/j.bpj.2010.03.012` (cit. on p. 13).

[90] D. Yang, A. Ward, K. Halvorsen, and W. P. Wong, “Multiplexed single-molecule force spectroscopy using a centrifuge”, *Nature Communications*, vol. 7, p. 11 026, Mar. 2016. DOI: `10.1038/ncomms11026` (cit. on p. 13).

[91] T. Hoang, D. S. Patel, and K. Halvorsen, “A wireless centrifuge force microscope (CFM) enables multiplexed single-molecule experiments in a commercial centrifuge”, *Review of Scientific Instruments*, vol. 87, no. 8, p. 083 705, Aug. 2016. DOI: `10.1063/1.4961477` (cit. on p. 13).

[92] G. Binnig, C. F. Quate, and C. Gerber, “Atomic force microscope”, *Physical Review Letters*, vol. 56, no. 9, pp. 930–933, Mar. 1986. doi: `10.1103/physrevlett.56.930` (cit. on p. 13).

[93] P. Hansma, V. Elings, O. Marti, and C. Bracker, “Scanning tunneling microscopy and atomic force microscopy: Application to biology and technology”, *Science*, vol. 242, no. 4876, pp. 209–216, Oct. 1988. doi: `10.1126/science.3051380` (cit. on p. 13).

[94] M. Radmacher, R. Tillamnn, M. Fritz, and H. Gaub, “From molecules to cells: Imaging soft samples with the atomic force microscope”, *Science*, vol. 257, no. 5078, pp. 1900–1905, Sep. 1992. doi: `10.1126/science.1411505` (cit. on p. 13).

[95] M. Radmacher, M. Fritz, H. Hansma, and P. Hansma, “Direct observation of enzyme activity with the atomic force microscope”, *Science*, vol. 265, no. 5178, pp. 1577–1579, Sep. 1994. doi: `10.1126/science.8079171` (cit. on p. 13).

[96] E. Florin, V. Moy, and H. Gaub, “Adhesion forces between individual ligand-receptor pairs”, *Science*, vol. 264, no. 5157, pp. 415–417, Apr. 1994. doi: `10.1126/science.8153628` (cit. on pp. 13, 29).

[97] S. B. Smith, Y. Cui, and C. Bustamante, “Overstretching b-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules”, *Science*, vol. 271, no. 5250, pp. 795–799, Feb. 1996. doi: `10.1126/science.271.5250.795` (cit. on p. 13).

[98] M. Rief, “Reversible unfolding of individual titin immunoglobulin domains by AFM”, *Science*, vol. 276, no. 5315, pp. 1109–1112, May 1997. doi: `10.1126/science.276.5315.1109` (cit. on p. 13).

[99] S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl, and H. E. Gaub, “Single-molecule cut-and-paste surface assembly”, *Science*, vol. 319, no. 5863, pp. 594–596, Feb. 2008. doi: `10.1126/science.1151424` (cit. on pp. 13, 35).

[100] S. K. Kufer, M. Strackharn, S. W. Stahl, H. Gumpp, E. M. Puchner, and H. E. Gaub, “Optically monitoring the mechanical assembly of single molecules”, *Nature Nanotechnology*, vol. 4, no. 1, pp. 45–49, Nov. 2008. doi: `10.1038/nnano.2008.333` (cit. on pp. 13, 36).

[101] P. C. Nickels, B. Wünsch, P. Holzmeister, W. Bae, L. M. Kneer, D. Grohmann, P. Tinnefeld, and T. Liedl, “Molecular force spectroscopy with a DNA origami–based nanoscopic force clamp”, *Science*, vol. 354, no. 6310, pp. 305–307, Oct. 2016. doi: `10.1126/science.aah5974` (cit. on p. 13).

[102] C. Grashoff, B. D. Hoffman, M. D. Brenner, R. Zhou, M. Parsons, M. T. Yang, M. A. McLean, S. G. Sligar, C. S. Chen, T. Ha, and M. A. Schwartz, “Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics”, *Nature*, vol. 466, no. 7303, pp. 263–266, Jul. 2010. doi: `10.1038/nature09198` (cit. on p. 13).

[103] L. S. Wong, J. Thirlway, and J. Micklefield, “Direct site-selective covalent protein immobilization catalyzed by a phosphopantetheinyl transferase”, *Journal of the American Chemical Society*, vol. 130, no. 37, pp. 12 456–12 464, Sep. 2008. doi: `10.1021/ja8030278` (cit. on p. 13).

[104] J. L. Zimmermann, T. Nicolaus, G. Neuert, and K. Blank, "Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments", *Nat. Protoc.*, vol. 5, no. 6, pp. 975–985, 2010. DOI: `10.1038/nprot.2010.49` (cit. on pp. 13, 21).

[105] N. Hino, A. Hayashi, K. Sakamoto, and S. Yokoyama, "Site-specific incorporation of non-natural amino acids into proteins in mammalian cells with an expanded genetic code", *Nature Protocols*, vol. 1, no. 6, pp. 2957–2962, Jan. 2007. DOI: `10.1038/nprot.2006.424` (cit. on p. 13).

[106] A. J. de Graaf, M. Kooijman, W. E. Hennink, and E. Mastrobattista, "Nonnatural amino acids for site-specific protein conjugation", *Bioconjugate Chemistry*, vol. 20, no. 7, pp. 1281–1295, Jul. 2009. DOI: `10.1021/bc800294a` (cit. on p. 13).

[107] E. Durner, W. Ott, M. A. Nash, and H. E. Gaub, "Post-Translational Sortase-Mediated attachment of High-Strength force spectroscopy handles", *ACS Omega*, vol. 2, no. 6, pp. 3064–3069, 2017. DOI: `10.1021/acsomega.7b00478` (cit. on pp. 13, 21).

[108] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, and M. A. Nash, "Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox", *J. Struct. Biol.*, vol. 197, no. 1, pp. 3–12, 2017. DOI: `10.1016/j.jsb.2016.02.011` (cit. on pp. 15, 16, 23, 29, 39).

[109] P. Hinterdorfer, W. Baumgartner, H. J. Gruber, K. Schilcher, and H. Schindler, "Detection and localization of individual antibody-antigen recognition events by atomic force microscopy.", *Proceedings of the National Academy of Sciences*, vol. 93, no. 8, pp. 3477–3481, 1996. DOI: `10.1073/pnas.93.8.3477` (cit. on p. 16).

[110] M. Ludwig, W. Dettmann, and H. E. Gaub, "Atomic force microscope imaging contrast based on molecular recognition", *Biophysical Journal*, vol. 72, no. 1, pp. 445–448, 1997. DOI: `10.1016/S0006-3495(97)78685-5` (cit. on p. 16).

[111] P. Hinterdorfer and Y. F. Dufrêne, "Detection and localization of single molecular recognition events using atomic force microscopy", *Nat. Methods*, vol. 3, no. 5, pp. 347–355, 2006. DOI: `10.1038/nmeth871` (cit. on p. 16).

[112] A. Raab, W. Han, D. Badt, S. J. Smith-Gill, S. M. Lindsay, H. Schindler, and P. Hinterdorfer, "Antibody recognition imaging by force microscopy", *Nature Biotechnology*, vol. 17, no. 9, pp. 902–905, 1999. DOI: `10.1038/12898` (cit. on p. 16).

[113] S. Zhang, H. Aslan, F. Besenbacher, and M. Dong, "Quantitative biomolecular imaging by dynamic nanomechanical mapping", *Chem. Soc. Rev.*, vol. 43, no. 21, pp. 7412–7429, 2014. DOI: `10.1039/C4CS00176A` (cit. on p. 16).

[114] D. Alsteens, V. Dupres, S. Yunus, J. P. Latgé, J. J. Heinisch, and Y. F. Dufrêne, "High-resolution imaging of chemical and biological sites on living cells using peak force tapping atomic force microscopy", *Langmuir*, vol. 28, no. 49, pp. 16 738–16 744, 2012. DOI: `10.1021/la303891j` (cit. on p. 16).

[115] M. Pfreundschuh, D. Alsteens, M. Hilbert, M. O. Steinmetz, and D. J. Müller, "Localizing chemical groups while imaging single native proteins by high-resolution atomic force microscopy", *Nano Letters*, vol. 14, no. 5, pp. 2957–2964, 2014. DOI: `10.1021/nl5012905` (cit. on p. 16).

[116] D. Alsteens, M. Pfreundschuh, C. Zhang, P. M. Spoerri, S. R. Coughlin, B. K. Kobilka, and D. J. Müller, "Imaging G protein-coupled receptors while quantifying their ligand-binding free-energy landscape", *Nature Methods*, vol. 12, no. 9, pp. 845–851, 2015. DOI: `10.1038/nmeth.3479` (cit. on p. 16).

[117] M. Pfreundschuh, D. Alsteens, R. Wieneke, C. Zhang, S. R. Coughlin, R. Tampé, B. K. Kobilka, and D. J. Müller, "Identifying and quantifying two ligand-binding sites while imaging native human membrane receptors by AFM", *Nature Communications*, vol. 6, 2015. DOI: `10.1038/ncomms9857` (cit. on p. 16).

[118] M. Dong, S. Husale, and O. Sahin, "Determination of protein structural flexibility by microsecond force spectroscopy", *Nature Nanotechnology*, vol. 4, no. 8, pp. 514–517, 2009. DOI: `10.1038/nnano.2009.156` (cit. on p. 16).

[119] M. Dong and O. Sahin, "A nanomechanical interface to rapid single-molecule interactions", *Nature Communications*, vol. 2, no. 1, 2011. DOI: `10.1038/ncomms1246` (cit. on p. 16).

[120] M. A. Jobst, C. Schoeler, K. Malinowska, and M. A. Nash, "Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy", *Journal of Visualized Experiments*, no. 82, pp. 1–10, Dec. 2013 (cit. on pp. 18, 150).

[121] E. M. Puchner, G. Franzen, M. Gautel, and H. E. Gaub, "Comparing proteins by their unfolding pattern", *Biophys. J.*, vol. 95, no. 1, pp. 426–434, 2008. DOI: `10.1529/biophysj.108.129999` (cit. on pp. 18, 36).

[122] M. A. Jobst, L. F. Milles, C. Schoeler, W. Ott, D. B. Fried, E. A. Bayer, H. E. Gaub, and M. A. Nash, "Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy", *Elife*, vol. 4, e10319, 2015. DOI: `10.7554/eLife.10319` (cit. on pp. 18, 35, 105).

[123] W. Ott, M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, and H. E. Gaub, "Elastin-like polypeptide linkers for Single-Molecule force spectroscopy", *ACS Nano*, vol. 11, no. 6, pp. 6346–6354, 2017. DOI: `10.1021/acsnano.7b02694` (cit. on pp. 18, 23, 75).

[124] C. Ortiz and G. Hadziioannou, "Entropic elasticity of single polymer chains of poly(methacrylic acid) measured by atomic force microscopy", *Macromolecules*, vol. 32, no. 3, pp. 780–787, Feb. 1999. DOI: `10.1021/ma981245n` (cit. on pp. 19–21).

[125] L. Livadaru, R. R. Netz, and H. J. Kreuzer, "Stretching Response of Discrete Semiflexible Polymers", *Macromolecules*, vol. 36, no. 10, pp. 3732–3744, 2003. DOI: `10.1021/ma020751g` (cit. on p. 19).

[126] W. Kuhn, "Über die Gestalt Fadenförmiger Moleküle in Lösungen", *Kolloid-Zeitschrift*, vol. 68, no. 1, pp. 2–15, Feb. 1934. DOI: `10.1007/BF01451681` (cit. on p. 19).

[127] E. Guth and H. Mark, “Zur innermolekularen Statistik, insbesondere bei Kettenmolekülen I”, *Monatshefte für Chemie*, vol. 65, no. 1, pp. 93–121, Dec. 1934. DOI: `10.1007/BF01522052` (cit. on p. 19).

[128] H. A. Kramers, “The behavior of macromolecules in inhomogeneous flow”, *The Journal of Chemical Physics*, vol. 14, no. 7, pp. 415–424, Jul. 1946. DOI: `10.1063/1.1724163` (cit. on p. 19).

[129] S. Smith, L. Finzi, and C. Bustamante, “Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads”, *Science*, vol. 258, no. 5085, pp. 1122–1126, Nov. 1992. DOI: `10.1126/science.1439819` (cit. on p. 19).

[130] R. Petrosyan, “Improved approximations for some polymer extension models”, *Rheologica Acta*, vol. 56, no. 1, pp. 21–26, Jan. 2017. DOI: `10.1007/s00397-016-0977-9` (cit. on pp. 20, 21).

[131] O. Kratky and G. Porod, “Röntgenuntersuchung gelöster fadenmoleküle”, *Recl. Trav. Chim. Pays-Bas*, vol. 68, no. 12, pp. 1106–1122, 1949. DOI: `10.1002/recl.19490681203` (cit. on pp. 20, 21).

[132] S. Brinkers, H. R. C. Dietrich, F. H. de Groote, I. T. Young, and B. Rieger, “The persistence length of double stranded DNA determined using dark field tethered particle motion”, *The Journal of Chemical Physics*, vol. 130, no. 21, p. 215 105, Jun. 2009. DOI: `10.1063/1.3142699` (cit. on p. 20).

[133] J. F. Marko and E. D. Siggia, “Stretching DNA”, *Macromolecules*, vol. 28, no. 26, pp. 8759–8770, Dec. 1995. DOI: `10.1021/ma00130a008` (cit. on pp. 20, 21).

[134] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, “Entropic elasticity of lambda-phage DNA”, *Science*, vol. 265, no. 5178, pp. 1599–1600, 1994. DOI: `10.1126/science.8079175` (cit. on p. 20).

[135] J. G. Kirkwood and J. Riseman, “The intrinsic viscosities and diffusion constants of flexible macromolecules in solution”, *The Journal of Chemical Physics*, vol. 16, no. 6, pp. 565–573, Jun. 1948, ISSN: 0021-9606. DOI: `10.1063/1.1746947` (cit. on p. 21).

[136] J. Riseman and J. G. Kirkwood, “The rotatory diffusion constants of flexible molecules”, *The Journal of Chemical Physics*, vol. 17, no. 5, pp. 442–446, May 1949. DOI: `10.1063/1.1747286` (cit. on p. 21).

[137] T. Odijk, “Stiff chains and filaments under tension”, *Macromolecules*, vol. 28, no. 20, pp. 7016–7018, Sep. 1995. DOI: `10.1021/ma00124a044`. arXiv: `1011.1669` (cit. on p. 21).

[138] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, and R. R. Netz, “Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-initio Theory”, *Phys. Rev. Lett.*, vol. 94, no. 4, pp. 048301-1–048301-4, 2005. DOI: `10.1103/PhysRevLett.94.048301` (cit. on p. 21).

[139] R. Walder, M.-A. LeBlanc, W. J. V. Patten, D. T. Edwards, J. A. Greenberg, A. Adhikari, S. R. Okoniewski, R. M. A. Sullan, D. Rabuka, M. C. Sousa, and T. T. Perkins, "Rapid characterization of a mechanically labile alpha-helical protein enabled by efficient site-specific bioconjugation", *Journal of the American Chemical Society*, vol. 139, no. 29, pp. 9867–9875, Jul. 2017. DOI: `10.1021/jacs.7b02958` (cit. on p. 21).

[140] F. Oesterhelt, M. Rief, and H. E. Gaub, "Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water", *New J. Phys.*, vol. 1, no. 6, pp. 1–11, 1999. DOI: `10.1088/1367-2630/1/1/006` (cit. on pp. 21, 22).

[141] W. Ott, T. Nicolaus, H. E. Gaub, and M. A. Nash, "Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation", *Biomacromolecules*, vol. 17, no. 4, pp. 1330–1338, 2016. DOI: `10.1021/acs.biomac.5b01726` (cit. on p. 23).

[142] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, and M. A. Nash, "Ultrastable cellulosome-adhesion complex tightens under load", *Nat. Commun.*, vol. 5, no. 5635, pp. 1–8, 2014. DOI: `10.1038/ncomms6635` (cit. on pp. 23, 29, 30, 35, 54).

[143] S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer, and H. E. Gaub, "Single-molecule dissection of the high-affinity cohesin-dockerin complex", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 50, pp. 20 431–20 436, 2012. DOI: `10.1073/pnas.1211929109` (cit. on pp. 23, 29, 30).

[144] F. Baumann, M. S. Bauer, L. F. Milles, A. Alexandrovich, H. E. Gaub, and D. A. Pippig, "Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy", *Nature Nanotechnology*, vol. 11, no. 1, pp. 89–94, 2016. DOI: `10.1038/nnano.2015.231` (cit. on pp. 23, 29, 30).

[145] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez, "Stepwise unfolding of titin under force-clamp atomic force microscopy", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 2, pp. 468–472, 2001. DOI: `10.1073/pnas.021321798` (cit. on p. 23).

[146] A. Noy and R. W. Friddle, "Practical single molecule force spectroscopy: How to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope", *Methods*, vol. 60, no. 2, pp. 142–150, 2013. DOI: `10.1016/j.ymeth.2013.03.014` (cit. on pp. 23, 24).

[147] G. I. Bell, "Models for the specific adhesion of cells to cells", *Science*, vol. 200, no. 4342, pp. 618–627, 1978. DOI: `10.1126/science.347575` (cit. on pp. 23, 24, 26).

[148] E. Evans and K. Ritchie, "Dynamic strength of molecular adhesion bonds", *Biophys. J.*, vol. 72, no. 4, pp. 1541–1555, 1997. DOI: `10.1016/S0006-3495(97)78802-7` (cit. on pp. 23, 24, 26, 28).

[149] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, "Molecular dynamics study of unbinding of the avidin-biotin complex", *Biophys. J.*, vol. 72, no. 4, pp. 1568–1581, 1997. DOI: `10.1016/S0006-3495(97)78804-0` (cit. on pp. 23, 24).

[150] O. Dudko, G. Hummer, and A. Szabo, "Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments", *Physical Review Letters*, vol. 96, no. 10, p. 108 101, 2006. DOI: `10.1103/PhysRevLett.96.108101` (cit. on pp. 23, 28).

[151] R. W. Friddle, A. Noy, and J. J. De Yoreo, "Interpreting the widespread nonlinear force spectra of intermolecular bonds", *Proceedings of the National Academy of Sciences*, vol. 109, no. 34, pp. 13 573–13 578, Aug. 2012. DOI: `10.1073/pnas.1202946109` (cit. on pp. 24, 27).

[152] F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring, "High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations", *Science*, vol. 342, no. 6159, pp. 741–743, 2013. DOI: `10.1126/science.1239764` (cit. on pp. 24, 27).

[153] C. Schoeler, R. C. Bernardi, K. H. Malinowska, E. Durner, W. Ott, E. A. Bayer, K. Schulten, M. A. Nash, and H. E. Gaub, "Mapping mechanical force propagation through biomolecular complexes", *Nano Lett.*, vol. 15, no. 11, pp. 7370–7376, 2015. DOI: `10.1021/acs.nanolett.5b02727` (cit. on pp. 24, 29, 30, 35).

[154] J. J. Hood, "On the Laws of Chemical Change", *Phil. Mag.*, vol. 6, no. 38, pp. 371–383, Nov. 1878 (cit. on p. 24).

[155] J. J. Hood, "On the Influence of Heat on the Rate of Chemical Change", *Phil. Mag.*, vol. 20, no. 125, pp. 323–328, Oct. 1885 (cit. on p. 24).

[156] M. J. H. van't Hoff, "Etudes de dynamique chimique", *Recueil des Travaux Chimiques des Pays-Bas*, vol. 3, no. 10, pp. 333–336, Sep. 1884. DOI: `10.1002/recl.18840031003` (cit. on p. 24).

[157] S. Arrhenius, "Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren", *Zeitschrift für Physikalische Chemie*, vol. 4, pp. 226–248, 1889 (cit. on p. 24).

[158] A. Christiansen and H. Kramers, "Über die Geschwindigkeit chemischer Reaktionen", *Zeitschrift für Physikalische Chemie*, vol. 104, pp. 451–469, 1923 (cit. on p. 24).

[159] H. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions", *Physica*, vol. 7, no. 4, pp. 284–304, Apr. 1940. DOI: `10.1016/S0031-8914(40)90098-2` (cit. on p. 24).

[160] A. Einstein, "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen", *Annalen der Physik*, vol. 322, no. 8, pp. 549–560, 1905. DOI: `10.1002/andp.19053220806` (cit. on p. 24).

[161] P. Langevin, "Sur la théorie du mouvement brownien", *C. R. Acad. Sci. (Paris)*, vol. 146, pp. 530–533, 1908. DOI: `10.1119/1.18725`. arXiv: `0009448 [cond-mat]` (cit. on p. 24).

[162] P. Hänggi, P. Talkner, and M. Borkovec, “Reaction-rate theory: fifty years after Kramers”, *Reviews of Modern Physics*, vol. 62, no. 2, pp. 251–341, Apr. 1990. DOI: `10.1103/RevModPhys.62.251` (cit. on pp. 24, 25).

[163] P. Talkner and J. Łuczka, “Rate description of Fokker-Planck processes with time-dependent parameters”, *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 69, no. 4, p. 10, 2004. DOI: `10.1103/PhysRevE.69.046109`. arXiv: `0307498 [cond-mat]` (cit. on p. 25).

[164] S. N. Zhurkov, “Kinetic concept of the strength of solids”, *International Journal of Fracture Mechanics*, vol. 1, pp. 311–323, Dec. 1965. DOI: `10.1007/BF00962961` (cit. on p. 25).

[165] E. Evans, “Probing the relation between force–lifetime–and chemistry in single molecular bonds”, *Annu. Rev. Biophys. Biomol. Struct.*, vol. 30, pp. 105–128, 2001. DOI: `10.1146/annurev.biophys.30.1.105` (cit. on p. 26).

[166] G. Hummer and A. Szabo, “Kinetics from Nonequilibrium Single-Molecule Pulling Experiments”, *Biophysical Journal*, vol. 85, no. 1, pp. 5–15, Jul. 2003. DOI: `10.1016/S0006-3495(03)74449-X` (cit. on p. 27).

[167] J. T. Bullerjahn, S. Sturm, and K. Kroy, “Theory of rapid force spectroscopy”, *Nature Communications*, vol. 5, 2014. DOI: `10.1038/ncomms5463` (cit. on p. 27).

[168] R. W. Friddle, P. Podsiadlo, A. B. Artyukhin, and A. Noy, “Near-Equilibrium Chemical Force Microscopy”, *The Journal of Physical Chemistry C*, vol. 112, no. 13, pp. 4986–4990, Apr. 2008. DOI: `10.1021/jp7095967` (cit. on p. 27).

[169] Z. Tshiprut, J. Klafter, and M. Urbakh, “Single-Molecule Pulling Experiments: When the Stiffness of the Pulling Device Matters”, *Biophysical Journal*, vol. 95, no. 6, pp. L42–L44, Sep. 2008. DOI: `10.1529/biophysj.108.141580` (cit. on pp. 27, 28).

[170] E. B. Walton, S. Lee, and K. J. Van Vliet, “Extending Bell’s Model: How Force Transducer Stiffness Alters Measured Unbinding Forces and Kinetics of Molecular Complexes”, *Biophysical Journal*, vol. 94, no. 7, pp. 2621–2630, Apr. 2008. DOI: `10.1529/biophysj.107.114454` (cit. on pp. 27, 28).

[171] C. Friedsam, A. K. Wehle, F. Kühner, and H. E. Gaub, “Dynamic single-molecule force spectroscopy: Bond rupture analysis with variable spacer length”, *Journal of Physics: Condensed Matter*, vol. 15, no. 18, S1709–S1723, Apr. 2003. DOI: `10.1088/0953-8984/15/18/305` (cit. on p. 28).

[172] C. Ray, J. R. Brown, and B. B. Akhremitchev, “Rupture force analysis and the associated systematic errors in force spectroscopy by AFM”, *Langmuir*, vol. 23, no. 11, pp. 6076–6083, 2007. DOI: `10.1021/la070131e` (cit. on p. 28).

[173] O. K. Dudko, G. Hummer, and A. Szabo, “Theory, analysis, and interpretation of single-molecule force spectroscopy experiments”, *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15 755–15 760, 2008. DOI: `10.1073/pnas.0806085105` (cit. on p. 28).

[174] C. Ray, J. R. Brown, and B. B. Akhremitchev, “Correction of Systematic Errors in Single-Molecule Force Spectroscopy with Polymeric Tethers by Atomic Force Microscopy”, *The Journal of Physical Chemistry B*, vol. 111, no. 8, pp. 1963–1974, Mar. 2007. DOI: `10.1021/jp065530h` (cit. on p. 28).

[175] L. B. Freund, “Characterizing the resistance generated by a molecular bond as it is forcibly separated”, *Proceedings of the National Academy of Sciences*, vol. 106, no. 22, pp. 8818–8823, Jun. 2009. DOI: `10.1073/pnas.0903003106` (cit. on p. 28).

[176] L. F. Milles, E. A. Bayer, M. A. Nash, and H. E. Gaub, “Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium perfringens”, *The Journal of Physical Chemistry B*, vol. 121, no. 15, pp. 3620–3625, Apr. 2017. DOI: `10.1021/acs.jpcb.6b09593` (cit. on pp. 29, 30).

[177] C. Verbelen, H. J. Gruber, and Y. F. Dufrêne, “The NTA–His6 bond is strong enough for AFM single-molecular recognition studies”, *Journal of Molecular Recognition*, vol. 20, no. 6, pp. 490–494, Nov. 2007. DOI: `10.1002/jmr.833` (cit. on p. 29).

[178] F. Rico and V. T. Moy, “Energy landscape roughness of the streptavidin–biotin interaction”, *Journal of Molecular Recognition*, vol. 20, no. 6, pp. 495–501, Nov. 2007. DOI: `10.1002/jmr.841` (cit. on p. 29).

[179] J. P. Junker and M. Rief, “Single-molecule force spectroscopy distinguishes target binding modes of calmodulin”, *Proceedings of the National Academy of Sciences*, vol. 106, no. 34, pp. 14 361–14 366, Apr. 2009. DOI: `10.1073/pnas.0904654106`. eprint: `arXiv:1408.1149` (cit. on p. 29).

[180] S. M. Sedlak, M. S. Bauer, C. Kluger, L. C. Schendel, L. F. Milles, D. A. Pippig, and H. E. Gaub, “Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry”, *PLOS ONE*, vol. 12, no. 12, A. Pastore, Ed., e0188722, Dec. 2017. DOI: `10.1371/journal.pone.0188722` (cit. on pp. 29, 30).

[181] J. Morfill, K. Blank, C. Zahnd, B. Luginbühl, F. Kühner, K.-E. Gottschalk, A. Plückthun, and H. E. Gaub, “Affinity-Matured Recombinant Antibody Fragments Analyzed by Single-Molecule Force Spectroscopy”, *Biophysical Journal*, vol. 93, no. 10, pp. 3583–3590, Nov. 2007. DOI: `10.1529/biophysj.107.112532` (cit. on p. 29).

[182] G. Neuert, C. Albrecht, E. Pamir, and H. Gaub, “Dynamic force spectroscopy of the digoxigenin-antibody complex”, *FEBS Letters*, vol. 580, no. 2, pp. 505–509, Jan. 2006. DOI: `10.1016/j.febslet.2005.12.052` (cit. on pp. 29, 30).

[183] J. Morfill, F. Kühner, K. Blank, R. A. Lugmaier, J. Sedlmair, and H. E. Gaub, “B-S Transition in Short Oligonucleotides”, *Biophysical Journal*, vol. 93, no. 7, pp. 2400–2409, Oct. 2007. DOI: `10.1529/biophysj.107.106112` (cit. on pp. 29, 30).

[184] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, “Sequence-dependent mechanics of single DNA molecules”, *Nature Structural Biology*, vol. 6, no. 4, pp. 346–349, Apr. 1999. DOI: `10.1038/7582` (cit. on p. 29).

[185] P. Zugenmaier, *Crystalline Cellulose and Cellulose Derivatives – Characterization and Structures*. Springer-Verlag, 2008, ISBN: 9783540739340 (cit. on p. 33).

[186] R. Lamed, E. Setter, and E. A. Bayer, “Characterization of a cellulose-binding, cellulase-containing complex in clostridium thermocellum”, *J. Bacteriol.*, vol. 156, no. 2, pp. 828–836, 1983 (cit. on p. 33).

[187] E. A. Bayer, J.-P. Belaich, Y. Shoham, and R. Lamed, “The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides”, *Annu. Rev. Microbiol.*, vol. 58, pp. 521–554, 2004. DOI: `10.1146/annurev.micro.57.030502.091022` (cit. on p. 33).

[188] A. B. Boraston, E. Kwan, P. Chiu, R. A. J. Warren, and D. G. Kilburn, “Recognition and hydrolysis of noncrystalline cellulose”, *J. Biol. Chem.*, vol. 278, no. 8, pp. 6120–6127, 2003. DOI: `10.1074/jbc.M209554200` (cit. on p. 33).

[189] D. Guillén, S. Sánchez, and R. Rodríguez-Sanoja, “Carbohydrate-binding domains: Multiplicity of biological roles”, *Applied Microbiology and Biotechnology*, vol. 85, no. 5, pp. 1241–1249, 2010. DOI: `10.1007/s00253-009-2331-y` (cit. on p. 33).

[190] K. Ruel, Y. Nishiyama, and J.-P. Joseleau, “Crystalline and amorphous cellulose in the secondary walls of arabidopsis”, *Plant Sci.*, vol. 193-194, pp. 48–61, 2012. DOI: `10.1016/j.plantsci.2012.05.008` (cit. on p. 33).

[191] W. W. Navarre and O. Schneewind, “Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria”, *Mol. Microbiol.*, vol. 14, no. 1, pp. 115–121, 1994. DOI: `10.1111/j.1365-2958.1994.tb01271.x` (cit. on p. 33).

[192] E. Roberts, J. Eargle, D. Wright, and Z. Luthey-Schulten, “Multiseq: Unifying sequence and structure data for evolutionary analysis”, *BMC Bioinformatics*, vol. 7, no. 1, p. 382, 2006. DOI: `10.1186/1471-2105-7-382` (cit. on p. 34).

[193] A. Mechaly, S. Yaron, R. Lamed, H.-P. Fierobe, A. Belaich, J.-P. Belaich, Y. Shoham, and E. A. Bayer, “Cohesin-dockerin recognition in cellulosome assembly: Experiment versus hypothesis”, *Proteins: Structure, Function, and Genetics*, vol. 39, no. 2, pp. 170–177, May 2000. DOI: `10.1002/(sici)1097-0134(20000501)39:2<170::aid-prot7>3.0.co;2-h` (cit. on p. 34).

[194] A. L. Carvalho, F. M. V. Dias, J. A. M. Prates, T. Nagy, H. J. Gilbert, G. J. Davies, L. M. A. Ferreira, M. J. Romao, and C. M. G. A. Fontes, “Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex”, *Proceedings of the National Academy of Sciences*, vol. 100, no. 24, pp. 13 809–13 814, Nov. 2003. DOI: `10.1073/pnas.1936124100` (cit. on p. 34).

[195] E. A. Bayer, J.-P. Belaich, Y. Shoham, and R. Lamed, “The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides”, *Annual Review of Microbiology*, vol. 58, no. 1, pp. 521–554, Oct. 2004. DOI: `10.1146/annurev.micro.57.030502.091022` (cit. on p. 34).

[196] M. A. Nash, S. P. Smith, C. M. Fontes, and E. A. Bayer, “Single versus dual-binding conformations in cellulosomal cohesin–dockerin complexes”, *Current Opinion in Structural Biology*, vol. 40, pp. 89–96, Oct. 2016. DOI: `10.1016/j.sbi.2016.08.002` (cit. on p. 34).

[197] A. L. Carvalho, F. M. V. Dias, T. Nagy, J. A. M. Prates, M. R. Proctor, N. Smith, E. A. Bayer, G. J. Davies, L. M. A. Ferreira, M. J. Romao, C. M. G. A. Fontes, and H. J. Gilbert, “Evidence for a dual binding mode of dockerin modules to cohesins”, *Proceedings of the National Academy of Sciences*, vol. 104, no. 9, pp. 3089–3094, Feb. 2007. DOI: `10.1073/pnas.0611173104` (cit. on p. 35).

[198] B. A. Pinheiro, M. R. Proctor, C. Martinez-Fleites, J. A. M. Prates, V. A. Money, G. J. Davies, E. A. Bayer, C. M. G. A. FontesM, H.-P. Fierobe, and H. J. Gilbert, "The Clostridium cellulolyticum Dockerin displays a dual binding mode for its cohesin partner", *Journal of Biological Chemistry*, vol. 283, no. 26, pp. 18 422–18 430, Apr. 2008. DOI: `10.1074/jbc.m801533200` (cit. on p. 35).

[199] K. Cameron, S. Najmudin, V. D. Alves, E. A. Bayer, S. P. Smith, P. Bule, H. Waller, L. M. A. Ferreira, H. J. Gilbert, and C. M. G. A. Fontes, "Cell-surface attachment of bacterial multienzyme complexes involves highly dynamic protein-protein anchors", *Journal of Biological Chemistry*, vol. 290, no. 21, pp. 13 578–13 590, Apr. 2015. DOI: `10.1074/jbc.m114.633339` (cit. on p. 35).

[200] O. Salama-Alber, M. K. Jobby, S. Chitayat, S. P. Smith, B. A. White, L. J. W. Shimon, R. Lamed, F. Frolow, and E. A. Bayer, "Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: Binding fidelity, promiscuity, and structural buttresses", *J. Biol. Chem.*, vol. 288, no. 23, pp. 16 827–16 838, 2013. DOI: `10.1074/jbc.M113.466672` (cit. on p. 35).

[201] M. Hammel, H.-P. Fierobe, M. Czjzek, S. Finet, and V. Receveur-Bréchot, "Structural insights into the mechanism of formation of cellulosomes probed by small angle x-ray scattering", *Journal of Biological Chemistry*, vol. 279, no. 53, pp. 55 985–55 994, Oct. 2004. DOI: `10.1074/jbc.m408979200` (cit. on p. 35).

[202] M. Hammel, H.-P. Fierobe, M. Czjzek, V. Kurkal, J. C. Smith, E. A. Bayer, S. Finet, and V. Receveur-Bréchot, "Structural basis of cellulosome efficiency explored by small angle x-ray scattering", *Journal of Biological Chemistry*, vol. 280, no. 46, pp. 38 562–38 568, Sep. 2005. DOI: `10.1074/jbc.m503168200` (cit. on p. 35).

[203] R. D. Piner, J. Zhu, F. Xu, S. Hong, and C. A. Mirkin, ""Dip-pen"nanolithography", *Science*, vol. 283, no. 5402, pp. 661–663, Jan. 1999. DOI: `10.1126/science.283.5402.661` (cit. on p. 35).

[204] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns", *Nature*, vol. 440, no. 7082, pp. 297–302, Mar. 2006. DOI: `10.1038/nature04586` (cit. on pp. 35, 36).

[205] C. Steinhauer, R. Jungmann, T. L. Sobey, F. C. Simmel, and P. Tinnefeld, "DNA origami as a nanoscopic ruler for super-resolution microscopy", *Angewandte Chemie International Edition*, vol. 48, no. 47, pp. 8870–8873, Nov. 2009. DOI: `10.1002/anie.200903308` (cit. on p. 35).

[206] D. A. Pippig, F. Baumann, M. Strackharn, D. Aschenbrenner, and H. E. Gaub, "Protein–DNA chimeras for nano assembly", *ACS Nano*, vol. 8, no. 7, pp. 6551–6555, Jun. 2014. DOI: `10.1021/nn501644w` (cit. on p. 35).

[207] K. R. Erlich, F. Baumann, D. A. Pippig, and H. E. Gaub, "Strep-tag II and monovalent strep-tactin as novel handles in single-molecule cut-and-paste", *Small Methods*, vol. 1, no. 8, p. 1 700 169, Jul. 2017. DOI: `10.1002/smtd.201700169` (cit. on p. 35).

[208] T. Cordes, M. Strackharn, S. W. Stahl, W. Summerer, C. Steinhauer, C. Forthmann, E. M. Puchner, J. Vogelsang, H. E. Gaub, and P. Tinnefeld, "Resolving single-molecule assembled patterns with superresolution blink-microscopy", *Nano Letters*, vol. 10, no. 2, pp. 645–651, Feb. 2010. DOI: `10.1021/nl903730r` (cit. on p. 36).

[209] M. Strackharn, S. W. Stahl, E. M. Puchner, and H. E. Gaub, "Functional assembly of aptamer binding sites by single-molecule cut-and-paste", *Nano Letters*, vol. 12, no. 5, pp. 2425–2428, Apr. 2012. DOI: `10.1021/nl300422y` (cit. on p. 36).

[210] D. C. Duffy, J. C. McDonald, O. J. A. Schueller, and G. M. Whitesides, "Rapid prototyping of microfluidic systems in poly(dimethylsiloxane)", *Analytical Chemistry*, vol. 70, no. 23, pp. 4974–4984, Dec. 1998. DOI: `10.1021/ac980656z` (cit. on p. 36).

[211] S. R. Quake, "From micro- to nanofabrication with soft materials", *Science*, vol. 290, no. 5496, pp. 1536–1540, Nov. 2000. DOI: `10.1126/science.290.5496.1536` (cit. on p. 36).

[212] D. Gerber, S. J. Maerkl, and S. R. Quake, "An in vitro microfluidic approach to generating protein-interaction networks", *Nature Methods*, vol. 6, no. 1, pp. 71–74, Dec. 2008. DOI: `10.1038/nmeth.1289` (cit. on p. 36).

[213] M. Otten, W. Ott, M. A. Jobst, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, and H. E. Gaub, "From genes to protein mechanics on a chip", *Nat. Methods*, vol. 11, no. 11, pp. 1127–1130, 2014. DOI: `10.1038/nmeth.3099` (cit. on pp. 36, 127).

[214] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, and J. Kjems, "Self-assembly of a nanoscale DNA box with a controllable lid", *Nature*, vol. 459, no. 7243, pp. 73–76, May 2009. DOI: `10.1038/nature07971` (cit. on p. 36).

[215] S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih, "Self-assembly of DNA into nanoscale three-dimensional shapes", *Nature*, vol. 459, no. 7245, pp. 414–418, May 2009. DOI: `10.1038/nature08016` (cit. on p. 36).

[216] J.-M. Arbona, J.-P. Aimé, and J. Elezgaray, "Cooperativity in the annealing of DNA origamis", *The Journal of Chemical Physics*, vol. 138, no. 1, p. 015 105, Jan. 2013. DOI: `10.1063/1.4773405` (cit. on p. 37).

[217] J.-P. J. Sobczak, T. G. Martin, T. Gerling, and H. Dietz, "Rapid folding of DNA into nanoscale shapes at constant temperature", *Science*, vol. 338, no. 6113, pp. 1458–1461, Dec. 2012. DOI: `10.1126/science.1229919` (cit. on p. 37).

[218] K. E. Dunn, F. Dannenberg, T. E. Ouldridge, M. Kwiatkowska, A. J. Turberfield, and J. Bath, "Guiding the folding pathway of DNA origami", *Nature*, vol. 525, no. 7567, pp. 82–86, Aug. 2015. DOI: `10.1038/nature14860` (cit. on p. 37).

[219] A. B. Farimani, P. Dibaeinia, and N. R. Aluru, "DNA origami–graphene hybrid nanopore for DNA detection", *ACS Applied Materials & Interfaces*, vol. 9, no. 1, pp. 92–100, Dec. 2016. DOI: `10.1021/acsami.6b11001` (cit. on p. 37).

[220] F. Praetorius and H. Dietz, "Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes", *Science*, vol. 355, no. 6331, eaam5488, Mar. 2017. DOI: `10.1126/science.aam5488` (cit. on p. 37).

[221] D. Han, X. Qi, C. Myhrvold, B. Wang, M. Dai, S. Jiang, M. Bates, Y. Liu, B. An, F. Zhang, H. Yan, and P. Yin, "Single-stranded DNA and RNA origami", *Science*, vol. 358, no. 6369, eaao2648, Dec. 2017. DOI: `10.1126/science.aao2648` (cit. on p. 37).

[222] F. Praetorius, B. Kick, K. L. Behler, M. N. Honemann, D. Weuster-Botz, and H. Dietz, "Biotechnological mass production of DNA origami", *Nature*, vol. 552, no. 7683, pp. 84–87, Dec. 2017. DOI: `10.1038/nature24650` (cit. on p. 37).

[223] B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, and J. P. K. Doye, "Introducing improved structural properties and salt dependence into a coarse-grained model of DNA", *The Journal of Chemical Physics*, vol. 142, no. 23, p. 234 901, Jun. 2015. DOI: `10.1063/1.4921957` (cit. on p. 37).

Acknowledgements

THIS DOCTORAL THESIS is not only written by me as a doctoral candidate, but also made possible by a large number of predecessors, coworkers, staff, and of course professors and post-docs. Therefore, I owe my thanks to the following people:

Hermann Gaub, who made it possible for me to explore, learn and apply countless scientific technologies and abilities that are integral to scientific careers. He always gives his students the freedom to plan end execute their projects on their own behalf, but stands on their side, when they require his experience and expertise, for all of which I am very grateful. **Michael Nash**, who initially took me in as one of his first team members, and showed me countless important aspects of scientific life. **Edward Bayer**, who gave important impulses for many of the projects of this thesis and provided initial molecules for the studies as well as helpful and valuable insights.

My immediate research colleagues, who also became dear and valued friends over the past few years. **Ellis Durner**, who not only shared my passion for creating devices and developing instruments, but also showed me many insights into electronics and microcontroller programming. **Lukas Milles**, whose pythonic skills made me learn an all new and powerful language, and enabled massive data reduction among our projects. **Thomas Nicolaus**, who helped a lot in the lab, and provided many insights and almost exclusively cherfulness. **Wolfgang Ott**, did extensive proofreading, collaborated in uncountable subprojects, taught many physicists molecular biotechnology skills. **Tobias Verdorfer**, who also did excellent development on the instruments and really brought life into the group. Amongst many professional skills and methods we learned, shared, refined and applied, I cannot leave unmentioned countless discussions and activities ranging from slightly to severly off-topic.

Project partners, who entrusted me with their work participated in fruitful and interesting discussions and opened yet another field of research to me: **David Smith**, **Martin Sajfutdinow** and **Megan Engel**. Coworkers who, despite the (spatial or project-wise) distance still contributed severly to my daily life in the lab, or to projects of this thesis: **Fabian Baumann**, **Magnus Bauer**, **Katherine Erlich**, **Diana Pippig**, **Constantin Schöler**, **Eduard Unterauer**, and **Willem Vanderlinden**, and also all other lab members. Of course, **Angelika Kardinal**, **Sylvia Kreuzer** and **Sabine Faust** who ran the lab and organized everything around it.

This thesis is typeset in LaTeX and compiled with pdflatex. Despite being heavily altered, many parts of the typesetting template for this thesis are still originally based on the template by **Jordan Suchow**.

I furthermore want to thank **Elena** and **Alex** for their valuable and continuous company, and last but not least, I want to thank my sisters **Andrea** and **Verena** and my parents **Erna** and **Karl-Heinz** for their long-lasting support.