Single-Molecule Directed Assembly and Enzyme Investigation

Katherine Erlich

München 2018

Single-Molecule Directed Assembly and Enzyme Investigation

Katherine Erlich

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

vorgelegt von
Katherine Erlich
aus Vallejo, CA, USA

München, den 13.09.2018

Erstgutachter: Prof. Dr. Hermann E. Gaub
Zweitgutachter: Prof. Dr. Philip Tinnefeld
Tag der mündlichen Prüfung: 19.10.2018

Zusammenfassung

Einzelmolekül-Cut-and-Paste (SMC&P) ist eine vielseitige Technik im Feld der Nanotechnologie, die mittels der Kontrolle von einzelnen Molekülen Bottom-up-Assemblierungen ermöglicht. Mit Hilfe eines Rasterkraftmikroskops können Moleküle in beliebigen Mustern auf einer Oberfläche mit einer Genauigkeit in der Größenordnung von Nanometern angeordnet werden. So angeordnete Moleküle können dann über Fluoreszenzmikroskopie detektiert werden, was eine zeitaufgelöste Untersuchung auf Einzelmolekülebene ermöglicht.

Die vorliegende Arbeit verfolgt die Integration funktioneller Enzyme als Transfermolekül in SMC&P. Im Prinzip können Enzyme über SMC&P auf einer Oberfläche angeordnet und deren Aktivität über floureszente Reaktionsprodukte beobachtet werden. Dies eröffnet spannende neue Möglichkeiten Enzyme auf Einzelmolekülebene zu untersuchen. Beispielsweise können Einzelmolekülmethoden wie SMC&P zeitaufgelöstes Enzymverhalten offenlegen, wie zum Beispiel die Existenz von Verzögerungsphasen oder spontanem Aktivitätsanstieg. Zusätzlich würde SMC&P auch eine statistische Analyse der Verteilung der katalytischen Aktivität in einer Population von Enzymen ermöglichen. Solche statistischen Analysen sind in Untersuchungen eines Gesamtensembles von Molekülen inhärent unmöglich, da nur das durchschnittliche Verhalten einer Population zu einem bestimmten Zeitpunkt berichtet beobachtet werden kann. Schließlich bietet die Fähigkeit, Moleküle auf einer Oberfläche spezifisch anzuordnen, die Möglichkeit, das dynamische katalytische Zusammenspiel zu untersuchen, wenn Enzyme in verschiedenen Mustern angeordnet sind, wodurch Informationen über geometrie- und dichteabhängige Prozesse gewonnen werden können. Werden mehrere verschiedene Enzyme zu einem multifunktionalen Netzwerk arrangiert, könnte das Zusammenspiel zwischen Enzymen, die für verschiedene Reaktionsstufen in komplexen biologischen Prozessen verantwortlich sind, besser verstanden und charakterisiert werden.

Die zugrundeliegenden Arbeit wird zunächst durch einen Überblick über relevante Hintergrundinformationen in den wissentschaftlichen Kontext eingeordnet. Dies umfasst sowohl die Funktion von Enzymen aus einer biophysikalischen Perspektive, als auch eine Einführung in Technik und Theorie von Einzelmolekültechniken. Weiter werden die biochemischen, biophysikalischen und biotechnologischen Eigenschaften jedes näher beleuchteten Enzyms genauer diskutiert. Die benötigten Schlüsselmethoden in Molekularbiologie, Einzelmolekül-Kraftspektroskopie und Fluoreszenmikroskopie werden dargelegt.

Es wurden umfangreiche Strategien für die Fluoreszenz- oder Biolumineszenz-Detektion jedes untersuchten Enzyms entworfen, um letztlich, Netzwerke von Enzymen mittels SMC&P zu schaffen. Gründliche Tests und Charakterisierungen potentiell interessanter Enzyme und deren jeweiligen Fluoreszenz-Auslesesysteme werden vorgestellt und diskutiert. Neben der Entwicklung von Strategien zur Enzymaktivitätsmessung wurde das Repertoire der SMC&P-Handhabungsstrategien erweitert. Darüber hinaus wird die erfolgreiche Implementierung eines komplett DNA-freien SMC&P-Systems demonstriert, das insbesondere für die Bottom-up-Assembly von DNA-bindenden Enzymen von vorteil ist. Dieser methodische Fortschritt ermöglicht ein potentiell neues Paradigma für zukünftige SMC&P-Experimente.

Preface

Single-molecule cut-and-paste (SMC&P) is a versatile technique in the field of nanotechnology, merging bottom-up assembly and single-molecule directed control. With the aid of an atomic force microscope, molecules of interest are arranged in specifically designed patterns on a surface with precision on the order of nanometers. Arranged molecules may then be detected via fluorescence, enabling time-resolved analysis of single molecules.

Soon after SMC&P was successfully tested as a proof of principle, it was quickly realized that this technique had great potential to investigate single-molecule activity beyond mere fluorescence. For example, DNA aptamers were assembled on the surface, forming sites in which a target ligand's structure was stabilized and its fluorescence enhanced [1]. Furthermore, SMC&P offers absolute control of the placement of molecules on a surface, thereby circumventing many difficulties that arise from stochastic immobilization. For example, SMC&P enables precise placement of molecules within the center of nanoaperatures [2], decreasing the heterogeneity of fluorescence intensity and lifetime.

The work presented in this thesis pursued integration of functional enzymes into SMC&P as transfer cargo. In concept, enzymes can be arranged on a surface via SMC&P, and their activities monitored via fluorescent output. This opens up several exciting possible avenues of enzyme investigations on the order of single molecules. Firstly, single-molecule measurements such as SMC&P can reveal time-resolved enzyme behavior, such as bursts of activity or periods of stalling. Additionally, with the assumption that the nontrivial process of protein folding leads to a spectrum of final folded states - and accordingly a spectrum of final functional states - SMC&P would also enable assessment of the distribution of catalytic activity within a population of enzymes. Such statistical analyses are inherently impossible in bulk studies, where only the mean behavior within a population at any given moment is reported.

Lastly, the ability to specifically arrange molecules on a surface offers the possibility to investigate dynamic catalysis when enzymes are arranged in various patterns, gleaning information into geometry- and density-dependent processes. Moreover, were several different enzymes arranged into a multifunctional network, the interplay between enzymes that are responsible for different steps in complex biological processes could be better understood and characterized.

Relevant background information to provide context of the scientific work is discussed in Chapters 1 - 3. An overview of enzyme function from a biophysical perspective is presented in Chapter 1. This chapter also discusses the heterogeneity of enzyme populations and the corresponding investigative value of single-enzyme analysis. Chapter 2 provides an introduction to the technology and theory pertinent to single-molecule probing. In particular, the AFM, total internal reflection fluorescence microscopy, and SMC&P are discussed. Chapter 3 elaborates on the the core library of enzymes of interest investigated in this work. The biochemical, biophysical, and biotechnological properties of each enzyme are discussed at length, providing a basis to understand the function of the enzymes within the framework of this thesis as well as the scientific community's general interest in these particular enzymes.

The key methodologies utilized in this work are outlined in Chapters 4 and 5. Extensive molecular biology techniques were employed to design, express, purify and test enzymes of interest as well as related materials. Chapter 4 provides a broad description of these techniques, with particular attention given to non-standard methodologies or unique protocols developed for this work. Chapter 5 briefly describes the methods employed in single-molecule force spectroscopy and fluorescence microscopy within this work.

The principle results in this thesis are detailed in Chapters 6 - 8. Unique strategies for fluorescent or bioluminescent readout were explored for each enzyme of interest, with the eventual goal of integrating networks of enzymes of interest into SMC&P. Rigorous testing and characterization of enzymes of interest and their respective fluorescence readout systems are presented and discussed in Chapter 6.

In addition to development of enzyme activity measurement strategies, the repertoire of SMC&P handling strategies was expanded. The incorporation of monovalent *Strep*-Tactin and its peptide ligand *Strep*-Tag II as a handling system between cantilever and transfer construct is presented in Chapter 7. Furthermore, the successful development of completely DNA-free SMC&P - which may prove advantageous for bottom-up assembly of DNA-binding enzymes - is introduced in Chapter 8. This methodological advancement in particular represents a potentially new paradigm for future SMC&P experiments.

Contents

Part I

Scientific Context

Chapter 1
Enzymes: Biological Machinery

Overview

Chemical reactions are an integral component of life at all scales, from the nanoscopic to the macroscopic. Energy sources must be broken down, new molecular structures built, and biophysical information relayed. Even the simplest organisms yet discovered - arguably single-celled bacteria - undertake hundreds of distinct chemical reactions in order to survive and replicate.

Through billions of years of evolution and natural selection, nature has found a remarkable strategy to coordinate chemical reactions within biological systems: encoded into an organism's genome and synthesized from the elemental building blocks of amino acids and nucleic acids, enzymes are highly specialized biological catalysts with a vast array of functions. Virtually all chemical steps of metabolism, signal transduction, and cell growth utilize enzymes in order to precisely and efficiently carry out the required reactions.

The production and activity of enzymes is undeniably of great interest across a broad range of fields, from medicine and pharmacology to renewable energy and basic physics. The following section seeks to lay out the fundamental principles of enzymes from biophysical and biochemical perspectives, as well as contextualize the motivation for the results presented in this thesis.

1.1 Catalysis within Biological Systems

In its most fundamental sense, a chemical reaction occurs when the electrons from one atom's orbital shell are transferred to or shared with another atom, and typically results in the formation or destruction of a covalent bond between the two atoms. The primary physical force driving such a process is the release of energy that leaves the system in a more energetically stable

configuration. The thermodynamic favorability of a process can be expressed with the change in Gibbs free energy, which is given by:

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

where the change in free energy ΔG is a function of the change in enthalpy ΔH, temperature T and the change in entropy ΔS. For any given process, if $\Delta G < 0$ then the system overall releases energy into the surroundings and is considered to be exergonic. In other words, this process is energetically favorable and it will likely occur spontaneously without the addition of external energy. This is in contrast with an endergonic process with $\Delta G > 0$, which requires an input of external energy in order to proceed.

However, even if a process has net change in free energy $\Delta G_{net} < 0$ and is overall exergonic, the forward reaction may require the system to first proceed through an intermediate state of higher energy. This initial endergonic step prevents the complete reaction from proceeding spontaneously. This so-called energy barrier between the starting and end states has an associated activation energy E_a that must be overcome in order for the reaction to proceed. Random thermal energy fluctuations on the order of $k_B T$ alone may allow the system to overcome the barrier, although the rate of the reaction may possibly be so slow to the point of being statistically impossible to occur. For such chemical reactions, catalysts are crucial for overcoming the energy barrier and driving the reaction forward.

For a non-spontaneous reaction, the formation of the intermediate state between reactants and products is typically the slowest step, as it is the highest-energy state the system must occupy throughout this process. In general, catalysts propel otherwise slow reactions forward by stabilizing the intermediate state and therefore lowering the activation energy required to overcome the energy barrier (Figure 1.1). The conversion of reactants R to intermediates I and finally to products P can be understood schematically by:

$$\mathrm{R} \underset{}{\overset{k_1}{\rightleftharpoons}} \mathrm{I} \longrightarrow \mathrm{P} \tag{1.2}$$

whereas the conversion in the presence of a catalyst C can be understood schematically by:

$$\mathrm{C} + \mathrm{R} \rightleftharpoons \mathrm{C}\cdot\mathrm{R} \overset{k_2}{\rightleftharpoons} \mathrm{C}\cdot\mathrm{I} \longrightarrow \mathrm{C}\cdot\mathrm{P} \longrightarrow \mathrm{C} + \mathrm{P} \tag{1.3}$$

such that the rates of the limiting steps of intermediate formation obey $k_1 << k_2$. The formation of the intermediate, whether it occurs spontaneously or catalyzed, is considered a reversible process. The final step of the product formation is usually fast and considered irreversible, especially for reactions that have $\Delta G_{net} << 0$. Importantly, the catalyst is not consumed and does not change ΔG_{net} of the reaction. However, by simply lowering the energy barrier, catalysts are capable of expediting chemical reactions by many orders of magnitude.

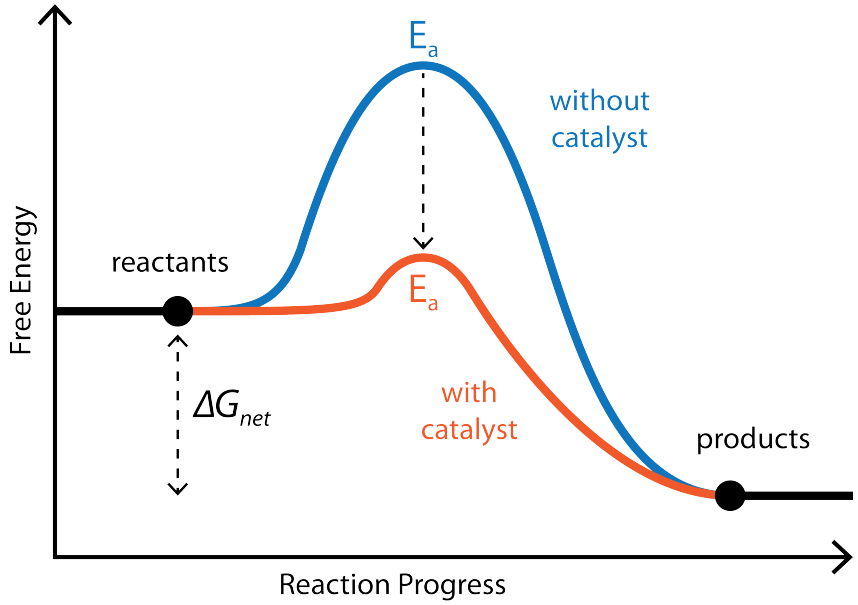


Figure 1.1: Visual representation of a chemical reaction. The chemical conversion of reactants to products requires overcoming an energy barrier with activation energy E_a. The introduction of a catalyst lowers E_a, and thereby increases the likelihood that the reaction proceeds. The net change in free energy ΔG_{net} of the reaction is unaffected by the presence or absence of the catalyst. However, as the catalyzed reaction has a much lower energy barrier that can more easily be overcome by thermal fluctuations, the rate of the reaction is likely several orders of magnitude faster with the catalyst than without.

Biological systems require programmable machinery that can efficiently and specifically catalyze chemical reactions under biologically relevant conditions. Moreover, it is crucial that these catalysts are regulated according to the organism's momentary needs. Nature's elegant solution is for organisms to synthesize their own genetically-encoded catalysts - *i.e.*, enzymes - from fundamental biological building blocks such as amino acids and nucleic acids.

Enzymes perform several basic functions that enable them to catalyze such chemical reactions. Firstly, like most catalysts, enzymes stabilize a transition state between substrate and product. The enzyme forms temporary bonds with the substrates via reactive or charged motifs, producing an intermediary transition state. The products finally form as the reaction completes and are ejected from the enzyme's active site, thereby recovering the original state of the enzyme. The conversion of substrates S to intermediates I and finally to products P in the presence of enzyme E can be understood schematically with an expansion of the Michaelis-Menten model of enzyme activity:

$$\mathrm{E} + \mathrm{S} \rightleftharpoons \mathrm{E{\cdot}S} \rightleftharpoons \mathrm{E{\cdot}I} \longrightarrow \mathrm{E{\cdot}P} \longrightarrow \mathrm{E} + \mathrm{P} \tag{1.4}$$

Compared to the same reaction in the absence of the enzyme, the intermediate state is much more likely to occur, as the enzyme stabilizes this configuration and thus decreases the activation energy required to overcome the energy barrier. In other words, enzymes have a high binding specificity for the transition state, as was first proposed by Linus Pauling [3]. Many biological

reactions can effectively only proceed with the assistance of an enzyme, as the timescales of the uncatalyzed reactions are extremely slow. For example, the yeast enzyme OMP decarboxylase catalyzes the decarboxylation of orotidine monophosphate to form uridine monophosphate with an estimated reaction rate of 18 ms, while the uncatalyzed reaction has an estimated reaction rate on the order of 78 million years [4]. This attribute of transition state stabilization is shared by most catalysts, as per their definition. However, enzymes possess other abilities that enable them to further facilitate chemical reactions in ways that inorganic catalysts often cannot.

Enzymes often couple an energetically unfavorable reaction to a favorable one, such that the negative ΔG_{net} of the overall process enables the thermodynamically unfavorable reaction to occur. Reactive energy carrier molecules - namely adenosine triphosphate (ATP) - are frequently used as energetic fuel for endergonic reactions. Enzymes hydrolyze the γ-phosphate or β-pyrophosphate of ATP - both highly exergonic reactions - while simultaneously catalyzing another reaction that on its own would have $\Delta G_{net} < 0$ [5]. This step is critical to many metabolic reactions, as energy must be expended in order to construct complex forms of higher energy states from much simpler molecules.

Lastly, if the chemical reaction requires two or more substrates, enzymes will often bind both components with high affinity so as to increase the local concentration of the substrates. This is critical for reactions where at least once substrate is found in relatively low concentrations, and Brownian diffusion would not reliably bring the two reactants into contact long enough for a reaction to occur.

Through a combination of these three major properties of enzymes, biological systems are able to undertake the chemical reactions necessary for their function. Furthermore, enzymes offer a means of biology-based regulation of these reactions. In contrast to spontaneous reactions, enzyme-catalyzed reactions can only occur at biologically-relevant timescales in the presence of an enzyme, which is subject to regulation by the host organism. Enzymes therefore function like on/off switches for such reactions, allowing tight control of biochemistry by modulating enzyme expression. Gene expression is subject to intricate systems of transcriptional and translational regulation, which are informed by parameters such as environmental factors or cell cycle stage. Enzymes are also sensitive to signals in the form of other molecules or covalent modifications such as phosphorylation, which may induce activation or inhibition depending on the enzyme in question. These signals can fine-tune the function and activity of enzymes, producing a remarkably elegant network of biochemical reactions and responses.

1.2 Orders of Protein Structure

Protein structure - and by extension enzyme structure - is conceptualized as four levels of organization (Figure 1.2). The primary structure of a protein is the linear sequence of amino acid residues that form the polypeptide, with synthesis starting at the amino terminus and ending at the carboxyl terminus. The protein is initially highly unstructured and takes the conformation of a random coil. Hydrogen bonds, ionic interactions, and van der Waals interactions arise

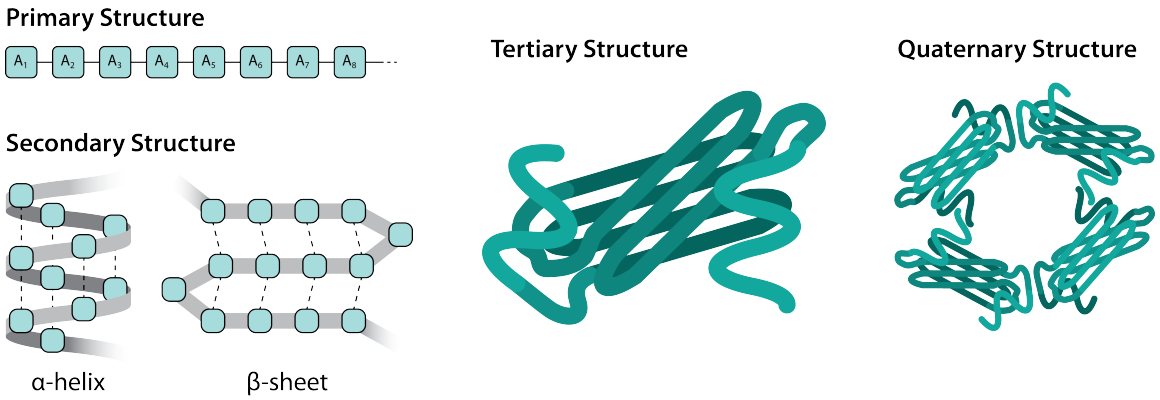


Figure 1.2: The four levels of protein structure. The primary structure consists only of the linear sequence of amino acid residues that form the polypeptide chain. The secondary structure results from the formation of noncovalent bonds between neighboring residues into stable motifs, such as α-helices and β-sheets. The hydrophobic residues cluster towards each other to minimize their contact with the aqueous environment and are shielded by hydrophilic residues, forming the final folded protein and the tertiary structure. The quaternary structure is a multimeric complex assembled from distinct subunits.

between neighboring residues as the polypeptide grows. These weak noncovalent interactions confer a favorable decrease in enthalpic energy of the protein and drive the assembly of secondary structures. Some of the more canonical motifs, such as α-helices and β-sheets, are ubiquitously found in most proteins. Under oxidizing conditions, rigid disulfide bonds may additionally form between nearby cysteine residues. Simultaneously, if a stretch of amino acids cannot form such stabilizing bonds, it will instead remain in an unstructured flexible loop that does not adopt a predictable shape. These so-called intrinsically disordered proteins range in scope from linker regions between structured domains to fully unstructured proteins.

Although individual secondary structure motifs are likely quite stable, the interactions between secondary structures linked by flexible loops allow the protein to adopt additional conformations and further stabilize into a tertiary structure. In particular, the hydrophobic residues cluster together in the core of the protein, shielded by hydrophilic motifs that can favorably interact with the surrounding aqueous environment. The primary thermodynamic force driving this process is a net gain in entropy compared to the partially folded conformation. Although the completely folded and stabilized conformation has fewer degrees of freedom on its own, the exposed hydrophobic residues of the partially folded conformation create a highly structured hydration shell in the aqueous environment, as they cannot form hydrogen bonds with the water molecules. This hydration shell limits the free conformational states of the water molecules, making this state entropically unfavorable. Consequently, sequestering the hydrophobic residues from the aqueous environment confers a relative increase in entropy and decrease in free energy.

Finally, the quaternary structure arises should the folded protein assemble into a complex with other folded proteins. While some proteins do not possess quaternary structures, as they do not normally form stable complexes with other subunits, most proteins naturally form multimeric complexes and are fully functional only when the complex is assembled. For

example, the subunits of the oxygen-carrying hemoglobin complex are only functional in the assembled tetramer, thus the quaternary structure is critical to the protein's function [6, 7].

1.3 Protein Folding as a Thermodynamic Process

Similar to the progress of chemical reactions, the folding of a nascent protein into a three-dimensional structure is based on the system exploring new conformations to achieve a state of lower energy and therefore higher stability. The so-called "native" state of a protein is the conformation it occupies when it is optimally folded for stability and functionality. Remarkably, some proteins can fold spontaneously and rapidly into the native state, as demonstrated by Anfinsen *et al.* [6, 8], even when chemically denatured and renatured in the absence of protein translation machinery found *in vivo*. However, as noted by Cyrus Levinthal, the large number of degrees of freedom in an unfolded polypeptide chain produces an astronomical number of conformational states that the protein may adopt [9–11]. This so-called Levinthal's paradox predicts that random sampling of conformation states to reach the native state in such a complex system would require timescales far beyond the bounds of what is possible for biological systems. In order to efficiently adopt the native state, protein folding must therefore be a thermodynamically-guided process rather than relying on random sampling.

It should be noted that the exact mechanisms by which proteins fold from a disordered state to the final native state are hotly debated. In general, two major theories have emerged. The classical view proposes that a nascent protein is guided through a series of discrete intermediate states along a single defined pathway [12]. More recent elaborations informed by experimental investigations describe small cooperative units of folded protein motifs termed "foldons", which sequentially fold to achieve the native state (or unfold to achieve a denatured state). From a thermodynamic perspective, this model proposes ladder-like steps of conformational free energy that the protein occupies in between the disordered and native states [13, 14].

Alternatively, the energy landscape theory of protein folding proposes that it is for most proteins a nonlinear process. Within this theoretical framework, the conversion of the unstructured protein into a stably-folded state is often abstracted with an energy landscape (Figure 1.3). Here, the entire protein is funneled simultaneously as an ensemble - rather than sequentially as segments - towards states of increasingly lower free energy [15, 16]. The dynamic nature of proteins allows them to adopt momentary structural conformations - *i.e.* conformational substates. As a protein folds, it can sample nearby substates in the thermodynamic energy landscape with an overall tendency to adopt states of lower free energy, although it can overcome small energetic barriers with thermal energy fluctuations. This system explores substates of increasingly lower free energy, and the protein folds and stabilizes until it reaches a local energetic minimum from which it cannot escape [17–19].

Molecules that fold into a local energetic minimum with significant energy barriers will remain in such a stable conformation, even if that conformation is not the absolute minimum or native state. Such is the case with misfolded proteins, *i.e.* stably-folded proteins in a non-functional conformation. A non-functional protein may have improperly aligned residues in its

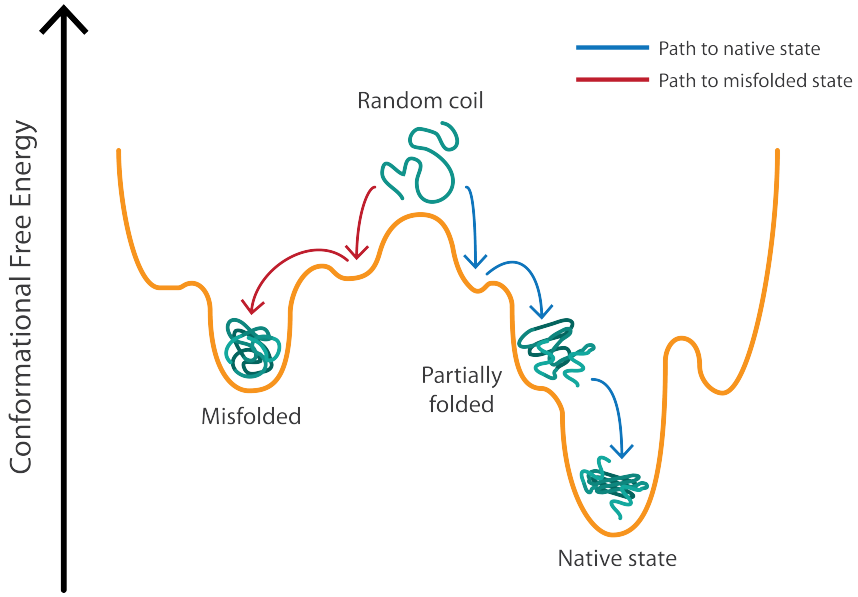


Figure 1.3: Spontaneous protein folding according to energy landscape theory. Newly-translated proteins begin as unstructured random coils. Nearby amino acid residues form secondary structures via weak noncovalent interactions, and the hydrophobic residues are shielded from the aqueous environment in the center of the protein globule. The protein explores substates of increasingly lower free energy and is able to surmount small energy barriers due to random thermal energy fluctuations. Refolding continues until the protein adopts a conformation in a significant local energy minimum from which it cannot escape. This stable conformation may be the native state (blue pathway) or a misfolded state (red pathway). Once the protein occupies a stable state, it is unlikely to escape from the energy minimum unless external energy is applied to the system.

binding site or a disordered scaffold. In an extreme case of a denatured protein, the secondary structures may be completely dismantled and the hydrophobic residues exposed such that the protein precipitates from solution. Without the application of external energy to push the protein into a conformation of higher energy, the protein cannot refold and will remain non-functional.

In nature, proteins may not reliably fold into a functional native state merely by sampling nearby substates, as the energetic barriers between the random coil state and the native state are too large. To assist the folding of such proteins, molecular chaperone proteins bind and stabilize them in various transition states, thereby guiding the nascent protein into the native state. Chaperones may also facilitate the assembly of some highly intricate multi-molecular complexes such as nucleosomes. Even under ideal conditions - *i.e.* biologically relevant temperature, salt concentration, oxidation, etc. - proteins in a nascent random coil or molten globule state are not guaranteed to spontaneously adopt the correct native conformations.

Moreover, the production of chaperone proteins by the host organism is often a response to environmental stresses. The free energy associated with a given substate depends on many factors, such as ambient temperature and salt concentration. A change in the conditions under which the protein folds will affect the relative stability of each substate and modulate the energetic landscape of folding. Significant changes to the ambient conditions may alter the folding landscape so dramatically that a nascent protein is no longer able to spontaneously fold

into the native state. Similarly, a shift to hostile conditions may cause previously folded proteins to spontaneously unfold and denature, as the native state is no longer thermodynamically favored.

To combat these effects, biological systems again make use of molecular chaperones. Various specialized chaperones stabilize the native state conformation of other proteins in order to discourage denaturation under stressful conditions. Such conditions activate the genes for these chaperones via regulatory pathways that respond to external stimuli. Consequently, the organism is able to respond directly to environmental changes that would otherwise interfere with correct protein folding. Chaperone proteins allow biological systems to carefully manage the intricate process of protein folding, especially when spontaneous formation of the native state is not possible [7, 20, 21].

According to the energy landscape theory of protein folding, the maturation of proteins into various folded conformations can be broadly described as a nonlinear process. It is probability dependent according to the thermodynamic landscape and sensitive to any factors that affect the free energy of substates within the folding landscape. However, biological systems have developed strategies to facilitate the production of functional proteins. Particularly stable substructures - *e.g.* α helices and β sheets - form in the initial stages of protein folding, followed by rapid funneling into energetic minima. Additional assistance from chaperone proteins propels nascent or misfolded proteins into the correct states. The combination of thermodynamics and coordinated intervention allows biological systems to reliably produce and organize proteins into functional folded states.

1.4 Enzyme Structure and Function

Enzymes are classically understood as protein-based, although more recently it was discovered that catalytic RNA molecules termed ”ribozymes” are capable of catalyzing biochemical reactions and play pivotal roles in regulation of gene expression [22]. Additionally, a small population of enzymes such as the ribosome are formed as a complex of protein and nucleic acids. As the results in this work concern exclusively protein-based enzymes, the following section focuses on the maturation of this class of enzymes and neglects ribozymes and hybrid enzymes. However, all enzymes - regardless of their exact composition - possess a direct connection between structure and function.

Enzymes in their folded state are often globular proteins composed of hundreds of amino acids, although only a small subset thereof directly participates in catalysis. The active site contains a multitude of residues whose coordination allows them to bind and orient the substrates. This requires precise positioning of critical residues that not only have affinity for the substrates, but additionally stimulate formation of the transition state to catalyze the reaction. There may also exist other binding sites for regulatory molecules that enhance or suppress enzyme activity by inducing conformational changes that effect the geometry of the active site. The rest of the protein acts as a scaffold to support the active site and other secondary binding sites (Figure 1.4).

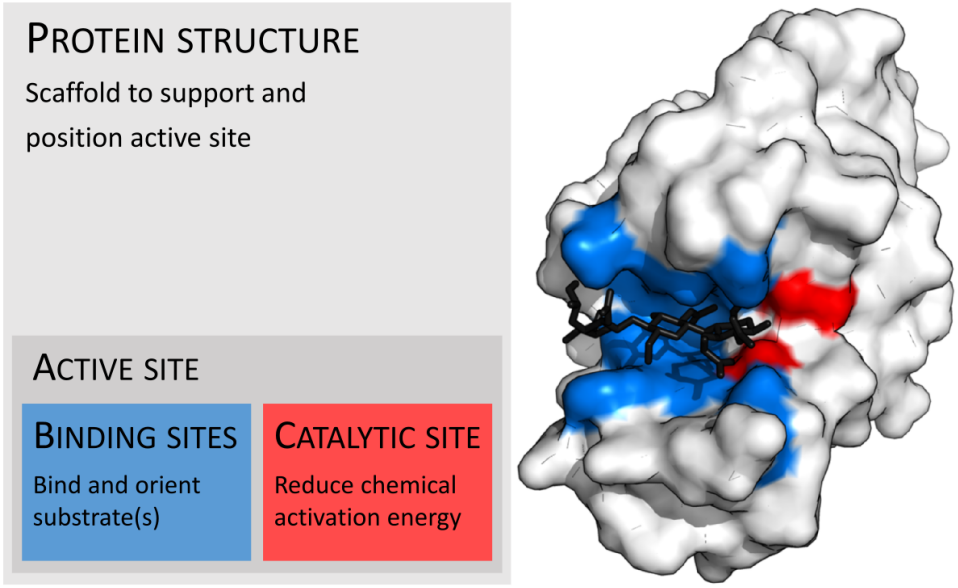


Figure 1.4: Structure of an enzyme bound to a substrate. Most of the protein structure (gray) serves as a scaffold to support active residues. The binding site (blue) contains residues that form noncovalent interactions with the substrate (black) in order to hold it in place. The active site (red) contains the catalytic residues that participate direction the chemical reaction. || Image created by Thomas Shafee and used with permission under the Creative Commons [by Thomas Shafee [CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)], from Wikimedia Commons].

Mutations in the catalytic residues will frequently completely abolish catalytic activity. If a catalytic residue is removed due to a deletion mutation, the enzyme may not form the required transient bonds with the substrate in order to stabilize the intermediate state. Similarly, if a residue in the binding site is replaced with another with a markedly different side-chain - *e.g.* a polar and charged glutamate residue replaced with a nonpolar alanine residue - the enzyme may lose affinity for the substrate entirely.

The active site of enzymes and binding sites of proteins in general is certainly crucial for their function. However, the importance of the scaffold cannot be ignored. Mutations in the scaffolding outside the catalytic residues can also have dramatic effects on enzyme function, as the scaffold is responsible for aligning the residues responsible for binding and catalysis. Even a minor change to the orientation of the catalytic residues can disrupt enzyme function if the catalytic residues can no longer correctly bind the substrate or stabilize the intermediate state [18].

However, mutations are not the only source of variable enzyme structure. Individual proteins with identical primary structures may present radically different higher-ordered structures. The momentary environmental conditions of the system influence the energy landscape. For example, elevated temperature alters the relative free energy of various conformational states, thereby making the native state energetically less stable than a denatured state for many proteins. The presence of molecular chaperones may additionally guide the protein towards particular substates. Consequently, the final folded state of a protein cannot necessarily be predicted from primary structure alone.

Furthermore, the energy landscape is itself a statistical description of the potential states a protein may occupy [17]; the thermal energy of the system confers a small degree of randomness and allows the system to explore nearby substates without the need of a specific external driving force. The energy landscape of protein folding is generally biased towards the native state as proteins dynamically sample the folding energy landscape. However, this allows for multiple folding paths for the same nascent protein. These paths may be discrete, branched, converging, or diverging, although a unified consensus on the exact mechanisms has so far not been reached by the scientific community. Regardless, a feature that is widely accepted is that for any protein there exist a multitude of final folded substates of nearly equal free energy [15]. As protein folding is probability-driven, two individual protein molecules can conceivably possess significantly different final structures - and hence different functional states.

Given that a population of folded proteins - and more specifically enzymes - will display hidden heterogeneities in their structures, the question remains of how this spectrum of folded substates affects their function. As bulk studies can inherently only report a mean statistic, they are ill-equipped to address the potential distribution of behavior. On the other hand, assessment of isolated individuals within a population enables assembly of more detailed statistical analysis beyond a simple mean value.

1.5 Insights from Single-Molecule Perspectives

Enzymes are one of the most ubiquitous and indispensable elements found in all known life. Despite their significance, much of their exact biophysical properties are yet uncharacterized. Specifically, the relationships between the amino acid or nucleic acid sequence, three-dimensional structure and dynamic behavior of the folded enzyme are avidly researched across many fields. Concerted efforts in crystallographic studies [23–29], mutant studies, molecular dynamics simulations, and *de novo* enzyme design [30, 31] have indeed elucidated a wealth of information about specific enzymes of special interest.

Unfortunately, no single technique can serve as a universal method of exploring this topic. For example, crystallographic studies provide only static structural information of variable resolution, and efforts to crystallize many enzymes such as T4 DNA Ligase from the T4 bacteriophage are frequently unsuccessful. Simultaneously, it is not a trivial task to reliably predict the broad behavior of an protein based solely on its sequence or even its folded structure. *In silico* studies such as atomic-level molecular dynamics simulations of smaller proteins are indeed able to predict their final structure in good agreement with experimentally derived structures based solely on the primary sequence of amino acids [32, 33]. However, as the conformational complexity of proteins increases quickly with polypeptide length, hardware limitations make it impractical to similarly predict the structure of of large and dynamic proteins. Moreover, computational demand is such that simulations of highly complex systems are often confined to relatively narrow sampling spaces - *e.g.* modeling the conformation of a receptor-ligand pair under force based on known crystal structures [34–36] - and over very

short timescales. In short, a comprehensive understanding of the biophysical mechanics of enzyme activity requires cooperative input from diverse disciplines and perspectives.

Within a population of enzyme molecules, the nonlinear and quite nuanced process of protein folding inevitably leads to a heterogeneous population of folded enzymes that will range from entirely misfolded to optimally folded into the native state. Considering the direct connection between enzyme structure and function, this also strongly suggests that a population of molecules will display a distribution of catalytic activity. At the same time, most standard means of assessing enzyme activity utilize bulk analysis methods, such as monitoring the rate of product buildup within a reaction volume. One primary disadvantage of such studies is that they are by nature only able to report the average behavior within an ensemble, making it impossible to easily determine the distribution of activities between molecules. Additionally, bulk studies lack the capacity to time-resolve steps within a chemical reaction. For example, a hypothetical catalytic mechanism with constant moderate speed and another mechanism with intermittent bursts of high activity and stalling could display identical average catalytic rates, even on the level of a single enzyme molecule without time-resolution.

From this dilemma emerges a niche for time-resolved single-molecule enzyme analysis [37, 38]. While bulk studies are extremely useful for characterizing the average behavior within an ensemble, a clearer picture of the distribution of activity could elucidate much about the hidden heterogeneity of specific enzymes or protein folding in general. For example, a Gaussian distribution and a bimodal distribution with identical mean values would likely be indistinguishable in bulk assays. However, their implications for energy landscapes – a central energy funnel versus two significant energetic minima, respectively – or discrete folding pathways would strongly influence our understanding of the protein folding process. Such detailed statistical descriptions are only possible through single molecule analysis, where the behavior of individuals can be isolated from a broader population. Therefore, a more complete view of proteins and enzymes demands strategies for single molecule probing.

Chapter 2
Single Molecule Probing

Overview

The isolation of the behavior of single molecules - in contrast with bulk or ensemble averaging - sheds light on the distribution of molecular behavior, and by extension, the distribution of molecular characteristics. In particular, single-molecule force spectroscopy has been an invaluable tool in identifying alternative binding modes or unfolding pathways of biomolecules, which would not be possible through ensemble techniques that can only report the average behavior. These discoveries have notable implications in free energy landscapes of complex molecular behavior of proteins - *e.g.* titin refolding. Within the field of fluorescence microscopy, observation of single molecules enables statistical analysis of molecular behavior as well as time-resolved reaction and binding steps - several parameters that are largely inaccessible in bulk measurements. Additionally, single molecule studies offer the opportunity to better identify behavior of interest within a potentially noisy system, for example by using fingerprint domains in force spectroscopy. At the frontier of nanotechnology, bottom-up directed assembly of single molecules offers incredible potential for the arrangement and study of molecules with nanometer-control of their positioning.

The study of single molecules is made possible through many elegant techniques that allow researchers to probe and manipulate matter down to the nano- and atomic-scale. The work here employs several techniques utilizing the atomic force microscope. Consequently, the following section seeks to provide a technical and theoretical foundation of these techniques as well as contextualize the work in the broader field of single-molecule studies.

2.1 AFM-based Single Molecule Force Spectroscopy

The atomic force microscope (AFM), preceded by the scanning tunneling microscope, was first experimentally implemented in 1986 by Binnig *et al.* as a tactile imaging technique with

sensitivity on the atomic scale [39]. In AFM imaging, the bending of the cantilever as it scans or taps the surface reports the local height of the sample with resolution on the order of angstroms. The measured heights of all point are then assembled into a topographical image, enabling high-resolution imaging far below the diffraction limit of traditional light-based microscopy. Moreover, the AFM can image samples under biologically-relevant conditions without the need for extra preparatory processing, unlike most forms of electron microscopy. The decades following its introduction saw extensive development and refinement of the AFM as a tool for not only atomic-scale imaging, but also force spectroscopy.

The AFM utilizes a cantilever with a very fine tip (typically a few nanometers in diameter) to mechanically probe a surface and any molecules that might be immobilized on it. A fundamental principle of cantilever-based probing is the relationship between the applied force and the corresponding induced bend in the cantilever. Assuming the cantilever acts as an ideal spring, the bend of the cantilever under force is determined by Hooke's Law:

$$F = kx \tag{2.1}$$

where the force F required to displace one end of a spring by a distance x is determined by the spring constant k. After calibrating the spring constant of a given cantilever, it is possible to calculate the force applied for a given displacement, or vice-versa.

Modern iterations of the AFM - especially when used to probe biological molecules under buffer - most often detect the cantilever deflection by a laser beam reflected from the back of the cantilever to a quadrant photodiode. As the cantilever bends due to an applied force, the laser is reflected at a slight angle, which is detected by the photodiode as increased or decreased signal in the quadrants. Hence, the force applied to the cantilever at any given moment is readily measured via the cantilever deflection (Figure 2.1a). It is also necessary to calibrate the inverse optical lever sensitivity (InvOLS) to translate the photodiode output in volts to an absolute distance in nanometers. Given a known spring constant k, this information can then be used directly in AFM imaging or used to calculate the force exerted on the cantilever via Hooke's Law.

Although initially envisioned as an imaging tool, the was AFM used to mechanically probe surface-immobilized biomolecules only a few years after its invention [40–42]. Together with similar force-transducing techniques such as magnetic tweezers [43] and optical tweezers [44, 45], diverse biomolecules were investigated for their mechanical properties during this time. Since then, the AFM has proven itself an invaluable tool for the mechanical probing of biomolecules via single-molecule force spectroscopy (SMFS) [46, 47]; specific or nonspecific adhesion of the cantilever tip to molecules on the surface enables investigation of diverse molecular characteristics, such as protein unfolding and refolding [48], receptor-ligand unbinding [36, 49], effects of mutations or ligand binding on mechanostability [50, 51], and DNA sequence-dependent stability [52, 53].

In AFM-based SMFS, the cantilever first binds to surface-immobilized molecules upon approach to the surface. As the cantilever retracts away from the surface, a force is exerted and transduced through the molecules spanning the cantilever tip and the surface, thereby

bending the lever and deflecting the laser spot proportional to the force (Figure 2.1a). The lever continues to retract, thereby increasing the applied force and stretching linker molecules and free peptide chains, until one of two events happens: (1) one or more motifs within the path of force propagation suddenly unfold and extend, *e.g.* protein unfolding, or (2) the noncovalent bonds between the cantilever and the surface-immobilized molecules completely rupture, *e.g.* a receptor-ligand complex unbinds. Either case will allow the cantilever to relax, and it will return to an undeflected state. In the case of mere extension of bound molecules, the continued retraction of the lever will again apply a force until the noncovalent bonds finally rupture and sever the force propagation pathway (Figure 2.1b-c). The cantilever is then free to probe another spot on the surface. With optimized protocols recently developed for high-throughput probing, tens of thousands of force traces are acquired in an experiment, enabling the acquisition of large data sets and exceptional statistical analysis.

Historically, AFM-based SMFS has utilized the strategy of nonspecific pulling of polyproteins. When mechanically probed and fully extended from both termini, the polyprotein - *e.g.* immunoglobulin (Ig)-like tandem repeats of titin - produces predictable sawtooth-like unfolding patterns [54, 55]. This strategy was extended to probe the unfolding of other proteins, where a domain of interest was embedded between Ig domain repeats in a recombinant chimeric construct, and the observed extra peak corresponded to the unfolding of the domain of interest [56–60]. Although this strategy does enable identification of single interactions with reasonable certainty of the pulling geometry, the overall interaction efficiency of nonspecific polyprotein pulling is low - especially when compared to more recently-developed techniques. This creates a practical difficulty in the acquisition of large data sets. Polyprotein expression also has drawbacks from a molecular biology perspective, as recombinant expression in bacteria is often limited by an upper threshold of total construct size. Considering that an employed sequence of Ig repeats alone is typically around 100 kDa in size, successful expression of the full chimeric construct is often not possible, especially for larger proteins of interest. Lastly, the unfolding forces that can be probed are limited by the strength of nonspecific cantilever adhesion, thereby prohibiting measurement of mechanically exceptionally stable motifs.

More recent efforts in AFM-based SMFS benefit from several programmable assets that increase probing specificity as well as statistical analysis. First and foremost, functionalized cantilevers that present a receptor or ligand with specific affinity for the surface-immobilized molecules is crucial for probing in well-defined and consistent geometries. As an unfunctionalized cantilever tip readily adheres indiscriminately to any surface-immobilized molecules without preference for motif or interface, likely only a very small percent of the data set represents the intended probing geometry. In contrast, utilizing a robust receptor-ligand pair split between the cantilever and surface greatly increases the reliability of mechanical probing, as receptor-ligand binding is favored over nonspecific adhesion. Exemplary receptor-ligand pairs that present high affinity and specificity include engineered antibody-peptide interactions [61, 62], streptavidin-biotin binding [49, 63–65] and cellulosomal adhesion complexes [34, 35, 66, 67].

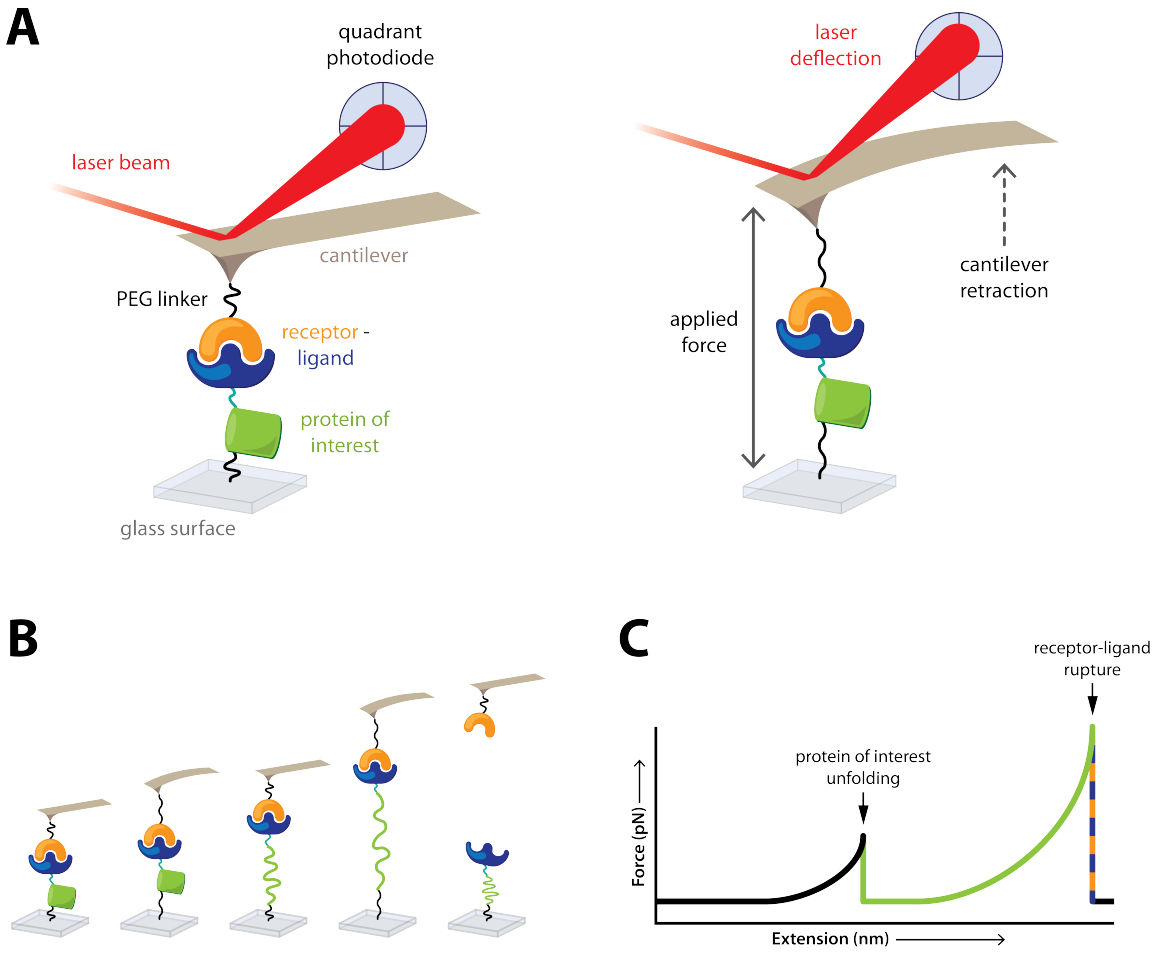


Figure 2.1: AFM-based single molecule force spectroscopy, shown here in a cartoon schematic. (a) Immobilized molecules are probed via specific handles on both the cantilever and surface and the force is measured via the laser deflection. The receptor-ligand complex (orange and blue) forms when the cantilever approaches the surface. Since there is no tension applied to the system, the cantilever is relaxed and the laser is centered on a photodiode. As the lever retracts, tension is applied to the system and the probed molecules are stretched. This in turn creates a slight bend in the cantilever and deflects the laser at an angle. Since the bend of the lever is determined by the spring constant, it is possible to calculate the force exerted on the lever and in turn molecules at any moment based on the displacement of the laser. (b) The mechanical probing unfolds internal domains in the polyprotein and finally ruptures the receptor-ligand complex. As the cantilever retracts, force propagates through the series of molecules and stretches the polymers. First the PEG linkers (black) extend, and the force increases as the cantilever pulls further away. Eventually the exerted force is large enough to unfold the domain of interest (green). As the cantilever continues to retract, the unfolded domain of interest extends with a contour length increment that corresponds to the length of the domain's peptide backbone minus the length of the folded protein. Finally, the high force pulls apart the receptor-ligand complex. The force drops to baseline levels as the cantilever is completely relaxed, at which point it probes another molecule and repeats the process. (c) Corresponding force traces highlight the stretching of PEG linkers (black trace), sudden unfolding and extension of the protein of interest (green trace), and unbinding of the receptor-ligand complex (orange and blue trace).

A key feature of SMFS is the utilization of previously characterized domains that demonstrate predictable unfolding patterns, such as Ig repeats. Often referred to as ”fingerprint” domains, these features help to identify single, specific interactions among data sets where typically 80-99 % of curves contain measurement artifacts - *i.e.* curves with no interaction, multiple interactions, or unspecific adhesion events [47]. This eliminates some of the need for ”guess-work” of which curves reflect specific interactions and which do not. While polyprotein SMFS employs the sawtooth patterns of titin Ig repeats as a fingerprint, receptor-ligand based SMFS ideally includes analogous fingerprint domains. One or more fingerprints may be included within the force propagation pathway - usually in a recombinant chimera construct with the surface- or cantilever-immobilized proteins - as a reporter for both specific adhesion and single-molecule probing. They also often demonstrate fast and reliable refolding, which is especially useful for fingerprinting cantilever-immobilized molecules over thousands of repeated probing events [34, 36, 47, 49, 68, 69]. Compared to massive Ig repeats, fingerprints typically consist of a single relatively small domain. From a molecular biology perspective, smaller and readily soluble fingerprints are highly advantageous for chimeric protein expression and stability. For example, the fourth filamin domain from *Dictyostelium discoideum* termed ddFLN4 is a convenient 11 kDa, and has demonstrated robust expression in recombinant constructs with other proteins or peptides of interest [36, 49, 70].

Owing to improved robustness and statistical yield from recent AFM-based SMFS developments, the distribution of protein behavior can be more completely understood. For example, alternative binding modes [68] and unfolding pathways [35] in the mechanically stable cellulosomal proteins were recently identified, thereby demonstrating the unique investigative value of single-molecule studies. By nature, ensemble averaging is only able to report mean behavior, while single-molecule studies enable identification of distinct behaviors within the same system. This process requires above all large amounts of usable data. In the case of AFM-based SMFS, improvements to experimental strategies - namely receptor-ligand SMFS - have indeed increased the efficiency of specific single-molecule probing. Additionally, robust strategies to identify exactly these specific single-molecule events - namely the inclusion of fingerprint domains - greatly enhances data sorting. Ultimately, such rigorous work undertaken to improve AFM-based SMFS has yielded remarkable properties of specific protein systems.

2.2 Models for Mechanical Protein Probing

The following section briefly discusses key models within force spectroscopy pertaining to polymer elasticity and rupture force distributions. Details regarding the mathematical and theoretical framework of force spectroscopy analysis can be found in several in-depth review articles and theses with elegant derivations of the mathematical models [47, 71–74].

In a very broad sense, these models seek to examine mechanically-driven biomolecular unfolding or unbinding processes within a thermodynamic framework. Protein folding and unfolding along a certain reaction coordinate can be abstracted as a free energy landscape, as the precise behavior of a peptide chain with potentially millions of atoms in variable environments

cannot be trivially predicted or calculated. In SMFS, molecules are mechanically loaded such that the free energy landscape is altered. This can increase the probability that they escape stable, folded conformations and are driven into unfolded or unbound states that are normally less likely due to large energy barriers.

The theoretical framework is founded in the seminal work of Bell [75], and was later nearly simultaneously established by Evans and Ritchie [76] and Izrailev *et al.* [77] This so-called Bell-Evans model predicts a log-linear relationship between the loading rate under which a protein is mechanically stretched and the most probable rupture force, and is commonly used in SMFS analysis. In the context of AFM-based SMFS, the cantilever behaves as a Hookean spring with a known spring constant that is used to measure the force exerted on surface-immobilized molecules, and the force grows with increasing retraction distance. Simultaneously, the surface-immobilized molecules and linkers act as entropic springs as they are extended, producing an overall non-linear increase in force with increasing distance. Assuming a constant increase in force over time $F(t)$, the Bell-Evans model provides a general expression for the experimentally measured distribution of rupture forces $p(F)$:

$$p(F) = \frac{k(F)}{\dot{F}(F)} \exp\left(- \int_0^F \frac{k(F')}{\dot{F}(F')} \mathrm{d}F' \right) \tag{2.2}$$

where $p(F)$ is derived from the force-dependent unbinding rate $k(F)$. Due to the non-linear increase in force, the loading rate is accordingly not constant during cantilever retraction at constant speed. Assuming that the system is in equilibrium at sufficiently slow retraction speeds, the loading rate just prior to an unfolding or rupture event may be calculated from the given force traces. It is further assumed that only the loading rate immediately prior to an unfolding or rupture event affects the most probably rupture force. These assumptions are crucial to the analysis of nonlinearly elastic force transducers - *e.g.* PEG linkers or peptides - which would otherwise be unaddressable by the Bell-Evans model in constant speed mode.

Polymer elasticity models are often employed to describe the extension of semi-flexible polymers. Within the context of SMFS, these models are used to extract contour length information from the raw end-to-end distance between cantilever and surface. In particular, the worm-like chain (WLC) model accurately expresses the distribution of conformational extension states of biological polymers, such as DNA and proteins, as a function of the polymer's persistence length. However, the WLC model has no analytical solution for the extension of the polymer as a function of force. An approximation of the WLC model [78] is used to describe the mechanical stretching of an unfolded protein domain reasonably well when subjected to forces of less than approximately 200 pN. Within the framework of the WLC model, the free contour length parameter L_c - *i.e.* the length of the polypeptide along the contour of the polymer chain

in a given folded state - can be solved as a function of the persistence length L_p, extension x and force F:

$$\frac{FL_p}{k_BT} = \frac{1}{4}\left(1 - \frac{x}{L_c}\right)^{-2} - \frac{1}{4} + \frac{x}{L_c} \tag{2.3}$$

As the validity of the simple WLC model has an upper force threshold, enhanced models such as the freely rotating chain model [79] incorporate other parameters in an effort to identify a unifying model that is valid across a large range of forces.

2.3 Single-Molecule Fluorescence

The property of fluorescence - which is found both in nature as well as synthetic molecules - occurs when a molecule absorbs the energy from an incoming photon of a particular wavelength and emits a photon of a higher wavelength as it relaxes back to a ground state. This spectral shift occurs as a result of non-radiative energetic transitions before photon emission, which cause the emitted photon to have a lower energy (and therefore higher wavelength) than the initial photon whose energy was absorbed (Figure 2.2).

Fluorescence microscopy relies on this spectral shift to selectively image molecules of particular spectral properties, and enables detection of diverse biomolecules of interest. As the excitation and emission paths are filtered to permit only a narrow range of wavelengths, typically only the fluorescent molecules of interest are readily detected. These properties grant fluorescence microscopy several advantages when compared to other microscopy techniques. For example, fluorescence microscopy enables detection of structures that are otherwise invisible in conventional microscopy, such as single molecules. Simultaneously, fluorescence microscopy enables high-resolution imaging of active ”live” samples under native conditions, in contrast to the highly involved fixation steps required for electron microscopy. The indispensable usefulness of fluorescence microscopy has led to the development of numerous specialized techniques with applications from cell biology to nanotechnology [80–83].

Total internal reflection fluorescence (TIRF) microscopy emerged within the field of fluorescence microscopy [84], and provided a strategy to capture images with extremely high signal-to-noise ratio in a narrow excitation volume. This microscopy technique fundamentally relies on the refractive and reflective properties of media interfaces to create a narrow excitation field via an evanescent wave (Figure 2.3). In a typical TIRF microscope, a coverglass slide rests above an objective with an interface of immersion oil in between. The sample on the coverglass slide contains an aqueous solution of low refractive index and fluorescent molecules. The sample is then illuminated with an incoming excitation beam, usually from a diode laser. The beam approaches the sample at an incidence angle θ, which is measured with respect to the normal at the refractive boundary. For an interface of two media with different refractive

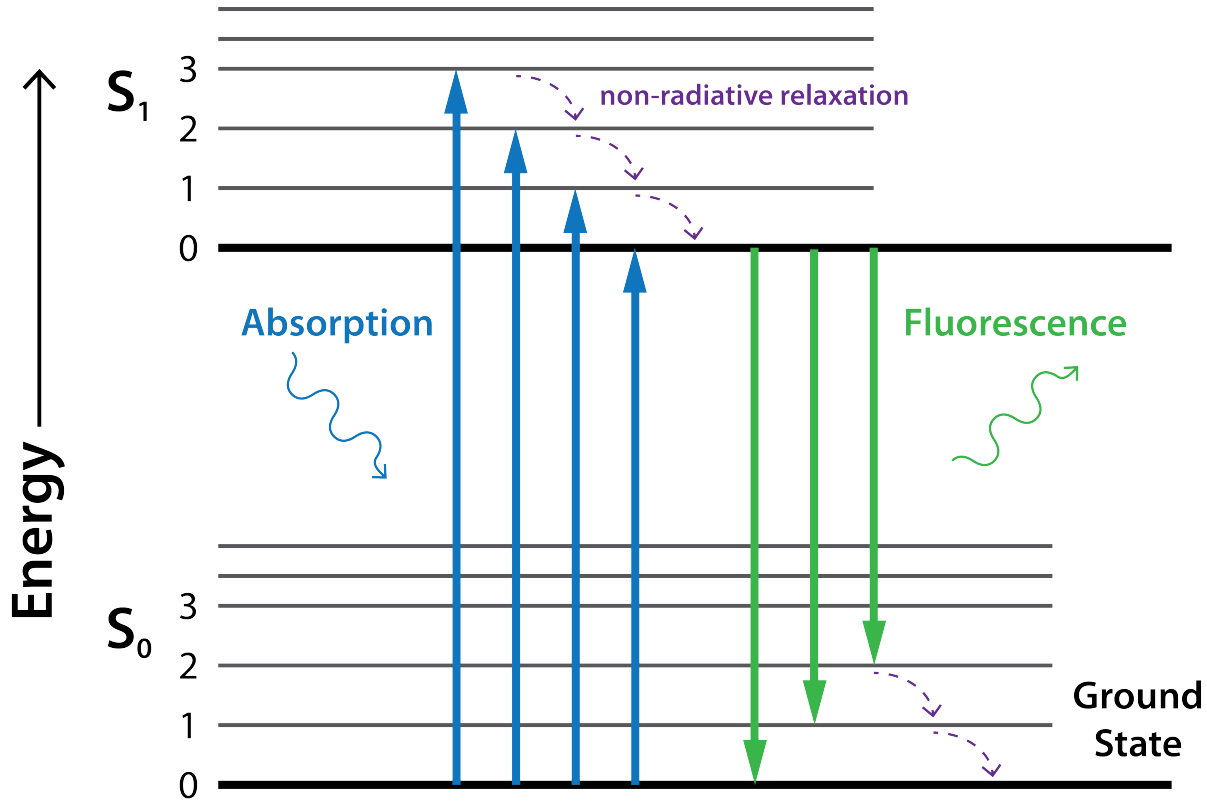


Figure 2.2: Jablonski diagram illustrating the principle of fluorescence. The energy from an incoming photon is absorbed, thereby causing the molecule to reach an excited electronic state of higher energy. As the molecule relaxes, some of the energy is dissipated through nonradiative transitions. The photon that is emitted through a fluorescent radiative transition consequently has a lower energy (and therefore higher wavelength) than the initial absorbed photon.

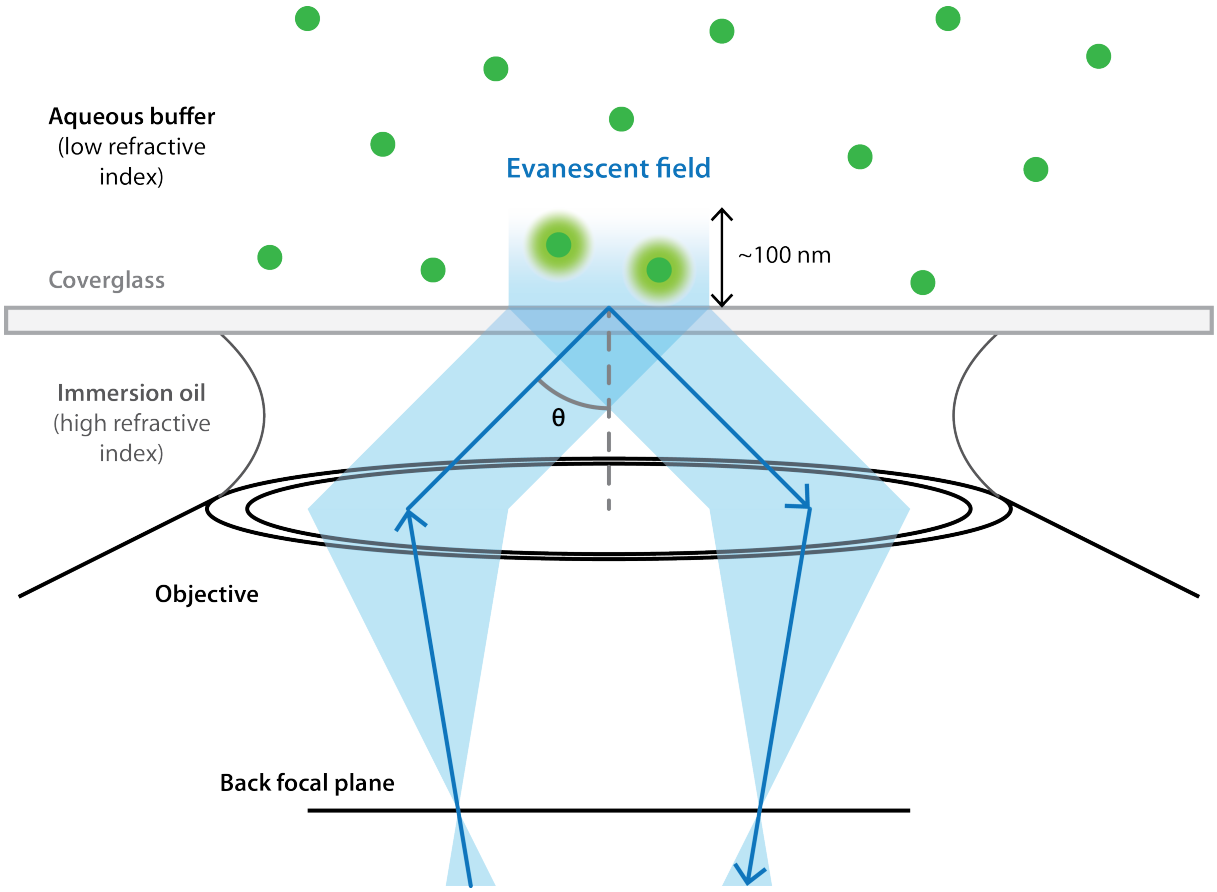


Figure 2.3: Principle of TIRF microscopy. The excitation beam approaches the sample at an incidence angle θ. The excitation beam can achieve total internal reflection within the coverglass slide when θ passes a threshold of a critical angle. As the aqueous sample buffer has a significantly lower refractive index than the immersion oil, the excitation beam cannot pass through the sample-glass interface at this angle, and is instead reflected back to the objective. A subsequent evanescent wave transmits beyond the sample-glass interface and decays exponentially. In a typical TIRF setup, this enables excitation of fluorescent molecules within a very narrow field of approximately 100 nm above the coverglass slide and minimizes background signal from molecules in solution.

indices n_1 and n_2, the angles of incidence θ_1 and refraction θ_2 of light that passes through the interface is determined by Snell's law:

$$n_1 sin(\theta_1) = n_2 sin(\theta_2) \tag{2.4}$$

If $n_1 > n_2$ and the incidence angle θ_1 is increased, the refractive angle θ_2 increases more quickly. The critical angle θ_c is defined as the incident angle when the refractive angle is 90°, at which point no light is transmitted through the interface. By solving for θ_1 when $\theta_2 = 90°$, Snell's law can be rearranged to give the critical angle:

$$\theta_1 = \theta_c = arcsin\left(\frac{n_2}{n_1}\right) \tag{2.5}$$

If the $\theta > \theta_c$, the excitation beam can achieve total internal reflection within the coverglass slide. The excitation beam is able to readily pass through the oil-glass interface at this angle, as the oil has a high refractive index matched to the refractive indices of the coverglass slip and microscope objective. However, the low refractive index of the aqueous solution prevents propagation of the excitation beam through the sample-glass interface, and the beam is instead reflected back to the objective. A subsequent evanescent wave transmits beyond the sample-glass interface, which is able to excite fluorescent molecules within its range. As the evanescent wave decays exponentially from its plane of origin, only molecules within a very limited field of approximately 100 nm from the coverglass surface are efficiently excited. This presents a major advantage of minimizing background signal from fluorophores in solution, thereby enabling precise imaging of molecules within a narrow excitation volume. This technique has been a valuable asset in systems that normally suffer from high background signal, including whole-cell imaging [80, 81, 84, 85] and single-molecule fluorescence studies [82, 86, 87].

Förster resonance energy transfer (FRET) occurs between two nearby fluorescent molecules - a donor and an acceptor (Figure 2.4). The donor molecule is first excited by incoming photons. From this excited state, the donor molecule then relaxes by transferring energy to a nearby acceptor molecule via nonradiative dipole-dipole coupling, rather than emitting the energy as a photon. The acceptor molecule is then excited to a higher energy state and releases a photon of higher wavelength as it relaxes. Importantly, as dipole-dipole coupling potential decays exponentially with distance between the donor and acceptor molecules, FRET can only efficiently occur between molecules within approximately 10 nm or less of each other and in an amenable orientation. Additionally, the emission spectrum of the donor and absorption spectrum of the acceptor must have significant overlap in order to produce efficient FRET [82, 88]. The strict requirements to achieve FRET of the donor and acceptor pair make this detection method highly selective for specific fluorescent molecules of interest within close proximity of each other, enabling FRET to be used as a kind of molecular ruler for systems that may adopt several conformational states [89, 90].

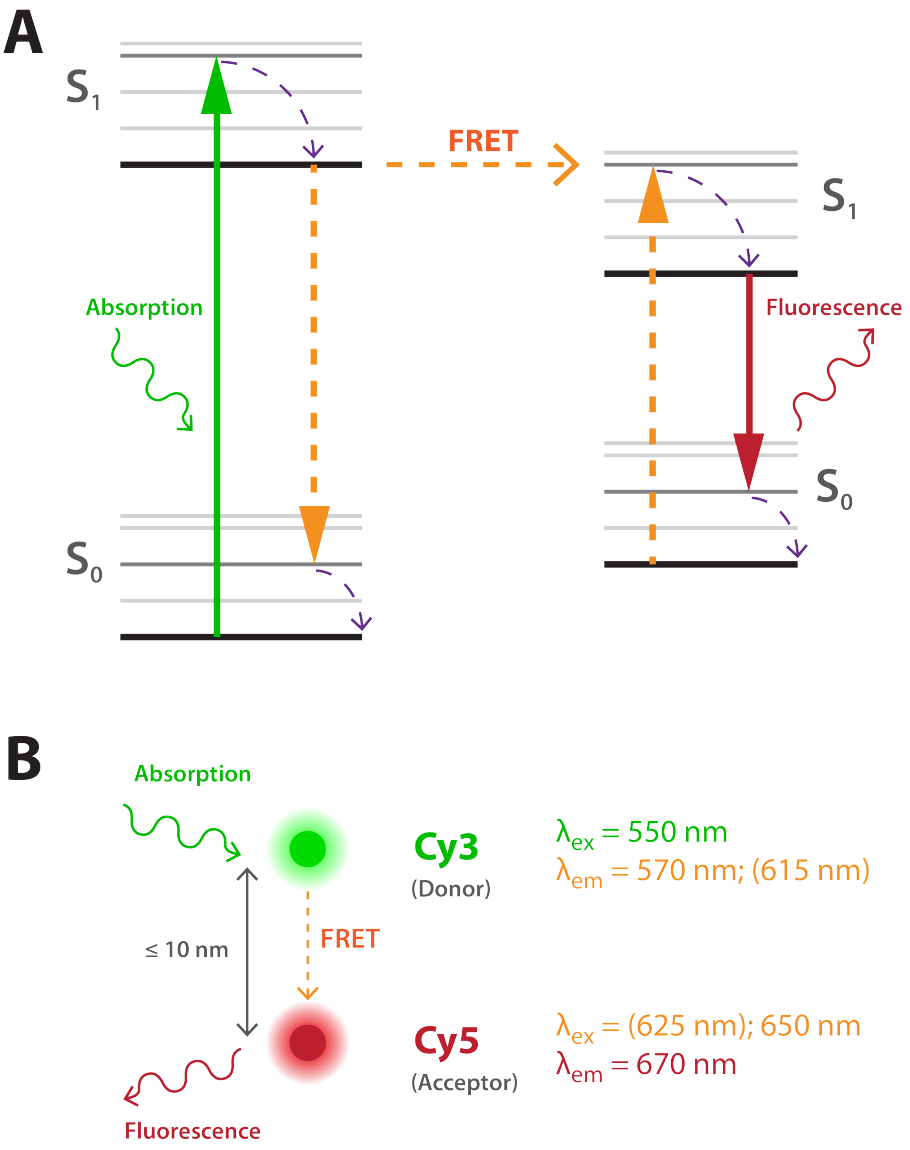


Figure 2.4: Jablonski diagram illustrating the principle of FRET-based fluorescence. (a) FRET occurs as nonradiative energy transfer via dipole-dipole coupling between two nearby fluorescent molecules. A donor molecule first absorbs incoming light and is excited to a higher energy state. As the donor molecule relaxes, it transfers the energy to a nearby acceptor molecule rather than emitting another photon. The excited donor molecule then fluoresces as it relaxes, releasing a photon of significantly higher wavelength than the initially absorbed photon. (b) A commonly implemented FRET pair consists of Cy3 as the donor and Cy5 as the acceptor, whose respective emission and absorption spectra have sufficient overlap for FRET. As dipole-dipole coupling potential decays exponentially with distance between the donor and acceptor, FRET is limited to molecules within approximately 10 nm or less of each other.

2.4 Single-Molecule Cut-and-Paste

Bottom-up assembly of biological structures is a key pursuit of nanoscience. With alluring potential applications in diverse fields across basic research, industry and medicine, many elegant solutions to this challenge have been devised in recent years.

For example, the creation of engineered protein modules demonstrates a means of directing molecular segregation via cell membrane-like structures [91]. Arguably the most widely-implemented programmable biological assembly technique to date is DNA origami [92–94]. Originally conceived of in the 1980's [95] and made readily accessible by modern DNA synthesis, DNA origami is the self-assembly of DNA structures in a variety of shapes and sizes based on specifically designed complementary single DNA strands. DNA origami has long been established as a proof-of-principle, primarily in the creation of remarkably intricate structures. However, recent efforts within this field have focused on developing DNA origami as a functional tool for *e.g.* activated drug delivery [96] and as a scaffold to arrange molecules of interest for spatial organization studies [94, 97].

In contrast to self-assembly of nanostructures, directed assembly offers the possibility of absolute spatial control of molecules of interest. Single-molecule cut-and-paste (SMC&P) - which combines bottom-up assembly with remarkable control on the level of single molecules - utilizes the AFM as a tool for molecular positioning rather than solely as a tool for force spectroscopy. In principle, molecules that are noncovalently immobilized in a so-called "depot" area on a glass surface are first picked up by the cantilever. The cantilever then travels to a different area on the glass surface and the molecules of interest are deposited in a so-called "target" area. As the precision of placement is determined by instrument, piezo stages with accuracy on the order of 10 nm are implemented for lateral positioning control. This process is repeated potentially hundreds of times in a row, with the cantilever picking up fresh molecules from the depot and placing them in new locations in the target (Figure 2.5). The resulting pattern of arranged transfer construct molecules is imaged with TIRF microscopy (Figure 2.6), thereby confirming robust and precise placement of molecules on a surface.

This system presents several strict prerequisites. First and foremost, the transfer construct must be specifically immobilized to the surface and cantilever in order to control the pulling geometry as well as most probable rupture force. Secondly, this system critically relies on a programmable rupture force hierarchy. The most probable rupture force between the transfer construct and the depot area (F_D), cantilever handle (F_C) and target area (F_T) must be tuned such that $F_D < F_C < F_T$. This enables the probabilistic yet reliable sequential transfer of the construct from depot to cantilever to target. Lastly, the transfer construct requires a fluorescent label, *e.g.* a synthetic fluorophore, or domain, *e.g.* GFP, in order to image the pattern with TIRF microscopy, although this is independent of the transfer process.

SMC&P has previously been successfully carried out with a variety of immobilization and handling strategies and fluorescent imaging strategies. While the first published SMC&P experiment was relatively simple in design with fluorescently labeled DNA [98], SMC&P schemes have grown to include functional DNA-RNA hybrid aptamer assembly [1], nanoparticle self-

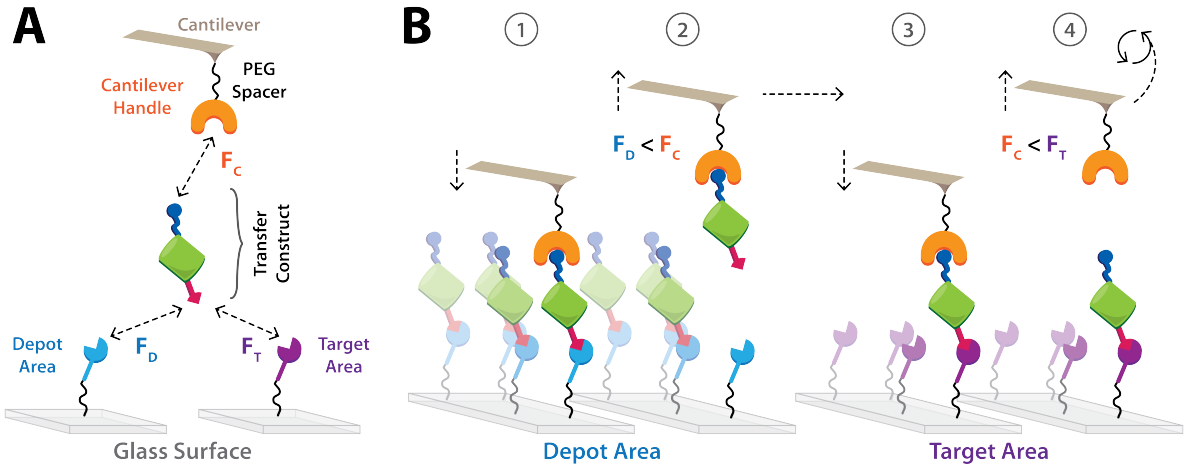


Figure 2.5: Schematic of the principle of SMC&P. (a) A transfer construct is designed and expressed with tags or domains that specifically bind to a cantilever handle as well as surface-immobilized molecules in the depot and target areas. The most probable rupture forces between the transfer construct and the depot area (F_D), cantilever handle (F_C) and target area (F_T) are established experimentally via SMFS. (b) A force hierarchy governs the repeatable transfer of molecules in SMC&P. The force hierarchy is tuned such that $F_D < F_C < F_T$. The cantilever tip approaches the depot area, which allows the cantilever handle to strongly bind a surface-immobilized transfer construct (1). Retraction of the cantilever tip pulls a transfer construct from the surface by rupturing its bond with the depot area (2). The loaded cantilever with transfer construct cargo travels to the target area and approaches the surface, allowing the noncovalent bond with the target area to form (3). The cantilever again retracts, rupturing the bond between the cantilever handle and the transfer construct. Having deposited a molecule with high precision in the target area, the unloaded cantilever travels back to the depot area to start a new SMC&P cycle (4).

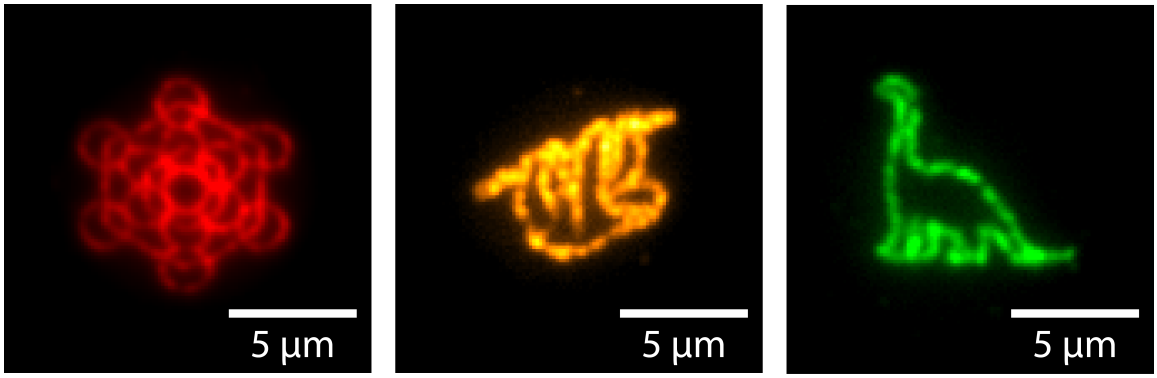


Figure 2.6: Fluorescently labeled DNA arranged by SMC&P. Single-stranded DNA modified with the fluorophore Atto-647N was specifically immobilized on the surface and handled by the cantilever via complementary DNA sequences. The molecules were arranged over hundreds of transfer cycles in the pattern of the Kavli Foundation logo (red), a sloth hanging from a branch (orange), and a dinosaur (green). The images are composed of 20 stacked frames from TIRF microscopy acquisition (0.12 s exposure time at $\approx 10\,\mathrm{W/cm^2}$ with red laser excitation.

assembly [99], protein-DNA hybrid assembly [86, 100, 101], and assembly within nanoapertures of zero mode waveguides [2]. Most recently, SMC&P schemes with a new cantilever handling strategy [102] and an entirely DNA-free SMC&P system [103] were successfully implemented, which are discussed at length in Chapters 7 and 8, respectively.

As was initially envisioned for SMC&P, a paramount goal of the development of this technique is the incorporation of enzymes in the transfer construct, and the subsequent live imaging of catalytic turnover via fluorescent products. Similar to SMFS, SMC&P offers a unique opportunity to observe single-enzyme behavior and assess the distribution of behavior in contrast with ensemble averaging. Moreover, the precise spatial positioning of SMC&P enables assembly of enzyme networks in various geometries and two-dimensional densities.

2.5 Microfluidics

When working with extremely small volumes or small surface areas, it is often advantageous to design microfluidic systems to facilitate proper surface functionalization with high precision. A common strategy is to affix a polydimethylsiloxane (PDMS) chip to a glass surface and use a pump or vacuum system to apply molecules of interest to the sample channels. Extremely complex microfluidic chips have been designed that also include a series of control channels that close or open valves in the sample channels [104].

Such a strategy has proven exceptionally useful in SMC&P (Figure 2.7). As a high flux of buffer can be pumped through the channels with minimal user intervention, the surface can be both better and more easily washed after applying molecules of interest. This is of particular interest especially in the case of fluorescent molecules that create increasing background signal buildup over time from initial nonspecific adhesion to the depot area followed by dissociation and re-association in the target area. Additionally, by designing the depot and target areas within a relatively small distance of each other, the cantilever travel distance is minimized and any effects of a tilted surface on z-piezo travel distance are limited. Moreover, in the years when SMC&P was first developed, the lateral piezo stage of the hybrid TIRF/AFM setup had a limited range of 100 µm, therefore necessitating a means of functionalizing the depot and target areas within a few micrometers of each other.

However, as the technical aspects of the TIRF/AFM setup have been improved, the lateral piezo stage range has been enhanced to now travel distances of 1 - 2 cm, allowing a much greater surface area to be sampled. This in turn allows for a much thicker PDMS barrier between the channels. With the increased surface area between the barrier and the glass surface, the likelihood of a leak between the channels is reduced, thereby improving experimental throughput of this technique.

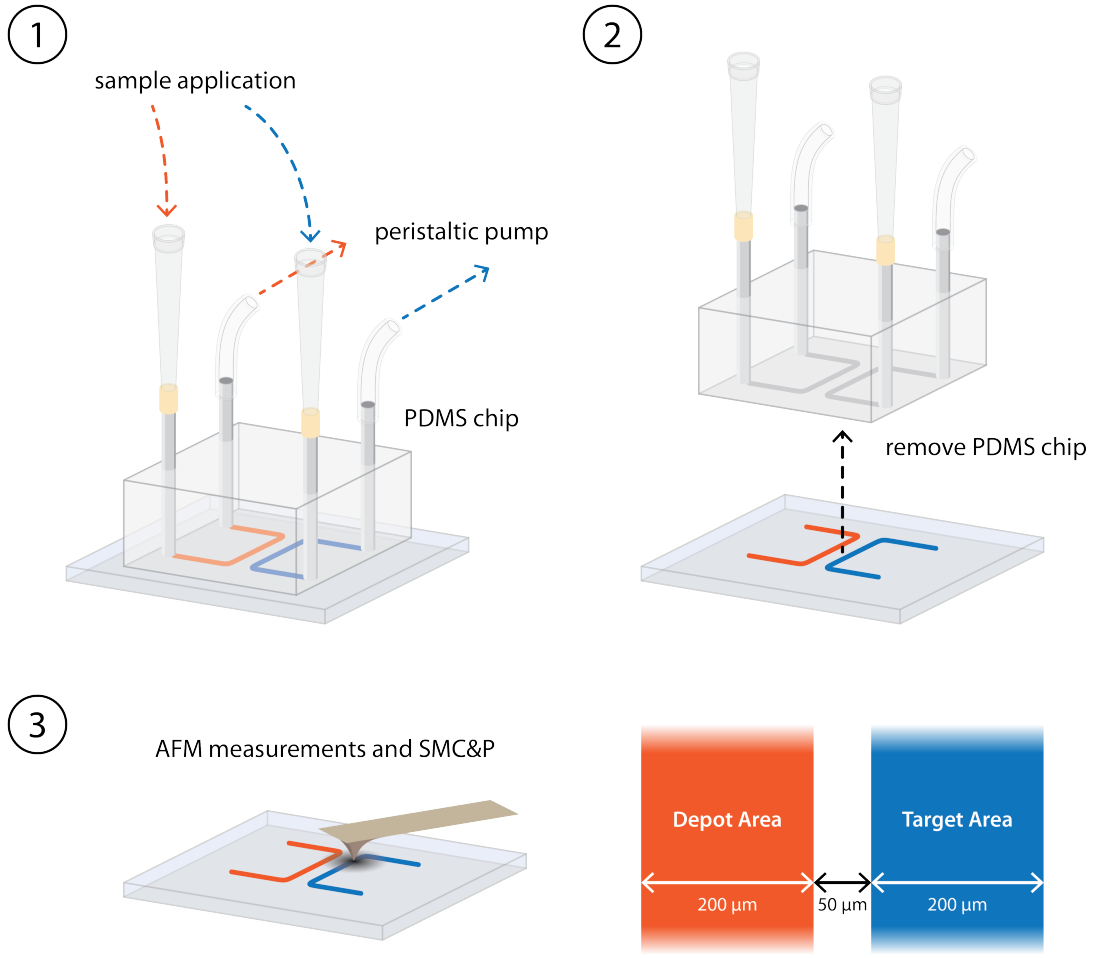


Figure 2.7: Surface preparation with a microfluidic system. Channels are assembled on the surface to minimize the travel distance of the AFM head. A PDMS chip with a reverse-mold of the channel design is affixed to the glass surface with inlet- and outlet-cannulas inserted to intersect with the channels. Additionally, the inlet cannulas are connected to fresh pipette tips via flexible tubing, and the outlet cannulas are connected to a peristaltic pump via flexible tubing. Various liquid samples containing the depot or target anchor molecules, transfer construct, passivation agents, or wash buffer are applied to the inlet cannulas and pumped through the channels with the peristaltic pump system. After the surface is prepared, the PDMS system is removed and the surface submerged in buffer. The AFM cantilever is free to interact with both channels, which are typically on the order of 200 µm wide with a 50 µm barrier in between.

2.6 Streptavidin and *Strep*-Tactin in SMFS

The homotetrameric, tetravalent protein streptavidin (SA) from the bacterium *Streptomyces avidinii* binds with remarkably high affinity to biotin, a small molecule commonly found in biological systems, with measured dissociation constant K_d on the order of 1×10^{-14} M. As one of the longest-lived noncovalent bonds found in nature to date, this system quickly became a subject of great interest as both a model system of noncovalent interactions as well as a practical tool in bio-related fields [105, 106].

Moreover, extensive research and development of the SA-biotin system has yielded new ligands and engineered variants of SA. The Strep-Tag II (SII) tag (WSHPQFEK) - an improved version of its predecessor *Strep*-Tag - was developed as a short peptide ligand of SA [107] and is commonly used as a purification tag when fused to a protein of interest [108]. *Strep*-Tactin (ST), an engineered variant of SA with improved affinity for SII, was likewise developed to improve utility of this system [109]. From an investigative perspective, the *Strep*-family of interactions has been probed thoroughly using many techniques in order to understand how the two molecules form such a strong binding pair. Crystallographic analysis [110, 111], molecular dynamics simulations [112, 113], and single-molecule force spectroscopy [49, 63–65] have all readily been undertaken in order to elucidate the biophysical principles behind this system.

Despite the prevalence and robustness of SA and ST, the tetravalency exhibited by both of these complexes is disadvantageous in the context of single-molecule probing. For example, in AFM-based SMFS with wildtype SA or ST immobilized to the cantilever, there is a high probability to probe multiple molecules on the surface. Hence, only a very small fraction of the resulting data likely represents truly single-molecule interactions, thereby limiting the amount of usable data. Thus, optimized application of the *Strep*-family of molecules in single-molecule studies requires improved stoichiometric control of ligand binding.

A heterotetrameric, monovalent version of *Strep*-Tactin (mST) was introduced and implemented in AFM-based force spectroscopy by Baumann *et al.*. As the single functional subunit of mST is also responsible for covalent immobilization via a reactive terminal cysteine, mST enabled single-molecule probing in a well-defined geometry [69]. Interestingly, it was also found that the specific pulling geometry of SII strongly affected the rupture force distribution; SII fused to the N-terminus of a domain of interest resulted in significantly lower forces than a C-terminally fused tag. In this work, an N-terminally fused SII enabled specific handling in a lower force regime with a mST-functionalized cantilever. Similarly, a heterotetrameric, monovalent variant of streptavidin (mSA) that retained its high affinity for biotin was previously developed by Howarth *et al.*[114] This heterotetrameric variant was subsequently implemented in single-molecule force spectroscopy by Sedlak *et al.* [49] where it was discovered that the pulling geometry of mSA strongly influences the rupture force distribution [113]. Reminiscent of the behavior of SII, N-terminally anchored mSA (N-mSA) unbinds from biotin at significantly lower forces than C-terminally anchored mSA (C-mSA). Further development of streptavidin produced a stable monomeric, monovalent variant (mcSA2) [115]. Bauer et al. subsequently

employed a fusion of mcSA2 and ddFLN4 in AFM-SMFS, where the ddFLN4 motif served both as a solubility enhancer and a fingerprint domain.

In addition to SMFS-focused AFM applications, monovalent *Strep*-proteins have also lent themselves to SMC&P for precise positioning of molecules of interest on a surface. mST was used as a cantilever handle for SII-tagged transfer constructs [102], and the geometry-dependent rupture forces of mSA were utilized to form a low-force N-mSA depot area and a high-force C-mSA target area [103].

Chapter 3
Enzymes of Interest

Overview

The work presented here concerns itself with select enzymes of interest. To that end, this section seeks to provide pertinent background information to contextualize these enzymes and why they are broadly of interest for researchers.

In general, the selected enzymes have been already extensively investigated from many perspectives, and have many shared advantageous characteristics. Primarily, the exhaustive volume of research conducted previously helped to inform proper protocols for expression, handling and testing. The following enzymes are also monomeric and do not require special molecular chaperones for expression in B-strains of *Escherichia coli*. Furthermore, all enzymes participate in reactions that have fluorescent or luminescent products - either through their fundamental function as with luciferases or by design as with enzymes that could (but do not necessarily) interact with fluorescently-labeled DNA. This is essential to TIRF-based readout of SMC&P.

3.1 *Pfu* DNA Polymerase

All organisms with the exception of retroviruses store their genetic information in the form of DNA. Genome replication is a highly coordinated process involving dozens of enzymes, each with their specific roles to play in copying and repairing DNA.

During cell replication, the entire genome must be efficiently replicated with very low error rate. In both bacteria and eukaryotes, this task is typically split between several iterations of DNA polymerase - the enzyme primarily responsible for producing a new strand of DNA based on a template strand. This process occurs simultaneously for both strands in the double-helix, resulting in two helices each composed of a template and a newly replicated daughter strand.

Importantly, all known polymerases require an RNA or DNA primer in order to synthesize a new DNA strand from a template. In cells, this is accomplished by a primase, a sort of *de novo* polymerase that is incredibly slow and error-prone. Primase first creates a short RNA primer. After it dissociates, an elongation-specializing DNA polymerase is then able to begin rapid and accurate DNA replication from this primer sequence. The RNA primers are removed, leaving gaps in the daughter strand. A repair-specializing DNA polymerase fills in the gaps, and DNA ligase repairs the nicks, resulting in a double-stranded helix.

In addition to DNA replication, polymerases often possess exonuclease activity. This ability to excise the terminal nucleotides of a DNA strand - especially in the 3' $\rightarrow$ 5' direction - is crucial for proofreading, as all polymerases have some associated error rate [7].

Research into DNA polymerases is of great interest in biotechnology, especially with regards to advances in polymerase chain reaction (PCR) strategies. Originally developed in 1983, PCR has become one of the powerful and ubiquitously-used techniques in molecular biology [116]. Using a thermostable DNA polymerase, stretches of DNA may be amplified exponentially in an *in vitro* environment with only the addition of free nucleotides and short single-stranded DNA primers. The temperature-dependent melting and annealing of DNA is used to accomplish what is normally carried out by DNA helicase and DNA primase *in vivo*, and the thermostability of the DNA polymerase is suited to catalyzing the reaction under the elevated temperatures.

One of the most prominent wildtype thermostable DNA polymerases that has been widely implemented in PCR and other DNA replication techniques is *Pfu* DNA polymerase (*Pfu* DNAP) from the thermophilic archaeon species *Pyrococcus furiosus*. Compared to similar thermostable polymerases such as *Taq* DNA polymerase, *Pfu* DNAP has the lowest measured error rate, making it ideally suited to exactly amplifying DNA of interest [117].

DNA polymerases have undergone extensive mutagenesis and optimization research, and several biotechnology companies offer proprietary novel polymerases composed of domains from disparate sources that are exceptionally well-suited for PCR (*e.g.* Phusion and Q5 DNA Polymerase PCR systems). Similarly, there is great interest in developing DNA polymerases that can better tolerate the incorporation of non-canonical nucleotides. This is especially important in applications of DNA labeling with modified bases.

Recently, a variant of *Pfu* DNAP that is capable of incorporating specific labeled nucleotides was developed using rigorous directed evolution [118]. *Pfu* DNAP was strategically evolved using compartmentalized self-replication - a specialized self-evolution developed for polymerases. The polymerase amplifies its own sequence, which is then transcribed and translated *in vitro*, and the newly-translated enzyme contributes further to the sequence amplification. As the reaction volume is increasingly shifted to target conditions, only mutants that are functional under those conditions will replicate their own sequences to continue DNA amplification and protein expression. Thereby, it is possible to select for polymerase mutants that are functional under the desired conditions. In this particular case, canonical unlabeled nucleotide dCTP was substituted with Cy3- and Cy5-modified dCTP - two substrates that wildtype *Pfu* DNAP is unable to incorporate.

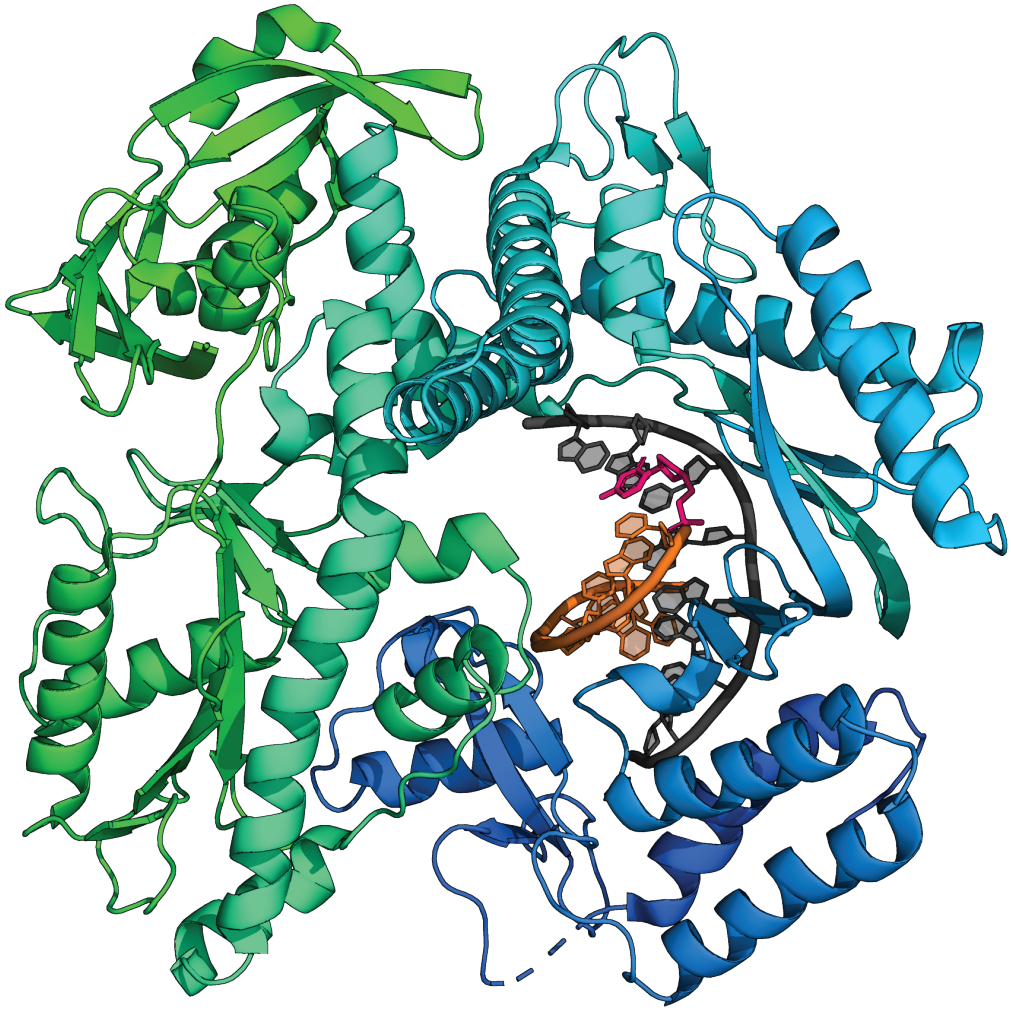


Figure 3.1: Structure of evolved *Pfu* DNA polymerase E10. The polymerase (green and blue spectrum) catalyzes DNA replication in the 5' → 3' direction. In order to bind and replicate the template strand (dark gray), Pfu-E10 requires a double-stranded primed sequence (orange). At the 3' terminus of the growing strand, the polymerase attempts to match free nucleotides (magenta) with the next nucleotide of the template strand. When a fit is achieved, the polymerase catalyzes the formation of a phosphodiester bond of the growing DNA oligomer. This evolved mutant possess the additional ability to incorporate Cy3- and Cy5-labeled dCTP nucleotides - substrates that wildtype *Pfu* DNAP is unable to use. || Pfu-E10 evolved polymerase PDB ID: 4AIL. DOI: 10.2210/pdb4AIL/pdb. || Wynne, S., Pinheiro, V., Holliger, P., Leslie, A. G. W. Structures of an Apo and a Binary Complex of an Evolved Archeal B Family DNA Polymerase Capable of Synthesising Highly Cy-Dye Labelled DNA. *PloS ONE*, 8(8):e70892, 2013.

First, three point mutations were strategically introduced to the polymerase to disable the 3' → 5' exonuclease activity and uracil-stalling functions, thereby making the polymerase more error-prone and increasing the mutation rate of its replicants. This *Pfu* DNAP mutant then amplified its own sequence with increasing concentrations of Cy-dye labeled dCTP, and mutants were screened in two steps. The best-evolved mutant E10 (Fig 3.1) remarkably demonstrated an ability to incorporate the Cy-dye labeled variants of dCTP just as efficiently as unlabeled dCTP. Thus, Pfu-E10 generated large DNA fragments up to 1 kb with all dCTP bases substituted with Cy3- or Cy5-labeled dCTP. Moreover, it was also demonstrated that the expanded substrate palette of Pfu-E10 was limited to the Cy-dye nucleotides rather than the polymerase simply having no selectivity for its substrates [118]. The solved crystal structure of Pfu-E10 illuminates some of the modified ways in which the polymerase interacts with its DNA and nucleotide substrates [28].

Impressively, the E10 mutant contains only five additional point mutations, meaning it deviates from the wildtype *Pfu* DNAP by only eight residues. This discrepancy is an exemplary demonstration of how small changes to the primary structure of a protein can have significant effects in the folded tertiary structure and hence the protein's function.

3.2 T7 RNA Polymerase

In most organisms, genomic DNA is used as a template for RNA transcription. This task is carried out by RNA polymerases, which respond to biochemical cues and specific promoter sequences to bind a DNA template and catalyze the polymerization of an RNA oligomer in the 5' → 3' direction. This step is essential for the production of messenger RNA - which subsequently provides a template for protein translation - as well as many subclasses of more recently-discovered small RNA oligomers [7].

Bacteriophages are specialized viruses that prey on bacteria cells. They rapidly infect a host cell and take advantage of the biochemically rich environment to copy their genome and produce new virus particles. Following massive bacteriophage reproduction, the host cell may be immediately lysed in order for the newly-produced virions to continue infection of other cells and begin the cycle anew. Alternatively, the virus may remain dormant in the host cell where its genome is integrated with the host DNA or established as a plasmid [119].

A key asset that enables bacteriophages to rapidly reproduce is a repertoire of highly efficient and relatively simple replication and gene expression machinery. One enzyme in particular that has gained exceptional attention is RNA Polymerase of the T7 phage (T7 RNAP) [120]. This enzyme is responsible for catalyzing the synthesis of RNA from a double-stranded DNA template (Figure 3.2). Transcription is initiated when the polymerase binds to a specific promoter sequence. T7 RNAP then melts the two DNA strands in order to read the individual nucleotides on the template strand. The enzyme additionally binds free ribonucleotides in solution and attempts to match them with the next base in the DNA sequence. When complementary bases align, T7 RNAP catalyzes the polymerization of the nascent RNA strand with the new nucleotide, and repeats the process for each successive template base.

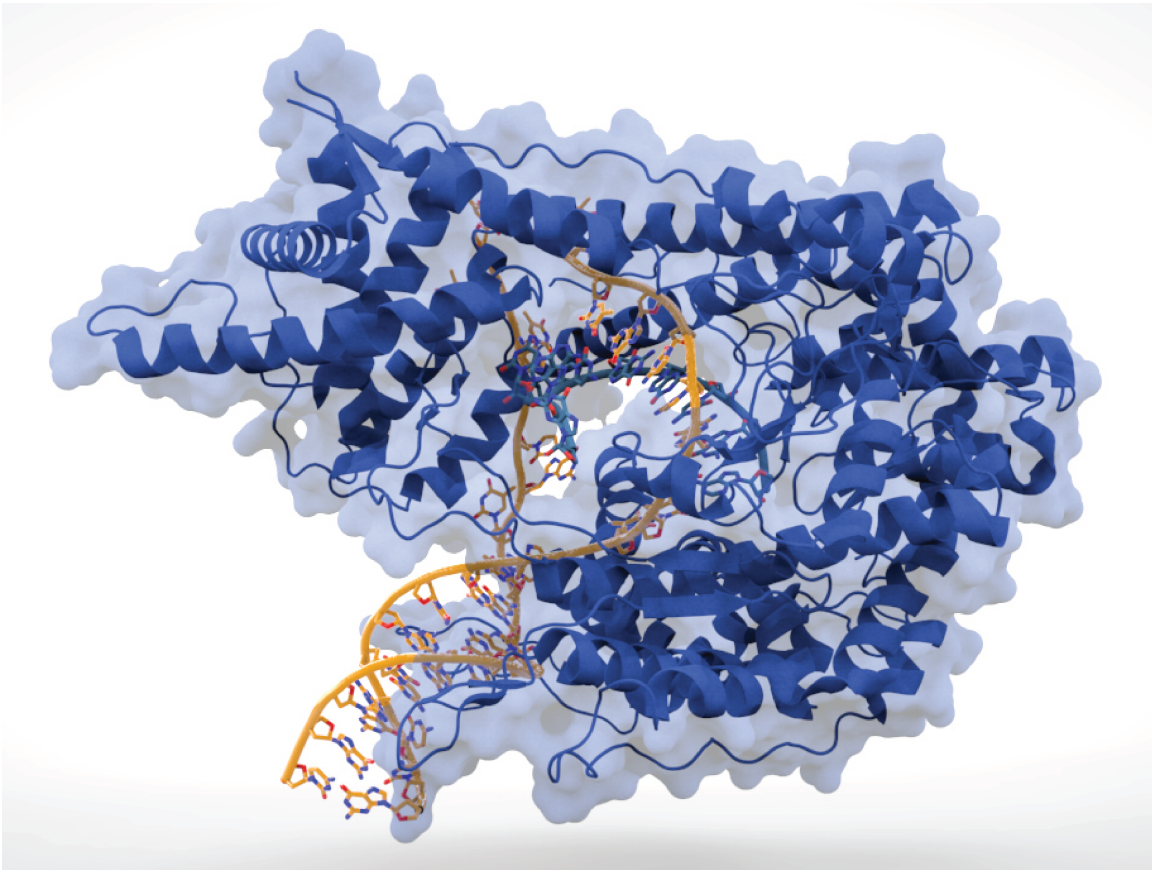


Figure 3.2: Structure of T7 RNA Polymerase bound to a DNA template. The 99 kDa T7 RNA polymerase enzyme (dark blue) binds a double-stranded DNA template (yellow) in a large binding cleft and catalyzes the production of single-stranded RNA (light blue) from the template. || Image created by Thomas Splettstoesser and used with permission under the Creative Commons [By Thomas Splettstoesser [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], from Wikimedia Commons].

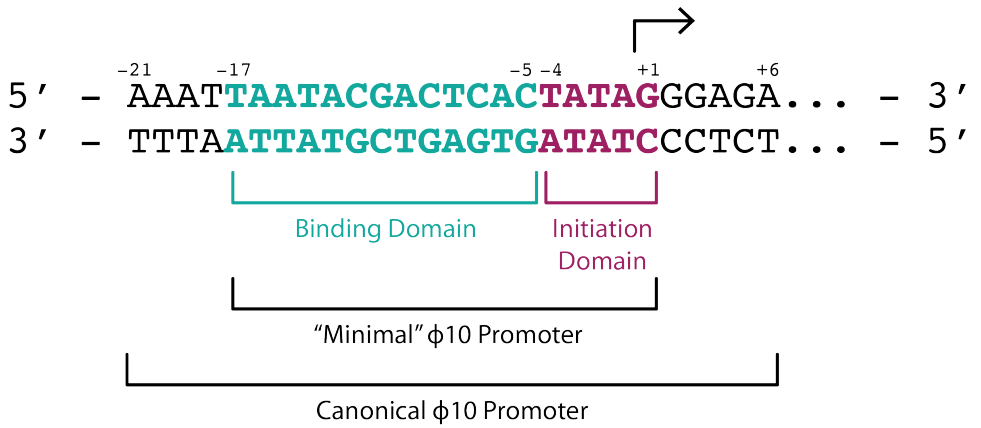


Figure 3.3: ϕ10 promoter sequence of T7 RNA Polymerase. The promoter is divided into two regions, each partially responsible for RNA transcription. The binding domain is essential for recognition by T7 RNAP. The initiation domain is responsible for commencing transcription, and T7 RNAP is especially sensitive to mutations in this domain. Importantly, there is a strong preference for a G nucleotide in the +1 position.

Soon after its discovery and isolation in 1970, T7 RNAP became a focus of studies as a model enzyme and prominently lent itself to applications in biotechnology. T7 RNAP possesses several traits that make it an exceptionally useful tool. First and foremost, T7 RNAP consists of a single 99 kDa protein that is sufficient to execute all steps of transcription. No other proteins are required for initiation, elongation or termination. This is in stark contrast to most other eukaryotic and even prokaryotic RNA polymerases that are composed of multiple subunits. Its monomeric quality negates the need for quaternary structure formation and hence greatly simplifies recombinant expression and purification compared to multimeric RNA polymerases.

Additionally, T7 RNAP exhibits unusually high specificity for a conserved promoter sequence (Figure 3.3). Even so much as a single nucleotide polymorphism in the promoter region can nearly completely suppress transcription [121]. This makes the T7 RNAP system especially advantageous for recombinant protein expression systems, as T7 RNAP requires exactly this sequence and will not engage in so-called ”leaky” transcription of unintended DNA targets. Similarly, T7 RNAP adheres very tightly to small group of transcription terminator sequences, which decreases the likelihood that a random DNA sequence within a gene of interest will prematurely provoke termination. The rigid DNA sequence requirements of T7 RNAP allow researchers exquisite control over recombinant gene expression.

Lastly, transcript elongation under T7 RNAP is especially fast and reliable. T7 RNAP is approximately five times faster than *E. coli* RNA polymerase and can synthesize RNA molecules thousands of bases long without dissociating from the DNA template and hence prematurely terminating transcription. The combination of these properties make T7 RNAP an indispensable tool for both *in vivo* and *in vitro* gene expression.

Extensive studies have investigated the biophysics behind this remarkable enzyme. Of particular note, the kinetics of promoter binding and transcription initiation from a biochemical perspective have been a major focus [122–125]. Understanding the mechanism by which T7 RNAP operates offers not only the possibility to improve the function of the enzyme, *e.g.* with

directed evolution, but also to understand RNA transcription in a broader sense with T7 RNAP as a model polymerase.

Crystallographic structural studies of T7 RNAP bound to the promoter under different conditions similarly expand an understanding of the enzyme's action [24, 27]. For example, Kennedy *et. al* examined the crystallized polymerase bound to a DNA promoter strand both with and without free nucleotides bound in the active site. Notably, nucleotide binding in the initiation phase has distinct mechanism from the elongation phase. The structures revealed a crucial DNA melting step mediated by the binding of two initiating GTP nucleotides to stabilize the open promoter in the initiation complex [27]. This observation provides an elegant explanation for a long-known feature of the T7 RNAP promoter sequence that the polymerase has a strong bias for a GTP at the +1 position of the nascent transcript [126].

Force spectroscopy has also been a helpful tool in characterizing this enzyme, especially with regards to DNA interactions. Studies utilizing magnetic tweezers [127] and optical tweezers [128] have investigated proposed mechanisms of T7 RNAP transcription initiation and found support for the "scrunching" model, where T7 RNAP exerts a force on downstream DNA to unwind the duplex and pull itself forward. The atomic force microscope was also used to probe the rupture force between T7 RNAP and DNA promoters of various sequences. The most probable rupture force appeared to be correlated with promoter strength with the Φ10 consensus sequence having the highest rupture force, thus offering a potential means of screening promoter sequences [129].

Ongoing efforts from diverse fields continuously add to a comprehensive understanding of this enzyme. These developments enhance its utility as a tool of biotechnology as well as a model enzyme. Implementation of T7 RNAP in a single-molecule analysis technique such as SMC&P would potentially bridge the gap between bulk kinetic assays and single-molecule force spectroscopy, thereby revealing a more complete understanding of this enzyme.

3.3 DNA Ligases

DNA repair is crucial to the longevity of genetic information across many generations of cell replication. Even without the effects of external damage sources such as UV light or heat, DNA routinely requires nick repair. For example, genomic DNA replication generates a series of fragmented DNA sequences in the lagging strand. These relatively short oligomers are bridged by the complementary template strand, making a nicked double-stranded DNA duplex. Although the DNA sequence may be technically complete on both strands, the nicks must be repaired to ensure duplex rigidity and stability [7].

The majority of DNA repair is handled by DNA ligase, a large class of enzymes found in all organisms that catalyze the formation of the phosphodiester bond in the DNA backbone between two adjacent nucleotides. Unlike polymerases, DNA ligases cannot add single nucleotides to a growing DNA strand. Rather, they sense nicks or breaks between DNA that is otherwise already double-stranded and join the 3'OH and 5'PO_4 termini of the two DNA fragments. Furthermore, ligases require a source of chemical energy in the form of an energy carrier;

DNA ligases from eubacteria employ oxidized nicotinamide adenine dinucleotide (NAD^+) as a high-energy cofactor, while most other species - eukaryotic, viral and archaebacterial - use adenosine triphosphate (ATP).

Although there is limited sequence homology across all ligases, there exist some characteristics that appear to be well-conserved. From a structural perspective, all ligases contain at the minimum a nucleotidyl transferase (NTase) domain and an oligonucleotide binding (OB) domain. The NTase domain comprises the primary catalytic site and contains the critical conserved lysine residue where self-adenylation occurs. The OB domain is required for DNA binding and polynucleotide recognition, as well as self-adenylation [130, 131].

Moreover, all ligases obey the same general catalytic mechanism. First, the ligase binds the energy carrier and catalyzes a self-adenylation reaction. Interestingly, this step occurs at a universally conserved lysine residue in the catalytic site. The activated enzyme then binds the damaged DNA, and transfers the adenylyl group from the active site to the 5' terminus of the nick. Finally, the ligase catalyzes the 3'OH terminus nucleophilic attack on the 5'PO_4 terminus, thereby ligating the DNA and releasing AMP [23, 26, 130–134]. The mechanism of DNA nick repair by NAD^+- ligases can be presented schematically with:

$$\begin{aligned} E + NAD^+ &\rightleftharpoons E \cdot NAD^+ \longrightarrow E_{pA} + NMN^+ \\ E_{pA} + DNA_p &\rightleftharpoons E_{pA} \cdot DNA_p \longrightarrow E \cdot DNA_{ppA} \\ E \cdot DNA_{ppA} + DNA_{OH} &\rightleftharpoons E \cdot [DNA_{ppA} + DNA_{OH}] \longrightarrow E + DNA_{ds} + AMP \end{aligned} \tag{3.1}$$

The ligase enzyme (E) binds NAD^+ and catalyzes a self-adenylation reaction to covalently join a conserved lysine residue with the adenylyl moiety (E_{pA}) and releases nicotinamide mononucleotide (NMN^+). The adenylated enzyme binds the DNA at the 5'PO_4 of the nick (DNA_p) and transfers the adenylyl group (DNA_{ppA}). Finally, the ligase catalyzes the nucleophilic attack from the 3'OH terminus of the nick (DNA_{OH}) to the adenylated terminus, thereby reforming a phosphodiester bond in the double-stranded DNA (DNA_{ds}) and releasing adenosine monophosphate (AMP).

Mechanistically, DNA repair by ATP-dependent ligases is very similar, with the exception of the first step; the ligase binds ATP and the self-adenylation reaction releases inorganic phosphate (PP_i):

$$\begin{aligned} E + ATP &\rightleftharpoons E \cdot ATP \longrightarrow E_{pA} + PP_i \\ E_{pA} + DNA_p &\rightleftharpoons E_{pA} \cdot DNA_p \longrightarrow E \cdot DNA_{ppA} \\ E \cdot DNA_{ppA} + DNA_{OH} &\rightleftharpoons E \cdot [DNA_{ppA} + DNA_{OH}] \longrightarrow E + DNA_{ds} + AMP \end{aligned} \tag{3.2}$$

Most DNA ligases are limited to sealing single-strand breaks in DNA, as with nicked DNA or sticky-overhang cohesive DNA. However, in the presence of high concentrations of

polyethylene glycol (PEG) or other crowding agents, a subset of ligases - in particular T4 DNA ligase - can ligate blunt-end DNA breaks [134, 135]. DNA ligases are also generally unable to ligate single-stranded DNA fragments, RNA strands, and DNA-RNA hybrid duplexes.

E. coli DNA Ligase (LigA) (Figure 3.4) is an NAD^+-dependent monomeric domain of 74 kDa. Genomic analysis of other species of bacteria has revealed that all species sequenced so far encode similar NAD^+-dependent ligases with a typical amino acid sequence identity of 35 - 50%. Additionally, crystal structures of DNA ligases from several other bacteria have been solved and display high degree of structural homology with LigA. Considering the similarities of LigA with other ligases analyzed so far, there is strong evidence to suggest that DNA ligases across most - if not all - species of bacteria possess comparable structures based on several conserved protein folds [132, 133].

Bacteriophage genomes encode relatively small but crucial DNA ligases. As with all ligases, their primary function is genomic DNA repair, especially during replication. Their fast activity and small size offer several advantages to laboratory research, thereby making them frequently the ligases of choice in molecular biology.

The two most commonly used bacteriophage ligases are T4 DNA ligase (T4 DNAL) and T7 DNA ligase (T7 DNAL). Both ligases readily ligate cohesive DNA ends, and T4 DNAL can also ligate blunt-ended DNA fragments in the presence of crowding agents such as PEG. These two ATP-dependent ligases are frequently employed in recombinant cloning to covalently join DNA sequences from disparate sources.

Compared to bacterial and eukaryotic ligases, bacteriophage ligases are significantly smaller - 55 kDa for T4 DNAL and 41 kDa for T7 DNAL. Additionally, both ligases are monomeric, making them amenable to recombinant expression and purification from *e.g. E. coli*.

Due to their widespread use in molecular biology and biotechnology, both of these phage ligases have been the focus of much interest from a perspective of improving their activity and expression. For example, a chimeric variant of T4 DNAL was designed to have an additional adenylate kinase domain. This extra domain provided two major advantages over wildtype T4 DNAL. First, the adenylate kinase domain catalyzed the conversion of ADP to ATP - the critically required energy carrier of the ligase. Therefore, the enzyme could ligate DNA when only ADP rather than ATP was provided in solution. Secondly, the chimera protein demonstrated significantly improved solubility and hence purification yield compared to T4 DNAL alone [136].

While the two ligases have limited amino acid sequence identity, the conserved catalytic lysine is present in both enzymes. Interestingly, to date there is no available crystal structure of T4 DNAL due to unsuccessful attempts at crystallization both with and without cofactor bound [135]. However, the crystal structure of T7 DNAL has been solved, and offers great detail into the mechanism of ligation for this enzyme [23] (Figure 3.5).

DNA ligases tend to have low tolerance for non-ideal DNA substrates, *i.e.* DNA that is mismatched, short, modified or damaged at or near the site of ligation. In a general sense, these alterations interfere with the binding of the ligase to the DNA substrate. Consequently, the ligase cannot properly orient the DNA to efficiently catalyze the phosphodiester bond formation.

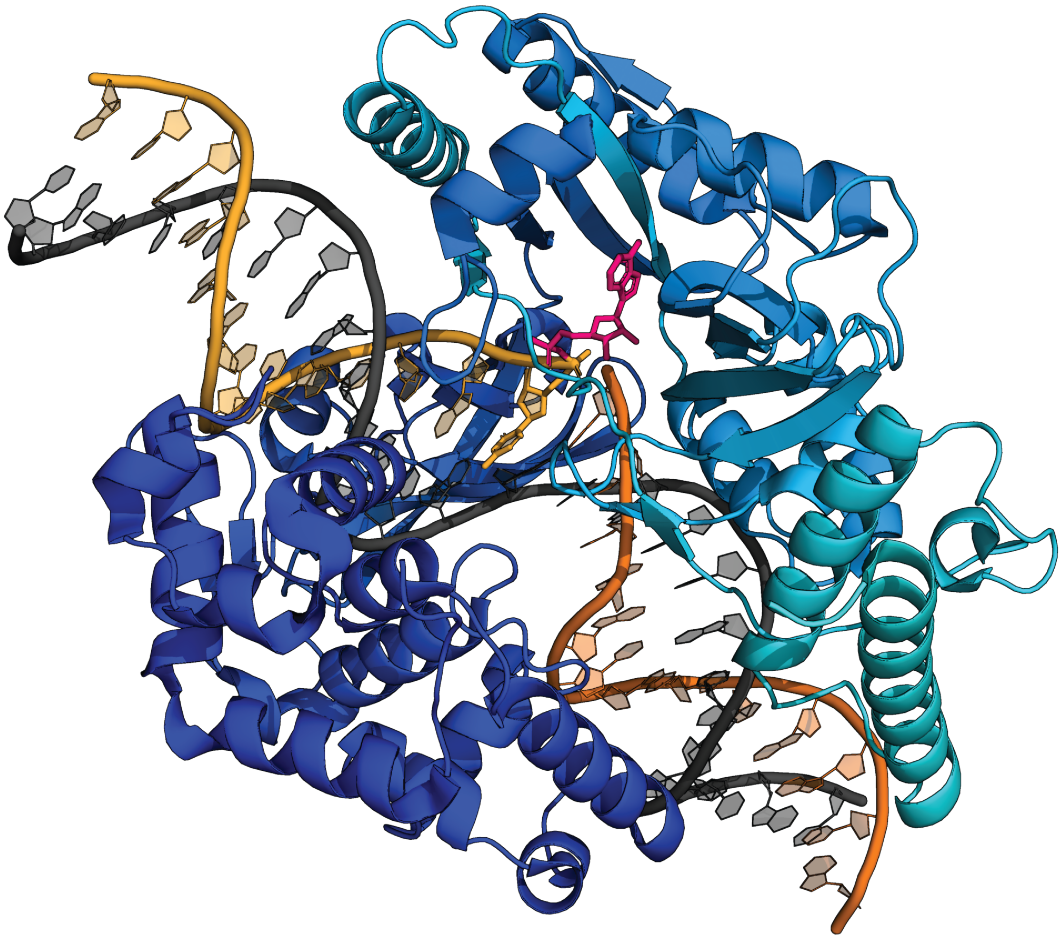


Figure 3.4: Crystal structure of *E. coli* DNA ligase LigA. The 74 kDa enzyme utilizes NAD^+ as a source of energy to catalyze the repair of nicked DNA. The ligase (blue spectrum) adenylates itself, then transfers the adenylate (magenta) to the 5'PO_4 terminus of the nicked DNA strand (yellow and orange). The DNA fragments are bridged by a continuous complementary DNA strand (dark gray). The ligase finally catalyzes the 3' terminus nucleophilic attack on the 5' terminus, thereby repairing the DNA backbone and releasing AMP. The structure above shows LigA bound to the adenylated DNA intermediate. || *E. coli* DNA ligase LigA PDB ID: 2OWO. DOI: 10.2210/pdb2OWO/pdb. || Nandakumar, J., Nair, P., Shuman, S. Last Stop on the Road to Repair: Structure of E. coli DNA Ligase Bound to Nicked DNA-Adenylate. *Molecular Cell*, 26(2):257-271, 2007.

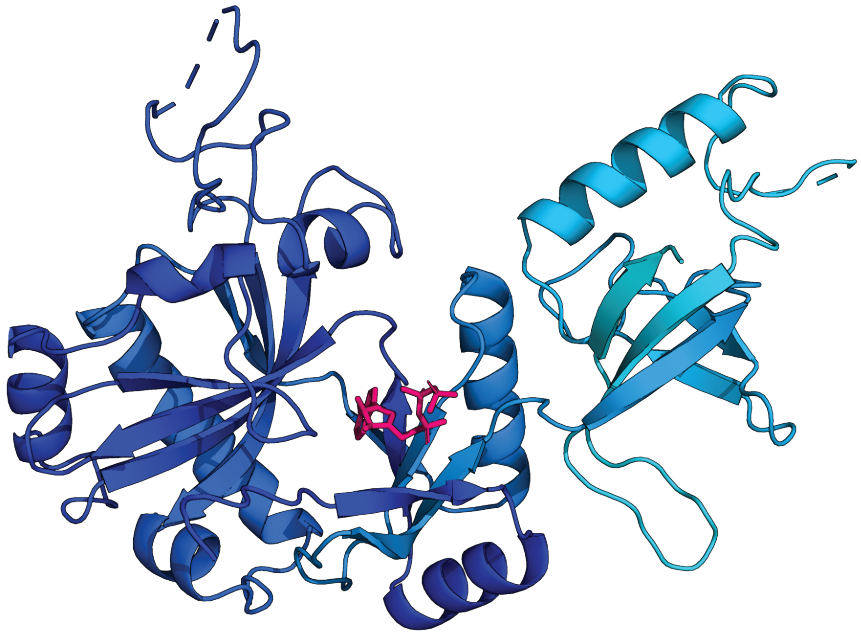


Figure 3.5: Crystal structure of T7 bacteriophage DNA ligase. The enzyme T7 DNAL (blue spectrum) utilizes ATP (magenta) as a source of energy to catalyze the repair of nicked DNA. || T7 bacteriophage DNA ligase PDB ID: 1A0I. DOI: 10.2210/pdb1A0I/pdb. || Subramanya, H., Doherty, A., Ashford, S. Crystal Structure of an ATP-Dependent DNA Ligase from Bacteriophage T7. *Cell*, 85(4):607-615, 1996.

Extensive research has been conducted to investigate and compare the tolerances of different ligases. For example, DNA ligase from the thermophilic bacterium *Thermus thermophilus* displays significantly decreased catalytic efficiency when ligating nicked DNA with an octameric 3'OH terminal strand and displayed no detectable ligation activity of an hexameric 3'OH terminal strand. In contrast, T4 DNAL and T7 DNAL were demonstrated to efficiently ligate the hexameric 3'OH terminal strand, demonstrating that the bacteriophage ligases have higher tolerances for short DNA [137].

The discrepancy in tolerance for DNA substrates between bacterial and bacteriophage ligases is largely attributed to their sizes. The length of DNA at which a DNA-binding protein adheres is referred to as its "footprint". Within this stretch of DNA, DNA-binding proteins tend to be highly selective of their substrates. Hence, a smaller footprint enables a protein to bind more diverse DNA sequences that might be prohibitive for proteins with larger footprints. LigA has an estimated DNA footprint of 19 bp [133]. The significantly smaller bacteriophage ligases correspondingly have smaller estimated footprints, with T4 DNAL at 11 bp [138] and T7 DNAL at 12-14 bp [131]. Compared to bacteriophage ligases, LigA is much more sensitive to DNA modifications or basepair mismatches close to the cite of ligation.

Ligases are indispensable tools in modern molecular biology. Additionally, they provide a unique opportunity to explore the relationship of structure and function. This is primarily informed by their highly conserved catalytic lysine, mechanism of action and basic domain structure across all forms of life. Numerous biochemical and biophysical disciplines contribute

valuable insight to this broad class of enzymes. This is of special relevance in cases where no crystal structure is available (as with T4 DNAL). Further investigation from diverse perspectives into ligases can elucidate more about their mechanisms of action.

3.4 Luciferases

Bioluminescence - or the ability to produce light from biomolecules without the addition of an external light source - is used by many species from diverse corners of the phylogenetic tree, including some bacteria, insects, fungi and marine invertebrates. The biological roles of bioluminescence varies from species to species, and includes predation, mate attraction, and inter-organismal signaling.

In a general sense, bioluminescence results from a chemiluminescent oxidation reaction between an enzyme and its corresponding substrate. The reaction that is observed in the abdomen of the click beetle genus *Elateridae* was first identified by the french pharmacologist Raphael Dubois in 1885, who termed the enzyme responsible for the reaction ”luciferase” and its corresponding substrate that is consumed in the reaction ”luciferine”. Although initially these terms referred only to these exact molecules, ”luciferase” and ”luciferin” have become catch-all terms for the enzyme and substrate, respectively, responsible for a bioluminescent reaction. Since their initial identification, luciferases and luciferins from dozens of species have been isolated and characterized [139, 140].

Nearly all identified luciferases and luciferins have very similar chemical mechanisms of action. The luciferase binds the small luciferin substrate, and catalyzes the oxidation of critical high-energy bonds within the luciferase. This reaction releases energy, some of which is emitted in the form of a photon of a particular wavelength that varies from species to species. As it is an oxidative process, all luciferase reactions require molecular oxygen. Some additionally require the energy carrier ATP or metal salts as cofactors to catalyze the reaction [140].

Luciferase and luciferase-like activity is reported in proteins originating from myriad species. Although not technically classified as a luciferase, the enzyme aequorin from the jellyfish *Aequorea victoria* possesses similar catalytic abilities; it oxidizes the luciferin coelenterazine to produce blue light of wavelength 465 nm. In contrast to luciferases, this enzyme stably binds coelenterazine without catalysis, and only upon activation by calcium ions does it oxidize its substrate and emit light. Aequorin is naturally closely associated with green fluorescent protein (GFP) - one of the most rigorously studied and developed molecules over the century - where the two molecules form a FRET pair. The energy released from aequorin upon oxidizing its substrate is absorbed by a nearby GFP, which then releases a photon of green light. The coordination of a luciferase enzyme and a fluorescent protein is well-documented in other species as well, such as the sea pansy *Renilla reniformis* [141] and several species of bacteria [142].

Interestingly, the overall structures of luciferases and luciferase-like photoproteins are not necessarily highly conserved, although they may catalyze very similar reactions. For example, aequorin and *R. reniformis* luciferase (RLuc) both consume coelenterazine as a luciferin substrate,

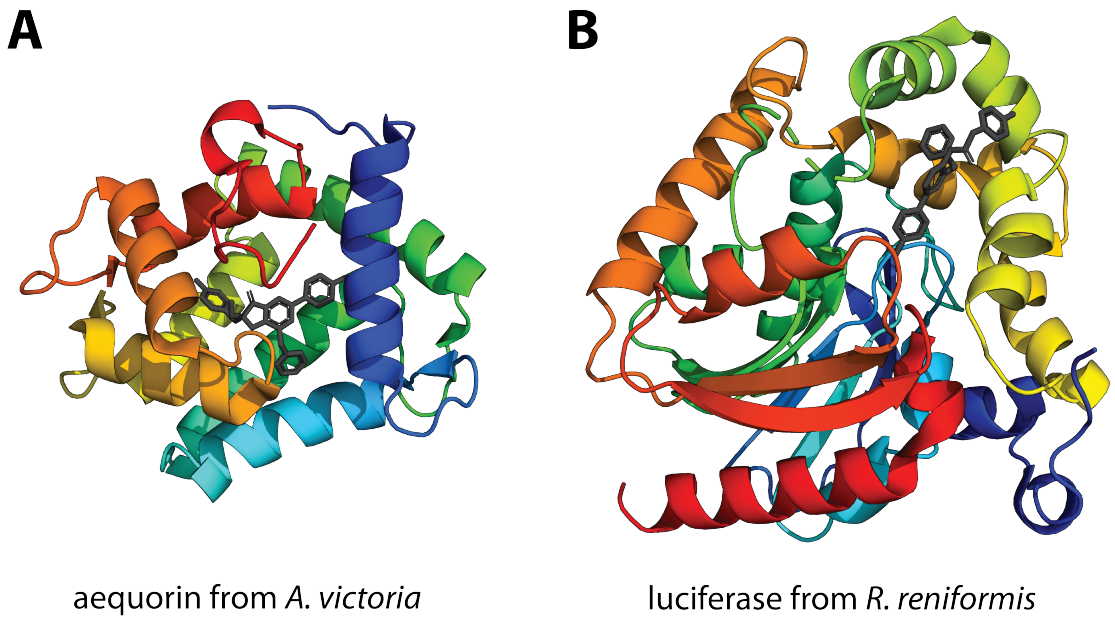


Figure 3.6: Comparison of the crystal structures of aequorin and sea pansy luciferase. Both aequorin from the jellyfish *A. victoria* (a) and RLuc from the sea pansy *R. reniformis* (b) catalyze the oxidation of coelenterazine into coelenteramide, which releases a photon in the blue spectrum. Additionally, both enzymes closely associate with a GFP and form a FRET pair. The structures here show the enzymes (rainbow spectrum) bound to analogues of coelenterazine (dark gray). Notably, 21 kDa aequorin is significantly smaller than 36 kDa sea pansy luciferase, and displays no stable β-sheet secondary structures. || *A. victoria* aequorin PDB ID: 1EJ3. DOI: 10.2210/pdb1EJ3/pdb. || Head, J., Inouye, S., Teranishi, K., Shimomura, O. The crystal structure of the photoprotein aequorin at 2.3 Å resolution. *Nature*, 405(6784):372-376, 2000. || *R. reniformis* RLuc PDB ID: 2PSJ. DOI: 10.2210/pdb2PSJ/pdb. || Loening, A., Fenn, T., Gambhir, S. Crystal Structures of the Luciferase and Green Fluorescent Protein from Renilla reniformis. *Journal of Molecular Biology*, 374(4):1017-1028, 2007.

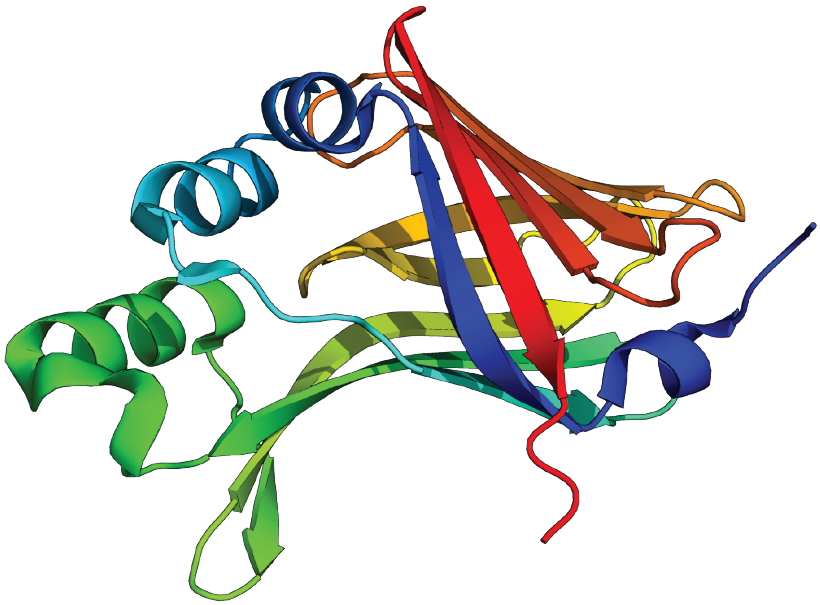


Figure 3.7: Crystal structure of NanoLuc. The 19 kDa NanoLuc engineered luciferase catalyzes the cleavage of the luciferin furimazine, which produces a photon of emission wavelength 460 nm. || PDB ID: 5IBO. DOI: 10.2210/pdb5IBO/pdb. || Lovell, S., Mehzabeen, N., Battaile, K.P., Wood, M.G., Encell, L.P., Wood, K.V. 1.95A resolution structure of NanoLuc luciferase.

they emit photons in the blue spectrum, and are naturally closely associated with a GFP. While the GFP from both species adopts a readily discernible β-barrel conformation containing a central chromophore, the structure of the luciferases is far less comparable. To illustrate the diversity of luciferase structures, the crystallographic structures of 21 kDa aequorin and 36 kDa RLuc are depicted in Figure 3.6. The enzymes possess not only visibly different scaffolds but also vastly different complements of catalytic residues [141, 143].

Unlike fluorescent proteins - where the lifetime of light production is limited by bleaching and quenching effects - luciferases produce light via chemiluminescence. In other words, since light is produced when the substrate is consumed, a single enzyme could continuously emit light for an indefinite length of time as long as there is ample substrate and the protein is not destroyed. This facet is especially useful when continuously imaging samples over timescales that normally bleach fluorophores or fluorescent proteins, or when an application of an external light source is not possible or practical for the experiment. Additional developments in biotechnology - especially in recombinant gene technology - have enabled expression of luciferases across species. This has been instrumental in widely employing luciferases for diverse applications in biological research. For example, luciferases have been extensively used to report gene expression, explore cell signaling, investigate protein-protein interactions, *in vivo* biolumniescence imaging and biolumniescence-based FRET (BRET).

Directed evolution and mutagenesis studies have improved the qualities of many wildtype luciferases, often by increasing protein stability and the relative luminescence. NanoLuc is an

engineered enzyme derived from the luciferase from deep-sea shrimp *Oplophorus gracilirostris*. The wild-type luciferase consists of a heterodimeric complex composed of 35 kDa and 19 kDa subunits. It was determined that the 19 kDa subunit is responsible for the catalytic activity of the luciferase, although it expresses poorly and is unstable without the 35 kDa subunit. Rigorous mutagenesis and structural optimization of the 19 kDa led to a stable and efficient luciferase termed NanoLuc for its relatively small size (Figure 3.7). The improved stability combined with extensive substrate evolution produced a luciferase system of unprecedented efficiency [140].

Part II

Methods

Chapter 4
Molecular Biology and Biotechnology

Overview

Most proteins and enzymes utilized in this work necessitated extensive preparation and development of suitable systems to assess their activity. Several unique assays are described in detail, specifically with regards to incorporating enzyme activity into single-molecule fluorescence measurements. However for techniques that are widely-used, broad descriptions of the methods used therein are given here.

4.1 Recombinant Cloning

Genes of interest were arranged in desired sequences and placed into plasmids suited for replication and protein expression within select strains of genetically engineered *E. coli*. All DNA was stored at 4 °C for short-term use or frozen at -20 °C for long-term storage.

4.1.1 Plasmids

The DNA constructs used within this work are based on the pET21a (+), pET28a (+), pGEX6p2, and pAC4 vectors. The appropriate proteins of interest, purification tags, handling tags, protease cleavage sites, aptamer sequences, linkers and miscellaneous other sequences were cloned within the multiple cloning sites of these vectors.

4.1.2 Primary Genes of Interest

The genes of interest presented in this work are derived from several sources. For most genes longer than approximately 50 nucleotides, DNA was originally commercially synthesized and delivered either as a linear sequence containing only the gene of interest or embedded in

a purified plasmid. Two major exceptions are the genes for *E. coli* DNA Ligase LigA and *P. furiosus* DNA Polymerase *Pfu* DNAP, which were both amplified via PCR directly from the genomic DNA of the respective species. Shorter sequences were introduced during the cloning process either by extension-overlap PCR or as double-stranded DNA fragments derived from complementary synthesized sequences.

The genes for the following proteins were assembled into plasmids via recombinant cloning:

- *Pfu* DNA Polymerase (*Pyrococcus furiosus*)
- Pfu-E10 evolved DNA polymerase (*Pyrococcus furiosus*)
- T7 RNA Polymerase (Bacteriophage T7)
- LigA DNA Ligase (*Escherichia coli*)
- T7 DNA Ligase (Bacteriophage T7)
- T4 DNA Ligase (Bacteriophage T4)
- NanoLuc evolved luciferase (*Oplophorus gracilirostris)*
- Superfolder GFP evolved fluorescent protein (*Aequorea victoria*)
- ddFLN4 motif (*Dictyostelium discoideum*) with a key C18S mutation

Additionally, a single-stranded RNA aptamer was previously evolved to bind the chromophore malachite green (MG), which stabilizes the aromatic groups of the chromophore via intercalation in the binding pocket [144]. The aptamer binds MG with a measured affinity of 800 nM, and binding of the aptamer to MG is estimated to increase its fluorescence quantum yield by a factor of 2360 [145].

A 38-nucleotide minimal version of the malachite green aptamer previously employed in rolling-circle tandem repeat amplification [146] was employed in rolling-circle transcription. Additionally, this sequence was cloned into the pET21a (+) *E. coli* amplification vector in iterations of one, two, and three repeats separated by TC-repeat linker sequences. To ensure transcription termination and optimum aptamer folding, the MG plasmids were linearized via restriction enzyme digest.

4.1.3 Tags for Protein Purification and Specific Handling

Several peptide tags were appended to proteins of interest to enable purification by affinity chromatography:

- The 6xHis tag is a short sequence of six histidine residues that is often appended to the N- or C-terminus of a protein of interest. Its small size but strong localized charge makes it a valuable tool for fast purification of many proteins by Nickel-Ion Immobilized Affinity Chromatography (Ni-IMAC). In this work, the 6xHis tag was coded into the N-terminus of several constructs.

- The Glutathione S-transferase (GST) tag binds with high affinity to glutathione (GSH), and consequently was utilized in combination with Glutathione Sepharose columns for high-purity affinity chromatography.

Several short tags and sequences were employed for covalent modification, immobilization, cantilever handling, or protein modulation:

- The ybbR tag is a peptide (DSLEFIASKLA) developed as a minimal sequence of a peptide carrier protein (PCP), a domain that in nature is covalently modified with Coenzyme A (CoA) by the enzyme Sfp phosphopantetheinyl transferase (Sfp) at a conserved serine residue [147–149]. While full PCP domains are on the order of 100 amino acids long, the ybbR tag is significantly shorter and hence more amenable to fusion with a protein of interest. The protein of interest can then be labeled with CoA and CoA-modified molecules such as biotin or DNA [86, 102]. Additionally, this tag enables site-specific covalent surface immobilization of proteins of interest in AFM-based single-molecule force spectroscopy experiments [47].

- A minimal variant of the GCN4(7P14P) peptide fragment, derived from the yeast transcription factor GCN4, constituted the GCN4 tag (YHLENEVARLKK) [150]. This tag binds with high affinity to the C11L34 single-chain variable fragment antibody, thereby providing a system to handle proteins with an antibody-functionalized cantilever [61, 86, 102].

- The short peptide Strep-Tag II (WSHPQFEK) binds to tetravalent StrepTactin [109] and monovalent StrepTactin (monoST) [69]. This tag was utilized as a specific pulling handle for a monoST-functionalized cantilever. Specifically, the low-force pulling geometry has an expected rupture force in a range compatible with DNA-immobilized molecules in SMC&P.

- The prototypical pathogenic adhesin SdrG binds to the 15-amino acid Fgβ sequence (NEEGFFSARGHRPLD) in a ”dock, lock, latch” pathway. The mechanical strength of this adhesion system has been extensively investigated by Milles *et al.*, and it was determined that the pulling geometry has a dramatic effect on the expected rupture force of this complex [36]. When N-terminally immobilized SdrG and Fgβ fused to the N-terminus of a protein of interest are mechanically probed, the expected rupture force is compatible with N- and C-terminally immobilized monovalent Streptavidin in SMC&P.

- The PreScission Protease cleavage sequence (LEVLFQ/GP) was added for controlled protease cleavage of the purification tag(s) from several proteins of interest. This was advantageous for removal of large tags (such as a GST tag) or highly charged tags that could interfere with protein function or binding (such as a 6xHis tag).

- Linker sequences consisting of glycine, serine and alanine repeats were inserted in select constructs between domains or tags to improve folding or tag accessibility.

4.1.4 DNA Amplification and Manipulation

An indispensable tool for cloning is polymerase chain reaction (PCR) of DNA sequences of interest. The technique enables not only exponential *in vitro* amplification of DNA fragments, but also modifications such as the addition of small tag sequences to the DNA at the termini of the sequences.

PCR reactions for subcloning were carried out with commercially-available PCR master-mixes, specifically the Phusion HF Master Mix and the Q5 HF Master Mix (New England Biolabs, Massachusetts, USA). Additionally, commercially synthesized DNA primers (metabion GmbH, Planegg, Germany; Eurofins Genomics GmbH, Ebersberg, Germany) were employed. Template DNA was derived from various sources, primarily plasmids.

Enzymatic cleavage and re-assembly of DNA for the purpose of recombinant cloning has undergone extensive research to develop new protocols and optimize existing ones. Genes of interest were assembled into plasmid vectors through several recombination strategies according to what was judged to be the most contextually straightforward strategy.

- Commercially-available restriction enzymes (FastDigest enzymes and buffers, Thermo Fisher Scientific) were utilized to cut plasmid DNA or PCR products at specific palindromic sequences, leaving sticky ends of linear DNA that were ligated by commercially-available T4 DNA Ligase (Thermo Fisher Scientific).

- Recently, an extremely robust novel DNA recombination method was first developed and described by Daniel G. Gibson [151], and this strategy is hence often referred to as "Gibson Assembly". Discrete fragments of DNA are first designed to have overlapping terminal sequences, and the DNA can be sourced from *in vitro* PCR products or miniprepped plasmids. The DNA fragments are then exposed to an exonuclease enzyme that excises a length of nucleotides at all 5' termini, thereby creating complementary single-stranded regions of DNA between fragments that allow them to anneal. Any gaps in the excised DNA are filled in by a DNA polymerase, and the nicks are sealed by a DNA ligase. In principle, this produces circular plasmids with newly-inserted genes of interest. In contrast to several other classical cloning strategies, *e.g.* restriction enzyme-based cloning, there is the possibility to seamlessly join genes and avoid DNA scars. However, a critical requirement is that each DNA fragment is long enough that the 5' $\rightarrow$ 3' exonuclease activity will not completely chew through the extent of a sequence. The commercially-available NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) containing all required enzymes was utilized in one-step Gibson Assembly of linear plasmid and PCR product DNA fragments.

- Blunt-ended DNA - such as PCR products - was also directly circularized without the creation of cohesive sticky ends via blunt-end ligation. The 5' termini of the DNA fragments were first phosphorylated by T4 Polynucleotide Kinase (New England Biolabs), and subsequently ligated by T4 DNA Ligase (New England Biolabs) in the presence of polyethylene glycol as a crowding agent.

All recombinant constructs were first sequenced in the multiple cloning site before proceeding to expression. DNA sequencing was carried out using cycle sequencing technology by Eurofins Genomics GmbH utilizing standard primers such as T7 (5' TAA TAC GAC TCA CTA TAG GG 3') and T7 term (5' CTA GTT ATT GCT CAG CGG T 3'), or custom primers for especially long constructs that required multiple sequencing reactions.

4.2 Bacteria Manipulation

Bacteria cultures provided a simple and cost-effective means of producing large quantities of plasmid DNA, isolating single products from recombinant DNA cloning, and expressing proteins of interest. All employed bacterial strains adhere to S1 safety standards for genetically engineered organisms.

4.2.1 Chemically Competent Bacterial Strains

Several chemically competent K12-strain and protease-deficient B-strain bacteria species were employed for DNA amplification as well as protein expression.

- K12 bacterial strains were utilized during the cloning process for plasmid repair and DNA amplification. In particular, DH5-Alpha cells (Thermo Fisher Scientific) were commonly chosen for transformation of newly-created plasmid constructs that required repair and amplification before sequencing. Mach1 $T1^R$ cells (Invitrogen, Thermo Fisher Scientific) were also chosen for their short replication time, enabling rapid growth of cell cultures.

- B-strain bacteria were primarily utilized for T7 expression of proteins of interest. BL21-CodonPlus (DE3)-RIPL cells (Stratagene, Agilent Technologies) as well as an improved alternative strain Nico21(DE3) (New England Biolabs) were transformed with sequenced plasmids containing genes for proteins of interest.

4.2.2 Transformation

Chemically competent cells were transformed with plasmid DNA containing genes of interest according to a standard protocol.

In brief, shots of chemically competent cells were stored at -80 °C and thawed on ice shortly before transformation. DNA samples were gently added to the shots and incubated on ice for approximately 10 min, followed by a heat-shock at 45 °C for 45 s. Antibiotic-free SOC growth medium was added to the cell shots, and the cells were incubated with shaking at 37 °C for up to 1 h to allow the transformed cells to produce the antibiotic-resistance genes encoded in the plasmids. Finally, cells were plated to LB-Miller agar plates containing growth nutrients as well as a selective antibiotic and incubated at 37 °C to obtain colonies. Single colonies could then be picked to grow cultures for miniprep or protein expression.

4.2.3 Cultures and Growth Media

Cells were grown in various enriched media in volumes ranging from 5 mL to 500 mL according to the purpose of the cultures.

For the purposes of amplifying plasmid DNA for sequencing, cloning or *in vitro* transcription reactions, transformed Mach1 or DH5-Alpha cells were grown in LB-Miller medium containing 50 µg/mL of the appropriate antibiotic - either Carbenicillin or Kanamycin. Cultures of volumes 5 mL - 20 mL were incubated at 37 °C with shaking for approximately 12 h. Cultures were then spun down to separate the cells from the medium. The medium was discarded and the DNA was isolated from the cell pellet using a QIAprep Miniprep Kit (Qiagen).

Expression cultures of BL21-CodonPlus or Nico21(DE3) cells were first grown as 20 mL overnight precultures in LB-Miller medium containing the appropriate antibiotic(s) at 37 °C with shaking. Precultures were then added to larger volumes of expression-appropriate medium of up to 500 mL total volume also containing the appropriate antibiotic(s). SB medium with the timed addition of IPTG to a concentration of 1 mM was used for induced expression for several constructs. Auto-induction medium (AI medium) based on a protocol originally developed by Studier [152] was used to support high-density growth of expression cultures as well as enable automatic induction of protein expression based on the cell consumption of nutrients in solution. A variation of AI medium that omitted lactose was also used. This formulation supported high-density cell cultures but required the addition of IPTG to 1 mM to induce protein expression, which was advantageous for proteins sensitive to aggregation or proteolysis.

Following protein expression, cultures were centrifuged to separate the cells from the medium. The medium was discarded and the cells were either directly resuspended in buffers for lysis and purification or frozen away as pellets at -80 °C for purification at a later date.

4.3 Protein Purification

Cell pellets derived from expression cultures were lysed using a combination of hypotonic buffers, gentle detergents, and intensive sonication. Following lysis, the lysate was centrifuged to separate the soluble and insoluble fractions. The soluble fraction - which normally contained the protein of interest - was filter-sterilized to remove larger particulates as well as any microbes. In order to obtain purified and concentrated stocks of the proteins of interest, several chromatography techniques - namely affinity chromatography - were used to isolate proteins based on engineered features. Following purification, select fractions for all constructs were analyzed via SDS-PAGE or native PAGE to isolate the fraction of optimal purity.

4.3.1 Ni-IMAC

Nickel-Ion Immobilized Affinity Chromatography (Ni-IMAC) was the most prominently used strategy in this work. To this end, proteins of interest were expressed with a 6xHis tag, a series of six histidine residues that possesses a strong local negative charge. The histidine residues

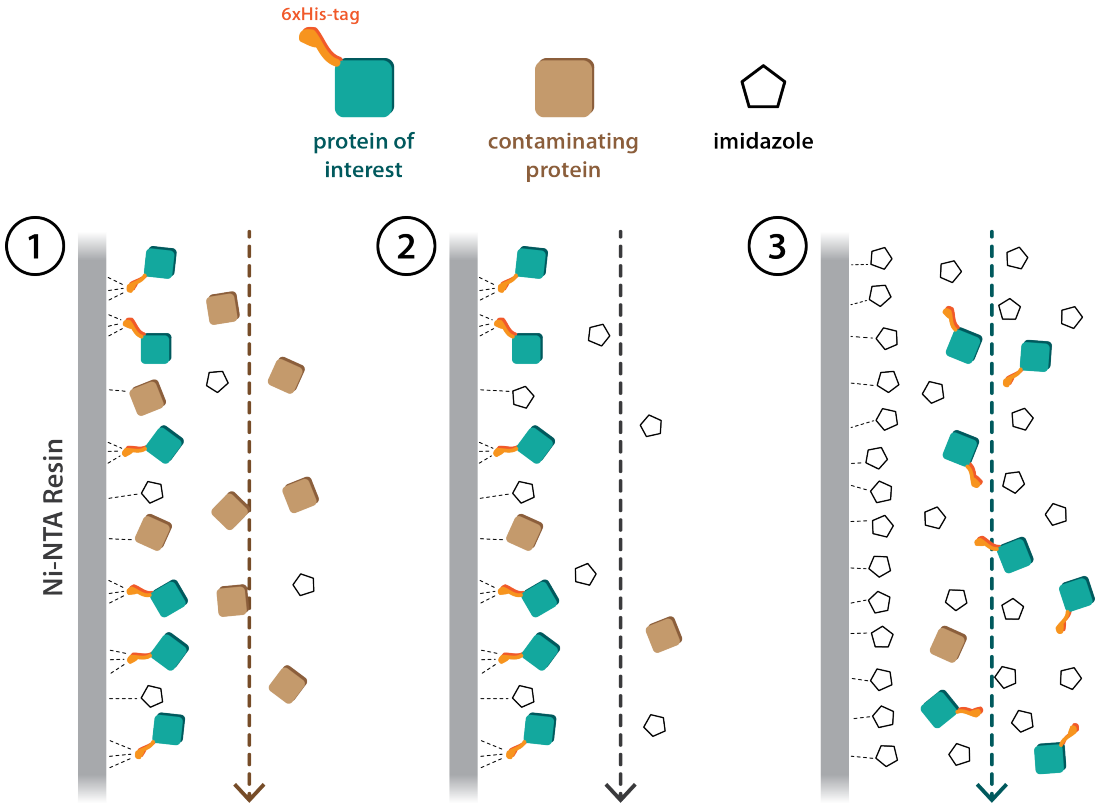


Figure 4.1: Purification of a 6xHis tagged protein by Ni-NTA column. The unpurified soluble fraction of the cell lysate is suspended in a buffer containing low concentrations of imidazole to discourage binding of untagged proteins to to the Ni-NTA resin. The column is first flushed with the unpurified soluble fraction, which contains the 6xHis tagged protein of interest as well as natively expressed proteins from the *E. coli* genome. The tagged protein binds with high affinity to the column Ni-NTA resin, while most other proteins flow through the column. The column is washed and then flushed with a buffer containing highly concentrated imidazole. The imidazole competes with the 6xHis tag for binding to the resin, and the protein of interest elutes with high purity. || Adapted from: Erlich, K. R. G4b Biophysics Practical Course Script. LMU München, 2017 (unpublished).

bind with high affinity to nickel ions, thereby immobilizing the tagged protein of interest in a Ni-NTA sepharose column (HisTrap HP and HisTrap FF columns, GE Healthcare) while most other proteins flow out of the column. The protein of interest is eluted with the application of imidazole, a molecule that resembles histidine and competes for binding to the nickel-embedded resin (Figure 4.1). The elution step may be executed either isocratically to obtain the highest concentration of protein or in a gradient to obtain more pure fractions of the protein.

The 6xHis tag was also removed from the protein of interest by directed protease cleavage following Ni-IMAC purification, as the highly charged 6xHis tag can interfere with protein function. Specifically, the binding of Strep-Tag II to monovalent StrepTactin is significantly hindered by the presence of a nearby 6xHis tag [69, 102].

4.3.2 Reverse-His Purification

In cases where many contaminating proteins co-eluted with the protein of interest in the first round of 6xHis tag-based purification, the construct protein could be further purified.

Following cleavage of the 6xHis tag via directed protease cleavage, several constructs were again flushed through a Ni-NTA column. In this case, the protein of interest elutes immediately at low concentrations of imidazole, while contaminating proteins that bind strongly to the column via charged motifs remain bound. The contaminating proteins can then be eluted later with high concentrations of imidazole. This offered a simple and fast strategy to increase the sample purity of 6xHis tagged proteins.

4.3.3 Glutathione S-Transferase

The Glutathione S-Transferase (GST) tag was additionally implemented to achieve exceptionally high protein purity after one purification step. The GST tag binds with high affinity to glutathione (GSH), which was immobilized in column via sepharose resin (GSTrap HP and GSTrap FF columns, GE Healthcare). GST tagged proteins were then eluted with high concentrations of GSH, which competes with the column resin for binding to the GST tag.

Of note, the GST tag is a very large tag with a mass of 26 kDa. Compared to the 6-amino acid 6xHis tag, this purification tag is often less compatible with proteins of interest, especially larger constructs due to *E. coli* generally having difficulty with expressing large proteins. Its bulkiness could also conceivably block binding sites or interfere with protein function. Therefore, the GST tag was always cleaved after purification.

4.3.4 Ion Exchange Chromatography

In the case where a molecule has an exceptionally high net surface charge - either positive or negative - ion exchange chromatography is ideally suited for purification. The column contains a charged resin that binds oppositely-charged molecules, while similarly-charged or neutral molecules pass through the column. As the sample is applied under conditions of low ionic buffer strength, oppositely-charged molecules initially bind strongly to the column. The bound molecules are eluted by application of a buffer with high ionic strength, such that the ions in solution compete for binding to the column resin.

DNA-binding proteins often have a net positive surface charge in order to interact favorably with negatively-charged DNA. Moreover, they also tend to have a specific affinity for heparin, a highly negatively-charged polymer that resembles that sugar-phosphate backbone of DNA. These proteins were purified by heparin column (HiTrap Heparin HP column, GE Healthcare) in a combination of heparin affinity and cation exchange chromatography. Proteins were eluted at high salt concentration where cations competed for binding to heparin.

Similarly, being a highly negatively-charged molecule due to its phosphate backbone, DNA and DNA-modified constructs were also purified via ion exchange. This was especially useful for separating DNA-modified constructs from unreacted CoA-DNA in solution (see *In Vitro* Post-Translational Modifications). DNA constructs were purified by anion exchange (HiTrap Q HP column, GE Healthcare), as their binding to the positively-charged resin was out-competed by highly concentrated anions in solution.

4.3.5 Size-Exclusion Chromatography

Purification according to size is exceptionally advantageous after labeling reactions with small molecules, specifically when the labeled molecule of interest and the small molecule label have markedly different sizes. The unpurified sample is applied to a column packed with a porous resin. Small molecules are able to enter the resin beads and their flow speed is hampered. On the other hand, larger molecules are not able to enter the beads. Instead, they pass by them and flow through quickly relative to smaller molecules. The eluted fractions are consequently separated according to size, with the largest molecules eluting first and the smallest molecules eluting last.

In particular, labeled constructs for SMC&P (see *In Vitro* Post-Translational Modifications) were purified via size exclusion chromatography in order to remove unreacted CoA-biotin and fluorescently-labeled maleimide, both of which would likely be sources of difficulty in SMC&P if left in solution with the transfer construct.

4.4 *In Vitro* Post-Translational Modifications

The proteins of interest used in this work relied heavily on *in vitro* covalent modifications. These modifications were undertaken with a series of labeling reactions followed by additional purification steps to remove the unreacted label from solution.

4.4.1 CoA Conjugation via Sfp

Single-molecule probing of immobilized molecules requires specific anchoring in order to maintain a conserved pulling geometry. To that end, the enzyme Sfp has been extensively used to covalently join CoA to a ybbR tag in both surface and cantilever functionalization [47]. This technique was similarly employed here in AFM-based SMFS experiments to directly immobilize proteins of interest to a CoA-functionalized surface or cantilever.

SMC&P largely takes advantage of the extremely high affinity of complementary DNA strands as well as the molecule's stability to noncovalently anchor molecules of interest on a surface. As such, the protein of interest is assembled into a transfer construct via the addition of a single-stranded DNA molecule whose complementary strand is covalently attached to the surface. Initially, this was accomplished with a chimera construct consisting of GFP and zinc-finger domain, which binds with high affinity to its consensus sequence [101]. After the Sfp-CoA labeling system was developed, the zinc-finger domain was replaced with the small ybbR tag to instead covalently join the transfer construct to CoA-modified DNA [86, 102]. A key advantage of covalently attached DNA is that there is no risk of the protein of interest and the DNA dissociating from each other in additional purification steps or during an SMC&P experiment. Additionally, the single-stranded DNA modification enables reliable and long-lasting immobilization of proteins of interest to the surface.

Similarly, the high affinity of biotin for Streptavidin was an asset in developing DNA-free SMC&P. CoA-modified biotin was covalently joined to a ybbR tag to enable specific and efficient immobilization of transfer constructs to a surface functionalized with monovalent Streptavidin.

Sfp labeling reactions were carried out in 1x Sfp Reaction Buffer (120 mM Tris HCl pH 7.5 at room temperature, 10 mM $MgCl_2$, 150 mM NaCl, 2 %(v/v) glycerol, 2 mM DTT). Covalent attachment of proteins of interest to CoA-functionalized surfaces or cantilevers was accomplished with approximately 10 µM Sfp and 10 µM ybbR tagged protein in solution. For the covalent attachment of proteins of interest to CoA-modified DNA or biotin, the CoA-modified molecules were added to final concentrations between 10 µM - 100 µM in solution. Reactions were incubated at room temperature for up to 2 h or at 4 °C for up to 48 h. Protein-conjugated surfaces and levers were rinsed with the appropriate buffer to remove Sfp and excess protein in solution. DNA- and biotin-labeled proteins were further purified by anion exchange chromatography and size exclusion chromatography, respectively.

4.4.2 Maleimide Conjugation

Synthetic fluorophores - especially more recently developed state-of the art dyes - have several advantages compared to fluorescent proteins. Primarily, they are often much brighter and more resistant to bleaching [153]. Their increased sensitivity and robustness when compared to fluorescent proteins makes them an attractive alternative for single-molecule imaging. Additionally, fluorescent proteins are subject to misfolding and require maturation for the formation of the chromophore. While synthetic dyes can be quenched via chemical reaction - *e.g.* Cy5 is reversibly quenched with TCEP [154] - they are generally more stable than fluorescent proteins.

Finally, as SMC&P constructs are necessarily chimeric, a bulky fluorescent protein tag could conceivably create problems for construct expression or protein function. Synthetic dyes are comparatively small, and a wide range of labeling techniques are available to post-translationally attach them to a protein of interest via a small tag.

Considering the advantages of synthetic fluorophores, a transfer construct for SMC&P was designed for such an imaging strategy. A ddFLN4 motif was designed with a C-terminal reactive cysteine (Cys), which in a reduced state possesses an active thiol that spontaneously forms a stable thioether bond with a maleimide reactive group. Cys was reacted with commercially available maleimide modified with a Cy5 or Atto647N fluorescent prosthetic group (Sigma-Aldrich). Labeling reactions were carried out in 1x Maleimide Reaction Buffer (30 mM Tris-HCl pH 7.2 at room temperature, 150 mM NaCl). Approximately 150 µM purified ddFLN4 and 2.5 mM fluorescently-labeled maleimide were incubated overnight at 4 °C, followed by biotinylation for SMC&P and purification by size-exclusion chromatography.

Maleimide-Cys chemistry was additionally implemented for direct immobilization of molecules to a surface or cantilever. As described previously, this strategy was used to immobilize monovalent *Strep*-Tactin [69, 102] and Streptavidin [49], which are not compatible with a ybbR tag. In order to immobilize these molecules to the surface or cantilever, the purified

proteins were briefly reduced with TCEP beads or liquid TCEP for 30 min at room temperature. TCEP beads were then removed by membrane spin column purification, and liquid TCEP was removed by desalting spin column purification. Monovalent *Strep*-Tactin and Streptavidin were then applied to the surface or cantilever and incubated for up to 1 h at room temperature. Excess protein was rinsed away with the appropriate buffer.

The single-chain antibody fragment C11L34Ser, which binds the GCN4 tag with high affinity, was also attached to the cantilever via maleimide-Cys chemistry as described previously [61, 86, 100, 101]. As folded antibodies often have disulfide bridges in their structure, no additional reducing agent was applied to to C11L34Ser before immobilization. The purified antibody was applied directly to the cantilever and incubated for up to 1 h at room temperature. Excess antibody was rinsed away with the appropriate buffer.

4.5 *In Vitro* Reactions and Analysis

Several strategies for monitoring enzyme activity were carried out, with special emphasis on developing a means of monitoring fluorescent products or output signals. The setup of a typical reaction for each enzyme is described here.

4.5.1 *Pfu* DNA Polymerase Replication

In order to compare activity between *Pfu* DNAP constructs under various conditions, PCR screens were devised. A typical PCR screen was prepared as follows:

***Pfu* DNAP PCR**
1x *Pfu* DNAP Reaction Buffer
200 µM each dNTP
200 nM - 2 µM dA-Cy5, dC-Cy3 or dC-Cy5 (optional)
2.5 µM forward primer
2.5 µM reverse primer
5 - 50 pM plasmid template DNA
1 nM - 1 µM *Pfu* DNAP

The thermal cycler program used in each PCR depends very heavily on the annealing temperature of the primers to the template as well as the length of the DNA sequence to be amplified. Below is an example of a typical thermal cycler program:

PCR Amplification Program
94 °C → 2 min
repeat cycle x35:
    94 °C → 30 s
    58 °C → 30 s
    72 °C → 2 min

72 °C → 15 min
10 °C → hold

However, the elevated temperatures of a typical *Pfu* DNAP PCR are not relevant for AFM- and SMC&P experiments, which occur at room temperature. Therefore, it was necessary to devise a means of detecting DNAP activity at lower temperatures that does not rely on thermally-driven DNA melting. A single-stranded template oligomer was annealed to a much shorter complementary strand to act as a primer for the polymerase, thereby enabling low-temperature replication. Importantly, in contrast to a typical PCR, this strategy does not yield exponential amplification of the sequence of interest. Low-temperature non-exponential replication was typically prepared as follows and incubated at 22 °C - 40 °C from 25 min to as long as overnight:

***Pfu* DNAP Low-Temperature Replication Reaction**
1x *Pfu* DNAP Low-Temperature Reaction Buffer
200 µM each dNTP
200 nM - 2 µM dA-Cy5, dC-Cy3 or dC-Cy5 (optional)
100 nM annealed primed DNA template
10 nM - 1 µM *Pfu* DNAP

Pfu DNAP reaction products were analyzed by gel electrophoresis. Agarose gels were stained with Roti-GelStain (Carl Roth GmbH, Karlsruhe, Germany) and subsequently imaged on a ChemiDoc MP (Bio-Rad Laboratories, Inc., California, USA). Gels were imaged with UV excitation of DNA or LED excitation of incorporated fluorophores, and an appropriate emission filter for the desired wavelength was used.

4.5.2 T7 RNA Polymerase *In Vitro* Transcription

T7 RNA polymerase activity was determined by *in vitro* transcription production of RNA aptamers [155]. Two types of DNA templates were generated (Figure 4.2): linearized plasmid templates for run-off transcription, or short circularized templates for rolling circle transcription. For run-off transcription (Figure 4.2a), a pET21a plasmid was cloned with one (1xMG), two (2xMG), or three (3xMG) repeats of the malachite green aptamer sequence downstream of the T7 RNA promoter sequence. Aptamer repeats were additionally separated by stretches of T and C bases to facilitate correct aptamer folding, as those regions have no self-complementarity and discourage inter-aptamer annealing. The recognition sequence of the restriction enzyme EcoRI was included at the downstream 3' end to enable plasmid linearization, thereby yielding transcripts of predetermined lengths.

Rolling-circle transcription of MG aptamers (Figure 4.2b) offers a unique strategy of aptamer generation. Furukawa *et al.* reported a transcription strategy that yields continuous aptamer repeat transcripts of potentially ten or more repeats long [146]. Single-stranded template DNA is precircularized with the aid of a splint strand, and optionally ligated with a 5' reactive phosphate

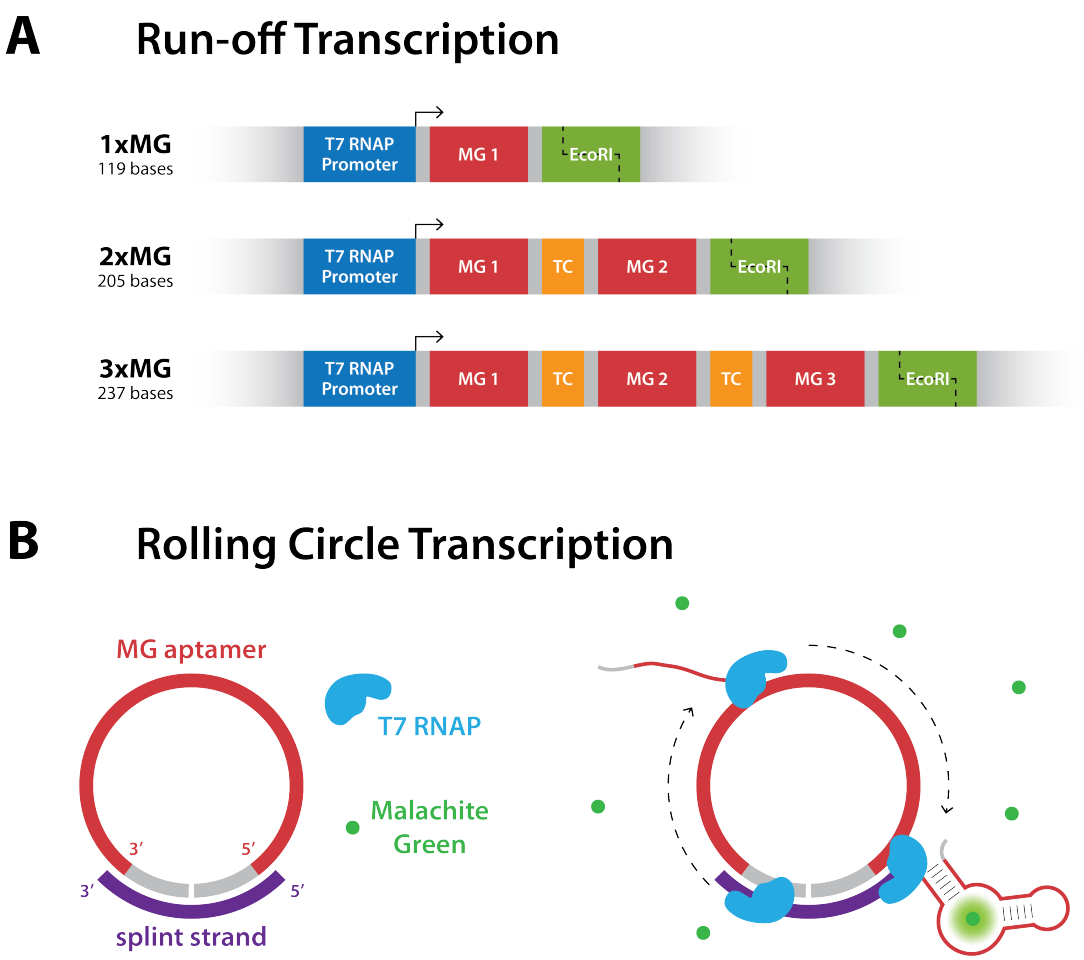


Figure 4.2: RNA aptamer generation strategies. (a) Run-off transcription was accomplished by linearizing several plasmid templates with the restriction enzyme EcoRI. This enabled generation of RNA aptamer transcripts of well-controlled lengths. The malachite green aptamer sequence (MG 1-3) was cloned into the pET21a plasmid downstream of the T7 RNAP promoter sequence in one, two, or three repeats. Sequential MG repeats in the 2xMG and 3xMG templates were separated by stretches of T and C bases (TC) to facilitate correct aptamer folding. (b) Rolling circle transcription, as described by Furukawa *et al.*[146], was accomplished with a synthetic DNA template. Single-stranded aptamer template DNA and single-stranded splint DNA are incubated to pre-circularize the aptamer template. As the template strand also contains a 5' reactive phosphate, it can then be ligated with T4 DNA ligase to form permanently circularized aptamer strands. T7 RNAP binds the double-stranded region of the template strand and transcribes the aptamer sequence, which binds malachite green and enhances its fluorescence. The polymerase continues transcription of aptamer repeats indefinitely until it eventually dissociates from the template strand.

and T4 DNA ligase. Although T7 RNAP strongly favors a double-stranded canonical promoter sequence to initiate transcription, Furukawa *et al.* reported efficient promoter-free transcription. In contrast with run-off transcription, rolling circle transcription produces transcripts of greatly variable lengths, depending on the dissociation probability of the polymerase from the template.

Both strategies utilize identical DNA sequences for the MG aptamer sequences. Concurrent with transcription, each aptamer repeat self-anneals as it is transcribed and forms a stem-loop-stem-loop structure that binds and stabilizes MG, thereby enhancing the fluorescence.

The single-stranded RNA oligomers were either assessed by gel electrophoresis, or used in conjunction with MG to produce a fluorescent signal with $\lambda_{ex} = 628\,\text{nm}$ and $\lambda_{em} = 654\,\text{nm}$ and measured in real-time on an Infinite M1000 PRO microwell plate reader (Tecan, Zurich, Switzerland). High-percentage agarose gels were stained with ethidium bromide and imaged on a ChemiDoc MP (Bio-Rad Laboratories, Inc., California, USA) with UV excitation of RNA. A typical *in vitro* transcription reaction was prepared as follows and incubated for variable times at 22 °C:

***In vitro* Transcription Reaction**
1x T7 RNAP Reaction Buffer
2 mM each rNTP
1.5 U Thermostable Inorganic Pyrophosphatase (New England Biolabs)
30 U RiboLock RNase Inhibitor (Thermo Fisher Scientific)
5 mM TCEP (Thermo Fisher Scientific)
5 nM template DNA
10 nM - 1 µM T7 RNAP
5% (v/v) glycerol
1 µM malachite green (Fluka Analytical) for plate reader assays

As T7 RNAP requires Mg^{+2} ions as a cofactor for transcription, *in vitro* transcription reactions were halted by the addition of EDTA to chelate Mg^{+2} ions. This was also crucial for the analysis of aptamer transcripts by gel electrophoresis, as Mg^{+2} ions catalyze the hydrolysis of RNA molecules, especially at elevated temperatures.

Following incubation, 25 mM EDTA was added to samples of the *in vitro* transcription reactions to chelate Mg^{+2} ions to prevent RNA degradation at high temperatures. Formamide was added to a final volume of 50 %(v/v) to stabilize a single-stranded conformation of RNA transcripts. Samples were then heated at 70 °C for 10 min to melt annealed RNA transcripts, followed by a short incubation on ice at 2 min. RiboRuler Low Range Ladder RNA standard (Thermo Scientific, Massachusetts, USA) was also used to estimate the size of RNA transcripts from gel electrophoresis. Typically, a 2 %(w/v) agarose gel without any staining agent in the gel matrix was prepared, and gels were run until good separation of RNA transcripts was achieved. Gels were then stained with incubation in ethidium bromide and imaged with UV excitation.

4.5.3 DNA Ligase Ligation and FRET Signaling

DNA ligase activity under various conditions was determined by DNA band-shift, fluorophore colocalization and FRET assays. DNA duplexes composed of individually synthesized single-stranded DNA oligomers were annealed separately by gradual cooling from 80 °C to room temperature. Sets of duplexes with cohesive sticky ends and 5' reactive phosphates were then ligated and analyzed by native PAGE. Gels were imaged on a ChemiDoc MP (Bio-Rad Laboratories, Inc., California, USA) with LED excitation of fluorophores and an appropriate emission filter for the desired wavelength. DNA ligase reactions were prepared as follows and incubated at 22 °C for 2 - 12 h:

***E. coli* DNAL Ligation Reaction**
1x *E. coli* DNA Ligase Reaction Buffer
250 nM - 1 µM each annealed DNA duplex
25 nM - 1 µM *E. coli* DNAL

T7 DNAL Ligation Reaction
1x T7 DNA Ligase Reaction Buffer
250 nM - 1 µM each annealed DNA duplex
25 nM - 1 µM T7 DNAL

T4 DNAL Ligation Reaction
1x T4 DNA Ligase Reaction Buffer (Thermo Fisher Scientific, Massachusetts, USA)
250 nM - 1 µM each annealed DNA duplex
25 nM - 1 µM T4 DNAL

4.5.4 Luciferase Luminescence

Engineered NanoLuc and wildtype firefly luciferase LuLic are both able to cleave the luciferin coelenterazine in a reaction that releases a photon. NanoLuc has greatly increased affinity for the luciferin furimazine, resulting in greater catalytic activity and hence brighter overall luminescence. Luciferase luminescence reactions were prepared as follows and incubated at 22 °C for up to 2 h:

NanoLuc Luminescence Reaction
1x NanoLuc Reaction Buffer
60 µM luciferin (coelenterazine or furimazine)
50 pM - 50 nM luciferase (NanoLuc or LuLic)

Chapter 5
AFM Experiment Setup

Overview

AFM experiments were prepared according to robust protocols used previously in numerous SMFS and SMC&P experiments [36, 51, 69, 86, 102, 156, 157].

5.1 Cantilever Preparation

Cantilevers (Biolever mini BL-AC40TS, Olympus, Japan; MLCT, Bruker, Massachusetts, USA) were oxidized in a UVOH 150 LAB UV-ozone cleaner (FHR Anlagenbau GmbH, Ottendorf-Okrilla, Germany). Silanization was then accomplished by incubation in (3-Aminopropyl) dimethylethoxysilane (ABCR, Karlsruhe, Germany, 50 %(v/v)inEthanol) for 2 min. Cantilevers were then washed sequentially in toluene, isopropanol, and finally deionized water to remove excess silane, and subsequently baked at 80 °C for 30 min to catalyze the covalent attachment of silane to the cantilever surface. Silanized cantilevers were then incubated in 25 mM heterobifunctional polyethylene glycol (PEG) crosslinker [158, 159] with N-hydroxy succinimide and maleimide groups (molecular weight 5000 Da, Rapp Polymere, Tübingen, Germany) dissolved in either 150 mM sodium borate buffer, pH 8.5 or 50 mM HEPES, pH 7.5. The N-hydroxy succinimide group forms a stable covalent bond with primary amine groups on the silanized cantilever, and the maleimide group is free to react with thiol or cysteine groups and form a stable thioester bond with molecules of interest. Unreacted PEG crosslinkers were rinsed off with deionized water.

PEGylated cantilevers were either reacted directly to handling molecules, or further functionalized with Coenzyme A (CoA) dissolved in Coupling Buffer for 1 h. Unreacted molecules of interest were rinsed off with an appropriate measurement buffer, and unreacted CoA was washed off with deionized water. CoA-functionalized levers were then reacted via Sfp to ybbR-tagged handling molecule. In a typical reaction, cantilevers were incubated for 1 - 2 h

in a solution of 1x Sfp Reaction Buffer containing 100 µM ybbR-tagged molecule and 5 µM Sfp Synthase. Unreacted handling molecules were washed off with an appropriate measurement buffer.

5.2 Surface Preparation

Coverglass slips were utilized as surfaces to immobilize molecules of interest for AFM probing. Surfaces were first sonicated in 50 %(v/v) aqueous 2-propanol for 15 min and oxidized in a solution of 50 %(v/v) hydrogen peroxide (30%) and sulfuric acid for 30 min. They were then washed in deionized water, dried in a nitrogen stream and silanized by incubation in (3-Aminopropyl) dimethylethoxysilane (ABCR, Karlsruhe, Germany, 1.8 %(v/v) in Ethanol) for 1 h. Silanized surfaces were then rinsed sequentially in toluene, isopropanol, and finally deionized water to remove excess silane, and subsequently baked at 80 °C for 30 min to catalyze the covalent attachment of silane to the glass surface. Surfaces were stored long-term under inert Ar_2 to prevent oxidative degradation of reactive amine groups.

Surfaces were freshly functionalized immediately before experimentation. Silanized surfaces were optionally incubated in 150 mM sodium borate buffer, pH 8.5 for 30 min in order to deprotonate primary amine groups prior to PEGylation.

For SMFS experiments, a silicone mask with 1.5 mm-diameter wells spaced 1 mm apart was sonicated for 15 min first in deionized water and then 50 %(v/v) aqueous 2-propanol. The mask was then affixed to the silanized surface and briefly baked at 60 °C to improve mask adhesion. The wells were incubated in 25 mM heterobifunctional PEG crosslinker [158, 159] with N-hydroxy succinimide and maleimide groups (molecular weight 5000 Da, Rapp Polymere, Tübingen, Germany) dissolved in either 150 mM sodium borate buffer, pH 8.5 or 50 mM HEPES, pH 7.5. Unreacted PEG crosslinkers were rinsed off with deionized water. PEGylated surfaces were either reacted directly to handling molecules, or further functionalized with Coenzyme A (CoA) dissolved in Coupling Buffer for 1 h. Unreacted molecules of interest were rinsed off with an appropriate measurement buffer, and unreacted CoA was rinsed off with deionized water. CoA-functionalized surfaces were then reacted via Sfp to ybbR-tagged molecule. In a typical reaction, cantilevers were incubated for 1 - 2 h in a solution of 1x Sfp Reaction Buffer containing 100 µM ybbR-tagged molecule and 100 µM Sfp Synthase. Unreacted molecules were washed off with an appropriate measurement buffer.

For SMC&P experiments, a PDMS microfluidic system – based on the system described by Kufer et al. [98] – was assembled. The PDMS chip was affixed to the glass surface either after whole-surface PEGylation (as per an older protocol, similar to SMFS surface preparation) or directly to the unPEGylated surface prior to in-channel PEGylation (as per an updated protocol). In either case, the PDMS chip was bonded to the surface briefly at 60 °C for 10 min. For in-channel PEGylation, the Depot and target channels were incubated in 25 mM heterobifunctional PEG crosslinker [158, 159] with N-hydroxy succinimide and maleimide groups (molecular weight 5000 Da, Rapp Polymere, Tübingen, Germany) dissolved in either 150 mM sodium borate

buffer, pH 8.5 or 50 mM HEPES, pH 7.5. Unreacted PEG crosslinkers were flushed from the channels with filtered deionized water.

Functionalized depot and target channels were finally covalently modified with freshly-reduced anchoring molecules with reactive terminal cysteine or thiol groups. Anchoring molecules were incubated in the channels for 1 h. Both channels were then flushed with filtered PBS to remove unbound anchoring molecules. 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 fluorescent transfer construct was diluted to an approximate concentration of 1 nM in 1x PBS with 0.05 g/L BSA and 0.01% TWEEN20 and incubated in the depot channel for 1 h. The depot channel was then extensively flushed with 1x PBS to clear the solution and remove unbound- or nonspecifically-bound transfer construct. The microfluidic system was then removed and the surface submerged in an appropriate measurement buffer.

5.3 AFM Measurement

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, were used. 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 and calibrated in solution using the equipartition theorem [40, 160]. Pulling velocities were set in a range from 200 nm/s to 3200 nm/s for both SMFS and SMC&P experiments. The positioning feedback accuracy is ±3 nm, although long-term deviations may arise due to thermal drift. For SMC&P experiments, typical times for one Cut-and-Paste cycle amount to approximately 3 s.

5.4 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-color laser (Toptica Photonics, Gräfelfing, Germany), which is capable of emitting light at 405 nm, 488 nm, 561 nmand 640 nm through one single fiber mode. Specifically, blue excitation at 488 nm, green excitation at 561 nm, and red excitation at 640 nm with estimated intensities of approximately 10 W/cm^2 were utilized to monitor the fluorescence of various molecules. 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 multicolor 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.

5.5 SMC&P Experiments

SMC&P patterns were written typically in 200 - 1000 transfer cycles with 100 - 200 nm spacing between each deposition point. The pulling speed in the depot was set to 2,000 - 3.200 nm/s and in the target to 200 nm/s. This corresponds to approximate surface contact times [161] (dependent on approach/retraction velocity, indentation force and substrate stiffness) of 5 ms and 80 ms, respectively, and should allow for ligand binding. Considering a single handling molecule being bound to the cantilever tip and estimating its localization in a half sphere with r = 30 nm (approximate length of PEG 5000 Da linker), the local concentration of cantilever-immobilized molecules would be in the micromolar range, which is usually several orders of magnitude higher than the K_d of high-affinity receptor-ligand pairs. 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.

Rupture forces and loading rates were evaluated from AFM force distance curves that were recorded for each pickup and deposition process utilizing the WLC model [162] with quantum mechanical correction (force spectroscopy data were evaluated in Python 2.7, Python Software Foundation).

Part III

Results

Chapter 6
Enzyme Activity and Fluorescent Systems

Summary

A unifying paradigm of fluorescence was chosen to measure single-molecule catalytic activity, as it is both feasible with SMC&P and adaptable to diverse readout approaches. However, most enzymes do not typically participate in fluorescent or bioluminescent processes. Hence, strategies for fluorescence-based readout of catalytic activity were specifically developed and characterized for various enzymes. The assays were initially tested and optimized in bulk, with an eventual goal of integrating the enzymes into SMC&P with fluorescent readouts. Conceivably, if an enzyme could be implemented in SMC&P as part of the transport cargo, the activity of individual enzymes could be observed on the surface in relatively simple single-enzyme reactions with a fluorescent output signal, or in tandem with other enzymes in a network.

In this chapter, the primary strategies of acquiring fluorescence readout for the enzymes of interest are described. The major results from the development of these enzyme activity assays are subsequently presented and discussed. Additionally, a strategy for improved expression of a recombinant enzyme are described. Details regarding the molecular biology techniques used to prepare all enzymes as well as all measurement methodologies ares found in Chapter 4. Buffers and DNA sequences are found in B.

6.1 Pfu-E10 Incorporation of Labeled Nucleotides

A strategy for detection of DNA polymerase (DNAP) activity was devised (Figure 6.1). The evolved DNAP Pfu-E10 can incorporate Cy3- and Cy5-labeled dCTP into a growing DNA strand during replication. Hence, a primed DNA template was incubated with Pfu-E10 DNAP in the presence of unlabeled dNTPs as well as dCTP-Cy3 and dCTP-Cy5, resulting in replicated DNA adorned with Cy3 and Cy5 labels. In order to observe simple fluorescence, only dCTP-Cy3 or

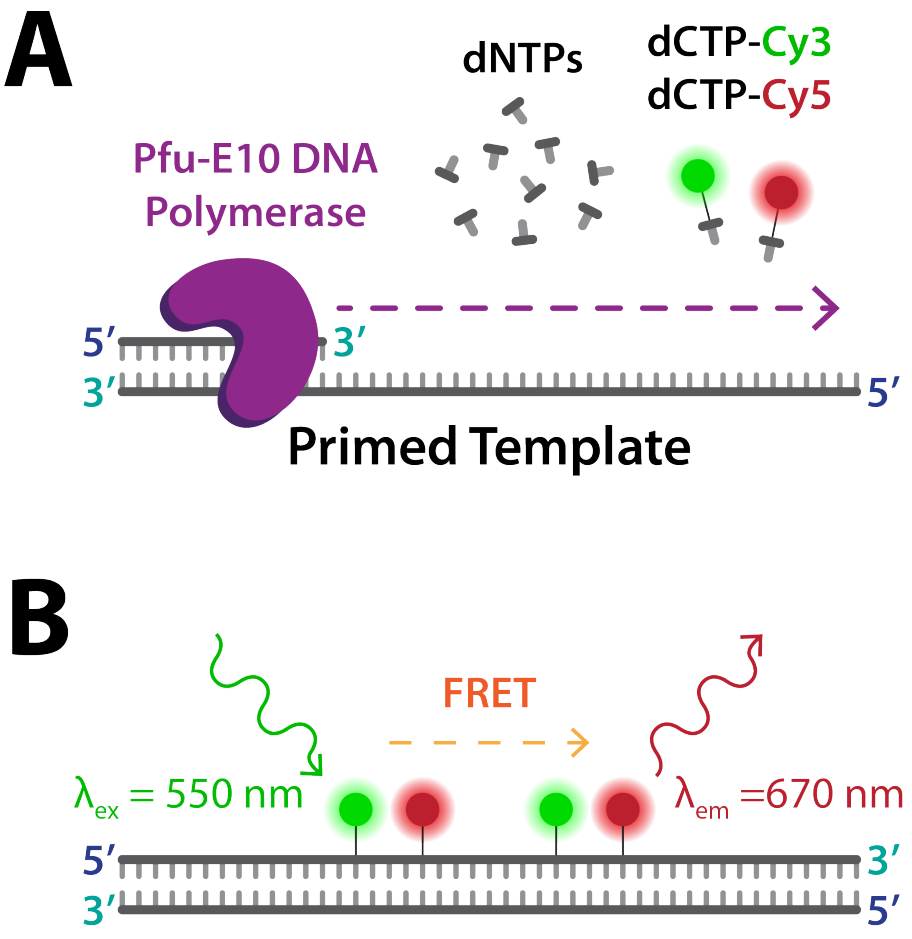


Figure 6.1: Incorporation of labeled nucleotides by Pfu-E10 evolved polymerase. (a) This strategy utilizes Pfu-E10, a primed single-stranded DNA template, unlabeled dNTPs and Cy3- and Cy5-labeled dCTP. (b) Pfu-E10 polymerizes a growing DNA strand labeled with Cy3 and Cy5 molecules, enabling observation of a FRET signal from nearby fluorophores. Alternatively, dCTP-Cy3 or dCTP-Cy5 may be omitted to instead observe simple fluorescence.

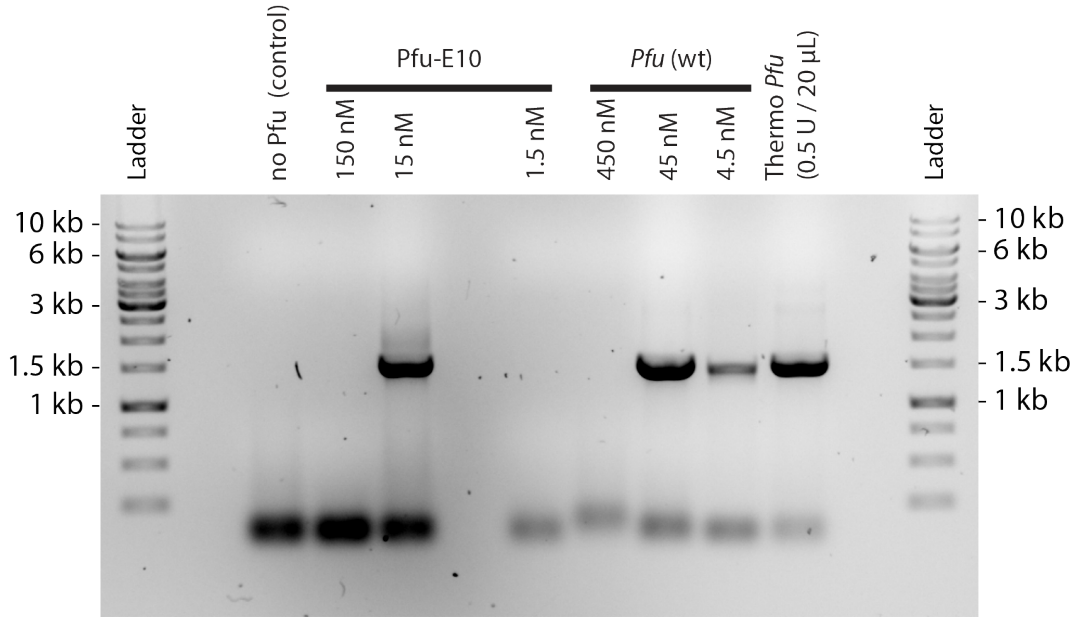


Figure 6.2: PCR with *Pfu* variants, analyzed by agarose gel electrophoresis. PCRs were performed with recombinant Pfu-E10 and wildtype *Pfu* DNA polymerases at various concentrations, as well as commercially-available wildtype *Pfu* as a positive control. GeneRuler 1kb DNA Ladder was used as a mass standard. Expected product size is 1488 bp.

dCTP-Cy5 were included in the reaction. However, by stochastic incorporation of dCTP-Cy3 and dCTP-Cy5 simultaneously in the replication reaction, it would then be possible to observe FRET-based fluorescence from nearby Cy3 and Cy5 fluorophores. Importantly, as SMC&P is undertaken at room temperature, a strategy for low-temperature replication was required. Single-stranded DNA was pre-annealed to a short primer sequence to allow DNA replication in the absence of heat cycling.

The activity of recombinant tagged wildtype *Pfu* and Pfu-E10 polymerases was first compared to confirm the activity of the evolved mutant (Figure 6.2). Both polymerases were challenged to exponentially amplify a 1488 bp DNA sequence via PCR, and the reactions were examined with agarose gel electrophoresis. Identical reactions with varying concentrations of each polymerase were run using a robust plasmid template, primers, and reaction protocol. A reaction with commercially-available native *Pfu* DNAP (Thermo Scientific, Massachusetts, USA) was additionally run as a positive control. GeneRuler 1kb DNA Ladder (Thermo Scientific) was used as a mass standard to estimate the size of observed bands from each reaction.

Reactions from both wildtype *Pfu* and evolved Pfu-E10 with intermediate concentrations of polymerase display PCR activity comparable to the positive control reaction, and the major product with an expected size of 1488 bp appears readily visible. Importantly, this indicates that the evolved polymerase has comparable activity to the wildtype variant, and is in good agreement with the observed behavior of Pfu-E10 previously demonstrated by Ramsay et al. [118]

Interestingly, both polymerases tested in a dilution series appear to only facilitate exponential amplification of PCR products within a limited concentration range. At a low enough concentration, the polymerase presumably becomes the limiting factor in exponential amplification. Simply too little DNA is replicated with each cycle, and the final mass of DNA is below the detection limits of gel electrophoresis. This is evidenced by the reaction with 4.5 nM wildtype *Pfu*, which does produce a visible product band, albeit in a noticeably lower concentration than the positive control reaction. Additionally, reactions loaded with higher concentrations appear unable to efficiently amplify DNA in PCR. One possible explanation for this behavior is that higher concentrations of DNAP affect primer stability. Alternatively, individual DNAP molecules may begin to compete with each other for DNA binding to the point that the replication complex is not sufficiently stable for rapid DNA amplification. This is somewhat remarkable, as it indicates that the optimal concentration of *Pfu* DNAP for successful PCR is a relatively narrow window of a single order of magnitude. In both cases of too high or too low DNAP concentration, the absence of a product band is also not likely explained by degradation of PCR products; all visible bands appear sharp, and a smear that is typical of degraded oligomers is not visible in the gel.

The evolved Pfu-E10 DNAP was then tested to confirm its ability to incorporate Cy3- and Cy5-labeled dCTP nucleotides, as well as its specificity for exactly those molecules. This was accomplished by PCR, and the products were assessed by agarose gel electrophoresis (Figure 6.3). The reactions contained 200 200 µM of each unlabeled dNTP supplemented with relatively low concentrations (200 nM - 2 µM) of fluorescently labeled dNTPs. The concentration of labeled dNTPs was kept low in comparison to unlabeled dNTPs to prevent polymerase stalling in the event that a particular labeled nucleotide is not an optimal substrate. The gel was then imaged sequentially with UV light for DNA detection, followed by green LED excitation for Cy3 and finally red LED excitation for Cy5. The composite multichannel image shows the overlap of the signals from the three channels.

Analysis of background signal contributed by the free unincorporated fluorescently-labeled nucleotides shows that each contributes a large signal in their respective channels, visible as diffuse clouds of fluorescence at these exposure settings. Notably, the stock of Cy5-labeled dATP appears to be partially hydrolyzed or degraded, as a fluorescent signal is visible both lower in the gel near the labeled dCTPs as well as higher in the gel near the PCR products. Additionally, all samples were loaded with 6x DNA Gel Loading Dye (Thermo Scientific), which contains the dye molecules bromophenol blue and xylene cyanol. These dyes contribute background signals of their own in both the red and green channels, visible as diffuse clouds at approximately 4 kb and 500 bp, respectively. Interestingly, both of these dyes produce a negative signal in the UV channel. Despite these sources of background signal, this likely does not interfere with detection of DNA.

All PCRs produce a clearly visible DNA band with an estimated size of approximately 1500 bp, consistent with the expected size of 1488 bp. Additionally, reactions with Cy5-labeled dCTP results in a clear colocalizing band in the red channel (lanes 2, 5 and 8). While reaction with Cy3-labeled dCTP at a higher concentrations results in an increase in signal from the

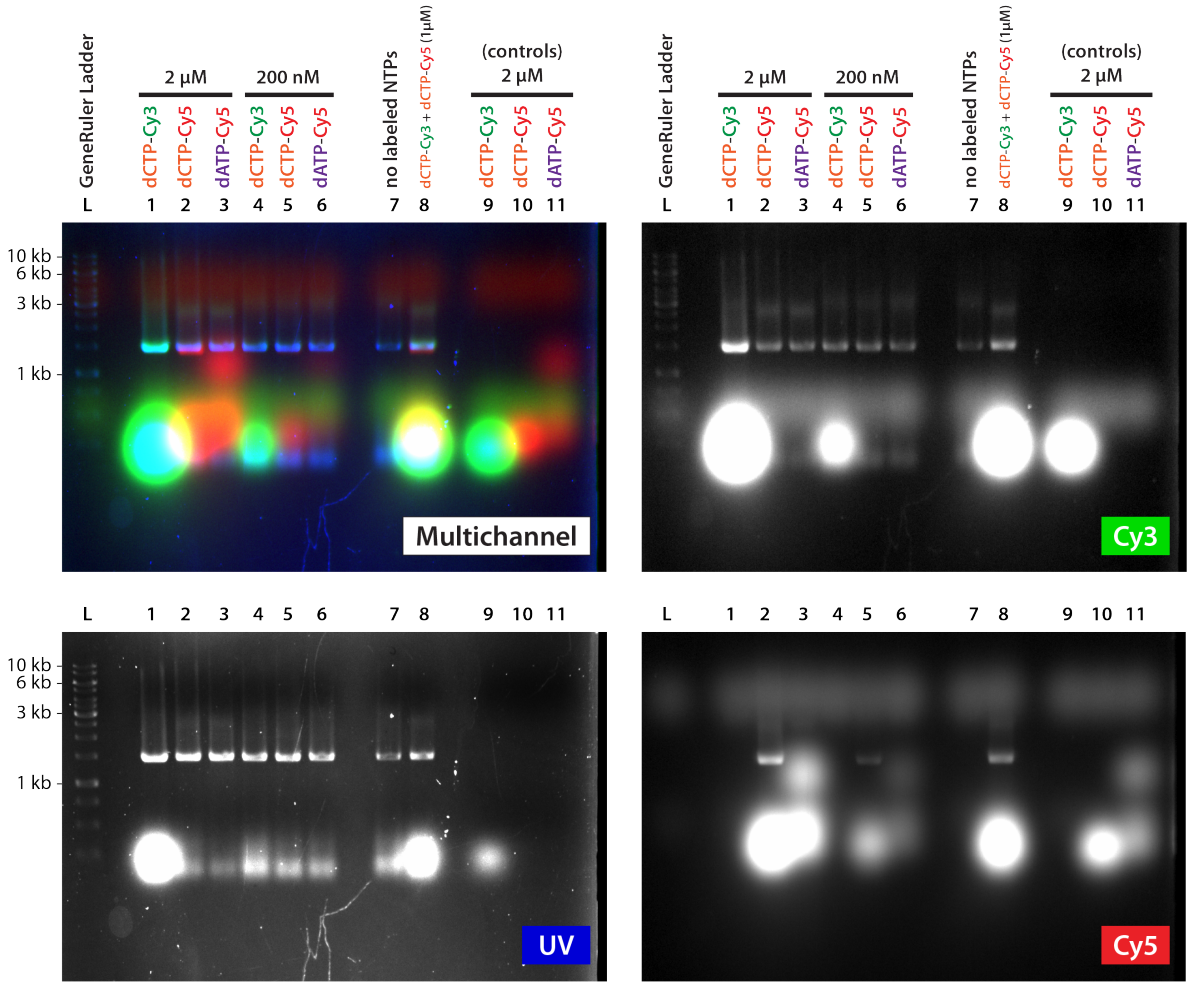


Figure 6.3: Incorporation of labeled nucleotides by Pfu-E10, analyzed by agarose gel electrophoresis. PCR reactions were performed with Pfu-E10 in a standard *Pfu* reaction protocol additionally supplemented with relatively low concentrations (200 nM - 2 µM) of various labeled nucleotides (lanes 1 - 8). Reactions were incubated separately with dCTP-Cy3, dCTP-Cy5 and dATP-Cy5 at concentrations of 2 µM (lanes 1 - 3) and 200 nM (lanes 4 - 6), as well as with both dCTP-Cy3 and dCTP-Cy5 at a concentration of 1 µM (lane 8) and with all fluorescent nucleotides omitted (lane 7). As additional controls, stock solutions of the individual labeled nucleotides were analyzed to pinpoint sources of background signal unassociated with PCR activity (lanes 9 - 11). GeneRuler 1kb DNA Ladder (L) was used as a mass standard. The gel was imaged with UV exposure for DNA, green LED excitation for Cy3 and red LED excitation for Cy5. The multichannel image was assembled by simple overlay of the three channels, with UV channel in blue, green LED channel in green, and red LED channel in red. Expected product size is 1488 bp.

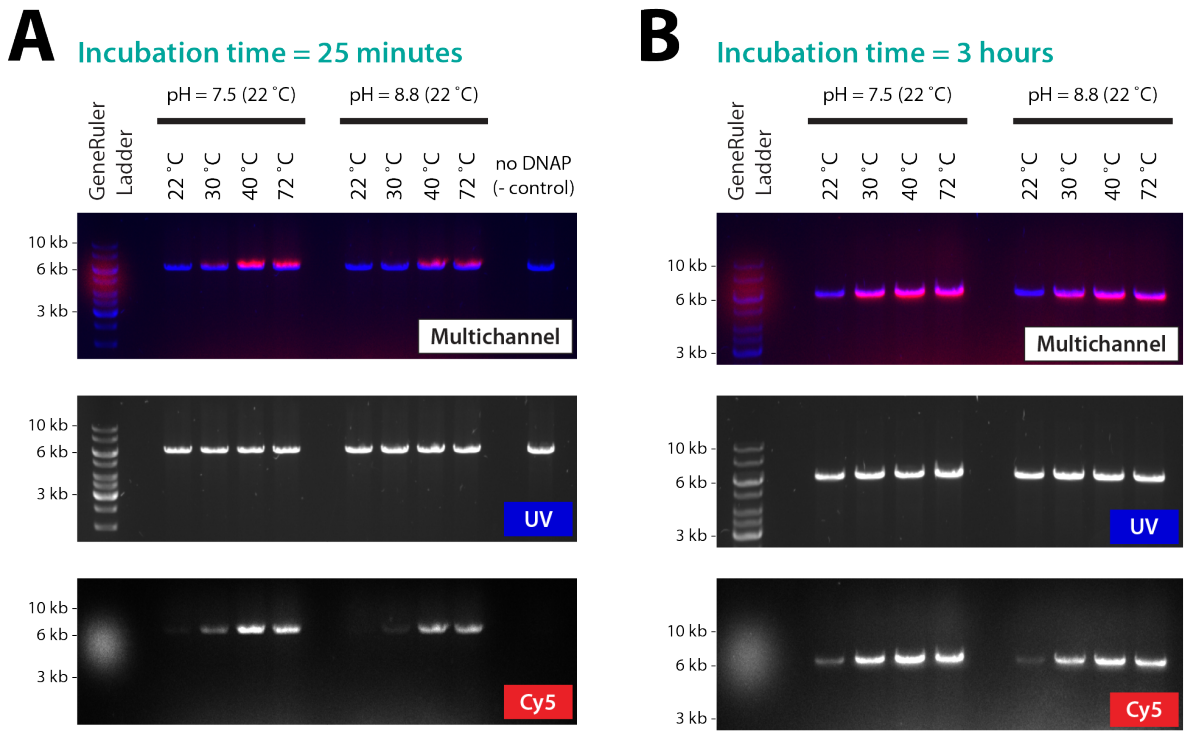


Figure 6.4: Temperature dependence of Pfu-E10 DNAP, analyzed by agarose gel electrophoresis. Linearized plasmid was labeled with dCTP-Cy5 by Pfu-E10 at various temperatures and with reaction buffers of two set pH values at 22 °C. Additionally, a negative control sample that lacked DNAP to carry out the polymerization reaction was included. Samples were analyzed after (a) 25 min incubation, and later after (b) 3 h incubation. The gels were imaged with UV exposure for DNA and red LED excitation for Cy5. The multichannel images were assembled by simple overlay of the two channels, with UV channel in blue and red LED channel in red. Plasmid sequence length is 6103 bp.

visible band in the green channel (lanes 1 and 8), it is unclear if a distinct Cy3 signal is similarly present with lower concentrations of dCTP-Cy3 (lane 4). This is likely due to fluorescent crosstalk of DNA stained with Roti-GelStain; this nucleotide stain is optimally imaged with UV light, but it is evidently still excited by the green LED, as the reaction without any fluorophores present (lane 7) also produces a signal in the green channel. However, incubation at higher concentrations (1 - 2 µM) of Cy3 results in a clear increase in signal from those bands in the green channel. Therefore, it may be inferred that Cy3 is also incorporated at lower concentrations (200 nM), although the signal is lost in the background crosstalk from DNA staining.

Pfu-E10 appears readily able to utilize base-labeled dCTP-Cy3 and dCTP-Cy5. In contrast, the reactions with Cy5-labeled dATP produced no detectable colocalization signal in the red channel, thereby indicating that Pfu-E10 is unable to utilize this molecule as a substrate for DNA replication. These findings are in good agreement with Ramsay et al. [118], where it was determined that Pfu-E10 has high specificity for dCTP-Cy3 and dCTP-Cy5 and is not freely able to incorporate other labeled nucleotides.

As *Pfu* DNAP is originally sourced from a thermophilic bacterium, it is best suited for catalysis at higher temperatures. This is exhibited by its application in PCR, where the optimal

extension temperature is typically 72 °C and the enzyme remains stable above 95 °C. However, SMC&P experiments are undertaken at ambient room temperature, which is typically 18 - 24 °C. Therefore, the ability of the polymerase to operate in this temperature range was examined (Figure 6.4). Moreover, the pH of buffers typically fluctuates over such large temperature ranges. A standard buffer system for *Pfu* DNAP typically has a basic pH at room temperature (pH ≈ 8.8 at 22 °C), and when used in PCR at higher temperatures the pH drops to within a more neutral range (pH ≈ 7.5 at 72 °C). Hence, in order to best facilitate DNAP activity at lower temperatures, the buffer system also required tuning.

The ability of Pfu-E10 to polymerize sticky overhangs was challenged with a range of temperatures, most of which far below the optimal catalysis temperature of the enzyme. Additionally, two buffer systems were tested, one calibrated to pH 7.5 at 22 °C and the other to pH 8.8 at 22 °C. A plasmid was first linearized with the restriction enzyme XhoI. This enzyme recognizes the palindromic sequence 5' CTCGAG 3' and cleaves the phosphodiester bonds of both strands between C_{+1} and T_{+2}. The resulting 5' sticky overhangs may be filled in by DNAP with the sequence 5' TCGA 3'. The reactions were supplied with 10 mM dTTP and 20 µM dCTP-Cy5, and individual reactions were incubated at 22 °C, 30 °C, 40 °C or 72 °C. Samples of the reactions were analyzed after 25 min of incubation, followed by further analysis after 3 h of incubation. The gels were then imaged sequentially with UV light for DNA detection, followed by red LED excitation for Cy5. The composite multichannel image shows the overlap of the signals from the two channels.

After only 25 min, extensive labeling is observed at higher temperatures under both buffer systems. However, only very minimal labeling is observed at 22 °C and 30 °C, with stronger labeling occurring in the lower-pH buffer. After 3 h of incubation, the labeling for all reactions appears to be superior - as would be expected for relatively slow biological reactions. Of the two reactions at 22 °C, the reaction undertaken in the buffer of lower pH appears to be more efficient, in agreement with the 25 min-samples.

It is not surprising that Pfu-E10 is less active at lower temperatures; the structure of wildtype *Pfu* DNAP was evolved to not only withstand but thrive in higher temperatures, and accordingly it likely contains many features that are at lower temperatures are too rigid to function at peak performance. Experimental conditions of SMC&P are therefore not ideal for this enzyme. However, as seen at lower temperatures in Figure 6.4, the enzyme activity is still in a detectable range.

In order to maximize enzyme detection on the level of single molecules at low temperatures, a DNA template with a much longer stretch of single-stranded sequence was designed. Two synthetic DNA molecules - a template strand and a much shorter primer strand - were annealed to form a complex consisting of a 16 bp double-stranded primed region followed by 117 bases of a single-stranded template. According to this intended strategy, DNAP would bind the double-stranded primed region and proceed to replicate the single-stranded region. However, a 117-base stretch of single-stranded DNA has a high likelihood of forming transient hairpins or dimers, which could interfere with DNAP activity in the absence of thermally-driven DNA melting.

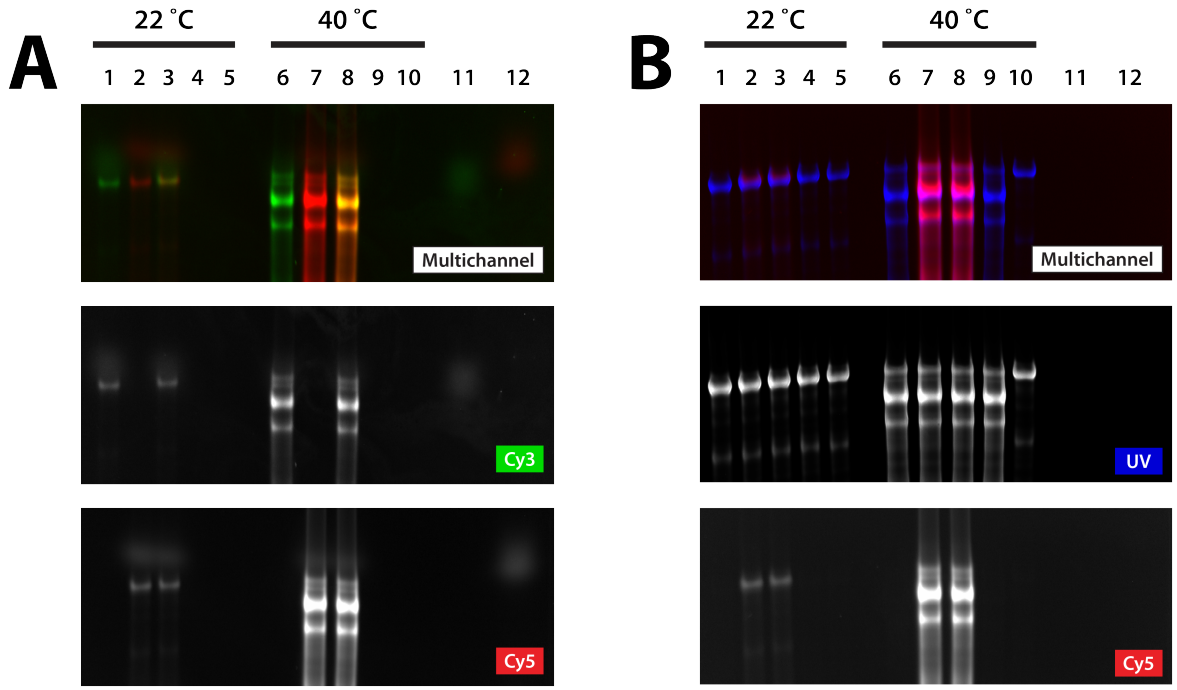


Figure 6.5: Low-temperature DNAP template, analyzed by native PAGE. (a) Pfu-E10 was incubated with the low-temperature template in reactions supplemented with dCTP-Cy3 (lanes 1 and 6), dCTP-Cy5 (lanes 2 and 7) and both Cy3- and Cy5-labeled dCTP (lanes 3 and 8). Control reactions without labeled nucleotides (lanes 4 and 9) and additionally without Pfu-E10 (lanes 5 and 10) were also prepared. Reactions were incubated for 3 h at 22 °C and 40 °C. Samples of all reactions were analyzed by native PAGE, with free dCTP-Cy3 (lane 11) and dCTP-Cy5 (lane 12) additionally analyzed to control for sources of background signal. The gel was imaged with green LED excitation for Cy3 and red LED excitation for Cy5. The multichannel image was assembled by simple overlay of the two channels, with green LED channel in green and red LED channel in red. (b) Following green- and red-channel capture, the same gel was incubated in a solution of ethidium bromide to directly stain DNA. The gel was imaged with UV exposure for DNA and red LED excitation for Cy5. The multichannel image was assembled by simple overlay of the two channels, with UV channel in blue and red LED channel in red. As ethidium bromide is strongly fluorescent in the green channel due to crosstalk, it was not possible to capture a signal from Cy3 at this stage. The noticeable bend in the gel arises from matrix relaxation when removed from the plastic casing, causing the gel to slowly deform over time.

Pfu-E10 was challenged to replicate the single-stranded stretch of the template DNA and incorporate labeled nucleotides (Figure 6.5). One set of reactions was incubated at 22 °C and an otherwise identical set were incubated at 40 °C. The primed template was incubated with Pfu-E10, free dNTPs, as well as labeled dCTP: dCTP-Cy3, dCTP-Cy5, or both dCTP-Cy3 and dCTP-Cy5 simultaneously. Additionally, control reactions with only unlabeled dNTPs were prepared. The annealed DNA template in the absence of Pfu-E10 and labeled nucleotides was examined to control for a band-shift in the DNA. Free dCTP-Cy3 and dCTP-Cy5 were also examined to identify any sources of background signal from unincorporated labeled nucleotides. Due to the much smaller size of this DNA template compared to PCR products or plasmids, the reactions were analyzed by native PAGE instead of agarose gel electrophoresis to improve resolution of separated DNA molecules.

Following reaction with Pfu-E10 at both 22 °C and 40 °C, fluorescent bands from Cy3 and Cy5 are visible, indicating incorporation of fluorescent nucleotides and at least partial DNA replication. However, as evidenced by the lower signal in both the green and red channels in Figure 6.5a as well as the lack of a noticeable band shift compared to unreacted template strand in Figure 6.5b, Pfu-E10 is only minimally active at 22 °C.

In agreement with the results observed in Figure 6.4, Pfu-E10 activity quickly increases at only slightly higher temperatures, in this case 40 °C - visible by strong fluorescent labeling and a noticeable band shift. The reactions incubated at 40 °C with Pfu-E10 present display a marked downwards band shift compared to the unreacted template.

This behavior might be somewhat surprising, as DNA is expected to travel in a gel matrix according to mass, with smaller molecules traveling faster and migrating further down the gel. Although the replicated DNA is indeed more massive than unreplicated DNA, the migration patterns of single-stranded and double-stranded DNA are not comparable. While double-stranded DNA reliably migrates according to size under native conditions, the migration pattern of single-stranded DNA is far less predictable. This is due to the variable final folded states of single-stranded DNA - which may form hairpins or dimers between molecules, depending on the degree of self-complementarity. Therefore, the unreacted template DNA is not necessarily expected to migrate faster than the replicated DNA, even though it is less massive.

Taking this into account, it may be surmised that the major bands in lanes 6 - 9 constitute the fully replicated template that is now a fully double-stranded helix decorated with Cy3 and Cy5 labels. However, the smear pattern in these lanes indicates that replication was not completed for many of the templates complexes. This is likely due to a combination of slower polymerase activity at lower temperatures combined with the intrinsic off-rate between the polymerase and the DNA template, resulting in fewer bases successfully replicated before the polymerase unbinds from the template strand.

Furthermore, the major bands in lanes 1 - 4 and the upper-most bands in lanes 6 - 9 migrate indistinguishably from the unreacted template samples in lanes 5 and 10. In the case of the reactions incubated at 40 °C, this indicates that a significant fraction of the template DNA molecules are nearly completely unreplicated. In the case of the reactions incubated at 22 °C, this suggests that this template can only be very minimally replicated at room temperature. This

may be due simply to extremely low DNAP activity at this temperature. Another possibility is that at such low temperatures, self-annealed hairpin structures or dimers between template molecules may be stable enough that the template DNA does not reliably adopt a single-stranded conformation - which is required for replication to occur. These two explanations are not mutually exclusive, and is likely a combination of the two that results in low or inconsistent DNAP activity.

Nevertheless, the strategy of a primed template for low-temperature DNA replication is in principle a viable substrate for Pfu-E10 DNAP. Further optimization of the template strand - possibly by decreasing the self-complementarity to encourage a truly single-stranded conformation - could increase Pfu-E10 activity even at room temperature.

6.2 T7 RNA Polymerase *In Vitro* Transcription Reactions

A system to detect the activity of T7 RNA polymerase (T7 RNAP) by an increase in the fluorescence of malachite green (MG) was designed and extensively developed. The results presented in Section **??** largely use data that were originally presented in the author's master thesis [155]. The work therein ties very directly into more recent results, as well as constitutes a basis for an argument for DNA-free SMC&P. Hence, select data have been adapted for discussion.

Recombinant tagged T7 RNAP constructs were expressed in *E. coli* and purified using a variety of affinity chromatography strategies. The catalytic activities of recombinant T7 RNAP constructs were determined by *in vitro* transcription (IVT) reaction. Reactions were incubated at 22 °C to match the temperature under which SMC&P experiments are undertaken (ambient temperature of approximately 18 - 24 °C). RNA aptamer products were examined directly by gel electrophoresis or indirectly by MG fluorescence over time with $\lambda_{ex} = 628\,\text{nm}$ and $\lambda_{em} = 654\,\text{nm}$ on a plate reader. The production of RNA aptamers of well-defined lengths was achieved by run-off transcription (Figure 6.6).

As fluorescence is measured in arbitrary units (AU) in plate reader assays, the numerical fluorescent readout is not intended to be understood in absolute terms. Reactions measured together within one experiment were carried out under identical conditions, except for tested variables. However, individual experiments generally had slightly different conditions in either buffer conditions, reagent conditions or enzyme concentration. Moreover, as the employed plate reader device is subject to ambient temperature fluctuations, the measurement temperature was likely also varied between experiments. Lastly, the excitation laser intensity or detection gain is likely varied. Therefore, distinct experiments should not be compared to each other, and data should only be numerically compared within a single experiment.

A purified T7 RNAP construct with N-terminal ybbR and C-terminal GCN4 tags was first assessed for general enzyme activity, and to confirm as a proof of principle that the designed IVT scheme functions to produce RNA aptamers that can enhance MG fluorescence (Figure 6.7). A pET21a plasmid with a single MG aptamer sequence downstream of the T7 promoter sequence was designed and cloned. Following amplification in *E. coli*and subsequent miniprep, the

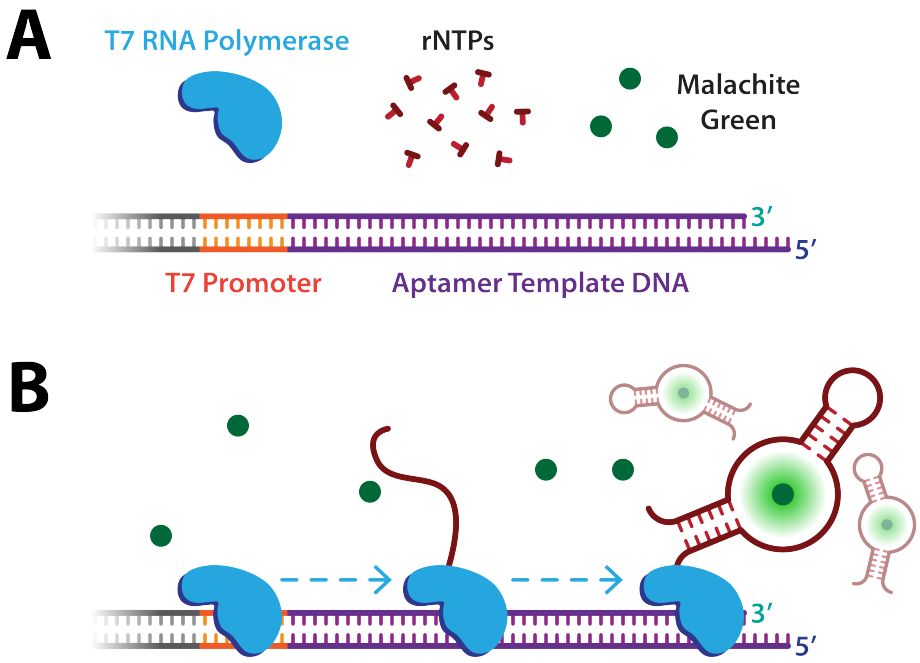


Figure 6.6: RNA aptamer generation by run-off transcription. (a) This strategy employs T7 RNAP, a linearized DNA template with the sequence for an MG aptamer, rNTPs and MG. (b) Linearized plasmid templates enabled production of RNA aptamers of well-defined lengths by T7 RNAP. RNA transcripts spontaneously anneal due to self-complementarity. MG intercalates in the binding pocket and its structure is stabilized, resulting in a significant increase in MG fluorescence.

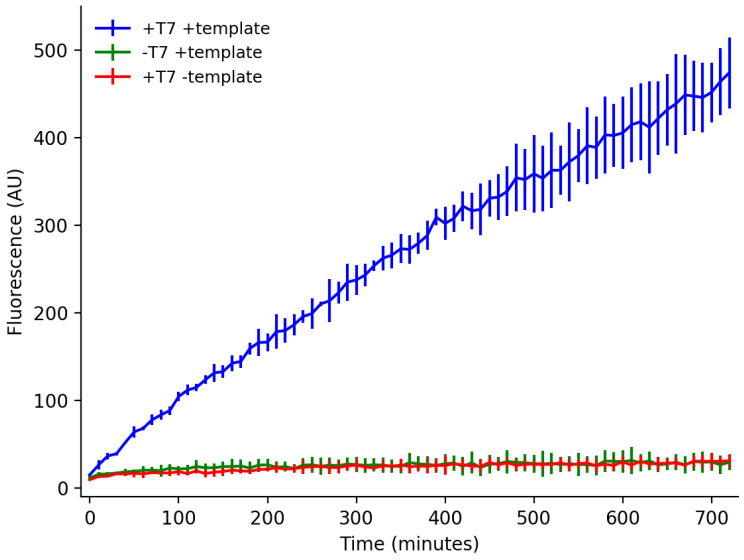


Figure 6.7: MG aptamer generation by IVT reaction, examined with MG fluorescence via plate reader. IVT reactions contained T7 RNAP and the DNA template containing the MG aptamer sequence (blue trace), or omitted either T7 RNAP (green trace) or the DNA template (red trace). Measurements were performed in triplicate, and error bars display the standard deviation. Originally presented in [155] and the data have been adapted for this work.

plasmid was linearized via an EcoRI restriction site downstream of the MG aptamer sequence to enable run-off transcription. The T7 RNAP construct was incubated with the linearized plasmid in the presence of free ribonucleotides to allow transcription of the MG aptamer sequence. Free MG was additionally present in solution, which provided a fluorescent readout upon correct aptamer folding and subsequent binding to MG. The fluorescence was measured via plate reader in intervals of 10 min. As negative controls to identify possible background sources of fluorescence, this reaction was duplicated with either the DNA template or T7 RNAP omitted from the reaction. The negative control reactions displayed no significant increase in fluorescence over time, therefore indicating that the buffer, unbound MG (at least at this concentration), and other reaction components do not on their own significantly contribute to the measured fluorescence. In contrast, the complete reaction with T7 RNAP and the DNA template produced a steadily increasing fluorescence signal over hundreds of minutes. This demonstrates that the IVT reaction is robust over this timescale and at this temperature of 22 °C.

Additionally, as the fluorescent signal remains roughly linear and does not reach a plateau at these timescales, two things may be inferred: First, the monomeric ribonucleotides in solution do not appear to be limiting, even after > 10 h of continuous catalysis. Secondly, the lifetime of the DNA template, T7 RNAP, MG and RNA aptamers themselves is at least long enough that they do not spontaneously degrade in solution at this timescale and under these conditions.

Several additional pET21a plasmids with MG aptamer repeats were designed and cloned. Aptamer sequence repeats in the DNA templates were separated by regions of T and C nucleotides to encourage proper folding of individual aptamer repeats and discourage inter-aptamer annealing. The resulting library of DNA templates contained either a single aptamer sequence (1xMG), two consecutive repeats (2xMG) or three consecutive repeats (3xMG). Addition of more aptamer repeats was not achieved during the cloning process, likely due to the increasingly repetitive nature of the plasmids in the multiple cloning region.

The IVT reaction efficiency of each DNA template was assessed by gel electrophoresis (Figure 6.8). Equimolar amounts of each DNA template were incubated in otherwise identical IVT reactions at 22 °C for 4 hours. An additional reaction lacking T7 RNAP was included as a negative control. IVT samples and an agarose gel was prepared according to the standard protocol found in Chapter 4. RiboRuler Low Range Ladder RNA standard (Thermo Scientific, Massachusetts, USA) was used as a mass standard to estimate the size of the transcripts from each reaction.

As the DNA templates are linearized plasmids, each reaction should produce RNA transcripts of well-defined lengths. Based on the sequence length between the +1 position in the T7 promoter and the EcoRI restriction digest site, the DNA templates have expected transcript lengths of 119 bases for 1xMG, 205 bases for 2xMG and 237 bases for 3xMG. Analysis of RNA products shows a single clean band of a uniform length from each reaction. Moreover, the size of each product is in the expected range. The relative difference in signal between each band is furthermore to be expected; considering that initiation is the rate-limiting step in T7 RNAP transcription, it is likely that transcripts are produced in equimolar amounts in each reaction,

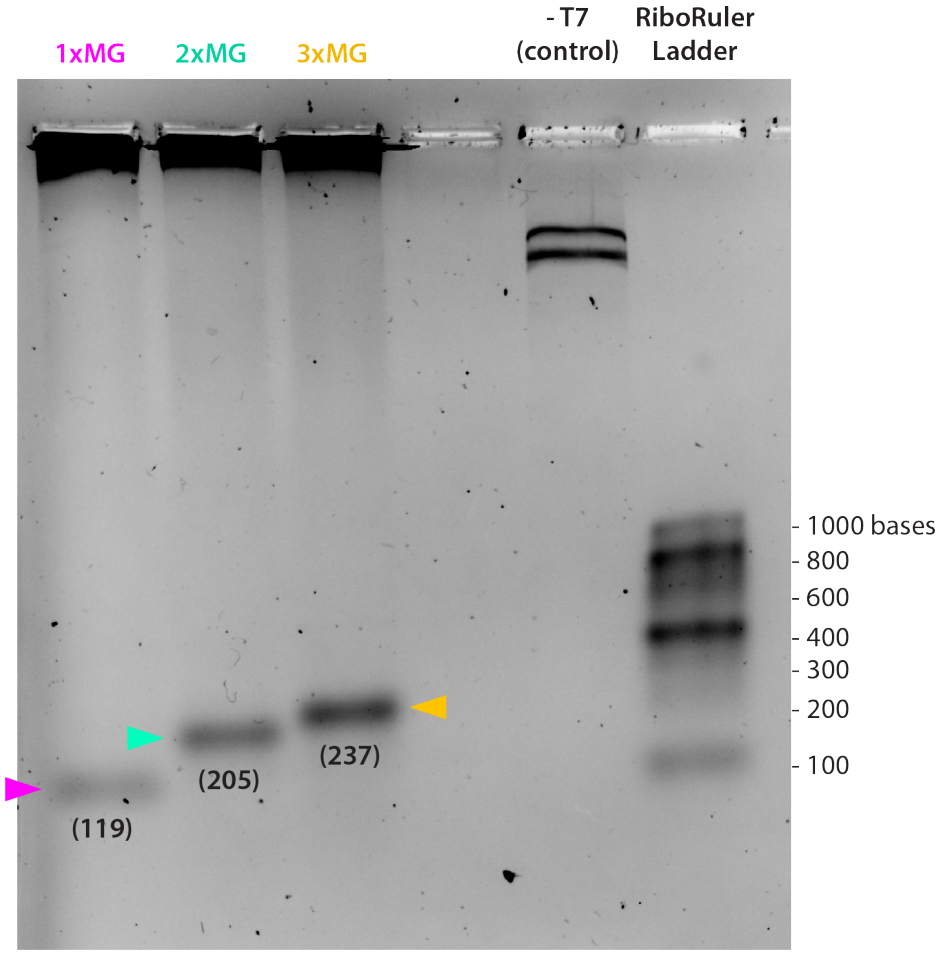


Figure 6.8: IVT RNA products, examined by agarose gel electrophoresis. Identical reactions with equimolar amounts of linearized 1xMG, 2xMG or 3xMG plasmid template were prepared, in addition to a negative control reaction that omitted T7 RNAP (-T7). Bands are indicated with arrows for 1xMG (magenta), 2xMG (turquoise) and 3xMG (orange) IVT reactions. The negative control reaction produces no observable nucleotide products. Expected transcript sizes in bases are displayed below the visible band of that sample. RiboRuler Low Range Ladder was additionally loaded as an RNA standard for size estimation.

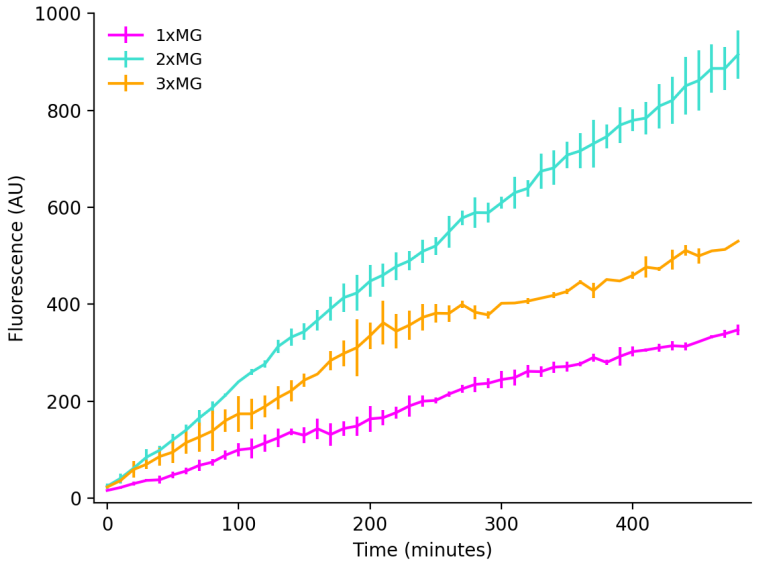


Figure 6.9: Fluorescence enhancement efficiency of variable MG repeats, examined with MG fluorescence via plate reader. Identical reactions with equimolar amounts of linearized 1xMG (magenta trace), 2xMG (turquoise trace) or 3xMG (orange trace) plasmid template were examined. Measurements were performed in triplicate, and error bars display the standard deviation. Originally presented in [155] and the data have been adapted for this work.

and the longer transcripts will have a higher mass and hence produce a stronger signal in gel electrophoresis. Taken together, this demonstrates a robust, precise IVT strategy to produce transcripts of well-defined sizes.

Interestingly, a large signal of nucleic acids is found in the pockets loaded with T7 RNAP-containing samples. This signal is likely from the plasmid template stuck in the pocket due to a strong interaction with T7 RNAP. This is further evidenced by the presence of two large bands that have migrated into the gel in the control reaction. As the samples were heat-shocked in the presence of 50 %v/v formamide - which stabilizes a single-stranded conformation of nucleic acids by deionizing the polymers - it is likely that the two bands result from a fully melted single-strands in the lower band and partially- or fully-duplexed DNA in the upper band. Alternatively, it has been previously shown that it is possible to separate complementary individual strands of DNA plasmid under denaturing conditions by agarose gel electrophoresis [163]. The two observed strands could therefore also both result from fully single-stranded DNA molecules that together form the duplexed plasmid.

The relative efficacy of MG fluorescence enhancement from each DNA template - and hence aptamer sequence - was examined via plate reader fluorescence measurements (Figure 6.9). Equimolar amounts of each DNA template were incubated in otherwise identical IVT reactions at 22 °C, and the fluorescence was measured in intervals of 10 min.

The three linearized DNA templates displayed noticeably different levels of MG fluorescence enhancement. As would be expected, the 1xMG transcripts confer the lowest fluorescence

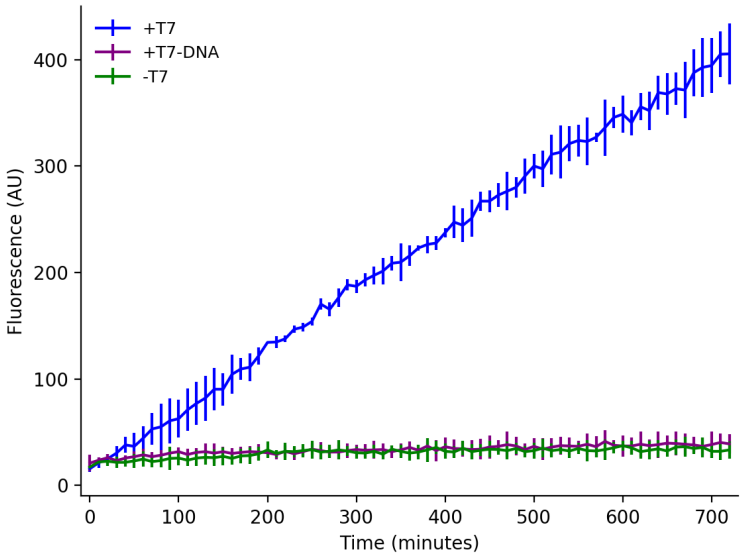


Figure 6.10: Activity of T7 RNAP and DNA-modified T7 RNAP, examined with MG fluorescence via plate reader. IVT reactions with equimolar amounts of unmodified T7 RNAP (blue trace) and DNA-modified T7 RNAP (purple trace) were prepared, as well as a negative control with T7 RNAP entirely omitted (green trace). Measurements were performed in triplicate, and error bars display the standard deviation. Originally presented in [155] and the data have been adapted for this work.

increase. Additionally, the 2xMG transcripts appear to be approximately twice as efficient as the 1xMG transcripts. Considering that T7 RNAP likely produces transcript molecules at similar rates in each reaction, resulting in similar molar amounts of each transcript, it may be inferred that the 2xMG transcripts are able to efficiently fold into two aptamers and stabilize two MG molecules per transcript.

In contrast with what might be predicted, the 3xMG transcripts appear to be only of intermediate efficiency. As gel electrophoresis analysis of the transcripts shows that the greatest mass of RNA transcripts is produced from the 3xMG reaction (Figure 6.8), the explanation that IVT of the 3xMG reaction is simply less efficient can likely be excluded. This then suggests that while the 3xMG transcripts are efficiently produced by T7 RNAP, they do not efficiently fold into the correct conformation and bind MG. As the 3xMG transcripts are inherently more repetitive than the 1xMG and 2xMG transcripts, it is likely that the 3xMG transcripts self-anneal into structures other than the intended aptamer repeats, and hence cannot increase MG fluorescence. This may be due to an inability to properly stabilize MG in the binding pocket, a decreased affinity for MG, or a combination of the two.

In order to implement T7 RNAP in SMC&P, a means of specific noncovalent surface immobilization was required. The ybbR tag of T7 RNAP was therefore covalently reacted with single-stranded DNA modified with a terminal CoA group via Sfp. The IVT activity of the modified polymerase was then assessed and compared to unmodified T7 RNAP (Figure 6.10). Equimolar amounts of modified and unmodified T7 RNAP were incubated in otherwise

identical IVT reactions with the 2xMG plasmid template at 22 °C. A negative control reaction that omitted T7 RNAP was also prepared. The fluorescence was measured in intervals of 10 min. As observed previously, IVT reaction with unmodified T7 RNAP increases the measured fluorescence significantly and steadily over time. In contrast, the reaction with DNA-modified T7 RNAP does not differ significantly from the negative control reaction with the polymerase entirely omitted.

Samples of the plate reader-measured IVT reactions were subsequently examined via agarose gel electrophoresis (Figure 6.11). Samples from two of the replicates of the reaction containing unmodified T7 RNAP were run as duplicates. A sample from one of the replicates of the reaction containing DNA-modified T7 RNAP and of the control reaction that omitted T7 RNAP were also examined. RiboRuler Low Range Ladder RNA standard was used as a mass standard to estimate the size of the transcripts from each reaction.

Consistent with the plate reader measurement of MG fluorescence, IVT reactions with unmodified T7 RNAP produce a single readily detectable transcript consistent with the expected size from the 2xMG plasmid template, while the reaction with DNA-modified T7 RNAP does not produce detectable transcripts. Interestingly, it is also observed that the presence of DNA-modified T7 RNAP does not hinder the migration of the plasmid template in the agarose gel, while unmodified T7 RNAP causes much of the plasmid template to be either trapped in the pocket or smeared near the top of the gel.

Although the unmodified T7 RNAP construct appears to be a functional and efficient enzyme, the covalent attachment of DNA appears to nullify enzyme activity. As hinted at by the migration of behavior of the plasmid template in agarose gel electrophoresis, this is likely due to an inability of DNA-modified T7 RNAP to bind the plasmid template. Rather, the single-stranded DNA modification out-competes the plasmid template for the polymerase's active site due to its covalent attachment to the polymerase and therefore effective local concentration that can be estimated to be at least in the mM range. In this situation, it may be inferred that the polymerase is able to only fleetingly bind the plasmid template before it is effectively kicked off by the covalently attached DNA anchor. Considering that the initiation step is the rate-limiting step in the T7 RNAP transcription reaction, it is likely that hardly any transcripts of any length are produced by DNA-modified T7 RNAP.

Historically, SMC&P had relied on single-stranded DNA to immobilize cargo molecules of interest on a surface in the target area. This strategy is not necessarily compatible with all molecules of interest, as demonstrated by significantly reduced activity from DNA-modified T7 RNAP.

Considering that T7 RNAP has a high affinity for DNA, the covalently-attached DNA may compete with the IVT template DNA for binding to enzyme. Additionally, as the DNA anchor likely has affinity for the enzyme's binding pocket, the DNA anchor is consequently less free to interact with its complementary DNA strands on the depot and target surfaces. This would likely affect the efficiency of SMC&P, resulting in fewer successful transfer events. In conclusion, T7 RNAP likely cannot be integrated into SMC&P given a DNA-based immobilization scheme.

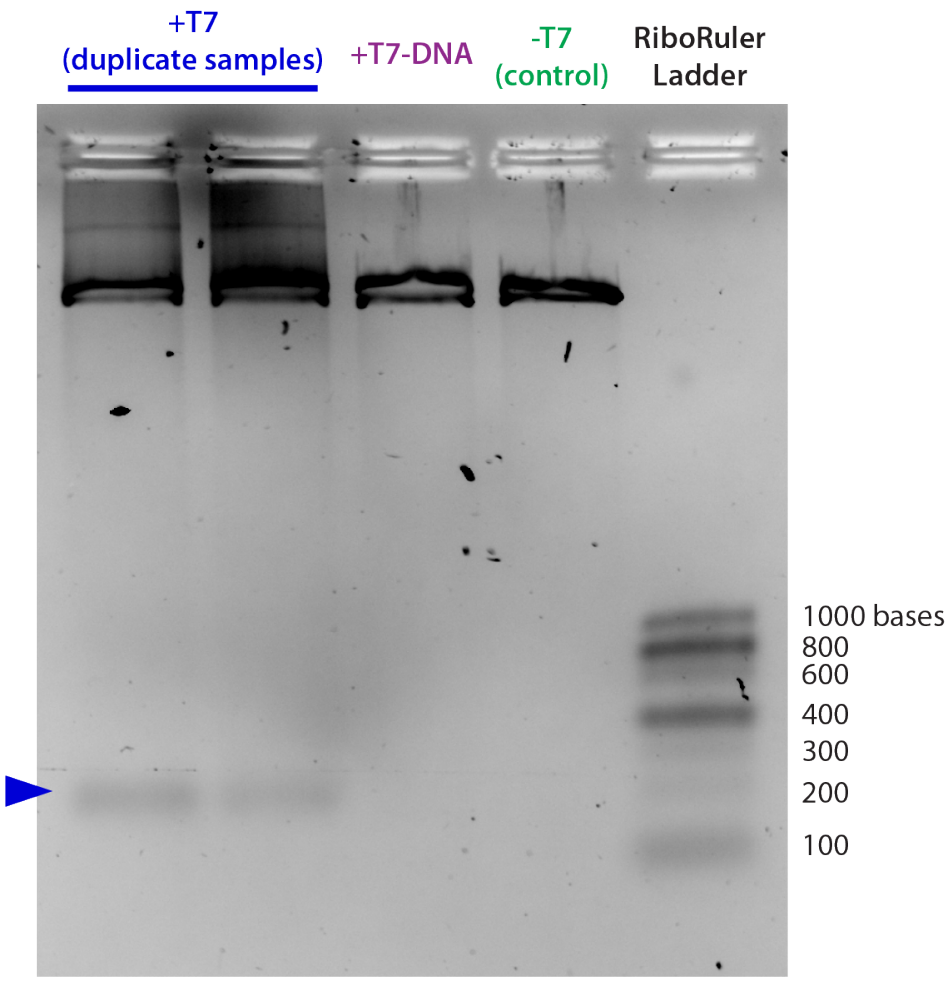


Figure 6.11: IVT reactions and T7 RNAP comparison, examined by agarose gel electrophoresis. Identical reactions with 2xMG plasmid template and equimolar amounts of unmodified T7 RNAP (+T7) and DNA-modified T7 RNAP (+T7-DNA) were prepared, in addition to a negative control reaction that omitted T7 RNAP (-T7). 2xMG IVT reaction RNA product is indicated with a blue arrow. Both the DNA-modified T7 RNAP reaction and the negative control reaction produces no observable nucleotide products. RiboRuler Low Range Ladder was additionally loaded as an RNA standard for size estimation.

Moreover, it is conceivable that a DNA-reliant immobilization strategy is similarly not viable for DNA-binding enzymes in general.

Alternatively, a DNA-free SMC&P strategy could potentially circumvent these difficulties. By limiting the unwanted affinity of the enzyme cargo and the immobilization molecules, the enzyme would be free to interact with the target ligands. Simultaneously, the surface immobilization handles would be better able to interact with the depot and target surfaces. Such a strategy of DNA-free SMC&P was successfully implemented, and is discussed at length in Chapter 8.

6.3 DNA Ligase Nick-Sealing of Labeled DNA

A system to detect DNA Ligase (DNAL) activity through fluorescence was contrived (Figure 6.12). As a general strategy, two pre-annealed DNA duplexes with fluorophores that together formed a FRET pair were ligated together by DNAL. Each duplex was designed to have complementary 5' sticky-overhangs and 5' phosphates to enable cohesive end-joining and subsequent nick-sealing ligation. After ligation, the covalently joined DNA formed a single duplex with two nearby fluorophores that together produced a FRET signal. DNA duplexes were necessarily designed in sets such that they possessed complementary sticky overhangs as well as FRET-compatible fluorophore pairs.

Ligation reactions were then analyzed by native PAGE rather than agarose gel electrophoresis, as the PAGE offers better resolution and separation of extremely short DNA complexes. The native conditions also preserved the structure of duplexes with long stretches of complementary bases but simultaneously enabled separation of unligated duplexes, which had only four basepairs of cohesive overlap. Duplexes - both ligated and unligated - were separated according to size, and the separated bands were then imaging with LED-based excitation of the fluorescent labels. Ligation is indicated by a band-shift from lower in the gel (smaller DNA that migrates faster) to higher in the gel (ligated DNA that is larger and migrates slower). Additionally, a FRET signal between nearby Cy3 and Cy5 labels was acquired with green LED excitation and then imaging with the red emission filter. Relative FRET efficiency was compared between duplex pairs with variable distances between the Cy3 and Cy5 labels.

Of specific interest was the effect of a fluorescent label in the footprint or near the active site of the ligase. The potential steric clash between the ligase and the fluorophore would presumably interfere with substrate recognition or transition state stabilization, resulting in compromised ligation efficiency. Considering that FRET is limited to a maximum distance between fluorophores of approximately 10 nm, the ligation reaction was tuned to identify a minimal distance between fluorophores that simultaneously does not drastically hinder ligase activity.

E. coli DNAL LigA was first challenged to ligate pre-annealed duplexes with cohesive sticky ends and a variable location of the Cy3 label (Figure 6.13). A reaction with a duplex pair lacking Cy3 served as a control reaction to benchmark LigA efficiency without any interfering Cy3 label. The control reaction and the reaction with Cy3 14 bp from the nearest nick site appear to

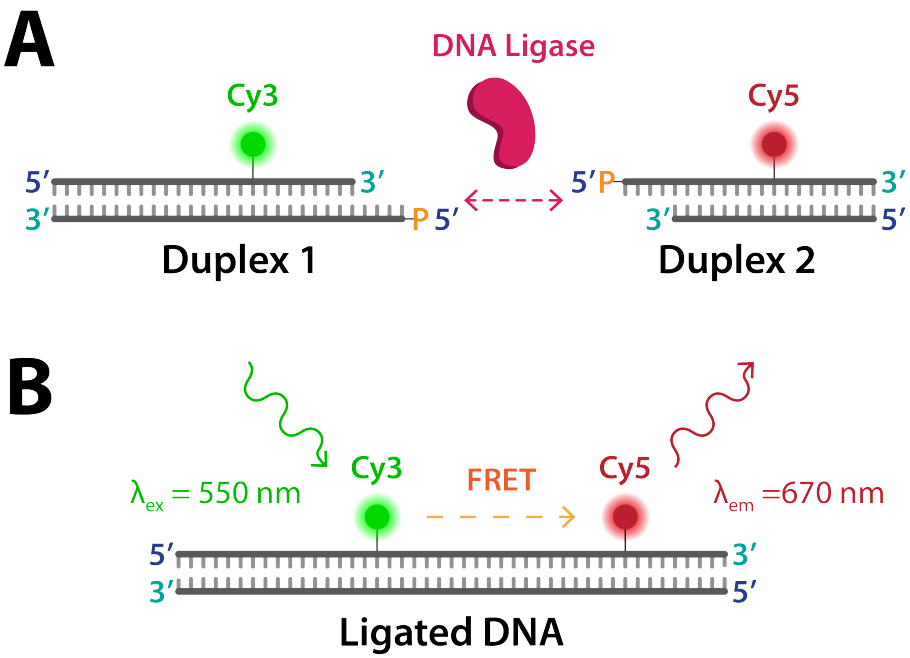


Figure 6.12: Ligation of fluorescently-labeled DNA by DNA Ligase. (a) Two pre-annealed duplexes are utilized, both with complementary sticky overhangs and 5' reactive PO_4. The duplexes are also both fluorescently labeled such that the two fluorophores together form a FRET pair (for example, Cy3 and Cy5 as shown here). DNA Ligase joins the sticky overhangs by sealing the nick in the backbone between the 5' phosphate and the 3' end of the other duplex. (b) Ligated duplexes form a single double-stranded duplex with two fluorophores that form a FRET pair.

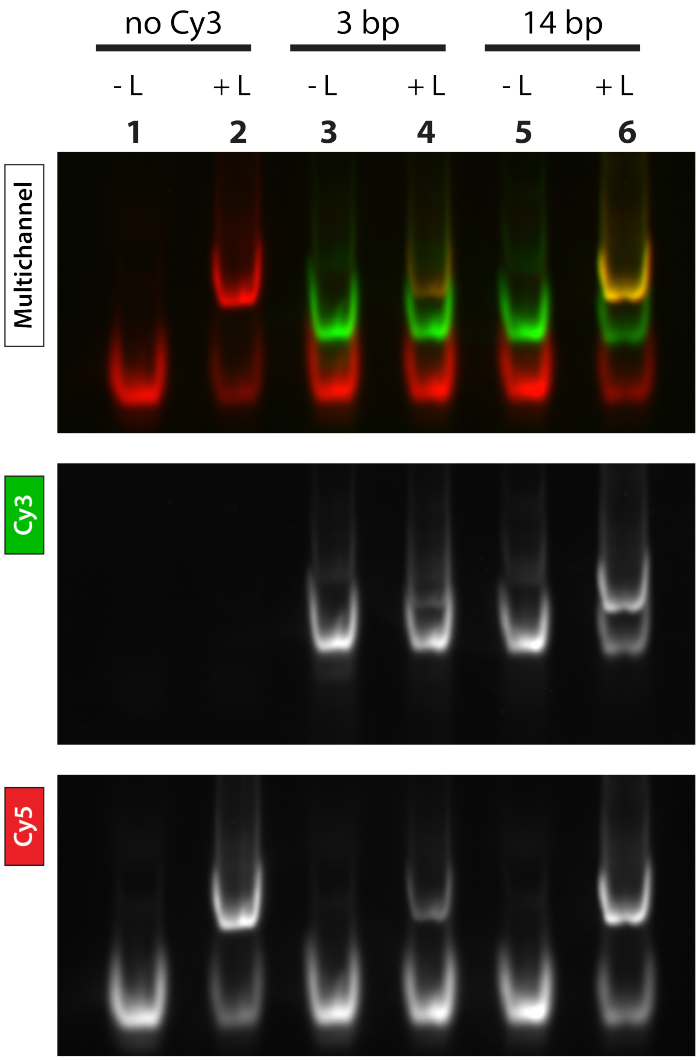


Figure 6.13: LigA ligation efficiency of fluorescent DNA, analyzed by native PAGE. Equimolar amounts of two pre-annealed duplexes with cohesive sticky ends were incubated with LigA (lanes 2, 4 and 6), as well as in the absence of DNAL as negative controls to check for unligated DNA migration (lanes 1, 3 and 5). A duplex with Cy5 13 bp from the nearest nick was used in all reactions, and was ligated to a complementary duplex in three variations: lacking Cy3 (lanes 1 and 2), with Cy3 3 bp from the nearest nick (lanes 3 and 4), and with Cy3 14 bp from the nearest nick (lanes 5 and 6). The gel was imaged with green LED excitation for Cy3 and red LED excitation for Cy5. The multichannel image was assembled by simple overlay of the two channels, with the green LED channel in green and the red LED channel in red.

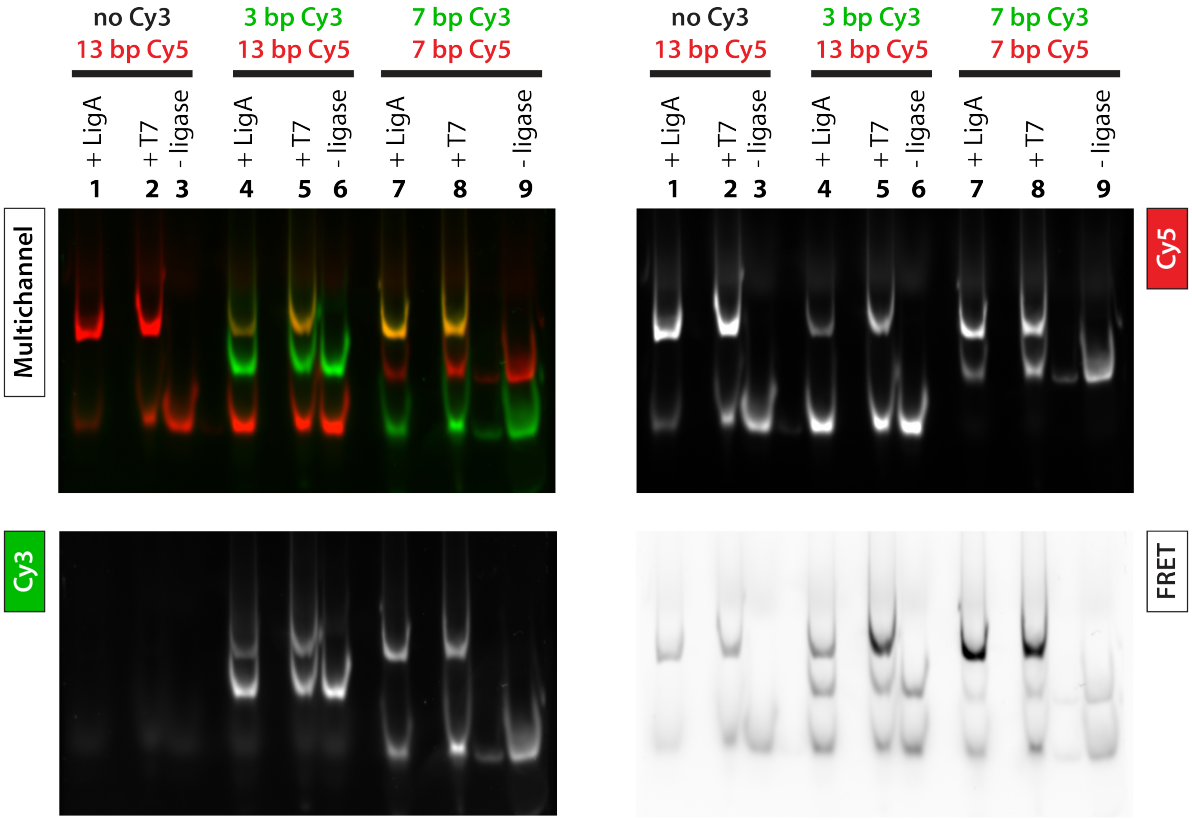


Figure 6.14: Tolerances of ligases for labeled DNA, analyzed by native PAGE. Equimolar amounts of two pre-annealed duplexes with cohesive sticky ends were incubated with LigA (lanes 1, 4 and 7), T7 DNAL (lanes 2, 5 and 8) as well as in the absence of DNAL as negative controls to check for unligated DNA migration (lanes 3, 6 and 9). Ligases were challenged to ligate duplex pairs with Cy3 and Cy5 the following distances from the nearest nick site: no Cy3 and Cy5 13 bp away (lanes 1 - 3); Cy3 3 bp away and Cy5 13 bp away (lanes 4 - 6); and both Cy3 and Cy5 7 bp away (lanes 7 - 9). The gel was imaged with green LED excitation and a 605/50 emission filter for Cy3, and red LED excitation and a 695/55 emission filter for Cy5. The multichannel image was assembled by simple overlay of the two channels, with the green LED channel in green and the red LED channel in red. FRET signal was acquired by combining green LED excitation and a 695/55 emission filter.

have similar ligation efficiencies. Considering that the footprint of LigA is estimated to be 19 bp and centered on the nick site [133], it is not surprising that the presence of fluorophores well outside of the ligase footprint appears to have no effect on ligation efficiency (compare lanes 2 and 6). In contrast, a Cy3 label only 3 bp from the nearest nick site (lane 4) has a noticeable adverse effect on the ligation efficiency, visible by the less bright higher band in both the green and red channels when compared to the same reaction with Cy3 much further from the nick site. Additionally, the lower band in the red channel that represents unligated Cy5-labeled DNA is comparatively bright in the reaction with Cy3 only 3 bp away from the nick. As this Cy3 label is well within the footprint of the ligase, it may be surmised that the lower ligation efficiency results from steric clash between the fluorophore and the ligase.

The relative tolerances of LigA and T7 DNAL for DNA modifications near the nick sites were directly compared (Figure 6.14). Ligation reactions were incubated with LigA and T7

DNAL, as well as in the absence of ligase as negative controls. Ligases were challenged to ligate duplex pairs containing Cy3 and Cy5 the following distances, respectively, from the nearest nick site: no Cy3 and Cy5 13 bp away; Cy3 3 bp away and Cy5 13 bp away; and both Cy3 and Cy5 7 bp away.

Compared to LigA, T7 DNAL has a much smaller footprint of approximately 12 - 14 bp centered asymmetrically on the nick site [131]. T7 DNAL is therefore predicted to have a better tolerance for fluorophores closer to the nick site, as the ligase binds fewer nucleotides bordering the nick. In agreement with the results presented in Figure 6.13, LigA appears to have significantly decreased catalytic activity when incubated with a duplex pair containing Cy3 a mere 3 bp away from the nearest nick site. T7 DNAL displays a less severe hindrance, as evidenced by the comparably brighter band in the multichannel image and stronger signal in the FRET channel. However, this reaction appears overall to be less efficient for both ligases compared to the reactions without Cy3 present.

This poses an interesting dilemma from the perspective of designing a FRET-based system of detection: how best to move the fluorophores closer together to increase FRET efficiency while simultaneously retaining high levels of ligase activity? To further investigate this question, another duplex was designed to have the Cy3 and Cy5 labels placed symmetrically 7 bp away from the nearest nick sites. It is also worth pointing out that Cy3 is located on the shorter strand, and Cy5 on the longer strand, although this should have no practical effects other than swapping the fluorescence pattern when separating DNA by size. Even though these two fluorophores are on average closer to the nick sites than the duplex with Cy3 3 bp away and Cy5 13 bp away, ligation efficiency appears to be much higher for this pair in the case of both ligases. Furthermore, the shorter distance between the fluorophores is expected to improve FRET efficiency independent of ligation efficiency. The combination of improved ligation as well as better FRET leads to a much stronger FRET signal for this duplex pair.

A final duplex was designed with Cy3 and Cy5 located 5 bp away from the nearest nick sites. The FRET efficiency of this duplex was compared to previously tested duplexes (Figure 6.15). As T7 DNAL appears to have higher ligation efficiencies than LigA when fluorophores are located near the nick sites, only T7 DNAL was used in this assay. T7 DNAL was challenged to ligate duplexes with Cy3 and Cy5 the following distances from the nearest nick site were ligated: Cy3 14 bp away and Cy5 13 bp away; both Cy3 and Cy5 7 bp away; and both Cy3 and Cy5 5 bp away. Taking into account the 4 bp sticky overlap from the cohesive duplex ends and a distance conversion factor of 0.34 nm/bp, the total nanometer distances between duplex pairs are approximately 10.5 nm, 6.1 nm and 4.8 nm, respectively.

Impressively, T7 DNAL appears to efficiently ligate the duplex pair with fluorophores each 5 bp away from the nick sites. As this fluorophore pair is only 4.8 nm apart, it is not surprising that this reaction displays the highest FRET efficiency of the three duplexes examined. Notably, for the fluorophore pair that is 4.8 nm apart, the signal strongly drops in the green channel, which in turn creates an orange color in the multichannel image. This is attributed to high FRET efficiency, leading Cy3 to fluoresce very little and instead transfer much of the absorbed energy in a non-radiative manner to nearby Cy5. In this case, ligation efficiency is best estimated by

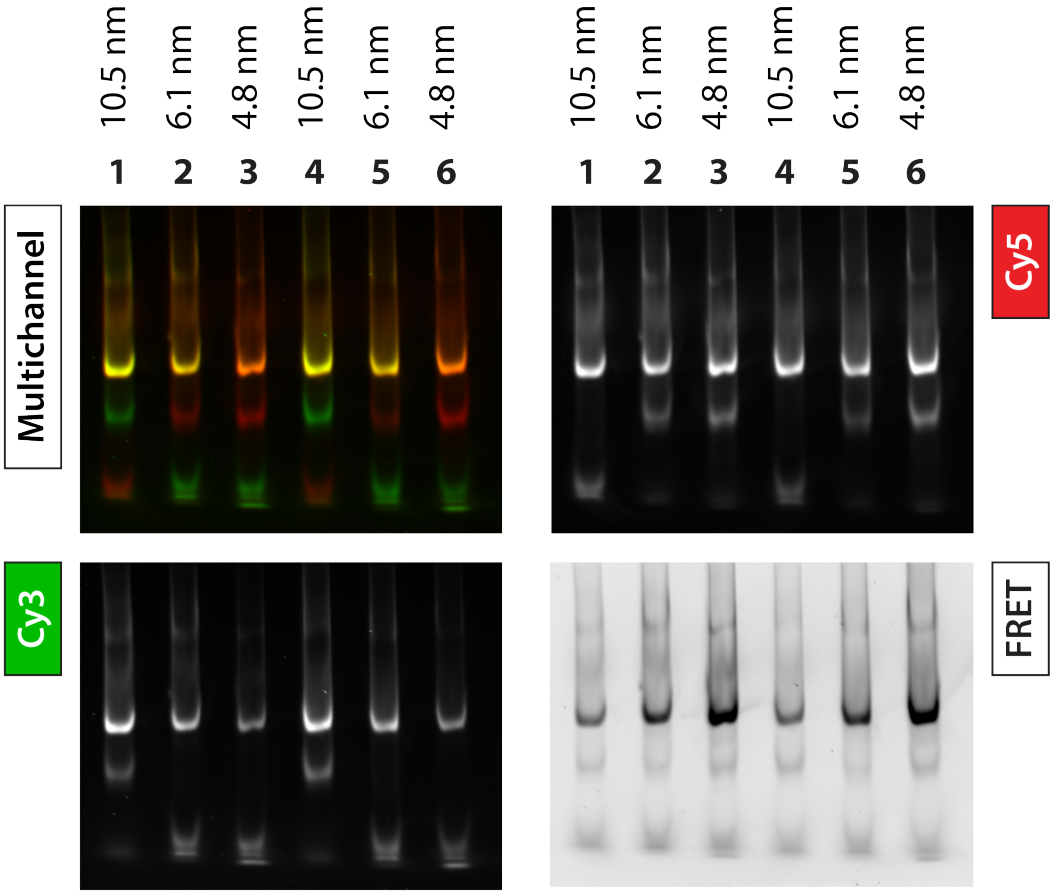


Figure 6.15: FRET efficiencies of ligated duplex pairs, analyzed by native PAGE. Equimolar amounts of two pre-annealed duplexes with cohesive sticky ends were incubated with T7 DNAL, and the relative FRET efficiency of each ligated duplex pair was compared. Each reaction was run in duplicate, and replicates are in good agreement with each other. Duplex pairs with Cy3 and Cy5 the following distances from the nearest nick site were ligated: Cy3 14 bp away and Cy5 13 bp away (lanes 1 and 4); both Cy3 and Cy5 7 bp away (lanes 2 and 5); and both Cy3 and Cy5 5 bp away (lanes 3 and 6). Taking into account the 4 bp sticky overlap from the cohesive duplex ends, the total nanometer distances between duplex pairs are approximately 10.5 nm, 6.1 nm and 4.8 nm, respectively. The gel was imaged with green LED excitation and a 605/50 emission filter for Cy3, and red LED excitation and a 695/55 emission filter for Cy5. The multichannel image was assembled by simple overlay of the two channels, with the green LED channel in green and the red LED channel in red. FRET signal was acquired by combining green LED excitation and a 695/55 emission filter.

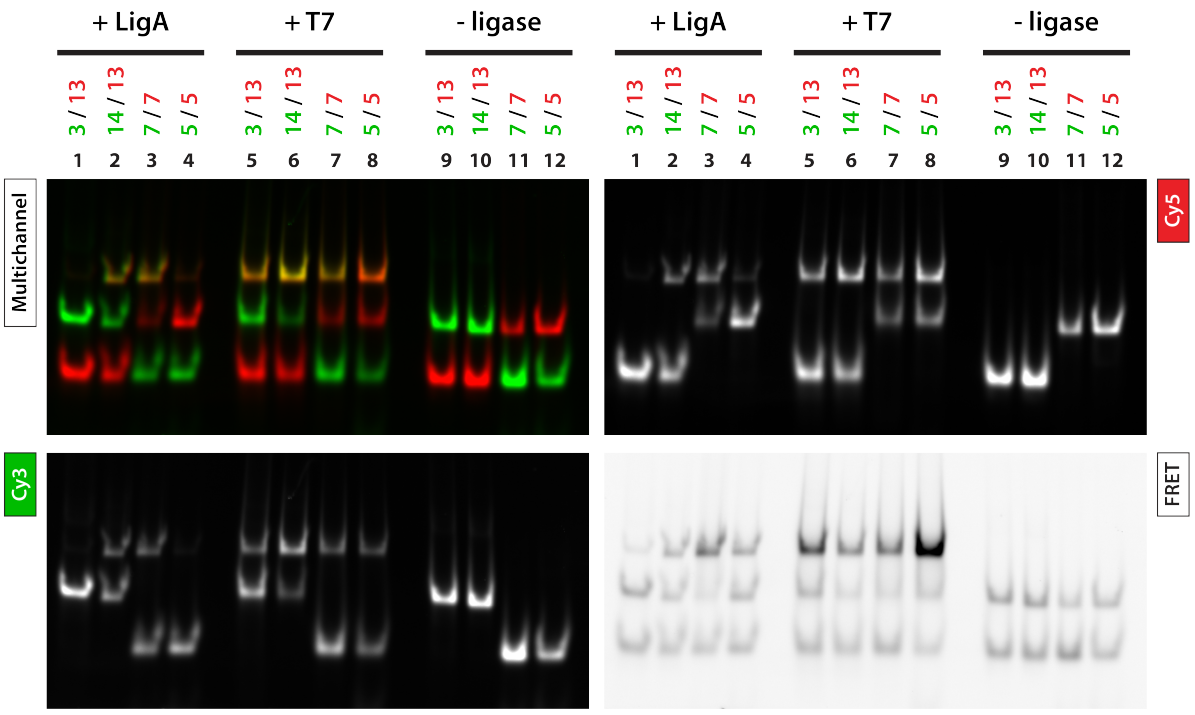


Figure 6.16: Ligation and FRET efficiencies of labeled DNA, analyzed by native PAGE. Equimolar amounts of two pre-annealed duplexes with cohesive sticky ends were incubated with LigA (lanes 1 - 4), T7 DNAL (lanes 5 - 8) as well as in the absence of DNAL as negative controls to check for unligated DNA migration (lanes 9 - 12). Ligases were challenged to ligate duplex pairs with Cy3 and Cy5 the following distances from the nearest nick site: Cy3 14 bp away and Cy5 13 bp away (lanes 1, 5 and 9); Cy3 3 bp away and Cy5 13 bp away (lanes 2, 6 and 10); both Cy3 and Cy5 7 bp away (lanes 3, 7 and 11); and both Cy3 and Cy5 5 bp away (lanes 4, 8 and 12). The gel was imaged with green LED excitation and a 605/50 emission filter for Cy3, and red LED excitation and a 695/55 emission filter for Cy5. The multichannel image was assembled by simple overlay of the two channels, with the green LED channel in green and the red LED channel in red. FRET signal was acquired by combining green LED excitation and a 695/55 emission filter.

comparing the amount of unligated DNA that still remains in the reactions, as the high FRET efficiency alters the appearance of the ligated product band.

The ligation efficiency and FRET efficiency of Cy3- and Cy5- labeled duplex pairs incubated with LigA or T7 DNAL are summarized in Figure 6.16. T7 DNAL appears to have high catalytic activity for all tested duplex, including duplexes with fluorescent labels within 3 - 5 basepairs of a ligation site. In contrast, LigA activity is significantly hindered by fluorophores near the active site.

This discrepancy in fluorophore tolerance between LigA and T7 DNAL cannot be explained by footprint size alone. T7 DNAL is expected to interact directly with nucleotides within 3 - 5 bp on the 3'OH side and 7 - 9 bp on the 5'PO_4 side of a nick site [131]. This footprint encompasses several tested duplexes (such as the duplex with Cy3 only 3 bp away from the nick), although T7 DNAL does not display the same degree of impeded catalytic activity. Rather, it is likely the three-dimensional interface of ligases and DNA that explains this difference. The smaller

size of T7 DNAL compared to LigA not only gives it a shorter footprint, but also inevitably decreases the surface contact area between the ligase and the DNA substrate; LigA wraps around the DNA helix in almost a 360° arc [133], while T7 DNAL is restricted to only one face of the helix. The difference in binding geometry likely plays a role in sensitivity to fluorophores near the nick sites. Fluorescent labels on the bases of nucleotides typically project out into the major groove of DNA in order to minimize steric clash between the bulky fluorophore and the surrounding DNA. If T7 DNAL binds the nicked DNA on the opposite side of the helix and hence does not make physical contact with the fluorophore, then ligation efficiency will likely not be affected - even if the fluorophore is technically within the footprint area of the ligase.

Furthermore, in the experiments presented here, the employed duplexes both have a 5' reactive PO_4 at the designed nick sites to facilitate complete ligation on both strands. Therefore, it is conceivable that many of the ligated duplexes observed in gels are in fact still nicked at one of the sites.

A potentially valuable continuation of this series of bulk reactions would be to incubate pairs of duplexes in which only one of the duplexes has a reactive 5'PO_4 for DNA ligation. Considering that when both duplexes have a 5' reactive PO_4, it is possible that only one of the two nicks is sealed. In cases where fluorophores are positioned asymmetrically around the nick sites - *e.g.* the duplex pair with Cy3 spaced 3 bp away and Cy5 spaced 13 bp away from the nearest nick sites - one nick may be efficiently sealed but the other not. The partially-ligated duplexes would still migrate as a larger DNA complex in the gel, thereby incorrectly reporting the deleterious effects of one fluorophore on the ligation efficiency. By instead testing the nick sites individually, the effects of specific modifications while simultaneously investigating FRET efficiency between two fluorophores may be better understood. Similarly, the strand location of a fluorophore relative to the nick site (*i.e.* same strand or opposite strand) likely also affects the ligation efficiency, therefore varying the fluorophore position in that respect could also further elucidate DNAL behavior.

The most successful ligase scheme so far tested - T7 DNAL and a duplex pair with fluorophores spaced 5 bp symmetrically around the nick sites - is a promising candidate as an enzyme system to be implemented in SMC&P. Further optimization of the system could reveal additional fluorophore geometries that enhance FRET efficiency or better preserve ligase activity. Additionally, the fluorophores themselves may be swapped out for versions with better fluorescent properties and longer lifetimes, *e.g.* Atto550 and Atto647N. Nevertheless, the combination of well-preserved ligase activity of T7 DNAL and remarkable FRET efficiency of nearby Cy3 and Cy5 overall makes this reaction already a potentially highly efficient FRET-based readout system on the single-molecule level.

6.4 NanoLuc Luciferase Bioluminescence

The activity of the engineered luciferase NanoLuc was first compared in bulk to firefly luciferase (LuLic) to confirm its greatly improved luminescence. Reactions containing the luciferin coelenterazine and NanoLuc or LuLic were incubated at room temperature, and the absorbance

was measured over time via plate reader. In agreement with what is predicted for the engineered luciferase, NanoLuc was significantly more luminescent than the wildtype luciferase when compared at equal concentrations. This is due primarily to increased catalytic turnover on the part of NanoLuc.

A microfluidics system designed for SMC&P was utilized to immobilize NanoLuc on a glass surface with a well-defined border between the enzyme-containing channels and extra-channel area of the surface. NanoLuc molecules were covalently immobilized via ybbR tag to a CoA-functionalized surface, and a buffer solution containing high concentrations of the luciferin furimazine was added to the surface. For the purposes of imaging, the surface was first adjusted to the focal plane of the objective by TIRF illumination with the blue laser, which tends to excite nonspecifically adsorbed contaminating molecules. This provided a means of focusing to the surface in the absence of specifically immobilized fluorescent molecules of interest. The laser was then shut off completely, and the blue filter set was removed from the Optosplit to allow all photons in the blue range to be detected in the emission path.

Although in bulk experiments it was indeed observed that NanoLuc produces significantly more bioluminescence signal compared to wildtype luciferases, this was still insufficient to observe luciferase activity on the level of single-molecules with the optics setup available.

6.5 Improved Expression of T7 RNA Polymerase

Although recombinant expression of a protein of interest in *E. coli* has several advantages - namely rapid protocols and relatively inexpensive materials - it is a strategy that is often ill-suited for many proteins. In particular, large proteins are frequently express in a denatured state that forms an insoluble precipitate during purification. In both eukaryotes and prokaryotes, complex proteins often require chaperone proteins to assist with proper folding [7, 20, 21], and such chaperones from another organism are likely not natively present in *E. coli*. The absence of such necessary chaperones greatly increases the likelihood that a protein does not fold correctly and instead folds directly into a denatured state. Although there are methods to extract the protein of interest from the insoluble fraction and refold them into a soluble state, initially soluble proteins are nevertheless preferred.

Elaborate specialized protocols are often developed to optimize expression and purification of particular proteins. Additionally, efficient solubility tags - *i.e.* highly soluble and relatively small peptides or proteins that may be fused to a protein of interest to improve its solubility - are highly sought after. Solubility tags may also have secondary functions in addition to solubilizing the expressed fusion construct. For example, Liu *et al.* fused an adenylate kinase domain to the ATP-dependent ligase T4 DNAL to both greatly increase the expression efficiency of the ligase as well as confer an ability of the fusion construct to convert ADP to ATP [136].

Wild-type as well as 6xHis-tagged T7 RNAP generally exhibits very poor solubility in *E. coli* expression systems, and consequently yields relatively low concentrations of purified functional protein. Like many individual enzymes, detailed protocols have been devised to improve expression and purification of T7 RNAP [164, 165].

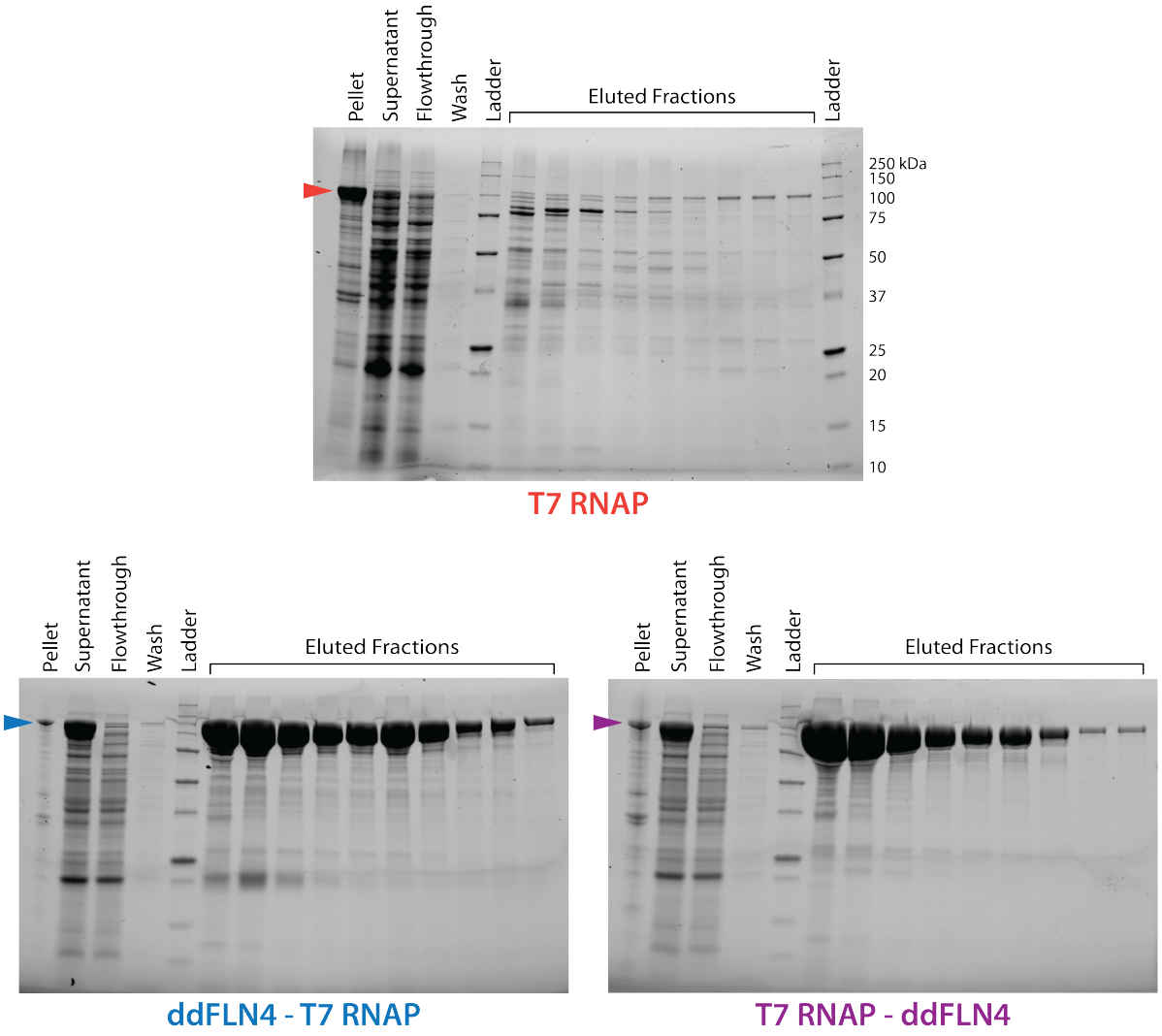


Figure 6.17: Improved expression and purification yield of T7 RNAP by fusion with ddFLN4, analyzed by SDS PAGE. T7 RNAP (100 kDa) and two fusions of T7 RNAP and ddFLN4 (118 kDa) were expressed and purified according to identical protocols. Following cell lysis and high-speed centrifugation to separate the soluble supernatant and insoluble pellet, T7 RNAP (orange arrow) is found mostly in the pellet with very little apparent in the supernatant. T7 RNAP was then purified by Ni-IMAC, and eluted fractions consequently contain low concentrations of protein. In contrast, two fusions of T7 RNAP with an N-terminal ddFLN4 domain (ddFLN4-T7 RNAP) and a C-terminal ddFLN4 domain (T7 RNAP-ddFLN4) were found mostly in the soluble supernatant fraction with much less found in the pellet, and subsequent eluted fractions contain high concentrations of protein. The same mass standard ladder (Precision Plus Unstained Ladder, Bio-Rad) was used in each gel for reference, and comparable volumes of each sample type were loaded into each gel.

The 11 kDa domain termed ddFLN4 has previously demonstrated solubility tag-like properties in the expression of monomeric Streptavidin [70], in addition to serving as a robust fingerprint domain in AFM-based SMFS [36, 49, 70]. When fused either N-terminally or C-terminally to T7 RNAP, ddFLN4 similarly increased both T7 RNAP solubility as well as fractionation yield in Ni-IMAC (Figure 6.17). Ni-IMAC purification of 6xHis-tagged and ddFLN4-fused T7 RNAP yielded peak fractions of approximately 50 µM - 150 µM, which is a 100-fold increase compared to 6xHis-tagged T7 RNAP peak fractions of approximately 500 nM - 1.5 µM.

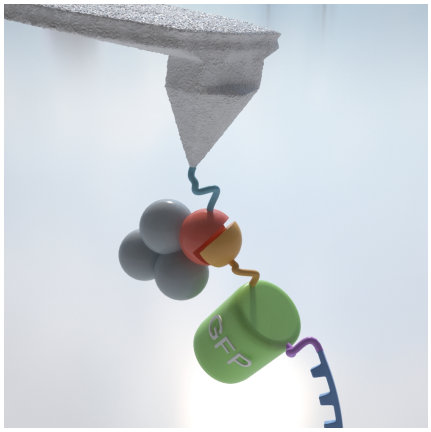
Normally an enzyme with mediocre yields from recombinant expression in bacteria, ddFLN4 appears to greatly improve solubility of this enzyme. The addition of the small ddFLN4 motif may therefore be an attractive strategy to improve expression efficiency of enzymes in general, provided there is care taken to the positioning of the ddFLN4 motif within the fully chimeric construct.

Chapter 7

Strep-Tag II and Monovalent *Strep*-Tactin as Novel Handles in Single-Molecule Cut-and-Paste

Summary

Directed spatial assembly of single molecules on a surface presents an opportunity to precisely control the positioning, density and geometry of molecules of interest within an ensemble. In contrast to bulk averaging, this enables detection and analysis of individual behavior within such a designed ensemble. The AFM-based technique of Single-Molecule Cut-and-Paste has been developed to arrange a variety of biomolecules on a surface through different handling strategies. This technique requires cantilever- and surface-handles that simultaneously adhere to a prerequisite rupture force hierarchy, and also do not cross-interact with each other or the transported molecules. As the molecules of interest diversify, so too must the handling methods to accommodate their unique characteristics. We demonstrate that a previously-developed monovalent variant of *Strep*-Tactin and its corresponding *Strep*-Tag II peptide ligand comprise a viable cantilever handling complex for SMC&P. Ultimately, this expansion to the SMC&P toolbox increases the system's versatility for new molecules of interest yet to be studied.



This chapter was published by Erlich *et al.* [102] in *Small Methods* and adapted with permission from Wiley-VCH.

7.1 Introduction

The frontier of nanoscale studies frequently presents unexpected challenges that must be overcome with innovation. As such, universally applicable approaches often do not exist, and instead diverse methods or tools must be developed. Bottom-up synthetic biology employs fundamental biological components as the building blocks for artificial biological systems with novel characteristics. A major endeavor of this broad field is to develop unique molecular organization techniques, such as engineered protein modules [91] and enzyme cascades assembled on DNA origami scaffolds [97]. Single-Molecule Cut-and-Paste (SMC&P) is one such organization technique, merging bottom-up assembly with control on the level of single molecules. SMC&P utilizes the single-molecule force spectroscopy (SMFS) and lateral surface positioning of atomic force microscopy (AFM) [39, 166, 167] to deposit molecules of interest in arbitrary patterns on a functionalized glass surface with nanometer-precision. There exist several key challenges in SMFS that are constantly improved upon: signal-to-noise limits of data resolution, and specific handling. Advances in existing tools, such as modified cantilevers in AFM-based SMFS [168], can greatly improve data quality and expose previously inaccessible levels of detail. Precise and versatile control of molecules of interest can likewise open up new avenues of study. For example, protein-labelled DNA tethering strategies in optical tweezers experiments offer additional flexibility in handling biomolecules for mechanical measurements [169]. Similarly, SMC&P requires robust immobilization and handling schemes for the specific and controlled arrangement of diverse biological agents. With a more expansive repertoire of handling strategies, its components could potentially be tailored to enable arrangement of any molecule of interest in a well-defined orientation.

During each SMC&P cycle, a non-covalently immobilized transfer molecule of interest is picked up from a depot area via a cantilever-coupled affinity handle. The cantilever relocates the transfer molecule to a target area, where it then deposits said molecule and is recycled back to the depot area to repeat the process. Arrays of molecules are assembled with precise localization in the target area, where their properties such as their fluorescent behavior in an ensemble or as individuals can be analyzed. For example, hybrid DNA-RNA molecules were specifically arranged and immobilized on a surface via complementary oligonucleotides. The resulting duplexes formed aptamers that stabilized the structure of a target dye molecule, enabling it to produce a fluorescent signal upon binding the SMC&P-arranged constructs [1]. Proteins have also been integrated into SMC&P, including constructs containing modified green fluorescent protein (GFP) [170] with different surface-immobilization strategies and a protein-based handle for the cantilever [86, 101]. Directed placement of molecules within the nanoapertures of zero mode waveguides with SMC&P is also possible. These nanoapertures facilitate measurements in a confined volume and improve background fluorescence when compared to conventional fluorescence microscopy. Importantly, the precise placement of molecules via SMC&P decreases the heterogeneity of fluorescence intensity and lifetime that results from stochastic immobilization and quenching effects from the metallic sidewalls [2]. In combination with fluorescence microscopy, SMC&P therefore presents new opportunities to

examine biomolecular behavior on the single-molecule level with precise control of surface location and environment.

SMC&P fundamentally relies on a hierarchy of rupture forces of the specific interactions between the transfer molecule and the depot area (F_D), the cantilever (F_C) and the target area (F_T) such that $F_D < F_C < F_T$. Consequently, this system demands a palette of selective immobilization and pickup methods so that the forces involved are tunable to fit this hierarchy. Furthermore, an ideal SMC&P scheme uses orthogonal agents that avoid unwanted cross-reactivity. Even small peptide tags and single-stranded DNA anchors could have adverse interactions between themselves or with molecules of interest, which in turn could interfere with SMC&P efficiency or molecule behavior. Therefore, a modular approach to construct design confers a degree of plasticity and promotes integration of any molecule of interest into SMC&P. By expanding the selection of cantilever- or surface-handles, the system can be further adapted to specific force and mechanistic requirements.

Strep-Tactin, an engineered variant of streptavidin, is a tetravalent complex that specifically binds with high affinity to the short peptide *Strep*-tag II (SII) [107, 109]. A monovalent version of *Strep*-Tactin (monoST) was recently developed and implemented in SMFS as a cantilever-immobilized handle for protein constructs harboring a SII-peptide [69]. The rupture forces of SII:monoST were found to be dependent on loading rate as well as location of SII in the protein construct, with N-terminal SII resulting in much lower rupture forces than C-terminal SII under identical loading rates. Its tethering geometry-dependent force regime and tunable rupture forces make the SII:monoST complex a compelling candidate as a handling system for SMC&P. Moreover, its addition to the growing SMC&P toolbox advances the technique towards the ultimate goal of enabling precise arrangement of any molecule of interest.

7.2 Materials and Experimental Methods

Preparation of Monovalent Strep-Tactin: A heterotetrameric monovalent version of *Strep*-Tactin (monoST) was designed, expressed, purified and reconstituted as previously described by Baumann et al. [69] In brief, the non-functional subunits and the single functional subunit harboring a 6xHis tag and a reactive Cysteine residue were separately expressed in *E. coli* BL21(DE3)-CodonPlus cells. Inclusion bodies were dissolved and denatured, and the dissolved inclusion body fractions of the non-functional and functional subunits were mixed in a ratio of 10:1 respectively. Subunits were refolded by slowly and drop-wise adding to a reservoir of 1x phosphate-buffered saline (PBS). The assembled monoST was purified by Ni-IMAC affinity chromatography. The fractions containing monoST were isolated and dialyzed against 1x PBS. Purified monoST was long-term stored at 4 °C in the presence of TCEP beads.

Preparation of sfGFP Construct and DNA Coupling: A superfolder Green Fluorescent Protein (sfGFP) [170] transfer construct was designed, expressed and purified as previously described in Baumann et al. [69] In brief, the construct harbors an N-terminal *Strep*-tag II (SAWSHPQFEK = SII) [107, 109] and a C-terminal ybbR-tag (DSLEFIASKLA) [148, 149] to enable specific cantilever handling and DNA coupling, respectively. The GFP gene was cloned into a modified pET28a

vector that contains an N-terminal 6xHis-tag followed by a PreScission Protease cleavage site (PreSc). The resulting fusion protein (6xHis-PreSc-SII-sfGFP-ybbR) was expressed in E.coli BL21(DE3)-CodonPlus cells. The sfGFP construct was obtained in the soluble fraction after cell lysis and purified by Ni-IMAC affinity chromatography. Selected fractions of purified protein were then dialyzed overnight against storage buffer (50 mM Tris HCl pH 7.5, 150 mM NaCl, 2 mM DTT, 5 %(v/v) Glycerol) and stored long-term at -80 °C. The sfGFP construct was covalently coupled to DNA via the enzyme Sfp transferase as similarly described by Pippig et al. [86], which is slightly altered from the protocol of Yin et al. [149] PreScission Protease, Sfp transferase, and Coenzyme A-modified transfer DNA (biomers.net GmbH, Ulm, Germany) were incubated with the purified 6xHis-PreSc-SII-sfGFP-ybbR construct at room temperature for 2 h for simultaneous cleavage of the 6xHis tag and covalent coupling of the ybbR tag to DNA. The reaction was filtered and then stored on ice until application in a microfluidic system.

Preparation of Cantilevers: MLCT cantilevers (Bruker, Camarillo, USA) were silanized in 3-(Aminopropyl)dimethylethoxysilane and subsequently functionalized with a heterobifunctional PEG crosslinker [158, 159] with N-hydroxy succinimide and maleimide groups (MW 5000 Da). Cantilevers were covalently coupled to monoST.

Preparation of Glass Surfaces: Glass cover slips were silanized in (3-Aminopropyl)dimethylethoxysilane and subsequently functionalized with a heterobifunctional PEG crosslinker with N-hydroxy succinimide and maleimide groups (MW 5000 Da). Thiol-modified depot and target DNA was reduced and then purified by ethanol precipitation. A PDMS microfluidic system – based on the system described by Kufer et al. [98] – was fixed on the PEGylated cover glass. Depot and target channels were functionalized with their respective reduced DNA, and the sfGFP-DNA chimera construct was incubated in the depot channel for 1 h. The depot channel was then flushed with 1x PBS to remove unbound- or nonspecifically-bound sfGFP. The microfluidic system was then removed and the surface submerged in 1x PBS.

AFM/TIRFM Measurements: SMC&P experiments were carried out on a combined AFM/TIRFM setup, as described previously [156]. The patterns were written in 395 - 487 transfer cycles with 150 - 250 nm spacing between each deposition point. The pulling speed in the depot was set to 2000 nm/s and in the target to 200 nm/s. Rupture forces and loading rates were evaluated from AFM force distance curves that were recorded for each pickup and deposition process utilizing the WLC model [162] with quantum mechanical correction. Blue laser excitation at wavelength 488 nm with an estimated intensity of approximately 10 W/cm^2 was utilized to monitor the GFP fluorescence. Fluorescent images were evaluated and processed with the analysis software ImageJ.

7.3 Results and Discussion

A construct consisting of GFP with an N-terminal SII and a C-terminal ybbR tag was previously expressed and purified, and it was also demonstrated in SMFS that the rupture forces of an N-terminal SII and monoST were lower than the force required to unfold GFP [69]. This construct was here employed in SMC&P, where the relatively low rupture forces of N-terminal

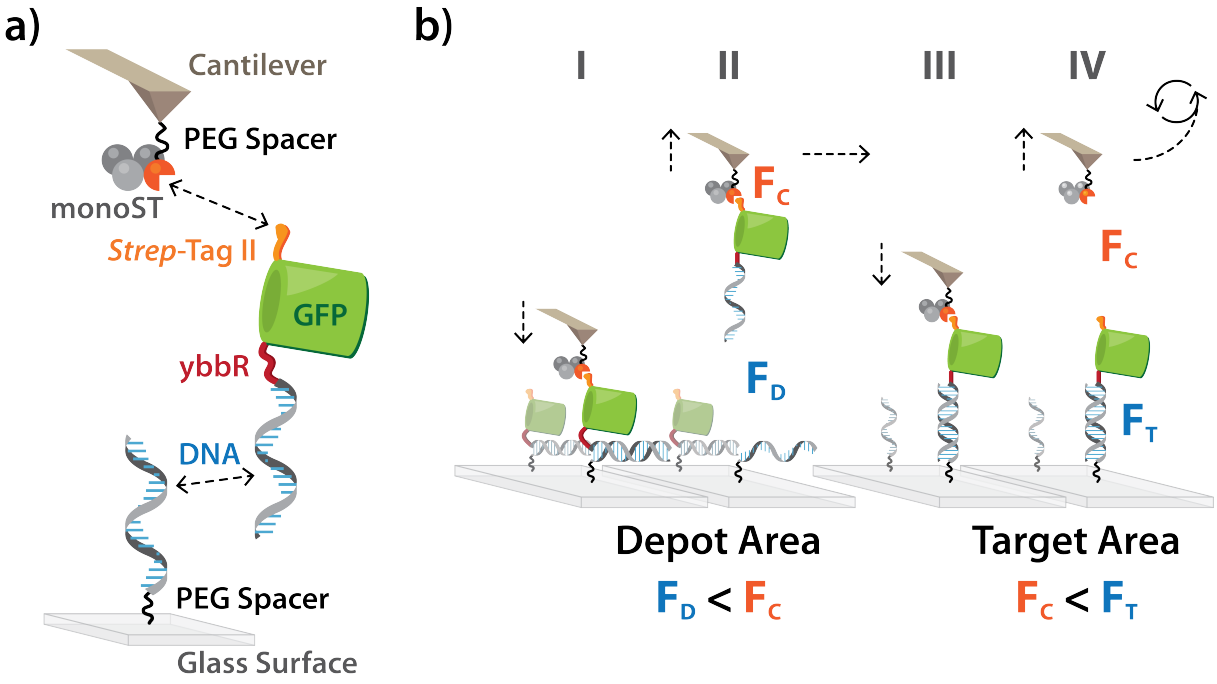


Figure 7.1: Monovalent *Strep*-Tactin and the GFP transfer construct employed in SMC&P. a) The chimeric transfer construct consists of a GFP molecule with an N-terminal SII as well as a C-terminal ybbR tag, which is then covalently coupled to 3'-modified CoA single-stranded DNA via reaction catalyzed by Sfp Synthase. The DNA anchor binds non-covalently to the surface via a complementary DNA strand. monoST is covalently coupled to the cantilever, and its single functional subunit targets SII of the transfer construct. b) Repeatable transfer cycling of SMC&P depends on a force hierarchy determined by DNA hybridization geometry and the SII:monoST interaction. The cantilever approaches the depot surface, and monoST binds to SII of a transfer molecule immobilized via complementary DNA in zipper orientation (I). The cantilever retracts and removes the transfer molecule as the DNA unzips (II). The loaded cantilever then transports the transfer molecule to the target surface where the transfer molecule binds to complementary DNA in shear orientation (III). Retraction of the cantilever ruptures the SII:monoST complex, and the cantilever is recycled back to the depot area to repeat the process (IV). The rupture forces of the transfer construct with the DNA in the depot area (F_D), monoST on the cantilever (F_C) and DNA in the target area (F_T) are tuned such that $F_D < F_C < F_T$.

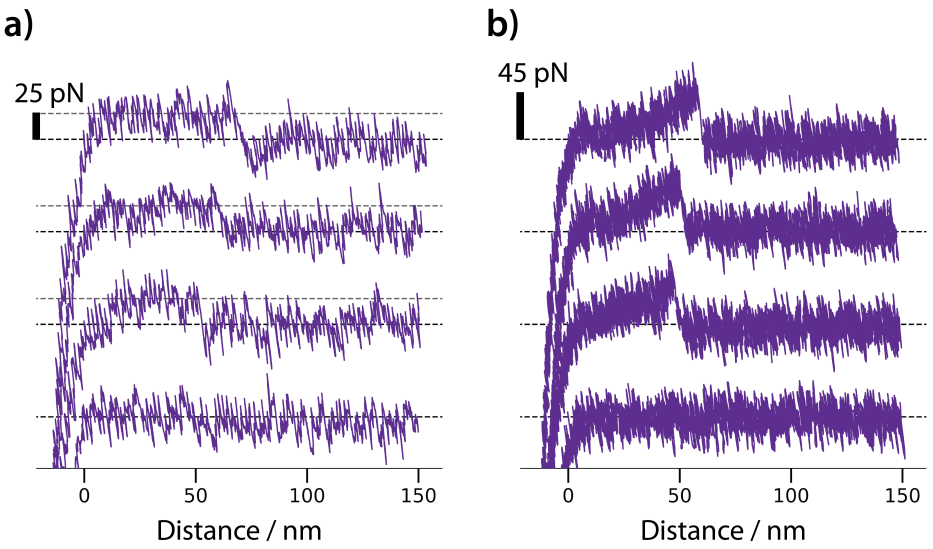


Figure 7.2: Typical force-distance curves of single-molecule a) depot pickup and b) target deposition events. Depot pickup events occur when the transfer construct-surface DNA complex is unzipped, resulting in a plateau of relatively constant force at ≈25 pN (depicted by the dashed grey lines). Target deposition events occur when the newly established transfer construct-surface DNA complex in shear orientation remains intact, and instead the SII:monoST complex unbinds. This results in a gradual buildup of force and a sharp unbinding peak at ≈SI45pN. In some cases, no molecules were picked up or deposited, which is reflected in the zero-force curves (bottom-most traces). The baseline for each curve of SI0pN force is depicted by the dashed black lines.

SII and monoST were advantageous for preserving the fold and thus fluorescence of GFP during the transport process. The C-terminal ybbR tag was covalently modified with 3'-CoA single-stranded DNA by the Phosphopantetheinyl Transferase enzyme Sfp Synthase [148]. Hybridization of the DNA anchor to a complementary DNA strand in either zipper- or shear-orientation enabled noncovalent surface immobilization to the depot and target areas, respectively (Figure 7.1a).

SMC&P was executed in a custom-built hybrid AFM/TIRFM instrument [156]. Chimeric transfer constructs were picked up from the depot area via a monoST-coupled cantilever and deposited into the target area over 395 - 487 consecutive SMC&P cycles. The technique fundamentally relies on a hierarchy of rupture forces between the transfer molecule and the depot storage molecule (F_D), the cantilever handle (F_C) and the target storage molecule (F_T) such that $F_D < F_C < F_T$. The rupture forces of DNA duplexes are tuned via duplex length and pulling geometry (zipper vs. shear) [98]. Additionally, the rupture forces of protein-protein interactions are influenced by loading rates [47, 61, 69]. In this way, it is possible to optimize a scheme to consistently transport transfer constructs and regenerate the cantilever (Figure 7.1b).

The distinct behaviors of the depot DNA duplex and the SII:monoST complex upon unbinding and rupture are illustrated in their respective force curve patterns (Figure 7.2). When pulled apart in zipper orientation, the basepair-by-basepair unzipping of DNA is described quantitatively by an equilibrium thermodynamic model [52, 171], leading to a plateau of constant force. During a pickup event from the depot area, the 40 bp DNA duplex melts at ≈25 pN, which is consistent with previous SMC&P experiments using this same DNA duplex [86, 101]. The

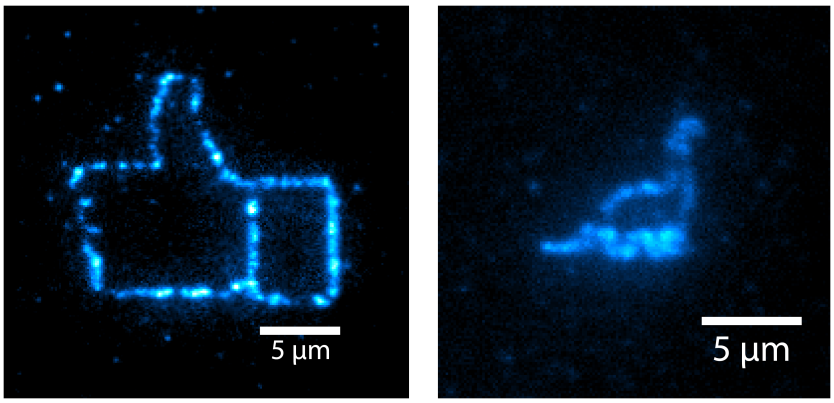


Figure 7.3: TIRFM image of GFP molecules after SMC&P in thumbs up and dinosaur patterns. The images are composed of the average pixel intensity of 30 stacked frames from TIRFM acquisition (0.12 s exposure time at ≈10 W/cm^2) with a blue laser. The patterns consist of 487 deposited molecules spaced 250 nm apart and 395 deposited molecules spaced 150 nm apart, respectively.

non-equilibrium unbinding of the SII:monoST interaction occurs at forces significantly greater than this value at the given loading rates [69]. Consequently, the transfer-depot DNA duplex melts while the SII:monoST bond stays intact, allowing for consistent pickup of molecules from the depot. The transfer construct is then transported to the target area where it binds to the surface-immobilized target DNA in shear orientation. The geometry of the 40 bp DNA duplex confers a substantial increase in rupture force due to force propagation through all basepairs (a most probable unbinding force of ≈65 pN at the observed loading rates around SI300pN/s) [172]. Meanwhile, the SII:monoST complex ruptures in a non-equilibrium process at ≈45 pN at this pulling speed and apparent loading rate. Thus, the transfer construct detaches from the cantilever upon retraction and remains deposited in the target area. The cantilever-coupled monoST complex is now free to pick up a new transfer construct in repeated SMC&P cycles. This therefore presents the opportunity to carefully tune the expected rupture force of a protein-based handle by varying the loading rate, while simultaneously incurring minimal or no effect on the expected rupture force of a DNA-based tether. Hence, the SMC&P force hierarchy can be further reinforced by adjusting the pulling speed in each step of the cycle to maximize the difference in rupture force of the cantilever handle and surface tether. Examples of single-molecule pickup and deposition events demonstrate the plateau-like force curves observed from basepair-by-basepair unzipping of DNA in the depot area (Figure 7.2a) and the worm-like chain [78] stretching behavior of the PEG linkers followed by a sharp unbinding peak of the SII:monoST complex in the target area (Figure 7.2b). Notably, no additional force barriers are observed, which is consistent with the GFP and *Strep*-Tactin fold staying intact throughout the transfer process.

Following completion of the SMC&P sequence, the GFP-containing transfer molecules arranged in patterns of a thumbs up and a dinosaur were imaged by TIRF microscopy (Figure 7.3), producing clearly discernible outlines. The previously demonstrated longevity of the monoST complex after hundreds of pulling events [69] is confirmed here with successful

SMC&P transport over 487 and 395 consecutive cycles, respectively. Moreover, the rupture forces exhibited by the SII:monoST complex are in a range that is compatible with the well-characterized depot-transfer and target-transfer DNA duplex unbinding. Patchiness in the pattern can be partially attributed to the limited photostability of GFP, likely causing a fraction of the transfer molecules to photobleach during purification and experimental setup before imaging. There are also cases where a cycle fails to transport a transfer construct, as SMC&P and the underlying rupture forces are probability-dependent. Surface defects and densities can also influence the efficiency of SMC&P. However, the corresponding force-distance curves in every cycle can control for this; an inherently non-fluorescent or bleached transfer construct produces a deposition force curve but no fluorescence signal, and a failed transport cycle produces neither (e.g. the bottom-most force traces in Figure 7.2). Such analysis was previously executed by Pippig et al. to evaluate a widely-spaced grid pattern of individual GFP molecules deposited by SMC&P [86]. The same strategy could conceivably be applied to any SMC&P experiment in which it is necessary to determine exactly which deposition points contain the transported molecules of interest.

7.4 Conclusion and Outlook

SMC&P is an attractive strategy for the study of enzyme activity, as it enables precise placement of molecules of interest on a surface with known positioning, in contrast to stochastic surface immobilization. Moreover, the pattern of molecules to be arranged is completely arbitrary, allowing unrestricted pattern design for the investigation of the effects of relative geometry within networks or clusters of enzymes. Single-molecule enzyme analyses may prove to be indispensable for gaining new insights into the dynamic nature of enzyme networks, such as cellulosomes. These cellulose-degrading complexes are utilized by many species of cellulolytic bacteria, and have the critical characteristic that cellulase enzymes are strategically arranged on a scaffold to increase the efficiency of the catalytic network [173, 174]. As the spatial organization of the enzymes is a key part of their function *in vivo*, similarly taking into account their relative geometry could provide new understanding of the cellulosomal components on a single-molecule level. Therefore, SMC&P may be an advantageous strategy for directed arrangement and investigation of this enzyme network.

Identification of diverse surface-immobilization methods is likewise a prerequisite for SMC&P's versatility. An intriguing potential use of the SII:monoST complex in SMC&P would leverage its terminus-dependent rupture force regimes to create SII-harboring depot and target regions. A monoST-containing transfer construct could be transported from the N-terminal SII depot (lower rupture forces) to the C-terminal SII target (higher rupture forces) via a cantilever tag of some intermediate rupture force. This presents an opportunity to forego DNA-based anchoring systems, which could be crucial for studying DNA-binding enzymes such as polymerases or ligases.

We have demonstrated that the SII:monoST complex is a viable handle for SMC&P. Although this study serves mainly as a proof of concept, it is invaluable to the expansion of SMC&P

that diverse immobilization and cantilever handling options are available. Arrangement of molecules that perform biological functions – such as enzymes or aptamers – requires careful consideration of the unique properties, requirements or limitations of each molecule of interest. Unexpected secondary interactions between a tethering system and an enzyme could hinder SMC&P transport or interfere with enzymatic activity. In other words, it is possible that universal means of surface immobilization or cantilever handling may ultimately not exist. It is therefore advantageous to have a modular system with exchangeable components so that no molecule of interest must be excluded from study. With the addition of the SII:monoST handle system to the SMC&P toolbox, we have expanded the handling options available and given this technique a new degree of flexibility.

Chapter 8

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

Summary

Single-molecule cut-and-paste presents an unprecedented opportunity for the bottom-up assembly of biomolecular networks, thereby enabling time-resolved analysis of complex biological systems in defined geometries on the single-molecule level. However, this technique requires versatile handling systems to enable arrangement of diverse molecules. The novel DNA-free scheme described here facilitates assembly of enzymes and enzyme networks.

8.1 Introduction

The spatial organization 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 [91] and prominently DNA origami [94, 97] – arguably the most widely-used self-assembly technique. Notably, a recent novel drug-delivery strategy via activated DNA origami showed potent tumor-inhibiting activity [96], 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 utilizes an atomic force microscope (AFM) cantilever tip to pick up and deposit single molecules with nanometer precision at defined positions on a surface.

This chapter was adapted from a manuscript that is currently in preparation and is likely to be submitted for peer-review in November 2018 [103]

SMC&P relies on a pre-programmed force hierarchy to facilitate the transfer of molecules from the depot area to the cantilever tip 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 molecules within nanoapertures such as zero-mode waveguides, thereby circumventing complications such as interference from metallic sidewalls that inevitably result from stochastic immobilization [2].

Previous iterations of SMC&P have undertaken arrangements and time-resolved fluorescent measurements of various biomolecules, including labeled DNA, DNA aptamers, green fluorescent protein, nanoparticle recognition sites, and diverse handling tags [1, 86, 98–102, 175] – 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 emergent technique, SMC&P compels improvements and developments to increase its robustness. In particular, SMC&P has previously relied on DNA anchoring of molecules to the surface. Although this strategy confers reliable and stable immobilization of transfer molecules, its scope is limited. A DNA-based approach is unsuited for the study of DNA-binding molecules – such as DNA-binding enzymes – as the molecule of interest would bind its own covalently attached DNA anchor, thereby impacting its behavior as well as SMC&P transfer. Arrangement of proteins additionally requires synthesizing protein-DNA hybrid molecules – a process that is often not straightforward. Moreover, SMFS analysis in SMC&P has also previously had limited applicability; probed molecules have lacked fingerprint domains to identify specific single events, and the low-force regimes of the handling systems were partly overlaid with the instrument noise.

Here, we present a revamped 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, instead relying on a protein-ligand interaction for surface immobilization. Simultaneously, a reliable fingerprint domain and increased rupture forces significantly improve SMFS analysis of SMC&P transfer both in real-time and in statistical analyses.

8.2 Materials and Experimental Methods

The experiments described in the manuscript were performed on an AFM/TIRFM hybrid, the details of which may be found in Gumpp et al. [156].

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 [40, 160]. 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-and-paste cycle amount to approximately 3 s in these experiments.

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-color 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 utilized 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 multicolor 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.

Preparation of Monovalent Streptavidin: Monovalent Streptavidin (mSA) with N-terminal immobilization (N-mSA) was previously expressed, purified and assembled by Sedlak et al.[49] C-terminally immobilized 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) [114]. 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 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 β-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

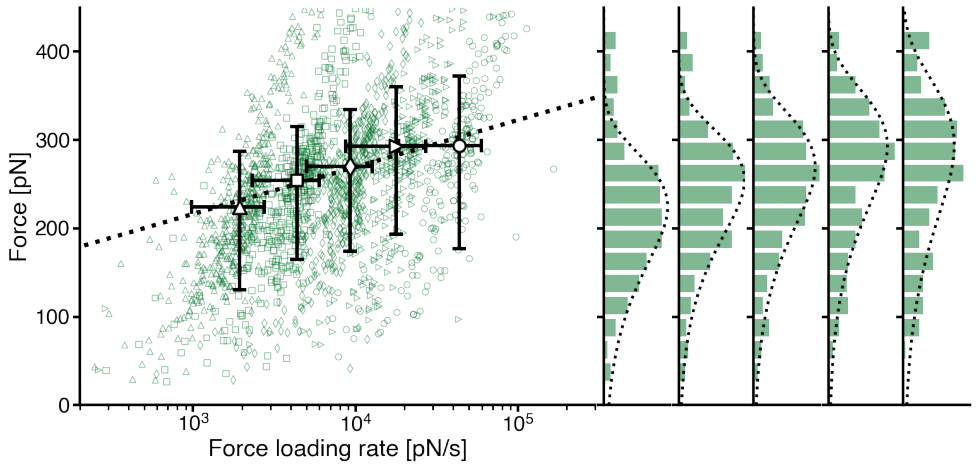


Figure 8.1: Dynamic force spectrum of the rupture of N-terminally immobilized SdrG and an N-terminal Fgβ tag measured with AFM-based SMFS. In this geometry, the SdrG:Fgβ bond ruptures between approximately 200 – 300 pN, depending on the loading rate.

fractions were analyzed by gel electrophoresis. The eluted monovalent streptavidin was dialyzed against phosphate buffered saline, pH 7.4 and stored at 4 °C.

Preparation of SdrG: SdrG N2N3 was previously expressed and purified as described by Milles et al.[36] In brief, SdrG was expressed in a pET28a vector with a 6xHis-tag and an N-terminal ybbR tag for covalent immobilization to CoA. 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 [152] 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 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 to be stored at -80 °C until used in experiments. The final protein concentration was 848 µM as measured by the absorbance at 280 nm via NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA).

N-terminally immobilized 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 (Figure 8.1).

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 was designed with several handling and purification tags. The construct harbors 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 harbors an internal ybbR-tag (DSLEFIASKLA) to covalently modify the protein with Biotin. A

Sortase tag (LPETGG) was also included, although not directly utilized 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 [152] containing 100 µg/mL Kanamycin and grown at 37 °C for 20 h. As the medium is designed to automatically induce expression when the cell density reaches a critical point such that all glucose is consumed, no addition of an inducing agent *e.g.* IPTG was necessary.

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 8.1).

Selected fractions from the major peak in the chromatogram were analyzed 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 binds to tryptophan residues and is activated by exposure to UV light. 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 approximately 700 µM as measured by the absorbance at 280 nm via NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA).

Cy5- and Biotin-Labeling of ddFLN4: The ddFLN4 transfer construct was modified first with Cy5 in a cysteine-maleimide reaction, followed by Biotin in an Sfp-catalyzed 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

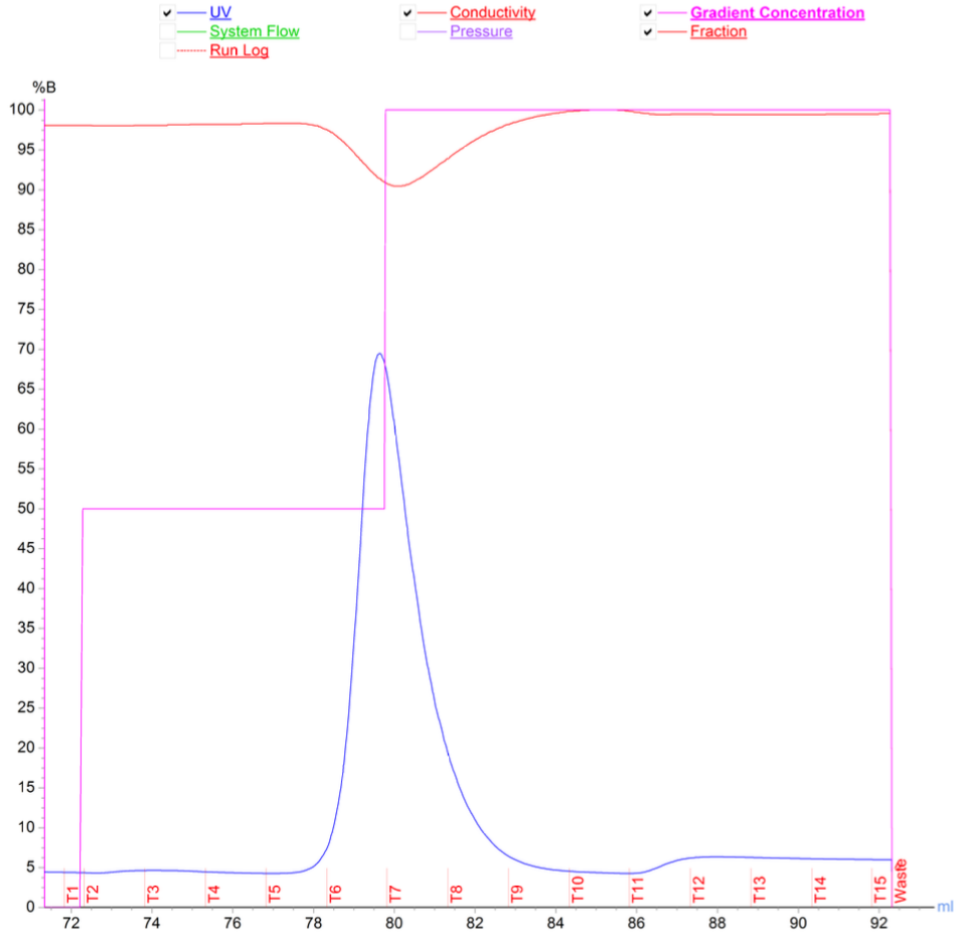


Figure 8.2: Chromatogram of ddFLN4 construct purification by Ni-IMAC. The 6xHis-tagged protein was purified by step gradient and eluted with high imidazole. Fractions 6 - 8 span the majority of the major peak, with a smaller peak spanning fractions 11 - 13. Note that there is a 6 mL delay in reported 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.

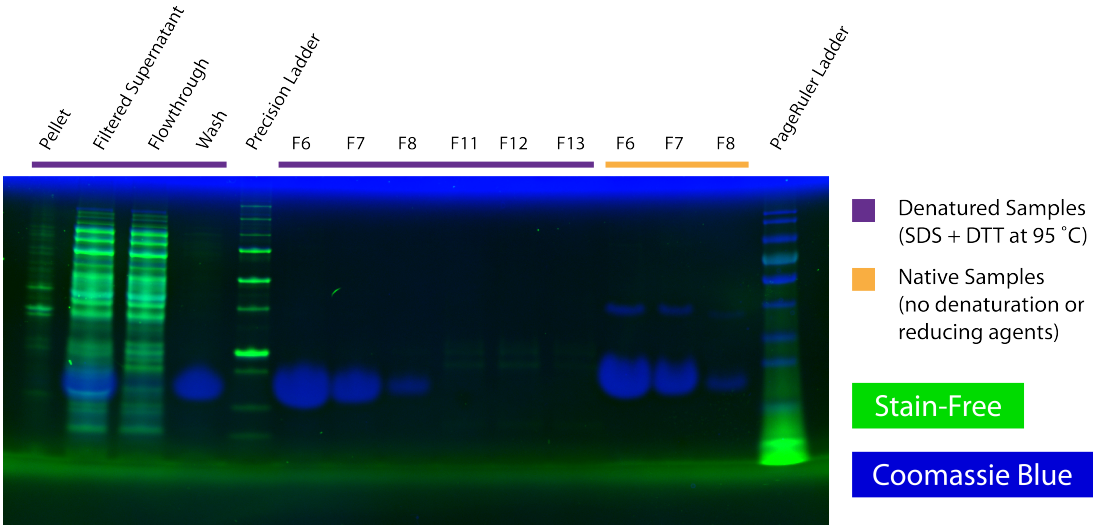


Figure 8.3: SDS-PAGE and Native PAGE analysis of the ddFLN4 protein construct purification by Ni-IMAC. The ddFLN4 construct was obtained in the soluble fraction and efficiently eluted by step gradient purification. Fractions 6 - 8 contained the majority of the protein, visible by the overloaded blue bands. These fractions were additionally analyzed under native conditions to examine the extent of dimerization 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.

of 7 nmol of purified ddFLN4 transfer construct protein and 50 nmol Mal-Cy5 in 1x Cysteine-Maleimide Reaction 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 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-labeled 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 8.3), and fractions were collected in 100 µL increments.

Fractions from the most prominent peaks were analyzed via SDS-PAGE to determine which peak contained Cy5-labeled ddFLN4 transfer construct (Figure 8.4). 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 analyzed to identify the optimal fractions (Figure 8.5). Fractions from the major peak as well as secondary peaks were also assessed for Biotin labeling via Native PAGE (Figure 8.6). Selected fractions that demonstrated efficient labeling with both Cy5 and Biotin were pooled (2B12-2B10; 2B9-2B7;

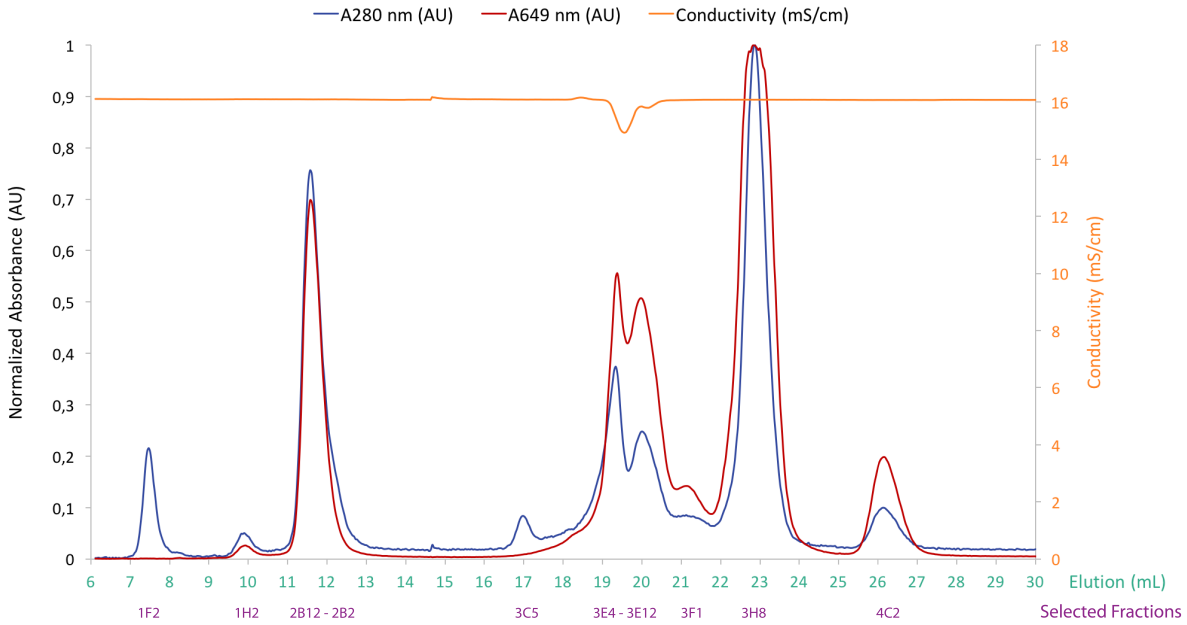


Figure 8.4: Chromatogram of Cy5- and biotin-labeled ddFLN4 transfer construct purification by size-exclusion chromatography. The absorbance at 280 nm and 649 nm is used to estimate the concentrations of protein and Cy5, respectively. Selected fractions (bottom purple labels) were further analyzed by SDS-PAGE.

2B6-2B3) and stored at -80 °C at a final concentration of approximately 1 µM as measured by the absorbance at 649 nm via NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA).

Preparation of Cantilevers: Cantilevers (BL-AC40TS, BioLever mini, Olympus, Japan) were oxidized in a UVOH 150 LAB UV-ozone cleaner (FHR Anlagenbau GmbH, Ottendorf-Okrilla, Germany). Silanization 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 min, the silanized cantilevers were placed in 25 mM heterobifunctional polyethylene glycol crosslinkers of 5000 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 1 h. 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 transferase, 10 mM magnesium chloride and 50 mM HEPES at pH 7.5. Sfp transferase covalently joins Coenzyme A on the surface and the ybbR-tagged proteins [148]. The functionalized cantilevers were washed and stored in phosphate buffered saline.

Preparation of Glass Surfaces: Glass cover slips were sonicated in 50 %(v/v) 2-propanol in filtered H_2O for 15 min and oxidized in a solution of 50 %(v/v) hydrogen peroxide (30 %) and

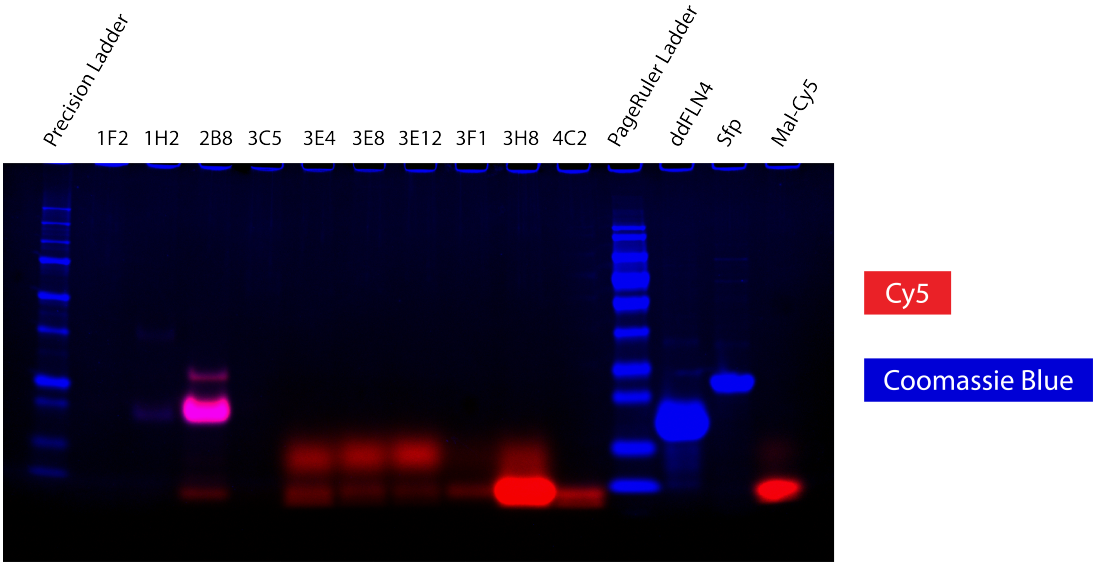


Figure 8.5: SDS-PAGE analysis of selected fractions of labeled ddFLN4 transfer construct purification. In addition to fractions from size-exclusion chromatography, samples of unreacted ddFLN4, Sfp and Mal-Cy5 were included as controls. The major peak spanning fractions 2B12 – 2B2 contained a high concentration of Cy5-labeled ddFLN4 as well as a small amount of co-eluted Cy5-labeled Sfp by-product.

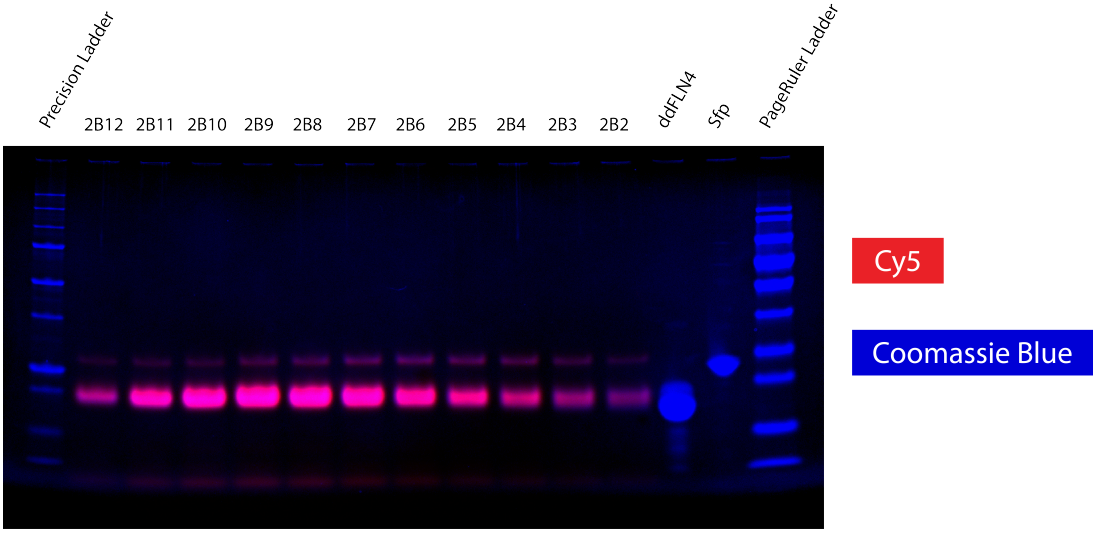


Figure 8.6: SDS-PAGE analysis of selected fractions from the major peak of labeled ddFLN4 transfer construct purification. In addition to fractions from size-exclusion chromatography, samples of unreacted ddFLN4 and Sfp were included as controls. The fractions contain a high concentration of labeled ddFLN4 as well as a lesser amount of co-eluted Cy5-labeled Sfp.

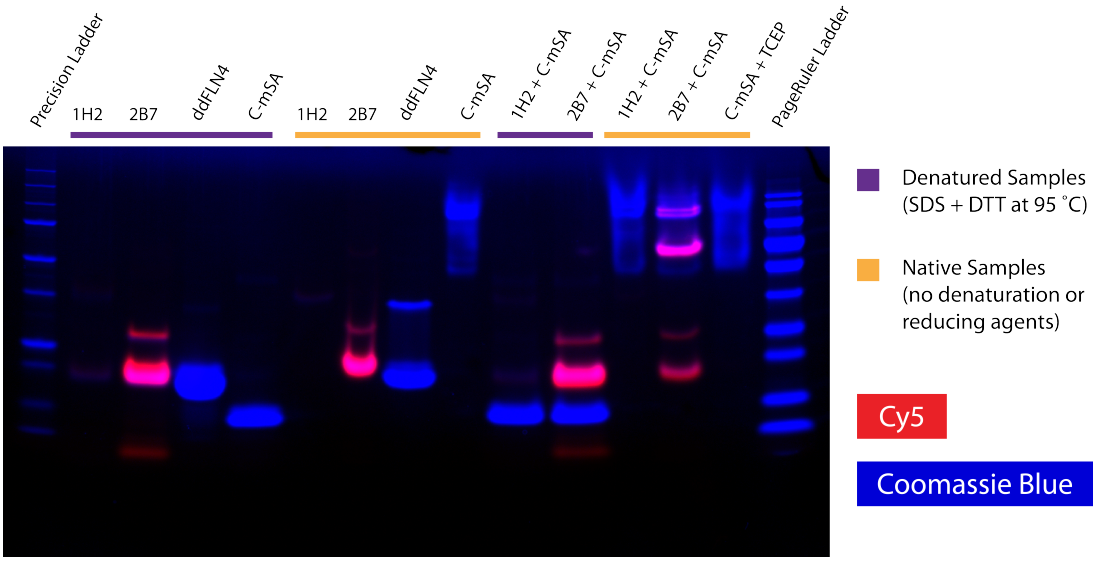


Figure 8.7: SDS-PAGE and Native PAGE analysis of selected fractions from labeled ddFLN4 transfer construct purification and binding to C-mSA. Fraction 2B7 from the major peak of Cy5-labeled ddFLN4 binds C-mSA under native conditions, visible by a band-shift in both Cy5-labeled ddFLN4 and C-mSA. This strongly suggests that at least the majority of Cy5-labeled ddFLN4 is also biotin-labeled.

sulfuric acid for 30 min. They were then washed in filtered H_2O, dried in a nitrogen stream and then silanized by incubating for 1 h in (3-Aminopropyl)dimethylethoxysilane (ABCR, Karlsruhe, Germany, 1.8 %(v/v) in Ethanol). The silanized 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. [98] – was fixed on the aminosilanized 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 [158, 159] with N-hydroxy succinimide and maleimide groups (molecular weight 5000 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 (mSA) 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 Desalting Columns, 7K MWCO (Thermo Fisher Scientific, Waltham, MA, USA). Freshly-reduced mSA was immediately applied to the PEG-functionalized 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 labeled 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 nonspecifically-bound ddFLN4. The microfluidic system was then removed and the surface submerged in PBS.

SMC&P Experiment: The rocket pattern was written in 442 transfer cycles with 200 nm spacing between each deposition point. The pulling speed in the depot was set to 3200 nm/s and in the target to 200 nm/s. This corresponds to approximate surface contact times [161] (dependent on approach/retraction velocity, indentation force and substrate stiffness) of 5 ms and 80 ms, respectively, and should allow for ligand binding. Considering a single SdrG molecule being bound to the cantilever tip and estimating its localization 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 (400 nM) [36] and the mSA:biotin interaction (<1 nM) [49]. 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.

Rupture forces and loading rates were evaluated from AFM force distance curves that were recorded for each pickup and deposition process utilizing the WLC model [162] with quantum mechanical correction (force spectroscopy data were evaluated in Python 2.7, Python Software Foundation).

8.3 Results and Discussion

Monovalent streptavidin (mSA) – a heterotetrameric complex that binds the small molecule biotin with high affinity – was recently employed in AFM-based SMFS [49]. 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 [113]. This geometry-dependent behavior was exploited in SMC&P to immobilize the transfer molecule with both low- and high-rupture forces via the same small biotin label. The adhesin SD-repeat protein G (SdrG) N2N3 domain from *Staphylococcus epidermidis* binds the short peptide from human fibrinogen β (Fgβ) with remarkably high AFM-measured rupture forces of 2 nN when probed in the native geometry of C-terminally immobilized SdrG and C-terminally pulled Fgβ18. When probed in the non-native geometry of N-terminally immobilized SdrG, the unbinding forces are in the range of 200 - 300 pN, depending on the loading rate (Figure S1). These binding pairs of N-mSA:biotin in the depot area, 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 re-folding as a low-force fingerprint in AFM-based SMFS [70, 176, 177]. A transfer construct consisting of a modified ddFLN4 motif was designed, expressed and purified with several key additional sequences. Namely, an N-terminal Fgβ peptide sequence – which would accordingly be pulled C-terminally – enabled specific handling by an SdrG-coupled

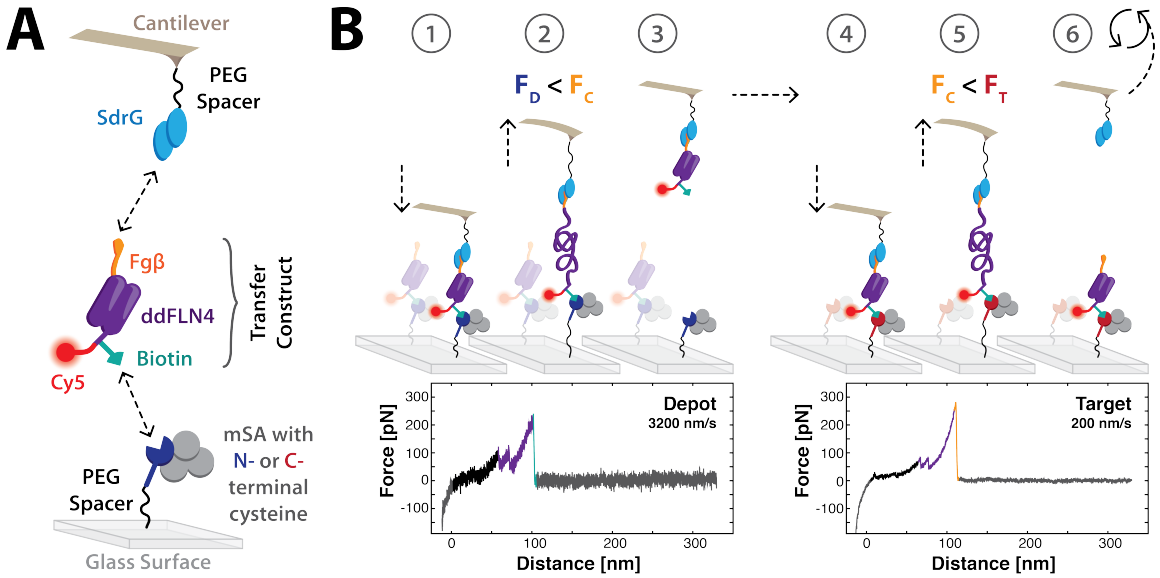


Figure 8.8: Schematic of the molecules used in SMC&P and the mechanism of SMC&P cycling. (A) SdrG is N-terminally immobilized to the cantilever tip, and monovalent Streptavidin with an N- or C-terminal reactive cysteine is immobilized on a glass surface. The chimeric transfer construct is composed of a ddFLN4 domain, which contains an N-terminal Fgβ 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 immobilization on a Streptavidin-functionalized 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 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 SdrG binds the Fgβ tag of an immobilized transfer construct (1). As the cantilever tip 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 SdrG:Fgβ bond ruptures. The unloaded cantilever tip is recycled back to the depot area to repeat the process (6). Force-distance curves of specific single-molecule interactions show the ddFLN4 unfolding pattern (purple traces) and a higher final peak associated with the rupture of N-mSA:Biotin in the depot (green trace) or SdrG:Fgβ in the target (orange trace).

cantilever tip. The ddFLN4 domain also harbored at its C-terminus a ybbR tag followed by a C-terminal reactive cysteine to enable covalent modification with coenzyme A (CoA)-biotin and maleimide-Cy5, respectively. The final transfer construct consisted of an efficiently labeled Fgβ-ddFLN4-biotin-Cy5 chimera (details of purification and labeling may be found in Materials and Experimental Methods) that binds to mSA via biotin and is imaged in TIRF microscopy via Cy5 (Figure 8.8a).

A custom-built hybrid AFM/total-internal-reflection fluorescence (TIRF) microscope was employed here for SMC&P [156]. The depot area consists of N-mSA covalently attached to the surface, and transfer constructs that are specifically immobilized via the biotin label. In the target area, C-mSA is covalently attached to the surface. A cantilever tip functionalized with 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), SdrG:Fgβ on the cantilever tip (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 is fundamental to SMC&P (Figure 8.8b). Transfer construct molecules bound to N-mSA in the depot area are pulled by an SdrG-coupled cantilever tip. The forces required to rupture both the N-mSA:biotin bond and the 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 bond ruptures, the force load is decreased, and the ddFLN4 motif undergoes rapid refolding. The cantilever tip, loaded with the transfer construct cargo, is then moved to the target area. As the cantilever tip approaches the surface, the C-mSA:biotin bond forms, thereby immobilizing the transfer construct to the surface again. The cantilever tip retracts and again unfolds ddFLN4, visible in single-molecule force-distance curves. As the C-mSA:biotin bond is stronger, the SdrG:Fgβ bond eventually ruptures. The ddFLN4 motif of the immobilized transfer construct again rapidly refolds, and the cantilever cycles back to the depot to repeat the process. Force-distance curves captured during SMC&P reflect the unfolding of the ddFLN4 fingerprint domain followed by a final rupture of either mSA:biotin or SdrG:Fgβ.

The retraction velocities of the cantilever tip 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. The rupture force of the SdrG:Fgβ bond demonstrates a stronger dependence on loading rate (Figure 8.1) compared to the N- and C-mSA:biotin bond [113]. This difference in loading rate dependence was exploited to favor the rupture of the lower-force binding pair and hence relocate the transfer construct. Fast retraction (3200 nm/s) in the depot made it possible to increase the likelihood of the rupture of N-mSA:biotin over SdrG:Fgβ, while slow retraction (200 nm/s) in the target favored the rupture of SdrG:Fgβ over C-mSA:biotin. Observed final rupture peaks in both the depot and the target correspond to the approximate expected rupture forces for the two receptor-ligand pairs at the given loading rates (Figure 8.9).

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 (Figure 8.10). Fluorescent immobilized molecules

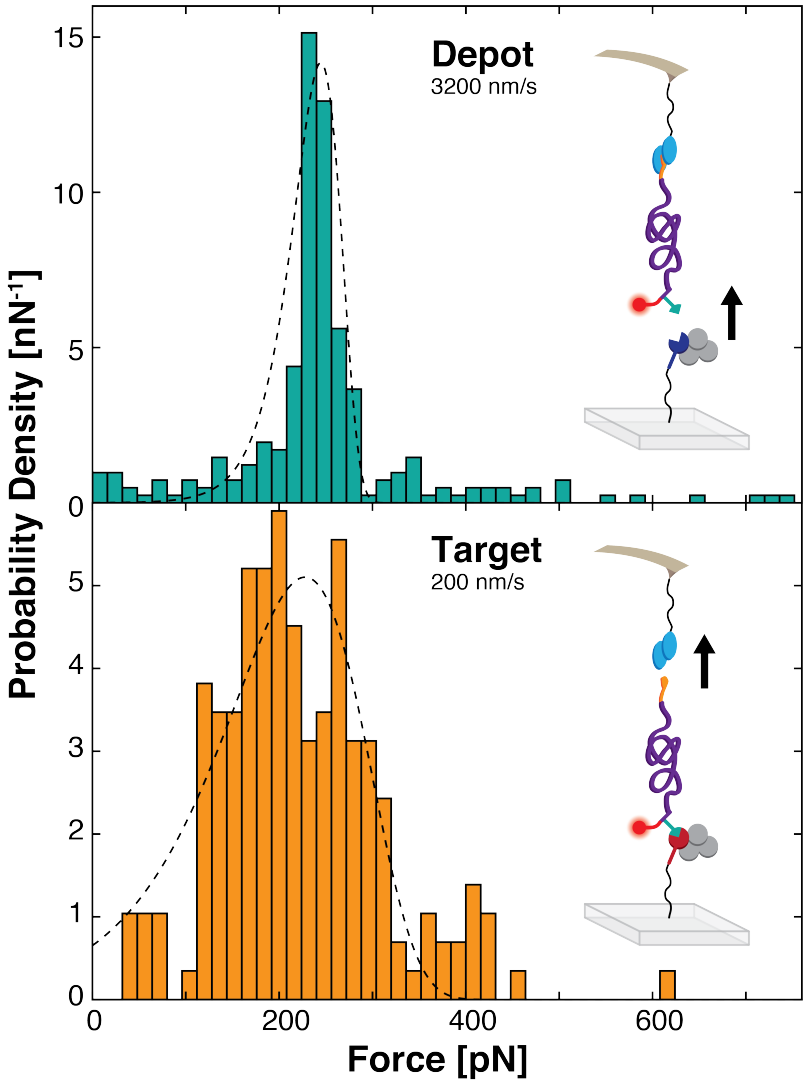


Figure 8.9: Forces associated with the final peaks of force traces observed in the depot (top) and target (bottom) areas during SMC&P. The depot curves correspond to the unbinding of the N-mSA:biotin, while the target curves correspond to the unbinding of SdrG:Fgβ. Each complex has an expected rupture 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 Bell-Evans theory (dashed lines) [75, 76].

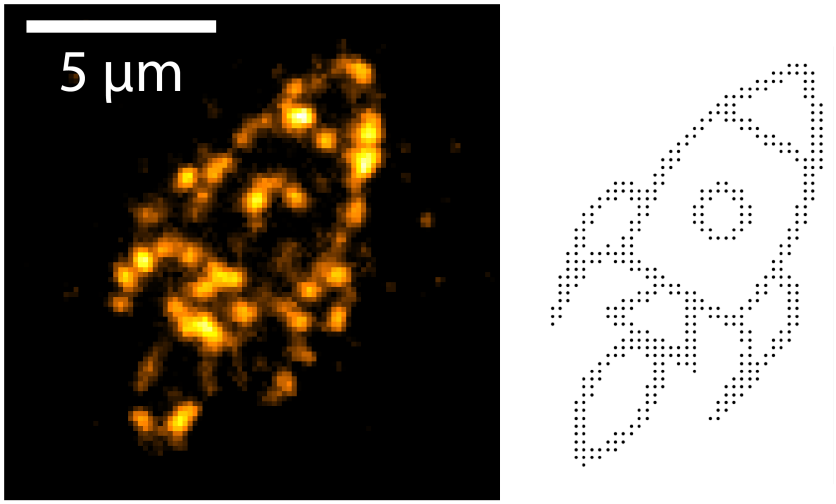


Figure 8.10: Cy5-labeled transfer molecules arranged in a rocket ship pattern by SMC&P and imaged with TIRF microscopy. The image is composed of the average pixel intensity of 20 stacked frames (0.12 s exposure time at approximately 10 W/cm^2) with red laser excitation at 640 nm (left). The pattern consists of 442 points spaced 200 nm apart (right).

were detected via Cy5 excitation at 640 nm and imaged with TIRF microscopy. Patchiness in the pattern may be partially due to incomplete labeling or photobleaching of transfer constructs during purification and experimental setup. Additionally, as the underlying rupture forces in SMC&P are probability-dependent, there are cases where a cycle fails to transport any molecules. Similarly, there is a certain probability that transfer constructs bound to the cantilever tip dissociate during transport. Surface defects and 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 nonfluorescent 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 utilized 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 a covalently-attached DNA anchor with high affinity. 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 interact with the surface. The immobilization strategy presented here is likely orthogonal for most biomolecules, thereby significantly increasing the versatility of the system.

Post-translational labeling of proteins with nucleotides in a controlled manner is also not a trivial process. As performed previously in SMC&P, proteins may be labeled with CoA-DNA via a ybbR tag and reaction with Sfp (as was performed here similarly for labeling with biotin), although this is a step that is necessarily performed post-translationally and *in vitro*. In contrast, biotin labeling may be performed *in vivo* during protein production with additional recombinant factors, such as an AviTag. Similarly, Cy5-labeling may be replaced with a fluorescent protein

domain, such as green fluorescent protein. On the other hand, the utilized strategy of cysteine-based labeling forgoes a need to create large chimeric protein constructs and enables fluorescent imaging of any protein of interest.

The introduction of the small ddFLN4 fingerprint domain is also exceptionally useful for force trace analysis. As a well-characterized and reliable fingerprint, ddFLN4 improves algorithmic curve sorting to isolate single and specific pulling events. Moreover, ddFLN4 was demonstrated to improve solubility of otherwise insoluble recombinant proteins [70]. In order to protect a domain of interest from the force propagation pathway, it 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 [157].

8.4 Conclusion and Outlook

Single-molecule studies offer unique perspectives in the investigation of biomolecular function. For example, spatial arrangement of isolated molecules opens up the possibility to examine time-resolved steps within a chemical reaction. Measurements of enzyme activity in bulk – the most common strategy of assessing enzyme activity, often by monitoring product accumulation – are by nature only capable of reporting the average behavior within a population. In contrast, single-molecule studies provide an avenue to examine the distribution of behavior within an ensemble. Considering the direct connection between enzyme structure and function as well as the nuanced process of protein folding, it is likely that a population of molecules will display heterogeneous final folds and accordingly a distribution of catalytic activity. Single-molecule approaches therefore offer insights into enzyme function that are otherwise inscrutable.

SMC&P enables precise arrangement of networked molecules on a surface in well-defined geometries as well as within the centers of nanoapertures, demonstrating the unique potential of this technique. However, the previously established DNA-based SMC&P immobilization 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 behavior as well as SMC&P efficiency. The mSA:biotin system introduced here offers an immobilization strategy that is orthogonal to the function of most enzymes. Furthermore, the diverse reactive tags allow for flexible construct design, and the utilized construct's ddFLN4 fingerprint enhances SMFS analysis. The advances demonstrated here set a methodological foundation for the single-molecule arrangement and analysis of diverse networked molecules, thereby providing a means to gain a more comprehensive understanding of biomolecular function.

Part IV

Conclusion and Outlook

Chapter 9
Enzyme Networks by Design

A key goal across the broad field of nanotechnology is the precise and reliable control of molecules of interest, especially within complex systems. The spatiotemporal potential of specifically arranged enzyme networks cannot be overlooked; the knowledge gleaned from single-enzyme and networked-enzyme studies would elucidate much about geometry- and density- dependence, time-resolved behavior, activity heterogeneity, and mechanistic steps in catalytic processes.

An expanded repertoire of handling strategies improves the adaptability of low-force SMC&P that employs DNA anchors between the transfer construct and the surface. Although a DNA-based immobilization system may not be compatible with all molecules of interest - specifically DNA-binding enzymes - the extreme chemical and thermodynamic stability of double-stranded DNA is an invaluable asset for maintaining patterns of immobilized molecules over very long timescales. Therefore, systems in which DNA does not interfere with functionality may still benefit from DNA-based immobilization systems or low-force SMC&P, and by extension the improved versatility of handling options granted by monoST:SII.

Nevertheless, the implementation of DNA as an anchoring strategy is not universally applicable. As demonstrated with T7 RNAP, a covalently-attached DNA strand can interfere with enzyme function. Simultaneously, should an enzyme of interest bind its covalently-attached DNA anchor, the DNA strand is no longer free to readily form a bond with complementary DNA on the depot- or target- surfaces, thereby decreasing the efficiency of initial immobilization as well as SMC&P.

Importantly, SMC&P has been further developed to include a completely DNA-free system. The introduction of the mSA:biotin interaction as a long-lived surface immobilization paradigm and SdrG:Fgβ as a cantilever handle paves the way for enzymes to be successfully incorporated into SMC&P. This strategy offers orthogonality between the cargo and handling systems, thereby increasing the prospect of high SMC&P efficiency as well as enzyme activity. Furthermore, the implementation of a ddFLN4 fold as a reliable fingerprint improves statistical analysis of

SMC&P experiments that were previously not feasible, and the improved solubility gained from ddFLN4 will likely prove advantageous for enzymes and proteins of interest that otherwise prove difficult to recombinantly express.

The results presented in this work concerning fluorescent readout strategies for enzymes of interest constitute the first steps in further development of this dynamic technique for the arrangement of enzyme networks. In addition to continued optimization of the assays, advances may surely be made to other aspects of this system. For example, surface passivation may be improved and tuned to minimize background signal, especially in systems where fluorescent molecules in solution have a high probability to adhere nonspecifically to the surface and create confounding background signal.

The work presented here lays a foundation for bottom-up directed assembly of enzyme networks via SMC&P. Given that a population of enzymes will likely present a distribution of final folded substates - and consequently a distribution of behavior - hidden heterogeneities of enzyme activity may be identified with single-molecule analysis. Furthermore, by arranging enzymes in specific patterns on a surface, the kinetic behavior of enzymes acting in concert can be probed with time resolution on the single-molecule level. With new advances in nanotechnology - and SMC&P specifically - a deeper understanding of the nuanced behavior of biomolecules may be revealed.

Part V

Appendix

Appendix A
Research Article in Small Methods Chapter 7 in Original Form

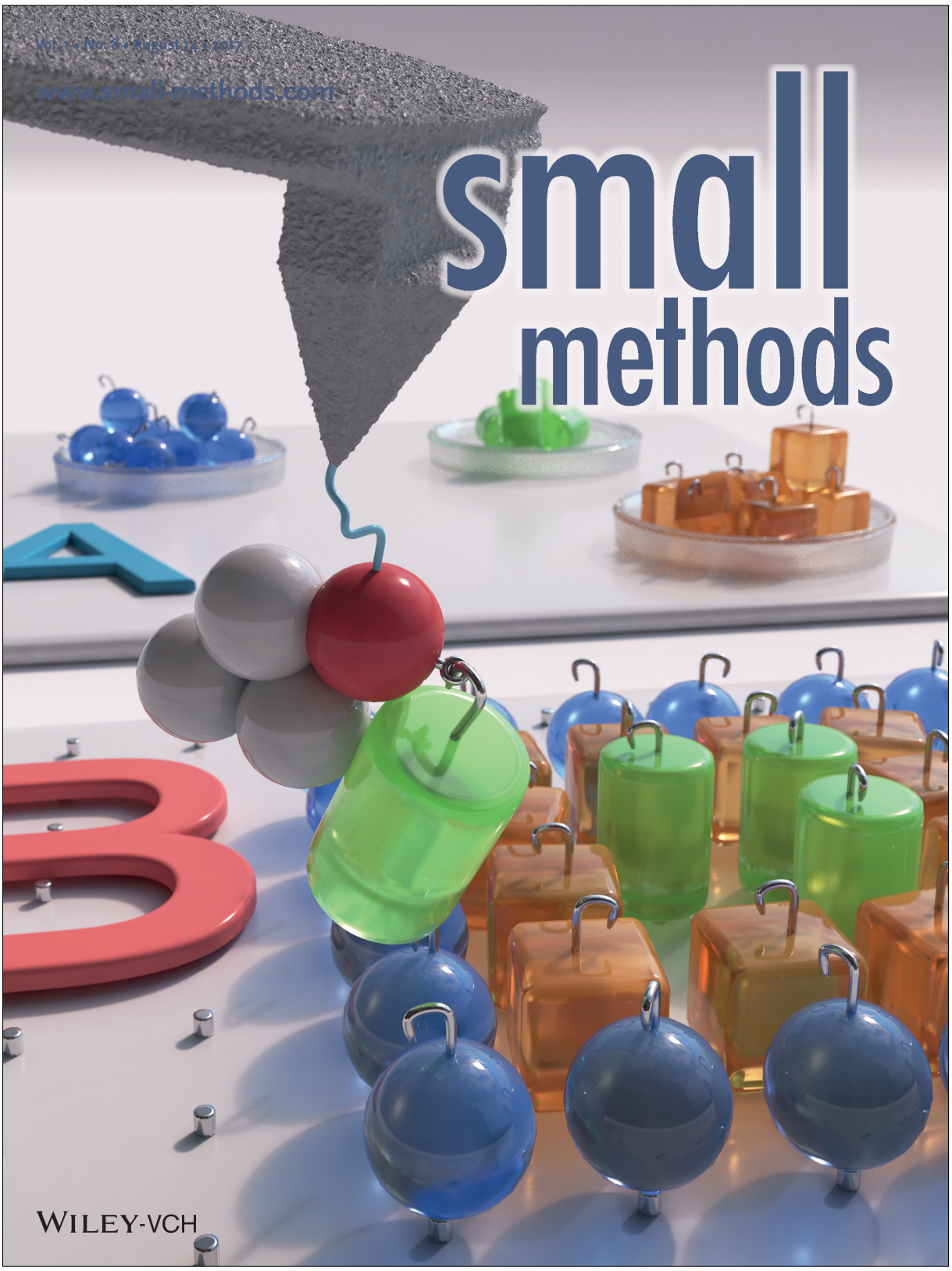
Strep-Tag II and Monovalent *Strep*-Tactin as Novel Handles in Single-Molecule Cut-and-Paste

by

Katherine R. Erlich, Fabian Baumann, Diana A. Pippig, and Hermann E. Gaub

published in

Small Methods, 1(8):1700169 DOI: 10.1002/smtd.201700169 (2017)
Reprinted from Erlich *et al.* [102] with permission from Wiley-VCH


small
methods
www.small-methods.com
WILEY-VCH

COMMUNICATION

Single-Molecule Spatial Arrangement

small methods

www.small-methods.com

Strep-Tag II and Monovalent *Strep*-Tactin as Novel Handles in Single-Molecule Cut-and-Paste

*Katherine R. Erlich, Fabian Baumann, Diana A. Pippig, and Hermann E. Gaub**

Directed spatial assembly of single molecules on a surface presents an opportunity to precisely control the positioning, density, and geometry of molecules of interest within an ensemble. In contrast to bulk averaging, this enables detection and analysis of individual behavior within such a designed ensemble. The atomic force microscopy (AFM)-based technique of single-molecule cut-and-paste (SMC&P) facilitates the arrangement of a variety of biomolecules on a surface through different handling strategies. This technique requires cantilever- and surface-handles that simultaneously adhere to a prerequisite rupture force hierarchy, and also do not cross-interact with each other or the transported molecules. As the molecules of interest diversify, so too must the handling methods to accommodate their unique characteristics. Here, it is demonstrated that a previously developed monovalent variant of *Strep*-Tactin and its corresponding *Strep*-Tag II peptide ligand comprise a viable cantilever handling complex for SMC&P. Ultimately, this expansion to the SMC&P toolbox increases the system's versatility for new molecules of interest yet to be studied.

The frontier of nanoscale studies frequently presents unexpected challenges that must be overcome with innovation. As such, universally applicable approaches often do not exist, and instead diverse methods or tools must be developed. Bottom-up synthetic biology employs fundamental biological components as the building blocks for artificial biological systems with novel characteristics. A major endeavor of this broad field is to develop unique molecular-organization techniques, such as engineered protein modules[1] and enzyme cascades assembled on DNA-origami scaffolds.[2] Single-molecule cut-and-paste (SMC&P) is one such organization technique, merging bottom-up assembly with control on the level of single molecules. SMC&P utilizes the single-molecule force spectroscopy (SMFS) and lateral surface positioning of atomic force microscopy (AFM)[3–5] to deposit molecules of interest in arbitrary patterns on a functionalized glass surface with nanometer-precision. There exist several key challenges in SMFS that are constantly improved upon: signal-to-noise limits of data resolution, and specific handling. Advances in existing tools, such as modified cantilevers in AFM-based SMFS,[6] can greatly improve data quality and expose previously inaccessible levels of detail. Precise and versatile control of molecules of interest can likewise open up new avenues of study. For example, protein-labeled DNA-tethering strategies in optical-tweezers experiments offer additional flexibility in handling biomolecules for mechanical measurements.[7] Similarly, SMC&P requires robust immobilization and handling schemes for the specific and controlled arrangement of diverse biological agents. With a more expansive repertoire of handling strategies, its components could potentially be tailored to enable arrangement of any molecule of interest in a well-defined orientation.

During each SMC&P cycle, a noncovalently immobilized transfer molecule of interest is picked up from a depot area via a cantilever-coupled affinity handle. The cantilever relocates the transfer molecule to a target area, where it then deposits said molecule and is recycled back to the depot area to repeat the process. Arrays of molecules are assembled with precise localization in the target area, where their properties such as their fluorescent behavior in an ensemble or as individuals can be analyzed. For example, hybrid DNA–RNA molecules were specifically arranged and immobilized on a surface via complementary oligonucleotides. The resulting duplexes formed aptamers that stabilized the structure of a target dye molecule, enabling it to produce a fluorescent signal upon binding the SMC&P-arranged constructs.[8] Proteins have also been integrated into SMC&P, including constructs containing modified green fluorescent protein (GFP)[9] with different surface-immobilization strategies and a protein-based handle for the cantilever.[10,11] Directed placement of molecules within the nanoapertures of zero-mode waveguides with SMC&P is also possible. These nanoapertures facilitate measurements in a confined volume and improve background fluorescence when compared to conventional fluorescence microscopy. Importantly, the precise placement of molecules via SMC&P decreases the heterogeneity of fluorescence intensity and lifetime that results from stochastic immobilization and quenching effects from the metallic sidewalls.[12] In combination with fluorescence microscopy, SMC&P therefore presents new opportunities to examine biomolecular behavior on the single-molecule level with precise control of surface location and environment.

SMC&P fundamentally relies on a hierarchy of rupture forces of the specific interactions between the transfer molecule and the depot area (F_D), the cantilever (F_C), and the target area

K. R. Erlich, Dr. F. Baumann, Dr. D. A. Pippig, Prof. H. E. Gaub
Center for NanoScience and Department of Physics
Ludwig Maximilians University of Munich
Amalienstraße 54, 80799 Munich, Germany
E-mail: gaub@lmu.de

DOI: 10.1002/smtd.201700169

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

small methods
www.small-methods.com

(F_T) such that $F_D < F_C < F_T$. Consequently, this system demands a palette of selective immobilization and pickup methods so that the forces involved are tunable to fit this hierarchy. Furthermore, an ideal SMC&P scheme uses orthogonal agents that avoid unwanted cross-reactivity. Even small peptide tags and single-stranded DNA anchors could have adverse interactions between themselves or with molecules of interest, which in turn could interfere with SMC&P efficiency or molecule behavior. Therefore, a modular approach to construct design confers a degree of plasticity and promotes integration of any molecule of interest into SMC&P. By expanding the selection of cantilever- or surface-handles, the system can be further adapted to specific force and mechanistic requirements.

Strep-Tactin, an engineered variant of streptavidin, is a tetravalent complex that specifically binds with high affinity to the short peptide *Strep*-tag II (SII).[13] A monovalent version of *Strep*-Tactin (monoST) was recently developed and implemented in SMFS as a cantilever-immobilized handle for protein constructs harboring an SII-peptide.[14] The rupture forces of SII:monoST were found to be dependent on loading rate as well as location of SII in the protein construct, with N-terminal SII resulting in much lower rupture forces than C-terminal SII under identical loading rates. Its tethering geometry-dependent force regime and tunable rupture forces make the SII:monoST complex a compelling candidate as a handling system for SMC&P. Moreover, its addition to the growing SMC&P toolbox advances the technique toward the ultimate goal of enabling precise arrangement of any molecule of interest.

A construct consisting of GFP with an N-terminal SII and a C-terminal ybbR tag was previously expressed and purified, and it was also demonstrated in SMFS that the rupture forces of an N-terminal SII and monoST were lower than the force required to unfold GFP.[14] This construct was here employed in SMC&P, where the relatively low rupture forces of N-terminal SII and monoST were advantageous for preserving the fold and thus fluorescence of GFP during the transport process. The C-terminal ybbR tag was covalently modified with 3′-coenzyme A (CoA) single-stranded DNA by the phosphopantetheinyl transferase enzyme Sfp synthase.[15] Hybridization of the DNA anchor to a complementary DNA strand in either zipper- or shear-orientation enabled noncovalent surface immobilization to the depot and target areas, respectively (**Figure 1**a).

SMC&P was executed in a custom-built hybrid AFM/total-internal-reflection fluorescence microscopy (TIRFM) instrument.[16] Chimeric transfer constructs were picked up from the depot area via a monoST-coupled cantilever and deposited into the target area over 395 consecutive SMC&P cycles. The technique utilizes a hierarchy of rupture forces between the transfer molecule and the depot storage molecule (F_D), the cantilever handle (F_C), and the target storage molecule (F_T) such that $F_D < F_C < F_T$. The rupture forces of DNA duplexes are tuned via duplex length and pulling geometry (zipper vs shear).[17] Additionally, the rupture forces of protein–protein interactions are influenced by loading rates.[14,18,19] In this way, it is possible to optimize a scheme to consistently transport transfer constructs and regenerate the cantilever (Figure 1b).

The distinct behaviors of the depot DNA duplex and the SII:monoST complex upon unbinding and rupture are illustrated in their respective force curve patterns (**Figure 2**). When pulled apart in zipper orientation, the basepair-by-basepair

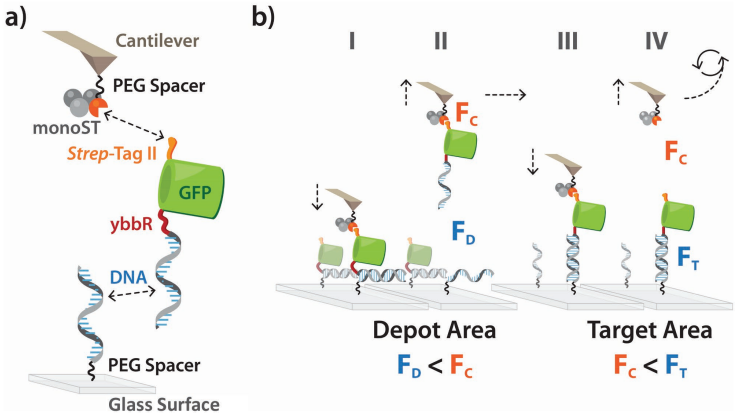


Figure 1. Monovalent *Strep*-Tactin and the GFP transfer construct employed in SMC&P. a) The chimeric transfer construct consists of a GFP molecule with an N-terminal SII as well as a C-terminal ybbR tag, which is then covalently coupled to 3′-CoA single-stranded DNA via reaction catalyzed by Sfp synthase. The DNA anchor binds noncovalently to the surface via a complementary DNA strand. The cantilever is covalently coupled to the single functional subunit of monoST, which targets SII of the transfer construct. b) Repeatable transfer cycling of SMC&P depends on a force hierarchy determined by DNA hybridization geometry and the SII:monoST interaction. The cantilever approaches the depot surface, and monoST binds to SII of a transfer molecule immobilized via complementary DNA in zipper orientation (I). The cantilever retracts and removes the transfer molecule as the DNA unzips (II). The cantilever then transports the transfer molecule to the target surface where the transfer molecule binds to complementary DNA in shear orientation (III). Retraction of the cantilever ruptures the SII:monoST complex, and the cantilever is recycled back to the depot area to repeat the process (IV). The rupture forces of the transfer construct with the DNA in the depot area (F_D), monoST on the cantilever (F_C) and DNA in the target area (F_T) are tuned such that $F_D < F_C < F_T$.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

small methods
www.small-methods.com

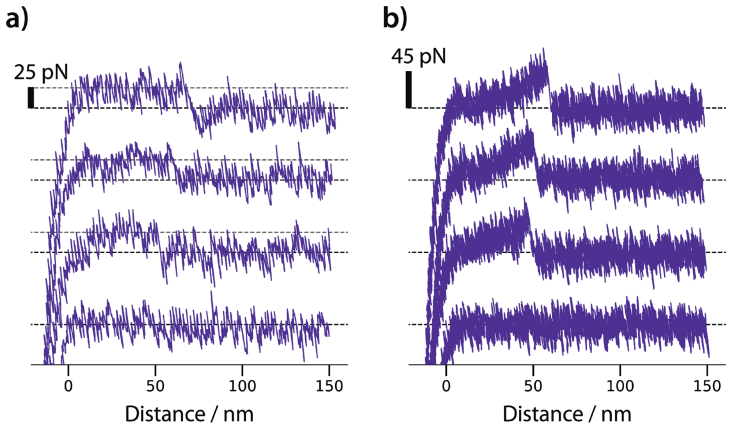


Figure 2. a,b) Typical force–distance curves of single-molecule depot pickup (a) and target deposition events (b). Depot pickup events occur when the transfer construct-surface DNA complex is unzipped, resulting in a plateau of relatively constant force at ≈25 pN (depicted by the dashed gray lines). Target deposition events occur when the newly established transfer construct–surface DNA complex in shear orientation remains intact, and instead the SII:monoST complex unbinds. This results in a gradual buildup of force and a sharp unbinding peak at ≈45 pN. In some cases, no molecules were picked up or deposited, which is reflected in the zero-force curves (bottom-most traces). The baseline for each curve of 0 pN force is depicted by the dashed black lines.

unzipping of DNA is described quantitatively by an equilibrium thermodynamic model,[20,21] leading to a plateau of constant force. During a pickup event from the depot area, the 40 bp DNA duplex melts at ≈25 pN, which is consistent with previous SMC&P experiments using this same DNA duplex.[10,11] The non-equilibrium unbinding of the SII:monoST interaction occurs at forces significantly greater than this value at the given loading rates.[14] Consequently, the transfer-depot DNA duplex melts while the SII:monoST bond stays intact, allowing for consistent pickup of molecules from the depot. The transfer construct is then transported to the target area where it binds to the surface-immobilized target DNA in shear orientation. The geometry of the 40 bp DNA duplex confers a substantial increase in rupture force due to force propagation through all basepairs (a most probable unbinding force of ≈65 pN at the observed loading rates around 300 pN s^{-1}).[22] Meanwhile, the SII:monoST complex ruptures in a non-equilibrium process at ≈45 pN at this pulling speed and apparent loading rate. Thus, the transfer construct detaches from the cantilever upon retraction and remains deposited in the target area. The cantilever-coupled monoST complex is now free to pick up a new transfer construct in repeated SMC&P cycles. This therefore presents the opportunity to carefully tune the expected rupture force of a protein-based handle by varying the loading rate, while simultaneously incurring minimal or no effect on the expected rupture force of a DNA-based tether. Hence, the SMC&P force hierarchy can be further reinforced by adjusting the pulling speed in each step of the cycle to maximize the difference in rupture force of the cantilever handle and surface tether. Examples of single-molecule pickup and deposition events demonstrate the plateau-like force curves observed from basepair-by-basepair unzipping of DNA in the depot area (Figure 2a) and the worm-like chain (WLC)[23] stretching behavior of the poly(ethylene glycol) (PEG) linkers followed by a sharp unbinding peak of the SII:monoST complex in the target area (Figure 2b). Notably, no additional force barriers are observed, which is consistent with the GFP and *Strep*-Tactin fold staying intact throughout the transfer process.

Following completion of the SMC&P sequence, the GFP-containing transfer molecules arranged in a dinosaur pattern were imaged by TIRFM (**Figure 3**), producing a clearly discernible outline. The previously demonstrated longevity of the monoST complex after hundreds of pulling events[14] is confirmed here with successful SMC&P transport over 395 consecutive cycles. Moreover, the rupture forces exhibited by the SII:monoST complex are in a range that is compatible with the well-characterized depot-transfer and target-transfer DNA duplex unbinding. Patchiness in the pattern can be partially attributed to the limited photostability of GFP, likely causing a fraction of the transfer molecules to photobleach during purification and experimental setup before imaging. There are also cases where a cycle fails to transport a transfer construct, as SMC&P and the underlying rupture forces are probability-dependent. Surface defects and densities can also influence the efficiency of SMC&P. However, the corresponding force–distance curves in every cycle can control for this; an inherently nonfluorescent or bleached transfer construct produces a deposition force curve but no fluorescence signal, and a failed transport cycle produces neither (e.g., the bottom-most force traces in Figure 2). Such analysis was previously executed by Pippig et al. to evaluate a widely spaced grid pattern of individual GFP molecules deposited by SMC&P.[11] The same strategy could conceivably be applied to any SMC&P experiment in which it is necessary to determine exactly which deposition points contain the transported molecules of interest.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

small methods
www.small-methods.com

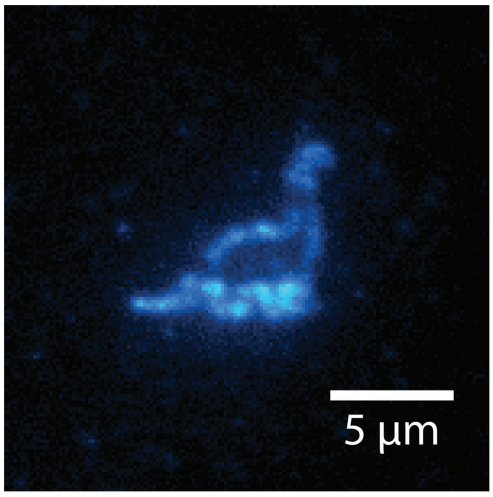


Figure 3. TIRFM image of GFP molecules after SMC&P in a dinosaur pattern. The image is composed of the average pixel intensity of 30 stacked frames from TIRFM acquisition (0.12 s exposure time at ≈10 W cm^{-2}) with a blue laser. The pattern consists of 395 deposited molecules spaced 150 nm apart.

SMC&P is an attractive strategy for the study of enzyme activity, as it enables precise placement of molecules of interest on a surface with known positioning, in contrast to stochastic surface immobilization. Moreover, the pattern of molecules to be arranged is completely arbitrary, allowing unrestricted pattern design for the investigation of the effects of relative geometry within networks or clusters of enzymes. Single-molecule enzyme analyses may prove to be indispensable for gaining new insights into the dynamic nature of enzyme networks, such as cellulosomes. These cellulose-degrading complexes are utilized by many species of cellulolytic bacteria, and have the critical characteristic that cellulase enzymes are strategically arranged on a scaffold to increase the efficiency of the catalytic network.[24,25] As the spatial organization of the enzymes is a key part of their function in vivo, similarly taking into account their relative geometry could provide new understanding of the cellulosomal components on a single-molecule level. Therefore, SMC&P may be an advantageous strategy for directed arrangement and investigation of this enzyme network.

Identification of diverse surface-immobilization methods is likewise a prerequisite for SMC&P's versatility. An intriguing potential use of the SII:monoST complex in SMC&P would leverage its terminus-dependent rupture force regimes to create SII-harboring depot and target regions. A monoST-containing transfer construct could be transported from the N-terminal SII depot (lower rupture forces) to the C-terminal SII target (higher rupture forces) via a cantilever tag of some intermediate rupture force. This presents an opportunity to forego DNA-based anchoring systems, which could be crucial for studying DNA-binding enzymes such as polymerases or ligases.

We have demonstrated that the SII:monoST complex is a viable handle for SMC&P. Although this study serves mainly as a proof of concept, it is invaluable to the expansion of SMC&P that diverse immobilization and cantilever handling options are available. Arrangement of molecules that perform biological functions—such as enzymes or aptamers—requires careful consideration of the unique properties, requirements or limitations of each molecule of interest. Unexpected secondary interactions between a tethering system and an enzyme could hinder SMC&P transport or interfere with enzymatic activity. In other words, it is possible that universal means of surface immobilization or cantilever handling may ultimately not exist. It is therefore advantageous to have a modular system with exchangeable components so that no molecule of interest must be excluded from study. With the addition of the SII:monoST handle system to the SMC&P toolbox, we have expanded the handling options available and given this technique a new degree of flexibility.

Experimental Section

Detailed information on all sections can be found in the Supporting Information.

Preparation of Monovalent Strep-Tactin: A heterotetrameric monovalent version of *Strep*-Tactin (monoST) was designed, expressed, purified, and reconstituted as previously described by Baumann et al.[14] In brief, the nonfunctional subunits and the single functional subunit harboring a 6×His tag and a reactive Cysteine residue were separately expressed in *Escherichia coli* (*E. coli*) BL21(DE3)-CodonPlus cells. Inclusion bodies were dissolved and denatured, and the dissolved inclusion body fractions of the nonfunctional and functional subunits were mixed in a ratio of 10:1, respectively. Subunits were refolded by slowly and dropwise adding to a reservoir of 1× phosphate-buffered saline (PBS). The assembled monoST was purified by Ni-IMAC affinity chromatography. The fractions containing monoST were isolated and dialyzed against 1× PBS. Purified monoST was long-term stored at 4 °C in the presence of tris(2-carboxyethyl)phosphine (TCEP) beads.

Preparation of Superfolder Green Fluorescent Protein (sfGFP) Construct and DNA Coupling: An sfGFP[9] transfer construct was designed, expressed, and purified as previously described by Baumann et al.[14] In brief, the construct harbors an N-terminal *Strep*-tag II (SAWSHPQFEK = SII)[13] and a C-terminal ybbR-tag (DSLEFIASKLA)[15,26] to enable specific cantilever handling and DNA coupling, respectively. The GFP gene was cloned into a modified pET28a vector that contains an N-terminal 6×His-tag followed by a PreScission Protease cleavage site (PreSc). The resulting fusion protein (6×His-PreSc-SII-sfGFP-ybbR) was expressed in *E. coli* BL21(DE3)-CodonPlus cells. The sfGFP construct was obtained in the soluble fraction after cell lysis and purified by Ni-IMAC affinity chromatography. Selected fractions of purified protein were then dialyzed overnight against storage buffer (50×10^{-3} M Tris HCl pH 7.5, 150×10^{-3} M NaCl, 2×10^{-3} M dithiothreitol (DTT), 5% glycerol) and stored long-term at −80 °C. The sfGFP construct was covalently coupled to DNA via the enzyme Sfp transferase as similarly described by Pippig et al.,[11] which is slightly altered from the protocol of Yin et al.[26] PreScission Protease, Sfp transferase, and CoA-modified transfer DNA (biomers.net GmbH, Ulm, Germany) were incubated with the purified 6×His-PreSc-SII-sfGFP-ybbR construct at room temperature for 2 h for simultaneous cleavage of the 6×His tag and covalent coupling of the ybbR tag to DNA. The reaction was filtered and then stored on ice until application in a microfluidic system.

Preparation of Cantilevers: MLCT cantilevers (Bruker, Camarillo, USA) were silanized in 3-(aminopropyl)dimethylethoxysilane and subsequently functionalized with a hetero-bifunctional PEG crosslinker[27,28] with *N*-hydroxy succinimide and maleimide groups (MW 5000). Cantilevers were covalently coupled to monoST.

Preparation of Glass Surfaces: Glass cover slips were silanized in (3-aminopropyl)dimethylethoxysilane and subsequently functionalized with a hetero-bifunctional PEG crosslinker with *N*-hydroxy succinimide and maleimide groups (MW 5000). Thiol-modified Depot and Target DNA was reduced and then purified by ethanol precipitation. A poly(dimethylsiloxane) (PDMS) microfluidic system—based on the system described by Kufer et al.[17]—was fixed on the PEGylated cover glass. Depot and Target channels were functionalized with their respective reduced DNA, and the sfGFP-DNA chimera construct was incubated in the Depot channel for 1 h. The Depot channel was then flushed with 1× PBS to remove unbound- or nonspecifically bound sfGFP. The microfluidic system was then removed and the surface submerged in 1× PBS.

AFM/TIRFM Measurements: SMC&P experiments were carried out on a combined AFM/TIRFM setup, as described previously.[16] The dinosaur pattern was written in 395 transfer cycles with 150 nm spacing between each deposition point. The pulling speed in the depot was set to 2 µm s^{-1} and in the target to 0.2 µm s^{-1}. Rupture forces and loading rates were evaluated from AFM force–distance curves that were recorded for each pickup and deposition process utilizing a quantum mechanically corrected WLC model.[29] Blue laser excitation at 488 nm with an estimated intensity of ≈10 W cm^{-2} was utilized to monitor the GFP fluorescence. Fluorescent images were evaluated and processed with the analysis software ImageJ.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by the European Research Council (Cellufuel, Advanced Grant No. 294438) and the German Research Foundation (SFB 1032-A01). The authors thank Ellis Durner and Markus A. Jobst for AFM/TIRFM hardware and software support; Angelika Kardinal and Thomas Nicolaus for laboratory support; and Christoph Hohmann of Nanosystems Initiative Munich (NIM) for graphic design assistance and production.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

AFM, monovalent *Strep*-Tactin, single-molecule cut-and-paste, single-molecule fluorescence, spatial arrangement

Received: April 27, 2017
Revised: May 13, 2017
Published online: July 18, 2017

[1] S. Hirschi, M. Stauffer, D. Harder, D. J. Müller, W. Meier, D. Fotiadis, *Chimia* **2016**, *70*, 398.
[2] A. Rajendran, E. Nakata, S. Nakano, T. Morii, *ChemBioChem* **2017**, *18*, 696.
[3] G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **1986**, *56*, 930.
[4] M. Radmacher, R. W. Tillmann, M. Fritz, H. E. Gaub, *Science* **1992**, *257*, 1900.
[5] M. Radmacher, M. Fritz, H. G. Hansma, P. K. Hansma, *Science* **1994**, *265*, 1577.
[6] D. T. Edwards, T. T. Perkins, *J. Struct. Biol.* **2017**, *197*, 13.
[7] V. S. Jadhav, D. Brüggeman, F. Wruck, M. Hegner, *Beilstein J. Nanotechnol.* **2016**, *7*, 138.
[8] M. Strackharn, S. W. Stahl, E. M. Puchner, H. E. Gaub, *Nano Lett.* **2012**, *12*, 2425.
[9] J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **2006**, *24*, 79.
[10] M. Strackharn, D. A. Pippig, P. Meyer, S. W. Stahl, H. E. Gaub, *J. Am. Chem. Soc.* **2012**, *134*, 15193.
[11] D. A. Pippig, F. Baumann, M. Strackharn, D. Aschenbrenner, H. E. Gaub, *ACS Nano* **2014**, *8*, 6551.
[12] S. F. Heucke, F. Baumann, G. P. Acuna, P. M. D. Severin, S. W. Stahl, M. Strackharn, I. H. Stein, P. Altpeter, P. Tinnefeld, H. E. Gaub, *Nano Lett.* **2013**, *14*, 391.
[13] S. Voss, A. Skerra, *Protein Eng.* **1997**, *10*, 975.
[14] F. Baumann, M. S. Bauer, L. F. Milles, A. Alexandrovich, H. E. Gaub, D. A. Pippig, *Nat. Nanotechnol.* **2016**, *11*, 89.
[15] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15815.
[16] H. Gumpp, S. W. Stahl, M. Strackharn, E. M. Puchner, H. E. Gaub, *Rev. Sci. Instrum.* **2009**, *80*, 063704.
[17] S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl, H. E. Gaub, *Science* **2008**, *319*, 594.
[18] J. Morfill, K. Blank, C. Zahnd, B. Luginbühl, F. Kühner, K. E. Gottschalk, A. Plückthun, H. E. Gaub, *Biophys. J.* **2007**, *93*, 3583.
[19] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, M. A. Nash, *J. Struct. Biol.* **2016**, *197*, 3.
[20] U. Bockelmann, B. Essevaz-Roulet, F. Heslot, *Phys. Rev. Lett.* **1997**, *79*, 4489.
[21] R. Krautbauer, M. Rief, H. E. Gaub, *Nano Lett.* **2003**, *3*, 493.
[22] J. Morfill, F. Kühner, K. Blank, R. A. Lugmaier, J. Sedlmair, H. E. Gaub, *Biophys. J.* **2007**, *93*, 2400.
[23] C. Bustamante, J. F. Marko, E. D. Siggia, S. Smith, *Science* **1994**, *265*, 1599.
[24] E. A. Bayer, J. P. Belaich, Y. Shoham, R. Lamed, *Annu. Rev. Microbiol.* **2004**, *58*, 521.
[25] S. Moraïs, J. Stern, A. Kahn, A. P. Galanopoulou, S. Yoav, M. Shamshoum, M. A. Smith, D. G. Hatzinikolaou, F. H. Arnold, E. A. Bayer, *Biotechnol. Biofuels* **2016**, *9*, 164.
[26] J. Yin, A. J. Lin, D. E. Golan, C. T. Walsh, *Nat. Protoc.* **2006**, *1*, 280.
[27] J. L. Zimmerman, T. Nicolaus, G. Neuert, K. Blank, *Nat. Protoc.* **2010**, *5*, 975.
[28] E. Celik, V. T. Moy, *J. Mol. Recogn.* **2012**, *25*, 53.
[29] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, R. R. Netz, *Phys. Rev. Lett.* **2005**, *94*, 048301.

small methods

Supporting Information

for *Small Methods*, DOI: 10.1002/smtd.201700169

Strep-Tag II and Monovalent *Strep*-Tactin as Novel Handles in Single-Molecule Cut-and-Paste

*Katherine R. Erlich, Fabian Baumann, Diana A. Pippig, and Hermann E. Gaub**

WILEY-VCH

Supporting Information

Strep-Tag II and Monovalent _Strep_-Tactin as Novel Handles in Single-Molecule Cut-and-Paste

*Katherine R. Erlich, Fabian Baumann, Diana A. Pippig, and Hermann E. Gaub**

K. R. Erlich, Dr. F. Baumann, Dr. D. A. Pippig, Prof. Dr. H. E. Gaub
Center for Nanoscience and Department of Physics of Ludwig Maximilians University of Munich, Amalienstraße 54, 80799 Munich, Germany
E-mail: gaub@lmu.de

The experiments described in the manuscript were performed on an AFM/TIRFM hybrid, the details of which may be found in Gumpp *et al.*[1] This supporting information specifies methods, materials and additional data that are relevant for the conduction of the measurements discussed in the main text.

AFM Measurements

A custom built AFM head and an Asylum Research MFP3D controller (Asylum Research, Santa Barbara, USA), which provides ACD and DAC channels as well as a DSP board for setting up feedback loops, were used. 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). MLCT cantilevers (Bruker, Camarillo, USA; 20 nm nominal tip radius, pyramidal-shaped probe) were chemically modified (see Preparation of Cantilevers) and calibrated in solution using the equipartition theorem.[2,3] Pulling velocities were set to 2 μm/s in the depot and 0.2 μm/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.

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-color 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, blue excitation at 488nm with an estimated intensity of approximately 10 W/cm^2 was utilized to monitor the GFP 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 multicolor 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.

Preparation of Monovalent *Strep*-Tactin

A heterotetrameric monovalent version of *Strep*-Tactin (monoST) was designed, expressed, purified and reconstituted as previously described by Baumann *et al.*[4] In brief, the non-functional subunits and the single functional subunit harboring a 6xHis tag and a reactive Cysteine residue were separately expressed in *E. coli* BL21(DE3)-CodonPlus cells (Agilent Technologies, Inc., Santa Clara, CA, USA). As individual monomers of *Strep*-Tactin are not readily soluble, both subunits formed inclusion bodies, which were dissolved and denatured in solubilization buffer (6 M Guanidinium HCl, 20 mM Tris HCl pH 7.5). The dissolved inclusion body fractions of the non-functional and functional subunits were mixed in a ratio of 10:1 respectively by mass (which is also approximately a 10:1 molar ratio, considering their comparable molecular mass). Refolding of the subunits was accomplished by slowly and drop-wise adding the denatured protein mixture to a

reservoir of 1x phosphate-buffered saline (PBS) from a 10x stock (Roche GmbH) and 10 mM β-Mercaptoethanol, followed by overnight stirring at 4 °C to increase refolding and assembly of the monomers into tetramers. The assembled monoST was subsequently filtered to remove aggregates, and loaded to a 5 mL HisTrap FF column (GE Healthcare) for Ni-IMAC purification. The bound molecules in the column were eluted over a linear gradient ranging from 10 to 300 mM Imidazole (in 1x PBS, 10 mM β-Mercaptoethanol). The fractions containing monoST were isolated and dialyzed against 1x PBS. Free reducing agent was omitted from the dialysis solution as it would later interfere with Mal-PEG immobilization. Therefore, bead-immobilized TCEP was added to the dialysis tubing. Following dialysis, monoST was long-term stored at 4 °C in the presence of TCEP beads. Immediately prior to cantilever immobilization, monoST was incubated with fresh TCEP beads to improve Cysteine reactivity (see Preparation of Cantilevers).

Preparation *Strep*-tag II Fused sfGFP Construct

A superfolder Green Fluorescent Protein (sfGFP)[5] transfer construct was designed, expressed and purified as previously described in Baumann *et al*.[4] In brief, the construct harbors an N-terminal *Strep*-tag II (SAWSHPQFEK = SII) and a C-terminal ybbR-tag (DSLEFIASKLA)[6,7] to enable specific cantilever handling and DNA coupling, respectively. The GFP gene was PCR amplified from a synthetic template (Lifetechnologies, Paisley, UK) with primers containing the respective tag coding sequences. The construct was cloned into a modified pET28a vector (GE Healthcare, Little Chalfont, UK), which contains an N-terminal 6xHis-tag followed by a PreScission Protease cleavage site (PreSc), by means of NdeI and XhoI restriction sites. The resulting fusion protein (6xHis-PreSc-SII-sfGFP-ybbR) was expressed in *E.coli* BL21(DE3)-CodonPlus cells. To this end, 1 L of SB medium was inoculated with 10 ml of an overnight culture and grown at 37 °C. When an OD_{600} of 0.7 had been reached, overnight expression at 18 °C was induced by adding 0.25 mM IPTG. Cells were lysed in 50 mM Tris HCl pH 7.5, 150 mM NaCl, 10mM imidazole, 5% Glycerol, by sonification. The SII-sfGFP-ybbR construct was obtained in the soluble fraction and purified by Ni-IMAC affinity chromatography on a 5mL HisTrap HP column (GE Healthcare, Little Chalfont, UK) via gradient elution from 10 mM to 250 mM imidazole. Selected fractions of purified protein were then dialyzed overnight against 50 mM Tris HCl pH 7.5, 150 mM NaCl, 2 mM DTT, 5% Glycerol, and finally stored at -80 °C at a final concentration of ~20 μM.

DNA-Coupling and PreScission Digestion of sfGFP Construct

The sfGFP construct was covalently coupled to DNA via the enzyme Sfp transferase as similarly described by Pippig *et al*.,[8] which is slightly altered from the protocol of Yin *et al.*[7] Importantly, PreScission protease was concurrently applied to remove the N-terminal 6xHis-tag to ensure SII-accessibility. 3'-Coenzyme A-modified transfer DNA (CoA-DNA) was synthesized by biomers.net GmbH (Ulm, Germany). Lyophilized DNA was dissolved in ddH_2O to a concentration of 100 mM and stored at -20°C. The combination coupling-digestion reaction consisted of 160 pmol SII-sfGFP-ybbR, 100 pmol CoA-DNA, 150 pmol Sfp transferase, and 60 pmol PreScission protease in a total volume of 30 μL in 1x Sfp buffer (120 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, 2% Glycerol, 2 mM DTT), and was incubated at room temperature for 2 hours. As the amount of CoA-DNA in solution was limiting, the reaction yielded a high percentage of a DNA-coupled SII-sfGFP-ybbR chimera construct, in addition to a lesser fraction of uncoupled SII-sfGFP-ybbR that cannot bind specifically to the Depot surface. The reaction was further diluted in 1x Sfp buffer to a total volume of 100 μL, and filtered with a 0.45 μm centrifugal filter unit to remove large aggregates that

could interfere with the microfluidic channels. The filtered reaction was then stored on ice until application in a microfluidic system (see Preparation of Glass Surfaces).

Preparation of Cantilevers

MLCT cantilevers (Bruker, Camarillo, USA) were oxidized in a UVOH 150 LAB UV-ozone cleaner (FHR Anlagenbau GmbH, Ottendorf-Okrilla, Germany) and silanized by incubation for 2 min in (3-Aminopropyl)dimethylethoxysilane (ABCR, Karlsruhe, Germany, 50% v/v in Ethanol). Cantilevers were washed sequentially in toluene, 2-propanol and ddH_2O and then baked at 80°C for 30 minutes. The silanized cantilevers were incubated in sodium borate buffer (150 mM, pH 8.5) for 30 minutes in order to deprotonate primary amine groups. Subsequently, the cantilevers were incubated for 30 minutes in a solution of a heterobifunctional PEG crosslinker[9,10] with N-hydroxy succinimide and maleimide groups (MW 5000, Rapp Polymere, Tübingen, Germany) dissolved to 30mM in sodium borate buffer. Unbound PEG was washed from the cantilevers in filtered H_2O. Prior to covalent binding to the cantilevers, monoST was incubated with fresh TCEP beads for 30 minutes, and then filtered with a 0.45 μm centrifugal filter unit to remove the beads. Cantilevers were then incubated in freshly-reduced monoST at room temperature for 1 h. Finally, cantilevers were washed and stored in 1x PBS.

Preparation of Glass Surfaces

Glass cover slips were sonicated in 50% (v/v) 2-propanol in filtered H_2O for 15 min and oxidized 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 silanized by incubating for 1 h in (3-Aminopropyl)dimethylethoxysilane (ABCR, Karlsruhe, Germany, 1.8% v/v in Ethanol). The silanized surfaces were incubated in sodium borate buffer (150 mM, pH 8.5) for 30 minutes in order to deprotonate primary amine groups. Subsequently, the surfaces were incubated for 30 minutes in a solution of a heterobifunctional PEG crosslinker[9,10] with N-hydroxy succinimide and maleimide groups (MW 5000, Rapp Polymere, Tübingen, Germany) dissolved to 30 mM in sodium borate buffer. Unbound PEG was washed from the surfaces in filtered H_2O. Thiol-modified Depot and Target DNA was reduced with 1mM TCEP and then purified by ethanol precipitation. DNA pellets were dissolved in phosphate buffer (pH 7.2, 50 mM NaCl, 10 mM EDTA). A PDMS microfluidic system – based on the system described by Kufer *et al.*[11] – was fixed on the PEGylated cover glass, and the freshly-reduced Depot and Target DNA were pumped through the two respective channels and incubated for 1 h. Both channels were then flushed with filtered H_2O to remove unbound DNA. Importantly, the Target channel was then disconnected from the pump and left with filtered H_2O in the channel to prevent drying and discourage nonspecific surface adsorption. The Depot channel was then flushed with 1 mg/mL filtered BSA and then 0.05% TWEEN 20. The sfGFP-DNA chimera construct was pumped into the Depot channel and incubated for 1 h. The Depot channel was then flushed again with 1x PBS to remove unbound- or nonspecifically-bound sfGFP. The microfluidic system was then removed and the surface submerged in 1x PBS.

SMC&P Experiment

The dinosaur pattern was written in 395 transfer cycles with 150 nm spacing between each deposition point. The pulling speed in the depot was set to 2 μm/s and in the target to 0.2 μm/s. This corresponds to approximate surface contact times[12] (dependent on approach/retraction velocity,

indentation force and substrate stiffness) of 8 ms and 80 ms, respectively, and should allow for ligand binding (compare k_{on}(DNA) > 10^4 $M^{-1}s^{-1}$ and k_{on}(*Strep*-tactin and SII) > 10^4 $M^{-1}s^{-1}$).[12–15] Considering a single monoST molecule being bound to the cantilever tip and estimating its localization in a half sphere with r = 30 nm (length of PEG linker), the local concentration of antibody would be in the mM range. This is several orders of magnitude higher than the measured K_d for the SII:monoST interaction of 2.3 μM.[4] 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.

Rupture forces and loading rates were evaluated from AFM force distance curves that were recorded for each pickup and deposition process utilizing a quantum mechanically corrected WLC model[16] (force spectroscopy data was evaluated in Python 2.7, Python Software Foundation).

DNA Oligomer Sequences

thiolated depot oligomer

5’ SH - TTT TTT CAT GCA AGT AGC TAT TCG AAC TAT AGC TTA AGG ACG TCA A 3’

thiolated target oligomer

5’ CAT GCA AGT AGC TAT TCG AAC TAT AGC TTA AGG ACG TCA ATT TTT T - SH 3’

CoA-modified transfer oligomer for protein coupling

5’ TTG ACG TCC TTA AGC TAT AGT TCG AAT AGC TAC TTG CAT GTT TTT TTT TTT - CoA 3’

References

[1] H, Gumpp, S. W. Stahl, M. Strackharn, E. M. Puchner, H. E. Gaub. Ultrastable Combined Atomic Force and Total Internal Reflection Fluorescence Microscope [Corrected]. *Rev. Sci. Instrum.* **2009**, *80*, 063704.

[2] E. Florin. Sensing Specific Molecular Interactions with the Atomic Force Microscope. *Biosens. Bioelectron.* **1995**, *10*, 895.

[3] H. J. Butt, M. Jaschke. Calculation of Thermal Noise in Atomic-Force Microscopy. *Nanotechnology* **1995**, *6*, 1.

[4] F. Baumann, M. S. Bauer, L. F. Milles, A. Alexandrovich, H. E. Gaub, D. A. Pippig. Monovalent *Strep*-Tactin for strong and site-specific tethering in nanospectroscopy. *Nature Nanotech.* **2016**, *11*, 89.

[5] J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79.

[6] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15815.

[7] J. Yin, A. J. Lin, D. E. Golan, C. T. Walsh. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **2006**, *1*, 280.

[8] D. A. Pippig, F. Baumann, M. Strackharn, D. Aschenbrenner, H. E. Gaub. Protein-DNA chimeras for nano assembly. *ACS nano* **2014**, *8*, 6551.

[9] J. L. Zimmerman, T. Nicolaus, G. Neuert, K. Blank. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nature protocols* **2010**, *5*, 975.

[10] E. Celik, V. T. Moy. Nonspecific interactions in AFM force spectroscopy measurements. *Journal of molecular recognition : JMR* **2012**, *25*, 53.

[11] S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl, H. E. Gaub. Single-Molecule Cut-and-Paste Surface Assembly. *Science* **2008**, *319*, 594.

[12] S. Guo, N. Lad, C. Ray, B. B. Akhremitchev. Association Kinetics from Single Molecule Force Spectroscopy Measurements. *Biophys. J.* **2009**, *96*, 3412.

[13] K. Tawa, D. Yao, W. Knoll. Matching Base-Pair Number Dependence of the Kinetics of DNA-DNA Hybridization Studied by Surface Plasmon Fluorescence Spectroscopy. *Biosens. Bioelectron.* **2005**, *21*, 322.

[14] M. R. Henry, P. Wilkins Stevens, J. Sun, D. M. Kelso. Real-Time Measurements of DNA Hybridization on Microparticles with Fluorescence Resonance Energy Transfer. *Anal. Biochem.* **1999**, *276*, 204.

[15] M. H. van Es, J. Tang, J. Preiner, P. Hinterdorfer, T. H. Oosterkamp. Single molecule binding dynamics measured with atomic force microscopy. *Ultramicroscopy* **2014**, *140*, 32.

[16] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, R. R. Netz. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments *Versus* Ab-Initio Theory. *Phys. Rev. Lett.* **2005**, *94*, 048301.

Appendix B
Buffers and DNA Sequences

B.1 Buffers for Purification and Reactions

All buffers were prepared at room temperature. Following complete dissolution of solid components and pH adjustment, all buffers were filter-sterilized via vacuum filtration through micropore filter paper or via syringe filtration through a 0.22 µm sterile filter. Most buffers were stored at 4 °C, with the exception of ATP- and NADH-containing buffers stored at -20 °C.

B.1.1 Protein Purification Buffers

- **His Lysis and Wash Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  10 mM imidazole
  pH = 7.5 at 22 °C

- **His Elution Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  250 mM imidazole
  10 % (v/v) glycerol
  pH = 7.5 at 22 °C

- **T7 RNAP His Lysis Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  10 mM imidazole
  4 µg/mL leupeptin

40 µg/mL PMSF
100 µg/mL lysozyme
5 mM β-mercaptoethanol, freshly added
pH = 7.8 at 22 °C

0.08 % (w/v) deoxycholic acid added after initial incubation in lysis buffer

- **T7 RNAP His Wash Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  10 mM imidazole
  5 mM β-mercaptoethanol, freshly added
  pH = 7.8 at 22 °C

- **T7 RNAP His Elution Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  250 mM imidazole
  5 mM β-mercaptoethanol, freshly added
  pH = 7.8 at 22 °C

- **GST Lysis and Wash Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  1 mM DTT
  5 % v/v glycerol
  pH = 7.5 at 22 °C

- **GST Elution Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  1 mM DTT
  25 mM glutathione, freshly added
  5 % v/v glycerol
  pH = 7.5 at 22 °C

- **GST Regeneration Buffer**
  100 mM Tris-HCl
  5 M guanidinium HCl
  100 mM EDTA
  pH = 8.0 at 22 °C

- **Ion Exchange Low-Salt Buffer**
  30 mM Tris-HCl

150 mM NaCl
1 mM DTT
5 % v/v glycerol
pH = 7.5 at 22 °C

- **Ion Exchange High-Salt Buffer**
  30 mM Tris-HCl
  1 M NaCl
  1 mM DTT
  5 % v/v glycerol
  pH = 7.5 at 22 °C

- **Size-Exclusion Chromatography Buffer**
  50 mM HEPES
  200 mM NaCl
  10 % v/v glycerol
  pH = 7.5 at 22 °C

- **Dialysis Buffer**
  30 mM Tris-HCl
  150 mM NaCl
  10 % (v/v) glycerol
  (+ 1 mM DTT for select proteins)
  pH = 7.5 at 22 °C

- **T4 DNAL Dialysis Buffer**
  30 mM Tris-HCl
  50 mM KCl
  1 mM DTT
  0.1 mM EDTA
  20 % (v/v) glycerol
  pH = 7.5 at 22 °C

- ***Pfu* DNAP Dialysis Buffer**
  30 mM Tris-HCl
  100 mM KCl
  1 mM DTT
  0.1 mM EDTA
  10 % (v/v) glycerol
  pH = 8.2 at 22 °C

B.1.2 Reaction Buffers

- **Coupling Buffer**
  50 mM $NaPO_4$
  50 mM NaCl
  10 mM EDTA
  pH = 7.2 at 4 °C

- **2x Maleimide Reaction Buffer**
  60 mM Tris-HCl
  300 mM NaCl
  pH = 7.5 at 22 °C

- **10x Sfp Reaction Buffer**
  1.2 M Tris-HCl
  100 mM $MgCl_2$
  1.5 M NaCl
  20 mM DTT
  20 % (v/v) glycerol
  pH = 7.5 at 22 °C

- **10x *Pfu* DNAP Reaction Buffer**
  200 mM Tris-HCl
  100 mM $(NH_4)_2SO_4$
  100 mM KCl
  20 mM $MgSO_4$
  1 mg/mL BSA
  1 % (v/v) Triton X-100
  pH = 8.8 at 22 °C

- **10x *Pfu* DNAP Low-Temperature Reaction Buffer**
  200 mM Tris-HCl
  100 mM $(NH_4)_2SO_4$
  100 mM KCl
  20 mM $MgSO_4$
  pH = 7.7 at 22 °C

- **5x T7 RNAP Reaction Buffer**
  200 mM Tris-HCl
  30 mM $MgCl_2$
  50 mM NaCl
  10 mM Spermidine
  pH = 7.9 at 22 °C

- **10x *E. coli* DNAL Reaction Buffer**
  300 mM Tris-HCl
  40 mM $MgCl_2$
  260 µM NADH
  10 mM DTT
  0.5 mg/mL BSA
  pH = 8.0 at 22 °C

- **2x T7 DNAL Reaction Buffer**
  132 mM Tris-HCl
  20 mM $MgCl_2$
  2 mM ATP
  2 mM DTT
  15 % w/v PEG 3000
  pH = 7.6 at 22 °C

- **10x NanoLuc Reaction Buffer**
  1 M Tris-HCl
  1.5 M KCl
  pH = 8.0 at 22 °C

B.2 DNA Sequences and Synthetic Oligomers

Synthetic DNA oligomers were typically purified by HPLC and delivered as lyophilized pellets. The DNA was dissolved in ddH_2O and stored at 4 °C, unless stated otherwise. Covalent modifications include coenzyme A (CoA), 5' phosphate (P), 5' or 3' thiol (HS), and terminal or internal fluorophores (Cy3, Cy5, Atto532, Atto550, Atto647N).

B.2.1 *Pfu* DNA Polymerase Low-Temperature Template

As AFM experiments - and specifically SMC&P experiments - are carried out at room temperature, it is not possible to rely on thermal melting of DNA strands for DNAP reactions as in PCR. Simultaneously, *Pfu* DNAP requires a double-stranded primer region to initiate DNA replication. Therefore, a DNA duplex with a double-stranded primer region and a single-stranded replication region was designed. The template and primer strands were pre-annealed prior to implementation in DNA polymerase reactions by mixing equimolar amounts and incubating at 80 °C followed by slow cooling to 22 °C.

DNAP Template (biomers GmbH, Ulm, Germany)
5' GAG CTA AAC GCT CAC CGT AAT GGT CAG CCA GAG TGT TGA ACG GGA TAG CAG CCA GTT CGA TGT CAG AGA AGT CGT TCT TAG CGA TGT TAA TCG TGT TCA TAG CCG AGA ATT CCC CTG GAC CCT GAA ACA GCA C 3'

DNAP Primer (metabion GmbH, Planegg, Germany)
5’ GTG CTG TTT CAG GGT C 3’

B.2.2 Malachite Green Aptamer Sequences

MG Aptamer DNA Template Forward Sequence
5’ GGA TCC CGA CTG GCG AGA GCC AGG TAA CGA ATG GAT CC 3’

MG-4 RNA Aptamer (IBA GmbH, Göttingen, Germany)
5’ rGrGrA rUrCrC rCrGrA rCrUrG rGrCrG rArGrA rGrCrC rArGrG rUrArA rCrGrA rArUrG rGrArU rCrC 3’

MG Rolling Circle Template (metabion GmbH, Planegg, Germany)
5’ P-GGC CAC AGG ATC CAT TCG TTA CCT GGC TCT CGC CAG TCG GGA TCC ACG TAC C 3’

MG Rolling Circle Splint (metabion GmbH, Planegg, Germany)
5’ CCT GTG GCC GGT ACG TGG 3’

B.2.3 T7 RNA Polymerase Promoter DNA

Double-stranded promoters were pre-annealed by mixing equimolar amounts of complementary strands and incubating at 80 °C followed by slow cooling to 22 °C. The hairpin promoter sequence was self-annealed by incubating at 80 °C followed by rapid cooling on ice to 4 °C.

T7 RNAP Forward Promoter Full (Eurofins Genomics GmbH, Ebersberg, Germany)
5’ GAG ACC ACA CGG CTA ATA CGA CTC ACT ATA GGA AGC G 3’

T7 RNAP Forward Promoter Short (Eurofins Genomics GmbH, Ebersberg, Germany)
5’ GAG ACC ACA CGG CTA ATA CGA CTC ACT ATA 3’

5’ HS T7 RNAP Reverse Promoter (biomers GmbH, Ulm, Germany)
5’ HS-AAA AAA CGC TTC CTA TAG TGA GTC GTA TTA GCC GTG TGG TCT C 3’

3’ HS T7 RNAP Reverse Promoter (biomers GmbH, Ulm, Germany)
5’ CGC TTC CTA TAG TGA GTC GTA TTA GCC GTG TGG TCT CAA AAA A-HS 3’

Hairpin T7 RNAP Promoter (biomers GmbH, Ulm, Germany)
5’ HS-AAA AAA CTT CCT ATA GTG AGT CGT ATT AGC CGT GTG GTC TCA AAA AAG AGA CCA CAC GGC TAA TAC GAC TCA CTA TAG GAA GTT TTT T 3’

B.2.4 SMC&P DNA

Synthetic DNA oligomers for SMC&P were purchased from biomers.net (biomers GmbH, Ulm, Germany).

AFM DNA
5' HS-TTT TTT TTT TCT GCA GGA ATT CGA TAT CAA 3'

Depot DNA
5' HS-TTT TTT TTT TTA GCT ATT CGA ACT ATA GCT TAA GGA CGT C 3'

Target DNA
5' TAG CTA TTC GAA CTA TAG CTT AAG GAC GTC TTT TTT TTT T-HS 3'

Atto647N Transfer DNA
5' Atto647N-TTT TTG ACG TCC TTA AGC TAT AGT TCG AAT AGC TAT TTT TTT TCA TCG ATA AGC TTG ATA TCG AAT TCC TGC AGT TTT T 3'

CoA Protein Transfer DNA, stored at -20 °C
5' TTG ACG TCC TTA AGC TAT AGT TCG AAT AGC TAC TTG CAT GTT TTT TTT TTT TTT-CoA 3'

CoA-Cy5 Protein Transfer DNA, stored at -20 °C
5' Cy5-TTG ACG TCC TTA AGC TAT AGT TCG AAT AGC TAC TTG CAT GTT TTT TTT TTT TTT-CoA 3'

B.2.5 DNA Ligase Substrates

Synthetic DNA oligomers for DNA ligation reactions and FRET studies were purchased from biomers.net (biomers GmbH, Ulm, Germany), unless stated otherwise. Duplexes of complementary strands were pre-annealed by mixing equimolar amounts and incubating at 80 °C followed by slow cooling to 22 °C.

Modifications include 5' phosphate (P), terminal thiol (HS), and internal fluorescently labeled thymine bases (dT-Cy3, dT-Cy5, dT-Atto532, dT-Atto550, dT-Atto647N). In the case of fluorescently-labeled DNA molecules, the number of bases from the nick site to the fluorophore is additionally listed.

Duplex 1A - 14 bp
5' GTT CGA CCT AAT GGT AGC TGC CTC AAC GAC T(dT-Cy3)C GTC ATC CAC TCG 3'

Duplex 1A - no Cy3 (metabion GmbH, Planegg, Germany)
5' GTT CGA CCT AAT GGT AGC TGC CTC AAC GAC TTC GTC ATC CAC TCG 3'

Duplex 1A 2 - 3 bp
5' GTT CGA CCT AAT GGT AGC TGC CTC AAC GAC TTC GTC ATC CAC (dT-Cy3)CG 3'

Duplex 1B
5’ P-AGT ACG AGT GGA TGA ACGA AGT CGT TGA GGC AGC TAC CAT TAG GTC GAA C 3’

Duplex 2A
5’ TGA AGC AGT TGC AGA TGG AGT GCG G 3’

Duplex 2B - 13 bp
5’ P-TAC TCC GCA CTC CAT C(dT-Cy5)G CAA CTG CTT CA 3’

Duplex 3A - 7 bp
5’ TGA AGC AGT TGC AGA TGG (dT-Cy3)GA GCG G 3’

Duplex 3B
5’ P-TAC TCC GCT CAC CAT CTG CAA CTG CTT CA 3’

Duplex 4A - 7 bp
5’ GTT CGA CCT AAT GGT AGC TGC CTC AAC GAC TTC GTG AC(dT-Cy5) AAC TCG 3’

Duplex 4B
5’ P-AGT ACG AGT TAG TCA CGA AGT CGT TGA GGC AGC TAC CAT TAG GTC GAA C 3’

Duplex 5A - 5 bp
5’ TGA AGC AGT TGC AGA TGG AG(dT-Cy3) GCG G 3’

Duplex 5B
5’ P-TAC TCC GCA CTC CAT CTG CAA CTG CTT CA 3’

Duplex 5B HS
5’ P-TAC TCC GCA CTC CAT CTG CAA CTG CTT CA-HS 3’

Duplex 5C - 5 bp
5’ TGA AGC AGT TGC AGA TGG AG(dT-Atto532) GCG G 3’

Duplex 6A - 5 bp
5’ GTT CGA CCT AAT GGT AGC TGC CTC AAC GAC TTC GTA AGA C(dT-Cy5)C TCG 3’

Duplex 6B
5’ P-AGT ACG AGA GTC TTA CGA AGT CGT TGA GGC AGC TAC CAT TAG GTC GAA C 3’

Duplex 6C - 5 bp
5’ GTT CGA CCT AAT GGT AGC TGC CTC AAC GAC TTC GTA AGA C(dT-Atto647N)C TCG 3’

FRET Long Transfer
5’ GAC GTC CTT AAG CTA TAG TTC GAA TAG C(dT-Atto550)A C(dT-Atto647N)T GCA TG 3’

Bibliography

[1] Strackharn, M., Stahl, S. W., Puchner, E. M., and Gaub, H. E. Functional Assembly of Aptamer Binding Sites by Single-Molecule Cut-and-Paste. *Nano Letters*, 12(5):2425–2428, 2012.

[2] Heucke, S. F., Baumann, F., Acuna, G. P., Severin, P. M. D., Stahl, S. W., Strackharn, M., Stein, I. H., Altpeter, P., Tinnefeld, P., and Gaub, H. E. Placing Individual Molecules in the Center of Nanoapertures. *Nano Letters*, 14(2):391–395, 2014.

[3] Amyes, T. L. and Richard, J. P. Specificity in Transition State Binding: The Pauling Model Revisited. *Biochemistry*, 52(12):2021–2035, 2013.

[4] Radzicka, A. and Wolfenden, R. A proficient enzyme. *Science*, 267(5194):90–93, 1995.

[5] Knowles, J. R. Enzyme-catalyzed phosphoryl transfer. *Annu. Rev. Biochem.*, 49:877–919, 1980.

[6] Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096): 223–230, 1973.

[7] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. *Molecular Biology of the Cell, 4th edition.* Garland Science, New York, NY, 2002.

[8] Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proceedings of the National Academy of Sciences*, 47(9):1309–1314, 1961.

[9] Levinthal, C. Are there pathways for protein folding? *Journal de Chimie Physique*, 65: 44–45, 1968.

[10] Levinthal, C. How to Fold Graciously. In *Mössbauer Spectroscopy in Biological Systems Proceedings*, volume 24, pages 22–24, 1969. doi: citeulike-article-id:380320. URL `http://www.cc.gatech.edu/{~}turk/bio{_}sim/articles/proteins{_}levinthal{_}1969.pdf`.

[11] Honig, B. Protein folding: From the levinthal paradox to structure prediction. *Journal of Molecular Biology*, 293(2):283–293, 1999.

[12] Bai, Y., Sosnick, T., Mayne, L., and Englander, S. Protein folding intermediates: native-state hydrogen exchange. *Science*, 269(5221):192–197, 1995.

[13] Englander, S. W. and Mayne, L. The nature of protein folding pathways. *Proceedings of the National Academy of Sciences*, 111(45):15873–15880, 2014.

[14] Englander, S. W. and Mayne, L. The case for defined protein folding pathways. *Proceedings of the National Academy of Sciences*, 114(31):8253–8258, 2017.

[15] Frauenfelder, H., Parak, F., and Young, R. D. Conformational Substates in Proteins. *Annual Review of Biophysics and Biophysical Chemistry*, 17(1):451–479, 1988.

[16] Leopold, P. E., Montal, M., and Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences*, 89(18):8721–8725, 1992.

[17] Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, 1997.

[18] Onuchic, J. N. and Wolynes, P. G. Theory of protein folding. *Current Opinion in Structural Biology*, 14(1):70–75, 2004.

[19] Dill, K. A. and MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science*, 338 (6110):1042–1046, 2012.

[20] Ellis, R. J. Molecular chaperones: assisting assembly in addition to folding. *Trends in Biochemical Sciences*, 31(7):395–401, 2006.

[21] Mamipour, M., Yousefi, M., and Hasanzadeh, M. An overview on molecular chaperones enhancing solubility of expressed recombinant proteins with correct folding. *International Journal of Biological Macromolecules*, 102:367–375, 2017.

[22] Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31(1):147–157, 1982.

[23] Subramanya, H. S., Doherty, A. J., Ashford, S. R., and Wigley, D. B. Crystal Structure of an ATP-Dependent DNA Ligase from Bacteriophage T7. *Cell*, 85(4):607–615, 1996.

[24] Cheetham, G. M. T., Jeruzalmi, D., and Steitz, T. a. Structural basis for initiation of transcription from an RNA polymerase–promoter complex. *Nature*, 399(6731):80–83, 1999.

[25] Odell, M., Sriskanda, V., Shuman, S., and Nikolov, D. B. Crystal Structure of Eukaryotic DNA Ligase–Adenylate Illuminates the Mechanism of Nick Sensing and Strand Joining. *Molecular cell*, 6(5):1183–1193, 2000.

[26] Doherty, A. J. and Suh, S. W. Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Research*, 28(21):4051–4058, 2000.

[27] Kennedy, W. P., Momand, J. R., and Yin, Y. W. Mechanism for De Novo RNA Synthesis and Initiating Nucleotide Specificity by T7 RNA Polymerase. *Journal of Molecular Biology*, 370(2):256–268, 2007.

[28] Wynne, S. A., Pinheiro, V. B., Holliger, P., and Leslie, A. G. W. Structures of an Apo and a Binary Complex of an Evolved Archeal B Family DNA Polymerase Capable of Synthesising Highly Cy-Dye Labelled DNA. *PLoS ONE*, 8(8):e70892, 2013.

[29] Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS Journal*, 280(22):5705–5736, 2013.

[30] Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St.Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science*, 329(5989):309–313, 2010.

[31] Zanghellini, A. de novo computational enzyme design. *Current Opinion in Biotechnology*, 29(1):132–138, 2014.

[32] Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. How Fast-Folding Proteins Fold. *Science*, 334(6055):517–520, 2011.

[33] Perez, A., Morrone, J. A., Brini, E., MacCallum, J. L., and Dill, K. A. Blind protein structure prediction using accelerated free-energy simulations. *Science Advances*, 2(11): e1601274–e1601274, 2016.

[34] Schoeler, C., Malinowska, K. H., Bernardi, R. C., Milles, L. F., Jobst, M. A., Durner, E., Ott, W., Fried, D. B., Bayer, E. A., Schulten, K., Gaub, H. E., and Nash, M. A. Ultrastable cellulosome-adhesion complex tightens under load. *Nature Communications*, 5(1):5635, 2014.

[35] Schoeler, C., Bernardi, R. C., Malinowska, K. H., Durner, E., Ott, W., Bayer, E. A., Schulten, K., Nash, M. A., and Gaub, H. E. Mapping Mechanical Force Propagation through Biomolecular Complexes. *Nano Letters*, 15(11):7370–7376, 2015.

[36] Milles, L. F., Schulten, K., Gaub, H. E., and Bernardi, R. C. Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science*, 359(6383):1527–1533, 2018.

[37] Lu, H. P., Xun, L., and Xie, X. S. Single-Molecule Enzymatic Dynamics. *Science*, 282 (5395):1877–1882, 1998.

[38] English, B. P., Min, W., van Oijen, A. M., Lee, K. T., Luo, G., Sun, H., Cherayil, B. J., Kou, S. C., and Xie, X. S. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nature Chemical Biology*, 2(2):87–94, 2006.

[39] Binnig, G. and Quate, C. F. Atomic Force Microscope. *Physical Review Letters*, 56(9): 930–933, 1986.

[40] Florin, E.-L., Rief, M., Lehmann, H., Ludwig, M., Dornmair, C., Moy, V., and Gaub, H. Sensing specific molecular interactions with the atomic force microscope. *Biosensors and Bioelectronics*, 10(9-10):895–901, 1995.

[41] Lee, G. U., Kidwell, D. A., and Colton, R. J. Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. *Langmuir*, 10(2):354–357, 1994.

[42] Lee, G., Chrisey, L., and Colton, R. Direct measurement of the forces between complementary strands of DNA. *Science*, 266(5186):771–773, 1994.

[43] Smith, S., Finzi, L., and Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 1992.

[44] Block, S. M., Goldstein, L. S. B., and Schnapp, B. J. Bead movement by single kinesin molecules studied with optical tweezers. *Nature*, 348(6299):348–352, 1990.

[45] Svoboda, K., Schmidt, C. F., Schnapp, B. J., and Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365(6448):721–727, 1993.

[46] Müller, D. J. and Dufrêne, Y. F. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. *Nature Nanotechnology*, 3(5):261–269, 2008.

[47] Ott, W., Jobst, M. A., Schoeler, C., Gaub, H. E., and Nash, M. A. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. *Journal of Structural Biology*, 197(1):3–12, 2017.

[48] He, C., Hu, C., Hu, X., Hu, X., Xiao, A., Perkins, T. T., and Li, H. Direct Observation of the Reversible Two-State Unfolding and Refolding of an α/β Protein by Single-Molecule Atomic Force Microscopy. *Angewandte Chemie International Edition*, 54(34):9921–9925, 2015.

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

[50] Verdorfer, T., Bernardi, R. C., Meinhold, A., Ott, W., Luthey-Schulten, Z., Nash, M. A., and Gaub, H. E. Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. *Journal of the American Chemical Society*, 139(49):17841–17852, 2017.

[51] Verdorfer, T. and Gaub, H. E. Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy. *Scientific Reports*, 8(1):9634, 2018.

[52] Krautbauer, R., Rief, M., and Gaub, H. E. Unzipping DNA Oligomers. *Nano Letters*, 3(4): 493–496, 2003.

[53] Albrecht, C., Blank, K., Lalic-Mülthaler, M., Hirler, S., Mai, T., Gilbert, I., Schiffmann, S., Bayer, T., Clausen-Schaumann, H., and Gaub, H. E. DNA: a programmable force sensor. *Science*, 301(5631):367–370, 2003.

[54] Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–1112, 1997.

[55] Fowler, S. B., Best, R. B., Toca Herrera, J. L., Rutherford, T. J., Steward, A., Paci, E., Karplus, M., and Clarke, J. Mechanical Unfolding of a Titin Ig Domain: Structure of Unfolding Intermediate Revealed by Combining AFM, Molecular Dynamics Simulations, NMR and Protein Engineering. *Journal of Molecular Biology*, 322(4):841–849, 2002.

[56] Oberhauser, A. F., Marszalek, P. E., Erickson, H. P., and Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, 393(6681):181–185, 1998.

[57] Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., and Fernandez, J. M. Mechanical and chemical unfolding of a single protein: a comparison. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7):3694–3699, 1999.

[58] Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H. E., and Müller, D. J. Unfolding pathways of individual bacteriorhodopsins. *Science*, 288(5463):143–146, 2000.

[59] Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F., and Fernandez, J. M. The mechanical stability of ubiquitin is linkage dependent. *Nature Structural & Molecular Biology*, 10(9):738–743, 2003.

[60] Dietz, H., Berkemeier, F., Bertz, M., and Rief, M. Anisotropic deformation response of single protein molecules. *Proceedings of the National Academy of Sciences*, 103(34): 12724–12728, 2006.

[61] Morfill, J., Blank, K., Zahnd, C., Luginbühl, B., Kühner, F., Gottschalk, K.-E., Plückthun, A., and Gaub, H. E. Affinity-Matured Recombinant Antibody Fragments Analyzed by Single-Molecule Force Spectroscopy. *Biophysical Journal*, 93(10):3583–3590, 2007.

[62] Kulik, A. J., Lekka, M., Lee, K., Pyka-Fościak, G., and Nowak, W. Probing fibronectin–antibody interactions using AFM force spectroscopy and lateral force microscopy. *Beilstein Journal of Nanotechnology*, 6(1):1164–1175, 2015.

[63] Florin, E., Moy, V., and Gaub, H. Adhesion forces between individual ligand-receptor pairs. *Science*, 264(5157):415–417, 1994.

[64] Lee, G. U., Kidwell, D. A., and Colton, R. J. Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. *Langmuir*, 10(2):354–357, 1994.

[65] Moy, V., Florin, E., and Gaub, H. Intermolecular forces and energies between ligands and receptors. *Science*, 266(5183):257–259, 1994.

[66] Gunnoo, M., Cazade, P.-A., Galera-Prat, A., Nash, M. A., Czjzek, M., Cieplak, M., Alvarez, B., Aguilar, M., Karpol, A., Gaub, H., Carrión-Vázquez, M., Bayer, E. A., and Thompson, D. Nanoscale Engineering of Designer Cellulosomes. *Advanced Materials*, 28(27):5619–5647, 2016.

[67] Ott, W., Jobst, M. A., Bauer, M. S., Durner, E., Milles, L. F., Nash, M. A., and Gaub, H. E. Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano*, 11 (6):6346–6354, 2017.

[68] Jobst, M. A., Milles, L. F., Schoeler, C., Ott, W., Fried, D. B., Bayer, E. A., Gaub, H. E., and Nash, M. A. Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *eLife*, 4(OCTOBER2015):1–19, 2015.

[69] Baumann, F., Bauer, M. S., Milles, L. F., Alexandrovich, A., Gaub, H. E., and Pippig, D. A. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nature Nanotechnology*, 11(1):89–94, 2016.

[70] Bauer, M. S., Milles, L. F., Sedlak, S. M., and Gaub, H. E. Monomeric streptavidin : a versatile regenerative handle for force spectroscopy. *bioRxiv preprint*, pages 1–16, 2018.

[71] Schöler, C. *Molecular Mechanisms for Mechanostability in Biomolecular Complexes*. Dissertation, Ludwig-Maximilians-Universität München, 2017.

[72] Baumann, F. *Atomic Force Microscopy for High-Specificity Manipulations of Proteins and High-Throughput Analysis in Nanoapertures*. PhD thesis, Ludwig-Maximilians-Universität München, 2017.

[73] Ott, W. B. *Single Molecule Force Spectroscopy with Biological Tools*. PhD thesis, Ludwig-Maximilians-Universität München, 2017.

[74] Jobst, M. A. *Multiplexed Single Molecule Observation and Manipulation of Engineered Biomolecules*. PhD thesis, Ludwig-Maximilians-Universität München, 2018.

[75] Bell, G. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.

[76] Evans, E. and Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophysical Journal*, 72(4):1541–1555, 1997.

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

[78] Bustamante, C., Marko, J., Siggia, E., and Smith, S. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994.

[79] Livadaru, L., Netz, R. R., and Kreuzer, H. J. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules*, 36(10):3732–3744, 2003.

[80] Axelrod, D. Total Internal Reflection Fluorescence Microscopy in Cell Biology. *Traffic*, 2 (11):764–774, 2001.

[81] Mattheyses, A. L., Simon, S. M., and Rappoport, J. Z. Imaging with total internal reflection fluorescence microscopy for the cell biologist. *Journal of Cell Science*, 123(21):3621–3628, 2010.

[82] Shashkova, S. and Leake, M. C. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Bioscience Reports*, 37(4):BSR20170031, 2017.

[83] Castell, O. K., Dijkman, P. M., Wiseman, D. N., and Goddard, A. D. Single molecule fluorescence for membrane proteins. *Methods*, (February):0–1, 2018.

[84] Axelrod, D. Cell-substrate Contacts Illuminated by Total-Internal Reflection Fluorescence. *Journal of Cell Biology*, 89(9):141–145, 1981.

[85] Martin-Fernandez, M. L., Tynan, C. J., and Webb, S. E. D. A 'pocket guide' to total internal reflection fluorescence. *Journal of Microscopy*, 252(1):16–22, 2013.

[86] Pippig, D. A., Baumann, F., Strackharn, M., Aschenbrenner, D., and Gaub, H. E. Protein–DNA Chimeras for Nano Assembly. *ACS Nano*, 8(7):6551–6555, 2014.

[87] Kudalkar, E. M., Davis, T. N., and Asbury, C. L. Single-Molecule Total Internal Reflection Fluorescence Microscopy. *Cold Spring Harbor Protocols*, 2016(5):pdb.top077800, 2016.

[88] Helms, V. Principles of Computational Cell Biology. 2008.

[89] Zheng, J. FRET and Its Biological Application as a Molecular Ruler. In Jue, T., editor, *Biomedical Applications of Biophysics*, pages 119–136. Humana Press, Totowa, NJ, 2010. ISBN 978-1-60327-232-2. doi: 10.1007/978-1-60327-233-9_5. URL `http://link.springer.com/10.1007/978-1-60327-233-9http://link.springer.com/10.1007/978-1-60327-233-9{_}5`.

[90] Nickels, P. C., Wünsch, B., Holzmeister, P., Bae, W., Kneer, L. M., Grohmann, D., Tinnefeld, P., and Liedl, T. Molecular force spectroscopy with a DNA origami–based nanoscopic force clamp. *Science*, 354(6310):305–307, 2016.

[91] Hirschi, S., Stauffer, M., Harder, D., Müller, D. J., Meier, W., and Fotiadis, D. Engineering and Assembly of Protein Modules into Functional Molecular Systems. *CHIMIA International Journal for Chemistry*, 70(6):398–401, 2016.

[92] Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440 (7082):297–302, 2006.

[93] Seeman, N. C. Structural DNA Nanotechnology: Growing Along with Nano Letters. *Nano Letters*, 10(6):1971–1978, 2010.

[94] Simmel, F. C. DNA-based assembly lines and nanofactories. *Current Opinion in Biotechnology*, 23(4):516–521, 2012.

[95] Seeman, N. C. Nucleic acid junctions and lattices. *Journal of Theoretical Biology*, 99(2): 237–247, 1982.

[96] Li, S., Jiang, Q., Liu, S., Zhang, Y., Tian, Y., Song, C., Wang, J., Zou, Y., Anderson, G. J., Han, J.-Y., Chang, Y., Liu, Y., Zhang, C., Chen, L., Zhou, G., Nie, G., Yan, H., Ding, B., and Zhao, Y. A DNA nanorobot functions as a cancer therapeutic in response to a molecular trigger in vivo. *Nature Biotechnology*, 36(3):258–264, 2018.

[97] Rajendran, A., Nakata, E., Nakano, S., and Morii, T. Nucleic-Acid-Templated Enzyme Cascades. *ChemBioChem*, 18(8):696–716, 2017.

[98] Kufer, S. K., Puchner, E. M., Gumpp, H., Liedl, T., and Gaub, H. E. Single-Molecule Cut-and-Paste Surface Assembly. *Science*, 319(5863):594–596, 2008.

[99] Puchner, E. M., Kufer, S. K., Strackharn, M., Stahl, S. W., and Gaub, H. E. Nanoparticle Self-Assembly on a DNA-Scaffold Written by Single-Molecule Cut-and-Paste. *Nano Letters*, 8(11):3692–3695, 2008.

[100] Strackharn, M., Stahl, S. W., Severin, P. M. D., Nicolaus, T., and Gaub, H. E. Peptide-Antibody Complex as Handle for Single-Molecule Cut & Paste. *ChemPhysChem*, 13(4): 914–917, 2012.

[101] Strackharn, M., Pippig, D. a., Meyer, P., Stahl, S. W., and Gaub, H. E. Nanoscale Arrangement of Proteins by Single-Molecule Cut-and-Paste. *Journal of the American Chemical Society*, 134(37):15193–15196, 2012.

[102] Erlich, K. R., Baumann, F., Pippig, D. A., and Gaub, H. E. Strep-Tag II and Monovalent Strep-Tactin as Novel Handles in Single-Molecule Cut-and-Paste. *Small Methods*, 1(8): 1700169, 2017.

[103] Erlich, K. R., Sedlak, S. M., and Gaub, H. E. DNA-Free Directed Assembly in Single-Molecule Cut-and-Paste. *Manuscript in preparation*, 2018.

[104] Rockel, S., Geertz, M., and Maerkl, S. J. MITOMI: A Microfluidic Platform for In Vitro Characterization of Transcription Factor–DNA Interaction. pages 97–114. 2012. doi: 10.1007/978-1-61779-292-2_6. URL `http://link.springer.com/10.1007/978-1-61779-292-2{_}6`.

[105] Wilchek, M., Bayer, E. A., and Livnah, O. Essentials of biorecognition: The (strept)avidin–biotin system as a model for protein–protein and protein–ligand interaction. *Immunology Letters*, 103(1):27–32, 2006.

[106] Laitinen, O. H., Nordlund, H. R., Hytönen, V. P., and Kulomaa, M. S. Brave new (strept)avidins in biotechnology. *Trends in Biotechnology*, 25(6):269–277, 2007.

[107] Schmidt, T. G., Koepke, J., Frank, R., and Skerra, A. Molecular Interaction Between the Strep-tag Affinity Peptide and its Cognate Target, Streptavidin. *Journal of Molecular Biology*, 255(5):753–766, 1996.

[108] Schmidt, T. G. and Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols*, 2(6):1528–1535, 2007.

[109] Voss, S. and Skerra, A. 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(8):975–982, 1997.

[110] Kawato, T., Mizohata, E., Meshizuka, T., Doi, H., Kawamura, T., Matsumura, H., Yumura, K., Tsumoto, K., Kodama, T., Inoue, T., and Sugiyama, A. Crystal structure of streptavidin mutant with low immunogenicity. *Journal of Bioscience and Bioengineering*, 119(6): 642–647, 2015.

[111] Zhang, M., Biswas, S., Deng, W., and Yu, H. The Crystal Structure of Monovalent Streptavidin. *Scientific Reports*, 6(1):35915, 2016.

[112] Liu, F., Zhang, J. Z. H., and Mei, Y. The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Scientific Reports*, 6(1):27190, 2016.

[113] Sedlak, S. M., Schendel, L. C., Melo, M. C. R., Gaub, H. E., and Bernardi, R. C. Direction Matters: Monovalent Streptavidin under Tension. *Manuscript in preparation*, 2018.

[114] Howarth, M., Chinnapen, D. J.-F., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L., El-Husseini, A., and Ting, A. Y. A monovalent streptavidin with a single femtomolar biotin binding site. *Nature Methods*, 3(4):267–273, 2006.

[115] DeMonte, D., Drake, E. J., Lim, K. H., Gulick, A. M., and Park, S. Structure-based engineering of streptavidin monomer with a reduced biotin dissociation rate. *Proteins: Structure, Function, and Bioinformatics*, 81(9):1621–1633, 2013.

[116] Bartlett, J. M. S. and Stirling, D. A Short History of the Polymerase Chain Reaction. In *PCR Protocols*, pages 3–6. Humana Press, New Jersey, 2003. doi: 10.1385/1-59259-384-4:3. URL `http://link.springer.com/10.1385/1-59259-384-4:3`.

[117] Cline, J., Braman, J. C., and Hogrefe, H. H. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, 24(18):3546–3551, 1996.

[118] Ramsay, N., Jemth, A.-S., Brown, A., Crampton, N., Dear, P., and Holliger, P. CyDNA: Synthesis and Replication of Highly Cy-Dye Substituted DNA by an Evolved Polymerase. *Journal of the American Chemical Society*, 132(14):5096–5104, 2010.

[119] McGrath, S. and van Sinderen, D. *Bacteriophage: Genetics and Molecular Biology, 1st edition.* Caister Academic Press, Wymondham, UK, 2007. ISBN 978-1-904455-14-1.

[120] Borkotoky, S. and Murali, A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *International Journal of Biological Macromolecules*, 118:49–56, 2018.

[121] Klement, J. F., Moorefiedl, M. B., Jorgensen, E., Brown, J. E., Risman, S., and McAllister, W. T. Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. *Journal of Molecular Biology*, 215(1):21–29, 1990.

[122] Újvári, A. and Martin, C. T. Thermodynamic and Kinetic Measurements of Promoter Binding by T7 RNA Polymerase. *Biochemistry*, 35(46):14574–14582, 1996.

[123] Bandwar, R. P. and Patel, S. S. The Energetics of Consensus Promoter Opening by T7 RNA Polymerase. *Journal of Molecular Biology*, 324(1):63–72, 2002.

[124] Bandwar, R. P., Jia, Y., Stano, N. M., and Patel, S. S. Kinetic and Thermodynamic Basis of Promoter Strength: Multiple Steps of Transcription Initiation by T7 RNA Polymerase Are Modulated by the Promoter Sequence. *Biochemistry*, 41(11):3586–3595, 2002.

[125] Ramírez-Tapia, L. E. and Martin, C. T. New Insights into the Mechanism of Initial Transcription. *Journal of Biological Chemistry*, 287(44):37352–37361, 2012.

[126] Chamberlin, M. and Ring, J. Characterization of T7-Specific Ribonucleic Acid Polymerase. 1. General Properties of the Enzymatic Reaction and the Template Specificity of the Enzyme. *The Journal of Biological Chemistry*, 248(6):2235–2244, 1973.

[127] Revyakin, A., Liu, C., Ebright, R. H., and Strick, T. R. Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science*, 314(5802):1139–1143, 2006.

[128] Skinner, G. M., Kalafut, B. S., and Visscher, K. Downstream DNA Tension Regulates the Stability of the T7 RNA Polymerase Initiation Complex. *Biophysical Journal*, 100(4): 1034–1041, 2011.

[129] Zhang, X., Yao, Z., Duan, Y., Zhang, X., Shi, J., and Xu, Z. Investigation of specific interactions between T7 promoter and T7 RNA polymerase by force spectroscopy using atomic force microscope. *Biochemical Journal*, 475(1):319–328, 2018.

[130] Doherty, A. J. and Wigley, D. B. Functional domains of an ATP-dependent DNA ligase 1 1Edited by A. R. Fersht. *Journal of Molecular Biology*, 285(1):63–71, 1999.

[131] Doherty, A. J. and Dafforn, T. R. Nick recognition by DNA ligases. *Journal of Molecular Biology*, 296(1):43–56, 2000.

[132] Wilkinson, A., Smith, A., Bullard, D., Lavesa-Curto, M., Sayer, H., Bonner, A., Hemmings, A., and Bowater, R. Analysis of ligation and DNA binding by Escherichia coli DNA ligase (LigA). *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1749(1):113–122, 2005.

[133] Nandakumar, J., Nair, P. A., and Shuman, S. Last Stop on the Road to Repair: Structure of E. coli DNA Ligase Bound to Nicked DNA-Adenylate. *Molecular Cell*, 26(2):257–271, 2007.

[134] Bauer, R. J., Zhelkovsky, A., Bilotti, K., Crowell, L. E., Evans, T. C., McReynolds, L. A., and Lohman, G. J. S. Comparative analysis of the end-joining activity of several DNA ligases. *PLOS ONE*, 12(12):e0190062, 2017.

[135] Mueser, T. C., Hinerman, J. M., Devos, J. M., Boyer, R. A., and Williams, K. J. Structural analysis of bacteriophage T4 DNA replication: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virology Journal*, 7(1):359, 2010.

[136] Liu, X., Huang, A., Luo, D., Liu, H., Han, H., Xu, Y., and Liang, P. Use of adenylate kinase as a solubility tag for high level expression of T4 DNA ligase in Escherichia coli. *Protein Expression and Purification*, 109:79–84, 2015.

[137] Pritchard, C. Effects of base mismatches on joining of short oligodeoxynucleotides by DNA ligases. *Nucleic Acids Research*, 25(17):3403–3407, 1997.

[138] Ng, P.-S. and Bergstrom, D. E. Protein-DNA footprinting by endcapped duplex oligodeoxyribonucleotides. *Nucleic Acids Research*, 32(13):e107, 2004.

[139] Fraga, H. Firefly luminescence: A historical perspective and recent developments. *Photochemical & Photobiological Sciences*, 7(2):146, 2008.

[140] England, C. G., Ehlerding, E. B., and Cai, W. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. *Bioconjugate Chemistry*, 27(5):1175–1187, 2016.

[141] Loening, A. M., Fenn, T. D., and Gambhir, S. S. Crystal Structures of the Luciferase and Green Fluorescent Protein from Renilla reniformis. *Journal of Molecular Biology*, 374(4): 1017–1028, 2007.

[142] Baldwin, T. O., Christopher, J. A., Raushel, F. M., Sinclair, J. F., Ziegler, M. M., Fisher, A. J., and Rayment, I. Structure of bacterial luciferase. *Current Opinion in Structural Biology*, 5 (6):798–809, 1995.

[143] Head, J. F., Inouye, S., Teranishi, K., and Shimomura, O. The crystal structure of the photoprotein aequorin at 2.3 Å resolution. *Nature*, 405(6784):372–376, 2000.

[144] Grate, D. and Wilson, C. Laser-mediated, site-specific inactivation of RNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11): 6131–6, 1999.

[145] Babendure, J. R., Adams, S. R., and Tsien, R. Y. Aptamers Switch on Fluorescence of Triphenylmethane Dyes. *Journal of the American Chemical Society*, 125(48):14716–14717, 2003.

[146] Furukawa, K., Abe, H., Abe, N., Harada, M., Tsuneda, S., and Ito, Y. Fluorescence generation from tandem repeats of a malachite green RNA aptamer using rolling circle transcription. *Bioorganic & Medicinal Chemistry Letters*, 18(16):4562–4565, 2008.

[147] Yin, J., Liu, F., Li, X., and Walsh, C. T. Labeling Proteins with Small Molecules by Site-Specific Posttranslational Modification. *Journal of the American Chemical Society*, 126 (25):7754–7755, 2004.

[148] 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. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences*, 102(44):15815–15820, 2005.

[149] Yin, J., Lin, A. J., Golan, D. E., and Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nature Protocols*, 1(1):280–285, 2006.

[150] Zahnd, C., Spinelli, S., Luginbühl, B., Amstutz, P., Cambillau, C., and Plückthun, A. Directed in Vitro Evolution and Crystallographic Analysis of a Peptide-binding Single Chain Antibody Fragment (scFv) with Low Picomolar Affinity. *Journal of Biological Chemistry*, 279(18):18870–18877, 2004.

[151] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, 2009.

[152] Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1):207–234, 2005.

[153] Stockert, J. C. and Blazquez-Castro, A. *Fluorescence Microscopy in Life Sciences*. BENTHAM SCIENCE PUBLISHERS, 2017. ISBN 9781681085180. doi: 10.2174/9781681085180101. URL http://www.eurekaselect.com/158339/volume/1.

[154] Vaughan, J. C., Dempsey, G. T., Sun, E., and Zhuang, X. Phosphine Quenching of Cyanine Dyes as a Versatile Tool for Fluorescence Microscopy. *Journal of the American Chemical Society*, 135(4):1197–1200, 2013.

[155] Erlich, K. R. *Integration of a Functional Enzyme into Single Molecule Cut & Paste*. PhD thesis, Ludwig-Maximilians-Universität München, 2014.

[156] Gumpp, H., Stahl, S. W., Strackharn, M., Puchner, E. M., and Gaub, H. E. Ultrastable combined atomic force and total internal fluorescence microscope. *Review of Scientific Instruments*, 80(6):063704, 2009.

[157] Durner, E., Ott, W., Nash, M. A., and Gaub, H. E. Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. *ACS Omega*, 2(6):3064–3069, 2017.

[158] Zimmermann, J. L., Nicolaus, T., Neuert, G., and Blank, K. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nature Protocols*, 5(6):975–985, 2010.

[159] Celik, E. and Moy, V. T. Nonspecific interactions in AFM force spectroscopy measurements. *Journal of Molecular Recognition*, 25(1):53–56, 2012.

[160] Butt, H. J. and Jaschke, M. Calculation of thermal noise in atomic force microscopy. *Nanotechnology*, 6(1):1–7, 1995.

[161] Guo, S., Lad, N., Ray, C., and Akhremitchev, B. B. Association Kinetics from Single Molecule Force Spectroscopy Measurements. *Biophysical Journal*, 96(8):3412–3422, 2009.

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

[163] Hegedüs, É., Kókai, E., Kotlyar, A., Dombrádi, V., and Szabó, G. Separation of 1–23-kb complementary DNA strands by urea–agarose gel electrophoresis. *Nucleic Acids Research*, 37(17):e112, 2009.

[164] Grodberg, J. and Dunn, J. J. ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *Journal of Bacteriology*, 170(3): 1245–1253, 1988.

[165] Rio, D. C. Expression and Purification of Active Recombinant T7 RNA Polymerase from E. coli. *Cold Spring Harbor Protocols*, 2013(11):pdb.prot078527–pdb.prot078527, 2013.

[166] Radmacher, M., Tillamnn, R. W., Fritz, M., and Gaub, H. E. From Molecules to Cells: Imaging Soft Samples with the Atomic Force Microscope. *Science*, 257(5078):1900–1905, 1992.

[167] Radmacher, M., Fritz, M., Hansma, H., and Hansma, P. Direct observation of enzyme activity with the atomic force microscope. *Science*, 265(5178):1577–1579, 1994.

[168] Edwards, D. T. and Perkins, T. T. Optimizing force spectroscopy by modifying commercial cantilevers: Improved stability, precision, and temporal resolution. *Journal of Structural Biology*, 197(1):13–25, 2017.

[169] Jadhav, V. S., Brüggemann, D., Wruck, F., and Hegner, M. Single-molecule mechanics of protein-labelled DNA handles. *Beilstein Journal of Nanotechnology*, 7(1):138–148, 2016.

[170] Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 24 (1):79–88, 2006.

[171] Bockelmann, U., Essevaz-Roulet, B., and Heslot, F. Molecular Stick-Slip Motion Revealed by Opening DNA with Piconewton Forces. *Physical Review Letters*, 79(22):4489–4492, 1997.

[172] Morfill, J., Kühner, F., Blank, K., Lugmaier, R. A., Sedlmair, J., and Gaub, H. E. B-S transition in short oligonucleotides. *Biophysical Journal*, 93(7):2400–2409, 2007.

[173] Bayer, E. A., Belaich, J.-P., Shoham, Y., and Lamed, R. The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annual Review of Microbiology*, 58(1):521–554, 2004.

[174] Moraïs, S., Stern, J., Kahn, A., Galanopoulou, A. P., Yoav, S., Shamshoum, M., Smith, M. A., Hatzinikolaou, D. G., Arnold, F. H., and Bayer, E. A. Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability. *Biotechnology for Biofuels*, 9(1):164, 2016.

[175] Kufer, S. K., Strackharn, M., Stahl, S. W., Gumpp, H., Puchner, E. M., and Gaub, H. E. Optically monitoring the mechanical assembly of single molecules. *Nature Nanotechnology*, 4(1):45–49, 2009.

[176] Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A., and Rief, M. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nature Structural and Molecular Biology*, 11(1):81–85, 2004.

[177] Milles, L. F., Bayer, E. A., Nash, M. A., and Gaub, H. E. Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium perfringens. *The Journal of Physical Chemistry B*, 121(15):3620–3625, 2017.

List of Figures

Acknowledgments

I would like to personally and sincerely thank the following people, whose contributions to the work of this dissertation and my doctoral studies in general cannot be overlooked:

Prof. Dr. Hermann E. Gaub first and foremost for the incredible opportunity to pursue my PhD with this group under your guidance. I'm not sure exactly what I expected when I wrote you that first timid email asking for a position here, but I don't think I could have ever predicted an experience so rewarding.

Magnus Bauer for many helpful discussions, as well as an abundant positive energy even when life is at its most abominable.

Dr. Fabian Baumann for the support with SMC&P and fruitful discussions surrounding nanotechnology troubleshooting.

Ellis Durner for the countless times I've gotten assistance on all things large and small - including but not limited to technical support, experimental setup, dissertation feedback and educating me on the finer points of coffee - and for matching my affinity for memes and doing it tomorrow.

Dr. Markus Jobst for improvements to AFM setups, creating the most beautiful LaTeX templates, and a mutual adoration of nonsense cartoon shows with too many colors.

Angelika Kardinal for the extensive lab support as well as being an excellent office mate and source of encouragement.

Sylvia Kreuzer for running the administrative side of things with great aplomb, providing the Gummibärchen when we need them most, and always knowing how to calm us down when we have little organizational emergencies.

Achim Löf for an exquisitely dry wit, taking trash parties very seriously, and seeing both the sunny side and the other side.

Lukas Milles for carefully reading my dissertation and providing many helpful suggestions, assisting with coding and SMFS analysis, and for maintaining solidarity in the Aeropress when we are surrounded on all sides by espresso fanatics.

Thomas Nicolaus for preparing buffers, maintaining the lab, and silanizing… however many hundreds of glass slides for me. But more importantly, for that thumpin bass and the many engaging chats about very important and professional topics, such as audio-visual computational entertainment and how Mutter isn't actually that scary.

Dr. Wolfgang Ott for a seemingly endless wealth of knowledge and expertise of molecular biology and biotechnology, and a willingness to share it with others to make our lab experiences much smoother.

Marilena Pinto for being an absolute life-saver and organizational angel for CeNS and the

SFB, and without whom I would still be wandering around the KVR like a lost puppy.

Dr. Diana Pippig for invaluable mentorship in the first four years of my work.

Leonard Schendel for the many discussions of fluorescence microscopy, and for all of the hard work put into improving and maintaining the Seuss.

Steffen Sedlak for the excellent collaboration and easygoing attitude, and for tolerating my idea of entertaining films.

Eduard Unterauer for having the patience of a saint while people (including myself) constantly grind coffee directly behind you, and for the mooncake.

Dr. Tobias Verdorfer for the assistance with designing a new PDMS chip for SMC&P, the general Tobi-energy, and most importantly the Schokokiste - which constituted a frightening fraction of my calories in the months of writing the thesis.

Dr. Philipp Walker for being the best running buddy, tea break buddy, trash film buddy, language-learning buddy, and pöbeln buddy. And if you doubt that any of this is true, wann habe ich letztes mal was falsches gesagt?

I would also like to thank the whole **Gambi Crew** for creating such a wonderful working atmosphere! I received help and insight from all of my colleagues on numerous occasions, from scientific input and technical assistance to translations and coffee break conversations.

And last but not least, **my family - Jessica, Jennifer and Kim -** for your love, support, and sticking by me despite the distance. We may never find the parking lot or remember the cake, but we'll always have our terrible sense of humor.