THE UNIVERSITY OF CHICAGO

DEVELOPMENT OF RNA POLYMERASE BIOSENSORS AND THEIR APPLICATIONS IN SYNTHETIC BIOLOGY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY
JULIA ROSE ZINKUS-BOLTZ

CHICAGO, ILLINOIS
MARCH 2020

I dedicate this thesis to my family and friends, whose love and support helped make this work possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to take this opportunity to show appreciation for all of the individuals who have shaped my experience and have supported me during my doctoral studies.

First and foremost, I would like to thank my advisor Professor Bryan Dickinson for being a key part of my research experience. I will always be appreciative of the multiple lines of communication he established to discuss research progress, career progress, and general lab management. I was especially impressed when he opened up an avenue for constructive feedback, and even more impressed when he acknowledged our concerns, directly addressing them, and describing how we as a lab would move forward. The ability to hear constructive criticism, acknowledge areas of improvement, and take steps to improve, will always be something I admire about him.

I also want to thank current and past Dickinson lab members - Jinyue, Rahul, Michael, Jane, Krysten, Somayeh, Clémence, Nate, Simone, Kaitlin, Jeff, Tian, Saara-Anne, Victoria, Yang, Tong, Fernando, Ian, Pablo, Evan, Dan, Alice, Emily, Liyuan, PJ, Tony, Woojin, Michelle, Noah, and Vivian - who have truly made my experience in the Dickinson lab a time I will always treasure. Every single member has been so supportive, providing technical support, advice, and expertise in all things relating to experimental design and troubleshooting, as well as, advice and encouragement for all areas of research, career development, and life. One gesture that will always be dear to my heart is when the Dickinson lab ”dissertation fairies” banded together, providing me with words of encouragement, a cheery ”lab” plant, and chocolate covered coffee beans as I assembled my thesis document. I have been truly blessed to work alongside them.

I would also like to take the opportunity to thank the individuals who have played a vital role in the success and completion of my research projects. First, I would like to thank Jinyue for being my mentor when I first joined the lab. She took the time to show me the necessary skills and techniques I would need in order to be a successful researcher, going above and beyond providing advice and encouragement when I encountered chal-

lenges both in my research and interpersonal relationships. She also played a a vital role in the development of the split T7 RNA polymerase biosensor; she ran all of the PACE evolutions, as I watched, assisted, and learned how to do so in the process, which has been instrumental in all of my subsequent projects. I would also like to thank Craig, who provided much needed advice and technical support in the design and execution of my high throughput sequencing experiment to interrogate the KRAS/RAF protein-protein interaction interface. He also was instrumental in analyzing the data; he wrote unique Python scripts to both extract and transfer data into Excel, so I could analyze the outcome of my selection. I also want to thank Liyuan for assisting in the generation of large scale antibody mimetic phage libraries, which I have subsequently used in preliminary work to evolve a *de novo* protein-protein interaction partner.

I also want to thank my friends, who have been so encouraging and supportive. My childhood friends, Amanda, Mackenzi, Kellie, and Agnieszka, who have always been there for me during every stage of my life. My friends from college, Jenny and David, who pursued their own PhDs and who have provided love, encouragement, and advice during every stage of my PhD. And the friends I have made here, Ryan, Rebecca, and Polina, who have been nothing but encouraging and supportive as we all navigate PhD life together.

I would also like to thank my family for all of the love and support they have given me throughout my academic career. My Grandma Jackie was a true inspiration - her dream was to support her grandchildren to obtain a college education, and I believe she would be truly proud of me. My Grandpa John, her husband, has also been so encouraging throughout the entire course of my PhD, asking about my progress and taking a genuine interest in the work I am doing. Also, my Nonna for being so encouraging and supportive, leading me to believe that I could achieve anything I set my mind to. I would also like to thank my parents, David and Palmira, for being encouraging and supportive in every aspect of my life - being instrumental in helping plan my wedding and providing much needed help with home ownership and home improvement projects. And most importantly,

I would like to thank my husband Matthew who has been there every step of the way. He has been there providing food, love, care and support in my successes as well as my failures, and I couldn't ask for a better partner to spend the rest of my life with.

I would also like to thank the different funding sources that have helped make my research possible, including the University of Chicago Pre-doctoral Training Program in Chemistry & Biology (funded through the National Institute of Health, T32 GM008720) and the Otto and Valerie Windt Memorial Graduate Fellowship. I would also like to acknowledge the Windt Graduate Student Travel Award for providing me with financial support to present my research at the Spring 2019 American Chemical Society National Meeting & Expo.

Finally, I would like to thank all of the individuals who have proofread my thesis: Professor Bryan Dickinson, Professor Joseph Piccirilli, Professor Bozhi Tian, Dr. David Dolivo, and Dr. Somayeh Ahmadiantehrani. These individuals are truly the unsung heroes of this thesis, providing valuable guidance and insight into the overall structure and content.

ABSTRACT

Protein-protein interactions (PPIs) are crucial for many diverse cellular processes, and dysregulated PPIs are often implicated in disease states. Therefore, monitoring PPIs is critical to understanding underlying biological processes and disease. Many different techniques and tools have been developed to monitor PPIs, including biosensors. Biosensors, which are composed of biological components, contain detection and response elements to facilitate transducing a biochemical signal or event into a detectable output. Certain biosensors with the ability to generate genetic outputs or protein outputs have been utilized for monitoring PPIs and synthetic biology applications including, the generation of synthetic genetic circuits and biosynthetic pathways, diagnostic tools and therapeutics, and sensors for industrial applications. However, the extensive, system-specific engineering and optimization required for many of these biosensors precludes their use in other biosensor designs and applications.

Because of the utility of biosensors to sense, monitor, and impact cellular biology, there is a growing demand in many diverse fields for biosensors that are highly characterized and broadly applicable. My thesis work aimed to generate such a biosensor for PPIs; a sensor that was robust, versatile, and capable of being implemented in orthogonal systems and contexts without the need for extensive re-optimization for each new PPI. By utilizing the T7 RNA polymerase (RNAP) as a scaffold, we developed a new protein fragment complementation assay (PCA)-based biosensor for the detection of PPIs. To optimize the properties of the split T7 RNAP biosensor scaffold, we utilized the directed evolution platform phage assisted continuous evolution (PACE) with a dual positive and negative selection scheme. The resultant split T7 RNAP PCA biosensor scaffold was characterized with multiple PPIs to show its versatility, including light-inducible and small molecule-inducible PPIs; and its orthogonality was demonstrated by testing in both *E. coli* and mammalian cells. The applicability of the split T7 RNAP biosensor scaffold for novel functions was demonstrated by generating a selection based scheme to interrogate the

PPI interface of the KRAS/RAF PPI. Without the need to optimize any component of the split T7 RNAP system, a selection of different RAF variants against KRAS was conducted in order to identify key residues in the interaction interface. Preliminary work with utilizing the split T7 RNAP biosensor scaffold for directed evolution applications was also explored by using stable protein scaffolds, termed antibody mimetics, as starting points to evolve a new PPI partner for the inflammatory bowel disease (IBD) biomarker calprotectin. From the original characterization and these additional applications, we were able to demonstrate how the split T7 RNAP biosensor is a useful tool that can be applied to diverse applications, and should find broad utility in synthetic biology applications.

LIST OF PUBLICATIONS†

1. **Zinkus-Boltz, J.**; DeValk, C.; Dickinson, B. C. A phage-assisted continuous selection approach for deep mutational scanning of protein-protein interactions. *ACS Chemical Biology* **2019**, *14*, 2757-2767, DOI: 10.1021/acschembio.9b00669.

2. Jones, K.; **Zinkus-Boltz, J.**; Dickinson, B. C. Recent advances in developing and applying biosensors for synthetic biology. *Nano Futures* **2019**, *3*, 042002, DOI: 10.1088/2399-1984/ab4b78.

3. Pu, J.*; **Zinkus-Boltz, J.***; Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nature Chemical Biology* **2017**, *13*, 432–438, DOI: 10.1038/nchembio.2299.

†. The following chapters of this dissertation contain sections and figures adopted from the listed publications with modifications. Chapter 1: publication 2; Chapter 2: publication 3; Chapter 3: publication 1; Chapter 4: publication 2.

∗. Denotes equal contribution.

CHAPTER 1

DEVELOPMENT AND APPLICATIONS OF BIOSENSORS FOR SYNTHETIC BIOLOGY

1.1 Introduction

Protein-protein interactions (PPIs) are important for diverse cellular functions, including metabolism, signaling, and cellular organization.[1] As a consequence, dysregulated PPIs are often key drivers of disease.[2–5] Due to the critical nature of PPIs for biological systems, many methods have been developed to study PPIs. These methods include genetic approaches (e.g. yeast 2-hybrids i.e. biosensors), biochemistry approaches (e.g. co-immunopreciptation and cross-linking), mass spectrometry-based approaches (e.g. proteomics), and imaging-based approaches (e.g. FRET and confocal microscopy).[6,7] These methods have increased the understanding of PPIs, providing a deeper insight into how cells function.

Of the aforementioned techniques, very few have the ability to both monitor and impact cellular functions. Biosensors, which are comprised of both sensor and reporter domain elements with the ability to transduce input signals and events into tractable outputs, are one of the few tools to study PPIs that have this potential. Crucially, the generation and characterization of biosensors to measure, analyze, and control cellular functions has enabled the field of synthetic biology to design complex genetic circuits[8] and biosynthetic pathways,[9–12] diagnostic tools and therapeutics,[13,14] and sensors for industrial applications.[15] Though exciting, the extensive, systems-specific engineering and optimization required for many of these functionalities and applications has thus far meant that many biosensors remain useful only in their systems of original design and optimization.[16,17]

Due to the unique nature of biosensors to sense, monitor, and impact cellular functions based on input signals, there is a growing demand in many diverse fields for biosensors. Biosensors with the capacity to sense PPIs are routinely made and utilized

both to understand underlying cellular biology and generate cellular changes based on measured PPIs. This chapter will briefly introduce biosensor design and characterization, followed by discussing different biosensors used to detect PPIs, including resonance energy transfer (RET) biosensors, protein fragment complemntation assays (PCAs), engineered cell surface receptors, and n-hybrid biosensors. I will then discuss some of the current challenges and opportunities in the field of biosensor design for PPIs, specifically the need for biosensors to interface with cellular processes for synthetic biology applications, and summarize the layout of my thesis.

1.2 Biosensor design and characterization

Biosensors are comprised of both sensor and reporter domain elements with the ability to recognize an input signal and generate an output for either detection or driving a cellular change. Various underlying biomolecular scaffolds, usually first identified and characterized in natural systems, have been utilized (e.g., RNA, transcription factors, proteins, and enzymes), forming a diverse set of sensors capable of detecting a broad array of inputs, from small molecules to proteins, and even light, to yield an equally varied set of readout/output signals, such as fluorescence, luminescence, gene expression, and enzymatic activity.[18–21] Rational, informed design of novel, synthetic biosensors is not trivial and requires significant effort in first characterizing and then optimizing properties that affect overall performance, such as the output dynamic range, input specificity and sensitivity, and reliable, orthogonal utility in different conditions and genetic contexts (Figure 1.1). This last point is especially important when considering porting biosensors between original and target systems.[22,23] To highlight some of the considerations associated with the design and characterization of a new biosensor, two classic, highly-reviewed biosensor systems will be briefly introduced: transcription factor (TF)-based biosensors and riboswitches (Figure 1.2).

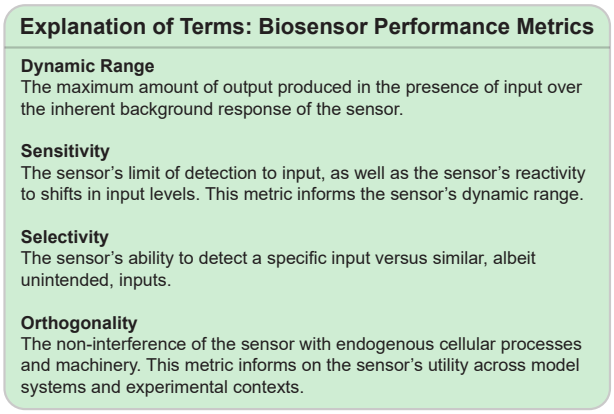

Figure 1.1: Biosensor Performance Parameters.

Although available since the 1950s,[24] biosensor technologies have only relatively recently made the leap from electronic to biological parts. Small molecule-inducible microbial promoters quickly emerged as relatively straight-forward biomolecular options that were readily harnessed for biosensing applications. TF-based biosensors utilize naturally evolved TFs to either activate or repress reporter gene expression.[18] Identified via genome mining, TF/promoter pairs or repressor/operator pairs typically respond to small molecules or metabolites (e.g., arabinose, IPTG, anhydrotetracycline, acrylate, glucarate, erythromycin, naringenin, etc.)[22,25] to modulate transcriptional activity (Figure 1.2a). Newer, synthetic TF biosensor designs involve engineering and linking together ligand sensing and DNA binding domains to alternately recruit transcriptional machinery or block transcription based on the presence and absence of input signal, respectively.[26–28] More complex systems involve integrating transcription factors into biosynthetic pathways to help regulate and increase the production of metabolites.[29] This reliance on the system's transcriptional machinery, however, denotes a major limiting factor in the design of TF-based biosensors. Transcriptomic methods can and have been used to help identify ligand-responsive transcriptional elements,[30–32] but their finite system-wide diversity can present challenges for specificity.[18] Transference of transcriptional elements between heterologous systems also tends to be unreliable, due to unforeseen losses in efficiency in non-native systems.[18,33] Despite this, TF-based biosensors are widely used, and have

found utility in detecting metabolite production in biosynthetic pathways and organisms via high-throughput screening (HTS) techniques, such as fluorescence assisted cell sorting (FACS).[34,35] For a more extensive look at TF-based biosensors, see the following recent reviews: 18,34.

Composed entirely of RNA, genetically-encoded riboswitches, or RNA switches, are typically divided into aptamer and expression platform segments and allosterically inform reporter gene expression (Figure 1.2b). The unique biophysical properties of RNA enable binding of the aptamer portion to its cognate ligand, typically a small molecule, which induces a conformational change in the expression platform to modulate a genetic response, (e.g., translational initiation, translational termination, or transcript decay via ribozymatic self-cleavage).[36] Although naturally-occurring, any number of predictive secondary structure or biophysical property algorithms can be used to rationally design synthetic riboswitches *de novo*.[37–41] In addition, evolution of aptamers from selection-based platforms, such as systematic evolution of ligands by exponential enrichment (SELEX),[42–44] and annotated repositories of known RNA aptamer, riboswitch, and ribozyme sequence libraries (for example: https://www.aptagen.com/aptamer-index) have facilitated generation of novel riboswitches.[45,46] Their deceptively simple design (i.e., aptamer and expression platform sequences tethered with a flexible linker sequence) belies the extensive effort required for designing and synthesizing a riboswitch with the correct desired functionality. For example, while directed evolution strategies, such as phage-assisted continuous evolution (PACE), can be harnessed to tune the dynamic range of an existing riboswitch, they have yet to be used to successfully alter ligand specificity.[47] A recent comprehensive rational design strategy further elucidates the different considerations and optimization steps needed to generate a functional riboswitch.[48] Nevertheless, HTS and FACS approaches[49–55] have contributed to an increasing speed of synthetic RNA switch design and development. For a more extensive review of riboswitches, see: 20,34,56.

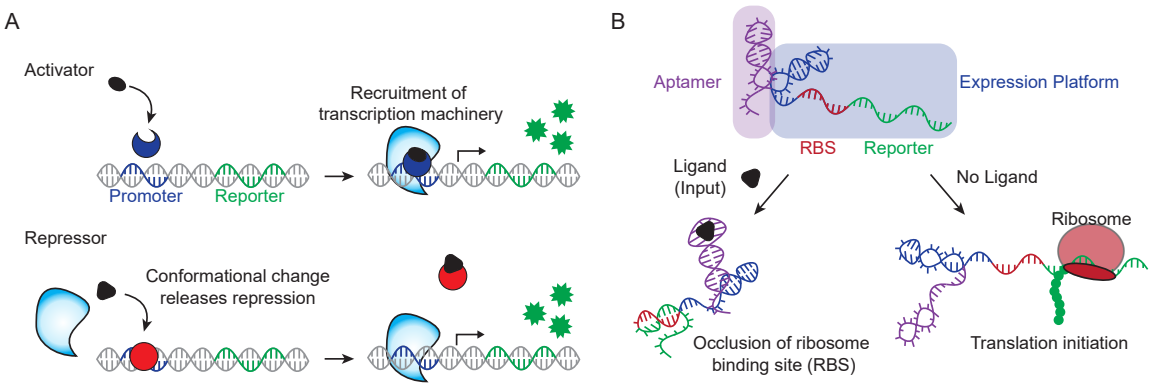

Figure 1.2: Classic biosensor platforms produce gene reporter outputs in response to ligand inputs. A) For basic transcription factor-based biosensors, ligand-mediated changes in protein conformation either facilitate (top) or block (bottom) transcription factor binding to gene promoters. In both cases, the cell's transcriptional machinery is recruited to express a reporter gene. B) In this example of a riboswitch, ligand-induced structural alterations obstruct ribosome binding and reporter gene translation (left), while the absence of ligand permits reporter gene translation (right).

1.3 Resonance energy transfer (RET) biosensors

Resonance energy transfer (RET) biosensors utilize either fluorescence-RET (FRET) or bioluminescence-RET (BRET) to detect PPIs *in vivo*. In resonance energy transfer, a donor molecule is excited and, in a non-radiative energy transfer process, excites an acceptor. In FRET biosensors, fluorescent proteins or fluorophores are used as the donor and acceptor molecules, where the excitation of the donor is generated by light; in BRET, a luciferase and fluorescent protein (or fluorophore) are used as a donor and receptor, respectively, and excitation of the donor is generated through cleavage of a small molecule to produce bioluminescence. Because these biosensors rely on RET, the distance between the donor and acceptor molecules can be quantified based upon the ratio of emission of the donor and/or acceptor. Because of this, care needs to be taken when designing these biosensors to ensure both orientation and linkers are optimized to facilitate generation and observation of a robust FRET/BRET signal.[57–59]

FRET and BRET biosensor designs rely on protein fusions of the acceptor and donor to two PPI partners, and have been used to study many different PPIs including GPCR

meditated PPIs.[57–59] Due to the utility of both FRET and BRET to detect PPIs *in vivo*, improvements in the design of both donor and acceptor molecules are routinely done.[60–62] For example, a BRET platform based upon NanoLuc luciferase was developed by the Wood lab to improve upon previous BRET platforms that utilized Renilla luciferase. In this study they fused the donor, NanoLuc, and the acceptor, a Halo-Tag conjugated to a chloroalkane derivative of nonchloro TOM (NCT) dye, to PPI partners to generate a robust NanoBRET sensor with minimized spectral overlap between the donor and acceptor.[60] With this improved NanoBRET biosensor they were able to monitor the β-arrestin 2 (ARRB2)/vasopressin receptor 2 (AVPR2) PPI *in vivo*. RET based biosensors can also be incorporated into more complex biosensor designs to improve biosensor properties through the generation of AND gates. In a recent report by the Ting laboratory, they utilized NanoLuc BRET to improve their previously reported 'Specific Protein Association tool giving transcriptional Readout with rapid Kinetics' (SPARK) biosensor for the detection of PPIs. By incorporating NanoLuc as the donor for BRET in their biosensor design, they eliminated the need for external light stimulation for their sensor, and were able to reduce background activation in the absence of a PPI in high expression conditions.[63]

These two examples using the improved luciferase NanoLuc showcase how RET biosensors can be utilized to detect PPIs *in vivo*. However, due to the nature of the light output generated by RET biosensors, they cannot directly be utilized to integrate with host cellular processes and drive cellular changes. More complex biosensor designs that incorporate RET with other elements like protease cleavage to detect PPIs, as demonstrated by the SPARK biosensor designed by the Ting lab,[63] have this potential, but are limited in applicability due to the extensive design and engineering required in order to tune biosensor properties.

1.4 Protein fragment complementation assays (PCAs)

Split reporters are routinely used to monitor biological interactions and assembly processes. For example, split RNA aptamer systems have been utilized to observe cleavage of RNA ribozymes *in vitro*,[64] RNA transcription and hybridization *in vitro*,[65] and RNA-RNA interactions *in vivo*.[66] Analogous protein fragment complementation assays (PCAs) are routinely used to measure PPIs through the use of split reporters,[67] reviewed in: 68–70. Complementary split reporters are fused to interacting proteins that direct assembly of the reporter to generate an output signal (Table 1.1).

PCA Scaffold	Output
Luciferases	Bioluminescence
Fluorescent proteins	Fluorescence
Peroxidases (HRP and APEX)	Reactive radical (fluorophore, proximity/EM probe)
Biotin ligase	Proximal protein labeling
Proteases	Proteolysis (cell signaling)
RNA aptamers	Fluorescence

Table 1.1: A summary of split reporter scaffolds. HRP, horseradish peroxidase; APEX, ascorbate peroxidase; EM, electron microscopy.

For example, split fluorescent proteins and luciferases are routinely used to detect PPIs *in vitro* and *in vivo* through the generation of fluorescence and photons, respectively.[71,72] Due to the utility of these split reporters as imagining agents, an array of different split fluorescent and split luciferase/luciferin pairs have been developed for the ability to analyze multiple PPIs in a single system.[71–74] A recent study by the Paganetti lab utilized both bipartite- and tripartite-split GFP to analyze the multimerization of the RNA-binding protein TDP-43, which is a protein associated with frontotemporal dementia and amyotrophic lateral sclerosis. They were able to visualize multimerization using live cell imaging, and with follow-up immunoprecipitation assays utilizing GFP complementation, identified the N-terminal region and specific residues of TDP-43 critical for multimerization.[75] Importantly, these tools enable numerous applications beyond PPI analysis, including HTS of PPI inhibitors, detection of protein biomarkers, monitoring changes in pro-

tein conformation, and assessment of protein solubility.[71,76–78] Due to their broad utility, developmental engineering of new split luciferase and fluorescent protein systems are routinely conducted.[79–82]

Other PCA biosensors based upon split enzymes that rely on output signals other than light have been developed to detect PPIs.[68] More recent split enzyme designs utilize proximity labeling to facilitate the detection of protein interactions. Both peroxidases and biotin ligases have been engineered as PCA biosensors facilitating proximity labeling of nearby proteins and RNA, allowing for visualization of transient PPIs, protein complexes, or RNA-protein interactions.[83–86] For example, split-BioID, which was engineered from a BirA biotin ligase, has been utilized for the visualization of transient PPIs.[83,84] The Bollen group utilized the engineered split-BioID to interrogate the protein interactome for protein phosphatase PP1 and the putative PPI partners NIPP1 and RepoMan. In their analysis, they identified a list of transient PPI interaction partners for PP1, and they were able to show that split-BioID had lower background biotinylation compared with full length BioID.[83] Due to the applicability of proximity labeling to detect protein complexes and transient PPIs, these split enzyme reporters will find broad utility in the detection and mapping of the protein interactome.

Other PCA scaffolds based upon split proteases, such as tobacco etch virus (TEV), are also regularly utilized to monitor PPIs. For example, reassembly of split TEV can either proteolytically cleave and release TFs for reporter gene expression, or proteolytically activate an inactive reporter protein.[87,88] For example, in a study conducted by the Rossner lab which utilized proteolytic cleavage of a transcirption factor to monitor PPIs, interactions between different PPIs at both the membrane and in the cytosol were examined. Notably, they were able to detect the preferential heterodimerization required for $GABA_B$ G protein–coupled receptor (GPCR) activation and neuregulin-1 (Nrg-1)-mediated ErbB receptor tyrosine kinase (RTK) dimerization.[87]

Split proteases have also become important components of engineered gene circuits

in synthetic biology due to the ability of split proteases to model cell signaling through the release of a transcription factor or the degradation of a reporter protein.[89,90] For example, Jerala and colleagues developed a panel of split proteases, fused with orthogonally dimerizing coiled coils (CC), that recognize specific cleavage sites to model Boolean logic functions and signaling pathways in mammalian cells.[90] A similar approach from the Elowitz group utilizes degrons fused to output reporters via a protease cleavage site, creating protease-repressible and -activatable platforms in which the degron is exposed or removed resulting in degradation or stabilization of the reporter.[91] To regulate the protease that controls the degron/reporter platform, they split the protease such that enzyme complementation was dependent upon an additional, separate protease. This protease-regulated protease was interfaced with Boolean logic gates for multiple inputs and specific outputs.

Of the aforementioned PCA biosensor scaffolds, many suffer from only being applicable for the detection of PPIs due to the nature of their output being unable to integrate with host cellular processes. Split protease PCA biosensors are unique in that they have the capacity to directly integrate with host cellular processes through either the release of a transcirption factor to drive gene expression or degredation of a reporter protein modeling native cell signaling networks, which has facilitated their use in a diverse array of synthetic biology applications.

1.5 Engineered cell surface receptors

Extracellular ligand-mediated activation of cell-surface receptors transmits environmental information to intracellular signaling networks, triggering various downstream cellular processes, including gene expression, cell growth, and secretion of small molecules/cytokines. Developing orthogonal cell-surface receptors is therefore a tractable approach to either report on or modify intracellular signaling. G-protein coupled receptors

(GPCRs), in particular, possess broad utility in this aspect. Synthetic GPCRs, activated only in the presence of an otherwise pharmacologically inert small molecule, have been engineered through directed evolution,[92–94] chimeric design,[95] and site-directed mutagenesis.[92,96] Designs that detect protein ligands or PPIs have also been developed due to the utility of cell surface receptors to directly integrate into downstream cascade activation and/or gene expression outputs for a diverse array of applications including therapeutics,[97–101] histological patterning,[102,103] and synthetic biology.

For example, the Cornish lab identified and repurposed fungal mating peptide/GPCR pairs into orthogonal signaling interfaces in yeast to assemble and pattern communication topologies, building a synthetic, interdependent cell community.[104] Though modeled in yeast, this scalable, modular system can likely be transported to mammalian cells to enable peptide signaling-mediated control of cellular behavior. By incorporating a proteolytic cleavage-dependent component, Lee and colleagues developed the Tango assay to detect both receptor activation and PPIs. This assay, inspired by the mechanism of Notch receptor signaling (Figure 1.3a-b), relies on the expression of both a GPCR fused to a transcription factor, and an arrestin-TEV protease fusion protein.[105] Ligand-mediated receptor activation leads to the TEV-mediated proteolytic cleavage of the GPCR-anchored transcription factor and expression of reporter genes.[105] Tango assays have been used to measure the activity of 300+ GPCRs.[106] Moreover, this platform design has been extended to additional receptor classes, including receptor tyrosine kinases and steroid hormone receptors.[105] Another Notch pathway-inspired approach, synNotch, retains the small self-cleaving proteolytic region of Notch but replaces the extracellular and intracellular receptor domains to orthogonally detect a broad array of cell-surface proteins to produce customized responses (Figure 1.3c).[102] The synNotch platform has enabled cell engineering efforts with applications in cancer immunotherapy[107,108] and self-organizing tissue engineering.[103]

Other groups have looked to build a more generalizable approach to extracellular

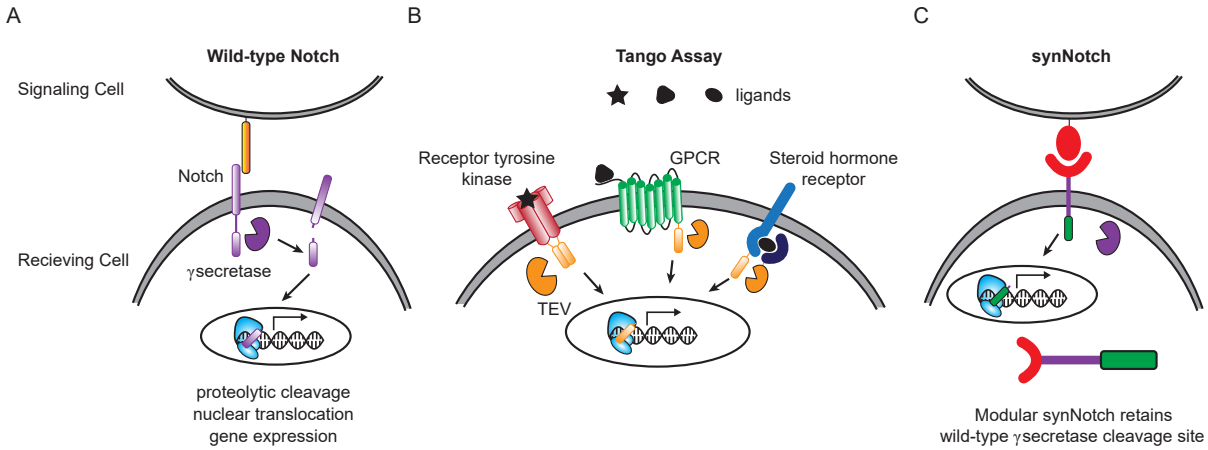

Figure 1.3: Comparison of Notch-based biosensor platforms. A) Upon binding with cognate cell-surface signaling proteins, the intracellular domain of wild-type Notch receptors is proteolytically cleaved by γ-secretase, producing a nuclear-translocating transcription factor that promotes gene expression. B) Fusion of a transcription factor to receptor tyrosine kinases, G protein-coupled receptors (GPCRs), or steroid hormone receptors via a linker containing a tobacco etch virus (TEV) protease cut site forms the basis of the Tango Assay. Ligand-mediated receptor activation recruits the TEV protease, which cleaves the transcription factor and subsequently drives reporter gene expression. C) The highly modular synNotch biosensor platform utilizes the wild-type Notch γ-secretase cleavage site but allows for customization of both the receptor and transcription factor elements. As with wild-type Notch, activation of the synNotch receptor by its cognate signaling protein recruits γ-secretase to proteolytically free the transcription factor and activate gene expression.

biosensing by developing customizable, synthetic cell surface receptors. The Modular Extracellular Sensor Architecture (MESA) platform was developed by the Leonard lab, and similar to Tango, MESA functions via extracellular ligand-mediated receptor dimerization and subsequent proteolytic cleavage and release of a transcription factor.[89] The extensive, swappable modularity of MESA allows both sensor optimization and customizable recognition of new ligands. This plug-and-play flexibility was expanded to sense a physiologically relevant input to stimulate a cell-based therapeutic response. MESA receptors engineered with single-chain variable fragments from a vascular endothelial growth factor (VEGF) antibody,[109] dimerize in response to VEGF, release dead Cas9 (dCas9)-TF, a catalytically inactive Cas9 fused to a transcriptional activation domain, resulting in secretion of interleukin-2. Multiplexing of MESA receptors expands the potential to

direct diverse cellular functions.[110] In the same vein, Chen and colleagues have recently elaborated upon existing chimeric antigen receptor (CAR)-expressing T-cell technologies to recognize soluble ligands[111] (previous CAR platforms recognize only surface-bound ligands).[112–119] This development is significant, as it allows targeting of secreted cytokines, shed tumor antigens, and other soluble factors associated with pathologic environments by CAR-expressing T-cell-based therapies.

As with the previously discussed split protease PCA biosensors, engineered cell surface receptors have a direct inroad to interfacing with host cellular processes through protein signaling networks. This has allowed for the modulation of cellular responses for both the detection of ligand-receptor interactions and the engineering of complex synthetic biology systems for diverse applications, including the design CAR-expressing T-cell therapies.[111–119] While the detection of new ligand targets requires substantial design efforts and optimization, advances in modular designs such as the MESA platform developed by the Leonard lab,[89,109,110] will help ameliorate design efforts and expand the utility of engineered cell surface receptors for both the detection of PPIs and utility in synthetic biology applications.

1.6 n-hybrid biosensors

n-hybrid-based biosensors have been designed for the detection of DNA-protein (1-hybrid), protein-protein (2-hybrid), protein-small molecule (3-hybrid), and protein-RNA (3-hybrid) interactions. These biosensors link the detection of the DNA-protein, protein-protein, protein-small molecule, or protein-RNA interaction to the expression of a genetically encoded reporter to produce an output signal.[120–125] In the originally designed yeast 2-hybrid assay, “bait” and “prey” proteins are fused to the DNA binding domain of Gal4 and the transcriptional activator domain of Gal4, respectively, and upon association of the fused bait and prey proteins, transcription of the genetically encoded reporter oc-

curs.[121] Yeast 2-hybrid approaches have been utilized for the detection of binary PPIs on a genome scale for multiple organisms,[126] including humans (*Homo sapiens*),[127] bacteria (*Escherichia coli*),[128] yeast (*Saccharomyces cerevisiae*),[129,130] nematodes (*Caenorhabditis elegans*),[131] and fruit flies (*Drosophila melanogaster*).[132] Due to the utility of yeast 2-hybrid to link PPIs to genetic outputs, new 2-hybrid designs have emerged. Other DNA binding proteins such as transcription activator like effectors (TALEs) and nuclease-deficient Cas9 proteins (dCas9) have been re-purposed to facilitate the design of 2-hybrid biosensors.[133–136]

These newer designs have broad utility in synthetic biology because these 2-hybrid systems generate genetic outputs that can directly impact biological systems. TALE-based 2-hybrid biosensors utilize the DNA targeting ability of TALEs and a transcirptional activator each fused to protein partners to detect PPIs. For example, Feng and colleagues generated a biosensor using the blue light-inducible protein dimerization domains CRY2 and CIB1 fused to a TALE protein or to a transcriptional activator VP64, respectively, to generate a light-inducible biosensor. In primary mouse neurons, they were able to show reversible target gene expression of endogenous Grm2 in the presence of light, and target histone modifications by swapping out the transcriptional activator VP64 for either histone methyltransferase and deacetylase enzymes.[134]

dCas9-based 2-hybrid biosensors utilize the DNA recognition capabilities of the nuclease-deficient Cas9 enzyme and subsequent single guide RNA (sgRNA)-mediated targeting of the dCas9 fusion to a gene of interest to modulate transcription to generate several gene activation (CRISPRa) and repression (CRISPRi) biosensor platforms,[135,137,138] reviewed in 136. For example, the light-activated CRISPR-Cas9 effector (LACE) system developed by Polstein and colleagues fused the light-inducible CRY2 and CIB1 domains from the plant *A. thaliana* to the transcriptional activator VP64 and dCas9, respectively, for CRISPRa functionality following light stimulation at levels comparable to a direct, dCas9-VP64 fusion.[135] Importantly, the light-inducible gene expression was re-

versible, spatially constrained, and non-cumulative over multiple rounds of light stimulation and removal, although a slight increase in background expression over time was noted.[135] A similar approach was also utilized by the Sato laboratory, with a CRY2-fused p65 transcriptional activator associating with the CIB1-fused dCas9 under blue light stimulation to drive gene expression.[138]

Further utility of Cas9-based 2-hybrid biosensors for PPIs was explored by conferring a dimension of orthogonality to Cas9 biosensor platforms. Gao et al coupled dCas9 enzymes from different bacterial species (*S. pyogenes* and *S. aureus*) with different small molecule-inducible PPI pairs fused with either a transcriptional activator (VPR) or a repressor (KRAB).[137] Although the group observed disparities in the level of gene expression modulation between the two species of dCas9, possibly due to either an unoptimized sgRNA or lower DNA binding affinity for the *S. aureus*-derived dCas9, CRISPRa and CRISPRi outputs were both reversible and minimally cross-reactive. This orthogonality enabled the generation of Boolean logic OR and AND gates, as well NOR and NAND gates.[137] This particular synthetic circuit design expanded the utility of dCas9-based synthetic circuitry beyond the previous dCas9-mediated control of gene activation.[139–142]

These improvements in the design of 2-hybrid based approaches using other DNA binding domains (TALEs and dCas9) have expanded the utility of n-hybird biosensors by facilitating their deployment in other host systems, and control of native gene expression, as seen with sgRNA-targeted dCas9-based 2-hybrid biosensors. These features will greatly expand the utility of these biosensors in a diverse array of applications, including the design of novel synthetic biology systmes.

1.7 Challenges and opportunities in the development of biosensors for PPIs

The utility of biosensors to sense, report on, and impact biology in response to endogenous and exogenous signals has prompted the generation of a wide array of novel biosensors. However, many biosensor designs cannot be used for synthetic biology applications due to either the nature of their output or due to the limited versatility of biosensor designs for new PPIs and applications. When examining biosensors for PPIs, it is important to note that they fall into two distinct modalities: those that can only sense PPIs, and those that can both sense and impact biology.

Biosensors with non-genetic or non-protein outputs can only monitor PPIs and cannot be utilized in synthetic biology applications. RET biosensors and PCA biosensors based on split fluorescent proteins, split luciferases, split peroxidases, and split ligases, produce either light or small molecule cleavage events for proximity labeling, so they cannot directly interface with biological systems. Due to this, other PCA designs have been explored. Split proteases, such as split TEV, have overcome this limitation by facilitating the design of artificial protein signaling networks based on cleavage of cognate protease cut sites to release either transcription factors to produce genetic outputs, or to activate inactive proteins to generate artificial signaling networks. These particular systems while quite robust, still require optimization to limit background activation in the absence of PPIs or to prevent cross-talk between similar PPI partners, as seen with the synthetic protein signaling network generated with CC domains by Jerala and colleagues.[90]

Biosensors with the ability to monitor PPIs and directly impact biological systems generate either a genetic output or protein output that can impact biology. Engineered cell surface receptors and 2-hybrid biosensors generate either genetic outputs or protein outputs, and have been applied to many different synthetic biology applications, including the generation of complex genetic circuits[8] and biosynthetic pathways in organisms,[9–12]

development of diagnostic tools and therapeutics.[13,14] However, the extensive, systems-specific engineering and optimization required for many of these PPI biosensors, limits their implementation to the system of their original design and optimization.[16,17] For example, traditional yeast 2-hybrid systems have known issues with both false negatives (~70-90%) and false positives,[126,143] requiring careful experimental design and extensive optimization of linker lengths, fusion orientations, and fusion geometries to generate a new 2-hybrid system.[126,144–147] However, recent innovations in modular design (i.e., 'plug-and-play' approaches to building biosensors), random and targeted mutagenesis, screening-based selection, directed evolution technologies,[42] and antibody mimetics[148] have each partially ameliorated design, development, and optimization phases for individual biosensors.

Improvements in the original design of biosensors are needed, as well as extensive characterization of new biosensor designs. By incorporating modularity in the design, biosensor scaffolds can be utilized for multiple different PPIs with limited need for re-optimzation of the original design. Extensively characterizing biosensors with multiple different PPIs and in multiple different genetic contexts will generate more information regarding biosensor performance, greatly facilitating the implementation of biosensors for novel synthetic biology functions. Due to this, there is a growing need for improved designs and extensive characterize of versatile PPI biosensor scaffolds, with large dynamic ranges, high specificity and selectivity, and orthogonality for diverse synthetic biology applications.

1.8 Scope of thesis

Due to the unique nature of biosensors to sense, monitor, and impact cellular functions based on input signals, there is a growing demand in many diverse fields for biosensors. Because of the broad applicability of biosensors that can integrate with biological systems, biosensors with genetic outputs or protein signaling outputs are routinely

generated; however, system specific-designs, often requiring re-optimization for new PPIs or functions, limit their utility in many diverse applications. Due to this, there is a growing demand for versatile PPI biosensors, with large dynamic ranges, high specificity and selectivity, and orthogonality for diverse synthetic biology applications.

The goal of my thesis work was to address this limitation and design a robust, versatile, and orthogonal biosensor capable of generating a genetic output for monitoring PPIs and for other applications in synthetic biology and chemical biology, such as directed evolution. In chapter 2, I discuss the design and characterization of a versatile 'plug and play' split T7 RNAP based biosensor for the detection of PPIs in both *E. coli* and mammalian cells, and for such diverse applications as detection and synthetic biology through the generation of RNA aptamer, mRNA, and RNAi outputs. In chapter 3, I discuss an application of the split T7 RNAP biosensor to generate a selection platform for the interrogation of PPI interfaces utilizing deep mutation scanning to identify key residues in the KRAS/RAF PPI, highlighting residues with both an intolerance and a high tolerance to mutations. In appendix A, I highlight some of the preliminary work in the utilization of the split T7 RNAP biosensor to evolve new PPI partners for the development a diagnostic for a biomarker in inflammatory bowel disease. In chapter 4, I will summarize the work I have done towards the generation of a versatile and orthogonal biosensor using the split T7 RNAP scaffold as well as future applications and challenges associated with this biosensor design.

CHAPTER 2

EVOLUTION OF A SPLIT RNA POLYMERASE AS A VERSATILE BIOSENSOR PLATFORM

2.1 Introduction

Diverse areas of chemical biology and biotechnology, including directed evolution, synthetic biology, and bioengineering, require methods to link chemical and biochemical processes to defined genetic outputs.[11,100,149,150] RNA is a particularly useful output owing to the availability of technologies to drive cellular responses on the basis of computed nucleic acid signals.[151–153] In nature, transcription factors detect target activities and drive genetic responses, and many naturally occurring transcription factors, such as the Lac and Tet repressors, have been repurposed as gene expression control elements.[11,133] However, because the scope of detection of such natural systems is limited, several engineered alternative technologies have been developed.

n-hybrid systems, such as those for detection of protein-DNA (one-hybrid), protein–protein (two-hybrid), protein-RNA (three-hybrid), and protein-small-molecule (three-hybrid) interactions, are among the approaches most commonly deployed for genetic sensor design.[121,123–125] The development of programmable DNA binding domains such as transcription activator-like effectors (TALEs) and nuclease-deficient Cas9 (dCas9) have revolutionized one- and two-hybrid approaches.[133–136] However, a challenge with n-hybrid systems is that they must be carefully tuned and optimized for each new interaction, and more complex multi-component systems such as three-hybrids often lack sensitivity and have low signal-to-noise ratios.[143,154] Although one-hybrids (and two-hybrids, in some cases) can be used for synthetic biology purposes, these methods are generally not suitable for applications that require a high level of control and dynamic range. A primary alternative to n-hybrid approaches is riboswitches, RNA-based elements that drive translational outputs on the basis of an aptamer's interaction with a target ligand.[155,156]

Riboswitches have become an important tool for designing systems that respond to target inputs, which are usually small molecules but can also be proteins.[157] Although riboswitches can work well with targets for which aptamers can be created, protein-based biochemical activities are generally beyond their scope of detection, and translating these tools in mammalian systems often results in diminished performance.[154] Outside of these general strategies, a host of other synthetic biology parts have been developed for highly specialized activities.[26,100] Therefore, there is a need for a general method to transduce endogenous chemical and biochemical information into DNA or RNA for subsequent storage or integration with engineered regulatory systems.

Engineered polymerases have been proposed as a strategy to respond to and measure endogenous biochemical processes.[158,159] Inspired by this concept, we developed protease-responsive RNAPs (PRs) to respond to protease activity by production of defined RNA outputs.[160] We showed that PRs can encode multidimensional protease activities in defined sequences of RNA in both prokaryotic and mammalian cells. In principle, RNAPs provide a new platform for biosensor creation, but engineering such complex enzymes is challenging, and the inhibitor-based design strategy used for PRs is fundamentally limited to protease activities.

We envisioned a new biosensor system based on recently reported 'split' T7 RNAPs,[161,162] in which N- and C-terminal components of the T7 RNAP spontaneously assemble to form a functional RNAP enzyme. We aimed to engineer activity-responsive RNAPs (ARs) in which assembly is not spontaneous but instead depends on fused interaction partners, similarly to other protein fragment complementation technologies (PFCs).[68] PFC involves tethering potential binding partners to complementary halves of a split protein marker, such as a fluorescent protein.[163] Generally, the protein binding partners must be expressed at much higher levels than endogenous conditions to achieve a detectable signal, and only two or fewer signals can be monitored at once.[73] More importantly, the fluorescence output in PFC cannot be integrated into downstream synthetic control sys-

tems. We reasoned that if a proximity-dependent split RNAP could be developed, genetic biosensors could be created through a 'plug-and-play' approach in which target interacting domains can be swapped in, resulting in a GFP-like platform for biosensor engineering (Figure 2.1a).

In this study we engineered a proximity-dependent split T7 RNAP sensor using continuous molecular evolution. We then demonstrated the versatility and ease of use of the platform by creating robust light- and small-molecule-responsive genetic sensors. To illustrate the power of polymerase-based biosensors, we showed that multidimensional protein–protein interactions (PPIs) can be monitored using ARs. Finally, we confirmed that ARs can trigger RNA nanostructure formation, protein synthesis, and gene knockdown in mammalian cells using a small-molecule-triggered AR biosensor. The AR platform greatly simplifies and expands genetic circuit creation and opens up new opportunities in protein engineering, synthetic biology, and bioengineering.

2.2 Results

2.2.1 Biophysical feasibility of proximity-dependent split RNAPs

Two key biophysical requirements for our proposed AR strategy are that fused protein domains do not sterically interfere with the split RNAP and that interactions of fused domains can influence the RNAP assembly process. We chose to deploy T7 RNAP split at position 179 because (i) the N-terminal half is small, (ii) structural data indicate that this position is solvent exposed and removed from the DNA-binding face of the protein, and (iii) mutations that influence DNA promoter specificity are C terminal to this position.[162,164,165]

First, we validated that the two RNAP halves spontaneously assemble using an *Escherichia coli* luciferase reporter system (Figure 2.1b and Figure B.1 and Table B.1). We then fused leucine zipper peptides that form a tight interaction with one another (ZA and ZB)[166,167] to the split RNAP halves. Fusion of ZA or ZB to only the N-terminal or C-

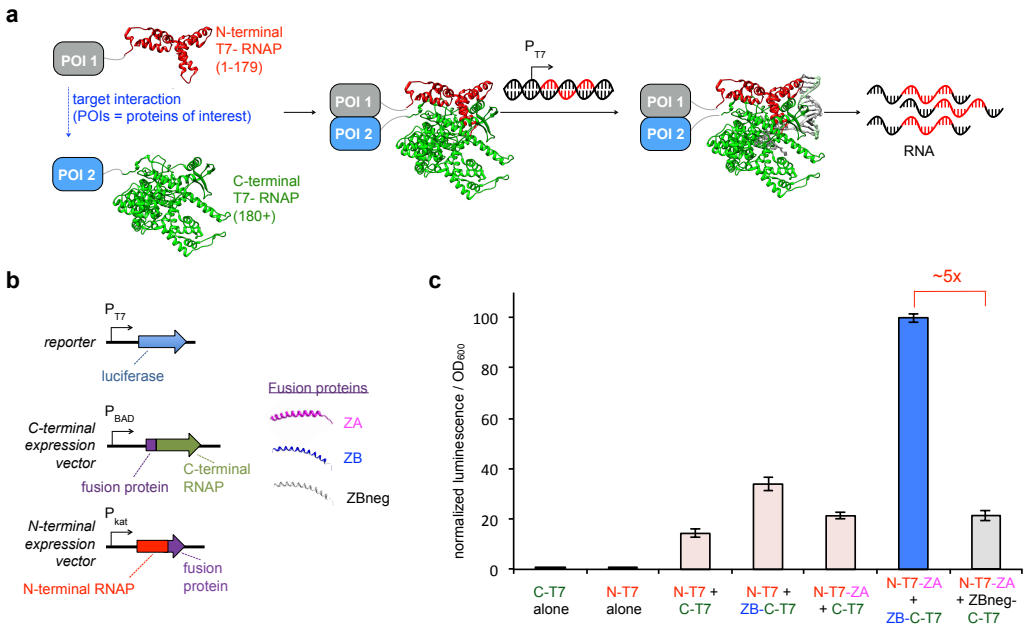

Figure 2.1: Design and biophysical feasibility of activity-responsive RNAPs (ARs) based on proximity-dependent split RNAPs. (a) Schematic of AR design. Split T7 RNAP assembles into a functional RNAP when proteins of interest (POIs) fused to each half interact, resulting in transcription of a user-defined sequence of RNA from a supplied DNA substrate. P_{T7}, T7 promoter. (b) Vectors designed to test split RNAPs *in vivo*. N-terminal split RNAP (red) and C-terminal split RNAP (green) were fused to ZA (pink), ZB (blue), or ZBneg (gray). P_{kat}, kat promoter. (c) Transcriptional output of split RNAPs with fusion proteins assayed in *E. coli* using the vectors shown in b. Cells were induced for 2 h with arabinose then analyzed for luminescence. Error bars, s.e.m., n = 4 biological replicates. Data normalized to signal N-T7-ZA-ZB-C-T7 interaction.

terminal RNAP, respectively, did not dramatically affect spontaneous RNAP assembly, indicating that the split RNAP can tolerate fusions. However, fusion of both RNAP halves to the interaction partners resulted in a 5-fold enhancement in transcription, indicating that additional pendant interactions enhance split RNAP assembly (Figure 2.1c). A triple mutant of ZB (ZBneg) that weakens the interaction confirmed that the enhancement in transcription was due to the fused PPI. We note that T7 RNAP undergoes large conformational changes during the course of its enzymatic activity [168], including dramatic structural changes in the N-terminal RNAP half. It is therefore not obvious that the RNAP can tolerate fusions or that the interactions between fusions can modulate the assembly process while maintaining enzymatic viability. However, because these preliminary data confirmed

that the split enzyme is amenable to controlled assembly, we turned our attention to engineering the split RNAP to be more dependent on the fused interaction partners.

2.2.2 Development of an evolution system to optimize split RNAPs

For split RNAPs to work as a platform for biosensor design, the RNAP assembly process needs to be more dependent on fused interaction partners. Achieving this involves tuning the assembly of the RNAP halves while maintaining all other aspects of RNAP enzymatic function, including DNA binding, nucleotide binding, and RNA synthesis. Such a protein engineering problem presents substantial challenges, but these can in principle be overcome by molecular evolution. We chose to deploy PACE, a rapid evolution system[169] that has been used to evolve RNAP promoter specificity, protein–DNA interactions, protease activities, and PPIs.[150,164,169–174] Briefly, PACE involves providing an evolving gene of interest to M13 bacteriophage, linking the life cycle of the phage to an activity of interest to be evolved in the target gene, and then propagating the virus until the activity evolves. Expression of gIII, a required phage gene, is the basis of the life cycle link.

We envisioned a PACE system for the evolution of selective assembly for split RNAPs using leucine zipper peptides as a model PPI. In this system, the phage would carry an evolving N-terminal RNAP variant fused to ZA, and the host *E. coli* cells would express two different C-terminal variants, each with orthogonal DNA promoter specificity, fused to either ZB or ZBneg. Assembly of the evolving phage-carried ZA-fused N-terminal RNAP with the ZB-fused C-terminal RNAP in the host cells would result in enhanced phage propagation, whereas assembly with the ZBneg-fused C-terminal RNAP would decrease phage propagation. We postulated that this simultaneous positive–negative selection would result in the most robust evolutionary outcome. We would use relatively long (6–8 amino acids), unstructured linkers to tether the fusion proteins to the RNAP halves, enforcing the evolution of a mechanism for proximity-dependent RNAP assembly that is less dependent on geometry and linker composition, which we hypothesized would result

in a more versatile biosensor platform.

To develop this new PACE system, we engineered M13 phage by replacing gIII with N-terminal T7 RNAP fused to ZA, which is the target of the evolution (Figure 2.2a). *E. coli* cells were engineered with a 'positive accessory' plasmid (posAP) that expresses a ZB-fused C-terminal T7 RNAP variant (C-term CGG RNAP) containing seven point mutations that allow it to act selectively on the CGG promoter over the T7 promoter[161,165] and CGG promoter–driven gIII (Figure B.1i, j). We also engineered a negative accessory plasmid (negAP), which expresses wild-type C-terminal T7 RNAP fused to ZBneg and a dominant-negative form of gIII under control of the T7 promoter[171] (Figure B.1k). Therefore, if an evolving N-terminal RNAP variant assembles efficiently with the T7 and CGG C-terminal RNAP halves regardless of the fusion protein, phage production will be blocked. However, phage encoding N-terminal variants that selectively assemble with the ZB-fused CGG C-terminal half will replicate more efficiently and continue to mutate until the interaction is optimized (Figure 2.2b). The only differences between the positive and negative selection are whether or not the fused peptides interact, which is based on three point mutations in ZBneg, and the seven mutations that alter DNA binding of the CGG C-terminal RNAP, which are not at the protein–protein interface and are not expected to alter RNAP assembly.

2.2.3 Evolution of a proximity-dependent split RNAP using PACE

After cloning and validating the system components, we initiated PACE. We modified the positive and negative selection pressures by carefully tuning the system components (Figure 2.2c and Table B.2) and monitored the progress of the evolution by activity-dependent plaque assays and genetic analysis of the evolving phage (Table 2.1). Specifically, we altered the concentrations of the on-target and off-target interactions and the strength of selection of the RNA output by tuning the ribosome-binding sites (RBSs) controlling each system component. After 3–4 d of PACE on a given target, we would use

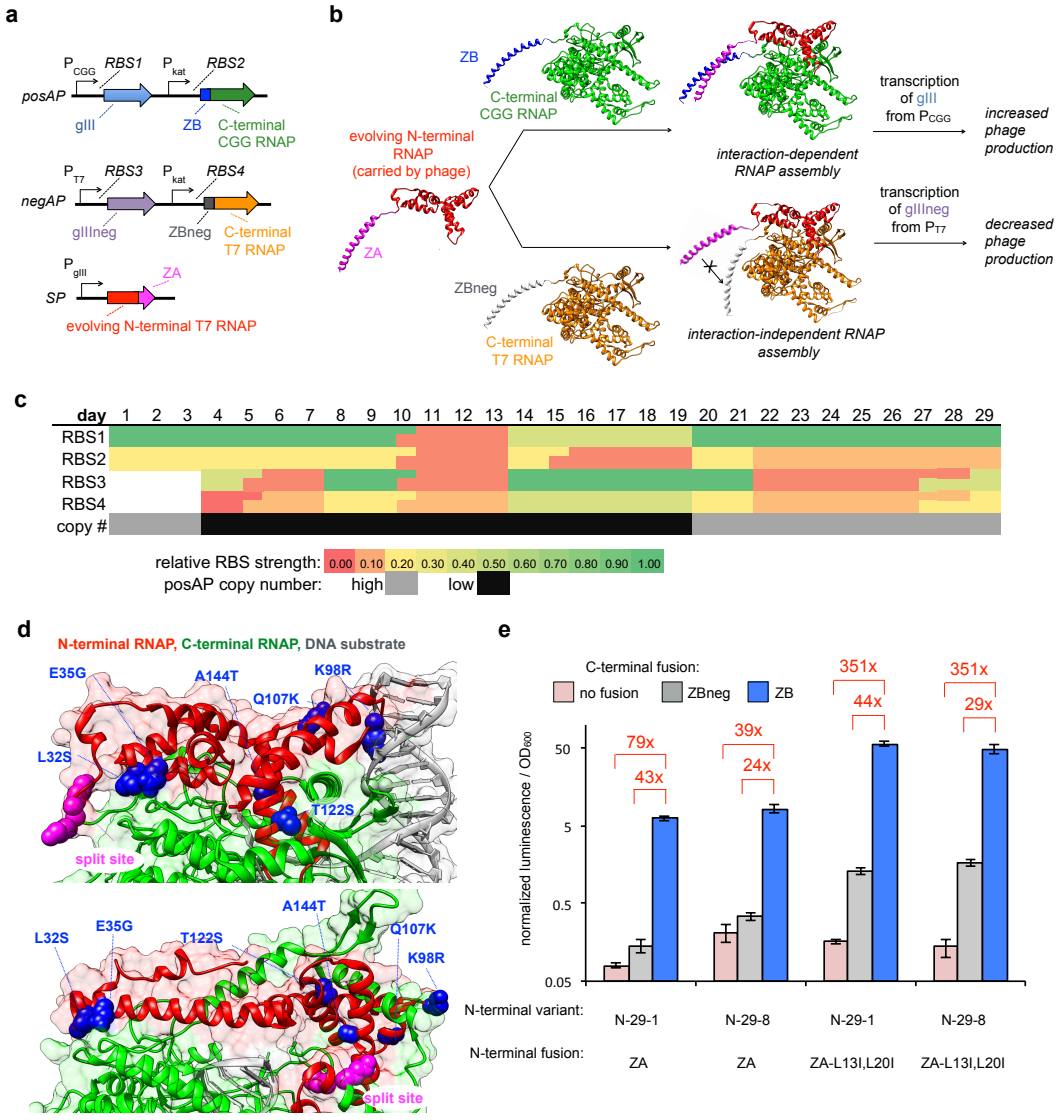

Figure 2.2: Evolution of a proximity-dependent split RNAP for PPI detection. (a) Vectors for PACE system for proximity-dependent RNAPs. P_{CGG}, CGG promoter; P_{T7}, T7 promoter; P_{gIII}, gIII promoter, P_{kat}, kat promoter. (b) Mechanism of PACE system for proximity-dependent RNAPs. Phage carries an evolving N-terminal RNAP fused to ZA, which can assemble with either a C-terminal RNAP variant fused to ZB to allow phage replication or to the noninteracting ZBneg, which hinders phage production by producing a dominant-negative form of gIII (gIIIneg). (c) Schematic of evolution parameters used during PACE. The RBS strengths[175] controlling expression of the C-terminal RNAPs, RBSs controlling proteins on vectors shown in a, and posAP copy number were carefully tuned during a 29-d evolution. (d) Mapping the mutations of N-29-1 onto T7 RNAP crystal structures. Top, initiation complex (PDB 1QLN); bottom, elongation complex (PDB 1H38). (e) Transcriptional reporter assay of N-29-1 and N-29-8 fused to ZA or the evolved ZA double mutant, interacting with either the C-terminal RNAP alone (pink) or fused to ZB (blue) or ZBneg (gray). Error bars, s.e.m., n = 4 biological replicates. Data are plotted on a logarithmic scale to show background.

day 3																					
N-3-1																					
N-3-2																					
N-3-3																					
N-3-4									D87N							S128N					
N-3-5		D26Y											I109N								
N-3-6																					
N-3-7																					
N-3-8																					
day 7																					
N-7-1			L32S			E63K					K98R	Q107K									
N-7-2			L32S			E63K					K98R	Q107K									
N-7-3			L32S								K98R	Q107K	I109T								
N-7-4						E63K					K98R	Q107K									
N-7-5						E63K					K98R	Q107K									
N-7-6			L32S							E91G	K98R	Q107K									
N-7-7			L32S			E63K					K98R	Q107K									
N-7-8	E25D		L32S							E91G	K98R	Q107K									
day 8																					
N-8-1			L32S							E91G	K98R	Q107K									
N-8-2			L32S			E63K					K98R	Q107K									
N-8-3						E63K					K98R	Q107K									
N-8-4			L32S								K98R	Q107K	I109T							G152D	
N-8-5			L32S			E63K					K98R	Q107K									
N-8-6						E63K					K98R	Q107K									
N-8-7			L32S	E35D						E91G	K98R	Q107K									
N-8-8			L32S			E63K					K98R	Q107K									
day 13																					
N-13-1			L32S	E35G						E91G	K98R	Q107K						A144T			
N-13-2						E63K					K98R	Q107K									
N-13-3		D26G				E63K					K98R	Q107K						A144T			A159S
N-13-4			L32S	E35G						E91G	K98R	Q107K									
N-13-5			L32S	E35G						E91G	K98R	Q107K									
N-13-6			L32S	E35G						E91G	K98R	Q107K									
day 21																					
N-21-1			L32S	E35G						E91G	K98R	Q107K									
N-21-2			L32S	E35G						E91G	K98R	Q107K							A149T		
N-21-3			L32S	E35G						E91G	K98R	Q107K									
N-21-4			L32S	E35G	A49S					E91G	K98R	Q107K							A149T		
N-21-5			L32S	E35G						E91G	K98R			A124S							
N-21-6			L32S	E35G						E91G	K98R	Q107K									
N-21-7			L32S	E35G			A83T			E91G	K98R	Q107K									
N-21-8			L32S	E35G			A83T			E91G	K98R	Q107K									
N-21-9			L32S	E35G			A83T			E91G	K98R	Q107K									
N-21-10			L32S	E35G		E63K			D87E		K98R	Q107K	T122S					A144T			
day27																					
N-27-2			L32S	E35G						E91G	K98R	Q107K		A124S							
N-27-3			L32S	E35G						E91G	K98R	Q107K		A124S							
N-27-4			L32S	E35G						E91G	K98R	Q107K		A124S							
N-27-5			L32S	E35G		E63K					K98R	Q107K	T122S					A144T			
N-27-6			L32S	E35G		E63K					K98R	Q107K	T122S					A144T			
N-27-7			L32S	E35G		E63K						Q107K	T122S					A144T			
N-27-8			L32S	E35G		E63K		N86S				Q107K	T122S					A144T			
day29																					
N-29-1			L32S	E35G							K98R	Q107K	T122S					A144T			
N-29-2			L32S	E35G		E63K					K98R	Q107K	T122S		T127A		A136D	A144T			
N-29-3			L32S	E35G							K98R	Q107K	T122S					A144T			
N-29-4			L32S	E35G		E63K					K98R	Q107K	T122S					A144T			
N-29-5			L32S	E35G		E63K					K98R	Q107K	T122S					A144T			
N-29-7			L32S	E35G		E63K					K98R	Q107K	T122S					A144T			
N-29-8			L32S	E35G		E63K					K98R	Q107K	T122S					A144T			

Table 2.1: Mutational analysis of evolving split RNAP. Single phage sequenced during the course of the PACE experiment and coding mutations are shown for a set of variants assayed at each time point. The final variants selected for further assay (N-29-1 and N-29-8) are highlighted yellow.

activity-dependent plaque assays to choose the subsequent evolutionary targets. This process continued for 29 d, after which the N-terminal RNAP converged on two main variants, one with six-mutations (N-29-1) and one with seven (N-29-8), with several of the mutations near the interface between the halves of the RNAP (Figure 2.2d). Several mutations that are prevalent in the phage population are present in either solvent-exposed regions of the structure or, less predictably, at the protein–DNA interface.[168,176,177] This suggests epistatic interactions between mutations that tune the protein–protein interface with key mutations elsewhere in the protein, further illustrating why our unbiased directed evolution strategy is optimal for tuning a complex molecular machine such as an RNAP.

Assays of the two primary variants that emerged from PACE in the luciferase transcription reporter system revealed that the background level of transcription with the ZBneg control was much lower than with ZB (Figure 2.2e). Further genetic analysis revealed that ZA, which we assumed was already fully optimized for interaction with ZB, also evolved during PACE, converging on two leucine to isoleucine substitutions (L13I and L20I). Fusion of this ZA variant into N-29-1 and N-29-8 resulted in a dramatic enhancement in assembly of the split RNAP with the ZA and ZB partners but maintained low levels of background with the ZBneg control, with variant N-29-1 showing a 44-fold increase in RNA synthesis based on the interaction (Figure 2.2e). This observation demonstrates that our split RNAP PACE system can be deployed to optimize biomolecular interactions similarly to recent two-hybrid PACE systems.[170]

We hypothesized that the actual background of the evolved split RNAP was lower than that measured with the original ZA and ZB owing to enhancement of the interaction of ZA (L13I, L20I) with both ZB and ZBneg as a result of the isoleucine substitutions. To test this, we assayed the transcriptional output of N-29-1 and N-29-8 fused to ZA (L13I, L20I) with an unfused C-terminal RNAP. As expected, we found a dramatically lower background signal, indicating a measurable affinity for ZA (L13I, L20I) and ZBneg. Therefore, the actual dynamic range of ZA (L13I, L20I)–ZB-driven N-29-1 is >350-fold higher than that of

the background assembly without a fused PPI (Figure 2.2e). These results indicate that the assembly of the evolved RNAP is not only dependent on fused interaction partners but is also dynamically sensitive to the affinity of the interaction. Most notably, PACE yielded the N-terminal RNAP variant N-29-1 for the AR strategy.

2.2.4 Activity-responsive RNAPs (ARs) as light and small-molecule biosensors

Having optimized and validated the AR system for PPIs, we next sought to explore the generality of the approach by developing inducible PPI systems, and therefore targeted light- and small molecule-activated AR sensors. To create a light-activated RNAP, we appended the light-oxygen-voltage 2 (LOV2)–SsrA fusion variant to N-29-1 and SspB from the improved light-induced dimer (iLID-nano) system[178] to the C-terminal RNAP (Figure 2.3a) without additional optimization of linkers, geometry, or concentrations, and assayed the fusions in *E. coli*. Illumination with blue LED light, which induces dimerization of the iLID-nano system, resulted in a 26-fold enhancement in transcriptional output of the light-activated AR, whereas control fusions did not show a light response (Figure 2.3b). The dynamic range of this non-optimized system is notable because the reported light-induced difference in affinity of the iLID-nano system proteins is only 36-fold.[178]

Next, to further explore the versatility of ARs to detect a three hybrid-like interaction, we sought to engineer a small-molecule-activated RNAP. We chose the rapamycin-induced dimerization of the proteins FKBP-rapamycin binding domain (FRB) and FK506 binding protein (FKBP), a workhorse small-molecule-induced dimerization system for many applications,[179] as the next target for our split RNAP system. There are methods to control RNA synthesis with small molecules in prokaryotic systems, but few such methods function well in mammalian systems,[180,181] and many have issues with signal-to-noise ratios and background signal. We replaced the peptide fusions on the *E. coli*

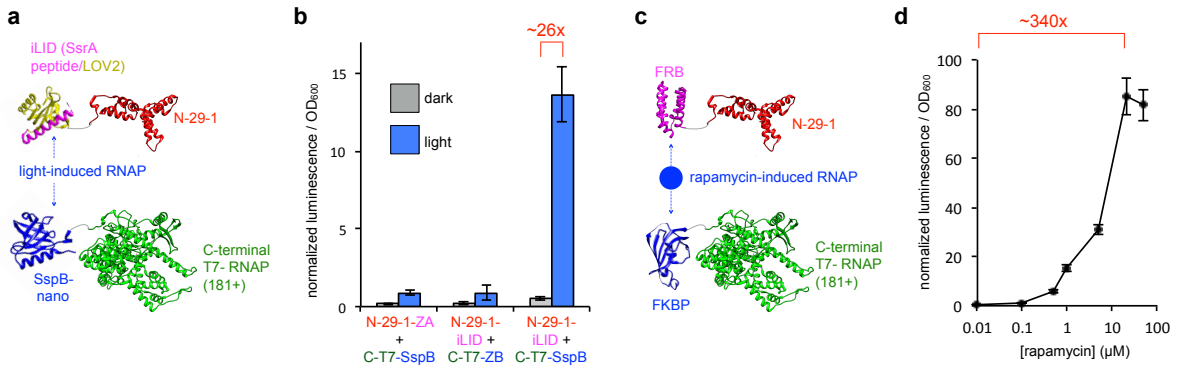

Figure 2.3: Small-molecule- and light-responsive activity-responsive RNAPs (ARs). (a) Light-activated RNAP design using the iLID-nano system. (b) Transcription response of the light-activated RNAP system in *E. coli*. Cells were transformed with expression vectors for the halves of the light-inducible RNAP and a reporter vector then either kept in the dark or illuminated with blue LED light for 3 h before transcriptional analysis. (c) Small-molecule-responsive RNAP design using FRB and FKBP. (d) Transcription response of the rapamycin-inducible RNAP system in *E. coli*. Cells were transformed with expression vectors for the halves of the small-molecule-inducible RNAP and a reporter vector then induced with rapamycin for 3 h before transcription analysis. Error bars, s.e.m., n = 4 biological replicates (b,d).

expression vectors for the C-terminal RNAP and N-29-1 with FKBP and FRB, respectively, again without optimization of linkers, concentration, geometry, or any other system component (Figure 2.3c). We observed a dramatic, dose-responsive increase in RNA synthesis, assessed in the *in vivo* luciferase transcription reporter assay, upon treatment with rapamycin, with a 340-fold enhancement in RNA synthesis and essentially undetectable background (Figure 2.3d).

Compared with traditional n-hybrid strategies that require extensive linker and geometric optimization,[170] the ARs appeared to be much less dependent on linkers. Using the rapamycin-inducible AR as a prototype, we experimentally assessed the effect of linker length on the split RNAP assembly process by varying the linker lengths of the fusion proteins from 2 to 14 amino acids. Transcriptional assays revealed minimal linker-length dependency, and the longer linkers performed slightly better (Figure 2.4). This might be due to the large conformational changes that occur at the split site (Figure 2.2d), as they may be better accommodated by longer linkers. Regardless of the mechanism, these data

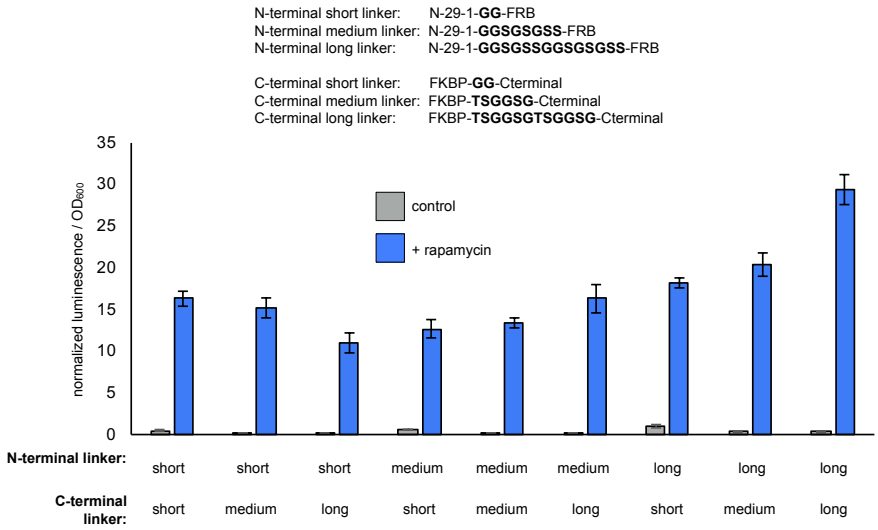

Figure 2.4: Linker effects of AR system. Transcription response of the rapamycin-inducible RNAP system in *E. coli*. Cells transformed with expression vectors for the two halves of the small molecule-inducible RNAP and a reporter vector, then induced with either DMSO or 20 µM rapamycin for 3 h prior to transcription analysis (error bars std. error, n = 5). The linkers tethering the binding domains to each half of the split RNAP were varied as indicated in the figure.

demonstrate the ability to swap new binding domains into the AR system with minimal optimization to create biosensors.

2.2.5 Activity-responsive RNAPs (ARs) can monitor multidimensional PPI networks

One core advantage of an RNAP-based biosensor for analyzing endogenous molecular interactions is that the signals are encoded in an output RNA. Fluorescent-based split biosensors have problems with spectral overlap, differential binding affinities, and linker dependencies, limiting their capability for multidimensional analysis.[182] RNA, however, potentially permits highly multidimensional analysis. To explore this possibility, we next sought to test whether ARs with orthogonal RNA outputs could be used to monitor dynamic, multidimensional PPIs. We engineered a well controlled, synthetic trimolecular PPI network with an inducible change in interactions to validate the concept. We first en-

gineered N-29-1 fused to both ZA and FRB (FZ-N), which we could then deploy with a C-terminal CGG RNAP variant fused to ZB (Z-C_G) or ZBneg (Z_{neg}-C_G) and C-terminal T7 RNAP fused to FKBP (F-C_7) (Figure 2.6a). In this design, interaction between FZ-N and Z-C_G should produce an RNA signal from the CGG promoter, whereas interaction between FZ-N and F-C_7 should produce an RNA signal from the T7 promoter (Figure 2.6b).

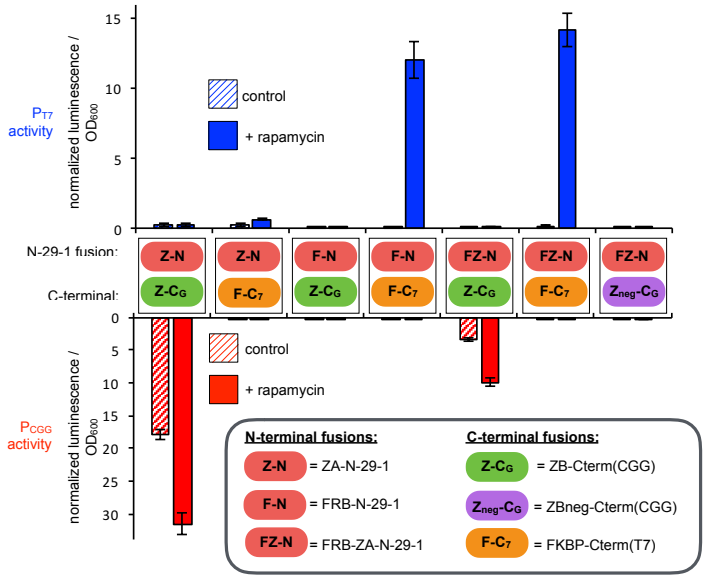

Figure 2.5: Validation of trimolecular interaction system. Transcriptional response and selectivity in *E. coli* of bimolecular interactions of the synthetic parts shown in Fig.2.6a. Cells transformed with expression vectors for N-29-1 fused to ZA ("Z-N", p5-74), N-29-1 fused to FRB ("F-N", p5-79), or N-29-1 fused to FRB and ZA ("FZ-N", pJin200); either Z-CG (p3-13), Zneg-CG (p4-32), or F-C7 (p5-39); and a reporter vector with either T7- (p2-22) or CGG promoter (p2-64) driven luciferase. The cells were treated with either DMSO or 20 µM rapamycin for 2 h prior to transcriptional analysis (error bars std. dev., n = 4). As predicted, the FZ-N-F-C7 interaction is rapamycin-inducible and selectively drives transcription from the T7 promoter, while the FZ-N-Z-CG interaction is constitutive and selective for the CGG promoter.

We first validated the engineered trimolecular model using our luciferase reporter system with one interaction at a time, demonstrating that each interaction can be monitored and selectively acts only on the prescribed promoter (Figure 2.5). Next, to simultaneously monitor both interactions in the same cells, we redesigned the vector system to express both F-C_7 and Z-C_G or Z_{neg}-C_G along with FZ-N and produce DsRed from the CGG promoter and luciferase from the T7 promoter, allowing both transcriptional outputs to be

monitored. We found that the FZ-N–Z-C_G interaction is dependent on the zipper peptides and is unresponsive to rapamycin, whereas the FZ-N–F-C_7 interaction is induced upon rapamycin addition (Figure 2.6c). These data confirm that orthogonal C-terminal variants can be used along with N-29-1 to monitor multidimensional PPI networks in live cells.

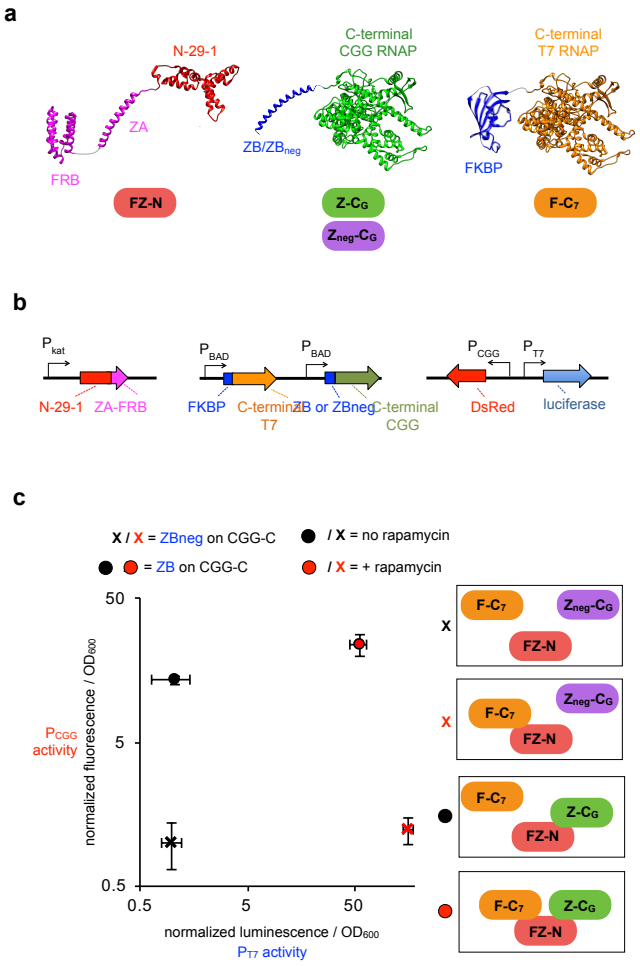

Figure 2.6: Multidimensional PPI detection by activity-responsive RNAPs (ARs). (a) Design of a synthetic trimolecular protein interaction network using FZ-N and Z-CG or Zneg-CG and F-C7. In the absence of rapamycin, FZ-N–Z-CG should be the dominant PPI, driving a CGG-promoter output. In the presence of rapamycin, FZ-N–F-C7 should also be present. (b) Vectors used to monitor the two PPIs simultaneously. (c) Simultaneous monitoring of PPIs in the same cells. E. coli cells were transformed with expression vectors, induced with DMSO or 10 µM rapamycin for 5 h, then analyzed for luminescence and DsRed fluorescence. Error bars, s.e.m., n = 4 biological replicates.

2.2.6 *Activity-responsive RNAPs (ARs) can control RNA synthesis in mammalian cells*

Finally, we assayed the ability of the ARs to function in mammalian cells using the rapamycin-inducible system as an exemplar. We generated rapa-T7, a vector that expresses the rapamycin inducible AR and contains a T7-promoter-driven gene of interest (GOI) output circuit (Figure 2.7a). To confirm that RNA was being generated and measure the kinetics of AR activation, we first deployed a fluorescent aptamer (F30-2xdBroccoli)[183] as the GOI output of the rapa-T7 vector (rapa-T7-F30-2xdBroccoli) to enable visualization of RNA synthesis with fluorescence microscopy. Treatment of HEK293T cells transfected with the rapa-T7-F30-2xdBroccoli vector with 100 nM rapamycin for 30 min resulted in robust enhancement of intracellular fluorescence compared to cells not treated with rapamycin (Figure 2.7b), demonstrating the fast kinetics of a T7 RNAP-based biosensor. Next, to assess whether we could trigger protein production and to assay processivity of the evolved split RNAP, we set IRES-driven GFP mRNA (mRNA(GFP)) as the GOI on the rapa-T7 vector (rapa-T7-mRNA(GFP)). Again, the background fluorescence in cells transfected with rapa-T7-mRNA(GFP) was low, but addition of 10 nM rapamycin resulted in a strong enhancement in GFP fluorescence (Figure 2.7c).Finally, to test dose responsiveness and whether ARs could trigger genetic changes to the cell, we tested whether RNA interference (RNAi) is a viable output. For this, we set small hairpin RNA (shRNA) targeting GFP as the GOI in the rapa-T7 vector (rapa-T7-shRNA(GFP)), cotransfected cells with both a constitutive GFP expression vector and the rapa-T7-shRNA(GFP) vector, and analyzed GFP production by flow cytometry. Induction with rapamycin resulted in a dose-dependent knockdown of GFP signal (Figure 2.7d). Collectively, these results demonstrate that ARs function in mammalian cells and can trigger a variety of outputs via the RNA signal.

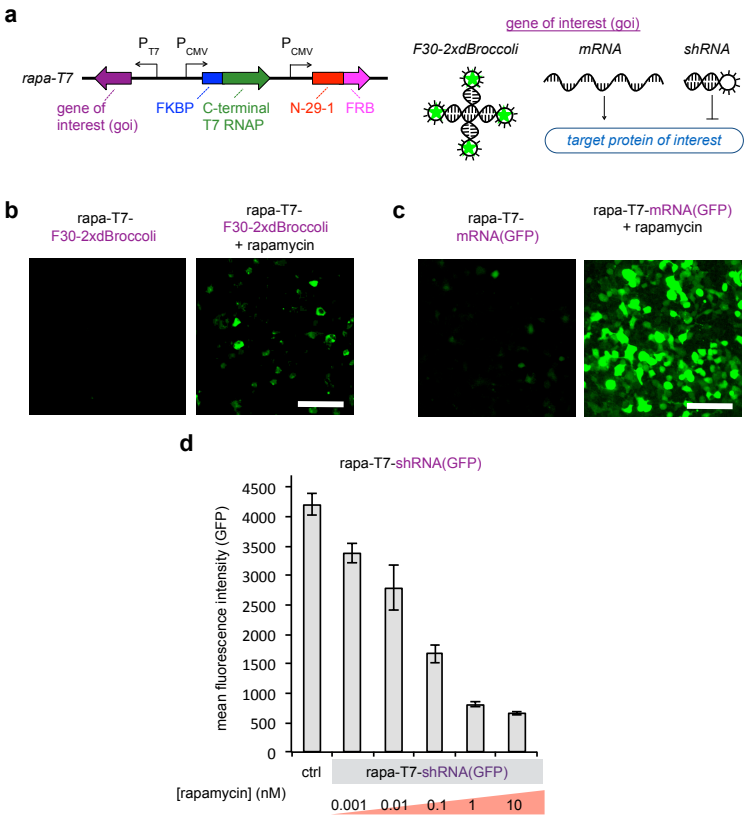

Figure 2.7: Activity-responsive RNAPs (ARs) can trigger a variety of outputs in mammalian cells. (a) Design of the rapa-T7 vector for rapamycin-induced transgene expression in mammalian systems. (b) Validation of the rapa-T7 vector with a fluorescent RNA aptamer as the output. HEK293T cells were transfected with rapa-T7-F30-2xdBroccoli and induced with 0 or 100 nM rapamycin for 30 min in the presence of 20 µM DHFBI-1T then analyzed by fluorescence microscopy. (c) Validation of the rapa-T7 vector with mRNA as the output. HEK293T cells transfected with rapa-T7-mRNA(GFP) were induced with 0 or 10 nM rapamycin overnight then analyzed by fluorescence microscopy. (d) GFP fluorescence (arbitrary units) analyzed by flow cytometry to validate rapa-T7 with shRNA as the output. HEK293T cells were transfected with a GFP expression vector and a rapa-T7-shRNA(GFP) vector and induced with varying concentrations of rapamycin for 44 h. Error bars, s.e.m., n = 3 biological replicates. Scale bars, 100 µm (b,c).

2.3 Discussion

Here we present the design, optimization, and deployment of a new split RNAP as a versatile biosensor platform. After discovering that fused PPIs can modulate the assembly of T7 RNAP split at position 179, we developed a PACE system to evolve an optimal proximity-dependent split RNAP. PACE yielded a variant with a 350-fold dynamic range

based on fused leucine zipper peptide PPIs. We then explored the versatility of the split RNAP system by generating light- and small-molecule-activated RNAPs through swapping in different combinations of fusion proteins. Without any optimization, the resulting sensors had between 26- and >300-fold dynamic range. Moreover, we used a model trimolecular PPI system to demonstrate that multidimensional PPI networks can be monitored using orthogonal RNAP sensors with different RNA outputs. Finally, we demonstrated the versatility of RNAP-based sensors by showing that a small-molecule-driven RNA output can be used to synthesize fluorescent RNA nanostructures, generate proteins, or knock down genes in mammalian cells.

A challenge with traditional n-hybrid approaches is that the linker lengths, compositions, and geometries of the parts need to be carefully tuned for each new interaction to be interrogated, measured, selected, or evolved. In practice, this often means months of cloning, screening, and careful optimization, and for some targets, steric or geometric concerns preclude detection. With this in mind, we sought to create the ARs in a manner that is less dependent on these variables, which we postulated would streamline the process of engineering new sensors. From a first approximation, tuning the assembly of a split RNAP to be more dependent on fusion proteins would involve weakening the affinity of the two RNAP halves. However, such a thermodynamic design is likely to result in a high dependency on linkers and composition of the fusion proteins, which would make up the loss of binding energy. Therefore, we deliberately made the evolution of the ARs more challenging by tethering the fusion zipper peptides to the RNAP halves with long (6–8 amino acids), flexible linkers. In this way, the RNAP assembly process would be likely to evolve altered thermodynamics and kinetics of assembly, because the flexible linkers would not allow sufficient thermodynamic gain in binding energy between the fusion proteins. Further studies delineating the mechanism of controlled assembly of the evolved proximity-dependent RNAPs could test this hypothesis and improve engineering efforts. However, our results showing that the AR system worked well with diverse new binding

partners suggest that it functions as we expected, whatever the evolved mechanism of RNAP assembly is. Moreover, our observation that linker length had little effect on RNAP activity suggests that thermodynamics is not the primary driving force.

Although PACE is a powerful directed-evolution platform, a fundamental limitation of the technology, and all *in vivo* evolution systems, is the challenge of designing genetic circuits that link target activities to fitness.[150] ARs provide a new, robust mechanism to link target activities to gene expression and fitness for applications in directed evolution. It will be interesting to explore whether other target protein-interaction domains can replace the leucine zipper peptides for engineering by continuous evolution. Even more broadly, our system can also be adopted to other three-hybrid-like approaches, where different types of 'baits', such as small molecules or RNA, are displayed on the C-terminal RNAP half. Finally, because this new PACE system utilizes continuous and simultaneous positive and negative selection pressures, the approach should have advantages in terms of evolving selectivity in target proteins, as off-target interactions can be displayed on the orthogonal C-terminal RNAP and drive negative selection.

The AR system provides a new approach to monitor and respond to molecular interactions, molecules, and external cues in live cells. Combined with the ability to generate ARs with orthogonal DNA promoter specificity, these tools open the possibility of performing highly multidimensional interaction network analysis, presaging a new approach to cell analysis using high-throughput sequencing.[159,160] ARs should have advantages in terms of sensitivity owing to the signal amplification of the RNA output made possible by PCR. Unlike fluorescence-based technologies, which are solely for analysis, the RNA outputs of the ARs can be engineered to store information or drive cell fate changes on the basis of measured events. Engineering AR-based gene circuits that simultaneously measure and compute both pathologically relevant endogenous events and external cues provides a new strategy to develop 'smart' genetic therapies. Although we demonstrated PPI-, small-molecule-, and light-activated systems here, in principle any existing fluorescent

protein-based probes could be immediately transported into the AR system for integration with synthetic biology applications. We anticipate that ARs will provided a simplified and more robust strategy for engineering gene circuits for applications in screening, directed evolution, and synthetic biology.

2.4 Methods

2.4.1 Cloning

All plasmids were constructed by Gibson Assembly[184] from PCR products generated using Q5 Hot Start DNA Polymerase (NEB) or Phusion Polymerase. Phage were cloned by Gibson Assembly of the split N-terminal RNAP-ZA fusion into a previously optimized SP phage backbone[172] and transformation into 1059 cells,[171] which supply gIII in an activity-independent manner. After overnight growth in medium, the supernatant was isolated by centrifugation and plaque assays were performed on 1059 cells. Single plaques were selected for overnight growth and sequencing to identify clonal phage samples with the correct insert. All plasmids and phage were sequenced at the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. All new vectors are described in table B.1 and maps for each plasmid are shown in figure B.1.

2.4.2 Sequence of split RNAP fusions

Listed below are the structures and sequences of the leucine zipper peptide fusions, iLID-nano light-induced dimerization fusions, and rapamycin-induced dimerization constructs used in this study. Linkers are enclosed in square brackets. The three point mutations between ZB and ZBneg are underlined and were obtained from previous studies,[167] and the two mutations in ZA that evolved during PACE are also underlined.

N-ZA: RNAP(1-179)-[GGSGSGSS]-ALKKELQANKKELAQLKWELQALKKELAQ

C-ZB: MASEQLEKKLQALEKKLAQLEWKNQALEKKLAQ-[TSGGSG]-RNAP(180+)

C-ZBneg: MASEQLEKELQALEKELAQLKWKNQALEKKLAQ- [TSGGSG]-RNAP(180+)

N-ZA-(L13I,L20I): N-28-1(1-179)-[GGSGSGSS]-ALKKELQANKKEIAQLKWEIQALKKELAQ

N-term-iLID: N-28-1(1-179)-[GGSGSGSS]-iLID

SspB-C-term: SspB nano-[TSGGSG]-RNAP(180+)

N-term-FRB: N-28-1(1-179)-[GGSGSGSS]-FRB

FKBP-C-term: FKBP-[TSGGSG]-RNAP(180+)

N-ZA-FRB: RNAP(1-179)-[GGSGSGSS]-ZA-[GGSAGGSG]-FRB

2.4.3 In vivo transcription assays of split RNAPs

N- and C-terminal halves of the RNAP were cloned into expression vectors, with the N-terminal RNAP under a constitutive promoter and the C-terminal RNAP under the arabinose inducible promoter. S1030 cells[171] were transformed by electroporation with three plasmids: (i) an N-terminal RNAP expression plasmid, (ii) a C-terminal expression plasmid, and (iii) a reporter plasmid that encodes luciferase under control of the T7 promoter. The transformed cells were then plated onto agar plates (15 g/L in LB) with 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, and 10 mM glucose. Single colonies were grown to saturation overnight at 37 °C, and each well of a 96-well deep-well plate containing 0.54 mL of LB with antibiotics and 10 mM arabinose was inoculated with 60 µL of the overnight culture. After growth with shaking at 37 °C for 2 h, 150 µL of each culture was transferred to a 96-well black-wall, clear-bottom plate (Costar), and luminescence and OD_{600} was measured on a Synergy H4 Hybrid Reader (BioTek). The data were analyzed by dividing the luminescence values by the background-corrected OD_{600} value, then subtracting out the background from the reporter vector alone. All val-

ues were then normalized to the wild-type split RNAP fused to ZA and ZB (Figure 2.1c), which was assigned an arbitrary value of 100, allowing the values from each luminescence plot to be compared to one another. For the light-activated system, the experiment was performed identically except upon outgrowth for 3 h, cells in the light condition were illuminated with a homemade blue LED lightbox and cultured at 25 °C, and cells in the dark condition were cultured at 37 °C; this was done to correct for heat output from the light source to maintain similar temperatures in both conditions. For the rapamycin-inducible system, the experiment was performed identically, except upon outgrowth, rapamycin was added for 3 h, then luminescence was analyzed. Sample size ($n = 4$ biological replicates for each condition) was determined on the basis of previous work using a similar *in vivo* luciferase reporter system and provided excellent reproducibility both between biological replicates on a given day and between days of experimental replicates.

2.4.4 Phage-assisted continuous evolution (PACE)

PACE was carried out to evolve the split T7 RNAP variants using a modified version of previously described methods.[160] *E. coli* strain S1030 was transformed by electroporation with combinations of the positive accessory plasmid (posAP), negative accessory plasmid (negAP), and mutagenesis plasmid (MP).[172] 5 mL starter cultures were grown overnight in LB supplemented with antibiotics and 10 mM glucose. Chemostats (100 mL sterile bottles) containing 80 mL of Davis rich medium[171] were inoculated with 2 mL starter culture and grown at 37 °C with magnetic stir-bar agitation. At approximately OD_{600} 1.0, fresh Davis rich medium was pumped in at 60-80 mL per h, with a waste needle set at 80 mL. 10 µL phage was used to seed a fresh lagoon (25 mL flask with a rubber septum). To initiate the evolution, a monoclonal phage population was used. Waste needles were set to maintain the lagoon volume at 15-20 mL, and host cell cultures were flowed in at 15–17 mL per h. Arabinose (10% w/v in water) was added directly to lagoons via syringe pump at 1.0 mL per h to induce mutagenesis. A lagoon sample was taken from the waste withdrawal line every

24 h and centrifuged, and the supernatant was stored at 4 °C. The complete evolutionary protocol is described in table B.2After the completion of each leg of the evolution, activity-dependent plaque assays were used to select the next evolutionary target, and the PACE experiment was again initiated as described, using 10 µL phage from the previous endpoint of the evolution. The strength of the positive and negative selection pressures were varied by altering the ribosome binding sites (RBSs)[175] controlling the expression of each of the C-terminal target fusions, gIII and a dominant-negative form of gIII (gIIIneg). Mixed selection pressures from multiple posAP and negAP sets at a given time point were used as appropriate to enhance the likelihood of successful evolution.[160,164,174]

2.4.5 Sequence and activity analysis of variants from PACE

Phage samples were boiled for 10 min to lyse the phage and release the genomes. PCR was then used to amplify the DNA library containing the N-terminal RNAP variants, which was then subcloned into vector p3-7. Single colonies were picked from the transformation and subjected to analysis by Sanger sequencing. The results of the sequence analysis during the course of the evolution are shown in table 2.1.Variants N-29-1 and N-29-8 cloned into vector p3-7 were subjected to analysis by the luciferase assays as shown in figure 2.2e. In order to sequence potential mutations that occurred in the peptide fusion in the phage, single plaques from an activity-independent plaque assay were picked, grown overnight, boiled, and analyzed by Sanger sequencing.

2.4.6 In vivo split GFP assay

S1030 cells were transformed by electroporation with two plasmids: (i) a constitutive N-terminal GFP–FRB fusion and (ii) an arabinose inducible FKBP–C-terminal GFP fusion. The transformed cells were then plated onto agar plates (15 g/L in LB) with 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, and 10 mM glucose. Single colonies were

then grown to saturation overnight at 37 °C, and then each well of a 96-well deep-well plate containing 0.54 mL of LB with antibiotics, 10 mM glucose or 10 mM arabinose, and varying concentrations of rapamycin (0 nM, 1 nM, 10 nM, 0.1 µM, 1 µM, 20 µM, or 100 µM) were inoculated with 60 µL overnight culture. After growth with shaking at 37 °C for 3 h, 6 h, or 30 h, 150 µL culture was transferred to a 96-well deep-well plate. The cultures were centrifuged (10 min, 25 °C, 2,000 RCF) and washed with 1 mL PBS three times before suspension in 150 µL PBS and transfer to a 96-well black-wall, clear-bottom plate (Costar). GFP fluorescence (excitation 485 ± 20 nm, emission 516 ± 20 nm) and OD_{600} were measured on a Synergy Neo2 Microplate Reader (BioTek). The data were analyzed by dividing the background-corrected GFP fluorescence values by the background-corrected OD_{600} value. All values were then normalized to the 0 nM rapamycin conditions. (sample size n $=$ 5 biological replicates for each condition).

2.4.7 Dual reporter PPI in vivo detection assays

FRB and ZA peptide were fused to the N-terminal RNAP (N-29-1) in an *E. coli* expression vector to form pJin200. FKBP was fused to the C-terminal of T7 RNAP and ZB or ZBneg was fused to C-terminal of CGG RNAP in one vector to construct the dual C-terminal vectors. Then the CGG promoter–driven DsRed-Express2 circuit was added to p2-22, which contained a T7 promoter–driven luciferase gene to generate pJin216 as the dual reporter vector. To test the multidimensional system, pJin200, pJin216, and either pJin207 or pJin208 were transformed into S1030 cells. The transformed cells were then plated onto agar plates (15 g/L in LB) with 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, and 10 mM glucose. Single colonies were cultured overnight in LB liquid medium with the same antibiotics and 10 mM glucose. The next day, 50 µL overnight cultures were transferred to a deep 96-well plate containing LB with same antibiotics, 10 mM arabinose with 0 or 20 µM rapamycin. After shaking at 37 °C for 5 h, 150 µL culture was transferred to a 96-well deep-well plate. The samples were

centrifuged, washed and prepared in the same way as for the split GFP assay. Then the OD_{600}, luminescence and DsRed-Express2 fluorescence (excitation 555 $\pm$ 15 nm, emission 590 $\pm$ 15 nm) were measured on a Synergy Neo2 Microplate Reader (BioTek). The data were analyzed by dividing the background-corrected luminescence values and DsRed-Express2 fluorescence background-corrected values by the background-corrected OD_{600} value. All luminescence values were normalized to the 0 µM rapamycin conditions with the ZB-fused C-terminal variant, and DsRed-Express2 fluorescence values were normalized to 0 µM rapamycin conditions with the ZBneg-fused C-terminal variant.

2.4.8 Cell culture

HEK293T cells (ATCC) were maintained in DMEM (high glucose, L-glutamine, phenol red, sodium pyruvate; obtained from Gibco (11995-065) or Hyclone (SH30081.01)) supplemented with 10% FBS (Gibco/Life Technologies, qualified US origin) and 1% penecillin–streptomycin (P/S, Gibco/Life Technologies). As HEK293T cells are listed in the database of commonly misidentified cell lines maintained by ICLAC (https://iclac.org/databases/cross-contaminations/), we obtained fresh cells from ATCC, which were frozen down at an early passage (passage 5) in individual aliquots. The cells were then used for $<$ 25 passages for all experiments. Multiple biological replicates were performed with cells from different passages and freshly thawed aliquots. There was no testing for mycoplasma infection or further authentication because early-passage cells were used for all experiments.

2.4.9 Imaging mammalian AR activation by fluorescence microscopy

HEK293T cells cultured in DMEM (high glucose, L-glutamine, phenol red, pyruvate; Gibco/Life Technologies) supplemented with 10% FBS (Gibco/Life Technologies, qualified US origin) were plated on an 8-well coverglass slide (Labtek) and transfected with 600 ng

rapa-T7 vector (pJin141 or p6-8) using 1.5 µL of Lipofectamine 2000 (Thermo Fisher Scientific) using the standard protocol. For the rapa-T7-F30-2xdBroccoli (pJin141) experiments, 100 nM rapamycin or DMSO control was added along with 20 µM DHFB1-1T for 30 min before imaging. For the rapa-T7-mRNA(GFP) (p6-8) experiments, 10 nM rapamycin or DMSO control was added to the sample 20 h after transfection, and then incubated for an additional 24 h. The cells were imaged on an Olympus BX53 microscope using a GFP filter set and a ×10 objective. Each image for a given condition was processed using identical conditions to adjust brightness and contrast to a level where background fluorescence was observed for control samples in ImageJ (NIH).

2.4.10 Flow cytometry

HEK293T cells were cultured in DMEM (high glucose, L-glutamine, phenol red, pyruvate; Gibco/Life Technologies) supplemented with 10% FBS (Gibco/Life Technologies). The day before transfection, cells were passaged and plated at 50,000 cells per well in a 48-well plate (NEST Biotechnology). After 19 h, 50 ng RFP plasmid (p3-62), 200 ng GFP plasmid (p1-53), and 400 ng rapa-T7-shRNA(GFP) vector (pJin140) were transfected into cells using 1.5 µL Lipofectamine 2000 (Thermo Fisher Scientific) using the standard protocol. 30 min after transfection, DMEM supplemented with FBS and either DMSO or rapamycin was added to the wells so the final rapamycin concentration was 0 nM, 0.001 nM, 0.01 nM, 0.1 nM, 1 nM, or 10 nM. 29 h after transfection, the medium was replaced with fresh medium including the previously added concentration of rapamycin. 44 h after transfection, the cells were trypsinized and suspended in DMEM supplemented with FBS and rapamycin, then analyzed on a LSR-Fortessa 4-15 (BD digital instrument, 488 nm laser, 530 $\pm$ 30 nm filter for GFP, and 610 $\pm$ 20 nm filter for RFP). Mean GFP fluorescence intensity was calculated for HEK293T cells expressing RFP using FlowJo. Reported values are average mean GFP fluorescence intensity values taken from 3 separate replicate samples. Sample size (n $=$ 3 biological replicates for each condition) was determined by

initial trial experiments to find the spread in the data.

CHAPTER 3

A PHAGE-ASSISTED CONTINUOUS SELECTION APPROACH FOR DEEP MUTATIONAL SCANNING OF PROTEIN-PROTEIN INTERACTIONS

3.1 Introduction

Protein-protein interactions (PPIs) are critical to cellular signaling, metabolism, and cellular organization.[1] Dysregulated PPIs are often key drivers of disease[2–5] and are therefore compelling targets for therapeutic development.[185–187] PPI dysregulation can emerge from either mis-regulated protein synthesis or post-translational modifications, or mutations in one of the binding partners, resulting in aberrant, disease-contributing interactions.[188] In either case, understanding the molecular determinants of PPIs is crucial to establishing the biophysical basis of disease and assisting in the creation of therapeutic PPI modulators.

Traditionally, a reductionist approach is deployed to understand PPIs, which generally involves alanine scanning to identify key sites for affinity and specificity of the interface between binding partners.[189,190] While these methods have generated a wealth of information, such as the identification of PPI hot spots,[191] they are labor intensive, requiring functional biochemical characterization of each variant, which often limits the number of variants characterized. The advent of display technologies, such as phage[192] and yeast display,[193] coupled with advances in two-hybrid approaches for the detection of PPIs,[194] has allowed the gathering of large amounts of information about the genotype-phenotype relationships of biomolecules[195] and PPIs.[196] These and other evolutionary and selection-based approaches, and their resultant large data sets, paint a more nuanced picture of how individual mutations correlate with altered function.

The application of high-throughput sequencing methods to understand sequence-

function relationships in biomolecules is referred to as deep mutational scanning (DMS).[197,198] DMS generates libraries of protein variants whose individual functional consequences are assayed through selection experiments. Based upon the relative enrichment of each variant in the selection, as monitored by high throughput sequencing (HTS), the functional consequence of each mutation can be deduced. A key component of every DMS experiment is the design and characterization of a selection platform for the desired protein and its relevant properties.[199] Display-based platforms,[200–202] and other assays that directly link protein function to an observable phenotype, such as cell viability[203] or growth,[204] have been used to screen protein libraries.

Continuous evolution methods,[150] which link defined biochemical fitness to viral life cycles, allow rapid and large-scale evolution and selection. However, the *in vivo* nature of viral selection systems requires robust genetic selection systems for a biochemical activity of interest. Phage-assisted continuous evolution (PACE)[169] and phage-assisted non-continuous evolution (PANCE),[205,206] are especially powerful technologies in which large populations of gene-encoding protein variants can be evolved or selected for over time, based on defined biochemical properties. For example, PACE has been used to evolve T7 RNAP promoter specificity,[169,174] protease specificity,[207] drug resistance,[172] Cas proteins,[208] and several other biochemical machines.[150,209] PACE has also been used to evolve PPIs using a bacterial 2-hybrid-based selection system.[170] While this evolutionary system is powerful, it is limited in terms of requiring optimization for each new PPI target, necessitating re-design for each new target of interest.

The design and optimization of a relevant selection platform is often the most challenging and labor-intensive step in DMS experiments and, we reasoned, is one of the key reasons why powerful selection approaches such as PACE and PANCE have not found more widespread use. To address this limitation, we designed a versatile new phage-assisted continuous selection (PACS) system to interrogate the binding interface between interacting proteins. We developed PACS in the context of probing the KRAS/RAF PPI,

due to its high value as a target for cancer therapies; both KRAS and RAF are known oncogenes, with roughly 20% of all cancers harboring mutations in KRAS.[210,211] Practically, we chose to use the KRAS4b isoform due to its high mutation rate across different cancers,[212] and the RAS binding domain (RBD) of RAF, which is the key part of the RAF protein that binds RAS isoforms.[213,214]

Our new system relies on proximity-dependent split RNA polymerase (RNAP) biosensors,[215] greatly simplifying and streamlining the genetic sensing of PPIs *in vivo* compared with more traditional 2-hybrid approaches.[216] Using these split RNAP biosensors, we designed a PACS system that links replicating bacteriophages to the binding of bacterial-expressed KRAS to phage-expressed RAF. After validating the system, we generated libraries of phage-encoded RAF variants and subjected them to PACS. Analysis by HTS revealed whether each site in RAF is tolerant to mutation, and if so, which mutations are tolerated. Aside from providing new insights into the KRAS/RAF interface from the RAF perspective, our new method is likely broadly applicable, facilitating rapid understanding of the mutational landscape of PPIs using DMS.

3.2 Results

3.2.1 Engineering split RNAPs to detect the KRAS/RAF PPI

In PACE, the viral life cycle of M13 bacteriophage is linked to a target activity through the production of gIII, a required phage gene that is moved from the phage genome into host *Escherichia coli* cells.[169] The phage must therefore infect the host cell and produce a phage-encoded protein variant capable of activating gIII production in the host cell in order to replicate. Therefore, the key to developing a system for a desired target activity is linking the target activity of interest, in our case RAF and KRAS binding, to the production of gIII. To do this, we deployed our previously-reported split RNAP-based biosensor platform, based on an evolved variant of split T7 RNAP, that only forms a functional enzyme and

produces a defined RNA output when proteins fused to each half of the split enzyme bind one another.[215] Our versatile split RNAP biosensor technology is able to monitor different PPIs without requiring re-optimization, as evidenced by its successful deployment to detect various PPIs, including PPIs from the BCL-2 family of proteins[217] and small molecules.[218]

First, we assayed whether our split RNAP biosensor can detect the PPI between KRAS and RAF in *E. coli* using a modified version of our previously-reported *in vivo* luciferase assay, in which interactions between KRAS and RAF should drive luciferase production.[215] We designed the *in vivo* luciferase assay with three plasmids: 1) an N-terminal expression plasmid containing the N-terminal fragment of the split RNAP ($RNAP_N$) with a C-terminal fusion of either RAF or a control, non-interacting partner, 2) a C-terminal expression plasmid that constitutively expresses the C-terminal RNAP fragment ($RNAP_C$) with a N-terminal fusion of either KRAS or an off-target protein interaction partner, and 3) a reporter vector with the bacterial luciferase gene (luxAB)[219] under control of the T7 RNAP promoter (Figure 3.1a). Under these conditions, only interaction of the two fusion proteins should result in assembly of the split T7 RNAP biosensor, expression of luciferase, and generation of luminescence (Figure 3.1b).

After cloning and optimizing the system components, we then measured luciferase output in reporter *E. coli* cells expressing either on-target (KRAS/RAF) or control (ZB/RAF; ZB is a leucine zipper peptide that does not bind RAF)[166,215] PPI fusions. As expected, the KRAS/RAF combination yielded robust luciferase signal (Figure 3.1c). Absence of either KRAS or RAF on either half of the biosensor resulted in ~20-fold less signal, confirming KRAS/RAF PPI-dependent gene expression. To further validate the system, we mutated RAF with R89L a known RAS interaction-disrupting mutation.[220] Again, as expected, $RNAP_N$ fused to RAF R89L produced levels of luciferase activity comparable with that of off-target controls (Figure 3.1c). After validating the use of our split biosensor and our *in vivo* luciferase assay for the KRAS/RAF PPI, we next sought to deploy the biosensors in a new PACS system.

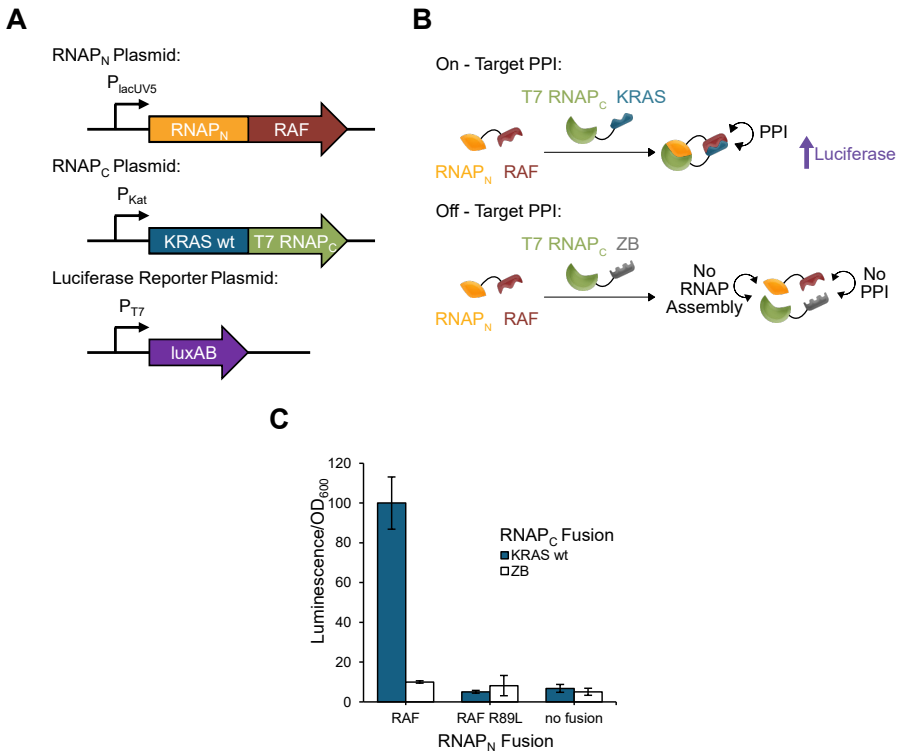

Figure 3.1: Detecting the KRAS/RAF PPI with split RNAP biosensors. (A) Plasmids for the *in vivo* luciferase assay to monitor the KRAS/RAF PPI. (B) Schematic of on-target and off-target PPI partners and their expected output in the *in vivo* luciferase assay system; only pendent protein fusions that interact with each other can drive the reassembly of the split RNAP biosensor to produce the luciferase enzyme and luminescence. (C) Luminescence signal for the on-target and off-target PPIs as monitored by the *in vivo* luciferase assay (error: std. dev., n = 4).

3.2.2 Designing a PACS system for PPIs

To adopt the biosensors into a PACS selection system, we cloned *E. coli* selection plasmids and modified phage to encode the appropriate system components (Figure 3.2a). We modified a previously-optimized M13 bacteriophage genome[215] by removing gIII and replacing it with our split RNAP$_N$ biosensor C-terminally fused to either RAF or an off-target protein as a control. The selection plasmids contained in the *E. coli* selection cells were used to modulate the expression of gIII through the assembly of our split RNAP biosensors, including: 1) a "positive selection plasmid" that contains the CGG RNAP$_C$ half of our RNAP biosensor, which is a known variant of the T7 RNAP that recognizes an orthogonal

CGG RNAP promoter,[165] N-terminally fused to KRAS, and gIII under the control of the CGG RNAP promoter; and 2) a "negative selection plasmid" that contains the T7 $RNAP_C$ half of our RNAP biosensor N-terminally fused to an off target leucine zipper peptide, ZBneg,[166,215] and contains a dominant negative form of gIII, gIIIneg,[221] controlled by the cognate T7 RNAP promoter. With these selection plasmids, only a KRAS-binding $RNAP_N$-fusion should drive reassembly of the split CGG RNAP biosensor and gIII expression, generating infectious phage progeny. Non-PPI-driven assemblies would activate both the positive and negative selection plasmids, resulting in the production of gIIIneg and non-infectious phage. In other words, the negative selection plasmid prevents the phage from "cheating" the selection by maintaining proximity-dependency in $RNAP_N$.

After constructing the selection cells and phage, we validated the ability of the system to discriminate between phage encoding the known binding partner of KRAS, RAF, and an off-target protein, an isoleucine zipper peptide termed iZA,[166,215] using activity-dependent plaque assays. Because this assay informs upon the ability of a given phage to replicate and infect neighboring cells by forming "plaques," we were able to directly monitor if the proteins fused to the $RNAP_N$ biosensor contained in the phage activated the positive selection plasmid in the selection cells, producing gIII. Indeed, only the phage containing $RNAP_N$ fused to RAF were able to form distinct plaques on the selection cells, validating the system components (Figure 3.2b).

Next, we sought to validate the selection system in a mock PACS experiment using spike-in samples to enrich an active phage variant from an excess of inactive variants (Figure 3.2c). To do this, we seeded a phage vessel with 1,000-fold excess iZA phage to RAF phage. Fresh *E. coli* selection cells containing the selection plasmids were continuously flowed in from a chemostat, allowing constant diluting inflow and waste outflow. We monitored the selection over the course of 24 h, sampling the phage at various time points. We assayed the relative distribution of the inactive iZA phage versus the active RAF phage by PCR amplification of the phage samples, using primers that selectively amplify the

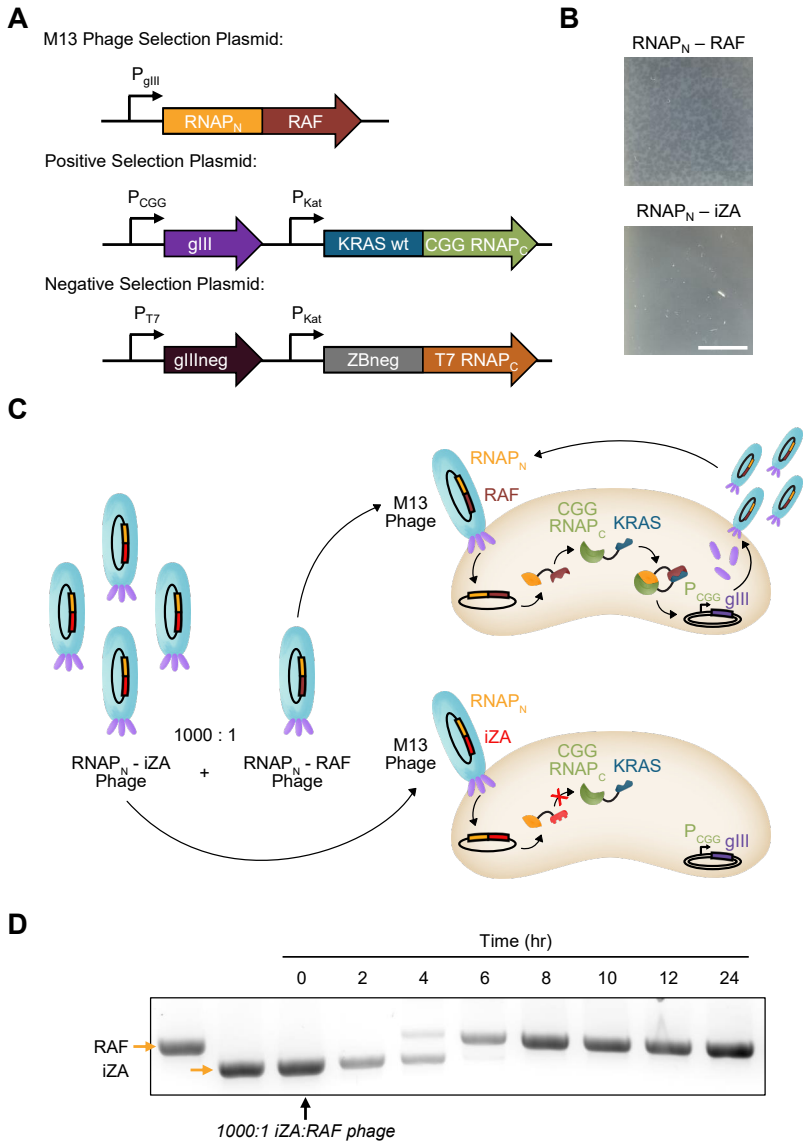

Figure 3.2: Design of a phage-assisted continuous selection (PACS) system for PPIs. (A) Plasmids for the *in vivo* PACS system to monitor the KRAS/RAF PPI. (B) Plaque assays of on-target and off-target PPI fusions on the PACS selection cells (scale bar: 1 cm). (C) Schematic of the mock selection with and the on-target RAF phage and off-target iZA phage; only the RAF phage will be able to replicate on the selection cells and produce gIII to generate infectious phage progeny during the course of the selection. (D) DNA agarose gel depicting the relative fraction of iZA and RAF phage over the course of the mock selection experiment.

phage-encoded gene product. Indeed, although there was no detectable RAF in the initial phage population used to seed the experiment, the population of RAF-encoding phage increased over time, while the inactive iZA phage were diluted out, with complete washout and enrichment after 6 h (Figure 3.2d).

3.2.3 Validating PACS system for PPIs by evolution

After validating the system components in a mock selection, we next sought to confirm the system was capable of evolving PPIs from non-interacting variants. In addition to the aforementioned R89L variant, we cloned a series of RAF mutants that we predicted would disrupt the KRAS/RAF PPI, informed by either literature or intuition from the crystal structure.[222] These variants included: K84A, K84E, Q66A, and Q66K. We first assayed the mutants in the luciferase reporter assay, which revealed that alanine mutations at positions 66 and 84 had lower activity compared to wild-type (wt) RAF, but higher activity than the R89L null mutation (Figure 3.3a). The K84E and Q66K mutations, both of which dramatically change the charge at the protein surface, had activity similar to R89L, suggesting abrogation of the PPI. We cloned each mutant into RAF phage for activity-dependent plaque assays. As expected, based on the reporter assay, K84A and Q66A showed slightly smaller, "weaker" plaques than wt RAF, indicating weaker affinity for KRAS. In contrast, R89L phage produced no plaques, indicating no detectable PPI with KRAS, consistent with its known loss of interaction affinity (Figure 3.3b).[222] The K84E and Q66K phage generated extremely small, "diminished" plaques compared to the alanine mutants at those positions, indicating further reduction in affinity to KRAS. Although only qualitative, these plaque assay experiments suggest both the RNAP biosensors and phage selection are capable of detecting RAF variants with weak but measurable binding affinity and are capable of discriminating between the relative binding affinities.

We next aimed to evolve a non-interacting RAF variant to regain its KRAS-binding capabilities. To do this, we coupled the PACS selection cells (Figure 3.2a) with the

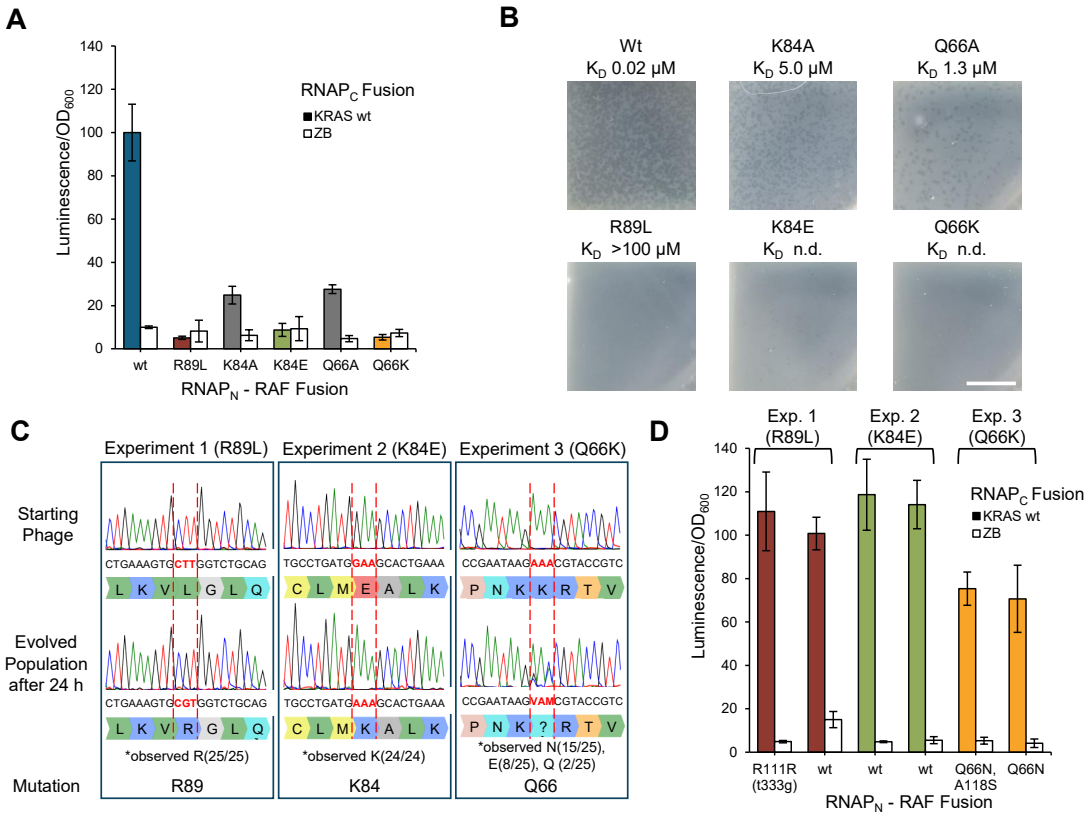

Figure 3.3: Validation of PACS PPI selection system by evolving inactive RAF variants to bind KRAS. (A) Luminescence signal for wild type (wt) RAF and RAF variants with known or suspected lower binding affinity to wt KRAS on the wt KRAS or off-target PPIs (error: std. dev., n = 4). The RAF mutants used for additional testing were color coded: red: R89L; green: K84E; yellow: Q66K. (B) Plaque assays of the different RAF variants on the PACS selection cells (scale bar: 1 cm). (C) Bulk Sanger sequencing of the phage populations before and after 24 h of PACE. The number of individual colonies from the evolved phage population after 24 h with the indicated amino acid at the mutated position are listed as "*observed X(#/#)." (D) Luminescence signal for two evolved RAF variants from each experiment on the wt KRAS or off-target PPIs (error: std. dev., n = 4). The two randomly selected RAF variants for experiment 2 had the same genotype. wt indicates reversion to wt nucleotide sequence, and (n#n) indicate nucleotide changes for silent mutations.

previously-reported *in vivo* mutagenesis plasmid MP6,[223] to develop a new PPI PACE system. We chose to initiate the experiment with phage encoding the R89L, Q66K, or K84E RAF variants, which started with either undetectable or diminished activity, to rescue wt-like binding phenotypes. We seeded three separate phage vessels, each with one population of either R89L, Q66K, or K84E RAF phage, then ran PACE for 24 h under KRAS PPI selection. At the end of the experiment, all three lagoons still contained robust

populations of phage, which we collected and Sanger-sequenced to assess the mutations in the bulk library population (Figure 3.3c).

The phage populations that began with R89L and K84E both reverted back to wt through evolution. Unsurprisingly for the R89L-starting population, only one codon was observed due to only one codon, “CTT,” requiring a single nucleotide base substitution to encode arginine, while the other five degenerate codons for arginine required two or three nucleotide base substitutions. Similarly, for the K84E-starting population, it was not surprising that only one codon was observed because of the two degenerate codons encoding lysine, only one, “GAA,” required a single nucleotide base substitution to revert back to lysine. In contrast, more than one codon was observed in the bulk sequencing library for the Q66K-starting population. This was surprising, considering that only a single nucleotide base substitution would change the mutant codon “AAA” to one of the two degenerate codons encoding for glutamine. Based upon the bulk sequencing data, there are six possible codons, encoding for six different amino acids, that could be present in the final population: K (the starting mutant amino acid), Q (the wt amino acid), N, H, D, and E. Of these possible amino acids, asparagine was the only polar amino acid, containing the same amide functional group and differing only in the length of the carbon chain in the R group, as compared with the wt glutamine. Also, we reasoned that either aspartic acid or glutamate would be present in the population due to the potential to form an ionic interaction with the proximal R41 and K42 residues on KRAS. We therefore expected that these would be some of the species observed upon further assessment of individual variants from the bulk population. Indeed, upon sequencing individual variants, Q66N, Q66E, and Q66Q variants were observed (Figure 3.3c).

We next subcloned individual variants from each of the populations to test KRAS binding in the luciferase reporter assay. As expected, the initial R89L and K84E populations contained mostly wt phage, with some additional silent mutations that performed similarly to wt in the reporter assay (Figure 3.3d). The Q66K phage population contained multiple

variants, including the Q66N mutation predicted from the bulk sequencing results. The Q66N variants had luciferase activity levels greater than the starting Q66K mutant, but slightly less than wt, indicating partial recovery of wt-like KRAS binding affinity. Taken together, these experiments confirm that the selection vectors can be used to either select for or evolve RAF variants that bind KRAS for either PACS or PACE, respectively. Next, we aimed to test whether the system could be used in a high-throughput DMS experiment starting with a phage-encoded RAF library.

3.2.4 Deep mutational scanning (DMS) of the KRAS/RAF binding interface

To demonstrate how our PACS system could easily be integrated into the DMS workflow,[199] we generated a library of RAF variants introducing mutations along the interface of the KRAS/RAF interaction based upon crystal structures of these proteins, focusing on amino acid residues 52-90. We generated the library utilizing error prone PCR (epPCR) mutagenesis methods,[224] with the level of mutagenesis tuned such that double nucleotide mutations were predominantly observed in the final library (Methods section 3.4.2, Table 3.1 and 3.2). To ensure that the generated library still contained active variants, a small subset of the library was cloned into the N-terminal luciferase assay plasmid (Figure 3.1a) and tested (Figure 3.4). The assayed variants showed varying degrees of activity compared to wt RAF, ranging from no activity, similar to the R89L null variant, to indiscernible from wt.

Next, we subjected three replicate populations of the RAF library to our PACS platform (Figure 3.6a) to enrich for RAF variants with the ability to bind KRAS. Phage library samples were collected at multiple time points throughout the course of the 72 h selection experiment. We cloned the collected phage samples from the end of the selection at 72 h into the N-terminal expression plasmid (Figure 3.1a) and tested them in the luciferase as-

	Sanger Sequencing (8 colonies)		HTS dataset (0 h)	
	Calculated	Estimated error	Calculated	Estimated error
Transitions/Transversion	1.3	0.1	1.2	0
AT→GC/GC→AT	2	0	2.8	0
A→N, T→N (%)	71.4	3.6	75.5	4.2
G→N,C→N (%)	28.6	3.6	24.5	4.2
AT→GC (%)	57.1	3.6	60.2	3.9
GC→AT (%)	28.6	3.6	21.8	0.3
Transitions Total (%)	57.1	3.6	53.7	2
A→G, T→C (%)	35.7	2.5	36.5	1.9
G→A, C→T (%)	21.4	2.5	17.2	0.3
Transversions Total (%)	42.9	2.5	46.3	2.4
A→T, T→A (%)	14.3	0	15.4	0.5
A→C, T→G (%)	21.4	2.5	23.7	3.4
G→C, C→G (%)	0	0	2.7	0
G→T, C→A (%)	7.1	2.5	4.5	0.1

Table 3.1: Error prone RAF library mutational biases. Nucleotide substitution biases in the error prone (ep) library calculated using the Mutanalyst program[225] for both Sanger sequencing data (8 colonies) and HTS data of the initial ep RAF phage library (0 h time point).

Number of Amino Acid Mutaitons	Predicted	Observed	Percent Coverage (%)
0	1	1	100.0
1	741	297	40.1
2	267501	5369	2.0
3	125368802	4282	0.0

Table 3.2: Error prone RAF library coverage. Number of predicted and observed RAF variants with unique single, double, and triple amino acid substitutions in the error prone RAF library determined from the HTS dataset (0 h time point).

say to confirm that the selection experiment had enriched for RAF variants with the ability to bind KRAS (Figure 3.5). The assayed RAF variants all had activity above that of the RAF R89L null variant, demonstrating that the selection had functioned as anticipated, i.e., active variants were enriched over the course of PACS.

We subjected the collected phage samples from all three replicate experiments to HTS analysis to perform DMS, with time-dependent information, of the mutated binding interface. Practically, only the region that was subjected to mutagenesis was generated into sequencing amplicons to allow for paired end reads, improving the overall quality of the data by reducing error associated with sequencing. In our data set, we were able to observe 1279 unique protein variants over the course of the experiment. Functional scores were calculated for each observed variant in the HTS dataset based upon the

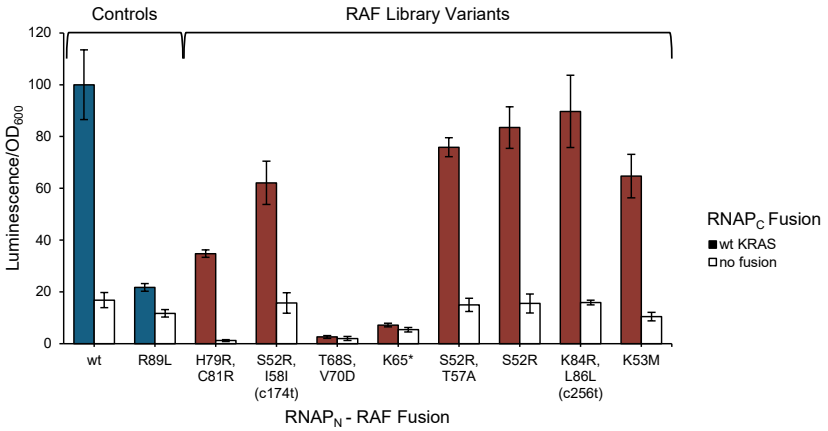

Figure 3.4: Activity of error prone RAF library variants. Luminescence signal for RAF error prone (ep) library variants cloned into the N-terminal expression plasmid tested in the *in vivo* luciferase assay (error: std. dev., n = 4). (n#n) indicates nucleotide changes for silent mutations.

observed counts and were plotted to show their distribution over time for the different variants (Figure 3.6b). Based upon this data, we determined that a minimum of 24 h was required to achieve stable functional scores for the different variants; earlier time points showed greater variability in the tabulated functional scores. Although the 72 h dataset overall performed best (most linearity between affinity and functional score), even the 24 h data was quite good, showcasing the rapid speed of data acquisition made possible by this technology.

To test if the calculated functional scores from PACS corresponded to the relative binding affinity, we compared literature-reported K_D values[222] for various RAF mutants observed in our HTS dataset against our calculated functional scores (Figure 3.6c). There is a clear correlation between the functional scores and the relative binding affinities of the different RAF mutants ($R^2 = 0.96$; $p = 0.0035$). We also verified that our *in vivo* luciferase assay could inform upon RAF properties, such as stability and expression, as well as approximate binding affinity, by plotting each variant's luminescence signal against the corresponding calculated functional score from the HTS datasets (Figure 3.6d). Taken together, these data confirm that PACS enriched active RAF variants capable of binding KRAS, resulting in datasets delineating relative fitness across a range of RAF mutants.

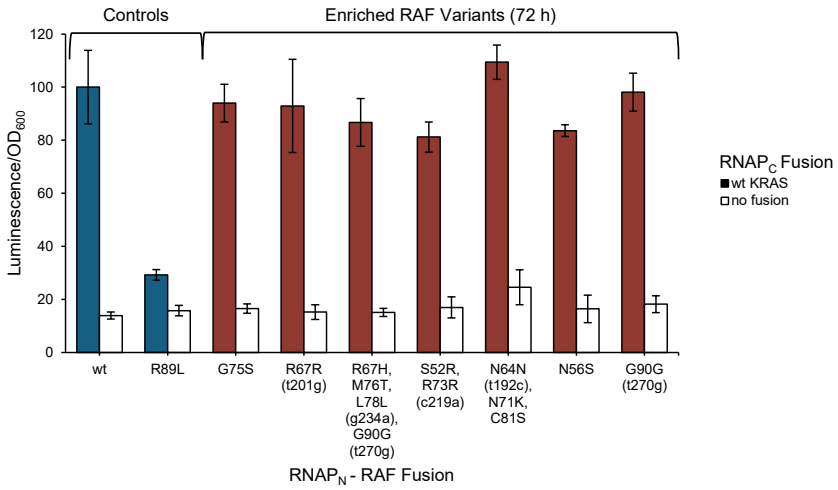

Figure 3.5: PACS selection validation. Luminescence signal for RAF variants collected at the 72 h time point from the PACS selection (red) and cloned into the N-terminal expression plasmid tested in the *in vivo* luciferase assay (error: std. dev., n $=$ 4). (n#n) indicates nucleotide changes for silent mutations.

We were also pleasantly surprised that the enrichment values correlated quite well and linearly to in vitro affinity measurements. This correlation indicates that our PACS system provides more nuanced information about relative affinities.

3.2.5 Analysis of the KRAS/RAF binding interface

From our DMS dataset, we were able to examine the relative functional consequence of different amino acid substitutions along the mutagenized region by calculating the enrichment values for mutants observed along the binding interface relative to wild type (wt) at the 72 h time point (Figure 3.7a). Many of the 39 amino acid positions explored in our mutant library favored positively charged side chains, with either the wt amino acid (aa) being a positively charged residue or positions enriching for positive residues (16 total positive aa: 9 positions with a wt positive aa; 10 positions enriching for a positive aa, 3 of these positions having a wt positive aa), and very few favored negatively charged side chains (3 total aa: 1 position with a wt negative aa, 2 positions enriching for a negative aa), which corresponds with the overall structure of RAF having a large number of ba-

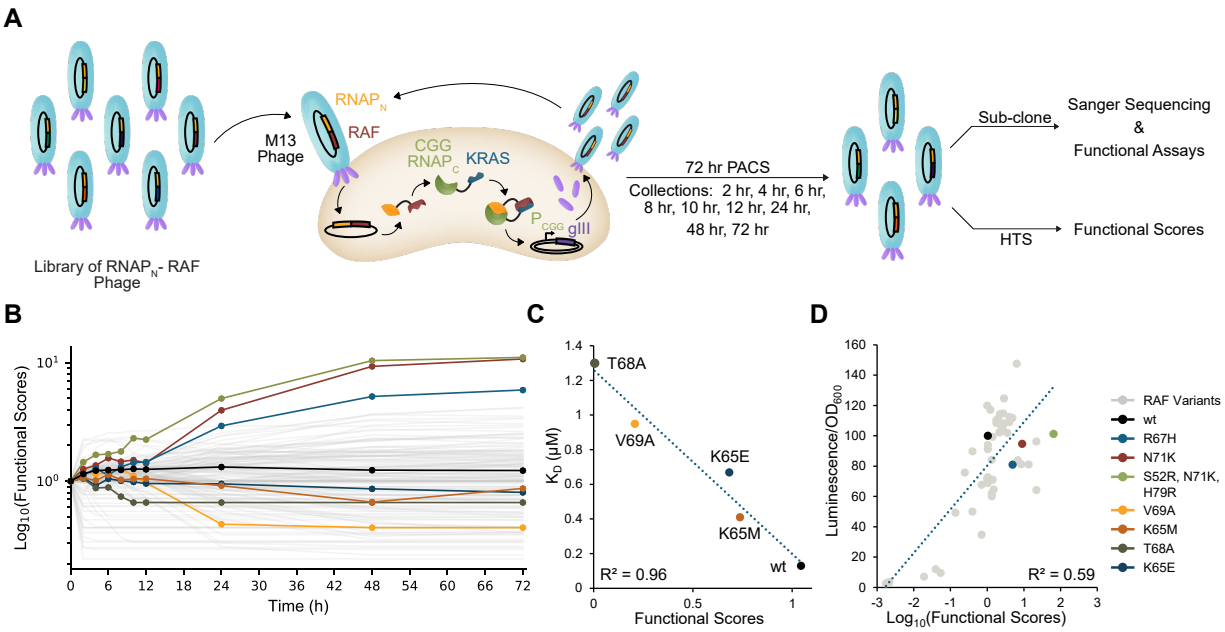

Figure 3.6: Deploying PACS system for DMS of RAF. (A) Schematic of the work-flow for the PACS DMS experiment for an error prone RAF mutant library. (B) Plot of time-dependent functional scores for all observed RAF variants in a single replicate lagoon from the PACS DMS experiment. The RAF variants in B-D are color-coded according to the key in D. (C) Plot of the correlation between literature reported K_D values and their corresponding calculated functional scores observed from a single replicate lagoon from the PACS DMS experiment ($R^2 = 0.96$, p value $= 0.0035$). (D) Plot of the correlation between *in vivo* luciferase assay values for tested RAF variants and their corresponding calculated functional scores observed from a single replicate lagoon in the PACS DMS experiment ($R^2 = 0.59$, p value < 0.0001).

sic residues (K and R) along the binding interface with KRAS.[226,227] Also, many of the positions along the binding interface showed no substitutions for which there was an increased enrichment relative to the wt aa across experimental replicates, including but not limited to: 59, 61, 62, 63, 65, 66, 69, 77, 82, 84, and 89. As expected, the amino acids at positions 89, 84, and 66 that are reported in the literature as being key in the binding interface of KRAS and RAF, showed the least tolerance to substitutions with only the wt sequence being favored and all other observed amino acids de-enriching at these positions.[220,222,226] For example, position 66 de-enriches for all positively charged residues (K, R, H) likely due to charge repulsion with the proximal residues R41 and K42 on KRAS.

Other positions that tolerated at least one substitution with similar aa properties

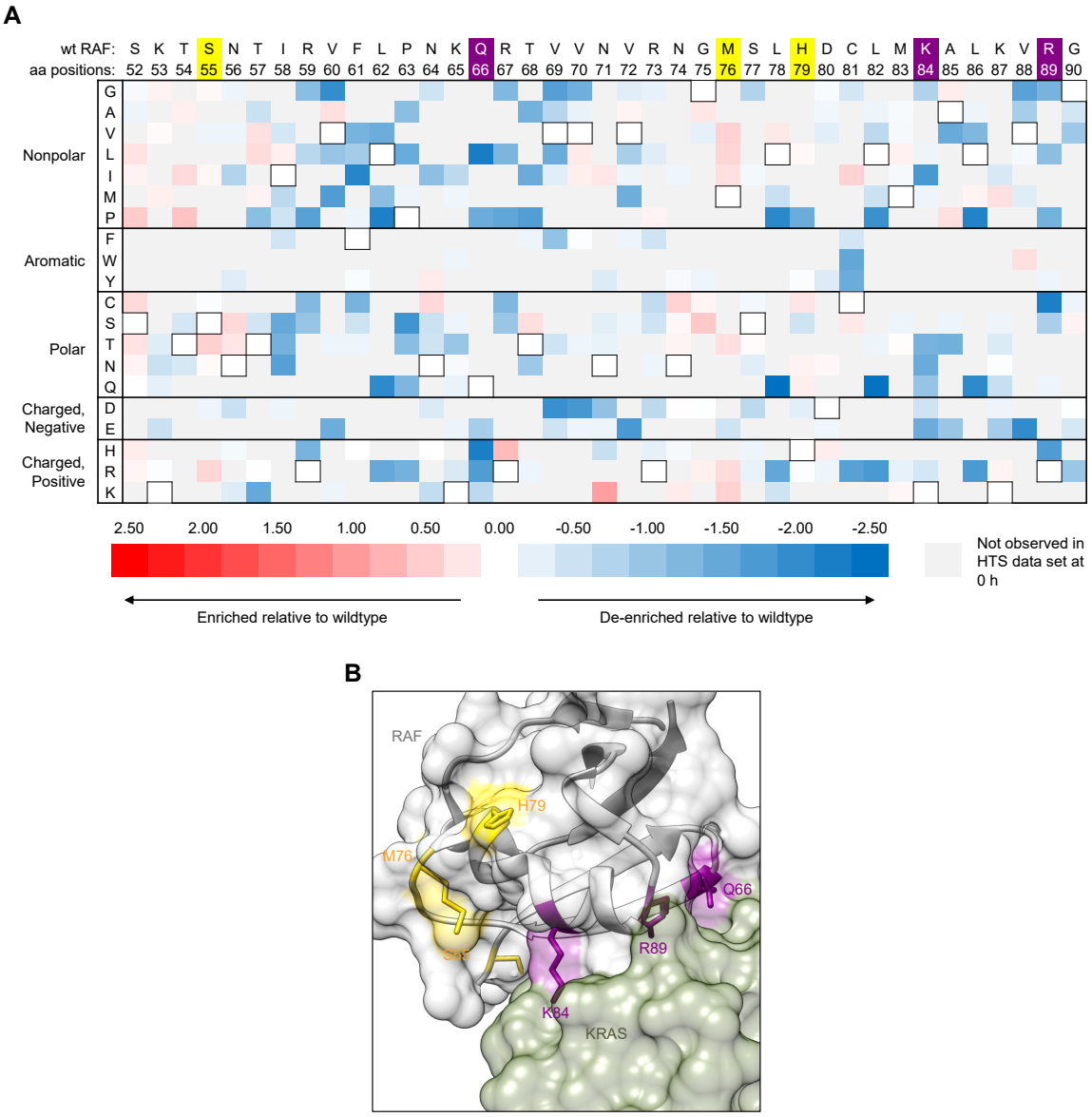

Figure 3.7: Analysis of DMS data for RAF. (A) Heat map of relative enrichment values at 72 h for each amino acid (aa) compared to the wild type (wt) aa. The wt aa is indicated with a thick border, and any aa not observed at 0 h in the HTS dataset is colored gray. (B) Crystal structure of the KRAS/RAF PPI (PDB: 4G0N). Positions 66, 84, and 89, which are key residues in the KRAS/RAF PPI, showed intolerance to aa substitutions in the HTS dataset, and are indicated in purple. Positions 55, 76, and 79 showed a high degree of aa substitution tolerance in the HTS data, and are indicated in yellow.

groups to wt aa included: 56, 58, 60, 67, 68, 70, 72, 78, 85, and 86. Residues 56 and 68 favored polar residues, similar to the wt aa asparagine or threonine, respectively, and are directly involved in stabilizing the β-sheet structure. Residue 68 was unique in that it only favored the polar residues with alcohol moieties on the R groups (wt S, or T) most likely due to the formation of hydrogen bonds with the proximal residue D38 on KRAS. Positions 58, 60, 70, 72, and 86 all favored non-polar residues; only favoring a subset of the residues of similar sizes, which appears to be key in the formation of the hydrophobic core of RAF. For example, position 70 only seemed to favor nonpolar aliphatic residues, but not small nonpolar side chains like glycine or alanine, possibly due to the need to reduce flexibility in the β-sheet structure and increase the overall surface area of binding between KRAS and RAF. Position 85 was unique in that it enriched for both glycine and proline, but not other non-polar residues. It is interesting that these two particular amino acids were enriched because they have very different effects on structure; glycine is highly flexible while proline is the most rigid in terms of conformation. Residue 85 is located in a turn of an α-helix requiring the backbone to change direction by either introducing a proline to introduce such a change, or the flexible glycine to allow for such a change.

There were a few aa positions that tolerated aa substitutions with properties that differed from the wt aa: 53, 54, 57, 64, 73, 74, 80, 88, and 90. For example, position 53 enriched for the nonpolar amino acids glycine, valine, and methionine, but in the wt protein it is lysine. Other aa positions tolerated a diverse set of aa substitutions, in particular positions 52, 55, 71, 75, 76, 79, 81, 83, 87. For example, aa position 71 tolerated nonpolar and positively charged residues. In particular, lysine was the most enriched aa at this position possibly due to the formation of an ionic interaction with the proximal residue D66 on KRAS. Positions 52, 55, 76, and 79 all enriched for multiple types of amino acids, including those from nonpolar, polar, and charged categories. Position 76 was unique in that it had the greatest number of aa substitutions among all the replicates that were tolerated without a clear categorical preference: valine, leucine, isoleucine, threonine, lysine,

and arginine. It is interesting to note that this particular residue, as well as residue 79, is located on the loop region between an α-helix and a β-sheet, possibly accounting for the varied amino acids tolerated at these positions (Figure 3.7b).

3.3 Discussion

While not every amino acid at each position was sampled, potentially due to either bias present in error prone mutagenesis methods,[224,228] or from the limited sequencing coverage of each time point examined over the course of the experiment, we did observe trends associated with certain positions. Known key residues in the KRAS/RAF interaction showed intolerance to amino acid substitutions, while residues not directly involved in the interface, located in loop regions between an α-helix and a β-sheet, had more variability in the amino acids tolerated at those positions. Deeper sequencing or improved mutagenesis methods[229] would likely improve the depth of the datasets and number of unique variants assayed. Nevertheless, these experiments confirm the new PACS platform is able to rapidly select for RAF variants capable of biding KRAS, recapitulating known invariable positions identified in RAF. We were also able to show how the functional scores calculated correlated to binding affinity of different RAF variants based on reported K_D values[222] and observed luciferase activity *in vivo*, facilitating rapid characterization of variants obtained through the selection experiment.

Due to the plug and play nature of the RNAP-based biosensor, other PPIs can be similarly examined in a high-throughput manner. Previous demonstrations using the RNAP-based biosensor to study PPIs of the BCL2 family proteins and BH3 only proteins, show that it is possible to swap in different protein domains and observe interactions under different conditions.[217] The only requirements for a given target are that the proteins are expressible in *E. coli* and can tolerate fusion to the split RNAP biosensor. Notably, this latter limitation has recently been mitigated through the evolution of the C-terminal half of

the RNAP biosensor to tolerate C-terminal fusions.[230] This method could also be utilized to facilitate the interrogation of very similar proteins, such as members of a protein family, to determine differences in highly similar protein variants and to identify key residues and properties that are unique to individual members. Another key benefit of this method is the ease with which it can be converted into a directed evolution campaign by the addition of *in vivo* mutagenesis to convert the PACS system into PACE. Because of this feature, not only can information about the binding interface be determined, but a *de novo* protein binder could also be evolved in response to given selection parameters. More broadly, this and other methods under development[231,232] will continue to apply evolutionary and high-throughput approaches to interrogate biomolecular interactions, moving away from traditional reductionist approaches toward a more comprehensive understanding of the sequence-function relationships in protein machines.

3.4 Methods

3.4.1 Cloning

All plasmids used in this study were generated using Gibson assembly of PCR products generated with NEB Q5 high fidelity polymerase. All constructs were validated via Sanger sequencing performed by the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping facility. All constructs are listed in tables B.3, B.4, B.5, B.6, and B.7 with links to online vector maps. The wt KRAS4b gene was obtained from the Frederick National Laboratory for Cancer Research.

3.4.2 Optimization of $MnCl_2$ concentration and cycle number for the RAF error prone library

To optimize the concentration of $MnCl_2$, a surrogate *E. coli* plasmid containing the identical gene construct $RNAP_N$ C-terminally fused to RAF (16-77) was used in place of the 14-70 phage vector. This was done to have a direct link between the number of variants observed and the number of transformations, which is not possible using the phage vector and phage cloning due to the replication of genotypes that would arise from phage outgrowth.

A small suite of $MnCl_2$ concentrations were explored to tune the level of mutagenesis in the resulting gene to be approximately 2 nucleotide substitutions per gene. 50 µL error prone PCR reactions were set up with the following final concentrations: 1 ng/µL template plasmid, 0.2 µM forward primer, 0.2 µM reverse primer, 1x standard NEB Taq buffer, 0.2 mM dNTPs, varrying amounts of $MnCl_2$, and 0.025 U/µL NEB Taq DNA polymerase. The following $MnCl_2$ concentrations were tested in the error prone PCR reaction: 100 µM, 200 µM, 300 µM, 400 µM, 500 µM, and 750 µM. The thermocycler conditions utilized an initial denaturation step (95 °C, 30 sec) followed by 20 cycles of denaturation (95 °C, 30 sec), annealing (58 °C, 30 sec) and extension (72 °C, 1 min), and a final extension (72 °C, 2 min).

The error prone PCR product was then treated with 1 µL NEB DpnI enzyme in the PCR buffer for 1 hour, and purified using a Zymo DNA clean and concentrator kit. The insert was then Gibson assembled into the PCR amplified 16-77 vector backbone, transformed into chemicompetent 10βcells, and plated onto agar plates (15 g/L in LB) containing antibiotics. 8 colonies from each of the tested $MnCl_2$ conditions were picked and subjected to 25 µL colony PCR reactions to amplify up the relevant gene region for Sanger sequencing. Briefly, 25 µL colony PCR reactions were set up with the following final concentrations: 1 colony/25 µL, 0.2 µM forward primer, 0.2 µM reverse primer, 1x standard NEB Taq buffer,

0.2 mM dNTPs, 0.0125 U/µL NEB Taq DNA polymerase. The thermocycler conditions utilized an initial denaturation step (95 °C, 3 min) followed by 25 cycles of denaturation (95 °C, 45 sec), annealing (60 °C, 15 sec) and extension (68 °C, 50 sec), and a final extension (68 °C, 2 min). Figure 3.8A shows the number of nucleotide mutations observed for each colony that had the correct sized gene product as determined by a DNA agarose gel of the colony PCR product.

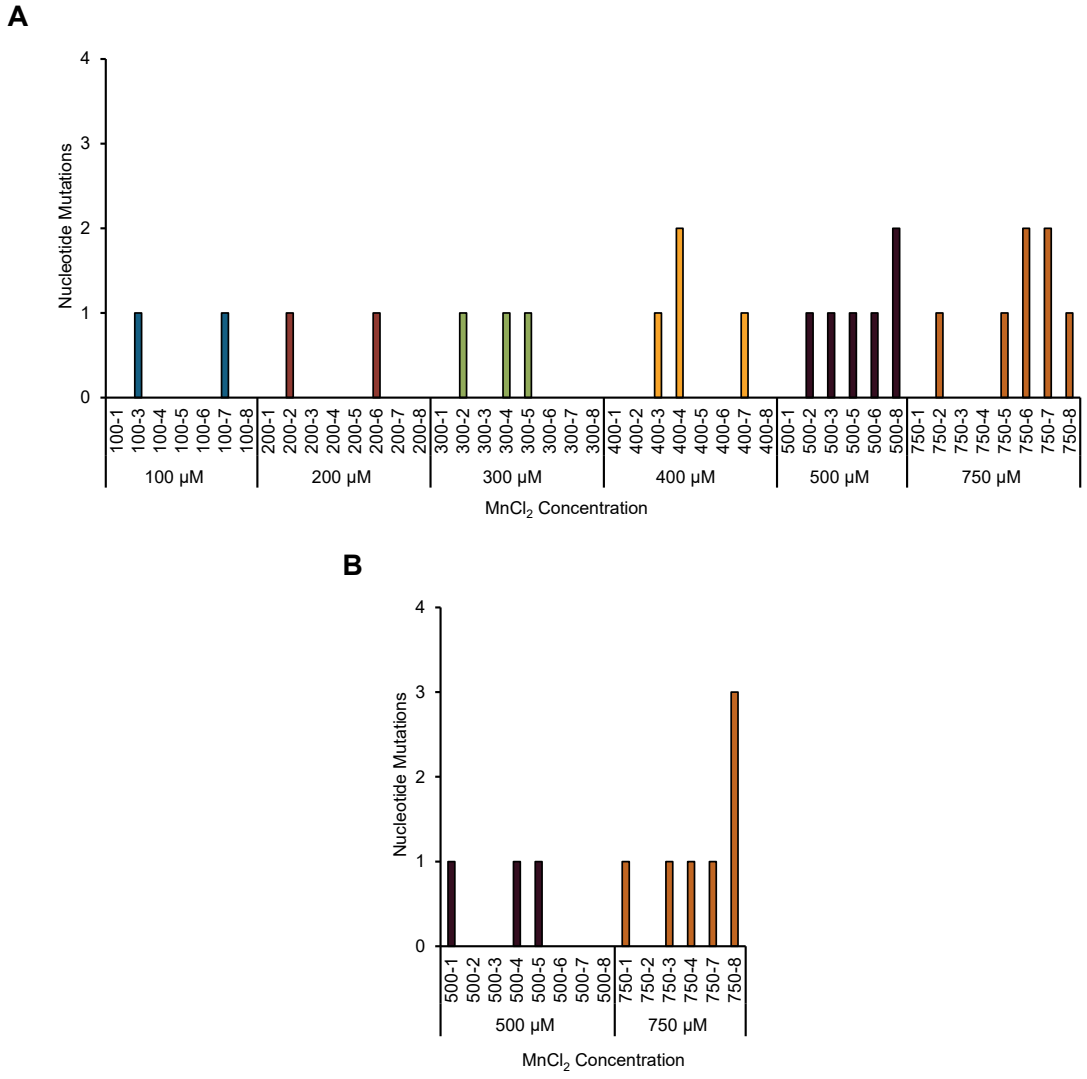

Figure 3.8: Optimization of error prone PCR conditions. (A) Number of nucleotide mutations observed in Sanger sequencing for each colony generated using inserts made with 20 cycles of error prone PCR with the indicated $MnCl_2$ concentrations. (B) Number of nucleotide mutations observed in Sanger sequencing for each colony generated using inserts made with 30 cycles of error prone PCR with the indicated $MnCl_2$ concentrations.

Using the two highest $MnCl_2$ concentrations, the number of PCR cycles was optimized by following the error prone PCR conditions outlined above and changing the number of cycles for the repeat denature, annealing, and extension step from 20 cycles to 30 cycles. Again, the error prone PCR product was then treated with 1 µL NEB DpnI enzyme in the PCR buffer for 1 hour, and purified using a Zymo DNA clean and concentrator kit.

The insert was then Gibson assembled into PCR amplified 16-77 vector backbone, transformed into chemicompetent 10βcells, and plated onto agar plates (15 g/L in LB) containing antibiotics. 8 colonies from each of the tested $MnCl_2$ conditions were picked and subjected to 25 µL colony PCR reactions using the colony PCR recipe and thermocycler conditions outlined above to amplify up the relevant gene region for Sanger sequencing. Figure 3.8B shows the number of nucleotide mutations observed for each colony that had the correct sized gene product as determined by a DNA agarose gel of the colony PCR products.

Based on these tests, the concentration of $MnCl_2$ used for the RAF library generation was 750 µM because it gave the highest fraction of mutations observed for all of the colonies tested, and the highest number of mutations for each transcript. 30 cycles of error prone PCR were chosen to increase the likelihood of observing double mutations.

3.4.3 RAF phage library generation

The insert was prepared by introducing mutations into the RAF gene using error prone PCR (epPCR).[224] Briefly, 50 µL epPCR reactions were set up with the following final concentrations: 1 ng/50 µL 14-70 template plasmid, 0.2 µM forward primer, 0.2 µM reverse primer, 1x standard NEB Taq buffer, 0.2 mM dNTPs, 750 µM $MnCl_2$, and 0.025 U/µL NEB Taq polymerase. The thermocycler conditions utilized an initial denaturation step (95 °C, 30 sec) followed by 30 cycles of denaturation 95 °C, 30 sec), annealing (58 °C, 30 sec) and extension (72 °C, 1 min), and a final extension (72 °C, 2 min). The PCR product was digested with 1 µL NEB DpnI enzyme in the PCR reaction buffer for 1 h at 37 °C, and then

purified using a Zymo clean and concentrator kit.

The $RNAP_N$-fused wt RAF phage vector (14-70) backbone was amplified overnight in 1059 cells,[171] which replicate phage lacking gIII, and purified using a Qiagene plasmid purification kit.

Inserts were subcloned into the 14-70 vector using the XhoI and PstI sites with NEB XhoI enzyme and NEB PstI-HF enzyme following the standard NEB double digest protocol. Briefly, 30 µL restriction digestion reactions were set up for both the insert and backbone with the following final concentration for each component: 1 µg/30 µL insert or backbone DNA, 20 U/30 µL NEB XhoI enzyme, 20 U/30 µL NEB PstI-HF enzyme, and 1x NEB cut smart buffer. These were kept at 37 °C overnight, run on an agarose gel, gel extracted, and purified using a Qiagene gel purification kit.

The T4-mediated ligation reactions were performed overnight at 16 °C using NEB T4 DNA ligase following the standard NEB ligation protocol to generate roughly 1 µg of the final RAF DNA library. Briefly, 100 µL ligation reactions were set up with the following final concentration for each component: 1 µg/100 µL restriction digested backbone, 100 ng/100 µL restriction digested insert, 5 U/100 µL NEB T4 DNA ligase, and 1x NEB T4 DNA ligase buffer. After overnight ligation, these were purified with a Zymo DNA clean and concentrator kit. All resulting plasmids were suspended in 30 µL water.

The phage library was prepared by electroporating (1800 V, 5 msec, 0.2 cm cuvettes, 50 µL cells, 143 ng DNA, 7 transformations total) 1 µg ligated vector product into 350 µL electrocompetent 1059 cells,[171] which were recovered in 25 mL of SOC media for 1 h. Afterwards, the recovered cells were used to inoculate 250 mL super broth to allow for phage library growth overnight. The phage containing supernatant was collected by centrifugation and sterile filtration to generate a RAF phage library.

3.4.4 Plaque assays

To quantify phage titers, 1059 activity-independent cells[171] were used; while S1030 cells[171] transformed with the positive selection plasmid (31-69) and the negative selection plasmid (20-6) were used to quantify the activity of different phage containing $RNAP_N$ protein fusions on KRAS. For all experiments, *E. coli* were grown to an optical density (OD_{600}) of approximately 0.6.

Phage dilutions (1/50, 1/12500, 1/125000, 1/6250000) were made by serially diluting a phage sample in water. Plaque assays were performed by mixing 50 µL of each phage dilution with 50 µL of cells, suspending these solutions in 55 °C warmed soft agar (7 g agar/1 L LB), and plating this mixture onto quartered solid agar plates without antibiotics (15 g agar/1 L LB). These were then grown in a 37 °C incubator overnight, and phage plaques were counted and imaged to determine the relative plaque forming units (PFU).

3.4.5 Luciferase assays

Three separate vectors were constructed: 1) a N-terminal expression plasmid containing the $RNAP_N$ fragment C-terminally fused to either RAF or no protein partner upstream of the UV5 IPTG inducible promoter,[233] 2) a C-terminal expression plasmid containing the T7 $RNAP_C$ fragment N terminally-fused to either KRAS or an off target protein partner upstream of the constitutive P_{Kat} promoter, and 3) a luciferase reporter plasmid containing the bacterial luxAB gene[219] upstream of the T7 RNAP promoter. The N-terminal, C terminal, and the luciferase reporter plasmids were electroporated into S1030 cells,[171] and plated onto agar (15 g/L in LB) plates containing 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, and 7.5 µg/mL tetracycline. Individual colonies were then picked and grown overnight to saturation with shaking at 37 °C in 500 µL of 2xYT media containing 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, 7.5 µg/mL tetracycline in a 96, deep-well plate. A new

culture was then prepared using 60 µL of the overnight culture to inoculate 540 µL of fresh 2xYT media containing 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, 7.5 µg/mL tetracycline and 5 µM IPTG. After 3 h of growth at 37 °C with shaking, a 150 µL aliquot was transferred to a black-walled, clear-bottomed, 96-well plate and read on a Biotek Synergy Neo2 plate reader, monitoring both the optical density at 600 nm (OD_{600}) and the luminescence signal.

Luciferase activity was reported as normalized luminescence signal/OD_{600} for each well. Values are reported as the arithmetic average of replicate wells ($n = 4$) with error reported as the standard deviation of the replicates. In order to compare different days, the normalized luminescence signal/OD_{600} for a given sample was divided by the normalized luminescence signal for the wt KRAS/RAF interaction partners, which was given an arbitrary value of 100.

3.4.6 PACS & PACE

PACS was performed utilizing a modified version of the previously described PACE method.[215] S1030 cells were transformed by electroporation with the positive selection plasmid (31-69) and the negative selection plasmid (20-6). A 5 mL starter culture was grown overnight to saturation in LB media supplemented with 50 µg/mL carbenicillin, 40 µg/mL kanamycin, and 7.5 µg/mL tetracycline. Chemostats (250 mL sterile bottles) containing 150 mL of Davis rich medium[215] were inoculated with 5 mL starter culture and grown at 37 °C with magnetic stir-bar agitation. When the OD_{600} reached ~0.6, fresh Davis rich medium was pumped in at 120-150 mL per h, with a waste needle set at 150 mL. Selection phage containing $RNAP_N$ protein fusions were used to seed fresh lagoons (25 mL flask with a rubber septum) in triplicate. For the mock PACS experiment, 40 µL of a 1000:1 mixture of $RNAP_N$ – iZA to $RNAP_N$ – RAF phage were used to inoculate the lagoons; for the PACS DMS experiment, 1 mL of the $RNAP_N$ – RAF error prone phage library was used to inoculate the lagoons. Waste needles were set to maintain the lagoon volume at

20-25 mL, and host cell cultures were flowed in at 20-25 mL per h. Lagoon samples were taken from the waste withdrawal line, centrifuged, sterile filtered, and the supernatant was stored at 4 °C. For the mock PACS experiment, lagoon samples were collected at 2 h, 4 h, 6 h, 8 h, 10 h, 12 h, and 24 h. For the DMS PACS experiment, lagoon samples were collected at 2 h, 4 h, 6 h, 8 h, 10 h, 12 h, 24 h, 48 h, and 72 h.

PACE was performed in a similar fashion to PACS experiments, with some differences. S1030 cells were transformed by electroporation with the positive selection plasmid (31-69), the negative selection plasmid (20-6), and the additional *in vivo* mutagenesis plasmid MP6.[223] A 5 mL starter culture was grown to saturation overnight in LB media supplemented with 50 μg/mL carbenicillin, 33 μg/mL chloramphenicol, 40 μg/mL kanamycin, 7.5 μg/mL tetracycline, and 20 mM glucose. For the PACE reversion experiment, 40 μL of single clonal phage populations of each $RNAP_N$ – RAF R89L, K84E, or Q66K phage were used to inoculate sperate fresh lagoons. Mutagenesis was initiated by the direct addition of arabinose (20% w/v in water) into each lagoon at a rate of 0.5 mL per h. Lagoon samples were taken at 24 h.

3.4.7 Preparing PACS library amplicons for HTS

Amplicons for HTS on the Illumina MiSeq platform were prepared by amplifying the gene of interest from the phage supernatant of the initial RAF phage library and each time point collected during the PACS DMS experiment using PCR. For the first PCR (PCR 1), additional diversity was added with a variable sequence of 6-9 nucleotides, and regions complementary to the i7 and i5 Illumina TruSeq adaptors were added using the indicated primers below. Briefly, 50 μL PCR reactions set up with the following final concentrations: 1 μL phage supernatant, 0.5 μM forward primer, 0.5 μM reverse primer, 1x standard NEB Q5 HF buffer, 0.2 mM dNTPs, 0.02 U/μL NEB Q5 HF polymerase. The thermocycler conditions utilized an initial denaturation step (98 °C, 1 min) followed by 5 cycles of denaturation (98 °C, 10 sec), annealing (60 °C, 15 sec) and extension (72 °C, 10 sec), and

a final extension (72 °C, 2 min). The PCR product was purified using a Zymo clean and concentrator kit, and eluted in 20 µL of water.

Primers for PCR 1:

The italic sequence is the region complementary to the immediate upstream and downstream region of the mutagenized RAF gene. The bold region is the random sequence of 6 (pictured) to 9 nucleotides. The underlined region in the forward and reverse primers are the complementary regions for the Illumina TruSeq i7 adaptor and the i5 adaptor, respectively.

Forward primer:

5' – GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT – **NNNNNN** – *TCTG GCTCTGGCTCGAGC* – 3'

Reverse primer:

5' – ACACTCTTTCCCTACACGACGCTCTTCCGATCT – **NNNNNN** – *ACAACA CTCCGGCTGCAG* – 3'

The i7 and i5 Illumina TruSeq adaptors were added using PCR to give each time point from each replicate lagoon a unique i7/i5 barcode combination to allow for sample multiplexing. Briefly, 50 µL PCR reactions set up with the following final concentrations: 1.25 µL purified PCR 1 product, 0.5 µM forward primer, 0.5 µM reverse primer, 1x standard NEB Q5 HF buffer, 0.2 mM dNTPs, 0.02 U/µM NEB Q5 HF polymerase. The thermocycler conditions utilized an initial denaturation step (98 °C, 1 min) followed by 10 cycles of denaturation (98 °C, 10 sec), annealing (58 °C, 15 sec) and extension (72 °C, 10 sec), and a final extension (72 °C, 2 min). The PCR product was visualized using DNA agarose gel to verify the generation of the correctly sized product, and was purified using a Zymo clean and concentrator kit, and eluted in 20 µL of water.

The dsDNA concentration was determined using a QuBit dsDNA high sensitivity (HS) assay kit (Invitrogen, Ref Q32854). The purified PCR products were then pooled into a single DNA library by combining equal amounts of each PCR product into a single sample.

The combined library was then purified using a Zymo clean and concentrator kit to remove additional salts as measured by a nanodrop (260/230 ratio: 2.0; 260/280 ratio: 1.8), then the dsDNA concentration was determined using the QuBit dsDNA HS assay kit (Invitrogen, Ref Q32854). The library was then diluted to a final concentration of 4 nM with water, which was stored in $-20\,^{\circ}$C.

The amplicon library was prepared the day-of for HTS on the Illumina MiSeq platform using the Illumina MiSeq v2 reagent kit (300 cycles, MS-102-2002) and the associated "MiSeq: Denature and Dilute Libraries Guide" from Illumina. The procedures outlining how to prepare a final library concentration of 10 pM for the Illumina MiSeq V2 kit with a 25% PhiX spike in control were used. This mixture was kept on ice until run on the Illumina MiSeq instrument following the procedures outlined in the "Sequencing Analysis Viewer v2.4" guide for paired end reads.

3.4.8 HTS data processing

Initial data processing was performed on the Illumina MiSeq instrument using the Sequencing Analysis Viewer v2.4 software to demultiplex pooled samples based on their i7 and i5 indexes.

The demultiplexed fastq.gz files were downloaded, unzipped, and subjected to a standard data processing pipeline using a Python script (available upon reasonable request). Briefly, paired end reads were merged using the FLASH program[234] using the default parameters. Afterwards, all reads were trimmed to the open reading frame of the RAF protein sequenced (aa 52-90). Next the length of the trimmed reads and quality scores were determined. Those below a length of 100 nucleotides (trimmed sequence length is 117) or with a greater than 5% chance of not containing a sequencing error were removed from the data set. The nucleotide sequences were then translated into amino acid sequences, and each amino acid sequence with their total observed count was stored as our final data set used for further analysis.

3.4.9 Functional score calculation

The final data set was used for functional score calculation, which were calculated by first determining the frequency of each sequence for a given time point. To avoid issues with future data processing of variants that were unobserved in the initial library, sequence counts for each variant *x* at time *t* had an arbitrary value of 1 added.

$$Frequency\ of\ variant\ x\ at\ time\ t\ = \frac{(sequence\ counts\ for\ variant\ x\ at\ time\ t) + 1}{total\ sequence\ counts\ at\ time\ t}$$

Functional scores were then determined by dividing the frequency of variant *x* at time *t*, by the frequency of variant *x* at time 0 (the observed frequency in the initial library).

$$Functional\ score\ of\ variant\ x\ at\ time\ t\ = \frac{frequency\ of\ variant\ x\ at\ time\ t}{frequency\ of\ variant\ x\ at\ time\ 0}$$

3.4.10 Error for calculated functional scores

The error was determined by assuming the observed counts for each variant followed a Poisson distribution where the variance for a given measurement is equal to the mean. Therefore, the variance for the observed counts was equal to the observed counts, and the standard deviation was equal to the square root of the observed counts.

$$\sigma^2 = \mu$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\mu}$$

The error for the calculated functional scores was determined using the generic equation for error propagation for multiplication and division, where *a*, *b*, *c*, and *d* represent the counts used in the calculation of a functional score, depicted below. The 95th confidence interval or the 95th percentile of the error was used for each calculated functional score, which was obtained by multiplying the error by 1.96.

$$Q = \frac{ab}{cd}$$

$$\frac{\delta Q}{|Q|} = \sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \left(\frac{\delta c}{c}\right)^2 + \left(\frac{\delta d}{d}\right)^2}$$

$$\delta Q = |Q|\sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \left(\frac{\delta c}{c}\right)^2 + \left(\frac{\delta d}{d}\right)^2}$$

$$\delta Q = |Q|\sqrt{\left(\frac{\sqrt{a}}{a}\right)^2 + \left(\frac{\sqrt{b}}{b}\right)^2 + \left(\frac{\sqrt{c}}{c}\right)^2 + \left(\frac{\sqrt{d}}{d}\right)^2}$$

$$\delta Q = |Q|\sqrt{\left(\frac{1}{\sqrt{a}}\right)^2 + \left(\frac{1}{\sqrt{b}}\right)^2 + \left(\frac{1}{\sqrt{c}}\right)^2 + \left(\frac{1}{\sqrt{d}}\right)^2}$$

$$\delta Q = |Q|\sqrt{\left(\frac{1}{a}\right) + \left(\frac{1}{b}\right) + \left(\frac{1}{c}\right) + \left(\frac{1}{d}\right)}$$

It is important to note that the calculated error and observed sequence counts are inversely related, meaning that the fewer observed sequence counts results in larger error due to the increased level of uncertainty associated with lower sequence counts. Put more simply, as the number of counts approaches the stochastic noise of the instrument, it becomes more challenging to actually quantify it presence in the population above the noise.

3.4.11 Generation of global functional score plots

Global plots of functional scores over time were generated using a Python script. All sequences that appeared at any time over the course of the selection were given an arbitrary functional score of 1 at time 0 h to account for variants that appeared at later time points but were unobserved at the initial time point. Frequencies of each variant were

calculated using the modified equation below. Any sequence below a calculated frequency of 0.001 was given a value of 0.001; this value was used because it was determined to be the minimum threshold frequency above the stochastic noise observed in the dataset. Functional scores were calculated as above, and plotted as the $\log_{10}$ of the functional score.

$$Modified\ frequency\ of\ variant\ x\ at\ time\ t\ = \frac{sequence\ counts\ for\ variant\ x\ at\ time\ t}{total\ counts\ at\ time\ t}$$

3.4.12 Generation of amino acid distribution heat maps

Amino acid distributions at each position for a given time point were determined using a Python script. An additional data processing step was performed on the final data set to remove any truncated proteins that had a premature stop codon; this was done by removing any sequences without an exact sequence length of 39 amino acids. The fraction of each amino acid present at each position was determined for each time point; only sequences containing single amino acid changes were used in this analysis. These values were exported into excel where heat maps where generated.

The relative enrichment values for a particular amino acid *z* at a given position were calculated using the 72 h time point from the selection. This was done by taking the ratio of the percentage of amino acid *z* at 72 h over the percentage of amino acid *z* at 0 h in reference to the ratio of the wt amino acid (see equation below). In order to account for species that de-enriched below the detection limit of the HTS dataset at 72 h, an arbitrary percentage equivalent to 1 observation amongst the total sequence counts was used for all amino acids that were observed in the 0 h dataset. The relative enrichment values were utilized to construct color coded heat maps showing the functional consequence of these amino acid substitutions. Any amino acid that wasn't observed in the data set at 0 h was colored gray.

$$Relative\ enrichment\ value\ = \log_{10}\left(\frac{\%\ aa\ at\ position\ x\ at\ 72h}{\%\ aa\ at\ position\ x\ at\ 0\ h}\right) - \log_{10}\left(\frac{\%\ wt\ aa\ at\ position\ x\ at\ 72\ h}{\%\ wt\ aa\ at\ position\ x\ at\ 0}\right)$$

3.4.13 Error for calculated relative enrichment scores

The error for relative enrichment values or the log of functional scores was determined by deriving the error equation from the relative enrichment equation (depicted below), utilizing the same assumptions about the Poisson distribution as used for the calculated functional scores. The 95th confidence interval of the error was used for each calculated functional score, which was obtained by multiplying the error by 1.96.

$$E = \log_{10}\left(\frac{a/b}{c/d}\right) = \log_{10} a - \log_{10} b - \log_{10} c + \log_{10} d$$

$$\frac{\partial E}{\partial a} = \frac{1}{a \ln(10)} = \frac{k}{a}; k = \frac{1}{\ln(10)}$$

$$\delta E = \sqrt{\left(\frac{\partial E}{\partial a}\delta a\right)^2 + \left(\frac{\partial E}{\partial b}\delta b\right)^2 + \left(\frac{\partial E}{\partial c}\delta c\right)^2 + \left(\frac{\partial E}{\partial d}\delta d\right)^2}$$

$$\delta E = \sqrt{\left(\frac{k}{a}\delta a\right)^2 + \left(\frac{-k}{b}\delta b\right)^2 + \left(\frac{-k}{c}\delta c\right)^2 + \left(\frac{k}{d}\delta d\right)^2}$$

$$\delta E = \frac{1}{\ln(10)}\sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \left(\frac{\delta c}{c}\right)^2 + \left(\frac{\delta d}{d}\right)^2}$$

$$\delta E = \frac{1}{\ln(10)}\sqrt{\left(\frac{\sqrt{a}}{a}\right)^2 + \left(\frac{\sqrt{b}}{b}\right)^2 + \left(\frac{\sqrt{c}}{c}\right)^2 + \left(\frac{\sqrt{d}}{d}\right)^2}$$

$$\delta E = \frac{1}{\ln(10)} \sqrt{\left(\frac{1}{\sqrt{a}}\right)^2 + \left(\frac{1}{\sqrt{b}}\right)^2 + \left(\frac{1}{\sqrt{c}}\right)^2 + \left(\frac{1}{\sqrt{d}}\right)^2}$$

$$\delta E = \frac{1}{\ln(10)} \sqrt{\left(\frac{1}{a}\right) + \left(\frac{1}{b}\right) + \left(\frac{1}{c}\right) + \left(\frac{1}{d}\right)}$$

CHAPTER 4

SUMMARY AND PERSPECTIVES

PPIs are critical for a diverse array of cellular processes, and dysregulation of PPIs are often key drivers of disease. Due to their critical nature it is important to study PPIs. One tool that has been developed for this purpose is biosensors. These tools possess the ability to transduce input biochemical interactions into detectable outputs, and have been composed of array of different biological scaffolds and yield an equally diverse set of outputs. These are unique in that some of the available outputs include genetic or protein outputs, facilitating altering underlying biology.

Biosensors available for PPIs have a diverse arrange of outputs including light, proximity labeling, genetic, and protein outputs. Of these, RET- and split fluorescent protein- and split luciferase enzymes-, and split ligase and split peroxidase enzyme-based biosensors only produce light (fluorescence, luminescence) or proximity labeling (cleavage of reactive small molecule), limiting their applicability in PPI detection. Split proteases (e.g. split TEV), engineered cell surface receptors, and n-hybrid based-biosensors can directly integrate into host cell processes by linking biochemical signals and events into either protein outputs or genetic outputs. These biosensors have readily been utilized for a diverse set of applications in synthetic biology, including the generation of genetic circuits and biosynthetic pathways, therapeutics and diagnostics, and sensors for industrial applications. However, many of these sensors with the capability to impact biology, suffer from requiring extensive re-optimization for each new PPI or for each new application.

In my thesis work, I addressed this limitation by designing a versatile new “plug-and-play” biosensor based upon the T7 RNA polymerase (T7 RNAP). The T7 RNAP has long been used as a robust transcriptional system in a variety of *in vitro* and *in vivo* molecular biology and synthetic biology applications.[235] The highly-specific T7 RNAP drives expression only from its cognate T7 promoter (P_{T7}), making it orthogonal to native *E. coli* and eukaryotic promoters and polymerases.[236,237] Taken together with the enzyme’s strong

transcriptional activity,[237] this ultimately results in a high signal-to-noise ratio.

From this, we decided to explore developing the split T7 RNAP into a protein fragment complementation assay (PCA) biosensor scaffold for the detection of PPIs. The utility of PCA systems is dictated by the dynamic range of the split reporter, whereby background assembly in the absence of an input signal is low, and assembly driven by pendant fusion domains results in robust activation. The efficient, spontaneous assembly of split T7 RNAP fragments confers a very high "background" signal, precluding their utility as PCA biosensors. To address this limitation, we initiated a program to engineer "proximity-dependent" split T7 RNAPs as a new PCA-based approach for generating robust biosensors that encode molecular interactions in RNA outputs.

In chapter 2, the design and optimization of the split T7 RNAP biosensor is described. Using the previously reported split site at 179, we began to design our split T7 RNAP biosensor. We chose to evolve the N-terminal RNAP half ($RNAP_N$) of the split T7 RNAP due to its smaller size compared to the C-terminal RNAP half ($RNAP_C$), it was solvent exposed and removed from DNA binding, and the $RNAP_C$ contained mutations responsible for promoter recognition. To optimize the split T7 RNAP components to only assemble in the presence of fused interacting PPI partners, we chose to use previously characterized isoluecine zipper peptides as model PPIs to modulate the assembly. Linkers were also considered in the design, and long flexible GGS linkers were chosen to limit dependence on linker length and composition to make up the loss in affinity between the split T7 RNAP halves. With these design considerations and PPI partners in hand, we initiated a directed evolution campaign using phage assisted continuous evolution (PACE) to evolve the desired properties of split T7 RNAP PCA biosensor. Practically, we designed a dual positive and negative selection for PACE based on interacting and non interacting isoleucince zipper PPIs to allow for robust selection for desired properties, namely assembly only when pendant protein fusions interacted.

After 29 days of PACE, the split T7 RNAP had converged on two distinct genotypes,

which were termed 29-1 and 29-8. Notably, the 6 mutations (29-1) and 7 mutations (29-8) observed for each genotype were not localized at the interface between the two split T7 halves, some mutations occurred at solvent exposed positions or at the protein-DNA interface. The mutations observed would not have been predicted based upon rational design of the split enzyme, showing the power of the unbiased directed evolution approach PACE to optimize properties of the split T7 RNAP sensor. To show the versatility of the evolved split T7 RNAP biosensor (29-1 was used for characterization), small molecule- and light-inducible PPI partners were used in place of the zipper peptide fusions to develop both small molecule- and light-activated biosensors. These were tested, and showed robust activation in the presence of small molecule and light, indicating that the evolved split T7 RNAP could function as a ”plug-and-play” biosensor with the ability to swap in new PPI partners. The small molecule-inducible 29-1 split T7 RNAP biosensor was then characterized in both *E. coli* and mammalian systems to show its orthogonality.

The split 29-1 T7 RNAP biosensor was used in subsequent work described in chapter 3. In this study, the split T7 RNAP biosensor was utilized to generate a selection scheme to interrogate the binding interface between KRAS/RAF using deep mutational scanning (DMS). For a DMS selection experiment, there is a need to link the biochemical signal or event, in this case the binding of two PPI partners, to a genetic output. Because the split T7 RNAP biosensor has RNA as an output, simply fusing two PPI partners to each half of the biosensor to modulate its assembly and produce RNA effectively generates a selection system. Typically, optimization is required for each new selection system with known active and inactive variants to ensure expected signals are observed and that there is a sizeable dynamic range between these two variants. Based on our previous experience, we chose to directly fuse the wild type RAF and known null binding variant RAF R89L to the $RNAP_N$ and wild type KRAS to the $RNAP_C$ to assess the selection system’s parameters. Without the need to optimize linkers or any other components of the originally designed split T7 RNAP biosensor, we were able to observe ~20 fold difference between these

two different PPI partners in a luciferase reporter assay, and clear distinction between an interaction and no interaction in plaque assays.

After validating the split T7 biosensor selection system, a library of RAF variants was generated and put through 72 h of phage assisted continuous selection (PACS) to enrich for RAF variants with binding to KRAS. These variants were then assessed using a DMS workflow via high throughput sequencing to gain a deeper understanding of how different variants enriched in the population over the course of time. From the variants observed at the end of the selection, we were able to show that calculated enrichment scores obtained from our selection correlated with known K_D measurements from the literature,[222] and our luciferase reporter assay also correlated with calculated enrichment scores. We then looked at the amino acids observed at the binding interface at 72 h to inform upon the properties of different mutations at the interaction interface between RAF and KRAS. Heat maps were constructed based on relative enrichment values for each amino acid along the binding interface, and satisfyingly, intolerance to amino acid substitution was observed for known hot spot residues at positions 89, 84, and 66. Interestingly, we were able to identify amino acid positions along the binding interface that allowed for a diverse array of amino acid substitutions from multiple different amino acid groups, identifying amino acid position 76 as having the most diverse array of enriched residues. With this split T7 RNAP biosensor selection platform and PACS, we were able to quickly generate a selection scheme and identify key residues in the KRAS/RAF binding interaction without the need to re-optimize any component of the split T7 RNAP biosensor platform. This further demonstrates the versatility of the split T7 RNAP biosensor for use in diverse applications, such as DMS.

In preliminary work outlined in appendix A, I applied the split T7 RNAP biosensor scaffold toward the evolution of a new PPI partner from a naive protein scaffold. In this work, I used reported antibody mimetic scaffolds and nanobodies as starting points for a *de novo* PPI partner for the inflammatory bowel disease (IBD) biomarker calprotectin.

Using an evolution scheme similar to our PACS scheme, we fused either a library of antibody mimetics or nanobodies to the $RNAP_N$ and the biomarker calprotectin to the $RNAP_C$ to evolve a new PPI partner. After 72 h of PACE evolution, I was able to show convergence towards a single genotype in the adhiron phage population and upon assessment of activity in an *in vivo* luciferase reporter assay, the identified genotype was able to drive reassembly of the split T7 RNAP biosensor and generate luminescence signal comparable to the signal observed for the zipper peptide controls. Biochemical characterization of the PPI of this adhiron variant and calprotectin is ongoing, but these preliminary results demonstrate how the split T7 RNAP biosensors can be utilized to drive the evolution of a new PPI partner, in addition to its ability to be used to sense PPIs to either monitor or impact biology.

The proximity-dependent split T7 RNAP-based sensor scaffold essentially allows measurement and encoding of input biomolecular interactions into a defined RNA output. For example, the split T7 RNAP biosensor scaffold naturally lends itself for use in detecting PPIs. Unlike optically-based PPI detection methods, however, the RNA-based output is both exceedingly sensitive (even small changes in output signal can be detected via PCR amplification) and highly multidimensional (due to the information storage capacity of nucleic acids). To demonstrate this concept, the split T7 RNAP fragments were used to observe the interactive interplay between the BCL2 family of apoptotic regulatory proteins, ultimately monitoring and describing a one-by-four competitive PPI interaction network via RT-qPCR of four barcoded RNA output signals.[217] Additionally, we were able to measure pharmacological engagement of the network in a time-dependent manner, showing both on- and off-target effects of PPI inhibitors in live cells. Notably, using RNA and other genetic output signals to store biochemical information in live cells opens up interesting opportunities for bioanalysis.

Aside from producing RNA signals that store information for subsequent analysis, the proximity-dependent split RNAP systems can also be used to drive cellular changes

based on measured input signals. For example, the split RNAP biosensors can be used in mammalian cells to trigger production of mRNA, and therefore protein synthesis, shRNA, and therefore gene knockdown, and guide RNA (gRNA) for CRISPR/Cas9-mediated gene knockout.[215,218] Moreover, multiple small molecule-inducible split T7 RNAP-based biosensors were used to drive different Cas9-mediated genome engineering responses based on disparate small molecule input signals.

Few existing biosensor systems have the dynamic range and input versatility as split T7 RNAP-based biosensors. Moreover, because T7 RNAP directly produces RNA from its own orthogonal, cognate promoter, it is removed from the host cell transcriptional machinery, allowing these sensors to be ported between different host systems with little need for further optimization after the initial characterization, generation, and optimization efforts. Future work expanding on both the possible inputs and outputs of split T7 RNAP-based biosensors will continue to increase the utility of the scaffold in the synthetic biology for the generation of genetic circuits to both detect and impact underlying biology in whichever host system is being studied. Finally, split T7 RNAP can also be repurposed into directed evolution platforms to link phenotypes to genotypes, by modulating the assembly of the T7 RNAP according to the phenotype of interest allowing evolution of interesting new biophysical protein properties (e.g. stability),[209] and, in principle, even evolving new interactions as introduced in appendix A.

APPENDIX A

TOWARD THE EVOLUTION OF A *DE NOVO* BINDER FOR THE IBD BIOMARKER CALPROTECTIN

A.1 Introduction

While generating the versatile split T7 RNA polymerase (RNAP) biosensor platform, which was described in chapter 2, we discovered the pendant protein fused to the N-terminal half of the split T7 RNAP ($RNAP_N$) accumulated mutations. During the course of the phage assisted continuous evolution (PACE), the pendant ZA leucine zipper peptide had two leucine to isoleucine mutations emerge, resulting in increased affinity for its protein interaction partner ZB, and an improved dynamic range as assayed by the luciferase reporter assay (Figure 2.2).[215] Based upon this observation, we explored utilizing the split T7 RNAP biosensor to evolve protein-protein interaction (PPI) partners. We envisioned fusing PPI partners to each half of the split T7 RNAP to modulate the reassembly of the split T7 RNAP to drive a transcriptional output, which could be integrated into a directed evolution system, such as PACE.[169]

For our PPI partner evolution, we chose to evolve a *de novo* protein interaction partner instead of improving the properties of a known PPI partner, as previously demonstrated for the ZA/ZB leucine zipper peptides.[215] While examining potential proteins to use in our evolution, we looked for protein scaffolds that were monomeric and had no post translational modifications or disulfide bonds, which would facilitate their soluble expression in *E. coli* cells to facilitate their use in a split T7 RNAP biosensor-based PACE system. We chose to use reported antibody mimetic scaffolds due to their many favorable properties, such as their relatively small size (<20 kDa) for easy integration into the phage genome, their high mutation-tolerant structural stability, and their ability to be easily expressed in bacterial systems.[238] Additionally, the use of antibody mimetic scaffolds in both the design of diagnostics and therapeutics have previously been demonstrated,[238–242]

with a Kunitz domain (DX-88 (Ecallantide)) and knottin (Linaclotide (Linzess)) being approved for the treatment of hereditary angioedema and inflammatory bowel disease (IBD), respectively.[240]

For our particular study, we decided to utilize the reported adhiron and fynomer antibody mimetic protein scaffolds because they are easily expressed in *E. coli* and are highly stable (adhiron: T_m ~100 °C,[243] fynomer: T_m ~70 °C).[244] The adhiron (also known as affimer) protein scaffold is derived from a consensus sequence of phytocystatin proteins, and has been used in the development of high affinity binding reagents for many proteins,[243,245] including SUMO proteins,[246] Fcγ receptors,[247] actin,[248] and BCL family proteins.[249] Also, the adhiron scaffold has been developed into imaging reagents for DNA-PAINT,[250] and as detection reagents for biomarkers[251–253] and small molecules.[254]

The fynomer protein scaffold is derived from the SH3 domain of the human Fyn tyrosine kinase and has been evolved to bind several different proteins, including IL-17A,[255] fibronectin,[244] albumin,[256] chymase,[257] BACE-2,[258] CD33,[259] and HER2.[260] This particular protein scaffold has also been developed into bi-specific targeting molecules through fusions with antibodies, termed FynomAb.[259–262] FynomAB have improved pharmacokinetics compared to the fynomer protein alone and have been explored as potential therapeutics, with one bispecific FynomAb undergoing a phase I and II clinical trials for the treatment of plaque psoriasis.[261]

We also explored utilizing antibody fragments as starting points in our evolution of a *de novo* protein binder due to their well defined epitope regions and their proven ability to generate novel protein binders. In particular, we chose to explore nanobodies. Nanobodies are an antibody fragment comprised of the variable region (VHH) derived from the heavy chain antibodies (HCAbs) found in camelids.[263,264] Due to the small size of nanobodies compared to full length HCAbs and IgG antibodies and IgG antibody fragments, these have been explored for use as high affinity binding reagents for applications in structural biology, basic research, diagnostic tools, and therapeutic reagents,[265–268] with proto-

cols available for the production of humanized nanobodies for clinical applications.[269,270] Some recent examples of nanobody technology include fusing the pro-drug cisplatin to an anti-EGFR nanobody protein construct to facilitate localized unmaking of this therapeutic at tumor sites in mice,[271] and the development of nanobody-based chimeric antigen receptor (CAR)-expressing T-cells for the treatment of solid tumors.[272]

For our study, we decided to evolve *de novo* protein binders for the inflammatory bowel disease (IBD) biomarker calprotectin.[273] This particular biomarker is used to monitor inflammation in the gastrointestinal tract in IBD patients, in conjugation with other methods, such as colonoscopies and patient reported symptoms.[274–278] Due to the nature of IBD, the ability to rapidly detect inflammation to inform treatment decisions is important in the management of the disease and improving disease outcomes.[279] Currently, ELISA-based kits or other antibody-based detection methods are available to detect calprotectin.[280,281] However, many of these tests are limited with regards to long turnaround times and issues with patient compliance with sample collection.[282] The recently developed point-of-care IBDoc Calprotectin Home Testing kit helps ameliorate some of these issues, but data is still limited on its widespread use.[283–285]

Due to the importance of monitoring levels of inflammation in IBD patients, we wanted to develop a point-of-care test to monitor calprotectin levels that could be utilized by patients early detection of inflammation, even in the absence of other symptoms. In order to accomplish this, we would need to develop a way to both detect calprotectin and generate an easily detectable output to quantify calprotectin levels. Ultimately, we decided to develop a biosensor based on protein fragment complementation (PFC) strategies,[67–70] using either split luciferase[71,76,81,286] or spit GFP,[77,79,80,287] to generate light as our output for easy monitoring. This would require evolving new protein interaction partners for calprotectin, ideally on two different regions of the calprotectin heterodimer, to facilitate split enzymatic reporter reassembly and generate our chosen output, light (Figure A.1). Below I will outline the work I have done to generate a *de novo* protein binder for

calprotectin, and outline the future work that needs to be done to generate a calprotectin biosensor.

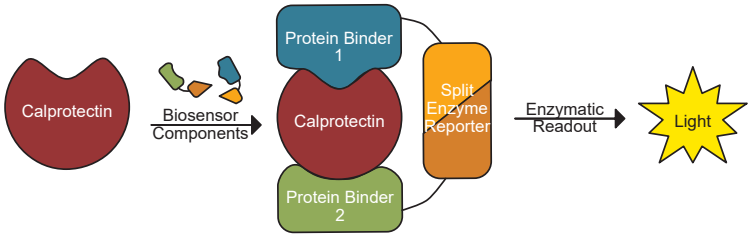

Figure A.1: Schematic of a protein fragment complementation (PFC) calprotectin biosensor. Protein binder 1 and protein binder 2 are each fused to one half of a split enzymatic reporter. Upon protein binder 1 and protein binder 2 binding to calprotectin, the split enzyme reporter reassembles to produce a readily-detectable output signal, such as light.

A.2 Preliminary results

Following similar designs to the *in vivo* luciferase assay and PACE system[288] described in chapter 3, we developed both an *in vivo* luciferase assay and PACE system using our split T7 RNAP biosensor to both monitor binding properties and evolve a protein binder for calprotectin, respectively.

For the *in vivo* luciferase assay, we developed a 3-plasmid system to detect an interaction between potential protein binders and calprotectin: 1) a N-terminal expression plasmid with the N-terminal half of our split T7 RNAP biosensor ($RNAP_N$) C-terminally fused to either a naive protein scaffold (adhiron,[243] fynomer,[257] or nanobody[289]), an off-target PPI partner, or no fusion partner; 2) a C-terminal expression plasmid with the C terminal half of our split T7 RNAP biosensor ($RNAP_C$) N-terminally fused to a modified calprotectin heterodimer (see Method A.3.2), off-target PPI partner, or no fusion partner; and 3) a reporter plasmid with the bacterial luciferase luxAB[219] gene under control of the T7 promoter (Figure A.3a). Only when the protein fusion on the $RNAP_N$ binds the protein fusion on the $RNAP_C$ will the split T7 RNAP biosensor reassemble and transcribe the

luxAB gene upstream of the T7 promoter to produce luminescence.

For our PACE evolution, we developed M13 bacteriophage and *E. coli* accessory plasmids to evolve our naive protein scaffolds to bind calprotectin. We modified a previously optimized M13 bacteriophage genome[215] by removing gIII and replacing it with our $RNAP_N$ C-terminally used to one of the naive protein scaffolds (fynomer, adhiron, nanobody) (Figure A.2a). For our *E. coli* cells, we developed *E. coli* accessory plasmids to modulate expression of gIII based upon the assembly of our split RNAP biosensors: 1) a "positive accessory plasmid" (posAP) containing the CGG $RNAP_C$ half of our split RNAP biosensor, which is a known variant of the T7 RNAP that recognizes an orthogonal CGG RNAP promoter, N-terminally fused to our modified calprotectin construct, and gIII under control of the CGG RNAP promoter; and 2) a "negative accessory plasmid" (negAP) containing the T7 $RNAP_C$ half of our split RNAP biosensor and gIIIneg, a dominant negative form of gIII that poisons the ability of phage to produce infectious progeny, under control of the T7 RNAP promoter (Figure A.2a). Only when a PPI occurs between calprotectin and a protein with affinity for calprotectin would the split fragments of the CGG RNAP reassemble to transcribe gIII, which would result in the production of infectious phage progeny. Any other assemblies would result in activation of both the posAP and negAP, resulting in the production of gIIIneg, which would poison the ability phage to generation infectious progeny.

Due to the lack of known interactions between the selected naive protein scaffolds and calprotectin in the literature and preliminary *in vivo* luciferase assay tests, we decided to generate libraries of each of the naive proteins selected to prevent premature termination of the evolution (wash out of the phage) do to low phage replication rates in PACE. This was done based on a previous report showing increased selection stringency associated with the continuous flow of PACE compared to iterative rounds of phage propagation in phage-assisted non-continuous evolution (PANCE) for the evolution of a methanol dehydrogenase enzyme; only after rounds of evolution in the PANCE system was phage

replication supported in the continuous flow of PACE.[206]

The generated phage libraries for each protein scaffold were then tested in auxiliary plaque assays to determine the ability of each library to both infect and activate the accessory plamsids in *E. coli* cells. Based upon initial plaque assay results with our weakest posAP containing cells (p15a orgin, 27-55), we realized that our adhiron protein library and fynomer protein library contained variants with some level of affinity for calprotectin prior to evolution (Figure A.2b). However, as we increased the stringency of our system by decreasing the gene expression of the $RNAP_C$ protein fusion in the posAP by modulating the copy number of the plasmid, the ability to detect distinct plaques diminished, indicating that we could tune the selection pressure over the course of an evolution to improve the binding characteristics of this adhiron protein binder.

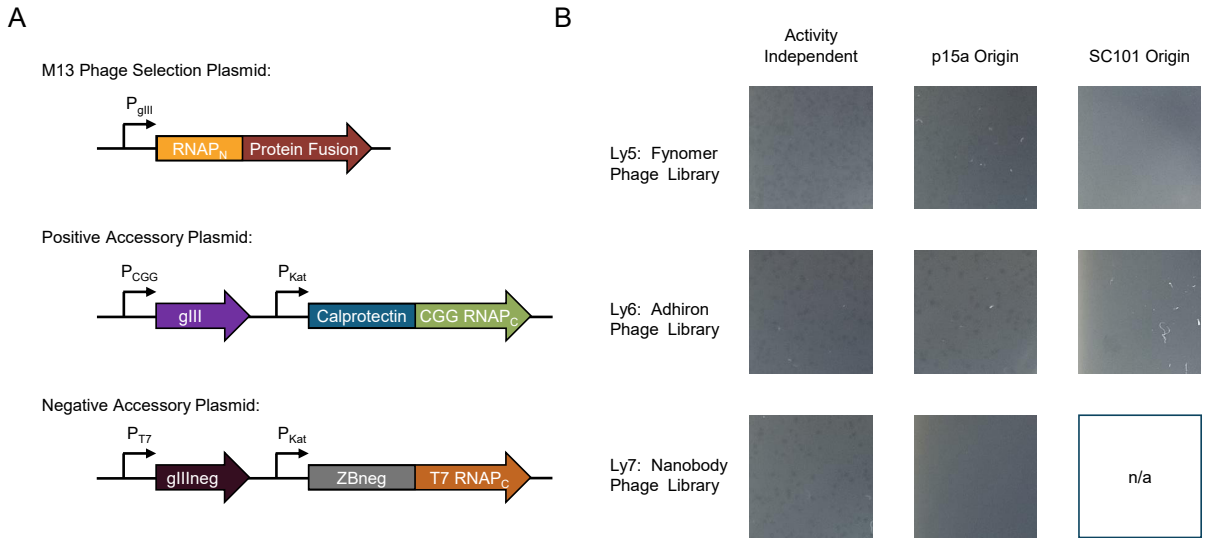

Figure A.2: Design of a phage-assisted continuous evolution (PACE) system to evolve a calprotectin PPI partner. (A) Plasmids for the *in vivo* PACE experiment to evolve a calprotectin binder from a naïve protein scaffold. (B) Plaque assays of starting phage libraries on activity depended and independent cells. Activity independent plaque assays were performed on 1059 cells;[171] activity dependent plaque assays were performed with S1030 cells[171] transformed with a negative accessory plasmid (Jin 409), and either a p15a origin (27-55) or sc101 (28-41) origin calprotectin positive accessory palsmid.

After demonstrating that members of the adhiron and fynomer phage library contained variants with affinity for calprotectin, we performed a PACE experiment to improve

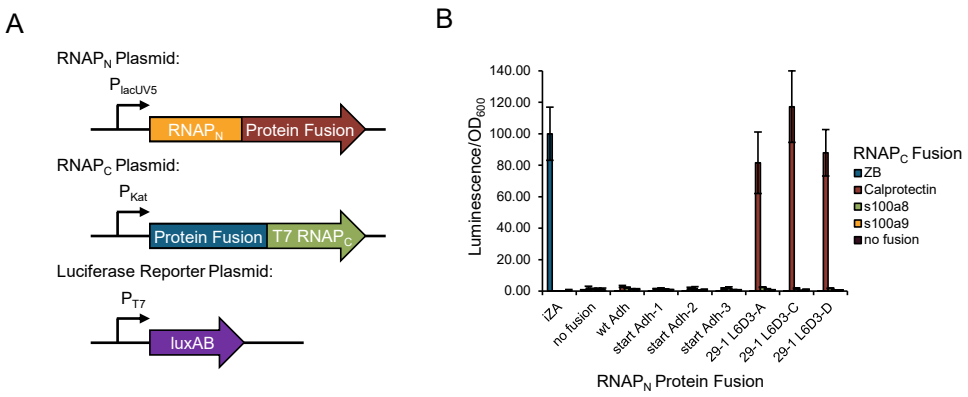

Figure A.3: Detecting the interaction between adhiron protein variants and calprotectin using the *in vivo* luciferase assay. (A) Plasmids for the *in vivo* luciferase assay to monitor the interaction between a potential protein binder and calprotectin. (B) Activity of adhiron variants (wild type adhiron: wt Adh, starting adhiron library variants: start Adh-#, evolved adhiron variants: 29-1 L6D3-X) on off-target and on-target fusions as monitored by the *in vivo* luciferase assay; only pendent protein fusions that interact with each other can drive the reassembly of the split RNAP biosensor to produce the luciferase enzyme and luminescence. Luminescence signal for the tested PPIs as monitored by the *in vivo* luciferase assay (error: std. dev., n= 3).

the binding characteristics of these variants. After 24 h of evolution, the fynomer phage library was unable to support phage replication and the phage population dropped below the detection limit of our plaque assays (~10^4). After 72 h of evolution, the adhiron phage library still contained a robust phage population (~10^8). A small subset of the adhiron variants from the end of this PACE evolution were then sub-cloned into the N-terminal expression vector for the luciferase assay and tested. The results from the *in vivo* luciferase assay were promising; there was a detectable interaction between the isolated adhiron variants and calprotectin comparable to the iZA/ZB isoleucine zipper peptide PPI control (Figure A.3b). This result suggested that we had evolved an adhiron variant that could bind calprotectin. Further examination of the bulk phage population via sequencing showed that the phage population converged on a single genotype after 72 h (Figure A.4), indicating that this particular variant was capable of binding calprotectin and out-competed all other variants that occurred over the course of the evolution.

These two observations prompted experiments to biochemically validate and charac-

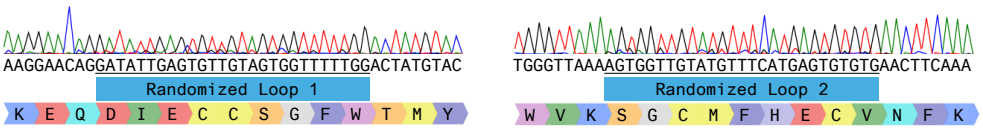

Figure A.4: Bulk sequencing of the evolved adhiron phage population after 72 h of PACE. The annotated randomized loop 1 and 2 are the same regions as indicated in the original paper describing the adhiron protein scaffold.[243]

terize the binding interaction between the identified adhrion variants and calprotectin.

In order to biochemically characterize the interaction between the identified adhiron variants and calprotectin, these proteins were cloned into expression vectors, purified, and analyzed for binding in dot blot assays. Under the conditions utilized, there was no observable interaction between the selected adhiron variant and calprotectin, when compared to an positive control antibody which recognizes the native protein structure of calprotectin[290] (Figure A.5). After assessing the approximate amount of available adhiron protein used in the experiment, it was determined that the protein solution used contained only 0.05 µM total adhiron protein, which is 20 fold less protein than the highest calprotectin concentration used in the dot blot assay. To ensure that adhiron protein amount wasn't a limiting factor in detecting a binding interaction, dot blots with higher concentrations of the adhiron proteins (10 fold excess or more) will be used to assess binding.

This work is ongoing with many different avenues being explored to both evolve and characterize new PPI partners for calprotectin. Improvements in the design of *in vitro* binding assays to detect the PPI between an evolved protein binder and calprotectin are currently being explored, by assessing different protein concentrations and salt conditions. Other avenues to continue to evolve the protein are being explored, including continuing the evolution on our current phage population to improve binding characteristics before assessing it in the *in vivo* luciferase assay and *in vitro* dot blot assay, as well as exploring selection-based directed evolution systems, such as phage display, to enrich for variants with the ability to bind calprotectin.

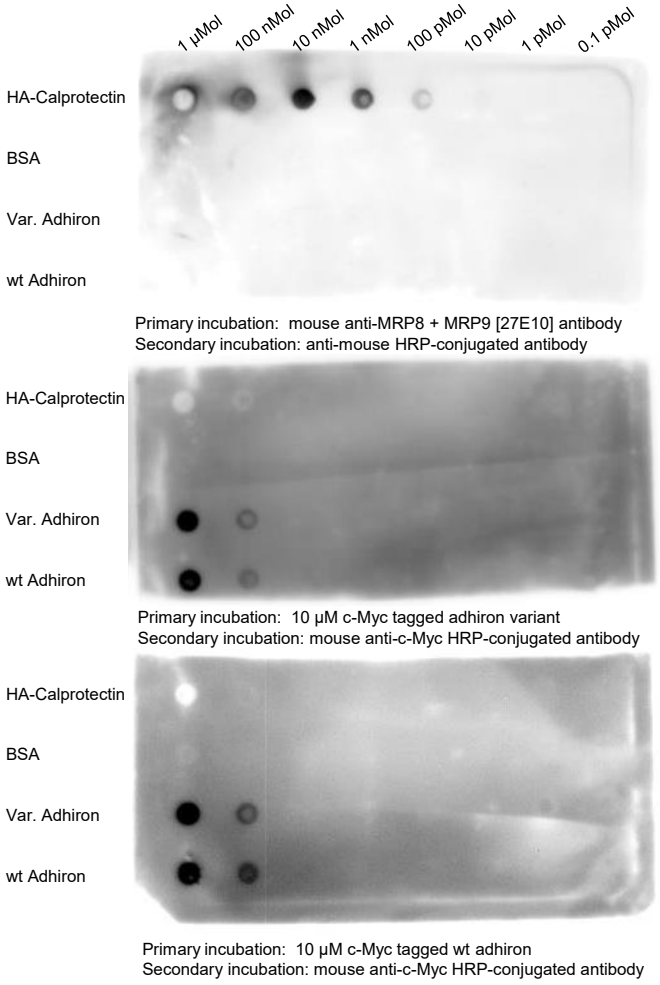

Figure A.5: Detecting the binding interaction between an evolved adhiron protein variant and calprotectin. Each dot blot was spotted with a dilution series of purified HA-tagged fused calprotectin, BSA, purified c-Myc-tagged adhiron variant, and purified c-Myc-tagged wt adhrion protein at the indicated concentrations. These blots were then incubated with the indicated primary antibody or protein solution overnight, and the indicated secondary antibody solutions for 2 h. These were then visualized using a chemiluminescent substrate and imaged on a Fluor Chem R (Protein Simple).

A.3 Methods

A.3.1 General cloning

All plasmids used in this study were generated using Gibson assembly[184] of PCR products generated with NEB Q5 high fidelity polymerase or Phusion polymerase. All constructs were validated via Sanger sequencing performed by the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping facility. All constructs cloned for and/or used in this study are listed in tables B.8, B.9, B.10, B.11, B.12, and B.13.

A.3.2 Calprotectin protein constructs

Calprotectin is a heterodimer of the s100a8 (MRP8) and s100a9 (MRP12) proteins. A fused version of calprotectin was generated based upon crystal structures to ensure equal expression of the two proteins in the heterodimer to minimize the formation of homodimers.[291] We also generated fused versions for the two homodimers following the same design as for the calprotectin heterodimer.

The constructs for the s100a8 and s100a9 proteins, as well as, the fused protein dimers are depicted below. The s100a8 protein is in italics, the linker is in bold, and the s100a9 protein is underlined.

s100a8: *MLTELEKALNSIIDVYHKYSLIKGNFHAVYRDDLKKLLETECPQYIRKK GADVWFKELDINTDGAVNFQEFLILVIKMGVAAHKKSHEESHKE*

s100a9: MTCKMSQLERNIETIINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNF LKKENKNEKVIEHIMEDLDTNADKQLSFEEFIMLMARLTWASHEKMHEGDEGP GHHHKPGLGEGTP

Calprotectin fusion (12aa linker): S100a9 – **GGSGGSGGSGGS** – *S100a8*

S100a8 fusion (12 aa linker): *s100a8* – **GGSGGSGGSGGS** – *s100a8*

S100a9 fusion (12 aa linker): s100a9 – **GGSGGSGGSGGS** – s100a9

A.3.3 Preparation of super-electrocompetent ss1059 cells

ss1059 cells were generated by transforming in the PSP promoter gIII vector from the previously reported 1059 cells[171] into ss320 cells* and plating them onto a 2xYT bacterial agar plate (15 g agar/1 L 2xYT) containing 50 µg/mL carbenicillin.

A single colony from this transformation was then grown to saturation overnight in 50 mL of super broth containing 50 µg/mL carbenicillin. The 50 mL overnight culture was used to inoculate 2 L of super broth without antibiotics. This was grown with shaking at 37 °C until the cells reached an OD_{600} of ~0.6. The cells were then placed on ice for ~30 min, and were kept on ice or at 4 °C for the remaining procedure. These were then centrifuged (2000 xg, 10 min, 4 °C) to remove the supernatant, and they were subsequently washed twice with 200 mL 1 mM HEPES buffer, and once with 250 mL H_2O. The cells were then suspended in 2 mL H_2O. Glycerol was added to the cells to be ~10-15% of the total volume; then the cells were flash frozen in liquid nitrogen, and stored in the -80 °C.

The colony forming units (CFU) of the prepared ss1059 cells was determined with a surrogate *E. coli* plasmid of similar size to the phage vector (6-22) to confirm the transformation efficiency prior to transformation of the phage library. This was done by electroporating the same ratio of DNA to cells as for the phage library transformation (1800 V, 5 msec, 0.2 cm cuvettes, 50 µL cells, ~143 ng DNA). The cells were recovered in 3.6 mL of SOC media for 1 h, serially diluted in 2xYT media (dilution factors: 1x, 10x, 100x, 1000x, 10000x, 100000x, 1000000x, and 10000000x), and 10 µL of each dilution was plated onto a 2xYT bacterial agar plate (15 g agar/1 L 2xYT) containing antibiotics. The following morning the CFU was determined based on the equation below.

*. The ss320 cells were a gift from the Piccirilli lab.

$$CFU = \frac{\left(\frac{\#\ colonies}{volume\ plated\ (10\ \mu L)}\right) \times Dilution\ factor \times Volume\ recovery\ media\ (3600\ \mu L)}{DNA\ used\ in\ transformation\ (0.143\ \mu g)}$$

A.3.4 Antibody mimetics and nanobody phage library generation

Three different peptide libraries were generated via restriction cloning methods. The extact protocol for each library is outlined below.

Adhiron phage library generation

The backbone was prepared by PCR amplification of the M13 bacteriophage vector containing the wild type adhiron protein[243] (12-42) with NEB Phusion DNA polymerase using primers containing EcoRI and BamHI restriction enzyme cut sites (JZ-285 and JZ-422). Briefly, 50 µL PCR reactions were set up with the following final concentrations to generate ~1 µg of product: 1 ng/50 µL 12-42 template plasmid, 0.5 µM forward primer (JZ-285), 0.5 µM reverse primer (JZ-422), and 1x NEB Phusion Hot Start Flex 2x Master Mix. The thermocycler conditions utilized an initial denaturation step (95 °C, 1 min) followed by 30 cycles of denaturation (95 °C, 7 sec), annealing (59.5 °C, 15 sec) and extension (72 °C, 3 min 5 sec), and a final extension (72 °C, 5 min). The PCR product was digested with 1 µL NEB DpnI enzyme in the PCR reaction buffer for 3 h at 37 °C. These were then purified with a Zymo clean and concentrator kit.

The insert library was prepared by PCR amplification of a DNA oligo containing 11 NNK randomized codons (JZ-405) with NEB Q5 DNA polymerase using primers containing EcoRI and BamHI restriction enzyme cut sites (JZ-408 and JZ-409). Briefly, 50 µL PCR reactions were set up with the following final concentrations to generate ~1 µg of product: 0.1 µM JZ-405 template primer, 0.5 µM forward primer (JZ-408), 0.5 µM reverse primer (JZ-409), and 1x NEB Q5 reaction buffer, 0.2 mM dNTPs, and 0.02 U/µL NEB Q5 DNA polymerase. The thermocycler conditions utilized an initial denaturation step (95 °C, 1

min) followed by 7 cycles of denaturation (95 °C, 7 sec), annealing (59 °C, 15 sec) and extension (72 °C, 8 sec), followed by 23 cycles of denaturation (95 °C, 7 sec), annealing (70 °C, 15 sec) and extension (72 °C, 8 sec), and a final extension (72 °C, 2 min). These were then run on a DNA agarose gel, and gel purified with a Zymo gel purification kit.

Inserts were subcloned into the backbone using the BamHI and EcoRI cut sites with NEB BamHI-HF enzyme and NEB EcoRI-HF enzyme following the standard NEB double digest protocol. Briefly, 50 µL restriction digestion reactions were set up for both the insert and backbone with the following final concentration for each component: 1 µg/50 µL insert or backbone DNA, 0.2 U/µL NEB BamHI-HF enzyme, 0.2 U/µL NEB EcoRI-HF enzyme, and 1x NEB cut smart buffer. These were kept at 37 °C for 1 h, then purified using a Zymo DNA clean and concentrator kit.

The T4-mediated ligation reactions were performed at 25 °C for 45 min using NEB T4 DNA ligase following the standard NEB ligation protocol to generate roughly 1 µg of the final adhiron DNA library. Briefly, 50 µL ligation reactions were set up with the following final concentration for each component: 200 ng/µL restriction digested backbone, 50 ng/µL restriction digested insert, 20 U/µL NEB T4 DNA ligase, and 1x NEB T4 DNA ligase buffer. After ligation, these were purified with a Zymo DNA clean and concentrator kit. All resulting plasmids were suspended in 30 µL water.

The phage library was prepared by electroporating (1800 V, 5 msec, 0.2 cm cuvettes, 50 µL cells, ~143 ng DNA, 7 transformations total) 1 µg ligated vector product into 350 µL super-electrocompetent ss1059 cells (as described above), which were recovered in 25 mL of SOC media for 1 h. Afterwards, the recovered cells were used to inoculate 250 mL super broth to allow for phage library growth overnight. The phage containing supernatant was collected by centrifugation and sterile filtration to generate an adhiron phage library.

Fynomer phage library generation

The backbone was prepared by PCR amplification of the M13 bacteriophage vector containing the wild type fynomer protein^{257} (11-40) with NEB Q5 DNA polymerase using primers containing XhoI and BamHI restriction enzyme cut sites (JZ-285 and JZ-286). Briefly, 50 µL PCR reactions were set up with the following final concentrations to generate ~1 µg of product: 1 ng/50 µL 11-40 template plasmid, 1 µM forward primer (JZ-285), 1 µM reverse primer (JZ-286), 1x NEB Q5 reaction buffer, 0.2 mM dNTPs, and 0.02 U/µL NEB Q5 DNA polymerase. The thermocycler conditions utilized an initial denaturation step (95 °C, 1 min) followed by 30 cycles of denaturation (95 °C, 7 sec), annealing (63.5 °C, 15 sec) and extension (72 °C, 2 min 54 sec), and a final extension (72 °C, 5 min). The PCR product was digested with 1 µL NEB DpnI enzyme in the PCR reaction buffer for 3 h at 37 °C. These were then purified with a Zymo clean and concentrator kit.

The insert library was prepared by PCR amplification of a DNA oligo containing 12 NNK randomized codons (JZ-404) with NEB Phusion DNA polymerase using primers containing XhoI and BamHI restriction enzyme cut sites (JZ-280 and JZ-281). Briefly, 50 µL PCR reactions were set up with the following final concentrations to generate ~1 µg of product: 0.1 µM JZ-404 template primer, 0.5 µM forward primer (JZ-280), 0.5 µM reverse primer (JZ-281), and 1x NEB Phusion Hot Start Flex 2x Master Mix. The thermocycler conditions utilized an initial denaturation step (95 °C, 1 min) followed by 7 cycles of denaturation (95 °C, 7 sec), annealing (59 °C, 15 sec) and extension (72 °C, 8 sec), followed by 23 cycles of denaturation (95 °C, 7 sec), annealing (70 °C, 15 sec) and extension (72 °C, 8 sec), and a final extension (72 °C, 5 min). These were then purified with a Zymo clean and concentrator kit.

Inserts were subcloned into the backbone using the BamHI and XhoI cut sites with NEB BamHI-HF enzyme and NEB XhoI enzyme following the standard NEB double digest protocol. Briefly, 50 µL restriction digestion reactions were set up for both the insert and backbone with the following final concentrations for each component: 1 µg/50 µL insert or

backbone DNA, 0.2 U/μL NEB BamHI-HF enzyme, 0.2 U/μL NEB XhoI enzyme, and 1x NEB cut smart buffer. These were kept at 37 °C for 1 h, then purified using a Zymo DNA clean and concentrator kit.

The T4-mediated ligation reactions and fynomer phage library generation were performed as described for the adhiron phage library.

Nanobody phage library generation

The backbone was prepared by PCR amplification of the M13 bacteriophage vector containing a previously reported nanobody protein[289] (11-41) with NEB Q5 DNA polymerase using primers containing EcoRI and BamHI restriction enzyme cut sites (JZ-430 and JZ-412). Briefly, 50 μL PCR reactions were set up with the following final concentrations to generate ~1 μg of product: 1 ng/50 μL 11-41 template plasmid, 0.5 μM forward primer (JZ-430), 0.5 μM (JZ-412), 1x NEB Q5 reaction buffer, 0.2 mM dNTPs, and 0.02 U/μL NEB Q5 DNA polymerase. The thermocycler conditions utilized an initial denaturation step (95 °C, 1 min) followed by 25 cycles of denaturation (95 °C, 7 sec), annealing (58.9 °C, 15 sec) and extension (72 °C, 3 min 15 sec), and a final extension (72 °C, 6 min). The PCR product was digested with 1 μL NEB DpnI enzyme in the PCR reaction buffer for 3 h at 37 °C. These were then purified with a Qiagen reaction cleanup kit.

The insert library was prepared by PCR amplification of a DNA oligo containing 12 NNK randomized codons (JZ-407) with NEB Q5 DNA polymerase using primers containing XhoI and BamHI restriction enzyme cut sites (JZ-414 and JZ-415). Briefly, 50 μL PCR reactions were set up with the following final concentrations to generate ~1 μg of product: 0.1 μM JZ-407 template primer, 0.5 μM forward primer (JZ-414), 0.5 μM reverse primer (JZ-415), 1x NEB Q5 reaction buffer, 0.2 mM dNTPs, and 0.02 U/μL NEB Q5 DNA polymerase. The thermocycler conditions utilized an initial denaturation step (95 °C, 1 min) followed by 30 cycles of denaturation (95 °C, 7 sec), annealing (60 °C, 15 sec) and extension (72 °C, 8 sec), and a final extension (72 °C, 2 min). These were then run on a DNA agarose gel,

and gel purified with a Qiagen gel purification kit.

Inserts were subcloned into the backbone using the BamHI and EcoRI cut sites with NEB BamHI-HF enzyme and NEB EcoRI-HF enzyme following the standard NEB double digest protocol. Briefly, 50 µL restriction digestion reactions were set up for both the insert and backbone with the following final concentrations for each component: 1 µg/50 µL insert or backbone DNA, 0.2 U/µL NEB BamHI-HF enzyme, 0.2 U/µL NEB EcoRI-HF enzyme, and 1x NEB cut smart buffer. These were kept at 37 °C for 1 h, then purified using a Zymo DNA clean and concentrator kit.

The backbone was then treated with antarctic phosphatase following the standard NEB protocol. Briefly, 50 µL reactions were set up with the following final concentrations for each component: 1 µg/50 µL restriction digested backbone DNA, 1x NEB antarctic phosphatase reaction buffer, and 0.1 U/µL NEB antarctic phosphatase. This was kept at 37 °C for 45 minutes, then these were purified with a Zymo clean and concentrator kit.

The T4-mediated ligation reactions and nanobody phage library generation were performed as described for the adhiron phage library.

A.3.5 Plaque assays

To quantify phage titers, 1059 activity-independent cells[171] were used; while S1030 cells[171] transformed with a positive accessory plasmid (see table B.12) and a negative accessory plasmid (Jin 409, see table B.12) were used to quantify the activity of different phage containing $RNAP_N$ protein fusions on calprotectin. For all experiments, *E. coli* were grown to an optical density (OD_{600}) of approximately 0.6.

Phage dilutions (1/50, 1/12500, 1/125000, 1/6250000) were made by serially diluting a phage sample in water. Plaque assays were performed by mixing 50 µL of each phage dilution with 50 µL of cells, suspending these solutions in 55 °C warmed soft agar (7 g agar/1 L LB), and plating this mixture onto quartered solid agar plates without antibiotics (15 g agar/1 L LB). These were then grown in a 37 °C incubator overnight, and phage

plaques were counted and imaged to determine the relative plaque forming units (PFU).

A.3.6 Luciferase assays

Three separate vectors were constructed: 1) a N-terminal expression plasmid containing the $RNAP_N$ fragment C-terminally fused to the wild type or variant antibody mimetic (adhiron, fynomer) or camelid proteins (nanobody), an off target protein partner, or no protein partner upstream of the UV5 IPTG inducible promoter,[233] 2) a C-terminal expression plasmid containing the T7 $RNAP_C$ fragment N-terminally fused to either calprotectin, s100a8 fused homodimer, s100a9 fused homodimer, an off target protein partner, or no protein partner upstream of the constitutive P_{Kat} promoter, and 3) a luciferase reporter plasmid containing the bacterial luxAB gene[219] upstream of the T7 RNAP promoter. The N-terminal, C-terminal, and the luciferase reporter plasmids were electroporated into S1030 cells,[171] and plated onto agar (15 g/L in LB) plates containing 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, and 7.5 µg/mL tetracycline. Individual colonies were then picked and grown overnight to saturation with shaking at 37 °C in 500 µL of 2xYT media containing 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, 7.5 µg/mL tetracycline in a 96, deep-well plate. A new culture was then prepared using 60 µL of the overnight culture to inoculate 540 µL of fresh 2xYT media containing 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 33 µg/mL chloramphenicol, 7.5 µg/mL tetracycline and 1 µM IPTG. After 3 h of growth at 37 °C with shaking, a 150 µL aliquot was transferred to a black-walled, clear-bottomed, 96-well plate and read on a Biotek Synergy Neo2 plate reader, monitoring both the optical density at 600 nm (OD_{600}) and the luminescence signal.

Luciferase activity was reported as normalized luminescence signal/OD_{600} for each well. Values are reported as the arithmetic average of replicate wells ($n = 3$) with error reported as the standard deviation of the replicates.

A.3.7 PACE

PACE was performed utilizing a modified version of the previously described method.[215] S1030 cells were transformed by electroporation with a positive accessory plasmid (27-55), a negative accessory plasmid (Jin 409), and the *in vivo* mutagenesis plasmid MP6.[223] A 5 mL starter culture was grown overnight to saturation in LB media supplemented with 50 μg/mL carbenicillin, 33 μg/mL chloramphenicol, 40 μg/mL kanamycin, 7.5 μg/mL tetracycline, and 20 mM glucose. Chemostats (250 mL sterile bottles) containing 150 mL of Davis rich medium[215] were inoculated with 5 mL starter culture and grown at 37 °C with magnetic stir-bar agitation. When the OD_{600} reached ~0.6, fresh Davis rich medium was pumped in at 120-150 mL per h, with a waste needle set at 150 mL. Selection phage containing $RNAP_N$ protein fusions were used to seed fresh lagoons (25 mL flask with a rubber septum) in triplicate. For the initial PACE experiment 1 mL of each phage library was used to inoculate separate lagoons. Waste needles were set to maintain the lagoon volume at 20-25 mL, and host cell cultures were flowed in at 20-25 mL per h. Mutagenesis was initiated by the direct addition of arabinose (20% w/v in water) into each lagoon at a rate of 0.5 mL per h. Lagoon samples were collected every 24 h from the waste withdrawal line, centrifuged, sterile filtered, and the supernatant was stored at 4 °C.

A.3.8 Protein expression and purification of calprotectin

The fused calprotectin construct was cloned into a pET30 protein expression vector containing an N-terminally fused 6xHis tag and an N-terminally fused 3xHA tag using the procedures outlined for the general cloning procedures to generate vector 29-56.

The 29-56 vector was transformed into electrocompetent BL21 cells and streaked onto a bacterial agar plate (15 g agar/1 L LB) containing 40 μg/mL kanamycin. A single colony was then used to inoculate 8 mL of 2xYT media supplemented with 40 μg/mL kana-

mycin and 1% glucose and grown overnight. 5 mL of this overnight culture was then used to inoculate 500 mL of 2xYT media supplement with 40 µg/mL kanamycin, and this was allowed to grow at 37 °C with shaking until the OD_{600} reached ~0.6. The cells were chilled on ice for 1 h, and then protein expression was induced with 0.5 mM IPTG and grown overnight at 16 °C with shaking.

The cells were kept on ice for the remaining procedure. The cells were pelleted (1500 xg, 15 min, 4 °C), then suspended in 20 mL of lysis buffer (50 mM Tris, 1 M NaCl, 10 mM TCEP, 20% glycerol, pH 7.5) containing protease inhibitors (AEBSF (1 mM), aprotinin (200 nM), bestatin (10 µM), E64 (20 µM), leupeptin (100 µM), pepstatin A (20 µM)). The cells were then lysed using sonication (70% amplitude, 1 min total time (1 sec on, 1 sec off pulses), 3x). Lysates were cleared by centrifugation (12000 rpm, 40 min, 4 °C). Cleared lysates were incubated with His60 Ni Superflow Resin (Takara) for 1 h at 4 °C with constant gentle agitation. After 1 h, the resin was washed with lysis buffer and eluted with a gradient imidazole elution (10 mM-250 mM). Fractions containing the calprotectin protein were identified via SDS-PAGE gel, and concentrated using Ultra-15 Centrifugal Filter Units with 10 kDa cutoff (Amicon, EMD Millipore). After sufficient concentration, the combined fractions were desalted and buffer exchanged into protein storage buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% glycerol, 1 mM DTT) using Sephadex G-25 in PD-10 Desalting Columns (GE Healthcare Life Sciences). The protein solution was aliquoted and stored at -80 °C.

A.3.9 Protein expression and purification of wt adhiron and adhiron variants

The wild type adhiron protein and an adhiron variant identified as having the ability to bind calprotectin as determined from the *in vivo* luciferases assay were cloned into a pET30 protein expression vector containing an C-terminally fused 6xHis tag and an N-

terminally fused c-Myc tag using the procedures outlined for the general cloning procedures to generate vectors 31-35 and 31-36, respectively.

The adhiron protein expression vector was transformed into chemicompetent Rosetta cells and streaked onto bacterial agar plates (15 g agar/1 L LB) containing 40 µg/mL kanamycin and 33 µg/mL chloramphenicol. A single colony was then used to inoculate of 50 mL of 2xYT media supplemented with 40 µg/mL kanamycin, 33 µg/mL chloramphenicol, 1% glucose and grown overnight. 50 mL of this overnight culture was then used to inoculate 500 mL of 2xYT media supplement with 40 µg/mL kanamycin and 33 µg/mL chloramphenicol, and this was allowed to grow at 37 °C with shaking until the OD_{600} reached ~0.6. The cells were chilled on ice for 1 h, and then protein expression was induced with 0.5 mM IPTG and grown overnight at 16 °C with shaking.

The cells were kept on ice for the remaining procedure. The cells were pelleted (1500 xg, 15 min, 4 °C), then suspended in 20 mL of lysis buffer (50 mM Tris, 1 M NaCl, 10 mM TCEP, 20% glycerol, pH 7.5) containing protease inhibitors (AEBSF (1 mM), aprotinin (200 nM), bestatin (10 µM), E64 (20 µM), leupeptin (100 µM), pepstatin A (20 µM)). The cells were then lysed using sonication (70% amplitude, 1 min total time (1 sec on, 1 sec off pulses), 3x). Lysates were cleared by centrifugation (12000 rpm, 40 min, 4 °C). Cleared lysates were incubated with His60 Ni Superflow Resin (Takara) for 1 h at 4 °C with constant gentle agitation. After 1 h, the resin was washed with lysis buffer and eluted with a gradient imidazole elution (10 mM-250 mM). Fractions containing the calprotectin protein were identified via SDS-PAGE gel, and concentrated using Vivaspin 20 centrifugal concentrators with 3 kDa cutoff (Sartorius). After sufficient concentration, the combined fractions were desalted and buffer exchanged into protein storage buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% glycerol, 1 mM DTT) using Sephadex G-25 in PD-10 Desalting Columns (GE Healthcare Life Sciences). The protein solution was aliquoted and stored at -80 °C.

A.3.10 Western blot

Purified HA tagged calprotectin and c-Myc tagged adhiron protein solutions were used to confirm the ability to detect the protein tags for subsequent analysis.

Protein loading buffer (50 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.05% bromphenol blue, 100 mM DTT) was added to 1 µg of the indicated purified protein, and boiled for 10 min at 95°C and loaded onto a 10% or 12% SDS-PAGE gel. After stacking at 70 V, the gel was run at 120 V until the dye front reached the bottom, and the proteins were transferred onto a PVDF membrane (Millipore) and blocked in 5% nonfat milk in TBST for 1 hr at room temperature.

The membrane was rinsed with TBST and then incubated with primary antibody solution at 4 °C with gentle rocking overnight. The membrane was rinsed with TBST and then incubated with secondary antibody solutions at room temp with gentle rocking for 1-2 h. The membrane was then rinsed with wash buffer and visualized using the SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Scientific, Ref# 34577) and imaged on a Fluor Chem R (Protein Simple).

HA-tagged calprotectin proteins were detected using 1:1000 anti-HA [F-7] antibody (Santa Cruz Biotechnology, sc-7392), followed by 1:1000 anti-mouse HRP-conjugated antibody (Santa Cruz Biotechnology, sc-358914). Adhiron proteins were detected using 1:1000 mouse anti-c-Myc [9e10] antibody (Santa Cruz Biotechnology, sc-40), followed by 1:1000 anti-mouse HRP-conjugated antibody (Santa Cruz Biotechnology, sc-358914).

A.3.11 Dot blot

Protein solutions were prepared by taking purified HA-tagged calprotectin protein, purified c-Myc-tagged adhiron proteins, or commercially available protein (recombinant calprotectin, Biolegend, Cat# 753404; recombinant RAGE-Fc chimera protein, Biolegend, Cat# 769504) and diluting it to the indicated concentrations in PBS buffer.

Nitrocellulose membrane and filter paper were cut to size and wetted with TBS buffer for 10 min. Proteins were then applied to the nitrocellulose membrane using a 96-well Bio-Dot apparatus (BioRad) using gentle vacuum filtration. The membrane was then removed from the apparatus and blocked using fresh blocking solution (TBST, 2 mM $CaCl_2$, 5% milk) at room temp with gentle rocking for 1 h.

The membrane was rinsed with wash buffer (TBST, 2 mM $CaCl_2$), cut, and then incubated with primary antibody solution at 4 °C with gentle rocking overnight. The membrane was rinsed with wash buffer and then was incubated with secondary antibody solutions at room temp with gentle rocking for 2 h. The membrane was then rinsed with wash buffer and visualized using the SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Scientific, Ref# 34577) and imaged on a Fluor Chem R (Protein Simple).

Calprotectin proteins were detected using 1:10 mouse anti-MRP8 + MRP9 [27E10] antibody (Abcam, ab17050), followed by 1:1000 anti-mouse HRP-conjugated antibody (Santa Cruz Biotechnology, sc-358914). Adhiron proteins were detected using 1:1000 mouse anti-c-Myc [9e10] antibody (Santa Cruz Biotechnology, sc-40), followed by 1:1000 anti-mouse HRP-conjugated antibody (Santa Cruz Biotechnology, sc-358914).

Test conditions to detect a binding interaction between calprotectin and the evolved adhiron variants were performed by using a 10 µM solution (or indicated concentration) of purified c-Myc-tagged adhiron variant (or c-Myc-tagged wt adhiron as a negative control), followed by 1:1000 mouse anti-c-Myc [9e10] HRP-conjugated antibody (Santa Cruz Biotechnology, sc-40 HRP).

APPENDIX B

VECTORS

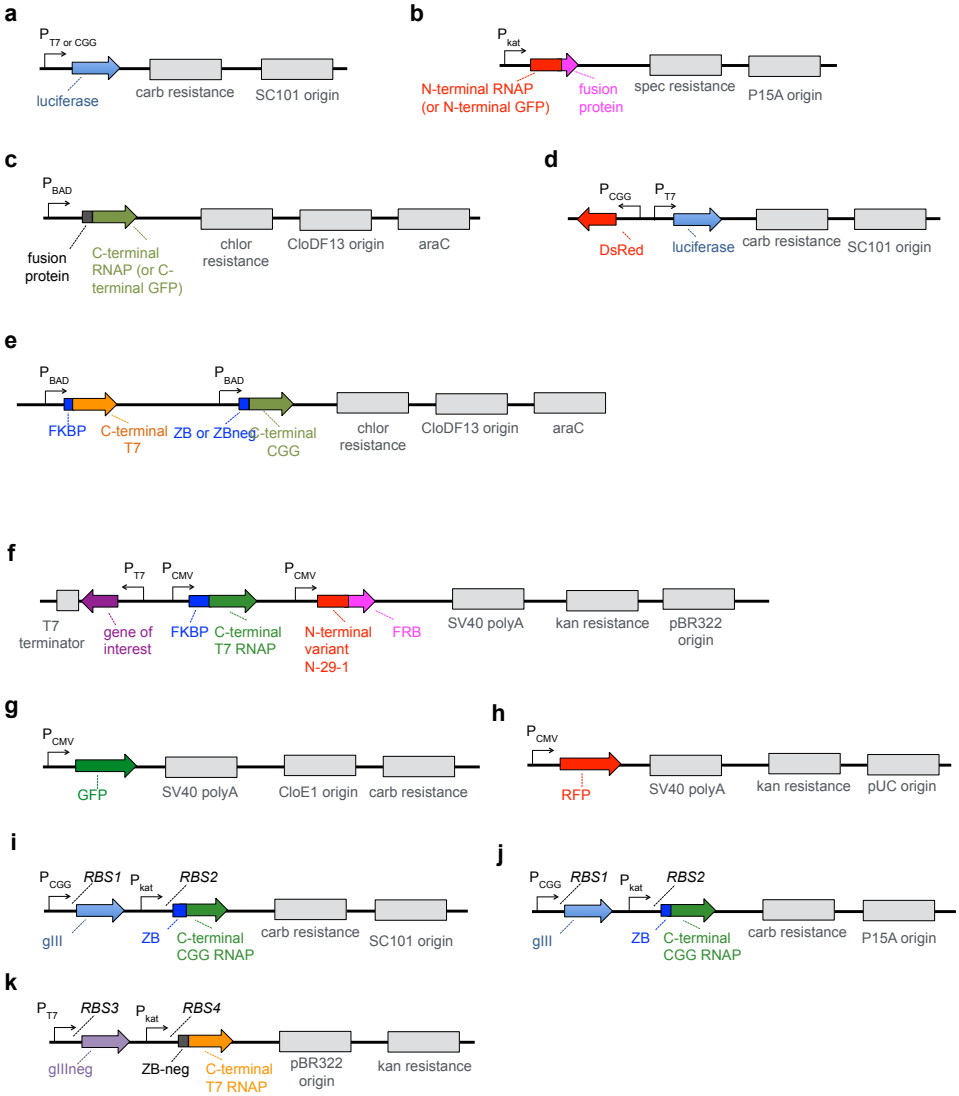

Figure B.1: Vector maps for all constructs used in the proximity-dependent split RNAP platform study. Vector maps corresponding to the vectors utilized in this study outlined in table B.1 and B.2.

Vector name	Antibiotic resistance	Origin	Purpose	Map
p2-22	carb/amp	SC101	pT7 luciferase reporter plasmid	a
pJin129	spec	P15a	Split N-terminal T7 RNAP expression plasmid	b
p3-7	spec	P15a	Split N-terminal T7 RNAP-linker-ZA expression plasmid	b
p2-55	chlr	CloDF13	Split C-terminal T7 RNAP expression plasmid	c
p2-39	chlr	CloDF13	ZB-linker-split C-terminal T7 RNAP expression plasmid	c
p2-57	chlr	CloDF13	ZBneg-linker-split C-terminal T7 RNAP expression plasmid	c
p5-71	spec	P15a	N-29-1 split N-terminal T7 RNAP variant-linker-ZA expression plasmid	b
p5-72	spec	P15a	N-29-8 split N-terminal T7 RNAP variant-linker-ZA expression plasmid	b
p5-74	spec	P15a	N-29-1 split N-terminal T7 RNAP variant-linker-ZA (I13I, L20I) expression plasmid	b
p5-75	spec	P15a	N-29-8 split N-terminal T7 RNAP variant-linker-ZA (I13I, L20I) expression plasmid	b
p5-79	spec	P15a	N-29-1 split N-terminal T7 RNAP variant-GGSGSGSS-FRB expression plasmid	b
p6-27	spec	P15a	N-29-1 split N-terminal T7 RNAP variant-GGSGSGSS-iLID expression plasmid	b
p6-29	chlr	CloDF13	SspB Nano-linker-Split C-terminal T7 RNAP expression plasmid	c
p5-39	chlr	CloDF13	FKBP-TSGGSG-Split C-terminal T7 RNAP expression plasmid	c
p5-40	chlr	CloDF13	FKBP-GG-Split C-terminal T7 RNAP expression plasmid	c
p7-68	chlr	CloDF13	FKBP-TSGGSGTSGGSG-Split C-terminal T7 RNAP expression plasmid	c
p5-70	spec	P15a	N-29-1 split N-terminal T7 RNAP variant-GG-FRB expression plasmid	b
p7-69	spec	P15a	N-29-1 split N-terminal T7 RNAP variant-GGSGSSGGSGSGSS-FRB expression plasmid	b
pJin200	spec	P15a	N-29-1 split N-terminal T7 RNAP variant-linker-ZA-linker-FRB expression plasmid	b
p3-13	chlr	CloDF13	ZB-linker-split C-terminal CGG RNAP expression plasmid	c
p4-32	chlr	CloDF13	ZB-linker-split C-terminal CGG RNAP expression plasmid	c
p2-64	carb/amp	SC101	pCGG luciferase reporter plasmid	a
pJin216	carb/amp	SC101	pT7 luciferase, pCGG RFP, reporter plasmid	d
pJin207	chlr	CloDF13	ZB-C-terminal CGG/FKBP-C-terminal T7 expression plasmid	e
pJin208	chlr	CloDF13	ZBneg-C-terminal CGG/FKBP-C-terminal T7 expression plasmid	e
pJin141	kan	pBR322	rapa-T7-F30-2xdBroccoli	f
p6-8	kan	pBR322	rapa-T7- mRNA(GFP)	f
pJin140	kan	pBR322	rapa-T7-shRNA(GFP)	f
p1-53	carb/amp	CloE1	eGFP expression vector	g
p3-62	kan	pUC	mRFP expression vector	h
p8-61	chlr	CloDF13	FKBP-linker-split C-terminal mEGFP (158+) expression plasmid	c
p8-62	spec	P15a	split N-terminal mEGFP (1-157)-linker-FRB expression plasmid	b
pJin210	chlr	CloDF13	FKBP-linker-split C-terminal scEGFP (158+) expression plasmid	c
pJin211	spec	P15a	split N-terminal scGFP (1-157)-linker-FRB expression plasmid	b

Table B.1: List of all non-PACE constructs used in the proximity-dependent split RNAP platform study. Vector maps for each construct type are shown in figure B.1

evolutionary date	positive AP					negative AP				
	Vector name	Map	origin	RBS1	RBS2	Vector name	Map	origin	RBS3	RBS4
day1	pJin69	i	SC101	SD8	sd8					
day2	pJin69	i	SC101	SD8	sd8					
day3	pJin69	i	SC101	SD8	sd8					
day4	pJin177	j	P15A	SD8	sd8	pJin173	k	pBR322	SD4	sd2
day5	pJin177	j	P15A	SD8	sd8	pJin173	k	pBR322	SD4	sd2
	pJin177	j	P15A	SD8	sd8	pJin172	k	pBR322	sd5	sd6
day6	pJin177	j	P15A	SD8	sd8	pJin172	k	pBR322	sd5	sd6
day7	pJin177	j	P15A	SD8	sd8	pJin172	k	pBR322	sd5	sd6
day8	pJin177	j	P15A	SD8	sd8	pJin104	k	pBR322	SD8	sd8
day9	pJin177	j	P15A	SD8	sd8	pJin104	k	pBR322	SD8	sd8
day10	pJin177	j	P15A	SD8	sd8	pJin104	k	pBR322	SD8	sd8
	pJin182	j	P15A	sd5	sd5	pJin172	k	pBR322	sd5	sd6
day11	pJin182	j	P15A	sd5	sd5	pJin172	k	pBR322	sd5	sd6
day12	pJin182	j	P15A	sd5	sd5	pJin172	k	pBR322	sd5	sd6
day13	pJin182	j	P15A	sd5	sd5	pJin172	k	pBR322	sd5	sd6
day14	pJin185	j	P15A	SD4	sd8	pJin194	k	pBR322	SD8	SD4
day15	pJin185	j	P15A	SD4	sd8	pJin194	k	pBR322	SD8	SD4
	pJin196	j	P15A	SD4	sd5	pJin194	k	pBR322	SD8	SD4
day16	pJin196	j	P15A	SD4	sd5	pJin194	k	pBR322	SD8	SD4
day17	pJin196	j	P15A	SD4	sd5	pJin194	k	pBR322	SD8	SD4
day18	pJin196	j	P15A	SD4	sd5	pJin194	k	pBR322	SD8	SD4
day19	pJin196	j	P15A	SD4	sd5	pJin194	k	pBR322	SD8	SD4
day20	pJin70	i	SC101	SD8	sd8	pJin104	k	pBR322	SD8	sd8
day21	pJin70	i	SC101	SD8	sd8	pJin104	k	pBR322	SD8	sd8
day22	pJin178	i	SC101	SD8	sd6	pJin172	k	pBR322	sd5	sd6
day23	pJin178	i	SC101	SD8	sd6	pJin172	k	pBR322	sd5	sd6
day24	pJin178	i	SC101	SD8	sd6	pJin172	k	pBR322	sd5	sd6
day25	pJin178	i	SC101	SD8	sd6	pJin172	k	pBR322	sd5	sd6
day26	pJin178	i	SC101	SD8	sd6	pJin172	k	pBR322	sd5	sd6
day27	pJin178	i	SC101	SD8	sd6	pJin172	k	pBR322	sd5	sd6
	pJin178	i	SC101	SD8	sd6	pJin171	k	pBR322	SD4	sd8
day28	pJin178	i	SC101	SD8	sd6	pJin172	k	pBR322	sd5	sd6
	pJin178	i	SC101	SD8	sd6	pJin171	k	pBR322	SD4	sd8
day29	pJin178	i	SC101	SD8	sd6	pJin171	k	pBR322	SD4	sd8

Table B.2: Full evolutionary protocol for PACE experiment to evolve proximity-dependent split RNAP. Vector names and details are provided for each day of PACE. Vector maps for the posAP and negAP vectors are shown in figure B.1i, j, and k. Two sets of posAP/negAP vectors listed on the same day indicates that a mixed selection pressure was utilized, in which two types of host cells, each containing one set of the posAP/negAP plasmids were added to a lagoon simultaneously. The relative RBS strengths were obtained from previous studies.[175]

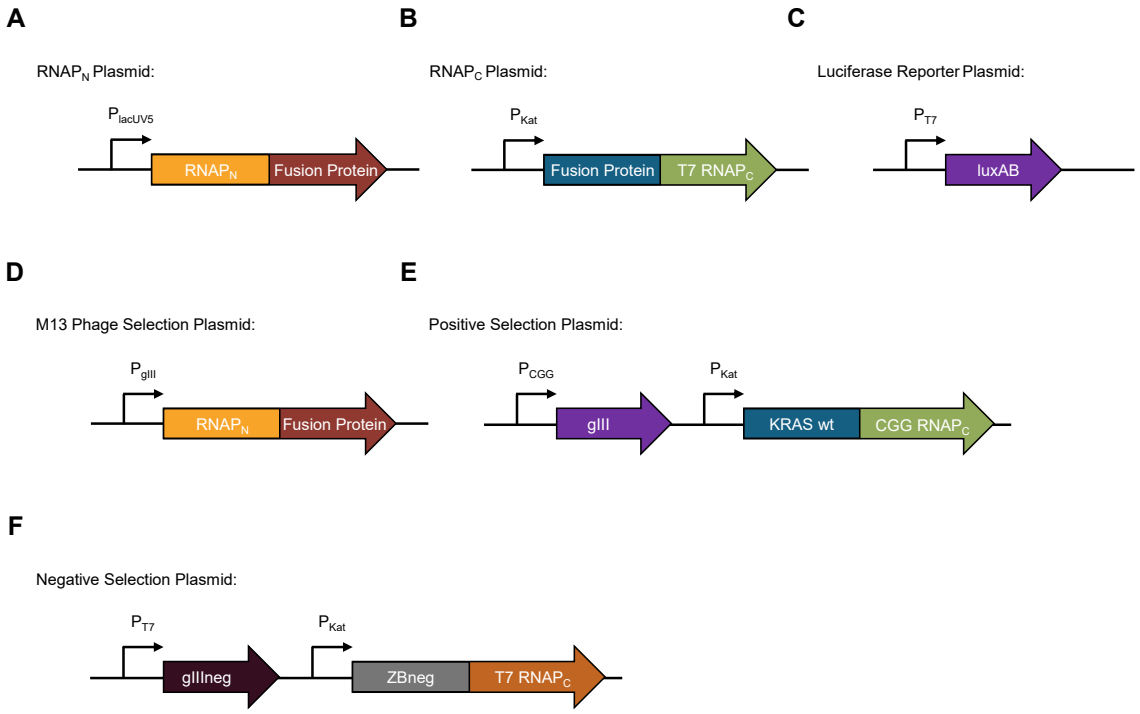

Figure B.2: Vector maps for all constructs used in the PACS DMS study. Vector maps corresponding to the vectors utilized in this study outlined in tables B.3, B.4, B.5, B.6, and B.7.

Label	Description	RNAP Fusion	Map	Link
31-a	SP, 29 1 N-terminal RNAP - linker - iZA	iZA	D	https://benchling.com/s/seq-9gmqhxn6yrBYyTl3jW5m
14-70	SP, 29-1 N-terminal RNAP - linker - RAF phage vector	RAF wt	D	https://benchling.com/s/seq-1070ExyDnJO7BrOnsgFS
22-13	SP, 29-1 N-terminal RNAP - linker - RAF K84A phage vector	RAF K84A	D	https://benchling.com/s/seq-io8RcEaSFhHUmya9nl63
22-14	SP, 29-1 N-terminal RNAP - linker - RAF Q66A phage vector	RAF Q66A	D	https://benchling.com/s/seq-HQfyNWTxI1Exc2eF0JxD
22-17	SP, 29-1 N-terminal RNAP - linker - RAF R89L phage vector	RAF R89L	D	https://benchling.com/s/seq-hut0nurkPZqjWb1tkvHb
22-18	SP, 29-1 N-terminal RNAP - linker - RAF K84E phage vector	RAF K84E	D	https://benchling.com/s/seq-iPkmnelfGAk7ur0voc5U
22-19	SP, 29-1 N-terminal RNAP - linker - RAF Q66K phage vector	RAF Q66K	D	https://benchling.com/s/seq-vQwTTzQ2D6bXpclvraPL

Table B.3: M13 bacteriophage plasmids used in PACS DMS study. Vector maps for each construct are shown in figure B.2.

Label	Description	RNAP Fusion	Map	Link
20-06	negAP, pBR322 origin T7 promoter sd5 gIIIneg, 2xVSV terminator, pKat promoter sd2 ZBneg - linker - T7 C terminal RNAP	ZBneg	F	https://benchling.com/s/seq-n6PTtqn0Aeu07HaoE1PM
31-69	posAP, p15a origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 (minus t) wt KRAS4b CDS - linker - CGG C terminal RNAP	wt KRAS4b	E	https://benchling.com/s/seq-PyAGyWTh5uzPjDkV07bP
MP6	MP6 *in vivo* mutagenesis plasmid[223]	n/a	n/a	https://www.addgene.org/69669/

Table B.4: PACS & PACE plasmids used in PACS DMS study. Vector maps for each construct are shown in figure B.2.

Table B.5: N-terminal expression plasmids for the luciferase assay used in PACS DMS study. Vector maps for each construct are shown in figure B.2.

Label	Description	RNAP Fusion	Map	Link
16-77	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS	RAF wt	A	https://benchling.com/s/seq-iTjydEb4awl5YJcCYFsE
22-20	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS R89L mutant	RAF R89L	A	https://benchling.com/s/seq-zBhVBLIF5IKLZBS5t1Ae
23-36	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS K84E mutant	RAF K84E	A	https://benchling.com/s/seq-Mn17whd6Vuzw7jNq10xJ
23-37	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS Q66K mutant	RAF Q66K	A	https://benchling.com/s/seq-ClFOP1vMR69aAebeYuxR
23-38	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS K84A mutant	RAF K84A	A	https://benchling.com/s/seq-4SHbK7SAEitV6Mof6tSA
23-39	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS Q66A mutant	RAF Q66A	A	https://benchling.com/s/seq-CxKVCQIaJnMVT6kHkRz4
34-45	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS L78P mutant	RAF L78P	A	https://benchling.com/s/seq-s8S6FgpqtXJgarqcF6pe
34-46	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS F61L mutant	RAF F61L	A	https://benchling.com/s/seq-DwtU3g1IAgFTmxIX4KK1

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
34-47	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS K84R mutant	RAF K84R	A	https://benchling.com/s/seq-V0nPFA7D1yUk8ER8smU1
34-48	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS L82P mutant	RAF L82P	A	n/a
34-49	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS L62P mutant	RAF L62P	A	n/a
34-54	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS H79R mutant	RAF H79R	A	n/a
34-55	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, N71K, H79R mutant	RAF S52R, N71K, H79R	A	https://benchling.com/s/seq-WckwAjgu9nK83K2ZfLgQ
34-56	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, H79R mutant	RAF S52R, H79R	A	n/a
34-57	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS N71K mutant	RAF N71K	A	https://benchling.com/s/seq-RBSilVp1k1ukEUX2QPak
34-58	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS N71K, H79R mutant	RAF N71K, H79R	A	n/a

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
34-59	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R mutant	RAF S52R	A	n/a
34-60	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, N71K mutant	RAF S52R, N71K	A	n/a
34-61	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, R67H mutant	RAF S52R, R67H	A	n/a
34-62	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS R67H mutant	RAF R67H	A	https://benchling.com/s/seq-66rs2z2Sw6Fgp1skWQU3
35-54	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS T54I mutant	RAF T54I	A	n/a
35-55	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, C81S mutant	RAF S52R, C81S	A	n/a
35-56	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS H79L mutant	RAF H79L	A	n/a
35-57	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS N71I mutant	RAF N71I	A	n/a

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
35-73	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS G75S mutant	RAF G75S	A	n/a
35-74	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS M76K mutant	RAF M76K	A	n/a
35-75	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS M76T mutant	RAF M76T	A	n/a
35-76	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS G75D mutant	RAF G75D	A	n/a
35-77	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS M76V mutant	RAF M76V	A	n/a
35-78	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS T68S mutant	RAF T68S	A	n/a
35-79	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS K87R mutant	RAF K87R	A	n/a
35-80	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS N56S mutant	RAF N56S	A	n/a

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
35-81	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS M76L mutant	RAF M76L	A	n/a
36-01	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS C81S mutant	RAF C81S	A	n/a
36-02	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, G75S mutant	RAF S52R, G75S	A	n/a
36-04	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, M76T mutant	RAF S52R, M76T	A	n/a
36-05	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, G75D mutant	RAF S52R, G75D	A	n/a
36-06	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, M76V mutant	RAF S52R, M76V	A	n/a
36-07	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, T68S mutant	RAF S52R, T68S	A	n/a
36-10	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, M76L mutant	RAF S52R, M76L	A	n/a

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
180607 LAD1-11	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS	RAF wt	A	n/a
180607 LAD1-15	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS A85A mutant	RAF A85A	A	n/a
180607 LBD1-11	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS L102P mutant	RAF L102P	A	n/a
180607 LBD1-12	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS L102P, A110T mutant	RAF L102P, A110T	A	n/a
180607 LCD1-11	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS Q66N mutant	RAF Q66N	A	n/a
180607 LCD1-14	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS Q66E mutant	RAF Q66E	A	n/a
190626 LA-2	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS R111R mutant	RAF R111R	A	n/a
190626 LA-4	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS	RAF wt	A	n/a
190626 LB-2	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS	RAF wt	A	n/a

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
190626 LB-8	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS	RAF wt	A	n/a
190626 LC-2	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS Q66N, A118S mutant	RAF Q66N, A118S	A	n/a
190626 LC-6	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS Q66N mutant	RAF Q66N	A	n/a
gd-1.3	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS R67H, M76T, L78L, G90G mutant	RAF R67H, M76T, L78L, G90G	A	n/a
gd-1.4	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, R73R mutant	RAF S52R, R73R	A	n/a
gd-2.5	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS N64N, N71K, C81S mutant	RAF N64N, N71K, C81S	A	n/a
gd-2.8	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS N56S mutant	RAF N56S	A	n/a
gd-3.5	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS	RAF wt	A	n/a

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
gd-3.6	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS G90G mutant	RAF G90G	A	n/a
Lib.2	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, T57A mutant	RAF S52R, T57A	A	n/a
lib.3	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R mutant	RAF S52R	A	n/a
lib.5	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS K84R, L86L mutant	RAF K84R, L86L	A	n/a
lib.7	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS K53M mutant	RAF K53M	A	n/a
Lib.b	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS H79R, C81R mutant	RAF H79R, C81R	A	n/a
Lib.c	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS S52R, I58I mutant	RAF S52R, I58I	A	n/a
Lib.f	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS T68S, V70D mutant	RAF T68S, V70D	A	n/a

Continued on next page

Table B.5 - continued from previous page

Label	Description	RNAP Fusion	Map	Link
Lib.h	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS K65* mutant	RAF K65*	A	n/a
wt-2.4	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS G75S mutant	RAF G75S	A	n/a
wt-3.2	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - RAF (52-131) CDS R67R mutant	RAF R67R	A	n/a

Label	Description	RNAP Fusion	Map	Link
16-27	p15A origin pKat wt KRAS4b CDS - linker - C-terminal T7RNAP	wt KRAS4b	B	https://benchling.com/s/seq-fFAPfbngpNpuXRTwF8IQ
16-32	p15A origin pKat ZB peptide - linker - C-terminal T7RNAP	ZB	B	https://benchling.com/s/seq-LibSN1MPXN3ALXwzQ3Eq
16-33	p15A origin pKat C-terminal T7RNAP (no RNAP Protein Fusion)	no fusion	B	https://benchling.com/s/seq-G4klSsHqCaR94H04blbC

Table B.6: C-terminal expression plasmids for the luciferase assay used in PACS DMS study. Vector maps for each construct are shown in figure B.2.

Label	Description	RNAP Fusion	Map	Link
2-22	pJC173b - AP pSC101 pT7 SD8 RBS luxAB	n/a	C	https://benchling.com/s/seq-ARzLVz9yIlqDti0VsbiG

Table B.7: Reporter plasmids for the luciferase assay used in PACS DMS study. Vector maps for each construct are shown in figure B.2.

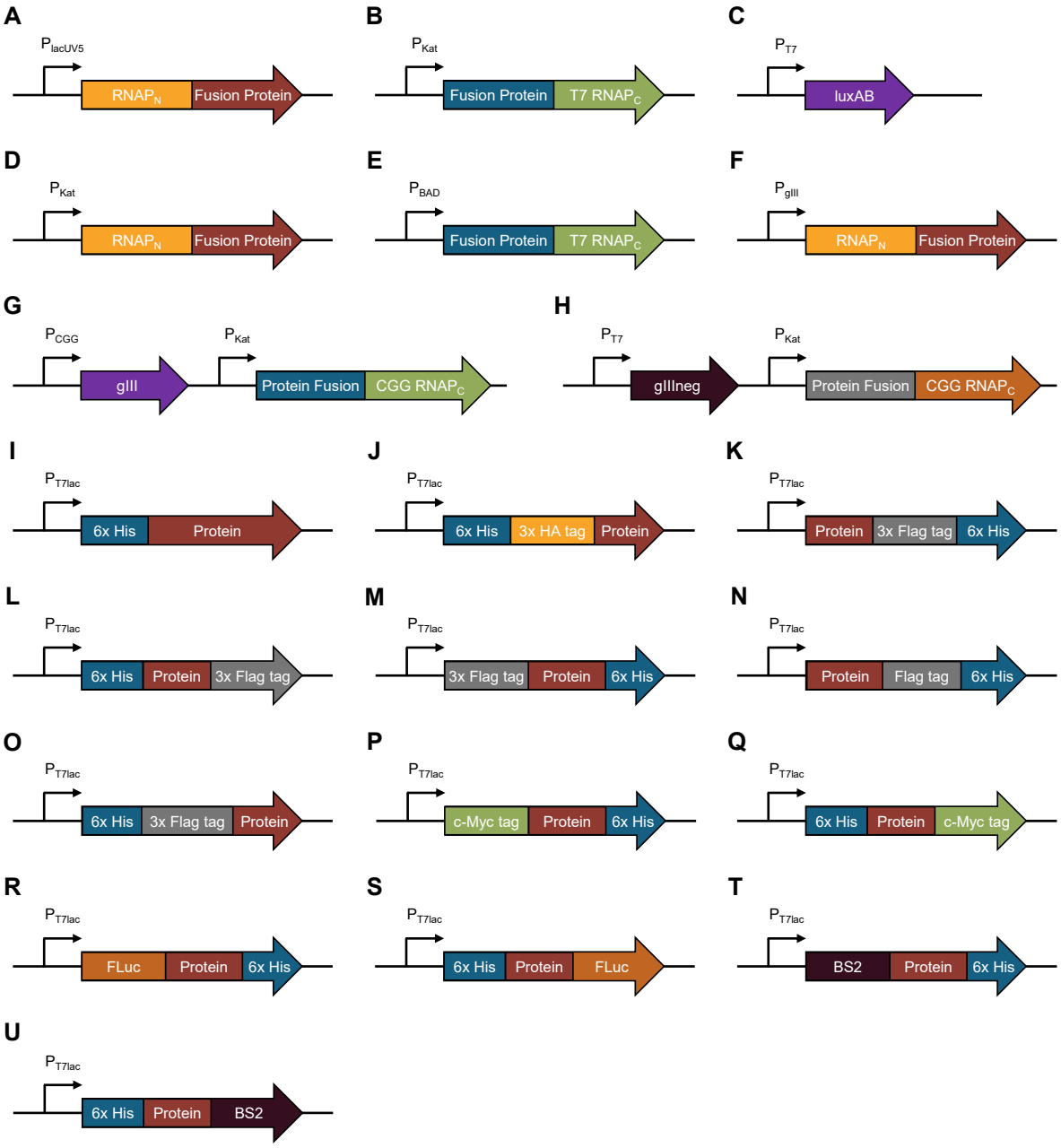

Figure B.3: Vector maps for all cloned constructs for the *de novo* calprotectin binder study. Vector maps corresponding to the vectors utilized in this study outlined in tables B.8, B.9, B.10, B.11, B.12, and B.13.

Label	Description	Protein Fusion	Map
13-52	p15A origin pKat 29-1 PACE N-terminal T7RNAP variant - linker - Adhiron variant	Adhiron variant	D
13-53	p15A origin pKat 29-1 PACE N-terminal T7RNAP variant - linker - Fynomer variant	Fynomer variant	D
13-54	p15A origin pKat 29-1 PACE N-terminal T7RNAP variant - linker - BC2 nanobody without BamHI cutsite	BC2 nanobody	D
16-75	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant (no RNAP Protein Fusion)	no fusion	A
16-76	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - isoleucine ZA peptide	iZA	A
25-13	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - fynomer E12M, R14K, E16M, L19L, N31H, S33S variant	Fynomer variant	A
25-14	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - adhiron	wt Adhiron	A
25-15	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - nanobody	BC2 nanobody	A
28-63	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - adhiron variant (E36E, Loop1 GSCMRSNMT, K71K, L72L, Loop2 VEARYRGCR)	Adhiron variant	A
28-64	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - adhiron variant (E36E, Loop1 ILVKLNNII, K71K, L72L, Loop2 KAEEGAFSD)	Adhiron variant	A
28-65	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - adhiron variant (E36E, Loop1 KSREDRDQG, K71K, L72L, Loop2 VSRKWGYNS)	Adhiron variant	A
40-18	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - adhiron variant (E36E, Loop1 DIECCSGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV)	Adhiron variant	A
40-50	pCloDF13 origin UV5 IPTG inducible promoter PACE 29-1 N-terminal T7RNAP variant - linker - adhiron variant (E36E, Loop1 DIECCSGFW, K71K, L72L, Loop2 SGCMFHECV)	Adhiron variant	A

Table B.8: N-terminal expression plasmids cloned for the luciferase assay for the *de novo* calprotectin binder study. Vector maps for each construct are shown in figure B.3.

Label	Description	Protein Fusion	Map
16-32	p15A origin pKat ZB peptide - linker - C-terminal T7RNAP	ZB	B
16-33	p15A origin pKat C-terminal T7RNAP (no RNAP Protein Fusion)	no fusion	B
26-81	pCloDF13 origin PBAD calprotectin heterodimer (s100a9 - 6aa linker - s100a8) - linker C-terminal T7RNAP	calprotectin fusion (6aa linker)	E
27-20	pCloDF13 origin PBAD calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker C-terminal T7RNAP	calprotectin fusion (12aa linker)	E
27-21	p15A origin pKat calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - C-terminal T7RNAP	calprotectin fusion (12aa linker)	B
27-33	p15A origin pKat calprotectin heterodimer (s100a9 - 6aa linker - s100a8) - linker - C-terminal T7RNAP	calprotectin fusion (6aa linker)	B
28-61	p15A origin pKat s100a8 - linker - C-terminal T7RNAP	s100a8	B
28-62	p15A origin pKat s100a9 - linker - C-terminal T7RNAP	s100a9	B
29-24	p15A origin pKat s100a8 homodimer (12aa linker) - linker - C-terminal T7RNAP	s100a8 fusion (12aa linker)	B
29-25	p15A origin pKat s100a9 homodimer (12aa linker)- linker - C-terminal T7RNAP	s100a9 fusion (12aa linker)	B

Table B.9: C-terminal expression plasmids cloned for the luciferase assay for the *de novo* calprotectin binder study. Vector maps for each construct are shown in figure B.3.

Label	Description	Protein Fusion	Map
2-22	pJC173b - AP pSC101 pT7 SD8 RBS luxAB	n/a	C

Table B.10: Reporter plasmids cloned for the luciferase assay for the *de novo* calprotectin binder study. Vector maps for each construct are shown in figure B.3.

Label	Description	Protein Fusion	Map
11-40	SP, 29-1 N-terminal RNAP - linker - fynomer phage vector	wt Fynomer	F
11-41	SP, 29-1 N-terminal RNAP - linker - BC2 nanobody phage vector	BC2 nanobody	F
12-42	SP, Day 29-1 N-terminal RNAP - linker - Adhiron phage vector	wt Adhiron	F

Table B.11: Phage plasmids cloned for the *de novo* calprotectin binder study. Vector maps for each construct are shown in figure B.3.

Table B.12: PACE plasmids cloned for the *de novo* calprotectin binder study. Vector maps for each construct are shown in figure B.3.

Label	Description	Protein Fusion	Map
20-01	neg AP, pBR322 origin T7 promoter sd8 gIIIneg, 2xVSV terminator, pKat promoter sd8 ZBneg - linker - T7 C terminal RNAP	Zbneg	H
20-02	neg AP, pBR322 origin T7 promoter sd8 gIIIneg, 2xVSV terminator, pKat promoter sd6 ZBneg - linker - T7 C terminal RNAP	Zbneg	H
20-03	neg AP, pBR322 origin T7 promoter sd8 gIIIneg, 2xVSV terminator, pKat promoter sd5 ZBneg - linker - T7 C terminal RNAP	Zbneg	H
20-04	neg AP, pBR322 origin T7 promoter sd8 gIIIneg, 2xVSV terminator, pKat promoter sd2 ZBneg - linker - T7 C terminal RNAP	Zbneg	H
20-05	neg AP, pBR322 origin T7 promoter sd5 gIIIneg, 2xVSV terminator, pKat promoter sd5 ZBneg - linker - T7 C terminal RNAP	Zbneg	H
20-06	neg AP, pBR322 origin T7 promoter sd5 gIIIneg, 2xVSV terminator, pKat promoter sd2 ZBneg - linker - T7 C terminal RNAP	Zbneg	H
25-31	p15a origin, pCGG SD8 gIII, 2xVSV terminator, pKat SD8 (minus t) calprotectin heterodimer (s100a9 - 6aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (6aa linker)	G
25-32	pSC101 origin, pCGG SD8 gIII, 2xVSV terminator, pKat SD8 calprotectin heterodimer (s100a9 - 6aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (6aa linker)	G
25-33	p15a origin, pCGG SD8 gIII, 2xVSV terminator, pKat SD8 (minus t) s100a8 (monomer of calprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a8	G
25-34	pSC101 origin, pCGG SD8 gIII, 2xVSV terminator, pKat SD8 s100a8 (monomer of calprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a8	G
25-35	p15a origin, pCGG SD8 gIII, 2xVSV terminator, pKat SD8 (minus t) s100a9 (monomer of calprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a9	G
25-36	pSc101 origin, pCGG SD8 gIII, 2xVSV terminator, pKat SD8 s100a9 (monomer of calprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a9	G
26-71	negAP, pBR322, terminator-ZBneg neg C term T7-SD8-sd5	Zbneg	H
26-72	negAP, pBR322, terminator-ZBneg neg C term T7-SD8-sd6	Zbneg	H

Continued on next page

Table B.12 - continued from previous page

Label	Description	Protein Fusion	Map
27-19	p15a origin, pCGG SD8 gIII, 2xVSV terminator, pKat SD8 (minus t) calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
27-22	p15a origin, pCGG SD8 gIII, pKat SD8 (minus t) calprotectin heterodimer (s100a9 - 6aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (6aa linker)	G
27-49	p15a origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 (minus t) calprotectin heterodimer (s100a9 - 6aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (6aa linker)	G
27-50	pSC101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 calprotectin heterodimer (s100a9 - 6aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (6aa linker)	G
27-51	p15a origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 (minus t) s100a8 (monomer of calprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a8	G
27-52	pSC101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 s100a8 (monomer of calprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a8	G
27-53	p15a origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 (minus t) s100a9 (monomer of claprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a9	G
27-54	pSc101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 s100a9 (monomer of calprotectin heterodimer) - linker - CGG C terminal RNAP, posAP	s100a9	G
27-55	p15a origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 (minus t) calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
28-41	pSC101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
28-80	pSc101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD4 Calprotectin (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
28-81	pSc101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat sd8 Calprotectin (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
29-01	pSc101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat sd6 Calprotectin (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
29-02	pSc101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat sd5 Calprotectin (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G

Continued on next page

Table B.12 - continued from previous page

Label	Description	Protein Fusion	Map
29-26	p15a origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 (minus t) s100a8 homodimer (12aa linker) - linker - CGG C terminal RNAP, posAP	s100a8 fusion (12aa linker)	G
29-27	pSC101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 s100a8 homodimer (12aa linker) - linker - CGG C terminal RNAP, posAP	s100a8 fusion (12aa linker)	G
29-28	p15a origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 (minus t) s100a9 homodimer (12aa linker) - linker - CGG C terminal RNAP, posAP	s100a9 fusion (12aa linker)	G
29-29	pSc101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat SD8 s100a9 homodimer (12aa linker) - linker - CGG C terminal RNAP, posAP	s100a9 fusion (12aa linker)	G
29-39	negAP, pBR322, terminator-ZBneg neg C term T7-SD8-sd8	Zbneg	H
30-21	negAP, pBR322, terminator-ZBneg neg C term T7-SD8-SD4	Zbneg	H
30-22	negAP, pBR322, terminator-ZBneg neg C term T7-SD8-SD8	Zbneg	H
31-04	pSc101 origin, pCGG SD8 recoded gIII, 2xVSV terminator, pKat sd2 calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
31-05	pSc101 origin, pCGG sd8 recoded gIII, 2xVSV terminator, pKat sd8 calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
31-06	pSc101 origin, pCGG sd8 recoded gIII, 2xVSV terminator, pKat sd6 calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
31-07	pSc101 origin, pCGG sd7 recoded gIII, 2xVSV terminator, pKat sd6 calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
31-08	pSc101 origin, pCGG sd8 recoded gIII, 2xVSV terminator, pKat sd5 calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
31-09	pSc101 origin, pCGG sd8 recoded gIII, 2xVSV terminator, pKat sd2 calprotectin heterodimer (s100a9 - 12aa linker - s100a8) - linker - CGG C terminal RNAP, posAP	calprotectin fusion (12aa linker)	G
31-10	neg AP, pBR322 origin T7 promoter SD8 gIIIneg, 2xVSV terminator, pKat promoter sd5 s100a8 homodimer (12aa linker) - linker - T7 C terminal RNAP	s100a8 fusion (12aa linker)	H
31-11	neg AP, pBR322 origin T7 promoter SD8 gIIIneg, 2xVSV terminator, pKat promoter sd8 s100a8 homodimer (12aa linker) - linker - T7 C terminal RNAP	s100a8 fusion (12aa linker)	H

Continued on next page

Table B.12 - continued from previous page

Label	Description	Protein Fusion	Map
31-12	neg AP, pBR322 origin T7 promoter SD8 gIIIneg, 2xVSV terminator, pKat promoter sd5 s100a9 homodimer (12aa linker) - linker - T7 C terminal RNAP	s100a9 fusion (12aa linker)	H
31-13	neg AP, pBR322 origin T7 promoter SD8 gIIIneg, 2xVSV terminator, pKat promoter sd8 s100a9 homodimer (12aa linker) - linker - T7 C terminal RNAP	s100a9 fusion (12aa linker)	H
Jin408	negAP, pBR322, 2xVSV terminator, ZBneg-C term T7-sd5-sd6	Zbneg	H
Jin409	negAP, pBR322, 2xVSV terminator, ZBnegC term T7-SD4-sd8	Zbneg	H

Table B.13: Protein expression plasmids cloned for the *de novo* calprotectin binder study. Vector maps for each construct are shown in figure B.3.

Label	Description	Protein Fusion	Map
28-58	pET-30a 6xHis-Calprotectin (s100a9 - 12aa linker - s100a8)	calprotectin fusion (12aa linker)	I
28-59	pET-30a 6xHis-s100a8	s100a8	I
28-60	pET-30a 6xHis-s100a9	s100a9	I
29-56	pET-30a 6xHis-3xHA-Calprotectin (s100a9 - 12aa linker - s100a8)	calprotectin fusion (12aa linker)	J
29-57	pET-30a 6xHis-3xHA-s100a8 homodimer (12aa linker)	s100a8 fusion (12aa linker)	J
29-58	pET-30a 6xHis-3xHA-s100a9 homodimer (12aa linker)	s100a9 fusion (12aa linker)	J
28-53*	pET-30a Adhiron-3xFlag-6xHis	wt Adhiron	K
28-54*	pET-30a Adhiron variant (E36E, Loop1 DIECCSGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) -3xFlag-6xHis	Adhiron variant	K
30-10*	pET-30a 6xHis - Adhiron variant (E36E, Loop1 DIECCSGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - 3xFlag	Adhiron variant	L
30-11*	pET-30a 6xHis - Adhiron - 3xFlag	wt Adhiron	L
30-12*	pET-30a 3xFlag - Adhiron variant (E36E, Loop1 DIECC-SGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - 6xHis	Adhiron variant	M
30-13*	pET-30a 3xFlag - Adhiron - 6xHis	wt Adhiron	M

Continued on next page

∗. Flag-tagged and 3x flag-tagged proteins were undetectable alongside a 3x flag-tagged positive control using using 1:1000 mouse anti-flag [FGR4] HRP-conjugated antibody (Invitrogen, MA1-91878-HRP).

Table B.13 - continued from previous page

Label	Description	Protein Fusion	Map
30-14*	pET-30a Adhiron variant (E36E, Loop1 DIECCSGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) -Flag-6xHis	Adhiron variant	N
30-15*	pET-30a Adhiron-Flag-6xHis	wt Adhiron	N
30-16*	pET-30a 6xHis - 3xFlag - Adhiron variant (E36E, Loop1 DIECCSGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV)	Adhiron variant	O
30-17*	pET-30a 6xHis-3xFlag-Adhiron	wt Adhiron	O
31-35	pET-30a c-Myc tag- Adhiron variant (E36E, Loop1 DIECC-SGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - 6xHis	Adhiron variant	P
31-36	pET-30a c-Myc tag - Adhiron - 6xHis	wt Adhiron	P
31-37	pET-30a 6xHis - Adhiron variant (E36E, Loop1 DIECC-SGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - c-Myc tag	Adhiron variant	Q
31-38	pET-30a 6xHis - Adhiron - c-Myc tag	wt Adhiron	Q
31-39	pET-30a FLuc - Adhiron variant (E36E, Loop1 DIECC-SGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - 6xHis	Adhiron variant	R
31-40	pET-30a FLuc - Adhiron - 6xHis	wt Adhiron	R
31-41	pET-30a 6xHis - Adhiron variant (E36E, Loop1 DIECC-SGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - FLuc	Adhiron variant	S
31-42	pET-30a 6xHis - Adhiron - FLuc	wt Adhiron	S
31-43	pET-30a BS2 - Adhiron variant (E36E, Loop1 DIECC-SGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - 6xHis	Adhiron variant	T
31-44	pET-30a BS2 - Adhiron - 6xHis	wt Adhiron	T
31-45	pET-30a 6xHis - Adhiron variant (E36E, Loop1 DIECC-SGFW, Y59C, K71K, L72L, Loop2 SGCMFHECV) - BS2	Adhiron variant	U
31-46	pET-30a 6xHis - Adhiron - BS2	wt Adhiron	U

REFERENCES

1. Braun, P.; Gingras, A.-C. History of protein-protein interactions: From egg-white to complex networks. *PROTEOMICS* **2012**, *12*, 1478–1498, DOI: 10.1002/pmic.201100563.

2. Vidal, M.; Cusick, M. E.; Barabási, A.-L. Interactome networks and human disease. *Cell* **2011**, *144*, 986–998, DOI: 10.1016/j.cell.2011.02.016.

3. Sahni, N.; Yi, S.; Taipale, M.; Fuxman Bass, J. I.; Coulombe-Huntington, J. et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **2015**, *161*, 647–660, DOI: 10.1016/j.cell.2015.04.013.

4. Jubb, H. C.; Pandurangan, A. P.; Turner, M. A.; Ochoa-Montaño, B.; Blundell, T. L. et al. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Progress in Biophysics and Molecular Biology* **2017**, *128*, 3–13, DOI: 10.1016/j.pbiomolbio.2016.10.002.

5. Laddach, A.; Ng, J. C.-F.; Chung, S. S.; Fraternali, F. Genetic variants and protein-protein interactions: a multidimensional network-centric view. *Current Opinion in Structural Biology* **2018**, *50*, 82–90, DOI: 10.1016/j.sbi.2017.12.006.

6. Ngounou Wetie, A. G.; Sokolowska, I.; Woods, A. G.; Roy, U.; Loo, J. A. et al. Investigation of stable and transient protein-protein interactions: Past, present, and future. *Proteomics* **2013**, *13*, 538–557, DOI: 10.1002/pmic.201200328.

7. Ngounou Wetie, A. G.; Sokolowska, I.; Woods, A. G.; Roy, U.; Deinhardt, K. et al. Protein-protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. *Cellular and Molecular Life Sciences* **2014**, *71*, 205–228, DOI: 10.1007/s00018-013-1333-1.

8. Wang, B.; Kitney, R. I.; Joly, N.; Buck, M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications* **2011**, *2*, 508, DOI: 10.1038/ncomms1516.

9. Rogers, J. K.; Taylor, N. D.; Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Current Opinion in Biotechnology* **2016**, *42*, 84–91, DOI: 10.1016/j.copbio.2016.03.005.

10. Rogers, J. K.; Church, G. M. Genetically encoded sensors enable real-time observation of metabolite production. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 2388–2393, DOI: 10.1073/pnas.1600375113.

11. Zhang, J.; Jensen, M. K.; Keasling, J. D. Development of biosensors and their application in metabolic engineering. *Current Opinion in Chemical Biology* **2015**, *28*, 1–8, DOI: 10.1016/j.cbpa.2015.05.013.

12. Eggeling, L.; Bott, M.; Marienhagen, J. Novel screening methods–biosensors. *Current Opinion in Biotechnology* **2015**, *35*, 30–36, DOI: 10.1016/j.copbio.2014.12.021.

13. Kitada, T.; DiAndreth, B.; Teague, B.; Weiss, R. Programming gene and engineered-cell therapies with synthetic biology. *Science* **2018**, *359*, DOI: 10.1126/science.aad1067.

14. Turner, A. P. F. Biosensors: sense and sensibility. *Chemical Society Reviews* **2013**, *42*, 3184–3196, DOI: 10.1039/c3cs35528d.

15. Kirsch, J.; Siltanen, C.; Zhou, Q.; Revzin, A.; Simonian, A. Biosensor technology: recent advances in threat agent detection and medicine. *Chemical Society Reviews* **2013**, *42*, 8733–8768, DOI: 10.1039/c3cs60141b.

16. Way, J. C.; Collins, J. J.; Keasling, J. D.; Silver, P. A. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell* **2014**, *157*, 151–161, DOI: 10.1016/j.cell.2014.02.039.

17. Gardner, T. S. Synthetic biology: from hype to impact. *Trends in Biotechnology* **2013**, *31*, 123–125, DOI: 10.1016/j.tibtech.2013.01.018.

18. Carpenter, A.; Paulsen, I.; Williams, T. Blueprints for Biosensors: Design, Limitations, and Applications. *Genes* **2018**, *9*, 375, DOI: 10.3390/genes9080375.

19. Kaczmarski, J. A.; Mitchell, J. A.; Spence, M. A.; Vongsouthi, V.; Jackson, C. J. Structural and evolutionary approaches to the design and optimization of fluorescence-based small molecule biosensors. *Current Opinion in Structural Biology* **2019**, *57*, 31–38, DOI: 10.1016/j.sbi.2019.01.013.

20. Müller, S.; Strohbach, D.; Wolf, J. Sensors made of RNA: tailored ribozymes for detection of small organic molecules, metals, nucleic acids and proteins. *IEE Proceedings. Nanobiotechnology* **2006**, *153*, 31–40, DOI: 10.1049/ip-nbt:20050047.

21. Shcherbakova, D. M.; Shemetov, A. A.; Kaberniuk, A. A.; Verkhusha, V. V. Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools. *Annual Review of Biochemistry* **2015**, *84*, 519–550, DOI: 10.1146/annurev-biochem-060614-034411.

22. Rogers, J. K.; Guzman, C. D.; Taylor, N. D.; Raman, S.; Anderson, K. et al. Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Research* **2015**, *43*, 7648–7660, DOI: 10.1093/nar/gkv616.

23. Arkin, A. P. A wise consistency: engineering biology for conformity, reliability, predictability. *Current Opinion in Chemical Biology* **2013**, *17*, 893–901, DOI: 10.1016/j.cbpa.2013.09.012.

24. Bhalla, N.; Jolly, P.; Formisano, N.; Estrela, P. Introduction to biosensors. *Essays in Biochemistry* **2016**, *60*, 1–8, DOI: 10.1042/EBC20150001.

25. Lutz, R.; Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research* **1997**, *25*, 1203–1210, DOI: 10.1093/nar/25.6.1203.

26. Feng, J.; Jester, B. W.; Tinberg, C. E.; Mandell, D. J.; Antunes, M. S. et al. A general strategy to construct small molecule biosensors in eukaryotes. *eLife* **2015**, *4*, e10606, DOI: 10.7554/eLife.10606.

27. Juárez, J. F.; Lecube-Azpeitia, B.; Brown, S. L.; Johnston, C. D.; Church, G. M. Biosensor libraries harness large classes of binding domains for construction of allosteric transcriptional regulators. *Nature Communications* **2018**, *9*, 3101, DOI: 10.1038/s41467-018-05525-6.

28. Jester, B. W.; Tinberg, C. E.; Rich, M. S.; Baker, D.; Fields, S. Engineered Biosensors from Dimeric Ligand-Binding Domains. *ACS Synthetic Biology* **2018**, *7*, 2457–2467, DOI: 10.1021/acssynbio.8b00242.

29. Yang, Y.; Lin, Y.; Wang, J.; Wu, Y.; Zhang, R. et al. Sensor-regulator and RNAi based bifunctional dynamic control network for engineered microbial synthesis. *Nature Communications* **2018**, *9*, 3043, DOI: 10.1038/s41467-018-05466-0.

30. Kenyon, C. J.; Walker, G. C. DNA-damaging agents stimulate gene expression at specific loci in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **1980**, *77*, 2819–2823, DOI: 10.1073/pnas.77.5.2819.

31. Shi, S.; Choi, Y. W.; Zhao, H.; Tan, M. H.; Ang, E. L. Discovery and engineering of a 1-butanol biosensor in Saccharomyces cerevisiae. *Bioresource Technology* **2017**, *245*, 1343–1351, DOI: 10.1016/j.biortech.2017.06.114.

32. Dahl, R. H.; Zhang, F.; Alonso-Gutierrez, J.; Baidoo, E.; Batth, T. S. et al. Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotechnology* **2013**, *31*, 1039–1046, DOI: 10.1038/nbt.2689.

33. Wang, M.; Li, S.; Zhao, H. Design and engineering of intracellular-metabolite-sensing/regulation gene circuits in Saccharomyces cerevisiae. *Biotechnology and Bioengineering* **2016**, *113*, 206–215, DOI: 10.1002/bit.25676.

34. Mahr, R.; Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Applied Microbiology and Biotechnology* **2016**, *100*, 79–90, DOI: 10.1007/s00253-015-7090-3.

35. Cheng, F.; Tang, X.-L.; Kardashliev, T. Transcription Factor-Based Biosensors in High-Throughput Screening: Advances and Applications. *Biotechnology Journal* **2018**, *13*, e1700648, DOI: 10.1002/biot.201700648.

36. Schmidt, C. M.; Smolke, C. D. RNA Switches for Synthetic Biology. *Cold Spring Harbor Perspectives in Biology* **2019**, *11*, a032532, DOI: 10.1101/cshperspect.a032532.

37. Puton, T.; Kozlowski, L. P.; Rother, K. M.; Bujnicki, J. M. CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Research* **2013**, *41*, 4307–4323, DOI: 10.1093/nar/gkt101.

38. Domin, G.; Findeiß, S.; Wachsmuth, M.; Will, S.; Stadler, P. F. et al. Applicability of a computational design approach for synthetic riboswitches. *Nucleic Acids Research* **2017**, *45*, 4108–4119, DOI: 10.1093/nar/gkw1267.

39. Espah Borujeni, A.; Mishler, D. M.; Wang, J.; Huso, W.; Salis, H. M. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Research* **2016**, *44*, 1–13, DOI: 10.1093/nar/gkv1289.

40. Beisel, C. L.; Bayer, T. S.; Hoff, K. G.; Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Molecular Systems Biology* **2008**, *4*, 224, DOI: 10.1038/msb.2008.62.

41. Golynskiy, M. V.; Koay, M. S.; Vinkenborg, J. L.; Merkx, M. Engineering protein switches: sensors, regulators, and spare parts for biology and biotechnology. *ChemBioChem* **2011**, *12*, 353–361, DOI: 10.1002/cbic.201000642.

42. Bayat, P.; Nosrati, R.; Alibolandi, M.; Rafatpanah, H.; Abnous, K. et al. SELEX methods on the road to protein targeting with nucleic acid aptamers. *Biochimie* **2018**, *154*, 132–155, DOI: 10.1016/j.biochi.2018.09.001.

43. Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818–822, DOI: 10.1038/346818a0.

44. Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505–510, DOI: 10.1126/science.2200121.

45. Thodima, V.; Pirooznia, M.; Deng, Y. RiboaptDB: a comprehensive database of ribozymes and aptamers. *BMC Bioinformatics* **2006**, *7 Suppl 2*, S6, DOI: 10.1186/1471-2105-7-S2-S6.

46. Lee, J. F.; Hesselberth, J. R.; Meyers, L. A.; Ellington, A. D. Aptamer database. *Nucleic Acids Research* **2004**, *32*, D95–100, DOI: 10.1093/nar/gkh094.

47. Shaver, Z. M.; Bent, S. S.; Bilby, S. R.; Brown, M.; Buser, A. et al. Attempted use of PACE for riboswitch discovery generates three new translational theophylline riboswitch side products. *BMC Research Notes* **2018**, *11*, 861, DOI: 10.1186/s13104-018-3965-6.

48. Ogawa, A. Rational design of artificial riboswitches based on ligand-dependent modulation of internal ribosome entry in wheat germ extract and their applications as label-free biosensors. *RNA* **2011**, *17*, 478–488, DOI: 10.1261/rna.2433111.

49. Wittmann, A.; Suess, B. Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Molecular bioSystems* **2011**, *7*, 2419–2427, DOI: 10.1039/c1mb05070b.

50. Lynch, S. A.; Gallivan, J. P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Research* **2009**, *37*, 184–192, DOI: 10.1093/nar/gkn924.

51. Fowler, C. C.; Brown, E. D.; Li, Y. A FACS-based approach to engineering artificial riboswitches. *ChemBioChem* **2008**, *9*, 1906–1911, DOI: 10.1002/cbic.200700713.

52. Xiu, Y.; Jang, S.; Jones, J. A.; Zill, N. A.; Linhardt, R. J. et al. Naringenin-responsive riboswitch-based fluorescent biosensor module for Escherichia coli co-cultures. *Biotechnology and Bioengineering* **2017**, *114*, 2235–2244, DOI: 10.1002/bit.26340.

53. Liang, J. C.; Chang, A. L.; Kennedy, A. B.; Smolke, C. D. A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity. *Nucleic Acids Research* **2012**, *40*, e154, DOI: 10.1093/nar/gks636.

54. Sharma, V.; Nomura, Y.; Yokobayashi, Y. Engineering complex riboswitch regulation by dual genetic selection. *Journal of the American Chemical Society* **2008**, *130*, 16310–16315, DOI: 10.1021/ja805203w.

55. Townshend, B.; Kennedy, A. B.; Xiang, J. S.; Smolke, C. D. High-throughput cellular RNA device engineering. *Nature Methods* **2015**, *12*, 989–994, DOI: 10.1038/nmeth.3486.

56. Machtel, P.; Bąkowska-Żywicka, K.; Żywicki, M. Emerging applications of riboswitches - from antibacterial targets to molecular tools. *Journal of Applied Genetics* **2016**, *57*, 531–541, DOI: 10.1007/s13353-016-0341-x.

57. Ciruela, F. Fluorescence-based methods in the study of protein-protein interactions in living cells. *Current Opinion in Biotechnology* **2008**, *19*, 338–343, DOI: 10.1016/j.copbio.2008.06.003.

58. Lohse, M. J.; Nuber, S.; Hoffmann, C. Fluorescence/bioluminescence resonance energy transfer techniques to study G-protein-coupled receptor activation and signaling. *Pharmacological Reviews* **2012**, *64*, 299–336, DOI: 10.1124/pr.110.004309.

59. Algar, W. R.; Hildebrandt, N.; Vogel, S. S.; Medintz, I. L. FRET as a biomolecular research tool - understanding its potential while avoiding pitfalls. *Nature Methods* **2019**, *16*, 815–829, DOI: 10.1038/s41592-019-0530-8.

60. Machleidt, T.; Woodroofe, C. C.; Schwinn, M. K.; Méndez, J.; Robers, M. B. et al. NanoBRET–A Novel BRET Platform for the Analysis of Protein-Protein Interactions. *ACS Chemical Biology* **2015**, *10*, 1797–1804, DOI: 10.1021/acschembio.5b00143.

61. Bajar, B. T.; Wang, E. S.; Lam, A. J.; Kim, B. B.; Jacobs, C. L. et al. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Scientific Reports* **2016**, *6*, 20889, DOI: 10.1038/srep20889.

62. Rainey, K. H.; Patterson, G. H. Photoswitching FRET to monitor protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 864–873, DOI: 10.1073/pnas.1805333116.

63. Kim, C. K.; Cho, K. F.; Kim, M. W.; Ting, A. Y. Luciferase-LOV BRET enables versatile and specific transcriptional readout of cellular protein-protein interactions. *eLife* **2019**, *8*, e43826, DOI: 10.7554/eLife.43826.

64. Ausländer, S.; Fuchs, D.; Hürlemann, S.; Ausländer, D.; Fussenegger, M. Engineering a ribozyme cleavage-induced split fluorescent aptamer complementation assay. *Nucleic Acids Research* **2016**, *44*, e94, DOI: 10.1093/nar/gkw117.

65. Chandler, M.; Lyalina, T.; Halman, J.; Rackley, L.; Lee, L. et al. Broccoli Fluorets: Split Aptamers as a User-Friendly Fluorescent Toolkit for Dynamic RNA Nanotechnology. *Molecules* **2018**, *23*, 3178, DOI: 10.3390/molecules23123178.

66. Alam, K. K.; Tawiah, K. D.; Lichte, M. F.; Porciani, D.; Burke, D. H. A Fluorescent Split Aptamer for Visualizing RNA-RNA Assembly In Vivo. *ACS Synthetic Biology* **2017**, *6*, 1710–1721, DOI: 10.1021/acssynbio.7b00059.

67. Michnick, S. W.; Remy, I.; Campbell-Valois, F. X.; Vallée-Bélisle, A.; Pelletier, J. N. et al. Detection of protein-protein interactions by protein fragment complementation strategies. *Methods in Enzymology* **2000**, *328*, 208–230, DOI: 10.1016/s0076-6879(00)28399-7.

68. Shekhawat, S. S.; Ghosh, I. Split-protein systems: beyond binary protein-protein interactions. *Current Opinion in Chemical Biology* **2011**, *15*, 789–797, DOI: 10.1016/j.cbpa.2011.10.014.

69. Michnick, S. W. Exploring protein interactions by interaction-induced folding of proteins from complementary peptide fragments. *Current Opinion in Structural Biology* **2001**, *11*, 472–477, DOI: 10.1016/S0959-440X(00)00235-9.

70. Kodama, Y.; Hu, C.-D. Bimolecular fluorescence complementation (BiFC): a 5-year update and future perspectives. *BioTechniques* **2012**, *53*, 285–298, DOI: 10.2144/000113943.

71. Azad, T.; Tashakor, A.; Hosseinkhani, S. Split-luciferase complementary assay: applications, recent developments, and future perspectives. *Analytical and Bioanalytical Chemistry* **2014**, *406*, 5541–5560, DOI: 10.1007/s00216-014-7980-8.

72. Zhang, X.-E.; Cui, Z.; Wang, D. Sensing of biomolecular interactions using fluorescence complementing systems in living cells. *Biosensors & Bioelectronics* **2016**, *76*, 243–250, DOI: 10.1016/j.bios.2015.07.069.

73. Hu, C.-D.; Kerppola, T. K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nature Biotechnology* **2003**, *21*, 539–545, DOI: 10.1038/nbt816.

74. Verhoef, L. G. G. C.; Mattioli, M.; Ricci, F.; Li, Y.-C.; Wade, M. Multiplex detection of protein-protein interactions using a next generation luciferase reporter. *Biochimica et biophysica acta* **2016**, *1863*, 284–292, DOI: 10.1016/j.bbamcr.2015.11.031.

75. Foglieni, C.; Papin, S.; Salvadè, A.; Afroz, T.; Pinton, S. et al. Split GFP technologies to structurally characterize and quantify functional biomolecular interactions of FTD-related proteins. *Scientific Reports* **2017**, *7*, 14013, DOI: 10.1038/s41598-017-14459-w.

76. Hattori, M.; Ozawa, T. Split luciferase complementation for analysis of intracellular signaling. *Analytical Sciences* **2014**, *30*, 539–544, DOI: 10.2116/analsci.30.539.

77. Cabantous, S.; Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nature Methods* **2006**, *3*, 845–854, DOI: 10.1038/nmeth932.

78. Wade, M.; Méndez, J.; Coussens, N. P.; Arkin, M. R.; Glicksman, M. A. In *Assay Guidance Manual*; Sittampalam, G. S., Coussens, N. P., Brimacombe, K., Grossman, A., Arkin, M. et al. , Eds.; 2004.

79. Cabantous, S.; Nguyen, H. B.; Pedelacq, J.-D.; Koraïchi, F.; Chaudhary, A. et al. A new protein-protein interaction sensor based on tripartite split-GFP association. *Scientific Reports* **2013**, *3*, 2854, DOI: 10.1038/srep02854.

80. Ozawa, T.; Takeuchi, T. M.; Kaihara, A.; Sato, M.; Umezawa, Y. Protein splicing-based reconstitution of split green fluorescent protein for monitoring protein-protein interactions in bacteria: improved sensitivity and reduced screening time. *Analytical Chemistry* **2001**, *73*, 5866–5874, DOI: 10.1021/ac010717k.

81. Dixon, A. S.; Schwinn, M. K.; Hall, M. P.; Zimmerman, K.; Otto, P. et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chemical Biology* **2016**, *11*, 400–408, DOI: 10.1021/acschembio.5b00753.

82. Shyu, Y. J.; Liu, H.; Deng, X.; Hu, C.-D. Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions. *BioTechniques* **2006**, *40*, 61–66, DOI: 10.2144/000112036.

83. De Munter, S.; Görnemann, J.; Derua, R.; Lesage, B.; Qian, J. et al. Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions. *FEBS Letters* **2017**, *591*, 415–424, DOI: 10.1002/1873-3468.12548.

84. Schopp, I. M.; Amaya Ramirez, C. C.; Debeljak, J.; Kreibich, E.; Skribbe, M. et al. Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nature Communications* **2017**, *8*, 15690, DOI: 10.1038/ncomms15690.

85. Martell, J. D.; Yamagata, M.; Deerinck, T. J.; Phan, S.; Kwa, C. G. et al. A split horseradish peroxidase for the detection of intercellular protein-protein interactions

and sensitive visualization of synapses. *Nature Biotechnology* **2016**, *34*, 774–780, DOI: 10.1038/nbt.3563.

86. Han, Y.; Branon, T. C.; Martell, J. D.; Boassa, D.; Shechner, D. et al. Directed Evolution of Split APEX2 Peroxidase. *ACS Chemical Biology* **2019**, *14*, 619–635, DOI: 10.1021/acschembio.8b00919.

87. Wehr, M. C.; Laage, R.; Bolz, U.; Fischer, T. M.; Grünewald, S. et al. Monitoring regulated protein-protein interactions using split TEV. *Nature Methods* **2006**, *3*, 985–993, DOI: 10.1038/nmeth967.

88. Wintgens, J. P.; Rossner, M. J.; Wehr, M. C. Characterizing Dynamic Protein-Protein Interactions Using the Genetically Encoded Split Biosensor Assay Technique Split TEV. *Methods in Molecular Biology* **2017**, *1596*, 219–238, DOI: 10.1007/978-1-4939-6940-1_14.

89. Daringer, N. M.; Dudek, R. M.; Schwarz, K. A.; Leonard, J. N. Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS Synthetic Biology* **2014**, *3*, 892–902, DOI: 10.1021/sb400128g.

90. Fink, T.; Lonzarić, J.; Praznik, A.; Plaper, T.; Merljak, E. et al. Design of fast proteolysis-based signaling and logic circuits in mammalian cells. *Nature Chemical Biology* **2019**, *15*, 115–122, DOI: 10.1038/s41589-018-0181-6.

91. Gao, X. J.; Chong, L. S.; Kim, M. S.; Elowitz, M. B. Programmable protein circuits in living cells. *Science* **2018**, *361*, 1252–1258, DOI: 10.1126/science.aat5062.

92. Strader, C. D.; Gaffney, T.; Sugg, E. E.; Candelore, M. R.; Keys, R. et al. Allele-specific activation of genetically engineered receptors. *The Journal of Biological Chemistry* **1991**, *266*, 5–8.

93. Armbruster, B. N.; Li, X.; Pausch, M. H.; Herlitze, S.; Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 5163–5168, DOI: 10.1073/pnas.0700293104.

94. Dong, S.; Rogan, S. C.; Roth, B. L. Directed molecular evolution of DREADDs: a generic approach to creating next-generation RASSLs. *Nature Protocols* **2010**, *5*, 561–573, DOI: 10.1038/nprot.2009.239.

95. Coward, P.; Wada, H. G.; Falk, M. S.; Chan, S. D.; Meng, F. et al. Controlling signaling with a specifically designed Gi-coupled receptor. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 352–357, DOI: 10.1073/pnas.95.1.352.

96. Claeysen, S.; Joubert, L.; Sebben, M.; Bockaert, J.; Dumuis, A. A single mutation in the 5-HT4 receptor (5-HT4-R D100(3.32)A) generates a Gs-coupled receptor activated exclusively by synthetic ligands (RASSL). *The Journal of Biological Chemistry* **2003**, *278*, 699–702, DOI: 10.1074/jbc.C200588200.

97. Nissim, L.; Wu, M. R.; Pery, E.; Binder-Nissim, A.; Suzuki, H. I. et al. Synthetic RNA-Based Immunomodulatory Gene Circuits for Cancer Immunotherapy. *Cell* **2017**, *171*, 1138–1150 e15, DOI: 10.1016/j.cell.2017.09.049.

98. Roybal, K. T.; Lim, W. A. Synthetic Immunology: Hacking Immune Cells to Expand Their Therapeutic Capabilities. *Annual Review of Immunology* **2017**, *35*, 229–253, DOI: 10.1146/annurev-immunol-051116-052302.

99. Fischbach, M. A.; Bluestone, J. A.; Lim, W. A. Cell-based therapeutics: the next pillar of medicine. *Science Translational Medicine* **2013**, *5*, 179ps7, DOI: 10.1126/scitranslmed.3005568.

100. Lienert, F.; Lohmueller, J. J.; Garg, A.; Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature Reviews Molecular Cell Biology* **2014**, *15*, 95–107, DOI: 10.1038/nrm3738.

101. Slomovic, S.; Pardee, K.; Collins, J. J. Synthetic biology devices for in vitro and in vivo diagnostics. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 14429–14435, DOI: 10.1073/pnas.1508521112.

102. Morsut, L.; Roybal, K. T.; Xiong, X.; Gordley, R. M.; Coyle, S. M. et al. Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* **2016**, *164*, 780–791, DOI: 10.1016/j.cell.2016.01.012.

103. Toda, S.; Blauch, L. R.; Tang, S. K. Y.; Morsut, L.; Lim, W. A. Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science* **2018**, *361*, 156–162, DOI: 10.1126/science.aat0271.

104. Billerbeck, S.; Brisbois, J.; Agmon, N.; Jimenez, M.; Temple, J. et al. A scalable peptide-GPCR language for engineering multicellular communication. *Nature Communications* **2018**, *9*, 5057, DOI: 10.1038/s41467-018-07610-2.

105. Barnea, G.; Strapps, W.; Herrada, G.; Berman, Y.; Ong, J. et al. The genetic design of signaling cascades to record receptor activation. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 64–69, DOI: 10.1073/pnas.0710487105.

106. Kroeze, W. K.; Sassano, M. F.; Huang, X.-P. P.; Lansu, K.; McCorvy, J. D. et al. PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nature Structural & Molecular Biology* **2015**, *22*, 362–369, DOI: 10.1038/nsmb.3014.

107. Roybal, K. T.; Rupp, L. J.; Morsut, L.; Walker, W. J.; McNally, K. A. et al. Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits. *Cell* **2016**, *164*, 770–779, DOI: 10.1016/j.cell.2016.01.011.

108. Roybal, K. T.; Williams, J. Z.; Morsut, L.; Rupp, L. J.; Kolinko, I. et al. Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors. *Cell* **2016**, *167*, 419–432.e16, DOI: 10.1016/j.cell.2016.09.011.

109. Schwarz, K. A.; Daringer, N. M.; Dolberg, T. B.; Leonard, J. N. Rewiring human cellular input-output using modular extracellular sensors. *Nature Chemical Biology* **2017**, *13*, 202–209, DOI: 10.1038/nchembio.2253.

110. Hartfield, R. M.; Schwarz, K. A.; Muldoon, J. J.; Bagheri, N.; Leonard, J. N. Multiplexing Engineered Receptors for Multiparametric Evaluation of Environmental Ligands. *ACS Synthetic Biology* **2017**, *6*, 2042–2055, DOI: 10.1021/acssynbio.6b00279.

111. Chang, Z. L.; Lorenzini, M. H.; Chen, X.; Tran, U.; Bangayan, N. J. et al. Rewiring T-cell responses to soluble factors with chimeric antigen receptors. *Nature Chemical Biology* **2018**, *14*, 317–324, DOI: 10.1038/nchembio.2565.

112. Hombach, A.; Heuser, C.; Sircar, R.; Tillmann, T.; Diehl, V. et al. An anti-CD30 chimeric receptor that mediates CD3-zeta-independent T-cell activation against Hodgkin's lymphoma cells in the presence of soluble CD30. *Cancer Research* **1998**, *58*, 1116–1119.

113. Lanitis, E.; Poussin, M.; Hagemann, I. S.; Coukos, G.; Sandaltzopoulos, R. et al. Redirected antitumor activity of primary human lymphocytes transduced with a fully human anti-mesothelin chimeric receptor. *Molecular Therapy* **2012**, *20*, 633–643, DOI: 10.1038/mt.2011.256.

114. Nolan, K. F.; Yun, C. O.; Akamatsu, Y.; Murphy, J. C.; Leung, S. O. et al. Bypassing immunization: optimized design of "designer T cells" against carcinoembryonic antigen (CEA)-expressing tumors, and lack of suppression by soluble CEA. *Clinical Cancer Research* **1999**, *5*, 3928–2941.

115. Westwood, J. A.; Murray, W. K.; Trivett, M.; Haynes, N. M.; Solomon, B. et al. The Lewis-Y carbohydrate antigen is expressed by many human tumors and can serve as a target for genetically redirected T cells despite the presence of soluble antigen in serum. *Journal of Immunotherapy* **2009**, *32*, 292–301, DOI: 10.1097/CJI.0b013e31819b7c8e.

116. Ma, Q.; DeMarte, L.; Wang, Y.; Stanners, C. P.; Junghans, R. P. Carcinoembryonic antigen-immunoglobulin Fc fusion protein (CEA-Fc) for identification and activation of anti-CEA immunoglobulin-T-cell receptor-modified T cells, representative of a new class of Ig fusion proteins. *Cancer Gene Therapy* **2004**, *11*, 297–306, DOI: 10.1038/sj.cgt.7700685.

117. Carpenter, R. O.; Evbuomwan, M. O.; Pittaluga, S.; Rose, J. J.; Raffeld, M. et al. B-cell Maturation Antigen Is a Promising Target for Adoptive T-cell Therapy of Multiple Myeloma. *Clinical Cancer Research* **2013**, *19*, 2048–2060, DOI: 10.1158/1078-0432.CCR-12-2422.

118. McGuinness, R. P.; Ge, Y.; Patel, S. D.; Kashmiri, S. V.; Lee, H. S. et al. Anti-tumor activity of human T cells expressing the CC49-zeta chimeric immune receptor. *Human Gene Therapy* **1999**, *10*, 165–173, DOI: 10.1089/10430349950018968.

119. Chmielewski, M.; Hahn, O.; Rappl, G.; Nowak, M.; Schmidt-Wolf, I. H. et al. T cells that target carcinoembryonic antigen eradicate orthotopic pancreatic carcinomas without inducing autoimmune colitis in mice. *Gastroenterology* **2012**, *143*, 1095–107.e2, DOI: 10.1053/j.gastro.2012.06.037.

120. Deplancke, B.; Dupuy, D.; Vidal, M.; Walhout, A. J. M. A gateway-compatible yeast one-hybrid system. *Genome Research* **2004**, *14*, 2093–2101, DOI: 10.1101/gr.2445504.

121. Fields, S.; Song, O.-k. K. A novel genetic system to detect protein-protein interactions. *Nature* **1989**, *340*, 245–6, DOI: 10.1038/340245a0.

122. Di Lallo, G.; Castagnoli, L.; Ghelardini, P.; Paolozzi, L. A two-hybrid system based on chimeric operator recognition for studying protein homo/heterodimerization in Escherichia coli. *Microbiology* **2001**, *147*, 1651–1656, DOI: 10.1099/00221287-147-6-1651.

123. Putz, U.; Skehel, P.; Kuhl, D. A tri-hybrid system for the analysis and detection of RNA–protein interactions. *Nucleic Acids Research* **1996**, *24*, 4838–4840, DOI: 10.1093/nar/24.23.4838.

124. SenGupta, D. J.; Zhang, B.; Kraemer, B.; Pochart, P.; Fields, S. et al. A three-hybrid system to detect RNA-protein interactions in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 8496–8501, DOI: 10.1073/pnas.93.16.8496.

125. Baker, K.; Bleczinski, C.; Lin, H.; Salazar-Jimenez, G.; Sengupta, D. et al. Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 16537–16542, DOI: 10.1073/pnas.262420099.

126. Rajagopala, S. V. Mapping the Protein-Protein Interactome Networks Using Yeast Two-Hybrid Screens. *Advances in Experimental Medicine and Biology* **2015**, *883*, 187–214, DOI: 10.1007/978-3-319-23603-2_11.

127. Rolland, T.; Taşan, M.; Charloteaux, B.; Pevzner, S. J.; Zhong, Q. et al. A proteome-scale map of the human interactome network. *Cell* **2014**, *159*, 1212–1226, DOI: 10.1016/j.cell.2014.10.050.

128. Rajagopala, S. V.; Sikorski, P.; Kumar, A.; Mosca, R.; Vlasblom, J. et al. The binary protein-protein interaction landscape of Escherichia coli. *Nature Biotechnology* **2014**, *32*, 285–290, DOI: 10.1038/nbt.2831.

129. Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S. et al. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **2000**, *403*, 623–627, DOI: 10.1038/35001009.

130. Yu, H.; Braun, P.; Yildirim, M. A.; Lemmens, I.; Venkatesan, K. et al. High-quality binary protein interaction map of the yeast interactome network. *Science* **2008**, *322*, 104–110, DOI: 10.1126/science.1158684.

131. Gong, W.; Shen, Y.-P. P.; Ma, L.-G. G.; Pan, Y.; Du, Y.-L. L. et al. Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant physiology* **2004**, *135*, 773–82, DOI: 10.1104/pp.104.042176.

132. Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B. et al. A protein interaction map of Drosophila melanogaster. *Science* **2003**, *302*, 1727–1736, DOI: 10.1126/science.1090289.

133. Copeland, M. F.; Politz, M. C.; Johnson, C. B.; Markley, A. L.; Pfleger, B. F. A transcription activator-like effector (TALE) induction system mediated by proteolysis. *Nature Chemical Biology* **2016**, *12*, 254–260, DOI: 10.1038/nchembio.2021.

134. Konermann, S.; Brigham, M. D.; Trevino, A. E.; Hsu, P. D.; Heidenreich, M. et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **2013**, *500*, 472–476, DOI: 10.1038/nature12466.

135. Polstein, L. R.; Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology* **2015**, *11*, 198–200, DOI: 10.1038/nchembio.1753.

136. Nuñez, J. K.; Harrington, L. B.; Doudna, J. A. Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chemical Biology* **2016**, *11*, 681–688, DOI: 10.1021/acschembio.5b01019.

137. Gao, Y.; Xiong, X.; Wong, S.; Charles, E. J.; Lim, W. A. et al. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nature Methods* **2016**, *13*, 1043–1049, DOI: 10.1038/nmeth.4042.

138. Nihongaki, Y.; Yamamoto, S.; Kawano, F.; Suzuki, H.; Sato, M. CRISPR-Cas9-based photoactivatable transcription system. *Chemistry & Biology* **2015**, *22*, 169–174, DOI: 10.1016/j.chembiol.2014.12.011.

139. Kiani, S.; Chavez, A.; Tuttle, M.; Hall, R. N.; Chari, R. et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nature Methods* **2015**, *12*, 1051–1054, DOI: 10.1038/nmeth.3580.

140. Kiani, S.; Beal, J.; Ebrahimkhani, M. R.; Huh, J.; Hall, R. N. et al. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nature Methods* **2014**, *11*, 723–726, DOI: 10.1038/nmeth.2969.

141. Nissim, L.; Perli, S. D.; Fridkin, A.; Perez-Pinera, P.; Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular Cell* **2014**, *54*, 698–710, DOI: 10.1016/j.molcel.2014.04.022.

142. Jusiak, B.; Cleto, S.; Perez-Piñera, P.; Lu, T. K. Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. *Trends in Biotechnology* **2016**, *34*, 535–547, DOI: 10.1016/j.tibtech.2015.12.014.

143. Martin, F. Fifteen years of the yeast three-hybrid system: RNA-protein interactions under investigation. *Methods* **2012**, *58*, 367–375, DOI: 10.1016/j.ymeth.2012.07.016.

144. Braun, P.; Tasan, M.; Dreze, M.; Barrios-Rodiles, M.; Lemmens, I. et al. An experimentally derived confidence score for binary protein-protein interactions. *Nature Methods* **2009**, *6*, 91–97, DOI: 10.1038/nmeth.1281.

145. Chen, Y.-C. C.; Rajagopala, S. V.; Stellberger, T.; Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nature Methods* **2010**, *7*, 667–668, DOI: 10.1038/nmeth0910-667.

146. Stellberger, T.; Häuser, R.; Baiker, A.; Pothineni, V. R.; Haas, J. et al. Improving the yeast two-hybrid system with permutated fusions proteins: the Varicella Zoster Virus interactome. *Proteome Science* **2010**, *8*, 8, DOI: 10.1186/1477-5956-8-8.

147. Koegl, M.; Uetz, P. Improving yeast two-hybrid screening systems. *Briefings in Functional Genomics & Proteomics* **2007**, *6*, 302–312, DOI: 10.1093/bfgp/elm035.

148. Raina, M.; Sharma, R.; Deacon, S. E.; Tiede, C.; Tomlinson, D. et al. Antibody mimetic receptor proteins for label-free biosensors. *The Analyst* **2015**, *140*, 803–810, DOI: 10.1039/c4an01418a.

149. Ruder, W. C.; Lu, T.; Collins, J. J. Synthetic biology moving into the clinic. *Science* **2011**, *333*, 1248–1252, DOI: 10.1126/science.1206843.

150. Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **2015**, *16*, 379–394, DOI: 10.1038/nrg3927.

151. Benenson, Y. RNA-based computation in live cells. *Current Opinion in Biotechnology* **2009**, *20*, 471–478, DOI: 10.1016/j.copbio.2009.08.002.

152. Green, A. A.; Silver, P. A.; Collins, J. J.; Yin, P. Toehold switches: de-novo-designed regulators of gene expression. *Cell* **2014**, *159*, 925–939, DOI: 10.1016/j.cell.2014.10.002.

153. Brophy, J. A. N.; Voigt, C. A. Principles of genetic circuit design. *Nature Methods* **2014**, *11*, 508–520, DOI: 10.1038/nmeth.2926.

154. Church, G. M.; Elowitz, M. B.; Smolke, C. D.; Voigt, C. A.; Weiss, R. Realizing the potential of synthetic biology. *Nature Reviews Molecular Cell Biology* **2014**, *15*, 289–294, DOI: 10.1038/nrm3767.

155. Yen, L.; Svendsen, J.; Lee, J.-S.; Gray, J. T.; Magnier, M. et al. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* **2004**, *431*, 471–476, DOI: 10.1038/nature02844.

156. Winkler, W. C.; Breaker, R. R. Regulation of bacterial gene expression by riboswitches. *Annual Review of Microbiology* **2005**, *59*, 487–517, DOI: 10.1146/annurev.micro.59.030804.121336.

157. Culler, S. J.; Hoff, K. G.; Smolke, C. D. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* **2010**, *330*, 1251–1255, DOI: 10.1126/science.1192128.

158. Lee, J. H.; Daugharthy, E. R.; Scheiman, J.; Kalhor, R.; Yang, J. L. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **2014**, *343*, 1360–1363, DOI: 10.1126/science.1250212.

159. Zamft, B. M.; Marblestone, A. H.; Kording, K.; Schmidt, D.; Martin-Alarcon, D. et al. Measuring Cation Dependent DNA Polymerase Fidelity Landscapes by Deep Sequencing. *PLoS ONE* **2012**, *7*, e43876, DOI: 10.1371/journal.pone.0043876.

160. Pu, J.; Chronis, I.; Ahn, D.; Dickinson, B. C. A Panel of Protease-Responsive RNA Polymerases Respond to Biochemical Signals by Production of Defined RNA Outputs in Live Cells. *Journal of the American Chemical Society* **2015**, *137*, 15996–15999, DOI: 10.1021/jacs.5b10290.

161. Segall-Shapiro, T. H.; Meyer, A. J.; Ellington, A. D.; Sontag, E. D.; Voigt, C. A. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Molecular Systems Biology* **2014**, *10*, 742, DOI: 10.15252/msb.20145299.

162. Shis, D. L.; Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 5028–5033, DOI: 10.1073/pnas.1220157110.

163. Kerppola, T. K. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation. *Chemical Society Reviews* **2009**, *38*, 2876–2886, DOI: 10.1039/b909638h.

164. Dickinson, B. C.; Leconte, A. M.; Allen, B.; Esvelt, K. M.; Liu, D. R. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 9007–9012, DOI: 10.1073/pnas.1220670110.

165. Ellefson, J. W.; Meyer, A. J.; Hughes, R. A.; Cannon, J. R.; Brodbelt, J. S. et al. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nature Biotechnology* **2014**, *32*, 97–101, DOI: 10.1038/nbt.2714.

166. Ghosh, I.; Hamilton, A. D.; Regan, L. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *Journal of the American Chemical Society* **2000**, *122*, 5658–5659, DOI: 10.1021/ja994421w.

167. Magliery, T. J.; Wilson, C. G. M.; Pan, W.; Mishler, D.; Ghosh, I. et al. Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *Journal of the American Chemical Society* **2005**, *127*, 146–157, DOI: 10.1021/ja046699g.

168. Steitz, T. A. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Current Opinion in Structural Biology* **2009**, *19*, 683–690, DOI: 10.1016/j.sbi.2009.09.001.

169. Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **2011**, *472*, 499–503, DOI: 10.1038/nature09929.

170. Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P. et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **2016**, *533*, 58–63, DOI: 10.1038/nature17938.

171. Carlson, J. C.; Badran, A. H.; Guggiana-Nilo, D. A.; Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* **2014**, *10*, 216–222, DOI: 10.1038/nchembio.1453.

172. Dickinson, B. C.; Packer, M. S.; Badran, A. H.; Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nature Communications* **2014**, *5*, 5352, DOI: 10.1038/ncomms6352.

173. Hubbard, B. P.; Badran, A. H.; Zuris, J. A.; Guilinger, J. P.; Davis, K. M. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature Methods* **2015**, *12*, 939–942, DOI: 10.1038/nmeth.3515.

174. Leconte, A. M.; Dickinson, B. C.; Yang, D. D.; Chen, I. a.; Allen, B. et al. A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* **2013**, *52*, 1490–1499, DOI: 10.1021/bi3016185.

175. Ringquist, S.; Shinedling, S.; Barrick, D.; Green, L.; Binkley, J. et al. Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Molecular Microbiology* **1992**, *6*, 1219–1229, DOI: 10.1111/j.1365-2958.1992.tb01561.x.

176. Cheetham, G. M.; Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **1999**, *286*, 2305–2309, DOI: 10.1126/science.286.5448.2305.

177. Tahirov, T. H.; Temiakov, D.; Anikin, M.; Patlan, V.; McAllister, W. T. et al. Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature* **2002**, *420*, 43–50, DOI: 10.1038/nature01129.

178. Guntas, G.; Hallett, R. A.; Zimmerman, S. P.; Williams, T.; Yumerefendi, H. et al. Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 112–117, DOI: 10.1073/pnas.1417910112.

179. Rivera, V. M.; Clackson, T.; Natesan, S.; Pollock, R.; Amara, J. F. et al. A humanized system for pharmacologic control of gene expression. *Nature Medicine* **1996**, *2*, 1028–1032.

180. Gossen, M.; Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 5547–5551, DOI: 10.1073/pnas.89.12.5547.

181. Paulmurugan, R.; Padmanabhan, P.; Ahn, B.-C.; Ray, S.; Willmann, J. K. et al. A novel estrogen receptor intramolecular folding-based titratable transgene expression system. *Molecular Therapy* **2009**, *17*, 1703–1711, DOI: 10.1038/mt.2009.171.

182. Kerppola, T. K. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annual Review of Biophysics* **2008**, *37*, 465–487, DOI: 10.1146/annurev.biophys.37.032807.125842.

183. Filonov, G. S.; Kam, C. W.; Song, W.; Jaffrey, S. R. In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chemistry & Biology* **2015**, *22*, 649–660, DOI: 10.1016/j.chembiol.2015.04.018.

184. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **2009**, *6*, 343–345, DOI: 10.1038/nmeth.1318.

185. Petta, I.; Lievens, S.; Libert, C.; Tavernier, J.; De Bosscher, K. Modulation of Protein-Protein Interactions for the Development of Novel Therapeutics. *Molecular Therapy* **2016**, *24*, 707–718, DOI: 10.1038/mt.2015.214.

186. Valeur, E.; Jimonet, P. New Modalities, Technologies, and Partnerships in Probe and Lead Generation: Enabling a Mode-of-Action Centric Paradigm. *Journal of Medicinal Chemistry* **2018**, *61*, 9004–9029, DOI: 10.1021/acs.jmedchem.8b00378.

187. Valeur, E.; Narjes, F.; Ottmann, C.; Plowright, A. T. Emerging modes-of-action in drug discovery. *MedChemComm* **2019**, *10*, 1550–1568, DOI: 10.1039/C9MD00263D.

188. Taipale, M. Disruption of protein function by pathogenic mutations: common and uncommon mechanisms. *Biochemistry and Cell Biology* **2019**, *97*, 46–57, DOI: 10.1139/bcb-2018-0007.

189. Cunningham, B. C.; Wells, J. A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **1989**, *244*, 1081–1085, DOI: 10.1126/science.2471267.

190. Weiss, G. A.; Watanabe, C. K.; Zhong, A.; Goddard, A.; Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 8950–8954, DOI: 10.1073/pnas.160252097.

191. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot spots–a review of the protein-protein interface determinant amino-acid residues. *Proteins* **2007**, *68*, 803–812, DOI: 10.1002/prot.21396.

192. Goodyear, C. S.; Silverman, G. J. Phage-Display Methodology for the Study of Protein-Protein Interactions: Overview. *Cold Spring Harbor Protocols* **2008**, *3*, pdb.top48, DOI: 10.1101/pdb.top48.

193. Younger, D.; Berger, S.; Baker, D.; Klavins, E. High-throughput characterization of protein-protein interactions by reprogramming yeast mating. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 12166–12171, DOI: 10.1073/pnas.1705867114.

194. Lin, J.-S.; Lai, E.-M. In *Bacterial Protein Secretion Systems: Methods and Protocols*; Journet, L., Cascales, E., Eds.; Springer New York: New York, NY, 2017; pp 177–187, DOI: 10.1007/978-1-4939-7033-9_14.

195. Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **2009**, *10*, 866–876, DOI: 10.1038/nrm2805.

196. Erffelinck, M.-L.; Ribeiro, B.; Perassolo, M.; Pauwels, L.; Pollier, J. et al. A user-friendly platform for yeast two-hybrid library screening using next generation sequencing. *PLoS ONE* **2018**, *13*, e0201270, DOI: 10.1371/journal.pone.0201270.

197. Fowler, D. M.; Fields, S. Deep mutational scanning: a new style of protein science. *Nature Methods* **2014**, *11*, 801–807, DOI: 10.1038/nmeth.3027.

198. Araya, C. L.; Fowler, D. M. Deep mutational scanning: assessing protein function on a massive scale. *Trends in Biotechnology* **2011**, *29*, 435–442, DOI: 10.1016/j.tibtech.2011.04.003.

199. Fowler, D. M.; Stephany, J. J.; Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nature Protocols* **2014**, *9*, 2267–2284, DOI: 10.1038/nprot.2014.153.

200. Starita, L. M.; Young, D. L.; Islam, M.; Kitzman, J. O.; Gullingsrud, J. et al. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics* **2015**, *200*, 413–422, DOI: 10.1534/genetics.115.175802.

201. Araya, C. L.; Fowler, D. M.; Chen, W.; Muniez, I.; Kelly, J. W. et al. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 16858–16863, DOI: 10.1073/pnas.1209751109.

202. Fowler, D. M.; Araya, C. L.; Fleishman, S. J.; Kellogg, E. H.; Stephany, J. J. et al. High-resolution mapping of protein sequence-function relationships. *Nature Methods* **2010**, *7*, 741–746, DOI: 10.1038/nmeth.1492.

203. Stiffler, M. A.; Subramanian, S. K.; Salinas, V. H.; Ranganathan, R. A Protocol for Functional Assessment of Whole-Protein Saturation Mutagenesis Libraries Utilizing High-Throughput Sequencing. *Journal of Visualized Experiments* **2016**, e54119, DOI: 10.3791/54119.

204. Roscoe, B. P.; Thayer, K. M.; Zeldovich, K. B.; Fushman, D.; Bolon, D. N. A. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *Journal of Molecular Biology* **2013**, *425*, 1363–77, DOI: 10.1016/j.jmb.2013.01.032.

205. Suzuki, T.; Miller, C.; Guo, L.-T.; Ho, J. M. L.; Bryson, D. I. et al. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nature Chemical Biology* **2017**, *13*, 1261–1266, DOI: 10.1038/nchembio.2497.

206. Roth, T. B.; Woolston, B. M.; Stephanopoulos, G.; Liu, D. R. Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. *ACS Synthetic Biology* **2019**, *8*, 796–806, DOI: 10.1021/acssynbio.8b00481.

207. Packer, M. S.; Rees, H. A.; Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nature Communications* **2017**, *8*, 956, DOI: 10.1038/s41467-017-01055-9.

208. Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **2018**, *556*, 57–63, DOI: 10.1038/nature26155.

209. Wang, T.; Badran, A. H.; Huang, T. P.; Liu, D. R. Continuous directed evolution of proteins with improved soluble expression. *Nature Chemical Biology* **2018**, *14*, 972–980, DOI: 10.1038/s41589-018-0121-5.

210. Prior, I. A.; Lewis, P. D.; Mattos, C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Research* **2012**, *72*, 2457–2467, DOI: 10.1158/0008-5472.CAN-11-2612.

211. Davies, H.; Bignell, G. R.; Cox, C.; Stephens, P.; Edkins, S. et al. Mutations of the BRAF gene in human cancer. *Nature* **2002**, *417*, 949–954, DOI: 10.1038/nature00766.

212. Hobbs, G. A.; Der, C. J.; Rossman, K. L. RAS isoforms and mutations in cancer at a glance. *Journal of Cell Science* **2016**, *129*, 1287–1292, DOI: 10.1242/jcs.182873.

213. Scheffler, J. E.; Waugh, D. S.; Bekesi, E.; Kiefer, S. E.; LoSardo, J. E. et al. Characterization of a 78-residue fragment of c-Raf-1 that comprises a minimal binding domain for the interaction with Ras-GTP. *The Journal of Biological Chemistry* **1994**, *269*, 22340–22346.

214. Vojtek, A. B.; Hollenberg, S. M.; Cooper, J. A. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* **1993**, *74*, 205–214, DOI: 10.1016/0092-8674(93)90307-c.

215. Pu, J.; Zinkus-Boltz, J.; Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nature Chemical Biology* **2017**, *13*, 432–438, DOI: 10.1038/nchembio.2299.

216. Jones, K. A.; Zinkus-Boltz, J.; Dickinson, B. C. Recent advances in developing and applying biosensors for synthetic biology. *Nano Futures* **2019**, *3*, 042002, DOI: 10.1088/2399-1984/ab4b78.

217. Pu, J.; Dewey, J. A.; Hadji, A.; LaBelle, J. L.; Dickinson, B. C. RNA Polymerase Tags To Monitor Multidimensional Protein-Protein Interactions Reveal Pharmacological Engagement of Bcl-2 Proteins. *Journal of the American Chemical Society* **2017**, *139*, 11964–11972, DOI: 10.1021/jacs.7b06152.

218. Pu, J.; Kentala, K.; Dickinson, B. C. Multidimensional Control of Cas9 by Evolved RNA Polymerase-Based Biosensors. *ACS Chemical Biology* **2018**, *13*, 431–437, DOI: 10.1021/acschembio.7b00532.

219. Kirchner, G.; Roberts, J. L.; Gustafson, G. D.; Ingolia, T. D. Active bacterial luciferase from a fused gene: expression of a Vibrio harveyi luxAB translational fusion in bacteria, yeast and plant cells. *Gene* **1989**, *81*, 349–354, DOI: 10.1016/0378-1119(89)90195-9.

220. Fabian, J. R.; Vojtek, A. B.; Cooper, J. A.; Morrison, D. K. A single amino acid change in Raf-1 inhibits Ras binding and alters Raf-1 function. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 5982–5986, DOI: 10.1073/pnas.91.13.5982.

221. Bennett, N. J.; Rakonjac, J. Unlocking of the filamentous bacteriophage virion during infection is mediated by the C domain of pIII. *Journal of Molecular Biology* **2006**, *356*, 266–273, DOI: 10.1016/j.jmb.2005.11.069.

222. Block, C.; Janknecht, R.; Herrmann, C.; Nassar, N.; Wittinghofer, A. Quantitative structure-activity analysis correlating Ras/Raf interaction in vitro to Raf activation in vivo. *Nature Structural Biology* **1996**, *3*, 244–251, DOI: 10.1038/nsb0396-244.

223. Badran, A. H.; Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nature Communications* **2015**, *6*, 8425, DOI: 10.1038/ncomms9425.

224. Cadwell, R. C.; Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods and Applications* **1992**, *2*, 28–33.

225. Ferla, M. P. Mutanalyst, an online tool for assessing the mutational spectrum of epPCR libraries with poor sampling. *BMC Bioinformatics* **2016**, *17*, 152, DOI: 10.1186/s12859-016-0996-7.

226. Marshall, M. Interactions between Ras and Raf: key regulatory proteins in cellular transformation. *Molecular Reproduction and Development* **1995**, *42*, 493–499, DOI: 10.1002/mrd.1080420418.

227. Campbell-Valois, F.-X.; Tarassov, K.; Michnick, S. W. Massive sequence perturbation of the Raf ras binding domain reveals relationships between sequence conservation, secondary structure propensity, hydrophobic core organization and stability. *Journal of Molecular Biology* **2006**, *362*, 151–171, DOI: 10.1016/j.jmb.2006.06.061.

228. Vanhercke, T.; Ampe, C.; Tirry, L.; Denolf, P. Reducing mutational bias in random protein libraries. *Analytical Biochemistry* **2005**, *339*, 9–14, DOI: 10.1016/j.ab.2004.11.032.

229. Moore, C. L.; Papa, L. J.; Shoulders, M. D. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo. *Journal of the American Chemical Society* **2018**, *140*, 11560–11564, DOI: 10.1021/jacs.8b04001.

230. Pu, J.; Disare, M.; Dickinson, B. C. Evolution of C-Terminal Modification Tolerance in Full-Length and Split T7 RNA Polymerase Biosensors. *ChemBioChem* **2019**, *20*, 1547–1553, DOI: 10.1002/cbic.201800707.

231. English, J. G.; Olsen, R. H. J.; Lansu, K.; Patel, M.; White, K. et al. VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells. *Cell* **2019**, *178*, 748–761, DOI: 10.1016/j.cell.2019.05.051.

232. Berman, C. M.; Papa, L. J.; Hendel, S. J.; Moore, C. L.; Suen, P. H. et al. An Adaptable Platform for Directed Evolution in Human Cells. *Journal of the American Chemical Society* **2018**, *140*, 18093–18103, DOI: 10.1021/jacs.8b10937.

233. de Boer, H. A.; Comstock, L. J.; Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences of the United States of America* **1983**, *80*, 21–25, DOI: 10.1073/pnas.80.1.21.

234. Magoč, T.; Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963, DOI: 10.1093/bioinformatics/btr507.

235. Wang, W.; Li, Y.; Wang, Y.; Shi, C.; Li, C. et al. Bacteriophage T7 transcription system: an enabling tool in synthetic biology. *Biotechnology Advances* **2018**, *36*, 2129–2137, DOI: 10.1016/j.biotechadv.2018.10.001.

236. Chamberlin, M.; McGrath, J.; Waskell, L. New RNA polymerase from Escherichia coli infected with bacteriophage T7. *Nature* **1970**, *228*, 227–231, DOI: 10.1038/228227a0.

237. Studier, F. W.; Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **1986**, *189*, 113–130, DOI: 10.1016/0022-2836(86)90385-2.

238. Yu, X.; Yang, Y.-P.; Dikici, E.; Deo, S. K.; Daunert, S. Beyond Antibodies as Binding Partners: The Role of Antibody Mimetics in Bioanalysis. *Annual Review of Analytical Chemistry* **2017**, *10*, 293–320, DOI: 10.1146/annurev-anchem-061516-045205.

239. Vazquez-Lombardi, R.; Phan, T. G.; Zimmermann, C.; Lowe, D.; Jermutus, L. et al. Challenges and opportunities for non-antibody scaffold drugs. *Drug Discovery Today* **2015**, *20*, 1271–1283, DOI: 10.1016/j.drudis.2015.09.004.

240. Simeon, R.; Chen, Z. In vitro-engineered non-antibody protein therapeutics. *Protein & Cell* **2018**, *9*, 3–14, DOI: 10.1007/s13238-017-0386-6.

241. Owens, B. Faster, deeper, smaller-the rise of antibody-like scaffolds. *Nature Biotechnology* **2017**, *35*, 602–603, DOI: 10.1038/nbt0717-602.

242. Wuo, M. G.; Arora, P. S. Engineered protein scaffolds as leads for synthetic inhibitors of protein–protein interactions. *Current Opinion in Chemical Biology* **2018**, *44*, 16–22, DOI: 10.1016/j.cbpa.2018.05.013.

243. Tiede, C.; Tang, A. A. S.; Deacon, S. E.; Mandal, U.; Nettleship, J. E. et al. Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Engineering, Design & Selection* **2014**, *27*, 145–155, DOI: 10.1093/protein/gzu007.

244. Grabulovski, D.; Kaspar, M.; Neri, D. A novel, non-immunogenic Fyn SH3-derived binding protein with tumor vascular targeting properties. *Journal of Biological Chemistry* **2007**, *282*, 3196–3204, DOI: 10.1074/jbc.m609211200.

245. Tiede, C.; Bedford, R.; Heseltine, S. J.; Smith, G.; Wijetunga, I. et al. Affimer proteins are versatile and renewable affinity reagents. *eLife* **2017**, *6*, 1–35, DOI: 10.7554/eLife.24903.

246. Hughes, D. J.; Tiede, C.; Penswick, N.; Tang, A. A.-S. S.; Trinh, C. H. et al. Generation of specific inhibitors of SUMO-1- and SUMO-2/3-mediated protein-protein interactions using Affimer (Adhiron) technology. *Science Signaling* **2017**, *10*, 1–14, DOI: 10.1126/scisignal.aaj2005.

247. Robinson, J. I.; Baxter, E. W.; Owen, R. L.; Thomsen, M.; Tomlinson, D. C. et al. Affimer proteins inhibit immune complex binding to FcγRIIIa with high specificity through competitive and allosteric modes of action. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, E72–E81, DOI: 10.1073/pnas.1707856115.

248. Lopata, A.; Hughes, R.; Tiede, C.; Heissler, S. M.; Sellers, J. R. et al. Affimer proteins for F-actin: Novel affinity reagents that label F-actin in live and fixed cells. *Scientific Reports* **2018**, *8*, 1–15, DOI: 10.1038/s41598-018-24953-4.

249. Edwards, T. A.; Miles, J. A.; Hobor, F.; Taylor, J. L.; Tiede, C. et al. Selective Affimers Recognize BCL-2 Family Proteins Through Non-Canonical Structural Motifs. *bioRxiv Biochemistry* **2019**, DOI: 10.1101/651364.

250. Schlichthaerle, T.; Eklund, A. S.; Schueder, F.; Strauss, M. T.; Tiede, C. et al. Site-Specific Labeling of Affimers for DNA-PAINT Microscopy. *Angewandte Chemie* **2018**, *57*, 11060–11063, DOI: 10.1002/anie.201804020.

251. Zhurauski, P.; Arya, S. K.; Jolly, P.; Tiede, C.; Tomlinson, D. C. et al. Sensitive and selective Affimer-functionalised interdigitated electrode-based capacitive biosensor for Her4 protein tumour biomarker detection. *Biosensors & Bioelectronics* **2018**, *108*, 1–8, DOI: 10.1016/j.bios.2018.02.041.

252. Johnson, A.; Song, Q.; Ko Ferrigno, P.; Bueno, P. R.; Davis, J. J. Sensitive affimer and antibody based impedimetric label-free assays for C-reactive protein. *Analytical Chemistry* **2012**, *84*, 6553–6560, DOI: 10.1021/ac300835b.

253. Xie, C.; Tiede, C.; Zhang, X.; Wang, C.; Li, Z. et al. Development of an Affimer-antibody combined immunological diagnosis kit for glypican-3. *Scientific Reports* **2017**, *7*, 9608, DOI: 10.1038/s41598-017-10083-w.

254. Koutsoumpeli, E.; Tiede, C.; Murray, J.; Tang, A.; Bon, R. S. et al. Antibody Mimetics for the Detection of Small Organic Compounds Using a Quartz Crystal Microbalance. *Analytical Chemistry* **2017**, *89*, 3051–3058, DOI: 10.1021/acs.analchem.6b04790.

255. Silacci, M.; Baenziger-Tobler, N.; Lembke, W.; Zha, W.; Batey, S. et al. Linker length matters, fynomer-Fc fusion with an optimized linker displaying picomolar IL-17A inhibition potency. *The Journal of Biological Chemistry* **2014**, *289*, 14392–14398, DOI: 10.1074/jbc.M113.534578.

256. Bertschinger, J.; Grabulovski, D.; Neri, D. Selection of single domain binding proteins by covalent DNA display. *Protein Engineering, Design and Selection* **2007**, *20*, 57–68, DOI: 10.1093/protein/gzl055.

257. Schlatter, D.; Brack, S.; Banner, D. W.; Batey, S.; Benz, J. et al. Generation, characterization and structural data of chymase binding proteins based on the human Fyn kinase SH3 domain. *mAbs* **2012**, *4*, 497–508, DOI: 10.4161/mabs.20452.

258. Banner, D. W.; Gsell, B.; Benz, J.; Bertschinger, J.; Burger, D. et al. Mapping the conformational space accessible to BACE2 using surface mutants and cocrystals with Fab fragments, Fynomers and Xaperones. *Acta Crystallographica Section D Biological Crystallography* **2013**, *69*, 1124–1137, DOI: 10.1107/s0907444913006574.

259. Klupsch, K.; Baeriswyl, V.; Scholz, R.; Dannenberg, J.; Santimaria, R. et al. COVA4231, a potent CD3/CD33 bispecific FynomAb with IgG-like pharmacokinetics for the treatment of acute myeloid leukemia. *Leukemia* **2019**, *33*, 805–808, DOI: 10.1038/s41375-018-0249-z.

260. Brack, S.; Attinger-Toller, I.; Schade, B.; Mourlane, F.; Klupsch, K. et al. A Bispecific HER2-Targeting FynomAb with Superior Antitumor Activity and Novel Mode of Action. *Molecular Cancer Therapeutics* **2014**, *13*, 2030–2039, DOI: 10.1158/1535-7163.mct-14-0046-t.

261. Silacci, M.; Lembke, W.; Woods, R.; Attinger-Toller, I.; Baenziger-Tobler, N. et al. Discovery and characterization of COVA322, a clinical-stage bispecific TNF/IL-17A inhibitor for the treatment of inflammatory diseases. *mAbs* **2016**, *8*, 141–149, DOI: 10.1080/19420862.2015.1093266.

262. Wuellner, U.; Klupsch, K.; Buller, F.; Attinger-Toller, I.; Santimaria, R. et al. Bispecific CD3/HER2 Targeting FynomAb Induces Redirected T Cell-Mediated Cytolysis with High Potency and Enhanced Tumor Selectivity. *Antibodies* **2015**, *4*, 426–440, DOI: 10.3390/antib4040426.

263. Muyldermans, S.; Baral, T. N.; Retamozzo, V. C.; Baetselier, P. D.; Genst, E. D. et al. Camelid immunoglobulins and nanobody technology. *Veterinary Immunology and Immunopathology* **2009**, *128*, 178–183, DOI: 10.1016/j.vetimm.2008.10.299.

264. Muyldermans, S. Nanobodies: natural single-domain antibodies. *Annual Review of Biochemistry* **2013**, *82*, 775–797, DOI: 10.1146/annurev-biochem-063011-092449.

265. Dmitriev, O. Y.; Lutsenko, S.; Muyldermans, S. Nanobodies as Probes for Protein Dynamics in Vitro and in Cells. *The Journal of Biological Chemistry* **2016**, *291*, 3767–3775, DOI: 10.1074/jbc.R115.679811.

266. Meyer, T. D.; Muyldermans, S.; Depicker, A.; De Meyer, T.; Muyldermans, S. et al. Nanobody-based products as research and diagnostic tools. *Trends in Biotechnology* **2014**, *32*, 263–270, DOI: 10.1016/j.tibtech.2014.03.001.

267. Wang, Y.; Fan, Z.; Shao, L.; Kong, X.; Hou, X. et al. Nanobody-derived nanobiotechnology tool kits for diverse biomedical and biotechnology applications. *International Journal of Nanomedicine* **2016**, *11*, 3287–3303, DOI: 10.2147/IJN.S107194.

268. Hu, Y.; Liu, C.; Muyldermans, S. Nanobody-Based Delivery Systems for Diagnosis and Targeted Tumor Therapy. *Frontiers in Immunology* **2017**, *8*, 1442, DOI: 10.3389/fimmu.2017.01442.

269. Vincke, C.; Loris, R.; Saerens, D.; Martinez-Rodriguez, S.; Muyldermans, S. et al. General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *The Journal of Biological Chemistry* **2009**, *284*, 3273–3284, DOI: 10.1074/jbc.M806889200.

270. Moutel, S.; Bery, N.; Bernard, V.; Keller, L.; Lemesre, E. et al. NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies. *eLife* **2016**, *5*, 1–31, DOI: 10.7554/eLife.16228.

271. Huang, H.; Wu, T.; Shi, H.; Wu, Y.; Yang, H. et al. Modular design of nanobody-drug conjugates for targeted-delivery of platinum anticancer drugs with an MRI contrast agent. *Chemical Communications* **2019**, *55*, 5175–5178, DOI: 10.1039/c9cc01391a.

272. Xie, Y. J.; Dougan, M.; Jailkhani, N.; Ingram, J.; Fang, T. et al. Nanobody-based CAR T cells that target the tumor microenvironment inhibit the growth of solid tumors in immunocompetent mice. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 7624–7631, DOI: 10.1073/pnas.1817147116.

273. Walsham, N. E.; Sherwood, R. A. Fecal calprotectin in inflammatory bowel disease. *Clinical and Experimental Gastroenterology* **2016**, *9*, 21–29, DOI: 10.2147/CEG.S51902.

274. Heida, A.; Park, K. T.; van Rheenen, P. F. Clinical Utility of Fecal Calprotectin Monitoring in Asymptomatic Patients with Inflammatory Bowel Disease. *Inflammatory Bowel Diseases* **2017**, *23*, 894–902, DOI: 10.1097/mib.0000000000001082.

275. Rokkas, T.; Portincasa, P.; Koutroubakis, I. E. Fecal calprotectin in assessing inflammatory bowel disease endoscopic activity: a diagnostic accuracy meta-analysis. *Journal of Gastrointestinal and Liver Diseases* **2018**, *27*, 299–306, DOI: 10.15403/jgld.2014.1121.273.pti.

276. Rubin, D. T.; Ananthakrishnan, A. N.; Siegel, C. A.; Sauer, B. G.; Long, M. D. ACG Clinical Guideline: Ulcerative Colitis in Adults. *The American Journal of Gastroenterology* **2019**, *114*, 384–413, DOI: 10.14309/ajg.0000000000000152.

277. Mak, W. Y.; Buisson, A.; Andersen, M. J.; Lei, D.; Pekow, J. et al. Fecal Calprotectin in Assessing Endoscopic and Histological Remission in Patients with Ulcerative Colitis. *Digestive Diseases and Sciences* **2018**, *63*, 1294–1301, DOI: 10.1007/s10620-018-4980-0.

278. Theede, K.; Holck, S.; Ibsen, P.; Kallemose, T.; Nordgaard-Lassen, I. et al. Fecal Calprotectin Predicts Relapse and Histological Mucosal Healing in Ulcerative Colitis. *Inflammatory Bowel Diseases* **2016**, *22*, 1042–1048, DOI: 10.1097/MIB.0000000000000736.

279. Peyrin-Biroulet, L.; Sandborn, W.; Sands, B. E.; Reinisch, W.; Bemelman, W. et al. Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): Determining Therapeutic Goals for Treat-to-Target. *The American Journal of Gastroenterology* **2015**, *110*, 1324–1338, DOI: 10.1038/ajg.2015.233.

280. Labaere, D.; Smismans, A.; Olmen, A. V.; Christiaens, P.; D'Haens, G. et al. Comparison of six different calprotectin assays for the assessment of inflammatory bowel disease. *United European Gastroenterology Journal* **2014**, *2*, 30–37, DOI: 10.1177/2050640613518201.

281. Freeman, K.; Willis, B. H.; Fraser, H.; Taylor-Phillips, S.; Clarke, A. Faecal calprotectin to detect inflammatory bowel disease: a systematic review and exploratory meta-analysis of test accuracy. *BMJ Open* **2019**, *9*, e027428, DOI: 10.1136/bmjopen-2018-027428.

282. P121. Patients' point of view regarding acceptability and usefulness of inflammatory bowel diseases monitoring tools: results from a nationwide multicentre study (the ACCEPT study). Journal of Crohn's and Colitis. 2016; p S149, DOI: 10.1093/ecco-jcc/jjw019.240.

283. Heida, A.; Knol, M.; Kobold, A. M.; Bootsman, J.; Dijkstra, G. et al. Agreement Between Home-Based Measurement of Stool Calprotectin and ELISA Results for Monitoring Inflammatory Bowel Disease Activity. *Clinical Gastroenterology and Hepatology* **2017**, *15*, 1742–1749, DOI: 10.1016/j.cgh.2017.06.007.

284. Hejl, J.; Theede, K.; Møllgren, B.; VikkelsøMadsen, K.; Heidari, A. et al. Point of care testing of fecal calprotectin as a substitute for routine laboratory analysis. *Practical Laboratory Medicine* **2018**, *10*, 10–14, DOI: 10.1016/j.plabm.2017.11.002.

285. Bello, C.; Roseth, A.; Guardiola, J.; Reenaers, C.; Ruiz-Cerulla, A. et al. Usability of a home-based test for the measurement of fecal calprotectin in asymptomatic IBD patients. *Digestive and Liver Disease* **2017**, *49*, 991–996, DOI: 10.1016/j.dld.2017.05.009.

286. Ozawa, T.; Kaihara, A.; Sato, M.; Tachihara, K.; Umezawa, Y. Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Analytical Chemistry* **2001**, *73*, 2516–2521, DOI: 10.1021/ac0013296.

287. Blakeley, B. D.; Chapman, A. M.; McNaughton, B. R. Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein-protein interactions in vivo. *Molecular bioSystems* **2012**, *8*, 2036–2040, DOI: 10.1039/c2mb25130b.

288. Zinkus-Boltz, J.; DeValk, C.; Dickinson, B. C. A Phage-Assisted Continuous Selection Approach for Deep Mutational Scanning of Protein–Protein Interactions. *ACS Chemical Biology* **2019**, *14*, 2757–2767, DOI: 10.1021/acschembio.9b00669.

289. Braun, M. B.; Traenkle, B.; Koch, P. A.; Emele, F.; Weiss, F. et al. Peptides in headlock–a novel high-affinity and versatile peptide-binding nanobody for proteomics and microscopy. *Scientific Reports* **2016**, *6*, 19211, DOI: 10.1038/srep19211.

290. Hessian, P. A.; Fisher, L. The heterodimeric complex of MRP-8 (S100A8) and MRP-14 (S100A9). Antibody recognition, epitope definition and the implications for structure. *European Journal of Biochemistry* **2001**, *268*, 353–363, DOI: 10.1046/j.1432-1033.2001.01894.x.

291. Futami, J.; Atago, Y.; Azuma, A.; Putranto, E. W.; Kinoshita, R. et al. An efficient method for the preparation of preferentially heterodimerized recombinant S100A8/A9 coexpressed in Escherichia coli. *Biochemistry and Biophysics reports* **2016**, *6*, 94–100, DOI: 10.1016/j.bbrep.2016.03.009.