Adaptation of HUH endonucleases for protein-DNA conjugation

A THESIS
SUBMITTED TO THE FACULTY OF
THE UNIVERSITY OF MINNESOTA
BY

Klaus N. Lovendahl

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Prof. Wendy R. Gordon

October 2018

Abstract

The conjugation of proteins to DNA is a useful technique for basic research and medicine. Current methods for conjugation are limited in both specificity and efficiency, as they generally depend on nonspecific exogenous chemical crosslinkers or reversible DNA-binding proteins. In this work we have adapted a natural protein domain, known as the HUH domain, which targets DNA and forms a covalent phosphotyrosine adduct with the DNA backbone. We show that HUH domains react efficiently with specific sequences of unmodified single-stranded DNA, and we have identified five proteins which react orthogonally with distinct DNA sequences. We demonstrate a number of applications for this technology. HUH fusion proteins can be use to label proteins both *in vitro* and in cultured mammalian cells with excellent specificity. Fusing an HUH domain to the gene-editing Cas9 protein enhances the integration of a single-stranded donor DNA template in mammalian cells. Surface-adhered HUH protein can be used to tether DNA molecules for sensitive measurements in a single-molecule magnetic tweezer. Additionally, we show that an HUH domain protein can be rationally mutated to increase its stability and reactivity. This work presents a novel technology with wide applicability and presents several avenues for future work.

Acknowledgments

Hideki Aihara
Eric Aird
Rahul Banerjee
Thomas Bohl
Robert Evans III
Irina Goldstein
Suzy Gustafson & Sam Robertson
Amanda Hayward
Corey Knoot
Anna Komor
John Lee
Karen Howe & Chris Lovendahl
Ari Lovendahl
Bella, Felix, & Uli Lovendahl
Nisse Lovendahl & Matt Robertson
Guillermo Marques
Megan McCarthy
Katie Muratore & Mark Smalley
John Oja
Lien Phung
Brent Rivard
Melanie Rogers
Kassidy Tompkins
Alina Zdelchik

And, of course, Wendy Gordon. Thank you for all your guidance and encouragement.

Contents

List of Figures v

List of Tables vi

1 Introduction 1

1.1 Information flow in nature and biotechnology 1
1.2 The example of immunodetection 2
1.3 Current methods 3

2 Development of the HUH tags 9

2.1 Introduction 9
2.2 Results 12
2.3 Conclusions 18
2.4 Current and future work: Cas9-HUH fusions for gene editing 21
2.5 Experimental procedures 24
2.6 Supplemental material from Lovendahl et al. *JACS* 2017 29

3 Rational improvement of the miniMobA relaxase 42

3.1 Improvement of enzymes for human chemistry 42
3.2 Results 47
3.3 Conclusions 51
3.4 Experimental procedures 52

4 HUH conjugation in magnetic tweezers 54

4.1 Introduction 54
4.2 Results 59
4.3 Conclusions and future directions 64

4.4 Experimental procedures 64

5 References **68**

List of abbreviations **81**

Supplementary data **82**

Supplement to Chapter 2 82

Supplement to Chapter 3 85

Supplement to Chapter 4 86

List of Figures

1.1 Detection methods in immunoassay 3

2.1 Conserved structure of the HUH endonucleases 10
2.2 Chemistry of phosphotyrosine bond formation 11
2.3 Yield and rate of the HUH reaction 13
2.4 Sequence specificity of the HUH reaction 15
2.5 Applications of purified HUH-fusion proteins 16
2.6 Labeling of fusion proteins in mammalian cells 18
2.7 Cell surface labeling on live cells 19
2.8 Cas9-HUH fusions increase genome editing 22
2.9 HUH reaction optimization using a 96-well plate assay 30
2.10 Testing stability and reversibility of HUH-DNA conjugates. 31
2.11 Full yield gels 32
2.12 Optimization of mMobA target sequence 33
2.13 Cas9-HUH RNP transfection 34
2.14 Reactions of HUH-tagged proteins with DNA origami structures 35
2.15 Soluble lysate labeling 36
2.16 Fixed cell imaging controls 37
2.17 HUH-tagged Notch receptors function normally 38
2.18 One-pot labeling of cell-surface HUH-tags 39
2.19 Timecourse of cell-surface labeling 40
2.20 Live-cell fluorescent labeling of HUH-tagged intracellular protein 41

3.1 Designed mMobA mutations 44
3.2 Reaction efficiency of mMobA mutant proteins 48
3.3 Thermal denaturation of mMobA mutant proteins 50

4.1 Magnetic tweezers schematic and example images 56
4.2 Enzyme activity in magnetic tweezers 57
4.3 Overhang DNA by spacer PCR 59
4.4 Overhang DNA by self-assemby 61
4.5 Anchoring of spacer PCR DNA 62
4.6 Single-molecule tracking of mMobA-anchored DNA 63

1 Uncropped gel images 85
2 Immobilization of overhang PCR DNA 88
3 Raw traces of mMobA-tethered DNA 89

List of Tables

1.1 Existing methods for protein-DNA conjugation 6

2.1 Proteins studied in this work; see 4.4 for full details 12

3.1 Calculated melting temperatures for mMobA mutants and the fused SUMO tag 50

1 Raw data for fig. 4.3 87

Introduction 1

Information flow in nature and biotechnology

The fundamental basis of this work is laid out by Francis Crick in 1958, in the first statement of the 'Central Dogma':

> *...once 'information' has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible*[1].

The key point is the capacity of each class of molecules to store and transmit information, and the importance of that information. In nature proteins can function as 'black boxes'[2] - the key to their existence is what they do rather than how they do it. As long as $A \rightarrow B$, it does not matter which enzyme performs the conversion - an enzyme can be replaced by a homolog with the same function from a distinct species[3]. In contrast, information storage is the purpose of DNA - it must be readable to produce proteins and permit reproduction and writeable to drive evolution.

The one-way flow of information becomes a problem, however, in the field of biotechnology. In research and medicine it is vitally important to understand the *how* of an event. For example, antibodies are involved in the body's immune response to infection as well as medical diagnostic tests, but in very different ways. The immune response is activated by an antibody binding to a foreign molecule; while the binding region of the antibody varies to match the antigen, activation of the immune system is mediated by the conserved Fc region. Medical diagnostics, however, seek to provide information on the specific infection in order to guide treatment. For instance, treatment of a viral infection is heavily dependent on the class of virus - RNA vs. DNA genome, for instance. In essence, biotechnology at-

tempts to run the Central Dogma in reverse - to extract human-readable information from a function.

The example of immunodetection

An illuminating example is provided by the traditional method of protein detection and quantification by labeled antibodies in an enzyme-linked immunosorbent assay (ELISA). In essence, every immunoassay converts a molecular recognition event - the binding of an antibody to antigen - into a human-readable format, whether production of light, change of color, or increase in fluorescence. In practice, a molecule of interest is detected by use of an antibody specific to that molecule conjugated with some detectable group. The labeled antibody is added to a complex mixture containing the target and allowed to bind; unbound antibody is washed away, and the remaining bound antibody is detected. In the traditional form, this is a horseradish peroxidase (HRP) or alkaline phosphatase (AP) enzyme, which can process substrates to produce light or a change in color. The use of antibody conjugates [a] has been a powerful technique for the detection of biological molecules since the 1960's; the early methods of fluorescent[4], chemiluminescent, and colorimetric[5] detection are still in use today. These detection methods have high sensitivity but are limited to a single target.

Multiplex detection can be achieved by replacing a single luminescence/colorimetric signal with multicolor fluorescence. Upon excitation, the labeled target molecule releases detectable fluorescence; using suitable excitation and emission filters, multiple wavelengths, and thus, targets, can be detected. The number of targets detectable by fluorescence is limited by the need for fluorophores and detector bandwidth in different colors; moving beyond the *de facto* standard of red, green, and blue remains a challenge due to spectral overlap between fluorophores. While complex spectral mixing techniques have been developed to expand this range into the hundreds[6], detection beyond ten colors remains challenging[7].

Expanding detection to, potentially, the entire proteome is possible through DNA barcoding. DNA-antibody conjugates were first introduced in 1992 for use in immuno-PCR[8]. Later work expanded this approach by integrating rolling-circle replication for greater sensitivity and multiplex detection[9] and using DNA ligation-based amplification to increase sensitivity[10]. However, retrieving the full information encoded by the DNA was only made

[a] While the example of antibodies is used here, this is only because they are the historical standard. The same points could apply to any recognition protein.

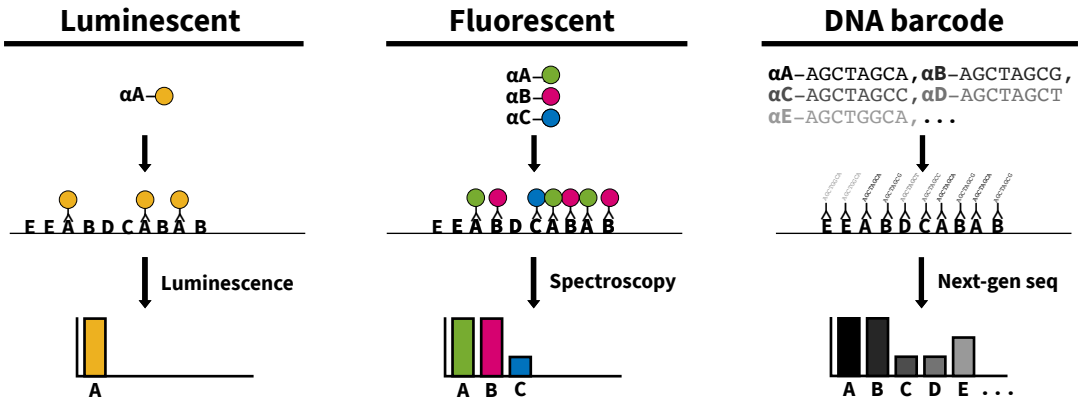

Figure 1.1: **Comparison of protein-conjugates for detection** a. Traditional luminscent detectino converts a single binding signal into the production of light. b. Multiple fluorescently-labeled antibodies can detect their targets in a single reaction. c. DNA-labeled antibodies can be quantified by next-gen sequencing, expanding the detection range into the thousands.

possible by modern next-gen sequencing, which can sequence and quantify thousands of strands in a single reaction. A strand of DNA of length N base-pairs has 4^N possible sequences; an 8-mer encodes $4^8 = 65,536$ unique sequences, sufficient to assign a unique code to every species in the proteome[11]. This approach was implemented in 2017 to quantify 82 proteins on single cells[12].

This example illustrates the multiplexing power of DNA labeling as a replacement for traditional labeling methods. By replacing light or fluorescent read out with specific sequences the detection range can be expanded into the thousands, allowing fine-grained study of complex samples.

Current methods

Existing strategies to link DNA to proteins can be roughly classified into those which are strong (here meaning covalently bonded) but nonspecific vs. those which are specific but not strong.

Chemical crosslinkers

Covalent bonds can be formed using crosslinking reagents specific for chemical groups found on proteins and DNA. A crosslinker is a bifunctional molecule containing two or more reactive groups joined by some linker; they can be either homobifunctional, with the same reactive group on both termini, or heterobifunctional, presenting different reactive groups. The most common reactive groups for protein-DNA crosslinking are *N*-hydroxysuccinimide (NHS), which reacts with primary amines, and maleimide, which reacts with reduced thiols[13]. Both molecules are almost completely non-specific, attacking any free amine or reduced thiol. This generality means that they are the only method for labeling naturally-occurring proteins, such as antibodies. However, labeling of non-antibody proteins is difficult as they have many fewer 'free' residues which can be labeled without perturbing function.

The naturally-occurring groups available for chemical labeling on proteins are the primary amines of lysine and the N-terminus, the thiol side chain of cysteine, and to a lesser extent the carboxyl groups of aspartate, glutamate, and the C-terminus. DNA does not naturally contain easily-attackable groups; linking is generally accomplished by adding amine or thiol groups to synthetic oligonucleotides used for PCR or hybridization. The 5' phosphate of DNA can be labeled using the carbodiimide EDC with imidazole as a catalyst[14], but this reaction is not widely used. DNA can be non-specifically attacked by intercalating molecules such as psoralen, which can form adducts to pyrimidine bases upon exposure to UV light[15]; again, this method is not widely used.

In practice, crosslinking is generally performed with a heterobifunctional crosslinker in a two-step reaction scheme. A commonly used linker is SMCC[b], containing NHS and maleimide groups, which will be used as an example; the full range of crosslinking schemes cannot be addressed here, although a number are listed in Table 1.1. The problem of specificity is addressed, as much as is possible, by separating the reactions. DNA containing a 5'-amine group (commercially available from most suppliers) is mixed with the crosslinker first and allowed to react with the NHS group of the SMCC, then the product is purified to remove unreacted crosslinker. Because the amine is synthetic there is only one group per DNA molecule, and no thiols are present; this prevents unwanted inter-DNA linking. The

[b] Succinimidyl 4-(*N*-maleimidomethyl) cyclohexane-1-carboxylate, available from Thermo Fisher Scientific and other suppliers

purified linker-DNA is then added to the protein, where the maleimide group reacts with free thiols on the target protein. Purification and validation is then performed; a final yield of approximately 50% is common[c]. This reaction could obviously be reversed, with thiol DNA and targeting protein amine groups; almost every reaction is different, and must be optimized to the specific protein and its intended purpose.

'Click' chemistry

The relatively new class of 'click' reactions, typified by the Copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC)[39,40] are ideally suited for bioconjugation, being highly efficient, specific, and completely bioorthogonal - neither azide nor alkyne groups normally occur in biological systems. For these reasons they merit special attention, and are the focus of intense research.

Owing to their small size, both azides[25] and alkynes[27] can be incorporated into proteins using modified amino acids[d] or oligonucleotides using modified bases[41].

A drawback to the azide-alkyne reaction is its requirement for the toxic catalyst Cu^{I}. This led to the development of the 'strain-promoted', copper-free click reactions using cyclooctynes[42], which avoid the catalyst but at a substantial cost in reaction rate.

The newer inverse-electron-demand Diels-Alder reaction between *trans*-cyclooctene and tetrazine[43] is both copper-free and extremely fast, making it ideal for bioconjugation. This reaction has been successfully performed in live mice[44], and tetrazine-modified amino acids have been successfully incorporated into proteins[26,45].

Given their usefulness, the click reactions are an area of intense and ongoing research. However, they are an improvement on the crosslinking methods described above rather than a novel form - any alkyne reacts with any azide. If the functional groups can be controlled the reaction is very useful, but the possiblities for multiplex labeling are still limited.

Specific biological attachment

Chemical crosslinkers target a single group for an irreversible reaction, generating a single strong bond; in contrast, biological recognition is mediated by a large number of weak

[c] author's experience

[d] Unnatural amino acids for labeling are a vast and rapidly-evolving field, which cannot be covered here; interested readers are directed to the excellent review by Lang and Chin[26].

Table 1.1: Existing methods for protein-DNA conjugation; interested readers can find more examples in the Crosslinking Technical Handbook published by Thermo Fisher Scientific[16]

Crosslink	Method	DNA group	Protein group	Notes	Ref.
Amine-amine	Glutaraldehyde (GA)	amine	Lys, N-terminus	targets endogenous protein groups	17
	bis-*N*-hydroxysuccinimide	amine	Lys, N-terminus	targets endogenous protein groups	
Amine-thiol	*N*-hydroxysuccinimide-maleimide	amine	Cys	targets endogenous protein groups	18,19
	N-hydroxysuccinimide-maleimide	thiol	Lys, N-terminus	targets endogenous protein groups	20
Amine-carboxylate	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)	amine	Asp, Glu, C-terminus	targets endogenous protein groups	
Amine-phosphate	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)	5'-phosphate	Lys, N-terminus	targets endogenous protein groups, targets endogenous DNA groups, uses imidazole catalyst	14
Aldehyde-thiol	maleimide-hydrazide	thiol	aldehyde	aldehydes can be formed by oxidizing glycans	21
Amine-thymine	*N*-hydroxysuccinimide-psoralen	thymine	Lys, N-terminus	psoralen intercalates nonspecifically into dsDNA, where it can form adducts upon activation with UV light	15
Thiol-thiol	dithiol, bis-maleimide	thiol	Cys	targets endogenous protein groups	22–24
'Click' chemistry	azide-alkyne	alkyne	azidohomoalanine	uses Cu^{1+} catalyst	25,26
	azide-alkyne	azide	homopropargylglycine	uses Cu^{1+} catalyst	26,27
	trans-cyclooctene - tetrazine				26
Biotin-biotin	streptavidin/biotin	biotin	biotin or Strep-tag/SBP	tetravalent streptavidin can bridge multiple biotins	28,29
Protein fusion	SNAP-tag	O^6-benzylguanine	SNAP-tag fusion		30,31
	CLIP-tag	O^2-benzylcytosine	CLIP-tag fusion		32
	HaloTag	chloroalkane	HaloTag fusion		33
	Streptavidin/biotin	Biotin	streptavidin fusion	streptavidin protein can be fused to proteins	34
	Zinc-finger	specific DNA sequence	zinc-finger fusion		35,36
DNA fusion	Aptamer	aptamer	specific protein		37,38

contacts, which in sum can produce a strong interaction.

For example, the model DNA-binding protein Zif268 targets the nucleotide sequence GCGTGGGCG and binds with a dissociation constant of 0.17nM[46]. This interaction is mediated by seventeen direct contacts between protein side chains and both the DNA bases and phosphate backbone[46]; mutation of a single hydrogen-bonding arginine reduces the affinity over 100-fold[46]. This cooperativity creates a selectivity filter - if even a slightly wrong DNA sequence is bound, the protein dissociates quickly and can seek out its correct target. This selectivity is unnecessary *in vitro* - the experimenter applies their own filter by only supplying the correct sequence. This limits the efficiency of coupling when seeking to form protein-DNA complexes, as there is still a dissociation rate from the correct sequence.

Natural interactions have been exploited for protein-DNA construction to varying degrees of success. Engineered zinc-finger proteins, which target a designed DNA sequence, have been employed to position proteins on a DNA origami structure[35], although with low yield ($\tilde{5}0\%$). Using the reverse strategy, protein-binding DNA aptamers have been incorporated into DNA structures to specifically immobilize proteins[37]. Both of these methods are highly limited, requiring existing aptamers or zinc-fingers as well as space to incorporate them, making them impractical for small structures.

The same general model of summed weak interactions holds for proteins which bind small molecules, with some notable exceptions.

Special attention should be paid to the avidin proteins, primarily streptavidin, which bind the vitamin biotin with a dissociation constant in the femtomolar range[47]. Although not covalent, for most laboratory applications this interaction is essentially irreversible. In addition the protein is extremely stable (T_m = 75.5°C) and arranged in a tetramer with two binding sites per side, creating a versatile 'nano-hub' which can link multiple biotinylated molecules[48]. Biotin-conjugated oligonucleotides are widely available, and proteins can be biotinylated using reactive esters or by enzymatic means[49]. A notable limitation of this system is found in force-probing experiments, as the off-rate increases dramatically with applied force[50].

Specific covalent bonds can be formed between fusion proteins containing the SNAP[30], CLIP[32], or HALO[51] tags. These proteins target small molecules (see 1.1), which can be attached to DNA using NHS or maleimide esters. However, all are proprietary technologies and conjugation must be performed in the lab, adding significant cost and time to proce-

dures.

Both avidin/biotin and the covalent bond-forming proteins are still limited by chemistry, in the same way as chemical crosslinkers. Both the recognition and reaction of these molecules are determined by the basic chemistry of the molecules they target, and thus is very difficult or impossible to extend for greater diversity in targeting.

The ideal: combine strength and specificity

The ideal method for DNA-protein conjugation would combine the specificity of a DNA sequence target, as in the zinc-fingers, with the strength of a covalent bond. In our research we have adapted a group of proteins, the HUH endonucleases, which seem to fit this criteria. In this work we investigate the utility of these proteins as fusion tags for linking proteins to DNA in a variety of contexts.

Development of the HUH tags

2

> Good art should elicit a response of *'Huh? Wow!'* as opposed to *'Wow! Huh?'*
>
> EDWARD RUSCHA

Adapted with permission from Lovendahl et al. J. Am. Chem. Soc. 2017[52]

Introduction

The HUH endonucleases are a large superfamily of proteins involved in the cleavage and ligation of single-stranded DNA (ssDNA). The hallmarks of the family are the conserved HUH, or 'His-hydrophobic-His', motif, which coordinates a divalent metal ion, and one or more catalytic tyrosine residues which attacks the phosphate backbone to nick the DNA[a]. In the natural DNA processing cycle, the HUH protein nicks a segment of ssDNA at a specific point in the genome (known as the *ori*) using the catalytic tyrosine, generating a 5'-phosphotyrosine bond and a free 3'-hydroxyl group. The 3'-OH is then used to initiate replication using the standard replication machinery, leading to the production of new ssDNA. Following complete replication of the circular template, the 5' and 3' ends of the DNA are ligated, either by the attached HUH protein or another free protein, creating a closed, circular single-stranded genome.

[a]For an in-depth treatment of the basic biology of the HUH endonucleases see Chandler et al. [53].

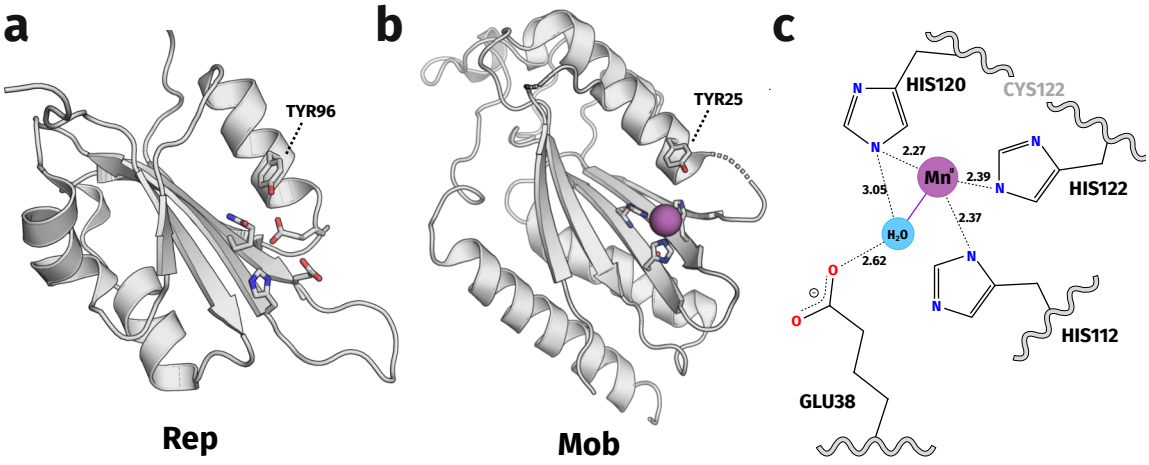

Figure 2.1: **Conserved structure of HUH endonucleases from disparate sources** a. NMR structure of the rep domain from Porcine circovirus 2 (PDB ID 2HW0). b. X-ray structure of the relaxase domain of mobilization protein A from plasmid R1162 (PDB ID 2NS6). c. Active site of the R1162 mobilization protein A

The mechanism by which HUH proteins recognize their target sequence varies, but generally involves DNA hairpins which form in the ssDNA. Although only a few HUH-DNA complex structures are known, all use hairpin structures and recognition of both the DNA bases and backbone over a large area of the protein's surface[54–58].

While the primary sequence of family members is highly variable, all known examples share a similar three-dimensional structure of a central β-sheet which presents the HUH metal-coordinating motif opposite a catalytic tyrosine on a single α-helix (fig. 2.1, a&b). This structure is found in HUH proteins from such disparate sources as porcine circovirus 2 (PCV2), which causes illness in domestic pigs, and the broad host-range plasmid R1162, originally isolated from *Pseudomonas aeuruginosa*[59,60].

The HUH motif coordinates a divalent metal ion with the aid of another residue on an adjacent β-sheet. Fig.2.1c shows a typical metal coordination site from the MobA relaxase. A manganese ion is coordinated by the imidazole groups of histidine residues, two from the HUH motif and one located on an adjacent β-strand. A fourth coordination site is occupied by a water molecule held between the manganese and a glutamate residue. The metal-binding site is typical for a divalent metal[61], and has essentially no specificity. HUH proteins have been crystallized with copper[56], magnesium[62], manganese[60], nickel[56,57],

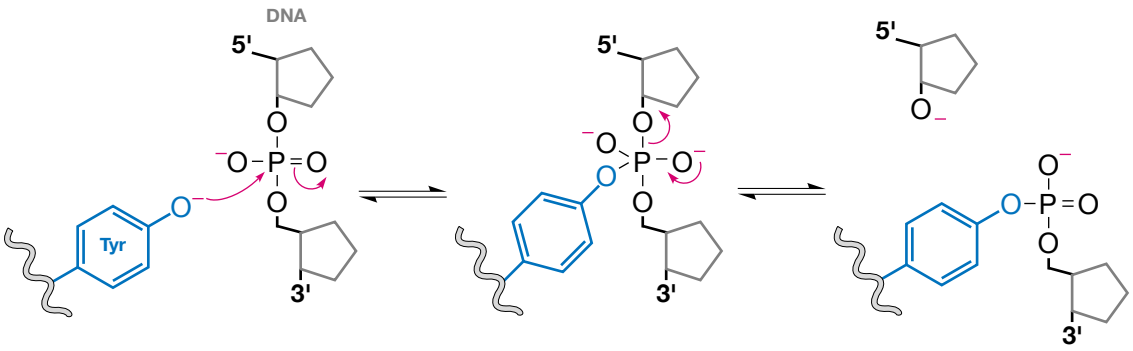

Figure 2.2: **Chemistry of phosphotyrosine bond formation** The reaction is initiated by the hydroxyl group of the tyrosine residue (blue), which attacks the phosphorous atom in the DNA backbone. A pentavalent intermediate is formed, then broken by the release of the 5' fragment of the DNA, leaving the tyrosine linked to the phosphate.

and zinc[54,56], although variation in the metal ion bound has drastic effects on activity[56,63]. The 'correct' metal ion used *in vivo* is unknown, although it is believed to be magnesium.

Despite giving the family its name, the HUH motif itself varies widely. As it is found in a β-sheet the central hydrophobic residue faces away from the active site, and thus has no role in catalysis; its generally hydrophobic character seems to be a matter of reinforcing the core of the protein. The relaxase-type protein MobA from plasmid R1162 has a cysteine in the central position, although close sequence homologues have an alanine in this position (data not shown), and the structurally similar TraI relaxase has a threonine[60]. The PCV2 Rep domain (fig. 2.1a) contains an unusual 'HLQ' sequence, resulting in a very low affinity ($K_d > 3\mu M$) for divalent metal[59].

The chemistry of the HUH reaction with DNA is relatively simple (fig. 2.2). The catalytic tyrosine acts as the nucleophile; the hydroxide group of the tyrosine side chain must be deprotonated in order to function. The target phosphate group of the DNA backbone is brought in close proximity to the catalytic tyrosine residue, which then attacks the phosphorous atom. After the formation of a theorized pentavalent intermediate the 5' fragment of the DNA is released, leaving the tyrosine linked to the phosphate.

The first HUH protein identified, gene A protein (gpA*), was found in bacteriophage ϕX174 from genetic mutants unable to produce ssDNA[64]. Isolation of the protein showed that it associates with target DNA at the 5' phosphoryl end through an 'apparently covalent'

attachment resistant to boiling or harsh chemicals[65]. This was also the first (and, until this work, only) HUH protein exploited for biotechnology. Akter et al. constructed a fusion of gpA* and an antibody-binding domain for use as the detection reagent in a 'sandwich' ELISA, where the target is detected by amplification of the DNA linked to the protein. Although limited, this provided an important proof of concept for the adaptation of isolated HUH domains.

Results

Selection of the initial test proteins

The initial HUH proteins selected for testing were identified from literature, focusing on proteins with known structures (Table 2.1). The set of six proteins included two from the Rep family and four from the relaxase family in order to test a range of proteins with potentially different properties. From this list, two homologs were identified from sequence databases, in order to test the feasibility of 'bioprospecting'[67] for new HUH proteins with desirable properties. In this case, we sought to find homologs with a lower isoelectric point in order to prevent non-specific ionic interactions. Using this method, we identified homologs of the viral Rep protein PCV2 and the relaxase RepB.

Table 2.1: Proteins studied in this work; see 4.4 for full details

Protein	***ori* sequence**
FBNYV	AAGTATTACCAGAAA
miniMobA	CCAGTTTCTCGAAGAGAAACCGGTAAGTGCACCCTCCC
NES	ACGCGAACGGAACGTTCGCATAAGTGCGCCCTTACGGGATTTAAC
PCV2	AAGTATTACCAGAAA
RepB	TGCTTCCGTACTACGACCCCCCA
TraI	TTTGCGTGGGGTGTGGTGCTTT
DCV (PCV2 homolog)	AAGTATTACCAGAAA
RepBm (RepB homolog)	TGCTTCCGTACTACGACCCCCCA

Biochemical characterization in vitro

The proteins were initially expressed in *E. coli* by standard methods, and purified to apparent homogeneity by metal-affinity followed by size-exclusion chromatography. All were monomeric and well-behaved in solution, and could be purified in high yield and stored at high concentration.

Because the nicking activity results in the formation of a covalent adduct the reaction could be studied by a simple SDS-PAGE gel-shift assay - covalent bonding causes an increase in mass, resulting in slower migration on a gel. This assay was used to determine the yield, rate, and specificity of the nicking/bonding reaction *in vitro*.

As previously demonstrated, all HUH proteins could form covalent adducts when incubated with divalent metal and ssDNA bearing their *ori* sequence; removal of the metal by EDTA chelation blocked adduct formation (fig. 2.3a.).

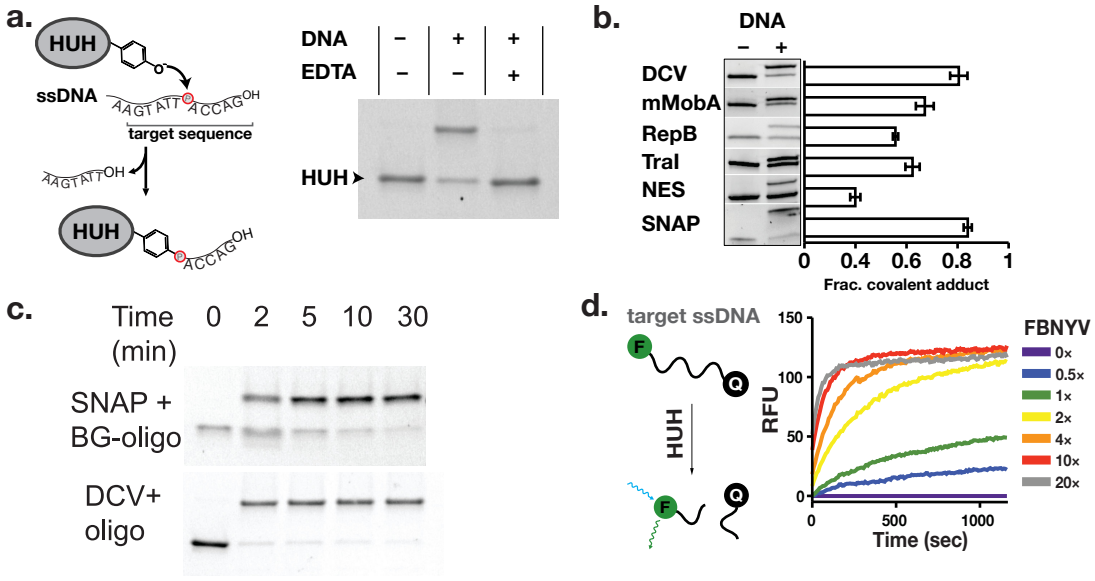

Figure 2.3: **Yield and rate of the HUH reaction** a. The HUH reaction and an example SDS-PAGE gel-shift result. b. Extent of reaction in 15 minutes at 37°C. c. Rate of the reaction and comparison to the SNAP-tag, assessed by gel-shift. d. Rate of reaction and nicking extent assessed by fluorescence unquenching.

The total extent of protein labeling was assessed by incubating each protein with a ten-fold excess of its target at 37°C for 15 minutes, and quantified the products by gel-shift (fig. 2.3b.). As a baseline we compared the HUH reaction to the reaction of purified SNAP-

tag protein, acquired from New England Biolabs (NEB), with a benzylguanine-labeled oligonucleotide. All the proteins showed appreciable labeling, with a range from 25 to 80%; only one protein, NES, had a yield below 50%. The viral Rep proteins showed the highest yield, comparable to the commercial SNAP-tag. It is worth noting that the SNAP-tag, like most commercialized proteins, has been extensively modified to increase its activity[68,69], while the HUH proteins have not.

The gel-shift assay was then used to determine the rate of the HUH reaction, again compared to the SNAP-tag. Proteins were incubated at 37°C with a ten-fold excess of target oligonucleotide and quenched with SDS-PAGE loading buffer at the indicated times (fig. 2.3c). The reaction of the Rep protein with target DNA is extremely fast, reaching near-completion within two minutes, compared to the SNAP-tag's time of 15-30 minutes. The reactions of the relaxase proteins were slower but still occurred within the 15-30 minute time scale (fig. 2.9).

For greater precision and higher throughput we used a fluorescence unquenching assay to measure the rate of reaction in 96-well format. In this assay, the target *ori* sequence is flanked by a fluorescent fluorescein dye and a dark 'quencher' molecule, which absorbs the energy from a nearby fluorophore but does not emit any fluorescence of its own (fig. 2.3d). Initially the fluorescein emission is absorbed by the quencher, preventing fluorescence; upon reaction, the DNA is split and the molecules move apart, allowing detection. This assay also allowed us to measure the total extent of DNA cleavage, as opposed to the amount of protein labeling. The results confirmed the gel-shift assay findings - the Rep reaction is extremely fast. At high protein:DNA ratios, the reaction reaches completion in under five minutes; a reaction ratio as low as 2:1 could reach completion, albeit at a slower rate.

We used this assay to explore the optimal conditions for HUH-endonuclease activity, finding an optimal pH range of 7-8 for both classes of HUH, and determined that common additives such as reducing agents (5mM β-mercaptoethanol (BME)), calcium, BSA (5% w/v), and surfactant (0.05% v/v Triton X-100) do not appreciably affect the enzymatic activity (fig. 2.9).

A major advantage of the HUH labeling reaction is that it targets specific DNA sequences, and thus could be used for 'one-pot' reactions of multiple proteins. We tested the sequence specificity of the labeling reaction by incubating each protein with a ten-fold ex-

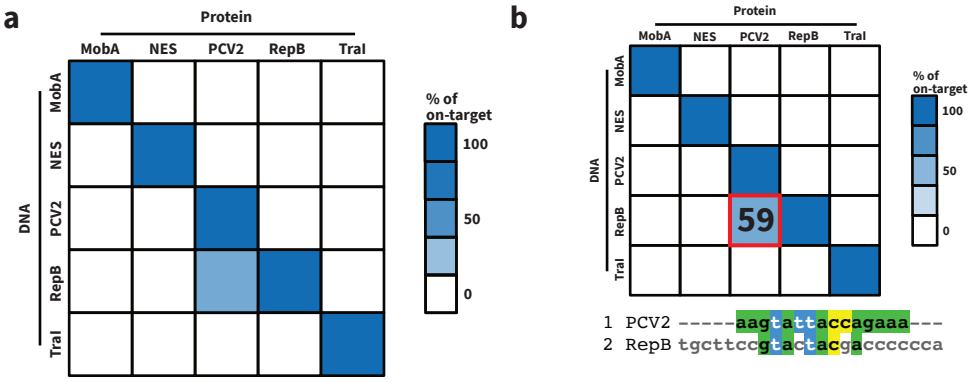

Figure 2.4: **Sequence specificity of the HUH reaction** a. Heatmap showing reactivity of each protein on each DNA target; normalized as a percentage of the 'correct' reaction. b. PCV2 protein cross-reacts with RepB target, but not vice-versa.

cess of every target DNA, and quantified the products by gel-shift. Figure 2.4 displays this data in heat-map form; each reaction is normalized as a percentage of the reaction with the correct target (fig. 2.11). The results show that the HUH proteins have excellent specificity, with one important exception. None of the relaxase-type proteins showed any detectable cross-reactivity in this assay, while the Rep-family PCV2 reacted to a large extent with the highly similar RepB target sequence (fig. 2.4b). Curiously, the opposite was not true - RepB protein had no activity on PCV2 target. This suggests that the non-specific reactivity is a feature of the Rep family proteins. This establishes an important distinction between the two families: the relaxase family is highly specific but less reactive, while the Rep proteins are highly reactive but with lower specificity. Whether these features are related - ie, if Reps require low specificity to achieve high yield - is unknown, and merits further investigation.

These results show that the HUH proteins can match or exceed a commonly used conjugation protein for bioconjugation *in vitro*, with greater versatility and without the need for chemical labeling of the DNA.

Applications of purified HUH-fusion proteins

As a proof-of-concept for this strategy, we prepared a recombinant protein comprised of tandem HUH-tags linked by a protein sequence containing an MMP2 proteolytic cleavage site (fig. 2.8c). We then added two oligos with distinct target sequences containing Cy3 or

Cy5 dyes. Fluorescent imaging shows robust dual labeling of the protein, and treating the protein with MMP2 protease results in splitting of the dual-tagged protein into two single tagged proteins.

The next example we thought of is Cas9, which is routinely delivered to cells as a recombinant protein via transfection or electroporation as a ribonucleoprotein (RNP) with in vitro transcribed guide RNA[70]. We reasoned that we could use the HUH-tag to specifically label the RNP with organic fluorophores for quantitation purposes or for potential multiplexed covalent labeling of genomic loci with an organic fluorophore or other modification that can be encoded on an oligo. We first tested that the recombinant PCV-Cas9 (fig. 2.13a) was functional by measuring its ability to cleave and disrupt GFP fluorescence in a cell line stably expressing GFP (fig. 2.8b). The number of GFP-expressing cells was reduced by approximately 80% for both Cas9 and PCV-Cas9 (fig. 2.13c-d). In addition to exhibiting similar levels of GFP knockdown, the PCV fusion tag allowed us to also codeliver a Cy5-labeled oligo into the cells (fig. 2.8b). Interestingly, while the Cy5-oligo tethered to Cas9 was most efficiently delivered into cells when delivered with cationic lipid, an appreciable amount of the RNP-oligo complex also entered the cells even in the absence of cationic lipid, suggesting that the additional positive charges from the PCV enhance its ability to cross the cell membrane (fig. 2.13b).

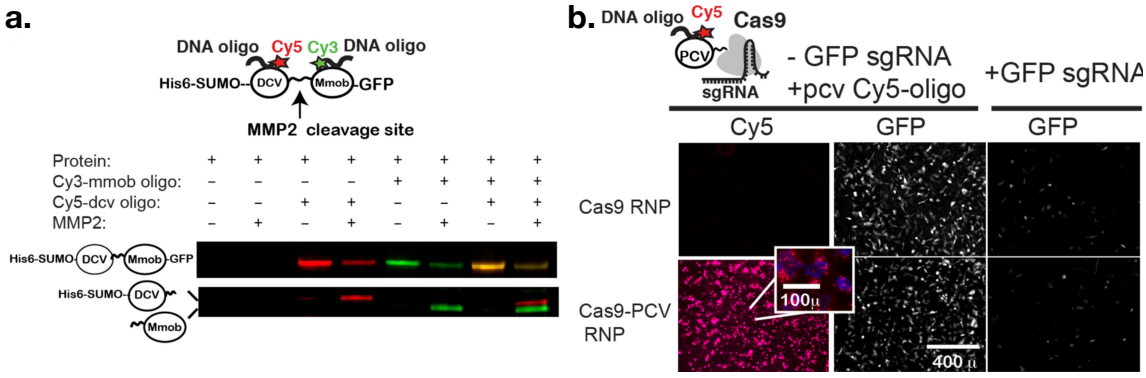

Figure 2.5: **Applications of purified HUH-fusion proteins** a. Construction of a biosensor for matrix-metalloprotease 2 using the DCV and mMobA HUH proteins. b. Labeling of Cas9 protein for genome labeling and editing.

As a final *in vitro* test, we attempted to react the HUH proteins with both small and large DNA nanostructures functionalized with HUH target sequences (fig. 2.14). The puri-

fied protein was reacted with the structure and the products analyzed by agarose gel electrophoresis in standard 0.5×TBE buffer supplemented with 0.1% SDS to denature the protein; addition of the protein caused a mass shift in the DNA. These results are in line with other groups' published findings[71,72], which were published both as this work was in preparation for publication (Sagredo et al.[71]) and afterwords (Bernardinelli and Högberg[72]).

Use of HUH fusion proteins in cultured mammalian cells

We initially transfected human U2OS cells with vectors constitutively expressing HUH proteins to identify which could be correctly expressed. Standard preparation of cell lysates followed by incubation with a 100nM solution of 3'-TAMRA target oligonucleotides and Mn^{2+} showed labeling of mMobA and TraI by fluorescent SDS-PAGE (fig. 2.15). Later experiments with fusion proteins showed that the RepB homolog could also be expressed (named RepBm, for its expression in mammalian cells).

To assess the use of HUH-tags for labeling in fixed cells and effects on cellular localization, we fused TraI and mMobA to the N-terminus of human β-actin and expressed the constructs in U2OS cells. Labeling the fixed TraI/mMobA-β-actin cells with 3'-Alexa647 target oligos showed expected labeling of both actin filaments and cytoplasmic actin; counterstaining with phalloidin-Alexa488 showed that the fusion protein was efficiently incorporated into actin filaments (fig. 2.6). Control cells, transfected with EGFP-β-actin and mock labeled with either fluorescent target, showed no fluorescence in the far-red region, indicating that nonspecific interactions are not responsible for labeling (fig. 2.16).

We next tested HUH-tags in labeling proteins on the surface of live cells. We first fused mMobA and RepBm to the N-terminus of cell-surface Notch receptors also containing a FLAG-tag. Both fusions exhibited good cell-surface trafficking in U2OS cells and normal capability to signal in comparison to a SNAP-fused Notch receptor, as shown by labeling the FLAG-epitope tag with an APC-conjugated antibody and in luciferase signaling assays (fig. 2.17). Treating the cells expressing HUH-fusion tags with magnetic beads coated with oligos bearing the target sequence (2.7a.) only labeled the HUH-tag containing receptors. To test the orthogonality of labeling HUH-tags in the cellular context, we constructed a synthetic transmembrane receptor with both mMobA and RepBm in the extracellular domain (2.7b.). Addition of fluorescent oligonucleotides to the culture media, as well as Mn^{2+} and salmon sperm DNA to block non-specific interactions, resulted in fast and efficient label-

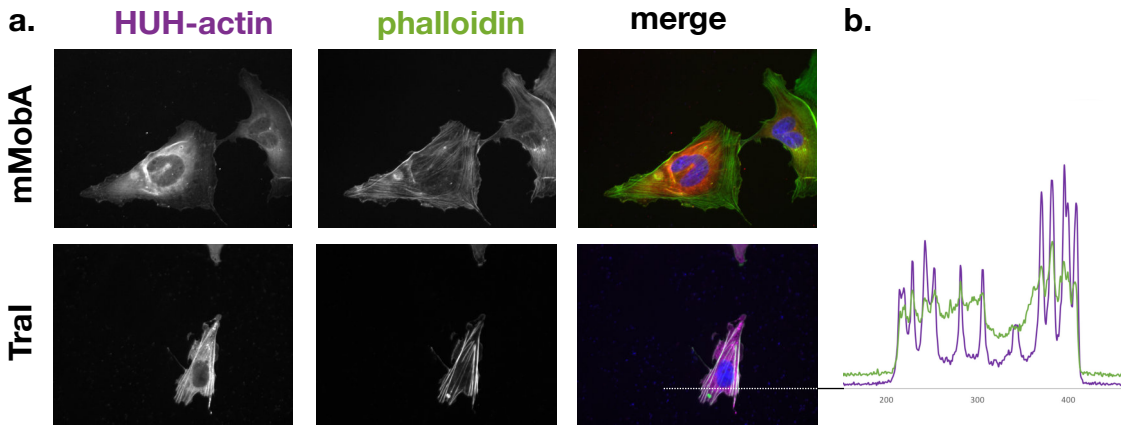

Figure 2.6: **Labeling of fusion proteins in mammalian cells** a. Labeling of mMobA and TraI-fused actin in fixed U2OS cells using fluorescent oligonucleotides; the counterstain phalloidin specifically labels actin filaments. b. Linescan of the merged TraI images showing colocalization of actin filaments.

ing of the cell surface. Cells expressing receptors containing only one HUH protein were labeled only by the correct target sequence, and counterstaining with α-HA-Dylight488 showed that untransfected cells were not labeled (fig. 2.18). Time course imaging of the labeling reaction showed that visible labeling was achieved within ten seconds, and the number and intensity of labeled cells increased to a maximum after 15-20 minutes (fig. 2.19).

Live-cell fluorescent labeling of intracellular targets is hampered by the fact that oligonucleotides do not freely travel into cells. We performed proof-of-concept live-cell labeling of cells transfected with TraI-β-actin by delivering its fluorophore-conjugated target oligo using cationic lipid (fig. 2.20). Delivery of the fluorescent oligo to mock-transfected cells showed very little fluorescent signal, while the transfected cells were clearly labeled in the cytoplasm. Intracellular labeling could be further optimized by exploring alternative methods to deliver oligos into cells.

Conclusions

We have shown that HUH endonucleases can be adapted into useful fusion tags for DNA-protein conjugation. They efficiently form covalent bonds with a specific DNA sequences,

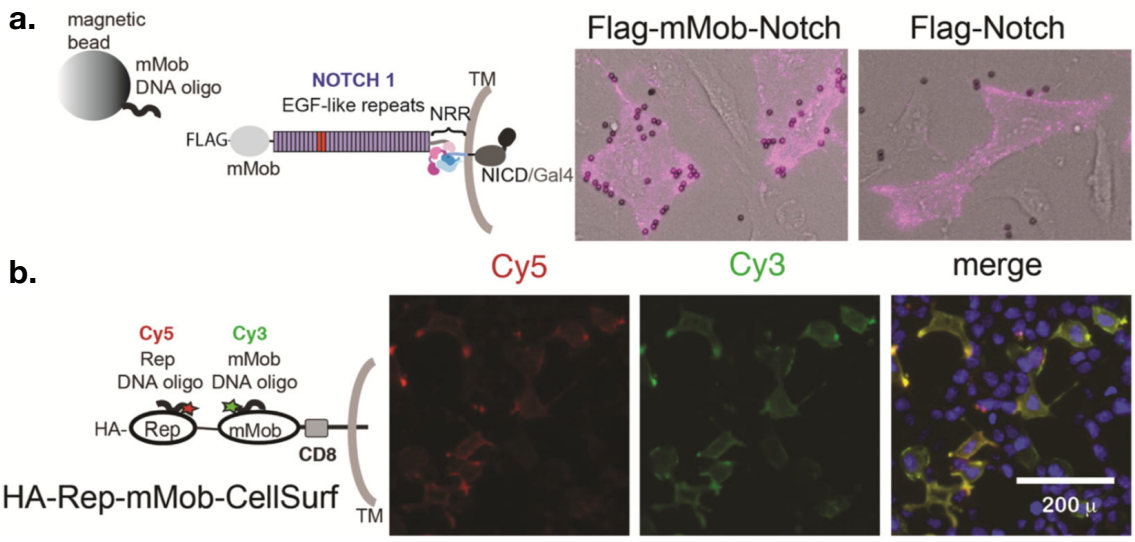

Figure 2.7: **Labeling of cell surface proteins in live cells** a. Labeling of Notch receptor fusions with magnetic beads. b. Multiplex labeling of a synthetic mMobA-RepBm fusion protein on the cell surface.

eliminating the need for modified DNA and allowing multiplex 'one-pot' labeling. We have demonstrated a variety of applications *in vitro* and in cultured cells, using N- and C-terminal fusions in the cytoplasm and cell surface without disturbing protein function. The reaction is compatible with a variety of *in vitro* conditions, standard cell culture media, and cellular lysis and fixation.

We believe that this technology could greatly expand the protein-labeling toolkit in many applications. The small size of the labeling reagent, relative to antibodies, could be very useful for super-resolution imaging, especially in combination with small recombinant affinity reagents such as nanobodies or affimers. Because they are active in cell lysate HUH fusions could be directly immobilized or labeled without further purification, allowing faster and more informative experiments.

Finally, the fact that the Rep family proteins can achieve complete labeling suggests that the other, more specific relaxases could be engineered, either rationally or by directed evolution, for higher activity. The SNAP-tag used as our 'benchmark' labeling protein was extensively engineered before becoming commercially available; the fact that the HUH proteins can meet or exceed its effectiveness before modification is extremely promising

for their future possibilities.

Current and future work: Cas9-HUH fusions for gene editing

Reproduced with permission from Aird et al. Comms. Bio. 2018[73]
©2018 Springer Nature Limited

A number of projects incorporating HUH fusion proteins are ongoing in the Gordon lab. The most advanced, led by graduate student Eric Aird, is highlighted here; a full reprint is included in 4.4. I began this project during the initial work on the HUH proteins; I developed the basic techniques for purifying and working with Cas9 and fusions *in vitro* and in cells and provided support to the work. Credit for the final product is due entirely to Eric, who lead the team that brought the work to fruition.

The ability of CRISPR-Cas9 to specifically target and cleave sites in the genome to produce a double-stranded break (DSB) has made it a critical tool in genome editing[74,75]. Mammalian cells repair the DSB predominantly through two pathways: non-homologous end joining (NHEJ) or homology-directed repair (HDR). The more frequent NHEJ pathway results in the formation of small insertions or deletions (indels) at the DSB site, while the alternative HDR pathway can be utilized to insert (or 'knock-in') exogenous DNA sequences into the genome[76]. For genome-editing HDR is frequently desired but is limited by the low efficiency of recombination.

It has long been known that single-stranded DNA is an efficient substrate for genomic incorporation at break sites[77]. Given that we have previously shown that Cas9-PCV2 fusions can be efficiently labeled with oligonucleotides[52], we sought to extend this method to tethering of an HDR template DNA to the protein, thus co-localizing the donor to the break site and potentially increasing the efficiency of incorporation.

We tested the activity of fusions of the Rep protein PCV2 to Cas9 at the N- (PCV-Cas9) and C-termini (Cas9-PCV). To assess reactivity of the fusions with long ssODNs suitable for HDR insertion, the formation of the protein-DNA conjugate was assessed using SDS-PAGE, as previously described. An equimolar ratio of either protein to DNA resulted in ¿60% covalent complex formation (by gel densitometry) after 15 min at room temperature. The ability of Cas9 to cleave DNA was assessed *in vitro* and found to not be perturbed by the addition of PCV at either terminus.

For in-cell testing we used a recently-described luminescence activity assay (Promega HiBit) to monitor the in-frame insertion of a 13 amino acid split-luciferase complement peptide (HiBiT) into endogenous loci. Adding the recombinant N-terminal nanoluciferase

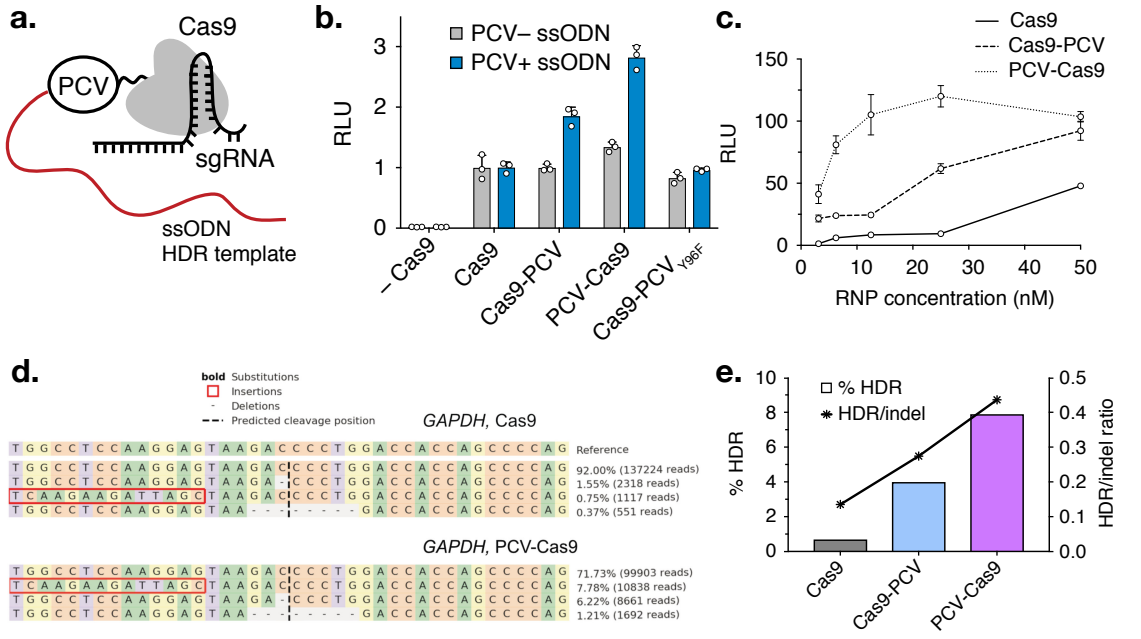

Figure 2.8: **Cas9-HUH fusions increase genome editing** a. The fusion protein used in this work. b. Knock-in of a luciferase complement peptide using Cas9 and Cas9-HUH fusions. c. Luciferase peptide knock-in as a function of transfected protein complex. d. Deep-sequencing results from knock-in of luciferase peptide at the endogenous *GAPDH* locus in HEK293T cells. e. Bar graph of data from d. with addition of the HDR/indel ratio, a measure of editing precision.

fragment (LgBiT) reconstitutes activity and produces light upon addition of substrate; the intensity of light emitted is thus a relative measure of HDR efficiency. RNPs targeting the 3?-end of *GAPDH* were assembled *in vitro* and transfected into HEK-293T cells with or without ssODN containing the PCV recognition sequence (fig. 2.8b). When using an ssODN lacking the PCV recognition sequence (PCV- ssODN), all versions of Cas9 resulted in similar luminescence levels when assayed 48h post-transfection. However, upon addition of an ssODN containing the PCV recognition sequence (PCV+ ssODN), a substantial two to three-fold change in luminescence is observed for cells transfected with Cas9-PCV fusions. The increase in luminescence was not observed when catalytically inactive PCV (Y96F) was fused to Cas9, suggesting the specific attachment of ssODN to Cas9 via PCV is responsible for the increase in HDR frequency.

The increased editing is even more pronounced when transfecting small amounts of RNP (fig. 2.8c). At 25nM RNP concentration the fusions provided a 15- to 30-fold increase in peptide insertion. This is very encouraging for therapeutic applications of this technique, where protein delivery is tightly limited.

To directly assay the insertion of the HiBit at an endogenous gene we targeted the *GAPDH* locus in HEK293T cells and analyzed the results by deep sequencing. Use of the fusion proteins gave a 5- to 11-fold increase in the specific insertion of the peptide compared to non-fused Cas9 (fig. 2.8d); with PCV-Cas9 this brought the absolute editing rate to 7.78%, compared to 0.75% using Cas9. Fusion also increased the ratio of HDR to indel formation (2.8e), indicating that these results are due to shifting the repair process to favor HDR, rather than a general increase in editing.

These results are highly encouraging, and have been the subject of great interest by outside groups, judging by requests for the protein and plasmids. A similar approach was recently published using the SNAP-tag, which gave similar results[78]. As noted before, using the SNAP-tag requires chemical modification of the donor DNA, which adds difficulty and cost and could have deleterious effects if used *in vivo*. Our method, which requires only an addition of 15bp to the ssODN, could have wide application both in basic science and medicine.

Experimental procedures

Materials All coding sequences were obtained as codon-optimized synthetic DNA from Life Technologies or Integrated DNA Technologies (IDT). Staple oligonucleotides for the DNA origami construct were purchased from Life Technologies. All other oligonucleotides were purchased from IDT. All restriction enzymes were purchased from NEB. Electrophoresis supplies were purchased from Bio-Rad unless otherwise specified.

HUH proteins Linear coding DNA was inserted into vector pTD68/6×HIS-SUMO at the BamHI and XhoI sites using restriction-based ligation, In-Fusion (Clontech), or Hi-Fi DNA Assembly (NEB). Sequenced constructs were transformed into *Escherichia coli* (*E. coli*) BL21(DE3) cells and grown in lysogeny broth (LB) supplemented with 100µg/mL ampicillin; at OD_{600} 0.8 the cells were induced with 500µM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and allowed to express for 3h at 37°C or overnight at 18°C. The cells were harvested by centrifugation at 4,000×G, the pellet resuspended in 50mM Tris-HCl, pH 8.0, 350mM NaCl, 5mM BME, 10mM imidazole, and lysed by sonication on ice. Insoluble material was removed by centrifugation at 24,446×G, 4°C, for one hour. Soluble protein was batch-bound to Nickel-NTA agarose (Thermo Scientific), washed with five column volumes of 50mM Tris-HCl, pH 8.0, 1M NaCl, 5mM BME, 10mM imidazole, and eluted in lysis buffer containing 250mM imidazole. The purified protein was dialyzed overnight against 50mM Tris-HCl, pH 8.0, 350mM NaCl, and 5mM BME or directly concentrated for further purification. To remove the 6×HIS-SUMO fusion tag, HIS-tagged Ulp1 protease (gift of Amanda Hayward, produced in-house) was included in the dialysis bag and incubated overnight at 4°C; the protease and 6×HIS-SUMO were then removed by running the solution over nickel-NTA agarose before further purification. Proteins were purified by size-exclusion chromatography using a Bio-Rad SEC650 column equilibrated in 50mM Tris-HCl, pH 8.0, 200mM NaCl, 2mM EDTA. All concentration and buffer-exchange procedures were performed using GE Vivaspin columns, 10K MWCO.

Cas9 proteins Cas9 was expressed from the plasmid pET15/SP-Cas9 (a gift from Niels Geijsen[79], Addgene plasmid #62731). PCV-Cas9 was constructed by inserting the Cas9 coding sequence between the 6×HIS-SUMO and PCV2 proteins at the BamHI site in pTD68/6×HIS-SUMO-PCV2 using In-Fusion cloning. Cas9 and Cas9-PCV2 were puri-

fied according to the method of Anders & Jinek[80]. Proteins were expressed in *E. coli* BL21(DE3) grown in autoinduction media[81] for approximately 8 hours at 37°C, then shifted to 25°C for 24 hours. Cells were collected by centrifugation at 4,000×G, resuspended in 20mM Tris-HCl, pH 8.0, 500mM NaCl, 5mM imidazole, 0.4mM AEBSF, and lysed by sonication on ice. Insoluble material was removed by centrifugation at 24,446×G, 4°C, for one hour. Soluble protein was bound to Ni-NTA agarose, washed with 15 column volumes lysis buffer, and eluted in 20mM Tris, pH 8.0, 500mM NaCl, 500mM imidazole. The eluted protein was dialyzed against 20mM HEPES, pH 7.5, 150mM KCl, 1mM EDTA, 1mM Dithiothreitol (DTT), 10% glycerol overnight at 4°C. The 6×HIS-SUMO tag of Cas9-PCV2 was removed with recombinant Ulp1 protease during dialysis, as described above. Dialyzed protein was bound to a HiTrap SP HP cation-exchange column (1mL, GE Healthcare) equilibrated in 20mM HEPES, pH 7.5, 100mM KCl, and eluted with a linear gradient from 0.1-1M KCl. Purified protein was snap-frozen in aliquots.

SDS-PAGE of reactions between HUH-tags and ssDNA oligos Unless otherwise noted, gel-shift assays were performed in 50mM HEPES, pH 8, 50mM NaCl, 1mM $MgCl_2$ and 1mM $MnCl_2$ (HUH buffer), incubated at 37°C for 15 minutes, and stopped with 4× Sodium dodecyl sulfate (SDS) loading buffer. The reactions were analyzed by electrophoresis on 4-20% polyacrylamide gels in Tris-Glycine-SDS buffer and stained with either Coomassie Blue or Bio-Rad Stain-Free gels. For comparison of covalent adduct formation of SNAP and DCV, 25pmol of SNAP/DCV proteins were mixed with 100pmol respective DNA-oligo in 50mM HEPES, pH8, 50mM NaCl, and 5mM BME or HUH buffer, respectively; 4× SDS loading buffer was added at indicated times to stop the reaction. Specificity reactions of HUH proteins with each target oligo were performed in HUH buffer with 150mM NaCl. For one-pot reaction of double-HUH tagged protein, HUH reaction was performed as previously, and reactions were divided in half. To one half (minus MMP-2), MMP assay buffer was added (50mM Tris-HCl, pH 7.5, 10mM $CaCl_2$, 150mM NaCl, 0.05% (w/v) Brij-35). To other half, 36.7ng activated MMP-2 was added. 220ng of MMP-2 was activated using 2 mM APMA in MMP assay buffer for 1 hour at 37 C. Reactions were stopped with 4× SDS loading buffer and run on 4-20% SDS-PAGE. Fluorescence was detected with a Typhoon FLA9500 imager (GE), using the standard Cy3 and Cy5 channels.

Fluorescence unquenching assays Oligonucleotides were purchased with a 5' quencher and 3' fluorescein or Cy3 and dissolved at 100μM in water. Oligos were diluted to designated concentration (125-500nM) in water and 50μl was added to wells in black 96-well plates (Corning) Proteins were dissolved at designated concentration in desired buffer, and 50μl added to wells containing fluorophore-quencher oligo. Fluorescence of fluorescein or Cy3 was measured on a fluorescence plate reader (Molecular Devices, Gemini). For experiments using different buffers, each trace was corrected for fluorescence of oligo alone in designated buffer.

Oligonucleotide labeling Amino-modified oligonucleotides were obtained with standard desalting purification and resuspended in MilliQ water to 200μM concentration. N-hydroxysuccinimide (NHS) ester dyes (Life Technologies) and resuspended to 10mg/mL in anhydrous dimethyl sulfoxide (DMSO). Labeling was perfomed by mixing 20μL dye solution, 20μL DNA, 20μL 0.5M HEPES, pH 8.5, and 40μL water and incubating the mixture overnight at room temperature. Excess dye was removed by repeated ethanol precipitation and purification using G-50 spin columns (IBI Scientific). The SNAP substrate was prepared as above using an amino oligo and the NHS-ester of benzylguanine (NEB). The reaction was purified on a DNA-Pac anion-exchange column and concentrated using 3k MWCO centrifigal filters.

DNA origami labeling The six-helix bundle construct was designed using CadNano2[82]. Staple strands were mixed at 10-fold excess with 10nM m13mp18 scaffold in TEM Buffer (10mM Tris-HCl, pH 8.0, 1mM EDTA, 10mM $MgCl_2$) and folded by cooling from 80°C to 60°C over 80 minutes, then 60°C to 24°C over 15 hours. Excess staples were removed by diluting the reaction ten-fold in TEM buffer and concentrating it using 100K MWCO columns (AmiCon) spun at 1,000×G, with two changes of buffer. 1nM six-helix bundle was incubated with 10-fold excess of the selected proteins under standard reaction conditions. The products were analyzed on 2% agarose in 0.5X TBE + 11mM $MgCl_2$ and stained with SYBR Safe (Invitrogen).

Mammalian vector construction Constitutive expression vectors (denoted pcDNA3_NAME) were constructed by inserting the coding sequence into the BamHI site of pcDNA3 (Invitrogen) using Hi-Fi DNA Assembly (NEB). Actin vectors were constructed by inserting

the coding sequence of human β-actin into pcDNA3_mTraI36 and pcDNA3_mmobA using BamHI and XhoI, to create a C-terminal in-frame fusion. For cell-surface fusions, existing Flag-Notch1-Gal4 Notch vectors[83] were cut with KpnI between the FLAG-tag and EGF-1, and the codon-optimized HUH-tag was inserted by In-Fusion cloning. The tandem HUH-tag cell surface receptor, called HA-Rep-mMob-CellSurf, was cloned by inserting DNA fragments containing protein G[84], RepBm, a 40 amino acid linker, mMobA, and a CD8 domain into an existing vector encoding a portion of the DLL4 cell surface receptor[83] and an N-terminal HA-tag and FKBP domain in place of the FKBP domain.

Cell lysate labeling HEK293T cells were grown in DMEM + 10%FBS to 90% confluency in 12-well plates and transfected with 1µg pcDNA3 vector using Lipofectamine 3000 (Life Technologies). Transfected cells were grown for 48 hours before being lysed with 300µL Pierce IP Lysis Buffer (Thermo Scientific) according to manufacturer's instructions. 10µL of cell lysate was incubated at 37°C for 30 minutes with 1µL TAMRA-labeled target DNA with or without the addition of 20mM $MgCl_2$ and $MnCl_2$. The reactions were separated by SDS-PAGE and imaged using a Typhoon FLA9500 imager.

Fixed cell labeling U2OS cells were grown either on glass coverslips in 6-well dishes or 12-well chambered coverglass (MatTek) at 37°C with 5% CO_2. At 30-50% confluence the cells were transfected using Lipofectamine 3000 as previously described. After 24 hours of expression the cells were fixed and permeabilized by the following protocol: 15 minute fixation in 4% paraformaldehyde (Thermo Fisher) in CBS (10mM MES, pH 6.1, 138mM KCl, 3mM $MgCl_2$, 2mM EGTA, 0.32M sucrose), 3×2 minute wash with TBS + 0.3M glycine, permeabilization with Permeabilization/Blocking buffer (TBS + 0.025% saponin, 1% BSA, 5mM $MgCl_2$, 5mM $MnCl_2$), 30 minute labeling by addition of 100nM Alexa 647 oligo to the permeabilization buffer, 2×3 minute wash with TBS + 0.5M NaCl, 1 minute wash with TBS + 2 drops of NucBlue Fixed-Cell Stain (Life Technologies), mounting in Slowfade Diamond (Life Technologies). Cells were imaged on an EVOS FL-AUTO microscope. Analysis and contrast adjustment was performed in ImageJ[85].

Live cell surface labeling U2OS or HEK293T cells were transiently transfected with full-length Notch receptors harboring an N-terminal FLAG-tag plus mMobA, RepBm, or SNAP fusion tag and intracellular Gal4 fusion (for transcriptional assays[86]), or tandem cell sur-

face construct HA-Rep-mMob-CellSurf, in 96-well plates as previously described. 100ng of plasmid was used per well, or 10ng for reporter assays. 24-48 hours later, cells were washed 2x with PBS, and labeling solution added. Standard labeling solution used a base of standard DMEM + 10% FBS, 1% penicilling/streptomycin, 0.5mM $MnCl_2$, 0.5mM $MgCl_2$, 1:20 Salmon Sperm DNA, and 200-250nM fluorescent oligonucleotide. APC-anti-Flag or Dylight488-anti-HA (Pierce) were added as required at 1:750. Reactions were performed at 37°C for 20 minutes. Cells were washed twice with PBS and media was replaced with Fluorobrite DMEM (Life Technologies)+ 10% FBS + 2ug/ml Hoescht. Cells were imaged on an EVOS FL-AUTO microscope. Analysis and contrast adjustment was performed in ImageJ[85]. Luciferase assays were performed by co-transfecting luciferase reporter plasmids, and plating cells in wells coated with 10µg/ml Jagged-1 (R&D Systems). Cells were lysed and Dual Luciferase Assay (Promega) was performed according to manufacturer's instructions.

Guide RNA preparation The EGFP sgRNA was synthesized using an EnGen®sgRNA Synthesis Kit, *S. pyogenes* (NEB), with the primer GFP_AS_sgRNA_1, or *in vitro* transcribed from a PCR product derived from a PCR template kindly provided by Eric Hendrickson. The product was purified by ethanol precipitation

GFP targeting and analysis GFP knockdown comparison of Cas9 and PCV2-Cas9 was performed in a doxycycline-inducible GFP cell line, a kind gift from Peter Gordon. GFP was inserted into the pLVX-TetOne vector (Clontech) by In-Fusion cloning and the resultant vector used to generate lentiviral particles. HT1080 cells (ATCC) were transduced with lentiviral particles for 48 hours and then selected with puromycin in order to generate doxycycline-inducible GFP HT1080 cells. Cells were seeded to approx. 70% confluency in clear-bottom 96-well plates. 10 pmol of Cas9 or PCV2-Cas9 were mixed with 50pmol 3' Cy5-labeled PCV2 target oligo for 5 minutes in Opti-MEM supplemented with 1mM $MgCl_2$. Reactions were divided in half and water or 10pmol GFP sgRNA added for 10 minutes at room temperature. Reactions were divided in half and 0.5µl RNAiMax (Life Technologies) added to half of the reactions for 15 minutes. RNP/liposome mixes were then added to cells in antibiotic-free DMEM + 10% FBS. 12 hours later, 1µM doxycycline was added to the wells. Cells were imaged on an EVOS FL-AUTO microscope 4-10 hours later. Analysis of GFP and Cy5 intensities was performed in ImageJ[85].

Supplemental material from Lovendahl et al. *JACS* 2017

Reproduced with permission from Lovendahl et al. J. Am. Chem. Soc. 2017[52]

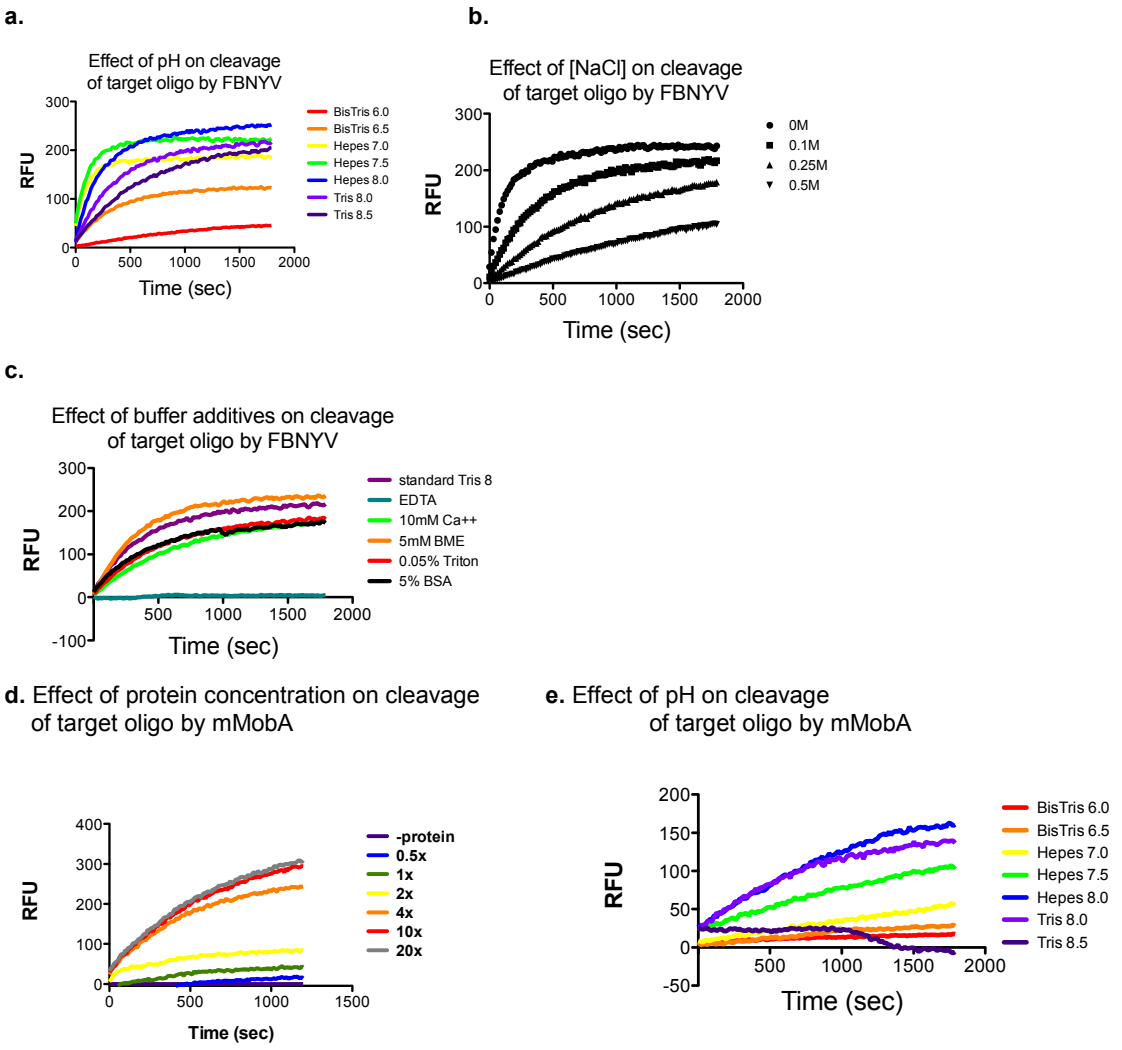

Figure 2.9: **Testing optimal conditions of FBNYV (a-c) and mMobA (d, e) nicking reaction using a molecular beacon** a. Buffers of varying pH: 100mM designated buffer, 100mM NaCl, 1mM Mn^{2+}, 1mM Mg^{2+}. b. Buffers of varying [NaCl] in 100mM Tris 8, 1mM Mg/Mn. c. Buffer additives in 100mM Tris-HCl, pH 8, 100mM NaCl, 1mM Mg/Mn. d. Cleavage rates of DNA oligo as a function of excess protein. All traces resulted from subtracting a blank containing no added MmobA. Buffer was 50mM Tris-HCl, pH 8, 100mM NaCl, 1mM Mn^{2+}, 1mM Mg^{2+} e. Buffers of varying pH: 100mM designated buffer, 100mM NaCl, 1mM Mn^{2+}, 1mM Mg^{2+}. Each trace resulted from subtracting Buffer+oligo control from Buffer + oligo + HUH-protein.

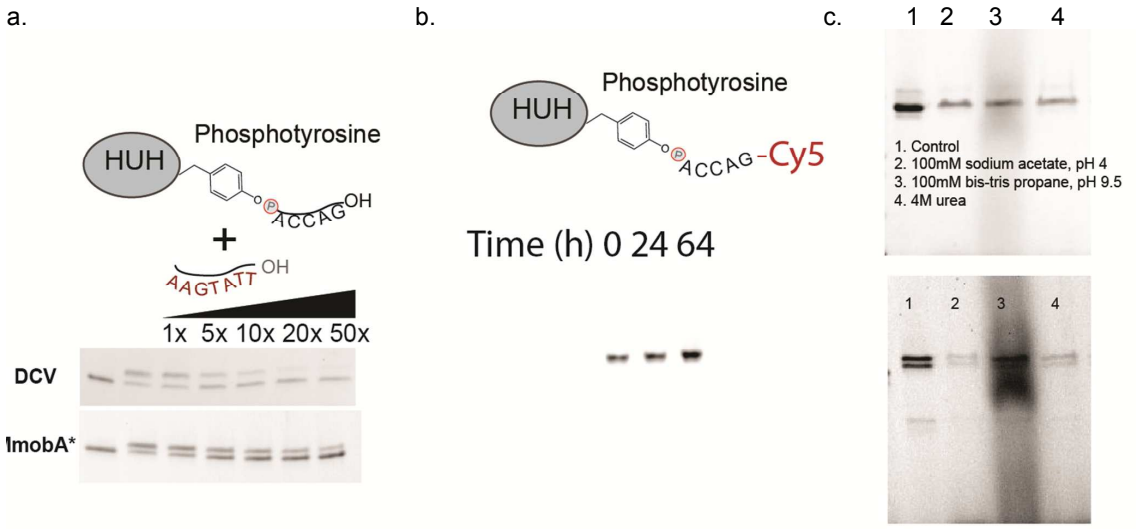

Figure 2.10: **Testing stability and reversibility of HUH-DNA conjugates.**
a. 10pmol of SUMO-DCV or SUMO-mMobA was reacted with 100pmol of target DNA in standard HUH-conditions for 15 minutes at 37°C. Oligos with free 3'-OH corresponding to the product of HUH-endonuclease activity that is not covalently attached were added in increasing amounts of molar excess with respect to HUH protein to test reversibility of protein-DNA conjugates formed in vitro. Both conjugates can be reversed at high molar excess of free 3'-OH containing oligo, which could be useful for some applications. Quantities of free 3'-OH containing oligo on the order of the stoichiometry of the HUH protein ($\approx$1X) do not reverse the protein-DNA conjugates.
b. SUMO- PCV was reacted with 10-fold excess of its target oligo conjugated to Cy5 under standard HUH conditions. The reaction was then mixed with 10-fold excess of conditioned media from HEK293T cells to simulate physiologic conditions. Aliquots were quenched at several timepoints and the presence of the HUH-oligo conjugate visualized by Typhoon imaging an SDS-PAGE gel. The band representing the intact PCV-oligo-Cy5 can only be present if the protein and oligo are both intact. There is no significant degradation after 64 hours.
c. PCV and mMobA-DNA conjugates were treated to pH 4, pH 9.5, 4M urea for 1 hour, with not significant loss of protein-DNA conjugate.

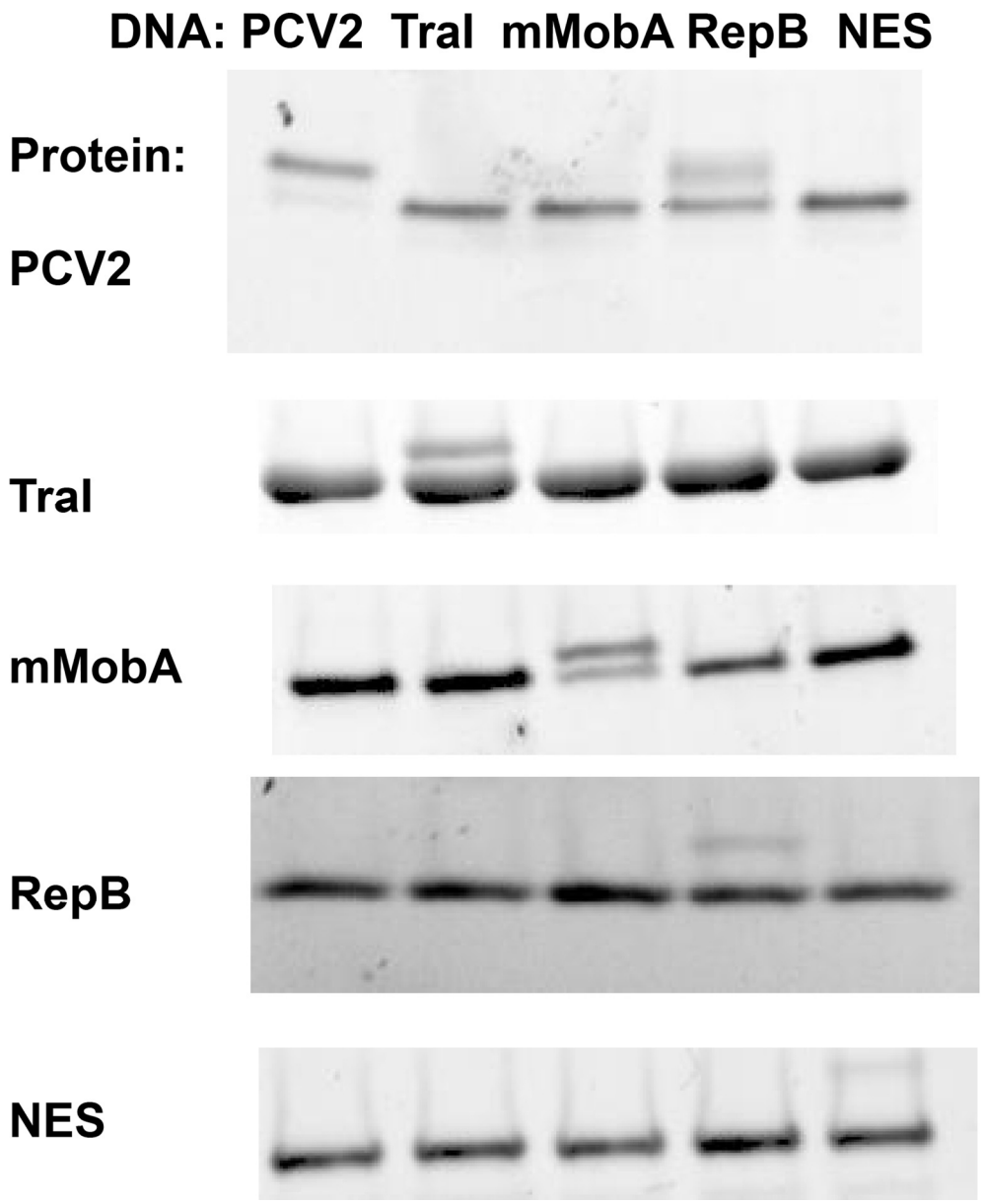

Figure 2.11: **Full gels used for yield calculations**
HUH-reactions were performed as described in Methods. Gel bands were quantitated using ImageJ.

a. Sequences of oligos:

```
mmob0       CCAGTTTCTCGAAGAGAAACCGGTAAGTGCG -CCCTCCC
mmob1       CCAGTTTCTCGAAGAGAAACCGGGAAGTGCG -CCCTCCC
mmob2              CTCGAAGAGAAACCGGTAAGTGCG -CCCTCCC
mmob3       CCAGTTTCTCGAAGAGAAACCGGTAAGTGCA -CCCTCCC
mmob4       CCAGTTTCTCGAAGAGAAACCGGTAAGTGCG -GCCTCCC
mmob5:      CCAGTTTCTCGAAGAGAAACCGGTAAATGCG -CCCT
mmob6:      CCAGTTTCTCGAAGAGAAACCGGTAAATGCG -CCCTTATAAGCGGGAGATTCGTCCTCATA
Trai:                        TTTGCGTGGGGTGT -GGTGCTTT
NES1:          ACGCGAACGGAACGTTCGCATAAGTGCG -CCCTTACGGGATTTAAC
NES2:          ACGCGAACGGAACGTTCGCATAAGTGCA -CCCTTACGGGATTTAAC
NES3:                      GTTCGCATAAGTGCG -CCCTTACGGGATTTAAC
```

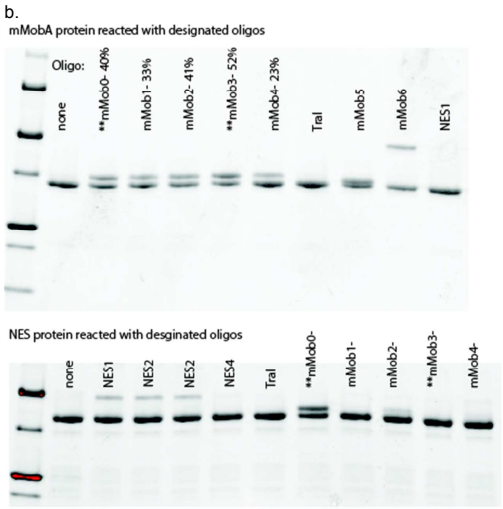

Figure 2.12: **Optimization of mMobA target sequence**
a. Sequences of oligos used in this figure aligned at nick site. b. (top) mMobA or (bottom) NES was reacted with approx. 5-fold excess of denoted oligos for 15 minutes at 37 degrees and run on SDS-PAGE. For oligos mMob0-4 reacting with mMobA protein, % covalent adduct was calculated by quantitation of bands in ImageJ. Please note **- oligo mMob0 contains the original ori sequence for mMobA. This forms a covalent adduct with mMobA to a 40% extent (top) BUT also cross reacts extensively with NES (bottom). A single bp substitution in oligo mMob3 enhances covalent adduct to 52% (top) and abrogates cross-reactivity with NES (bottom).

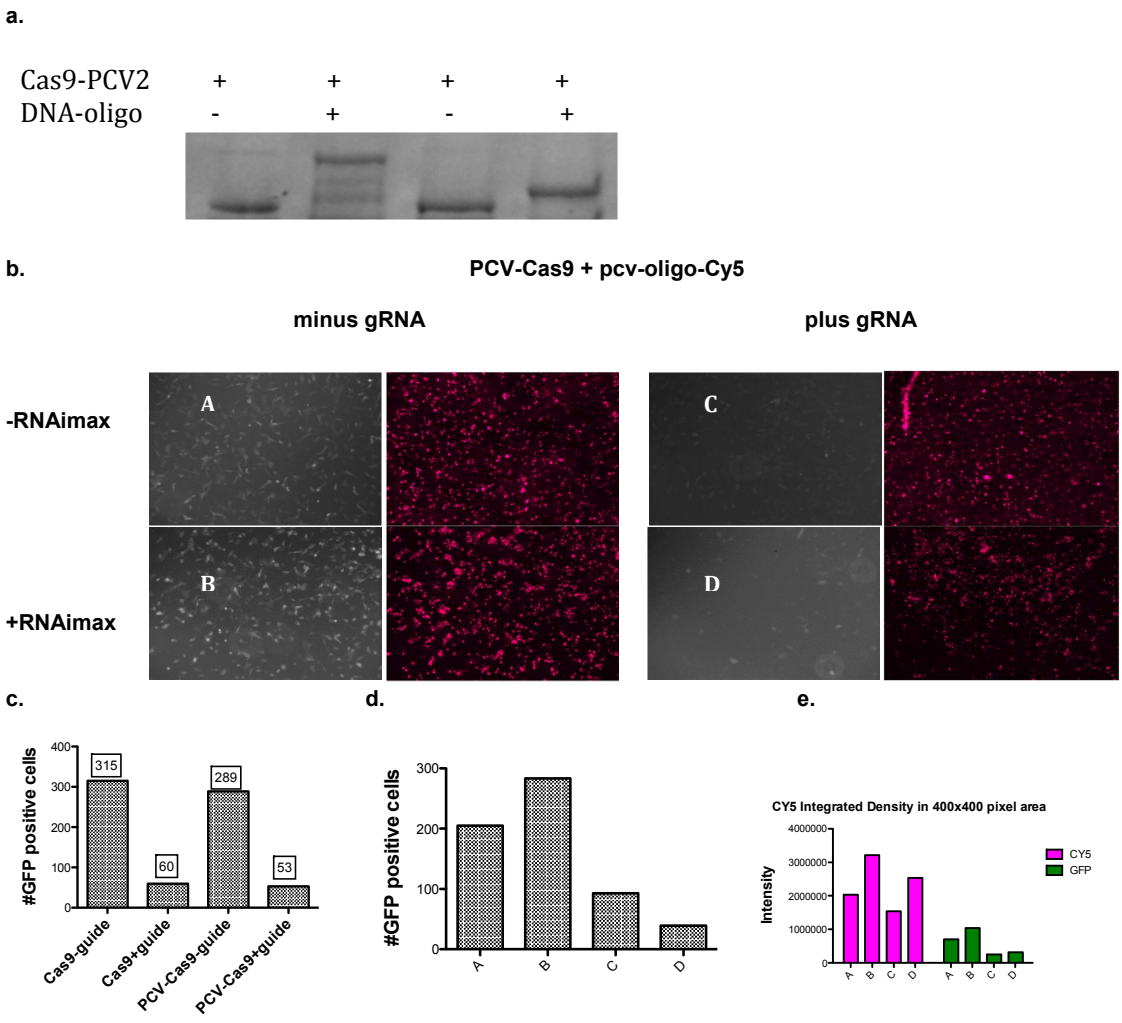

Figure 2.13: **Cas9-HUH RNP transfection**
a. Recombinant PCV-Cas9 reaction with 200bp oligo containing target site and standard target oligo, using standard HUH reaction conditions, as described in Methods. b-e. PCV-Cas9 + pcv-target oligo- Cy5 knockdown of GFP. HCT119 cells stably expressing inducible GFP were treated in 96-well plates with 10pmol of PCV-Cas9 pre-reacted with 50pmol Cy5 oligo, plus or minus 10pmol GFP-gRNA, plus or minus 0.5µL RNAimax. 12 hours later, GFP was induced with doxicyclin. 12 hours later, cells were imaged on an EVOS-FL-AUTO at 10x. (c,d,e) To quantitate GFP knockdown, GFP images (raw images shown above) were loaded into ImageJ as a stack, and converted to 8-bit greyscale. A 2? Gaussian blur was applied, followed by background subtraction. The images were threshholded, and then the number of cells counted (c,d). For intensity analysis (e), a 400x400 pixel region was analyzed for integrated intensity density for background subtracted images (GFP) or raw images (CY5).

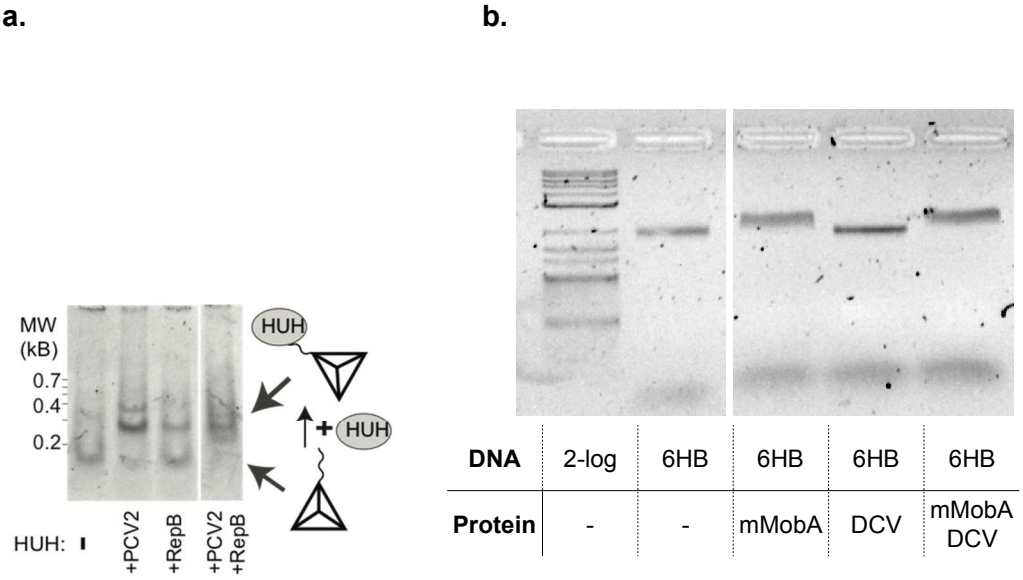

Figure 2.14: **Reactions of HUH-tagged proteins with DNA origami structures.** a. Proteins were incubated in 10-fold excess with a DNA tetrahedron bearing target ssDNA on its corners; the reaction products were analyzed by 5% TBE-PAGE supplemented with 0.1% SDS and stained with SYBR Gold (Invitrogen). b. The DNA origami six-helix bundle (6HB) was folded and purified by repeated centrifugation in AmiCon 100k MWCO filters. The structure was incubated with the indicated HUH proteins for 30 minutes at 37°C with 2mM MnCl2. Reaction products were separated on 2% agarose in 0.5X TBE supplemented with 11mM MgCl2.

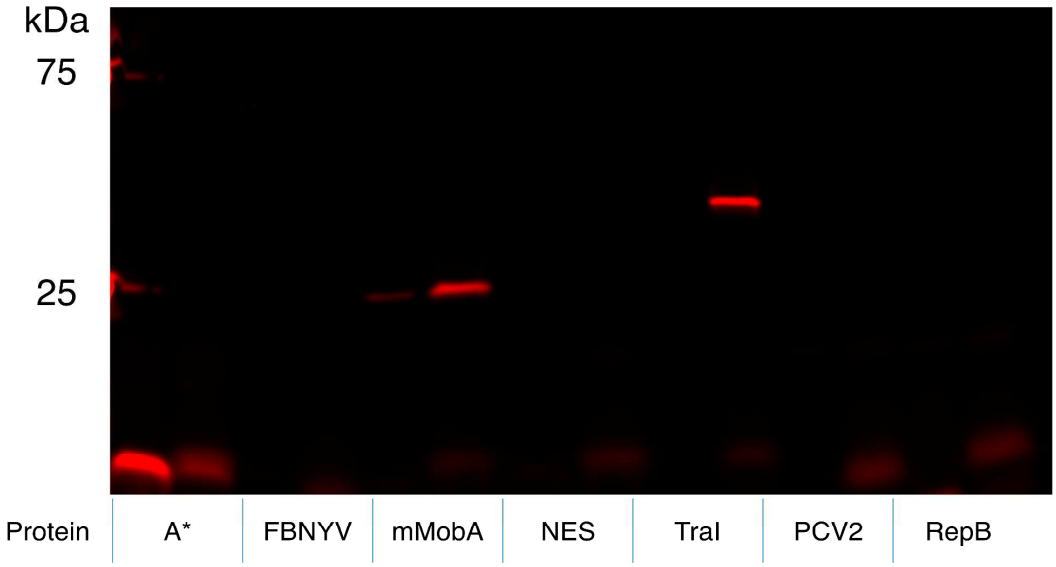

Figure 2.15: **Soluble lysate labeling** HEK293T cells expressing the indicated proteins were lysed and reacted with 1mM TAMRA-labeled target oligonucleotide in the presence of either 10mM NiCl2 (left lane) or 10mM MgCl2 and 10mM MnCl2 (right lane) and imaged using Typhoon Imager. A$*$ is the GeneA HUH-endonuclease domain, which was not used in any other experiments in the manuscript.

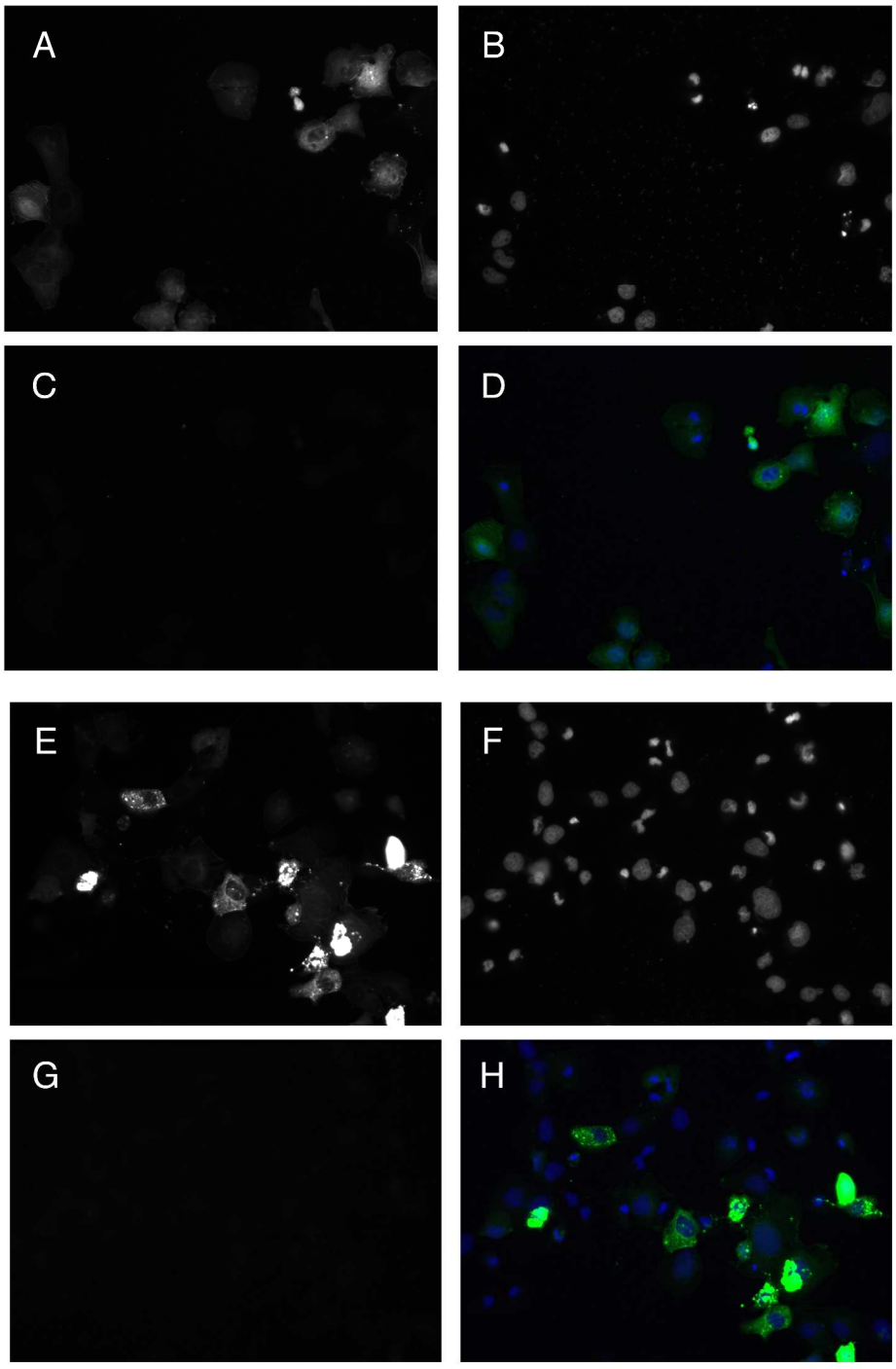

Figure 2.16: **Fixed cell imaging controls**

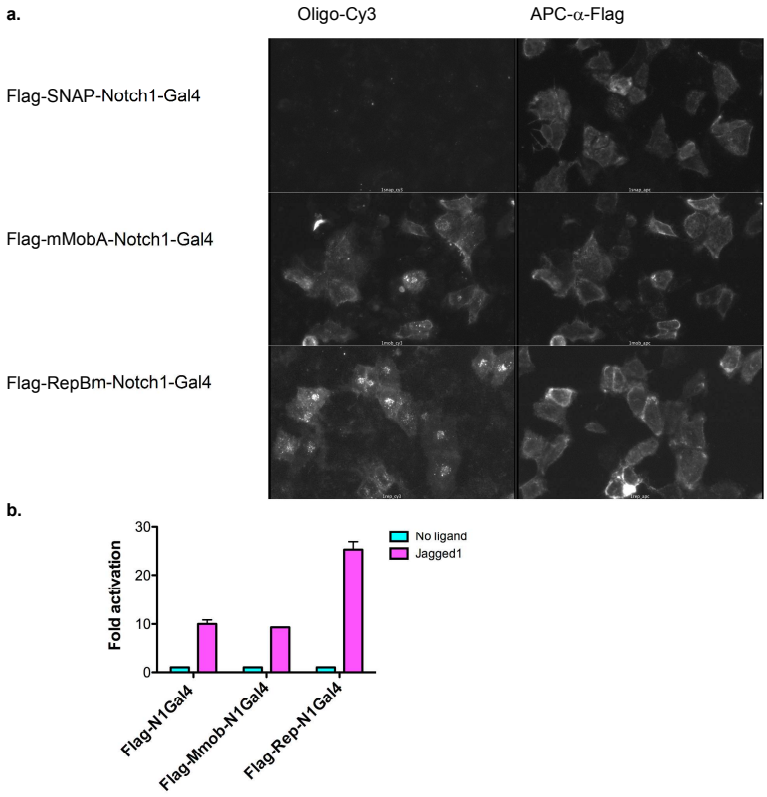

Figure 2.17: **HUH-tagged Notch receptors function normally** a. Cell surface expression and labeling of full- length Notch receptors encoded with N-terminal HUH-tags. U2OS cells were transfected with N-terminal Flag- tagged Notch constructs also containing SNAP, mMobA, and RepBm fusion tags between the Flag tag and EGF1. 24 hours post-transfection, cells were labeled for 30 minutes with an APC-a-Flag antibody (1:750) and a 3?-Cy3 oligo containing mMob or Rep target sequence. The cells expressing SNAP-tagged receptors were labeled with the mMob Cy3 oligo. b. Luciferase reporter assay of HUH-tagged Notch molecules with plated ligand. U2OS cells were reverse-transfected with Flag-, Flag-mMob- or Flag-RepBm-Notch1-Gal4 and reporter constructs containing a Gal4 response element upstream of Firefly luciferase, and a control renilla luciferase plasmid in 96 wells containing PBS or wells precoated with 10µg/ml Jagged1 in PBS. 24 hours later, luciferase reporter assays were performed according to manufacturers instructures (Promega Dual luciferase). Luciferase activity was calculated from the ratio of firefly to renilla luciferase luminescence. Fold activation was calculated by dividing luciferase activity on plated Jagged-1 by luciferase activity of cells plated on untreated wells. Error bars reflect the standard error of triplicate measurements.

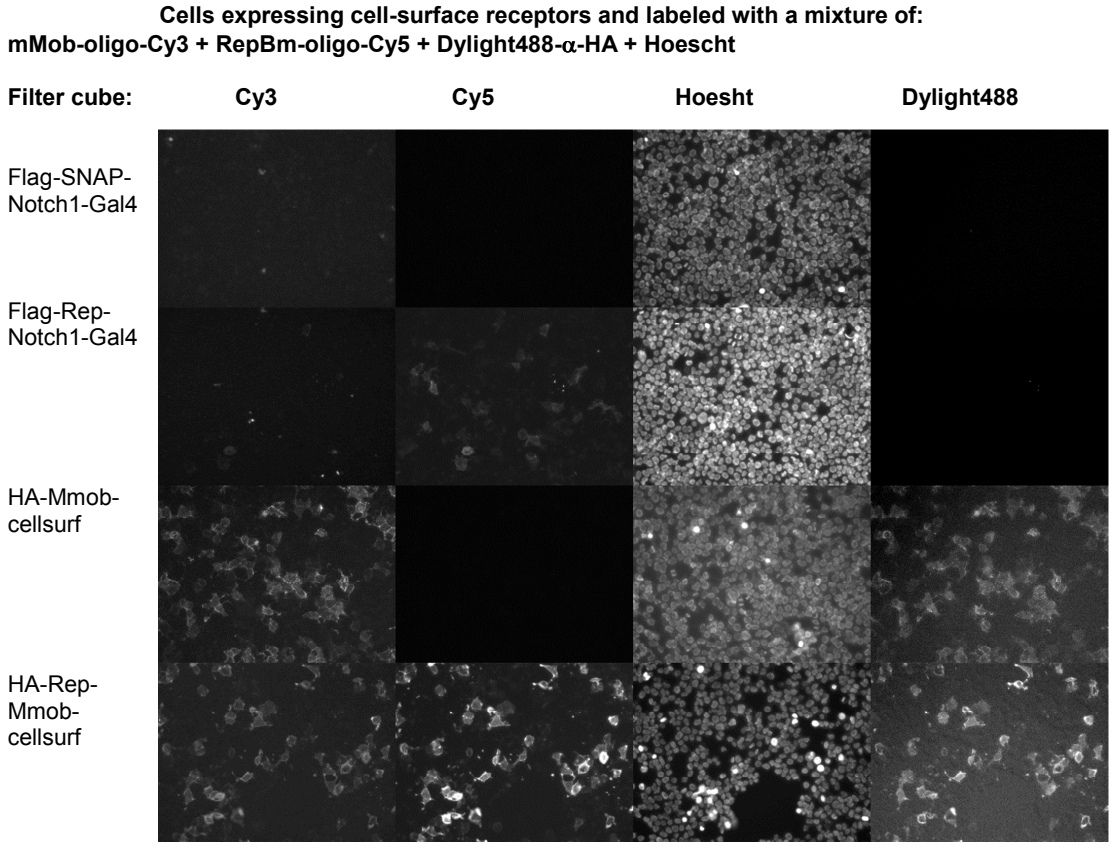

Figure 2.18: **One-pot labeling of cell-surface HUH-tags** HEK293T cells were reverse transfected with cell-surface receptors harboring HUH and epitope tags. After 24 hours, media was removed and cells labeled with a mixture of Cy3 and Cy5 HUH-target oligos, a Dylight488 conjugated anti-HA antibody. After one wash, media was replaced with imaging media containing no phenol red and Hoechst stain was added (1:2500). Images were collected on an EVOS-FL-AUTO with a 10x objective under same light settings for every image. Images were opened in ImageJ. Low and high intensity values for images were set to equal for each filter cube: RFP:20-2000, Cy5: 150-1000, GFP: 200-550, Dapi: 500-4000.

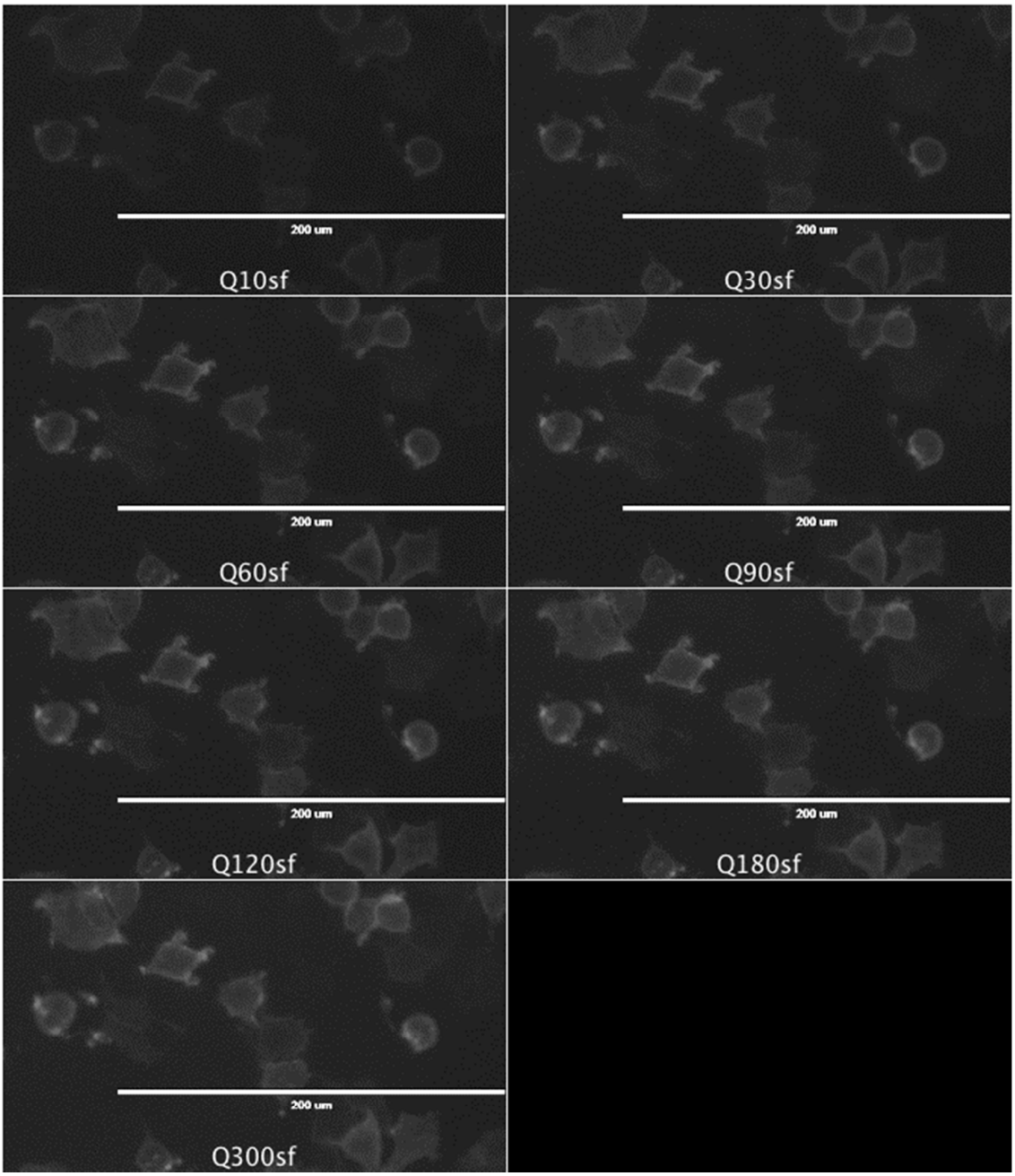

Figure 2.19: **Timecourse of cell-surface labeling** HEK293T cells were transiently transfected with Flag- Rep-Mmob-CD8-TM. 24 hours post-transfection, media was replaced with labeling solution using the Cy3 mMobA quencher oligo, and successive images were immediately collected at 10sec, 30sec, 60sec, 90sec, 120sec, 180sec and 300sec. Images were imported into ImageJ as a stack, and a montage created.

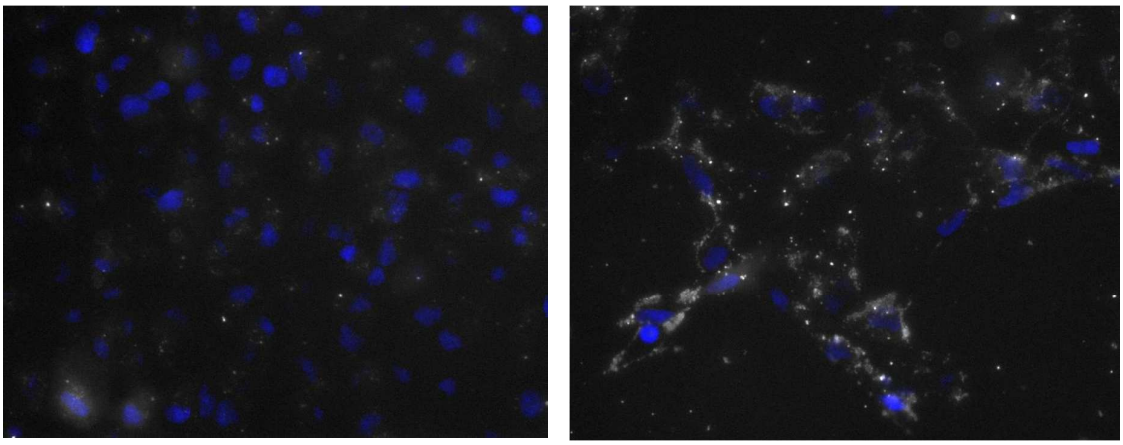

Figure 2.20: **Live-cell fluorescent labeling of HUH-tagged intracellular protein** U2OS cells were transiently reverse-transfected with TraI-β-actin in clear-bottom 96-well plates using standard Lipofectamine 3000 protocols. 24 hours post-transfection, full-media containing 1μM TraI quencher oligo complexed with Oligofectamine cationic lipid was added to untransfected (left) and transfected cells (right). Five hours later, cells were washed once with PBS and Fluorobrite DMEM containing Hoechst stain was added. Images in the GFP and DAPI channels were collected on an EVOS FL-AUTO microscope. Images were imported into ImageJ and the DAPI images false colored blue.

Rational improvement of the miniMobA relaxase

3

Improvement of enzymes for human chemistry

Protein engineering is an area of great importance for many areas of industrial chemistry and biotechnology[87]. As discussed in 1.1, the needs of nature and biotechnology are not necessarily aligned. Both tend to favor enzymes with fast reaction rate and high turnover. However, natural selection breeds for enzymes that are 'good enough' in a specific context; biotechnology seeks 'as good as possible' in a very different context. Thus, protein engineering can involve both improvement of activity and adaptation to *in vitro* conditions such as high temperature or organic solvents[87].

While the field of protein engineering is vast and complex, the experimental approaches used to improve proteins can be roughly divided into 'directed evolution' and 'rational design'[88].

Directed evolution uses sequence diversification to generate a library of mutant proteins, which is then subjected to artificial selection to enrich for a phenotype of interest[89]. The two primary methods for selection are phage[90] and yeast[91] display, which present a mutant protein on the surface of a bacteriophage or yeast cell, respectively, where it can undergo some reaction. The mutant library is constructed such that each virion or cell displays a single mutant, so that the selected mutant clones can be further cultured for repeated rounds of selection. Because it relies on random mutagenesis, directed evolution does not require any structural information about the protein; however, the practical limitations of the experiment prevent screening of every possible mutant, which may lead to 'missed' opportunities for improvement.

Rational design draws on sequence and structural information to design mutants with the aim of creating or improving a desired function[92]. Mutations can be designed manually

by the experimenter or computationally. If the chemical mechanism of a protein is known, sites for mutation can be selected from a known structure; Rosetta modeling is widely used to computationally screen mutants to narrow the possibilites for a selected site[93–95]. As important residues tend to be conserved through evolution, stabilizing mutations can be inferred by analyzing multiple sequence alignments in a process known as consensus design[92,96].

In practice, modern protein engineering almost always uses a hybrid or 'semi-rational', approach, in which sequence and structure information are used to design targeted libraries which then undergo artificial selection[87,97].

Both the HALO- and SNAP-tag underwent protein engineering before being commercialized. The HaloTag was created by mutating the active site of a natural haloalkane dehalogenase to trap the covalent protein-substrate adduct[51]. The SNAP-tag was improved by several rounds of targeted mutagenesis followed by phage display[69], producing a mutant with 52-fold increased activity and a 17°C increase in melting temperature[68]. This mutant was further modified and evolved to yield the orthogonal CLIP-tag[32].

Evolution of the HUH proteins presents a special problem as it is difficult, though not impossible, to select for formation of the covalent bond. In one of the few examples of this strategy, Keeble et al.[98] used phage-display with a stringent washing step of low pH and detergent to select for isopeptide bond formation by the engineered SpyCatcher protein, resulting in a two- to three-fold improvement in activity.

The minimal relaxase domain of MobA[52,60] (miniMobA, or mMobA) presents an attractive target for improvement by rational mutagenesis. As stated above, it is inherently difficult to select for a covalent bond formation, and thus evolution would require a specialized protocol. A high-resolution structure of mMobA is available[60], and a large number of homologous sequences are available in the UniProt refprot database to guide the selection of mutations. Finally, the chemistry performed in the HUH reaction is fairly simple, using a single metal ion to polarize a bond and prime it for nucleophilic attack, suggesting that it should be possible to rationally predict the effect of active site mutations on the reaction efficiency.

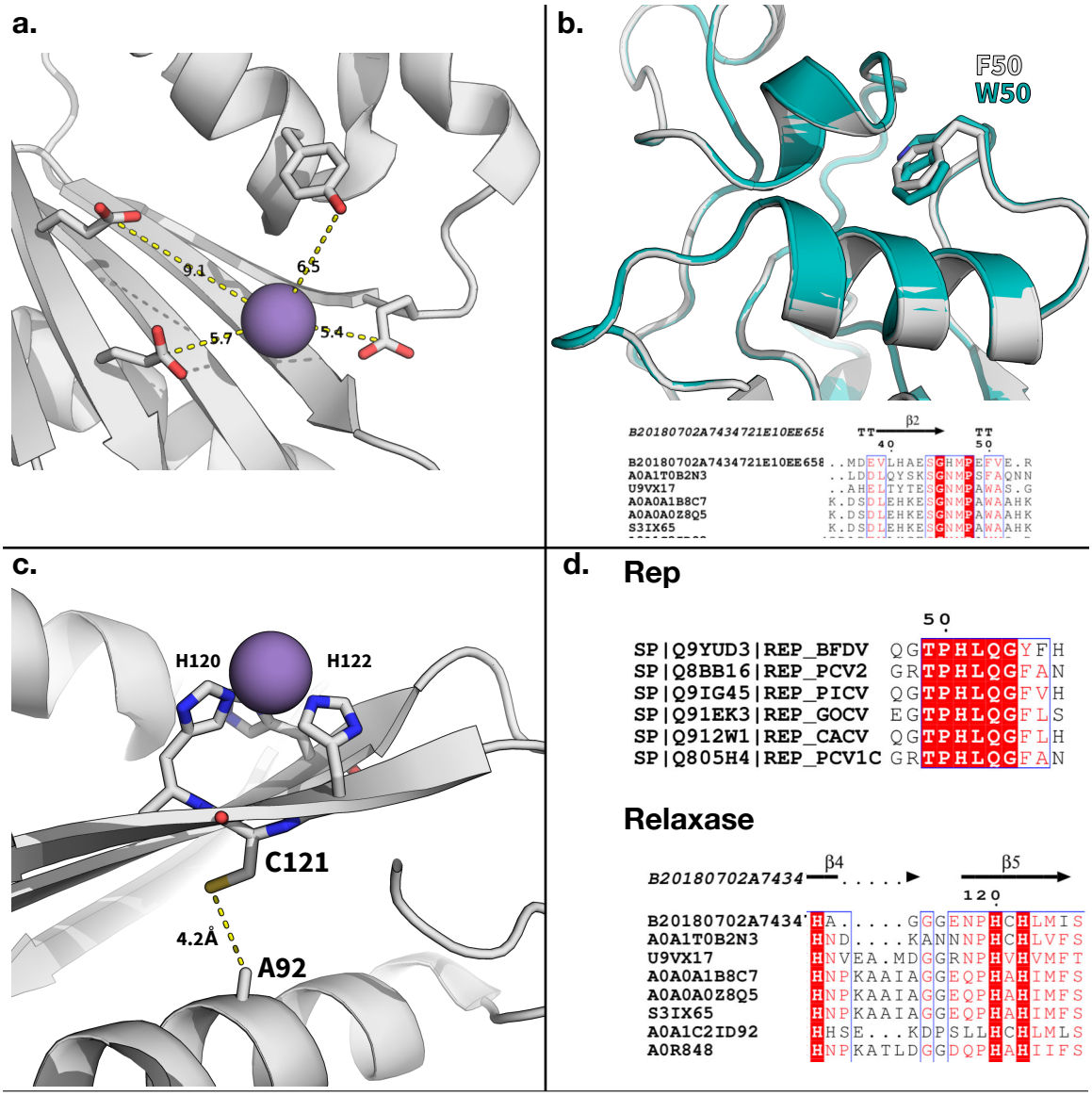

Figure 3.1: **Designed mMobA mutation sites** See text for details

Design of mutants

Active site mutants: Y32F, E38A, E74A, E76A As discussed previously, the aims of proteins *in vivo* and *in vitro* can diverge drastically. In their natural role, the phosphotyrosine bond formed in the HUH reaction is reversible, leading to religation of the DNA strand. For use as labeling tags, it would be ideal to abolish this reverse reaction for highly efficient labeling.

The HUH proteins vary in sequence as well as activity, which provides clues to potential mutants of interest. The viral Rep proteins are uniformly highly efficient at labeling, reaching close to complete reaction very rapidly. In contrast, the mMobA reaction reaches a maximum of approximately 50%. This suggests that 'remodeling' the active site of mMobA to more closely resemble that of the Reps could improve the efficiency, without affecting sequence specificity. Notably, all structurally-characterized Rep proteins have only a single tyrosine near the catalytic metal, while mMobA has two - Y25, known to form the phosphotyrosine, and Y32, whose function is unknown. Monzingo et al.[60] first constructed the Y32F mutant protein and found that it could mediate plasmid mobilization in *E. coli* cells at wild-type levels, but did not test its reaction efficiency *in vitro*.

The phosphotyrosine is not necessarily broken by another tyrosine; in theory, any nucleophilic atom can attack and break the bond. In the RepA protein of plasmid pC194 this reaction is mediated by a glutamate residue near the catalytic tyrosine, which coordinates an attacking water molecule[99]. mMobA has three glutamates - E38, E74, and E76 - in close ($\leq$ 10Å) proximity to the metal site, where they could potentially promote hydrolysis of the covalent adduct. To assess their impact on the stability of the complex these residues were mutated to alanine as a neutral mutation. Preventing breakage of the phosphotyrosine adduct could lead to more efficient labeling of the protein.

Space-filling mutant: F50W The F50W mutant was identified computationally by both the RosettaVIP[100] and Protein Repair One-Stop Shop (PROSS)[101] webservers. Both programs identify stabilizing mutations by using Rosetta macromolecular modeling[102] to identify 'holes' in the hydrophobic core of the protein where packing is less than optimal, then attempt to fill them with mutant residues; the VIP method selects residues solely based on calculated energy, while the PROSS method incorporates multiple sequence alignments to select mutations. The F50W mutation was independently identified by both methods,

and sequence alignment of homologous proteins shows that F in this position is unusual. F50 lies on a minimally-structured loop on the opposite side of the protein from the active site, where its side chain extends into the core of the protein. Mutating this residue to W maintains its hydrophobic character while increasing its size, potentially creating a stronger hydrophobic core.

'U' mutants: A92C, C121A The HUH motif in mMobA is an unusual 'HCH' sequence; the cysteine, C121, is not conserved in close homologs, suggesting that it does not serve a functional purpose. Two mutants were designed to test this. The first, C121A, simply replaces the cysteine with an alanine, which preserves the hydrophobic character while eliminating the thiol side chain. This mutation is both a test of the function of this residue and a useful practical mutation - C121 is the only cysteine present in mMobA, and removing it would allow the introduction of an N- or C-terminal cysteine for labeling with maleimide.

The second, A92C, attempts to introduce a disulfide bond into the backbone of the protein. The thiol of C121 comes into close proximity with the Cβ of A92, suggesting that an added γS could form a bond with γS of C121. The Cβ-Cβ distance is 4.8Å, where previous successful efforts to introduce disulfides used sites with Cβ-Cβ distance of $\leq$ 5.5Å [103]. If successful, this could dramatically increase the stability of the protein - introduction of a non-native disulfide into T4 lysozyme gave an 11°C increase in melting temperature with only a 4% loss of activity [104,105].

Metal-binding mutants: C121A-H122Q The viral Rep proteins use an 'HUQ' motif rather than HUH, and this motif is conserved among sequenced homologs, suggesting that it has a functional significance. The central 'U' residue in the Reps is L, and thus the mMobA C121 was also mutated to A, as above, to mimic this without causing a large perturbation in the structure.

Circular permutation Circular permutation is a protein engineering strategy in which the primary sequence of the protein is 'reorganized' by genetically fusing the N- and C-termini, while breaking the chain in some other location. This, theoretically, preserves most of the native three-dimensional structure of the protein. Circular permutation can result in simple effects, such as reduced proteolytic susceptibility, or can allow drastic changes such as insertion of entire new domains into the protein [106].

The N- and C-termini of mMobA are both semi-flexible, unstructured regions, which in the folded structure are only 11.2Å apart. The permutated constructed was designed to bridge the N- and C-termini using a flexible `GGSGG` linker, for a loop length of approximately 17.5Å. The peptide chain was then broken between S12 and R13, located in a loop which is invisible in the crystal structure and thus likely unstructured.

The end goal of this process was to move the N-terminus closer to the catalytic tyrosine, Y25, placing any fusion protein closer to the labeling site. This would be particularly useful for force spectroscopy experiments, in which the initial β-sheet could unfold and distort the results.

Results

Expression and purification

All mutants were expressed as 6×HIS-SUMO fusions using the autoinduction method[81] and purified by standard metal-affinity chromatography followed by size-exclusion chromatography. Between the affinity and size-exclusion steps the samples were treated with 1mM EDTA to remove any bound metal. Both the wild-type and mutants showed reversible aggregation in low-salt (150mM) buffer, as was previously observed[52]; use of 1M NaCl in the size-exclusion and storage buffer kept the samples in good condition. All the constructs expressed at high levels and were easily purified; the samples could be stored at 4°C for over a month with no apparent loss of activity.

Initial testing showed that all the mutants retained some activity, with the exception of the circularly-permutated construct, which was totally inactive and was excluded from further quantitations.

Reaction yield measurements

The total yield of the labeling reaction was measured as previously[52], using a 10:1 excess of substrate DNA to protein in a standard buffer containing 1mM $MgCl_2$ and 1mM $MnCl_2$. The remaining reactions were quantified by SDS-PAGE gel-shift assays as previously described, quantified using Fiji; the results are shown in fig.

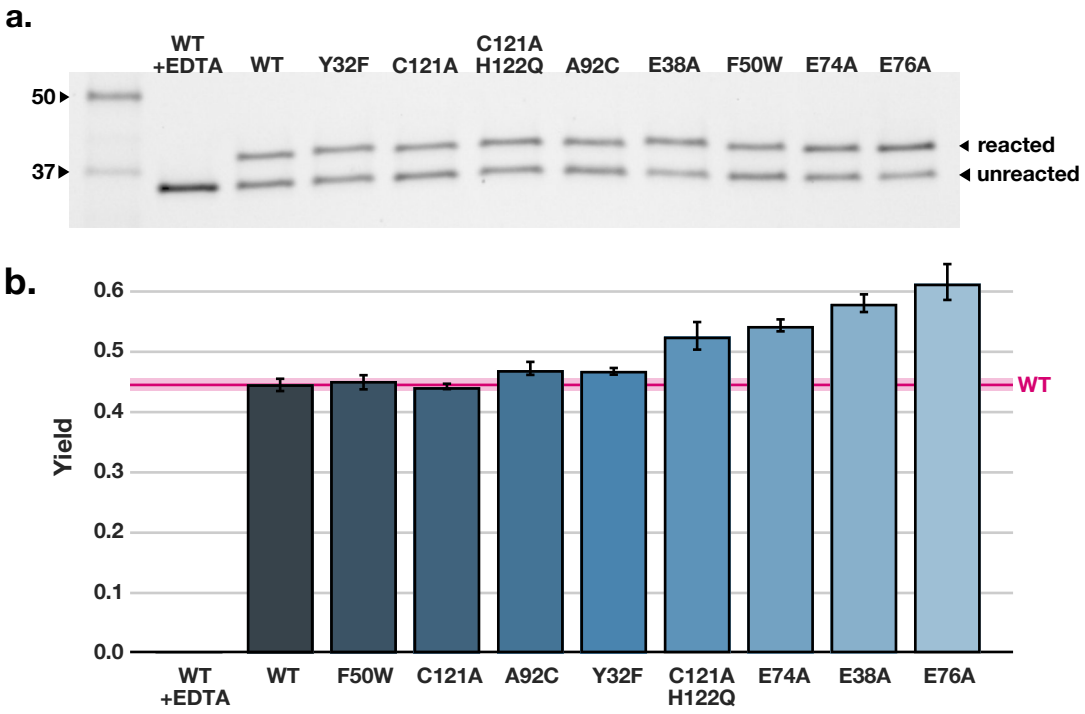

Figure 3.2: **Reaction efficiency of mMobA mutant proteins** Labeling efficiency as measured by SDS-PAGE gel-shift. **a.** Example SDS-PAGE image of reaction products; full uncropped images of source gels shown in fig. 1. **b.** Quantification of $n = 3$ gel-shift assays as shown in a., rearranged to show ascending efficiency from left to right. Error bars represent the 95% confidence interval as calculated by the bootstrap method.

The wild-type reaction efficiency averaged 44%, in line with previous results (ch2). As expected, the mutations designed to increase stability - F50W, C121A, and A92C - did not affect the reaction yield. The Y32F mutant showed no significant increase, indicating that it is not involved in the reaction. The C121A H122Q mutant increased yield to approximately 52%, a small but significant increase over the wild-type. All three glutamate mutations increased the yield. The largest increase was seen with the E76A mutant, which raised the yield to 60%, a 50% increase over the wild-type. The E38A mutant was nearly as effective, averaging approximately 57%. E74A also increased the yield to a lower extent, which is reasonable considering that it is the farthest from the active site; still, it gave a significant increase to the yield.

These results suggest that the phosphotyrosine adduct is highly susceptible to hydrolytic attack mediated by carboxyl groups near the active site, and that removing these residues can significantly increase the efficiency of the reaction.

Structural stability

The structural stability of the mutants was assessed by determining their melting temperature using differential scanning fluorimetry [107,108]. In this assay, the protein sample is mixed with a hydrophobic dye, SYPRO Orange™ (Thermo Fisher); upon binding to a hydrophobic region, such as unfolded protein, the dye increases in fluorescence approx. 500%. Using a real-time PCR instrument, the sample can be heated while monitoring the fluorescence, giving a readout of protein stability vs. temperature. The fluorescence increases initially as the protein's hydrophobic core becomes exposed, peaks, then decreases as the fully unfolded protein aggregates. The inflection point of the fluorescence curve is the melting temperature of the protein; this can be most easily calculated and visualized by finding the peak of the derivative curve.

3.3a. shows the raw fluorescence traces of the mutant proteins, while b. shows the derivatives; the calculated melting points are shown in 3.1. All the mutants were analyzed as fusions to the yeast SUMO protein, which provides an internal standard for the experiment; the SUMO melting point remains constant at 57-58°C across all samples.

The melting point of wild-type mMobA was calculated as 44°C, in good agreement with the previously reported value of 42.7°C[63]. The active site mutants all melted at slightly lower temperatures, 42.5-43.5°C, indicating that the structure of the proteins are well preserved.

The only mutant with a pronounced effect on stability was the C121A H122Q mutant, which fell to 38.5°C. Curiously, the C121A single mutant showed only a small decrease, to 42.5°C, indicating that this decrease was due primarily to the H122Q substitution; the slightly longer glutamine side chain could distort the active site, causing instability. As precautions were taken during the purification process to remove all bound metal it is unlikely that this decrease is caused by a change in affinity for metal, although the possibility cannot be ruled out as the metal content of the purified samples was not measured.

The F50W space-filling mutation was successful, giving a melting temperature increase of 3°C, to 47°C. For comparison, the greatest increase from a single mutation in

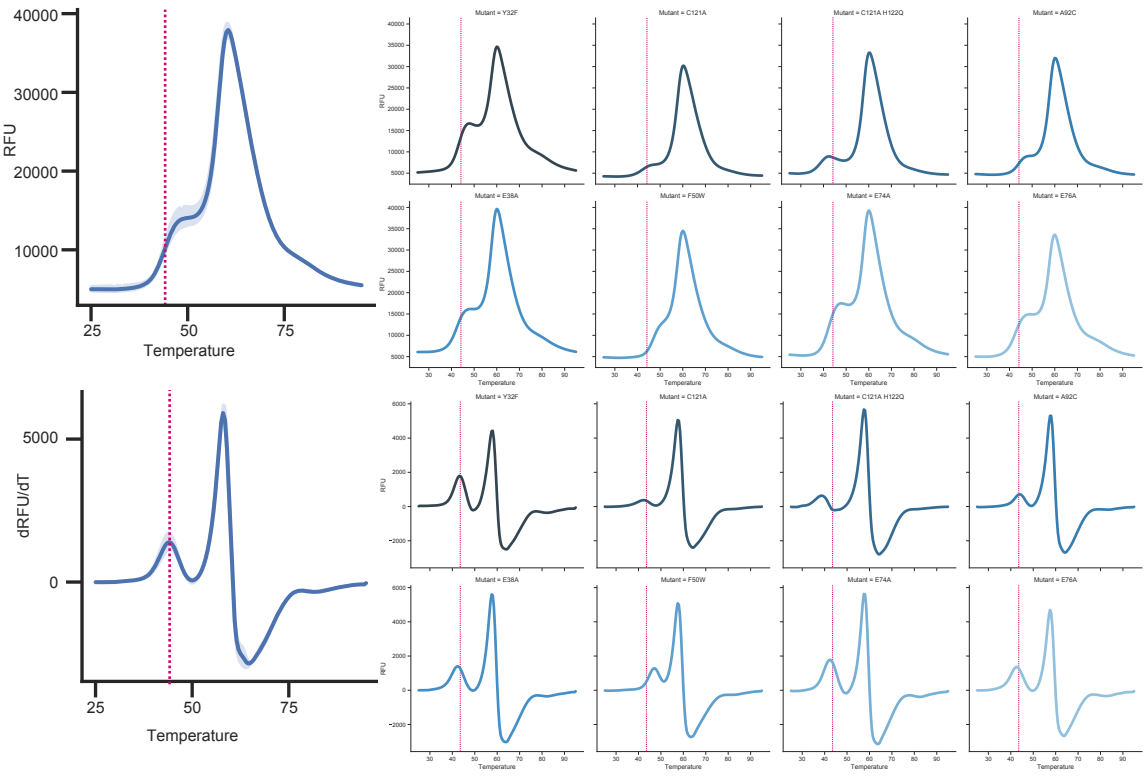

Figure 3.3: **Thermal denaturation of mMobA mutant proteins** 2µM mutant protein was heat denatured in 10mM HEPES, pH 7.4, 0.6M NaCl with 1.75× SYPRO Orange. Fluorescence increase reflects the hydrophobic binding of the dye. **a.** Raw fluorescence traces. **b.** Derivative traces.

Table 3.1: Calculated melting temperatures for mMobA mutants and the fused SUMO tag

Sample	mMobA T_m	SUMO T_m
WT	44.17 ± 0.29	58
Y32F	43.67 ± 0.29	57.83 ± 0.29
C121A	42.5	57.83 ± 0.29
C121A H122Q	38.5	57.5
A92C	44	58
E38A	42.17 ± 0.29	57.5
F50W	47	57.67 ± 0.29
E74A	42.5	57.5
E76A	42.5	57.67 ± 0.29

T4 lysozyme was 5.1°C[104].

The engineered disulfide mutant A92C was not successful. The calculated melting temperature remained unchanged, and analysis of the protein by SDS-PAGE showed no change in native vs. reducing conditions. This is not surprising, as disulfide engineering is inherently unpredictable.

Conclusions

These results are highly encouraging. The obvious next step (currently ongoing in the Gordon lab) is to combine promising single mutations to determine which will work synergistically to improve yield.

I predict that the glutamate mutations, all of which increase the yield on their own, will have additive effects when combined. These residues are fairly well-separated from each other, suggesting that they do not interact.

The F50W substitutions, being quite far from the active site, can likely be included with the yield-enhancing mutations to counteract the slight decrease in stability (as measured by DSF) caused by these mutations. This prediction is supported by the fact that F50W has no effect on reaction efficiency (3.2), indicating that it reinforces the packing of the protein core without affecting the active site.

By the same logic, the C121A mutation merits further investigation. The alanine substitution did not significantly alter yield or stability, suggesting that the cysteine at this position is not functionally important. Inserting larger hydrophobic residues in this site (I/L/V) could strengthen the hydrophobic core, analogous to F50W.

These hypotheses will require careful testing; protein folding and function are extremely complex, and predictions are only predictions. Even so, these results give an excellent starting point for further investigation.

Experimental procedures

Mutant design Manual design was performed using Pymol[109]. RosettaVIP was run on the ROSIE web server[110] for five cycles with three tries per cycle; no residues were excluded. F50W was the only mutant identified. PROSS was run on the PROSS web server[101] with default settings; seven mutated sequences were designed, all of which included the F50W mutation.

Mutant construction All mutants with the exception of the circular permutation construct were generated by inverse PCR[111] using KLD Enzyme Mix (NEB) and verified by sequencing by Genewiz using the primer 'T7 Term', sequence 5'-GCTAGTTATTGCTCAGCGG-3'.

The circular permutation mutant was designed manually and obtained as synthetic DNA from IDT.

Protein expression & purification All proteins were expressed as fusions with a 6×HIS-SUMO protein from the vector pTD68/6×HIS-SUMO, as described previously. Proteins were expressed in *E. coli* NiCo21(DE3) (NEB) grown in autoinduction media[81] for approximately 18 hours at 37°C. Cells were collected by centrifugation at 4,000×G, resuspended in Tris-buffered saline (TBS) + 0.2mg/mL lysozyme, and lysed by sonication for five minutes at 4°C. Insoluble material was removed by centrifugation at 24,446×G, 4°C, for one hour. fast protein liquid chromatography (FPLC) was performed using a Bio-Rad NGC chromatography system. Soluble protein was bound to a 5mL HisTrap FF Crude column (GE Life Sciences). The resin was washed with 10column volumes (CV) 20mM Tris-HCl, pH 7.4, 1M NaCl, 20mM imidazole, 5% glycerol before eluting in 15mL of 95% wash buffer / 5% 20mM Tris-HCl, 4M imidazole, pH 8.1. The full elution was dialyzed against 2L 10mM HEPES, pH 7.4, 1M NaCl, 10% glycerol, 0.5mM EDTA overnight at 4°C.

EDTA was added to the dialyzed protein to 1mM and the mixture was concentrated approximately 3-fold before purification by size-exclusion chromatography on a Bio-Rad SEC70 column in 10mM HEPES, pH 7.4, 1M NaCl, 5% glycerol. Peak fractions were mixed and flash frozen in aliquots. 10µM working stocks were prepared in SEC buffer.

Yield assays Reactions were performed with 1µM protein and 10µM substrate DNA in 20mM HEPES, pH 8, 100mM NaCl, 0.5% glycerol, 1mM $MgCl_2$, 1mM $MnCl_2$. Substrate DNA (100µM in TE) was diluted to 10µM in 20mM HEPES, pH 8, 1mM $MgCl_2$, 1mM $MnCl_2$. EDTA was added to the substrate solution at 10mM final concentration for the negative control '+EDTA' sample. 1/10 volume of 10µM stock of protein containing 1M NaCl were added to the substrate solution, giving a final NaCl concentration of 100mM. Unless otherwise specified, reactions were for 30 minutes at 37°C. The reaction was stopped by addition of 4× NuPAGE™LDS Sample Buffer; the products were denatured at 80°C for five minutes and 5µL loaded on a 4-20% polyacrylamide gel in Tris-Glycine-SDS buffer (Bio-Rad #456-8096). The gel was run for 40 minutes at 200V constant voltage. In-gel stain-free compound was activated with UV light for five minutes. The stained gel was imaged on a Gel-Doc EZ (Bio-Rad) and exported in 16-bit TIFF. Quantification was performed using Fiji[85]. The image was background-subtracted using a rolling-ball radius of 50 pixels and smoothing disabled, and each lane was analyzed using the Gel Analyzer tool. Reaction yield was calculated as the integrated density of the reacted band divided by the integrated density of both bands - $Y = \frac{R}{R+U}$.

Thermal denaturation Thermal stability was measured by differential scanning fluorimetry on a Bio-Rad C1000 Touch thermal cycler equipped with a CFX96 Real-Time System (Aihara lab, University of Minnesota). 1/5 volume of 10µM stock of protein containing 1M NaCl were added to 20mm HEPES, pH 7.4, 500mM NaCl, giving a final NaCl concentration of 600mM. SYPRO™Orange (Thermo Fisher Scientific, 5000× concentrate in DMSO) was diluted 1:400 in same. 30µL of protein solution and 5µL dye solution were mixed at room temperature. The mixture was held at 25°C for two minutes, then heated by 0.5°C every 30 seconds to 95°C. Data was analyzed and melting temperature calculated using Bio-Rad CFX Manager software.

Acknowledgements Thanks to Hideki Aihara for allowing us use of his equipment, and Thomas Bohl for assistance in setting it up.

HUH conjugation in magnetic tweezers

4

Introduction

Single molecule biology

The development of single-molecule techniques and their application to biology has yielded unique and valuable information. Rather than measuring the population average of a trait of interest, single-molecule experiments examine the individual molecules in the population, revealing the heterogeneity of a sample.

For instance: for an enzyme preparation of n molecules with total activity A_T, is the activity of the preparation homogeneous among all the protein molecules, ie $A_T = n \times \frac{A_T}{n}$, or is the activity $A_T = 0.1n \times 10A_T + 0.9n \times 0$, or some more complex formula? This cannot be answered by solution experiments.

Generally a single-molecule experiment is performed by anchoring the molecule(s) of interest to a surface, usually glass or mica. This reduces a three-dimensional solution to a two-dimensional array, spatially separating the molecules so that they can be analyzed individually. Detection of individual units is achieved by linking the molecules to detectable groups. The sensitivity of single-molecule detection creates its own difficulties - spurious interactions that would be masked in other assays can be indisinguishable from the true signal of interest. Further difficulty arises from the anchoring necessary for spatial separation.

Magnetic tweezers

In a magnetic tweezer (or magnetic trap) a paramagnetic or superparamagnetic particle, usually called a 'bead', is placed in a magnetic field ($\vec{B}$) produced by permanent magnets

or, occasionally, electromagnets[112]. The magnetic field induces a magnetic moment in the particle, which creates a force proportional to the gradient of the field, oriented perpendicular to the direction of $\vec{B}$[113]. Importantly, the strength of the magnetic field decays over distance in accordance with the Biot-Savart Law[114]; thus, the force exerted on the particle can be varied simply by moving the magnets in relation to the sample[a]. For millimeter-scale permanent magnets and micron-scale magnetic beads the field has a characteristic length scale of approximately 1mm; magnets can be easily moved to adjust the force over the biologically relevant range of $10^{-3} - 10^{2}$ pN[115]. Because the length scale is relatively large, the apparatus does not require sophisticated feedback systems to control the positioning, making it ideal for the study of very low forces.

Magnetic tweezers have a number of practical advantages over other single-molecule techniques. Magnets can be used to apply torque by rotating the field, making them essential in the study of DNA supercoiling and relaxation[116,117]. An ideal, 'perfect' superparamagnetic bead would not experience torque; in reality, beads contain small inconsistencies from manufacturing, giving their magnetic moment anisotropy which aligns with the magnetic field[118].

Magnetic tweezers are relatively simple and inexpensive to set up and operate, as they do not require any modification of the optical path in a standard inverted microscope. Because the molecules of interest are tethered to large (micron-scale) beads sophisticated illumination and detection are not necessary - an IKEA desk lamp and a USB webcam have been used successfully[119]. At the other extreme, magnets can be integrated into other sophisticated optical setups, such as single-molecule TIRF[120] or FRET[121] microscopes, and optical tweezers[122].

Crucially, magnetic tweezers provide much higher throughput than other single-molecule techniques. A single experiment can track several hundred beads, sufficient to extract statistically significant measurements of reactions such as ligand binding[123] or proteolysis[83].

DNA handles for precise manipulation

Unlike optical tweezers, which can trap objects in three dimensions, a magnetic trap is only one-dimensional; it is, in fact, physically impossible for magnetic fields to stably trap

[a]In a configuration using electromagnets the field strength can be varied by adjusting the applied current[112]

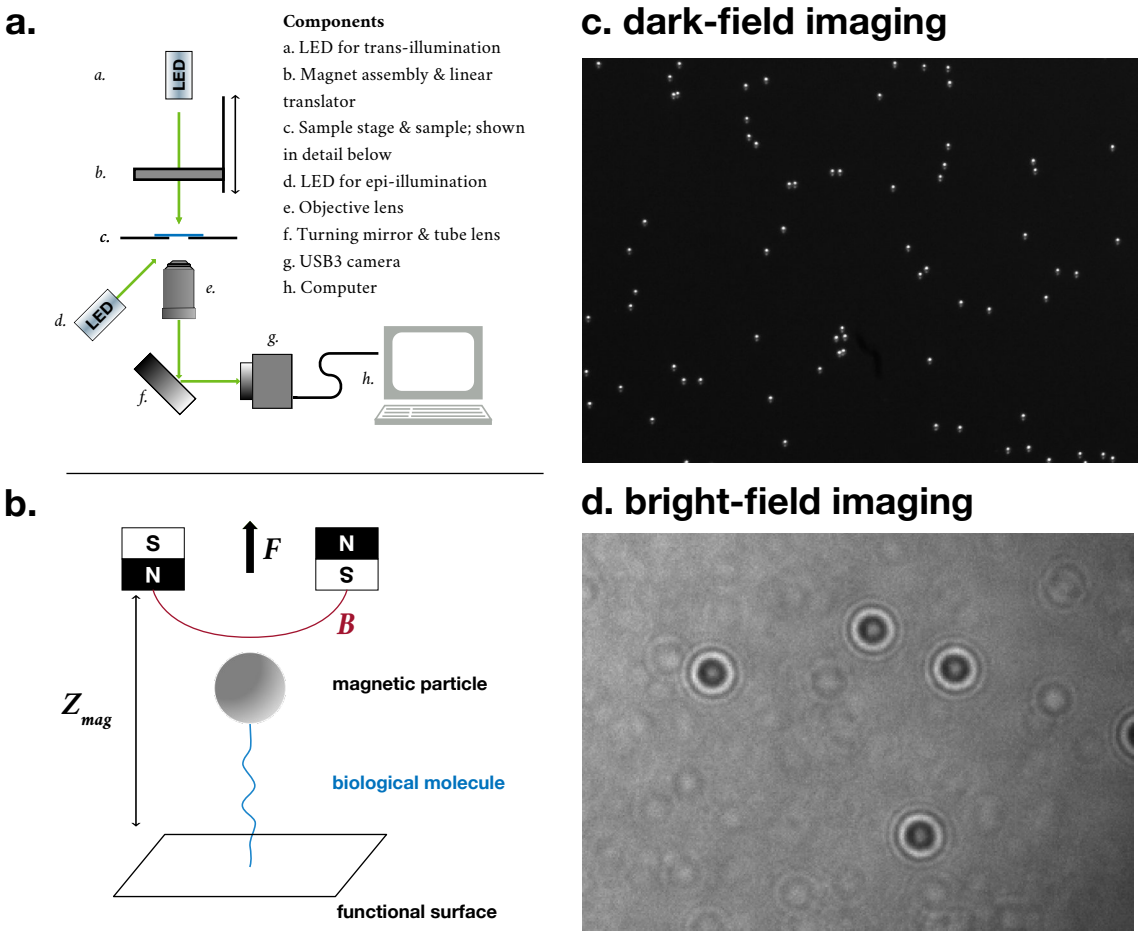

Figure 4.1: **Magnetic tweezers** a. Components and layout of a basic magnetic tweezer apparatus; the optical path shown is simplified, and can be modified in many ways b. Forces applied on the sample c. Dark-field image of 2.8µm magnetic Dynabeads captured using epi-illumination d. Bright-field image of 2.8µm magnetic Dynabeads captured using *trans*-illumination

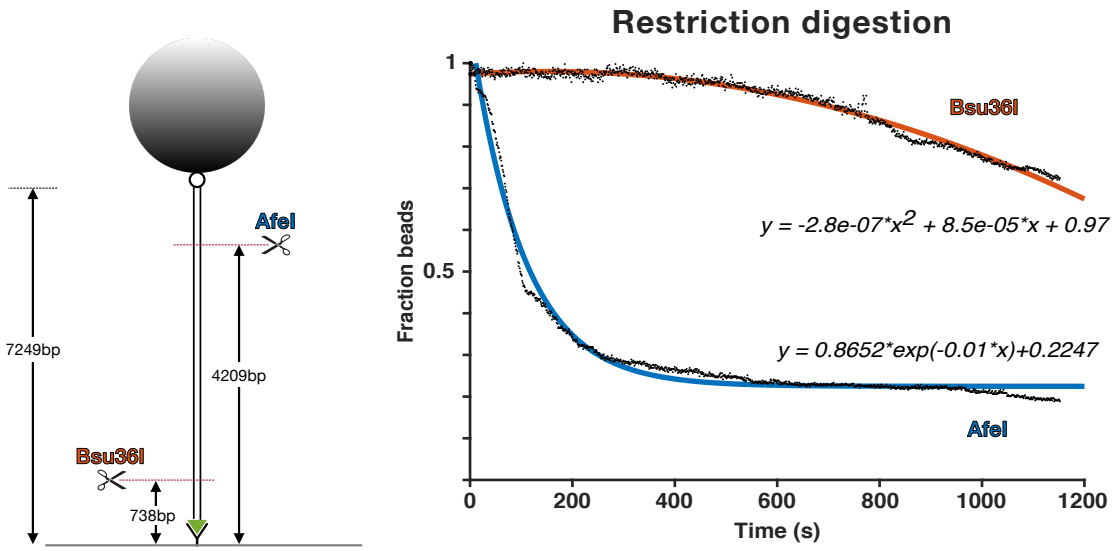

Figure 4.2: **Enzyme activity close to a surface** a. DNA substrate (not to scale): the construct is attached to the surface via digoxigenin/α-digoxigenin interaction and the bead via streptavidin/biotin. b. Collected at 1Hz under 1pN force at 22.5°C.

a particle[118,124]. Due to this limitation, the molecule of interest must be physically tethered or it will simply move towards the magnet. This is accomplished by anchoring the molecule to a solid surface, nearly always a glass microscope coverslip functionalized in some way to bind the molecule.

This causes difficulties when attempting to study fragile biological molecules. Both the magnetic bead and the solid surface present non-biological surfaces which can have unpredictable effects on the surrounding area. For example, even well-passivated surfaces can exert measurable forces on optically-trapped beads up to 100nm above the surface[125].

An example of the difficulty in studying enzymatic reactions using magnetic tweezers is provided in fig. 4.2. A 7kb (2.4µm) DNA substrate in a magnetic trap was digested with two different restriction enzymes: Bsu36I, which cuts ≈250nm from the surface, and AfeI, which cuts ≈1400nm from the surface. The enzyme concentrations were normalized by activity, as given by the manufacturer. Tracking the reaction by counting the number of beads remaining attached to the surface shows that AfeI cuts the distant site with fast, exponential kinetics, as expected; the Bsu36I reaction close to the surface proceeds very slowly and does not fit well to an exponential equation.

Due to this, it has been demonstrated that attaching small proteins or DNA structures to long polymer 'handles' to separate them from interference provides the cleanest and most accurate data[126]. By far the most widely-used material for this purpose is DNA[19,23]. As a polymer DNA is ideally suited for this role; it is one of the stiffest polymers, with a persistence length[b] of 50nm[127], corresponding to a sequence length of 150bp; a DNA handle sufficient to provide several hundred nanometers of separation can be easily prepared by PCR. Currently, attachment of the DNA to proteins is generally achieved by crosslinking using bifunctional linkers such as SMCC[19] or disulfide bond formation between engineered cysteines and thiol-modified DNA[23,128]. Attachment of the DNA to the surface and beads is nearly always achieved using biotin-DNA and streptavidin beads combined with digoxigenin-DNA and anti-digoxigenin surface, or vice versa[19,128].

Chemical labeling has obvious limitations in specificity and efficiency, as noted previously. Perhaps more importantly, they are not usable in complex mixtures such as cell lysate or conditioned media, precluding their use in advanced techniques such as single-molecule pulldown[129]. Given that the HUH proteins function in both these contexts, we hypothesized that they could be used for anchoring of DNA for single-molecule experiments.

The synthesis of DNA handles for HUH attachment presents a technical problem. Attachment handles are most often generated by PCR using modified primers[19,23] or modified nucleotides[130]. This is not possible for HUH-target handles, as the targeting group is itself DNA; if appended to a primer the opposite strand will simply be extended to the end, creating dsDNA rather than the necessary ssDNA. Two techniques were used to overcome this problem, one adapting PCR and the other using DNA self-assembly.

PCR DNA handles were generated by incorporating a polyethylene-glycol spacer group into the primer between the HUH target and the hybridization region; the PEG spacer acts as a block to extension[131], leaving the HUH target sequence single-stranded. While PCR is efficient and convenient for small and medium-size DNA, its efficiency quickly decreases for longer segments (>3000bp). For the generation of long fragments up to 7000bp we employed a previously-described method for DNA self-assembly[20]. In this method, single-stranded m13 phage genomic DNA is hybridized with a mix of primers covering the entire length, creating a double-stranded product. By this method, addressable single-stranded sites can be added at any point along the sequence.

[b] see 4.4 for a very short introduction to polymer physics

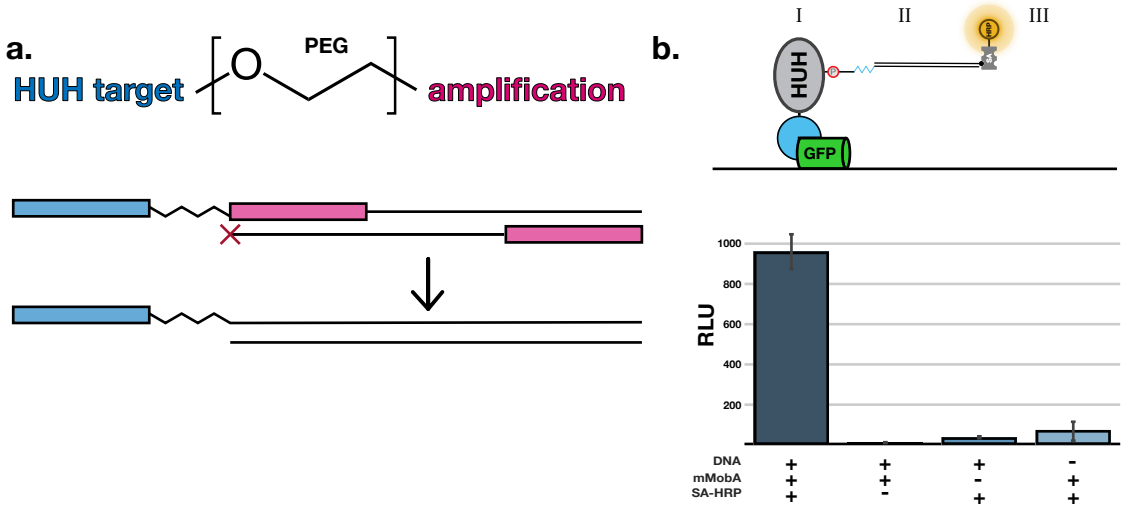

Figure 4.3: **Spacer PCR** a. Schematic for the spacer PCR reaction; the PEG linker used in this work contained six repeats of the ethylene glycol block, for a total of 18 carbons long b. Analysis of overhang DNA attachment by ELISA; full data is presented in table 1.

Results

Construction of overhang DNA handles by PCR

A 1728bp fragment of m13 phage DNA was amplified by spacer PCR with a six-unit PEG spacer; primers were designed containing the mMobA, PCV2, and RepB target sequences. The amplification was successful, with no apparent decrease in yield compared to a 5'-amine primer (4.3). Reacting the mMobA protein with the DNA in solution and analyzing the products by agarose gel gave unclear results; the result was a smeared product rather than a clear gel-shift. However, the smear only appeared for the combination of mMobA protein and mMobA target overhang; amine-labeled fragments or fragments containing the PCV2 overhang did not show a change on addition of mMobA protein.

For a more definite demonstration of linking a microplate ELISA assay was used (fig. 4.3b). Wells of a polystyrene 96-well plate were coated with GFP overnight. After thorough blocking a fusion protein of anti-GFP nanobody and mMobA was added to the wells. A purified overhang DNA with a 5' biotin on the opposite end was added to the wells to react with mMobA protein, washed thoroughly, and detected with horseradish peroxidase-

conjugated streptavidin by chemiluminescence. The combination of mMobA protein and target DNA produced a strong signal; omission of either protein or DNA abolished the signal to nearly background levels. This provided convincing evidence that the PCR product contained the target DNA and could react with mMobA protein.

Construction of overhang DNA by self-assembly

In the self-assembly protocol developed by Halvorsen et al.[20] the single-stranded genome of phage m13 is linearized by the restriction enzyme BtsCI, then annealed with oligos with cover the entire length of the fragment to produce nicked, double-stranded DNA. The full length of m13 DNA is 7249bp; although this length could be reduced by restriction digestion, the full length was used here.

The fragment was assembled with a 5' end bearing the mMobA or PCV2 target sequence and the 3' end bearing a biotin. The assembly was performed and analyzed as previously described by Chandrasekaran et al.[132]; the assembly reaction appeared to fully label the single-stranded DNA, as evidenced by agarose gel.

Attempts to analyze protein-DNA conjugation were unclear, as above; the combination of mMobA protein and target DNA produced a blurring effect on the band which was not seen with PCV2 target DNA, suggesting that linking was occurring. Repeated attempts to verify the linking by ELISA, as above, were unsuccessful, with all wells showing similar, very high luminescence. This could be attributed to residual single-stranded DNA in the sample, which caused a high level of non-specific adhesion in the wells - the PCR product was purified by spin column chromatography and was presumably nearly free of contaminant DNA, while the self-assembled product was not. Thus, the self-assembled product was tested in the magnetic tweezer assay without this confirmation.

'Direct' anchoring of PCR fragments

Both overhang DNA fragments were anchored for single-molecule analysis using mMobA protein on nitrocellulose-coated glass. Two strategies were employed - a one step 'direct' method, in which the DNA and protein were mixed and added to the glass together, and a two step method in which the protein was first adhered to the glass before adding the DNA.

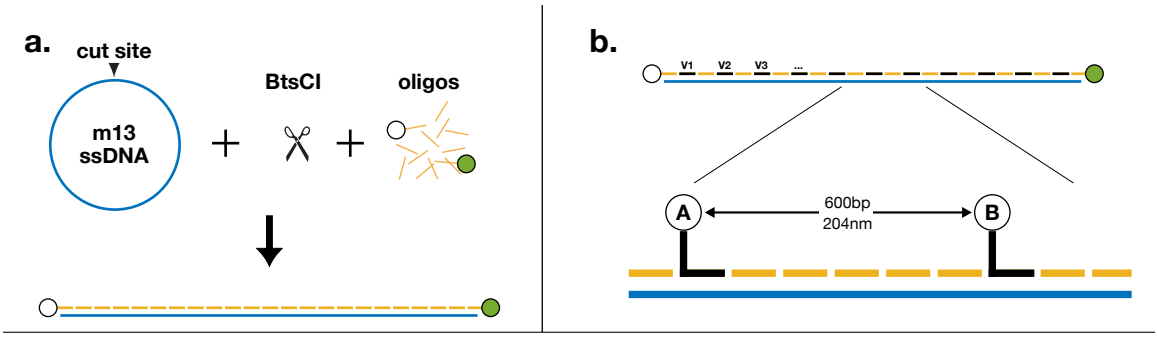

Figure 4.4: **Self-assembly** a. Schematic for the assembly reaction, adapted from Halvorsen et al.[20]. b. Functionalization of the DNA. Because the the DNA is self-assembled any site on the duplex can be functionalized with molecules.

The 'direct' method exploits the differential interaction of protein and DNA with the nitrocellulose surface. DNA is uniformly negatively-charged and thus does not bind strongly to the negatively-charge nitrocellulose. Proteins present a mix of positive and negative charges, as well as hydrophobic patches, which can form strong adhesive interactions with the surface[133]. Thus, after binding the non- or weakly-interacting free DNA can be washed away with buffer, while the strongly-bonded protein and protein-DNA conjugate remain attached.

The DNA and protein were reacted for 30 minutes at 37°C, then added to the glass and incubated for a further 30 minutes at room temperature. Following extensive washing and blocking streptavidin-coupled magnetic beads were added to attach to the biotinylated DNA. Further washing and application of force with the magnets were used to remove unbound beads.

The reaction worked well, although it required extensive washing ($>$ 20 chamber volumes). The MobA protein and MobA DNA produced a dense field of monodisperse particles, while replacing the DNA with an off-target version or removing the protein gave a nearly clear field (fig. 4.5). A number of buffers were tested for removing unbound DNA and beads. Initial tests were successful using alternating washes of Tris-EDTA buffer (TE) and TE + 1M NaCl; usually three cycles of buffer were sufficient. Addition of the competitor polyanion dextran sulfate at 0.1mg/mL was even more effective, requiring less wash volume. The most succesful was a solution of PBS containing 10mg/mL Roche Blocking Reagent, a proprietary mixture specifically marketed for DNA hybridization studies using

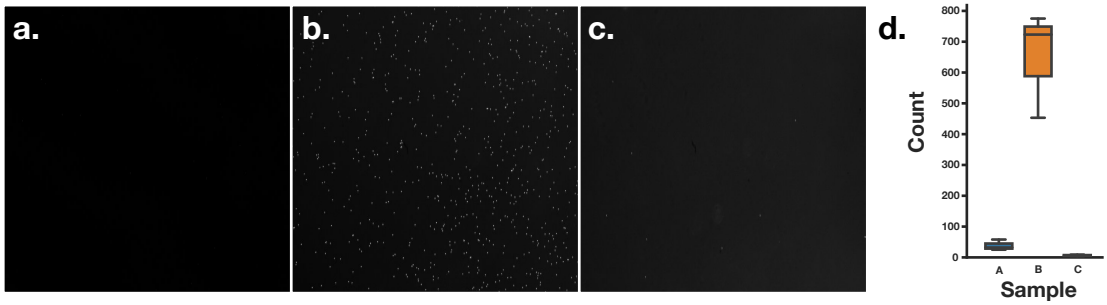

Figure 4.5: **Direct anchoring of overhang DNA** Full image set shown in fig. 2. a. MobA protein + PCV2-target DNA b. MobA protein + MobA-target DNA c. no protein + MobA-target DNA d. Quantification

nitrocellulose[c].

The direct anchoring method was also performed successfully with the self-assembled product, although with significantly higher background binding; as in the ELISA described earlier, this could be a result of free single-stranded DNA causing strong non-specific adhesion.

Although this data is encouraging, the presence and absence of beads is a somewhat limited reporter of successful anchoring; strong, non-specific interactions cannot be completely ruled out by this experiment. Thus the interaction was further analyzed by single-molecule tracking on the longer, self-assembled product.

Two-step anchoring of self-assembled DNA

The two-step protocol was used with the longer self-assembled DNA to obtain high-resolution information on the HUH anchoring interaction.

The mMobA protein was immobilized on nitrocellulose glass and blocked thoroughly with PBS + Roche Blocking Reagent. The self-assembled DNA was then added in HUH reaction buffer containing 1mM MgCl2 and 1mM MnCl2 to react with the adhered protein.

[c] Although officially non-specified, the protein component of this reagent is most likely casein; the product is a white powder, and analysis by SDS-PAGE showed a protein of ≈25kDa, consistent with Bovine κ-casein (data not shown). The non-protein components are unknown

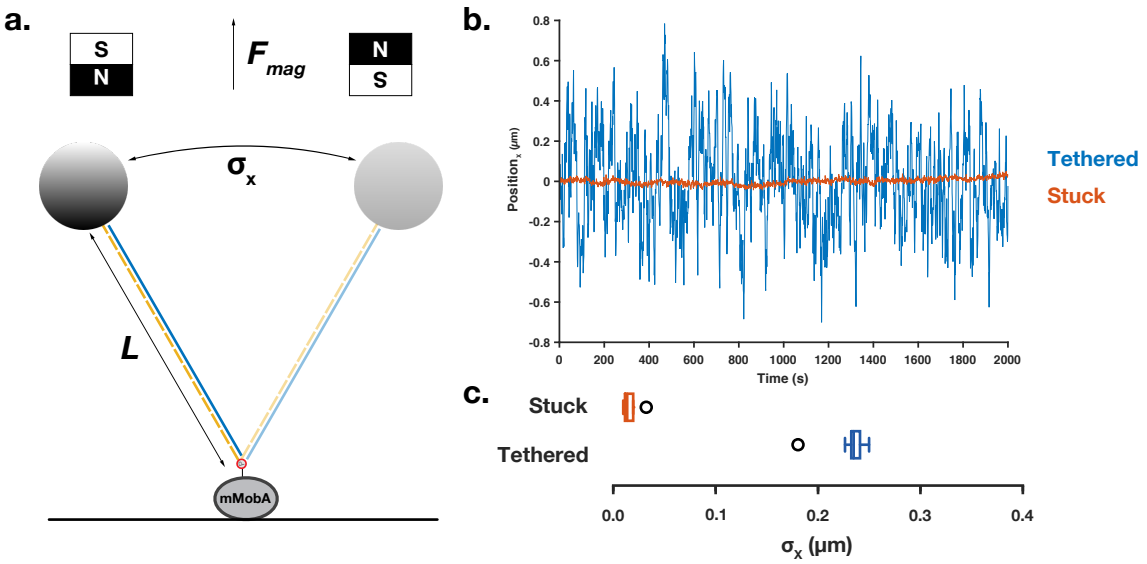

Figure 4.6: **Single-molecule tracking of mMobA-anchored DNA** a. Schematic for the tethered molecule; not drawn to scale b. Tracking results for a single bead, compared to a single stuck bead; for full data set see fig. 3. c. Calculated standard deviation (σ) for all traces.

Unreacted DNA was washed out with buffer and streptavidin-coated magnetic beads were added to reacted with the biotin DNA.

The tethered beads were tracked at 10Hz for 200s by dark-field microscopy at very low (≈ 0.14pN) force to allow large motions of the bead. Subpixel localization of the beads was performed in MATLAB by the radial center method of Parthasarathy [134]; full localization traces are shown in fig.3. Using the inverted-pendulum simplification [135] the variance of the position can be calculated at a known force:

$$F = \frac{k_B T \times l}{\sigma^2} \implies \sigma = \sqrt{\frac{k_B T \times l}{F}}$$

For a 7249bp fragment $l = 2464$nm, giving $\sigma_{calc} = 277$nm, assuming $k_B T = 4.1$pN nm at room temperature. Analysis of $n = 12$ beads gave $\sigma_{meas} = 237$nm, in good agreement with the theoretical prediction and indicating the DNA molecules are specifically tethered by their ends. Tracking of $n = 4$ 'stuck' beads in the same chamber showed almost no variance in the position, indicating low drift in the instrumentation. Likewise, tracking of beads in chambers containing no mMobA protein or mMobA protein and PCV2-target

DNA showed close to no variance, indicating that any beads in these channels are non-specifically attached.

Conclusions and future directions

These data indicate that HUH anchoring is a viable strategy for DNA tethering in magnetic tweezers or other single-molecule techniques. The bonds formed were stable and resistant to high force, and high-resolution tracking demonstrated that the tethered DNA could freely move. This method could be very widely-applicable, as analysis of DNA dynamics remains the most active area of magnetic tweezer research.

While the biotin/streptavidin interaction is highly stable and essentially irreversible under standard conditions ($F_{max} \approx 200$pN), the widely-used digoxigenin/α-digoxigenin interaction ruptures much closer to the biological force regime ($F_{max} \approx 25$pN)[136,137]. The formation of covalent, 'invincible' tethers is an area of great interest, particularly for the study of DNA dynamics where higher forces are required. However, all previous work relies on chemical functionalization for anchoring, which is much more complex than the antibody adsorption method[136,138]. Thus the combined ease of use and high force stability offered by HUH immobilization could be of great use to many groups.

The obvious future direction is the use of HUH proteins for the other side of the tether - using an HUH protein to fix a fused protein of interest to a DNA tether for precise analysis. This is an area of ongoing work in the Gordon laboratory; although the data are not presented here early work is encouraging.

Experimental procedures

Materials Unless otherwise stated: molecular biology materials and enzymes were from New England BioLabs, coverslips were Thermo Scientific Gold Seal, chemicals were from Fisher Scientific, oligonucleotides were from Integrated DNA Technologies,

Surface preparation 22×40mm coverslips were placed in a Coplin jar, rinsed three times with hot distilled water, and sonicated for thirty minutes in 2% Hellmanex III (Hellma Analytics). The coverslips were then rinsed three times with warm distilled water, blown dry with filtered air, and incubated at 80°C for five to ten minutes in a drying oven.

Nitrocellulose solution was prepared by diluting 2% collodion (Electron Microscopy Sciences #12620-50, 0.45µm filtered) to a concentration of 0.2% in 100% ethanol; solution was used within a week of preparation. 2µL of nitrocellulose was pipetted onto one edge of a clean coverslip, spread evenly across the surface with the side of 200µL pipette tip, and cured at 80°C for ten minutes. Preparted coverslips were stored in a dust-free glass Petri dish and used within a week of preparation.

Flow chamber assembly Flow chambers were assembled using Kapton®polyimide double-sided tape, thickness 0.096" (McMaster-Carr #7361A41). Strips of tape were arranged forming a channel of approximately 20µL volume on a nitrocellulose coverslip and covered with a non-functionalized 22×22mm coverslip; the assembled was pressed firmly and incubated at 80°C for five to ten minutes to seal.

Flow chamber use Solution exchange was performed manually by pipetting into one end of the channel and wicking from the other end using Kimwipes or filter paper. Incubations were performed at room temperature in a humidified chamber, unless otherwise specified.

PCR assembly of overhang DNA A 1728bp fragment of m13mp18 genomic DNA was amplified using an 18-carbon PEG spacer forward primer (IDT code `/iSp18/`) and a biotinylated reverse primer in *Taq*®2× Master Mix, according to manufacturer's protocol. Typical reaction volume was 4×50µL, pooled and purified using a single PCR purification column. The product was analyzed by agarose gel electrophoresis.

Self-assembly of overhang DNA The 7249bp construct was self-assembled from single-stranded m13mp18 DNA by the method of Chandrasekaran et al. [132]; see Halvorsen et al. [20] for a complete list of primers. The 5' end of the construct was functionalized with an extended oligonucleotides with a miniMobA or PCV2 recognition sequence or a 5' digoxigenin; the 3' end was functionalized with a 3' biotinylated oligonucleotide. The product was analyzed by agarose gel electrophoresis and used without purification.

Overhang DNA ELISA Wells of a 96-well polystyrene plate (Corning) were coated with 1µg/mL sfGFP in carbonate/bicarbonate buffer (pH 9.6) at 4°C overnight. The wells were washed two times with PBS then blocked overnight at 4°C with PBS + 0.1% Tween 20 +

2% BSA. Wells were washed two times with wash buffer (TBS + 0.1% Tween 20 + 0.2% BSA). 50µL of antiGFP-mmoba (0.6mg/mL) was added and incubated for thirty minutes at room temperature, followed by three washes. 2µL of handle DNA in 50µL HUH buffer (50mM TrisHCl, pH 8, 0.1M NaCl, 1mM MnCl2) was added and incubated for one hour at 37°C, followed by five washes. 100µL wash buffer with Streptavidin-HRP (Pierce, 1:4000) was added and incubated thirty minutes at room temperature, followed by five washes. 100µL SuperSignal™ELISA Pico (Thermo Scientific 37069) was added and incubated for two minutes with gentle agitation. Chemiluminescence was detected on a Molecular Devices LMaxII 384, set to one second integration time.

Magnetic bead preparation Streptavidin-coated M-280 Dynabeads™(Invitrogen, 10mg/mL) were diluted in 200µL of blocking buffer, vortexed briefly, and collected using a magnet rack; this was repeated four times, for a total of five washes. The beads were then collected and resuspended to a final concentration of 0.1mg/mL in working buffer.

Digoxigenin-tethered magnetic trap A flow cell as described above was filled with 20µg/mL polyclonal α-digoxigenin (from sheep, Roche #11333089001) in PBS + 0.02% sodium azide and incubated for thirty minutes. The chamber was flushed with 50µL PBS, then filled with PBS + 0.02% sodium azide + 10mg/mL Blocking Reagent (Roche #11096176001) and incubated for one hour. The chamber was rinsed with 100µL 1$\times$ CutSmart buffer (NEB) before adding 20µL magnetic beads in CutSmart buffer. After two minutes the chamber was rinsed with 200µL CutSmart buffer before imaging.

'One-step' miniMobA immobilization Overhang DNA at 200-500pM was mixed with an approximately 20-fold excess of mMobA protein in HUH buffer (10mM HEPES, pH 8, 1mM $MgCl_2$, 1mM $MnCl_2$) and incubated for thirty minutes at 37°C. The mixture was added to a flow cell chamber and incubated for thirty minutes at room temperature. The chamber was flushed with 50µL HUH buffer, then filled with PBS + 0.02% sodium azide + 10mg/mL Blocking Reagent and incubated for one hour. The chamber was rinsed with 100µL 1$\times$ CutSmart buffer (NEB) before adding 20µL magnetic beads in CutSmart buffer. After two minutes the chamber was rinsed with 200µL CutSmart buffer before imaging.

'Two-step' miniMobA immobilization A flow cell chamber was filled with a 100nM solution of mMobA protein in HUH buffer and incubated for thirty minutes. The chamber was rinsed with 50µL of HUH buffer, filled with PBS + 0.02% sodium azide + 10mg/mL Blocking Reagent and incubated for one hour. The chamber was rinsed with 50µL of HUH buffer and a 250pM solution of DNA was added and incubated for one hour. The chamber was rinsed with 100µL CutSmart buffer before adding 20µL magnetic beads in CutSmart buffer. After two minutes the chamber was rinsed with 200µL CutSmart buffer before imaging.

Restriction enzyme digestion Enzyme was diluted to 0.2U/µL in CutSmart buffer and added to the flow chamber while imaging at room temperature using a pipette and wicking filter paper. The timing of fluid exchange was identified by blurring of the image.

Single-particle tracking Wide-field images were cropped manually in FIJI to isolate single particles for analysis. Subpixel location of the particles was determined by the radial centers method of Parthasarathy [134] using the MATLAB script `radialcenter.m`, available at the Parthasarathy group's page.

Magnetic trap calibration Bacteriophage λ genomic DNA (48502bp) was labeled at one end by ligating the oligonucleotide `5Phos_LAMBDA_3Bio` using T4 ligase for ninety minutes at room temperature. The opposite was labeled with digoxigenin by end-filling using Klenow Fragment (3'→5' exo-) and digoxigenin-11-dUTP (Roche #11093088910) at 37°C for two hours. The reaction was purified by repeatedly diluting and re-concentrating the product using a 100K MWCO spin filter. The final retentate was diluted to 500pM & heated @ 55°C, 5:00. The product was examined for integrity on an agarose gel.

Single-particle traces were analyzed to determine the variance in the position, which was used to calculate the applied force using the inverted pendulum model derived by Strick et al. [135]:

$$F = \frac{k_B T \times l}{\langle \sigma^2 \rangle}$$

where $k_B T = 4.1$pN nm and $l = 16.4$µm for λ-DNA. Image analysis and calculations were performed in MATLAB.

References

5

[1] F H CRICK. On protein synthesis. *Symp Soc Exp Biol*, 12:138–163, 1958. URL http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=13580867&retmode=ref&cmd=prlinks.

[2] Mario Bunge. A General Black Box Theory. *Philosophy of Science*, 30(4):346–358, 1963. doi: 10.1086/287954. URL https://www.journals.uchicago.edu/doi/10.1086/287954.

[3] Haihong Wang and John E Cronan. Functional Replacement of the FabA and FabB Proteins of Escherichia coli Fatty Acid Synthesis by Enterococcus faecalis FabZ and FabF Homologues. *J. Biol. Chem.*, 279(33):34489–34495, August 2004. doi: 10.1074/jbc.M403874200. URL http://www.jbc.org/content/279/33/34489.full.

[4] Albert H Coons. Histochemistry with Labeled Antibody. In G H Bourne and J F Danielli, editors, *International Review of Cytology*, pages 1–23. Academic Press, 1956. URL http://www.sciencedirect.com/science/article/pii/S0074769608625656.

[5] Stratis Avrameas. Coupling of enzymes to proteins with glutaraldehyde: Use of the conjugates for the detection of antigens and antibodies. *Immunochemistry*, 6(1):43–52, 1969. URL http://www.sciencedirect.com/science/article/pii/0019279169901773.

[6] Alex M Valm, Rudolf Oldenbourg, and Gary G Borisy. Multiplexed Spectral Imaging of 120 Different Fluorescent Labels. *PLoS ONE*, 11(7):e0158495, 2016. doi: 10.1371/journal.pone.0158495.

[7] Stephen P Perfetto, Pratip K Chattopadhyay, and Mario Roederer. Seventeen-colour flow cytometry: unravelling the immune system. *Nature Reviews Immunology*, 4(8):648–655, August 2004. doi: 10.1038/nri1416. URL https://www.nature.com/articles/nri1416.

[8] T Sano, C L Smith, and C R Cantor. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science*, 258(5079):120–122, October 1992. doi: 10.1126/science.1439758. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1439758.

[9] Barry Schweitzer, Steven Wiltshire, Jeremy Lambert, Shawn O'Malley, Kari Kukanskis, Zhengrong Zhu, Stephen F Kingsmore, Paul M Lizardi, and David C Ward. Immunoassays with rolling circle DNA amplification: A versatile platform for ultrasensitive antigen detection. *Proceedings of the National Academy of Sciences*, 97(18):10113–10119, August 2000. doi: 10.1073/pnas.170237197. URL http://www.pnas.org/cgi/doi/10.1073/pnas.170237197.

[10] Simon Fredriksson, Mats Gullberg, Jonas Jarvius, Charlotta Olsson, Kristian Pietras, Sigrún Margrét Gústafsdóttir, Arne Östman, and Ulf Landegren. Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology*, 20(5):473–477, 2002. doi: 10.1038/nbt0502-473. URL http://www.nature.com/articles/nbt0502-473.

[11] Mathias Wilhelm, Judith Schlegl, Hannes Hahne, Amin Moghaddas Gholami, Marcus Lieberenz, Mikhail M Savitski, Emanuel Ziegler, Lars Butzmann, Siegfried Gessulat, Harald Marx, Toby Mathieson, Simone Lemeer, Karsten Schnatbaum, Ulf Reimer, Holger Wenschuh, Martin Mollenhauer, Julia Slotta-Huspenina, Joos-Hendrik Boese, Marcus Bantscheff, Anja Gerstmair, Franz Faerber, and Bernhard Kuster. Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502):582–587, 2014. doi: 10.1038/nature13319.

[12] Vanessa M Peterson, Kelvin Xi Zhang, Namit Kumar, Jerelyn Wong, Lixia Li, Douglas C Wilson, Renee Moore, Terrill K McClanahan, Svetlana Sadekova, and Joel A Klappenbach. Multiplexed quantification of proteins and transcripts in single cells. *Nature Biotechnology*, 35(10):936–939, August 2017. doi: 10.1038/nbt.3973. URL http://www.nature.com/doifinder/10.1038/nbt.3973.

[13] Nicholas Stephanopoulos and Matthew B Francis. Choosing an effective protein bioconjugation strategy. *Nature Chemical Biology*, 7(12):876–884, December 2011. doi: 10.1038/nchembio.720. URL https://www.nature.com/articles/nchembio.720.

[14] M J Halloran and C W Parker. The preparation of nucleotide-protein conjugates: carbodiimides as coupling agents. *The Journal of Immunology*, 96(3):373–378, March 1966. URL http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=4160422&retmode=ref&cmd=prlinks.

[15] Markus Kurz, Ke Gu, Amal Al-Gawari, and Peter A Lohse. cDNA–Protein Fusions: Covalent Protein–Gene Conjugates for the In Vitro Selection of Peptides and Proteins. *ChemBioChem*, 2(9):666–672, September 2001. doi: 10.1002/1439-7633(20010903)2:9⟨666::AID-CBIC666⟩3.0.CO;2-{\#}. URL https://onlinelibrary.wiley.com/doi/full/10.1002/1439-7633%2820010903%292%3A9%3C666%3A%3AAID-CBIC666%3E3.0.CO%3B2-%23.

[16] Thermo Scientific. Crosslinking Technical Handbook. Technical report, Thermo Fisher Scientific, October 2012. URL https://tools.thermofisher.com/content/sfs/brochures/1602163-Crosslinking-Reagents-Handbook.pdf.

[17] Halina Borel, Takeshi Sasaki, David B Stollar, and Yves Borel. Conjugation of DNA fragments to protein carriers by glutaraldehyde: Immunogenicity of oligonucleotide-hemocyanin conjugates. *Journal of Immunological Methods*, 67(2):289–302, 1984. doi: 10.1016/0022-1759(84)90469-1. URL http://linkinghub.elsevier.com/retrieve/pii/0022175984904691.

[18] Berea A R Williams and John C Chaput. *Synthesis of Peptide-Oligonucleotide Conjugates Using a Heterobifunctional Crosslinker*, volume 18. John Wiley & Sons, Inc., Hoboken, NJ, USA, May 2001. ISBN 0471142700. doi: 10.1002/0471142700.nc0441s42. URL http://doi.wiley.com/10.1002/0471142700.nc0441s42.

[19] Duyoung Min, Mark A Arbing, Robert E Jefferson, and James U Bowie. A simple DNA handle attachment method for single molecule mechanical manipulation experiments. *Protein Science*, 25 (8):1535–1544, June 2016. doi: 10.1002/pro.2952. URL http://doi.wiley.com/10.1002/pro.2952.

[20] Ken Halvorsen, Diane Schaak, and Wesley P Wong. Nanoengineering a single-molecule mechanical switch using DNA self-assembly. *Nanotechnology*, 22(49):494005–9, November 2011. doi: 10.1088/0957-4484/22/49/494005. URL http://stacks.iop.org/0957-4484/22/i=49/a=494005?key=crossref.2edf23f2d98758bda483307c527fdf85.

[21] Igor A Kozlov, Peter C Melnyk, Katie E Stromsborg, Mark S Chee, David L Barker, and Chanfeng Zhao. Efficient strategies for the conjugation of oligonucleotides to antibodies enabling highly sensitive protein detection. *Biopolymers*, 73(5):621–630, 2004. doi: 10.1002/bip.20009. URL http://doi.wiley.com/10.1002/bip.20009.

[22] D R Corey and P G Schultz. Generation of a hybrid sequence-specific single-stranded deoxyribonuclease. *Science*, 238(4832):1401–1403, December 1987. doi: 10.1126/science.3685986. URL http://www.sciencemag.org/cgi/doi/10.1126/science.3685986.

[23] Ciro Cecconi, Elizabeth A Shank, Frederick W Dahlquist, Susan Marqusee, and Carlos Bustamante. Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J*, 37(6):729–738, January 2008. doi: 10.1007/s00249-007-0247-y. URL http://link.springer.com/10.1007/s00249-007-0247-y.

[24] Yuhe Renee Yang, Soma Dhakal, Zhao Zhao, Minghui Liu, Ting Zhang, Nils G Walter, Hao Yan, and Jinglin Fu. Assembly of multienzyme complexes on DNA nanostructures. *Nature Protocols*, 11(11): 2243–2273, October 2016. doi: 10.1038/nprot.2016.139. URL http://dx.doi.org/10.1038/nprot.2016.139.

[25] K L Kiick, E Saxon, D A Tirrell, and C R Bertozzi. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proceedings of the National Academy of Sciences*, 99(1):19–24, 2001. doi: 10.1073/pnas.012583299.

[26] Kathrin Lang and Jason W Chin. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.*, 114(9):4764–4806, February 2014. doi: 10.1021/cr400355w. URL http://pubs.acs.org/doi/10.1021/cr400355w.

[27] Kimberly E Beatty, Fang Xie, Qian Wang, and David A Tirrell. Selective Dye-Labeling of Newly Synthesized Proteins in Bacterial Cells. *J. Am. Chem. Soc.*, 127(41):14150–14151, 2005. doi: 10.1021/ja054643w.

[28] P Bechtluft, R G H van Leeuwen, M Tyreman, D Tomkiewicz, N Nouwen, H L Tepper, A J M Driessen, and S J Tans. Direct Observation of Chaperone-Induced Changes in a Protein Folding Pathway. *Science*, 318(5855):1458–1461, 2007. doi: 10.1126/science.1144972.

[29] Fatemeh Moayed, Alireza Mashaghi, and Sander J Tans. A Polypeptide-DNA Hybrid with Selective Linking Capability Applied to Single Molecule Nano-Mechanical Measurements Using Optical Tweezers. *PLoS ONE*, 8(1):e54440–7, January 2013. doi: 10.1371/journal.pone.0054440. URL http://dx.plos.org/10.1371/journal.pone.0054440.

[30] Antje Keppler, Susanne Gendreizig, Thomas Gronemeyer, Horst Pick, Horst Vogel, and Kai Johnsson. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology*, 21(1):86–89, December 2002. doi: 10.1038/nbt765. URL http://www.nature.com/articles/nbt765.

[31] Brian S Goodman and Samara L Reck-Peterson. Engineering Defined Motor Ensembles with DNA Origami. Elsevier, 2014. ISBN 9780123979247. doi: 10.1016/b978-0-12-397924-7.00010-8. URL http://linkinghub.elsevier.com/retrieve/pii/B9780123979247000108.

[32] Arnaud Gautier, Alexandre Juillerat, Christian Heinis, Ivan Reis Corrêa Jr., Maik Kindermann, Florent Beaufils, and Kai Johnsson. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chemistry & Biology*, 15(2):128–136, February 2008. doi: 10.1016/j.chembiol.2008.01.007. URL http://linkinghub.elsevier.com/retrieve/pii/S1074552108000410.

[33] Marie-Eve Aubin-Tam, Adrian O Olivares, Robert T Sauer, Tania A Baker, and Matthew J Lang. Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. *Cell*, 145(2):257–267, 2011. doi: 10.1016/j.cell.2011.03.036. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867411003138.

[34] Nobuhide Doi and Hiroshi Yanagawa. STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro. *FEBS Letters*, 457(2):227–230, August 1999. doi: 10.1016/S0014-5793(99)01041-8. URL https://febs.onlinelibrary.wiley.com/doi/full/10.1016/S0014-5793%2899%2901041-8.

[35] Eiji Nakata, Fong Fong Liew, Chisana Uwatoko, Shigeki Kiyonaka, Yasuo Mori, Yousuke Katsuda, Masayuki Endo, Hiroshi Sugiyama, and Takashi Morii. Zinc-Finger Proteins for Site-Specific Protein Positioning on DNA-Origami Structures. *Angewandte Chemie International Edition*, 51(10):2421–2424, 2012. doi: 10.1002/anie.201108199.

[36] Tien Anh Ngo, Eiji Nakata, Masayuki Saimura, and Takashi Morii. Spatially Organized Enzymes Drive Cofactor-Coupled Cascade Reactions. *J. Am. Chem. Soc.*, 138(9):3012–3021, 2016. doi: 10.1021/jacs.5b10198.

[37] Rahul Chhabra, Jaswinder Sharma, Yonggang Ke, Yan Liu, Sherri Rinker, Stuart Lindsay, and Hao Yan. Spatially Addressable Multiprotein Nanoarrays Templated by Aptamer-Tagged DNA Nanoarchitectures. *J. Am. Chem. Soc.*, 129(34):10304–10305, 2007. doi: 10.1021/ja072410u.

[38] Sherri Rinker, Yonggang Ke, Yan Liu, Rahul Chhabra, and Hao Yan. Self-assembled DNA nanostructures for distance-dependent multivalent ligand–protein binding. *Nature Nanotechnology*, 3(7):418–422, June 2008. doi: 10.1038/nnano.2008.164. URL http://www.nature.com/articles/nnano.2008.164.

[39] Vsevolod V Rostovtsev, Luke G Green, Valery V Fokin, and K Barry Sharpless. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition*, 41(14):2596–2599, 2002. doi: 10.1002/1521-3773(20020715)41:14⟨2596::aid-anie2596⟩3.0.co;2-4. URL http://doi.wiley.com/10.1002/1521-3773%2820020715%2941%3A14%3C2596%3A%3AAID-ANIE2596%3E3.0.CO%3B2-4.

[40] Christian W Tornøe, Caspar Christensen, and Morten Meldal. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.*, 67(9):3057–3064, May 2002. doi: 10.1021/jo011148j. URL http://pubs.acs.org/doi/abs/10.1021/jo011148j.

[41] Harita Rao, Arun A Tanpure, Anupam A Sawant, and Seergazhi G Srivatsan. Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post- transcriptional chemical functionalization. *Nature Protocols*, 7(6):1097–1112, May 2012. doi: 10.1038/nprot.2012.046. URL http://dx.doi.org/10.1038/nprot.2012.046.

[42] Nicholas J Agard, Jennifer A Prescher, and Carolyn R Bertozzi. A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.*, 126(46):15046–15047, November 2004. doi: 10.1021/ja044996f. URL http://pubs.acs.org/doi/abs/10.1021/ja044996f.

[43] Melissa L Blackman, Maksim Royzen, and Joseph M Fox. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.*, 130(41):13518–13519, October 2008. doi: 10.1021/ja8053805. URL http://pubs.acs.org/doi/abs/10.1021/ja8053805.

[44] Raffaella Rossin, Sandra M van den Bosch, Wolter ten Hoeve, Marco Carvelli, Ron M Versteegen, Johan Lub, and Marc S Robillard. Highly Reactive trans-Cyclooctene Tags with Improved Stability for Diels–Alder Chemistry in Living Systems. *Bioconjugate Chem.*, 24(7):1210–1217, June 2013. doi: 10.1021/bc400153y. URL http://pubs.acs.org/doi/10.1021/bc400153y.

[45] Robert J Blizzard, Dakota R Backus, Wes Brown, Christopher G Bazewicz, Yi Li, and Ryan A Mehl. Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid. *J. Am. Chem. Soc.*, 137(32):10044–10047, August 2015. doi: 10.1021/jacs.5b03275. URL http://pubs.acs.org/doi/10.1021/jacs.5b03275.

[46] Monicia Elrod-Erickson and Carl O Pabo. Binding Studies with Mutants of Zif268. *J. Biol. Chem.*, 274(27):19281–19285, 1999. doi: 10.1074/jbc.274.27.19281. URL http://www.jbc.org/lookup/doi/10.1074/jbc.274.27.19281.

[47] O H Laitinen, V P Hytönen, H R Nordlund, and M S Kulomaa. Genetically engineered avidins and streptavidins. *Cell. Mol. Life Sci.*, 63(24):2992–3017, November 2006. doi: 10.1007/s00018-006-6288-z. URL http://link.springer.com/10.1007/s00018-006-6288-z.

[48] C Niemeyer. Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR. *Nucleic Acids Research*, 27(23):4553–4561, 1999. doi: 10.1093/nar/27.23.4553. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/27.23.4553.

[49] Michael Fairhead and Mark Howarth. Site-Specific Biotinylation of Purified Proteins Using BirA. In Arnaud Gautier and Marlon J Hinner, editors, *Site-Specific Protein Labeling: Methods and Protocols*, pages 171–184. Springer New York, New York, NY, 2015. URL f6z9mj.

[50] Claire E Chivers, Estelle Crozat, Calvin Chu, Vincent T Moy, David J Sherratt, and Mark Howarth. A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat Methods*, 7 (5):391–393, 2010. doi: 10.1038/nmeth.1450.

[51] Georgyi V Los, Lance P Encell, Mark G McDougall, Danette D Hartzell, Natasha Karassina, Chad Zimprich, Monika G Wood, Randy Learish, Rachel Friedman Ohana, Marjeta Urh, Dan Simpson, Jacqui Méndez, Kris Zimmerman, Paul Otto, Gediminas Vidugiris, Ji Zhu, Aldis Darzins, Dieter H Klaubert, Robert F Bulleit, and Keith V Wood. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chemical Biology*, 3(6):373–382, June 2008. doi: 10.1021/cb800025k. URL http://pubs.acs.org/doi/abs/10.1021/cb800025k.

[52] Klaus N Lovendahl, Amanda N Hayward, and Wendy R Gordon. Sequence-Directed Covalent Protein–DNA Linkages in a Single Step Using HUH-Tags. *J. Am. Chem. Soc.*, 139(20):7030–7035, May 2017. doi: 10.1021/jacs.7b02572. URL http://pubs.acs.org/doi/10.1021/jacs.7b02572.

[53] Michael Chandler, Fernando de la Cruz, Fred Dyda, Alison B Hickman, Gabriel Moncalian, and Bao Ton-Hoang. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature Reviews Microbiology*, 11(8):525–538, 2013. doi: 10.1038/nrmicro3067.

[54] Alicia Guasch, María Lucas, Gabriel Moncalian, Matilde Cabezas, Rosa Pérez-Luque, F Xavier Gomis-Rüth, Fernando de la Cruz, and Miquel Coll. Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nature Structural & Molecular Biology*, 10(12):1002–1010, 2003. doi: 10.1038/nsb1017.

[55] Chris Larkin, Saumen Datta, Matthew J Harley, Brian J Anderson, Alexandra Ebie, Victoria Hargreaves, and Joel F Schildbach. Inter- and Intramolecular Determinants of the Specificity of Single-Stranded DNA Binding and Cleavage by the F Factor Relaxase. *Structure*, 13(10):1533–1544, October 2005. doi: 10.1016/j.str.2005.06.013. URL http://linkinghub.elsevier.com/retrieve/pii/S096921260500273X.

[56] Roeland Boer, Silvia Russi, Alicia Guasch, María Lucas, Alexandre G Blanco, Rosa Pérez-Luque, Miquel Coll, and Fernando de la Cruz. Unveiling the Molecular Mechanism of a Conjugative Relaxase: The Structure of TrwC Complexed with a 27-mer DNA Comprising the Recognition Hairpin and the Cleavage Site. *Journal of Molecular Biology*, 358(3):857–869, 2006. doi: 10.1016/j.jmb.2006.02.018.

[57] Jonathan S Edwards, Laurie Betts, Monica L Frazier, Rebecca M Pollet, Stephen M Kwong, William G Walton, W Keith Ballentine, Julianne J Huang, Sohrab Habibi, Mark Del Campo, Jordan L Meier, Peter B Dervan, Neville Firth, and Matthew R Redinbo. Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. *Proceedings of the National Academy of Sciences*, 110(8):2804–2809, February 2013. doi: 10.1073/pnas.1219701110. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1219701110.

[58] Radoslaw Pluta, D Roeland Boer, Fabián Lorenzo-Díaz, Silvia Russi, Hansel Gómez, Cris Fernández-López, Rosa Pérez-Luque, Modesto Orozco, Manuel Espinosa, and Miquel Coll. Structural basis of a histidine-DNA nicking/joining mechanism for gene transfer and promiscuous spread of antibiotic resistance. *Proceedings of the National Academy of Sciences*, 114(32):E6526–E6535, August 2017. doi: 10.1073/pnas.1702971114. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1702971114.

[59] Susana Vega-Rocha, In-Ja L Byeon, Bruno Gronenborn, Angela M Gronenborn, and Ramón Campos-Olivas. Solution Structure, Divalent Metal and DNA Binding of the Endonuclease Domain from the Replication Initiation Protein from Porcine Circovirus 2. *Journal of Molecular Biology*, 367(2):473–487, 2007. doi: 10.1016/j.jmb.2007.01.002. URL http://linkinghub.elsevier.com/retrieve/pii/S0022283607000034.

[60] Arthur F Monzingo, Angela Ozburn, Shuangluo Xia, Richard J Meyer, and Jon D Robertus. The Structure of the Minimal Relaxase Domain of MobA at 2.1 Å Resolution. *Journal of Molecular Biology*, 366(1):165–178, 2007. doi: 10.1016/j.jmb.2006.11.031.

[61] Heping Zheng, Maksymilian Chruszcz, Piotr Lasota, Lukasz Lebioda, and Wladek Minor. Data mining of metal ion environments present in protein structures. *Journal of Inorganic Biochemistry*, 102(9):1765–1776, 2008. doi: 10.1016/j.jinorgbio.2008.05.006.

[62] Saumen Datta, Chris Larkin, and Joel F Schildbach. Structural Insights into Single-Stranded DNA Binding and Cleavage by F Factor TraI. *Structure*, 11(11):1369–1379, 2003. doi: 10.1016/j.str.2003.10.001.

[63] Shuangluo Xia and Jon D Robertus. Effect of divalent ions on the minimal relaxase domain of MobA. *Archives of Biochemistry and Biophysics*, 488(1):42–47, August 2009. doi: 10.1016/j.abb.2009.06.004. URL http://dx.doi.org/10.1016/j.abb.2009.06.004.

[64] Ethel S Tessman. Mutants of bacteriophage S13 blocked in infectious DNA synthesis. *Journal of Molecular Biology*, 17(1):218–236, 1966. doi: 10.1016/s0022-2836(66)80104-3. URL http://linkinghub.elsevier.com/retrieve/pii/S0022283666801043.

[65] S Eisenberg and A Kornberg. Purification and characterization of phiX174 gene A protein. A multifunctional enzyme of duplex DNA replication. *J. Biol. Chem.*, 254(12):5328–5332, June 1979. URL http://www.jbc.org/content/254/12/5328.abstract.

[66] Farhima Akter, Masayasu Mie, Sebastian Grimm, Per-Åke Nygren, and Eiry Kobatake. Detection of Antigens Using a Protein–DNA Chimera Developed by Enzymatic Covalent Bonding with phiX Gene A*. *Analytical Chemistry*, 84(11):5040–5046, 2012. doi: 10.1021/ac300708r.

[67] Michael J Liszka, Melinda E Clark, Elizabeth Schneider, and Douglas S Clark. Nature Versus Nurture: Developing Enzymes That Function Under Extreme Conditions. *Annual Review of Chemical and Biomolecular Engineering*, 3(1):77–102, 2012. doi: 10.1146/annurev-chembioeng-061010-114239.

[68] Birgit Mollwitz, Elizabeth Brunk, Simone Schmitt, Florence Pojer, Michael Bannwarth, Marc Schiltz, Ursula Rothlisberger, and Kai Johnsson. Directed Evolution of the Suicide Protein O6-Alkylguanine-DNA Alkyltransferase for Increased Reactivity Results in an Alkylated Protein with Exceptional Stability. *Biochemistry*, 51(5):986–994, January 2012. doi: 10.1021/bi2016537. URL http://pubs.acs.org/doi/10.1021/bi2016537.

[69] Alexandre Juillerat, Thomas Gronemeyer, Antje Keppler, Susanne Gendreizig, Horst Pick, Horst Vogel, and Kai Johnsson. Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chemistry & Biology*, 10(4):313–317, April 2003. doi: 10.1016/S1074-5521(03)00068-1. URL http://linkinghub.elsevier.com/retrieve/pii/S1074552103000681.

[70] John A Zuris, David B Thompson, Yilai Shu, John P Guilinger, Jeffrey L Bessen, Johnny H Hu, Morgan L Maeder, J Keith Joung, Zheng-Yi Chen, and David R Liu. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature Biotechnology*, 33(1):73, 2015.

[71] Sandra Sagredo, Tobias Pirzer, Ali Aghebat Rafat, Marisa A Goetzfried, Gabriel Moncalian, Friedrich C Simmel, and Fernando de la Cruz. Orthogonale Assemblierung von Proteinen auf DNA-Nanostrukturen mithilfe von Relaxasen. *Angewandte Chemie*, 128(13):4421–4425, 2016. doi: 10.1002/ange.201510313.

[72] Giulio Bernardinelli and Björn Högberg. Entirely enzymatic nanofabrication of DNA–protein conjugates. *Nucleic Acids Research*, 45(18):e160–e160, August 2017. doi: 10.1093/nar/gkx707. URL http://academic.oup.com/nar/article/45/18/e160/4082854.

[73] Eric J Aird, Klaus N Lovendahl, Amber St Martin, Reuben S Harris, and Wendy R Gordon. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Communications Biology*, 1(1):816, 2018. doi: 10.1038/s42003-018-0054-2. URL http://www.nature.com/articles/s42003-018-0054-2.

[74] M Jinek, K Chylinski, I Fonfara, M Hauer, J A Doudna, and E Charpentier. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096):816–821, August 2012. doi: 10.1126/science.1225829. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1225829.

[75] Martin Jinek, Alexandra East, Aaron Cheng, Steven Lin, Enbo Ma, and Jennifer Doudna. RNA-programmed genome editing in human cells. *eLife*, 2:273–9, January 2013. doi: 10.7554/elife.00471. URL https://elifesciences.org/articles/00471.

[76] A Kakarougkas and P A Jeggo. DNA DSB repair pathway choice: an orchestrated handover mechanism. *The British Journal of Radiology*, 87(1035):20130685, 2014. doi: 10.1259/bjr.20130685.

[77] Fuqiang Chen, Shondra M Pruett-Miller, Yuping Huang, Monika Gjoka, Katarzyna Duda, Jack Taunton, Trevor N Collingwood, Morten Frodin, and Gregory D Davis. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods*, 8(9):753–755, 2011. doi: 10.1038/nmeth.1653.

[78] Natasa Savic, Femke CAS Ringnalda, Helen Lindsay, Christian Berk, Katja Bargsten, Yizhou Li, Dario Neri, Mark D Robinson, Constance Ciaudo, Jonathan Hall, Martin Jinek, and Gerald Schwank. Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *eLife*, 7:1–18, May 2018. doi: 10.7554/eLife.33761. URL https://elifesciences.org/articles/33761.

[79] Diego S D'Astolfo, Romina J Pagliero, Anita Pras, Wouter R Karthaus, Hans Clevers, Vikram Prasad, Robert Jan Lebbink, Holger Rehmann, and Niels Geijsen. Efficient Intracellular Delivery of Native Proteins. *Cell*, 161(3):674–690, 2015. doi: 10.1016/j.cell.2015.03.028. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867415003153.

[80] Carolin Anders, Ole Niewoehner, and Martin Jinek. In Vitro Reconstitution and Crystallization of Cas9 Endonuclease Bound to a Guide RNA and a DNA Target. In *Methods in Enzymology*, pages 515–537. Methods in Enzymology, 2015. doi: 10.1016/bs.mie.2015.02.008.

[81] F William Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1):207–234, 2005. doi: 10.1016/j.pep.2005.01.016.

[82] Shawn M Douglas, Adam H Marblestone, Surat Teerapittayanon, Alejandro Vazquez, George M Church, and William M Shih. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Research*, 37(15):5001–5006, June 2009. doi: 10.1093/nar/gkp436. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkp436.

[83] Wendy R Gordon, Brandon Zimmerman, Li He, Laura J Miles, Jiuhong Huang, Kittichoat Tiyanont, Debbie G McArthur, Jon C Aster, Norbert Perrimon, Joseph J Loparo, and Stephen C Blacklow. Mechanical Allostery: Evidence for a Force Requirement in the Proteolytic Activation of Notch. *Developmental Cell*, 33(6):729–736, 2015. doi: 10.1016/j.devcel.2015.05.004. URL http://linkinghub.elsevier.com/retrieve/pii/S1534580715003202.

[84] A M Gronenborn, D R Filpula, N Z Essig, A Achari, M Whitlow, P T Wingfield, and G M Clore. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science*, 253(5020):657–661, August 1991. doi: 10.1126/science.1871600. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1871600.

[85] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: an open-source platform for biological-image analysis. *Nat Methods*, 9(7):676–682, July 2012. doi: 10.1038/nmeth.2019. URL https://www.nature.com/articles/nmeth.2019.

[86] Gary Struhl and Atsuko Adachi. Nuclear Access and Action of Notch In Vivo. *Cell*, 93(4):649–660, 1998. doi: 10.1016/s0092-8674(00)81193-9. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867400811939.

[87] U T Bornscheuer, G W Huisman, R J Kazlauskas, S Lutz, J C Moore, and K Robins. Engineering the third wave of biocatalysis. *Nature*, 485(7397):185–194, 2012. doi: 10.1038/nature11117. URL http://www.nature.com/doifinder/10.1038/nature11117.

[88] Zhen Chen and An-Ping Zeng. Protein engineering approaches to chemical biotechnology. *Current Opinion in Biotechnology*, 42:198–205, 2016. doi: 10.1016/j.copbio.2016.07.007. URL http://linkinghub.elsevier.com/retrieve/pii/S0958166916301756.

[89] Michael S Packer and David R Liu. Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16(7):379–394, 2015. doi: 10.1038/nrg3927.

[90] G P Smith. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705):1315–1317, June 1985. doi: 10.1126/science.4001944. URL http://www.sciencemag.org/cgi/doi/10.1126/science.4001944.

[91] Eric T Boder and K Dane Wittrup. Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology*, 15(6):553–557, June 1997. doi: 10.1038/nbt0697-553. URL http://www.nature.com/doifinder/10.1038/nbt0697-553.

[92] Benjamin T Porebski and Ashley M Buckle. Consensus protein design. *Protein Engineering, Design and Selection*, 29(7):245–251, 2016. doi: 10.1093/protein/gzw015.

[93] Gautam Dantas, Colin Corrent, Steve L Reichow, James J Havranek, Ziad M Eletr, Nancy G Isern, Brian Kuhlman, Gabriele Varani, Ethan A Merritt, and David Baker. High-resolution Structural and Thermodynamic Analysis of Extreme Stabilization of Human Procarboxypeptidase by Computational Protein Design. *Journal of Molecular Biology*, 366(4):1209–1221, 2007. doi: 10.1016/j.jmb.2006.11.080. URL http://linkinghub.elsevier.com/retrieve/pii/S0022283606016391.

[94] Will Sheffler and David Baker. RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design and validation. *Protein Science*, 18(1):NA–NA, 2008. doi: 10.1002/pro.8. URL http://doi.wiley.com/10.1002/pro.8.

[95] Eric M Brustad and Frances H Arnold. Optimizing non-natural protein function with directed evolution. *Current Opinion in Chemical Biology*, 15(2):201–210, 2011. doi: 10.1016/j.cbpa.2010.11.020. URL http://linkinghub.elsevier.com/retrieve/pii/S1367593110001948.

[96] Boris Steipe, Britta Schiller, Andreas Plückthun, and Stefan Steinbacher. Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain. *Journal of Molecular Biology*, 240(3):188–192, 1994. doi: 10.1006/jmbi.1994.1434. URL http://linkinghub.elsevier.com/retrieve/pii/S0022283684714343.

[97] Stefan Lutz. Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology*, 21(6):734–743, 2010. doi: 10.1016/j.copbio.2010.08.011. URL http://linkinghub.elsevier.com/retrieve/pii/S0958166910001540.

[98] Anthony H Keeble, Anusuya Banerjee, Matteo P Ferla, Samuel C Reddington, Irsyad N A Khairil Anuar, and Mark Howarth. Evolving Accelerated Amidation by SpyTag/SpyCatcher to Analyze Membrane Dynamics. *Angewandte Chemie International Edition*, 56(52):16521–16525, December 2017. doi: 10.1002/anie.201707623. URL https://onlinelibrary.wiley.com/doi/full/10.1002/anie.201707623.

[99] Marie-Françoise Noirot-Gros and Stanislav D Ehrlich. Change of a Catalytic Reaction Carried Out by a DNA Replication Protein. *Science*, 274(5288):777–780, November 1996. doi: 10.1126/science.274.5288.777. URL http://www.sciencemag.org/cgi/doi/10.1126/science.274.5288.777.

[100] Benjamin Borgo and James J Havranek. Automated selection of stabilizing mutations in designed and natural proteins. *Proceedings of the National Academy of Sciences*, 109(5):1494–1499, January 2012. doi: 10.1073/pnas.1115172109. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1115172109.

[101] Adi Goldenzweig, Moshe Goldsmith, Shannon E Hill, Or Gertman, Paola Laurino, Yacov Ashani, Orly Dym, Tamar Unger, Shira Albeck, Jaime Prilusky, Raquel L Lieberman, Amir Aharoni, Israel Silman, Joel L Sussman, Dan S Tawfik, and Sarel J Fleishman. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell*, 63(2):337–346, 2016. doi: 10.1016/j.molcel.2016.06.012. URL http://linkinghub.elsevier.com/retrieve/pii/S109727651630243X.

[102] Rhiju Das and David Baker. Macromolecular Modeling with Rosetta. *Annual Review of Biochemistry*, 77(1):363–382, 2008. doi: 10.1146/annurev.biochem.77.062906.171838.

[103] L J Perry and R Wetzel. Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science*, 226(4674):555–557, November 1984. doi: 10.1126/science.6387910. URL http://www.sciencemag.org/cgi/doi/10.1126/science.6387910.

[104] Walter A Baase, Lijun Liu, Dale E Tronrud, and Brian W Matthews. Lessons from the lysozyme of phage T4. *Protein Science*, 19(4):631–641, 2010. doi: 10.1002/pro.344.

[105] Alan A Dombkowski, Kazi Zakia Sultana, and Douglas B Craig. Protein disulfide engineering. *FEBS Letters*, 588(2):206–212, January 2014. doi: 10.1016/j.febslet.2013.11.024. URL https://febs.onlinelibrary.wiley.com/doi/full/10.1016/j.febslet.2013.11.024.

[106] Ying Yu and Stefan Lutz. Circular permutation: a different way to engineer enzyme structure and function. *Trends in Biotechnology*, 29(1):18–25, 2011. doi: 10.1016/j.tibtech.2010.10.004. URL http://linkinghub.elsevier.com/retrieve/pii/S0167779910001782.

[107] Nataša Poklar, Jurij Lah, Mateja Salobir, Peter Maček, and Gorazd Vesnaver. pH and Temperature-Induced Molten Globule-Like Denatured States of Equinatoxin II: A Study by UV-Melting, DSC, Far- and Near-UV CD Spectroscopy, and ANS Fluorescence†. *Biochemistry*, 36(47):14345–14352, 1997. doi: 10.1021/bi971719v. URL http://pubs.acs.org/doi/abs/10.1021/bi971719v.

[108] Frank H Niesen, Helena Berglund, and Masoud Vedadi. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9):2212–2221, 2007. doi: 10.1038/nprot.2007.321.

[109] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. November 2015. URL https://pymol.org/2/.

[110] Sergey Lyskov, Fang-Chieh Chou, Shane Ó Conchúir, Bryan S Der, Kevin Drew, Daisuke Kuroda, Jianqing Xu, Brian D Weitzner, P Douglas Renfrew, Parin Sripakdeevong, Benjamin Borgo, James J Havranek, Brian Kuhlman, Tanja Kortemme, Richard Bonneau, Jeffrey J Gray, and Rhiju Das. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS ONE*, 8(5):e63906, 2013. doi: 10.1371/journal.pone.0063906.

[111] L Gama and G E Breitwieser. Generation of epitope-tagged proteins by inverse PCR mutagenesis. *BioTechniques*, 26(5):814–816, May 1999. URL http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10337467&retmode=ref&cmd=prlinks.

[112] Charlie Gosse and Vincent Croquette. Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level. *Biophys. J.*, 82(6):3314–3329, June 2002. doi: 10.1016/S0006-3495(02)75672-5. URL http://dx.doi.org/10.1016/S0006-3495(02)75672-5.

[113] Iwijn De Vlaminck and Cees Dekker. Recent Advances in Magnetic Tweezers. *Annu. Rev. Biophys.*, 41(1):453–472, 2012. doi: 10.1146/annurev-biophys-122311-100544.

[114] Richard P Feynman, Robert B Leighton, and Matthew L Sands. *The Feynman lectures on physics (online edition)*. http://www.feynmanlectures.caltech.edu/, Pasadena, CA, 1963.

[115] Keir C Neuman and Attila Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods*, 5(6):491–505, June 2008. doi: 10.1038/nmeth.1218. URL http://www.nature.com/articles/nmeth.1218.

[116] T Lionnet, J F Allemand, A Revyakin, T R Strick, Omar A Saleh, D Bensimon, and Vincent Croquette. Single-Molecule Studies Using Magnetic Traps. *Cold Spring Harbor Protocols*, 2012(1):pdb.top067488–pdb.top067488, December 2011. doi: 10.1101/pdb.top067488. URL http://www.cshprotocols.org/cgi/doi/10.1101/pdb.top067488.

[117] Yeonee Seol, Marie-Paule Strub, and Keir C Neuman. Single molecule measurements of DNA helicase activity with magnetic tweezers and t-test based step-finding analysis. *Methods*, 105(C):119–127, August 2016. doi: 10.1016/j.ymeth.2016.04.030. URL http://dx.doi.org/10.1016/j.ymeth.2016.04.030.

[118] Keir C Neuman, T Lionnet, and J F Allemand. Single-Molecule Micromanipulation Techniques. *Annu. Rev. Mater. Res.*, 37(1):33–67, August 2007. doi: 10.1146/annurev.matsci.37.052506.084336. URL http://www.annualreviews.org/doi/10.1146/annurev.matsci.37.052506.084336.

[119] Kelly Williams, Brendan Grafe, Kathryn M Burke, Nathan Tanner, Antoine M van Oijen, Joseph Loparo, and Allen C Price. A single molecule DNA flow stretching microscope for undergraduates. *American Journal of Physics*, 79(11):1112–1120, 2011. doi: 10.1119/1.3620410.

[120] Yeonee Seol and Keir C Neuman. Combined Magnetic Tweezers and Micro-mirror Total Internal Reflection Fluorescence Microscope for Single-Molecule Manipulation and Visualization, 2017.

[121] Felix E Kemmerich, Marko Swoboda, Dominik J Kauert, M Svea Grieb, Steffen Hahn, Friedrich W Schwarz, Ralf Seidel, and Michael Schlierf. Simultaneous Single-Molecule Force and Fluorescence Sampling of DNA Nanostructure Conformations Using Magnetic Tweezers. *Nano Lett.*, 16(1):381–386, 2015. doi: 10.1021/acs.nanolett.5b03956.

[122] L Sacconi, G Romano, R Ballerini, M Capitanio, M De Pas, M Giuntini, D Dunlap, L Finzi, and F S Pavone. Three-dimensional magneto-optic trap for micro-object manipulation. *Optics Letters*, 26(17): 1359, 2001. doi: 10.1364/ol.26.001359.

[123] Claudia Danilowicz, Derek Greenfield, and Mara Prentiss. Dissociation of Ligand–Receptor Complexes Using Magnetic Tweezers. *Analytical Chemistry*, 77:3023–3028, April 2005. doi: 10.1021/ac050057. URL https://pubs.acs.org/doi/abs/10.1021/ac050057%2B.

[124] S Earnshaw. On the nature of the molecular forces which regulate the constitution of the luminiferous ether. *Trans. Camb. Phil. Soc*, 7:97–112, 1842.

[125] Erik Schäffer, Simon F Nørrelykke, and Jonathon Howard. Surface Forces and Drag Coefficients of Microspheres near a Plane Surface Measured with Optical Tweezers. *Langmuir*, 23(7):3654–3665, March 2007. doi: 10.1021/la0622368. URL http://pubs.acs.org/doi/abs/10.1021/la0622368.

[126] M Manosas, J D Wen, P T X Li, S B Smith, Carlos Bustamante, I Tinoco Jr, and F Ritort. Force Unfolding Kinetics of RNA using Optical Tweezers. II. Modeling Experiments. *Biophys. J.*, 92(9):3010–3021, May 2007. doi: 10.1529/biophysj.106.094243. URL http://dx.doi.org/10.1529/biophysj.106.094243.

[127] Steven B Smith, Yujia Cui, and Carlos Bustamante. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science*, 271(5250):795–799, February 1996. doi: 10.1126/science.271.5250.795. URL http://www.sciencemag.org/cgi/doi/10.1126/science.271.5250.795.

[128] Duyoung Min, Kipom Kim, Changbong Hyeon, Yong Hoon Cho, Yeon-Kyun Shin, and Tae-Young Yoon. Mechanical unzipping and rezipping of a single SNARE complex reveals hysteresis as a force-generating mechanism. *Nature Communications*, 4(1):1705, April 2013. doi: 10.1038/ncomms2692. URL https://www.nature.com/articles/ncomms2692.

[129] Ankur Jain, Ruijie Liu, Yang K Xiang, and Taekjip Ha. Single-molecule pull-down for studying protein interactions. *Nature Protocols*, 7(3):445–452, March 2012. doi: 10.1038/nprot.2011.452. URL https://www.nature.com/articles/nprot.2011.452.

[130] Bojk A Berghuis, Mariana Köber, Theo van Laar, and Nynke H Dekker. High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers. *Methods*, 105:90–98, 2016. doi: 10.1016/j.ymeth.2016.03.025. URL http://linkinghub.elsevier.com/retrieve/pii/S1046202316300597.

[131] Kelly P Williams and David P Bartel. PCR product with strands of unequal length. *Nucleic Acids Research*, 23(20):4220–4221, 1995. doi: 10.1093/nar/23.20.4220. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/23.20.4220.

[132] Arun Richard Chandrasekaran, Oksana Levchenko, Dhruv S Patel, Molly MacIsaac, and Ken Halvorsen. Addressable configurations of DNA nanostructures for rewritable memory. *Nucleic Acids Research*, 45(19):11459–11465, August 2017. doi: 10.1093/nar/gkx777. URL http://academic.oup.com/nar/article/45/19/11459/4097620.

[133] T I Přistoupil, M Kramlová, and J Štěrbíková. On the mechanism of adsorption of proteins to nitrocellulose in membrane chromatography. *Journal of chromatography A*, 42:367–375, 1969. doi: 10.1016/s0021-9673(01)80636-1. URL http://linkinghub.elsevier.com/retrieve/pii/S0021967301806361.

[134] Raghuveer Parthasarathy. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat Methods*, 9(7):724–726, 2012. doi: 10.1038/nmeth.2071.

[135] T R Strick, J F Allemand, D Bensimon, A Bensimon, and Vincent Croquette. The Elasticity of a Single Supercoiled DNA Molecule. *Science*, 271(5257):1835–1837, March 1996. doi: 10.1126/science.271.5257.1835. URL http://www.sciencemag.org/cgi/doi/10.1126/science.271.5257.1835.

[136] Richard Janissen, Bojk A Berghuis, David Dulin, Max Wink, Theo van Laar, and Nynke H Dekker. Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments. *Nucleic Acids Research*, 42(18):e137–e137, August 2014. doi: 10.1093/nar/gku677. URL http://academic.oup.com/nar/article/42/18/e137/2434509/Invincible-DNA-tethers-covalent-DNA-anchoring-for.

[137] G Neuert, C Albrecht, E Pamir, and H E Gaub. Dynamic force spectroscopy of the digoxigenin-antibody complex. *FEBS Letters*, 580(2):505–509, 2005. doi: 10.1016/j.febslet.2005.12.052.

[138] Jorine M Eeftens, Jaco van der Torre, Daniel R Burnham, and Cees Dekker. Copper-free click chemistry for attachment of biomolecules in magnetic tweezers. *BMC Biophysics*, pages 1–7, August 2017. doi: 10.1186/s13628-015-0023-9. URL http://dx.doi.org/10.1186/s13628-015-0023-9.

[139] C G Baumann, S B Smith, V A Bloomfield, and Carlos Bustamante. Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences*, 94(12):6185–6190, 1997. doi: 10.1073/pnas.94.12.6185. URL http://www.pnas.org/cgi/doi/10.1073/pnas.94.12.6185.

List of abbreviations

AP alkaline phosphatase 2

BME β-mercaptoethanol 14, 24, 25

CV column volumes 52

DTT Dithiothreitol 24

E. coli *Escherichia coli* 24, 52

ELISA enzyme-linked immunosorbent assay 2, 11

FPLC fast protein liquid chromatography 52

gpA* gene A protein 11

HRP horseradish peroxidase 2

IDT Integrated DNA Technologies 24

IPTG Isopropyl β-D-1-thiogalactopyranoside 24

NEB New England Biolabs 13, 24, 52

PCV2 porcine circovirus 2 10, 11

SDS Sodium dodecyl sulfate 25, 27

ssDNA single-stranded DNA 9, 11, 13

TBS Tris-buffered saline 52

Supplementary data

Supplement to Chapter 2

Amino acid sequences of HUH-tags

DCV Rep protein from Muscovy duck circovirus
UniProt D2JZX8
MAKSGNYSYKRWVFTINNPTFEDYVHVLEFCTLDNCKFAIVGEEKGANGTPHLQG
FLNLRSNARAAALEESLGGRAWLSRARGSDEDNEEYCAKESTYLRVGEPVSKGRS
S

FBNYV Master replication protein from Faba bean necrotic yellows virus (isolate SV292-88)
UniProt O39828
PDB 2HWT
MARQVICWCFTLNNPLSPLSLHDSMKYLVYQTEQGEAGNIHFQGYIEMKKRTSLA
GMKKLIPGAHFEKRRGTQGEARAYSMKEDTRLEGPWEYGEFVP

miniMobA Minimal relaxase domain of mobilization protein A from *Escherichia coli* plasmid R1162
UniProt P07112
PDB 2NS6
MAIYHLTAKTGSRSGGQSARAKADYIQREGKYARDMDEVLHAESGHMPEFVERP
ADYWDAADLYERANGRLFKEVEFALPVELTLDQQKALASEFAQHLTGAERLPYT
LAIHAGGGENPHCHLMISERINDGIERPAAQWFKRYNGKTPEKGGAQKTEALKPK
AWLEQTREAWADHANRALERAGH

NES Nicking enzyme from *Staphylococcus aureus* plasmid pLW1043
UniProt Q53632
PDB 4HT4
MAMYHFQNKFVSKANGQSATAKSAYNSASRIKDFKENEFKDYSNKQCDYSEILLP
NNADDKFKDREYLWNKVHDVENRKNSQVAREIIIGLPNEFDPNSNIELAKEFAESL
SNEGMIVDLNIHKINEENPHAHLLCTLRGLDKNNEFEPKRKGNDYIRDWNTKEKH
NEWRKRWENVQNKHLEKNGFSVRVSADSYKNQNIDLEPTKKEGWKARKFEDETG

PCV2 Replication-associated protein from Porcine circovirus 2
UniProt Q8BB16
PDB 2HW0
SPSKKNGRSGPQPHKRWVFTLNNPSEDERKKIRDLPISLFDYFIVGEEGNEEGRTPH
LQGFANFVKKQTFNKVKWYLGARCHIEKAKGTDQQNKEYCSKEGNLLMECGAPR
SQGQR

RepB Replication protein B from *Streptococcus agalactiae* plasmid pMV158
UniProt P13921
PDB 3DKY
MAKEKARYFTFLLYPESIPSDWELKLETLGVPMAISPLHDKDKSSIKGQKYKKAHY
HVLYIAKNPVTADSVRKKIKLLLGEKSLAMVQVVLNVENMYLYLTHESKDAIAKK
KHVYDKADIKLINNFDIDRY

RepBm Plasmid replication protein from *Streptococcus pneumoniae*
UniProt A0A0T8A2Q2
MSEKKEIVKGRDWTFLVYPESAPENWRTILDETFMRWVESPLHDKDVNADGEIKK
PHWHILLSSDGPITQTAVQKIIGPLNAPNAQKVGSAKGLVRYMVHLDNPEKYQYSL
DEIVGHNGADVASYFELTA

TraI Multifunctional conjugation protein from *Escherichia coli* F plasmid
P14565
PDB 1P4D (apo structure), 2A0I (DNA-bound)
MMSIAQVRSAGSAGNYYTDKDNYYVLGSMGERWAGRGAEQLGLQGSVDKDVFT

RLLEGRLPDGADLSRMQDGSNRHRPGYDLTFSAPKSVSMMAMLGGDKRLIDAHN
QAVDFAVRQVEALASTRVMTDGQSETVLTGNLVMALFNHDTSRDQEPQLHTHAV
VANVTQHNGEWKTLSSDKVGKTGFIENVYANQIAFGRLYREKLKEQVEALGYETE
VVGKHGMWEMPGVPVEAFSGRSQTIREAVGEDASLKSRDVAALDTRKSKQHVDP
EIKMAEWMQTLKETGFDIRAYRDAADQRADLRTLTPGPASQDGPDVQQAVTQAIA
GLSER

Supplement to Chapter 3

Figure 1: **Uncropped gel images from yield assays** Gels 1 & 2 were performed on the same day; gel 3 was performed on the following day. Raw TIFF exports were converted to 8-bit format for display but are otherwise unaltered.

Supplement to Chapter 4

Polymer physics terms

(return to 4.1.3)

Persistence length (P) The characteristic length under which a polymer can be considered a flexible beam; above the persistence length a polymer must be treated as a chain.

Given its phosphate backbone DNA is a polyelectrolyte, and the persistence length of DNA is affected by the ionic strength of the environment. At physiological salt $P_{dsDNA} \approx 500\mathring{A}$, which at very low salt concentration rises to $\approx 900\mathring{A}$ [139]. The persistence length of single-stranded DNA is $\approx 7.5\mathring{A}$ [127].

Double-stranded DNA can be well described by a simple freely-jointed chain (FJC) model, in which the polymer is made up of stiff segments of length $2P$ (known as the Kuhn length) which are oriented randomly relative to each other. An excellent introduction to the freely-jointed chain model can be found here, courtesy of Prof. Cory Simon, Oregon State University

Contour length The length of a polymer at its maximum possible extension. The contour length of B-form dsDNA is 0.34nm/bp [127]

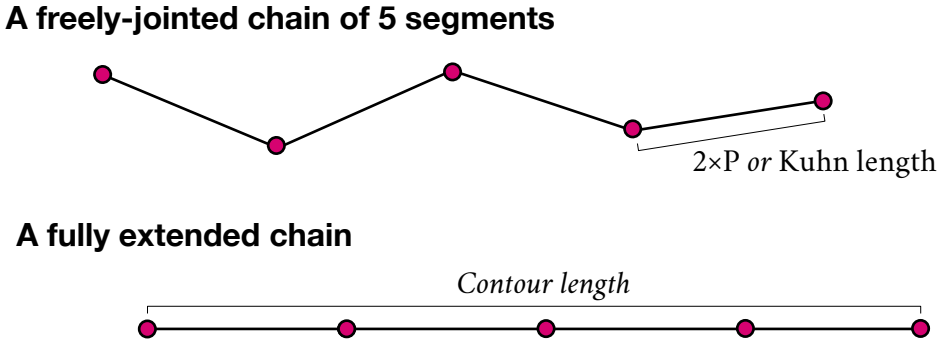

Table 1: Raw data for fig. 4.3

DNA	MobA	HRP	RLU
+	+	+	1046.6593 873.5887 962.4217
+	+	-	11.6934 9.3336 13.3691
+	-	+	42.7776 38.8705 33.1304
-	+	+	21.2786 115.0723 86.8769

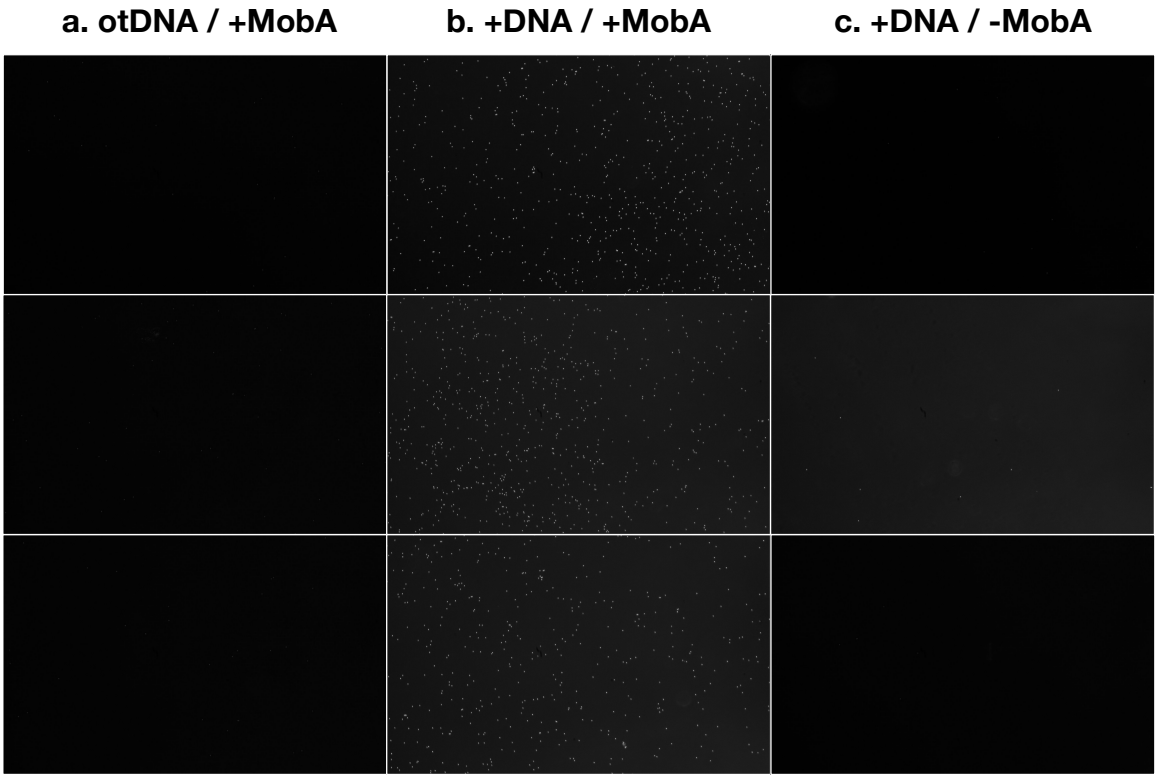

Figure 2: **Immobilization of overhang PCR DNA** a. MobA protein + PCV2-target DNA b. MobA protein + MobA-target DNA c. no protein + MobA-target DNA

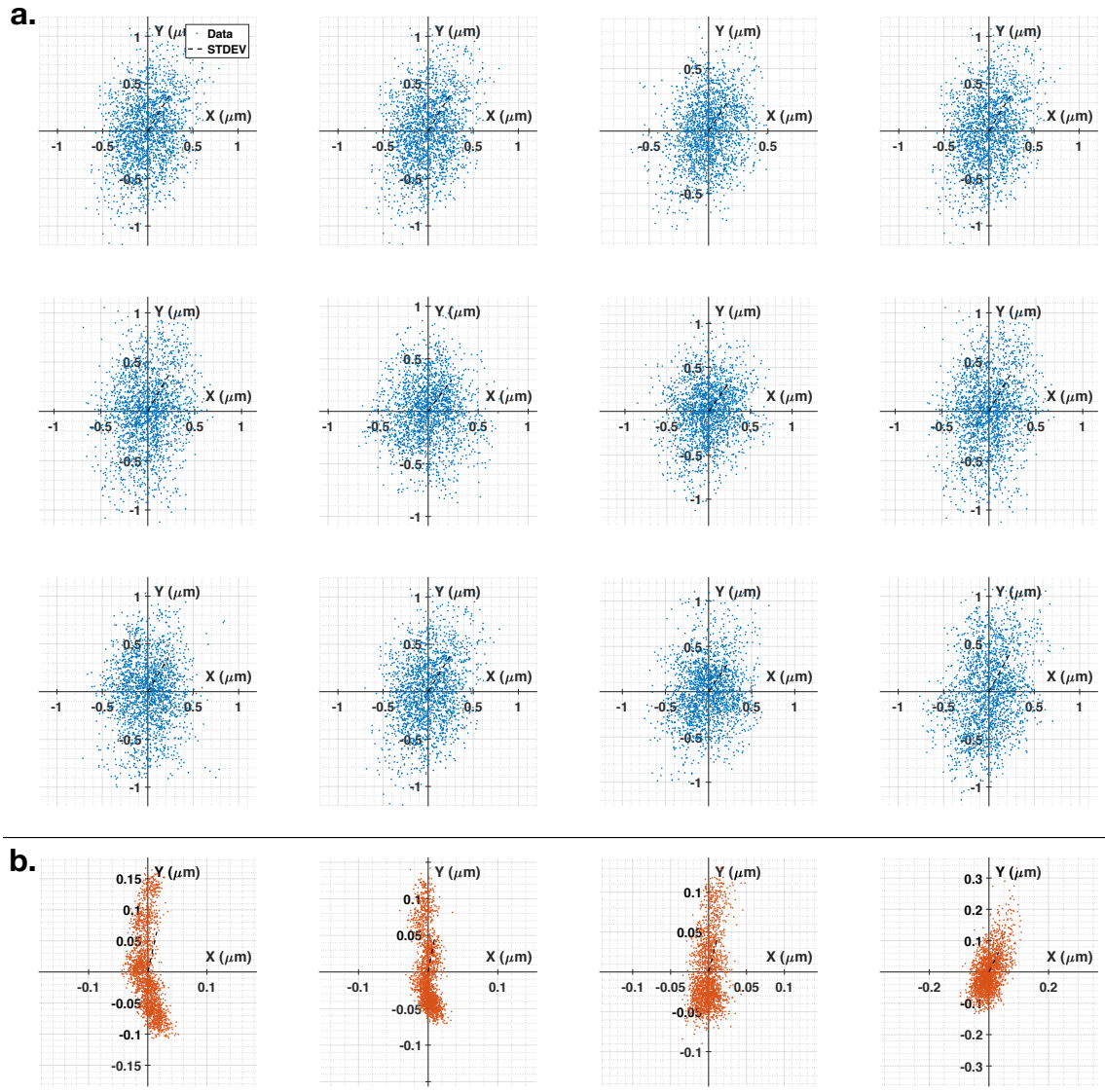

Figure 3: **Raw traces of mMobA-tethered DNA** a. Subpixel traces of tethered beads; dashed line indicates σ. b. Traces of stuck beads; note axis scale.