Mechanics of the Streptavidin/Biotin Interaction

Steffen M. Sedlak

München 2019

Mechanics of the Streptavidin/Biotin Interaction

Steffen M. Sedlak

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Steffen M. Sedlak
aus Göppingen

München, den 18. Juli 2019

Erstgutachter: Prof. Dr. Hermann Gaub
Zweitgutachter: Prof. Dr. Joachim Rädler
Tag der mündlichen Prüfung: 27. August 2019

Zusammenfassung

Das Protein Streptavidin bindet das kleine Molekül Biotin mit außerordentlich hoher Affinität. Die Interaktion ist spezifisch, langlebig und robust unter verschiedensten Bedingungen. Biotin kann recht einfach an Proteine, Nukleinsäuren oder Nanopartikel gekoppelt werden. Darum wird dieses Rezeptor-Liganden-System für vielzählige Anwendungen in der Biotechnologie, Nanotechnologie und Medizin verwendet. Viele moderne diagnostische Verfahren, wie zum Beispiel Immunoassays, sind auf die stabile Interaktion zwischen Streptavidin und Biotin angewiesen. In der Biophysik dient die Streptavidin/Biotin-Interaktion als molekulares Verankerungssystem, um Moleküle oder Zellen für kraftspektroskopische Untersuchungen (mit optischen oder magnetischen Pinzetten, akustischer oder atomarer Kraftmikroskopie) zu immobilisieren.

In den letzten 25 Jahren sind sehr viele Anstrengungen unternommen worden, um die mechanische Stabilität der engen Bindung von Biotin an Streptavidin zu verstehen. Hinsichtlich der Reaktion des Rezeptor-Liganden-Komplexes auf Trennung durch äußere Kräfte sind widersprüchliche Ergebnisse publiziert worden.

In der vorliegenden Arbeit wurden biomolekulare Protein-Engineering-Methoden verwendet, um individuell angepasste Streptavidinkonstrukte mit definierter Stöchiometrie und Markierungen für eine kontrollierte kovalente Anbindung herzustellen. Mit nanotechnologischen Instrumenten und Methoden, um einzelne Moleküle zu manipulieren, wurden die mechanischen Eigenschaften der Interaktion zwischen Streptavidin und Biotin untersucht. Die Experimente wurden durch Molekulardynamiksimulationen ergänzt, um auch die zu Grunde liegenden molekularen Prozesse zu beleuchten.

Die Verwendung von monovalentem Streptavidin [1] ermöglichte es, das Entbinden von Biotin von einer einzelnen Streptavidinuntereinheit in einer kontrollierten Weise zu messen. Bei Untersuchungen von monovalentem Streptavidin mit Rasterkraftmikroskopie-basierter Einzelmolekülkraftspektroskopie in einer definierten Anbindungsgeometrie wurde eine im Vergleich zu früheren Studien verhältnismäßig schmale Verteilung von Entbindungskräften gemessen [2]. Durch die Verwendung von C- beziehungsweise N-terminal angebundenem monovalentem Streptavidin in Kraftspektroskopieexperimenten wurde außerdem gezeigt, wie verschiedene Kraftladungsgeometrien Entbindungskräfte beeinflussen. Molekulardynamiksimulationen trugen dazu bei, einen Entbindungsweg aufzudecken, der eine partielle Entfaltung der funktionalen Untereinheit von Streptavidin beinhaltet [3]. Die unterschiedlichen Entbindungskräfte für die verschiedenen Anbindungsgeometrien wurden ausgenutzt, um eine Krafthierarchie zu schaffen, die es erlaubt, einzelne Moleküle mithilfe der Spitze des Federbalkens des Rasterkraftmikroskops zu manipulieren, um Proteine in definierten Strukturen im Nanometerbereich auf einer Oberfläche anzuordnen [4]. Dies kann in der Zukunft dafür verwendet werden, Enzymnetzwerke auf Einzelmolekülebene in beliebigen Anordnungen zu untersuchen.

Die Experimente wurden schließlich von monovalentem auf tetravalentes Streptavidin ausgeweitet. Es wurde entdeckt, dass sich die vier Biotin-Bindungstaschen von Streptavidin unter Kraft verschieden verhalten, obwohl sie thermodynamisch gleich sind: Wenn man eine Untereinheit an ihrem C-Terminus befestigt und ein Biotin aus einer der Bindungstaschen zieht, ist die Symmetrie zwischen den Untereinheiten gebrochen und die maximale Rückstell-

kraft, die sich im Proteinkomplex aufbaut, bevor dieser nachgibt, ist für die verschiedenen Bindungstaschen signifikant verschieden [3]. Mithilfe von Molekulardynamiksimulationen konnte gezeigt werden, dass unter Kraft – je nach Anbindungsgeometrie – Biotin und anschließende molekulare Linker gegen einen flexiblen Teil der Bindungstasche drücken. Dies führt zu einer kraftinduzierten Konformationsänderung der Bindungstasche, beeinträchtigt ihre strukturelle Integrität und mindert die Entbindungskraft von Biotin erheblich. Diese Ergebnisse konnten in Experimenten mit konstanter Kraft in magnetischen Pinzetten bestätigt werden [5]. Dieser deutliche Lebensdauerunterschied der Bindung für verschiedene Anbindungsgeometrien ist von Interesse, wenn man eine besonders langlebige Anbindung von Molekülen in einem Analyseverfahren benötigt, bei dem äußere Kräfte (sowohl Zug- als auch Scherkräfte) auf diese wirken.

Die Einblicke, die im Rahmen dieser Arbeit in die mechanische Stabilität der Streptavidin-Biotin-Interaktion gewonnen wurden, haben dazu beigetragen, die früheren Widersprüche zwischen verschiedenen Publikationen zu verstehen und miteinander in Einklang zu bringen: Für verschiedene Kraftladungsgeometrien können für dasselbe Rezeptor-Liganden-System erheblich unterschiedliche Entbindungskräfte gemessen werden. Durch die Ergebnisse dieser Arbeit konnte außerdem gefolgert werden, dass es enge Verflechtungen von Entfaltung und Entbindung im Bezug auf Ligandendissoziation unter Kraft geben kann: Es ist durchaus möglich, dass ein Rezeptor zunächst teilweise entfaltet wird, bevor die Bindung zum Liganden nachgibt. Darüber hinaus wurde gezeigt, dass kraftinduzierte Konformationsänderungen der Bindungstasche, die in einer Schwächung der Bindung resultieren, auftreten können, wenn Streptavidin in einer unvorteilhaften Anbindungsgeometrie verwendet wird. Das bessere Verständnis der mechanischen Stabilität des Biotin-Streptavidin-Systems dient nun als Wegweiser, robuste und unter Kraft stabile Anbindungsgeometrien für Streptavidin in Kraftspektroskopiemessungen zu verwenden. Die Ergebnisse dieser Arbeit können in Zukunft dazu beitragen, Eigenschaften von Streptavidin mit Protein-Engineering-Methoden für bestimmte Anwendungen besser anzupassen und weiter zu optimieren.

Abstract

The homotetrameric protein streptavidin (SA) binds the small molecule biotin with an extraordinarily high affinity. Their interaction is specific, long-lived and robust under a wide range of conditions. Biotin can readily be attached to proteins, nucleic acids or nanoparticles. This receptor-ligand system has thus found wide application in biotechnology, nanotechnology and medicine. Many modern diagnostic techniques, such as immunoassays, rely on the stable interaction of SA and biotin. In biophysics, the SA/biotin interaction serves as a popular anchoring tool to immobilize molecules or cells for force spectroscopy experiments using optical or magnetic tweezers, acoustic or atomic force microscopy.

Within the last 25 years, a lot of effort has been put into understanding the mechanical stability of the tight binding of biotin to SA. Contradictory results on the response of the receptor-ligand complex, when forcibly separated, have been published.

In this thesis, biomolecular protein engineering methods were applied to design and prepare customized SA constructs with defined stoichiometry and handles enabling covalent, site-specific immobilization and labeling for single-molecule or bulk assays. Using tools from nanotechnology to manipulate proteins on the single-molecule level, the mechanical properties of the receptor-ligand interaction between SA and biotin were investigated. Steered molecular dynamics simulations complemented the experiments to shed light on the underlying molecular processes.

The development of monovalent SA [1] made it possible to investigate the unbinding of biotin from a single SA subunit in a controlled manner: Probing monovalent SA by atomic force microscopy-based single-molecule force spectroscopy in a well-defined tethering geometry, a relatively narrow distribution of unbinding forces (compared with previous studies) was observed [2]. Tethering monovalent SA by the C- or N-terminus of its functional subunit, it was demonstrated how different force-loading geometries influence unbinding forces. Steered molecular dynamics simulations helped to reveal an unbinding pathway that involves partial unfolding of SA's functional subunit [3]. Taking advantage of the difference in unbinding forces, a force hierarchy was created that allowed to manipulate single molecules with the cantilever tip of the atomic force microscope to create nanoscale arrangements of proteins on a surface [4]. In the future, this can be used to investigate enzyme networks on the single molecule level in arbitrary arrangements.

The study was finally extended from monovalent to tetravalent SA. Although the four biotin binding pockets of SA are thermodynamically equal, the subunits behave differently when force is applied: Tethering one of the subunits C-terminally and pulling on one of the biotin molecules, SA's symmetry is broken and the maximum restoring force, which builds up before the complex ruptures, is significantly different for the four subunits [3]. With the help of steered molecular dynamics simulations, it was demonstrated that depending on the tethering geometry, biotin and the adjacent linker molecules get pushed against a flexible part of the binding pocket. This induces a conformational change in the binding pocket, impedes its structural integrity and significantly decreases biotin's unbinding force.

These results were confirmed in constant force experiments using magnetic tweezers [5]. The significant difference in bond lifetime for the different tethering geometries is of interest when aiming for particularly long-lived interactions in any assay in which forces (*e.g.* shear or tensile forces) act on the SA/biotin bond.

The insights into the mechanical stability of the SA/biotin interaction that were gained in this work have contributed to understand and to reconcile the contradictory results of previous publications: For different force-loading geometries, significantly different unbinding forces can be measured for the same receptor-ligand system. With the results of this work, it could further be concluded that in the context of ligand dissociation under force, unfolding and unbinding can be intertwined: It is altogether possible that a receptor is partially unfolded before the binding to the ligand gives in. Beyond that, it was shown that if SA is anchored in a disadvantageous geometry, force-induced conformational changes of the binding pocket can occur, resulting in a substantial weakening of the binding. The better understanding of the mechanical stability of the SA/biotin system guides the way to use robust and stable tethering geometries of SA in force spectroscopy measurements. In the future, the results of this work can contribute to modify certain characteristics of SA by protein engineering to further optimize this extraordinary protein for specific applications.

Contents

Part I

Introduction

Introduction

Force spectroscopy is at the heart of mechanobiology, which investigates the response of cells to mechanical signals and aims at understanding relevant molecular mechanisms. The term ”force spectroscopy” comprises various biophysical techniques to investigate mechanical properties of single macromolecules, macromolecular structures or individual bonds between them. The streptavidin (SA)/biotin interaction was among the first receptor-ligand systems to be probed by atomic force microscopy (AFM)-based single-molecule force spectroscopy (SMFS) [6–8]. Over the years, the SA/biotin system has proved to be a reliable handle and has become a standard anchoring tool to specifically link different molecular entities.

SA is a key molecule in biotechnology. Myriads of methods in biochemistry, immunology and molecular biology rely on the robust and specific interaction of SA with the small molecule biotin. With its extraordinarily high affinity in the femtomolar range, this receptor-ligand system is among the – with respect to affinity – strongest non-covalent interactions in nature. Therefore, it has become (and still is) a popular research object. Hereby, the focus is not only on the comprehension of the molecular binding mechanism but also on modifications of certain characteristics of the system for specific applications.

Regarding SMFS experiments, the SA/biotin system is probably the most studied receptor-ligand system (*cf.* Manuscript M1, Section 1.5). Technological advances in biomolecular engineering on SA, such as the development of SA of distinct valencies [1] or the incorporation of a unique cysteine for site-specific immobilization [9], yet paved the way to refine how SMFS on the SA/biotin system is conducted (*cf.* Publication P1, Section 3.1). In the context of this thesis, protein engineering was employed to design SA with different numbers of functional subunits and various anchoring points for site-specific immobilization. By this, new insights into the dependence of the mechanical stability of the SA/biotin interaction on the force-loading geometry were gained (*cf.* Publication P2, Manuscripts M2 and M3, Sections 3.2, 3.3 and 3.4). Findings of these studies were transferred to nanotechnological applications: Monovalent SA in distinct tethering geometries was implemented into single-molecule cut-and-paste to realize a DNA-free assay (*cf.* Publication P3, Section 4.1). In addition, an especially long-lived tethering geometry was identified and incorporated into a magnetic tweezers protocol to enable reliable measurements of single molecules at comparably high constant forces for many hours (*cf.* Manuscript M4, Section 4.2). Beyond the optimization of the established tetrameric SA anchor, a monomeric SA serving as a molecular handle for force spectroscopy was developed and characterized (*cf.* Manuscript M5, Section 5.1).

Part II

Scientific Context

Chapter 1

The Streptavidin/Biotin Interaction: One of the Strongest Non-Covalent Interactions

Streptavidin (SA) was first isolated from *Streptomycetes* in 1964 by Chaiet and Wolf [10]. It is similar to avidin, a protein that is abundant in hen egg white. The biological function of avidin (and SA) is not understood but it is assumed that it inhibits bacterial growth by tightly binding the small molecule biotin (vitamin B_7 or vitamin H), which is an important cofactor for essential metabolic processes [11]. The interaction between avidin (or SA) and biotin is one of the strongest non-covalent interactions in nature [12].

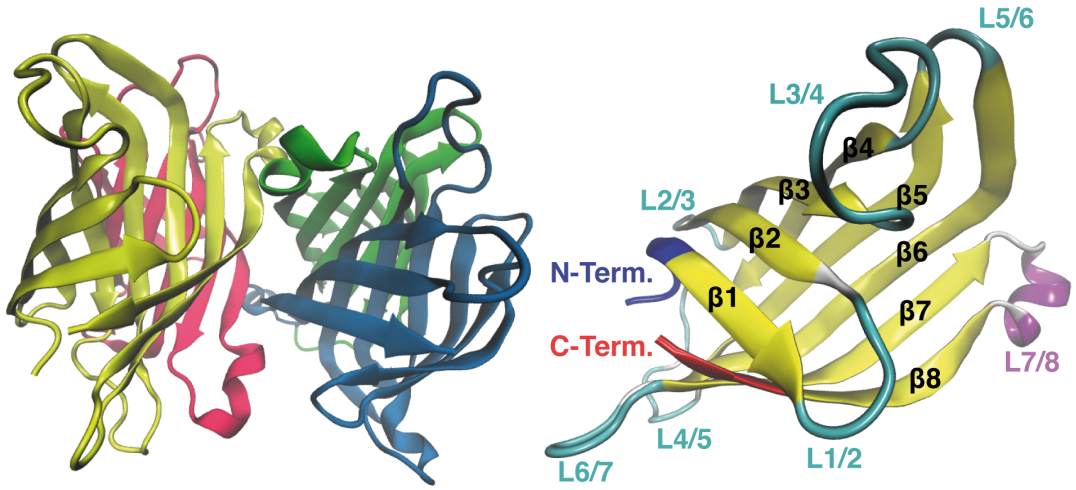
In the 1970s, first methods to biotinylate peptides and proteins were developed [13, 14]. Since then the extraordinarily strong interaction between biotin and avidin has evolved into a popular tool in biochemistry. In 1986, Argarana *et al.* identified the SA gene [15]. In contrast to avidin, SA is not glycosylated and its isoelectric point is at a near-neutral pH. Compared to avidin, SA is thus easier to handle for biotechnological applications. To date, many assays in nanotechnology, biochemistry and medicine rely on the SA/biotin interaction.

1.1 Structure of Streptavidin

The wildtype SA consists of 159 amino acids. In *E.coli*, it is however shortened by proteolytic digestion to the core SA subunit, which consists of amino acids 13-139 [17, 18]. Core SA has a total molecular weight of 52.8 kDa. In this thesis, recombinant core SA is used. Two groups independently solved the crystal structure of core SA [11, 19]. It is depicted in Figure 1.1a. Consisting of four identical subunits (shown in different colors), SA is a homotetramer. As two subunits form a tightly bound dimer and two dimers then form a less tightly bound tetramer, SA is also described as a dimer of dimers [20].

Each subunit consists of an eight-stranded anti-parallel β-barrel. It is illustrated in Figure 1.1b. β-strands are marked in yellow, the loops connecting them are highlighted in cyan. The loop connecting the seventh and eighth β-strand has an α-helical secondary structure and is colored in purple.

In the tetramer, the longer (10-13 residues) β-strands (β6-β8), which are closer to the C-terminus (highlighted in red), mediate the interactions with the neighboring subunits



(a) SA tetramer. **(b)** SA subunit.

Figure 1.1: Crystal structure of SA (PDB code: 1SWC [16]). (a) Four identical SA subunits (colored differently) comprise the SA tetramer. The red and yellow polypeptide chain, as well as the blue and the green one, form a tightly bound dimer by interactions mediated through the long β-strands. These dimers are then associated into the tetrameric structure. (b) An anti-parallel β-barrel forms the SA subunit. The eight β-strands are highlighted in yellow and numbered beginning at the N-terminus (blue) with β1 and ending at the C-terminus (red) with β8. The loops in between the β-strands are labeled accordingly, *e.g.* L1/2. The ends of the polypeptide chain are located in close proximity, the biotin binding site is situated at the opposite end of the β-barrel.

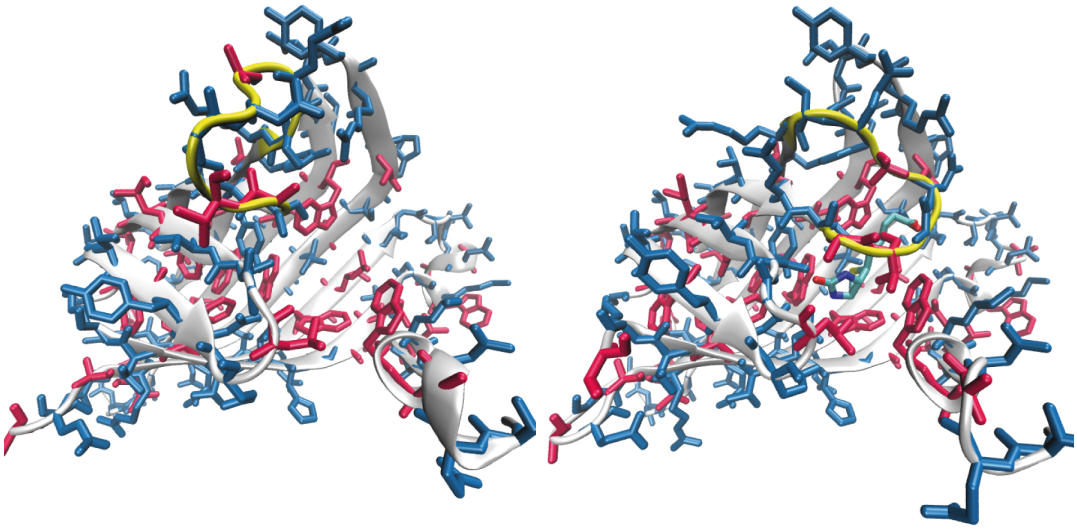
[19], while the shorter (5-7 residues) β-strands β1-β4, which are closer to the N-terminus (highlighted in blue), are outwardly oriented. In addition, interactions between the loops L4/5 and L7/8 stabilize the assembly of subunits into the tetrameric structure. One tryptophan, W120, loacted in the L7/8-loop reaches into the neighboring β-barrel [11]. The L3/4-loop is flexible in its orientation and can undergo a conformational change upon binding of biotin.

1.2 Binding of Biotin to Streptavidin

The binding of the small molecule biotin, which is also known as vitamin B_7 or vitamin H, to SA is one of the strongest non-covalent interactions known. Its dissociation constant K_D is in the femtomolar range (4×10^{-14} M) [21].

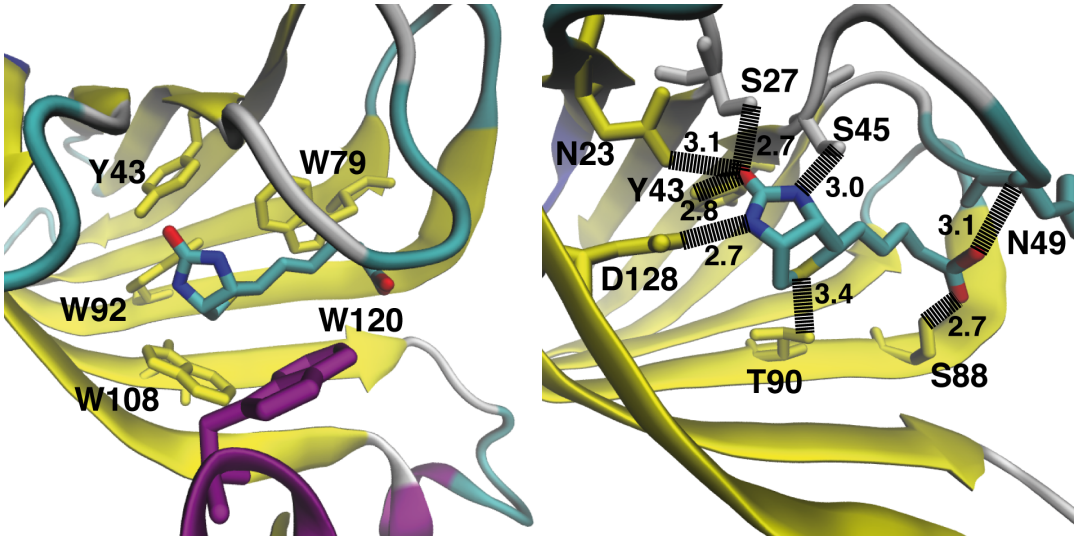
Every SA subunit can bind one biotin molecule. The wildtype SA tetramer can thus bind four biotin molecules in total. Hereby, each biotin molecule is fixed within a hydrophobic cavity inside one of SA's β-barrels. In Figure 1.2a, hydrophobic residues are highlighted in red, while polar residues are marked in blue. It is obvious that hydrophobic amino acid side-chains predominantly point into the β-barrel, while the polar ones are directed outwards.

Upon binding, SA undergoes a conformational change: The polypeptide loop L3/4, which connects β-strands β3 and β4, is closing over the biotin molecule like a lid [22]. The hydrophobic tryptophan residue W120 of a neighboring subunit reaches into the binding pocket and complements the L3/4-loop from the side [23]. In Figure 1.2b, the closed conformation is shown. The L3/4-loop is highlighted in yellow. Biotin (shown in cyan) is buried inside the



(a) SA in its apo form. **(b)** SA in its holo form.

Figure 1.2: Conformational change in SA upon binding of biotin (PDB code: 1SWC and 1SWE [16]). Polar residues are shown as blue sticks, while hydrophobic residues are shown in red. The L3/4-loop that closes over biotin is highlighted in yellow. Biotin is shown in cyan. The L7/8-loop from a neighboring subunit is also depicted, as it contains residue W120, which reaches into the β-barrel and complements the binding pocket.



(a) Steric confinement. **(b)** Hydrogen bond network.

Figure 1.3: Binding of biotin to SA (PDB code: 1SWE [16]). (a) Bulky side-chains confine biotin's position within the β-barrel. In particular, the opening of the β-barrel is blocked by tryptophan W120, which is provided from a neighboring subunit. (b) Eight direct hydrogen bonds hold biotin in place. The bond length is given in Ångström.

β-barrel. This is energetically favorable, as biotin displaces polar solvent molecules from the hydrophobic binding pocket.

Within the hydrophobic cavity, bulky aromatic amino acids confine biotin (*cf.* Figure 1.3a). In particular, the tryptophan residues W79, W92, W108 and W120 (from the neighboring subunit) contour the binding site [24–26].

In addition to these hydrophobic interactions, a cooperative network of hydrogen bonds that keeps biotin at its position has been identified [19, 27, 28]. In total, there are eight direct hydrogen bonds between SA and biotin. The SA residues that form these bonds are further stabilizes by a second layer of hydrogen bonds between different residues of the SA subunit. This network of hydrogen bonds and its influence on the binding energetics have been studied in detail. In particular, mutations of N23, S27, Y43, S45 and D128 increase the off-rate [27, 29, 30].

In summary, three effects contribute to the tight binding of biotin to SA: (i) the conformational change closing the L3/4-loop over biotin and absorbing it in the hydrophobic cavity; (ii) the steric confinement by bulky side-chains; (iii) the network of hydrogen bonds fixing biotin in its position.

1.3 Biotin and other Ligands

Biotin is an important cofactor for several enzymes. Biotin was discovered because of its tight binding to the egg-white protein avidin, which is similar to SA. It was first identified as a protection factor against a syndrome called ”egg white injury”, which results from high dietary intake of raw egg-white [31–33]. For humans, biotin deficiency results in hair loss and dermatitis [34]. This is why biotin is also called vitamin H (for German ”Haut und Haare”, skin and hair).

The structure of biotin was reported in 1956 [35]. It is depicted in Figure 1.4. When binding to SA, biotin’s five-ring structure is deeply buried in the hydrophobic pocket within SA’s β-barrel and a network of hydrogen bond is formed (*cf.* Figure 1.3b): The top oxygen atom forms three hydrogen bonds with side-chains of residues N23, S27 and Y43. In addition, hydrogen bonds between the secondary amines and residues S45 and D128 form. The sulfur atom forms a hydrogen bond with residue T90. The oxygen atoms of the carboxyl group can form hydrogen bonds with residues N49 and S88. The carboxyl group is usually used to link biotin to other biomolecules. In this case, at least one of the hydrogen bonds is lost.

Besides biotin, other ligands that are able to bind to SA exist. The biotin analogs desthiobiotin and iminobiotin are depicted in Figure 1.4. For desthiobiotin, the thioether part of the lower five-ring is missing, *i.e.* the hydrogen bond between the sulfur and T90 cannot be formed. For iminobiotin, the top oxygen is replaced by an amine group. The protonation of this amine group is dependent on pH. By this, also the occurrence of hydrogen bonds with N23, S27 and Y43 is altered and the binding affinity becomes pH-dependent [36, 37]. Besides biotin analogs, a number of peptide tags can also bind to SA [38–40].

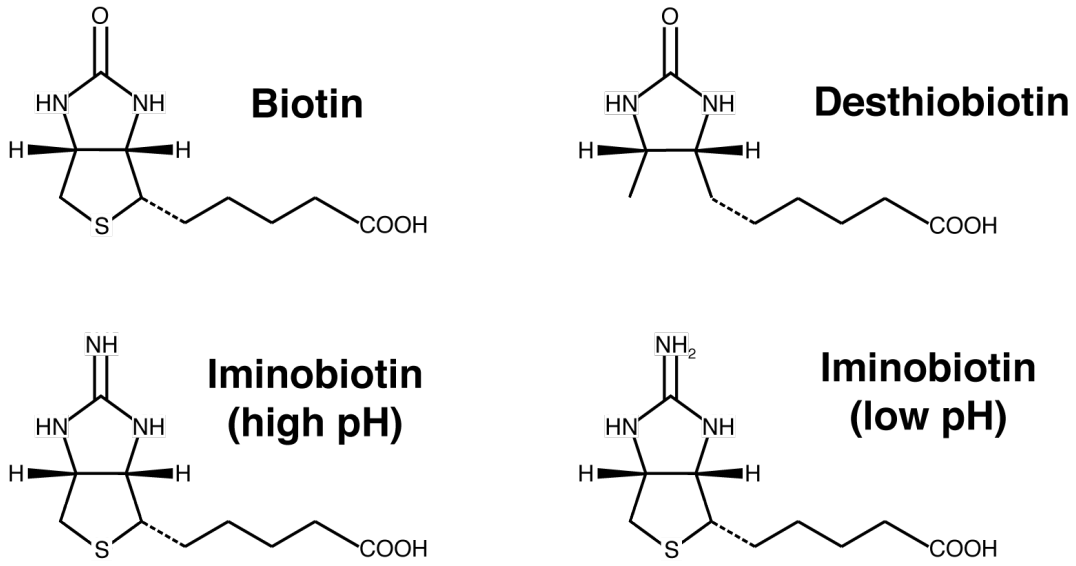


Figure 1.4: Structure of biotin and several biotin analogs. All can bind to SA.

1.4 Biomolecular Engineering of Streptavidin

With its outstanding high affinity, the binding of biotin to SA has become a paradigm for receptor-ligand interactions [12]. Countless mutation studies on the SA/biotin system have been performed to understand the origins of the extraordinarily high affinity of the binding [24, 25, 41], to improve or alter binding properties [1, 42–44] and to develop engineered versions of SA that are of value for commercial applications [45–47]. Over the years, the system has evolved into a widely used tool in biochemistry, nanotechnology and medicine.

Listing over one-hundred mutations and their effects on the binding affinity, a review paper by Laitinen *et al.* gives a good overview on mutation studies on SA and avidin [48]. Exchanging only a few amino acids, the properties of the binding can be drastically altered. In the following, a few exemplary engineered version of SA are described.

1.4.1 Strep-Tactin

Strep-Tactin (ST) is a modified version of SA developed by Voss and Skerra [46]. It is optimized to bind to the eight amino acid sequence Strep-tag II. Nowadays, it serves as a biotechnological tool for protein purification [49] Proteins to which a Strep-Tag II is fused are loaded onto a ST column. While these proteins bind to the column, other proteins and contaminations are washed out from the column. Eventually, biotin, which has an higher affinity for ST than Strep-Tag II, is used to elute the protein of interest from the column. As Strep-tag II can be genetically encoded, it is also a popular choice for anchoring molecules in SMFS [9, 50, 51].

ST and SA differ by only three amino acids: Residues 44-47, reading ESAV for SA, are mutated to VTAR in ST. All mutations are situated within the flexible L3/4-loop, which ranges from residues 44 to 53. In SA, this loop closes over biotin upon binding. In ST, the mutations cause this loop to stay in its open conformation resulting in higher affinity for the Strep-Tag II peptide [52].

1.4.2 Traptavidin

Traptavidin (TA) is a modified version of SA developed by Chivers *et al.* It has increased thermostability and binds biotin with a tenfold lower off-rate. Therefore, it is of interest for the immobilization of molecules in assays requiring stable long-term measurements or high temperature. The on-rate for biotin is yet also lowered by a factor of two [53].

The only difference to SA are two mutations in the L3/4-loop, namely S52G and R53D. Crystallographic data suggest that these mutations make the L3/4-loop less flexible and fix it in the closed conformation, even in the absence of biotin (apo-form) [12]. This might explain the lower on- and off-rates of biotin.

1.4.3 Non-Functional Streptavidin Subunit

Howarth *et al.* developed a SA subunit with (for most applications) negligible affinity ($K_D = 1.2\,\mu M$) for biotin [1]. It was created by further developing a SA double mutant [37] into a SA triple mutant: N23A, S27D, S45A. While the mutation S45D is located within the L3/4-loop, the other two are located at the ends of the L1/2-loop, *i.e.* towards the N-terminus of the subunit. The mutations of N23A and S45A impede the formation of hydrogen bonds that usually emerge between SA and biotin [1]. The mutation S27D creates steric hindrance with residue A46 to prevent the L3/4-loop from closing [54]. In addition, crystallographic data suggest that a hydrogen bond network involving S27A, S45A, A46, N49 and S52 stabilizes the open conformation of the L3/4-loop, hindering the loop closure over biotin [54]. Thereby, the binding affinity for biotin is drastically reduced. Combining non-functional (dead) SA subunits (DA) with functional SA subunits, SA variants of distinct valencies can be created [1, 9, 55].

1.4.4 Monomeric Streptavidin

SA is a tetrameric protein and can bind four ligand molecules. For certain applications, a defined 1:1-stoichiometry is yet desirable [1, 2, 9]. To achieve a 1:1-stoichiometry, monomeric version of SA (mcSA) have been developed [44, 56]. As single SA subunits are not stable on their own, special stable stand-alone SA domains have been created using biomolecular engineering techniques [57, 58]. For all versions of mcSA, SA's high affinity for biotin is yet significantly lowered [44]. Details are given in Manuscript M5.

1.4.5 Monovalent Streptavidin

Combining three dead and one functional subunits a heterotetrameric monovalent SA (mSA) can be created [1]. For this, the different subunits are expressed separately, denatured, mixed in a 3:1-ratio and finally refolded. By size-exclusion, mSA is then purified to obtain a version of SA with 1:1-stoichiometry. In a similar manner, non-functional, divalent, trivalent SA can be created. In contrast to mcSA, these versions still have femtomolar affinity for biotin. For this thesis, a slightly different protocol is used to create SA of distinct valencies: One subunit is outfitted with a polyhistidine tag to allow for purification by affinity chromatography (*cf.* Section 2.1.1).

1.5 Manuscript M1: Force Spectroscopy on Streptavidin

In 1994, the first SMFS experiments of the SA/biotin interaction strength were performed [6–8]. Over the last 25 years, a lot of effort has been put into investigating the mechanical stability of this receptor-ligand interaction with its extraordinarily high affinity by AFM-based SMFS. In Manuscript M1, the development of the field of AFM-based SMFS over this time is reviewed, focusing on important trends and describing their influence on research performed on the SA/biotin interaction. In particular, improvements in immobilization strategy, the development of the theoretical framework and the challenges involved in measuring true single molecule interactions are discussed. This review also features new data comparing the attachment strategy used for the first SMFS measurements of the SA/biotin interaction [8] with the site-specific tethering employed recently [3]. Showing how tethering geometry influences rupture forces, this work reconciles seemingly contradictory results reported in various studies over the years. At the end of this review, an outlook on how the field of SMFS might develop in the next decade is provided. In the context of this thesis, this review gives an overview of SMFS on SA/biotin. Technical details, however, will be discussed in the following chapters.

25 Years of Force Spectroscopy on the Streptavidin/Biotin interaction

by

Steffen M. Sedlak, Lukas F. Milles and Hermann E. Gaub

in preparation

25 Years of Force Spectroscopy on the Streptavidin/Biotin Interaction

Steffen M. Sedlak,[1] Lukas F. Milles,[1] and Hermann E. Gaub[1]

[1]Department of Physics and Center for NanoScience, LMU Munich, Germany

Abstract

With the measurement of forces between individual receptor ligand pairs with the atomic force microscope (AFM), reported in 1994 (1-3), the field of single-molecule force spectroscopy (SMFS) emerged. Since then, instrumentation and experimental techniques were improved, a theoretical framework was put forward, sophisticated data analysis methods were developed, and complementary techniques, such as steered molecular dynamics simulations, evolved. The first interaction examined by means of SMFS was the binding of the small molecule biotin to the homotetrameric protein streptavidin (SA). Reviewing SMFS studies on the SA/biotin interaction conducted by different groups over the last 25 years, we illustrate important developments in the field of SMFS, identify present challenges and point out perspectives for the future.

Introduction

With femtomolar affinity, the interaction between the small molecule biotin (vitamin H) and the homotetrameric protein streptavidin (SA) is (regarding affinity) one of the strongest non-covalent interactions. The SA tetramer consists of four β-barrel-shaped subunits. Each can bind a biotin molecule. Several factors contribute to SA's high affinity for biotin: once bound into a hydrophobic patch inside a β-barrel subunit, a network of hydrogen bond is formed around it. Additionally, a binding-induced conformational change in SA closes a flexible peptide over the binding pocket. As biotin is straightforward to conjugate to macromolecules or nanoparticles, SA is an established, key molecule in nanotechnology widely used for labeling, detection or immobilization. Nowadays, SA/biotin has become a standard receptor-ligand system for single-molecule force spectroscopy (SMFS) experiments.

Its wide, also commercial, availability combined with its extremely high affinity and considerable mechanostability, made the SA/biotin interaction an attractive target to fundamentally understand its properties, not only in SMFS experiments. Indeed, it is a commonly investigated proof-of-principle system. It was the first receptor-ligand systems to be studied by SMFS and as the field of SMFS evolved, has been revisited many times to establish methodological improvements (in experimental techniques, instrumentation, data analysis or complementary techniques). A variety of experimental and theoretical methods have been used by a number of labs to investigate SA/biotin mechanics. Over the last 25 years a range of – partly contradictory – results has been reported in the literature. Suspected reasons for the variance of force ranges proposed range from: the history of the bond (4), the underlying energy landscape (5), the influence of linker molecules (6), SA's tetravalency (7), the stability of SA itself (8), and the force-loading direction (9). Here, we do not provide a list of reported unbinding forces, see previous reviews (7, 10). Instead, given these proposals for a mechanistic understanding of SA/biotin, we focus on the developments in the field and put related functional studies of the SA/biotin interaction into context.

Pioneering work: The first measurements of forces between individual receptor-ligand pairs

In 1994, three landmark studies published the first measurements of forces between individual receptor-ligand pairs using the atomic force microscope (AFM). Lee *et al.* employed biotinylated bovine serum albumin (BBSA) to coat mica surfaces and glass microspheres epoxy-glued to an AFM cantilever tip (1). The surfaces were subsequently incubated with SA. Bringing the glass microsphere and the mica surface into contact, interaction forces between biotin and SA upon cantilever retraction were measured. Lee stated: "The magnitude and distribution of the observed adhesive forces suggest they result from individual streptavidin-biotin interactions".

Florin *et al.* and Moy *et al.* directly functionalized the AFM cantilever tip using BBSA, which was then incubated with avidin or SA, respectively (2, 3). Interaction forces were measured between the functionalized tip probing biotinylated agarose beads on the surface. Crucially, the adhesive force was determined from the final rupture event's force peak, and not the maximum force reached in initial adhesion due to multiple attachments of the tip. Using an autocorrelation function on the observed rupture force histograms resulted in the "elementary force quantum of 160±20 pN" of the interaction. Thus, the dissociation forces between individual ligand-receptor pairs could be determined.

Here, forces were still chosen as a representative for the binding enthalpy, an intuitive, yet overly simplistic view. Energy landscapes of protein mechanics and those of binding thermodynamics are not directly related.

Improvements in immobilization strategy: From non-specific sticking to well-defined covalent chemistry

To measure the interaction forces between single ligand-receptor pairs in AFM-based force spectroscopy, the molecules have to be specifically immobilized on cantilever tip and sample surface. Stepwise improvements to non-specific attachment employed in early work have been introduced over the years. Very different immobilization strategies have been developed and validated over the years. An overview is provided in Table 1.

The first force spectroscopy measurements on the SA/biotin interaction used biotinylated bovine serum albumin (BBSA), which was non-specifically adsorbed onto both tip and sample surface. Either BBSA covered surface could then be incubated with SA. Due to its tetravalency, SA attached to the BBSA's biotins through any number of its four binding pockets, leaving the others vacant. Once tip and surface are brought into contact, the vacant binding sites on SA may bind to the free BBSA's biotins presented on tip or sample surface, respectively. With the bonds under investigation formed, the cantilever was retracted from the surface at constant velocity. The interaction force could be measured, as the SA/biotin bonds (either on the tip or on the surface) ruptured. With the resulting multitude in potential force-loading geometries depicted in Figure 1A, the unbinding events could not be unambiguously assigned to a single geometry.

The passive adsorption of BBSA on the surface presents an additional uncertainty: In some cases not the SA/biotin bond, but the interaction between BBSA and the surface might be broken, yet interpreted as a biotin/SA event. First improvements in attachment strategy, thus aimed to establish a much stronger, covalent link between the biotin molecules and the sample or tip surface, respectively.

In 1998, a covalent functionalization of biotin to the cantilever apex was realized: S. Wong *et al.* performed an impressive set of experiments using carbon nanotube tips (11). The attachment of individual biotin molecules at the end of a single nanotube enabled direct measurement of single-molecule interactions, eliminating the need for an autocorrelation function to correct for multiple bonds breaking. To date, the use of a well-defined covalent chemistry is – regrettably – not a routinely enforced standard in the field of force spectroscopy. In this way, S. Wong *et al.* were ahead of their respective time. Yet, the use of carbon nanotube tips was experimentally too laborious to find widespread application.

In 1999, Merkel *et al.* employed flexible polyethylene glycol (PEG) polymer linkers with N-hydroxy succinimide (NHS) on the one end and biotin on the other end to prepare a biomembrane force probe (12). NHS forms a covalent bond to amine groups on beads or surfaces functionalized with amino-silane.

The use of linker molecules, such as PEG, is advantageous for passivation of the sample surfaces reducing undesirable interactions between cantilever tip and sample surface. The behavior of single polymer chains under force has been intensively studied by experiments and theoretically modeled (13, 14). Single polymer chains are modeled well as entropic springs (*e.g.* worm-like (15), freely-jointed (16), or freely-rotating chain (17)). These models can be used to extract the contour length of an unfolding of protein domains in AFM-based

SMFS experiments (18). Nowadays, the use of polymer models to transform force-extension traces into contour length space is a standard procedure to determine the size of contour length increments upon unfolding of protein domains (19).

All immobilization strategies described so far employed the ligand biotin itself to anchor SA (Figure 1A; Table 1A), essentially force-loading SA from two or more binding pockets. A second approach of probing the SA/biotin interaction by force spectroscopy is to directly immobilize SA to a surface or cantilever tip (Table 1B,C). The resulting force loading geometry is depicted in Figure 1B.

In 1996, Allen *et al*. implemented this geometry using SA-coated polystyrene microtitre wells as sample surface, improving the experimental setup by covalently, yet still non-specifically, attaching SA to the surface (20). In the following years, various studies developed different attachment strategies: In 1999, J. Wong *et al.* performed measurements of the SA/biotin interaction using a phospholipid bilayer to attach SA and biotinylated PEG (21). In 2002, Stevens *et al.* developed a dextran-based immobilization strategy for biomolecular force measurements and tested it on the SA/biotin system (22). In 2006, de Odrowaz Piramowicz *et al.* used an immobilization protocol in which glutaraldehyde crosslinked SA to the amine-functionalized cantilever tip functionalized, while biotin was attached to a glass surface coated with poly-L-lysine (23). Although successful, these strategies have not been adopted widely.

In 2009 and 2010, Taninaka *et al.* and Teulon *et al.* independently reconsidered dynamic (*i.e.* force loading rate varying) force spectroscopy on SA/biotin (7, 24-26). Both employed coupling of thiol groups to gold surfaces to immobilize short carbon chains as molecular linkers for biotin. Both used SA's reactive amines for coupling: Teulon *et al.* used EDC to establish peptide bonds with primary amines in SA. Taninaka *et al.* established thioether bonds to commercially available maleimide-SA, prepared by adding SMCC (a short NHS-Maleimide crosslinker) to SA. Although these coupling strategies are covalent, their force-loading geometry is ambiguous, as SA exhibits more than one accessible amine.

To overcome this restriction, protein engineering techniques defining unambiguous tethering geometries had to be employed. In 2011, Kim *et al.* expressed SA subunits genetically fused to His-tagged Titin I27 Ig-domains, which assembled into tetrameric SA decorated with a tail of Ig folds (8). They studied the mechanical stability of the SA tetramer itself (Figure 1D; Table 1D): Immobilizing these constructs on a His-tag-binding Ni-NTA-functionalized surface while probing with a Ni-NTA-functionalized cantilever. The fusion proteins with the Ig-fold tail defined a site-specific loading of the SA tetramer (not of a biotin/SA bond), because the force could only be applied from the N-termini. However and crucially, Kim *et al.* realized that different pulling geometries across the SA subunit interfaces were possible.

In 2015, Baumann *et al.* developed the current state-of-the-art in SA immobilization in SMFS (Figure 1C; Table 1C) (27). Employing a method developed by Howarth *et al.* (28), they were able to create a tetrameric but monovalent version of Streptactin (ST). ST is highly homologous to SA, but binds the short StrepTag II peptide. The monovalent ST developed contains only one functional, *i.e.* ligand binding, subunit. The remaining three subunits have been mutated to not bind the ligand. In addition, they introduced a unique cysteine in the single functional subunit to allow for site-specific immobilization of the molecule onto a maleimide-functionalized surface. Thus, the SA/biotin interaction could for the first time be tethered in a well-defined and unambiguous force loading geometry (29).

Development of a theoretical framework and the onset of dynamic force spectroscopy

Building on Kramer's theory for reaction kinetics (30) and Bell's model for reversible bonds between molecules (31), Evans and Ritchie as well as Izrailev *et al.* developed a theoretical framework for force measurements on ligand-receptor pairs in 1997 (32, 33). Originally developed to describe the biotin/avidin interaction as force-induced dissociation across a single energy barrier, today it is routinely used in SMFS and referred to as Bell-Evans theory.

Evans and Ritchie concluded "that strengths of molecular linkages must be tested under controlled [force] loading over a wide range of rates". In 1999, the first of such a dynamic force spectrum (most probable rupture force against the logarithm of force loading rate) of the SA/biotin interaction was measured: Merkel *et al.* employed a biomembrane force probe (BFP) to explore the energy landscapes of the SA/biotin interaction and avidin/biotin interaction, respectively (12). Varying the retraction velocity and the force constant of the BFP, rupture forces could be measured over a wide range of force loading rates to obtain the log-linear dependency of the rupture force on the force-loading rate. A prominent feature emerged: a "kink" in slopes of rupture forces in the dynamic force spectrum, interpreted as two energy barriers to be overcome when breaking the SA/biotin bond.

In the AFM, the force constant, *i.e.* the spring constant of the cantilever, cannot be changed. Instead, to probe with different force loading rate, the retraction velocity of the cantilever is varied. In 2000, Yuan *et al.* determined the dynamic force spectrum of the SA/biotin interaction using the AFM (34). Lo *et al.* independently reproduced these results in the following year (35).

In 2007, Rico *et al.* elucidated the influence of temperature on the SA/biotin rupture forces (5). Increasing the temperature from 17°C to 37°C, lowered the rupture forces, in agreement with Bell-Evans theory.

Over the years extended models for molecular unbinding under force were proposed. These models solved some issues of Bell-Evans theory, for example the assumption of a constant loading rate. Interested readers may refer to additional literature (36-41).

The challenge of measuring single molecule interactions – a constant companion

Ensuring that only a single receptor-ligand interaction is measured in an SMFS experiment is a methodological challenge, that accompanied the research on the SA/biotin interaction over the years (42). In 1994, an autocorrelation function was used to extract a "force quantum" from the histogram of rupture forces, to arrive at the strength of a single bond from data containing multiple binding events.

The first approach to tackle the problem of possible multiple interactions was to block most of the SA's binding sites by adding free biotin (2, 3, 20) or to block most of the biotin molecules on the surfaces by adding SA (12). Comparing the number successful tethers with the number of total approach-retraction cycles conducted, Poisson statistics were used to estimate the likelihood of forming a single bond.

In 1998, S. Wong *et al.* used an elegant approach (11): By attaching a single carbon nanotube modified with again just a single reactive site to the cantilever apex, the number of ligands on tip was virtually limited to one.

In 1999, Lo *et al.* introduced a Poisson statistical analysis method to extract rupture forces of an individual receptor ligand interaction from data acquired when force-loading multiple receptor ligand pairs in parallel (43). With the rise of dynamic force spectroscopy, they applied this method once more in 2001 to data recorded using different retraction velocities (35). In 2008, Erdmann *et al.* performed BFP experiments on the biotin/SA system deliberately aiming for multiple interactions in force traces to develop a model that can extract individual rupture forces from such data (44).

Teulon *et al.* reviewed previous SMFS studies on SA/biotin in 2011 with regard to discrepancies in measured rupture forces. They attributed these differences in previous studies to multiple bonds measured in parallel (7). Their own data, measured in a different force-loading geometry than the work referenced, substantiated this argument.

In the same year, Kim *et al.* performed landmark experiments (8): Instead of probing the biotin/SA interaction strength, SA tetramer stability was probed by determining their unfolding force. SA was force-loaded from the N-termini of two different subunits by AFM SMFS. The authors noted "that the strength of the dimer-dimer interface in streptavidin is comparable or even weaker (at the examined stretching rates) than the biotin-streptavidin unbinding force that was reported to range from 100 pN to 450 pN when measured by AFM". To identify single-molecule interactions and apply force to the construct, I27 domains of titin and staphylococcal nuclease modules had been genetically fused to the N-terminus of the SA subunits. The distinct unfolding patterns of these domains served as fingerprints and allowed to discern specifically from non-specifically probed interactions.

For mechanical ligand dissociation from an avidin-like protein, Baumann *et al.* were the first to employ a fingerprint domain to identify of single-molecule interactions in 2015 (27). For the biotin/SA interaction, a fingerprint was not introduced until 2017 (29).

Although single molecule interactions can be reliably identified by the unfolding pattern of a fingerprint domain SA's tetravalency inherently causes multiple possible force-loading geometries. Pulling biotin from different SA subunits results in different force propagation pathways through the SA tetramer and thus in different rupture forces. Baumann *et al.* also addressed this issue (for the interaction of Streptactin and Strep-Tag II) in 2015 (27). Building on a tetrameric but monovalent version of SA that Howarth *et al.* had published in 2006 (28), Baumann *et al.* introduced a single unique cysteine into the only functional subunit of monoSA allowing for site-specific immobilization of Streptactin by maleimide chemistry, and only a single possible force-loading geometry. Sedlak *et al.* continued this strategy and published the first monodisperse, *i.e.* single defined geometry, SMFS data of SA/biotin in 2017 (29). In 2018, Sedlak *et al.* showed that biotin/SA rupture forces are dependent on the force loading geometry: N- or C-terminal tethering of SA result in rupture forces differing by around a factor of two. In 2019, Erlich *et al.* used this force hierarchy to arrange single molecules into a predefined nanoscale pattern by AFM-based single molecule cut & paste (45).

From data interpretation to an understanding at the molecular level

Receptor-ligand rupture forces obtained from a force spectroscopy experiment are not sufficient to elucidate a molecular mechanism with atomic detail. To interpret such forces complementary techniques providing additional details are needed. Besides the aforementioned dynamic force spectroscopy measurements, providing insights into the

energy landscape of a system, varied conditions, modifications of the ligand molecule or mutants in the receptor molecules can contribute to a better understanding.

Already in 1994, Moy *et al.* not only investigated the interaction between SA and its ligand biotin, but also between SA and desthiobiotin, a biotin variant lacking the sulfur atom in the tetrahydrothiophene ring. In the following year, Chilkoti *et al.* used AFM-based SMFS to study variants of SA, in which tryptophan residues were mutated to phenylalanine or alanine (46). Further AFM-based SMFS experiments with mutated versions of SA were performed by Yuan *et al.* (34), Chivers *et al.* (47) and Baumann *et al.* (27). The latter two employed Traptavidin and Streptactin, respectively. These SA variants contain mutations within the flexible peptide loop that closes over the binding pocket for improved off-rate and peptide tag binding, respectively. Both are highly relevant for biotechnological applications. To obtain biotin/SA with a 1:1-stoichiometry, a monomeric version of SA was designed (48-52). Bauer *et al*. investigated its mechanics by AFM-based SMFS in 2018 (53).

The development of tetrameric yet monovalent (one funcitonal, three non-functional subunits) SA by Howarth *et al.* in 2006 (28) and site-specific immobilization strategies for monovalent Streptactin by Baumann *et al.* in 2015 (27), enabled Sedlak *et al.* to study the unbinding of biotin from SA with unambiguous immobilization geometry through a cysteine residue (29). Through moving the cysteine to either N- or C-terminus the biotin/SA interaction could be probed in different force-loading geometries, resulting in twofold differing rupture forces. Complementary simulations helped to shed light on the underlying molecular mechanism.

Steered molecular dynamics (SMD) simulations are an ideal complement to AFM-based SMFS experiments. The first SMD simulation of the SA/biotin complex were conducted in 1996 by Grubmüller *et al.* (54). Back then, only a single SA subunit could be simulated. Izrailev *et al.* performed the first simulation of the unbinding of biotin from the full (avidin) in 1997 (33). Restrictions in computational power and AFM instrumentation resulted in a large difference in force loading, as simulation runtimes were too short to reach experimentally accessible pulling velocities. For the biotin/SA system only this year, the gap was finally bridged combining high-speed AFM measurements and comparably slow-speed SMD simulations (55, 56).

Comparability – force transducer stiffness and calibration issues

Other factors can influence the unbinding forces reported in AFM-based SMFS studies and complicate the comparability between the numerous experiments performed over the last 25 years. These certainly are partially responsible for some disagreements in the literature.

In 2006, Thormann *et al.* tested which influence the linker used for tethering had on dynamic force spectra with SA/biotin as a model system (57). Their results suggest that unbinding forces are influenced by the type and length of linker molecules. For long PEG linkers, lower unbinding forces were observed. Additionally, Thormann *et al.* reported that the previously reported kink in the dynamic force spectrum could not be observed with long PEG linkers.

In 2008, Walton *et al.* performed a combination of AFM-based SMFS experiments and steered molecular dynamics simulations to probe the influence of force transducer (or force probe, such as an AFM cantilever) stiffness on unbinding forces and binding kinetics (6), using the SA/biotin interaction as a model. The study suggests that stiffer cantilevers result in higher unbinding forces. Caution must be applied in comparing the data directly, as the force-loading rate is calculated as a product of cantilever stiffness and retraction velocity (and not

extracted from the slope of the force extension trace). Fundamentally, it cannot be denied that “the effective stiffness of the force transducer *k* can significantly perturb the energy landscape and the apparent unbinding force of the complex for sufficiently stiff force transducers” (6). An exact analytical description of the effect of the force probe properties on measured forces remains a key challenge that the field must address.

The correct calibration of cantilever spring constants is vital for AFM-based SMFS. Incorrect values for the spring constants are directly translated into incorrect force values. The two most common methods to calibrate the spring constant of the AFM-cantilever were proposed by Hutter and Bechhoefer in 1993 (58) (further developed in 1995 by Butt and Jaschke (59)), commonly called the thermal method, and by Sader *et al.* in 1999 (60), respectively. Both are non-destructive and can be performed *in situ*. In 2011, eight labs performed a round robin experiment to compare the calibration of the cantilever’s spring constant on ten different AFM instruments (61), reporting uncertainty levels of about 7-15 % depending on the applied method. A more recent publication on cantilever calibration methods also reports a stiffness deviation of at least 7.7 % (62).

To circumvent calibration issues, Baumann *et al.* introduced a fingerprint protein domain in series with the receptor-ligand system (in this case the StrepTactin/Strep-Tag II system, which is closely related to SA/biotin) in 2015 (27). This fingerprint domain unfolds prior to the rupture of the complex, thereby giving rise to a distinct unfolding pattern. The unfolding force can be used as an internal force reference allowing to compare rupture forces observed in the same measurements for different protein complexes or buffer conditions relative to the unfolding force of the fingerprint domain. Sedlak *et al.* adopted this method for measurements of the rupture forces of the SA/biotin complex (29, 63).

In the same year, Verdorfer *et al.* introduced a method to directly compare forces measured for different protein domains by immobilizing them spatially separated on a single sample surface and interrogating them with a single AFM cantilever (64). Thus, issues of the cantilever calibration are irrelevant and absolute forces may be compared. Sedlak *et al.* used this method to compare monovalent SA tethered either from N- or C-terminus in 2018 (9).

Here, we compared these two well-defined unambiguous force-loading geometries of monovalent SA (Figure 1C) with the tethering geometry reminiscent of the very first SMFS measurements on biotin/SA in 1994. In these, both cantilever tip and surface where functionalized with biotin and SA was only added afterwards (Figure 1A). To circumvent aforementioned issues of cantilever calibration, and to compare absolute forces we immobilized the three different SA variants, and anchoring geometries at different areas on a single surface and probed all with the same cantilever. Clearly, the rupture forces for the different immobilization strategies and thus for the different force-loading geometries drastically differ (Figure 2). Details on the experimental procedure and a more detailed interpretation are provided in the Supplementary Information. In turn, we believe that this dependence of the unbinding forces on force-loading geometry is the principal reason behind the wide range of rupture forces reported for the SA/biotin system. Without precise control of the geometry of force application, multiple force propagation pathways are probed and their rupture forces superimposed will produce very different force spectra although the same interaction is probed.

Outlook

In this review, we provided an overview about research performed by (mainly AFM-based) force spectroscopy on the biotin/SA system in the last 25 years. We illustrated the development of the field of force spectroscopy using measurements of the biotin/SA as reference. We could show how certain problems were solved as experimental techniques, instrumentation and theoretical framework were further developed. Certain challenges still have to be tackled.

Experimentally, a widespread use of site-specific covalent chemistry, SA of defined valencies and suitable fingerprint domains is desirable. Beyond that, Ott *et al.* recently extended the SMFS toolbox by introducing elastin-like polypeptides (ELP) as molecular linkers more suited to probe protein systems (58). Furthermore, the role of adjacent linkers tethering a biotin interacting with SA, especially on SA's lid region, is worthy of investigation.

For the theoretical framework, we identified the challenge that to date, a comprehensive theory of the coupling of the cantilever properties and molecular linkers into the system of interest is still missing. Some effects, like the effect of cantilever stiffness or linker length on unbinding forces, have been studied but a complete understanding of the interplay between components used to determine mechanical properties of a system, that inadvertently influences said measurement, remains a challenge for the field in general.

Regarding complementary SMD simulations, further investigations into the consistency between SMD simulations and experiments is required, also in regard to biotin's force field. Using site-specific tethering geometries, the experimental force-loading geometry can be exactly reproduced in SMD simulations (9). With the increase in computational resources, SMD simulations can nowadays already be performed at the same speed as the fastest AFM-based SMFS experiments (56). We anticipate that the currently narrow overlap of timescales between *in silico* and *in vitro* will widen considerably over the next years, and such overlaps may become routine to SMFS investigations.

For the AFM instrumentation, the wider use of force-ramp setup might be advantageous. In force-ramp mode, measurements are performed at a constant force-loading rate, more closely to the Bell-Evans theory. Furthermore, the combination of AFM-based SMFS with single-molecule imaging techniques, down to the superresolution, will provide novel approaches for single-molecule experiments in the field of mechanobiology. In this context, the implementation of nanostructured surfaces, such as zero-mode waveguides, into AFM-based SMFS is a relevant technological advancement. Another improvement of the AFM instrument has been recently demonstrated by Edwards *et al.* (65): With the use of tailored AFM cantilevers, the temporal resolution of AFM-based force spectroscopy can be significantly improved. Better temporal resolution will allow for observation of substeps in unbinding or unfolding pathways. Combined with SMD simulations, this will help to deepen our understanding of the action of mechanical forces on molecular systems.

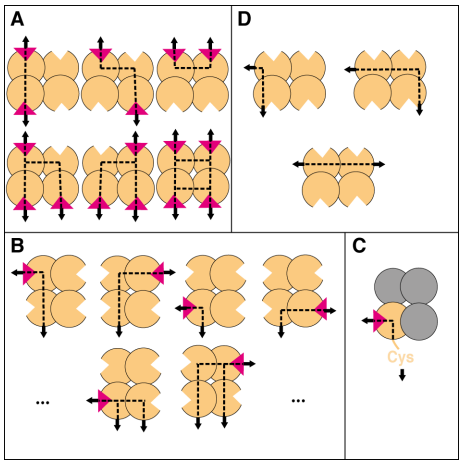


Figure 1. Immobilization strategies result in different force loading geometries. Biotin (pink) is pulled out of functional SA subunits (light orange). Black arrows indicate potential anchoring points. Black dotted lines schematically indicate potential force propagation pathways. (A) Immobilizing biotin on both cantilever tip and sample surface and then incubating with SA, results in six different geometries, because SA can be anchored by one, two or three biotin molecules. (B) Direct immobilization of SA on the surface results in four force loading geometries, when SA is anchored by one distinct point. Otherwise, if there are multiple anchoring sites, numerous force loading geometries might be probed in a single measurement. (C) Site specific immobilization of a monovalent SA results in an unambiguous, well-defined force loading geometry. (D) Pulling on different SA subunits, the stability of the SA tetramer itself is probed.

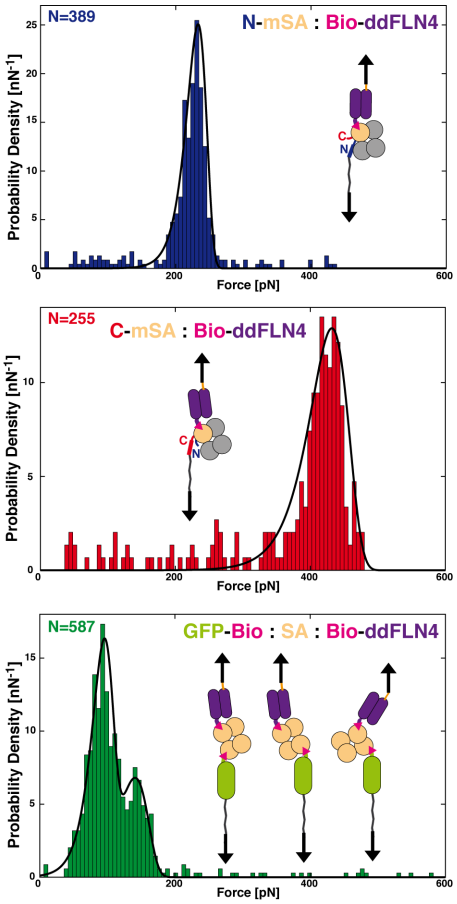


Figure 2. Force-loading geometries results in different unbinding forces. Rupture force histograms for different tethering geometries are shown. All data were recorded in a single measurement with the same cantilever. Different tethering geometries were applied at different spots on the same sample surface. Data obtained for direct tethering of monovalent SA at its N- or C-terminus is shown is shown in blue or red, respectively. Data for tethering of SA by biotin is shown in green. As green histogram overlays with unfolding forces for the GFP and ddFLN4 fingerprint domains, force-extension traces were not first sorted for the fingerprint pattern but the raw data were plotted. The relevant point still becomes clear: Different immobilization strategies result in different force-loading geometries and thus in different unbinding forces.

Table 1. Development of Immobilization Strategies.

A) Immobilization of biotin on both sample surface and cantilever tip

Year	First Author	Immobilization on sample surface	Immobilization on cantilever tip
1994 1995	Lee (1) Chilkoti (46)	Mica : BBSA : SA	Si_3N_4 : Glass bead : BBSA
1994 1994 2000	Moy (3) Florin (2) Yuan (34)	BBSA agarose bead	Si_3N_4 : BBSA : SA
1998	Wong, S. (11)	Mica : BBSA : SA	Au : acrylic adhesive : carbon nanotube (with COOH) + EDC + Biotinamido-$(CH_2)_5$- NH_2
1999 2001	Lo (43) Lo (35)	glass : BBSA	Si_3N_4 : BBSA : SA
1999	Wong, J. (21)	a) mica : biotinyl. phospholipid bilayer b) mica : phos.lipid bilayer + PEG-Bio	mica : biotinylated phospholipid bilayer : SA
1999 2008	Merkel* (12) Erdmann* (44)	glass + NH_2-silane + NHS-PEG-Biotin : SA	glass + NH_2-silane + NHS-PEG-Biotin : SA
2006	Thormann (57)	a) glass : BBSA : SA b) glass + NH_2-silane + NHS-PEG-Biotin : SA	Si_3N_4 : BBSA
2007 2010	Rico (5) Chivers (47)	glass : SA: biotinylated agarose beads	Si_3N_4 : BBSA : SA
2008	Walton (6)	Mica + NH_2-silane +EDC + BBSA : SA	Si_3N_4 + NH_2-silane +EDC + BBSA
2010	Taninaka (25)	Au : HS-$(CH_2)_8$-NH_2 + NHS-PEG-Biotin : SA	Au : HS-$(CH_2)_8$-NH_2 + NHS-PEG-Biotin
2019	This work	glass + NH_2-silane + NHS-PEG-MAL + GFP-Bio : SA	Si_3N_4 + NH_2-silane + NHS-PEG-MAL + CoA + ybbR-SdrG : Fgβ-ddFLN4-Biotin

*biomolecular force probe measurements

B) Non-specific covalent immobilization of SA

Year	Author	Immobilization on sample surface	Immobilization on cantilever tip
1996	Allen (20)	SA-coated microtitre wells	Si_3N_4 : BBSA
2002	Stevens (22)	glass + NH_2-silane + NHS + EDC + Succinoylated dextran + EDC + NHS + SA	Si_3N_4 + NH_2-silane + NHS + EDC + Succinoylated dextran + EDC + NHS + BBSA
2006	De Odrowaz Piramowicz (23)	glass : poly-L-lysine + glutaraldehyde + Biotin	Si_3N_4 + NH_2-silane + glutaraldehyde + SA
2010	Teulon (7)	Au : HS-(CH_2)10-COOH + EDC + SA	Au : HS-$(CH_2)_2$-NH_2 + NHS-PEG-Biotin
2009 2010	Taninaka (24) Taninaka (25, 26)	a) Au : HS-(CH_2)10-SH + Mal-SA b) Au : HS-PEG-SH + Mal-SA	Au : HS-$(CH_2)_8$-NH_2 + NHS-PEG-Biotin
2019	Rico (56)	agarose layer : SA-coated agarose beads	Si_3N_4 + silane-PEG-Biotin

C) Site-specific covalent immobilization of SA

Year	Author	Immobilization on sample surface	Immobilization on cantilever tip
2015	Baumann (27)	glass + NH_2-silane + NHS-PEG-MAL + CoA + ybbR-GFP-Bio	Si_3N_4 + NH_2-silane + NHS-PEG-MAL + Cys-mST
2017 2018	Sedlak (29) Sedlak (9)	glass + NH_2-silane + NHS-PEG-MAL + Cys-mSA	Si_3N_4 + NH_2-silane + NHS-PEG-MAL + Cys-ddFLN4-Biotin
2018 2019	Sedlak (9) this work	glass + NH_2-silane + NHS-PEG-MAL + Cys-mSA	Si_3N_4 + NH_2-silane + NHS-PEG-MAL + CoA + ybbR-SdrG : Fgβ-ddFLN4-Biotin

D) Tearing SA apart

Year	Author	Immobilization on sample surface	Immobilization on cantilever tip
2011	Kim (8)	glass + SH-silane + MAL-C3-NTA : Ni^{2+} : HIS-tagged Ig-SA	Au : HS-$(CH_2)_{11}$-EG_3-NTA

"-": covalently bound entity; "+": covalently reacting; ":": non-covalent binding; "EDC": 1-Ethyl-3-(3-dimethyl-amino-propyl)-carbo-diimide; Mal: maleimide; "Cys": unique cysteine residue; "CoA": Coenzyme A; "ybbR": ybbR-tag; "GFP": green fluorescent protein; "ddFLN4": *Dictyostelium discoideum's* fourth filamin domain; "SdrG": SD repeat protein G; "Fgβ": short peptide from human fibrinogen β; "mSA": monovalent Streptavidin; "mST": monovalent Streptactin

References

1. Lee, G., D. Kidwell, and R. Colton. 1994. Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. Langmuir 10:354-357.
2. Florin, E. L., V. T. Moy, and H. E. Gaub. 1994. Adhesion forces between individual ligand-receptor pairs. Science 264:415-417.
3. Moy, V. T., E. L. Florin, and H. E. Gaub. 1994. Intermolecular forces and energies between ligands and receptors. Science 266:257-259.
4. Pincet, F., and J. Husson. 2005. The solution to the streptavidin-biotin paradox: the influence of history on the strength of single molecular bonds. Biophys J 89:4374-4381.
5. Rico, F., and V. T. Moy. 2007. Energy landscape roughness of the streptavidin-biotin interaction. J Mol Recognit 20:495-501.
6. Walton, E. B., S. Lee, and K. J. Van Vliet. 2008. Extending Bell's model: how force transducer stiffness alters measured unbinding forces and kinetics of molecular complexes. Biophys J 94:2621-2630.
7. Teulon, J. M., Y. Delcuze, M. Odorico, S. W. Chen, P. Parot, and J. L. Pellequer. 2011. Single and multiple bonds in (strept)avidin-biotin interactions. J Mol Recognit 24:490-502.
8. Kim, M., C. C. Wang, F. Benedetti, M. Rabbi, V. Bennett, and P. E. Marszalek. 2011. Nanomechanics of streptavidin hubs for molecular materials. Adv Mater 23:5684-5688.
9. Sedlak, S. M., L. C. Schendel, M. Cardoso dos Reis Melo, D. A. Pippig, Z. Luthey-Schulten, H. E. Gaub, and R. C. Bernardi. 2018. Direction Matters – Monovalent Streptavidin:Biotin Complex under Load. Nano Letters.
10. Zlatanova, J., S. M. Lindsay, and S. H. Leuba. 2000. Single molecule force spectroscopy in biology using the atomic force microscope. Prog Biophys Mol Biol 74:37-61.
11. Wong, S. S., E. Joselevich, A. T. Woolley, C. L. Cheung, and C. M. Lieber. 1998. Covalently functionalized nanotubes as nanometre-sized probes in chemistry and biology. Nature 394:52-55.
12. Merkel, R., P. Nassoy, A. Leung, K. Ritchie, and E. Evans. 1999. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. Nature 397:50-53.
13. Oesterhelt, F., M. Rief, and H. E. Gaub. 1999. Single molecule force spectroscopy by AFM indicates helical structure of poly (ethylene-glycol) in water. New Journal of Physics 1:6.
14. Hugel, T., M. Rief, M. Seitz, H. E. Gaub, and R. R. Netz. 2005. Highly stretched single polymers: atomic-force-microscope experiments versus ab-initio theory. Phys Rev Lett 94:048301.
15. Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. Science 265:1599.
16. Ortiz, C., and G. Hadziioannou. 1999. Entropic Elasticity of Single Polymer Chains of Poly(methacrylic acid) Measured by Atomic Force Microscopy. Macromolecules 32:780-787.
17. Livadaru, L., R. R. Netz, and H. J. Kreuzer. 2003. Stretching response of discrete semiflexible polymers. Macromolecules 36:3732-3744.
18. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276:1109-1112.

19. Puchner, E. M., G. Franzen, M. Gautel, and H. E. Gaub. 2008. Comparing proteins by their unfolding pattern. Biophys J 95:426-434.
20. Allen, S., J. Davies, A. C. Dawkes, M. C. Davies, J. C. Edwards, M. C. Parker, C. J. Roberts, J. Sefton, S. J. Tendler, and P. M. Williams. 1996. In situ observation of streptavidin-biotin binding on an immunoassay well surface using an atomic force microscope. FEBS Lett 390:161-164.
21. Wong, J., A. Chilkoti, and V. T. Moy. 1999. Direct force measurements of the streptavidin-biotin interaction. Biomol Eng 16:45-55.
22. Stevens, M. M., S. Allen, M. C. Davies, C. J. Roberts, E. Schacht, S. J. B. Tendler, S. VanSteenkiste, and P. M. Williams. 2002. The Development, Characterization, and Demonstration of a Versatile Immobilization Strategy for Biomolecular Force Measurements. Langmuir 18:6659-6665.
23. de Odrowaz Piramowicz, M., P. Czuba, M. Targosz, K. Burda, and M. Szymonski. 2006. Dynamic force measurements of avidin-biotin and streptavdin-biotin interactions using AFM. Acta Biochim Pol 53:93-100.
24. Taninaka, A., O. Takeuchi, and H. Shigekawa. 2009. Site-Selective Anatomy of Step-by-Step Reactions in Ligand–Receptor Bonding Processes Using Dynamic Force Spectroscopy. Applied Physics Express 2:085002.
25. Taninaka, A., O. Takeuchi, and H. Shigekawa. 2010. Reconsideration of dynamic force spectroscopy analysis of streptavidin-biotin interactions. Int J Mol Sci 11:2134-2151.
26. Taninaka, A., O. Takeuchi, and H. Shigekawa. 2010. Hidden variety of biotin-streptavidin/avidin local interactions revealed by site-selective dynamic force spectroscopy. Phys Chem Chem Phys 12:12578-12583.
27. Baumann, F., M. S. Bauer, L. F. Milles, A. Alexandrovich, H. E. Gaub, and D. A. Pippig. 2016. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. Nat Nanotechnol 11:89-94.
28. Howarth, M., D. J. Chinnapen, K. Gerrow, P. C. Dorrestein, M. R. Grandy, N. L. Kelleher, A. El-Husseini, and A. Y. Ting. 2006. A monovalent streptavidin with a single femtomolar biotin binding site. Nat Methods 3:267-273.
29. Sedlak, S. M., M. S. Bauer, C. Kluger, L. C. Schendel, L. F. Milles, D. A. Pippig, and H. E. Gaub. 2017. Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. PLOS ONE 12:e0188722.
30. Kramers, H. A. 1940. Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 7:284-304.
31. Bell, G. I. 1978. Models for the specific adhesion of cells to cells. Science 200:618-627.
32. Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. Biophys J 72:1541-1555.
33. Izrailev, S., S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. 1997. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J 72:1568-1581.
34. Yuan, C., A. Chen, P. Kolb, and V. T. Moy. 2000. Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy. Biochemistry 39:10219-10223.
35. Lo, Y.-S., Y.-J. Zhu, and T. P. Beebe. 2001. Loading-Rate Dependence of Individual Ligand–Receptor Bond-Rupture Forces Studied by Atomic Force Microscopy. Langmuir 17:3741-3748.
36. Hummer, G., and A. Szabo. 2003. Kinetics from nonequilibrium single-molecule pulling experiments. Biophys J 85:5-15.
37. Dudko, O. K., G. Hummer, and A. Szabo. 2006. Intrinsic rates and activation free energies from single-molecule pulling experiments. Phys Rev Lett 96:108101.

38. Ray, C., J. R. Brown, and B. B. Akhremitchev. 2007. Correction of systematic errors in single-molecule force spectroscopy with polymeric tethers by atomic force microscopy. J Phys Chem B 111:1963-1974.
39. Ray, C., J. R. Brown, and B. B. Akhremitchev. 2007. Rupture force analysis and the associated systematic errors in force spectroscopy by AFM. Langmuir 23:6076-6083.
40. Dudko, O. K., G. Hummer, and A. Szabo. 2008. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. Proc Natl Acad Sci U S A 105:15755-15760.
41. Freund, L. B. 2009. Characterizing the resistance generated by a molecular bond as it is forcibly separated. Proc Natl Acad Sci U S A 106:8818-8823.
42. Johnson, K. C., and W. E. Thomas. 2018. How Do We Know when Single-Molecule Force Spectroscopy Really Tests Single Bonds? Biophys J 114:2032-2039.
43. Lo, Y.-S., N. D. Huefner, W. S. Chan, F. Stevens, J. M. Harris, and T. P. Beebe. 1999. Specific Interactions between Biotin and Avidin Studied by Atomic Force Microscopy Using the Poisson Statistical Analysis Method. Langmuir 15:1373-1382.
44. Erdmann, T., S. Pierrat, P. Nassoy, and U. S. Schwarz. 2008. Dynamic force spectroscopy on multiple bonds: Experiments and model. EPL (Europhysics Letters) 81:48001.
45. Erlich, K. R., S. M. Sedlak, M. A. Jobst, L. F. Milles, and H. E. Gaub. 2019. DNA-free directed assembly in single-molecule cut-and-paste. Nanoscale 11:407-411.
46. Chilkoti, A., T. Boland, B. D. Ratner, and P. S. Stayton. 1995. The relationship between ligand-binding thermodynamics and protein-ligand interaction forces measured by atomic force microscopy. Biophys J 69:2125-2130.
47. Chivers, C. E., E. Crozat, C. Chu, V. T. Moy, D. J. Sherratt, and M. Howarth. 2010. A streptavidin variant with slower biotin dissociation and increased mechanostability. Nat Methods 7:391-393.
48. Laitinen, O. H., H. R. Nordlund, V. P. Hytönen, S. T. H. Uotila, A. T. Marttila, J. Savolainen, K. J. Airenne, O. Livnah, E. A. Bayer, M. Wilchek, and M. S. Kulomaa. 2003. Rational Design of an Active Avidin Monomer. Journal of Biological Chemistry 278:4010-4014.
49. Wu, S.-C., and S.-L. Wong. 2005. Engineering Soluble Monomeric Streptavidin with Reversible Biotin Binding Capability. Journal of Biological Chemistry 280:23225-23231.
50. Lim, K. H., H. Huang, A. Pralle, and S. Park. 2013. Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. Biotechnology and Bioengineering 110:57-67.
51. DeMonte, D., E. J. Drake, K. H. Lim, A. M. Gulick, and S. Park. 2013. Structure-based engineering of streptavidin monomer with a reduced biotin dissociation rate. Proteins: Structure, Function, and Bioinformatics 81:1621-1633.
52. Demonte, D., C. M. Dundas, and S. Park. 2014. Expression and purification of soluble monomeric streptavidin in Escherichia coli. Applied Microbiology and Biotechnology 98:6285-6295.
53. Bauer, M. S., L. F. Milles, S. M. Sedlak, and H. E. Gaub. 2018. Monomeric streptavidin: a versatile regenerative handle for force spectroscopy. bioRxiv.
54. Grubmuller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 271:997-999.
55. Rico, F., L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring. 2013. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. Science 342:741-743.
56. Rico, F., A. Russek, L. Gonzalez, H. Grubmuller, and S. Scheuring. 2019. Heterogeneous and rate-dependent streptavidin-biotin unbinding revealed by high-

speed force spectroscopy and atomistic simulations. Proc Natl Acad Sci U S A 116:6594-6601.
57. Thormann, E., P. L. Hansen, A. C. Simonsen, and O. G. Mouritsen. 2006. Dynamic force spectroscopy on soft molecular systems: improved analysis of unbinding spectra with varying linker compliance. Colloids Surf B Biointerfaces 53:149-156.
58. Hutter, J. L., and J. Bechhoefer. 1993. Calibration of atomic-force microscope tips. Review of Scientific Instruments 64:1868-1873.
59. Butt, H. J., and M. Jaschke. 1995. Calculation of thermal noise in atomic force microscopy. Nanotechnology 6:1.
60. Sader, J. E., I. Larson, P. Mulvaney, and L. R. White. 1995. Method for the calibration of atomic force microscope cantilevers. Review of Scientific Instruments 66:3789-3798.
61. te Riet, J., A. J. Katan, C. Rankl, S. W. Stahl, A. M. van Buul, I. Y. Phang, A. Gomez-Casado, P. Schon, J. W. Gerritsen, A. Cambi, A. E. Rowan, G. J. Vancso, P. Jonkheijm, J. Huskens, T. H. Oosterkamp, H. Gaub, P. Hinterdorfer, C. G. Figdor, and S. Speller. 2011. Interlaboratory round robin on cantilever calibration for AFM force spectroscopy. Ultramicroscopy 111:1659-1669.
62. Brand, U., S. Gao, W. Engl, T. Sulzbach, S. W. Stahl, L. F. Milles, V. Nesterov, and Z. Li. 2017. Comparing AFM cantilever stiffness measured using the thermal vibration and the improved thermal vibration methods with that of an SI traceable method based on MEMS. Measurement Science and Technology 28:034010.
63. Sedlak, S. M., L. C. Schendel, M. C. R. Melo, D. A. Pippig, Z. Luthey-Schulten, H. E. Gaub, and R. C. Bernardi. 2019. Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. Nano Letters 19:3415-3421.
64. Verdorfer, T., R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash, and H. E. Gaub. 2017. Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. J Am Chem Soc 139:17841-17852.
65. Edwards, D. T., J. K. Faulk, M. A. LeBlanc, and T. T. Perkins. 2017. Force Spectroscopy with 9-mus Resolution and Sub-pN Stability by Tailoring AFM Cantilever Geometry. Biophys J 113:2595-2600.

Chapter 2
Methods

2.1 Protein Preparation

2.1.1 Preparation of Streptavidin Constructs

Since SA is a tetrameric protein, its preparation is not straightforward. In order to create heterotetrameric SA constructs, different subunits have to be expressed separately. In addition, the cell lysis is complicated, because SA subunits form inclusion bodies in *E.coli*. After a comparatively long lysis protocol, the different subunits are denatured and then mixed in a defined ratio, before they are diluted in refolding buffer and finally purified using affinity chromatography.

Cloning of Streptavidin Constructs

While the first SA subunits (His-Cys-ST and DA), have been amplified from synthetic templates [9], all other subunits were created by modifying and further developing the His-Cys-ST or DA subunits by means of molecular cloning. In particular, a blunt-end cloning and a QuikChange protocol were applied for site-directed mutagenesis. A complete list of all subunits is given in Table 2.1. All variants of SA subunits were cloned into pET21a(+) or pET28a(+) vectors (Merck KGaA, Darmstadt, Germany). Both vectors contain an antibiotic resistance (to carbenicillin for pET21a and to kanamycin for pET28a). They also contain the lactose repressor gene *lacI* and the lac operator to allow for induction of expression by isopropyl β-D-1-thiogalactopyranoside (IPTG). The SA subunits were cloned between the T7-promotor and T7-terminator region to allow for expression by T7 RNA polymerase.

For the QuikChange protocol, forward and reverse primers were designed such that the amino acids to be substituted were in the middle of the complementary primers. On each side of the substituted amino acids, the primers had an overlap with the backbone of 15-25 bp resulting in a melting temperature of about 55°C. For the polymerase chain reaction (PCR), 10 µl of 2x Phusion MasterMix (New England Biolabs, Ipswich, USA), 1 µl 10 ng/µl plasmid DNA, 0.5 µl 1 pmol/µl of each primer, 0-2 µl dimethyl sulfoxide (DMSO) and 8-6 µl ultrapure water were mixed. In a thermocycler, the mixture was first heated to 98°C for 5 minutes for initial denaturation of the DNA. Then, 30 cycles of denaturation, annealing and elongation

were performed. For denaturation, the mixture was heated to 98°C for 20 s. Annealing was performed at 55°C for 20 s. For the Phusion polymerase, elongation was performed at 72°C. The length of the elongation step depended on the size of the plasmid and was estimated by 15 s/kbp. After the last PCR cylce, the mixture was kept at 72°C for 10 minutes to allow for final elongation. PCR products were circular plasmids. To digest template DNA, 1 µl of the restriction enzyme DpnI (Thermo Fisher Scientific, Waltham, USA) was added and incubated at 37°C for at least 30 minutes. The template DNA, which was of bacterial origin, was methylated and therefore cut by this enzyme. The PCR product was not methylated and thus not affected by the DpnI digest. To deactivate the enzyme, the mixture was heated to 80°C for 5 minutes.

For blunt-end cloning, forward and reverse primers were designed without any overlap. Amino acids to be substituted or added were encoded on one of the primers at the end. For the PCR, 10 µl of 2x Phusion MasterMix, 1 µl 10 ng/µl plasmid DNA, 0.5 µl 1 pmol/µl of each primer, 0-2 µl dimethyl sulfoxide (DMSO) and 8-6 µl ultrapure water were mixed. In a thermocycler, the mixture was first heated to 98°C for 5 minutes for initial denaturation of the DNA. Then, 30 cycles of denaturation, annealing and elongation were performed. For denaturation, the mixture was heated to 98°C for 20 s. Annealing was performed for 20 s. The annealing temperature depended on the melting temperature of the primers, the polymerase (here Phusion) and the amount of DMSO added. For the Phusion polymerase, elongation was performed at 72°C. The length of the elongation step depended on the size of the plasmid and was estimated by 15 s/kbp. After the last PCR cylce, the mixture was kept at 72°C for 10 minutes to allow for final elongation. PCR products were linear DNA strands. The success of a PCR could thus be monitored by agarose gel electrophoresis. To obtain a circular plasmid, the blunt ends of the linear DNA fragments had to be ligated. For this, 4.5 µl of the unpurified PCR product were mixed with 1 µl 10x CutSMART buffer (New England Biolabs, Ipswich, USA), 1 µl 10 mM adenosine triphosphate (New England Biolabs, Ipswich, USA), 0.5 µl polyethylene glycol of 4-6 kDa molecular weight (Thermo Fisher Scientific, Waltham, USA) and 1 µl of the enzymes DpnI, T4 Polynucleotide Kinase (New England Biolabs, Ipswich, USA), T4 DNA Ligase (Thermo Fisher Scientific, Waltham, USA). Using a thermocycler the mixture was first kept at 37°C for 15 minutes to allow the DpnI to digest all of the remaining methylated template DNA. Then, the temperature was set to 22°C for 45 minutes. At this temperature, T4 PNK added phosphates on the 5' ends of the linear DNA. The DNA Ligase then ligated the phosphorylated blunt ends. Finally, the temperature was set to 80°C to inactivate enzymes by denaturation.

After digestion (for the blunt-end protocol after ligation) *E.coli* DH5αcells were transformed with 2-4 µl of the PCR product mixture, plated on an LB agar plate containing the appropriate antibiotic (carbenicillin for pET21a or kanamycin for pET28a) and grown overnight at 37°C.

Several 5 ml LB medium batches containing the appropriate antibiotic were inoculated with bacteria from a single colony on the agar plate and grown in a shaker overnight at 37°C. DNA was extracted from the cells using a commercial miniprep kit (QIAprep Spin Miniprep Kit; QIAGEN, Hilden, Germany) and sent for sequencing (to Eurofins Genomics, Ebersberg, Germany) using the T7 promotor as primer to check if the mutagenesis was successful.

Table 2.1: List of all SA/TA/ST/DA subunits cloned in the context of this thesis. The sequence of the subunits (N- to C-terminus) is provided in short notation: SA/TA/ST/DA indicate the corresponding protein sequence. Cys indicates a unique cysteine residue. His indicates a polyhistidine tag. G- and -LPETGG indicate an N-terminal glycine or a C-terminal sortase motif – both can be used for sortase-mediated linking. Note: the subunits contain additional amino acids without any function between the SA/TA/ST/DA domain and the tags (*e.g.* GS-repeats).

construct	construct	construct
ST	DA	SA
Cys-ST	Cys-DA	G-His-SA
ST-Cys-His	DA-Cys-His	SA-Cys-His
G-His-Cys-ST	G-His-Cys-DA	G-His-Cys-SA
G-His-ST	G-His-DA	G-His-SA
G-His-ST-Cys	G-His-DA-Cys	G-His-SA-Cys
	G-His-Cys-DA-LPETGG	G-His-Cys-SA-LPETGG
TA		His-Cys-SA-LPETGG
TA-Cys-His		Cys-SA-Cys-His
G-His-Cys-TA		Cys-SA-His

Expression of Streptavidin Constructs

BL21-CodonPlus (DE3)-RIPL or NiCo21(DE3) competent cells were transformed with the pET-vectors endocing for a certain SA subunit. 15 ml LB medium containing the appropriate antibiotic were inoculated with the bacteria and incubated overnight in a shaker at 37°C.

For expression, either SB medium or auto-induction medium was used. When using SB medium, 300-500 ml SB medium, 6-10 ml KH_2PO_4 and the appropriate antibiotic were inoculated with the preculture until an optical density at 600 nm OD_{600} of 0.1 was reached. The expression cultures were incubated in a shaker at 37°C, until an OD_{600} of 0.9 was reached. To induce protein expression 0.2 mM IPTG was added. The temperature of the expression cultures was then lowered to 18°C. When using auto-induction medium, 300-500 ml of auto-induction medium was inoculated with the preculture. After incubating the expression cultures in a shaker at 37°C for 5-6 h , the temperature was lowered to 18°C. Monitoring the OD_{600} and adding IPTG was not needed. After 16 h at 18°C, the cells were harvested. For this, the expression cultures were centrifuged at 20,000 × rcf for 15 minutes. The supernatant was discarded. At this point, the cell pellets could be stored at -80°C. Alternatively, one could directly proceed with the cell lysis.

Purification of Strepavidin Constructs

The cell pellets were weighted and suspended in 5 ml/g Bacterial Protein Extraction Reagent (BPER; Thermo Fisher Scientific, Waltham, USA). 100 µg/ml Lysozyme and 10 µg/ml DNase I were added. The suspensions were put on a rolling shaker at 4°C for 20 minutes.

For complete disintegration of the cells, the suspensions were sonicated on ice. The lysates were then centrifuged at 50,000 × rcf for 30 minutes. The supernatants were discarded and the newly formed pellets were suspended in 10 ml/g lysis buffer (phosphate buffered saline (PBS; Sigma-Aldrich, St. Louis, USA), 1 mM DTT, 0.1% Triton X-100, pH 7.4). The

sonication, centrifugation and resuspension steps were repeated until the supernatants were clear liquids.

After discarding the supernatants, the remaining pellets (mainly consisting of inclusion bodies) were suspended in protein denaturation buffer (PBS, 6 M guanidine hydrochloride, pH 7.5). For a last time, the suspensions were sonicated and centrifuged. This time, the supernatants (containing the denatured SA subunits) were kept, while the insoluble parts forming a pellets were discarded. The supernatants were filtered through a sterile 0.22 µm filter. Then, their absorbance at 280 nm was measured. The denatured subunits were thoroughly mixed in a 1:10 ratio (subunits with polyhistidine tag : subunits without polyhistidine tag) according to their absorbance at 280 nm. The mixture was then slowly (drop-by-drop) diluted in 500 ml refolding buffer (PBS, 10 mM β-Mercaptoethanol) at 4°C. The subunits were allowed to refold into their native structure, forming tetrameric SA. Usually, the proteins were kept in refolding buffer over night.

The refolded SA was purified by immobilized metal ion affinity chromatography. For this, a HisTrap FF 5-ml column (GE Healthcare, Chicago, USA) was equilibrated with washing buffer (PBS, 20 mM imidazole, pH 7.5). This column is packed with Sepharose (cross-linked agarose beads) and Nickel ions (nickel-nitrilotriacetic acid). The refolding buffer containing the SA construct was centrifuged at 20,000 × rcf for 10 min and subsequently filtered through a sterile 0.22 µm filter, before it was loaded onto the column. The SA construct got bound to the column by its polyhistidine tag. The imidazole side-chains of the histidines chelated the nickel ions and thus the construct was immobilized. While pH 8.0 would be most effective for the chelation, binding to the column at the lower pH of the refolding buffer was efficient enough. Usually, about 1 mg of the SA construct were obtained.

After loading the column, it was washed with 50 to 100 ml of washing buffer. The free imidazole in the buffer competed with the histidines. At the low imidazole concentration, unspecifically bound proteins were eluted from the column. To elute the SA construct from the column, the imidazole concentration was linearly increased over a range of 100 ml to a final concentration of 250 mM. Comprising a single polyhistidine tag, the desired SA construct was eluted at an imidazole concetration of about 100 mM. Wrong combinations of SA subunits, comprising more than one polyhistidine tags, were eluted later, while unspecifically adsorbed proteins, *e.g.* combinations of SA subunits without polyhistidine tags, were eluted earlier. Collecting 1 ml fractions of the eluate, the desired SA construct was separated from the other, unwanted variants. By monitoring the adsorption of the eluate at 280 nm, protein containing fractions were registered.

Fractions containing the desired SA construct were identified using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). For this, three types of samples were prepared for every absorption peak: One, where biotinylated green flourescent protein (GFP) was added before mixing with loading buffer to check the valency of the SA construct. One, where the protein was only mixed with loading buffer. One, where the protein was mixed with loading buffer and then heated to 95°C. If heated that much, SA is disintegrated into its subunits and denatured. The SDS in the loading buffer attaches to the poypeptide chain providing a net charge that is proportional to the molecular weight of the subunit. By this, the correct fraction could be identified by binding the right number of biotinylated GFP, by having the correct molecular weight and by the 3:1 ratio (without/with polyhistidine tag) of the subunits.

Table 2.2: List of all SA variants prepared in the context of this thesis. The sequence of the subunits (N- to C-terminus) is provided in short notation: SA/TA/ST/DA indicate the corresponding protein sequence. Cys indicates a unique cysteine residue. His indicates a polyhistidine tag. G- and -LPETGG indicate an N-terminal glycine or a C-terminal sortase motif – both can be used for sortase-mediated linking. Note: the subunits contain additional amino acids without any function between the SA/TA/ST/DA domain and the tags (*e.g.* GS-repeats).

label	subunit A	subunit B	subunit C	subunit D
His-Cys-mST	DA	DA	DA	His-Cys-ST
His-Cys-tST	ST	ST	ST	His-Cys-ST
G-His-mST-Cys	DA	DA	DA	G-His-ST-Cys
mST-Cys-His	DA	DA	DA	ST-Cys-His
tST-Cys-His	ST	ST	ST	ST-Cys-His
His-Cys-dSA	DA	DA	DA	His-Cys-DA
His-Cys-mSA	DA	DA	DA	His-Cys-SA
His-Cys-tSA	SA	SA	SA	His-Cys-SA
His-tSA	SA	SA	SA	His-SA
G-His-mSA-Cys	DA	DA	DA	G-His-SA-Cys
His-mSA-Cys	DA	DA	DA	His-SA-Cys
Cys-mSA-His	DA	DA	DA	Cys-SA-His
dSA-Cys-His / 0SA	DA	DA	DA	DA-Cys-His
mSA-Cys-His / 1SA	DA	DA	DA	SA-Cys-His
2SA-Cys-His / 2SA	DA	DA	SA-Cys-His	SA-Cys-His
3SA-Cys-His / 3SA	SA	SA	SA	DA-Cys-His
tSA-Cys-His / 4SA	SA	SA	SA	SA-Cys-His
His-Cys-mTA	DA	DA	DA	His-Cys-TA
His-Cys-tTA	TA	TA	TA	His-Cys-TA
mTA-Cys-His	DA	DA	DA	TA-Cys-His
tTA-Cys-His	TA	TA	TA	TA-Cys-His

The corresponding fractions were then pooled and the desired SA construct was dialyzed against PBS at 4°C. In contrast to other proteins that are rapidly frozen and then stored at -80°C, the SA constructs were stored in PBS at 4°C, because freezing and thawing might harm their tetravalent structure. A list of all SA variants produced in the context of this thesis are listed in Table 2.2.

2.2 Binding Kinetics

2.2.1 Receptor-Ligand Interactions

A ligand is a molecule that can form a complex with another molecule, a so-called receptor. In the biological context, mainly proteins are denoted as receptors. Proteins are the most versatile class of macromolecules that fulfill many different vital functions in biological systems. Proteins consist of an amino acid chain with a distinct sequence that is folded in a certain three-dimensional structure, called conformation. This structure is however not

rigid but contains flexible parts. External factors as temperature, salt concentration or pH can induce changes of the conformation.

Small molecules, sugars, lipids, DNA or other proteins can act as ligands. The part of the protein, where the ligand molecule binds, is called binding site. The protein's amino acid side-chains are arranged to allow for highly specific interaction: ligand and binding site are often complementary in shape. Ligand binding can induce conformational changes in the receptor protein. Thereby, ligands can serve as a signal and trigger certain action or inhibit a protein's function. In a biological system, this can result in complex signal cascades, *e.g.* the addition of a certain sugar molecule to the culture medium can induce protein expression within bacteria (*cf.* Section 2.1.1).

Usually, the interaction between receptor proteins and ligand molecules is non-covalent and relies on Van der Waals forces, hydrophobic interactions, hydrogen bonds and ionic bonds. Thermodynamically, ligand binding is a reversible process, *i.e.* dissociation of a ligand from the receptor is possible. In the following, the theoretical framework and experimental techniques for ligand-receptor interaction are described.

2.2.2 Binding Kinetics

The concentrations of bound and free ligands in presence of receptor molecules are determined by rate-dependent processes and can be described by receptor-ligand kinetics. When a certain concentration of receptors $[R]$ and ligands $[L]$ is mixed, they will form complexes $[RL]$ with a certain rate k_{on}. At the same time, ligands spontaneously dissociate from receptor molecules with the rate k_{off}.

$$[R] + [L] \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} [RL] \tag{2.1}$$

By this, a dynamic equilibrium is reached. The concentrations of free receptors, free ligands and receptor-ligand complexes are then constant. Mathematically, the following condition has to be met:

$$\frac{d[RL]}{dt} = -k_{off} \cdot [RL] + k_{on} \cdot [R] \cdot [L] \overset{!}{=} 0 \tag{2.2}$$

Obviously, the on-rate $k_{on} = [\mathrm{s}^{-1}\mathrm{M}^{-1}]$, is dependent on ligand concentration, whereas the off-rate $k_{off} = [\mathrm{s}^{-1}]$ does not dependent on the ligand concentration. The rates, an thus the equilibrium position, can be influenced by environmental conditions (pH, temperature, salt concentration). For example, if the on-rate depends on the diffusion of ligands within the buffer solution containing the receptor molecules, it will be dependent on temperature, for the diffusion coefficient is temperature-dependent.

From the off-rate, the average lifetime of a single bond can be defined as $\tau = \frac{1}{k_{off}}$. After this time, 63% of all ligands will have dissociated ($1 - \exp(-1) = 0.63$).

Equation 2.2 can be rewritten as

$$K_D = \frac{[R][L]}{[RL]} = \frac{k_{off}}{k_{on}}. \tag{2.3}$$

By this, just as for chemical reactions, an equilibrium constant is defined. It is called binding constant or dissociation constant K_D. It is a measure for the affinity, *i.e.* the binding strength

of a receptor-ligand interaction. Although equilibrium constants are actually dimensionless numbers, K_D is commonly given in units of concentration, molar. The ligand concentration $[L] = K_D$ is the concentration at which 50% of the receptors have a ligand bound (under the condition $[R] \gg K_D$), as can be seen from

$$f = \frac{[RL]}{[R] + [RL]} = \frac{L}{K_D + L} \tag{2.4}$$

where f is the fraction of bound receptors.

There are also receptor molecules that can bind more than one ligand. If the on- and off-rates for all binding sites are identical and independent of each other, the previous equations can be rewritten as follows:

$$[R] + n \cdot [L] \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} [RL_n] \tag{2.5}$$

$$K_D = \frac{[R][L]^n}{[RL_n]} = \frac{k_{off}^n}{k_{on}^n} \tag{2.6}$$

In this context, the terms allostery and cooperativity are important. Allostery means that ligand binding influences the conformation of the receptor molecule in such a way that functional properties at another site are altered. Cooperativity describes how binding of one ligand increases or decreases the chance of binding a another ligand.

2.2.3 Fluorescence Anisotropy

Fluorescence Anisotropy (FA) is a technique to characterize the binding behavior of receptor-ligand interactions (*cf.* Section 5.1). Measurements at equilibrium and kinetic measurements can be performed. The equilibrium constant (dissociation constant) as well as on- and off-rates can be determined.

For an FA experiments, fluorescently labeled ligand molecules are excited with polarized light. The light emitted from the fluorophores is analyzed using polarization filters. The intensities of the parallel $I_{\parallel}$ and orthogonal components $I_{\perp}$ are measured. Here, parallel means same polarization for the incoming and outgoing light, while orthogonal means that the angle between excitation and emission filter reads 90°. The anisotropy A is defined as the difference between the two components divided by the total intensity:

$$A = \frac{I_{\parallel} - G \cdot I_{\perp}}{I_{\parallel} + 2 \cdot G \cdot I_{\perp}} \tag{2.7}$$

Here, G is an instrument specific parameter which takes into account that the sensitivity for the different components might be different. It has to be calibrated before the actual measurement.

FA relies on the dynamics on the fluorophores. It compares two timescales: The average lifetime of the excited state of the fluorophore and the timescale on which its orientation changes. For example, if the excited state is short-lived and the rotation of the molecule is very slow, the orientation of the molecule does not change much between excitation and de-excitation. The polarization of the incoming and outgoing light will be highly correlated;

the anisotropy reads one. On the other hand, if the timescale on which the molecule rotates, *i.e.* loses its orientation, is much shorter than the average lifetime of the excited state, the polarization of the exciting light is not correlated with the re-emitted light. On average, light is emitted isotropically and the anisotropy reads zero.

The dynamics of a ligand molecule change, when it gets bound to a receptor molecule. This effect is more pronounced if a small ligand binds to a comparatively large receptor molecule. For the experiments conducted in the context of this thesis, a short single-stranded DNA labeled with a dye on the one and small ligand on the other end is mixed with receptor molecules in such a ratio that the all ligands are bound. Then, unlabeled ligands are added in excess to the solution. Monitoring the anisotropy over time, the unbinding of the fluorescently labeled ligands from the receptor, *i.e.* the off-rate k_{off}, is measured, because they are most likely replaced by the unlabeled ligands added in excess. To correct for evaporation or bleaching effects, it is advisable to preform control measurements in parallel: one control with labeled ligands only (also used for G-factor reference) and a second control with labeled ligands mixed with a large excess of receptor molecules.

2.2.4 Connection to Thermodynamic Potentials

The connection between binding kinetics and thermodynamics was first described by Van't Hoff [59] and Arrhenius [60] who set up the following equation:

$$k_{off} = A \cdot \exp\left(-\frac{\Delta G}{k_B T}\right) \tag{2.8}$$

Arrhenius connected the off-rate k_{off} with temperature $k_B T$ and the Gibb's free energy ΔG, which he initially called activation energy. The prefactor A was determined in 1940 by Kramers [61]. In the following paragraph, a short derivation of Kramers result is given for a particle that moves in an external force field and is subject to Brownian motion. At the beginning the particle is trapped in a potential well but can at some point escape by overcoming an energy barrier. The probability of escape is calculated for the one-dimensional overdamped case, which is relevant in the context of this thesis (*i.e.* when a ligand is disturbed, moved slightly out of the binding pocket, it moves back without overshooting beyond the equilibrium position). More general approaches are discussed in a review by Hänggi *et al.* [62].

Escape from a One-Dimensional Potential Well In Figure 2.1, the energy landscape of a receptor-ligand interaction along a reaction coordinate x is depicted. In the bound state, the ligand is trapped in the metastable state at x_b. The energy barrier at the transition state $x^{\ddagger}$ must be overcome to escape the binding potential, *i.e.* for dissociation of the ligand from the receptor, and to transition to the unbound state x_u.

In this one-dimensional case, the binding potential around x_b is approximated as an harmonic potential $V_b(x)$ with the force constant κ_b

$$V_b(x) = \frac{1}{2}\,\kappa_b\,(x - x_b)^2 \tag{2.9}$$

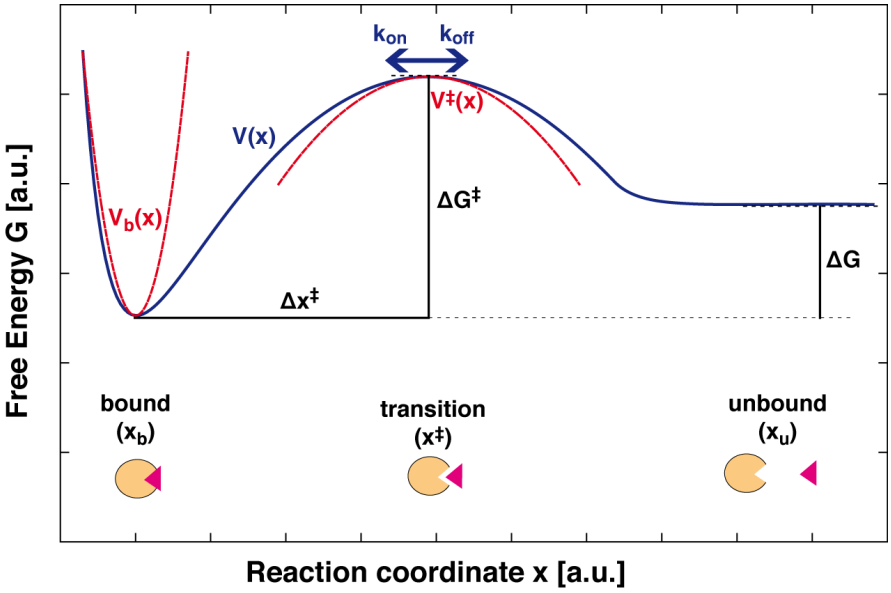


Figure 2.1: Schematic of an energy landscape. A particle trapped in the meta-stable bound state at x_b has to overcome the energy barrier of $\Delta G^{\ddagger}$ (at the transition state $x^{\ddagger}$) to transition into the unbound state x_u. The difference in Gibbs free energy between bound and unbound state is given by ΔG. For the theoretical description, the potential well and the energy barrier are approximated by harmonic potentials $V_b(x)$ and $V^{\ddagger}(x)$, respectively.

to calculate the probability flux J from x_b to x_u. At time $t = 0$, the probability density $P(x, t = 0)$ of the ligand within this potential is normalized to $\int dx P(x) = 1$. It reads

$$P(x, t = 0) = \sqrt{\frac{\kappa_b}{2\pi k_B T}} \exp\left(\frac{-\kappa_b (x - x_b)^2}{2 k_B T}\right), \tag{2.10}$$

where k_B is the Boltzmann constant and T the absolute temperature.

Using the one-dimensional Fokker-Planck equation (for the over-damped case)

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial J}{\partial x} = \frac{1}{\gamma}\frac{\partial}{\partial x}\left(P(x,t)\frac{\partial V(x)}{\partial x}\right) + D\frac{\partial^2 P(x,t)}{\partial x^2} \tag{2.11}$$

the flux is given by

$$\begin{aligned} J &= \frac{1}{\gamma}\left(P(x,t)\,\frac{\partial V(x)}{\partial x}\right) + D\,\frac{\partial P(x,t)}{\partial x} \\ &= \frac{k_B T}{\gamma}\exp\left(-\frac{V(x)}{k_B T}\right)\frac{\partial}{\partial x}\left(P(x,t)\cdot\exp\left(\frac{V(x)}{k_B T}\right)\right). \end{aligned} \tag{2.12}$$

This can be rewritten as

$$J\exp\left(\frac{V(x)}{k_B T}\right) = \frac{k_B T}{\gamma}\frac{\partial}{\partial x}\left(P(x,t)\cdot\exp\left(\frac{V(x)}{k_B T}\right)\right). \tag{2.13}$$

The rearrangement of the terms allows to easily integrate both sides of the equation from x_b to x_u. (We will now assume that the probability density only changes slowly over time, *i.e.*

that the flux J is independent of x.) On the right hand side, the integral can be solved using that $P(x_u) = 0$:

$$\int_{x_b}^{x_u} dx \frac{\partial}{\partial x}\left(P(x,t)\cdot\exp\left(\frac{V(x)}{k_BT}\right)\right) = \exp\left(\frac{V(x_b)}{k_BT}\right)P(x_b) - \exp\left(\frac{V(x_u)}{k_BT}\right)P(x_u)$$
$$= \sqrt{\frac{\kappa_b}{2\pi k_BT}} \tag{2.14}$$

To solve the integral on the left hand side, we approximate the transition state barrier, which dominates the integral, by the harmonic potential

$$V^{\ddagger}(x) = \Delta G^{\ddagger} - \frac{1}{2}\kappa^{\ddagger}(x - x^{\ddagger})^2. \tag{2.15}$$

Here, $\kappa^{\ddagger}$ is the force constant of the harmonic potential, $\Delta G^{\ddagger}$ is the difference in Gibbs free energy of the meta-stable state and the transition state. The integral reads:

$$\begin{aligned}\int_{x_b}^{x_u} dx\ \exp\left(\frac{V(x)}{k_BT}\right) &= \int_{x_b}^{x_u} dx \exp\left(\frac{\Delta G^{\ddagger} - \frac{1}{2}\kappa^{\ddagger}(x-x^{\ddagger})^2}{k_BT}\right)\\ &= \exp\left(\frac{\Delta G^{\ddagger}}{k_BT}\right)\int_{x_b}^{x_u} dx \exp\left(\frac{-\kappa^{\ddagger}(x-x^{\ddagger})^2}{2k_BT}\right)\\ &= \exp\left(\frac{\Delta G^{\ddagger}}{k_BT}\right)\sqrt{\frac{2\pi k_BT}{\kappa^{\ddagger}}}\end{aligned} \tag{2.16}$$

In the last step, the limit $x_b \to -\infty$ and $x_u \to \infty$ was applied, assuming that the contributions to the integral over $\exp\left(\frac{-\kappa^{\ddagger}(x-x^{\ddagger})^2}{2k_BT}\right) \to 0$ are negligible.

Putting all terms together, the following expression for the flux is obtained:

$$\begin{aligned}J &= \frac{k_BT}{\gamma}\sqrt{\frac{\kappa_b}{2\pi k_BT}\frac{\kappa^{\ddagger}}{2\pi k_BT}}\exp\left(-\frac{\Delta G^{\ddagger}}{k_BT}\right)\\ &= \frac{\sqrt{\kappa^{\ddagger}}}{\gamma}\frac{\sqrt{\kappa_b}}{2\pi}\exp\left(-\frac{\Delta G^{\ddagger}}{k_BT}\right)\\ &= \frac{\omega^{\ddagger}}{\gamma}\frac{\omega_b}{2\pi}\exp\left(-\frac{\Delta G^{\ddagger}}{k_BT}\right) = k_{off}\end{aligned} \tag{2.17}$$

In the last step, the force constants were replaced by the angular frequencies $\omega_b = \sqrt{\frac{\kappa_b}{m}}$ corresponding to the harmonic potentials in the meta-stable state and at the transition state. Applying an ergodicity argument, the flux in phase space J corresponds to the off-rate k_{off}.

2.2.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is an experimental technique to examine the binding of ligands to receptor molecules. The experiments are conducted at constant pressure and

constant temperature, *i.e.* in an isothermal and isobaric environment. For these conditions, the Gibbs free energy is used to describe the thermodynamic state of the system. In most cases, heat, the so-called binding enthalpy ΔH, is released when a ligand binds to a receptor. In addition, the number of possible microstates decreases, resulting in a decrease in entropy ΔS. The change in the system's Gibbs free energy ΔG, which is also related to the dissociation constant K_D reads:

$$\Delta G = \Delta H - T \cdot \Delta S = R \cdot T \cdot \ln(K_D) \tag{2.18}$$

Here, T is the absolute temperature and R the universal gas constant. From ITC data, binding enthalpy ΔH, stoichiometry n and dissociation constant K_D can be extracted. Using Equation 2.18 changes in Gibbs free energy ΔG and ΔS can also be determined. In contrast to FA, rates cannot be determined by ITC.

All ITC experiments described in this thesis have been conducted in a MicroCal iTC_{200} system (Malvern Instruments Ltd., Malvern, UK). It consists of two identical cells made from Hastelloy C-276 (a nickel-chromium-molybdenum alloy with good corrosion resistance and heat conductivity), which are integrated into an adiabatic case. The sample cell is filled with a known concentration of receptor molecules dissolved in buffer, while the reference cell is filled with ultrapure water. Both cells are heated separately but very accurately kept at the same constant temperature.

During the measurement, small amounts of ligand molecules are titrated into the sample cell. Upon ligand binding, heat is released. Therefore, less power is needed to keep the temperature of the sample cell constant. Recording the difference in power required to hold both sample and reference cell at the same temperature, the binding enthalpy can thus be measured. For better distribution of molecules within the cell and for better heat transport, the solution in the sample cell is continuously stirred.

Knowing the volume of the cell, the concentrations of receptor and ligand solutions and the volume of ligand solution added to the cell, the ratio of receptor and ligand molecules within the cells can be calculated. Adding more and more ligands to the cell, the receptors get saturated and the released heat decreases. Plotting released heat against the ratio of receptor and ligands in the cell, the binding stoichiometry and the dissociation constant, which is related to the slope of the saturation of the receptor molecules, can be determined.

To perform a good measurement, it is important to dissolve both receptors and ligands in exactly the same buffer. Mismatched buffers cause entropy of mixing, when adding ligands to the sample cells, thus falsifying the measurement. Beyond that, the concentrations of receptor and ligand have to be chosen correctly (depending on the stoichiometry and the details of the experimental setup) and carefully adjusted.

In the context of this thesis (*cf.* Section 3.1), it should be mentioned that the dissociation constant of biotin and SA is beyond the detection limit of the ITC instrument: Because of the high affinity, the saturation of the receptor molecules is reached abruptly, making it impossible to reasonably fit the transition range. Using smaller concentrations or adding ligands in smaller steps would reduce the released heat impairing the signal-to-noise ratio of the measurement.

2.2.6 Atomic Force Microscopy

Atomic force microscopy (AFM) is a type of scanning probe microscopy. It was first realized in 1986 by Binning *et al.* [63]. Scanning probe microscopes raster scan a surface using a small probe. Monitoring the interactions between the probe and the surface, information about the surface properties is obtained – just like blind people can read braille when moving their finger across embossed printing. Using nanometer-sized probes precisely controlled by piezoelectric elements and fast control circuits, some scanning probe microscopy techniques can reach atomic resolution.

Images taken by scanning probe microscopy are different from optical images. Optical microscopy provides information on how a surface or objects on a sample surface interact with, for most cases how they reflect, light. Scanning probe microscopy, on the other hand, provides information on how the surface or objects on a sample surface interacted with the probe that is used for scanning, *e.g.* a scanning tunneling microscope provides information about the local density of states while a magnetic resonance force microscope provides information about resonant spins on the surface [64].

Principle of the Atomic Force Microscope

For AFM, a micrometer-long cantilever with a sharp tip at the end is used as probe and scanned across the sample surface. The probe material is often silicon nitride. The interaction of the tip with the surface results in deflection of the cantilever. The deflection of the cantilever is monitored by reflecting a laser off the back of the cantilever onto a position sensitive device (PSD). The differential voltage measured by the PSD is proportional to the deflection of the cantilever. AFM can resolve structures smaller than one nanometer – for comparison: the resolution limit of conventional optical microscopy is about 250 nm (Abbe limit).

Measurement Modes

The simplest way of imaging a surface by AFM is the contact mode. For this, the cantilever tip is brought into close proximity of the surface using piezoelectric elements. High attractive forces cause the cantilever tip to snap in on the surface. Scanning the tip across the surface, the cantilever is bend. The deflection is monitored by the PSD and a two-dimensional image is obtained. Admittedly, scratching the cantilever tip across the surface might not be the best idea, because it can damage both cantilever tip and surface – especially, in the case of soft, *e.g* biological, samples. Therefore, two other imaging modes have been developed and are widely used: non-contact and tapping mode. For both, a small piezoelectric element is used to drive the cantilever.

In non-contact mode, the cantilever is oscillated at its resonance frequency. Forces between the cantilever tip and the surface alter the resonance frequency of the cantilever. A feedback-loop is used to adjust the drive frequency. Scanning across the surface while recording the change in resonance frequency, a two-dimensional image is obtained. As the cantilever tip is not in direct contact with the surface, surface and cantilever tip are not damaged. Non-contact mode has difficulties in ambient conditions: thin films of liquid on the sample shield some forces so that the change of the resonant frequency is lowered. Bringing

the cantilever closer to the surface, it might snap in on the surface. Hence, non-contact mode is mostly applied when measuring in vacuum.

This issue is resolved by the tapping mode. Hereby, the cantilever is driven slightly below its resonance frequency. Instead of changing the drive frequency, when scanning across the surface, the distance between tip and surface is adjusted. The feedback-loop alters this distance such that the amplitude of the oscillation is kept constant. By this, a continuous snap in on the surface is avoided while the distance between cantilever tip and sample surface is still low enough to sense all important interactions: electrostatic forces, van-der-Waals forces, adhesion forces (owed to the water meniscus) and repulsive forces mostly due to Pauli's exclusion principle.

Imaging Biomolecules with the AFM

To image biomolecules by AFM, they have to be trapped on a surface. For the imaging experiments performed in the context of this thesis (*cf.* Section 3.4), mica surfaces are used. Mica is a sheet silicate that can be easily cleaved using adhesive tape to obtain a flat and clean imaging substrate. The freshly-cleaved mica substrate is incubated for 30 s with 0.01% (w/v) poly-L-lysine of 0.5-2 kDa molecular weight dissolved in ultrapure water. The substrate is then thoroughly rinsed with 30 ml ultrapure water and dried with nitrogen. The poly-L-lysine coated substrate is then incubated for 30 s with a mixture of biotinylated DNA and SA and cautiously rinsed with 30 ml ultrapure water and dried with nitrogen. The negatively charged DNA is pulled down and trapped on the positively charged poly-L-lysine. SA that has one or several biotinylated DNA strands bound is also trapped on the surface. The dried substrate is stable over several days and can be imaged by AFM.

For this thesis (*cf.* Section 3.4), AFM-imaging has been performed on a commercial MFP-3D AFM (Asylum Research, Santa Barbara, USA; now: Oxford Instruments plc, Abingdon, UK) using 60 µm long silicon cantilevers AC160TS (Olympus, Tokyo, Japan) with a nominal resonance frequency of 300 kHz and a nominal spring constant of 26 pN. The tetrahedral tip has a length of 14 µm and a tip radius of 7 nm. Imaging was conducted in air with tapping mode, scanning either 512×512 or 1024×1024 points with a scan-rate of 1-3 Hz. Image sizes varied from 0.25 to 4 μm^2.

2.3 Force Spectroscopy

Force spectroscopy experiments aim to measure the force required to stretch or twist (bio-)polymers, to unfold protein domains or to separate a ligand from a receptor molecule. In an SMFS experiment, a molecule or receptor-ligand complex of interest is tethered between a sample surface and a probe, such as an AFM cantilever tip or a magnetic bead. Upon separation of probe and surface, the tethered molecules have to make up for the increasing distance. A restoring force builds up in the system. As force increases, the tethered molecules might rotate to relax, migrate to an entropically less favorable state or unfold to release additional contour length. At some point, the distance between sample and tip and thus the restoring force is large enough to break the complex. In principle, there are three measurement modes: First, constant-speed experiments, in which sample and probe are separated at a constant velocity. Second, force-ramp experiments, in which a fast force-feedback is employed to

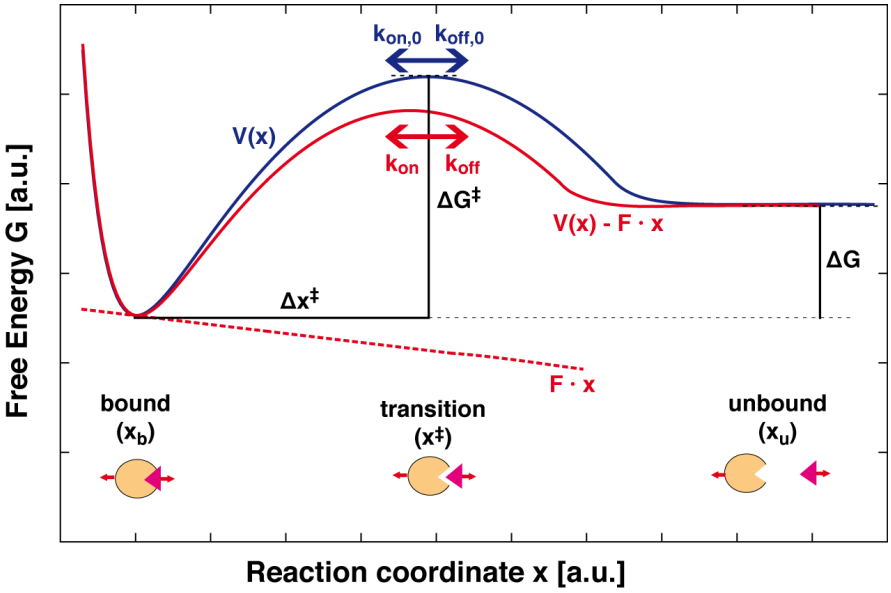


Figure 2.2: Effect of force on the energy landscape. A particle trapped in the meta-stable bound state at x_b has to overcome the energy barrier of $\Delta G^{\ddagger}$ (at the transition state $x^{\ddagger}$) to transition into the unbound state x_u. Application of an external force F tilts the energy landscape and lowers the energy barrier, resulting in a change of binding kinetics, *i.e.* of on- k_{on} and off-rates k_{off}. The difference in Gibbs free energy between bound and unbound state is given by ΔG.

linearly increase the restoring force that builds up. Third, force-clamp experiments, in which the distance between probe and sample is kept at a certain level, while the response of the tethered molecules is recorded, *e.g.* unfolding and refolding. The experiments described in this thesis were either performed at constant speed (AFM-based SMFS) or at constant force (magnetic tweezers measurements). In the following section, the theoretical framework for the behavior of a receptor-ligand system under force is derived. Subsequently, AFM-based SMFS and magnetic tweezers are explained.

2.3.1 Applying Force to a Receptor-Ligand System

The derivation, described in this section, follows Izrailev *et al.* [65] and Friedsam *et. al* [66]. When a receptor-ligand system is forcibly separated, a resisting force builds up and the energy landscape is tilted as depicted in Figure 2.2. By this, the energy barrier $\Delta G^{\ddagger}$ is lowered by $F\Delta x$. The off-rate $k_{off}(F)$ reads:

$$k_{off}(F) = \frac{\omega^{\ddagger}}{\gamma}\frac{\omega_0}{2\pi}\exp\left(-\frac{\Delta G^{\ddagger} - F\Delta x}{k_B T}\right) = k_{off,0}\exp\left(\frac{F\Delta x}{k_B T}\right) \tag{2.19}$$

Note, that $\Delta G^{\ddagger}$ is the Gibbs free energy to the transition state, which is not to be confused with the Gibbs free energy of the binding ΔG. Furthermore, effects of the force on the curvatures $\omega^{\ddagger}$ and ω_0 and distance to the transition state Δx are neglected here.

Assuming that the number of bonds N_b changes with a rate $k_{off}(F(t))$, the change in number of bonds over time is given by

$$\frac{dN_b}{dt} = -k_{off}(F(t))\,N_b \tag{2.20}$$

This differential equation is easily solved. To derive a bond rupture probability, the number of ruptured (unbound) bonds N_u is more useful. It is easy to rewrite the equation. For simplicity, we assume that N_b and N_u add up to one. In addition, the integration over time is substituted by an integration over force.

$$N_b(t) = \exp\left(-\int_0^t dt'\, k_{off}(t')\right) \tag{2.21}$$

$$N_u(t) = 1 - N_b(t) = 1 - \exp\left(-\int_0^t dt'\, k_{off}(t')\right) \tag{2.22}$$

$$N_u(F) = 1 - \exp\left(-\int_0^F dF'\, \frac{dt}{dF}\, k_{off}(F')\right) \tag{2.23}$$

To calculate the bond rupture probability $p(F)$, the number of ruptured bonds is derived with respect to force. Assuming a constant loading rate $\dot{F} = \frac{dF}{dt}$, the integral is analytically solved.

$$\begin{aligned} p(F) = \frac{dN_u(F)}{dF} &= \frac{1}{\dot{F}}\, k_{off}(F)\, \exp\left(-\int_0^F dF' \frac{dt}{dF} k_{off}(F')\right) \\ &= \frac{k_{off,0}}{\dot{F}}\, \exp\left(\frac{F\Delta x}{k_B T}\right) \exp\left(-\int_0^F dF' \frac{k_{off,0}}{\dot{F}} \exp\left(\frac{F\Delta x}{k_B T}\right)\right) \\ &= \frac{k_{off,0}}{\dot{F}}\, \exp\left(\frac{F\Delta x}{k_B T}\right) \exp\left(-\frac{k_{off,0}}{\dot{F}} \frac{\Delta x}{k_B T}\left(1 - \exp\left(\frac{F\Delta x}{k_B T}\right)\right)\right) \end{aligned} \tag{2.24}$$

The maximum of the above probability density function is given by

$$F(\dot{F}) = \frac{k_B T}{\Delta x}\, \ln\left(\frac{\dot{F}}{k_{off}} \frac{\Delta x}{k_B T}\right) \tag{2.25}$$

These derivations were first published in 1997 by Evans and Ritchie [67] and Izrailev *et al.* [65] in the same issue of the Biophysical Journal. Nowadays, it is known as Bell-Evans theory. There are various extensions of this theory, such as considering stochastic fluctuations of the force [68], possible rebinding events [69, 70], alternations of the location of the transition state [71], the stiffness of force transducer [72, 73] or arbitrary time-dependence of different parameters of the energy landscape [74]. It is yet beyond the scope of this thesis to go into more detail, because for all results discussed in the following only the basic Bell-Evans theory was applied.

2.3.2 AFM-based Single-Molecule Force Spectroscopy

In an AFM-based single-molecule force spectroscopy (SMFS) experiment, ligand molecules are immobilized on an AFM cantilever, while receptor molecules are immobilized on a sample surface. When the cantilever tip is approached to the surface, ligand molecules on the apex of the AFM cantilever tip interact with the receptor molecules on the surface and form non-covalent bonds. Retracting the cantilever from the surface, the force needed to unbind the ligand from the receptor is measured. This is accomplished by reflecting a laser from

the back of the cantilever onto a four-quadrant position sensitive device (PSD) – the force between tip an surface results in a bending of the cantilever which results in a shift of the laser reflex on the PSD.

An alternative to attaching ligands to the cantilever tip and their counterparts to the surface is to attach one kind of receptors to the cantilever and another kind of receptors to the surface. Interactions can be established by adding a chimeric protein construct to the measurement buffer that contains both kinds of ligands and is thus able to bind the receptors on both tip and surface.

To prevent surface effects, linker molecules are introduced between the surface material and the ligand or receptor molecules, respectively. In this thesis, polyethylene glycol (PEG) or elastin-like polypeptide (ELP) [75] linkers are used.

To obtain single-molecule interactions, the surface density of receptor molecules is adjusted. In addition, so-called fingerprint protein domains are used. These are connected in series with the receptor-ligand system, *i.e.* when force is applied by retracting the cantilever tip from the surface, they can unfold before the ligand unbinds from the receptor molecule. The characteristic unfolding pattern of these domains is used to identify pure single-molecule interactions. The unfolding force of the fingerprint domain can also be used as internal force reference. In this thesis, the green fluorescent protein (GFP) [76], the fourth filamin domain of *Dictyostelium discoideum* (ddFLN4) [77] and the alpha-helical protein FIVAR [78], which have all been well characterized before, are used as fingerprint domains.

Cantilever and surface functionalization

For a successful force spectroscopy experiment, receptor and ligand molecules have to be attached to cantilever tip and surface, respectively. The used linker molecules (PEG and ELPs) as well as proteins, which are polypeptide chain folded into a distinct three-dimensional structure, are linear polymers. The task of surface and cantilever tip functionalization can thus be compared to knotting together different kind of strings in a well-defined manner: First, one kind of strings is attached to the surface, then the next sort of strings is knotted to the first one. Step-by-step, different layers are added. In the case of molecular chains, the different sorts molecules are also added one by one but they are not connected by knotting. Instead, covalent linkages have to be established between them. This can be achieved by bioconjugate techniques, by enzyme-mediated linkages or by genetic fusion, *i.e.* co-expressing different protein domains as a single chimeric construct. To obtain a well-defined tethering geometry, all linkages have to be orthogonal to each other, *i.e.* with minimal cross-reactivity, and site-specific, *i.e.* the reaction site within a molecule must be unique. In the following, the methods that have been mainly employed in the context of this thesis are briefly summarized. Detailed descriptions of the different functionalization protocols are provided in the materials and methods sections of the associated publications (*cf.* Chapter 3).

Bioconjugation Two molecules can be covalently connected by a chemical reaction between functional groups. The reactions of N-hydroxysuccinimide to primary amines and of thiol groups to maleimide groups have been extensively used in the context of this thesis.

N-hydroxysuccinimide (NHS) reacts with primary amines (NH_2) to form a stable amide bond, if the pH is in the range of 7.2 to 9. This reaction is used to covalently link PEG (with

an NHS group at one end) to the aminosilanized cantilever tip or glass surface, respectively. Usually, the reaction is carried out in a 50-100 mM HEPES buffer at pH 7.5 for 30-45 minutes. Afterwards, it is important to properly rinse off unreacted NHS groups to prevent unwanted reactions with primary amines on proteins that are added later.

Maleimide groups react with sulfhydryl groups to form stable thioether bonds, if the pH is in the range of 6.5 to 7.5. This reaction is used to covalently link PEG (with a malemide group at one end) to cysteine residues on proteins or to Coenzyme A (CoA), as both contain a thiol group. Usually, the reaction is carried out in coupling buffer (50 mM Na_2PO_4, 50 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA), pH 7.2) for one hour. For a successful reaction, it is important to reduce disulfide bridges between proteins to make the thiol accessible for the maleimide. For this, the proteins are incubated with Bond-Breaker TCEP solution (Thermo Fisher Scientific, Waltham, USA) before the reaction with maleimide.

Enzyme-mediated linkage Besides direct chemical reactions, linking reactions can be mediated by enzymes. In the context of this thesis, the enzymes Sfp phosphopantetheinyl transferase (Sfp) that links the peptide with the sequence DSLEFIASKLA (ybbR-tag) to CoA [79] and sortase A (eSrt) that establishes a covalent connection between proteins, in particular between an LPETG-motif and an N-terminal glycine residue [80]. The two reactions are bio-orthogonal and can in principle be carried out in parallel.

Sfp catalyzes the reaction between CoA and 11-residue ybbR-tag. Hereby, CoA is split in the middle of its diphosphate group. The phosphate located towards the cysteamine is connected to the hydroxymethyl sidechain of the first serine within the ybbR-tag. Adenosine 3',5'-biphosphate is produced as by-product. Sfp requires Mg^{2+} ions to function. The reaction is usually carried out in a 10 mM Tris buffer supplemented with 10 mM $MgCl_2$ at pH 7.5 for one hour at room temperature or 37°C.

eSrt, a pentamutant of wild-type sortase A, catalyzes the reaction between an LPETG-motif and an N-terminal glycine. Hereby, the glycine of the LPETG is cut away and replaced by the N-terminal one. Since everything beyond the glycine of the LPETG-motif is cut away, it is advisable to place the LPETG-motif at the very C-terminus of the protein that is to be connected. To prevent cross-linking, it is important that there is only one kind of protein with an N-terminal glycine – in the case of SA, only one subunit with an N-terminal glycine. In contrast to the Sfp reaction, the eSrt reaction is reversible in the sense that a newly established linkage can be cut again. This effect can be reduced by adding several glycines to the N-terminus. eSrt requires Ca^{2+} ions to function. The reaction is usually carried out in a 20 mM HEPES buffer supplemented with 150 mM NaCl and 1 mM $CaCl_2$ at pH 7.4 for about 30 minutes at room temperature. The incubation time is dependent on the eSrt and substrate concentration. The reaction is stopped by rinsing off eSrt and by exchanging the buffer with an EDTA-containing one to chelate remaining Ca^{2+} ions.

Biotinylation In the context of this thesis, two enzyme-mediated approaches have been used for biotinylation, the reaction of covalently connecting a biotin to another molecule. Because of their importance for research on SA, they are discussed separately.

In most cases, a commercial CoA-Biotin (New England Biolabs, Ipswich, USA) is applied, which in principle consists of a maleimide-PEG_2-biotin, where the maleimide has formed a thio-ether bond with the thiol group of CoA. PEG and biotin are connected by a peptide

bond, which has been formed by a reaction between biotin's carboxyl group and an amine group at the end of the short PEG linker. The aforementioned Sfp reaction is used to connect CoA-Biotin to a ybbR-tag, which is genetically fused to a protein domain.

An alternative method for biotinylation is the use of the enzyme BirA, which is able to attach a biotin to the peptide with the sequence GLNDIFEAQKIEWHE (AviTag) [81]. Hereby, a peptide bond is formed between biotin's carboxyl group and the amine group at the end of the side chain of the lysine residue within the AviTag. Biotinylation by BirA can be carried out *in vitro* [82] or *in vivo* [83]. For the *in vitro*-biotinylation, adenosine triphosphate, biotin and the enzyme BirA have to be added to the AviTag containing protein construct. The reaction is irreversible. Usually, the reaction is allowed to proceed for one hour at room temperature. Excess biotin has to be removed using Zeba Spin Desalting Columns (Thermo Fisher Scientific, Waltham, USA). For the *in vivo*-biotinylation, the competent *E.coli* strain CVB101, which contains the BirA gene on a pACYC184 vector, is used. Protein constructs to be biotinylated are cloned into a pAC4 vector that contains the AviTag. Biotin has to be added to the expression culture. Excess biotin is removed in the protein purification process.

As an alternative to biotinylation, Strep-Tag II [52] should be mentioned. This peptide with the sequence WSHPQFEK can be genetically fused to chimeric protein constructs. It has a much lower affinity for SA than biotin but is still a useful handle for force spectroscopy measurements [9].

Performing AFM-based SMFS

To perform an AFM-based SMFS experiment, the functionalized sample surface is mounted onto the xy-stage of the AFM setup. The functionalized cantilever is mounted on a glass cone using vacuum grease. A drop of measurement buffer is added to prevent the cantilever tip from drying out during mounting. The glass cone is then mounted on the AFM-head and placed in measurement buffer above an inverted microscope connected to a camera.

An infrared laser is focused on the back of the cantilever and reflected onto the PSD. The lateral deflection on the PSD can be adjusted manually with a micrometer screw moving the PSD. The vertical deflection, *i.e.* the one that is important for force measurements, can be adjusted by the control software using a small motor that changes the PSD's position. For the relaxed cantilever, the PSD position is adjusted such that both deflections are zero. If, due to thermal drift, the vertical deflection for the relaxed cantilever exceeds a certain deflection tolerance, the measurement software resets the PSD position.

The AFM-head is then moved to its measurement position above the sample surface. The sample surface can be moved with a long-range stage that allows reaching different spots on the surface, where different proteins have been immobilized – millimeters apart from each other. In addition, a short-range stage moves the sample surface in a snail-like pattern by a hundred nanometers within the same spot to expose a new surface area for every approach-retraction cycle of the cantilever tip. When the xy-stage is moved large distances the cantilever tip is retracted from the surface using the z-motor to prevent the tip from damage by unintentional surface contact.

The approach and retraction of the cantilever tip is conducted by the z-piezo that moves the AFM-head up and down. The nominal range of the z-piezo is 12 µm, in a normal measurement the distance between tip and surface is varied over less than 1 µm. The distance

Table 2.3: Typical measurement parameters for AFM-based SMFS. Approach velocity v_a, retraction velocity v_r, retraction distance d, dwell time t_{dwell}, approach sampling rate SR_a and retraction sampling rate SR_r for different measurement variations. The indentation force was usually set to 100 pN and then adjusted if necessary as the cantilever is only calibrated after the measurement.

v_a (nm/s)	v_r (nm/s)	d (nm)	t_{dwell} (s)	SR_a (Hz)	SR_r (Hz)
4000	200	350	0	12500	3000
4000	400	350	0	12500	6000
4000	800	350	0	12500	12000
4000	1600	350	0	12500	18000
4000	3200	350	0	12500	24000
4000	6400	350	0	12500	48000
4000	9600	350	0	12500	50000
4000	12800	350	0	12500	50000

can be changed with velocities ranging from 50 nm/s to 13 µm/s. For force spectroscopy measurements, the limiting factor is the read-out frequency of the PSD. Its maximum value is 50 kHz. The sampling rate has to be adjusted to the retraction velocity to obtain a reasonable amount of data points.

For the approach and retraction of the cantilever tip, not only velocities and sampling rates are set in the control software. In addition, the retraction distance and an indentation force have to be set. The cantilever tip is pressed into the surface until this trigger force is reached, then it is retracted immediately from the surface. Longer surface contact can be accomplished, if a dwell-time is set. Typical values for a force spectroscopy measurement are provided in Table 2.3.

An Exemplary Approach-Retraction Cycle The events occurring during an ideal approach-retraction cycle are depicted in Figure 2.3. The cantilever tip is pressed into the surface, a single bond between the molecules on the cantilever tip and the surface forms. Retracting the cantilever from the surface, different events occur: stretching of molecular linkers and unfolding of different domains is observed before the bond finally ruptures.

Single-molecule interaction is not observable for all approach-retraction cycles. In most cases, no specific interaction can be seen. In some cases, multiple interactions, with more than a single fingerprint unfolding pattern, are observed. These curves are excluded from data analysis. In a good AFM-based SMFS experiment about 7% of the approach-retraction cycles can be classified as specific single-molecule interactions.

A Typical Force-Extension Trace A typical force-extension trace, corresponding to the approach-retraction cycle described in the previous paragraph is depicted in Figure 2.4a. To illustrate the different events that occur when retracting the cantilever tip from the surface, the different parts of the force curve are highlighted in different colors: Green indicates the stretching of linker molecules. Unfolding events of the fingerprint domain are highlighted in blue, unbinding of the ligand from the receptor molecule in yellow. Parts where both unfolded fingerprint domains and linker molecules get stretched are displayed in cyan. The

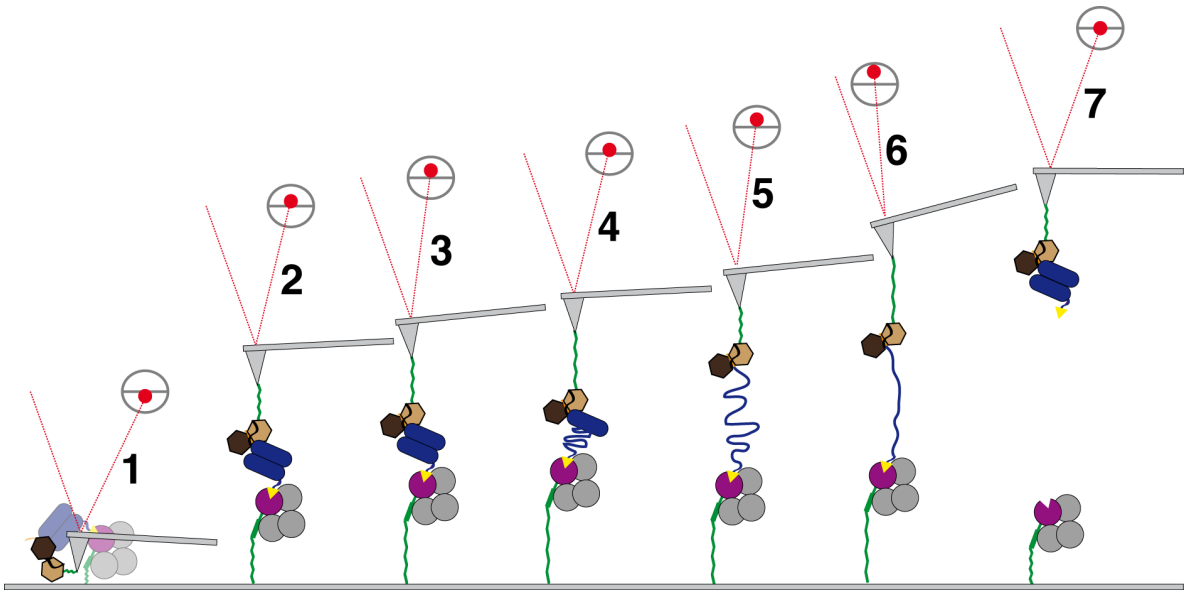


Figure 2.3: Illustration of a typical approach-retraction cycle. (1) The cantilever tip is pressed into the surface resulting in a negative deflection of the cantilever, such that the laser reflex on the PSD (red spot) is shifted downwards. A bond between the molecules on the cantilever tip and those on the surface is formed. (2) Retracting the cantilever from the surface, the molecular linker molecules (green) get stretched. The cantilever is now deflected in the opposite direction; the position of the laser reflex on the PSD is shifted upwards. (3) Retracting further, the force on the cantilever increases resulting in a larger deflection. The laser reflex on the PSD is shifted further. (4) At a certain force, the first part of the fingerprint domain (blue) unfolds, adding additional contour length to the system. Stress on the linkers is released and the force drops, resulting in a lower deflection of the cantilever. The laser spot on the PSD moves down a bit. Retracting the cantilever further, the second part of the fingerprint domain unfolds and releases additional contour length. (5) Now, both molecular linkers and the unfolded polypeptide chain of the fingerprint domain get stretched. (6) The force increases further, the laser reflex gets shifted further upwards on the PSD. (7) Finally, the non-covalent receptor ligand bond ruptures. Stress is released. The cantilever is no longer deflected. The laser reflex moves back to the middle of the PSD. The fingerprint domain folds back into its native state.

part where the cantilever tip is pressed into the surface and the final baseline, where no force is exerted on the cantilever, are depicted in black. From the force-extension trace, unbinding and unfolding forces are extracted. They are given by the local force maxima (force peaks).

Fitting a straight line through the baseline, the point of zero force is determined. The first intersection of the force-extension trace and the baseline yields the zero point of the extension axis. The loading rate for each unfolding or unbinding event, *i.e.* the increase of force over time, is determined from the force-extension trace by a linear fit to the last 3 nm just before the corresponding force peak (either for the two unfolding or the unbinding event). The loading rate is then given by $\dot{F} = \frac{\delta F}{\delta x} \cdot v_r$.

The contour length of the stretched polymer can be determined using polymer elasticity model. The worm-like chain (WLC) model as proposed by Bustamante *et al.* [84] is commonly used.

$$F(x) = \frac{k_B T}{l_p}\left(\frac{1}{4(1-\frac{x}{l_c})^2} + \frac{x}{l_c} - \frac{1}{4}\right) \tag{2.26}$$

Here, l_p and l_c are persistence and contour length of the polymer, k_B is the Boltzmann constant, and T the absolute temperature. The persistence length is the length over which the orientation of tangent vectors to the polymer is correlated. If a fixed value is chosen for l_p and T, the equation can be solved for $l_c(x, F)$ and every point of the force extension trace is mapped into contour length space [85]. For this, the temperature is set to $T = 298\,\mathrm{K}$ and the persistence is chosen as $l_p = 0.365\,\mathrm{nm}$, which corresponds to the distance between to carbon-α atoms in a fully stretched polypeptide. Binning the frequency of occurrence of different contour lengths, distinct contour length increments are detectable (*cf.* Figure 2.4b). Knowing the underlying molecular structure, these increments can be attributed to the unfolding of certain domains. For example, ddFLN4 consists of 101 amino acids so that the total contour length increment upon unfolding reads $101 \cdot 0.365\,\mathrm{nm} = 36.9\,\mathrm{nm}$. Adding the two increments in Figure 2.4b, yields 35 nm and thus agrees nicely.

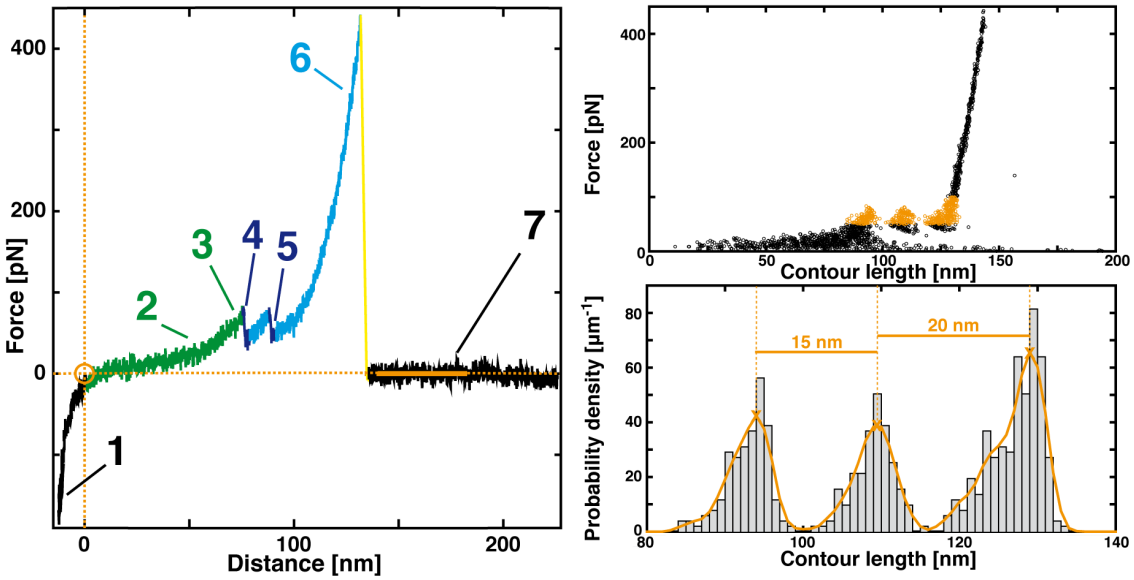
Force Measurement and Cantilever Calibration The raw data of a force-extension trace recorded by the AFM is the voltage applied to the z-piezo U_z and the deflection voltage measured on the PSD U_{defl}. To obtain useful physical units, these two voltages have to be translated into force F and extension x.

The z-piezo is calibrated using a gold surface mounted parallel to the relaxed cantilever, *i.e.* with an angle of 8°, because this is how the cantilever is mounted onto the glass cone. Defocusing the laser such that it is not only reflected off the back of the cantilever but also off the gold surface, an interference pattern is registered on the PSD when varying the z-piezo voltage, *i.e.* the z-position of the cantilever. Knowing the laser's wavelength, the z-sensitivity z_{sens} can be determined from the interference pattern. The distance between the cantilever tip and the surface z is then calculated by Equation 2.27.

$$z[\mathrm{nm}] = U_z[\mathrm{V}] \cdot z_{sens}[\frac{\mathrm{nm}}{\mathrm{V}}] \tag{2.27}$$

The zero point of the extension x is determined from the force-extension trace as the intersection of the force curve with the force baseline (*cf.* Figure 2.4a).

To translate U_{defl} into force F, the thermal method by Hutter *et al.* and its extensions by Butt *et al.* are used [86, 87]. Here, the cantilever is assumed as a one-dimensional Hookean



(a) An exemplary force-extension trace.

(b) Transformation into contour length space.

Figure 2.4: An exemplary force-extension trace and its transformation into contour length space. The numbers in (a) correspond to the events described in Figure 2.3. The green part of the force extension trace illustrates the stretching of the molecular linkers. The unfolding events of the fingerprint domain are shown in blue. The stretching of both unfolded fingerprint polypeptide chain and linker molecules is highlighted in cyan. The yellow line marks the unbinding of the ligand molecule from the receptor. The black part at the end of the curve is given by the signal from the PSD in the absence of force and therefore used to fit a baseline (orange line) to define the zero-point on the force axis. The first intersection of the force-extension trace with the baseline (marked by an orange circle) is used to set the zero-point on the extension axis. The negative forces in the beginning of the force curve (also marked in black) result from pushing the cantilever tip into the surface until the trigger point (here: about -180 pN) is reached. Using the WLC model, every (x, F)-point can be mapped into contour length space – upper panel in (b). The WLC model is best suited for forces between 50 pN and 100 pN, which are highlighted in orange. In the lower panel of (b), the occurrence of the corresponding contour lengths is given (grey bars). Data are fitted with a kernel density estimate (orange line). Local maxima (orange crosses) correspond to the unfolding and unbinding events in (a). Contour lengths increments of 15 nm and 20 nm (orange bars) upon unfolding of the fingerprint domain are detected.

spring. Using the equipartition theorem, it is possible to deduce the spring constant k of the cantilever from the Fourier transformation of the thermal fluctuations of the cantilever. In addition, the inverse optical lever sensitivity *InvOLS* has to be calibrated to convert the deflection voltage into a deflection distance. *InvOLS* is calibrated by pressing the cantilever firmly into the surface until a linear regime is reached, in which the deflection voltage is proportional to the z-piezo voltage, *i.e.* the position of the cantilever. In principle, fitting a straight line to this regime provides *InvOLS* (only a small correction factor is introduced to account for that the lever is not freely vibrating). The force F exerted on the cantilever is then calculated by

$$F[\mathrm{pN}] = U_{defl}[\mathrm{V}] \cdot InvOLS[\frac{\mathrm{nm}}{\mathrm{V}}] \cdot k[\frac{\mathrm{pN}}{\mathrm{nm}}]. \tag{2.28}$$

Since pressing the cantilever tip into the surface might destroy the tip's functionalization, the calibration is always performed after the measurement. At this point it should be noted that every cantilever has to be calibrated. Comparison between measurements performed with different tips are difficult, as the stiffness of the cantilever can slightly alter the measured forces [73]. To directly compare interactions between different proteins, it is thus advisable to measure them in a single multispot measurement with the same cantilever on the same surface.

2.3.3 Magnetic tweezers

Magnetic tweezers (MT) are an instrument to perform SMFS (*cf.* Section 4.2). They allow to apply low constant forces (0.001-100 pN [88]) over a long period of time (up to several days). In a MT experiment, the molecule on which force is exerted is clamped within a flow cell between a glass surface and a µm-sized magnetic bead using similar techniques as for AFM experiments (Section 2.2.6). Magnets placed above the flow cell exert an attractive force on the bead, thus pulling on the molecular chain clamped between bead and surface. The magnitude of the force can be altered by varying the distance between magnets and flow cell. For the correct relation between magnet position and exerted force, the instrument is calibrated measuring the thermal fluctuations of the bead and considering the equipartition theorem and Stokes' law for friction forces.

In an MT experiment, the position of the bead is tracked by a fast camera using an inverted microscope placed underneath the flow cell. The (vertical) z-position of the bead, which is equivalent to the extension in an AFM experiment, is the most important measurand. To determine the z-position, the bead is monitored slightly out of focus so that diffraction fringes are visible. Before the measurement, a look-up-table is created by altering the position of the objective while recording the bead's diffraction pattern at a constant force so that the position of the bead does not change. During the measurement, the position of the bead changes but the position of the objective is constant. Comparing the diffraction fringes with the look-up table, the z-position is obtained.

2.3.4 Total Internal Reflection Fluorescence Microscopy

Total Internal Reflection Fluorescence (TIRF) Microscopy is a technique to image fluorescently labeled samples on a surface. In fluorescence microscopy, the sample is exposed to light of a

certain wavelength. Fluorescent molecules, so-called fluorophores, absorb it and re-emit light of a longer wavelength. The excitation light is filtered out so that only the emission light reaches the detector. Excitation and emission wavelengths and the lifetime of the excited state are dependent on the type of fluorophore.

In TIRF, the excitation light is not guided to the fluorophores directly. Instead, the sample molecules, deposited on a thin glass slide, are illuminated from below at a large angle of incidence resulting in total reflection of the illumination light. An evanescent field excites only those fluorophores that are in close proximity of the surface (100-200 nm). By this, a good resolution along the optical axis is obtained, as the evanescent wave decays exponentially.

To achieve total reflection at the surface, large incident angles are needed. TIRF instrument requires a high numerical aperture of the objective. This is accomplished by the use of immersion oil.

In the context of this thesis, TIRF has been conducted in combination with AFM-based SMFS (*cf.* Section 4.1). A sophisticated technical design is used: AFM-based SMFS is performed with fluorescently labeled molecules immobilized on a thin glass surface, which is imaged with a four-color TIRF setup from below. These experiments are elaborate in their preparation, because a thin, thus fragile, glass surface is needed, and difficult to conduct, because the narrow TIRF focus level has to be fine-tuned and at the same time the AFM cantilever has to reach the surface but must not crash into it.

Part III

Results

Chapter 3

Single-Molecule Force Spectroscopy on the Streptavidin/Biotin Interaction

3.1 Publication P1: Monodisperse Measurement of the Biotin-Streptavidin Interaction Strength in a Well-Defined Pulling Geometry

Publication P1 describes how the SA/biotin interaction can be probed by AFM-based SMFS in a well-defined force-loading geometry. In particular, three decisive factors to obtain high quality SMFS data are identified: First, the usage of a site-specific covalent immobilization strategy to tether SA and its counterpart biotin to the cantilever tip and sample surface, respectively. Second, the application of a well-characterized fingerprint domain that enables to distinguish single-molecule unbinding events from unspecific or multiple interactions between the molecules on cantilever tip and sample surface. Third, the implementation of monovalent SA (mSA) to have full control over the tethering geometry, as mSA exhibits only a single functional SA subunit biotin can bind to.

To introduce mSA to the force spectroscopy community, the molecule is characterized in detail and compared with custom-made tetravalent SA as well as with a commercial SA using SDS PAGE. In addition, the binding behavior of the three SA variants is examined by ITC. It is demonstrated that the three versions of the molecule only differ in binding stoichiometry, while the binding enthalpy per biotin molecule is the same for all.

For the AFM-based SMFS experiments, mSA is immobilized on a surface by a unique cysteine, which has been introduced at the N-terminus of the functional subunit. The reaction of the thiol side-chain to a maleimide on the surface results in a covalent thioether bond. Therefore, mSA's force-loading geometry is well-defined. To the AFM cantilever tip, biotinylated ddFLN4 is covalently attached. The characteristic two-step unfolding pattern of this fingerprint domain is used to detect single-molecule interaction events. All force curves that do not show the pattern are excluded from the final data analysis. By this, a comparatively narrow force distribution (located at about 200 pN) is obtained. From the dynamic force spectrum, properties of the underlying energy landscape are extracted.

The comparison of the obtained SMFS data on mSA/biotin with previously published force spectroscopy data on SA/biotin suggests that SA's tetravalency most likely distorts the data reported in these studies. It is concluded that monodisperse data can only be measured using mSA in combination with site-specific immobilization, *i.e.* by realizing a well-defined pulling geometry.

Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry

by

Steffen M. Sedlak, Magnus S. Bauer, Carleen Kluger, Leonard C. Schendel, Lukas F. Milles, Diana A. Pippig, Hermann E. Gaub

published in

PLoS ONE 12(12):e0188722. doi:10.1371/journal.pone.0188722

PLOS ONE

RESEARCH ARTICLE

Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry

Steffen M. Sedlak, Magnus S. Bauer, Carleen Kluger, Leonard C. Schendel, Lukas F. Milles, Diana A. Pippig, Hermann E. Gaub*

Lehrstuhl für Angewandte Physik and Center for NanoScience (CeNS), Ludwig-Maximilians-Universität München, Munich, Germany

* gaub@lmu.de

OPEN ACCESS

Citation: Sedlak SM, Bauer MS, Kluger C, Schendel LC, Milles LF, Pippig DA, et al. (2017) Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. PLoS ONE 12(12): e0188722. https://doi.org/10.1371/journal.pone.0188722

Editor: Annalisa Pastore, National Institute for Medical Research, Medical Research Council, London, UNITED KINGDOM

Received: August 2, 2017

Accepted: November 7, 2017

Published: December 5, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Funding was provided by the European Research Council Advanced Grant (HEG) ERC-2011-ADG_20110310 to Hermann E. Gaub and the SFB1032 (HEG) of the Deutsche Forschungsgemeinschaft. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the

Abstract

The widely used interaction of the homotetramer streptavidin with the small molecule biotin has been intensively studied by force spectroscopy and has become a model system for receptor ligand interaction. However, streptavidin's tetravalency results in diverse force propagation pathways through the different binding interfaces. This multiplicity gives rise to polydisperse force spectroscopy data. Here, we present an engineered monovalent streptavidin tetramer with a single cysteine in its functional subunit that allows for site-specific immobilization of the molecule, orthogonal to biotin binding. Functionality of streptavidin and its binding properties for biotin remain unaffected. We thus created a stable and reliable molecular anchor with a unique high-affinity binding site for biotinylated molecules or nanoparticles, which we expect to be useful for many single-molecule applications. To characterize the mechanical properties of the bond between biotin and our monovalent streptavidin, we performed force spectroscopy experiments using an atomic force microscope. We were able to conduct measurements at the single-molecule level with 1:1-stoichiometry and a well-defined geometry, in which force exclusively propagates through a single subunit of the streptavidin tetramer. For different force loading rates, we obtained narrow force distributions of the bond rupture forces ranging from 200 pN at 1,500 pN/s to 230 pN at 110,000 pN/s. The data are in very good agreement with the standard Bell-Evans model with a single potential barrier at $\Delta x_0 = 0.38$ nm and a zero-force off-rate $k_{off,0}$ in the 10^{-6} s^{-1} range.

Introduction

With its low dissociation constant in the femtomolar range [1], its specificity, and its high stability under harsh conditions [2], the binding of the small molecule biotin to the homotetramer streptavidin (SA) is a popular and widely used tool in nanotechnology, biotechnology, and medicine. Especially after biotinylation became available [3], this receptor-ligand system found versatile applications, e.g. detection [4, 5] or capturing of biomolecules [6–9], and diverse other *in vivo* and *in vitro* methods. For single-molecule techniques, the tetravalency of

manuscript. SMS and MSB thank the Nanosystems Initiative Munich for support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

SA can however be disadvantageous, as it promotes clustering of biotinylated molecules. Single-molecule force spectroscopy (SMFS) [10], super-resolution imaging techniques, and analytical applications like surface plasmon resonance or switch sense technology [11] often require a 1:1 stoichiometry. Efforts have been directed at the development of monomeric versions of SA [12]. However, since the interplay between different subunits is important for the tight binding of biotin [13], monomeric SAs lack the outstanding affinity of wildtype SA [12]. In 2006, Howarth et al. [14] developed a tetrameric but monovalent streptavidin (mSA), by reconstituting one functional with three non-functional subunits (Fig 1A). mSA preserves femtomolar affinity towards biotin. Here, we present the implementation of mSA as a molecular anchor for atomic force microscopy (AFM)-based SMFS, which enables us to revisit the biotin:SA interaction in a very specific and monodisperse manner.

The interaction between biotin and tetravalent SA/avidin was the first receptor-ligand interactions probed by AFM-based SMFS [17–19]. It has become a model system for non-covalent receptor-ligand complexes and to study biorecognition processes [20]. In an AFM-based SMFS measurement, a functionalized AFM-cantilever decorated with ligand molecules is approached to a functionalized surface decorated with receptor molecules. A receptor-ligand complex is formed and when retracting the cantilever from the surface, the bending of the cantilever is recorded providing a measure for the force that the receptor-ligand complex can withstand, i.e. for its mechanical strength under load.

In 1994, Moy et al. [19] reported integer multiples of biotin:SA unbinding events and analyzed the relation between binding energies and unbinding forces. Biotinylated bovine serum albumin (BSA) was unspecifically adsorbed to both cantilever and sample surface. Bringing cantilever and surface in contact, SA that had been added to the solution could bind to a biotin on the cantilever and to one on the surface at the same time. Retracting the cantilever from the surface, the force needed to pull biotin and SA apart was recorded. The way load was applied to tetravalent SA in this experiment is schematically described in Fig 1B. Combinations of the geometries shown in this figure are also likely to occur. To obtain data at the single-molecule level, either the concentration of SA molecules was adjusted or free biotin was added to the solution.

Several groups independently repeated the experiment [18, 21]. Allen et al. slightly modified the setup by direct, yet unspecific, immobilization of SA to the sample surface [22]. In the following years, the biotin:SA interaction was modeled by MD simulations [23, 24] and theoretical descriptions for the process of unbinding were put forward [25–27]. In 1999, Merkel et al. [28] measured the biotin:SA interaction with a biomembrane force probe instrument. For the first time, measurements using different force loading rates were performed. On top of that, they introduced covalent attachment of biotin through polyethylene glycol (PEG) linkers. With a covalent immobilization strategy, detachment of biotin from the sample surfaces became unlikely, resulting in higher purity of the recorded data. The variety of possible pulling geometries, as depicted in Fig 1B, remained. Using the loading-rate dependence of rupture forces, the energy landscape of the biotin:SA binding was investigated. Dynamic force spectra of the receptor-ligand system were also recorded with the AFM using diverse attachment strategies, such as immobilization in a phospholipid bilayer [29] or a dextran-coated surface [30], by biotinylated BSA [31–33] or by cross-linking with glutaraldehyde [34]. In 2010, Taninaka et al. further improved the measurement procedure by binding both biotin and SA covalently with PEG spacers to sample and cantilever surface, respectively [35]. The way load is applied to the SA tetramer in this case is shown in Fig 1C.

Due to different ways the ligand binds to the receptor, AFM-based SMFS data can be dispersed when performing experiments using multivalent receptor molecules, such as SA, even if actual single-molecule interactions are probed. Pulling on the ligand, the force can propagate

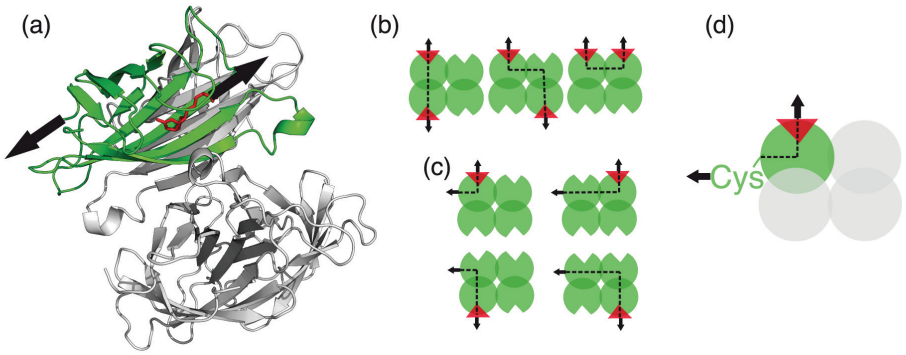


Fig 1. Possible pulling geometries for SA of different valencies. (a) Crystal structure of mSA (pdb identification code 5TO2 [15], overlaid with 1MK5 [16] to show the position of biotin). The functional subunit (green) with biotin (red) bound is stabilized by the three non-functional subunits (grey). Black arrows show the direction of the applied load for the AFM-based SMFS measurement. (b) Tetravalent SA consists of four functional subunits (green balls) each possessing a biotin (red triangles) binding site. In previous experiments, SA has been attached to a biotinylated surface resulting in a variety of possible pulling geometries: Across the strong interface, across the weak interface or diagonally across the tetramer. Having several functional binding pockets available, multiple binding to surface or cantilever can also occur. Black arrows indicate the pulling direction, black dotted lines possible ways force propagates through the molecule. (c) Attaching the tetravalent SA molecule covalently to the surface gives also rise to diverse pulling geometries. (d) In our experiments, we employ mSA consisting of one functional (green ball) and three non-functional subunits that are unable to bind biotin (grey balls). Having mSA tethered by a single N-terminal cysteine in the functional subunit, we pull biotin out of the binding pocket. The force only propagates through a single subunit.

https://doi.org/10.1371/journal.pone.0188722.g001

through the receptor molecule in different ways (Fig 1B and 1C). This results in a broad distribution of rupture forces. Furthermore, when the receptor molecule is composed of several non-covalently bound subunits, the data are distorted if the subunits of the receptor molecule get torn apart. In a SMFS experiment, a rupture of the receptor molecule itself cannot be distinguished from the unbinding of the ligand from the receptor. Beyond that, disrupted receptor tetramers may clog the cantilever thus preventing specific interaction resulting in low data yield.

From the crystal structure of wild-type SA, it can be reasoned that the SA monomers assemble into strongly associated dimers that form less stable tetramers [36]. Therefore, the different interfaces between the four subunits of a SA tetramer might be of different mechanical stability. Kim et al. [37] proved that the mechanical strength of the SA tetramer itself is highly dependent on the pulling geometry, i.e. on the way force is applied to the tetramer. Pulling on various control domains that were genetically fused to the N-termini of the SA monomers, they observed two distinct peaks in the distribution of rupture forces of the tetramer [37]. The two peaks can be assigned to a rupture across the strong interface between two subunits forming a dimer and to the rupture across the weak interface between the two dimers forming the tetramer. Interestingly, the force peaks of around 100 pN and 400-500 pN overlap with the range of unbinding forces reported for the biotin:SA interaction [18, 19, 21, 22, 28–32, 35, 38–40].

Non-equilibrium unbinding forces are loading rate dependent [41]. Any comparison of unbinding forces on an absolute scale, especially when measured with different setups under different conditions, is to be treated with caution. Nevertheless, it is conceivable that SMFS experiments with biotin and tetravalent SA are to some extend distorted by the potential rupture of the tetramer before unbinding biotin from SA. To examine the behavior of the biotin:

SA interaction under load, it is therefore important to overcome the problem of SA's tetravalency.

We therefore implement mSA to perform high-throughput AFM-based SMFS experiments for probing the mechanical stability of the biotin:SA system in a well-defined pulling geometry, no longer distorted by the receptor's multivalency. The quality of the data is further improved by the use of protein calibration domains for identification of single interactions. The unfolding patterns of the calibration domains that are enzymatically fused to ligand or receptor molecule verify single rupture events. When unfolding under the applied load before the receptor-ligand complex ruptures, they yield a specific unfolding force, which serves as internal reference for force calibration, and a defined length increment that is taken as an indicator for single receptor-ligand unbinding.

For site-selective immobilization of SA, we genetically modified the functional subunit of mSA. Although wildtype SA does not contain any cysteine residues, the SA tetramer was found to be of high stability under conditions, which are usually denaturing [42]. In contrast to many other proteins, the interaction between the subunits is not mediated by disulfide bridges but originates from a network of hydrogen bonds and hydrophobic interactions. We thus introduced a single cysteine at the N-terminus of the functional subunit of mSA for site-selective immobilization by conventional thiol-maleimide coupling [43]. We thereby created a stable molecular anchor for biotinylated (bio-)molecules with femtomolar affinity and well-defined stoichiometry. This well-defined single anchor point together with the monovalency of the biotin mSA interaction defines an unambiguous force propagation path. It enables us to perform AFM-based SMFS experiments in which the force only propagates through a single subunit of SA (Fig 1D).

Materials and methods

Gene construction, protein expression and purification

A detailed description of expression and purification is provided in the supplement (S1 Appendix). SA and mutant SA (deficient in biotin binding) constructs containing an N-terminal polyhistidine-tag (His-tag) for purification were cloned into pET vectors (Novagen, EMD Millipore, Billerica, USA). Constructs contained an N-terminal cysteine for site-specific immobilization, except for the subunits that were not meant to attach to AFM-cantilever surface or the glass coverslip. SA subunits with and without cysteine and His-tag and mutant SA subunits were expressed separately in *E. coli* BL21(DE3)-CodonPlus (Agilent Technologies, Santa Clara, USA). The constructs formed inclusion bodies that were isolated as described previously [44]. To reconstitute mSA and to guarantee a 1:3 ratio of functional to non-functional SA subunits in the final tetramer, inclusion bodies were solubilized in 6 M guanidine hydrochloride and then mixed at a 1:10 ratio prior to refolding and purification via the His-tag. To obtain tetravalent SA with a unique cysteine coupling site, the construct containing the cysteine residue as well as a His-tag was mixed with functional SA devoid of either.

The *Dictyostelium discoideum* fourth filamin domain (ddFLN4) construct with an N-terminal ybbR-tag [45] and a C-terminal cysteine (the internal cysteine 18 was mutated to serine) was cloned into pET vectors (Novagen, EMD Millipore, Billerica, USA). After expression in *E. coli* BL21(DE3)-CodonPlus (Agilent Technologies Santa Clara, USA) and lysis, purification was achieved by immobilized metal ion affinity chromatography (Ni-IMAC).

The superfolder green fluorescent protein (GFP) construct with an N-terminal cysteine and a C-terminal ybbR-tag was cloned into pET vectors (Novagen, EMD Millipore, Billerica, USA) and expressed in *E. coli* BL21(DE3)-CodonPlus (Agilent Technologies Santa Clara, USA). Purification was performed by Ni-IMAC.

Biotinylation of protein constructs

GFP and ddFLN4 constructs were biotinylated using the ybbR-tag/Sfp-Synthase system [45]. For the GFP construct, 18 μM GFP-ybbR were incubated with 60 μM CoA-Biotin (New England BioLabs) and 9 μM Sfp Synthase in a solution of 10 mM $MgCl_2$ and 50 mM HEPES at pH 7.5 for 1 h at 37˚C. To clean the solution from remaining CoA-Biotin, a buffer exchange to phosphate buffered saline (PBS; Sigma-Aldrich, Saint Louis, USA) was performed with Zeba Spin Desalting Columns (Thermo Scientific, Rockford, USA) with 7K MWCO according to the manufacturer's instructions. For the ddFLN4 construct, the incubation was performed at room temperature. All other steps were done in the same way as for GFP.

SDS-PAGE

Gel electrophoresis was performed using Any kD Mini-PROTEAN TGX Precast Protein Gels (Bio-Rad, Hercules, USA) in TRIS-based running buffer (2.5 mM TRIS, 200 mM glycerol, 3.5 mM SDS). For lanes 2–4, we heated 0.6 μM SA dissolved in loading buffer (50 mM TRIS, pH 8.0, 2.5% SDS, 5% glycerol, 0.005% bromophenol blue, 2.5% β-mercaptoethanol) for 5 minutes to 95˚C. For the other SA containing lanes, we used about 1.5 μM. For lanes 10–13, we added 1 μl of the purified Sfp reaction mixture containing both biotinylated and un-biotinylated GFP. We employed Precision Plus Unstained Protein Standards (Bio-Rad Laboratories, Hercules, USA) as molecular weight standards. The gel was run at room temperature with a constant current of 25 mA. The gel was analyzed with a ChemiDoc MP Imaging System (Bio-Rad Laboratories, Hercules, USA).

Isothermal titration calorimetry

The calorimetric experiments were carried out with a Malvern MicroCal ITC200 (Malvern, UK). SA samples were equilibrated with PBS using Zeba Spin Desalting Columns (Thermo Scientific, Rockford, USA) with 40K MWCO following the manufacturer's instructions. The concentration was determined by spectrophotometry with a NanoDrop 1000 (Thermo Scientific, Rockford, USA) using an extinction coefficient of ε_{280} = 167,760 $M^{-1}cm^{-1}$ calculated from the protein sequence using the SIB bioinformatics resource portal [46]. Biotin (Sigma-Aldrich, St. Louis, USA) was dissolved in PBS. For all measurement, the same stock solution of biotin was used. For mSA, a tenfold excess of biotin was titrated into the sample cell. For tetravalent SA, we used a ratio of 1:40, resulting in a final molar ratio of 1:8. All experiments were performed at 25˚C.

Functionalization of cantilevers and coverslips

AFM cantilevers (Biolever Mini, Olympus, Tokyo, Japan) and glass coverslips were silanized as described by Zimmermann et al. [43]. They were incubated with 25 mM heterobifunctional PEG (Rapp Polymere, Tübingen, Germany) with a molecular weight of 5 kDA equipped with an N-Hydroxysuccinimide (NHS) group and a maleimide group dissolved in a 50 mM HEPES solution at pH 7.5 for 45 minutes. The PEG spacers ensure passivation of glass cover slip and AFM-cantilevers and allow for specific sample immobilization. The coverslips were washed in ultrapure water and mounted into AFM sample holder. A 3.5 μl droplet of monovalent or tetravalent SA was deposited on the surface. The cantilevers were washed in ultrapure water and then placed in a 15 μl drop of the purified biotinylated ddFLN4 construct. For an efficient reaction of thiol with maleimide groups which forms stable thioester bonds, we reduced the thiol groups of SA and ddFLN4 construct by adding Immobilized TCEP Disulfide Reducing Gel (Thermo Scientific, Rockford, USA) in a v/v ratio of 1:6 and incubated for 1 h. The gel was

removed with the help of an Ultrafree-MC, HV 0.45 μm centrifugal filter (Merck Millipore, Darmstadt, Germany) directly before adding the proteins to the coverslips or cantilevers. During the formation of the thioester bonds, the samples were kept in a humidity chamber to prevent evaporation. After 1.5 h, the cantilevers were washed twice in PBS and the surfaces were rinsed with 50 ml PBS to flush out unbound protein.

AFM-based single-molecule force spectroscopy experiments

The experiments were performed with a custom-built AFM as described by Gumpp et al. [47]. The cantilevers were approached to the surface and after short contact, retracted at constant velocities of 200 nm/s, 800 nm/s, 2,000 nm/s, 5,000 nm/s, and 10,000 nm/s. To always probe a different spot on the surface, it was horizontally moved by 100 nm after each approach. For calibration of the cantilevers, we employed the equipartition theorem [48]. Baumann et al. [44] and Milles et al. [49] provide detailed descriptions of experimental SMFS procedures and SMFS data analysis.

Results and discussion

Size and functionality of mSA constructs with terminal cysteine is maintained

After expression and purification, we checked size and quality of the SAs with SDS polyacrylamide gel electrophoresis (Fig 2). Heating mSA and tetravalent SA (tSA) for 5 min to 95˚C, the tetramers fall apart into monomers of approximately 14 kDa (Fig 2B). The higher band can be assigned to the monomer with the additional His-tag and we confirmed the expected ratio between the monomers to be 1:3. Commercially available SA from *Streptomyces avidinii* (sSA) shows only one slightly larger and broader band. In contrast to the recombinantly expressed core SA monomer that consist of 123 residues, the SA monomer from *Streptomyces avidinii* contains 183 amino acids. In a posttranslational digest process, it is cut down to core SA.

The size of the tetramers can be estimated from unheated samples (Fig 2C). For mSA and tSA band size is slightly below the expected 54 kDa. Bands at double size are attributed to two tetramers connected via disulfide bridges between their cysteine residues. sSA shows several smeared out bands of larger size, caused by an incomplete posttranslational digest. The lowest one corresponds to core SA (54 kDa).

To illustrate the binding stoichiometry of the SAs to biotin, we added biotinylated GFP to mSA, tSA, and sSA (Fig 2D and 2E). Since the biotinylation of GFP has been incomplete, bands of unbound SA and bands of GFP without biotin are still visible. All SAs having a single GFP bound appear at the same size of about 70 kDa. Valencies of the different SA can be determined from the number of bands. For mSA, only one band with a single biotinylated GFP bound is seen. For sSA, four bands are clearly visible. Because of dimerized tetramers binding one or several biotinylated GFPs, additional bands appear for tSA.

Modifications of mSA do not change biotin binding properties

We compared the binding properties of our modified mSA with tSA and sSA by isothermal titration calorimetry (Fig 3). Because of the high affinity of biotin to SA, we could only conclude that the dissociation constant K_D is lower than 1 nM. The binding enthalpy per mole of added biotin (ΔH_{mSA} = -26 kcal/mol, ΔH_{tSA} = -25 kcal/mol, ΔH_{sSA} = -26 kcal/mol) and the binding stoichiometry (N_{mSA} = 0.95, N_{tSA} = 4.31, N_{sSA} = 4.31) confirmed that the functional subunit of our modified mSA is capable of binding biotin in the same manner as the subunits of sSA, while the binding of biotin to the mutated non-functional subunits is negligible. The

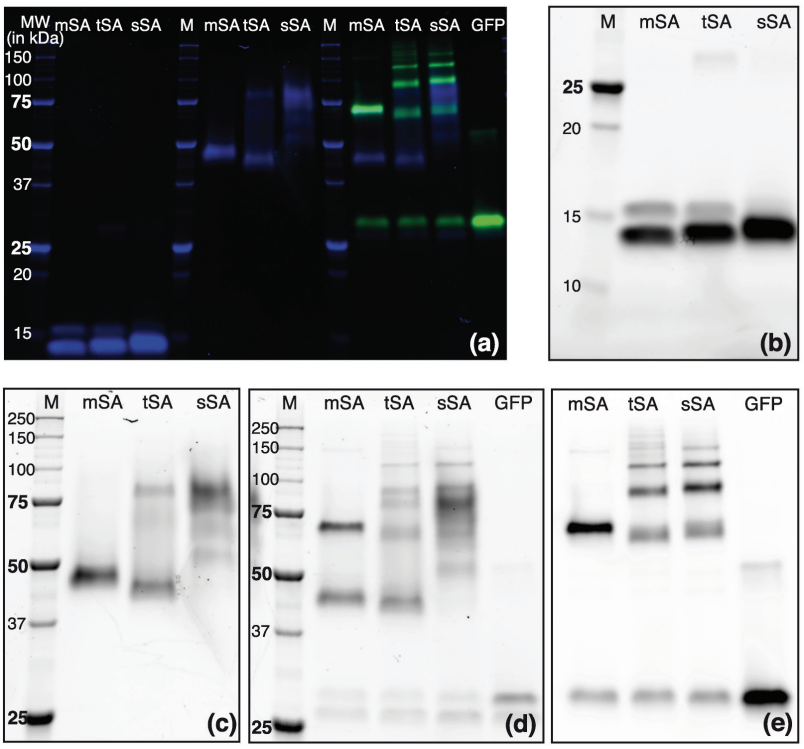


Fig 2. SDS-PAGE of mSA, tSA and commercial SA from *streptomyces avidinii* (sSA). (a) Overview of differently treated SAs with and without addition of biotinylated GFP on a stain-free polyacrylamide gel. Overlay of images taken with UV light excitation (blue) and illumination with a blue LED source (green). Parts of this image are inverted and shown in detail (b-d UV-excitation; e: GFP-channel): (b) Denatured SA samples (5 min at 95˚C). Decomposition into monomers (14 kDa) is visible. His-tagged subunits appear larger. sSA subunits are smeared out. (c) Untreated SA samples which maintain tertiary structure. (d,e) Addition of biotinylated GFP to untreated SA samples. Valencies of SAs are visible as different numbers of GFPs are bound. The lowest band in (d) corresponds to Sfp Synthase (26 kDa).

https://doi.org/10.1371/journal.pone.0188722.g002

measured enthalpies are also in line with previously reported values [50]. This implies that the modifications at the N-terminus of the functional subunit do not impede the binding of biotin. We therefore argue that structure and function of the sSA are preserved for our monovalent and tetravalent versions with N-terminal modifications.

AFM-based SMFS using mSA as a handle

Using reconstituted mSA in combination with a calibration domain, we were able to perform SMFS with a well-defined pulling geometry that are not distorted by SA's multivalency. In our experiments, force propagates only through a single subunit of the SA tetramer (Fig 1D). Therefore, no tension across any interface within the tetramer, which could cause dissociation of the tetramer into its subunits, is applied. The measurement process is illustrated in Fig 4. To ensure the specificity of the probed interaction, we used the unfolding pattern of biotinylated

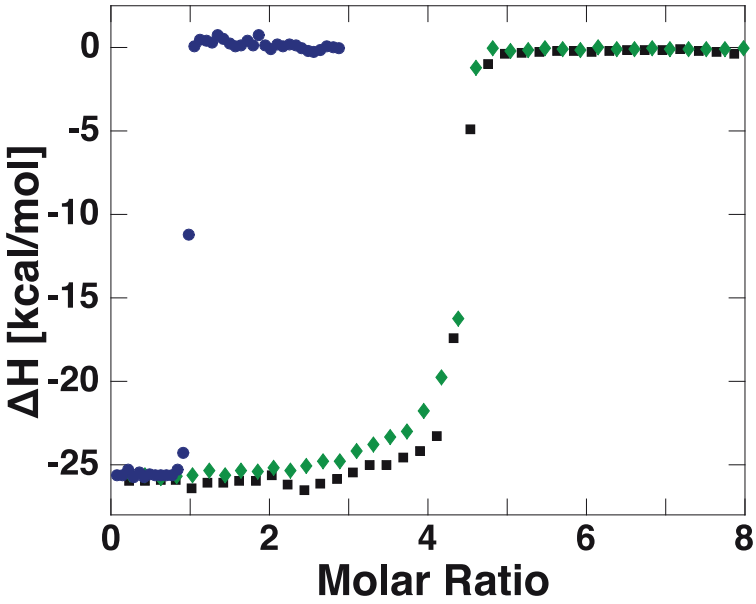


Fig 3. Isothermal titration calorimetry of biotin and SAs of different valency. The binding of biotin to different SAs was measured with isothermal titration calorimetry. The binding stoichiometry of mSA and biotin was determined as $N = 0.95$ (blue circles). The measured binding stoichiometry of the engineered tetravalent version (green diamonds) $N = 4.31$ is in good agreement with the value of commercial SA isolated from *Streptomyces avidinii* (black squares) $N = 4.29$. Within the limits of the measurement's accuracy, the binding enthalpies of the different SAs are the same ($\Delta H = -26$ kcal/mol for monovalent, $\Delta H = -25$ kcal/mol for tetravalent and $\Delta H = -26$ kcal/mol for commercial SA), confirming that the N-terminal modifications do not interfere with the binding of biotin.

https://doi.org/10.1371/journal.pone.0188722.g003

ddFLN4 [51] to identify single molecule rupture events. Because ddFLN4 folds back into its native state when the force drops after unbinding of biotin from mSA, it was used as a calibration domain on the cantilever, while mSA was immobilized on the surface. We use this attachment strategy for probing the biotin:mSA interaction, because we can probe a new mSA molecule, which has not yet been exposed to pulling forces, for every force-distance curve. Only those force curves that showed the specific unfolding pattern of the calibration domain were considered in subsequent data analysis procedures.

Analysis of AFM-based SMFS data

In an AFM experiment, about 5,000 force extension traces were recorded of which about 1,100 showed interaction. A larger data set of over 50,000 traces obtained in a 15 h measurement is shown in the supplement (S3 Appendix). To prove reliability and reproducibility of the control domain's unfolding pattern, an overlay of all 575 force-distance curves that feature the distinct unfolding pattern of ddFLN4 before biotin unbinds from mSA is shown in Fig 5A.

For every data bin along the extension axis, we selected the force bin with the highest value to obtain a characteristic force-extension curve. The curve consists of three parts: First, only the PEG-spacers on the cantilever and the surface are stretched (Fig 4). Then ddFLN4 unfolds in two distinct steps. Using the worm-like chain model for semi-flexible polymers [52] to fit

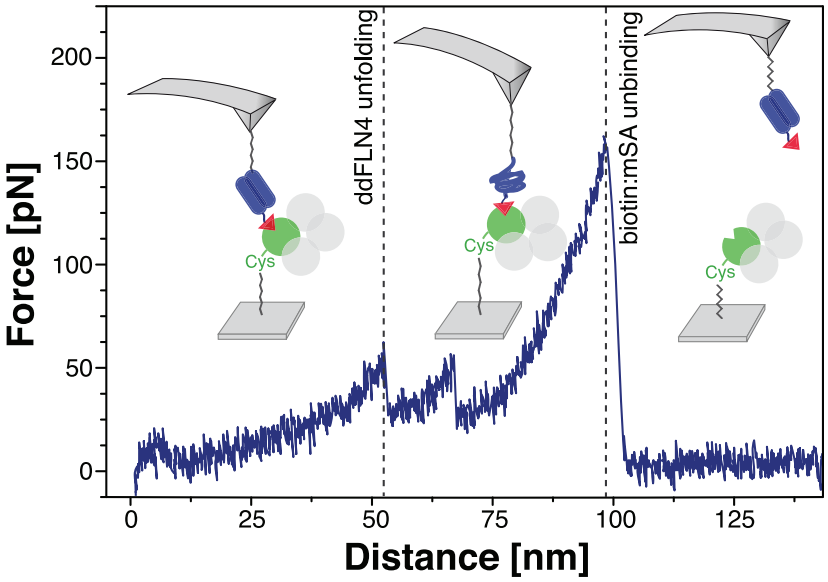


Fig 4. Investigation of the mechanical stability of the biotin:mSA binding with a well-defined pulling geometry. The functionalized cantilever tip is approached to the surface and a bond between biotin (red triangle) and mSA (green and gray balls) is formed. First, only the PEG (grey lines) spacers are stretched, when retracting the cantilever with constant speed from the surface. At forces of about 60 pN, the ddFLN4 (blue) unfolds in a characteristic two-step process that is used to identify single-molecule interactions. PEG spacers and the polypeptide chain are then further stretched until biotin unbinds from mSA under the applied load. The force drops and ddFLN4 folds back into its native state. As an example, one of the recorded force-distance curves (pulled at 800 nm/s) is shown in blue. More force-distance curves are shown in the supplement (S2 Appendix).

https://doi.org/10.1371/journal.pone.0188722.g004

this characteristic curve (black lines in Fig 5A), we deduced persistence lengths and contour lengths of the stretched construct for the different unfolding steps of the calibration domain. As the PEG-spacers undergo a conformational transition from cis to trans above forces of about 100 pN [53, 54] resulting in a linear force extension relation, we restricted the WLC fit to the part of the curve with forces lower than 100 pN. We find persistence lengths of 0.240 nm for the PEG-stretch, 0.265 nm and 0.282 nm for the subsequent parts. The fitted contour lengths of 80.7 nm, 96.4 nm, and 113.5 nm are in good agreement with theoretical estimations. From the molecular weights, we estimated the lengths of the two PEG-spacers to be about 31 nm to 40 nm each and the total contour length increment resulting from ddFLN4 unfolding to be 36 nm (S4 Appendix).

From the worm-like chain model, an expression for the contour length as a function of persistence length, force and extension can be derived [55]. Assuming a constant persistence length of 0.26 nm, we translated every data point of the characteristic curve (Fig 5A) into contour length space (S5 Appendix). In Fig 5B, the corresponding histogram of contour lengths is shown. Three pronounced peaks with maxima at 79.5 nm, 96.5 nm and 113.5 nm are visible, confirming the correct assignment of the different parts of the force-extension curve to different parts of our molecular construct.

We probed the biotin:mSA complex with five different retraction velocities (200 nm/s, 800 nm/s, 2,000 nm/s, 5,000 nm/s and 10,000 nm/s). The distributions of the resulting forces of the

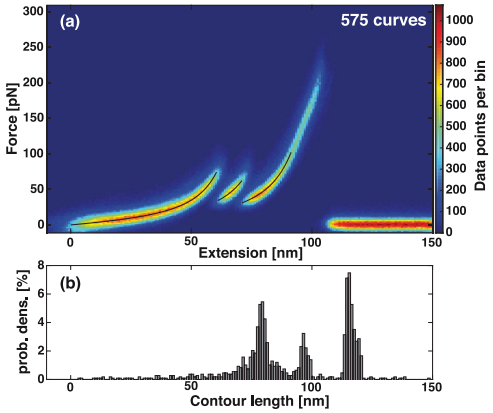


Fig 5. Overlay of force-extension curves and transformation into contour length space. (a) The 575 force-extension curves for which the characteristic unfolding pattern of ddFLN4 was visible are overlaid. We fit the three parts of the curve independently with the worm-like chain polymer model (black lines). (b) Using the mean persistence length of the worm-like chain fits, each point of the force extension curve is translated into contour length space. From the histogram, the contour lengths of the stretched constructs corresponding to the three parts of the force curve are determined.

https://doi.org/10.1371/journal.pone.0188722.g005

biotin:mSA unbinding and the ddFLN4 unfolding are depicted in Fig 6. The histograms of the forces corresponding to the two subsequent ddFLN4 unfolding steps exhibit defined peaks at 60-80 pN. For biotin:mSA unbinding force histograms, a sharp peak at about 200 pN is found. Its exact position depends on the applied loading rate. To obtain exact values, all force histograms were fitted with Bell-Evans models [25, 41] yielding the most probable rupture force, off-rates and distance to the transition state (S6 Appendix).

The dynamic force spectrum is shown in Fig 7. Force loading rates were determined by fitting a linear slope over the last 3 nm before unfolding and unbinding force peaks in the force-extension curves. In the semi-logarithmic plot, the centers of gravity of force and loading rate distributions for the ddFLN4 unfolding and the biotin:mSA unbinding are fitted by a straight line. This linear dependence of unfolding or rupture forces on the loading rate is given by Bell-Evans theory (S5 Appendix). From slope and y-intercept, the distance to transition state Δx_0 and the zero-force off-rate $k_{off,0}$ can be determined. For the ddFLN4-unfolding, we find $\Delta x_0 = (0.76 \pm 0.05)$ nm and $k_{off,0} = 8 \times 10^{-4}$ s^{-1} for the first unfolding peak and $\Delta x_0 = (0.56 \pm 0.02)$ nm and $k_{off,0} = 5 \times 10^{-2}$ s^{-1} for the subsequent peak. The distance to the transition state of the biotin:mSA unbinding reads $\Delta x_0 = (0.38 \pm 0.02)$ nm and the zero-force off-rate is determined as $k_{off,0} = 3 \times 10^{-6}$ s^{-1}. The off-rate is in good agreement with the value obtained in an off-rate assay ($k_{off,exp} = 6.1 \times 10^{-5}$ s^{-1}) [14]. Previous studies reported a kink in the force-loading rate dependence that was attributed to two potential barriers in the binding potential [28]. For the range of loading rates we applied and for the specific geometry that we used to load the complex, we could not observe this feature.

Conclusion

Even though binding of biotin to SA is widely used as a tool and has been extensively studied previously, the unbinding forces reported in the literature scatter substantially. With the development of mSA and progress in AFM-based SMFS it became possible to study the mechanical

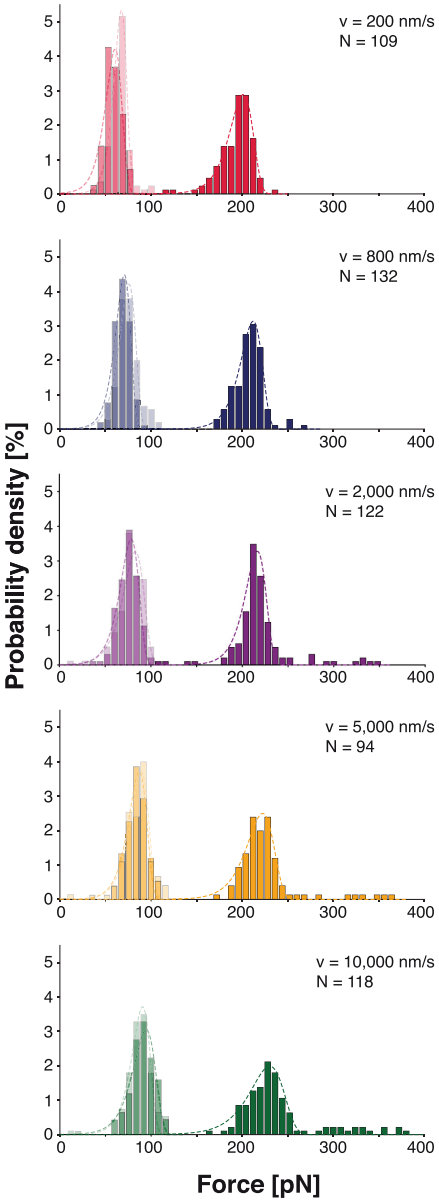


Fig 6. Unfolding forces of ddFLN4 and unbinding forces of biotin and mSA for different pulling velocities. The distribution of the forces of the first (transparent bars in the background) and second (semi-transparent bars) step of the ddFLN4 unfolding gives rise to two distinct peaks at approximately 85 pN and 75 pN. The biotin:mSA unbinding forces (opaque bars) are distributed more broadly but exhibit a clear maximum

at about 200 pN depending on the applied force loading rate. The experiment was carried out with a cantilever with a spring constant of 73.9 pN/nm. The dashed lines show independent fits of Bell-Evans distributions to the force histograms.

https://doi.org/10.1371/journal.pone.0188722.g006

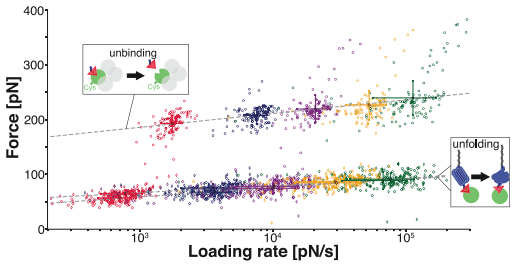


Fig 7. Bell-Evans plot of unfolding and unbinding forces. For all specific single-molecule interactions, the unbindig forces of biotin:mSA (circles) and the forces of the first (diamonds) and second (squares) step of the ddFLN4 unfolding are plotted against the loading rates at the corresponding force peak. The data are equal to the one shown in Fig 6 and the same color code is used. The dashed lines are linear fits to the centers of gravity (shown as filled circles, diamonds and squares) of the distributions of forces and loading rates, respectively. The colored crosses indicate the corresponding standard deviations.

https://doi.org/10.1371/journal.pone.0188722.g007

stability of the biotin:SA complex in a better defined way. Relating to previous measurements of the unbinding of biotin from tetravalent SA, we illustrated how multivalency of receptor molecules can distort SMFS data of receptor-ligand unbinding. We presented AFM-based SMFS data of the unbinding of biotin from monovalent SA with a 1:1-stoichiometry in a distinct pulling geometry, in which the force only propagated through a single subunit of the SA tetramer. The main improvements of our measurements contributing to the high quality of our data are covalent immobilization of both receptor and ligand molecules, the use of a calibration domain to verify single-molecule interaction events, and exact control over the attachment geometry by a single distinct anchoring site and monovalent receptor molecules.

Beyond that, we introduced a new tethering strategy for the use of mSA not only in force spectroscopy but also in many other single-molecule applications. The immobilization of mSA by implementing a single cysteine at the terminus of the functional subunit provides an anchoring site for sulfhydryl-reactive chemical groups, i.e. an anchoring site that is orthogonal to the interaction with biotin. In contrast to defined divalent SA [56] that can serve as a molecular hub for biotinylated molecules, mSA engineered with a single terminal cysteine on the functional subunit allows for controlled immobilization of biotinylated biomolecules or nanoparticles providing a 1:1-binding site.

Supporting information

S1 Appendix. Streptavidin preparation.
(PDF)

S2 Appendix. Exemplary force-distance curves.
(PDF)

S3 Appendix. Long-term SMFS measurement.
(PDF)

S4 Appendix. Estimating the contour lengths of PEG and ddFLN4.
(PDF)

S5 Appendix. Formulas.
(PDF)

S6 Appendix. Fitted Bell-Evans distributions shown in Fig 6.
(PDF)

S7 Appendix. Sequences of protein constructs.
(PDF)

S8 Appendix. Measuring with mSA immobilized on the cantilever.
(PDF)

Acknowledgments

The authors thank M. A. Jobst for discussions and the AFM control software, F. Baumann for support with the AFM experiments, and A. Kardinal and T. Nicolaus for laboratory support.

Author Contributions

Conceptualization: Steffen M. Sedlak, Diana A. Pippig.

Data curation: Steffen M. Sedlak, Leonard C. Schendel.

Formal analysis: Steffen M. Sedlak, Magnus S. Bauer, Carleen Kluger, Leonard C. Schendel, Diana A. Pippig.

Funding acquisition: Hermann E. Gaub.

Investigation: Steffen M. Sedlak, Magnus S. Bauer, Carleen Kluger, Leonard C. Schendel.

Methodology: Steffen M. Sedlak, Magnus S. Bauer, Lukas F. Milles.

Project administration: Diana A. Pippig, Hermann E. Gaub.

Resources: Lukas F. Milles, Diana A. Pippig, Hermann E. Gaub.

Software: Steffen M. Sedlak, Magnus S. Bauer, Lukas F. Milles.

Supervision: Diana A. Pippig, Hermann E. Gaub.

Validation: Steffen M. Sedlak, Carleen Kluger.

Visualization: Steffen M. Sedlak, Magnus S. Bauer, Carleen Kluger.

Writing – original draft: Steffen M. Sedlak, Carleen Kluger, Diana A. Pippig, Hermann E. Gaub.

Writing – review & editing: Steffen M. Sedlak.

References

1. Green NM. Avidin and streptavidin. Methods Enzymol. 1990; 184:51–67. PMID: 2388586.
2. Gonzalez M, Argarana CE, Fidelio GD. Extremely high thermal stability of streptavidin and avidin upon biotin binding. Biomol Eng. 1999; 16(1–4):67–72. PMID: 10796986.
3. Bayer EA, Zalis MG, Wilchek M. 3-(N-Maleimido-propionyl)biocytin: a versatile thiol-specific biotinylating reagent. Anal Biochem. 1985; 149(2):529–36. PMID: 3935007.

4. Howarth M, Takao K, Hayashi Y, Ting AY. Targeting quantum dots to surface proteins in living cells with biotin ligase. Proc Natl Acad Sci U S A. 2005; 102(21):7583–8. https://doi.org/10.1073/pnas.0503125102 PMID: 15897449; PubMed Central PMCID: PMCPMC1129026.
5. Spate AK, Dold JE, Batroff E, Schart VF, Wieland DE, Baudendistel OR, et al. Exploring the Potential of Norbornene-Modified Mannosamine Derivatives for Metabolic Glycoengineering. Chembiochem. 2016; 17(14):1374–83. https://doi.org/10.1002/cbic.201600197 PMID: 27147502.
6. Aschenbrenner D, Baumann F, Milles LF, Pippig DA, Gaub HE. C-5 Propynyl Modifications Enhance the Mechanical Stability of DNA. Chemphyschem. 2015; 16(10):2085–90. https://doi.org/10.1002/cphc.201500193 PMID: 25982589.
7. Ta H, Keller J, Haltmeier M, Saka SK, Schmied J, Opazo F, et al. Mapping molecules in scanning far-field fluorescence nanoscopy. Nat Commun. 2015; 6:7977. https://doi.org/10.1038/ncomms8977 PMID: 26269133; PubMed Central PMCID: PMCPMC4557268.
8. Kriegel F, Ermann N, Forbes R, Dulin D, Dekker NH, Lipfert J. Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers. Nucleic Acids Res. 2017. https://doi.org/10.1093/nar/gkx280 PMID: 28460037.
9. Walder R, LeBlanc MA, Van Patten WJ, Edwards DT, Greenberg JA, Adhikari A, et al. Rapid Characterization of a Mechanically Labile alpha-Helical Protein Enabled by Efficient Site-Specific Bioconjugation. J Am Chem Soc. 2017; 139(29):9867–75. https://doi.org/10.1021/jacs.7b02958 PMID: 28677396.
10. 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(1):3–12. https://doi.org/10.1016/j.jsb.2016.02.011 PMID: 26873782.
11. Knezevic J, Langer A, Hampel PA, Kaiser W, Strasser R, Rant U. Quantitation of affinity, avidity, and binding kinetics of protein analytes with a dynamically switchable biosurface. J Am Chem Soc. 2012; 134(37):15225–8. https://doi.org/10.1021/ja3061276 PMID: 22946661.
12. Qureshi MH, Wong SL. Design, production, and characterization of a monomeric streptavidin and its application for affinity purification of biotinylated proteins. Protein Expr Purif. 2002; 25(3):409–15. PMID: 12182820.
13. Sano T, Cantor CR. Intersubunit contacts made by tryptophan 120 with biotin are essential for both strong biotin binding and biotin-induced tighter subunit association of streptavidin. Proc Natl Acad Sci U S A. 1995; 92(8):3180–4. PMID: 7724536; PubMed Central PMCID: PMCPMC42129.
14. Howarth M, Chinnapen DJ, Gerrow K, Dorrestein PC, Grandy MR, Kelleher NL, et al. A monovalent streptavidin with a single femtomolar biotin binding site. Nat Methods. 2006; 3(4):267–73. https://doi.org/10.1038/NMETHXXX PMID: 16554831; PubMed Central PMCID: PMCPMC2576293.
15. Zhang M, Biswas S, Deng W, Yu H. The Crystal Structure of Monovalent Streptavidin. Sci Rep. 2016; 6:35915. https://doi.org/10.1038/srep35915 PMID: 28000673; PubMed Central PMCID: PMCPMC5175265.
16. Hyre DE, Le Trong I, Merritt EA, Eccleston JF, Green NM, Stenkamp RE, et al. Cooperative hydrogen bond interactions in the streptavidin-biotin system. Protein Sci. 2006; 15(3):459–67. https://doi.org/10.1110/ps.051970306 PMID: 16452627; PubMed Central PMCID: PMCPMC2249767.
17. Florin EL, Moy VT, Gaub HE. Adhesion forces between individual ligand-receptor pairs. Science. 1994; 264(5157):415–7. PMID: 8153628.
18. Lee G, Kidwell D, Colton R. Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. Langmuir. 1994; 10(2):354–7. doi: citeulike-article-id:3733610 https://doi.org/10.1021/la00014a003
19. Moy VT, Florin EL, Gaub HE. Intermolecular forces and energies between ligands and receptors. Science. 1994; 266(5183):257–9. PMID: 7939660.
20. Wilchek M, Bayer EA, Livnah O. Essentials of biorecognition: the (strept)avidin-biotin system as a model for protein-protein and protein-ligand interaction. Immunol Lett. 2006; 103(1):27–32. https://doi.org/10.1016/j.imlet.2005.10.022 PMID: 16325268.
21. Chilkoti A, Boland T, Ratner BD, Stayton PS. The relationship between ligand-binding thermodynamics and protein-ligand interaction forces measured by atomic force microscopy. Biophys J. 1995; 69(5):2125–30. https://doi.org/10.1016/S0006-3495(95)80083-4 PMID: 8580356; PubMed Central PMCID: PMCPMC1236446.
22. Allen S, Davies J, Dawkes AC, Davies MC, Edwards JC, Parker MC, et al. In situ observation of streptavidin-biotin binding on an immunoassay well surface using an atomic force microscope. FEBS Lett. 1996; 390(2):161–4. PMID: 8706850.
23. Grubmuller H, Heymann B, Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science. 1996; 271(5251):997–9. PMID: 8584939.

24. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J. 1997; 72(4):1568–81. https://doi.org/10.1016/S0006-3495(97)78804-0 PMID: 9083662; PubMed Central PMCID: PMCPMC1184352.

25. Evans E, Ritchie K. Dynamic strength of molecular adhesion bonds. Biophys J. 1997; 72(4):1541–55. https://doi.org/10.1016/S0006-3495(97)78802-7 PMID: 9083660; PubMed Central PMCID: PMCPMC1184350.

26. Friedsam C, Wehle AK, Kühner F, Gaub HE. Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. Journal of Physics: Condensed Matter. 2003; 15(18):S1709.

27. 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(41):15755–60. https://doi.org/10.1073/pnas.0806085105 PMID: 18852468; PubMed Central PMCID: PMCPMC2572921.

28. Merkel R, Nassoy P, Leung A, Ritchie K, Evans E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. Nature. 1999; 397(6714):50–3. https://doi.org/10.1038/16219 PMID: 9892352.

29. Wong J, Chilkoti A, Moy VT. Direct force measurements of the streptavidin-biotin interaction. Biomol Eng. 1999; 16(1–4):45–55. PMID: 10796984.

30. Stevens MM, Allen S, Davies MC, Roberts CJ, Schacht E, Tendler SJB, et al. The Development, Characterization, and Demonstration of a Versatile Immobilization Strategy for Biomolecular Force Measurements. Langmuir. 2002; 18(17):6659–65. https://doi.org/10.1021/la0202024

31. Yuan C, Chen A, Kolb P, Moy VT. Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy. Biochemistry. 2000; 39(33):10219–23. PMID: 10956011.

32. Lo Y-S, Zhu Y-J, Beebe TP. Loading-Rate Dependence of Individual Ligand−Receptor Bond-Rupture Forces Studied by Atomic Force Microscopy. Langmuir. 2001; 17(12):3741–8. https://doi.org/10.1021/la001569g

33. Rico F, Moy VT. Energy landscape roughness of the streptavidin-biotin interaction. J Mol Recognit. 2007; 20(6):495–501. https://doi.org/10.1002/jmr.841 PMID: 17902095.

34. de Odrowaz Piramowicz M, Czuba P, Targosz M, Burda K, Szymonski M. Dynamic force measurements of avidin-biotin and streptavdin-biotin interactions using AFM. Acta Biochim Pol. 2006; 53(1):93–100. PMID: 16410837.

35. Taninaka A, Takeuchi O, Shigekawa H. Reconsideration of dynamic force spectroscopy analysis of streptavidin-biotin interactions. Int J Mol Sci. 2010; 11(5):2134–51. https://doi.org/10.3390/ijms11052134 PMID: 20559507; PubMed Central PMCID: PMCPMC2885099.

36. Laitinen OH, Nordlund HR, Hytonen VP, Kulomaa MS. Brave new (strept)avidins in biotechnology. Trends Biotechnol. 2007; 25(6):269–77. https://doi.org/10.1016/j.tibtech.2007.04.001 PMID: 17433846.

37. Kim M, Wang CC, Benedetti F, Rabbi M, Bennett V, Marszalek PE. Nanomechanics of streptavidin hubs for molecular materials. Adv Mater. 2011; 23(47):5684–8. https://doi.org/10.1002/adma.201103316 PMID: 22102445; PubMed Central PMCID: PMCPMC3837471.

38. Lo Y-S, Huefner ND, Chan WS, Stevens F, Harris JM, Beebe TP. Specific Interactions between Biotin and Avidin Studied by Atomic Force Microscopy Using the Poisson Statistical Analysis Method. Langmuir. 1999; 15(4):1373–82. https://doi.org/10.1021/la981003g

39. Chivers CE, Crozat E, Chu C, Moy VT, Sherratt DJ, Howarth M. A streptavidin variant with slower biotin dissociation and increased mechanostability. Nat Methods. 2010; 7(5):391–3. https://doi.org/10.1038/nmeth.1450 PMID: 20383133; PubMed Central PMCID: PMCPMC2862113.

40. Hu Q, Yang H, Wang Y, Xu S. Quantitatively resolving multivalent interactions on a macroscopic scale using force spectroscopy. Chem Commun (Camb). 2016; 52(18):3705–8. https://doi.org/10.1039/c5cc10535h PMID: 26864087; PubMed Central PMCID: PMCPMC4767602.

41. Bell GI. Models for the specific adhesion of cells to cells. Science. 1978; 200(4342):618–27. PMID: 347575.

42. Bayer EA, Ehrlich-Rogozinski S, Wilchek M. Sodium dodecyl sulfate-polyacrylamide gel electrophoretic method for assessing the quaternary state and comparative thermostability of avidin and streptavidin. Electrophoresis. 1996; 17(8):1319–24. https://doi.org/10.1002/elps.1150170808 PMID: 8874057.

43. 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(6):975–85. https://doi.org/10.1038/nprot.2010.49 PMID: 20448543.

44. Baumann F, Bauer MS, Milles LF, Alexandrovich A, Gaub HE, Pippig DA. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. Nat Nanotechnol. 2016; 11(1):89–94. https://doi.org/10.1038/nnano.2015.231 PMID: 26457965.

45. 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(44):15815–20. https://doi.org/10.1073/pnas.0507705102 PMID: 16236721; PubMed Central PMCID: PMCPMC1276090.

46. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012; 40(Web Server issue):W597–603. https://doi.org/10.1093/nar/gks400 PMID: 22661580; PubMed Central PMCID: PMCPMC3394269.

47. Gumpp H, Stahl SW, Strackharn M, Puchner EM, Gaub HE. Ultrastable combined atomic force and total internal reflection fluorescence microscope [corrected]. Rev Sci Instrum. 2009; 80(6):063704. https://doi.org/10.1063/1.3148224 PMID: 19566207.

48. Hutter JL, Bechhoefer J. Calibration of atomic-force microscope tips. Review of Scientific Instruments. 1993; 64(7):1868–73. https://doi.org/10.1063/1.1143970

49. 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. 2016. https://doi.org/10.1021/acs.jpcb.6b09593 PMID: 27991799.

50. Chilkoti A, Stayton PS. Molecular Origins of the Slow Streptavidin-Biotin Dissociation Kinetics. Journal of the American Chemical Society. 1995; 117(43):10622–8. https://doi.org/10.1021/ja00148a003

51. 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(1):81–5. https://doi.org/10.1038/nsmb705 PMID: 14718927.

52. Bustamante C, Marko JF, Siggia ED, Smith S. Entropic elasticity of lambda-phage DNA. Science. 1994; 265(5178):1599–600. PMID: 8079175.

53. 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(1):6.

54. 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(6):6346–54. https://doi.org/10.1021/acsnano.7b02694 PMID: 28591514.

55. Puchner EM, Franzen G, Gautel M, Gaub HE. Comparing proteins by their unfolding pattern. Biophys J. 2008; 95(1):426–34. https://doi.org/10.1529/biophysj.108.129999 PMID: 18550806; PubMed Central PMCID: PMCPMC2426622.

56. Fairhead M, Krndija D, Lowe ED, Howarth M. Plug-and-play pairing via defined divalent streptavidins. J Mol Biol. 2014; 426(1):199–214. https://doi.org/10.1016/j.jmb.2013.09.016 PMID: 24056174; PubMed Central PMCID: PMCPMC4047826.

S1 Appendix.

Streptavidin preparation

Streptavidin Cloning

SA variants were obtained by site-directed mutagenesis of plasmids encoding Strep-Tactin constructs, whose sequence is similar to streptavidin [1], using a polymerase chain reaction and subsequent blunt-end ligation. By DNA sequencing (Eurofins Genomics, Ebersberg, Germany), we checked all final open reading frames.

Streptavidin Expression

The different SA subunits were expressed separately in *E.coli* BL21(DE3)-CodonPlus cells (Agilent Technologies, Santa Clara, USA). Plasmids encoding for different SA constructs, were transferred into *E.coli* BL21(DE3)-CodonPlus cells. Cells were grown at 37°C in pure LB Medium to build up antibiotic resistance, spread on an agar plate containing the appropriate antibiotic, and grown for 18 h at 37°C. We inoculated a preculture (8 ml LB medium, 1:1000 antibiotic) and grew the cells for 15 h at 37°C. We added preculture to the expression medium (500 ml SB medium with 20 mM KH_2PO_4 and 1:1000 antibiotic) until an optical density (absorbance at 600 nm) $OD_{600} = 0.1$ was reached. The expression culture was grown at 37°C until the optical density read $OD_{600} = 0.8$. After adding 1:5000 IPTG, the culture was grown for 15 h at 18°C. Then, it was centrifuged at 24,000 × g for 15 min. A bacterial pellet formed and was stored at -80 °C.

Streptavidin Purification

During all steps, samples were kept at 4 °C or on ice, respectively. Bacterial pellets for functional and non-functional subunits were weighed and then lysed separately in 5 ml Bacterial Protein Extraction Reagent (B-PER; Thermo Scientific, Rockford, USA) per gram bacterial pellet. We added 1 mg Lysozyme (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) and 50 µg DNase I (Roche Diagnostics GmbH, Mannheim, Germany) per gram bacterial pellet and placed the tube with the solution for 20 min on a rolling shaker. To lyse the bacteria completely, each of the dissolved pellets was sonicated. We then centrifuged the solutions with 60,000 × g for 30 min. As our protein formed inclusion bodies, we discarded the supernatants and resuspended each pellet in lysis buffer (PBS, 1 mM DTT, 0.1 % Triton X-100, pH 7.4). Sonication, centrifugation and resuspension steps were repeated until the supernatants were clear solutions. Each pellet was then resuspended in a denaturation buffer (PBS, 6 M guanidine hydrochloride, pH 7.5), sonicated and centrifuged. We kept the supernatants and measured the absorption at 280 nm. The solutions were then mixed in a ratio of 1:10 (functional subunits with His-tag to non-functional subunits) according to the measured absorption. We slowly pipetted the mixture into 500 ml of refolding buffer (PBS, 10 mM β-mercaptoethanol, pH 7.4) and placed it on a magnetic stirrer for 15 h.

The solution was centrifuged at 14,000 × g for 10 min. The supernatant was filtered through a hydrophilic 0.22 µm MF-Millipore Membrane and loaded on a 5 ml HisTrap FF (GE Healthcare, Little Chalfont, UK) that had been equilibrated with binding buffer (PBS, 10 mM imidazole, pH 7.4). After washing the loaded column with binding buffer, the recovery of the protein was accomplished using a gradient elution (elution buffer: PBS, 250 mM imidazole, pH 7.4). The flow through was fractionated. Fractions were analyzed using absorption

S1 Appendix.

spectroscopy and gel electrophoresis. Fractions containing SA were dialyzed against PBS and stored at 4 °C.

References

1. Baumann F, Bauer MS, Milles LF, Alexandrovich A, Gaub HE, Pippig DA. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. Nat Nanotechnol. 2016;11(1):89-94. doi: 10.1038/nnano.2015.231. PubMed PMID: 26457965.

S2. Appendix.

Exemplary force-distance curves

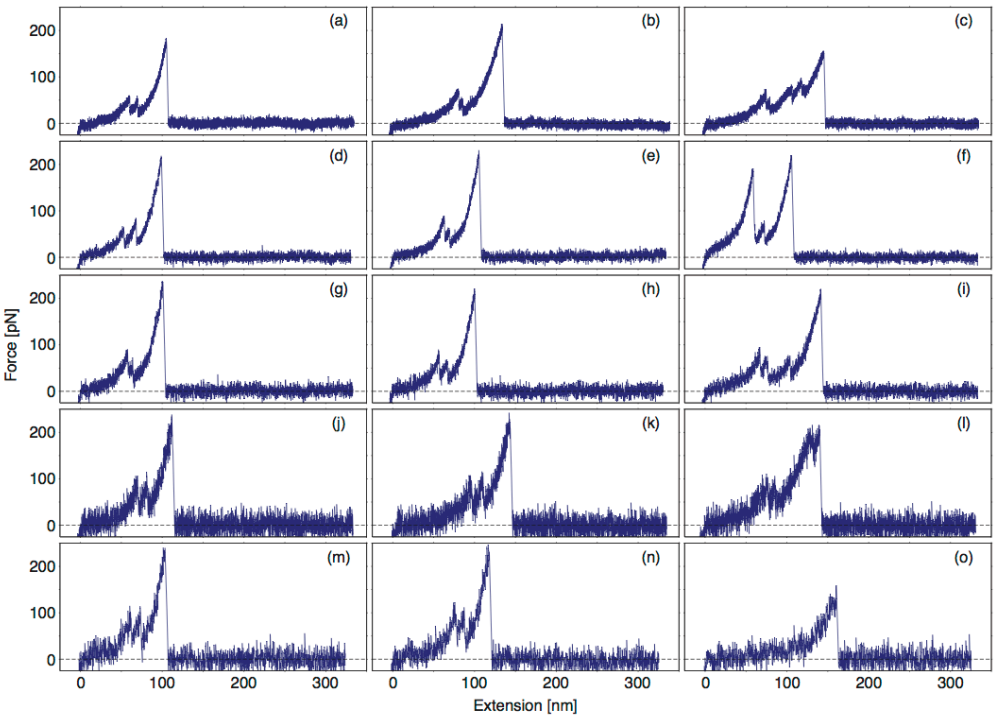


Fig A. Exemplary force-distance curves. Force-extension data recorded for different pulling velocities: (a-c) 200 nm/s, (d-f) 800 nm/s, (g-i) 2,000 nm/s, (j-l) 5,000 nm/s, and (m-o) 10,000 nm/s. The curves depicted in the left and in the middle column show a clear unfolding pattern of the calibration domain indicating specific single-molecule interaction. The curves depicted in the right column show interaction, but no clear unfolding pattern of the calibration domain is visible. These curves were thus not considered for further evaluation. The curves in (c), (i), and (l) are most probably caused by interaction of more than one biotin:mSA pair. For the curve in (f), a ddFLN4-like pattern is visible, but the unfolding force of the calibration domain is too high. Unspecific sticking of PEG or pulling with unfolded ddFLN4 may have caused the curve shown in (o).

S3 Appendix.

Long-term SMFS measurement

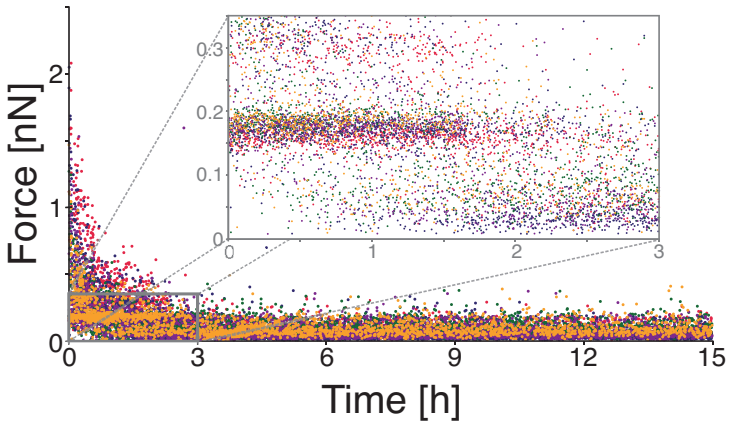


Fig A. Interaction of cantilever and surface over the course of the measurement. The force of the last peak in all force-extension curves that showed interaction between cantilever and surface are plotted over time. The different colors correspond to the different retraction velocities, with the color-coding being the same as in the main text. At the beginning of the measurement, multiple interactions give rise to high rupture forces. During the first 2.5 h (inset), a lot of specific single-molecule interactions are present resulting in a band of colored circles at about 200 pN. Wear out effects of cantilever and surface functionalization cause an increase of unspecific low-force interaction. For some of these, ddFLN4 unfolding is seen causing a small but broad unbinding peak at 100-160 pN in the histogram of rupture forces (Fig B).

S3 Appendix.

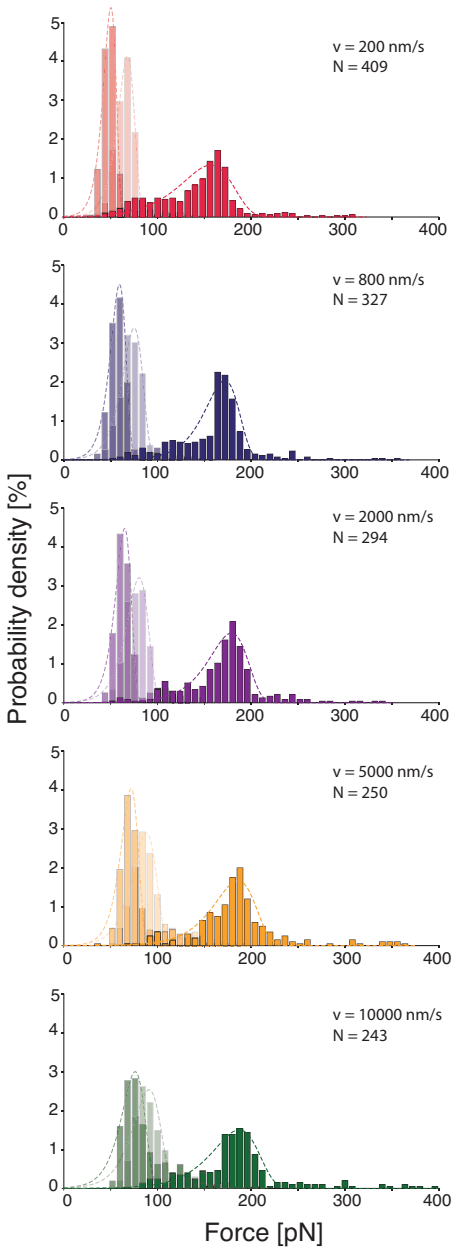


Fig B. Force histograms for a 15 h measurement. Unfolding and unbinding forces are plotted in the same manner as for Figure 6. For this experiment, the spring constant of the cantilever was 53 pN/nm. For this measurement, a second peak at lower forces is visible for the unbinding forces. From the course of the measurement (Fig A), it is obvious that the amount of low unbinding forces is insignificant in the first 2.5 h of the measurement. Therefore, the second peak cannot be caused by different binding states of biotin and mSA. The absence of a second binding state is further substantiated by the fact that for the lower unbinding forces, the unfolding forces of ddFLN4 are not shifted towards lower forces. As suggested by Schoeler et al. [1], such a bias occurs if there is an overlap of the probability distributions corresponding to unfolding and unbinding. Since they mostly occur for the slow retraction velocities, i.e. for long surface contact, we attribute these low unbinding forces to unspecific sticking of the cantilever to the surface resulting in ddFLN4 like force-extension patterns. Specific interaction at high forces was yet still detectable after 15 h of continuous measurement at room temperature. The time scale for the undisturbed interaction, i.e. without the additional low unbinding forces, is still sufficient for all immobilization and labeling applications of mSA envisioned in the main text.

S3 Appendix.

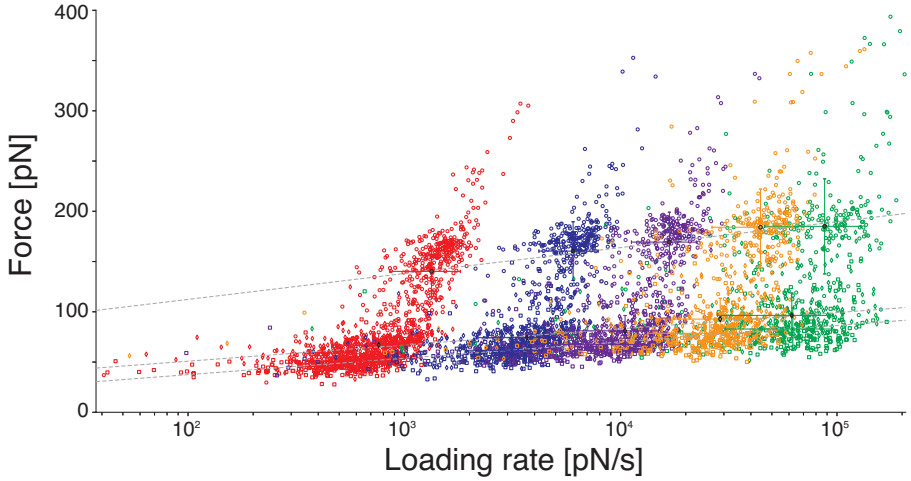


Fig C. Bell-Evans plot for a 15 h measurement. Data and color-coding are the same as in Fig B. Unfolding forces of ddFLN4 are plotted with open squares and diamonds, unbinding forces for biotin:mSA with open circles. Dashed lines are linear fits to the centers of gravity (shown as filled circles and diamonds) of the distributions of forces and loading rates, respectively. Colored crosses indicate the corresponding standard deviations. We find $\Delta x_0 = (0.59 \pm 0.06)$ nm and $k_{off,0} = 1 \times 10^{-2}$ s^{-1} for the first unfolding step of ddFLN4, $\Delta x_0 = (0.58 \pm 0.04)$ nm and $k_{off,0} = 7 \times 10^{-2}$ s^{-1} for the second unfolding step of ddFLN4, and $\Delta x_0 = (0.37 \pm 0.03)$ nm and $k_{off,0} = 4 \times 10^{-4}$ s^{-1} for the rupture of the biotin:mSA-complex.

References

1. Schoeler C, Verdorfer T, Gaub HE, Nash MA. Biasing effects of receptor-ligand complexes on protein-unfolding statistics. Phys Rev E. 2016;94(4-1):042412. doi: 10.1103/PhysRevE.94.042412. PubMed PMID: 27841541.

S4 Appendix.

Estimating the contour lengths of PEG and ddFLN4

In our experiments we use polyethylene glycol with a molecular weight of 5,000 g/mol. The molar mass of PEG is given by $(18.02 + 44.05 \times n)$ g/mol, where n is the number of subunits. For PEG5000, the number of subunits is $n = 113$. The net length of a segment is reported to be in the range of 0.278 nm to 0.358 nm depending on the orientation of the bonds [1]. We thus estimate the contour length of a PEG5000 polymer to be in the range of 31 nm to 40 nm. In this estimation, *N*-Hydroxysuccinimide and maleimide are not considered.

Our ddFLN4 consists of 101 amino acids. Assuming a length of 0.36 nm per amino acid, the contour length of the pure ddFLN4 reads 36 nm. We are neither taking into account additional length caused by linkers nor are we correcting for the end-to-end-distance of the folded ddFLN4, when considering the contour length increment upon unfolding.

References

1. 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(1):6.

S5 Appendix.

Formulas

Bell-Evans distribution

$$p(F) = \frac{k_{off,0}}{\dot{F}} \cdot e^{\frac{F\Delta x_0}{k_B T}} \cdot e^{\left(\frac{k_{off,0} k_B T}{\dot{F}\ \Delta x_0}\left(1 - e^{\left(\frac{F\Delta x_0}{k_B T}\right)}\right)\right)}$$

Loading-rate ($\dot{F}$) dependence of unbinding or unfolding force (F^*)

$$F^*\left(\dot{F}\right) = \frac{k_B T}{\Delta x_0} \log\left(\frac{\dot{F}}{k_{off,0}} \frac{\Delta x_0}{k_B T}\right)$$

Worm-like chain model

$$F(x) = \frac{kT}{4p}\left(\left(1 - \frac{x}{l}\right)^{-2} - 1 + 4\frac{x}{l}\right)$$

Transformation into contour length space

$$L(F,x) = Re\left(\frac{xkT}{6Fp}\left(3 + 4\frac{Fp}{kT} + \frac{4\left(\frac{Fp}{kT}\right)^2 - 3\frac{F}{kT} + 9}{f(kT,F,p)} + f(kT,F,p)\right)\right)$$

$$f(kT,F,p) = \left(27 - \frac{27Fp}{2kT} + \left(6\frac{Fp}{kT}\right)^2 - \left(2\frac{Fp}{kT}\right)^3 + \frac{3}{2}\sqrt{-3\left(\frac{Fp}{kT}\right)^2\left(\left(4\frac{Fp}{kT} - 3\right)^3 - 108\right)}\right)^{\frac{1}{3}}$$

S6 Appendix.

Fitted Bell-Evans distributions shown in Fig 6

First ddFLN4 unfolding peak					
v **[nm/s]**	**200**	**800**	**2,000**	**5,000**	**10,000**
$\dot{F}$ **[pN/s]**	768	3,519	10,080	29,010	66,710
F **[pN]**	66	75	83	89	90
Δx_0 **[nm]**	0.56	0.47	0.37	0.42	0.42
$k_{off,0}$ **[s^{-1}]**	7×10^{-3}	7×10^{-2}	6×10^{-1}	4×10^{-1}	7×10^{-1}
Second ddFLN4 unfolding peak					
v **[nm/s]**	**200**	**800**	**2,000**	**5,000**	**10,000**
$\dot{F}$ **[pN/s]**	701	3,609	9,841	29,820	76,030
F **[pN]**	59	70	78	87	94
Δx_0 **[nm]**	0.47	0.50	0.41	0.42	0.34
$k_{off,0}$ **[s^{-1}]**	1×10^{-1}	9×10^{-2}	4×10^{-1}	5×10^{-1}	2
Biotin:mSA unbinding peak					
v **[nm/s]**	**200**	**800**	**2,000**	**5,000**	**10,000**
$\dot{F}$ **[pN/s]**	1,736	7,469	20,680	52,390	111,900
F **[pN]**	201	212	217	222	230
Δx_0 **[nm]**	0.33	0.35	0.37	0.28	0.22
$k_{off,0}$ **[s^{-1}]**	2×10^{-5}	9×10^{-6}	8×10^{-6}	1×10^{-3}	2×10^{-1}

Table A. Fitted Bell-Evans distributions shown in Fig 6. To the histograms shown in Fig 6, Bell-Evans distributions were fitted. Mean loading rate used for the fit, most probable rupture force determined from the fit, and fitting parameters (distance to transition state and zero-force off-rate) are listed for the five retraction velocities and the different force peaks.

S7 Appendix.

Sequences of protein constructs

Functional core SA subunit with an N-terminal His-tag (green) and a unique cysteine (cyan):

MGSSHHHHHHMCGSEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRY
VLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQW
LLTSGTTEANAWKSTLVGHDTFTKVKPSAAS

Functional core SA subunit:

MEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDG
SGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKS
TLVGHDTFTKVKPSAAS

Non-functional core SA subunit with three mutations (red; N23A, S27D, S45A):

MEAGITGTWYAQLGDTFIVTAGADGALTGTYEAAVGNAESRYVLTGRYDSAPATDG
SGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKS
TLVGHDTFTKVKPSAAS

YbbR-tagged (magenta) ddFLN4 construct with N-terminal His-tag (green) and C-terminal cysteine (cyan). A cysteine that could potentially be accessible for binding to maleimide was mutated to serine (red; C18S):

MDSLEFIASKLAHHHHHHGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRT
DGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPK
TVTVKPAPGSC

YbbR-tagged (yellow) superfolder GFP construct with N-terminal His-tag (green) and cysteine (cyan) for tethering. A cysteine that could potentially be accessible for binding to maleimide was mutated to serine (red; C48S):

MGSSHHHHHHLEVLFQGPGHMCGSGSMSKGEELFTGVVPILVELDGDVNGHKFSVR
GEGEGDATIGKLTLKFISTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSA
MPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYN
FNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQTVLSKDPNEKRDHMVLHEYVNAAGITHGMDELYKSGSGSASDSLEFIASKLA

S8 Appendix.

Measuring with mSA immobilized on the cantilever

To test the stability of mSA as an anchor for SMFS, we also performed measurements in the opposite configuration, i.e. attaching mSA to the cantilever and biotinylated proteins to the surface (Figures S5 and S6). In this configuration, refolding of the control domain is unnecessary, because for every force-distance curve a new calibration domain is available on the surface. We used biotinylated GFP, whose unfolding pattern is well characterized [1], as calibration domain.

For these measurements, the distribution of rupture forces is much broader and slightly shifted to lower forces compared to the measurements with mSA on the surface. As we find the same effect, when immobilizing biotinylated ddFLN4 on the surface, we suspect shift and broadening of the distributions to be caused by slow degradation of the mSA molecules on the cantilever. This could imply that in this specific pulling geometry unbinding of biotin involves partial unfolding of the functional mSA subunit. To probe this hypothesis, steered molecular dynamics simulations could be helpful, but this is beyond the scope of this study.

S8 Appendix.

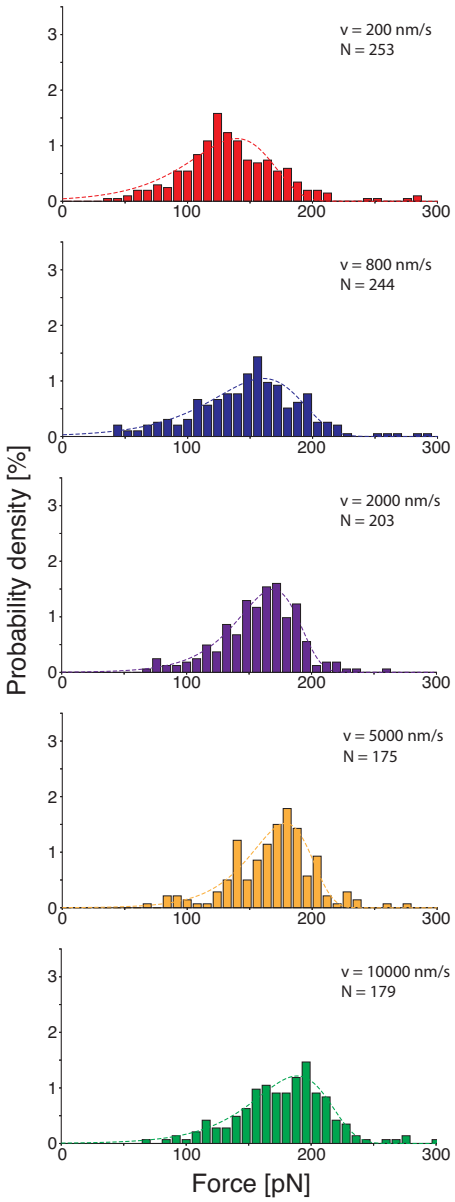


Fig A. Force Histograms, when measuring with mSA immobilized on the cantilever. For this measurement, mSA was immobilized on the cantilever and biotinylated GFP was attached to the surface. The spring constant of the cantilever was $k = 69.8$ pN/nm. The dashed lines show independent fits of Bell-Evans distributions to the force histograms.

S8 Appendix.

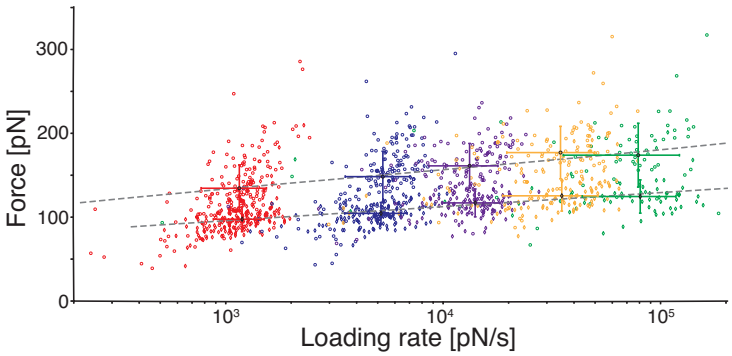


Fig B. Dynamic Force Spectrum for unfolding of GFP and unbinding of biotin:mSA. Color-coding is the same as in Fig A. Unfolding forces of GFP are plotted with open diamonds, rupture forces of the complex with open circles. Dashed lines are linear fits to the centers of gravity (shown as filled circles and diamonds) of the distributions of forces and loading rates, respectively. Colored crosses indicate the corresponding standard deviations. We find $\Delta x_0 = (0.56 \pm 0.08)$ nm and $k_{off,0} = 2 \times 10^{-4}\ s^{-1}$ for the unfolding of GFP and $\Delta x_0 = (0.39 \pm 0.05)$ nm and $k_{off,0} = 3 \times 10^{-4}\ s^{-1}$ for the rupture of the biotin:mSA-complex.

References

1. Dietz H, Rief M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. Proc Natl Acad Sci U S A. 2004;101(46):16192-7. doi: 10.1073/pnas.0404549101. PubMed PMID: 15531635; PubMed Central PMCID: PMCPMC528946.

3.2 Publication P2: Direction Matters: Monovalent Streptavidin/Biotin Complex under Load

Publication P2 describes how the force required to unbind biotin from mSA depends on the force loading geometry. mSA is tethered to a sample surface either by the N-terminus or by the C-terminus of the functional subunit using a unique cysteine introduced at the corresponding position in the amino acid sequence. The unbinding of biotin from these variants of mSA is examined by AFM-based SMFS. Complementary steered molecular dynamics (SMD) simulations contribute to understand resulting differences in unbinding forces on the molecular level.

To allow for direct comparison, both versions of mSA are immobilized on the same surface. On the cantilever tip, the introduction of a second receptor-ligand system, namely the Fgβ-SdrG-system, enables stable long-term measurements with large statistics. Hereby, SdrG is covalently attached to the cantilever tip, while its counterpart Fgβ is fused to biotinylated ddFLN4 domains, which are added to the measurement buffer. Fgβ unbinds from SdrG at much higher forces but has a much lower binding affinity than the SA/biotin interaction. This ensures continuous exchange of the biotinylated ddFLN4 molecules on the cantilever tip. Thereby, it prevents permanent clogging of the tip by stray SA molecules that come off the surface. To reliably exclude influences of this change in the experimental setup (compared with Publication P1), control experiments with direct covalent attachment of biotinylated ddFLN4 to the cantilever tip are performed.

In both cases, biotin unbinds from mSA tethered by the N-terminus of the functional subunit at forces of about 200 pN, while the bond between C-terminally tethered mSA and biotin withstands forces of about 400 pN (at a loading-rate of about 10 nN/s). Binding enthalpy and binding stoichiometry of both mSA variants are checked by ITC. Here, in the absence of force, no difference in binding behavior is seen. Therefore, it is concluded that the force-loading geometry is the pivotal difference resulting in the distinct unbinding forces for N- and C-terminally tethered mSA.

To get insights into the underlying molecular mechanism, SMD simulations are performed using the same force loading geometry as in the experiment. For N-terminally tethered mSA, an unbinding pathway that involves partial unfolding of the functional subunit is identified: β-strands β1 and β2 get torn apart upon N-terminal force application. This impairs the structural integrity of the binding pocket and results in lower unbinding forces for biotin compared to the unbinding of biotin from a C-terminally loaded mSA, for which no partial unfolding is observed.

While in the SMD simulations also high force unbinding events are observed for the N-terminal force-loading geometry, these are not detected in the experiment. On the other hand, lower unbinding forces for the C-terminal geometry are occasionally observed in the experiment, while in the SMD simulations, the structural integrity of the binding pocket is never disrupted for C-terminal force loading. This difference between SMD simulations and experiments is attributed to the fact that the two techniques examine different timescales. Due to restrictions on computational resources, the SMD simulations have to be performed at higher loading rates than the experiments. It is therefore possible that the two unbinding pathways are differently pronounced for experiment and simulation: In the experiment, the structural integrity of the N-terminal β-strands is always disrupted, resulting in low

unbinding forces. In the SMD simulation, this is not always the case. For the C-terminal force loading, the structural integrity of the binding pocket is always preserved in the simulation, while for the experiment this might not have always be the case. Yet, both AFM-based SMFS data and SMD simulations agree that force-loading direction matters for the mSA/biotin.

Direction Matters: Monovalent Streptavidin/Biotin Complex under Load

by

Steffen M. Sedlak*, Leonard C. Schendel*, Marcelo C. R. Melo, Diana A. Pippig, Zaida Luthey-Schulten, Hermann E. Gaub, and Rafael C. Bernardi

published in

Nano Letters 2019, 19, 3415-3421, doi:10.1021/acs.nanolett.8b04045

*These authors contributed equally to this work. SMS: Protein preparation, ITC, AFM-based SMFS (single receptor-ligand system), data analysis, writing of manuscript. LCS: AFM-based SMFS (double receptor-ligand system), data analysis, writing of manuscript

NANO LETTERS
Cite This: Nano Lett. 2019, 19, 3415−3421
Letter
pubs.acs.org/NanoLett

Direction Matters: Monovalent Streptavidin/Biotin Complex under Load

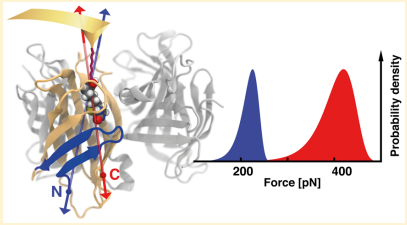
Steffen M. Sedlak,[⊥,†] Leonard C. Schendel,[⊥,†] Marcelo C. R. Melo,[‡,§] Diana A. Pippig,[†] Zaida Luthey-Schulten,[‡,§,∥] Hermann E. Gaub,[†] and Rafael C. Bernardi*[,‡]

[†]Lehrstuhl für Angewandte Physik and Center for NanoScience (CeNS), Ludwig-Maximilians-Universität München, Amalienstrasse 54, 80799 Munich, Germany

[‡]Beckman Institute for Advanced Science and Technology, [§]Center for Biophysics and Quantitative Biology, and [∥]Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Novel site-specific attachment strategies combined with improvements of computational resources enable new insights into the mechanics of the monovalent biotin/streptavidin complex under load and forced us to rethink the diversity of rupture forces reported in the literature. We discovered that the mechanical stability of this complex depends strongly on the geometry in which force is applied. By atomic force microscopy-based single molecule force spectroscopy we found unbinding of biotin to occur beyond 400 pN at force loading rates of 10 nN/s when monovalent streptavidin was tethered at its C-terminus. This value is about twice as high than that for N-terminal attachment. Steered molecular dynamics simulations provided a detailed picture of the mechanics of the unbinding process in the corresponding force loading geometries. Using machine learning techniques, we connected findings from hundreds of simulations to the experimental results, identifying different force propagation pathways. Interestingly, we observed that depending on force loading geometry, partial unfolding of N-terminal region of monovalent streptavidin occurs before biotin is released from the binding pocket.



KEYWORDS: *Streptavidin/biotin, single-molecule force spectroscopy, atomic force microscopy, molecular dynamics, machine learning*

The interaction of the small molecule biotin with the protein streptavidin (SA) is widely used for noncovalent, yet stable bonding in nanotechnology, biotechnology, and medicine.[1] The robustness of SA and the SA/biotin complex over a wide range of conditions, the comparatively easy fusion of biotin to nucleic acids, proteins, dyes, other macromolecules or nanoparticles, and the extraordinarily high affinity of the interaction make the complex a superior choice for immobilization, labeling, or detection of molecules.[2−5]

Recombinant core streptavidin monomers consist of the residues 13−159 of wild-type streptavidin and form a stable tetramer (Figure 1A). Every streptavidin subunit consists of a β-barrel in which a biotin molecule can be bound. The β-barrel is built up from eight antiparallel β-strands. The four β-strands located toward the N-terminus are considerably shorter (5−7 amino acids) than the four β-strands situated toward the C-terminus (10−13 amino acids). The four long β-strands and the residues in between mainly mediate the interaction with the other subunits. The short α-helix between seventh and eighth β-strand exhibits a tryptophan residue (TRP120) that reaches into a neighboring subunit and stabilizes this neighboring biotin binding pocket.[6−8]

The binding of biotin induces a conformational change in the molecule: The flexible loop between third and fourth β-strand (L3/4; residues 45−52) closes like a lid over the binding pocket.[9] Crystal structures of open and closed conformation have been solved.[10] Loop closure is vital for the tight binding of biotin. By mutating three residues (N23A, S27D, S45A) that are important for a stable closed loop conformation (cf. Supporting Information), Howarth et al. engineered a SA subunit with negligible affinity toward biotin (Figure S3).[11] Interestingly, all mutated residues are located between the L3/4-loop and the N-terminus.

Combining three nonfunctional subunits with one functional subunit, defined monovalent streptavidin (mSA) enabling a 1:1 binding stoichiometry can be created. Recently, the crystal structure of mSA was solved (Figure S4).[12] Crystallographic data suggest that in the nonfunctional subunit, the L3/4-loop is fixed in an open state−similar to the open state of wild-type apo-SA.

Over the last decades, scientists put a lot of effort in investigating the mechanical properties of this outstanding, noncovalent interaction. It was the first receptor ligand system where binding forces between individual molecules were measured by

Received: October 9, 2018
Revised: October 18, 2018
Published: October 22, 2018

DOI: 10.1021/acs.nanolett.8b04045
Nano Lett. 2019, 19, 3415−3421

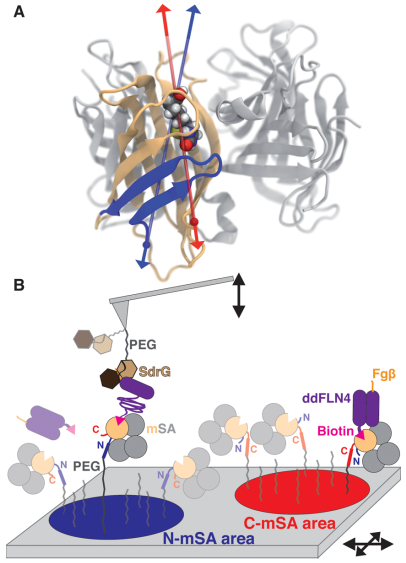


Figure 1. (A) Crystal structure of monovalent streptavidin (PDB 5TO2,[12] biotin from PDB 1MK5).[43] Biotin is bound in the functional subunit (light orange). The other subunits (gray) are genetically engineered to not bind biotin. Blue and red balls mark, respectively, the N- and C-terminus where mSA is tethered. Blue and red lines indicate the force loading directions. N-terminal region β-strands are highlighted in blue. (B) Experimental setup for AFM-based SMFS. At different surface areas, N-mSA and C-mSA are immobilized using PEG-spacers. Biotinylated (magenta) ddFLN4 (purple) is added to the solution and binds to the functional subunit of mSA (light orange ball). When the cantilever tip, functionalized with SdrG (brown hexagons), is approached to the surface, the Fgβ-peptide (orange) fused to ddFLN4 can bind to SdrG. Retracting the cantilever tip from the surface, ddFLN4 unfolds before biotin unbinds from mSA. Details of attachment chemistry and measurement process are provided in Figures S1 and S2.

atomic force microscopy (AFM)-based single molecule force spectroscopy (SMFS):[13,14] Unspecifically adsorbed biotinylated bovine serum albumin was immobilized on both cantilever and sample surface, while streptavidin was added to the buffer solution.[15] In subsequent studies, the experimental setup was improved using, for example, covalent attachment of biotin, polyethylene glycol linkers,[16] or other attachment strategies.[17−22] Later, covalent attachment of both biotin and streptavidin to cantilever or sample surface was accomplished.[23,24] Nowadays, the streptavidin/biotin system serves as a standard molecular anchoring system in AFM-based SMFS,[25,26] but also in optical tweezers,[27] magnetic tweezers,[28] and acoustic force spectroscopy experiments.[29]

Avidin/biotin and SA/biotin complexes were also fundamental in the initial development of steered molecular dynamics (SMD) simulations with both complexes among the first ones investigated by this technique.[30,31] Even before the advent of SMD, theoretical models have been put forward to describe the underlying molecular mechanism of the system.[32−35] Molecular dynamics (MD) simulations provided insights into different aspects of the interaction.[9,36,37] However, to investigate SA/biotin mechanics, the center of mass of the SA molecule has been kept fixed in previous SMD studies, which is different from the experimental force loading geometry.[30,38] In the literature, a large number of experimental and theoretical results, including supposedly contradictory studies, can be found.[39,40] On a molecular scale, a complete understanding of how biotin unbinds from the SA binding pocket under force is to date still missing.

For this study, we produced two different variants of mSA adding a unique cysteine either on the N-terminus (N-mSA) or the C-terminus (C-mSA) of the functional subunit (Figure 1A). The cysteine is utilized for site-specific covalent tethering. Additionally, the functional subunit was equipped with a polyhistidine tag used for purification. To ensure that the modifications do not affect the binding of biotin, we performed isothermal titration calorimetry (Figure S6).

In our experiments, the two different mSA variants were immobilized a few millimeters apart from each other on a glass slide by site-specific thiol maleimide coupling to polyethylene glycol (PEG) spacers (Figure 1B). The covalent immobilization of different proteins on the same surface is advantageous, because all are probed with the same cantilever tip. This allows for direct comparison of relative forces, thus avoiding issues of cantilever calibration or measurement conditions.[41,42]

We used the fourth filamin domain of *Dictyostelium discoideum* (ddFLN4) as fingerprint domain to identify single-molecule interactions, because it unfolds at forces lower than biotin unbinding from mSA.[44−46] We performed measurements with biotinylated ddFLN4 directly covalently attached to the cantilever tip (Figures S11−13). However, the high affinity of the mSA/biotin interaction causes a rapid loss of interaction as the cantilever tip gets clogged by mSA that was nonspecifically adsorbed to the surface.

To prevent cantilever clogging and to obtain better statistics, we introduced a second receptor−ligand pair (Figure 1B). While the surface was functionalized with mSA, the cantilever was functionalized with the adhesin SD-repeat protein G (SdrG) from *Staphylococcus epidermidis*.[47,48] After about a thousand approach−retraction cycles, biotinylated ddFLN4, to which short peptide from human fibrinogen β (Fgβ) had been genetically fused, was added to the measurement buffer. These molecules bound to the mSA on the surface via the biotin. The SdrG domain on the cantilever tip could pick up the Fgβ-peptide. Because the SdrG/Fgβ interaction can withstand a nearly 10-fold higher force than the mSA/biotin interaction,[48] we only measure the unbinding of biotin from mSA without bias from the SdrG/Fgβ interaction. On the other hand, the lower affinity of the SdrG/Fgβ interaction allows for a continuous exchange of the complexes at the tip and by means of this prevents permanent clogging of the cantilever tip. Even after 75 000 approach−retraction cycles, we still observed specific interactions between proteins immobilized on tip and surface (Figure 2).

The characteristic two-step unfolding pattern of ddFLN4 is used to identify single-molecule interactions, that is, a single biotin molecule binding to a single mSA molecule. In Figure 3A, two exemplary force−extension traces for single-molecule interaction on the area where N-mSA or C-mSA were immobilized are depicted (cantilever retraction velocity: 1,600 nm/s). Although the ddFLN4 unfolding is observed at the same force (Figures S7, S8), the final force peaks reach different values. These last peaks are attributed to the unbinding of biotin from mSA. Selecting all force curves that clearly show single-molecule interaction, we plotted mSA/biotin unbinding force histograms

DOI: 10.1021/acs.nanolett.8b04045
Nano Lett. 2019, 19, 3415−3421

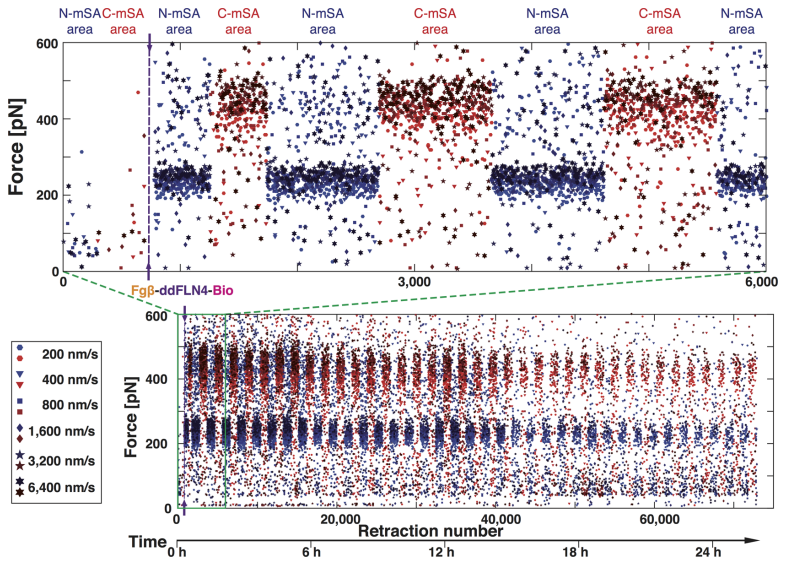


Figure 2. Course of a measurement. The final unbinding forces for all retractions of the cantilever tip from the surface are shown. Interactions on the surface area with the C-mSA are plotted in red colors; interactions in the N-mSA area are shown in blue colors. The darker the color is, the higher the cantilever retraction velocity is. The beginning of the measurement is shown on top. The Fgβ−ddFLN4−Biotin construct was added after 960 approaches, indicated by the purple dashed line and arrows. At the beginning of the measurement, high unbinding forces for N-mSA are also observed which are attributed to multiple interactions.

for both attachment geometries and all six retraction velocities (Figure 3B). We used Bell-Evans theory to fit the peaks of the distributions (Tables S1, S2).[49,50] While biotin unbinds from N-terminally tethered mSA at forces of about 200 pN, its binding to C-terminally tethered mSA is mechanically more stable and withstands forces of more than 400 pN. Fitting the dynamic force spectrum (Figure 3C), we could draw conclusions about coarse features of the binding energy landscape: by a factor of 2, the potential well is narrower for C-mSA compared to N-mSA.

To reveal the underlying molecular mechanism of the mSA/biotin interaction, ensuring statistical reliability, we performed 150 SMD production runs, which combined account for 19 μs. Simulations were performed using QwikMD[51] and GPU-accelerated NAMD.[52,53] In previous SMD studies, usually the center of mass of the SA molecules was kept at a fixed position, which does not resemble the experimental conditions. In our SMD simulations, we hold mSA either by the C-terminus or the N-terminus of the functional subunit and pulled biotin out of the binding pocket (for details on the preparation of the system, cf. Supporting Information), which is in agreement with the experimentally applied force loading geometry (Figure 1A). While for C-mSA, a unimodal force distribution was observed (Figure 4A), N-mSA showed a bimodal behavior (Figure 4B,C). For 9 out of 25 SMD replicas performed at 5000 μm/s pulling speed, the structural integrity of the N-terminal β-sheet was destroyed, before biotin left the binding pocket (Figure 4E). This structural rearrangement weakens the stability of the N-terminal β-sheet structure and thus results in lower final unbinding forces, blurring the boundaries between unbinding and unfolding. In one case, due to an extended simulation time we even observed how streptavidin regains its native fold when the force drops after biotin has left the pocket. The number of H-bonds between the first and the second β-strand provides a measure for the structural integrity (Figure 4G). If the N-terminal β-sheet structure stays intact, the number of H-bonds stays constant over time and high unbinding forces can be reached. The small unfolding observed in the simulations is beyond the resolution of our experimental setup. As the force loading rate dependence of an unfolding or unbinding event can be completely different than the one of a direct unbinding event, the simulations can be favoring the latter type of event while the experiments the former.

The simulations provide a detailed picture of the unbinding process, with atomic spatial resolution and femtosecond time resolution. Using correlation-based network analysis (Figures S14−16),[54] we analyzed the force propagation profiles, identifying which amino acids and domains of the molecules transmit force.[55] For C-mSA (Figure 5A−C), force either propagates through the long C-terminal β-strand, or through the N-terminal β-sheet structure, near the first hairpin between β-strands 1 and 2. These pathways indicate that mSA is structurally stable from both biotin sides when force is applied at the C-terminus, comparable with a claw. For N-mSA (Figure 5D−F), on the other hand, force is only rarely transmitted through the long C-terminal β-strands. Force propagates mainly through the shorter N-terminal β-strands. As the tension is high over the first and the second β-strand, high rupture forces can be reached if this region stays intact (Figure 5F). If the first two β-strands get torn apart (Figure 5E), the N-terminal structure loosens, mSA releases its grip on biotin, and biotin leaves the binding pocket.

DOI: 10.1021/acs.nanolett.8b04045
Nano Lett. 2019, 19, 3415−3421

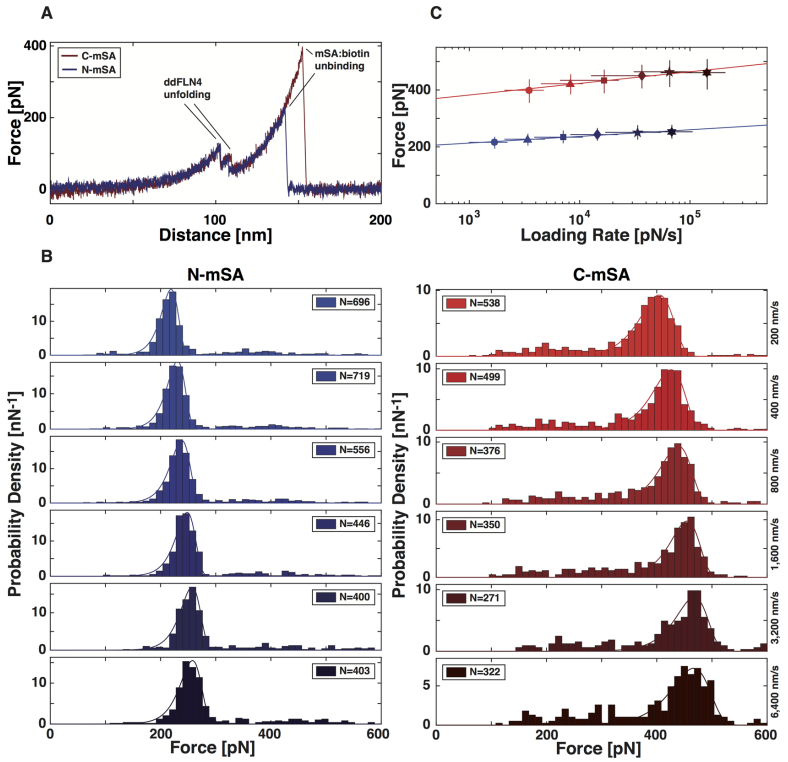


Figure 3. Analysis of force curves showing characteristic unfolding pattern. (A) Exemplary force extension traces measured at a retraction velocity of 1600 nm/s for C-mSA (red) and N-mSA (blue) displaying the characteristic two peak unfolding pattern of ddFLN4. Only traces showing this pattern are selected for further analysis. (B) Force histograms of mSA/biotin unbinding for six different retraction velocities. Peaks are fitted with Bell-Evans distributions (solid lines). (C) For all retraction velocities, the most probable unbinding force is plotted against the most probable loading rate and fitted according to Bell-Evans theory. From the fit, distance to transition state Δx_0 and zero-force off-rate $k_{off,0}$ are determined. N-mSA: $\Delta x_0 = 0.41$ nm, $k_{off,0} = 7.7 \times 10^{-8}$ s^{-1}; C-mSA: $\Delta x_0 = 0.23$ nm, $k_{off,0} = 2.5 \times 10^{-8}$ s^{-1}. Error bars are given by the full width at half-maximum of the peak of the corresponding distribution.

When there is no more tension on the mSA subunit, the native N-terminal structure is retrieved. The importance of the N-terminal structural integrity is in line with the fact that all mutations needed to generate the nonfunctional subunit, are at prominent positions within the N-terminal β-sheet structure (cf. Supporting Information).

SMD trajectories were also employed to investigate the contact between biotin and SA. Using PyContact,[56] we created a map of the interactions between ligand and receptor. Initially, the contact score was analyzed throughout the whole simulation time for each of the trajectories. To better understand the differences in an equilibrium versus a force-loaded regime, we compared the contact score over trajectory windows under no force load and under high-force load. The analysis was performed for all 50 slow pulling trajectories performed at 5000 μm/s pulling speed (25 for N-terminal pulling and 25 for C-terminal pulling). Additionally, the root-mean-square fluctuation (RMSF) was also analyzed in the same trajectory windows. Because of the large amount of data generated in such analysis, a "big-data" strategy of dimensionality reduction had to be adopted. The analysis was performed using python libraries through Jupyter Notebook.[57] Commonly known as machine learning techniques, our approach employed mutual information theory to identify the amino acid residues that were "force-active". These residues were coupled to changes in force and could indicate possible key points of force regulation. Indeed, most of these residues had been previously identified as key-players in the mechanism of SA/biotin interaction (Tables S3 and S4).

Combined, the analysis of the SMD trajectories indicate that the partial unfolding for N-terminal force loading is the cause of the lower forces seen for N-mSA compared with C-mSA in the experiments. The second N-terminal pulling unbinding pathway seen in the simulations is only rarely observed in the experiments, as indicated by the small number of high-force events in Figure 3B. On the one hand, this might be due to the much faster pulling speeds of the simulations. In the experiment, the force loading rates are at least four orders of magnitude

DOI: 10.1021/acs.nanolett.8b04045
Nano Lett. 2019, 19, 3415−3421

Nano Letters Letter

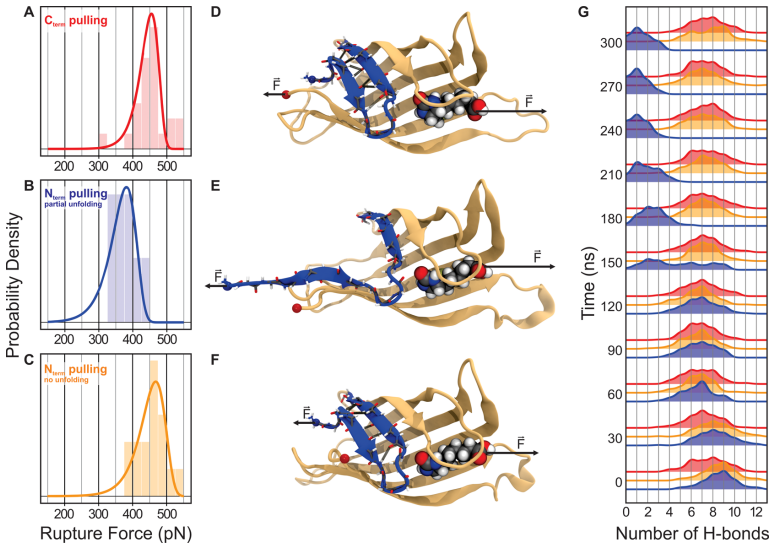


Figure 4. Results of SMD simulations. Unbinding force histogram for C-terminal attachment of mSA shows a unimodal distribution (red, A) ($N = 25$). For N-terminal attachment of mSA, two unbinding force peaks are observed: One at lower forces (blue, B) ($N = 9$) and one at higher forces (yellow, C) ($N = 16$). For C-terminal attachment of mSA, the structural integrity of the N-terminal β-sheet (marked in blue) is preserved (D). For N-terminal attachment of mSA, the structure of the N-terminal β-sheet can be destroyed before biotin unbinds from mSA, resulting in lower unbinding forces (E). If it stays intact, higher unbinding forces are reached (F). The number of hydrogen bonds between the first and the second N-terminal β-strand is a good measure to differentiate both cases (G). For the C-terminal attachment of mSA, it stays roughly constant over the timespan of the simulation (red). For N-terminal attachment, the contact is either broken completely (blue) or only slightly attenuated (yellow).

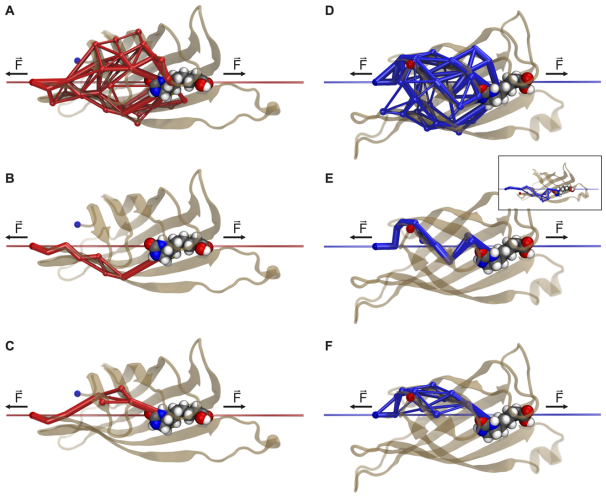


Figure 5. Force propagation pathways through the functional mSA subunit. (A) Overlay of the force propagation pathways for simulation replicas with C-terminal loading (Video S2) ($N = 25$). Force propagates through C-terminal β-sheets (B) or also through N-terminal β-sheets (C). (D) Overlay of all force propagation pathways for all simulation replicas with N-terminal loading (Video S3) ($N = 25$). Force propagates through N-terminal β-sheets. If the structural integrity of the N-terminal β-sheets is destroyed, the unbinding forces are low (E). If the N-terminal structure stays intact, higher unbinding forces can be reached (F). The thickness of the pathway edges represents the probability of force propagating through the particular edge. The probability was normalized for each simulation, leading to the same maximum thickness (maximum information pathway) for each simulation replica.

DOI: 10.1021/acs.nanolett.8b04045
Nano Lett. 2019, 19, 3415−3421

lower. The N-terminal β-sheet structure is held under tension for a much longer time, such that the unzipping of the first from the second β-strand is more likely. On the other hand, the molecular linker of the biotin to Coenzyme A (for details of the biotinylation, cf. Supporting Information) is not considered in the simulations (Figure S5) because there is no crystal structure for the linker and in addition missing force field parameters could introduce a source of imprecision. In previous combined AFM SMD studies, it was shown that only a complete simulation of all molecular linkers in proximity of the protein of interest provided an excellent agreement between experimental and simulated forces.[48] It is yet reasonable to assume that the additional interaction of the linker between biotin and Coenzyme A with mSA increases the final unbinding forces of biotin from mSA. Such interaction would favor the N-mSA unzipping/low force unbinding pathway over the high force unbinding pathway even more, also explaining the different in force distribution between simulation and experiment.

In this study, experiments and simulations were used hand-in-hand, providing a detailed picture of the system mechanics with the atomistic detail of the simulation, substantiated by the large statistical content of experiments. The nearly twofold difference in unbinding forces that we report for biotin in the two well-defined N- and C-terminal tethering geometries of mSA is nicely matched by the 2-fold reduction of the binding potential width as revealed by the Bell-Evans analysis of the rate dependence of the unbinding forces. Because we measured by ITC the same binding energy for the mSA/biotin complexes in both tethering geometries, we can conclude that our force histograms represent largely homogeneous ensembles of unbinding modes. The analysis of these modes by steered SMD revealed that in the case of the C-terminally tethered mSA the forced separation of biotin can be described best by a rupture process, leaving the molecular structure of the mSA binding pocket largely intact. The N-terminally tethered mSA, however, shows in a significant number of traces a marked structural change, a local unfolding of the binding pocket. We assume that the much slower time scale of the AFM-based SMFS favors the low force unfolding path. This partial unfolding results in a substantial widening of the potential energy landscape accompanied by a reduction of the unbinding force for N-mSA compared to C-mSA. In view on our results, it is worth noting that the widespread of SA/biotin unbinding forces reported in the literature[35,40] may have arisen from a multiplicity of force propagation geometries due to the nonspecific immobilization of the terameric streptavidins used in these investigations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.8b04045.

Materials and methods, results of isothermal titration calorimetry, results of SMFS with covalent attachment of ddFLN4, fit parameter of Bell-Evans fits, results of mutual information analysis, protein sequences (PDF)

Steered molecular dynamics simulations. Analyses of C- versus N-terminal pulling reveal that the instability of N-terminal region during the latter pulling is fundamental for the differences observed in force traces (AVI)

Exposé of all force propagation pathways for all simulation replicas with C-terminal loading (AVI)

Exposé of all force propagation pathways for all simulation replicas with N-terminal loading (AVI)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rcbernardi@ks.uiuc.edu.

ORCID

Steffen M. Sedlak: 0000-0002-7037-1793
Leonard C. Schendel: 0000-0002-1986-2693
Marcelo C. R. Melo: 0000-0001-6901-1646
Hermann E. Gaub: 0000-0002-4220-6088
Rafael C. Bernardi: 0000-0003-0758-2026

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Author Contributions

⊥S.M.S. and L.C.S. contributed equally.

Funding

The European Research Council (ERC advanced grant Cellufuel Grant 294438) and the Deutsche Forschungsgemeinschaft (DFG, Sonderforschungsbereich 1032, Teilprojekt A01) provided funding for this work. Furthermore, it was supported by the National Institutes of Health (NIH) Grant P41-GM104601, "Center for Macromolecular Modeling and Bioinformatics". R.C.B. is partially supported by National Science Foundation (NSF) Grant MCB-1616590, "Molecular Modeling of Bioenergetic Systems". Molecular dynamics simulations made use of GPU-accelerated nodes of Blue Waters supercomputer as part of the Petascale Computational Resource (PRAC) grant "The Computational Microscope", which is supported by the National Science Foundation (award number ACI-1713784). The state of Illinois and the National Science Foundation (awards OCI-0725070 and ACI-1238993) support Blue Waters sustained-petascale computing project.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank A. Kardinal and T. Nicolaus for laboratory support, L.F. Milles for providing the SdrG/Fgβ-system, and M. Scheurer for assistance with integrating PyContact and Jupyter. The authors acknowledge M.S. Bauer and C. Kluger for helpful discussions. S.M.S. thanks the Nanosystems Initiative Munich for support.

■ ABBREVIATIONS

SA, streptavidin; AFM, atomic force microscopy; SMFS, single-molecule force spectroscopy; SMD, steered molecular dynamics; mSA, monovalent streptavidin; N-mSA, N-terminally tethered monovalent streptavidin; C-mSA, C-terminally tethered monovalent streptavidin; ddFLN4, fourth filamin domain of *Dictyostelium discoideum*; SdrG, SD-repeat protein G from *Staphylococcus epidermidis*; Fgβ, a short peptide from human fibrinogen β

■ REFERENCES

(1) Dundas, C. M.; Demonte, D.; Park, S. *Appl. Microbiol. Biotechnol.* **2013**, *97* (21), 9343−53.
(2) Wilchek, M.; Bayer, E. A. *Anal. Biochem.* **1988**, *171* (1), 1−32.
(3) Bayer, E. A.; Ben-Hur, H.; Wilchek, M. *Methods Enzymol.* **1990**, *184*, 80−9.
(4) Green, N. M. *Methods Enzymol* **1990**, *184*, 51−67.

DOI: 10.1021/acs.nanolett.8b04045
Nano Lett. 2019, 19, 3415−3421

(5) Laitinen, O. H.; Nordlund, H. R.; Hytonen, V. P.; Kulomaa, M. S. *Trends Biotechnol.* **2007**, *25* (6), 269−77.
(6) Freitag, S.; Le Trong, I.; Chilkoti, A.; Klumb, L. A.; Stayton, P. S.; Stenkamp, R. E. *J. Mol. Biol.* **1998**, *279* (1), 211−21.
(7) Lim, K. H.; Huang, H.; Pralle, A.; Park, S. *Biotechnol. Bioeng.* **2013**, *110* (1), 57−67.
(8) Scholl, Z. N.; Yang, W.; Marszalek, P. E. *ACS Nano* **2015**, *9* (2), 1189−97.
(9) Bansal, N.; Zheng, Z.; Song, L. F.; Pei, J.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **2018**, *140* (16), 5434−5446.
(10) Freitag, S.; Le Trong, I.; Klumb, L.; Stayton, P. S.; Stenkamp, R. E. *Protein Sci.* **1997**, *6* (6), 1157−1166.
(11) Howarth, M.; Chinnapen, D. J.; Gerrow, K.; Dorrestein, P. C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y. *Nat. Methods* **2006**, *3* (4), 267−73.
(12) Zhang, M.; Biswas, S.; Deng, W.; Yu, H. *Sci. Rep.* **2016**, *6*, 35915.
(13) Florin, E. L.; Moy, V. T.; Gaub, H. E. *Science* **1994**, *264* (5157), 415−7.
(14) Lee, G.; Kidwell, D.; Colton, R. *Langmuir* **1994**, *10* (2), 354−357.
(15) Moy, V. T.; Florin, E. L.; Gaub, H. E. *Science* **1994**, *266* (5183), 257−9.
(16) Merkel, R.; Nassoy, P.; Leung, A.; Ritchie, K.; Evans, E. *Nature* **1999**, *397* (6714), 50−3.
(17) Wong, J.; Chilkoti, A.; Moy, V. T. *Biomol. Eng.* **1999**, *16* (1−4), 45−55.
(18) Stevens, M. M.; Allen, S.; Davies, M. C.; Roberts, C. J.; Schacht, E.; Tendler, S. J. B.; VanSteenkiste, S.; Williams, P. M. *Langmuir* **2002**, *18* (17), 6659−6665.
(19) Yuan, C.; Chen, A.; Kolb, P.; Moy, V. T. *Biochemistry* **2000**, *39* (33), 10219−23.
(20) Lo, Y.-S.; Zhu, Y.-J.; Beebe, T. P. *Langmuir* **2001**, *17* (12), 3741−3748.
(21) Rico, F.; Moy, V. T. *J. Mol. Recognit.* **2007**, *20* (6), 495−501.
(22) de Odrowaz Piramowicz, M.; Czuba, P.; Targosz, M.; Burda, K.; Szymonski, M. *Acta Biochim Pol* **2006**, *53* (1), 93−100.
(23) Taninaka, A.; Takeuchi, O.; Shigekawa, H. *Int. J. Mol. Sci.* **2010**, *11* (5), 2134−51.
(24) Sedlak, S. M.; Bauer, M. S.; Kluger, C.; Schendel, L. C.; Milles, L. F.; Pippig, D. A.; Gaub, H. E. *PLoS One* **2017**, *12* (12), e0188722.
(25) Ott, W.; Jobst, M. A.; Schoeler, C.; Gaub, H. E.; Nash, M. A. *J. Struct. Biol.* **2017**, *197* (1), 3−12.
(26) Walder, R.; LeBlanc, M. A.; Van Patten, W. J.; Edwards, D. T.; Greenberg, J. A.; Adhikari, A.; Okoniewski, S. R.; Sullan, R. M. A.; Rabuka, D.; Sousa, M. C.; Perkins, T. T. *J. Am. Chem. Soc.* **2017**, *139* (29), 9867−9875.
(27) Tych, K. M.; Jahn, M.; Gegenfurtner, F.; Hechtl, V. K.; Buchner, J.; Hugel, T.; Rief, M. *J. Phys. Chem. B* **2018**.
(28) Kriegel, F.; Ermann, N.; Forbes, R.; Dulin, D.; Dekker, N. H.; Lipfert, J. *Nucleic Acids Res.* **2017**, *45*, 5920.
(29) Sitters, G.; Kamsma, D.; Thalhammer, G.; Ritsch-Marte, M.; Peterman, E. J.; Wuite, G. J. *Nat. Methods* **2015**, *12* (1), 47−50.
(30) Grubmuller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271* (5251), 997−9.
(31) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72* (4), 1568−81.
(32) Chilkoti, A.; Stayton, P. S. *J. Am. Chem. Soc.* **1995**, *117* (43), 10622−10628.
(33) Chilkoti, A.; Boland, T.; Ratner, B. D.; Stayton, P. S. *Biophys. J.* **1995**, *69* (5), 2125−30.
(34) Pincet, F.; Husson, J. *Biophys. J.* **2005**, *89* (6), 4374−81.
(35) Guo, S.; Ray, C.; Kirkpatrick, A.; Lad, N.; Akhremitchev, B. B. *Biophys. J.* **2008**, *95* (8), 3964−3976.
(36) Song, J.; Li, Y.; Ji, C.; Zhang, J. Z. H. *Sci. Rep.* **2015**, *5*, 7906.
(37) Liu, F.; Zhang, J. Z. H.; Mei, Y. *Sci. Rep.* **2016**, *6*, 27190.
(38) Rico, F.; Russek, A.; Gonzalez, L.; Grubmuller, H.; Scheuring, S. **2018** arXiv preprint arXiv:1808.07122.
(39) Zlatanova, J.; Lindsay, S. M.; Leuba, S. H. *Prog. Biophys. Mol. Biol.* **2000**, *74* (1−2), 37−61.
(40) Teulon, J. M.; Delcuze, Y.; Odorico, M.; Chen, S. W.; Parot, P.; Pellequer, J. L. *J. Mol. Recognit.* **2011**, *24* (3), 490−502.
(41) Verdorfer, T.; Bernardi, R. C.; Meinhold, A.; Ott, W.; Luthey-Schulten, Z.; Nash, M. A.; Gaub, H. E. *J. Am. Chem. Soc.* **2017**, *139* (49), 17841−17852.
(42) Verdorfer, T.; Gaub, H. E. *Sci. Rep.* **2018**, *8* (1), 9634.
(43) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. *Protein Sci.* **2006**, *15* (3), 459−67.
(44) Schwaiger, I.; Kardinal, A.; Schleicher, M.; Noegel, A. A.; Rief, M. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 81−5.
(45) Milles, L. F.; Bayer, E. A.; Nash, M. A.; Gaub, H. E. *J. Phys. Chem. B* **2017**, *121* (15), 3620−3625.
(46) Bauer, M. S.; Milles, L. F.; Sedlak, S. M.; Gaub, H. E. **2018**, *bioRxiv*.
(47) Herman, P.; El-Kirat-Chatel, S.; Beaussart, A.; Geoghegan, J. A.; Foster, T. J.; Dufrene, Y. F. *Mol. Microbiol.* **2014**, *93* (2), 356−68.
(48) Milles, L. F.; Schulten, K.; Gaub, H. E.; Bernardi, R. C. *Science* **2018**, *359* (6383), 1527−1533.
(49) Bell, G. I. *Science* **1978**, *200* (4342), 618−27.
(50) Evans, E.; Ritchie, K. *Biophys. J.* **1997**, *72* (4), 1541−55.
(51) Ribeiro, J. V.; Bernardi, R. C.; Rudack, T.; Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Schulten, K. *Sci. Rep.* **2016**, *6*, 26536.
(52) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781−802.
(53) Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. *Nat. Methods* **2018**, *15* (5), 351−354.
(54) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6620−5.
(55) Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E. *Nano Lett.* **2015**, *15* (11), 7370−6.
(56) Scheurer, M.; Rodenkirch, P.; Siggel, M.; Bernardi, R. C.; Schulten, K.; Tajkhorshid, E.; Rudack, T. *Biophys. J.* **2018**, *114* (3), 577−583.
(57) Kluyver, T.; Benjamin, R.-K.; Fernando, P.; Brian, G.; Matthias, B.; Jonathan, F.; Kyle, K.; Jessica, H.; Jason, G.; Sylvain, C.; Paul, I.; Damián, A.; Safia, A.; Carol, W.; Jupyter Development, T. In *Jupyter Notebooks—a publishing format for reproducible computational workflows*; Positioning and Power in Academic Publishing: Players, Agents and Agendas, 20th International Conference on Electronic Publishing, Göttingen, Germany, 2016.

DOI: 10.1021/acs.nanolett.8b04045
Nano Lett. 2019, 19, 3415−3421

Supplementary Information

Direction Matters – Monovalent Streptavidin:Biotin Complex under Load

Steffen M. Sedlak,[‡,1] Leonard C. Schendel,[‡,1] Marcelo C. R. Melo,[2,3] Diana A. Pippig,[1] Zaida Luthey-Schulten,[2,3,4] Hermann E. Gaub,[1] and Rafael C. Bernardi[2]*

[1]Lehrstuhl für Angewandte Physik and Center for NanoScience (CeNS), Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 Munich, Germany

[2]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[3]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[4]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Table of contents

Extended table of contents

I. Materials and methods

I.1 Preparation of proteins

I.1.A Preparation of monovalent streptavidin

A detailed description of the expression, lysis and purification lab protocols is given by Sedlak *et al*.,[1] here we just provide a short overview. We used different streptavidin (SA) monomers encoded on different pET vectors: SA monomers with an N-terminal cysteine (Cys-SA), SA monomers with a C-terminal cysteine (SA-Cys), and SA monomers without cysteine but with three mutations in the binding pocket (N23A, S27D, S45A) preventing the binding of biotin (dSA). SA-Cys and Cys-SA further contained a polyhistidine tag. All monomers were expressed separately in *E. coli* BL21(DE3)-CodonPlus. SA plasmids were transferred to *E. coli* BL21(DE3)-CodonPlus cells (Agilent Technologies Inc., Santa Clara, USA) and expressed in SB medium. After cell lysis and purification of inclusion bodies, the monomers were denatured and mixed to obtain a 1:10 ratio of functional to non-functional subunits using dSA and either Cys-SA or SA-Cys. By slowly diluting the mixture into phosphate buffered saline (PBS; Sigma-Aldrich, Saint Louis, USA), the subunits refolded and tetrameric SA was formed. Monovalent streptavidin (mSA) with a unique, either N- or C-terminal cysteine (N-mSA or C-mSA) was purified by immobilized metal ion affinity chromatography making use of the polyhistidine tag on the functional subunit.

I.1.B Preparation of fingerprint domains

As a fingerprint domain for AFM-based force spectroscopy, the well-characterized fourth filamin domain of *Dictyostelium discoideum* (ddFLN4)[2] was used. Two different constructs were prepared: For the first, we added an N-terminal cysteine for site-specific immobilization and mutated the internal cysteine 18 to serine. For the second construct, we exchanged the N-terminal cysteine for an Fgβ-petide. For both, we cloned a ybbR-tag[3] to the C-terminus which we used to covalently attach a Coenzyme A modified biotin via a Sfp phoshopantetheinyl transferase-catalyzed reaction. The ddFLN4 construct was expressed in *E. coli* BL21(DE3)-CodonPlus and purified by immobilized metal ion affinity chromatography.

I.1.C Biotinylation of the ddFLN4-construct

45 μM ddFLN4-construct, 50 μM CoA-Biotin (New England Biolabs, Ipswich, Massachusetts, USA), and 5 μM Sfp Synthase dissolved in Sfp buffer (10 mM TRIS, 10 mM MgCl2, pH 7.5) were allowed to react for 1 h at 37°C. The reaction product was purified using Zeba Spin Desalting Columns (Thermo Fisher Scientific, Waltham, USA) with a molecular weight cut off of 7 kDa equilibrated with coupling buffer (50 mM NaCl, 50 mM NaHPO4, 10 mM EDTA, pH 7.2) according to the manufacturer's instruction.

I.2. Isothermal titration calorimetry

Isothermal titration calorimetry experiments were conducted with a Malvern Microcal ITC200. We diluted 8 mg of biotin (Sigma-Aldrich, St. Louis, USA) in 40 ml PBS to obtain an 818.6 μM stock solution. Using Zeba Spin Columns with a molecular weight cut off of 40 kDa according to the manufacturer's protocol, we performed a buffer exchange to have C-mSA or N-mSA, respectively, dissolved in the same PBS. We determined the protein concentration using the absorption at 280 nm and an absorption coefficient of 167,760 $M^{-1}cm^{-1}$. We diluted biotin with PBS to obtain tenfold molar excess, which resulted in a final molar ratio of about 2:1 (mSA:biotin) in the ITC measurement. For both mSA variants, we performed three independent measurements.

I.3. AFM-based SMFS measurements

I.3.A Surface functionalization

Bifunctional polyethylene glycol of 5,000 Da having an N-hydroxysuccinimide group at one end and a maleimide group on the other (NHS-PEG5000-MAL, Rapp Polymere, Tübingen, Germany) was dissolved in 50 mM HEPES at pH 7.5 and immediately used to incubate aminosilanized glass slides.[4] After one hour, the glass slides were thoroughly washed in ultrapure water.

5 μM N-mSA and 5 μM C-mSA were supplemented with 1 mM Bond-Breaker TCEP Solution (Thermo Fisher Scientific). After one hour, the mixture was purified using Zeba Spin Desalting Columns (Thermo Fisher Scientific, Waltham, USA) with a molecular weight cut off of 40 kDa equilibrated with coupling buffer (50 mM NaCl, 50 mM NaHPO4, 10 mM EDTA, pH 7.2) according to the manufacturer's instruction.
Using silicon masks, 10 μl droplets of the prepared C-mSA and N-mSA were placed on the surfaces. After a one-hour incubation, the surfaces were thoroughly washed using PBS, to rinse off unbound mSA.

I.3.B Cantilever functionalization

As for the surfaces, aminosilanized[4] BioLever mini (Olympus Corporation, Tokyo, Japan) were first incubated with heterobifunctional polyethylene glycol having a N-hydroxysuccinimide on the one end and a maleimide group on the other end and then washed in ultrapure water.
To couple the SD-repeat protein G from *staphylococcus epidermidis* (SdrG),[5] the cantilevers were incubated with 1 mM Coenzyme A diluted in coupling buffer for one hour. The sulfhydryl reacts with the maleimide to form a stable thioether bond. After washing in ultrapure water, the cantilevers were placed in Sfp buffer containing13 μM SdrG with a C-terminal ybbR-tag, 5 μM Sfp Synthase for at least one hour. The Sfp Synthase covalently coupled the ybbR-tag to the Coenzyme A. Finally, the levers were washed and stored in PBS.

To couple ddFLN4 to the cantilever, the Coenzyme A step was omitted and the unique cysteine of the biotinylated ddFLN4 constructs was used to specifically and covalently couple to the maleimide on the cantilever.

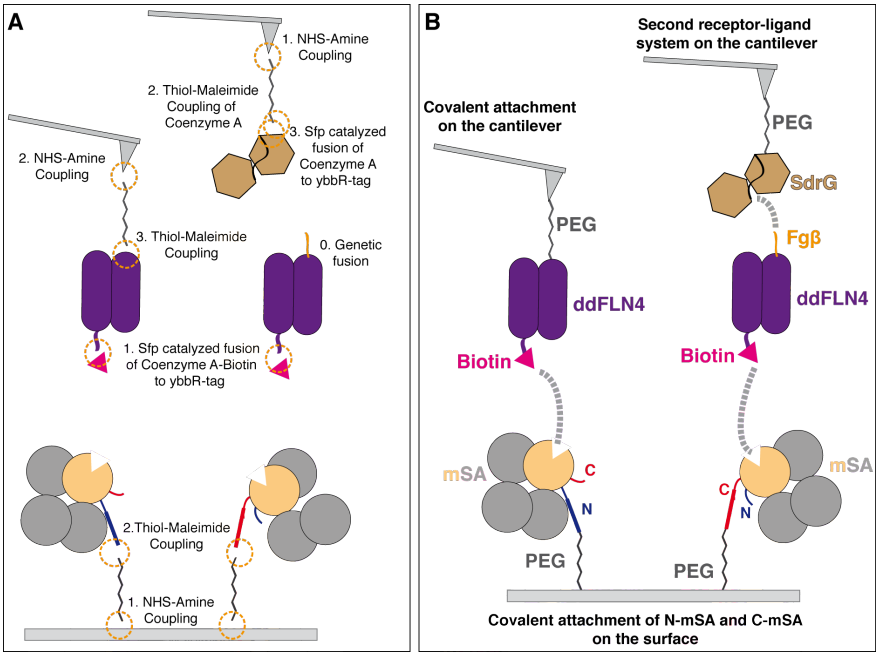


Figure S1. Surface and cantilever functionalization. (A) Aminosilanized cantilever tips and glass surfaces are functionalized separately by applying several subsequent chemical and enzyme mediated reaction steps. (B) The two attachment strategies: Covalent attachment of the biotinylated ddFLN4 to the cantilever tip or with the second receptor-ligand system on the cantilever, where biotinylated Fgβ-ddFLN4 constructs were added to the buffer solution. In both cases, N-mSA and C-mSA were covalently attached to the surface. The grey dashed lines indicate non-covalent receptor-ligand interactions.

I.3.C AFM-based force spectroscopy measurements

AFM-based force spectroscopy measurements were performed on a custom-built AFM[6] controlled with a self-written Igor Pro 6 software operating a commercial MFP3D controller. Both mSA constructs were immobilized on different spots on the same surface. The cantilever tip was either covalently functionalized with biotinylated fingerprint proteins or functionalized with the SdrG. In the latter case, biotinylated ddFLN4 constructs equipped with an N-terminal Fgβ-Peptide were added to the solution. After short contact with the surface, the cantilever was retracted with constant velocities and force-distance curves were recorded. After each retraction, the surface was moved 100 nm to provide an unused area to the cantilever for the next approach. After a few hundred approach-retraction-cycles, the surface was moved to expose the spot, where the other mSA construct had been immobilized, to the cantilever tip. S-Using the same cantilever for pulling mSA:biotin in two different but distinct geometries, facilitates direct comparison of both configurations.

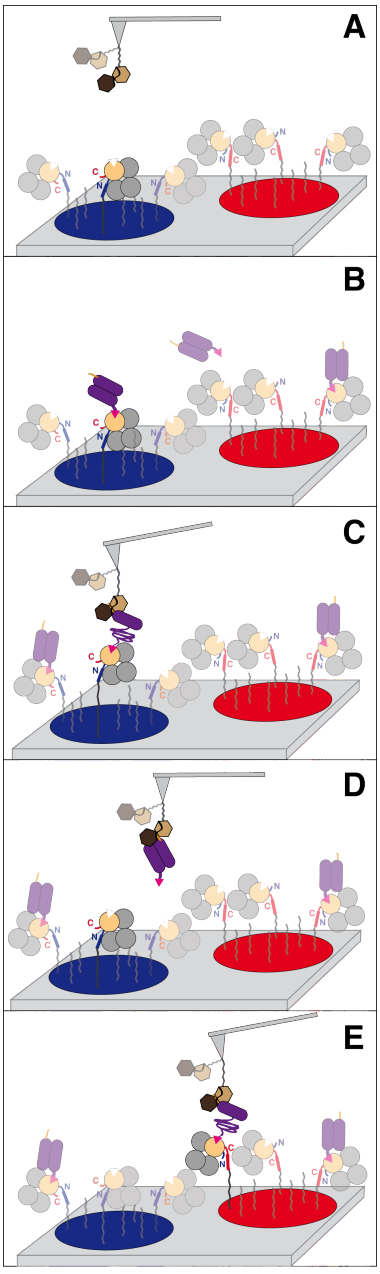


Figure S2. Detailed description of the AFM measurement process. (A) Different areas of the surface, a few millimeters apart from each other, are functionalized with N-mSA (blue are) and C-mSA (red area). The cantilever is functionalized with SdrG. To ensure that there is no specific interaction between SdrG and mSA, we perform a few hundred approach-retraction cycles. (B) A low concentration of Fgβ-ddFLN4-biotin construct is added to the solution and binds to some of the mSA molecules on the surface. (C) Approaching the cantilever tip, SdrG binds to an Fgβ-Peptide. Retracting the cantilever, ddFLN4 unfolds and finally biotin unbinds from mSA. (D) The ddFLN4 refolds and stays on the lever. Moving the surface, other mSA molecules are probed with the same fingerprint domain. At some point, the Fgβ unbinds from SdrG and another ddFLN4 is picked up from the surface. (E) Every few hundred approach-retraction cycles, the surface is moved so that the cantilever tip interacts with the other area on the surface.

I.4. Sequences of the protein constructs

Functional SA subunit with C-terminal cysteine (magenta) and polyhistidine tag (green):
M**EAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATD GSGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAW KSTLVGHDTFTKVKPSAAS**CLE**HHHHHH**

Functional SA subunit with N-terminal cysteine (magenta) and C-terminal polyhistidine tag (green):
MCGS**EAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPA TDGSGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANA WKSTLVGHDTFTKVKPSAAS**LE**HHHHHH**

Non-functional SA subunit with the three mutations N23A, S27D, S45A (red):
M**EAGITGTWYAQLGDTFIVTAGADGALTGTYEAAVGNAESRYVLTGRYDSAPATD GSGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAW KSTLVGHDTFTKVKPSAAS**

ddFLN4 construct with N-terminal ybbR-tag (blue) and polyhistidine tag (green) and C-terminal cysteine (magenta). The internal cysteine 18 is mutated to serine (red):
M**DSLEFIASKLA**HHHHHH**GSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHR TDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGF PKTVTVKPAP**GSC

ddFLN4 construct with N-terminal Fgβ-peptide (cyan), C-terminal polyhistidine tag (green) and ybbR-tag (blue):
MAT**NEEGFFSARGHRPLD**GSGSGSGSAGTGSG**ADPEKSYAEGPGLDGGESFQPSKFK IHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVIN LTLDGDNVNGFPKTVTVKPAP**SG**HHHHHH**GS**DSLEFIASKLA**LPETGG

The following subunits and mSA constructs made thereof were used to ensure that the position of the polyhistidine tag is not influencing the rupture forces measured with AFM-based SMFS.

Functional SA subunit with N-terminal cysteine (magenta) and polyhistidine tag (green):
MGSS**HHHHHHHH**M**C**GS**EAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESR YVLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQ WLLTSGTTEANAWKSTLVGHDTFTKVKPSAAS**

Functional SA subunit with C-terminal cysteine (magenta) and N-terminal polyhistidine tag (green):
MGSS**HHHHHHHH**MGS**EAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRY VLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQ WLLTSGTTEANAWKSTLVGHDTFTKVKPSAAS**C

I.5. In silico force spectroscopy

Employing advanced run options of QwikMD,[7] our *in silico* approach followed previously published protocols, which were recently published for filamins,[8] cellulosomes,[9] and adhesins.[5]

I.5.A Simulation setup

The structure of a monovalent *Streptomyces avidinii* streptavidin (mSA) had been solved by means of X-ray crystallography at 1.65 Å resolution and was available at the protein data bank (PDB: 5TO2).[10] As this structure does not contain the biotin bound to streptavidin, the structure of the tetravalent *S. avidinii* streptavidin bound to biotin (PDB: 1MK5), which was also solved by means of X-ray crystallography at 1.4 Å resolution,[11] was used to place the biotin on to its binding site at chain D of the mSA. The alignment and placing of the biotin into the monovalent structure was performed using VMD.[12] Employing advanced run options of QwikMD,[7] the structure was solvated and the net charge of the system was neutralized in a 0.15 mol/l sodium

chloride solution. In total, approximately 275,000 atoms were simulated in each simulation. The CHARMM36 force field,[13] along with the TIP3 water model[14] was used to describe all systems.

I.5.B Equilibrium molecular dynamics simulations

All MD simulations in the present study were performed employing the GPU-accelerated NAMD molecular dynamics package.[15] The simulations were performed assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for temperature and pressure coupling, the latter kept at 1 bar. A distance cut-off of 11.0 Å was applied to short-range non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method.[16] The equations of motion were integrated using the r-RESPA multiple time step scheme[17] to update the van der Waals interactions every step and electrostatic interactions every two steps. The time step of integration was chosen to be 2 fs for all simulations performed. Before the MD simulations, all systems were submitted to an energy minimization protocol for 5,000 steps. An MD simulation with position restraints in the protein backbone atoms and biotin non-hydrogen atoms was performed for 10 ns. To allow for a total relaxation of the system and to make sure biotin was stable in the streptavidin pocket, a 100 ns simulation in equilibrium, where no external forces were applied, was performed. The MD protocol served to pre-equilibrate the system before the steered molecular dynamics (SMD) simulations were performed.

I.5.C Steered molecular dynamics simulations

With structures properly equilibrated and checked, SMD simulations[18] were performed using a constant velocity stretching (SMD-CV protocol), employing three different pulling speeds: 5.0, 0.5 and 0.05 Å/ns. Simulations were performed restraining the position of the carbon C1 of biotin while pulling the C_α of either the N- or C-terminal amino acid residue, GLY16 or LYS134, respectively. As recently shown, for an accurate investigation of a pulling experiment, many simulation replicas are necessary, with simulation and experiments performed as similarly as possible.[5] Here, for both configurations, many replicas were performed. For each, N- and C-terminal pulling, 25 replicas were performed at 0.05 Å/ns pulling speed, with each of the

50 pulling simulations ranging from 350 to 400 ns total simulation time. Additionally, for C-terminal pulling, 50 replicas were performed at 0.5 Å/ns pulling speed, and 50 replicas at 5.0 Å/ns. In total, approximately 19 μs of production SMD were performed using GPU-accelerated XK nodes of the NCSA/Blue Waters supercomputer. The SMD procedure is equivalent to attaching one end of a harmonic spring to the end of a molecule and pulling on the end of the other molecule with another spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the z-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the z-axis. Owing to the flexibility of the experimentally employed linkers connecting the domains of interest and the fingerprint domains, this approach reproduces the experimental protocol.

I.5.D Simulation analysis

Simulation force-time traces were analyzed analogously to experimental data. For each simulation, the rupture force was determined as the highest force of a trace and the force loading rate was determined as a linear fit to the force versus time traces immediately before rupture. Analyses of force traces and MD trajectories, except for the force propagation analyses, were carried out employing python scripts taking advantage of Jupyter Notebooks.[19] Particularly, MDAnalysis,[20] and PyContact[21] were employed for trajectory analysis together with in-house scripting wrappers, which collected information from all simulation replicas. Mutual information coefficients were calculated[22] to identify observables that are closely related to the force traces of the simulated replicas. This process, known as “feature selection”, took into consideration contacts between biotin and streptavidin (calculated using PyContact), and RMSF values (calculated using MDAnalysis).

Force propagation analyses were performed using dynamical network analysis, which is implemented in VMD’s[12] Network View plugin.[23] A network was defined as a set of nodes, all α-carbons, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and two monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (non-hydrogen) from the two monomers are within 4.5 Å of each other for at least 75% of the frames analyzed. Filtering this network, one can investigate allosteric

signaling.[24] Allostery can be understood in terms of pathways of residues that efficiently transmit energy, here in the form of mechanical stress, between different binding sites.[25] The dynamical networks were constructed from 20 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as w_{ij}=-log(|C_{ij}|), where C is the correlation matrix calculated with Carma.[26] Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was -log(0.5)=0.69. As previously demonstrated by our group,[25] Pearson correlation is ideal for force propagation calculation.

II. Supporting data

II.1. Streptavidin and biotin structure

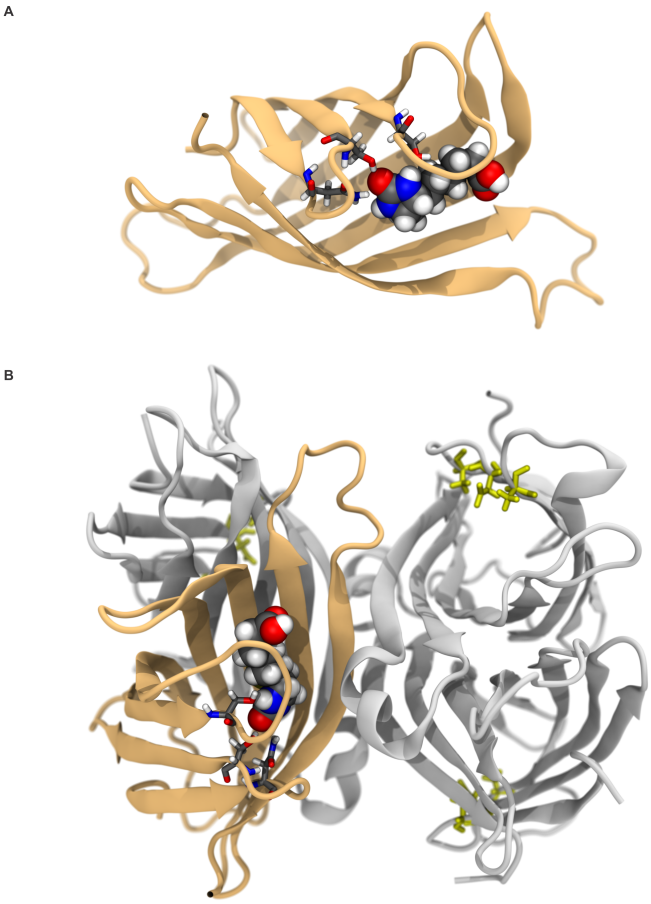


Figure S3. Crystal structure of mSA. (A) Functional SA subunit with biotin in the binding pocket. Asn23, Ser27, and Ser45 (the residues that are mutated in the non-functional subunits) are shown as sticks and colored by element. They are all located towards the N-terminus. (B) mSA tetramer. The functional subunit is shown as before. Non-functional subunits are depicted in grey. Mutated residues are highlighted as yellow sticks.

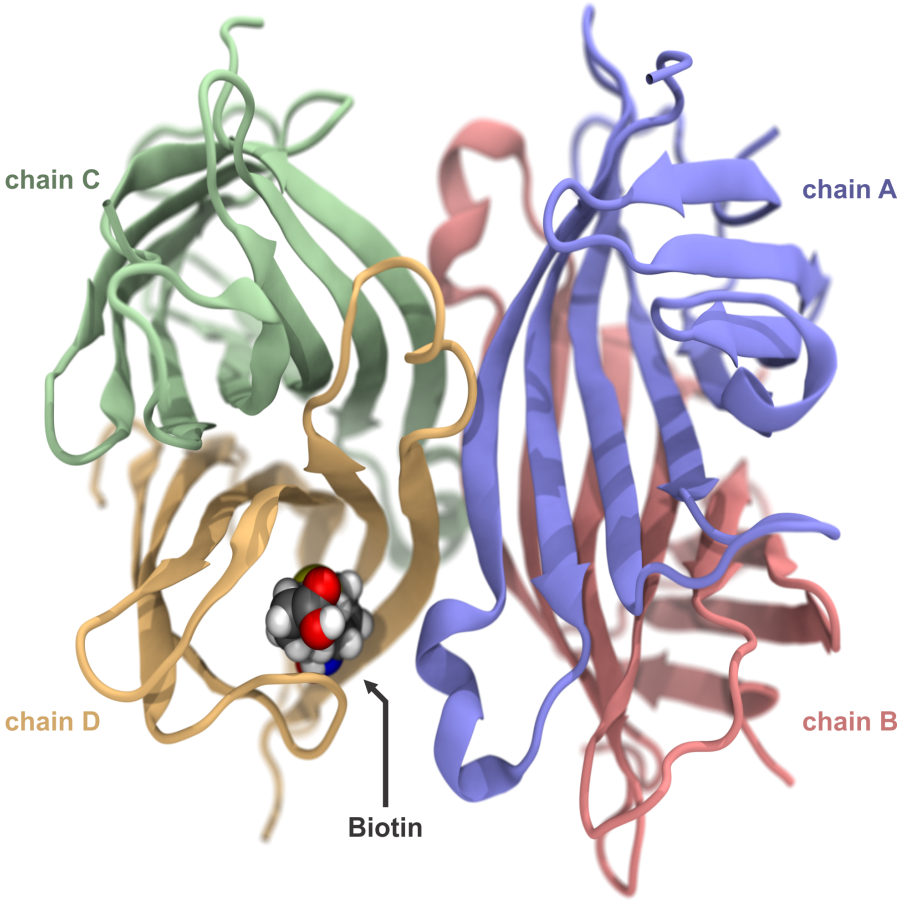


Figure S4. Tetrameric structure of mSA. The functional subunit (chain D) of SA with biotin bound is shown in bright orange. The non-functional subunits are shown in light purple (chain A), red (chain B), and green (chain C). Residues in the helical loop of chain A stabilize the biotin binding.

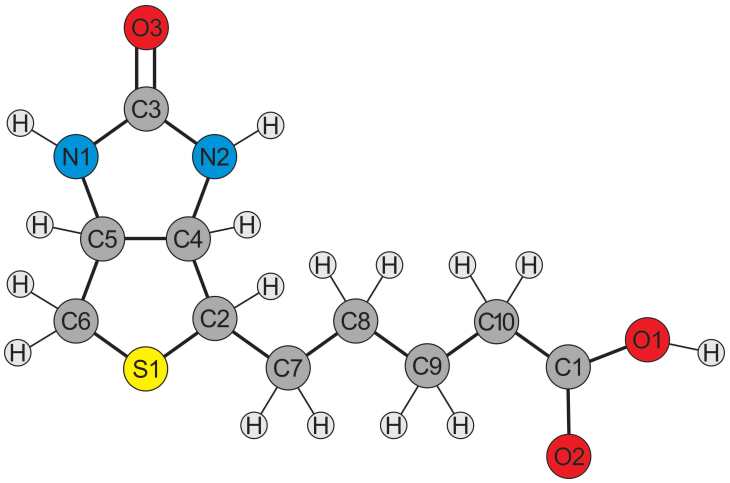


Figure S5. Representation of the chemical structure of the biotin. The structure shows the naming of the non-hydrogen atoms as they used in the simulations and analysis.

II.2. Isothermal Titration Calorimetry

To ascertain that the observed force difference results from the tethering geometry and is not caused by the tags fused to mSA, we performed isothermal titration calorimetry at 25°C with both mSA variants (Fig. S1). Titrating biotin into the reaction chamber containing either N-mSA or C-mSA, we determined the reaction enthalpy and stoichiometry of the binding. For comparability, we used the same biotin stock solution for all measurements. Every mSA was measured three times and the data were fitted individually. No significant difference in stoichiometry or binding enthalpy was observed. Mean and standard error of the three fits to N-mSA resulted in N=1,0±0,1 and ΔH=(-25,2±0,2) kcal/mol. For the fits to the C-mSA data, we obtained N=0,98±0,04 and ΔH=(-25,3±0,2) kcal/mol. For both mSA, the binding affinity was outside of the measurement range of our instrument but an upper limit of 1 nM could be assigned. The data are in good agreement with previously reported values for mSA. The binding enthalpy is comparable to wildtype SA.[1, 27] We thus conclude that in the absence of force, both N-mSA and C-mSA have one subunit that binds biotin in the way the SA wildtype does.

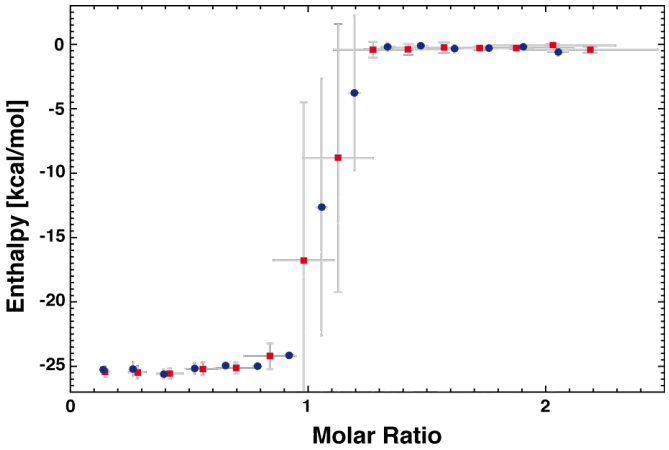


Figure S6. Isothermal titration calorimetry. The reaction enthalpy of biotin with N-mSA (blue circles) and C-mSA (red squares) was measured at 25°C. For all measurements the same biotin stock solution was used. The grey error bars are given by the standard deviation between three measurements.

II.3. Additional SMFS data of measurement shown in main text

In AFM-based SMFS, the use of so-called fingerprint domains is important. First, true single-molecule interactions are identified by the distinct unfolding pattern (force drop and contour length increment) of the fingerprint domain. Second, the unfolding force of the fingerprint domain serves as internal force reference. In our case, we used the well-characterized ddFLN4 as fingerprint domain. It unfolds in a two-step unfolding process. Histograms of the two corresponding unfolding force peaks are plotted in Sup. S-Fig. S7 and Sup. S-Fig. S8. The force-extension traces analyzed here are the same as in Fig. 3B. We do not observe any significant difference in unfolding forces of ddFLN4, when pulling on the C-mSA or the N-mSA area.

For the measurement shown, we performed 72,894 approach-retraction-cycles. In 26,245 cases (36%), we observed interactions between tip and surface higher than 50 pN. We identified specific single-molecule interactions in 5576 force-extension traces using the ddFLN4 fingerprint (7.6%).

In Sup. S-Fig. S9, we provide unbinding force histograms of final unbinding peaks observed in force-extension traces with interactions higher than 50 pN. (The final unbinding force can still be lower than 50 pN.) In these histograms, we did not use the ddFLN4 fingerprint pattern for sorting and thus show the unsorted data. Nevertheless, the specific final unbinding force peaks of the mSA:biotin interaction are clearly visible. Thus, the difference in unbinding force between N-mSA and C-mSA is already visible in the raw data. We interpret this as follows: Even if there are incomplete ddFLN4 unfolding patterns, additional unspecific or multiple specific interactions between surface and cantilever tip, the final force peak in the force-extension trace will, in most cases, be caused by a single biotin molecule unbinding from a single mSA molecule.

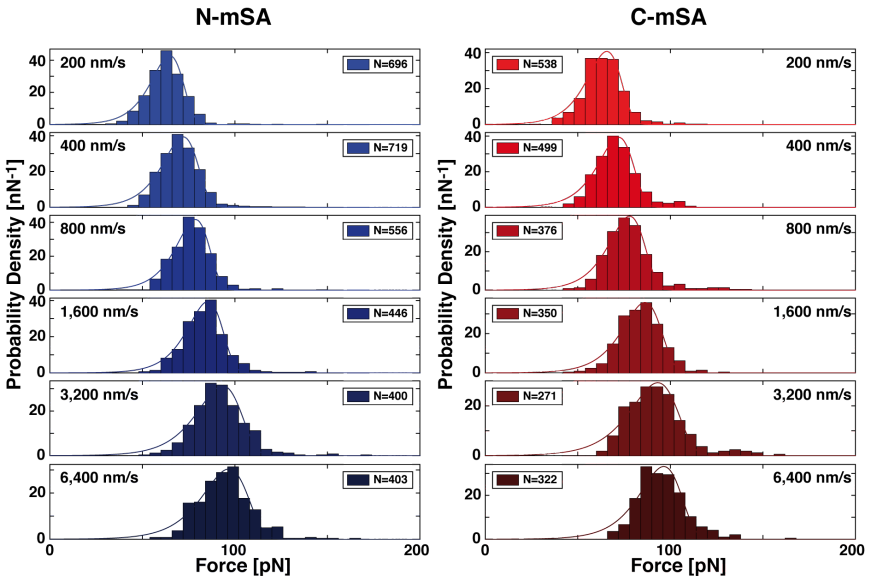


Figure S7. Unfolding forces of ddFLN4. Forces of the first unfolding peak of ddFLN4 in the case of N-terminal and C-terminal mSA surface attachment. Although the unbinding forces of biotin from C-mSA and N-mSA are quite different, the unfolding force of the fingerprint domain is not affected. The force extension curves used here are the same as for Figure 2. The number of interactions for every histogram is given in the boxes.

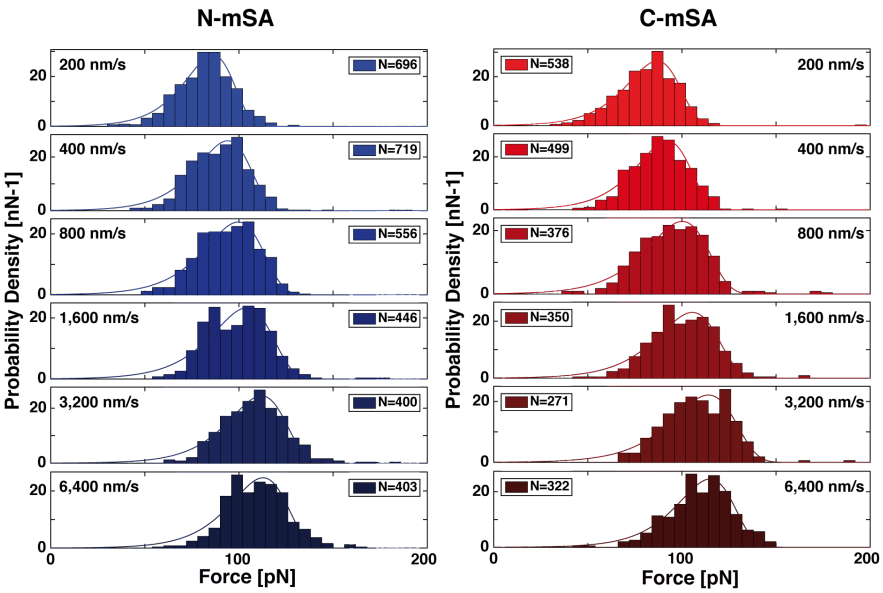


Figure S8. Unfolding forces of ddFLN4. Forces of the second unfolding peak of ddFLN4 in the case of N-terminal (blue colors) and C-terminal (red color) mSA surface attachment. Although the unbinding forces of biotin from C-mSA and N-mSA are quite different, the unfolding force of the fingerprint domain is not affected. The force extension curves used here are the same as for Figure 3B. The number of interactions for every histogram is given in the boxes.

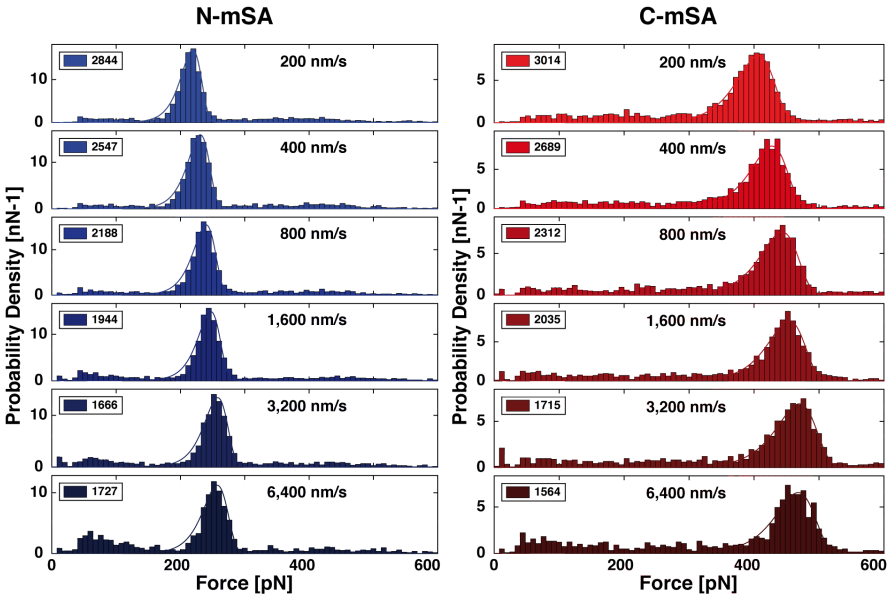


Figure S9. Unbinding forces of biotin from mSA. All final unbinding forces for interactions larger than 50 pN are binned into histograms and fitted with Bell-Evans distributions. N-terminal and C-terminal surface attachment of mSA is shown. Data are not sorted by the fingerprint pattern, *i.e.* they can contain multiple interactions or incomplete unfolding. The fact that these histograms still resemble the ones in Figure 3B implies that the final rupture peak in a force extension curve is in most cases caused by a single biotin unbinding from a single mSA. The number of interactions for every histogram is given in the boxes.

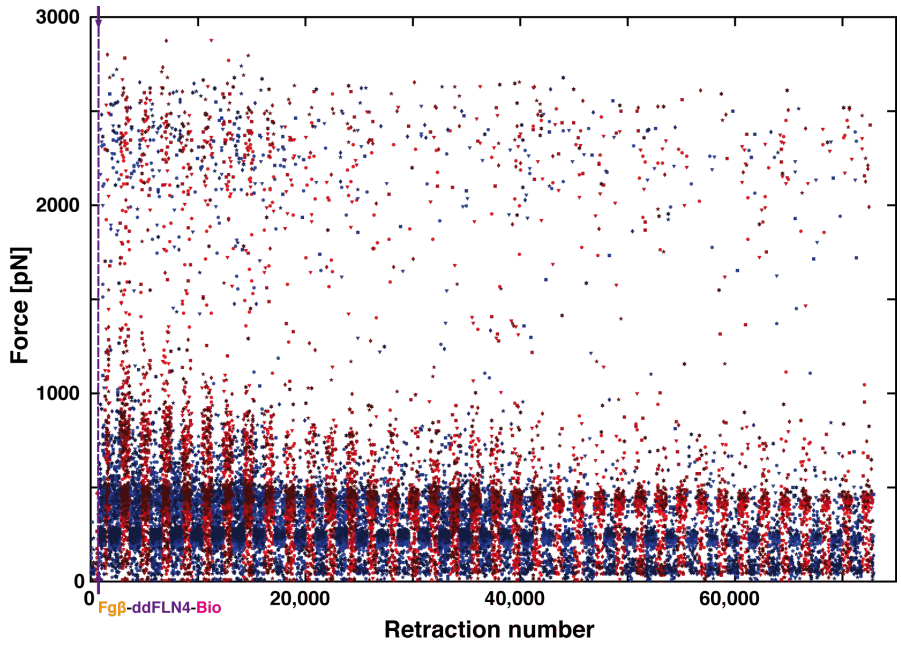


Figure S10. Course of the measurement over the complete range of forces. Here, we show the final unbinding forces for all interactions higher than 50 pN. N-mSA data are shown in blue colors, C-mSA data in red colors. Darker colors represent faster retraction velocities. The two dense bands at 200 pN or 400 pN, respectively, contain most of the specific single-molecule interactions that show the specific fingerprint unfolding pattern. At the beginning of the measurement multiple interactions of mSA:biotin are visible (about 400 pN for N-mSA, about 800 pN for C-mSA). In addition, in rare cases Fgβ:SdrG unbinding, probably caused by multiple and unspecific surface interactions, is visible at about 2500 pN. At the end of the measurement, the number of non-specific interactions at low forces increases, probably because the quality of the cantilever tip or its functionalization slowly decreases over time.

II.4. AFM-based SMFS with covalent attachment of ddFLN4

The high affinity of biotin for SA can be disadvantageous for SMFS experiments. The cantilever tip can easily get clogged when SA that was unspecifically adsorbed to the surface binds to the biotinylated proteins on the cantilever tip. S-In this study, we introduced a second receptor-ligand pair (SdrG:Fgβ) with a lower affinity to prevent clogging of the cantilever tip. S-In doing so, we obtained good statistics and were able to perform daylong measurements.

The use of a second receptor-ligand pair is counter-intuitive, as one could argue that mSA:biotin and unbinding SdrG:Fgβ cannot be distinguished or that the SdrG:Fgβ system somehow influences the behavior of the mSA:biotin unbinding. The first concern is negligible, because in the force loading geometry used here, SdrG:Fgβ was shown to withstand forces of 2 nN.[5] To tackle the second concern, we performed additional control measurement without the second receptor-ligand system, covalently attaching biotinylated ddFLN4 to the cantilever tip. S-For this, we again immobilized both mSA variants at different areas on the surface, a few millimeters apart from each other (Sup. S-Fig. S11A). In this way, we could probe both variants with the same cantilever. The cantilever tip, covalently functionalized with biotinylated ddFLN4, is approached to the surface. Biotin binds to mSA on the surface. Retracting the cantilever, the force needed to unbind biotin from mSA is measured.

ddFLN4 on the cantilever unfolds at lower forces than biotin unbinds from mSA. The characteristic two-step unfolding pattern is used to identify single-molecule interactions, *i.e.* a single biotin molecule binding to a single mSA molecule. In Sup. S-Fig. S11B, two exemplary force-extension traces for single-molecule interaction on the area where N-mSA or C-mSA were immobilized are depicted. In both cases, the AFM cantilever was retracted with a velocity of 800 nm/s. While the ddFLN4 unfolding is observed at the same force of about 60-80 pN for both traces, the final force peak is situated at different forces. This last peak is attributed to the unbinding of biotin from mSA.

Selecting all force curves that clearly show single-molecule interaction, we can plot all the mSA:biotin unfolding forces for all approach retraction cycles (Sup. S-Fig. S11C). Out of the first 6,000 traces, 2,640 showed interaction (44%). We could identify 557 single-molecule interactions (9.3%). The loss in interaction frequency over time is due to clogging of the lever by mSA that had been unspecifically adsorbed to the surface. Already a low amount of mSA is sufficient to obstruct the measurement. Force curves measured on the N-mSA area are depicted

in blue colors, while data collected on the C-mSA area is shown in red colors. Faster retraction velocities are represented by darker colors: 200 nm/s (circles), 800 nm/s (triangles), 3,200 nm/s (squares). Already from this data set, it is obvious that the mechanical stability of the mSA:biotin interaction is different depending on mSA tethering. Furthermore, comparing the data with the one in the main matter, we can show that the introduction of the second receptor-ligand system has no significant influence on the unbinding forces. It only helps to get better statistic by preventing lever clogging.

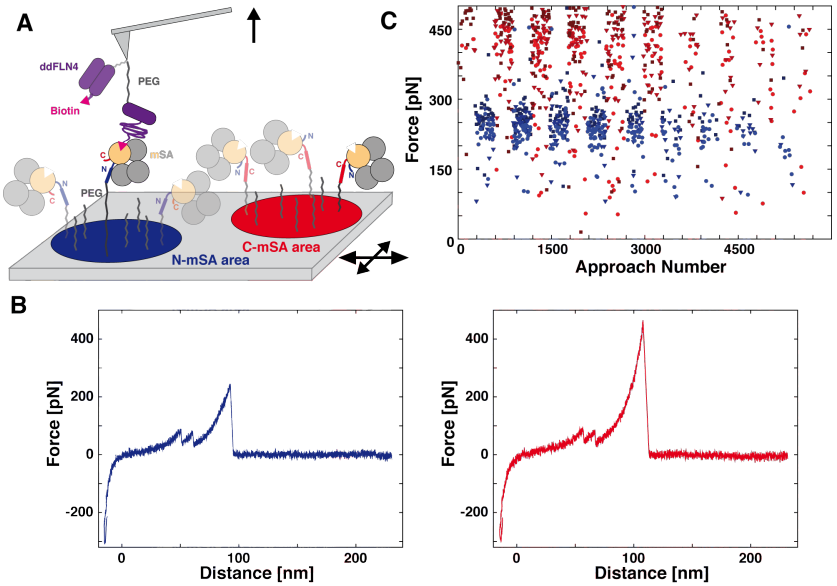


Figure S11. SMFS Experiment with N-mSA and C-mSA. (A) Schematic of the experiment: In different areas on the surface, N-mSA (blue) or C-mSA (red) is immobilized via PEG-linkers. With the AFM cantilever tip, decorated with biotinylated ddFLN4 (purple), both areas are probed. When the cantilever approaches the surface, biotin (magenta) binds to mSA. Upon retraction of the cantilever, ddFLN4 unfolds before biotin unbinds from mSA. (B) Typical force-extension traces for N-mSA (left, blue) and C-mSA (right, red) at 800 nm/s: The characteristic unfolding pattern of ddFLN4 serves as a fingerprint to identify single mSA:biotin interactions. (C) Every 300 approach-retraction cycles, the cantilever is moved to the other area. Final unbinding forces for mSA:biotin are plotted for different retraction velocities. C-mSA: red colors; N-mSA: blue colors; 200 nm/s bright circles; 800 nm/s triangles; 3,200 nm/s dark squares.

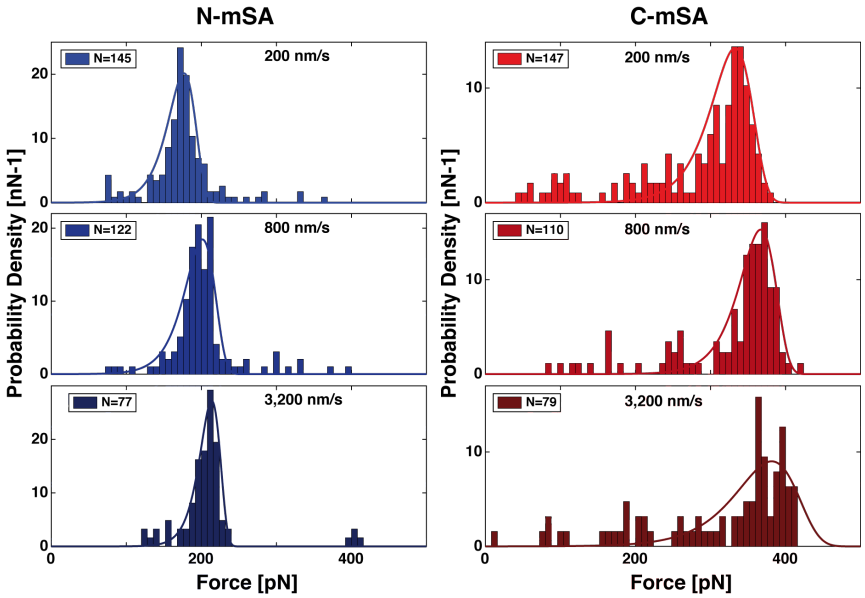


Figure S12. Unbinding force histograms for experiments with covalent attachment of ddFLN4 to the cantilever tip. S-Forces recorded on the N-mSA spot are colored in blue colors, forces recorded on the C-mSA area in red colors. Although the overall yield is much lower than for non-covalent Fgβ:SdrG tethering, due to cantilever clogging, the distribution of rupture forces are qualitatively comparable with the ones in Figure 3B.

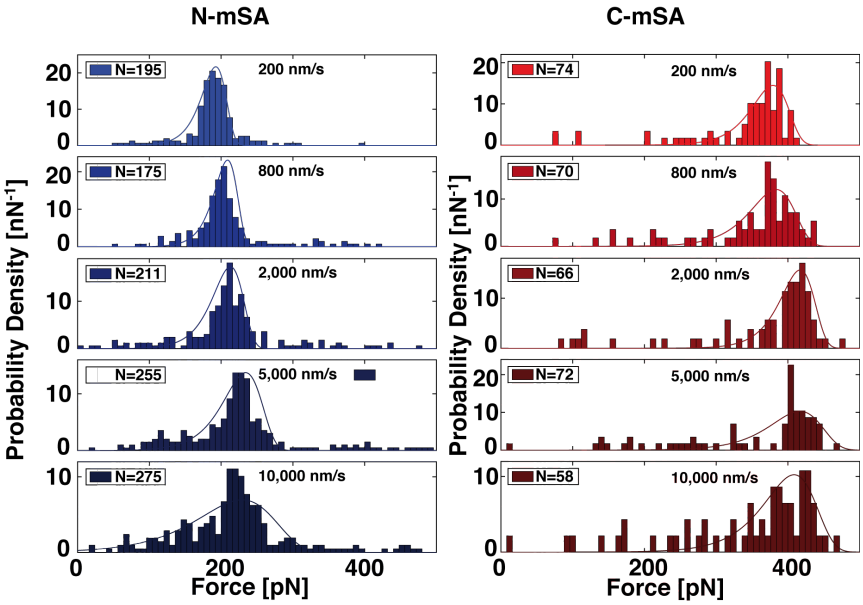


Figure S13. Unbinding force histograms for experiments with covalent attachment of ddFLN4 to the cantilever over a wider range of force loading rates. Forces recorded on the N-mSA spot are colored in blue colors, forces recorded on the C-mSA area in red colors. Although the overall yield is lower than for non-covalent Fgβ:SdrG tethering, due to cantilever clogging, the distribution of rupture forces are qualitatively comparable with the ones in Figure 3B. Over the range of retraction velocities, and thus force loading rates, applied here a transition from one unbinding pathway to the other, which might be expected for N-mSA, could not be resolved.

III. Data analysis

III.1. Fitting of the Bell-Evans model

Bell, Evans, and Ritchie developed a model to characterize how an external force affects the dissociation of molecular bonds.[28-29] Under the assumption of a constant force loading rate *r*, the probability density function of bond rupture under force is calculated. Although the force loading rate is not constant for AFM force spectroscopy using constant retraction velocity, as is the case for our experiments, the model is still often used to fit rupture force histograms. Here, we employed the Bell-Evans model to fit the rupture force distributions in Figure 3B:

$$p(F) = s \cdot a \cdot \exp(b \cdot F) \cdot \exp\left(\frac{a}{b} \cdot (1 - \exp(b \cdot F))\right)$$

As the probability density function *p(F)* is normalized to unity, we employed a factor *s* to take into account that there is a fraction of rupture forces outside the main peak. To determine the factor *s*, we first identified the most probable rupture force <*F*>, *i.e.* the maximum of the peak, by fitting a kernel density estimate to the unbinding force histogram and taking the maximum. We then divided the number of unbinding forces within an arbitrarily chosen range of 100 pN below and above the peak force by the number of total forces in the histogram to obtain the factor *s*.

From the fitting parameters *a* and *b*, physical parameters can be derived:

$$a = \frac{k_{off,0}}{r}$$

$$b = \frac{x_0}{k_B T}$$

Here, k_B is the Boltzmann factor, T the absolute temperature, x_0 the distance to transition state, r the force loading rate, and $k_{off,0}$ the off-rate at zero force. The physical parameters derived here, especially the zero-force off-rate, have to be taken with caution. As stated previously, the force loading rate is not constant. To determine the zero-force off-rates, we divided the fitting parameter *a* by the most probable force loading rate <*r*>. The latter was determined by fitting the last 3 nm before the rupture peak in every the force-extension curve, plotting a histogram of these fitted loading rates and finally taking the maximum of a kernel density estimate fitted to this histogram. The distance to transition state was determined by multiplying the fitting parameter *b* with the Boltzmann factor k_B and the temperature T = 300 K.

These parameters can also be derived from the force loading rate plot (Figure 3C). Here, we employed the following equation:

$$< F >= \frac{k_B T}{x_0} \log \left(\frac{r}{k_{off,0}} \frac{x_0}{k_B T}\right)$$

The fitting error of the zero-force off-rate is quite large due to the exponential relation of loading rate and force. The values for the distance to transition state differ by a factor of two depending on whether they are derived from the dynamic force spectrum or from the probability density function to a single rupture force distribution. Nevertheless, the relative difference between N-SA and C-SA is the same, namely, that the distance to transition state is twice as large for N-SA, indicating that the potential well is narrower for C-terminal loading.

Table S1. Fitting parameters for Figure 3B.

	v_R [nm/s]	s [%]	a [$\times 10^3$]	b [$\times 10^{10}$]
N-SA	**200**	72.4	26.4	5.92
	400	72.4	31.6	5.58
	800	71.4	74.4	5.24
	1600	68.4	43.1	5.89
	3200	65.3	58.7	6.34
	6400	58.3	61.1	5.98
C-SA	**200**	61.4	23.0	3.50
	400	59.1	6.0	3.65
	800	58.9	4.8	3.50
	1600	57.9	3.0	3.59
	3200	56.4	5.6	3.32
	6400	52.8	6.5	3.37

Table S2. Physical parameters extracted from the data in Figure 3B.

	v_R [nm/s]	<F> [pN]	<r> [pN/s]	x_0 [nm]	$k_{off,0}$ [s^{-1}]
N-SA	**200**	216	1,500	0.24	4.0e-5
	400	228	3,300	0.23	1.0e-4
	800	237	7,000	0.21	5.2e-4
	1600	243	14,400	0.24	6.2e-7
	3200	254	31,400	0.26	1.8e-6
	6400	253	56,700	0.24	3.5e-6
C-SA	**200**	403	3,500	0.14	8.1e-5
	400	423	7,700	0.15	4.6e-5
	800	438	15,900	0.14	7.6e-5
	1600	449	34,400	0.15	1.0e-7
	3200	464	72,900	0.13	4.1e-7
	6400	461	148,000	0.14	9.6e-7

IV. Supporting simulation data

IV.1. Cross-correlation and network analysis

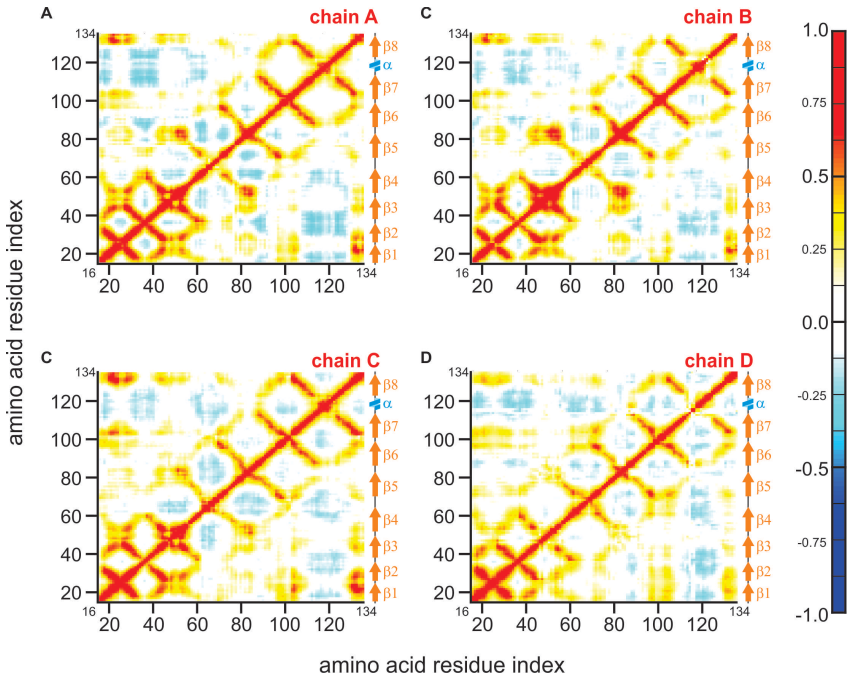


Figure S14. Cross-correlation matrices for C-terminal pulling. The plots show the cross-correlation between the residues within each of the four SA subunits. The β-sheet structure results in high correlation between the corresponding residues (red lines close to the diagonals). For chain D, this structure is slightly less pronounced but is in principle preserved. In particular, the high cross-correlation between the first and the second β-strand is mostly maintained (bottom left corner).

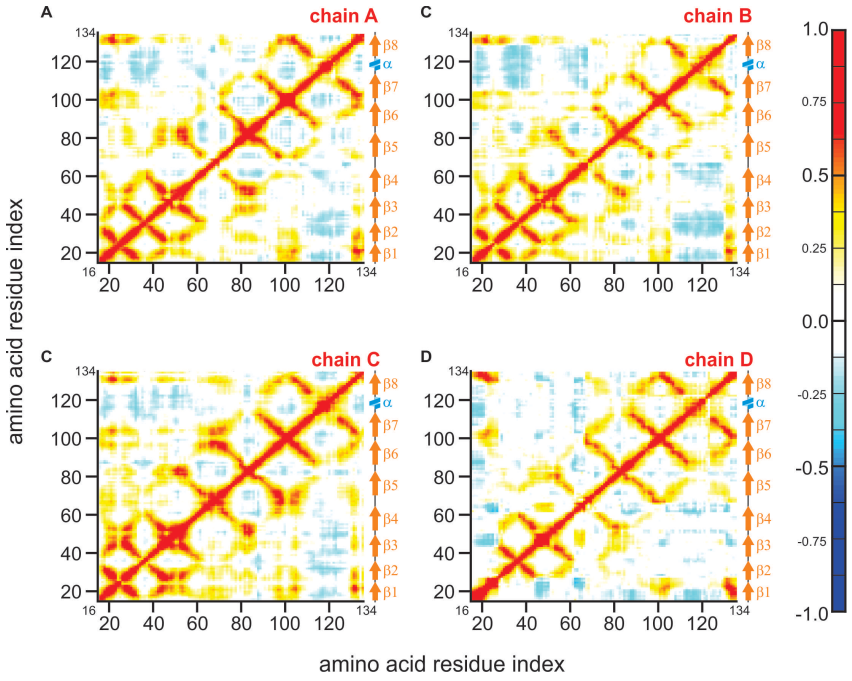


Figure S15. Cross-correlation matrices for N-terminal pulling. The plots show the cross-correlation between the residues within each of the four SA subunits. The β-sheet structure results in high correlation between the corresponding residues (red lines close to the diagonals). For chain D, this structure is less pronounced. In particular, the high cross-correlation between the first and the second β-strand is lost (bottom left corner of D). Instead, high cross-correlation between the first (N-terminal) and eighth (C-terminal) β-strand is observed. This is in line with the rearrangement of the first β-strand under force shown in Fig. 4.

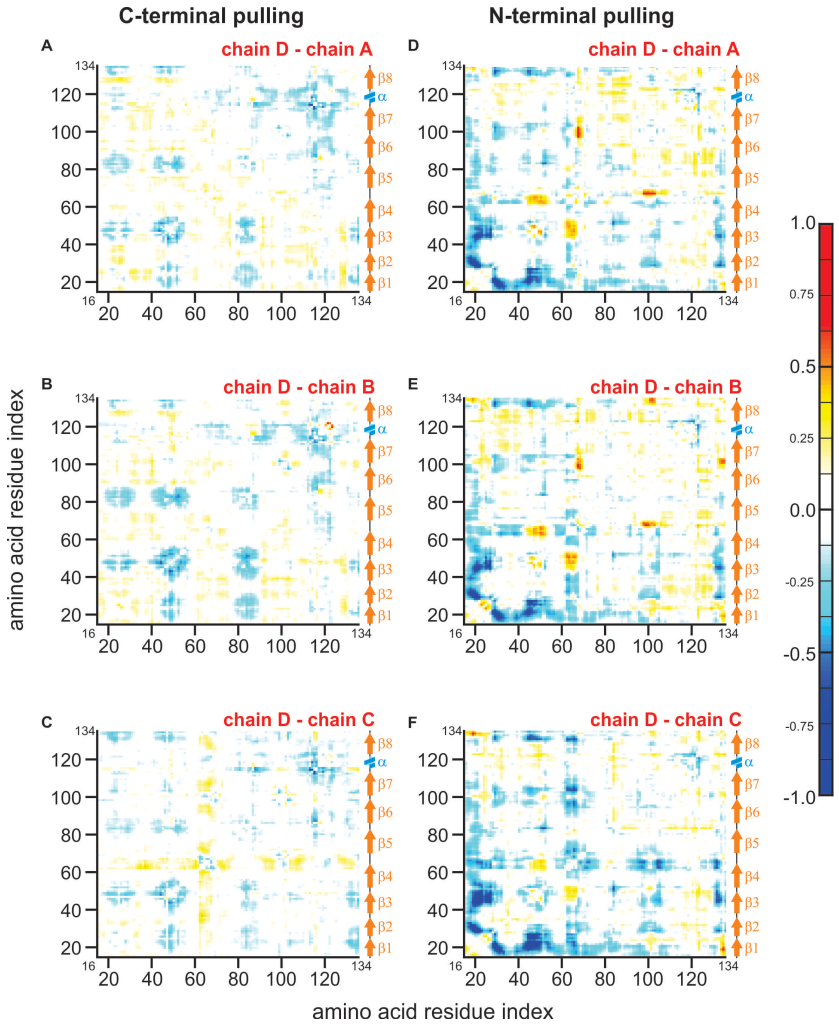


Figure S16. Difference in cross-correlation matrices between pulling and non-pulling chains. The cross-correlation matrices shown before are subtracted here. Differences in cross-correlation between the different subunits become visible. In addition, C-terminal (A-C) and N-terminal force loading (D-F) can directly be compared. The structural rearrangement for N-mSA is clearly visible in the bottom left corner of the plots D-F. For C-mSA, differences in cross-correlation are overall less pronounced. For some linker regions, *e.g.* between the fourth and fifth β-strand, cross-correlations are higher in the subunit that is under force load (chain D).

IV.2. Mutual information analysis (RMSF, Contacts, Force)

Among all biotin-streptavidin contacts, and the residues with highest RMSF values, the feature selection methodology listed observables with from highest to lowest mutual information (MI) coefficient. The observables with highest MI are analyzed here, and listed in Tables S3 and S4.
SER27 constitutes one of three mutations used to weaken biotin binding to create the non-functional subunits utilized to generate mSA.[30] It appears to play a major role in both N- and C-terminal pulling experiments. The three amino acids ASN49, TYR43 and SER27 were identified to generate an H-bond network perpendicular to the force loading during unbinding of biotin.[31] This way, only a simultaneous rupture allows biotin in a final step to escape the binding pocket. Also, Liu et al. showed that SER27, TYR43, THR90 contribute to binding affinity.[32] Dixon and Kollman demonstrated the importance of TRP79 and SER27 for SA biotin binding strength.[33] The van der Waals contacts of the tryptophan residues TRP79, TRP108, TRP120 (TRP120 of subunit A) are also known to contribute to the high affinity.[34] Residues GLY16, ILE17, and THR18 are the residues at the N-terminus forming the first β-strand that is ripped away from the second β-strand when force is applied.

Table S3. Results of Mutual Information Analysis. A score function has been developed and applied to identify the most important residues of mSA for force propagation in the case of N- or C-terminal pulling of mSA. Marked in red are the amino acid residues that presented a correlation between RMSF and forces, while the other residues present a correlation between a contact to biotin and force.

N-mSA			C-mSA		
Residue	Biotin Atom	Score	Residue	Biotin Atom	Score
SER27	O3	1.000000	SER27	O3	1.000000
TRP108	C6	0.778314	TRP108	C5	0.880444
ALA63		0.758845	TRP79	C8	0.866975
THR90	S1	0.753362	TYR43	O3	0.781555
TYR43	O3	0.687164	TRP79	C3	0.735351
ASN49	O2	0.650896	TRP79	C9	0.683922
GLY16		0.626958	THR90	S1	0.662114
TRP79	C9	0.616018	TRP79	C10	0.624008
ILE17		0.614522	TRP79	C6	0.579682
TRP120	C7	0.570383	TRP108	C6	0.562329
TRP79	C8	0.565346	ASN49	O1	0.553145
THR18		0.516329	TRP79	N1	0.537884
THR66		0.510152	TRP120	O3	0.524486

Table S4. Results of Mutual Information Analysis C-terminal pulling versus N-terminal pulling. Marked in red are the amino acid residues that presented a correlation between RMSF and forces, while the other residues present a correlation between a contact to biotin and force. The mutual information scores are not as high as those observed in Table S3, reinforcing that the mechanism that differentiates C- versus N-terminal pulling is related to partial unfolding and not to the contacts or other structural fluctuations.

Residue	Biotin Atom	Score
ASN118		0.337043
SER45	O3	0.218223
TRP79	C6	0.136393
TRP108	C5	0.12872
TRP79	C2	0.115025
GLY98		0.112509
ASN49	O1	0.10985
THR90	S1	0.104352
TYR83		0.088686
ALA117		0.072231
TRP79	C5	0.061104
GLY68		0.055532
TRP79	N1	0.04377
ASN49	O2	0.039842
SER69		0.03954
TRP120	O3	0.027222
GLY99		0.023102
ALA100		0.018072
TRP79	C8	0.011375

V. Supplementary references

1. Sedlak, S. M.; Bauer, M. S.; Kluger, C.; Schendel, L. C.; Milles, L. F.; Pippig, D. A.; Gaub, H. E., Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. *PLOS ONE* **2017,** *12* (12), e0188722.

2. Schwaiger, I.; Kardinal, A.; Schleicher, M.; Noegel, A. A.; Rief, M., A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat Struct Mol Biol* **2004,** *11* (1), 81-5.

3. Yin, J.; Straight, P. S-D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T., Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* **2005,** *102* (44), 15815-20.

4. Zimmermann, J. L.; Nicolaus, T.; Neuert, G.; Blank, K., Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat Protoc* **2010,** *5* (6), 975-85.

5. Milles, L. F.; Schulten, K.; Gaub, H. E.; Bernardi, R. C., Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science* **2018,** *359* (6383), 1527-1533.

6. Gumpp, H.; Stahl, S. W.; Strackharn, M.; Puchner, E. M.; Gaub, H. E., Ultrastable combined atomic force and total internal reflection fluorescence microscope [corrected]. *Rev Sci Instrum* **2009,** *80* (6), 063704.

7. Ribeiro, J. V.; Bernardi, R. C.; Rudack, T.; Stone, J. E.; Phillips, J. C.; Freddolino, P. S-L.; Schulten, K., QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts. *Sci Rep* **2016,** *6*, 26536.

8. Seppala, J.; Bernardi, R. C.; Haataja, T. J. K.; Hellman, M.; Pentikainen, O. T.; Schulten, K.; Permi, P.; Ylanne, J.; Pentikainen, U., Skeletal Dysplasia Mutations Effect on Human Filamins' Structure and Mechanosensing. *Sci Rep* **2017,** *7* (1), 4218.

9. Verdorfer, T.; Bernardi, R. C.; Meinhold, A.; Ott, W.; Luthey-Schulten, Z.; Nash, M. A.; Gaub, H. E., Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to

Characterize and Tune Cellulosomal Scaffoldin Mechanics. *J Am Chem Soc* **2017,** *139* (49), 17841-17852.

10. Zhang, M.; Biswas, S.; Deng, W.; Yu, H., The Crystal Structure of Monovalent Streptavidin. *Sci Rep* **2016,** *6*, 35915.

11. Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S-S., Cooperative hydrogen bond interactions in the streptavidin-biotin system. *Protein Sci* **2006,** *15* (3), 459-67.

12. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.

13. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. S-E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* **2012,** *8* (9), 3257-3273.

14. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.

15. Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z., NAMD goes quantum: an integrative suite for hybrid simulations. *Nat Methods* **2018,** *15* (5), 351-354.

16. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

17. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** *26* (16), 1781-802.

18. Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K., Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J* **1997,** *72* (4), 1568-81.

19. Kluyver, T.; Benjamin, R.-K.; Fernando, P.; Brian, G.; Matthias, B.; Jonathan, F.; Kyle, K.; Jessica, H.; Jason, G.; Sylvain, C.; Paul, I.; Damián, A.; Safia, A.; Carol, W.; Jupyter Development, T. In *Jupyter Notebooks—a publishing format for reproducible computational workflows*, Positioning and Power in Academic Publishing: Players, Agents and Agendas, 20th International Conference on Electronic Publishing, Göttingen, Germany, 2016; Göttingen, Germany, 2016.

20. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **2011,** *32* (10), 2319-27.

21. Scheurer, M.; Rodenkirch, P.; Siggel, M.; Bernardi, R. C.; Schulten, K.; Tajkhorshid, E.; Rudack, T., PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations. *Biophys J* **2018,** *114* (3), 577-583.

22. Kraskov, A.; Stogbauer, H.; Grassberger, P., Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys* **2004,** *69* (6 Pt 2), 066138.

23. Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z., Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A* **2009,** *106* (16), 6620-5.

24. Vanwart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E., Exploring residue component contributions to dynamical network models of allostery. *J Chem Theory Comput* **2012,** *8* (8), 2949-2961.

25. Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E., Mapping Mechanical Force Propagation through Biomolecular Complexes. *Nano Lett* **2015,** *15* (11), 7370-6.

26. Glykos, N. M., Software news and updates. Carma: a molecular dynamics analysis program. *J Comput Chem* **2006,** *27* (14), 1765-8.

27. Chilkoti, A.; Stayton, P. S-S., Molecular Origins of the Slow Streptavidin-Biotin Dissociation Kinetics. *Journal of the American Chemical Society* **1995,** *117* (43), 10622-10628.

28. Bell, G. I., Models for the specific adhesion of cells to cells. *Science* **1978,** *200* (4342), 618-27.

29. Evans, E.; Ritchie, K., Dynamic strength of molecular adhesion bonds. *Biophys J* **1997,** *72* (4), 1541-55.

30. Howarth, M.; Chinnapen, D. J. F.; Gerrow, K.; Dorrestein, P. S-C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y., A monovalent streptavidin with a single femtomolar biotin binding site. *Nature Methods* **2006,** *3*, 267.

31. Rico, F.; Russek, A.; Gonzalez, L.; Grubmuller, H.; Scheuring, S., Heterogeneous and rate-dependent streptavidin-biotin unbinding revealed by high-speed force spectroscopy and molecular dynamics simulations. *arXiv preprint arXiv:1808.07122* **2018**.

32. Liu, F.; Zhang, J. Z. H.; Mei, Y., The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Scientific Reports* **2016,** *6*, 27190.

33. Dixon, R. W.; Kollman, P., The free energies for mutating S27 and W79 to alanine in streptavidin and its biotin complex: The relative size of polar and nonpolar free energies on biotin binding. *Proteins: Structure, Function, and Bioinformatics* **1999,** *36* (4), 471-473.

34. Chilkoti, A.; Tan, P. S-H.; Stayton, P. S-S., Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. *Proceedings of the National Academy of Sciences* **1995,** *92* (5), 1754-1758.

3.3 Manuscript M2: Streptavidin/Biotin: Tethering Geometry defines Unbinding Mechanics

In Manuscript M2, new insights into the mechanics of the SA/biotin complex under load are reported, which allow rethinking the diversity of rupture forces reported in the literature on the SA/biotin interaction. Unraveling the mechanism of this complex binding is only possible by analyzing a large set of data produced with a combination of state-of-the-art *in silico* and *in vitro* force spectroscopy approaches.

25 years ago, the binding forces of this system have been resolved for the first time. Since then, people have been puzzled by the broad distribution of the measured unbinding forces, which are regularly reported also in nowadays publications. By taking the number of experimental and simulation replicas to a whole new level, the system is revisited in this work and the puzzle is resolved: Biotin molecules in the different pockets of the SA tetramer require different forces to be pulled out! The molecular mechanisms are analyzed and a strategy for monodisperse interaction is provided.

In this work, engineered SA molecules possessing one, three or four functional subunit are used in AFM-based SMFS experiments to study their interaction with biotin. A unique cysteine at the C-terminus of a certain subunit is introduced to enable site-specific immobilization onto a surface. With an AFM-cantilever tip, covalently functionalized with biotin, the different versions of SA are probed. Depending on which subunits biotin binds to, significantly different rupture forces, which are caused by four distinct force-loading geometries, are observed. As the exact molecular site of surface attachment is known, steered molecular dynamics simulations of the unbinding process that are in agreement with the experimental force-loading geometry are performed.

Experiments and simulations are used hand-in-hand, providing a detailed picture of the system mechanics with the atomistic detail of the simulations and the large statistical content of the experiments. With the results of the all-atom molecular dynamics simulations, a mechanism is discovered how, upon force loading, biotin and adjacent linker molecules induce conformational changes in the SA molecule, disturbing the mechanical integrity of the binding pocket and lowering the dissociation energy barrier. In summary, it is shown that the way in which SA is tethered is of utmost importance for the mechanical stability of the bond.

For most application in bio-nanotechnology, biotin is attached to a molecular linker so that the experiments actually mimic the predominant situation for myriads of assays using SA/biotin. It is expected that the findings make a large impact on the way SA is applied in future force spectroscopy experiments.

Streptavidin/Biotin: Tethering Geometry Defines Unbinding Mechanics

by

Steffen M. Sedlak, Leonard C. Schendel, Rafael Bernardi, and Hermann E. Gaub

submitted

Streptavidin/biotin: tethering geometry defines unbinding mechanics

Steffen M. Sedlak,[1] Leonard C. Schendel,[1] Rafael C. Bernardi,[2*] and Hermann E. Gaub[1*]

[1]Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 Munich, Germany.

[2]NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

*corresponding authors: R.C.B. rcbernardi@ks.uiuc.edu, H.E.G. gaub@lmu.de

Abstract/Introductory paragraph

Multivalency has made the specific and long-lived biotin/streptavidin (SA) interaction indispensable for immobilization, labeling, and detection of molecules in bionanotechnology. Although SA's four subunits have the same affinity, we find that the forces required to break the biotin/SA bond depend strongly on the attachment geometry. We measured unbinding forces of biotin from different SA subunits to range from 100 pN to over 400 pN. Using a wide-sampling approach, we carried out hundreds of all-atom steered molecular dynamics (SMD) simulations for the entire system, including molecular linkers. Our strategy revealed the molecular mechanism that leads to a fourfold difference in mechanical stability: Certain force-loading geometries induce conformational changes in SA's binding pocket lowering the energy barrier, which biotin has to overcome to escape the pocket. In summary, we reconcile the inconsistent unbinding forces reported in the literature, revealing force-loading direction as an important parameter in receptor-ligand mechanical stability, particularly for biotin/SA.

Main text

To this day, tremendous effort has been invested to probe the mechanical strength of a single biotin/SA interaction. Previous studies, varying in instrumentation and immobilization strategies, found a wide range of unbinding forces for the biotin/SA complex.[1, 2, 3, 4, 5, 6] The underlying molecular mechanism for the mechanical stability of this complex has also been extensively investigated using computational tools.[6, 7, 8, 9] To consolidate the discrepancies in the reported unbinding forces, we investigated the unbinding of biotin from different SA subunits with total control of subunit geometry[10] by building on state-of-the-art site-specific immobilization strategies,[11] parallelized AFM-based single-molecule force spectroscopy (SMFS) of different molecular species on a single sample surface,[12] and the development of SA mutants with defined valencies.[13]

To prepare non-functional (0SA), monovalent (1SA), trivalent (3SA), and tetravalent (4SA) SA, we expressed different SA subunits separately and assembled them in a defined ratio (Supplementary Fig. 1). In the following, the subunit that contains a unique cysteine, *i.e.* the one that is attached to the surface in SMFS, is always denoted as subunit D. The other subunits are denoted accordingly, as given by the crystal structure in Fig. 1a. For AFM-based SMFS experiments, the different SA variants were covalently and site-specifically tethered in millimeter-separated spots on a glass slide (Fig. 1b). To circumvent inconsistencies of cantilever calibration and measurement conditions, all SA variants were probed with the same cantilever allowing for reliable and precise comparison of the resulting unbinding forces.[12]

As previously established,[9] cantilever clogging was avoided by using a proxy receptor ligand system: The adhesin SD-repeat protein G (SdrG) from *Staphylococcus epidermidis* and its binding partner, a short peptide from human fibrinogen β (Fgβ), were employed, because their rupture forces far exceed those of biotin/SA interaction.[14] Data with covalent attachment of biotin to the cantilever tip is provided as Supplementary Fig. 2. To unambiguously identify single molecule unbinding events, we employed *Dictyostelium discoideum's* fourth filament domains (ddFLN4),[15, 16] with an N-terminal Fgβ-peptide and a C-terminal biotin to establish a molecular link between SdrG on the cantilever tip and SA on the sample surface (Fig. 1b).

Out of 80,000 binding attempts, around 10,000 retraction traces showed interactions with forces higher than 50 pN (Supplementary Fig. 3). About one fifth showed the distinct two-step unfolding pattern of ddFLN4 before rupturing the biotin/SA complex (Supplementary Fig. 4), which are plotted as histograms of unbinding forces in Fig. 1c. On the surface where

0SA, the non-functional control mutant, was immobilized, only two events (out of 20,000 attempts) showing a ddFLN4-like force curve pattern were observed demonstrating the low level of non-specific interactions in the assay. For 1SA, the unbinding force histogram exhibits a single, most probable rupture force of 440 pN, fitted well by a Bell-Evans distribution[8, 17] for dissociation of a single bond in a single step Markovian manner. In contrast, the unbinding force histogram of 3SA exhibits two peaks at lower forces with maxima at 100 pN and 210 pN. The histogram can be fitted by a cumulative function of two Bell-Evans distributions. The unbinding force histogram of events recorded on the 4SA area reveals a combination of both, 1SA and 3SA, namely three distinct unfolding force peaks. The red and green dashed lines in the lowest panel of Fig. 1c are weighted 1SA and 3SA fits from before. Using a cumulative function of three Bell-Evans distributions results in a comparable fit. Fit formula and parameters are provided in Supplementary Notes and Supplementary Table 1.

Combining the functional subunits of 1SA and 3SA leads to 4SA. The same is true for the force histograms: The combination of the rupture force histograms of 1SA and 3SA resembles the histogram of 4SA. Thus, we interpret the data by attributing the different rupture force peaks in the histogram to unbinding of biotin from different SA subunits. Evidently, unbinding from subunit D can be attributed to the highest rupture force peak at 440 pN, as 1SA only shows this single peak. The attachment of the tetramer to the surface via subunit D might explain the comparatively low relative frequency of this rupture force event in the 4SA histogram due to lower accessibility of the subunit D binding pocket. The two remaining force peaks thus stem from biotin unbinding from subunits A, B and C.

To elucidate the underlying mechanism, we performed all-atom constant-velocity SMD[18] simulations using the same force loading geometry as for the SMFS experiment. Simulations of a fully solvated biotin/SA complex (Supplementary Fig. 5) were prepared following QwikMD[19] protocols and carried out with GPU-accelerated NAMD.[20] A wide-sampling approach was taken where hundreds of fully independent simulations were carried out, accounting for over 30 μs of production SMD runs. For simplicity, we always anchored the molecular linker of biotin bound to subunit D (PDB:5TO2,[21] biotin from PDB:1MK5[22]) and pulled on one of the four subunits by its C-terminus. This reproduces the four different experimental force-loading geometries. Furthermore, the simulations include part of the linker, which connects biotin to the Fgβ-ddFLN4-construct (Supplementary Fig. 6). We found omitting the linker yields significantly different results (Supplementary Fig. 7), presumably due to missing interactions between the linker and SA.

During SMD simulations, as the pulling and anchoring points are gradually separated at constant pulling velocity, the complex is free to rotate into an orientation maximizing the distance between the attachment points. This orientation defines the direction in which gradually a restoring force builds up in the molecular complex upon further separation. In Fig. 2a, the crystal structure of biotin/SA complex is depicted. For the binding pocket, a surface representation is chosen to illustrate the spatial confinement of biotin. The four colored lines connect biotin's carboxyl group with the C-termini, indicating the different initial force-loading directions. Upon stretching, the molecular linker approximately aligns along this line. SA tightly encapsulates biotin, except for biotin's carboxyl group to which the molecular linker is covalently attached. For pulling on subunit D, which showed the strongest unbinding forces in the experiment, the initial force-loading direction points straight through the binding pocket cavity (yellow line in Fig. 2a). For the other subunits, the initial force-loading directions pierce through the binding pocket's confinement. Upon stretching, biotin will be pushed against parts of the enclosing binding pocket. We hypothesized that this levering of biotin or the adjacent molecular linker against flexible parts of SA, destabilizes the binding pocket and interferes with its structural integrity resulting in lower unbinding forces.

Binding of biotin to SA is mediated by hydrophobic interactions, a network of hydrogen bonds and a conformational change in the SA subunit:[23] A flexible peptide loop between the third and the fourth β-strand (L3/4-loop) closes over the binding pocket like a lid and buries biotin inside the pocket.[24] Calculations performed by Bansal *et al.* showed that this conformational change accounts for about 75% of the change in free energy upon biotin binding.[25] In the analysis of our SMD data, we therefore focused on this vital contribution of the lid to biotin binding. We propose that for the three weaker attachment geometries (anchoring of subunit A, B or C) the L3/4-loop is, under load, forced towards its open conformation. By analyzing SMD trajectories, we observed that the lid indeed opens up before biotin dissociation, particularly in those simulations where subunit A or C were probed. To illustrate the mechanism of force-induced lid opening, we depicted different stages of the SMD simulation (for pulling of subunit C) in Fig. 2b-f and Supplementary Fig. 8.

Beyond this phenomenological description, the wide-sampling SMD strategy allowed statistical treatment of the SMD data (Fig. 3). By plotting rupture force histograms for the SMD data (Fig. 3b), we see that the SMD results agree qualitatively with the experimental SMFS data: The force needed to unbind biotin from subunit D is the highest (510 pN), while the unbinding forces from subunit B are lower (450 pN). The unbinding from subunit A and subunit C is observed at similar forces of about 340 pN and 360 pN. Plotting a histogram

combing forces from all domains shows that subunit A and subunit C results are nearly indistinguishable (Supplementary Fig. 7).

To monitor the position of the L3/4 loop, we introduced a distance-based (Fig. 3c-d) and an angle-based metric (Supplementary Fig. 9). For the former, we measured the distance between the α-carbon of residue GLY48 (tip of the L3/4 loop) and the α-carbon of residue LEU124 (middle of β-strand β8, Supplementary Fig. 10). Tracking this metric over time for single representative trajectories (Fig. 3c), we found that for subunit A and C, the distance abruptly increases about 10 ns before the complex ruptures, which indicates that the L3/4-loop opens (Supplementary Video 1, 3). Subunit B exhibits a similar but much less pronounced behavior (Supplementary Video 2), while for subunit D, the distance is constant up to the point of rupture (Supplementary Video 4). A histogram over all 100 replicas confirms this trend (Fig. 3d): While for subunit D, the distributions at the beginning of the force loading (grey) and around the rupture (red) are almost congruent; they differ significantly for the other three subunits, particularly for subunit A and C.

To investigate how force propagates through the receptor-ligand complex, we employed a cross-correlation based network analysis.[26] From thermodynamic fluctuation theory, one can infer that paths with high correlation of motion can be isolated to describe the paths along which force propagates through the system.[26, 27] In Fig. 4, the force propagation pathways through the SA tetramer are depicted. Whereas clear differences between the four force-loading geometries are evident, one can observe that force propagation pathways for subunit B and D, are quite similar within subunit D. The network model suggests that interactions between receptor and ligand are highly correlated in multiple sites of the subunit D β-barrel, as it was previously shown for C-terminal pulling of subunit D.[9] Since for force loading of subunit A, B and C, the force has to propagate through the SA tetramer, it is in principle imaginable that not the biotin/SA interaction, but the SA tetramer structure ruptures, as suggested by Kim *et al.*[28] While we cannot rule out such a process for our AFM-based SMFS experiments, any indication for rupturing of the SA tetramer was absent in the SMD simulations.

Our study reconciles the conflictingly wide range of rupture forces reported in previous force spectroscopy studies on the biotin/SA interaction from a more complete perspective, showing that for four different well-defined tethering geometries, the experimental unbinding forces can vary fourfold. Anchoring of SA via unspecific pull-down by reactive amines or similar groups as it is done in many commercial products might result in an even wider range of

unbinding forces. Therefore, we show that the way in which SA is tethered is of critical importance for the force propagation path in the complex and thus for the mechanical stability of the biotin/SA interaction. Diversity in binding forces were shown to be caused by different force-loading geometries and the accompanying induction of conformational changes caused by pushing biotin against the flexible L3/4 peptide loop. We demonstrated that for SMD simulations it is important to consider the experimental force-loading geometry, and take explicitly into account molecules that may be interfering with the receptor-ligand interaction, such as the biotin linker molecule. In summary, our findings encourage to reconsider how SA is tethered in future force spectroscopy experiments: With site-specific anchoring and consideration of resulting force loading geometries, higher mechanical stability of the biotin/SA bond can be achieved in future investigations. Likewise, since biotin is attached to a molecular linker for most applications in bionanotechnology, our experimental and computational design follows the predominant scenario for assays using biotin/SA complexes and should be used to guide new developments whenever these complexes might be under mechanical stress.

Acknowledgements

The authors thank A. Kardinal and T. Nicolaus for laboratory support, L.F. Milles for providing the SdrG/Fgβ-system, D.A. Pippig for cloning of the SA subunits, M.C.R. Melo for help with SMD analysis scripts and J.C. Fischer for assistance in one of the control measurements. The authors acknowledge M.S. Bauer, A. Löf, C. Kluger, L.F. Milles and K.R. Erlich for helpful discussions. S.M. Sedlak thanks the Nanosystems Initiative Munich for support. The Deutsche Forschungsgemeinschaft (DFG, Sonderforschungsbereich 1032, Teilprojekt A01) provided funding for this work. Furthermore, this work was supported by the National Institutes of Health (NIH) grant P41-GM104601, "Center for Macromolecular Modeling and Bioinformatics." R.C.B. is partially supported by National Science Foundation (NSF) grant MCB-1616590, "Molecular Modeling of Bioenergetic Systems." Molecular dynamics simulations made use of GPU-accelerated nodes of Blue Waters supercomputer as part of the Petascale Computational Resource (PRAC) grant "The Computational Microscope", which is supported by the National Science Foundation (award number ACI-1713784). The state of Illinois and the National Science Foundation (awards OCI-0725070 and ACI-1238993) support Blue Waters sustained-petascale computing project.

Author contribution

S.M.S. prepared SA constructs. S.M.S., and L.C.S. performed and analyzed single-molecule force spectroscopy experiments. R.C.B. performed and analyzed molecular dynamics simulations. H.E.G. and R.C.B. supervised the study. S.M.S., and L.C.S. drafted the manuscript. All authors contributed with writing the final version of the manuscript.

Competing interests

The authors declare no competing interests.

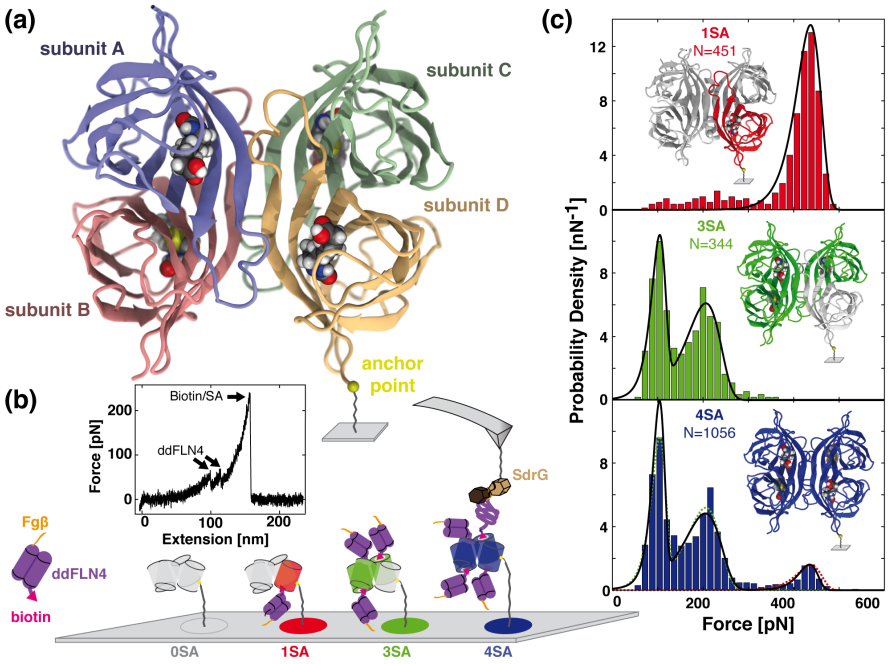


Figure 1. Force spectroscopy of the biotin/SA complex with different valencies. (a) Crystal structure of SA. SA comprises four subunits, each consisting of a β-barrel into which a biotin molecule can be bound. At the C-terminus of subunit D, a unique cysteine is used as anchor point for site-specific covalent immobilization by maleimide-polyethylene glycol (PEG) linkers onto a functionalized glass surface. (b) Combining non-functional (light grey cylinders) and functional subunits (colored cylinders) allows preparation of SA of different valencies. These different SA variants are immobilized at different areas on the surface: 0SA (grey), 1SA (red), 3SA (green), and 4SA (blue) are all examined with the same cantilever. Biotinylated ddFLN4 (purple) with an N-terminal Fgβ-peptide (orange) is added to the solution. While biotin (magenta) binds to SA on the surface, the Fgβ-peptide can bind to the SdrG-domain (brown) immobilized on the cantilever. Retracting the cantilever, ddFLN4 unfolds and biotin is pulled out of the binding pocket, while the force is recorded. A typical force-extension trace is shown in the inset. (c) After sorting the force curves for specific interactions, *i.e.* for those showing the specific unfolding pattern of ddFLN4, unbinding force histograms are plotted and fitted with Bell-Evans distributions: 1SA (red) is fitted with a single Bell-Evans-distribution. To fit 3SA (green), a double Bell-Evans distribution is needed. 4SA (blue lines) is fitted with a triple Bell-Evans. Furthermore, a combination of distributions of 1SA and 3SA can be fitted (red and green dotted lines).

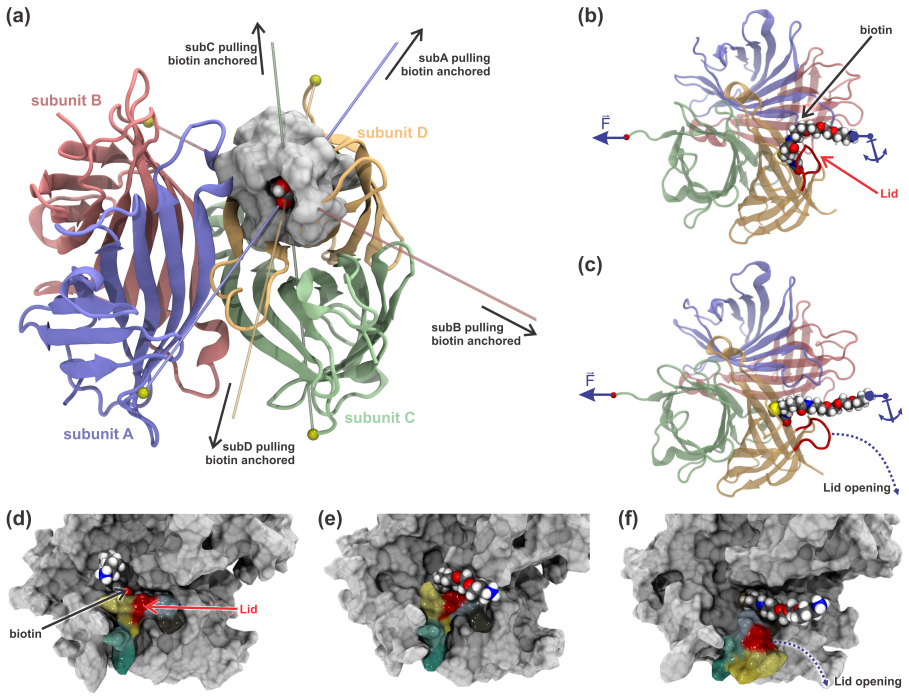


Figure 2. Direction dependent unbinding and lid opening. (a) Schematics of the force-loading geometries. To simplify MD-simulations, biotin bound in subunit D (shown with surface representation) was anchored by the end of its molecular linker, while one of four subunits A-D was pulled by its C-terminus. Colored lines indicate the four resulting force-loading directions (polymeric biotin-linker not shown). (b,c) The structure of SA stretched via its subunit C and the end of the polymeric linker of biotin bound in subunit D is shown prior to (b) and after (c) lid opening just prior to bond rupture. (d-f) Surface representation of SA shows how the stretching of biotin and its linker during subunit C pulling – from initial conformation at time 0 ns (d), to time 32 ns (e), to time 54 ns (f) – induces conformational changes in the binding pocket's lid (colored by amino acid sequence).

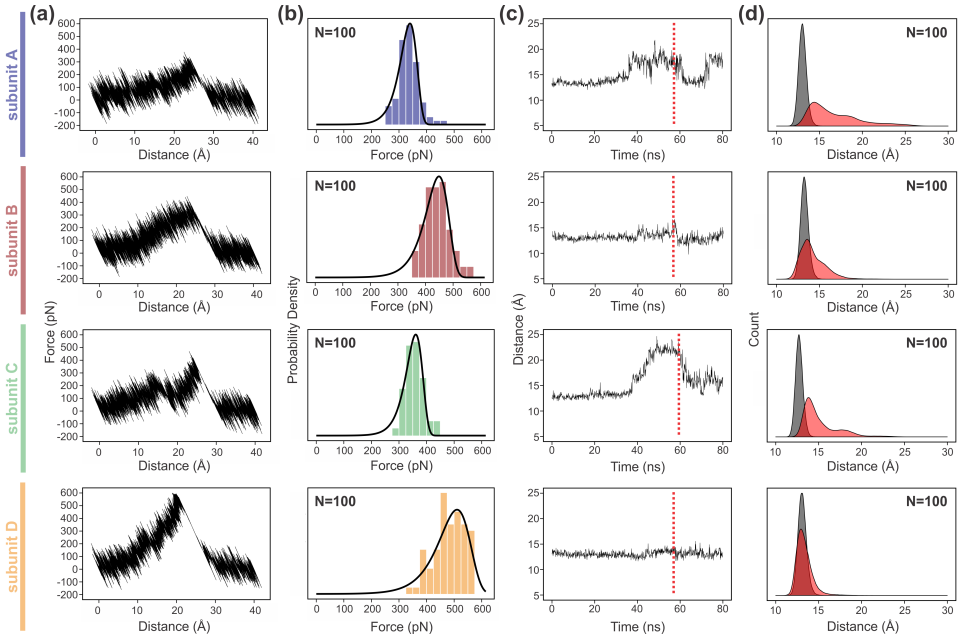


Figure 3. Results of SMD simulations. Pulling C-termini of SA subunits while holding molecular linker of biotin bound to pocket in subunit D. (a) Exemplary force extension traces for the four geometries. (b) Resulting rupture force histograms fitted with Bell-Evans distributions. (c) Exemplary plots of the distance metric for the L3/4-opening (distance between α-carbons of GLY48 and LEU124 residues) over time. The red dashed lines denote the moment at which biotin leaves the pocket. (d) Histograms of the distance metric for the L3/4-opening for the first 10 ns of the simulation (unloaded condition, grey) and for 10 ns just prior to the point of rupture (loaded condition, red).

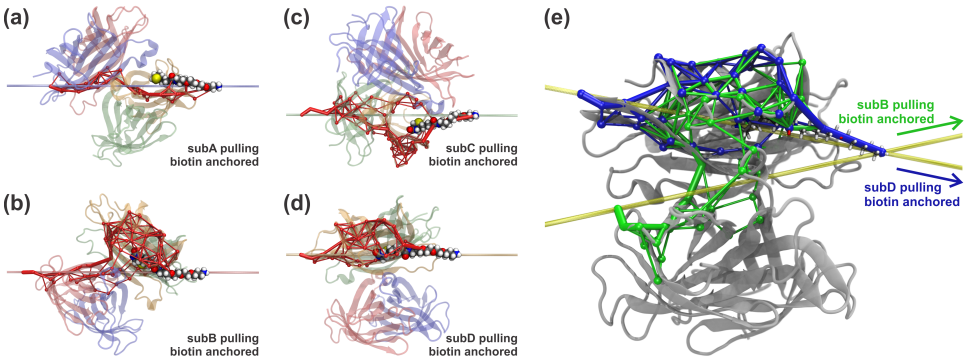


Figure 4. Force propagation pathways through the SA tetramer. (a-d) The force propagation pathway is shown for the different subunits close to the point of rupture. Force propagation pathways were obtained from cross-correlation based network analysis calculated for all 100 replicas in a force-loaded condition. α-carbon atoms serve as nodes that are connected by tubes of different diameters corresponding to how likely it is to have force transferred between them. SA is rotated to align the directions of force application horizontally. (e) Overlay of the force propagation pathway of subunit B and D. Within subunit D, the two are similar. For subunit D, a strong correlation is found between the molecular linker of biotin and the fourth β-strand of subunit D, revealing a stabilization of the biotin/SA interaction pocket.

References

1. Moy VT, Florin EL, Gaub HE. Intermolecular forces and energies between ligands and receptors. *Science* 1994, **266**(5183)**:** 257-259.
2. Lee G, Kidwell D, Colton R. Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. *Langmuir* 1994, **10**(2)**:** 354-357.
3. Chilkoti A, Boland T, Ratner BD, Stayton PS. The relationship between ligand-binding thermodynamics and protein-ligand interaction forces measured by atomic force microscopy. *Biophys J* 1995, **69**(5)**:** 2125-2130.
4. Merkel R, Nassoy P, Leung A, Ritchie K, Evans E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 1999, **397**(6714)**:** 50-53.
5. Teulon JM, Delcuze Y, Odorico M, Chen SW, Parot P, Pellequer JL. Single and multiple bonds in (strept)avidin-biotin interactions. *J Mol Recognit* 2011, **24**(3)**:** 490-502.
6. Rico F, Russek A, Gonzalez L, Grubmuller H, Scheuring S. Heterogeneous and rate-dependent streptavidin-biotin unbinding revealed by high-speed force spectroscopy and atomistic simulations. *Proc Natl Acad Sci U S A* 2019, **116**(14)**:** 6594-6601.
7. Grubmuller H, Heymann B, Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* 1996, **271**(5251)**:** 997-999.
8. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J* 1997, **72**(4)**:** 1568-1581.
9. 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* 2018.
10. 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**(12)**:** e0188722.
11. 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**(1)**:** 3-12.
12. Verdorfer T, Bernardi RC, 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, **139**(49)**:** 17841-17852.
13. Howarth M, Chinnapen DJ, Gerrow K, Dorrestein PC, Grandy MR, Kelleher NL*, et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods* 2006, **3**(4)**:** 267-273.
14. Milles LF, Schulten K, Gaub HE, Bernardi RC. Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science* 2018, **359**(6383)**:** 1527-1533.
15. 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**(1)**:** 81-85.
16. Bauer MS, Milles LF, Sedlak SM, Gaub HE. Monomeric streptavidin: a versatile regenerative handle for force spectroscopy. *bioRxiv* 2018.
17. Evans E, Ritchie K. Dynamic strength of molecular adhesion bonds. *Biophys J* 1997, **72**(4)**:** 1541-1555.
18. Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* 2001, **11**(2)**:** 224-230.
19. 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.
20. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E*, et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* 2005, **26**(16)**:** 1781-1802.
21. Zhang M, Biswas S, Deng W, Yu H. The Crystal Structure of Monovalent Streptavidin. *Sci Rep* 2016, **6:** 35915.

22. Hyre DE, Le Trong I, Merritt EA, Eccleston JF, Green NM, Stenkamp RE*, et al.* Cooperative hydrogen bond interactions in the streptavidin-biotin system. *Protein Sci* 2006, **15**(3)**:** 459-467.
23. Weber PC, Ohlendorf DH, Wendoloski JJ, Salemme FR. Structural origins of high-affinity biotin binding to streptavidin. *Science* 1989, **243**(4887)**:** 85.
24. Freitag S, Le Trong I, Klumb L, Stayton PS, Stenkamp RE. Structural studies of the streptavidin binding loop. *Protein Sci* 1997, **6**(6)**:** 1157-1166.
25. Bansal N, Zheng Z, Song LF, Pei J, Merz KM, Jr. The Role of the Active Site Flap in Streptavidin/Biotin Complex Formation. *J Am Chem Soc* 2018, **140**(16)**:** 5434-5446.
26. 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**(11)**:** 7370-7376.
27. Greene RF, Callen HB. On the formalism of thermodynamic fluctuation theory. *Physical Review* 1951, **83**(6)**:** 1231.
28. Kim M, Wang CC, Benedetti F, Rabbi M, Bennett V, Marszalek PE. Nanomechanics of streptavidin hubs for molecular materials. *Adv Mater* 2011, **23**(47)**:** 5684-5688.

Methods

Preparation of Proteins

All proteins sequences used are provided as Supplementary Notes. GFP, used for polyacrylamide electrophoresis, or ddFLN4 were cloned into pET vectors and expressed as described by Sedlak *et al.*[10] Recombinant GFP and ddFLN4 proteins contain a ybbR-tag that was employed for biotinylation using SfP-Synthase[29] as described in Erlich *et al*.[30]

For cloning and expression, we follow a protocol provided by Baumann *et al.*[31] The four different SA subunits were cloned into pET vectors. Subunits were expressed separately, then denatured, mixed and purified by Ni-IMAC.

For example, to obtain 3SA, we denatured non-functional SA subunits (with polyhistidine tag and a unique cysteine at their C-terminus) and mixed them with denatured functional subunits without tags in a 1:10 ratio. After protein refolding, we employed Ni-IMAC to select for SA with a single polyhistidine tag, *i.e.* 3SA. A complete description is given by Sedlak, Schendel *et al*.[9]

To ascertain the number of functional subunits per SA, we added biotinylated GFP to the different SA variants and performed SDS polyacrylamide gel electrophoresis. The different SA variants (0SA, 1Sa, 3SA, 4SA) were mixed with biotinylated GFP. We allowed the proteins to bind to each other (about 10 min) before adding loading buffer. Proteins were loaded them onto an Any kD Mini-Protean TGX Stain-Free Protein Gel (Bio-Rad, Hercules, USA).

Surface Preparation

Heterobifunctional polyethylene glycol of 5,000 Da molecular weight was dissolved to 25 mM in a 50 mM HEPES buffer at pH 7.5 and added onto an amino-silanized glass slide.[32] During 30 min of incubation, the NHS group on one end of the PEG-linker formed a stable amide bond with the amines on the glass slide. After washing off unbound PEG using ultrapure water, a silicon mask was placed on the surface and at different spots 10 µl of the reduced SAs dissolved in coupling buffer were added onto the surface. The SA's unique cysteines reacted with the maleimide group on the other end of the polyethylene glycol to form a stable thioetherbond. A graphical illustration of the process is given in the supplementary information of Sedlak, Schendel *et al*.[9]

Cantilever Preparation

Bifunctional polyethylene glycol of 5,000 Da having an N-hydroxysuccinimide group at one end and a maleimide group on the other (NHS-PEG5000-MAL, Rapp Polymere, Tübingen, Germany) was dissolved in 50 mM HEPES at pH 7.5 and immediately used to incubate amino-silanized BioLever mini (Olympus Corporation, Tokyo, Japan).[32] After one hour, the cantilevers were thoroughly washed in ultrapure water and then placed in 25 µl droplets of Coenzyme A (CoA) dissolved in coupling buffer (50 mM NaCl, 50 mM $NaHPO_4$, 10 mM EDTA, pH 7.2). After one hour, the cantilevers were thoroughly washed in ultrapure water and then placed in 25 µl droplets of the Sfp reaction mix (10 µl 10x Sfp buffer (10 mM $MgCl_2$, 50 mM HEPES, pH 7.5), 5 µl 100 µM Sfp Synthase, 40 µl 32.5 µM SdrG-ybbR construct,[14] 45 µl MiliQ H_2O). After at least one-hour incubation time, the cantilevers were thoroughly washed in PBS and stored in PBS. A graphical illustration of the process is given in the supplementary information of Sedlak, Schendel *et al.*[9]

For covalent attachment of the biotinylated ddFLN4-domain, the biotinylated ddFLN4-construct with the C-terminal cysteine was used and coupled to the maleimide instead of the CoA. After at least one-hour incubation time, the cantilevers were thoroughly washed in PBS and stored in PBS.

AFM-based Single-Molecule Force Spectroscopy Experiments

The AFM-based single-molecule force spectroscopy measurements were performed with a custom-built AFM controlled by an MFP-3D controller (Asylum Research, Santa Barbara, USA)

and a self-written routine programmed in Igor Pro 6 (WaveMetrics, Oregon, USA). The cantilevers were approached to the surface with 3000 nm/s and after short contact (indentation of 100 pN) retracted with a constant velocity of 800 nm/s. Read-out of the distance and cantilever deflection was performed at 12,000 Hz. The cantilever was retracted 350 nm. After each approach-retraction cycle the surface was moved 100 nm in lateral direction to expose a fresh surface area to the cantilever tip. All measurements were performed in PBS, pH 7.4 in ambient conditions. Cantilevers were calibrated following the thermal noise method as described by te Riet *et al*.[33]
For measurements with a second receptor ligand system on the cantilever tip, we first performed about 1,000 approach-retraction cycles to ensure the absence of unspecific interaction between the SA on the surface and the SdrG on the cantilever tip. We then placed the mounted AFM cantilever tip in PBS containing the biotinylated Fgβ-ddFLN4 construct at a concentration in the low nanomolar range for two minutes. By this, some ddFLN4 gets adsorbed to the cantilever tip. We then transferred the AFM head back onto the sample surface and continued with the approach-retraction cycles, now measuring specific interactions. An alternative approach that also worked is to directly add the diluted biotinylated Fgβ-ddFLN4 construct to the measurement buffer.
For measurements with several surface areas, where different proteins are immobilized, the cantilever tip was retracted a few µm from the surface after 250-2,000 approach-retraction cycles. Then, the surface was moved a few millimeters in lateral direction so that the next surface area could be probed. The cantilever was approach automatically and the probing of the surface continued.

AFM-based Single-Molecule Force Spectroscopy Data Analysis
Using the cantilever spring constant, the optical lever sensitivity as well as the z-piezo sensitivity the deflection voltage and the z-piezo voltage are translated into force and distance, respectively. Then, the cantilever bending correction is preformed and the value for zero-force and zero-distance are determined for each force-extension trace. After denoising, each force-extension trace is translated into contour length space. Detecting force peaks, force-extension traces are sorted to identify those that show the correct increase in contour-length corresponding to the distinct two-step unfolding of the ddFLN4 fingerprint domain[34]. Rupture and unfolding forces for each surface area are analyzed separately and plotted as histograms.

Molecular Dynamics Simulations
Employing advanced run options of VMD's[35] QwikMD[19] plugin, our in silico approach followed established protocols that were previously employed to investigate mechanical properties of SA[9], filamins[36, 37], cellulosomes[12], and adhesins[14].

System setup
The structure of a monovalent *Streptomyces avidinii* SA (mSA) had been solved by means of X-ray crystallography at 1.65 Å resolution and was available at the protein data bank (PDB: 5TO2)[21]. As this crystallographic structure does not contain a biotin bound to the binding pocket, the structure of the tetravalent *S. avidinii* SA bound to biotin (PDB: 1MK5)[22], solved at 1.4 Å resolution, was used to place the biotin on to its binding site at chain D of the mSA. The polyethylene glycol (PEG_3) molecular linker employed in the experiments was designed with VMD's Molefacture[35] plugin. The alignment and placing of the biotin with linker into the monovalent structure was performed using VMD[35] based on the alignment of the aforementioned crystal structures. Employing the quantum mechanics/molecular mechanics (QM/MM) tools of QwikMD[19], we performed a short 10 ps long hybrid QM/MM MD simulation with NAMD[20, 38] and MOPAC[39], using a 0.5 fs integration time step. The classical CHARMM36 force field[40] was employed to represent the SA atoms, while the biotin and its linker were treated with quantum mechanics at PM7 level[41]. This

QM/MM simulation was performed without presence of solvent molecules and keeping the SA and the biotin non-hydrogen atoms with position restraints, allowing only for the linker to search for a plausible conformation. The biotin with its linker was than parameterized for classical molecular dynamics (MD) simulations using CHARMM General Force Field (CGenFF)[42]. Employing advanced run options of QwikMD[19], the structure resulting from the QM/MM simulation was solvated and the net charge of the system was neutralized in a 0.15 mol/l sodium chloride solution. In total, approximately 275,000 atoms were simulated in each of the classical MD simulation. The CHARMM36 force field[40], along with the TIP3 water model[43] was used to describe all systems.

Equilibrium molecular dynamics simulations
All classical MD simulations were performed in GPU-accelerated XK nodes of the NCSA/Blue Waters supercomputer, employing the NAMD molecular dynamics package[20]. All simulations were performed assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for temperature and pressure coupling, the latter kept at 1 bar. A distance cut-off of 11.0 Å was applied to short-range non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method[44]. The equations of motion were integrated using the r-RESPA multiple time step scheme[20] to update the Lennard-Jones interactions every step and electrostatic interactions every two steps. The time step of integration was chosen to be 2 fs for all simulations performed. Before the MD simulations, an energy minimization was performed for 5,000 steps. An MD simulation with position restraints in the protein backbone atoms and biotin and linker non-hydrogen atoms was performed for 10 ns. To allow for a total relaxation of the system and to make sure biotin and its linker were stable in the SA pocket, a 100 ns simulation in equilibrium, where no external forces were applied, was performed. The MD protocol served to pre-equilibrate the system before the steered molecular dynamics (SMD) simulations were performed.

Steered molecular dynamics simulations
With structures properly equilibrated and checked, SMD simulations[8] were performed using a constant velocity stretching (SMD-CV protocol) at 0.5 Å/ns. The SMD procedure is equivalent to attaching one end of a harmonic spring to the end of a molecule and pulling on the end of the other molecule with another spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the z-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the z-axis. Owing to the flexibility of the experimentally employed linkers connecting the domains of interest and the fingerprint domains, this approach reproduces the experimental protocol. Simulations were performed restraining the position of the terminal nitrogen of the biotin linker while pulling the C_α of each subunit's C-terminal amino acid residue. For all four configurations, many simulation replicas were performed in a wide-sampling approach. For each subunit pulling, 100 replicas were performed, with each of the simulations accounting for 80 ns total simulation time. In total, 32 μs of production SMD were performed.

Simulation Data Analysis
Simulation force-time traces were analyzed analogously to experimental data. For each simulation, the rupture force was determined as the highest force of a trace and the force-loading rate was determined as a linear fit to the force versus time traces immediately before rupture. Analyses of force traces and MD trajectories, except for the force propagation analyses, were carried out employing python scripts taking advantage of Jupyter Notebooks[45]. Particularly, VMD[35], MDAnalysis[46], and PyContact[47] were employed for trajectory analysis together with in-house scripting wrappers, which collected information from all simulation replicas. Force propagation

analyses were performed using dynamical network analysis, which is implemented in VMD's Network View plugin[48]. A network was defined as a set of nodes, all α-carbons plus three atoms of the biotin and its linker, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and two monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (non-hydrogen) from the two monomers are within 4.5 Å of each other for at least 75% of the frames analyzed. Filtering this network, one can investigate allosteric signaling.[38, 49] Allostery can be understood in terms of pathways of residues that efficiently transmit energy, here in the form of mechanical stress, between different binding sites.[26] The dynamical networks were constructed from 10 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as $w_{ij}=-\log(|C_{ij}|)$, where C is the correlation matrix calculated with Carma[50]. Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was -log(0.5)=0.69. As previously demonstrated by our group,[26] Pearson correlation is ideal for force propagation calculation.

References

29. 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**(44)**:** 15815-15820.
30. Erlich KR, Sedlak SM, Jobst MA, Milles LF, Gaub HE. DNA-free directed assembly in single-molecule cut-and-paste. *Nanoscale* 2019, **11**(2)**:** 407-411.
31. Baumann F, Bauer MS, Milles LF, Alexandrovich A, Gaub HE, Pippig DA. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat Nanotechnol* 2016, **11**(1)**:** 89-94.
32. 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**(6)**:** 975-985.
33. te Riet J, Katan AJ, Rankl C, Stahl SW, van Buul AM, Phang IY*, et al.* Interlaboratory round robin on cantilever calibration for AFM force spectroscopy. *Ultramicroscopy* 2011, **111**(12)**:** 1659-1669.
34. 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* 2016.
35. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996, **14**(1)**:** 33-38, 27-38.
36. Haataja TJK, Bernardi RC, Lecointe S, Capoulade R, Merot J, Pentikainen U. Non-syndromic Mitral Valve Dysplasia Mutation Changes the Force Resilience and Interaction of Human Filamin A. *Structure* 2019, **27**(1)**:** 102-112 e104.
37. Seppala J, Bernardi RC, Haataja TJK, Hellman M, Pentikainen OT, Schulten K*, et al.* Skeletal Dysplasia Mutations Effect on Human Filamins' Structure and Mechanosensing. *Sci Rep* 2017, **7**(1)**:** 4218.
38. 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**(5)**:** 351-354.
39. Stewart JJ. MOPAC: a semiempirical molecular orbital program. *J Comput Aided Mol Des* 1990, **4**(1)**:** 1-105.
40. Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* 2013, **34**(25)**:** 2135-2145.
41. Stewart JJ. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J Mol Model* 2013, **19**(1)**:** 1-32.
42. Vanommeslaeghe K, MacKerell AD, Jr. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J Chem Inf Model* 2012, **52**(12)**:** 3144-3154.

43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 1983, **79**(2)**:** 926-935.
44. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* 1993, **98**(12)**:** 10089-10092.
45. Kluyver T, Benjamin R-K, Fernando P, Brian G, Matthias B, Jonathan F*, et al.* Jupyter Notebooks—a publishing format for reproducible computational workflows. Positioning and Power in Academic Publishing: Players, Agents and Agendas, 20th International Conference on Electronic Publishing; 2016 2016; Göttingen, Germany; 2016.
46. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* 2011, **32**(10)**:** 2319-2327.
47. 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**(3)**:** 577-583.
48. Eargle J, Luthey-Schulten Z. NetworkView: 3D display and analysis of protein.RNA interaction networks. *Bioinformatics* 2012, **28**(22)**:** 3000-3001.
49. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A* 2009, **106**(16)**:** 6620-6625.
50. Glykos NM. Software news and updates. Carma: a molecular dynamics analysis program. *J Comput Chem* 2006, **27**(14)**:** 1765-1768.

3.4 Manuscript M3: The Tetravalency of Streptavidin Results in Distinct Lifetimes of Biotin Unbinding Under Constant Load

In Manuscript M3, the lifetime of the SA/biotin interaction under constant force and its dependence on the tethering geometry are investigated. 1SA, 3SA or 4SA are C-terminally tethered to superparamagnetic beads, which are then used in magnetic tweezers experiments. Using ddFLN4 as a fingerprint domain to identify beads that are tethered by a single SA/biotin interaction, the survival time of single SA/biotin bonds at 65 pN is recorded. In the molecular chain between the surface of the flow cell and the superparamagnetic bead, the SA/biotin interaction is the only non-covalent bond. This assures that, when the bead is detaching, it is the SA/biotin bond that breaks.

AFM imaging and ITC measurements are employed to complement the magnetic tweezers experiments: To verify the different valencies of the three SA variants, AFM imaging of biotinylated doublestranded 250 bp long DNA bound to SA is conducted. Obviously, 4SA binds four biotinylated DNA strands, while 3SA cannot bind more than three and 1SA only binds a single biotinylated DNA strand. ITC experiments are performed to ascertain that – in the absence of force – the binding behavior for all SA variants is the same. All three SA variants exhibited the same binding enthalpy per mole of biotin added.

For the measurements of the survival time of the SA/biotin interaction under constant force, it is observed that a single exponential fits the survival fraction of 1SA, while for 3SA and 4SA, this is not possible. At least two off-rates have to be used to reasonably fit these data, *i.e.* a double, triple or fourfold exponential fit is needed. These results demonstrate that the force-loading geometry is also for constant force measurements an important factor. For 1SA, there is only a single force-loading geometry, as biotin can only bind to one subunit. For 3SA (4SA), three (four) different force-loading geometries exist.

This interpretation is further substantiated by the fact that a combination of the lifetimes determined for 3SA and 1SA fits the 4SA data. The tethering by 1SA is especially long-lived, with a lifetime of about 8 h at 65 pN, while for 3SA nearly all beads rupture within the first hour. The 4SA data are a combination of both: About 75% rupture within the first hour; the remaining 25% are long-lived, comparable with 1SA. These findings nicely agree with the result for the AFM-based SMFS experiments presented in Manuscript M2 (Section 3.3). The underlying molecular mechanism is most likely the same.

The tetravalency of streptavidin results in distinct lifetimes of biotin unbinding under constant load

by

Achim Löf*, Steffen M. Sedlak*, Sophia Gruber, Martin Benoit, Jan Lipfert, and Hermann E. Gaub

in preparation

*These authors contributed equally to this work. AL: AFM imaging, functionalization of magnetic beads, MT experiments, data analysis, writing of manuscript. SMS: Protein design and preparation, preparation of biotinylated DNA, functionalization of magnetic beads, ITC measurements, data analysis, writing of manuscript

The tetravalency of streptavidin results in distinct lifetimes of biotin unbinding under constant load

Achim Löf[1,¶], Steffen M. Sedlak[1,¶], Sophia Gruber[1], Martin Benoit[1], Jan Lipfert[1], and Hermann E. Gaub[1,*]

[1]Department of Physics and Center for NanoScience, LMU Munich, Germany

[¶] These authors contributed equally to this work.

*** To whom correspondence should be addressed:**
Hermann E. Gaub, Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 Munich, Germany,
Tel: +49 89 2180 3172, e-mail: gaub@lmu.de

Abstract

The binding of the small molecule biotin to the homotetrameric protein streptavidin (SA) is an easy-to-handle and widely applicable tool for immobilization, labeling, and detection of molecules in nanotechnology, biotechnology, and medicine. Biotin's interaction with SA is specific, long-lived, and robust under harsh conditions. In single-molecule force spectroscopy (SMFS) experiments, SA is employed as a molecular handle to study the behavior of biotinylated molecules or molecular complexes under force. For SMFS measurements, it is critical to have a well-characterized and ideally monodisperse attachment, since otherwise interpretation of lifetimes in constant force or rupture forces in constant speed measurements is complicated or impossible. Recent data suggest that SA/biotin bonds under load show strong directional dependence and multi-exponential lifetimes. Here, we use our recently established constant force magnetic tweezers (MT) assay to comprehensively characterize the lifetimes of single biotin/SA interactions under constant force, using magnetic beads functionalized with SA variants of different valencies. For site-specific tethering of a certain SA subunit, we observe significant differences in lifetimes for the biotin/SA bond under force, depending on which subunit biotin is bound to. In particular, we identified an especially long-lived tethering geometry with a lifetime of approximately 8 h at 65 pN. We anticipate that our results will influence the way SA is applied in force assays: Control of valency and tethering geometry will provide monodisperse data that are easier to analyze. In addition, using the right tethering geometry, longer lifetimes under force and higher mechanical stability of the biotin/SA bond can be achieved.

Introduction

The non-covalent, high-affinity binding of the small molecule biotin to streptavidin (SA) is ubiquitously used in a variety of biological, chemical, biophysical and pharmaceutical applications (1, 2). Biotin can readily be covalently attached to nucleic acids (3, 4), proteins (5, 6) or linker molecules (7). SA is stable over a wide range of conditions and easy to handle (1). Owing to the specificity of the binding, as well as the robustness of the complex, the interaction has in particular become a popular tool in the context of single-molecule force spectroscopy (SMFS). It serves as a molecular handle to anchor molecules of interest and to apply force to them (8-12). The exceptionally long lifetime of the biotin/SA complex under exposure to external forces has even been demonstrated to facilitate constant-force SMFS experiments lasting for several days in a recent study employing magnetic tweezers (MT) (13).

SA's tetravalency yet poses a caveat in such applications, since it is ambiguous to which subunit biotin binds. This results in four different force-loading geometries, depending on which SA subunit biotin is bound to (Fig. 1) (14). In addition, commercially available SA-

coated magnetic beads appear to feature various attachment geometries (13). In atomic force microscope (AFM)-based constant speed SMFS experiments, we recently showed that the force needed to unbind biotin from the SA binding pocket is heavily dependent on the force-loading direction (15): Tethering SA by a single residue and pulling biotin out of one of the binding pockets results in different force-loading geometries. For some of these pulling directions, the SA subunit is deformed such that the energy barrier of the binding is decreased. This causes lower biotin unbinding forces. The influence of the tethering geometry of SA on the stability of the SA/biotin interaction under constant forces has not yet been investigated.

In this study, we directly measure the lifetime of the biotin/SA interaction under constant force in an MT assay. We employ variants of SA that have different valencies and a unique tethering point to restrict and control the number of possible force-loading geometries for the SMFS measurement. AFM imaging provides direct evidence for the valencies of different SA variants. We demonstrate that the lifetime of the biotin/SA bond under constant load strongly depends on the force-loading geometry. The different stabilities give rise to multi-exponential lifetime distributions. By using a single well-defined geometry, a single-exponential lifetime is recorded. This finding is of particular importance for force spectroscopy measurements. In addition, it might affect other biotin/SA-based assays involving force, such as flow experiments.

Materials and Methods

Expression of SA constructs
Functional and non-functional SA subunits were expressed separately and then mixed to obtain SA of different valencies: Tetravalent SA (4SA), trivalent SA (3SA), monovalent SA (1SA) and non-functional SA (0SA). 4SA and 1SA contained one functional subunit with a unique cysteine for surface immobilization at its C-terminus. 3SA and 0SA contained one non-functional subunit with a unique cysteine for surface immobilization at its C-terminus. To select for the correct SA, this subunit further contained a polyhistidine tag for purification by nickel-immobilized metal ion affinity chromatography.
All SA subunits were cloned into pET vectors (Merck Millipore, Burlington, USA). SA plasmids were transferred to *E. coli* BL21(DE3)-CodonPlus cells (Agilent Technologies Inc., Santa Clara, USA) and expressed in SB medium. 15 ml of preculture, which was grown overnight at 37°C, was used to inoculate 500 ml of SB medium containing the appropriate antibiotic. Cells were grown at 37°C. At an OD_{600} of 0.8, expression was induced with 0.02 M IPTG and the temperature was reduced to 18°C for 16 h. The cultures were spun down so that bacterial pellets formed, which were then stored at -80°C.

Purification of SA constructs
All steps except for chromatography were performed on ice or at 4°C, respectively. SA cell pellets were thawed, suspended in 5 l/g Bacterial Protein Extraction Reagent (B-PER, Thermo Fisher Scientific, Waltham, MA) and incubated with 1 µg lysozyme and 0.05 µg DNAse I per gram bacterial pellet on a rolling shaker for 20 minutes. To ensure full break-up, cells were subsequently sonicated. The lysed cells were then centrifuged. The supernatants were discarded. The pellets were resupended in lysis buffer (PBS, 1 mM DTT, 0.1% Triton X-100). Sonification, centrifugation and resuspension were repeated four to five times until the supernatant was a clear liquid. The pellets were then resuspended in denaturing buffer (PBS, 6 M guanidine hydrochloride), sonicated and centrifuged. This time, supernatants contained the protein. Supernatants were filtered through a sterile 0.22 µm filter. Then, the absorption at 280 nm was determined. Denatured subunits were mixed in a 1:10 ratio (subunits with and without polyhistidine tag). The mixture was then slowly diluted into 500 ml PBS and stirred over night using a magnetic stirrer. This solution containing the refolded and reassembled SA was loaded onto a Ni-NTA column (HisTrap FF, GE Healthcare, Chicago, USA). We employed a gradient elution to elute SA from the column and selected for those SA containing a single polyhistidine tag (elution fractions were checked by SDS-PAGE). Elution fractions containing the protein were dialyzed against PBS and then stored in PBS at 4°C.

Expression and purification of ddFLN4
The recombinant ddFLN4 construct expressed in *E.coli* (with the internal cysteine at position 18 mutated to serine) was a kind gift from Lukas F. Milles. At its C-terminus, the ddFLN4 construct possesses a polyhistidine-tag for purification and a ybbR-tag. At its N-terminus, the construct possesses a short linker sequence (MGTGSGSGSGSAGTGSG) with the terminal methionine being followed by a single glycine. Due to efficient cleavage of the methionine by *E.coli* methionine aminopeptidases, the glycine is available for Sortase-catalyzed ligation.
The ddFLN4 gene was synthesized codon-optimized for expression in *E.coli* as a linear DNA fragment (GeneArt – ThermoFisher Scientific, Regensburg, Germany), and inserted into pET28a vectors via Gibson assembly (New England Biolabs, Frankfurt, Germany). Protein expression in *E.coli* NiCo21 (DE3) (New England Biolabs) and purification via the polyhistidine-tag were carried out as previously described in detail(16).

Preparation of biotinylated DNA
To prepare biotinylated DNA, we performed a polymerase chain reaction using a biotinylated and a normal DNA primer, designed such that a 250 bp DNA strand was amplified from the DNA template. For purification of the PCR product, we performed size-exclusion chromatography.

AFM imaging

SA constructs were reduced using 50 mM of dithiothreitol and mixed with biotinylated 250 bp double-stranded DNA in PBS buffer, with DNA being in large excess to ensure that SA molecules with the maximum possible number of bound DNA strands can be observed. A 1:10 SA:DNA stoichiometry was chosen for 4SA and 3SA, and a 1:4 stoichiometry for 1SA and 0SA, with a final DNA concentration of approximately 4 nM.

Preparation of poly-l-lysine (PLL) coated mica substrates for AFM imaging was performed analogously to a recently described protocol[10]. After at least 1 h of incubation, 20 µl of the SA–DNA mix were incubated on a PLL-coated substrate for 30 s, which was subsequently rinsed with water and finally dried in a gentle stream of nitrogen. The positively charged PLL allows for stable attachment of negatively charged DNA and of DNA-streptavidin complexes. Free streptavidin without bound DNA strands, however, does not stably attach to the substrate.

AFM images of 1 µm x 1 µm or 2 µm x 2 µm and 1024 x 1024 pixels were recorded in tapping mode in air, using an MFP-3D AFM (Asylum Research, Santa Barbara, CA) and cantilevers with silicon tips (AC160TS, Olympus, Japan), possessing a nominal spring constant of 26 N/m and a resonance frequency of approximately 300 kHz. Raw image data were processed using SPIP software (v6.5.1; Image Metrology, Denmark). Image processing involved plane correction (third order polynomial plane-fitting), line-wise flattening (according to the histogram alignment routine), and Gaussian smoothing (for zoom-ins only).

Isothermal titration calorimetry

ITC was performed at 25°C on a MicroCal iTC200 (Malvern, Worcestershire, UK). Biotin was dissolved in PBS to obtain a stock solution. SA was dissolved in exactly the same buffer, using Zeba Spin Desalting Columns with a molecular weight cut-off of 40 kDa (Thermo Fisher Scientific, Waltham, USA) for buffer exchange. Concentration of SA was determined using a spectrophotometer (NanoDrop 1000, Thermo Fisher Scientific, Waltham, USA), the absorption at 280 nm, and a molar attenuation coefficient of ε_{280}=167,760 $M^{-1}cm^{-1}$. SA was filled into the measurement cell and biotin was titrated in. A 10-fold concentration of biotin was used for 1SA, a 30-fold excess for 3SA, and a 40-fold excess for 4SA, as the ratio of the measurement cell volume to the total titrant volume is five to one.

Functionalization of magnetic beads with SA constructs

5 µM of 1SA, 3SA, or 4SA were supplemented with 5 mM Bond-Breaker TCEP Solution (Thermo Fisher Scientific). After half an hour, the mixture was purified using Zeba Spin Desalting Columns (Thermo Fisher Scientific) with a molecular weight cut-off of 40 kDa equilibrated with coupling buffer (50 mM NaCl, 50 mM $NaHPO_4$, 10 mM EDTA, pH 7.2) according to the manufacturer's instruction.

Bifunctional polyethylene glycol of 5,000 Da having an N-hydroxysuccinimide group at one end and a maleimide group at the other (NHS-PEG5000-MAL, Rapp Polymere, Tübingen, Germany) was dissolved in 50 mM HEPES, pH 7.5, to a final concentration of 25 mM and immediately used to incubate superparamagnetic beads with amine groups (Dynabeads M-270 Amine, Invitrogen/Thermo Fisher). After 45 min, beads were washed extensively with DMSO and water. Beads were then incubated with the respective SA construct in coupling buffer for 90 min and extensively washed with measurement buffer (20 mM HEPES, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1% (v/v) Tween-20, pH 7.4).

Magnetic tweezers setup

MT experiments were performed on a previously described custom MT setup (11, 13). The setup employs a pair of permanent magnets (5×5×5 mm^3 each; W-05-N50-G, Supermagnete, Switzerland) in vertical configuration (17). The distance between magnets and flow cell (and, thus, the force) is controlled by a DC-motor (M-126.PD2; PI Physikinstrumente, Germany). An LED (69647, Lumitronix LED Technik GmbH, Germany) is used for illumination. A 40x oil immersion objective (UPLFLN 40x, Olympus, Japan) and a CMOS sensor camera with 4096×3072 pixels (12M Falcon2, Teledyne Dalsa, Canada) allow to image a large field of view of approximately 440×330 μm^2 at a frame rate of 58 Hz. Images

are transferred to a frame grabber (PCIe 1433; National Instruments, Austin, TX) and analyzed with open-source tracking software (18) The bead tracking accuracy of the setup is ≈0.6 nm in (*x, y*) and ≈1.5 nm in *z* direction. For creating the look-up table required for tracking the bead positions in *z*, the objective is mounted on a piezo stage (Pifoc P-726.1CD, PI Physikinstrumente). Force calibration was conducted as described by te Velthuis *et al.* (19) based on the fluctuations of long DNA tethers. Importantly, for the small extension changes on the length scales of our protein tethers, the force stays virtually constant, with the relative change in force due to tether stretching or protein unfolding being $< 10^{-4}$ (13). Force deviations due to magnetic field inhomogeneities across the full range of the field of view are < 3% (13). The largest source of force uncertainty is the bead-to-bead variation, which is on the order of ≤ 10% for the beads used in this study (13, 17, 20, 21).

Magnetic Tweezers experiments

Preparation of flow cells was performed as recently described (13). In brief, aminosilanized glass slides were functionalized with elastin-like polypeptide (ELP) linkers (22), possessing a single cysteine at their N terminus as well as a C-terminal Sortase motif, via a small-molecule crosslinker with a thiol-reactive maleimide group (sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate; Sulfo-SMCC, Thermo Fisher Scientific). Flow cells were then assembled from an ELP-functionalized slide as bottom and a non-functionalized glass slide with two small holes for inlet and outlet as top, with a layer of cut-out parafilm (Pechiney Plastic Packaging Inc., Chicago, IL) in between to form a channel. Flow cells were incubated with 1% casein solution (Sigma-Aldrich) for 1 h and flushed with 1 ml (approximately 20 flow cell volumes) of buffer (20 mM HEPES, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.4).

CoA-biotin (New England Biolabs) was coupled to the ybbR-tag of the ddFLN4 construct in a bulk reaction in the presence of 5 µM sfp phosphopantetheinyl transferase and 10 mM $MgCl_2$ at 37°C for 60 min. Afterwards, ddFLN4 was diluted to a final concentration of approximately 20 nM in 20 mM HEPES, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.4, and incubated in the flow cell in the presence of 2 µM Sortase A for 30 min. Subsequently, the flow cell was flushed with 1 ml of buffer. Finally, beads functionalized with the respective SA construct were incubated in the flow cell for 60 s, and unbound beads were flushed out with 2 ml of measurement buffer.

At the beginning of each measurement, the tethered beads were subjected to two 5-min intervals of a constant force of 25 pN to allow for identification of specific, single-tethered beads by the characteristic three-state unfolding pattern of ddFLN4 (13, 23). Importantly, essentially no specific beads ruptured during this phase of the measurement. After 30 s at a low resting force of 0.5 pN, beads were subjected to a constant force of 65 pN for either 15 h (4SA and 1SA) or 5 h (3SA), and the time until bead rupture was monitored.

All measurements were performed at room temperature (≈22°C).

Analysis of Magnetic Tweezers measurements

Lifetimes were determined from the survival fraction *vs.* time data based on > 50 bead rupture events for each SA construct. In the case of 1SA-functionalized beads, the data were well described by a single-exponential decay, and the corresponding lifetime was determined by a linear fit to all data points in logarithmic representation. Data obtained for beads functionalized with 3SA or 4SA were piecewise fitted by linear regression as described in the main text.

Results and Discussion

To facilitate systematic investigation of the influence of the pulling geometry on the stability of the biotin/SA complex, we prepared tetra-, tri-, and monovalent variants of SA. These comprise four, three, and one functional subunit(s), while the remaining subunits are incapable of biotin binding (4SA, 3SA, and 1SA; **Fig. 2A**) due to three mutations located around the binding pocket (N23A, S27D, S45A) (24). In addition, a variant consisting of four non-functional subunits (0SA) was prepared. All variants possess a single cysteine residue at the C-terminus of their subunit D, allowing for site-specific immobilization (8, 14) [cite Sedlak2019]. For 3SA and 0SA, subunit D is non-functional, whereas for 1SA and 4SA, it is functional (**Fig. 2A;** for details on protein engineering see **Methods**).

To verify the valency of the different variants, we incubated them with biotinylated, short (250 bp) double-stranded DNA and directly visualized the resulting SA/biotinylated DNA complexes by AFM imaging (**Fig. 2B** and **Supplementary Figs. S1-S4**). An excess of biotinylated DNA over SA (approximately ten-fold for 4SA and 3SA, and four-fold for 1SA and 0SA) was used to ensure that SA molecules with DNA strands bound to all functional subunits could be observed. Indeed, a maximum of four, three, and one bound biotinylated DNA strand(s) was observed for 4SA, 3SA and 1SA, respectively, confirming the expected valencies. In the case of 0SA, no SA/biotinylated DNA complexes were observed.

Next, we performed isothermal titration calorimetry (ITC) measurements to compare the binding behavior of biotin to the different constructs in the absence of force (**Fig. 2C**). In principle, ITC allows to determine the stoichiometry, the affinity and the binding enthalpy. The results for the binding stoichiometry agree well with the AFM imaging results: For 1SA, we obtain a value of 1.0±0.2, for 3SA 3.3±0.5 and for 4SA 3.8±0.7. The uncertainty of the value increases with the number of available binding pockets, because deviations in protein concentration then have a larger impact. The largest contribution to the measurement errors results from the uncertainties in concentration. The measurement errors given here are minimum and maximum values, assuming an uncertainty of 10% in protein concentration and of 5% in biotin concentration. Due to limitations of our instrument and the very high affinity of biotin to streptavidin, the binding constant could not be obtained. We can only provide an upper limit of K_D=1 nM. To allow for good comparability, we used the same biotin stock solution for all measurements. Per binding site, we obtained binding enthalpies of -(25.0±1.3) kcal/mol for 1SA, -(25.6±1.4) kcal/mol for 3SA and -(26.1±1.3) kcal/mol for 4SA. These results agree well with enthalpies measured in previous studies (14, 25). Within experimental errors, the binding enthalpies for all SA variants are the same, further suggesting that in the absence of force all subunits are equivalent with regard to biotin binding and that no effects of binding geometries come into play.

To directly measure the lifetime of the SA/biotin interaction under constant force, and in particular to investigate the influence of different force-loading geometries, we performed MT measurements using the different SA variants (**Fig. 3**). In MT, the system of interest is tethered between the bottom surface of a flow cell and a superparamagnetic bead (**Fig. 3A**). By applying a magnetic field, generated by permanent magnets, a constant force is exerted on the bead and thereby also on the tether. By tracking the 3D position of the bead, the extension of the tether can be determined with nanometer resolution. Importantly, with our MT setup we can track approximately 100 beads in parallel, allowing for generating reliable good statistics in a short amount of time (13). In addition, MT provide excellent force and drift stability, facilitating long measurement durations (13).

For the measurement, the small protein domain ddFLN4 (4^{th} F-actin crosslinker filamin rod domain of *Dictyostelium discoideum* (23)) was biotinylated and covalently coupled to the bottom surface of a flow cell by an elastin-like polypeptide linker (22). The different SA variants (4SA, 3SA, or 1SA) were site-specifically and covalently immobilized on magnetic beads *via* polyethylene glycol (PEG) linkers (by reacting the C-terminal cysteine of subunit D with a thiol-reactive maleimide group on the PEG linker; **Fig. 3A**). Upon flushing the SA-functionalized beads into the flow cell, one of the functional subunits of the respective SA construct bound to the biotinylated ddFLN4, thereby tethering the magnetic bead to the surface. Upon force application, the molecular linkers get stretched and the ddFLN4 unfolds in a characteristic three-state manner (13, 26). This distinct two-step unfolding pattern serves as fingerprint to identify specific, single-tethered beads, *i.e*. beads that are bound to the surface *via* a single biotin/SA interaction. To measure the lifetime of this interaction, beads were subjected to a constant force of 65 pN and the time until bead rupture was recorded. The rupture event is attributed to the unbinding of biotin from SA, as this is the only non-covalent bond within the tether connecting bead and surface.

In the case of 1SA, only subunit D (attached to the bead *via* its C-terminus) is capable of biotin binding. All 1SA-functionalized beads are thus tethered in the same geometry, resulting in one well-defined force-loading direction. In line with this consideration, the monovalent variant exhibited a single-exponential survival time distribution. This indicates a single population with a remarkably long lifetime of approximately 8.0 h (28800 s; 95% CIs: 27900-29800 s), as inferred from a linear fit to the logarithm of the survival fraction as a function of time (**Fig. 3C**, red). This value is in good agreement with the one of 6.7 h reported recently for a smaller data set (13).

3SA is complementary to the 1SA variant, in the sense that all but the attached subunit are functional, so that three different pulling geometries are possible for 3SA. Consequently, 3SA is expected to exhibit three different populations of lifetimes. In the logarithmic representation of the survival time data (**Fig. 3C**, green), two approximately linear regimes

are visible, suggesting the existence of two populations with distinct lifetimes (**Fig. 3C**, inset). As the first, i.e. shorter-lived, population comprises approximately two thirds of all unbinding events, it appears plausible to assume that this population comprises two of the three different pulling geometries, which exhibit lifetimes that are too similar to be distinguished in this experiment. Linear fits to the two approximately linear regimes in logarithmic space (data points at times below and above 500 s, respectively) yielded lifetimes of approximately 7.5 min (453 s; 422-488 s) and 36 min (2160 s; 1930-2460 s), respectively. Two data points at approximately 3 h were as outliers not taken into account for fitting. Remarkably, the lifetime of the shortest-lived population at 65 pN is almost two orders of magnitude (64-fold) lower than the one observed for the 1SA construct.

For 4SA, both the force-loading geometries for 3SA and 1SA are possible. In line with this, both short- and long-lived populations were observed in the MT measurements on the 4SA variant (**Fig. 3C**, blue). After approximately 10 min at 65 pN, unbinding of about 60% of all initially tethered beads had already occurred. This finding is in excellent agreement with the shortest-lived population observed for 3SA. In contrast, many of the remaining tethers survived for several hours. These long-lived tethers likely correspond to the force-loading geometry of the 1SA variant. This assumption is corroborated by the fact that the decay of those beads that had survived the first two hours of the measurement closely matches the one observed in the measurements on 1SA. Fitting the data points at times above 2 h (blue line in **Fig. 3C**) yielded a lifetime of 10.0 h (36200 s; 28500-49400 s), agreeing well with the lifetime obtained for 1SA, within the 95% confidence intervals of the respective fits. Finally, the data on 4SA are consistent with the existence of another population of lifetimes corresponding to the longer-lived of the two populations observed for 3SA, although in the intermediate regime of times between approximately 10 min and 2 h, the number of observed unbinding events is not sufficient for a formal analysis. Taken together, the observations for the 4SA variant are fully in line with the 1SA and 3SA data.

Combined, the above findings confirm the hypothesis that the lifetime of biotin unbinding from SA under constant force strongly depends on the tethering geometry. This finding is in agreement with results obtained by AFM-based constant speed SMFS experiments and can likely be attributed to the same molecular mechanism: For certain pulling directions, the SA binding pocket itself is deformed before biotin leaves the pocket. This alters the energy landscape of the binding and results in lower unbinding forces for constant speed measurements [Cite Sedlak2019], and in shorter lifetimes for constant force experiments.

More importantly, from an application perspective, the force-loading geometry that yields the longest lifetime corresponds to pulling biotin out of the binding pocket of the subunit that is C-terminally tethered. The lifetime for this geometry is, at the force probed here, almost two orders of magnitude higher than for the other possible geometries. Thus, it would be

highly beneficial to utilize this geometry in applications for which high force stability is desirable. Importantly, this can straightforwardly be realized employing the 1SA variant used in our experiments.

Finally, it is noteworthy that the lifetimes obtained for the site-specifically attached 4SA used here were, both for the longest- and for the shortest-lived population, approximately four-fold higher than the respective lifetimes estimated for commercially available SA-coated beads (Dynabeads M-270 Streptavidin, Invitrogen/Thermo Fisher), as recently described (13). This difference may be explained considering that for the custom SA constructs, force was applied from the C-terminus used for attachment, whereas in the case of commercially available beads the attachment is likely not site-specific, resulting in a variety of pulling geometries. In line with this, it has recently been demonstrated in AFM SMFS experiments that a 1SA/biotin complex can withstand markedly higher forces when loaded with force from the C-terminus as compared to pulling from the N-terminus (15).

Conclusion

In this study, we showed that the lifetime of the SA/biotin interaction subjected to constant force strongly depends on the force-loading geometry. Different geometries arise from binding of biotin to one of the four binding pockets of SA and result in lifetimes that differ by almost two orders of magnitude. For force spectroscopy experiments utilizing the biotin/SA interaction as a handle, and in particular for constant force measurements, it is therefore beneficial to implement a specific tethering geometry that yields long lifetimes and thus allows for long measurement durations even at high forces. The tethering geometry that we identified as the one yielding the longest lifetimes can be easily realized in experiments by employing the 1SA variant presented in this study, thus providing a straightforward approach to highly specific and stable SMFS experiments.

Acknowledgements

The authors thank Thomas Nicolaus and Angelika Kardinal for laboratory assistance, and Wolfgang Ott, Magnus S. Bauer, and Lukas F. Milles for providing ELP linkers and the ddFLN4 construct, respectively. Philipp U. Walker is acknowledged for helpful discussions. This study was supported by research funding from the German Research Foundation (DFG) to the project “Unraveling the Mechano-Regulation of Von Willebrand Factor” (project number 386143268).

Author Contributions

All authors designed the experimental approach. S.M.S. engineered the streptavidin constructs, prepared biotinylated DNA, and performed ITC measurements. A.L. conducted

AFM imaging. A.L., S.M.S. and S.G. functionalized magnetic beads. A.L. and S.G. performed MT experiments. All authors contributed to data interpretation and writing of the manuscript.

Competing Interests Statement

The authors declare no competing interests.

References

1. Laitinen, O. H., V. P. Hytonen, H. R. Nordlund, and M. S. Kulomaa. 2006. Genetically engineered avidins and streptavidins. Cell Mol Life Sci 63:2992-3017.
2. Dundas, C. M., D. Demonte, and S. Park. 2013. Streptavidin-biotin technology: improvements and innovations in chemical and biological applications. Appl Microbiol Biotechnol 97:9343-9353.
3. Didenko, V. V. 1993. Biotinylation of DNA on membrane supports: a procedure for preparation and easy control of labeling of nonradioactive single-stranded nucleic acid probes. Anal Biochem 213:75-78.
4. Kriegel, F., N. Ermann, R. Forbes, D. Dulin, N. H. Dekker, and J. Lipfert. 2017. Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers. Nucleic Acids Res.
5. Bayer, E. A., and M. Wilchek. 1990. Protein biotinylation. Methods Enzymol 184:138-160.
6. Cull, M. G., and P. J. Schatz. 2000. Biotinylation of proteins in vivo and in vitro using small peptide tags. Methods Enzymol 326:430-440.
7. Cannizzaro, S. M., R. F. Padera, R. Langer, R. A. Rogers, F. E. Black, M. C. Davies, S. J. Tendler, and K. M. Shakesheff. 1998. A novel biotinylated degradable polymer for cell-interactive applications. Biotechnol Bioeng 58:529-535.
8. Baumann, F., M. S. Bauer, L. F. Milles, A. Alexandrovich, H. E. Gaub, and D. A. Pippig. 2016. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. Nat Nanotechnol 11:89-94.
9. Ott, W., M. A. Jobst, C. Schoeler, H. E. Gaub, and M. A. Nash. 2017. Single-molecule force spectroscopy on polyproteins and receptor-ligand complexes: The current toolbox. J Struct Biol 197:3-12.
10. Lipfert, J., J. W. Kerssemakers, T. Jager, and N. H. Dekker. 2010. Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. Nat Methods 7:977-980.
11. Walker, P. U., W. Vanderlinden, and J. Lipfert. 2018. Dynamics and energy landscape of DNA plectoneme nucleation. Physical Review E 98:042412.
12. Ganim, Z., and M. Rief. 2017. Mechanically switching single-molecule fluorescence of GFP by unfolding and refolding. Proc Natl Acad Sci U S A 114:11052-11056.
13. Löf, A., P. U. Walker, S. M. Sedlak, T. Obser, M. A. Brehm, M. Benoit, and J. Lipfert. 2018. Modular, ultra-stable, and highly parallel protein force spectroscopy in magnetic tweezers using peptide linkers. bioRxiv:491977.
14. Sedlak, S. M., M. S. Bauer, C. Kluger, L. C. Schendel, L. F. Milles, D. A. Pippig, and H. E. Gaub. 2017. Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. PLOS ONE 12:e0188722.

15. Sedlak, S. M., L. C. Schendel, M. C. R. Melo, D. A. Pippig, Z. Luthey-Schulten, H. E. Gaub, and R. C. Bernardi. 2019. Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. Nano Letters 19:3415-3421.
16. Milles, L. F., E. A. Bayer, M. A. Nash, and H. E. Gaub. 2016. Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium perfringens. J Phys Chem B.
17. Lipfert, J., X. Hao, and N. H. Dekker. 2009. Quantitative modeling and optimization of magnetic tweezers. Biophys J 96:5040-5049.
18. Cnossen, J. P., D. Dulin, and N. H. Dekker. 2014. An optimized software framework for real-time, high-throughput tracking of spherical beads. Rev Sci Instrum 85:103712.
19. Te Velthuis, A. J. W., J. W. J. Kerssemakers, J. Lipfert, and N. H. Dekker. 2010. Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. Biophysical journal 99:1292-1302.
20. De Vlaminck, I., T. Henighan, M. T. van Loenhout, D. R. Burnham, and C. Dekker. 2012. Magnetic forces and DNA mechanics in multiplexed magnetic tweezers. PLoS One 7:e41432.
21. Ostrofet, E., F. S. Papini, and D. Dulin. 2018. Correction-free force calibration for magnetic tweezers experiments. Sci Rep 8:15920.
22. Ott, W., M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, and H. E. Gaub. 2017. Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. ACS Nano 11:6346-6354.
23. Schwaiger, I., A. Kardinal, M. Schleicher, A. A. Noegel, and M. Rief. 2004. A mechanical unfolding intermediate in an actin-crosslinking protein. Nat Struct Mol Biol 11:81-85.
24. Howarth, M., D. J. Chinnapen, K. Gerrow, P. C. Dorrestein, M. R. Grandy, N. L. Kelleher, A. El-Husseini, and A. Y. Ting. 2006. A monovalent streptavidin with a single femtomolar biotin binding site. Nat Methods 3:267-273.
25. Chilkoti, A., and P. S. Stayton. 1995. Molecular Origins of the Slow Streptavidin-Biotin Dissociation Kinetics. Journal of the American Chemical Society 117:10622-10628.
26. Bauer, M. S., L. F. Milles, S. M. Sedlak, and H. E. Gaub. 2018. Monomeric streptavidin: a versatile regenerative handle for force spectroscopy. bioRxiv.
27. Basu, S., A. Finke, L. Vera, M. Wang, and V. Olieric. 2019. Making routine native SAD a reality: lessons from beamline X06DA at the Swiss Light Source. Acta Crystallographica Section D 75:262-271.
28. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33-38, 27-38.

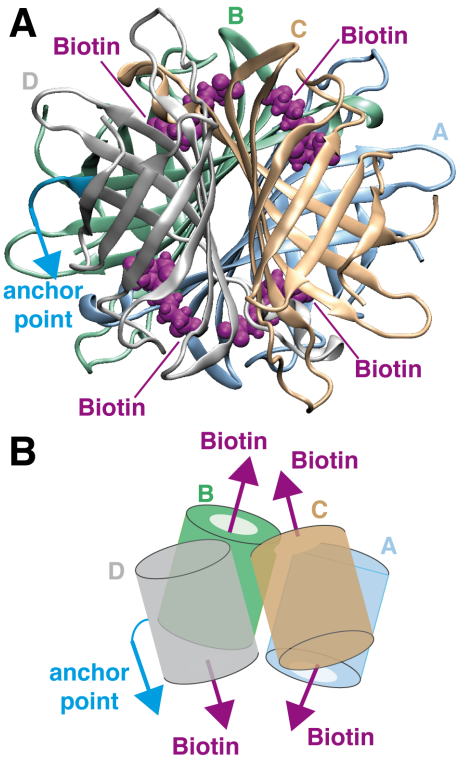


Figure 1. SA's tetravalency results in different force-loading geometries. A Crystal structure of the SA tetramer (PDB-ID: 6M9B (27), rendered using VMD (28)) with the four subunits shown in different colors. Four bound biotin molecules are shown in purple. The light blue arrow marks the anchor point (C-terminus of subunit D). **B** Schematic representation of the tetramer structure. The colored barrels represent the four subunits. Arrows indicate how force builds up in the SMFS experiments: The light blue arrow marks the C-terminus of subunit D used for site-specific immobilization. Purple arrows indicate the four possible directions of pulling biotin out of the different binding pockets. Depending on which subunit biotin is bound to, force propagates differently through the complex.

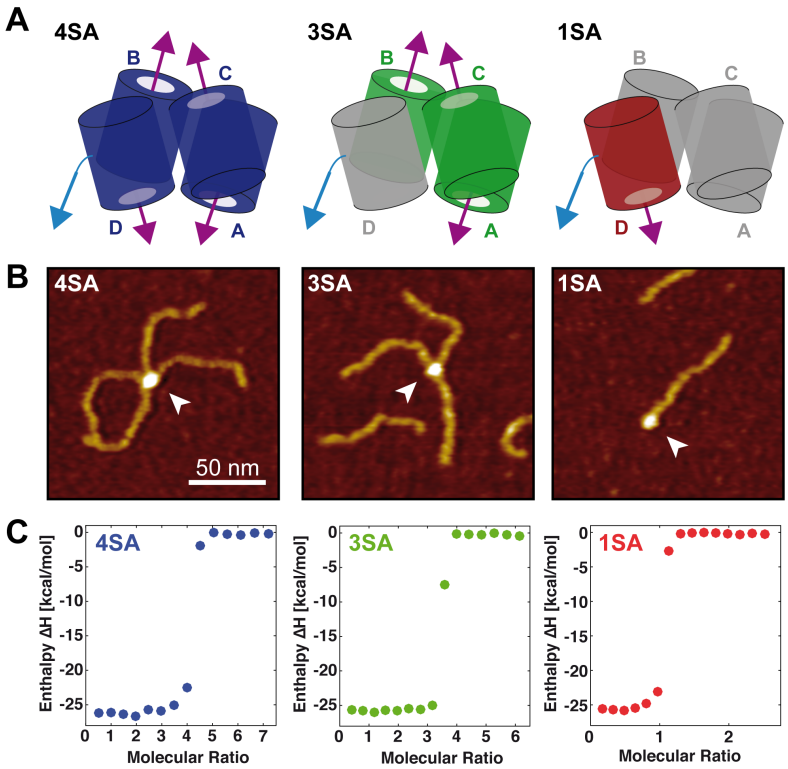


Figure 2. SA variants with different valencies. A Schematic structure of SA constructs with different valencies. 4SA (left), 3SA (middle), and 1SA (right) possess four, three and one functional subunit(s) (colored), respectively. The remaining subunits (gray) are incapable of binding biotin. All constructs possess a single C-terminal cysteine at their subunit D – nonfunctional for 3SA, functional for 1SA and 4SA– for site-specific immobilization (light blue line). In our SMFS experiments, force is applied to the constructs from the immobilization site (light blue arrows) and from the biotin being pulled out of one of the functional subunits (possible directions indicated by purple arrows). **B** Exemplary AFM images of 4SA (left), 3SA (middle), and 1SA (right) with the maximal number (four, three, and one, respectively) of biotinylated DNA strands bound. Arrowheads mark the SA molecules. Height range of color scale is 2 nm. **C** Isothermal titration calorimetry was employed to verify that the biotin binding behavior is the same for all different SA variants in the absence of force. While 4SA, 3SA and 1SA differ in binding stoichiometry, the binding enthalpy per mole biotin added is the same. Due to the high affinity of the complex, it is not possible to determine a binding constant from the data.

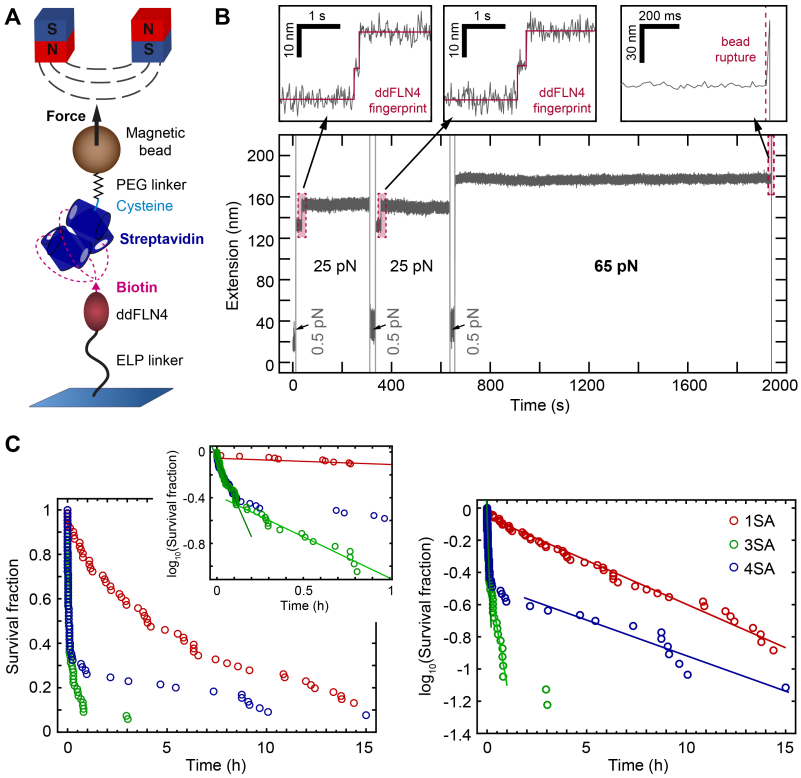


Figure 3. Lifetime of the SA/biotin interaction under constant force. A Schematic of MT experiments (not to scale). SA (4SA, 3SA or 1SA) is site-specifically and covalently immobilized on magnetic beads via the single C-terminal cysteine at its subunit D using a PEG linker with a thiol-reactive maleimide group. Biotinylated ddFLN4 is covalently immobilized on the bottom slide of the MT flow cell via an ELP linker. Binding of the biotin to one of the functional subunits of the respective streptavidin construct thus tethers the beads by a single biotin/SA bond. Force is exerted on the magnetic beads by permanent magnets positioned above the flow cell. **B** Exemplary time trace of the tether extension during an MT measurement. At the beginning of the measurement, beads are subjected to two 5-min intervals of a constant force of 25 pN, allowing for unfolding of ddFLN4 in a characteristic three-state manner (left and middle zoom-in), which serves as fingerprint to identify specific, single-tethered beads. Short low force intervals (0.5 pN) allow for ddFLN4 refolding. Tethers are then subjected to a constant force of 65 pN and the time until bead rupture due to unbinding of biotin from streptavidin is monitored (right zoom-in). **C** Survival fractions at 65 pN as a function of time for 1SA (red), 3SA (green), and 4SA (blue). Data are shown both linearly scaled (left) and logarithmically (right and inset). Lines are linear fits to the data in logarithmic space as explained in the text.

Supplementary Information

Sequences of the protein constructs

Functional SA subunit:

MEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTAL
GWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
VKPSAAS

Functional SA subunit with C-terminal cysteine and His-tag:

MEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTAL
GWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
VKPSAASCLEHHHHHH

Non-functional SA subunit:

MEAGITGTWYAQLGDTFIVTAGADGALTGTYEAAVGNAESRYVLTGRYDSAPATDGSGTAL
GWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
VKPSAAS

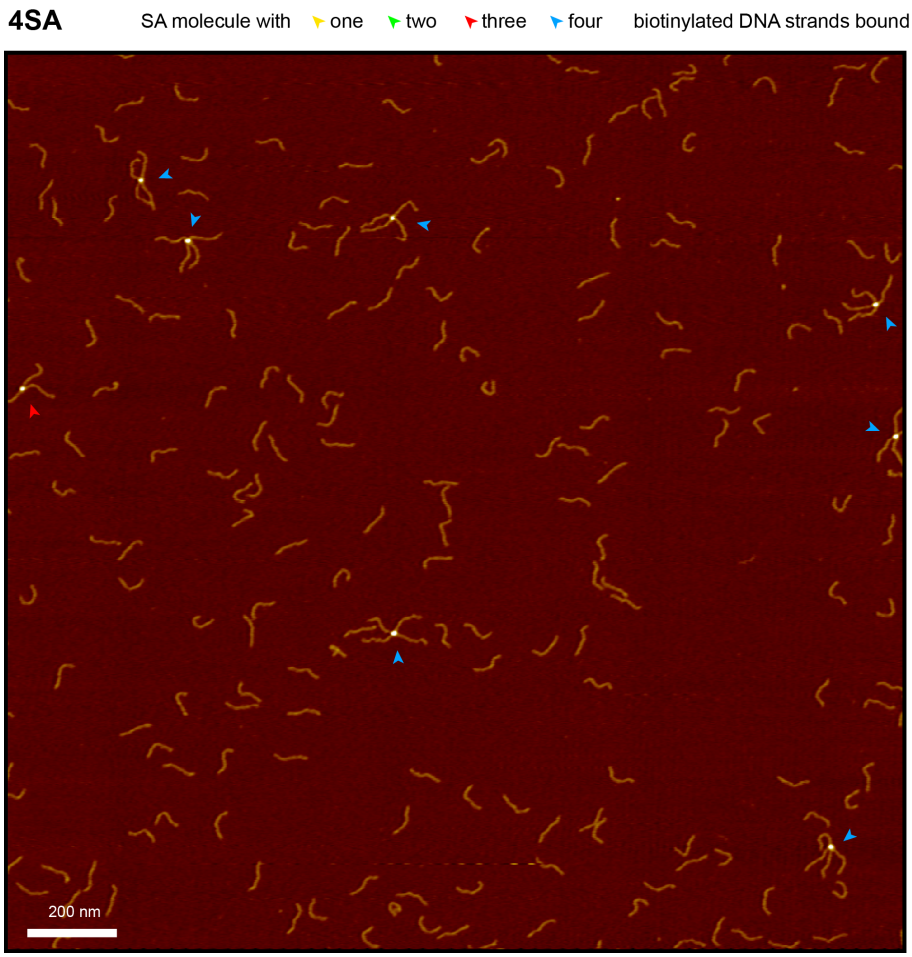
Non-functional SA subunit with C-terminal cysteine and His-tag:

MEAGITGTWYAQLGDTFIVTAGADGALTGTYEAAVGNAESRYVLTGRYDSAPATDGSGTAL
GWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
VKPSAASCLEHHHHHH

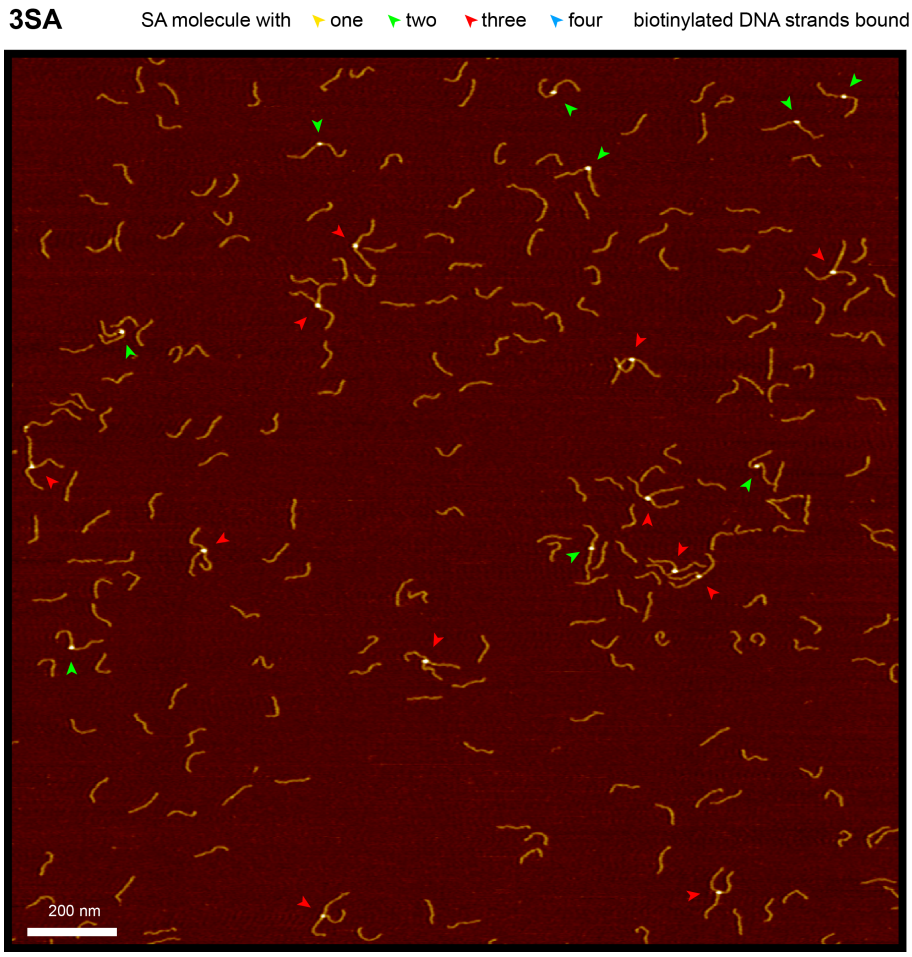
ddFLN4 (C18S) construct with N-terminal glycine and short linker sequence, and C-terminal His-tag and ybbR-tag:

MGTGSGSGSGSAGTGSGADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDG
FVV TIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPSG
HHHHHHGSDSLEFIASKLA

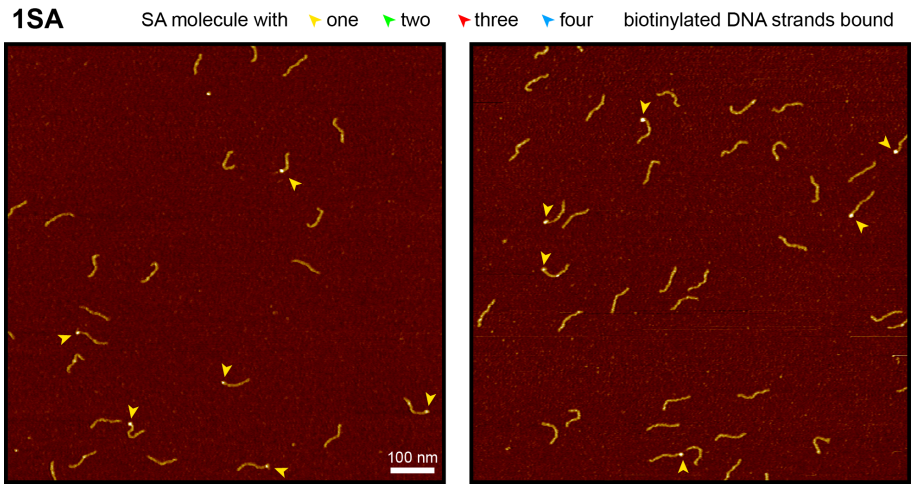
Supplementary Figures



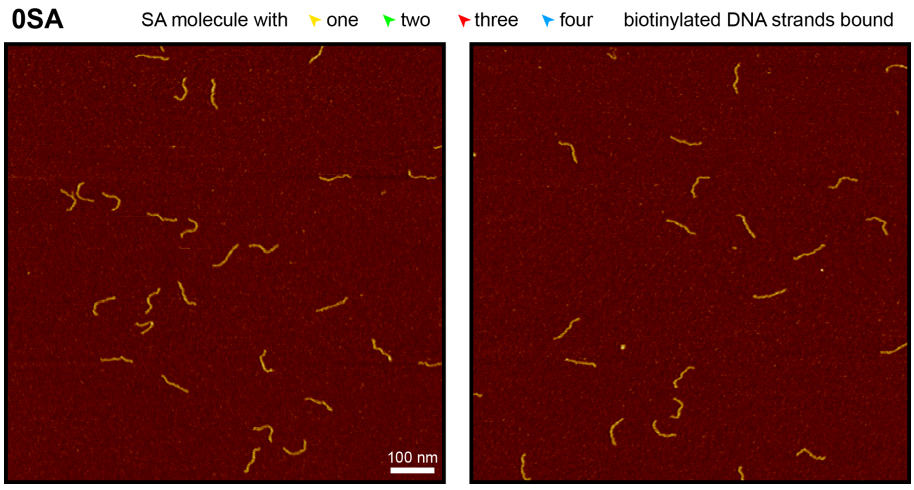
Supplementary Figure S1. AFM imaging of 4SA with biotinylated DNA.
Representative AFM images of 4SA and biotinylated 250 bp dsDNA after incubation in a 1:10 stoichiometry. Arrowheads mark streptavidin molecules, with the color of arrowheads indicating the number of bound DNA strands (yellow – one, green – two, red – three, blue – four). For 4SA, up to four bound strands were observed. Height range of color scale is 2.5 nm.



Supplementary Figure S2. AFM imaging of 3SA with biotinylated DNA.
Representative AFM images of 3SA and biotinylated 250 bp dsDNA after incubation in a 1:10 stoichiometry. Arrowheads mark streptavidin molecules, with the color of arrowheads indicating the number of bound DNA strands (yellow – one, green – two, red – three, blue – four). For 3SA, up to three bound strands were observed. No 3SA molecules with four strands were observed. Height range of color scale is 2 nm.



Supplementary Figure S3. AFM imaging of 1SA with biotinylated DNA.
Representative AFM images of 1SA and biotinylated 250 bp dsDNA after incubation in a 1:4 stoichiometry. Arrowheads mark streptavidin molecules, with the color of arrowheads indicating the number of bound DNA strands (yellow – one, green – two, red – three, blue – four). Only 1SA molecules with a single bound DNA strand were observed. Height range of color scale is 2 nm.



Supplementary Figure S4. AFM imaging of 0SA with biotinylated DNA.
Representative AFM images of 0SA and biotinylated 250 bp dsDNA after incubation in a 1:4 stoichiometry. Only free DNA strands not bound to 0SA were observed. Free 0SA molecules could not be observed since they do not stably attach to the positively charged poly-L-lysine coated mica substrate. Height range of color scale is 2 nm.

Chapter 4

The Streptavidin/Biotin System in Nanotechnological Applications

4.1 Publication P3: DNA-free Directed Assembly in Single-Molecule Cut-and-Paste

In Publication P3, the difference in unbinding forces for biotin unbinding from differently tethered mSA, as described in Publication P2, is employed to establish a force hierarchy to manipulate single biomolecules using the AFM. A so-called single-molecule cut and paste (SMC&P) assay is created. This technique uses the cantilever tip of the AFM to pick up molecules from a certain surface area (depot) and to deposit them at another surface area (target). The nanometer precision of the AFM allows arranging single molecules into a well-defined pattern. By this, distinct spatial assembly of, in principle many different, molecules can be created with nanoscale precision. In the long run, the development of this technique is aimed at designing and investigating enzyme networks on the single-molecule level from scratch. The high spatial resolution of the AFM is complemented with high temporal resolution by imaging fluorescently labeled biomolecules by TIRF microscopy in a combined AFM-TIRF setup.

The SMC&P setup described in Publication P3 features (i) a versatile shuttle construct, (ii) a protein fingerprint domain to monitor both pick-up and deposition process and (iii) a completely DNA-free attachment strategy to extend the applicability of the SMC&P assay to enzymes that have DNA as substrate.

This advancement of the SMC&P toolbox uses the mSA/biotin interaction for the immobilization of molecules on the sampe surface. The mSA/biotin system stands out with its high affinity, low off-rate and great stability. It is thus well suited for stable long-term attachment in both depot and target area. On the cantilever side, the SdrG/Fgβ-system, employed in a non-native pulling geometry, complements the assay. With N-mSA/biotin (unbinding force at about 200 pN) in the depot, non-native SdrG/Fgβ (about 300 pN) on the cantilever tip and C-mSA/biotin (about 400 pN) in the target, a force hierarchy ($F_{depot} < F_{cantilever} < F_{target}$) is created that allows for repeated transfer of molecules from the depot to cantilever tip to the target area. The efficiency of the process is optimized by adjusting the retraction velocity,

and thereby changing the force-loading rate, to minimize the overlap of the corresponding unbinding force distribution, *i.e.* favoring unbinding events from the surface (depot) or from the cantilever tip (target).

The advantage of using mSA in both depot and target is that for immobilization to the sample surface only one small ligand molecule (biotin) has to be attached to the molecule, which is to be transported. A shuttle construct is designed, to which in the future potential enzymes of interests can be fused. It consists of an N-terminal Fgβ-peptide fused to the well-characterized ddFLN4 domain, which serves as a fingerprint to monitor successful pick-up and deposition events in depot and target (after pick-up, it rapidly refolds on the cantilever tip). For biotinylation, an AviTag and a ybbR-tag are available. The shuttle construct's sequence further contains a polyhistidine tag for purification and a unique cysteine, which is employed for fluorescent labeling using a maleimide dye. At the C-terminus of the shuttle construct, a Sortase-motif will facilitate future linkages to enzymes, which are to be investigated by means of SMC&P.

The feasibility of the described SMC&P assay is verified by arranging the fluorescently labeled shuttle constructs in a distinct pattern in the target area and subsequently imaging them by TIRF microscopy.

DNA-free directed assembly in single-molecule cut-and-paste

by

Katherine R. Erlich*, Steffen M. Sedlak*, Markus A. Jobst, Lukas F. Milles, and Hermann E. Gaub

published in

Nanoscale, 2019,11, 407-411, doi:10.1039/C8NR08636B

*These authors contributed equally to this work. KRE: Protein design and preparation (shuttle construct), surface preparation, SMC&P, TIRF imaging, data analysis, writing of manuscript. SMS: Protein design (shuttle construct and SA) and preparation (SA), AFM-based SMFS, cantilever preparation, SMC&P, data analysis, writing of manuscript

Nanoscale

COMMUNICATION

View Article Online
View Journal

Cite this: DOI: 10.1039/c8nr08636b

Received 25th October 2018,
Accepted 1st December 2018
DOI: 10.1039/c8nr08636b

rsc.li/nanoscale

DNA-free directed assembly in single-molecule cut-and-paste†

Katherine R. Erlich, ‡ Steffen M. Sedlak, ‡ Markus A. Jobst, Lukas F. Milles and Hermann E. Gaub*

Single-molecule cut-and-paste facilitates bottom-up directed assembly of nanoscale biomolecular networks in defined geometries and enables analysis with spatio-temporal resolution. However, arrangement of diverse molecules of interest requires versatile handling systems. The novel DNA-free, genetically encodable scheme described here utilises an orthogonal handling strategy to promote arrangement of enzymes and enzyme networks.

The spatial organisation of molecules is of key interest in both single-molecule studies as well as the broader field of nanotechnology. Arrangement of biomolecular structures may be accomplished *via* two general approaches: self-assembly and directed assembly. The former strategy encompasses a wide range of programmable structures, including engineered protein modules[1] and prominently DNA origami.[2,3] Notably, a recent novel drug-delivery strategy *via* activated DNA origami showed potent tumour-inhibiting activity,[4] demonstrating the profound utility of spatially arranged molecules.

Directed assembly of single molecules is possible with single-molecule cut-and-paste (SMC&P), merging bottom-up spatial assembly and exceptionally precise control of molecular positioning. This technique utilises an atomic force microscope (AFM) cantilever tip to pick up and deposit single molecules with nanometre precision at defined positions on a surface. SMC&P relies on a pre-programmed force hierarchy to facilitate the transfer of molecules from the depot area to the cantilever to the target area. The handled molecules are probed *via* single-molecule force spectroscopy (SMFS), which provides critical feedback of the success of the transfer, and the assembled pattern is imaged *via* total internal reflection fluorescence (TIRF) microscopy. Additionally, SMC&P enables precise arrangement of single molecules within nanoapertures such as zero-mode waveguides, thereby circumventing complications that arise from stochastic immobilisation such as heterogeneity of fluorescence intensity and lifetime caused by interference from metallic sidewalls.[5]

Previous iterations of SMC&P have undertaken arrangements and time-resolved fluorescent measurements of various biomolecules, including labelled DNA, DNA aptamers, green fluorescent protein, nanoparticle recognition sites, and diverse handling tags,[6,7,8–12] demonstrating the versatility of this technique. Furthermore, SMC&P presents a unique opportunity for investigation of enzymes and enzyme networks on the level of single molecules, arranged with precisely controlled geometry *via* directed assembly.

As an emerging technique, SMC&P compels continuous developments to increase its robustness and broaden its scope. In particular, SMC&P has previously relied on DNA to anchor molecules to the surface. Although this strategy confers reliable and stable immobilisation of transfer molecules, its scope is limited. A DNA-based approach presents difficulties for the arrangement of molecules with affinity for DNA, which would bind the covalently attached DNA anchors. This secondary interaction would both decrease SMC&P transfer efficiency as well as impact the behaviour of the molecules of interest. Consequently, this strategy is particularly unsuited for the study of DNA-binding proteins and enzymes. Moreover, the synthesis of protein–DNA hybrid molecules required for protein arrangement is often laborious. SMFS analysis in SMC&P has previously also had limited applicability; probed molecules have lacked fingerprint domains to identify specific single-molecule events, and the low-force regimes of the handling systems were partly overlaid with the instrument noise.

Here, we present a revised strategy that greatly expands the SMC&P toolbox, improves the technique's versatility and makes substantial progress towards SMC&P-based investigation of enzyme networks. The newly developed system is DNA-free, and is instead based on a protein-small molecule interaction for surface immobilisation. Simultaneously, a

Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 München, Germany.
E-mail: gaub@lmu.de
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c8nr08636b
‡These authors contributed equally to this work.

reliable fingerprint domain and increased rupture forces significantly enhance SMFS analysis of SMC&P transfer both in real-time and subsequent statistical analyses.

Results and discussion

Monovalent streptavidin (mSA), a heterotetrameric complex that binds the small molecule biotin with ultrahigh affinity, was recently employed in AFM-based SMFS.[13] Anchored by a single functional subunit in a well-defined pulling geometry, it was additionally discovered that the tethering geometry of mSA strongly influences the rupture force of the mSA : biotin bond; N-terminally tethered mSA (N-mSA) unbinds from biotin at forces around 200 pN, while C-terminally tethered mSA (C-mSA) unbinds around 450 pN, in both cases depending on force loading rate.[13] This geometry-dependent behaviour was exploited in SMC&P to immobilise the transfer molecule with both low- and high-rupture forces *via* the same small biotin label (Fig. 1). The adhesin SD-repeat protein G N2N3 domain (SdrG) from *Staphylococcus epidermidis*[14] binds a short peptide from the N-terminus of human fibrinogen β[15] (Fgβ) with remarkably high AFM-measured rupture forces of over 2 nN when probed in the native geometry of C-terminally immobilised SdrG and C-terminally pulled Fgβ.[14,16] When probed in a non-native geometry of N-terminally immobilised SdrG (N-SdrG) and C-terminally pulled Fgβ, we found that the loading rate-dependent unbinding forces are in the range of 250 pN. These binding pairs of N-mSA : biotin in the depot area, N-SdrG : Fgβ in a non-native geometry on the cantilever tip, and C-mSA : biotin in the target area form the force hierarchy required for SMC&P.

The fourth filamin domain from *Dictyostelium discoideum* (ddFLN4) demonstrates reliable and rapid refolding as a low-force fingerprint in AFM-based SMFS.[17–19] A transfer construct consisting of a modified ddFLN4 motif was designed, recombinantly expressed, and purified with several key additional protein tags. Namely, an N-terminal Fgβ peptide sequence, which is accordingly pulled C-terminally, enables specific handling by an N-SdrG-coupled cantilever. The ddFLN4 domain also harbours a ybbR tag[20] at its C-terminus, followed by a C-terminal reactive cysteine to enable covalent modification with Coenzyme A-biotin and maleimide-Cy5, respectively. The final transfer construct consists of an efficiently labelled Fgβ-ddFLN4-biotin-Cy5 chimera (details of purification and labelling in Supplement) that binds to mSA *via* biotin and N-SdrG *via* Fgβ, and is imaged in TIRF microscopy

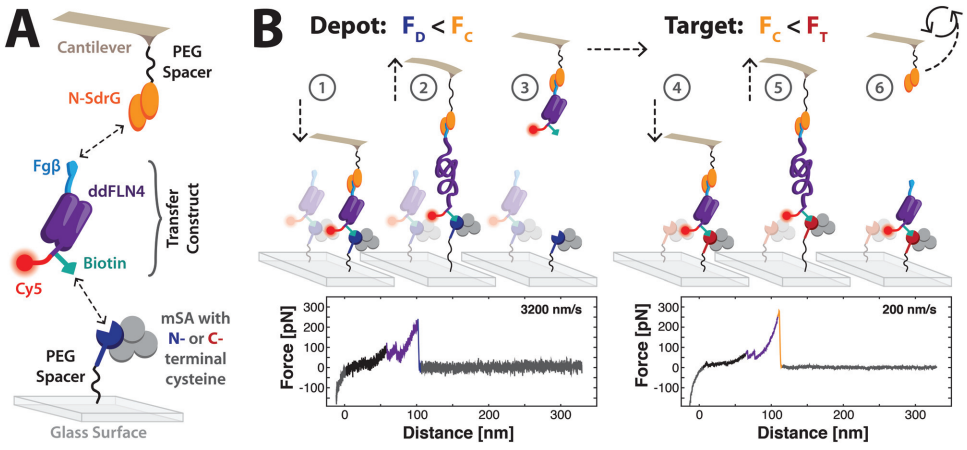


Fig. 1 Schematic of the molecules used in SMC&P and the mechanism of SMC&P cycling. (A) SdrG is N-terminally immobilised to the cantilever tip, and monovalent streptavidin with an N- or C-terminal reactive cysteine is immobilised on a glass surface. The chimeric transfer construct is composed of a ddFLN4 domain, which contains an N-terminal Fgβ (*i.e.* C-terminally pulled) tag for specific handling by the cantilever tip. At its C-terminus, the protein is additionally modified with biotin *via* a ybbR tag for specific immobilisation on a streptavidin-functionalised surface and a Cy5 fluorophore for fluorescence imaging. (B) A force hierarchy governs the repeatable transfer of molecules in SMC&P. The force required to rupture the N-mSA : Biotin bond in the depot (F_D), the N-SdrG : Fgβ bond in this geometry on the cantilever tip (F_C) and the C-mSA : Biotin bond in the target (F_T) are tuned such that $F_D < F_C < F_T$. The cantilever tip approaches the depot surface and N-SdrG binds the Fgβ tag of an immobilised transfer construct (1). As the cantilever retracts, the ddFLN4 domain unfolds under force (2). The molecules are pulled in series until the N-mSA : Biotin bond finally ruptures, releasing the transfer construct and allowing ddFLN4 to rapidly re-fold (3). The cantilever tip loaded with the transfer construct cargo travels to the target area and approaches, allowing the C-mSA : Biotin bond to form (4). The cantilever tip again retracts and unfolds ddFLN4 (5) until the comparatively weak N-SdrG : Fgβ bond ruptures. The unloaded cantilever is recycled back to the depot area to repeat the process (6). Force–distance curves of specific single-molecule interactions show the two-step ddFLN4 unfolding pattern (purple traces) and a higher final peak associated with the rupture of N-mSA : Biotin in the depot (blue trace) or N-SdrG : Fgβ in the target (orange trace).

via Cy5 (Fig. 1a). Importantly, the forces applied during the SMC&P process do not propagate through the transferred molecule of interest, here the fluorescent dye at the very C-terminus.

A custom-built hybrid AFM/TIRF microscope was employed for SMC&P.[21] The depot area consists of N-mSA covalently attached to the surface, and transfer constructs that are specifically immobilised *via* the biotin label. In the target area, C-mSA is covalently attached to the surface. A cantilever tip functionalised with N-terminally immobilised N-SdrG picks up transfer construct molecules from the depot area and deposits them in the target area, a process that critically relies on a well-defined hierarchy of rupture forces. The most probable rupture forces of N-mSA : biotin in the depot (F_D), N-SdrG : Fgβ on the cantilever (F_C) and C-mSA : biotin in the target (F_T) are tuned such that $F_D < F_C < F_T$, thereby enabling reliable transfer of molecules from the depot area to the cantilever tip to the target area.

Repeatable cycling throughout the cut-and-paste process is essential to SMC&P (Fig. 1b). Transfer construct molecules bound to N-mSA in the depot area are pulled by an N-SdrG-coupled cantilever. The forces required to rupture both the N-mSA : biotin bond and the N-SdrG : Fgβ are large enough that the ddFLN4 motif is fully unfolded, visible in single-molecule force–distance curves. Eventually the weaker non-covalent bond of N-mSA : biotin ruptures, the force load drops, and the ddFLN4 motif rapidly refolds. The cantilever, loaded with the transfer construct cargo, is then moved to the target area. As the cantilever approaches the surface, the C-mSA : biotin interaction forms, thereby immobilising the transfer construct to the surface again. The cantilever retracts and again unfolds ddFLN4, visible in single-molecule force–distance curves. As the C-mSA : biotin bond is stronger, the N-SdrG : Fgβ bond eventually ruptures. The ddFLN4 domain of the immobilised transfer construct again rapidly refolds, and the cantilever is moved back to the depot to repeat the process. Force–distance curves captured during SMC&P reflect the unfolding of the two-step ddFLN4 fingerprint domain followed by a final rupture of either mSA : biotin or N-SdrG : Fgβ.

The retraction velocities of the cantilever in the depot and the target area were tuned to decrease the overlap of the rupture force probability distributions of the two probed binding pairs (Fig. 2a). The rupture force of the N-SdrG : Fgβ bond demonstrates a stronger dependence on force loading rate compared to the N- and C-mSA : biotin bond.[24] This difference in loading rate dependence was exploited to favour the rupture of the lower-force binding pair and hence relocate the transfer construct. Fast retraction (3200 nm s^{-1}) in the depot made it possible to increase the likelihood of the rupture of N-mSA : biotin over N-SdrG : Fgβ, while slow retraction (200 nm s^{-1}) in the target favoured the rupture of N-SdrG : Fgβ over C-mSA : biotin. Observed final rupture peaks in both the depot (Fig. 2b) and the target (Fig. 2c) correspond to the approximate expected rupture forces and respective distribution spreads for the two receptor–ligand pairs at the given loading rates. Due to the broad distribution of N-SdrG : Fgβ

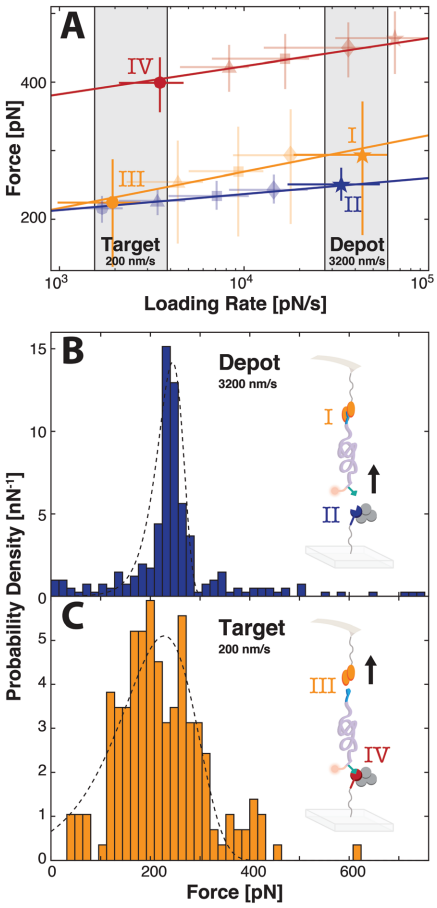


Fig. 2 Dynamic force spectra and forces associated with the final peaks of force traces observed during SMC&P. (A) The dynamic force spectra of the rupture of N-mSA : biotin (blue), N-SdrG : Fgβ (orange), and C-mSA : biotin (red) display a variable dependence of rupture force on loading rate. Pulling at 3200 nm s^{-1} favours the unbinding of N-mSA : biotin (II) rather than N-SdrG : Fgβ (I), while pulling at 200 nm s^{-1} strongly favours the unbinding of N-SdrG : Fgβ (III) rather than C-mSA : biotin (IV). The N-SdrG : Fgβ force spectrum was measured with covalent attachment of ddFLN4 to the surface (*cf.* Supplement). The mSA : biotin data are taken from Sedlak *et al.*[24] Error bars are given by the full-width at half maximum of the corresponding distributions. Regressions are fitted with the Bell–Evans model. Symbols indicate the cantilever retraction velocity: 200 nm s^{-1} (circles), 400 nm s^{-1} (triangles), 800 nm s^{-1} (squares), 1600 nm s^{-1} (diamonds), and 3200 nm s^{-1} (stars). (B) Force–distance curves captured during SMC&P in the depot correspond to the unbinding of the N-mSA : biotin (II), (C) while the target curves correspond to the unbinding of N-SdrG : Fgβ (III). Each complex has an expected rupture force of approximately 200 pN at the given respective loading rates. Forces were binned with a width of 16 pN. The histograms are fitted by the Bell–Evans formula for the distribution of the rupture forces (dashed lines).[25,26]

unbinding, there is an overlap with the rupture force distribution of N-mSA : biotin, even at a pulling speed of 3200 nm s^{-1}. Consequently, in the depot there is a chance that N-SdrG : Fgβ instead of N-mSA : biotin unbinds. Although it is not possible to distinguish between these two rupture events based solely on force curves, the probability of rupturing N-mSA : biotin may be further favoured by additionally increasing the pulling speed. With the narrow N-mSA : biotin rupture force distribution, the pick up process in the depot at 3200 nm s^{-1} is efficient enough to reliably transport molecules.

As a proof of principle, molecules were transferred *via* SMC&P and arranged in the target area in a 442-point pattern of a rocket ship (Fig. 3). Fluorescent immobilised molecules were detected *via* Cy5 excitation at 640 nm and imaged with TIRF microscopy. Patchiness in the pattern may be partially due to incomplete labelling or photobleaching of transfer constructs during purification and experimental setup. Additionally, as the underlying rupture forces in SMC&P are dependent on rupture force probability distributions that are not perfectly separated, there are cases where a cycle fails to transport any molecules. Similarly, there is a certain probability that transfer constructs bound to the cantilever dissociate during transport. Surface defects and uneven densities may also influence the efficiency of SMC&P, resulting in heterogeneously distributed mSA. However, these variations may be controlled for by a combination of force–distance curves and fluorescent signal; a successfully transferred non-fluorescent molecule, due to absence or bleaching of Cy5, produces a deposition curve in the target but no fluorescent signal in TIRF microscopy, while an unsuccessful transport cycle produces neither.

The complement of molecules utilised in SMC&P here offers several advantages compared to previous iterations. Importantly, this system is DNA-free – a key improvement required for the assembly of DNA-binding proteins and enzymes that would likely bind covalently attached DNA anchors. Not only would this potentially interfere with protein function, SMC&P efficiency could be impacted as well by reducing the likelihood that the DNA anchor is free to hybridise with its immobilised complementary strand. The immobilisation strategy presented here is likely orthogonal for most biomolecules, thereby significantly increasing the versatility of the system. Moreover, strategic integration of a domain of interest would protect it from the force propagation pathway. The domain could be simply inserted C-terminally of the ybbR tag either *via* direct chimeric expression as a continuous peptide chain, or post-translationally, *e.g. via* Sortase tag-mediated covalent joining.[22]

The introduction of the small ddFLN4 fingerprint domain is also exceptionally useful for force trace analysis. As a well-characterised and reliable fingerprint, ddFLN4 improves algorithmic curve sorting to isolate single and specific pulling events. Additionally, ddFLN4 was demonstrated to improve solubility of otherwise insoluble proteins[19] – a common difficulty of recombinant protein expression and purification. Furthermore, post-translational labelling of proteins with nucleotides in a controlled manner is not a trivial process. As performed previously in SMC&P, proteins may be labelled with CoA-DNA *via* a ybbR tag and reaction with Sfp (as was performed here similarly for labelling with biotin). This is a step that is necessarily performed post-translationally and *in vitro*. In contrast, biotin labelling may be performed *in vivo* during protein production with additional recombinant factors, such as an AviTag.[23] Similarly, Cy5-labelling may be replaced with a genetically encoded cargo such as a fluorescent protein domain, *e.g.* green fluorescent protein. On the other hand, the utilised strategy of cysteine-based labelling forgoes a need to create large chimeric protein constructs and enables fluorescent imaging of any protein of interest.

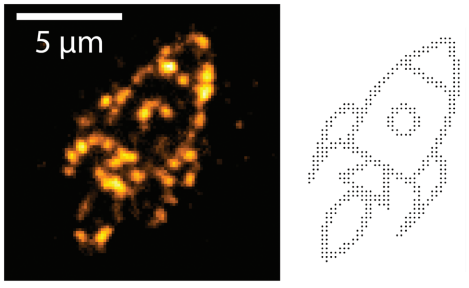


Fig. 3 Cy5-labeled transfer molecules arranged in a rocket ship pattern by SMC&P and imaged with TIRF microscopy. The image is composed of the average pixel intensity of 20 stacked frames (0.12 s exposure time at ~10 W cm^{-2}) with red laser excitation at 640 nm (left). The pattern consists of 442 points spaced 200 nm apart (right).

Conclusions

Single-molecule approaches offer invaluable insights into the function of biomolecules. SMC&P enables precise arrangement of networked molecules on a surface in well-defined geometries as well as within the centres of nanoapertures, demonstrating the unique potential of this technique to investigate the spatio-temporal coordination within enzyme networks. However, the previously established DNA-based SMC&P immobilisation system necessarily limits the range of molecules that may be arranged by bottom-up assembly. DNA-binding proteins and enzymes would likely display unwanted interactions with the covalently attached DNA anchor, thereby impacting both enzyme behaviour as well as SMC&P efficiency. The mSA : biotin system introduced here offers an immobilisation strategy that is orthogonal to the function of most enzymes. Furthermore, the diverse reactive tags allow for flexible construct design, and the utilised construct's ddFLN4 fingerprint simplifies SMFS data analysis. The advances demonstrated here set a methodological foundation for the precise single-molecule arrangement of diverse biomolecules, and enzymes in particular. Thus, we provide a means to study their behaviour as

isolated molecules as well as in an organised network to gain a more comprehensive understanding of enzyme function.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors gratefully acknowledge Ellis Durner for discussions and assistance with the TIRF instrument, Leonard C. Schendel for the glucose oxidase system, and Diana A. Pippig for providing some of the mSA subunits. Furthermore, the authors thank Angelika Kardinal and Thomas Nicolaus for laboratory support. Steffen M. Sedlak thanks the Nanosystems Initiative Munich for support. The authors gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, Ga 309/11-1).

Notes and references

1 S. Hirschi, M. Stauffer, D. Harder, D. J. Muller, W. Meier and D. Fotiadis, *Chimia*, 2016, **70**, 398–401.
2 F. C. Simmel, *Curr. Opin. Biotechnol*, 2012, **23**, 516–521.
3 A. Rajendran, E. Nakata, S. Nakano and T. Morii, *ChemBioChem*, 2017, **18**, 696–716.
4 S. Li, Q. Jiang, S. Liu, Y. Zhang, Y. Tian, C. Song, J. Wang, Y. Zou, G. J. Anderson, J. Y. Han, Y. Chang, Y. Liu, C. Zhang, L. Chen, G. Zhou, G. Nie, H. Yan, B. Ding and Y. Zhao, *Nat. Biotechnol.*, 2018, **36**, 258–264.
5 S. F. Heucke, F. Baumann, G. P. Acuna, P. M. Severin, S. W. Stahl, M. Strackharn, I. H. Stein, P. Altpeter, P. Tinnefeld and H. E. Gaub, *Nano Lett.*, 2014, **14**, 391–395.
6 S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl and H. E. Gaub, *Science*, 2008, **319**, 594–596.
7 E. M. Puchner, S. K. Kufer, M. Strackharn, S. W. Stahl and H. E. Gaub, *Nano Lett.*, 2008, **8**, 3692–3695.
8 S. K. Kufer, M. Strackharn, S. W. Stahl, H. Gumpp, E. M. Puchner and H. E. Gaub, *Nat. Nanotechnol.*, 2009, **4**, 45–49.
9 M. Strackharn, S. W. Stahl, P. M. Severin, T. Nicolaus and H. E. Gaub, *Chemphyschem*, 2012, **13**, 914–917.
10 M. Strackharn, S. W. Stahl, E. M. Puchner and H. E. Gaub, *Nano Lett.*, 2012, **12**, 2425–2428.
11 D. A. Pippig, F. Baumann, M. Strackharn, D. Aschenbrenner and H. E. Gaub, *ACS Nano*, 2014, **8**, 6551–6555.
12 K. R. Erlich, F. Baumann, D. A. Pippig and H. E. Gaub, *Small Methods*, 2017, **1**, 1700169.
13 S. M. Sedlak, M. S. Bauer, C. Kluger, L. C. Schendel, L. F. Milles, D. A. Pippig and H. E. Gaub, *PLoS One*, 2017, **12**, e0188722.
14 P. Herman, S. El-Kirat-Chatel, A. Beaussart, J. A. Geoghegan, T. J. Foster and Y. F. Dufrene, *Mol. Microbiol.*, 2014, **93**, 356–368.
15 K. Ponnuraj, M. G. Bowden, S. Davis, S. Gurusiddappa, D. Moore, D. Choe, Y. Xu, M. Hook and S. V. L. Narayana, *Cell*, 2003, **115**, 217–228.
16 L. F. Milles, K. Schulten, H. E. Gaub and R. C. Bernardi, *Science*, 2018, **359**, 1527–1533.
17 I. Schwaiger, A. Kardinal, M. Schleicher, A. A. Noegel and M. Rief, *Nat. Struct. Mol. Biol.*, 2004, **11**, 81–85.
18 L. F. Milles, E. A. Bayer, M. A. Nash and H. E. Gaub, *J. Phys. Chem. B*, 2017, **121**(15), 3620–3625.
19 M. S. Bauer, L. F. Milles, S. M. Sedlak and H. E. Gaub, 2018, bioRxiv.
20 J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15815–15820.
21 H. Gumpp, S. W. Stahl, M. Strackharn, E. M. Puchner and H. E. Gaub, *Rev. Sci. Instrum.*, 2009, **80**, 063704.
22 E. Durner, W. Ott, M. A. Nash and H. E. Gaub, *ACS Omega*, 2017, **2**, 3064–3069.
23 S. S. Ashraf, R. E. Benson, E. S. Payne, C. M. Halbleib and H. Gron, *Protein Expression Purif.*, 2004, **33**, 238–245.
24 S. M. Sedlak, L. C. Schendel, M. Cardoso dos Reis Melo, D. A. Pippig, Z. Luthey-Schulten, H. E. Gaub and R. C. Bernardi, *Nano Lett.*, 2018, DOI: 10.1021/acs.nanolett.8b04045.
25 G. I. Bell, *Science*, 1978, **200**, 618–627.
26 E. Evans and K. Ritchie, *Biophys. J.*, 1997, **72**, 1541–1555.

Electronic Supplementary Material (ESI) for Nanoscale.

Supplementary Information

DNA-Free Directed Assembly in Single-Molecule Cut-and-Paste

Katherine R. Erlich,[‡a] Steffen M. Sedlak,[‡a] Markus A. Jobst,[a] Lukas F. Milles,[a] and Hermann E. Gaub[*a]

[a]Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 München

[‡]These authors contributed equally to this work.

Correspondence to: H.E.G. gaub@lmu.de

Table of contents

1. Protein Expression, Purification, and Labelling

1.a. Preparation of Monovalent Streptavidin

Monovalent Streptavidin (mSA) with N-terminal immobilisation (N-mSA) was previously expressed, purified and assembled by Sedlak *et al*.[1] C-terminally immobilised mSA (C-mSA) was created here using the same protocol. In brief, three different streptavidin subunits were designed: a functional subunit with a polyhistidine tag and a single cysteine at its N-terminus, a functional subunit with a polyhistidine tag and a single cysteine at its C-terminus, and a non-functional subunit (N23A, S27D, S45A).[2]

Functional SA subunit with C-terminal cysteine (orange) and His-tag (green):

MEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTALGWTVA WKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAASCLEH HHHHH

Functional SA subunit with N-terminal cysteine (orange) and His-tag (green):

MGSSHHHHHHMCGSEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAP ATDGSGTALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHD TFTKVKPSAAS

Non-functional SA subunit (mutated residues in red):

MEAGITGTWYAQLGDTFIVTAGADGALTGTYEAAVGNAESRYVLTGRYDSAPATDGSGTALGWTVA WKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAAS

The three different subunits were cloned into pET vectors and expressed separately in *E.coli* BL21(DE3)-CodonPlus. In the following steps, the different subunits were treated separately: Each harvested cell pellet was dissolved in B-PER reagent. Lysozyme and DNAse I was added. Full cell lysis was achieved by sonication. Inclusion bodies formed and were regained by centrifuging the solution at 20,000 x g for 30 min and discarding the supernatant. The inclusion bodies containing pellet was suspended in washing buffer (phosphate buffered saline, 0.1% Triton X-100, 1 mM DTT). The centrifugation and washing was repeated until the supernatant was clear. The inclusion bodies were then dissolved in denaturation buffer (phosphate buffered saline, 6 M guanidine hydrochloride, pH 7.5). Non-functional and functional subunits (either with N- or C-terminal tags, again treated separately in the next steps) were mixed in a 10:1 ratio as given by the absorption at 280 nm. Refolding into streptavidin tetramers was accomplished by slowly dissolving the mixtures in 500 ml refolding buffer (phosphate buffered saline, 10 mM beta-mercaptoethanol) and stirring it at 4°C overnight. The refolding solution was centrifuged to remove precipitated protein, filtered with a 0.22 µm cellulose filter, and loaded onto a 5 ml HisTrap FF column (GE Healthcare Life Sciences, Little Chalfont, UK). Monovalent streptavidin was eluted from the column using a linear gradient from 10 mM to 250 mM imidazole. Elution fractions were analysed by gel electrophoresis. The eluted monovalent streptavidin was dialysed against phosphate buffered saline, pH 7.4 and stored at 4°C.

1.b. Preparation of N-SdrG

SdrG N2N3 was expressed and purified as previously described by Milles *et al.*[3] In brief, ybbr-SdrG-6xHIS was expressed in a pET28a vector with a 6xHis-tag and an N-terminal ybbR tag for covalent immobilisation to Coenzyme A.

The protein sequence is (ybbR tag in blue, His tag in green):

MATDSLEFIASKLATEQGSNVNHLIKVTDQSITEGYDDSDGIIKAHDAENLIYDVTFEVDDKVKSGDTM TVNIDKNTVPSDLTDSFAIPKIKDNSGEIIATGTYDNTNKQITYTFTDYVDKYENIKAHLKLTSYIDKSKVP NNNTKLDVEYKTALSSVNKTITVEYQKPNENRTANLQSMFTNIDTKNHTVEQTIYINPLRYSAKETNVN ISGNGDEGSTIIDDSTIIKVYKVGDNQNLPDSNRIYDYSEYEDVTNDDYAQLGNNNDVNINFGNIDSPY IIKVISKYDPNKDDYTTIQQTVTMQTTINEYTGEFRTASYDNTIAFSTSSGQGQGDLPPEKTELKLPRSR HHHHHH

A 5 ml preculture of LB medium containing 50 µg/ml Kanamycin grown overnight at 37°C was inoculated in 200 ml ZYM-5052 autoinduction medium[4] containing 100 µg/ml Kanamycin and grown at 37°C for 6 h, then at 18°C overnight. Cells were harvested by centrifugation at 8,000 x g, and pellets were stored frozen at -80°C until purification.

All purification steps were performed at 4°C or on ice when possible. The bacteria pellet was resuspended in a Lysis Buffer and cells were lysed through sonication followed by centrifugation at 40,000 x g for 45 min. The supernatant was applied to a Ni-NTA column for purification by Ni-IMAC and eluted with a buffer containing 200 mM imidazole. Protein-containing fractions were concentrated in centrifugal filters, exchanged into measurement buffer by desalting columns, and frozen in aliquots with 10% (v/v) glycerol in liquid nitrogen

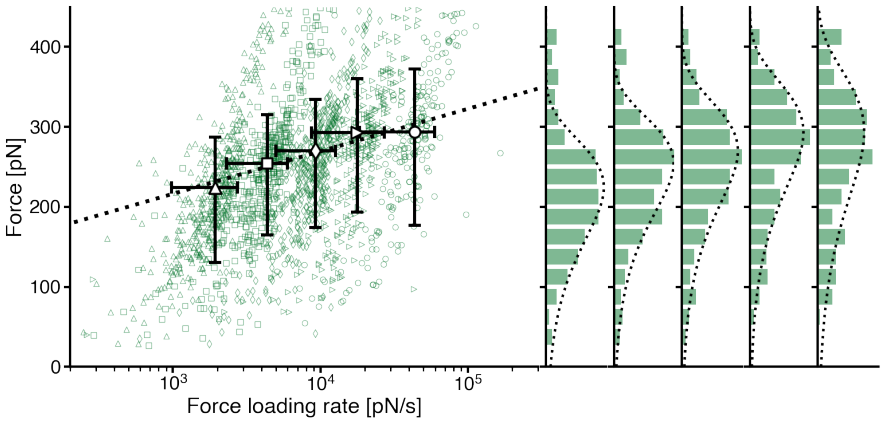


Figure S1. Dynamic force spectrum of the rupture of N-terminally immobilised SdrG and an N-terminal (*i.e.* C-terminally pulled) Fgβ-tag measured with AFM-based SMFS at retraction velocities from left to right: 200 nm/s, 400 nm/s, 800 nm/s, 1600 nm/s, 3200 nm/s. In this geometry, the N-SdrG:Fgβ bond ruptures between approximately 200-300 pN, depending on the loading rate.

to be stored at -80°C until used in experiments. The final protein concentration was 848 M as measured by the absorbance at 280 nm via NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA).

N-terminally immobilised SdrG bound to its target peptide ligand Fgβ as an N-terminal tag was additionally probed with AFM-based SMFS to acquire a dynamic force spectrum of the rupture force.

1.c. Preparation ddFLN4 Transfer Protein

A transfer construct whose main fold consists of the fourth filamin domain from *Dictyostelium discoideum* (ddFLN4) with a crucial C18S mutation to prevent disulfide bond formation and undesired reaction to the Maleimide-dye used here was designed with several handling and purification tags. The construct harbours an N-terminal Fgβ-tag (NEEGFFSARGHRPLD) to enable direct binding to SdrG. An internal 6xHis tag was included for purification by Ni-IMAC. The construct also harbours an internal ybbR-tag (DSLEFIASKLA) to covalently modify the protein with Biotin. A Sortase tag (LPETGG) was also included,

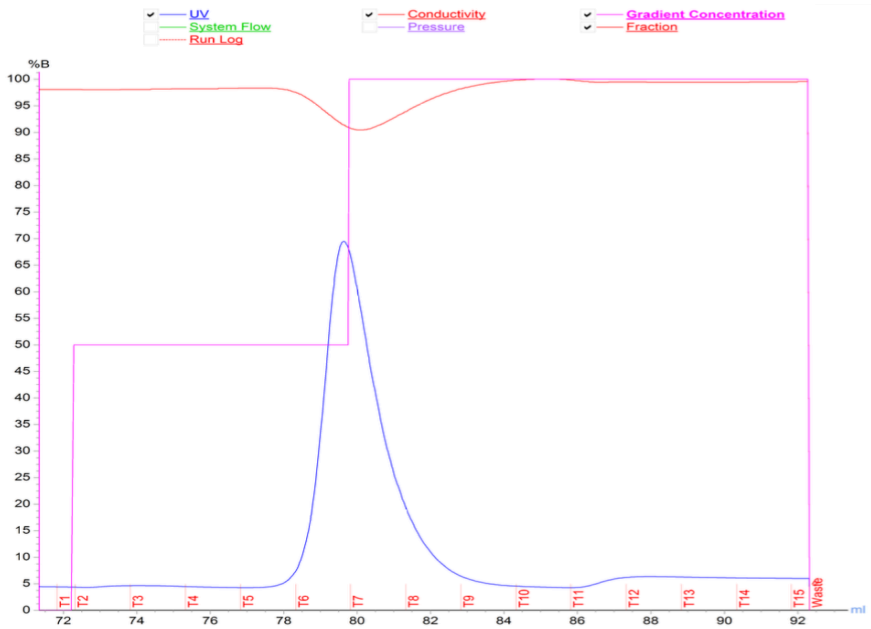


Figure S2. Chromatogram of ddFLN4 construct purification by Ni-IMAC. The 6xHis-tagged protein was purified by step gradient and eluted with high imidazole. Fractions 6-8 span the majority of the major peak, with a smaller peak spanning fractions 11-13. Note that there is a 6 ml delay in reported percentage elution buffer as the program does not take into account tubing length and column dead volume. Therefore, the major peak is observed at 50% elution buffer, or 135 mM imidazole.

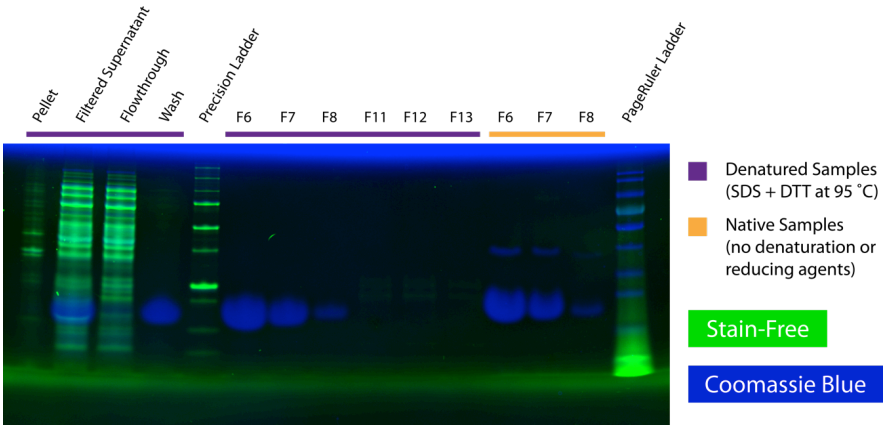


Figure S3. SDS-PAGE and Native PAGE analysis of the ddFLN4 protein construct purification by Ni-IMAC. The ddFLN4 construct was obtained in the soluble fraction and efficiently eluted by step gradient purification. Fractions 6-8 contained the majority of the protein, visible by the overloaded blue bands. These fractions were additionally analysed under native conditions to examine the extent of dimerisation via C-terminal cysteine. Most of the protein is in a reduced state and therefore does not require further reduction before maleimide-Cy5 coupling. Precision Unstained Ladder (Bio Rad) and PageRuler Prestained Ladder (Thermo Fisher Scientific) were used as molecular mass markers

although not directly utilised in this work. Lastly, a C-terminal cysteine (Cys) enabled covalent modification with Cy5. The ddFLN4 gene was PCR amplified from a synthetic template with primers containing the respective tag coding sequences. The construct was cloned into a modified pET28a vector (GE Healthcare Life Sciences, Little Chalfont, UK). The resulting fusion protein (Fgβ-ddFLN4-6xHis-ybbR-LPETGG-Cys) was expressed in *E.coli* Nico(DE3)-RIPL cells. A preculture of 5 ml LB containing 50 µg/ml Kanamycin was grown overnight at 37°C for 16 h. The preculture was then inoculated in 500 ml of ZYM-5052 autoinduction medium[4] containing 100 µg/ml Kanamycin and grown at 37°C for 20 h.

All purification steps were performed at 4°C or on ice when possible. Following expression, cells were separated from the medium by centrifugation at 500 x g for 20 min. Cells were then resuspended in His Lysis Buffer (30 mM Tris-HCl pH 7.8, 150 mM NaCl, 20 mM imidazole) and lysed by pulse sonication. The soluble fraction and insoluble fractions were separated by centrifugation at 20,000 x g for 45 min. The transfer construct was obtained in the soluble fraction and filtered with a 0.22 µm syringe filter. The filtered supernatant was purified by Ni-IMAC on a 5 ml HisTrap HP Ni-NTA column (GE Healthcare Life Sciences, Little Chalfont, UK) via step gradient elution from 20 mM to 250 mM imidazole (His Elution Buffer: 30 mM Tris-HCl pH 7.8, 150 mM NaCl, 250 mM imidazole) using an Äkta Start HPLC (GE LifeSciences, Little Chalfont, UK), producing a chromatogram with a single major peak (Figure S2).

Selected fractions from the major peak in the chromatogram were analysed by SDS-PAGE and Native PAGE. Samples were loaded to a Mini-PROTEAN TGX Stain-Free Precast Gel (Bio-

Rad Laboratories, CA, USA), which contains within its matrix a proprietary imaging molecule that is activated by exposure to UV light and then specifically labels tryptophan residues. While this imaging method is much faster than traditional coomassie staining, proteins that have no tryptophan residues (such as the ddFLN4 construct) do not produce a signal. Therefore, after first imaging with the stain-free method, the gel was additionally stained with coomassie blue. This has the advantage of enabling direct discrimination between the ddFLN4 construct and other co-eluting proteins. Gels were imaged with a ChemiDoc MP (Bio-Rad Laboratories) using stain-free imaging as well as and coomassie blue imaging. The images were overlaid using Image Lab software (Bio-Rad Laboratories).

Glycerol (10% v/v final concentration) was directly added to fractions 6-8 of purified protein. Reducing agents were omitted, as their presence would presumably interfere with cysteine-maleimide coupling later. The protein was finally stored at -80°C at a final concentration of about 700 µM as measured by the absorbance at 280 nm via NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA).

1.d. Cy5- and Biotin-Labelling of ddFLN4

The ddFLN4 transfer construct was modified first with Cy5 in a cysteine-maleimide reaction, followed by Biotin in an Sfp-catalysed transferase reaction. Cy5 Maleimide Mono-Reactive Dye (Mal-Cy5, Sigma Aldrich) was dissolved in DMSO to a stock concentration of 5 mM and stored at -20°C. The cysteine-maleimide reaction consisted of 7 nmol of purified ddFLN4 transfer construct protein and 50 nmol Mal-Cy5 in 1x Cysteine-Maleimide Reaction

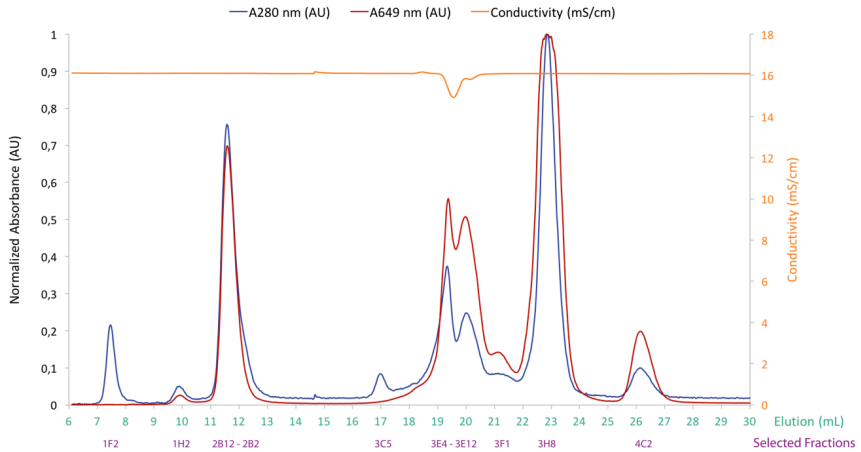


Figure S4. Chromatogram of Cy5- and biotin-labelled ddFLN4 transfer construct purification by size-exclusion chromatography. The absorbance at 280 nm and 649 nm is used to estimate the concentrations of protein and Cy5, respectively. Selected fractions (bottom purple labels) were further analysed by SDS-PAGE

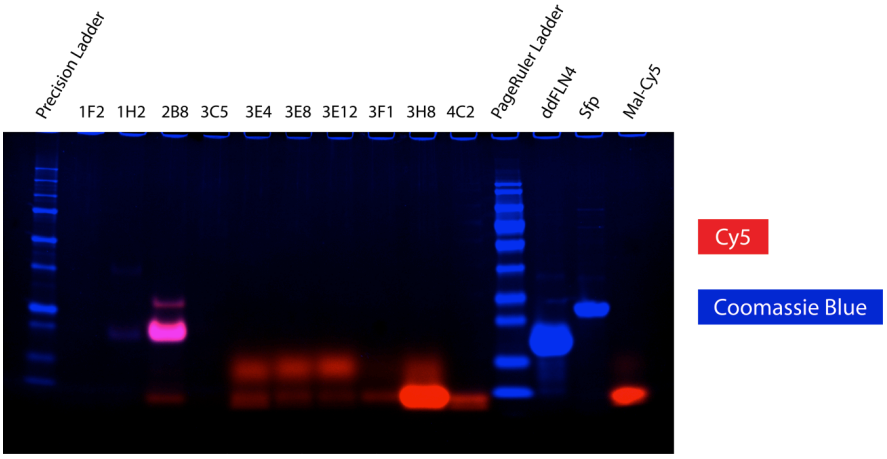


Figure S5. SDS-PAGE analysis of selected fractions of labelled ddFLN4 transfer construct purification. In addition to fractions from size-exclusion chromatography, samples of unreacted ddFLN4, Sfp and Mal-Cy5 were included as controls. The major peak spanning fractions 2B12-2B2 contained a high concentration of Cy5-labelled ddFLN4 as well as a small amount of co-eluted Cy5-labelled Sfp by-product.

Buffer (30 mM Tris-HCl pH 7.2, 150 mM NaCl) in a total volume of 40 µl at room temperature. The reaction was incubated at room temperature for 1 h followed by overnight incubation at 4°C. Subsequently, 9 nmol CoA-Biotin (Sigma Aldrich) and 1 nmol Sfp transferase were added to the reaction volume. Sfp Buffer Reaction Buffer was added to a 1x concentration (120 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, 2% Glycerol, 2 mM DTT) to give a final total volume of 100 µl. The reaction was incubated at 37°C for 1 h and then overnight at 4°C.

In order to isolate the dual-labelled ddFLN4 transfer construct, the reaction volume was purified by size-exclusion chromatography using an Äkta Explorer HPLC (GE LifeSciences, Little Chalfont, UK). A Superdex 75 Increase 10/300 GL column (GE Healthcare Life Sciences, Little Chalfont, UK) was first equilibrated with Size Exclusion Buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 10% v/v glycerol). The unpurified protein was loaded to the column and eluted in Size Exclusion Buffer. Chromatograms of the absorbance at 280 nm and 649 nm were collected during purification (Figure S4), and fractions were collected in 100 µl increments.

Fractions from the most prominent peaks were analysed via SDS-PAGE to determine which peak contained Cy5-labelled ddFLN4 transfer construct (Figure S4). Gels were imaged with a ChemiDoc MP (Bio-Rad Laboratories) using Epi-red LED excitation and 695/55 nm emission filter to detect Cy5 and coomassie blue imaging for protein detection. The images were overlaid using Image Lab software (Bio-Rad Laboratories). The major peak was further analysed to identify the optimal fractions (Figure S5). Fractions from the major peak as well as secondary peaks were also assessed for Biotin labelling via Native PAGE (Figure S6).

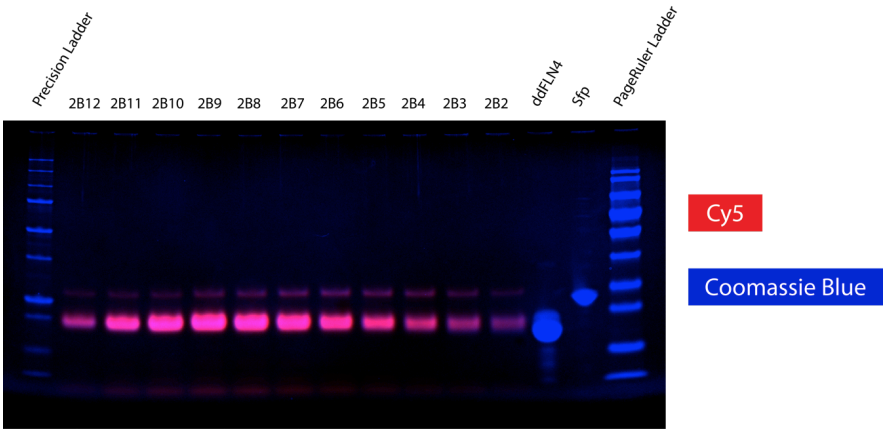


Figure S6. SDS-PAGE analysis of selected fractions from the major peak of labelled ddFLN4 transfer construct purification. In addition to fractions from size-exclusion chromatography, samples of unreacted ddFLN4 and Sfp were included as controls. The fractions contain a high concentration of labelled ddFLN4 as well as a lesser amount of co-eluted Cy5-labelled Sfp.

Selected fractions that demonstrated efficient labelling with both Cy5 and Biotin were pooled (2B12-2B10; 2B9-2B7; 2B6-2B3) and stored at -80°C at a final concentration of about 1 µM as measured by the absorbance at 649 nm via NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA).

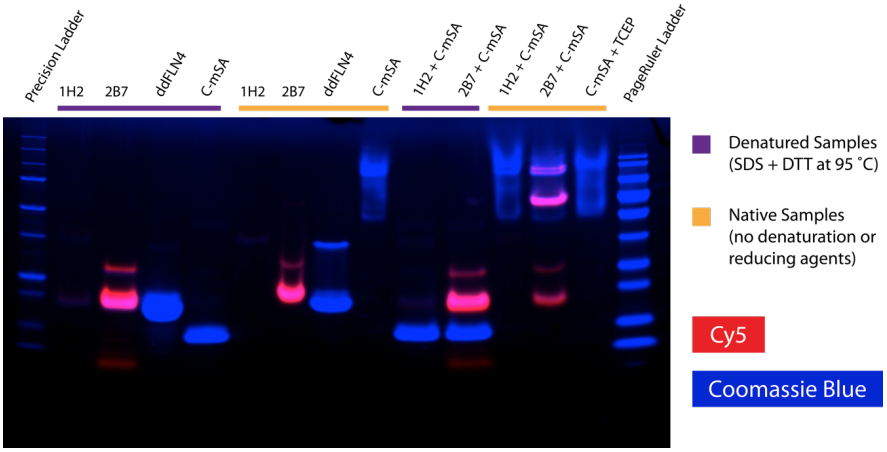


Figure S7. SDS-PAGE and Native PAGE analysis of selected fractions from labelled ddFLN4 transfer construct purification and binding to C-mSA. Fraction 2B7 from the major peak of Cy5-labelled ddFLN4 binds C-mSA under native conditions, visible by a band-shift in both Cy5-labelled ddFLN4 and C-mSA. This strongly suggests that at least the majority of Cy5-labelled ddFLN4 is also biotin-labelled.

2. Surface preparation

2.a. Preparation of Cantilevers

Cantilevers (BL-AC40TS, BioLever mini, Olympus, Japan) were oxidised in a UVOH 150 LAB UV-ozone cleaner (FHR Anlagenbau GmbH, Ottendorf-Okrilla, Germany). Silanisation was accomplished by incubating the cantilevers in (3-Aminopropyl)dimethylethoxysilane (ABCR, Karlsruhe, Germany, 50% v/v in Ethanol) for 2 min. Cantilevers were washed in toluene, then in isopropanol, and finally in ultrapure water and finally baked at 80°C for 45 min. For 30 minutes, the silanised cantilevers were placed in 25 mM heterobifunctional polyethylene glycol crosslinkers of 5,000 Da molecular weight (Rapp Polymere, Tübingen, Germany) dissolved in 50 mM HEPES at pH 7.5. The amines on the cantilevers reacted with the N-hydroxy succinimide on the one end of the crosslinkers. Using ultrapure water unreacted crosslinkers were washed off, before the cantilevers were placed in 1 mM Coenzyme A dissolved in coupling buffer (50 mM sodium phosphate, 50 mM sodium chloride, 10 mM EDTA, pH 7.2) for one hour. The maleimide on the other end of the PEG crosslinker and the thiol of the Coenzyme A formed a stable thioester bond. Unreacted Coenzyme A was washed off by ultrapure water. For several hours, the Coenzyme A-coated cantilevers were incubated with an Sfp-reaction mix containing 85 µM ybbR-SdrG, 3 µM Sfp Synthase, 10 mM magnesium chloride and 50 mM HEPES at pH 7.5. Sfp Synthase covalently joins Coenzyme A on the surface and the ybbR-tagged proteins.[5] The functionalised cantilevers were washed and stored in phosphate buffered saline.

2.b. Preparation of Glass Surfaces

Glass cover slips were sonicated in 50% (v/v) 2-propanol in filtered H_2O for 15 min and oxidised in a solution of 50% (v/v) hydrogen peroxide (30%) and sulfuric acid for 30 min. They were then washed in ddH_2O, dried in a nitrogen stream and then silanised by incubating for 1 h in (3-Aminopropyl)dimethylethoxysilane (ABCR, Karlsruhe, Germany, 1.8% v/v in Ethanol). The silanised surfaces were incubated in sodium borate buffer (150 mM, pH 8.5) for 30 min in order to deprotonate primary amine groups.

A PDMS microfluidic system – based on the system described by Kufer *et al.* [6] – was fixed on the aminosilanised glass and bonded briefly at 60°C for 10 min. The depot and target channels were incubated with a solution of a heterobifunctional PEG crosslinker [7, 8] with N-hydroxy succinimide and maleimide groups (molecular weight 5,000 Da, Rapp Polymere, Tübingen, Germany) dissolved to 30 mM in 100 mM HEPES pH 8.0 for 20 min. Unbound PEG was flushed from the channels with filtered H_2O.

Concurrently with assembling and functionalizing the microfluidics channels, Monovalent Streptavidin was reduced for covalent attachment to maleimide. Streptavidin with a reactive cysteine at the N-terminus (N-mSA) and at the C-terminus (C-mSA) was incubated in 5 mM TCEP at room temperature for 1 h, followed by buffer-exchange to PBS via Zeba

Spin Deslating Columns, 7K MWCO (Thermo Fisher Scientific, Waltham, MA, USA). Freshly reduced mSA was immediately applied to the PEG-functionalised microfluidic system, with N-mSA in the Depot channel and C-mSA in the Target channel.

N- and C-mSA were incubated in the channels for 1 h. Both channels were then flushed with filtered PBS to remove unbound mSA. The channels were then flushed with 0.1 mg/ml filtered BSA and 0.05% TWEEN20 in PBS to passivate the surface and discourage nonspecific adsorption. The labelled ddFLN4 transfer construct was diluted to an approximate concentration of 1 nM in PBS with 0.05 mg/ml BSA and 0.01% TWEEN20 and incubated in the depot channel for 1 h. The depot channel was then extensively flushed with PBS to clear the solution and remove unbound or non-specifically bound ddFLN4. The microfluidic system was then removed and the surface submerged in PBS.

3. Experimental Procedures

The experiments described in the manuscript were performed on an AFM/TIRFM hybrid, the details of which may be found in Gumpp *et al.*[9] This supporting information specifies methods, materials and additional data that are relevant for the conduction of the measurements discussed in the main text.

3.a. AFM Measurements

Measurements employed a custom-built AFM head and an Asylum Research MFP3D controller (Asylum Research, Santa Barbara, USA), which provides ADC and DAC channels as well as a DSP board for setting up feedback loops. Software for the automated control of the AFM head and xy-piezos during the force spectroscopy measurements was programmed in Igor Pro (Wave Metrics, Lake Oswego, USA). BioLever Mini (Olympus, Tokyo, Japan) cantilevers were chemically modified (see Preparation of Cantilevers) and calibrated in solution using the equipartition theorem.[10, 11] Pulling velocities were set to 3200 nm/s in the depot and 200 nm/s in the target area. The positioning feedback accuracy is ±3 nm. However, long-term deviations may arise due to thermal drift. Typical times for one Cut & Paste cycle amount to approximately 3 s in these experiments.

3.b. TIRF Microscopy

The fluorescence microscope of the hybrid instrument excites the sample through the objective in total internal reflection mode. A Nikon Apochromat 100x NA1.49 oil immersion objective (CFI Apochromat TIRF, Nikon, Japan) was employed. Laser excitation was achieved with a fiber-coupled Toptica iChrome MLE-LFA four-colour laser (Toptica Photonics, Gräfelfing, Germany), which is capable of emitting light at 405 nm, 488 nm, 561 nm and 640 nm through one single fiber mode. Specifically, red excitation at 640 nm with an

estimated intensity of approximately 10 W/cm^2 was utilised to monitor the Cy5 fluorescence. Emitted light from the sample was separated from the laser light with a Chroma quad line zt405/488/561/640rpc TIRF dichroic mirror (Chroma, Bellows Falls, VT, USA) and focused with a 20 cm tube lens. Separation of different emission wavelengths for simultaneous multicolour imaging was achieved by a Cairn Research Optosplit III (Cairn Research, Faversham, UK). Images were recorded with a back-illuminated Andor iXon DV860 DCS-BV EMCCD camera (Andor, Belfast, Ireland) in frame transfer mode with 1 MHz readout rate at a frame rate of 10 Hz. The camera was cooled and operated at -80°C. Fluorescent images were evaluated and processed with the analysis software ImageJ.

3.c. SMC&P Experiment

The rocket pattern was written in 442 transfer cycles with 200 nm spacing between each deposition point. The retraction veolcity in the depot was set to 3,200 nm/s and in the target to 200 nm/s. This corresponds to approximate surface contact times[12] (dependent on approach/retraction velocity, indentation force and substrate stiffness) of 5 ms and 80 ms, respectively, sufficient for ligand binding. Considering a single N-SdrG molecule being bound to the cantilever tip and estimating its localisation in a half sphere with r = 30 nm (approximate length of PEG5000 linker), the local concentration of SdrG would be in the μM range. This is several orders of magnitude higher than the measured K_d for the SdrG:Fgβ interaction (about 400 nM)[13] and the mSA:biotin interaction (<1 nM).[1] Taking further into account that bond formation is not diffusion-limited for the SMC&P experiment, successful attachment is very likely even at the given, short contact times.

4. Supplementary References

1. S. M. Sedlak, M. S. Bauer, C. Kluger, L. C. Schendel, L. F. Milles, D. A. Pippig and H. E. Gaub, *PLOS ONE*, 2017, **12**, e0188722.
2. M. Howarth, D. J. Chinnapen, K. Gerrow, P. C. Dorrestein, M. R. Grandy, N. L. Kelleher, A. El-Husseini and A. Y. Ting, *Nat Methods*, 2006, **3**, 267-273.
3. L. F. Milles, K. Schulten, H. E. Gaub and R. C. Bernardi, *Science*, 2018, **359**, 1527-1533.
4. F. W. Studier, *Protein Expression and Purification*, 2005, **41**, 207-234.
5. J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter and C. T. Walsh, *Proc Natl Acad Sci U S A*, 2005, **102**, 15815-15820.
6. S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl and H. E. Gaub, *Science*, 2008, **319**, 594-596.
7. J. L. Zimmermann, T. Nicolaus, G. Neuert and K. Blank, *Nat Protoc*, 2010, **5**, 975-985.
8. E. Celik and V. T. Moy, *J Mol Recognit*, 2012, **25**, 53-56.
9. H. Gumpp, S. W. Stahl, M. Strackharn, E. M. Puchner and H. E. Gaub, *Rev Sci Instrum*, 2009, **80**, 063704.

10. E. L. Florin, M. Rief, H. Lehmann, M. Ludwig, C. Dornmair, V. T. Moy and H. E. Gaub, *Biosensors and Bioelectronics*, 1995, **10**, 895-901.
11. H. J. Butt and M. Jaschke, *Nanotechnology*, 1995, **6**, 1.
12. S. Guo, N. Lad, C. Ray and B. B. Akhremitchev, *Biophys J*, 2009, **96**, 3412-3422.
13. K. Ponnuraj, M. G. Bowden, S. Davis, S. Gurusiddappa, D. Moore, D. Choe, Y. Xu, M. Hook and S. V. L. Narayana, *Cell*, 2003, **115**, 217-228.

4.2 Manuscript M4: Modular, Ultra-stable and Highly Parallel Protein Force Spectroscopy in Magnetic Tweezers Using Peptide Linkers

In Manuscript M4, the successful development of a protocol is described to reliably clamp single protein domains between a magnetic bead and a glass surface within a flow cell to perform force spectroscopy using magnetic tweezers. The confinement of the system of interest within a flow cell entails some challenges with respect to surface preparation: Thoroughly rinsing the sample between different steps of the immobilization protocol is hampered by the enclosed design. In particular, unspecific sticking of micrometer-sized superparamagnetic beads to PEG-linkers on the surface has been identified to cause severe problems.

To overcome these challenges, a protocol involving elastin-like polypeptide (ELP)-linkers, detergents added to buffer solutions, and the attachment of biotinylated proteins by Sortase-mediated linking to the sample before the addition of SA-coated beads has been developed. Its applicability to two different systems, namely the well-characterized ddFLN4 domain and the von Willebrand factor, a large glycoprotein involved in hemostasis, is demonstrated by successful force spectroscopy measurements. The used magnetic tweezers setup allows to track more than 100 beads in parallel so that large statistics are obtained in a single measurement. The SA/biotin interaction is well suited for tethering the beads to the molecule of interest, as there are many ways to reliably attach the small biotin molecule covalently to proteins. In addition, the high stability and the low off-rate of the SA/biotin system allow for stable long-term measurements – the feasibility of a week-long measurement has been demonstrated.

However, using commercial SA-coated beads, the survival fraction of tethered beads cannot be fitted with a single exponential. Instead, at least two different off-rates are observed. Using beads decorated with mSA tethered site-specifically, the lifetime of the beads can be fitted by a single off-rate. In addition, the average lifetime of the mSA-coated beads is more than four times higher compared with commercial SA-coated beads. This shows that a defined tethering geometry is advantageous for single-molecule force spectroscopy.

Modular, ultra-stable and highly parallel protein force spectroscopy in magnetic tweezers using peptide linkers

by

Achim Löf, Philipp U. Walker, Steffen M. Sedlak, Tobias Obser, Maria A. Brehm, Martin Benoit, and Jan Lipfert

submitted to

bioRxiv 491977; doi: https://doi.org/10.1101/491977

Modular, ultra-stable, and highly parallel protein force spectroscopy in magnetic tweezers using peptide linkers

Achim Löf [1], Philipp U. Walker [1], Steffen M. Sedlak [1], Tobias Obser [2], Maria A. Brehm [2], Martin Benoit [1,], and Jan Lipfert [1,*]*

[1] Department of Physics and Center for NanoScience, LMU Munich, 80799 Munich, Germany
[2] Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

[*]**To whom correspondence should be addressed:**

Martin Benoit, Department of Physics and Center for NanoScience, LMU Munich, Amalienstrasse 54, 80799 Munich, Germany, Tel: +49 89 2180 3133, Fax: +49 89 2180 2050, e-mail: martin.benoit@physik.uni-muenchen.de

or

Jan Lipfert, Department of Physics and Center for NanoScience, LMU Munich**,** Amalienstrasse 54, 80799 Munich, Germany, Tel: +49 89 2180 2005, Fax: +49 89 2180 2050, e-mail: jan.lipfert@lmu.de

Abstract

Single-molecule force spectroscopy has provided unprecedented insights into protein folding, force-regulation, and function. Here, we present a modular magnetic tweezers force spectroscopy approach that uses elastin-like polypeptide linkers to provide a high yield of protein tethers. Our approach extends protein force spectroscopy into the low force (<1 pN) regime and enables ultra-stable measurements on many molecules in parallel. We present (un-)folding data for the single protein domain ddFLN4 and for the large multi-domain dimeric protein von Willebrand factor (VWF) that is critically involved in primary hemostasis. The measurements reveal exponential force-dependencies of unfolding and refolding rates, directly resolve the stabilization of the VWF A2 domain by Ca^{2+}, and discover transitions in the VWF C-domain stem at low forces that likely constitute the first steps of VWF activation *in vivo*. Our modular attachment approach will enable precise and multiplexed force spectroscopy measurements for a wide range of proteins in the physiologically relevant force regime.

Introduction

Mechanical forces acting on proteins or ligand-receptor pairs are an integral part of many biological processes. Often the physiological function of proteins is critically regulated by force: examples include the mechano-activation of enzymes, force-regulated exposure of cryptic binding sites, and force-dependent unfolding and refolding of protein domains as "strain absorbers" to dissipate mechanical stress (*1*, *2*). A well-studied protein "strain absorber" is the fourth domain of the F-actin crosslinking filamin rod of *Dictyostelium discoideum* (ddFLN4), which exhibits extraordinarily fast refolding, facilitated by an intermediate state along the refolding pathway (*3*, *4*). Another remarkable example of physiological force regulation occurs in the large, multimeric glycoprotein von Willebrand factor (VWF) in the vasculature. VWF's hemostatic function is regulated by increased hydrodynamic forces occurring upon blood vessel injury. Activation of VWF relies on a complex interplay of force-induced conformational changes both of single domains and of the large-scale protein conformation (*5*–*7*), while down-regulation of VWF is based on mechano-enzymatic cleavage at a cryptic binding site only accessible upon unfolding of VWF's A2 domain (*8*). While many of the individual transitions in VWF have been probed in detail, the overall picture of how full-length VWF reacts to external forces in the blood stream remains incomplete. Since hydrodynamic peak forces grow as the square of the contour length (*5*, *8*), transitions that release contour length at low forces are expected to be particularly relevant for VWF's physiological function as they will set of a cascade of increasing forces that trigger additional transitions with further contour length release. Recent work using AFM imaging has suggested transitions in the VWF C-domain stem that, however, could not be detected in AFM-based force spectroscopy, due to its limited force resolution (*9*, *10*).

Most insights into the mechanical properties and regulation of proteins and their complexes at the single-molecule level have been obtained from force spectroscopy experiments using atomic force microscopes (AFM) or optical tweezers (OT). While AFM and OT force

spectroscopy measurements have provided unprecedented insights, they also have important shortcomings (*11*). AFM measurements cannot resolve forces below ~10 pN; OT provide excellent spatio-temporal resolution even for forces down to ~1 pN, but are not capable of measuring many molecules in parallel (*11*). In addition, both AFM and OT intrinsically control position and not force, such that constant-force measurements require active feedback.

Magnetic tweezers (MT) are a single-molecule force spectroscopy technique that can overcome these shortcomings. In MT, molecules of interest are tethered between a surface and superparamagnetic beads (*11–13*) (**Supplementary Fig. S1**). External magnetic fields exert precisely controlled forces (*14*) in the range of ~0.01-100 pN and previous work has demonstrated camera-based tracking for ~10s-100s of nucleic acid-tethered beads simultaneously (*15–17*) in (*x*,*y*,*z*) with ~nm-spatial resolution, and, recently, also up to ≤ ms-temporal resolution (*18–20*). MT naturally operate in constant force mode, i.e. the applied force is constant during the measurement (to within 0.01%; **Supplementary Fig. S2**), as long as the external magnetic field is not actively changed, with excellent sensitivity in particular at low forces. In addition, MT enable long-term, stable, and robust measurements and do not suffer from heating or photo-damage (*11*).

Despite these advantages, MT so far have mainly been employed to investigate nucleic acid tethers. A key challenge in applying MT to protein force spectroscopy remains to tether ~nm-sized proteins between much larger, ~µm-sized beads and the surface, while avoiding unspecific surface interactions and ideally with a large number of usable tethers in each field of view. Previous MT studies on proteins, therefore, mostly employed large protein constructs, often polyproteins with repeats of e.g. titin Ig or protein L domains (*21–24*). Current strategies for attaching proteins to the surface in MT are either based on antibodies (*25–29*) or His-tag Cu^{2+}-NTA chemistry (*30*, *31*), or on covalent linkage, either of Halo-tag fusion proteins to a surface coated with Halo-tag amine ligands (*21–23*, *32*) or using the Spy-tag-SpyCatcher system (*24*, *33*). Non-covalent attachment has the disadvantage of limited force stability compared to covalent attachment. Attachment via fusion proteins without a specific linker potentially complicates the analysis due to unfolding and refolding transitions of the proteins used for attachment (e.g. the Halo-tag (*21*)) or as spacers (e.g. flanking titin Ig-domains (*24*, *33*)). In addition, attachment via fusion proteins appears to suffer from a low number of usable tethers, as so far there are no reports of multiplexed protein unfolding and refolding measurements.

Here, we present a versatile, modular protein attachment strategy for single-molecule MT force spectroscopy. Our tethering protocol uses an elastin-like polypeptide (ELP) linker (*34*) that ensures efficient attachment to the surface while minimizing unspecific interactions, both critical prerequisites for high-throughput parallel measurements. In our approach, the protein of interest requires only short (1 and 11 amino acids [aa]) peptide tags for coupling to the linker and bead, respectively, avoiding the need for large fusion proteins and providing a general attachment strategy that is independent of protein size. We demonstrate the versatility of our attachment strategy by applying it to a small, single protein domain, ddFLN4 (100 aa), and a very large, multi-domain protein, dimeric full-length VWF (≈4000 aa). For both proteins, we achieve a high yield of specific tethers, i.e. a large number of single-molecule tethers that exhibit characteristic unfolding and refolding signatures and can be measured in parallel in a single field of view. Our highly-parallel ultra-stable measurements of repeated unfolding and refolding resolve outstanding questions about the respective folding pathways

and stabilities. In addition, we leverage the ability of our assay to apply constant forces over extended periods of time to many molecules in parallel to probe the stability of the biotin-streptavidin receptor-ligand system. We anticipate our tethering strategy to be applicable to a wide range of proteins, and, furthermore, expect it to be of immediate use for other parallel force spectroscopy techniques, such as single-molecule centrifugation (*35*, *36*) or acoustic force spectroscopy (*37*), extending their capabilities towards multiplexed protein force spectroscopy.

Results

Site-specific and efficient tethering of proteins with elastin-like polypeptide linkers

Our attachment strategy uses an unstructured ELP linker (*34*, *38*) with a contour length of ≈120 nm and functional groups at its termini that we utilize as spacer for immobilizing the protein of interest to the bottom glass slide of the flow cell and to reduce unspecific protein-surface (*34*) and bead-surface interactions (**Fig. 1A**). The ELP linker is attached to a glass slide functionalized with thiol-reactive maleimide groups via an N-terminal cysteine (see **Methods** for details of the coupling protocol). The ELP linker carries a C-terminal *LPETGG* motif that allows for site-specific and covalent ligation to the protein of interest via an N-terminal glycine residue in a reaction catalyzed (*39*) by the enzyme sortase A. For coupling to the bead, the protein of interest is further engineered to carry an 11-aa ybbR-tag (*40*) at its C-terminus that is covalently attached to coenzyme A–biotin in the sfp phosphopantetheinyl transferase reaction. Finally, the biotin-label forms a high-affinity non-covalent bond to streptavidin-functionalized beads. Our approach requires only short peptide tags on the protein of interest that can readily be introduced by standard molecular cloning methods and have been shown to be compatible with expression and folding of a large range of proteins (*39–43*).

Here, we apply our tethering protocol to two very different protein systems: the small ddFLN4 domain and large full-length dimeric VWF. We obtained comparable and efficient tethering of beads with a large number of specific, single-molecule tethers in both cases. Typically, in a single field of view (≈440 × 330 μm^2) of our MT setup (**Fig. 1B**; see **Methods** and **Supplementary Figs. S1 and S2** for details on the setup) 50 to 100 tethered beads are tracked in parallel, of which 30 to 50 tethers exhibit characteristic unfolding and refolding signatures (**Fig. 1C**). The beads that do not show characteristic signatures are likely anchored to the surface by multiple protein tethers, since in control measurements without the protein of interest added, there is essentially no unspecific binding of beads to the surface (0-1 beads per field of view). The fraction of specific tethers attached via a single protein can be increased by decreasing the density of proteins immobilized on the surface. However, decreasing the protein concentration will also result in a decrease of the number of single-tethered beads. Optimizing our conditions, we achieved fractions of up to ≈60% specific, single-protein tethers, while still obtaining a large number of tethered beads at the same time. The most efficient flow cell exhibited 50 specific out of 85 beads within the single field of view measured.

Three-state folding and unfolding of ddFLN4

We first applied our tethering protocol to the Ig-fold ddFLN4 domain (**Fig. 2A**), which exhibits a characteristic three-state unfolding pattern that has been extensively studied in AFM experiments (*3*, *4*) and is routinely employed as a molecular fingerprint in AFM force spectroscopy experiments (*44–46*). To characterize unfolding and (re-)folding in our MT assay, we recorded time traces of tether extension under different, constant forces. In a typical measurement (**Fig. 2B)**, the force is increased from an initial low value (0.5 pN) that allows for (re-)folding, to a high value (25 pN in **Fig. 2B**) that promotes unfolding, and subsequently decreased to a moderate value (6.5 and 7.5 pN in **Fig. 2B**) to directly monitor refolding. Subsequently, this cycle is repeated multiple times with variable force levels to collect statistics. Unfolding and refolding of ddFLN4 were observed as clear double-steps in the traces, i.e., as an increase or decrease of the tether extension in two separate steps that we interpret as transition between the native (N) and intermediate (I) and between the intermediate and unfolded (U) states, respectively (**Fig. 2B**, insets). We analyzed the changes in extension for the transitions N↔I and I↔U as well as for the full transition N↔U for many different clamped forces (**Fig. 2C**). The resulting force–extension profiles are well described by fits of the worm-like chain (WLC) model with a fixed persistence length of 0.5 nm, in accordance with a previous AFM study (*4*), yielding contour length values (mean ± SD) of 15.0 ± 0.1 nm, 18.3 ± 0.1 nm, and 31.9 ± 0.2 nm, in excellent agreement with values reported from AFM (*3*, *4*).

Our data are fully consistent with previous work that found unfolding of the ddFLN4 domain to proceed via a mandatory, short-lived intermediate state: In a first unfolding step, strands A and B (42 aa; blue in **Fig. 2A**) detach and unfold, with strands C-G (58 aa; red in **Fig. 2A**) forming a less stable intermediate state(*3*), which quickly unfolds in the second unfolding step. Folding of ddFLN4 was also suggested to proceed via an intermediate state that is most likely structurally identical or very similar to the intermediate populated during unfolding (*4*). In our data set, data from unfolding (data points >8 pN) and refolding (data points ≤8 pN) are well described by a single WLC curve, confirming that the intermediate states populated during unfolding and folding are structurally very similar or identical. Importantly, no other features except the double-steps originating from ddFLN4 were observed in the force range probed (**Supplementary Fig. S3**), showing that the other components of our tethering strategy do not interfere with the measurements.

Our force clamp measurements allowed us to directly determine the rates of all transitions (**Methods, Supplementary Fig. S4**). For unfolding (**Fig. 2D**), we observed the rate for the first transition, N→I, to increase with increasing force from ≈ 0.001 s^{-1} at 9 pN to ≈ 0.2 s^{-1} at 31 pN. We fitted the rates to a single-barrier kinetic model, in which the rate is given by $k(F) = k_0 \exp(F \cdot \Delta x / k_B T)$, where F is the applied force, k_0 the rate at zero force, Δx the distance to the transition state, k_B the Boltzmann constant and T the temperature (*47*). We find $k_{0,N\rightarrow I}$ = 1.5 ×10^{-4} s^{-1} (1.3 to 1.8 ×10^{-4} s^{-1}) and $\Delta x_{N\rightarrow I}$ = 0.99 ± 0.03 nm (mean ± SD for all rates and Δx values). The measured rates for full unfolding N→U are essentially identical to those for the transition N→I (**Fig. 2D**), owing to the fact that the rates for the second transition, I→U ($k_{0,I\rightarrow U}$ = 0.25 s^{-1} [0.19 to 0.32 s^{-1}]; $\Delta x_{I\rightarrow U}$ = 0.32 ± 0.04 nm), are at least one order of magnitude faster (**Fig. 2D**), implying that the transition N→I is the rate-limiting step for

unfolding. The three-fold difference between $\Delta x_{N\rightarrow I}$ and $\Delta x_{I\rightarrow U}$ reflects that the difference between the rates N→I and I→U decreases with increasing force.

For refolding in the force range from 2 pN to 8 pN (**Fig. 2E**), the rates for the first substep U→I ($k_{0,U->I}$ = 0.52 s^{-1} [0.34 to 0.79 s^{-1}]; $\Delta x_{U->I}$ = 3.6 ± 0.3 nm) are again essentially identical to the rates for full refolding U→N (**Fig. 2E**) and rates for the second transition I→N (1500 s^{-1} [950 to 2500 s^{-1}]; $\Delta x_{I->N}$ of 3.5 ± 0.3 nm) are at least three orders of magnitude higher compared to the first transition, implying that again the first transition, U→I , is rate-limiting (**Fig. 2F**). For forces below 6 pN, the intermediate state was too short-lived to be reliably detected, so that rates were determined separately only for forces ≥ 6 pN.

We compare our force clamp results to the rates at zero force reported previously from fits of a single barrier kinetic model to constant pulling speed AFM measurements (*4*). For unfolding, the rates for the second step $k_{0,I->U}$ are in excellent agreement (0.25 and 0.33 s^{-1} in MT and AFM, respectively), yet the zero force rates for the first step $k_{0,N->I}$ appear, in contrast, to differ significantly (1.5 $\times 10^{-4}$ *vs.* 0.28 s^{-1}). However, in AFM measurements with extremely low pulling speeds of 1 nm/s a markedly higher mechanical strength of ddFLN4 has been observed and explained by switching to a second unfolding pathway (*48*). The reported zero-force rate (for full unfolding) from AFM of 1.4 $\times 10^{-4}$ s^{-1} is in excellent agreement with our results. Thus, our data support the existence of a second unfolding pathway and suggest that differences between the two pathways can be largely attributed to the first unfolding step N→I.

For refolding, a direct comparison is less straightforward, as refolding in AFM experiments has been measured at zero force and not under load (*4*). The rates obtained from MT and AFM differ significantly ($k_{0,U->I}$: 0.52 *vs.* 55 s^{-1}; $k_{0,I->N}$: 1500 *vs.* 179 s^{-1}), which might indicate different folding pathways in the presence and absence of force. Intriguingly, however, in both cases the same intermediate state appears to be populated during folding. Whereas the first step of folding –and thereby also full folding– is markedly slowed down by force, the second step of folding is almost 10-fold sped up, suggesting a pre-alignment of the structured portion of the intermediate state by force that allows for faster folding of strands A and B in the second folding step. Since ddFLN4 *in vivo* is positioned within actin-crosslinking filamin and under tensile load, it appears plausible that a force-induced pre-alignment of the intermediate state might play a physiological role.

Ultra-stable equilibrium measurements of ddFLN4 unfolding and refolding

By determining the force at which the fitted rates for full unfolding and refolding (**Fig. 2D,E**; black lines) intersect, we predicted the equilibrium force at which the probabilities of ddFLN4 being in the unfolded and folded states are equal to be approximately 7.3 pN (**Fig. 3A**). We tested this prediction by measuring at a constant force of 7.5 pN close to the predicted equilibrium force. Since the predicted rates at equilibrium are only ~3 h^{-1} (**Fig. 3A**), we performed very long measurements (up to 55 h; **Fig. 3B**), harnessing the excellent force and drift stability of MT. We observed repeated transitions between the unfolded and folded states, with the system spending approximately half of the time in each of the two states, as expected for a measurement close to equilibrium. Examining the traces close to equilibrium in detail, we observe repeated transitions not only N↔U via the I state (**Fig. 3C**, left and middle trace), but also from the U and N states into the I state that return to the initial state (**Fig. 3C**,

right trace), strongly suggesting that the same intermediate state is populated during unfolding and folding. Finally, we note that even for the very long measurements reported here, no significant change of ddFLN4's force response over time was observed, indicating reliable, correct refolding of the domain without any hysteresis effects, as verified for a ≈10 h-long measurement comprising 35 unfolding and refolding cycles and for a long equilibrium measurement at constant force (**Supplementary Fig. S5**). The long-term stability combined with its very characteristic three-state unfolding signature make ddFLN4 an ideal fingerprint for the identification of single-molecule tethers.

Lifetime of biotin-streptavidin interactions for multi- and monovalent streptavidin

Having established ddFLN4 as an ideal fingerprint for the identification of specific single-molecule tethers and having demonstrated the ability to apply constant forces over extended periods of time to multiple tethers in parallel, we utilized our assay to investigate not only protein folding and refolding, but also ligand-protein receptor interactions. As a proof-of-concept measurement, and to validate our tethering approach, we directly probed the stability of the high-affinity, non-covalent biotin-streptavidin interaction under constant force. Since all other linkages in our tethering protocol consist of mechanically stable covalent bonds, we used ddFLN4-tethered beads to apply different high forces (45-65 pN) to the biotin-streptavidin bond and monitored the time until bead rupture, only taking into account beads that showed the specific ddFLN4 unfolding signature in two short force plateaus of 25 pN at the beginning of the measurement. Importantly, the number of beads that ruptured already during these initial short plateaus was small (< 3.5%). For commercially available streptavidin-coated beads (Dynabeads M-270 Streptavidin, Invitrogen), we found the survival fraction to decay with time in a complex, multi-exponential fashion (**Fig. 4A**) for all forces probed, suggesting the existence of several populations of the biotin–streptavidin interaction with different lifetimes. To quantify the lifetimes involved, we fit the fastest and slowest decaying populations by linear regression to the logarithm of the first and last 20% of data points, respectively (lines in **Fig. 4A**). Over the studied force range, the lifetime of the fastest decaying population ranged from ≈100 s at 65 pN to ≈2100 s at 45 pN, whereas the lifetime of the slowest decaying population was ~50-fold higher, increasing from ≈5,000 s at 65 pN to ≈68,000 s at 45 pN (**Fig. 4C**). For both populations, the lifetime was found to increase exponentially with decreasing force (**Fig. 4C**). Already for a force of 20 pN, extrapolated lifetimes are well above a day and off-rates at zero force are in the range of 10^{-7} to 10^{-8} s^{-1}, consistent with the fact that beads remained stably bound for hours or days in our force spectroscopy measurements at forces ≤ 20 pN (**Fig. 2,3, and 5**).

We hypothesized that the different populations and multi-exponential lifetimes for commercially available streptavidin-functionalized beads originate from the biotin-streptavidin complex being loaded with force in different geometries that result from the tetravalency of streptavidin (*46*). Indeed, for measurements with custom-made beads functionalized with a monovalent version of streptavidin (*46*) in a well-defined geometry using a C-terminal tag (*49*), the survival fraction was well described by a single-exponential decay (**Fig. 4B**). We chose immobilization of the monovalent streptavidin construct via the C-terminus of its functional subunit, as it has been recently demonstrated in AFM force spectroscopy measurements that the monovalent streptavidin-biotin complex can withstand markedly higher forces when loaded with force from the C-terminus as compared to pulling

from the N-terminus (*49*). Indeed, we found the lifetime of the custom-made monovalent streptavidin beads to be 24,000 s (≈6.7 h) at 65 pN (**Fig. 4B**), similar to and even exceeding that of the 20% longest-lived commercially available beads (**Fig. 4C**).

Force clamp measurements on full-length VWF dimers

Having demonstrated our attachment approach on a small well-characterized protein, we next applied it to large (≈500 kDa) dimeric constructs of full-length VWF. Dimers, the smallest repeating subunits of VWF multimers, consist of two multi-domain monomers that are C-terminally linked via disulfide bonds and have a contour length of ≈130 nm between their N-termini (*5*, *50*) (**Fig. 5A**). Since different peptide tags at the two N-termini are required for attaching dimers in the desired pulling geometry (**Fig. 5A**), we genetically engineered heterodimers consisting of two different monomers that are N-terminally modified with a ybbR-tag or a sortase motif *GG*, respectively (**Methods**). After tethering in the MT, we recorded time traces of VWF dimers with alternating plateaus of high force (**Fig. 5B**, 6-20 pN) and moderate force (**Fig. 5B**, 2-5 pN). In most cases, we observed two unfolding and two refolding steps in the recorded high and moderate force traces, respectively, with extension values matching the expected values for unfolding of the A2 domains (≈180 aa each) that were previously probed in isolation in OT (*8*, *51*). Observation of domain (un-)folding only for the two A2 domains is consistent with the prediction that all domains of VWF except A2 are protected against unfolding by long-range disulfide bonds (*50*, *52*) and with the results of recent AFM studies (*9*, *10*).

In addition to the steps attributed to A2 unfolding and refolding, we less frequently also observed larger steps (**Supplementary Fig. S6;** 70-80 nm at ~11 pN), which we attribute to the dissociation of a strong intermonomer interaction mediated by the D4 domains that has recently been identified in AFM force measurements in approximately one-half of all VWF dimers under near-physiologic conditions (*9*, *10*). Consistent with their assignment to the D4-mediated intermonomer interaction, the large unfolding steps occur much less frequently in the absence of divalent ions, which have been shown to be critical for the intermonomer interaction (*9*, *10*), and are absent for mutant constructs lacking the D4 domain (delD4; **Supplementary Fig. S6,S7**). The dissociation of the intermonomer interaction was in some cases –after intermittent relaxation to a low force– observed repeatedly for the same molecule, implying reversibility of the interaction (**Supplementary Fig. S6**). Whereas in the constant pulling speed AFM measurements dissociation of this interaction had always occurred at much higher forces than –and therefore after– A2 unfolding, in our constant force measurements we observed dissociation of the D4-mediated intermonomer interaction in the same force range as A2 unfolding, suggesting a pronounced force-loading rate dependence for the intermonomer interaction. In fact, in the constant-force measurements we repeatedly observed dissociation of this interaction even before unfolding of one or both of the A2 domains (**Supplementary Fig. S6)**. Importantly, this implicates a likely important role of the D4-mediated intermonomer interaction for regulation of VWF's hemostatic activity at physiologically relevant forces in the bloodstream.

Calcium binding stabilizes the VWF A2 domain

We next used our assay to elucidate the controversial impact of calcium on A2 stability (*51*, *53*, *54*). We performed measurements both in buffer mimicking the physiological pH and salt

concentrations of the vasculature ('near-physiologic'; pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, and 1 mM $CaCl_2$) and in buffer lacking divalent ions and supplemented with 10 mM EDTA. First, we analyzed the change in extension upon A2 unfolding and refolding for different constant forces. For both buffer conditions, the resulting force–extension profiles (**Fig. 5C**), combining data from unfolding (data points >6.5 pN) and from refolding (data points <5 pN), are well described by a single WLC curve. The WLC fits yielded values for contour length and persistence length of 75.0 nm and 0.42 nm (95% CI: 70.8–79.2 nm and 0.37–0.46 nm) for near physiological-buffer, and of 68.5 nm and 0.50 nm (62.7–74.3 nm and 0.41–0.58 nm) for the EDTA buffer, and thus show no significant difference, indicating that calcium has no effect on the extension of the unfolded state. A WLC fit to the combined data from both buffer conditions (inset in **Fig. 5C**) yielded contour and persistence length values of 71.9 nm and 0.45 nm (68.3-75.4 nm and 0.41–0.50 nm). The contour length increments determined from the MT measurements on full-length dimeric VWF are in excellent agreement with OT unfolding studies on isolated A2 domains (*8*, *51*, *53*), suggesting that complete A2 unfolding is not obstructed by the presence of other domains. Control measurements using the same attachment protocol and ddFLN4 under the same buffer conditions found no difference in the force-response for the different buffer conditions (**Supplementary Fig. S8**).

Next, we studied the kinetics of A2 unfolding and refolding. In the case of unfolding, rates are approximately two- to fourfold higher for the EDTA buffer in the force range probed, 6.5-17 pN (**Fig. 5D**, circles). For both buffer conditions, rates increase exponentially with increasing force, with a slightly stronger dependence on force for the EDTA condition. Fitting a single-barrier kinetic model yielded values for the unfolding rate at zero force $k_{unf,0} = 3.6 \times 10^{-5}$ s^{-1} (1.8 to 7.1 $\times 10^{-5}$ s^{-1}) and 7.8 $\times 10^{-5}$ s^{-1} (5.1 to 12 $\times 10^{-5}$ s^{-1}) and distances to the transition state $\Delta x_{unf} = 2.45 \pm 0.22$ nm and 2.60 ± 0.15 nm in the presence and absence of Ca^{2+}, respectively. The rates measured in our constant force assay are two orders of magnitude slower than the rates determined in near-physiological buffer in OT measurements on isolated A2 domains. While in principle this difference might indicate stabilization of A2 by neighboring domains, we deem it likely that it at least partially results from the transformation of rupture force distributions measured in the OT using the Dudko-Hummer-Szabo method (*55*), which is sensitive to the elastic response of employed flexible linkers. In our traces we did not observe any features that correspond to dissociation of potential interactions of A2 with neighboring domains. For refolding against external forces of 2-5 pN, rates are approximately two- to sixfold higher in the presence of Ca^{2+} (**Fig. 5D**, triangles) and decrease exponentially with force, with a more pronounced force dependence in the presence of Ca^{2+}, which is reflected by the higher value of Δx of 6.80 ± 0.56 nm compared to 4.73 ± 0.26 nm in the absence of Ca^{2+}. The refolding rate at zero force in the presence of Ca^{2+} $k_{ref,0} = 5.1$ s^{-1} (2.9 to 8.7 s^{-1}) is 20-fold higher than in the absence of Ca^{2+}, $k_{ref,0} = 0.23$ s^{-1} (0.18 to 0.28 s^{-1}), indicating that calcium substantially speeds up folding of A2.

Taken together, our results demonstrate that A2 is stabilized by the presence of Ca^{2+} by increasing the refolding rate and stabilizing against unfolding compared to the conditions without Ca^{2+}. The observed increases in the refolding rates in our experiments are in quantitative agreement with a previous report using OT on isolated A2 domains (*51*). Importantly, we directly observe refolding under mechanical load even in the absence of Ca^{2+} (**Supplementary Fig. S9**), in contrast to a previous study (*53*). The role of Ca^{2+} in the stabilization against unfolding is controversial: We observe a modest reduction in the

unfolding rate by Ca^{2+}, which is consistent with the low-force data found in one OT study (*51*), which, however, reported no statistically significant change in the unfolding rate with and without Ca^{2+} overall, possibly as their assay might have lacked the sensitivity to resolve small differences. In contrast, we find no evidence for a long-lived intermediate in the unfolding pathway in the presence of Ca^{2+} that was reported by another study using OT (*53*). Finally, we occasionally observed tethers that only showed the unfolding and refolding signal of one A2 domain (**Supplementary Fig. S10**). In such tethers, refolding of one A2 domain may be inhibited due to cis-trans isomerization of a cis-proline, as reported in a previous OT study (*8*).

Transitions in the VWF stem at low forces

AFM imaging (*9*, *10*) and electron microscopy (*56*) suggest that the VWF stem consisting of six C-domains can open and close in a zipper-like fashion (**Fig. 6A,B**). However, transitions of the VWF stem have not been observed directly. To probe for interactions in the VWF stem, we subjected dimeric VWF tethers that had shown the characteristic A2 unfolding pattern to low constant forces. At forces of ~1 pN, we observed repeated, reversible transitions with a maximum contour length increase of ~50-60 nm that is consistent with fully unzipping and rezipping the VWF stem (**Fig. 6C**). Increasing force in the range 0.6-1.4 pN systematically shifted the population towards higher tether extensions, which we interpret as less compact "unzipped" conformations of the VWF stem. Importantly, such transitions were never observed for ddFLN4-tethered beads. In addition, the large change in extension at very low forces makes it appear highly unlikely that the observed transitions originate from domain unfolding events (for comparison, a tether extension of < 5 nm would be expected for unfolding of an A2 domain at a force of ~1 pN, due to the WLC stretching behavior of the unfolded protein chain).

The observed transitions occurred between multiple levels and featured more than two distinct states, as expected from the observation by AFM imaging that multiple C-domain interactions can contribute to the opening and closing of the stem. However, a reliable assignment of the transitions will require additional analysis and measurements under a range of solution conditions and using VWF mutants. Using high-speed bead tracking (1 kHz; **Fig. 6D**), we observed transitions to occur on time scales ≤ 1 s. Our observations are consistent with the predicted (*9*) occurrence of unzipping transition in the VWF stem at forces of ~1 pN, close to the value for the onset of VWF multimer elongation predicted by Brownian hydrodynamics simulations (*57*). Molecular interactions that break and release contour length at such low forces are expected to be particularly relevant for VWF's physiological function as these are likely the first interactions to open under shear flow and to set off a cascade of increased contour length and increased force, since hydrodynamic peak forces grow as the square of the contour length (*5*, *8*).

Discussion

We have introduced a novel approach for single-molecule force spectroscopy measurements on proteins using MT. Our protocol enables multiplexed measurements through high-yield,

ultra-stable tethering of proteins requiring only short peptide tags. As a proof-of-concept measurement, we probed the three-state unfolding and refolding of ddFLN4. Our measurements at constant force are overall in excellent agreement with constant pulling speed AFM experiments and confirm the existence of a low-loading rate pathway for unfolding (*4*, *48*). Applying our method to the large, force-regulated protein VWF confirms several structural transitions previously observed in AFM and OT force-spectroscopy measurements and reveals how Ca^{2+} stabilizes the A2 domain. In addition, our measurements reveal transitions in the VWF stem at low forces of ≈1 pN, which likely constitute critical first steps in the stretch response of VWF under physiological shear flow.

Using the ability of our assay to apply constant forces over long periods of time to multiple tethers in parallel, we probed the stability of streptavidin-biotin bonds, a widely used ligand-receptor system. We found that commercially available, multivalent streptavidin (far) exceeds the requirements of typical constant force force-spectroscopy measurements (with lifetimes ≥ 1 day at forces ≤ 20 pN), but has a multi-exponential lifetime distribution. Monovalent, site-specifically attached streptavidin, in contrast, exhibits a single-exponential lifetime distribution with extremely high force stability, making it an attractive approach for force spectroscopy on systems that require high forces over extended periods of time. Ultimately, one could also replace the biotin-streptavidin bond with a covalent linkage to even further enhance the force and also chemical stability of the attachment protocol.

In conclusion, we have demonstrated the versatility and power of a new approach for single-molecule protein force spectroscopy measurements using MT. Our method provides a high yield of ultra-stable specific single-molecule tethers that can be probed in parallel at constant forces over extended periods of time. Given ongoing improvements in camera technology, we expect that the number of protein tethers that can be measured in parallel will further increase by at least an order of magnitude in the near future. In addition, we anticipate that our tethering strategy will enable multiplexed protein force spectroscopy using other single-molecule methods such as acoustic and centrifugal force spectroscopy as well. Since the approach is modular and only requires minimal modifications to the protein of interest, we anticipate it to be applicable to a wide range of proteins. We expect MT force spectroscopy to in particular give access to the physiologically relevant low force (< 1 pN) regime and to provide a wealth of novel insights into the mechanics and force-regulation of proteins.

Materials and Methods

Preparation of ddFLN4 constructs

Recombinant ddFLN4 expressed in *E.coli* (with the internal cysteine at position 18 mutated to serine) was a kind gift from Lukas Milles (LMU Munich). At its C-terminus, the ddFLN4 construct possesses a polyhistidine-tag for purification and a ybbR-tag (*40*). At its N-terminus, the construct has a short linker sequence (*MGTGSGSGSGSAGTGSG*) with the terminal methionine followed by a single glycine. Due to efficient cleavage of the methionine by *E.coli* methionine aminopeptidases, the glycine is expected to be available for sortase-catalyzed ligation.

The ddFLN4 gene was synthesized codon-optimized for expression in *E.coli* as a linear DNA fragment (GeneArt – ThermoFisher Scientific, Regensburg, Germany), and inserted into pET28a Vectors via Gibson assembly (*58*) (New England Biolabs, Frankfurt, Germany). Protein expression in *E.coli NiCo21 (DE3)* (New England Biolabs) and purification via the polyhistidine-tag were carried out as previously described in detail(*45*).

Preparation of hetero-bifunctional VWF dimer constructs

For preparation of hetero-bifunctional VWF dimers two different types of monomers were co-expressed, which at their N-termini –subsequent to a required signal peptide– possess either a ybbR-tag (*40*) or an N-terminal strep-tag II for purification (*59*), followed by a tobacco etch virus (TEV) protease cleavage site (*60*) and the sortase motif *GG* (*39*). The TEV site serves two purposes: first, to remove the strep-tag after purification, as it might otherwise interact with Streptavidin on the magnetic beads during measurements, and second, to free the sortase motif *GG*, which must be located terminally for the sortase reaction. Both monomer constructs lack the VWF pro-peptide (domains D1 and D2) in order to abolish linkage of dimers into larger multimers. For delD4 dimers, additionally the D4 domain is deleted in both monomers. For AFM images shown in Fig. 5, dimeric VWF constructs consisting of two identical monomers, possessing a Strep-tag at their N-termini, were used.

Plasmid construction was carried out analogously to a procedure previously described(*9*). For expression, $2{\cdot}10^6$ HEK 293 cells in a 75 cm^2 flask (DSMZ, Braunschweig, Germany) were transfected in Dulbecco's modified Eagle's medium (Life Technologies, Darmstadt, Germany) containing 10% fetal bovine serum (Life Technologies), 2 µg of each of the two plasmids, and 15 µl Lipofectamine 2000 (Life Technologies). 24 h after transfection, cells were transferred into selection medium containing 500 µg/ml G418 (Invivogen, Toulouse, France) and 250 µg/ml Hygromycin B (Invivogen). After 2–3 weeks, the polyclonal cell culture was seeded for expression. After 72 h of cell growth, the medium was exchanged against OPTIPRO-SFM (Life Technologies) for serum-free collection of secreted recombinant VWF. The culture supernatant was collected after 72 h and concentrated using Amicon Ultra-15 MWCO 100 kDa (Merck, Schwalbach, Germany).

Dimeric constructs were purified via a HiTrap StrepTrap affinity chromatography column (GE Healthcare) using the AEKTA Explorer system (GE Healthcare). As running buffer, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.4, was used. Elution buffer additionally contained 2.5 mM d-desthiobiotin. Eluates were buffer exchanged (to the running buffer) and concentrated by centrifuge filtration using Amicon Ultra MWCO 100 kDa (Merck Millipore). All purified VWF dimers were further inspected by AFM imaging and showed no structural differences as compared to dimeric VWF constructs with different peptide tags or without tags used in previous studies (*9*, *10*).

Preparation of ELP linkers

Recombinant ELP linkers expressed in *E.coli NiCo21 (DE3)* were a kind gift from Wolfgang Ott (LMU Munich). The ≈300 aa ELP linker with a contour length of ≈120 nm used in this study has the sequence $[(VPGEG)\text{-}(VPGVG)_4\text{-}(VPGAG)_2\text{-}(VPGGG)_2\text{-}(VPGEG)]_6$ and possesses a single N-terminal cysteine and the C-terminal sortase recognition motif *LPETGG*. Cloning, expression and purification have been described (*34*, *38*), and can be performed

using standard procedures for the production of recombinant proteins. Plasmids are provided at Addgene by Ott et al. (Addgene accession number 90472 for the ELP linker used in this study).

Attachment chemistry and flow cell preparation

Functionalization of glass slides with the ELP linkers described above followed the protocol by Ott et al. (*34*). Glass slides were first silanized with 3-(aminopropyl)dimethylethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany), and then coated with 10 mM of a sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate cross-linker with a negligible contour length of 0.83 nm (Sulfo-SMCC, Thermo Fisher Scientific Inc.), dissolved in 50 mM HEPES, pH 7.5. Subsequently, ELP linkers were linked to the thiol-reactive maleimide groups via the single cysteine at their N-terminus in coupling buffer consisting of 50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH 7.2. Afterwards, 10 mM L-cysteine dissolved in coupling buffer were added to saturate potentially remaining unreacted maleimide groups. Finally, non-magnetic polystyrene beads (Polybead Microspheres 3 µm; Polysciences GmbH, Hirschberg, Germany) dissolved in ethanol were baked onto the slides at ≈70 °C for ≈5 min for use as reference beads. After each step, slides were extensively rinsed with ultrapure water. Flow cells were assembled from an ELP-functionalized cover slip as the bottom surface and a non-functionalized cover slip with two small holes for inlet and outlet as the top, with a layer of cut-out parafilm (Pechiney Plastic Packaging Inc., Chicago, IL) as a spacer to form a (~4 mm wide and 50 mm long) flow channel. Flow cells were assembled by heating on a hot plate to ≈70 °C for ≈2 min. Assembled flow cells can be stored under ambient conditions for weeks.

Prior to experiments, the flow cells were incubated with 1% casein solution (Sigma-Aldrich) for 1 h and afterwards flushed with 1 ml (approximately 20 flow cell volumes) of buffer (20 mM HEPES, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.4). CoA-biotin (New England Biolabs) was coupled to the ybbR-tag on the protein of interest in a bulk reaction in the presence of 5 µM sfp phosphopantetheinyl transferase and 10 mM $MgCl_2$ at 37 °C for 60 min. In the case of VWF, subsequently TEV protease was added to a final concentration of approximately 25 µM and incubated for 30-60 min. Dithiothreitol (DTT) present in the storage buffer of TEV protease was removed beforehand using desalting columns (Zeba Spin 40 K MWCO, Thermo Scientific Inc.). Afterwards, protein was diluted to a final concentration of approximately 10 nM (VWF dimers) or 25 nM (ddFLN4) in 20 mM HEPES, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.4, and incubated in the flow cell in the presence of 1-2 µM sortase A for 30 min. Subsequently, the flow cell was flushed with 1 ml of buffer.

Magnetic beads –either Dynabeads M-270 streptavidin (Invitrogen) or beads functionalized with monovalent streptavidin (see below)– in measurement buffer containing 0.1% (v/v) Tween-20 (Sigma-Aldrich) were incubated in the flow cell for 60 s, and unbound beads were flushed out with 2 ml of measurement buffer. All measurements were performed at room temperature (≈22 °C).

Starting with silanized glass slides, complete flow cell preparation takes less than 7 h. In addition, flow cells functionalized with ELP linkers, but not yet incubated with casein and protein, can be prepared in advance and stored at room temperature for weeks without loss of

functionality. Starting with ELP-functionalized flow cells, measurements can be started within 120 min.

Preparation of monovalent streptavidin

Tetrameric, but monovalent streptavidin (mSA) consisting of three mutant subunits deficient in biotin binding and one functional subunit, possessing at its C-terminus a polyhistidine-tag for purification and a single cysteine for site-specific immobilization, was prepared as described in detail by Sedlak et al. (*46*, *49*). In brief, functional and mutant subunits were cloned into pET vectors (Novagen, EMD Millipore, Billerica, USA) and separately expressed in *E.coli BL21(DE3)-CodonPlus* (Agilent Technologies, Santa Clara, USA). Resulting inclusion bodies were solubilized in 6 M guanidine hydrochloride. Functional and mutant subunits were then mixed at a 1:10 ratio prior to refolding and purification via the polyhistidine-tag, in order to ensure a 1:3 ratio of functional to non-functional subunits in the final tetrameric streptavidin construct.

Site-specific, covalent immobilization of monovalent streptavidin on magnetic beads

Magnetic beads with surface amine groups (Dynabeads M-270 Amine, Invitrogen; these beads are otherwise identical to Dynabeads M-270 Streptavidin) were functionalized with 25 mM of 5-kDa NHS–polyethylene glycol (PEG)–maleimide linkers with reactive NHS and maleimide end groups (Rapp Polymere, Tübingen, Germany) in 50 mM HEPES, pH 7.5, and afterwards extensively washed first with DMSO and then with water. The mSA constructs possessing a single cysteine as described above were reduced with 5 mM TCEP bond breaker solution (Thermo Fisher) and afterwards buffer exchanged to coupling buffer using desalting columns (Zeba Spin 40 K MWCO, Thermo Scientific Inc.). Beads were then incubated with mSA in coupling buffer for 90 min and extensively washed with measurement buffer.

Buffers

All measurements on ddFLN4 and measurements on VWF dimers under 'near-physiologic' conditions were performed in buffer containing 20 mM HEPES, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1% Tween-20, at pH 7.4. Measurements without calcium were performed in EDTA buffer containing 20 mM HEPES, 150 mM NaCl, 10 mM EDTA, 0.1% Tween-20, at pH 7.4. Before measurements in the absence of calcium, the flow cell was incubated with EDTA buffer for 2 h. Control measurements at acidic pH were performed in 20 mM sodium-acetate, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1% Tween-20, at pH 5.5.

Magnetic tweezers setup

Measurements were performed on a custom MT setup described by Walker et al. (*61*). A schematic and an image of the setup are given in **Supplementary Fig. S1.** The setup uses a pair of permanent magnets (5×5×5 mm^3 each; W-05-N50-G, Supermagnete, Switzerland) in vertical configuration (*14*). The distance between magnets and flow cell (and, therefore, the force; **Supplementary Fig. S2**) is controlled by a DC-motor (M-126.PD2; PI Physikinstrumente, Germany). For illumination, an LED (69647, Lumitronix LED Technik GmbH, Germany) is used. Using a 40x oil immersion objective (UPLFLN 40x, Olympus, Japan) and a CMOS sensor camera with 4096×3072 pixels (12M Falcon2, Teledyne Dalsa,

Canada), a large field of view of approximately 440 × 330 μm^2 can be imaged at a frame rate of 58 Hz. For measurements with an acquisition rate of 1 kHz, a reduced field of view of 1792 × 280 pixels was used. Images are transferred to a frame grabber (PCIe 1433; National Instruments, Austin, TX) and analyzed with an open-source tracking software (*15*). The bead tracking accuracy of our setup was determined to be ≈0.6 nm in (*x, y*) and ≈1.5 nm in *z* direction, as determined by tracking non-magnetic polystyrene beads, with a diameter comparable to M270 beads (3 μm), after baking them onto the flow cell surface. For creating the look-up table required for tracking the bead positions in *z*, the objective is mounted on a piezo stage (Pifoc P-726.1CD, PI Physikinstrumente). Force calibration was performed as described by te Velthuis et al. (*62*) based on the fluctuations of long DNA tethers. The final force calibration, i.e. the dependence of the force applied to a bead on the distance between magnets and flow cell, is shown in **Supplementary Fig. S2**, together with an example trace showing the DNA B-S overstretching transition at the expected force of ≈65 pN. Importantly, for the small extension changes on the length scales of our protein tethers, the force stays constant to very good approximation, with the relative change in force due to tether stretching or protein (un-)folding being $< 10^{-4}$ (**Supplementary Fig. S2**). We verified the uniformity of the magnetic field across the field of view and found the change in force across the full range of the field of view to be < 3% (**Supplementary Fig. S2**). The largest source of force uncertainty is the bead-to-bead variation, which we found to be on the order of ≤ 10% for the beads used in this study (**Supplementary Fig. S2**), in line with several previous reports (*14*, *63*, *64*).

AFM imaging

For AFM imaging, a dimeric VWF construct possessing a strep-tag at both N-termini was used. Preparation of substrates for AFM imaging was performed as recently described (*9*, *10*). In brief, 5 μg/ml of VWF dimers in near-physiologic buffer were incubated on a poly-L-lysine-coated mica substrate for 30 s, which was subsequently rinsed with water and finally dried in a gentle stream of nitrogen. AFM images of 1 μm x 1 μm and 1024 x 1024 pixels were recorded in tapping mode in air, using an MFP-3D AFM (Asylum Research, Santa Barbara, CA) and cantilevers with silicon tips (AC160TS, Olympus, Japan), possessing a nominal spring constant of 26 N/m and a resonance frequency of approximately 300 kHz. Raw image data were processed using SPIP software (v6.5.1; Image Metrology, Denmark). Image processing involved plane correction (third order polynomial plane-fitting), line-wise flattening (according to the histogram alignment routine), and Gaussian smoothing.

Data analysis

All data analysis was carried out using custom-written Matlab scripts (Matlab v.R2015b; The MathWorks Inc., Natick, MA) incorporated into a custom Matlab GUI. We obtained tether extension *vs.* time by subtracting the *z*-position of the reference bead from the *z*-position of the protein-tethered bead. All traces shown and analyzed are the raw extension *vs.* time traces recorded at 58 Hz, used without any filtering or smoothing. For ddFLN4 measurements, only beads that in unfolding force plateaus repeatedly showed a double-step with a short-lived intermediate state were taken into account for further analysis. Similarly, for VWF measurements, only beads repeatedly exhibiting two steps of equal height corresponding to unfolding of the A2 domains in unfolding force plateaus were analyzed, unless otherwise

noted. Unfolding and refolding behavior for ddFLN4 and VWF under the different reported buffer conditions were observed in at least 3 independently prepared flow cells in all cases.

To determine the position of steps, we employed the step-finding algorithm by Kerssemakers et al. (*65*), and the corresponding change in extension was determined as the difference between the average extensions of the adjacent 1000 frames recorded before and after the step, respectively (fewer frames were used if the 1000-frame interval contained another step). Extensions of folding and unfolding (sub)steps were histogrammed for each clamped force (1 nm binning for ddFLN4, and 3 nm and 2 nm binning for VWF A2 unfolding and refolding, respectively), and fitted with Gaussians. Error bars in figures report the FWHM of the fits, divided by the square root of the respective counts. The resulting force–extension profiles were fitted to the WLC model of polymer elasticity (an approximation to this model with less than 1% relative error was used for fitting (*66*)). In the case of ddFLN4, a fixed persistence length of 0.5 nm was used to enable direct comparison with results from an AFM study by Schwaiger et al. (*4*). In the case of VWF A2, both persistence length and contour length were free fit parameters.

To determine the unfolding or refolding rates $k(F)$ at a given constant force F, the respective fraction of observed unfolding or refolding events as a function of time was fitted with the exponential expression $1 - a \exp(-kt) + b$ (**Supplementary Fig. S4**), where the free parameters a and b can compensate for events that were missed due to the finite measurement time or due to the finite time of motor movement when setting the force. However, such missed events were rare and parameters a and b were close to 1 and 0, respectively. Error bars on rates in figures indicate 95% confidence bounds of fits. In the case of VWF, only events corresponding to steps with extensions ≤ 60 nm were taken into account to ensure that only A2 unfolding events –and not dissociation of the D4-mediated intermonomer interaction (see **Supplementary Fig. S6**)– are analyzed.

The force dependence of unfolding and refolding rates was described by a single barrier kinetic model: $k(F) = k_0 \exp(F\Delta x/k_BT)$, with the rate at zero force k_0 and the distance to the transition state Δx as fit parameters. Fitting was carried out as linear fits to the natural logarithm of the data. Error margins for k_0 and Δx given in the text correspond to 1 SD.

For bead rupture measurements, lifetimes at different constant forces were determined from the survival fraction *vs*. time data based on > 35 rupture events for each condition. In the case of mSA-beads, data were described by a single-exponential decay, and the corresponding lifetime was determined by a linear fit to the natural logarithm of the data. In the case of the more complex decay behavior observed for commercial streptavidin-coated beads, lifetimes for the fastest- and slowest-decaying populations were estimated by linear fits to the natural logarithm of the first and last 20% of data points, respectively. The dependence of estimated lifetimes on force was again described by the single barrier kinetic model introduced above.

Acknowledgements

General: We thank Wolfgang Ott and Lukas F. Milles for providing ELP and ddFLN4 constructs, respectively, Gesa König and Thomas Nicolaus for laboratory assistance, and Ellis Durner, Reinhard Schneppenheim, and Hermann E. Gaub for helpful discussions.

Funding: This study was supported by research funding from the German Research Foundation (DFG) to the projects "Unraveling the Mechano-Regulation of Von Willebrand Factor" (project number 386143268), "Shear Flow Regulation of Hemostasis - Bridging the Gap between Nanomechanics and Clinical Presentation" (SHENC, project FOR1543), SFB 863 project A11, and the Nanosystems Initiative Munich.

Author contributions: A.L., P.U.W., S.M.S., M.A.B., M.B. and J.L. designed experiments. T.O. designed and prepared VWF constructs. S.M.S. prepared mSA constructs. A.L. prepared samples. A.L., P.U.W., and S.M.S. performed experiments. P.U.W. wrote analysis software. A.L. and P.U.W. analyzed results. A.L. prepared figures. A.L., M.B. and J.L. wrote the manuscript. All authors discussed the data and critically reviewed the manuscript.

Competing interests: The authors declare no competing interests.

Data and materials availability: All data presented in the main and supplementary figures are available from the authors upon request. The employed MT setup control and bead-tracking software, developed by Cnossen et al. (*15*), is open-source and available from http://www.github.com/jcnossen/BeadTracker. Custom-written Matlab software for analysis of magnetic tweezers data is available from the authors upon request. Plasmids for the ELP linkers used in this study are provided at Addgene by Ott et al. (Addgene accession number 90472).

References

1. E. M. Puchner, H. E. Gaub, Single-Molecule Mechanoenzymatics. *Annu. Rev. Biophys.* **41**, 497–518 (2012).
2. V. Vogel, Mechanotransduction involving multimodular proteins: converting force into biochemical signals. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 459–88 (2006).
3. I. Schwaiger, A. Kardinal, M. Schleicher, A. A. Noegel, M. Rief, A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat. Struct. Mol. Biol.* **11**, 81–5 (2004).
4. I. Schwaiger, M. Schleicher, A. A. Noegel, M. Rief, The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep.* **6**, 46–51 (2005).
5. T. A. Springer, Von Willebrand factor, Jedi knight of the bloodstream. *Blood*. **124**, 1412–1426 (2014).
6. A. Löf, J. P. Müller, M. A. Brehm, A biophysical view on von Willebrand factor activation. *J. Cell. Physiol.* **233**, 799–810 (2018).
7. H. Fu *et al.*, Flow-induced elongation of von Willebrand factor precedes tension-dependent activation. *Nat. Commun.* **8**, 324 (2017).
8. X. Zhang, K. Halvorsen, C.-Z. Zhang, W. P. Wong, T. A. Springer, Mechanoenzymatic Cleavage of the Ultralarge Vascular Protein von Willebrand Factor. *Science*. **324**, 1330–1334 (2009).
9. J. P. Müller *et al.*, Force Sensing by the Vascular Protein von Willebrand Factor is Tuned by a Strong Intermonomer Interaction. *Proc. Natl. Acad. Sci.* **113**, 1208–1213 (2016).
10. J. P. Müller *et al.*, pH-dependent interactions in dimers govern the mechanics and structure of von Willebrand Factor. *Biophys. J.* **111**, 312–322 (2016).
11. K. C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods*. **5**, 491–505 (2008).
12. S. B. Smith, L. Finzi, C. Bustamante, Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*. **258**, 1122–6 (1992).
13. T. R. Strick, J. F. Allemand, D. Bensimon, A. Bensimon, V. Croquette, The elasticity of a

single supercoiled DNA molecule. *Science*. **271**, 1835–7 (1996).
14. J. Lipfert, X. Hao, N. H. Dekker, Quantitative modeling and optimization of magnetic tweezers. *Biophys. J.* **96**, 5040–9 (2009).
15. J. P. Cnossen, D. Dulin, N. H. Dekker, An optimized software framework for real-time, high-throughput tracking of spherical beads. *Rev. Sci. Instrum.* **85**, 103712 (2014).
16. N. Ribeck, O. A. Saleh, Multiplexed single-molecule measurements with magnetic tweezers. *Rev. Sci. Instrum.* **79**, 094301 (2008).
17. I. De Vlaminck *et al.*, Highly Parallel Magnetic Tweezers by Targeted DNA Tethering. *Nano Lett.* **11**, 5489–5493 (2011).
18. B. M. Lansdorp, S. J. Tabrizi, A. Dittmore, O. A. Saleh, A high-speed magnetic tweezer beyond 10,000 frames per second. *Rev. Sci. Instrum.* **84**, 044301 (2013).
19. D. Dulin *et al.*, High Spatiotemporal-Resolution Magnetic Tweezers: Calibration and Applications for DNA Dynamics. *Biophys. J.* **109**, 2113–25 (2015).
20. A. Huhle *et al.*, Camera-based three-dimensional real-time particle tracking at kHz rates and Ångström accuracy. *Nat. Commun.* **6**, 5885 (2015).
21. H. Chen *et al.*, Dynamics of Equilibrium Folding and Unfolding Transitions of Titin Immunoglobulin Domain under Constant Forces. *J. Am. Chem. Soc.* **137**, 3540–3546 (2015).
22. S. Haldar, R. Tapia-Rojo, E. C. Eckels, J. Valle-Orero, J. M. Fernandez, Trigger factor chaperone acts as a mechanical foldase. *Nat. Commun.* **8**, 668 (2017).
23. I. Popa *et al.*, A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J. Am. Chem. Soc.* **138**, 10546–10553 (2016).
24. D. Spadaro *et al.*, Tension-Dependent Stretching Activates ZO-1 to Control the Junctional Localization of Its Interactors. *Curr. Biol.* **27**, 3783–3795.e8 (2017).
25. A. S. Adhikari, E. Glassey, A. R. Dunn, Conformational dynamics accompanying the proteolytic degradation of trimeric collagen I by collagenases. *J. Am. Chem. Soc.* **134**, 13259–65 (2012).
26. X. J. A. Janssen *et al.*, Torsion stiffness of a protein pair determined by magnetic particles. *Biophys. J.* **100**, 2262–7 (2011).
27. E. Pérez-Ruiz *et al.*, Probing the Force-Induced Dissociation of Aptamer-Protein Complexes. *Anal. Chem.* **86**, 3084–3091 (2014).
28. A. S. Adhikari, J. Chai, A. R. Dunn, Mechanical Load Induces a 100-Fold Increase in the Rate of Collagen Proteolysis by MMP-1. *J. Am. Chem. Soc.* **133**, 1686–1689 (2011).
29. A. van Reenen, F. Gutiérrez-Mejía, L. J. van IJzendoorn, M. W. J. Prins, Torsion Profiling of Proteins Using Magnetic Particles. *Biophys. J.* **104**, 1073–1080 (2013).
30. H. Chen *et al.*, Mechanical perturbation of filamin A immunoglobulin repeats 20-21 reveals potential non-equilibrium mechanochemical partner binding function. *Sci. Rep.* **3**, 1642 (2013).
31. M. Yao *et al.*, Force-dependent conformational switch of α-catenin controls vinculin binding. *Nat. Commun.* **5**, 4525 (2014).
32. M. Yao *et al.*, The mechanical response of talin. *Nat. Commun.* **7**, 11966 (2016).
33. S. Le *et al.*, Mechanotransmission and Mechanosensing of Human alpha-Actinin 1. *Cell Rep.* **21**, 2714–2723 (2017).
34. W. Ott *et al.*, Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano*. **11**, 6346–6354 (2017).
35. K. Halvorsen, W. P. Wong, Massively parallel single-molecule manipulation using centrifugal force. *Biophys. J.* **98**, L53-5 (2010).
36. D. Yang, A. Ward, K. Halvorsen, W. P. Wong, Multiplexed single-molecule force spectroscopy using a centrifuge. *Nat. Commun.* **7**, 11026 (2016).
37. G. Sitters *et al.*, Acoustic force spectroscopy. *Nat. Methods*. **12**, 47–50 (2015).
38. W. Ott, T. Nicolaus, H. E. Gaub, M. A. Nash, Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation. *Biomacromolecules*. **17**, 1330–1338 (2016).
39. C. S. Theile *et al.*, Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1800–7 (2013).
40. J. Yin *et al.*, Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci.* **102**, 15815–15820 (2005).
41. M. W.-L. Popp, H. L. Ploegh, Making and Breaking Peptide Bonds: Protein Engineering Using

Sortase. *Angew. Chemie Int. Ed.* **50**, 5024–5032 (2011).
42. W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, M. A. Nash, Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. *J. Struct. Biol.* **197**, 3–12 (2017).
43. E. Durner, W. Ott, M. A. Nash, H. E. Gaub, Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. *ACS Omega*. **2**, 3064–3069 (2017).
44. L. F. Milles, K. Schulten, H. E. Gaub, R. C. Bernardi, Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science*. **359**, 1527–1533 (2018).
45. L. F. Milles, E. A. Bayer, M. A. Nash, H. E. Gaub, Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen *Clostridium perfringens*. *J. Phys. Chem. B*. **121**, 3620–3625 (2017).
46. S. M. Sedlak *et al.*, Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. *PLoS One*. **12**, e0188722 (2017).
47. E. Evans, K. Ritchie, Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72**, 1541–1555 (1997).
48. Z. T. Yew, M. Schlierf, M. Rief, E. Paci, Direct evidence of the multidimensionality of the free-energy landscapes of proteins revealed by mechanical probes. *Phys. Rev. E*. **81**, 031923 (2010).
49. S. M. Sedlak *et al.*, *Nano Lett.*, in press, doi:10.1021/acs.nanolett.8b04045.
50. Y. F. Zhou *et al.*, Sequence and structure relationships within von Willebrand factor. *Blood*. **120**, 449–458 (2012).
51. A. J. Xu, T. A. Springer, Calcium stabilizes the von Willebrand factor A2 domain by promoting refolding. *Proc. Natl. Acad. Sci.* **109**, 3742–3747 (2012).
52. C. Baldauf *et al.*, Shear-induced unfolding activates von Willebrand factor A2 domain for proteolysis. *J. Thromb. Haemost.* **7**, 2096–2105 (2009).
53. A. J. Jakobi, A. Mashaghi, S. J. Tans, E. G. Huizinga, Calcium modulates force sensing by the von Willebrand factor A2 domain. *Nat. Commun.* **2**, 385 (2011).
54. M. Zhou *et al.*, A novel calcium-binding site of von Willebrand factor A2 domain regulates its cleavage by ADAMTS13. *Blood*. **117**, 4623–31 (2011).
55. O. K. Dudko, G. Hummer, A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci.* **105**, 15755–15760 (2008).
56. Y.-F. Zhou *et al.*, A pH-regulated dimeric bouquet in the structure of von Willebrand factor. *EMBO J.* **30**, 4098–111 (2011).
57. S. Lippok *et al.*, Shear-Induced Unfolding and Enzymatic Cleavage of Full-Length VWF Multimers. *Biophys. J.* **110**, 545–554 (2016).
58. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*. **6**, 343–345 (2009).
59. T. G. Schmidt, A. Skerra, The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2**, 1528–1535 (2007).
60. J. Phan *et al.*, Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**, 50564–72 (2002).
61. P. U. Walker, W. Vanderlinden, J. Lipfert, Dynamics and energy landscape of DNA plectoneme nucleation. *Phys. Rev. E*. **98**, 042412 (2018).
62. A. J. W. Te Velthuis, J. W. J. Kerssemakers, J. Lipfert, N. H. Dekker, Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. *Biophys. J.* **99**, 1292–302 (2010).
63. I. De Vlaminck, T. Henighan, M. T. J. van Loenhout, D. R. Burnham, C. Dekker, Magnetic Forces and DNA Mechanics in Multiplexed Magnetic Tweezers. *PLoS One*. **7**, e41432 (2012).
64. E. Ostrofet, F. S. Papini, D. Dulin, Correction-free force calibration for magnetic tweezers experiments. *Sci. Rep.* **8**, 15920 (2018).
65. J. W. J. Kerssemakers *et al.*, Assembly dynamics of microtubules at molecular resolution. *Nature*. **442**, 709–712 (2006).
66. R. Petrosyan, Improved approximations for some polymer extension models. *Rheol. Acta*. **56**, 21–26 (2017).
67. P. Fucini, C. Renner, C. Herberhold, A. A. Noegel, T. A. Holak, The repeating segments of the F-actin cross-linking gelation factor (ABP-120) have an immunoglobulin-like fold. *Nat. Struct.*

Biol. **4**, 223–30 (1997).

68. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–8, 27–8 (1996).

Figures

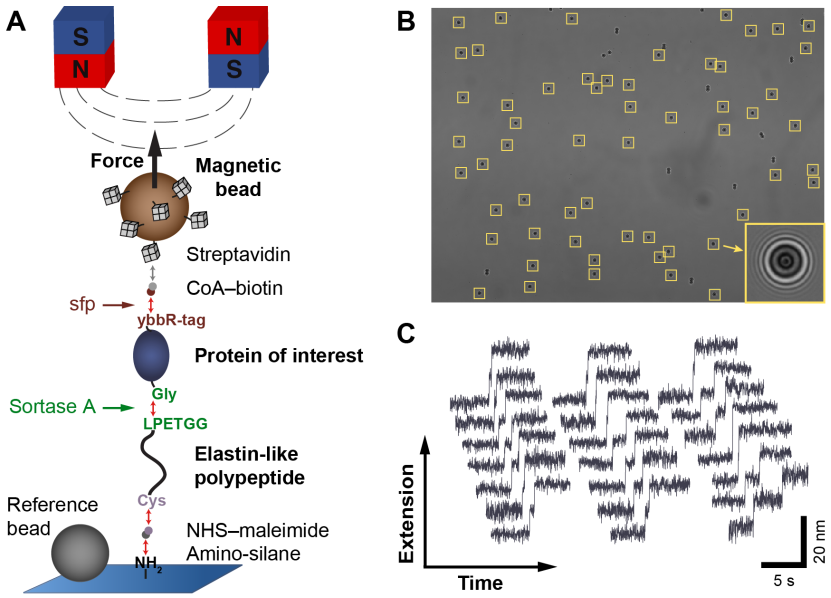


Fig. 1. Attachment protocol for highly-parallel force spectroscopy on proteins in magnetic tweezers. (A) Schematic of the strategy for tethering a protein of interest between the bottom glass slide of the flow cell and a magnetic bead (not to scale). An ELP linker with a single cysteine at its N-terminus is coupled to the amino-silanized glass slide via a small-molecule NHS–maleimide cross-linker. After covalent coupling of coenzyme A (CoA)–biotin to the ybbr-tag at the C-terminus of the protein in a bulk reaction catalyzed by sfp phosphopantetheinyl transferase, the protein is covalently ligated to the ELP linker via one (or more) glycines at its N-terminus in a reaction mediated by sortase A, which selectively recognizes the C-terminal *LPETGG* motif of the ELP linker. Finally, a streptavidin-coated magnetic bead is bound to the biotinylated protein via the high-affinity biotin–streptavidin interaction. Red and grey double arrows indicate covalent and non-covalent bonds, respectively. Forces are exerted on the magnetic bead by permanent magnets positioned above the flow cell. Non-magnetic polystyrene beads baked onto the surface are used as reference beads for drift correction. (**B)** Representative field of view. Yellow boxes indicate approximately 60 beads marked for tracking. The enlarged image of one bead shows the diffraction ring pattern used for 3D bead tracking. (**C)** Example tether extension time traces showing the characteristic three-state unfolding pattern of ddFLN4. All traces shown were recorded in parallel from different beads within the same field of view at a constant force of 21 pN.

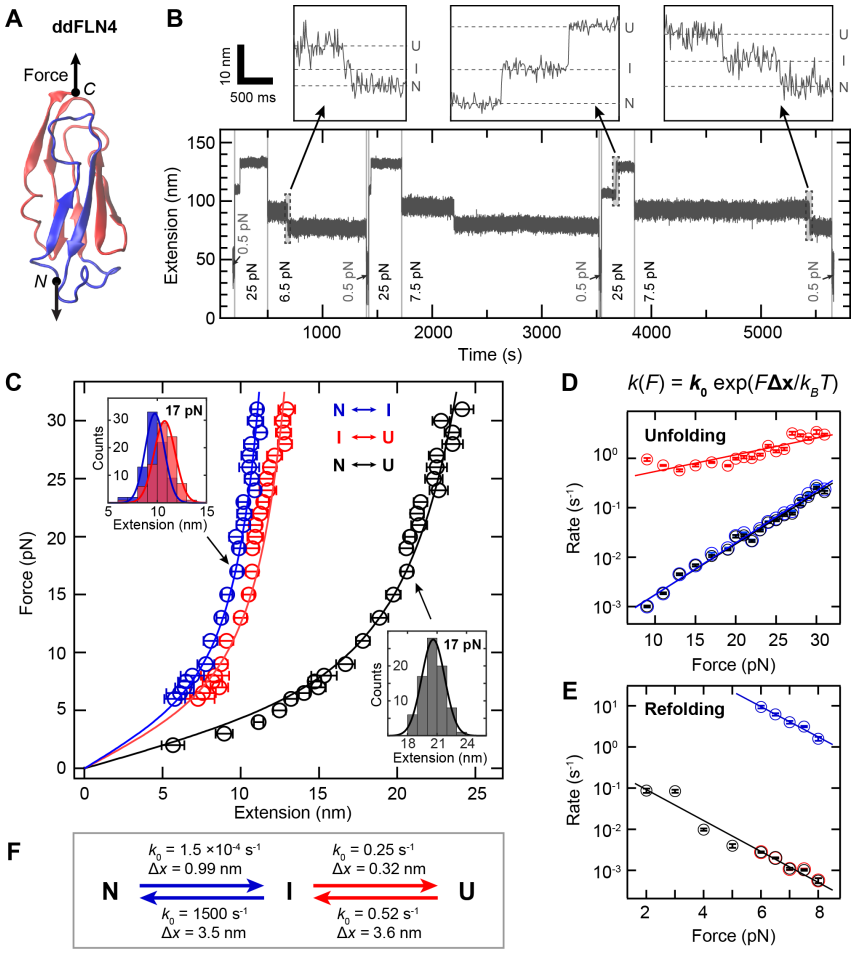


Fig. 2. Three-state folding and unfolding of ddFLN4 at constant force
(A) Structure of ddFLN4 (PDB: 1KSR(*67*), rendered using VMD(*68*)), with strands A and B rendered in blue and strands C-G, forming the structured portion of the intermediate state, in red. Arrows indicate the direction of force acting on ddFLN4's termini during MT experiments. **(B)** Extension *vs.* time trace of a ddFLN4 tether subjected to alternating intervals of high force (here 25 pN) allowing for ddFLN4 unfolding, intermediate force (here 6.5 or 7.5 pN) allowing for direct observation of refolding, and low force (0.5 pN) to ensure refolding before the next cycle. Unfolding and refolding of ddFLN4 via the mandatory intermediate state are observed as upward or downward double-steps in the traces, respectively. Dashed lines in insets indicate extension levels corresponding to the native (N), intermediate (I) and unfolded (U) states, respectively. **(C)** Force–extension profiles of ddFLN4 for the transitions N↔I (blue) and I↔U (red), and for full (un)folding N↔U (black). Data points are obtained by Gaussian fits to step extension histograms (insets) at each constant force. Data points above 8 pN are from unfolding (based on 68-131 events obtained from 27-36 independent tethers),

data points up to 8 pN from refolding (54-159 events from 26-39 independent tethers). Error bars correspond to the FWHM of Gaussian fits, divided by the square root of counts. Lines are fits of the WLC model. **(D)** Rates of unfolding at different constant forces for the three transitions. Color code as in panel C. Error bars correspond to 95% confidence intervals of exponential fits to the fraction of observed events as a function of time (**Methods, Supplementary Fig. S4**). Lines are fits of a single-barrier kinetic model. **(E)** Rates of refolding at different constant forces. Color code, error bars and fits analogous to panel D. **(F)** Fitted rates at zero force k_0 and distances to the transition state Δx for the unfolding and refolding transitions as determined from the fits of a single-barrier kinetic model shown in panels D and E.

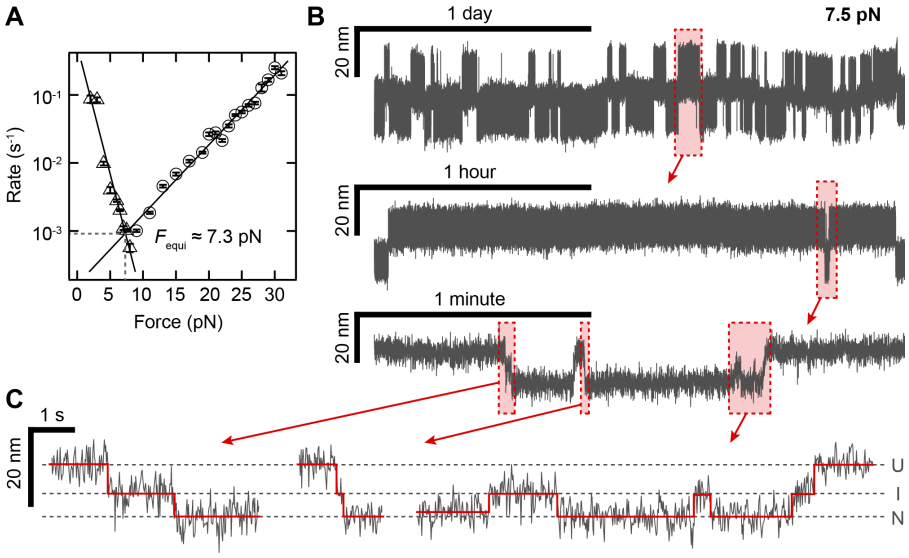


Fig. 3. Long and stable ddFLN4 folding and unfolding traces close to equilibrium
(A) Force dependence of the rates for complete unfolding (circles) and refolding (triangles) as determined in Fig. 2d-e. The intersection of the linear fits predicts the equilibrium force $F_{equi} \approx 7.3$ pN at which the probabilities of ddFLN4 being in the folded and unfolded state are equal. **(B)** 55 h-long extension *vs.* time trace of a ddFLN4 tether subjected to a constant force of 7.5 pN and zooms into indicated segments of the trace. **(C)** Zooms show not only full unfolding and folding transitions, but also transitions from the native to the intermediate state and back. Dashed lines indicate average extension levels corresponding to native (N), intermediate (I), and unfolded (U) state, respectively. Red lines indicate positions of transitions between states and extension levels in each state, as determined by the step-finding algorithm employed for analysis.

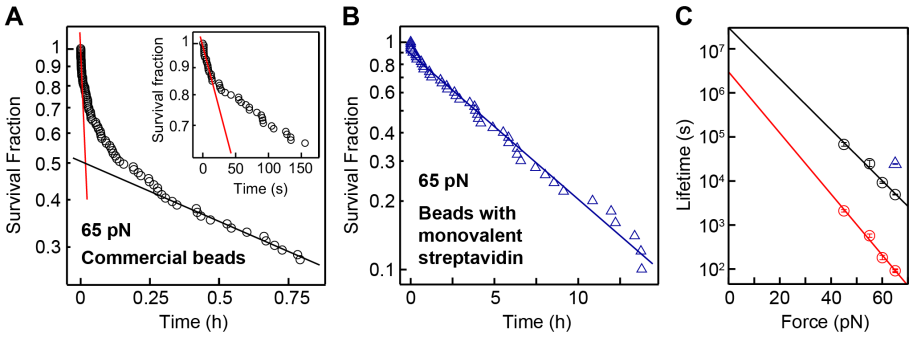


Fig. 4. Bond lifetimes of beads functionalized with different streptavidin variants
(A) Survival fraction as a function of time for commercially available beads coated with tetravalent streptavidin (Dynabeads M-270 Streptavidin, Invitrogen), tethered by ddFLN4 and subjected to a constant force of 65 pN. The survival fraction decays in a complex, multi-exponential fashion, suggesting the existence of several populations of biotin–streptavidin interactions with different lifetimes. Red and black lines are linear fits to the logarithm of the first and last 20% of data points, respectively, to estimate off-rates of the fastest- (inset) and slowest-decaying populations. **(B)** Survival fraction as a function of time for ddFLN4-tethered beads functionalized with a monovalent streptavidin variant, measured at 65 pN. The blue line is a linear fit to the natural logarithm of all data points. Note the markedly different scale of the time axis. **(C)** Estimated lifetime of the biotin–streptavidin interaction at different constant forces for the fastest- and slowest-decaying population of commercial beads with tetravalent streptavidin (red and black circles, respectively), and for beads with monovalent streptavidin (blue triangle). Error bars correspond to 1 SD; lines are fits of a single-barrier kinetic model. The number of measured, specifically tethered beads per condition was between 44 and 118, of which 35 to 86 ruptured during the duration of the measurement. The measurement duration was 15 h for beads with monovalent streptavidin and for the commercial beads 1 h at 65 pN, 3 h at 55 and 60 pN, and 15 h at 45 pN).

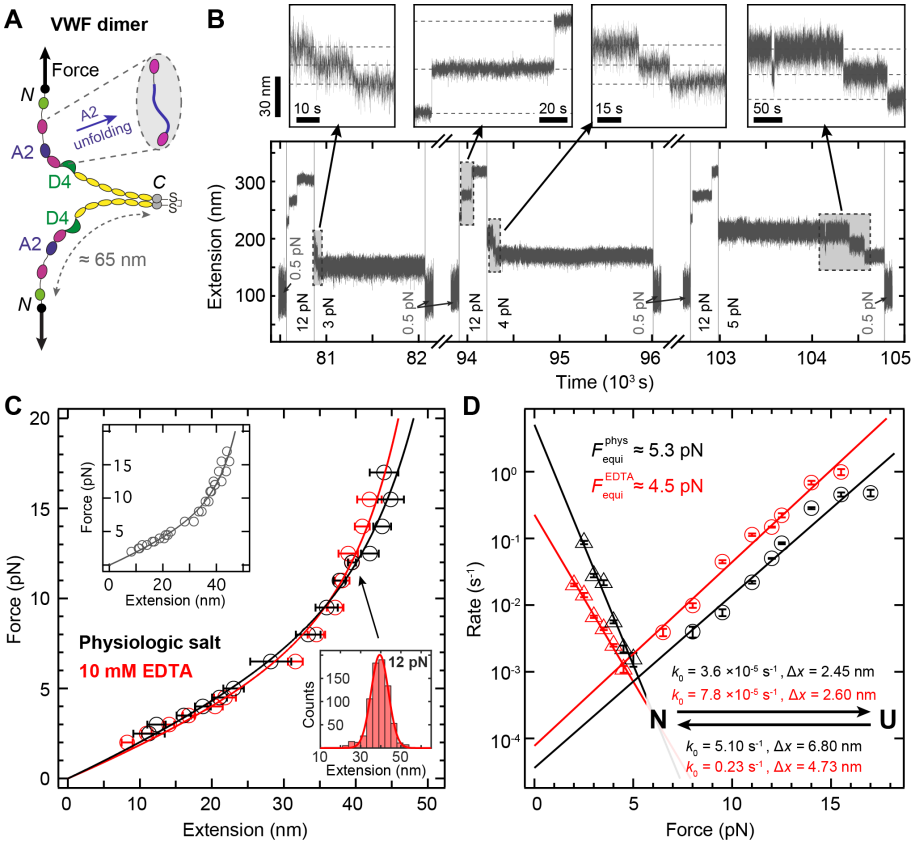


Fig. 5. Folding and unfolding of A2 domains within VWF dimers

(A) Schematic structure of a VWF dimer, consisting of two ≈65 nm, multi-domain monomers C-terminally connected by disulfide bonds. The two A2 domains, which can unfold under force (inset), are shown in blue. Arrows indicate the direction of force acting on the two N-termini during MT experiments. **(B)** Segments from a ≈30-hour long extension *vs.* time trace of a VWF dimer tether subjected to alternating intervals of high force (here 12 pN), allowing for A2 unfolding, of intermediate force (here 3, 4 or 5 pN), allowing for direct observation of A2 refolding, and of low force (0.5 pN) to ensure refolding. Unfolding and refolding of the two A2 domains are observed as two independent positive or negative steps in the trace, respectively. Dashed lines in the insets indicate extension levels with none, one, or both of the A2 domains unfolded. **(C)** Force–extension curves of A2 (un)folding, in near-physiological buffer containing Ca^{2+} (black) and in buffer without Ca^{2+} and with 10 mM EDTA (red). Data points are obtained by Gaussian fits to step extension histograms (lower right inset) at each constant force. Data points above 5 pN are from unfolding (based on 62-632 and 40-747 events for the near-physiologic and EDTA case, respectively, obtained from 13-53 independent tethers), data points up to 5 pN are from refolding (41-120 and 49-158 events for

the near-physiologic and EDTA case, respectively, obtained from 10-19 independent tethers). Error bars correspond to the FWHM of Gaussian fits, divided by the square root of counts. Lines are fits of the WLC model. Upper left inset shows a global WLC fit to all data points. **(D)** Rates of unfolding (circles) and refolding (triangles) at different constant forces for near-physiologic (black) and EDTA (red) buffer. Error bars correspond to 95% confidence intervals of exponential fits to the fraction of observed events as a function of time (**Methods**). Lines are fits of a single-barrier kinetic model, yielding rates at zero force k_0 and distances to the transition state Δx for unfolding and refolding as indicated.

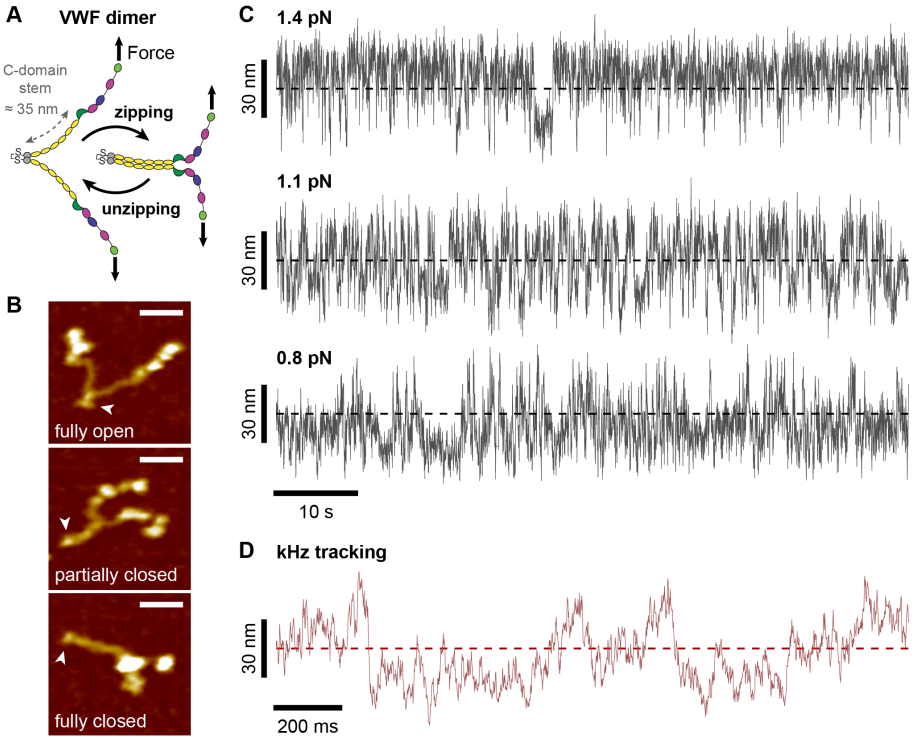


Fig. 6. Low-force unzipping and zipping of the C-domain stem in VWF dimers.
(A) Schematic of closing and opening of the stem (domains C1-C6, yellow) of VWF dimers in a zipper-like fashion due to interactions between the C-domains. **(B)** AFM images of single VWF dimers adsorbed onto a surface under near-physiologic buffer conditions. Arrowheads indicate C-terminal ends of dimers, where the two constituting monomers are linked. In thermal equilibrium and in the absence of force, dimer stems exhibit conformations ranging from fully open to fully closed. It should be noted that in approximately one-half of all dimers, the stem region is firmly closed by the D4-mediated intermonomer interaction (*9*, *10*). Scale bars are 20 nm; height range of color scale is 2.4 nm. **(C)** Extension *vs*. time traces of the same VWF dimer tether subjected to low forces. Fast reversible transitions between a maximum and minimum value of extension, approximately 60 nm apart, are observed that we attribute to closing and opening of the C-domain stem. Dashed lines indicate the midpoint between the two extreme values of extension. At a force of ≈1.1 pN, the system spends approximately half of the time above and below the midpoint. Traces are raw data recorded at 58 Hz. **(D)** Segment of an extension *vs*. time trace of the same tether shown in panel c, at 1.3 pN, recorded at an acquisition rate of 1 kHz. The measurements with high temporal resolution confirm that the observed transitions are not jumps between two discrete extension levels, but rather gradual transitions with several intermediate extension levels, in line with a zipper-like closing and opening of the stem with several pairs of interacting C-domains.

Supplementary Materials

Supplementary Figures

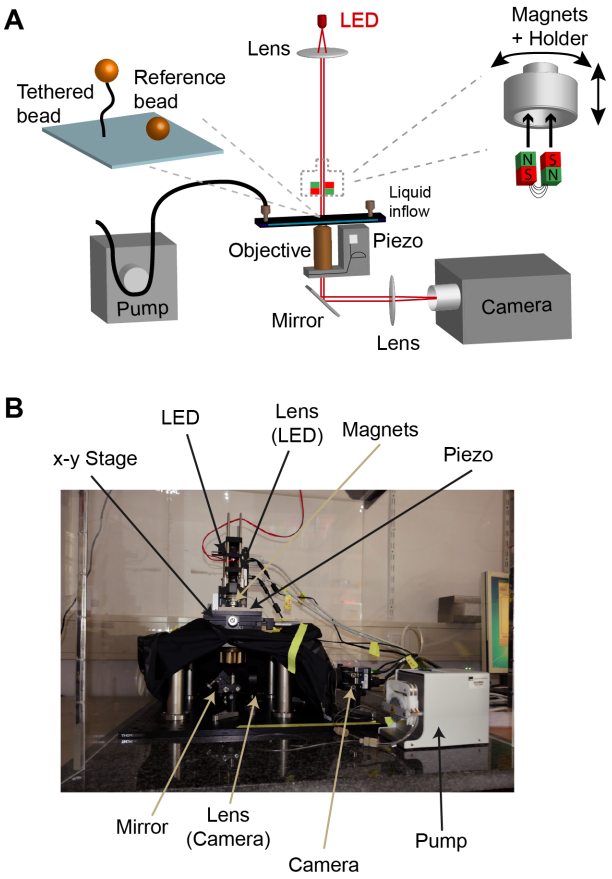


Fig. S1. Magnetic tweezers setup. (A) Schematic of the MT setup. Proteins are tethered between a magnetic bead and the bottom surface of the flow cell, which is illuminated using an LED. A large field of view is imaged using a 40x oil-immersion objective and a CMOS sensor camera. For creating the look-up table necessary to track the z position of the beads, the objective is mounted on a piezo stage. A set of two cubic permanent magnets is positioned above the flow cell. The distance between magnets and flow cell can be adjusted using a DC-motor in order to adjust the force applied to the magnetic beads. A peristaltic pump allows for flushing the flow cell. For technical details of the different components, see **Methods**. **(B)** Image of the MT setup with essential components being highlighted.

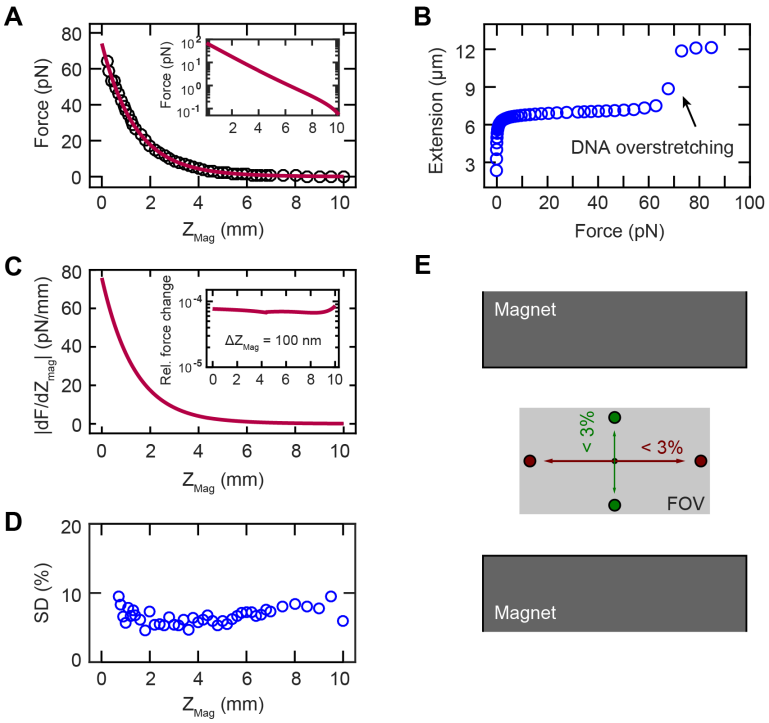


Fig. S2. Force calibration of the MT setup. (A) Force acting on the magnetic beads used in this study (Dynabeads M270) as a function of the distance Z_{Mag} between the magnets and the flow cell. Forces were calibrated using the method described by Velthuis et al. (*62*), based on the Brownian fluctuations of long (here 21 kbp) double-stranded (ds) DNA tethers. Data points are mean forces determined from 16 DNA tethers. The green line is the final fit of the dependence of force on the magnet distance. **(B)** Exemplary trace of a 21 kbp dsDNA tether, showing the B-S overstretching transition at the expected force of ≈65 pN, confirming the force calibration from analysis of the transverse fluctuations. (**C**) Absolute value of the derivative of the force with respect to Z_{Mag}. The inset shows the relative force change for extension changes in z direction of 100 nm –larger than any (un-)folding steps in our measurements–, which was found to be $< 10^{-4}$ for all forces, as calculated from the expression for $|dF/dZ_{Mag}|$. **(D)** Bead-to-bead force variation. Independently performing the calibration procedure for 16 different DNA tethers, we found the standard deviation of the force from the mean value to be ≲ 10% over the whole range of magnet distances, indicating a bead-to-bead force variation of ≲ 10%, in line with previous reports (*14*, *63*, *64*). **(E)** Force uniformity across the field of view. To verify that the magnetic field is uniform and thus the forces do not vary significantly across the field of view (FOV), we repeatedly performed the force calibration procedure for the same DNA tether at different positions at the edges of the FOV, as schematically indicated by circles, and in the middle of the FOV. For each of four independently measured DNA tethers, changes in force were found to be < 3% both along the axis parallel to and the axis perpendicular to the gap between the magnets (not drawn to scale).

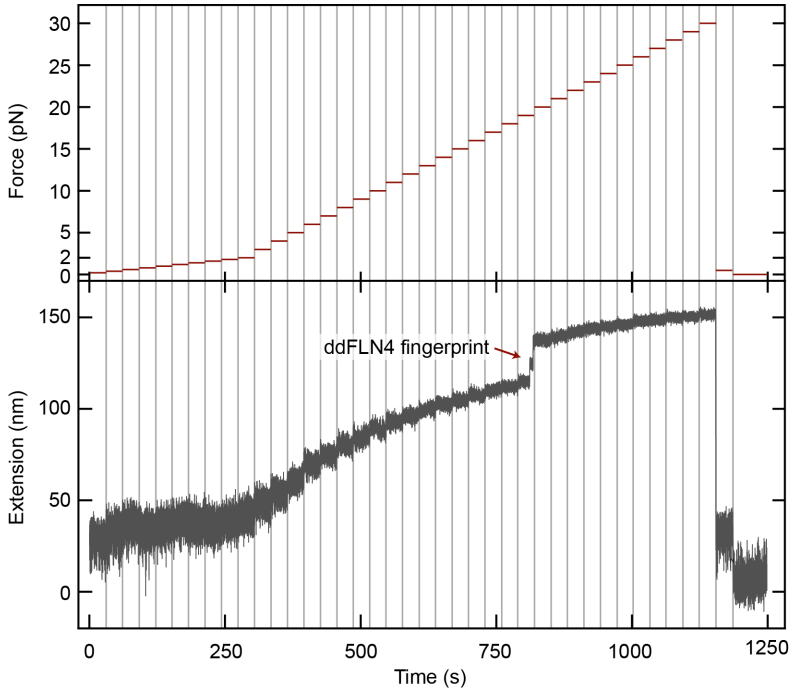


Fig. S3. Extension of ELP linker-protein tethers. Exemplary extension trace (bottom) of a ddFLN4-ELP linker complex tethered between glass surface and magnetic bead as shown in Fig. 1 in the main text, recorded while the force was increased stepwise every 30 s (indicated by red lines; top), in steps of 0.2 pN between 0.2 and 2 pN, and in steps of 1 pN between 2 and 30 pN. Afterwards, the tether was relaxed to 0.5 pN to allow for refolding of ddFLN4 and further relaxed to zero force to determine the zero position of extension. No peculiar features –in particular no steps– were observed over the entire probed force range, with exception of the characteristic ddFLN4 unfolding pattern, which served to identify specific single-tethered beads. This finding shows that the ELP linker does not cause any signals that may interfere with analysis of the specific signals of the measured protein of interest.

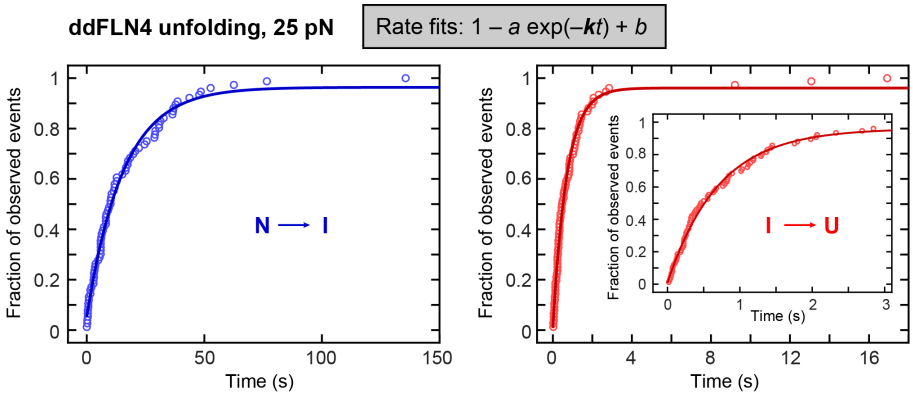


Fig. S4. Determination of rates from the observed unfolding and refolding events. Concept of rate determination from the fraction of observed unfolding or refolding events as a function of time. Shown here as an example are the fractions of observed unfolding events *vs*. time for the two substeps of ddFLN4 unfolding at 25 pN, i.e. for the transitions from the native (N) to the intermediate (I) state (left, blue) and from the intermediate to the unfolded (U) state (right, red). To obtain the unfolding rate k of a transition at constant force F, the fraction of observed unfolding events as a function of time t is fit to the expression $1 - a \exp(-kt) + b$ (lines), where the free parameters a and b can compensate for events that were missed due to the finite measurement time or due to the finite time of motor movement when setting the force. As a rule, parameters a and b were close to 1 and 0, respectively.

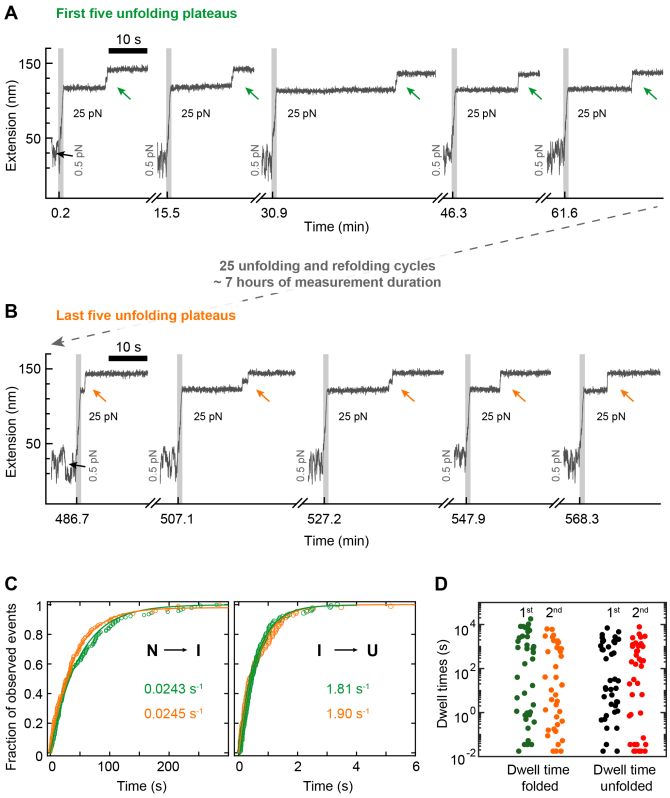


Fig. S5. ddFLN4 does not exhibit hysteresis upon repeated unfolding and refolding. (A,B) Segments from a ≈10 h-long measurement on ddFLN4 tethers with repeated alternating unfolding and refolding plateaus (similar to the data shown in **Fig. 2B**). Reliable unfolding and refolding was observed throughout the entire measurement. Shown here for one exemplary tether are the first five **(A)** and the last five **(B)** unfolding plateaus (all at 25 pN; unfolding events marked by arrows), which were separated by 25 cycles of unfolding and refolding, corresponding to ≈7 h of measurement duration. We analyzed the same 31 ddFLN4 tethers, separately for the first five and last five unfolding plateaus. The obtained mean extension values for the two unfolding transitions N→I and I→U both varied by less than 4%. Furthermore, the measured unfolding rates matched very closely. **(C)** Fits and unfolding rates are shown in green and orange for the first five and last five plateaus, respectively. Both for the first step of unfolding, N→I (left panel), and for the second step of unfolding, I→U (right panel), rates deviated by less than 5%, well within the 95% confidence intervals of the fits. Our data thus indicate that no significant hysteresis effects occur for ddFLN4 even after tens of unfolding/refolding cycles and spending an extended period of time in the unfolded state. **(D)** Analysis of a long (~50 h) trace at constant force close to the equilibrium point (the trace shown in **Fig. 3B**). The dwell times in the folded and unfolded states were quantified and are shown separately for the first and second halves of the trace. The distributions for the two halves of the trace for both folded and unfolded states are identical, within experimental error (as assessed by a two-sample Kolmogorov–Smirnov test with $p = 0.51$ and $p = 0.53$, respectively).

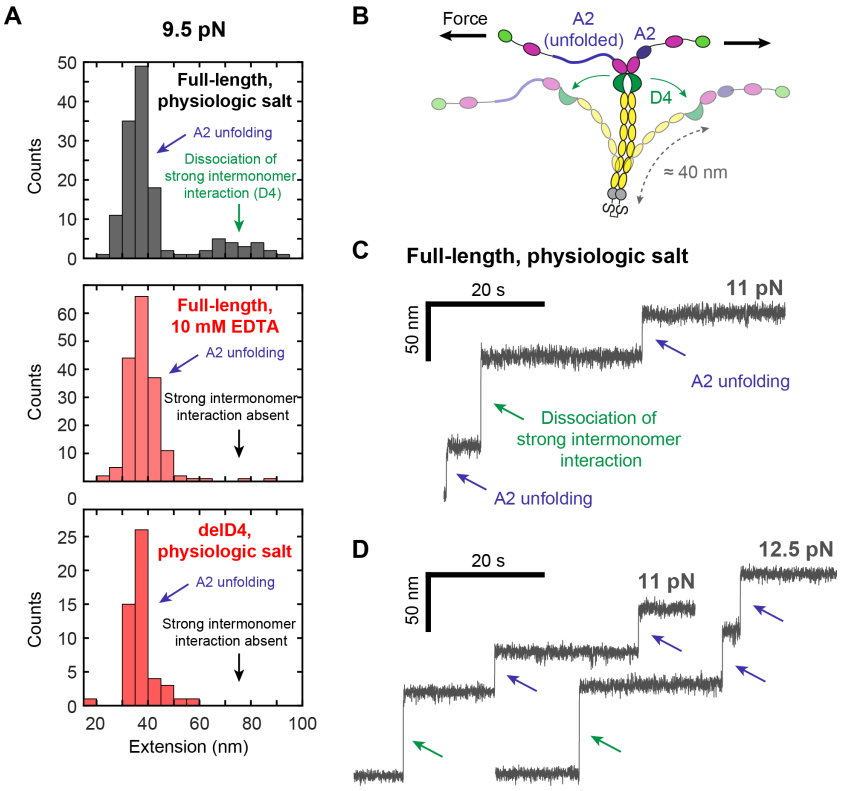


Fig. S6. Dissociation of D4-mediated intermonomer interaction in VWF dimers. (A) Extension histograms of steps observed in traces of VWF dimers recorded at a force of 9.5 pN, for full-length dimers in the presence of divalent ions (top) or in the presence of 10 mM EDTA (middle), and for dimers with a deletion of the D4 domain (delD4, see also **Supplementary Fig. S7**) in the presence of divalent ions (bottom). In the case of the full-length dimers, in the presence of divalent ions a broad peak at extension values of roughly 70-80 nm is observed in addition to the peak associated with A2 unfolding, centered at ca. 36 nm. In the presence of EDTA, or for the delD4 construct, in contrast, only the peak associated with A2 unfolding is observed. The length increase by 70-80 nm, the sensitivity to removal of divalent ions by EDTA, and the involvement of the D4 domain are in line with the dissociation of a strong intermonomer interaction mediated by VWF's D4 domain that has recently been identified in AFM force measurements on VWF dimers (*9*, *10*). **(B)** Schematic of dimer opening. Dissociation of an intermonomer interaction mediated by the D4 domain (green) leads to the opening of the closed stem region of VWF (yellow) and thus a release of formerly hidden length of approximately 80 nm. Dimer opening occurs independently of A2 (blue) unfolding, since the A2 domains are not shielded from force by the D4-mediated interaction. **(C)** Exemplary extension trace of a full-length dimer exhibiting unfolding of both A2 domains and dimer opening, recorded at 11 pN. **(D)** Extension traces from the same VWF dimer tether, probed at different forces and repeatedly exhibiting dimer opening, implying reversibility of the D4-mediated intermonomer interaction.

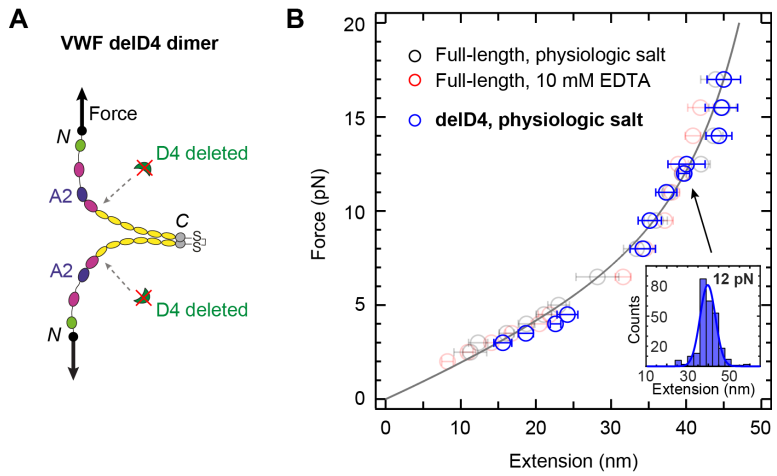


Fig. S7. Measurements on VWF dimers with deletion of the D4 domain. (A) Schematic structure of a VWF dimer with deletion of both D4 domains (delD4 dimer). The two A2 domains are shown in blue. Arrows indicate the direction of force acting on the two N termini during MT experiments. **(B)** Force–extension profile of A2 unfolding and refolding, recorded for the delD4 construct in near-physiologic buffer at pH 7.4 (blue symbols). The force–extension profile closely matches those obtained for the full-length construct in near-physiologic buffer and in buffer with 10 mM EDTA (co-plotted with lower opacity in black and red, respectively), as presented in Fig. 4c in the main text. The line is the global WLC fit to all data from the full-length construct, as presented in the inset in Fig. 4c in the main text. Data points are obtained by Gaussian fits to step extension histograms (inset) at each constant force. Data points above 5 pN are from unfolding, data points up to 5 pN from refolding. Error bars correspond to the FWHM of Gaussian fits, divided by the square root of counts.

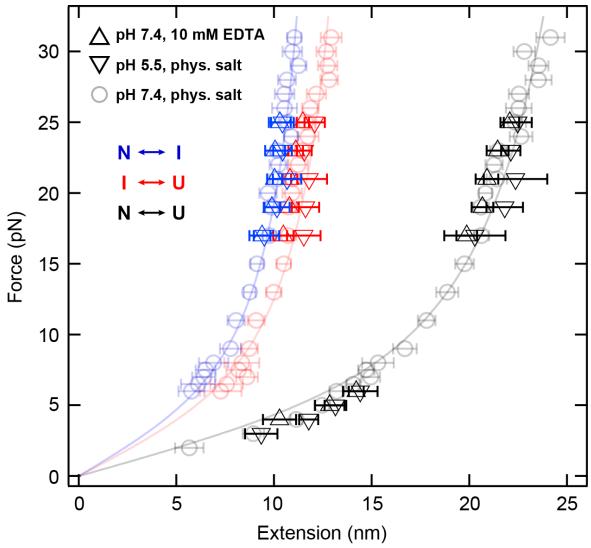


Fig. S8. ddFLN4 unfolding and refolding under varied salt and pH conditions. Extension of ddFLN4 unfolding and refolding steps at different constant forces and under varied buffer conditions. Extensions of the transitions between the native state and the intermediate state (blue) as well as between the intermediate and the unfolded state (red) are shown separately in addition to the full extension between native and unfolded state (black). Data points at forces up to 8 pN are from refolding, data points at forces above 8 pN from unfolding measurements. Co-plotted with lower opacity are the data obtained for near-physiological buffer conditions (pH 7.4, with divalent ions; circles) as shown in Fig. 2c in the main text and the respective WLC fits (lines). Force–extension data sets obtained at pH 7.4 in the presence of 10 mM EDTA (upward triangles) and at acidic pH 5.5 in the presence of divalent ions (downward triangles) both are within measurement uncertainty identical to the ones obtained for near-physiologic buffer conditions. Error bars correspond to the FWHM of Gaussian fits, divided by the square root of counts.

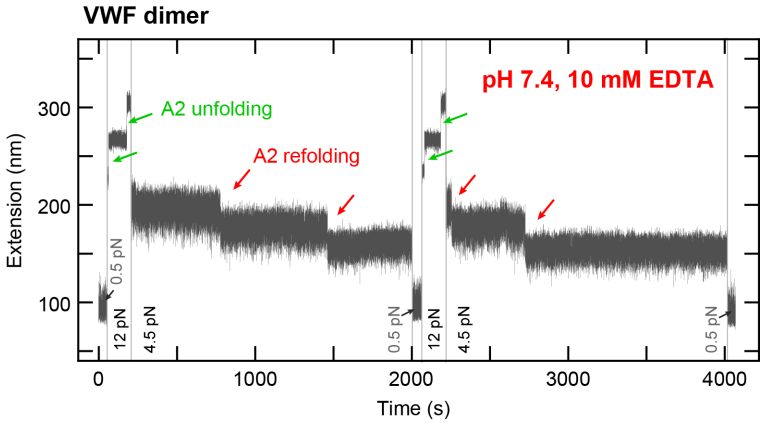


Fig. S9. Refolding of the VWF A2 domain under mechanical load in the absence of Ca^{2+}. Extension *vs.* time trace of a VWF dimer tether subjected to alternating intervals of high force (here 12 pN), allowing for A2 unfolding, of intermediate force (here 4.5 pN), allowing for direct observation of A2 refolding, and of low force (0.5 pN) to ensure refolding, in buffer without Ca^{2+} and with 10 mM EDTA. Unfolding and refolding of the two A2 domains are observed as two independent positive or negative steps in the trace, respectively. Direct observation of refolding steps (marked by red arrows) shows that A2 can refold under significant mechanical load even in the absence of Ca^{2+}.

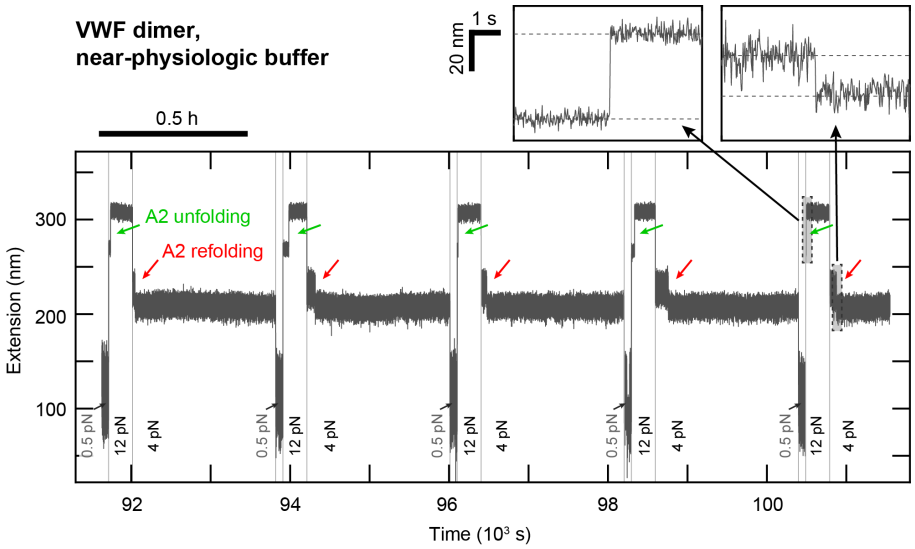


Fig. S10. Occasional inhibition of A2 refolding in VWF dimer tethers. Segment of a ≈30 h-long extension *vs.* time trace of a VWF dimer tether subjected to alternating intervals of high force (here 12 pN), allowing for A2 unfolding, and of different intermediate forces (4 pN shown here), allowing for direct observation of A2 refolding, recorded under near-physiologic buffer conditions. The shown tether exhibits the unfolding and refolding signal (marked by arrows) of only one of the two A2 domains. Otherwise, the tether does not show any differences to regular tethers exhibiting signals of both A2 domains. In particular, the observed A2 unfolding and refolding steps were indistinguishable (see insets). In such tethers exhibiting only one A2 signal, which occurred only occasionally, refolding of one of the A2 domains may be inhibited due to cis-trans isomerization of a cis-proline, as reported in a previous OT study (*8*).

Chapter 5

Biomolecular Engineering on Streptavidin

5.1 Manuscript M5: Monomeric Streptavidin: A Versatile Regenerative Handle for Force Spectroscopy

In Manuscript M5, a novel molecular handle for force spectroscopy based on a monomeric variant of SA (mcSA2) is introduced. mcSA2 only consists of a single subunit. It is designed in such a way that it still folds into a stable β-barrel and that it still binds to biotin. The tetrameric structure of SA entails an elaborate protein preparation protocol: Different polypeptide chains have to be expressed separately, denatured, mixed and diluted into a refolding buffer allowing always four chains to come together and to fold into the native tetrameric SA structure.

In contrast, the new force handle construct consists of a single polypeptide chain making the production process much easier. The construct comprises different tags for immobilization and purification, a ddFLN4 fingerprint domain and the mcSA2. The design of the construct allows *in situ* refolding of the construct, *i.e.* if a cantilever tip, on which the handle is immobilized, gets clogged, it can be placed into denaturing conditions. This causes dissociation of molecules sticking to the cantilever tip without destroying the molecular handle. It will refold when it is put back into measurement buffer.

Although mcSA2 has a much lower affinity for biotin than mSA, the interaction is still strong enough to perform useful force spectroscopy experiment, as it withstands forces of about 200 pN. As for mSA, the forces needed to unbind biotin from mcSA2 are dependent on the way mcSA2 is tethered. The difference between N- and C-terminal tethering of mcSA2 is yet less pronounced than for mSA. With fluorescence anisotropy measurements, it is shown that biotin's off-rate, in absence of force, is the same for both versions of mcSA2.

Monomeric streptavidin: a versatile regenerative handle for force spectroscopy

by

Magnus S. Bauer, Lukas F. Milles, Steffen M. Sedlak, and Hermann E. Gaub

submitted to

bioRxiv 276444; doi: https://doi.org/10.1101/276444

Monomeric streptavidin: a versatile regenerative handle for force spectroscopy

Magnus S. Bauer[1], Lukas F. Milles[1], Steffen M. Sedlak[1], and Hermann E. Gaub[1]

[1]Lehrstuhl für Angewandte Physik and Center for NanoScience (CeNS), Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence to: gaub@lmu.de

Abstract

Most avidin-based handles in force spectroscopy are tetravalent biotin binders. Tetravalency presents two issues: multiple pulling geometries as well as multiple targets bound simultaneously. Additionally, such tetravalent handles require elaborate purification protocols in order to reassemble. A stoichiometric, monomeric variant of streptavidin (mcSA2) had been engineered previously. It is readily expressed and purified, and it binds biotin with a nanomolar K_D. For atomic force microscopy-based single-molecule force spectroscopy (AFM-SMFS), we fused the monomeric streptavidin with a small protein domain as an experimental fingerprint and to improve solubility. A ybbR-tag was additionally included for covalent site-specific tethering. Rupture forces of the mcSA2:biotin complex were found to be in a comparable range above 150 pN at force loading rates of 1E4 pN/s as for previously published, tetravalent streptavidin:biotin systems. Additionally, when tethering mcSA2 from its C-terminus, rupture forces were found to be slightly higher than when tethered N-terminally. Due to its monomeric nature, mcSA2 could also be chemically denatured and subsequently refolded - and thus regenerated during an experiment, in case the handle gets misfolded or clogged. We show that mcSA2 features a straightforward expression and purification with flexible tags, high stability, regeneration possibilities and an unambiguous pulling geometry. Combined, these properties establish mcSA2 as a reliable handle for single-molecule force spectroscopy.

Introduction

Avidin-based handles have a long and successful history in biotechnology. They are widely applied as tagging and pull-down handles due to their femtomolar affinity towards the small molecule biotin, low off-rate, broad availability, and easy handling. As the first receptor-ligand system probed in atomic force microscopy-based single-molecule force spectroscopy (AFM-SMFS) studies (1,2), they still enjoy great popularity as handles to apply force to biomolecular systems.

Avidin (3) and similar molecules, such as streptavidin (4) or strep-tactin (5), are tetramers composed of four separate subunits, each capable of binding a single biotin molecule with high affinity. However, for some applications there is yet a need for precise control over stoichiometry. Considerable effort went into the design of a monovalent variant of streptavidin, a tetramer with only one single biotin binding subunit (6). For SMFS studies, an identical approach guaranteeing a well-defined tethering with 1:1 binding stoichiometry and specific pulling geometry was pursued by assembling a functional streptavidin subunit with three non-functional subunits (7). An analogous approach has been established for strep-tactin to tether a single strep-tag II peptide (8). These approaches achieve monovalent binding behavior but still require tetrameric structure to retain function. Additionally, they rely on elaborate purification procedures to assemble the tetrameric structure.

Recently, Park and colleagues undertook the effort to engineer a monomeric streptavidin - a solitary, yet functional streptavidin subunit. Monomeric variants inherently have some disadvantages compared to their tetrameric equivalents, among them lower biotin affinity, low solubility and problems with aggregation (9,10). To overcome these issues, Lim et al. engineered a monomeric streptavidin (mcSA) as a chimera based on structural homology modeling of streptavidin and rhizavidin, a dimeric protein that binds biotin using residues from only a single subunit (11). The resulting biotin affinity of 2.8 nM is the highest among non-tetrameric streptavidin. DeMonte et al. crystalized mcSA, analyzed it in detail, and improved it further by some mutations in the binding pocket (12). The resulting mcSA2 has a 20-40% lower off-rate. Adding solubility tags optimized the expression procedure (13).

In this study, we employ mcSA2 and combine it with the 4th filamin domain from *Dictyostelium discoideum* (ddFLN4) as both a molecular fingerprint for SMFS and a solubility enhancer. Additionally, an N- or C-terminal polyhistidine purification tag and a ybbR-tag (14) for site-specific covalent immobilization were included. We describe a straightforward expression and purification protocol under denaturing conditions to eliminate biotin already present in the binding pocket beforehand, followed by refolding of the fusion protein via dialysis. We test the new mcSA2 force handle in AFM-SMFS and show that the mcSA2:biotin complex withstands forces comparable to the streptavidin:biotin interaction

and is also showing two different force regimes by pulling from the molecule's N- or C-terminus. Additionally, the monomeric nature of the employed handles entail a unique feature: it can be completely denatured and refolded *in situ* making it superior to tetrameric biotin handles. For example, if clogged by stray biotin or trapped in misfolded states, the mcSA2 handle can be regenerated by recovering its binding ability. This property results in higher data yield and better statistics as it allows performing AFM-SMFS experiments with a single cantilever for several days without loss of interaction.

Results and Discussion

Applicability of the handle for force spectroscopy

To probe the applicability and long term stability of mcSA2 as a handle for force spectroscopy AFM-SMFS measurements were performed. We investigated two similar constructs to examine the mechanical characteristics of the unbinding of biotin from mcSA2 under force application on its different termini: an mcSA2 with the ddFLN4 fingerprint and the ybbR-tag on the N-terminus (geometry N, ybbR-ddFLN4-mcSA2) and an mcSA2 with the fingerprint domain and the immobilization tag on its C-terminus (geometry C, mcSA2-ddFLN4-ybbR) as depicted in Figure 1A,B.

The handles were covalently linked to AFM cantilevers and probed against a biotinylated surface (cf. materials and methods, Figure 1B). Single unbinding events could be identified by the characteristic unfolding pattern of ddFLN4, which includes a shielded substep (Figure 1C). The recurring unfolding pattern assured that the large number of specific mcSA2:biotin interaction events are pulled specifically by a single handle in a well-defined geometry, and thus shows that the handle can be implemented as a reliable force handle in SMFS experiments. The resulting forces of 150-200 pN needed for detaching a single biotin from the mcSA2 binding pocket are comparable to what has been reported for the streptavidin:biotin interaction (1,7,15). Using different retraction velocities, a dynamic force spectrum was obtained and fitted as a single bond dissociation over an energy barrier according to Bell (16) and Evans (17). For geometry N, the fit yielded a distance to the transition state x_0 = 0.42 nm and a zero-force off-rate $k_{off,0}$ = 7.7 × 10^{-6} s^{-1}. For geometry C, x_0 = 0.37 nm and $k_{off,0}$ = 6.1 × 10^{-6} s^{-1} were obtained. Over the broad range of loading rates, unbinding forces for the C-terminally tethered mcSA2 are higher than those for the N-terminally tethered mcSA2 as correctly as it could be determined with two different cantilevers.

Comparison of N- and C-terminal pulling geometry

Calibration errors and changes in force due to differing spring constants between individual cantilevers can render comparison of experimental force data – especially when addressing small force differences – unreliable. To compare rupture forces of mcSA2:biotin loaded in geometry N and C, we thus performed measurements with one single cantilever by immobilizing the two different constructs of the mcSA2 handle at two separate spots on one functionalized glass slide (Figure 2A). This way both geometries can be probed with the same cantilever with one consistent spring constant of 139.2 pN/nm in order to yield directly comparable force values. To ensure single-molecule interactions, we introduced an additional fingerprint domain on the cantilever: the refolding, alpha-helical protein FIVAR (derived from "Found In Various Architectures") domain (18) from the pathogen C*lostridium perfringens* that is known to unfold at forces of 50-60 pN (Figure 2B). Biotinylation was accomplished using an AviTag sequence (19), which is covalently modified with a biotin during protein expression (cf. Materials and Methods). Covalent and site-specific tethering was again achieved employing a ybbR-tag.

In this SMFS experiment, the cantilever alternated between surface areas with mcSA2 tethered in geometry N and C for every 300 approaches. While the unfolding forces of the fingerprint domains remained the same for both tethering geometries, we found the mcSA2:biotin interaction to be significantly stronger for geometry C than for geometry N throughout all varied retraction velocities. The most probable rupture forces in pN according to the Bell-Evans-model for each geometry is shown in Figure 2C. The most probable forces for geometry C consistently exceeded those for geometry N by 30 – 40 pN. Fitting the dynamic force spectrum with the Bell-Evans-model, the N-terminal tethering yielded a distance to the transition state x_0 = 0.39 nm and a zero-force off-rate $k_{off,0}$ = 1.2 × 10^{-5} s^{-1}, while x_0 = 0.35 nm and $k_{off,0}$ = 5.3 × 10^{-6} s^{-1} was obtained for the C-terminal tethering. These results agree well with the results obtained for the mcSA2 handles on the cantilever from Figure 1D.

Characterization of affinity

To determine whether the difference in unbinding forces for the two different geometries emerges from the way the mcSA2 molecule is loaded or by a conformational difference resulting from the addition of ddFLN4 to the termini, we performed fluorescence anisotropy experiments. In a competition assay, we measured the off-rates for both constructs in solution, thus in the absence of external force (Figure 3). Measurements of mcSA2 with ddFLN4 on the N- and C-terminus yielded off-rates of 1.05 × 10^{-4} s^{-1} and 1.08 × 10^{-4} s^{-1}, respectively. Regarding the measurement's accuracy, the off-rates of both constructs are

considered to be equal. Therefore, we conclude that the difference in unbinding force during AFM-SMFS is determined solely by the way force is applied to the handle and thus the trajectory chosen to overcome the binding energy barrier rather than the position of the ddFLN4 fingerprint itself.

Regeneration of the mcSA2 handle

In AFM-SMFS experiments, a streptavidin handle on the cantilever may occasionally pick up biotinylated molecules that were unspecifically adsorbed to the sample surface. The high affinity of the streptavidin:biotin interaction is in this case particularly disadvantageous, because biotinylated molecules block the binding pockets of the handle almost irreversibly. Once a cantilever is clogged, the interaction with the biotinylated molecules on the surface is lost and they cannot be investigated any further. To regenerate such a clogged handle, we placed the cantilever in 6 M guanidine hydrochloride to denature the mcSA2 handle, releasing biotinylated molecules from its binding pocket. Subsequent gentle washing steps in phosphate buffered saline facilitates the refolding of the handle into its functional state. The ddFLN4 fingerprint also rapidly refolds. Using this protocol, we could recover mcSA2 from clogged or misfolded states and regain tethering activity on the surface.

In our experiment, we regenerated the handle up to 3 times but the regeneration steps are not limited to that. Resuming the SMFS measurement, no significant change in unfolding or rupture forces was detectable (Figure 4).

Conclusion

Building on monomeric streptavidin, we could establish a highly specific handle for biotin-binding that is straightforward to produce and employ in force spectroscopy experiments. Additionally, mcSA2 is a long-lived tethering handle, enhanced in its performance even further as it can be regenerated by refolding. Our study shows that mcSA2 can be a significant asset for SMFS and related applications. Combined with site-specific anchoring, it permits high data yields, whenever biotinylation is possible.

We could also show the importance of anchoring positions for the stability of a receptor-ligand interaction since this changes the trajectory chosen in the binding energy landscape to overcome the energy barrier. Therefore precise control of the pulling geometry changes the interaction's mechanostability, permitting to switch the addressed force range. In conclusion, its robustness and versatility renders mcSA2 an excellent choice for force spectroscopy measurements.

Materials and Methods

Protein Expression and Purification - Gene construction and cloning

mcSA2 was expressed and purified with a fingerprint and solubility enhancer, the 4th filamin domain of *Dictyostelium discoideum* (ddFLN4). This small Ig-like fold expresses well and refolds rapidly. By varying the position of the ybbR-tag, used for covalent protein pulldown, two different tethering geometries could be examined: Geometry N with mcSA2 on the C-terminus (ybbR-ddFLN4-mcSA2) and geometry C with mcSA2 on the N-terminus (mcSA2-ddFLN4-ybbR). These constructs were cloned using the Gibson assembly strategy into pET28a vectors. The ybbR-HIS-FIVAR-AviTag was cloned into a pAC4 vector.

Both constructs were expressed in NiCo Cells (New England Biolabs) in autoinduction Media under Kanamycin resistance. Harvested cell pellets were resuspended in 50 mM TRIS, 50 mM NaCl, 10 % (w/v) Glycerol, 0.1 % (v/v) Triton X-100, 5 mM MgCl2 at pH 8.0. To enhance cell lysis, 100 µg/ml lysozyme and 10 µg/ml DNase were added. The solution was then sonicated for 2 x 8 min. The lysed cells were spun down for 10 min at 7000 rpm in a precooled centrifuge at 4°C. Solid guanidine hydrochloride was added to the supernatant to a concentration of 6 M to completely unfold the construct to release any bound biotin. The denatured construct was purified by immobilized metal ion affinity chromatography using a HisTrap FF column (GE Healthcare). Once the protein was bound to the column, it was extensively washed with denaturing buffer to remove any stray biotin present. Finally the protein was eluted with 200 mM Imidazole. The purified protein was refolded by three rounds of dialyzation against Phosphate buffered saline (PBS) overnight and finally, after the addition of 10% glycerol, flash frozen in liquid nitrogen, to be stored at -80°C.

ybbR-FIVAR-AviTag on a pAC4 vector was expressed in E. Coli CVB101 (Avidity LLC), supplemented with biotin in the expression medium in autoinduction media and was purified identically, although non-denaturing conditions.

Surface functionalization for the AFM measurement

The preparation of the experiments comprises two similar immobilization protocols. Either for the mcSA2 or FIVAR-Biotin construct with ybbR-tag or the NHS-PEG-Biotin on a glass/cantilever surface. The experiments were designed to either have mcSA2 on the cantilever and NHS-PEG-Biotin or FIVAR-Biotin on the surface or vice versa. Immobilization of mcSA2 to cantilever or glass surface is identical to the protocol used for the attachment of FIVAR. (14,20)

Preparation of Cantilevers

For aminosilanization of the cantilevers (BioLever Mini obtained from Olympus, Japan) they were first oxidized in a UV-ozone cleaner (UVOH 150 LAB, FHR Anlagenbau GmbH, Ottendorf-Okrilla, Germany) and subsequently silanized for 2 minutes in (3-Aminopropyl)dimethylethoxysilane (ABCR, Karlsruhe, Germany; 50 % v/v in Ethanol). For rinsing, the cantilevers were stirred in 2-Propanol (IPA), ddH_2O and afterwards dried at 80°C for 30 minutes. After that the cantilevers were incubated in a solution of 25 mM heterobifunctional PEG spacer (MW 5000, Rapp Polymere, Tübingen, Germany) solved in 50 mM HEPES for 30 minutes. Subsequent to rinsing with ddH_2O, the surfaces were incubated in 20 mM Coenzyme A (Calbiochem) dissolved in coupling buffer (sodium phosphate, pH 7.2) to react with the maleimide groups. After that the levers get rinsed with ddH_2O. Then the ybbR-tag of the mcSA2 (at 5-50 µM) construct (in PBS supplemented with 10 mM $MgCl_2$) is attached covalently by a sfp (at 2 µM) catalyzed reaction to the CoA. After 30 min to 2 h the protein is covalently connected resulting in an unambiguous, site-specific pulldown. Finally, the cantilevers were rinsed thoroughly and stored in 1 x PBS.

For the preparation of PEG Biotin (5000 Da) cantilevers pegylation protocols were identical, only that NHS-PEG-Biotin instead of NHS-PEG-Maleimide was applied for 1 h.

For the preparation of FIVAR cantilevers the mcSA2 construct was substituted for the FIVAR construct. Similar concentrations of protein were used.

Preparation of Glass Surfaces

Before aminosilanization the glass surfaces were cleaned by sonication in 50 % (v/v) Isopropanol (IPA) in ultrapure water for 15 minutes. For oxidation the glass surfaces were soaked for 30 minutes in a solution of 50 % (v/v) hydrogen peroxide (30 %) and sulfuric acid. Afterwards they were thoroughly washed in ultrapure water and then blown dry in a gentle nitrogen stream. Silanization is achieved by incubating in (3-Aminopropyl)dimethylethoxysilane (ABCR, Karlsruhe, Germany, 1.8 % v/v in Ethanol) while gently shaking. Thereafter, surfaces were washed again in IPA and ultrapure water and then dried at 80°C for 40 minutes, to be stored under Argon for weeks.

To attach mcSA2 to the glass surface heterobifunctional Polyethyleneglycol (PEG, 5000 Da, dissolved in 100 mM HEPES pH 7.5 at 25 mM for 30 min) spacers were used to avoid unspecific interactions between the cantilever and the glass surface. The PEG spacers had an N-hydroxysuccinimide (NHS) group on one side, for attachment to the aminosilanized surface. The other end provided a Maleimide group for subsequent coupling to the thiol group of Coenzyme A (CoA, 1 mM in 50 mM sodium phospahte, 50 mM NaCl, 10 mM EDTA, pH 7.2, incubated for 1 h). Through a reaction catalyzed by sfp (at 2 µM) the CoA was covalently connected to the ybbR-tag of the mcSA2 (at 5-50 µM) construct (in PBS

supplemented with 10 mM $MgCl_2$ for 30 min to 2 h), resulting in an unambiguous, site-specific pulldown.

For the preparation of PEG Biotin (5000 Da) surfaces pegylation protocols were identical, only that NHS-PEG-Biotin instead of NHS-PEG-Maleimide was applied for 1 h.

For the preparation of FIVAR surfaces the mcSA2 construct was substituted for the FIVAR construct. Similar concentrations of protein were used.

AFM-SMFS

Adapted from Milles et al. (18):

AFM-SMFS data was acquired on a custom-built AFM operated in closed loop by a MFP3D controller (Asylum Research, Santa Barbara, CA, USA) programmed in Igor Pro 6 (Wavemetrics, OR, USA). Cantilevers were briefly (<200 ms) and gently (< 200 pN) brought in contact with the functionalized surface and then retracted at constant velocities ranging from 0.2, 0.8, 1.6, 2.0, 3.2, 5.0, 6.4 to 10.0 µm/s for a dynamic force spectrum. After each curve acquired, the glass surface was moved horizontally by at least 100 nm to expose an unused, fresh surface spot. Typically, 50000 - 100000 curves were recorded per experiment. If quantitative comparisons of absolute forces were required, a single cantilever was used to move between multiple spatially separated spots to be probed on the same surface (created using the protocol described above). To calibrate cantilevers the Inverse Optical Cantilever Sensitivity (InvOLS) was determined as the linear slope of the most probable value of typically 40 hard (>2000 pN) indentation curves. Cantilevers spring constants were calculated using the equipartition theorem method with typical spring constants between 90-160 pN nm-1. A full list of calibrated spring constants from experiments presented in this work is provided below, as the stiffness of the cantilever, may influence the complex rupture and domain unfolding forces measured. Experiments and spring constants of cantilevers for data shown:

Measurement	**Spring constant [pN/nm]**	**Force [pN] @ 800 nm/s**
geometry C - surf_biotin_lv_mcSAddFLN4ybbR (Figure 1D)	56.2	204.2
geometry N surf_biotin_lv_ybbRddFLN4mcSA2 (Figure 1D)	120.9	179.9
both geometries - surf_mcSA2bothmulti_lv_yFIVARbiotin (Figure 2C)	139.2	187.2 / 218

SMFS data analysis

Adapted from Milles et al. (18):

Data analysis was carried out in Python 2.7 (Python Software Foundation). Laser spot drift on the cantilever relative to the calibration curve was corrected via the baseline noise (determined as the last 5 % of datapoints in each curve) for all curves and smoothed with a moving median (windowsize 300 curves). The inverse optical lever sensitivity (InvOLS) for each curve was corrected relative to the InvOLS value of the calibration curve.

Raw data were transformed from photodiode and piezo voltages into physical units with the cantilever calibration values: The piezo sensitivity, the InvOLS (scaled with the drift correction) and the cantilever spring constant (k).

The last rupture peak of every curve was coarsely detected and the subsequent 15 nm of the baseline force signal were averaged and used to determine the curve baseline, that was then set to zero force. The origin of molecule extension was then set as the first and closest point to zero force. A correction for cantilever bending, to convert extension data in the position of the cantilever tip was applied. Bending was determined through the forces measured and was used on all extension datapoints (x) by correcting with their corresponding force datapoint (F) as

$xcorr = x - F/k$.

To detect unfolding or unbinding peaks, data were denoised with Total Variation Denoising (TVD, denoised data is not shown in plots), and rupture events detected as significant drops in force relative to the baseline noise.

Rupture force histograms for the respective peaks and dynamic force spectra were assembled from all curves showing the fingerprint unfolding, or (if applicable) a specific fingerprint domain, and/or a clean complex rupture event. The most probable loading rate of all complex rupture or domain unfolding events was determined with a KDE, bandwidth chosen through the Silverman estimator. This value was used to fit the unfolding or rupture force histograms with the Bell-Evans model for each pulling velocity. A final fit was performed through the most probable rupture forces and loading rates for each pulling velocity to determine the distance to the transition state $\Delta x0$ and natural off-rate at zero force $k_{off,0}$.

Fluorescence Anisotropy Measurement

For fluorescence anisotropy measurements, biotinylated fluorescently labeled single-stranded DNA was mixed with the mcSA2 constructs in a 1:1 ratio. The change in anisotropy upon the addition of a more than 100-fold excess of biotin was recorded for 2,5 h.

Fluorescence anisotropy measurements were carried out in Corning 384 well plates. For passivation, the wells were incubated with 5 mg/ml bovine serum albumin dissolved in

phosphate buffered saline (PBS) (Sigma-Aldrich, Saint Louis, USA) for 2 h. After removing the passivation solution by turning the plates upside down, the wells were flushed twice with ultrapure water.

The protein constructs were filtered with a 0.45 µm centrifuge filter (Merck Millipore, Darmstadt, Germany) according to the manufacturer's instructions. To match the buffers, we employed Zeba Spin Desalting Columns (Thermo Scientific, Rockford, USA) with 7K MWCO using PBS to equilibrate the columns following the manufacturer's protocol.

The concentrations of the constructs were determined with a NanoDrop 1000 (Thermo Scientific, Rockford, USA) UV-Vis spectrophotometer using the absorption peak at 280 nm and an extinction coefficient of 41035 $M^{-1}cm^{-1}$ calculated from the protein sequence using the "ExPASy: SIB bioinformatics resource portal" (21). We used 17 bp long single-stranded DNA oligonucleotides labeled with Biotin at the 5'-end and a ATTO 647N dye ot the 3'-end purchased from IBA (IBA GmbH, Göttingen, Germany).

We prepared 40 µl of 30 nM biotinylated fluorescently labeled DNA and the same amount of protein construct dissolved in PBS containing 1 mM DTT. As G-factor and measurement blank, we used 40 µl PBS with 1 mM DTT added. G-factor reference also contained 30 nM of the biotinylated fluorescently labeled DNA. After measuring the anisotropy in absence free biotin, we added 10 µl 818 µM Biotin dissolved in PBS to all wells and recorded the anisotropy every five seconds for 2.5 h.

Acknowledgements

Support for this work was provided by the ERC Advanced Grant CelluFuel. The authors thank D.A. Pippig, F. Baumann, M.A. Jobst for helpful discussions, M. Freitag for experimental assistance, K. Erlich for proof reading and A. Kardinal and T. Nicolaus for laboratory support.

Author contributions

M.S.B.: Conceptualization, Data curation, Software, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review and editing
L.F.M.: Conceptualization, Data curation, Software, Investigation, Visualization, Writing—original draft, Writing—review and editing
S.M.S.: Data curation, Formal analysis, Investigation, Writing—original draft, Writing—review and editing
H.E.G.: Conceptualization, Supervision, Funding acquisition, Investigation, Project administration, Writing—review and editing

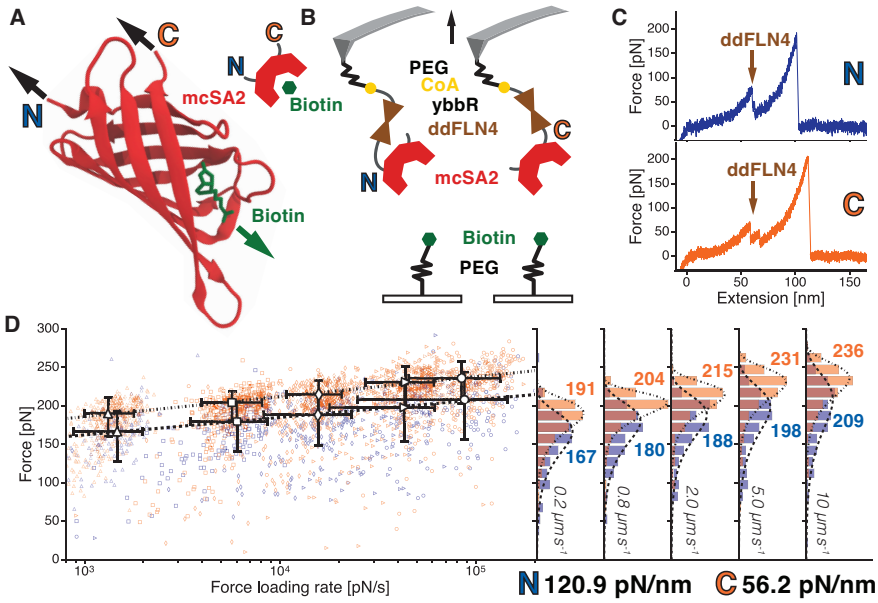


Figure 1 - Characterization of the mcSA2 handle by AFM-based SMFS. *Panel A: the crystal structure adapted from protein database (PDB) entry 4JNJ (12) and schematic of mcSA2 (red) and biotin (green) with pulling geometries N (blue, pulled from N-terminus) and C (orange, pulled from C-terminus). Panel B: a schematic of the attachment chemistry is depicted. Both constructs are immobilized on an aminosilanized cantilever with heterobifunctional NHS-PEG-maleimide linkers. On the maleimide side of the PEG, a CoA is attached for an sfp phosphopantetheinyl transferase (sfp) catalyzed reaction with the ybbR-tag of the mcSA2 handle constructs. The likewise aminosilanzed glass surface is functionalized with a heterobifunctional NHS-PEG-biotin linker. Panel C: two exemplary curves for both geometries N (top) and C (bottom) with its characteristic ddFLN4 fingerprint. Panel D: a dynamic force spectrum and force histograms of both geometries N (blue) and C (orange) indicating a similar force loading rate dependence but with generally higher forces for geometry C. The forces indicated in the histograms show the most probable force in pN according to the Bell-Evans-model. In this experimental setup the different force datasets had to be recorded with two separate cantilevers in order to probe the long term stability of the handles in both geometries on the cantilever. Since e.g. deviations in the cantilevers' spring constants (bottom right) hinder to compare forces directly in absolute values, both tethering geometries were additionally measured with a single cantilever in one measurement for better comparability as shown in Figure 2.*

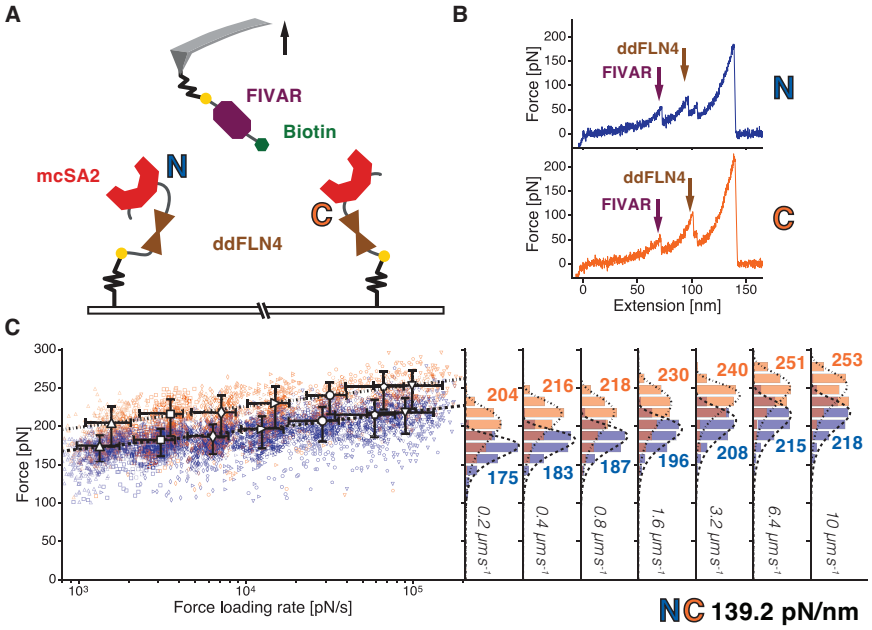


Figure 2 - Direct comparison of unbinding forces for two different tethering geometries N and C. *Panel A: to compare the unbinding forces of the two tethering scenarios, both geometries N (blue, pulled from N-terminus) and C (orange, pulled from C-terminus) were immobilized on separate spots on a surface and were probed using the same cantilever harboring a FIVAR domain with a Biotin attached. Panel B: two exemplary curves for both geometries N (top) and C (bottom) with its characteristic FIVAR and ddFLN4 fingerprint. Panel C: the data were recorded within one experiment by switching between the two spots every 300 curves. This resulted in a dynamic force spectrum and force histograms for both geometries, allowing direct comparison of unbinding forces for both geometries N and C. The forces indicated in the histograms show the most probable force in pN according to the Bell-Evans-model. The spring constant of the cantilever (139.2 pN/nm) used to pull both geometries is shown on the bottom right.*

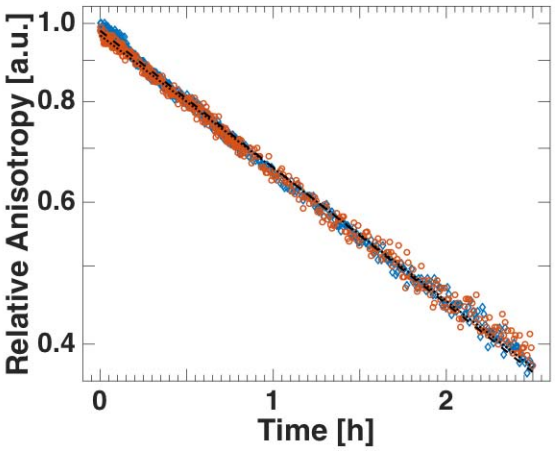


Figure 3 - Off-rates for two different tethering geometries. *For geometry C (orange circles) and geometry N (blue diamonds), the relative anisotropy is plotted over time. Fitting the off-rates yields 0.000108 s-1 × t - 0.208 for geometry C (black dotted line) and 0.000105 s-1 × t - 0.342 for geometry N (black dashed line). Hence, no significant difference for the off-rates is observed. (Here, relative anisotropy denotes the logarithm of the present anisotropy difference between sample and reference divided by the difference at the moment of biotin addition, t=0.)*

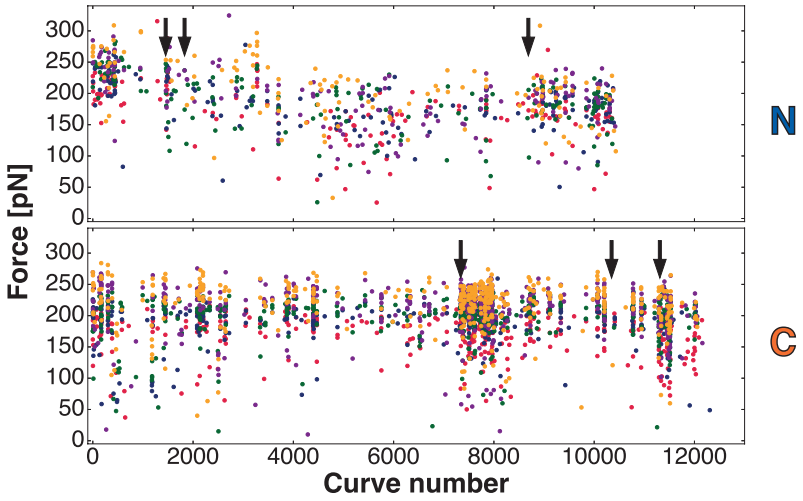


Figure 4 - Regeneration of the mcSA2 handle. *During the course of an AFM-SMFS measurement, the pulling handle eventually gets clogged with excess biotin picked up from the surface or is brought into a misfolded state rendering it unable to bind biotin any more. Due to its monomeric nature mcSA2 is able to be unfolded in 6 M guanidine hydrochloride and subsequently refolded in phosphate buffered saline in order to resume the measurement. These regeneration steps are indicated with black arrows. The Graph shows the force of mcSA2:biotin rupture in pN vs. curve number from the dataset shown in Figure 1D. Each curve number contains one pulling cycle of five retraction speeds of 200 nm/s (red), 800 nm/s (blue), 2000 nm/s (green), 5000 nm/s (purple), 10000 nm/s (orange). After a regeneration step, the ability to bind biotin is recovered - shown by the increased number of interactions recorded after the black arrows. This worked well with both geometries N (top panel) and C (bottom panel).*

References

1. Florin, E. L., Moy, V. T., and Gaub, H. E. (1994) Adhesion forces between individual ligand-receptor pairs. *Science* **264**, 415-417
2. Lee, G., Kidwell, D., and Colton, R. (1994) Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. *Langmuir* **10**, 354-357
3. Eakin, R. E., McKinley, W. A., and Williams, R. J. (1940) Egg-white injury in chicks and its relationship to a deficiency of vitamin h (biotin). *Science* **92**, 224
4. Tausig, F., and Wolf, F. J. (1964) Streptavidin—A substance with avidin-like properties produced by microorganisms. *Biochemical and Biophysical Research Communications* **14**, 205-209
5. Voss, S., and Skerra, A. (1997) Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Engineering, Design and Selection* **10**, 975-982
6. Howarth, M., Chinnapen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L., El-Husseini, A., and Ting, A. Y. (2006) A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods* **3**, 267-273
7. Sedlak, S. M., Bauer, M. S., Kluger, C., Schendel, L. C., Milles, L. F., Pippig, D. A., and Gaub, H. E. (2017) Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. *PLOS ONE* **12**, e0188722
8. Baumann, F., Bauer, M. S., Milles, L. F., Alexandrovich, A., Gaub, H. E., and Pippig, D. A. (2016) Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat Nanotechnol* **11**, 89-94
9. Laitinen, O. H., Nordlund, H. R., Hytönen, V. P., Uotila, S. T. H., Marttila, A. T., Savolainen, J., Airenne, K. J., Livnah, O., Bayer, E. A., Wilchek, M., and Kulomaa, M. S. (2003) Rational Design of an Active Avidin Monomer. *Journal of Biological Chemistry* **278**, 4010-4014
10. Wu, S.-C., and Wong, S.-L. (2005) Engineering Soluble Monomeric Streptavidin with Reversible Biotin Binding Capability. *Journal of Biological Chemistry* **280**, 23225-23231
11. Lim, K. H., Huang, H., Pralle, A., and Park, S. (2013) Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnology and Bioengineering* **110**, 57-67
12. DeMonte, D., Drake, E. J., Lim, K. H., Gulick, A. M., and Park, S. (2013) Structure-based engineering of streptavidin monomer with a reduced biotin dissociation rate. *Proteins: Structure, Function, and Bioinformatics* **81**, 1621-1633
13. Demonte, D., Dundas, C. M., and Park, S. (2014) Expression and purification of soluble monomeric streptavidin in Escherichia coli. *Applied Microbiology and Biotechnology* **98**, 6285-6295
14. Yin, J., Straight, P. D., McLoughlin, S. M., Zhou, Z., Lin, A. J., Golan, D. E., Kelleher, N. L., Kolter, R., and Walsh, C. T. (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* **102**, 15815-15820
15. Wong, J., Chilkoti, A., and Moy, V. T. (1999) Direct force measurements of the streptavidin-biotin interaction. *Biomol Eng* **16**, 45-55
16. Bell, G. I. (1978) Models for the specific adhesion of cells to cells. *Science* **200**, 618-627
17. Evans, E., and Ritchie, K. (1997) Dynamic strength of molecular adhesion bonds. *Biophys J* **72**, 1541-1555
18. Milles, L. F., Bayer, E. A., Nash, M. A., and Gaub, H. E. (2016) Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium perfringens. *The Journal of Physical Chemistry B*, *121*(15), 3620–3625.

19. Beckett, D., Kovaleva, E., and Schatz, P. J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Science : A Publication of the Protein Society* **8**, 921-929
20. Zimmermann, J. L., Nicolaus, T., Neuert, G., and Blank, K. (2010) Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat Protoc* **5**, 975-985
21. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* **40**, W597-603

Supplementary information

Protein sequences

ybbr-HIS-ddFLN4-mcSA2
MDSLEFIASKLAHHHHHHGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGG
DGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAP
GSGSGSGSAEAGITGTWYNQHGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRY
NGTKLEWRVEWNNSTENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQGQDT
FTKVK

mcSA2-ddFLN4-HIS-ybbr
MSAEAGITGTWYNQHGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRYNGTKLE
WRVEWNNSTENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQGQDTFTKVKG
SGSGSGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPV
DPVMVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSELKLPRSRHH
HHHHGSLEVLFQGPDSLEFIASKLA

ybbr-HIS-FIVAR-AviTag (+ Biotin)
MADSLEFIASKLAHHHHHHGSSGSDKTNLGELINQGKSLLDESVEGFNVGEYHKGAKDGLT
VEINKAEEVFNKEDATEEEINLAKESLEGAIARFNSLLIEESTSGTASGLNDIFEAQKIEWHE

Part IV

Conclusions and Outlook

Summary: New Insights into the Mechanics of the Streptavidin/Biotin Interaction

The SA/biotin system is widely used in bionanotechnology. Nowadays, SA also often serves as molecular handle to apply force to biotinylated proteins, DNA or macromolecular structures. First AFM-based SMFS experiments on the SA/biotin interaction have been performed 25 years ago. Since then, the interaction has been repeatedly studied by AFM-based SMFS and a wide range of partly contradictory results has been reported (*cf.* Manuscript M1, Section 1.5).

In this thesis, I employed methods from biomolecular engineering to refine AFM-based SMFS on the SA/biotin interaction and to put previous results into context. Combining a monovalent, but tetrameric variant of SA, with site-specific immobilization strategies, I was able to gain new insights into the mechanical stability of the SA/biotin interaction, *i.e.* its behavior under tensile force. Controlling the stoichiometry of SA and its immobilization site, *i.e.* the force-loading geometry, in AFM-based SMFS measurements, I obtained monodisperse force spectroscopy data of an intrinsically multivalent, and thus polydisperse, system (*cf.* Publication P1, Section 3.1).

I first focused on the unbinding of biotin from a single functional SA subunit (within a monovalent SA tetramer). Tethering mSA either at the N- or at the C-terminus of the functional subunit, I observed a large difference in biotin unbinding force (*cf.* Publication P2, Section 3.2). Complementary SMD simulations helped to unravel what is causing the different behavior for the different tethering geometries on the molecular level: For N-terminal force loading, the first two β-strands of the functional subunit are pulled apart before biotin is pulled out of the pocket. This partial unfolding of the SA subunit impedes the structural integrity of the binding pocket and results in lower unbinding forces. For C-terminal force loading, on the other hand, the binding pocket stays intact so that higher forces are needed to drag biotin out of SA's binding pocket. This work (*cf.* Publication P2, Section 3.2) underlines the importance of the force-loading geometry for the mechanical stability of a receptor-ligand interaction. Furthermore, the observation that partial unfolding of the receptor's binding pocket precedes the unbinding of the ligand from the binding pocket is a fundamental discovery that changes our concept of ligand-receptor systems under force, showing that unbinding and unfolding are not necessarily two separate processes.

I made use of the difference in biotin unbinding force between N- and C-terminally tethered mSA to establish a new SMC&P assay (*cf.* Publication P3, Section 4.1). Complementing the N-terminally tethered mSA in the depot area and the C-terminally tethered mSA in the target area with an SdrG (in a non-native geometry) on the AFM cantilever, a force hierarchy was created that allows to transfer biotinylated protein domains equipped with an N-terminal Fgβ-peptide from the depot to the cantilever tip to the target area. Exploiting the nanometer precision of the AFM, the proteins can be assembled in arbitrary

nanoscale patterns in the target area. In various aspects, this SMC&P assay is superior to previous versions: First, it is completely DNA-free and would thus, in principle, allow studying networks of DNA processing enzymes. Second, compared to previous assay, the unbinding forces for all involved receptor-ligand systems are significantly higher, which makes the use of a fingerprint domain possible. This builds an internal feedback into the assay that enables on-line monitoring of the transfer process. Third, our transfer construct offers several opportunities to covalently link enzymes of interest. This is only possible because the long-lived and stable SA/biotin interaction is used for both the depot and the target, while other attempts to realize a pure protein-based SMC&P assay tried to build up a force hierarchy using three different receptor-ligand systems. The transfer construct can thus be used as a shuttle, to which enzymes of interest are attached, in future SMC&P experiments.

Besides the force propagation through a single SA subunit, I also studied the force propagation through a fully functional SA tetramer (*cf.* Manuscript M2, Section 3.3). For 4SA, site-specifically tethered by a unique cysteine at one of the C-termini, I observed three different force peaks in the unbinding force histogram. Using SA of different valencies, I substantiated the hypothesis that these different unbinding forces originate from binding of biotin to the different SA subunits. An explanation for the processes within the SA molecule that result in these different forces could be found using SMD simulations for the different force-loading geometries caused by binding of biotin to the different subunits of the SA tetramer: Again, it is the force-loading geometry that matters. Depending on the direction in which biotin is pulled out of the binding pocket, biotin itself or the adjacent molecular linker can be pushed against the flexible L3/4-loop, which acts as a lid closing over the binding pocket. The lid is thereby forced towards its open conformation. This lowers the mechanical stability of the SA/biotin bond such that biotin can leave the pocket at lower forces. Although this appears to be similar to the mechanism observed for mSA, it is yet different: For N-terminally tethered mSA, partial unfolding of the receptor molecule lowers the energy barrier that has to be overcome to unbind the ligand. Here, when a restoring force builds up in 4SA, the interaction with the ligand itself induces a conformational change within the receptor molecule lowering the energy barrier.

For SA of different valencies, I performed not only AFM-based SMFS, but also studied the SA/biotin system under constant force using magnetic tweezers (*cf.* Manuscript M3, Section 3.4). The results agree well with those for AFM-based SMFS: Different lifetimes of the SA/biotin interaction were observed – most likely, due the same mechanism as described in the previous paragraph – and could be attributed to the binding of biotin to the different SA subunits. Identifying an especially long-lived binding geometry, a tenfold increase in lifetime (compared with commercially available SA-coated superparamagnetic beads) was obtained by site-specific coupling of mSA by the C-terminus of the functional subunit. I could thereby contribute to implement mSA as a stable force-handle for long-term measurements in magnetic tweezers (*cf.* Manuscript M4, Section 4.2). Beyond that I also contributed to the development and characterization of mcSA2 as a new handle for AFM-based SMFS experiments, described in the last part of the result section of this thesis (*cf.* Manuscript M5, Section 5.1).

Outlook: The Eternal Dilemma of Science

The eternal dilemma of science is that each answer leads to new questions. In this thesis, I answered some questions about the mechanical stability of the SA/biotin interaction. In this section, I will point out both remaining issues and new questions, propose future experiments and provide ideas on how to tackle the involved challenges.

In Publication P2 (Section 3.2), differences in unbinding pathways for the experiment and the SMD simulation were reported: For N-terminal tethering of mSA, high unbinding forces were observed in the simulation but not in the experiment. We took an educated guess and attributed this difference to the large difference in force-loading rates between the experiment and the simulations. Extrapolating the force-loading rates, the forces in the simulations should yet be much higher. Recently, another group published results on high-speed AFM on the SA/biotin interaction bridging the gap in force-loading rates between simulation and experiment. Their simulations showed even lower unbinding forces. This might indicate that the force-field parameters for biotin – in both our and their simulations – are not ideal. Unfortunately, as described in Manuscript M1 (Section 1.5), they used a different force-loading geometry so that our question about the high unbinding forces seen in our simulation could not be answered. High-speed AFM-based SMFS experiments with our site-specific N-terminal attachment strategy could still verify our educated guess and would be another step towards resolving the puzzle of the SA/biotin interaction. Another experimental approach would be to use protein engineering to prevent the N-terminal unfolding and thus block this low force unbinding pathway.

In Manuscript M2 (Section 3.3), it was shown that the closed L3/4-loop is vital for the mechanical stability of the SA/biotin interaction and its conformation highly depends on the force-loading direction. Another idea for future experiments would thus be to perform similar experiments using TA or ST, as both of them have mutations within L3/4-loop influencing its conformation. These experiment could provide further insights into the loop dynamics and could potentially lead to the development of a molecular force handle with even longer lifetimes than for mSA/biotin.

One of the short-comings of Manuscript M3 (Section 3.4) is that it could not be finally concluded that it is always a SA/biotin bond that ruptures. In principle, it would also be possible that instead the SA tetramer ruptures. To be sure about the molecular mechanism and to exclude SA tetramer rupture, it would therefore be interesting to probe the mechanical stability of the SA tetramer itself in the presence and absence of biotin. For this, the protocol for SA preparation must be modified to allow for creation of SA that comprises three different subunits (instead of two, as has been the case for all SA used in this thesis). One could probably genetically fuse one of the subunits (X) to a protein domain, *e.g.* ddFLN4, as has been performed for mcSA2 (*cf.* Manuscript M5, Section 5.1). The second type of subunit (Y) would be labeled with a polyhistidine tag and the third type of subunit (Z) would have no label at all. Purification of the right tetramers (consisting of X, Y, Z and Z) would be performed by Ni-IMAC and subsequent size-exclusion chromatography. Anchoring of SA to

the cantilever would be performed *via* X, and anchoring to the surface *via* Y. SA could then be pulled apart and the mechanical stability of the SA tetramer could be determined. Depending on which of X, Y and Z are functional or non-functional, stabilization of the SA tetramer by ligand binding could be investigated. As the assembly of the subunits into the SA tetramer is stochastic, the exact force-loading geometry would not be controllable in this experiment. Instead, one would have to decipher the observed rupture forces afterwards, probably using complementary SMD simulations. A work-around would be to use fluorescently labels or quenchers on the different subunit and to perform combined TIRF/AFM-based SMFS experiments in zero-mode waveguides.

Besides studying the stability of the SA tetramer under force, experiments on the mechanical stability of a single subunit within the tetramer could be performed in the future. The comparison of the unfolding forces of an SA, TA or ST subunit in the presence or absence of different ligands, such as biotin, desthiobiotin, iminobiotin or StrepTag II, with thermodynamic parameters like binding enthalpy, affinity or free energy might yield some valuable insights into the relation of unbinding force and binding affinity. Beyond that, these experiments could provide additional information about cooperativity in biotin binding between the different subunits. Probing the mechanical stability of subunit D in 1SA, 3SA, 4SA in the presence and absence of biotin, conclusions about allosteric effects within the SA tetramer might be drawn in the future.

In Manuscript M3 and M4 (Sections 3.4 and 4.2), mSA is introduced as a reliable force handle in magnetic tweezers experiments on proteins. In the described protocol, the unique cysteine is used to site-specifically couple SA to maleimide groups at the end of PEG linkers. Since the use of PEG in magnetic tweezers flow cells is problematic, as it promotes unspecific sticking of the magnetic beads to the surface, it might be advantageous to also replace the PEG on the beads by ELP linkers. For this, the thiol-maleimide coupling would have to be replaced by an enzyme-linked coupling to an mSA variant with a C-terminal sortase motif on the functional subunit. The preparation of such a mSA variant has already been successful. As the enzyme-linked tethering to ELP might be less efficient compared with the established thiol-maleimide coupling, the practicability of the sortase-mediated linkage of SA to ELPs has to be thoroughly investigated.

Closing Remark: Nanos Gigantum Humeris Insidentes

Regarding SMFS experiments, the SA/biotin system is probably the most studied receptor-ligand system. This raises the question why there is another thesis on the ”Mechanics of the Streptavidin/Biotin Interaction”. The answer is almost trivial: Recent technological advances enable to dig deeper and to shed more light on the mechanical stability of the SA/biotin interaction. This motivation pretty much resembles the ones given for previous studies and it can be readily assumed that the same argument will apply in the future, when new experimental tools become available. I am well aware of this and do certainly not claim to have finally resolved everything about the SA/biotin system. But I contributed with new results and elucidated certain aspects of the interaction. Just like a puzzle piece, this work will finally fit in with all former and future studies. Considering all SMFS experiments on the SA/biotin system conducted over the last 25 years, I am certain that my work on the SA/biotin system will not be the last but is, at least for now, the latest.

Part V

Appendix

Chapter 6

Further Publications

6.1 Publication P4: Quantitative Evaluation of Statistical Errors in Small-Angle X-ray Scattering Measurements

Quantitative Evaluation of Statistical Errors in Small-Angle X-Ray Scattering Measurements

by

Steffen M. Sedlak, Linda K. Brützel, Jan Lipfert

published in

Journal of Applied Crystallography (2017). 50, 621-630

research papers

JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Quantitative evaluation of statistical errors in small-angle X-ray scattering measurements

Steffen M. Sedlak, Linda K. Bruetzel and Jan Lipfert*

Department of Physics, Nanosystems Initiative Munich, and Center for NanoScience, LMU Munich, Amalienstrasse 54, Munich, 80799, Germany. *Correspondence e-mail: jan.lipfert@lmu.de

Received 13 January 2017
Accepted 23 February 2017

Edited by J. Hajdu, Uppsala University, Sweden

Keywords: small-angle X-ray scattering; SAXS; measurement errors; simulations; scattering profiles; hybrid pixel detectors.

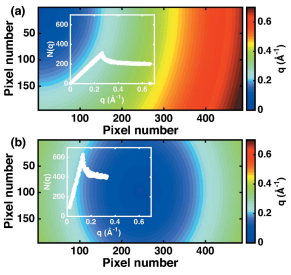
Supporting information: this article has supporting information at journals.iucr.org/j

A new model is proposed for the measurement errors incurred in typical small-angle X-ray scattering (SAXS) experiments, which takes into account the setup geometry and physics of the measurement process. The model accurately captures the experimentally determined errors from a large range of synchrotron and in-house anode-based measurements. Its most general formulation gives for the variance of the buffer-subtracted SAXS intensity $\sigma^2(q) = [I(q) + \text{const.}]/(kq)$, where $I(q)$ is the scattering intensity as a function of the momentum transfer q; k and const. are fitting parameters that are characteristic of the experimental setup. The model gives a concrete procedure for calculating realistic measurement errors for simulated SAXS profiles. In addition, the results provide guidelines for optimizing SAXS measurements, which are in line with established procedures for SAXS experiments, and enable a quantitative evaluation of measurement errors.

1. Introduction

Small-angle X-ray scattering (SAXS) is a powerful technique to probe the structure, conformations and dynamics of biological macromolecules and their complexes in solution (Vachette *et al.*, 2003; Svergun & Koch, 2003; Lipfert & Doniach, 2007; Putnam *et al.*, 2007; Hura *et al.*, 2009; Blanchet & Svergun, 2013). A particular advantage of the SAXS technique is the ability to study macromolecules under virtually arbitrary solution conditions, from (near) physiological to highly denaturing. The ability to probe even complex and/or partially folded macromolecules and their assemblies in solution has made SAXS a popular tool for structural biology. SAXS data are routinely used in increasingly complex analyses, ranging from the traditional Guinier fits (Guinier, 1939) and regularized Fourier transformations (Glatter, 1977; Moore, 1980; Svergun, 1992), to *ab initio* shape reconstructions of proteins (Svergun, 1999; Walther *et al.*, 2000) and nucleic acids (Lipfert, Das *et al.*, 2007; Lipfert, Chu *et al.*, 2007), and hybrid methods that incorporate data from a combination of measurement modalities (Grishaev *et al.*, 2005; Putnam *et al.*, 2007; Rambo & Tainer, 2013*a*,*b*; Schindler *et al.*, 2016; Chen & Hub, 2015; Tuukkanen *et al.*, 2016; Bernadó *et al.*, 2007; Konarev *et al.*, 2016).

In a typical SAXS measurement, the macromolecules of interest are dissolved in an appropriate buffer and a scattering pattern is recorded (Fig. 1*a*). Subsequently, the scattering pattern of pure buffer is measured. After circular averaging of the two scattering patterns, the buffer profile is subtracted from the scattering profile of the macromolecular sample to obtain the final scattering profile, which is used for further processing and analysis (Figs. 1*b* and 1*c*). In the measurement process a range of systematic and statistical errors are, at least



OPEN ACCESS

potentially, introduced. Possible sources of measurement error include (i) problems with sample preparation, purification and homogeneity, (ii) radiation damage of the sample during measurement, (iii) scattering contributions from components of the setup such as the sample cell, beamstop and X-ray windows, (iv) the inherent beam divergence, (v) errors due to detector noise and counting statistics, and (vi) errors in buffer subtraction, for example, due to a mismatch in the beam intensity or in buffer composition as well as alterations of the setup between buffer and sample measurements.

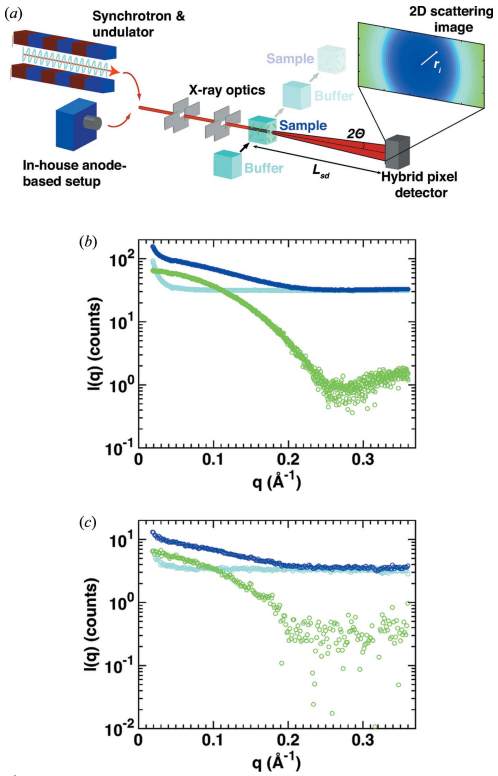


Figure 1
Principle of biological SAXS measurements. (*a*) Schematic of a SAXS setup. At a synchrotron, electrons passing through an undulator (or wiggler or bending magnet) produce X-rays; alternatively an anode source is used at in-house setups. The beam is collimated and directed at a measurement cell filled with either protein sample or buffer only. A hybrid pixel detector records two-dimensional scattering images, which are transformed to one-dimensional scattering profiles. (*b*)–(*c*) One-dimensional scattering profiles from the sample (cytochrome *c* at 8 mg ml^{-1}; dark blue) and buffer (light blue) and the resulting buffer-subtracted scattering profiles (green) obtained at (*b*) a synchrotron source (exposure time 1 s; BM29, ESRF, Grenoble) and (*c*) an in-house source (Bruetzel *et al.*, 2016) (exposure time 2 h; Department of Physics, LMU Munich).

Despite the fact that SAXS profiles are applied in increasingly sophisticated analyses, there is currently no widely accepted and tested model for the errors in SAXS profiles. A solid understanding and quantification of the errors in SAXS measurements are desirable for several reasons: (i) to quantify the reliability of SAXS measurements and to assess the goodness of fit of, for example, a model against experimental data; (ii) to quantitatively compare and optimize different setups; (iii) to simulate SAXS profiles including the appropriate noise.

In particular, in the context of simulating SAXS profiles, different models for the error on scattering profiles have been put forward. A popular choice of model for the error on SAXS profiles is to add Gaussian noise to the scattering intensity in every q bin with zero mean and a constant standard deviation (Bernadó *et al.*, 2007; Schindler *et al.*, 2016; Förster *et al.*, 2008; Pinfield & Scott, 2014), which is often expressed as a percentage of the forward scattering intensity $I(0)$ or the scattering intensity at the highest scattering angle $I(q_{\max})$. This choice of a constant Gaussian error corresponds to setting the variance $\sigma^2(q) = a^2$, where a is a constant. Values for a described in the literature (Förster *et al.*, 2008; Pinfield & Scott, 2014; Zettl *et al.*, 2016) range from $0.3I(q_{\max})$ to $10^{-4}I(0)$. Given that the scattering intensity for biological macromolecules tends to decrease with increasing q, the choice of a constant standard deviation for all q values corresponds to a (often much) larger relative error at higher q. An alternative choice of model is to introduce Gaussian noise with a q-dependent standard deviation $\sigma(q)$; one such model (Stovgaard *et al.*, 2010) proposed the use of $\sigma(q) = I(q)(q + \alpha)\beta$ with constants $\alpha = 0.15$ and $\beta = 0.3$. A similar model is used by the program *FoXS* to estimate the uncertainties of the scattering intensity when computing a SAXS profile from a crystal structure (Schneidman-Duhovny *et al.*, 2010), though *FoXS* uses different constants and additionally employs a Poisson distribution. Similarly, the *FoXS* web server (Schneidman-Duhovny *et al.*, 2010) will assume errors distributed according to a Poisson distribution with an expectation value (which is equal to the variance for Poisson distributions) of 10, unless the user provides an experimental measurement error.

Comparing the models currently described in the literature with experimental data (see Fig. S1 in the supporting information), we find that they fail to quantitatively capture the experimentally observed errors for the entire q range. Here, we first derive and then test a new model for the measurement errors in SAXS experiments that provides an accurate description of experimental data over a large range of measurement parameters.

2. Materials and methods

2.1. Samples for SAXS measurements

Cytochrome *c*, lysozyme and bovine serum albuminum (BSA) were purchased from Sigma–Aldrich and applied without further purification. The lyophilized powder of each

protein was weighed to prepare a stock solution of the highest concentration and diluted to the required concentrations. Cytochrome *c* was dissolved in 100 m*M* acetate buffer pH 4.6, with 0.5 *M* guanidinium hydrochloride added. For lysozyme a 40 m*M* acetate buffer pH 4.5, with 150 m*M* NaCl added, was prepared. BSA was dissolved in 50 m*M* HEPES pH 7.5, 50 m*M* KCl. Prior to the measurements, both buffer and sample solutions were filtered through 0.22 µm syringe filters (Thermo Scientific, USA) and centrifuged at 13 500 r min^{-1} for 10 min in a tabletop centrifuge (Eppendorf, Germany). For in-house SAXS measurements sample and buffer solutions were degassed in a desiccator at a pressure level of 30 mbar (3 kPa) for 20 min to prevent the formation of air bubbles in the sample chamber during experiments. 80 µl of sample or buffer solution was loaded into the sample chambers. For synchrotron measurements 35 µl of sample or buffer solution was used.

2.2. SAXS data acquisition

In-house SAXS measurements were performed with an Mo GeniX^{3D} microfocus X-ray tube (Xenocs SA, Sassenage, France) combined with FOX2D single reflection optics delivering a monochromatic beam with an X-ray energy of 17.4 keV (Bruetzel *et al.*, 2016). The sample–detector distance was set to ∼1.11 m, yielding usable q values between 0.05 and 0.35 Å^{-1}. We used a PILATUS 100K detector (DECTRIS Ltd, Switzerland) for X-ray detection. For each experiment, sample and buffer profiles were collected with three to five exposures of 2 h each.

All synchrotron data, except for the data presented in Figs. 6 and 7, were collected at beamline BM29 at the ESRF in Grenoble at an X-ray energy of 12.5 keV and a sample–detector distance of 2.87 m, resulting in a usable q range of 0.05–0.35 Å^{-1} (Pernot *et al.*, 2013). We used a PILATUS 1M (DECTRIS Ltd, Switzerland) detector for data acquisition. Data were collected in 'flow' mode at room temperature with ten measurement frames at an exposure time of between 1 and 4 s in 'multibunch mode' or 'low bunch mode'.

3. Results and discussion

We propose a new model for the errors in a typical SAXS measurement and evaluate the model against measured SAXS data from a range of experimental setups that employ hybrid pixel detectors (Broennimann *et al.*, 2006; Henrich *et al.*, 2009). While sample quality and (the absence of) radiation damage are critical factors in any SAXS measurement (Hura *et al.*, 2009; Jeffries *et al.*, 2016; Dyer *et al.*, 2014), they tend to be specific to the sample under investigation (Hopkins & Thorne, 2016). In this work, we will focus, therefore, on errors that are intrinsic to the SAXS measurement process, *i.e.* statistical errors resulting from photon counting statistics. Systematic errors and radiation damage are not treated further here. We note that the errors considered in our model are unavoidable in any physical measurement and constitute a best-case scenario, which is most relevant for simulations of SAXS profiles. All test measurements reported in this work use well characterized samples that are pure and monodisperse and do not suffer from radiation damage under the measurement conditions employed. Most calibration measurements used cytochrome *c*, a protein typical of weakly scattering biological samples that has been used as a scattering standard previously (Bruetzel *et al.*, 2016). Data were collected at state-of-the-art synchrotron-based (Pernot *et al.*, 2013; Lipfert *et al.*, 2006; Beno *et al.*, 2001) and in-house anode-based SAXS setups (see §2 for details). We confirmed the absence of radiation damage by partitioning the total exposure time of each measurement into frames and testing for significant changes in the scattering curves in subsequent frames (Fig. S2).

Our new error model is based on the following assumptions: (i) scattering and photon counting are Poisson processes; (ii) the scattering intensity of the buffer profile is approximately constant over the whole q range; (iii) buffer and sample measurements have independent statistical errors.

3.1. Counting statistics

Raw SAXS data are two-dimensional images (Fig. S3) providing the number of counts per pixel n_i. To obtain a scattering profile, every pixel is assigned to the appropriate momentum transfer value q and the one-dimensional intensities for the sample $I_s(q)$ and the buffer $I_b(q)$ (in units of counts) are calculated by averaging over all $N(q)$ pixels belonging to the same q bin:

$$I_{s,b}(q) = \frac{1}{N(q)} \sum_{i=1}^{N(q)} n_i. \tag{1}$$

Assuming that the individual pixels have independent statistical errors σ_i, the variance of the intensity (*i.e.* of the sample mean) at a given value of q is given by (a detailed derivation is provided in the supporting information)

$$\begin{aligned} \sigma^2_{s,b}(q) &= \frac{1}{N^2(q)} \sum_{i=1}^{N(q)} \sigma_i^2 \\ &= \frac{1}{N^2(q)} \sum_{i=1}^{N(q)} [n_i - I_{s,b}(q)]^2. \end{aligned} \tag{2}$$

While the second line of equation (2) is applied to evaluate and quantify errors from experimental data, the more general formulation in the first line will be used in the next steps. Assuming that scattering and photon counting are Poisson processes, the mean and the variance of the distribution of counts are equal, which results in

$$\sigma_i^2 = I_{s,b}(q), \tag{3}$$

$$\sigma^2_{s,b}(q) = \frac{I_{s,b}(q)}{N(q)}. \tag{4}$$

We find the relationship in equation (4) to be valid for both in-house and synchrotron data over the entire measured q range (Fig. 2). For the synchrotron data, missing pixels between the different detector modules (Fig. S3) cause small increases in the variance at specific q values due to the

decreases in the number of pixels in these q bins, which are correctly reproduced by the model in equation (4) (Fig. 2*a*). For the in-house data, there are some outliers in the variance for large q, which result from broken pixels (Fig. 2*b*).

3.2. Buffer subtraction

In biological SAXS experiments, a buffer profile is subtracted from the sample profile to obtain the macromolecular scattering curve, which is used for further analysis:

$$I(q) = I_s(q) - I_b(q). \quad (5)$$

Assuming buffer and sample measurements to be independent, we have to propagate the uncertainties by adding the variances:

$$\sigma_I^2(q) = \sigma_s^2(q) + \sigma_b^2(q). \quad (6)$$

For simplicity, we approximate the buffer profile to be constant over the whole q range. This is a good approximation for all but the lowest q values (Figs. 1*b* and 1*c*) and we find that in practice it works well over the entire q range considered in our measurements (see below). It is convenient to relate the buffer profile intensity to the sample profile (at an arbitrary q value q_{arb}) by introducing a contrast factor c:

$$I_b(q) = cI_s(q_{arb}). \quad (7)$$

Since the buffer profile is approximately constant for intermediate to large q values, q_{arb} can be chosen arbitrarily within this constant buffer range (Figs. 1*b* and 1*c*). The choice of q_{arb} then sets the values of $I_s(q_{arb})$ and c, such that the $cI_s(q_{arb})$ is constant, and defines the level of the constant buffer intensity. Equations (4), (5) and (7) allow us to rewrite equation (6) as (a step-by-step derivation can be found in the supporting information)

$$\begin{aligned}\sigma_I^2(q) &= \frac{I(q)}{N(q)} + \frac{I_b(q)}{N(q)} + \frac{I_b(q)}{N(q)} \\ &= \frac{I(q)}{N(q)} + \frac{2cI_s(q_{arb})}{N(q)} \\ &= \frac{1}{N(q)}\left[I(q) + \frac{2cI(q_{arb})}{(1-c)}\right]. \end{aligned} \quad (8)$$

The second term in the sum in equation (8) represents the constant buffer intensity. We note that it is constant and independent of q_{arb}. The choice of q_{arb} affects the values of c and $I(q_{arb})$, but not the overall value of the second term. From an experimental point of view, it might appear unnecessarily complicated to not keep the buffer and sample intensities explicitly; however, for simulations of experimental noise for computed scattering profiles the formulation of equation (8) is very convenient. Typical calculations of theoretical SAXS profiles [*e.g.* using *CRYSOL* (Svergun *et al.*, 1995) or *FoXS* (Schneidman-Duhovny *et al.*, 2010)] from crystal structures do not generate separate buffer $I_b(q)$ and sample $I_s(q)$ scattering profiles and only the final intensity $I(q)$ is provided. Therefore, the form of equation (8) is advantageous, because it only contains $I(q)$, $N(q)$ and c. The number of pixels per q bin $N(q)$ can be approximated as shown in the following section. The contrast c between sample and buffer intensity at a certain, arbitrary, scattering vector q_{arb} has to be estimated and we provide typical values derived from experimental data in Table 1 and Table S1 in the supporting information.

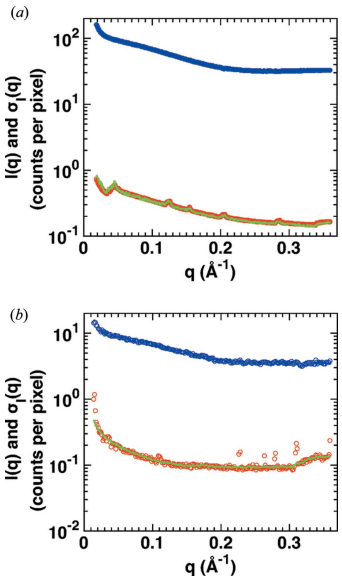


Figure 2
Counting statistics of SAXS profiles before buffer subtraction. SAXS measurements of cytochrome *c* (at 8 mg ml^{-1}) at (*a*) a synchrotron source (exposure time 1 s; BM29, ESRF, Grenoble) and (*b*) an in-house source (Bruetzel *et al.*, 2016) (exposure time 2 h; Department of Physics, LMU Munich). Scattering profiles of the protein samples after circular averaging are shown as blue circles; the corresponding standard errors of the mean σ_s computed from the counts in the individual pixels [equation (2)] are shown as red circles. The green line is the square root of the intensity divided by the number of pixels per q bin $\{\sigma_s(q) = [I_s(q)/N(q)]^{1/2}\}$.

Table 1
Typical values for k and c for in-house and synchrotron SAXS experiments.

Typical values (q_{arb} = 0.2 Å^{-1}).

	$I(q)$ (in counts)	k (in Å)	$c(q_{arb})$
Synchrotron	1–100	4500	0.85
In-house setup	0.1–10	4500	0.90

3.3. Effects of the setup geometry

Equation (8) states an inverse proportionality of the variance and the number of pixels per q bin, which in turn is determined by the setup and detector geometry. Especially for low count rates and small detector dimensions, frequently encountered at in-house setups, the setup geometry is of great importance to achieve good data quality. The exact binning of

pixels is subject to some freedom, but for a standard SAXS geometry where the detector is placed orthogonally to the beam and the q range is linear, *i.e.* the size of the q bins is constant, one finds

$$N(q) \propto 2\pi r = 2L_{\rm sd}\pi \tan\left[2\sin^{-1}\left(\frac{\lambda q}{4\pi}\right)\right] \simeq L_{\rm sd}\lambda q, \quad (9)$$

where $L_{\rm sd}$ is the sample–detector distance, r the distance of a pixel to the beam centre on the detector and λ the X-ray wavelength. This geometrical relation is illustrated in Fig. 1(*a*). We use the convention of $q = 4\pi\sin(\theta)/\lambda$ for the absolute value of the momentum transfer q, where 2θ is the angle between the incident and the scattered beam.

The detector dimensions restrict the values for r. For small detectors, r is tightly confined and the exact setup geometry is relevant for the number of pixels per q bin. By varying the sample–detector distance and/or the position of the beam centre on the detector, the available q range and the number of pixels per q bin (and thus the variance) can be changed (Fig. 3 and Figs. S3–S5). For larger detectors, r is less constrained for the same q range. In general, the number of pixels per q bin $N(q)$ can approximated by $N(q) = kq$, at least for small scattering angles, so that

$$\sigma_I^2(q) = \frac{1}{kq}\left[I(q) + \frac{2cI(q_{\rm arb})}{(1-c)}\right]. \quad (10)$$

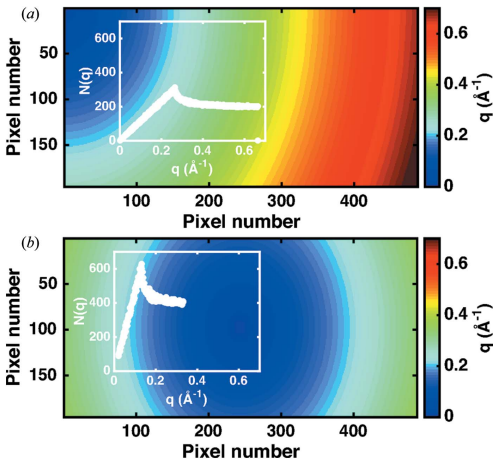


Figure 3
Measurement geometry and the number of pixels per q bin. Assignment of the pixels in two dimensions to q bins for different positions of the beam centre on a PILATUS 100K detector, with an X-ray energy of 17.4 keV and a sample–detector distance of 1.11 m. (*a*) Configuration with the beam centre aligned on the top left corner of the detector. (*b*) Setup with the beam centred on the detector. The insets show the resulting pixels per q bin $N(q)$. Further beam centre positions are depicted in Fig. S4.

Comparison with experimental data allows for determination of c and k. The value of c will depend on the choice of $q_{\rm arb}$ and on the sample. In contrast, k depends predominantly on the setup geometry.

3.4. Measurement errors for in-house and synchrotron measurements

First, we consider a single sample frame and a single buffer frame only (using 24 mg ml^{-1} cytochrome c as a representative test sample; red circles in Fig. 4). The final scattering profile $I(q)$ and the corresponding variances $\sigma^2(q)$ are calculated by circular averaging of the pixels for both the sample and buffer profiles and subsequent buffer subtraction while propagating the errors, according to equations (2), (5) and (6). We compare the experimental data with our model using the exact $N(q)$ [equation (8); black line in Fig. 4] as well as the approximation [equation (10); green line in Fig. 4] with $q_{\rm arb}$ = 0.2 Å^{-1}. As described in §3.2, the value for $q_{\rm arb}$ can be chosen arbitrarily. Our choice of $q_{\rm arb}$ = 0.2 Å^{-1} is motivated by the

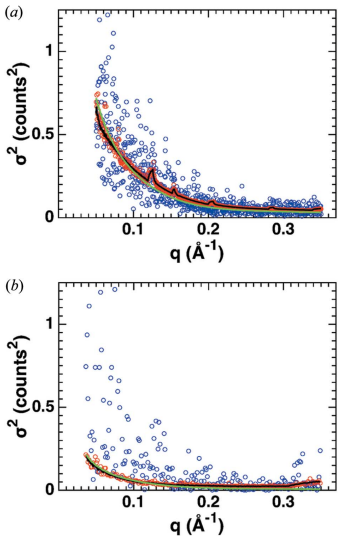


Figure 4
SAXS measurement errors after buffer subtraction. SAXS measurement of cytochrome c (24 mg ml^{-1}) obtained (*a*) at a sychrotron source (BM29, ESRF, Grenoble) and (*b*) from our in-house source (Department of Physics, LMU Munich). The propagated standard errors for the buffer-subtracted measurement from single exposures of sample and buffer are shown as red symbols. The black lines are co-plots of equation (8) with the red data, using the exact number of pixels per q bin for the respective setups and determining the contrast factor c by dividing the buffer by the sample intensity at $q_{\rm arb}$ = 0.2 Å^{-1} [synchrotron: c = 0.69, $I(q_{\rm arb})$ = 14.20; in-house setup: c = 0.68, $I(q_{\rm arb})$ = 2.30]. The green lines show fits of equation (10) to the data, with k as a free fitting parameter (synchrotron: k = 6104; in-house setup: k = 4298). The blue data points show the variance in the intensity determined from repeat exposures (synchrotron: ten exposures of 4 s each; in-house setup: three exposures of 3 h each) for comparison.

observation that for smaller scattering angles the assumption of constant buffer intensity becomes inaccurate while for larger angles the number of pixels per q bin decreases and thus the measurement errors in I_b and I_s increase. Since $I(q_{arb})$ and c are determined from the sample and buffer scattering intensities at q_{arb}, there are no free parameters if the exact number of pixels per q bin is used; if the approximation in equation (10) is used, the only free fitting parameter is k. We find an excellent agreement between our model and the experimental data if the exact number of pixels per q bin is taken into account (compare red data and black lines in Fig. 4). For the synchrotron data a good fit is achieved even if the approximation for the number of pixels per q bin is used [equation (10); green line in Fig. 4(*a*)]. For the in-house data, if the linear approximation for the number of pixels per q bin [equation (10)] is used, the fit still captures the right trend and magnitude, but is less convincing. Consequently, it is preferable to use the exact number of pixels per q bin [equation (8)] for in-house data, mostly owing to the smaller detector.

As an alternative way to estimate the measurement errors, we computed the variances of the buffer-subtracted scattering intensities from repeat exposures (blue dots in Fig. 4). We first perform circular averaging on every single sample and buffer frame. Then, we form pairs of sample and buffer profiles. Using equation (5), we calculate a final scattering profile for each pair. Now, we determine the variances between the resulting scattering intensities in each q bin. We find the variances computed from repeat exposures to be distributed more broadly compared with the errors estimated from single frames (Fig. 4, compare blue to red points), which is likely to be due to the still comparatively low number of frames. Note that with a larger number of frames the variance from the repeated exposures more closely resembles the estimate from one pair of frames (compare Figs. 4*a* and 4*b*). Importantly, the functional dependence of the variance on q is very similar for the two estimates.

3.5. Experimental errors and optimization of SAXS measurements

To demonstrate the applicability of our error model and to obtain quantitative estimates of the errors in SAXS profiles under a range of conditions, we collected buffer-subtracted scattering profiles for cytochrome c at a state-of-the-art synchrotron beamline with varying sample concentrations, exposure times and beam intensities (§2). For this analysis, image frames were stacked to create a single sample and a single buffer image, on which circular averaging was subsequently performed. To determine the experimental errors, we computed the mean and variance of the scattering intensity in each q bin [equations (2) and (6); Fig. 5]. The total scattering intensities differ for varying measurement conditions. Thus, the absolute values of the variances (Fig. 5*a*) do not directly reflect the quality of the SAXS data. As a better and more readily interpreted measure of the signal-to-noise ratio, we therefore focus on the standard deviation relative to the scattering intensity (Fig. 5*b*) to discuss the scaling of the observed errors with measurement parameters.

For our experimental data, we find the relative error (for a given flux and protein concentration) to only depend on the total (flux-corrected) exposure time t_{exp} and to scale as $t_{exp}^{-1/2}$ (inset in Fig. 5*a*). t_{exp} can be increased by increasing either the flux, the exposure time per frame or the number of frames. In practice, the exposure time per frame should not be chosen to be too short: for detectors with non-negligible read-out noise (in particular CCD detectors) the count rate per frame should

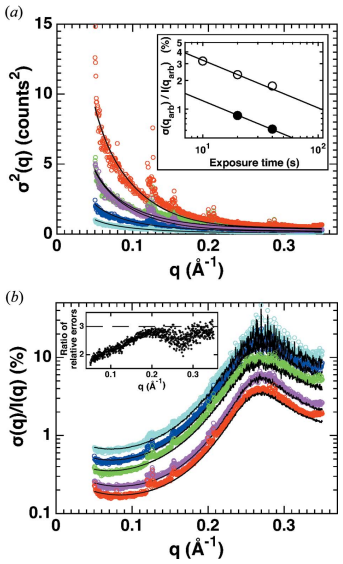


Figure 5
Dependence of SAXS measurement errors on concentration, exposure time and flux. Mean and variances of the buffer-subtracted scattering intensities were determined from repeat exposures under a range of measurement conditions at synchrotron beamline BM29, ESRF, Grenoble (coloured symbols). The absolute variances (*a*) and the relative errors (*b*) are shown. Measurement conditions (protein concentration, number of exposures and length of each exposure) were 8 mg ml^{-1}, 5 × 2 s (cyan); 8 mg ml^{-1}, 10 × 2 s (blue); 8 mg ml^{-1}, 10 × 4 s (green); 24 mg ml^{-1}, 10 × 4 s (red); 24 mg ml^{-1}, 10 × 4 s at half beam intensity (magenta). Fits of our error model with c and k as fitting parameters are shown as black solid lines. The fitting parameters are (using q_{arb} = 0.2 Å^{-1}) $I(q_{arb})$ = 12.21, k = 6134, c = 0.8750 (cyan data); $I(q_{arb})$ = 24.20, k = 5873, c = 0.8757 (blue data); $I(q_{arb})$ = 45.77, k = 5273, c = 0.8737 (green data); $I(q_{arb})$ = 142.08, k = 4822, c = 0.6894 (red data); $I(q_{arb})$ = 75.49, k = 5507, c = 0.7186 (magenta data). When fitting a straight line to the number of pixels per q bin $N(q)$, a value for k of the same order of magnitude is obtained (k = 4387; Fig. S5). The inset in panel (*a*) shows the relative error at q_{arb} = 0.2 Å^{-1} as a function of the (flux-corrected) exposure time. The solid lines are fits of a relationship $\alpha t_{exp}^{-1/2}$, which provide an excellent description of the data for both the low- (8 mg ml^{-1}; open symbols) and high-concentration (24 mg ml^{-1}; solid symbols) data. From the fit we find α(8 mg ml^{-1})/α(24 mg ml^{-1}) = 2.7, close to the concentration ratio. The inset in panel (*b*) shows the ratio of relative errors between 8 and 24 mg ml^{-1} data at the same t_{exp} (green divided by red data from the main panel), which is generally close to 3 (indicated as the dashed line), but varies with q.

not be too low; for any detector system processing (too) many frames can be cumbersome for data handling and processing. On the other hand, individual exposure times should not be chosen to be too long, either, as otherwise radiation damage might occur within one frame which is difficult to detect.

For a given t_{exp}, increasing the protein concentration can reduce the relative errors. Indeed, we find a significant reduction of the relative error for all q values at higher protein concentrations for our data set, approximately linear in protein concentration. We note, however, that the dependence of the relative error on protein concentration is complex, since changing the protein concentration will not affect the scattering profile of the buffer and the corresponding contributions to the buffer-subtracted profile (Fig. 5*b*, inset). The reduction in relative error with increasing protein concentration would suggest always measuring at the highest possible protein concentration. However, in practice, increasing the sample concentration can be challenging or inadvisable, since high protein concentration can give rise to sample aggregation or interparticle interference effects in the scattering profiles (Lipfert *et al.*, 2009; Jeffries *et al.*, 2016; Dyer *et al.*, 2014).

Our results also suggest guidelines to optimize the setup geometry for particular SAXS measurements. If, for example, the focus is on global conformational changes requiring especially low q values to determine reliable R_g values, a long sample–detector distance with a centrally arranged beam is preferable. In particular, for in-house setups with restrictions in beam intensity and detector dimensions, one should consider increasing the number of pixels for the respective q bins by positioning the detector accordingly (Fig. 3).

While the general recommendations for SAXS measurements obtained here are in line with established guidelines, we note that the quantitative analysis of variance can serve as a diagnostic for experimentalists to test and improve their measurements. Importantly, under all conditions investigated here, our model accurately describes the experimental errors with appropriately chosen parameters (Fig. 5, solid lines; parameters are given in the figure legend; see Fig. S6 and Table S1 for data on additional proteins).

3.6. Experimental errors from independent repeat measurements

So far our analysis has focused on the errors encountered in measurements of a single aliquot of sample solution, albeit consisting of multiple exposures and using properly matched buffer measurements. We note that, while it is good practice to record multiple exposures to check for radiation damage and to carry out control measurements with dilutions of the same sample solution stock to test for interparticle interference and aggregation effects, it is common to use the buffer-subtracted scattering profile from a single aliquot of sample solution in subsequent SAXS analyses. Nonetheless, it is instructive to determine the level of variation encountered in independent repeat measurements (each involving multiple exposures and buffer subtraction) of aliquots drawn from the same stock solution (sometimes called 'technical repeats') or even repeat measurements of independently prepared stock solutions (often called 'biological repeats') (Krzywinski *et al.*, 2014). In particular, for high-flux synchrotron sources where the counting errors can be minimized (Fig. 5), we expect the variability from technical repeats or, ultimately, independent biological repeats to provide a more realistic assessment of the true measurement error.

A priori, the biological variability of independently prepared solutions depends strongly on the nature of the sample and method of preparation, which is beyond the focus of this study. Here, we therefore investigate the variability observed in technical repeats using multiple aliquots of sample and buffer solution. For each q bin, we computed the mean of and the variance between scattering profiles of technical repeats, which were independently recorded, circularly averaged and buffer subtracted, for a range of biological samples (Fig. 6). Our data set includes cytochrome *c*, full-length wt myosin VI (Spink *et al.*, 2008) and *n*-dodecyl-β-D-maltoside micelles (Lipfert, Columbus *et al.*, 2007) that give rise to very distinct scattering profiles (Figs. 6*a*, 6*d* and 6*g*). We note that, while the variance observed in repeat measurements on different aliquots is still fundamentally constrained by the arguments outlined in the sections above, there can be additional contributions to the errors (*e.g.* variations in fluid handling, sample cells and synchrotron flux between the different measurements) and, *a priori*, one would expect deviations from the error model given by equation (10). Indeed, we observe relative errors from repeat measurements on independent aliquots (Figs. 6*c*, 6*f* and 6*i*) that are higher than the errors seen for measurements on single aliquots (Fig. 5). Nonetheless, we find that the errors obtained from repeat measurements on independent aliquots are still reasonably well described by the model of equation (10), when considering both the variances (Fig. 6*b*, 6*e* and 6*h*) and the relative errors (Fig. 6*c*, 6*f* and 6*i*). Here, the values of $I(q_{arb})$, c and k lose their physical interpretation, so that it is reasonable to condense $I(q_{arb})$ and c into a single constant const. and rewrite equation (10) as $\sigma^2(q) = [I(q) + \text{const.}]/(kq)$. This is the basic functional form of our model, where k and const. are treated purely as fitting parameters without any direct physical interpretation. The results (Fig. 6) suggest that the functional form of our model adequately captures the variability even for technical repeats, which, in turn, implies that our model provides a fairly general description of measurement variability for simulating measurement errors. As our model only takes into account Poisson noise and no other sources of error, this result might be surprising and raises the question of to what extent the deviations between technical repeats are still dominated by counting statistics. For ideal technical repeat measurements (without errors caused by buffer or concentration mismatch or differences in alignment of the setup or the beam) with modern noise-free hybrid pixel detectors, the remaining errors are statistical errors due to counting statistics. This suggests that significant deviations of the variances between technical repeat measurements from the functional form of our model may be used to identify systematic errors, such as buffer mismatch.

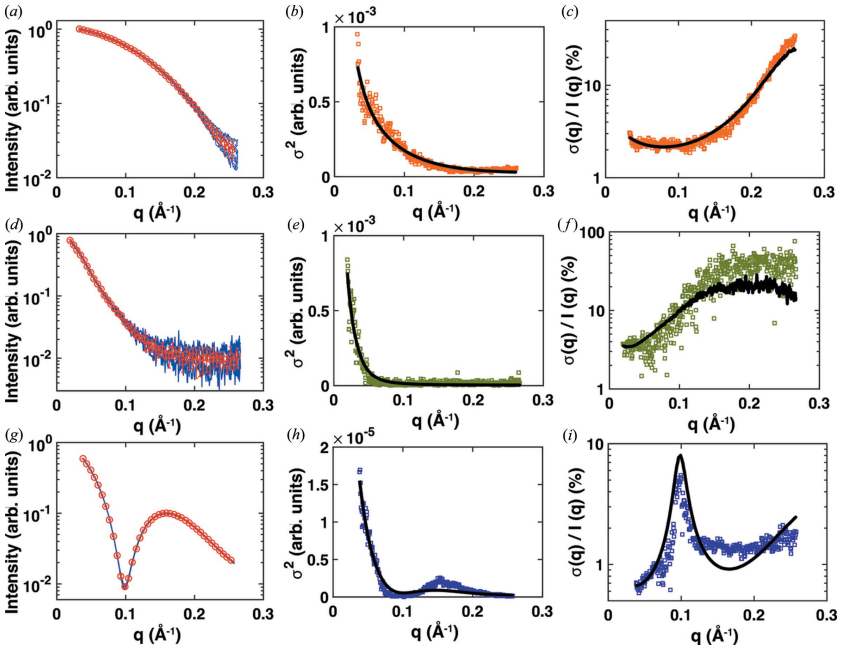


Figure 6
Errors from independent repeat measurements for a range of biological samples. SAXS measurements for 8 mg ml^{-1} cytochrome *c* (*a*)–(*c*), 0.9 mg ml^{-1} full-length wt myosin VI (*d*)–(*f*) and *n*-dodecyl-*β*-D-maltoside micelles at a detergent concentration of 45 m*M* (*g*)–(*i*). All data were collected at beamline 12ID at the Advanced Photon Source, Argonne, IL, USA, using a CCD detector (Mar CCD165), an X-ray energy of 12.0 keV and a custom-made sample environment (Lipfert *et al.*, 2006; Beno *et al.*, 2001; Lipfert, Columbus *et al.*, 2007; Spink *et al.*, 2008). The panels on the left (*a*), (*d*) and (*g*) show the individual scattering profiles as blue lines (nine profiles for cytochrome *c*, four for myosin VI and five for dodecyl-maltoside) and the mean and standard deviation for every tenth q bin as red symbols and error bars. The middle (*b*), (*e*) and (*h*) and right (*d*), (*f*) and (*i*) panels show the variances and relative errors obtained from the experimental data as symbols and the best fit of the model defined by $\sigma^2(q) = [I(q) + \text{const.}]/(kq)$ as black lines.

3.7. Recommendations for simulating errors for theoretical SAXS data

A number of increasingly powerful analysis techniques have been and are being developed to analyse SAXS data (see §1). Not only do these techniques require a precise treatment of experimental errors, but very often they rely on simulated SAXS data for testing and performance evaluation (Schindler *et al.*, 2016; Bernadó *et al.*, 2007; Pinfield & Scott, 2014; Zettl *et al.*, 2016). There are several programs to compute SAXS profiles from high-resolution structures (Svergun *et al.*, 1995; Schneidman-Duhovny *et al.*, 2010; Poitevin *et al.*, 2011; Ravikumar *et al.*, 2013; Chen & Hub, 2015). However, to simulate realistic SAXS profiles that are representative of the experimental situation, it is important to add errors to the calculated scattering profiles (Rambo & Tainer, 2013*b*). Here, we provide a concrete procedure for simulating SAXS data with realistic errors. Starting from an ideal, error-free SAXS profile computed from a high-resolution structure, we have to first scale the theoretical scattering intensity to a number of counts per q bin representative of real SAXS measurements. The best agreement between experiment and modelling is achieved using the exact number of pixels per q bin $N(q)$ to estimate the standard deviation and thus considering the exact measurement geometry (Figs. S7*e* and S7*f*). However, if experimental details are unknown or to generate 'generic' yet realistic errors, the approximation stated in equation (10) can be used. Recommended values for synchrotron and in-house measurements are provided in Table 1.

We recommend the following procedure [a MATLAB (The MathWorks Inc., Natick, MA, USA) code is given in Fig. S8]:

(i) Compute a theoretical scattering profile $I_t(q)$ from a crystal structure (using *CRYSOL*, *FoXS* or another program).

(ii) Normalize the scattering profile by dividing it by $I_t(0)$.

(iii) Multiply the scattering profile by a factor of 100 (10) to scale to a realistic number of counts for a synchrotron (in-house) setup.

(iv) Calculate the variance $\sigma_t^2(q)$ by equation (10) using k = 4500, c = 0.85 and $q_{\rm arb}$ = 0.2 Å^{-1}. [Here, $q_{\rm arb}$ is chosen for typical SAXS geometries resulting in a q range covering up to $q_{\rm max} \simeq 0.35$ Å^{-1}. The values of k and c were estimated from the experimental data set (Table S1) and can be adjusted to match sample, buffer and setup geometry.]

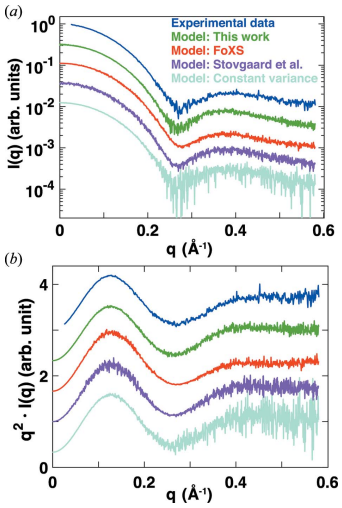


Figure 7
Comparison of experimental data and models for errors in SAXS data. (*a*) Intensity *versus* q and (*b*) Kratky representation [$qI(q)$ *versus* q] of scattering profiles for cytochrome *c*. Experimental data (Lipfert *et al.*, 2006) for 8 mg ml^{-1} cytochrome *c* in 100 m*M* acetate buffer (pH 4.6) with 0.5 *M* guanidinium hydrochloride added recorded at beamline 12ID at the Advanced Photon Source, Argonne, IL, USA are shown in blue. Simulated scattering profiles were computed using *CRYSOL* (Svergun *et al.*, 1995) (green, magenta and cyan profiles) or *FoXS* (Schneidman-Duhovny *et al.*, 2010) (red profile) from the crystal structure (Sanishvili *et al.*, 1995) of cytochrome *c* (PDB accession code 1crc). Simulated noise was added to the computed profiles using (i) the procedure described in §3.7 of this work (green); (ii) the errors from the *FoXS* webserver (red); (iii) the error model of Stovgaard *et al.* (2010), $\sigma(q) = I(q)(q + \alpha)\beta$ with the constants adjusted to $\alpha = 0.15$ and $\beta = 0.2$ (magenta); and (iv) Gaussian noise with a constant variance set at $\sigma = 0.5\%I(0)$ (cyan).

(v) Compute a scattering profile with errors $I_c(q)$ employing a random-number generator and a Gaussian distribution with mean $I_t(q)$ and standard deviation $\sigma_t(q)$.

We have simulated SAXS profiles with noise added using the procedure outlined above and have found that they closely resemble experimental data and provide a more realistic description compared with previously used error models (Fig. 7 and Fig. S7).

4. Summary

We have derived a new error model for SAXS data which incorporates the measurement process and the setup geometry, and thereby correctly describes the magnitude and scaling of the measurement errors. We have demonstrated its broad applicability to a range of samples, setups, exposure times and sample concentrations. Moreover, we provide guidelines on how to employ the model to simulate uncertainties and to model realistic noise for theoretical scattering profiles. The theoretical scattering profiles simulated using our protocol closely resemble experimental data and we expect our model to be widely applicable to generate synthetic test data sets for the validation of new SAXS modelling approaches. As our model is based on a few simple and general assumptions, we anticipate that similar arguments can be applied to other techniques that employ hybrid photon couting detectors such as correlated X-ray scattering (Mendez *et al.*, 2014).

Acknowledgements

We thank Dr Adam Round for support at beamline BM29 at the ESRF in Grenoble, Dr Sönke Seifert for support with measurements at beamline 12ID at Argonne National Laboratory, and Stefan Fischer and Bert Nickel for useful discussions. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences. The authors acknowledge the Center for Nanoscience (CeNS) and DFG-SFB 1032 (Nanoagents) for financial support.

References

Beno, M. A., Jennings, G., Engbretson, M., Knapp, G. S., Kurtz, C., Zabransky, B., Linton, J., Seifert, S., Wiley, C. & Montano, P. A. (2001). *Nucl. Instrum. Methods Phys. Res. Sect. A*, **467–468**, 690–693.
Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). *J. Am. Chem. Soc.* **129**, 5656–5664.
Blanchet, C. E. & Svergun, D. I. (2013). *Annu. Rev. Phys. Chem.* **64**, 37–54.
Broennimann, Ch., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2006). *J. Synchrotron Rad.* **13**, 120–130.
Bruetzel, L. K., Fischer, S., Salditt, A., Sedlak, S. M., Nickel, B. & Lipfert, J. (2016). *Rev. Sci. Instrum.* **87**, 025103.
Chen, P. C. & Hub, J. S. (2015). *Biophys. J.* **108**, 2573–2584.
Dyer, K. N., Hammel, M., Rambo, R. P., Tsutakawa, S. E., Rodic, I., Classen, S., Tainer, J. A. & Hura, G. L. (2014). *Methods Mol. Biol.* **1091**, 245–258.
Förster, F., Webb, B., Krukenberg, K. A., Tsuruta, H., Agard, D. A. & Sali, A. (2008). *J. Mol. Biol.* **382**, 1089–1106.
Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.
Grishaev, A., Wu, J., Trewhella, J. & Bax, A. (2005). *J. Am. Chem. Soc.* **127**, 16621–16628.
Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.
Henrich, B., Bergamaschi, A., Broennimann, C., Dinapoli, R., Eikenberry, E. F., Johnson, I., Kobas, M., Kraft, P., Mozzanica, A. & Schmitt, B. (2009). *Nucl. Instrum. Methods Phys. Res. Sect. A*, **607**, 247–249.
Hopkins, J. B. & Thorne, R. E. (2016). *J. Appl. Cryst.* **49**, 880–890.
Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., Tsutakawa, S. E., Jenney, F. E. Jr, Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S., Scott, J. W., Dillard, B. D., Adams, M. W. W. & Tainer, J. A. (2009). *Nat. Methods*, **6**, 606–612.
Jeffries, C. M., Graewert, M. A., Blanchet, C. E., Langley, D. B., Whitten, A. E. & Svergun, D. I. (2016). *Nat. Protoc.* **11**, 2122–2153.
Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2016). *J. Appl. Cryst.* **49**, 953–960.
Krzywinski, M., Altman, N. & Blainey, P. (2014). *Nat. Methods*, **11**, 977–978.
Lipfert, J., Chu, V. B., Bai, Y., Herschlag, D. & Doniach, S. (2007). *J. Appl. Cryst.* **40**, s229–s234.
Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A. & Doniach, S. (2007). *J. Phys. Chem. B*, **111**, 12427–12438.

Lipfert, J., Das, R., Chu, V. B., Kudaravalli, M., Boyd, N., Herschlag, D. & Doniach, S. (2007). *J. Mol. Biol.* **365**, 1393–1406.
Lipfert, J. & Doniach, S. (2007). *Annu. Rev. Biophys. Biomol. Struct.* **36**, 307–327.
Lipfert, J., Herschlag, D. & Doniach, S. (2009). *Methods Mol. Biol.* **540**, 141–159.
Lipfert, J., Millett, I. S., Seifert, S. & Doniach, S. (2006). *Rev. Sci. Instrum.* **77**, 046108.
Mendez, D., Lane, T. J., Sung, J., Sellberg, J., Levard, C., Watkins, H., Cohen, A. E., Soltis, M., Sutton, S., Spudich, J., Pande, V., Ratner, D. & Doniach, S. (2014). *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130315.
Moore, P. B. (1980). *J. Appl. Cryst.* **13**, 168–175.
Pernot, P. *et al.* (2013). *J. Synchrotron Rad.* **20**, 660–664.
Pinfield, V. J. & Scott, D. J. (2014). *PLoS One*, **9**, e95664.
Poitevin, F., Orland, H., Doniach, S., Koehl, P. & Delarue, M. (2011). *Nucleic Acids Res.* **39**, W184–W189.
Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). *Q. Rev. Biophys.* **40**, 191–285.
Rambo, R. P. & Tainer, J. A. (2013*a*). *Annu. Rev. Biophys.* **42**, 415–441.
Rambo, R. P. & Tainer, J. A. (2013*b*). *Nature*, **496**, 477–481.
Ravikumar, K. M., Huang, W. & Yang, S. (2013). *J. Chem. Phys.* **138**, 024112.
Sanishvili, R., Volz, K. W., Westbrook, E. M. & Margoliash, E. (1995). *Structure*, **3**, 707–716.
Schindler, C. E., de Vries, S. J., Sasse, A. & Zacharias, M. (2016). *Structure*, **24**, 1387–1397.
Schneidman-Duhovny, D., Hammel, M. & Sali, A. (2010). *Nucleic Acids Res.* **38**, W540–W544.
Spink, B. J., Sivaramakrishnan, S., Lipfert, J., Doniach, S. & Spudich, J. A. (2008). *Nat. Struct. Mol. Biol.* **15**, 591–597.
Stovgaard, K., Andreetta, C., Ferkinghoff-Borg, J. & Hamelryck, T. (2010). *BMC Bioinformatics*, **11**, 429.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Svergun, D. I. & Koch, M. H. J. (2003). *Rep. Prog. Phys.* **66**, 1735–1782.
Tuukkanen, A. T., Kleywegt, G. J. & Svergun, D. I. (2016). *IUCrJ*, **3**, 440–447.
Vachette, P., Koch, M. H. & Svergun, D. I. (2003). *Methods Enzymol.* **374**, 584–615.
Walther, D., Cohen, F. E. & Doniach, S. (2000). *J. Appl. Cryst.* **33**, 350–363.
Zettl, T., Mathew, R. S., Seifert, S., Doniach, S., Harbury, P. A. & Lipfert, J. (2016). *Nano Lett.* **16**, 5353–5357.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information

JOURNAL OF APPLIED CRYSTALLOGRAPHY

Volume 50 (2017)

Supporting information for article:

Quantitative evaluation of statistical errors in small-angle X-ray scattering measurements

Steffen M. Sedlak, Linda K. Bruetzel and Jan Lipfert

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-1

S1. Detailed derivation of our error model

Equation 2 follows from the general consideration that for a set of independent measurements, the standard error of the mean is given by

$$\bar{\sigma} = \sigma/\sqrt{N}$$

In our case the counts n_i in every pixel i belonging to the same q-bin are considered as $N(q)$ independent measurements. Since the buffer or sample scattering intensity $I_{s/b}(q)$ for a certain scattering angle q are calculated by taking the mean of the photons recorded by all $N(q)$ pixels belonging to the same q-bin, the corresponding errors $\sigma_{s/b}(q)$ and variances $\sigma^2_{s/b}(q)$ can be calculated as

$$\sigma_{s/b}(q) = \sigma_i/\sqrt{N(q)}$$

$$\sigma_{s/b}{}^2(q) = \sigma^2{}_i/N(q)$$

Here, we assumed that all pixels belonging to the same q-bin have the same variance σ_i. As this might not be the case, the best approximation is to average over all variances σ_i, so that

$$\sigma_{s/b}{}^2(q) = 1/N(q)\, 1/N(q) \sum_{i=1}^{N(q)} \sigma^2{}_i$$

While we employ this equation to derive our model, we use the expression

$$\sigma_i{}^2 = (n_i - I(q))^2$$

to experimentally determine the variance σ_i^2, i.e. by squaring the difference between the counts in a certain pixel and the mean intensity in the corresponding q-bin:

In the derivation of Equation 8, the following steps are used:

$$I_s(q_{arb}) = I(q_{arb}) + I_b(q_{arb})$$

$$I_s(q_{arb}) = I(q_{arb}) + c\, I_s(q_{arb})$$

$$I(q_{arb}) = I_s(q_{arb}) - c\, I_s(q_{arb})$$

$$I(q_{arb}) = (1 - c\,)\, I_s(q_{arb})$$

$$I_s(q_{arb}) = \frac{I(q_{arb})}{(1-c)}$$

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-2

Table S1 Experimentally fitted parameters. Experimentally determined fit parameters for *k*, *c* and $I(q_{arb})$ for different measurement conditions at in-house and synchrotron setups are determined by fitting Equation 10 to the deviations between intensities determined from several frames of sample and buffer measurements on which circular averaging has been performed individually, i.e. the variance between repeat exposures. While *c* and $I(q_{arb})$ are well-defined, the free parameter *k* scatters more broadly, because it incorporates many different parameters (Equation 9).

Experimental fits (q_{arb} = 0.2 Å^{-1})

Synchrotron (λ = 0.9919 Å)	conc. (mg/ml)	exp. time (s)	frames	*I(q)* (counts)	*k* (Å)	$I(q_{arb})$ (counts)	$c(q_{arb})$	$I_b(q)^*$
Lysozyme	5	1	10	1-100	2756	3.5	0.87	46.8
L_{sd} = 2.864 m	10	1	10	1-100	2593	6.5	0.79	48.9
(multibunch)	20	1	10	1-100	2563	13.0	0.65	48.3
Cytochrome *c*	2	4	10	1-100	4551	1.2	0.97	77.6
L_{sd} = 2.872 m	8	4	9	1-100	4274	4.5	0.87	60.2
(multibunch)	24	4	9	1-100	5520	14.1	0.69	62.8
Cytochrome *c*	2	4	10	0.1-10	6858	0.5	0.97	32.3
L_{sd} = 2.872 m	8	4	10	0.1-10	6327	2.4	0.89	38.8
(low bunch)	24	4	10	0.1-10	5213	7.5	0.72	38.6
In-house (λ = 0.7085 Å)	conc. (mg/ml)	exp. time (s)	frames	*I(q)* (counts)	*k* (Å)	$I(q_{arb})$ (counts)	$c(q_{arb})$	$I_b(q)^*$
Cytochrome *c*	2	7200	4	0.01-1	2572	-0.17	1.06	6.0
L_{sd} = 1.109 m	8	7200	4	0.1-10	3363	0.48	0.87	6.4
	24	7200	3	1-100	5934	1.27	0.71	6.2
Lysozyme	5	7200	5	0.1-10	3284	0.33	0.92	7.6
L_{sd} = 1.109 m	10	7200	4	1-100	4133	0.59	0.87	7.9
	20	7200	5	1-100	4833	1.12	0.71	5.5
BSA	5	7200	4	0.1-10	3663	0.43	0.91	8.7
L_{sd} = 1.109 m								

Synchrotron data were taken in multibunch mode (170 mA beamcurrent) and low bunch mode (90 mA beamcurrent).

*The buffer level $I_b(q)$ is calculated by $(2\,c\,I(q_{arb})) / (1-c)$.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-3

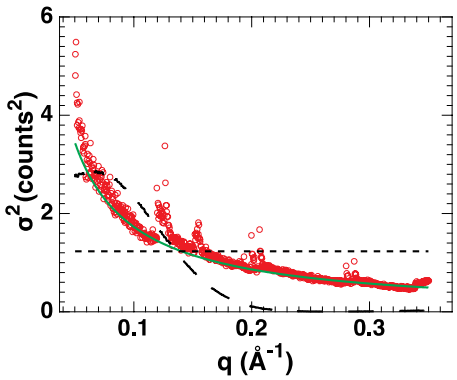


Figure S1 Comparison of different error models. Measurement errors obtained from repeat exposures of cytochrome *c* (8 mg/ml, 10 frames of 4 s exposure time each; BM29, ESRF, Grenoble) are fitted with different error models for comparison. The variance of the intensity determined from the 10 exposures is shown as red circles. The model from (Stovgaard *et al.*, 2010) with fitted values $\alpha = 0.0520$, $\beta = 0.0279$ is shown as the black dashed line. The best fitting constant variance $\sigma^2 = 1.232$ is shown as a black dotted line. The best fit of our new model with $k = 3681$, $I(q_{arb}) = 45.77$, $c = 0.8737$, $q_{arb} = 0.2$ Å^{-1} is shown as the green line.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-4

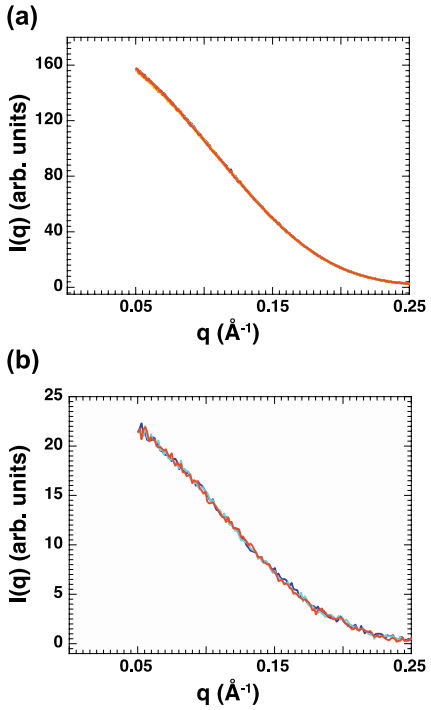


Figure S2 Absence of radiation damage. Overlay of profiles of repeat exposures of cytochrome *c* (24 mg/ml) recorded at (a) a synchrotron source (BM29, ESRF, Grenoble) and (b) at our in-house source (Department of Physics, LMU Munich). No significant differences between the scattering profiles, in particular at small angles, are observed. This confirms the absence of radiation damage. For the other data used in this study the absence of radiation damage was confirmed in the same way.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-5

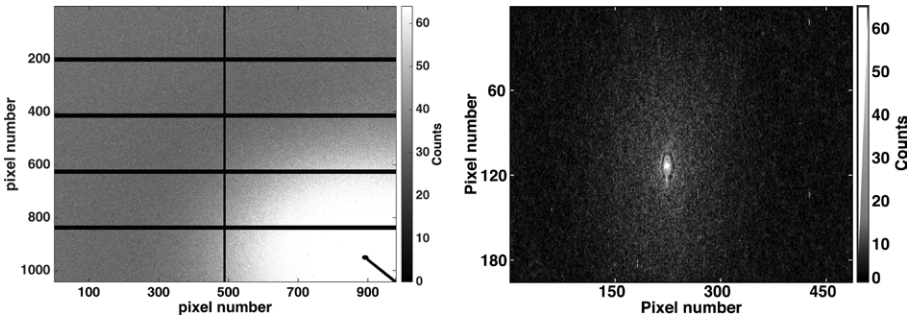


Figure S3 Typical 2D detector images from synchrotron and our in-house setup. (a) Scattering pattern of cytochrome *c* (8 mg/ml) recorded at beam line BM29, ESRF, Grenoble. The detector (Pilatus 1M) is built up from ten detector arrays. In between, pixels are missing, resulting in slightly increased measurement errors for certain angles (Figure 2). The shadow of beamstop and beamstop mounting in the lower right corner is masked out. (b) Scattering pattern of cytochrome *c* (8 mg/ml) using a Pilatus 100k detector at an in-house setup at the Department of Physics, LMU Munich. The bright spot and the dark ring in the very middle of the image are caused by the direct beam and the semitransparent beamstop and are masked out when processing the data.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-6

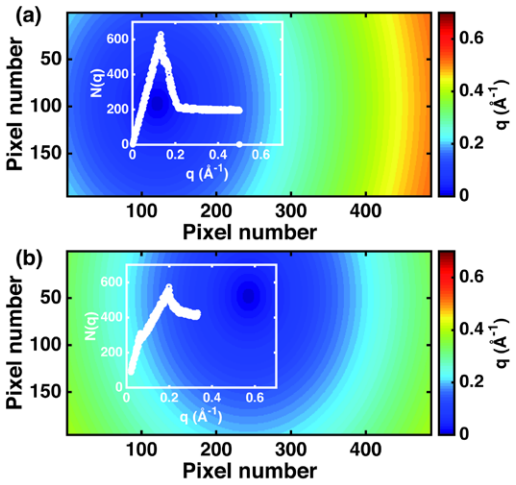


Figure S4 Assignment of pixels to *q*-bins. Depending on the position of the beam center on the detector, different pixels are assigned to different *q*-bins. The number of pixels per *q*-bin and the size of the *q*-range can be varied by rearranging the detector position. The corresponding numbers of pixel per *q*-bin *N(q)* are shown in the insets.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-7

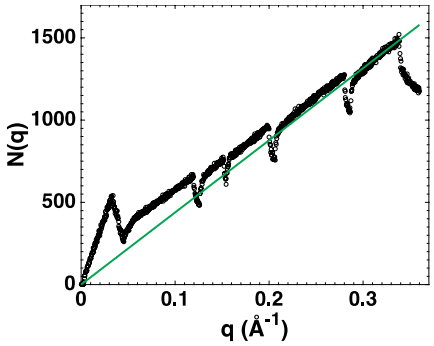


Figure S5 Number of pixels per *q*-bin *N(q)* for a synchrotron measurement. The number of pixels per *q*-bin *N(q)* (black circles) for the measurements shown in Figure 5. The detector image is depicted in Supplementary Figure S3a Missing pixels between the different detector arrays result in a decrease of number of pixels *N(q)* for certain scattering angles *q*. The green line is the best fit of the form $N(q) = k\,q$ with $k = 4387$. The value of k is in good agreement with the values obtained from fitting the variances (Figure 5).

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-8

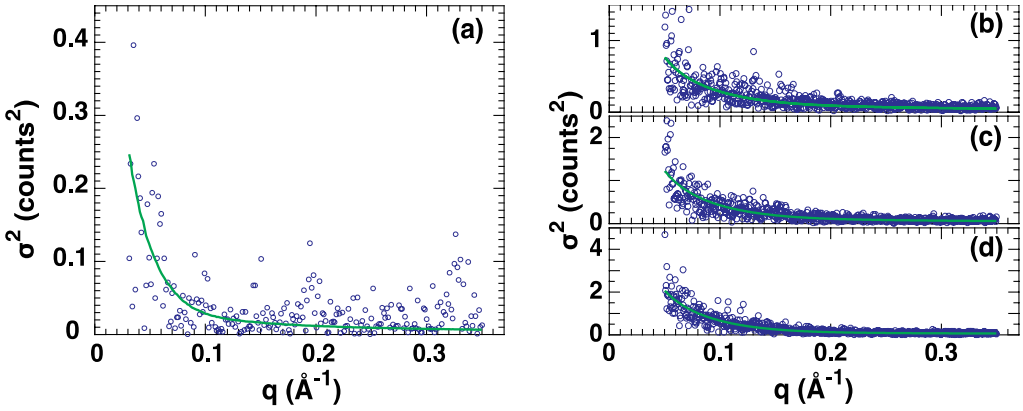


Figure S6 Analysis of SAXS measurement errors for bovine serum albumin and lysozyme. Measurement errors of (a) BSA SAXS data obtained at our in-house setup at the Department of Physics, LMU Munich and (b)-(d) lysozyme SAXS data obtained at a synchrotron beamline (BM29, ESRF, Grenoble) are fitted with our model (similar to the analysis presented in Figure 4 of the main text). Blue circles: variance computed from 4 repeat exposures (of 2 h each) using 5 mg/ml BSA (a) and 10 repeat exposures (of 1.0 s each) using 5 mg/ml (b), 10 mg/ml (c) and 20 mg/ml (d) lysozyme. Green lines: fits using Equation 10. The corresponding fit parameters are listed in Table S1.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-9

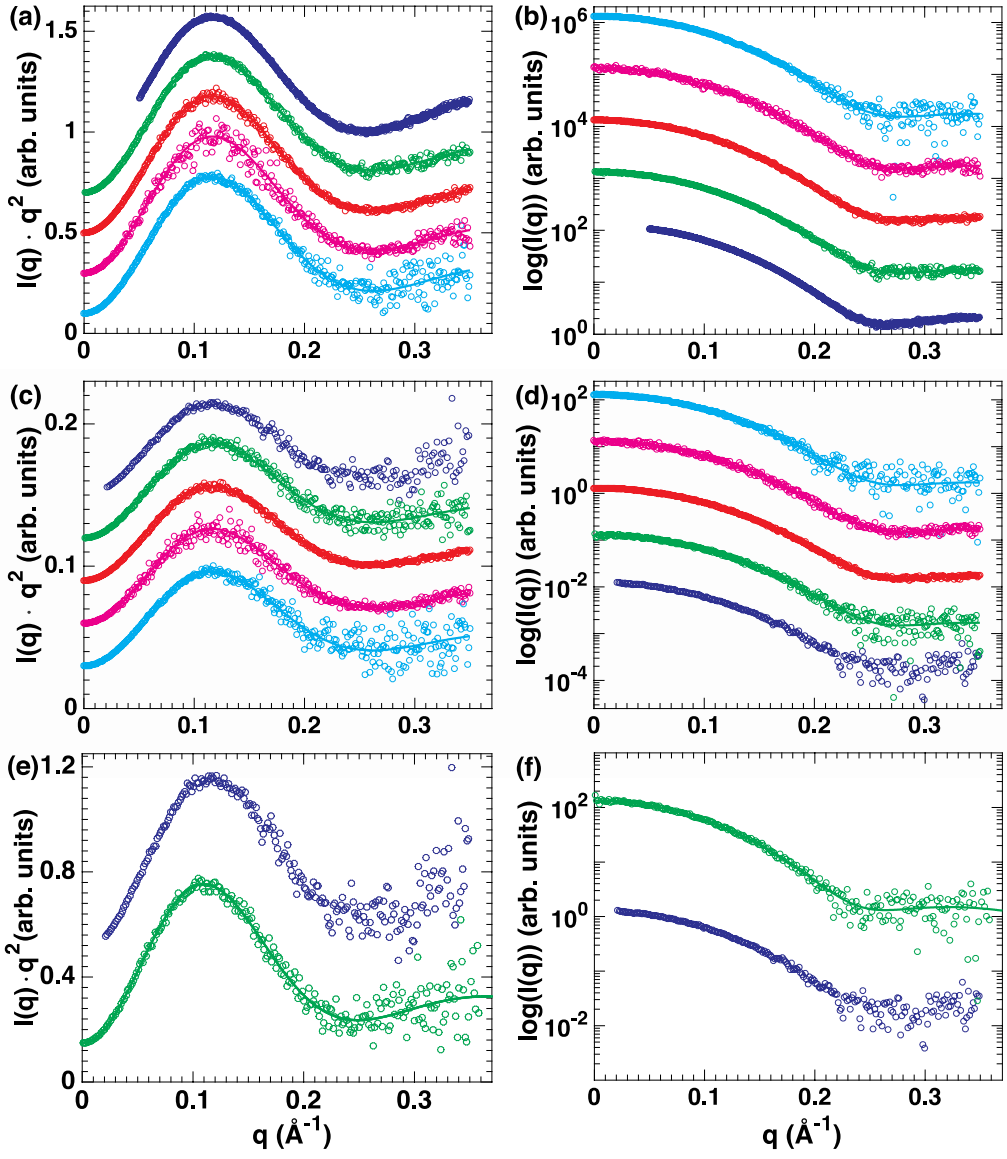


Figure S7 Comparison of noise models for simulated scattering profiles. Simulated and experimental SAXS profiles for lysozyme. Dark blue data in all panels are experimental data measured using 10 mg/ml lysozyme (see Materials and Methods for details). Data in panel (a) and (b) are for synchrotron meaurements (10 frames of 1 s exposure time each; BM29, ESRF, Grenoble). Data in panels (c) - (f) are for in-house data (4 frames of 2 h exposure time each; Department of Physics, LMU

J. Appl. Cryst. (2017). **50**, doi:10.1107/S1600576717003077 Supporting information, sup-10

Munich). Panels (a), (c), and (e) show the data in Kratky representation; panels (b), (d), and (f) as log(I) vs. *q*. All profiles are scaled and vertically offset for clarity. The solid lines are theoretical scattering profile computed from the crystal structure (PDB ID: 6LYZ) using FoXS (Schneidman-Duhovny *et al.*, 2010) ((a)-(d)) and CRYSOL (Svergun *et al.*, 1995) ((e) and (f)). The intensities were scaled to mimic experimentally encountered values according the protocol outlined in the main text. The circles color matched to the solid lines are calculated scattering profiles with simulated noise added. They were created by taking a random number for every *q*-bin from a normal distribution with mean $I(q)$ and standard deviation $\sigma(q)$ according to the different error models: 1) Constant standard deviation $\sigma(q) = 0.005 \cdot I(0)$ shown in cyan; 2) Stovgaard's model (Stovgaard *et al.*, 2010) with $\sigma(q) = I(q) \cdot (q + 0.15) \cdot 0.3$ shown in magenta; 3) the variance provided by the program FoXS in red; 4) the new model derived in this work in green. Simulated data using our new model in panel (a) and (b) used $k = 4500$ and $c = 0.85$; simulated data in panel (c) and (d) used $k = 4500$ and $c = 0.90$; simulated data in (e) and (f) used the exact number of pixels per *q*-bin $N(q)$ and $c = 0.90$. The model with constant variance (cyan) tends to underestimate the error at low *q* and/or overestimate the error at high *q*. The model by Stovgaard *et al.* (magenta) and the FoXS model (red) tend to overestimate the errors at intermediate *q* compared to the level of scatter at high *q*.

J. Appl. Cryst. (2017). 50, doi:10.1107/S1600576717003077 Supporting information, sup-11

```
% read in the data so that the scattering angle is
% stored in q and the the scattering intensity in I

k=4500; c=0.85;
I=(100.*I)./(I(1,1);
[~, idx]=min(abs(0.2-q));
Iarb=I(idx);
s=sqrt((1./(k.*q)).*(I+(2*c*Iarb)./(1-c)));
Ie=normrnd(I,s);
```

Figure S8 MATLAB code to simulate realistic errors onto theoretical profiles. Providing the momentum transfer vector *q* and the respective theoretical scattering intensities in a vector *I*, this code can be employed to model realistic noise. The scattering intensities with errors added are stored in the vector *Ie*, the standard error in the vector *s*. The parameters shown are for a typical synchrotron measurement.

6.2 Publication P5: Conformational Changes and Flexibility of DNA Devices Observed by Small-Angle X-ray Scattering

Conformational Changes and Flexibility of DNA Devices Observed by Small-Angle X-ray Scattering

by

Linda K. Bruetzel, Thomas Gerling, Steffen M. Sedlak, Philipp U. Walker, Wenjun Zheng, Hendrik Dietz, Jan Lipfert

published in

Nano Letters 2016, 16(8), 4871-9

NANO LETTERS
Letter
pubs.acs.org/NanoLett

Conformational Changes and Flexibility of DNA Devices Observed by Small-Angle X-ray Scattering

Linda K. Bruetzel,[†] Thomas Gerling,[‡] Steffen M. Sedlak,[†] Philipp U. Walker,[†] Wenjun Zheng,[§] Hendrik Dietz,[‡] and Jan Lipfert*[,†]

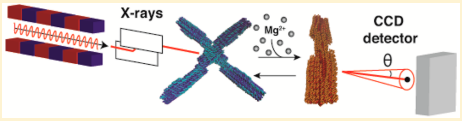
[†]Department of Physics, Nanosystems Initiative Munich, and Center for Nanoscience, LMU Munich, Amalienstrasse 54, 80799 Munich, Germany

[‡]Physik Department, Walter Schottky Institute, Technische Universität München, Am Coulombwall 4a, 85748 Garching near Munich, Germany

[§]Physics Department, State University of New York at Buffalo, Buffalo, New York 14260, United States

Supporting Information

ABSTRACT: Self-assembled DNA origami nanostructures enable the creation of precisely defined shapes at the molecular scale. Dynamic DNA devices that are capable of switching between defined conformations could afford completely novel functionalities for diagnostic, therapeutic, or engineering applications. Developing such objects benefits strongly from experimental feedback about conformational changes and 3D structures, ideally in solution, free of potential biases from surface attachment or labeling. Here, we demonstrate that small-angle X-ray scattering (SAXS) can quantitatively resolve the conformational changes of a DNA origami two-state switch device as a function of the ionic strength of the solution. In addition, we show how SAXS data allow for refinement of the predicted idealized three-dimensional structure of the DNA object using a normal mode approach based on an elastic network model. The results reveal deviations from the idealized design geometries that are otherwise difficult to resolve. Our results establish SAXS as a powerful tool to investigate conformational changes and solution structures of DNA origami and we anticipate our methodology to be broadly applicable to increasingly complex DNA and RNA devices.



KEYWORDS: *DNA origami, small-angle X-ray scattering, SAXS, normal modes, conformational changes*

A fundamental aim of nanotechnology is to design synthetic objects that can adopt specific conformational states and carry out functions at the molecular scale, for example, in transport, signal transduction, or molecular circuitry. Molecular self-assembly of DNA is a particularly successful approach towards creating versatile structures at the nanometer scale.[1−4] When using the DNA origami technique, a several kilobase long circular single-stranded scaffold strand is folded into custom target shapes with the assistance of hundreds of short single-stranded staple strands. By exploiting the specificity of DNA base pairing, precisely controlled shapes reaching over 100 nm in size and molecular weights of several MDa can be created.[2,5−8]

While an important initial focus in the design of self-assembled DNA structures was to create static objects of well-defined shapes,[1−3,6] more complex functions require dynamic 3D nanostructures that can undergo controlled conformational changes. Examples of dynamic DNA origami structures include a DNA box with a closable lid,[9] a DNA nanorobot,[10] a reconfigurable plasmonic nanostructure,[11] or a DNA tweezers.[12] Such dynamic DNA structures are promising candidates for applications ranging from nanoengineering[13,14] to medical diagnostics and therapeutics.[15,16] An important challenge in this context is the precise control over the 3D shape and mechanical flexibility of the target design in solution to achieve desired functionality.

So far, structural characterization of DNA origami structures has predominantly relied on atomic force microscopy (AFM) imaging[17,18] and negative-stain transmission electron microscopy (TEM).[6,19−21] While these techniques are well suited to image static structures, they both rely on immobilizing samples on a surface and involve steps such as drying or staining the samples, which renders the solution conditions far from physiological. Cryo-electron microscopy provides less harsh conditions and has recently been successfully applied to DNA origami structures[9,22] but still requires immobilized samples embedded in vitrified ice, potentially biasing the conformation of the sample and making it difficult to detect conformational changes upon variation in solution conditions.

In contrast, small-angle X-ray scattering (SAXS) can probe molecular conformations and transitions and provides low-resolution structural information on molecules and molecular assemblies in solution.[23,24] As SAXS can operate under virtually arbitrary solution conditions,[25,26] the technique is ideally suited

Received: March 30, 2016
Revised: June 28, 2016
Published: June 29, 2016

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

to detect conformational changes triggered by changes in solution environment, such as ionic strength, denaturant, temperature, or ligand binding. SAXS has proven very powerful to detect the large structural changes associated with the folding of proteins[27−29] and nucleic acids[30−32] but can also readily detect more subtle conformational changes, for example, triggered by the binding of small-molecule ligands.[33−36]

Recently, Gerling et al.[37] established a framework based on shape-complementary recognition for the programmable and reversible assembly and disassembly of complex 3D shapes built from DNA. One important example involves a dynamic "switch" device, where multiple weak base stacking interactions were exploited to change conformations between a closed and an open state as a function of temperature or ionic strength of the solution. Here, we use small-angle X-ray scattering (SAXS) to probe the structure and conformational changes of the switch device in solution. In particular, we detect and evaluate the conformational changes upon variations in solution conditions and observe quantitative agreement with solution-based Förster resonance energy transfer (FRET) measurements. In addition, we refine the 3D structure of the switch objects against the scattering data using a normal mode based fitting procedure and find evidence for swelling and structural rearrangements away from idealized DNA helix geometries. Together, our results establish SAXS as a powerful technique to probe the structures and conformations of DNA devices.

DNA Origami Switch Samples Based on Base Stacking Interactions. We performed SAXS measurements on three different variants of a DNA origami switch object that is based on shape-complementarity and base stacking interactions.[37] A dynamic variant of the switch (switch D) can undergo conformational changes between an x-shaped open and a rectangular-shaped closed state (Figure 1; see Supporting Information and Supplementary Figures S1−S3 for details). This variant consists of two rigid bundles of DNA double helices arranged in a honeycomb lattice that form the two arms of the structure. The arms are connected in the middle of the structure by a single Holliday junction that acts as a pivot point for the rotational degree of freedom (Figure 1). The structure of the closed state is prescribed by shape-complementary patterns of double helical protrusions (red domains, Figure 1) and recessions (blue domains, Figure 1) that can precisely dock into each other when the two arms of the switch object come close together. The closed state is stabilized by up to 16 short-range stacking interactions of the terminal bases of shape-complementary surface topographies. The conformational equilibrium sensitively depends on ambient conditions such as the salt concentration or the temperature of the solution.[37] TEM images of the switch D variant show that at low salt concentrations the great majority of structures assumes the open state, while at high salt concentrations switch D particles predominantly populate the closed state.[37] As reference structures, we employed two static variants of the switch object that are permanently locked in the open and closed states, respectively. In the static closed variant (switch C) stacking interactions are replaced by stronger hybridization interactions of 3-bases-long single-stranded overhangs of corresponding staple strands holding the two arms of the switch object in the closed conformational state (Figure 1, right). In the static open variant (switch O) all stacking interactions are deactivated and the two arms are connected by additional crossovers holding them at an opening angle of ∼90° (Figure 1, left).

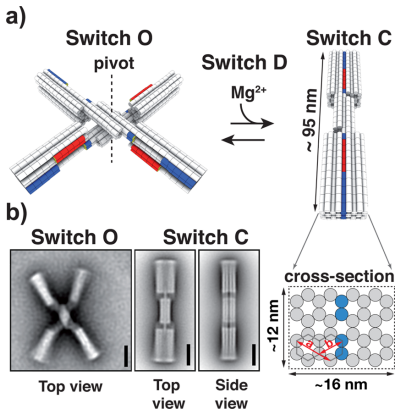


Figure 1. Illustrations of static and dynamic switch devices. (a) Schematics of the switch devices used in this study. The dynamic switch object (switch D) changes from an x-shaped open to a rectangular-shaped closed conformation upon addition of magnesium ions. Shape-complementary protrusions and recessions are indicated by the red and blue DNA double helical domains, respectively. Static switch variants are locked in the open (switch O, left) and closed (switch C, right) state. The schematic of the cross-sectional area of switch C indicates the horizontal and vertical dimensions including interhelical distances of a = 6 nm and b = 4 nm, which give rise to a peak in the scattering profiles of switch C and switch O. (b) Corresponding average negative-stain TEM micrographs of switch O in the presence of 5 mM $MgCl_2$ and of switch C at a $MgCl_2$ concentration of 25 mM. Scale bars, 20 nm.

DNA Origami Structures Give Rise to High Signal-to-Noise SAXS Profiles at 25−100 nM Concentrations. To estimate the minimum concentrations required for synchrotron-based SAXS measurements on our large (∼16 000 nucleotides (nt) or ∼5 MDa) DNA origami structures, we used prior SAXS data of smaller nucleic acids in combination with extrapolation based on a scaling relationship (see Supporting Information). We analyzed the concentrations used for SAXS measurements that resulted in a sufficient signal-to-noise ratio for structural analyses (which we loosely define as analyses that go beyond Guinier fitting of the lowest q values) for a range of nucleic acid samples (Figure 2a, blue symbols). The data set ranges from an 8 nt DNA[38] to a large (∼400 nt) ribozyme[31] and includes both RNA[35,36,39,40] and DNA samples,[41,42] as well as data for an ∼14 knt DNA origami structure[9] recorded at an in-house X-ray source. The data are well described by a scaling relationship of the form $c \sim MW^{-\nu}$, where c is the required concentration, MW is the molecular weight, and the scaling exponent ν was fitted to be ν = 1.30 (Figure 2a, dashed line, and Supporting Information). The scaling relation predicts that concentrations of ∼10−50 nM are sufficient to obtain a good scattering signal for an ∼16 knt DNA structure. Experimentally, we indeed obtained good signal-to-noise scattering profiles for concentrations as low as 25 nM of the DNA origami structures (Figure 2b and Supplementary Figure S4) that are in excellent agreement with the predicted scaling relationship (Figure 2a, red star). Additional measurements at 50 and 100 nM concentration display even higher signal-to-noise ratios (especially in the

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

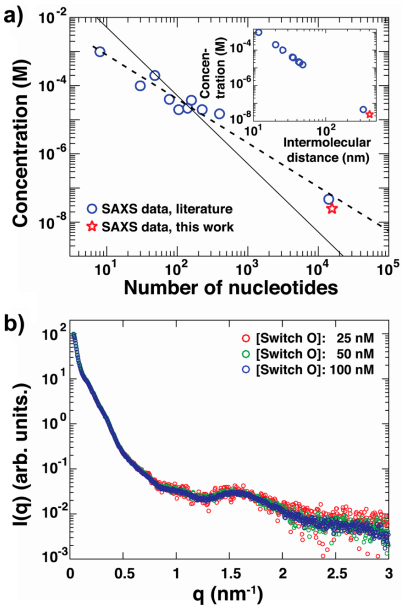


Figure 2. Concentration requirements and SAXS signals for large DNA origami structures. (a) Concentrations required to obtain a suitable SAXS signal, as a function of molecule size (in number of nucleotides), for a range of nucleic acid samples investigated previously (blue circles). The solid line is a fit of the relationship a/nt^2, where nt is the number of nucleotides and a is a fitting constant. The dashed line is a fit of the relationship b/nt^{ν} where b and ν are fitting constants. From the best fit we find $\nu \sim 1.30$. The red star corresponds to measurements of the DNA origami switch samples in this study that were guided by the scaling behavior. Inset: Intermolecular distances calculated for the required SAXS concentrations of the different nucleic acid samples. (b) Averaged scattering profiles of the switch O measured at three different concentrations: 25 nM (red circles), 50 nM (green circles), 100 nM (blue circles). Data are scaled by concentration.

higher q-range) and are superimposable after scaling by concentration, indicating the absence of aggregation, radiation damage or interparticle interference (Figure 2b and Supplementary Figure S4). Interparticle interference effects occur if the particles in solutions are, on average, sufficiently close to interact, for example, via excluded volume or electrostatic effects. We note that interparticle interference effects are expected to be (even) weaker for larger macromolecular assemblies as the typical intermolecular distances increase for higher molecular weights due to the lower required concentration (Figure 2a, inset). For instance, the average intermolecular distance of a 24 bp DNA sample (radius of gyration $R_g \sim 2.4$ nm) measured at a concentration of 0.2 mM is around 20 nm, whereas for our DNA origami objects ($R_g \sim$ 28 nm, see below) measured at 25 nM it is around 400 nm. Remarkably, due to their large size, the DNA origami objects give rise to scattering profiles with a dynamic range of ≥4 orders of magnitude in intensity with features identifiable up to $q \approx 3$ nm^{-1}. We note that while the scaling argument and extrapolation shown here only provide a rough estimate of the required sample concentrations, we anticipate that it can provide a useful guideline to other SAXS experiments on nucleic acid assemblies as well.

SAXS Reveals Structural Features of DNA Origami Objects. For a first structural characterization, we analyzed the scattering data from the static switch samples, which serve as reference samples for the dynamic switch variant. When comparing the scattering profiles of the switch O and switch C samples, we observe significant differences in the q-range below $q < 0.5$ nm^{-1} (Figure 3a), which is in line with global

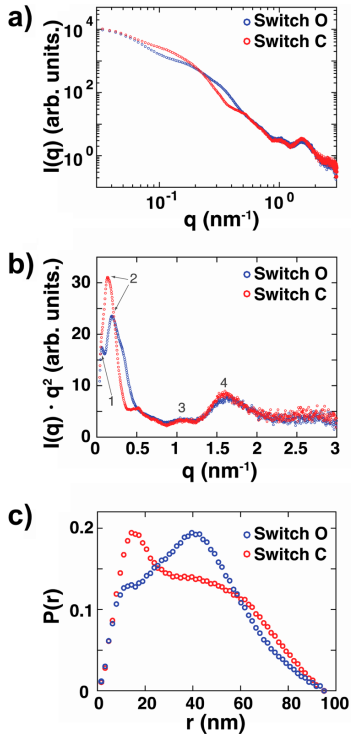


Figure 3. SAXS data reveal conformational features for open and closed switch objects. (a) Double-logarithmic representation of scattering intensity profiles obtained from the switch O and switch C sample. (b) Kratky representation of the data from (a) scaled by a constant factor. Numbers indicate peaks, which are described in the main text. (c) Pair distance distribution function $P(r)$ calculated from data shown in (a) assuming a maximum particle dimension D_{max} of 95 nm. $P(r)$ functions are normalized to equal areas.

structural differences in the open and closed states. For higher q, corresponding to smaller length scales, the scattering curves largely coincide, exhibiting two distinct peaks.

We performed a Guinier analysis of the scattering profiles in the low q-range to determine the overall radii of gyration (R_g) (see Supporting Information and Supplementary Figure S5). We obtained an R_g of (27.9 ± 0.1) nm for the switch C (Table

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

1), which is in excellent agreement with a theoretical R_g of 28 nm, derived from approximating the closed switch as a rectangular beam, with $R_g^{theo} = 1/3[(W/2)^2 + (H/2)^2 + (L/2)^2]^{1/2}$, where W, H, and L are the width, height, and length of the object, respectively (Figure 1). For the switch O sample, we found an average R_g of (29.0 ± 0.2) nm (Table 1), overall similar to switch C, which is expected as the approximate distances from the center of mass are conserved upon the transition from the closed to the open state.

Table 1. Comparison of the Radius of Gyration (R_g) and the Cross-Sectional R_g (R_c) for the Static and Dynamic Versions of the Switch Object Derived from Experimental and Theoretical Scattering Profiles[a]

Sample	R_g (nm)	R_c (nm)
Switch C	27.9 (±0.1)	6.7 (±0.1)
Switch O	29 (±0.2)	4.8 (±0.0)
Switch D30	28.1 (±0.1)	6.0 (±0.1)
Switch D05	27.5 (±0.2)	4.8 (±0.0)
CRYSOL (closed)	28.2[b]	6.4
CRYSOL (open)	29.5[b]	4.4

[a]Experimental data correspond to averaged results from concentration scaled scattering profiles for sample concentrations of 25, 50, and 100 nM. [b]Values were determined from Guinier fits of the predicted scattering profiles in the fitting range $q_{max} \cdot R_g < 1.3$.

For elongated rod-like particles, where the axial dimension is much larger than the radial dimension (as is the case for the DNA origami structures investigated in this work), the scattering intensity can be factorized in an axial and radial scattering component.[43] Analysis of the intermediate q-range then permits the calculation of the radius of gyration for the radial cross-section (R_c) (see Supporting Information). We obtained an average R_c value for the switch C of (6.7 ± 0.1) nm corresponding to a radius of the cross-section $R \sim 9.4$ nm, which is in good agreement with the cross-sectional dimensions of the design model (Figure 1). The switch O can be thought of as being assembled from two rods where the cross-section is half of the size as for the switch C sample. Here, a smaller average R_c value of 4.8 nm, corresponding to a radius $R \sim 6.8$ nm, is fully consistent with the expected reduction of the cross-sectional area when the switch changes from the closed to an open conformation.

A Kratky representation ($q^2I(q)$ versus q) of the scattering data of switch C and switch O reveals a number of peaks that can be related to structural features (Figure 3b). The peak and shoulder at lowest q-values ("1", Figure 3b) for the switch O and switch C samples, respectively, at $q \sim 0.06$ nm^{-1} are related to the overall dimensions of the objects ($d \sim 100$ nm) and to their R_g via $q \approx (3)^{1/2}/R_g \approx 0.06$ nm^{-1}. The major peaks ("2", Figure 3b) at $q \approx 0.14$ nm^{-1} and at $q \approx 0.19$ nm^{-1} for the switch C and the switch O sample, respectively, are related to the maximum of the cross-sectional intensity expected at $q_{max} = 1/R_c$. The fitted R_c values of 6.7 nm for the switch C and 4.8 nm for the switch O sample (Table 1) suggest $q_{max} \sim 0.15$ nm^{-1} and $q_{max} \sim 0.2$ nm^{-1}, which is in very good agreement with the observed peak positions in the Kratky plot. In the higher q-range, both scattering profiles display a small and broad peak ("3", Figure 3b) at $q \sim 1.0$ nm^{-1} ($d \sim 6.3$ nm, Figure 1 red arrow a) and a more pronounced peak ("4", Figure 3b) at $q \sim 1.6$ nm^{-1} ($d \sim 3.9$ nm, Figure 1 red arrow b), which corresponds to the distances between and within the honeycomb lattice, respectively (Figure 1). These values are in

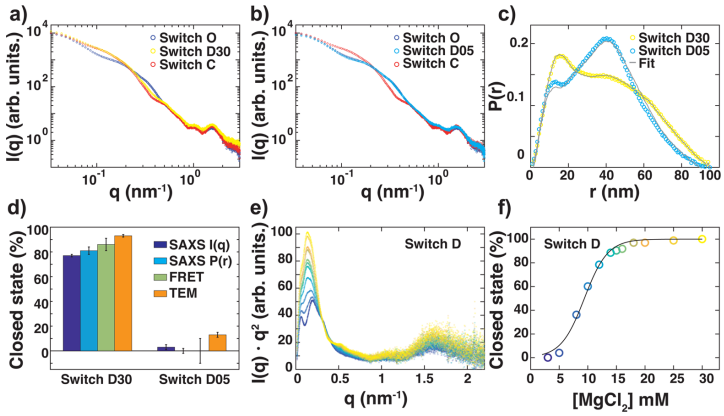


Figure 4. Characterization of conformational states of dynamic switch structures. (a) Comparison of the scattering profile from the switch D30 (yellow) to the scattering profiles of the static switch samples switch O (blue) and switch C (red). (b) Scattering profile from the switch D05 sample (cyan) in comparison to scattering curves from the static structures (same color code as in (a)). (c) $P(r)$ functions of the dynamic switch variants (cyan, yellow circles) and the resulting two-state model fits (gray lines). (d) Comparison of the relative fractions of the closed states determined from the scattering profiles (blue bars), the $P(r)$ functions (cyan bars), ensemble FRET (green bars), and TEM imaging (orange bars) for the switch D30 and switch D05 samples, corresponding to $MgCl_2$ concentrations of 30 mM and 5 mM, respectively. For TEM imaging the highest $MgCl_2$ concentration was 25 mM. (e) Kratky representation of the scattering profiles of switch D samples for varying $MgCl_2$ concentrations: 3 (dark blue,bottom), 5, 8, 10, 12, 14, 15, 16, 18, 20, 25, and 30 mM (light yellow, top). Data are normalized to the intensity at zero scattering angle and scaled by a constant factor. (f) Fraction of closed switch particles for $MgCl_2$ titration experiments shown in (e), determined from a two-state model. Solid lines represent a two-state model with a free energy term that depends linearly on the $MgCl_2$ concentration.

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

approximate agreement with the theoretical values and the relative number of these distances is approximately the same for both conformations, consistent with the similarity of the scattering curves in the higher q-regime. We note that features relating to the structure of single DNA helices (such as their diameter, the minor groove/major groove periodicity, and the spacing between base pairs) occur on even shorter length scales and thus correspond to q-values ≥ 3 nm^{-1}, which have been probed in wide-angle X-ray scattering measurements,[44] but are not the focus of the present work.

To more directly visualize the contribution of features on various length scales, we calculated the pair distance distribution function $P(r)$ (see Supporting Information and Supplementary Figure S6), which describes a histogram of all pairwise distances r within the sample (Figure 3c). For both static open and closed structures, we find a maximum pairwise distance D_{max} of 95 nm, which is in good agreement with the expected maximum distance from the designed structures. The shape of the $P(r)$ function obtained for the switch C variant is peaked at low r with a long tail out to higher r, characteristic of an elongated object. In contrast, the $P(r)$ of the switch O exhibits an overall more Gaussian shape, characteristic of a more globular object. In the switch C $P(r)$ function, we observe a well-defined peak at an intraparticle distance of 16 nm, which corresponds to the maximum transverse distance of the closed state (Figure 1). This peak is not apparent in the $P(r)$ function of the switch O sample, as expected, because the opening of the switch reduces the transverse distance to 8 nm. For the open conformation we find a smaller feature at 12 nm that is related to the height of the switch object and also contains contributions from the maximum transverse distance of ~8 nm (Figure 1), which become more exposed in the open state. The dominant $P(r)$ peak for the switch O, however, occurs around 40 nm, the distance associated with the length of each of the two opened arms.

Conformational Populations of the Dynamic Switch Variant. Having demonstrated that SAXS clearly reveals the large-scale conformational changes between the open and closed versions of the static switch object, we next analyzed the conformational states of the dynamic version of the switch (switch D) at high (30 mM) and low (5 mM) magnesium chloride ($MgCl_2$) concentrations (Figure 4a,b and Supplementary Figure S8). In general, the scattering profile from an ensemble is given by the sum of the scattering profiles for the individual components, weighted by their relative occupancy. In the case of a two-state system, the scattering profile can be described by a linear superposition of the two states

$$I(q) = f_1 I_1(q) + f_2 I_2(q) \quad (1)$$

$I_1(q)$ and $I_2(q)$ are the scattering profiles and the coefficients f_1 and f_2 are fractional occupancies of states 1 and 2, respectively. Using the scattering profiles of the switch O and switch C objects for the open and closed states, we fitted the scattering profiles of the dynamic variant at 30 mM $MgCl_2$ (switch D30) and 5 mM $MgCl_2$ (switch D05) as a linear superposition of the two states (Supplementary Figure S8). Under both conditions, the two-state fits provide an overall excellent description of the experimental data, suggesting that the conformations of the dynamic switch variant can be well approximated by a two-state model featuring the open and closed states. For both samples, slight deviations of the fit become apparent at higher q-values, which might imply that there exist structural differences within the internal honeycomb lattice between the dynamic and static versions. This might be attributed to the different concentrations of $MgCl_2$ in the sample solutions, which have an impact on structural integrity and flexibility owing to its efficacy in screening interhelical repulsion and stabilizing DNA Holliday junctions (see also below).[45,46] In addition, previous TEM studies on the switch D05 sample revealed a slightly reduced opening angle compared to the fixed opening angle of 90° for the switch O sample,[37] which might cause some additional differences in the scattering profiles.

Complementary to analyzing $I(q)$, we applied a two-state model analogous to eq 1 to the $P(r)$ functions (Figure 4c). We find that the $P(r)$ function of the switch D30 sample can be described accurately by the two-state model. For the $P(r)$ function of the switch D05 sample again slight deviations between the two-state model and the data are observable, but overall the two-state description is still accurate.

The fitted parameters f_1 and f_2 in eq 1 provide a direct measure of the relative populations of the two states. Figure 4d shows the relative populations of the closed conformation determined from the scattering intensity and $P(r)$ fits (the corresponding populations of the open conformation are the complement to 100%). From the $I(q)$ fits, we find a population of (77 ± 1)% in the closed state for the switch D30 sample, in agreement with the expectation that screening of electrostatic repulsion at high salt concentration should lead to a predominant population of the closed conformation. In contrast, the occupancies derived for the switch D05 sample are (3 ± 2)% for the closed state, which is in line with the prediction that electrostatic repulsion at lower ionic strength favors the open configuration. The $P(r)$ fits gave identical results within experimental error (Figure 4d). These findings are further supported by the fact that the fitted cross-sectional radii of gyrations of the switch D object in 5 and 30 mM $MgCl_2$ are close to values determined for the switch O and switch C conformations, respectively (Table 1).

The results of the SAXS analyses can be compared to data obtained from ensemble FRET measurements and TEM imaging on switch D particles at varying $MgCl_2$ concentrations[37] (see Supporting Information and Supplementary Figure S7). Data from solution-based ensemble FRET measurements are in good agreement within experimental errors with the SAXS results (Figure 4d). From TEM imaging data, higher fractional occupancies for the closed state were obtained compared to the solution-based methods: (93 ± 1)% of the objects were identified to be in the closed state at a $MgCl_2$ concentration of 25 mM and (13 ± 2)% of closed particles were found at a $MgCl_2$ concentration of 5 mM. The deviations of the TEM-determined fractions to the solution-based values are modest but statistically significant for the SAXS derived values (Figure 4d) and might be related to several factors. First, for the TEM analysis switch D particles were picked from TEM images for each salt condition and manually assigned to be either open or closed; errors were determined from binomial counting statistics. This process might introduce a slight bias, as overlapping objects could not be classified and as partially closed switch objects were considered as closed. Second, TEM imaging requires immobilization of samples on a surface potentially affecting their conformation. Furthermore, the staining process for TEM imaging can alter the global shape of the particles.[47] In addition, single-molecule FRET experiments, which likewise require surface immobilization of the switch D particles, gave similar results as the TEM data.[37] Taken together, the data suggest that

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

surface immobilization and/or staining might create a modest bias toward the closed conformation and can give rise to a small population of partially closed conformations, possibly due to direct surface interactions or excluded volume effects.

In order to further test whether the transition from the open to the closed state of the dynamic switch variant upon the addition of $MgCl_2$ ions can be described as a two-state process, we performed SAXS experiments on switch D samples for varying $MgCl_2$ concentrations ranging from 3 to 30 mM (Figure 4e). The scattering profiles at different $MgCl_2$ concentrations exhibit an iso-scattering point around $q \sim 0.3$ nm^{-1} tentatively suggesting that the conformational transition can be described as a two-state process. For a more quantitative analysis, we performed a two-state fit of the scattering profiles at each $MgCl_2$ concentration according to eq 1 and fitted the resulting populations by a thermodynamic model (using Equations S8 and S9 in Supporting Information) assuming a linear dependence of the free energy ΔG on the ion concentration c. From a least-squares fit we obtained for ΔG_0 = 1.2 kcal/mol at the reference ion concentration of 5 mM and the slope m_c = −0.3 kcal/(mol·mM), which is in good agreement with values based on ensemble FRET measurements (Figure 4f and Supplementary Figure S7). In addition, the two state-fits yield a good fit of the full scattering profiles over the entire range of $MgCl_2$ concentrations (Supplementary Figure S8). These findings show that the switch D transition from the open to the closed state can be described adequately, at least at the current level of signal-to-noise, by a two-state model employing a single open and closed conformation, without the need to introduce intermediates states or conformations.

Overall, we find quantitative agreement between SAXS and solution FRET derived population estimates and approximate agreement with the TEM derived values, confirming the switching mechanism in the dynamic switch variant. Our findings highlight the importance of solution-based techniques when performing structural characterization of complex DNA structures.

Comparison of Experimental SAXS Data to Idealized Models and Model Refinement. In addition to detecting conformational transitions and providing global measures of size and shape (such as R_g, R_c, and D_{max}), SAXS can provide information about the full 3D solution structure of macromolecules and their assemblies.[24,48] Even though the resolution of SAXS experiments is typically insufficient to compute a unique structure, it is possible to test and refine structural models against experimental SAXS data.[49−52] First, we compared our experimental data to scattering profiles of the switch O and switch C samples predicted from idealized atomistic models generated by CanDo[53] (see Supporting Information). The computed profiles from the CanDo models reproduce the overall shape of the experimental curves and reveal similar characteristic peaks (Figure 5a,b; Supplementary Figure S9). In addition, we obtain R_g and R_c values from the theoretical scattering curves, which are in good agreement with the experimentally determined values (Table 1, Supplementary Table S1). However, small but systematic deviations between the experimental and theoretical profiles are apparent. There is an additional peak in the theoretical scattering patterns for the open state at $q \sim 0.26$ nm^{-1} and the peaks that are visible in both experimental and theoretical curves are shifted, mostly to higher q in the theoretical curves compared to experiment. Furthermore, the ratios of the peak intensity values at low and high q differ between the experimental and theoretical curves.

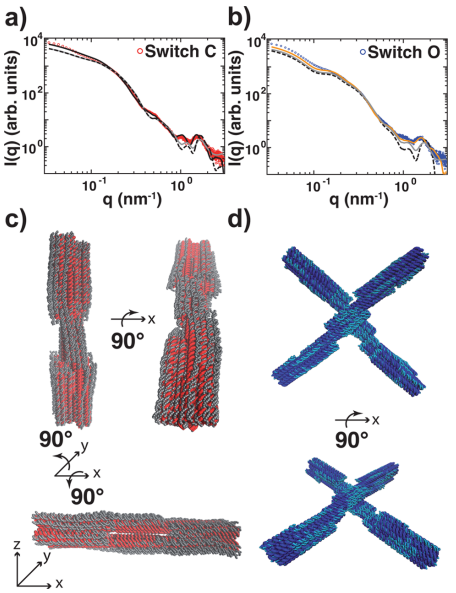


Figure 5. Normal mode-based refinement of DNA origami structures against SAXS data. (a) Data for the switch C construct and (b) the corresponding results for the switch O sample. Experimental scattering profiles are shown as red or blue circles. Scattering profiles predicted from the initial CanDo derived models using all atoms and the software CRYSOL are shown as gray lines and using a one-bead-per-base representation as dashed black lines. Scattering profiles for the final models (computed using the one-bead-per-base representation) after normal mode based refinement are shown as black (for switch C) and orange (for switch O) lines. (c,d) Initial models for the switch C and switch O objects as red and blue tubes and the final models after normal mode refinement as gray and cyan spheres, respectively.

In addition, we determined the $P(r)$ functions from the theoretical data and calculated a histogram of distances directly from the atomistic model coordinates (Supplementary Figure S9). In comparison to the experimental data, the peaks are more pronounced and deviations from the experimental peak positions are observable.

We note that even though the different methods to compute scattering profiles from the structures exhibit some differences (see Supporting Information and Supplementary Figure S9), they do give overall very similar results and show comparable deviations from the experimental data, suggesting that the details of the scattering computations are relatively unimportant and can not explain the observed differences to the experimental data. In principle, both the hydration layer of partially ordered water molecules around a macromolecule in solution[49,54] and the ion atmosphere around charged nucleic acids[55,56] contribute to the scattering profile. For simple DNA duplexes, the effect of the ion atmosphere has been studied in detail and while the ion cloud's contribution to the scattering is pattern is relatively minor, the radius of gyration typically increases by a few angstroms.[55,57] We have performed electrostatic calculations using linearized Poisson−Boltzmann

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

theory (Supporting Information and Supplementary Figures S10 and S11) to compare the electrostatic potential in the vicinity of the DNA switch object with a simple DNA duplex. Our results suggest that the electrostatic potential and, consequently, the ion density around our DNA origami structures are only slightly elevated and overall similar in magnitude and spatial extent compared to a single double-stranded DNA helix (Supplementary Figures S10 and S11), consistent with previous reports in the literature.[58,59,61] Taken together, these observations suggest that for the very large DNA structures considered in this work contributions from the ion atmosphere to the scattering profile are small or negligible. In addition, we tested whether altering the density of the solvent or the contrast of the hydration layer in the range of physically plausible values would explain the observed differences between the CanDo derived models and our experimental data, but again we found that while changing the hydration shell gives rise to small changes in the scattering profiles these changes are insufficient to account for the observed differences (Supplementary Figure S12).

Combined, the differences between experimental and predicted scattering profiles indicate that the switch objects adopt conformations in solutions that differ from the idealized models generated by CanDo. Such deviations have been suggested previously, for example, Pan et al.[53] found an average root-mean-square deviation (RMSD) of 3.2 Å between the CanDo derived model and the crystal structure of a DNA tensegrity motif. In general, electrostatic repulsion between adjacent helix bundles or at crossovers resulting in the bowing out of double helical domains[2,4] can lead to local displacements of nucleobase positions. Theoretical calculations and experimental evidence based on TEM data suggest an important role of flexibility for several DNA origami structures,[60−62] indicating maximum root-mean-square fluctuation amplitudes of a few nanometers.[60] In addition, a cryo-EM structure of a DNA origami object observed deviations between the idealized structure and the experimentally determined density map.[22]

There is currently no established method to refine DNA origami structures quantitatively against experimental data. A considerable challenge in this regard is the large size of our switch objects that renders refinement, for example, based on all-atom molecular dynamics[61,63] challenging. As a computationally tractable approach, we turned to normal mode refinement of the CanDo derived model against the experimental SAXS data using an elastic network model. Normal mode analysis[64−68] based on coarse-grained elastic network models has proven to describe large-scale conformational changes surprisingly well as compared to considerably more complex approaches[66] and has been applied to deform macromolecular structures to fit and refine experimental data from cryo-EM,[69] X-ray crystallography,[67,70] and SAXS data.[71−73] We iteratively refined the switch C and switch O structures against the experimental SAXS data by normal mode based deformations (see Supporting Information). The resulting structures yield significantly better fits to the data (Figure 5a,b); the goodness-of-fit statistic X^2 (defined in Equation S4 in the Supporting Information) is reduced from 0.5% to 0.06% and from 6.7% to 1.9% for the switch C and switch O structures, respectively. We find that for the refined structures the highly symmetric lattice structure is significantly deformed (Figure 5c,d). In comparison to the initial models, some parts in the refined closed and opened switch objects swell and bulge out. This effect is especially pronounced in double helices around the center of the structure, where the two arms are connected to each other (Supplementary Figure S13). In addition, the refined structures show the helices at the ends and sides of the arms slightly bend outward (Supplementary Figure S13). Interestingly, these effects are more pronounced in the switch C compared to the switch O structure. The RMSD for the refined switch C structure compared to the initial model is 22.3 Å; for the switch O, the refined structure has an RMSD of 8.4 Å relative to the starting model. The larger deformations in the switch C object compared to switch O might be due to the more compact structure and, therefore, higher charge density, that would make electrostatic repulsion more relevant for this object. Taken together, these data suggest an important role of flexibility and local deformations in DNA origami objects, which has to be considered when designing complex origami structures.

In summary, we have demonstrated the ability for SAXS to sensitively monitor conformational changes of self-assembled DNA origami objects in solution. SAXS provides a number of advantages. First, being a solution-based technique, SAXS is free of potential biases and perturbations from the proximity of a surface. Second, SAXS is a label free method, without the need to chemically modify the structure of interest. Third, SAXS reads out the global conformation of molecules or molecular assemblies in solution, as defined by their electron density, thus avoiding concerns whether, for example, variations in fluorescence might stem from local conformational changes or photophysical effects upon changes in solution condition. Taken together, these advantages render SAXS a very promising novel approach for detecting conformational states of dynamic DNA origami objects and we anticipate that many of the techniques' capabilities that were previously demonstrated in other contexts can be extended toward monitoring conformational changes in DNA nanostructures, including temperature controlled[74] and/or time-resolved SAXS[31,32,75] measurements and the detection and characterization of structural intermediates and molecular ensembles.[36,39,76]

Quantitative comparison of the experimental SAXS data to theoretical profiles derived from 3D models of the DNA objects reveal considerable flexibility and deformations away from the idealized "design" structure. Such deformations will have to be taken into account for high-resolution designs in the future. In addition, this work highlights the ability of SAXS to critically test structural models against solution-based data, even for very large DNA objects, which constitutes a promising approach that is complementary to the more routinely used methods.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.6b01338.

Supplementary Methods, Table S1, and Supplementary Figures S1−S13.(PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Jan.Lipfert@lmu.de. Phone: +49-80-2180-2005.

Author Contributions

L.K.B., T.G., H.D., and J.L. designed the study; T.G. assembled and purified samples; L.K.B., S.M.S., and P.W. performed SAXS measurements; W.Z. performed structure refinement. All

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

authors analyzed data, contributed to writing the paper, and have given approval to the final version of the manuscript.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft through grants provided within Gottfried-Wilhelm Leibniz Program, the Excellence Clusters CIPSM (Center for Integrated Protein Science Munich), NIM (Nanosystems Initiative Munich), and the Sonderforschungsbereich SFB863.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Keyao Pan and Dr. Marc Delarue for helpful discussions and support on atomistic modeling and normal-mode analysis, Dr. Adam Round for assistance at beamline BM29 at the ESRF (Grenoble, France), Dr. Alessandro Spilotros and Nelly Hajizadeh for help at beamline P12 at DESY (Hamburg, Germany), Dr. Marianne Liebi at beamline X12SA, cSAXS at the Swiss Light Source (Switzerland), and Florian Praetorius for scaffold DNA preparations.

■ ABBREVIATIONS

SAXS, small-angle X-ray scattering; FRET, (Förster) fluorescence resonance energy transfer; AFM, atomic force microscopy; Switch C, closed switch; switch O, open switch; switch D, dynamic switch

■ REFERENCES

(1) Seeman, N. C. *Nature* **2003**, *421* (6921), 427−431.
(2) Rothemund, P. W. K. *Nature* **2006**, *440* (7082), 297−302.
(3) Winfree, E.; Liu, F.; Wenzler, L. A.; Seeman, N. C. *Nature* **1998**, *394* (6693), 539−544.
(4) Castro, C. E.; Kilchherr, F.; Kim, D.-N.; Shiao, E. L.; Wauer, T.; Wortmann, P.; Bathe, M.; Dietz, H. *Nat. Methods* **2011**, *8* (3), 221−229.
(5) Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M. *Nature* **2009**, *459* (7245), 414−418.
(6) Dietz, H.; Douglas, S. M.; Shih, W. M. *Science* **2009**, *325* (5941), 725−730.
(7) Benson, E.; Mohammed, A.; Gardell, J.; Masich, S.; Czeizler, E.; Orponen, P.; Högberg, B. *Nature* **2015**, *523* (7561), 441−U139.
(8) Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. *Science* **2012**, *338* (6111), 1177−1183.
(9) Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Cristiano, L. P.; Oliveira; Pedersen, J. S.; Birkedal, V.; Besenbacher, F.; Gothelf, K. V.; Kjems, J. *Nature* **2009**, *459* (7243), 73−76.
(10) Douglas, S. M.; Bachelet, I.; Church, G. M. *Science* **2012**, *335* (6070), 831−834.
(11) Kuzyk, A.; Schreiber, R.; Zhang, H.; Govorov, A. O.; Liedl, T.; Liu, N. *Nat. Mater.* **2014**, *13* (9), 862−866.
(12) Liu, M.; Fu, J.; Hejesen, C.; Yang, Y.; Woodbury, N. W.; Gothelf, K.; Liu, Y.; Yan, H. *Nat. Commun.* **2013**, *4*, 2127.
(13) Zhang, D. Y.; Seelig, G. *Nat. Chem.* **2011**, *3* (2), 103−113.
(14) Marras, A. E.; Zhou, L.; Su, H.-J.; Castro, C. E. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (3), 713−718.
(15) Kuzuya, A.; Ohya, Y. *Acc. Chem. Res.* **2014**, *47* (6), 1742−1749.
(16) Zhao, Y.-X.; Shaw, A.; Zeng, X.; Benson, E.; Nyström, A. M.; Högberg, B. *ACS Nano* **2012**, *6* (10), 8684−8691.
(17) Zhang, F.; Jiang, S.; Wu, S.; Li, Y.; Mao, C.; Liu, Y.; Yan, H. *Nat. Nanotechnol.* **2015**, *10* (9), 779−784.
(18) Aghebat Rafat, A.; Pirzer, T.; Scheible, M. B.; Kostina, A.; Simmel, F. C. *Angew. Chem., Int. Ed.* **2014**, *53* (29), 7665−7668.
(19) Langecker, M.; Arnaut, V.; Martin, T. G.; List, J.; Renner, S.; Mayer, M.; Dietz, H.; Simmel, F. C. *Science* **2012**, *338* (6109), 932−936.
(20) Han, D.; Pal, S.; Nangreave, J.; Deng, Z.; Liu, Y.; Yan, H. *Science* **2011**, *332* (6027), 342−346.
(21) Schreiber, R.; Luong, N.; Fan, Z.; Kuzyk, A.; Nickels, P. C.; Zhang, T.; Smith, D. M.; Yurke, B.; Kuang, W.; Govorov, A. O.; Liedl, T. *Nat. Commun.* **2013**, *4*, 2948.
(22) Bai, X.-C.; Martin, T. G.; Scheres, S. H. W.; Dietz, H. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (49), 20012−20017.
(23) Koch, M. H. J.; Vachette, P.; Svergun, D. I. *Q. Rev. Biophys.* **2003**, *36* (02), 147−227.
(24) Lipfert, J.; Doniach, S. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36* (1), 307−327.
(25) Rambo, R. P.; Tainer, J. A. *Annu. Rev. Biophys.* **2013**, *42* (1), 415−441.
(26) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. *Q. Rev. Biophys.* **2007**, *40* (3), 191−285.
(27) Eliezer, D.; Jennings, P. A.; Wright, P. E.; Doniach, S.; Hodgson, K. O.; Tsuruta, H. *Science* **1995**, *270* (5235), 487−488.
(28) Akiyama, S.; Takahashi, S.; Kimura, T.; Ishimori, K.; Morishima, I.; Nishikawa, Y.; Fujisawa, T. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (3), 1329−1334.
(29) Doniach, S. *Chem. Rev.* **2001**, *101* (6), 1763−1778.
(30) Fang, X.; Littrell, K.; Yang, X.-J.; Henderson, S. J.; Siefert, S.; Thiyagarajan, P.; Pan, T.; Sosnick, T. R. *Biochemistry* **2000**, *39* (36), 11107−11113.
(31) Russell, R.; Millett, I. S.; Tate, M. W.; Kwok, L. W.; Nakatani, B.; Gruner, S. M.; Mochrie, S. G. J.; Pande, V.; Doniach, S.; Herschlag, D.; Pollack, L. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (7), 4266−4271.
(32) Herschlag, D.; Russell, R.; Millett, I. S.; Doniach, S. *Nat. Struct. Biol.* **2000**, *7* (5), 367−370.
(33) Tuukkanen, A. T.; Svergun, D. I. *FEBS J.* **2014**, *281* (8), 1974−1987.
(34) Bessman, N. J.; Bagchi, A.; Ferguson, K. M.; Lemmon, M. A. *Cell Rep.* **2014**, *9* (4), 1306−1317.
(35) Rambo, R. P.; Tainer, J. A. *RNA* **2010**, *16* (3), 638−646.
(36) Lipfert, J.; Das, R.; Chu, V. B.; Kudaravalli, M.; Boyd, N.; Herschlag, D.; Doniach, S. *J. Mol. Biol.* **2007**, *365* (5), 1393−1406.
(37) Gerling, T.; Wagenbauer, K. F.; Neuner, A. M.; Dietz, H. *Science* **2015**, *347* (6229), 1446−1452.
(38) Sim, A. Y. L.; Lipfert, J.; Herschlag, D.; Doniach, S. *Phys. Rev. E* **2012**, *86* (2), 021901.
(39) Ali, M.; Lipfert, J.; Seifert, S.; Herschlag, D.; Doniach, S. *J. Mol. Biol.* **2010**, *396* (1), 153−165.
(40) Lipfert, J.; Ouellet, J.; Norman, D. G.; Doniach, S.; Lilley, D. M. J. *Structure* **2008**, *16* (9), 1357−1367.
(41) Bai, Y.; Das, R.; Millett, I. S.; Herschlag, D.; Doniach, S. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (4), 1035−1040.
(42) Meisburger, S. P.; Sutton, J. L.; Chen, H.; Pabit, S. A.; Kirmizialtin, S.; Elber, R.; Pollack, L. *Biopolymers* **2013**, *99* (12), 1032−1045.
(43) Kratky, O.; Glatter, O. *Small Angle X-ray Scattering*; Academic Press: London, 1982; pp 1−267.
(44) Zuo, X.; Cui, G.; Merz, K. M.; Zhang, L.; Lewis, F. D.; Tiede, D. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (10), 3534−3539.
(45) Chiu, T. K.; Dickerson, R. E. *J. Mol. Biol.* **2000**, *301* (4), 915−945.
(46) Martin, T. G.; Dietz, H. *Nat. Commun.* **2012**, *3*, 1103.
(47) Frank, J. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31* (1), 303−319.
(48) Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L., II; Tsutakawa, S. E.; Jenney, F. E., Jr; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S.-J.; Scott, J. W.; Dillard, B. D.; Adams, M. W. W.; Tainer, J. A. *Nat. Methods* **2009**, *6* (8), 606−612.
(49) Svergun, D.; Barberato, C.; Koch, M. *J. Appl. Crystallogr.* **1995**, *28* (6), 768−773.
(50) Schneidman-Duhovny, D.; Hammel, M.; Sali, A. *Nucleic Acids Res.* **2010**, *38*, W540−W544.

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

(51) Poitevin, F.; Orland, H.; Doniach, S.; Koehl, P.; Delarue, M. *Nucleic Acids Res.* **2011**, *39*, W184−W189.
(52) Ravikumar, K. M.; Huang, W.; Yang, S. *J. Chem. Phys.* **2013**, *138* (2), 024112−024118.
(53) Pan, K.; Kim, D.-N.; Zhang, F.; Adendorff, M. R.; Yan, H.; Bathe, M. *Nat. Commun.* **2014**, *5*, 5578−7.
(54) Svergun, D. I.; Richard, S.; Koch, M. H.; Sayers, Z.; Kuprin, S.; Zaccai, G. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (5), 2267−2272.
(55) Pollack, L. *Annu. Rev. Biophys.* **2011**, *40* (1), 225−242.
(56) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. *Annu. Rev. Biochem.* **2014**, *83* (1), 813−841.
(57) Meisburger, S. P.; Pabit, S. A.; Pollack, L. *Biophys. J.* **2015**, *108* (12), 2886−2895.
(58) Gavryushov, S.; Zielenkiewicz, P. *Biophys. J.* **1998**, *75* (6), 2732−2742.
(59) Yoo, J.; Aksimentiev, A. *Biophys. J.* **2014**, *106* (2), 695a.
(60) Kim, D. N.; Kilchherr, F.; Dietz, H.; Bathe, M. *Nucleic Acids Res.* **2012**, *40* (7), 2862−2868.
(61) Yoo, J.; Aksimentiev, A. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (50), 20099−20104.
(62) Sedeh, R. S.; Pan, K.; Adendorff, M. R.; Hallatschek, O.; Bathe, K.-J.; Bathe, M. *J. Chem. Theory Comput.* **2016**, *12* (1), 261−273.
(63) Maffeo, C.; Yoo, J.; Aksimentiev, A. *Nucleic Acids Res.* **2016**, *44*, 3013−3019.
(64) Miyashita, O.; Gorba, C.; Tama, F. *J. Struct. Biol.* **2011**, *173* (3), 451−460.
(65) López-Blanco, J. R.; Miyashita, O.; Tama, F.; Chacón, P. *Normal Mode Analysis Techniques in Structural Biology*; John Wiley & Sons, Ltd: Chichester, U.K., 2001; pp 1−12.
(66) Tama, F.; Sanejouand, Y. H. *Protein Eng., Des. Sel.* **2001**, *14* (1), 1−6.
(67) Delarue, M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 40−48.
(68) Hinsen, K. *Proteins: Struct., Funct., Genet.* **1998**, *33* (3), 417−429.
(69) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (16), 9319−9323.
(70) Schröder, G. F.; Levitt, M.; Brunger, A. T. *Nature* **2010**, *464* (7292), 1218−1222.
(71) Gorba, C.; Tama, F. *Bioinf. Biol. Insights* **2010**, *4*, 43−54.
(72) Zheng, W.; Tekpinar, M. *Biophys. J.* **2011**, *101* (12), 2981−2991.
(73) Panjkovich, A.; Svergun, D. I. *Phys. Chem. Chem. Phys.* **2016**, *18* (8), 5707−5719.
(74) Pérez, J.; Vachette, P.; Russo, D.; Desmadril, M.; Durand, D. *J. Mol. Biol.* **2001**, *308* (4), 721−743.
(75) Pollack, L. *Biopolymers* **2011**, *95* (8), 543−549.
(76) Baird, N. J.; Westhof, E.; Qin, H.; Pan, T.; Sosnick, T. R. *J. Mol. Biol.* **2005**, *352* (3), 712−722.

DOI: 10.1021/acs.nanolett.6b01338
Nano Lett. 2016, 16, 4871−4879

Supplementary Information to "Conformational Changes and Flexibility of DNA Devices Observed by Small-Angle X-Ray Scattering"

Linda K. Bruetzel[†], Thomas Gerling[‡], Steffen M. Sedlak[†], Philipp U. Walker[†], Wenjun Zheng[§], Hendrik Dietz[‡], Jan Lipfert[†,*]

[†]Department of Physics, Nanosystems Initiative Munich, and Center for Nanoscience, LMU Munich, Amalienstrasse 54, 80799 Munich, Germany

[‡]Physik Department, Walter Schottky Institute, Technische Universität München, Am Coulombwall 4a, 85748 Garching near Munich, Germany

[§]Physics Department, State University of New York at Buffalo, Buffalo, New York 14260, United States

[*]Corresponding author: Jan Lipfert; Email: Jan.Lipfert@lmu.de; Phone: +49-89-2180-2005

Supplementary Information consisting of

- **Materials and Methods**
- **Supplementary Table S1**
- **Supplementary Figures S1-S13**
- **Supplementary References**

Materials and Methods

DNA origami assembly and purification

DNA origami objects were designed using caDNAno v.02.[1] Three different variants of the switch object were prepared for the SAXS experiments (Supplementary Figure S1-S3): two static variants that were permanently locked either in the open state (switch O) or the closed state (switch C), and a dynamic variant with 16 activated stacking interactions (switch D) (Figure 1). Each structure contained 16128 nucleotides. The scaffold DNA (p8064) was prepared as previously described.[2] Staple DNA strands were synthesized by solid-phase chemical synthesis (Eurofins Genomics GmbH, Ebersberg, Germany; HPSF purification). DNA origami objects were self-assembled by subjecting the one-pot reaction mixture to a thermal annealing ramp using a thermal cycling device (TETRAD; MJ Research – now Biorad).[3] The reaction mixture contained 50 nM scaffold DNA (p8064), 200 nM of each staple DNA strand, folding buffer (1 mM EDTA, 5 mM TrisBase, 5 mM NaCl; pH 8) and 20 mM $MgCl_2$. After a 15 min-long thermal denaturation step at 65 ˚C, the thermal annealing ramp covered the temperature interval [58 – 55 ˚C] with a rate of 1 ˚C/90 min. Excess staple DNA strands were removed from the reaction mixture by performing two rounds of polyethylene glycol (PEG) precipitation.[4] The resulting pellets were dissolved in folding buffer (1 mM EDTA, 5 mM TrisBase, 5 mM NaCl; pH 8) containing 5 mM $MgCl_2$. To allow for equilibration, all samples were incubated at 40 ˚C and 400 rpm overnight. Residual PEG was removed from the samples by performing three rounds of ultrafiltration (30K Amicon Ultra-0.5mL; Merck Millipore). Filters were equilibrated by adding 500 µl folding buffer containing either 5, 15, or 30 mM $MgCl_2$ at 2000 x *g* and 25 ˚C for 2 minutes. Then, 50 µl sample was mixed with 450 µl folding buffer and centrifuged at 8000 x *g* and 25 ˚C for 15 minutes. The flow-throw was discarded and 480 µl of folding buffer was added to the recovered sample. The concentration of the different samples was determined using a spectrophotometer (NanoDrop 8000; Thermo Scientific). All samples were measured at three different DNA origami object concentrations (25 nM, 50 nM, and 100 nM), prepared by dilution using the appropriate buffers. Samples of the switch O and switch C were measured in folding buffer (1 mM EDTA, 5 mM TrisBase, 5 mM NaCl; pH 8) containing 15 mM $MgCl_2$. Samples of the switch D variant were measured in folding buffer containing either 5 mM or 30 mM $MgCl_2$.

Scaling relationship for the concentration requirements for nucleic acid SAXS measurements

For monodisperse solutions and in the absence of interparticle interference effects, the forward scattering intensity *I(0)* scales linearly with sample concentration *c* and quadratically with the molecular weight *MW*:[5,6]

$$I(0) = K\, c\, (\Delta\rho)^2\, p^2\, (MW)^2 \qquad \text{(S1)}$$

$\Delta\rho$ and p are the average electron contrast and the partial specific volume of the molecule and are approximately constant for all nucleic acid molecules.[6] K is an instrument specific constant, which is typically determined from comparison to a molecular weight standard. Even though Equation 1 only strictly holds for the forward scattering intensity, it provides a rough estimate of the scattering signal expected in a SAXS measurement at a given sample concentration and molecular weight. For the set of samples analyzed in this work (Figure 2a), the concentrations required for good quality SAXS measurements follows roughly the anticipated $\sim MW^{-2}$ relationship (Figure 2a, solid line) from Equation S1. Treating the exponent as a free parameter, i.e. fitting the data to a relationship $\sim MW^{-\nu}$, yields a better fit with $\nu = 1.30$ (Figure 2a, dashed line). A shallower dependence than $\sim MW^{-2}$ on the number of nucleotides can be rationalized by considering the fact that SAXS profiles for nucleic acids are maximal at low q and fall off for higher q. Importantly, the fall off with increasing q is more rapid for larger structures, suggesting that for these structures higher concentrations than suggested by the simple $\sim MW^{-2}$ scaling are required to obtain a decent signal at higher q-values. Our DNA origami measurements at 25 nM (Figure 2a, red star) are in excellent agreement with the extrapolated scaling relationship with $\nu = 1.30$ (Figure 2a, dashed line); including the origami data point into the fit yields a nearly identical scaling exponent of $\nu = 1.36$.

SAXS measurements

SAXS measurements were performed at beamline BM29, ESRF, Grenoble[7] at an X-ray wavelength λ of 0.99 Å, using a sample-to-detector distance of 2.87 m and a Pilatus 1M detector, resulting in a q-range of 0.03 to 5 nm^{-1} ($q = 4\pi \cdot \sin(\theta)/\lambda$, where θ is the total scattering angle). For each sample concentration ten runs with an exposure time of 1 s in 'flow' mode were conducted at room temperature. SAXS data of dynamic switch samples for Mg^{2+} titration experiments were collected at beamline P12 (DESY, Hamburg, Germany[8]) at an X-ray wavelength λ of 1.18 Å and a sample-to-detector distance of 6 m, resulting in a

q-range of 0.03 to 2.2 nm^{-1}. Data were acquired in 'flow' mode with an exposure time of 95 ms and 150 frames per sample at room temperature. For each experiment buffer samples were measured using identical procedures before and after each sample measurement. Sample and buffer data from each run were analyzed for radiation damage; no damage was observed in any of the measurements. Matching sample and buffer profiles were averaged and buffer profiles were subtracted for background correction. Unless otherwise noted the scattering profiles shown in this work correspond to data from averaged and buffer subtracted intensity profiles measured at a DNA origami concentration of 100 nM.

SAXS data analysis

We performed Guinier analyses to obtain radii of gyrations for all measured DNA origami structures, by fitting the logarithm of the scattering intensity as a function of q^2 to a straight line for small values of q.[9] Due to the large size of the DNA origami objects, we extended the fitting range criterion to $q_{max} \cdot R_g < 1.6$,[10] enabling us to obtain reasonable estimates of the forward scattering intensity and radii of gyration that, nonetheless, should still be treated as approximations (Supplementary Figure S5a).

In addition to determining the global R_g of the object from the scattering signal at very low q, DNA origami structures investigated in this work can be approximated as rod-like particles with an axial length L (~95 nm) and a radial cross section A (~20 nm and ~14 nm for switch C and switch O, respectively) (Figure 1). The total scattering intensity is approximated by:[11]

$$I(q) = I(q)_{axial} * I(q)_{cross} = \frac{\pi \cdot L}{q} * A^2 \Delta\rho^2 \exp\left(-\frac{q^2 R_c^2}{2}\right) = \frac{a}{q} \exp\left(-\frac{q^2 R_c^2}{2}\right) \quad \text{(S2)}$$

where the first factor is related only to the axial component and the remaining part corresponds to the cross-sectional scattering with an electron density contrast $\Delta\rho$. The pre-factors can be combined into a single fitting constant a. Equation S2 is valid in the range of $q_{min} = 2\pi/R_g$ and $q_{max} = 2\pi/R_c$, which corresponds to a q-range of ~ 0.2 – 0.9 nm^{-1} (1.3 nm^{-1}) for the switch C (switch O) object. Values for a and R_c were obtained by performing a least squares analysis in the valid q-range (Supplementary Figure S5b,c). The R_c value can be used for calculating the corresponding radius according to $R_c^2 = R^2/2$ when describing the switch object by a cylinder model with R = 10 nm (corresponding to a diagonal D = 20 nm of the cross-sectional area of the closed arm) and R = 7 nm (corresponding to the diagonal $D \approx 14$ nm of the cross-sectional area of the open arm).

Furthermore, we computed the pair distance distribution function $P(r)$ as described by Moore[12] using an indirect Fourier transformation in terms of a sine series expansion, based on the Shannon sampling theorem.[13] The large size of the DNA nanostructure is beneficial in the

context of the series expansion, since the maximum number of series coefficients (a_n) is given by $n_{max} = \frac{q_{max} \cdot D_{max}}{\pi}$, where D_{max} describes the maximum particle dimension. The calculation of *P(r)* requires a fixed value for D_{max}, which we expected to be around 95 nm for the switch object (Figure 1). We tested different values for D_{max} ranging from 60 – 130 nm for all switch samples by calculating a chi-squared value (χ^2), which describes the discrepancy between the experimental data and the fit, for each D_{max}. The χ^2 values decay approximately exponentially with increasing D_{max} (Supplementary Figure S6) until they plateau for $D_{max} > 90$ nm. For D_{max} values > 95 nm no change in the overall shape of the *P(r)* was observable, thus a D_{max} of 95 nm was taken as the best estimate of D_{max}. *P(r)* functions shown were normalized to give equal surface areas.

Structural parameters described above as well as the two-state model fitting (see main text) were calculated from scattering profiles averaged from 10 independent runs for all three sample concentrations. The related errors represent the standard deviation.

Computing SAXS profiles from atomic models of DNA origamis

We generated atomistic models for the open and closed switch object with the CanDo software assuming idealized DNA helix and junction geometries.[14] CanDo simulations were run using the CanDo webserver (http://cando-dna-origami.org); computation time was dependent on the load of the server and the design of the structure, but typically in the range of hours. There are several methods to predict a scattering profile from an atomistic model that differ in various aspects of the computation. We used the programs *CRYSOL*[15] and *FOXS*[16] for calculation of the theoretical scattering curves from atomistic models. *CRYSOL*[15] computes the scattering intensity using a spherical harmonics expansion and scattering contributions from the hydration shell around the molecule are taken into account by assuming a homogenous 3 Å thick border layer with a default density contrast value of 0.03 e/Å^3. The program *FOXS*[16] evaluates the theoretical scattering profile from the Debye formula and the particle hydration layer is modeled as a function of surface accessibility.[17] As a complementary and more simplistic approach, we utilized a custom written routine in C, adapted from the program SAXS3D,[18] to determine theoretical scattering profiles based on a coarse-grained representation of the switch objects including only scattering centers per DNA base. For the Debye formula routine, only one particle (placed at the phosphorus position) per base was used and the *q*-range was set to 0 - 3 nm^{-1} including 300 datapoints. *CRYSOL* was run in interactive mode, setting the order of harmonics to the maximum value of 50, given the large size of the switch object. The number of points in the theoretical curve was fixed to 800

within a q-range from 0 - 3 nm^{-1}; the remaining parameters were set to default values, without fitting the theoretical curve to the experimental data. *FOXS* was executed in default mode using the same number of points and q-range as applied in *CRYSOL*. Calculated scattering profiles were fitted to the experimental data of the switch C and switch O sample by performing a linear fit including a constant offset (Figure 5a,b and Supplementary Figure S9). To test the influence of the ion shell surrounding the switch object on the shape of the theoretical scattering profiles we investigated solvent density values ranging from 0.334 e/$Å^3$ (default, corresponding to the solvent density of water) and 0.344 e/$Å^3$ in *CRYSOL* (Supplementary Figure S12) and from 0.307 e/$Å^3$ (minimum) – 0.388 e/$Å^3$ (maximum) in *FOXS* (data not shown). In addition, we varied the contrast of the solvation shell surrounding the DNA origami; i.e. we varied the difference in electron density between the hydration layer and bulk solution, testing values from 0.06 to 0.25 e/$Å^3$ (0.03 e/$Å^3$ is the default value) in *CRYSOL*. The latter is based on a literature value reported for experiments on Mg^{2+} ions dissolved in water.[19] Increasing the contrast or solvent density to even higher numbers would not correspond to physically plausible solution conditions.

Electrostatic potential calculations and estimates of the ion atmosphere

To estimate the extent of the ion atmosphere around the DNA origami objects used in this work, in particular in comparison to simple double-stranded DNA molecules for which the role of ion scattering has been investigated in detail previously,[20-23] we performed simple electrostatic calculations. We calculated the electrostatic potential with a custom-written MATLAB script based on the Debye-Hückel/Poisson-Boltzmann approximation and on the atomistic model of a 35 bp DNA and the switch C, including only the positions of the phosphate atoms of the DNA backbone. Each phosphate atom was described by its position (x_i,y_i,z_i) and modeled as a negatively charged point charge. Moreover, we assumed a Debye-Hückel exponential screening factor to account for the ionic screening due to mobile, dissolved ions. The resulting screened electrostatic potential at a certain position r_i is given by the sum of the electrostatic potential over all phosphate atoms:

$$\Phi(r)_{screen} = \sum_i \frac{-q}{4\pi\varepsilon_0\varepsilon_r r_i} \exp\left(\frac{-r_i}{\lambda_D}\right) \quad \text{(S3)}$$

with the charge of q = 1.602 x 10^{-19} C the vacuum permittivity ε_0 = 8.85 x 10^{-12} F m^{-1}, the relative permittivity of water ε_r = 80.4 and a Debye length λ_D= 9.9 Å corresponding to the high-salt experimental buffer condition for the switch object consisting of 5 mM NaCl and

30 mM $MgCl_2$. We calculated the corresponding ion concentrations around the 35 bp DNA and the switch C assuming the Boltzmann distribution:

$$c^{\pm} = c_0^{\pm} \exp\left(\frac{-q\,\Phi(r)_{screen}}{k\,T}\right) \quad \text{(S4)}$$

where c_0 correspond to the initial ion concentration , kT = 4.11x 10^{-21} J and $\Phi(r)_{screen}$ was computed by Equation S3.

Normal mode refinement of models against SAXS data

To refine the initial model against the experimental SAXS data, we employed a flexible fitting method based on a coarse-grained (one-bead-per-residue) nucleic acid representation and a modified elastic network model that allows large-scale conformational changes while maintaining pseudobonds and secondary structures.[24] This method optimizes a pseudoenergy that combines the modified elastic network model energy with a SAXS-fitting score and a collision energy that penalizes steric collisions. The optimization process effectively uses a weighted combination of normal modes to iteratively improve the fitting of SAXS data. To apply this method to a large DNA object, the following modifications and improvements have been made to the methods described previously.[24] First, each DNA nucleotide is represented by a bead located at the C4' atomic position. All pairs of DNA beads within a cutoff distance of 35 Å are linked to build an elastic network model (see Equation 1 of Ref.[24]). The coarse-grained form factors for DNA nucleotides are taken from the Fast-SAXS-Pro program.[25] Second, to stabilize the local structure of double-stranded DNA, for nucleotide i and i' that form a base pair, additional springs are added between the following pairs: (i, i'), (i±1, i'), (i, i'±1), (i-1, i+1), (i'-1, i'+1). Third, the SAXS fitting score (Equation 10 of Ref.[24]) is modified to the following:

$$E_{SAXS} = f_{SAXS} \cdot N \cdot \mathrm{X}^2 \quad \text{(S5)}$$

with a chi-squared of:

$$\mathrm{X}^2 = \min_c \left\{ \sum_i \frac{[c \cdot I_m(q_i) - I_t(q_i)]^2}{\sigma_i^2} \Big/ \sum_i \frac{[I_t(q_i)]^2}{\sigma_i^2} \right\} \quad \text{(S6)}$$

The constant pre-factor f_{SAXS} = 30, N is the number of DNA beads and q_i is the scattering vector uniformly sampled between 0 and 3 nm^{-1} with an increment of 0.025 nm^{-1}, I_m is the model SAXS profile, I_t is the target SAXS profile measured experimentally, and σ is the experimental error of I_t. Fourth, no hydration shell is modeled, which is expected to have negligible effect on SAXS fitting especially for large molecular systems.[17] Fifth, to reduce memory usage for the large systems, all Hessian matrices except H_{SAXS} in Equation 11 of

Ref.[24] are stored in the sparse matrix format, and the H_{SAXS} term is omitted. The linear equation in Equation 11 of Ref.[24] is solved using the CHOLMOD suite (http://www.cise.ufl.edu/research/sparse). Normal mode refinement calculations were run on an Intel Xeon Processor L5520 (8M Cache, 2.26 GHz); the full refinement of a structure using 209 *q*-values required 120 h of computational time.

Ensemble FRET measurements via donor quenching

Ensemble FRET experiments in solution on DNA origami switch objects were conducted and analyzed as described in the supplementary information of Gerling *et al*.[26] Fluorescently labeled switch particles exhibit low and high FRET signals upon a conformational change from the open to the closed state, respectively. In order to dissect the populations of the closed and open conformation of the switch D sample as a function of $MgCl_2$ concentration, ensemble FRET measurements were performed on switch C, switch O and switch D samples while titrating $MgCl_2$ concentrations in the range of 5 mM to 25 mM (Supplementary Figure S7), as published previously.[26] Ensemble FRET data of the switch C and switch O sample serve as reference samples for the closed and open state, respectively.

To compare the fractions of populations derived from ensemble FRET experiments to the fractions obtained from the two-state model for the switch D sample of the SAXS data, each titration curve was fitted up to a $MgCl_2$ concentration range of 30 mM, assuming a two-state model where the resulting ensemble FRET value (E_{FRET}) is given by:

$$E_{FRET} = f_c \cdot E_c + f_o \cdot E_o \tag{S7}$$

E_c and E_o correspond to ensemble FRET values and the coefficients f_c and f_o are fractional occupancies of the closed and open state, respectively. From statistical thermodynamics of a two-state system it follows that $f_c + f_o = 1$ where f_c is calculated as follows:

$$f_c = \frac{1}{1+\exp\left(-\frac{\Delta G_c}{kT}\right)} \tag{S8}$$

The overall free energy difference between the open and closed state $\Delta G(MgCl_2)$ is given by:

$$\Delta G(MgCl_2) = \Delta G_0 + m_c \cdot c(MgCl_2) \tag{S9}$$

where ΔG_0 is the free energy difference at a reference $MgCl_2$ concentration of 5 mM, and m_c represents the cation sensitivity parameter. We obtained a ΔG_0 = 1.5 kcal/mol and m_c = -0.4 kcal/(mol mM) from a least-square fitting of the experimental data.

Values from the fitted titration curves at low (5 mM) and high (30 mM) $MgCl_2$ concentrations were used to determine the fraction of closed switch D particles. Analogous to the two-state model approach applied on the SAXS data, the ensemble FRET value of the switch D sample (E_d) can be described by a linear superposition of the open and closed states represented by ensemble FRET values of switch O (E_{so}) and switch C (E_{sc}), respectively:

$$E_d(MgCl_2) = f_c \cdot E_{sc}(MgCl_2) + f_o \cdot E_{so}(MgCl_2) \quad \text{(S10)}$$

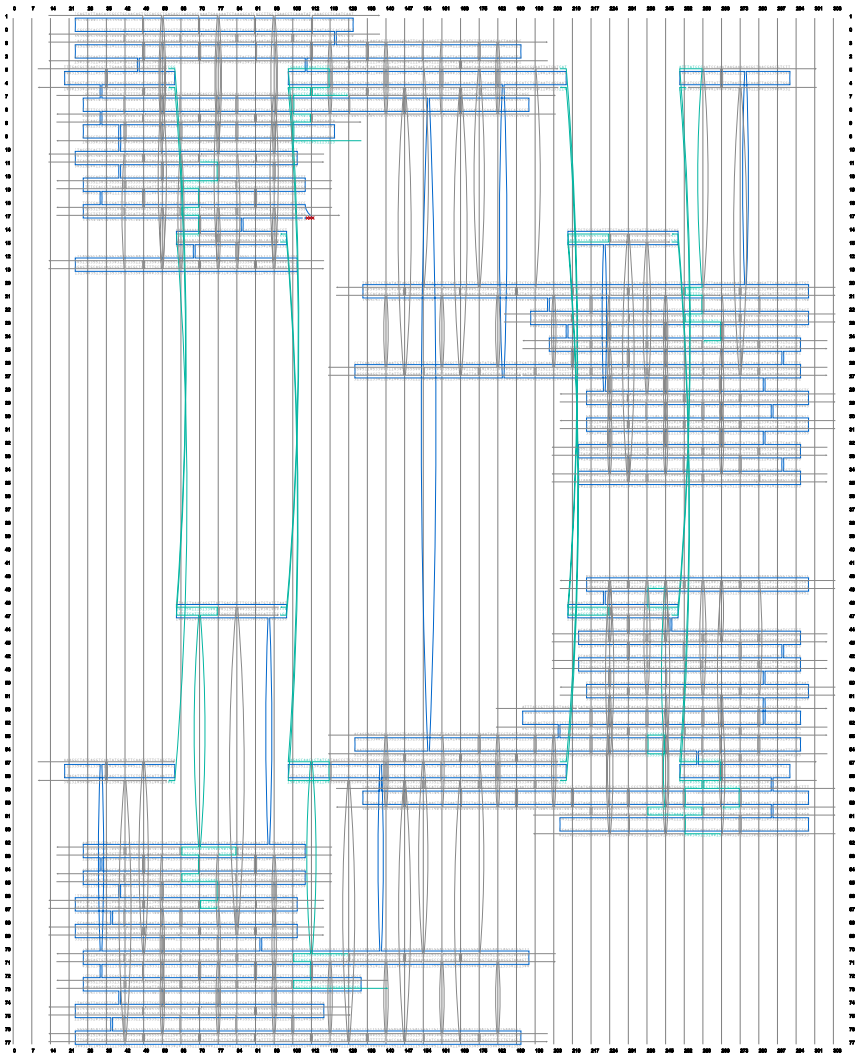
For the evaluation of the conformational state of the switch D sample at low and high $MgCl_2$ concentrations, we averaged fitted ensemble FRET values for each sample taking values at the exact concentration (e.g. E_d(5 mM)) and the values of the precedent and subsequent concentration (i.e. E_d(4 mM) and E_d(6 mM)). Based on these values a least squares fit was performed to determine the closed fraction of the switch D sample at high (30 mM) and low (5 mM) $MgCl_2$ concentrations (Figure 4d). Errors were calculated based on a propagation of uncertainty.

Supplementary Table

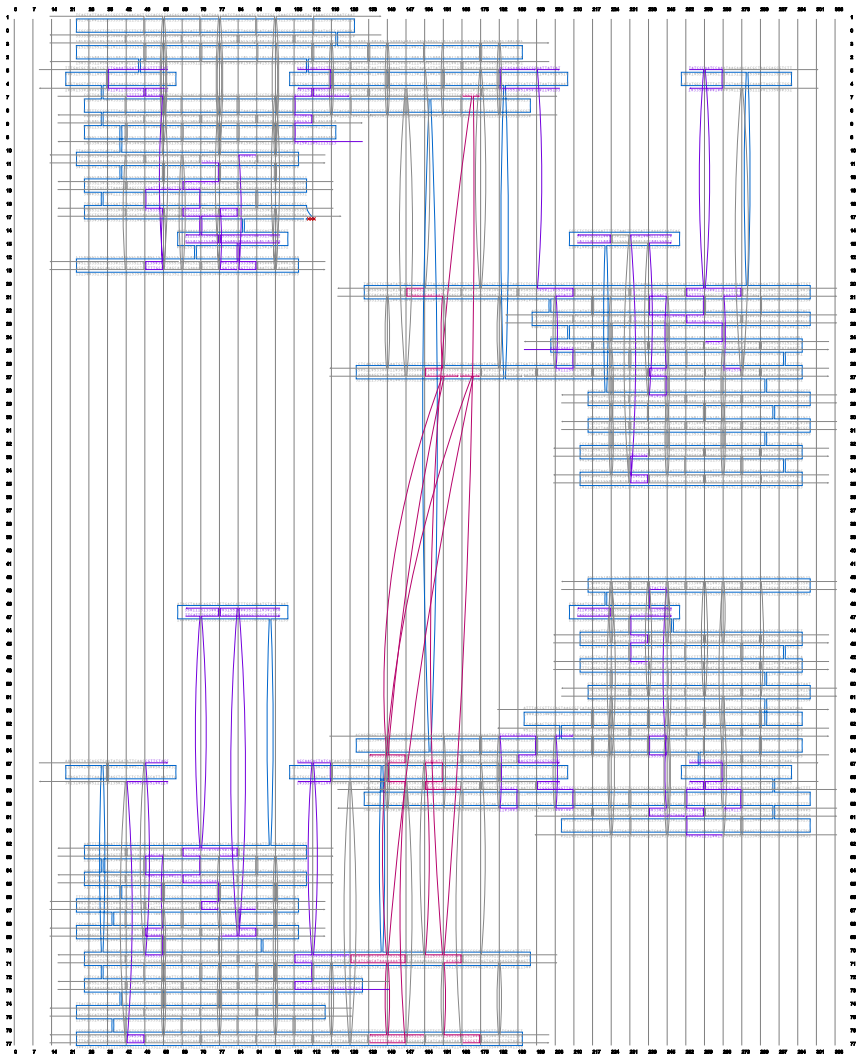
Supplementary Table S1. Comparison of the radius of gyration (R_g) and the cross-sectional R_g (R_c) determined from the theoretical scattering profiles calculated with *CRYSOL, FOXS* and a custom written *C* script for the static switch samples. For Guinier analysis the fitting range $q_{max}R_g < 1.3$ was used and for R_c calculations a q-range of ~ 0.2 – 0.9 nm^{-1} (1.3 nm^{-1}) for the switch C (switch O) object was defined.

Sample	R_g (nm)	R_c (nm)
CRYSOL (closed)	28.2	6.4
FOXS (closed)	28.1	6.3
Custom written (closed)	27.6	6.4
CRYSOL (open)	29.5	4.4
FOXS (open)	29.4	4.4
Custom written (open)	29.4	4.4

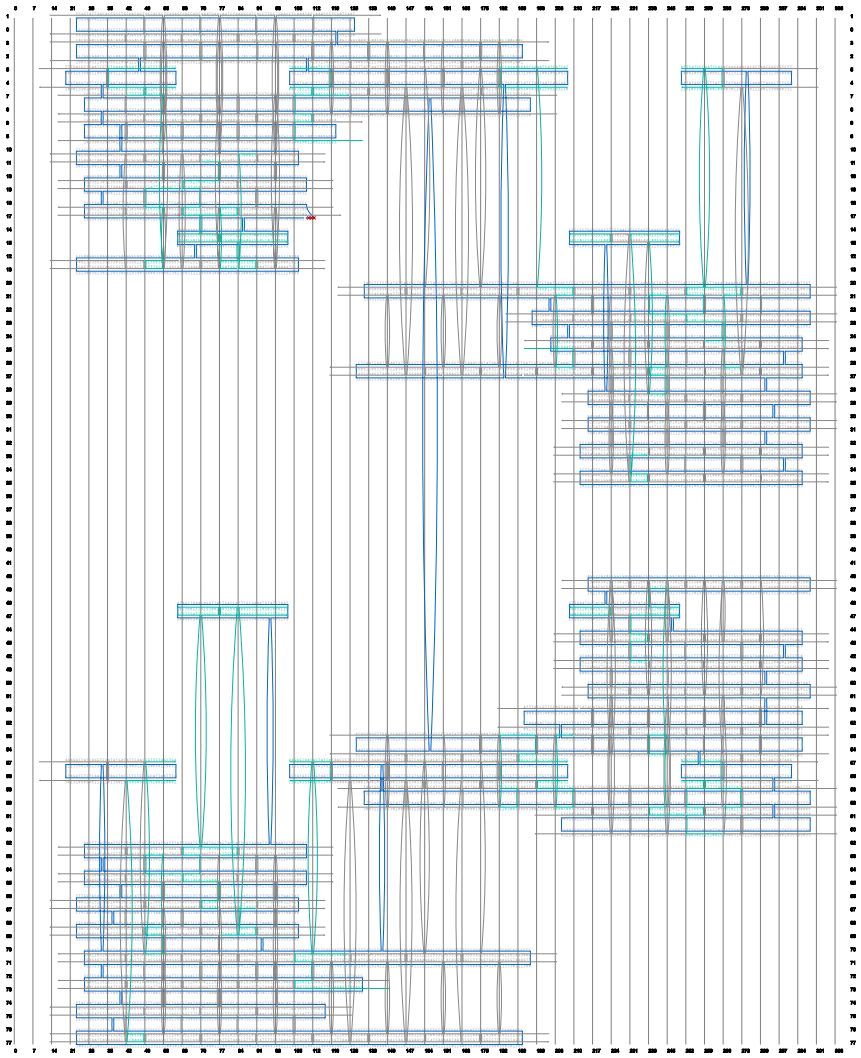
Supplementary Figures



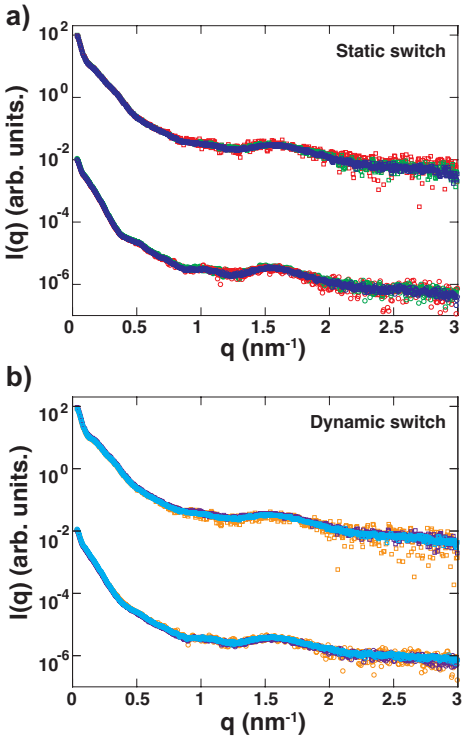
Supplementary Figure S1. Strand diagram of the static closed switch (switch C) variant. Scaffold (shown in blue) and staple layout of the switch C variant. Cyan: 3-bases sticky end. Generated with caDNAno v0.2.



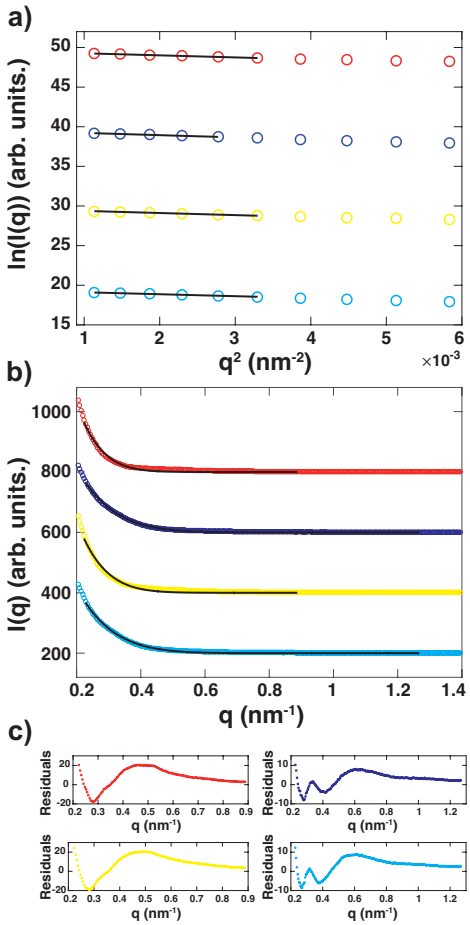
Supplementary Figure S2. Strand diagram of the static open switch (switch O) variant. Scaffold (shown in blue) and staple layout of the switch O variant. Purple: stacking deactivated; Pink: additional connections between the upper and bottom arm. Generated with caDNAno v0.2.



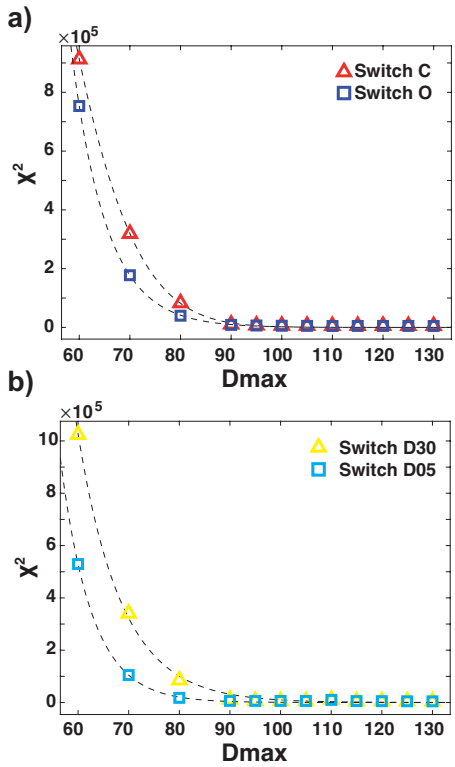
Supplementary Figure S3. Strand diagram of the dynamic (switch D) variant. Scaffold (shown in blue) and staple layout of the dynamic switch variant with 16 activated stacking interactions. Cyan: stacking activated. Generated with caDNAno v0.2.



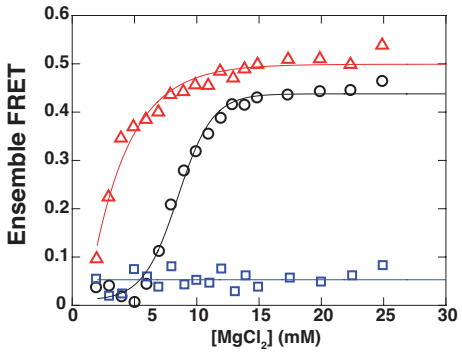
Supplementary Figure S4. Concentration scaled scattering profiles for the static (switch C and switch O, top) and the dynamic (switch D30 and switch D05, bottom) switch variants. (a) Averaged scattering profiles for the switch C (circles, bottom) and switch O (squares, top) measured at varying concentrations: 25 nM (red), 50 nM (green) and 100 nM (blue). (b) Averaged scattering profiles for the switch D30 (circles, bottom) and switch D05 (squares, top) for applied concentrations of 25 nM (orange), 50 nM (purple) and 100 nM (cyan). Profiles are scaled by their concentration and the lower scattering profiles are vertically offset for clarity.



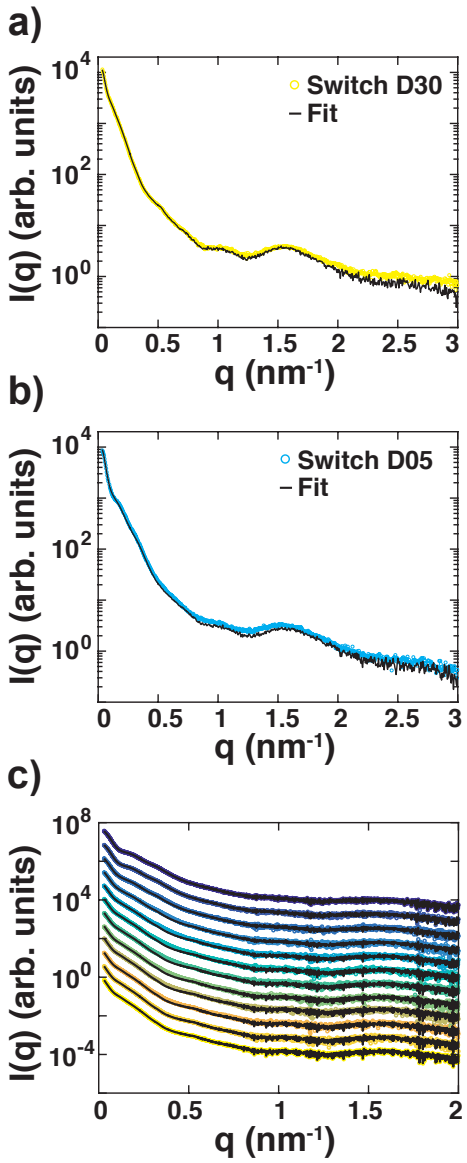
Supplementary Figure S5. Guinier analysis and fits for the cross-sectional scattering intensity of static and dynamic switch samples. (a) Guinier representation of the experimental scattering data for switch C (red, top), switch O (blue), switch D30 (yellow) and switch D05 (cyan, bottom). The Guinier fits are indicated by black lines covering a q-range of $qR_g < 1.6$. Profiles are vertically offset for clarity. (b) Fits of the cross-sectional scattering intensities to experimental data shown in (a) (same color code as in (a)) for the q-range $q_{min} = 2\pi/R_g$ and $q_{max} = 2\pi/R_c$, where the R_c is given by $R_c^2 = R^2/2$. For switch C and switch D30 a radius of R = 10 nm and for switch O and switch D05 a radius of R = 7 nm was assumed for the fit. Profiles are vertically offset for clarity. (c) Residuals for data from (b) (same color code as in (b)).



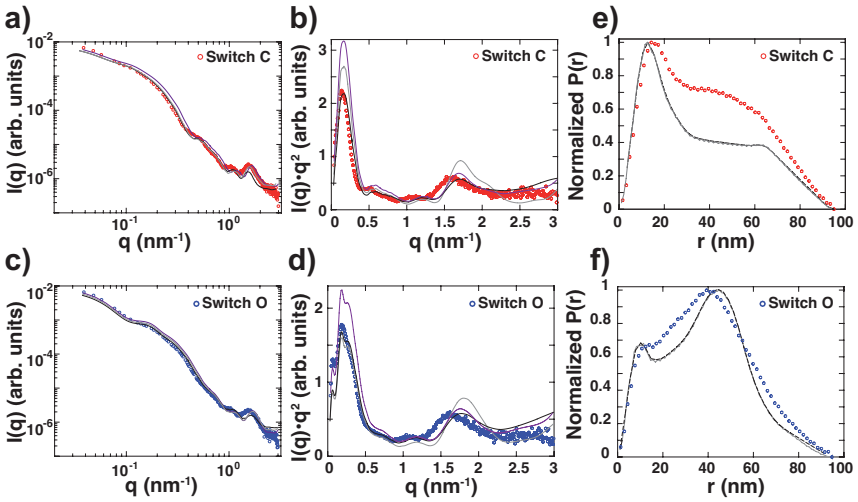
Supplementary Figure S6. Determination of an adequate D_{max} value used for the computation of the pair distance distribution function *P(r)*. (a) χ^2 describing the discrepancy between the experimental data and the fit as a function of different D_{max} values evaluated for the static switch versions switch C and switch O. (b) Corresponding analysis for the dynamic switch versions switch D30 and switch D05. For all samples a mimimun of χ^2 around 95 nm was found, which was applied to calculate the *P(r)* function for each sample.



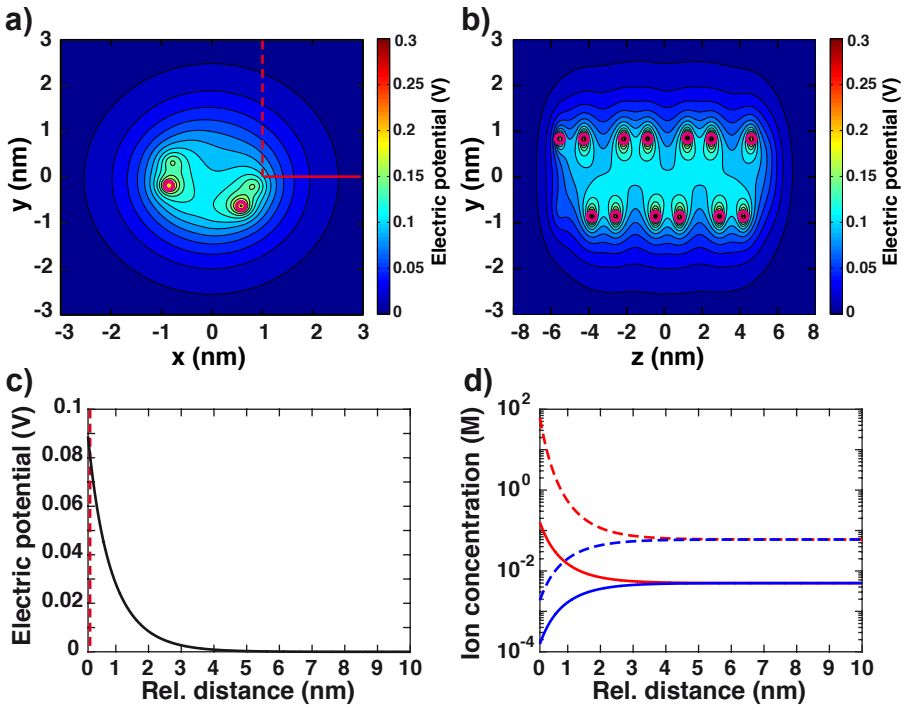
Supplementary Figure S7. Ensemble FRET measured via donor quenching for a titration of $MgCl_2$ in solutions containing switch C (red triangles), switch D (black circles) and the dynamic switch variant with all click contacts deactivated (blue squares). Solid lines represent a two-state model with a free energy term that depends linearly on the $MgCl_2$ concentration (see Equation S9).



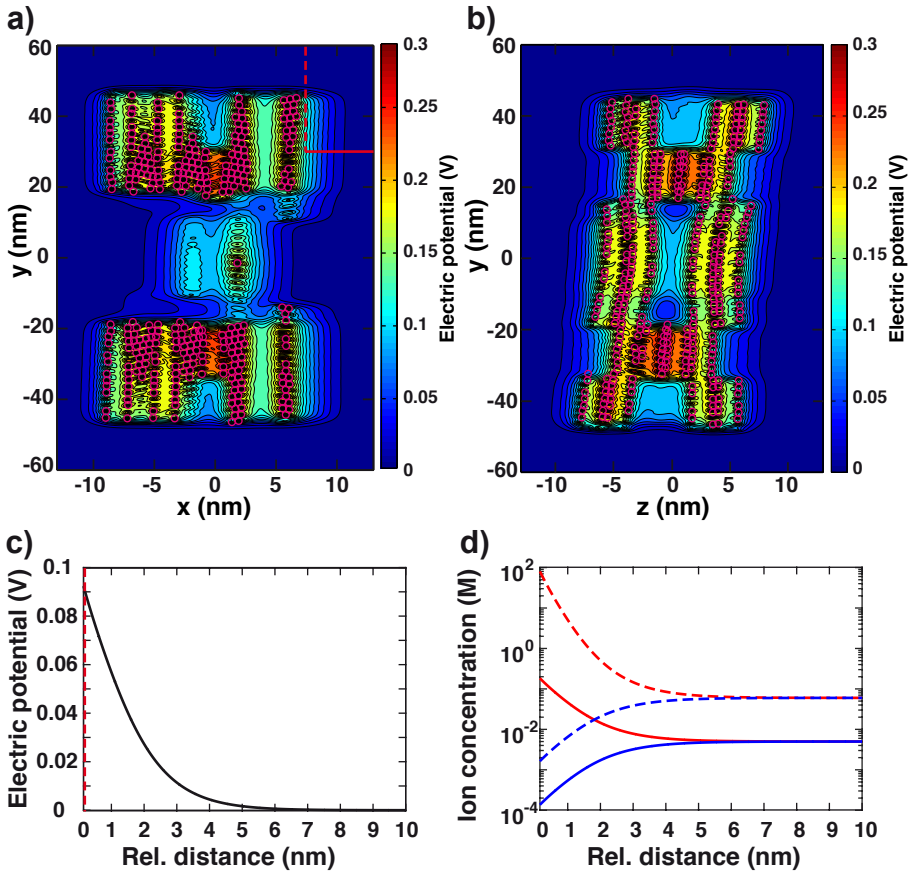
Supplementary Figure S8. Fits of two-state models for the dynamic switch variants (switch D). (a) Scattering profile (yellow) of the switch D30 sample (30 mM $MgCl_2$) and the fitted profile (black) obtained from the two-state model (see main text). (b) Scattering profile (cyan) of the switch D05 sample (5 mM $MgCl_2$) and the fitted profile of the two-state model (black). (c) Scattering profiles of switch D for varying $MgCl_2$ concentrations: 3 (dark blue, bottom), 5, 8, 10, 12, 14, 15, 16, 18, 20, 25 and 30mM (light yellow, top) and fitted profiles from a two-state model (see main text). Data are vertically offset for clarity.



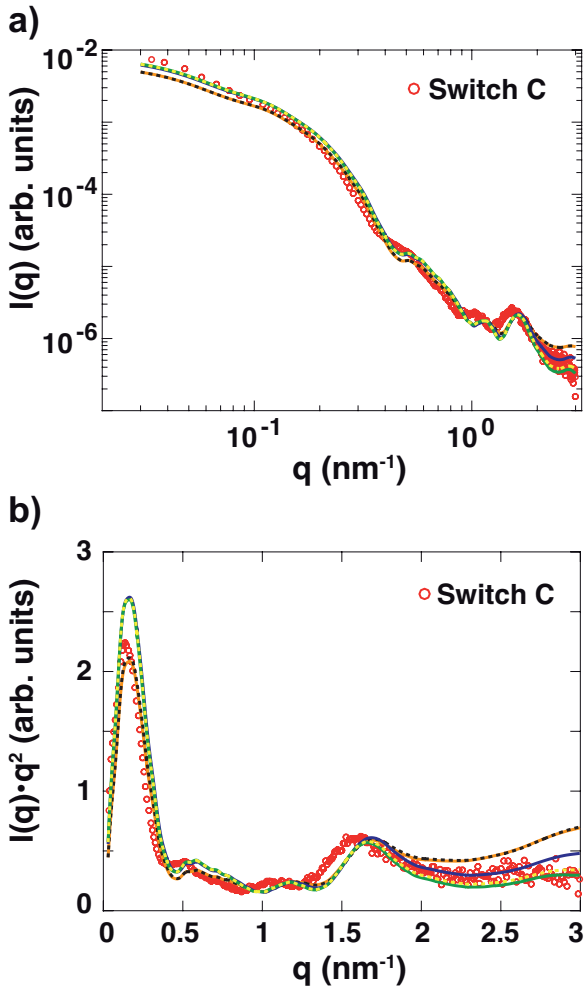
Supplementary Figure S9. Comparison of the experimental and theoretical scattering profiles and P(r) functions that were predicted from the atomistic models derived from CanDo of the static switch samples. (a) Fits including an additional constant in the theoretical scattering profiles calculated with the program *CRYSOL* (black line), *FOXS* (gray line) and a custom written *C* script (purple line) to the experimental data of the switch C sample (red circles). For the custom written routine only one bead per base was assumed. (b) Kratky representation of the data shown in (a) (same color code as in (a)). (c) Fits including an additional constant in the theoretical scattering profiles calculated with the program *CRYSOL*, *FOXS* and a custom written *C* script (same color code and computational settings as in (a)) to the experimental data of the switch O sample (blue circles). (d) Kratky representation of the data shown in (c) (same color code as in (c)). (e,f) P(r) functions of switch C (red circles) and switch O (blue circles) based on experimental scattering data and from profiles derived from *FOXS* (gray line). Black dashed lines correspond to histograms of distances calculated directly from the atomistic models. P(r) data were normalized to unity.



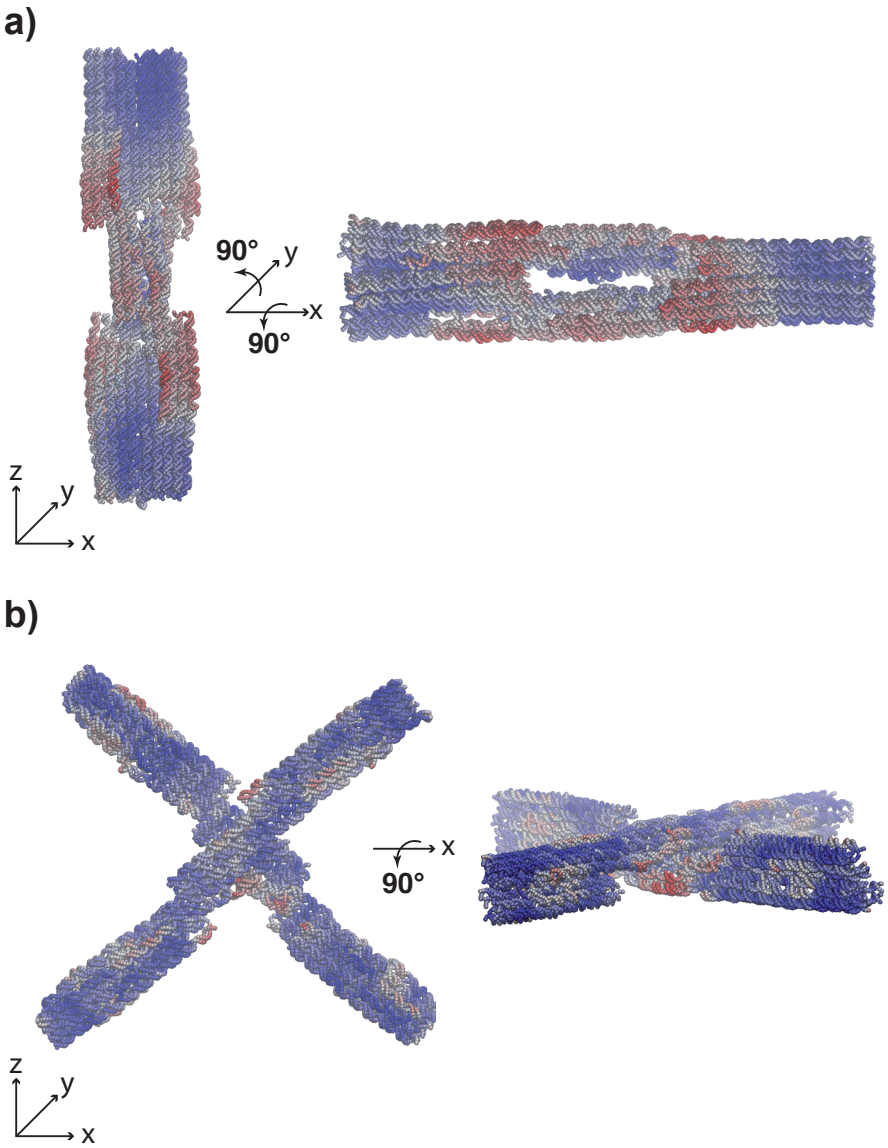
Supplementary Figure S10. Electrostatic potential and ion distribution calculations to estimate the ion atmosphere. (a,b) Contour plots of the screened electrostatic potential for an atomistic model considering only phosphate charges of a 35 bp DNA, showing cross-sections in the x-y plane (a) and y-z plane (b). Magenta circles indicate positions of phosphate atoms. (c) Electric potential along the x-axis (for y,z = 0 nm) corresponding to a solvent-accessible area outside the 35 bp DNA (indicated by the vertical (dashed) and horizontal red lines). (d) Concentration of mono- and divalent ions for the same area as in (c) corresponding to buffer conditions of 5 mM NaCl and 30 mM $MgCl_2$: Mg^{2+} (red dashed line), 2 x Cl^- (blue dashed line), Na^+ (red line) and Cl^- (blue line).



Supplementary Figure S11. Electrostatic potential and ion distribution calculations to estimate the ion atmosphere. (a,b) Contour plots of the screened electrostatic potential for an atomistic model considering only phosphate atoms of switch C, showing cross-sections in the x-y plane (a) and y-z plane (b). Magenta circles indicate positions of phosphate atoms. (c) Electric potential along the x-axis (for y = 30 nm; z = 0 nm) corresponding to a solvent-accessible area outside the switch C structure (indicated by the vertical (dashed) and horizontal red lines). (d) Concentration of mono- and divalent ions for the same area as in (c) corresponding to buffer conditions of 5 mM NaCl and 30 mM $MgCl_2$: Mg^{2+} (red dashed line), 2 x Cl^- (blue dashed line), Na^+ (red line) and Cl^- (blue line).



Supplementary Figure S12. Computations of the scattering profiles with varying hydration shell conditions in *CRYSOL*. Linear fits including a constant offset of theoretical scattering profiles calculated with *CRYSOL* with contrast values of the hydration shell of 0.06 e/Å^3 (black dashed line), 0.12 e/Å^3 (orange line) and 0.25 e/Å^3 (green line) (solvent density = 0.334 e/Å^3 for all three profiles), with a solvent density value of 0.344 e/Å^3 and default contrast (blue line) and with a solvent density value of 0.344 e/Å^3 setting the contrast value to 0.25 e/ Å^3 (yellow dashed line). (b) Data from (a) in Kratky representation.



Supplementary Figure S13. Visualization of the switch deformations from the normal mode refinement. (a) Local deformations for switch C (defined as RMSD against the initial CanDo derived structure) are indicated by the colored code where blue and red areas correspond to minimal and maximal deformations, respectively. (b) Local deformations for switch O, same color code as in (a).

Supplementary References

(1) Douglas, S. M.; Marblestone, A. H.; Teerapittayanon, S.; Vazquez, A.; Church, G. M.; Shih, W. M. *Nucl. Acids Res.* **2009**, *37* (15), 5001–5006.
(2) Kick, B.; Praetorius, F.; Dietz, H.; Weuster-Botz, D. *Nano Lett.* **2015**, *15* (7), 4672–4676.
(3) Sobczak, J.-P. J.; Martin, T. G.; Gerling, T.; Dietz, H. *Science* **2012**, *338* (6113), 1458–1461.
(4) Stahl, E.; Martin, T. G.; Praetorius, F.; Dietz, H. *Angewandte Chemie* **2014**, *126* (47), 12949–12954.
(5) Svergun, D. I.; Koch, M. H. J. *Rep. Prog. Phys.* **2003**, *66* (10), 1735–1782.
(6) Lipfert, J.; Herschlag, D.; Doniach, S. *Methods Mol. Biol.* **2009**, *540*, 141–159.
(7) Pernot, P.; Round, A.; Barrett, R.; De Maria Antolinos, A.; Gobbo, A.; Gordon, E.; Huet, J.; Kieffer, J.; Lentini, M.; Mattenet, M.; Morawe, C.; Mueller-Dieckmann, C.; Ohlsson, S.; Schmid, W.; Surr, J.; Theveneau, P.; Zerrad, L.; McSweeney, S. *J Synchrotron Radiat* **2013**, *20* (Pt 4), 660–664.
(8) Blanchet, C. E.; Spilotros, A.; Schwemmer, F.; Graewert, M. A.; Kikhney, A.; Jeffries, C. M.; Franke, D.; Mark, D.; Zengerle, R.; Cipriani, F.; Fiedler, S.; Roessle, M.; Svergun, D. I. *J. Appl. Cryst (2015).* **2015**, 48, 431-443.
(9) Guinier, A. *Ann. Phys. (Paris)* **1939**, 12, 161.
(10) Hammel, M.; Yu, Y.; Mahaney, B. L.; Cai, B.; Ye, R.; Phipps, B. M.; Rambo, R. P.; Hura, G. L.; Pelikan, M.; So, S.; Abolfath, R. M.; Chen, D. J.; Lees-Miller, S. P.; Tainer, J. A. *J. Biol. Chem.* **2010**, *285* (2), 1414–1423.
(11) Kratky, O.; Glatter, O. *Academic Press, London* **1982**, 1–267.
(12) Moore, P. B. *J Appl Crystallogr* **1980**, *13* (2), 168–175.
(13) Shannon, C. E. *Bell System Technical Journal* **2013**, *27* (4), 623–656.
(14) Pan, K.; Kim, D.-N.; Zhang, F.; Adendorff, M. R.; Yan, H.; Bathe, M. *Nature Communications* **2014**, *5*, 5578–7.
(15) Svergun, D.; Barberato, C.; Koch, M. *J Appl Crystallogr* **1995**, *28* (6), 768–773.
(16) Schneidman-Duhovny, D.; Hammel, M.; Sali, A. *Nucl. Acids Res.* **2010**, *38* (Web Server issue), W540–W544.
(17) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. *Biophys. J.* **2013**, *105* (4), 962–974.
(18) Walther, D.; Cohen, F. E.; Doniach, S. *J Appl Crystallogr* **2000**, *33* (2), 350–363.
(19) Waluyo, I.; Huang, C.; Nordlund, D.; Bergmann, U.; Weiss, T. M.; Pettersson, L. G. M.; Nilsson, A. *J Chem Phys* **2011**, *134* (6).
(20) Meisburger, S. P.; Pabit, S. A.; Pollack, L. *Biophysj* **2015**, *108* (12), 2886–2895.
(21) Andresen, K.; Das, R.; Park, H. Y.; Smith, H.; Kwok, L. W.; Lamb, J. S.; Kirkland, E. J.; Herschlag, D.; Finkelstein, K. D.; Pollack, L. *Phys. Rev. Lett.* **2004**, *93* (24).
(22) Das, R.; Mills, T. T.; Kwok, L. W.; Maskel, G. S.; Millett, I. S.; Doniach, S.; Finkelstein, K. D.; Herschlag, D.; Pollack, L. *Phys. Rev. Lett.* **2003**, *90* (18), 188103.
(23) Pollack, L. *Biopolymers* **2011**, *95* (8), 543–549.
(24) Zheng, W.; Tekpinar, M. *Biophys. J.* **2011**, *101* (12), 2981–2991.
(25) Ravikumar, K. M.; Huang, W.; Yang, S. *The Journal of Chemical Physics* **2013**, *138* (2), 024112–024118.
(26) Gerling, T.; Wagenbauer, K. F.; Neuner, A. M.; Dietz, H. *Science* **2015**, *347* (6229), 1446–1452.

6.3 Publication P6: A Mo-anode-based In-house Source for Small-Angle X-ray Scattering Measurements of Biological Macromolecules

A Mo-anode-based In-house Source for Small-Angle X-ray Scattering Measurements of Biological Macromolecules

by

Linda K. Bruetzel, Stefan Fischer, Annalena Salditt, Steffen M. Sedlak, Bert Nickel, and Jan Lipfert

published in

Review of Scientific Instruments

REVIEW OF SCIENTIFIC INSTRUMENTS **87**, 025103 (2016)

A Mo-anode-based in-house source for small-angle X-ray scattering measurements of biological macromolecules

Linda K. Bruetzel, Stefan Fischer, Annalena Salditt, Steffen M. Sedlak, Bert Nickel, and Jan Lipfert[a)]
Department of Physics, Nanosystems Initiative Munich, and Center for Nanoscience, Ludwig-Maximilians-University Munich, Amalienstr. 54, 80799 Munich, Germany and Geschwister-Scholl Platz 1, 80539 Munich, Germany

(Received 17 November 2015; accepted 15 January 2016; published online 8 February 2016)

We demonstrate the use of a molybdenum-anode-based in-house small-angle X-ray scattering (SAXS) setup to study biological macromolecules in solution. Our system consists of a micro-focus X-ray tube delivering a highly collimated flux of 2.5×10^6 photons/s at a beam size of 1.2×1.2 mm^2 at the collimation path exit and a maximum beam divergence of 0.16 mrad. The resulting observable scattering vectors q are in the range of 0.38 Å^{-1} down to 0.009 Å^{-1} in SAXS configuration and of 0.26 Å^{-1} up to 5.7 Å^{-1} in wide-angle X-ray scattering (WAXS) mode. To determine the capabilities of the instrument, we collected SAXS data on weakly scattering biological macromolecules including proteins and a nucleic acid sample with molecular weights varying from ~12 to 69 kDa and concentrations of 1.5–24 mg/ml. The measured scattering data display a high signal-to-noise ratio up to q-values of ~0.2 Å^{-1} allowing for an accurate structural characterization of the samples. Moreover, the in-house source data are of sufficient quality to perform *ab initio* 3D structure reconstructions that are in excellent agreement with the available crystallographic structures. In addition, measurements for the detergent decyl-maltoside show that the setup can be used to determine the size, shape, and interactions (as characterized by the second virial coefficient) of detergent micelles. This demonstrates that the use of a Mo-anode based in-house source is sufficient to determine basic geometric parameters and 3D shapes of biomolecules and presents a viable alternative to valuable beam time at third generation synchrotron sources. © *2016 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4940936]

I. INTRODUCTION

Small-angle X-ray scattering (SAXS) is a powerful tool to investigate the structure and interactions of biological macromolecules in solution.[1–3] SAXS has the important advantage of being a solution-based technique, thus obviating the need for sample crystallization and enabling studies of biological macromolecules in a range of solution conditions, from (near-) physiological to highly denaturing.[4] In the past, SAXS data have frequently been used to determine basic parameters of macromolecules in solution, such as the radius of gyration[5,6] (R_g) and the maximum intramolecular distance[7] (D_{max}). Determination of, e.g., R_g under varying solution conditions has provided important insights into protein[4] and RNA folding[8] and into the nature of the unfolded states.[9,10] Nonetheless, the utility of SAXS data has been tremendously enhanced in the last two decades through the increasing availability of algorithms to determine and to compare the (low resolution) 3D structures of macromolecules from 1D scattering profiles. In particular, a number of algorithms now make it possible to obtain low resolution 3D "bead" models from SAXS data for proteins[11–13] and for nucleic acids[14] without any other prior information about the sample. In addition, if SAXS data can be combined with prior information from, e.g., FRET,[15] NMR,[16] crosslinking, or known crystal structures[3,6] even more refined molecular models can be obtained.

Current SAXS measurements often rely on state-of-the-art synchrotron sources, in particular, due to their high photon flux and tunability. Nonetheless, in-house anode-based sources remain an important and attractive alternative,[17] in particular, given the limited availability of measurement time at synchrotron user facilities and the considerable logistic challenges to carry out measurements at an-often far away-off-site location. In principle, it is possible to compensate the reduced flux at lab-sources at least partially by extended counting times, e.g., by increasing exposure times from ~1 s typical for biological samples at high-flux synchrotrons to 10^3–10^4 s. However, this approach only works if the signal-to-noise ratio is high, i.e., if the background noise does not increase too much for long integration times. Therefore, the question which energy range is best suited for SAXS measurements of macromolecules in solution is tightly connected to the question which energy range provides the best signal-to-noise ratio for these conditions.

Currently, most in-house based sources employ copper (Cu) anodes with K_α radiation at 8.0 keV, corresponding to a wavelength of 1.54 Å. Their application for solution SAXS measurements on weakly scattering biological samples has been already proven[17] and also *ab initio* reconstructions with

[a)]Author to whom correspondence should be addressed. Electronic mail: Jan.Lipfert@lmu.de

the programs DAMMIF[18] and GASBOR[12] could be performed successfully.[19,20] An alternative to Cu anodes are molybdenum (Mo; K_α = 17.4 keV) anode sources, which provide shorter wavelength X-rays compared to Cu, with a characteristic wavelength of 0.71 Å. Mo-anode sources have been employed to investigate macromolecules, powders or thin film alloys by SAXS,[21] wide-angle X-ray scattering (WAXS),[21–23] grazing-incidence X-ray scattering (GISAXS),[24] crystallography,[25] and diffractometry.[26,27] However, a detailed description and analysis of a Mo-anode-based in-house setup for SAXS measurements on biological macromolecules in solution is still lacking. Here, we present a comprehensive specification and characterization of a Mo-anode in-house source for SAXS measurements on proteins, nucleic acids, and detergent micelles. The shorter wavelength of Mo has a number of potential advantages: First, since the absorption coefficient for X-rays decreases sharply with increasing energy, higher energies cause less radiation damage in the sample.[28] Second, scattering from air and window materials in the beam path is also reduced at higher X-ray energies. Third, the reduced absorption coefficient means that the optimal beam path (μ^{-1}) in the sample is longer for higher X-ray energies, e.g., for water μ^{-1}~10 mm for Mo and μ^{-1}~1 mm for Cu radiation, which can be advantageous for samples handling, i.e., for sample environments that benefit from larger sample dimensions. Fourth, since the magnitude of the scattering vector q is inversely proportional to the X-ray wavelength λ, a shorter wavelength is highly beneficial to perform WAXS measurements, where high q values are desired. Thus, a shorter wavelength as given by Mo anodes, facilitates the combination of SAXS and WAXS measurements within a single setup, which can be advantageous for structural studies on biological samples such as proteins or peptides.[29,30]

We test our Mo-anode setup on a panel of typical, weakly scattering biological samples, including several proteins (bovine serum albumin, horse heart cytochrome *c*, and chicken egg white lysozyme), a nucleic acid sample (24 bp DNA duplex), and a micelle forming detergent (decyl-maltoside; DM). These samples have been investigated previously at third generation synchrotron sources and (except for the micelle sample) have known crystallographic structures, enabling a critical comparison and evaluation of our in-house data. The results suggest that our Mo-anode-based source achieves good signal-to-noise even on weakly scattering samples; the data are of sufficient quality to carry out standard SAXS analyses, such as Guinier fitting of the R_g, and to obtain *ab initio* 3D shape reconstructions for the protein and nucleic acid samples that exhibit good agreement with the known crystallographic structures. In addition, the data permit to fit a two-component ellipsoid model to the DM micelle data and to determine the size, shape, and interactions of the detergent micelles in solution.

II. THE X-RAY SETUP

In brief, the in-house setup consists of a microfocus X-ray source with multilayer optics corresponding to the K_α line of the target, a collimation path with two scatterless slits, a motorized sample stage, two exchangeable vacuum tubes, and a hybrid pixel detector (Fig. 1). The individual components are described in detail in Secs. II A–II E. The setup is optimized for SAXS measurements but can also be used for WAXS and diffraction measurements, as shown previously[31] and discussed only briefly here.

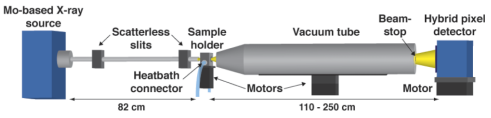


FIG. 1. Schematic layout of the in-house setup for SAXS measurements. For further details see Section II in the text.

A. X-ray source and collimation path

Our system consists of a Mo GeniX^{3D} microfocus X-ray tube (Xenocs SA, Sassenage, France) combined with FOX2D single reflection optics delivering a monochromatic and highly stable beam with an X-ray energy of 17.4 keV. The flux is typically around 2.5×10^6 photons/s at the sample stage. For collimation the beam enters an 82 cm long, fully evacuated collimation path closed by a 25 μm thick Kapton foil at the end. Collimation is achieved by integrating two partially motorized scatterless aperture slits (Xenocs SA, Sassenage, France),[32] one upstream right at the mirror and the second at the tube exit. The scatterless slits consist of a rectangular single Ge-crystal substrate bonded to a metal base with a large tapering angle away from the beam, which significantly reduces parasitic scattering and enhances resolution compared to conventional X-ray apertures.[32] Moreover, their insertion leads to a simplified optical design in comparison to previous implementations, which required three apertures.[33] With this optical configuration, we achieve a highly collimated (horizontal divergence: 0.12 mrad, vertical divergence: 0.16 mrad FW20%M) beam with a size of approximately 1.2×1.2 mm^2 at the collimation path exit.

B. Sample cell and sample stage

The sample stage is positioned 5 cm in front of the collimation path exit. It consists of a fully motorized platform where six stepper motors allow moving the stage in horizontal and vertical directions with 5 μm and 0.1 μm precision, respectively, and enable rotation of the stage about all three axes with 0.005° precision. A rectangular aluminum sample holder with two slots for sample chambers (adapted from Ref. 34) is connected to the platform for successive automated measurements of sample and buffer solution (Fig. S1 of the supplementary material).[35] The bottom part of the sample holder is connected to a circulating water bath (F12-MA, Julabo, Germany) via rubber tubes enabling temperature control of the sample cells in the range of 4–70 °C (±0.8 °C). For SAXS measurements conducted at room temperature, we used polyvinyl chloride (PVC) based sample chambers. The cylindrical observation volume of the chambers is filled with sample solution via two small inlets with a diameter of 0.6 mm from the top. This design prevents the formation of air bubbles and minimizes evaporation during measurements. In

order to achieve the optimum scattering signal, the chamber length corresponds to the absorption length of Mo in water, equal to the absorption coefficient, which is approximately 10 mm[36] The diameter of the cylindrical observation volume is 3 mm so that parasitic scattering due to interactions of the incoming beam with the PVC is avoided and the opening angle is large enough so that scattering events happening at the beginning of the chamber can still be detected. The overall sample volume is roughly 80 μl. The windows of the sample chamber are made of 25 μm thick potassium aluminosilicate (muscovite mica) sheets (Goodfellows Cambridge Ltd., UK) that are glued on both openings of the cell with two-component epoxy (UHU Ltd., Germany). The use of mica as window material only introduces ~6% attenuation (assuming an attenuation length of 800 μm for mica[36]) and it suppresses evaporation of sample solution during measurements. Furthermore, the windows do not cause a detectable background signal (see the text and Fig. S2 of the supplementary material).[35] For temperature-controlled experiments, we used sample chambers made from aluminum for improved thermal coupling. The sample holder can be removed to perform calibration measurements with glassy carbon, silver behenate (AgBe), and lanthanum hexaboride (LaB_6). AgBe and LaB_6 are stored in aluminum chambers closed by 25 μm thick Kapton foils. The glassy carbon sample is fixed in an aluminum holder and mounted directly below the AgBe chamber. The calibration materials can be placed into the beam by a motor.

C. Evacuated flight path and beamstop

Our setup for SAXS measurements can be switched between two sample-detector distances of 110 cm and 250 cm, corresponding to q-ranges of 0.014–0.38 Å^{-1} and 0.009–0.15 Å^{-1}, respectively. A distance of 250 cm corresponds to more pixels covering the low q-range. For the two measurement modes, vacuum tubes with lengths of 95 cm and 180 cm, respectively, are placed between the sample stage and the detector in order to reduce air scattering (Fig. 1). A detailed analysis of various sources of background signals and, in particular, of the influence of air scattering on the SAXS data is given in the text and Fig. S2 of the supplementary material.[35] Kapton foils with a thickness of 25 μm at the front and 50 μm at the end seal the ends of each vacuum tube. The vacuum tube has a diameter of 3 cm at the front and 10 cm at the back. We integrate beamstops at the end of the vacuum tubes by gluing circular lead tapes with diameter of 3 mm (for the 95 cm vacuum tube) and 4 mm (for the 180 cm vacuum tube) at the center of the Kapton foil inside the vacuum tube. The entire vacuum tube can be moved by two stepper motors in vertical and horizontal directions allowing for accurate alignment of the beamstop. This configuration is advantageous, as it does not introduce any additional shadow effects from a beamstop holder, and as it avoids air scattering compared to a beamstop positioned outside of the vacuum. Moreover, the lead tape is slightly transparent to the beam so that fluctuations in the beam position can be detected. In the WAXS geometry, the sample-to-detector distance is set to 32 cm resulting in a q-range of 0.26–5.7 Å^{-1}. Due to the relatively short sample-to-detector distance, we do not employ an evacuated flight path.

D. Detector

For X-ray detection, we use a CMOS hybrid pixel detector (Pilatus 100K, Dectris Ltd, Switzerland) with a sensor thickness of 1000 μm yielding a quantum efficiency at molybdenum K_α-energy of 76%, which is limited by the absorption of silicon (μ^{-1} (Mo) ~700 μm for silicon). The detector area consists of 487×195 pixels with a pixel size of 172 μm in both directions, which leads to a total size of 83.8×33.5 mm^2 (width × height). The dynamic range is 20 bits, corresponding to 1 048 576 photons. Hybrid pixel detectors are single photon counters with the advantage of low background and the absence of dark noise.[37] For the SAXS configuration with the short vacuum tube, the detector is attached to a stepper motor that moves the detector in vertical direction. This stepper motor is fixed to a custom-made focusing rail, which can be moved manually in order to align the detector in the horizontal direction. For the configuration with the long sample-detector distance, the detector is fixed in vertical position such that the beam is centered and can be adjusted manually in the horizontal dimension. For WAXS measurements, the detector is placed on a motorized stage (BiSlide, Velmex, Inc.). The motorized stage can be moved in horizontal and vertical direction with a travel range of 25.4 cm for automated scanning and stitching of the detector images.

E. Software

We control the instrument components and perform data acquisition using the UNIX-based software package "spec" (Certified Scientific Software, Cambridge, USA) which is widely used for X-ray scattering and diffraction experiments at synchrotrons and laboratory systems. The "spec" can directly communicate with the Pilatus detector via macros (downloaded from the Dectris website: www.dectris.com). Furthermore, a custom-written Matlab routine displays the live image of the detector for fast and easy adjustment.

III. MATERIALS AND METHODS

A. Calibration standards

We utilize silver behenate (AgBe; VWR International, Germany) to calibrate the beam center position and sample-to-detector distance for all small-angle measurements (Fig. S3 of the supplementary material).[35,38] In addition, we use a pre-calibrated 1 mm thick glassy carbon sample (kindly provided by Dr. Jan Ilavsky, APS, Argonne National Laboratory, USA) for the calibration of the recorded intensity to absolute scattering cross section units[39] of cm^{-1} and sr^{-1}, which enables the comparison of scattering data from different instruments. For the calibration of the wide-angle configuration, we use lanthanum hexaboride (LaB_6; SRM 660c, NIST) (Fig. S3 of the supplementary material).[35]

TABLE I. Overview of samples with corresponding concentrations and buffers used for SAXS measurements.

Sample	Number of residues/nucleotides	Molecular weight (kDa)	Concentrations (mg/ml)	Buffer
BSA	583	69.0	5	50 mM HEPES, pH = 7.5, 50 mM KCl
Cytochrome *c*	104	12.4	2, 8, 24	100 mM acetate buffer, pH = 4.6, 0.5M guanidinium hydrochloride
Lysozyme	129	14.3	5, 10, 20	40 mM acetate buffer, pH = 4.5, 150 mM NaCl
24 bp DNA duplex	48	~14.6	1.5, 4.4	50 mM sodium 3-(*N*-morpholino) propanesulfonic acid, pH = 7.0, 150 mM NaCl

B. Sample preparation

A 24 bp DNA duplex sample was assembled from chemically synthesized oligonucleotides (Metabion, Germany) and prepared as described previously.[40] Bovine serum albuminum (BSA), cytochrome *c*, and lysozyme were purchased from Sigma-Aldrich and used without further purification. Detailed information about the employed buffers and sample concentrations is listed in Table I. For concentration series, a stock solution of the highest concentration was prepared by weighing out the lyophilized protein powder and diluted to the required concentrations. Both buffer and sample solutions were filtered through 0.22 μm syringe filters (Thermo scientific, USA). Prior to the measurements, sample solutions were centrifuged at 13 500 rpm for 10 min in a tabletop centrifuge (Eppendorf, Germany). Sample and buffer solutions were degassed in a desiccator at a pressure level of 30 mbar for 20 min to avoid the formation of air bubbles in the sample chamber during measurements. For each measurement, 80 μl of sample or buffer solution was loaded into the sample chambers.

C. Measurement procedures

Prior to each experimental run, scattering curves of AgBe and glassy carbon were measured to determine the sample-detector distance with mm-accuracy and to calibrate the scattering curves on an absolute scattering scale. Sample chambers were placed in the sample holder and aligned such that the incoming X-ray beam penetrates the chamber at its center as follows: The sample stage is scanned vertically and horizontally in a range of 5 mm. At each position, a 1 s exposure is recorded with the beamstop removed and the intensity is integrated. The intensity stays approximately constant when the X-ray beam penetrates the observation volume and drops off rapidly when the beam is clipped by the sides of the sample chamber, allowing for an accurate determination of the center position.

Biological SAXS measurements were performed at room temperature and exposure times were set to 1–3 h with 3–6 repeats each, resulting in a total exposure time of up to 24 h for each measurement. Matching SAXS profiles of each repeat were used for data averaging as described in Sec. III D. For concentration series, we used the same chamber, which was rinsed with deionized water and buffer solution before filling it with fresh sample solution. Matching buffer profiles were collected using identical settings and procedures. For selected SAXS experiments, dynamic light scattering measurements on a NANO-flex® 180° instrument (Particle Metrics GmbH, Germany) were performed to test for possible aggregation. No aggregation was observed for any of the tested samples.

D. Data processing and evaluation

The two-dimensional detector images were processed with a macro including the command "remove outliers" of the software ImageJ (National Institutes of Health, USA) in order to remove artefacts, which appear as small bright spots of only a few pixels in the detector image, due to background radiation as, for instance, cosmic rays. The "remove outliers" algorithm replaces a pixel value by the median of adjacent pixel values if it deviates from the median by more than the threshold value. We used a radius of 7 for the pixel area to calculate the median and a threshold value of 50. By setting "which outlier" to "bright" only pixels that are brighter than the median of the surrounding are replaced. Next, we used the Igor Pro plugin NIKA[41] to reduce the 2D detector data into a one-dimensional scattering intensity. First, the sample-to-detector distance and the beam center were refined based on the AgBe scattering data. Then, circular averaging of both sample and buffer images was performed without any additional corrections. We further used custom-written MATLAB scripts to inspect the scattering data for aggregation or radiation damage, to perform data averaging, buffer subtraction, to define the usable q-range and for calibration of the data to exposure time, concentration, and absolute intensity. In addition, the MATLAB scripts performed a Guinier analysis to determine the radius of gyration (R_g) by iterative linear regression within the q-range of the data limited by $q \cdot R_g < 1.3$. Unless otherwise noted, the profiles shown represent averaged scattering data resulting from 3 repeats of 2 h exposures.

E. Theoretical scattering curves and *ab initio* low resolution reconstructions

For comparison of the experimental SAXS data, we calculated theoretical scattering profiles for our panel of scattering standards based on their atomic coordinates using the program *CRYSOL*[42] in default mode. The crystallographic structures of the protein samples were obtained from the protein data bank,[43] with PDB accession codes 4F5S for BSA, 1HRC for cytochrome *c* and 6LYZ for lysozyme. For the 24 bp DNA, a PDB file with the atomic coordinates was generated using the 3DNA package.[44]

We used the program DAMMIF[18] to generate *ab initio* three-dimensional models from the scattering data. DAMMIF represents the particle as an assembly of identical beads

inside a search volume. It employs a simulated annealing protocol to determine a compact interconnected model whose scattering pattern fits the experimental data. The particle distance distribution function $P(r)$ generated from the ATSAS software[45] was used as input file. For each tested molecule, 20 independent runs in the "slow" mode were performed using default parameters and assuming no symmetry. Next, we averaged the 20 models for each molecule using DAMAVER[46] comprising a sequence of programs: first, the low resolution models from DAMMIF were aligned based on their axes of inertia using a normalized spatial discrepancy (NSD) criterion.[47] The NSD value provides a quantitative measure of similarity between different models. A NSD value of zero corresponds to identical objects and values exceeding 1 refer to objects that systematically differ from one another. If pairwise NSD values are in the range between zero and one then the models are classified as structurally similar. The aligned bead models were averaged and filtered by removing loosely connected beads. For the next steps, the reconstructed file with the lowest NSD value was chosen. We used the pdb2vol program from the SITUS[48] package (version 2.7.2) to convert aligned bead models to electron density maps. Finally, we aligned the models to the corresponding crystal structures,[43] again by minimizing the NSD value between both structures. Molecular graphics were prepared using visual molecular dynamics.[49]

IV. RESULTS AND DISCUSSION

To explore the capabilities of our setup, we conducted a number of test measurements using a panel of biological macromolecules as measurement standards that comprise horse heart cytochrome *c*, chicken egg white lysozyme, BSA, and a 24 bp DNA construct (Table I). All selected macromolecules have known high-resolution structures and have been used as scattering standards previously.[14,50–52] They span a range of molecular weights but, in general, have relatively small sizes and, consequently, scatter weakly. Therefore, they are ideally suited to characterize our setup within the described q-range and constitute rigorous test cases for typical biological samples for SAXS measurement.

A. Exposure time and concentration analysis

We initially carried out a set of test measurements to determine concentration requirements, optimal exposure times, and possible radiation damage effects. We performed concentration- and exposure time series on the scattering standard samples listed in Table I (except for BSA). Fig. 2(a) shows SAXS profiles for three different concentrations (5 mg/ml, 10 mg/ml, and 20 mg/ml) of lysozyme, which has been previously characterized in synchrotron based SAXS experiments.[10,52] The concentration scaled data are superimposable and exhibit no evidence of radiation damage or interparticle effects such as aggregation or interparticle interference. Kratky plots, where the scattering intensity weighted by q^2 is plotted against q, are shown for all concentrations (Fig. 2(b)). The Kratky representation is frequently used to represent scattering data of macromolecular ensembles,[4] where a well-folded homogeneous particle will exhibit a parabolic curve and an unfolded particle will give rise to a hyperbolic curve. For all three concentrations of lysozyme, the scattering profiles display a pronounced peak indicating that the protein is well-

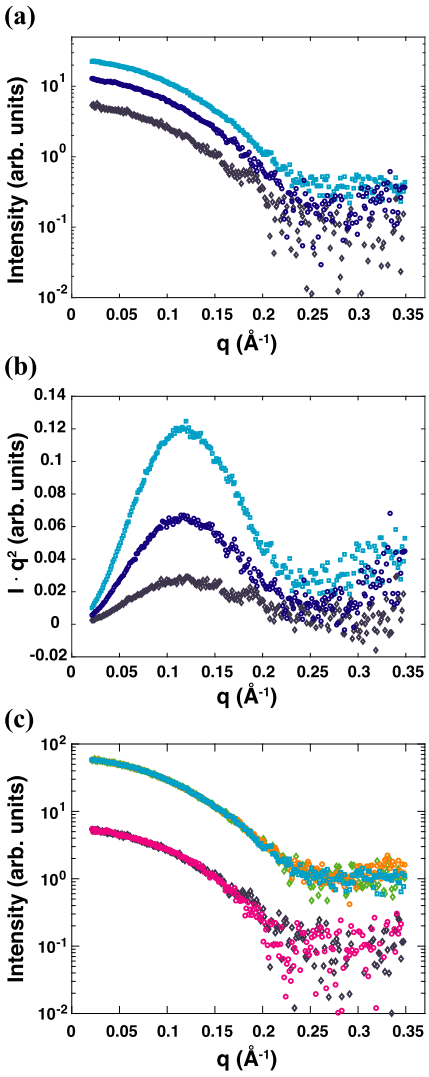


FIG. 2. Effects of protein concentration and exposure time on scattering profiles. Scattering profiles shown are for lysozyme (see Table I for details). (a) Averaged scattering data at concentrations of 20 mg/ml (cyan, squares), 10 mg/ml (blue, circles), and 5 mg/ml (grey, diamonds) for three repeats of 2 h. (b) Kratky plots ($q^2 \cdot I$ vs. q) for the data from panel (a). (c) Averaged scattering data of lysozyme at a concentration of 20 mg/ml for six repeats of 0.5 h (green, diamonds), three repeats of 1 h (orange, circles) and two repeats of 2 h (cyan, squares) and a concentration of 5 mg/ml with exposure times of 2 h for three repeats (grey, diamonds) and 3 h for two repeats (magenta, circles). Data are scaled by exposure time.

behaved in its folded state. Although the signal-to-noise ratio decreases for lower concentrations, we still obtain reasonable quality data at the lowest concentration of 5 mg/ml. In addition, we examined different exposure times and number of repeats for the highest and lowest concentrations used for our test molecules (Fig. 2(c)). We found that for lysozyme (c = 20 mg/ml) six exposures of 30 min each lead already to decent signal for q-values below 0.2 Å^{-1}. However, we observed an improvement in data quality when using three repeats of 2 h, resulting in a signal-to-noise ratio suitable for structural analysis as described in Section IV C. Even longer exposure times were tested for the lowest concentration but did not increase the data quality. This is probably due to a higher level of background noise attributed to background radiation, which is also integrated over time. Similar results were obtained for cytochrome c measured at concentrations of 2, 8, and 24 mg/ml (Fig. S4 of the supplementary material).[35]

B. Comparison of in-house data and synchrotron data

We compared the data obtained at our in-house source with data collected at the beamline BM29 at the ESRF in Grenoble (for proteins) and at the beamline 12-ID-B of the Advanced Photon Source (APS), Argonne, Illinois (for the 24 bp DNA), both third generation synchrotron light sources with instruments designed for biological SAXS measurements in solution (Fig. 3). ESRF data were collected in the "flow" mode at room temperature with an exposure time of 1 s. APS data were collected in a static sample cell at room temperature with an exposure time of 1 s. Matching data from ten runs were averaged. For q values <0.2 Å^{-1}, the synchrotron scattering profiles are closely approximated by those acquired on the in-house setup. However, for larger q values, the signal-to-noise ratio decreases faster for the in-house data, as one would expect.

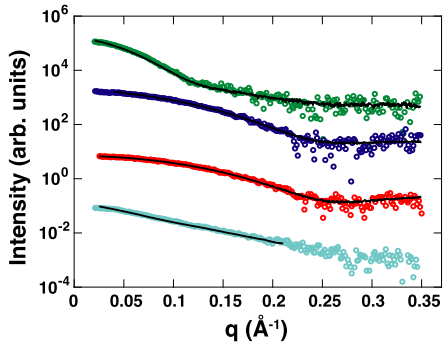


FIG. 3. Comparison of in-house source and synchrotron-based SAXS data. Scattering profiles acquired at our in-house source (shown in color) and measured at synchrotron sources (shown as black lines) for BSA (green, top), lysozyme (blue), cytochrome c (red), and 24 bp DNA (cyan, bottom) with concentrations from Table II. In-house data correspond to averaged data from three repeats with 2 h exposure time. Synchrotron data were averaged from 10 runs with 1 s exposure time. The synchrotron data for lysozyme had to be cut at a q-value of 0.04 Å^{-1} due to problems with the flow cell. DNA data were taken at another beamline with a maximum q-value of 0.21 Å^{-1}. Profiles are vertically offset for clarity.

C. Structural characterization and *ab initio* modeling of proteins and nucleic acids

The radius of gyration (R_g) and forward scattering intensity at zero angle ($I(0)$) are two parameters routinely extracted from SAXS data by Guinier analysis, where a straight line is fitted to the logarithm of the scattering intensity plotted as a function of q^2 for the lowest scattering angles. The R_g gives an overall measure for the size of the molecule; $I(0)$ is used to calculate the molecular weight (MW) and to evaluate sample monodispersity.[1,6] We performed Guinier analyses of the scattering profiles for every concentration and exposure time (see Fig. 4 for examples). For the molecular weight determination we employed BSA as reference sample. The Guinier plots for all of our test samples exhibit good linearity (Fig. 4) and the forward scattering intensities scale linearly with sample concentration, indicating the absence of interparticle interference effects or aggregation. We obtained radii of gyration from our experimental data that are consistent with literature values reported for the native state of each tested macromolecule (Table II). Molecular weight estimates from the forward scattering (Table II) are in good agreement (within experimental error) with the molecular weights expected from the primary structure of the monomeric samples. The error of the molecular weight determination in Table II is dominated by uncertainties in the macromolecular concentrations of approximately 10% relative error.

For further comparison, we calculated theoretical scattering profiles from the crystal structures (Figs. 5(a)-5(d)) for each tested molecule and determined the radius of gyration from the predicted scattering profiles based on the crystal structures (Table II). The experimental data are in excellent agreement with the theoretical scattering profiles and the overall scattering features of each molecule are observable. The resulting chi-squared values (χ^2), which characterize the "goodness-of-fit," are all around 0.1. Moreover, the

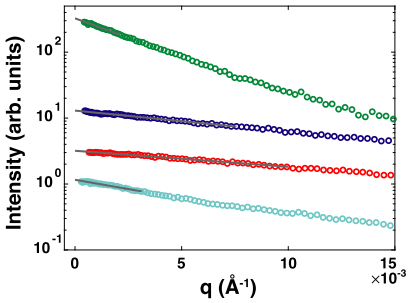


FIG. 4. Guinier analysis of biological samples. Guinier representation of the experimental scattering data for BSA (green, top), lysozyme (blue), cytochrome c (red) and 24 bp DNA (cyan, bottom). The Guinier fits are indicated by grey lines covering a q-range of $qR_g < 1.3$. Profiles are vertically offset for clarity.

TABLE II. Radii of gyration and molecular weights for protein and nucleic acid samples used in this study.

Sample	PDB used	Concentration (mg/ml)	R_g (Å)[a]	R_g (Å)[b]	R_g (Å)[c]	R_g (Å)[d]	MW[a] (kDa)	MW[b] (kDa)
BSA	4F5S	5	29.0 (±0.8)	29.9 (±0.8)[52]	27.3	28.1	...	69.0
Cytochrome *c*	1HRC	8	13.4 (±0.1)	13.8 (±0.3)[53]	12.6	13.2	11.5 (±1.1)	12.4
Lysozyme	6LYZ	10	14.6 (±0.4)	14.3 (±0.4)[52]	14.6	14.7	14.8 (±1.3)	14.3
24 bp DNA	...	1.5	23.3 (±2.7)	24.2 (±0.5)[14]	24.8	21.8	15.5 (±2.5)	14.6

[a]Values for radii of gyration and molecular weights determined in this study.
[b]Taken from experimental data reported in the literature.
[c]Calculated from the theoretical scattering profiles.
[d]From the reconstruction fit files.

theoretically predicted R_g values are in the range of the experimental and the literature values for all test samples.

Over the last two decades, the development of algorithms for *ab initio* reconstructions of low resolution three-dimensional electron density maps from one-dimensional scattering profiles has significantly enhanced the capabilities of SAXS measurements.[11–13,54] In addition, *ab initio* reconstructions can be combined with atomistic structures derived by NMR or X-ray crystallography or other sources of structural information to enhance or validate models for both proteins and nucleic acids.[14,55,56] To determine whether the data collected at our Mo-based in-house source are of sufficient quality to obtain 3D structure reconstructions of typical biological macromolecules, we performed *ab initio* modeling for all macromolecules of our test panel using the software DAMMIF (see Sec. III). The reconstructions converged to solutions that fit the experimental scattering profiles very well (Figs. 5(a)-5(d)). The scattering profiles from the models are in very good agreement with the experimental data over the whole q-range with χ^2 values below 0.1. However, they slightly deviate from the theoretical scattering profiles of the crystal structures for q-values above 0.25 Å^{-1}. For all reconstructions, the pairwise NSD values for independent reconstruction runs never exceeded 1, indicating that the reconstruction algorithm is stable and converges onto similar structures in each run. The final *ab initio* generated models were compared and aligned to corresponding crystal structures. Figs. 5(e)-5(h) show the bead models of each molecule rendered as smooth transparent surfaces and the superimposed crystal structures as black ribbons (proteins) and stick (24 bp DNA duplex) representations. The overall shapes and sizes of the proteins were reproduced well. For BSA (Fig. 5(e)), the reconstructed density fits nicely to the triangular-like shape of the protein. The surface of the density map is rough with several small indentations reproducing the high amount of alpha-helices present in native BSA. For lysozyme and cytochrome *c* (Figs. 5(f) and 5(g)), we obtained reconstructions representing their globular shape, which are in good agreement with the protein sizes. The reconstructed density of the 24 bp DNA duplex (Fig. 5(h)) corresponds reasonably to the overall cylindrical shape of a duplex. The length of the duplex gets reproduced well, whereas small deviations for the diameter of the reconstructions are observable. However, the periodicity of the major and minor grooves is visible in the reconstruction.

D. Determining the shape, size, and interactions of detergent micelles

Micelles are aggregates of amphiphilic molecules in aqueous solution where the hydrophilic head groups face

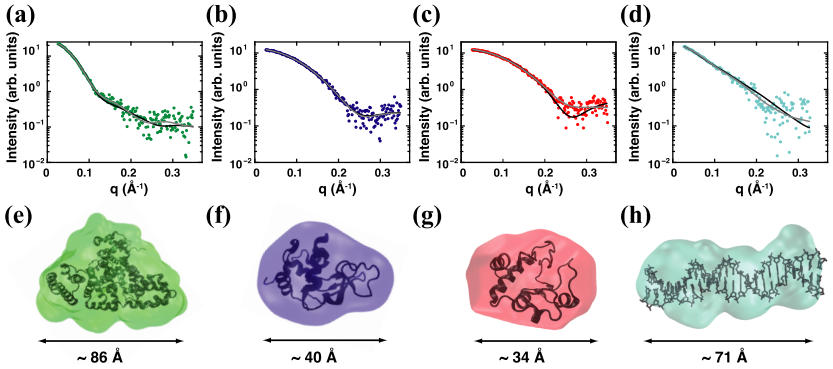


FIG. 5. Comparison of crystal structures and *ab initio* 3D shape reconstructions for protein and DNA samples. (a)-(d) Comparison of experimental (colors; same color code as in Figs. 3 and 4) and theoretical scattering profiles that were predicted from the crystal structures (black lines) and fitted scattering profiles from *ab initio* 3D reconstructions (grey lines). (e)-(h) Models obtained from *ab initio* 3D structure reconstructions for BSA (green), lysozyme (blue), cytochrome *c* (red), and 24 bp DNA (cyan). The maximum dimension D_{max} of each molecule is indicated below each molecule and was derived by calculating the pair distance distribution function $P(r)$ from the experimental scattering profiles.

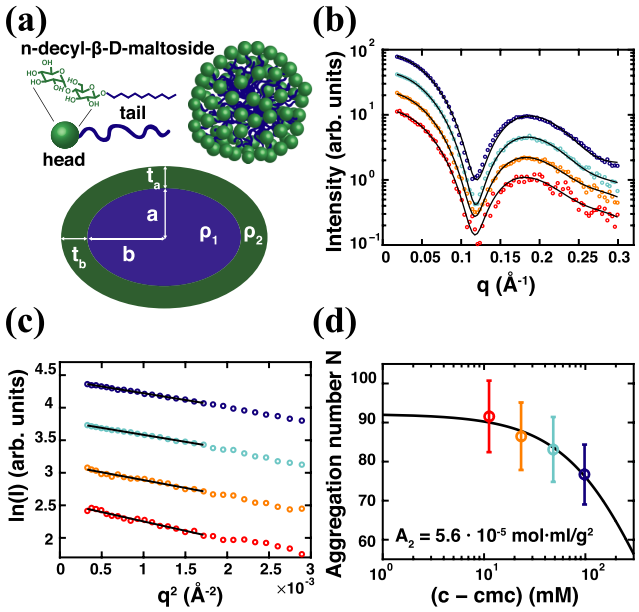


FIG. 6. Characterization of the size, shape, and interactions of DM micelles. (a) Chemical structure of n-decyl-β-D-maltoside forming micelles and the schematic of the two-component ellipsoid model. a and b are the dimensions and ρ_1 the electron density of the hydrophobic core. t_a and t_b are the thicknesses and ρ_2 the electron density of the head group region. The figure shows the case of an oblate ellipsoid with a <b. (b) Experimental data for different DM concentrations of 100 mM (blue, top), 50 mM (cyan), 25 mM (orange), and 12.5 mM (red, bottom) and the corresponding fits (black lines). (c) Guinier analysis for DM data shown in (b). (d) Apparent aggregation numbers N obtained from the extrapolated forward scattering intensity and Eq. S2 in the supplementary material[35] (circles, same color code as in (b)). The solid line is a fit to the model of Eq. S6 and A_2 is the fitted second virial coefficient determined from the fit, indicative of weak repulsive interactions between the DM micelles in solution at higher concentrations.

outward and hydrophilic tail groups are segregated in the interior (Fig. 6(a)). Micelle forming detergents are employed in a large range of biochemical and industrial applications. In particular, detergents are commonly used as mimetics of the cell membrane for the solubilization and structural characterization of membrane proteins.[57,58] However, the choice of a suitable detergent for membrane protein solubilization still remains a major hurdle.[59–61] SAXS has been used extensively to characterize the size and shapes of both membrane protein-detergent complexes-formed by a membrane protein surrounded by detergents-and of "empty" micelles.[62–66] To test to what extent our in-house source is capable of revealing the shape and size of detergent micelles, we recorded scattering profiles at different concentrations of n-decyl-β-D-maltoside (DM), a detergent featuring a maltose head group and a ten carbon single-chain alkyl tail (Fig. 6(a)), which is routinely used for membrane protein solubilization and has been characterized by SAXS in several previous studies.[65,67] We obtain decent signal-to-noise down to a concentration of 12.5 mM DM (Fig. 6(b)). The scattering profiles are well described by a two-component ellipsoid model (Figs. 6(b) and 6(d)), which features a core corresponding to the hydrophobic portion of the micelles formed by the tail groups and of a shell corresponding to the hydrated head groups (Fig. 6(a); see supplementary material for details of the model).[35] The size parameters obtained from the fits of the two-component ellipsoid model reveal oblate micelles with the short axis of the core of ~12.8 Å and the long axes of ~22.2 Å and a thickness of the hydrophilic shell of ~7 Å, in excellent agreement with previous work.[65,67] In addition, we performed a Guinier analysis of the data (Fig. 6(c)) and determined apparent aggregation numbers (i.e., the number of detergent monomers per micelle) from the fitted forward scattering intensities by comparison with a scattering standard as described by Lipfert *et al.* (see supplementary material for details).[35,65] For the measured concentrations, we find radii of gyration in the range of 25.5 Å (±0.4 Å) and aggregation numbers from the forward scattering intensity in the range of ~90 for the lowest concentration (Fig. 6(d)), in excellent agreement with the number calculated from the size of the hydrophic core volume of ~26 nm^3 determined from the two-component ellipsoid fit and with previous measurements, which indicate aggregation numbers in the range of 85–95 monomers per micelle.[65,67] The apparent aggregation numbers show a small, but systematic decrease with increasing detergent concentration (Fig. 6(d)). This decrease in apparent aggregation number could be indicative of DM micelles shrinking with increasing detergent concentrations, which is however unlikely, or due to inter-particle interference effects. The latter results, in particular, from repulsive interactions of the micelles in solution, e.g., due to excluded volume effects, which become more relevant at higher concentrations. Similar decreases in the apparent

aggregation number with increasing detergent concentrations had been seen for a range of uncharged and, in particular, charged detergents previously.[65] Here, we present a new model that describes the apparent aggregation number as a function of detergent concentration in terms of intrinsic, true aggregation number and of the second virial coefficient, a parameter that characterizes the interparticle interactions in solution (see the text and Fig. S5 of the supplementary material).[35] Our DM data are well described by the model (Fig. 6(d), solid line), with a fitted intrinsic aggregation number of $N(c_0) = 92$ and a fitted second virial coefficient of $A_2 = 5.6 \times 10^{-5}$ mol ml/g^2, which indicates weak repulsive interactions. In summary, the DM data suggest that our in-house setup is fully able to reveal the size, shape, and overall interactions of typical detergent micelles.

V. CONCLUSION

We have presented a Mo-anode-based in-house SAXS setup for structural analysis of macromolecules covering a broad range of sizes, shapes, compositions (proteins/DNA/micelles), and scattering properties. Our system contains a Mo-based microfocus X-ray tube with an integrated multilayer mirror delivering a stable monochromatic beam. By using two scatterless slits for collimation, a highly collimated X-ray beam of low beam divergence is generated. The typical flux at the sample stage is around 2.5×10^6 photons/s. Due to the reduced air scattering for Mo-radiation, the sample chambers do not have to be placed in vacuum. Our sample holder contains two sample chambers with observation volumes of 80 μl allowing subsequent automated SAXS measurements of sample and buffer. In addition, the sample chambers can be temperature controlled within a temperature range of 4–70 °C (±0.8 °C). By using the hybrid pixel detector PILATUS 100K, weakly scattering signals can be detected. Our system allows us to perform SAXS measurements on a broad range of weakly scattering biological macromolecules at concentrations comparable to synchrotron based SAXS measurements within 2 h. The achievable scattering vectors for SAXS measurements cover a range of 0.009–0.38 Å^{-1}, such that macromolecules with a size of up to ~30 nm can be structurally characterized. By performing in-house SAXS measurements on a test set of molecules including several proteins and DNA, we demonstrate that the data are of adequate quality to determine *ab initio* low resolution 3D structures of the macromolecules, which were in very good agreement with previously reported structures. Our scattering data were also consistent with theoretical data calculated from the atomic structures of our test molecules. In addition, we demonstrate the instrument's ability to obtain high quality data for detergent micelles commonly used in membrane protein studies and we describe a novel simple model that enables us to determine the micelle aggregation number and second virial coefficient from SAXS data at different detergent concentrations. In general, the significantly lower flux (at least five orders of magnitude) of current Mo-based in-house sources compared to synchrotron sources necessitates much longer integration times (hours compared to seconds, respectively) and limits the practically achievable signal-to-noise ratio, in particular, at larger q values. However, these disadvantages are partially offset by the much greater availability and reduced measurement logistics of an in-house instrument. In summary, our results suggest that Mo-anode-based in-house SAXS experiments are a viable alternative to other anode materials and allow studying many aspects of weakly scattering biological samples.

ACKNOWLEDGMENTS

We thank Dr. Adam Round for support at beamline BM 29 at the ESRF in Grenoble, Dr. Jan Ilavsky (APS) for the calibrated glassy carbon specimen, Dr. Sönke Seifert for support with measurements at beamline 12ID at Argonne National Lab, Dr. Linda Columbus for discussions and for providing detergent samples, and Julian Nguyen for help with the fitting of the micelle data. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences. The authors acknowledge the Center for Nanoscience (CeNS), DFG-SFB 1032 (Nanoagents) and the Nanosystems Initiative Munich (NIM) for financial support.

[1]J. Lipfert and S. Doniach, Annu. Rev. Biophys. Biomol. Struct. **36**, 307 (2007).
[2]M. H. J. Koch, P. Vachette, and D. I. Svergun, Q. Rev. Biophys. **36**, 147 (2003).
[3]C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, Q. Rev. Biophys. **40**, 191 (2007).
[4]S. Doniach, Chem. Rev. **101**, 1763 (2001).
[5]A. Guinier, Ann., Phys. (Paris) **12**, 161 (1939).
[6]D. I. Svergun and M. H. J. Koch, Rep. Prog. Phys. **66**, 1735 (2003).
[7]D. I. Svergun, J. Appl. Crystallogr. **25**, 495 (1992).
[8]D. Herschlag, R. Russell, I. S. Millett, and S. Doniach, Nat. Struct. Biol. **7**, 367 (2000).
[9]A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, Phys. Rev. E **86**, 021901 (2012).
[10]P. Bernado, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun, J. Am. Chem. Soc. **129**, 5656 (2007).
[11]D. I. Svergun, Biophys. J. **76**, 2879 (1999).
[12]D. I. Svergun, M. V. Petoukhov, and M. H. J. Koch, Biophys. J. **80**, 2946 (2001).
[13]D. Walther, F. E. Cohen, and S. Doniach, J. Appl. Crystallogr. **33**, 350 (2000).
[14]J. Lipfert, V. B. Chu, Y. Bai, D. Herschlag, and S. Doniach, J. Appl. Crystallogr. **40**, s229 (2007).
[15]J. Lipfert, J. Ouellet, D. G. Norman, S. Doniach, and D. M. J. Lilley, Structure **16**, 1357 (2008).
[16]A. Grishaev, J. Wu, J. Trewhella, and A. Bax, J. Am. Chem. Soc. **127**, 16621 (2005).
[17]J. S. Pedersen, J. Appl. Crystallogr. **37**, 369 (2004).
[18]D. Franke and D. I. Svergun, J. Appl. Crystallogr. **42**, 342 (2009).
[19]G. B. Mortuza, T. Cavazza, M. F. Garcia-Mayoral, D. Hermida, I. Peset, J. G. Pedrero, N. Merino, F. J. Blanco, J. Lyngsø, M. Bruix, J. S. Pedersen, I. Vernos, and G. Montoya, Nat. Commun. **5**, 5072 (2014).
[20]D. M. Dupont, C. K. Thuesen, K. A. Bøtkjær, M. A. Behrens, K. Dam, H. P. Sørensen, J. S. Pedersen, M. Ploug, J. K. Jensen, and P. A. Andreasen, PLoS One **10**, e0119207 (2015).
[21]W. Vogel, Surf. Sci. **156**, 420 (1985).
[22]T. Takamuku, A. Yamaguchi, D. Matsuo, M. Tabata, T. Yamaguchi, T. Otomo, and T. Adachi, J. Phys. Chem. B **105**, 10101 (2001).
[23]M. Matsugami, T. Takamuku, T. Otomo, and T. Yamaguchi, J. Phys. Chem. B **110**, 12372 (2006).
[24]N. Metoki, W. Donner, and H. Zabel, Phys. Rev. B **49**, 17351 (1994).
[25]M. Cianci, J. R. Helliwell, D. Moorcroft, A. Olczak, J. Raftery, and P. J. Rizkallah, J. Appl. Crystallogr. **37**, 555–564 (2004).
[26]T. Mühge, T. Zeidler, Q. Wang, C. Morawe, N. Metoki, and H. Zabel, J. Appl. Phys. **77**, 1055 (1995).

[27]S. Sato, T. Nishi, E. Matsubara, M. Imafuku, N. Nishiyama, Y. Waseda, and A. Inoue, Sci. Technol. Adv. Mater. **3**, 69 (2008).
[28]S. Kuwamoto, S. Akiyama, and T. Fujisawa, J. Synchrotron Radiat. **11**, 462 (2004).
[29]A. Rösler, H.-A. Klok, I. W. Hamley, V. Castelletto, and O. O. Mykhaylyk, Biomacromolecules **4**, 859 (2003).
[30]A. Martel, M. Burghammer, R. J. Davies, E. Di Cola, C. Vendrely, and C. Riekel, J. Am. Chem. Soc. **130**, 17070 (2008).
[31]J. A. Sichert, Y. Tong, N. Mutz, M. Vollmer, S. Fischer, K. Z. Milowska, R. García Cortadella, B. Nickel, C. Cardenas-Daw, J. K. Stolarczyk, A. S. Urban, and J. Feldmann, Nano Lett. **15**, 6521 (2015).
[32]Y. Li, R. Beck, T. Huang, M. C. Choi, and M. Divinagracia, J. Appl. Crystallogr. **41**, 1134 (2008).
[33]A. Bergmann, D. Orthaber, G. Scherf, and O. Glatter, J. Appl. Crystallogr. **33**, 869 (2000).
[34]J. Lipfert, I. S. Millett, S. Seifert, and S. Doniach, Rev. Sci. Instrum. **77**, 046108 (2006).
[35]See supplementary material at http://dx.doi.org/10.1063/1.4940936 for further details on scattering background analysis and detergent micelle characterization.
[36]B. L. Henke, E. M. Gullikson, and J. C. Davis, At. Data Nucl. Data Tables **54**, 181 (1993).
[37]P. Kraft, A. Bergamaschi, C. Broennimann, R. Dinapoli, E. F. Eikenberry, B. Henrich, I. Johnson, A. Mozzanica, C. M. Schlepuetz, P. R. Willmott, and B. Schmitt, J. Synchrotron Radiat. **16**, 368 (2009).
[38]T. N. Blanton, T. C. Huang, H. Toraya, C. R. Hubbard, S. B. Robie, D. Louër, H. E. Göbel, G. Will, R. Gilles, and T. Raftery, Powder Diffr. **10**, 91 (1995).
[39]F. Zhang, J. Ilavsky, G. G. Long, J. P. G. Quintana, A. J. Allen, and P. R. Jemian, Metall. Mater. Trans. A **41**, 1151 (2009).
[40]Y. Bai, R. Das, I. S. Millett, D. Herschlag, and S. Doniach, Proc. Natl. Acad. Sci. U. S. A. **102**, 1035 (2005).
[41]J. Ilavsky, J. Appl. Crystallogr. **45**, 324 (2012).
[42]D. Svergun, C. Barberato, and M. Koch, J. Appl. Crystallogr. **28**, 768 (1995).
[43]H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[44]X. J. Lu and W. K. Olson, Nucleic Acids Res. **31**, 5108 (2003).
[45]M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, and D. I. Svergun, J. Appl. Crystallogr. **45**, 342 (2012).
[46]V. V. Volkov and D. I. Svergun, J. Appl. Crystallogr. **36**, 860 (2003).
[47]M. B. Kozin and D. I. Svergun, J. Appl. Crystallogr. **34**, 33 (2001).
[48]W. Wriggers, R. A. Milligan, and J. A. McCammon, J. Struct. Biol. **125**, 185 (1999).
[49]W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[50]S. Akiyama, S. Takahashi, T. Kimura, K. Ishimori, I. Morishima, Y. Nishikawa, and T. Fujisawa, Proc. Natl. Acad. Sci. U. S. A. **99**, 1329 (2002).
[51]D. J. Segel, A. Bachmann, J. Hofrichter, K. O. Hodgson, S. Doniach, and T. Kiefhaber, J. Mol. Biol. **288**, 489 (1999).
[52]E. Mylonas and D. I. Svergun, J. Appl. Crystallogr. **40**, 245 (2007).
[53]D. J. Segel, A. L. Fink, K. O. Hodgson, and S. Doniach, Biochemistry **37**, 12443 (1998).
[54]P. Chacón, J. F. Díaz, F. Morán, and J. M. Andreu, J. Mol. Biol. **299**, 1289 (2000).
[55]J. J. P. Perry and J. A. Tainer, Methods **59**, 363 (2013).
[56]M. Hammel, Y. Yu, B. L. Mahaney, B. Cai, R. Ye, B. M. Phipps, R. P. Rambo, G. L. Hura, M. Pelikan, S. So, R. M. Abolfath, D. J. Chen, S. P. Lees-Miller, and J. A. Tainer, J. Biol. Chem. **285**, 1414 (2010).
[57]M.le. Maire, P. Champeil, and J. V. Moller, Biochim. Biophys. Acta **1508**, 86 (2000).
[58]S. Hiller, R. G. Garces, T. J. Malia, V. Y. Orekhov, M. Colombini, and G. Wagner, Science **321**, 1206 (2008).
[59]A. M. Seddon, P. Curnow, and P. J. Booth, Biochim. Biophys. Acta **1666**, 105 (2004).
[60]L. Columbus, J. Lipfert, K. Jambunathan, D. A. Fox, A. Y. L. Sim, S. Doniach, and S. A. Lesley, J. Am. Chem. Soc. **131**, 7320 (2009).
[61]J. Bowie, Curr. Opin. Struct. Biol. **11**, 397 (2001).
[62]S. H. Chen, Annu. Rev. Phys. Chem. **37**, 351 (1986).
[63]R. T. Zhang, P. A. Marone, P. Thiyagarajan, and D. M. Tiede, Langmuir **15**, 7510 (1999).
[64]C. Dupuy, X. Auvray, and C. Petipas, Langmuir **13**, 3965 (1997).
[65]J. Lipfert, L. Columbus, V. B. Chu, S. A. Lesley, and S. Doniach, J. Phys. Chem. B **111**, 12427 (2007).
[66]L. Columbus, J. Lipfert, H. Klock, I. Millett, S. Doniach, and S. A. Lesley, Protein Sci. **15**, 961 (2006).
[67]R. C. Oliver, J. Lipfert, D. A. Fox, R. H. Lo, S. Doniach, and L. Columbus, PLoS One **8**, e62488 (2013).

Supplementary Information for "A Mo-Anode-Based In-House Source for Small-Angle X-Ray Scattering Measurements of Biological Macromolecules"

Linda K. Bruetzel[1], Stefan Fischer[1], Annalena Salditt[1], Steffen Sedlak[1], Bert Nickel[1], and Jan Lipfert[1,a)]

[1]*Department of Physics, Nanosystems Initiative Munich, and Center for Nanoscience, Ludwig-Maximilians-University Munich, Amalienstr. 54 and Geschwister-Scholl Platz 1, 80799 and 80539 Munich, Germany*

a) Author to whom correspondence should be addressed. Electronic mail: Jan.Lipfert@lmu.de

Supplementary Information consisting of

- **Supplementary Text**
- **Supplementary Figures S1-S5**

Analysis of contributions to the background signal in Molybdenum-anode-based SAXS experiments

When performing SAXS experiments the reduction of the overall background, mainly caused by detector noise, natural background radiation and parasitic scattering, is of major concern. The latter originates from apertures, window materials and air in the beam path. For our in-house setup the main contribution of parasitic scattering can be related to windows and air scattering, as the collimation of the beam employs scatterless slits. The correct choice of window material and the evacuation of the flight path can reduce the amount of parasitic scattering. In order to estimate the overall background, we performed SAXS measurements using mica and Kapton as window materials commonly used for X-ray scattering experiments (both with window thicknesses of 25 μm), under conditions when the flight path between the sample and detector was either fully evacuated or in air. To mimic the regular measurement conditions of a SAXS experiment, we used deionized water as a sample solution. Moreover, we measured the natural background when the X-ray tube is turned off. For all measurements exposure times were set to 2h with three repeats and circular integrated data were averaged. The intensity was transformed to counts per second. The natural background level recorded by the detector is ~ 10^{-4} counts/s (Supplementary Fig. S2a), which is well (by at least a factor of three) below the intensity level obtained from SAXS measurements on the water sample, suggesting that despite the increased sensor thickness of our detector (~ three times higher than the sensor thickness of the Pilatus 100 K detector (i.e. 320 μm) commonly used for Cu-anode based SAXS measurements), the natural background radiation recorded by the detector do not limit our measurements. Comparing the results obtained for SAXS measurements on the water sample with the flight path (95 cm) either in air or fully evacuated (Supplementary Fig. S2a), we find significant deviations of the scattering curvature for the non-evacuated measurement, in the low q-region. The q-dependence of the air scattering is influenced by two effects. If the flight path is not evacuated, air molecules along the entire flight path will

contribute to the scattering signal. However, scattering by molecules close to the beamstop will only contribute to the signal at high q, since their low q contribution is blocked by the beamstop (Supplementary Fig. S2b). This shadowing effect of the beamstop leads to the decrease of the scattering intensity in the low q-region for the air measurements (Supplementary Fig. S2). The mean background levels for the higher q-range are in the range of ~ 3.3 x 10^{-3} and ~ 4.5 x 10^{-4} counts/s for the air measurements and the evacuated experiments, respectively. The difference between both datasets is ~ 2.9 x 10^{-3} counts/s, implying that per cm of non-evacuated distance the background due to parasitic scattering is increased by ~ 3 x 10^{-5} counts/s. This relatively low level of air scattering suggests that a few cm of non-evacuated flight path result in a background level in the order of the natural background. Therefore, it is for instance not necessary to evacuate the sample environment, as it is done for some Cu-anode-based SAXS setups. Additionally, we find no significant change in parasitic scattering for the different window materials.

Determination of detergent micelle size, shape, and interactions from SAXS measurement

Two-component ellipsoid model for detergent micelles

In the absence of significant interparticle interference effects (as is the case for our lowest detergent concentrations), the scattering intensity profile from monodisperse micelles can be well approximated by the scattering form factor of a two-component ellipsoid model (Fig. 6a) that features a core with electron density ρ_1, symmetry axis of length a and orthogonal axes of length b and a shell of electron density ρ_2 and thicknesses along the a and b dimensions of t_a and t_b, respectively.[1] For micelles, the electron density of the core, which corresponds to the region occupied by the hydrophobic tail groups, is typically less than the electron density of the solvent ρ_s, *i.e.* $\rho_1 < \rho_s$. In contrast, the electron density of the outer shell, which corresponds to the hydrated head groups, is usually larger than that of the solvent, such that $\rho_2 > \rho_s$. For $a < b$ the micelle is oblate and for $a > b$ it is prolate.

The form factor of the two component ellipsoid model is given by:

$$P(q) = \int_0^1 \left(3V_1(\rho_1 - \rho_2)\frac{j_1(u_1)}{u_1} + 3(V_1 + V_2)(\rho_2 - \rho_s)\frac{j_1(u_2)}{u_2}\right)^2 dx \qquad \text{(S1)}$$

with $u_1 = q\cdot(a^2x^2 + b^2(1 - x^2))^{1/2}$, $u_2 = q\cdot((a+t_a)^2x^2 + (b+t_b)^2(1 - x^2))^{1/2}$, the core volume $V_1 = 4\pi ab^2/3$, the total volume $V_1 + V_2 = 4\pi(a+t_a)(b+t_b)^2/3$, and j_1 being the first order spherical Bessel function. We fitted Eqn. S1 to the experimental scattering data (Fig. 6b) using custom written Matlab routines as described in Lipfert et al.[1] In the fits, we held the solvent density and the density of the hydrophobic core fixed at ρ_s = 0.34 e/Å^3, the approximate electron density of water with 150 mM NaCl added at room temperature, and ρ_1 = 0.273 e/Å^3, the electron density of the hydrophobic core computed from the Tanford volume of the hydrocarbon chain.[1] In addition, we assumed equal thicknesses of the outer shell in all dimensions, i.e. kept $t_a = t_b$ in the fits. Consequently, the free fitting parameters were a, b, t_a, ρ_2, as well as an overall scaling constant and a constant offset. The fits yielded values for the density of the outer shell of $\rho_2 \approx 0.45$ e/Å^3, which is slightly less then the theoretical value from the chemical composition and the density of the head group alone (=0.52 e/Å^3),[1] as would be expected, since ρ_2 represents the average electron density of the hydrated head group layer. The fitted values for a, b, and t_a are given in the main text.

Determination of micelle aggregation numbers from the forward scattering intensity

The forward scattering intensity determined from Guinier analysis (Fig. 6c) can be related to a scattering standard and to the expected intensity from a monomer to determine the (apparent) aggregation numbers of micelles, i.e. the number of detergent monomers in a micelle: [1,2]

$$N = \frac{I(0)_{det}}{I(0)_{mon}} = \frac{I(0)_{det}}{Kc\,(\rho_{det} - \rho_s)^2 V_{mon}^2} \qquad \text{(S2)}$$

$I(0)_{det}$ is the experimentally determined forward scattering intensity obtained from Guinier analysis of the data (Fig. 6c), c the detergent monomer concentration corrected for the critical micelle concentration $c = c_{mon} - cmc$ (cmc = 1.8 mM for DM[1]), V_{mon} and ρ_{det} are the molecular volume of a detergent monomer and its average electron density, both computed from the published specific densities as described.[1] K is a proportionality constant determined from measurements of our protein molecular weight standards of known concentration,

electron density, and molecular mass (Table I). Applying Eqn. S2, we find the apparent aggregation numbers *N* for DM as a function of detergent concentration (Fig. 6d, symbols).

Determination of true micelle aggregation numbers and second virial coefficients from the concentration dependence of the apparent aggregation numbers

In the *N* vs. concentration data, we observe a change in the apparent aggregation number with increasing detergent concentration (see Fig. 6d and similar observations for a range of detergents in Refs.[1,2]). *A priori,* there are two possible reasons for a change of apparent aggregation number with detergent concentration: i) the actual aggregation number might change with increasing detergent concentration or ii) interparticle interference effects might influence the scattering profiles at higher concentrations, in particular at low *q*, which in turn would influence the extrapolated forward scattering intensity and thus the measured apparent aggregation number. In cases where the apparent aggregation number decreases with increasing concentration, as is the case for DM (Fig. 6d), the second explanation is much more likely, since it is unlikely that micelles would shrink with increasing detergent concentration and it is plausible that micelles repel in solution, via excluded volume and (for charged detergent head group) electrostatic interactions.

Here, we present an analysis framework to fit apparent aggregation number vs. detergent concentration data to determine the aggregation number in the absence of interparticle interference effects and the second virial coefficient due to micelle-micelle interactions. The second virial coefficient is the first term in an expansion describing deviations from non-interacting "ideal gas-like" particles. Positive values of the second virial coefficient correspond to repulsive interactions between the particles and negative values are characteristic of attractive interactions. Generally, the solution structure factor is the change of the concentration-normalized scattering intensity compared to the scattering intensity obtained at a concentration c_0 ("infinite dilution") at which interparticle interference is negligible (see e.g. Equation 31 of Ref. [3]):

$$S(q,c) = (c_0 \cdot I(q,c)) / (c \cdot I(q,c_0)) \qquad \text{(S3)}$$

The solution structure factor at $q = 0$ is related to the second virial coefficient A_2 (see e.g. Equation 35 of Ref. [3]):

$$1/S(q=0,c) = 1 + 2 \cdot M \cdot A_2 \cdot c \qquad \text{(S4)}$$

to first order in the concentration c, where M is the molecular mass of the of the solute. Combining Equations (S3) and (S4), the change in concentration normalized forward scattering intensity (again to first order in c) is given by:

$$I(0,c) = (c/c_0) \cdot I(q,c_0) / (1 + 2 \cdot M \cdot A_2 \cdot c) \qquad \text{(S5)}$$

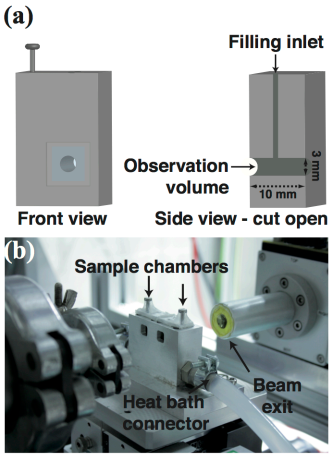
For the apparent aggregation numbers determined from the forward scattering intensity (see Equation 4 of Ref. [1]) this implies:

$$N(c) = N(c_0) / (1 + 2 \cdot M \cdot A_2 \cdot c) \qquad \text{(S6)}$$

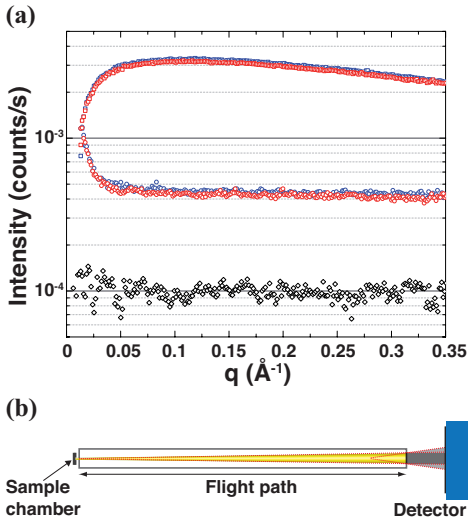
where $N(c)$ is the apparent aggregation number determined at concentration c and $N(c_0)$ is the "true" aggregation number determined in the absence of interparticle effects. We again take the concentration c as the monomer concentration corrected for the critical micelle concentration *cmc*, i.e. $c = c_{mon} - cmc$. If we take the concentration in g/ml and the molecular weight in Daltons (g/mol), A_2 has units of mol·ml/g^2. For proteins in solution, A_2 tends to have a magnitude in the range of 10^{-3} to 10^{-5} mol·ml/g^2 and can have a positive or negative sign, depending on solution conditions.[4] Values of A_2 of approximately $-5 \cdot 10^{-4}$ mol·ml/g^2 are characteristic of the so-called "crystallization slot", typical of solution conditions that promote crystal formation.[5] Eqn. S6 can be directly fit to experimental data of aggregation number vs. concentration, treating $N(c_0)$ and A_2 as free parameters. The model of Eqn. S6 provides an excellent description of our DM data (Fig. 6d, solid line) with $N(c_0) = 92$ and $A_2 = 5.6 \cdot 10^{-5}$ mol·ml/g^2.

In order to test the general applicability of the model derived above, we analyzed aggregation number vs. concentration data published previously by Lipfert et al.[1] for all detergents for which a decrease of apparent aggregation number with increasing concentration was observed (Supplementary Fig. S5). The model of Eqn. S6 provides an excellent description of the observed behaviors. In most cases, the fitted values for $N(c_0)$ are in close agreement with the previously reported values that were simply based on the lowest measured concentrations. In some cases, the fitted numbers for $N(c_0)$ actually match better with the

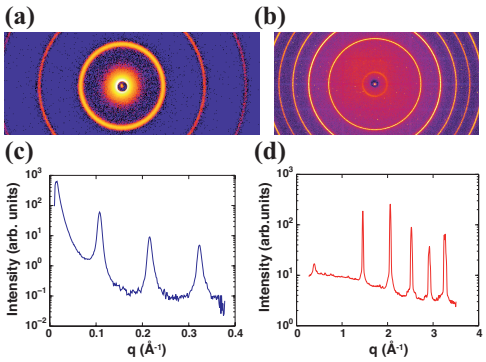
values determined from the two-component ellipsoid models than the previously determined numbers from the forward scattering, in particular for the detergents *n*-decylphosphocholine (FC-10) and *n*-dodecylphosphocholine (FC-12). The fitted values of the second virial coefficient are well within the range of values determined for proteins in solution under conditions that are not conducive for crystallization. It is noticeable that the second virial coefficient for *n*-decyl-β-D-maltoside (DM) and *n*-dodecyl-β-D-maltoside (DDM) with their non-ionic maltose head groups are significantly smaller than those determined for the ionic 1-palmitoyl-2-hydroxy-sn-glycero-3-[phosphor-rac-(1-glycerol)] (LPPG) or zwitterionic detergents (FC-10, FC-12, DHPC), suggesting that in the latter cases electrostatic repulsion plays an important role.



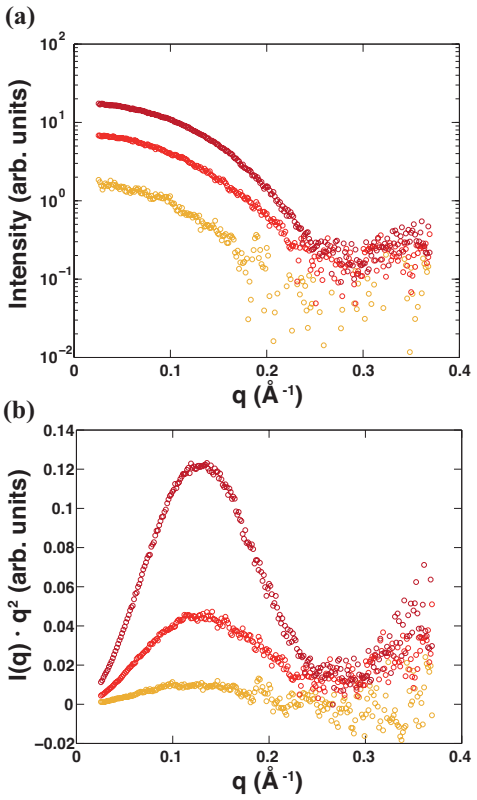
Supplementary Figure S1. Sample chamber and sample stage of the in-house SAXS setup. (a) Schematic front and side view of the sample chamber with an observation volume of 71 mm^3. (b) Image of the sample stage with heat bath connectors containing two sample chambers.



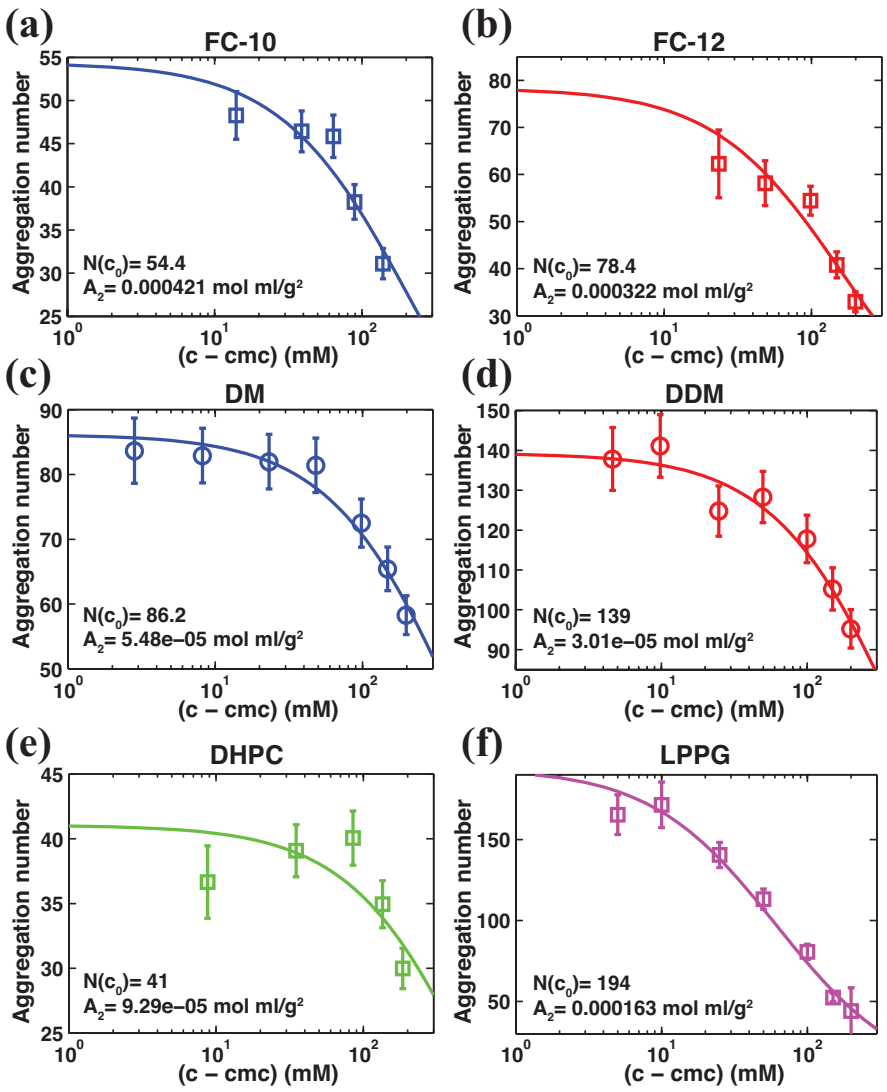
Supplementary Figure S2. Background measurements to estimate parasitic scattering in SAXS experiments. (a) Black diamonds (bottom) indicate the natural background level, measured with the Pilatus 100 K detector when the X-ray tube is turned off. Scattering curves (middle) of a deionized water sample when the flight path is fully evacuated using mica (red dots) and Kapton (blue dots) as window materials. Scattering curves (top) for the same samples described previously when the flight path is in air (mica: red squares; Kapton: blue squares). Data correspond to averaged scattering profiles from three runs with an exposure time of 2h each. (b) Schematic layout of the sample-to-detector setup for background measurements. When the flight path is fully evacuated only the primary beam (yellow) is blocked by the beamstop resulting in a beamstop shadow in the detector image (dark grey area). For a non-evacuated flight path, scattering from air molecules (illustrated by the red dashed lines) leads to an additional effect of partial shading in the detector image (light grey area) explaining the intensity decrease in (a, top) in the low q-region.



Supplementary Figure S3. Calibration standards for SAXS and WAXS measurements. (a,b) Detector images for measurements of the calibration standards silver behenate (AgBe) (a) and lanthanum hexaboride (LaB_6) (b) for SAXS and WAXS, respectively. (c,d) Measured diffraction curves for the AgBe (c) and LaB_6 (d) measurements. The first small peak in the LaB_6 diffraction pattern results from the Kapton foil used as window material.



Supplementary Figure S4. Scattering profiles for horse heart cytochrome *c* at different protein concentrations. (a) Scattering profiles for protein concentrations of 24 mg/ml (dark red, top), 8 mg/ml (red, middle) and 2 mg/ml (orange, bottom), averaged from three repeats of 2 h each. (b) Kratky representation of the data from (a).



Supplementary Figure S5. Determination of the true aggregation number $N(c_0)$ and the second virial coefficient A_2 from fits of the apparent aggregation number vs. concentration for different detergents. Symbols are aggregation numbers determined from the forward scattering intensity in SAXS measurements taken from Lipfert et al.[1] Solid lines are fits of the model in Eqn. S6 to the data. The fitted values for $N(c_0)$ and A_2 are shown as insets in each panel. Data are for FC-10 (panel a), FC-12 (panel b), DM (panel c), DDM (panel d), DHPC (panel e), and LPPG (panel f).

Supplementary References

1. Lipfert, J., Columbus, L., Chu, V.B., Lesley, S.A. & Doniach, S. Size and shape of detergent micelles determined by small-angle X-ray scattering. *Journal of Physical Chemistry B* **111**, 12427-12438 (2007).

2. Oliver, R.C. et al. Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLoS One* **8**, e62488 (2013).

3. Svergun, D.I. & Koch, M.H.J. Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735 (2003).

4. Bonneté, F., Finet, S. & Tardieu, A. Second virial coefficient: variations with lysozyme crystallization conditions. *Journal of Crystal Growth* **196**, 403-414 (1999).

5. George, A. & Wilson, W.W. Predicting protein crystallization from a dilute solution property. *Acta Crystallographica Section D* **50**, 361-365 (1994).

Bibliography

(1) Howarth, M. et al. (2006). A monovalent streptavidin with a single femtomolar biotin binding site. *Nature methods 3*, 267.

(2) Sedlak, S. M. et al. (2017). Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. *PloS one 12*, e0188722.

(3) Sedlak, S. M. et al. (2019). Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. *Nano Letters 19*, PMID: 30346175, 3415–3421.

(4) Erlich, K. R. et al. (2019). DNA-free directed assembly in single-molecule cut-and-paste. *Nanoscale 11*, 407–411.

(5) Löf, A. et al. (2018). Modular, ultra-stable, and highly parallel protein force spectroscopy in magnetic tweezers using peptide linkers. *bioRxiv*, 491977.

(6) Lee, G. U., Kidwell, D. A., and Colton, R. J. (1994). Sensing discrete streptavidin-biotin interactions with atomic force microscopy. *Langmuir 10*, 354–357.

(7) Florin, E.-L., Moy, V. T., and Gaub, H. E. (1994). Adhesion forces between individual ligand-receptor pairs. *Science 264*, 415–417.

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

(9) Baumann, F. et al. (2016). Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nature nanotechnology 11*, 89.

(10) Chaiet, L., and Wolf, F. J. (1964). The properties of streptavidin, a biotin-binding protein produced by Streptomycetes. *Archives of biochemistry and biophysics 106*, 1–5.

(11) Hendrickson, W. A. et al. (1989). Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proceedings of the National Academy of Sciences 86*, 2190–2194.

(12) Chivers, C. E. et al. (2011). How the biotin–streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *Biochemical Journal 435*, 55–63.

(13) Bayer, E. A., Wilchek, M., and Skutelsky, E. (1976). Affinity cytochemistry: The localization of lectin and antibody receptors on erythrocytes via the avidin-biotin complex. *FEBS letters 68*, 240–244.

(14) Hofmann, K., Finn, F. M., and Kiso, Y. (1978). Avidin-biotin affinity columns. General methods for attaching biotin to peptides and proteins. *Journal of the American Chemical Society 100*, 3585–3590.

(15) Argarana, C. E. et al. (1986). Molecular cloning and nucleotide sequence of the streptavidin gene. *Nucleic acids research 14*, 1871–1882.

(16) Freitag, S. et al. (1997). Structural Studies of the streptavidin binding loop. *Protein Science 6*, 1157–1166.

(17) Pähler, A et al. (1987). Characterization and crystallization of core streptavidin. *Journal of Biological Chemistry 262*, 13933–13937.

(18) Bayer, E. A. et al. (1989). Postsecretory modifications of streptavidin. *Biochemical Journal 259*, 369–376.

(19) Weber, P. C. et al. (1989). Structural origins of high-affinity biotin binding to streptavidin. *Science 243*, 85–88.

(20) Kim, M. et al. (2011). Nanomechanics of streptavidin hubs for molecular materials. *Advanced Materials 23*, 5684–5688.

(21) Green, N. M. In *Methods in enzymology*; Elsevier: 1990; Vol. 184, pp 51–67.

(22) Stenkamp, R. E. et al. (1997). Structural studies of the streptavidin binding loop. *Protein Science 6*, 1157–1166.

(23) Sano, T., and Cantor, C. R. (1995). Intersubunit contacts made by tryptophan 120 with biotin are essential for both strong biotin binding and biotin-induced tighter subunit association of streptavidin. *Proceedings of the National Academy of Sciences 92*, 3180–3184.

(24) Chilkoti, A., and Stayton, P. S. (1995). Molecular origins of the slow streptavidin-biotin dissociation kinetics. *Journal of the American Chemical Society 117*, 10622–10628.

(25) Chilkoti, A., Tan, P. H., and Stayton, P. S. (1995). Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. *Proceedings of the National Academy of Sciences 92*, 1754–1758.

(26) Stayton, P. S. et al. (1999). Streptavidin–biotin binding energetics. *Biomolecular engineering 16*, 39–44.

(27) Hyre, D. E. et al. (2006). Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein science 15*, 459–467.

(28) Liu, F., Zhang, J. Z., and Mei, Y. (2016). The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Scientific reports 6*, 27190.

(29) Klumb, L. A., Chu, V., and Stayton, P. S. (1998). Energetic roles of hydrogen bonds at the ureido oxygen binding pocket in the streptavidin- biotin complex. *Biochemistry 37*, 7657–7663.

(30) Freitag, S. et al. (1999). X-ray crystallographic studies of streptavidin mutants binding to biotin. *Biomolecular engineering 16*, 13–19.

(31) Bateman, W. G. (1916). The digestibility and utilization of egg-proteins. *Proceedings of the Society for Experimental Biology and Medicine 13*, 155–156.

(32) Boas, M. A. (1927). The effect of desiccation upon the nutritive properties of egg-white. *Biochemical Journal 21*, 712.

(33) Gyorgy, P. et al. (1939). The curative factor (vitamin H) for egg white injury, with particular reference to its presence in different foodstuffs and in yeast. *Journal of Biological Chemistry 131*, 733–744.

(34) Rushton, D. (2002). Nutritional factors and hair loss. *Clinical and Experimental Dermatology: Clinical dermatology 27*, 396–404.

(35) Traub, W (1956). Crystal structure of biotin. *Nature 178*, 649.

(36) Bayer, E. A. et al. (1986). An improved method for the single-step purification of streptavidin. *Journal of biochemical and biophysical methods 13*, 103–112.

(37) Reznik, G. O. et al. (1998). A streptavidin mutant with altered ligand-binding specificity. *Proceedings of the National Academy of Sciences 95*, 13525–13530.

(38) Schmidt, T. G., and Skerra, A. (1993). The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. *Protein Engineering, Design and Selection 6*, 109–122.

(39) Keefe, A. D. et al. (2001). One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. *Protein expression and purification 23*, 440–446.

(40) Bessette, P. H., Rice, J. J., and Daugherty, P. S. (2004). Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Engineering Design and Selection 17*, 731–739.

(41) Freitag, S. et al. (1998). Structural studies of binding site tryptophan mutants in the high-affinity streptavidin-biotin complex. *Journal of molecular biology 279*, 211–221.

(42) Hyre, D. E. et al. (2000). Ser45 plays an important role in managing both the equilibrium and transition state energetics of the streptavidin–biotin system. *Protein Science 9*, 878–885.

(43) Aslan, F. M. et al. (2005). Engineered single-chain dimeric streptavidins with an unexpected strong preference for biotin-4-fluorescein. *Proceedings of the National Academy of Sciences 102*, 8507–8512.

(44) Lim, K. H. et al. (2011). Engineered streptavidin monomer and dimer with improved stability and function. *Biochemistry 50*, 8682–8691.

(45) Chilkoti, A. et al. (1995). Engineered chimeric streptavidin tetramers as novel tools for bioseparations and drug delivery. *Nature Biotechnology 13*, 1198.

(46) Voss, S., and Skerra, A. (1997). Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein engineering 10*, 975–982.

(47) Peters, V., and Rehm, B. H. (2008). Protein engineering of streptavidin for in vivo assembly of streptavidin beads. *Journal of biotechnology 134*, 266–274.

(48) Laitinen, O. et al. (2006). Genetically engineered avidins and streptavidins. *Cellular and Molecular Life Sciences CMLS 63*, 2992–3017.

(49) Schmidt, T. G., and Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature protocols 2*, 1528.

(50) Müller, J. P. et al. (2016). Force sensing by the vascular protein von Willebrand factor is tuned by a strong intermonomer interaction. *Proceedings of the National Academy of Sciences 113*, 1208–1213.

(51) Erlich, K. R. et al. (2017). Strep-Tag II and Monovalent Strep-Tactin as Novel Handles in Single-Molecule Cut-and-Paste. *Small Methods 1*, 1700169.

(52) Korndörfer, I. P., and Skerra, A. (2002). Improved affinity of engineered streptavidin for the Strep-tag II peptide is due to a fixed open conformation of the lid-like loop at the binding site. *Protein science 11*, 883–893.

(53) Chivers, C. E. et al. (2010). A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nature methods 7*, 391.

(54) Zhang, M. et al. (2016). The Crystal Structure of Monovalent Streptavidin. *Scientific reports 6*, 35915.

(55) Fairhead, M. et al. (2014). Plug-and-play pairing via defined divalent streptavidins. *Journal of molecular biology 426*, 199–214.

(56) Qureshi, M. H. et al. (2001). Development and characterization of a series of soluble tetrameric and monomeric streptavidin muteins with differential biotin binding affinities. *Journal of Biological Chemistry 276*, 46422–46428.

(57) Laitinen, O. H. et al. (2003). Rational design of an active avidin monomer. *Journal of Biological Chemistry 278*, 4010–4014.

(58) Wu, S.-C., and Wong, S.-L. (2005). Engineering soluble monomeric streptavidin with reversible biotin binding capability. *Journal of Biological Chemistry 280*, 23225–23231.

(59) Van't Hoff, J. H., *Etudes de dynamique chimique*; Muller: 1884; Vol. 1.

(60) Arrhenius, S. (1889). Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Zeitschrift für physikalische Chemie 4*, 226–248.

(61) Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica 7*, 284–304.

(62) Hänggi, P., Talkner, P., and Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Reviews of modern physics 62*, 251.

(63) Binnig, G., Quate, C. F., and Gerber, C. (1986). Atomic Force Microscope. *Phys. Rev. Lett. 56*, 930–933.

(64) Züger, O., and Rugar, D. (1993). First images from a magnetic resonance force microscope. *Applied Physics Letters 63*, 2496–2498.

(65) Izrailev, S. et al. (1997). Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical journal 72*, 1568–1581.

(66) Friedsam, C. et al. (2003). Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. *Journal of Physics: Condensed Matter 15*, 1709.

(67) Evans, E., and Ritchie, K. (1997). Dynamic strength of molecular adhesion bonds. *Biophysical journal 72*, 1541–1555.

(68) Hummer, G., and Szabo, A. (2003). Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical journal 85*, 5–15.

(69) Dudko, O. et al. (2003). Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proceedings of the National Academy of Sciences 100*, 11378–11381.

(70) Friddle, R. W., Noy, A., and De Yoreo, J. J. (2012). Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proceedings of the National Academy of Sciences 109*, 13573–13578.

(71) Dudko, O. K., Hummer, G., and Szabo, A. (2006). Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical review letters 96*, 108101.

(72) Tshiprut, Z, Klafter, J, and Urbakh, M (2008). Single-molecule pulling experiments: When the stiffness of the pulling device matters. *Biophysical journal 95*, L42–L44.

(73) Walton, E. B., Lee, S., and Van Vliet, K. J. (2008). Extending Bell's model: how force transducer stiffness alters measured unbinding forces and kinetics of molecular complexes. *Biophysical journal 94*, 2621–2630.

(74) Freund, L. (2009). Characterizing the resistance generated by a molecular bond as it is forcibly separated. *Proceedings of the National Academy of Sciences 106*, 8818–8823.

(75) Ott, W. et al. (2017). Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano 11*, PMID: 28591514, 6346–6354.

(76) Dietz, H., and Rief, M. (2004). Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proceedings of the National Academy of Sciences 101*, 16192–16197.

(77) Schwaiger, I. et al. (2004). A mechanical unfolding intermediate in an actin-crosslinking protein. *Nature structural & molecular biology 11*, 81.

(78) Milles, L. F. et al. (2017). Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium perfringens. *The Journal of Physical Chemistry B 121*, PMID: 27991799, 3620–3625.

(79) Yin, J. et al. (2005). Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences 102*, 15815–15820.

(80) Durner, E. et al. (2017). Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. *ACS Omega 2*, 3064–3069.

(81) Beckett, D., Kovaleva, E., and Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Science 8*, 921–929.

(82) Czerwinski, M. et al. (2009). Construction of an agglutination tool: recombinant Fab fragments biotinylated in vitro. *New biotechnology 26*, 215–221.

(83) Ashraf, S. S. et al. (2004). A novel multi-affinity tag system to produce high levels of soluble and biotinylated proteins in Escherichia coli. *Protein expression and purification 33*, 238–245.

(84) Bustamante, C. et al. (1994). Entropic elasticity of lambda-phage DNA. *Science*, 1599–1599.

(85) Puchner, E. M. et al. (2008). Comparing proteins by their unfolding pattern. *Biophysical journal 95*, 426–434.

(86) Hutter, J. L., and Bechhoefer, J. (1993). Calibration of atomic-force microscope tips. *Review of Scientific Instruments 64*, 1868–1873.

(87) Butt, H.-J., and Jaschke, M. (1995). Calculation of thermal noise in atomic force microscopy. *Nanotechnology 6*, 1.

(88) Neuman, K. C., and Nagy, A. (2008). Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature methods 5*, 491.

(89) Bauer, M. S. et al. (2018). Monomeric streptavidin: a versatile regenerative handle for force spectroscopy. *bioRxiv*, 276444.

(90) Sedlak, S. M., Bruetzel, L. K., and Lipfert, J. (2017). Quantitative evaluation of statistical errors in small-angle X-ray scattering measurements. *Journal of applied crystallography 50*, 621–630.

(91) Bruetzel, L. K. et al. (2016). Conformational changes and flexibility of DNA devices observed by small-angle X-ray scattering. *Nano letters 16*, 4871–4879.

(92) Bruetzel, L. K. et al. (2016). A Mo-anode-based in-house source for small-angle X-ray scattering measurements of biological macromolecules. *Review of Scientific Instruments 87*, 025103.

List of Publications and Manuscripts

Publications in Peer-Reviewed Journals

1. Bruetzel LK, Gerling T, **Sedlak SM**, Walker PU, Zheng W, Dietz H, Lipfert J. Conformational changes and flexibility of DNA devices observed by small-angle X-ray scattering. Nano letters. 2016 Jul 27;16(8):4871-9.
(*cf.* Section 6.2)

2. Bruetzel LK, Fischer S, Salditt A, **Sedlak SM**, Nickel B, Lipfert J. A Mo-anode-based in-house source for small-angle X-ray scattering measurements of biological macromolecules. Review of Scientific Instruments. 2016 Feb 8;87(2):025103.
(*cf.* Section 6.3)

3. **Sedlak SM**, Bruetzel LK, Lipfert J. Quantitative evaluation of statistical errors in small-angle X-ray scattering measurements. Journal of applied crystallography. 2017 Apr 1;50(2):621-30.
(*cf.* Section 6.1)

4. **Sedlak SM**, Bauer MS, Kluger C, Schendel LC, Milles LF, Pippig DA, Gaub HE. Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. PloS one. 2017 Dec 5;12(12):e0188722.
(*cf.* Section 3.1)

5. Erlich KR,* **Sedlak SM**,* Jobst MA, Milles LF, Gaub HE. DNA-free directed assembly in single-molecule cut-and-paste. Nanoscale. 2019;11(2):407-11.
(*cf.* Section 4.1)

6. **Sedlak SM**,* Schendel LC,* Melo MC, Pippig DA, Luthey-Schulten Z, Gaub HE, Bernardi RC. Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. Nano letters. 2019;19(6):3415-3421.
(*cf.* Section 3.2)

Preprints

7. Bauer MS, Milles LF, **Sedlak SM**, Gaub HE. Monomeric streptavidin: a versatile regenerative handle for force spectroscopy. bioRxiv. 2018 Jan 1:276444.
(*cf.* Section 5.1)

8. Löf A, Walker PU, **Sedlak SM**, Obser T, Brehm MA, Benoit M, Lipfert J. Modular, ultra-stable, and highly parallel protein force spectroscopy in magnetic tweezers using peptide linkers. bioRxiv. 2018 Jan 1:491977.
(*cf.* Section 4.2)

In Preparation

9. **Sedlak SM**, Schendel LC, Gaub HE, Bernardi RC. Streptavidin/biotin: Tethering Geometry defines Unbinding Mechanics.
(*cf.* Section 3.3)

10. Kluger C,* Braun L,* **Sedlak SM**, Bauer MS, Miller K, Milles LF, Pippig DA, Gaub HE, Vogel V. Directional asymmetry of force transmission across vinculin binding sites.
(in preparation - not included in this thesis)

11. **Sedlak SM**, Milles LF, Gaub HE. 25 years of force spectroscopy on the streptavidin/biotin interaction.
(*cf.* Section 1.5)

12. Löf A,* **Sedlak SM**,* Gruber S, Benoit M, Lipfert J, Gaub HE. The tetravalency of streptavidin results in distinct lifetimes of biotin unbinding under constant load.
(*cf.* Section 3.4)

13. Schendel LC,* **Sedlak SM**,* Gaub HE. Reinforcing Streptavidin.
(in preparation - not included in this thesis)

List of Figures

List of Tables

Danksagung

Vieles in der Experimentalphysik braucht Teamwork. Meistens lernt man Dinge nicht aus dem Textbuch, sondern von seinen Kollegen. Man entwickelt gemeinsam Ideen, tauscht Erfahrungen aus und diskutiert Ergebnisse. In meiner Doktorandenzeit durfte ich viele großartige Menschen kennenlernen und eng mit ihnen zusammenarbeiten. Oft sind daraus Freundschaften entstanden, bei denen ich mir sicher bin, dass sie mir über das Studentendasein hinaus erhalten bleiben werden. Viele Menschen haben mich während meiner Doktorarbeit begleitet und in vielfältiger Weise unterstützt. Dafür bin ich von Herzen dankbar. Einige will ich im Folgenden besonders erwähnen:

Hermann: Danke, dass ich bei Dir meine Doktorarbeit schreiben konnte, dass wir so viel über das SA/Biotin-System diskutiert haben, dass ich die Freiheit hatte, eigene Ideen auszuprobieren, dass ich vieles in Bezug auf mein Projekt selbst entscheiden durfte und dass Du immer eine offene Tür und ein offenes Ohr hattest. Danke für die entspannte Arbeitsatmosphäre an Deinem Lehrstuhl – nicht zuletzt durch die Ausflüge ins Oberland und das Sorgen fürs leibliche Wohl.

Magnus: Danke, dass Du mir so viel beigebracht hast – über Biophysik, Proteine, Kraftspektroskopie, Programmieren, Wissenschaft, aber auch über Politik, Philosophie, das Leben. Die Diskussionen mit Dir waren sehr wertvoll und Du hast sehr viel zum Gelingen dieser Arbeit beigetragen. Mit Deinem rechtschaffenen Charakter und Deiner ruhigen geradlinigen Art bist Du ein echtes Vorbild.

Leo: Danke, dass Du Dich so sehr beim SA/Biotin-Projekten eingebracht hast. Besonders, wenn ich an unseren Daten zweifelte, hast Du mich immer wieder bestärkt und geholfen, noch unklare Ergebnisse in den größeren Kontext einzuordnen, um neue Ideen fürs weitere Vorgehen zu entwickeln.

Rafael: Danke für die produktive Zusammenarbeit, die vielen guten Diskussionen über Skype und Deine Besuche bei uns. Danke, dass Du Dir so viel Zeit genommen und zu all unseren Extrawünschen nie ”Nein” gesagt hast.

Diana: Danke, dass Du mich überzeugt hast, bei Euch anzufangen, dass Du mich für das SA/Biotin-Projekt begeistert hast, dass Du mir viel über biomolekulare Methoden beigebracht hast und dass Du bei Fragen immer erreichbar warst.

Carleen: Danke für die gemeinsamen Projekte, die Diskussionen über das SA/Biotin-Projekt, die Einführung in die Arbeit mit HEK-Zellen, das Korrekturlesen und dafür, dass man mit Dir auch die ausgefallensten Ideen diskutieren kann.

Achim: Danke für die gemeinsamen Experimente mit magnetischen Pinzetten, fürs AFM-Imaging und für die gemeinsame Betreuung von Antonias Projekt, für Dein wertvolles Feedback zur Doktorarbeit und all die gemeinsame Zeit fernab der Wissenschaft.

Katy: Danke, dass Du mit mir das Streptavidin/Biotin-System in SMC& P implementiert hast. Danke für Deine Hilfe zu den Feinheiten der englischen Sprache.

Lukas: Danke für die fca-Software, Deine Hilfsbereitschaft, Deine Ideen, Denkanstöße und guten Empfehlungen. Deine Belesenheit und Dein Wissen waren mir oft eine Hilfe.

Ellis: Danke für all die Sachen, die Du am Lehrstuhl für uns alle machst, für Dein direktes Feedback, Deine ehrliche Meinung und Deine Tipps zu Laborprotokollen, Messmethoden und Datenanalyse.

Tom und **Angelika**: Danke fürs Exprimieren, Klonieren, Silanisieren und all die anderen unzähligen Laborarbeiten, für die Beratung zu Laborprotokollen, für die Musik und für die Smoothies.

Sylvia: Danke für die Unterstützung bei administrativen Angelegenheiten, für die guten Gespräche morgens beim Kaffee und natürlich auch für den Kaffee selbst.

Fabi, **Cons**, **Klara**, **Markus**, **Tobi** und **Wolfgang**: Danke, für alles, was Ihr am Lehrstuhl an Methoden, Vorgehensweisen, Instrumenten und Software etabliert habt und meinen Kollegen und mir weitergegeben habt. Vieles, was ich in meiner Arbeit einfach so verwenden konnte, habt Ihr entwickelt.

Antonia, **Jonas** und **Philipp**: Danke, dass Ihr immer motiviert und fleißig wart, dass Ihr zuverlässig im Labor gearbeitet habt und dass Ihr die richtigen Fragen gestellt und mit mir diskutiert habt.

Jan: Danke, dass Du mich für die Biophysik begeistert hast und dass ich noch immer im Lipfertlab mitmischen darf. Danke für die Projekte mit Linda, Thomas und Samuel, für die SA/Biotin-Messungen mit der Magnetischen Pinzette und für Deine Unterstützung.

Philipp: Danke für die Unterstützung am Anfang meiner Doktorarbeit, für die gemeinsame Zeit im Büro, für Deinen Sarkasmus, Deine Ironie, Deinen Humor, die vielen Aktionen außerhalb der Uni wie die Reise nach Belgien und die Hilfe beim Umzug.

Tine: Danke für die gute gemeinsame Zeit im Labor, für die gute Laune und für die Ausflüge zur Eisdiele, in die Berge und nach Belgien.

Anki: Danke für die gute Zeit im Büro und außerhalb der Uni: WG-Partys, Berge, Silvester, Oktoberfest und der Berch. Du bist einfach top.

Linda: Danke, dass ich bei Deinen Projekten mitarbeiten durfte und dass Du mir geholfen hast, mein SAXS-Fehlermodell zu veröffentlichen.

Thomas: Danke für die gemeinsame Zeit im Büro, Kekse und Kaffee und dafür, dass Du mich zu den SAXS-Messungen zum ESRF mitgenommen hast.

Samuel: Danke für eine grandiose Messzeit mit SAXS-Experimente am ESRF. Isso.

Sophia: Danke für die Unterstützung mit den magnetischen Pinzetten und den superparamagnetischen Kugeln.

Beno und der **Band UnCeNSiert**: Danke fürs miteinander Krach machen, für viele Stunden chaotischer Bandproben und für die unzähligen "öffentlichen Generalproben", sei es bei der Absolventenfeier der Fakultät, CeNS-Partys oder bei Konferenzen wie z.B. in Venedig.

Selina: Danke für die vielen lustigen Aktionen mit unseren Nachbarn, für gute Partys und für Leben in unserer Bude. Bei all der Wissenschaft warst Du einfach immer die Person, die halt weiß, worauf es im Leben ankommt.

Basti, **Konrad**, **Philip** und **Dima**: Danke für die gemeinsame Zeit mit Euch in den letzten neun Jahren, für Studium, Feiern, Wandern, Skifahren, Musik, Theater, Reisen, und vieles mehr – für Eure Freundschaft.

Meinen **Eltern**, meinen **Geschwistern** und meiner ganzen **Familie**: Danke für Eure Unterstützung und für die zahlreichen Besuche bei Euch, die immer wie ein erholsamer Kurzurlaub mit Rundumverpflegung waren. Es ist so gut, dass ich Euch habe und immer auf Euch zählen kann.

Tabea: Danke, dass Du mich immer unterstützt, dass Du für mich da bist und dass ich mich auf Dich verlassen kann. Danke für Dein Vertrauen und Deine Liebe.