Force and Affinity in Cellulosomal Complexes

Ellis Durner

München 2019

Force and Affinity in Cellulosomal Complexes

Dissertation

an der

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

vorgelegt von

Ellis Durner

aus Wangen im Allgäu

München, 15.11.2019

Erstgutachter: Prof. Dr. Hermann E. Gaub
Zweitgutachter: Prof. Dr. Joachim Rädler
Tag der mündlichen Prüfung: 19.12.2019

Zusammenfassung

In dieser Arbeit werden die molekularen Mechanismen der Organisation des Cellulosoms - ein komplexes extrazelluläres Proteinnetzwerk - als Modellsystem für Protein-Protein Interaktionen mittels biophysikalischer Methoden untersucht. Dieses extrazelluläre Organell ermöglicht bestimmten Bakterien die Zersetzung von Cellulose, indem es Enzyme und Cellulose-Bindedomänen auf gerüstartigen Proteinstrukturen in synergistischer Weise kombiniert. Die einzelnen Komponenten werden hierbei von einer Klasse von Rezeptor-Liganden-Paaren namens Cohesin- Dockerin in ihrer Stöchiometrie und Anordnung funktionell kombiniert.

Ein Teil dieser Arbeit besteht in der Entschlüsselung der molekularen Bindemechanismen des Cohesins CohE, welches das Bakterium *Ruminococcus flavefaciens* mit seinem Cellusom verbindet. Durch die Kombination von Einzelmolekül-Kraftspektroskopie mit Molekulardynamik-Simulationen konnte die aussergewöhnliche Belastbarkeit der Interaktionen von CohE mit zwei homologen Dockerinen entschlüsselt werden. Hierbei wurde insbesondere der Einfluss der Kraftpropagation innerhalb eines Proteinkomplexes auf dessen mechanische Widerstandsfähigkeit untersucht. Die physiologische Verankerung über den carboxyl-Terminus von CohE erwies sich als deutlich robuster im Vergleich zu einer nicht nativen N-terminalen Verankerung.

Um den Kontrast zwischen hoher mechanischer Belastbarkeit bei moderaten Affinitäten im nano- bis mikromolaren Bereich besser verstehen zu können, wandte ich mich der Bestimmung der kinetischen Ratenkonstanten k_{off} und k_{on} zu, deren Quotient die Gleichgewichtskonstante bildet. Während es eine kleine Dissoziationskonstante dem Bakterium ermöglichen würde die von ihm exprimierte Nanomaschinerie fest an sich zu binden, könnte ein höheres k_{off} und k_{on} einen dynamischeren Austausch von Cellulosomen innerhalb des Mikrobioms ermöglichen. Zusätzlich stellte sich die Frage, ob die Verankerungsgeometrie auch in Abwesenheit von Kraft Einfluss auf das Bindeverhalten nehmen würde. Nachdem initiale Messungen mittels Oberflächenplasmonenresonanzspektroskopie inkonsistent waren, wurde eine neuartige, enzymbasierte Kopplungsstrategie für oberflächengebundene Affinitätsbestimmungen entwickelt. Hiermit konnte CohE funktional und spezifisch auf Sensoroberflächen immobilisiert werden. Es zeigte sich, dass in Abwesenheit von externer Kraft die Verankerungsgeometrie von CohE keinen Einfluss auf das Bindeverhalten hat. Dies bestärkt im Umkehrschluss die Hypothese, dass mechanische Stabilitäten stets geometrieabhängig zu untersuchen sind.

Im Rahmen dieser Arbeit wurden auch methodische Verbesserungen in der Einzelmolekülkraftspektroskopie erzielt. Zum einen wurde eine Strategie entwickelt, um Proteindomänen zeitsparend *in vitro* zu exprimieren und ohne weitere Aufreinigung spezifisch auf Objektträgern zu verankern. Die darauffolgende enzymatische Peptidligation eines Dockerins via Sortase A erlaubt es nun, mit hohem Durchsatz Entfaltungsstudien an Proteinen mithilfe der Cohesin-Dockerin Interaktion durchzuführen.

Weiterhin ermöglichte es dieselbe Sortase-vermittelte Peptidligation, die gängigen Polyethylenlinker durch Elastin-ähnliche Peptide zu ersetzen. Dies verhindert Artefakte, die sonst durch Polyethylenlinker bei Protein-Kraftspektroskopie über 100 pN entstünden.

Zuletzt wurde der Entfaltungsprozess einer Cohesin-Domäne aus *Acetivibrio cellulolyticus* untersucht, deren Familie in vorangegangenen Studien teils bimodale Entfaltungskraftverteilungen zeigte. Durch die Kombination zweier Messmodi konnte die Kraft-Ladungsrate über fünf Größenordnungen variiert werden. Es konnte gezeigt werden, dass das dabei beobachtete Verhalten mit einer Konformationsänderung während der experimentellen Zeitskala zwischen verschiedenen, gefalteten Konformationen konsistent ist.

Contents

Part I

Context

1
Introduction

Understanding molecular interactions is a key challenge in the ongoing effort to unravel the self-organizing nature of life. Proteins navigate a vast conformational space until they adopt a functional, folded state - a topic that may at first glance not seem best explored by the reductionist approach of a physics. In recent decades however, techniques have emerged that allow us to manipulate and probe single biomolecules, to investigate how they interact with each other, or how they respond to external stimuli. The invention and development of techniques such as single-molecule force spectroscopy or single-molecule fluorescence methods provide access to the length and timescales relevant to biomolecular interactions. More importantly, they allow us to see more diverse behaviors otherwise hidden by ensemble averaging. This however does not mean that more traditional ensemble methods of investigating biomolecular interactions such as isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR) should be disregarded. It is rather a combination of all these tools that enables us to gain a better understanding of how the building blocks of life interact.

Within this thesis, the cellulosome, a large extracellular network of proteins expressed by certain bacteria, that aids in the degradation of cellulose, is used as a model system for studying protein-protein interactions. It consists of a multitude of scaffold-like proteins, enzymes and cellulose-binding domains, interconnected by a class of receptor-ligand pairs called cohesin-dockerin.

Publication 3.1 investigates the remarkable mechanical strength of a cohesin-dockerin pair, which non-covalently anchors cellulose-binding proteins to the bacterial cell wall of the cellulose degrading bacterium *Ruminococcus flavefaciens*. Steered all-atom molecular dynamic simulations revealed that - contrary to intuition - the surface contact area between the binding partners increases under load.

Publication 3.2 provides further mechanistic insight into how this protein complex is able to withstand the remarkable forces measured. To study how force propagates through the complex, the cohesin was anchored in two geometries, via the carboxyl-terminus (which in nature is coupled to the bacterium's cell wall), as well as via the amine-terminus. A substantial decrease in force resilience in the non-native anchoring geometry was revealed. By analyzing all-atom molecular dynamics simulations with thermodynamic fluctuation theory, force propagation pathways could be determined. This helped to elucidate the mechanics that stabilize the complex, revealing pathways with strong components normal to the pulling direction.

Based on the knowledge gained from and with the tools developed for these two publications, a homologous receptor-ligand pair was investigated in **publication 3.3**. It non-covalently anchors a large multi-domain cellulosome scaffold to *R. flavefaciens*. From MD-simulations it was predicted to be even stronger than the system characterized in publications 3.1 and 3.2. This was confirmed experimentally, whereas the affinity of this complex was found to be lower than for the weaker complex. This highlights the importance of distinguishing between the force-induced unbinding along a certain reaction coordinate, and the sum of all unbinding pathways, which give rise to the off-rate of a complex in the absence of force.

Driven by these results, we sought to obtain a more complete understanding of the role these complexes serve in nature. The contrast between very high force resilience and moderate affinities in the nano- to micro-molar range (see publication 3.3) as determined by ITC prompted the question which kinetic rate constants k_{on} and k_{off} give rise to them. While low off-rates would allow the bacterium to hold on to the nano-machinery it expressed, higher off-rates might provide for a more flexible composition of the cellulosome. As initial measurements conducted via SPR (using non-specific amine coupling) were inconsistent, a novel coupling strategy for surface-based affinity assays was developed, detailed in **publication 3.4**. Using a biolayer interferometer (BLI) as measurement platform, three different enzymatic coupling methods were used to covalently and site-specifically immobilize a protein of interest. Making use of the same short tags employed in the SMFS experiments of this thesis, this circumvents problems that arise from the conventionally employed non-specific coupling via primary amines. This way, cohesin E (*R.f.*) was functionally immobilized on BLI sensors in both geometries studied in publication 3.2. It was found that there was no apparent difference between immobilization geometries in BLI, *i.e.* in the absence of force, in contrast to under force. These findings underline the importance and influence of the force loading geometry and the resulting force propagation in biomolecular complexes.

Publication 3.5 investigates the unfolding behavior of a cohesin domain from *Acetovibrio cellulolyticus*. Verdorfer et al. [1, 2] found some members of this family of cohesins to exhibit bimodal unfolding force distributions, which became unimodal upon ligand binding. To assess whether this was caused by a heterogeneity in folding states or unfolding pathways, a combination of constant loading rate and constant speed SMFS was conducted. This allowed recording of unfolding forces over a very wide range of loading rates. The shape of the unfolding force distributions changed in a force loading rate-dependent manner - consistent with the domain switching between at least two different folded states within the experimental timescale. Monte Carlo simulations modelling a simple system consisting of two folded states, which interchange in a non force-dependent manner, were qualitatively consistent with the recorded SMFS data. While this behavior has been described before in a theoretical study by Pierse and Dudko [3], this is the first time such a behavior was shown for a protein domain by SMFS. During the course of this dissertation, efforts to improve throughput and data quality in SMFS led to publications 3.6 and 3.7.

Publication 3.6 presents a streamlined approach to expressing proteins of interest via *in vitro* transcription and translation reactions and immobilizing them on a sample surface spatially separated via silicon microwell masks. Subsequently, CttA-XDoc - which is part of the high-force receptor-ligand system described in publications

3.1, 3.2 and 3.3 - is enzymatically ligated to the protein of interest. This allows for high specificity and throughput, while eliminating the need to express the protein of interest as a fusion protein with the pulling handle.

Publication 3.7 replaces the in SMFS commonly used polyethylene-glycol (PEG) linkers with elastin-like polymers (ELP), improving three aspects of protein based SMFS. As their amino acid sequence is genetically encoded, they can be designed to have terminal sequences recognized by peptide-ligases such as Sortase A. Not only does this provide a site-specific means of anchoring proteins, it also assures that only full-length ELPs are functionalized, ensuring linker monodispersity. Using ELPs as linkers also results in a homogeneous system, where the whole stretched polymer is a peptide. This is especially advantageous for measurements where forces of more than $\sim 100\,\mathrm{pN}$ are reached, as PEG exhibits a force-dependent phase transition around that force.

Refinements to AFM instrumentation as well as careful data analysis allowed the investigation of a possible mechano-activation of focal adhesion kinase (FAK) (**publication 3.8**). FAK is a signaling molecule involved in cell adhesion and migration. By combining *in silico* and *in vitro* SMFS, it is shown that mechano-activation can below the unfolding force of FAK. This is consistent with a proposed force - or distance - induced conformational change of FAK that activates it.

2
Context

2.1 The Cellulosome

First described by R. Lamed and E. A. Bayer in a 1983 study [4], the cellulosome is a bacterial multi-domain, extracellular protein network capable of efficent plant fiber degradation. It is organized by, and constructed around a receptor-ligand interaction termed cohesin-dockerin. A cellulosome typically contains a variety of different enzymes and adhesion proteins fused to dockerin domains, which non-covalently bind to the cohesin-bearing cellulosome backbones. These backbones are multi-domain proteins called scaffoldins, which usually contain several cohesin domains, separated by peptide linkers. Scaffoldins also may contain Dockerin domains themselves, allowing for an intricate stacking of different scaffoldins. This self-assembling nature allows bacteria to fine tune the catalytic activity to their substrate. But it also makes it an excellent toolbox for the thriving field of synthetic biology, as the cellulosome provides many orthogonal cohesin-dockerin pairs, which differ in affinity, specificity, robustness against force, and size. Recently, David et al. have discovered a member of the human gut microbiome, *Ruminococcus champanellensis* [5], to express cellulosomes.

2.2 Affinity Determination and Kinetic Rate Measurements

The correct determination of intermolecular affinities is essential for many areas of research. The perfect technique to determine binding behavior should not suffer from unspecific surface binding effects, require little material, be sensitive enough to detect low molecular-weight changes, provide kinetic rate data, and be label-free. Because no actually existing technique satisfies all of these qualities, some systems may require more than one technique to reliably characterize an interaction. Quantitative methods to determine intermolecular affinities can be divided by three criteria - is one of the binding partners immobilized to a surface, can they resolve kinetic rates (or do they just give access to the equilibrium binding constant), and are they label-free. Within this work, isothermal calorimetry (ITC), as well as biolayer interferometry (BLI), were employed to study the binding behavior of receptor-ligand systems in the absence of force.

2.2.1 Isothermal Titration Calorimetry

While ITC has the advantage of being a truly label-free technique, requiring no molecular modifications and being largely unaffected by suface adsorpion, it does not provide kinetic rate data (with the exception of the recently developed (but debated) kinetic ITC [6]). It works by tracking the heat differential required to keep temperature constant in two liquid cells, one of which contains reference buffer, and one which contains sample to be analyzed. Consequently, ligand binding in the sample cell needs to produce a sufficiently large temperature change to be observable. Therefore, the change in Gibbs free energy upon binding $\Delta G = \Delta H - T\Delta S$ must have a sufficient enthalpic component to be observable by ITC. At the same time, the binding sigmoid's sharpness is a function of receptor concentration in the cell. The cell concentration needs to be tuned such that sufficient data points are acquired around the sigmoid's turning point, which provides an upper limit for concentrations that can be chosen. Therefore, very high affinity binders (roughly $K_d < 1\,\text{nM}$) cannot be reliably probed via ITC (although this limit will shift with cell volume).

2.2.2 Biolayer Interferometry

Observing ligand binding in both, biolayer interferometry as well as surface plasmon resonance (SPR), is based on changes in optical density at a sensor surface upon binding. Surface based techniques, such as BLI, SPR or quartz crystal microbalance often provide high sensitivity. They are suitable to study high-affinity binders, as the desired signal correlates to the occupancy of the surface-bound receptors. This comes at the cost of having to manage unspecific surface binding effects. In order to mitigate these issues, great effort has gone into the development of surface coatings designed to minimize unspecific adsorption [7]. Nevertheless, the experimentalist often has to painstakingly optimize buffer conditions and passivation reagents for every new interaction investigated. Referencing the measurement signal with a second signal acquired from a non-functionalized surface does not provide a means to fully control for unspecific surface adsorption, because the ligand concentration at a sensor surface drastically increases upon specific receptor binding. This often causes unspecific surface binding to increase with the occupancy of surface-coupled receptors. A second challenge for surface based assays is the surface immobilization of the molecules to be studied. While SPR and BLI often is termed label-free, surface immobilization often involves antibody-capture methods or covalent modification to primary amines. A possible solution for this issue is presented in publication 3.4. Here, enzyme-based site-specific ligation reactions are used to covalently functionalize BLI sensors under mild reaction conditions.

Analyzing Kinetic Rate Measurements

Kinetic Rates for a one-to-one binding scenario between receptor and ligand can be extracted in a straight-forward manner by fitting equations 2.1 and 2.2. Here, y_{max} represents the maximal binding signal, that can be achieved when all receptors are occupied with a ligand, and y_{off} is the binding signal at the beginning of ligand dissociation. Ideally, for several concentrations of ligand, association and dissociation cycles are recorded, and the fitted globally. This means that k_{on}, k_{off}, and y_{max} are

optimized for data from all concentrations simultaneously. However, if the sensor's binding capacity degrades over the course of the measurements, it might become necessary to optimize y_{max} for each concentration separately. Figure 2.2 gives an example of BLI data showing binding and unbinding of a receptor-ligand interaction under variation of ligand concentations.

$$y = (y_0 - y_{off}) * e^{-k_{off}*(t-t_{0,dissociation})} \tag{2.1}$$

$$y = y_{max} \frac{1}{1 + \frac{k_{off}}{k_{on}*[c]}} * \left(1 - e^{-([c]k_{on}+k_{off})*(t-t_{0,association})}\right) \tag{2.2}$$

If full ligand dissociation cannot be achieved after each measurement cycle (for example because of very small off-rates), one option is to force ligand unbinding by altering buffer conditions. However, this so called regeneration procedure needs to be established for each interaction anew, and it should be ensured that binding curves are repeatable after regeneration. Sometimes, no viable regeneration conditions can be found. In these cases, another measurement protocol can be employed, that does not require complete ligand dissociation before moving on to another ligand concentration. Figure 2.3 gives an example for this measurement protocol. In order to correctly analyze such data, one needs to account for the already occupied receptor population at the beginning of each association cycle. To do so, equation 2.2 solved for

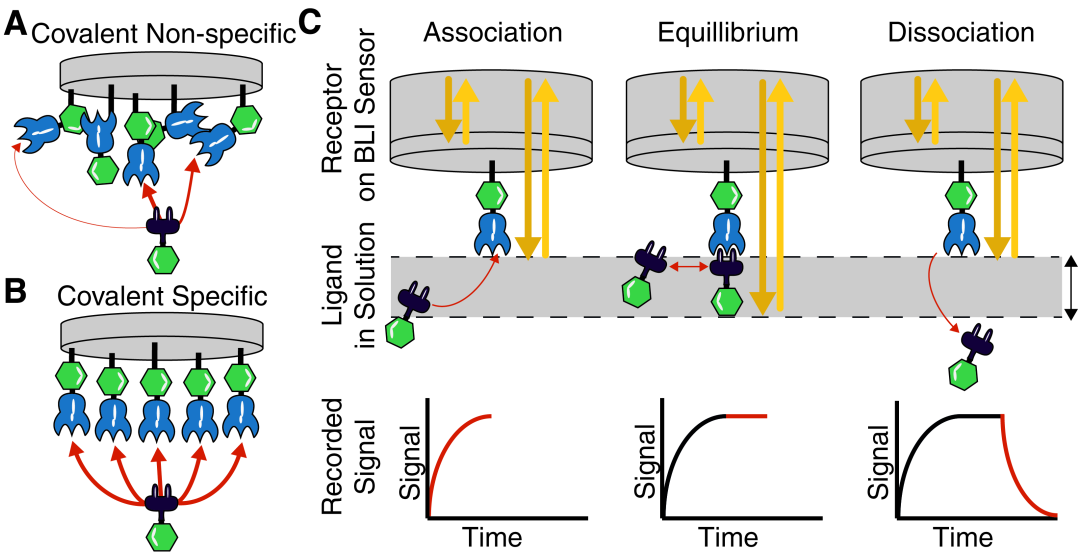


Figure 2.1. Illustration of affinity measurements via BLI. (**A**) Non-specific immobilization of the receptor on the sensor in different geometries as a result of several accessible amine-groups. (**B**) Specific and site-directed immobilization of a receptor to a sensor. All receptors are homogeneously orientated. The red arrows in (A) represent different binding geometries with possibly different kinetics, whereas specific attachment (B) provides a uniform population of binders. (**C**) The principle of a BLI kinetic experiment. A receptor-functionalized sensor is immersed into a ligand solution. The increasing signal shows binding of the ligand. When the sensor signal has reached a steady state, the rates of ligand association and dissociation are equal—the system has reached equilibrium. The sensor is then moved to a buffer solution, the receptor starts to dissociate and the detected signal decreases again. Figure adapted from publication 3.4.

the time t, yielding equation 2.3. This can be used to calculate a 'time offset' for every association step, by inserting parameters k_{on}, k_{off} and the current concentration [c]. The so calculated t_{offset} can then be inserted into equation 2.4.

$$t_{offset} = \frac{-ln\left(1 - \left(\frac{y_{start}}{y_{max}*[c]}\right) * \left(\frac{k_{off}}{k_{on}} + [c]\right)\right)}{k_{on} * [c] + k_{off}} \tag{2.3}$$

$$y = y_{max}\frac{1}{1 + \frac{k_{off}}{k_{on}*[c]}} * \left(1 - e^{-([c]k_{on}+k_{off})*(t-t_{0,association}+t_{offset})}\right) \tag{2.4}$$

Equations 2.3 and 2.3 can now again be simultaneously fitted to the data under variation of k_{on}, k_{off}, and y_{max}.

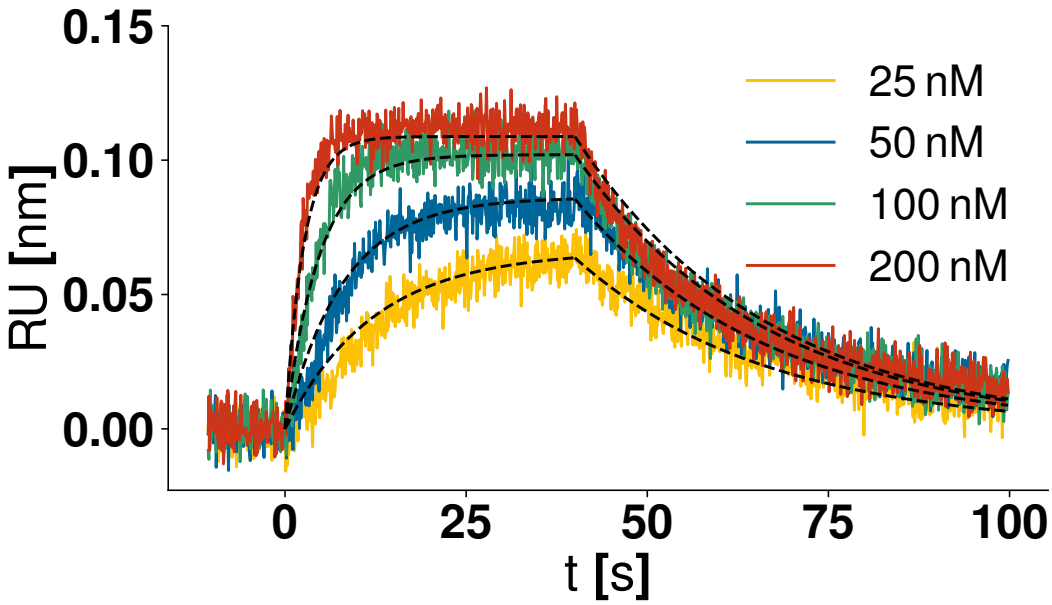


Figure 2.2. This figure illustrates a typical series of kinetic rate measurements via BLI. After each dissociation, sufficient time is given for all ligand to dissociate. (Data from publication 3.4).

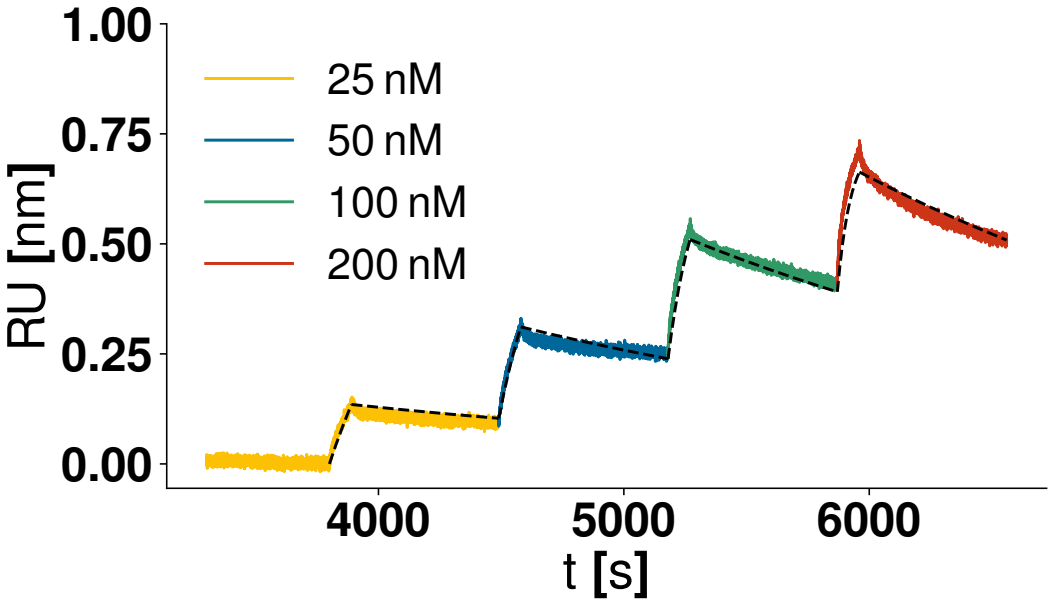


Figure 2.3. An example for a so called kinetic titration series, recorded by BLI. Binding of a receptor-ligand interaction is measured under variation of ligand concentration. While the time given for ligand dissociation between association streps needs to be sufficient for correct determination of the off-rate, complete ligand dissociation does not need to be reached in this measurement protocol. (Data taken from publication 3.4).

2.3 Single-Molecule Force Spectroscopy

Single-molecule force spectroscopy today encompasses a variety of techniques, capable of studying molecular force responses over a wide range of forces. While forces below ~ 25 pN are best studied by magnetic or optical tweezer assays, AFM-based SMFS is uniquely suited to study systems in the force range from ~ 50 pN up to nanonewtons [8]. This makes it an ideal technique to study the mechanical properties of cellulosomal components, which can differ strongly in their capability to withstand forces. For example, a cohesin-dockerin pair from *Clostridium perfringens* [9] was found to unbind at approximately ~ 60 pN. These unbinding forces are surpassed more than ten-fold by the cohesin-dockerin systems studied within this thesis (using similar pulling velocities and force probes). AFM based SMFS is most commonly conducted in three different measurement modes: Either the distance between cantilever and surface is increased with constant velocity ("constant speed"), or the force is either increased linearly over time ("force ramp") or kept constant ("force clamp").

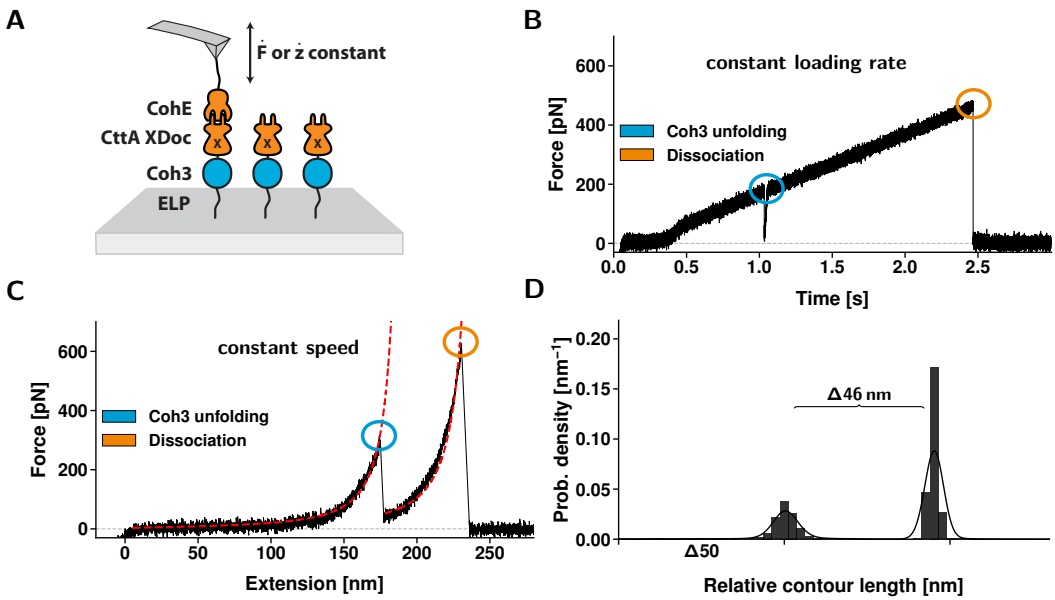


Figure 2.4. Ilustration showing the experimental arrangement of a receptor-ligand based SMFS experiment, as well as example curves for the constant speed and constant loading rate measurement modes. (**A**) Schematic illustration of the experimental setup. The Ctta XDoc:CohE (*R.f*) receptor ligand system was used to specifically and reliably probe Coh3 (*A.c.*), which was expressed as a fusion protein with XDoc. After each measurement, the surface is moved laterally to avoid probing the same Coh3 twice and the retraction speeds and loading rates are varied. (**B, C**) Example traces showing Coh3 unfolding in constant loading rate and constant speed mode. Unfolded peptide behaves as an entropic spring, as fits of a worm-like chain model (red lines) to both stretches in (C) illustrate. (**D**) Contour length transformation of the constant loading rate trace in (B), illustrating how Coh3 unfolding events are assigned if they match the expected contour length increment of 46 nm. Figure adapted from publication 3.5.

2.3.1 Measurement Modes

Constant Speed

Because of the ease of implementation, this is by far the most commonly used measurement protocol. Accurate distance control is easily achieved using piezo-actuators with additional position sensors, and by controlling the distance, the duration of a measurement cycle is limited by the time it takes for the retraction distance to approach the contour length of the stretched molecule. SMFS experiments usually rely on linkers to mitigate surface interactions, which have a force response similar to an entropic spring. Therefore, force increases with distance in a non-linear fashion.

Force Ramp

AFM-based force ramp SMFS requires a much more carefully designed feedback mechanism in comparsion to constant speed SMFS. Abrupt changes in force proceeding e.g. protein unfolding events demand a well tuned control loop in order to restore the set force quickly, without overshooting it. Piezoelectric elements that exhibit small hystereses are favorable, allowing purpose-build instruments to compensate for sudden changes in force within $\sim 10\,\mathrm{ms}$ [10]. Importantly, force ramp mode allows to record data at comparably low loading rates, as demonstrated in publication 3.5. This effect is especially pronounced for high-force complexes, as in constant speed mode force increases steeply, when the extension approaches the contour length of the system probed.

Force Clamp

In contrast to constant speed or constant loading rate mode, force clamp SMFS provides direct access to life time statistics. This is essential when investigating phenomena like catch-bonding [11], where the defining criterium is a bond lifetime that increases with force, setting it apart from common slip-bonds. However, force clamp SMFS becomes challenging when an investigated system consists of several components with differing life times.

2.3.2 Kinetic Rate Models and Polymer Elasticity

In an effort to quantitatively assess force-biased molecular unfolding or unbinding, kinetic rate models of varying complexity have been developed [3, 12–14]. The most commonly employed model was developed by Bell, Evans, and Schulten. It models a unbinding or unfolding process as the thermal crossing of a one-dimensional free energy barrier, separating two states. Equation 2.5 gives an expression for the off-rate from the bound state. In the Bell picture [15], ν_0 is the microscopic attempt frequency, that can be combined with the energy barrier height ΔG to form the zero-force off-rate k^0_{off} along the reaction coordinate Δx. The application of force now reduces the energy barrier height, thus increasing the off-rate.

$$k_{off}(F) = \nu_0 \left(\frac{F\Delta x}{k_B T} - \frac{\Delta G}{k_B T} \right) = k^0_{off} exp \left(\frac{F\Delta x}{k_B T} \right) \tag{2.5}$$

To be applicable by the experimentalist, theory must be relatable to quantities that can be reliably extracted from measured data. Under the assumption of a constant force loading rate $\dot{F}$, analytical expressions for the barrier crossing forces (eq. 2.6) as well as for the most probable crossing force (eq. 2.7) can be derived.

$$p(F) = \frac{k^0_{off}}{\dot{F}} \exp\left(\frac{\Delta x F}{k_B T} - k^0_{off} k_B T \frac{\exp\left(\frac{\Delta x F}{k_B T}\right) - 1}{\Delta x \dot{F}}\right) \tag{2.6}$$

$$F_{most\_probable}(\dot{F}) = \frac{k_B T}{\Delta x} \log\left(\frac{\dot{F}\Delta x}{k^0_{off} k_B T}\right) \tag{2.7}$$

Linker Corrections to the Bell-Evans Model

While these expressions are directly applicable to data recorded in force ramp mode, constant speed experiments will usually not provide constant force loading rates due to the elastic response of linker molecules. Several groups have sought to quantify the resulting deviations [16, 17], and to provide experimentalists with theory consistent with non-constant loading rates that result from entropic springs. Ray et al. have derived an analytical expression for a bond rupture distribution $p(F)$, that is derived for the force-distance response of a worm-like chain [18]. Considering the forces reached in some receptor-ligand systems (e.g. publication 3.3), this approach can be used to derive a similar expression for a polymer model that describes the elastic response of a peptide linker in the high-force limit more accurately. Using the polymer elasticity response for high forces from Livadaru et al. [19], $x = L_c\left(1 - \frac{k_b T}{cFb}\right)$ and using $F_b = \frac{k_b T}{b}$ as a lower bound, the bond survival probability $S(t)$ can be calculated analogous to Ray et al. [18]:

$$\begin{aligned} S(t) &= \exp\left[-\int \frac{1}{\dot{f}} k^0_{off} \exp(f/F\dagger) df\right] \\ &= \exp\left[-k^0_{off}\left(\int_0^F \frac{1}{k_c v} e^{f/F^\dagger} df + \int_{F_b}^F \frac{F_b}{f^2} e^{f/f^\dagger} df\right)\right] \\ &= \exp\left[-\frac{k^0_{off} F^\dagger}{v k_c}\left(e^{F/F^\dagger}\left(1 - \frac{L_c F_b k_c}{c F_\dagger F}\right) + \frac{L_c F_b k_c}{c F^{\dagger 2}}\left(\frac{F^\dagger}{F_b} e^{F_b/F^\dagger} + \mathrm{Ei}\left(\frac{F}{F^\dagger}\right) - \mathrm{Ei}\left(\frac{F_b}{F^\dagger}\right)\right) - 1\right)\right] \end{aligned} \tag{2.8}$$

Here, k_c denotes the cantilever spring constant, Ei the exponential integral, and v the retraction velocity. By performing the derivative $-\frac{dS(F)}{dF} = p(F)$, we get Eq.2.9.

$$\begin{aligned} p(F) = &\frac{k^0_{off}}{v k_c}\left(1 + \frac{L_c F_b k_c}{c F^2}\right) \exp\left[\frac{F}{F^\dagger} - \frac{k^0_{off}}{v k_c}\left(F^\dagger e^{F/F^\dagger}\left(1 - \frac{L_c F_b k_c}{c F^\dagger F}\right) + \right.\right. \\ &\left.\left. \frac{k_c L_c}{c} e^{F_b/F^\dagger} + \frac{L_c F_b k_c}{c F^\dagger}\left(\mathrm{Ei}\left(\frac{F}{F^\dagger}\right) - \mathrm{Ei}\left(\frac{F_b}{F^\dagger}\right)\right) - F^\dagger\right)\right] \end{aligned} \tag{2.9}$$

Part II

Results

3
Research Articles

3.1 Load-Tightening of a Protein-Protein Binding Interface

This publication explores the mechanical resilience of a receptor-ligand pair, which anchors the cellulose-adhesion machinery of a cellulose degrading bacterium to its cell wall. Its ability to withstand exceptionally high forces, which may have evolved to withstand the turbulent environment of the bovine rumen, is explored with a combined approach of AFM based SMFS and all-atom MD simulations. Comparing how the contact surface area between the recetor-ligand pair changes when under force, we surprisingly find that it increases with force. My contribution to this work includes the preparation of Cohesin E in its native anchoring geometry as well as performing SMFS experiments and analyzing the recorded data.

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

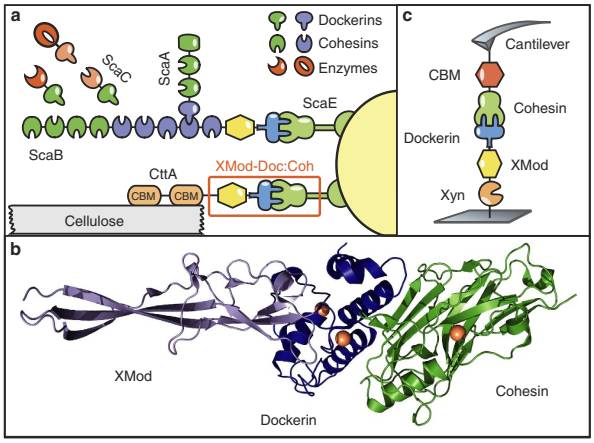


Figure 1 | System overview. (**a**) Schematic of selected components of the *R. flavefaciens* cellulosome. The investigated XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose is highlighted in orange. (**b**) Crystal structure of the XMod-Doc:Coh complex. Ca^{2+} ions are 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.

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms6635 ARTICLE

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 $k_{off} = 7.3 \times 10^{-7}\,s^{-1}$ for an intact XMod, and $\Delta x = 0.19$ nm and $k_{off} = 4.7 \times 10^{-4}\,s^{-1}$ 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 $k_{off} = 2.6 \times 10^{-6}\,s^{-1}$. The loading

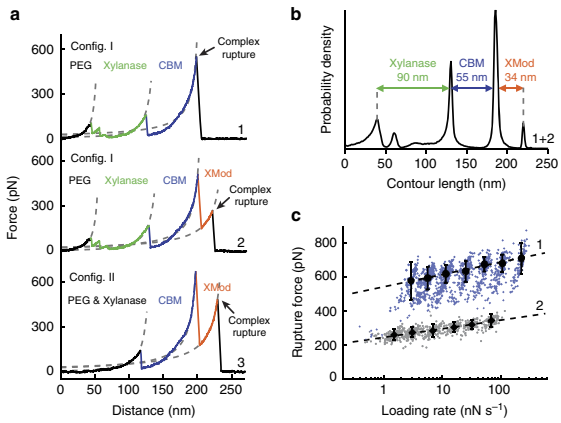


Figure 2 | Experimental SMFS unfolding traces. (**a**) Unfolding fingerprints from pulling configuration I (curves 1 & 2) and configuration II (curve 3). The QM-FRC model (dashed lines) was used to estimate the contour lengths of the unfolded modules. (**b**) Contour length histogram obtained from 127 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.

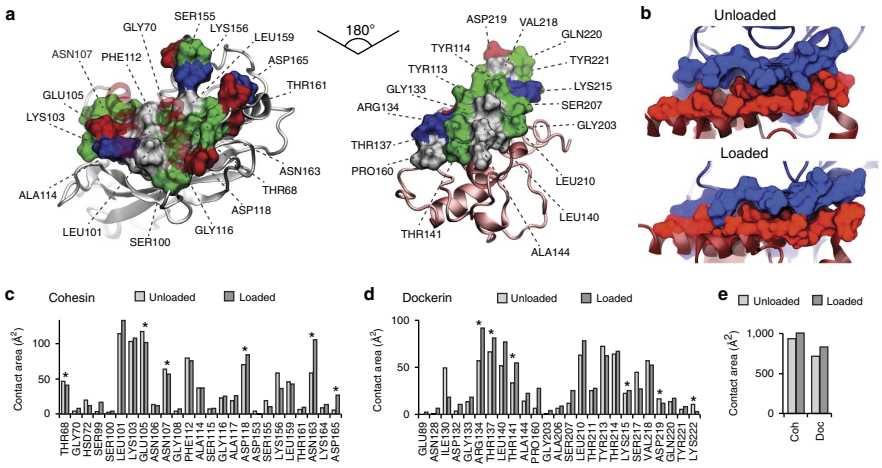


Figure 3 | Analysis of binding interface and catch bond mechanism from SMD. (**a**) Surface plots for the main interacting residues of Coh (left) and Doc (right). Hydrophobic residues are shown in grey, polar residues in green, and negative and positive residues in red and blue, respectively. Both Coh 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 $\sim 4{,}000$ 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

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms6635 ARTICLE

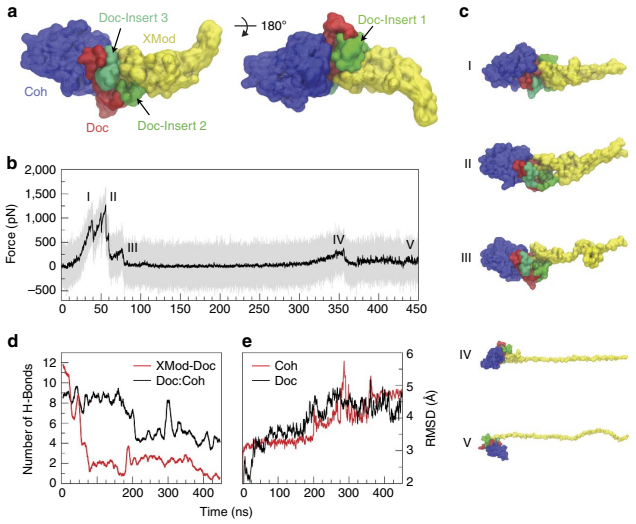


Figure 4 | SMD shows unfolding of XMod destabilizes Doc:Coh binding interface. XMod was unfolded by moving the harmonic restraint to the C terminus of XMod while the N terminus was moved at 0.625 Å ns^{-1}. (**a**) Surface representation of XMod-Doc:Coh complex with Doc insert sequences. Coh is shown 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.

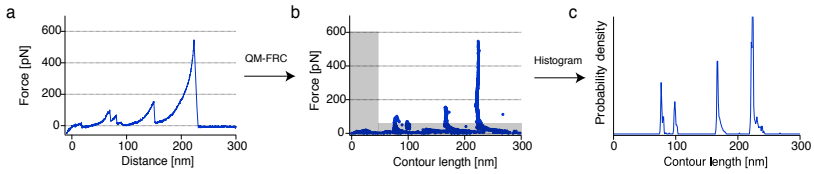
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

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

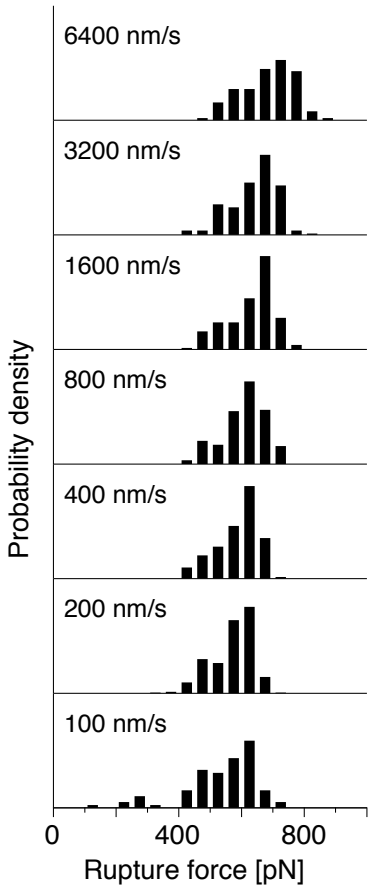
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

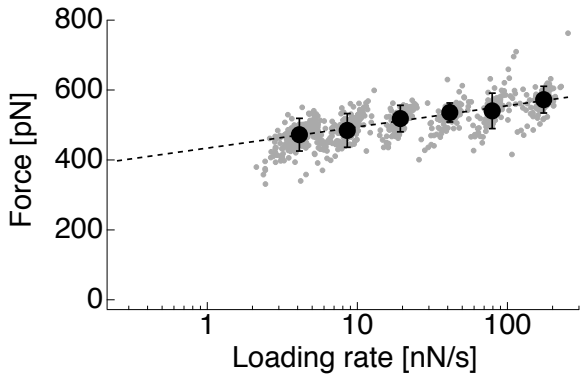


Supplementary Fig. 1: Assembly of contour length histograms. **a** Force-extension traces are transformed into contour length space using a QM-corrected FRC model with parameters $\gamma = 41°$, and $b = 0.11\,\text{nm}$. **b** In force-contour length space, force and contour length thresholds are applied 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.



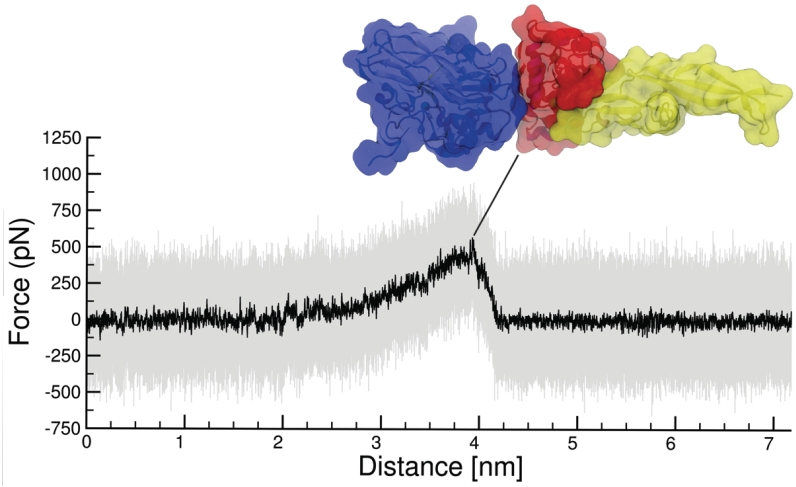
Supplementary Fig. 2: Complex rupture force histograms for pulling speeds ranging from 100 $\mathrm{nm\,s^{-1}}$ to 6400 $\mathrm{nm\,s^{-1}}$. Pulling speeds are indicated next to the histograms. Only traces with an intact XMod were taken into account (no XMod unfolding observed, corresponding to Fig.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

3



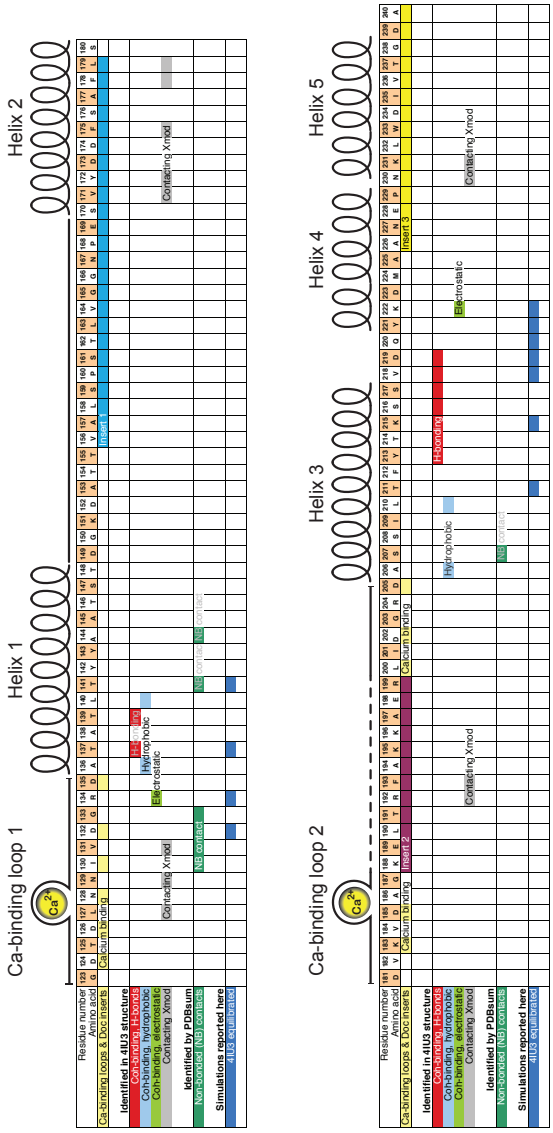
Supplementary Fig. 3: Dynamic force spectrum for XMod unfolding obtained from 654 force-extension traces. The gray points show single XMod unfolding events. Black circles represent the most probable rupture forces and loading rates obtained by Gaussian fitting at each pulling speed. Error bars are ± 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 $k_{off} = 2.6 \times 10^{-6}\,\mathsf{s}^{-1}$.

4



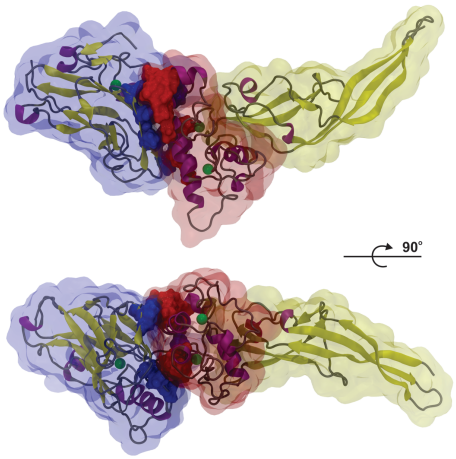
Supplementary Fig. 4: Force distance trace obtained by SMD at a pulling speed of 0.25 Å ns^{-1}. Force values at each time step are shown in gray, with average force calculated every 200 ps in black. The inset is a snapshot of the XMod-Doc:Coh complex immediately prior to rupture. XMod is shown in yellow,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

5



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.

6



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 $\Delta L_E = N_A \cdot 0.365\,\mathrm{nm} - L_F$, where L_F is the folded length, N_A is the number of amino acids, and $0.365\,\mathrm{nm}^2$ 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

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

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.

9

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).

3.2 Force Propagation and the Influence of Anchoring Geometry

In an effort to better understand the molecular mechanisms, which allow the CohE:CttA-XDoc receptor-ligand pair to withstand remarkable forces (as we had found in a previous study (3.1)), in this article we investigate how force propagates through a protein complex under load. Using all-atom MD simulations, we use network-based correlation analysis to identify stiff paths through which force is transmitted. We find that for the natively occuring anchoring geometry, force is directed with strong normal components through the complex' binding interface. When anchoring the Cohesin non-natively at its N-terminus, the complexes ability to withstand force is compromised. I contributed to this manuscript by performing SMFS experiments together with Constantin Schöler. Additionaly, Wolfgang Ott and I cloned and expressed CohE with tags in both anchoring geometries, native and non-native. I also contributed to the analysis and interpretation of SMFS data.

Mapping Mechanical Force Propagation through Biomolecular Complexes

Constantin Schoeler[†], Rafael C. Bernardi[†], Klara H. Malinowska,
Ellis Durner, Wolfgang Ott, Edward A. Bayer,
Klaus Schulten, Michael A. Nash, and Hermann E. Gaub

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

published in

Nano Lett. 2015, 15, 7370-7376

NANO LETTERS

Letter

pubs.acs.org/NanoLett

Mapping Mechanical Force Propagation through Biomolecular Complexes

Constantin Schoeler,[†,#] Rafael C. Bernardi,[‡,#] Klara H. Malinowska,[†] Ellis Durner,[†] Wolfgang Ott,[†,§] Edward A. Bayer,[∥] Klaus Schulten,[‡,⊥] Michael A. Nash,[*,†] and Hermann E. Gaub[†]

[†]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

[‡]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

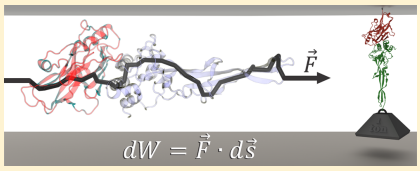
[§]Center for Integrated Protein Science Munich (CIPSM), University of Munich, 81377 Munich, Germany

[∥]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

[⊥]Department of Physics, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Here we employ single-molecule force spectroscopy with an atomic force microscope (AFM) and steered molecular dynamics (SMD) simulations to reveal force propagation pathways through a mechanically ultrastable multidomain cellulosome protein complex. We demonstrate a new combination of network-based correlation analysis supported by AFM directional pulling experiments, which allowed us to visualize stiff paths through the protein complex along which force is transmitted. The results implicate specific force-propagation routes nonparallel to the pulling axis that are advantageous for achieving high dissociation forces.



KEYWORDS: *Force propagation, single molecule force spectroscopy, steered molecular dynamics, network analysis, cohesin−dockerin*

Mechanical forces play a fundamental role in biological systems. Cells are able to sense and respond to mechanical cues in their environment by, for example, modulating gene expression patterns,[1] reshaping the extracellular matrix,[2] or exhibiting differential biochemical activities.[3] At the molecular level, these behaviors are governed by mechanically active proteins. Such proteins are able to sense and respond to force by undergoing conformational changes,[4] exposing cryptic binding sequences,[5] acting synergistically with ion channels,[6] or modulating their function in a variety of ways.[7−9]

Experimental methods including AFM single-molecule force spectroscopy (SMFS) allow direct measurement of molecular mechanical properties. These studies have demonstrated the importance of the shear topology involving parallel breakage of hydrogen bonds in providing mechanical stability to protein folds.[10,11] Many globular domains and protein complexes also exhibit a directional dependence in unfolding mechanics, consisting of stiff and soft axes.[12−18] Pulling geometry can be defined by controlling the positions of the chemical linkages between protein monomer units through a variety of bioconjugate techniques.

Primary sequences of mechanically active proteins are extremely diverse, essentially rendering them undetectable by conventional bioinformatics approaches. Yet, another computational approach, namely, molecular dynamics (MD), allows sampling of structural conformations of large and frequently mechanostable protein complexes.[19,20] Analysis of these conformations from MD trajectories have recently led to the development of network-based correlation methods for investigating signal transmission and allosteric regulation in proteins.[21−23] In network models, local correlations of positional fluctuations in a protein are represented as a web of inter-residue connections. Within such a network, the behavior of nodes that are highly correlated and within close physical proximity can be analyzed to obtain the shortest path between two network nodes (i.e., amino acids). This analysis helps to identify which connecting residues are most important for intramolecular communication.[23−25] Examination of multiple pathways, also known as suboptimal paths, within an acceptable deviation from the optimal path helps to detect the web of nodes critical for transmission of information.

Among MD methods, steered molecular dynamics (SMD) simulations in which external forces are used to explore the response and function of proteins have become a powerful tool especially when combined with SMFS.[6] SMD has been successfully employed in a wide range of biological systems, from the investigation of protein mechanotransduction,[5,26] to

Received: July 9, 2015
Revised: August 5, 2015
Published: August 11, 2015

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Nano Letters Letter

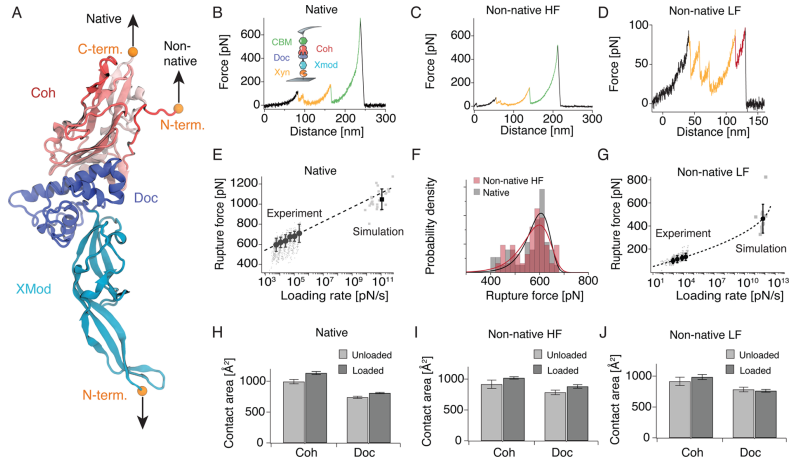


Figure 1. Single molecule force spectroscopy and steered molecular dynamics of XMod-Doc:Coh in two pulling configurations. (A) Crystal structure of the XMod-Doc:Coh complex (PDB 4IU3) with orange spheres marking the termini where force was applied. (B) Experimental unfolding trace for the native pulling configuration at a pulling speed of 1600 nm s^{-1}. The inset shows a schematic of the pulling geometry. Unfolding signatures of the Xyn and CBM marker domains are marked in orange and green, respectively. (C) Experimental unfolding trace for the non-native high force class obtained at a pulling speed of 700 nm s^{-1}. (D) Experimental unfolding trace for the non-native low force class obtained at a pulling speed of 700 nm s^{-1}. The additional 17−19 nm contour length increment attributed to N-terminal Coh unfolding is shown in red. (E) Dynamic force spectrum for XMod-Doc:Coh unbinding in the native geometry obtained from experiment and simulations. Gray points and squares represent the rupture force/loading rate pairs obtained from experiment and simulation, respectively. Black circles represent the most probable rupture force/loading rate obtained from Gaussian fits to the experimental data at six pulling speeds. The black square shows the mean rupture force and loading rate for the simulated rupture events. (F) Rupture force histograms obtained at a pulling speed of 800 nm s^{-1} for the native (gray, $n = 46$) and non-native high force class (red, $n = 48$). Fitted probability densities $p(F)$ are shown as solid black and red lines. Data for both pulling configurations were obtained with the same cantilever to minimize calibration errors. (G) Dynamic force spectrum for XMod-Doc:Coh unbinding in the non-native low force class obtained from experiments and simulation. The same representation as in (E) is used. (H,I,J) Unloaded and loaded surface contact areas for the different pulling geometries ((H) native, (I) non-native high force class, and (J) non-native low force class).

permeability of membrane channels,[27,28] and the characterization of protein−receptor interactions.[29] SMD simulations have also been used to study force propagation through proteins by employing force distribution analysis (FDA).[30,31] In FDA, all pairwise forces, which are usually calculated in MD simulations, are stored in $N \times N$ matrices, where N is the number of atoms.[32] These pairwise forces can then be used to assess a protein's response to a mechanical or allosteric signal.[33] In the FDA approach, atoms under mechanical strain are identified by subtracting forces of both loaded and unloaded states for each pair of interacting atoms.[31] However, to achieve a sufficient signal-to-noise ratio, FDA will often require exhaustive sampling of the conformational space.[32,34] FDA, therefore, requires more computational resources than usual SMD studies, which are frequently already computationally demanding. There is therefore a clear need for new analysis methods that enable visualization of force propagation pathways from a single SMD trajectory.

Here we implemented a novel combination of SMD, network-based correlation analysis, and thermodynamic fluctuation theory, supported by AFM-SMFS experiments to study force propagation through a protein complex subjected to different pulling geometries. We chose an ultrastable receptor−ligand interaction as a model system because of its remarkably high mechanical stability,[29] which effectively improves the signal-to-noise ratio. This complex consists of two interacting protein domains called cohesin (Coh) and dockerin (Doc) that maintain bacterial adhesion of *Ruminococcus flavefaciens* to cellulosic substrates. Doc is found within the same polypeptide chain as a stabilizing ancillary domain called X-module (XMod), located N-terminally of Doc. Based on its position with the *R. flavefaciens* cellulosomal network, Coh is mechanically anchored *in vivo* at its C-terminal end to the cell surface. Our prior work demonstrated that, when force is applied to the complex in the native configuration (i.e., C-terminal Coh, N-terminal XMod-Doc anchor points), the complex is extremely stable, exhibiting high rupture forces of 600−750 pN at loading rates from 1−100 nN s^{-1}.[29] Since the bulk equilibrium affinity of the complex is an unremarkable 20 nM,[35] we hypothesized that the high mechanostability is explained by a catch bond mechanism. AFM rupture force data and SMD simulations supported this prediction, where it was observed that the contact surface area of the two proteins increased as mechanical force was applied.

To characterize the mechanisms behind Coh:Doc high stability, here we additionally pulled the complex apart in a non-native configuration (i.e., N-terminal Coh, N-terminal XMod-Doc anchor points). In the non-native pulling

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

configuration, we found that the complex dissociated along two competing pathways with very different mechanical characteristics.

Our new dynamic network analysis protocol reveals how different mechanical behaviors are attributable to differences in the direction of force transmission across the binding interface. Together, the experiments and simulations depict a simple physical mechanism for achieving high complex rupture forces: the complex directs force along pathways orthogonal to the pulling axis.

Single-Molecule Pulling Experiments and SMD. For SMFS experiments, XMod-Doc was produced as a fusion protein with an N-terminal Xylanase (Xyn) domain. Coh was produced as either an N- or C-terminal fusion domain with a carbohydrate binding module (CBM). These fusion domains were used for site specific immobilization to a glass surface and AFM cantilever to achieve the two loading configurations shown in Figure 1A and further served as marker domains with known unfolding length increments to validate single-molecule interactions and sort SMFS data traces.[36]

For the native pulling configuration found *in vivo*, Coh-CBM and XMod-Doc are loaded from their C- and N-termini, respectively (Figure 1A). A representative unbinding trace for the native pulling configuration is shown in Figure 1B. We measured the loading rate dependence of complex rupture using both experimental and SMD data sets (unbinding trace from SMD shown in Figure 3A) and plotted them on a combined dynamic force spectrum (Figure 1E). The linear Bell model produced fit parameters for the effective distance to the transition state $\Delta x = 0.13$ nm, and the zero-force off rate $k_{off} = 4.7 \times 10^{-4}$ s^{-1}. Both experimental and simulation data are well described by a single Bell expression, despite the differences in loading rates between experiments and simulation. The observation suggests that the application of force does not significantly change Δx for this particular configuration.

To test the influence of pulling geometry on mechanical stability, we performed SMFS and SMD on the system where Coh was pulled from the opposite terminus (i.e., non-native N-terminus, cf. Figure 1A). Unlike the native pulling geometry, this geometry exhibited two clearly distinct unbinding pathways that are characterized by different force ranges (high or low) at which the complex dissociated. We refer to these pathways as non-native high force (HF) (Figure 1C) and non-native low force (LF) (Figure 1D).

AFM data traces classified as non-native HF showed similar characteristics as those in the native pulling configuration (cf. Figure 1B,C,F). The non-native LF traces, however, exhibited a markedly different unfolding behavior (Figure 1D). Xyn unfolding (highlighted in orange) was regularly observed, but CBM unfolding was only very rarely observed. The complex usually did not withstand forces high enough to unfold CBM when rupturing along the non-native LF path. Among non-native LF curves, we regularly found an additional contour length increment of 17−19 nm consistent with unfolding of ∼60 amino acids located at the N-terminus of Coh. This unfolding occurred immediately following Xyn unfolding (Figure 1D, red), or alternatively prior to Xyn unfolding, or with a substep (Supplementary Figure S1). Taken together, it appears that partial Coh unfolding from the N-terminus destabilizes the complex, causing lower rupture forces (Figure 1G).

The experimental rupture forces from the non-native HF class were indistinguishable from those arising in the native configuration. To confirm this, we performed additional measurements where both Coh configurations were alternately probed with the same Xyn-XMod-Doc functionalized cantilever (Supplementary Figure S2), eliminating inaccuracies introduced through multiple cantilever calibration. Most probable rupture forces at a pulling speed of 800 nm s^{-1} of 606 and 597 pN for the native configuration and non-native HF class, respectively, were determined in the Bell Evans model (Figure 1F, Supplementary eq S2), demonstrating that the native and non-native HF classes are experimentally indistinguishable.

For the LF class, we analyzed the final complex rupture event and plotted the combined dynamic force spectrum (Figure 1G). Here, simulated and experimentally observed data were not well described by a single Bell expression. In such cases nonlinear models have been developed to obtain kinetic and energetic information from dynamic force spectra.[37,38] To fit the combined data, we used the nonlinear Dudko−Hummer−Szabo (DHS) model (Supplementary eq S3) and obtained values of $\Delta x = 0.42$ nm and $k_{off} = 0.005$ s^{-1}. The DHS model further provides the free energy difference ΔG between the bound state and the transition state as a fit parameter, which was found to be $\Delta G = 129\ k_B T$. The model fit produced a distance to transition that was much longer than observed for the native configuration. Independent SMD simulations for the non-native pulling configuration were found to also lead to HF and LF unbinding scenarios (see below, Figure 4A,D, respectively).

The differential solvent contact area was calculated from SMD simulations to estimate the intermolecular contact area in the Doc:Coh complex. In the native configuration, the simulated Doc:Coh contact area increased by 14% and 9% for Coh and Doc, respectively (Figure 1H). For the non-native HF class, the contact area increased by 11% and 12% for Coh and Doc, respectively (Figure 1I). In the non-native LF class, the contact area increased by only 7% for Coh and decreased by 3% for Doc (Figure 1J). Evidently, an increased surface contact area for Doc in the native and non-native HF pathways correlated with high mechanostability of the system.

Force Propagation Theory: A Simple Model. To further understand the observed unbinding pathways, we sought to identify paths through the molecule along which the externally applied load propagates. From thermodynamic fluctuation theory,[39,40] it is known that the correlation of fluctuations of atoms i and j and the force $\mathbf{F}_i$ on atom i are related through

$$\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{1}$$

where $\Delta \mathbf{r}_i = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$ and $\mathbf{r}_i$ is the position of atom i. The derivative on the right-hand side of eq 1 states that neighboring atoms i and j will move with high correlation due to an external force $\mathbf{F}_i$ acting on atom i if the coupling between them is strong. Hence, a given element of a correlation matrix $M_{ij} = \langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle$ will be large in the case of a strong interaction potential between i and j. When force is propagated through a molecule, soft degrees of freedom will be stretched out along the path of force propagation, while stiff degrees become more important for the dynamics of the system.

Consequently, paths with high correlation of motion describe the paths along which force propagates through the system. To illustrate this behavior for a toy system, we employed the NAMD[41] SMD[42] constant velocity protocol to a test pattern of identical spheres connected with harmonic springs of different

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Nano Letters Letter

stiffness (Figure 2A). The position of one sphere was fixed during the simulation, while another sphere on the opposite side of the structure was withdrawn at constant velocity. The strained structure at the end of the simulation is shown in Figure 2B. We assigned weights to the lines between spheres according to the Pearson correlation coefficient C_{ij} (Supplementary eq S4) between those network nodes (Figure 2C). The Pearson correlation coefficient differs from the left-hand side of eq 1 by a normalization factor $(\langle \Delta \mathbf{r}_i^2(t)\rangle \langle \Delta \mathbf{r}_j^2(t)\rangle)^{-1/2}$ and was chosen to make our analysis mathematically more tractable. For a detailed discussion on this choice of correlation measure, see Supporting Information. In a harmonic potential approximation, the equipartition theorem can be applied to this normalization factor resulting in the following expression for C_{ij}:

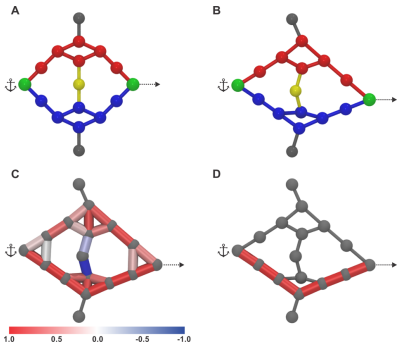


Figure 2. Network analysis test simulation. (A) Simulated pattern of atoms depicted by spheres. Connecting lines between atoms represent harmonic springs with different stiffnesses (red, k; blue, $5k$; yellow, $7.5k$; black, $10k$). The green atom was fixed (anchor), while a second green atom was withdrawn at constant speed (arrow). Black and yellow atoms and their adjacent springs were introduced to maintain the general shape of the pattern. (B) Deformed sphere pattern at the end of the simulation. (C) Edges between nodes are weighted by the corresponding correlation matrix elements. (D) The path with highest correlation of motion is shown in red.

$$C_{ij} = \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i}\sqrt{k_{i,\mathrm{eff}}k_{j,\mathrm{eff}}} \tag{2}$$

where $k_{i,\mathrm{eff}} = \left(\frac{1}{k_{x_i}} + \frac{1}{k_{y_i}} + \frac{1}{k_{z_i}}\right)^{-1}$ and k_{x_i} is the curvature of the potential on atom i in the x direction. For a full derivation, see Supporting Information. Equation 2 illustrates how Pearson correlation is a suitable measure to identify the stiff paths in our simple model. We then used dynamical network analysis implemented in VMD[49] to find the path of highest correlation (Figure 2D). As expected from eq 1, we found this path to be the one connected by the stiff springs.

Force Propagation through XMod-Doc:Coh Complex. The simple pattern of spheres validated our general approach of using local correlations to identify load-bearing pathways through networks. We next employed dynamical network analysis to understand force propagation through the XMod-Doc:Coh complex.

The dynamic networks for the native configuration (unloaded and loaded) are shown in Figure 3B,C, respectively. While the network shows multiple suboptimal paths in the unloaded scenario, the loaded case exhibits a well-defined main path along which force propagates through the system. Interestingly, in the loaded configuration, force propagates through both binding helices of Doc, which results in a force path with large normal components to the unbinding axis close to the binding interface as illustrated in Figure 3D. It had been shown for another ultrastable protein, namely, silk crystalline units, that curving force paths distribute tension through the entire system.[31] A strategy that assumes an indirect path would therefore allow the system to have more time to absorb the tension from the applied force. The result here supports the view that directing the force along a path with significant perpendicular components to the pulling axis leads to high mechanical stability. In a simple mechanical picture, a certain amount of mechanical work, namely $dW = \mathbf{F}\cdot d\mathbf{s}$, is required to separate the two binding interfaces by a distance Δz and break the interaction. In this simplified picture, $d\mathbf{s}$ points along the unbinding axis, whereas the force $\mathbf{F}$ is locally largely perpendicular to this direction. Consequently, a larger force is required to break the interaction than in a scenario where the force path would point along the unbinding axis.

To validate this picture, we repeated the same analysis for the non-native HF and non-native LF pathways. The HF simulation (Figure 4A) exhibited only a small stretching of the flexible N-terminal region of Coh and complex dissociation at approximately 800 pN and a pulling distance around 10 nm. However, the LF case shown in Figure 4D exhibited a stepwise N-terminal Coh unfolding, dissociating at a force of about 480 pN at a pulling distance of about 25 nm. This behavior confirmed our assignment of the experimentally observed 17–19 nm contour length increment to Coh unfolding up to residue 62 in PDB 4IU3.

While the experimental data did not show a detectable difference between the native configuration and the non-native HF class, the propagation of force takes place along a different pathway (Figure 4B). For N-terminal Coh pulling, helix 3 of Doc is not involved in the propagation of force as it is for the native geometry. In the native configuration, force propagates through the center of Coh, while for non-native HF the path is shifted toward the side of the molecule. Despite these differences, there is a common feature between the native and non-native HF pathways. At the binding interface, the pathway again shows pronounced components perpendicular to the unbinding axis (cf. Figure 4C), suggesting that this feature is indeed responsible for the exceptional mechanical strength observed for these two unbinding pathways.

Figure 4E shows the force propagation pathway for the non-native LF class prior to rupture. Due to the unfolding of the N-terminal Coh segment, the propagation of force is shifted even further away from the central portion of Coh than for the non-native HF class. Interestingly, force is propagated through the small helical segment of Coh (ALA167-GLN179), a portion of the molecule that is not involved in force propagation for any of the other analyzed trajectories. Unlike in the aforementioned scenarios, there is no pronounced tendency for perpendicular force components at the binding interface for the non-native LF class. In fact, the force is propagated along a path largely parallel to the pulling axis (cf. Figure 4F). In cases where force

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Nano Letters Letter

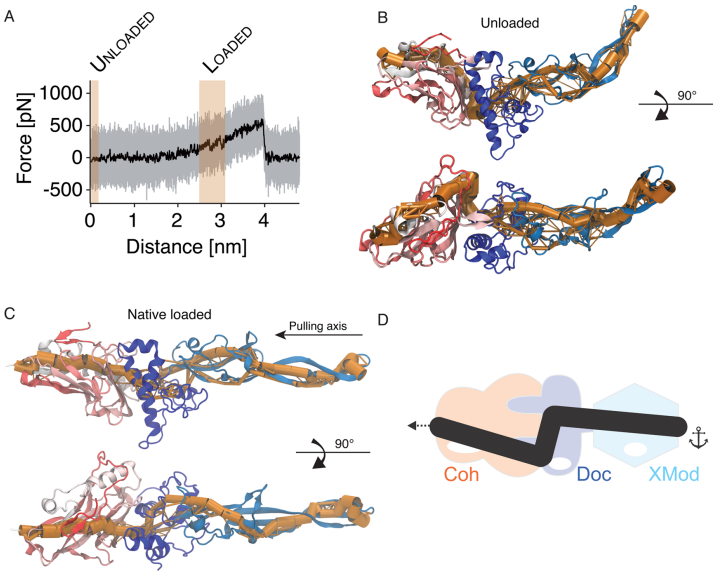


Figure 3. Force propagation through XMod-Doc:Coh in the native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray. The black line represents a moving average with a box size of 500 steps. The highlighted red areas denote the windows where dynamic networks and contact areas were calculated. (B) Network paths for the unloaded system. The thickness of the orange tube represents the number of suboptimal correlation paths passing between two nodes. (C) Network paths for the loaded system. A detailed 2D representation of the pathway, highlighting the amino acids present in the pathway, is shown in Supplementary Figure S5. (D) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis.

propagation occurs parallel to the pulling axis, as in Figure 4E, low mechanical stability was observed.

The aforementioned force propagation architecture along with the effect of increasing contact surface area upon mechanical loading combine for elevated mechanostability of the system.[29] In cases where we observed an N-terminal Coh unfolding of 62 amino acids in the non-native geometry, the system was no longer able to summon this mechanism, causing dissociation at much lower forces.

Previously, our groups have reported on a family of mechanically stable protein ligand receptor complexes that are key building blocks of cellulosomes,[29,44−46] the multi-enzyme complexes used by select anaerobic bacteria to digest lignocellulose. However, the molecular origins of the stability of these complexes remained largely unclear. An initial clue was obtained when, in a previous work, we were able to show that contact surface area of the two proteins increased as mechanical force was applied.[29] In a different study,[47] coarse-grained MD simulations showed much smaller rupture forces at similar loading rates both for native and non-native pulling than we report here. This disagreement is likely due to the inability of the coarse-grained model to capture the rearrangement of amino acid side chains observed here. As we demonstrated, force propagation calculation from network-based correlation analysis helped in investigating the dramatic effect on the mechanical stability of the Doc:Coh interaction when different pulling geometries are applied. Our methodological approach, to the best of our knowledge, has never been applied even though network analysis of SMD trajectories was performed before to probe the mechanism of allosteric regulation in imidazole glycerol phosphate synthase.[48]

In summary, for both unbinding cases where we observed high mechanostability, we found that across the binding interface, force propagated along paths with strong normal components to the pulling direction. Such a behavior was not observed for the non-native LF class, where, presumably due to N-terminal Coh unfolding, the system was no longer able to direct the force across the binding interface at high angles. From these findings, we conclude that the ultrastable complex formed by Coh and Doc achieves its remarkable mechano-stability by actively directing an externally applied force toward an unfavorable angle of attack at the binding interface, consequently requiring more force to achieve a given amount of separation along the pulling direction. Our results show that this mechanically stable complex uses an architecture that exploits simple geometrical and physical concepts from Newtonian mechanics to achieve high stability against external forces. The analytical framework derived here provides a basis for developing a deeper understanding of the functioning of various mechanoactive proteins that are crucial for physiolog-

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Nano Letters Letter

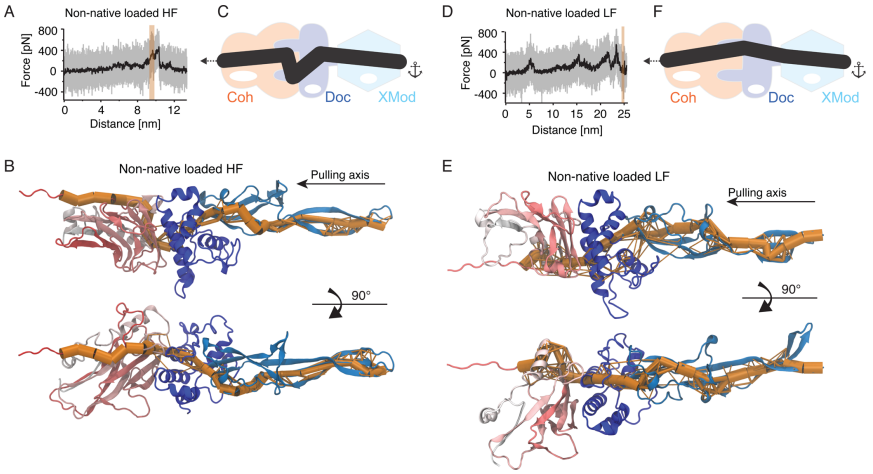


Figure 4. Force propagation through XMod-Doc:Coh in the non-native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500 steps. Note that this computational pulling experiment revealed a high-force behavior. (B) Network pathways calculated from dynamical network analysis for the non-native HF trajectory. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S6. (C) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis. (D) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500. This computational pulling experiment revealed partial Coh unfolding that led to LF behavior. (E) Network pathways for the non-native LF scenario. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S7. (F) Schematic model of force propagation across the Coh:Doc binding interface. Unlike in both HF scenarios, force propagates across the binding interface mostly along the unbinding axis.

ically relevant processes such as mechanotransduction, cellular mechanosensing, and pathogenesis. Additionally, it could provide a design platform for development of artificial mechanoactive systems with applications as tissue engineering scaffolds or components in engineered nanomaterials.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.5b02727.

Details of experimental materials and methods, steered molecular dynamics simulations, and additional theoretical discussion (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.nash@lmu.de.

Author Contributions

#These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Zaida Luthey-Schulten, Helmut Grubmüller, Marcelo Melo, and Jonathan Lai for helpful discussions. The authors gratefully acknowledge funding from an advanced grant to HEG from the European Research Council (Cellufuel Grant 294438). This work was supported by grants from the National Institutes of Health (NIH, 9P41GM104601 to K.S.) and from the Energy Biosciences Institute (UCB BP 2014OOO4J01 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 NERSC/Edison supercomputer as part of the DoE ALCC program. Resources of the National Energy Research Scientific Computing Center (NERSC) are supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. 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. The authors appreciate the support of the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nano-systems: CellulosomePlus Project number: 604530 and an ERA-IB Consortium (EIB.12.022), acronym FiberFuel. 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

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

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, and from the LMU Excellence Initiative.

REFERENCES

(1) Bao, G.; Suresh, S. *Nat. Mater.* **2003**, *2*, 715−725.
(2) Grodzinsky, A. J.; Levenston, M. E.; Jin, M.; Frank, E. H. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 691−713.
(3) Kumar, S.; Weaver, V. M. *Cancer Metastasis Rev.* **2009**, *28*, 113−127.
(4) Vogel, V.; Thomas, W. E.; Craig, D. W.; Krammer, A.; Baneyx, G. *Trends Biotechnol.* **2001**, *19*, 416−423.
(5) del Rio, A.; Perez-Jimenez, R.; Liu, R.; Roca-Cusachs, P.; Fernandez, J. M.; Sheetz, M. P. *Science* **2009**, *323*, 638−641.
(6) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144−1148.
(7) Engel, A.; Muller, D. J. *Nat. Struct. Biol.* **2000**, *7*, 715−718.
(8) Noy, A.; Friddle, R. W. *Methods* **2013**, *60*, 142−150.
(9) De Yoreo, J. J.; Chung, S.; Friddle, R. W. *Adv. Funct. Mater.* **2013**, *23*, 2525−2538.
(10) Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophys. J.* **1998**, *75*, 3008−3014.
(11) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109−1112.
(12) Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. *Biophys. J.* **2012**, *103*, 2361−2368.
(13) Lee, W.; Zeng, X.; Rotolo, K.; Yang, M.; Schofield, C. J.; Bennett, V.; Yang, W.; Marszalek, P. E. *Biophys. J.* **2012**, *102*, 1118−1126.
(14) Jagannathan, B.; Elms, P. J.; Bustamante, C.; Marqusee, S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17820−17825.
(15) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12724−12728.
(16) Carrión-Vázquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738−743.
(17) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731−737.
(18) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M. *J. Am. Chem. Soc.* **2013**, *135*, 12762−12771.
(19) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64−74.
(20) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(21) Ribeiro, A. S. T.; Ortiz, V. *J. Phys. Chem. B* **2015**, *119*, 1835−1846.
(22) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. *J. Chem. Theory Comput.* **2014**, *10*, 511−517.
(23) Van Wart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. *J. Chem. Theory Comput.* **2012**, *8*, 2949−2961.
(24) Alexander, R. W.; Eargle, J.; Luthey-Schulten, Z. *FEBS Lett.* **2010**, *584*, 376−386.
(25) Chennubhotla, C.; Bahar, I. *Mol. Syst. Biol.* **2006**, *2*, 1−13.
(26) Hytönen, V. P.; Vogel, V. *PLoS Comput. Biol.* **2008**, *4*, e24.
(27) Khalili-Araghi, F.; Gumbart, J.; Wen, P.-C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol.* **2009**, *19*, 128−37.
(28) Li, W.; Shen, J.; Liu, G.; Tang, Y.; Hoshino, T. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 271−281.
(29) 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. *Nat. Commun.* **2014**, *5*, 5635.
(30) Stacklies, W.; Vega, M. C.; Wilmanns, M.; Gräter, F. *PLoS Comput. Biol.* **2009**, *5*, e1000306.
(31) Xiao, S.; Stacklies, W.; Cetinkaya, M.; Markert, B.; Gräter, F. *Biophys. J.* **2009**, *96*, 3997−4005.
(32) Stacklies, W.; Seifert, C.; Graeter, F. *BMC Bioinf.* **2011**, *12*, 101.
(33) Seifert, C.; Gräter, F. *Biophys. J.* **2012**, *103*, 2195−2202.
(34) Palmai, Z.; Seifert, C.; Gräter, F.; Balog, E. *PLoS Comput. Biol.* **2014**, *10*, e1003444.
(35) Salama-Alber, O.; Jobby, M. K.; Chitayat, S.; Smith, S. P.; White, B. A.; Shimon, L. J. W.; Lamed, R.; Frolow, F.; Bayer, E. A. *J. Biol. Chem.* **2013**, *288*, 16827−16838.
(36) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. *Biophys. J.* **2008**, *95*, 426−434.
(37) Friddle, R. W.; Noy, A.; De Yoreo, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13573−13578.
(38) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
(39) Greene, R. F.; Callen, H. B. *Phys. Rev.* **1951**, *83*, 1231−1235.
(40) Erman, B. *Phys. Biol.* **2011**, *8*, 056003.
(41) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(42) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568−1581.
(43) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.
(44) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. *Nat. Methods* **2014**, *11*, 1127−1130.
(45) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20431−20436.
(46) Jobst, M. A.; Schoeler, C.; Malinowska, K. H.; Nash, M. A. *J. Visualized Exp.* **2013**, *82*, e50950.
(47) Wojciechowski, M.; Thompson, D.; Cieplak, M. *J. Chem. Phys.* **2014**, *141*, 245103.
(48) Amaro, R. E.; Sethi, A.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A. *Biochemistry* **2007**, *46*, 2156−2173.
(49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Supporting Information

Mapping mechanical force propagation through biomolecular complexes

Constantin Schoeler[a], Rafael C. Bernardi[b], Klara H. Malinowska[a], Ellis Durner[a], Wolfgang Ott[a,c], Edward A. Bayer[d], Klaus Schulten[b,e], Michael A. Nash[*a], and Hermann E. Gaub[a]

[a]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[b]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
[c]Center for Integrated Protein Science Munich (CIPSM), University of Munich, 81377 Munich, Germany
[d]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel
[e]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

[*] michael.nash@lmu.de

1 Materials and Methods

1.1 Site Directed Mutagenesis

We performed site-directed mutagenesis of *Ruminococcus 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*[1] were subjected to QuikChange mutagenesis to install the mutations described in the prior paper[2]. All mutagenesis products were confirmed by DNA sequencing analysis.

1.2 Expression and Purification of Cysteine-Mutated Xyn-XMod-Doc

The Xyn(T129C)-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 TBS buffer containing 20 mM 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 dialyzed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at $\sim$ 20°C. The concentrations of the protein stock solutions were determined to be $\sim$ 5 mg/mL by absorption spectrophotometry.

1.3 Expression and Purification of Coh-CBM and mutated Coh-CBM C63S

The Coh-CBM (C63S) fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicole containing ZYM-5052 media[3] 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 $\sim$ 5 mg/mL by absorption spectrophotometry.

1.4 Sample Preparation

Cantilevers and cover glasses were functionalized according to previously published protocols[4]. Briefly, 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 Xyl-XMod-Doc proteins were reduced for 2 hours at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with TBS measurement buffer, centrifuged at 850 rcf for 3 minutes, 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 TBS to remove

unbound proteins, and stored under TBS prior to force spectroscopy measurements. Site specific immobilization of the Coh-CBM-ybbR fusion proteins to PEGylated cantilevers or coverglasses was carried out according to previously published protocols[5]. Briefly, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer for 1h at room temperature. Levers or surfaces were then rinsed with 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 hours at room 37°. Finally, surfaces or levers were subjected to a final rinse with TBS and stored under TBS prior to measurement.

1.5 Single Molecule Force Spectroscopy Measurements

SMFS measurements were performed on a custom built AFM 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. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from 0.2–6.4 μm/s. 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 hours. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw datasets 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 with CBM and Xyn unfolding length increments. We sorted 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: the xylanase (∼89 nm) and the CBM (∼56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

1.6 Data Analysis

Data were analyzed using slight modifications to previously published protocols[4;6;7]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11\,\mathrm{nm}$ connected by a fixed angle $\gamma = 41°$ and and assembled into barrier position histograms using cross-correlation. 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 prior to 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 probability density fits to histograms of rupture forces and loading rates at each pulling speed.

1.7 Molecular Dynamics Simulations

Connecting dynamics to structural data from diverse experimental sources, molecular dynamics simulations allow one to explore off-equilibrium properties of protein structure complexes in unparalleled detail[8]. More specifically, molecular dynamics simulations have always been viewed as a general sampling method for the study of conformational changes[9]. The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97Å resolution and is available at the protein data bank (PDB:4IU3). The system was then solvated 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. Total system size was approximately 580k atoms. The MD simulations in the present study were performed employing the molecular dynamics package NAMD[10;11]. The CHARMM36 force field[12;13] along with the TIP3 water model[14] was used to describe all systems. The simulations were carried out 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)[15] method. The equations of motion were integrated using the r-RESPA multiple time step scheme[11] 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. The first two nanoseconds of the simulations served to equilibrate systems before the production runs, which varied from 200 ns to 1.3 μs in the different simulations. To characterize the coupling between dockerin and cohesin, we performed SMD simulations[16] of constant velocity stretching (SMD-CV protocol) with pulling speed of 0.25 Å/ns. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically, and moving a second restraint point, at the end of the Coh domain, 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 pulling spring is then monitored during the time of the molecular dynamics simulation. All analyses of MD trajectories were carried out employing VMD[17] and its plugins. Surface contact areas of interacting residues were calculated employing Volarea[18] implemented in VMD. The area is calculated using a probe radius defined as an *in silico* rolling sphere that is scanned around the area of the dockerin exposed to the cohesin and also the cohesin area exposed to the dockerin. The Network View plugin[19] on VMD[17] was employed to perform dynamical network analysis. A network was defined as a set of nodes, all α-carbons, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and 2 nonconsecutive monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (nonhydrogen) from the 2 monomers are within 4.5 Å of each other for at least 75% of the frames analyzed. As suggested by Sethi et al.[20], nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths. The dynamical networks were constructed from 20 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as $w_{ij} = -log\,(\mid C_{ij} \mid)$, where C_{ij} is the correlation matrix calculated with Carma[21]. Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was $-log\,(0.5) = 0.69$. To calculate the relevance of off-diagonal terms in the correlation matrix we employed Carma to calculate a correlation matrix where x, y, z components of each atom were considered independently.

2 Protein Sequences

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

2.1 HIS-Xyn(T128C)-XDoc

X-module
Dockerin type III

Xylanase
Linker or extra residues

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

2.2 Coh-CBM(C2A,C63S)-HIS-ybbR

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

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

2.3 CBM(T2C)-Coh-HIS

CBM (T2C)
CohIII
Linker or extra residues

M C 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 C D H A A I I G S N G S Y N G I T S

5

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 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 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 L
E H H H H H H

3 Supplementary Discussion

The Pearson correlation matrices of the Xmod-Doc:Coh complex before and after applying force in the native pulling configuration are presented in Supplementary Figure S3 and S4, respectively. For the unloaded complex, movements within Doc domain are seen to be highly correlated, while XMod is seen to be divided into two anti-correlated sub-domains, one comprising the β-sheet fragment close to the N-terminus (residues 5-15 and 45-66) and the other constituting the rest of the domain. Intra-domain correlations of Coh exhibit more a complex pattern to which both secondary (anti-parallel β-strands and β-sheet at the binding interface) and tertiary structure (vicinity of C- and N-termini) contribute. Some of the inter-domain correlations in the complex originate from spatial vicinity and direct interactions, specifically at the Doc:Coh binding interface and at XMod contacts with Doc inserts. However, coupling between distant parts of the complex is also present. For example, fluctuations of the non-binding part of Coh are correlated with the N-terminal part of XMod and strongly anti-correlated with Doc domain.

4 Supplementary Notes

4.1 Constant Barrier Distance Model

The constant barrier distance model[16], also referred to as the Bell-Evans model[22], is commonly used to estimate the distance to the transition state Δx and the natural off-rate k_0 of mechanically induced receptor ligand dissociation from single-molecule force spectroscopy experiments. It predicts that the most probable rupture force $\langle F \rangle$ is linearly dependent on the logarithm of the force loading rate[16]:

$$\langle F(r) \rangle = \frac{k_B T}{\Delta x} \ln \frac{\Delta x \cdot r}{k_0 k_B T} \tag{S1}$$

where k_B is Boltzmann's constant, T is the temperature and r is the loading rate at the point of rupture.

The probability density distribution of rupture forces at given loading rate r in this model is given as[16]:

$$p(F) = \frac{k_0}{r} \exp\left[\frac{\Delta x}{k_B T} F - \frac{k_0 \cdot k_B T}{\Delta x \cdot r}\left(e^{\frac{\Delta x}{k_B T} F} - 1\right)\right] \tag{S2}$$

6

4.2 Dudko-Hummer-Szabo Model

The Dudko-Hummer-Szabo (DHS)[23;24] model describes a non-linear dependence for the most probable rupture force on loading rate:

$$\langle F(r)\rangle = \frac{\Delta G}{\nu \Delta x}\left\{1-\left[\frac{k_B T}{\Delta G}\ln\left(\frac{k_B T k_0}{\Delta x r}e^{\frac{\Delta G}{k_B T}+\gamma}\right)\right]^{\nu}\right\} \tag{S3}$$

where ΔG is the free energy of activation and $\gamma = 0.577$ is the Euler-Mascheroni constant. The model parameter ν defines the single-well free-energy surface model used ($\nu = \frac{2}{3}$ for linear-cubic and $\frac{1}{2}$ for cusp free-energy. For $\nu = 1$ and $\Delta G \to \infty$ independent of ν the Eqs. (S1) and (S2) are recovered.

4.3 Pearson Correlation and covariance matrix

4.3.1 Validation

An $N \times N$ matrix of Pearson correlation coefficients C_{ij} (Supporting Eq. S4) was calculated from each atom's x, y, z position throughout the simulation trajectory, which inherently ignores off-diagonal elements of the atomic 3×3 submatrices D_{ij}^{mn} from the full normalized $3N \times 3N$ covariance matrix (*i.e.*, correlations along orthogonal axes are neglected, see Supporting Eqs. (S5) and (S6)) and Supporting Fig S8.

Although this quasi-harmonic approximation is commonly employed in correlation analysis[19;25–29], it is not *a priori* justified for complicated biomolecular interactions[30]. To validate the use of Pearson correlations, we therefore first analyzed independently the contributions from diagonal and off-diagonal elements of each 3 x 3 covariance submatrix for each pair of α-carbons within the structure (Fig. S9A and B). Both with and without applied force, the off-diagonal elements roughly follow Gaussian distributions centered around a correlation value of 0. Interestingly, as force was applied, the standard deviation of the distribution of off-diagonal correlation values decreased from $\sigma_{unloaded} = 0.45$ to $\sigma_{loaded} = 0.29$. This indicated a lesser influence of off-diagonal elements on the highly (anti-)correlated motion within the system under force (see Supporting Discussion 3). The diagonal elements of the sub-matrices that are used for calculating the Pearson correlation values showed a dramatically different behavior. Both in the unloaded and loaded state, the resulting distributions were strongly shifted towards highly correlated motion, and the shape of the distribution remained mostly unchanged after application of force. Since our analysis relies on the identification of paths of highest correlation through proximate residues, the quasi-harmonic approximation implied by the use of Pearson correlation is justified, especially for suboptimal pathway analysis. The resulting distributions of on- and off-diagonal matrix elements of each covariance submatrix for the loaded configuration HF class (Fig. S10A) and LF class (Fig. S10B) exhibited the same characteristics as previously described for the native configuration, with off-diagonal elements showing symmetric correlations around zero and diagonal elements showing highly correlated motions.

4.3.2 Supplementary Equations

The Pearson correlation coefficient C_{ij} used in our dynamical network analysis protocol is given by:

$$C_{ij} = \frac{\langle \Delta \mathbf{r}_i(t) \cdot \Delta \mathbf{r}_j(t)\rangle}{\left(\left\langle \Delta \mathbf{r}_i(t)^2\right\rangle \left\langle \Delta \mathbf{r}_j(t)^2\right\rangle\right)^{\frac{1}{2}}} \tag{S4}$$

where $\Delta\mathbf{r}_i(t) = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$.

The full $3N \times 3N$ covariance matrix M_{ij} for atoms i and j consists of 3×3 submatrices of the form:

$$\left\langle \Delta\mathbf{r}_i(t)\, \Delta\mathbf{r}_j(t)^T \right\rangle = M_{ij} = \begin{pmatrix} M_{ij}^{xx} & M_{ij}^{xy} & M_{ij}^{xz} \\ M_{ij}^{yx} & M_{ij}^{yy} & M_{ij}^{yz} \\ M_{ij}^{zx} & M_{ij}^{zy} & M_{ij}^{zz} \end{pmatrix} \tag{S5}$$

The full normalized correlation matrix is calculated from M_{ij}:

$$D_{ij}^{mn} = \frac{M_{ij}^{mn}}{\sqrt{M_{ij}^{mm} M_{ij}^{nn}}} \tag{S6}$$

Consequently, the Pearson correlation coefficient is calculated as the trace of the normalized 3×3 submatrices ($C_{ij} = \operatorname{Tr} D_{ij}$).

4.3.3 Derivation of Main Text Equation 2

Eq. 1 from the main text reads:

$$\left\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j^T \right\rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{S7}$$

Combining Eqs. (S7) and (S4) yields:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\left\langle \Delta\mathbf{r}_i^2(t) \right\rangle \left\langle \Delta\mathbf{r}_j^2(t) \right\rangle \right)^{-\frac{1}{2}} \tag{S8}$$

For an arbitrary potential $U_i(\mathbf{r})$ of atom i, a Taylor expansion around the potential minimum (set to be at 0) yields:

$$U_i(\mathbf{r}) = 0 + \underbrace{\mathbf{r}_i^T \nabla U(0)}_{=0} + \frac{1}{2} \mathbf{r}_i^T \underline{H}(0)\, \mathbf{r}_i + ... \tag{S9}$$

where $\underline{H}(0)$ is the Hessian matrix evaluated at the potential minimum. Assuming Schwarz' theorem holds for $U_i(\mathbf{r})$, $\underline{H}(0)$ is a symmetric matrix and therefore has real eigenvalues and orthonormal eigenvectors. Hence, a change to the eigenbasis of $\underline{H}(0)$ is a rotation of the coordinate system. In this new basis the Hessian is diagonal:

$$\underline{H}(0) \to \underline{H}'(0) = \begin{pmatrix} k_{x'} & 0 & 0 \\ 0 & k_{y'} & 0 \\ 0 & 0 & k_{z'} \end{pmatrix} \tag{S10}$$

This yields a simple expression for the second order term in Eq. (S9):

$$U_i(\mathbf{r}') = \frac{1}{2} \mathbf{r}'^T \underline{H}'(0)\, \mathbf{r}' = \frac{1}{2} \left(k_{x'} x'^2 + k_{y'} y'^2 + k_{z'} z'^2 \right) \tag{S11}$$

Now we inspect the normalization of C_{ij}:

$$\langle \Delta\mathbf{r}_i^2(t) \rangle = \langle \mathbf{r}_i^2(t) - 2\mathbf{r}_i(t) \langle \mathbf{r}_i(t) \rangle + \langle \mathbf{r}_i(t) \rangle^2 \rangle \tag{S12}$$

8

In the harmonic approximation of the potential of atom i, $\langle \mathbf{r}_i(t) \rangle = 0$, and therefore $\langle \Delta \mathbf{r}_i(t)^2 \rangle = \langle \mathbf{r}_i^2(t) \rangle$. In the basis of $\underline{H}'(0)$ this becomes:

$$\langle \mathbf{r}_i'^2(t) \rangle = \langle x_i'(t)^2 + y_i'(t)^2 + z_i'(t)^2 \rangle = \langle x_i'(t)^2 \rangle + \langle y_i'(t)^2 \rangle + \langle z_i'(t)^2 \rangle \tag{S13}$$

Applying the equipartition theorem to this result yields:

$$\langle x_i'(t)^2 \rangle = \frac{k_B T}{k'_{x_i}} \tag{S14}$$

And therefore:

$$\langle \Delta \mathbf{r}_i'(t)^2 \rangle = k_B T \left(\frac{1}{k'_{x_i}} + \frac{1}{k'_{y_i}} + \frac{1}{k'_{z_i}} \right) = \frac{k_B T}{k'_{i,eff}} \tag{S15}$$

Plugging this result into Eq. (S8), one finds:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\frac{k_B T}{k'_{i,eff}} \right)^{-\frac{1}{2}} \left(\langle \Delta \mathbf{r}_j(t)^2 \rangle \right)^{-\frac{1}{2}} \tag{S16}$$

Repeating the above steps for atom j yields the final result:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\frac{k_B T}{k'_{i,eff}} \right)^{-\frac{1}{2}} \left(\frac{k_B T}{k'_{j,eff}} \right)^{-\frac{1}{2}} \tag{S17}$$

$$= \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \sqrt{k'_{i,eff} \cdot k'_{j,eff}} \tag{S18}$$

5 Supplementary Figures

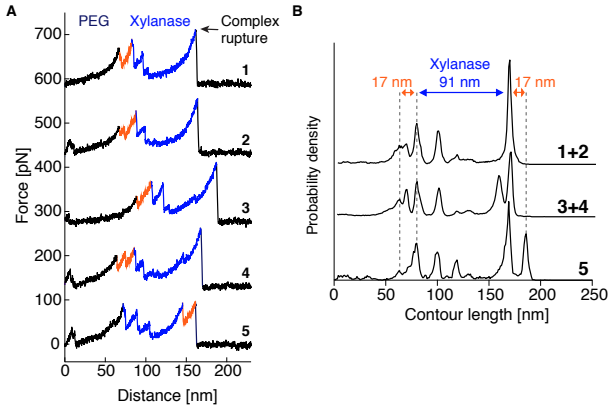


Fig. S1: SMFS of the non-native low force curve class. **A** Typical unfolding fingerprints. All traces showed a characteristic Xyn fingerprint (blue). A $17-19$ nm increment corresponding to partial N-terminal Coh unfolding (orange) occurs either prior to Xyn unfolding (traces 1-4), or just before complex rupture (trace 5). It was observed as a single event (traces 1,3 and 5) or showed substructure (traces 2 and 4). **B** Traces were grouped and assembled into contour length histograms. One or more of the unassigned increments combined into a $17-19$ nm increment.

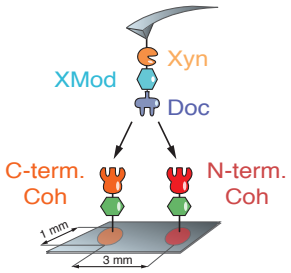


Fig. S2: Comparing the native geometry with the non-native high force class. Two exclude uncertainties in cantilever calibration when comparing the native geometry with the non-native HF class, we immobilized both Coh-CBM (native) and CBM-Coh (non-native) on two spatially separated spots on a single cover glass. These spots where then alternately probed with the same Xyn-XMod-Doc functionalized cantilever.

10

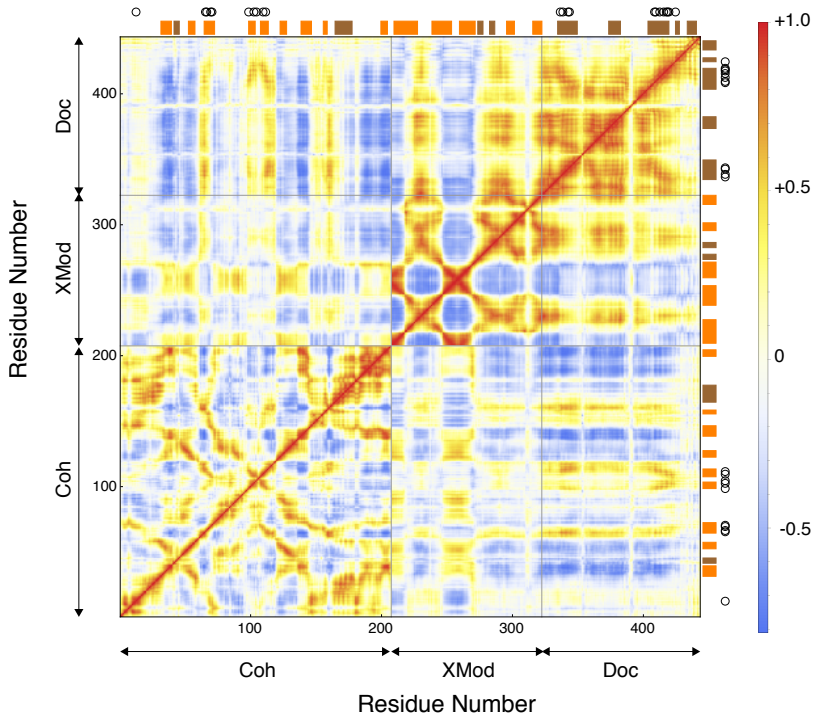


Fig. S3: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the unloaded Xmod-Doc:Coh complex. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from the Coh and Doc binding interface.

11

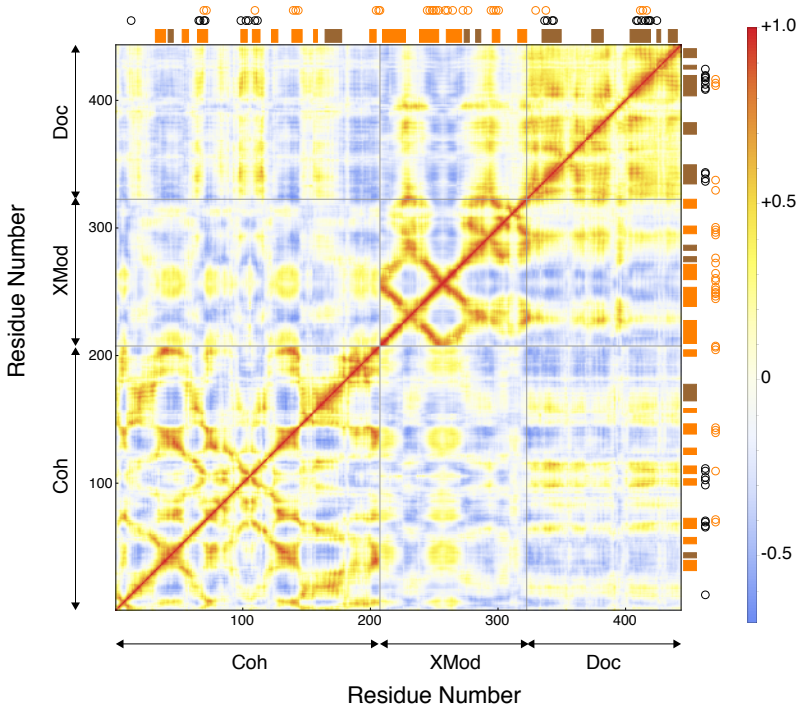


Fig. S4: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the Xmod-Doc:Coh complex loaded with force in the native pulling geometry. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from Coh and Doc binding interfaces and orange circles represent residues on the force propagation path.

12

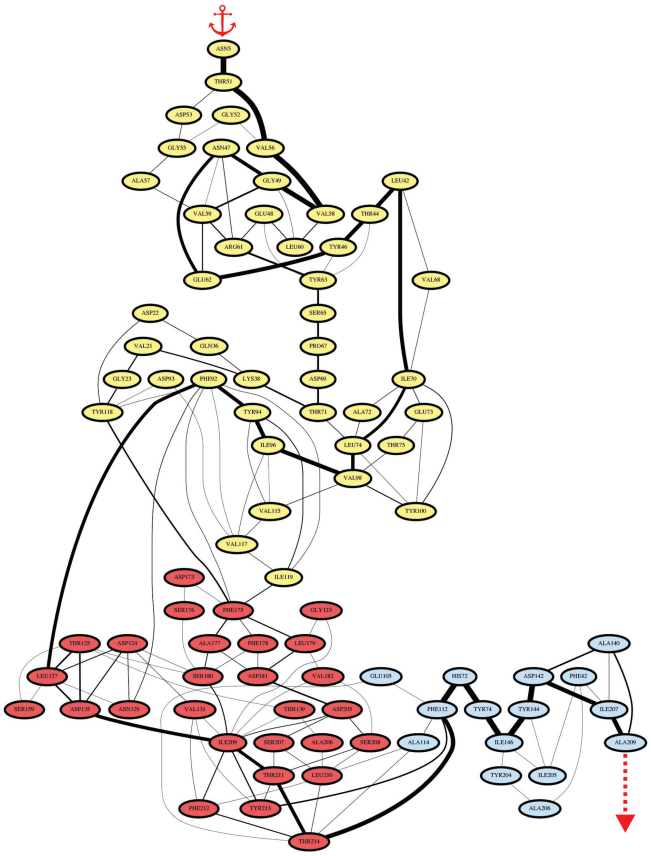


Fig. S5: Force propagation pathway through the loaded XMod-Doc:Coh complex in the native pulling geometry (N-terminal pulling of Xmod-Doc, C-terminal pulling of Coh) obtained from dynamical network analysis. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

13

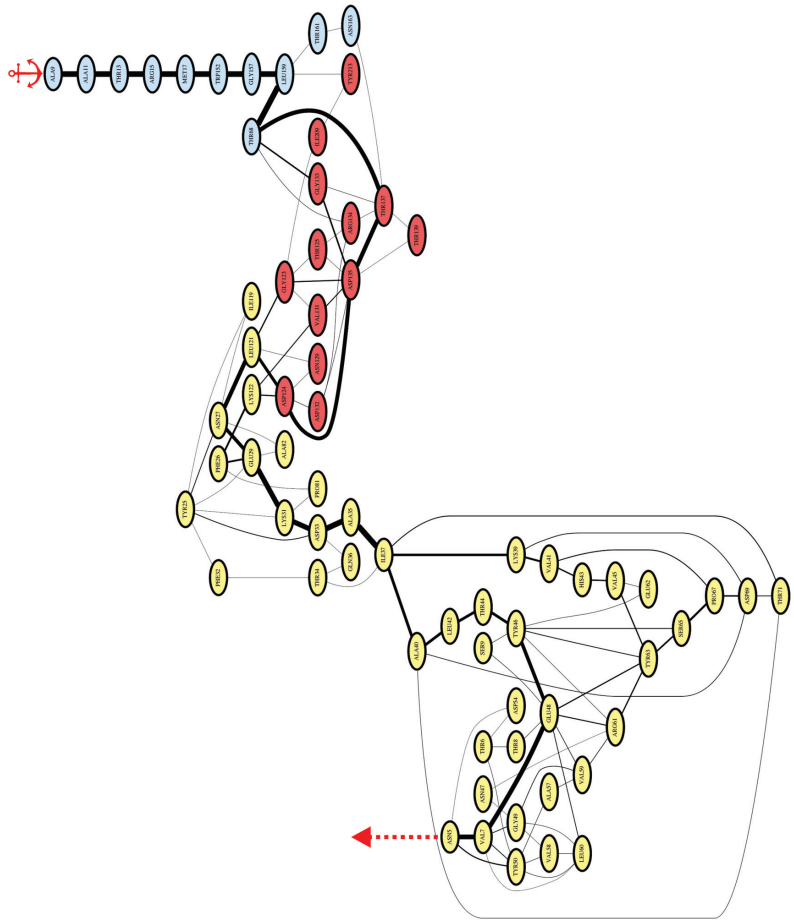


Fig. S6: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing high-force unbinding characteristics and no C-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

14

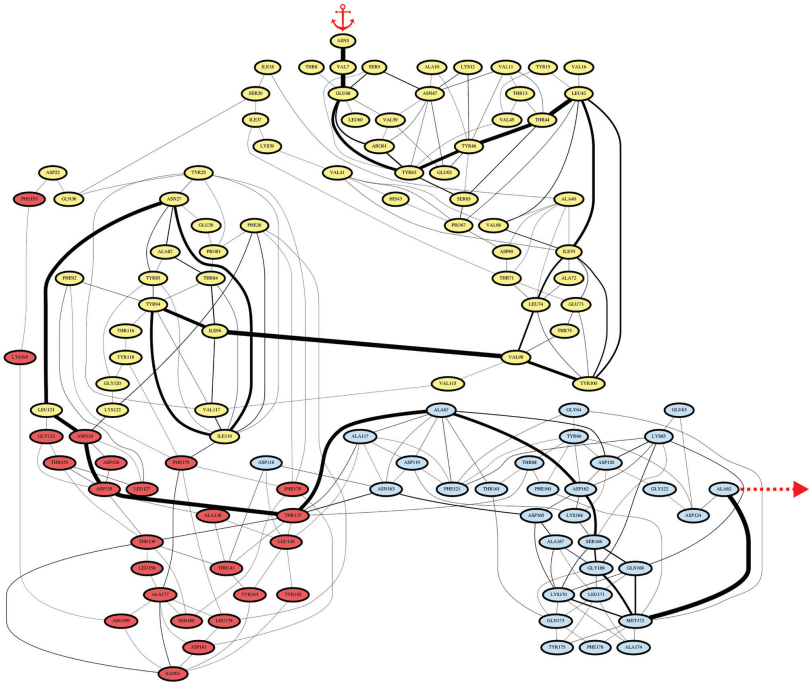


Fig. S7: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing low-force unbinding characteristics and partial N-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

15

$$M_{ij} = \begin{vmatrix}
M_{1x_1x_1} & M_{1x_1y_1} & M_{1x_1z_1} & M_{1x_2x_2} & M_{1x_2y_2} & M_{1x_2z_2} & M_{1x_3x_3} & M_{1x_3y_3} & M_{1x_3z_3} & M_{1x_4x_4} & M_{1x_4y_4} & M_{1x_4z_4} & M_{1x_5x_5} & M_{1x_5y_5} & M_{1x_5z_5} \\
M_{1y_1x_1} & M_{1y_1y_1} & M_{1y_1z_1} & M_{1y_2x_2} & M_{1y_2y_2} & M_{1y_2z_2} & M_{1y_3x_3} & M_{1y_3y_3} & M_{1y_3z_3} & M_{1y_4x_4} & M_{1y_4y_4} & M_{1y_4z_4} & M_{1y_5x_5} & M_{1y_5y_5} & M_{1y_5z_5} \\
M_{1z_1x_1} & M_{1z_1y_1} & M_{1z_1z_1} & M_{1z_2x_2} & M_{1z_2y_2} & M_{1z_2z_2} & M_{1z_3x_3} & M_{1z_3y_3} & M_{1z_3z_3} & M_{1z_4x_4} & M_{1z_4y_4} & M_{1z_4z_4} & M_{1z_5x_5} & M_{1z_5y_5} & M_{1z_5z_5} \\
M_{2x_1x_1} & M_{2x_1y_1} & M_{2x_1z_1} & M_{2x_2x_2} & M_{2x_2y_2} & M_{2x_2z_2} & M_{2x_3x_3} & M_{2x_3y_3} & M_{2x_3z_3} & M_{2x_4x_4} & M_{2x_4y_4} & M_{2x_4z_4} & M_{2x_5x_5} & M_{2x_5y_5} & M_{2x_5z_5} \\
M_{2y_1x_1} & M_{2y_1y_1} & M_{2y_1z_1} & M_{2y_2x_2} & M_{2y_2y_2} & M_{2y_2z_2} & M_{2y_3x_3} & M_{2y_3y_3} & M_{2y_3z_3} & M_{2y_4x_4} & M_{2y_4y_4} & M_{2y_4z_4} & M_{2y_5x_5} & M_{2y_5y_5} & M_{2y_5z_5} \\
M_{2z_1x_1} & M_{2z_1y_1} & M_{2z_1z_1} & M_{2z_2x_2} & M_{2z_2y_2} & M_{2z_2z_2} & M_{2z_3x_3} & M_{2z_3y_3} & M_{2z_3z_3} & M_{2z_4x_4} & M_{2z_4y_4} & M_{2z_4z_4} & M_{2z_5x_5} & M_{2z_5y_5} & M_{2z_5z_5} \\
M_{3x_1x_1} & M_{3x_1y_1} & M_{3x_1z_1} & M_{3x_2x_2} & M_{3x_2y_2} & M_{3x_2z_2} & M_{3x_3x_3} & M_{3x_3y_3} & M_{3x_3z_3} & M_{3x_4x_4} & M_{3x_4y_4} & M_{3x_4z_4} & M_{3x_5x_5} & M_{3x_5y_5} & M_{3x_5z_5} \\
M_{3y_1x_1} & M_{3y_1y_1} & M_{3y_1z_1} & M_{3y_2x_2} & M_{3y_2y_2} & M_{3y_2z_2} & M_{3y_3x_3} & M_{3y_3y_3} & M_{3y_3z_3} & M_{3y_4x_4} & M_{3y_4y_4} & M_{3y_4z_4} & M_{3y_5x_5} & M_{3y_5y_5} & M_{3y_5z_5} \\
M_{3z_1x_1} & M_{3z_1y_1} & M_{3z_1z_1} & M_{3z_2x_2} & M_{3z_2y_2} & M_{3z_2z_2} & M_{3z_3x_3} & M_{3z_3y_3} & M_{3z_3z_3} & M_{3z_4x_4} & M_{3z_4y_4} & M_{3z_4z_4} & M_{3z_5x_5} & M_{3z_5y_5} & M_{3z_5z_5} \\
M_{4x_1x_1} & M_{4x_1y_1} & M_{4x_1z_1} & M_{4x_2x_2} & M_{4x_2y_2} & M_{4x_2z_2} & M_{4x_3x_3} & M_{4x_3y_3} & M_{4x_3z_3} & M_{4x_4x_4} & M_{4x_4y_4} & M_{4x_4z_4} & M_{4x_5x_5} & M_{4x_5y_5} & M_{4x_5z_5} \\
M_{4y_1x_1} & M_{4y_1y_1} & M_{4y_1z_1} & M_{4y_2x_2} & M_{4y_2y_2} & M_{4y_2z_2} & M_{4y_3x_3} & M_{4y_3y_3} & M_{4y_3z_3} & M_{4y_4x_4} & M_{4y_4y_4} & M_{4y_4z_4} & M_{4y_5x_5} & M_{4y_5y_5} & M_{4y_5z_5} \\
M_{4z_1x_1} & M_{4z_1y_1} & M_{4z_1z_1} & M_{4z_2x_2} & M_{4z_2y_2} & M_{4z_2z_2} & M_{4z_3x_3} & M_{4z_3y_3} & M_{4z_3z_3} & M_{4z_4x_4} & M_{4z_4y_4} & M_{4z_4z_4} & M_{4z_5x_5} & M_{4z_5y_5} & M_{4z_5z_5} \\
M_{5x_1x_1} & M_{5x_1y_1} & M_{5x_1z_1} & M_{5x_2x_2} & M_{5x_2y_2} & M_{5x_2z_2} & M_{5x_3x_3} & M_{5x_3y_3} & M_{5x_3z_3} & M_{5x_4x_4} & M_{5x_4y_4} & M_{5x_4z_4} & M_{5x_5x_5} & M_{5x_5y_5} & M_{5x_5z_5} \\
M_{5y_1x_1} & M_{5y_1y_1} & M_{5y_1z_1} & M_{5y_2x_2} & M_{5y_2y_2} & M_{5y_2z_2} & M_{5y_3x_3} & M_{5y_3y_3} & M_{5y_3z_3} & M_{5y_4x_4} & M_{5y_4y_4} & M_{5y_4z_4} & M_{5y_5x_5} & M_{5y_5y_5} & M_{5y_5z_5} \\
M_{5z_1x_1} & M_{5z_1y_1} & M_{5z_1z_1} & M_{5z_2x_2} & M_{5z_2y_2} & M_{5z_2z_2} & M_{5z_3x_3} & M_{5z_3y_3} & M_{5z_3z_3} & M_{5z_4x_4} & M_{5z_4y_4} & M_{5z_4z_4} & M_{5z_5x_5} & M_{5z_5y_5} & M_{5z_5z_5}
\end{vmatrix}$$

Fig. S8: Full unnormalized covariance Matrix M_{ij} for a five atom system from which the full normalized covariance matrix is calculated according to Eq. (S6). On- and off-diagonal elements from one of the atomic submatrices are highlighted in yellow and blue, respectively.

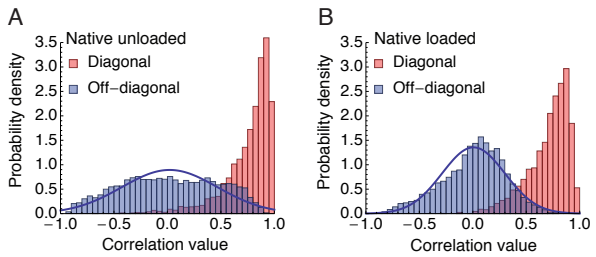


Fig. S9: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A**, the native unloaded, and **B** the native loaded, scenario.

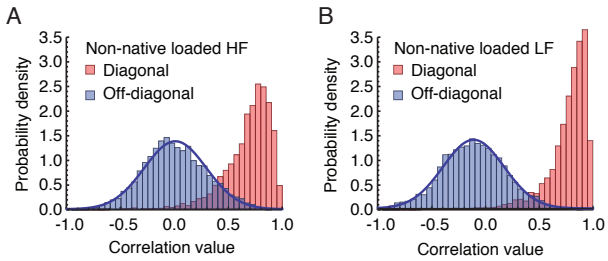
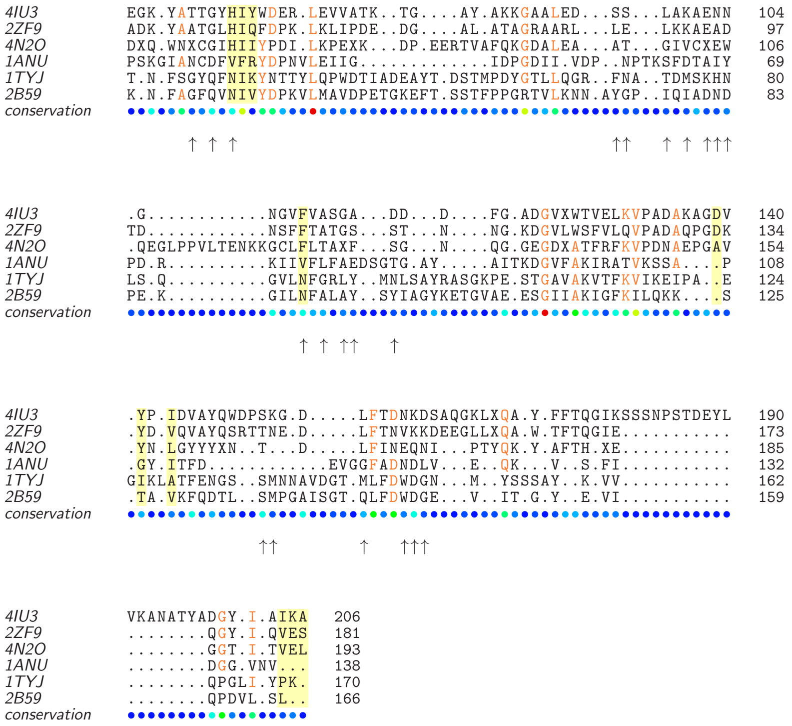


Fig. S10: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A,** the non-native HF, and **B** the non-native LF, scenario.

16



```
4IU3          EGK.YATTGYHIYWDER.LEVVATK..TG....AY.AKKGAALED...SS...LAKAENN   104
2ZF9          ADK.YAATGLHIQFDPK.LKLIPDE..DG....AL.ATAGRAARL...LE...LKKAEAD    97
4N2O          DXQ.WNXCGIHIIYPDI.LKPEXK...DP.EERTVAFQKGDALEA...AT...GIVCXEW   106
1ANU          PSKGIANCDFVFRYDPNVLEIIG.............IDPGDII.VDP..NPTKSFDTAIY    69
1TYJ          T.N.FSGYQFNIKYNTTYLQPWDTIADEAYT.DSTMPDYGTLLQGR..FNA..TDMSKHN    80
2B59          K.N.FAGFQVNIVYDPKVLMAVDPETGKEFT.SSTFPPGRTVLKNN.AYGP..IQIADND    83
conservation  ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●

                    ↑ ↑ ↑                                     ↑↑   ↑ ↑ ↑↑↑

4IU3          .G............NGVFVASGA...DD...D....FG.ADGVXWTVELKVPADAKAGDV   140
2ZF9          TD............NSFFTATGS...ST...N....NG.KDGVLWSFVLQVPADAQPGDK   134
4N2O          .QEGLPPVLTENKKGCLFLTAXF...SG...N....QG.GEGDXATFRFKVPDNAEPGAV   154
1ANU          PD.R..........KIIVFLFAEDSGTG.AY.....AITKDGVFAKIRATVKSSA....P   108
1TYJ          LS.Q..........GVLNFGRLY..MNLSAYRASGKPE.STGAVAKVTFKVIKEIPA..E   124
2B59          PE.K..........GILNFALAY..SYIAGYKETGVAE.ESGIIAKIGFKILQKK....S   125
conservation  ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●

                               ↑ ↑ ↑↑   ↑

4IU3          .YP.IDVAYQWDPSKG.D.....LFTDNKDSAQGKLXQA.Y.FFTQGIKSSSNPSTDEYL   190
2ZF9          .YD.VQVAYQSRTTNE.D.....LFTNVKKDEEGLLXQA.W.TFTQGIE...........   173
4N2O          .YN.LGYYYXN..T...D.....LFINEQNI...PTYQK.Y.AFTH.XE...........   185
1ANU          .GY.ITFD............EVGGFADNDLV...E..QK...V..S.FI...........   132
1TYJ          GIKLATFENGS..SMNNAVDGT.MLFDWDGN...M..YSSSAY..K.VV...........   162
2B59          .TA.VKFQDTL..SMPGAISGT.QLFDWDGE...V..IT.G.Y..E.VI...........   159
conservation  ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●

                           ↑↑        ↑   ↑↑↑

4IU3          VKANATYADGY.I.AIKA   206
2ZF9          ........QGY.I.QVES   181
4N2O          ........GGT.I.TVEL   193
1ANU          ........DGG.VNV...   138
1TYJ          ........QPGLI.YPK.   170
2B59          ........QPDVL.SL..   166
conservation  ●●●●●●●●●●●●●●●●●●
```

Fig. S11: Structure-aligned sequences of six crystallized cohesins. Residues on the force propagation path are highlighted in yellow. Arrows indicate binding residues. Residue conservation is color-coded from blue - lack of conservation, to red - residue fully conserved. Crystal structures used: 4IU3 ScaE *Rf* FD-1, 2ZF9 ScaE *Rf* strain 17, 4N2O CohG *Rf* FD-1, 1ANU CohC2 CipC *Ct*, 1TYJ CohA11 ScaA *Bc*, 2B59 SdbA *Ct*.

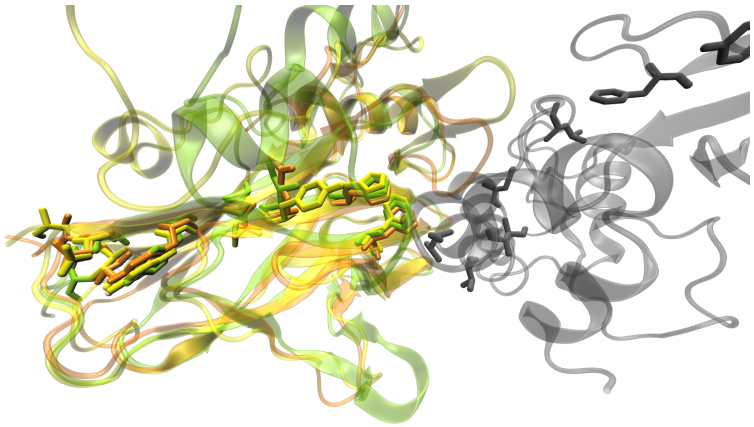


Fig. S12: Structure and sequence conservation of the force propagation pathway residues in Coh. CohE from the ScaE cell anchoring protein, *Rf* FD-1 used in this work (PDB 4IU3) is highlighted in green. Highly homologous structures of CohE from *Rf* strain 17 (PDB 2ZF9) and Coh G from *Rf* FD-1 (PDB 4A2O) are colored in orange and yellow, respectively. Residues lying in the force propagation path are shown as sticks. XDoc from the CttA *Rf* FD-1 scaffold used in this work is shown in gray.

References

[1] Orly Salama-Alber, Maroor K Jobby, Seth Chitayat, Steven P Smith, Bryan A White, Linda J W Shimon, Raphael Lamed, Felix Frolow, and Edward A Bayer. Atypical cohesin-dockerin complex responsible for cell-surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.*, 288(23):16827–16838, April 2013.

[2] 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. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.*, 5:1–8, December 2014.

[3] F William Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expres. Purif.*, 41(1):207–234, May 2005.

[4] Markus A Jobst, Constantin Schoeler, and Michael A Nash. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *J. Vis. Exp.*, 82(82):e50950, 2013.

[5] Jun Yin, Alison J Lin, David E Golan, and Christopher T Walsh. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.*, 1(1):280–285, June 2006.

[6] Stefan W Stahl, Michael A Nash, Daniel B Fried, Michal Slutzki, Yoav Barak, Edward A Bayer, and Hermann E Gaub. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.*, 109(50):20431–20436, December 2012.

[7] Elias M Puchner, Gereon Franzen, Mathias Gautel, and Hermann E Gaub. Comparing proteins by their unfolding pattern. *Biophys. J.*, 95(1):426–434, July 2008.

[8] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.*, 31:64–74, 2015.

[9] R C Bernardi, M C R Melo, and K Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta*, 1850(5):872–877, 2015.

[10] Laxmikant Kalé, Robert Skeel, Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.*, 151(1): 283–312, 1999.

[11] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005.

[12] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro E M Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell, Jr. 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(9):3257–3273, 2012.

[13] A D MacKerell, D Bashford, Bellott, R L Dunbrack, J D Evanseck, M J Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T K Lau, C Mattos, S Michnick, T Ngo, D T Nguyen, B Prodhom, W E Reiher, B Roux, M Schlenkrich, J C Smith, R Stote, J Straub, M Watanabe, J Wiórkiewicz-Kuczera, D Yin, and M Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.

[14] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926, 1983.

[15] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[16] S Izrailev, S Stepaniants, M Balsera, Y Oono, and K Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72(4):1568–1581, 1997.

[17] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996.

[18] João V Ribeiro, Juan A C Tamames, Nuno M F S A Cerqueira, Pedro A Fernandes, and Maria J Ramos. Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems. *Chem. Biol. Drug. Des.*, 82(6):743–755, December 2013.

[19] J Eargle and Z A Luthey-Schulten. NetworkView: 3D display and analysis of protein·RNA interaction networks. *Bioinformatics*, 28(22):3000–3001, 2012.

[20] Anurag Sethi, John Eargle, Alexis A Black, and Zaida Luthey-Schulten. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, 106(16):6620–6625, April 2009.

[21] N Glykos. Software news and updates carma: A molecular dynamics analysis program. *J. Comput. Chem.*, 27(14):1765–1768, 2006.

[22] Gerhard Hummer and Attila Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical Journal*, 85(1):5–15, July 2003.

[23] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):108101, March 2006.

[24] O K Dudko, G Hummer, and A Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *PNAS*, 105(41):15755–15760, October 2008.

[25] M Karplus and J N Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14:325–332, 1981.

[26] R M Levy, M Karplus, J Kushick, and D Perahia. Evaluation of the Configurational entropy for Proteins: Application to Molecular Dynamics Simulations of an α-Helix. *Macromolecules*, 17(7):1370–1374, 1984.

[27] Rommie E Amaro, Anurag Sethi, Rebecca S Myers, V Jo Davisson, and Zaida A Luthey-Schulten. A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry*, 46(8):2156–2173, February 2007.

[28] Rebecca W Alexander, John Eargle, and Zaida Luthey-Schulten. Experimental and computational determination of tRNA dynamics. *FEBS Lett.*, 584(2):376–386, January 2010.

[29] Rafael C Bernardi, Isaac Cann, and Klaus Schulten. Molecular dynamics study of enhanced Man5B enzymatic activity. *Biotechnol. Biofuels*, 7(83):1–8, 2014.

[30] Oliver F Lange and Helmut J Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins*, 62(4):1053–1061, March 2006.

3.3 Comparing two Cellulosomal Anchoring Interactions

This publication investigates the receptor-ligand pair ScaB-XDoc:CohE *(R.f)*, one of the strongest receptor:ligand interactions found to date. It is a homologue of the interaction between Ctta-XDoc and CohE, which are described in publications 3.1 and 3.2. Despite the very similar binding interfaces of CttA-XDoc and ScaB-XDoc and comparable binding constants, we find the latter to be substantially stronger. We attribute this increase in mechanical strength to the larger X-module domain of ScaB, which unfolds at higher forces and more seldomly than its CttA counterpart. Using all-atom MD simulations, we can corroborate the previously assumed role of the Xmodule as a mechanical stabilizer, and propose that it is essential for distributing force to the Dockerin's binding interface. I contributed to the design of the experimental part of the study, and performed SMFS experiments and analysis. Interpretation of the results as well as manuscript preparation was shared between Michael A. Nash, Rafael C. Bernardi, and myself.

Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy

Rafael C. Bernardi†, Ellis Durner†, Constantin Schoeler,
Klara H. Malinowska, Bruna G. Carvalho, Edward A. Bayer, Zaida Luthey-Schulten,
Hermann E. Gaub, and Michael A. Nash

†*these authors contributed equally to this publication*

published in

J. Am. Chem. Soc. 2019, 141, 14752-14763

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
Cite This: J. Am. Chem. Soc. 2019, 141, 14752−14763
pubs.acs.org/JACS

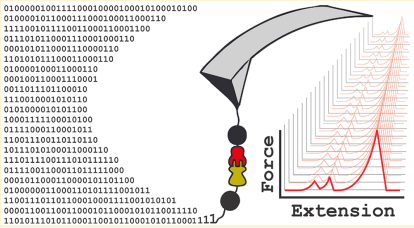
Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy

Rafael C. Bernardi,[*,†,○] Ellis Durner,[‡,○] Constantin Schoeler,[‡] Klara H. Malinowska,[‡] Bruna G. Carvalho,[§] Edward A. Bayer,[∥] Zaida Luthey-Schulten,[†,⊥] Hermann E. Gaub,[‡] and Michael A. Nash[*,∇,#]

[†]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[‡]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[§]School of Chemical Engineering, University of Campinas, 13083-852 Campinas, Brazil
[∥]Department of Biomolecular Sciences, Weizmann Institute of Science, 76100 Rehovot, Israel
[⊥]Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[∇]Department of Chemistry, University of Basel, 4058 Basel, Switzerland
[#]Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

Supporting Information

ABSTRACT: Can molecular dynamics simulations predict the mechanical behavior of protein complexes? Can simulations decipher the role of protein domains of unknown function in large macromolecular complexes? Here, we employ a wide-sampling computational approach to demonstrate that molecular dynamics simulations, when carefully performed and combined with single-molecule atomic force spectroscopy experiments, can predict and explain the behavior of highly mechanostable protein complexes. As a test case, we studied a previously unreported homologue from *Ruminococcus flavefaciens* called X-module-Dockerin (XDoc) bound to its partner Cohesin (Coh). By performing dozens of short simulation replicas near the rupture event, and analyzing dynamic network fluctuations, we were able to generate large simulation statistics and directly compare them with experiments to uncover the mechanisms involved in mechanical stabilization. Our single-molecule force spectroscopy experiments show that the XDoc-Coh homologue complex withstands forces up to 1 nN at loading rates of 10^5 pN/s. Our simulation results reveal that this remarkable mechanical stability is achieved by a protein architecture that directs molecular deformation along paths that run perpendicular to the pulling axis. The X-module was found to play a crucial role in shielding the adjacent protein complex from mechanical rupture. These mechanisms of protein mechanical stabilization have potential applications in biotechnology for the development of systems exhibiting shear enhanced adhesion or tunable mechanics.



INTRODUCTION

Molecular structure and motion can be studied using various techniques, but none offer a level of detail comparable to molecular dynamics (MD) simulations. By depicting atomic-level motions with femtosecond resolution, all-atom MD provides researchers with a unique computational microscope.[1] It is crucial, however, to validate simulations by direct comparison with experiments.[2,3] Once validated, MD can provide structural and mechanistic insights at extremely high resolution. One field in which insights gleaned from MD simulations can have a major impact is in the study of molecular biomechanics. Molecular biomechanics[4−6] is a topic that is challenging to study experimentally because molecular level mechanical properties remain hidden to ensemble averaging methods and bioinformatic methods, and require measurements with high spatial and temporal resolution. MD simulations can therefore aid experimentalists and provide mechanistic insights of molecules under force.

We previously reported on the extreme mechanostability among folded domains[7−9] and receptor−ligand complexes[10−16] derived from cellulose-degrading enzyme networks

Received: July 1, 2019
Published: August 29, 2019

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

Journal of the American Chemical Society Article

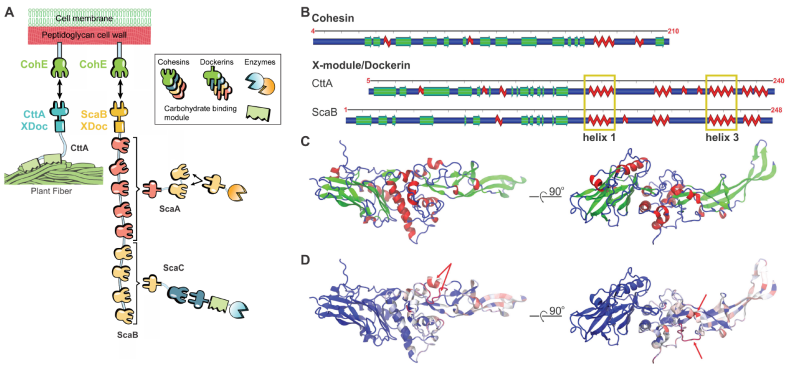


Figure 1. Homology model of ScaB-XDoc:CohE. (A) Schematic of the *Rf* cellulosome protein network for fiber degradation.[20] The two cell-surface anchoring complexes are CttA-XDoc:CohE and ScaB-XDoc:CohE. (B) Representation of secondary structures of Coh (CohE) and XDoc domains from both CttA (PDB ID: 4IU3) and ScaB (model protein). (C) The structural homology model of the ScaB-XDoc:CohE complex colored by secondary structure and (D) colored by sequence similarity (BLOSUM60) to CttA-XDoc:CohE. In (D), regions indicated by red arrows are ScaB-XDoc insert sequences not present in CttA.

known as cellulosomes. The binding partners within these networks belong to grouped families of conserved domains called Cohesin (Coh) and Dockerin (Doc),[17] comprising tens of dozens of known homologues that exhibit high affinity and specificity.[18] X-modules (Xmod) are Ig-like folds frequently found as tandem dyads adjacent to scaffold-borne Docs that, despite having been heavily studied, have no known function.[19] The most mechanostable Coh−Doc interaction reported to date is a complex from *Ruminococcus flavefaciens* (*Rf*) formed between the tandem dyad CttA-XDoc and the cell-wall anchored Cohesin E (CohE).[12,20] In its native pulling geometry, this complex was capable of resisting forces of 500−800 pN at loading rates ranging from 2 to 300 nN/s.[12]

Here, we report a related protein complex that is significantly stronger. Depicted schematically in Figure 1A, *Rf* scaffold B (ScaB) is found within the same cellulosome as the previously reported CttA-XDoc:CohE complex[12,15] and contains a C-terminal XDoc (ScaB-XDoc) that shares high sequence homology with CttA-XDoc. The role of ScaB-XDoc is to connect the large protein machinery of the *Rf* cellulosome to the cell wall via high-affinity binding to CohE.[21−23] CohE is itself covalently linked to the peptidoglycan outer cell wall by a sortase-mediated pathway.[24,25]

To investigate the adhesion forces at play in the ScaB-XDoc:CohE complex, we employed steered molecular dynamics (SMD)[26−28] simulations. Using a wide sampling approach, simulations were performed with many replicas for each of the systems, allowing us to analyze experiments and simulations in the same statistical framework. Wide sampling combined with dynamic network analysis[29] allowed us to visualize most probable deformation pathways through the protein architecture and understand how resistance to mechanical stress arises at the level of individual complexes. Simulation results were validated using single-molecule force spectroscopy (SMFS) experiments carried out with an atomic force microscope (AFM) over a range of loading rates. By recording the force required to break the complex for thousands of individual interactions, we built up sufficient statistics to analyze the interaction and unfolding pathways. We describe remarkable agreement between simulations and experiments, demonstrating that they probe fundamentally the same molecular process. Such an approach provided a strong basis for the molecular-level mechanistic descriptions that emerged from detailed analysis of MD simulations.

■ RESULTS AND DISCUSSION

Modeling. There were no structural data available on the ScaB-XDoc:CohE complex, so we built a homology model[30] based on CttA-XDoc. The primary sequences of CttA-XDoc and ScaB-XDoc are highly similar containing 47% sequence identity.[20] ScaB-XDoc is slightly longer (Figure 1B), containing small sequence insertions in the Xmod domain and Doc insert regions. Doc insert regions are sequence additions within the conserved Doc sequence that make contact with Xmod in the CttA-XDoc:CohE crystal structure.[20] It was suggested that these Doc inserts provide structural buttresses for the elongated stalk-like conformation of the Xmod.[20] We employed MODELLER 9.17[31] and used CttA-XDoc (PDB 4IU3)[20] as a template to obtain a model of the ScaB-XDoc structure. We then used VMD[32] to align helices 1 and 3 of the model structure (ScaB-XDoc) with those contained in the template (CttA-XDoc:CohE complex) to build a homology model of the ScaB-XDoc:CohE complex (see Figure 1C). For the two sequence gaps in the Doc insert loop regions marked by red arrows in Figure 1D, we performed a loop optimization protocol using ROSETTA.[33] The obtained structure was further refined with MD simulations. Equilibration for 100 ns was performed using NAMD[34,35] through its QwikMD interface.[36] Figure 1C,D shows the structure obtained after loop optimization and equilibration by MD simulation. The equilibration resulted in a stable complex with no major changes in conformation when compared to the initial structural model.

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

Journal of the American Chemical Society Article

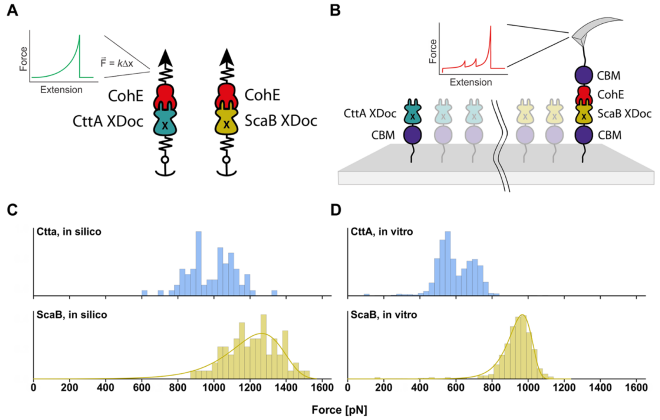


Figure 2. Mechanical dissociation of ScaB-XDoc:CohE and CttA-XDoc:CohE using in silico (left) and in vitro (right) approaches. (A) Schematic illustration of the setup for simulations. Constant velocity SMD simulations were performed at a pulling speed of 5.0 Å/ns. The N-terminus of ScaB-XDoc or CttA-XDoc was restrained in space with a virtual spring, while the C-terminus of CohE was pulled by another spring. The force applied to the spring was monitored during each time step of the simulation. (B) Schematic illustration of the experimental AFM setup, which is analogous to the simulation. Both protein complexes were probed with the same Coh-modified cantilever. (C) Comparison of complex rupture forces arising from mechanical dissociation of ScaB-XDoc:CohE or CttA-XDoc:CohE from in silico MD simulations. Histograms were assembled by taking the highest force reached in each simulation trajectory. (D) Comparison of the forces arising from forced dissociation of ScaB-XDoc:CohE and CttA-XDoc:CohE from in vitro single-molecule AFM at a pulling speed of 1600 nm/s.

Mechanical Dissociation in Silico and in Vitro. To evaluate the behavior of ScaB-XDoc:CohE under force, we first performed in silico SMD simulations[28] employing NAMD[34] through the QwikMD[36] interface. An atom of the ligand was attached to a dummy atom via a virtual spring of known stiffness. The dummy atom was then moved at constant velocity and the force was determined using Hooke's law. Simultaneously, an atom of the receptor was anchored to its initial position with another virtual spring of high stiffness. Specifically, the N-terminus of ScaB-XDoc was fixed and the C-terminus of CohE was pulled at a constant velocity of 5.0 Å/ns (Figure 2A). For comparison, CttA-XDoc:CohE was simulated under the same conditions. To capture the stochastic nature of the thermally driven unbinding process, 75 replicas were performed for each complex.

For in vitro AFM-SMFS experiments, CttA-XDoc and ScaB-XDoc were produced as purified recombinant fusion proteins in *E. coli* bearing an N-terminal carbohydrate binding module (CBM) and an N-terminal ybbR-tag. The ybbR tag allowed for covalent and site-specific surface immobilization, catalyzed by 4′-phosphopantetheinyl transferase (Sfp synthase), while the CBM domain served as a fingerprint/marker domain with known unfolding force and contour length increment that was used for filtering large-scale AFM-SMFS data sets as previously described.[15] CohE was also produced bearing a C-terminal CBM fusion domain and C-terminal ybbR-tag. These surface anchoring sites (N-terminus for XDoc, C-terminus for Coh) precisely mimic the orientation of these domains within the cellulosome network in vivo. Using microwell silicon masks, ybbR-CBM-ScaB-XDoc and ybbR-CBM-CttA-XDoc were immobilized at spatially separated locations on a single glass slide (Figure 2B). This allowed us to probe both proteins with the same cantilever, which was covalently modified with CohE as described previously.[8,13,37,38] In doing so, we could compare the interactions between the two Docs and CohE while circumventing errors that would arise from uncertainty in cantilever calibration, which are known to be on the order of ±10%.[39,40]

Upon contact between the AFM tip and the surface, an XDoc:CohE complex formed, and the cantilever was retracted at constant speed, resulting in unfolding of two CBM domains in series followed by rupture of the receptor−ligand interaction. This experiment was repeated tens of thousands of times and large-scale data sets of force vs extension curves were obtained and screened for 2× CBM unfolding contour length increments to positively identify single-molecule interactions.[10]

Force distributions of both in silico (Figure 2C) and in vitro (Figure 2D) experiments consist of the highest forces reached in each experimental single-molecule force−extension curve or simulation trial. In some cases, the Xmod unfolded at high force, followed by complex rupture at low force. For these simulated traces, the Xmod unfolding force was included in the histogram. If the complex dissociated with the Xmod remaining folded, the complex rupture force was included in the histogram. Both simulations and experiments show an unimodal force distribution for ScaB-XDoc:CohE that could be fitted with a two state Bell−Evans model, whereas CttA-XDoc:CohE yielded a bimodal distribution. The simulations predicted that the ScaB-XDoc:CohE interaction should withstand significantly higher forces than those of the CttA-XDoc:CohE interaction. Remarkably, the experiments con-

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

Journal of the American Chemical Society Article

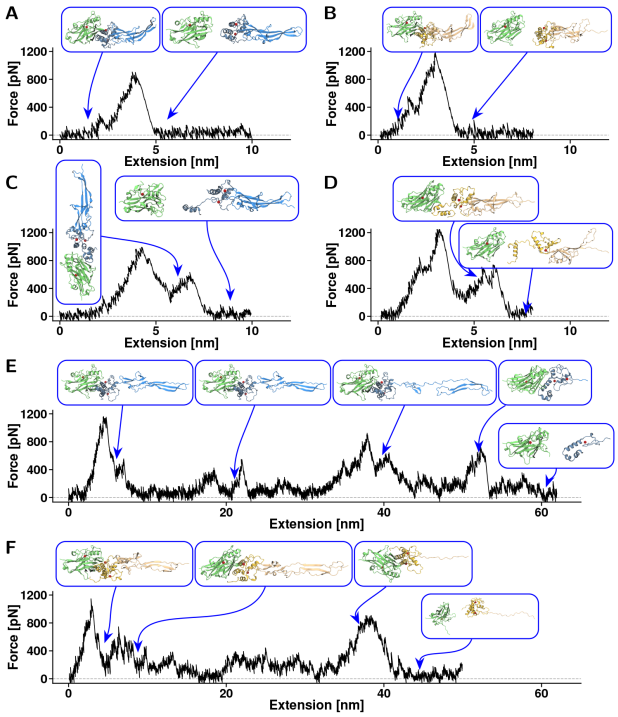


Figure 3. Exemplary force–extension traces from steered molecular dynamics simulations. Plots represent the three types of rupture events observed in CttA-XDoc:CohE and ScaB-XDoc:CohE simulations. (A) For CttA-XDoc:CohE, one-step rupture events were observed in 57 out of 75 replicas. (B) For ScaB-XDoc:CohE, one step rupture events were observed in 34 out of 75 replica simulations. In the one-step event, all three domains (Coh, Doc, Xmod) remained intact following rupture. (C) For CttA-XDoc:CohE, a destabilization of Doc during breakage was observed in 16 out of 75 replicas. (D) For ScaB-XDoc:CohE, Doc destabilization occurred in 39 out of 75 replicas. Doc partial unfolding led to multipeak force vs extension traces that were highly variable in replica simulations. (E) For CttA-XDoc:CohE, Xmod unfoding was observed in 2 out of 75 replicas. (F) For ScaB-XDoc:CohE, Xmod unfolding was also observed in 2 out of 75 replicas. Xmod unfolding led to multipeak traces that were variable in each trajectory. The first and largest peak in these cases corresponded to the loss of contact between Xmod and Doc insert regions.

firmed this finding, with ScaB-XDoc:CohE ruptures reaching forces of ∼1 nN, ∼50% higher than those observed for CttA-XDoc:CohE at the same loading rate. This finding places ScaB-XDoc:CohE among the most mechanically stable protein receptor–ligand complexes reported to date.

Results from simulations and experiments were found to be in very good agreement despite the 6 orders of magnitude difference in force loading rate. Although the agreement is very good, we have previously reported even better agreements for rupture events in bacteria adhesion complexes.[41] Here, our simulation results reveal a rupture force that is approximately 8% lower for the CttA-XDoc:CohE complex and 14% lower for the ScaB-XDoc:CohE complex, when compared to the expected forces based on extrapolations of the experimental results using a Bell–Evans model (see Supporting Information Figure S1). We believe that force field parameters and the use of a homology modeling protocol may be responsible for these small differences.

Both ScaB-XDoc:CohE and CttA-XDoc:CohE interactions were observed to dissociate along at least three different unbinding pathways, which can be seen in Figures 3 and 4A, B. MD simulations revealed the complexes were either broken without prior domain unfolding (one-step event, Figure 3A,B), at nearly the same time as the region connecting the Xmod to the Doc binding helices unfolds (Figure 3C,D), or accompanied by a previous complete Xmod domain unfolding (see Figure 3 E,F). The frequency of occurrence of each of these unbinding pathways is shown in Table 1. In the second class of unfolding trajectories (Figure 3C,D), contact was lost between Doc insert regions and Xmod followed by loss of Doc

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752–14763

Journal of the American Chemical Society Article

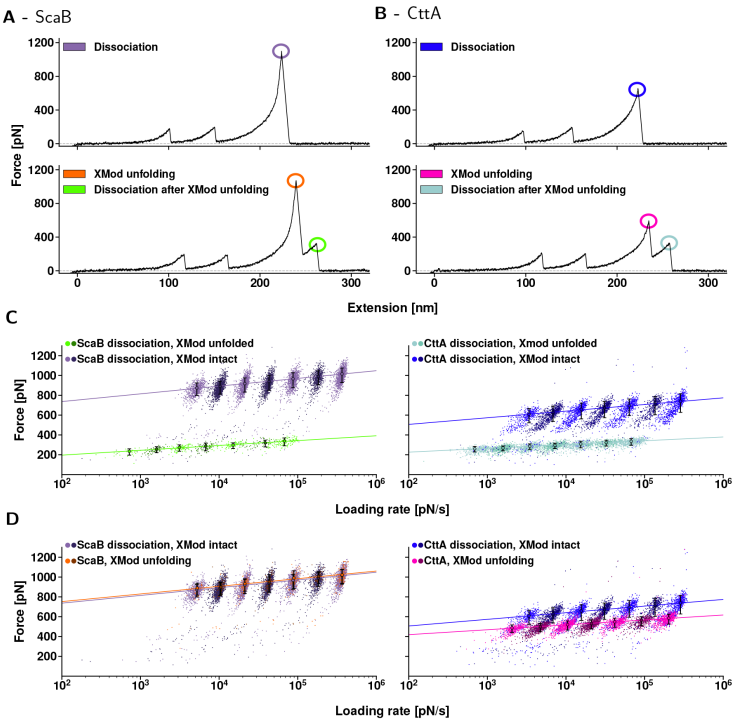


Figure 4. Dynamic force spectra of CttA-XDoc:CohE (left) ScaB-XDoc:CohE (right). (A, B) Typical force–extension traces. First, both CBM-domains unfolded. Their known unfolding behavior served as a fingerprint, indicating that single molecular complexes were probed. Then, for both ScaB-XDoc:CohE and CttA-XDoc:CohE, Xmod remained either folded until complex rupture (upper traces, purple and blue) or unfolded (lower traces, orange and pink) prior to complex rupture. When Xmod unfolding occurred, both complexes ruptured at markedly lower forces (bright green and teal). (C, D) Dynamic force spectra for each class of unfolding or unbinding events that are encircled in (A, B). The colors match the corresponding events in (A, B), and a different color saturation was chosen for every other pulling speed to display the resulting populations more clearly. Data were fitted with the two-state Bell–Evans model. (C) Complex rupture forces. In cases where Xmod remained folded, the ScaB-XDoc:CohE complex ruptured at markedly higher forces than did CttA-XDoc:CohE over the entire range of loading rates tested (purple vs blue). When the Xmod unfolded, the complexes showed nearly identical rupture behavior (bright green vs teal). (D) Comparison of the peak forces reached in both unbinding pathways. The data points either stem from complex rupture events for traces lacking Xmod unfolding, or from unfolding of the Xmod. Interestingly, the most probable unfolding force of ScaB's Xmod is about the same as the ScaB-XDoc:CohE complex dissociation forces that occur when Xmod remained intact (orange vs purple). The same was not true for CttA-XDoc:CohE, where Xmod unfolding forces were surpassed by complex dissociation forces with no prior Xmod unfolding (blue vs pink). The likelihood of observing Xmod unfolding prior to complex unbinding was only 7% for ScaB-XDoc, as compared to 43% for CttA-XDoc.

Table 1. Occurrence of Each Unbinding Pathway Observed in the SMD Simulations

system	one-step rupture	Doc concerted unfolding	Xmod unfolding
ScaB-XDoc:CohE	34/75	39/75	2/75
CttA-XDoc:CohE	57/75	16/75	2/75

secondary structure. In these cases, concerted Doc unfolding and complex rupture exhibited highly variable behavior among the various simulation replicas. It is worth emphasizing that the triggering factor for a possible concerted unbinding-unfolding was the loss of contact between Doc inserts and Xmod. As was proposed by Salama-Alber et al.,[20] these regions would presumably lend structural reinforcement to the stalk-like conformation of these Xmods. Our results agree with the proposed model, providing dynamical information to the atomic data. The sequence additions within the inset regions of the ScaB Doc domain provide for a stronger contact with XMod and are likely playing a role in the higher mechanical

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752–14763

Journal of the American Chemical Society Article

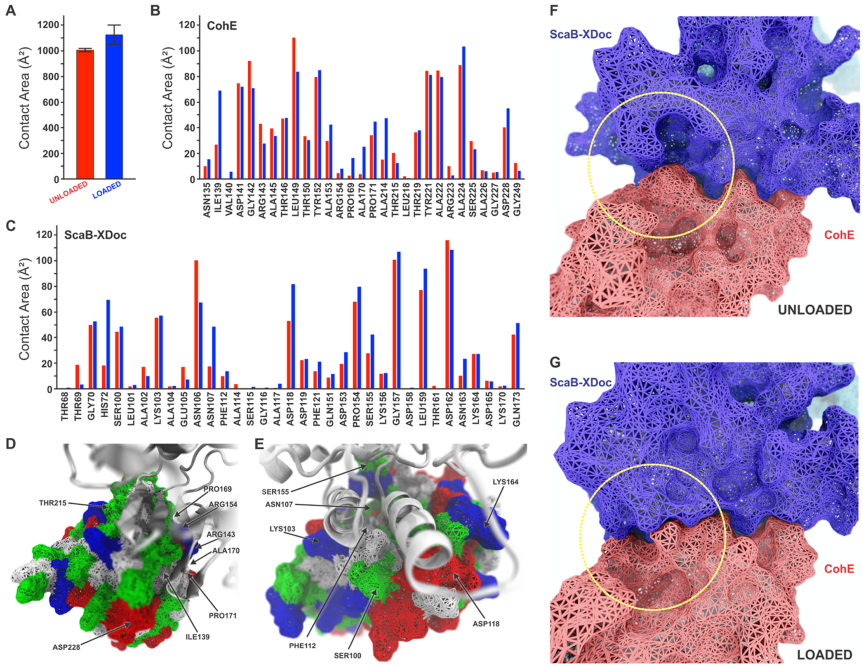


Figure 5. Analysis of ScaB-XDoc:CohE binding interface under force from SMD. (A) Average contact surface area between CohE and ScaB-XDoc under unloaded and loaded conditions. The contact area increases under force load condition. The contacts were calculated using PyContact from 4 ns trajectory windows that were combined for all 75 trajectories. Standard deviations were calculated from these combined windows and are presented as error bars. (B) Surface contact area of interacting residues of CohE under unloaded (red) and loaded (blue) conditions. (C) Surface contact area of interacting residues of Doc under unloaded (red) and loaded (blue) conditions. (D,E) Surface of Doc (D) and Coh (E) showing main interface residues. Hydrophobic residues are shown in white, polar residues in green, and negative and positive residues in red and blue, respectively. (F,G) Rearrangement of binding residues of Coh (red) and Doc (blue) under force. The yellow circle shows the region of the binding interface where a rearrangement increases the contact surface from the unloaded (F) to the loaded (G) configurations.

stability of the ScaB-XDoc:CohE complex as compared with the CttA-XDoc:CohE complex. However, we were not able to identify a clear unfolding path in these concerted events, and Figure 3C,D shows exemplary force traces with intermediates that have no statistical significance. These results indicated that, under high-tensile forces, Doc can unfold nearly simultaneously with complex dissociation. Yet some of the Coh:Doc contacts were present for a couple of nanoseconds before the complex broke apart completely (see Figure 3C,D). Thus, complex rupture shows a multistep force−distance trace, with the final separation of Coh and Doc happening as a shielded (lower force) event.

The frequency of occurrence of the various unbinding pathways obtained from simulations is shown in Table 1. These data indicate that ScaB-XDoc:CohE unbinding is more often accompanied by Doc unfolding than is CttA-XDoc:CohE unbinding. Since ScaB-XDoc and CttA-XDoc share nearly identical sequence and structure at the Coh:Doc binding interface, we attributed this difference to the higher forces reached for ScaB-XDoc:CohE.

In the AFM-SMFS experiments, we could distinguish only two unbinding behaviors for both ScaB-XDoc:CohE (Figure 4A) and CttA-XDoc:CohE (Figure 4B). One pathway that was consistently observed corresponded to simulation pathway 1 (one-step rupture), where Xmod remained folded prior to rupture of the complex at high force (Figure 4A and B, top). A second experimental pathway corresponding to simulation pathway 3 (Xmod unfolding) was also repeatedly observed in the experiments where Xmod completely unfolded prior to complex dissociation (Figure 4A and B, bottom). Following Xmod unfolding for both CttA-XDoc and ScaB-XDoc, the Doc:CohE complex was significantly destabilized and ruptured at lower forces. We did not detect AFM-SMFS traces corresponding to simulation pathway 2 (Doc concerted unfolding), likely due to the nanosecond lifetime of the intermediate state and the finite response time of the AFM

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

Journal of the American Chemical Society Article

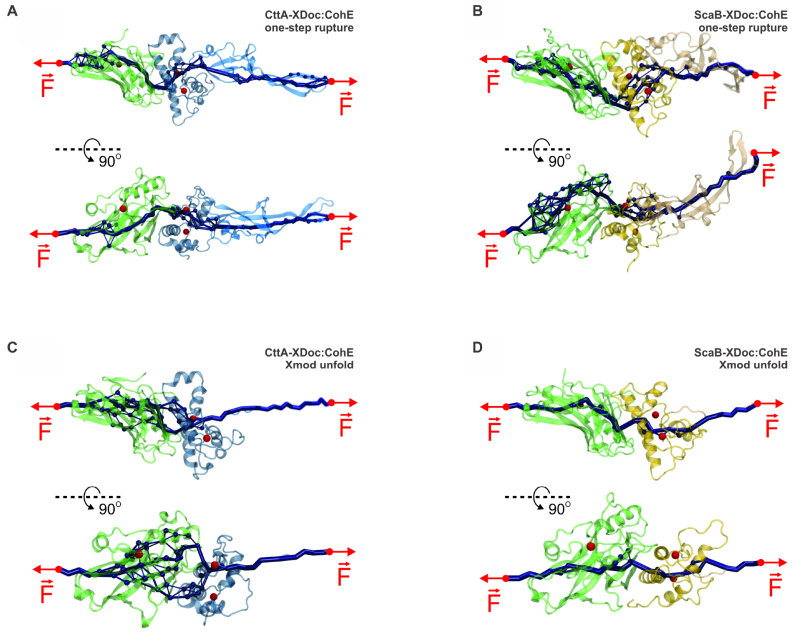


Figure 6. Evolution of force propagation pathways. Force propagation pathways were calculated from correlation-based network analysis carried out using SMD trajectories within 4.0 ns windows just prior to rupture. For one-step rupture of complexes CttA-XDoc:CohE (A) and ScaB-XDoc:CohE (B), force-propagation routes nonparallel to the pulling axis are observed near the Doc:Coh interface. For complex rupture following Xmod unfolding in CttA-XDoc:CohE (C) and ScaB-XDoc:CohE (D) simulation, we again observed redirection of force near XDoc:Coh interface. The thickness of the dark blue tubes represents the number of suboptimal correlation paths connecting two nodes. Thicker tubes correspond to higher likelihood of force propagating through a path. The observed changes in force direction, presented mostly as sharp turns in force propagation routes particularly near the interface, suggest that the ultrastable complexes formed by Coh and Doc achieve mechanostability by directing externally applied forces normal to the pulling axis. Consequently, more force is required to achieve a given separation along the pulling axis.

cantilever. The short-lived state likely transitioned to a completely unbound state before the released contour length of the Doc could be resolved by AFM.

From SMFS experiments, we analyzed force-loading rate data of rupture events obtained over a range of loading rates (i.e., dynamic force spectroscopy[42]), as shown in Figure 4C,D. Out of 16 034 single-molecule traces, unfolding of the Xmod occurred in only 7% of all in vitro traces for ScaB-XDoc, making this unfolding pathway much less frequent, as compared to 43% of all in vitro traces for CttA-XDoc. For both ScaB-XDoc and CttA-XDoc, complex dissociation forces were drastically reduced following Xmod unfolding (Figure 4C). Interestingly, when comparing ScaB-XDoc:CohE vs CttA-XDoc:CohE complexes both with unfolded Xmod domains, higher mechanostability of ScaB-XDoc:CohE was no longer observed. Complex dissociation of ScaB-XDoc:CohE with an unfolded Xmod domain (Figure 4C, bright green) was not stronger than complex dissociation of CttA-XDoc:CohE with its Xmod unfolded (Figure 4C, teal). The similar behavior observed for the two systems when Xmod is unfolded is an indication that the higher forces reached by ScaB-XDoc:CohE stem from differences in the Xmod rather than from differences in the binding interface between Doc and Coh. For ScaB-XDoc:CohE, the dynamic force spectrum of XMod domain unfolding (Figure 4D, orange) was indistinguishable from that of the ScaB-XDoc:CohE complex rupture with an intact Xmod (Figure 4D, purple), indicating that both events may initially arise from a common energy barrier. The same is not true for CttA-XDoc:CohE, where Xmod unfolding events (Figure 4D, blue) occurred at lower forces than the complex dissociation events with no prior Xmod unfolding (Figure 4D, pink). Combining these two events into one histogram yields the bimodal distribution for CttA-XDoc:CohE as displayed in Figure 2D. Rupture force histograms of the data presented in Figure 4 are available in Supporting Information Figures S2–S5. It is worth noting that the exceptionally high rupture forces measured both in silico and in vitro are disproportionate to the equilibrium thermodynamic affinity for both complexes. The

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

K_D of CttA-XDoc:CohE obtained by isothermal titration calorimetry (see Supporting Information Figure S6) was 18 nM while the K_D of ScaB-XDoc:CohE was 149 nM, both of which are comparable to other Coh:Doc interactions. The discrepancy between the equilibrium analysis, where Ctta-XDoc:CohE exhibits higher affinity, and the mechanical analysis, where ScaB-XDoc:CohE exhibits higher rupture forces, highlights how mechanical dissociation can direct unbinding reactions along pathways that are distinct from those dominant in the absence of force.

High-Force Mechanism. An increase in contact area upon application of force was observed in previous SMD simulations of the CttA-XDoc:CohE complex.[12] Taking advantage of the recently developed PyContact,[43] we further analyzed our simulation trajectories by quantifying the contact surface area between CohE and ScaB-XDoc in both loaded and unloaded scenarios. For the loaded scenario, we analyzed all trajectories immediately prior to complex rupture over a simulation window of 4.0 ns. For the unloaded scenario, we analyzed equilibrium MD simulations, also over a 4.0 ns window. The total contact area was found to increase upon loading due to a rearrangement of interacting residues, as shown in Figure 5A. We further analyzed the contact surface area per residue (Figure 5B, C), and plotted the contact surface area for main interacting residues as shown in Figure 5D and E.

ScaB-XDoc (Figure 5D) and CohE (Figure 5E) both exhibit a binding interface comprising a hydrophobic core (white) surrounded by polar (green) and charged residues (blue, positive; red, negative). This residue pattern suggests that the hydrophilic side chains prevent access of solvent to the interior core. Such hot spots of binding are a common motif in protein−protein binding interfaces, referred to as O-rings.[44] Contact area between the two proteins increased upon mechanical loading based on rearrangement of side chains and closing off of the hydrophobic core to solvent (Figure 5F, G).

Exemplary Force Propagation Pathways. To investigate how strain propagates through the protein complex, we employed cross-correlation based network analysis to study one-step rupture events and rupture events following Xmod unfolding. From thermodynamic fluctuation theory, one can infer that paths with high correlation of motion describe the paths along which force propagates through a molecular system.[15,45] The relationship between fluctuations of atoms i and j and the force $\vec{F}_i$ on atom i can be described by

$$\langle \Delta\vec{r}_i \Delta\vec{r}_j^{\,T} \rangle = k_B T \frac{\partial \vec{r}_j}{\partial \vec{F}_i} \tag{1}$$

where $\Delta\vec{r}_i = \vec{r}_i(t) - \langle \vec{r}_i(t) \rangle$ and $\vec{r}_i$ is the position of atom i. If an external force $\vec{F}_i$ acts on atom i and the potential between atoms i and j is steep, the two atoms will move with high positional correlation. By calculating a given element of the correlation matrix $M_{ij} = \langle \Delta\vec{r}_i \Delta\vec{r}_j^{\,T} \rangle$, we can quantify the strength of an interaction potential between i and j. The deformation response of the molecular complex under applied force will be dominated by the stiff interaction potentials, while the soft potentials become insignificant.[15] The propagation of force through the molecules and across the binding interface can then be visualized along the pathways with high correlations of motion. The high-correlation path with the smallest length, namely, the smallest number of amino acid residues, is considered to be the optimal path. When multiple paths of similar length (number of nodes) are found to have high correlation, then these are considered suboptimal paths. It is important to notice that not all residues along these paths need be considered to have the same importance for force propagation. Instead, only residues or interactions that occur in the highest number of suboptimal pathways need to be conserved to guarantee an effective force propagation pathway in the complex.[15,29]

Although the quasi-harmonic approximation of Pearson correlation is commonly employed in correlation analysis of molecular systems,[46−51] it is not a priori justified for complicated biomolecular interactions.[52] However, since our analysis relies on the identification of paths of highest correlation through proximate residues, the quasi-harmonic approximation implied using Pearson correlation is justified, particularly for suboptimal pathway analysis.[15]

In Figure 6, the force propagation pathways through both ScaB-XDoc:CohE (Figure 6A, C) and CttA-XDoc:CohE (Figure 6B, D) complexes are depicted. Figure 6A and B shows correlation networks obtained from one-step rupture event trajectories for ScaB-XDoc:CohE and CttA-XDoc:CohE, respectively. Figure 6C and D shows correlation networks obtained from simulations exhibiting Xmod unfolding prior to rupture. The depicted paths through the system are those along which stress primarily propagates under load, obtained just prior to rupture over a simulation window of 4.0 ns.

For the one-step rupture trajectories, it was previously shown for CttA-XDoc:CohE that force propagated through both binding helices of the Doc and along pathways with large normal components with respect to the pulling axis.[12] The thickness of the lines in Figure 6 indicates how important these pathways are. That is calculated by considering the multiple possible paths of similar length with high-correlation. Important nodes (amino acid residues) are those that are present in a large number of paths. Likewise, important edges (a line connecting two nodes) are those edges that are part of a large number of paths. Therefore, the most important pathways are those that are shown with thicker lines. Although clear differences between the four force pathways shown in Figure 6 are evident, one can easily observe that these important pathways all have "sharp turns" near the interface. These force-propagation routes nonparallel to the pulling axis for ScaB-XDoc:CohE are advantageous for achieving high dissociation forces.

It was previously shown for another ultrastable protein, namely, silk crystalline units, that curving and branching force paths distributed tension through the entire system.[53] A strategy that assumes an indirect path would allow the system to more evenly utilize the interface between binding partners. Such behavior can be simplified in a mechanical picture, where a certain amount of mechanical work $dW = \vec{F}\cdot d\vec{s}$ is required to separate two binding interfaces by a distance Δz and break the interaction. If force $\vec{F}$ is locally perpendicular to the direction of the unbinding axis $d\vec{s}$, a larger force is required to break the interaction than in a scenario where the force path points along the unbinding axis.

Analysis of the correlation maps indicated that the calcium-binding loops were key in distributing forces through the Coh:Doc binding interface. In all scenarios, force propagated through the calcium loops. Even after unfolding of Xmod, force-propagation through the folded Doc remained largely unchanged. As shown in Figure 4C, complex rupture forces were greatly reduced when the Xmod was unfolded, but

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

Journal of the American Chemical Society Article

remained on the order of 200 pN at the given loading rates. Overall, our correlation analysis supports the view that directing force along a path with significant perpendicular component to the pulling axis leads to high mechanical stability.

■ CONCLUSION

The detailed molecular mechanisms of cellulosomes are of broad interest, particularly given their applications in biotechnology[54] and their recent discovery as part of the human gut microbiome.[55,56] The ScaB-XDoc:CohE receptor–ligand complex is essential for bacterial degradation of cellulose, serving as an adhesion domain and anchoring its organism to a primary carbon source.

The multistep computational pipeline we described to understand the ScaB-XDoc:CohE complex involved homology modeling, wide-sampling of steered molecular dynamics simulations, contact area analysis, and finally dynamic network analysis. We developed this pipeline in order to structurally characterize the ScaB-XDoc:CohE complex, computationally predict its high resistance to applied forces, and understand the molecular mechanisms involved in force resistance. We furthermore were able to confirm the predicted high forces experimentally using parallel high volume acquisition of single-molecule AFM force traces.

The additional strength observed for the ScaB-XDoc:CohE complex is in agreement with its function. The actual load on this complex, with its attached Doc-bearing enzymes and other subunits, would seem to justify the improved mechanical stability of the ScaB-XDoc:CohE complex over the CttA-XDoc:CohE. The primary differences in behavior between ScaB-XDoc and the previously characterized CttA-XDoc arise from the Xmod region, which is significantly longer in ScaB. Unfolding of the Xmod domain prior to complex rupture caused a strong destabilization of the complex, and eliminated the improved mechanical stability of ScaB-XDoc:CohE over CttA-XDoc:CohE complexes. As the ScaB's Xmod unfolded far less frequently and at higher forces than its CttA analogue, its larger structure rendered it more robust, improving the stabilizing effect it had on the Doc:CohE complex. Taken together, these results depict the Xmod's function as a structural support of the Doc's binding helices under high loads, acting in part as a mechanical shield to protect the adjacent Doc from unfolding. Our wide-sampling simulation approach with many simulation replicas was critical for interpreting the mechanical behavior of these receptor–ligand complexes, particularly in describing the mechanisms of mechanical stability. In the future, successful merger of SMD simulations with single-molecule mechanical experiments will benefit from the analysis pipeline presented here in order to gain insight into the fascinating interplay between equilibrium protein binding affinity and adhesion mechanics.

■ METHODS

Structural Model. The structure of the CttA-XDoc:CohE complex has been solved by means of X-ray crystallography at 1.97 Å resolution and is available in the Protein Data Bank (PDB: 4IU3).[20] The second system, namely, ScaB-XDoc, had no structure available and was solved by a homology modeling strategy.[30] The construction of Coh models was performed using MODELLER 9.17[31] software that employs spatial restriction techniques based on the 3D-template structure. The best model was selected by analyzing the stereochemical quality check using PROCHECK[57] and overall quality by ERRAT server.[58] The structures of two sequence gaps in the loop regions marked by arrows in Figure 1C were resolved by a loop optimization protocol employing ROSETTA.[33] Using NAMD[34,35] through its QwikMD interface,[36] structures were subjected to 100 ns of equilibrium MD to ensure conformational stability. All structures shown are from postequilibration MD simulations.

Molecular Dynamics Simulations. MD simulations were performed employing the NAMD molecular dynamics package.[34] The CHARMM36 force field[59] along with the TIP3 water model[60] was used to describe all systems. Simulations were carried out 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 cutoff of 14.0 Å was applied to short-range, nonbonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method. The equations of motion were integrated using the r-RESPA multiple time step scheme[34] 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.

In a wide-sampling strategy,[61] 75 steered molecular dynamics (SMD) runs were carried out for a total of 1.5 μs for each system. To characterize the coupling between dockerin and cohesin, SMD simulations[28] with constant velocity stretching (SMD-CV protocol) at 5.0 Å/ns pulling speed were carried out. In all simulations, SMD was employed by restraining the position the N-terminal of ScaB-XDoc or CttA-XDoc, while pulling on the C-terminus of CohE. The procedure is equivalent to attaching one end of a harmonic spring to the end of one domain and pulling on the other domain with a spring, analogous to the experiment. The force applied to the harmonic spring was then monitored during the time of the molecular dynamics simulation.

All analyses of MD trajectories were carried out employing VMD[32] and its plugins. Surface contact areas of interacting residues were calculated employing PyContact.[43] The Network View plugin[29] on VMD was employed to perform dynamical network analysis. A network was defined as a set of nodes comprising all α-carbons connected together by edges. Edges connect pairs of nodes if corresponding monomers are in contact, and two nonconsecutive monomers are said to be in contact if they fulfill a proximity criterion, namely, any heavy atoms (non-hydrogen) from the two monomers are within 4.5 Å of each other for at least 75% of the frames analyzed. As suggested by Sethi et al.,[29] nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths. Conceptually, suboptimal paths can be understood as allosteric signaling pathways or force propagation pathways. Suboptimal paths are defined as paths that are slightly longer than the optimal path, with a given suboptimal path visiting a node not more than once. These multiple communication paths are nearly equal in length, and not all residues along these paths need be considered important for allostery.

Instead, only residues or interactions that occur in the highest number of suboptimal pathways need to be conserved to guarantee an effective pathway for allosteric communication. The thickness of the edges connecting the nodes reveals the least and most used paths. Allostery can be understood in terms of pathways of residues that efficiently transmit energy, here in the form of mechanical stress,[15,62,63] between different binding sites. The dynamical networks were constructed from 4 ns windows of the total trajectories. Using the Floyd–Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was $-\ln 0.5 = 0.69$. To calculate the relevance of off-diagonal terms in the correlation matrix, we employed Carma to calculate a correlation matrix where x, y, and z components of each atom were considered independently.

Protein Preparation for Experiments. All proteins were expressed from pET28a vectors using standard induction and expression protocols in *NiCo21(DE3)RIPL* cells, which were cultivated in ZYM-5052 autoinduction media[64] supplemented with kanamycin and chloramphenicol. After pelleting, cells were lysed by sonication and then centrifuged at 4 °C, 39 000 rcf for 60 min. The

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752–14763

supernatant was filtered through a 0.22 PES membrane (Carl Roth + Co. KG, Karlsruhe, Germany) and applied to Ni-NTA columns (HisTrap FF, GE Healthcare Europe GmbH, Freiburg, Germany). After washing with 6 column volumes of a buffer containing 25 mM TRIS, pH 8.4, 300 mM NaCl, 20 mM imidazole, and 0.5% (v/v) Triton X-100, the bound fraction was eluted with a buffer containing 25 mM TRIS, pH 8.4, 300 mM NaCl, and 300 mM imidazole. All protein solutions were concentrated using Amicon centrifugal filter units (10K MWCO, Merck KGaA, Darmstadt, Germany), followed by buffer exchange against Ca-TBS (25 mM TRIS, pH 7.2, 75 mM NaCl, and 1 mM $CaCl_2$) buffer using ZebaSpin columns (Zeba spin desalting columns 7K, Thermo Fisher Scientific Inc.). Proteins were stored at −80 °C, with glycerol added to 10% (v/v).

Surface Functionalization. Glass surfaces and silicon nitride cantilevers (BioLever mini BL-AC40TS-C2, Olympus, Tokio, Japan) were silanized with (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany). Utilizing silicon masks (CultureWell Reusable Gaskets, Grace Bio-Laboratories, Bend, OR), two spatially separated spots on the silanized glass surfaces were PEGylated with α-maleimindo-hexanoic-ω-NHS polyethylene glycol (NHS-PEG5000-Mal, Rapp Polymere, Tübingen, Germany) dissolved into 25 mM in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer (HEPES), 50 mM, pH 7.5. Cantilevers were PEGylated using the same conditions. Next, the PEGylated surfaces and cantilevers were coupled to Coenzyme A (CoA, 1 mM) in sodium phosphate buffer, pH 7.2. Finally ybbR-CBM(C63S)-ScaB-XDoc or ybbR-CBM(C63S)-CttA-XDoc was covalently immobilized onto the two spatially separated spots on the glass slide via their ybbR-tags in an Sfp-catalyzed ligation at room temperature for 30 min. Each Dockerin was diluted to 0.5 μM in Ca-TBS supplemented with 20 mM $MgCl_2$, while the Sfp enzyme was added to 1 μM. CohE-CBM(C63S)-ybbR was coupled to cantilevers at a concentration of 20 μM under the same conditions.

Single-Molecule Force Spectroscopy. Measurements were performed in Ca-TBS buffer using custom built AFM instruments (driven vertically by PI-731 piezo actuators and laterally by a 25 × 25 mm piezomotor (U-751) in combination with a 100 × 100 nm (P-734) stage, Physik Instrumente, Germany) in conjunction with MFP-3D AFM controllers (Asylum Research, Santa Barbara, CA). Upon approaching the sample surface with the cantilever tip, the complex between CohE and either CttA-XDoc or ScaB-XDoc was formed and the cantilever was retracted from the surface at constant velocities of 100, 200, 400, 800, 1600, 3200, and 6400 nm/s. After each force−extension curve was acquired, the sample was moved laterally by 100 nm in order to probe a different molecule. Every several hundred measurements, the glass slide was moved laterally between protein spots, such that alternatingly CohE-ScaB-Doc and CohE-CttA-Doc complexes were probed throughout the measurement. In this manner, thousands of force−extension curves were automatically acquired over a measurement time of 24−72 h. Single-molecule interaction traces were identified by filtering the data sets using contour length analysis, and identifying only those traces in which two CBM unfolding events were observed.[11] Traces exhibiting two CBM unfolding length increments were then analyzed to create rupture event scatter plots describing the rupture of the XDoc:CohE complexes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b06776.

Protein sequences, ITC data, rupture force histograms, dynamic force spectrum combining in vitro and in silico experiments (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*rcbernardi@ks.uiuc.edu

*michael.nash@unibas.ch

ORCID

Rafael C. Bernardi: 0000-0003-0758-2026

Edward A. Bayer: 0000-0001-7749-5150

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

Michael A. Nash: 0000-0003-3842-1567

Author Contributions

○R.C.B. and E.D. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank Prof. Klaus Schulten (in memoriam), whose contribution to this work, particularly at its early stages, is of inestimable value. Support for this work was provided by the EU seventh Framework Programme NMP4-SL-2013-604530 (CellulosomePlus), the Nanosystems Initiative Munich and the ERC Advanced Grant CelluFuel. M.A.N. acknowledges support from an ERC Starting Grant (#715207). This work was supported by National Institutes of Health (NIH) Grant P41-GM104601, "Center for Macromolecular Modeling and Bioinformatics". R.C.B. is partially supported by the National Science Foundation (NSF) Grant MCB-1616590, "Molecular Modeling of Bioenergetic Systems", as well as the Energy Biosciences Institute. Equilibrium molecular dynamics simulations made use of ALCF/Mira and NERSC/Edison supercomputers as part of the DoE ALCC program. This research partially used resources of the Argonne Leadership Computing Facility (ALCF), which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. This research partially used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Steered molecular dynamics simulations were performed on GPU-accelerated nodes of Blue Waters supercomputer as part of the Petascale Computational Resource (PRAC) Grant The Computational Microscope, which is supported by the National Science Foundation (Award Number ACI-1713784). The state of Illinois and the National Science Foundation (Awards OCI-0725070 and ACI-1238993) support Blue Waters sustained-petascale computing project. We thank Lukas F. Milles for providing the force curve analysis software, as well as Thomas Nicolaus and Angelika Kardinal for laboratory assistance. We thank Marcelo C. R. Melo for the help with SMD analysis and Maximilian Scheurer for the help with PyContact.

■ REFERENCES

(1) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64−74.

(2) Olsson, S.; Wu, H.; Paul, F.; Clementi, C.; Noé, F. Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8265−8270.

(3) Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355−360.

(4) Schönfelder, J.; Alonso-Caballero, A.; De Sancho, D.; Perez-Jimenez, R. The life of proteins under mechanical force. *Chem. Soc. Rev.* **2018**, *47*, 3558−3573.

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

(5) Alonso-Caballero, A.; Schönfelder, J.; Poly, S.; Corsetti, F.; De Sancho, D.; Artacho, E.; Perez-Jimenez, R. Mechanical architecture and folding of E. coli type 1 pilus domains. *Nat. Commun.* **2018**, *9*, 2758.
(6) Milles, L. F.; Unterauer, E. M.; Nicolaus, T.; Gaub, H. E. Calcium stabilizes the strongest protein fold. *Nat. Commun.* **2018**, *9*, 4764.
(7) Gunnoo, M.; Cazade, P.-A.; Galera-Prat, A.; Nash, M. A.; Czjzek, M.; Cieplak, M.; Alvarez, B.; Aguilar, M.; Karpol, A.; Gaub, H.; Carrin-Vzquez, M.; Bayer, E. A.; Thompson, D. Nanoscale Engineering of Designer Cellulosomes. *Adv. Mater.* **2016**, *28*, 5619−5647.
(8) Verdorfer, T.; Bernardi, R. C.; Meinhold, A.; Ott, W.; Luthey-Schulten, Z.; Nash, M. A.; Gaub, H. E. Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. *J. Am. Chem. Soc.* **2017**, *139*, 17841−17852.
(9) Gunnoo, M.; Cazade, P.-A.; Orlowski, A.; Chwastyk, M.; Liu, H.; Ta, D. T.; Cieplak, M.; Nash, M.; Thompson, D. Steered molecular dynamics simulations reveal the role of Ca2+ in regulating mechanostability of cellulose-binding proteins. *Phys. Chem. Chem. Phys.* **2018**, *20*, 22674−22680.
(10) 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.
(11) 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**, No. e50950.
(12) 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. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **2014**, *5*, 5635.
(13) 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.
(14) 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*, No. e10319.
(15) Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E. Mapping mechanical force propagation through biomolecular complexes. *Nano Lett.* **2015**, *15*, 7370−7376.
(16) 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.
(17) 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.* **2013**, *23*, 686−694.
(18) Mechaly, A.; Yaron, S.; Lamed, R.; Fierobe, H.-P.; Belaich, A.; Belaich, J.-P.; Shoham, Y.; Bayer, E. A. Cohesin-dockerin recognition in cellulosome assembly: Experiment versus hypothesis. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 170−177.
(19) Fontes, C. M.; Gilbert, H. J. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* **2010**, *79*, 655−681.
(20) Salama-Alber, O.; Jobby, M. K.; Chitayat, S.; Smith, S. P.; White, B. A.; Shimon, L. J.; Lamed, R.; Frolow, F.; Bayer, E. A. Atypical Cohesin-Dockerin Complex Responsible for Cell Surface Attachment of Cellulosomal Components BINDING FIDELITY, PROMISCUITY, AND STRUCTURAL BUTTRESSES. *J. Biol. Chem.* **2013**, *288*, 16827−16838.
(21) Venditto, I.; et al. Complexity of the Ruminococcus flavefaciens cellulosome reflects an expansion in glycan recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 7136−7141.
(22) Bule, P.; Alves, V. D.; Israeli-Ruimy, V.; Carvalho, A. L.; Ferreira, L. M. A.; Smith, S. P.; Gilbert, H. J.; Najmudin, S.; Bayer, E. A.; Fontes, C. M. G. A. Assembly of Ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes. *Sci. Rep.* **2017**, *7*, 759−759.
(23) Israeli-Ruimy, V.; et al. Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions. *Sci. Rep.* **2017**, *7*, 42355.
(24) Rincon, M. T.; Čepeljnik, T.; Martin, J. C.; Lamed, R.; Barak, Y.; Bayer, E. A.; Flint, H. J. Unconventional mode of attachment of the Ruminococcus flavefaciens cellulosome to the cell surface. *Journal of bacteriology* **2005**, *187*, 7569−7578.
(25) Rincon, M. T.; Dassa, B.; Flint, H. J.; Travis, A. J.; Jindou, S.; Borovok, I.; Lamed, R.; Bayer, E. A.; Henrissat, B.; Coutinho, P. M.; Antonopoulos, D. A.; Miller, M. E. B.; White, B. A. Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, Ruminococcus flavefaciens FD-1. *PLoS One* **2010**, *5*, No. e12476.
(26) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **1997**, *72*, 1568−1581.
(27) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **1996**, *271*, 997−999.
(28) Isralewitz, B.; Gao, M.; Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224−230.
(29) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical networks in tRNA: protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.
(30) Goh, B. C.; Hadden, J. A.; Bernardi, R. C.; Singharoy, A.; McGreevy, R.; Rudack, T.; Cassidy, C. K.; Schulten, K. Computational methodologies for real-space structural refinement of large macromolecular complexes. *Annu. Rev. Biophys.* **2016**, *45*, 253−278.
(31) Webb, B.; Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2014**, *1137*, 1−15.
(32) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(33) Mandell, D. J.; Coutsias, E. A.; Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **2009**, *6*, 551−552.
(34) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(35) Melo, M. C.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. NAMD goes quantum: an integrative suite for hybrid simulations. *Nat. Methods* **2018**, *15*, 351.
(36) Ribeiro, J. V.; Bernardi, R. C.; Rudack, T.; Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Schulten, K. QwikMD—Integrative Molecular Dynamics Toolkit for Novices and Experts. *Sci. Rep.* **2016**, *6*, 26536.
(37) Durner, E.; Ott, W.; Nash, M. A.; Gaub, H. E. Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. *ACS Omega* **2017**, *2*, 3064−3069.
(38) Liu, H.; Ta, D. T.; Nash, M. A. Mechanical Polyprotein Assembly Using Sfp and Sortase-Mediated Domain Oligomerization for Single-Molecule Studies. *Small Methods* **2018**, *2*, 1800039.
(39) Slattery, A. D.; Blanch, A. J.; Ejov, V.; Quinton, J. S.; Gibson, C. T. Spring constant calibration techniques for next-generation fast-scanning atomic force microscope cantilevers. *Nanotechnology* **2014**, *25*, 335705.
(40) Clifford, C. A.; Seah, M. P. The determination of atomic force microscope cantilever spring constants via dimensional methods for nanomechanical analysis. *Nanotechnology* **2005**, *16*, 1666.
(41) Milles, L. F.; Schulten, K.; Gaub, H. E.; Bernardi, R. C. Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science* **2018**, *359*, 1527−1533.

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

(42) Merkel, R.; Nassoy, P.; Leung, A.; Ritchie, K.; Evans, E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* **1999**, *397*, 50−53.
(43) Scheurer, M.; Rodenkirch, P.; Siggel, M.; Bernardi, R. C.; Schulten, K.; Tajkhorshid, E.; Rudack, T. PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations. *Biophys. J.* **2018**, *114*, 577−583.
(44) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1−9.
(45) Greene, R. F.; Callen, H. B. On the formalism of thermodynamic fluctuation theory. *Phys. Rev.* **1951**, *83*, 1231.
(46) Karplus, M.; Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981**, *14*, 325−332.
(47) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. Evaluation of the Configurational entropy for Proteins: Application to Molecular Dynamics Simulations of an α-Helix. *Macromolecules* **1984**, *17*, 1370−1374.
(48) Amaro, R. E.; Sethi, A.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A. A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry* **2007**, *46*, 2156−2173.
(49) Alexander, R. W.; Eargle, J.; Luthey-Schulten, Z. Experimental and computational determination of tRNA dynamics. *FEBS Lett.* **2010**, *584*, 376−386.
(50) Eargle, J.; Luthey-Schulten, Z. A. NetworkView: 3D display and analysis of protein-RNA interaction networks. *Bioinformatics* **2012**, *28*, 3000−3001.
(51) Bernardi, R. C.; Cann, I.; Schulten, K. Molecular dynamics study of enhanced Man5B enzymatic activity. *Biotechnol. Biofuels* **2014**, *7*, 83.
(52) Lange, O. F.; Grubmüller, H. J. Generalized correlation for biomolecular dynamics. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 1053−1061.
(53) Xiao, S.; Stacklies, W.; Cetinkaya, M.; Markert, B.; Gräter, F. Mechanical Response of Silk Crystalline Units from Force-Distribution Analysis. *Biophys. J.* **2009**, *96*, 3997−4005.
(54) Artzi, L.; Bayer, E. A.; Moraïs, S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* **2017**, *15*, 83.
(55) Ben David, Y.; Dassa, B.; Borovok, I.; Lamed, R.; Koropatkin, N. M.; Martens, E. C.; White, B. A.; Bernalier-Donadille, A.; Duncan, S. H.; Flint, H. J.; Bayer, E. A.; Moraïs, S. Ruminococcal cellulosome systems from rumen to human. *Environ. Microbiol.* **2015**, *17*, 3407−3426.
(56) Cann, I.; Bernardi, R. C.; Mackie, R. I. Cellulose degradation in the human gut: Ruminococcus champanellensis expands the cellulosome paradigm. *Environ. Microbiol.* **2016**, *18*, 307−310.
(57) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(58) MacArthur, M. W.; Laskowski, R. A.; Thornton, J. M. Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy. *Curr. Opin. Struct. Biol.* **1994**, *4*, 731−737.
(59) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257.
(60) 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.* **1983**, *79*, 926−935.
(61) Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(62) Seppälä, J.; Bernardi, R. C.; Haataja, T. J. K.; Hellman, M.; Pentikäinen, O. T.; Schulten, K.; Permi, P.; Ylänne, J.; Pentikäinen, U. Skeletal Dysplasia Mutations Effect on Human Filamins' Structure and Mechanosensing. *Sci. Rep.* **2017**, *7*, 4218.
(63) Haataja, T. J.; Bernardi, R. C.; Lecointe, S.; Capoulade, R.; Merot, J.; Pentikäinen, U. Non-syndromic mitral valve dysplasia mutation changes the force resilience and interaction of human Filamin A. *Structure* **2019**, *27*, 102−112.
(64) Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif.* **2005**, *41*, 207−234.

DOI: 10.1021/jacs.9b06776
J. Am. Chem. Soc. 2019, 141, 14752−14763

Supporting information for: Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy

Rafael C. Bernardi*,†,@ Ellis Durner,‡,@ Constantin Schoeler,‡ Klara H. Malinowska,‡ Bruna G. Carvalho,¶ Edward A. Bayer,§ Zaida Luthey-Schulten,†,‖ Hermann E. Gaub,‡ and Michael A. Nash**,⊥,#

†*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 61801 Urbana, IL*

‡*Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany*

¶*School of Chemical Engineering, University of Campinas, 13083-852 Campinas, Brazil*

§*Department of Biomolecular Sciences, Weizmann Institute of Science, 76100 Rehovot, Israel*

‖*Department of Chemistry, University of Illinois at Urbana-Champaign, 61801 Urbana, IL*

⊥*Department of Chemistry, University of Basel, 4058 Basel, Switzerland*

#*Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland.*

@*these authors contributed equally*

E-mail: *rcbernardi@ks.uiuc.edu;**michael.nash@unibas.ch

S1

Protein Sequences

pET28a-ybbR-HRV3C-6xHIS-CBM(C63S)-CttA-XDoc

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASMANTPVSGNLKVEFYNSNPS
DTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWSDHAAIIGSNGSYNG
ITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSN
DYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVPSTQPVTTPPATTKPPATTI
PPSDDPNAVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTVYNEG
YTGDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYEIPVYYNNATL
KDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATTVALSP
STLVGGNPESVYDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTFYTKSS
VDQYKDMAANEPNKLWDIVTGDAEEE*

pET28a-ybbR-HRV3C-6xHIS-CBM(C63S)-ScaB-XDoc

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASMANTPVSGNLKVEFYNSNPS
DTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWSDHAAIIGSNGSYNG
ITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSN
DYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVPSTQPVTTPPATTKPPATTI
PPSDDPNAVVPATNSGDNVSVYYTIETVAGYYFSHDTGVRGNGEAGGFDKNQVVK
ITKYTKDKNGNIIAINDLDLANVNYNGYTPNKAYIDRFGDPAQNPTDQTLANFADN
FAYDIPVYYGGDQLVDENGQPLTVKAYIGVKGDTNLDFIVDGRDATATLTYYARVS
TDNYTEADTPISPAPFITGADDPLDDLAAFLSDVDTNEWDKDNWKLAREDRILDGR
DATNILTYYARASAGDGEYAGLDAQELWNTVVPNRFG*

pET28a-CohE-CBM(C63S)-HIS-ybbR

MGTALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSG
AEGKYATTGYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVAS
GADDDFGADGVMWTVELKVPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQ
GKLMQAYFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAGEPGSVVPSTQPVTTP
PATTKPPATTIPPSDDPNAMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSS

S2

AIDLSKLTLRYYYTVDGQKDQTFWSDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNN
ADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTA
YLNGVLVWGKEPGELKLPRSRHHHHHHGSLEVLFQGPDSLEFIASKLA*

S3

Supplementary Figures

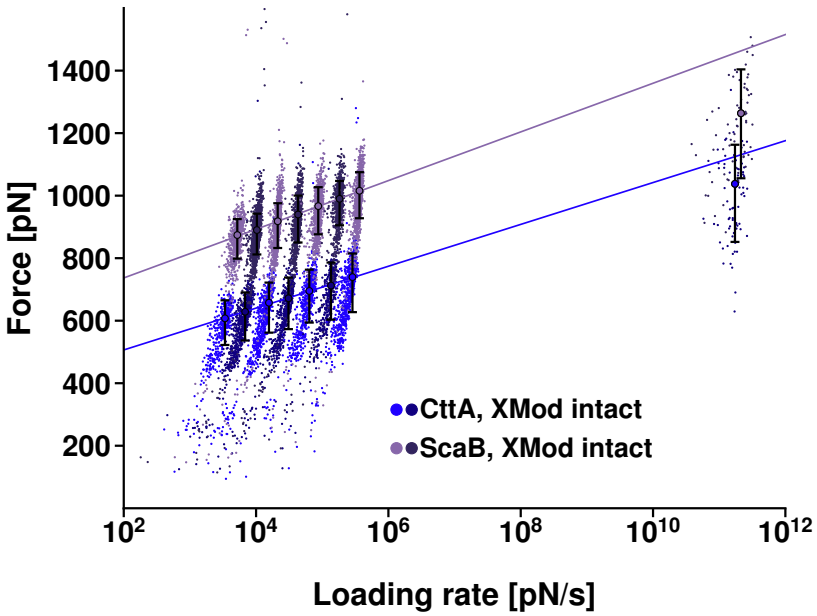


Figure S1: Dynamic force spectrum of curves with no apparent X-module unfolding fitted with a Bell Evans model. Extrapolating this fit to the simulated data reveals that the simulated rupture forces are approximately 8% lower than expected for the CttA-XDoc:CohE complex and 14% lower than expected for the ScaB-XDoc:CohE complex.

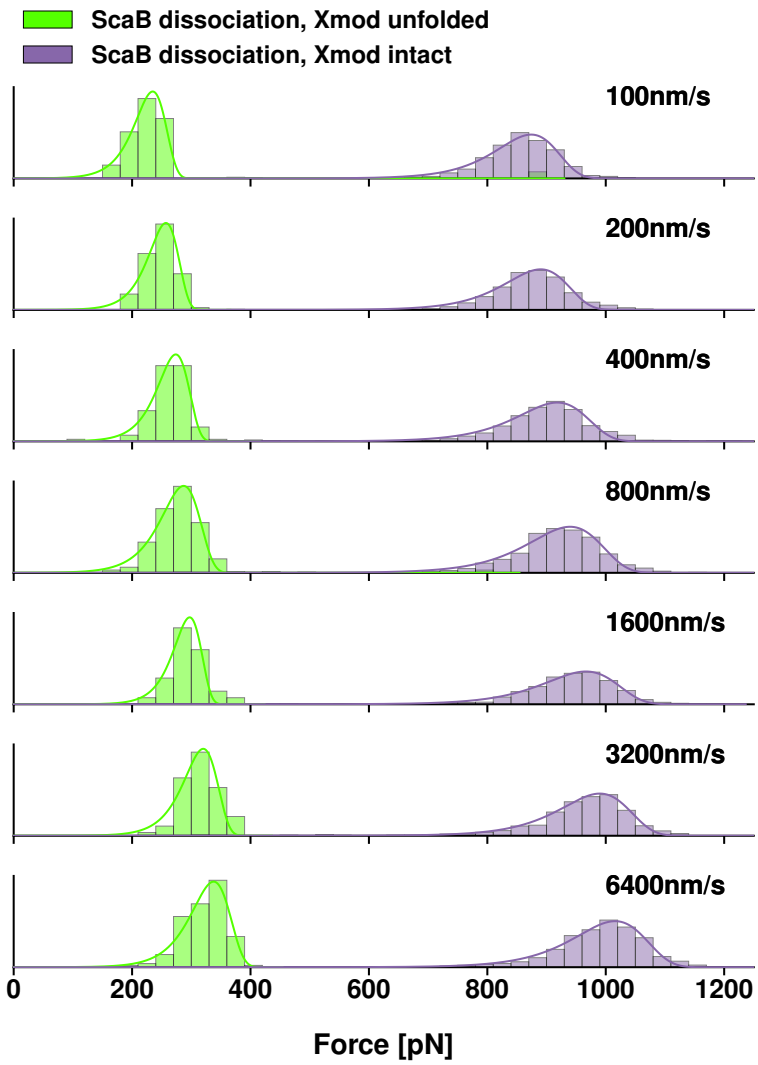


Figure S2: **Complex rupture forces of *Rf* ScaB-XDoc:CohE**. Green bars shows the rupture force of the complex following Xmod unfolding. Purple bars show the rupture force of the complex with Xmod intact.

S5

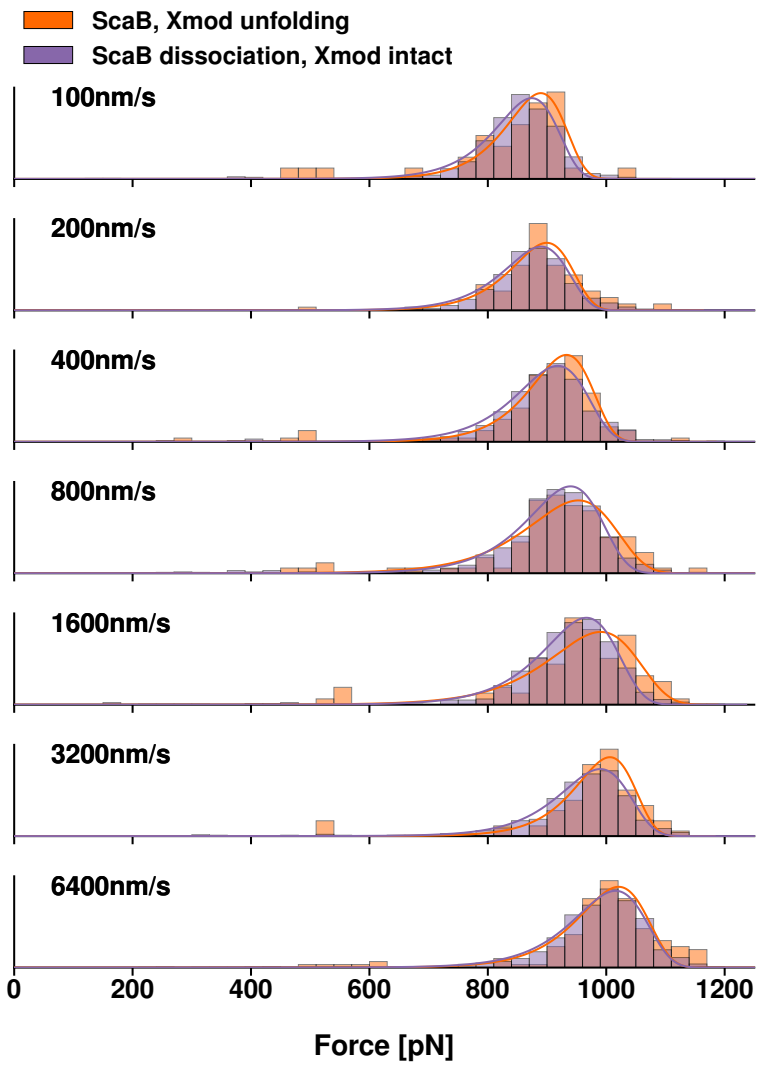


Figure S3: **Complex rupture forces with XMod intact vs. X-module unfolding forces for *Rf* ScaB-XDoc:CohE**. Orange bars show the unfolding force distribution of Xmod domains. Purple bars show the force distribution of complex rupture events with Xmod intact.

S6

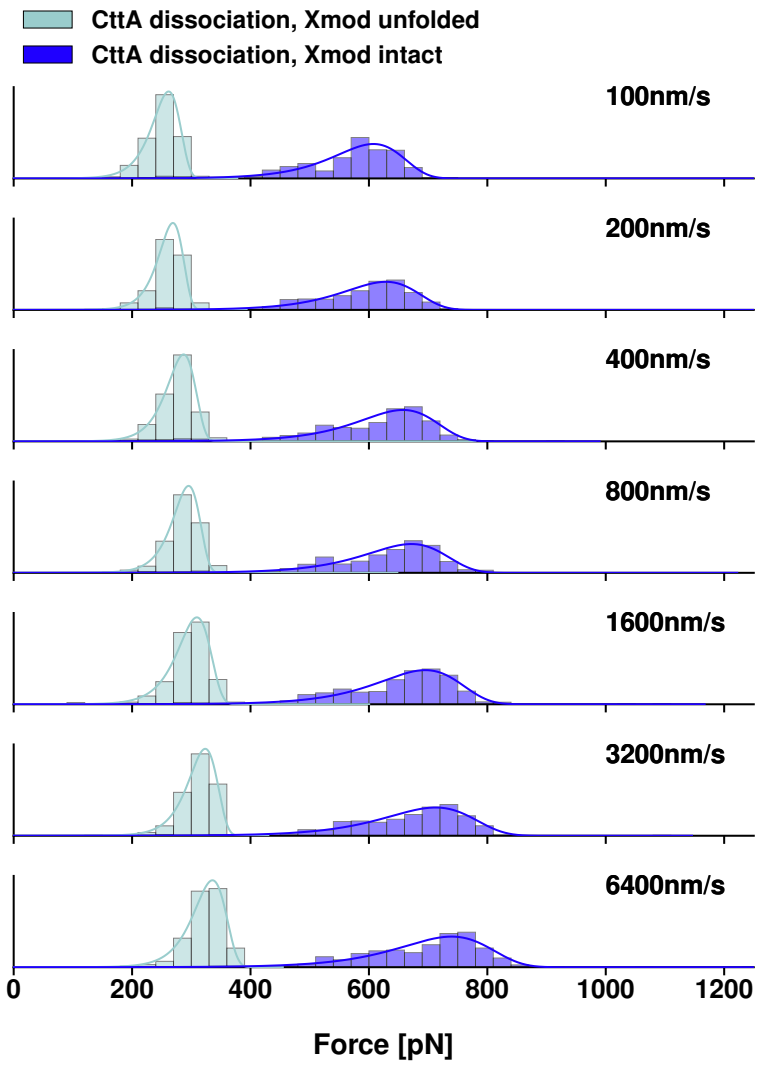


Figure S4: **Complex dissociation forces of *Rf* CttA-XDoc:CohE**. Teal bars shows the rupture forces of the complex following XMod unfolding. Blue bars show the rupture force distribution of the complex with XMod intact.

S7

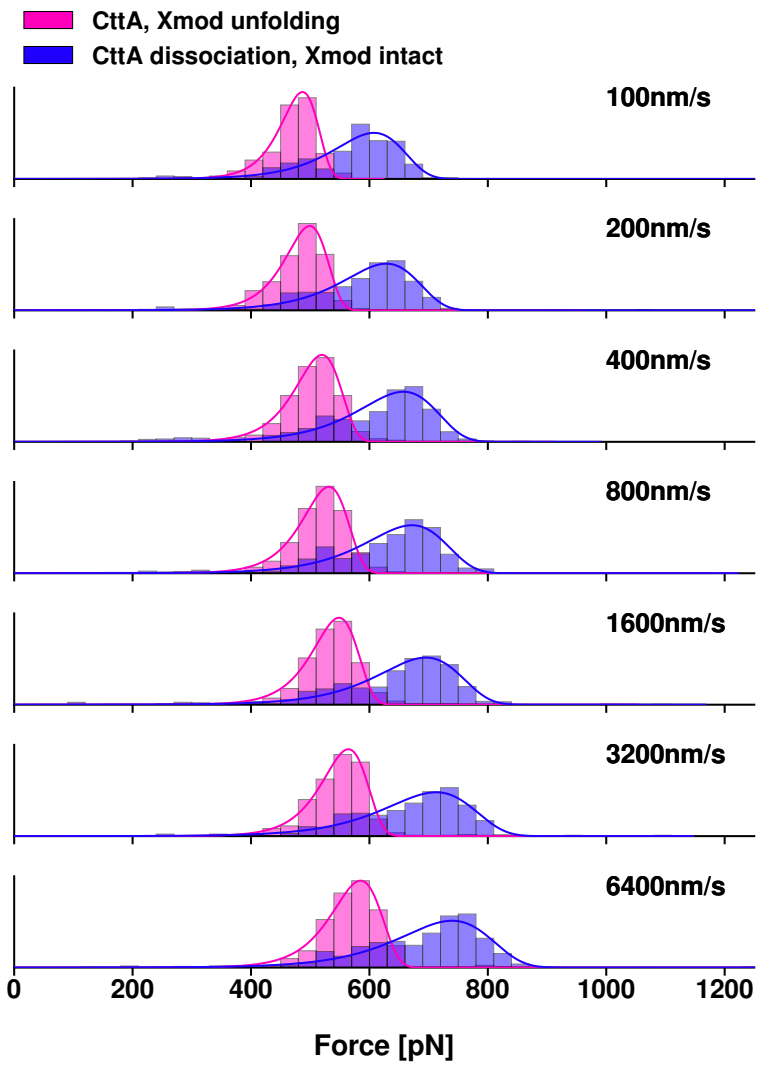


Figure S5: **Complex rupture forces with Xmod intact vs. X-module unfolding forces for *Rf* CttA-XDoc:CohE**. Pink bars show the unfolding force distribution of Xmod domains. Blue bars show the force distribution of complex rupture events with XMod intact.

S8

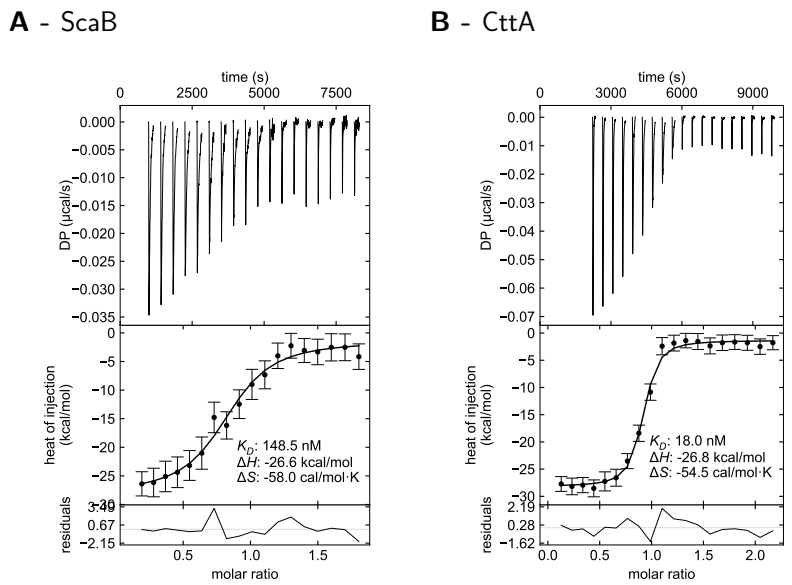


Figure S6: **Isothermal titration calorimetry. A:** The calorimeter's cell contained ybbR-CBM(C63S)-ScaB-XDoc at 4.04 µM, the injection syringe 35.0 µM CohE-NGL. **B:** The calorimeter's cell contained ybbR-CBM(C63S)-CttA-XDoc at 4.04 µM, the injection syringe contained 34.7 µM CohE-NGL.

S9

3.4 Surface Chemistry for Affinity Assays

In this publication, we utilize enzymatic, covalent surface immobilization strategies to perform kinetic rate measurements of receptor-ligand pairs on a biolayer interferometer. This allows us to provide the same anchoring geometry as in SMFS experiments. It also allows us to exclude possible artifacts introduced by the usually employed sensor functionalization via primary amines. My contribution to this publication was to the design of the study, the execution of experiments as well as as the writing of the manuscript. I also wrote analysis software capable of fitting kinetic models to the acquired data.

Enzyme-Mediated, Site-Specific Protein Coupling Strategies for Surface-Based Binding Assays

Wolfgang Ott[†], Ellis Durner[†], and Hermann E. Gaub

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

published in

Angew. Chem. Int. Ed. 2018, 57, 12666 –12669

GDCh Communications Angewandte International Edition Chemie

Protein Immobilization Hot Paper

International Edition: DOI: 10.1002/anie.201805034
German Edition: DOI: 10.1002/ange.201805034

Enzyme-Mediated, Site-Specific Protein Coupling Strategies for Surface-Based Binding Assays

*Wolfgang Ott[+], Ellis Durner[+], and Hermann E. Gaub**

Abstract: *Covalent surface immobilization of proteins for binding assays is typically performed non-specifically via lysine residues. However, receptors that either have lysines near their binding pockets, or whose presence at the sensor surface is electrostatically disfavoured, can be hard to probe. To overcome these limitations and to improve the homogeneity of surface functionalization, we adapted and optimized three different enzymatic coupling strategies (4′-phosphopantetheinyl transferase, sortase A, and asparaginyl endopeptidase) for biolayer interferometry surface modification. All of these enzymes can be used to site-specifically and covalently ligate proteins of interest via short recognition sequences. The enzymes function under mild conditions and thus immobilization does not affect the receptors' functionality. We successfully employed this enzymatic surface functionalization approach to study the binding kinetics of two different receptor–ligand pairs.*

The binding properties of receptor–ligand complexes have been studied in vitro with numerous assays developed during the last decades.[1–3] Mainly, covalent approaches for receptor immobilization have been established to precisely determine on-rate (k_{on}), off-rate (k_{off}), and equilibrium constant (K_d).[4] For these methods, the receptor is immobilized onto a surface and a change in signal upon ligand application is evaluated. While sometimes the terminology "ligand-analyte" is used, throughout this article the molecule immobilized to the sensor surface is called the receptor and its binding partner the ligand. In general, accessible side chains of corresponding amino acids (amine-, carboxyl-, or thiol-groups)[5] can be employed to covalently link the receptor to a surface. However, non-specific attachment requires an electrostatically driven surface pre-concentration step, where the pH and salt conditions of the buffer must be chosen such that the sensor surface and the receptors are predominantly oppositely charged. This pre-concentration step requires a buffer of low ionic strength in order to prevent screening of surface charges, which in turn may cause unfolding and aggregation issues.[6] Additionally, proteins with a low isoelectric point might not be sufficiently protonated, and thus remain negatively charged.

Another challenge with non-specific surface chemistry is that proteins often contain more than one reactive residue, which leads to inhomogeneous surface anchoring. Consequently, sensorgrams of a binding experiment represent the superimposed response of multiple populations of differently attached receptors. Varying attachment sites may strongly influence binding kinetics solely due to the molecules' orientation. The binding behaviour can be altered or binding may even be prevented, especially with receptors immobilized via reactive residues close to their binding interface (Figure 1 A).[7]

In this study, we expand the toolbox for surface functionalization by adapting advances in enzyme-based protein modification strategies to overcome the limitations of non-specific pull-down strategies in binding assays. The employed enzymes are a 4′-phosphopantetheinyl transferase from *Bacillus subtilis* (Sfp),[8] an evolved sortase A (SrtA) from *Staphylococcus aureus* (d59SrtA, P94R/D160N/D165A/K190E/K196T),[9] and an engineered asparaginyl endopeptidase from the plant *Oldenlandia affinis* (OaAEP1) (C247A).[10] All of these enzymes recognize specific amino acid sequences (tags) and covalently attach these tags to other

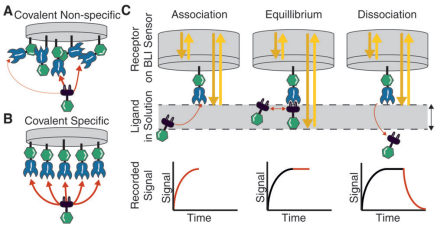


Figure 1. Schematic of BLI Kinetics. A) Non-specific immobilization of the receptor on the sensor in different geometries as a result of several accessible amine-groups. B) Specific and site-directed immobilization of a receptor to a sensor. All receptors are homogeneously orientated. The red arrows in (A) represent different binding geometries with possibly different kinetics, whereas specific attachment (B) provides a uniform population of binders. C) The principle of a BLI kinetic experiment. A receptor-functionalized sensor is immersed into a ligand solution. The increasing signal shows binding of the ligand. When the sensor signal has reached a steady state, the rates of ligand association and dissociation are equal—the system has reached equilibrium. The sensor is then moved to a buffer solution, the receptor starts to dissociate and the detected signal decreases again.

[*] Dr. W. Ott,[+] E. Durner,[+] Prof. Dr. H. E. Gaub
Lehrstuhl für Angewandte Physik and Center for NanoScience
Ludwig-Maximilians-Universität München
Amalienstrasse 54, 80799 Munich (Germany)
and
Center for Integrated Protein Science Munich
(CIPSM), Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: gaub@lmu.de
[+] These authors contributed equally to this work.
Supporting information (including experimental details) and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201805034.

GDCh Communications Angewandte International Edition Chemie

amino acid sequences (SrtA and OaAEP1) or to Coenzyme A (CoA; Sfp). In case of SrtA and OaAEP1, the tags have to be at the termini of the protein, whereas the ybbR-tag (11 amino acids) recognized by Sfp can also be internal (if accessible) since its ligation mechanism does not rely on peptidase activity. These tags can be fused to proteins and employed in surface pull-down strategies, hence allowing homogeneous loading of a surface (Figure 1 B). In single-molecule approaches, such as single-molecule force spectroscopy, site-specific reactions[11–13] are already well established and ensure reliable mechano-probing of receptor–ligand systems without removing the proteins from the surface or cantilever. We adapted these enzyme-based techniques, which enabled us to link a receptor of interest to a sensor surface in very mild reaction conditions while using only low micromolar quantities of receptor.

We chose a biolayer interferometer (BLI) as a development platform because of its fast and flexible assay format. However, it should be noted that the approach presented here is applicable to other surface sensitive techniques, such as surface plasmon resonance (SPR) or quartz crystal microbalance (QCM), since the receptor immobilization relies on the same chemistry. The underlying principle of a BLI makes use of light reflection at interfaces between media of different optical densities to analyse the spectral shift of interference signals upon binding—which effectively modifies the optical path length—to the sensor.[14,15] The interference signal changes whenever binding/unbinding to the sensor fibre occurs (Figure 1 C).

In order to establish our enzyme-based BLI binding assays, we selected two different systems (Table S1, Figure S1 in the Supporting Information). Firstly, we chose GFP-binding nanobodies (LaG9).[16] Nanobodies are small functional single-chain antibodies[17] and are popular tools in diagnostic as well as in therapeutic applications. As a second system, we chose the mechanically highly robust cohesin–dockerin type III complex (CohE–XDocIII) from *Ruminococcus flavefaciens*. As previous single-molecule force spectroscopy studies have shown, its unbinding behaviour under force depends on the anchoring geometry of the cohesin. When immobilized via its C-terminus, a most probable rupture force of around 700 pN (at 100 nN s^{-1})[18] is observed, in contrast to only 100 pN (at 0.7 nN s^{-1})[19] when anchored via its N-terminus. With the site-specific immobilization strategies presented here (Figure 2), we were able to probe the geometry dependence in the absence of force.

Experimental details, traces for Sfp-, SrtA- and OaAEP1-based sensor modifications (Figure S4–S19), and an overview of all possible immobilization geometries (Figure S20) can be found in the Supporting Information. Once the sensors were site-specifically loaded with the protein of interest, they were equilibrated in the same measuring buffer throughout all experiments.

In order to compare the different immobilization strategies, a kinetic binding series with each coupling approach was recorded. Figure 3 A shows an example sensorgram of an SrtA-based experiment. Despite using another GFP variant which differs in the binding epitope (Figure S21), we obtained similar binding kinetics to the reported ones ($K_d = 3.5$ nM, $k_{on} = 2.3 \times 10^6$ M^{-1} s^{-1}, $k_{off} = 8.0 \times 10^{-3}$ s^{-1}) determined with SPR[16] (compare Figure 3 B).

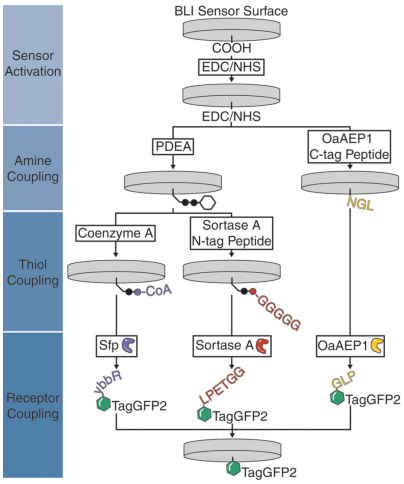
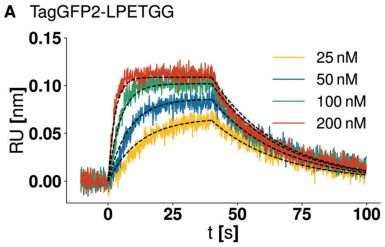


Figure 2. Overview of the different covalent, site-specific immobilization techniques. Left: Sfp catalyses the reaction between ybbR-tag of the TagGFP2 and coenzyme a (CoA). First, the amine-group of PDEA reacts with the EDC/NHS-activated carboxyl groups of the sensor. PDEA can then undergo a thiol exchange reaction with CoA, which presents a free thiol. Middle: SrtA links C-terminal LPETGG with N-terminal GGG. In the case shown here, a C-GGGGG peptide was reacted with the EDC/NHS-PDEA-activated sensor. Right: OaAEP1 recognizes the C-terminal amino acids NGL and fuses it to TagGFP2 containing the N-terminal amino acids GLP. EDC/NHS-activated sensors were reacted with the amine-groups of a KK-GSGS-NGL peptide. All three immobilization methods yield a homogeneous TagGFP2-modified sensor ready for binding kinetic measurements.

The obtained kinetic rates were independent of the functionalization method (specific and non-specific). The site-specific approach anchors proteins at their termini and decreases the chance of binding site obstruction (spatial separation of surface coupling and ligand binding), which thus allows us to determine the unaltered (un)binding rates. This increased reliance is an intrinsic advantage of our site-specific surface functionalization. Based on this, we can compare the data with the non-specifically anchored proteins and conclude that the multiple lysine anchoring possibilities do not obstruct the binding behaviour in the case of TagGFP2–LaG9 interaction. TagGFP2 contains 17 lysines that may take part in the non-specific immobilization procedure. Hence, it is not surprising that enough primary amines non-adjacent to the binding epitope that do not disturb binding are available as anchoring sites. Other receptor–ligand systems might be more strongly affected by the non-specific anchoring (see cohesin–dockerin interaction below). Especially if the surface area at the ligand binding site is charged such that it is electrostatically favoured to make surface contact during the pre-concentration step, the binding site could be obstructed.

GDCh Communications Angewandte International Edition Chemie



B Rates

Ligand	k_{on} [M^{-1} s^{-1}] $\times 10^5$	k_{off} [s^{-1}] $\times 10^{-3}$	K_d [M^{-1}] $\times 10^{-9}$
ybbR-TagGFP2	17.4 ± 1.2	36.3 ± 3.8	20.9 ± 0.7
TagGFP2-LPETGG	15.6 ± 1.2	35.2 ± 4.3	22.6 ± 2.7
GGG-TagGFP2	17.3 ± 0.8	32.6 ± 0.4	18.8 ± 0.7
GLP-TagGFP2	30.6 ± 1.3	70.6 ± 1.5	23.1 ± 0.7
unspecific	37.9 ± 5.0	48.8 ± 8.5	13.2 ± 3.3

Figure 3. Binding kinetics of the TagGFP2-LPETGG receptor with the GLP-LaG9-HIS ligand. A) An example sensorgram of LaG9 ligand binding to the TagGFP2-LPETGG receptor at different concentrations (25, 50, 100 and 200 nM). B) The kinetic rates obtained from performing global fits to sensorgrams for each immobilization method. Values and the respective standard errors are obtained from triplicates.

TagGFP2-NGL and TagGFP2-ybbr could not be fused to a BLI sensor. However, both tags were functional to fuse protein domains in an in vitro bulk reaction. Thus, it appears that both tags are sterically hindered by the GFP domain when used for surface functionalization. A longer linker between GFP domain and the recognition sequence could possibly provide both enzymes (Sfp, OaAEP1) more space to ligate the protein to the sensor.

This enzyme-based and site-specific surface reaction also allowed us to probe the inverse geometry with the nanobody now immobilized to the sensor. Two kinetic titration series, one using the SrtA-based and the other using the non-specific immobilization approach, were recorded (Figure S10). Fits deviated notably from a 1:1 binding model, which might be explained by either ligand–surface interactions or by potential avidity effects should TagGFP2 present more than one binding interface. The ability to site-specifically anchor both binding partners allows us to exclude that the deviations from a simple 1:1 binding model stem from heterogeneous surface preparation due to non-specific protein anchoring. Based on the conducted experiments, we were able to show that all three enzymes can be used for sensor functionalization.

The advantages of defined surface immobilization emerge more clearly when investigating the CohE–XDocIII interaction of *R. flavefaciens*. This cohesin–dockerin pair has already been characterized in bulk studies[20] as well as single-molecule studies.[18,19] However, we were not previously able to non-specifically immobilize the cohesin in a functional state (Figure S3), possible due to a lysine in its binding pocket.[18] However, not only were we able to attach the cohesin site-specifically and in biologically active form with the enzymatic approach, we were also able to do so from either terminus. This was of particular interest since the unbinding behaviour of this complex under external force was shown to strongly depend on the anchoring geometry of the cohesin. The complete sensorgram of the sensor modification can be found in Figure S11. Because of the evidently very low off-rate of the cohesin–dockerin complex, and because no regeneration conditions could be found to force ligand dissociation, we chose to perform a kinetic titration experiment. Full dissociation of the complex would take too long and by far exceed the four hours of total experiment time suggested for BLI. Longer measurements would suffer from evaporation of liquids in the microwell plate, thus falsifying concentrations. Initial experiments showed that sensor-drift effects seemed to exceed the actual ligand dissociation due to the low off-rate (Figure S12). Thus, as recommended,[21] we modified our protocol such that both sensors are loaded with a receptor; in our case they were functionalized site-specifically with cohesin. For referencing, one sensor was only dipped into measurement buffer while the other was presented with ligand. Despite having minimized drift, it became clear that the off-rate of the complex is too low to be assessed through BLI. Too little dissociation occurred so that adequate fitting of the data was not possible. For an exact determination of the off-rate, alternative techniques such as SPR or QCM might be more promising, since they are not limited by evaporation effects and can thus measure for extended periods of time. The apparent low off-rate is common in cohesin–dockerin systems.[22,23] However, a qualitative statement about the (un)binding behaviour of the cohesin–dockerin complex is possible, namely that on- and off-rates appear to be independent of the anchoring geometry (Figure 4).

This stands in contrast to force spectroscopy experiments,[18,19] where the anchoring geometry strongly changes the force necessary to dissociate cohesin and dockerin. By comparing the findings obtained by force spectroscopy with those from site-directed BLI, we can conclude that the different rupture forces are indeed a consequence of force propagation within the receptor–ligand complex, rather than an artefact caused by surface effects or the employed anchoring chemistry.

In summary, the presented strategy provides an efficient means to covalently and site-specifically couple receptors under mild reaction conditions. The employed tags are all small and should not influence the overall functionality of a protein. This makes it viable to use the same constructs for characterization in a surface-based assay as well as for other bulk and single-molecule studies. Moreover, these small tags can be further used for post-translational protein modifications, that is, attachment of a fluorescent dye[24] or an additional protein domain,[13] or as a pull-down technique.[13] Hence, label-free and label-dependent techniques can be used with the same batch of proteins. While the enzymatic approaches presented here are shown for sensor modification in BLI, they can be easily adopted for other assays, such as SPR or QCM. Overall, the site-directed and covalent immobilization techniques present a viable, easily implementable alternative to the non-specific approach. Additionally, no buffer conditions suitable for pre-concentration need to be

GDCh Communications Angewandte International Edition Chemie

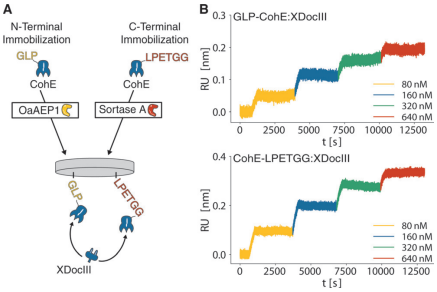


Figure 4. Comparison of N- and C-terminal immobilization of CohE. A) Site-specific surface chemistry enables N- and C-terminal anchoring to be distinguished. B) Kinetic titration series of an OaAEP1-immobilized GLP-CohE (Top) and a SrtA-immobilized CohE-LPETGG (Bottom) sensor with XDocIII ligand binding. SrtA was employed to covalently couple CohE-LPETGG to a polyglycine-modified sensor. OaAEP1 immobilized the GLP-CohE to an NGL-modified sensor. A kinetic titration series was performed by subsequently dipping the sensor into different concentrations of XDocIII (80, 160, 320 and 640 nM) with dissociation steps in measurement buffer.

found, which provides a faster way of establishing assays on new receptor–ligand interactions. Also, due to the specific nature of the surface coupling, signal arising from ligands that bind to non-specifically adsorbed receptors can be subtracted since the reference sensor can be prepared by simply omitting the coupling enzyme. Most importantly, receptor–ligand interactions that were previously inaccessible due to reactive residues in their binding interface or due to electrostatic repulsion can now be site-specifically immobilized and characterized with the enzymatic approaches.

Acknowledgements

This work was supported by the Advanced Grant "Cellufuel" of the European Research Council, the Deutsche Forschungsgemeinschaft through SFB 1032 and a grant by the German-Israeli Foundation for Scientific Research and Development. The authors thank Markus A. Jobst, Tobias Verdorfer and Katherine Erlich for helpful discussions and proofreading. The authors thank Thomas Nicolaus for laboratory assistance.

Conflict of interest

The authors declare no conflict of interest.

Keywords: biolayer interferometry · protein immobilization · protein ligation · site-specific protein modification · sortase A, Sfp, OaAEP1

How to cite: *Angew. Chem. Int. Ed.* **2018**, *57*, 12666–12669
Angew. Chem. **2018**, *130*, 12848–12851

[1] J.-P. Renaud, C.-W. Chung, U. H. Danielson, U. Egner, M. Hennig, R. E. Hubbard, H. Nar, *Nat. Rev. Drug Discovery* **2016**, *15*, 679–698.
[2] M. J. Eddowes, *Biosensors* **1987**, *3*, 1–15.
[3] J. Homola, S. S. Yee, G. N. Gauglitz, *Sens. Actuators B* **1999**, *54*, 3–15.
[4] M. Oliverio, S. Perotto, G. C. Messina, L. Lovato, F. De Angelis, *ACS Appl. Mater. Interfaces* **2017**, *9*, 29394–29411.
[5] H. H. Nguyen, J. Park, S. Kang, M. Kim, *Sensors* **2015**, *15*, 10481–10510.
[6] E. T. Gedig in *Handbook of Surface Plasmon Resonance* (Ed.: R. B. M. Schasfoort), Royal Society of Chemistry, London, **2017**, pp. 173–220.
[7] E. Steen Redeker, D. T. Ta, D. Cortens, B. Billen, W. Guedens, P. Adriaensens, *Bioconjugate Chem.* **2013**, *24*, 1761–1777.
[8] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15815–15820.
[9] B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13343–13348.
[10] R. Yang, Y. H. Wong, G. K. T. Nguyen, J. P. Tam, J. Lescar, B. Wu, *J. Am. Chem. Soc.* **2017**, *139*, 5351–5358.
[11] R. Walder, M.-A. LeBlanc, W. J. Van Patten, D. Edwards, J. A. Greenberg, A. Adhikari, S. R. Okoniewski, R. M. A. Sullan, D. Rabuka, M. C. Sousa, et al., *J. Am. Chem. Soc.* **2017**, *139*, 9867–9875.
[12] H. Lei, C. He, C. Hu, J. Li, X. Hu, X. Hu, H. Li, *Angew. Chem. Int. Ed.* **2017**, *56*, 6117–6121; *Angew. Chem.* **2017**, *129*, 6213–6217.
[13] E. Durner, W. Ott, M. A. Nash, H. E. Gaub, *ACS Omega* **2017**, *2*, 3064–3069.
[14] C. Hänel, G. Gauglitz, *Anal. Bioanal. Chem.* **2002**, *372*, 91–100.
[15] H.-M. Schmitt, A. Brecht, J. Piehler, G. Gauglitz, *Biosens. Bioelectron.* **1997**, *12*, 809–816.
[16] P. C. Fridy, Y. Li, S. Keegan, M. K. Thompson, I. Nudelman, J. F. Scheid, M. Oeffinger, M. C. Nussenzweig, D. Fenyö, B. T. Chait, et al., *Nat. Methods* **2014**, *11*, 1253–1260.
[17] V. Cortez-Retamozo, N. Backmann, P. D. Senter, U. Wernery, P. De Baetselier, S. Muyldermans, H. Revets, *Cancer Res.* **2004**, *64*, 2853–2857.
[18] 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, et al., *Nat. Commun.* **2014**, *5*, 1–8.
[19] C. Schoeler, R. C. Bernardi, K. H. Malinowska, E. Durner, W. Ott, E. A. Bayer, K. Schulten, M. A. Nash, H. E. Gaub, *Nano Lett.* **2015**, *15*, 7370–7376.
[20] O. Salama-Alber, M. K. Jobby, S. Chitayat, S. P. Smith, B. A. White, L. J. W. Shimon, R. Lamed, F. Frolow, E. A. Bayer, *J. Biol. Chem.* **2013**, *288*, 16827–16838.
[21] R. Tobias, S. Kumaraswamy, Application Note 14: Biomolecular Binding Kinetics Assays on the Octet Platform, Pall Life Sciences, **2013**.
[22] S. Jindou, A. Soda, S. Karita, T. Kajino, P. Béguin, J. H. D. Wu, M. Inagaki, T. Kimura, K. Sakka, K. Ohmiya, *J. Biol. Chem.* **2004**, *279*, 9867–9874.
[23] K. Cameron, S. Najmudin, V. D. Alves, E. A. Bayer, S. P. Smith, P. Bule, H. Waller, L. M. A. Ferreira, H. J. Gilbert, C. M. G. A. Fontes, *J. Biol. Chem.* **2015**, *290*, 13578–13590.
[24] W. Ott, T. Nicolaus, H. E. Gaub, M. A. Nash, *Biomacromolecules* **2016**, *17*, 1330–1338.

Manuscript received: April 30, 2018
Revised manuscript received: July 20, 2018
Accepted manuscript online: August 6, 2018
Version of record online: August 31, 2018

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Enzyme-Mediated, Site-Specific Protein Coupling Strategies for Surface-Based Binding Assays

*Wolfgang Ott[+], Ellis Durner[+], and Hermann E. Gaub**

anie_201805034_sm_miscellaneous_information.pdf

WILEY-VCH

SUPPORTING INFORMATION

Table of Contents

WILEY-VCH

SUPPORTING INFORMATION

Methods

All materials employed in this study were at least of analytical purity grade and purchased from Sigma Aldrich Chemie GmbH (Taufkirchen, Germany) and Carl Roth + Co. KG (Karlsruhe, Germany). All buffers were filtered (0.22 µm) and degassed *via* sonication. The pH of the buffers were adjusted at room temperature.

Cloning

The employed nanobody LaG9 was published and characterized by Firdy *et al.*[1]. It was necessary to modify the plain nanobody sequence with the corresponding tags for enzyme-catalyzed pulldown. All constructs were subcloned into a modified pET28a vector with Gibson assembly[2]. In order to perform Gibson assembly, the inserts were ordered as a gene string with overlapping sequences to the plasmid, up- and downstream of the insert or amplified with primers containing the overlaps. The Gibson assembly reaction was performed at 50°C for 1 hour (10 µl 2x HiFi MasterMix; New England Biolabs, Ipswich, MA, USA, mixed with 0.1 nmol of vector and 0.2 nmol of insert). Later modifications were done with a plasmid linearization reaction *via* PCR and recircularization reaction (4.5 µl of PCR product, 1 µl ATP (10 mM, Thermo Fisher Scientific Inc., Waltham, MA, USA), 0.5 µl PEG-6000 (Thermo Fisher Scientific Inc.), 1 µl CutSmart buffer (10x, New England Biolabs), 1 µl T4 Polynucleotidekinase (Thermo Fisher Scientific Inc.), 1 µl *Dpn*I (Thermo Fisher Scientific Inc.), 1 µl T4 Ligase (10 U/µl,Thermo Fisher Scientific Inc.). This reaction was performed for 15 minutes at 37°C and 45 minutes at 22°C.
DH5α cells (Life Technologies GmbH, Frankfurt, Germany) were thawed on ice, 1 µl of the reaction mix was added to the cells to transform them. After 30 minutes on ice a heat shock at 42°C for 1 minute was done. Finally, the cells grew shaking (850 rpm) for 1 hour at 37°C in 1 ml of SOC Medium. Usually 100 µl of the transformed culture was used to streak on a LB-Kanamycin containing plate. Cells were incubated overnight at 37°C. Plasmids of a small amount of clones (less than 5) were amplified and sent to sequencing to verify their sequence.

The gene for the asparaginyl endopeptidase OaAEP1 was ordered *via* gene string (Thermo Fisher Scientific Inc.) and subcloned *via* Gibson assembly. Sortase A (SrtA) was cloned and modified as described by Durner *et al.*[3]. YbbR-HIS-XDockerinIII (XdocIII), CohesinE-HIS-LPETGG and GLP-CohesinE-HIS (CohE) were created with the above described linearization and recircularization reaction. The underlying constructs were ybbR-HIS-CBM-XDocIII and CohE-CBM-HIS-ybbR[4]. 4´-Phosphopantetheinyl transferase (Sfp) was already available in the laboratory.

Protein Expression and Purification

NiCo21(DE3) (New England Biolabs) cells were transformed with the appropriate plasmid (50 ng), and incubated overnight at 37°C on LB-Agar plates with kanamycin (50 ng/µl). One clone was inoculated in 5 ml LB-Kanamycin liquid culture (37°C, 200 rpm, 12-16 hours). This starter culture was used to inoculate a larger 400 ml autoinduction media (ZYM-5052) culture (100 ng/µl kanamycin)[5]. Usually the cultures were incubated for 24 hours (4 hours 37°C, 20 hours 18°C; 120 rpm). Cells were harvested *via* centrifugation (6500 g, 20 minutes, 4°C), the supernatant discarded and the pellets frozen at -80°C until further use.
The frozen pellets were resolubilized in 40 ml lysis buffer (50 mM TRIS (tris(hydroxymethyl)aminomethane)-HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 10 % (v/v) glycerol, 0.1 % (v/v) Triton X-100 and supplemented with 10 µg/ml DNase I and 100 µg/ml lysozyme). To enhance the chemical lysis, the cells were sonicated with a sonication lance twice for 7 minutes on ice (50 % Power, 50 % Cycle; Bandelin Sonoplus GM 70, tip: Bandelin Sonoplus MS 73, Berlin, Germany). The cell lysate was centrifuged (30000 g, 4°C) for 1 hour. The supernatant was filtrated (0.45 µm) and applied to a Ni-NTA column (5 ml HisTrap FF, GE Healthcare Europe GmbH, Freiburg, Germany), since all proteins contained a HIS-tag. The protein was eluted after washing (25 mM TRIS-HCl pH 7.8, 500 mM NaCl, 20 mM imidazole, 0.25 % (v/v) Tween 20, 10 % (v/v) glycerol) with 6 column volumes with elution buffer (25 mM TRIS-HCl, pH 7.8, 500 mM NaCl, 300 mM imidazole, 0.25 % (v/v) Tween 20 (v/v), 10% (v/v) glycerol). The eluted proteins were checked on a SDS-gel for digestion and purity.
Nanobody expressing cultures were inoculated in the autoinduction media minus lactose to prevent autoinduction (ZYM-505). They were grown at 37°C until an OD_{600} of 4, induced with 0.1 mM IPTG and the temperature lowered to 18°C. Cells were harvested (6500 g, 10 minutes, 4°C) after 20 hours and directly lysed and purified. Nanobodies were purified with periplasmic extraction[6]. The cell pellet was resuspended in lysis buffer (100 mM TRIS-HCl, 500 mM Saccharose, 1 mM EDTA, pH 8.0) and incubated on a reaction tube roller (60 rpm, 4°C, 1 hour). A centrifugation step separated the insoluble part from the nanobody containing supernatant (3220 g, 4°C, 10 minutes). The supernatant was again centrifuged (30000 g, 4°C, 1 hour) and filtered (0.22 µm). The filtrated supernatant was applied on a Ni-NTA column as described above.
The buffer of the protein solution was exchanged with ZebaSpin columns (Zeba spin desalting columns 7K, Thermo Fisher Scientific Inc.) to 50 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$ and 20 % (v/v) of glycerol was added. Nanobodies were stored in 20 mM HEPES pH 8.0, 150 mM NaCl and 10 % (v/v) glycerol. Small aliquots of all proteins were flash frozen in liquid nitrogen and stored at -80°C.

WILEY-VCH

SUPPORTING INFORMATION

Production of OaAEP1

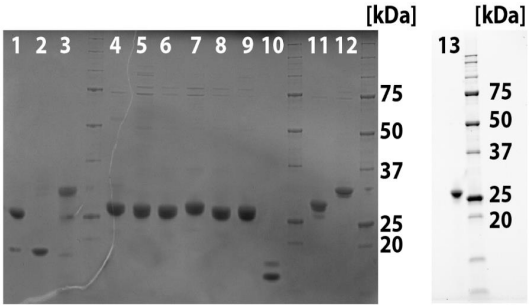
Expression and Purification of OaAEP1 roughly followed recently published protocols[7,8]. *SHuffle® T7* Competent *E. coli* (New England Biolabs) were transformed with a plasmid encoding for HIS-Ub-OaAEP1(C247A) and plated on LB-Agar plates with kanamycin (50 ng/µl). A 5 ml LB-Kanamycin liquid culture was inoculated with one colony (30°C, 200 rpm, 12-16 hours). A 1 l culture of ZYM-505 medium was inoculated with this preculture and grown to an OD_{600} of 4, when expression was induced by adding IPTG to a concentration of 100 µM and the temperature was lowered to 16°C. After 18 hours, cells were harvested by centrifugation (6500 g, 10 minutes, 4°C), the supernatant discarded and the pellets frozen at -80°C until purification.
For purification, the frozen pellet was resolubilized, lysed and applied to a Ni-NTA column as described above. Eluate containing HIS-Ub-OaAEP1(C247A) was pooled and its pH was adjusted to 4 by adding acetic acid, enabling the autocatalytic activation during which both termini of OaAEP1 are cleaved. After 10 hours at 37°C, precipitate was removed by centrifugation and the supernatant was concentrated using Amicon centrifugal filter units (10k MWCO, Merck KGaA, Darmstadt, Germany).
To remove cleaved fragments and to exchange the buffer (to 100 mM NaAcetate, 50 mM NaCl, pH 5.5), the concentrate was applied to a Superdex 75 Increase column (GE Healthcare Europe GmbH). Fractions containing cleaved OaAEP1(C247A) were pooled and supplemented with glycerol to 10% (v/v), then flash-frozen in liquid nitrogen and stored at -80°C.

Biophysical Parameters and SDS-Gels of the Employed Proteins

Table S1. Biophysical parameters of all employed proteins with molecular weight (MW), isoelectric point (pI) and their molecular extinction coefficient (280 nm, 483 nm)

	Protein	MW [Da]	pI	ε_{280}	ε_{483}
1	Sfp	27115	5.85	29130	
2	Sortase A	17408	7.99	14440	
3	OaAEP1	32630	4.96	55030	
4	ybbR-TagGFP2-HIS	29080	5.92	20650	56500
5	HIS-TagGFP2-ybbR	28797	6.00	20650	56500
6	GGG-TagGFP2-HIS	27880	6.09	20650	56500
7	HIS-TagGFP2-LPETGG	28177	6.00	20650	56500
8	GLP-TagGFP2-HIS	28034	6.09	20650	56500
9	His-TagGFP2-NGL	27906	6.09	20650	56500
10	GLP-LaG9-HIS	15648	9.39	28545	
11	CohE-HIS-LPETGG	24945	5.03	32890	
12	GLP-CohE-HIS	22987	5.06	32890	
13	ybbR-XDocIII-HIS	28472	4.45	27850	

Sequences of all employed proteins can be found at the end of this document.



1: Sfp (27.1 kDa)
2: Sortase A (17.4 kDa)
3: OaAEP1 (32.6 kDa)
4: ybbR-TagGFP2-HIS (29.1 kDa)
5: HIS-TagGFP2-ybbR (28.8 kDa)
6: GGG-TagGFP2-HIS (27.9 kDa)
7: HIS-TagGFP2-LPETGG (28.2 kDa)
8: GLP-TagGFP2-HIS (28.0 kDa)
9: HIS-TagGFP2-NGL (27.9 kDa)
10: GLP-LaG9-HIS (15.6 kDa)
11: CohE-HIS-LPETGG (24.9 kDa)
12: ybbR-HIS-Xmod-DocIII (28.5 kDa)
13: GLP-CohE-HIS (23.0 kDa)

Supplemental Figure S1. SDS-Gels of all employed proteins in this study. A final concentration of 5 µM of each protein was applied on a Mini-PROTEAN® TGX Stain-Free™ Precast Gel.

BLI Measurements

All measurements were done at 30°C with an Octet® K2 System (Pall ForteBio LLC.). Sensors functionalized with the receptor were equilibrated for 15 minutes in measurement buffer (25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) casein, 0.1 % (v/v) Tween-20). Two sensors were moved synchronously into wells with ligand. One of the two sensors was functionalized specifically with the receptor, the other was prepared as a reference sensor by omitting the ligating enzyme during functionalization so that no receptor is presented at the sensor surface.

The sensors of the BLI can also be functionalized offline once the reaction conditions are optimized with the optical readout. Offline functionalization means, that the sensors are manually moved from one well to the next one in the sensor tray. The offline functionalization is especially advantageous for the Octet K2, since it only can operate two sensors at a time.

The association was done for 40 seconds (TagGFP2:LaG9) or 400 seconds (cohesin:dockerin), after that the sensors moved back into the wells with measurement buffer and dissociation was recorded for more than 60 seconds (TagGFP2:LaG9) 45 minutes (cohesin:dockerin). Between dissociation and following association of the TagGFP2:LaG9 interaction a steady baseline was awaited (5 minutes).

For the CoA-chemistry, as well as when peptides presenting the N-terminal tags for SrtA or OaAEP1 ligation are used, sensor surfaces are modified *via* disulfide exchange to the PDEA (2-(2-pyridinyldithio) ethaneamine hydrochloride, GE Healthcare Europe GmbH) crosslinker molecule. The present disulfide bond renders it possible to regenerate the sensor by adding a reducing agent. However, it is not possible to measure the binding kinetics under reducing conditions, otherwise the receptor would detach from the sensor. If reducing conditions are required, then sensors can be functionalized *via* the C-terminal tags for SrtA or OaAEP1 *via* the amine-groups of the peptides.

WILEY-VCH

SUPPORTING INFORMATION

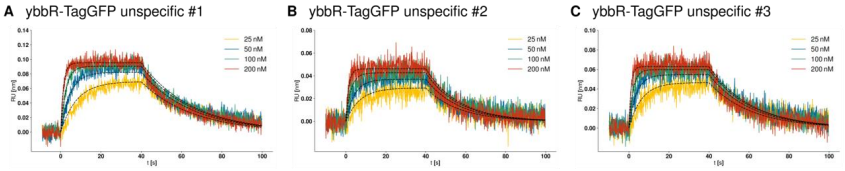
BLI Sensor Modification

Basic Sensor Preparation

For all receptor immobilisation strategies the AR2G - Amine Reactive 2nd Generation sensors (Pall ForteBio LLC, Fremont, CA, USA) were used. Sensors were equilibrated in 50 mM MES (2-(N-morpholino)ethanesulfonic acid) buffer, pH 6.0 for at least 15 minutes at room temperature. After proper rehydratation, they were activated to an amine reactive state by incubating the sensors with a mixture of 50 mM EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) and 50 mM NHS (N-Hydroxysuccinimide) (in 50 mM MES buffer, pH 6.0) for 10 minutes.

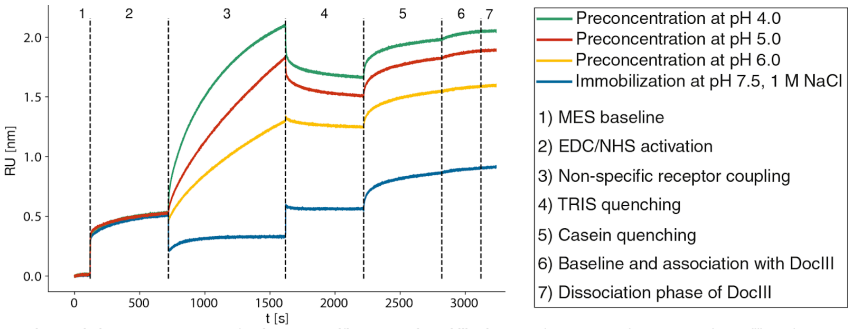
Non-specific Sensor Immobilization

An amine-reactive sensor was activated with EDC/NHS (50 mM in 50 mM MES, pH 6.0, 10 minutes). YbbR-TagGFP2-HIS was loaded to the sensor (100 nM in 50 mM MES, pH 6.0, 10 seconds). For the negative control, one sensor was dipped into only MES buffer. Two quenching steps followed (100 mM TRIS-HCl (pH 8.5) and 25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) casein, 0.1 % (v/v) Tween-20). After equilibrating the sensor in measurement buffer the binding experiment was started by dipping the sensors GLP-LaG9-HIS (25, 50, 100 and 200 nM). **Figure S2** shows the (un)binding behaviour of the non-specific immobilized TagGFP2 and the nanobody LaG9.



Supplemental Figure S2. Specific binding of GLP-LaG9-HIS to the non-specifically immobilized ybbR-TagGFP2. (A)-(C) shows the triplicate runs of GLP-LaG9-HIS vs. ybbR-TagGFP2. The employed concentrations of GLP-LaG9-HIS were: 25, 50, 100 and 200 nM.

A cohesin construct was diluted in different buffers with varying pH but below the cohE's isoelectric point. The EDC/NHS-activated sensor was immobilized with GGG-CohesinE-HIS (0.2 µM) in 10 mM Na-Acetate buffer, pH 4.0, 5.0, 6.0 for 10 minutes. Also a non pre-concentration run was tested with 5 µM GGG-CohE-HIS in 100 mM HEPES, pH 7.5 and 1 M NaCl. Two subsequent quenching steps followed with 100 mM TRIS-HCl (pH 8.5) and measuring buffer (25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) casein, 0.1 % (v/v) Tween-20) for 10 minutes each. None of the conditions could successfully immobilize functional CohE domains (**Figure S3, 3**). CohE also contains lysines in its binding pocket, hence it is possible that the immobilization signal produced was based on these (**Figure S3, Green, Red and Yellow Trace**). A second approach with high salt conditions to shield electrostatic interactions also was not viable to bind enough cohesin to the surface (**Figure S3, Blue Trace**).



Supplemental Figure S3. Sensorgram for the non-specific receptor immobilization. In order to react amine-groups to the equilibrated BLI-sensor **(1)** was activated with EDC/NHS **(2)**. Different non-specific immobilization techniques were probed **(3)** 0.2 nM cohesin was diluted in 10 mM Na-Acetate buffers with a pH range of 4 – 6 (green, red and yellow trace) to enable electrostatic attraction (below the pI of the cohesin). In blue an alternative approach is illustrated: shielding all electrostatic interactions by adding 1 M of NaCl and increasing the cohesin concentration to 5 µM. **(4)** TRIS-Quenching followed to disable all remaining, reactive EDC/NHS-groups. **(5)** Casein passivation followed to prevent non-specific interaction between sensor and ligand. **(6 and 7)** shows negligible association and dissociation of dockerin.

No binding signal could be detected while dipping into a well with dockerin (1000 nM concentration). The concentration of dockerin was many fold above the reported K_d of 20.83 nM[9]. As a consequence, it can be said that with the non-specific approaches no biofunctional cohesin could be immobilized, requiring alternative immobilization methods.

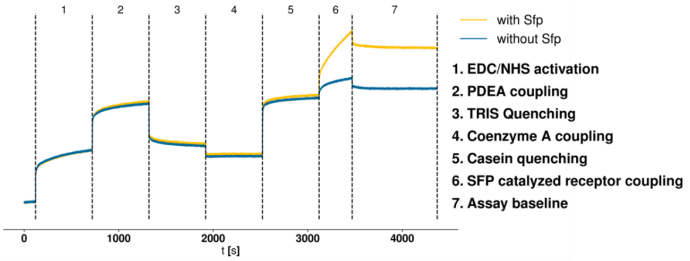
Specific Sensor Immobilization

If the specific protocol relied on a disulfide exchange reaction, the now amine reactive sensors were dipped for 10 minutes into a 40 mM solution of PDEA (dissolved in 50 mM borate buffer, pH 8.5) to covalently immobilize the thiol reactive compound to the biosensor. In order to quench remaining amine-reactive EDC/NHS-groups, the PDEA modified sensors were incubated for 10 minutes in 100 mM TRIS-HCl, pH 8.5.
The sensors for TagGFP2-experiments were functionalized and the interaction measured in one experiment. Sensors for Cohesin-experiments were functionalized and parked in measurement buffer before starting the actual binding experiment. This two step approach was necessary to undercut the recommended 4 hours timeframe for an experiment to avoid evaporation of the solutions in the microwell plate.
Sfp has the benefit that the ybbR-tag can be internally introduced in loops and unstructured regions. OaAEP1 works under acidic conditions and might be attractive to immobilize protein complexes with a low isoelectric point. Regarding TagGFP2 immobilization SrtA proved to be most robust.

Sfp-based Sensor Modification

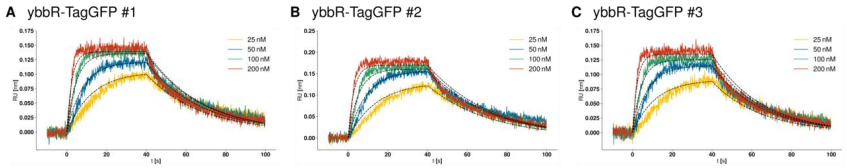
Sensors were modified with PDEA as described under basic sensor preparation. Since Coenzyme A (CoA, Merck KGaA Darmstadt, Germany) contains an accessible thiol-group, the PDEA modified sensors can react with them. 1 mM CoA in coupling buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH 7.2) was fused to the sensors for 10 minutes. A final quenching step (25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) casein, 0.1 % (v/v) Tween-20) passivated the remaining sensor surface against non-specific attachment (10 minutes). The receptor of choice with a ybbR-tag was covalently attached to the CoA by enzymatic ligation via Sfp (5 µM receptor, 1 µM Sfp in 25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 5 mM $MgCl_2$, 0.1 % (v/v) casein, 0.1 % (v/v) Tween-20) for 6 minutes until a surface density of ca. 0.6 - 0.8 nm was achieved. The sensors are now ready for kinetic binding measurements.

Figure S4 shows an example signal trace for the Sfp functionalization. A carboxyl-sensor, in MES-buffer equilibrated, dips into the EDC/NHS solution (**Figure S4, 1**). An increase in signal can be detected. The following PDEA step also shows binding (**Figure S4, 2**). Next an amine quenching is necessary, because CoA also contains amine-groups. When linked *via* its amine-groups, CoA cannot be ligated to the ybbR-tag (**Figure S4, 3**). The thiol exchange reaction between CoA and PDEA can now take place. It should be noted that CoA does not generate any signal upon binding at concentrations below 20 mM (**Figure S4, 4**). Next, the sensor needs to be passivated with a passivation agent suited for the receptor:ligand pair. In this case, casein worked best (**Figure S4, 5**). The site-specific Sfp reaction can now take place. In yellow a trace is shown with added Sfp, in blue only the receptor was added, Sfp was omitted (**Figure S4, 6**). The sensor of the blue trace is dipped into a well with only TagGFP2, hence the small increase in signal is due to non-specific binding. We therefore can correct for the content of non-specific binding during the enzyme-catalyzed reaction (**Figure S4, 6, Yellow Trace**). The positive binding signal can be later corrected with the non-specific adsorption value. The almost vertical jumps in the signal traces are due to changing buffer conditions which cause an abrupt change in the signal.



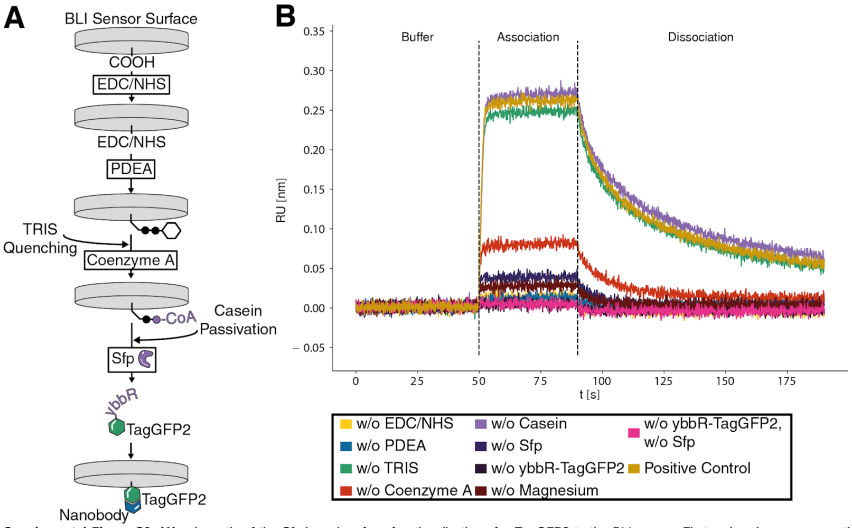
Supplemental Figure S4. Exemplary sensorgram for the immobilization of ybbR-TagGFP2. First the carboxyl-group is activated with EDC/NHS **(1)** which reacts with the amine-group of PDEA **(2)**. **(3)** TRIS quenches all the unreacted, but active EDC/NHS-groups. Coenzyme A **(4)** shows no increase in the sensorgram but nevertheless links to the PDEA *via* its thiol-group. **(5)** is a passivation step with casein to prevent non-specific adsorption of proteins to the sensor. **(6)** shows the specific Sfp reaction (yellow) and the non-specific adsorption of the protein mix to the sensor (blue). **(7)** is the equilibration to baseline in measurement buffer.

The functionalized sensors are then dipped into wells with GLP-LaG9-HIS with different concentrations (25, 50, 100 and 200 nM) and buffer to measure dissociation of the nanobody (**Figure S5**).



Supplemental Figure S5. Specific binding of GLP-LaG9-HIS to the immobilized ybbR-TagGFP2. (A)-(C) shows the triplicate runs of GLP-LaG9-HIS vs. ybbR-TagGFP2. The employed concentrations of GLP-LaG9-HIS were: 25, 50, 100 and 200 nM.

To exclude any non-specific binding artefacts, several controls were made (**Figure S6**). In an offline-immobilization experiment each of the single components was left out per immobilization run. After equilibration in measurement buffer, these sensors were dipped into a 500 nM LaG9 solution to probe for binding. Skipping crucial steps, such as activation of sensors with EDC/NHS (yellow trace), PDEA (blue tace) leads to no signal as the sensor cannot be subsequently modified. Skipping the Coenzyme A (red trace) functionalization step leaves PDEA-activated sensors which react with the TagGFP2 containing accessible cysteines. Sfp shows no enzymatic activity without Magnesium (dark red trace) as no binding signal is observed. Also, no signal is obtained when leaving Sfp (purple trace), ybbR-TagGFP2 (dark purple trace) or both (pink trace) from the specific loading reaction. Skipping the amine quenching reaction with TRIS (green trace) leds to no observable differences. This might be because subsequent reactions also contain reactants with primary amines, which are able to quench the EDC/NHS activated sensors or PDEA already reacted to all accessible EDC/NHS activated carboxyl-groups. Omitting the passivation step with casein prior to immobilization (light purple) has no impact, most probably because all subsequent buffers did contain casein, providing sufficient passivation.



Supplemental Figure S6. (A) schematic of the Sfp-based surface functionalization of a TagGFP2 to the BLI-sensor. First carboxyl-groups are activated with EDC/NHS and coupled with PDEA. PDEA is thiol reactive and couples in this case Coenzyme A. Coenzyme A is recognized by Sfp and covalently links a ybbR-containing TagGFP2. The sensor is now able to bind the corresponding nanobody (LaG9). **(B)** shows association and dissociation of nanobody to the TagGFP2 functionalized sensor. Each trace represents one control where one reagent was left out in the proper surface chemistry. Yellow ochre represents the positive control with all added substances.

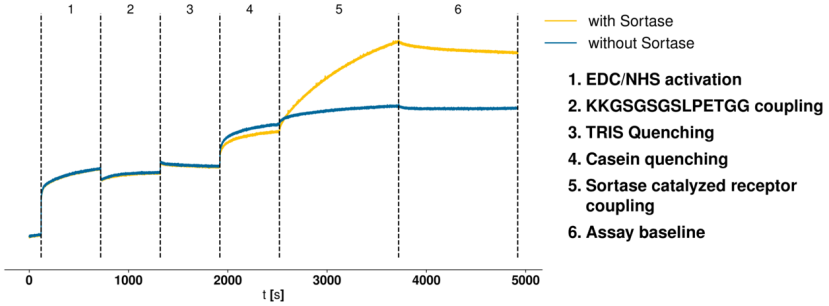
Sortase A- and OaAEP1-based Sensor Modification

For the peptide-based SrtA or OaAEP1-mediated sensor functionalizations, two different strategies were employed. Peptides containing the C-terminal SrtA or OaAEP1 recognition sequences were directly coupled to the sensors *via* primary amines.
For functionalizing the sensors with the N-terminal recognition sequences, the 'adapter molecule' PDEA was employed to pull-down the peptides bearing a cysteine at their C-termini *via* a disulfide exchange to the cysteines' SH-groups. It is important to note that the covalent ligation product is able to be recleaved again by SrtA. Since the reaction intermediate LPET・is prone to hydrolysis, functionalization can decay again for prolonged reaction times when sensors presenting the C-terminal recognition sequence LPETGG

are used. While OaAEP1 is a transpeptidase as well, production of shunt product due to hydrolysis seems to occur at a much slower rate when compared to SrtA, as we could not observe any degradation of the functionalization even at long reaction times.

Sortase A N-tag-based Modification

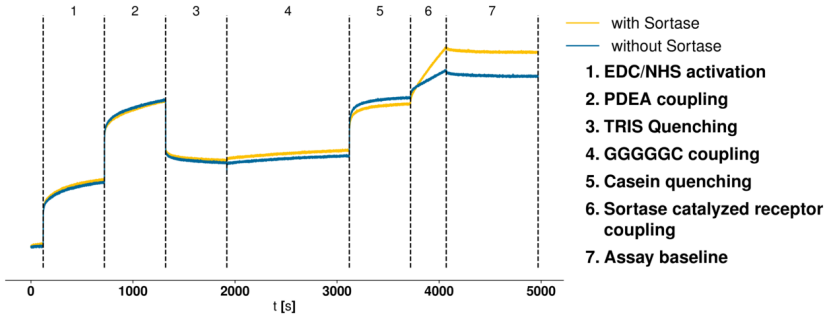
Sensors were brought to an amine reactive state as described under basic sensor preparation. Amine reactive sensors were then dipped for 10 minutes into a solution of 200 µM KKGSGSGSLPETGG peptide (GenScript, Piscataway, USA) in 10 mM HEPES-HCl, pH 7.2. LPETGG does not contain any amines besides its N-terminal one, so in order to improve reactivity, to increase the isoelectric point of the peptide (easier pre-concentration), and to add a linker sequence the sequence KKGSGSGS was added to the peptide (KKGSGSGS-LPETGG). With two lysines located at the N-terminus, conjugation can occur to any of the three amine-groups of the peptide, the C-terminal SrtA recognition sequence is connected by a 3xGS linker. To quench any potentially remaining amine-reactivity, sensors were then incubated in 100 mM TRIS-HCl, pH 8.5 for 10 minutes. To prevent non-specific adhesion in the following steps, sensors were then passivated (25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) casein 0.1 % (v/v) Tween-20). For functionalization with the desired receptor (exhibiting the N-terminal Sortase-Tag GGG), sensors were incubated with 10 µM receptor, 0.5 µM SrtA and 0.1 % (v/v) casein in 25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Tween-20) until a desired functionalization density (0.6 - 0.8 nm) is reached, usually for about 15 to 20 minutes. **Figure S7** shows an exemplary sensorgram of the immobilization of GGG-TagGFP2 with SrtA.



Supplemental Figure S7. Surface chemistry for N-terminal Sortase A-based sensor modification. In order to site-specifically couple TagGFP2 some prior steps are necessary: **(1)** the carboxyl-groups are activated with EDC/NHS. **(2)** amine-groups of the KKGSGSGSLPETGG-peptide couples to the activated carboxyl-groups. **(3)** remaining activated carboxyl-groups are quenched with TRIS. **(4)** the BLI sensor is passivated with casein to prevent any non-specific binding. **(5)** the yellow trace shows the coupling reaction of GGG-TagGFP2 with SrtA. The blue shows the non-specific adsorption of GGG-TagGFP2 to the sensor. **(6)** shows the baseline after enzyme-based coupling.

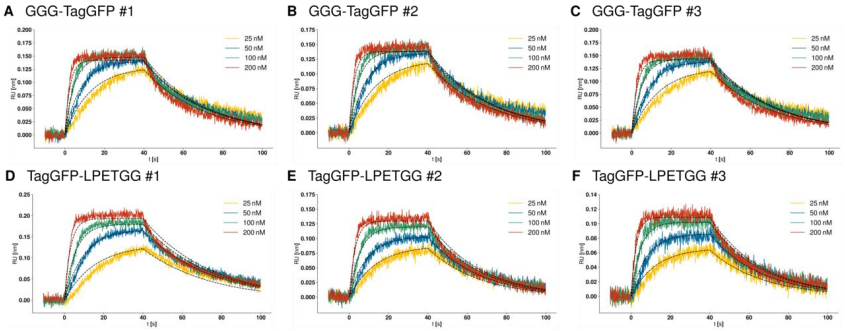
Sortase A C-tag-based Modification

Sensors were prepared to a thiol reactive state by modifying them with PDEA as described under basic sensor preparation. Thiol reactive sensors were then loaded with GGGGGC peptide by performing a disulfide exchange reaction, replacing the PDEA's leaving group with the peptide. This reaction was performed by incubating the sensors for 20 minutes in a solution of 200 µM peptide in 10 mM sodium acetate buffer at pH 4.5. Subsequently, sensors were then passivated (25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) casein 0.1 % (v/v) Tween-20). For functionalization with the desired receptor (exhibiting the C-terminal Sortase-Tag LPETGG), sensors were incubated with 5 µM receptor, 0.2 µM SrtA enzyme and 0.1 % (v/v) casein in 25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Tween-20) until a desired functionalization density is reached (0.6 - 0.8 nm), usually for about 5 to 10 minutes. **Figure S8** shows an exemplary sensorgram of the immobilization of TagGFP2-LPETGG with SrtA.



Supplemental Figure S8. Surface chemistry for C-terminal Sortase A-based sensor modification. In order to site-specifically couple TagGFP2 some prior steps are necessary: **(1)** the carboxyl-groups are activated with EDC/NHS. **(2)** amine-groups of the PDEA react with the EDC/NHS activated carboxyl-groups. **(3)** remaining activated carboxyl-groups are quenched with TRIS. **(4)** the thiol-group of the cysteine of the GGGGGC-peptide couples to the PDEA. **(5)** the BLI sensor is passivated with casein to prevent any non-specific binding. **(6)** the yellow trace shows the coupling reaction of TagGFP2-LPETGG with SrtA. The blue shows the non-specific adsorption of TagGFP2-LPETGG to the sensor. **(7)** shows the baseline after enzyme-based coupling.

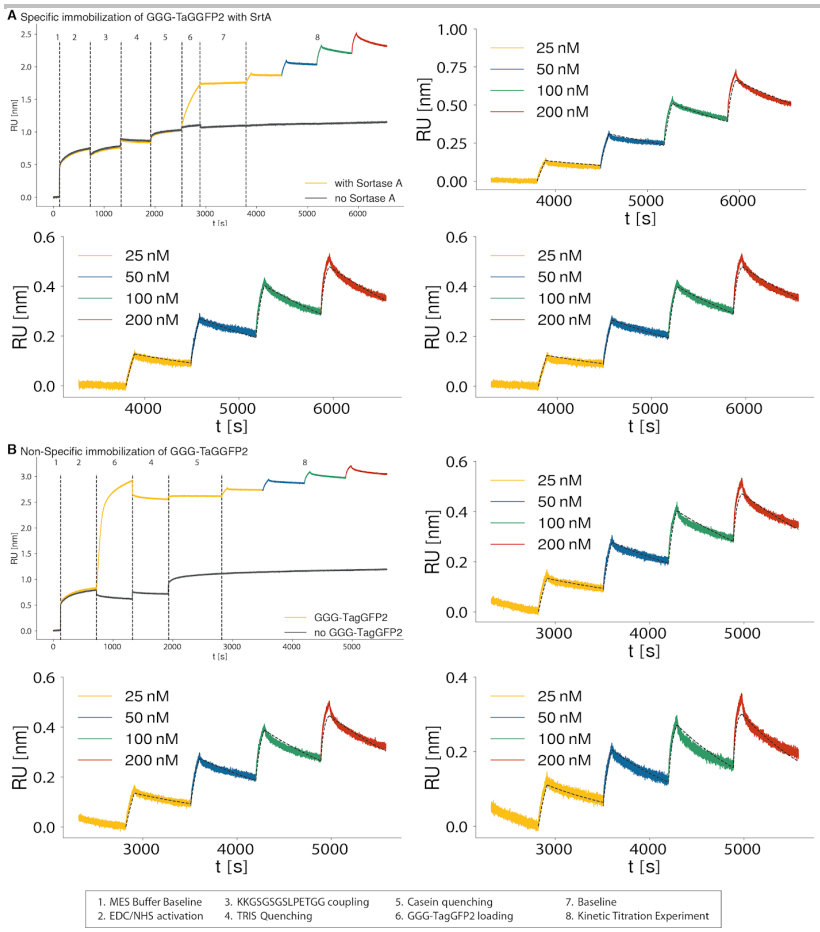
Independent of the TagGFP2 anchoring geometry (N- or C-terminal), the functionalized sensors are then dipped into wells with GLP-LaG9-HIS with different concentrations (25, 50, 100 and 200 nM) and buffer to measure dissociation of the nanobody (**Figure S9**).



Supplemental Figure S9. Specific binding of GLP-LaG9-HIS to the immobilized GGG-TagGFP2 and TaGFP2-LPETGG. (A)-(C) shows the triplicate runs of GLP-LaG9-HIS vs. GGG-TagGFP2. **(D)-(E)** shows the triplicate runs of GLP-LaG9-HIS vs. TagGFP2-LPETGG. The employed concentrations of GLP-LaG9-HIS were: 25, 50, 100 and 200 nM.

WILEY-VCH

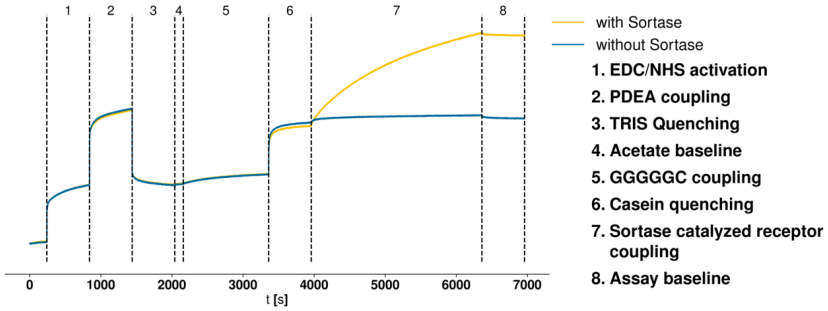
SUPPORTING INFORMATION



Supplemental Figure S10. Specific binding of ybbR-TagGFP2-HIS to the immobilized GGG-LaG9-HIS. (A) shows the sensorgram of the SrtA-based site-specific surface chemistry. **(B)** shows the non-specific immobilization run of GGG-LaG9-HIS. The three zoom in panels show results of the triplicate runs of the association and dissociation of ybbR-TagGFP2-HIS to the two different set of experiments. The employed concentrations of ybbR-TagGFP2-HIS were: 25, 50, 100 and 200 nM. It is worth noting, that the signal ratio between loaded TagGFP2 and obtained and specific binding signal of the LaG9 nanobody is much better with the SrtA-based strategy than the non-specific strategy. specific: Fitting of the kinetic titration series gave mean values of k_{on} =1.17±0.19 e+05 M^{-1} s^{-1}, k_{off}=5.06±0.47 e-04 s^{-1}, K_d=4.39±0.33 e-09 M^{-1} for the SrtA coupled LaG9. For the unspecifically immobilized LaG9 we obtained mean values: k_{on} =1.69±0.27 e+05 M^{-1} s^{-1}, k_{off}=7.16±0.14 e-04 s^{-1}, K_d=4.20±0.18 e-09 M^{-1}. The error is the standard deviation.

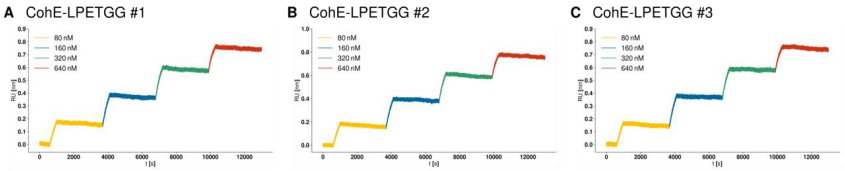
It should be noted that the fit slightly deviates from the dissociation behaviour which might be caused by diffusion limitation or mass transport effects[10].

Figure S11 shows the sensorgram of the C-terminal surface coupling of CohE *via* SrtA.



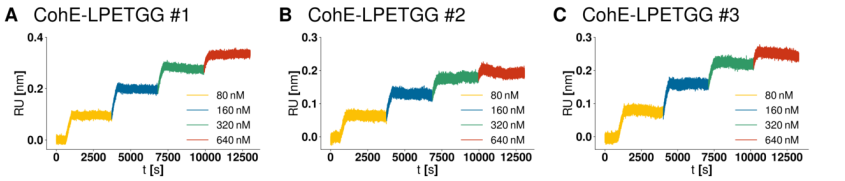
Supplemental Figure S11. Surface chemistry for C-terminal Sortase A-based sensor modification. In order to site-specifically couple CohE some prior steps are necessary: **(1)** the carboxyl-groups are activated with EDC/NHS. **(2)** amine-groups of the PDEA react with the EDC/NHS activated carboxyl-groups. **(3)** remaining activated carboxyl-groups are quenched with TRIS. **(4)** baseline in acetate buffer to equilibrate sensor **(5)** the thiol-group of the cysteine of the GGGGGC-peptide couples to the PDEA. **(6)** the BLI sensor is passivated with casein to prevent any non-specific binding. **(7)** the yellow trace shows the coupling reaction of CohE-LPETGG with SrtA. The blue shows the non-specific adsorption of CohE-LPETGG to the sensor. **(8)** shows the baseline after enzyme-based coupling.

The cohesin sensor was then dipped into dockerin containing wells to measure association and dissociation (**Figure S12**). Concentration of the dockerin wells were 80, 160, 320 and 640 nM.



Supplemental Figure S12. Specific binding of ybbR-HIS-XDocIII to the immobilized CohE-LPETGG. (A)-(C) shows the triplicate runs of CohE-HIS-LPETGG vs. ybbR-HIS-XDocIII. The employed concentrations of ybbR-HIS-XDocIII were: 80, 160, 320 and 640 nM.

In order to minimize drift effects both sensors (see above for a detailed C-terminal SrtA protocol) were functionalized with CohE-LPETGG (0.2 µM SrtA, 20 µM CohE-LPETGG in 25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.01 % (v/v) Tween-20, 0.1 % (v/v) casein). After baseline equilibration in measurement buffer (25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Tween-20, 0.1 % (v/v) casein) a kinetic titration experiment was recorded. In contrast to prior experiments one CohE-LPETGG sensor was dipped only in measurement buffer to account for drift effects.

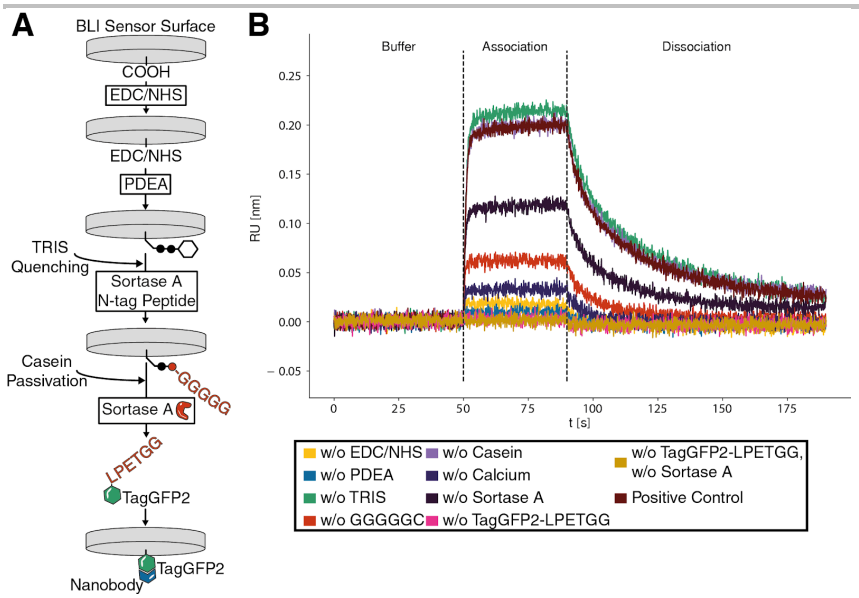


Supplemental Figure S13. Specific binding of ybbR-HIS-XDocIII to the immobilized CohE-HIS-LPETGG. (A)-(C) shows the triplicate runs of CohE-LPETGG-HIS vs. ybbR-HIS-XDocIII. The employed concentrations of ybbR-HIS-XDocIII were: 80, 160, 320 and 640 nM. The reference sensor was functionalized the same way as the measurement sensor. For the binding experiment the reference sensor was dipped only into measurement buffer to correct for drift effects.

SrtA controls were performed similar to the Sfp controls to validate the specific nature of the reaction. In an offline-immobilization experiment each of the single components were left out in one immobilization run. After equilibration in measurement buffer the functionalized sensors were dipped in to a 500 nM LaG9 solution to probe binding (**Figure S14**). Explanations for the sensorgrams for EDC/NHS, PDEA, TRIS, casein and leaving out enzyme and/or receptor protein from the reaction can be found prior **figure S6**. The reaction of skipping the GGGGG-C peptide (red trace) is similar to leaving out Coenzyme A, accessible cysteines of the TagGFP2 might react to the sensor. Leaving Calcium (purple trace) from the receptor loading reaction shows that the Sortase A is active enough to catalyze a specific binding reaction.

WILEY-VCH

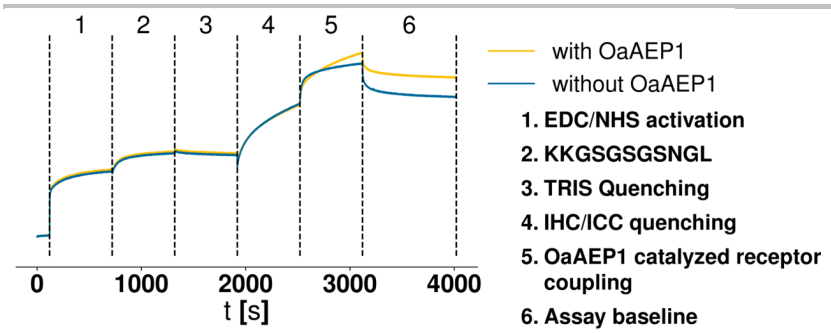
SUPPORTING INFORMATION



Supplemental Figure S14. (A) schematic of the Sortase A-based surface functionalization of TagGFP2 to the BLI-sensor. First carboxyl-groups are activated with EDC/NHS and coupled with PDEA. PDEA is thiol reactive and couples in this case a GGGGGC peptide. GGGGGC is recognized by SrtA and covalently links a LPETGG-containing TagGFP2. The sensor is now able to bind the corresponding nanobody (LaG9). **(B)** shows association and dissociation of a nanobody to the TagGFP2 functionalized sensor. Each trace represents one control where one reagent was left out in the proper surface chemistry. Dark red represents the positive control with all added substances.

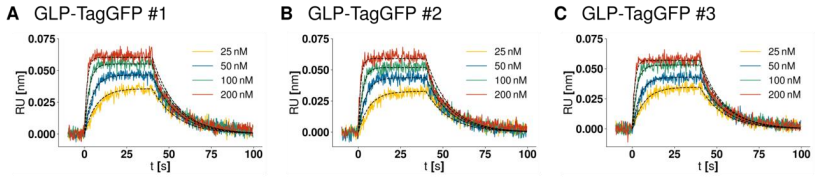
OaAEP1 N-tag-based Modification

Sensors were prepared to an amine reactive state as described under basic sensor preparation. As for the SrtA peptide a linker sequence the sequence KKGSGSGS was added to the peptide, to improve reactivity and increase the isoelectric point of the peptide. Amine reactive sensors were then coupled to KKGSGSGSNGL peptides by dipping them into a solution of 200 µM peptide in 10 mM HEPES-HCl at pH 7.2 (10 minutes). Hereafter, potentially remaining amine reactive groups were quenched with 100 mM TRIS-HCl, pH 8.5 for 5 minutes. Sensors were then passivated (OaAEP1 in 50 mM Na-Acetate, pH 5.5, 100 mM NaCl, 10 % (v/v) IHC/ICC Blocking Buffer, 0.1 % (v/v) Tween-20). For functionalization with the desired receptor (exhibiting the N-terminal OaAEP1-Tag GLP), sensors were incubated with 5 µM receptor, 0.2 µM OaAEP1 enzyme and 50 mM Na-Acetate, pH 5.5, 100 mM NaCl, 10 % (v/v) IHC/ICC Blocking Buffer, 0.1 % (v/v) Tween-20 until a desired functionalization density was reached, usually for about 5 to 10 minutes. **Figure S15** shows an exemplary sensorgram of the OaAEP1 surface immobilization of GLP-TagGFP2.



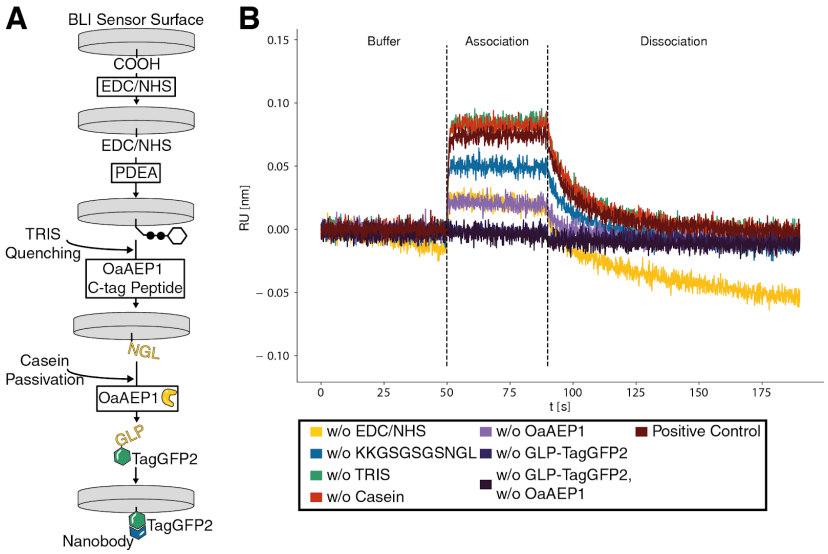
Supplemental Figure S15. Surface chemistry for N-terminal OaAEP1-based sensor modification. In order to site-specifically couple TagGFP2 some prior steps are necessary: **(1)** the carboxyl-groups are activated with EDC/NHS. **(2)** amine-groups of the KKGSGSGSNGL-peptide couples to the activated carboxyl-groups. **(3)** remaining activated carboxyl-groups are quenched with TRIS. **(4)** the BLI sensor is passivated with IHC/ICC blocking buffer to prevent any non-specific binding. **(5)** the yellow trace shows the coupling reaction of GLP-TagGFP2 with OaAEP1. The blue shows the non-specific adsorption of GLP-TagGFP2 to the sensor. **(6)** shows the baseline after enzyme-based coupling.

The TagGFP2 functionalized sensors were then dipped into wells with GLP-LaG9-HIS with different concentrations (25, 50, 100 and 200 nM) and buffer to measure dissociation of the nanobody (**Figure S16**).

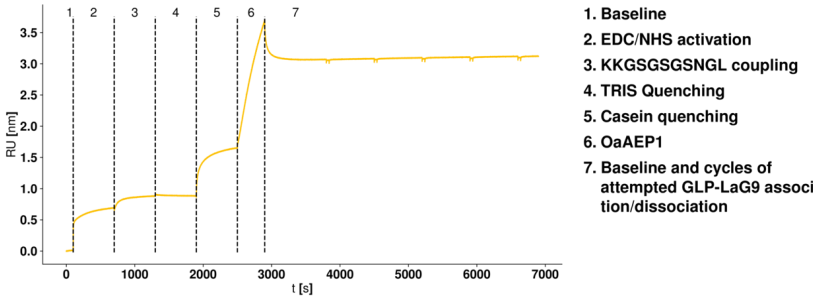


Supplemental Figure S16. Specific binding of GLP-LaG9-HIS to the immobilized GLP-TagGFP2. (A)-(C) shows the triplicate runs of GLP-LaG9-HIS vs. GLP-TagGFP2. The employed concentrations of GLP-LaG9-HIS were: 25, 50, 100 and 200 nM.

OaAEP1 controls were performed similar to the Sfp and SrtA controls to confirm the specific coupling reaction. In an offline-immobilization experiment each of the single components were left out in one immobilization run. After equilibration in measurement buffer the functionalized sensors were dipped in to a 500 nM LaG9 solution to probe binding (**Figure S17**). Explanations for the sensorgrams for EDC/NHS, PDEA, TRIS, casein and leaving enzyme and/or receptor protein from the reaction can be found prior **figure S6**. As OaAEP1 has no cofactor this reaction condition could not be probed. Besides the aforementioned controls, a further negative control was done. An OaAEP1-peptide modified sensor was dipped into an OaAEP1 containing mix. After dissociation and equilibration of the sensor in measuring buffer the sensor was moved into a LaG9 containing mix. No binding was observed at the employed concentrations (5 µM OaAEP1 and up to 0.2 µM TagGFP2), which are above the determined K_d (**Figure S18**).



Supplemental Figure S17. (A) schematic of the OaAEP1-based surface functionalization of TagGFP2 to the BLI-sensor. First carboxyl-groups are activated with EDC/NHS and coupled with an amine-containing peptide KKGSGSGSNGL. NGL is recgonized by OaAEP1 and covalently links a GLP-containing TagGFP2. The sensor is now able to bind the corresponding nanobody (LaG9). **(B)** shows association and dissociation of a nanobody to the TagGFP2 functionalized sensor. Each trace represents one control where one reagent was left out in the proper surface chemistry. Dark red represents the positive control with all added substances.

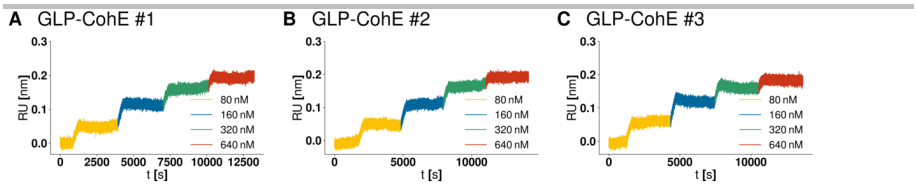


Supplemental Figure S18. Negative control for OaAEP1:GLP-Interaction. In order to control for a possible binding of the GLP-LaG9 ligand to unspecifically bound OaAEP1, a sensor was prepared as for immobilizing a GLP-tagged ligand, only that the ligand itself was ommited. In subsequent association attempts to GLP-LaG9 ligand in concentrations from 12.5 nM to 200 nM no binding could be observed.

In order to probe the N-terminal anchored CohE, an OaAEP1-based strategy was chosen (**Figure S19**). Since drifte effects should kept minimal, both sensors were functionalized with GLP-CohE (compare **Figure S13** for C-terminal anchoring). Conditions were 20 µM GLP-CohE, 0.5 µM OaAEP1 in 50 mM Na-Acetate, pH 5.5, 100 mM NaCl, 10 % (v/v) IHC/ICC Blocking Buffer (Thermo Fisher Scientific Inc.) 0.1 % (v/v) Tween-20. After baseline equilibration in measurement buffer (25 mM TRIS-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Tween-20, 0.1 % (v/v) casein) a kinetic titration experiment was recorded. The reference sensor was dipped only in measurement buffer to account for drift effects.

WILEY-VCH

SUPPORTING INFORMATION



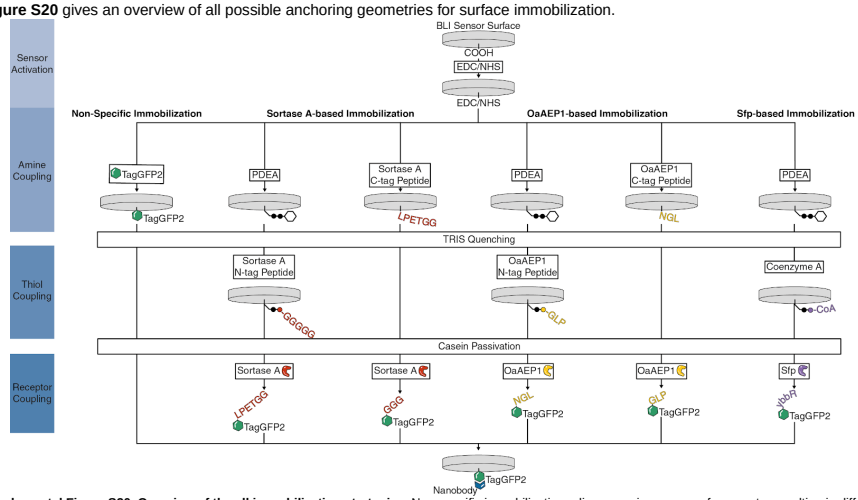
Supplemental Figure S19. Specific binding of ybbR-HIS-XDocIII to the immobilized GLP-CohE-HIS. (A)-(C) shows the triplicate runs of GLP-CohE-HIS vs. ybbR-HIS-XDocIII. The employed concentrations of ybbR-HIS-XDocIII were: 80, 160, 320 and 640 nM. The reference sensor was functionalized the same way as the measurement sensor. For the binding experiment the reference sensor was dipped only into measurement buffer to correct for drift effects.

OaAEP1 C-tag-based Modification

Sensors were prepared to a thiol reactive state by modifying them with PDEA as described under basic sensor preparation. Thiol reactive sensors were then loaded with GLPGSC (or GLPGSGSGSC) peptide by performing a disulfide exchange reaction, replacing the PDEA's leaving group with the peptide. This reaction was performed by incubating the sensors for 20 minutes in a solution of 200 µM peptide in 10 mM sodium acetate buffer at pH 4.5. Subsequently, sensors were then passivated (50 mM Na-Acetate, pH 5.5, 100 mM NaCl, 10 % (v/v) IHC/ICC Blocking Buffer, 0.1 % (v/v) Tween-20). For functionalization with the desired receptor (exhibiting the C-terminal OaAEP1-Tag NGL, sensors were incubated with 5 µM receptor, 0.5 µM OaAEP1 enzyme and 50 mM Na-Acetate, pH 5.5, 100 mM NaCl, 10 % (v/v) IHC/ICC Blocking Buffer, 0.1 % (v/v) Tween-20. Unfortunately, no functional TagGFP2 could be immobilized and no binding signal was detected.

Overview of all Immobilization Strategies

Figure S20 gives an overview of all possible anchoring geometries for surface immobilization.



Supplemental Figure S20. Overview of the all immobilization strategies. Non-specific immobilization relies on amine-groups of a receptor resulting in different orientation of the molecules on the surface. Three different site-specific, enzyme-based strategies are shown: **Left:**. SrtA links C-terminal LPETGG with N-terminal GGG. Either a KK-GSGSGS-LPETGG peptide or a GGGGG-C was reacted with the EDC/NHS sensor to any of its three N-terminal primary amines of the LPETGG-peptide or *via* PDEA in case of the GGGGG-peptide. **Middle:** OaAEP1 recognizes the C-terminal amino acids NGL and fuses it to proteins containing the N-terminal amino acids GLP. EDC/NHS activated sensors were reacted with the amine-groups of a KK-GSGS-NGL peptide or *via* PDEA for the GLP-C case. **Right:** Sfp catalyzes the reaction between ybbR-tag of a protein and Coenzyme A. In order to enable the reaction PDEA is used to crosslink EDC/NHS-activated sensors with Coenzyme A.

WILEY-VCH

SUPPORTING INFORMATION

Data Analysis

Data was read in from .frd files and then analyzed using a custom script written in Python (Python Software Foundation, Python Language Reference, version 2.7, available at http://www.python.org), utilizing the libraries NumPy, SciPy, and Matplotlib. The script is attached at the end of the supplemental information.
In brief, data from the reference sensor was subtracted from that of the functionalized sensor. Then, a global optimization of fitting parameters to the data from all concentration variations within one experiment was performed. Equations (1) and (2), describing association and dissociation, were fitted under global optimization of k_{off} and k_{on}. The parameter y_{max} describing the maximal response was fitted independently for each association/dissociation cycle, as depending on the sensors passivation, ligand may remain bound to the sensor indefinitely, thus modifying the sensors binding capacity.

(1) $y = \left(y_0 - y_{off}\right) * \exp\left(-k_{off} * \left(t - t_{0,dissociation}\right)\right)$

(2) $y = y_{max} * (1/(1 + \frac{k_{off}}{k_{on} * [c]}))) * (1 - \exp(-\left([c] * k_{on} + k_{off}\right) * \left(t - t_{0,association}\right))$

For fitting kinetic titration series, one needs to account for the ligand binding sites already occupied at the start of each association cycle. We implement this this by introducing an “offset” in time for each association cycle. Solving equation (2) for the time t yields equation (3), which allows this “time offset” t_{offset} to be calculated by inserting parameters k_{off}, k_{on} and the current concentration [c]. This t_{offset} can now be inserted into equation (4), describing the response of an association cycle in a kinetic titration series.

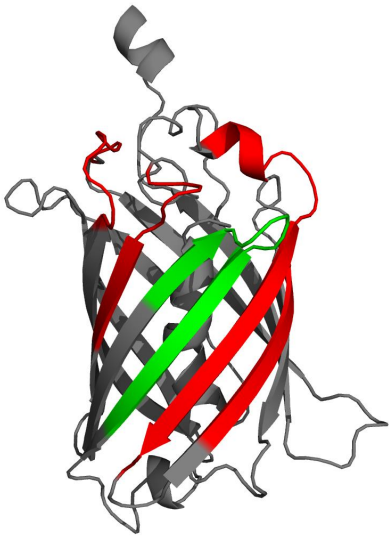
(3) $t_{offset} = -\log(1 - (y_{start}/(y_{max} * [c])) * ((k_{off}/k_{on}) + [c]))/((k_{on} * [c]) + k_{off})$

(4) $y = y_{max} * (1/(1 + k_{off}/(k_{on} * [c]))) * (1 - \exp(-\left([c] * k_{on} + k_{off}\right) * \left(t - t_{0,association}\right))$

Fitting kinetic titration series was performed using equations (1), (3) and (4) under global optimization of k_{off}, k_{on} and y_{max}.

SUPPORTING INFORMATION

```
GFP       : 100
TagGFP2   : 100
cons      : 100

GFP       SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFGYGVQCFAR
TagGFP2   SGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTTGKLPVPWPTLVTTLCYGIQCFAR

cons      * *****:*:**:*:******:****** ****** **** :******************: **:*****


GFP       YPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNY
TagGFP2   YPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEVKFEGDTLVNRIELKGKDFKEDGNILGHKLEYSF

cons      **:*** :************:***** *:***:****.**************** ***************.:


GFP       NSHNVYIMADKQKNGIKVNFKIRHNIEDGSVHLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDH
TagGFP2   NSHNVYIRPDKANNGLEANFKTRHNIEGGGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDH

cons      ******* .** :**::.*** *****.*.*:****** *.*:******:* *******: :*** ** ***


GFP       MVLLEFVTAAGITHGMDELYKGLEVLFQGPS
TagGFP2   MVLLESFSACCHTHGMDELY-----------

cons      ***** .:*.  ********
```



Supplemental Figure S21. Sequence Alignment of GFP in LaG9 study[1] and the TagGFP2 employed in this study. Top: Red and green highlight the sequence of the binding epitope of the LaG9 nanobody to the GFP. Red represents mismatching sequences from GFP to TagGFP2. Green highlights the overlapping sequence identity of both GFPs. **Bottom:** Homology model obtained with Phyre2 of TagGFP2[11,12]. Grey represents the sequence of the TagGFP2, red the mismatching part and in green the matching part of the binding epitope of LaG9 of GFP.

17

SUPPORTING INFORMATION

Protein Sequences

Enzymes

sfp-HIS
MKIYGIYMDRPLSQEENERFMTFISPEKREKCRRFYHKEDAHRTLLGDVLVRSVISRQYQLDKSDIRFS
TQEYGKPCIPDLPDAHFNISHSGRWVIGAFDSQPIGIDIEKTKPISLEIAKRFFSKTEYSDLLAKDKDEQT
DYFYHLWSMKESFIKQEGKGLSLPLDSFSVRLHQDGQVSIELPDSHSPCYIKTYEVDPGYKMAVCAAH
PDFPEDITMVSYEELLEASHHHHHH

Sortase A-HIS
MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATREQLNRGVSFAEENESLDDQNISIAGHTFIDRPNY
QFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTAVGVLDEQKGKDKQLTLITCDDYNEETGVWE
TRKIFVATEVKHHHHH

HIS-Ubiquitin-OaAEP1(C247A)
MHHHHHHGSGSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYN
IQKESTLHLVLRLRGG**ARDGDYLHLPSEVSRFFRPQETNDDHGEDSVGTRWAVLIAGSKGYANYRH**
QAGVCHAYQILKRGGLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVNAKN
FLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYADELNDALKKKHASGTYK
SLVFYLEACESGSMFEGILPEDLNIYALTSTNTTESSWAYYCPAQENPPPPEYNVCLGDLFSVAWLE
DSDVQNSWYETLNQQYHHVDKRISHASHATQYGNLKLGEEGLFVYMGSNPANDNYTSLDGNALTP
SSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAMSHRVHIDSSIKLIGKLLFGIEKCTEILNAVR
PAGQPLVDDWACLRSLVGTFETHCGSLSEYGMRHTRTIANICNAGISEEQMAEAASQACASIP

After cleavage
ARDGDYLHLPSEVSRFFRPQETNDDHGEDSVGTRWAVLIAGSKGYANYRHQAGVCHAYQILKRGG
LKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVNAKNFLAAILGNKSAITGGS
GKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYADELNDALKKKHASGTYKSLVFYLEACESGSMF
EGILPEDLNIYALTSTNTTESSWAYYCPAQENPPPPEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQ
YHHVDKRISHASHATQYGNLKLGEEGLFVYMGSNPANDNYTSLDGNALTPSSIVVNQRDAD

Recepor/Ligand

PelB-GLP-LaG9-HIS
KYLLPTAAAGLLLLAAQPAMAGLPGSGSGSADVQLVESGGGLVQAGGSLRL
SCAASGRTFSTSAMGWFRQAPGKEREFVARITWSAGYTAYSDSVKGRFTISRDKAKNTVY
LQMNSLKPEDTAVYYCASRSAGYSSSLTRREDYAYWGQGTQVTVSGSGSGSHHHHHH

ybbR-TagGFP2-HIS
MDSLEFIASKLALPESGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKF
ICTTGKLPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKY
KTRGEVKFEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFK
TRHNIEGGGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSA
CCHTHGMDELYRHHHHHH

HIS-TagGFP2-ybbR
MHHHHHHGSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTTGKL
PVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEVK
FEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIEG
GGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTHGM
DELYRDSLEFIASKLA

SUPPORTING INFORMATION

GGG-TagGFP2-HIS
MGGGSGSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTTGK
LPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEV
KFEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIE
GGGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTHG
MDELYRHHHHHH

HIS-TagGFP2-LPETGG
MHHHHHHGSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTTGKL
PVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEVK
FEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIEG
GGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTHGM
DELYRLPETGG

GLP-TagGFP2-HIS
MGLPGSGLPESGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTTG
KLPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGE
VKFEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNI
EGGGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTH
GMDELYRHHHHHH

HIS-TagGFP2-NGL
MHHHHHHGSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTTGKL
PVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEVK
FEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIEG
GGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTHGM
DELYRNGL

ybbR-HIS-XDocIII
MDSLEFIASKLAHHHHHHGVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTVYNEGYT
GDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYEIPVYYNNATLKDAEGNDATVTVYIG
LKGDTDLNNIVDGRDATATLTYYAATSTDGKDATTVALSPSTLVGGNPESVYDDFSAFLSDVKVDAGK
ELTRFAKKAERLIDGRDASSILTFYTKSSVDQYKDMAANEPNKLWDIVTGDAEEE

CohE-HIS-HRV3C-LPETGG
MALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATTGYHIYWD
ERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGADDDFGADGVMWTVELKVPADAKAGD
VYPIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAG
EPELKLPRSRHHHHHHGSLEVLFQGPLPETGG

GLP-CohE-HIS
M**GLP**GSGALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEG
KYATTGYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGADDDFGAD
GVMWTVELKVPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSS
SNPSTDEYLVKANATYADGYIAIKAGEP**HHHHHH**

<u>Peptides</u>

KKGSGSGSLPETGG
KKGSGSGSNGL
GLPGSC
GLPGSGSGSC
GGGGGC

SUPPORTING INFORMATION

Data Analysis Script

```
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
import sys, os
import re
import csv
import argparse
import StringIO
import string
import matplotlib.pylab as pylab
import xml.etree.ElementTree as ET
import struct
import glob

fontsize = 16
params = { 'font.size': fontsize,
           'legend.fontsize': fontsize,
           'axes.labelsize': fontsize,
           'axes.titlesize': fontsize,
           'xtick.labelsize': fontsize,
           'ytick.labelsize': fontsize,
           'axes.labelpad':     3.0,
           'text.usetex': True,
       'text.latex.preamble': [
            r"\usepackage{helvet}",
            r"\usepackage{sansmath}",
            r"\renewcommand\familydefault{\sfdefault}"
            r"\usepackage{sfmath}"
            r"\usepackage[detect-all]{siunitx}",
            r"\DeclareSIUnit\Molar{M}",
            r"\sisetup{mode = math, detect-family, detect-weight, exponent-product =
\cdot, math-rm=\mathsf, text-rm=\sffamily}",
            r"\boldmath"
            ]
           }
pylab.rcParams.update(params)
mpl.rc("figure", figsize=(12, 6))

lw=3
papercolors = [\
(0.988, 0.752, 0.015), #Yellow
(0, 0.403, 0.603), #Blue
(0.192, 0.6, 0.4),  #Green
(0.8, 0.207, 0.090),  #Bright Red
(0.529, 0.403, 0.666), #Bright Purple
(0.196, 0.141, 0.368), #Medium Purple
(0.180, 0.074, 0.196), #Dark Purple
(0.396, 0.078, 0.054), #Dark Red
(0.909, 0.192, 0.541), #Pink
(0.796, 0.6, 0.015), #Brown
(0.450, 0.760, 0.686), #Dark Teal
(0.603, 0.803, 0.803)]  #Bright Teal

##############################################################################
```

SUPPORTING INFORMATION

```
def string2float(value):
    value = value.replace(',','.')
    return float(value)

def plotadd(axis, dd, cycleparm, fit_global, i):
    sensor = cycleparm['sensor']
    coefoff = cycleparm['offrate_parms']
    if fit_global == True:
        koff = cycleparm['offrate_parms_global'][0]
        coefon = cycleparm['onrate_parms_global']
    else:
        koff = coefoff[0]
        coefon = cycleparm['onrate_parms']
    conc = cycleparm['conc']
    offratefit_offset = cycleparm['offratefit_offset']

    baseline_duration = 10# in seconds
    baseline_length = (baseline_duration/\
    (dd[sensor[0]][1]-dd[sensor[0]][0])).astype(int)
    start = cycleparm['assoc_start']-baseline_length
    concstring = '{0:.2e}'.format(conc)
    axis.plot(dd[sensor[0]][start:\
    cycleparm['dissoc_end']] - cycleparm['t0_assoc'],
    dd[sensor[0]+"_resp"][start:cycleparm['dissoc_end']],
    '-', color=papercolors[i], markevery=1,
    label=r'$\SI{{{}}}{{\nano\Molar}}$'.format(int(conc*1e9)),
    zorder=0)
####plot offrate####
    if offratefit_offset == True:
        plt.plot(dd[sensor[0]][cycleparm['dissoc_start']:\
    cycleparm['dissoc_end']] - cycleparm['t0_assoc'],
    offrate(dd[sensor[0]][cycleparm['dissoc_start']:\
    cycleparm['dissoc_end']],koff,
    cycleparm['y0_dissoc'], cycleparm['t0_dissoc'], coefoff[1]),linestyle="--",
    color="k", zorder=1)#, label='koff= {0:.2e}'.format(coefoff[0]))
    elif offratefit_offset == False:
        plt.plot(dd[sensor[0]][cycleparm['dissoc_start']:\
    cycleparm['dissoc_end']] - cycleparm['t0_assoc'],
    offrate(dd[sensor[0]][cycleparm['dissoc_start']:\
    cycleparm['dissoc_end']], koff, cycleparm['y0_dissoc'],
    cycleparm['t0_dissoc'], offset=0),linestyle="--",color="k",
    zorder=1)
    #, label='koff= {0:.2e}'.format(coefoff[0]))
####plot onrate####
    if cycleparm['fit_onrate'] == True:
        if coefon[1:].size > 1:#ymax unlinked
            plt.plot(dd[sensor[0]][cycleparm['assoc_start']:\
    cycleparm['assoc_end']] - cycleparm['t0_assoc'],
    onrate(dd[sensor[0]][cycleparm['assoc_start']:\
    cycleparm['assoc_end']], coefon[0], koff,
    coefon[1+i], conc, cycleparm['t0_assoc']),linestyle="--",
    color="k", zorder=1)
        else:
            plt.plot(dd[sensor[0]][cycleparm['assoc_start']:\
    cycleparm['assoc_end']] - cycleparm['t0_assoc'],
    onrate(dd[sensor[0]][cycleparm['assoc_start']:\
    cycleparm['assoc_end']], coefon[0], koff,
    coefon[1], conc, cycleparm['t0_assoc']),linestyle="--",
```

```
    color="k", zorder=1)
    #,label='kon= {0:.2e}'.format(coefon[0]))

def plotadd_kinetic(axis, dd, cycleparm, fit_global, i):
    sensor = cycleparm['sensor']
    coefoff = cycleparm['offrate_parms']
    if fit_global == True:
        koff = cycleparm['offrate_parms_global'][0]
        coefon = cycleparm['onrate_parms_global']
    else:
        koff = coefoff[0]
        coefon = cycleparm['onrate_parms']
    conc = cycleparm['conc']
    offratefit_offset = cycleparm['offratefit_offset']
    plt.plot(dd[sensor[0]][cycleparm['assoc_start']:\
    cycleparm['assoc_end']], onrate_kinetic(dd[sensor[0]]\
    [cycleparm['assoc_start']:cycleparm['assoc_end']],coefon[0],
    koff, coefon[1], cycleparm['ystart'], conc, cycleparm['t0_assoc']),
    linestyle="--",color="k")
    #,label='kon= {0:.2e}'.format(coefon[0]))

##offset in kinetic titration series doesn't make sense
#    if offratefit_offset == True:
#        plt.plot(dd[sensor[0]][cycleparm['dissoc_start']:\
#    cycleparm['dissoc_end']], offrate(dd[sensor[0]]\
#    [cycleparm['dissoc_start']:cycleparm['dissoc_end']],
#    koff, cycleparm['y0_dissoc'], cycleparm['t0_dissoc'], coefoff[1]),
#    linestyle="--",color="k")
#    #, label='koff= {0:.2e}'.format(coefoff[0]))
    if offratefit_offset == False:
        #for i in range(1, len(offrates)+1):
        plt.plot(dd[sensor[0]][cycleparm['dissoc_start']:\
    cycleparm['dissoc_end']], offrate(dd[sensor[0]]\
    [cycleparm['dissoc_start']:cycleparm['dissoc_end']],
    koff, cycleparm['y0_dissoc'], cycleparm['t0_dissoc'], offset=0),
    linestyle="--",color="k")
    #, label='koff= {0:.2e}'.format(coefoff[0]))

def plotadd_data(axis, dd, cycleparm, i):
    sensor = cycleparm['sensor']
    conc = cycleparm['conc']
    baseline_duration = 60# in seconds
    baseline_length = (baseline_duration/\
    (dd[sensor[0]][1]-dd[sensor[0]][0])).astype(int)
    if i == 0:
        start = cycleparm['baseline_start']
    else:
        start = cycleparm['assoc_start']
    axis.plot(dd[sensor[0]][start:\
    cycleparm['dissoc_end']], dd[sensor[0]+"_resp"]\
    [start:cycleparm['dissoc_end']],
    '-', color=papercolors[i], markevery=1, zorder=0,
    label=r'$\SI{{{}}}{{\nano\Molar}}$'.format(int(conc*1e9)))
    #label='{0:.2e}M'.format(conc))
#    axis.plot(time, response, color='black')
def plotrates(axis, onrate, onrate_error, offrate, offrate_error):
    kd = offrate/onrate
    onstring = "$k_{{on}}: \SI{{{0:.2e}}}{{\per\mole\per\second}} \pm
```

SUPPORTING INFORMATION

```
\SI{{{1:.2e}}}{{}}$".format(onrate,onrate_error)
    offstring =  "$k_{{off}}:
\SI{{{0:.2e}}}{{\per\second}}\pm\SI{{{1:.2e}}}{{}}$".format(offrate, offrate_error)
    kdstring =  "$K_d: \SI{{{}}}{{\per\\nano\Molar}}$".format(int(kd*1e9))
    axis.annotate(offstring+'\n'+onstring+'\n'+kdstring, xy=(0.5, 0.1),
    xycoords='axes fraction')

def plotformat(axis):
    axis.legend(loc='upper left', shadow=False, frameon=False)
    despine(axis, top=1, bottom=0, left=0,right=1,xaxis=0,yaxis=0)
#   plt.axvline()
    axis.set_xlabel('t $[\si{\second}]$')
    axis.set_ylabel('RU $[\si{\\nano\metre}]$')

def despine(axis,top=1,bottom=1,left=1,right=1,xaxis=1,yaxis=1):
    if xaxis:
         axis.xaxis.set_visible(False)
    if yaxis:
        axis.yaxis.set_visible(False)
    if top:
        axis.spines["top"].set_visible(False)
    if bottom:
        axis.spines["bottom"].set_visible(False)
    if left:
        axis.spines["left"].set_visible(False)
    if right:
        axis.spines["right"].set_visible(False)

def extract_steps(folder):
    assaydict={}
    datadict={}
    files = glob.glob(folder +'/*_*.frd')
#    files = [os.path.basename(x) for x in glob.glob(folder +'/*_*.frd')]
    for f in files:
        tree = ET.parse(f)
        root = tree.getroot()
        sensorcoord = root.find('./ExperimentInfo/SensorName').text
        assaydict[sensorcoord] = []
        xdatas = []
        ydatas = []
        for entry in root.iter('AssayXData'):
            length = int(entry.attrib['Points'])
            xdatas.append(unpack_data(entry.text))
            if length != len(xdatas[-1]):
                print "x data length not as expected"
                return 1
            assaydict[sensorcoord].append(length)
        for entry in root.iter('AssayYData'):
            length = int(entry.attrib['Points'])
            ydatas.append(unpack_data(entry.text))
            if length != len(ydatas[-1]):
                print "y data length not as expected"
                return 1
        datadict[sensorcoord] = np.hstack(xdatas)
        datadict[sensorcoord + '_resp'] = np.hstack(ydatas)
    return assaydict, datadict

def unpack_data(data):
```

23

```
    size = 4
    decoded = data.decode('base64')
    length = len(decoded)/4
    return struct.unpack('<{}f'.format(length), decoded)

def onrate_equ(t,kon, koff ,yequ, conc, t0_assoc):
    y = yequ*(1-np.exp(-((conc*kon)+koff)*(t-t0_assoc)))
    return y

def yequ_to_ymax(yequ, koff, kon, conc):
    return yequ/(conc/(conc+koff/kon))

def kinetic_timeoffset(ystart, kon, koff, maxresp, conc):
    timeoffset = (1-(ystart/(maxresp*conc))*((koff/kon)+conc))
    timeoffset = -np.log(timeoffset)/((kon*conc)+koff)
    return timeoffset

def onrate_back(t,kon, koff ,ymax, conc, t0_assoc):
    y = ymax*(conc/(conc+koff/kon))*(1-np.exp(-((conc*kon)+koff)*(t-t0_assoc)))
    return y


def onrate(t, kon, koff, ymax, conc, t0_assoc):
    y = ymax*(1/(1+koff/(kon*conc)))*(1-np.exp(-((conc*kon)+koff)*(t-t0_assoc)))
    return y

def onrate_kineticequ(t,kon, koff ,yequ, ystart, conc, t0_assoc):
    y = yequ*(1-np.exp(-((conc*kon)+koff)*(t-t0_assoc)))+ystart
    return y

def onrate_kinetic(t, kon, koff ,ymax, ystart, conc, t0_assoc):
    t_offset = kinetic_timeoffset(ystart, kon, koff, ymax, conc)
    y = ymax*(1/(1+koff/(kon*conc)))*(1-np.exp(-((conc*kon)+koff)\
    *(t-t0_assoc+t_offset)))
    return y

def offrate(t, koff, y0, t0, offset):
    y = (y0-offset)*np.exp(-koff*(t-t0)) + offset
    return y

def combined_global(t, *args):
    concs = args[0]
    t0s_on = args[1]
    t0s_off = args[2]
    t_on_idx = args[3]
    t_off_idx = args[4]
    ystarts = args[5]
    offsets = args[6]
    fitparms = args[7]
    kon_global = fitparms[0]
    ymaxs = fitparms[1:-1]
    koff_global = fitparms[-1]
    calc = []
    start = 0
    for i in range(0, len(concs)):
        if len(ystarts) > 0:#kinetic fit
            if i == 0:
                ystart = 0
```

```
            else:
                ystart = calc[-1][-1]
            calc.append(onrate_kinetic(t[start:(start+t_on_idx[i])],
    kon_global, koff_global, ymaxs[0], ystart, concs[i], t0s_on[i]))
        else:
            if len(ymaxs) > 1:
                calc.append(onrate(t[start:(start+t_on_idx[i])],
    kon_global, koff_global, ymaxs[i], concs[i], t0s_on[i]))
            else:
                calc.append(onrate(t[start:(start+t_on_idx[i])],
    kon_global, koff_global, ymaxs[0], concs[i], t0s_on[i]))
        start += t_on_idx[i]
        calc.append(offrate(t[start:(start+t_off_idx[i])],
    koff_global, calc[-1][-1], t0s_off[i], offsets[i]))
        start += t_off_idx[i]
    return np.hstack(calc)

def offrate_global(t, koff, *args):
    y0s = args[0]
    t0s = args[1]
    offsets = args[2]
    t_idx = args[3]
    offrate_calc = []
    start = 0
    for i, idx in enumerate(t_idx):
        offrate_calc.append(offrate(t[start:(start+idx)],
    koff, y0s[i], t0s[i], offsets[i]))
        start += idx
    return np.hstack(offrate_calc)

def onrate_global(t, *args):
    koff = args[0]
    concs = args[1]
    t0s = args[2]
    t_idx = args[3]
    ystarts = args[4]
    fitparms = args[5]
    kon_global = fitparms[0]
    ymaxs = fitparms[1:]
    onrate_calc = []
    start = 0
    for i, idx in enumerate(t_idx):
        if len(ystarts) > 0:#kinetic fit
            onrate_calc.append(onrate_kinetic(t[start:(start+idx)],
    kon_global, koff, ymaxs[0], ystarts[i], concs[i], t0s[i]))
        else:
            if len(ymaxs) > 1:
                onrate_calc.append(onrate(t[start:(start+idx)],
    kon_global, koff, ymaxs[i], concs[i], t0s[i]))
            else:
                onrate_calc.append(onrate(t[start:(start+idx)],
    kon_global, koff, ymaxs[0], concs[i], t0s[i]))
        start += idx
    return np.hstack(onrate_calc)

def displaysensors(experiment, sensor, dd, assaydict, displaysteps):
    fig, axis = plt.subplots()
    steptimes = []
```

```
    stepidx = []
    steps = -1
    minresp = np.min(np.hstack((dd[sensor[0]+"_resp"][:],
    dd[sensor[1]+"_resp"][:])))
    maxresp = np.max(np.hstack((dd[sensor[0]+"_resp"][:],
    dd[sensor[1]+"_resp"][:])))
    for  step in assaydict[sensor[0]]:
        steps += step
        stepidx.append(step)
        steptimes.append(dd[sensor[0]][steps])
    endidx = sum(stepidx[:displaysteps])
    axis.plot(dd[sensor[0]][:endidx], dd[sensor[0]+"_resp"][:endidx],
    '-', color=papercolors[0], markevery=1, zorder=1,
    label='with OaAEP1')
    axis.plot(dd[sensor[1]][:endidx], dd[sensor[1]+"_resp"][:endidx],
    '-', color=papercolors[1], markevery=1, zorder=1,
    label='without OaAEP1')
    plt.vlines(x=steptimes[:displaysteps], ymin=minresp-0.2, ymax=maxresp+0.2,
    color = 'black', linestyles='dashed')
#    for i in range(0, len(steptimes) -1):
    for i in range(0, displaysteps -1):
        textcoord = steptimes[i] + (steptimes[i+1] - steptimes[i])/2
        axis.text(textcoord, maxresp+0.2, "%d" % (i+1), rotation=0,
    horizontalalignment='center')
    despine(axis, top=1, bottom=0, left=1,right=1,xaxis=0,yaxis=1)
    axis.set_xlabel('t $[\SI{}{\second}]$')
    axis.set_ylabel('RU $[\SI{}{\\nano\metre}]$')
    plt.legend(bbox_to_anchor=(0.98,1), loc="upper left", frameon=False)

    plt.savefig(experiment + ".pdf", bbox_inches="tight")
    plt.show()
#    label='{0:.2e}M'.format(conc))
#    axis.plot(time, response, color='black')

def analyze_experiment(experiment, sensors, fit_kinetic=False, fit_onrate=True,
    reference=True, offratefit_offset = False, fit_global=False, uslocale=True,
    singlefit=False, skipsteps=0, fit_y0=False, ymax_linked=True,
    invert_sensors=False, koff_initial=1e-3, kon_initial=1e3,
    offset_initial=0.1, displayonly=False, displaysteps = 7,
    fit_stepwise = True):
    print(experiment.center(80, '*'))
    assaydict, dd = extract_steps(experiment)
    sensorresults = []

    for sensor in sensors:
        if displayonly == True:
            displaysensors(experiment, sensor, dd, assaydict, displaysteps)
        else:
            fig, axis = plt.subplots()
            sensorresults.append(fit_sensor(sensor, dd, assaydict,
    fit_kinetic, fit_onrate, reference, offratefit_offset, uslocale,
    singlefit, skipsteps, fit_y0, koff_initial, kon_initial, offset_initial,
    invert_sensors))
    if len(sensors) == 1 and fit_global == True:
        if fit_stepwise == True:
            koff_global_sol, koff_global_stderr = fit_global_off_onesensor(dd,
    sensors[0], sensorresults)
            kon_global_sol, kon_global_stderr = fit_global_on(dd, sensors[0],
```

```
    sensorresults, koff_global_sol[0], fit_kinetic, ymax_linked)
        else:
            kon_global_sol, kon_global_stderr, koff_global_sol,\
    koff_global_stderr = fit_global_combined(dd, sensors[0], sensorresults,
    fit_kinetic, ymax_linked)

        for j, resultdict in enumerate(sensorresults[0]):
            resultdict['offrate_parms_global'] = koff_global_sol
            resultdict['offrate_stderr_global'] = koff_global_stderr
            resultdict['onrate_parms_global'] = kon_global_sol
            resultdict['onrate_stderr_global'] = kon_global_stderr
        print(' global fitting '.center(80, '*'))
        print "ymaxtest", kon_global_sol[1:]
        print "global k_on: {0:.2e}, stdev: {1:.2e}".format(kon_global_sol[0],
        kon_global_stderr[0])
        print koff_global_sol, "test"
        print "global k_off: {0:.2e}, stdev: {1:.2e}".format(koff_global_sol[0],
        koff_global_stderr[0])
        print "global K_d: {0:.2e}".format(koff_global_sol[0]/kon_global_sol[0])
    if len(sensors) == 1 and displayonly == False:
        for j, resultdict in enumerate(sensorresults[0]):
            if fit_kinetic == True:
                plotadd_kinetic(axis, dd, resultdict, fit_global, j)
                plotadd_data(axis, dd, resultdict, j)
            else:
                plotadd(axis, dd, resultdict, fit_global, j)



    if len(sensors) > 1 and fit_kinetic == False and fit_global == False:
        for j, results in enumerate(sensorresults):
            for resultdict in results:
                plotadd(axis, dd, resultdict, fit_global, j)
    if singlefit == 1:
        print "handle me"
    if displayonly == False:
        plotformat(axis)
        ticks = [tick for tick in fig.gca().get_yticks() if tick >=0]
        fig.gca().set_yticks(ticks)
        plt.savefig(experiment + ".pdf", bbox_inches="tight")
    #plotrates(axis, kon_global_sol[0], kon_global_stderr[0],
    #koff_global_sol[0], koff_global_stderr[0])
    plt.show()
    return

def fit_global_off_onesensor(dd, sensor, sensorresults):
    globaloff_t = []
    globaloff_resp = []
    koffs_local = []
    y0s = []
    t0s = []
    offsets = []
    t_idx = []
    for i, resultdict in enumerate(sensorresults[0]):
        globaloff_t.append(dd[sensor[0]]\
    [resultdict['dissoc_start']:resultdict['dissoc_end']])
        globaloff_resp.append(dd[sensor[0]+"_resp"]\
    [resultdict['dissoc_start']:resultdict['dissoc_end']])
```

```
        koffs_local.append(resultdict['offrate_parms'][0])
        y0s.append(resultdict['y0_dissoc'])
        t0s.append(resultdict['t0_dissoc'])
        offsets.append(resultdict['offset_dissoc'])
        t_idx.append(globaloff_t[i].size)
    globaloff_resp_arr = np.hstack(globaloff_resp)
    globaloff_t_arr = np.hstack(globaloff_t)
    koff_med = np.median(np.asarray(koffs_local))

    coefoff,covaroff = curve_fit(lambda t, koff: \
    offrate_global(globaloff_t_arr, koff, y0s, t0s, offsets, t_idx),
    globaloff_t_arr, globaloff_resp_arr,
    p0=(1e-3), bounds=((0), (np.inf)), method='trf')
    return coefoff, np.sqrt(np.diag(covaroff))

def fit_global_combined(dd, sensor, sensorresults, fit_kinetic, ymax_linked):
    global_t = []
    global_resp = []
    kons_local = []
    koffs_local = []
    t0s_on = []
    t0s_off = []
    maxresponses = []
    concs = []
    t_on_idx = []
    t_off_idx = []
    ystarts = []
    offsets = []
    for i, resultdict in enumerate(sensorresults[0]):
        global_t.append(dd[sensor[0]]\
    [resultdict['assoc_start']:resultdict['assoc_end']])
        global_resp.append(dd[sensor[0]+"_resp"]\
    [resultdict['assoc_start']:resultdict['assoc_end']])

        kons_local.append(resultdict['onrate_parms'][0])
        maxresponses.append(resultdict['onrate_parms'][1])
        concs.append(resultdict['conc'])
        t0s_on.append(resultdict['t0_assoc'])
        t_on_idx.append(global_t[-1].size)

        global_t.append(dd[sensor[0]]\
    [resultdict['dissoc_start']:resultdict['dissoc_end']])
        global_resp.append(dd[sensor[0]+"_resp"]\
    [resultdict['dissoc_start']:resultdict['dissoc_end']])
        koffs_local.append(resultdict['offrate_parms'][0])
        t0s_off.append(resultdict['t0_dissoc'])
        offsets.append(resultdict['offset_dissoc'])
        t_off_idx.append(global_t[-1].size)
        if fit_kinetic == True:
            ystarts.append(resultdict['ystart'])
    global_resp_arr = np.hstack(global_resp)
    global_t_arr = np.hstack(global_t)
    koff_med = np.median(np.asarray(koffs_local))
    kon_med = np.median(np.asarray(kons_local))
    maxresponse_med = np.median(np.asarray(maxresponses))
    if ymax_linked == True:
        fitparams_start = (kon_med, maxresponse_med, koff_med)
        no_of_onparms = 2
```

```
    else:
        fitparams_start = (kon_med,) + tuple(maxresponses) + (koff_med,)
        no_of_onparms = 1+len(maxresponses)
    fitparams = np.empty_like(fitparams_start)
    coef,covar = curve_fit(lambda t, *fitparams: \
    combined_global(global_t_arr, concs, t0s_on, t0s_off, t_on_idx, t_off_idx,
    ystarts, offsets, fitparams), global_t_arr, global_resp_arr,
    p0=fitparams_start,# bounds=((0, 0), (np.inf, np.inf)),
    method='trf')
    ystart = 0
    for i, resultdict in enumerate(sensorresults[0]):
        if fit_kinetic == False:
            if ymax_linked == True:
                y0_dissoc = onrate(t0s_off[i], coef[0], coef[no_of_onparms],
    coef[1], concs[i], t0s_on[i])
            else:
                y0_dissoc = onrate(t0s_off[i], coef[0], coef[no_of_onparms],
    coef[1+i], concs[i], t0s_on[i])
        else:
            resultdict['ystart'] =  ystart
            if ymax_linked == True:
                y0_dissoc = onrate_kinetic(resultdict['tend_assoc'], coef[0],
    coef[no_of_onparms], coef[1], ystart,  concs[i], t0s_on[i])
                ystart = offrate(resultdict['tend_dissoc'],
    coef[no_of_onparms], y0_dissoc, t0s_off[i], offsets[i])
        resultdict['y0_dissoc'] =  y0_dissoc
    return (coef[:no_of_onparms], np.sqrt(np.diag(covar))[:no_of_onparms],
    coef[no_of_onparms:], np.sqrt(np.diag(covar))[no_of_onparms:])

def fit_global_on(dd, sensor, sensorresults, koff, fit_kinetic, ymax_linked):
    globalon_t = []
    globalon_resp = []
    kons_local = []
    maxresponses = []
    concs = []
    t0s = []
    t_idx = []
    ystarts = []
    for i, resultdict in enumerate(sensorresults[0]):
        globalon_t.append(dd[sensor[0]]\
    [resultdict['assoc_start']:resultdict['assoc_end']])
        globalon_resp.append(dd[sensor[0]+"_resp"]\
    [resultdict['assoc_start']:resultdict['assoc_end']])

        kons_local.append(resultdict['onrate_parms'][0])
        maxresponses.append(resultdict['onrate_parms'][1])
        concs.append(resultdict['conc'])
        t0s.append(resultdict['t0_assoc'])
        t_idx.append(globalon_t[i].size)
        if fit_kinetic == True:
            ystarts.append(resultdict['ystart'])
    globalon_resp_arr = np.hstack(globalon_resp)
    globalon_t_arr = np.hstack(globalon_t)
    kon_med = np.median(np.asarray(kons_local))
    maxresponse_med = np.median(np.asarray(maxresponses))
    if ymax_linked == True:
#        fitparams_start = np.array([kon_med, maxresponse_med])
        fitparams_start = (kon_med, maxresponse_med)
```

```
    else:
        fitparams_start = (kon_med,) + tuple(maxresponses)
        #fitparams_start = np.copy(maxresponses)
        #fitparams_start = np.insert(fitparams_start,0,kon_med)
    fitparams = np.empty_like(fitparams_start)
    coefon,covaron = curve_fit(lambda t, *fitparams: \
    onrate_global(globalon_t_arr, koff, concs,
    t0s, t_idx, ystarts, fitparams), globalon_t_arr, globalon_resp_arr,
    p0=fitparams_start,# bounds=((0, 0), (np.inf, np.inf)),
    method='trf')
    return coefon, np.sqrt(np.diag(covaron))


def fit_sensor(sensor, dd, assaydict, fit_kinetic, fit_onrate,
        reference, offratefit_offset, uslocale, singlefit, skipsteps, fit_y0,
        koff_initial, kon_initial, offset_initial, invert_sensors):
    print(' local fitting '.center(80, '*'))
    if reference == True:
        if invert_sensors == False:
            dd[sensor[0]+"_resp"] = dd[sensor[0]+"_resp"] - \
    dd[sensor[1]+"_resp"]
        else:
            dd[sensor[0]+"_resp"] = dd[sensor[1]+"_resp"] - \
    dd[sensor[0]+"_resp"]
    no_of_concentrations = len(sensor[2:])
    if singlefit == 1:
        no_of_measurements = 1
    elif fit_kinetic==True:
        no_of_measurements=(len(assaydict[sensor[0]])-1-skipsteps)/2
        #-1 for baseline
    else:
        no_of_measurements=(len(assaydict[sensor[0]])-skipsteps)/3
    #print "no_of_meas vs concs", no_of_measurements, no_of_concentrations
    if no_of_concentrations > no_of_measurements:
        print "more concentrations given than measurement cycles found in log"

    cycleparm = []
    testcounter = 0
    for i in range(0, no_of_measurements):
        conc = sensor[i+2]
        if singlefit == 0 and fit_kinetic == False:
            assoc_start =\
    int(np.sum(assaydict[sensor[0]][:(skipsteps+1+(i*3))]))
            baseline_start =\
    assoc_start - int(assaydict[sensor[0]][skipsteps+(i*3)])
            t0_baseline = dd[sensor[0]][baseline_start]
            dissoc_start =\
    int(np.sum(assaydict[sensor[0]][:(skipsteps+2+(i*3))]))
            dissoc_end = int(np.sum(assaydict[sensor[0]]\
    [:(skipsteps+3+(i*3))])) - 1
        elif fit_kinetic == True:
            assoc_start =\
    int(np.sum(assaydict[sensor[0]][:(skipsteps+1+(i*2))]))
            baseline_start =\
    assoc_start - int(assaydict[sensor[0]][skipsteps+(i*2)])
            t0_baseline = dd[sensor[0]][baseline_start]
            dissoc_start =\
    int(np.sum(assaydict[sensor[0]][:(skipsteps+2+(i*2))]))
```

```
            dissoc_end = int(np.sum(assaydict[sensor[0]]\
[:(skipsteps+3+(i*2))])) - 1
        elif singlefit == 1:
            baseline_start = int(np.sum(assaydict[sensor[0]][skipsteps]))
            t0_baseline = dd[sensor[0]][baseline_start]
            assoc_start = int(np.sum(assaydict[sensor[0]][:skipsteps+1]))
            dissoc_start = int(np.sum(assaydict[sensor[0]][:skipsteps+2]))
            dissoc_end = int(np.sum(assaydict[sensor[0]][:skipsteps+3])) - 1
        baseline_end = assoc_start - 1
        assoc_end = dissoc_start - 1

    if fit_kinetic == True:
            if i == 0:
                baseline_y = (np.mean(dd[sensor[0]+"_resp"]\
[(baseline_end-20):baseline_end]))
                dd[sensor[0]+"_resp"] = dd[sensor[0]+"_resp"] - baseline_y
                ystart = 0
            else:
                ystart = cycleparm[i-1]['dissoc_end_y']
        elif fit_kinetic == False:
            baseline_y = (np.mean(dd[sensor[0]+"_resp"]\
[(baseline_end-20):baseline_end]))
            dd[sensor[0]+"_resp"][baseline_start:dissoc_end] -= baseline_y
        t0_dissoc = dd[sensor[0]][dissoc_start]
        t0_assoc = dd[sensor[0]][assoc_start]
        tend_dissoc = dd[sensor[0]][dissoc_end]
        tend_assoc = dd[sensor[0]][assoc_end]
        y0_dissoc_avg =np.mean(dd[sensor[0]+"_resp"]\
[(dissoc_start-3):(dissoc_start+3)])
        ####fit offrate####
        if offratefit_offset == True:
            coefoff,covaroff = curve_fit(lambda t, koff, offset: \
offrate(t, koff, y0_dissoc_avg, t0_dissoc, offset),
dd[sensor[0]][dissoc_start:dissoc_end],
dd[sensor[0]+"_resp"][dissoc_start:dissoc_end],
p0=(koff_initial, offset_initial), bounds=((0, 0),
(np.inf, np.inf)), method='trf')
            offset = coefoff[1]
            y0_dissoc = y0_dissoc_avg
        elif offratefit_offset == False and fit_y0 == True:
            coefoff,covaroff = curve_fit(lambda t, koff, y0_dissoc_opt: \
offrate(t, koff,  y0_dissoc_opt, t0_dissoc, offset = 0),
dd[sensor[0]][dissoc_start:dissoc_end],
dd[sensor[0]+"_resp"][dissoc_start:dissoc_end],
p0=(koff_initial, y0_dissoc_avg), bounds=((0, 0), (np.inf, np.inf)),
method='trf', maxfev=1000000)
            offset = 0
            y0_dissoc = (coefoff[1])
            print "y0_after", coefoff[1]
        elif offratefit_offset == False and fit_y0 == False:
            coefoff,covaroff = curve_fit(lambda t, koff: \
offrate(t, koff,  y0_dissoc_avg, t0_dissoc, offset = 0),
dd[sensor[0]][dissoc_start:dissoc_end],
dd[sensor[0]+"_resp"][dissoc_start:dissoc_end],
p0=(koff_initial), bounds=((0), (np.inf)), method='trf', maxfev=1000000)
            y0_dissoc = y0_dissoc_avg
            offset = 0
        #offrate_errs = (np.sqrt(np.diag(covaroff)))
```

SUPPORTING INFORMATION

```
#    if offratefit_offset == True:
#        print "offset =  {0:.2e} +/- {1:2e}".format(coefoff[1],
#offrates_err[2])
    dissoc_end_y =\
offrate(dd[sensor[0]][dissoc_start:dissoc_end], coefoff[0],
y0_dissoc, t0_dissoc, offset = 0)[-1]
    if fit_onrate == True:
        if fit_kinetic == True:
            coefon, covaron = curve_fit(lambda t, kon, maxresponse: \
onrate_kinetic(t, kon, coefoff[0], maxresponse, ystart, conc, t0_assoc),
dd[sensor[0]][assoc_start:assoc_end],
dd[sensor[0]+"_resp"][assoc_start:assoc_end],
p0=(kon_initial, np.amax(dd[sensor[0]+"_resp"][baseline_end:assoc_end])),
bounds=((0, 0), (np.inf, np.inf)), method='trf', maxfev=100000)
        else:
            coefon, covaron = curve_fit(lambda t, kon, maxresponse: \
onrate(t, kon, coefoff[0], maxresponse, conc, t0_assoc),
dd[sensor[0]][assoc_start:assoc_end],
dd[sensor[0]+"_resp"][assoc_start:assoc_end],
p0=(kon_initial, np.amax(dd[sensor[0]+"_resp"][baseline_end:assoc_end])),
bounds=((0, 0), (np.inf, np.inf)), method='trf', maxfev=100000)
#               coefon[1] = yequ_to_ymax(coefon[1], coefoff[0], coefon[0], conc)
        kon_stderrs = np.sqrt(np.diag(covaron))
        koff_stderrs = np.sqrt(np.diag(covaroff))
        print "concentration: ", conc
        print "local k_on = {0:.2e}, stdev: {1:.2e}".format(coefon[0],
kon_stderrs[0])
        print "local ymax = {0:.2f}, stdev: {1:.2e}".format(coefon[1],
kon_stderrs[1])
        print "local k_off = {0:.2e}, stdev: {1:.2e}".format(coefoff[0],
koff_stderrs[0])
        print "local K_d: {0:.2e}".format(coefoff[0]/coefon[0])
        print('''.center(80, '-'))
    cycleparm.append(dict([
    ('t0_baseline', t0_baseline),
('t0_dissoc', t0_dissoc),
('t0_assoc', t0_assoc),
('tend_dissoc', tend_dissoc),
('tend_assoc', tend_assoc),
('baseline_start', baseline_start),
('assoc_start', assoc_start),
('dissoc_start', dissoc_start),
('baseline_end', baseline_end),
('baseline_y', baseline_y),
('y0_dissoc', y0_dissoc),
('dissoc_end_y', dissoc_end_y),
('assoc_end', assoc_end),
('offrate_parms', coefoff),
('offrate_covar', covaroff),
('onrate_parms', coefon),
('onrate_covar', covaron),
('dissoc_end', dissoc_end),
('conc', conc),
('sensor', sensor),
('offratefit_offset', offratefit_offset),
('offset_dissoc', offset),
('fit_onrate', fit_onrate)]))
```

```
        if fit_kinetic == True:
            cycleparm[i]['ystart'] = ystart
#        plotformat(axis, onrates, onrate_errors, offrates, offrate_errors)
    return cycleparm


sensors_180417_GGG_LaG9_vs_ybbR_TagGFP_4 = \
    [('G1', 'H1', 25e-9, 50e-9, 100e-9, 200e-9)]
analyze_experiment("180417_GGG-LaG9_vs_ybbR-TagGFP_4",
    sensors_180417_GGG_LaG9_vs_ybbR_TagGFP_4, fit_kinetic=True,
    fit_global=True, fit_stepwise=False, skipsteps=6, koff_initial=1e-4,
    kon_initial=1e5)

SortaseA_TagGFP_LPETGG_3 = [('C1', 'D1', 25e-9, 50e-9, 100e-9, 200e-9)]
analyze_experiment("SortaseA_TagGFP-LPETGG_3", SortaseA_TagGFP_LPETGG_3,
    fit_kinetic=False, ymax_linked = False, fit_global=True, fit_y0=False,
    offratefit_offset=False, skipsteps=10, reference=True, fit_stepwise=False)
```

References

[1] P. C. Fridy, Y. Li, S. Keegan, M. K. Thompson, I. Nudelman, J. F. Scheid, M. Oeffinger, M. C. Nussenzweig, D. Fenyö, B. T. Chait, et al., *Nat. Methods* **2014**, *11*, 1253–1260.
[2] D. G. Gibson, L. Young, R.-Y. Chuang, C. J. Venter, C. A. Hutchison III, H. O. Smith, *Nat. Methods* **2009**, *6*, 343–347.
[3] E. Durner, W. Ott, M. A. Nash, H. E. Gaub, *ACS Omega* **2017**, *2*, 3064–3069.
[4] 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, et al., *Nat. Commun.* **2014**, *5*, 1–8.
[5] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207–234.
[6] A. Skerra, A. Plückthun, *Science* **1988**, *240*, 1038–1041.
[7] K. S. Harris, T. Durek, Q. Kaas, A. G. Poth, E. K. Gilding, B. F. Conlan, I. Saska, N. L. Daly, N. L. van der Weerden, D. J. Craik, et al., *Nat. Commun.* **2015**, *6*, 10199.
[8] R. Yang, Y. H. Wong, G. K. T. Nguyen, J. P. Tam, J. Lescar, B. Wu, *J. Am. Chem. Soc.* **2017**, *139*, 5351–5358.
[9] O. Salama-Alber, M. K. Jobby, S. Chitayat, S. P. Smith, B. A. White, L. J. W. Shimon, R. Lamed, F. Frolow, E. A. Bayer, *J. Biol. Chem.* **2013**, *288*, 16827–16838.
[10] J. R. Rahn, R. B. Hallock, *Langmuir* **1995**, *11*, 650–654.
[11] L. a. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **2015**, *10*, 845–858.
[12] L. a. Kelley, M. J. E. Sternberg, *Nat. Protoc.* **2009**, *4*, 363–371.

Author Contributions

W.O. and E.D. designed and performed the experiments (W.O. constructed, expressed, purified and established all Sfp-based proteins. E.D. constructed, expressed, and purified all SrtA and OaAEP1-based proteins). W.O. and E.D. analyzed data. E.D. wrote the costum analysis software and fitted the data. W.O., E.D. and H.E.G. wrote and revised the manuscript.

3.5 Conformational Switching of a Protein Fold

In this article, we are able to show conformational transitions between different folded states of a cellulosomal protein by SMFS. We combine different measurement modes to acquire a very wide dynamic loading rate spectrum of the unfolding event of a Cohesin domain. This allows us to capture the features neccesary to distinguish the unfolding from at least two different states from the unfolding of one state via different unfolding pathways. Finally, we perform Monte Carlo simulations to illustrate that the observed behaviour is qualitatively consistent with a simple model of two folded states. Design and execution of this study was performed primarily by myself, as well as the drafting of the manuscript.

Conformational Transitions of a Protein Fold Observed by Dynamic Single-Molecule Force Spectroscopy

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

at the time of print, manuscript under review at PRL

Conformational Transitions of a Protein Fold Observed by Dynamic Single-Molecule Force Spectroscopy

Ellis Durner, Magnus S. Bauer, Markus A. Jobst,
Wolfgang Ott, Lukas F. Milles, and Hermann E. Gaub*
Lehrstuhl für Angewandte Physik and Center for Nanoscience,
Ludwig-Maximilians-Universität, 80799 Munich, Germany

1

Abstract

When studying the mechanical properties of protein folds by Single-Molecule Force Spectroscopy, in some cases non-homogeneous behavior is observed. In these cases, the underlying process of forced unfolding from two distinct states must be discerned from two unfolding pathways of a single state. For these processes, theoretical predictions have been made based on transition state models of unfolding trajectories along a free energy landscape induced by a given force loading rate. Here, we experimentally test these predictions with atomic force microscopy-based single-molecule force spectroscopy. Using a wide force-loading rate dynamic force spectrum and Monte Carlo simulations we experimentally validate the theoretical distinctions on a mechanostability-state-switching cohesin protein fold from Clostridia.

INTRODUCTION

Single-Molecule Force Spectroscopy has evolved into a key tool for the investigation of proteins under mechanical stress[1, 2]. The most commonly employed approach to explore the energy landscape of a protein fold or the force-induced unbinding of a receptor-ligand system, is to mechanically load the system under well defined boundary conditions. A force probe is used to stretch single molecules in a known anchoring geometry, which can be controlled by site-specific surface immobilization strategies [3–5]. Most commonly, this is done either at constant retraction velocities $\dot{z}$ or at constant force loading rates $\dot{F}$ and until the protein unfolds or the complex dissociates[6]. The resulting dynamic force spectrum can often be interpreted with the Bell-Evans model [7], where the most probable rupture force of a single energy barrier increases linearly with the logarithm of the force loading rate. While this will in most cases be a gross oversimplification of the roughly 3^N-dimensional energy landscape of a fold consisting of N amino acids, it describes the unbinding process from an energy minimum along a certain reaction coordinate well if there is only one dominant rate-limiting step. More recently, and with improved understanding of the molecular mechanisms, also non-linear dependencies have been predicted and observed [8–10]. However, all of these models are only consistent with a unimodal rupture force distribution at a given force loading rate. Recently, we have reported on the mechanostability of a family of cohesin-domains

* gaub@lmu.de

from *Acetovibrio cellulolyticus* (*A.c.*), which are part of the cellulose degrading machinery of these gram-positive bacteria [11–13]. We found, that while certain cohesins exhibited bimodal rupture force distributions, unimodality was recovered upon binding of a ligand dockerin. No dependence of this behavior on the loading rate was observed. We proposed, that a multitude of different folded states undergo a transition to a single folded state upon dockerin binding. In this study, we improve our understanding of this protein fold by resolving its unfolding behaviour under a very wide range of loading rates. We achieved this by observing cohesin 3 (Coh3) in both measurement modes of applying mechanical stress during a constant retraction velocity and under a constant force loading rate. We reveal that Coh3 adopts different folded states with distinct mechanical stabilities within the experimental timescale, a behavior which has been predicted[14], but was not experimentally observed for a single protein domain before.

RESULTS AND DISCUSSION

To gain a more in-depth understanding of the dynamic force response of *A.c.*'s CipA cohesin domains, we chose Coh3 for further study. It has previously been characterized as being neither particularly strong nor weak in terms of force resilience when compared to the other cohesin domains of the same scaffold[12], and it had not exhibited any bimodality in its rupture force distribution, with or without a dockerin bound[13]. We performed a combined constant velocity and constant loading rate experiment, probing Coh3 at four different loading rates and four different constant retraction velocities, all with the same cantilever. This enabled us to obtain a dynamic force spectrum of Coh3 unfolding events that covers over five orders of magnitude in loading rate. Figure 1 illustrates the experimental configuration and shows example traces of Coh3 unfolding events, which match the contour length increment expected for full unfolding of Coh3.

To our surprise, the dynamic unfolding behaviour deviates markedly from the previously recorded log-linear behaviour[15]. At loading rates below $\sim 1000\,\mathrm{pN\,s^{-1}}$, a downwards kink of the most probable rupture force at a given loading rate is observed (Figure 2A). To exclude the possibility for this to be an artefact from combining data of constant speed and constant loading rate measurements, we assembled the dynamic force spectrum of the CttA-XDoc:CohE (R.f) receptor-ligand system, which is used here as a specific pulling han-

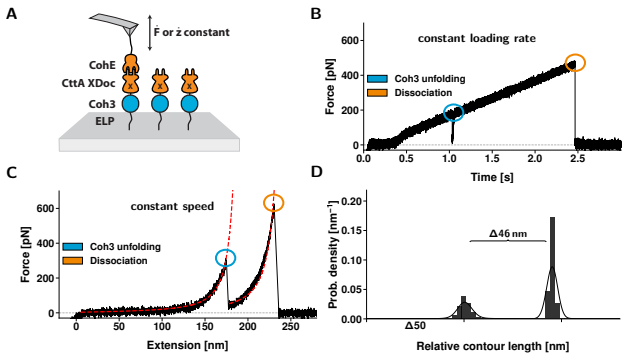


FIG. 1. (**A**) Schematic illustration of the experimental setup. The Ctta XDoc:CohE receptor ligand system was used to specifically and reliably probe Coh3, which was expressed as a fusion protein with XDoc. After each measurement, the surface is moved laterally to avoid probing the same Coh3 twice and the retraction speeds and loading rates are varied. (**B, C**) Example traces showing Coh3 unfolding in constant loading rate and constant speed mode. Unfolded peptide behaves as an entropic spring, as fits of a worm-like chain model (red lines) to both stretches in (C) illustrates. (**D**) Contour length transformation of the constant loading rate trace in (B), illustrating how Coh3 unfolding events are assigned if they match the expected contour length increment of 46 nm.

dle [5, 16]. Despite the loading rates not being constant in a constant velocity experiment due to linker molecules that act like entropic springs (such as the elastin-like polypeptide linkers employed here [17]), fits of the Bell-Evans model to the most probable rupture forces were in good agreement between the two measurement modes (see suppl. Figure S1).

By inspecting the unfolding force histograms of Coh3 for each measurement variation (Fig. 2B), we can gain a better understanding of the observed transition; while distributions at both the upper and the lower end of the loading rate spectrum appear unimodal, intermediate loading rates give rise to spectra that are distorted from what would be expected in the Bell-Evans model. Especially at $100\,\mathrm{nm\,s^{-1}}$, a bimodal rupture force distribution emerges, which is similar to the mutant Coh1(T107S) as investigated by Verdorfer et al[13]. There, it was suggested that either different conformations or different unfolding pathways give rise

to the bimodality of these distributions, and it was shown that the fold could be stabilized upon binding to the ligand dockerin. Given that unfolding rates for a single barrier can be expressed as:

$$k(F) = k^0_{off} \exp(\frac{F\Delta x}{k_B T}), \tag{1}$$

we can exclude that the bimodality can arise from one state with multiple unfolding pathways: As the unfolding probability p_i along a pathway i per time interval Δt is given as $p_i = k_i(F) * \Delta t$, it will increase monotonously with force. Consequently, the cumulative probability of n pathways

$$\sum_n p_i = \sum_n k_i(F) * \Delta t \tag{2}$$

will still result in a monotonously increasing rupture probability, and therefore a unimodal rupture force distribution. The data presented here not only require the existence of at least two different conformations for Coh3, which are distinct in their ability to withstand a stretching force. They also suggest that these states interconvert within the experimental timescale.

A simple model in line with the experimental data consists of two states with distinct single barriers to the unfolded state. Switching rates between these two states can be modeled as thermal crossings over a separating energy barrier with either a reaction coordinate affected by, or orthogonal to the applied force. The system is assumed to behave Markovian. To illustrate this, we performed Monte Carlo simulations for the simplest case: Figure 3 shows simulated data for a two-state system with fixed, non-force dependent rates for switching between the two folded states. The barriers' Δx and k_{off} for each state are taken from fits to either the two highest or the two lowest loading rates of the experimental dynamic force spectrum as indicated by the dashed lines in Figure 2A.

As no Monte Carlo fitting to the experimental data was performed due to the computational cost, rates for switching between the two folded states were chosen by hand. Hence, Figure 3 is merely demonstrates the general applicability of this simple two-state model[18], given the experimental data. A rigorous theoretical analysis of different possible cases of multi-state systems is provided in recent work by Pierse and Dudko[14]. In conclusion, we were able to experimentally observe a protein fold which needs to be described by theory beyond the standardly employed Bell Evans model. Our results are consistent with a two-state model as described in [14]. By combining constant loading rate and constant speed

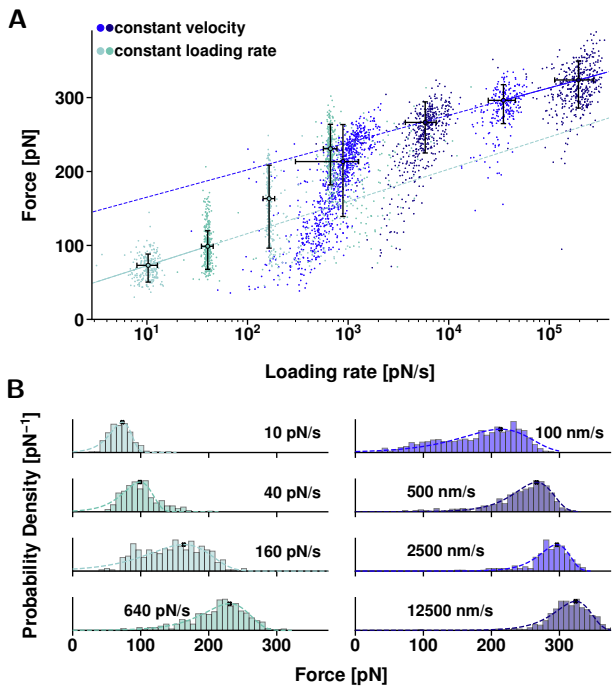


FIG. 2. (**A**) Dynamic force spectrum of Coh3 unfolding events, assembled from constant loading rate experiments performed at loading rates of $10\,\mathrm{pN\,s^{-1}}$, $40\,\mathrm{pN\,s^{-1}}$, $160\,\mathrm{pN\,s^{-1}}$, and $640\,\mathrm{pN\,s^{-1}}$ (mint green) and from constant velocity experiments performed at $100\,\mathrm{nm\,s^{-1}}$, $500\,\mathrm{nm\,s^{-1}}$, $2500\,\mathrm{nm\,s^{-1}}$, and $12\,500\,\mathrm{nm\,s^{-1}}$ (blue). To illustrate the deviation from a simple log-linear behavior, the Bell Evans model (Equation 3) was fitted to either the two lowest loading rate values or the two highest loading rate values and the resulting fits were then extrapolated and plotted as dashed lines. (**B**) Rupture force histograms of the Coh3 unfolding events presented in (A). While distributions at both low loading rates and high pulling speeds appear unimodal, intermediate measurement variations show a bimodal behaviour, especially pronounced at a velocity of $100\,\mathrm{nm\,s^{-1}}$. Distributions were fitted with the Bell Evans model (suppl. Equation 4), describing the unfolding probability at a given force. It is evident that the model is a poor fit to the experimental data , as the width of the distribution should not change for different loading rates.

6

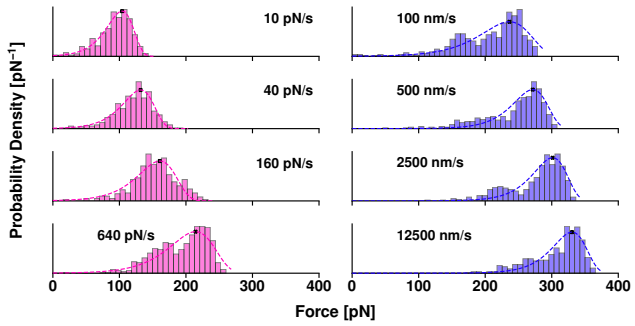


FIG. 3. Simulated unfolding events at different pulling speeds and loading rates for a complex that switches between two different folded states, independently of the applied force. Parameters Δx and k^0_{off} for the two states were chosen from the dynamic loading rate fits (see Figure 2A): $\Delta x_A = 0.25\,\mathrm{nm}$, $k^0_{offA} = 2.84 \times 10^{-5}\,\mathrm{s}^{-1}$, $\Delta x_B = 0.22\,\mathrm{nm}$, $k^0_{offB} = 1.1 \times 10^{-2}\,\mathrm{s}^{-1}$. The switching rates between the states are $k_{AB} = 4.8\,\mathrm{s}^{-1}$ and $k_{BA} = 19.2\,\mathrm{s}^{-1}$, chosen by hand to qualitatively illustrate the consistency of the model with the experimental data.

experimental modes, we were able to capture the non-monotonic trend in force distributions consistent with a two-state model. Thereby we could cover the wide range of loading rates[19] that is required to properly capture the dynamic force response of this protein domain, which had remained opaque in previous studies [12, 13].

The authors would like to thank Constantin Schöler for helpful discussions. We acknowledge funding from the German Research Foundation within the framework of the SFB1032.

[1] M. Krieg, B. M. Gaub, G. Fläschner, H. E. Gaub, Y. F. Dufrêne, G. J. L. Wuite, W. H. Roos, C. Gerber, D. Alsteens, and D. J. Müller, Nature Reviews Physics **1**, 41 (2018).

[2] H. Yu, M. G. W. Siewny, D. T. Edwards, A. W. Sanders, and T. T. Perkins, Science **355**, 945 (2017).

[3] Y. Taniguchi and M. Kawakami, Langmuir **26**, 10433 (2010).

[4] 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, Journal of

the American Chemical Society **139**, 9867 (2017).

[5] E. Durner, W. Ott, M. A. Nash, and H. E. Gaub, ACS Omega **2**, 3064 (2017).

[6] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, and M. A. Nash, Journal of Structural Biology **197**, 3 (2017).

[7] E. Evans and K. Ritchie, Biophysical Journal **72**, 1541 (1997).

[8] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).

[9] R. W. Friddle, A. Noy, and J. J. De Yoreo, Proceedings of the National Academy of Sciences **109**, 13573 (2012), https://www.pnas.org/content/109/34/13573.full.pdf.

[10] F. Rico, A. Russek, L. González, H. Grubmüller, and S. Scheuring, Proceedings of the National Academy of Sciences **116**, 6594 (2019), arXiv:1808.07122.

[11] Y. Hamberg, V. Ruimy-Israeli, B. Dassa, Y. Barak, R. Lamed, K. Cameron, C. M. Fontes, E. A. Bayer, and D. B. Fried, PeerJ **2**, e636 (2014).

[12] T. Verdorfer, R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash, and H. E. Gaub, J. Am. Chem. Soc. **0**, 0 (2017).

[13] T. Verdorfer and H. E. Gaub, Scientific Reports **8**, 10.1038/s41598-018-27085-x (2018).

[14] C. A. Pierse and O. K. Dudko, Phys. Rev. Lett. **118**, 088101 (2017).

[15] M. Schlierf and M. Rief, Biophysical Journal **90**, L33 (2006).

[16] 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, Nature Communications **5**, 5635 (2014).

[17] W. Ott, M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, and H. E. Gaub, ACS Nano **11**, 6346 (2017).

[18] M. Rief, J. M. Fernandez, and H. E. Gaub, Physical Review Letters **81**, 4764 (1998).

[19] F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring, Science **342**, 741 (2013).

Supplemental Material for
"Conformational Transitions of a Protein Fold Observed by Dynamic Single-Molecule Force Spectroscopy"

Ellis Durner, Magnus S. Bauer, Markus A. Jobst,
Wolfgang Ott, Lukas F. Milles, and Hermann E. Gaub
Lehrstuhl für Angewandte Physik and Center for Nanoscience,
Ludwig-Maximilians-Universität, 80799 Munich, Germany

S1

SUPPLEMENTARY INFORMATION

Protein Preparation for Experiments

Cloning

A modified pET28a vector encoding for GGG-HIS-G-Coh3(A.c.)-XDoc3[1, 2] was reconstituted via Gibson[3] assembly.

Protein expression and purification

GGG-HIS-G-Coh3(A.c.)-XDoc3, CohE-HIS-LPETGG[4] and d59 SortaseA(P94R/D160N/D165A/K190E/K196T) mutant[5] were expressed from modified pET28a vectors in NiCo21-(DE3)RIPL cells, which were cultivated in ZYM-5052 autoinduction media [6] supplemented with kanamycin and chloramphenicol. After pelleting, cells were lysed by sonication, then centrifuged at 4 °C, 39,000 rcf for 60 minutes. The supernatant was filtered through a 0.22 µm PES membrane (Carl Roth + Co. KG, Karlsruhe, Germany) and applied to Ni-NTA columns (HisTrap FF, GE Healthcare Europe GmbH, Freiburg, Germany). After washing with 6 column volumes of a buffer containing 25 mM TRIS, pH 8.4, 300 mM NaCl, 20 mM Imidazole and 0.25% (v/v) Tween-20, the bound fraction was eluted with a buffer containing 25 mM TRIS, pH 8.4, 300 mM NaCl and 300 mM imidazole. All protein solutions were concentrated using Amicon centrifugal filter units (10K MWCO, Merck KGaA, Darmstadt, Germany), followed by buffer exchange against Ca-TBS (25 mM TRIS, pH 7.2, 75 mM NaCl and 1 mM $CaCl_2$) buffer using ZebaSpin columns (Zeba spin desalting columns 7K, Thermo Fisher Scientific Inc.). Proteins were stored at −80 °C, with glycerol added to 10% (v/v).

Surface Functionalization

Glass surfaces and silicon nitride cantilevers (BioLever mini BL-AC40TS-C2, Olympus, Tokio, Japan) were silanized with (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany). Then a sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate) cross-linker (sulfo-SMCC, Thermo Fisher Scientific Inc.) was conjugated to the silanes amine groups in 50 mM 4-(2-hydroxyethyl)-1- piperazineethanesulfonic acid (HEPES) at pH 7.5. Next, the SMCC's maleimide groups were coupled to 50 µM Cys-

6xELP-LPETGG in case of the surfaces and to 50 µM GGG-6xELP-Cys[7] in case of the cantilevers. This reaction was carried out in 50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, with a pH of 7.2.

Single-Molecule Force Spectroscopy

Measurements were performed in Ca-TBS buffer using a custom built AFM instrument (sample driven vertically by PI-313 piezo actuator and laterally by a 100x100 nm stage (P-734.2CL), Physik Instrumente, Germany) in conjunction with a MFP-3D AFM controller (Asylum Research, Santa Barbara, USA) and a custom control software. Upon approaching the sample surface with the cantilever tip, the complex between CohE and CttA-XDoc was formed and the cantilever was retracted from the surface at constant velocities of $100\,\mathrm{nm\,s^{-1}}$, $500\,\mathrm{nm\,s^{-1}}$, $2500\,\mathrm{nm\,s^{-1}}$, and $12\,500\,\mathrm{nm\,s^{-1}}$ or at constant loading rates of $10\,\mathrm{pN\,s^{-1}}$, $40\,\mathrm{pN\,s^{-1}}$, $160\,\mathrm{pN\,s^{-1}}$, and $640\,\mathrm{pN\,s^{-1}}$. After each force-extension curve was acquired, the sample was moved laterally by 100 nm in order to not probe the same molecule twice on the surface side. Single-molecule interaction traces were identified by filtering the datasets using contour length analysis, and identifying only those traces in which a Coh3 unfolding events was observed[1, 8]. Loading-rates prior to Coh3 domain unfolding or complex dissociation were extracted by applying a linear fit to the last 3 nm in case of a constant velocity experiment. For constant loading rate experiments, a linear fit to the force over time was applied.
Dynamic force spectra were fitted to the Bell Evans model according to Equation 1.

$$F_{most\_probable}(\dot{F}) = \frac{k_B T}{\Delta x} \log(\frac{\dot{F}\Delta x}{k^0_{off} k_B T}) \tag{1}$$

Rupture force distributions were fitted to Equation 2, which gives the rupture force probability for a constant loading rate $\dot{F}$

$$p(F) = \frac{k^0_{off}}{\dot{F}} \exp(\frac{\Delta x F}{k_B T} - k^0_{off} k_B T \frac{\exp(\frac{\Delta x F}{k_B T}) - 1}{\Delta x \dot{F}}) \tag{2}$$

Monte Carlo simulations

We follow the approach described in Refs. ([9, 10]) by integrating $k(F) = k^0_{off} \exp(\frac{F\Delta x}{k_B T})$ over a timestep Δt. Here, k^0_{off} denotes the zero force off-rate and Δx is the distance to the

free energy barrier. As we are simulating a two state system, the probability to be in state A at $t = 0$ is given by $p_{A,initial} = \frac{1}{(k^0_{AB}/k^0_{BA})+1}$. We allowed for the timestep to be variable in order to keep the probability of overcoming the barrier within a certain timestep below a value of 0.0001. This enabled us to retain good precision without resorting to extremely small timesteps throughout a simulation run. For the constant speed mode, we chose to model the entropic stretching response of a 240 nm polymer linker according to Livadaru et al.[11]: $F(x) = \frac{k_B T}{cb(1-\frac{x}{L_c})}$. The loading rate prior to rupture was determined using the derivative $\dot{F}(x,v) = \frac{k_B T v}{cbL_c(1-\frac{x}{L_c})^2}$. For the constant loading rate mode, the current force at a certain time t was simply given by $F(t) = \dot{F}t$.

Protein Sequences

MGGG-HIS-G-Coh3(A.c)-CttA-XDoc(R.f)

MGGGHHHHHHGTGFTVSVDSVNGNVGEQIVIPVSFANIPANGISTADMTITYDSSKL
EYVSGVPGSIVTNPDVNFGINKETDGKLKVLFLDYTMSTGYISTSGVFTKVTFKVLS
SGGSTVGITGATFGDKNLGNVSATINAGSINGGVVPNTVTSAVKTQYVEIESVDGFY
FNTEDKFDTAQIKKAVLHTVYNEGYTGDDGVAVVLREYESEPVDITAELTFGDAT
PANTYKAVENKFDYEIPVYYNNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRD
ATATLTYYAATSTDGKDATTVALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELT
RFAKKAERLIDGRDASSILTFYTKSSVDQYKDMAANEPNKLWDIVTGDAEEE*

CohE(R.f)-HIS-HRV3C-LPETGG

MALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAE
GKYATTGYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGA
DDDFGADGVMWTVELKVPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQGK
LMQAYFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAGEPELKLPRSRHHHHHH
GSLEVLFQGPLPETGG*

S4

Supplementary Figures

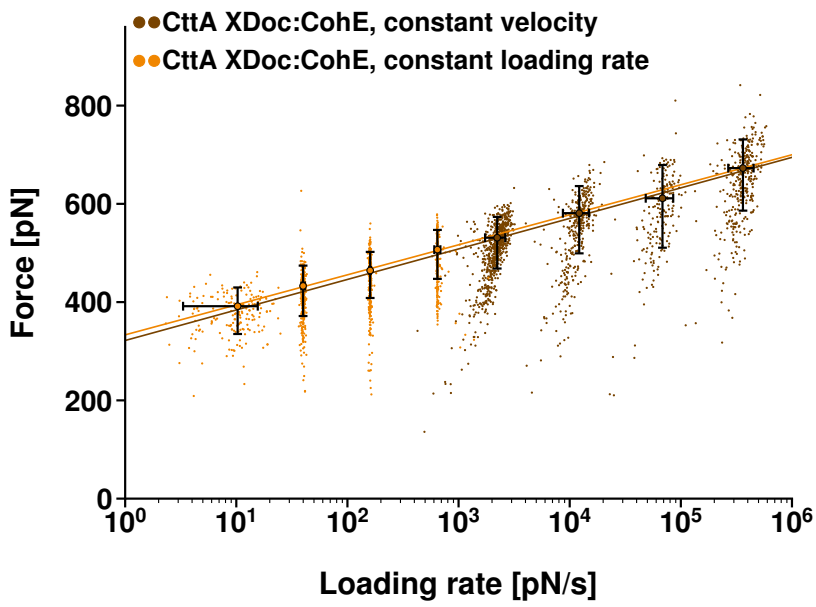


FIG. S1. Complex rupture forces of *Rf* CttA-XDoc:CohE with no prior XMod unfolding. Most probable loading rates and rupture forces of constant loading rate data and constant speed data were independently fitted with the Bell Evans model (see Equation 1).

S5

[1] T. Verdorfer, R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash, and H. E. Gaub, J. Am. Chem. Soc. **0**, 0 (2017).

[2] 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, Nature Communications **5**, 5635 (2014).

[3] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, Nature Methods **6**, 343 (2009).

[4] W. Ott, E. Durner, and H. E. Gaub, Angewandte Chemie International Edition **57**, 12666 (2018), https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201805034.

[5] I. Chen, B. M. Dorr, and D. R. Liu, Proceedings of the National Academy of Sciences of the United States of America **108**, 11399 (2011).

[6] F. W. Studier, Protein Expression and Purification **41**, 207 (2005).

[7] W. Ott, M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, and H. E. Gaub, ACS Nano **11**, 6346 (2017).

[8] E. M. Puchner, G. Franzen, M. Gautel, and H. E. Gaub, Biophysical journal **95**, 426 (2008).

[9] C. Schoeler, T. Verdorfer, H. E. Gaub, and M. A. Nash, Phys. Rev. E **94**, 042412 (2016).

[10] C. Friedsam, A. K. Wehle, F. K. hner, and H. E. Gaub, Journal of Physics: Condensed Matter **15**, S1709 (2003).

[11] L. Livadaru, R. R. Netz, and H. J. Kreuzer, Macromolecules **36**, 3732 (2003).

3.6 Immobilization Strategies for SMFS

When studying how protein domains react under force, careful design of the experimental arrangement is key. This article utilizes a combination of two orthogonal enzymatic coupling strategies, to covalently immobilize a protein of interest to a surface at one terminus, and to ligate a high force pulling handle to the other terminus. This allows to express the protein of interest with only two short tags, obviating the need for the classically employed polyprotein strategies. The high force receptor-ligand system CttA-XDoc:CohE enables specific and reliable probing of the anchored molecules. This article successfully demonstrates the compatibility of this immobilization and probing scheme with cell-free translation and transcription in microwells, directly on a sample slide, greatly reducing the amount of time and work spent on sample preparation. Design and execution of the study was performed by myself, as well as the drafting of the manuscript.

Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles

Ellis Durner, Wolfgang Ott, Michael A. Nash, and Hermann E. Gaub

published in

ACS Omega 2017, 2, 3064-3069

ACS OMEGA

Article

http://pubs.acs.org/journal/acsodf

Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles

Ellis Durner,[†] Wolfgang Ott,[†] Michael A. Nash,[‡,§] and Hermann E. Gaub*,[†]

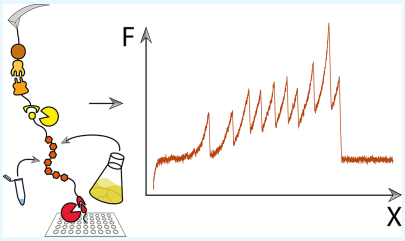
[†]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

[‡]Department of Chemistry, University of Basel, 4056 Basel, Switzerland

[§]Department of Biosystems Science & Engineering, Swiss Federal Institute of Technology (ETH-Zurich), 4058 Basel, Switzerland

Supporting Information

ABSTRACT: Single-molecule force spectroscopy greatly benefits from site-specific surface immobilization and specific probing with a functionalized cantilever. Here, we describe a streamlined approach to such experiments by covalently attaching mechanically stable receptors onto proteins of interest (POI) to improve pickup efficiency and specificity. This platform provides improved throughput, allows precise control over the pulling geometry, and allows for multiple constructs to be probed with the same ligand-modified cantilever. We employ two orthogonal enzymatic ligation reactions [sortase and phosphopantetheinyl transferase (Sfp)] to covalently immobilize POI to a pegylated surface and to subsequently ligate the POI to a mechanically stable dockerin domain at the protein's C-terminus for use as a high-strength pulling handle. Our configuration permits expression and folding of the POI to proceed independently from the mechanically stable receptor used for specific probing and requires only two short terminal peptide sequences (i.e., ybbR-tag and sortase C-tag). We applied this system successfully to proteins expressed using in vitro transcription and translation reactions without a protein purification step and to purified proteins expressed in *Escherichia coli*.



INTRODUCTION

In recent years, the field of single-molecule force spectroscopy (SMFS) has implemented many developments in bioconjugation to improve upon the classical approach of nonspecific pulling experiments by moving to specific, often covalent surface functionalization.[1,2] Traditionally, polyproteins are recombinantly expressed as fusion constructs framed by several repeats of marker domains of known unfolding patterns (often Ig-like domains) and nonspecifically deposited onto a surface.[3] A bare cantilever tip is then indented into the surface in an attempt to pickup and stretch single polyprotein chains on opposing ends by nonspecific adhesion. In case the number of domain unfoldings in the recorded data trace exceeds the number of domains on each side of the proteins of interest (POI), an N- to C-terminal stretching of the POI can be concluded.

In contrast to the nonspecific attachment, site-specific anchoring and probing approaches offer many advantages. They allow for homogeneous surface preparation as the immobilization geometry is defined; the usage of spacer molecules such as polyethyleneglycol (PEG) diminishes possible surface interaction effects. Drawbacks of unspecific probing—such as low-pickup efficiencies or the requirement of recombinant expression of large polyproteins—have been addressed by utilizing the receptor−ligand pairs as pulling handles to provide a specific interaction by which force can be applied to the POI. Systems such as StrepII-tag-Strep-Tactin,[4] streptavidin−biotin,[5,6] GCN4-peptide−antibody,[7] and cohesin−dockerin domains[8−10] are only a few of the interactions that have been employed for this purpose.

These pulling handles classically are genetically appended to the POI and expressed as fusion proteins. The fusion proteins are then covalently immobilized through one end of the POI and probed by ligand-functionalized cantilever-tips that recognize the respective receptor on the other end. A wide range of forces are accessible by utilizing short tags such as the StrepII-tag (116 pN at a loading rate of 4 nN s^{-1} if the tag is C-terminal and 46 pN at 4 nN s^{-1} if the tag is N-terminal[3]) and biotin (257 pN[5]), as well as that with larger handles such as the interaction between type-3 dockerin and cohesin E from *Ruminococcus flavefaciens*, reaching up to 700 pN at 100 nN s^{-1}.[10] These high-force interactions allow characterization of very stable proteins such as the unfolding of several Titin-Ig domains in series.

However, recombinant expression of a fusion between a (possibly large) POI and a large handle-protein (e.g., 29 kDa

Received: April 19, 2017
Accepted: May 26, 2017
Published: June 30, 2017

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

ACS Omega Article

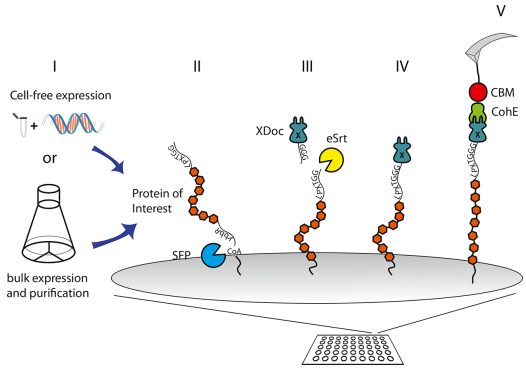


Figure 1. Schematic of the experimental setup. (I) POIs were either expressed in bulk or synthesized using a cell-free expression mix. POIs contained a ybbR-tag at the N-terminus and a sortase LPETGG tag at the C-terminus. (II) Surface-bearing PEG-coenzyme A is covalently modified with POIs via Sfp-catalyzed ligation. (III) Next, GGG-Doc is ligated to the POI at the C-terminal end using the LPETGG sortase-tag for use as a force spectroscopy pulling handle. (IV,V) Unfolding experiments are conducted by approaching and retracting a CohE-CBM-functionalized cantilever.

for CohE) can be cumbersome. The resulting fusion proteins might be insoluble or the correct folding of the POI might be affected by the presence of the fusion domains during translation and folding. Here, we utilize two orthogonal enzymatic ligation reactions to achieve sortase and phosphopantetheinyl transferase (Sfp)-mediated covalent surface attachment and post-translational modification of several POIs with dockerin handles by sortase-mediated[11,12] ligation. This allows the expression of only the protein domain of interest without risking to affect proper folding. The very robust interaction of type-3 dockerin and cohesin from *R. flavefaciens*was already shown to be functional over repeated measurements of about 24 h,[10] which is an important requirement for multiplexing atomic force microscopy (AFM) experiments.

Furthermore, we combined this technique with in vitro expression of the POI in a cell-free system. Because smaller proteins are, in general, expressed with higher yields,[13] the reduced size of the protein construct to be expressed is beneficial. This allows for a fast and easy workflow from plasmid DNA to covalently immobilized proteins containing mechanostable handles without the need for bulk expression. We anticipate that our approach will aid in highly parallel mechanical screening of mutant proteins, which benefits from the in vitro expression, obsoleting the need for protein purifications and benefitting from the enhancements in force spectroscopy throughput and robustness.

■ MATERIALS AND METHODS

Experimental Design. We selected Titin-Ig domains[14] and superfolder green fluorescent protein (sfGFP)[15] as the exemplary POIs for this study, as they are well-documented in the literature, and enable comparison with established methods. The POIs were cloned with a ybbR-peptide tag[16] at their N-terminus and a sortase A recognition sequence[17] LPETGG at their C-terminus. For force-spectroscopy handles, we used GGG-dockerin,[10] which was recombinantly expressed in *Escherichia coli*, purified, and ligated to the C-terminus of the POI using ligation with sortase A. While preliminary experiments were carried out with wild-type sortase A, an evolved mutant[18] was ultimately used because of its superior performance. On the cantilever side, CohE-CBM-ybbR was used and immobilized at the ybbR-tag via Sfp-catalyzed ligation.[16]

The two specific enzymatic recognition sites located on the termini of the POI ensure that only fully expressed proteins are measured in SMFS-experiments. Figure 1 shows a schematic overview of the experiment.

Cloning. Modified pET28a plasmids encoding for ybbR-His-XylanaseT6(T129C) (*Geobacillus stearothermophilus*)-Doc3 (*R. flavefaciens*), ybbR-His-sfGFP-DocI (*Clostridium thermocellum*), and Titin-Ig domains (repeats 27 to 32, repeat 34, human) were used as templates for polymerase chain reaction (PCR) with subsequent reconstitution by Gibson[19] assembly. The previously reported[18] d59 sortase(P94R/D160N/D165A/K190E/K196T) mutant was created by introducing the mutations via overlap extension PCR followed by ligating the linearized plasmid using Kinase−Ligase−DpnI (KLD) enzyme mix and KLD reaction buffer from the Q5 site-directed mutagenesis kit (New England Biolabs, MA, USA). The chemically competent *E. coli* DH5-α cells were transformed [Life Technologies GmbH, Frankfurt, Germany; 30 min on ice, 30 s heat shock at 42 °C followed by 37 °C for 1 h in a super optimal broth with catabolite repression medium] and plated on kanamycin-supplemented agar plates. For amino acid sequences, see the Supporting Information.

Protein Expression and Purification. All proteins were expressed in *NiCo21(DE3)RIPL* cells, which were cultivated in ZYM-5052 autoinduction media[20] supplemented with kanamycin and chloramphenicol. After pelleting, the cells were lysed by sonication and then centrifuged at 4 °C, 39 000 rcf for 60 min. The supernatant was filtered to 0.22 μm and applied to Ni-NTA columns (HisTrap FF, GE Healthcare Europe GmbH, Freiburg, Germany). After washing with 6 column volumes of a buffer containing 25 mM Tris, pH 8.4, 300 mM NaCl, 20 mM imidazole, and 0.5 vol % Triton X-100, the bound fraction was eluted with an elution buffer containing 25 mM Tris, pH 8.4, 300 mM NaCl, and 300 mM imidazole.

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

ACS Omega Article

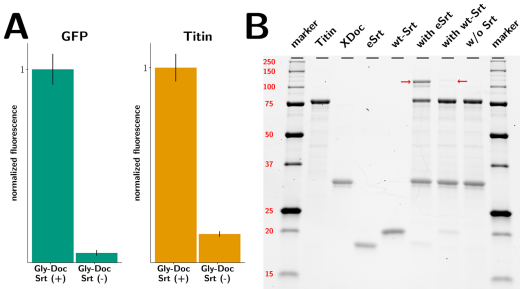


Figure 2. (A) Averaged fluorescence intensities of a CohE-CBM-ybbR-CoA647-labeled surface functionalized with ybbR-Titin-Ig-LPETGG and ybbR-sfGFP-LPETGG. Each protein was immobilized at two separate spots that were then incubated with either GGG-dockerin and sortase or GGG-dockerin but not with sortase. To test for successful ligation of dockerins, CohE-CBM-ybbR-CoA647 was allowed to bind for 10 min at 300 nM, then rinsed and imaged immediately afterward. Fluorescent intensities of each construct were normalized to the intensity of the sortase-positive spot. (B) SDS-PAGE demonstrating the ligation of GGG-dockerin to ybbR-Titin-LPETGG with wild-type sortase A (wt-Srt), pentamutant sortase A (eSrt), or no sortase as negative control. The red arrows are indicating the ligation products.

All protein solutions were concentrated using Amicon centrifugal filter units (10k MWCO, Merck KGaA, Darmstadt, Germany), followed by buffer exchange to Ca-TBS buffer (25 mM Tris, pH 7.2, 75 mM NaCl, and 1 mM $CaCl_2$) using polyacrylamide spin desalting columns. Proteins were stored at −80 °C with glycerol added to 10% (v/v). For cell-free expression, 25 μM reactions of PURExpress In Vitro Protein Synthesis Kit (New England Biolabs, Ipswich, Massachusetts) were incubated for 2 h at 37 °C, containing 300 ng plasmid DNA coding for the POIs.

In case of MGGG-His-Doc, the N-terminal methionine cleavage in *E. coli* was sufficient for the preparation of GGG-His-Doc, so that no additional protease digestion was necessary.

Surface Preparation. Surfaces and cantilevers for force spectroscopy were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany) and PEGylated with α-maleimide-hexanoic-ω-NHS PEG (NHS-PEG5000-Mal, Rapp Polymere, Tübingen, Germany) dissolved to 25 mM in 4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid buffer (HEPES), 50 mM, pH 7.5 to provide suitable conditions for NHS coupling. Then, the PEGylated surfaces and cantilevers were coupled to coenzyme A (CoA, 1 mM) in sodium phosphate buffer, pH 7.2.

Silicon nitride cantilevers (BioLever mini BL-AC40TS-C2, Olympus, Tokyo, Japan) were used as force probes. Silicone masks with a grid of 1 mm-diameter holes (CultureWell Reusable Gaskets, Grace Bio-Labs, Bend, OR, USA), were applied to the CoA-functionalized glass slides to create separated incubation wells. Each purified POI was diluted to 50 μM in Ca-TBS that was supplemented with 20 mM $MgCl_2$ and Sfp enzyme was added to 10 μM. The reaction mixtures were added to the single incubation wells in the mask, enabling covalent immobilization via Sfp-catalyzed ligation of CoA and the ybbR tags.

For cell-free expression of the POIs, the cell-free expression reaction mix (PURExpress, New England Biolabs, MA, USA) was prepared to contain 100 ng of plasmid DNA. The expression mix was incubated at 37 °C for 2 h, then supplemented with Sfp enzyme to 10 μM and directly applied to the micowells without further purification. Sfp ligation reactions were performed for 2 h at room temperature. After subsequent rinsing with Ca-TBS, the wells were incubated with 100 μM GGG-Doc protein and 10 μM sortase A for 1 h. After rinsing with Ca-TBS, the silicon mask was removed, providing an array of covalently linked proteins that were modified with the dockerin handle at one end.

The sortase-catalyzed ligation reactions for Figure 2 contained 10 μM ybbR-Titin-LPETGG, 10 μM GGG-Doc, and 10 μM of either wild-type d59 sortase or the evolved pentamutant.[18] The ligation reactions were incubated for 1 h at 37 °C.

For surface functionalization tests, CohE-CBM-ybbR was labeled with CoA647 (New England Biolabs, MA, USA) in a reaction containing 25 μM CoA647, 10 μM CohE-CBM-ybbR, and 2 μM Sfp in Ca-TBS supplemented with 20 mM $MgCl_2$. The labeling reaction was incubated for 4 h at 37 °C. Free dye and Sfp enzyme were removed via preparative gel filtration with Ca-TBS as the running buffer through a Yarra 3 μm SEC-3000 (Phenomenex, Torrance, California, USA) column. Appropriate fractions [evaluated via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)] were pooled, diluted to 3 μM, supplemented with glycerol to 10% (v/v), and stored at −80 °C.

Labeled surfaces were imaged using ChemiDoc MP (Bio-Rad, Hercules, California, USA), with 625(30) nm/695(55) nm emission/excitation filters. The exposure time was 30 s; for background subtraction, a blank and clean cover slip was imaged with the same settings and an average background signal was subtracted from the measured average intensities. Intensities were quantified via Image Lab 5.2 (Bio-Rad, Hercules, California, USA) volume tool.

Single-Molecule Force Spectroscopy. All data were obtained using Ca-TBS. Measurements were taken with custom-built instruments (driven by PI-731 piezo actuators, Physik Instrumente, Germany) in conjunction with MFP-3D AFM controllers (Asylum Research, Santa Barbara, USA). Upon approaching the sample surface with the cantilever tip, the complex between cohesin/dockerin (C/D) was formed, and the cantilever was retracted from the surface at a constant

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

ACS Omega Article

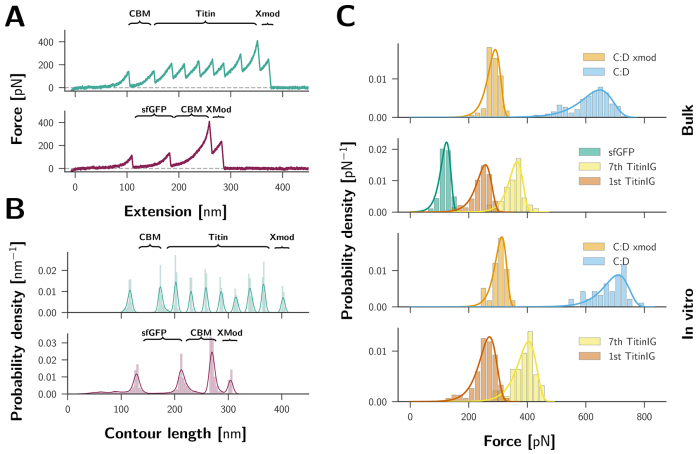


Figure 3. SMFS on Ctta-dockerin-labeled 7× Titin-Ig and sfGFP. (A) Force distance traces showing complete unfolding of the POI (Titin-Ig unfolding is shown in the upper trace, sfGFP in the lower trace). (B) Transformation of traces from (A) into contour-length space. (C) Force histograms of complex dissociation events and unfolding events of the POI: the upper two panels contain data from the bulk-expressed proteins and the lower two panels contain data from in vitro-expressed proteins. C/D complex dissociation can occur with [as in both sample traces shown in (A)] or without prior unfolding of the x-module, which is a subdomain of the dockerin, resulting in two populations of the dissociation forces. Each population was fitted with the Bell–Evans model.[23]

velocity of 800 nm s^{-1} while recoding the distance and cantilever deflection at a sampling rate of 12 500 Hz. After each force–extension curve was recorded, the sample was moved laterally by 100 nm to probe a different molecule. For data analysis, force–distance curves were transformed into contour length space using a freely rotating chain model with quantum mechanical corrections for peptide backbone stretching[21] and then sorted by contour length increments.[22] Loading rates prior to domain unfolding or complex dissociation were extracted by applying a linear fit to the last 3 nm before the respective event and then used in fitting the rupture-force histograms with the Bell–Evans model.[23]

■ RESULTS AND DISCUSSION

To test for successful surface functionalization, we incubated surfaces that had been prepared as described in the Materials and Methods section with fluorescently labeled cohesin. Figure 2A confirms that if sortase is ommited, no dockerin functionalization is achieved, whereas if sortase was present to perform the ligation reaction, binding of CoE-CBM-ybbR-CoA647 is observed. Figure 2B demonstrates successful ligation of GGG-dockerin to ybbR-Titin-LPETGG and illustrates the superior performance of the evolved sortase mutant d95/P94R/D160N/D165A/K190E/K196T[18] in comparison with wild-type sortase A.

Typical single-molecule force–distance unfolding patterns for the sortase-incubated spots are shown in Figure 3A. They exhibit the unbinding pattern of CohE–Doc dissociation as characterized in previous publications,[10] where dissociation can occur with or without unfolding of the dockerin subdomain called x-module. The resulting force–distance curves were transformed into contour length space and then sorted by comparing the observed unfolding increments (3B). Only curves exhibiting the 56 nm increment corresponding to a full unfolding of the CBM-domain were classified to be the result of probing a CohE–Doc complex. Furthermore, the curves were sorted to exhibit no more than one increment corresponding to the unfolding of sfGFP 79 nm and no more than seven increments corresponding to Titin-Ig unfolding 28 nM. These increments result from the added free contour length of the peptide chain upon unfolding the folded protein domains and match the previously reported values.[14,15,24] For these traces, unfolding forces of the domains of interest were histogrammed with a bin width of 20 pN (Figure 3C).

Despite its narrow tip apex, each cantilever is typically functionalized with multiple cohesin-anchors; hence, multiple receptor–ligand complexes can form if dockerin-decorated surface is densely populated. Therefore, we went for a rather sparse surface functionalization which can be tuned by the incubation times of Sfp and sortase-catalyzed ligation reactions and/or the substrate concentrations. Alternatively, cantilevers with blunter tips could be used when more interactions are desired. The achieved surface densities were in a suitable range for SMFS, sparse enough to avoid multiple interactions but dense enough to acquire good statistics. Probing attempts (1.24%) resulted in single molecule unfolding traces satisfying the outlined criteria. In total, 142 Titin-Ig and 92 sfGFP single molecule traces were obtained within 11 h of measurement with a single cantilever (spring constant: 0.093 N m^{-1}). If sortase had been omitted, no traces showing unfolding of CBM and one of the POI were recorded. For probing of in vitro-expressed Titin-Ig, 0.33% of attempts were successful, yielding 72 Titin-Ig unfoldings in 9 h of measurement, which was also probed with a single cantilever (spring constant: 0.097 N m^{-1}).

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064–3069

Figure 3C shows force histograms for unfolding events of sfGFP, the last of seven Titin-Ig domain to unfold and the complex dissociation itself. This was carried out for bulk-expressed and purified sfGFP and Titin-Ig, as well as for Titin-Ig expressed in the cell-free system. Complex dissociation events cluster into two populations that are characteristic of Doc/Coh unbinding.[10] The most probable forces at which the POI unfold are $124\left(^{+24}_{-16}\right)$ pN for sfGFP, $257\left(^{+36}_{-24}\right)$ pN for the first, and $365\left(^{+33}_{-22}\right)$ pN for the last Titin-Ig domain to unfold ($271\left(^{+42}_{-28}\right)$ and $404\left(^{+45}_{-30}\right)$ pN for Titin-Ig expressed in the cell-free system), the asymmetrical full widths at half maximum of the distributions are given in brackets. The most probable forces were determined by fitting each histogram of unfolding forces with the Bell–Evans model.[23]

The differences between the most probable unfolding forces observed for the POI expressed in the cell-free system and the bulk-expressed proteins are within tolerance of errors resulting from cantilever calibration.[25]

This method can be easily applied to any recombinantly expressed protein by adding the terminal peptide tags necessary for covalent surface attachment and post-translational sortase-mediated ligation. Owing to the terminal location of these tags, only nondigested and fully expressed proteins are probed. This is especially advantageous for cell-free expression systems, where the small quantity of expressed protein often makes the usually necessary affinity purification cumbersome.

CONCLUSIONS

We developed a method that enables acquisition of SMFS datasets of specifically probed and covalently immobilized single molecules. By post-translationally modifying the POI with the high-force interactions of the Coh/Doc receptor–ligand system via sortase ligation, we can probe even resilient proteins such as Titin-Ig domains with high specificity and throughput, improving on the nonspecific polyprotein method and eliminating the requirement of expressing the POI as large fusion constructs with handle domains. The modular system of post-translational attachment of the mechanostable pulling handle allowed us to probe different proteins with the same cantilever. We also applied this approach to proteins expressed in cell-free systems without further purification while still selecting for only fully expressed proteins owing to the specificity provided by the high-affinity pulling handle.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.7b00478.

Amino acid sequences (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: gaub@lmu.de (H.E.G.).

ORCID

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

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the EU 7th Framework Programme NMP4-SL-2013-604530 (CellulosomePlus) and the Society in Science—the Branco Weiss Fellowship from ETH Zürich. The authors thank M. A. Jobst for the instrument control software, L. F. Milles for providing the force curve analysis software, and Thomas Nicolaus as well as Angelika Kardinal for laboratory assistance.

REFERENCES

(1) 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.

(2) Alsteens, D.; Gaub, H. E.; Newton, R.; Pfreundschuh, M.; Gerber, C.; Müller, D. J. Atomic force microscopy-based characterization and design of biointerfaces. *Nat. Rev. Mater.* **2017**, *2*, 17008.

(3) Li, L.; Huang, H. H.-L.; Badilla, C. L.; Fernandez, J. M. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* **2005**, *345*, 817–826.

(4) 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.* **2016**, *11*, 89–94.

(5) Moy, V. T.; Florin, E.-L.; Gaub, H. E. Intermolecular forces and energies between ligands and receptors. *Science* **1994**, *266*, 257–259.

(6) Edwards, D. T.; Faulk, J. K.; Sanders, A. W.; Bull, M. S.; Walder, R.; LeBlanc, M.-A.; Sousa, M. C.; Perkins, T. T. Optimizing 1-μs-Resolution Single-Molecule Force Spectroscopy on a Commercial Atomic Force Microscope. *Nano Lett.* **2015**, *15*, 7091–7098.

(7) Morfill, J.; Neumann, J.; Blank, K.; Steinbach, U.; Puchner, E. M.; Gottschalk, K.-E.; Gaub, H. E. 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.* **2008**, *381*, 1253–1266.

(8) 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.

(9) 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.

(10) 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. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **2014**, *5*, 5635.

(11) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **1999**, *285*, 760–763.

(12) Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L. Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **2013**, *8*, 1787–1799.

(13) Gagoski, D.; Polinkovsky, M. E.; Mureev, S.; Kunert, A.; Johnston, W.; Gambin, Y.; Alexandrov, K. Performance benchmarking of four cell-free protein expression systems. *Biotechnol. Bioeng.* **2016**, *113*, 292–300.

(14) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **1997**, *276*, 1109–1112.

(15) Kufer, S. K.; Dietz, H.; Albrecht, C.; Blank, K.; Kardinal, A.; Rief, M.; Gaub, H. E. Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. *Eur. Biophys. J.* **2005**, *35*, 72–78.

(16) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically encoded short peptide tag for versatile protein labeling by Sfp

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064–3069

phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15815−15820.
(17) Popp, M. W.-L.; Antos, J. M.; Ploegh, H. L. Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. In *Current Protocols in Protein Science*; John Wiley & Sons, Inc., 2009; pp 15.3.1−15.3.9.
(18) Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11399−11404.
(19) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343−345.
(20) Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif.* **2005**, *41*, 207−234.
(21) Livadaru, L.; Netz, R. R.; Kreuzer, H. J. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules* **2003**, *36*, 3732−3744.
(22) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. Comparing proteins by their unfolding pattern. *Biophys. J.* **2008**, *95*, 426−434.
(23) Evans, E.; Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **1997**, *72*, 1541−1555.
(24) 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.
(25) Gibson, C. T.; Smith, D. A.; Roberts, C. J. Calibration of silicon atomic force microscope cantilevers. *Nanotechnology* **2005**, *16*, 234−238.

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

Post-translational sortase-mediated attachment of high-strength force spectroscopy handles

Ellis Durner,[†] Wolfgang Ott,[†] Michael A. Nash,[‡,¶] and Hermann E. Gaub[*,†]

†*Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany*

‡*Department of Chemistry, University of Basel, 4056 Basel Switzerland*

¶*Department of Biosystems Science & Engineering, Swiss Federal Institute of Technology (ETH-Zurich), 4058 Basel Switzerland*

E-mail: gaub@lmu.de

Supporting Information Available

Sequences

pET28a-MGGG-HIS-DocIII

MGGGHHHHHHGVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTV
YNEGYTGDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYEIPVYY
NNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATT
VALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTF
YTKSSVDQYKDMAANEPNKLWDIVTGDAEEE

pET28a-CohE-CBM(C63S)-HIS-ybbR

MGTALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSG
AEGKYATTGYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVAS

1

GADDDFGADGVMWTVELKVPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQ
GKLMQAYFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAGEPGSVVPSTQPVTTP
PATTKPPATTIPPSDDPNAMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSS
AIDLSKLTLRYYYTVDGQKDQTFWSDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNN
ADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTA
YLNGVLVWGKEPGELKLPRSRHHHHHHGSLEVLFQGPDSLEFIASKLA

ybbR-HRV3C-HIS-sfGFP-LPETGG

MDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASSKGEELFTGVVPILVELDGDVNG
HKFSVRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMK
RHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDG
NILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGITHGMDELYKLPETGG

ybbR-HRV3C-HIS-Titin-LPETGG

MDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASLIEVEKPLYGVEVFVGETAHFEI
ELSEPDVHGQWKLKGQPLAASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKS
AANLKVKELPLIFITPLSDVKVFEKDEAKFECEVSREPKTFRWLKGTQEITGDDRFE
LIKDGTKHSMVIKSAAFEDEAKYMFEAEDKHTSGKLIIEGIRLKFLTPLKDVTAKEK
ESAVFTVELSHDNIRVKWFKNDQRLHTTRSVSMQDEGKTHSITFKDLSIDDTSQIRV
EAMGMSSEAKLTVLEGDPYFTGKLQDYTGVEKDEVILQCEISKADAPVKWFKDGK
EIKPSKNAVIKADGKKRMLILKKALKSDIGQYTCDCGTDKTSGKLDIEDREIKLVRP
LHSVEVMETETARFETEISEDDIHANWKLKGEALLQTPDCEIKEEGKIHSLVLHNCR
LDQTGGVDFQAANVKSSAHLRVKPRVIGLLRPLKDVTVTAGETATFDCELSYEDIP
VEWYLKGKKLEPSDKVVPRSEGKVHTLTLRDVKLEDAGEVQLTAKDFKTHANLFV
KAPHVEFLRPLTDLQVREKEMARFECELSRENAKVKWFKDGAEIKKGKKYDIISKG
AVRILVINKCLLDDEAEYSCEVRTARTSGMLTVLECLPETGG

2

3.7 Linkers for SFMS

While linkers are beneficial to SMFS experiments in that they provide spacing of the molecules of interest from the surface they are anchored to, the most commonly employed PEG linkers can introduce artifacts themselves. When probing the elastic response of a polypeptide chain to force, polymer elasticity models allow us to access the total contour length of the stretched chain. If the stretched system is a mixed one, containing both polypeptide and PEG linkers, the differing elastic responses of each component, as well as the deviation of PEG from a purely entropic behaviour at higher forces, analysis of SMFS data becomes intricate. Here, we develop a genetically encoded ELP linker for force spectroscopy experiments. Not only does this ensure monodispersity in length, it is also ideal for force spectroscopy experiments on protein systems, as the ELP linkers are polypeptides themselves. The linkers allow for easy, covalent surface modification via a terminal Cysteine, and provide a Sortase recognition sequence at the other terminus, which is used to ligate a protein of interest under mild reaction conditions. We show that this linker system can be used to acquire high quality data, with a stretching response that more closely follows established polymer elasticity models than a mixed system. I contributed to this work by providing the pentamutant Sortase A variant used to covalently couple the ELP linkers to the molecules of interest. I also contributed to discussions about data analysis and polymer elasticity.

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 2017, 11, 6346-6354

ACSNANO

www.acsnano.org

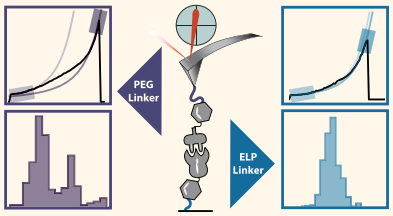
ARTICLE

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-

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

ACS Nano Article

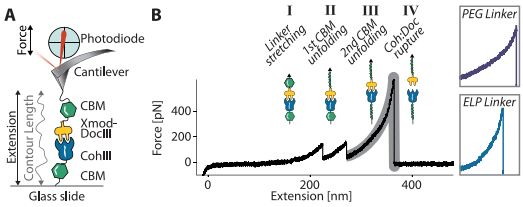


Figure 1. (A) SMFS configuration: Cantilevers are functionalized with CBM-Xmod-DocIII fusion proteins. Glass slides are modified with CohIII-CBM constructs. (B) Coh:Doc-based SMFS unfolding trace. Following Coh:Doc complex formation at zero extension, retraction of the cantilever results in mechanical 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.

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 ($ELP_{120\ nm}$, 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 ($PEG_{120\ nm}$). The maleimide groups of the

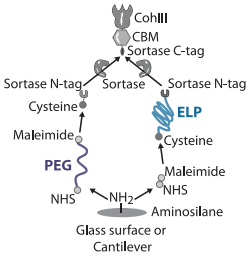


Figure 2. Comparison of immobilization strategies. For standard immobilization with PEG spacers, NHS chemistry was used to link PEG to amino-silanized surfaces. Protein constructs were then coupled *via* cysteine-sortase tag peptides to the maleimide end-groups on the PEG spacers. For immobilization 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.

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-$ELP_{120\ nm}$-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-$ELP_{120\ nm}$-LPETGG or to $PEG_{120\ nm}$-coupled Cys-LPETGG. Our enzyme-mediated protein immobilization approach has the advantage of site-specific linkages and results in a homogeneous

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

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

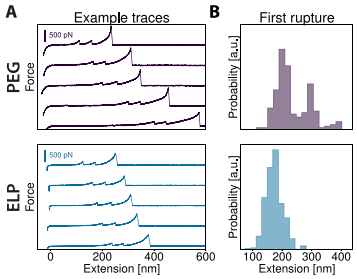


Figure 3. Comparison of dispersity of PEG and ELP linkers. (A) Typical force–extension traces for PEG (purple) and ELPs (blue). In the PEG linker experiment, the unfolding events occur over a wider range of absolute extension values, whereas unfolding events with ELP linkers occur over a narrow range. (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.

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

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

ACS Nano Article

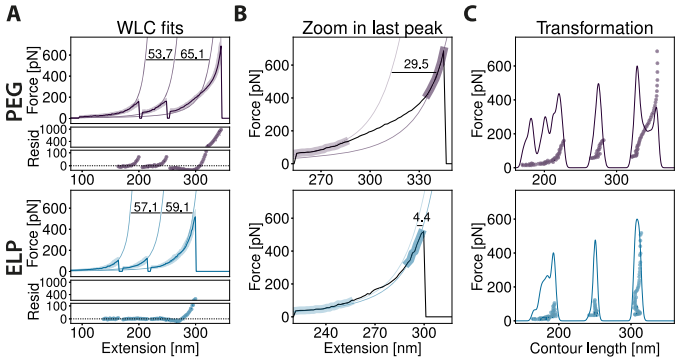


Figure 4. Elasticities of PEG and ELP linkers. (A) Superposition of multiple protein unfolding curves ("master curves") from SMFS experiments with PEG (purple, *N* = 73) and ELP linkers (blue, *N* = 151). The lower plots of each graph in panel A show the residuals of each WLC fit. Note that the residual 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

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

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 $[(VPGVG)_5\text{-}(VPGAG)_2\text{-}(VPGGG)_3]_6$ and is referred to as $ELP_{120\ nm}$.

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-$ELP_{120\ nm}$-LPETGG, 90475: Cys-$ELP_{40\ nm}$-LPETGG, 91571: GGG-$ELP_{40\ nm}$-Cys, 91572: GGG-$ELP_{120\ nm}$-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 *DHS*α 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 (15000g, 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 (15000g, 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 (3220g, 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 (3220g, 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

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

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

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

With the quantum mechanical correction,

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

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

⊥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-

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

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**.

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

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.

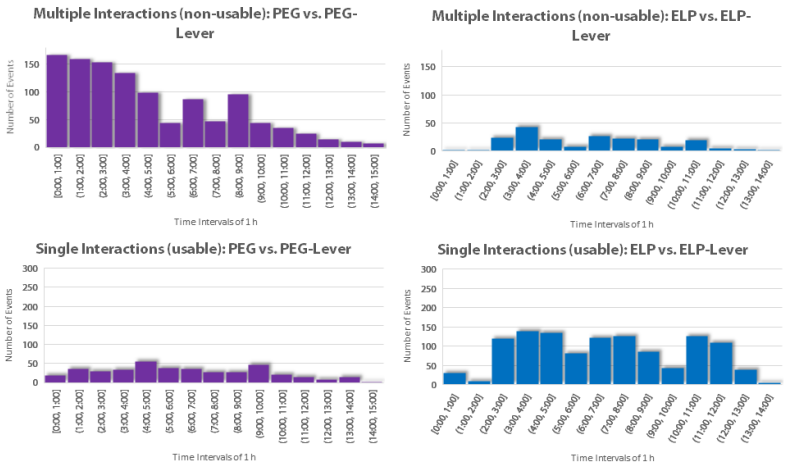
[c] Department of Chemistry, University of Basel, 4056 Basel, Switzerland.

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

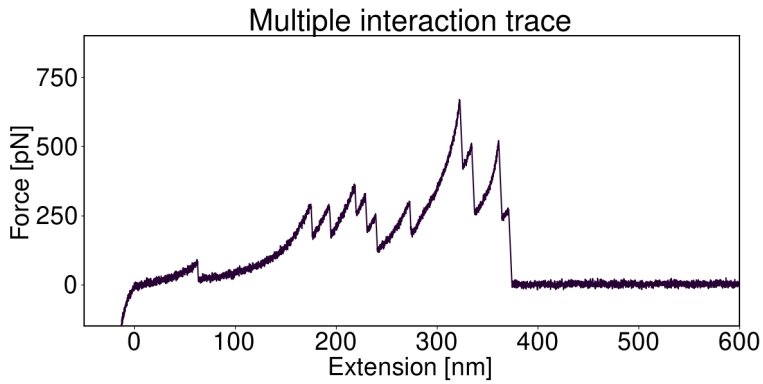
[⊥] These authors contributed equally to this work

[#] Corresponding author: gaub@lmu.de

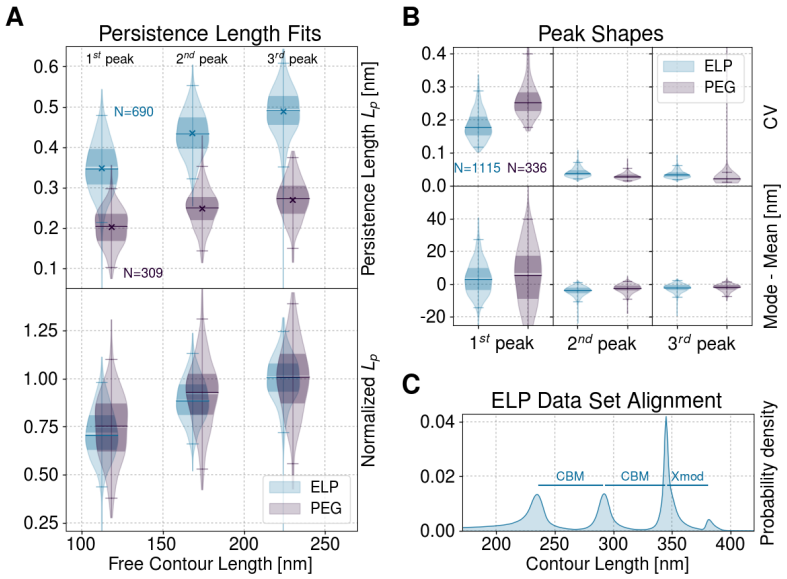
Supporting Information



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.

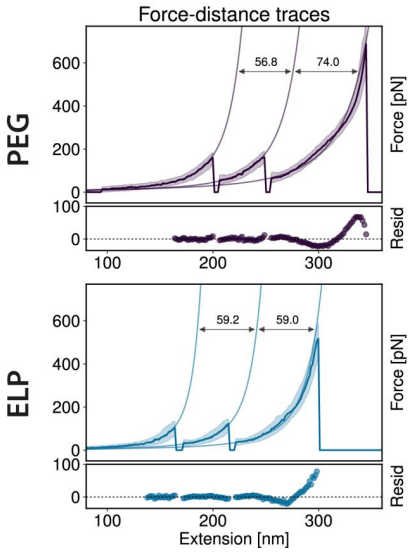


Supplemental Figure S3: Performance of contour length transformations. (A) Observed persistence lengths. Upper plot: observed persistence lengths preceding each CBM and Coh:DocIII unfolding/rupture peak as measured by WLC fits in the force range of 30 to 125 pN (ELP: 0.35, 0.44, and 0.49 nm; PEG: 0.20, 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

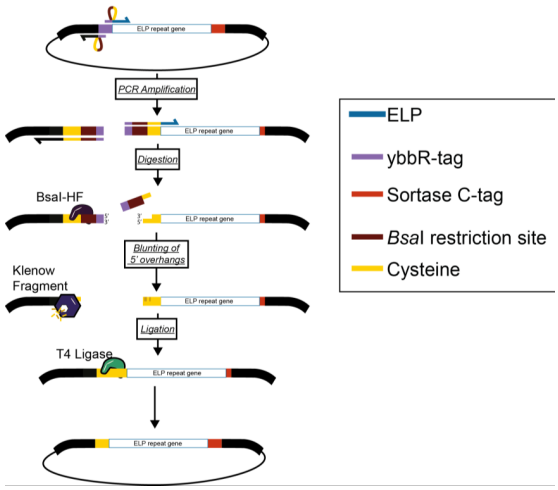
varying persistence lengths over the course of each pulling cycle. The persistence length as a measure for the stiffness of a polymer is lower for PEG than for ELP with bulky side chains and rotational restrictions of the peptide backbone. Comparable changes of persistence lengths over the course of 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.

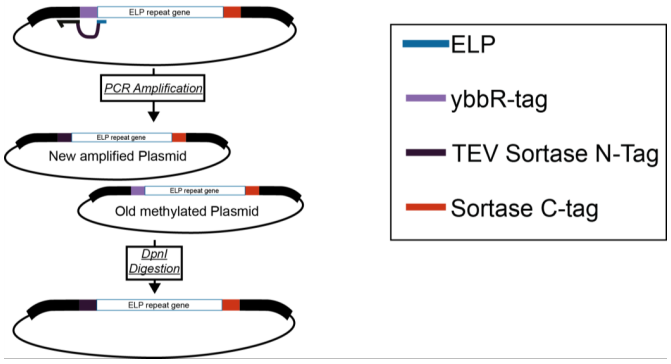
Cloning of ELP linkers. Standard PCR was used for amplification of DNA (Phusion High-Fidelity PCR Master Mix, Thermo Fisher Scientific Inc., Waltham, MA, USA). Melting temperatures were adjusted according to the employed primers (see **Table S1**, below).

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).



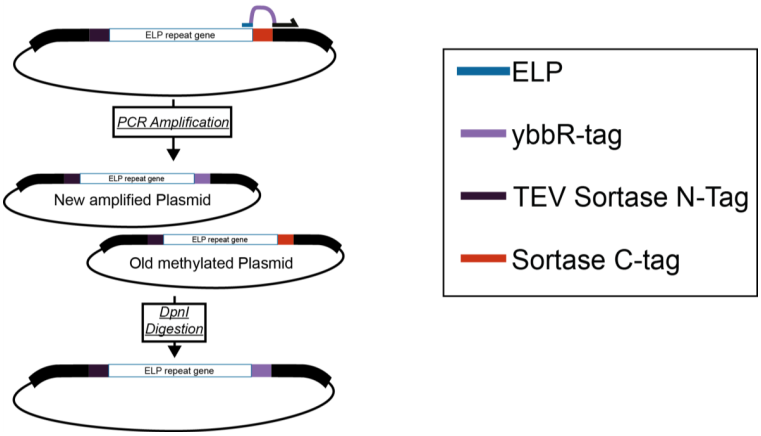
Supplemental Figure S5. Cloning scheme for Cys-ELP$_{120\ nm}$-LPETGG.

For the creation of the TEV-GGG-ELP$_{60\ nm}$-LPETGG plasmid, a plasmid encoding ybbR-ELP$_{60\ nm}$-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**).



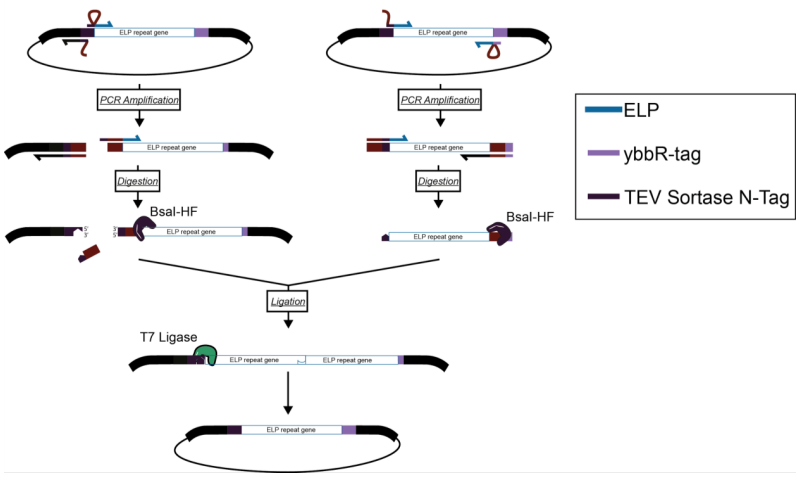
Supplemental Figure S6. Cloning scheme for TEV-GGG-ELP$_{60\ nm}$-LPETGG.

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.

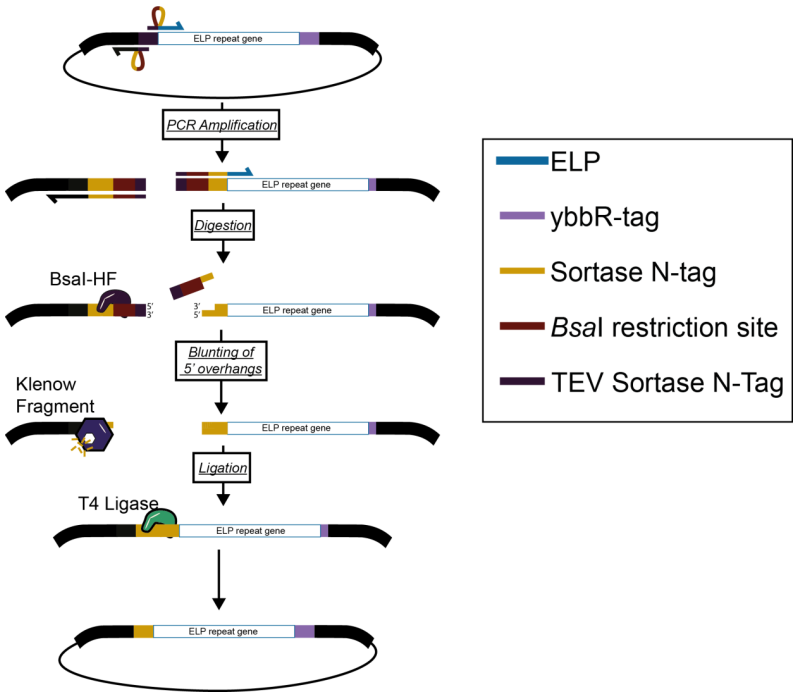
The ELP gene cassette was duplicated by insertion of a gene sequence encoding $[(VPGVG)_5$-$(VPGAG)_2$-$(VPGGG)_3]_3$ into the linearized vector containing TEV-GGG-$ELP_{60\ nm}$-ybbR. This was done by GoldenGate cloning.[5] For this purpose, both vector and insert were amplified with primers encoding 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-$ELP_{120\ nm}$-ybbR.

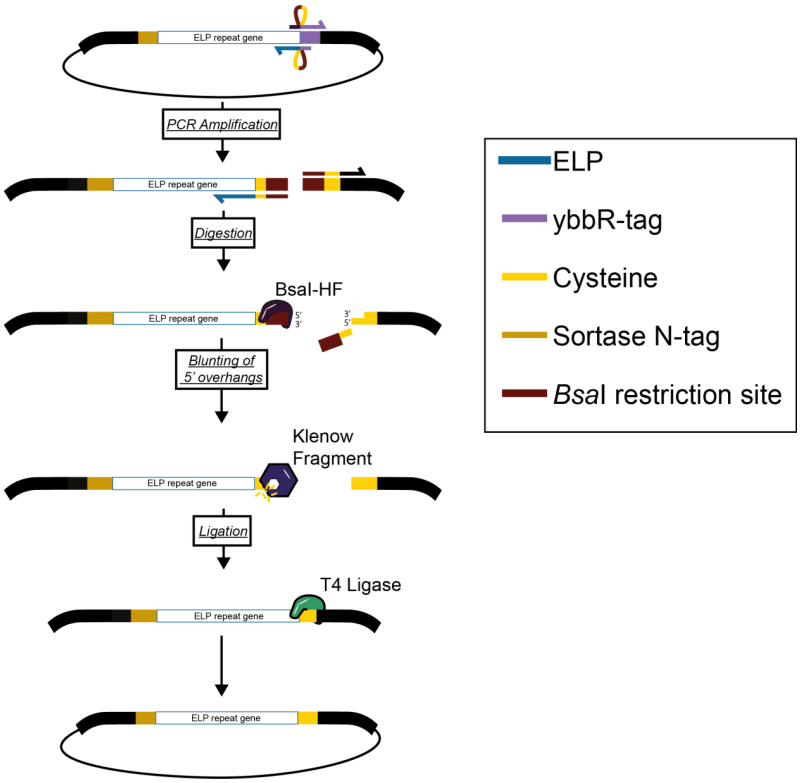
Experiments showed that the *E. coli* methionine aminopeptidases already fully digested the N-terminal methionine proceeding the polyglycine. Hence, removal of the TEV cleavage site was desired to simplify the ELP production process. This was achieved by a linearization reaction, *Bsa*I digestion and 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**).

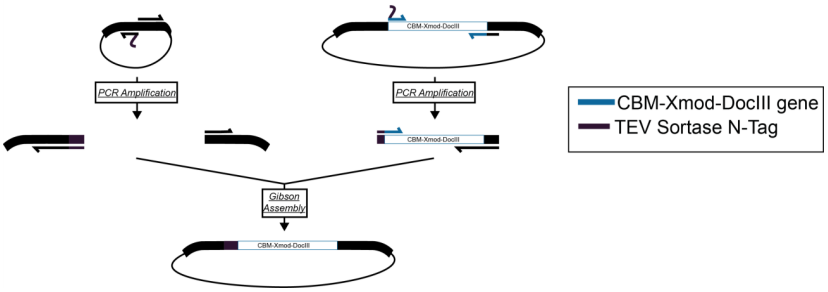


Supplemental Figure S10. Cloning scheme for GGG-ELP$_{120\ nm}$-Cys

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	GACTCTCTGGAATTCATCGCTTCTAAACTGGC TGGTCTCCTGCGTGCCGGGAGAAGGAG
REV BsaI ybbR	CCCGGCACAGCCAGTTTAGAAGCGATGAATTC CAGAGAGTCGGTCTCACATATGTATATC
Construction of TEV-GGG-ELP$_{60\ nm}$-LPETGG	
QuikChange Primer ybbR to TEV-GGG	GACACCAGGGACTCCTTCTCCCGGCACACCG CCCCCTCCCTGGAAGTACAGGTTTTCCATATG TATATCTCCTTC
Construction of TEV-GGG-ELP$_{60\ nm}$-ybbR	
QuikChange Primer LPETGG to ybbR	GACACCAGGGACTCCTTCTCCCGGCACACCG CCCCCTCCCTGGAAGTACAGGTTTTCCATATG TATATCTCCTTC
Construction of TEV-GGG-ELP$_{120\ nm}$-ybbR	
FW backbone BsaI	GAAAACCTGTACTTCCAGGGAGGGGGGTCTC GGGGTGTGCCGGGAGAAGGAG
REV backbone BsaI	ATATATGGTCTCGACCGCCCCCTCCCTGGAAG TACAGGTTTTC
FW insert TEV-GGG BsaI	CCAGGGAGGGGGGTCTCGCGGTGTGCCGGG AGAAGGAG
REV insert BsaI	TCGAGTTAAGCCAGTTTAGAAGCGATGAATTC CAGAGAGTCGGTCTCCACCCTCACCCGG
Construction of GGG-ELP$_{120\ nm}$-ybbR	
FW ELP GGG	GGGGGCGGTGTGCCGGGAG
REV BsaI TEV	GGCACACCGCCCCCTCCCTGGAAGTACAGGT TTTCGGTCTCACATATGTATATCTCCTTC

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$_{120\ nm}$-LPETGG	851370	24763.08	3.20	300 (307)	112.06
GGG-ELP$_{120\ nm}$-Cys	843030	24379.63	3.23	300 (304)	110.96

Protein Sequences

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

Sortase N-Tag
ELP
Cysteine

GGGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGV
PGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVG
VPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGV
GVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPG
AGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVP
GGGVPGEGC

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

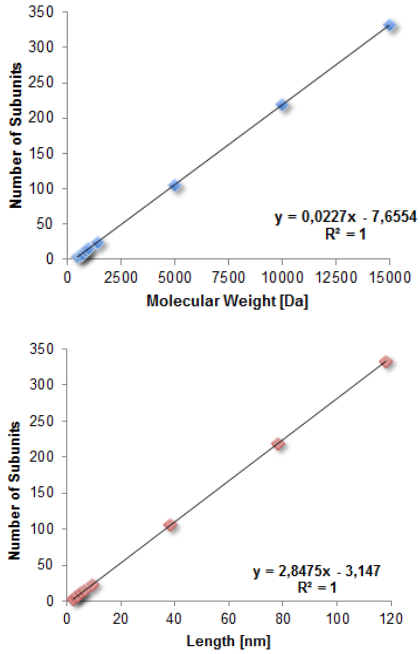
CBM

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.

3.8 Force Activation of Enzymes

This publication investigates the mechano-activation of focal adhesion kinase. In a combined in-vitro and in-silico approach, we show that mechanoactivation can occur at forces below the unfolding forces of FAK. Our findings are consistend with, and suggest that force-induced conformational changes of FAK may induce focal adhesion signaling. My contribution to this work was the construction of a high-speed AFM for SMFS experiments. This enabled experiments at retraction velocities of 12 800 nm s^{-1}, which resulted in loading rates beneficial in elevating unfolding events above the noise level.

Structural and mechanistic insights into mechanoactivation of focal adhesion kinase

Magnus S. Bauer, Fabian Baumann, Csaba Daday, Pilar Redondo,
Ellis Durner, Markus A. Jobst, Lukas F. Milles, Davide Mercadante, Diana A. Pippig
Hermann E. Gaub, Frauke Gräter, and Daniel Lietha

published in

6766–6774 | PNAS | April 2, 2019 | vol. 116 | no. 14

Structural and mechanistic insights into mechanoactivation of focal adhesion kinase

Magnus Sebastian Bauer[a,b], Fabian Baumann[a], Csaba Daday[c,d], Pilar Redondo[e], Ellis Durner[a], Markus Andreas Jobst[a], Lukas Frederik Milles[a], Davide Mercadante[c,d,1], Diana Angela Pippig[a,2], Hermann Eduard Gaub[a,1], Frauke Gräter[c,d,3], and Daniel Lietha[e,3,4]

[a]Lehrstuhl für Angewandte Physik, Nanosystems Initiative Munich and Center for Nanoscience, Ludwig-Maximilians-Universität München, 80799 Munich, Germany; [b]Center for Integrated Protein Science Munich, Ludwig-Maximilians-Universität München, 80799 Munich, Germany; [c]Heidelberg Institute for Theoretical Studies, 69118 Heidelberg, Germany; [d]Interdisciplinary Center for Scientific Computing, Heidelberg University, 69120 Heidelberg, Germany; and [e]Cell Signalling and Adhesion Group, Structural Biology Programme, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain

Edited by Melanie H. Cobb, University of Texas Southwestern Medical Center, Dallas, TX, and approved February 12, 2019 (received for review December 4, 2018)

Focal adhesion kinase (FAK) is a key signaling molecule regulating cell adhesion, migration, and survival. FAK localizes into focal adhesion complexes formed at the cytoplasmic side of cell attachment to the ECM and is activated after force generation via actomyosin fibers attached to this complex. The mechanism of translating mechanical force into a biochemical signal is not understood, and it is not clear whether FAK is activated directly by force or downstream to the force signal. We use experimental and computational single-molecule force spectroscopy to probe the mechanical properties of FAK and examine whether force can trigger activation by inducing conformational changes in FAK. By comparison with an open and active mutant of FAK, we are able to assign mechanoactivation to an initial rupture event in the low-force range. This activation event occurs before FAK unfolding at forces within the native range in focal adhesions. We are also able to assign all subsequent peaks in the force landscape to partial unfolding of FAK modules. We show that binding of ATP stabilizes the kinase domain, thereby altering the unfolding hierarchy. Using all-atom molecular dynamics simulations, we identify intermediates along the unfolding pathway, which provide buffering to allow extension of FAK in focal adhesions without compromising functionality. Our findings strongly support that forces in focal adhesions applied to FAK via known interactions can induce conformational changes, which in turn, trigger focal adhesion signaling.

atomic force microscopy | mechanobiology | focal adhesion signaling | protein kinase regulation | single-molecule force spectroscopy

Focal adhesions (FAs) are dense molecular assemblies that anchor cells via integrin receptors to the ECM and intracellularly connect to actin stress fibers (1). FAs not only form a structural link between the cell and its surroundings but also, are important for exchanging mechanical force cues and regulatory signals (2, 3). A key regulator in FAs is the nonreceptor tyrosine kinase focal adhesion kinase (FAK) that triggers FA signals on cell adhesion to the ECM. Apart from its function as a signaling kinase, it acts as a scaffolding hub for diverse interaction partners in FAs. Via its interactions and embedding in the FA complex, FAK is exposed to forces arising from inside or outside the cell. Cell-based studies show that increased forces exerted on FAs result in activation of FAK (4–6). Moreover, FAK seems to have an important force sensing role, since FAK is required for cells to respond to externally applied forces or to migrate toward stiffer substrates, which allows generation of higher forces in FAs (7). However, current studies lack a clear hint of whether FAK represents a first responder to force or is indirectly force activated by downstream signaling. An activation mechanism based on the direct application of mechanical force on an enzyme was previously described for the mammalian titin kinase, the related twitchin kinase in nematode (8, 9), and the smooth muscle myosin light-chain kinase (10), which are located in the load-bearing environment of the muscle sarcomeres. However, no nonmuscle enzyme was shown to be directly activated by mechanical force yet.

FAK is a multidomain protein that is subdivided into three major domains. The central catalytic kinase domain is flanked by an N-terminal 4.1 protein, Ezrin, Radixin, Moesin homology domain (FERM) and a C-terminal focal adhesion targeting (FAT) domain (Fig. 1*A*). In the basal state, the FERM and kinase domains interact to form a closed and autoinhibited conformation, where the active site and several regulatory phosphorylation sites are sequestered (Fig. 1*B*) (11). On integrin-mediated cell adhesion, the FAT domain targets FAK into FAs. Super-resolution optical microscopy has localized FAK to an

Significance

Nonreceptor tyrosine kinases are major players in cell signaling. Among them, focal adhesion kinase (FAK) is the key integrator of signals from growth factors and cell adhesion. In cancer, FAK is frequently overexpressed, and by promoting adhesion to the tumor stroma and ECM, FAK provides important signals for tumor invasion and metastasis. Although autoinhibitory mechanisms have previously been described, FAK activation in response to force generated by ECM attachment is currently not understood. Here, we use experimental and computational approaches to demonstrate how mechanical forces can induce conformational changes in FAK that result in activation. This mechanistic insight enables the design of alternative strategies for the discovery of potential anticancer drugs that inhibit both catalytic and scaffolding functions of FAK.

Author contributions: M.S.B., F.B., D.A.P., H.E.G., F.G., and D.L. designed research; M.S.B., F.B., C.D., P.R., E.D., M.A.J., L.F.M., D.M., D.A.P., H.E.G., F.G., and D.L. performed research; M.S.B., F.B., C.D., P.R., E.D., M.A.J., L.F.M., D.A.P., and D.L. contributed new reagents/analytic tools; M.S.B., F.B., C.D., P.R., D.A.P., and D.L. analyzed data; and M.S.B., C.D., H.E.G., F.G., and D.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Both the code and the datasets used in this study can be accessed at https://gitlab.physik.uni-muenchen.de/Magnus.Bauer/fak_analysis.

[1]Present address: Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

[2]Present address: Roche Diagnostics GmbH, 82377 Penzberg, Germany.

[3]To whom correspondence may be addressed. Email: gaub@lmu.de, frauke.graeter@h-its.org, or daniel.lietha@cib.csic.es.

[4]Present address: Cell Signalling and Adhesion Group, Structural and Chemical Biology, Centro de Investigaciones Biológicas (CIB), Spanish National Research Council (CSIC), 28040 Madrid, Spain.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820567116/-/DCSupplemental.

Published online March 15, 2019.

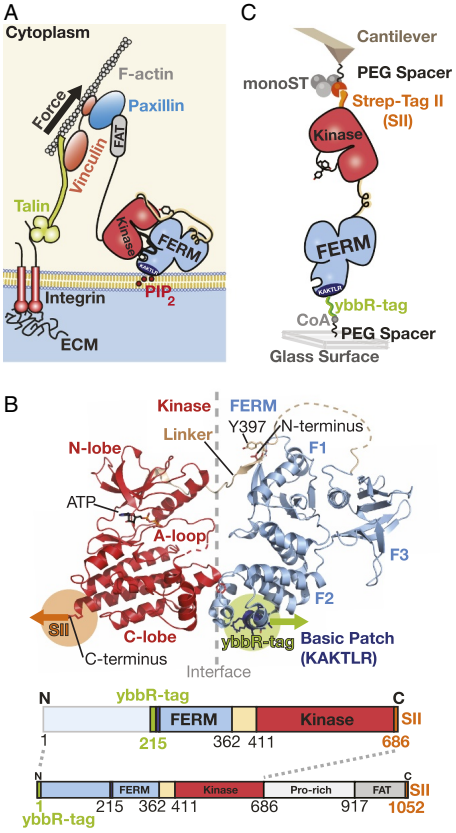


Fig. 1. Mimicking the attachment of FAK in FAs in the AFM force spectroscopy assay. (*A*) Schematic depiction of an FA connecting the ECM with the actin cytoskeleton. FAK is anchored to PIP2 in the lipid membrane via its N-terminal basic patch (amino acid sequence: "KAKTLR"; dark blue). On the other end, FAK is attached with its C-terminal FAT (gray) to paxillin. (*B*) The structure of autoinhibited FAK (Protein Data Bank ID codes 2J0J and 2IJM) is shown with placement of the affinity tags for AFM. The FERM domain directly attaches to the kinase domain and blocks the active site and phosphorylation of the activation loop (A loop). On the bottom, the schematic sequence of FAK used in the experiments (FK-FAK, residues 1–686) is shown with the placement of the 11-aa ybbR-tag in green for site-specific immobilization to CoA. Additionally, the SII in orange and the lipid binding basic patch in dark blue are depicted. The amino acid sequence from residues 1 to 215 is not stressed directly by force during the experiment and is, therefore, colored in light blue. The full-length FAK (1–1,052) with the C-terminal region lacking in FK-FAK was used for control measurements (*SI Appendix*, Fig. S1) and is depicted here to show the complete FAK protein. (*C*) The scheme illustrates the measurement setup with immobilized FAK on a PEG functionalized glass surface and monoST on the cantilever. The ybbR-tag ensures site-specific covalent attachment to the surface via an Sfp-catalyzed reaction with CoA connected to a PEG spacer. At the C terminus, FK-FAK is equipped with an SII for reversible tethering to monoST on the cantilever.

integrin signaling layer in FAs in close proximity to the plasma membrane (12). Accordingly, the FERM domain contains a stretch of basic residues forming a basic patch that interacts with phosphatidylinositol 4,5-bisphosphate (PIP2) (13, 14), which is clustered in the cell membrane at FA sites (15, 16). Activation of FAK can be initiated by an orchestrated process starting with PIP2 binding to the basic patch, resulting in exposure of a linker region containing the autophosphorylation site (14). After autophosphorylation, this site becomes a docking site for the Src kinase, which in turn, phosphorylates tyrosine residues in the activation loop of the FAK kinase. This results in full activation of FAK. The last step of Src phosphorylation is strongly promoted by prior separation of FERM and kinase domains (11). Here, we test the hypothesis that force is a key stimulus driving domain separation for FAK activation. With the N-terminal domain docked to the cell membrane and the C-terminal FAT domain tethered via paxillin and structural FA components to the actin cytoskeleton, FAK is likely exposed to stretching forces in FAs. Force-induced domain opening and activation were indeed already proposed and supported by molecular dynamics (MD) simulations, showing that forces applied to FAK result in rupture of autoinhibitory interactions and exposure of Src phosphorylation sites in the FAK activation loop (17). Furthermore and consistent with this hypothesis, FAK is known to undergo domain opening when localized to FAs in cells (13). Additionally, mutational disruption of autoinhibitory interactions and resulting FAK opening has experimentally been shown to strongly promote phosphorylation by Src (11).

In this study, we aim to obtain insight into the force response of FAK on a single-molecule level by means of atomic force microscopy (AFM)-based force spectroscopy in conjunction with force probe molecular dynamics (FPMD) simulations. We are able to measure highly reproducible force profiles recorded during FAK stretching and map-detected force events to structural features in FAK. By adapting experimental conditions, we are able to identify interface rupture of FERM and kinase domains as a discrete force peak. This is verified by control force profiles of FAK mutants lacking autoinhibitory FERM–kinase interactions and comparison with MD simulations. We conclude that the inhibitory interface ruptures at forces of around 25 pN for pulling speeds of 12,800 nm/s (or at smaller forces for slower pulling)—significantly before any domain unfolding—and that functionally important regions in FAK start to unfold at an extension of about 50 nm. Hence, we demonstrate that force activation of FAK can occur at physiological forces in FAs and that FAK maintains an active state when exposed to extensions expected to occur in FAs.

Results

Relative Mechanical Stabilities of FERM and Kinase Are ATP Dependent. In this study, we record force–distance traces for FAK containing FERM and kinase domain (residues 1–686; referred to hereafter as FK-FAK) (Fig. 1*C*) attached via short PEG linkers (425.39 Da) and pulling speeds of 800 nm/s. The curves show force peaks between 20 and 50 pN, with reliably recurring unfolding features and a final Strep-Tag II (SII) rupture at around 70 pN (Fig. 2). Force application guides the protein through a sequence of conformational states obeying a strict hierarchy as suggested by the heatmap, and the most probable unfolding curve in black (Fig. 2). All detected peaks above the FWHM of the noise level are highlighted with a colored triangle (the process is described in *Data Analysis*) (Fig. 2). These unfolding peaks indicate transitions between these conformational states and mark the unfolding of stable structural modules within the protein structure leading to a release of previously hidden contour lengths. To determine contour length increments of the polypeptide chain unfolded for each peak, we fitted the data before each rupture event with the worm-like chain model (18, 19).

BIOCHEMISTRY

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

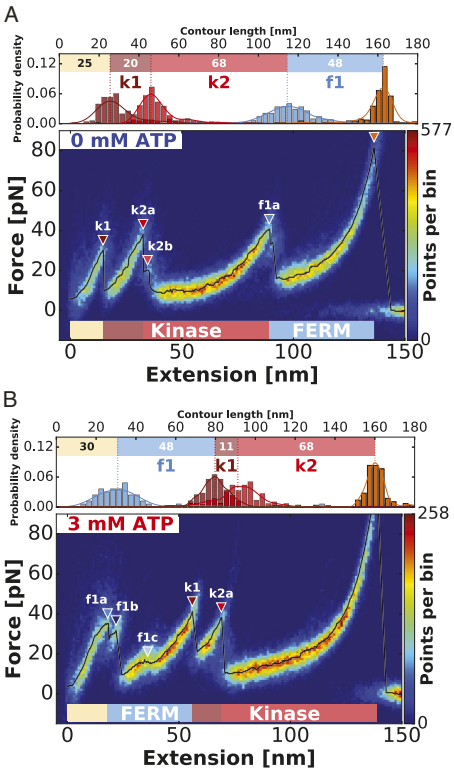


Fig. 2. Assignment of force peaks to structural modules of FERM and kinase domain (residues 1–686; FK-FAK) using AFM unfolding patterns. (*A*) The heatmap obtained from an overlay of 224 curves shows the typical unfolding pattern of FK-FAK in the absence of ATP, revealing low-force unfolding below 50 pN and final SII rupture at around 70 pN. In *Upper*, the increments between identified peaks are depicted, allowing assignment of the rupture events to structural modules in the crystal structure. (*B*) Depiction of the FK-FAK unfolding pattern by an overlay of 115 curves in the presence of 3 mM ATP. Both plots are created from one dataset recorded with the same cantilever, and therefore, they are directly comparable in absolute force. Although the unfolding pattern in *B* looks different from the one in *A*, the increments stay conserved. Comparison of the two conditions shows that the increment of f1 is swapping its position with k1 and k2 on ATP binding. This can be attributed to stabilization of the kinase domain on addition of ATP, shifting peak k1 to higher forces and thereby, changing the force hierarchy between FERM and kinase. (*A* and *B*) The black lines show the most probable unfolding patterns with all detected peaks above the FWHM of the noise level (described in *Data Analysis*) highlighted with colored triangles according to their assigned domain.

We determine contour length increments of 20 nm (k1), 68 nm (k2), and 48 nm (f1) for FK-FAK (Fig. 2*A* and Table 1).

Additionally, FK-FAK was stretched in presence of 3 mM ATP to probe the effect of ATP binding to the kinase domain. Although the presence of ATP substantially changes the unfolding pattern of FK-FAK (as observed in Fig. 2), the identified increments remain remarkably conserved [48 nm ($f1^{ATP}$), 11 nm ($k1^{ATP}$), and 68 nm ($k2^{ATP}$)] (Fig. 2*B* and Table 1). This suggests that the unfolding hierarchy changes on ATP addition as increment f1 moves from the back to the beginning of the curve (Fig. 2). The most probable rupture force of peak k1 (Fig. 2) significantly increases on ATP binding, corroborating its association with the kinase domain (20, 21). This is probably due to a stabilization of the part of the kinase domain involved in ATP binding, thereby yielding higher forces for peak k1 and consequently, changing the force hierarchy between FERM and kinase domain. The kinase domain maintains the internal sequence of unfolding (k1 then k2a) but unfolds after the FERM domain in the presence of ATP. Their sums of all contour lengths (88 and 79 nm, respectively) agree with expected values for the FAK kinase domain [91 nm for residues 422–686 (19, 22), subtracting 5 nm to account for the initial distance between residues 422 and 686 in the folded domains according to the crystal structure]; hence, we assign k1 and k2 to the kinase domain. The increment f1 is assigned to FERM unfolding, and a contour length of 48 nm reproduces an expected length of 50 nm for folded FERM from residues 216–362 plus an additional 12 residues for the inserted ybbR-tag, subtracting an initial folded distance of 6 nm. We, therefore, conclude that stabilizing the kinase domain by ATP binding results in a reversed unfolding sequence, with the kinase unfolding first in absence of ATP but as the last event in the presence of ATP. All determined contour length increments are in good agreement with defined structural features in the crystal structure of FK-FAK in Fig. 1*B* (11).

To further validate our assignment of unfolding increments, we used single-domain constructs with only the isolated FERM or kinase domains. The curves were analyzed as previously described (Figs. 3 and 4). The data confirm our assignment of the structural modules, with determined contour length increments matching well with the ones assigned in FK-FAK plots (Table 1). In agreement with our assignment in FK-FAK, the isolated kinase likewise unfolds in two main modules (k1 and k2a). Furthermore, we identify additional unfolding intermediates that were previously hidden in the noise. For the FERM domain, we now resolve three peaks with contour lengths 7 nm (f1a), 25 nm (f1b), and 14 nm (f1c). For the kinase domain, k2a is followed by another peak k2b, but it is too short to determine a proper contour length increment. Close inspection reveals that these intermediate states (k2b, f1b, f1c) are in fact also present in FK-FAK (Fig. 2). They can be detected in the beginning of the unfolding curve where the total free length is still short enough, leading to higher loading rates (as property of the worm-like chain model) and consequently, to higher forces, thereby allowing us to resolve more subtle peaks. These subtle peaks are, however, too small (and their extension is too short) to determine their contour length accurately. All determined contour length increments are summarized in Table 1.

Table 1. All measured contour length increments of the various FAK constructs used in this study

Protein segment	f1	k1	k2
FK-FAK −ATP	48	20	68
FK-FAK +ATP	48	11	68
FERM −ATP	46		
f1a/f1b/f1c	7/25/14		
FERM +ATP	46		
f1a/f1b/f1c	7/21/18		
Kinase −ATP		15	66
Kinase +ATP		14	66

The contour length increment values are given in nanometers and were determined as most probable values from a KDE.

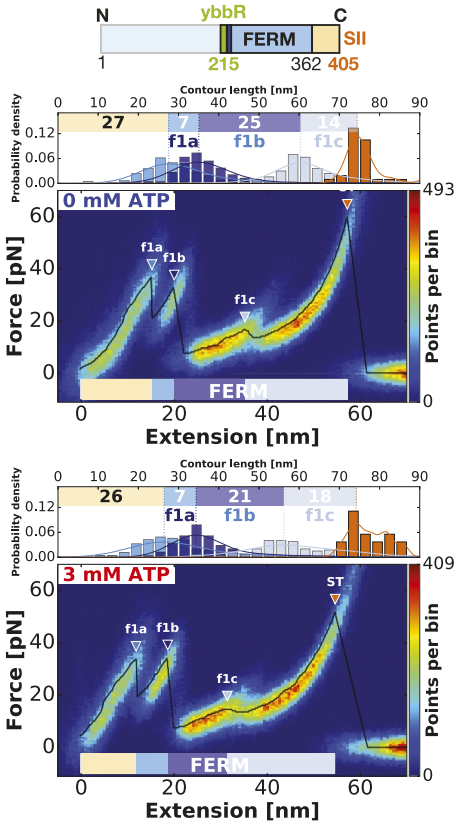


Fig. 3. Verification of the assignments of structural modules by probing single FERM (residues 1–405) domains. The heatmaps show single-FERM unfolding events without (*Upper*) and in the presence of 3 mM ATP (*Lower*). The increments and forces do not change in the presence of ATP as expected. The same peaks detected in detail here (f1b and f1c) can also be found in the FERM domain in Fig. 2*B*. Due to longer linkers at the end of the unfolding curve and therefore, lower loading rates leading to lower forces, they cannot be detected in Fig. 2*A*.

Rupture of FERM–Kinase Interface Precedes Domain Unfolding. The single-domain constructs demonstrated that subtle peaks can be hidden within prominent peaks. The interface opening between kinase and FERM likely represents such a small peak that is expected to release a contour length of only 20 nm. Assuming physiological activation under stress in vivo, the opening has to occur at relatively small forces and before any other unfolding, since it is shielding the remaining structure. However, since in this region of the force profiles, we do not detect an additional distinct force peak in FK-FAK profiles with a pulling speed of 800 nm/s, we conclude that domain separation may occur anywhere before the first unfolding event (i.e., below 40 pN in this loading rate regime) but likely is hidden in the measurement noise.

Aiming to resolve domain separation, we used higher pulling speeds of 12,800 nm/s to increase force responses and used long PEG (5,000 Da) for FAK attachment to facilitate analysis early in the force profile. This indeed enabled us to resolve an additional force peak at the beginning of the unfolding pattern (yellow triangles in Fig. 5*A*), where we expect domain separation to occur. To verify that this peak originates from the domain opening, we probed a mutant of FK-FAK (FK-FAKmut) with point mutations at the domain interface (Y180 A, M183 A). These mutations cause FERM and kinase domains to be permanently dissociated (14). As shown in Fig. 5*B*, this mutant lacks the low-force rupture event, which we identified as the interface opening in the FK-FAK wild-type profile

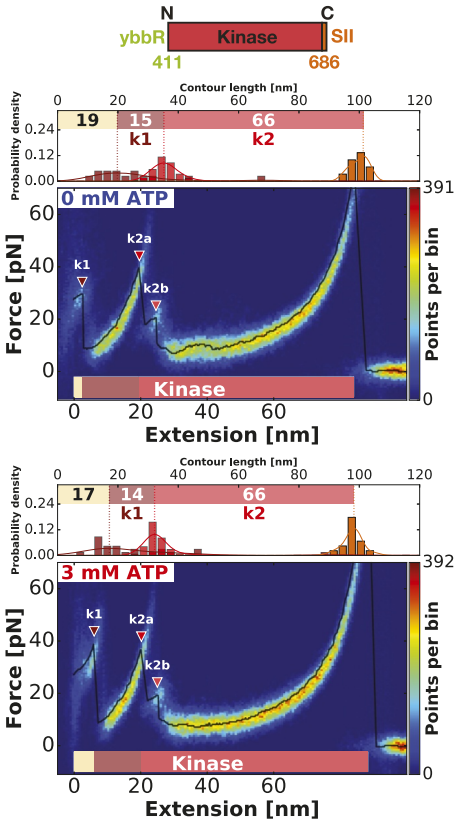


Fig. 4. Verification of the assignments of structural modules by probing single-kinase (residues 411–686) domains. The heatmaps show single-kinase unfolding events without (*Upper*) and in the presence of 3 mM ATP (*Lower*). The first peak k1 increases in force by addition of ATP, suggesting a binding event of ATP to the kinase domain. The same peak detected in detail here (k2b) can also be found in the kinase domain in Fig. 2*A*. Due to longer linkers at the end of the unfolding curve and therefore, lower loading rates leading to lower forces, they cannot be detected in Fig. 2*B*.

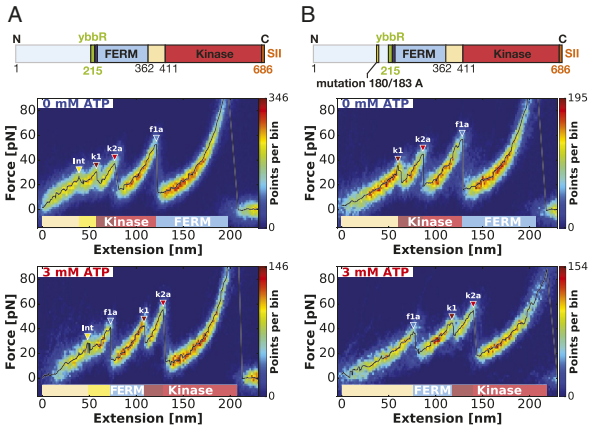


Fig. 5. Resolving the FERM–kinase domain interface rupture by comparing FK-FAK with a permanently open FK-FAKmut—recorded with long PEG (5,000-Da) linkers at a pulling speed of 12,800 nm/s. (*A*) The heatmap shows the same construct as measured in Fig. 2 but with improved force resolution in the investigated region. This way, an additional subtle peak in the beginning of the curve can be observed highlighted by yellow triangles. This peak is not detected in the permanently open FK-FAKmut construct measured in *B*. We conclude that this rupture event corresponds to the interface opening between FERM and kinase domain. With measurements in the presence of 3 mM ATP (*Bottom*), the same behavior is observed, hence showing no association of the rupture event with a protein domain and validating its assignment to interface opening. *SI Appendix*, Fig. S2 provides additional contour length histograms for the peaks detected in this graph.

(yellow triangles and Int label in Fig. 5*A*), but still retains all other unfolding features. As expected, domain dissociation in the FK-FAK wild type precedes any unfolding events. In conclusion, these data indicate that forced domain dissociation and thereby, conformational activation happens before any other subdomain unfolding at a force around 25 pN at pulling speeds of 12,800 nm/s.

MD Simulations Confirm Conformational Activation of FAK. Fig. 6 summarizes our unfolding simulations (detailed unfolding data are given in *SI Appendix*, Figs. S3–S11). Consistent with previously simulated results (17), the first event is invariably dissociation of the FERM–kinase interface. After this, there is considerable heterogeneity in our data, with most of the simulations featuring both domains being in the process of unfolding. However, it should be considered that pulling speeds in simulations are significantly faster than in AFM experiments, and Fig. 6*B* shows a clear trend of slower simulations featuring the experimentally observed hierarchy: the FERM domain unfolds before the kinase. We, therefore, restrict the following analysis to these cases (five at 0.1 m/s and two at 0.33 m/s).

The first force-induced conformational change is the loss of the FERM–kinase interface, giving 10 nm of extension (Fig. 6 and *SI Appendix*, Table S1). This is followed by another 12-nm extension due to the linker losing contact with the F1 lobe in the FERM domain (Fig. 6*B*). In agreement with simulations, domain separation of FAK was also experimentally observed as the first event (Fig. 5), and FERM–linker separation, which simulations show to require less force (*SI Appendix*, Fig. S11) and to occur shortly after domain separation, is not detected in experiments.

The FERM domain unfolding was observed experimentally to occur in three steps: one at 7–8 nm, a second one around 21–25 nm, and a third one around 14–19 nm. In our MD simulations, the FERM domain unfolding happens in two stages: first, lobe F3 unfolding is observed, amounting to 30 nm, and second, lobe F2 stretching is observed, corresponding to an increase of 9 nm. Since the F2 stretching never happens before F3 unfolding in our trajectories, we assign the first two events observed experimentally to F3 unfolding. The F2 stretching contributes to an increase in end-to-end distance of the protein by about 9 nm, which in experiments, amounts to around 13 nm considering the additional loop contained in the handle of the experimental FK-FAK construct. Kinase domain unfolding leads to FAK deactivation independent of the detailed sequence of events. We, therefore, discuss kinase unfolding events of the kinase subdomains observed in MD pulling simulations vs. AFM in *SI Appendix*. Importantly, the same unfolding sequences of various lobes of the FERM and kinase domains were observed in simulations performed on the subdomains only, further validating the experiments on individual domains. *SI Appendix* has details.

Discussion

Here, we report a detailed mechanical characterization of FAK by using an AFM setup to apply stretching forces on single FAK molecules and record force–extension profiles with high sensitivity. Combining our measurements with structural information of the FERM–kinase region of FAK (11) and FPMD simulations has allowed us to assign measured force peaks to unfolding of defined structural features in FAK. Importantly, increasing the force loading rate by applying high pulling speeds has enabled the detection of a low-force event corresponding to rupture of the autoinhibitory FERM–kinase interaction. We show that domain separation occurs at a low-force regime around 25 pN for pulling speeds of 12,800 nm/s. The fact that forces required for domain separation are much lower than those required for domain unfolding supports the hypothesis that tensile forces in FAs applied to N- and C-terminal regions in FAK can trigger activation via domain separation. Mutational dissociation of FERM and kinase domains has previously been shown to activate FAK (11).

PNAS

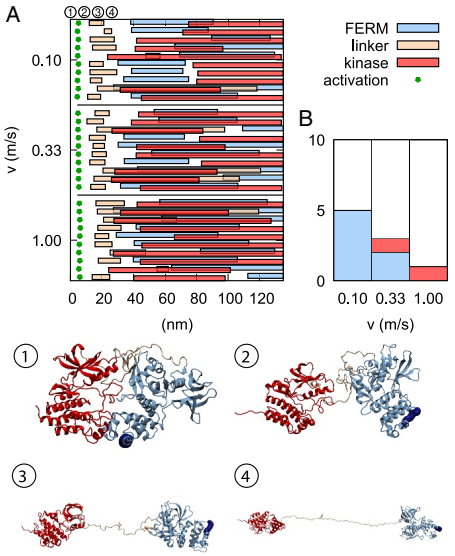


Fig. 6. Domain dissociation and linker detachment precede domain unfolding. (*A*) Rupture events across the 30 simulations are depicted. The green pentagons show the moment of activation (i.e., FERM–kinase dissociation; measured by a sudden increase in V215–K630 distance), and the three different types of bars show when unfolding of these elements begins and ends (monitored by end-to-end distance changes). (*B*) Preferred unfolding mechanism as a function of pulling velocity: FERM unfolds first before the kinase (blue), kinase unfolds first before the kinase (red), or a "mixed" unfolding pathway (white). The bars show proportions of the 10 simulations performed at that pulling velocity. Snapshots 1–4 show the process of conformational activation and linker release observed in MD simulations as illustrated by a sample trajectory at 0.1 m/s. In the native state (snapshot 1), the FERM domain is in contact with the kinase. After an initial kinase C-terminal helix unwinding (*SI Appendix*, Fig. S6), these two domains lose contact (snapshot 2). The interdomain linker gradually detaches and elongates, leading to a release of Tyr397 (snapshot 3). After an additional extension of 12 nm, the linker is fully stretched (snapshot 4). The approximate positions in time of these snapshots are shown in *A*.

We propose that tension forces in FAK are built up between the membrane-bound N-terminal FERM domain and the C-terminal FAT domain engaged via paxillin to vinculin and the actin cytoskeleton (Fig. 7). This scenario was also suggested by our previous MD simulations of the first conformational activation step (17), where FERM and kinase domains detached from one another via these attachment sites. These previous simulations explicitly included a PIP2-containing bilayer and indicated that, for membranes enriched in PIP2, such as is the case at FA sites (15, 16), the membrane–FERM interaction is mechanically significantly more stable than the autoinhibitory FERM–kinase interaction. Force applied at the FAK C terminus in the opposite direction of the membrane resulted in FERM–kinase separation independent of the pulling angle or loading rate (17). Consistent with AFM measurements and the MD simulations presented here, the previous membrane-containing simulations also suggest that domain separation occurs before domain unfolding. In the previous simulations as well as in the full unfolding simulations shown here, the force required for separating FERM and kinase domains is 150 pN; however, the much higher pulling speed in simulations (6 mm/s or higher) is known to cause overestimation of rupture forces. Our combined AFM experiments and MD simulations suggest that the FAK domain organization protects against unfolding of functional domains of FAK: that is, the kinase domain and the PIP2 binding site in the F2 lobe of FERM. First, FERM–kinase domain dissociation yielding around 10-nm extension is followed by around 12-nm-long extension (*SI Appendix*, Table S1, Linker-F1) due to the stretching of the interdomain linker including the Tyr397 phosphorylation site. This would indicate that the length of the linker further protects the kinase domain from forced unfolding after the dissociation of the FERM and kinase domains. Second, F3 unfolding involves an increase in extension of about 30 nm (*SI Appendix*, Table S1, F3 unfold) and happens when the F2 lobe is still capable of binding PIP2. All in all, this would mean that FAK has a "safety margin" up to a total length of 50 nm (10-nm domain dissociation, 12 nm of linker stretching plus at least 28 nm for F3 unfolding) where it is still catalytically active and also, capable of binding PIP2. Consistent with our model where force to the C terminus of FAK is applied via paxillin and vinculin (Fig. 7), vinculin is found to transition from a signaling layer close to the membrane in FAs (which also contains FAK) to a force transduction layer closer to actin (23). Both layers have an approximate thickness of 30 nm measured vertical to the membrane; therefore, the average movement of vinculin approximates 30 nm toward the force transduction layer. This suggests that the 50-nm safety margin appears sufficient to protect the average engaged FAK molecules from force-induced deactivation. The fraction of FAK molecules exceeding this margin would expect to unfold their F2 lobe, hence losing contact with the membrane but retaining an active kinase. At an average of 30-nm extension, our data suggest that FAK molecules, after they are extended, no longer experience significant stretching forces other than what is required to keep FAK in an extended conformation. Likely, forces generated in FAs are mainly carried by structural components, such as talin and vinculin. In contrast for FAK, force seems to act as an activation catalyst by operating a digital distance switch, which is "on" when forces stretch FAK into an open conformation.

In contrast to force activation of FA signaling, force-induced changes on structural FA components, such as talin or vinculin, have been characterized in detail. Mean forces experienced by talin and vinculin in FAs in cells have been determined to be in the range of 7–10 pN and around 2.5 pN, respectively (24, 25). However, force estimates from bulk measurements can be deceiving, since the load could be carried by only a fraction of molecules, whereas many others might not be engaged. Indeed, for integrins, average forces have initially been estimated at 1–2 pN (10); however, DNA-based tensions sensors that are irreversibly ruptured above a threshold force indicate peak forces of up to 40 pN for single integrin molecules (26). In our experiments, FERM–kinase separation occurs at around 25 pN at a pulling speed of 12,800 nm/s. Considering that average cellular force application in FAs is likely slower or even constant over certain time periods, it is highly likely that engaged FA molecules build up sufficient force (at a maximum of 40 pN for integrins) to separate FERM and kinase domains in FAK. Intriguingly, we find that unfolding of the FAK kinase domain in the more physiological ATP-loaded state occurs at around 50 pN; therefore, cellular forces in FAs seem well suited to allow FERM–kinase separation but not kinase unfolding, which are both prerequisites for FAK activation.

Force-induced separation of FERM and kinase domains will expose both the autophosphorylation site in the linker and the Src phosphorylation sites in the kinase activation loop. Mechanical extension of the Tyr397-containing linker might enhance Tyr397 exposure and autophosphorylation. This would likely

BIOCHEMISTRY
BIOPHYSICS AND COMPUTATIONAL BIOLOGY

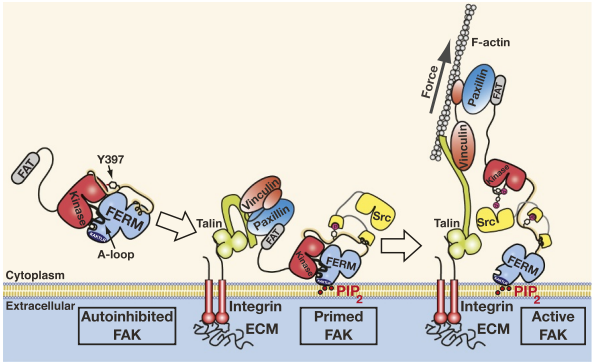


Fig. 7. Model of force-induced FAK activation. FAK is recruited into FAs via C-terminal FAT interactions with paxillin and talin. The N-terminal FERM domain docks via PIP2 to the lipid membrane to promote a primed FAK state where Tyr397 in the linker between FERM and kinase is autophosphorylated. Forces generated via the actin cytoskeleton pull FAK's C terminus away from the membrane, resulting in kinase release from the FERM domain and membrane. Src is recruited to autophosphorylated FAK and phosphorylates the exposed FAK activation loop to trigger full FAK activity.

only be the case in transphosphorylation mode, as force would act against folding back of Tyr397 into the active site of the same FAK molecule. However, we showed previously that membrane binding and resulting FAK oligomerization are sufficient to promote highly efficient FAK autophosphorylation also in the absence of force (14). This study also indicated significant membrane-induced conformational changes that expose the autophosphorylation site but apparently not the kinase active site, since contrary to domain separation, membrane binding did not catalytically activate FAK (14). Together, the two studies, therefore, support a model where initial membrane binding promotes a primed state of FAK by exposing the linker for efficient autophosphorylation, but subsequent buildup of tensile forces in FAK exposes the active site for efficient phosphorylation of the activation loop by Src (Fig. 7). It is the latter event that promotes full catalytic activity of FAK. It was recently shown that the kinase domain of FAK also contributes to binding to PIP2 membranes (27); therefore, force might be responsible for removing the kinase from both the FERM domain and the membrane. MD simulations indeed support such a scenario and found that the pulling angle can dictate which occurs first (17).

In conclusion, our mechanical analysis of FAK supports a model where physiological stretching forces in FAs can cause conformational changes in FAK, promoting its catalytic activation and thereby, triggering of FA signals. Multiple cellular studies have previously shown that FAK is activated in response to various mechanical stimuli (4–6), and our analysis on single FAK molecules demonstrates the feasibility of direct force activation of FAK. Force-induced activation of FA signals is highly relevant in disease. In tumors, stiffening of the stroma that allows increased force generation triggers strong adhesion signals that promote tumor invasion (28). Understanding the direct relation between tumor stiffness, force-induced adhesion signaling, and tumor invasion can, therefore, provide the basis for the development of specific agents targeting this mechanism.

Materials and Methods

AFM Setup for Characterizing FAK. To mimic physiological FAK stretching as occurring in FAs and to identify force-induced structural changes in FAK under stress, we developed an AFM-based single-molecule force spectroscopy assay. This allows for the detection of subtle force-induced events for FAK with high sensitivity during its guided stretching. We engineered FAK proteins to harbor affinity pulling handles for attachment to sample surface and cantilever. We introduced a ybbR-tag (29) for covalent linkage to the glass surface and an SII (30) for reversible tethering to an AFM cantilever tip functionalized with a monovalent Strep-Tactin (monoST) (26). Both attachments are formed via heterobifunctional PEG linkers (Fig. 1*C* has a schematic of FAK attachment to AFM). Before the experiment, FAK proteins are covalently immobilized to the glass surface. Typically, several thousand single-molecule AFM measurements are then performed, and force extension profiles are recorded by repeatedly approaching and retracting the functionalized AFM cantilever at constant speed. The measured curves are aligned and overlaid to generate heatmaps highlighting recurring features in the plots. These recurring unfolding events were identified by creating a most probable unfolding curve as described in *Data Analysis*.

In a previous study, the autoregulatory region of FAK was defined as FERM interacting with the kinase domain (14). Initial AFM experiments were conducted with full-length FAK (residues 1–1,052 in Fig. 1*B*, *Bottom*) and a construct containing only the FERM and kinase domains (residues 1–686, FK-FAK in Fig. 1*B*, *Bottom*), both equipped with affinity tags at their N and C termini. These experiments indicated that the FAT domain does not contribute to the force profile of the autoregulatory region (*SI Appendix*, Fig. S1); hence, subsequent experiments were only performed with FK-FAK. To mimic the physiological force path through the molecule, we introduced the N-terminal tag close to the lipid binding site, which in vivo attaches to the cell membrane. To prevent perturbation in protein folding, we inserted the 11-residue ybbR-tag into an unstructured loop immediately before the K216AKTLRK PIP2 binding site in the FERM domain. We confirm that these insertion mutants retain basal activity of wild-type FAK and that the FERM domain still maintains the ability to autoinhibit the catalytic activity of FAK (*SI Appendix*, Fig. S13). Previous MD simulations confirmed the PIP2–FERM linkage to be significantly more mechanically robust than the FERM–kinase interaction at relevant PIP2 concentrations (17), rationalizing the choice of a covalent handle to mimic of the FERM–membrane interaction.

FAK Expression. Chicken FAK constructs were engineered to contain the 11-aa ybbR-tag after V215, just before the K216-AKTLRK basic patch sequence (29), and the 8-aa SII tag (30) at the C terminus. FAK constructs containing full-length, FERM and kinase, or kinase-only regions were expressed by transient transfection of HEK293GnT1 cells using polyethyleneimine as a transfection agent (31). FERM-only constructs were expressed in *Escherichia coli* BL21 (DE3) as in ref. 32. All proteins were expressed with an N-terminal 6xHis tag. Initial purification was performed by Ni-chelate affinity purification (GE Healthcare) followed by protease cleavage to remove the 6xHis tag. Proteins containing an SII tag were further purified by

Strep-Tactin (GE Healthcare) affinity and size exclusion (Superdex 200; GE Healthcare) chromatography. Proteins without SII tag were further purified by anion exchange (Source 15Q; GE Healthcare) and size exclusion chromatography.

Sample Preparation for Surface and Cantilevers. The preparation of the experiment includes specifically immobilizing (29, 33) the FAK construct on the glass surface and functionalizing the cantilever with an monoST. This ensures a well-defined pulling geometry for minimizing multiple interactions. All FAK constructs harbored a ybbR-tag for covalent immobilization on a glass surface and an SII for binding to the monoST-functionalized cantilever (Fig. 1*C*). All measured constructs were derived from chicken FAK and expressed in HEK cells (compare with *FAK Expression*).

Both cantilevers and glass surfaces were passivated by short 425.39-Da [SM(PEG)2; PEGylated SMCC cross-linker; Thermo Scientific Pierce] or long 5,000-Da (molecular mass 5,000; Rapp Polymere) heterobifunctional PEG spacers to avoid unspecific interactions between the cantilever and the glass surface. The PEG spacers offer an *N*-hydroxy succinimide group on one side for attachment to the amino silanized surface of the cantilever. The other end provides a Maleimide (Mal) group for attachment of the thiol group found in the Cysteine of the monoST.

For silanization, the cantilevers were first oxidized in a UV ozone cleaner (UVOH 150 LAB; FHR Anlagenbau GmbH) and subsequently silanized for 2 min in (3-aminopropyl)dimethylethoxysilane [ABCR; 50% (vol/vol) in ethanol]. For rinsing, the cantilevers were stirred in 2-Propanol (IPA) in MilliQ and afterward, dried at 80 °C for 30 min. After that, the cantilevers were incubated in a solution of 25 mM heterobifunctional PEG spacer and 50 mM Hepes for 30 min (for short PEG first solved in half DMSO and then filled to 50 mM Hepes). Finally, the monoST was bound to the cantilevers for 1 h at room temperature followed by a washing step in 1× PBS. The functionalized cantilevers were stored in measurement buffer (40 mM Hepes, pH 7.4, 10 mM $MgCl_2$, 200 mM NaCl, 1 mM DTT) until use.

The preparation of the glass surfaces is in a lot of steps similar to the functionalization of the cantilevers as seen in Fig. 1*C*. The glass surfaces are amino silanized followed by a passivation with PEG linkers. The Mal of PEG offers a binding site for the thiol group of CoA. Via an Sfp-catalyzed reaction, the CoA can bind the ybbR-tag harbored by the FAK protein construct. This way, the protein gets attached and tethered in an uncompromisingly specific way.

Before silanization, the glass surfaces have to be cleaned by sonification in 50% (vol/vol) IPA in MilliQ for 15 min. For oxidation, the glass surfaces are soaked for 30 min in a solution of 50% (vol/vol) hydrogen peroxide (30%) and sulfuric acid. Afterward, they have to be thoroughly washed in MilliQ and then blown dry in an N2 stream. Then, the glass surfaces get silanized by incubating them in ABCR [1.8% (vol/vol) in ethanol]. Thereafter, they were washed again in IPA and MilliQ and then dried at 80 °C for 40 min. Then, the PEG is applied as described for the cantilevers. Subsequent to rinsing, the surfaces were incubated in 20 mM CoA (Calbiochem) dissolved in coupling buffer (sodium phosphate, pH 7.2) to react with Mal. After washing the glass surfaces, 8 μL of the FAK construct (20 μM) was mixed with 1 μL Sfp-synthase (132 μM) and 1 μL of 10× reaction buffer (100 mM Tris, pH 7.5, 100 mM $MgCl_2$); then, it was pipetted on the surfaces and incubated for 2 h at room temperature. Finally, the surfaces were rinsed thoroughly in measurement buffer (40 mM Hepes, pH 7.4, 10 mM $MgCl_2$, 200 mM NaCl, 1 mM DTT).

Force Spectroscopy Experiments. The AFM measurements were conducted on an Asylum research controller (Asylum Research) providing analog-to-digital converter and digital-to-analog converter channels as well as a digital signal processor board for setting up feedback loops. The controller operated either a custom-built AFM head (34) or an *xyz*-movable piezo-driven sample stage. Data were recorded automatically by cycling through the following steps: (*i*) approach of the functionalized AFM tip to the surface to allow coupling to the SII of surface-immobilized FAK; (*ii*) retraction of the AFM cantilever with nanometer precision at a constant speed and simultaneous recording of the mechanical force response with pN precision; and (*iii*) after monoST:SII separation, the piezo stage of the AFM is moved to probe a new spot on the sample surface in the next cycle. This process was operated by using an IgorPro6 (Wavemetrics) program controlling the *z* piezo in the AFM head (or sample stage) and the *xy* piezos. The surface is sampled in steps of 100-nm distance in a snail trace to avoid probing a spot multiple times. The BioLever Mini (BLAC40TS) cantilevers (Olympus; 10-nm nominal tip radius, sharpened probe) were indented with 180 pN, applying no additional dwell time. Cantilevers were chemically modified (compare with *Sample Preparation for Surface and Cantilevers*) and were calibrated after the measurement using the equipartition theorem method (28). The datasets were recorded in the course of a few hours and contained around 50,000–90,000 curves saved in hdf5 files for additional data analysis.

Previous studies on smooth muscle myosin light-chain kinase (10) were conducted with 5,000-Da PEG linkers at a pulling speed of 800 nm/s. For this study, these parameters did not provide sufficient force resolution for clearly identifying contour length increments of the subtle peaks measured. By using shorter PEG linkers (35) (in this case, 425 Da) and thereby, increasing the loading rate as seen by the molecule, rupture forces of the detected events could be increased (*SI Appendix*, Fig. S13). These higher forces enable reliable worm-like chain fits for accurate analysis of the contour length increments.

Data Analysis. To show the characteristic unfolding patterns of the probed FAK construct, heatmaps were assembled with all curves that showed the correct total contour length (indicating correct site-specific attachment) as well as the presence of the characteristic unfolding peaks. Denoised (based on Savitzky–Golay, length 35, two polynomial for 800 nm/s and length 21, two polynomial for 12,800 nm/s) force spectroscopy data were aligned manually in force–distance space (only by translating along the distance axis to account for length differences in PEG); they were binned from −15 to 150 nm in distance and from −15 to 150 pN in force for measurements with short PEG linkers (425.39 Da) and binned from −15 to 250 nm in distance and from −15 to 250 pN in force for measurements with long PEG linkers (5,000 Da) to create a heatmap. The number of bins (equal for both distance and force axis) is dependent on the curves contained in the heatmap (Fig. 2, 150 bins; Fig. 3, 250 bins; Fig. 4, 250 bins; and Fig. 5, 150 bins).

The denoised data points (Savitzky–Golay) in force–distance space were binned on the distance axis into 2.5-nm- (for Fig. 2), 3-nm- (for Figs. 3 and 4), and 3.5-nm-sized slices (for Fig. 5) (moving the slice window by 0.2 nm each step), and their densities on the force axis (*y* axis) were estimated by a kernel density estimate (KDE) with a bandwidth of 0.2 pN (compare with *SI Appendix*, Fig. S14). The resulting most probable values are then assembled to form the most probable unfolding curve (shown as a black line in *SI Appendix*, Fig. S14) and analyzed to find the most probable unfolding peaks. The FWHMs of the distance slices were then taken as selection criteria for the unfolding peaks. If slices contain a rupture event, the drop in force results in broad distributions, thereby clearly deviating from the noise level. The peaks were first detected by a simple peak detection based on taking the first order difference and then validated by the FWHM of the distance slices. To be accepted as a peak, the FWHM of the distance slices has to be above the FWHM of the KDE of the accumulated FWHMs of the distance slices of the curve, which gives a good representation of the noise level of the curve (compare with *SI Appendix*, Fig. S14). The procedure of assembling the most probable curve does not necessarily reproduce absolute rupture forces but yields a good result for the most probable and most representative pathway (36).

For additional analysis of the contour length increments, each stretch preceding an unfolding event is fitted with the worm-like chain model. This is done for every single curve contained in the heatmap. The most probable contour length for each peak is determined using a KDE. The increments between these most probable contour lengths were used to compare them with structural elements of the crystal structure.

Data analysis was completely carried out in Python 2.7 and is available online together with all used datasets (https://gitlab.physik.uni-muenchen.de/Magnus.Bauer/fak_analysis).

MD Simulations. We use the FK-FAK construct developed previously (17) and solvate the protein in a 150 × 10 × 10-nm box. The total system contains ∼1.5 million atoms, including 908 Na+ and 903 Cl− ions, corresponding to a ionic strength of 0.1 M. We use GROMACS (15), version 2016 for all of our simulations. As force field, we use Amber99SB-ILDN* force field (16) with Joung ions (27) and a transferable intermolecular potential with 3 points (TIP3P) water model (37). We use a time step of 2 fs and freeze all bonds in our simulations through a linear constraint solver (LINCS) procedure (BP Hess) of fourth order. Two Nosé–Hoover thermostats, one for protein and one for nonprotein atoms, were used with a time constant of 0.6 ps to keep the temperature at 300 K. An isotropic Parrinello–Rahman barostat with a time constant of 2 ps and a compressibility of 4.5×10^{-5} with a reference pressure of 1 atm was used for pressure coupling. Verlet neighbor lists with a cutoff of 1.0 nm were used with an initial frequency of 0.03 ps. These parameters were automatically updated during the simulations by GROMACS for optimal performance. For long-range electrostatics, we use a fourth-order particle mesh Ewald method (38) with a grid spacing of 0.16.

We perform a total of 30 pulling simulations, each with a spring constant of 830 pN/nm: 10 simulations each at the velocities of 1, 1/3, and 1/10 nm/ns. The simulations were performed in the presence of an ATP molecule and an Mg^{2+} ion. To obtain single-domain pulling simulations, we start from the coordinates of the full FK-FAK construct and keep only the residues in the relevant domains. We relax the structures in 100-ns equilibrium simulations and solvate the FERM domain and the kinase domain in $67 \times 9 \times 9$ and $100 \times 8.5 \times 8.5$ nm, respectively. These correspond to 534,000 (FERM) and 706,000 (kinase) atoms. In both cases, we remove the ATP molecule and the Mg^{2+} ion from the simulation. We pull only using the fastest pulling velocity (1.0 m/s) and otherwise, keep all parameters unchanged.

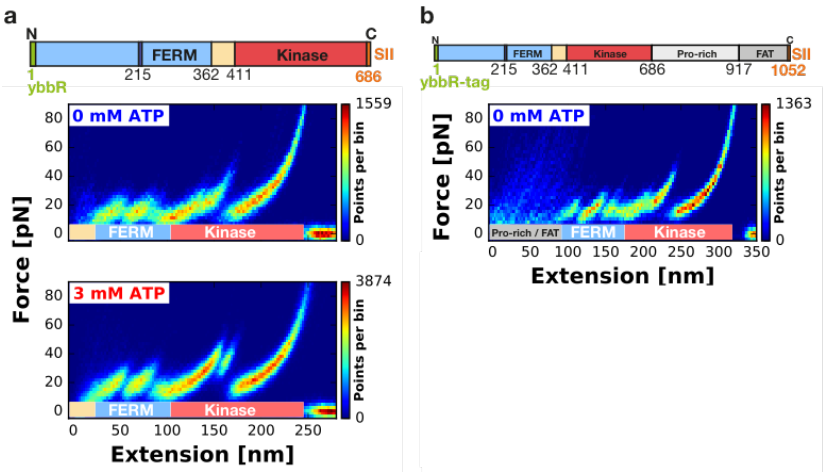
We quantify domain unfolding by measuring distances between residues as follows: for FERM, we used residues 216–362; for the linker, we used residues 362–418, and for the kinase, we used residues 418–686. We define the beginning and end of unfolding events as the times that the distances reach 10 and 45 nm for FERM, 7 and 15 nm for the linker, and 20 and 75 nm for the kinase. For the initial conformational activation, we also use a simple distance criterion: namely, whenever the distance 216–640 exceeds 10 nm. Since force profiles obtained in MD simulations include several intermediate ruptures, we identify peaks through a two-step procedure: (*i*) a Gaussian smoothing of the force profiles with an SD consistent with an extension of 0.1 nm and (*ii*) finding local maxima of the smoothed force profile in a window consistent with an extension of ± 10 nm.

ACKNOWLEDGMENTS. We thank Ivan Acebrón for help with activity measurements. M.S.B. acknowledges Leonard C. Schendel and Steffen M. Sedlak for experimental assistance; Angelika Kardinal and Thomas Nicolaus for laboratory support; Iris Ruider, Katherine Erlich, Marco Grison, and Wolfgang Ott for helpful discussions; Sylvia Kreuzer for administration; and the Nanosystems Initiative Munich for support. C.D. and F.G. are grateful for support from the state of Baden-Württemberg through high performance computing in Baden-Württemberg (bwHPC) and Deutsche Forschungsgemeinschaft (DFG) Grant INST 35/1134-1 FUGG. H.E.G. acknowledges funding from DFG Grant Sonderforschungsbereich 1032. F.G. acknowledges funding from the DFG through the research group SHENC (Shear Flow Regulation of Hemostasis—Bridging the Gap Between Nanomechanics and Clinical Presentation) and from the Klaus Tschira Foundation. D.L. acknowledges support from Spanish Ministry of Economy, Industry and Competitiveness Retos Grant BFU2016-77665-R cofunded by the European Regional Development Fund and Volkswagen Foundation Grant Az: 86 416-1. D.L. is the recipient of Worldwide Cancer Research Award 15-1177.

1. Winograd-Katz SE, Fässler R, Geiger B, Legate KR (2014) The integrin adhesome: From genes and proteins to human disease. *Nat Rev Mol Cell Biol* 15:273–288.
2. Galbraith CG, Yamada KM, Sheetz MP (2002) The relationship between force and focal complex development. *J Cell Biol* 159:695–705.
3. Sun Z, Guo SS, Fässler R (2016) Integrin-mediated mechanotransduction. *J Cell Biol* 215:445–456.
4. Seong J, et al. (2013) Distinct biophysical mechanisms of focal adhesion kinase mechanoactivation by different extracellular matrix proteins. *Proc Natl Acad Sci USA* 110:19372–19377.
5. Torsoni AS, Constancio SS, Nadruz WJ, Hanks SK, Franchini KG (2003) Focal adhesion kinase is activated and mediates the early hypertrophic response to stretch in cardiac myocytes. *Circ Res* 93:140–147.
6. Wong VW, et al. (2011) Focal adhesion kinase links mechanical force to skin fibrosis via inflammatory signaling. *Nat Med* 18:148–152.
7. Wang HB, Dembo M, Hanks SK, Wang Y (2001) Focal adhesion kinase is involved in mechanosensing during fibroblast migration. *Proc Natl Acad Sci USA* 98:11295–11300.
8. Puchner EM, et al. (2008) Mechanoenzymatics of titin kinase. *Proc Natl Acad Sci USA* 105:13385–13390.
9. von Castelmur E, et al. (2012) Identification of an n-terminal inhibitory extension as the primary mechanosensory regulator of twitchin kinase. *Proc Natl Acad Sci USA* 109:13608–13613.
10. Baumann F, et al. (2017) Increasing evidence of mechanical force as a functional regulator in smooth muscle myosin light chain kinase. *eLife* 6:e26473.
11. Lietha D, et al. (2007) Structural basis for the autoinhibition of focal adhesion kinase. *Cell* 129:1177–1187.
12. Kanchanawong P, et al. (2010) Nanoscale architecture of integrin-based cell adhesions. *Nature* 468:580–584.
13. Cai X, et al. (2008) Spatial and temporal regulation of focal adhesion kinase activity in living cells. *Mol Cell Biol* 28:201–214.
14. Goni GM, et al. (2014) Phosphatidylinositol 4,5-bisphosphate triggers activation of focal adhesion kinase by inducing clustering and conformational changes. *Proc Natl Acad Sci USA* 111:E3177–E3186.
15. Pronk S, et al. (2013) Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854.
16. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins* 78:1950–1958.
17. Zhou J, et al. (2015) Mechanism of focal adhesion kinase mechanosensing. *PLoS Comput Biol* 11:e1004593.
18. Bouchiat C, et al. (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys J* 76:409–413.
19. Hugel T, Rief M, Seitz M, Gaub HE, Netz RR (2005) Highly stretched single polymers: Atomic-force-microscope experiments versus ab-initiotheory. *Phys Rev Lett* 94:048301.
20. Hu X, Li H (2014) Force spectroscopy studies on protein-ligand interactions: A single protein mechanics perspective. *FEBS Lett* 588:3613–3620.
21. Verdorfer T, Gaub HE (2018) Ligand binding stabilizes cellulosomal cohesins as revealed by afm-based single-molecule force spectroscopy. *Sci Rep* 8:9634.
22. Dietz H, Rief M (2004) Exploring the energy landscape of gfp by single-molecule mechanical experiments. *Proc Natl Acad Sci USA* 101:16192–16197.
23. Case LB, Waterman CM (2015) Integration of actin dynamics and cell adhesion by a three-dimensional, mechanosensitive molecular clutch. *Nat Cell Biol* 17: 955–963.
24. Leonard TA, Hurley JH (2007) Two kinase family dramas. *Cell* 129:1037–1038.
25. Marko JF, Siggia ED (1995) Stretching DNA. *Macromolecules* 28:8759–8770.
26. Baumann F, et al. (2015) Monovalent strep-tactin for strong and site-specific tethering in nanospectroscopy. *Nat Nanotechnol* 11:89–94.
27. Joung IS, Cheatham TE (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* 112:9020–9041.
28. Hutter JL, Bechhoefer J (1993) Calibration of atomic-force microscope tips. *Rev Sci Instr* 64:1868–1873.
29. Yin J, Lin AJ, Golan DE, Walsh CT (2006) Site-specific protein labeling by sfp phosphopantetheinyl transferase. *Nat Protoc* 1:280–285.
30. Schmidt TG, Skerra A (2007) The strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* 2:1528–1535.
31. Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-ebna1 cells. *Nucleic Acids Res* 30:E9.
32. Ceccarelli DF, Song HK, Poy F, Schaller MD, Eck MJ (2006) Crystal structure of the ferm domain of focal adhesion kinase. *J Biol Chem* 281:252–259.
33. Zimmermann JL, Nicolaus T, Neuert G, Blank K (2010) Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat Protoc* 5:975–985.
34. Gumpp H, Stahl SW, Strackharn M, Puchner EM, Gaub HE (2009) Ultrastable combined atomic force and total internal fluorescence microscope. *Rev Sci Instr* 80: 063704.
35. Walder R, et al. (2017) Rapid characterization of a mechanically labile alpha-helical protein enabled by efficient site-specific bioconjugation. *J Am Chem Soc* 139:9867–9875.
36. Ott W, et al. (2017) Elastin-like polypeptide linkers for single-molecule force spectroscopy. *ACS Nano* 11:6346–6354.
37. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1998) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.
38. Darden T, York D, Pedersen L (1993) Particle mesh ewald: An n · log(n) method for ewald sums in large systems. *J Chem Phys* 98:10089–10092.

Supplementary Information

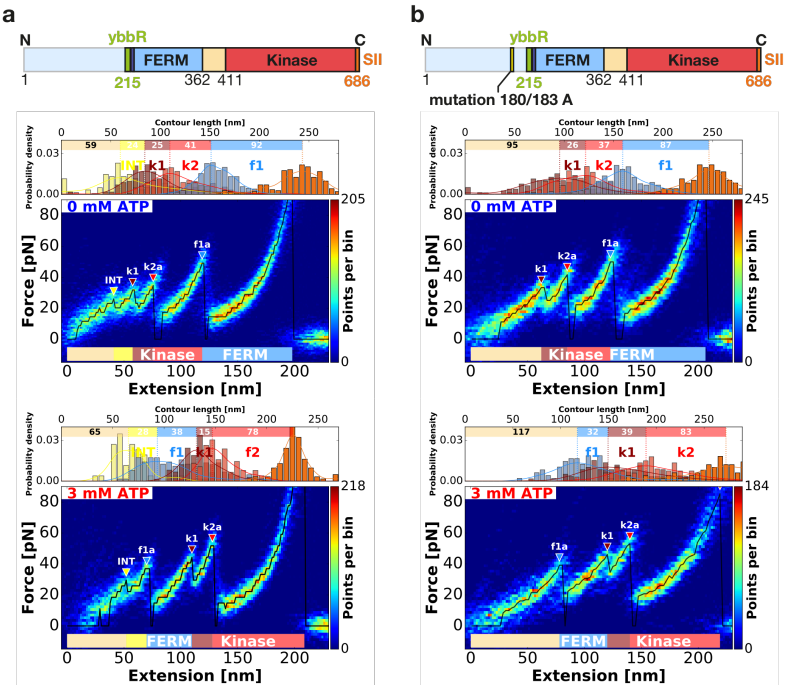
N-C terminal pulling of the FAK construct and full-length unfoldings



Supplementary Figure S1: *NC-terminal pulling of the FAK construct.* ***(a)*** *In initial experiments we attached FAK with tags placed at the N- and C-termini and retracted the cantilever at a speed of 800 nm/s with 5,000 Da PEG linkers. The top panel shows the unfolding without ATP present and the bottom with 3 mM ATP in the measurement buffer.* ***(b)*** *Probing of full-length FAK molecules (1-1052 amino acids) with 800 nm/s with 425.39 Da PEG resulted a longer unfolding pattern accounting for the longer total length. However, the pattern occurring at 100 nm is the same as in* ***(a)*** *as indicated by the labels below. This leads to the conclusion that the proline-rich region and FAT domain do not significantly contribute to unfolding of the autoinhibitory structure from (1-686 amino acids). This supports the findings of Goni et al.*

www.pnas.org/cgi/doi/10.1073/pnas.1820567116

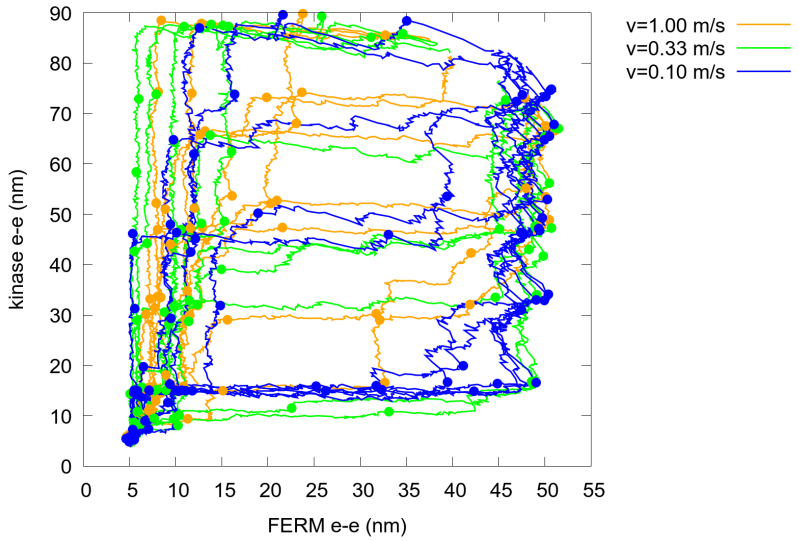
Contour length histograms for Figure 5



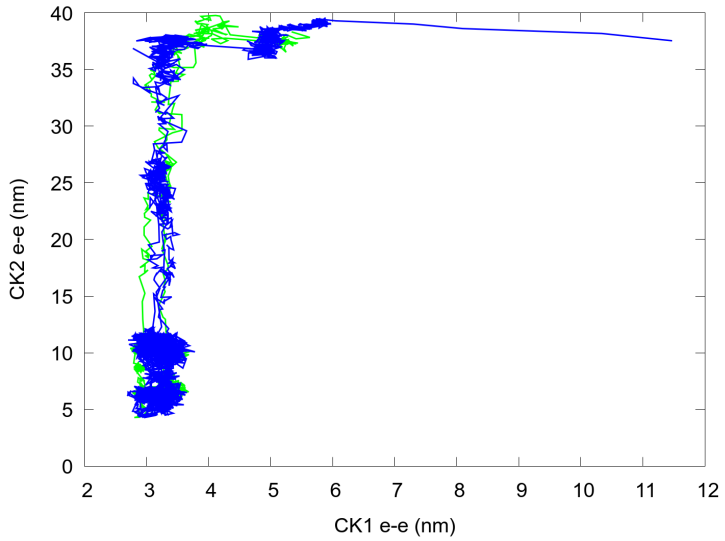
Supplementary Figure S2*: Figure 5 with analysis of contour length histograms.* *The Figure shows the same heatmaps as in Figure 5 for pullings with PEG 5,000 Da and 12,800 nm/s with added contour length histograms. Since the persistence length is changing too much over the course of the whole unfolding length, the increments are not very reliable. This is due to the long PEG linker (low persistence length) that is dominating the persistence length in the beginning of the curve and the increase in persistence length once parts of the protein get unfolded. It is possible to conduct WLC fits however the persistence length and contour length as fit parameters are not stable enough to produce comparable contour length increments. This is in contrast to the measurements with the short PEG 425.39 Da (Fig. 2, 3, 4) where only the first unfolding is dominated by the persistence length of the PEG and therefore yields comparable increments for further unfoldings.*

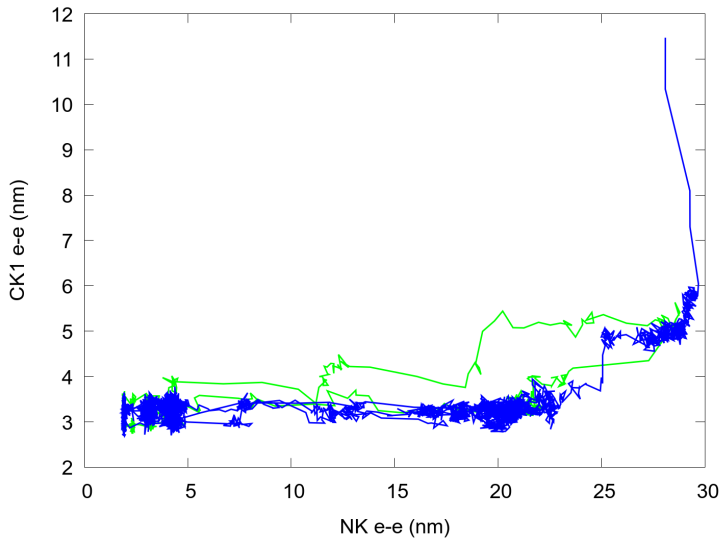
Kinase unfolding in molecular dynamics simulations

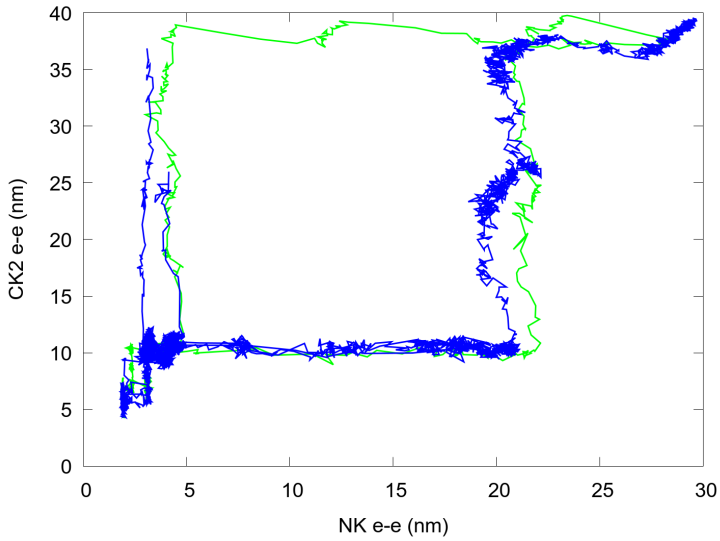
The kinase domain, in the presence of ATP, has been observed in AFM experiments to unfold in two stages. The first step amounts to 13 nm and the second one to 66 nm. In the MD simulations, the kinase domain stretches by 10 nm during or before FERM unfolding. 7 nm out of this stretching is due to a partial unfolding of the C-terminal region of the C-lobe, and another 3 nm is due to the lobes rearranging. While experimentally, kinase unfolding happens fully after the FERM domain unfolds, this partially unfolded state is observed in all of our simulations, and we consider it the most likely explanation to this first jump. The further unfolding of the kinase domain happens through numerous pathways, but we can see that the last part of the kinase that unfolds is the part of the C-lobe that is before the activation loop (cf. SI, where this subdomain is dubbed “CK1”). We note that kinase domain unfolding leads to FAK deactivation independent from the detailed sequence of events.



Supplementary Figure S3: *unfolding trajectories visualized in the "phase space" of FERM end-to-end distance and kinase end-to-end distance. Points are rupture peaks.*

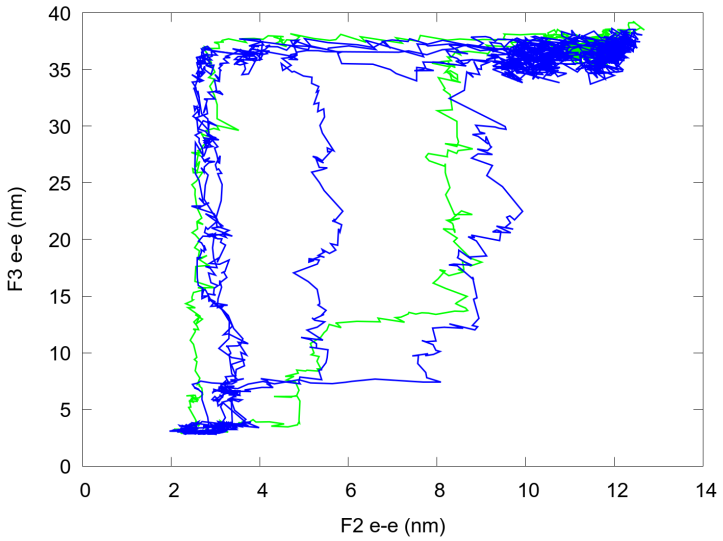

CK2 e-e (nm)
CK1 e-e (nm)
0
5
10
15
20
25
30
35
40
2
3
4
5
6
7
8
9
10
11
12


CK1 e-e (nm)
NK e-e (nm)
2
3
4
5
6
7
8
9
10
11
12
0
5
10
15
20
25
30

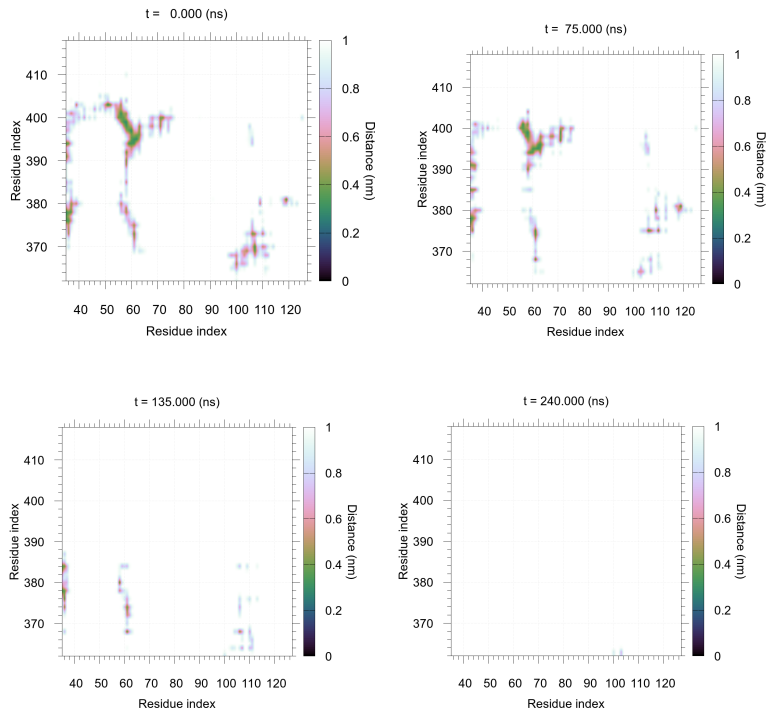


Supplementary Figures S4-6: *Order of unfolding between the three constitutive parts of the kinase domain.*

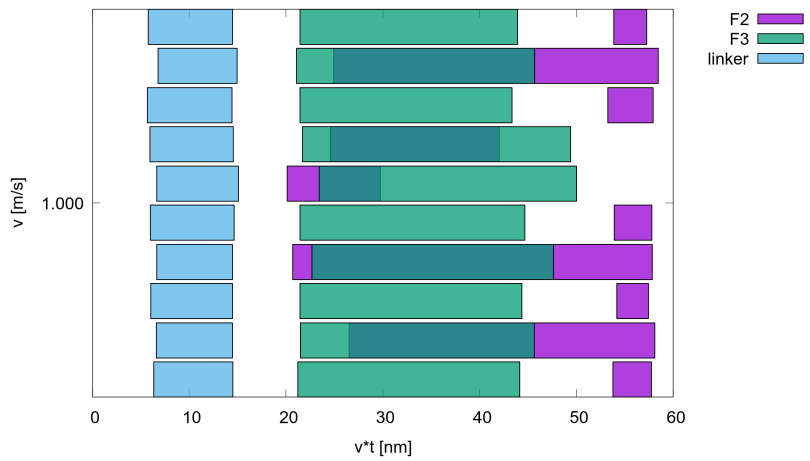
Only the 7 simulations consistent with the experiments are considered.



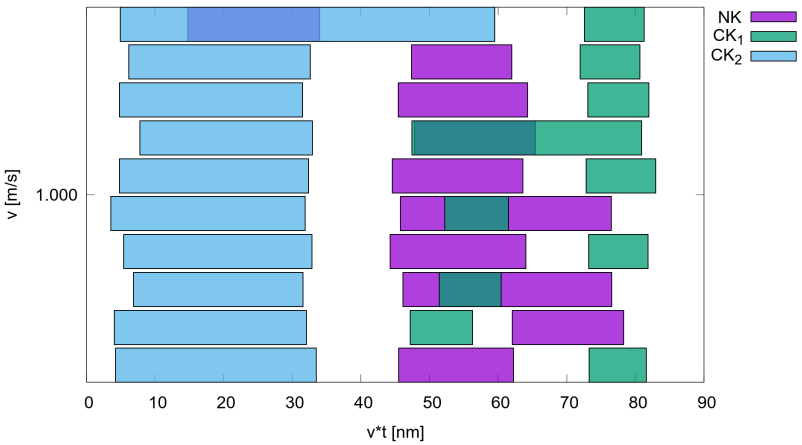
Supplementary Figure S7: *Order of unfolding between the two constitutive parts of the FERM domain. Only the 7 simulations consistent with the experiments are considered.*



Supplementary Figure S8: *contact maps of the linker-F1 interaction corresponding to the poses from Fig 5.*



Supplementary Figure S9: *the FERM-only unfolding simulations follow the same hierarchy of unfolding events as the unfoldings in the main text: The linker loses contact with lobe F1 first, then F3 unfolds in 8/10 cases, then F2 stretches.*

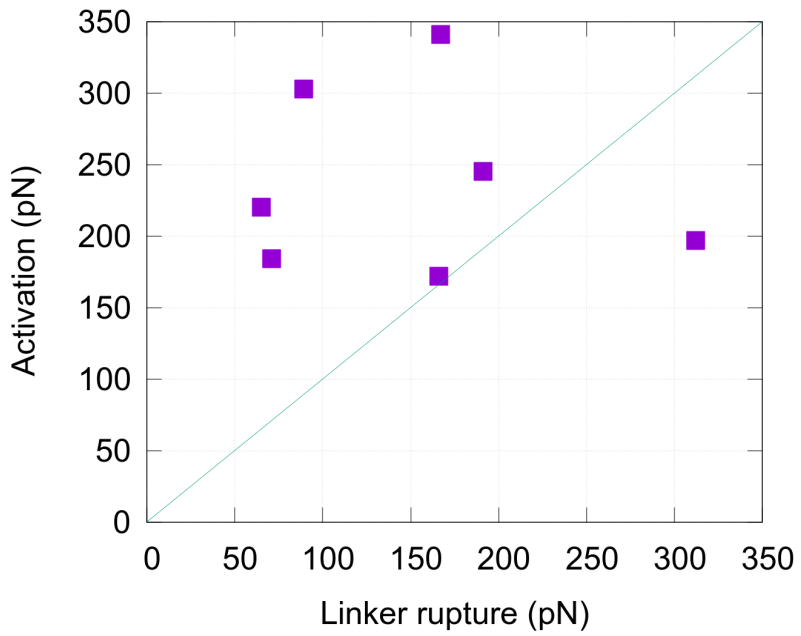


Supplementary Figure S10: *the kinase-only unfolding simulations follow the same hierarchy of unfolding events as the unfoldings in the main text: CK2 usually unfolds first, followed by NK and CK1.*

Event	MD (nm)	AFM (nm)	Assignment in AFM plots
Domain separation	10		first extension (prior to first unfolding)
Linker-F1 separation	12		first extension
F3 unfolding	30	29-32	f1a+f1b
F2 unfolding	13*	14-19	f1c
CK partial	10	13	k1
Kinase rest	70	68	k2

** 9 nm increase in the simulations and an estimated 4 nm from a loop region of the ybbR-tag included in the experimental construct*

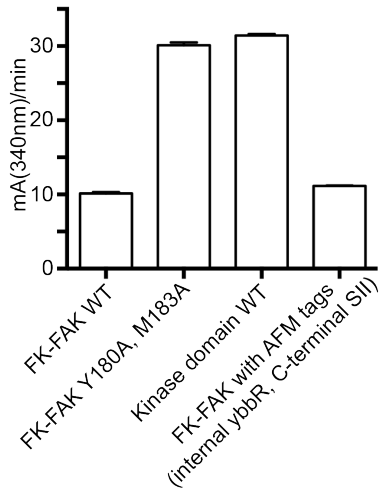
***Supplementary Table S1: Summary of the length changes observed in MD simulations (end-to-end distance changes) and AFM experiments (contour length increments)**. Due to the relatively high pulling speed in experiments (0.1m/s or higher), the MD increments can within the error of the two methods be assumed to be similar to the AFM increments.*



Supplementary Figure S11: Conformational activation requires more force than linker release. *We compare the rupture forces corresponding to conformational activation (y-axis) with those corresponding to the linker release (x-axis). Since 6/7 of our simulations required more force for activation than linker release, we conclude that linker rupture was not observed in experiments.*

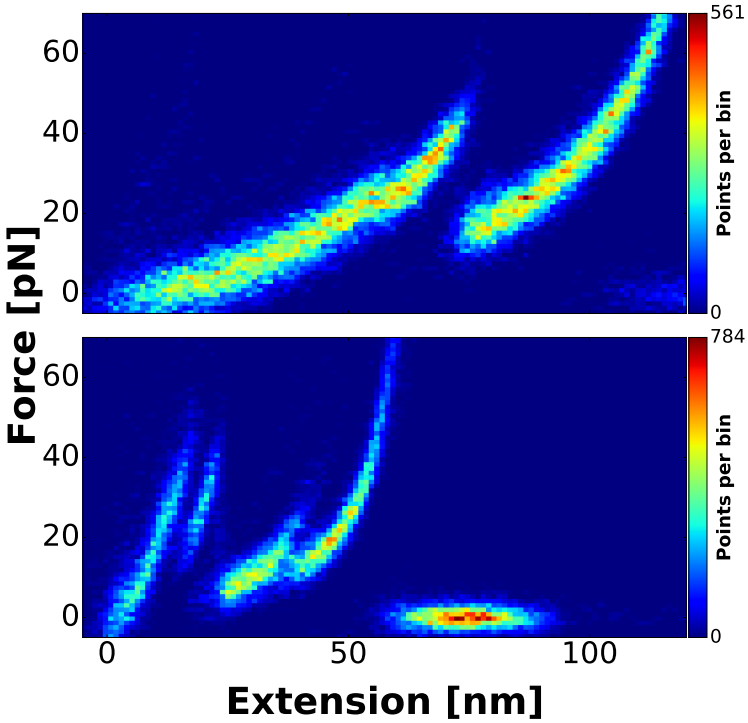
Kinase activity

An enzyme-coupled spectrophotometric assay was used to determine ATP turnover of FAK proteins as described by [2]. In brief, reactions were performed with 1 μM FAK, 2 mM MgCl2, 1 mM phosphoenolpyruvate, 0.25 mM NADH, 0.08 units/ L pyruvate kinase, 0.1 units/ L lactate dehydrogenase, and 100 μM E4Y (as polyGlu-Tyr, 4:1 Glu/Tyr; Sigma). Reactions were initiated with 1 mM ATP and NADH depletion was monitored by UV absorption at 340 nm.



Supplementary Figure S12: *Activity of FAK constructs was measured using a coupled kinase assay and readout of NADH consumption at 340 nm. Y180, M183 mutations disrupt autoinhibitory interactions between FERM and kinase domain . Introduction of tags for AFM experiments do not affect FAK activity or autoinhibition.*

Resolution increase with shorter PEGs

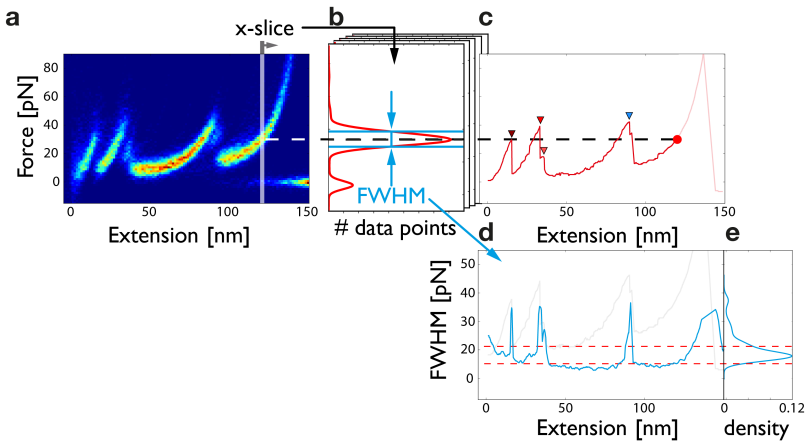


Supplementary Figure S13*: Comparison of the FERM construct with different linkers.* *Depicted is a comparison of the same FERM only construct (1-405 amino acids) between PEG 5,000 Da on top and PEG 425.39 Da on the bottom showing a much detailed unfolding pattern at 800 nm/s. This way it is possible to gain information on contour length increments not possible with the curves measured with the long PEG on top.*

In all previous experiments we used PEG with an average of 5,000 Da (long PEG), which has been used as a standard linker length in previous similar experiments.[3] Reducing the linker length to dimeric PEG (425.39 Da – short PEG), we indeed obtained greatly improved plots with reduced noise levels and increased force

signals. The increase in force signals we attribute to an increase in the average loading rates due to the the WLC behavior, resulting in higher force peaks. To even further boost the height of the force drops the loading rates where increased by using higher pulling velocities, 12,800 nm/s (fast pulling) instead of 800 nm/s (slow pulling) in some experiments. The experimental conditions indicated above (shorter linker, faster pulling) were applied accordingly in order to get enhanced results.

Most probable unfolding curve assembly and peak detection



Supplementary Figure S14*: Depiction of the most probable unfolding curve assembly and peak detection.* *For assembly of the most probable unfolding curve, the denoised data (Savitzky–Golay) in force-distance space are sliced in distance-axis slices (2.5 nm) with a moving slice window of 0.2 nm (a) and their densities (b) on the force axis (y-axis) were estimated by a kernel density estimate (KDE, bandwidth: 0.2 pN) (b). The most probable value is then plotted in c (red curve) to assemble the most probable unfolding curve. The FWHM of the most probable values of the KDEs in b are then plotted in d. Afterwards the KDE over the FWHMs of the distance-slices in d are shown in e, together with their FWHM (red dashed lines). This FWHM value describes the noise level of the most probable unfolding curve with points deviating showing unfolding events. The most probable unfolding curve can thereby be analyzed for most probable unfolding peaks (colored triangles on top of the red curve). A first selection is done by selecting peaks based on its first order difference. Then, the peaks are evaluated concerning their FWHM value in d. Only peaks above the FWHM of the noise level are accepted as peaks.*

1. Goni, G.M. et al. Phosphatidylinositol 4,5-bisphosphate triggers activation of focal adhesion kinase by inducing clustering and conformational changes. *Proc Natl Acad Sci U S A* **111**, E3177-86 (2014).
2. Lietha, D. et al. Structural basis for the autoinhibition of focal adhesion kinase. *Cell* **129**, 1177-87 (2007).
3. Baumann, F. et al. Increasing evidence of mechanical force as a functional regulator in smooth muscle myosin light chain kinase. *eLife* **6**, 621 (2017).

Part III

Appendix

A
List of Publications

[1] 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

[2] 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 Letters, 2015, DOI: 10.1021/acs.nanolett.5b02727

[3] 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

[4] **E. Durner**, W. Ott, M. A. Nash, and H. E. Gaub, "Post-translational sortase-mediated attachment of high-strength force spectroscopy handles", ACS Omega, 2017, DOI: 10.1021/acsomega.7b00478

[5] W. Ott*, **E. Durner***, and H. E. Gaub, "Site-Specific Protein Coupling Strategies for Surface-Based Binding Assays", Angewandte Chemie, 2018, DOI: 10.1002/ange.201805034

[6] M. S. Bauer, F. Baumann, C. Daday, P. Redondo, **E. Durner**, M. A. Jobst, L. F. Milles, D. Mercadante, D. A. Pippig, H. E. Gaub, F. Gräter, and Daniel Lietha, "Structural and mechanistic insights into mechanoactivation of focal adhesion kinase", Proceedings of the National Academy of Sciences, 2019, DOI: 10.1073/pnas.1820567116

[7] R. C. Bernardi*, **E. Durner***, C. Schoeler, K. H. Malinowska, B. G. Carvalho, E. A. Bayer, Z. Luthey-Schulten, H. E. Gaub, and M. A. Nash, "Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy", Journal of the American Chemical Society, 2019, DOI: 10.1021/jacs.9b06776

[8] **E. Durner**, M. S. Bauer, M. A. Jobst, W. Ott, L. F. Milles, and H. E. Gaub, "Conformational Transitions of a Protein Fold Observed by Dynamic Single-Molecule Force Spectroscopy", *under review at Physical Review Letters*

* These authors contributed equally to this work

Bibliography

[1] 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", *Journal of the American Chemical Society*, vol. 139, no. 49, pp. 17 841–17 852, 2017, ISSN: 15205126. DOI: 10.1021/jacs.7b07574 (cit. on p. 4).

[2] T. Verdorfer and H. E. Gaub, "Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy", *Scientific Reports*, vol. 8, no. 1, pp. 1–8, 2018, ISSN: 20452322. DOI: 10.1038/s41598-018-27085-x (cit. on p. 4).

[3] C. A. Pierse and O. K. Dudko, "Distinguishing Signatures of Multipathway Conformational Transitions", *Physical Review Letters*, vol. 118, no. 8, pp. 1–5, 2017, ISSN: 10797114. DOI: 10.1103/PhysRevLett.118.088101 (cit. on pp. 4, 13).

[4] R. Lamed, E. Setter, and E. A. Bayer, "Characterization of a cellulose-binding, cellulase-containing complex in clostridium thermocellum", *Journal of Bacteriology*, vol. 156, no. 2, pp. 828–836, 1983, ISSN: 00219193 (cit. on p. 7).

[5] Y. Ben David, B. Dassa, I. Borovok, R. Lamed, N. M. Koropatkin, E. C. Martens, B. A. White, A. Bernalier-Donadille, S. H. Duncan, H. J. Flint, E. A. Bayer, and S. Moraïs, "Ruminococcal cellulosome systems from rumen to human", *Environmental Microbiology*, vol. 17, no. 9, pp. 3407–3426, 2015, ISSN: 14622920. DOI: 10.1111/1462-2920.12868 (cit. on p. 7).

[6] P. Dumas, E. Ennifar, C. Da Veiga, G. Bec, W. Palau, C. Di Primo, A. Piñeiro, J. Sabin, E. Muñoz, and J. Rial, "Extending ITC to Kinetics with kinITC", *Methods in Enzymology*, vol. 567, pp. 157–180, 2016, ISSN: 15577988. DOI: 10.1016/bs.mie.2015.08.026 (cit. on p. 8).

[7] J. Y. Lichtenberg, Y. Ling, and S. Kim, "Non-specific adsorption reduction methods in biosensing", *Sensors (Switzerland)*, vol. 19, no. 11, pp. 1–17, 2019, ISSN: 14248220. DOI: 10.3390/s19112488 (cit. on p. 8).

[8] L. F. Milles, K. Schulten, H. E. Gaub, and R. C. Bernardi, "Molecular mechanism of extreme mechanostability in a pathogen adhesin", *Science*, vol. 359, no. 6383, pp. 1527–1533, 2018, ISSN: 10959203. DOI: 10.1126/science.aar2094 (cit. on p. 12).

[9] 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", *Journal of Physical Chemistry B*, vol. 121, no. 15, pp. 3620–3625, 2017, ISSN: 15205207. DOI: 10.1021/acs.jpcb.6b09593 (cit. on p. 12).

[10] M. Schlierf, H. Li, and J. M. Fernandez, "The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 19, pp. 7299–7304, 2004, ISSN: 00278424. DOI: 10.1073/pnas.0400033101 (cit. on p. 13).

[11] B. T. Marshall, M. Long, J. W. Piper, T. Yago, R. P. McEver, and C. Zhu, "Direct observation of catch bonds involving cell-adhesion molecules", *Nature*, vol. 423, no. 6936, pp. 190–193, 2003, ISSN: 00280836. DOI: 10.1038/nature01605 (cit. on p. 13).

[12] E. Evans and K. Ritchie, "Dynamic strength of molecular adhesion bonds.", *Biophysical journal*, vol. 72, no. 4, pp. 1541–55, Apr. 1997, ISSN: 0006-3495. DOI: 10.1016/S0006-3495(97)78802-7 (cit. on p. 13).

[13] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, "Molecular dynamics study of unbinding of the avidin-biotin complex", *Biophysical Journal*, vol. 72, no. 4, pp. 1568–1581, 1997, ISSN: 00063495. DOI: 10.1016/S0006-3495(97)78804-0 (cit. on p. 13).

[14] R. W. Friddle, A. Noy, and J. J. De Yoero, "Interpreting the widespread nonlinear force spectra of intermolecular bonds", *Proceedings of the National Academy of Sciences*, vol. 110, no. 11, Feb. 2013, ISSN: 0027-8424. DOI: 10.1073/pnas.1300714110 (cit. on p. 13).

[15] G. Bell, "Models for the specific adhesion of cells to cells", *Science*, vol. 200, no. 4342, pp. 618–627, May 1978, ISSN: 0036-8075. DOI: 10.1126/science.347575 (cit. on p. 13).

[16] 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, 2003, ISSN: 09538984. DOI: 10.1088/0953-8984/15/18/305 (cit. on p. 14).

[17] 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, ISSN: 07437463. DOI: 10.1021/la070131e (cit. on p. 14).

[18] 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", *Journal of Physical Chemistry B*, vol. 111, no. 8, pp. 1963–1974, 2007, ISSN: 15206106. DOI: 10.1021/jp065530h (cit. on p. 14).

[19] L. Livadaru, R. R. Netz, and H. J. Kreuzer, "Stretching Response of Discrete Semiflexible Polymers", *Macromolecules*, vol. 36, no. 10, pp. 3732–3744, May 2003, ISSN: 0024-9297. DOI: 10.1021/ma020751g (cit. on p. 14).

Acknowledgements

Without the support and help of many people, the work that led to this thesis would neither have been enjoyable, nor possible. I would therefore like to express my gratitude to:
Hermann Gaub for giving me the chance to explore science in a great and welcoming environment, for promoting a culture of collaboration and for sharing a wealth of knowledge in biophysics. **Michael Nash** for bringing together a group of people I am happy to have been part of, for giving me the freedom to explore my own projects, and for always providing motivation.
Special thanks go to my dear officemates and friends, **Markus Jobst**, **Lukas Milles**, **Wolfgang Ott**, and **Tobias Verdorfer**. Scientifically, I benefited immensely from working with you closely, sharing projects, teaching each other, and by having good (and sometimes heated) discussions. But most importantly for the companionship and for all the crazy things that happened (and will continue to happen) along the way.
Magnus Bauer for his dedication to science, questionable choice in software, for always being able to argue with you, and most importantly for being you (and eating lots of fish). **Fabian Baumann** for a memorable trip to Venice, helpful disscussions about optics, many mensa runs, and general good nature. **Edward Bayer** for introducing me to the cellulosome, a great collaboration, and for being always helpful and full of knowledge. **Rafael Bernardi** for fruitful collaborations on the schnitzel cornfields. **Katherine Erlich** for all the good times and countless other things, for helping me with many scientific and linguistic issues, and for always lending me an ear. **Angelika Kardinal** for keeping the lab running, and for spreading good mood. **Sylvia Kreuzer** and **Sabine Faust** for always being on top of administrative issues and keeping our backs free. And for all the sweets. **Thomas Nicolaus** for being a great help in many things and foremost for being a friend, without you much of this would not have been possible. **Franziska Kriegel** for being a great colleague and for exploring San Francisco with me. **Klara Malinowska** for shared projects in science and boulderwelt. **Thomas Perkins** for his contagious enthusiasm for force spectroscopy and helpful discussion. **Diana Pippig** for biochemical expertise as well as for excursions to the beer garden. **Leonard Schendel**, **Steffen Sedlak**, and **Achim Löw**, it was always fun to work at your sides. **Constantin Schöler** for many Stangen, for always being able to talk science, many hours of bouldering, and for being a good friend. **Willem Vanderlinden** for being a friendly neighbor, always willing to discuss science, with and without beer.
This thesis was set using LaTeX and is based on a template from Jordan Suchow, although heavily altered by Markus Jobst.
And last but not least, I want to thank: **My family** for the long-lasting support throughout the years, without which I would not be where I am now. **Anna Krautloher** for sharing everything with me, always encouraging me and for being the best partner I could ever wish for.