Probing and engineering mechanostable protein complexes

Inauguraldissertation
zur Erlangung der Würde eines Doktors der Philosophie vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Zhaowei Liu

Basel, 2021

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Michael Nash, Prof. Dr. Roderick Lim, und Prof. Dr. Jorge Alegre-Cebollada

Basel, den 19. Oktober 2021

Prof. Dr. Marcel Mayor

Dekan

Abstract

The mechanostability of proteins plays an important role in various biological processes, for example cell adhesion and pathogen invasion. Single-molecule force spectroscopy (SMFS) is a powerful tool to understand the molecular mechanisms of mechanostable proteins, gain a mechanistic insight into biological systems and also direct the engineering of biomolecules for desirable mechanical properties, for example enhanced mechanostability.

One family of highly mechanostable cell adhesion proteins is the dockerin (Doc)-cohesin (Coh) family from cellulosomes. Cellulosomes are large protein networks used by certain bacteria to bind and digest cellulose fibers. The interaction between Xmodule-dockerin B (XMod-DocB) and cohesin E (CohE) is responsible for attaching the cellulosome of human gut bacterium *R. champanellensis* (*Rc.*) to the cell wall and therefore is crucial for cellulosome function. SMFS is used to demonstrate that the XMod-DocB:CohE complex can be formed in two different conformations, a behavior known as ‘dual-binding mode’, and dissociates through three pathways with distinct mechanical stabilities under force. The complex preferably populates a high force pathway under increased force loading rate, precisely resembling a catch bond.

In addition to naturally occurring adhesion proteins, the mechanostability of antibodies and alternative scaffolds is also important for their functions in the context of antibody-coated therapeutic nanoparticles. Anticalin is a type of alternative scaffold developed to target various human cell surface receptors and small molecules related to diseases. One of its targets is cytotoxic T-lymphocyte antigen 4 (CTLA-4), an important target for tumor immunotherapy. Using SMFS combined with non-canonical amino acid incorporation and click chemistry, external pulling forces are applied to anticalin from eight different directions to dissociate it from CTLA-4 and characterize the geometric dependency of the unbinding energy landscape. The highest rupture force which is ~100% higher than the least mechanostable pulling geometry is found when pulling from residue 60 of anticalin.

The anisotropic response of proteins to mechanical forces can also be used to engineer naturally occurring protein-ligand systems and change their mechanical properties. Another Doc:Coh system from *Rc.*, DocG:CohE complex, dissociates in two pathways under force. The pulling geometry affects the rupture force in both pathways as well as the rate of entering each

pathway. When pulling from residue 13 of CohE, the complex exhibits a catch bond behavior, which is distinct from other measured pulling geometries, including the native pulling geometry.

In summary, SMFS is used here both to understand the underlying mechanisms of mechanostable protein-ligand complexes and to engineer them for higher mechanical stabilities as well as unique behaviors such as catch bonding.

Table of contents

Preface

Mechanical forces are constantly generated, maintained and propagated in biological systems at various levels, including individual molecules, cells and tissues. A major class of proteins with mechanical functions is adhesion proteins. Cells use adhesion proteins, for example cadherin and integrin, to adhere to the extracellular matrix, transduce mechanical forces and trigger downstream signaling reactions to regulate cell activity [1,2]. In pathogenesis, bacterial adhesins tightly adhere to target molecules on the host's cells and promote invasion [3,4]. Another class of bacterial adhesion proteins, cohesin and dockerin pairs, constitute a major building block of large extracellular protein networks known as cellulosomes. These receptor-ligands maintain the integrity and function of cellulosomal networks by resisting external shear forces [5,6].

In addition to native biomolecular systems, mechanical forces are also involved in various engineered biological systems. Examples of such artificial systems include drug-delivery nanoparticles [7,8], protein hydrogels [9,10], and biomimetic materials [11,12].

This thesis includes both of the aforementioned topics: mechanistic studies on native cell adhesion systems as well as engineering of mechanostable artificial systems. The scientific context of these two topics and the experimental and theoretical tools used in this thesis are introduced in chapter 1.

The main results are presented in chapters 2-4. Chapter 2 aims to gain mechanistic insight into the mechanical properties of a native bacterial cell adhesion system, the Xmodule-dockerin:cohesin complex from the cellulosome of a human gut bacterium, *Ruminococcus champanellensis* (*Rc.*). The unique dual-binding mode and high force catch bond behaviors of this system were demonstrated using a combination of single-molecule experiments and simulations. Chapter 3 focuses on an engineered alternative scaffold, anticalin. A novel AFM-SMFS method is presented to mechanically dissociate anticalin from its target by applying tension from eight different anchoring residues. Combined with simulations, these experiments demonstrate that the unbinding energy landscape and mechanical stability of this therapeutic protein complex is highly dependent on the force loading geometry. Chapter 4 uses the method developed in chapter 3 to dissociate another *Rc.* Doc:Coh complex from different directions and demonstrate that the complex exhibits a catch bond behavior when pulled from a specific direction, while the other measured pulling geometries exhibit slip bond behavior.

The results presented in this thesis are published in peer-reviewed journals (chapters 1, 2 and 3), or are in preparation for publication (chapter 4).

Research articles:

Chapter 2: Zhaowei Liu., Haipei Liu, Andres M. Vera, Rafael C. Bernardi, Philip Tinnefeld, and Michael A. Nash. High force catch bond mechanism of bacterial adhesion in the human gut. *Nat Commun* 11, 4321 (2020)

Contributions statement:

- Designed the experiments
- Performed biochemical preparations (molecular cloning, protein expression, protein labeling and protein purification) for smFRET and AFM-SMFS experiments
- Performed AFM-SMFS experiments and analyzed the data
- Wrote and edited the article manuscript

Chapter 3: Zhaowei Liu, Rodrigo A. Moreira, Ana Dujmović, Haipei Liu, Byeongseon Yang, Adolfo B. Poma, Michael A. Nash. Mapping Mechanostable Pulling Geometries of a Therapeutic Anticalin/CTLA-4 Protein Complex *Nano Letters* 10.1021/acs.nanolett.1c03584, (2021)

Contributions Statement:

- Designed the experiments
- Performed biochemical preparations (molecular cloning, protein expression, protein-peptide conjugation and protein purification) for AFM-SMFS experiments
- Performed AFM-SMFS experiments and analyzed the data
- Supervised the experiments as a part of Ana Dujmović’s master thesis project
- Wrote and edited the article manuscript

Chapter 4: in preparation for publication

Contributions Statement:

- Designed the experiments
- Performed biochemical preparations (molecular cloning, protein expression, protein-peptide conjugation and protein purification) for AFM-SMFS experiments
- Performed AFM-SMFS experiments and analyzed the data
- Supervised the experiments as a part of Ana Dujmović's master thesis project
- Drafted the article manuscript

Review article:

Chapter 1: Byeongseon Yang, Zhaowei Liu, Haipei Liu, and Michael A. Nash. Next generation methods for single-molecule force spectroscopy on polyproteins and receptor-ligand complexes. *Front. Mol. Biosci.* 7:85 (2020)

Contributions Statement:

- Drafted the 'Measurement configurations for AFM-SMFS', 'Polyproteins assembled by receptor-ligand complexes', and 'Pulling protocols and cantilever innovations in AFM-SMFS' sections of the manuscript
- Reviewed and edited the manuscript

I also contributed to the following works, the results of which are not presented in this thesis.

Byeongseon Yang, Haipei Liu, Zhaowei Liu, Regina Doenen, and Michael A. Nash. Influence of fluorination on single-molecule unfolding and rupture pathways of a mechanostable protein adhesion complex. *Nano Lett.*, 20, 12, 8940–8950 (2020)

Contributions Statement:

- Provided AFM data analysis tools
- Reviewed and edited the article manuscript

Haipei Liu, Zhaowei Liu, Byeongseon Yang, Joanan Lopez Morales, and Michael A. Nash. Optimal sacrificial domains in mechanical polyproteins: *S. epidermidis* adhesins are tuned for energy dissipation. *Submitted*. (2021)

Contributions Statement:

- Performed molecular cloning of fusion protein constructs
- Reviewed and edited the article manuscript

Chapter 1 Introduction

Part of this chapter (sections 1.1, 1.2, 1.3 and 1.4) was published in the following review article [13]:

Yang B., Liu Z., Liu H., and Nash M.A. Next Generation Methods for Single-Molecule Force Spectroscopy on Polyproteins and Receptor-Ligand Complexes. *Front. Mol. Biosci.* 7:85. (2020)

Adapted under a Creative Commons Attribution License (CC BY).

1.1 Measuring protein mechanics using AFM-SMFS

Single-molecule force spectroscopy (SMFS) is a well-established method that directly probes structural changes of macromolecules under the influence of mechanical force. The utilization of AFM in single-molecule biophysics has a history of over 20 years. The first AFM-SMFS measurements were carried out in the 1990s to study the unfolding kinetics of single immunoglobulin domains [14], the mechanical stabilities of protein-ligand complexes [15], and the strength of single covalent bonds [16]. With advancements in instrumentation, sample preparation, surface chemistry, and computer simulations, AFM-SMFS has become a very powerful tool in biophysics and has been used to study various native mechanical proteins including examples from muscle [14,17–19], hearing [20–24], blood coagulation [25–27], cell adhesion [28–30], the extracellular matrix [31,32], protein folding at the ribosomal exit tunnel [33–35], protein unfolding and proteolysis by the proteasome [36,37], and DNA/RNA molecular motors [38–41]. In addition to native mechanical proteins, SMFS has the potential to probe the biophysical stability of therapeutic antibodies and other binding scaffold proteins and predict their developability as drugs[42,43], and also contributed to the design and characterization of artificial biomaterials[10,12].

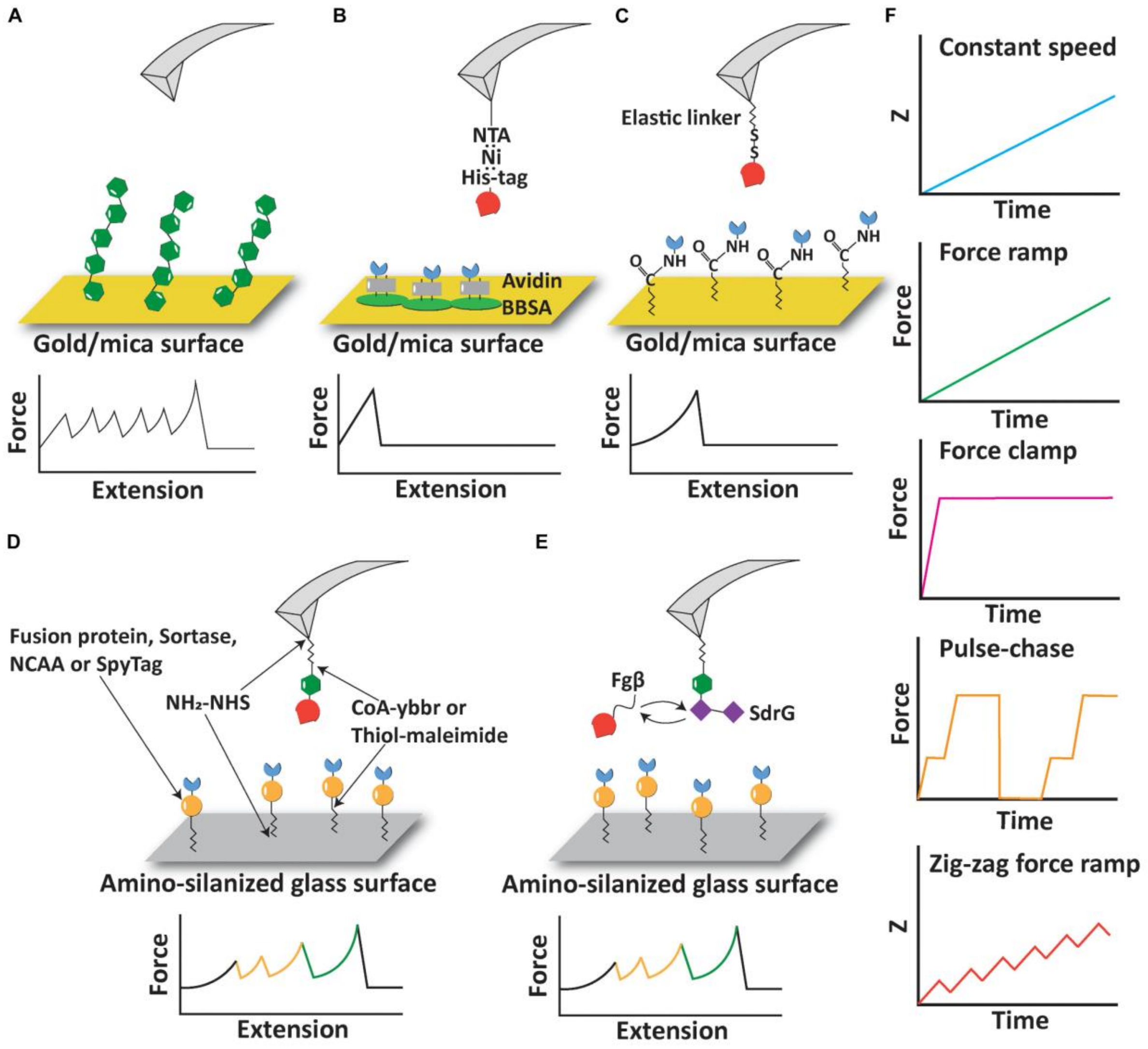


Figure 1.1 Experimental configurations and pulling protocols of AFM-SMFS. A: AFM measurement based on non-specific adsorption of proteins. **B:** Immobilization of proteins using non-covalent interactions including His:Ni and biotin:avidin. **C:** Immobilization of proteins using elastic linkers and covalent bonds. **D:** Covalent immobilization of proteins of interest and fingerprint domains using a variety of reactions and peptide tags. **E:** A free diffusion system allows continuous exchange of ligand molecules on the cantilever. **F:** Different pulling protocols used in AFM-SMFS. This figure was reprinted from Yang *et al.* [13] under a Creative Commons Attribution License (CC BY).

One of the earliest AFM-SMFS experimental setups, which is still used today, relies on non-specific adsorption of target proteins on surfaces, usually gold or mica [14,31,44,45]. The AFM cantilever tip is brought into contact with a surface sparsely decorated with adsorbed polyproteins, and with a low probability, a single molecule non-specifically adsorbs onto the AFM cantilever tip forming a tether between the cantilever and the surface, as shown in **Figure**

1.1A. The polyprotein controls the force loading geometry on individual domains between their N- and C-termini. This configuration has been successfully and widely used over the years [46–48]. However, there remain several limitations of the technique. One limitation is that the yield of usable single-molecule interaction curves is very low, sometimes well below 1%. This is because non-specific pickup of polyproteins is unpredictable, prone to spurious signals, and in many ways unreliable. Another limitation is that non-specific pickup procedures are not useful for probing receptor-ligand interactions because molecules that are picked up will clog the AFM tip and complicate data analysis.

Receptor-ligand interactions are a major class of protein-protein interactions that pose interesting objects of study. They have been reported both as the object of study as well as a tool for improving experimental yields in AFM-SMFS experiments using an approach referred to as “molecular handles” [49]. Early in the development of receptor-ligand handles for AFM-SMFS, several affinity tags commonly found in protein biochemistry labs were used to pick up molecules with the AFM tip. For example, high-affinity non-covalent interactions including Ni:His-tag and biotin:avidin have both been used as immobilization tags onto surfaces or as the object of study in AFM-SMFS [50–53], as shown in **Figure 1.1B**. One limitation of these non-covalent interactions as molecular handles for AFM-SMFS is the relatively low forces required to rupture these complexes. Both biotin/avidin and Ni-NTA/His-tag pairs break at around ~100–200 pN depending on the loading rate [15,54]. Therefore, the aforementioned AFM-SMFS configurations using non-covalent molecular handles are not suitable to probe mechanostable systems where the rupture or unfolding forces of the protein of interest (POI) exceed the rupture force of the molecular handle.

Another commonly used measurement configuration involves covalent bond formation between the POI and the surface. Since the rupture force of a covalent bond is >2 nN [16], covalent linkage to the surface establishes a link that is significantly more stable than typical receptor-ligand interactions or domain unfolding forces. Covalent linkage of proteins to surfaces/AFM tips is therefore a suitable setup for measuring mechanostable protein interactions and domain unfolding. Such an approach is also valuable when combined with the approaches mentioned, particular specific receptor-ligands as pulling handles. As shown in **Figure 1.1C**, disulfide bonds and EDC/NHS coupling reactions were used to covalently link cysteine or lysine residues to the surface via a polyethylene glycol (PEG) linker [51,55]. Disulfide bonds have also been used to measure the unfolding force of single protein domains under different pulling geometries [56]. However, strictly defining the pulling geometry in this case may be hampered by native cysteines or the multiplicity of lysines present on the POI.

For all the AFM-SMFS experimental configurations discussed above, a common limitation is the difficulty in discriminating single-molecule interactions from non-specific or multiple interactions occurring in the measurements [57,58]. Although the elastic linker attaching the protein to the surface helps to exclude short range non-specific adhesion [59], it is not sufficient to eliminate all background signals. To solve this problem, experimenters have identified a variety of protein domains which have characteristic unfolding patterns, well-defined contour lengths and unfolding forces that can serve as internal control modules to validate single-molecule interactions. These protein domains are known as “fingerprint domains” and have been used to screen for single receptor-ligand complex unbinding events from large datasets. We note that the fingerprint domains used for receptor-ligand SMFS should be chosen so that they unfold at a much lower range of forces than the unbinding event of the receptor-ligand under study in order to avoid the fingerprint biasing effect [60].

A typical AFM experimental setup to measure protein-ligand interactions with fingerprint domains is shown in **Figure 1.1D**. A polyprotein consisting of a fingerprint domain and the protein/ligand of interest is covalently immobilized on the AFM tip or the surface through an elastic linker, most often a poly(ethylene glycol) (PEG) linker [61] or more recently an elastin-like polypeptide (ELP)[62]. The POI can be expressed as a fusion protein with the fingerprint domain or covalently attached to the fingerprint domain and elastic linker using sortase or ybbr tags [62–64]. A broad range of receptor-ligand interactions including cohesin-dockerin [6,65–67], antibody-antigen [68] and bacterial adhesin-host interactions [4] have been studied with the help of fingerprint domains.

In the aforementioned experimental setup, the protein-ligand interaction can be lost due to irreversible unfolding of the protein molecule immobilized on the tip. In order to solve this problem, an exchangeable receptor-ligand pair, SdrG:Fgβ, was added between the receptor and ligand, as shown in **Figure 1.1E**. Two features of the SdrG:Fgβ complex are crucial to this experimental configuration: (1) the SdrG:Fgβ complex is able to withstand a force as high as 2 nN [4], which is in the same regime as a covalent bond and significantly larger than other receptor-ligand interactions. Therefore the receptor:ligand complex would always rupture without breaking the SdrG:Fgβ interaction; and (2) the affinity between SdrG and Fgβ is moderate (300–400 nM) [69]. Therefore, the receptor/ligand molecule attached to the tip is frequently exchanged based on the natural off-rate at equilibrium of this complex. A freely diffusing molecule can then re-bind the SdrG molecule on the tip and prevent the loss of interaction due to tip clogging or protein unfolding. This experimental setup has been used to characterize the mechanical properties of monovalent and tetravalent streptavidin:biotin complex [70,71]. A limitation of this method is that the N terminus of the Fgβ peptide has to be exposed to interact with SdrG, which restricts the geometry and necessitates that the Fgβ peptide is located at the N terminus of the freely diffusing molecule, and that the receptor-of-interest is situated at the C-terminal of the freely diffusing molecule.

Depending on the purpose of the AFM-SMFS measurement, the time-dependent evolution of force experienced by the POI can be controlled by applying various pulling protocols (**Figure 1.1F**). An early method still commonly in use today is referred to as “constant speed” mode, where the distance between the base of the AFM cantilever and the surface (z) is increased at a constant rate. This method only requires open loop positional control of the piezo element in the AFM and is therefore very straightforward to implement, however, open loop operations of piezo elements are generally not recommended due to piezo drift. Other commonly used methods include “force ramp” and “force clamp” modes. In these modes, the photodiode deflection signal is used in a feedback loop to adjust the piezo position such that the POI experiences a tension value set by the experimenter. In force ramp mode, the force is increased linearly with time [72,73]. Force clamp can be viewed as a subtype of force ramp with a ramp velocity equal to zero, and the force applied to the POI is held at a constant value [72,74]. Force ramp and force clamp modes can be used to directly observe force-dependent kinetics of protein unfolding and receptor-ligand complex rupture. Force ramp and force clamp protocols are more prone to external perturbations compared to the constant speed protocol, and the precision of force tuning is limited by many factors, including the response time of the cantilever, drift in the system, and the signal sampling frequencies.

Beyond force ramp, researchers have further developed pulse-chase protocols to study force-induced reactions that can modulate the length of proteins, such as disulfide reduction/oxidation [75–82], domain unfolding [74,83–86], elastic stretching [87], and the reversibility of such reactions. In pulse-chase protocols, force clamp is used to apply an initial force pulse to unfold a protein or a series of fingerprints/POI domains. The force pulse triggers a mechanochemical reaction of interest, for example, domain unfolding or disulfide bond cleavage by nucleophiles. The force is then quenched to zero or other sufficiently low value to allow the reverse reaction to take place. The occurrence of the back reaction is then characterized by applying a second force pulse and determining the fraction of event recurrence.

Another recently developed pulling protocol, zig-zag force ramp, has enhanced the ability of detecting protein unfolding intermediates[88,89]. The zig-zag force ramp protocol uses open loop piezo control to move the AFM tip away from the surface at a constant speed, followed by reversing direction and moving the tip closer to the surface in a two steps forward/one step backward manner. This updown cycle is repeated periodically at a low frequency of ~10 Hz, gradually increasing the distance between the tip and the surface in a stepwise fashion. The zig zag protocol was used to detect many intermediate folding states of bacteriorhodopsin not previously observable by conventional constant speed/force ramp measurements [88].

1.2 Surface chemistry in AFM-SMFS

As discussed above, covalent attachment chemistry enables measurement of highly mechanostable systems and prevents AFM tip clogging, and therefore is desired for AFM-SMFS measurements with receptor-ligand interactions as the objects of study. Chemical functionalization of cantilevers and substrate surfaces is usually required for further immobilization of target proteins. One way to prepare the substrate surface is using gold. Gold is a very stable and inert material and reacts readily with the thiol group on cysteine, forming a gold-sulfur bond so that thiol-containing molecules can be directly immobilized on gold surfaces. Gold-coated substrates and cantilevers are commercially available and also easily prepared. Due to the ease and convenience of this method, many AFM-SMFS measurements especially in the early years were performed using cysteine thiol-gold chemistry, and the technique remains in use today.

Another way to prepare the substrate surface is silanization. Silicon or silicon nitride cantilevers and glass have silanol groups on their surfaces, and these silanol groups can be functionalized with organic silanes carrying amine or carboxyl groups [90]. Aminosilanization has been widely applied and standardized for AFM-SMFS [61]. Further immobilization steps can be performed by reacting amino groups with an N-hydroxysuccinimide (NHS) group. In many of the biological immobilization protocols, aminosilane is the starting layer for further derivatization.

Proper flexible linkers are necessary for passivation of the surface to achieve very low non-specific interactions and for providing proper binding orientation with low steric hindrance away from the surface (**Figure 1.2A**). The most common linkers are PEG (polyethylene glycol) polymers. PEGs are linear, highly flexible with well-characterized elastic behavior, and also commercially available with a wide range of functional groups at the ends including NHS, maleimide and azide groups. PEGs provide well-passivated surfaces and provide functional groups for further derivatization. Some disadvantages of PEG include possible polydispersity and a trans-gauche to all-trans isomerization reaction that sets in around 300 pN of tension. This isomerization can distort contour length analysis for systems at high force [91,92]. More recently, elastin-like polypeptides (ELP) have been developed as linkers [62]. ELPs are composed of a repetitive GXGVP motif, where X can be any amino acid except proline. They are intrinsically disordered and provide added contour length and high flexibility, which are suitable for surface passivation. Also, since ELPs are encoded at the genetic level and expressed in bacteria, they are completely monodisperse with atomically defined lengths and compositions. These features make the use of ELPs a highly accurate measurement technique for analysis of contour length increments [62]. ELPs can be recombinantly expressed with the protein of interest as a fusion protein, or conjugated to the protein of interest via various ligation reactions, as shown in **Figure 1.2A**.

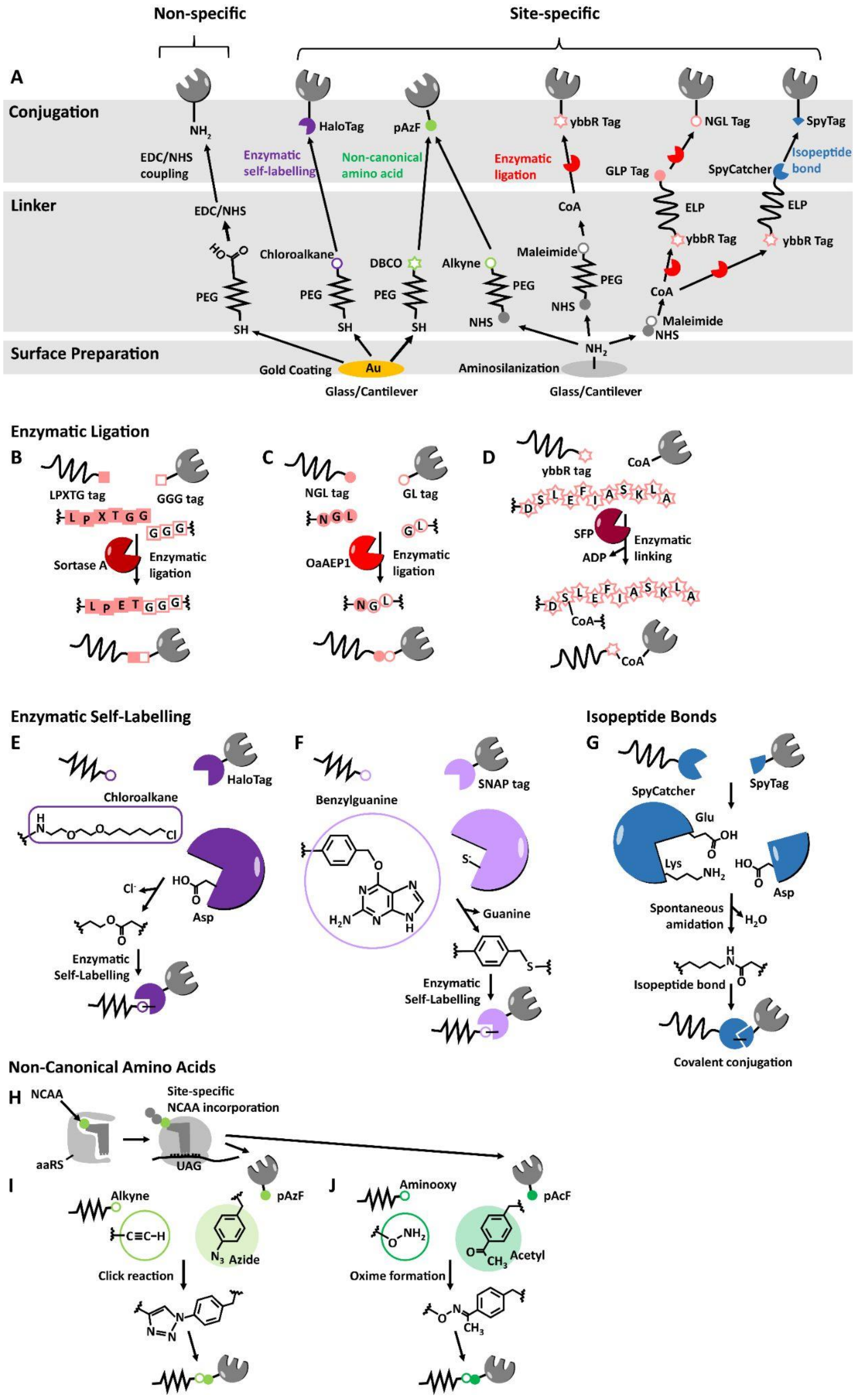


Figure 1.2 Surface chemistry, linkers, and site-specific immobilization methods for SMFS. A: Overview of cantilever and glass preparation for AFM-SMFS. Chemical functionalization of the

substrate surface by gold-coating or aminosilanization is followed by passivation and attachment of a suitable flexible linker (typically PEG or ELP) containing a functional end group. Target molecules can be further immobilized either non-specifically or site-specifically by several strategies: Non-specific immobilization through the reaction between lysine residues on protein surface and NHS ester; Enzymatic ligation using **(B)** LPXTG tag/GGG tag/Sortase A, **(C)** NGL tag/GL tag/OaAEP1, and **(D)** ybbR tag/CoA/SFP; Enzymatic self-labeling using **(E)** HaloTag with chloroalkane derivatives or **(F)** SNAP tag with benzyl group of benzylguanine; Spontaneous isopeptide bonds formation using **(G)** SpyTag/SpyCatcher, SnoopTag/Snoop catcher, and isopeptag/Pilin-C systems; Bioorthogonal click chemistry typically uses non-canonical amino acids, which are incorporated to the protein of interest by **(H)** amber suppression with **(I)** p-azidophenylalanine (pAzF) for click reactions with alkyne or DBCO compounds or **(J)** p-acetylphenylalanine (pAcF) for oxide formation with an aminooxy group. This figure was adapted from Yang *et al.* [13] under a Creative Commons Attribution License (CC BY).

A variety of reactions have been developed and used to covalently immobilize proteins on functionalized surfaces and AFM tips. One key distinction is between methods that allow for site-specific attachment at a known residue in the protein and those that result in a statistical distribution of anchor points within the molecule (i.e. non-specific covalent immobilization). One of the most widely used non-specific immobilization strategies relies on the reaction between the native lysine residues on proteins surface, which carries a free amino group, and functional groups on the surface (e.g. NHS ester) [93,94], as shown in **Figure 1.2A**. This approach is easy to use and does not require any additional modification to the protein of interest. However, it does not provide a precise control of the anchor point and results in heterogeneous force loading geometries in the measurements. Site-specific immobilization, on the other hand, allows precise control over the geometrical loading configuration with dramatic effects on the observed mechanical response of protein domains and receptor-ligand complexes. Site-specific methods can furthermore provide higher yields of usable force-extension curves than non-specific or random covalent immobilization procedures [95]. Site-specific conjugation can also reduce non-specific interactions since contaminating proteins in the sample are not linked to the surface during the conjugation reaction. This can provide higher accuracy, higher yield and generally more reliable results. The most widely used site-specific ligation reactions can be divided into four categories: enzymatic ligation, enzymatic self-labeling, isopeptide bond, and non-canonical amino acid ligation.

1.2.1 Enzymatic ligation tags

Certain enzymes are able to covalently ligate short peptide tags with other peptides or small molecules. Proteins of interest are typically expressed recombinantly with these peptide tags fused at the N- or C- terminus, and conjugated with elastic linkers for surface immobilization or with other protein molecules to assemble a polyprotein.

Sortase A from *S. aureus* recognizes an LPXTG tag at the C-terminus of a target protein, cleaves the bond between threonine and glycine, and ligates the target to a second protein containing an N-terminal oligo G motif (**Figure 1.2B**) [96]. One additional amino acid is required at the end of the LPXTG tag for proper binding of Sortase A. Depending on its accessibility, the N-terminal oligo G motif can contain between one and five glycines, however, three glycines (GGG tag) are generally sufficient. This system has been used for AFM-SMFS for immobilization of protein directly from cell lysate [97] or in systems where an LPETGG tag and GGG tag have been used to assemble polyproteins posttranslationally or to attach high-strength dockerin handles to proteins [63,64,98]. A similar enzymatic ligation system, asparaginyl endopeptidase isolated from the plant *Oldenlandia affinis* (OaAEP1) recognizes an NGL tag at the C-terminus of the target protein, cleaves the bond between asparagine and glycine, and ligates it to an N-terminal GL tag (**Figure 2.1C**) [99].

Another ligation enzyme, 4'-phosphopantetheinyl transferase (SFP) covalently links the first serine residue on the ybbR tag (D<u>S</u>LEFIASKLA) to coenzyme A (CoA), as shown in **Figure 2.1D** [100]. The thiol group on CoA is readily reactive to maleimide coated surfaces, which can be prepared by adding hetero-bifunctional PEG (NHS-PEG-maleimide) or sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC) to aminosilanized surfaces. Therefore, the ybbR tag/SFP system has been widely used in AFM-SMFS with a combination of aminosilanization [61,62,101,102].

1.2.2 Enzymatic self-labeling

In contrast to enzymatic ligation reactions, where the enzyme catalyzes the ligation between another two molecules, enzymes catalyzing self-labeling reactions covalently conjugate itself with the ligand molecule. Examples of enzymatic self-labeling include haloalkane dehydrogenase (HaloTag), which spontaneously forms a covalent ester bond with chloroalkane derivatives (**Figure 1.2E**) [103], and the hAGT or "SNAP" tag, which binds covalently to the benzyl group of benzylguanine, releasing guanine (**Figure 1.2F**) [104]. These self-labeling tags have been successfully used in AFM-SMFS [105–108].

1.2.3 Isopeptide bond

Isopeptide bonds are intramolecular covalent amide bonds formed outside of the protein backbone between amino acid side chains. Isopeptide bonds form spontaneously upon nucleophilic attack of a primary amine from a lysine side chain toward a carboxamide/carbonyl group of asparagine/aspartic acid in close proximity to a catalytic glutamic acid [109]. Proteins having isopeptide bonds have been engineered by dissecting the fold into two fragments and utilizing spontaneous covalent isopeptide bond formation upon fold reconstitution to site-specifically link targets together [110–112]. Isopeptide bond formation is fast, efficient, irreversible, and robust to diverse conditions [111], and is being increasingly used for site-specific immobilization of proteins for AFM-SMFS. The Spytag/Spycatcher system is perhaps the most well known isopeptide bond system, comprising the second immunoglobulin-like collagen adhesin domain of *S. pyogenes* which is stabilized by spontaneous isopeptide formation between Lys and Asp. This fold was rationally engineered and split into two parts: 13 amino acid SpyTag and the remainder of the domain, SpyCatcher (**Figure 1.2G**) [111]. SpyTag can be inserted at the protein terminus or internally in the sequence and remains reactive as long as it is accessible and can form the structure with SpyCatcher. For AFM-SMFS, the SpyTag/SpyCatcher system has been used to immobilize a cellulose binding module onto a cantilever [113], and conjugate an Xmodule-dockerin with ELP linker [114].

1.2.4 Non-canonical amino acids

Non-canonical amino acid (NCAA) incorporation is a sophisticated strategy to introduce new functional groups into proteins [115]. Natural amino acids cover only a very limited range of functional groups and because the same functional groups are repeatedly incorporated into multiple sites in typical protein, their chemical selectivity is poor. These limitations can be overcome by introducing unique bioorthogonal functional groups into target proteins via site-specific NCAA incorporation. To date, a variety of unique amino acids and their orthogonal aminoacyl-tRNA synthetase (aaRS) pairs have been developed [116]. The target NCAA with a unique functional group is recognized by a corresponding aaRS and takes part in the translational machinery at the site of a corresponding codon (typically the amber codon) (**Figure 1.2H**). Depending on the choice of the NCAA, site-specific immobilization for AFM-SMFS can be highly specific, bioorthogonal, and efficient. For example, click chemistry with an azide group is often used. NCAAs having azide groups such as p-azidophenylalanine (pAzF) are incorporated into target proteins at a desired site, and this target protein can be easily immobilized onto alkyne- or DBCO-terminated PEGylated surfaces (**Figure 1.2I**) [117–119]. Also, p-acetylphenylalanine (pAcF) can be introduced for immobilization to aminooxy-terminated PEGylated surface by oxime formation (**Figure 1.2J**) [120,121].

While many of the other methods described require longer peptide tags or require the ligation site to be located at the terminus of the protein, NCAA incorporation changes only a single amino acid and therefore minimally perturbes the target protein. Also NCAA-based attachment is not restricted to the protein terminus but can be achieved in the middle of the amino acid sequence, providing high flexibility in terms of selection of pulling positions for AFM-SMFS.

1.3 Theoretical models of protein unbinding energy profile

The conceptual free energy landscape is a high dimensional surface upon which proteins sample many conformations on their way to the folded/bound state. Due to the importance of protein folding/misfolding, conformational-sampling, and protein-ligand interactions in biological systems, quantifying energy landscapes is highly informative for the understanding of molecular behavior and can inform the development of new therapies. Using AFM-SMFS, we can use external force to perturb the energy landscape and measure the influence of force on transition rates from one state to another, and reconstruct the unfolding/unbinding energy landscape [122].

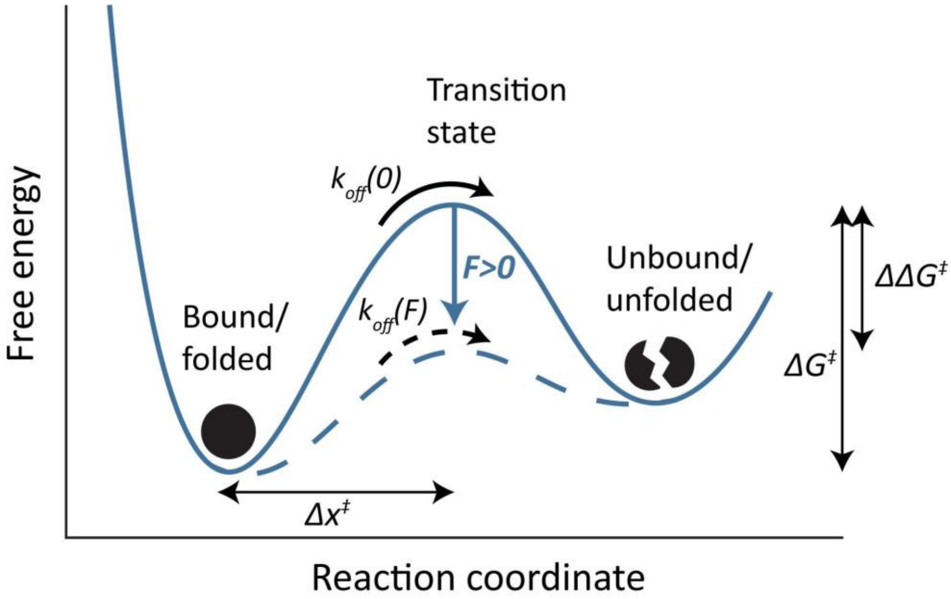


Figure 1.3 1D depiction of protein unbinding energy landscape with a single energy barrier in the presence of force. The protein crosses a single energy barrier to be dissociated or unfolded. In the absence of force, the spontaneous off rate k_0 is dependent on the activation energy $\Delta G^\ddagger$. External forces applied to the protein molecules lower the energy barrier and thus increase the off rate.

A simple and useful way of describing the protein unbinding/unfolding energy landscape is a 1D single energy barrier depiction, as shown in **Figure 1.3**. In this scenario, the protein only populates two states: the bound/folded and unbound/unfolded states, separated by an energy barrier, without additional intermediate states. While the true energy landscape might be more complex and contain additional intermediates, transition states and energy barriers, it is typically difficult to characterize the metastable intermediates and reconstruct the energy landscape in detail [122]. The single energy barrier model, however, provides a reliable approach to reconstructing the energy profile using AFM-SMFS and has been widely used to understand and predict the behavior of proteins.

The single energy barrier and the effect of external forces can be described by several parameters. The bound/folded protein needs to overcome an activation energy ($\Delta G^\ddagger$) to reach the transition state along the chemical reaction coordinate ($\Delta x^\ddagger$), and cross the energy barrier to reach the dissociated/unfolded state. The spontaneous off rate (k_0) is the dissociation/unfolding rate in the absence of force. In AFM-SMFS measurements, external pulling forces are applied to the protein and subsequently perturb the unbinding energy landscape. The pulling forces lower the energy barrier ($\Delta\Delta G^\ddagger$) and increase the off rate ($k_{off}(F)$). Depending on the pulling protocol of AFM-SMFS, the force-dependent energy landscape parameters can be measured based on the bond lifetime at different constant forces (force clamp) or rupture/unfolding force distribution at different force loading rates (force ramp/constant pulling speed). Several theoretical models have been developed to reconstruct the zero-force intrinsic energy landscape using AFM-SMFS data.

1.3.1 Bell-Evans (BE) model

The most widely used theoretical model is the Bell-Evans (BE) model, originally developed to study cell-cell adhesion forces. The BE model assumes that the distance between the initial state (bound/folded state) and the transition state (i.e. $\Delta x^{\ddagger}$) is independent of the force and therefore the external force linearly lowers the energy barrier [123,124]:

$$k_{\mathrm{off}}(F) = Ae^{-\frac{\Delta G^{\ddagger}-\Delta\Delta G^{\ddagger}}{k_B T}} = Ae^{-\frac{\Delta G^{\ddagger}-F\Delta x^{\ddagger}}{k_B T}} \qquad \text{Eq 1.1}$$

,where k_B is the Boltzmann constant, T is temperature, $\Delta G^{\ddagger}$ is the intrinsic height of the energy barrier in the absence of force.

The force-dependent off rate therefore increases exponentially with force:

$$k_{\mathrm{off}}(F) = k_0 e^{\frac{F\Delta x^{\ddagger}}{k_B T}} \qquad \text{Eq. 1.2}$$

Protein unfolding and complex rupture are stochastic processes and the rupture/unfolding force follows a certain probability density distribution. The general expression of the rupture/unfolding force distribution in constant pulling speed/force ramp measurement is:

$$p(F) = \frac{k_{off}(F)}{r}\exp\left(-\int_0^f \frac{k_{\mathrm{off}}(f)}{r} df\right) \qquad \text{Eq. 1.3}$$

, where r is the force loading rate.

Combining Eq. 1.2 and Eq. 1.3, the Bell-Evans model describes the rupture/unfolding force distribution as [125]:

$$p(F) = \frac{k_0}{r}\exp\left(\frac{F\Delta x^{\ddagger}}{k_BT} - \frac{k_0 \cdot k_BT \cdot (\exp(\frac{F\Delta x^{\ddagger}}{k_BT}) - 1)}{r\Delta x^{\ddagger}}\right) \quad \text{Eq. 1.4}$$

Eq. 1.4 also establishes the relationship between the most probable rupture/unfolding force and the force loading rate. The energy landscape parameters, i.e. spontaneous off rate $k_{off}(0)$ and distance to the transition state $\Delta x^{\ddagger}$, can be extracted by linearly fitting the most probable rupture/unfolding force against the logarithm of loading rate with Eq. 1.5 [123,124]:

$$F^{*} = \frac{k_BT}{\Delta x^{\ddagger}}\ln\left[\frac{r\Delta x^{\ddagger}}{k_0k_BT}\right] \qquad \text{Eq. 1.5}$$

, where F^* is the most probable rupture force at different loading rates.

1.3.2 Dudko-Hummer-Szabo (DHS) model

While the BE model has been widely and successfully used by the SMFS community for over two decades [6,14,114,126], its predictions deviate from the behaviors of certain systems due to oversimplifications. Dudko, Hummer and Szabo introduced a modified theoretical model (DHS model) to overcome these limitations [127–129].

As mentioned in section 1.3.1, the BE model assumes that the distance between the initial state (bound/folded state) and the transition state, $\Delta x^{\ddagger}$, is a constant independent of the external force (see Eq. 1.1). However, at very high forces, this distance is expected to be shortened, i.e. the position of the energy barrier moves closer to the initial state [129].

Instead of assuming a constant $\Delta x^{\ddagger}$, the DHS model uses Kramers theory to calculate the force-dependent off rate, as shown in Eq. 1.6 [128]:

$$k_{\mathrm{off}}(F) = k_0\left(1 - \frac{\upsilon F\Delta x^{\ddagger}}{\Delta G^{\ddagger}}\right)^{\frac{1}{\upsilon}-1}\exp\left[\beta\Delta G^{\ddagger}\left(1 - \left(1 - \frac{\upsilon F\Delta x^{\ddagger}}{\Delta G^{\ddagger}}\right)^{\frac{1}{\upsilon}}\right)\right] \qquad \text{Eq. 1.6}$$

, where $\beta^{-1} = k_BT$, and υ represents the shape of the energy barrier. Commonly used υ values include 0.5, which assumes a cusp-like potential and ⅔, which assumes a linear-cubic potential. In addition to k_0 and $\Delta x^{\ddagger}$, the height of the energy landscape $\Delta G^{\ddagger}$ is also considered in the DHS model. Note that when $\upsilon = 1$, Eq. 1.6 is the same as the BE model (Eq. 1.2). When the energy barrier is infinitely high ($\Delta G^{\ddagger} \rightarrow \infty$), the DHS and BE models are also equivalent.

An important limitation of Eq. 1.6 is that the equation is valid only when the force is below a critical force $F_C= \Delta G^{\ddagger}/\upsilon\Delta x^{\ddagger}$. The energy barrier disappears at the critical force, rendering the high energy barrier assumption of Kramers theory invalid and giving an incorrect k_{off} value $k_{off}(F_C)=0$ [128].

Similar to the derivation of Eq. 1.4, combination of Eq. 1.3 and Eq. 1.6 gives the distribution of rupture/unfolding force [128]:

$$p(F) = \frac{k_{off}(F)}{r} \exp\left(\frac{k_0}{\Delta x^{\ddagger} r}\right) \cdot \exp\left[-\frac{k_{off}(F)}{r\Delta x^{\ddagger}} \cdot \left(1 - \frac{\upsilon F \Delta x^{\ddagger}}{\Delta G^{\ddagger}}\right)^{1-\frac{1}{\upsilon}}\right] \qquad \text{Eq. 1.7}$$

Eq. 1.7 also describes the relationship between the mean rupture/unfolding force F_M and the force loading rate [128]:

$$F_{\mathrm{M}} = \frac{\Delta G^{\ddagger}}{\upsilon \Delta x^{\ddagger}}\left[1 - \left(\frac{1}{\Delta G^{\ddagger}} \ln \frac{k_0 e^{-\beta \Delta G^{\ddagger}}}{\Delta x^{\ddagger} l}\right)^{\upsilon}\right] \qquad \text{Eq. 1.8}$$

In contrast to the BE model, which gives a linear relationship between the force and the logarithm of loading rate, as well as the off rate and the logarithm of force, the DHS model describes non-linear force-loading rate and off rate-force relationships.

A practical approach to extracting energy profile parameters from SMFS data using DHS model is to fit the experimentally measured force-dependent off rate against the force. This approach is straightforward for constant force (force clamp) measurements, which directly measures the lifetime (the inverse of off rate) of the system under a given force. For constant speed or force ramp measurements, the force-dependent off rate should be calculated from the rupture/unfolding force distribution. By inverting Eq. 1.3, we have [129]:

$$k_{\mathrm{off}}(F) = \frac{P(F) r}{1 - \int_0^F P(f) df} \qquad \text{Eq. 1.9}$$

Eq. 1.9 is a model-free description of the force-dependent off rate, but this continuous equation is less practical for processing discrete data from experiments. The DHS model uses a discrete form of Eq. 1.9 to transform rupture/unfolding force histograms obtained constant speed and force ramp measurements into force-dependent off rates [129]:

$$k_{\mathrm{off}}(F_{\mathrm{k}}) = \frac{h_{\mathrm{k}} r(F_{\mathrm{k}})}{\left(\frac{h_{\mathrm{k}}}{2} + \sum_{i=k+1}^{N} h_{\mathrm{i}}\right) \Delta F} \qquad \text{Eq. 1.10}$$

,where $k_{\mathrm{off}}(F_{\mathrm{k}})$ is the off rate at the mean force of the kth bin in the rupture/unfolding force histogram, , $r(F_{\mathrm{k}})$ is the average loading rate of the kth bin, and h_{k} is the height of the kth bin. The height is calculated Eq. 1.11:

$$h_{\mathrm{k}} = \frac{C_{\mathrm{k}}}{C_{\mathrm{tot}}\Delta F} \quad \text{Eq. 1.11}$$

,where C_{k} is the number of counts in the kth bin, C_{tot} is the total number of counts in the histogram, and ΔF is the bin width of the histogram. The resulting off rate-force dependency can be fitted using Eq. 1.6 to extract the intrinsic energy profile parameters in the absence of force, including k_0, $\Delta x^{\ddagger}$, and $\Delta G^{\ddagger}$.

1.3.3 Friddle–De Yoreo model

Both the BE model and the DHS model rely on the assumption that the rupture and unfolding processes cross sufficiently high energy barriers and therefore are irreversible [124,128]. However, rebinding and refolding processes are not negligible in the low force regime, giving rise to nonlinearity in force-loading rate spectra. The Friddle-De Yoreo model overcomes this limitation by taking the reversibility of reactions into consideration and adding a rebinding/refolding rate to the model [130].

The unbinding/unfolding rate $k_{\mathrm{u}}(F)$ and the rebinding/refolding rate $k_{\mathrm{b}}(F)$ are given by Eq. 1.12 and Eq. 1.13 [130]:

$$k_{\mathrm{u}}(F) = k_u^0 \exp\left(\frac{F\Delta x^{\ddagger} - \frac{1}{2}k_{\mathrm{c}}\Delta x^{\ddagger 2}}{k_{\mathrm{B}}T}\right) \qquad \text{Eq. 1.12}$$

$$k_{\mathrm{b}}(F) = k_{\mathrm{u}}(F) \exp\left(\frac{\Delta G^{\ddagger} - \frac{F^2}{2k_{\mathrm{c}}}}{k_{\mathrm{B}}T}\right) \qquad \text{Eq. 1.13}$$

, where k_{c} is the spring constant of the AFM cantilever.

The force that gives $k_{\mathrm{u}}(F)=k_{\mathrm{b}}(F)$ is the unique force F_{eq}, at which the rebinding/refolding and the unbinding/unfolding processes are at equilibrium [130,131]:

$$F_{eq} = \sqrt{2k_{\mathrm{c}}\Delta G^{\ddagger}} \qquad \text{Eq. 1.14}$$

The relationship between the mean rupture/unfolding force and the force loading rate r is given by Eq. 1.15 [130,132]:

$$F_{\mathrm{M}} = F_{eq} + \frac{k_{\mathrm{B}}T}{\Delta x^{\ddagger}} \ln\left(1 + \frac{re^{-\gamma}}{k_{\mathrm{u}}(F_{eq})\frac{k_{\mathrm{B}}T}{\Delta x^{\ddagger}}}\right) \qquad \text{Eq. 1.15}$$

, where γ is Euler's constant.

1.4 Protein elasticity

As discussed above, due to the stochastic nature of protein unfolding and complex rupture, barrier crossing is observed to occur over a broad range of positions and forces when measured repeatedly. This makes it difficult to analyze pulling curves using only force-extension coordinates. The free contour length of a polyprotein, however, is a robust statistical parameter that represents the maximal length of physically possible extension in a given folding state. The contour length of the system will theoretically be the same for a given folding state, regardless of the force in the system at any given time. As such it is a robust means to visualize and analyze SMFS data, and can be used to identify unfolding events for a protein of interest. The additional contour length that is added to the tethered polyprotein following domain unfolding can be estimated simply by the length of the polypeptide released from the protein secondary/tertiary structures following protein unfolding. By knowing the amino acid sequence length of a domain, as well as its folded end-to-end length, we can generate expected values for the change in contour length that should be observed when a given domain unfolds. This is given by the equation $\Delta L_{\mathrm{c}} = (0.365\ \mathrm{nm/AA}) \times (\#\ \mathrm{AAs\ in\ POI}) - L_{\mathrm{f}}$, where ΔL_{c} is the expected contour length increment, 0.365 nm is the approximate contour length per amino acid of a protein, and L_{f} is the folded end-to-end length of the domain (typically <5 nm) [133–135].

Since receptor-ligand rupture typically results in loss of the tether between the cantilever and the surface, calculation of ΔL_{c} upon rupture does not have the same physical meaning for receptor-ligand rupture experiments as for domain unfolding experiments, however, ΔL_{c} calculations can be incorporated for fingerprinting of receptor-ligand interaction curves as well. Also, tethered protein receptor-ligand [27,136–138] and DNA [139,140] systems have been reported where the rupture of a molecular interaction results in extension of a flexible tether providing a known contour length increment. Therefore, ΔL_c analysis can be highly applicable not only to domain unfolding studies but also to receptor-ligand rupture experiments.

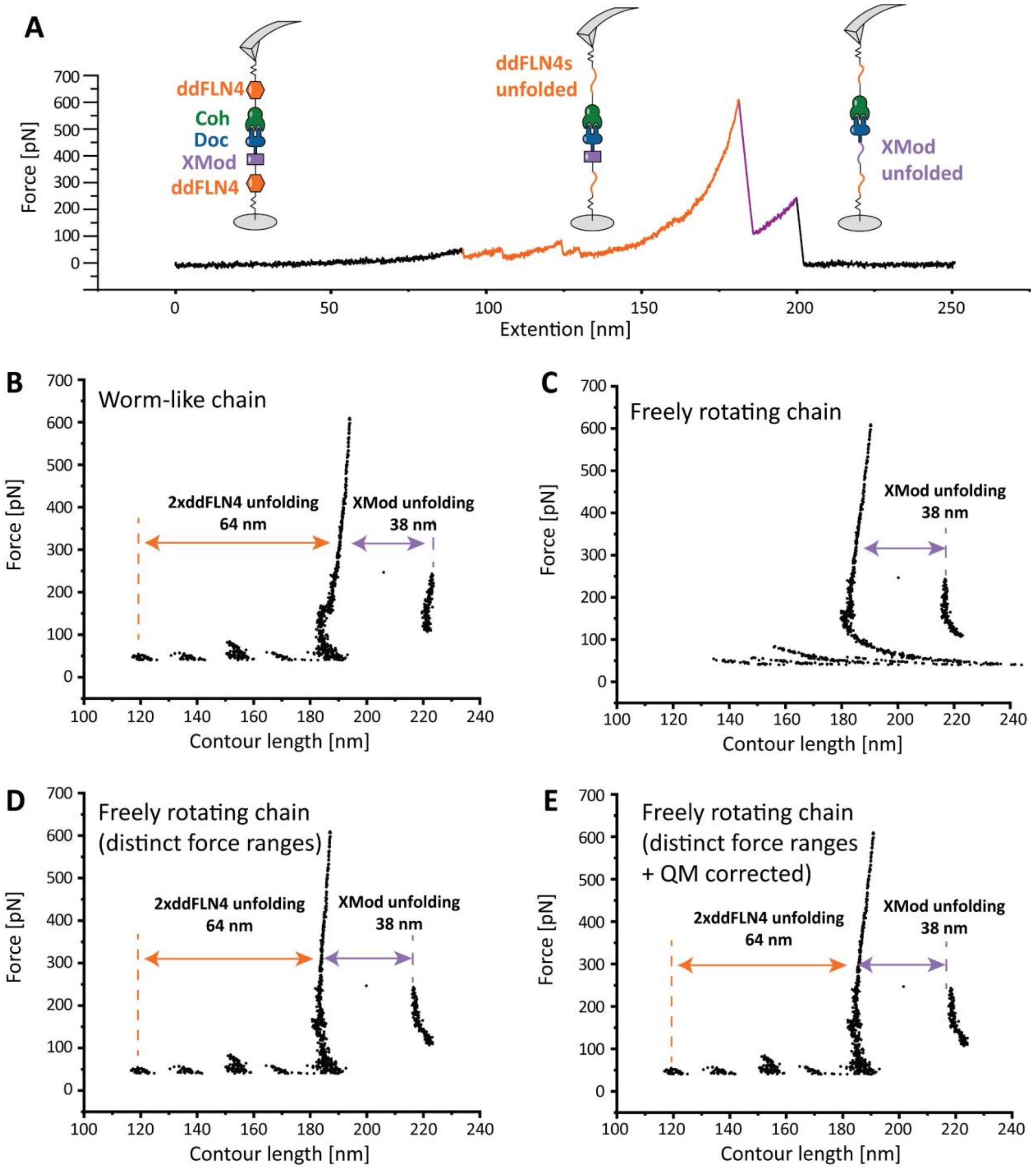


Figure 1.4 Contour length transformation of force-extension curves. A: A force-extension curve measured with *R. champanellensis* XMod-Doc:Coh complex. After initial stretching of the elastic linker (ELP), the ddFLN4 fingerprint domains first unfold (orange), followed by the unfolding of XMod (purple) before the complex ruptures. **B-E:** The force-extension curve in (**A**) was transformed to contour

length space using four elasticity models: worm like chain (**B**), freely rotating chain (**C**), segmented freely rotating chain (**D**) and quantum mechanics corrected freely rotating chain (**E**).

To calculate contour length increments (ΔL_c), polymer elasticity models such as the worm-like chain (WLC) [141], the freely jointed chain (FJC) [142], the freely rotating chain (FRC) [143], or quantum mechanics FRC (QM-FRC) [144] models are applied to transform the force-extension curve using a one-to-one mapping into force-contour length space. The elasticity models can have different force ranges over which they are more accurate.

A widely used model is an interpolation formula of the WLC [141], which is appropriate for ideal stiff chains (Eq. 1.16). This model mathematically describes the stretching of unfolded proteins, DNA, RNA, and other biopolymers reasonably well up to forces around 150 pN. However, for the high force regime over 150 pN, WLC gives less reliable transformation results, as shown in **Figure 1.4B**.

$$F = \frac{k_B T}{4 l_p}\left[\frac{4R}{L} - 1 + \left(1 - \frac{R}{L}\right)^{-2}\right] \qquad \text{Eq. 1.16}$$

, where F is the force, l_p is persistence length, R is the extension and L is the contour length.

To extend the theoretical treatment to higher force regimes, Livadaru *et al.* proposed an FRC model for semiflexible polymer chains made up of discrete segments [143], which has a better performance than WLC in the high force regime, but the results in the low force regime is much less reliable, as shown in **Figure 1.4C**. A more reliable way to use FRC is to divide the data into three distinct force ranges and use WLC to describe the low force regime, and FRC to describe the high force regime, as shown in Eq. 1.17. Note that the original reference [143] has two mistakes, which have been corrected in Eq. 1.17: the x in the first term should be in the denominator instead of numerator, and the number 2 in the third term should be in the numerator instead of in the denominator. The contour length transformation using Eq. 1.17 gives more accurate results than both WLC and FRC used individually when the data spans a wide range of force (**Figure 1.4D**)

$$L = \begin{cases} \frac{3k_BT}{xFa} & \text{for } \frac{Fb}{k_BT} < \frac{l}{p} \\ \frac{x}{1-\left(\frac{4Fl}{k_BT}\right)^{-0.5}} & \text{for } \frac{b}{l} < \frac{Fb}{k_BT} < \frac{l}{b} \\ \frac{x}{1-(\frac{2Fb}{k_BT})^{-1}} & \text{for } \frac{Fb}{k_BT} > \frac{l}{b} \end{cases} \qquad \text{Eq. 1.17}$$

, where $a = b\frac{1+cos\gamma}{(1-cos\gamma)cos\frac{\gamma}{2}}$ is the Kuhn length and $l = b\frac{cos\frac{\gamma}{2}}{|lncos\gamma|}$ is the persistence length.

For the same purpose, quantum mechanical (QM) corrections based on the WLC model were proposed to account for polypeptide backbone stretching in the high-force range of up to two nanonewtons [144]. The QM correction factor is calculated using Eq. 1.18:

$$\frac{L}{L_0} = 1 + \frac{-\gamma_1+\sqrt{{\gamma_2}^2+4\gamma_1F}}{2\gamma_2} \qquad \text{Eq. 18}$$

, where L is the corrected contour length, L_0 is the uncorrected contour length, F is the force, γ_1 and γ_2 are elastic coefficients, γ_1=27.4 nN and γ_2=109.8 nN (for peptides) [144].

Figure 1.4E shows the contour length transformation using the FRC model with quantum mechanical correction (QM-FRC). At this force, the corrected and uncorrected contour lengths differ by no more than 2%. For larger forces, the difference between the corrected and uncorrected contour lengths is higher (e.g. ~6% for force up to 2 nN).

Depending on the solvent environment, the effects of monomer side chains may become evident in the elastic response of individual biopolymers. A recent study showed that a more consistent fitting could be achieved using a new TSQM model that upgrades the previous modeling work with structure-relevant terms [145]. Given the importance of elastic stretching behavior in AFM-SMFS, isomerization reactions within monomer units of mixed synthetic/protein polymer systems can also become problematic, and blur contour length histograms. Intrinsically disordered elastin-like polypeptides (ELPs) have been incorporated as linkers to avoid the trans-gauche isomerization of PEG-linkers that occurs around 300 pN [62,91,92]. However, ELPs are also limited by the isomerization of prolines at 200-300 pN [146].

1.5 Catch bonding

As discussed in section 1.3, external pulling forces typically lower the energy barrier of protein-ligand dissociation and decrease the lifetime of protein-ligand complexes; such bonds are called slip bonds. However, it has been shown that for a limited number of protein-ligand complexes, primarily cell-cell adhesion proteins, the bond lifetime is found to increase with force over a certain force range. This counterintuitive behavior is called catch bonding. Catch bonds allow the cells to resist high shear stress, while maintaining the flexibility to detach from each other, migrate and diffuse under lower shear stress. SMFS measurements are able to distinguish catch bonds from slip bonds in both constant speed/force ramp mode and force clamp mode, and help explain the molecular mechanisms of catch bonding.

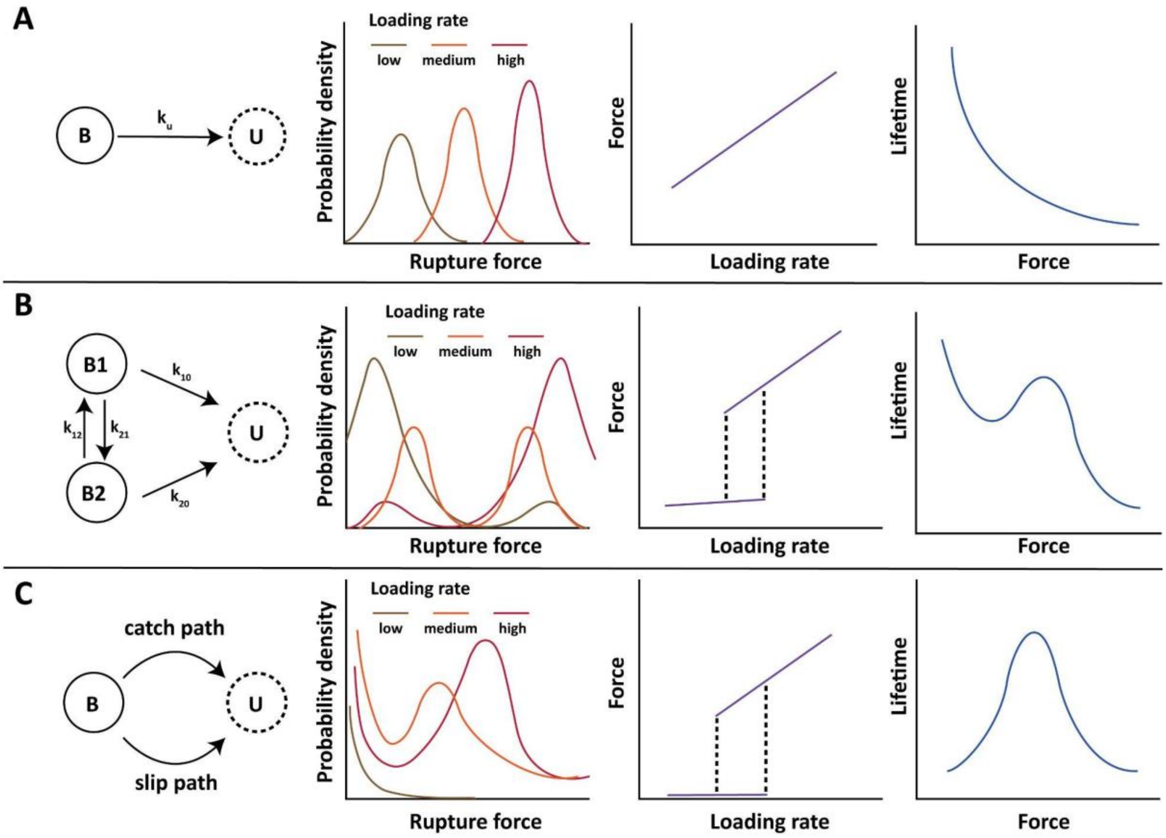


Figure 1.5 Comparison of slip bond, two-state two-pathway catch bond and one-state one-pathway catchbond. A: The behavior of slip bonds can be characterized by Bell-Evans model. In constant speed/force ramp measurements, the rupture force shows linear dependency on the logarithm of loading rate. In force clamp measurements, the bond lifetime decreases exponentially with force. **B:** Two-state

catch bonds have two binding states that are typically interchangeable. These two binding states have different off rates and rupture forces, and can be distinguished in constant speed/force ramp measurements, giving bimodal distributions of rupture force. The rupture events preferably populate the high force population at high loading rates. In force clamp measurements, the lifetime increases with force in a certain force regime. **C:** One-state two-pathway catch bonds have two possible dissociation paths: catch path and slip path. In constant speed/force ramp measurements, the two pathways also give rise to a bimodal rupture force distribution with the low force population centered at zero. In force clamp measurements, the lifetime increases with force at low forces when the catch pathway is dominant. When the force exceeds a critical force and the slip pathway becomes dominant, the lifetime decreases with force and the system behaves like a slip bond.

1.5.1 Slip bond

The behavior of single slip bonds under external pulling forces can be predicted very well by theoretical models such as Bell-Evans and Dudko-Hummer-Szabo models (see section 1.3). The following discussion is based on the simpler Bell-Evans model. As shown in **Figure 1.5A**, When measured with constant pulling speed or force ramp modes, slip bonds give rise to a single population of rupture force distribution, and the most probable rupture force increases linearly with the logarithm of force loading rate (see Eq. 1.5). The lifetime of slip bonds (i.e. the inverse of off rate) decreases exponentially with force (Eq. 1.2), which can be clearly shown in force clamp measurements. The interaction between biotin and streptavidin is a classical example of a slip bond [70,147].

1.5.2 Two-state two-pathway catch bond

One of the most widely used physical models to describe the underlying mechanism of catch bonding is the two-state two-pathway model. The receptor-ligand complex has two bound states and each bound state can dissociate and enter the unbound state with different off rates, i.e. each bound state behaves like a slip bond when considered separately. These two bound states have different off rates and thus also different rupture forces, meaning that they can be resolved using constant pulling speed/force ramp mode and give rise to a bimodal distribution of rupture force, as shown in **Figure 1.5B**. Unlike systems with multiple non-interchangeable and independent bound states, for example Doc:Coh complexes with dual-binding modes [148], the two bound states of two-state catch bonds are typically interchangeable, and the rate of entering one state from the other is dependent on the external force [149]. At high forces, the transition from the strong state (high rupture force) to the weak state (low rupture force) is suppressed and the rate of the reverse process is increased. As a result, the rupture force distribution of two-state catch bonds primarily populates the low-force population at low loading rate, and switches to the high-force population as loading rate increases. It is important to distinguish this low force-high force transition from the conventional loading rate dependency of rupture forces, where the most probable rupture force continuously increases with loading rate. The latter is a behavior of slip bonds that is well-described by the Bell-Evans model (see **Figure 1.5A** and Eq. 1.5). This behavior also distinguishes two-state catch bonds from independent two-state systems like dual-binding mode of Doc:Coh, where the prominence of each rupture force population is not dependent on the loading rate.

The two-state catch bond can also be observed using constant force measurements (**Figure 1.5B**, forth panel). When measured with force clamp mode, the transition from the weak state to the strong state is dominant in a certain force regime, giving rise to an increased lifetime with increasing force. It is worth noting that, above a critical force, the force-induced dissociation again dominates and the lifetime of the complex decreases with force like a slip bond. The catch bond behavior of several cell adhesion proteins can be well-described by the two-state two-pathway model [150–152].

1.5.3 One-state two-pathway catch bond

Another kinetic model describing catch bonds is the one-state two-pathway model, which has been used to describe selectin and myosin catch bonds [29,153–155]. In this model, the receptor-ligand complex has only one bound state, but dissociates in two pathways, a catch pathway and a slip pathway (see **Figure 1.5C**, first panel). The distance from the bound state to the transition state (i.e. $\Delta x^{\ddagger}$) in the catch pathway is negative, meaning that the external force increases, rather than decreases, the unbinding energy barrier [153]. The energy barrier of the catch pathway is relatively low at zero force, but increases with force, giving rise to longer lifetimes [156]. The other pathway (slip pathway) has the same behavior as a slip bond. Its zero-force energy barrier is higher than the catch pathway but is lowered by external force, giving rise to exponential decrease of lifetime (see Eq. 1.2). The catch pathway dominates the system at low force, and is inhibited by higher external force, resulting in the slip pathway being dominant at high forces [149]. Similar to the two-state catch bond, the coexistence of two pathways also gives rise to a bimodal distribution of rupture force in constant speed/force ramp measurements, and a transition to the high force population can be observed at high loading rate (**Figure 1.5C**). The rupture force population of the catch pathway is centered at zero and only the positive half of the peak can be observed [156]. When measuring one-state two-pathway catch bonds using force clamp mode, an increase in lifetime can be observed at low clamping forces, where the catch pathway is dominant, and the slip pathway dominates when the force exceeds a critical force, after which the bond lifetime decreases with force, similar to a slip bond.

1.6 Cellulosomal network and cohesin-dockerin complexes

Cellulose is one of the most abundant and prominent components of biomass [157] and can be degraded by a variety of fungi and bacteria to enter the carbon cycle [158]. Many cellulose-degrading bacteria rely on cellulosome, a large, complex, and multi-branched extracellular protein network to bind and digest cellulose fibers. Cellulosomes are composed of subunits called scaffoldins [159]. As shown in **Figure 1.6A**, the scaffoldin proteins typically contain multiple copies of cohesins (Cohs), a C-terminal dockerin (Doc), as well as other functional domains including cellulose binding modules (CBMs) and catalytic enzymes that digest the cellulose fibers [157,160,161]. The Cohs and Docs form high-affinity and highly mechanostable interactions, which crosslink the scaffoldin subunits to form a large cellulosomal network, and certain Doc:Coh pairs are responsible for attaching the extracellular cellulosome to the bacterial surface [159,162]. Therefore, the Doc:Coh complexes form the primary backbone of the cellulosome and play an important role in maintaining the integrity and function of cellulosomes.

Based on sequence similarities and binding specificities, Doc:Coh pairs can be classified into three types: type I, type II, and type III [163]. **Figure 1.6B-D** shows representative structures of Doc:Coh complexes of three different types [5,164,165]. The Docs are calcium binding proteins and contain two calcium binding loops. The binding of calcium induces Doc conformation change and is necessary for Docs to recognize and bind Cohs [166].

Certain Docs contain an Ig-like X-module (XMod) at their N-terminus (**Figure 1.6C and D, purple**). The XMods are known to enhance the affinity and mechanical stability of the Doc:Coh complex [6,164,165] and are important for Doc:Coh pairs under high shear stress, for example the Doc:Coh interactions at the cellulosome-cell wall interface.

The molecular mechanism of Doc:Coh interaction is essential to understand how complex cellulsosomal networks are assembled, resist external shear stress, and maintain highly dynamic and flexible structure at the same time to quickly respond to environmental changes. Among the various available biochemical and biophysical approaches, AFM-SMFS is a powerful tool to study Doc:Coh systems due to its capability to measure highly mechanostable proteins and gain a mechanistic insight into complex biological systems. AFM-SMFS has been used to study the mechanisms of highly mechanostable type III XMod-Doc:Coh systems and demonstrate that they can withstand a force up to 600 pN to 1 nN [6,67]. As a single-molecule approach, AFM-SMFS is uniquely suited to resolve distinct binding conformations of certain Doc:Coh complexes, a unique behavior known as ‘dual-binding mode’ [148]. In addition to the dissociation of Doc:Coh complexes, the mechanical properties of single Coh domains are also very interesting because tandem repeats of Cohs are the primary component of scaffoldin backbone and the stability of scaffoldins is highly dependent on the mechanostability of Cohs. AFM-SMFS was used to demonstrate the distinct mechanical stabilities of Cohs located at different positions of scaffoldins. The bridging Cohs between the CBM and the anchoring Doc are under higher mechanical stress and therefore are more resistant to forces compared to the hanging Cohs (see **Figure 1.6A**) [161]. Thanks to the high mechanical stability of Doc:Coh complexes, they are also widely used as a high-force handle in AFM-SMFS measurements to mechanically unfold other protein domains [10,64,167].

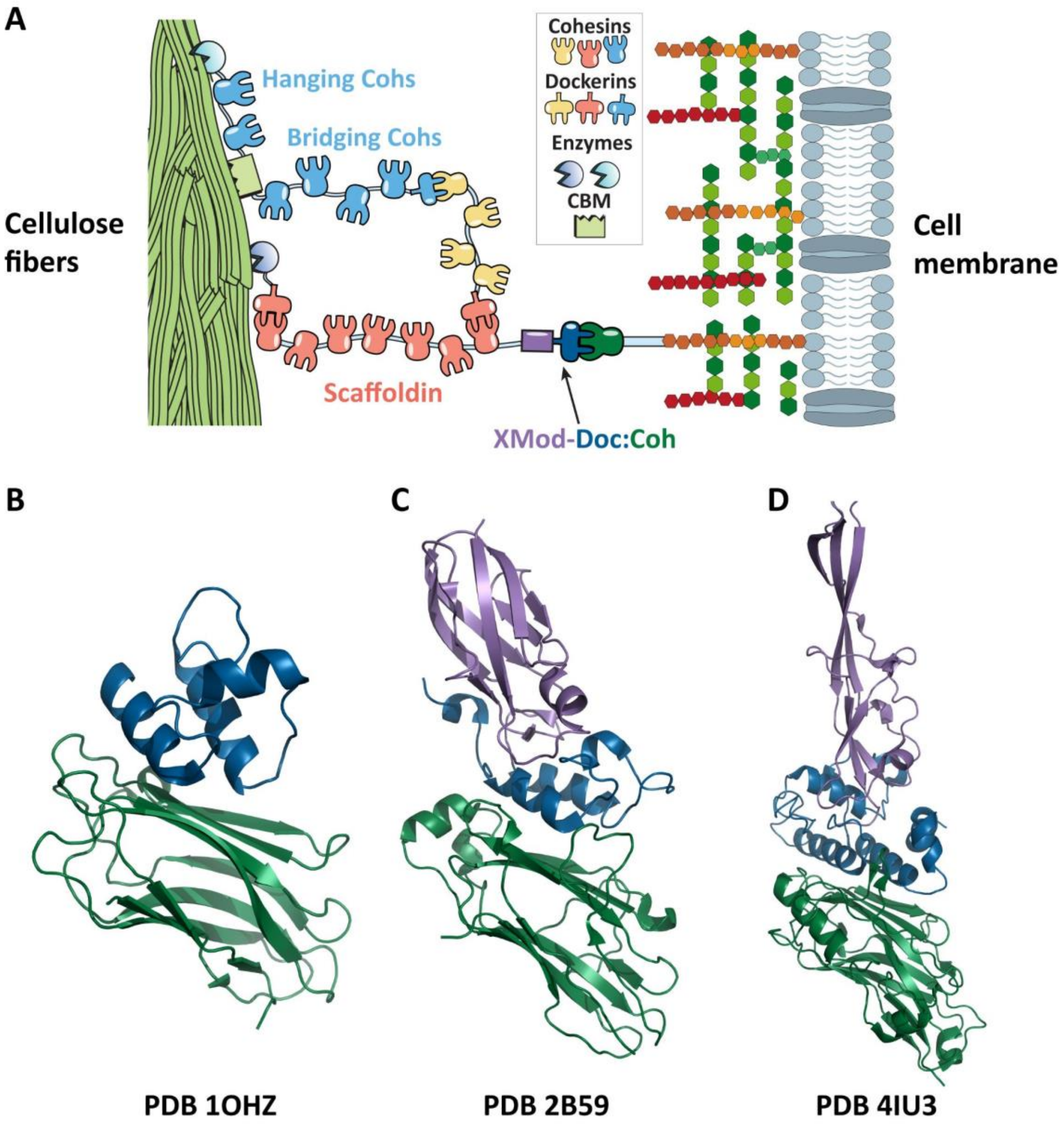


Figure 1.6 Cellulosome structure and representative Doc:Coh complex structures. A: Typical cellulosome scaffoldin subunits consist of cohesins (Cohs), dockers (Docs), cellulose binding modules (CBMs), and catalytic enzymes. Certain Docs have an adjacent X-module (XMod) at their N-terminus. Cohs located between the CBM and anchoring Doc of a scaffoldin are called bridging Cohs, which are subject to higher shear stress and found to be more mechanostable than the hanging Cohs [161]. **B-D**: Structures of type I (**B**), type II (**C**) and type III (**D**) Docs (blue) in complex with Cohs (green). The type II and III Docs shown here contain an X-module (XMod, purple) at the N-terminus [5,164,165]. Panel A was adapted from Liu *et al.* [168] under a Creative Commons Attribution 4.0 International License.

1.7 Mechanical anisotropy in biomolecules

The residues in protein domains and complexes form a complex network of interactions including covalent bonds, hydrogen bonds, hydrophobic interactions, van de Waals interactions, and electrostatic interactions. This network is anisotropic and therefore responds to external forces differently when forces are applied from different directions. As a result, force-induced protein unfolding and unbinding pathways are dependent on the pulling geometry. AFM-SMFS is capable of precisely controlling the force loading geometry on biomolecules, and therefore has been successfully used to demonstrate the anisotropic response to pulling forces of single protein domains, protein-ligand complexes, as well as other biomolecules, for example nucleic acids.

1.7.1 Single protein domains

Ubiquitin is one of the naturally occurring systems where the mechanical stability of proteins is regulated by force loading geometry. The ubiquitin monomers in ubiquitin chains are linked either between the N- and C-termini or the C-terminus and one of the four possible lysine residues via an isopeptide bond [169]. AFM-SMFS measurements showed that single ubiquitin domains can resist over three fold higher forces when pulled from the N- and C-termini, compared to Lys48-C-terminus linkage [170]. Another study used AFM to unfold a β-sheet protein, E2lip3, with two different pulling geometries, which gave rise to unfolding forces that differed by an order of magnitude [171]. Similar mechanical anisotropy was also shown in GFP. Single GFP domains were unfolded by pulling from five different directions and exhibited distinct unfolding forces (~5 fold difference) [56]. Combination of experimental and theoretical approaches has provided mechanistic insight into the anisotropic response of proteins to pulling forces. The anisotropic network model (ANM) was used to calculate the effective spring constant between different anchor residues on the three aforementioned proteins (ubiquitin, E2lip3, and GFP) and showed correlation between the calculated spring constant and measured unfolding forces [172]. Molecular dynamics (MD) simulations can help identify the protein unfolding pathways and, combined with experiments, provide information on the underlying mechanisms leading to the mechanical anisotropy of proteins [170,173,174].

1.7.2 Protein-ligand complexes

Directionally dependent responses to external forces have been shown not only in single protein domains, but also in protein-ligand complexes. The *R. flavefaciens* XMod-Doc:Coh complexes *in vivo* are anchored to the scaffoldins from the N-terminus of XMod-Doc and to the cell wall from the C-terminus of Coh. AFM-SMFS measurements resembling the native force loading geometry showed high rupture force up to ~600 pN [6]. A non-native pulling geometry, where the force is loaded to the N-terminus of Coh, gives rise to an additional low force rupture pathway at ~100 pN. Steered molecular dynamics (SMD) simulations were used to identify the force propagation route, which demonstrated the mechanisms of the low force pathway at the non-native pulling geometry, and explained how the complex achieves high mechanostability at native pulling geometry [65]. Another system widely used by the SMFS community, the biotin:streptavidin complex, was also shown to exhibit significant mechanical anisotropy. The monovalent streptavidin (mSA):biotin complex has ~2 fold higher rupture force when pulled apart from the C-terminus of mSA, compared to the N-terminus pulling geometry [70], The tetrameric streptavidin in complex with biotin also exhibited distinct force-induced unbinding pathways when force was applied to the complex from different streptavidin subunits [71]. SMD simulations also contributed to understanding the structural basis of mechanical anisotropy in biotin:streptavidin systems.

1.7.3 Nucleic acids

Similar to proteins, nucleic acids are also capable of forming complex secondary and tertiary structures and exhibit anisotropic response to forces. Double-stranded DNA can be unfolded by force in two directions: unzipping, where the base paired hydrogen bonds are broken in series, and shearing, where the hydrogen bonds are broken in parallel. The shearing force is typically larger (>50 pN) than unzipping force (~30 pN), showing a direction-dependent response to forces [175–177]. In terms of RNA molecules, AFM-SMFS measurements demonstrated the magnesium-dependent mechanical anisotropy of a three-way junction pRNA. In the presence of magnesium, the unfolding force of the RNA can be different by ~5 folds depending on the pulling geometry, but the mechanical anisotropy is lost in the absence of magnesium [178]. Since magnesium affects nucleic acid folding and structure, magnesium-dependent mechanical anisotropy was also observed in other RNAs [179].

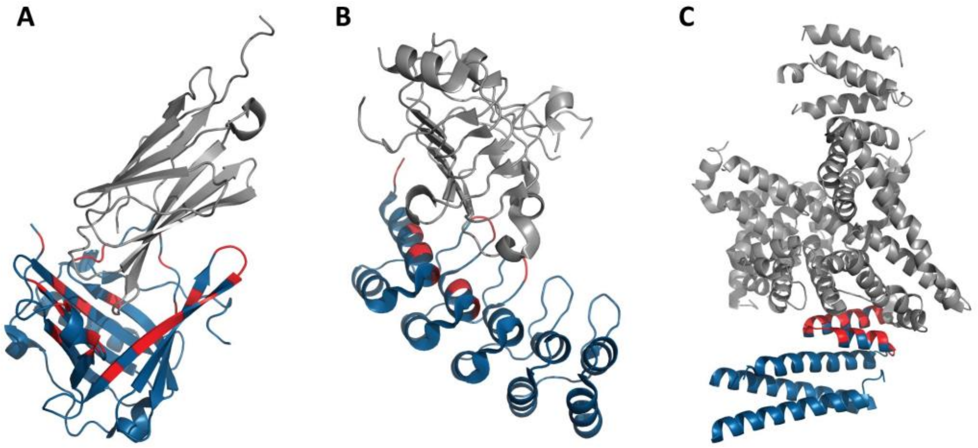
1.8 Non-immunoglobulin scaffolds and nanoparticles with biomedical applications

Antibody-based therapeutics have been successfully used to target various diseases including, but not limited to, cancer [180–182], HIV infection [183], sickle cell disease [184], and asthma [185]. The high affinity and specificity of antibodies towards their targets make them very useful tools for targeted drug delivery and diagnostic imaging, in the form of antibody-drug conjugates [186], or antibody-coated nanoparticles [187–189]. Antibodies can also directly block or neutralize their targets, without being conjugated with other molecules or nanoparticles [190–192]. While antibody-based therapy has achieved great success, there are certain disadvantages of immunoglobulin (Ig) scaffolds that limit in some cases the efficacy, stability, or developability of therapeutic antibodies based on full-length Igs. Igs are large proteins (~150 kD) with disulfide bonds and post-translational modifications (e.g. glycosylation). Their large size limits the efficiency of tissue penetration, which is a major limiting factor in tumor therapy [193], and decreases the clearance rate from the system, potentially leading to systematic accumulation [194]. The disulfide bonds and glycosylation pose challenges to the production of therapeutic antibodies. While the advancement of eukaryotic expression systems have made it easier and cheaper to produce full-length antibodies, production and analysis of uniformly glycosylated antibodies is still a challenge [195,196].

One approach to overcoming the aforementioned limitations is to develop therapeutic non-Ig scaffolds (alternative scaffolds). Similar to antibodies, non-Ig scaffolds specifically bind their targets with high affinity and can be potentially used for targeted drug delivery and imaging. However, they are typically smaller in size (no larger than 20 kD), have simpler structure with none or a few disulfide bonds, and lack glycosylation. These features make them easier to engineer and produce, and the small size, which leads to efficient tissue penetration, is very beneficial for tumor therapy [197].

1.8.1 Representative non-Ig scaffolds

Over one hundred non-Ig scaffolds have been developed so far [198]. They have diverse structures, a large variety of targets, and different routes of development. This section discusses three representative non-Ig scaffolds with different sources of parent protein: naturally occurring protein, consensus design, and de novo design, as shown in **Figure 1.7**.



Anticalin
Target: CTLA-4
Molecular weight: 22 kD
Origin: human lipocalin
Selection: phage display
PDB code: 3BX7

DARPin
Target: HER2
Molecular weight: 19 kD
Origin: ankyrin repeat protein
Selection: phage/ribosome display
PDB code: 4HRL

Alphabody
Target: MCL-1
Molecular weight: 15 kD
Origin: *de novo* design
Selection: phage display
PDB code: 6ZL1

Figure 1.7 Representative non-Ig scaffolds. The non-Ig scaffolds are shown in blue with the target binding residues labeled with red. Their targets are shown in grey. The three non-Ig scaffolds have different origins: anticalin **(A)** backbone was based on a naturally occurring protein, human lipocalin [199], DARPin **(B)** backbone was designed using consensus design, based on ankyrin repeat proteins [200], and alphabody **(C)** was developed using computer *de novo* design [201].

The most common approach to developing non-Ig scaffolds is to use a naturally occurring protein (usually human protein) as the parent protein, which typically binds a natural target *in vivo*. A random library is created to randomize the target binding residues of the parent protein. The library is selected using directed evolution techniques such as phage display, yeast/bacterial surface display and ribosome display [195] to develop high-affinity non-Ig scaffolds against a desired target in iterative selection-randomization cycles [202]. While affinity is the primary selection criterion in directed evolution of non-Ig scaffold, a number of other biochemical and biophysical properties also need to be considered, including specificity, solubility, thermostability, and, in the case of clinical applications, immunogenicity [195]. Anticalin is one of the non-Ig scaffolds developed using this approach. The parent protein is naturally occurring human protein lipocalin, which is involved in transportation of small molecules *in vivo*. **Figure 1.7A** shows one of the anticalins (blue). The ligand binding residues (red) were randomized to create a library, which was selected to target human T cell receptor CTLA-4 (grey) [199]. A number of other non-Ig scaffolds also have a naturally occuring parent protein, for example affibody, adnectin, and centyrin [203–205].

Consensus design strategy is also widely used in the development of non-Ig scaffolds. This approach takes a series of homologous protein sequences into consideration. The conserved residues in these sequences are considered important for the protein structure and kept unchanged in further directed evolution steps [206]. The variable residues, on the other hand, can be randomized and selected to bind certain targets. The conserved and variable residues can be identified by aligning the homologous sequences. Designed ankyrin repeat proteins (DARPins) are a successful example of consensus design. The sequences of naturally occurring ankyrin repeat proteins are aligned to identify the structural framework residues (conserved) and potential target binding residues (variable) to yield a consensus sequence [207,208]. A library is built by randomizing the variable residues and selected using ribosome display or phage display. An example of DARPin is shown in **Figure 1.7B** [200]. Other non-Ig scaffolds developed using consensus design are mostly repeat proteins, including armadillo repeat protein and repebody [209,210].

Computer *de novo* design also contributed to the development of non-Ig scaffolds. Alphabody (**Figure 1.18C**) is one of the most widely used *de novo* designed non-Ig scaffolds and has been successfully used for tumor drug delivery [201]. Recently, another *de novo* designed protein scaffold was developed to bind the SARS-CoV-2 spike receptor binding domain (RBD) with picomolar affinity [211].

1.8.2 Diagnostic and drug-delivery nanoparticles

Nanoparticles have been widely used in diagnostic imaging and targeted drug delivery, especially in tumor therapy [212–214]. The nanoparticles need to be transported to and accumulated at the target cells to be functional. One of the targeting strategies is to coat the nanoparticles with targeting moieties, which specifically bind cell surface receptors of target cells [215]. Antibodies and non-Ig scaffolds are useful targeting moieties because of their high affinity and specificity towards target molecules. Antibody-nanoparticle conjugates have been widely used in drug delivery [188,216], *in vivo* imaging [217,218], as well as *in vitro* diagnostics [219,220]. Non-Ig scaffold-coated nanoparticles are also widely used in biochemical applications. Nanobodies targeting the tumor marker epidermal growth factor receptor (EGFR) have been used to deliver drugs carried by nanoparticles [221] or nanogels [222] to tumor cells, and affibody-nanoparticle conjugates were used for *in vivo* imaging of tumors [223,224].

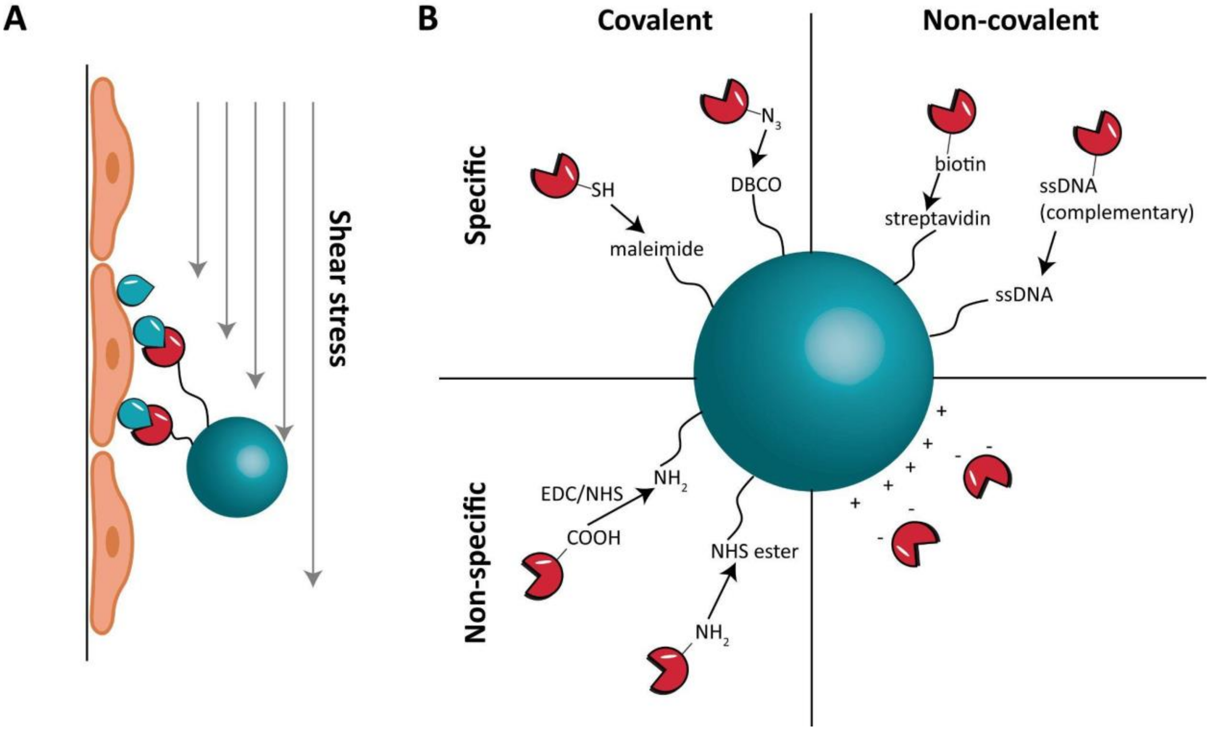


Figure 1.8 Binding of nanoparticles to target cells and nanoparticle coating strategies. A: Nanoparticles coated with targeting molecules bind the target cell through the interaction between the targeting molecule and the cell surface receptor. The nanoparticles are constantly under shear stress when binding the target cells. **B:** The targeting molecules can be conjugated with the nanoparticles covalently or noncovalently. Another way to categorize the conjugation strategies is based on site-specificity. The target cells in panel A were created with BioRender.com.

As shown in **Figure 1.8A**, when the nanoparticles coated with antibodies/non-Ig scaffolds bind the target cells under flow, they experience shear stress, which tends to dissociate the nanoparticles from the target cell surface [225–227]. Therefore, the interaction between the targeting molecule and the cell surface receptor, as well as the interaction between the target molecule and the nanoparticle, should have sufficient mechanical stability to resist the shear stress.

A number of different strategies have been developed to conjugate antibodies/non-Ig scaffolds with nanoparticles (see **Figure 1.8B**). The non-covalent conjugation strategies include simple physical adsorption [228], biotin-avidin interaction [229], and complementary DNA molecule hybridization [230]. However, one limitation of the non-covalent conjugation strategies is the low mechanical stability between the coating molecule and the nanoparticle. Covalent conjugation strategies are therefore preferred in this context. The most widely used covalent coupling reaction is the reaction between the amino groups on the antibody or non-Ig scaffold protein surface and the N-hydroxysuccinimide (NHS) ester group on the nanoparticle (the NHS ester can be generated from carboxyl groups using EDC/NHS reaction) [216,231,232]. However, this strategy is not site-specific and the orientation of the targeting molecule cannot be controlled, resulting in decreased efficacy of the nanoparticles [233]. Thiol-maleimide reaction, on the other hand, provides better site-specificity due to the low abundance of cysteines in natural proteins and is also widely used in protein-nanoparticle conjugates [234,235]. Since this approach is still limited by the native cysteines on the coating proteins, a better approach is to use bioorthogonal click chemistry, for example dibenzocyclooctyne (DBC)-azide reaction, which provides precise control over the orientation of proteins on the nanoparticle surface [236].

Chapter 2 Mechanics of bacterial adhesion complexes in the human gut

This chapter was published in:

Liu, Z., Liu, H., Vera, A.M., Bernardi R.C., Tinnefeld, P., and Nash, M.A. High force catch bond mechanism of bacterial adhesion in the human gut. *Nat Commun* 11, 4321 (2020)

Bacterial colonization of the human intestine requires firm adhesion of bacteria to insoluble substrates under hydrodynamic flow. Here we report the molecular mechanism behind an ultrastable protein complex responsible for resisting shear forces and adhering bacteria to cellulose fibers in the human gut. Using single-molecule force spectroscopy (SMFS), single-molecule FRET (smFRET), and molecular dynamics (MD) simulations, we resolve two binding modes and three unbinding reaction pathways of a mechanically ultrastable *R. champanellensis* (*Rc*) Dockerin:Cohesin (Doc:Coh) complex. The complex assembles in two discrete binding modes with significantly different mechanical properties, with one breaking at ~500 pN and the other at ~200 pN at loading rates from 1-100 nN s^{-1}. A neighboring X-module domain allosterically regulates the binding interaction and inhibits one of the low-force pathways at high loading rates, giving rise to a catch bonding mechanism that manifests under force ramp protocols. Multi-state Monte Carlo simulations show strong agreement with experimental results, validating the proposed kinetic scheme. These results explain mechanistically how gut microbes regulate cell adhesion strength at high shear stress through intricate molecular mechanisms including dual-binding modes, mechanical allostery and catch bonds.

2.1 Introduction

When cells adhere to surfaces under flow, adhesion bonds at the cell-surface interface experience mechanical tension and resist hydrodynamic drag forces. Because of this mechanical selection pressure, adhesion proteins have evolved molecular mechanisms to deal with tension in different ways. Most bonds not involved in force transduction *in vivo* have lifetimes that decay exponentially with applied force, a behavior well described by the classical Bell-Evans slip bond model [123,124,237]. Less intuitive are catch bonds [29,151,156,238], which are receptor-ligand interactions that serve as band pass filters for force perturbations, becoming stronger with applied force and weakening when force is released. When probed in constant force mode, the lifetime of a catch bond will rise as the force setpoint is increased. When probed in force ramp mode or constant speed mode, catch bonds typically give rise to bimodal rupture force distributions [239,240]. Different kinetic state models and network topologies can be used to describe catch bonds [241]. For example, mechanical allostery models such as the one-state two-pathway model [242], or two independent sites model [150,239] have been applied to mathematically describe catch bond behavior [150,152,156,239].

The *R. champanellensis (Rc)* cellulosome [160,243] is a bacterial protein complex found in the human gut that adheres to and digests plant fiber [244]. The large supramolecular complex is held together by dockerin:cohesin (Doc:Coh) interactions [245] which comprise a family of homologous high-affinity receptor-ligand pairs. A proposed topology of the *Rc* cellulosome is shown in **Figure 2.1** [160,243]. We focused here on the mechanical stability of the complex formed between DocB located at the C-terminus of scaffoldin B, and CohE which is covalently attached to the peptidoglycan cell wall. This complex anchors the base of the cellulosome network to the cell surface, and is required *in vivo* to maintain cell adhesion under hydrodynamic flow and applied shear stress. Single-molecule interactions between fiber substrates and cellulose binding modules (CBMs) have been reported to rupture at moderate forces (~50 pN) [113,246]. However, due to the multivalency of interactions, the mechanical requirements on the anchoring complex are more stringent. At the same time, the complex must be able to release cellulosome complexes to facilitate dynamic niche exploration and cellulosome shedding in response to new substrates. These seemingly mutually exclusive requirements of dynamics and strong adhesion motivated us to understand the molecular mechanism of the anchoring complex in more detail.

Rc-DocB belongs to the type III dockerin family and is the only *Rc* dockerin with an adjacent X-module domain (XMod; **Figure 2.1a**, purple) [160,243]. XMod was previously shown in a related system from *Ruminococcus flavefaciens* (*Rf*) to stabilize Doc and increase the mechanical stability of the XMod-Doc:Coh complex [6,65]. Furthermore, the binding helices of *Rc*-DocB are highly symmetric (**Figure 2.S1**, helices 1 and 3), a feature observed in types I and II Doc:Coh complexes that exhibit dual-binding modes [148,247–251]. Dual binding modes arise in certain Doc:Coh complexes where the complex populates two distinct binding conformations involving different sets of binding residues on Doc recognizing the same residues on Coh. In these systems, due to structural and sequence symmetry, Doc can be rotated 180° with respect to Coh to form an alternative bound conformation. However, dual binding mode behavior has not been previously reported in type III Docs such as the *Rc* XMod-Doc:Coh complex reported here.

Observing dual binding modes experimentally is extremely challenging using conventional bulk experiments because the two binding modes have nearly identical equilibrium binding affinity. Instead, we took a single-molecule approach which is uniquely suited for studying discrete heterogeneous systems. Single-molecule force spectroscopy with the atomic force microscope (AFM-SMFS) is able to explore a large force range up to several nN and has been used to characterize protein folding pathways [80,252,253] and receptor-ligand interactions [4,68,254,255]. Single-molecule FRET (smFRET) is capable of measuring distances at the molecular scale [256], and has been used to study protein dynamics [257,258] and to characterize structures of receptor-ligand complexes [259,260]. At the computational level, these experimental single-molecule approaches can be elaborated upon by employing molecular dynamics (MD) simulations [161]. When combined, these experimental and computational approaches can provide mechanistic insights into the dynamics of receptor-ligand complexes [4,67].

Here we used AFM-SMFS, smFRET, and MD simulations to study putative dual-binding modes and catch bond behavior of the *Rc* XMod-Doc:Coh complex. We developed a state map with experimentally measured transition rates to fully describe the system, and performed kinetic Monte Carlo simulations that recapitulate the experimental data. What emerges from this three state kinetic scheme is a picture of a unique adhesion bond that resembles a catch bond when probed under force ramp conditions, but maintains slip bonding under constant force.

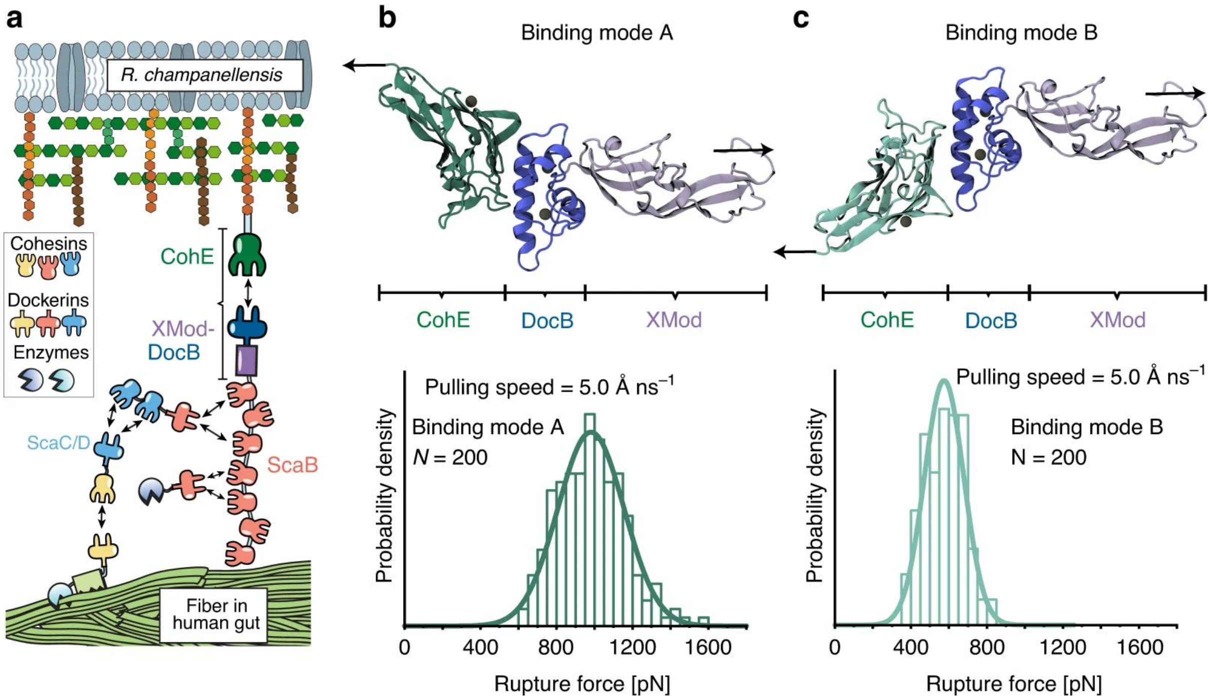


Figure 2.1 *Rc* XMod-Doc:Coh complex, dual binding modes, and molecular dynamics simulation of complex dissociation. a: The *Rc*-cellulosomal network is assembled through interactions between Doc and Coh domains. Cellulose binding domains, digestive enzymes, and structural scaffold proteins (Sca) self-assemble into a cellulosome complex, which binds and digests cellulose fibers in the human gut. XMod-DocB and CohE form a mechanically stable protein complex that anchors the cellulosomal network to the cell surface. **b and c:** Structural models showing the XMod-Doc:Coh complex in the two hypothesized binding modes. Green: Coh, blue: Doc, purple: XMod. Calcium ions are shown as black spheres. In both binding modes, the rupture forces observed for the 5 most stable models were measured by performing 200 steered molecular dynamics (SMD) replicas and plotting as histograms. The pulling directions are marked by black arrows. Rupture force histograms were fitted with Gaussian distributions. The most probable rupture force was 981 pN in binding mode A (panel **b**) and 575 pN in binding mode B (panel **c**) at a pulling speed of 5.0 Å/ns.

2.2 Results

2.2.1 XMod-Doc:Coh homology model and expression cassettes.

Since no structural information was available for the *Rc* XMod-Doc:Coh complex, we created homology models of each protein domain using Modeller 9.22 [261]. The structure of the *Rc* XMod-Doc domain was modeled based on the available structure of *Rf* CttA XMod-Doc (PDB 4IU3) [165], which shares 20% sequence identity (35% similarity) with the *Rc* domain.

The structure of the *Rc* Coh domain was modeled based on two different available structures from *Rf*, namely CohE (PDB 4IU3) with 15% sequence identity (28% similarity), and CohG (PDB 4WKZ) [262] with 18% sequence identity (34% similarity). Full amino acid sequences are given in the Supplementary Information. The selected templates share very high structural similarity with previously reported Coh and Doc domains. The 10 models with highest score from Modeller were selected for each domain/template pair, resulting in 10 models for the *Rc* XMod-Doc domain, and 20 models for the *Rc* Coh domain. Employing VMD [263], we assembled 200 models of the *Rc* XMod-Doc:Coh complex in each of the two binding modes (**Figure 2.1b and c**), building all possible combinations between XMod-Doc and Coh models. For binding mode A, the structure of the *Rf* XMod-Doc:Coh complex (PDB 4IU3) was employed to guide the *Rc* Coh:Doc interface alignment. To create a model for the hypothesized alternative binding mode B, Doc helix 1 from the homology model structure was used as a guide for the superposition of Doc helix 3. This alignment resulted in the XMod-Doc rotating 180° with respect to Coh. The models show that Doc binds Coh via the two Ca^{2+} binding loops and two binding helices (helices 1 and 3, see **Figure 2.S1**), forming a binding interface consisting of a hydrophobic center surrounded by hydrophilic amino acids, as shown in **Figure 2.S2**. This duplicated F-hand motif is consistent with those of other Doc domains which have been shown to exhibit dual binding modes [148,247,249,250].

We cloned polyproteins containing several modules for AFM-SMFS and purified them from *E. coli*. A ddFLN4 and an elastin-like polypeptide (ELP) were used as an unfolding fingerprint domain[264] and flexible linker [62,265–267], respectively. The Coh construct (N- to C-terminus) was Coh-ddFLN4-ELP-HIS-ybbr. The XMod-Doc construct (N- to C-terminus) was ybbr-ELP-ddFLN4-XMod-Doc-HIS. The ybbR tag facilitated site-specific and covalent linkage to the coverglass or cantilever tip [100]. The loading geometry with Coh pulled from its C-terminus and XMod-Doc pulled from its N-terminus precisely mimicked that experienced by the complex *in vivo*. Analysis of the equilibrium binding affinity of WT XMod-Doc:Coh using isothermal titration calorimetry (ITC) revealed $K_D = 1.0 \pm 0.3$ nM and a binding stoichiometry of 1:1. SDS-PAGE and mass spectrometry analysis indicated a molecular weight of 44 kDa for Coh construct and 55 kDa for XMod-Doc construct.

2.2.2 Steered molecular dynamics simulations reveal a weak and a strong binding mode.

To examine the stability of *Rc* XMod-Doc:Coh under mechanical load we carried out steered molecular dynamics (SMD) simulations [237] employing NAMD [268,269] and its QwikMD[270] interface. First, to test the stability of the 200 models of the complex in each binding mode, we performed equilibrium MD simulations for a combined simulation time of 2.0 µs (10 ns per model), followed by a combined 8.0 µs (40 ns per model) of SMD simulations at constant pulling velocity. These SMD simulations served as a metric to eliminate unsuitable structural models. We expected that good structural models should be stable under mechanical load, therefore, for each binding mode, we selected the 5 strongest complexes out of the 200 models. In fact, some of the 400 complexes were found not to be stable already after the equilibrium MD, and due to the low sequence identity of the templates, most of the models were not stable under mechanical load. A visual observation in VMD showed that many of these models had only partial contact between Coh and Doc following equilibrium MD. From the 5 strongest models for each binding mode, we performed 200 production SMD simulation replicas, using a similar protocol as previously described [4,67]. The simulations reveal that the dissociation of XMod-Doc:Coh occurs at clearly distinct forces for the two different binding modes, with mode A dissociating at ~981 pN, and mode B at ~575 pN, both at a 5.0 Å/ns pulling speed (**Figure 2.1b and c**).

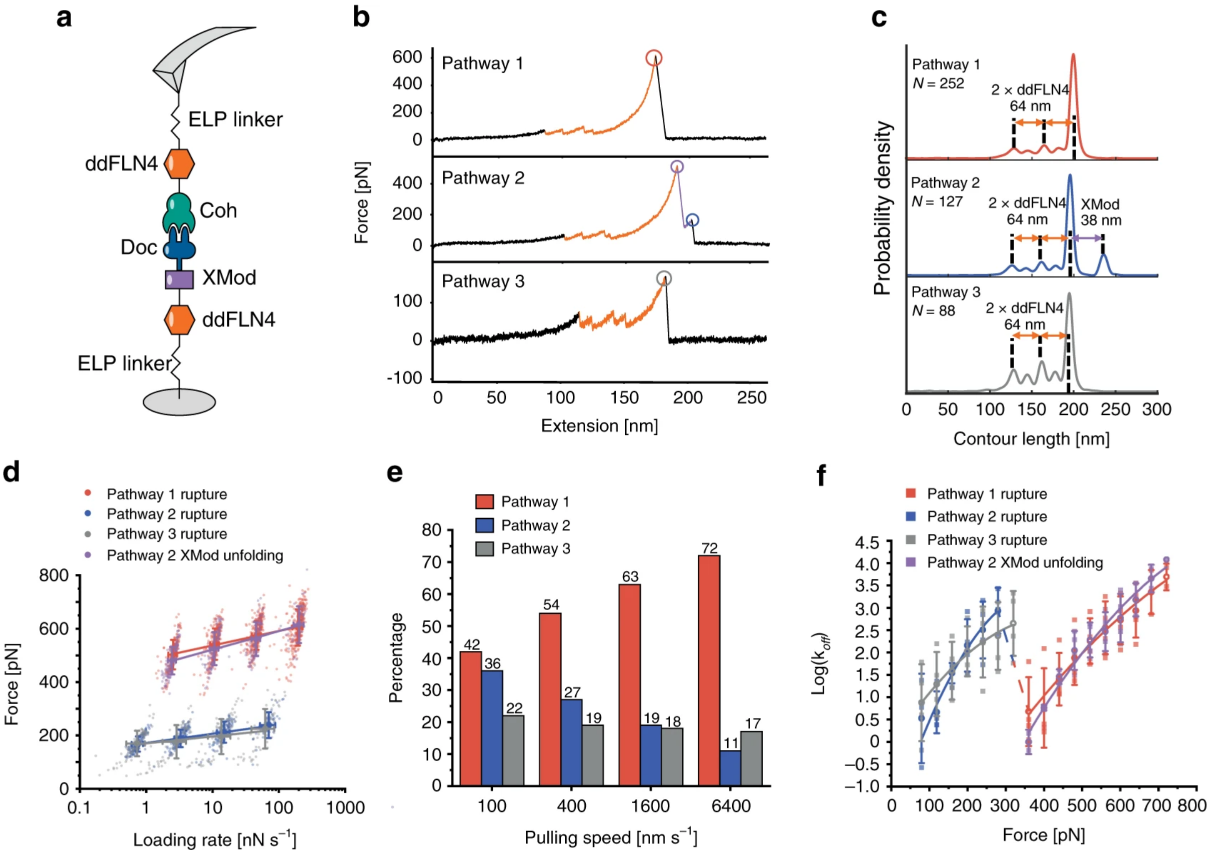


Figure 2.2 XMod-Doc:Coh unbinds along three pathways under mechanical load. **a:** Experimental configuration with Coh-ddFLN4-ELP immobilized on the AFM tip and ELP-ddFLN4-XMod-Doc immobilized on the surface. Immobilization was site-specific and covalent through a terminal ybbR tag on the ELP. **b:** Three different classes of force curves were repeatedly observed, corresponding to different pathways. In pathway 1 (P1), the complex ruptured at high force (500-600 pN, red) with the XMod remaining folded. In pathway 2 (P2), XMod unfolded (purple) followed by a low force rupture of the complex (blue). In pathway 3 (P3), the complex ruptured at low force rupture (grey) with the XMod remaining folded. Unfolding of the two ddFLN4 fingerprint domains (orange) was used to identify single-molecule traces. **c:** Combined contour length histograms for each unbinding pathway. Force-extension traces were transformed using a freely rotating chain elasticity model and aligned using cross-correlation analysis. Histograms show contour length increments resulting from unfolding of 2× ddFLN4 and XMod. **d:** Rupture force *vs.* loading rate plot showing final XMod-Doc:Coh complex rupture events obtained from the three pathways, as well as XMod unfolding events observed in P2. Error bars represent the standard deviation of rupture force and loading rate. Lines show linear Bell-Evans fits of the most probable rupture/unfolding force vs. logarithm of loading rate to obtain $\Delta x^{\ddagger}$ and k_0 (Eq. 2.2). Fitted $\Delta x^{\ddagger}$ and k_0 values are listed in Table 2.1. **e:** Percentages of the three pathways in all rupture events at different pulling speeds. **f:** Kinetic off-rate (k_{off}) *vs.* force for complex rupture and XMod unfolding events. Force-dependent off-rates calculated using the histogram transformation

method (Eq. 2.3) were plotted against force (shown in boxes). The average off-rates of four different pulling speeds were shown in open circles and fitted to the analytical expression (Eq. 2.5) with υ=0.5. Error bars represent the standard deviation of off-rate. The fitted $\Delta x^{\ddagger}$, $\Delta G^{\ddagger}$ and k_0 values are listed in Table 2.1.

2.2.3 Wild type XMod-Doc:Coh unbinds along 3 distinct pathways.

We performed AFM-SMFS with Coh covalently attached to the cantilever tip through its C-terminal ybbR tag and XMod-Doc covalently attached to the surface (**Figure 2.2a**) through its N-terminal ybbR tag. The XMod-Doc:Coh complex was formed by approaching the AFM tip to the surface and dwelling for 200 ms. After XMod-Doc:Coh complex formation, the cantilever base was retracted at constant speed and a force-extension curve was recorded. This procedure was repeated thousands of times typically over a 12 hours period to generate large datasets of force *vs.* extension curves. The recorded force curves were transformed into force vs. contour length space using a freely rotating chain (FRC) elasticity model (Eq. 2.1). We searched for the contour length pattern of ddFLN4, which contained ~32 nm of total contour length that resulted from a two-step unfolding pattern. Since one ddFLN4 molecule was contained in the surface-linked protein, and another one in the cantilever-linked protein, we only analyzed curves which contained in total two ddFLN4 unfolding fingerprints, thereby eliminating spurious signals.

We repeatedly observed three distinct unbinding pathways of the complex, as shown in **Figure 2.2b**. We refer to these as pathway 1 (P1), pathway 2 (P2), and pathway 3 (P3). We used cross-correlation analysis [134,271] to assemble superposition contour length histograms for each pathway (**Figure 2.2c**). These histograms all showed the distinct unfolding pattern of two ddFLN4 fingerprint domains, adding in total 64 nm of contour length to the system. P2 showed an additional 38 nm length increment which matched the expected value for XMod unfolding (116 XMod amino acids * 0.365 nm/amino acid - 5.3 nm folded length = 37 nm) (**Figure 2.2c, middle**). The contour length histograms were broadened by occasional unassigned unfolding events that were observed in all three pathways, which we attributed to partial unfolding of Coh or Doc. A representative sampling of these unassigned unfolding events are presented in **Figure 2.S3**.

Approximately 80% of curves were assigned to P1 or P2. In P1, unfolding of two ddFLN4 domains in series was followed by dissociation of XMod-Doc:Coh at high forces of ~500 pN (**Figure 2.2b, top)**. In P2, XMod unfolded at high forces, followed by the Doc-Coh complex rupture at low forces of ~200 pN (**Figure 2.2b, middle**). This indicated that XMod unfolding significantly destabilized the interaction between Doc and Coh in P2, giving rise to a shielded complex rupture event. P2 was reproduced in SMD simulations by deleting the XMod from the Doc:Coh complex in binding mode A, which showed a decrease in rupture forces (**Figure 2.S4**). The remaining 20% of curves were classified as P3, where XMod-Doc:Coh ruptured at low force (~200 pN, **Figure 2.2b, bottom**) and no XMod unfolding was observed. Based on these classifications, we hypothesized that P1 and P2 resulted from complexes with high mechanical stability, which were able to resist external forces as high as ~500 pN prior to high force complex rupture or XMod unfolding. In the cases where XMod unfolded, the Doc-Coh binding interaction became destabilized and ruptured at low force. P3 meanwhile represented a weaker Doc-Coh complex that ruptured at lower force (~200 pN) even without XMod unfolding. The existence of complexes with different mechanical stabilities was consistent with SMD simulation results **(Figure 2.1c)**, which showed that the dual-binding modes rupture at distinct forces.

2.2.4 Allosteric regulation by XMod gives rise to catch bonding in force ramp mode.

AFM measurements on WT XMod-Doc:Coh were carried out at pulling speeds of 100, 400, 1600, and 6400 nm/s, which allowed us to investigate the loading rate dependency of complex rupture and XMod unfolding in the various pathways (**Figure 2.2d and 10**). We used the Bell-Evans (BE) model (Eq. 2)[123,124] to analyze the experimental force-loading rate data and obtain the intrinsic off rate (k_0) and the distance to the transition state along the reaction coordinate ($\Delta x^{\ddagger}$) for the complex rupture events in each pathway, as well as for XMod unfolding along P2 (**Table 2.1**).

As shown in **Figure 2.2e**, the percentage of curves that were classified as P3 was independent of the pulling speed, maintaining a value of 17-22% across the range of speeds tested. This observation was consistent with the hypothesis that P3 belonged to a different binding mode than P1 and P2. Interestingly, the ratio between P1 and P2 was dependent on the pulling speed. The likelihood of P1 increased with increasing pulling speed from 100-6400 nm/s, while the likelihood of observing P2 decreased. This means that the complex preferentially populated the pathway with higher rupture force (P1) when pulled at higher loading rates. This switch from low rupture force P2 to high rupture force P1 at increasing loading rates is not to be confused with standard scaling based on Bell-Evans theory, which also predicts higher rupture forces at higher loading rates. This behavior, in contrast, represented a discrete non-linear switching from P2 to P1 with much higher rupture forces. Although P1 and P2 rupture events each individually scale as classical slip bonds as a function of the loading rate, the pathway switching behavior precisely mimics that of a catch bond [29,149,150,156,272] probed under force ramp conditions. In contrast to other reported catch bonds in the literature which occur at low force (<50 pN), XMod-Doc:Coh is activated at much higher forces (>300 pN).

Table 2.1 Kinetic parameters extracted from AFM-SMFS

Pathway	Event	k_0 [s^{-1}] (BE)	$\Delta x^{\ddagger}$ [nm] (BE)	$\Delta G^{\ddagger}$ [k_BT] (DHS)	k_0 [s^{-1}] (DHS)	$\Delta x^{\ddagger}$ [nm] (DHS)
1	High force rupture (XMod intact)	4.70×10^{-8}	0.178	22.8 (υ=0.5) 20.0 (υ=2/3)	9.39×10^{-5} (υ=0.5) 1.57×10^{-4} (υ=2/3)	0.146 (υ=0.5) 0.132 (υ=2/3)
2	Unfolding of XMod (measured)	9.94×10^{-6}	0.139	27.6 (υ=0.5) 24.0 (υ=2/3)	5.74×10^{-7} (υ=0.5) 4.79×10^{-6} (υ=2/3)	0.209 (υ=0.5) 0.168 (υ=2/3)
	Unfolding of XMod (corrected)	4.53×10^{-5}	0.116	30.1 (υ=0.5) 26.1 (υ=2/3)	2.66×10^{-6} (υ=0.5) 5.42×10^{-6} (υ=2/3)	0.173 (υ=0.5) 0.157 (υ=2/3)
	Low force rupture after XMod unfold	7.43×10^{-4}	0.277	13.9 (υ=0.5) 12.5 (υ=2/3)	0.0152 (υ=0.5) 0.101 (υ=2/3)	0.254 (υ=0.5) 0.171 (υ=2/3)
3	Low force rupture (XMod intact)	2.79×10^{-5}	0.366	8.46 (υ=0.5) 7.45 (υ=2/3)	0.831 (υ=0.5) 1.08 (υ=2/3)	0.131 (υ=0.5) 0.113 (υ=2/3)

The explanation for this apparent catch bond behavior under force ramp conditions is evident when looking at the loading rate dependency of XMod unfolding. The loading rate dependency of XMod unfolding is steeper than that of the complex rupture in P1 (**Figure 2.2d, Table 2.1**). Therefore, at high loading rates, far fewer complexes reach sufficiently high forces to unfold XMod prior to complex rupture, thus prohibiting the system from entering P2. This behavior is unique to this particular XMod-Doc:Coh system and was not observed in other Doc:Coh systems reported thus far [6,67].

We note that the experimentally observed values for XMod unfolding are slightly biased by the maximal stability of the receptor-ligand complex [60]. This ceiling effect is magnified at high loading rates (>100 nN/sec) because the XMod unfolding force increases and exceeds the maximal force that the complex can withstand. We attempted to measure the unbiased rupture force of XMod using a high rupture force handle, i.e. the SD-repeat protein G (SdrG) from *Staphylococcus epidermidis* in complex with the N-terminus of human fibrinogen β chain (Fgβ). As shown in **Figure 2.S6a**, the Fgβ-XMod-Doc-ddLFN4-ELP-His-ybbr immobilized on the glass surface binds the SdrG-ddFLN4-ELP-HIS-ybbr immobilized on the AFM tip. As the cantilever retracts, the ddLFN4 fingerprint domains and the XMod were unfolded, followed by the rupture of SdrG:Fgβ complex. The interaction between SdrG and Fgβ can withstand a force of ~2 nN[4], which is significantly higher than the unfolding force of XMod. Therefore this construct is expected to measure the XMod unfolding force without any biasing due to the XMod unfolding force exceeding the complex rupture force. However, as shown in **Figure 2.S6b**, the measured XMod unfolding force was ~300 pN, significantly lower than the unfolding force measured using the XMod-Doc:Coh complex. We attribute this decrease in XMod unfolding force to the change of force loading geometry. As shown in **Figure 2.S6a**, in the native XMod-Doc:Coh complex, the force was loaded to the XMod-Doc from its N-terminus and the Doc:Coh binding interface. However, in the SdrG:Fgβ-XMod-Doc complex, the force is loaded to the N- and C-termini of the XMod-Doc, leading to a decreased XMod unfolding force compared to the native force loading geometry. Since the current toolbox is not able to measure the unbiased XMod unfolding force experimentally, we used a theoretical tool to correct the XMod unfolding force distribution to take this biasing effect into account[60], as shown in **Figure 2.S7**. Using the Bell-Evans model, we obtained the kinetic parameters of XMod unfolding after bias correction (**Table 2.1**). This analysis confirmed what was observed in the rupture force *vs.* loading rate scatter plots, namely that XMod has a steeper loading rate dependency (lower $\Delta x^{\ddagger}$) than the high force rupture event in P1, and that these scaling differences give rise to catch bonding in force ramp/constant speed mode.

The rupture forces from three pathways as well as the XMod unfolding forces obtained at different pulling speeds were plotted as histograms (**Figure 2.S7 and 2.S8**), which were transformed into force-dependent off rate $k_{off}(F)$ using Eq. 3 and plotted against force. The force-dependent off-rates were fitted using Dudko-Hummer-Szabo (DHS) model (Eq. 2.5) [128,129] to extract the intrinsic barrier crossing rate (k_0), barrier height ($\Delta G^{\ddagger}$) and distance to the transition state along the reaction coordinate ($\Delta x^{\ddagger}$) of the various barrier-crossing events, as shown in **Figure 2.2f and 2.S9** and **Table 2.1**. The off rate *vs.* force plot was fitted using two different υ values in Eq. 2.5: $\upsilon = 0.5$ (**Fig. 2f**, assuming a cusp-like energy barrier) and $\upsilon = ⅔$ (**Figure 2.S9**, assuming a linear-cubic energy barrier). Both υ values generated good fits and extracted similar k_0, $\Delta G^{\ddagger}$ and $\Delta x^{\ddagger}$ values (**Table 2.1**), meaning that the model is applicable to the experimental data. In addition to the transformation shown in **Figure 2.2f**, we used Eq. 2.6 to calculate the force-dependent off-rate based on combining the histograms from P1 and P2 rupture events (**Figure 2.S10**) [273,274]. The rationale for combining the analysis of P1 and P2 rupture events was motivated by the evidence for dual binding modes (see below). Eq. 2.3 and Eq. 2.6 are essentially the same equation where Eq. 2.3 is the discrete form and Eq. 2.6 is the continuous form. Therefore, the force-dependent off-rate calculated from both separated and combined P1/P2 histograms (**Figure 2.2f, 14 and 15**) showed the same trend (**Figure 2.S10e**) and a crossover regime around 300 pN where the off rate decreased with increasing force, consistent with previously reported multi-pathway systems leading to catch bond behavior [273,274]. We note that comparison of the fitted energy landscape parameters between these two models in **Table 2.1** is complicated by the fact that the Bell-Evans model assumes $\Delta x^{\ddagger}$ is constant and independent of the force, while the DHS model assumes both $\Delta x^{\ddagger}$ and $\Delta G^{\ddagger}$ to be force-dependent. In practice we used DHS model to generate off-rates for each pathway at constant force with data derived from force ramp experiments, while Bell-Evans fitting was used for extracting the loading rate dependency of rupture events from constant speed measurements.

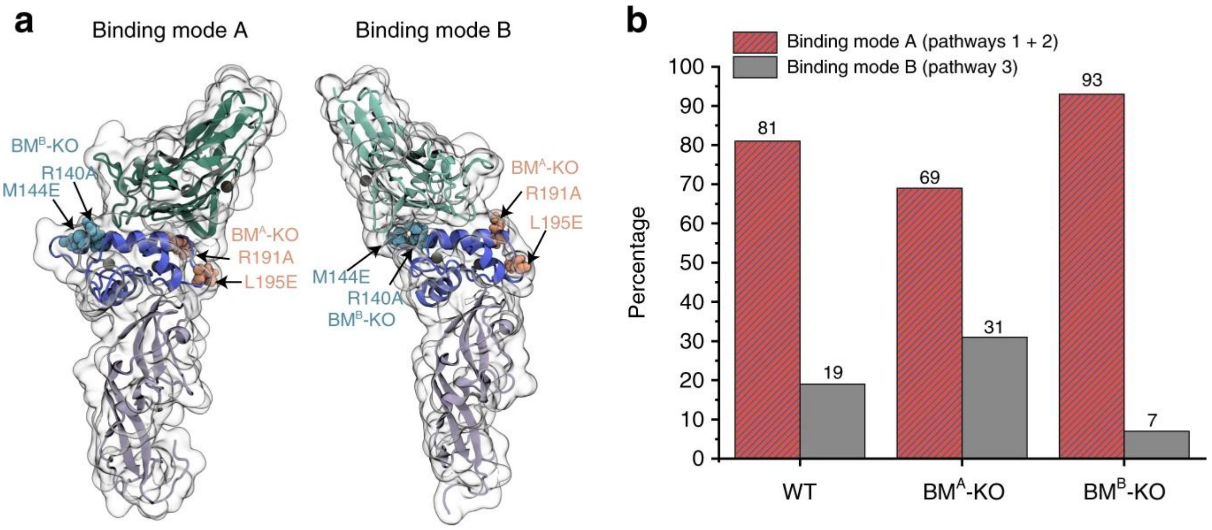


Figure 2.3 SMFS measurements of XMod-Doc binding mode mutants. a: Arginine 191 and leucine 195 on Doc were mutated to alanine and glutamic acid, respectively, to disrupt their interaction with Coh in binding mode A. Arginine 140 and methionine 144 were mutated to alanine and glutamic acid, respectively, to disrupt their interaction with Coh in binding mode B. **b:** Percentages of the two binding modes measured with WT XMod-Doc, BM^A-KO mutant and BM^B-KO mutant at 400 nm/s pulling speed. The BM^A-KO mutant preferentially populates binding mode B while the BM^B-KO mutant preferentially populates binding mode A.

2.2.5 AFM-SMFS evidence of dual-binding modes.

We hypothesized that P1 and P2 arose from one binding mode, while P3 arose from an alternative binding mode with lower mechanical stability. To test this, we sought to knock out specific binding modes by mutagenesis (**Figure 2.3a** and **2.6**). Using the structural models, we identified key Doc residues likely to be involved in each respective binding mode (**Figure 2.S1**), and designed mutations to disrupt electrostatics and hydrogen bonding. The mutant designed to knock out binding mode A contained R191A and L195E mutations, and is referred to as BM^A-KO. The mutant designed to knock out binding mode B contained R140A and M144E mutations and is referred to as BM^B-KO. Interactions between BM^A-KO or BM^B-KO and Coh were then measured using AFM-SMFS at 400 nm/s. For WT XMod-Doc:Coh, the percentage of P3 curves was typically ~20%. As shown in **Figure 2.3b** and **2.S11**, BM^A-KO resulted in a P3 curve percentage that increased to 31%. We attributed this increase in P3 probability to the destabilization of binding mode A, and slight preferential formation of binding mode B as compared to WT. This result indicated that binding mode B was likely associated with the low force pathway P3. However, the mutations were not able to completely knock out binding mode A.

BM^{B}-KO was more effective at knocking out binding activity, and decreased the percentage of P3 curves from 19% for WT down to 7% with a corresponding increase in P1 and P2 percentage. Despite the introduction of destabilizing mutations at the binding interface in BM^{B}-KO, we nonetheless obtained a system with higher stability and predominantly high force rupture pathways, a result that may seem counterintuitive but is explained by the presence of a weak binding mode B being knocked out or inhibited by the mutations. Based on these measurements with the binding mode knock-out mutants, we concluded that P1 and P2 are attributable to binding mode A, which is the strong binding mode, while P3 corresponds to binding mode B, which is the weak binding mode. In contrast to other Doc:Coh systems exhibiting dual-binding mode [148,249], the two binding modes of *Rc* XMod-Doc:Coh complex have significantly different mechanical stabilities with one rupturing at ~200 pN and the other able to withstand forces of ~500-600 pN.

This conclusion was further supported by a statistical analysis involving a biasing effect of an additional fingerprint domain [60] (see **Supplementary information note 2.2** and **Figure 2.S12**). We introduced an additional fingerprint domain (I27) whose unfolding force sits in between the P1 and P3 rupture events. If the multi-pathway dissociation behavior that we observed resulted from multiple unbinding reaction pathways originating from a single bound state, we would expect that the likelihood of observing an I27 unfolding event would be decorrelated from the pathway classification of the curve. We did not observe this, and instead the vast majority of curves that showed I27 unfolding terminated in a high force rupture event (P1) or XMod unfolding followed by low force rupture (P2). This indicated that complexes that ruptured in a low force rupture event (P3) were not sufficiently strong to unfold I27, consistent with P3 emerging from a discrete binding mode that was weaker than the P1 or P2 complexes, further substantiating the dual binding modes.

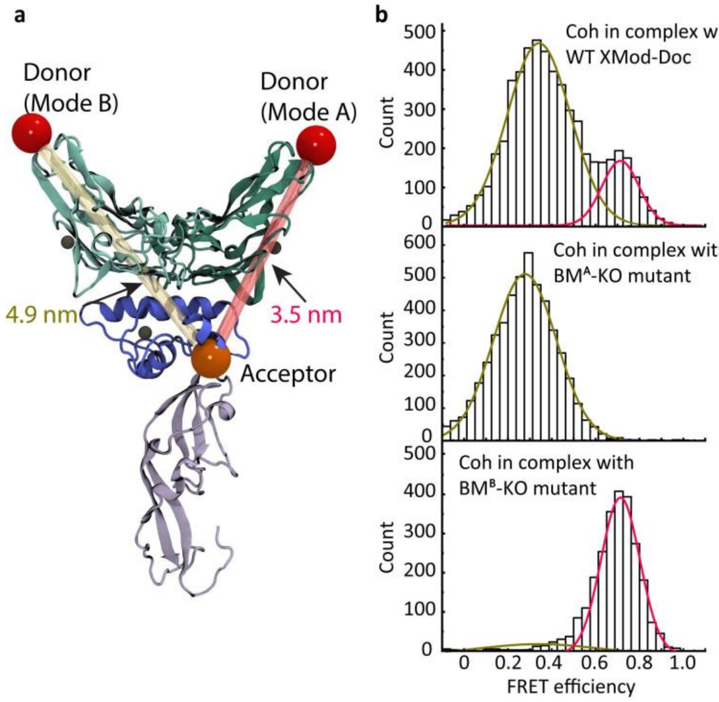


Figure 2.4 Dual binding modes are observed by smFRET. a: The C-terminus of Coh (position 154) was labeled with the FRET donor Cy3b while the C terminus of the XMod-Doc (position 199) was labeled with the FRET acceptor AF647. The donor-acceptor distance was smaller in binding mode A (3.5 nm) than the binding mode B (4.9 nm), giving rise to a higher FRET efficiency when the complex formed in binding mode A. **b:** FRET efficiency histograms measured using Cy3b-labeled WT XMod-Doc (top), BM^A-KO (middle), or BM^B-KO (bottom) complexed with AF647-labeled Coh. Efficiency histograms were fitted with single or double Gaussian distributions.

2.2.6 Single-molecule FRET evidence of dual-binding modes.

Based on differences in inter-residue distances in the two binding conformations, we used smFRET to observe the dual-binding modes. We introduced a point cysteine mutation at position 154 of Coh and covalently attached a FRET donor dye maleimide-Cy3b. Since XMod-Doc has native cysteines, we used amber suppression [275] to introduce a non-canonical azide at position 199 of XMod-Doc, and covalently attached DBCO-AF647. Based on the homology models (**Figure 2.4a**), the donor-acceptor distance is expected to be ~3.5 nm in binding mode A and ~4.9 nm in binding mode B. XMod-Doc:Coh complexes were formed by mixing labeled XMod-Doc and Coh in a 1:1 molar ratio and diluting them to ~200 pM. FRET efficiency of individual XMod-Doc:Coh complexes was measured on a confocal microscope and plotted into histograms (**Figure 2.4b**). A bimodal distribution was clearly observed in the FRET efficiency histogram of WT XMod-Doc:Coh, with mean FRET efficiencies of 0.34 and 0.71, corresponding to binding modes B and A, respectively. In addition to labeling and analyzing WT, we introduced the FRET acceptor dye into BM^A-KO and BM^B-KO mutants at position 199, and again measured FRET efficiency in complex with labeled Coh using the same protocol as for WT. We found that only the low FRET efficiency peak was observed in BM^A-KO, meaning that binding mode A corresponding to the high FRET efficiency peak was eliminated by the mutations. The FRET efficiency histogram of BM^B-KO complexed with Coh meanwhile showed predominantly the high FRET efficiency population, consistent with binding mode B being knocked out. Compared to AFM-SMFS, the binding mode A is much less prevalent in the smFRET measurement of wild-type complex and the BM^A-KO mutant knocks out the binding mode A much more efficiently in smFRET measurement. We attributed this difference to the acceptor dye destabilizing the complex in binding mode A but not binding mode B, which was supported by AFM-SMFS measurements between dye-labeled BM^A-KO and unlabeled Coh (**Figure 2.S13**).

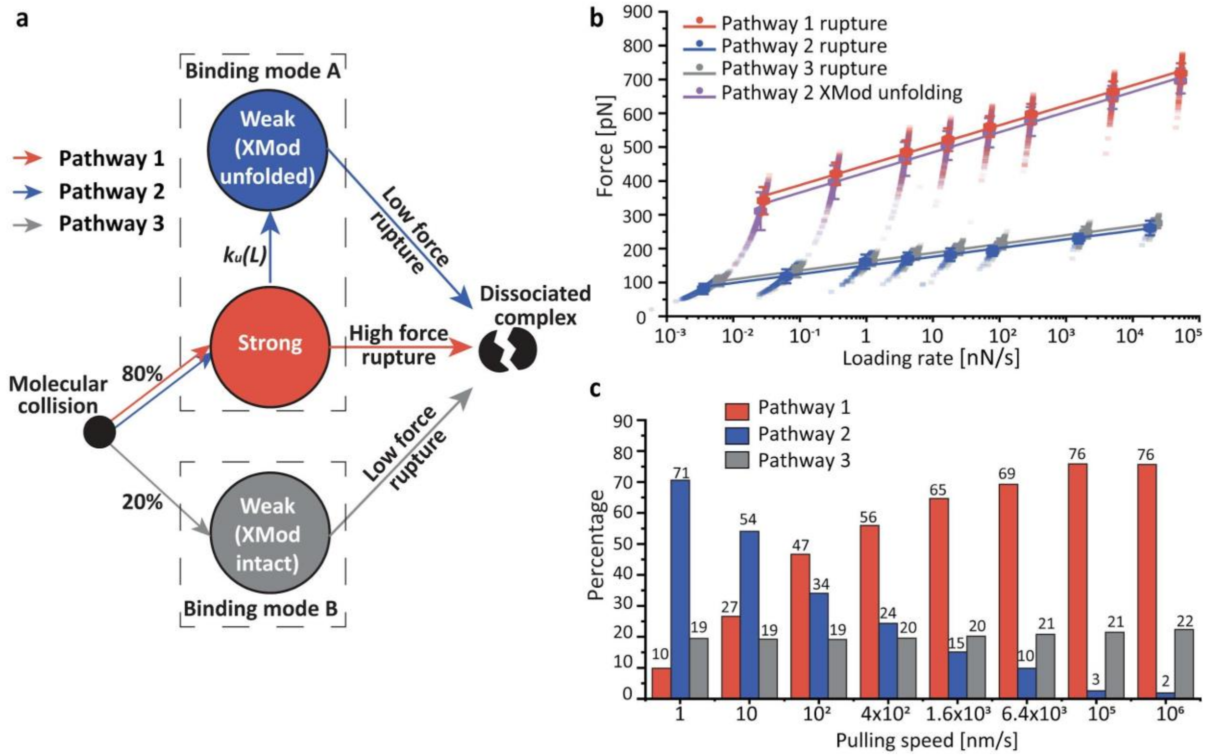


Figure 2.5 Multi-state kinetic model and Monte-Carlo simulation of mechanical rupture of XMod-Doc:Coh. a: The multi-state kinetic model postulates that upon molecular collision, the complex has ~80% probability of forming the strong binding mode A (pathways 1 and 2) and 20% probability of

forming the weaker binding mode B (pathway 3). Once the binding mode is set, there is no interconversion between the modes. In the strong binding mode, the complex can rupture at high force (pathway 1) with XMod remaining folded, or XMod can unfold prior to complex rupture according to a loading rate dependent unfolding rate $k_u(L)$ (pathway 2). In pathway 2, XMod unfolding destabilizes the complex, resulting in low force complex rupture. At increased loading rates, the XMod unfolding rate decreases so that pathway 2 is deactivated and the complex has a higher probability of unbinding along pathway 1 (high force rupture). In pathway 3, the complex ruptures at low force without reaching sufficiently high force to unfold XMod. **b and c:** Monte-Carlo simulation results. Force *vs.* extension curves at constant pulling speed were simulated to obtain the loading rate dependency of complex rupture and XMod unfolding events (**b**), as well as the percentages of the three unbinding pathways at different pulling speeds (**c**). The simulation was carried out at the same pulling speeds as the experiments (100, 400 ,1600, 6400 nm/s) and further extended over a range from 1 to 10^6 nm/s. Simulated force *vs.* loading rate plots were fitted with the Bell-Evans model to extract k_0 and $\Delta x^{\ddagger}$ values (**Table 2.S1**). The error bars in **(b)** represent the standard deviation of rupture forces.

2.2.7 Kinetic model and Monte Carlo simulations.

Combining the experimental results and MD simulations led us to propose a kinetic scheme for the unbinding mechanism of the *Rc* XMod-Doc:Coh complex that accounts for dual-binding modes as well as the catch bond behavior observed under force ramp conditions (**Figure 2.5a**). Our model postulates that there are two non-interconvertible bound states with different mechanical stabilities (binding modes A and B). Upon binding, the complex has an 80% probability of forming the more stable binding mode A, and a 20% probability of forming binding mode B. If the complex forms in binding mode B, the only escape pathway under load is P3 terminating in a low force rupture (~200 pN). When bound in binding mode A, the complex either ruptures at high force (P1), or enters a weakened state due to the unfolding of XMod (P2). The rate of entering the weaker state (P2) from the stronger state (P1) decreases as the loading rate increases because of the steeper loading rate dependency of XMod unfolding. This results in an increased proportion of P1 high rupture force curves when the complex is probed using force ramp conditions at high loading rates (>100 nN/sec), which is precisely what is observed in classical catch bonds.

We used kinetic rates obtained from AFM-SMFS combined with our proposed state model to simulate the system in a constant pulling speed scenario, identical to the experiments. We used worm-like chain elasticity theory and Monte-Carlo [276] to simulate force-extension curves of the XMod-Doc:Coh stretching, XMod unfolding and complex rupture (see Methods - Monte Carlo simulation). The loading rate dependency of the complex rupture force and XMod unfolding forces, as well as the rupture and unfolding force histograms from the simulated curves are shown in **Figure 2.5b**, **2.S14** and **2.S15**. The simulations showed remarkable agreement with experiment results both in terms of the rupture forces, and other observed trends (**Figure 2.2**, **2.S7** and **2.S8**). For example, our network Monte-Carlo modeling shows the same bimodality of the rupture force distributions, similar force magnitudes and similar ratios between the P1, P2, and P3 trajectories. Furthermore, the novel catch bond network topology that emerged in force ramp mode was also observed in the simulation.

The simulations further allowed us to probe a range of pulling speeds that were not accessible experimentally. We extended the range of pulling speeds in the simulations to get a clearer picture of the catch bond behavior. As shown in **Figure 2.5c**, at high loading rates, the complex predominantly ruptures along P1 due to the strengthening of XMod. At extremely slow pulling speeds, we see in the simulation that the P1 pathway is lost and the complex only exhibits P2 and P3 low force rupture behavior. The broad agreement of the simulation with the experimental results provided strong support for the proposed kinetic scheme.

2.3 Conclusions

We discovered a new mechanism by which bacteria achieve mechanically stable adhesion to crystalline fiber surfaces in the human gut, and resolved the dual binding modes of this complex using single-molecule techniques and all atom simulations. The kinetic scheme amounts to a novel multi-state catch bond mechanism in binding mode A (P1/P2 paths). The system starts in the high rupture force (P1, activated) state and has a certain probability of entering the low rupture force state (P2). The transition rate from P1 to P2 decreases with increasing loading rate, meaning that the low rupture force state is inhibited at high loading rates. Once the complex enters the low rupture force state, it cannot return to the high rupture force state. These features make our system distinct from the other two-state catch bond models [27,29,150–152]. Interestingly, the catch bond behavior emerges from a network of purely slip bonds/folds and only manifests under a force ramp or constant speed scenario. If this system is probed using constant force clamp conditions, there is no increase in lifetime as the clamping force is increased (**Figure 2.S16**).

To further clarify the description of the system as a catch bond, we note that when considering the XMod unfolding force and the binding interface together as a unified system, the maximal force that the system can withstand does not increase with increasing loading rate or clamping force (**Figure 2.S16**). This is due to the high forces required to unfold XMod and enter the low stability P2 pathway. Nevertheless, when considering the force at which the Doc:Coh binding interface breaks (i.e. the rupture event), we find the term catch bond appropriate. The force-dependent off rates (**Figure 2.2f, 2.S9** and **2.S10**) show catch bond behavior when considering the force at the time of the rupture event independently from bond history. This mechanism emerges due to the inhibition of P1 to P2 transfer rates at high loading rates (**Figure 2.2e** and **2.5c**), which caused P1 rupture events to become more frequent at higher loading rates.

Based on structural modelling and analysis, we predicted that the heterogeneity of unbinding pathways was attributable to two different binding conformations, binding modes A and B. AFM-SMFS and smFRET on mutant XMod-Doc constructs designed to specifically knockout binding mode A or B supported the presence of dual binding modes with different mechanical properties. The biological significance of the two binding modes is still unclear, however we speculate that the *Rc* bacterium might switch between the low (P3) and high force catch (P1/P2) adhesion modes based on post-translational modifications or environmental factors, for example fiber substrate composition or intraluminal pH of human colon [277,278], allowing the bacterium to respond to environment change. Our research demonstrates an entirely new degree of complexity by which bacteria regulate adhesion strength through molecular mechanisms such as dual-binding modes, mechanical allostery, and catch bonding.

2.4 Follow up study: *Candida auris* agglutinin-like sequence (ALS) adhesins

In addition to the maintenance of extracellular networks, adhesion proteins are involved in many other biological processes, an important one of which is pathogen-host interactions. Pathogens use cell-surface adhesins to specifically recognize and bind the receptors on the host cell surface to invade the host cells. The pathogen adhesins typically form highly mechanostable complexes with their ligands to resist shear stress and adhere tightly to the host cells.

Understanding the mechanisms of mechanostable pathogen adhesins helps developing new approaches to preventing and treating infections.

One of the adhesins used by pathogenic fungi is the agglutinin-like sequence (ALS) adhesins found in *Candida* yeasts. The ALS 9-2 of *C. albicans* was found to bind the C-terminus of human fibrinogen ɣ chain (Fgɣ) [279]. The ALS 9-2 is highly homologous to the other ALS adhesins found in *C. albicans*.

Another pathogenic *Candida* species, *C. auris*, was first identified in 2009, and has become a major threat to public health worldwide [280,281]. Using BLAST (https://blast.ncbi.nlm.nih.gov/), we aligned the sequence of *C. albicans* ALS 9-2 against the currently available *C. auris* genome and proteome sequences and identified three potential *C. auris* ALS homologous sequences. The sequences of *C. albicans* ALS 9-2 and the potential *C. auris* ALS homologous sequences are:

>Candida albicans ALS 9-2

RKTITGVFNSFDSLTWTRSVEYVYKGPETPTWNAVLGWSLNSTTADPGDTFTLILPCV
FKFITTQTSVDLTADGVSYATCDFNAGEEFTTFSSLSCTVNSVSVSYARVSGTVKLPIT
FNVGGTGSSVDLADSKCFTAGKNTVTFMDGDTKISTTVDFDASPVSPSGYITSSRIIPS
LNKLSSLFVVPQCENGYTSGIMGFVASNGATIDCSNVNIGISKGLNDWNFPVSSESFSY
TKTCTSTSITVEFQNVPAGYRPFVDAYISAENIDKYTLTYANEYTCENGNTVVDPFTL
TWWGYKNSEADSDGDVIVV

>QEL63120.1 Hypothetical Candida auris ALS 1 (Identity: 32%, homology: 50%)

AKLQSGVFNGIKSITPSDNRRPEQPSWHATVSWEIKPAMGVQEGDTFTLHMPYVYKF
TSSSNTLQLTAGGQVVANCNLYSGENIVGYSEVQCTATAAAANAGTFTGDVTFPFTF
NAGSTSDEVNLEAAGVWKSGQNTVTWSDGDKTFSTTVDFNPGASSIIQGSPENGVYG
LRKMVSLNINQHYLMGPSCPYDGQYGRLEISNPSPGVGFDCSSLAGAITDQVNDWYF
PKTAEKIGVNIDSCSSYQATVSFSNLPAGFRPYININAAIPNVASFRSSNTYSYNFVCGG
RQLGQSSIAWVMYNNGNTGSGGDFKPV

>QEO23751.1 Hypothetical Candida auris ALS 2 (Identity: 33%, homology: 49%)

ASVKEGIFDKILSITPPEDSLPETPSWSATVEWSFDEKNGVKAGDTFVLHMPYVYKFT
SGTRAVDLVADSVTFANCDLFSGDNVVAYSELKCTATSACEKVNSAKGTVEFPFTFN
AGSSSDKATLEASTVWHAGSNVVKWTDGSKELHSEVYFDEGNPYLFSGSVEHGVY

WLRKAVMRNTNQHLVLGPSCDCDGMSGYIEIQNPDYGVELDCSSVVGTITKQINDFY
FPESAESCSVHVDECCATRIRVSFDNIPSGFRPYINVDSAIPHYDFNSRNQYSYGFSCG
GIELCDSLWTNWYMYKDGETGGDGEFNSI

>PSK75301.1 Hypothetical Candida auris ALS 3 (Identity: 28%, homology: 46%)

AAPQTGVFTSIDSLAPFDVAWPMMPGWDATVSWHINSSMGMKDGDTFFLRIPFVIEF
NTDESSIQMSDGANTLANCVLTPGENLVPYSEVKCTATTQVEAVQSSSGTITFPIVFN
AGFSAQESDLEAANHWSTGSNTLEWTDGSNTLTHPITFVGGTMSAFNGRPKRGILDQ
RSFVSTNTIRQFLMGPLCQSSDMSGELSIENLSEEAPFDCDSITMAMSNQINAWYFPQT
ADEAEATIVSCSAAGVNVAFSNLPAGFRPYINIDATKKIAVSEIDNIYHYNFTCNGAEL
SDSIFAAWDQFFSDDTEEDDTLTQV

All of the three hypothetical *C. auris* ALS adhesin sequences have ~30% identity and ~50% homology with the *C. albicans* ALS 9-2. We speculate that the hypothetical *C. auris* ALS adhesins and *C. albicans* ALS 9-2 have similar behavior: they are promiscuous binders targeting Fgɣ peptide.

The *C. auris* ALS sequences and *C. albicans* ALS 9-2 were cloned to bacterial and yeast expression vectors. Both the *C. albicans* ALS 9-2 and the hypothetical *C. auris* ALS protein 1 could be expressed using both *E. coli* shuffle strain and *Pichia pastoris* cells in soluble form, but bulk experiments on size-exclusion column (SEC) did not detect formation of ALS adhesin:Fgɣ peptide complex. This result is inconsistent with previously reported results where *C. albicans* ALS 9-2 was shown to form complex with Fgɣ peptide [279]. Possible reasons leading to this inconsistency include: the affinity between ALS and Fgɣ is below the detection limit of SEC; the binding pocket of ALS adhesin was blocked by other binders due to the promiscuous nature of ALS; or the expressed ALS adhesins were not active.

Future troubleshooting will first concentrate on the *C. albicans* ALS 9-2 and identify the reason why the expressed adhesin did not bind the reported ligand Fgɣ. After solving this problem, the next steps will be to identify the natural ligand of hypothetical *C. auris* ALS adhesins and study the mechanisms of ALS adhesin-ligand interactions using bulk and single-molecule experiments.

2.5 Methods

All reagents were at least of analytical purity grade and were purchased from Sigma-Aldrich (St. Louis, MO, USA), Thermo Fisher Scientific (Waltham, MA, USA), GE Healthcare (Chicago, IL, USA), New England Biolabs (Ipswich, MA, USA) or ABCR GmbH (Karlsruhe, Germany).

Synthetic genes were purchased from Thermo Fisher Scientific.

Primers for cloning were purchased from Microsynth AG (Balgach, Switzerland).

All buffers were filtered through a 0.2 μm polyethersulfone membrane filter (Sarstedt, Nuembrecht, Germany) prior to use. The pH of all buffers was adjusted at room temperature.

2.5.1 Homology modeling and molecular dynamics simulations

To the best of our knowledge, the structure of the XMod-Doc domain from Scaffoldin B of *Ruminococcus champanellensis* and its binding partner, the CohE domain from the same bacterium, has not been solved by experimental means. Using homologous structures available in the Protein Data Bank (www.pdb.org), we employed Modeller 9.22 [261] to obtain a homology model of the two *R. champanellensis* cellulosomal domains. Modeller works by setting spatial restriction to the atomic positions of the model protein, based on 3D-template structures. Using the *Rc* XMod-Doc:Coh protein sequences we performed a protein BLAST [282], finding one satisfactory homologue template for the *Rc* XMod-Doc domain, and two for the *Rc* Coh domain. For all the templates, the sequence identity was observed to be low: 20% identity between *Rc* XMod-Doc and the *R. flavefaciens* XMod-Doc (PDB 4IU3) template; 15% identity between *Rc* Coh domain and the *Rf* Coh E (PDB 4IU3) template; 18% identity between *Rc* Coh domain and the *Rf* Coh G (PDB 4WKZ) template. Likewise, the sequence similarities were also found to be small, with 35%, 28%, and 34% respectively. Regarding the templates, both the 4WKZ [262] and the 4IUI3 [165] structures were solved by means of x-ray crystallography, with a resolution of 1.79Å and 1.97Å, respectively.

Using Modeller, we generated 10 structural models for the *Rc* XMod-Doc domain based on its template, and 20 structural models for the *Rc* Coh domain based on its two templates (10 models for each template). Using VMD [263], the structure of the *Rf* XMod-Doc:Coh complex was used as a guide to fit all 200 possible combinations of the *R. champanellensis* model structures into binding mode A. For the binding mode B, first an inverted *Rf* XMod-Doc:Coh binding was created by superimposing Doc helix 1 with helix 3, and helix 3 with helix 1, creating a 180^{o} rotated Coh structure. VMD was then used again to fit all the possible 200 models of *R. champanellensis* to this inverted *R. flavefaciens* structure. Typically the best structural model could be selected by employing tools like PROCHECK [283] and ERRAT server [284], however, due to the low sequence identity and similarity, we adopted a strategy of using molecular dynamics (MD) to thoroughly test all the homology models.

Employing QwikMD[270], all 400 model structures were subjected to 5 ns of equilibrium MD to ensure conformational stability. Although after a visual inspection we could see that many of the structural models were not stable following MD simulation, we chose to use a more systematic metric to select the best structural models, namely, selecting, for each of the binding modes, the 5 most stable models under load. For that we performed 20 ns of steered molecular dynamics (SMD) for each of the 400 model structures, pulling the complex apart. The simulations revealed that the complexes would rupture at a wide-range of forces, and the 5 models with highest rupture forces for each binding mode were selected as the best models.

To investigate the stability of the best structural models, we performed another set of SMD simulations using the 5 best models as initial structures in what we call an *in silico* force-spectroscopy approach [161]. Using a wide-sampling strategy, 200 steered molecular dynamics (SMD) replicas were carried out for a total of 8 μs for each binding mode, using the 5 different initial structures. All SMD simulations [237] were performed with a constant velocity protocol using 5.0 Å/ns as the pulling speed. In all simulations, SMD was employed by restraining the position of the N-terminal of XMod-Doc domain, while pulling on the C-terminus of the Coh domain.

To reproduce the scenario where the XMod had unfolded, we performed another set of SMD simulations where the XMod was removed using QwikMD. Using a wide-sampling strategy, 20 steered molecular dynamics (SMD) replicas were carried out for a total of 800 ns. This new set of SMD was performed by restraining the position of the N-terminal of Doc domain, while pulling on the C-terminus of the Coh domain.

In our study, all MD and SMD simulations were performed employing the NAMD molecular dynamics package [268,269]. The CHARMM36 force field [285] was employed to describe all simulations, using an explicit TIP3 water model [286]. Simulations were performed at the NpT ensemble, in periodic boundary conditions. Temperature was kept at 300K using Langevin dynamics for temperature coupling, while a Langevin piston was employed to hold pressure at 1 bar. A distance cut-off of 14.0 Å was applied to short-range, non-bonded interactions, whereas the particle-mesh Ewald (PME) method was employed for long-range electrostatic interactions. The equations of motion were integrated using a 2 fs time step for all simulations performed. All simulations were analyzed using VMD [263] and its plugins. Surface contact areas of interacting residues/domains were studied using PyContact [287].

2.5.2 Cloning

The constructs for AFM measurements were ybbr-ELP-ddFLN4-XMod-Doc-HIS and Coh-ddFLN4-ELP-HIS-ybbr. A ddFLN4 domain was inserted into a pET28a vector containing ybbr-HIS-ELP (for XMod-Doc) or ELP-HIS-ybbr (for Coh) so that the ELP linker was located between the ddFLN4 and the ybbr tag. The XMod-Doc synthetic gene was inserted to the C terminus of ddFLN4 using Gibson assembly and the Coh synthetic gene was inserted to the N terminus of ddFLN4 using restriction digestion cloning (NdeI and BamHI sites). The sequences of the inserted genes were confirmed by Sanger sequencing (Microsynth AG). The His-tag on the XMod-Doc constructs were then moved to the C terminus of the construct.

Protein samples for ITC measurement were prepared by removing ELP and ddFLN4 domains from the AFM measurement constructs.

The Coh smFRET construct was prepared by adding an Avi-tag to the N terminus of the ITC construct and introducing E154C mutation to Coh.

The WT XMod-Doc smFRET construct was prepared by replacing the serine at position 199 with an Amber codon (TCC→TGA). The smFRET constructs of the XMod-Doc binding mode mutants were prepared by adding the same mutations as the corresponding AFM constructs to the WT XMod-Doc smFRET construct.

2.5.3 Protein expression and purification

All protein samples used for AFM and ITC as well as Coh used in smFRET were expressed in NiCo21 (DE3) cells (New England Biolabs). Cells were cultured in TB (terrific broth) medium containing 50 µg/mL kanamycin until OD600 reached ~0.6. Protein expression was induced by adding 0.5 mM IPTG to the culture, followed by incubating at 20 °C overnight.

Cells were harvested and lysed using sonication. The cell lysate was pelleted and the supernatant was loaded onto a His-tap FF 5 mL column (GE Healthcare) and washed with TBS buffer supplemented with calcium (TBS-Ca, 25mM Tris, 72mM NaCl, 1 mM $CaCl_2$, pH 7.2). Bound protein was eluted using TBS-Ca buffer containing 500 mM imidazole. Eluted protein was further purified using size-exclusion column (GE Healthcare). Protein solutions for long term storage were concentrated using a Vivaspin 6 centrifugal filter (GE Healthcare) and stored in 35 % (v/v) glycerol at -20 °C. The concentration of the protein stocks were determined to be ~40 µM using UV absorption spectrophotometry.

2.5.4 Amber suppression

The Doc smFRET constructs were expressed in BL21Star (DE3) cells using amber codon suppression[275]. The pET28a vector carrying the Doc smFRET construct was co-transformed with plasmid pEVOL-pAzF (a gift from Peter Schultz, Addgene plasmid # 31186). Cells were grown in LB medium containing 50 µg/mL kanamycin and 25 µg/mL chloramphenicol until OD600 reached ~0.8. Cells were then pelleted, washed with M9 minimal medium and resuspended in M9 medium containing 50 µg/mL kanamycin, 25 µg/mL chloramphenicol, 0.2 mg/ml p-azido-l-phenylalanine (pAzF) and 0.02% arabinose. The culture was incubated at 37 °C for 1 h and then 1 mM IPTG was added to the culture, followed by incubating at 16 °C overnight. The expressed protein was extracted and purified using the same protocol as for the AFM constructs.

2.5.5 AFM sample preparation

The preparations of AFM measurement samples were conducted according to previously published protocols[271]. Biolever mini AFM cantilevers (Bruker, Billerica, MA, USA) and cover glasses were cleaned by UV-ozone treatment (cantilevers) or piranha solution (cover glasses), and silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce amine groups on the surface. The silanized cantilevers and cover glasses were subsequently incubated with 10 mg/mL sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC) solution for 30 min at room temperature in order to introduce maleimide groups on the surface. After incubating with sulfo-SMCC, the cantilevers and glasses were cleaned with ddH_2O and immediately incubated with 20 mM coenzyme A (CoA) solution for 2 h at room temperature and then cleaned with ddH_2O. CoA-coated cantilevers and cover glasses were incubated with Coh-ddFLN4-ELP-ybbR and ybbR-ELP-ddFLN4-XMod-Doc fusion proteins, respectively, in the presence of ~5 µM Sfp (phosphopantetheinyl transferase) enzyme and 10 mM $MgCl_2$ for 2 h at room temperature. After incubation, cantilevers and glass surfaces were intensively rinsed with TBS-Ca buffer and stored under TBS-Ca before measurement.

2.5.6 AFM-SMFS measurements

SMFS measurements were performed on a Force Robot AFM (JPK instruments, Berlin, Germany). Cantilever spring constants (ranging from 0.07 N/m to 0.1 N/m) were calibrated using the contact-free method. A control experiment was done showing that the contact-free calibration method gave the same result as contact-based method. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from 100 nm/s to 6400 nm/s. In a typical measurement around 5,000-10,000 force-extension curves were obtained with a single cantilever in an experimental run of 10-20 h. The majority of the data were unusable curves due to lack of interactions, multiple interactions or nonspecific adhesion of molecules to the cantilever tip. However, ~10% of the curves showed single-molecule interactions. We filtered the data by searching for the two-step unfolding patterns and the 64 nm contour length increment of two ddFLN4 fingerprint domains.

2.5.7 AFM data analysis

AFM data were analyzed using a combination of Python scripts, R scripts (R foundation, available at https://www.r-project.org/, utilizing packages readr and ggplot2 and user interface R Studio, available at https://www.rstudio.com/), and Origin 2018 (OriginLab).

Force-extension curves were transformed into contour length space using freely rotating chain (FRC) model, which assumes bonds of length b are connected by a fixed angle Υ. The force-extension curves were transformed to contour length L using Eq. 1.17 [143].

The force-extension curves were screened using the ~64 nm contour length increment of two ddFLN4 fingerprint domains.

The most probable rupture force of the complex and unfolding force of XMod was fitted linearly against the logarithm of loading rate (r_f) to extract the zero-force off rate k_0 and the distance to the energy barrier $\Delta x^{\ddagger}$ using Eq. 1.5, as explained by the Bell-Evans model [123,124].

The stochastic nature of complex rupture and domain unfolding leads to biasing effect[60], where the XMod unfolding event cannot be observed after the complex rupture. Therefore the fitted parameters of the XMod unfolding were corrected by simulation of the forced pulling process. With fixed energy barrier parameters for the complex, k_0 and $\Delta x^{\ddagger}$ for XMod unfolding were adjusted using least square fitting method to yield the closest ratio of curves showing the XMod unfolding referred to the experimental observations at different pulling speeds.

The rupture force of the complex and unfolding force of XMod were plotted into histograms, transformed into force-dependent off rate values and fitted using the Dudko-Hummer-Szabo model [128,129], as explained below.

Histograms were plotted using equal bin width ΔF = 40 pN. For one histogram containing N bins, starting from F_0 and ending at $F_N = F_0 + N\Delta F$. The k_{th} bin can be directly transformed into the force-dependent rate constant value using Eq. 1.10.

Based on Kramers theory, the force-dependence of $k_{off}(F)$ can be written as Eq. 1.6.

Smooth rupture force histograms (kernel density estimation) were transformed into force-dependent off rate using Eq. 1.9, which is the continuous form of Eq. 1.6.

2.5.8 Single-molecule FRET sample and chamber preparation

The Coh smFRET construct was reduced by adding 5 mM TCEP (tris(2-carboxyethyl)phosphine) and incubating at room temperature for 30 min. The reduced protein was mixed with 20-fold excess of maleimide-Cy3b (GE Healthcare) and incubated at room temperature for 1 h followed by incubation at 4 °C overnight. The XMod-Doc smFRET constructs incorporated with p-azido-phenylalanine were labeled by mixing with 5-fold excess of DBCO-AF647 (Jena Bioscience, Jena, Germany) and incubating at room temperature for 1 h, followed by incubation at 4 °C overnight. The labeled Coh and XMod-Doc constructs were purified using a desalting column followed by size-exclusion column.

smFRET experiments were carried out in Lab Tek chambers (Lab-Tek II chambered coverglass system, Thermo Fisher Scientific). Prior to the measurement, chambers were passivated with 1 mg/ml BSA (PAA Laboratories GmbH, Germany) for at least 1 hour. BSA solution was removed only before the measurement and the chamber was washed twice with PBS and once with measurement buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM $CaCl_2$, 2 mM Trolox/Trolox quinone and 1% glucose).

2.5.9 Single-molecule FRET measurements

Solution smFRET experiments were performed on a PIE-based[288] home built confocal microscope based on an Olympus IX-71 inverted microscope. Two pulsed lasers (639 nm, 80 MHz, LDH-D-C-640; 532 nm, 80 MHz, LDH-P-FA-530B, both from PicoQuant GmbH) were altered on the nanosecond timescale by a multichannel picosecond diode laser driver (PDL 828 “Sepia II”, PicoQuant GmbH, Berlin, Germany) with an oscillator module (SOM 828, PicoQuant GmbH). The lasers were coupled into a single mode fiber (P3-488PM-FC, Thorlabs GmbH, Dachau, Germany) to obtain a Gaussian beam profile. Circular polarized light was obtained by a linear polarizer (LPVISE100-A, Thorlabs GmbH) and a quarter-wave plate (AQWP05M- 600, Thorlabs GmbH, Dachau, Germany). The light was focused by an oil-immersion objective (UPLSAPO100XO, NA 1.40, Olympus Deutschland GmbH) onto the sample. The sample was moved by a piezo stage (P-517.3CD, Physik Instrumente (PI) GmbH & Co. KG, Karlsruhe, Germany) controlled by a E-727.3CDA piezo controller (Physik Instrumente (PI) GmbH & Co. KG, Karlsruhe, Germany). The emission was separated from the excitation beam by a dichroic beam splitter (z532/633, AHF analysentechnik AG) and focused onto a 50 μm pinhole (Thorlabs GmbH). The emission light was split by a dichroic beam splitter (640DCXR, AHF analysentechnik AG) into a green (Brightline HC582/75, AHF analysentechnik AG; RazorEdge LP 532, Laser 2000 GmbH) and red (Shortpass 750, AHF Analysentechnik AG; RazorEdge LP 647, Laser 2000 GmbH) detection channel. Emission was focused onto avalanche photodiodes (SPCM-AQRH-14-TR, Excelitas Technoligies GmbH & Co. KG) and signals were registered by a time-correlated single photon counting (TCSPC)-unit (HydraHarp400, PicoQuant GmbH, Berlin, Germany). The setup was controlled by a commercial software package (SymPhoTime64, Picoquant GmbH, Berlin, Germany). Excitation powers of 36 µW and 25 µW were used for donor and acceptor lasers (as measured in front of the entrance of the microscope).

Labeled Coh and XMod-Doc samples were mixed in a molar ratio of 1:1 at a concentration of 1 µM, incubated for 1 minute, and finally diluted in the chamber to a concentration of 200 pM.

2.5.10 Single-molecule FRET data analysis

smFRET burst selection was performed using a sliding time window burst search algorithm, with a time window of 500 µs and a minimum of 4 photons per time window. A threshold for burst detection of 40 photons was used[289]. In order to sort out photobleaching and blinking events, ALEX-2CDE [290] and ǀTDX-TAAǀ filters [291] were used. Doubly-labeled XMod-Doc:Coh complexes were further selected by keeping the stoichiometry parameter between 0.2 and 0.8. Accurate FRET efficiencies [256,292] were calculated from fluorescence intensities as:

$$E = \frac{I_{DA} - \alpha I_{DD} - \delta I_{AA}}{\gamma I_{DD} + I_{DA} - \alpha I_{DD} - \delta I_{AA}} \qquad \text{Eq. 2.1}$$

where I_{DA}, I_{AA} and I_{DD} are the background-corrected photon counts in the acceptor channel after donor excitation, the acceptor channel after acceptor excitation, and the donor channel after donor excitation. The α and δ correction parameters are calculated from donor only and acceptor only subpopulations and accounts for spectral cross talk and direct excitation of the donor dye. The different detection efficiencies and quantum yields of fluorophores are corrected with the γ correction factor [256,292].

2.5.11 ITC measurement

The titration was carried out at 25 °C using a VP-ITC instrument[293]. The analyte was 16.1 µM Coh (lacking ELP linker and ddFLN4 domains) and the injectant was 126 µM XMod-Doc protein (lacking ELP linker and ddFLN4 domains). Both protein samples were in the TBS-Ca buffer. The titration was carried out by injecting XMod-Doc dropwise into the analyte. Each drop contained 10 µL XMod-Doc solution and there was 5 min retention time between two consecutive drops so that the system could equilibrate after injecting a drop. The power required to maintain equal temperature between the sample cell and the reference cell (filled with water) was recorded. The titration was terminated after 27 injections, when the analyte (Coh) was fully saturated by the injectant (XMod-Doc).

2.5.12 Monte Carlo simulation

A Monte Carlo approach based on Kramers theory was used to validate the multi-state kinetic model. The receptor-ligand dissociation in combination with fingerprint domain unfolding was simulated in a constant pulling speed protocol. Briefly, the XMod-Doc:Coh complex was randomly assigned a binding mode to be either binding mode A (80% possibility) or binding mode B (20% possibility). The corresponding kinetic parameters (k_0 and $\Delta x^{\ddagger}$, see Table 1) extracted from AFM-SMFS were used for the simulation. A series of force values $F(t_i)$ was generated on an evenly distributed extension axis $X(t_i)$ using a worm-like chain (WLC) model [141]. Due to the fact that the constant pulling speed protocol is achieved by the constant speed pulling of the AFM head instead of the AFM tip, a bending correction was done by converting the molecular extension $X(t_i)$ to the AFM head height $H(t_i)$ using Eq. 2.2:

$$H(t_i) = X(t_i) + \frac{F(t_i)}{k} \qquad \text{Eq. 2.2}$$

where k is the spring constant of the AFM cantilever. Then the time series could be generated based on the pulling speed V:

$$t_{i+1} = t_i + \frac{H(t_{i+1}) - H(t_i)}{V} \qquad \text{Eq. 2.3}$$

During each time slice($\Delta t = t_{i+1} - t_i$), the probability of XMod-Doc:Coh rupture or protein domain unfolding was calculated using the following equation:

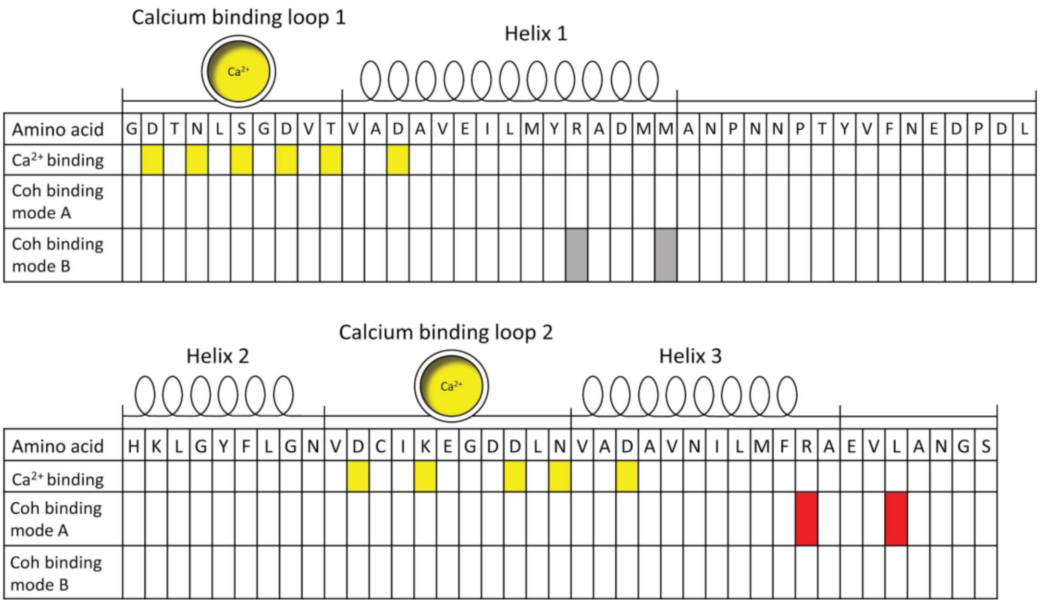
$$P(F) = 1 - e^{-k_{off}(F)\Delta t} \qquad \text{Eq. 2.4}$$

where $k_{off}(F)$ can be drawn from Eq. 1.2 following the Bell-Evans model.

The dissociation probability is compared to a random number between zero and unity. If the random number is smaller than $P(F)$ the rupture or unfolding event occurs and the corresponding force is recorded as the rupture or unfolding force. For each pulling speed, 1000 curves were generated and a histogram was drawn for the complex rupture force as well as the XMod unfolding force (**Figure 2.S14** and **2.S15**). For simulation under force clamp conditions, a constant force was used and 1000 curves were generated to calculate the lifetime of the complex under each applied force (**Figure 2.S16**).

2.6 Supplementary information

2.6.1 Supplementary figures



Amino acid	G	D	T	N	L	S	G	D	V	T	V	A	D	A	V	E	I	L	M	Y	R	A	D	M	M	A	N	P	N	N	P	T	Y	V	F	N	E	D	P	D	L
Ca^{2+} binding		🟨		🟨		🟨		🟨		🟨			🟨																												
Coh binding mode A																																									
Coh binding mode B																					▨				▨																

Amino acid	H	K	L	G	Y	F	L	G	N	V	D	C	I	K	E	G	D	D	L	N	V	A	D	A	V	N	I	L	M	F	R	A	E	V	L	A	N	G	S
Ca^{2+} binding											🟨			🟨				🟨		🟨			🟨																
Coh binding mode A																															🟥				🟥				
Coh binding mode B																																							

Figure 2.S1 Amino acid sequence and secondary structure elements of dockerin. The residues involved in calcium binding are shown in yellow, the residues mutated to knock out the binding mode A are shown in red and the residues mutated to knock out the binding mode B are shown in grey.

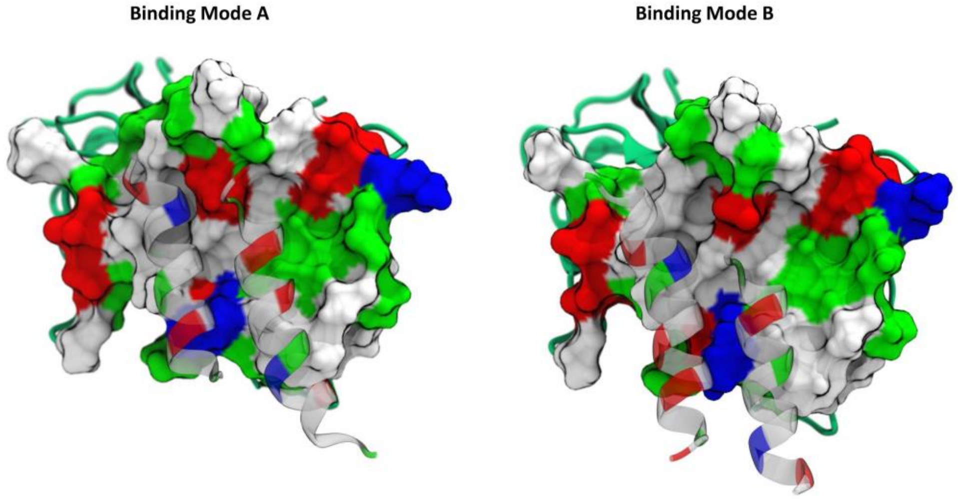


Figure 2.S2 Model of binding interface between dockerin and cohesin in both binding modes. In both binding modes, the dockerin and cohesin residues at the binding interface have complementary physical properties, forming a hydrophobic core surrounded by hydrophilic residues. Red: negatively charged residues; blue: positive charged residues; white: non-polar residues and green: polar residues.

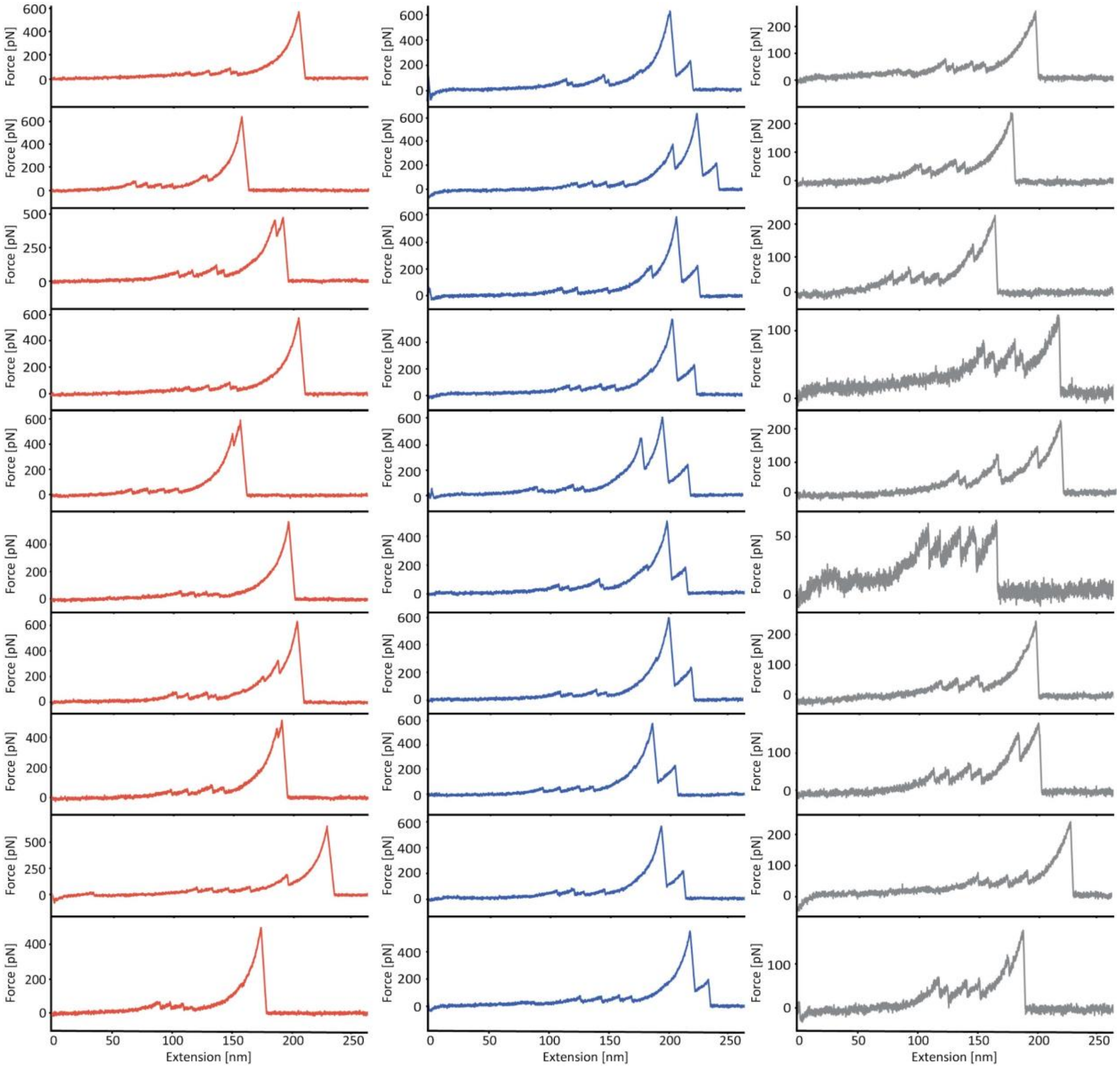


Figure 2.S3 Example force-extension curves obtained in AFM-SMFS measurements. Examples of pathways 1 (left column, red), 2 (middle column, blue) and 3 (right column, grey) force-extension curves obtained in AFM-SMFS measurements of WT XMod-Doc:Coh complex. Some force curves showed unassigned unfolding events between ddFLN4 unfolding and complex rupture or XMod unfolding. These unfolding events broadened the contour length histogram and can be attributed to partial unfolding of Coh or Doc domains.

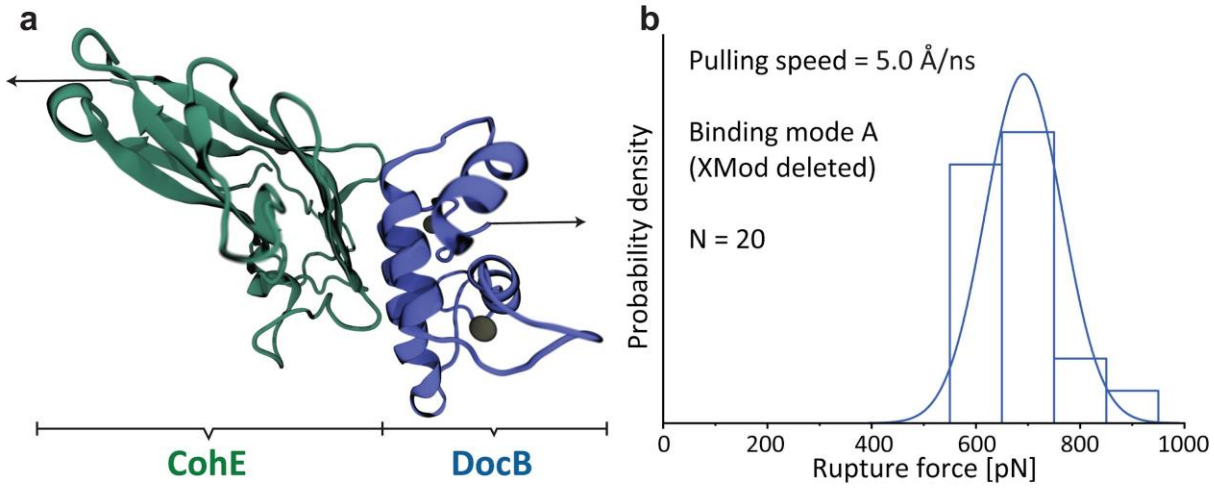


Figure 2.S4 SMD simulations of Doc:Coh complex lacking the XMod. a: Structural model showing the Doc:Coh complex in binding mode A. The XMod was deleted from the Doc to mimic the complex in pathway 2. The pulling directions are marked with black arrows. **b:** The rupture force of Doc:Coh complex in binding mode A was measured by performing 20 SMD replicas at 5.0 Å/ns pulling speed (same as Fig. 1 bc) and plotted in a histogram. The histogram was fitted with a Gaussian distribution, giving the most probable rupture force of 692 pN.

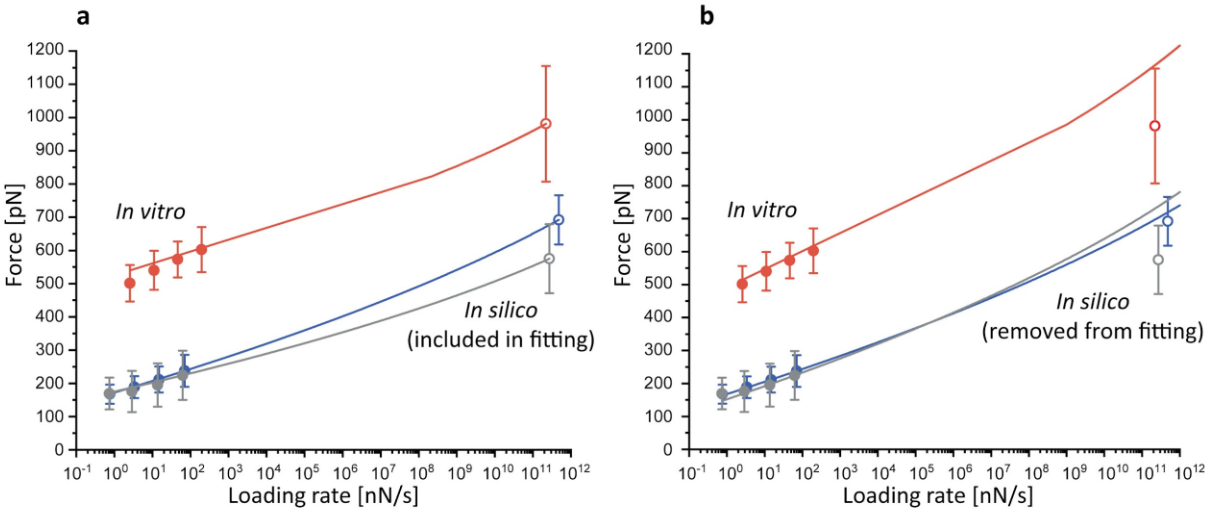


Figure 2.S5 Rupture force *vs.* loading rate plot from *in vitro* and *in silico* measurements. The most probable rupture forces measured with AFM-SMFS (*in vitro*) and SMD simulations (*in silico*) were plotted against the loading rate and fitted using the Dudko-Hummer-Szabo model (see supplementary note 3.3) for pathways 1 (red), 2 (blue) and 3 (grey). The SMD simulation results (open circles) were included in the fit **(a)** or excluded from the fit **(b)**. In both cases, the fitted curve showed the same trend and the SMD data was in good agreement with the experimental data.

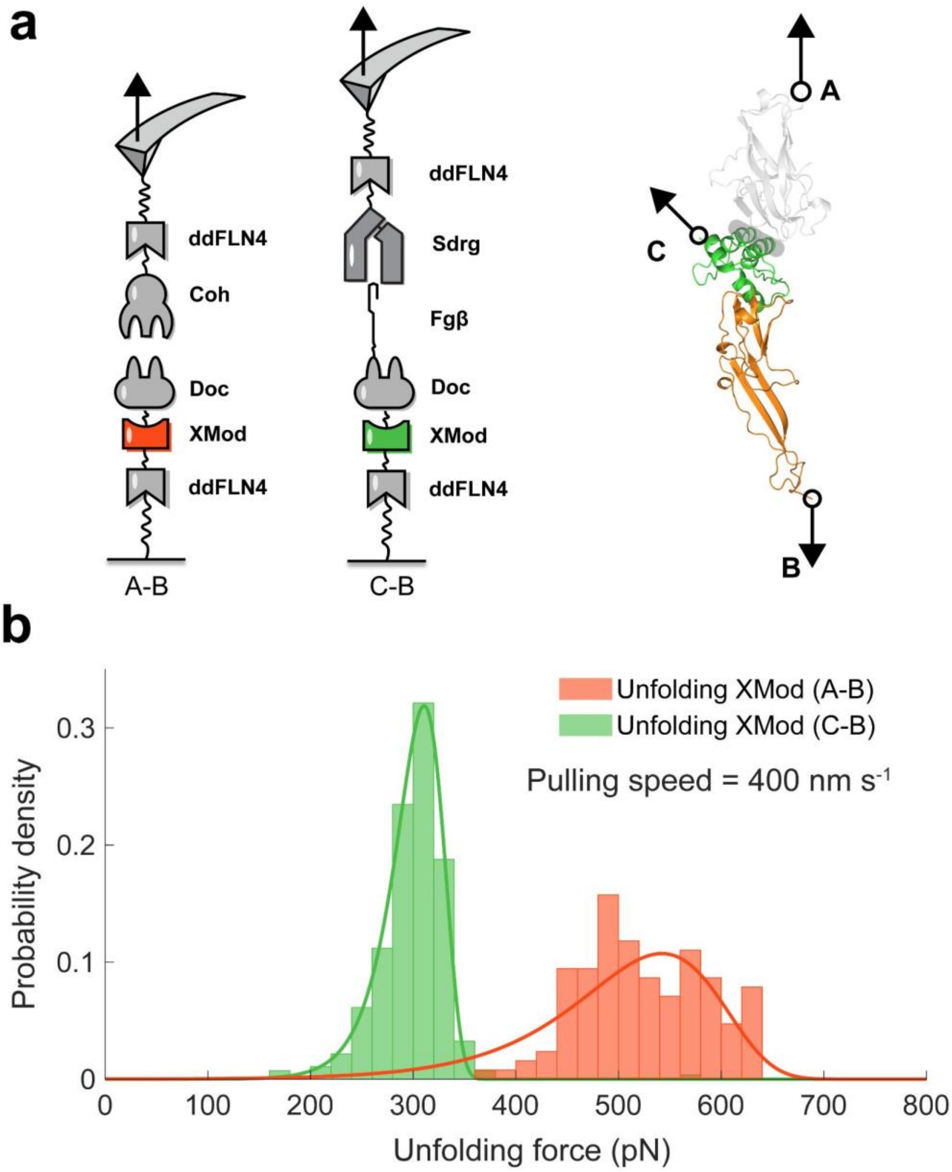


Figure 2.S6 Measuring unbiased XMod unfolding force using SdrG:Fgβ complex. a: Force loading geometries of XMod-Doc. In the native pulling geometry, the force is loaded to the XMod.Doc:Coh complex from the C-terminus of Coh (A) and N-terminus of XMod-Doc (B). In the SdrG:Fgβ-XMod-Doc construct, the force is loaded to the N- and C- termini of XMod-Doc (B and C). **b:** Different force loading geometries gave rise to different XMod unfolding forces. The measured unfolding force is around 550 pN in the native force loading geometry (A-B), but the change of pulling geometry (C-B) significantly decreased the unfolding force of XMod to around 300 pN.

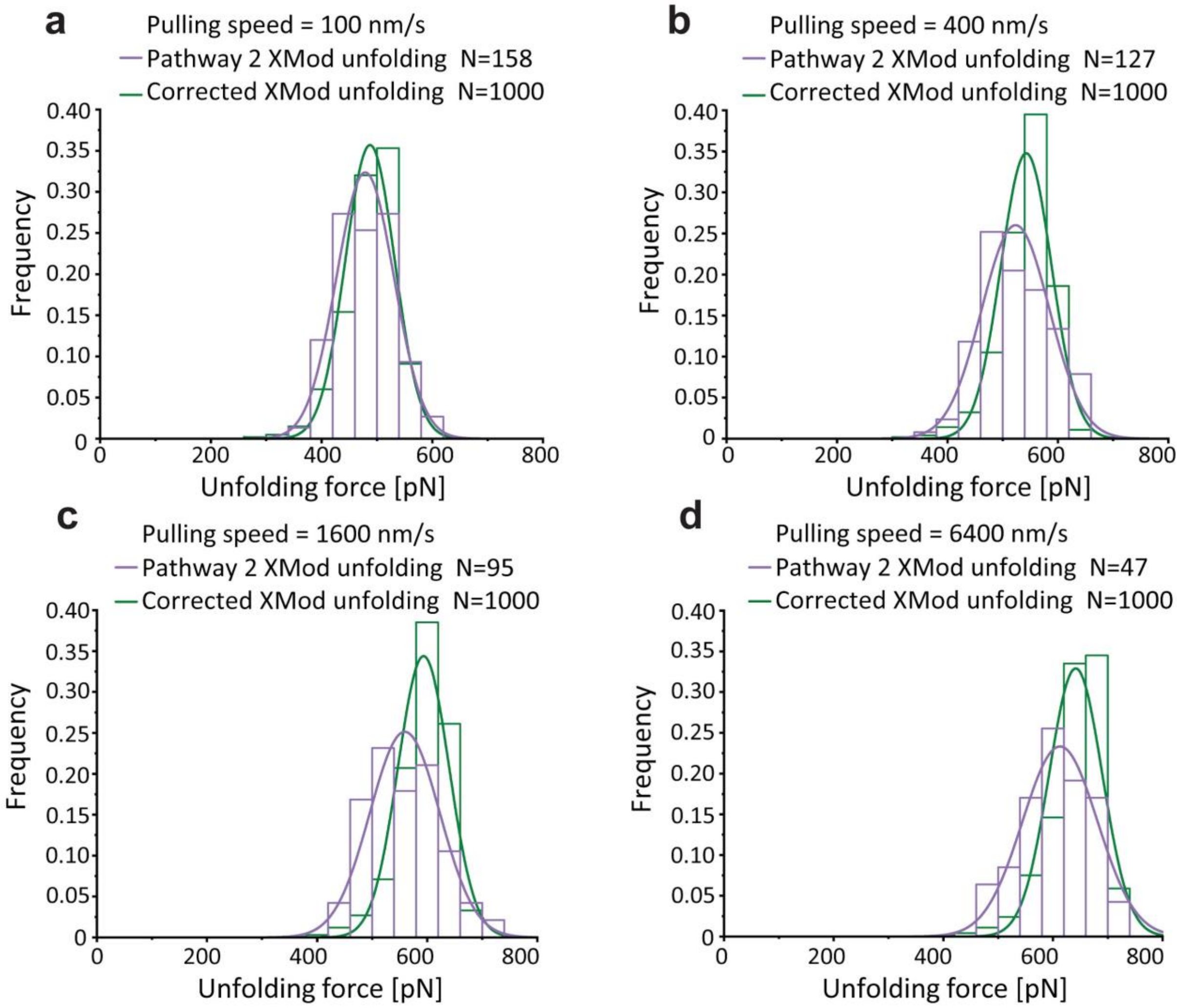


Figure 2.S7 Correction of XMod unfolding force taking biasing effect into account. XMod unfolding forces measured in pathway 2 under four different pulling speeds were plotted as histograms and fitted with Gaussian distributions (purple). The XMod unfolding force distributions were corrected using iterative fitting which takes the biasing effect[60] into account (green).

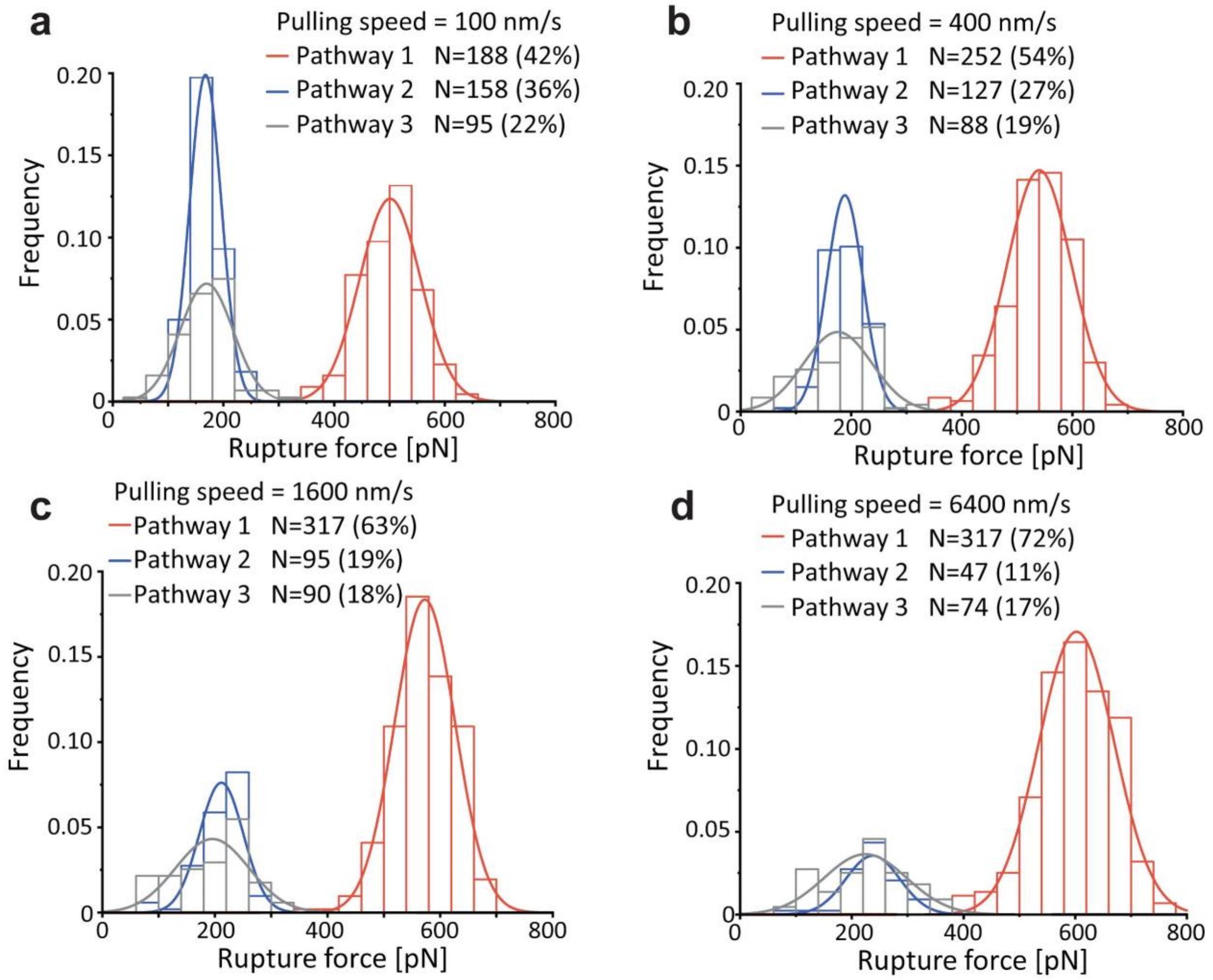


Figure 2.S8 XMod-Doc:Coh complex rupture force histograms at different pulling speeds. The three different pathways are plotted in different colors: High rupture force (P1, red); XMod unfolded, low rupture force (P2, blue); and XMod folded, low rupture force (P3, grey). Each histogram was fitted with a Gaussian distribution. Loss of the P2 population at higher pulling speeds drives catch bond behavior in the force ramp scenario.

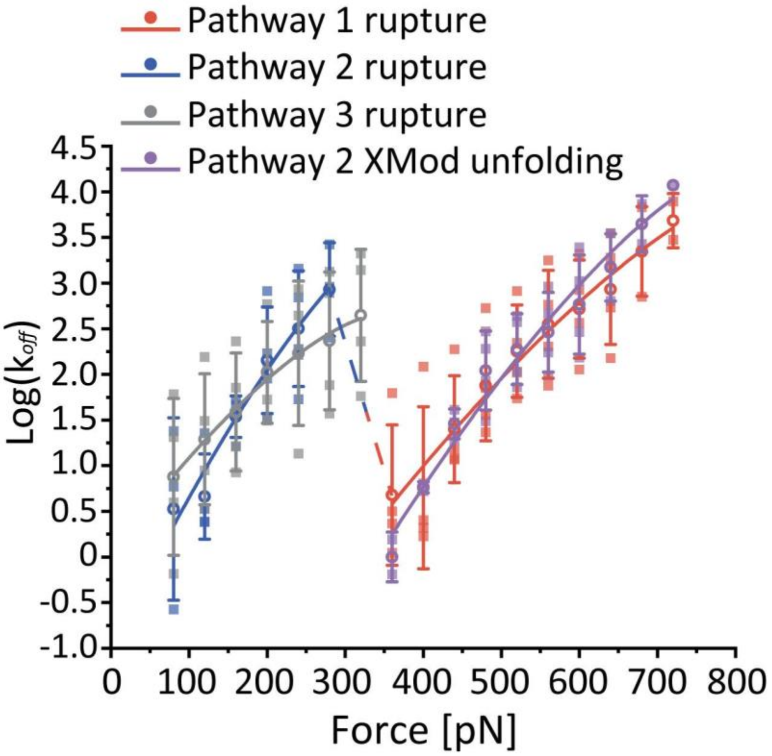


Figure 2.S9 Fitting the force-dependent off rate data with υ=2/3. The force-dependent off-rate from Figure 2.2f was fitted to the analytical expression (Eq. 2.5) with υ=2/3. Error bars represent the standard deviation of off-rate. The fitted $\Delta x^{\ddagger}$, $\Delta G^{\ddagger}$ and k_0 values are listed in Table 2.1.

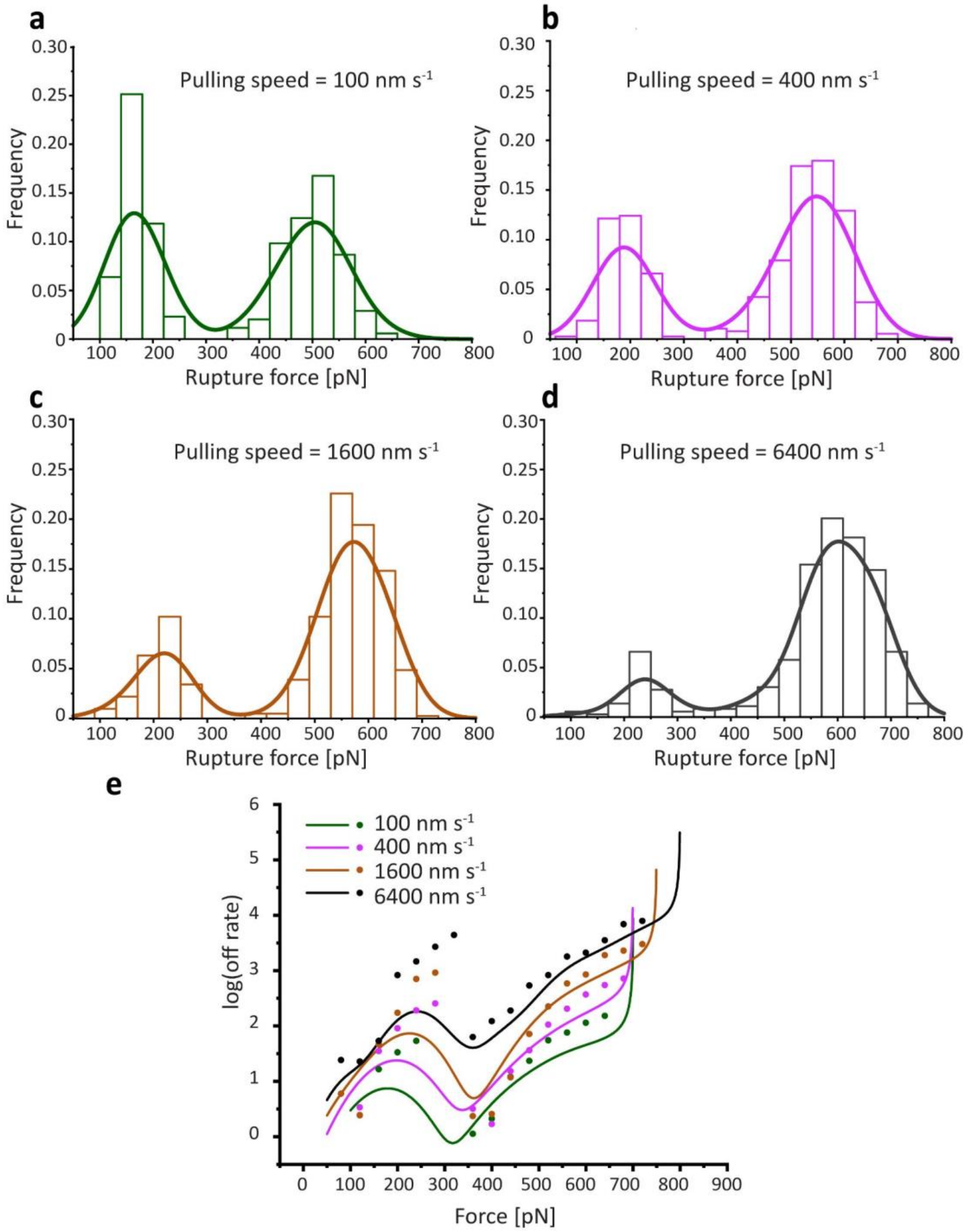


Figure 2.S10 Combined rupture force histograms of pathways 1 and 2 and calculation of force-dependent off rate. a-d: Combined rupture force histograms of pathways 1 and 2 at different pulling speeds smoothed with kernel density estimation *P(F)*. **e:** Force-dependent off rate calculated from the kernel density estimation of combined histograms using Eq. 2.6 (solid line), overlaid with the off rate calculated using Eq. 2.3 (dots). The two equations are essentially the same, where Eq. 2.3 is the discrete form and Eq. 2.6 is the continuous form.

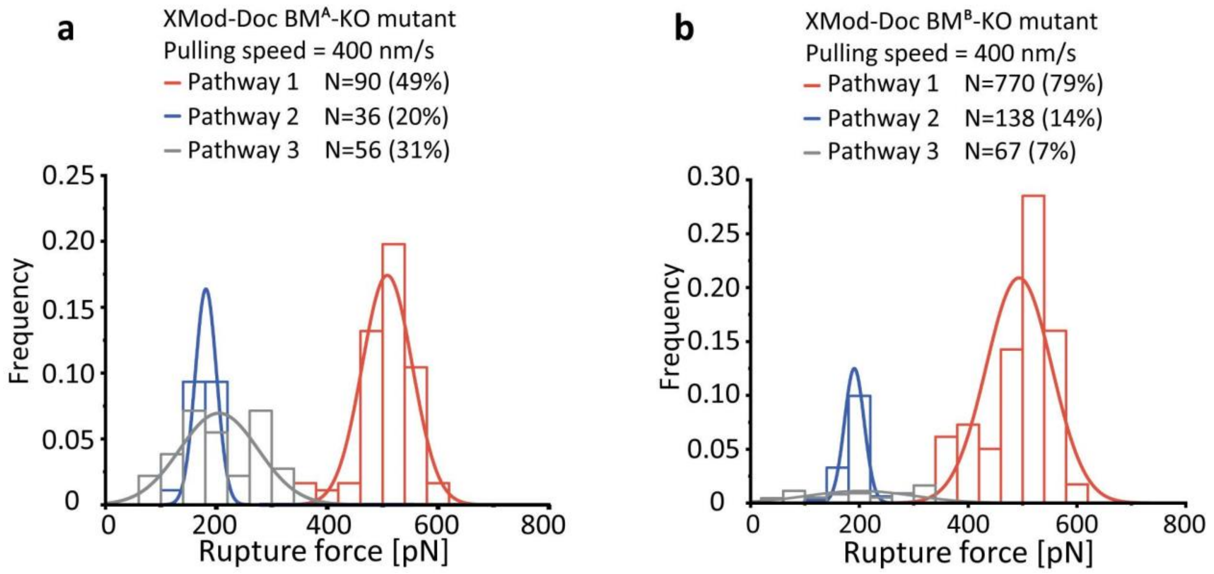


Figure 2.S11 Rupture force histograms of binding mode mutants. a: The rupture force histogram measured using BM^A-KO mutant, showing a decreased percentage of pathways 1 and 2 curves. **b:** The rupture force histogram measured using BM^B-KO mutant, showing a decreased percentage (nearly total loss) of pathway 3 curves.

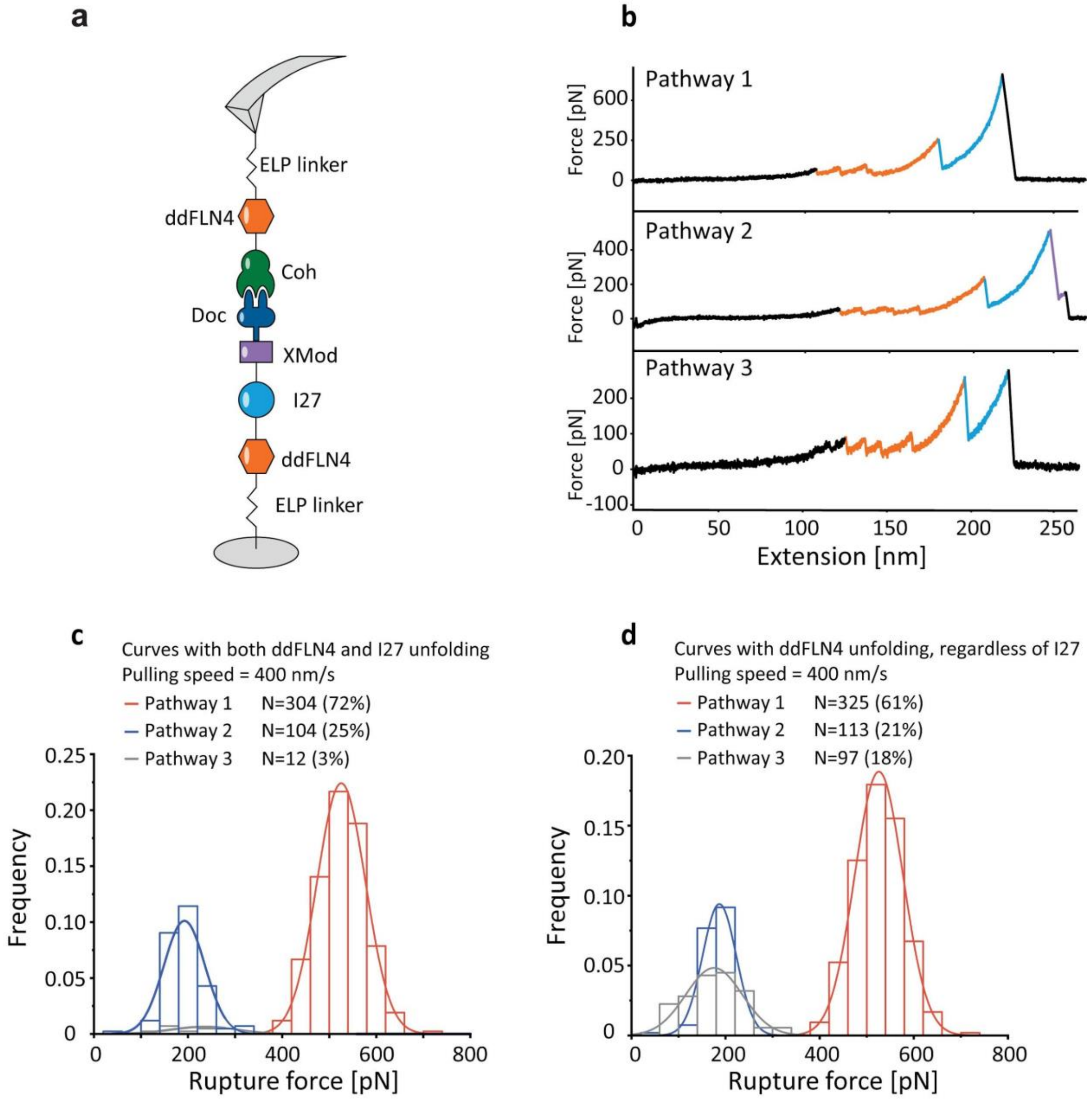


Figure 2.S12 AFM measurements with I27 fingerprint domain. a: AFM setup using the I27 biasing effect to demonstrate two binding modes. **b:** Example force-extension curves showing unfolding of 2x ddFLN4 (in orange) and I27 (in blue) in all three pathways. **c:** Rupture force histogram of force curves filtered with both ddFLN4 and I27 fingerprint domains shows that complexes capable of unfolding I27 rarely (3%) dissociated along pathway 3. **d:** Rupture force histogram of force curves filtered with only ddFLN4 showed that pathway 3 was prevalent in the dataset to the same degree as for WT (~18%), but these curves lacked I27 unfolding events.

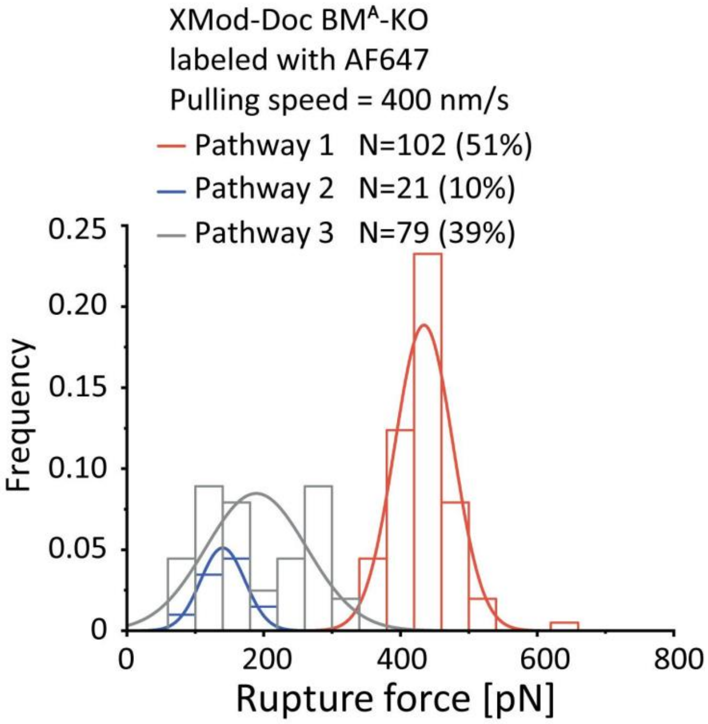


Figure 2.S13 AFM measurement of AF647 labeled BM^A-KO. Given the significant molecular weight of the FRET acceptor dye DBCO-AF647 (~1100 grams/mol) and its proximity to the binding interface in binding mode A, we sought to further understand the influence of dye labeling on binding. We used AFM-SMFS to measure the rupture forces between unlabeled wild-type Coh and BM^A-KO labeled with AF647 at 400 nm/s pulling speed. Binding mode A (P1 + P2) consists of 61% of the force curves, which is lower than the unlabeled BM^A-KO mutant (69%), indicating a decrease in the on-rate of the complex in binding mode A. In addition, the most probable rupture force of P1 was found to be 434 pN for the fluorophore labeled BM^A-KO complex, which is significantly lower than the P1 rupture force measured using unlabeled BM^A-KO (**Figure 2.S10a**, 508 pN). The decrease of rupture force in binding mode A for the dye-labeled construct indicated an increase in the intrinsic off-rate of binding mode A as compared with the unlabeled construct. Therefore, we concluded that the AF647 fluorophore at the C-terminus of XMod-Doc slightly destabilized the complex in binding mode A and decreased the binding affinity in this binding mode. However, fluorophore labeling did not have a significant influence on the rupture force of P3, which corresponds to binding mode B. As a consequence, the binding mode A population was less prevalent than binding mode B in the smFRET measurement. The observed ratios of the two binding modes probed by smFRET reflect the equilibrium scenario, whereas binding mode ratios probed by AFM are governed by differences in on-rates only.

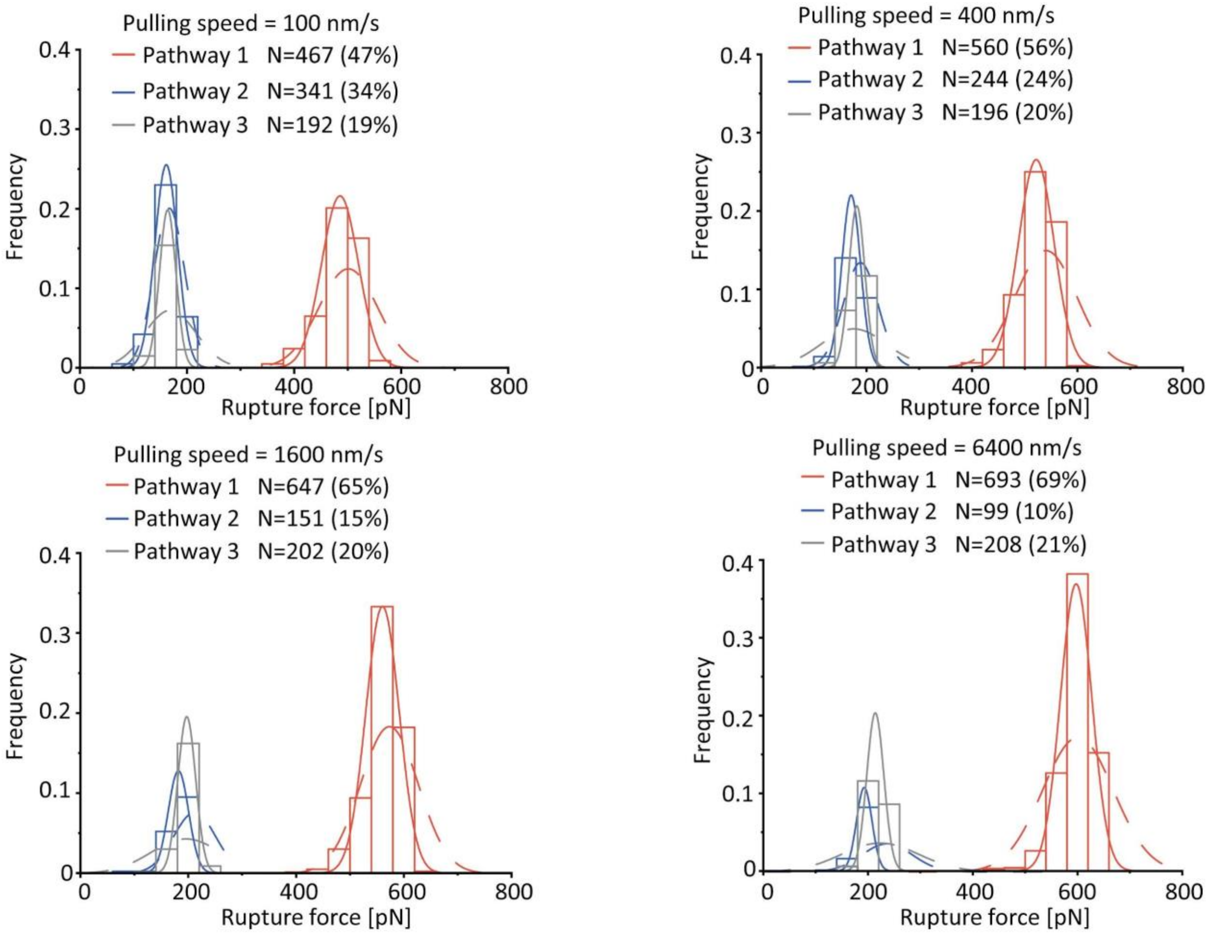


Figure 2.S14 Monte-Carlo simulation results of complex rupture forces. The complex rupture forces were simulated at different pulling speeds and plotted as histograms. The histograms were fitted with Gaussian distributions (solid lines). The corresponding experimental rupture force distributions are shown as dashed lines.

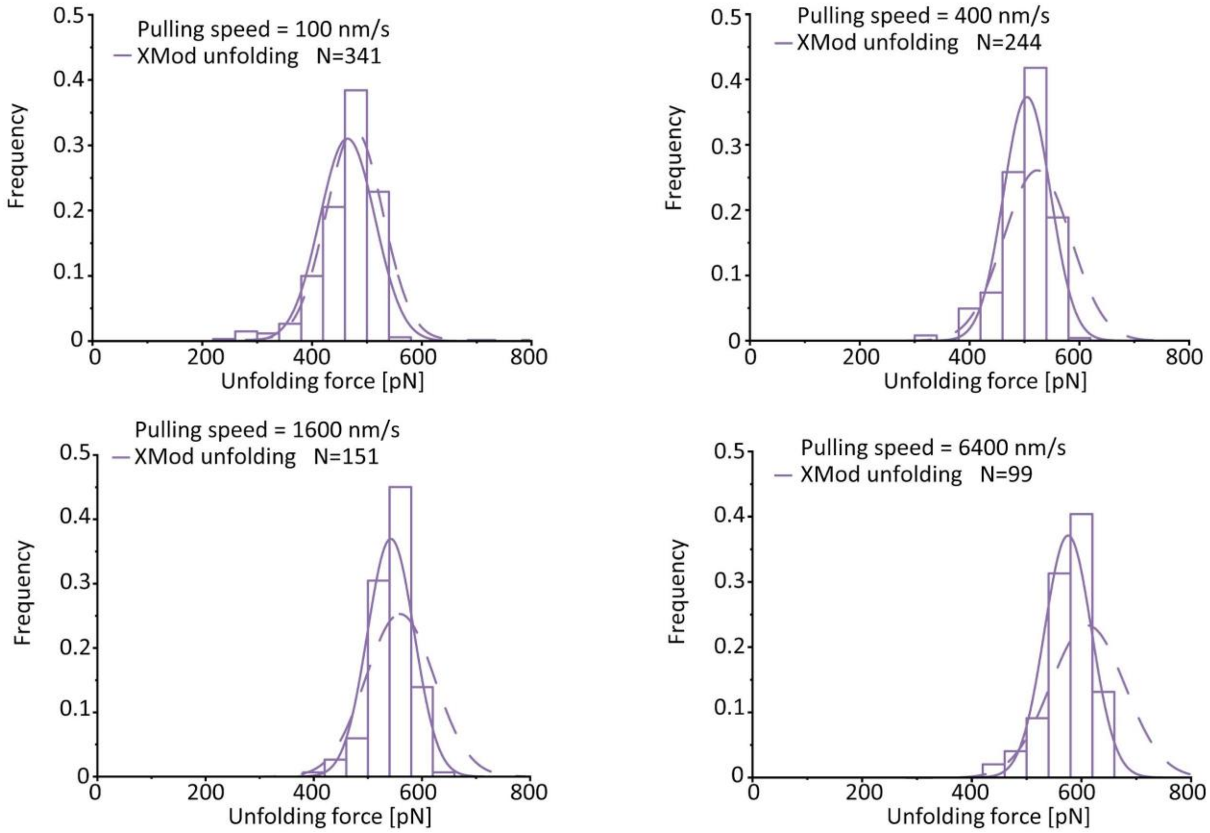


Figure 2.S15 Monte-Carlo simulation results on XMod unfolding force. The XMod unfolding forces in pathway 2 were extracted from Monte Carlo simulations at different pulling speeds and plotted into histograms. The histograms were fitted with Gaussian distributions (solid lines). The corresponding experimental unfolding force distribution is shown as a dashed line.

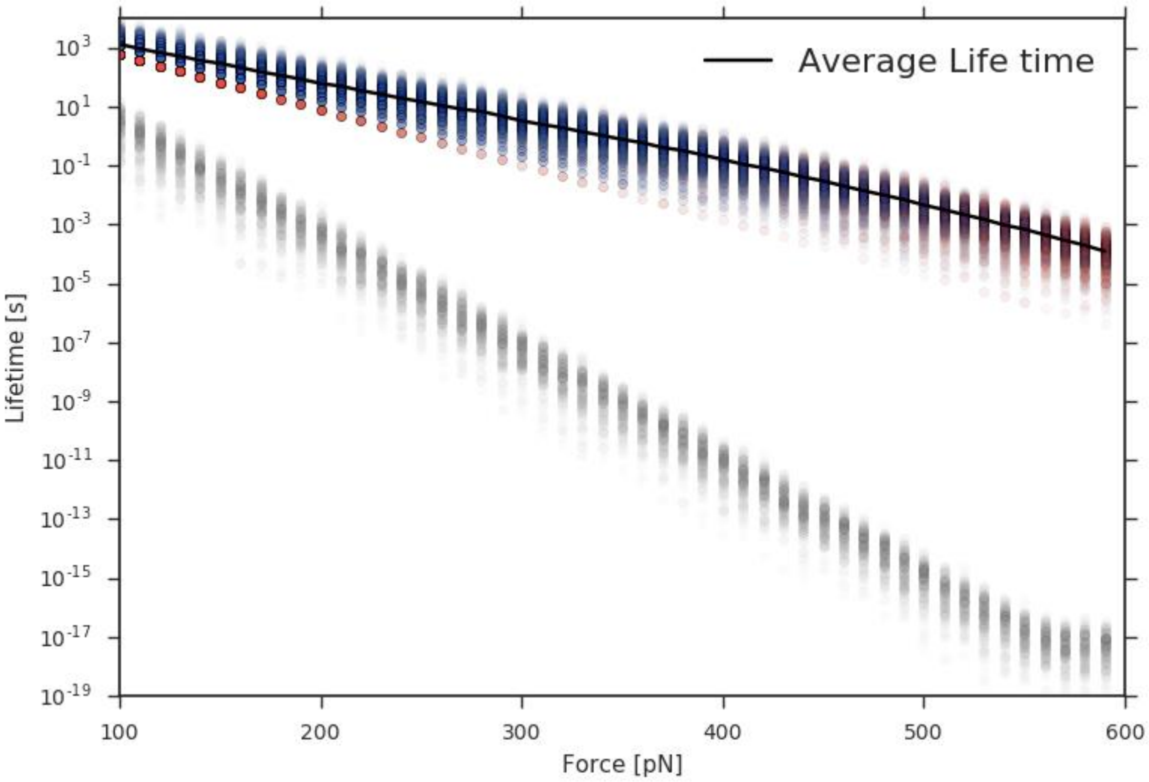


Figure 2.S16 Force clamp Monte-Carlo simulation. The force-dependent lifetime of the complex was simulated in force clamp mode (red: pathway 1, blue: pathway 2, grey: pathway 3). The average lifetime (black line) does not exhibit catch bond behavior for the system probed in force clamp mode.

2.6.2 Supplementary table

Table 2.S1 Kinetic parameters extracted from Monte-Carlo simulations

Pathway	Event	k_0 [s^{-1}] (Bell-Evans)	$\Delta x^{\ddagger}$ (Bell-Evans)
1	High force rupture (XMod intact)	1.05×10^{-6}	0.162
2	Unfolding of XMod (measured)	2.81×10^{-6}	0.161
	Low force rupture after XMod unfold	1.34×10^{-4}	0.371
3	Low force rupture (XMod intact)	5.95×10^{-5}	0.366

2.6.3 Supplementary notes

Supplementary note 2.1 Amino acid sequences

Color codes: ybbr-tag, his-tag, ddFLN4, I27, ELP linker, XMod, Doc, Coh, Avi-tag, inter-domain linker, mutations (AzF = p-azido-phenylalanine).

Addgene #153439: pET28a-ybbr-ELP-ddFLN4-XMod-Doc-HIS (wild type)

GTDSLEFIASKLAWGSGHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGV
GVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPG
VGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVP
GVGVPGVGVPGEGVPGEGVPGWPSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDP
DGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDN
VNGFPKTVTVKPAPGSGSGSGSSGSETTTETSVSQSFSFTEPTRVCYWSHDNRPFDLE
GLKVILQETIFIIHEDGTVVEAEEQPAAVDITEQCMTIFCNDKQVATPEDTYETQIYKY
MLEFAYEGNPDVIVGTHQVLIGVKGDTNLSGDVTVADAVEILMYRADMMANPNNP

TYVFNEDPDLHKLGYFLGNVDCIKEGDDLNVADAVNILMFRAEVLANGSATWEQIVGYDLIDPSDESTFRTIEKLGSHHHHHH

Addgene #153440 pET28a-ybbr-ELP-ddFLN4-XMod-Doc-HIS (BMA-KO)

GTDSLEFIASKLAWGSGHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWPSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSGSSGSETTTETSVSQSFSFTEPTRVCYWSHDNRPFDLEGLKVILQETIFIIHEDGTVVEAEEQPAAVDITEQCMTIFCNDKQVATPEDTYETQIYKYMLEFAYEGNPDVIVGTHQVLIGVKGDTNLSGDVTVADAVEILMYRADMMANPNNPTYVFNEDPDLHKLGYFLGNVDCIKEGDDLNVADAVNILMFAAEVEANGSATWEQIVGYDLIDPSDESTFRTIEKLLGSHHHHHH

Addgene #153441 pET28a-ybbr-ELP-ddFLN4-XMod-Doc-HIS (BM^{B}-KO)

GTDSLEFIASKLAWGSGHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWPSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSGSSGSETTTETSVSQSFSFTEPTRVCYWSHDNRPFDLEGLKVILQETIFIIHEDGTVVEAEEQPAAVDITEQCMTIFCNDKQVATPEDTYETQIYKYMLEFAYEGNPDVIVGTHQVLIGVKGDTNLSGDVTVADAVEILMYAADMEANPNNPTYVFNEDPDLHKLGYFLGNVDCIKEGDDLNVADAVNILMFRAEVLANGSATWEQIVGYDLIDPSDESTFRTIEKLGSHHHHHH

Addgene #153442 pET28a-ybbr-HIS -ELP-ddFLN4-I27-XMod-Doc (wild type)

GTDSLEFIASKLAHHHHHHWGSGHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWPSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLAASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKSAANLKVKELGSGSGSGSSGSETTTETSVSQSFSFTEPTRVCYWSHDNRPFDLEGLKVILQETIFIIHEDGTVVEAEEQPAAVDITEQCMTIFCNDKQVATPEDTYETQIYKYMLEFAYEGNPDVIVGTHQVLIGVKGDTNLSGDVTVADAVEILMYRADMMANPNNPTYVFNEDPDLHKLGYFLGNVDCIKEGDDLNVADAVNILMFRAEVLANGSATWEQIVGYDLIDPSDESTFRTIEKL

Addgene #153443 pET28a-ybbr-XMod-Doc (WT, S199AzF)-HIS

GTDSLEFIASKLAGSSGSETTTETSVSQSFSFTEPTRVCYWSHDNRPFDLEGLKVILQETIFIIHEDGTVVEAEEQPAAVDITEQCMTIFCNDKQVATPEDTYETQIYKYMLEFAYEGNPDVIVGTHQVLIGVKGDTNLSGDVTVADAVEILMYRADMMANPNNPTYVFNEDPDLHKLGYFLGNVDCIKEGDDLNVADAVNILMFRAEVLANGAzFATWEQIVGYDLIDPSDESTFRTIEKLGSHHHHHH

Addgene #153444 pET28a-ybbr-XMod-Doc (BM^{A}-KO, S199AzF)-HIS

GTDSLEFIASKLAGSSGSETTTETSVSQSFSFTEPTRVCYWSHDNRPFDLEGLKVILQETIFIIHEDGTVVEAEEQPAAVDITEQCMTIFCNDKQVATPEDTYETQIYKYMLEFAYEGNPDVIVGTHQVLIGVKGDTNLSGDVTVADAVEILMYRADMMANPNNPTYVFNEDPDLHKLGYFLGNVDCIKEGDDLNVADAVNILMFAAEVEANGAzFATWEQIVGYDLIDPSDESTFRTIEKLLGSHHHHHH

Addgene #153445 pET28a-ybbr-XMod-Doc (BM^{B}-KO, S199AzF)-HIS

GTDSLEFIASKLAGSSGSETTTETSVSQSFSFTEPTRVCYWSHDNRPFDLEGLKVILQETIFIIHEDGTVVEAEEQPAAVDITEQCMTIFCNDKQVATPEDTYETQIYKYMLEFAYEGNPDVIVGTHQVLIGVKGDTNLSGDVTVADAVEILMYRADMMANPNNPTYVFNEDPDLHKLGYFLGNVDCIKEGDDLNVADAVNILMFRAEVLANGAzFATWEQIVGYDLIDPSDESTFRTIEKLGSHHHHHH

Addgene #153446 pET28a-Coh-ddFLN4-ELP-HIS-ybbr

ADGAAKLSMDQKFAEPGETVEIALNLENFDASWTGLEFLVNYDPKLEVALDGAGDIDYSYGDAIGAMGKKISVGGAISKDLTADGLKGFAFAWGTATAISGNGQLGVFKFTVPADAQPGDEFPVNLTVNVGSFIDANKENIPFETVNGWIKIKEEGSGSGSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWRGHHHHHHGSDSLEFIASKLA

Addgene #153447 pET28a-Avi-Coh (E154C)-HIS

GLNDIFEAQKIEWHEGSGSADGAAKLSMDQKFAEPGETVEIALNLENFDASWTGLEFLVNYDPKLEVALDGAGDIDYSYGDAIGAMGKKISVGGAISKDLTADGLKGFAFAWGTATAISGNGQLGVFKFTVPADAQPGDEFPVNLTVNVGSFIDANKENIPFETVNGWIKIKCEGSHHHHHH

Supplementary note 2.2 Quantifying dual-binding mode behavior using fingerprint domain biasing effect

As shown in **Figure 2.S12a**, a titin I27 domain, which under our conditions has an unfolding force around ~200 pN[14,63], was inserted between ddFLN4 and WT XMod-Doc as an additional fingerprint domain. The interaction between XMod-Doc and Coh was then probed using AFM-SMFS in the presence of two ddFLN4 and one I27 fingerprint domain. Unfolding of I27 was identified by its unfolding force of ~200 pN and the contour length increment of ~28 nm, as shown in **Figure 2.S12b**. Force-extension curves were screened based on the contour length increments given by two ddFLN4 domains and one I27 domain and then sorted into the aforementioned three pathways based on the rupture force of the complex and the folding state of XMod. The rupture force of each pathway was plotted in a rupture force histogram (**Figure 2.S12c**). In binding mode A, the complex was able to resist an external force up to ~500 pN prior to rupture, which was larger than the force required to unfold I27. Therefore, pathways 1 and 2 were not biased by the additional I27 domain and were still observable in the dataset. However, binding mode B has relatively low mechanical stability and ruptures prior to I27 unfolding. Therefore, the frequency of pathway 3 was significantly decreased to only 3% when using I27 as an additional fingerprint domain for curve selection. However, when screening the force curves only based on the two ddFLN4 domains regardless of whether the curves contained I27 unfolding or not, the frequency of different pathways in the screened curves (**Figure 2.S12d**) was the same as the construct lacking I27 (**Figure 2.S7b**). This observation further demonstrated that pathways 1 and 2 (high force curves) belong to a different discrete binding mode than pathway 3 (low force curves). The two binding modes have different mechanical stabilities and cannot be converted to the other one on the timescale of the AFM-SMFS curve (~1 second).

Supplementary note 2.3 Dudko-Hummer-Szabo model describing the loading rate dependency of rupture forces

The rupture force-loading rate plot in **Figure 2.S5** was fitted using the following equation [128]**:**

$$F = \frac{\Delta G^{\ddagger}}{\upsilon \Delta x^{\ddagger}}[1 - (\frac{1}{\Delta G^{\ddagger}} \ln \frac{k_0 e^{-\beta \Delta G^{\ddagger}}}{\Delta x^{\ddagger} l})^{\upsilon}]$$

where k_0 is the intrinsic off rate in the absence of force, $\Delta x^{\ddagger}$ is the distance to the energy barrier, $\Delta G^{\ddagger}$ is the height of the energy barrier in the absence of force, $\beta^{-1}=k_B T$, and $\upsilon = 0.5$, assuming the shape of the free-energy surface is cusp.

Chapter 3 Mapping mechanostable pulling geometries of a therapeutic anticalin/CTLA-4 protein complex

This chapter was published in:

Zhaowei Liu, Rodrigo A. Moreira, Ana Dujmović, Haipei Liu, Byeongseon Yang, Adolfo B. Poma, Michael A. Nash. Mapping Mechanostable Pulling Geometries of a Therapeutic Anticalin/CTLA-4 Protein Complex *Nano Letters* 10.1021/acs.nanolett.1c03584 (2021)

We used single-molecule AFM force spectroscopy (AFM-SMFS) to screen residues along the backbone of a non-antibody protein binding scaffold (lipocalin/anticalin), and determine the optimal anchor point that maximizes binding strength of the interaction with its target (CTLA-4). By incorporating non-canonical amino acids into anticalin, and using click chemistry to attach an Fgβ peptide at internal sequence positions, we were able to mechanically dissociate anticalin from CTLA-4 by pulling from eight different anchoring residues using an AFM cantilever tip. We found that pulling on the anticalin from residue 60 or 87 resulted in significantly higher rupture forces and a decrease in koff by 2-3 orders of magnitude over a force range of 50-200 pN. Five of the six internal anchor points gave rise to complexes that were significantly more stable than N- or C-terminal anchor points, rupturing at up to 250 pN at loading rates of 0.1-10 nN sec^{-1}. Anisotropic network modelling and molecular dynamics simulations using Gō-MARTINI helped to explain the geometric dependency of mechanostability. These results demonstrate that optimization of attachment residue position on therapeutic binding scaffolds can provide large improvements in binding strength, allowing for mechanical affinity maturation without requiring mutation of binding interface residues.

3.1 Introduction

Mechanical anisotropy refers to the variety of mechanical responses that manifest when force is applied to macromolecules from different directions. A classic example is the mechanically induced dissociation of double stranded DNA/RNA double helices and hairpins

[175–177]. Depending on the pulling points, DNA can be unzipped at low forces (~30 pN) where the base paired hydrogen bonds are broken in series, or sheared apart at high forces (>50 pN) where the hydrogen bonds are broken in parallel. Since the sequence is identical in both scenarios, the hybridization energy is equal and the effect is entirely attributable to an anisotropic mechanical response of the double helix. This mechanical property of DNA has been used to build force hierarchies, enabling one-by-one assembly of complex molecular patterns on surfaces using single-molecule cut-and-paste with the atomic force microscope [294].

Folded protein domains similarly exhibit a heterogeneity of mechanical responses depending on the pulling geometry. This was shown for several folded domains including GFP [56], ubiquitin [170], E2lip3 [171] and GB1 [173]. A recent study also reported application of force from an internal sequence position to transmembrane bacteriorhodopsin to dislodge transmembrane helices in a defined order from the membrane and understand intermediate folding states in that system [88].

The force required to dissociate receptor-ligand binding interfaces has also been shown by our group and others to strongly depend on whether the receptor is pulled from the N- or C-terminus, as was shown for cohesin-dockerin and streptavidin-biotin systems [65,70,71]. Differences in shear *vs.* zip geometry have also been reported for protein-based coiled coils [295]. For protein-protein interactions, however, to the best of our knowledge all previous studies were limited to comparing N- and C-terminal anchor points.

Structure-based heuristics are already in place for predicting the mechanical stability of folded domains stretched between their N- and C-termini [132,296–298]. In general, high alpha helix content is associated with low unfolding forces and long unfolding barrier positions, while high beta strand content is associated with protein folds that are generally more mechanically robust with short unfolding barrier positions. Conserved structural motifs referred to as mechanical clamps are also known to impart mechanostability to folded domains [161,299,300]. However, these general trends may change when alternative anchor points are considered. Furthermore, for protein-protein interactions no such heuristics are available. It remains unclear how pulling geometry modulates the binding strength of protein-protein complexes. Single-molecule AFM force spectroscopy (AFM-SMFS) provides a powerful tool to study the force response of protein-protein complexes. State-of-the-art AFM setups are able to precisely control the pulling geometry of receptor-ligand systems and measure a wide range of forces [13,301,302].

For applications in targeted drug delivery using nanoparticles, liposome or engineered viruses, the mechanical force that a receptor molecule can withstand while remaining bound to its target ligand is a potentially valuable optimization parameter. One current trend in biotherapeutics is exploring beyond full length IgG antibodies and into the realm of non-antibody scaffolds. Compared to conventional monoclonal antibodies, non-antibody scaffolds are smaller in size, easier to manufacture and exhibit low immunogenicity [194,303]. Scaffolds such as anticalins, DARPins and adnectins can be diversified at the genetic level and evolved in vitro to bind many different molecular targets [198]. Anticalins are a class of non-antibody scaffolds derived from naturally occurring lipocalins [304]. The anticalins share homologous backbone structures (**Figure 3.S1**), but their ligand binding loops are engineered to specifically bind diverse molecular targets, including proteins, peptides and small molecules [199,305,306]. One of the targets, cytotoxic T-lymphocyte antigen 4 (CTLA-4), is involved in negative regulation of T-cell immune function and maintenance of immune homeostasis [307,308], and represents an important target for cancer immunotherapy.

Here, we used AFM-SMFS to study the response of a non-antibody binding scaffold (lipocalin/anticalin) bound to its target (CTLA-4) and mechanically dissociated under a variety of pulling geometries. We systematically scanned the anchoring residue on anticalin from the N- to the C-terminus, targeting flexible loop regions located between secondary structural elements, and quantified the unbinding energy landscape for each anchor point. Our experimental approach was combined with molecular dynamics (MD) simulations, which showed consistency with the experimental results and helped us gain mechanistic insight into the unbinding pathways of the anticalin:CTLA-4 complex under different loading geometries. What emerges is a clear picture of geometrically optimal anchor points for protein receptor-ligands: pulling from central positions results in highly stable cooperative interactions that break at high forces, while pulling from the termini results in peeling-like behavior where anticalin loses contacts with CTLA-4 and breaks at low forces. These features amount to a molecular unbinding mechanism analogous to a suction cup being centrally pulled or peripherally peeled off a surface.

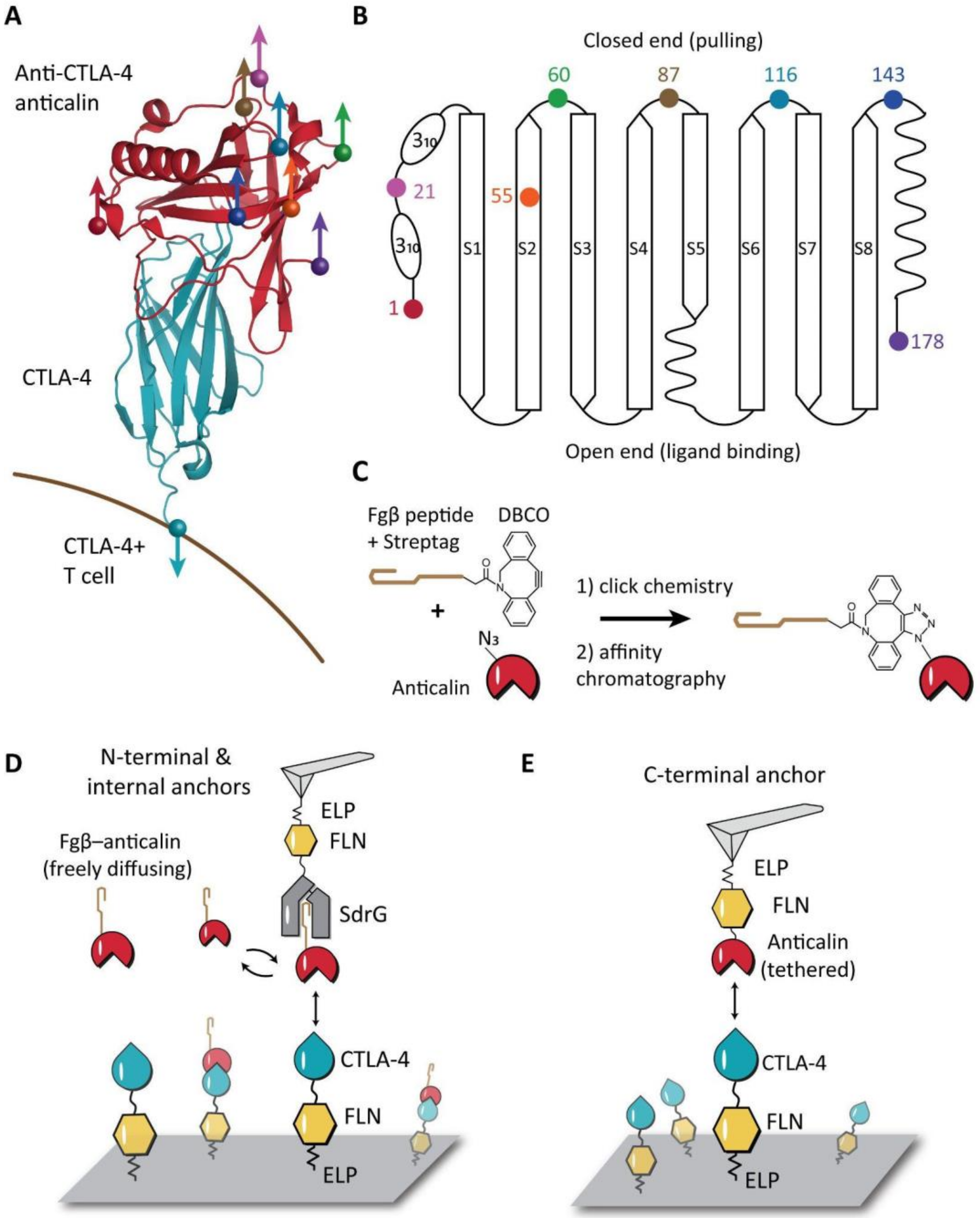


Figure 3.1 Anchor point selection and AFM-SMFS measurement setups. A: Structure of CTLA-4 in complex with anticalin (PDB code 3BX7). Anchor points on the anticalin are shown as colored spheres. The anchor point on CTLA-4 is fixed at the C-terminus, mimicking the natural tethering geometry on the cell surface. **B:** Anticalin has a central β-barrel, consisting of eight anti-parallel β-strands (S1-S8) connected by short flexible linkers. The anchor points shown as colored dots were chosen at the closed end of the β-barrel to avoid interference with the binding interface. **C:** Bioorthogonal conjugation of a fibrinogen β (Fgβ) peptide to the anticalin. The residue at the selected anchor point was replaced by p-azido-phenylalanine using amber suppression to introduce an azide group. The azide was covalently linked with a synthetic peptide comprising Fgβ-StrepTag and a C-

terminal DBCO group using click chemistry. **D:** AFM measurement setup for testing N-terminal and internal anchor points with freely diffusing Fgβ-anticalin. Anticalin conjugated with Fgβ was added to the measurement buffer to final concentration of 1 μM. SdrG-FLN-ELP-ybbr was immobilized on the AFM tip and the ligand (CTLA-4-FLN-ELP-ybbr) was immobilized on the glass surface covalently via ybbr tag. **E:** AFM measurement setup with tethered anticalin for probing C-terminal anticalin anchor point. Anticalin-FLN-ELP-ybbr was immobilized on the cantilever and CTLA4-FLN-ELP-ybbr was immobilized on the glass surface.

3.2 Results

3.2.1 Selection of anchor points, protein expression and AFM measurement setup

The anticalin targeting CTLA-4 was derived from human neutrophil gelatinase-associated lipocalin (NGAL). The structure of anticalin in complex with the extracellular domain of CTLA-4 is shown in **Figure 3.1A** (PDB code 3BX7) [199]. CTLA-4 is a human T cell receptor with the C-terminus of the extracellular domain anchored to the cell surface. As shown in **Figure 3.S1**, the structure of anticalin is highly homologous to other previously reported lipocalin folds [309–313], comprising a β-barrel formed by eight anti-parallel β-strands, flanked by helical regions at the N- and C-termini [314,315]. The two ends of the barrel are the open end, which is involved in ligand binding, and the closed end. Based on the protein structure, we selected eight anchor points on anticalin, scanning the sequence length through eight different pulling geometries. These anchor points were the N- and C- termini (residues 1 and 178), five residues located along flexible linkers connecting the β-strands (residues 21, 60, 87, 116 and 143), and one located within β-strand S2 (residues 55). All of the internal anchor points were located at the closed end of the barrel (**Figure 3.1A and B**). N- and C- terminal anchor points were towards the open end of the barrel closer to the bound CTLA-4 ligand.

Each anchor point on the anticalin molecule corresponds to a precisely defined pulling geometry in the AFM measurements. In order to attach the internal anchor points (residues 21, 55, 60, 87, 116 and 143) to the AFM tip, we combined non-canonical amino acid (NCAA) incorporation, click chemistry and a recently-reported AFM-SMFS setup with freely diffusing receptor molecules [13,70]. As shown in **Figure 3.1C**, the residue at the selected anchor point was replaced by a p-azido-L-phenylalanine using amber suppression to introduce an azide group [275] during protein translation in *E. coli*. A synthetic peptide comprising N-terminus fibrinogen β (Fgβ) peptide, Strep-tag and C-terminus dibenzocyclooctyne (DBCO) group was subsequently conjugated with the azide group on the anticalin using copper-free click chemistry [316]. The reaction product was purified with size-exclusion and Strep-trap columns to remove excess peptide and unreacted anticalin. Successful conjugation of the peptide increased the molecular weight of anticalin by 3 kD, as confirmed by SDS-PAGE and mass spectrometry (**Figure 3.S2**). Each anticalin with Fgβ clicked onto a given residue was expressed and purified, and measured in separate AFM experiments.

The AFM experimental setup with freely diffusing anticalin is shown in **Figure 3.1D**. CTLA-4 was cloned to the N-terminus of a polyprotein, followed by an FLN unfolding fingerprint domain [264], an elastin-like peptide (ELP) elastic linker, and a ybbr tag at the C-terminus. The ybbr tag facilitated site-specific and covalent surface immobilization of the polyprotein [100] to the glass surface. The C-terminus of CTLA-4 was tethered to the glass surface, mimicking the natural pulling geometry on the cell surface. SD-repeat protein G (SdrG) from *S. epidermidis* which binds Fgβ was cloned into the polyprotein SdrG-FLN-ELP-ybbr and immobilized on the AFM tip.

The anticalin with Fgβ peptide conjugated to the selected anchor point was added to the measurement buffer to final concentration of ~1 μM, which saturated immobilized CTLA-4 on the surface. SdrG on the cantilever can bind and unbind with Fgβ-anticalin in solution. A significant portion of the time, SdrG on the cantilever is free and when it approaches the surface, a 3-member complex forms, consisting of cantilever-borne SdrG bound to Fgβ-anticalin, which is itself in turn bound to CTLA-4. Alternatively, if SdrG was occupied by an Fgβ-anticalin molecule when the cantilever approached and indented the surface, mechanical contact can stimulate Fgβ ligand exchange, allowing the same 3-member complex to form. If no such exchange occurs, the AFM trace results in no specific interactions being recorded.

Despite the rapid on/off exchange of SdrG:Fgβ complexes at equilibrium, when placed under tension these complexes are extremely mechanically stable and can withstand forces as high as 2 nN [4]. When retracting the cantilever the significantly weaker anticalin:CTLA-4 complex was therefore the first to break, leaving the Fgβ conjugated anticalin on the cantilever. Since the affinity between SdrG and Fgβ is mediocre (Kd ~400 nM) [69], the anticalin on the cantilever quickly exchanged with the freely diffusing anticalin molecules in the solution, preventing the AFM tip from clogging [70]. Tens of thousands of approach-retract cycles could be performed in this format repeatedly using a range of constant pulling speeds from 100 nm s^{-1} to 800 nm s^{-1} to build up large statistics. In this measurement format, it is important to note that the cantilever and surface molecules are always freshly probed, so the refoldability of the cantilever-borne molecules does not play a role.

For the N-terminal anchor point, anticalin was cloned and expressed with the Fgβ peptide at its N-terminus and measured using the freely diffusing setup. However, since the SdrG:Fgβ complex is itself directionally dependent, it was not possible to use Fgβ as a C-terminal anchor point. To anchor the anticalin from the C-terminus, we used an AFM setup with tethered anticalin, where a polyprotein containing anticalin at the N-terminus (anticalin-FLN-ELP-ybbr) was directly immobilized on the cantilever, as shown in **Figure 3.1E**. The protocol for AFM cantilever approaching, dwelling and retracting was kept the same as for the setup with freely diffusing anticalin.

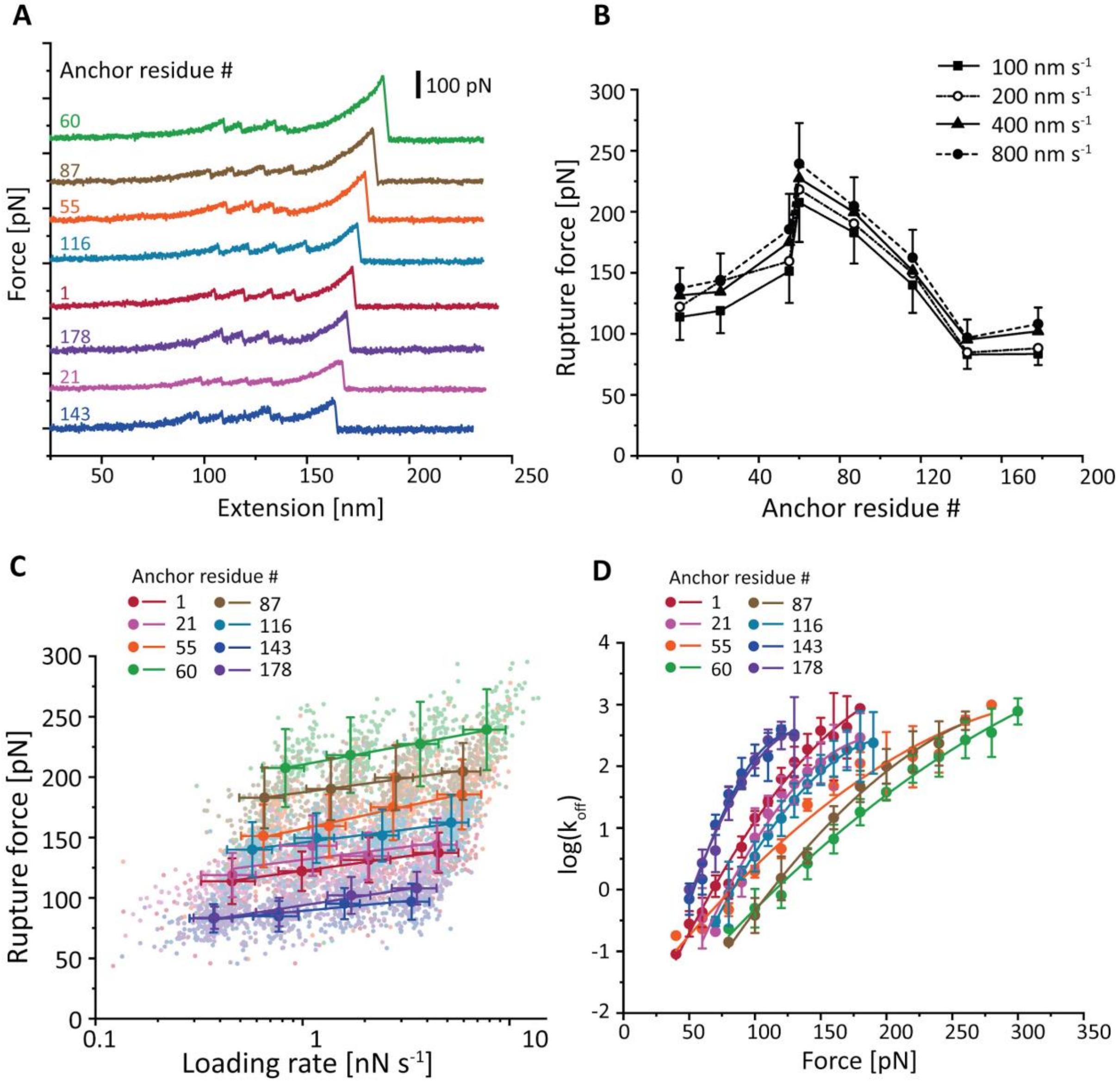


Figure 3.2 Dependency of anticalin:CTLA-4 complex stability on anticalin anchor position. A: Example AFM force-extension traces measured with eight different pulling geometries. Each trace shows unfolding of 2 FLN fingerprint domains and rupture of the anticalin:CTLA-4 complex for a given anchor residue on anticalin. **B:** Most probable rupture forces of the anticalin:CTLA-4 complex at various pulling speeds were plotted against the anchor residue number on anticalin. Error bars represent standard deviation of rupture forces measured at 100 nm s^{-1} (minus) and 800 nm s^{-1} (plus) pulling speeds. **C:** Most probable rupture forces measured at different pulling speeds were plotted against the logarithm of average loading rate and fitted linearly to extract the zero-force off rate k_0 and distance to the transition state $\Delta x^{\ddagger}$. Error bars represent the standard deviation of rupture forces and loading rates. **D:** Force-dependent off rate of anticalin:CTLA-4 complex was plotted against force and fitted using Eq. 6 to extract k_0, $\Delta x^{\ddagger}$ and $\Delta G^{\ddagger}$. Error bars represent the standard deviation of off rates measured at four different pulling speeds.

3.2.2 Different pulling geometries gave rise to diverse unbinding energy profiles

In a typical AFM measurement (~12 h), around 10,000 force-extension curves were recorded and transformed into contour length space using a freely rotating chain (FRC) elasticity model [143]. The curves were filtered based on the two-step unfolding pattern and 32 nm contour length increment of two FLN fingerprint domains [264]. In both AFM setups with freely diffusing or tethered anticalin, one FLN was on the cantilever and another was on the cover glass. We only analyzed the curves containing two FLN unfolding events to rule out nonspecific interactions and multiple parallel interactions between the tip and glass surface. Example force-extension curves of different pulling geometries are shown in **Figure 3.2A and 3.S3**. Intermediate unfolding events were observed in ~9% of selected force curves, including all eight pulling geometries (**Figure 3.S3**). We aligned the contour length histograms of all the selected curves using cross-correlation analysis. The resulting superposition histogram (**Figure 3.S4**) [134,271] showed contour length increments corresponding to the two FLN domains, which together added 64 nm of additional contour length to the system.

Rupture forces of the anticalin:CTLA-4 complex for each Fgβ anchoring residue number were measured in separate AFM data collection runs at four different pulling speeds ranging from 100 to 800 nm s^{-1}. The forces required to dissociate anticalin from CTLA-4 were plotted as histograms, as shown in **Figure 3.S5**. The histograms were fitted to extract the most probable rupture forces, which were plotted against the anchor residue number on anticalin (**Figure 3.2B**). It is clear from the plot that the mechanical stability of the anticalin:CTLA-4 complex is highly dependent on the anticalin anchor residue number. N- and C-terminal pulling points were among the lowest in stability, rupturing at 100-125 pN. The stability significantly rose for anchor points located near the middle of the protein sequence, for example at residues 60 and 87. Peak stability (~225 pN) was achieved when anticalin was anchored at residue 60, between β-strands S2 and S3.

We used Bell-Evans (BE) [123,317] and Dudko-Hummer-Szabo (DHS) [128,129] models to extract the unbinding energy profile parameters for each pulling geometry. As shown in **Figure 3.2C**, the rupture forces measured at each pulling speed were plotted against the logarithm of loading rate, and the most probable rupture forces were linearly fitted against the average loading rate at a given pulling speed to extract the zero force off-rate (k_0) and the distance to the energy barrier ($\Delta x^‡$) using the BE model. We next used the DHS model (**Figure 3.2D**) to transform the rupture force histograms into force-dependent off-rates (Eq. 4), and fitted the resulting plot using Eq. 6 to obtain the k_0, $\Delta x^‡$ and the height of the energy barrier ($\Delta G^‡$). The k_0, $\Delta x^‡$ and $\Delta G^‡$ values obtained using both models are listed in **Table 3.1**.

It is clear from **Figure 3.2** and **Table 3.1** that the complex crossed unbinding energy barriers with significantly different heights and shapes when pulled from different anchor points. Depending on the anchor residue number, the kinetic off-rate k_{off} at a given force can vary by 2-3 orders of magnitude (**Figure 3.2D**, 100 pN). Both DHS and BE models extracted short $\Delta x^‡$ values for anchor residue 60 (DHS: 0.25 ± 0.02 nm; BE 0.30 ± 0.01 nm) indicating a short steep energy barrier when tension was applied through the middle of the beta barrel on anticalin.

3.2.3 Analysis with anisotropic network model

To help interpret our experimental observations, we applied an anisotropic network model (ANM) to calculate the effective spring constant between different pairs of residues [318,319] in the protein complex. The ANM has been previously applied to several isolated protein domains including α-amylase inhibitor, green fluorescence protein (GFP), human ubiquitin and *E. coli* pyruvate dehydrogenase (PDH) E2lip3 domain, and the results were compared with MD simulations and experimental data [172,320]. Here we applied the ANM to the anticalin:CTLA-4 complex and derived effective spring constants between pairs of residues located on separate chains within the crystal structure. One residue was held fixed at the C-terminus of CTLA-4 while the other was scanned through all the residues on anticalin **(Figure 3.S6)**. To compare the predicted force constants from the ANM with the rupture forces and energy landscape parameters, we focused on the eight anchor residues on anticalin which were experimentally measured by AFM-SMFS **(Table 3.S1)**. As shown in **Figure 3.S7**, the correlation between the experimentally measured rupture forces and the effective spring constant is not significant ($p = 0.42$). However, the k_0 parameter derived using the DHS model, along with the $\Delta x^‡$ value derived from the BE model were both statistically correlated ($p < 0.05$) with the effective spring constants between pairs of residues predicted by the ANM. The $\Delta G^‡$ parameter derived from the DHS model was weakly correlated ($p = 0.07$). Both $\Delta G^‡$ (DHS) and

$\Delta x^{\ddagger}$ (BE and DHS) increased with increasing effective force constant, while k_0 (DHS model) decreased with increasing effective force constant. The k_0 calculated using the BE model, however, had negligible correlation with the effective force constant. We note that the unbinding energy profile measured for anticalin anchored from residue 55 was considered an outlier and excluded from the aforementioned correlation analysis. Residue 55 is the only anchor point located in a rigid region (β-strand) of anticalin. All other anchor residues are located at termini or within flexible loops. Due to its comparatively high structural rigidity, position 55 gave rise to the highest effective spring constant predicted by ANM, and resulted in a unique unbinding energy profile with a medium to low energy barrier and short $\Delta x^{\ddagger}$. Excluding residue 55, several of the experimentally determined energy landscape parameters were statistically correlated with the effective spring constant determined by the ANM approach. These findings provide evidence that ANM can be used to predict the anisotropic mechanical properties of protein-protein complex binding interfaces **(Figure 3.S6 and 3.S7)**.

Table 3.1 Unbinding energy landscape parameters of the anticalin:CTLA-4 complex under different pulling geometries.

Anchor point on anticalin	log(k0) (DHS)	log(k0) (BE)	$\Delta x^{\ddagger}$ [nm] (DHS)	$\Delta x^{\ddagger}$ [nm] (BE)	$\Delta G^{\ddagger}$ [k_BT] (DHS)
1	-3.0±0.2	-3.1±0.3	0.50±0.05	0.40±0.03	16.8±0.6
21	-4.2±0.5	-4±3	0.63±0.09	0.4±0.2	18±1
55	-2.0±0.3	-2.4±0.3	0.27±0.04	0.26±0.02	13.9±0.9
60	-2.7±0.2	-4.7±0.2	0.25±0.02	0.30±0.01	17±1
87	-3.7±0.3	-6.0±0.6	0.38±0.03	0.41±0.03	17.8±0.4
116	-4.0±0.1	-5±1	0.57±0.02	0.44±0.07	17.2±0.2
143	-4.4±0.6	-3±1	1.0±0.1	0.6±0.1	18±1
178	-3.8±0.6	-1.6±0.4	0.8±0.1	0.36±0.04	17±1

3.2.4 Utilizing experimental data to improve hybrid all-atom and Gō-MARTINI descriptions of protein binding interfaces

Next we analyzed the anticalin:CTLA-4 system using a structure-based coarse grained model that combined the MARTINI force field [321,322] with a Gō-like description of all native contacts that maintain secondary and tertiary structures in the protein. The latter model has been used in previous studies on mechanical stability of single protein domains [323–325] and protein aggregates [326–328]. Combined Gō-MARTINI now allows analysis of large conformational changes in proteins, and the characterization of protein mechanical properties [329,330]. Gō-MARTINI simulations of anticalin:CTLA-4 dissociation under different pulling geometries at first failed to reproduce the rupture force profile as a function of residue position as obtained by SMFS (**Figure 3.S8**). This indicated an incomplete representation of the energetics at the interface, a known issue in MARTINI force fields [331]. Hence, an additional set of native contacts (NCs) obtained from the crystal structure (PDB code 3BX7) were included in the Gō-MARTINI model to better describe the binding interface between anticalin and CTLA-4 (**Figure 3.S8A**) using Gō contact map determination (http://pomalab.ippt.pan.pl/GoContactMap/). However, this minimal description of contacts was again insufficient to capture the experimental SMFS profiles (**Figure 3.S8B**). To provide qualitative agreement with experiments and remain consistent with a minimal representation of the binding interface dictated by structure, an additional native contact was introduced. This contact was chosen through a systematic process considering contacts established within a cutoff distance equal to 0.9 nm between any pair of Cα-atoms laying at the interface. Then, we performed pulling simulations for each of them to identify the one which improved the rupture profile. Thus, an additional contact between PRO102 (CTLA-4) and VAL66 (anticalin) was included in the contact map (CM). The updated CM improved the energetic description of the protein-protein interface and could be tuned to achieve simulations that reproduced the experimental rupture force profiles (**Figure 3.S8B**). In order to map the energetics of the all-atom MD simulation for the interface, the energy of the additional contact was tuned to reproduce atomistic energy in equilibrium. This step corrects the energy scale of the protein-protein interface via an energy parameter (ε_{core}). Note that the standard native contact energy ($\varepsilon_{Gō\text{-}MARTINI}$) is equal to 9.414 kJ mol^{-1} in the Gō-MARTINI model. The ε_{core} parameter was incrementally increased, and the system was evaluated via pulling simulations at different residue positions until the rupture force profile reached a qualitative agreement with the experiments (SMFS data from **Figure 3.2B**). The corresponding value of ε_{core} that allowed this reconciliation was achieved at $\varepsilon_{core} = 100$ kJ mol^{-1} (see **Figure 3.S8B**). This result was validated in an ensemble (n=100) of different pulling simulations (see **Figure 3.S9**) for each anchor point. Although the optimal ε_{core} value is an order of magnitude larger than $\varepsilon_{Gō\text{-}MARTINI}$, its value is bounded by the typical energy scales we found in all-atom MD simulation. The energetic decomposition for the anticalin:CTLA-4 complex allows us to verify the improvement given by the Gō-MARTINI description. **Table 3.S2** shows the energy values for each protein chain and the protein-protein interface, which was defined by amino acid residues that form contacts between CTLA-4 and anticalin in the native structure. On average the energy at the interface in all-atom MD corresponded to 4% of total energy. The same decomposition for the Gō-MARTINI model assigned an average value of about 6% of the total energy. Note that our model retains the energy ratio for CTLA-4 and anticalin in good agreement with all-atom MD. The stability of the anticalin:CTLA-4 complex is mediated by the stability of the interface. The number of NCs was 189, 275 and 33 NCs for CTLA-4, anticalin and the protein-protein interface during 100 ns MD simulation.

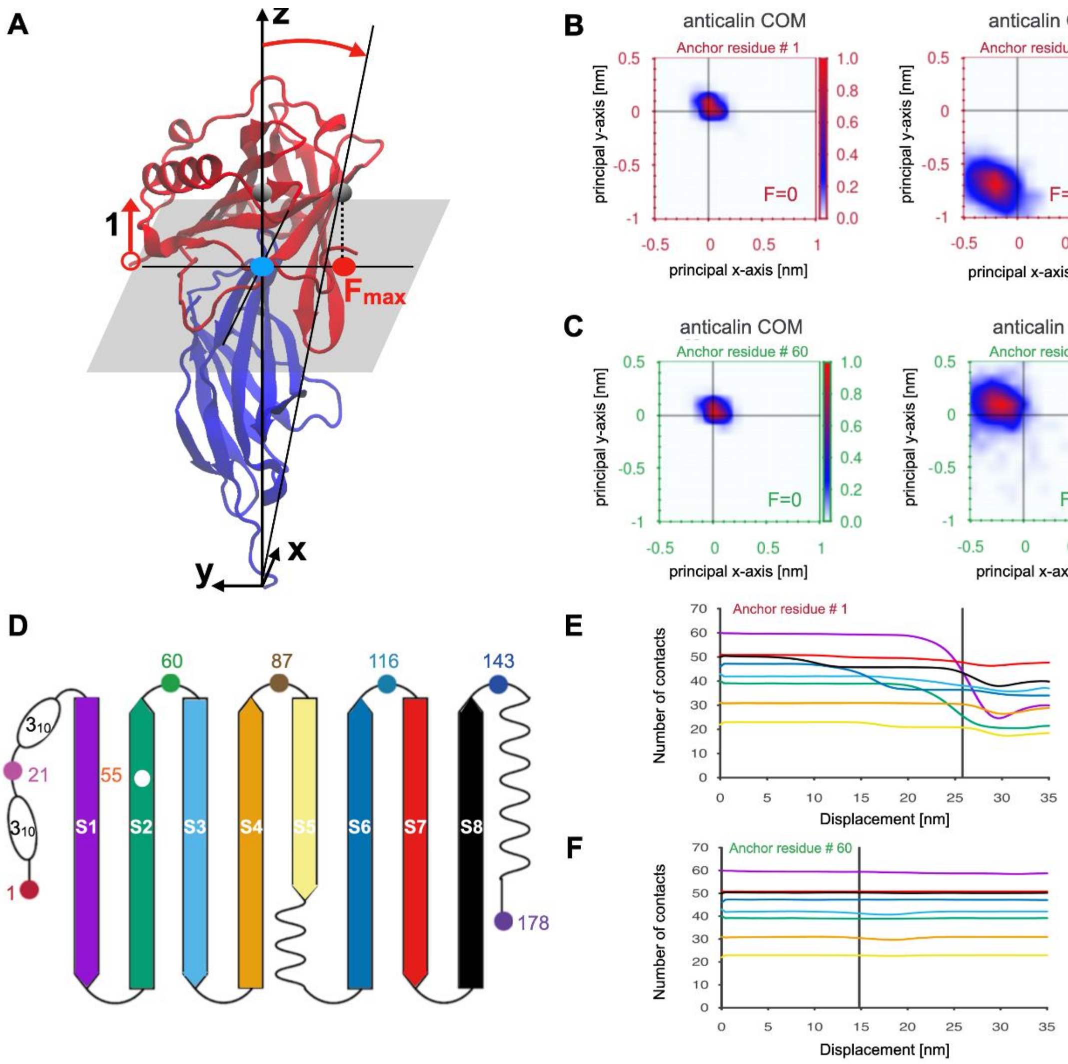


Figure 3.3 Molecular characterization of Gō-MARTINI trajectory for anticalin:CTLA-4 complex at different pulling geometries. A: The translation of anticalin COM for anchor residue 1 under a force applied along the z direction. The CTLA-4 was used as a reference system to define the normal plane. Blue circle denotes the starting anticalin COM position and the red one its translation along -y direction at F_{max}. **B and C**: The relative translation of the anticalin COM with respect to CTLA-4 molecule for two anticalin anchor residues 1 **(B)** and 60 **(C)** at F = 0 and F = F_{max}. Color bars indicate the probability of finding the COM in a given position along the X-Y plane which is perpendicular to the z direction of symmetry of the complex. **D:** The beta-sheet structure of the anticalin and its color representation. **E and F:** The intrachain native contact (NC) evolution for anticalin computed for each beta-sheet during the pulling process. Severe loss of contacts affects the anticalin for the pulling residue 1 **(E)**, whereas almost no loss of NC is reported for anchor residue 60 **(F)**. Color line is in agreement with the panel D.

3.2.5 Calibrated Gō-MARTINI model provides molecular insights into deformation pathways

With the calibrated model established, we next analyzed the trajectories to help explain the observed differences in binding strength as a function of anchor point. The first analysis we carried out was to monitor the position of the center-of-mass (COM) of anticalin during the pulling simulations (see **Figure 3.3A**). **Figure 3.3B and C** compares the COM position of anticalin at zero force and at the maximal force observed in the simulation (F_{max}) for anticalin anchor residues 1 (low stability) and 60 (high stability). For anchor residue 1, at F_{max} the COM of anticalin was clearly shifted in the negative y-direction by ~0.75 nm and in the negative x-direction by ~0.2 nm. However, when pulling anticalin from residue 60, the anticalin COM stayed close to its original position, translating slightly in the positive y-direction (~0.1 nm) and negative x-direction (~0.2 nm). These differences suggest a scenario where pulling from the N-terminus results in a peeling-like behavior of anticalin off of CTLA-4, while pulling from position 60 results in a well-aligned system that cooperatively breaks without translating. Analysis of the xy translation of anticalin COM was carried out for each anchor residue under pulling simulations (**Figure 3.S10**). These plots show that COM translation behavior is distinct for each anchor point. The lowest stability anchor point tested experimentally (residue #143) also shows a broad distribution of translation values for anticalin COMs at F_{max} suggesting significant deformation of the complex and rearrangement under tension.

The second MD analysis we carried out was to analyze the number of NCs (**Figure 3.3E and F and Figure 3.S11**) lost in different regions of anticalin when it was pulled from different directions. When pulling from residue 1, NCs were steadily lost in N-terminal beta strands S1, S2, S3 as well as S6 prior to rupture. However, when pulling from residue 60, few to no intramolecular NCs were lost in the anticalin (**Figure 3.3F**). The anticalin COM shift is more pronounced when several anticalin beta-sheets lose some of the stabilizing NCs (**Figure 3.S11**). Our analysis of the intrachain NCs suggest that breaking NCs in the beta strands makes the anticalin more flexible and its COM samples new positions through different pathways. Furthermore, pathways involving partial unfolding processes and severe loss of NCs were observed for anchor residues 1 and 21 (**Figure 3.S11**). The NCs on the binding interface also behave differently depending on the pulling geometry (**Table 3.S3 and Figure 3.S12**). The interface NCs were lost at different rates with different pulling geometries and the number of remaining interface NCs at F_{max} (immediately prior to rupture) varies across the simulations and shows positive correlation ($p<0.05$) with the rupture force measured both *in vitro* and *in silico* (**Figure 3.S13**). In addition, a few non-native contacts (about five, see **Figure 3.S12**) are established after the rupture. However, the new protein-protein interactions established during the dissociation pathway are not strong enough to maintain the mechanical stability of anticalin:CTLA-4 complex. Based on the simulation analyses, we conclude that the persistence of the original set of interface NCs, the translation of the anticalin COM and the loss of beta strand structure explain the geometric dependency of the mechanical properties of the complex.

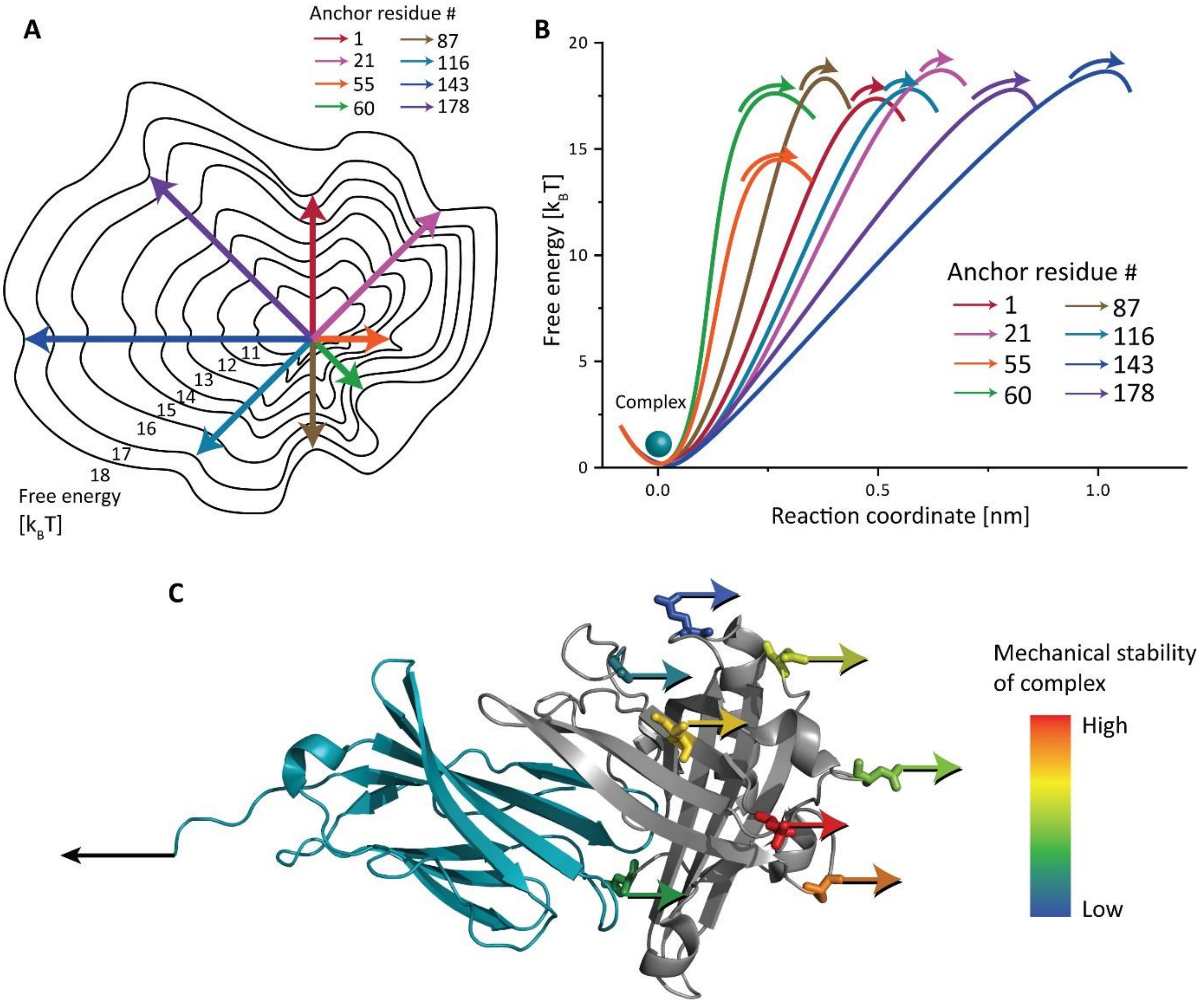


Figure 3.4 Depictions of anticalin:CTLA-4 complex unbinding energy landscape as a function of molecular pulling geometry. A: Energy landscape depiction where anchor point residues are represented as cardinal and ordinal directions of a compass. Under a constrained pulling geometry, the complex is forced to traverse different unbinding pathways across the energy landscape. These different

paths give rise to energy barriers with diverse heights and shapes. **B:** 1D depiction of unbinding energy barrier heights and positions calculated using the DHS model for each pulling geometry (see **Table 3.1**). **C:** Anchor points on the anticalin colored based on mechanical stability of the complex pulled through that position. The most and least mechanostable anchor points on the anticalin are residues 60 (red) and 143 (dark blue), respectively.

3.3 Discussion and outlook

Here, we reported a novel AFM-SMFS experimental method to covalently click a mechanostable peptide handle to internal residues of the non-antibody scaffold anticalin/lipocalin and to measure the rupture forces of anticalin:CTLA-4 complexes using an AFM setup with freely diffusing molecules. Using this method, we observed how the anticalin:CTLA-4 complex responds to external forces applied from different directions and found that the mechanical stability of the complex is highly dependent on the pulling geometry. When pulling from anticalin residue 60, the complex could withstand a high force up to 200 pN, which is ~100% higher than the least mechanically stable pulling geometry (residue 143). To confirm that the different constructs have similar equilibrium affinity, we measured the dissociation constant between CTLA-4 and three anticlain mutants using microscale thermophoresis (MST). As shown in **Table 3.S4**, although the anticalin E60AzF and E143AzF mutants have distinct responses to forces when pulled from different geometries, they have similar affinity towards CTLA-4 at equilibrium. The exception is the I55AzF mutant, which has a slightly lower affinity with CTLA-4. The anticalin mutants were conjugated with a bulky molecule using DBCO-azide reaction in both SMFS (with Fgβ peptide) and MST (with AF647 dye) measurements. Since residue 55 is on a β-strand, which is a part of the anticalin β-barrel structure, the mutation to pAzF and conjugation of peptide or fluorescent dye may have slightly destabilized the anticalin:CTLA-4 complex. Therefore, the I55AzF mutant gave rise to a lower energy barrier in SMFS measurement (see **Table 3.1** and **Figure 3.S7**) as well as lower equilibrium affinity in bulk experiments compared to other anticalin constructs.

Figure 3.4A and B illustrates the complex dissociation energy landscape. In the absence of force, the complex has the same free energy in different pulling geometry scenarios. When external force is applied to the complex from different anchor points, the complex unbinds through different pathways. These pathways correspond to unbinding energy barriers with different shapes and heights, giving rise to distinct responses to forces. It is worth noting that the energy barrier of the most mechanostable pulling geometry (anchor residue 60) is not the highest, but its exceptionally short $\Delta x^{\ddagger}$ contributed to the high resistance to external force. On the contrary, although the pulling geometry with anchor residue 143 has the highest energy barrier, the long $\Delta x^{\ddagger}$ made it the least mechanostable pulling geometry. Another interesting pulling geometry, anchor residue 55, has a unique energy landscape with the lowest energy barrier and a short $\Delta x^{\ddagger}$, giving rise to a mediocre rupture force among all pulling geometries. Therefore, the mechanical stability of the complex is determined by an interplay between the height and the shape of unbinding energy barriers.

We further used the anisotropic network model (ANM) to calculate the force constant between the eight anchor residues on the anticalin and the C-terminal anchor point of CTLA-4. While the effective force constant has very weak correlation with the rupture force, it is correlated with many of the energy profile parameters calculated using both Bell-Evans and Dudko-Hummer-Szabo models. However, residue 55 of anticalin, which is located on a rigid β strand while all the other anticalin anchor points are at the termini or in flexible loops, is an outlier in all of these correlations. This indicates that the AMN approach has certain limitations and additional factors, for example secondary structure, should be accounted for when applying this method.

The computational approach for investigating the nanomechanical stability of the anticalin:CTLA-4 complex under different pulling directions was parametrized by tuning the interface energy. This rendered the Gō-MARTINI approach a very predictable model for the study of large conformational changes of protein complexes at much cheaper computational cost than in regular SMD simulation allowing larger sampling of pathways. This model not only allowed the reproduction of the phenomenological one-to-one experimental SMFS profile for anticalin:CTLA-4 complex, but also unveiled microscopic pathways and mechanistic explanations of the higher mechanical stability observed when pulling anticalin at residue 60. Our computational study explained this stability in terms of translations of anticalin COM, and loss of NCs in anticalin. The simulations show that certain pulling directions can partially unfold the anticalin and destabilize the interface while other positions maintain the close contact of the proteins at the interface.

The anticalin against CTLA-4 shares very high structural similarity with other previously reported engineered anticalins (**Figure 3.S1**). In the future, it will be interesting to see if the mechanically optimal pulling point (residue 60) is generalizable to other anticalin:ligand complexes. Our research demonstrated that the mechanical stability of protein-ligand systems can be tuned by precisely controlling the loading geometry, without changing equilibrium binding properties. This suggests a new paradigm for affinity maturation of non-antibody scaffolds by correctly choosing the anchor points. Such an approach could be particularly beneficial for targeting therapeutic nanoparticles or imaging probes which can exert shear forces onto binding interfaces.

3.4 Follow-up study: anticalin targeting CD98hc

As discussed in section 3.3, it would be interesting to study the mechanically anisotropic behaviors of other anticalins to generalize the conclusions of this study. Another anticalin, namely anticalin P3D11, targets human cell receptor CD98 heavy chain (CD98hc) [310]. Overexpression of CD98 is observed in several cancers [332–334]. Therefore, CD98 is considered as a tumor marker and a potential target for tumor diagnosis and drug delivery. Like the CTLA-4 anticlain, the anticalin targeting CD98hc was also based on the human lipocalin LCN2 [310]. These two anticalins have highly homologous structures and exactly the same number of amino acids (178 amino acids) and share 81% amino acid sequence identity. The structures of anticalins in complex with CTLA-4 and CD98hc are shown in **Figure 3.S14AB** and the sequence alignment between the two anticalins is shown in **Figure 3.S14C**, showing that the non-conservative regions are mostly on the open end of the anticalin β-barrel.

The rupture force of the anticalin:CD98hc complex was measured at four different pulling speeds (100, 200, 400 and 800 nm s^{-1}) using the same AFM setup shown in **Figure 3.1D**, with Fgβ inserted at the N-terminus of anticalin and expressed as a fusion protein. The CD98hc was expressed with a SpyTag at the N-terminus (SpyTag-CD98hc) and covalently conjugated with ybbr-His-ELP-FLN-SpyCatcher protein using the spontaneous isopeptide bond between the SpyTag and SpyCatcher (see **Figure 3.S15A**). The conjugated protein was immobilized on the glass surface via the ybbr tag. This configuration pulls the CD98hc from its N-terminus, precisely mimicking the pulling geometry *in vivo*.

The rupture force between anticalin and CD98hc (pulling from N-terminus of anticalin) was plotted against the loading rate and fitted linearly to extract the distance to the transition state $\Delta x^{\ddagger}$ and zero-force off rate k_0, as described by Eq. 1.5. The rupture force-loading rate plot is shown in Figure 3.S15B. In addition to the BE model, the DHS model was also used to analyze the data. The rupture force histograms (**Figure 3.S15C**) were transformed to force-dependent off rate (Eq. 1.10 and Eq. 1.11) and plotted against the force (**Figure 3.S15D**). Eq. 1.6 was used to fit the off rate-force plot to extract $\Delta x^{\ddagger}$, $\Delta G^{\ddagger}$ and k_0, which are listed in **Table 3.S5**.

Table 3.S5 provides a comparison between the energy profile parameters of two different anticalins in complex with their target, CD98hc and CTLA-4, respectively. The parameters were measured with the same anchor point on the anticalin side (N-terminus) and calculated using both BE and DHS models. While the BE model gives distinct k_0 values between these two anticalin:ligand complexes, the $\Delta x^{\ddagger}$ values given by the BE model and the parameters given by the DHS model are very similar. In addition, the rupture force of these two complexes are also very similar (136.6 pN for anticalin:CD98hc and 113.8 pN for anticalin:CTLA-4 at 100 nm s^{-1}). These results indicate that the mechanical properties of these anticalin:ligand complexes is primarily decided by the anticalin, since the two ligands (CD98hc and CTLA-4) have significantly different structures (see **Figure 3.S14A**) and the binding interface residues are different in these two complexes.

These two anticalins have exactly the same residue number and highly similar structures, making it very easy to design parallel experiments and compare the results. The next steps would be to dissociate anticalin from CD98hc by pulling from other residues on the anticalin, including the same positions as the anticalin:CTLA-4 complex (residue numbers 21, 55, 60, 87, 116, 143 and 178) and other residues to expand the study to other structural elements (helices and other β-strands). The difficulty is that the anticalin against CD98hc has very low expression yield compared to the anticalin targeting CTLA-4. As a result, the expression failed when replacing the anchor residues of CD98hc anticalin with AzF using amber suppression, which further decreased the yield. One approach to enhancing protein expression yield is to express the target protein fused with another well-folded domain (also known as carrier protein) to increase the solubility and expression yield of the target protein. The fusion domain (carrier protein) can be subsequently cleaved using protease, if a protease cleavage tag is added between these two domains. One of the successful carrier proteins is xylanase, which has been used in the expression of dockerins [335]. A xylanase-anticalin construct was designed and cloned (see **Supplementary note 3.2**), but it did not enhance the yield sufficiently to produce anticalin mutants using amber suppression. Other fusion domains could be used in the future to enhance the expression yield, for example maltose-binding protein (MBP) [336].

3.5 Methods

3.5.1 Expression and purification of Fgβ-anticalin and anticalin-ddFLN4-ELP-ybbr

The constructs were cloned in a pET28a vector, which was used to transform NiCo21(DE3) competent cells (New England Biolabs, Ipswich, MA, USA). The cells were grown in terrific broth (TB) medium containing 50 µg mL^{-1} kanamycin at 37 °C until OD reached ~ 0.6. The protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and incubating at 20 °C overnight. The cells were subsequently pelleted and lysed using sonication. The cell lysate was loaded to a His-trap column (GE Healthcare, IL, USA), washed with phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 2 mM KH_2PO_4, pH = 7.4) with 20 mM imidazole and eluted with PBS buffer supplemented with 500 mM imidazole. The eluate was further purified with Superdex Increase 200 10/300 GL size-exclusion column (GE Healthcare).

3.5.2 Expression and purification of CTLA-4-ddFLN4-ELP-ybbr

CTLA-4 contains two disulfide bonds and therefore we used a system to facilitate cytoplasmic disulfide bond formation in E.coli (CyDisCo) to express the CTLA-4-ddFLN4-ELP-ybbr fusion protein. We co-transformed NiCo21(DE3) competent cells with pET28a vector carrying CTLA-4-ddFLN4-ELP-ybbr and pMJS205 vector carrying sulfhydryl oxidase Erv1p and disulfide bond isomerase PDI[337]. The transformed cells were grown in TB medium containing 50 µg mL^{-1} kanamycin and 34 µg mL^{-1} chloramphenicol at 37 °C until OD reached ~ 0.8. The expression of Erv1p, PDI and CTLA-4-ddFLN4-ELP-ybbr was induced by adding 0.5 mM IPTG, followed by incubating at 20 °C overnight. The CTLA-4-ddFLN4-ELP-ybbr protein was purified using the same procedure as the Fgβ-anticalin construct.

3.5.3 Amber suppression

Anticalin constructs with internal anchor points were expressed using an amber suppression system. The wild-type anticalin was cloned to a pET28a vector and the codon encoding the amino acid at the selected anchor point was mutated to an amber codon (TAG). NiCo21(DE3) competent cells were co-transformed using the expression vector and plasmid pEVOL-pAzF (addgene #31186). Cells were grown in 500 mL LB medium supplemented with 50 µg mL^{-1} kanamycin and 25 µg mL^{-1} chloramphenicol at 37 °C until OD reached 0.8, washed with 500 mL M9 minimal medium and transferred to 250 mL M9 medium supplemented with 50 µg mL^{-1} kanamycin, 25 µg mL^{-1} chloramphenicol, 0.2 mg ml^{-1} p-azido-l-phenylalanine (pAzF) and 0.02% arabinose. The culture was incubated at 37 °C for 1 h and was supplemented with 1 mM IPTG, followed by incubation at 20 °C overnight. The cells were lysed and the protein was purified using the same procedure as the wild-type Fgβ-anticalin construct.

3.5.4 Conjugation of Fgβ peptide and anticalin

5x molar excess of Fgβ-StrepTag-DBCO peptide (JPT Peptide Technologies GmbH, Berlin, Germany) was added to AzF-incorporated anticalin. The mixture was incubated at room temperature with shaking for 1 h, followed by incubation at 4 °C overnight. The reaction mixture was purified with size-exclusion column (SEC) to remove the excess peptide. The protein purified with SEC was loaded to a Strep-Trap column (GE Healthcare) and eluted using PBS buffer with 2.5 mM desthiobiotin to remove unreacted anticalin.

3.5.5 Microscale thermophoresis measurements

The anticalin I55AzF, E60AzF and E143AzF mutants were labeled with AF647 fluorescent dye (Jena Bioscience, Jena, Germany) by mixing the protein with 5x molar excess of DBCO-AF647, incubating at room temperature for 1 h, and incubating at 4 °C overnight. The labeled proteins were separated from unreacted free dye molecules using size-exclusion column.

The titration samples were prepared by mixing 10 nM labeled anticalin with a series of unlabeled CTLA-4 with different concentrations ranging from 110 pM to 3.6 µM. The microscale thermophoresis (MST) traces of each titration sample were measured using a Nanotemper Monolith NT.115 (NanoTemper Technologies GmbH, Munich, Germany). The temperature-related fluorescence intensity change was recorded for each sample and plotted against the concentration of CTLA-4 to derive the dissociation constant between anticalin and CTLA-4.

3.5.6 Surface chemistry for AFM measurements

Cover glasses and Biolever mini (Olympus) AFM cantilevers were amino-silanized with (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany). The silanized surfaces were incubated in 10 mg/mL sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC, Thermo Fisher Scientific, Waltham, MA, USA) at room temperature for 30 min, followed by incubation in 200 μM coenzyme A (CoA, Sigma-Aldrich, St. Louis, MO, USA) at room temperature for 2 h. The cover glasses and cantilevers were extensively washed with ddH_2O after each incubation step. The CoA-coated cover classes and cantilevers were incubated in CoA-ybbr reaction mixture, consisting of ~40 μM ybbr-tagged protein, 5 μM Sfp (phosphopantetheinyl transferase) enzyme and 20 mM $MgCl_2$, at room temperature for 2 h and subsequently washed with PBS buffer and stored in PBS until further use.

3.5.7 AFM measurements

AFM-SMFS measurements were performed on a Force Robot AFM (Bruker, Billerica, MA, USA). The contact-free method was used to calibrate the cantilever spring constants (ranging from 0.02 N m^{-1} to 0.14 N m^{-1}) and detector sensitivity. The cantilever and cover glass were submerged in PBS buffer. In the measurements using freely diffusing anticalin, ~ 1 μM Fgβ-conjugated anticalin was added to the measurement buffer. The cantilever was approached to the glass surface, dwelled for 200 ms and retracted at a constant speed ranging from 100 nm s^{-1} to 800 nm s^{-1} (100 nm s^{-1}, 200 nm s^{-1}, 400 nm s^{-1} and 800 nm s^{-1}). The x-y position of the stage was moved by 100 nm after each approach-retraction cycle to probe a new molecule on the surface.

3.5.8 AFM data analysis

The recorded force-extension curves were transformed to contour length using the freely rotating chain (FRC) model. This model assumes that the polymer chain consists of bonds with length b and fixed angle Υ, as described by Eq. 1.17 [143].

The force-extension curves were subsequently filtered based on the 64 nm contour length increment from the unfolding two FLN fingerprint domains. The final rupture forces and force loading rates were extracted from the selected force-extension curves. The rupture forces measured at different pulling speeds were plotted in histograms and the rupture force distribution $p(F)$ was fitted with Bell-Evans model (Eq. 1.4) [317] to extract the most probable rupture force.

The rupture force was plotted against the logarithm of loading rate and the most probable rupture force at each pulling speed was fitted against the average loading rate using a linear model to extract k_0 and $\Delta x^{\ddagger}$, as described by the Bell-Evans model (Eq. 1.5) [123,317].

In addition to the Bell-Evans model, we used the Dudko-Hummer-Szabo model to calculate the energy barrier parameters [128,129]. All rupture force histograms were plotted with equal binwidth $\Delta F = 20$ pN. Each bin of the rupture force histogram was transformed into force-dependent off rate using Eq. 1.10. The calculated force-dependent $k_{off}(F)$ was plotted against force and fitted using Eq. 1.6.

3.5.9 Calculation of effective spring constant using anisotropic network model

The effective spring constants were calculated using the MechStiff function of the ProDy Python package [172,338]. The Python code (see **Supplementary Note 3.3**) was adapted from the mechanical stiffness calculations tutorial of ProDy, available at http://prody.csb.pitt.edu/tutorials/mech_stiff/sm.html. The first five residues of anticalin are missing in the protein structure. Therefore the residue 6 of anticalin was considered as the N-terminus when calculating the force constant between the C- terminus of CTLA-4 and the anticalin residues.

3.5.10 Computational modelling

The crystallographic structure of the anticalin:CTLA-4 complex was taken from the deposited structure (PDB code 3BX7). Several missing residues were reconstructed using MODELLER [261] and UCSF Chimera [339]. In order to obtain a well-defined native contact map for the Gō-MARTINI simulation we employed an all-atom MD description in explicit water of the anticalin:CTLA-4 complex. The system was simulated by the CHARMM36c force field in GROMACS software [340]. The long-range electrostatic interactions were evaluated using the particle mesh Ewald (PME) method [341] with a grid size of less than 0.12 nm. A time step equal to 2 fs was used for integration of the potential function. The system was solvated with TIP3P water molecules and Na+ and Cl- ions were added to achieve the concentration of 150 mM and neutralize the whole system. Pre-equilibration steps were employed by first running an energy minimization step followed by equilibration in NVT ensemble at 300 K and final equilibration step was done in NPT ensemble. The temperature and the pressure of the system were set to 300 K and 1 bar using V-Rescale thermostat[342] and Berendsen barostat coupling methods. The production run was carried out at the same thermodynamics parameters of temperature and pressure using V-Rescale and Parrinello-Rahman algorithms respectively for 100 ns. The molecular trajectory was used to calculate the native contact map. In total we got 1000 maps based on the rCSU+OV contact map determination protocol. Only contacts with high frequencies (i.e. larger than 0.9) were considered. The set of contacts chosen avoid placing contact in flexible loops or other highly flexible regions in each protein chain.

The atomistic system was mapped onto a coarse-grained resolution via the MARTINI approach [321] and the secondary and tertiary structure were retained by the Gō-MARTINI interactions [329] with an effective depth (ε) of Lennard-Jones (LJ) potential equal to 9.414 kJ mol^{-1} and the corresponding σ_{ij} of the LJ potential is determined by the following relationship, $\sigma_{ij} = d/2^{1/6}$, with d being the Cα–Cα distance between a pair of residues that form a native contact. The set of native contacts or Gō-MARTINI interactions were determined through our local server (http://pomalab.ippt.pan.pl/GoContactMap/). From the all-atom simulation in equilibrium we obtained the following set of contacts for each protein region 189, 275 and 33 for CTLA-4, anticalin, and anticalin:CTLA-4 interface, respectively. The coarse-grained model was pre-equilibrated similarly to the all-atom MD model.

3.5.11 Characterization of the mechanical stability and unbinding process of proteins by Gō-MARTINI

The nanomechanics of the protein complexes were investigated through a Gō-MARTINI approach that is based on the combination of the MARTINI model [321] and the Gō-like model [323,325,328]. As a result the Gō-MARTINI model has been used to sample large conformational changes in proteins and their aggregates [329]. The former method offers high computational efficiency, while the latter gives an accurate (as much as all-atom MD) energetic description of the process as long as an accurate native contact map is determined. In order to quantify the differences in mechanical stability for the anticalin:CTLA-4 complex, we performed stretching simulations. The pulling direction was chosen along the end-to-end vector connecting the Cα-atom of the CTLA-4 C-terminus and an anchor point in anticalin that was pulled along this vector. Moreover, additional beads have to be attached to those Cα-atoms with the spring constant being 37.6 kJ mol^{-1} nm^{-2}, which is a reasonable value of the AFM cantilever stiffness in protein stretching studies. The typical time scale in the pulling simulations was on the order of 60 ns with a velocity of 5×10^{7} nm s^{-1}, with rupture of the complex generally occurring at an earlier stage. Although this value is still far from the experimental values of cantilever velocities (100-800 nm s^{-1}), it represents a significant computational improvement compared to all-atom SMD whose typical pulling speed is about an order of magnitude larger. In practice, the Gō-MARTINI model allows characterization of the process occurring in the SMFS experiment at almost the same resolution as conventional all-atom MD.

In AFM-SMFS, multiple proteins or protein complexes can exhibit several peaks in the force-distance profile, which signal the partial unfolding of one protein domain or component. Because of the limitations on time resolution and cantilever sensitivity, intermediate unfolding states are often missed in the experimental datasets. However, in the case of our computational model we can observe these intermediate states with a better resolution and assign to each of them a force peak. The largest of these force peaks, F_{max}, defines the characteristic largest rupture force in the system before dissociation or rupture of the protein-protein interface. Also, we can carry out an analysis of the presence of native contact at the rupture force. A contact will be present if the native distance, *d*, between two Cα atoms is less than 1.5 σ_{ij}.The analysis of the mechanical stability required a set of 100 independent trajectories. To fine-tune our data and reproduce the experimental AFM-SMFS rupture profile (see **Figure 3.2B**), the strength of the contact between residues VAL66 and PRO102 in anticalin and CTLA-4 was increased. The energy scale (ε_{core}) that optimized the agreement with the SMFS experiment was in the range of 100 kJ mol^{-1} and was bounded by protein-protein interface energy. The determination of non-native contacts in Gō-MARTINI simulations was constructed based on the overlaps of the enlarged VdW radii centered on Cα atoms. VdW parameters were taken according to our previous work[323]. If the distance of a pair of Cα atoms is smaller than the sum of their enlarged VdW radius, then we count the pair as a non-native contact. In order to characterize the unbinding process in Gō-MARTINI simulations we devised a reference system using the CTLA-4 as no large fluctuations were observed for that molecule. Then, anticalin was aligned with respect to CTLA-4. This allowed us to calculate the main axis of symmetry along the pulling z-direction and define the plane perpendicular to it in the x-y coordinates. For each snapshot in the molecular trajectory the center-of-mass of the anticalin was monitored and its changes were projected onto this plane (see **Figure 3.S10**).

3.6 Supplementary information

3.6.1 Supplementary figures

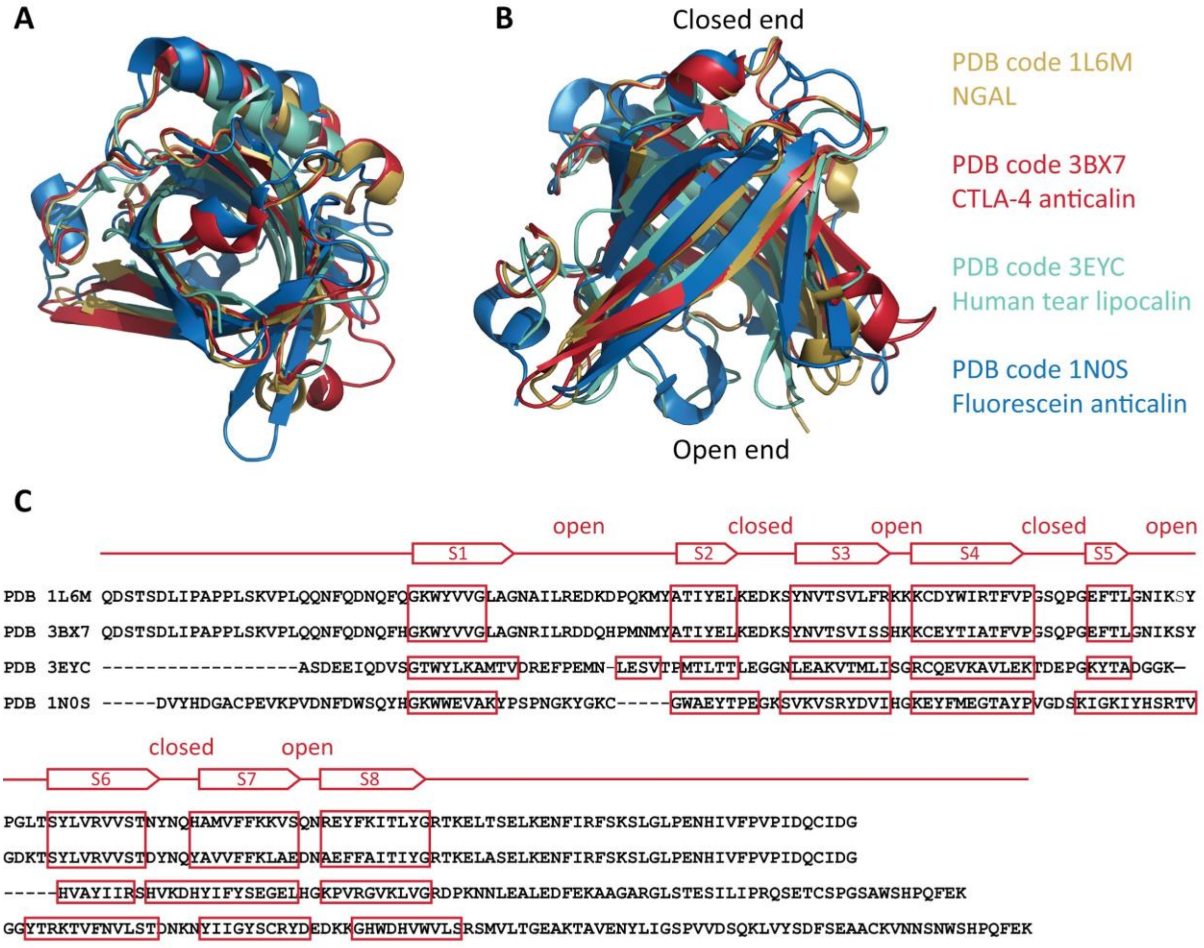


Figure 3.S1 Structure and sequence alignments between anticalin targeting CTLA-4 and other lipocalin folds. A and B: The structure of anticalin targeting human CTLA-4 (PDB 3BX7) was aligned with neutrophil gelatinase-associated lipocalin (NGAL, PDB 1L6M), human tear lipocalin (PDB 3EYC) and an anticalin binding fluorescein (PDB 1N0S). The alignment is shown in top view (panel A, from the closed end) and side view (panel B). Lipocalins share the same β-barrel structure formed by eight β-strands. **C:** Sequence alignment of lipocalins. The anti-(CTLA-4) anticalin was derived from NGAL and has a very high sequence homology (84% identity) with NGAL. The different residues are mainly on the open end, which was engineered to bind CTLA-4. The human tear lipocalin and fluorescein anticalin, which was derived from bilin-binding protein (BBP), have very low sequence homology but high structure homology with CTLA-4 anticalin and NGAL.

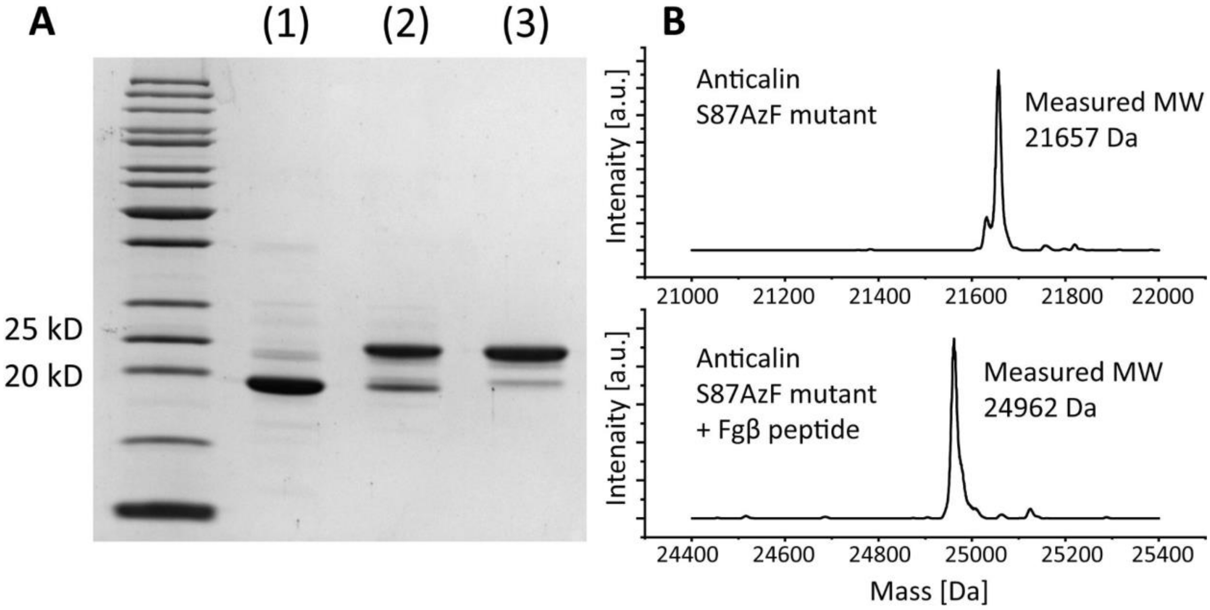


Figure 3.S2 Successful conjugation of Fgβ to anticalin demonstrated by SDS-PAGE and MS. A: SDS-PAGE analysis of anticalin S87AzF mutant protein (lane 1) and the product of the conjugation reaction with Fgβ-StrepTag-DBCO peptide before (lane 2) and after (lane 3) Strep-Trap column purification. Successful conjugation of the peptide increased the protein molecular weight by ~3 kD. Unreacted S87AzF protein was mostly removed by the Strep-Trap column. **B:** Mass spectrometry measurements on anticalin S87AzF mutant before and after conjugation with Fgβ-StrepTag-DBCO. The theoretical molecular weight of S87AzF mutant is 21,659 Da. After conjugation with the Fgβ peptide, the molecular weight increased by 3,305 Da, exactly matching the molecular weight of the synthetic Fgβ-StrepTag-DBCO peptide.

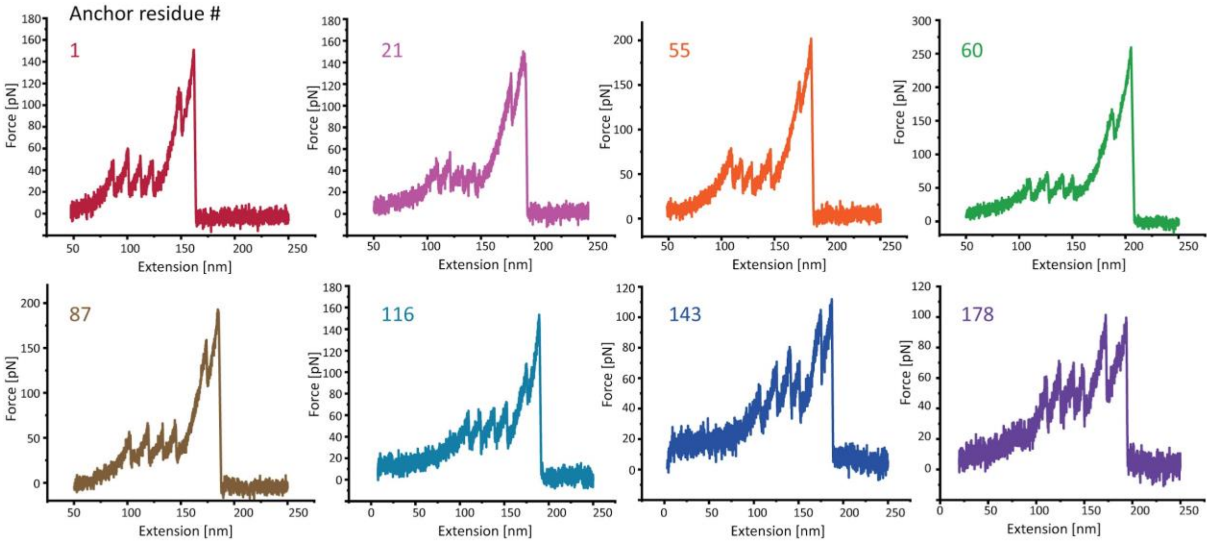


Figure 3.S3 Example force-extension curves with intermediate unfolding steps. Around 9% of the observed curves have intermediate unfolding events prior to complex rupture. Intermediate unfolding events were observed in all the eight pulling geometries suggesting the unfolding is attributable to partial unfolding of CTLA-4.

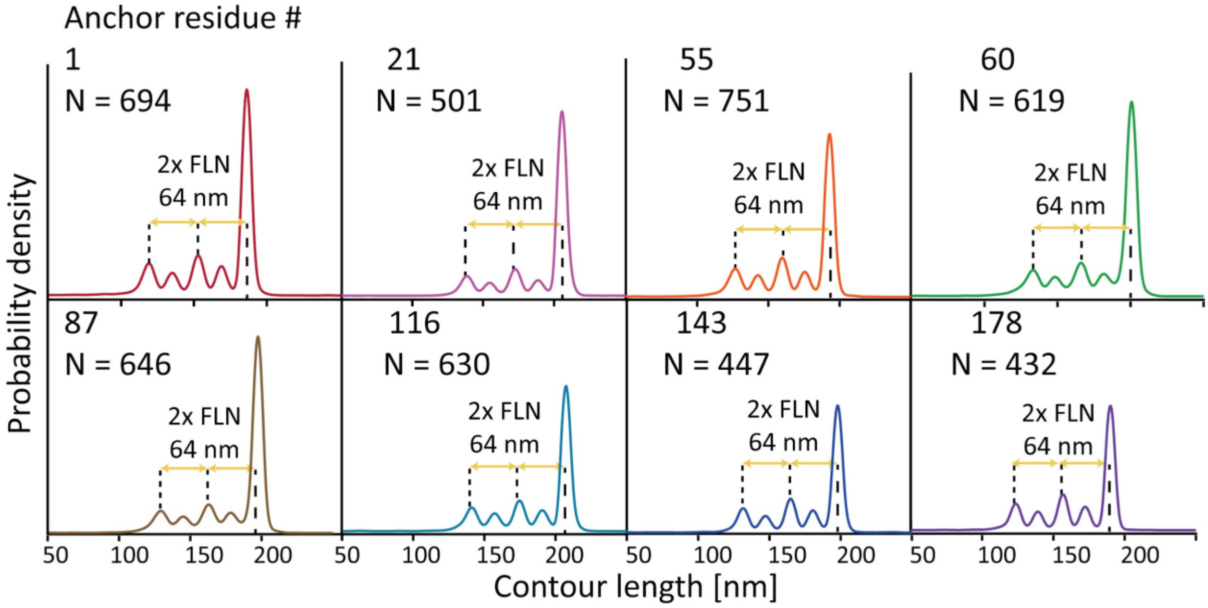


Figure 3.S4 Combined contour length histograms measured at different pulling geometries. All histograms show the unfolding of two FLN fingerprint domains, giving rise to 64 nm in total contour length prior to dissociation of the anticalin:CTLA-4 complex.

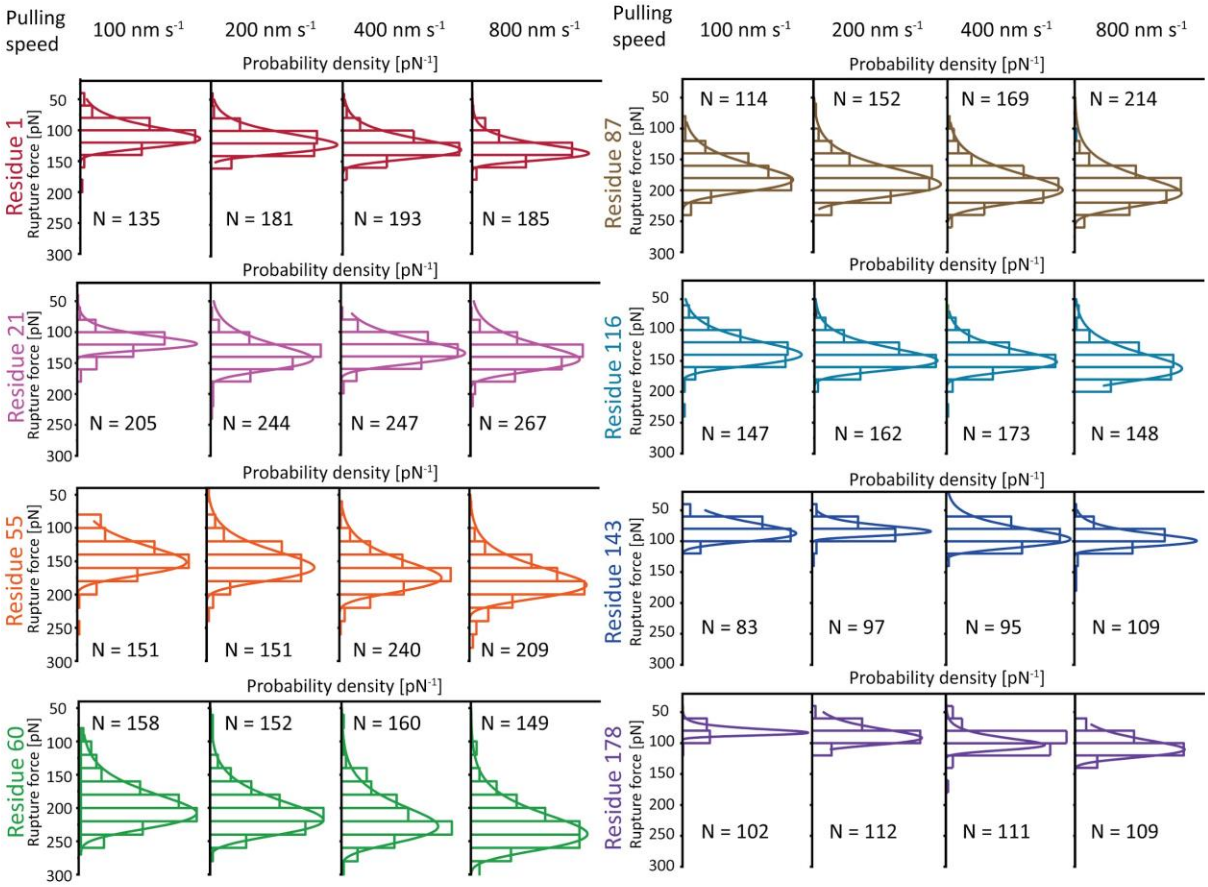


Figure 3.S5 Rupture force histograms of the anticalin:CTLA-4 complex under different pulling geometries. The histograms were fitted with the closed form expression of the Bell-Evans model to extract the most probable rupture force.

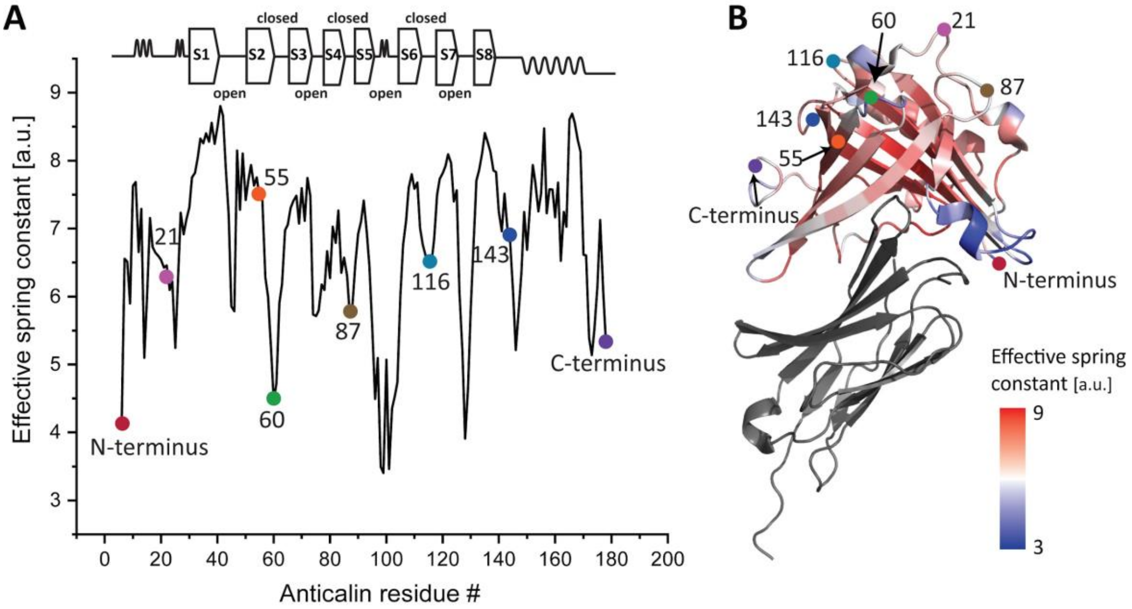


Figure 3.S6 Effective force constant between the C-terminus of CTLA-4 and all anticalin residues. **A:** The effective force constant is plotted against the anticalin residue number and mapped to the secondary structure. Rigid regions (β-strands and helices) have higher force constants and the flexible regions have lower force constants. **B:** Heat map of the CTLA-4:anticalin structure (PDB 3BX7) showing the effective force constant between anticalin residues and the CTLA-4 C-terminus. The eight anchor residues on anticalin are highlighted.

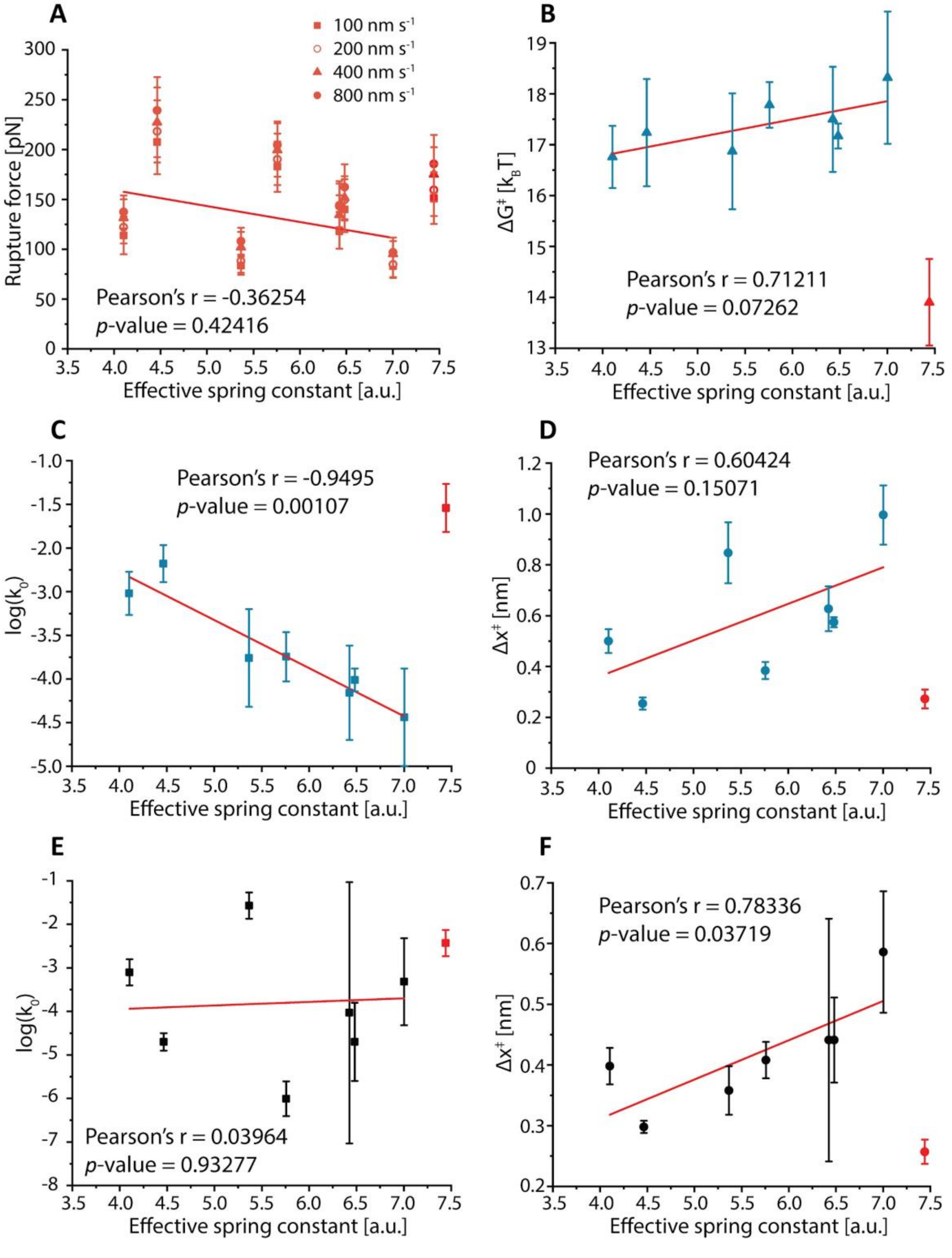


Figure 3.S7 Correlations between rupture force or unbinding energy profile and effective spring constant between anchor points. A: the rupture force does not significantly correlate with the effective spring constant. **B-D:** the $\Delta G^{\ddagger}$ (**B**) and $\Delta x^{\ddagger}$ (**D**) calculated using DHS model are positively correlated with the effective spring constant while the k_0 (**C**) is negatively correlated with the effective spring constant. **E and F:** the k_0 calculated using BE model (**E**) does not have significant correlation with effective spring constant while the $\Delta x^{\ddagger}$ (**F**) is positively correlated.

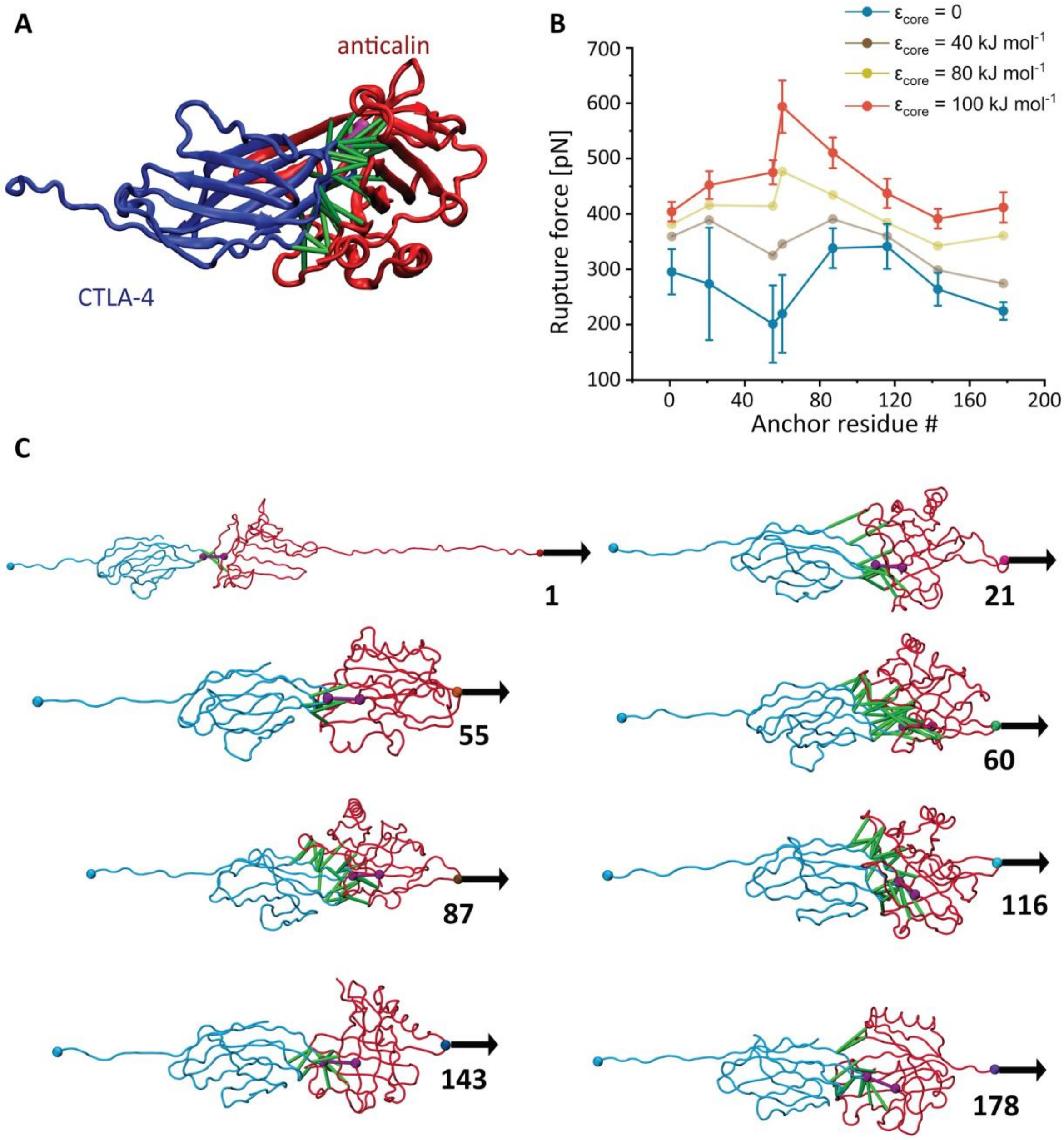


Figure 3.S8 Molecular dynamics (MD) simulations and *in silico* force spectroscopy. A: Crystallographic structure of the anticalin:CTLA-4 complex (PDB code 3BX7) with CTLA-4 shown in blue and anticalin in red. High frequency native contacts ($\varepsilon_{\text{Gō-MARTINI}}$ = 9.414 kJ mol^{-1} and ε_{core} = 100 kJ mol^{-1}) determined in all-atom MD simulation that describe the protein-protein binding interface are shown as green and purple solid lines. **B:** Force-residue profile with ε_{core} equal to 0.0 (blue), 40.0 (brown), 80.0 (yellow) and 100.0 (red) kJ mol^{-1} with standard deviation as the error bars shown for ε_{core}= 0 and ε_{core}= 100. **C:** Different conformations of the complex are captured by coarse-grained simulations at the rupture force. The most relevant native contacts with $\varepsilon_{\text{Gō-MARTINI}}$ and ε_{core} that contribute to the stability of the pulling geometry at the interface of the complex are highlighted as green and purple solid lines respectively. Number below the arrow indicates the anchor residue number.

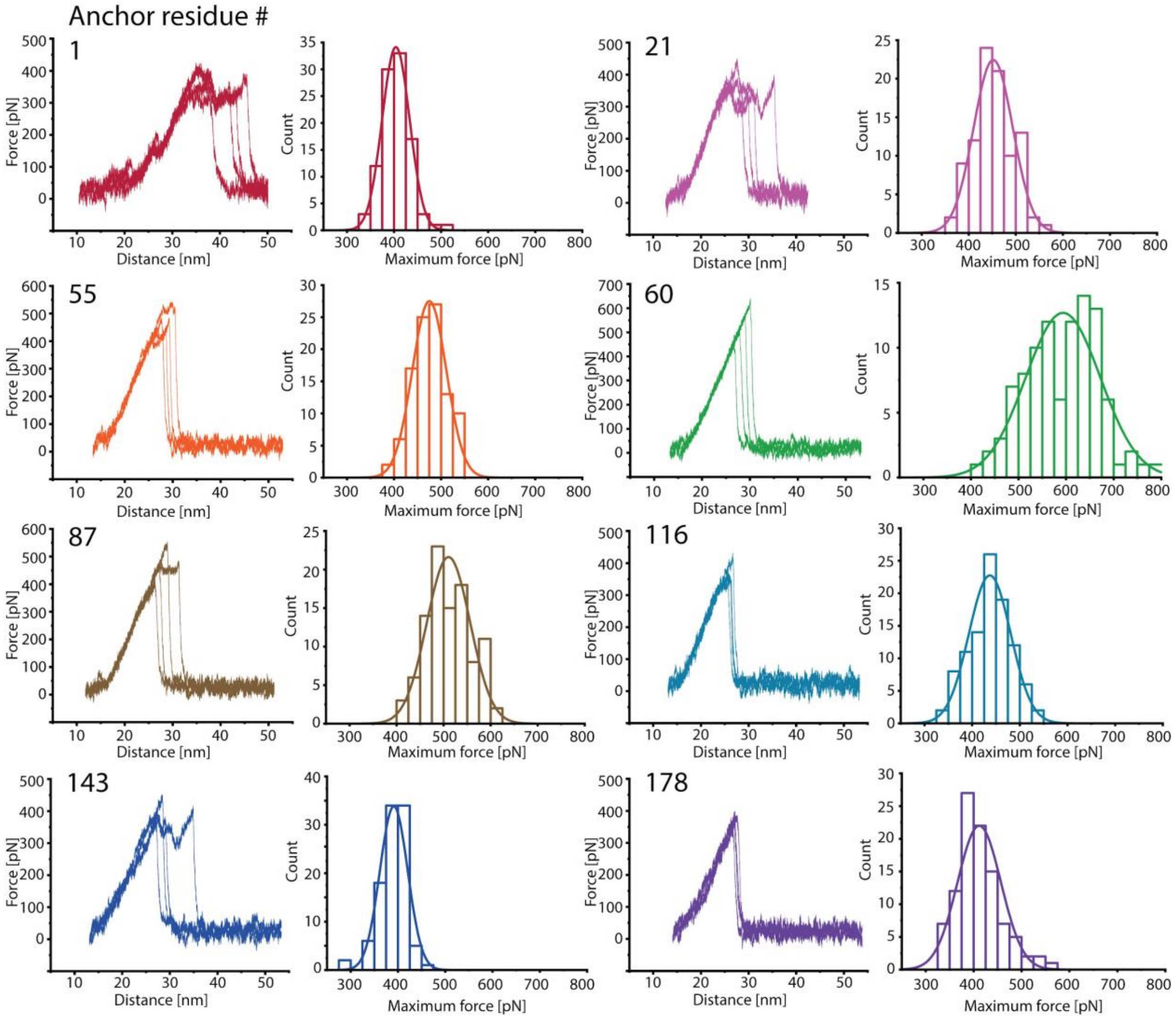


Figure 3.S9 Example force-extension curves and rupture force histograms of anticalin:CTLA-4 complex at different pulling geometries in Gō-MARTINI simulation. The pulling studies were carried out at 5×10^7 nm s^{-1}. Each histogram was obtained based on an ensemble of 100 trajectories, with four of the trajectories shown here as examples.

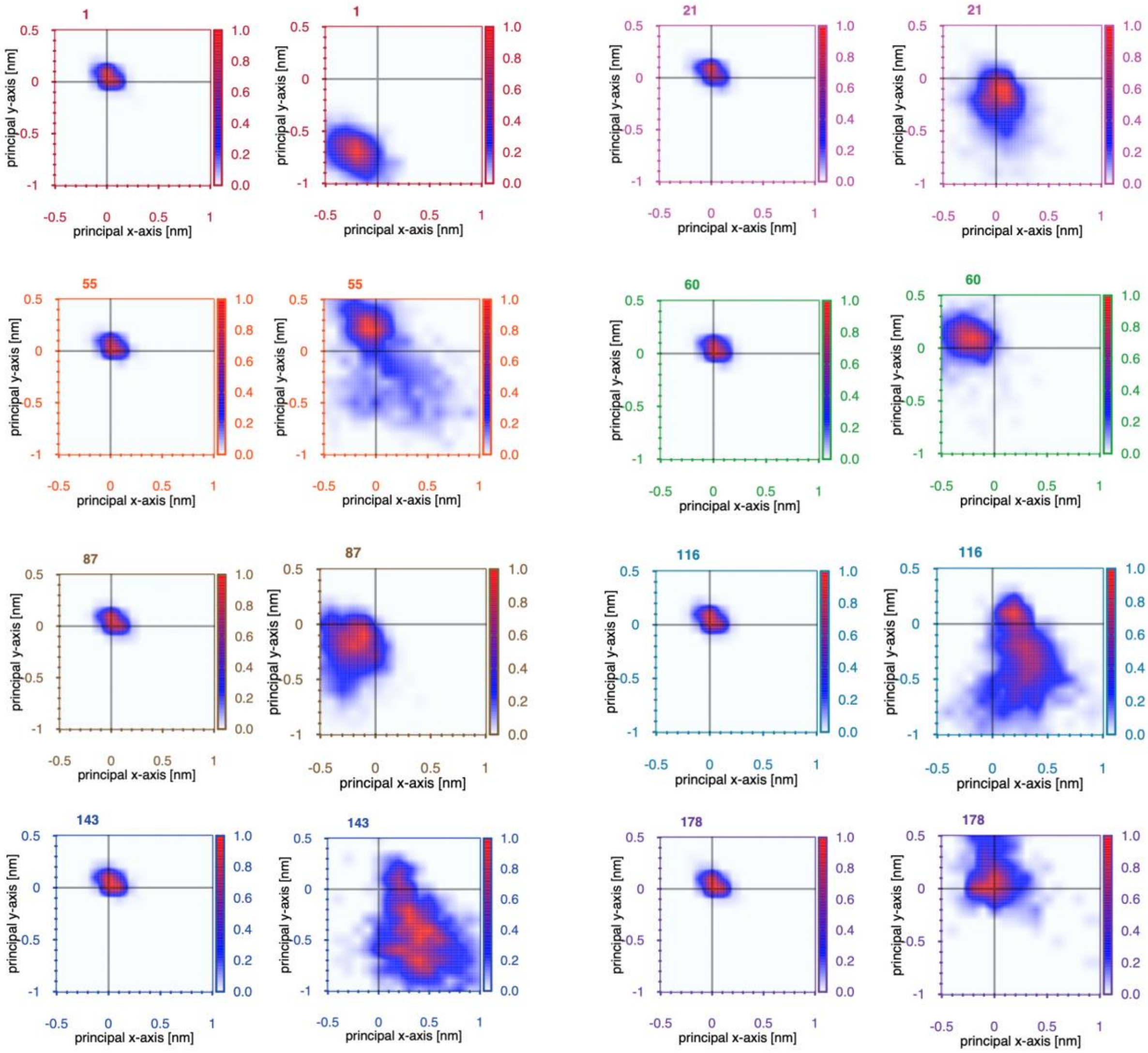


Figure 3.S10 Relative motion of the anticalin COM during Gō-MARTINI stretching simulations of the anticalin:CTLA-4 complex. Data is presented for each pulling geometry at zero applied force (left side) and at F_{max} (right side). Color bars indicate the probability to find the COM at a given position on the X-Y plane which is perpendicular to the z direction of symmetry of the complex. Numbers at the top left corner represent the anchor residue number.

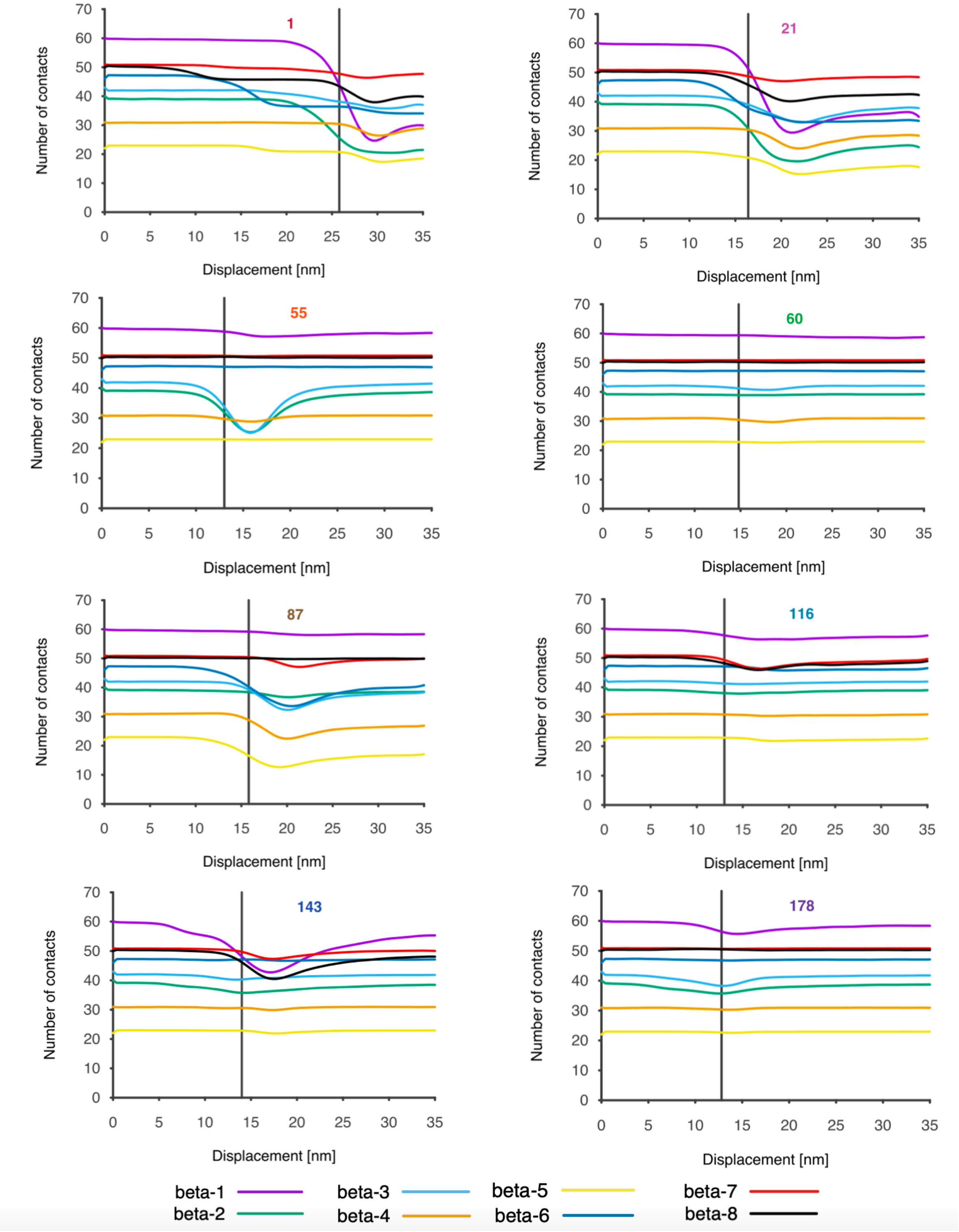


Figure 3.S11 Evolution of the anticalin intrachain native contacts (NC) during Gō-MARTINI stretching simulations of anticalin:CTLA-4 complex. Each color line represents the set of NC for each beta-sheet of the anticalin. Large deviations from the original set involve partial unfolding and a gain in flexibility of the anticalin. Vertical black lines show the position of breaking of the interface contacts at rupture force. Numbers at the top represent the anchor residue number.

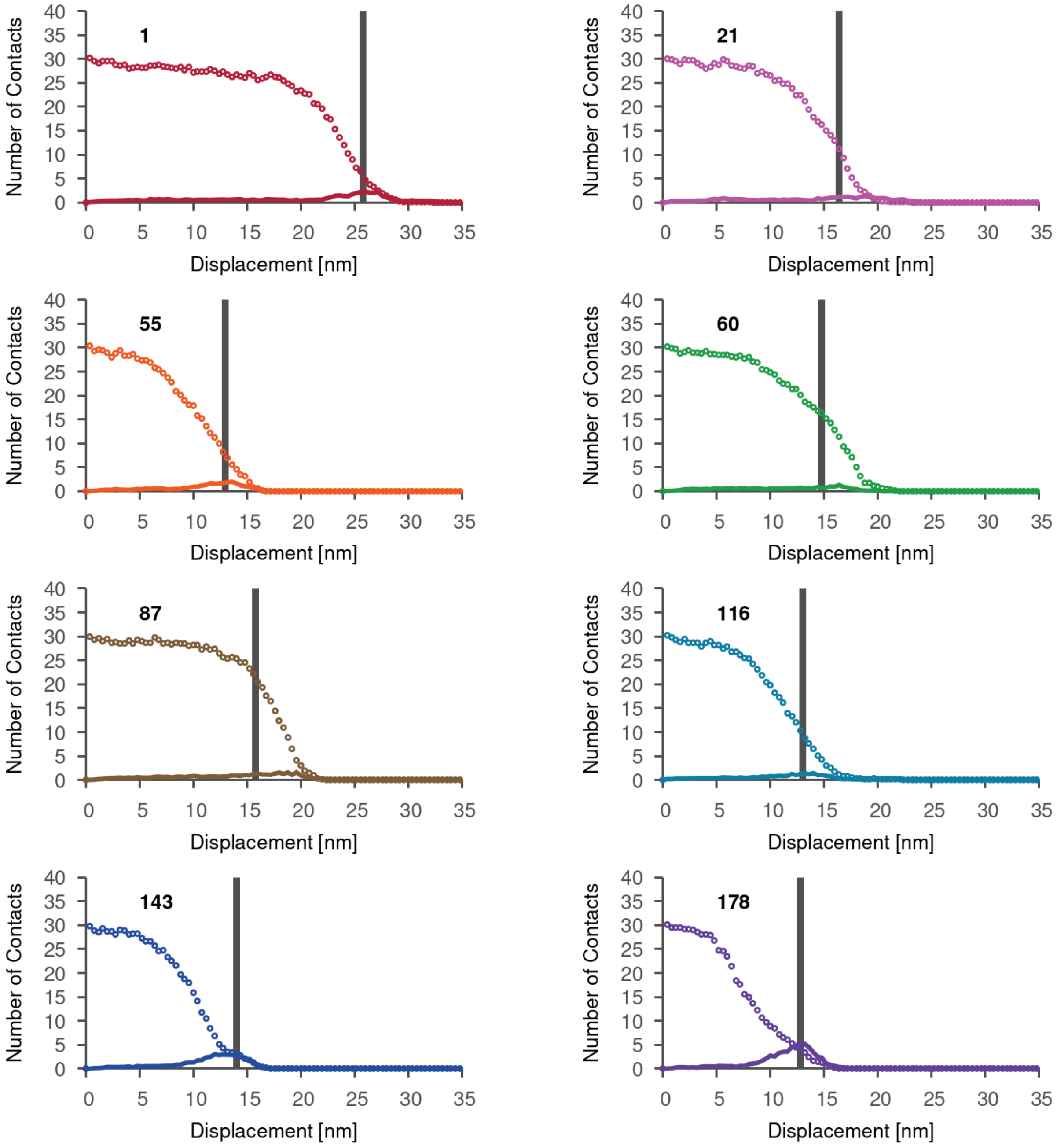
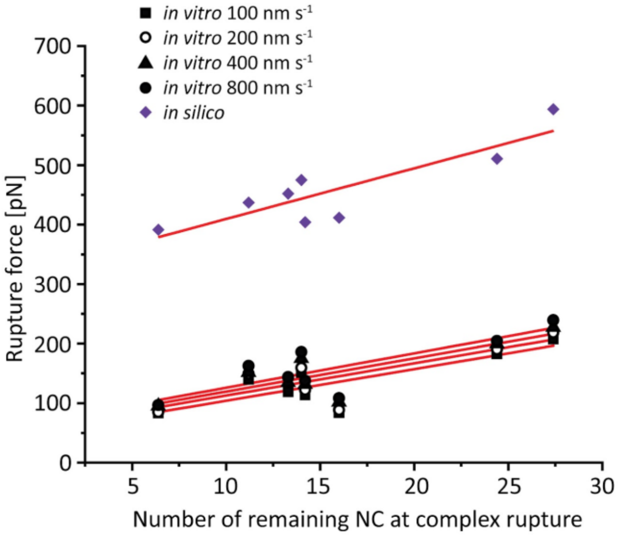


Figure 3.S12 Profile of native (open circles) and non-native (solid line) interface contacts during Gō-MARTINI stretching simulations of anticalin:CTLA-4 complex. The vertical black line shows the position of the native contacts at rupture force. Native contacts and non-native contacts are calculated in the ensemble of pulling trajectories (n=100). Error bars are given by the size of the symbol. Number next to native contact profile represents the anchor residue number.



Correlation	Pearson's r	*p*-value
in vitro 100 nm s^{-1}	0.818	0.013
in vitro 200 nm s^{-1}	0.803	0.016
in vitro 400 nm s^{-1}	0.831	0.011
in vitro 800 nm s^{-1}	0.819	0.013
in silico	0.872	0.005

Figure 3.S13 Rupture forces *vs.* number of remaining NC at complex rupture. The rupture forces measured both *in vitro* and *in silico* are positively correlated ($p<0.05$) with the number of remaining NC at complex rupture.

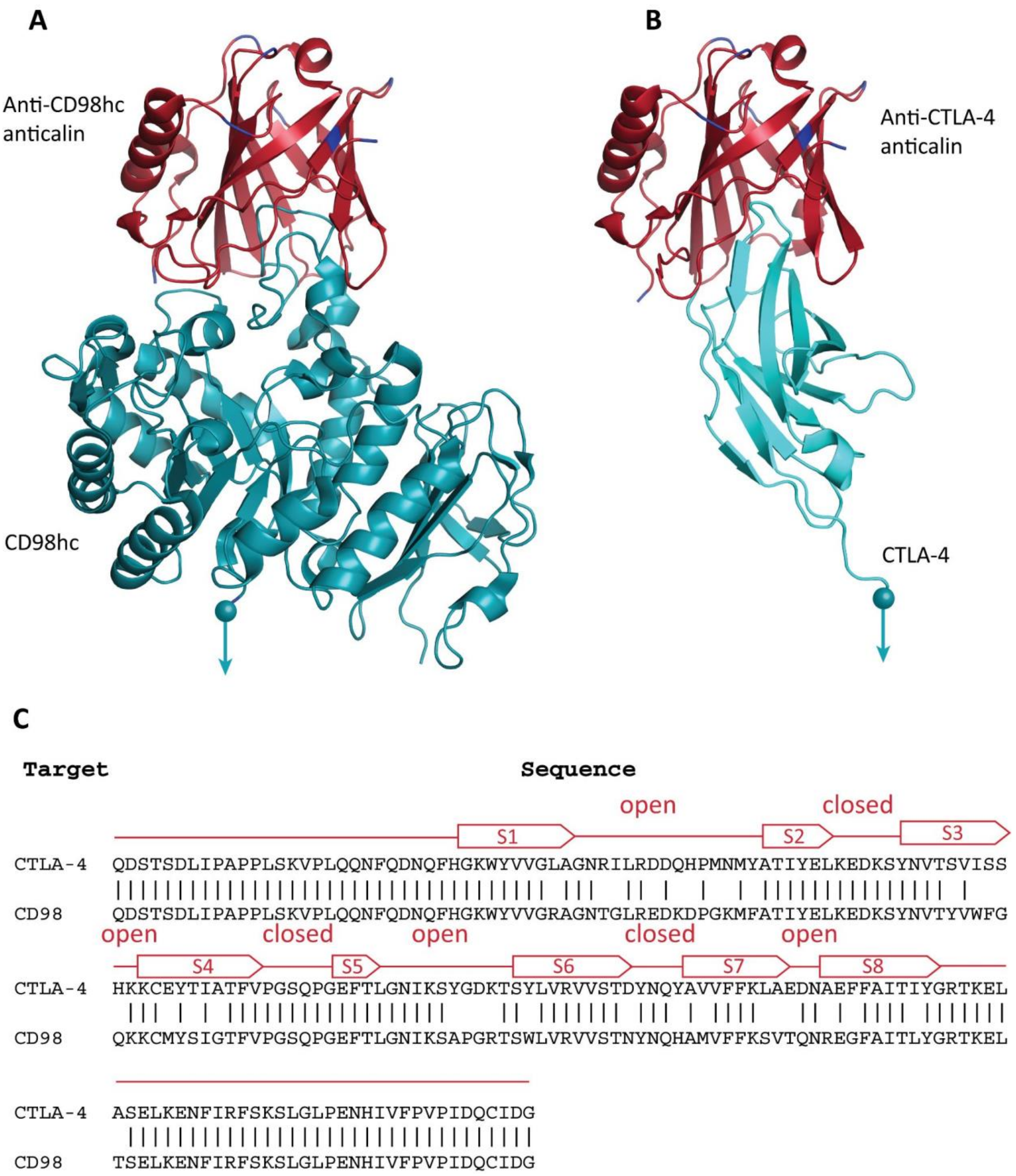


Figure 3.S14 Structure and sequence comparison between anticalins targeting CD98hc and CTLA-4. A: Anticalin (red) in complex with CD98hc (blue) (PDB code 6S8V). B: Anticalin (red) in complex with CTLA-4 (blue) (PDB code 3BX7). **C:** Sequence alignment between the two anticalins. The sequence identity is 81% with the non-conservative regions mostly on the open end of the anticalin structure.

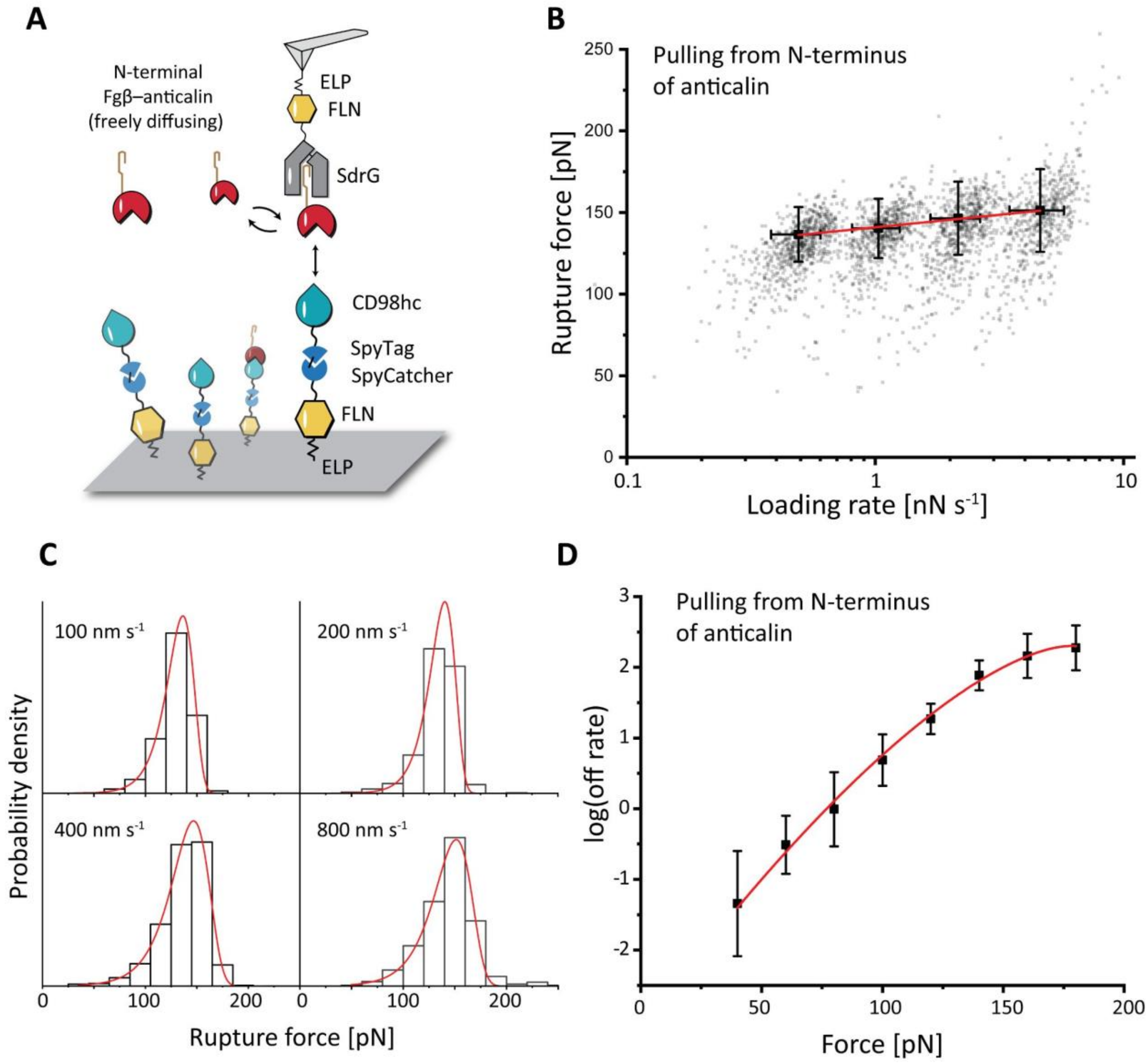


Figure 3.S15 AFM-SMFS measurement of anticalin:CD98hc complex. A: Experimental setup with freely diffusing anticalin expressed with N-terminal Fgβ and CD98hc conjugated with SpyCatcher-ELP-His-ybbr via N-terminal SpyTag. The CD98hc was subsequently immobilized on the glass surface via the ybbr tag. **B:** Rupture force-loading rate plot. The most probable rupture force was linearly fitted against the average of logarithm loading rate measured at four different pulling speeds. Error bars show the standard deviation of rupture force and loading rate. **C:** Rupture force histograms of anticalin:CD98hc measured at four pulling speeds. The histograms were fitted using Bell-Evans model (Eq.1.4). **D:** The rupture force histograms in panel C were transformed to force-dependent off rate using Eq. 1.10 and Eq. 1.11. The off rates were plotted against force and fitted using Eq. 1.6 to extract the energy landscape parameters $\Delta x^{\ddagger}$, $\Delta G^{\ddagger}$ and k_0.

3.6.2 Supplementary tables

Table 3.S1 Effective spring constant between the C terminus of CTLA-4 and different anchor points on anticalin.

Anchor residue #	1	21	55	60	87	116	143	178
Effective spring constant [a.u]	4.10	6.43	7.44	4.46	5.76	6.48	7.00	5.37

Table 3.S2 Non-bonded (VdW and coulomb interactions) energy contribution for each protein chain in the anticalin:CTLA-4 complex and protein-protein interface at different levels of representation. Energy value is given in kJ/mol and next to it in parenthesis as a percentage of the total energy nonbonded energy in the system.

Level of description	**CTLA-4**	**Interface**	**Anticalin**
All-atom MD	-2672.90 (37%)	-259.60 (4%)	-4255.20 (59%)
Gō-MARTINI	-3996.00 (36%)	-629.10 (6%)	-6410.70 (58%)

Table 3.S3 Statistics of the total number of native contacts (NC) present in each protein component of anticalin:CTLA-4 complex at different pulling geometries. Average values of NC and the standard deviations are calculated in the ensemble of 100 pulling simulations at the rupture force.

Anchor point on anticalin	**CTLA-4**	**Interface**	**Anticalin**
1	185.0±2.2	14.2±11.5	200.2±30.1
21	185.0±2.3	13.3±8.5	228.0±25.3
55	184.0±3.1	14.0±7.2	240.2±13.0

60	179.1±6.0	27.4±3.0	265.1±3.4
87	181.0±5.0	24.4±4.4	241.0±18.0
116	184.3±2.4	11.2±7.0	250.1±10.0
143	184.3±2.4	6.4±5.0	231.0±22.0
178	185.0±2.1	16.0±4.0	252.0±5.0

Table 3.S4 Dissociation constants between CTLA-4 and anticalin mutants

Anticalin mutant	I55AzF	E60AzF	E143AzF
Dissociation constant [nM]	140±68	82±50	59±31

Table 3.S5 Comparison of energy landscape parameters between two anticalins in complex with CD98hc and CTLA-4.

Target	k_0 [s^{-1}] (DHS)	k_0 [s^{-1}] (BE)	$\Delta x^{\ddagger}$ [nm] (DHS)	$\Delta x^{\ddagger}$ [nm] (BE)	$\Delta G^{\ddagger}$ [k_BT] (DHS)
CD98hc	$(7.2\pm0.4)\times10^{-4}$	$(1\pm1)\times10^{-7}$	0.45±0.02	0.62±0.05	14.1±0.5
CTLA-4	$(1.0\pm0.5)\times10^{-3}$	$(8\pm5)\times10^{-4}$	0.50±0.05	0.40±0.03	16.8±0.6

3.6.3 Supplementary notes

Supplementary Note 3.1 Amino acid sequences and Addgene accession codes for anticalin:CTLA-4 system

Color code: anticalin, CTLA-4, FLN, ELP, His-tag, ybbr tag, Fgβ, Streptag, SdrG

CTLA4-FLN-ELP-His-ybbr (Addgene accession code 168038)

HVAQPAVVLASSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNELTFLDDSICTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYLGIGNGTQIYVIDPEPGSGSGSGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWRGHHHHHHGSDSLEFIASKLA

Anticalin-FLN-ELP-His-ybbr (Addgene accession code 168039)

QDSTSDLIPAPPLSKVPLQQNFQDNQFHGKWYVVGLAGNRILRDDQHPMNMYATIYELKEDKSYNVTSVISSHKKCEYTIATFVPGSQPGEFTLGNIKSYGDKTSYLVRVVSTDYNQYAVVFFKLAEDNAEFFAITIYGRTKELASELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGSGSGSGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWRGHHHHHHGSDSLEFIASKLA

Fgβ-anticalin-His-ybbr (Addgene accession code 168040)

FFSARGHRPLDGSGSQDSTSDLIPAPPLSKVPLQQNFQDNQFHGKWYVVGLAGNRILRDDQHPMNMYATIYELKEDKSYNVTSVISSHKKCEYTIATFVPGSQPGEFTLGNIKSYGDKTSYLVRVVSTDYNQYAVVFFKLAEDNAEFFAITIYGRTKELASELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGSGSGSHHHHHHGSDSLEFIASKLA

Fgβ-StrepTag-DBCO peptide

NEEGFFSARGHRPLDGSWSHPQFEKGSGSC-DBCO

Anticalin mutant-His (Addgene accession codes 168041-168046)

21 55 60
QDSTSDLIPAPPLSKVPLQQNFQDNQFHGKWYVVGRAGNTGLREDKDPGKMFATIYELKED
87 116
KSYNVTSVISSHKKCEYTIATFVPGSQPGEFTLGNIKSYGDKTSYLVRVVSTDYNQYAVVFFK
143
LAEDNAEFFAITIYGRTKELASELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGSGSGSHHHHH
H

Mutants for internal anchor point measurements (residues 21, 55, 60, 87, 116 or 143) were made by replacing the amino acid at the corresponding position with p-azido-L-phenylalanine (AzF) using amber suppression (see Methods-amber suppression section).

SdrG-FLN-ELP-His-ybbr (Addgene accession code 168047)

EQGSNVNHLIKVTDQSITEGYDDSDGIIKAHDAENLIYDVTFEVDDKVKSGDTMTVNI
DKNTVPSDLTDSFAIPKIKDNSGEIIATGTYDNTNKQITYTFTDYVDKYENIKAHLKLT
SYIDKSKVPNNNTKLDVEYKTALSSVNKTITVEYQKPNENRTANLQSMFTNIDTKNH
TVEQTIYINPLRYSAKETNVNISGNGDEGSTIIDDSTIIKVYKVGDNQNLPDSNRIYDYS
EYEDVTNDDYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNKDDYTTIQQTVTMQTTI
NEYTGEFRTASYDNTIAFSTSSGQGQGDLPPEGSGSGSGSADPEKSYAEGPGLDGGES
FQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEA
GDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSHGVGVPGMGVPGVGVPGVGVPG
VGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVP
GVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGV
PGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWRGHHHHHHGSDSLEFIAS
KLA

Supplementary Note 3.2 Amino acid sequences and Addgene accession codes for anticalin:CD98hc system

Color code: anticalin, CD98hc, FLN, ELP, His-tag, ybbr tag, Fgβ, SpyTag/SpyCatcher, Xylanase

Fgβ-anticalin-His

FFSARGHRPLDGSGSQDSTSDLIPAPPLSKVPLQQNFQDNQFHGKWYVVGRAGNTGLREDKDPGKMFATIYELKEDKSYNVTYVWFGQKKCMYSIGTFVPGSQPGEFTLGNIKSAPGRTSWLVRVVSTNYNQHAMVFFKSVTQNREGFAITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGSGSGSHHHHHH

SpyTag-CD98-His

AHIVMVDAYKPTKGSGSASWSHPQFEKGAELPAQKWWHTGALYRIGDLQAFQGHGAGNLAGLKGRLDYLSSLKVKGLVLGPIHKNQKDDVAQTDLLQIDPNFGSKEDFDSLLQSAKKKSIRVILDLTPNYRGENSWFSTQVDTVATKVKDALEFWLQAGVDGFQVRDIENLKDASSFLAEWQNITKGFSEDRLLIAGTNSSDLQQILSLLESNKDLLLTSSYLSDSGSTGEHTKSLVTQYLNATGNRWCSWSLSQARLLTSFLPAQLLRLYQLMLFTLPGTPVFSYGDEIGLDAAALPGQPMEAPVMLWDESSFPDIPGAVSANMTVKGQSEDPGSLLSLFRRLSDQRSKERSLLHGDFHAFSAGPGLFSYIRHWDQNERFLVVLNFGDVGLSAGLQASDLPASASLPAKADLLLSTQPGREEGSPLELERLKLEPHEGLLLRFPYAAGSGSHHHHHH

ybbr-His-ELP-FLN-SpyCatcher

DSLEFIASKLAHHHHHHWGSGHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWPSGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSGSVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHI

Xylanase-anticalin-His

KNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIV
AENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEG
KPMVNECDPVKREQNKQLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRN
SPWYQIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEG
VPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAI
PKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKGSQDSTSDLIPAPPLSK
VPLQQNFQDNQFHGKWYVVGRAGNTGLREDKDPGKMFATIYELKEDKSYNVTYV
WFGQKKCMYSIGTFVPGSQPGEFTLGNIKSAPGRTSWLVRVVSTNYNQHAMVFFKS
VTQNREGFAITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGSGSGSH
HHHHH

Supplementary Note 3.3 Python code used to calculate the effective force constants

```
#Python version 3.8.2
#Code adapted from http://prody.csb.pitt.edu/tutorials/mech_stiff/sm.html
from prody import *
from pylab import *
ion()
anticalin, header = parsePDB('3bx7', header=True)
calphas = anticalin.ca
anm = ANM('CTLA4_anticalin ANM analysis')
anm.buildHessian(calphas, cutoff=13.0)  #cutoff in Å
anm.calcModes(n_modes='all')
stiffness = calcMechStiff(anm, calphas)
calcStiffnessRange(stiffness)
writeVMDstiffness(stiffness, anticalin, [122], [1,20], filename='Forceconstants')
```

Chapter 4 The catch bond behavior of human gut bacterial adhesion protein is dependent on anchor geometry

This chapter is in preparation for publication

A number of biomolecules exhibit mechanical anisotropy, meaning that they behave differently under external forces applied from different directions. Previous studies showed that the force loading geometry has significant impact on the mechanostabilities of nucleic acids, single protein domains and protein-ligand complexes. Here, we used AFM single-molecule force spectroscopy (AFM-SMFS) in combination of click chemistry to dissociate a high-force bacterial adhesion protein complex found in human gut bacterium *Ruminococcus champanellensis*, dockerin G:cohesin E (Doc:Coh) complex from five different anchor geometries. At the native anchor geometry (C-terminus of Coh and N-terminus of Doc), the complex dissociates in two pathways with different rupture forces. The anchor geometry affects the rupture forces of the two pathways as well as the rate of entering each pathway. When anchored at residue 13 of Coh and the N-terminus of Doc, the complex exhibits a catch bond behavior, which is not present at the native anchor geometry or other non-native anchor geometries.

4.1 Introduction

Catch bonding is a unique behavior found in a limited number of protein-ligand complexes, mostly cell adhesion proteins, including FimH [343], selectin [29], cadherin-catenin-actin complex [152], and cohesin:dockerin complex [168]. In contrast to conventional slip bonds, the bond lifetime of catch bonds increases with external pulling forces in a certain range [149,156]. This counterintuitive behavior allows cells to adhere tightly under high shear stress and dissociate to move freely under low shear stress. Single-molecule force spectroscopy (SMFS) is a powerful tool to study catch bonds and provide mechanistic information. The force clamp mode of SMFS directly measures the lifetime of protein-ligand complexes under different forces, and the measured lifetime of catch bonds typically increases with force when the clamping force is in the catch regime [152,344,345]. When probed using the force ramp or constant speed mode of SMFS, the rupture force of catch bonds exhibits a bimodal distribution and the prevalence of each population is dependent on the loading rate [168,346]. The distribution mainly populates the low rupture force population at low loading rates and moves to the high rupture force population when loading rate increases. It is important to distinguish this behavior from the loading rate dependency of rupture forces in slip bonds, as described by the Bell-Evans model [123,124]. The most probable rupture force of slip bonds increases continuously with loading rate, but the switch from the low force population to the high force population in catch bonds is discrete and takes place on a much broader range of forces.

Anisotropic response to external forces has been shown in other systems including single protein domains [56,170,173] and protein-ligand complexes [65,70,71,347], where the different force loading geometries give rise to new unfolding pathways and distinct mechanical stabilities. However, the concept of mechanical anisotropy has never been explored together with catch bonding. Here, we showed an anchor geometry-dependent catch bond for the first time, where a *Ruminococcus champanellensis* (*Rc.*) dockerin:cohesin (Doc:Coh) complex exhibits catch bond behavior only at a specific anchor geometry, but not the other anchor geometries measured.

Doc:Coh complexes are high-force cell adhesion systems found in the extracellular cellulosomal networks of fiber-digesting bacteria. Certain Doc:Coh complexes are known to exhibit catch bond behavior. Steered molecular dynamics (SMD) showed the binding interface area of a XModule (XMod)-Doc:Coh complex from *Ruminococcus flavefaciens* (*Rf.*) increases under load, suggesting a catch bond mechanism [6]. Another XMod-Doc:Coh complex from *Rc.* exhibits catch bond behavior when probed using constant pulling speed mode [168]. In the latter case, the allosteric regulation of the XMod at the N-terminus of the Doc gives rise to a high force and a low force unbinding pathway, and inhibits the low force pathway under high loading rate, which is crucial to the catch bond behavior. Here, we used atomic force microscope (AFM)-SMFS to dissociate another *Rc.* Doc:Coh complex (DocG:CohE), which does not have an XMod, and showed that the complex dissociates in two pathways with different mechanical stabilities without being regulated by an XMod. A newly developed approach combining AFM-SMFS and click chemistry [347] was used to dissociate the complex at different well-controlled anchor geometries and demonstrated that the rupture force of each pathway, as well as the rate of entering each pathway, is dependent on the anchor geometry.

The complex exhibits a catch bond behavior at one specific anchor geometry, whereas the other anchor geometries, including the native anchor geometry, exhibit slip bond behavior.

The mechanical anisotropy of biomolecules provided a novel approach to molecular engineering. By systematically screening anchor points, the unbinding and unfolding pathways of biomolecules could be altered, giving rise to new mechanical properties such as catch bonding, and distinct mechanical stabilities at different anchor points. This approach potentially has significant impact on the behavior of biomolecules under external forces, while requiring only minimal changes to the amino acid sequences of proteins.

4.2 Results

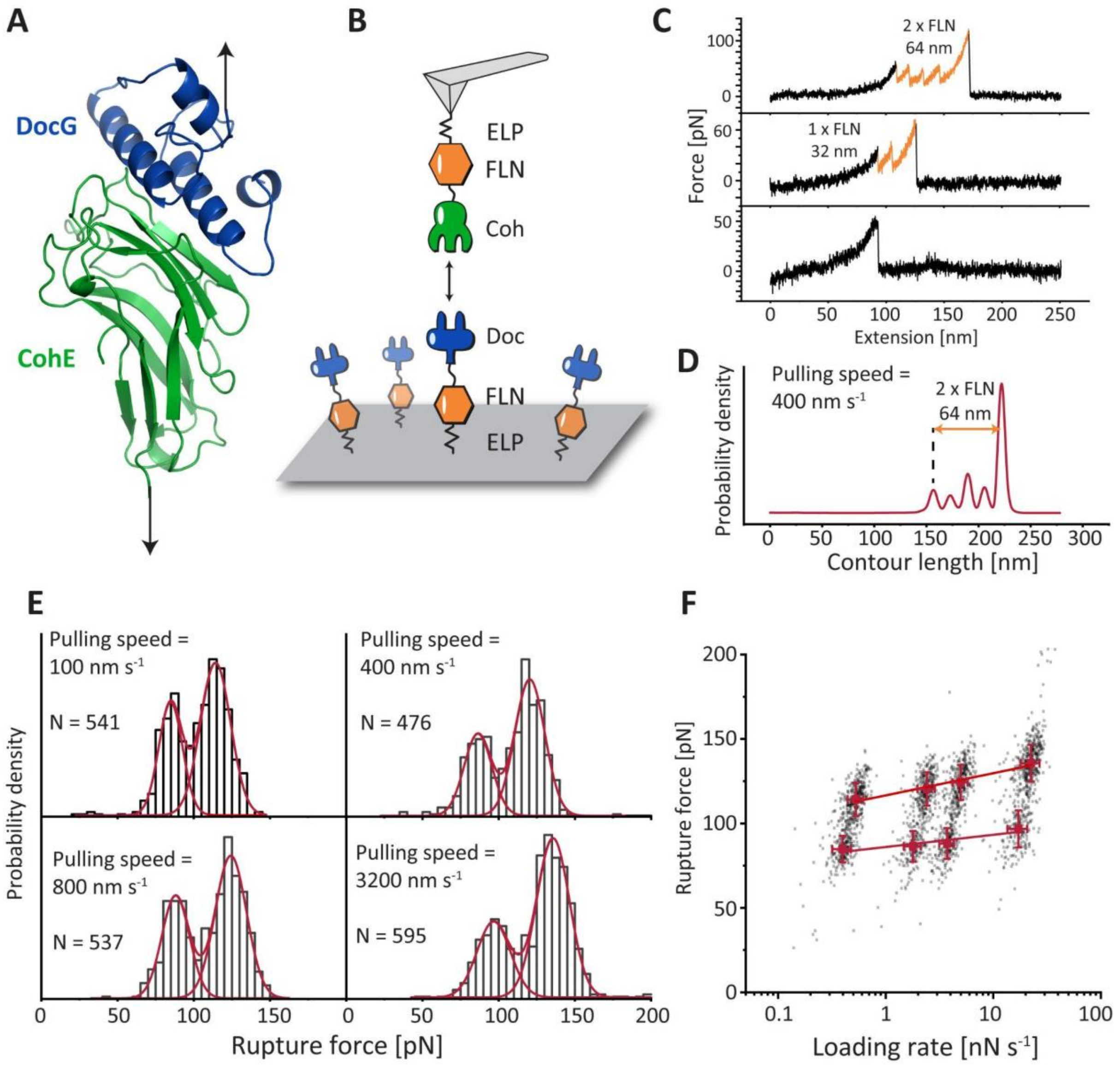


Figure 4.1 Structural model and AFM measurements of Doc:Coh complex at native pulling geometry. A: Structural model of *Rc*. DocG:CohE complex. The DocG model was built with the SWISS-MODEL server. The DocG structural model was aligned with the structural model of *Rc*. XMod-DocB:CohE to build a structure of DocG:CohE complex. **B:** AFM measurement setup for native pulling geometry. The Coh-FLN-ELP-ybbr construct was immobilized on the AFM tip and the ybbr-ELP-FLN-Doc construct was immobilized on the glass surface. AFM tip was approached to the glass surface and retracted to dissociate the Doc:Coh complex at the native pulling geometry. **C:** Example force-extension curves. Most of the curves ramped up to sufficient force to unfold both FLN domains, giving rise to 64 nm contour length increment (top panel). Less than 5% of the curves contain zero or one FLN unfolding events due to the Doc:Coh complex rupturing prior to the unfolding of one or two FLNs. **D:** Combined contour length histogram of all force-extension curves measured at 400 nm s^{-1} pulling speed. **E:** Rupture force histograms of Doc:Coh complex measured at native pulling geometry and four different pulling speeds. Each histogram was fitted with two gaussian peaks to calculate the most probable rupture force of each population. **F:** Force-loading rate plot. The most probable rupture forces of each unbinding pathway (high force: red, low force: blue) were plotted and linearly fitted against the logarithm of loading rate to extract energy landscape parameters (k_0 and $\Delta x^{\ddagger}$). Error bars show the standard deviation of loading rates and forces.

4.2.1 The *Rc.* Doc:Coh complex dissociates in two pathways at native pulling geometry

There is no structural information on the *Rc*. DocG:CohE complex and therefore we used homologous modelling approaches to build structural models. The structural model of DocG was built using SWISS-MODEL server (template: PDB 4WKZ) [348–352], and the DocG structure was aligned with a previously reported *Rc*. XMod-DocB:CohE structural model [168] to build a structural model for the DocG:CohE complex, as shown in **Figure 4.1A**.

The force is loaded to the N-terminus of Doc and the C-terminus of Coh *in vivo*. To mimic the native force loading geometry on AFM, the *Rc*. Doc and Coh were cloned to bacterial expression vectors as fusion proteins: ybbr-ELP-FLN-Doc and Coh-FLN-ELP-ybbr, and expressed in *E. coli*. The elastin-like peptide (ELP) is an intrinsically disordered protein and is used here as an elastic linker to separate the proteins from the surfaces [62]. The filamin (FLN) domain serves as a fingerprint domain to help identify single-molecule interactions [264]. The ybbr tag was used to covalently immobilize the Doc and Coh constructs on the glass surface and AFM tip, respectively [100], as shown in **Figure 4.1B**. The AFM cantilever was approached to the surface and dwelled for 200 ms to form a Doc:Coh complex between the tip and the glass surface. The cantilever was then retracted with a constant speed, ranging between 100 nm s^{-1} and 3200 nm s^{-1}, to apply pulling force to the Doc:Coh complex from the N-terminus of Doc and the C-terminus of Coh, precisely mimicking the native pulling geometry. A force-extension curve was recorded in each approach-retraction cycle. Over 10,000 curves are typically recorded in a measurement (~12 h). These curves were transformed into a contour length space using the freely rotating chain (FRC) elasticity model [143]. Each ELP linker in the Doc and Coh constructs is 170 amino acids long, and the total contour length of two ELP linkers is 170 × 2 × 0.365 nm/amino acid = 124 nm. The force curves are filtered based on the contour length at complex rupture. Only the force curves showing fully stretched ELP linkers, i.e. the contour length at final rupture is longer than 124 nm, are taken into further analysis, to exclude non-specific interactions between the tip and the surface. A majority of the filtered force curves show the unfolding events of two FLNs, as shown in **Figure 4.1C** (top). Each FLN unfolds in two steps and adds ~32 nm of contour length to the system when fully unfolded. Unfolding of two FLNs gave rise to ~64 nm of contour length increment, which is shown in the combined contour length histogram of all filtered curves (**Figure 4.1D**) [134,271]. A small fraction of the force curves (<5%) did not ramp up to sufficient force to unfold one or both of the FLNs (see **Figure 4.1C**, middle and bottom).

The rupture forces of the Doc:Coh complex at native pulling geometry were measured at four different pulling speeds (100, 400, 800 and 3200 nm s^{-1}) and plotted in histograms (**Figure 4.1E**). The histograms show a bimodal distribution, suggesting two different unbinding pathways with different rupture forces. The rupture force histograms were fitted with two gaussian distributions to extract the most probable rupture forces of each pathway. The most probable rupture forces were fitted linearly against the logarithm of loading rate to extract the energy landscape parameters, the zero-force off rate k_0 and the distance to the transition state $\Delta x^{\ddagger}$, as described by Eq. 1.5 [123,124]. The calculated parameters are shown in **Table 4.1**.

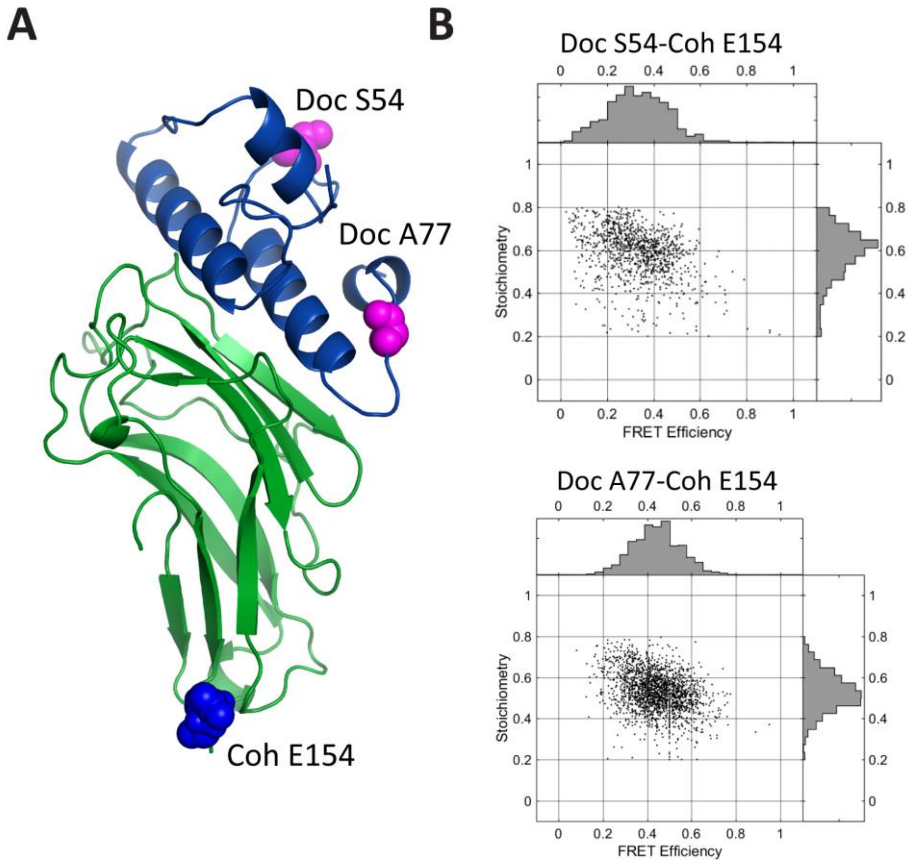


Figure 4.2 Single-molecule FRET shows single binding mode. A: The FRET donor (Cy3b) was conjugated to one of the two labeling sites of Doc (S54 or A77). The FRET acceptor (AF647) was conjugated to the residue E154 of Coh. **B:** smFRET measurements with the donor labeled at residue S54 or A77 of Doc showed a unimodal FRET efficiency distribution, meaning that the complex populates one binding conformation at equilibrium.

4.2.2 Single-molecule FRET demonstrates a single binding mode of Doc:Coh complex

A limited number of Doc:Coh complexes are known to exhibit a dual-binding mode behavior. The complex can be formed in two distinct conformations, where the Doc is rotated by 180° on the binding interface [249], as shown in **Figure 4.S1**. AFM-SMFS has been used to resolve the two binding modes and demonstrated that the different binding conformations of certain Doc:Cohs have different mechanical stabilities [148,168]. Therefore, the Doc:Coh complex has two potential binding conformations, but it is not clear if it populates both of them or only one of them. To test if the two unbinding pathways of the *Rc*. Doc:Coh complex observed here originated from one or two binding modes, we used single-molecule FRET (smFRET) to obtain conformational information on the Doc:Coh system. smFRET has been used to detect dual-binding modes in other Doc:Coh systems [168,353]. Here, we labeled the Coh with the FRET acceptor, AF647, and the Doc with the FRET donor, Cy3b. The Coh was labeled at the C-terminus (residue E154) and the Doc was labeled at the helix 2 (residue S54) and the C-terminus (residue A77), respectively (**Figure 4.2A**). The labeled Coh and Doc were mixed at picomolar concentration and the FRET efficiencies of single Doc:Coh complexes were measured on a confocal microscope. The measured FRET efficiencies were plotted in histograms, as shown in **Figure 4.2B**. The FRET efficiencies measured with FRET donor labeled at S54 and A77 of Doc both showed a unimodal distribution. The smFRET measurements demonstrated that the Doc:Coh complex does not exhibit dual-binding mode behavior, i.e. the complex populates only one of the two potential binding conformations. However, it remains to be seen which binding conformation is populated by the complex. smFRET has the potential of providing structural information of the complex and solving this question. As shown in **Figure 4.S1**, the two putative binding conformations can be discriminated by comparing the distances between the C-terminus of Coh (the FRET acceptor labeling site) and the C-terminus and helix 2 of Doc (the FRET donor labeling sites). However, the smFRET measurement using Doc labeled at S54 gave relatively low yield and broad FRET efficiency distribution due to low sample quality. Repeating the measurement using newly prepared samples would potentially provide more reliable results.

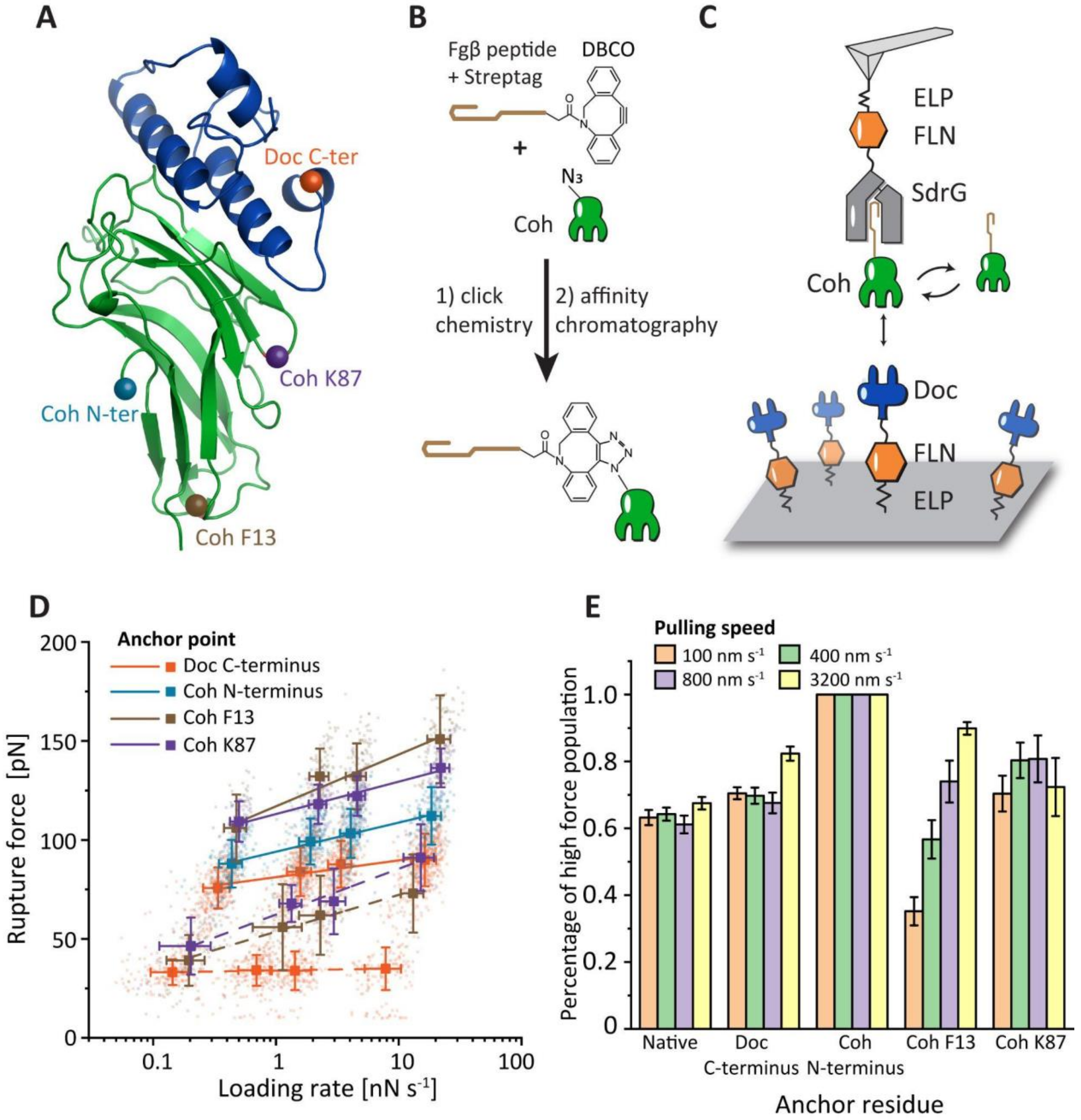


Figure 4.3 Catch bond behavior at non-native pulling geometry. A: Non-native anchor points selected on the Doc (C-terminus) and Coh (N-terminus, residue F13 and residue K87). **B:** For internal anchor points on the Coh, the selected anchor residue was replaced with azidophenylalanine using amber suppression. Fgβ was covalently conjugated with Coh at the anchor point via the click reaction between the azide and DBCO groups. **C:** AFM measurement setup for internal anchor points. Coh conjugated with Fgβ was added to the AFM measurement buffer and bound with the SdrG immobilized on the AFM tip. Doc was immobilized on the glass surface with the native anchor geometry (from N-terminus). The AFM cantilever was approached to the glass surface to form a 3-member SdrG:Coh:Doc complex. The AFM cantilever was subsequently retracted at a constant speed to dissociate the Doc:Coh complex at the selected pulling geometry. The rupture force SdrG:Fgβ is ~20 fold higher than Doc:Coh and

therefore the final rupture events on force-extension curves are contributed to Doc:Coh rupture. **D:** Force-loading rate plot of non-native pulling geometries. The rupture forces measured at different non-native pulling geometries were plotted against loading rate. The rupture forces exhibited unimodal (Coh N-terminus anchor point) or bimodal (all other pulling geometries) distributions and the most probable rupture force of each population was fitted against loading rate with Eq. 1.5 to extract k_0 and $\Delta x^{\ddagger}$ (solid line: high force population, dashed line: low force population). **E:** Percentage of high rupture force population measured at different anchor points and different pulling speeds. Among all the native and non-native pulling geometries, pulling from F13 of Coh is the only one exhibiting a catch bond behavior, where the prevalence of high rupture force population increases with loading rate. The error bars represent the standard deviation of the high force population percentage.

4.2.3 Change of pulling geometry gives rise to a catch bond behavior

To understand how the Doc:Coh complexes responses to external forces loaded from different directions, we measured the rupture force distributions as well as the loading rate dependency of the Doc:Coh at non-native pulling geometries by changing the anchor point on the Doc or the Coh. The measured non-native anchor points include the C-terminus of Doc, the N-terminus of Coh, and internal residues F13 and K87 of Coh. In these cases, the binding partners were anchored at the native anchor points (N-terminus of Doc or C-terminus of Coh).

To dissociate the Doc:Coh complex by pulling the Doc at the C-terminus, or pulling the Coh at the N-terminus, Doc-FLN-ELP-ybbr and ybbr-ELP-FLN-Coh constructs were cloned into bacterial expression vectors and expressed using *E. coli*. The Doc-FLN-ELP-ybbr and ybbr-ELP-FLN-Coh proteins were immobilized on the same AFM setups as the native pulling geometry measurement (**Figure 4.1B**). For the C-terminus anchor geometry of Doc, Doc-FLN-ELP-ybbr was immobilized on the glass surface and Coh-FLN-ELP-ybbr was immobilized on the AFM tip. For the N-terminus anchor geometry of Coh, ybbr-ELP-FLN-Coh was immobilized on the AFM tip and ybbr-ELP-FLN-Doc was immobilized on the glass surface. Similar to the native pulling geometry, the AFM cantilever was approached to the surface and dwelled for 200 ms to form a Doc:Coh complex and retracted at a constant speed (ranging from 100 nm s^{-1} to 3200 nm s^{-1}) to dissociate the complex at the selected pulling geometry and measure the force-extension curves.

The internal residues of Coh (F13 and K87) were anchored to the AFM tip using an AFM experimental setup combined with click chemistry (**Figure 4.3A and B**) [347]. The anchor residue on the Coh was replaced by a noncanonical amino acid, azidophenylalanine, using amber suppression. The azide at the selected anchor point was covalently conjugated with a Fgβ-StrepTag-DBCO peptide using the click reaction between the azide and dibenzocyclooctyne (DBCO) groups. Fgβ is the N-terminus peptide of the human fibrinogen β chain, which binds SdrG, a bacterial adhesin from *Staphylococcus epidermidis*. SdrG-FLN-ELP-ybbr protein was immobilized on the AFM cantilever and ybbr-ELP-FLN-Doc was immobilized on the glass surface. The Coh conjugated with Fgβ was purified using size-exclusion and StrepTrap columns to remove the excess peptide and unreacted Coh, and subsequently added to the AFM measurement buffer (TBS Ca) to a final concentration of ~100 nM. A 3-member complex was formed as the AFM cantilever approached the surface and dwelled for 200 ms, consisting of SdrG, Fgβ-StrepTag-Coh and Doc. The cantilever was subsequently retracted at a constant speed (100-3200 nm s^{-1}) to apply tension to the Doc:Coh complex from the N-terminus of Doc and the selected internal residue of Coh (F13 or K87) until the Doc:Coh ruptured. The rupture force of Fgβ:SdrG complex is around 2 nN [4], which is ~20 fold higher than the Doc:Coh rupture force. Therefore the final rupture event recorded on the force-extension curves always came from the Doc:Coh complex, as the system could not ramp up to sufficiently high forces to rupture the Fgβ:SdrG complex.

The collected force-extension curves were filtered using the same criterion as the native pulling geometry measurement, i.e. all the selected force curves contain the stretching of a single pair of ELP linkers. The rupture forces of Doc:Coh complex at different non-native pulling geometries were plotted against loading rate (**Figure 4.3C**). All the non-native pulling geometry measured, except pulling from the C-terminus of Coh, show a bimodal distribution of rupture forces (**Figure 4.S2**). The most probable rupture force of each population was linearly fitted against the logarithm of loading rate to extract k_0 and $\Delta x^{\ddagger}$ using Eq. 1.5 (see **Table 4.1**) [123,124].

The pulling geometry affects the rupture forces as well as the prevalence of rupture force populations. Among all the pulling geometries, pulling from residue F13 of Coh gives rise to a unique catch bond behavior, where the percentage of high rupture force population increases with pulling speed, as shown in **Figure 4.3D**. The prevalence of high rupture force population is independent from the pulling speed at other pulling geometries. Anchoring at the C-terminus of Coh gives rise to an unimodal rupture force distribution and therefore the percentage of high force population is considered to be 100%.

The pulling geometry at residue F13 of Coh and the N-terminus of Doc gives rise to two unbinding pathways under force. The low rupture force pathway has a higher prevalence at low loading rate and the rupture force distribution moves to the high rupture force pathway as loading rate increases, with the high force pathway dominant at high loading rates, as shown in **Figure 4.3D and E**. This behavior precisely resembles previously reported catch bonds [149,156,168,346], and is unique among all the five measured pulling geometries of the Doc:Coh complex.

Table 4.1 Energy landscape parameters of Doc:Coh complex at different pulling geometries

Pulling geometry and pathway	**log(k_0)**	**$\Delta x^{\ddagger}$ [nm]**
Native high force	-7±1	0.73±0.09
Native low force	-9±3	1.3±0.4
Doc C-terminus high force	-7±2	1.1±0.2
Doc C-terminus low force	-31±7	10±2
Coh N-terminus	-4.2±0.3	0.65±0.04
Coh F13 high force	-2.5±0.6	0.37±0.06
Coh F13 low force	-0.8±0.2	0.51±0.04
Coh K87 high force	-4.8±0.6	0.58±0.06
Coh K87 low force	-0.7±0.3	0.41±0.05

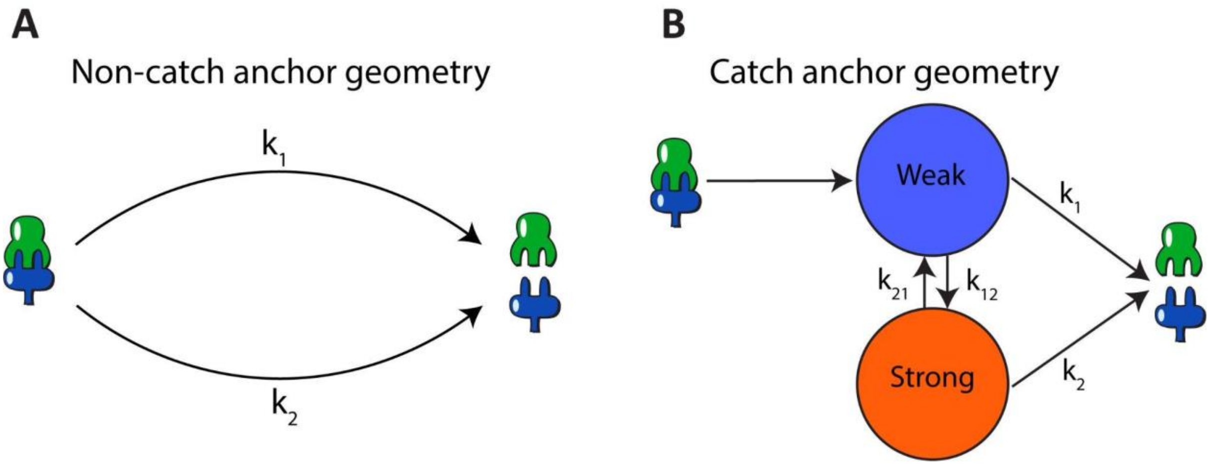


Figure 4.3 Kinetic model of non-catch bond and catch bond pulling geometries. A: The complex dissociates in one or two (non-interchangeable) pathways at non-catch pulling geometries. The probability of entering each pathway and the energy landscape of each pathway are dependent on the pulling geometry. **B:** At catch pulling geometry (pulling from residue 13 of Coh), the complex enters a weak state upon binding. It either dissociates from the weak state or enters a strong state, which has higher mechanostability than the weak state. The rate of entering the strong state from the weak state is force-dependent.

4.3 Discussions and outlook

The anisotropic response of proteins to external forces has been widely discussed for single protein domains as well as protein-ligand complexes and previously research demonstrated that the shapes and heights of unfolding and unbinding pathways are dependent on the pulling geometry, giving rise to distinct unfolding and rupture forces at different pulling geometries [56,347]. Here, we used AFM-SMFS to dissociate a Doc:Coh complex from different directions. The complex ruptures in two pathways with distinct rupture forces at four of the five measured pulling geometries, including the native pulling geometry. The anchor geometry affects both the mechanostabilities of different pathways and the probability of entering each pathway. Single-molecule FRET measurements demonstrated that the complex populates a single binding conformation.

Among the measured anchor geometries, the residue F13 of Coh gives rise to a unique catch bond behavior. At this anchor geometry, the complex dissociates in two pathways when probed using constant pulling speed mode. In contrast to other anchor geometries, the prevalence of each pathway is dependent on the force loading rate, where the high force pathway becomes more prevalent at high loading rates, precisely resembling a catch bond behavior. Kinetic models were developed to describe the behavior of the Doc:Coh complex under external forces at non-catch bonding and catch bonding anchor geometries (**Figure 4.4**).

In addition to the constant speed AFM-SMFS measurements reported here, force clamp measurement is another approach to demonstrate catch bond behavior. We used the same AFM-SMFS setup as constant pulling speed measurements (**Figure 4.2C**) to measure the lifetime of the Doc:Coh complex with constant pulling force applied to the complex from the N-terminus of Doc and residue F13 of Coh. The clamping force was set to 30, 40, 50, 60, 70 and 80 pN. As shown in **Figure 4.S3**, the measured lifetime has a wide distribution, and the recorded clamping force shows a continuous distribution, indicating that the drifting of AFM cantilever made it difficult to precisely control the clamping force at the desired values. Therefore a more reliable approach is to use Monte Carlo simulations to simulate the force clamp measurements based on the constant pulling speed data and the kinetic models (**Figure 4.4**)

Combining the experimental results with computer simulations, including steered molecular dynamics (SMD) and Monte Carlo simulations, would potentially provide additional mechanistic insight into the system, verify the kinetic model, and expand the conclusions to scenarios that are difficult to measure experimentally, including force ramp measurements at extremely high and low loading rates, and force clamp measurements.

4.4 Methods

4.4.1 Protein expression and purification

The Coh F13AzF and Coh K87AzF mutants, as well as the Coh (F13AzF)-FLN-ELP-ybbr and Coh (K87AzF)-FLN-ELP-ybbr constructs, were expressed using amber suppression. C321.ΔA.exp bacteria (gift from George Church, Addgene bacterial strain # 49018) [354] were transformed with pQE80l vector carrying the gene of the mutant, and pEVOL-pAzF plasmid (gift from Peter Schultz, Addgene plasmid # 31186) [275]. The transformed cells were grown in LB medium at 37 °C with ampicillin and chloramphenicol until OD reached ~0.5. Arabinose was added to the medium to final concentration of 0.02% and the culture was incubated at 37 °C for 1 h, followed by addition of isopropyl β-d-1-thiogalactopyranoside (IPTG) to final concentration of 1 mM to induce protein expression. The culture was subsequently incubated at 20 °C overnight.

The other proteins were expressed using NiCo21 (DE3) bacteria. The bacterial cells were transformed using pET28a vector carrying the gene of interest and grown in LB medium at 37 °C in the presence of 50 μg mL^{-1} kanamycin until OD reached ~0.6. IPTG was added to the medium to final concentration of 0.5 mM to induce protein expression. The culture was subsequently incubated at 20 °C overnight.

To purify the proteins of interest, the bacterial cells were pelleted, resuspended in TBS Ca buffer (25mM Tris, 72mM NaCl, 1 mM $CaCl_2$, pH 7.2), and lysed using sonication. The cell lysate was centrifuged at 4 °C 18000 g for 30 min. The supernatant was loaded to a His-trap column and washed using TBS Ca buffer with 20 mM imidazole. The target protein was eluted using TBS Ca buffer with 500 mM imidazole. The eluent was subsequently purified using a size-exclusion column in TBS Ca buffer.

4.4.2 Structural modelling

The structural model of DocG was built using the SWISS-MODEL server, available at https://swissmodel.expasy.org/interactive . The template was searched with BLAST against the SWISS-MODEL template library and the selected template structure was PDB 4WKZ (sequence identity 29.11%, sequence similarity 33%, coverage 94%). The structural model was built based on the sequence alignment between the target and the template using ProMod3 [350]. The structural model of CohE in complex with XMod-DocB was previously reported [168]. The structure of DocG:CohE complex was built by aligning the structural model of DocG to the CohE:XMod-DocB structure using PyMol.

4.4.3 Constant pulling speed AFM measurement

Constant pulling speed AFM measurements were performed on a Force Robot AFM (Bruker, Billerica, MA, USA). Biolever mini (Olympus) AFM cantilevers were amino-silanized with (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany), incubated in 10 mg/mL sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-

carboxylate (sulfo-SMCC, Thermo Fisher Scientific, Waltham, MA, USA) at room temperature for 30 min, extensively washed with ddH_2O, followed by incubation in 200 μM coenzyme A (CoA, Sigma-Aldrich, St. Louis, MO, USA) at room temperature for 2 h. The ybbr-tagged protein (Coh or SdrG) was immobilized on the cantilever by incubating the CoA coated cantilever with a reaction mixture consisting of ~40 μM ybbr-tagged protein, 5 μM Sfp (phosphopantetheinyl transferase) enzyme and 20 mM $MgCl_2$, at room temperature for 2 h.

The cover glasses were silanized and coated with CoA using the same protocol as the cantilevers. CoA-coated cover glasses were incubated with 500 nM ybbr-tagged Doc in the presence of 5 μM Sfp and 20 mM $MgCl_2$ at room temperature for 2 h to immobilize the Doc on the cover glasses.

The cover glasses and AFM cantilever coated with target proteins were extensively washed using TBS Ca buffer and submerged in TBS Ca buffer for AFM measurements. The cantilever spring constants (ranging from 0.02 N m^{-1} to 0.14 N m^{-1}) and detector sensitivity were calibrated using the contact-free mode of the Force Robot AFM. In the measurements using freely diffusing Coh, ~ 100 nM Fgβ-conjugated Coh was added to the measurement buffer. The cantilever was approached to the glass surface, dwelled for 200 ms and retracted at a constant speed ranging from 100 nm s^{-1} to 3200 nm s^{-1} (100 nm s^{-1}, 400 nm s^{-1}, 800 nm s^{-1} and 3200 nm s^{-1}). The x-y position of the sample stage was moved by 100 nm after the retraction, so that a new molecule was probed at the next approach.

4.4.4 AFM data analysis

The recorded force-extension curves were transformed to contour length using the freely rotating chain (FRC) mode described by Eq. 1.17 [143] and were subsequently filtered based on the stretching of the two ELP linkers on the AFM tip and the glass surface. The final rupture forces and force loading rates were calculated for each of the selected curves. The rupture forces measured at different pulling speeds were plotted in histograms and fitted using one or two Gaussian peaks, depending on the anchor geometry. The most probable rupture force of each pathway was plotted against the logarithm of loading rate and fitted using a linear model to extract k_0 and $\Delta x^{\ddagger}$, as described by the Bell-Evans model (Eq. 1.5) [123,317].

4.5 Supplementary information

4.5.1 Supplementary figures

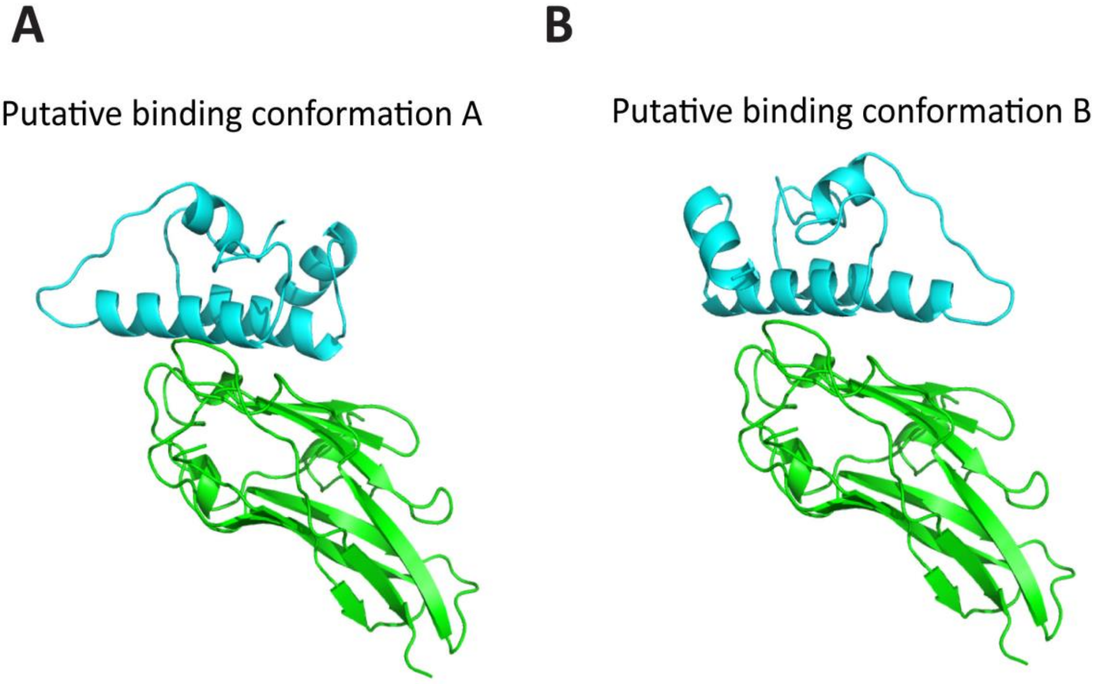


Figure 4.S1 Putative binding conformations of the complex. A: Putative binding conformation A, where the N- and C-termini of the Doc are closer to the C-terminus of the Coh whereas the helix 2 of the Doc is farther away from the C-terminus of Coh. **B:** Putative binding conformation B, where the Doc is rotated by 180 degrees on the binding interface compared to binding conformation **B.** The helix 2 of Doc is therefore closer to the C-terminus of Coh than the N- and C-termini of Doc.

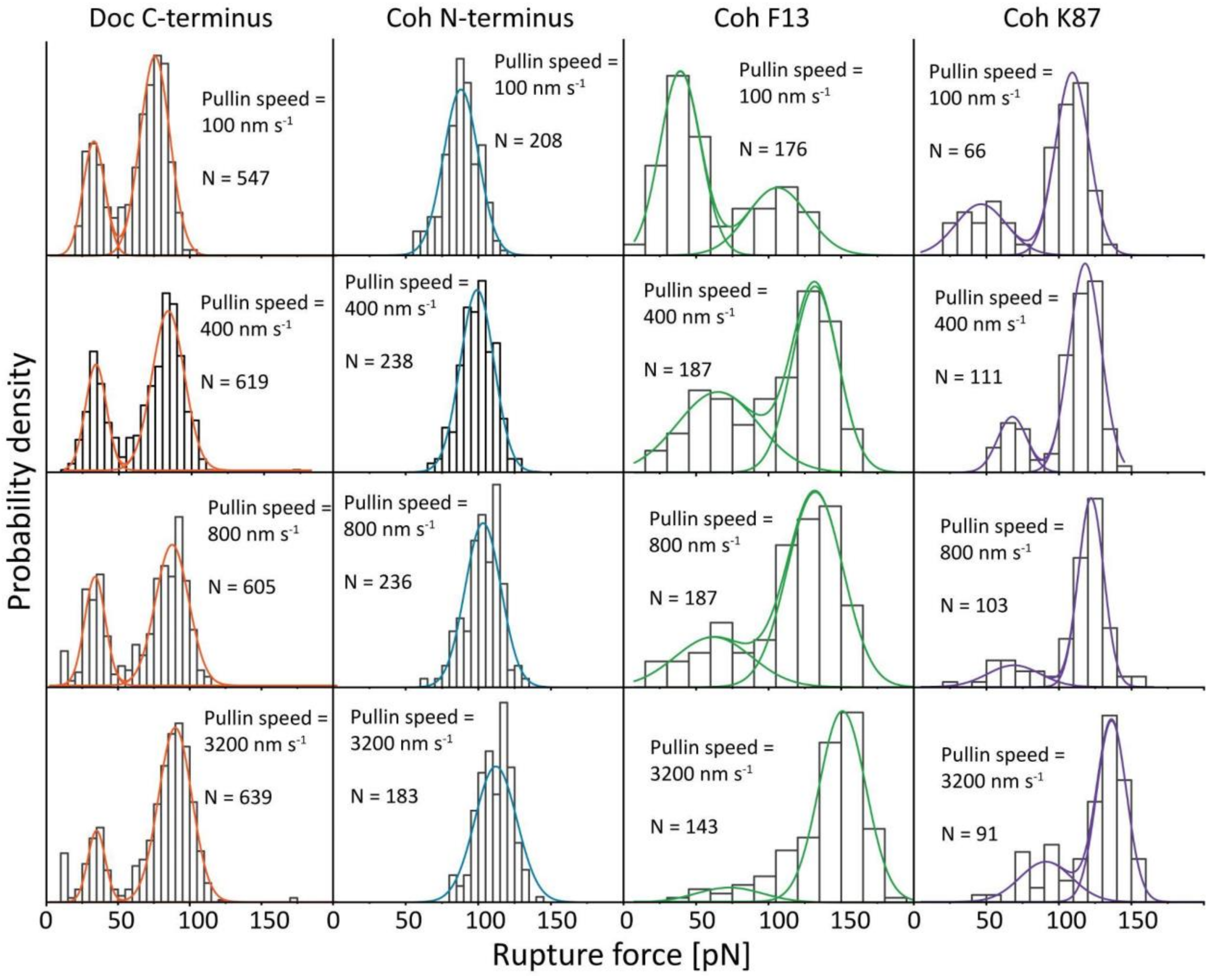


Figure 4.S2 Rupture force histograms measured at different anchor geometries and different pulling speeds. The histograms were fitted with one (Coh N-terminus anchor geometry) or two (other anchor geometries) Gaussian peaks.

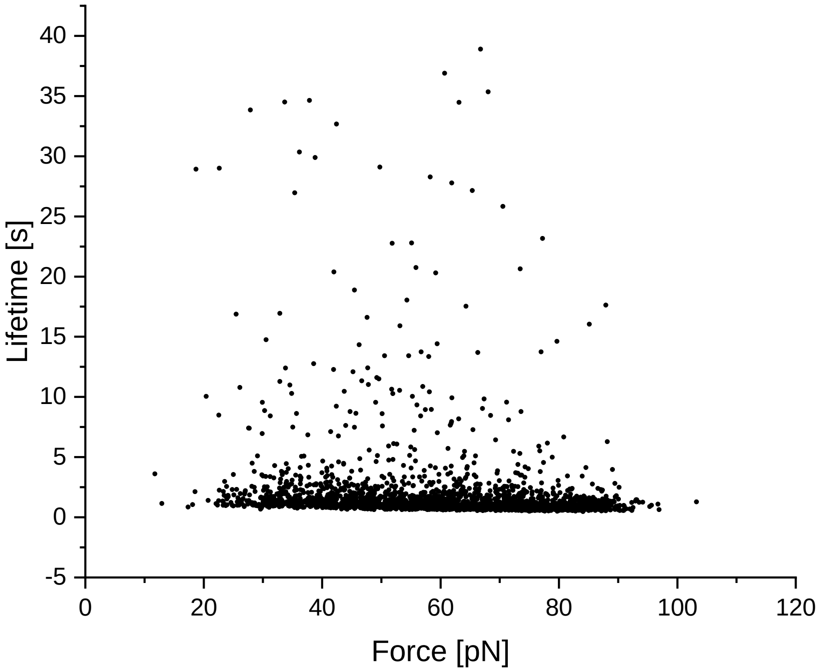


Figure 4.S3 Force clamp measurements of Doc:Coh complex with Doc anchored at the N-terminus and Coh anchored at residue F13. The clamping force was set at 30, 40, 50, 60, 70 and 80 pN.

4.5.2 Supplementary tables

Table 4.S1 Equilibrium affinity of the Doc:Coh complex

	Wild-type Coh	Coh F13AzF-Fgβ	Coh K87AzF-Fgβ
Dissociation constant [nM]	14±4	40±8	19±10

4.5.3 Supplementary notes

Supplementary note 1 Amino acid sequences

Color code:

ybbr tag, His tag, ELP linker, FLN, Doc, Coh

ybbr-His-ELP-FLN-Doc

MGTDSLEFIASKLAHHHHHHWGSGHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWPSGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSGSGVGDSLLRGDVDLDGDVDVADAVAVLQASAEQMVTGENPLSKDARFGADVNDDSRVDVSDAVLILQYSSMKIANPDADWDDLLG

Coh-FLN-ELP-His-ybbr

ADGAAKLSMDQKFAEPGETVEIALNLENFDASWTGLEFLVNYDPKLEVALDGAGDIDYSYGDAIGAMGKKISVGGAISKDLTADGLKGFAFAWGTATAISGNGQLGVFKFTVPADAQPGDEFPVNLTVNVGSFIDANKENIPFETVNGWIKIKEEGSGSGSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWRGHHHHHHGSDSLEFIASKLA

Doc-FLN-ELP-His-ybbr

GVGDSLLRGDVDLDGDVDVADAVAVLQASAEQMVTGENPLSKDARFGADVNDDSRVDVSDAVLILQYSSMKIANPDADWDDLLGGSGSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWRGHHHHHHGSDSLEFIASKLA

ybbr-His-ELP-FLN-Coh

MGTDSLEFIASKLAHHHHHHWGSGHGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGVGVPGMGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGEGVPGWPSGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSADGAAKLSMDQKFAEPGETVEIALNLENFDASWTGLEFLVNYDPKLEVALDGAGDIDYSYGDAIGAMGKKISVGGAISKDLTADGLKGFAFAWGTATAISGNGQLGVFKFTVPADAQPGDEFPVNLTVNVGSFIDANKENIPFETVNGWIKIKEE

Chapter 5 Conclusions and outlook

This work used AFM-SMFS, in combination with computer simulations and other biophysical and biochemical techniques, to probe and engineer mechanostable protein-ligand complexes, with the focus on Doc:Coh systems from the human gut bacterium *R. champanellensis* and non-Ig scaffold anticalin in complex with its target CTLA-4.

Doc:Coh complexes are high force cell adhesion proteins found in the extracellular cellulosomal networks of fiber-digesting bacteria. AFM-SMFS, SMD and smFRET provided comprehensive information on the molecular mechanism of *Rc.* XMod-DocB:CohE complex. This complex can be formed in two binding conformations, and dissociates in three pathways with distinct mechanical stabilities under external pulling force. The XMod at the N-terminus of the Doc tightens at high loading rates to suppress a low-force pathway, giving rise to a catch bond behavior in the high force binding conformation. The bacterial cells use this complex mechanism to regulate the cell adhesion strength and maintain high flexibility while achieving sufficient mechanical stability to resist shear stress from the environment. This work demonstrated the power of AFM-SMFS in understanding the mechanisms of naturally occurring high force protein-ligand complexes at their *in vivo* state, by precisely mimicking the native force loading geometry. The anticalin:CTLA-4 complex, on the contrary, is an engineered artificial system. Here, we developed a novel experimental method to combine AFM-SMFS and click chemistry, which was used to dissociate the anticalin:CTLA-4 complex from eight different directions and demonstrate that the unbinding energy landscape is highly dependent on the pulling geometry, giving rise to distinct rupture forces at different force loading geometries. Conventional AFM-SMFS methods are able to precisely control the anchor points on protein-ligand complexes, but the anchor residues are limited to the N- and C-termini. The method developed here expanded the measurable anchor points to the internal residues, providing a versatile platform to study mechanical anisotropy and optimize the mechanical stability of protein-ligand complexes.

The work on the *Rc.* DocG:CohE system represents both the probing (i.e. to mimic the native state of the complex) and engineering (i.e. to modify the properties of the molecules) aspects. At the native state, the complex forms in one binding mode and dissociates in two pathways with different stabilities under force. Using the AFM-SMFS method developed in the anticalin:CTLA-4 work, the DocG:CohE complex was dissociated at five different pulling geometries. The change of pulling geometry affects the rates of entering different pathways, as well as the unbinding energy landscapes, which determine the off rates of each pathway. The complex exhibits catch bond behavior at a specific non-native pulling geometry, meaning that new behaviors, for example catch bonding, can be introduced to protein-ligand systems by changing the pulling geometry.

The results presented here decoded the complex mechanisms of high force cell adhesion systems in the human gut, and demonstrated a novel aspect of molecular engineering: force loading geometry of protein-ligand complexes. While most of the molecular engineering works on protein-ligand systems focus on the binding interface, this work showed that the behavior of protein complexes can be modified by precisely controlling the force loading geometry, which does not require any change on the binding interface.

Acknowledgements

The successful completion of my PhD work, including the writing of this thesis, would not be possible without the support of many people. The first person I would like to thank is my advisor, Prof. Michael Nash for the opportunity of performing my PhD research in his lab. He taught me how to perform scientific research, how to communicate the results to the audience, and more importantly, how to be a scientist. I want to thank Prof. Roderick Lim, my second supervisor, and Prof. Jorge Alegre-Cebollada, the external expert on my PhD committee, for the insightful discussions on my research projects.

I would like to thank all the former and current Nash group members, who have created a warm, supportive, and inspiring environment in the lab and office. Special thanks to the AFM team: Haipei, Byeongseon, Fanny, Sean and Yang, for the support and discussions on the technique, the instruments, and data analysis. I want to thank the senior PhD students, Mariana and Ivan, for walking through the PhD journey together from the very beginning to the completion. Thanks to Regina, for keeping the lab always functional, and Claudia, for the administrative support. Thanks to all the current Nash lab members, especially Joanan, for the farewell gifts and lunch when I left Basel. I will always cherish the memory of the four years I spent in the Nash lab.

I would like to thank my collaborators: Prof. Rafael Bernardi, Prof. Philip Tinnefeld, Dr. Andrés Vera Gómez, Dr. Adolfo Poma and Dr. Rodrigo Moreira. It is a pleasure to work with them on various projects. Their impressive computational and experimental works as well as insightful discussions pushed these projects to the highest level.

Finally, I would like to thank my parents for their unconditional love and support in my entire life, and my partner Hanyu, for each and every minute we have been, and will be, spending together.

References

[1] Vining KH, Mooney DJ. Mechanical forces direct stem cell behaviour in development and regeneration. Nat Rev Mol Cell Biol 2017;18:728–42.

[2] Leckband DE, de Rooij J. Cadherin adhesion and mechanotransduction. Annu Rev Cell Dev Biol 2014;30:291–315.

[3] Foster TJ, Geoghegan JA, Ganesh VK, Höök M. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. Nat Rev Microbiol 2014;12:49–62.

[4] Milles LF, Schulten K, Gaub HE, Bernardi RC. Molecular mechanism of extreme mechanostability in a pathogen adhesin. Science 2018.

[5] Carvalho AL, Dias FMV, Prates JAM, Nagy T, Gilbert HJ, Davies GJ, et al. Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. Proceedings of the National Academy of Sciences 2003;100:13809–14.

[6] Schoeler C, Malinowska KH, Bernardi RC, Milles LF, Jobst MA, Durner E, et al. Ultrastable cellulosome-adhesion complex tightens under load. Nat Commun 2014;5:5635.

[7] Zhang Y, Yu J, Bomba HN, Zhu Y, Gu Z. Mechanical Force-Triggered Drug Delivery. Chem Rev 2016;116:12536–63.

[8] Zhang Y, Zhang J, Chen M, Gong H, Thamphiwatana S, Eckmann L, et al. A Bioadhesive Nanoparticle–Hydrogel Hybrid System for Localized Antimicrobial Drug Delivery. ACS Appl Mater Interfaces 2016;8:18367–74.

[9] Fang J, Mehlich A, Koga N, Huang J, Koga R, Gao X, et al. Forced protein unfolding leads to highly elastic and tough protein hydrogels. Nat Commun 2013;4:2974.

[10] Wu J, Li P, Dong C, Jiang H, Xue B, Gao X, et al. Rationally designed synthetic protein hydrogels with predictable mechanical properties. Nature Communications 2018;9. https://doi.org/10.1038/s41467-018-02917-6.

[11] Hickey RJ, Pelling AE. Cellulose Biomaterials for Tissue Engineering. Front Bioeng Biotechnol 2019;7:45.

[12] Lv S, Dudek DM, Cao Y, Balamurali MM, Gosline J, Li H. Designed biomaterials to mimic the mechanical properties of muscles. Nature 2010;465:69–73.

[13] Yang B, Liu Z, Liu H, Nash MA. Next Generation Methods for Single-Molecule Force Spectroscopy on Polyproteins and Receptor-Ligand Complexes. Front Mol Biosci 2020;7:85.

[14] Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 1997;276:1109–12.

[15] Florin EL, Moy VT, Gaub HE. Adhesion forces between individual ligand-receptor pairs. Science 1994;264:415–7.

[16] Grandbois M, Beyer M, Rief M, Clausen-Schaumann H, Gaub HE. How Strong Is a Covalent Bond? Science 1999;283:1727–30.

[17] Puchner EM, Alexandrovich A, Kho AL, Hensen U, Schäfer LV, Brandmeier B, et al. Mechanoenzymatics of titin kinase. Proc Natl Acad Sci U S A 2008;105:13385–90.

[18] Eckels EC, Haldar S, Tapia-Rojo R, Rivas-Pardo JA, Fernández JM. The Mechanical Power of Titin Folding. Cell Rep 2019;27:1836–47.e4.

[19] Rivas-Pardo JA, Eckels EC, Popa I, Kosuri P, Linke WA, Fernández JM. Work Done by Titin Protein Folding Assists Muscle Contraction. Cell Rep 2016;14:1339–47.

[20] Hazra JP, Sagar A, Arora N, Deb D, Kaur S, Rakshit S. Broken force dispersal network in tip-links by the mutations at the Ca2+-binding residues induces hearing-loss. Biochem

J 2019;476:2411–25.

[21] Lee G, Abdi K, Jiang Y, Michaely P, Bennett V, Marszalek PE. Nanospring behaviour of ankyrin repeats. Nature 2006;440:246–9.

[22] Bartsch TF, Hengel FE, Oswald A, Dionne G, Chipendo IV, Mangat SS, et al. Elasticity of individual protocadherin 15 molecules implicates tip links as the gating springs for hearing. Proc Natl Acad Sci U S A 2019;116:11048–56.

[23] Mulhall EM, Ward A, Yang D, Koussa MA, Corey DP, Wong WP. The Dynamic Strength of the Hair-Cell Tip Link Reveals Mechanisms of Hearing and Deafness. Biorxiv n.d. https://doi.org/10.1101/763847.

[24] Oroz J, Galera-Prat A, Hervás R, Valbuena A, Fernández-Bravo D, Carrión-Vázquez M. Nanomechanics of tip-link cadherins. Sci Rep 2019;9:13306.

[25] Zhmurov A, Brown AEX, Litvinov RI, Dima RI, Weisel JW, Barsegov V. Mechanism of Fibrin(ogen) Forced Unfolding. Structure/Folding and Design 2011;19:1615–24.

[26] Müller JP, Mielke S, Löf A, Obser T, Beer C, Bruetzel LK, et al. Force sensing by the vascular protein von Willebrand factor is tuned by a strong intermonomer interaction. Proc Natl Acad Sci U S A 2016;113:1208–13.

[27] Kim J, Zhang C-Z, Zhang X, Springer TA. A mechanically stabilized receptor-ligand flex-bond important in the vasculature. Nature 2010;466:992–5.

[28] Zhang X, Wojcikiewicz E, Moy VT. Force spectroscopy of the leukocyte function-associated antigen-1/intercellular adhesion molecule-1 interaction. Biophys J 2002;83:2270–9.

[29] Marshall BT, Long M, Piper JW, Yago T, McEver RP, Zhu C. Direct observation of catch bonds involving cell-adhesion molecules. Nature 2003;423:190–3.

[30] Evans EA, Calderwood DA. Forces and bond dynamics in cell adhesion. Science 2007;316:1148–53.

[31] Oberhauser AF, Marszalek PE, Erickson HP, Fernandez JM. The molecular elasticity of the extracellular matrix protein tenascin. Nature 1998;393:181–5.

[32] Oberhauser AF, Badilla-Fernandez C, Carrion-Vazquez M, Fernandez JM. The Mechanical Hierarchies of Fibronectin Observed with Single-molecule AFM. J Mol Biol 2002;319:433–47.

[33] Guinn EJ, Tian P, Shin M, Best RB, Marqusee S. A small single-domain protein folds through the same pathway on and off the ribosome. Proc Natl Acad Sci U S A 2018;115:12206–11.

[34] Samelson AJ, Bolin E, Costello SM, Sharma AK, O’Brien EP, Marqusee S. Kinetic and structural comparison of a protein’s cotranslational folding and refolding pathways. Sci Adv 2018;4:eaas9098.

[35] Goldman DH, Kaiser CM, Milin A, Righini M. Mechanical force releases nascent chain–mediated ribosome arrest in vitro and in vivo 2015.

[36] Cordova JC, Olivares AO, Shin Y, Stinson BM, Calmat S, Schmitz KR, et al. Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. Cell 2014;158:647–58.

[37] Aubin-Tam M-E, Olivares AO, Sauer RT, Baker TA, Lang MJ. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. Cell 2011;145:257–67.

[38] Bornschlögl T, Woehlke G, Rief M. Single molecule mechanics of the kinesin neck. Proc Natl Acad Sci U S A 2009;106:6992–7.

[39] Neuman KC, Abbondanzieri EA, Landick R, Gelles J, Block SM. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell 2003;115:437–47.

[40] Gelles J, Landick R. RNA polymerase as a molecular motor. Cell 1998;93:13–6.
[41] Milic B, Andreasson JOL, Hancock WO, Block SM. Kinesin processivity is gated by phosphate release. Proc Natl Acad Sci U S A 2014;111:14136–40.
[42] Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, et al. Biophysical properties of the clinical-stage antibody landscape. Proc Natl Acad Sci U S A 2017;114:944–9.
[43] Golinski AW, Holec PV, Mischler KM, Hackel BJ. Biophysical Characterization Platform Informs Protein Scaffold Evolvability. ACS Comb Sci 2019;21:323–35.
[44] Rief M, Pascual J, Saraste M, Gaub HE. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. J Mol Biol 1999;286:553–61.
[45] Oesterhelt F, Oesterhelt D, Pfeiffer M, Engel A, Gaub HE, Müller DJ. Unfolding pathways of individual bacteriorhodopsins. Science 2000;288:143–6.
[46] Schwaiger I, Schleicher M, Noegel AA, Rief M. The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. EMBO Rep 2005;6:46–51.
[47] Bornschlögl T, Rief M. Single-molecule protein unfolding and refolding using atomic force microscopy. Methods Mol Biol 2011;783:233–50.
[48] Scholl ZN, Marszalek PE. AFM-Based Single-Molecule Force Spectroscopy of Proteins. Methods Mol Biol 2018;1814:35–47.
[49] Otten M, Ott W, Jobst MA, Milles LF, Verdorfer T, Pippig DA, et al. From genes to protein mechanics on a chip. Nat Methods 2014;11:1127–30.
[50] Fritz J, Katopodis AG, Kolbinger F, Anselmetti D. Force-mediated kinetics of single P-selectin/ligand complexes observed by atomic force microscopy. Proc Natl Acad Sci U S A 1998;95:12283–8.
[51] Berquand A, Xia N, Castner DG, Clare BH, Abbott NL, Dupres V, et al. Antigen binding forces of single antilysozyme Fv fragments explored by atomic force microscopy. Langmuir 2005;21:5517–23.
[52] Dupres V, Menozzi FD, Locht C, Clare BH, Abbott NL, Cuenot S, et al. Nanoscale mapping and functional analysis of individual adhesins on living bacteria. Nat Methods 2005;2:515–20.
[53] Li J, Li Q, Potthoff S, Wei G, Lucio CC. Force spectroscopic detection of peptide cleavage by thrombin exploiting biotin–streptavidin interactions in a bio-sensing context. Anal Methods 2019;11:1102–10.
[54] Kienberger F, Kada G, Gruber HJ, Pastushenko VP, Riener C, Trieb M, et al. Recognition force spectroscopy studies of the NTA-His6 bond. Single Molecules 2000;1:59–65.
[55] Hinterdorfer P, Baumgartner W, Gruber HJ, Schilcher K, Schindler H. Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. Proc Natl Acad Sci U S A 1996;93:3477–81.
[56] Dietz H, Berkemeier F, Bertz M, Rief M. Anisotropic deformation response of single protein molecules. Proc Natl Acad Sci U S A 2006;103:12724–8.
[57] Guo S, Ray C, Kirkpatrick A, Lad N, Akhremitchev BB. Effects of multiple-bond ruptures on kinetic parameters extracted from force spectroscopy measurements: revisiting biotin-streptavidin interactions. Biophys J 2008;95:3964–76.
[58] Johnson KC, Thomas WE. How Do We Know when Single-Molecule Force Spectroscopy Really Tests Single Bonds? Biophys J 2018;114:2032–9.
[59] Tong Z, Mikheikin A, Krasnoslobodtsev A, Lv Z, Lyubchenko YL. Novel polymer linkers for single molecule AFM force spectroscopy. Methods 2013;60:161–8.
[60] Schoeler C, Verdorfer T, Gaub HE, Nash MA. Biasing effects of receptor-ligand complexes on protein-unfolding statistics. Phys Rev E 2016;94:042412.

[61] Zimmermann JL, Nicolaus T, Neuert G, Blank K. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. Nat Protoc 2010;5:975–85.
[62] Ott W, Jobst MA, Bauer MS, Durner E, Milles LF, Nash MA, et al. Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. ACS Nano 2017;11:6346–54.
[63] Liu H, Ta DT, Nash MA. Mechanical Polyprotein Assembly Using Sfp and Sortase-Mediated Domain Oligomerization for Single-Molecule Studies. Small Methods 2018;2:1800039.
[64] Durner E, Ott W, Nash MA, Gaub HE. Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. ACS Omega 2017;2:3064–9.
[65] Schoeler C, Bernardi RC, Malinowska KH, Durner E, Ott W, Bayer EA, et al. Mapping Mechanical Force Propagation through Biomolecular Complexes. Nano Lett 2015;15:7370–6.
[66] Milles LF, Bayer EA, Nash MA, Gaub HE. Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium perfringens. J Phys Chem B 2017;121:3620–5.
[67] Bernardi RC, Durner E, Schoeler C, Malinowska KH, Carvalho BG, Bayer EA, et al. Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy. J Am Chem Soc 2019;141:14752–63.
[68] Liu H, Schittny V, Nash MA. Removal of a Conserved Disulfide Bond Does Not Compromise Mechanical Stability of a VHH Antibody Complex. Nano Lett 2019;19:5524–9.
[69] Ponnuraj K, Bowden MG, Davis S, Gurusiddappa S, Moore D, Choe D, et al. A “dock, lock, and latch” Structural Model for a Staphylococcal Adhesin Binding to Fibrinogen. Cell 2003;115:217–28.
[70] Sedlak SM, Schendel LC, Melo MCR, Pippig DA, Luthey-Schulten Z, Gaub HE, et al. Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. Nano Lett 2019;19:3415–21.
[71] Sedlak SM, Schendel LC, Gaub HE, Bernardi RC. Streptavidin/biotin: Tethering geometry defines unbinding mechanics. Sci Adv 2020;6:eaay5999.
[72] Oberhauser AF, Hansma PK, Carrion-Vazquez M, Fernandez JM. Stepwise unfolding of titin under force-clamp atomic force microscopy. Proc Natl Acad Sci U S A 2001;98:468–72.
[73] Marszalek PE, Li H, Oberhauser AF, Fernandez JM. Chair-boat transitions in single polysaccharide molecules observed with force-ramp AFM. Proc Natl Acad Sci U S A 2002;99:4278–83.
[74] Popa I, Kosuri P, Alegre-Cebollada J, Garcia-Manyes S, Fernandez JM. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. Nat Protoc 2013;8:1261–76.
[75] Liang J, Fernández JM. Mechanochemistry: one bond at a time. ACS Nano 2009;3:1628–45.
[76] Alegre-Cebollada J, Kosuri P, Giganti D, Eckels E, Rivas-Pardo JA, Hamdani N, et al. S-glutathionylation of cryptic cysteines enhances titin elasticity by blocking protein folding. Cell 2014;156:1235–46.
[77] Alegre-Cebollada J, Perez-Jimenez R, Kosuri P, Fernandez JM. Single-molecule force spectroscopy approach to enzyme catalysis. J Biol Chem 2010;285:18961–6.
[78] Kosuri P, Alegre-Cebollada J, Feng J, Kaplan A, Inglés-Prieto A, Badilla CL, et al.

Protein folding drives disulfide formation. Cell 2012;151:794–806.
[79] Kahn TB, Fernández JM, Perez-Jimenez R. Monitoring Oxidative Folding of a Single Protein Catalyzed by the Disulfide Oxidoreductase DsbA. J Biol Chem 2015;290:14518–27.
[80] Beedle AEM, Mora M, Davis CT, Snijders AP, Stirnemann G, Garcia-Manyes S. Forcing the reversibility of a mechanochemical reaction. Nat Commun 2018;9:3155.
[81] Beedle AEM, Mora M, Lynham S, Stirnemann G, Garcia-Manyes S. Tailoring protein nanomechanics with chemical reactivity. Nat Commun 2017;8:15658.
[82] Giganti D, Yan K, Badilla CL, Fernandez JM, Alegre-Cebollada J. Disulfide isomerization reactions in titin immunoglobulin domains enable a mode of protein elasticity. Nat Commun 2018;9:185.
[83] Garcia-Manyes S, Brujić J, Badilla CL, Fernández JM. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. Biophys J 2007;93:2436–46.
[84] Garcia-Manyes S, Dougan L, Badilla CL, Brujic J, Fernández JM. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. Proc Natl Acad Sci U S A 2009;106:10534–9.
[85] Berkovich R, Garcia-Manyes S, Urbakh M, Klafter J, Fernandez JM. Collapse dynamics of single proteins extended by force. Biophys J 2010;98:2692–701.
[86] Echelman DJ, Alegre-Cebollada J, Badilla CL, Chang C, Ton-That H, Fernández JM. CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. Proc Natl Acad Sci U S A 2016;113:2490–5.
[87] Berkovich R, Hermans RI, Popa I, Stirnemann G, Garcia-Manyes S, Berne BJ, et al. Rate limit of protein elastic response is tether dependent. Proc Natl Acad Sci U S A 2012;109:14416–21.
[88] Jacobson DR, Uyetake L, Perkins TT. Membrane-Protein Unfolding Intermediates Detected with Enhanced Precision Using a Zigzag Force Ramp. Biophys J 2020;118:667–75.
[89] Nash MA. Zig Zag AFM Protocol Reveals New Intermediate Folding States of Bacteriorhodopsin. Biophys J 2020;118:538–40.
[90] Becke TD, Ness S, Sudhop S, Gaub HE, Hilleringmann M, Schilling AF, et al. Covalent Immobilization of Proteins for the Single Molecule Force Spectroscopy. J Vis Exp 2018. https://doi.org/10.3791/58167.
[91] Oesterhelt F, Rief M, Gaub HE. Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water. New Journal of Physics 1999;1:6–6. https://doi.org/10.1088/1367-2630/1/1/006.
[92] Liese S, Gensler M, Krysiak S, Schwarzl R, Achazi A, Paulus B, et al. Hydration Effects Turn a Highly Stretched Polymer from an Entropic into an Energetic Spring. ACS Nano 2017;11:702–12.
[93] Tan YH, Schallom JR, Ganesh NV, Fujikawa K, Demchenko AV, Stine KJ. Characterization of protein immobilization on nanoporous gold using atomic force microscopy and scanning electron microscopy. Nanoscale 2011;3:3395–407.
[94] Ebner A, Wildling L, Kamruzzahan ASM, Rankl C, Wruss J, Hahn CD, et al. A new, simple method for linking of antibodies to atomic force microscopy tips. Bioconjug Chem 2007;18:1176–84.
[95] Walder R, LeBlanc M-A, Van Patten WJ, Edwards DT, Greenberg JA, Adhikari A, et al. Rapid Characterization of a Mechanically Labile α-Helical Protein Enabled by Efficient Site-Specific Bioconjugation. J Am Chem Soc 2017;139:9867–75.
[96] Theile CS, Witte MD, Blom AEM, Kundrat L, Ploegh HL, Guimaraes CP. Site-specific

N-terminal labeling of proteins using sortase-mediated reactions. Nat Protoc 2013;8:1800–7.
[97] Srinivasan S, Hazra JP, Singaraju GS, Deb D, Rakshit S. ESCORTing proteins directly from whole cell-lysate for single-molecule studies. Anal Biochem 2017;535:35–42.
[98] Garg S, Singaraju GS, Yengkhom S, Rakshit S. Tailored Polyproteins Using Sequential Staple and Cut. Bioconjug Chem 2018;29:1714–9.
[99] Harris KS, Durek T, Kaas Q, Poth AG, Gilding EK, Conlan BF, et al. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. Nat Commun 2015;6:10199.
[100] Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, et al. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc Natl Acad Sci U S A 2005;102:15815–20.
[101] Baumann F, Bauer MS, Rees M, Alexandrovich A, Gautel M, Pippig DA, et al. Increasing evidence of mechanical force as a functional regulator in smooth muscle myosin light chain kinase. Elife 2017;6. https://doi.org/10.7554/eLife.26473.
[102] Edwards DT, LeBlanc M-A, Perkins TT. Modulation of a protein-folding landscape revealed by AFM-based force spectroscopy notwithstanding instrumental limitations. Proc Natl Acad Sci U S A 2021;118. https://doi.org/10.1073/pnas.2015728118.
[103] Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 2008;3:373–82.
[104] Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nat Biotechnol 2003;21:86–9.
[105] Taniguchi Y, Kawakami M. Application of HaloTag protein to covalent immobilization of recombinant proteins for single molecule force spectroscopy. Langmuir 2010;26:10433–6.
[106] Popa I, Berkovich R, Alegre-Cebollada J, Badilla CL, Rivas-Pardo JA, Taniguchi Y, et al. Nanomechanics of HaloTag tethers. J Am Chem Soc 2013;135:12762–71.
[107] Kufer SK, Dietz H, Albrecht C, Blank K, Kardinal A, Rief M, et al. Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. Eur Biophys J 2005;35:72–8.
[108] Fichtner D, Lorenz B, Engin S, Deichmann C, Oelkers M, Janshoff A, et al. Covalent and Density-Controlled Surface Immobilization of E-Cadherin for Adhesion Force Spectroscopy. PLoS ONE 2014;9:e93123. https://doi.org/10.1371/journal.pone.0093123.
[109] Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN. Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. Science 2007;318:1625–8.
[110] Zakeri B, Howarth M. Spontaneous intermolecular amide bond formation between side chains for irreversible peptide targeting. J Am Chem Soc 2010;132:4526–7.
[111] Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, et al. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc Natl Acad Sci U S A 2012;109:E690–7.
[112] Veggiani G, Nakamura T, Brenner MD, Gayet RV, Yan J, Robinson CV, et al. Programmable polyproteams built using twin peptide superglues. Proc Natl Acad Sci U S A 2016;113:1202–7.
[113] Griffo A, Rooijakkers BJM, Hähl H, Jacobs K, Linder MB, Laaksonen P. Binding Forces of Cellulose Binding Modules on Cellulosic Nanomaterials. Biomacromolecules 2019;20:769–77.
[114] Yang B, Liu H, Liu Z, Doenen R, Nash MA. Influence of Fluorination on Single-

Molecule Unfolding and Rupture Pathways of a Mechanostable Protein Adhesion Complex. Nano Lett 2020;20:8940–50.

[115] Kim CH, Axup JY, Schultz PG. Protein conjugation with genetically encoded unnatural amino acids. Curr Opin Chem Biol 2013;17:412–9.

[116] Wang L, Xie J, Schultz PG. Expanding the genetic code. Annu Rev Biophys Biomol Struct 2006;35:225–49.

[117] Deiters A, Cropp TA, Summerer D, Mukherji M, Schultz PG. Site-specific PEGylation of proteins containing unnatural amino acids. Bioorg Med Chem Lett 2004;14:5743–5.

[118] Maity S, Viazovkina E, Gall A, Lyubchenko Y. A Metal-free Click Chemistry Approach for the Assembly and Probing of Biomolecules. J Nat Sci 2016;2.

[119] Yu H, Heenan PR, Edwards DT, Uyetake L, Perkins TT. Quantifying the Initial Unfolding of Bacteriorhodopsin Reveals Retinal Stabilization. Angew Chem Int Ed Engl 2019;58:1710–3.

[120] Cho H, Daniel T, Buechler YJ, Litzinger DC, Maio Z, Putnam A-MH, et al. Optimized clinical performance of growth hormone with an expanded genetic code. Proc Natl Acad Sci U S A 2011;108:9060–5.

[121] Hallam TJ, Wold E, Wahl A, Smider VV. Antibody Conjugates with Unnatural Amino Acids. Molecular Pharmaceutics 2015;12:1848–62. https://doi.org/10.1021/acs.molpharmaceut.5b00082.

[122] Woodside MT, Block SM. Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu Rev Biophys 2014;43:19–39.

[123] Bell GI. Models for the specific adhesion of cells to cells. Science 1978;200:618–27.

[124] Evans E, Ritchie K. Dynamic strength of molecular adhesion bonds. Biophys J 1997;72:1541–55.

[125] Evans E. Probing the Relation Between Force—Lifetime—and Chemistry in Single Molecular Bonds. Annu Rev Biophys Biomol Struct 2001;30:105–28.

[126] Dietz H, Rief M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. Proc Natl Acad Sci U S A 2004;101:16192–7.

[127] Hummer G, Szabo A. Kinetics from nonequilibrium single-molecule pulling experiments. Biophys J 2003;85:5–15.

[128] Dudko OK, Hummer G, Szabo A. Intrinsic rates and activation free energies from single-molecule pulling experiments. Phys Rev Lett 2006;96:108101.

[129] Dudko OK, Hummer G, Szabo A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. Proc Natl Acad Sci U S A 2008;105:15755–60.

[130] Friddle RW, Noy A, De Yoreo JJ. Interpreting the widespread nonlinear force spectra of intermolecular bonds. Proc Natl Acad Sci U S A 2012;109:13573–8.

[131] Friddle RW, Podsiadlo P, Artyukhin AB, Noy A. Near-Equilibrium Chemical Force Microscopy. J Phys Chem C 2008;112:4986–90.

[132] Hughes ML, Dougan L. The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. Rep Prog Phys 2016;79:076601.

[133] Dietz H, Rief M. Protein structure by mechanical triangulation. Proc Natl Acad Sci U S A 2006;103:1244–7.

[134] Puchner EM, Franzen G, Gautel M, Gaub HE. Comparing proteins by their unfolding pattern. Biophys J 2008;95:426–34.

[135] Puchner EM, Gaub HE. Force and function: probing proteins with AFM-based force spectroscopy. Curr Opin Struct Biol 2009;19:605–14.

[136] 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 U S A 2009;106:13307–133310.
[137] Vera AM, Carrión-Vázquez M. Direct Identification of Protein-Protein Interactions by Single-Molecule Force Spectroscopy. Angewandte Chemie 2016;128:14176–9. https://doi.org/10.1002/ange.201605284.
[138] Milles LF, Gaub HE. Is mechanical receptor ligand dissociation driven by unfolding or unbinding? n.d. https://doi.org/10.1101/593335.
[139] Halvorsen K, Schaak D, Wong WP. Nanoengineering a single-molecule mechanical switch using DNA self-assembly. Nanotechnology 2011;22:494005.
[140] Yang D, Ward A, Halvorsen K, Wong WP. Multiplexed single-molecule force spectroscopy using a centrifuge. Nat Commun 2016;7:11026.
[141] Bustamante C, Marko JF, Siggia ED, Smith S. Entropic elasticity of lambda-phage DNA. Science 1994;265:1599–600.
[142] Ortiz C, Hadziioannou G. Entropic Elasticity of Single Polymer Chains of Poly(methacrylic acid) Measured by Atomic Force Microscopy. Macromolecules 1999;32:780–7. https://doi.org/10.1021/ma981245n.
[143] Livadaru L, Netz RR, Kreuzer HJ. Stretching Response of Discrete Semiflexible Polymers. Macromolecules 2003;36:3732–44.
[144] Hugel T, Rief M, Seitz M, Gaub HE, Netz RR. Highly stretched single polymers: atomic-force-microscope experiments versus ab-initio theory. Phys Rev Lett 2005;94:048301.
[145] Cai W, Lu S, Wei J, Cui S. Single-Chain Polymer Models Incorporating the Effects of Side Groups: An Approach to General Polymer Models. Macromolecules 2019;52:7324–30.
[146] Valiaev A, Lim DW, Oas TG, Chilkoti A, Zauscher S. Force-induced prolyl cis-trans isomerization in elastin-like polypeptides. J Am Chem Soc 2007;129:6491–7.
[147] Moy VT, Florin EL, Gaub HE. Intermolecular forces and energies between ligands and receptors. Science 1994;266:257–9.
[148] Jobst MA, Milles LF, Schoeler C, Ott W, Fried DB, Bayer EA, et al. Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. Elife 2015;4. https://doi.org/10.7554/eLife.10319.
[149] Thomas W. Catch bonds in adhesion. Annu Rev Biomed Eng 2008;10:39–57.
[150] Thomas W, Forero M, Yakovenko O, Nilsson L, Vicini P, Sokurenko E, et al. Catch-bond model derived from allostery explains force-activated bacterial adhesion. Biophys J 2006;90:753–64.
[151] Huang DL, Bax NA, Buckley CD, Weis WI, Dunn AR. Vinculin forms a directionally asymmetric catch bond with F-actin. Science 2017;357:703–6.
[152] Buckley CD, Tan J, Anderson KL, Hanein D, Volkmann N, Weis WI, et al. Cell adhesion. The minimal cadherin-catenin complex binds to actin filaments under force. Science 2014;346:1254211.
[153] Guo B, Guilford WH. Mechanics of actomyosin bonds in different nucleotide states are tuned to muscle contraction. Proc Natl Acad Sci U S A 2006;103:9844–9.
[154] Pereverzev YV, Prezhdo OV, Forero M, Sokurenko EV, Thomas WE. The two-pathway model for the catch-slip transition in biological adhesion. Biophys J 2005;89:1446–54.
[155] Pereverzev YV, Prezhdo OV, Thomas WE, Sokurenko EV. Distinctive features of the biological catch bond in the jump-ramp force regime predicted by the two-pathway model. Phys Rev E Stat Nonlin Soft Matter Phys 2005;72:010903.

[156] Thomas WE, Vogel V, Sokurenko E. Biophysics of catch bonds. Annu Rev Biophys 2008;37:399–416.

[157] Schwarz WH. The cellulosome and cellulose degradation by anaerobic bacteria. Appl Microbiol Biotechnol 2001;56:634–49.

[158] Bayer EA, Shoham Y, Lamed R. Cellulose-decomposing bacteria and their enzyme systems. Prokaryotes 2006;2:578–617.

[159] Bayer EA, Lamed R, White BA, Flint HJ. From cellulosomes to cellulosomics. Chem Rec 2008;8:364–77.

[160] Ben David Y, Dassa B, Borovok I, Lamed R, Koropatkin NM, Martens EC, et al. Ruminococcal cellulosome systems from rumen to human. Environ Microbiol 2015;17:3407–26.

[161] Verdorfer T, Bernardi R, Meinhold A, Ott W, Luthey-Schulten Z, Nash MA, et al. Combining in Vitro and in Silico Single Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. J Am Chem Soc 2017. https://doi.org/10.1021/jacs.7b07574.

[162] Bule P, Pires VM, Fontes CM, Alves VD. Cellulosome assembly: paradigms are meant to be broken! Curr Opin Struct Biol 2018;49:154–61.

[163] Karpol A, Jobby MK, Slutzki M, Noach I, Chitayat S, Smith SP, et al. Structural and functional characterization of a novel type-III dockerin from Ruminococcus flavefaciens. FEBS Lett 2013;587:30–6.

[164] Adams JJ, Pal G, Jia Z, Smith SP. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. Proc Natl Acad Sci U S A 2006;103:305–10.

[165] Salama-Alber O, Jobby MK, Chitayat S. Atypical Cohesin-Dockerin Complex Responsible for Cell Surface Attachment of Cellulosomal Components BINDING FIDELITY, PROMISCUITY, AND STRUCTURAL …. Journal of Biological 2013.

[166] Adams JJ, Webb BA, Spencer HL, Smith SP. Structural Characterization of Type II Dockerin Module from the Cellulosome of Clostridium thermocellum: Calcium-Induced Effects on Conformation and Target Recognition. Biochemistry 2005;44:2173–82. https://doi.org/10.1021/bi048039u.

[167] Song G, Tian F, Liu H, Li G, Zheng P. Pioglitazone Inhibits Metal Cluster Transfer of mitoNEET by Stabilizing the Labile Fe–N Bond Revealed at Single-Bond Level. J Phys Chem Lett 2021;12:3860–7.

[168] Liu Z, Liu H, Vera AM, Bernardi RC, Tinnefeld P, Nash MA. High force catch bond mechanism of bacterial adhesion in the human gut. Nat Commun 2020;11:4321.

[169] Weissman AM. Themes and variations on ubiquitylation. Nat Rev Mol Cell Biol 2001;2:169–78.

[170] Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. Nat Struct Biol 2003;10:738–43.

[171] Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, Smith DA, et al. Pulling geometry defines the mechanical resistance of a β-sheet protein. Nat Struct Mol Biol 2003;10:731–7.

[172] Eyal E, Bahar I. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. Biophys J 2008;94:3424–35.

[173] Li YD, Lamour G, Gsponer J, Zheng P, Li H. The molecular mechanism underlying mechanical anisotropy of the protein GB1. Biophys J 2012;103:2361–8.

[174] Lee W, Zeng X, Rotolo K, Yang M, Schofield CJ, Bennett V, et al. Mechanical anisotropy of ankyrin repeats. Biophys J 2012;102:1118–26.

[175] Krautbauer R, Rief M, Gaub HE. Unzipping DNA Oligomers. Nano Lett 2003;3:493–6.
[176] Rief M, Clausen-Schaumann H, Gaub HE. Sequence-dependent mechanics of single DNA molecules. Nat Struct Biol 1999;6:346–9.
[177] Woodside MT, Behnke-Parks WM, Larizadeh K, Travers K, Herschlag D, Block SM. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. Proc Natl Acad Sci U S A 2006;103:6190–5.
[178] Sun Y, Di W, Li Y, Huang W, Wang X, Qin M, et al. Mg2 -Dependent High Mechanical Anisotropy of Three-Way-Junction pRNA as Revealed by Single-Molecule Force Spectroscopy. Angewandte Chemie International Edition 2017;56:9376–80. https://doi.org/10.1002/anie.201704113.
[179] Niu X, Liu Q, Xu Z, Chen Z, Xu L, Xu L, et al. Molecular mechanisms underlying the extreme mechanical anisotropy of the flaviviral exoribonuclease-resistant RNAs (xrRNAs). Nat Commun 2020;11:5496.
[180] Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, et al. Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. N Engl J Med 2017;377:1919–29.
[181] Thatcher N, Hirsch FR, Luft AV, Szczesna A, Ciuleanu TE, Dediu M, et al. Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controlled phase 3 trial. The Lancet Oncology 2015;16:763–74. https://doi.org/10.1016/s1470-2045(15)00021-2.
[182] Baselga J, Cortés J, Kim S-B, Im S-A, Hegg R, Im Y-H, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med 2012;366:109–19.
[183] Emu B, Fessel J, Schrader S, Kumar P, Richmond G, Win S, et al. Phase 3 Study of Ibalizumab for Multidrug-Resistant HIV-1. N Engl J Med 2018;379:645–54.
[184] Ataga KI, Kutlar A, Kanter J, Liles D, Cancado R, Friedrisch J, et al. Crizanlizumab for the Prevention of Pain Crises in Sickle Cell Disease. N Engl J Med 2017;376:429–39.
[185] Nair P, Wenzel S, Rabe KF, Bourdin A, Lugogo NL, Kuna P, et al. Oral Glucocorticoid–Sparing Effect of Benralizumab in Severe Asthma. N Engl J Med 2017;376:2448–58.
[186] Beck A, Goetsch L, Dumontet C, Corvaïa N. Strategies and challenges for the next generation of antibody–drug conjugates. Nat Rev Drug Discov 2017;16:315–37.
[187] Chen F, Hong H, Zhang Y, Valdovinos HF, Shi S, Kwon GS, et al. In vivo tumor targeting and image-guided drug delivery with antibody-conjugated, radiolabeled mesoporous silica nanoparticles. ACS Nano 2013;7:9027–39.
[188] Xu S, Cui F, Huang D, Zhang D, Zhu A, Sun X, et al. PD-L1 monoclonal antibody-conjugated nanoparticles enhance drug delivery level and chemotherapy efficacy in gastric cancer cells. Int J Nanomedicine 2019;14:17–32.
[189] Shi M, Ho K, Keating A, Shoichet MS. Doxorubicin-conjugated immuno-nanoparticles for intracellular anticancer drug delivery. Adv Funct Mater 2009;19:1689–96.
[190] Rosenfeld PJ, Brown DM, Heier JS, Boyer DS, Kaiser PK, Chung CY, et al. Ranibizumab for neovascular age-related macular degeneration. N Engl J Med 2006;355:1419–31.
[191] Hodi FS, O’Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010;363:711–23.
[192] Migden MR, Rischin D, Schmults CD, Guminski A, Hauschild A, Lewis KD, et al.

PD-1 Blockade with Cemiplimab in Advanced Cutaneous Squamous-Cell Carcinoma. N Engl J Med 2018;379:341–51.

[193] Cruz E, Kayser V. Monoclonal antibody therapy of solid tumors: clinical limitations and novel strategies to enhance treatment efficacy. Biologics 2019;13:33–51.

[194] Richards DA. Exploring alternative antibody scaffolds: Antibody fragments and antibody mimics for targeted drug delivery. Drug Discov Today Technol 2018;30:35–46.

[195] Gebauer M, Skerra A. Engineered Protein Scaffolds as Next-Generation Therapeutics. Annu Rev Pharmacol Toxicol 2020;60:391–415.

[196] Zhang P, Woen S, Wang T, Liau B, Zhao S, Chen C, et al. Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. Drug Discovery Today 2016;21:740–65. https://doi.org/10.1016/j.drudis.2016.01.006.

[197] Deonarain MP, Yahioglu G, Stamati I, Pomowski A, Clarke J, Edwards BM, et al. Small-Format Drug Conjugates: A Viable Alternative to ADCs for Solid Tumours? Antibodies (Basel) 2018;7. https://doi.org/10.3390/antib7020016.

[198] Škrlec K, Štrukelj B, Berlec A. Non-immunoglobulin scaffolds: a focus on their targets. Trends Biotechnol 2015;33:408–18.

[199] Schönfeld D, Matschiner G, Chatwell L, Trentmann S, Gille H, Hülsmeyer M, et al. An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies. Proc Natl Acad Sci U S A 2009;106:8198–203.

[200] Jost C, Schilling J, Tamaskovic R, Schwill M, Honegger A, Plückthun A. Structural basis for eliciting a cytotoxic effect in HER2-overexpressing cancer cells via binding to the extracellular domain of HER2. Structure 2013;21:1979–91.

[201] Pannecoucke E, Van Trimpont M, Desmet J, Pieters T, Reunes L, Demoen L, et al. Cell-penetrating Alphabody protein scaffolds for intracellular drug targeting. Sci Adv 2021;7. https://doi.org/10.1126/sciadv.abe1682.

[202] Richter A, Eggenstein E, Skerra A. Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. FEBS Lett 2014;588:213–8.

[203] Feldwisch J, Tolmachev V, Lendel C, Herne N, Sjöberg A, Larsson B, et al. Design of an optimized scaffold for affibody molecules. J Mol Biol 2010;398:232–47.

[204] Ramamurthy V, Krystek SR Jr, Bush A, Wei A, Emanuel SL, Das Gupta R, et al. Structures of adnectin/protein complexes reveal an expanded binding footprint. Structure 2012;20:259–69.

[205] Diem MD, Hyun L, Yi F, Hippensteel R, Kuhar E, Lowenstein C, et al. Selection of high-affinity Centyrin FN3 domains from a simple library diversified at a combination of strand and loop positions. Protein Eng Des Sel 2014;27:419–29.

[206] Forrer P, Binz HK, Stumpp MT, Plückthun A. Consensus design of repeat proteins. Chembiochem 2004;5:183–9.

[207] Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 2003;332:489–503.

[208] Seeger MA, Zbinden R, Flütsch A, Gutte PGM, Engeler S, Roschitzki-Voser H, et al. Design, construction, and characterization of a second-generation DARP in library with reduced hydrophobicity. Protein Sci 2013;22:1239–57.

[209] Varadamsetty G, Tremmel D, Hansen S, Parmeggiani F, Plückthun A. Designed Armadillo repeat proteins: library generation, characterization and selection of peptide binders with high specificity. J Mol Biol 2012;424:68–87.

[210] Lee S-C, Park K, Han J, Lee J-J, Kim HJ, Hong S, et al. Design of a binding scaffold

based on variable lymphocyte receptors of jawless vertebrates by module engineering. Proc Natl Acad Sci U S A 2012;109:3299–304.

[211] Cao L, Goreshnik I, Coventry B, Case JB, Miller L, Kozodoy L, et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science 2020;370:426–31.

[212] Dang Y, Guan J. Nanoparticle-based drug delivery systems for cancer therapy. Smart Materials in Medicine 2020;1:10–9.

[213] Harmsen S, Wall MA, Huang R, Kircher MF. Cancer imaging using surface-enhanced resonance Raman scattering nanoparticles. Nat Protoc 2017;12:1400–14.

[214] Xu X, An H, Zhang D, Tao H, Dou Y, Li X, et al. A self-illuminating nanoparticle for inflammation imaging and cancer therapy. Sci Adv 2019;5:eaat2953.

[215] Yu MK, Park J, Jon S. Targeting strategies for multifunctional nanoparticles in cancer imaging and therapy. Theranostics 2012;2:3–44.

[216] Gu Z, Rolfe BE, Xu ZP, Campbell JH, Max Lu GQ, Thomas AC. Antibody-Targeted Drug Delivery to Injured Arteries Using Layered Double Hydroxide Nanoparticles. Advanced Healthcare Materials 2012;1:669–73. https://doi.org/10.1002/adhm.201200069.

[217] Hadjipanayis CG, Machaidze R, Kaluzova M, Wang L, Schuette AJ, Chen H, et al. EGFRvIII antibody-conjugated iron oxide nanoparticles for magnetic resonance imaging-guided convection-enhanced delivery and targeted therapy of glioblastoma. Cancer Res 2010;70:6303–12.

[218] Suwa T, Ozawa S, Ueda M, Ando N, Kitajima M. Magnetic resonance imaging of esophageal squamous cell carcinoma using magnetite particles coated with anti-epidermal growth factor receptor antibody. Int J Cancer 1998;75:626–34.

[219] Billingsley MM, Riley RS, Day ES. Antibody-nanoparticle conjugates to enhance the sensitivity of ELISA-based detection methods. PLOS ONE 2017;12:e0177592. https://doi.org/10.1371/journal.pone.0177592.

[220] Zhang L, Mazouzi Y, Salmain M, Liedberg B, Boujday S. Antibody-Gold Nanoparticle Bioconjugates for Biosensors: Synthesis, Characterization and Selected Applications. Biosens Bioelectron 2020;165:112370.

[221] Altintas I, Heukers R, van der Meel R, Lacombe M, Amidi M, van Bergen En Henegouwen PMP, et al. Nanobody-albumin nanoparticles (NANAPs) for the delivery of a multikinase inhibitor 17864 to EGFR overexpressing tumor cells. J Control Release 2013;165:110–8.

[222] Zhang Q, Ding F, Liu X, Shen J, Su Y, Qian J, et al. Nanobody-guided targeted delivery of microRNA via nucleic acid nanogel to inhibit the tumor growth. J Control Release 2020;328:425–34.

[223] Satpathy M, Wang L, Zielinski R, Qian W, Lipowska M, Capala J, et al. Active targeting using HER-2-affibody-conjugated nanoparticles enabled sensitive and specific imaging of orthotopic HER-2 positive ovarian tumors. Small 2014;10:544–55.

[224] Gao J, Chen K, Miao Z, Ren G, Chen X, Gambhir SS, et al. Affibody-based nanoprobes for HER2-expressing cell and tumor imaging. Biomaterials 2011;32:2141–8.

[225] Godoy-Gallardo M, Ek PK, Jansman MMT, Wohl BM, Hosta-Rigau L. Interaction between drug delivery vehicles and cells under the effect of shear stress. Biomicrofluidics 2015;9:052605.

[226] Kang T, Tran TT-T, Park C, Lee B-J. Biomimetic shear stress and nanoparticulate drug delivery. Int J Pharm Investig 2017;47:133–9.

[227] Hosta-Rigau L, Städler B. Shear stress and its effect on the interaction of myoblast cells with nanosized drug delivery vehicles. Mol Pharm 2013;10:2707–12.

[228] Sokolov K, Follen M, Aaron J, Pavlova I, Malpica A, Lotan R, et al. Real-time vital

optical imaging of precancer using anti-epidermal growth factor receptor antibodies conjugated to gold nanoparticles. Cancer Res 2003;63:1999–2004.

[229] Fahmy TM, Samstein RM, Harness CC, Mark Saltzman W. Surface modification of biodegradable polyesters with fatty acid conjugates for improved drug targeting. Biomaterials 2005;26:5727–36.

[230] Zhang K, Hao L, Hurst SJ, Mirkin CA. Antibody-linked spherical nucleic acids for cellular targeting. J Am Chem Soc 2012;134:16488–91.

[231] Ryan S, Kell AJ, van Faassen H, Tay L-L, Simard B, MacKenzie R, et al. Single-domain antibody-nanoparticles: promising architectures for increased staphylococcus aureus detection specificity and sensitivity. Bioconjug Chem 2009;20:1966–74.

[232] Sivaram AJ, Wardiana A, Howard CB, Mahler SM, Thurecht KJ. Recent advances in the generation of antibody–nanomaterial conjugates. Adv Healthc Mater 2018;7:1700607.

[233] Tripathi K, Driskell JD. Quantifying Bound and Active Antibodies Conjugated to Gold Nanoparticles: A Comprehensive and Robust Approach To Evaluate Immobilization Chemistry. ACS Omega 2018;3:8253–9.

[234] Steinhauser I, Spänkuch B, Strebhardt K, Langer K. Trastuzumab-modified nanoparticles: optimisation of preparation and uptake in cancer cells. Biomaterials 2006;27:4975–83.

[235] Hsieh C-L, Grange R, Pu Y, Psaltis D. Bioconjugation of barium titanate nanocrystals with immunoglobulin G antibody for second harmonic radiation imaging probes. Biomaterials 2010;31:2272–7.

[236] Greene MK, Richards DA, Nogueira JCF, Campbell K, Smyth P, Fernández M, et al. Forming next-generation antibody–nanoparticle conjugates through the oriented installation of non-engineered antibody fragments. Chemical Science 2018;9:79–87. https://doi.org/10.1039/c7sc02747h.

[237] Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J 1997;72:1568–81.

[238] Thomas WE, Trintchina E, Forero M, Vogel V, Sokurenko EV. Bacterial adhesion to target cells enhanced by shear force. Cell 2002;109:913–23.

[239] Pierse CA, Dudko OK. Distinguishing Signatures of Multipathway Conformational Transitions. Phys Rev Lett 2017;118:088101.

[240] Evans E, Leung A, Heinrich V, Zhu C. Mechanical switching and coupling between two dissociation pathways in a P-selectin adhesion bond. Proc Natl Acad Sci U S A 2004;101:11281–6.

[241] Prezhdo OV, Pereverzev YV. Theoretical aspects of the biological catch bond. Acc Chem Res 2009;42:693–703.

[242] Zhu C, Lou J, McEver RP. Catch bonds: physical models, structural bases, biological function and rheological relevance. Biorheology 2005;42:443–62.

[243] Moraïs S, Ben David Y, Bensoussan L, Duncan SH, Koropatkin NM, Martens EC, et al. Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, Ruminococcus champanellensis, reveals a fine-tuned system for cohesin-dockerin recognition. Environ Microbiol 2016;18:542–56.

[244] Artzi L, Bayer EA, Moraïs S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol 2017;15:83–95.

[245] Smith SP, Bayer EA, Czjzek M. Continually emerging mechanistic complexity of the multi-enzyme cellulosome complex. Curr Opin Struct Biol 2017;44:151–60.

[246] King JR, Bowers CM, Toone EJ. Specific Binding at the Cellulose Binding Module–Cellulose Interface Observed by Force Spectroscopy. Langmuir 2015;31:3431–40.

[247] Carvalho AL, Dias FMV, Nagy T, Prates JAM, Proctor MR, Smith N, et al. Evidence for a dual binding mode of dockerin modules to cohesins. Proc Natl Acad Sci U S A 2007;104:3089–94.

[248] Cameron K, Najmudin S, Alves VD, Bayer EA, Smith SP, Bule P, et al. Cell-surface Attachment of Bacterial Multienzyme Complexes Involves Highly Dynamic Protein-Protein Anchors. J Biol Chem 2015;290:13578–90.

[249] Nash MA, Smith SP, Fontes CM, Bayer EA. Single versus dual-binding conformations in cellulosomal cohesin–dockerin complexes. Curr Opin Struct Biol 2016;40:89–96.

[250] Pinheiro BA, Proctor MR, Martinez-Fleites C, Prates JAM, Money VA, Davies GJ, et al. The Clostridium cellulolyticum Dockerin Displays a Dual Binding Mode for Its Cohesin Partner. J Biol Chem 2008;283:18422–30.

[251] Brás JLA, Pinheiro BA, Cameron K, Cuskin F, Viegas A, Najmudin S, et al. Diverse specificity of cellulosome attachment to the bacterial cell surface. Sci Rep 2016;6:38292.

[252] Yu H, Siewny MGW, Edwards DT, Sanders AW, Perkins TT. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. Science 2017;355:945–50.

[253] Li J, Li H. Mechanical Unfolding Pathway of the High-Potential Iron-Sulfur Protein Revealed by Single-Molecule Atomic Force Microscopy: Toward a General Unfolding Mechanism for Iron-sulfur Proteins. J Phys Chem B 2018;122:9340–9.

[254] Cao Y, Yoo T, Li H. Single molecule force spectroscopy reveals engineered metal chelation is a general approach to enhance mechanical stability of proteins. Proc Natl Acad Sci U S A 2008;105:11152–7.

[255] Sumbul F, Rico F. Single-Molecule Force Spectroscopy: Experiments, Analysis, and Simulations. Methods Mol Biol 2019;1886:163–89.

[256] Hellenkamp B, Schmid S, Doroshenko O, Opanasyuk O, Kühnemuth R, Rezaei Adariani S, et al. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. Nat Methods 2018;15:669–76.

[257] Wang S, Vafabakhsh R, Borschel WF, Ha T, Nichols CG. Structural dynamics of potassium-channel gating revealed by single-molecule FRET. Nat Struct Mol Biol 2016;23:31–6.

[258] Koh HR, Roy R, Sorokina M, Tang G-Q, Nandakumar D, Patel SS, et al. Correlating Transcription Initiation and Conformational Changes by a Single-Subunit RNA Polymerase with Near Base-Pair Resolution. Mol Cell 2018;70:695–706.e5.

[259] Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature 2018;555:61–6.

[260] Holmstrom ED, Liu Z, Nettels D, Best RB, Schuler B. Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. Nature Communications 2019;10. https://doi.org/10.1038/s41467-019-10356-0.

[261] Webb B, Sali A. Protein Structure Modeling with MODELLER. Methods Mol Biol 2017;1654:39–54.

[262] Voronov-Goldman M, Yaniv O, Gul O, Yoffe H, Salama-Alber O, Slutzki M, et al. Standalone cohesin as a molecular shuttle in cellulosome assembly. FEBS Lett 2015;589:1569–76.

[263] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–8, 27–8.

[264] Schwaiger I, Kardinal A, Schleicher M, Noegel AA, Rief M. A mechanical unfolding intermediate in an actin-crosslinking protein. Nat Struct Mol Biol 2004;11:81–5.

[265] Ott W, Nicolaus T, Gaub HE, Nash MA. Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-

Mediated Enzymatic Ligation. Biomacromolecules 2016;17:1330–8.
[266] Ta DT, Vanella R, Nash MA. Magnetic Separation of Elastin-like Polypeptide Receptors for Enrichment of Cellular and Molecular Targets. Nano Lett 2017;17:7932–9.
[267] Ta DT, Vanella R, Nash MA. Bioorthogonal Elastin-like Polypeptide Scaffolds for Immunoassay Enhancement. ACS Appl Mater Interfaces 2018;10:30147–54.
[268] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–802.
[269] Melo MCR, Bernardi RC, Rudack T, Scheurer M, Riplinger C, Phillips JC, et al. NAMD goes quantum: an integrative suite for hybrid simulations. Nat Methods 2018;15:351–4.
[270] Ribeiro JV, Bernardi RC, Rudack T, Stone JE, Phillips JC, Freddolino PL, et al. QwikMD—integrative molecular dynamics toolkit for novices and experts. Sci Rep 2016;6:26536.
[271] Jobst MA, Schoeler C, Malinowska K, Nash MA. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. J Vis Exp 2013:e50950.
[272] Chakrabarti S, Hinczewski M, Thirumalai D. Phenomenological and microscopic theories for catch bonds. J Struct Biol 2017;197:50–6.
[273] Utjesanovic M, Matin TR, Sigdel KP, King GM, Kosztin I. Multiple Stochastic Pathways in Forced Peptide-Lipid Membrane Detachment. Scientific Reports 2019;9. https://doi.org/10.1038/s41598-018-36528-4.
[274] Matin TR, Utjesanovic M, Sigdel KP, Smith VF, Kosztin I, King GM. Characterizing the Locus of a Peripheral Membrane Protein–Lipid Bilayer Interaction Underlying Protein Export Activity in E. coli. Langmuir 2020;36:2143–52. https://doi.org/10.1021/acs.langmuir.9b03606.
[275] Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J Am Chem Soc 2002;124:9026–7.
[276] Rief M, Fernandez JM, Gaub HE. Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility. Phys Rev Lett 1998;81:4764–7.
[277] Yoav S, Barak Y, Shamshoum M, Borovok I, Lamed R, Dassa B, et al. How does cellulosome composition influence deconstruction of lignocellulosic substrates in Clostridium (Ruminiclostridium) thermocellum DSM 1313? Biotechnol Biofuels 2017;10:222.
[278] Fallingborg J. Intraluminal pH of the human gastrointestinal tract. Dan Med Bull 1999;46:183–96.
[279] Salgado PS, Yan R, Taylor JD, Burchell L, Jones R, Hoyer LL, et al. Structural basis for the broad specificity to host-cell ligands by the pathogenic fungus Candida albicans. Proc Natl Acad Sci U S A 2011;108:15775–9.
[280] Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. Candida auris sp. nov., a novel ascomycetous yeast isolated from the external ear canal of an inpatient in a Japanese hospital. Microbiol Immunol 2009;53:41–4.
[281] Rhodes J, Fisher MC. Global epidemiology of emerging Candida auris. Curr Opin Microbiol 2019;52:84–9.
[282] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.
[283] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.

[284] MacArthur MW, Laskowski RA, Thornton JM. Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy. Curr Opin Struct Biol 1994;4:731–7.

[285] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. 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;8:3257–73.

[286] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35.

[287] Scheurer M, Rodenkirch P, Siggel M, Bernardi RC, Schulten K, Tajkhorshid E, et al. PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations. Biophys J 2018;114:577–83.

[288] Müller BK, Zaychikov E, Bräuchle C, Lamb DC. Pulsed interleaved excitation. Biophys J 2005;89:3508–22.

[289] Nir E, Michalet X, Hamadani KM, Laurence TA, Neuhauser D, Kovchegov Y, et al. Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. J Phys Chem B 2006;110:22103–24.

[290] Tomov TE, Tsukanov R, Masoud R, Liber M, Plavner N, Nir E. Disentangling subpopulations in single-molecule FRET and ALEX experiments with photon distribution analysis. Biophys J 2012;102:1163–73.

[291] Kudryavtsev V, Sikor M, Kalinin S. Combining MFD and PIE for accurate single-pair Förster resonance energy transfer measurements 2012.

[292] Lee NK, Kapanidis AN, Wang Y, Michalet X, Mukhopadhyay J, Ebright RH, et al. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. Biophys J 2005;88:2939–53.

[293] Duff MR Jr, Grubbs J, Howell EE. Isothermal titration calorimetry for measuring macromolecule-ligand affinity. J Vis Exp 2011. https://doi.org/10.3791/2796.

[294] Kufer SK, Puchner EM, Gumpp H, Liedl T, Gaub HE. Single-molecule cut-and-paste surface assembly. Science 2008;319:594–6.

[295] Goktas M, Luo C, Sullan RMA, Bergues-Pupo AE, Lipowsky R, Vila Verde A, et al. Molecular mechanics of coiled coils loaded in the shear geometry. Chem Sci 2018;9:4610–21.

[296] Hoffmann T, Dougan L. Single molecule force spectroscopy using polyproteins. Chem Soc Rev 2012;41:4781–96.

[297] Hoffmann T, Tych KM, Hughes ML, Brockwell DJ, Dougan L. Towards design principles for determining the mechanical stability of proteins. Phys Chem Chem Phys 2013;15:15767–80.

[298] Crampton N, Brockwell DJ. Unravelling the design principles for single protein mechanical strength. Curr Opin Struct Biol 2010;20:508–17.

[299] Valbuena A, Oroz J, Hervás R, Vera AM, Rodríguez D, Menéndez M, et al. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc Natl Acad Sci U S A 2009;106:13791–6.

[300] Sikora M, Sułkowska JI, Cieplak M. Mechanical strength of 17,134 model proteins and cysteine slipknots. PLoS Comput Biol 2009;5:e1000547.

[301] Ott W, Jobst MA, Schoeler C, Gaub HE, Nash MA. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J Struct Biol 2017;197:3–12.

[302] Sedlak SM, Bauer MS, Kluger C, Schendel LC, Milles LF, Pippig DA, et al. Monodisperse measurement of the biotin-streptavidin interaction strength in a well-

defined pulling geometry. PLoS One 2017;12:e0188722.

[303] Bedford R, Tiede C, Hughes R, Curd A, McPherson MJ, Peckham M, et al. Alternative reagents to antibodies in imaging applications. Biophys Rev 2017;9:299–308.

[304] Rothe C, Skerra A. Anticalin® Proteins as Therapeutic Agents in Human Diseases. BioDrugs 2018;32:233–43.

[305] Eggenstein E, Eichinger A, -J. Kim H, Skerra A. Structure-guided engineering of Anticalins with improved binding behavior and biochemical characteristics for application in radio-immuno imaging and/or therapy. Journal of Structural Biology 2014;185:203–14. https://doi.org/10.1016/j.jsb.2013.03.009.

[306] Renders L, Budde K, Rosenberger C, van Swelm R, Swinkels D, Dellanna F, et al. First-in-human Phase I studies of PRS-080# 22, a hepcidin antagonist, in healthy volunteers and patients with chronic kidney disease undergoing hemodialysis. PLoS One 2019;14:e0212023.

[307] Lee KM, Chuang E, Griffin M, Khattri R, Hong DK, Zhang W, et al. Molecular basis of T cell inactivation by CTLA-4. Science 1998;282:2263–6.

[308] Walker LSK, Sansom DM. The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. Nat Rev Immunol 2011;11:852–63.

[309] Goetz DH, Holmes MA, Borregaard N, Bluhm ME, Raymond KN, Strong RK. The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. Mol Cell 2002;10:1033–43.

[310] Deuschle F-C, Morath V, Schiefner A, Brandt C, Ballke S, Reder S, et al. Development of a high affinity Anticalin® directed against human CD98hc for theranostic applications. Theranostics 2020;10:2172–87.

[311] Gebauer M, Schiefner A, Matschiner G, Skerra A. Combinatorial design of an Anticalin directed against the extra-domain b for the specific targeting of oncofetal fibronectin. J Mol Biol 2013;425:780–802.

[312] Korndörfer IP, Beste G, Skerra A. Crystallographic analysis of an “anticalin” with tailored specificity for fluorescein reveals high structural plasticity of the lipocalin loop region. Proteins 2003;53:121–9.

[313] Breustedt DA, Chatwell L, Skerra A. A new crystal form of human tear lipocalin reveals high flexibility in the loop region and induced fit in the ligand cavity. Acta Crystallogr D Biol Crystallogr 2009;65:1118–25.

[314] Flower DR. The lipocalin protein family: structure and function. Biochem J 1996;318 (Pt 1):1–14.

[315] Coles M, Diercks T, Muehlenweg B, Bartsch S, Zölzer V, Tschesche H, et al. The solution structure and dynamics of human neutrophil gelatinase-associated lipocalin. J Mol Biol 1999;289:139–57.

[316] Debets MF, van Berkel SS, Schoffelen S, Rutjes FPJT, van Hest JCM, van Delft FL. Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. Chem Commun 2010;46:97–9.

[317] Evans E, Ritchie K. Dynamic strength of molecular adhesion bonds. Biophys J 1997.

[318] Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–15.

[319] Eyal E, Lum G, Bahar I. The anisotropic network model web server at 2015 (ANM 2.0). Bioinformatics 2015;31:1487–9.

[320] Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to α-amylase inhibitor. Proteins: Struct Funct Bioinf 2000;40:512–24.

[321] Souza PCT, Thallmair S, Conflitti P, Ramírez-Palacios C, Alessandri R, Raniolo S, et al. Protein-ligand binding with the coarse-grained Martini model. Nat Commun 2020;11:3714.

[322] Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grunewald F, et al. Martini 3: A general purpose force field for coarse-grain molecular dynamics. Nat Methods 2021;in press. https://doi.org/10.1038/s41592-021-01098-3.

[323] Poma AB, Li MS, Theodorakis PE. Generalization of the elastic network model for the study of large conformational changes in biomolecules. Physical Chemistry Chemical Physics 2018;20:17020–8. https://doi.org/10.1039/c8cp03086c.

[324] Chwastyk M, Bernaola AP, Cieplak M. Statistical radii associated with amino acids to determine the contact map: fixing the structure of a type I cohesin domain in the Clostridium thermocellum cellulosome. Phys Biol 2015;12:046002.

[325] Senapati S, Poma AB, Cieplak M, Filipek S, Park PSH. Differentiating between Inactive and Active States of Rhodopsin by Atomic Force Microscopy in Native Membranes. Anal Chem 2019;91:7226–35.

[326] Poma AB, Chwastyk M, Cieplak M. Elastic moduli of biological fibers in a coarse-grained model: crystalline cellulose and β-amyloids. Phys Chem Chem Phys 2017;19:28195–206.

[327] Moreira RA, Chwastyk M, Baker JL, Guzman HV, Poma AB. Quantitative determination of mechanical stability in the novel coronavirus spike protein. Nanoscale 2020;12:16409–13.

[328] Poma AB, Guzman HV, Li MS, Theodorakis PE. Mechanical and thermodynamic properties of Aβ42, Aβ40, and α-synuclein fibrils: a coarse-grained method to complement experimental studies. Beilstein J Nanotechnol 2019;10:500–13.

[329] Poma AB, Cieplak M, Theodorakis PE. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. J Chem Theory Comput 2017;13:1366–74.

[330] Fontana F, Gelain F. Probing mechanical properties and failure mechanisms of fibrils of self-assembling peptides. Nanoscale Advances 2020;2:190–8.

[331] Alessandri R, Souza PCT, Thallmair S, Melo MN, de Vries AH, Marrink SJ. Pitfalls of the Martini Model. J Chem Theory Comput 2019;15:5448–60.

[332] Bajaj J, Konuma T, Lytle NK, Kwon HY, Ablack JN, Cantor JM, et al. CD98-Mediated Adhesive Signaling Enables the Establishment and Propagation of Acute Myelogenous Leukemia. Cancer Cell 2016;30:792–805.

[333] Furuya M, Horiguchi J, Nakajima H, Kanai Y, Oyama T. Correlation of L-type amino acid transporter 1 and CD98 expression with triple negative breast cancer prognosis. Cancer Science 2012;103:382–9. https://doi.org/10.1111/j.1349-7006.2011.02151.x.

[334] Wang Q, Tiffen J, Bailey CG, Lehman ML, Ritchie W, Fazli L, et al. Targeting amino acid transport in metastatic castration-resistant prostate cancer: effects on cell cycle, cell growth, and tumor development. J Natl Cancer Inst 2013;105:1463–73.

[335] Stahl SW, Nash MA, Fried DB, Slutzki M, Barak Y, Bayer EA, et al. Single-molecule dissection of the high-affinity cohesin–dockerin complex. Proc Natl Acad Sci U S A 2012;109:20431–6.

[336] Kapust RB, Waugh DS. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 1999;8:1668–74.

[337] Gaciarz A, Veijola J, Uchida Y, Saaranen MJ, Wang C, Hörkkö S, et al. Systematic screening of soluble expression of antibody fragments in the cytoplasm of E. coli. Microb Cell Fact 2016;15:22.

[338] Mikulska-Ruminska K, Kulik AJ, Benadiba C, Bahar I, Dietler G, Nowak W. Author Correction: Nanomechanics of multidomain neuronal cell adhesion protein contactin revealed by single molecule AFM and SMD. Sci Rep 2018;8:4291.

[339] Pettersen EF, Goddard TD, Huang CC. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of 2004.

[340] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1-2:19–25.

[341] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald: an N\textperiodcentered log (N) method for Ewald sums in large systems. J Chem Phys 1995;103:8577–93.

[342] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126:014101.

[343] Sauer MM, Jakob RP, Eras J, Baday S, Eriş D, Navarra G, et al. Catch-bond mechanism of the bacterial adhesin FimH. Nat Commun 2016;7:10738.

[344] Mathelié-Guinlet M, Viela F, Pietrocola G, Speziale P, Alsteens D, Dufrêne YF. Force-clamp spectroscopy identifies a catch bond mechanism in a Gram-positive pathogen. Nat Commun 2020;11:5431.

[345] Manibog K, Li H, Rakshit S, Sivasankar S. Resolving the molecular mechanism of cadherin catch bond formation. Nat Commun 2014;5:3941.

[346] Yakovenko O, Sharma S, Forero M, Tchesnokova V, Aprikian P, Kidd B, et al. FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. J Biol Chem 2008;283:11596–605.

[347] Liu Z, Moreira RA, Dujmovic A, Liu H, Yang B. Optimizing mechanostable anchor points of engineered lipocalin in complex with CTLA-4. bioRxiv 2021.

[348] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46:W296–303.

[349] Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, et al. The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res 2017;45:D313–9.

[350] Studer G, Tauriello G, Bienert S, Biasini M, Johner N, Schwede T. ProMod3—A versatile homology modelling toolbox. PLoS Comput Biol 2021;17:e1008667.

[351] Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics 2019;36:1765–71.

[352] Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep 2017;7:10480.

[353] Vera AM, Galera-Prat A, Wojciechowski M, Różycki B, Laurents DV, Carrión-Vázquez M, et al. Cohesin-dockerin code in cellulosomal dual binding modes and its allosteric regulation by proline isomerization. Structure 2021;29:587–97.e8.

[354] Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, et al. Genomically recoded organisms expand biological functions. Science 2013;342:357–60.