

Molecular engineering techniques for *in vitro* display technologies

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

Ayako Ohoka

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

Advisor: Casim A. Sarkar

February 2023

© Ayako Ohoka 2023

## Acknowledgements

---

My most sincere appreciation goes to my advisor, Casim, for shaping me into the scientist I am today. Your creative, out-of-the-box approach to asking questions and thinking about problems has helped me tremendously in pushing me out of my comfort zone and cultivating my curiosity and interest. You are also a phenomenal writer; I have learned so much from writing with you and will continue to polish my writing skills! Beyond grad school, I want to thank you for your continued support throughout all the other challenges that the past few years have brought, and for creating an inclusive and safe environment for all of us.

I have many great mentors to thank from the Sarkar Lab: Igor, I am so grateful to have been trained by you on countless skills and techniques, and for challenging me to persevere through tough experiments. To Wes, George, Fred, and Qiuge – thank you for training me on various techniques and helping me troubleshoot so many experiments. Paulina, I am so thankful to have worked on the TurboID project with you; our weekly talks (especially with Goobs supervising) have been such a delight and I've learned so much from working with you. Zi, Harish, Victor, Jennifer, Kristen, Kaylie, Riley – I've had so many fun memories spending time with you and wish you all the very best.

To my amazing parents – I cannot thank you enough for giving me the opportunity to pursue higher education and carve my own path. Living far away has been really tough at times but knowing that you two are cheering me on and continuing to support me has and will always keep me moving forward.

To my dear husband, Scott, for just about everything. This has been the hardest thing I've ever done so far, and from my best days to my worst days, coming home to you is always the highlight of my day. I appreciate you so much for all of the sacrifices that

you've made to let me pursue my grad school journey and fulfill my dreams of becoming a scientist. Thank you for taking care of me and being in my corner; I could not have done any of this without you.

To my siblings, Atsushi and Yuiko, for setting the bar high and inspiring me to continue working hard. Your encouragement and support (and love, I guess) have pushed me even when things are challenging. To all of my extended family in Japan, thank you for your continued support, and I hope to see you all soon.

To all of Scott's family, for your kindness, warmth, and encouragement throughout all of these years – I am so thankful to be a part of your family and am truly ecstatic to be moving close by!

To my BioE pals – LG, Mike, LP, Mallory, Rak, Ryan, and Blake – you guys are some of the coolest and smartest people I know, and I'm so grateful to call you my friends.

To Bowie, my favorite (cat) child – obviously you can't read this but thank you for always comforting me and being my constant companion.

And finally, thank you to my graduate committee – Paolo, Ben, and Brittany – for being a part of my graduate school journey, and to my funding sources – National Science Foundation Graduate Research Fellowship Program, University of Minnesota Doctoral Dissertation Fellowship, and various National Institutes of Health grants – for allowing me to conduct the research in this thesis.

## **Dedication**

---

To my incredible parents, Miharuru and Junko, for their everlasting support, love, and encouragement throughout all of my educational pursuits.

## Abstract

---

Biologics, which encompass peptides, proteins, nucleotides, and cell-based products are widely used in both therapeutics and diagnostic applications. Compared to small molecule drugs, biologics can achieve higher efficacy with lower systemic adverse side effects due to their complex three-dimensional structures which allow for better target selectivity and higher affinity. Despite the importance of biologics in human health, their discovery remains highly challenging, costly, and often unfruitful. In the past four decades, there has been an emergence of display technologies, in which a genotype (e.g., RNA) is connected to its corresponding phenotype (e.g., protein) and the library is typically screened or selected for binding against a target of interest. While these display technologies can streamline the discovery process since the amplification and readout of selected proteins can be carried out at the genetic level, current approaches could be further improved to address challenges in the ability to (1) perform selections in complex environments like cell surfaces and within a naïve library and (2) enrich for a more diverse pool of recovered binders, including ones with low-affinity that are typically lost.

The molecular engineering approaches and experimental protocols developed in this thesis work improve the signal-to-noise for the specific enrichment of protein binders in an array of biologics discovery platforms. Towards minimizing noise from nonspecific adsorption, we developed PEGylation strategies for stealthing DNA templates for biopanning applications in DNA display. Further, to enrich recovery of true binders and reduce noise from nonfunctional and nonspecific library binders, we introduced methods to incorporate photocrosslinking of the binder library to its target in mRNA display. To improve signal via higher binder recovery of low-affinity binders, we developed a facile cloning method that leverages rolling circle amplification to create homomultivalent protein

libraries for *in vitro* display technologies. Finally, in our collaborative protein work, we made efforts toward engineering biotin ligase fusions to identify new therapeutic targets and understand disease biology with proximity ligation using a blood-brain barrier cell model. Altogether, these advances in molecular engineering contribute to the broader toolbox for the discovery of new biologics.

## Table of Contents

---

List of Tables.....	ix
List of Figures.....	x
Chapter 1 : Introduction .....	1
1.1. Great Expectations: the growing field of biologics discovery .....	1
1.2. The role of directed evolution and display technologies for engineering biologics .....	2
1.3. <i>In vitro</i> display technologies for protein engineering applications.....	5
1.4. Beyond soluble antigens: performing affinity selections in complex environments .....	7
1.5. The diversity principle: enabling recovery of a diverse pool of binders in library selections	9
1.6. Age of discovery: developing new methods to identify new protein-protein interactions..	11
1.7. Overview of thesis work .....	12
Chapter 2 : Bioconjugation approaches for stealthing a DNA display particle for reduced nonspecific binding .....	15
2.1. Summary.....	15
2.2. Introduction .....	16
2.3. Results .....	19
2.3.1. Strategies for incorporating modifiable positions throughout a DNA template .....	20
2.3.2. Click-mediated bioconjugation techniques for functionalization of dUTPs .....	24
2.3.3. Biopanning strategies with PEGylated DNA templates .....	28
2.3.4. Analysis of non-specifically adsorbed DNA templates by qPCR .....	29
2.4. Discussion.....	31
2.5. Conclusion .....	37
2.6. Materials and Methods.....	37
2.6.1. T-spacing test .....	37
2.6.2. dUTP primer extension .....	38
2.6.3. Cell-free expression tests for primer-extended DNA templates.....	39
2.6.4. SPAAC bioconjugation experiments.....	39
2.6.5. CuAAC bioconjugation experiments and purification.....	40
2.6.6. Cell culture .....	41
2.6.7. Biopanning.....	41
2.6.8. Probe qPCR.....	41
Chapter 3 : Selection by photoaffinity crosslinking in complex environments (SPACE) .....	45
3.1. Acknowledgement and contribution .....	45
3.2. Summary.....	45
3.3. Introduction .....	46
3.4. Results .....	51
3.4.1. Summary of previous contributions .....	51
3.4.1.1. Synthetic nanobody library design.....	51
3.4.1.2. mRNA particle preparation .....	52
3.4.2. Development and optimization of library-target photocrosslinking conditions .....	54
3.4.3. Initial affinity selections and single clone screens of the nanobody library against model protein MBP .....	57
3.4.4. Efforts toward additional selections against MBP and other targets with second-generation nanobody library .....	59



3.4.5. Proof-of-principle experiments demonstrate successful crosslinking and recovery of mRNA displayed particle to target .....	61
3.5. Discussion.....	62
3.6. Conclusion .....	66
3.7. Materials and Methods.....	66
3.7.1. Expression and purification of engineered MRS5m tRNA synthetase .....	66
3.7.2. Expression and purification of mRNA display targets MBP-SNAP and sfGFP-SNAP .....	67
3.7.3. Biotinylation of purified MBP-SNAP and sfGFP-SNAP targets.....	68
3.7.4. Nanobody library construction .....	68
3.7.5. DARPIn cloning for proof-of-concept experiments.....	68
3.7.6. In vitro transcription .....	68
3.7.7. Preparation of mRNA for mRNA display.....	69
3.7.8. Modified PURExpress reactions .....	70
3.7.9. Affinity selection for nanobody library .....	72
3.7.10. RT-PCR and library reformatting .....	73
3.7.11. Single clone binding assay with MBP .....	74
3.7.12. Control affinity selections with Off7 <sub>D112M</sub> / WT and MBP.....	75
3.7.13. Preparation of cell lysate for panning .....	76
3.7.14. SDS-PAGE gel-shift crosslinking assay with Off7 <sub>D112M</sub> and MBP .....	76
Chapter 4 : Multivalent <i>in vitro</i> display and selection (MIDAS) .....	82
4.1. Summary.....	82
4.2. Introduction .....	83
4.3. Results .....	87
4.3.1. Strategies for rolling circle amplification of dsDNA template and mock DNA display particle synthesis and purification.....	87
4.3.2. Biopanning and analysis of binder-conjugated DNA templates.....	91
4.3.3. Modeling of bivalent binding kinetics using MVsim.....	93
4.3.4. Initial proof-of-principle panning experiments with SLP76 against Grb2 and Gads .	95
4.3.5. Modified RCA-enabled multivalent library synthesis and efforts toward affinity selections against monovalent or bivalent target.....	97
4.3.6 Proof-of-principle validation of bivalent RD particle binding through control selection .....	100
4.4 Discussion.....	102
4.4.1 RCA-enabled synthesis of in vitro multivalent libraries.....	102
4.4.2 Extended commentary and perspective: General design principles and considerations for multivalent therapeutics.....	105
4.5 Conclusion .....	111
4.6 Materials and Methods.....	111
4.6.1 Circularization and rolling circle amplification .....	112
4.6.2 Plasmid construction and protein expression .....	112
4.6.3 Preparation of mock DNA display particles .....	113
4.6.4 Biopanning.....	114
4.6.5 Library and control template cloning.....	115
4.6.6 Ribosome display .....	115
4.6.7 RT-qPCR .....	116

Chapter 5: TurboA $\beta$ : Proximity biotinylation approaches for identifying amyloid beta interactomes at the blood-brain barrier.....	122
5.1 Acknowledgement and contribution.....	122
5.2 Summary.....	122
5.3 Introduction.....	123
5.4.3 Protein characterization assays show functional activity of TurboID and some aggregation of TurboA $\beta$ constructs.....	132
5.4.4 Native PAGE assay shows aggregation of TurboA $\beta$ <sub>WT</sub> and TurboA $\beta$ <sub>F20P</sub> after 37 °C incubation.....	133
5.4.5 Characterization of known A $\beta$ binders in hCMEC/d3 model system and panning of TurboA $\beta$ constructs.....	134
5.4.6 In vitro assays for assessing TurboA $\beta$ functionality show lack of A $\beta$ -specific labeling.....	136
5.4.7 Current efforts toward creating new targeting peptide fusions.....	137
5.5 Discussion.....	140
5.6 Conclusion.....	143
5.7 Materials and Methods.....	144
5.7.1 Molecular cloning of TurboID-fusion constructs.....	144
5.7.2 Expression and purification of TurboID/miniTurboID fusion constructs.....	145
5.7.3 In vitro biotinylation test for assessing intramolecular labeling activity.....	147
5.7.4 Native PAGE assay.....	147
5.7.5 Additional experiments performed by Paulina Eberts.....	148
5.7.5.1 Immunostaining for confirming LRP1 and RAGE expression.....	148
5.7.5.2 TurboA $\beta$ validation on hCMECs via immunofluorescence.....	149
5.7.5.3 Cell-free assay for assessing Ab-TurboID fusion construct activity.....	149
5.7.5.4 Optimized cell-free assay for assessing activity of additional peptide-TurboID fusion constructs.....	150
Chapter 6: Conclusion, future directions, and outlook.....	154
Bibliography.....	159
Appendix: Additional experimental data for Chapter 3.....	172
A.1 Results and discussion.....	172
A.1.1 Efforts toward nanobody expression optimization prior to library selection.....	172
A.1.2 Preparation of cell surface targets for biopanning and protocol development for C-terminal tag pulldown onto solid support.....	177
A.2 Materials and Methods.....	179
A.2.1. Plasmids for cell transformations and transfections.....	179
A.2.2. Lemo21 expression test.....	179
A.2.3. Lemo21 expression, periplasmic extraction, and purification.....	180
A.2.4. T7 SHuffle expression and purification.....	181
A.2.5. Pulldown of SHuffle-expressed proteins onto MBP-beads.....	181
A.2.6. Transfection procedure.....	182
A.2.7. Pulldown of target proteins in the membrane fraction on BG-beads.....	183

## List of Tables

---

Table 1-1. Summary of various display technologies used for screening in directed evolution campaigns.....	4
Table 2-1. Calculations for PEG molecules used in this study. ....	35
Table 2-2. Synthetic templates used for thymine modification experiments. ....	43
Table 2-3. Primers used for qPCR.....	43
Table 3-1. Annotated nanobody library composition/diversity used for mRNA display.....	78
Table 3-2. List of primers used for mRNA display. ....	79
Table 3-3. List of amino acid sequences for proteins used in library and control selections. ....	80
Table 3-4. List of DNA sequences for Off7 control selections. ....	81
Table 4-1. Selected recently published multivalent therapeutic binders against trimeric Spike protein. ....	107
Table 4-2. List of amino acid sequences for proteins used in library and control selections. ....	117
Table 4-3. List of primers and gBlocks used in this study.....	118
Table 4-4. Comparison of estimated yields of DNA input for traditional ribosome display and homobivalent ribosome display.....	121
Table 5-1. Comparison of different proximity labeling enzymes. ....	126
Table 5-2. Repertoire of new peptide-TurboID fusions. ....	138
Table 5-3. Extended list of amino acid sequences used for creating TurboID fusion constructs.	151
Table 5-4. List of primers and gBlocks used for oligo-annealing, overlap-extension PCR, and subcloning.....	151
Table A-1. List of AA sequences used in this study.....	184

## List of Figures

---

Figure 1-1. Directed evolution cycle.....	3
Figure 2-1. Workflow for a DNA display panning selection.....	19
Figure 2-2. Gel electrophoresis analysis of azide-dUTP primer extension reaction parameters. .	21
Figure 2-3. sfGFP DNA templates used for IVTT expression tests. ....	22
Figure 2-4. Comparison of IVTT efficiency of modified and unmodified templates.....	23
Figure 2-5. Cycloalkyne and azide SPAAC experiments.....	25
Figure 2-6. PEGylation reaction products of a 427-bp DNA template. ....	28
Figure 2-7. Data summary for biopanning qPCR.....	31
Figure 3-1. Structural comparison of methionine to L-photo-methionine.....	49
Figure 3-2. Nanobody structure.....	50
Figure 3-3. Workflow for a SPACE display selection round.....	51
Figure 3-4. Additional mRNA preparation prior to selection.....	53
Figure 3-5. Comparison of photocrosslinking conditions.....	56
Figure 3-6. Single clone screens of binders recovered from library selection.....	58
Figure 3-7. Binding assay with selected clones.....	59
Figure 3-8. RT-PCR and single clone data from second library selection.....	60
Figure 3-9. Proof-of-concept demonstration of crosslinking-enriched binder recovery.....	62
Figure 4-1. Workflow for preparing a multivalent in vitro display particle.....	87
Figure 4-2. Circularization of dsDNA.....	88
Figure 4-3. Summary of steps for bioconjugation of protein to control n5T DNA template.....	91
Figure 4-4. Comparison of particle recovery per cell for G3-conjugated DNA templates.....	93
Figure 4-5. Dissociation kinetics from MVsim simulations.....	95
Figure 4-6. RD-qPCR data from control selections against monovalent or bivalent target.....	96
Figure 4-7. Circularization of ssDNA and overview of peptide library cloning.....	98
Figure 4-8. RT-PCR and RT-qPCR data from library selections.....	100
Figure 4-9. Test selections of SLP76 using our bivalent RD system.....	102
Figure 5-1. Workflow for a TurboAb panning assay.....	128
Figure 5-2. Amyloid beta structures.....	130
Figure 5-3. TurboA $\beta$ expression and purifications.....	132
Figure 5-4. In vitro biotinylation assay and analysis by WB.....	133
Figure 5-5. Native PAGE analysis of protein aggregates.....	134
Figure 5-6. Z-stack confocal images for surface receptor characterization on BBB model.....	135
Figure 5-7. Cell images showing treatment of hCMEC/d3 with Turbo constructs.....	136
Figure 5-8. Biotin labeling activity of TurboID constructs in <i>in vitro</i> assays.....	137
Figure 5-9. <i>In vitro</i> assay to assess nonspecific binding across different blocking conditions....	139
Figure A-1. Comparison of nanobody expression conditions.....	174
Figure A-2. Nanobody expression with T7 SHuffle and binding assay.....	176
Figure A-3. Validation of cell-surface target expression by flow cytometry.....	177
Figure A-4. Pulldown assays with enriched membrane fractions.....	179

## Chapter 1: Introduction

---

### 1.1. Great Expectations: the growing field of biologics discovery

Biologics, which are therapeutics derived from living organisms or through biological processes, are large and complex molecules that include common drugs like antibodies, enzymes, fusion proteins, and cell-based therapies.<sup>1</sup> Compared to chemically synthesized small molecule drugs, biologics can achieve higher efficacy with lower systemic adverse side effects due to their complex three-dimensional structures which allow for better target selectivity and higher affinity. The discovery of recombinant DNA technology, which enabled construction of DNA plasmids using restriction endonucleases, in 1973<sup>2</sup> and hybridoma technology, a technique used for production of antibodies by fusing antibody-producing cells with immortalized cancer cells, in 1975<sup>3</sup> are among several notable advancements in biotechnology that have revolutionized the development of biologics by allowing researchers to create them without the use of whole animals.

As a result, the share of biologics among novel drugs approved by the Food and Drug Administration (FDA) has grown over the last several decades.<sup>4</sup> In 2022, biologics license applications (BLAs) accounted for 41% (15 out of 37) of FDA drug approvals, the highest percentage in history.<sup>4</sup> Despite their importance in human health and potential impacts for treatment of a wide range of diseases, biologics discovery remains highly challenging and often unfruitful. The large molecular size and its inherent complexity requires arduous tasks like lengthy clinical trials and complex manufacturing processes, which amount to an average FDA approval timeline of ~12 years, with costs estimated around \$1 - \$1.8 billion USD per drug.<sup>5,6</sup>

On average, a majority of approved biologics are monoclonal antibodies (mAb) (~59%, from 2017 - 2022<sup>7-11,4</sup>), which are primarily discovered by hybridoma technology.

While successful, hybridoma technology requires immunization of animals with soluble antigens which greatly limits the pool of potential candidates, as many important therapeutic targets like membrane proteins cannot retain structural integrity outside the membrane. In addition, this technique further restricts therapeutic candidates to antibodies and antibody fragments, while other non-immunoglobulin protein scaffolds like affibodies<sup>12</sup> and designed ankyrin repeat proteins (DARPin)<sup>13,14</sup> may be of interest due to smaller size, which allows desirable qualities like easier manufacturing capabilities, better tissue penetration, and higher stability.<sup>15,16</sup>

## **1.2. The role of directed evolution and display technologies for engineering biologics**

For engineering of both antibody-based and alternative protein scaffolds, directed evolution has proven to be a highly robust and effective method that is broadly applicable to an array of ligands and targets. Mimicking natural evolutionary processes, directed evolution begins with a gene that encodes for a protein (e.g., DARPin) or a peptide that is diversified by random mutagenesis methods like error-prone PCR, or a semi-rational design, which uses structural and sequence information to build libraries based on consensus design, with mutations introduced at residues involved in putative binding or active sites (i.e., “hot spots”).<sup>17,18</sup> Then, the gene is translated to protein and screened or selected for improved biophysical or functional properties such as higher binding affinity or novel binding capability by applying user-defined selection pressures, and put into additional rounds of selection to further tune desirable qualities (Figure 1-1).<sup>19</sup> Screening or selection of large protein libraries ( $> 10^8$  molecules) is often needed to recover mutants that have achieved maximal fitness, and requires high-throughput methods.

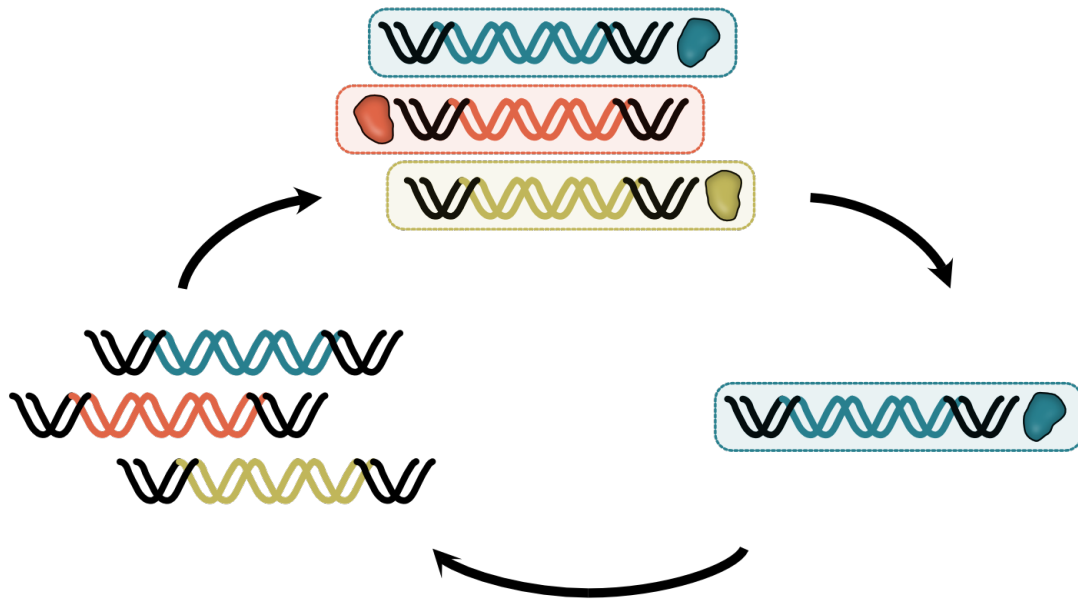


Figure 1-1. Directed evolution cycle.

(A) The user begins with a cDNA library, either from diversification of a protein-encoding gene (optional) or a semi-rational consensus design library with mutations introduced at strategic positions. (B) The gene is translated into corresponding proteins, then screened or selected for functional variants. (C) Upon exposure to various selection pressures, the pool with the desired properties is recovered, then replicated or amplified. The cycle, or round, is repeated until maximal fitness is achieved. (Figure adapted from Packer and Liu.<sup>19</sup>)

In the past four decades, display technologies in which a genotype (e.g., DNA or RNA) is connected to its corresponding phenotype (peptide or protein) to be panned against a protein of interest (POI) have emerged to streamline the screening process for large protein libraries since the amplification and readout of selected proteins can be carried out at the genetic level. Display technologies are often used for discovery of novel binders or affinity maturation of existing binders, and can be applied to many protein scaffolds.<sup>17</sup> Commonly used approaches include cell-based (*in vivo*) methods like phage display<sup>20</sup> and yeast display<sup>21</sup>, in which protein binders are displayed on the surface of a bacteriophage or yeast cell, as well as cell-free (*in vitro*) display methods like mRNA display<sup>22</sup> or ribosome display<sup>23</sup>, in which the genotype and phenotype are directly

associated through covalent linkages or via a tertiary (mRNA-ribosome-protein) translation complex (Table 1-1).

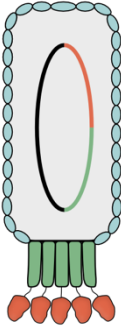
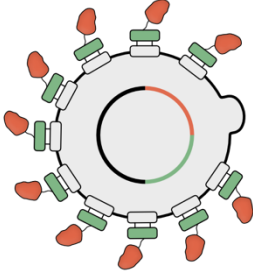


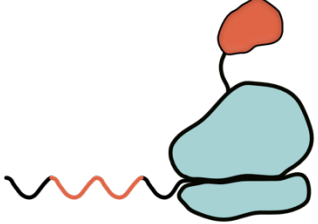
Display method	Max library size	Genotype-phenotype linkage	Ref.
Phage display 	$10^{11}$	Fusion to pIII, pVIII, or pIX coat protein	20,24
Yeast display 	$10^9$	Fusion to Aga2p surface protein, display via Aga1p-Aga2p disulfide linkage	21,25
DNA display 	$10^{10}$	Bioconjugation to DNA terminus	26
mRNA display 	$10^{14}$	Bioconjugation to mRNA terminus	22,27
Ribosome display 	$10^{14}$	mRNA-ribosome-protein ternary complex	23,28

Table 1-1. Summary of various display technologies used for screening in directed evolution campaigns.



In phage display, the gene encoding the binder library is typically fused to a coat protein such as pVIII or pIII in a plasmid and transformed into phage for binder display on the phage surface.<sup>20</sup> Similarly, in yeast display, the binder library is fused to a cell adhesion glycoprotein, Aga2p, and transformation allows for surface display of the binder library via disulfide bonds between Aga1p and Aga2p.<sup>21</sup> A unique advantage of yeast display, an eukaryotic system with similar protein expression pathways to those of mammalian cells, over other display methods is the conservation of some post-translational modifications and protein folding/secretory machinery.<sup>21,29</sup>

### **1.3. *In vitro* display technologies for protein engineering applications**

In contrast to the *in vivo* approaches described above, *in vitro* display systems typically use direct linkages between the genotype and phenotype, which allows for easy polymerase chain reaction (PCR)-based amplification and randomization directly after each round. In mRNA display, the 3'-terminus of the mRNA is ligated to small molecules that allow for covalent bioconjugation between the mRNA and the protein.<sup>21,30</sup> Puromycin, a naturally occurring antibiotic, is a commonly used molecule for covalent protein linkage in mRNA display. During translation, puromycin enters the ribosome and accepts the nascent polypeptide chain from the peptidyl-tRNA via ribosomal peptidyltransferase-catalyzed peptide bond formation, similar to other aa-tRNAs.<sup>31</sup> Instead of continued elongation, however, the puromycin incorporation terminates translation due to an uncleavable peptide bond (instead of the typical ester bond for aa-tRNAs), allowing the protein-puromycin product to bind covalently and causing the ribosome to disassemble.<sup>32</sup> This covalent linkage in mRNA display enables selections to be performed in more stringent conditions such as different pH, temperatures, and ionic strength<sup>33</sup> and has been

used for engineering of proteins like enzymes<sup>34</sup>, peptides<sup>35</sup>, and even libraries with non-natural amino acids.<sup>36</sup>

The binder library in ribosome display, another mRNA-based method, is C-terminally fused to a gene encoding an alpha-helical linker (i.e., spacer) and lacks a stop codon.<sup>37</sup> At the end of translation, the spacer pushes the protein out of the ribosomal tunnel and stalls due to omission of the stop codon and omission of polypeptide release factors and ribosome recycling factor, particularly in the PURE system.<sup>38</sup> The mRNA-ribosome-protein ternary complex is stabilized using  $Mg^{2+}$  and low temperature (4 °C), and used directly for panning experiments.<sup>23</sup> Unlike mRNA display, ribosome display lacks a covalent phenotype-genotype linkage which limits reaction conditions that can be used for selection experiments, although one notable study found that the complexes remain stable at 20 °C for several hours and can withstand incubations at 50 °C with only a ~2-fold loss in particles.<sup>39</sup> Further, the ribosome, which is primarily made of RNA, can also act as a solubility tag to prevent aggregation<sup>40</sup> and allow expression of complex proteins like leucine-rich repeats (LRRs).<sup>41</sup>

Finally, in DNA display, DNA libraries are terminally modified (either on the 5'- or 3'-end) with small molecules to allow for DNA-protein conjugation using interactions between biomolecules like biotin and streptavidin.<sup>26</sup> Protein-encoding genes are typically fused to protein tags like SNAP tags<sup>42</sup> or streptavidin<sup>43</sup> such that the expressed protein can link back to the DNA template via terminal modifications like benzylguanine (BG) or biotin, respectively. For particle synthesis, DNA templates are emulsified in droplets with *in vitro* transcription and translation (IVTT) mixtures, and the synthesized protein couples back to the template. In contrast to other *in vitro* methods, DNA display requires *in vitro* compartmentalization (i.e., emulsification in droplets) of the genetic material with translation machinery, which reduces its library size due to the high dilution factors needed

to ensure monoclonal occupancy in droplets.<sup>44</sup> However, the emulsification allows for a variety of bioconjugation techniques to be used; for example, the tetravalent biotin-streptavidin conjugation approach can display up to 4 copies of the protein.<sup>26</sup> In bead-surface display (BeSD), an extension of DNA display, encapsulation of beads in addition to the DNA and IVTT mix within the droplets enables up to  $10^6$  copies of proteins to be displayed.<sup>45</sup>

As outlined in the table, the main distinguishing factors between *in vivo* and *in vitro* are library size and linkage. *In vivo* methods typically have smaller library sizes (up to  $10^{11}$  molecules) due to transformation efficiency, while *in vitro* display methods (e.g., ribosome display) can accommodate extremely large protein libraries ( $10^{12}$  -  $10^{14}$  molecules).<sup>46</sup> In addition, although the direct linkage used by *in vitro* methods allow for small display particle sizes, they typically display only one copy of the binder (with exceptions for DNA display, as noted above); on the other hand, the genotype-phenotype linkage used by *in vivo* approaches enable display of tens or thousands of proteins on the cell surface.

Taken together, while *in vitro* display technologies have been used for successful selections, current approaches could be further improved to address shortcomings in the ability to (1) perform selections in complex environments like cell surfaces and within a large pool of naïve binders and (2) allow display of multiple protein domains to recover a diverse pool of selected sequences, including low-affinity binders.

#### **1.4. Beyond soluble antigens: performing affinity selections in complex environments**

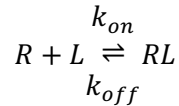
Much of biologics discovery and display methods have been performed against soluble target antigens such as cytokines<sup>47</sup> and large extracellular domains<sup>37</sup> of single-pass membrane proteins. Antigen preparation typically requires large scale recombinant

expression and purification in bacterial or mammalian systems, which is difficult for proteins like integral membrane proteins (IMPs) that are unable to retain their structure outside of the plasmid membrane despite their therapeutic potential as important regulators of various cell signaling events.<sup>48</sup> Interestingly, many biopanning approaches, in which selections are performed directly on live cell surfaces against membrane proteins, utilize *in vivo* display methods, most commonly phage display. Due to the complexity of the cell surface, however, subtractive screening (with negative selections on non-target expressing cells) and engineering of target proteins (e.g., fusions to fluorescent proteins, which will activate when bound) are needed to capture specific signal over nonspecific binding.<sup>49–52</sup> There are also approaches toward recapitulating cell membrane environments by addition of detergents or stabilization of the target proteins in lipid discs, but this is not broadly applicable for many targets that have multiple transmembrane regions such as G protein-coupled receptors (GPCRs) and require sophisticated lipid, cholesterol, and detergent compositions.<sup>53–56</sup>

A unique caveat to the large protein libraries that can be accommodated by *in vitro* display screens is that despite the increased probability of finding a high-affinity binder, more members must be screened in order to find true binders. While it is accepted that within a highly diverse library, the number of productive candidates may increase (as described by library *efficiency*, defined as the proportion of library members with desired therapeutic, manufacturing, and delivery properties), so does the number of library members with unfavorable qualities like hydrophobicity and instability (described as library *inefficiency*).<sup>57,58</sup> For example, an increase in the number of aromatic and arginine residues have been known to enhance nonspecific binding.<sup>59</sup> Thus, finding true binders in a noisy environment – in a pool of nonfunctional binders – is another challenge that current *in vitro* selection systems face.

### 1.5. The diversity principle: enabling recovery of a diverse pool of binders in library selections

The display technologies discussed thus far primarily select based on affinity, which is defined by how strongly a ligand binds to its receptor based on the simple association and dissociation model,



in which R is the receptor (i.e., target), L is the ligand, and RL is the receptor-ligand complex. The association rate constant,  $k_{on}$ , which defines the rate at which the receptor and ligand associate to form a receptor-ligand complex (units in  $\text{min}^{-1} \text{M}^{-1}$ ), is typically limited by encounter or diffusion. On the other hand,  $k_{off}$ , the dissociation rate constant, which is the rate at which the receptor-ligand complex dissociates (units in  $\text{min}^{-1}$ ), does not depend on local free ligand concentration, and is defined solely by the specific interaction between the receptor and the ligand. Together,  $k_{on}$  and  $k_{off}$  define  $K_D$  by the equation:

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

in which the  $K_D$ , the equilibrium dissociation constant, describes the affinity of a receptor and ligand under equilibrium conditions (units in M). The half-life of the receptor-ligand interaction is directly related to the dissociation constant and is described by<sup>60</sup>:

$$t_{\frac{1}{2}} = \frac{\ln(2)}{k_{off}}$$

For most selections, the display particles (i.e., ligand) are panned against the antigen (i.e., receptor or target) until the receptor-ligand complex is allowed to reach an equilibrium. Subsequently, any unbound or weakly bound ligands are washed away by replacing the buffer multiple times. By changing parameters such as target concentration and wash

duration, the selection stringency can be increased to select for the ligands that bind the strongest.

The affinity described above refers to a singular biomolecular (i.e., 'monovalent') interaction between a target and a ligand. Avidity, on the other hand, describes the synergistic effects of two or more (i.e., 'multivalent') interactions that enhance the apparent affinity.<sup>61</sup> In the context of multivalent biologics such as mAbs which are bivalent, the synergy results from increased local concentration due to proximity of the second ligand, which consequently increases the probability of a second target-ligand interaction. Overall, this leads to increased target residence time and effectively results in slower dissociation kinetics. In addition to monovalent affinities, valency (the number of ligands) and the linkage between binding domains are important features that control the dynamic network of multivalent binding.<sup>62,63</sup> As described in Section 1.3, *in vivo* approaches like phage display intrinsically take advantage of multivalency effects through their cell surface display; however, the valency and spatial arrangement of the displayed library members cannot be precisely controlled. Increased avidity is achieved only through the sheer number of attachment sites on a cell or bead surface, and thus may not exploit multivalency to their fullest potential.

Most *in vitro* display methods are monovalent, in which only one copy of the protein is linked to the genotype which can bias the selection toward high affinity binders, primarily dictated by having the slowest dissociation. This selection approach generally does not allow recovery of low-affinity binders, which are of therapeutic importance in some disease contexts such as cytokine-signaling which relies on dual high- and low-affinity binding for activation<sup>47</sup>, or tumor-targeting which greatly benefits from low-affinity binding for tissue penetration due to the binding-site barrier effect.<sup>64</sup> Together, we sought to address these

shortcomings in both *in vivo* and *in vitro* display methods for recovery of binders with weaker affinity.

### **1.6. Age of discovery: developing new methods to identify new protein-protein interactions**

As described in Section 1.2, most affinity selection campaigns involve panning a binder library against a POI. The discovery of therapeutically important POIs, much like the selections, is nontrivial and are identified through various techniques, such as from biomolecules involved in particular cellular signaling pathways or performing genome-wide association studies (GWAS) to study links between complex disease and genome variations.<sup>65</sup> Then, targets are experimentally verified by studying protein-protein interactions (PPIs) between the target and ligand by methods like affinity purification mass spectrometry (AP-MS), in which ligand-target interactions are enriched and isolated using affinity tags (either on the ligand or target) and identified through MS.<sup>66</sup> Facing challenges analogous to those described in previous sections, this approach cannot be broadly applied to membrane-bound targets or low affinity ligands due to instability of transmembrane targets outside their native contexts and the short half-life of low affinity interactions which can disrupt PPIs.

Proximity labeling, on the other hand, uses promiscuous labeling enzymes for covalent tagging of PPIs.<sup>67</sup> Proximity ligases are genetically fused to targeting ligands, which can be used to direct and spatially restrict tagging to relevant receptors and neighboring proteins, providing a quantitative “contour map” of the PPI in which the highest labeling occurs at the target with less labeling of adjacent proteins.<sup>68</sup> Although proximity labeling has primarily been used for intracellular applications, labeling at the cell surface could enable identification of possible transmembrane therapeutic targets involved in

extracellular signaling in their native, disease-relevant context for ligands with varying binding affinities.

### **1.7. Overview of thesis work**

The work presented in this thesis discusses molecular engineering techniques and panning approaches developed toward addressing current shortcomings in *in vitro* display technologies with proof-of-concept demonstrations and efforts made toward library selections. While varied in the display method used, all chapters introduce unique tools to modify or improve their respective display methods for broader, more diverse selection applications.

The first goal aimed to expand the capabilities of DNA (in Chapter 2) and mRNA displays (in Chapter 3) to perform selections in complex environments like cell surfaces and/or with a large naïve library. Chapter 2 describes bioconjugation approaches for stealthing a DNA display particle with polyethylene glycol (PEG), an inert polymer, to minimize nonspecific adsorption to the cell surface during panning experiments. In this project, we also developed novel approaches for purification of DNA display particles, biopanning on adherent cell surfaces, and analysis using quantitative polymerase chain reaction (qPCR) and demonstrated reduced background in proof-of-concept cell panning experiments. Chapter 3 discusses a new, modified mRNA display approach that incorporates photocrosslinking of the binder library to its target by using a photoactivatable analogue of the amino acid methionine. In this collaborative project, we further optimized a previously developed mRNA display protocol<sup>69</sup> to increase display particle yield, and introduced methods for designing a synthetic single-domain antibody (sdAb) library and performing crosslinking experiments. Although library selections were ultimately



unsuccessful in this project, proof-of-concept experiments with a known ligand-target interaction validate the experimental techniques developed.

The fourth chapter discusses efforts made toward our second goal of incorporating multivalency to recover a diverse pool of binders, including low affinity binders. Here, we developed a library synthesis method to genetically encode binders multivalently on a single template for broad applications in all *in vitro* display systems. We first confirmed its utility with a DNA display particle and developed corresponding cell panning methods, and used ribosome display to demonstrate its application in a selection system. Although we faced technical challenges with amplifications with library selections, control selections using binders with varying dissociation kinetics showed improved recovery of low affinity binders in ribosome display.

The final research chapter (Chapter 5) describes our work toward a related challenge for biologics discovery – the discovery of new therapeutic targets. For this project, we employed a biotin ligase enzyme, TurboID<sup>70</sup>, to engineer as a fusion construct to amyloid beta, a peptide implicated in the neurodegenerative disease, Alzheimer's disease (AD)<sup>71</sup>, with the goal of identifying new therapeutic targets and understanding disease biology with a blood-brain barrier cell model. The current data suggest issues with nonspecific labeling that need to be addressed for successful execution but continued work and future implementation of this approach can result in a novel approach toward target discovery and disease biology that is broadly applicable.

Overall, these chapters are tied together by an overarching goal of tackling important challenges in biologics discovery, which are identifying new therapeutic targets, being able to perform selections in therapeutically relevant contexts, and expanding the pool of binders that are recovered in a traditional selection system. The techniques

developed and optimized to address these shortcomings are also applicable to the greater molecular biology toolbox to be used as synthetic biology solutions for other challenges.

## Chapter 2: Bioconjugation approaches for stealthing a DNA display particle for reduced nonspecific binding

---

### 2.1. Summary

Most high-throughput selections for binder discovery using display technologies are performed against soluble targets in which the target proteins and antigens are recombinantly expressed, purified, and immobilized on a solid support. Many therapeutically important targets like cell surface receptors and other membrane-bound proteins, however, are difficult to express recombinantly as they rely on the unique cell membrane lipid composition for structural integrity and will often irreversibly denature and aggregate once taken out of the cell membrane. Further, these surface receptors often associate with cofactor proteins, have post-translational modifications, and are folded properly when presented in their native context within a live cell. While the most optimal approach is to perform panning experiments directly on the cell membrane, this is highly challenging due to the complexity of the cell surface. The selection environment is filled with non-target proteins, glycoproteins, and matrix proteins, which may lead to nonspecific adsorption of the selection particle.

In this project, we developed bioconjugation approaches to stealthing a DNA template with polyethylene glycol (PEG), a biologically inert polymer, to reduce nonspecific adsorption for downstream application in cell-surface DNA display selections and quantified nonspecific binding in control biopanning experiments. Our results showed that PEGylation of DNA templates reduces cell-binding in a modification-dependent manner, and that too much or too little PEGylation leads to higher adsorption with the most optimal condition from a moderate level of modifiable positions with high molecular weight PEG. Together, our results demonstrate a viable approach for modifying DNA selection particles and a new biopanning approach for directed evolution experiments.

## 2.2. Introduction

High-throughput screening of diverse gene libraries using display technologies has led to successful discovery of various biologics, including single-chain variable fragments (scFvs), designed ankyrin repeat proteins (DARPs), single-domain antibodies (i.e., VHH domains or nanobodies), and peptides.<sup>25,72</sup> In particular, affinity selections using these display methods typically involve (1) panning against a protein of interest (POI), (2) performing multiple wash cycles in which varying wash durations and buffer compositions are used, and (3) recovery of the strongest binding (i.e., slowest dissociating) library member for additional screening rounds or sequencing.<sup>37</sup> Despite their utility, many directed evolution campaigns are limited to binder selections against targets such as proteins that are inherently soluble (e.g., cytokines) or have large soluble extracellular domains<sup>72-74</sup>; selections against cell surface receptors and other integral membrane proteins (IMPs) are extremely challenging due to their low solubility in aqueous environments and instability outside of the lipid bilayer.<sup>75</sup> However, binder discovery against membrane proteins are highly desirable due to their immense therapeutic potential.

Membrane proteins, embedded in the surface of cells, are estimated to comprise only ~23% of the human proteome yet account for more than 60% of current drug targets.<sup>76</sup> Their wide array of functions, from transduction of signals into cells, transport of ions and other molecules, and modulation of biochemical cues between extracellular and intracellular environments make membrane proteins critical regulators of cell function and consequently, important drug targets as these processes are often implicated in various disease pathways.<sup>77</sup> Unsurprisingly, ~61% of all 127 antibodies marketed in the U.S. or E.U. target membrane proteins, although most of these antibodies target membrane proteins which have a simple transmembrane domain with large extracellular domains (ECDs), such as tyrosine kinase receptors.<sup>78</sup> Only three target more complex membrane

proteins, specifically G-protein coupled receptors (GPCRs)<sup>78</sup>, even though GPCR-targeting drugs account for ~35% of all FDA-approved drugs.<sup>79</sup>

Despite their therapeutic potential, ligand discovery against membrane proteins is still highly challenging and often unfruitful. Recombinant overexpression and purification of target proteins, commonly employed for antigen production for immunizations and selection panning, are difficult for membrane proteins as their surfaces are relatively hydrophobic and prone to denaturation and subsequent aggregation.<sup>80</sup> Further, they are often flexible and unstable, and may undergo post-translational modifications (e.g., glycosylation) which is generally achievable only by eukaryotic cells.<sup>80</sup> While there are successful examples of recombinantly expressed membrane protein stabilization through the addition of detergents and lipids to create membrane mimetic nanostructures like micelles and nanodiscs<sup>81</sup>, these can be technically challenging and not feasible for highly complex proteins with multiple transmembrane domains like GPCRs.<sup>82</sup> Lastly, truncations of soluble extracellular domains or additional stabilizing mutations may not be ideal, as native epitopes may not be preserved.<sup>50</sup>

With challenges in recombinant expression and optimal presentation most favorably achieved in the physiologically relevant lipid bilayer, library selections directly on cell surface membranes are highly desirable. However, although cell surface panning is an ideal selection format, cell surfaces are complex panning environments filled with non-target proteins, matrix proteins, and charged biomolecules like glycans, which increases the likelihood of nonspecific adsorption and other non-target interactions. In addition, many selection particles may be inherently too “sticky”; cell-based display particles such as bacteriophages and yeast cells are large in size and coated in native surface proteins, and even cell-free *in vitro* display particles like those used in mRNA and ribosome display have large mRNA molecules which may be sticky due to secondary structure and the

charged nature of mRNA. Indeed, many existing biopanning approaches require extensive washing steps and/or subtractive pre-panning steps in which the display particle is first panned against a non-target expressing cell, and the remaining pool is then panned against the target-expressing cell.<sup>50</sup> This not only requires the user to know their target *a priori*, but also limits the target protein space to membrane proteins that are able to be transfected into cell lines, which can be challenging depending on the complexity of the protein.<sup>83</sup> Finally, this pre-panning step can artificially decrease the library diversity if the display particle, but not the expressed binder, is adsorbing, which is possible for current display particles.

Given the library size limitations imposed by cell transformation in *in vivo* display methods and potential stickiness from large mRNA molecules used in ribosome and mRNA display, we aimed to modify a DNA display selection particle for panning on the live cell surface, a highly complex environment. Despite their relative advantage, DNA may still tend to nonspecifically adsorb to the cell surface, as exploited in cation-mediated DNA gene delivery<sup>84,85</sup> and cell transformations.<sup>86,87</sup> Thus, to minimize potential nonspecific adsorption of the DNA to the cell or panning surface and protect the template from potential nuclease activity as released by cells, we employed stealthing strategies on the DNA template to decrease noise and allow better recovery of meaningful binders. As the template which codes for its respective protein is much bigger than the protein itself (i.e., a 644 kDa template for a ~36 kDa protein for 330 amino acids) and may thus be more likely to interact with the cell surface when biopanned, we aimed to stealth the template throughout the entire structure. Our approach for stealthing is to decorate the template with methoxy-polyethylene glycol (mPEG) molecules that are covalently attached throughout the DNA sequence. PEG is often used to stealth nanoparticles, reduce protein binding, and protect biologics from degradation and is a suitable choice for selection on

cell surfaces.<sup>88</sup> For optimization of stealthing approaches, we systematically tested various bioconjugation techniques and varying degrees of modification. We further developed a biopanning strategy and optimized a subsequent quantitative polymerase chain reaction (qPCR) approach for analysis. Together, our results demonstrate a viable approach for modifying DNA selection particles and a new biopanning approach for directed evolution panning experiments (Figure 2-1).

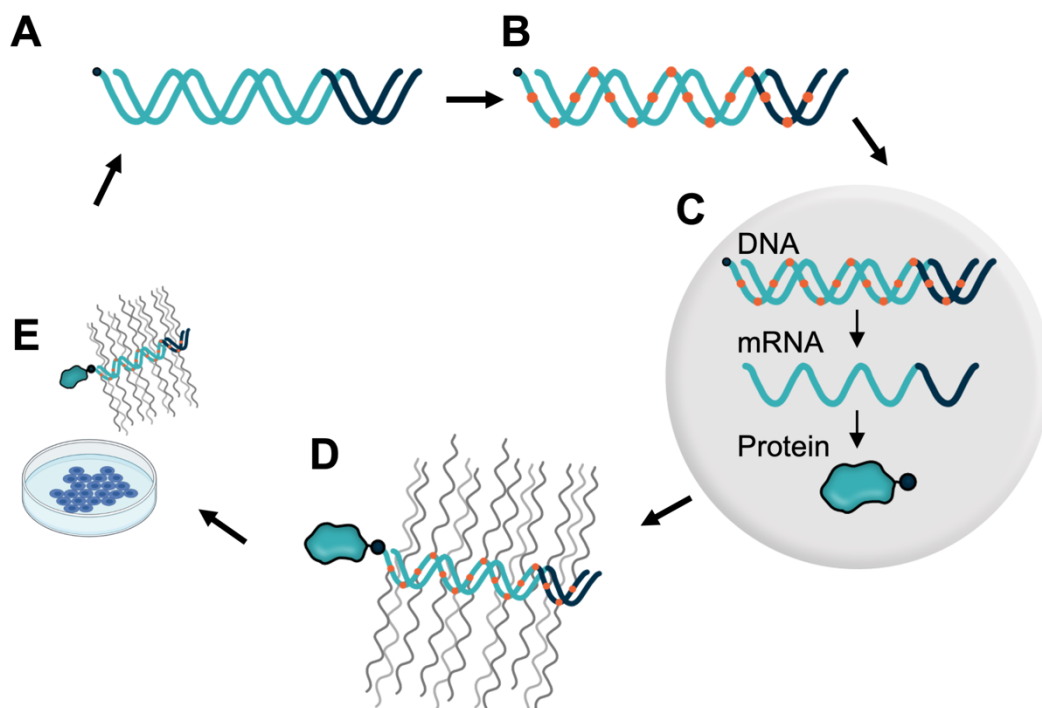


Figure 2-1. Workflow for a DNA display panning selection.

(A) DNA template (library fused to protein tag, for coupling protein back to DNA) is modified with (B) functional groups, then (C) *in vitro* transcribed and translated (IVTT) in droplets to retain monoclonality. Translated protein couples back to DNA via the protein tag. Next, templates are (D) PEGylated using the functional groups introduced before (E) subjected to biopanning. In this proof-of-concept study, we performed and optimized (A), (B), (D), and (E), with control experiments solely on the translation component of (C).

### 2.3. Results

Bioconjugation experiments and systematic optimizations were carried out using two different templates: (1) gene-encoding templates for initial testing of different nucleotide incorporation and corresponding protein expression efficiency of modified

templates, and (2) designed mock (not gene-encoding) templates for controlled testing of varying PEG modifications and subsequent testing in biopanning experiments with adherent cells.

### *2.3.1. Strategies for incorporating modifiable positions throughout a DNA template*

For introducing template-spanning modifiable positions, we used commercially available functionalized deoxyuridine triphosphates (dUTPs) that could be easily incorporated by a DNA polymerase during a primer extension step of a polymerase chain reaction (PCR). Following extension, the dUTPs could then be modified with functionalized PEG using bioconjugation techniques. To this end, we performed low-cycle PCR with the DNA templates using an equimolar mixture of dATP, dCTP, dGTP, and azide (N<sub>3</sub>)-functionalized dUTP. Beyond following the manufacturer's protocol for Vent (exo-) DNA polymerase-based PCR, the primer concentration, extension temperature, and cycling parameters were optimized for maximum dUTP-incorporated product which we assessed by gel electrophoresis. We reasoned that integration of bulky, functionalized dUTPs would increase the molecular weight of the DNA template considerably such that its relative mobility would be slower than that of unmodified DNA and would be visible on a nucleic acid gel.

In initial experiments, the azide-dUTP was mixed at ratios of 0%, 25%, 50%, 75%, and 100% with dTTP (unmodified) to assess the efficiency of incorporation. While modest band shifts were visible at 25% azide-dUTP, greater band shifts (i.e., higher molecular weight) were visible at 50% and 75% azide-dUTP, which may indicate even higher levels of azide-dUTP incorporation (Figure 2-2). Interestingly, at 100% azide-dUTP, there were two distinguishable products, one of which migrated at the same molecular weight as the 0% azide-dUTP product, and the rest at the highest molecular weight compared to the



other ratios tested. This result suggests some amplification inefficiency of the reaction when 100% azide-dUTP is used<sup>89</sup>, and that some templates did not participate in primer extension (Figure 2-2). Surprisingly, the absence of smeared DNA bands suggests uniform products – modified and unmodified – for the 100% azide-dUTP condition. For all following experiments, equimolar amounts of dATP, dCTP, dGTP, and azide-dUTPs are individually added to the reaction mixture such that any template that has successfully undergone primer extension will have all its thymine (i.e., functionalizable) positions modified with an azide. Although some slight smeared DNA bands were observed for some of these latter experiments with 100% azide-dUTP, most of the product appeared as distinct bands on the gel.

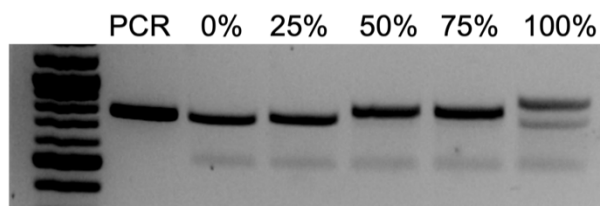


Figure 2-2. Gel electrophoresis analysis of azide-dUTP primer extension reaction parameters. Testing of different azide-dUTP and dTTP ratios (indicated from 0 - 100% azide-dUTP) show varying band shift trends, with higher azide-dUTP incorporation leading to higher molecular weight and slower migration on the gel. PCR refers to starting PCR amplified material (sfGFP-loTT, described below). The faint band visible in all primer extended sample is likely displaced ssDNA.

In addition to primer extension of the template with the functionalized dUTPs, the DNA template must remain “readable” by RNA polymerase for transcription as part of the preparation of a display particle downstream. For this study, we used gene-encoding DNA templates to assess the synthesis of functional protein from *in vitro* transcription and translation (IVTT) of modified DNA templates. We edited the nucleic acid sequence of a DNA template encoding super-folder green fluorescent protein (sfGFP) with two varying degrees of spacings between two thymine (TT) bases throughout the template using degenerate codons. We sought to test these two distributions to not only examine the

overall effect of bulky functional groups on the dUTPs, but also understand how the frequency at which these side group modifications occur throughout the DNA structure affects transcription and subsequent translation efficiency.

Intuitively, we hypothesized that too many modified bases in close proximity could lower transcription efficiency and thus lower the expression of the protein. (Hereafter the nomenclature for templates will be designated as “n<sub>x</sub>T” in which “x” indicates the number of bases following a thymine position. For example, ‘n<sub>5</sub>T’ corresponds to ‘VVVVT,’ in which V indicates an A, C, or G nucleotide following the IUPAC single letter code.) The first template, with higher consecutive thymine-thymine bases (‘TT’) throughout the template (sfGFP-hiTT), was designed to have fewer ‘n<sub>5</sub>T’ spacings, while the second template was designed with lower TT content (sfGFP-loTT) had more ‘n<sub>5</sub>T’ and ‘n<sub>6</sub>T’ spacings as allowable via codon degeneracy (Figure 2-3A, Table 2-2). Both the sfGFP-hiTT and sfGFP-loTT templates were compared to unmodified wild type (WT) sfGFP gene.

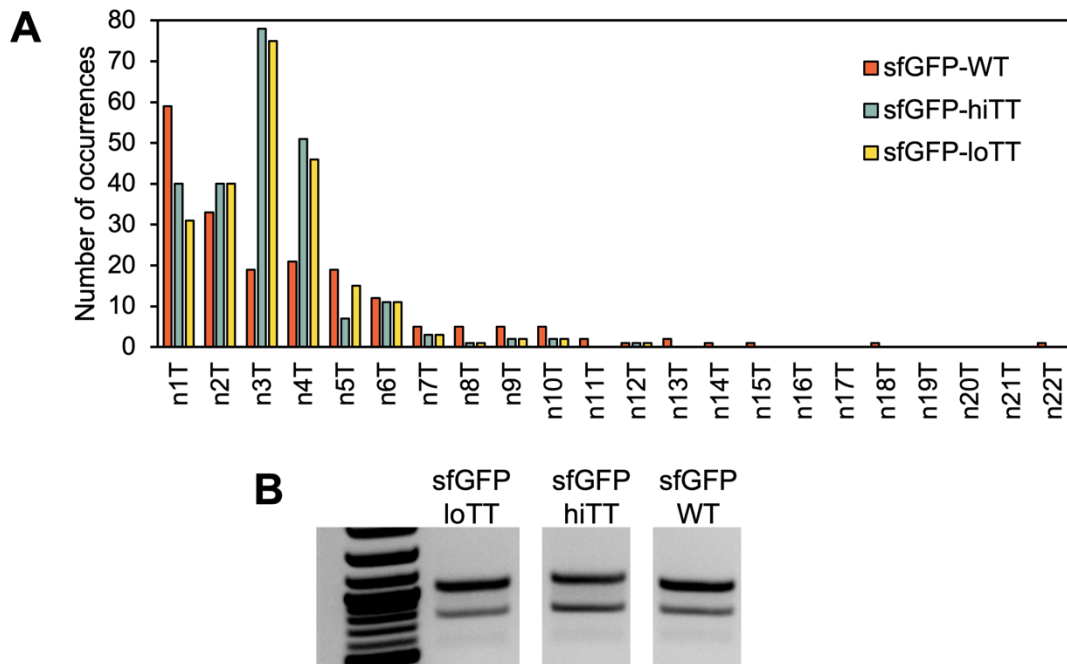


Figure 2-3. sfGFP DNA templates used for IVTT expression tests.

(A) Distribution of spacing between two thymine bases (TT) of sfGFP-hiTT and sfGFP-loTT

gene-encoding templates. (B) Different levels of azide-dUTP incorporated product (as inferred by band intensity) depending on thymine content.

The gene-encoding templates were amplified with azide-dUTPs as described in the experiments above. Based on the band intensity, the incorporation efficiency appeared to correlate with the number of TT positions in the DNA template, suggesting that the DNA polymerase was more accommodating of fewer incorporations (Figure 2-3B). After primer extension, the modified templates were gel extracted, purified, and put into an IVTT reaction. While the expression for all azide-dUTP templates was approximately 10-fold lower than that of an unmodified DNA template, the IVTT reactions with the modified template input still resulted in protein expression at levels above background. In line with our hypothesis, the efficiency of expression also correlated with the number of TT-modified positions in that the sfGFP-loTT template had higher expression than the sfGFP-hiTT.

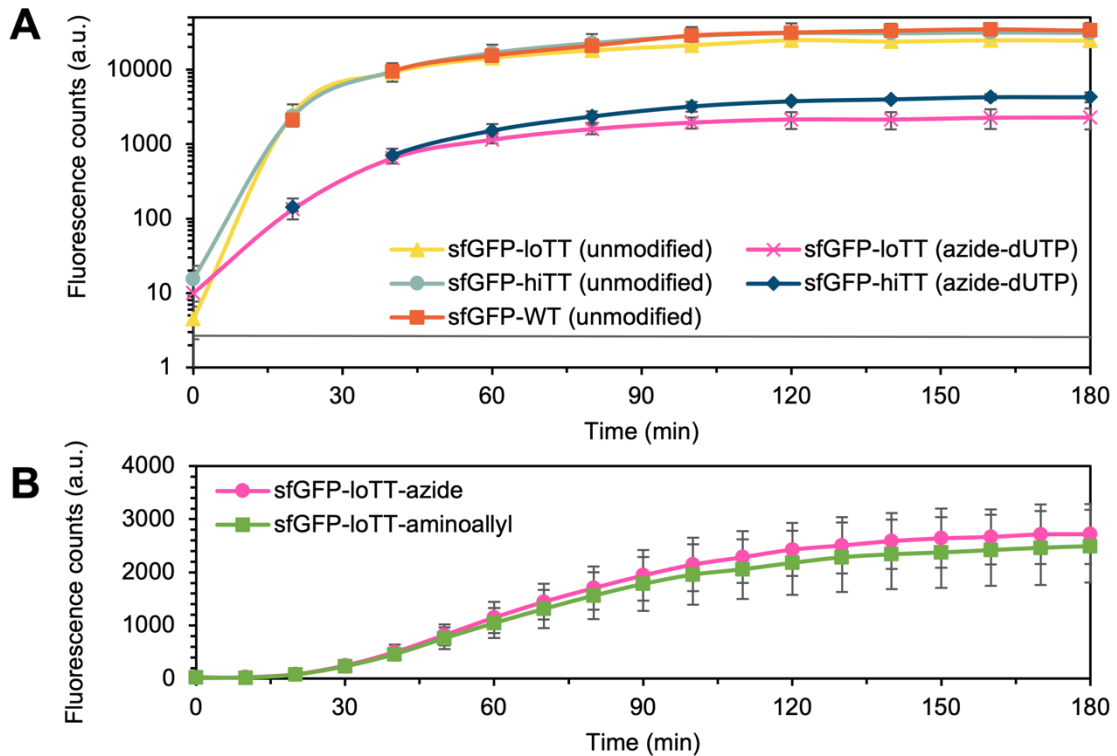


Figure 2-4. Comparison of IVTT efficiency of modified and unmodified templates. (A) Equal amounts of unmodified (PCR amplified) or modified (primer-extended with azide-dUTP) sfGFP-encoding DNA templates were subject to IVTT, and the protein expression was

analyzed by fluorescence detection. Unsurprisingly, unmodified templates resulted in the highest level of expression, while modified templates had ~10-fold lower expression. Although we anticipated the hiTT-dUTP sample to have lower levels of expression due to higher number of modifications, we hypothesized that the differences in expression between thymine content may be due to reduced codon optimization from using degenerate codons to design thymine spacings. (B) Comparison of sfGFP expression between aminoallyl-dUTP and azide-dUTP modified templates show similar efficiencies in functional expression.

Finally, in addition to the azide-dUTP, we also assessed another dUTP analog, aminoallyl-dUTP, for IVTT efficiency with the sfGFP-loTT template. We rationalized that the azide-dUTP may be more difficult to incorporate and subsequently express due to its larger size; however, IVTT data suggested that the sfGFP expression was relatively comparable for between templates modified with both dUTP analogs (Figure 2-4B). Thus, we proceeded with both templates for testing click chemistry-mediated bioconjugation of PEG molecules.

### *2.3.2. Click-mediated bioconjugation techniques for functionalization of dUTPs*

For our initial click-mediated bioconjugation approach, we opted for strain-promoted alkyne azide cycloaddition (SPAAC) chemistry as it is spontaneous, proceeds under physiological conditions, and more importantly, does not require metal catalysts. The SPAAC approach appeared preferable due to potential toxicity and side reactions promoted by metal catalysts in the presence of redox-sensitive biomolecules such as DNA and protein.<sup>90,91</sup> To accomplish this, we used the aminoallyl-dUTP primer extended templates to perform a two-step bioconjugation method: (1) amine-reactive crosslinking with aminoallyl-dUTP and a bifunctional N-hydroxysuccinimide-cycloalkyne molecule, then (2) strain-promoted [3 + 2] cycloaddition between an alkyne and an azide-functionalized PEG polymer. To introduce the alkyne groups needed for cycloaddition of our azide-PEGs, we performed coupling reactions with the bifunctional bicyclononyne-N-hydroxysuccinimide (BCN-NHS) and dibenzocyclooctyne-N-hydroxysuccinimide (DBCO-

NHS) esters (Figure 2-5A). We tested two different cycloalkynes, BCN and DBCO, as they differ in solubility and reactivity.<sup>92-94</sup> With the sfGFP-hiTT and sfGFP-loTT templates we tested a range of conditions including incubation duration, number of coupling reactions, and incubation temperature. With increasing number of coupling reactions, the relative base pair size (i.e., molecular weight) increased, although the difference decreased after 2 - 3 reactions relative to the first 1 - 2 reactions. As active esters can be easily hydrolyzed in aqueous environments, we believe that not all the input ester – while still in molar excess relative to the amine groups – reacts before being hydrolyzed, and that adding more ester in consecutive coupling reactions is the only method for increasing the efficiency of the reaction.<sup>90</sup>

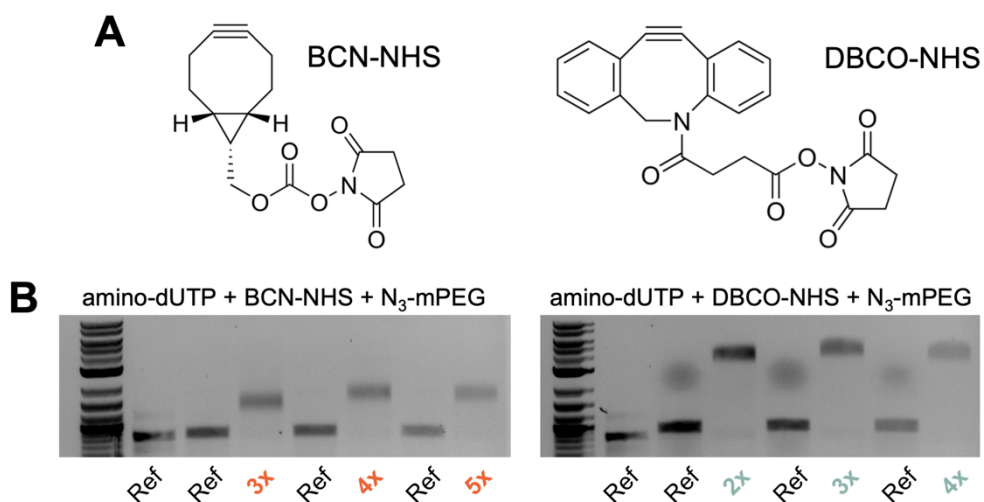


Figure 2-5. Cycloalkyne and azide SPAAC experiments.

(A) Comparison of BCN-NHS and DBCO-NHS structures used for amine-NHS coupling reactions. (B) Analysis of aminoallyl-dUTP + cycloalkyne-NHS + azide-PEG bioconjugation efficiency by gel electrophoresis. The same relative mass from each reaction was loaded in the gel for both bioconjugation approaches, and reference lane (Ref) shows molecular weight of starting material. Contrary to current literature, the reactivity for amine-NHS crosslinking using (A) BCN-NHS appears lower than (B) DBCO-NHS, which may be due to accessibility of the cyclooctyne in proximity with the DNA helix.<sup>92</sup> For both bifunctional alkyne-NHS molecules, the unmodified starting material is still present after multiple amine-NHS coupling reactions and the band intensity of modified DNA template decreases, suggesting loss of product and decreasing yield with repeated couplings.

Based on the band intensity, we also hypothesized that the yield was also decreasing with increased coupling reactions; thus, we switched our bioconjugation strategy to using the azide-dUTP to perform copper-catalyzed azide–alkyne cycloaddition (CuAAC) for PEGylation using alkyne-PEG. To adapt the click reaction to aqueous and biological conditions, the water-soluble  $\text{CuSO}_4$  and sodium ascorbate were chosen to be the copper source and reducing agent, respectively. Further, to minimize aforementioned concerns regarding metal catalyst toxicity to biomolecules, we included a nontoxic stabilizing ligand, Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), which intercepts peroxides and/or radicals derived from Cu/ascorbate/ $\text{O}_2$  that can oxidize histidine and other amino acid residues and also protects biomolecules from hydrolysis via Cu(II) byproducts.<sup>95</sup> Lastly, the reaction solution is buffered in a biologically compatible sodium phosphate buffer (pH 7.8), and the reaction is carried out for one hour at room temperature before being quenched by the addition of excess ethylenediaminetetraacetic acid (EDTA) to sequester the copper.

To develop the click-based PEGylation approach, we designed synthetic DNA templates with fixed thymine positions at every 3, 5, and 7 nucleotides (termed n3T, n5T, and n7T, respectively) and chose alkyne-PEG molecular weights of 350, 550, 750, 1K, 2K, and 5K for all three templates for systematic testing. While the three different template sequences mostly adhere closely to the designated fixed thymine positions, they all contain a common, short (90 bp), internal ‘n5T’ spacing region for subsequent analytical quantification using probe-based qPCR to negate any amplification differences (Table 2-2). In addition, the ‘n3T’ template contains seven ‘n4T’ spacing segments (highlighted in Table 2-2) that were necessary for gene synthesis purposes.

In contrast to PEGylation experiments using SPAAC, the CuAAC approach led to disappearance of the starting material after a single incubation step for 1 hour, which we

inferred to be complete reactivity. As expected, the slower migration pattern correlated with increasing PEG molecular weights. With the higher reactivity and more streamlined protocol, we proceed with the CuAAC approach for PEGylation of the n3T, n5T, and n7T DNA templates with the varying PEG molecular weights (Figure 2-6A).

Following PEGylation, the templates were purified using a modified gel extraction protocol. Traditional column purification methods in which DNA is adsorbed to silica membranes in the presence of chaotropic salts would be unusable in downstream applications with a DNA display particle (in which the phenotype is covalently coupled to the template); thus, we developed a milder gel extraction protocol in which the template is released from the agarose gel using agarase digestion.<sup>96</sup> For the digestion step, the gel percentage was kept low to allow for higher efficiency of agarase digestion while still allowing sufficient spatial separation between modified and unmodified templates. Consequently, due to the low agarose content, the gel was very delicate and unstable; thus, the gel was run at 4 °C to avoid gel melting from excessive heat during electrophoresis. After extraction and agarase digestion at 37 °C, the reaction mixture was further purified using a cellulose acetate filter-based column to physically separate the DNA in solution from the remaining gel solids in the mixture. In the last step of the modified purification protocol, the DNA was rebuffed into its final buffer (Buffer EB: Tris-HCl 10 mM, pH 8) using diafiltration columns. As the PEGylated DNA could no longer be quantified using standard absorption spectra measurements, the quantification was performed using gel quantification (Figure 2-6B).

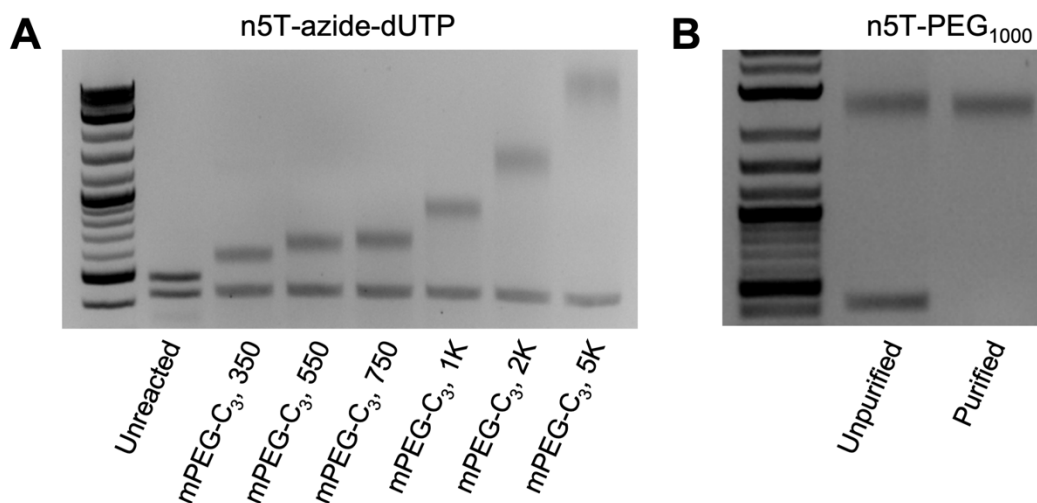


Figure 2-6. PEGylation reaction products of a 427-bp DNA template. Template contains 82 functionalizable positions spaced every 5 positions (n5T) throughout the template. (A) First lane is the non-clicked (unreacted) starting material, which is a mixture of unmodified template (bottom band) and the primer extended template (top band) which has increased in weight following azide-dUTP incorporation. The listed molecular weights indicate the PEG polymer used in the click reaction. (B) Reaction products from a click reaction (left lane), run alongside its purified PEGylated band (right lane).

### 2.3.3. Biopanning strategies with PEGylated DNA templates

Compared to most display technologies in which the selection particles are panned against purified, immobilized protein, the selection surface for biopanning is a cell membrane, a highly heterogeneous and complex landscape. In this study, we panned against adherent cells to enable high-throughput downstream applications by utilizing a 96-well plate and a corresponding automated washing procedure. Further, previous biopanning studies have shown successful selections against adherent cells<sup>97-99</sup>, even with cell culture treatment<sup>100</sup>, without increasing background binding. Although the particle modifications address some of the challenges of panning against a sticky cell surface, additional reaction conditions were introduced to present the DNA template in optimal selection conditions. For testing and optimizing biopanning strategies, we used an epithelial breast cancer cell line, SK-BR-3.



First, we tested various cell seeding densities to ensure a uniform monolayer for panning. We used a higher seeding density with a longer incubation period to achieve a confluent monolayer of cells on the well surface to reduce the number of particles that may interact with the empty space on the well surface (i.e., potentially sticky tissue culture treated plastic) as opposed to the cell surface. As these conditions led to growth of a heterogeneous secondary layer on top of the adherent monolayer of cells, loosely associated cells and cells that had grown on top of the monolayer were removed with light washing using ice-cold Dulbecco's phosphate buffered saline (DPBS) following plating.

Second, while it is desirable to keep the cells in a physiologically relevant temperature range (i.e., 25 - 37 °C), the particles were panned at 4 °C to prevent receptor cycling and endocytosis of the templates. All panning and washing steps were performed with cold DPBS to prevent spontaneous cell lysis caused by osmotic lysis. Next, to further minimize nonspecific adsorption of the selection particles to any open well surface and mitigate potential protease activity, 0.1% bovine serum albumin (BSA) was added to the biopanning buffer. Finally, for post-panning washes to reduce the number of free particles that are in solution, weakly bound, or adsorbed nonspecifically on the cell surface, we used an automated plate washer with mild shaking in between washes. The plate washing protocol was also modified to ensure gentle resuspension and aspiration to not disturb the cell monolayer. For cell recovery, we utilized a low-percentage trypsin solution (0.05%) with gentle agitation to collect the cells for qPCR and cell counting.

#### *2.3.4. Analysis of non-specifically adsorbed DNA templates by qPCR*

Following panning, the DNA templates that adhered nonspecifically to cell surfaces were quantified using qPCR. To directly measure the number of particles per cell, we collected cell suspensions and counted the cell density as well as put a relative volume of

the suspension directly in qPCR mix. qPCR analysis of templates was initially determined using SYBR Green, a dye that intercalates into dsDNA; however, addition of cell suspensions into the qPCR mix noticeably changed the overall signal and cycle threshold (Cq) values, possibly from the added salts and trypsin in the collection buffer. Thus, we shifted to a probe qPCR approach, which was much more sensitive in a complex environment in the presence of cell lysate and not affected by the presence of DPBS. As a cautionary measure, we utilized a Cy5-probe to prevent possible signal contamination from cellular autofluorescence.

To examine the effect of template stealthing on nonspecific adsorption, we used the PEGylated templates (no protein attachment) for biopanning rounds. As mentioned in Section 2.3.2, we used the n3T, n5T, and n7T templates (427 bp) conjugated with varying PEG sizes (350, 750, and 1000 Da) for biopanning experiments. Briefly, we panned approximately  $1 \times 10^{10}$  particles against the SK-BR-3 and compared the copies bound per cell. Since no proteins were attached, any association with the cells would be via nonspecific adsorption. After normalizing the copy number from suspensions samples to a control (i.e., unmodified DNA), the ratio of the number of PEGylated copies to unmodified was calculated. Generally, all templates with the high-molecular-weight PEG1000 exhibited the lowest amount of nonspecific adsorption relative to their unmodified (i.e., non-PEGylated) versions, while templates with the smaller PEG350 adsorbed the most except n7T with PEG350 (Figure 2-7A). The n5T template modified with PEG1000 had the least amount of nonspecific adsorption relative to non-PEGylated (Figure 2-7A), although n7T templates were overall the least adherent, which may be due to having the least number of modifications as the modifications, as the addition of PEG molecules may

disrupt the helical structure of the DNA, causing charged bases to be exposed (Figure 2-7B).

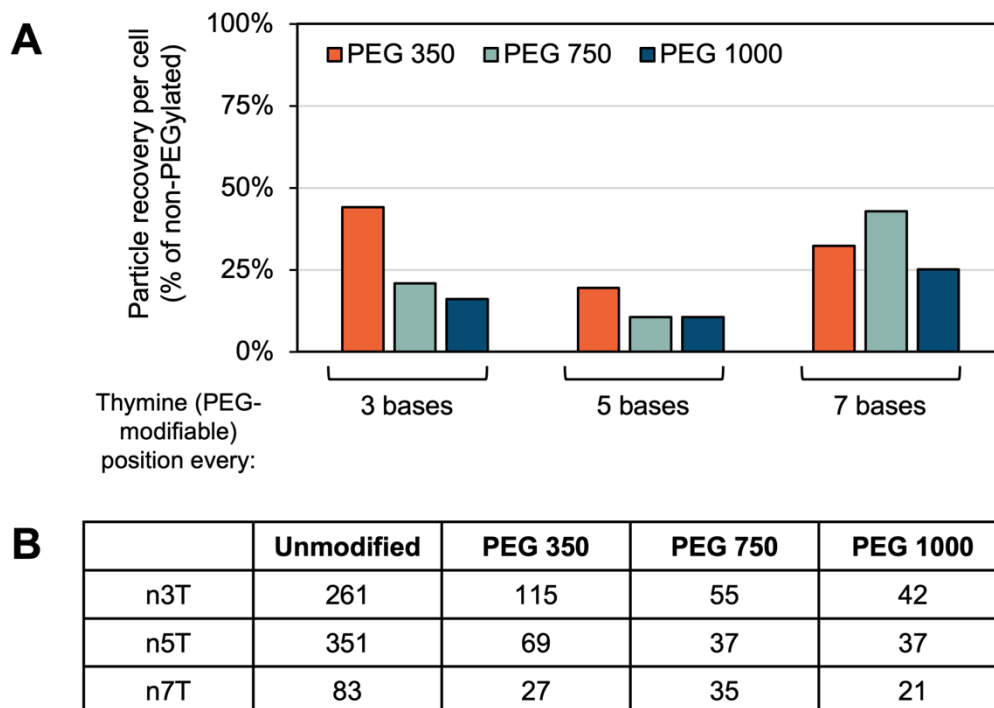


Figure 2-7. Data summary for biopanning qPCR.

(A) Comparison of particle recovery per cell for DNA templates with varying degrees of modification. Templates with thymine (PEG-modifiable) positions every 3, 5, and 7 bases, PEG molecular weights of 350, 550, 750, and 1000 were clicked onto the template using CuAAC. Templates with too many (n3T) modifications adsorbed to cell surfaces more than templates with a moderate level of modification (n5T and n7T). (B) Table shows corresponding total particle recovery per cell data from (A). The lowest adsorption relative to the unmodified (non-PEGylated) template was observed for high-molecular-weight PEG (1000) for the n5T template, while the lowest adsorption overall was the n7T template modified with PEG1000.

## 2.4. Discussion

Current approaches to protein-based selections on cell surfaces often require negative selections in which selection particles are panned against non-target-presenting cells to enrich the binder pool prior to selection panning.<sup>43,52,101–103</sup> Other biopanning protocols include using technically challenging methods like microfluidic devices to perform extensive washing steps to reduce the nonspecific background that are not easily

accessible.<sup>104,105</sup> While these can be helpful in removing binders that bind nonspecifically to cell surfaces, this can also preemptively limit the library diversity if the selection particle, rather than the binder, is driving the nonspecific binding. For many selection methods including DNA display, the non-binder portion of the selection particle is often much bigger than the protein binder itself and can therefore significantly interact with the selection surface. Interestingly, most literature regarding biopanning on live cells typically use *in vivo* (phage or yeast) display to the best of our knowledge. Taken together, we aimed to optimize components of a DNA display selection particle to improve selections in a complex environment.

Azide-dUTP incorporation using primer extension yielded two distinct products (modified and unmodified), most likely caused by amplification inefficiency from complete substitution of dTTP with dUTP. Indeed, other studies with modified dNTPs showed that product yield decreases substantially with increasing ratios of natural to modified dNTP, with 100% substitution resulting in complete or near-complete loss of product, although the specific modification of the dNTP (e.g., length of functional linker arm) appears to vary the efficiency.<sup>106–109</sup> Another consideration for decrease in amplification efficiency is the frequency of the modification; if the density of the modification is too high, steric hinderance and other molecular interactions between the modified dNTP can stall or dissociate DNA polymerases.<sup>110</sup> Accordingly, in our study, the yield for the desired dUTP-incorporated product increased with decreasing amounts of thymine positions; it may also be challenging for the DNA polymerase to incorporate bulky dUTP analogs.<sup>108</sup> Overall, while the dUTP analog incorporation was successful, the primer extension reaction may still be optimized for higher efficiency, as approximately half the starting material is unable to be extended with the azide-dUTP. In addition, as indicated by the IVTT data with the sfGFP-hiTT and sfGFP-loTT templates, there may also be an optimal thymine distribution

in the context of functional protein expression. Although higher spacing between two thymine positions (sfGFP-loTT) led to higher expression, the unmodified sfGFP-WT and sfGFP-hiTT had higher expression compared to unmodified sfGFP-loTT, which suggests that one must consider both thymine distribution (or other nucleotide base, if using other functionalized analogs) and codon optimization when designing and/or editing DNA sequences using codon degeneracy. Lastly, despite the lower expression for modified sfGFP-hiTT and sfGFP-loTT templates, the protein expression is still contextually relevant, as our modification approach still allows transcription and translation of the DNA template.

Both the strain-promoted and copper-catalyzed azide-alkyne cycloaddition approaches resulted in PEGylated DNA templates with varying migration behaviors on the nucleic acid gel; surprisingly, the BCN-NHS coupled DNA templates appeared to have lower efficiency of subsequent PEGylation. Compared to DBCO-NHS, BCN-NHS is expected to have nearly 10-fold higher reactivity ( $k_{\text{BCN}} = 2.0 - 2.9 \text{ M}^{-1} \text{ s}^{-1}$  vs.  $k_{\text{DBCO}} = 0.2 - 0.5 \text{ M}^{-1} \text{ s}^{-1}$ ), although the reaction kinetics vary on the azide substrate.<sup>92</sup> In this study, the proximity of the DBCO-NHS / BCN-NHS to the DNA template backbone, as well as the reaction conditions used to accommodate biomolecules may have contributed to the different reactivity of the bifunctional cyclooctyne-NHS molecules. For example, the bulkier DBCO may be farther away from the backbone and is therefore easier to access for the azide-PEG compared to the BCN, leading to higher reactivity. Although we performed a singular experiment with DBCO-dUTP in attempt to simplify the workflow and negate the amine-NHS coupling step altogether, as suspected, the much larger DBCO-dUTP analog was not able to be incorporated during primer extension without some addition of dTTP; thus, we proceeded with our azide-dUTP / alkyne-PEG bioconjugation approach.

Based on the CuAAC experiments, covalent linkage of the higher molecular weight PEG appeared to also increase the size of the DNA template. Templates clicked to large

PEGs (2000 and 5000 Da) stained more faintly than those clicked to smaller PEGs; we rationalized that the effective shielding of the DNA from high molecular weight PEG<sup>111</sup> led to reduced intercalation of the ethidium bromide (EtBr) dye used for imaging the gel. EtBr intercalation is also affected by the distance between bases (i.e., binding preferred in minor > major groove).<sup>112</sup> Here, the addition of bulky polymers at high density can cause steric hinderance<sup>113</sup> which may subsequently change the nucleotide spacing of the DNA, leading to decreased intercalation. Thus, we used templates modified with lower molecular weight PEGs for testing our biopanning protocol. The modified purification protocol for gel extraction of the PEGylated templates allowed for sufficient purity and retention of the modified templates. As later described in Chapter 4, we have also successfully applied this protocol for protein-conjugated DNA templates.

For preparation of the adherent cell panning surface, achieving a cell monolayer is an important step. Removal of non-monolayer cells is a critical point; since these cells may detach during panning as they are loosely attached and do not cover the well surface evenly. For cell recovery post-panning, we initially examined gentle removal options such as enzyme-free solutions, low-trypsin concentrations, citric saline, and cell scraping. However, in our efforts to streamline the panning protocol for easy implementation, we found that mild trypsinization with gentle shaking preserves cell viability and allows for quick cell collection.

In our systematic study, the effect of PEGylation on minimizing nonspecific adsorption varied with the degree of modification used. PEG, often conjugated to nanoparticles and proteins for stealthing, can switch between “mushroom” and “brush” models based on its density.<sup>114</sup> At low density, it can reside on the surface as mushroom-like structures that bunch close to the surface and provide surface coverage; contrastingly, at high density, PEG chains extend into brush-like shapes to avoid overlap with

neighboring PEG molecules. The controlling factor for PEG conformations is the distance between each PEG molecule (D) relative to their Flory radius (R<sub>F</sub>), which describes the approximate space that a polymer occupies, and is given by the equation,

$$R_F = aN^{\frac{3}{5}}$$

where “a” is the monomer size (i.e., persistence length, which is 0.35 nm for PEG) and N is the number of polymer units in the polymer.<sup>115</sup> PEG molecules can exist in multiple regimes: (1) at low densities (R<sub>F</sub> / D ≤ 1), PEG adopts a mushroom conformation; (2) at higher densities (R<sub>F</sub> / D > 1), PEG increasingly adopt a brush regime, until (3) it reaches a dense brush conformation when L / R<sub>F</sub> > 2.<sup>111</sup>

**A**

	<b>D (nm)</b>
n3T	1.02
n5T	1.70
n7T	2.38

**B**

	<b>R<sub>F</sub> (nm)</b>
PEG 350	1.21
PEG 750	1.92
PEG 1000	2.28

Table 2-1. Calculations for PEG molecules used in this study.

(A) Calculation of distance (D) between PEG positions based on 0.34 nm distance between DNA nucleotides.<sup>116</sup> (B) Calculation of Flory radii (R<sub>F</sub>) for each PEG molecular weight.

Based on the nucleotide distance (~0.34 nm between nucleotides<sup>116</sup>) (Table 2-1A) and molecular weight of the PEG used in this study (Table 2-1B), intermolecular distance (D) between PEG were calculated to be similar to R<sub>F</sub>. As such, the PEG polymers likely formed a combination of mushroom- and brush-like structures for most templates, with n3T modifications forming mostly brush-like PEG layers, and n7T modifications forming mushroom-like PEG structures. Many of these studies are performed on three-dimensional nanoparticle surfaces, however, and may be different for our two-dimensional DNA backbone.<sup>115,117</sup> In our study, we observed largest decreases in adsorption from templates with the high degrees of modification (n3T-PEG<sub>1000</sub> and n5T-PEG<sub>1000</sub>), which

may suggest that increased density of PEGylation is needed for adequate coverage of the DNA structure. The n5T template may provide a more ideal number of modifiable positions compared to n3T as dsDNA has a helical periodicity of 10-10.5 bases<sup>118</sup>, and the PEG molecules may point away from one another without much steric hinderance. Although n7T had the lowest levels of nonspecific adsorption overall, comparing the unmodified vs. PEGylated template shows minimal difference in adsorption, which suggests that the PEGylation density may be too low to achieve meaningful stealthing effects. The low adsorption of n7T, however, may be a result of other template-dependent factors, such as high GC-content (59%, compared to 52% for n5T and 46% for n3T) which may increase its stability<sup>119</sup> and less secondary structure as a result of fewer restrictions on thymine distribution.

While we were able to achieve up to 10-fold lower nonspecific adsorption of PEGylated templates compared to unmodified templates, it is important to note that the overall “stickiness” is quite low for DNA. For instance, while a mere 37 copies adhered per cell for n5T-PEG1000 modified templates, the unmodified n5T template still only had 351 copies adhered per cell, which is a relatively low background. If the cell surface target is overexpressed (e.g., HER2 receptors on SK-BR-3 cells, which is expressed at  $\sim 1.3 \times 10^6$  receptors per cell<sup>120</sup>), then the DNA modification may not be necessary, as the signal will vastly outweigh the noise from nonspecific binding. However, if the target cell surface protein is only expressed on the order of  $10^2$  copies per cell such as some GPCR classes<sup>121,122</sup>, binders may get lost within the noise of the selection, and DNA PEGylation may be advantageous for increasing the signal-to-noise.

Finally, full implementation of this DNA PEGylation approach for cell surface library selections using DNA display requires library generation, *in vitro* compartmentalization, and IVTT. Although establishing a reliable microfluidic platform for generating droplets



was beyond the scope of this study, the bioconjugation techniques, biopanning protocol, and analysis with probe qPCR can be applied to other *in vitro* display approaches such as mRNA/cDNA display (which behaved similarly to dsDNA in our experience) with minor modifications.

## **2.5. Conclusion**

Here, we developed and optimized a PEGylation technique for attaching stealthing PEG polymers to a DNA template. We observed varying degrees of reduced nonspecific adsorption on cell surfaces through our systematic testing of DNA template modification using fixed spacings and different molecular weights of PEG. Overall, we noted that the DNA itself was not very adsorptive on cell surfaces, although some modification may be helpful in selections against cell receptors such as integral membrane proteins that are lowly expressed and may fall within the noise of the selection. With continued biological and pharmaceutical interest, expanding the toolbox for direct cell surface panning experiments for ligand discovery against integral membrane proteins is an important challenge for *in vitro* display technologies. While the direct application would be DNA display particles in complex environments such as biopanning on cell surfaces, this technique can be applied to stealthing other DNA- or RNA-based therapeutics such as aptamers for prolonged circulation and reduced nonspecific adsorption.

## **2.6. Materials and Methods**

### *2.6.1. T-spacing test*

Genes encoding sfGFP-hiTT and sfGFP-loTT were ordered as gBlocks (IDT) (Table 2-2), digested using restriction enzymes XbaI and HindIII (New England Biolabs), and ligated into the pPUR vector (Clontech). The ligation product was transformed into

XL1-Blue *Escherichia coli* (*E. coli*) (Agilent) and individual clones were picked. The plasmids were purified using the QIAprep spin miniprep kit (Qiagen), and sequence-verified using Sanger sequencing.

### 2.6.2. dUTP primer extension

gBlocks (IDT) were first PCR amplified using Phusion Hot Start Flex following standard manufacturer's protocol. The products were purified using QIAquick PCR purification kit (Qiagen) and quantified using A260 measurements (BioTek). The purified PCR product was then put into a primer extension reaction with Vent (exo-) DNA polymerase (NEB), ThermoPol buffer (NEB), 5'-forward primer (IDT), and the deoxynucleotides dATP, dCTP, dGTP (NEB), and modified dUTP. The modified dUTPs used for the experiments were aminoallyl dUTP (Thermo Fisher Scientific), Azide-PEG4-aminoallyl-dUTP (Jena Biosciences), or 5-DBCO-dUTP (Jena Biosciences) with the following pipetting scheme (1X reaction):

Reagent (concentration)	Volume or mass
ThermoPol Reaction Buffer (10X)	2.5 $\mu$ L
dATP (10 mM)	0.5 $\mu$ L
dCTP (10 mM)	0.5 $\mu$ L
dGTP (10 mM)	0.5 $\mu$ L
Modified dUTP (10 mM)	0.5 $\mu$ L
Forward primer (1 $\mu$ M)	3.75 $\mu$ L
Template	200 ng
Vent (exo-) DNA polymerase (20 U/ $\mu$ L)	0.5 $\mu$ L
Nuclease free water	Up to 25 $\mu$ L
Total	25 $\mu$ L

The reaction was mixed and cycled using the following protocol:

Temperature	Duration	Cycles
95 °C	9.5 min	1X
95 °C	30 sec	4X
$T_a$ °C*	30 sec	
72 °C	45 sec	
72 °C	5 min	1X
4 °C	$\infty$	

\* $T_a$  of 53 - 55 °C were used, as determined by the NEB  $T_m$  calculator based on the primer

### 2.6.3. Cell-free expression tests for primer-extended DNA templates

PCR-amplified (unmodified) or primer-extended (dUTP-modified) DNA templates were put into 5  $\mu$ L PURExpress (NEB) reactions following the manufacturer's protocol. Reactions were incubated at 37 °C on a thermal cycler (Bio-Rad) and fluorescence measurements (excitation: 485 nm, emission: 515 nm) were taken every 20 min on a Cytation plate reader (Bio-Tek). To prevent reactions from cooling to room temperature between readings, the Cytation was pre-warmed to 37 °C, and an aluminum plate carrier used for transporting samples and taking measurements was warmed to 37 °C.

### 2.6.4. SPAAC bioconjugation experiments

(1*R*,8*S*,9*s*)-Bicyclo[6.1.0]non-4-yn-9-ylmethyl *N*-succinimidyl carbonate (BCN-NHS) (Millipore Sigma) and dibenzocyclooctyne-*N*-hydroxysuccinimidyl ester (DBCO-NHS) were dissolved in dimethyl sulfoxide (DMSO) (Millipore Sigma) at 50 mM and 10 mM working concentrations, respectively, aliquoted, and stored at -80 °C. The stocks were diluted to working stocks of 20 mM (BCN-NHS) or 5 mM (DBCO-NHS) with dry DMSO. For all post-primer extension purifications, the QIAquick PCR purification kit (Qiagen) was used with the following modifications: (1) the Buffer PE wash step was replaced with a phosphate wash buffer (70% v/v ethanol in 10 mM NaPO<sub>4</sub>, pH 7.9) and (2) Buffer EB elution step was replaced with either nuclease free water or 10 mM NaPO<sub>4</sub>, pH 7.9 to avoid Tris-based buffers.

Prior to performing the amine-NHS coupling reaction, the aminoallyl-dUTP primer extended DNA template (sfGFP loTT) was digested with NdeI to remove non-primer-extended (i.e., non-dUTP incorporated) product, then column purified. For the coupling reaction, 650 ng of the purified DNA was mixed with 500 mM NaPO<sub>4</sub>, pH 7.9 (10 mM final concentration), NHS-compound (either 1 mM BCN-NHS or 0.25 mM DBCO-NHS final concentration) and made up to 15  $\mu$ L with nuclease free water. The reaction was incubated

at room temperature for 3 hours. In between each coupling reaction, the mixture was column-purified on MinElute columns (Qiagen) and eluted in 15  $\mu$ L. An aliquot corresponding to 65 ng from each coupling reaction was saved for subsequent gel electrophoresis analysis. Following the final coupling step, the reaction was quenched with 50 mM glycine, then column purified after a 15 min room temperature incubation. After amine-NHS coupling, azido-dPEG<sub>23</sub>-amine (Quanta Biodesign) was added to the purified reaction (2 mM final) for PEGylation experiments, and the mixture was incubated overnight at room temperature. For analysis, the PEGylation reaction was mixed with 6X Loading dye (NEB) and run on a 0.8 - 1.2% agarose gel.

#### 2.6.5. CuAAC bioconjugation experiments and purification

Alkyne-mPEG of molecular weights (MWs) 350, 550, 750, 1000, 2000, and 5000 Da (Biochempeg) were resuspended in nuclease-free water at 1 - 10 mM stock concentrations and stored at -20 °C. The click chemistry reactions were performed with the following pipetting scheme:

Reagent	Final concentration
CuSO <sub>4</sub> / THPTA pre-mix*	0.4 mM (CuSO <sub>4</sub> ), 2 mM (THPTA)
Sodium ascorbate	4 mM
Alkyne-mPEG	0.2 mM
Sodium phosphate buffer	4 mM
Azide-modified DNA	60 - 200 ng
Nuclease free water	Up to 50 $\mu$ L
Total	50 $\mu$ L

\*Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) (Millipore Sigma) and copper(II)sulfate (Millipore Sigma) were pre-mixed at 10 mM and 2 mM concentrations, respectively, with nuclease-free water prior to mixing with the rest of the reaction.

The reaction was incubated for 1 hour at room temperature before being run on a 0.8% low-melting point agarose (Millipore Sigma) gel at 4 °C. To purify, the gel was mixed with 10% v/w  $\beta$ -Agarase I buffer (10X) (NEB) (e.g., 7.95  $\mu$ L buffer added to 79.5 mg agarose gel slice) and incubated for 10 min at 65 °C. The tube was transferred to a 42 °C

water bath for 10 - 15 min before 2 units of  $\beta$ -Agarase I (NEB) were added and incubated for 1 hour. The reaction was rebuffed with 100 kDa MWCO diafiltration columns (Amicon) using Buffer EB to remove digested agarose.

#### 2.6.6. Cell culture

SK-BR-3 cells were kindly provided by Professor Paolo Provenzano (Department of Biomedical Engineering, University of Minnesota). The cells were grown at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub> in McCoy's 5A medium (Gibco) containing 10% (v/v) fetal bovine serum (FBS) (Gibco). For all biopanning experiments, the cells were harvested at 80-90% confluency using 0.25% Trypsin-EDTA (Gibco) and resuspended in McCoy's media with 10% FBS.

#### 2.6.7. Biopanning

The SK-BR-3 cells were plated at  $4 \times 10^4$  cells/well in 96-well plates (Falcon) for 48 hours at 37 °C in growth media. Prior to panning, the cells were washed 3x with cold DPBS, and loosely attached cells were removed by shaking the well plate at 275 rpm for 20 min at 4 °C with 100  $\mu$ L DPBS. Panning buffer ( $1 \times 10^{10}$  DNA molecules in 50  $\mu$ L DPBS with 0.1% w/v BSA) was added to the well and panned for 2 hours at 4 °C on a plate shaker at 75 rpm. The plate was washed at 4 °C using the plate washer (BioTek) 13 times with DPBS. After pipetting out remaining DPBS, 40  $\mu$ L of resuspension buffer (0.05% trypsin-EDTA; Gibco) was added to the well, and the plate was incubated at 30 °C for 8 min, then at room temperature for ~10 min, or until the cells were suspended evenly. Then the cell suspension was collected in a 1.5 mL tube, made up to 50  $\mu$ L with resuspension buffer, and 10  $\mu$ L was taken for cell counting (Invitrogen).

#### 2.6.8. Probe qPCR

For all qPCR reactions, PrimeTime MasterMix (IDT) was used with 0.5  $\mu$ M forward and reverse primers and 0.2  $\mu$ M probe primer (IDT) in 10 - 20  $\mu$ L reaction volumes using the following cycling protocol:

<b>Temperature</b>	<b>Duration</b>	<b>Cycles</b>
95 °C	10 min	1X
95 °C	15 sec	40X
60 °C	45 sec	

For backcalculation of molecules in solution, the recovered C<sub>q</sub> was subtracted by 1, to account for only half the template (i.e., the non-PEGylated template strand) being amplifiable during the first cycle. On average, the C<sub>q</sub>s for primer extended and non-primer extended templates with same inputs differed by 1 - 3 cycles.

Table 2-2. Synthetic templates used for thymine modification experiments.

DNA name	DNA sequence (5' → 3')
sfGFP_hiTT (797 bp)	ATAGGGTCTAGAAATAAGGAGATATACATATGAGTAAAGGTGAAGAATTATTCAGTGGTGTCTG TCCCTATCCTCGTTGAATTAGATGGTGTATGTTAATGGTCATAAATTCAGTGTGCGTGGTGAGG GTGAAGGTGATGCTACTAATGGTAAATTAAGTCTTAAATTCATCTGTACTACTGGTAAATTCG CTGTCCCTTGGCCTACTCTCGTCACTACCTTAAGTATGGTGTTCATGCTTCAGTGTCTATC CTGATCATAATGAAACGTCATGACTTCTTCAAGTCTGCTATGCCTGAAGGTTACGTTTCAGGAAC GTACTATATCCTTCAAAGATGATGGTACTTATAAGACTCGTGTGAAGTAAAGTTCGAAGGTG ATACTCTCGTGAATCGTATTGAGCTTAAAGGTATTGACTTCAAAGAAGATGGTAATATCCTCG GTCATAAATTAGAGTATAACTTCAATAGTCATAATGTATACATCACTGCTGATAAAACAAAAGA ATGGTATTAAAGCTAACTTCAAATTCGTCATAATGTTGAAGATGGTCCGTTCAATTAGCTG ATCATTATCAACAAAATACCTTATCGGTGATGGTCCCTGCTCCTGCCTGATAATCATTACC TGAGTACTCAATCTAAATTTGCTAAAGATCCTAATGAAAAGCGTGATCATATGGTCCCTCTTG AGTTCGTAAGTGTGCTGGTATCACTCATGGTATGGATGAGCTCTATAAATAATAAGGATCCT AACCCTCTCTAAACGGAGGGGTTTCCCGAAGCTTGGCACT
sfGFP_loTT (797 bp)	ATAGGGTCTAGAAATAAGGAGATATACATATGAGTAAAGGTGAAGAATTATTCAGTGGTGTCTG TCCCTATCCTCGTTGAATTAGATGGTGTATGTTAATGGTCATAAATTCAGTGTGCGTGGTGAGG GTGAAGGTGATGCTACTAATGGTAAACTGACTCTGAAATTCATCTGTACTACTGGTAAACTCC CTGTCCCGTGGCCTACTCTCGTCACTACCTGACCTATGGTGTACAATGCTTCAGTGTCTATC CTGATCATAATGAAACGTCATGACTTCTTCAAGTCCGCTATGCCTGAAGGTTACGTTTCAGGAAC GTACTATATCCTTCAAAGATGATGGTACTTATAAGACTCGTGTGAAGTAAAGTTCGAAGGTG ATACTCTCGTGAATCGTATTGAGCTTAAAGGTATCGACTTCAAAGAAGATGGTAATATCCTCG GTCATAAATTAGAGTATAACTTCAATAGTCATAATGTATACATCACTGCTGATAAAACAAAAGA ATGGTATTAAAGCTAACTTCAAATCCGTCATAATGTTGAAGATGGTCCGTTCAATTAGCTG ATCATTATCAACAAAATACCTTATCGGTGATGGTCCCTGCTCCTGCCTGATAATCATTACC TGAGTACTCAATCTAAACTGAGTAAAGATCCTAATGAAAAGCGTGATCATATGGTCCCTCTTG AGTTCGTAAGTGTGCTGGTATCACTCATGGTATGGATGAGCTCTATAAATAATAAGGATCCT AACCCTCTCTAAACGGAGGGGTTTCCCGAAGCTTGGCACT
n3T (427 bp)	GCTCGTAGTGTGATACTCCTAGTAATCCTAATCGTGTCTCTGGTGTGATGGTGTAGTCTCT GGTACTGTGGTCTGCTAATCAATCGTCTAGTGTGGTAATCCTACTCCGTGCTAATGCTC ATAATGCTACTGGTAACTCCTAATCCTCATGCTAATGGTGTGCTCGTAAGTACTCATCGTACTAA TAATAATGGTCTACTGGTAACTCCTCATCGTGGTACTAGTCTACGGTCCGCTAAGATCAGA <u>TAAGTGATACTGATCCTCGTAATGCTACGGATCGCGTAACATCAAGTCACCTGTGCAATGGGA</u> <u>TGAGGTCCACGTAATAATCGTGGTAGTGTATCATAATGCGTTCATCGTACTAATCATCTGTATAA</u> <u>TCCATCGTGTATCATGATGGTAATCCTACTGGTGGTGTGATCATCTCTAG</u>
n5T (427 bp)	CTGAGGTACGCTGACATGCGATCGAATCGCGTTCGAGTAGGGTAGGCTGCCTACAGTGAGCTG GAATAAGATACCCTCAACTGGGATGCCATGCCCTGGAATGCCATGAAATAGAATCCGATGGGG TACAGTCCGCTGCACTAAGTACAATCGAGTAAAGCTCACGTACCATCCCATGACCTAACTCC GATCGGCTCACCTCGCGTACCCTAAACTCCACTGAGATCCAGTGGACTAAACTGCCCTGGGCT GGCGTGGCGTCCGCTAAGATCAGATAAGATCCACTCCAATAAACTAGACTCGGATCGCGTAAAC <b>ATCAAGTCACCT</b> CAATGGGATGAGGTCCACTGCGCTGCGATAAGATCAGCTGAAATAACATA CGCTCGAGTCAACTGCGATCGCGTGGCTCCGATCAAGTCCAGTGCATA
n7T (427 bp)	CACTAGAAAAGTCACAGATCCGCCATCAACCCTGGCGCATCAGAAATGAACGGTAGAAGGTCCG GCCTCAGCGCTGCGGCCCTACGCGCTGGAGGGTACGCACCTCGGAACTACAAGCTACAACGTCAA CAGTGGCCTAGACGCTCGCGCTGGACGGTCCGCAATCCGGACTCCGGGATCGCAACTCAA GAGTCCGGCTGCGAGGTCCGGCTAGAGCTAAAACTCGCGAGTCCAGCCCTCGGGGCTGCG CAGTAACGGTCCGCTAAGATCAGATAAGCGGTGAAAGCTGAAACCTAGGCGGATCGCGTAAACA <b>TCAAGTCACCT</b> CCCAATGGGATGAGGTCCACTCCACGCTCGGCCGTCAAACCTAACCCTAGA AACTCGACGCTCGACGCTGAAAAGTCAACCTCAAGCGTAACCGATACG

Highlighted section indicates “n4T” distribution within n3T; underlined and bolded sequences indicate shared sequences used for qPCR amplicon for outer primers and probe primers, respectively.

Table 2-3. Primers used for qPCR

Primer name	Description	Sequence (5' → 3')
n5_qPCR_fo	Forward primer for probe qPCR	GGTCCGCTAAGATCAGATAAG

n5_qPCR_re	Reverse primer for probe qPCR	GTGGACCTCATCCCATTG
n5_probe	5'-Cy5 fluorophore conjugated and 3'-Iowa Black quencher conjugated probe primer for probe qPCR	/5Cy5/CGGATCGCGTAACATCAAGTCACCT /3IAbRQSp/



## **Chapter 3: Selection by photoaffinity crosslinking in complex environments (SPACE)**

---

### **3.1. Acknowledgement and contribution**

Study conceptualization, synthetic library, mRNA display workflow and optimization, and several selection rounds were designed and performed by Dr. Igor Dodevski. I designed the nanobody expression, photocrosslinking optimization, single clone screening, several selection rounds, and control experiments with Dr. Igor Dodevski and performed experiments independently. I designed and performed all mammalian cell culture experiments and optimizations (in Appendix).

### **3.2. Summary**

As discussed in Chapter 2, selections against cell surfaces are an important yet highly challenging problem, and reducing nonspecific adsorption (i.e., noise) is a useful tool in improving the signal-to-noise ratio (SNR) in a complex selection system. In contrast to our study in the previous chapter in which we aimed to minimize overall nonspecific adsorption through PEGylation, our approach in this project for increasing the SNR was to enrich for true binder-target interactions by covalently crosslinking the binder to its panned target, pulling down the binder-target complex onto solid support, then performing harsh washing with denaturants to drastically reduce non-ligand interactions. By stabilizing the binder-target complex through a covalent linkage, the user can “lock in” the specific interaction which occurred through the binding domain and its target. Here, we developed an mRNA display selection system with incorporation of a photoactivatable amino acid for covalent ligand-target photocrosslinking following ultraviolet (UV) irradiation.

In initial selections against an array of soluble targets, a synthetic single-domain antibody (nanobody) library was used. A key feature of the library is a methionine-free

framework to enable strategic, rational incorporation of methionine positions in the binding paratope for subsequent crosslinking to targets using L-photo-methionine (pMet), a photo-activatable analog of methionine (Met).<sup>123</sup> Single Met (ATG) positions were incorporated at various locations within the nanobody scaffold near possible binding regions. For translation, a cell-free expression system was used with a custom amino acid mix of 19 natural amino acids (minus Met) and pMet.

While cell surface targets were prepared (HER2- and CXCR4-C-terminally tagged plasmids transfected into HEK293 cells), selections were initially performed on soluble protein targets. Despite performing two selection campaigns and analyzing dozens of single clones, we were unable to enrich for specific binders against any of our targets. Thus, instead of working with a naïve library, we used a known ligand-target interaction of the designed ankyrin repeat protein (DARPin) Off7 and maltose binding protein (MBP) for further proof-of-concept studies.<sup>14</sup> Again, we removed internal Met positions from the Off7 scaffold and mutated specific positions around the binding paratope to Met based on crystal structure information to utilize our pMet-enabled photocrosslinking approach. For a control mRNA display round with one of the Off7 mutants, D112M, against biotinylated MBP, we achieved 16- and 9-fold enrichment for UV-exposed samples over non-UV-exposed D112M and UV-exposed WT (no internal Met) samples, respectively. Our proof-of-concept results show dramatically reduced background and particle recovery using the SPACE platform.

### **3.3. Introduction**

mRNA display, an *in vitro* display technology, enables selection and directed evolution of protein binders from large, diverse libraries of up to  $10^{14}$  sequences.<sup>124</sup> In mRNA display, a cDNA library is transcribed into mRNA, which is subsequently ligated to

a short DNA spacer with a 3' puromycin molecule.<sup>22</sup> During translation, when the ribosome reaches the end of the gene (with or without the stop codon), the puromycin molecule, a tyrosyl-tRNA analogue, enters the A site on the ribosome and covalently attaches to the nascent polypeptide chain in the P site,<sup>22</sup> thereby covalently linking the protein to its cognate mRNA. The mRNA-protein fusion particle can be reverse transcribed to form mRNA-cDNA heteroduplex for added stability and reduced secondary structure<sup>125</sup>, although our lab has successfully shown in a previous study that the reverse transcription (RT) step may be delayed after selection to streamline the selection procedure and mitigate potentially adverse effects of elevated temperatures required for the RT process on displayed proteins.<sup>69</sup> The display particle can then be panned against a target of interest, washed to remove nonspecific binders, and recovered by RT-PCR for sequencing and/or additional selection rounds.<sup>125</sup>

As mentioned in Chapter 2, however, current *in vitro* display approaches including mRNA display are typically not amenable to panning in complex environments like cell surfaces despite the high therapeutic potential of drug discovery against the membrane surfaceome. Further, binder libraries such as those used in RNA-based display methods are usually naïve, synthetic/semi-synthetic libraries which have many (up to  $10^{15}$ ) unique sequences.<sup>46,126</sup> While the probability of identifying high-affinity binders increases with library size and diversity<sup>126</sup>, the number of non-functional library members with frameshifts, stop codons, and redundant and/or undesirable amino acids (i.e., noise) can also increase; such non-functional members may require the user to repeat additional selection rounds with extra selective pressures, but more likely ruin selection campaigns by outcompeting true binders.<sup>24,58</sup>

Reducing overall nonspecific adsorption (also noise) is a useful tool in improving the SNR in a selection system; in addition, one can also improve the SNR by dramatically

enriching the ligand-target interaction above background in a sea of non-functional and/or nonspecific binders as encountered in large naïve library selections or cell surface selections. Taking a different approach to our work in Chapter 2, our goal was to enrich for functional *and* specific binders by covalently crosslinking the target to its binder through photocrosslinking at the binder interface. Though there have been a few successful examples of photocrosslinking applied to *in vitro* display methods in recent literature<sup>127–129</sup>, they require detailed structural information of the target protein or have only been applied to DNA libraries (e.g., DNA aptamers or DNA-encoded libraries) due to limitations with crosslinking.

In our approach, we performed crosslinking of the binder library to its target at the binding interface using L-photo-methionine (pMet), which structurally resembles methionine but contains a photoactivatable diazirine ring that creates a reactive carbene that covalently crosslinks to nearby side chains after light-induced loss of nitrogen (Figure 3-1).<sup>123</sup> For both our library and proof-of-concept panning experiments, single pMet substitutions were strategically placed at various positions within or at the binding interface to allow for crosslinking as a result of antigen-specific binding. In addition, for the photo-reactive amino acid, our goal was to choose not only a commercially available amino acid that could easily be incorporated into an existing cell-free expression workflow for *in vitro* display, but also an amino acid that would require few substitutions within the binder framework. With these requirements in mind, we chose pMet over L-photo-Leucine (pLeu), another commonly used photoreactive amino acid.<sup>123</sup> Unlike other crosslinking molecules such as bifunctional chemical crosslinkers, the crosslinking is highly specific due to the short half-life of the activated-state carbene and the absence of any spacer molecule.<sup>123</sup> pMet, along with other photoactivatable amino acids, is nontoxic and has been used for protein-protein interaction studies in living cells.<sup>123</sup>

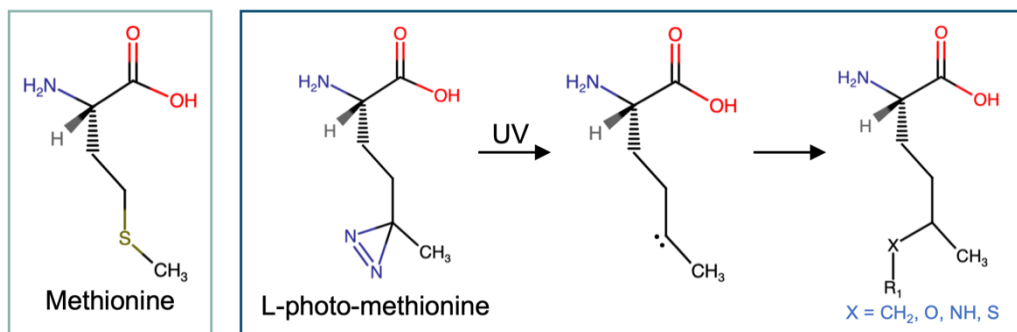


Figure 3-1. Structural comparison of methionine to L-photo-methionine.

The diazirine group on the pMet upon UV activation becomes a reactive carbene, allowing crosslinking to nearby molecules.<sup>130</sup>

For our application in library selections, by trapping all target binding-driven interaction of the library binder, crosslinking can greatly improve enrichment of diverse (i.e., both high- and low-affinity) binders over non-binders and allow selections against protein targets at low concentrations, such as those expressed on the surface of cells or protein targets in-solution. Additional target protein labeling (e.g., via C-terminal SNAP-tag, which forms covalent thiol bonds with O<sup>6</sup>-benzylguanine derivatives<sup>131</sup>) allows for complete covalent linkage from the mRNA to its encoded binder to the target; this linkage can be advantageous for performing harsh, denaturing washes to greatly minimize noise in the selection.

For our initial library selections, we used the variable domain of heavy-chain-only antibodies (VHH, also known as single-domain antibodies or nanobodies) found in the Camelidae family. Compared to conventional biologics such as monoclonal antibodies (mAbs) which are 150 kDa, nanobodies are much smaller (~14-15 kDa), which allows for not only easier production and high stability, but also the ability to bind to unique epitopes that are not accessible to mAbs (Figure 3-2A).<sup>132</sup> Their structure includes four conserved framework regions (FRs) flanking three complementarity determining regions (CDRs) (Figure 3-2B). While amino acid profiles and lengths of CDR1 and CDR2 are relatively

conserved, CDR3 can be highly variable in both amino acid composition and length.<sup>133</sup> These differences, particularly in CDR3, can change the binding surfaces, allowing the nanobody to take on concave, loop, or convex structures.<sup>134</sup> With increasing therapeutic interest<sup>135</sup> and many successful nanobody binder discoveries by directed evolution<sup>132,134,136,137</sup>, we aimed to engineer the nanobody scaffold for our binder selection campaign.

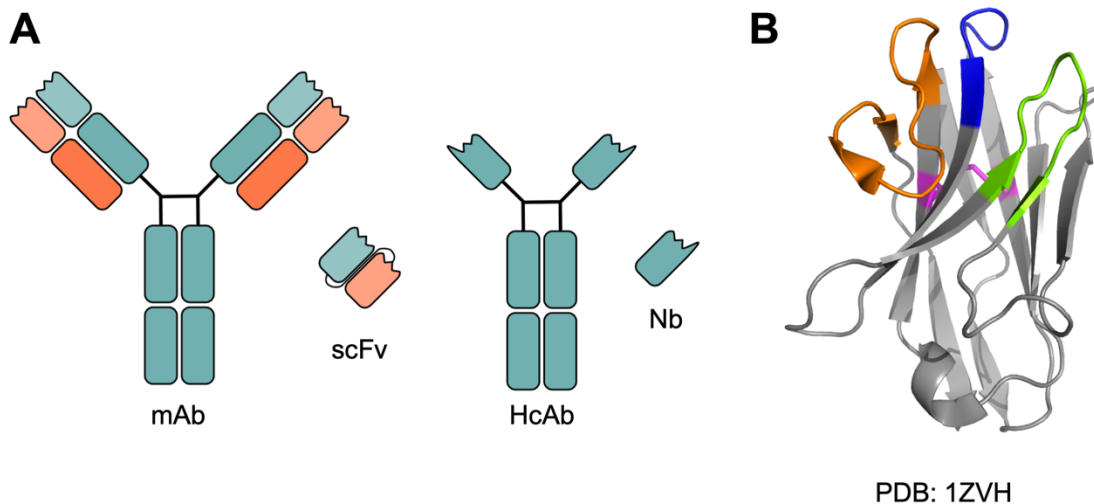


Figure 3-2. Nanobody structure.

(A) Comparison of mAb and single chain variable fragment (scFv) (left) to nanobody (Nb), derived from heavy chain antibody (HcAb) (right) (B) Detailed annotated structural information of a nanobody scaffold (taken from PDB: 1ZVH).<sup>138</sup> Four frameworks (FWs) flank three variable domains (CDRs); disulfide bond (labeled as sticks) bridging FW1 and FW3 help stabilize the fold.<sup>139,140</sup>

In this project, our goal was to develop a platform technology for discovery of synthetic nanobody binders in complex environments (such as on cell surfaces or within a highly diverse, large naïve library full of non-binders) by incorporating photocrosslinking of the nanobody library to its target (Figure 3-3). Key features of this approach are the (1) ability to engage targets in their relevant state or in low concentrations and (2) retaining a high signal-to-noise ratio.

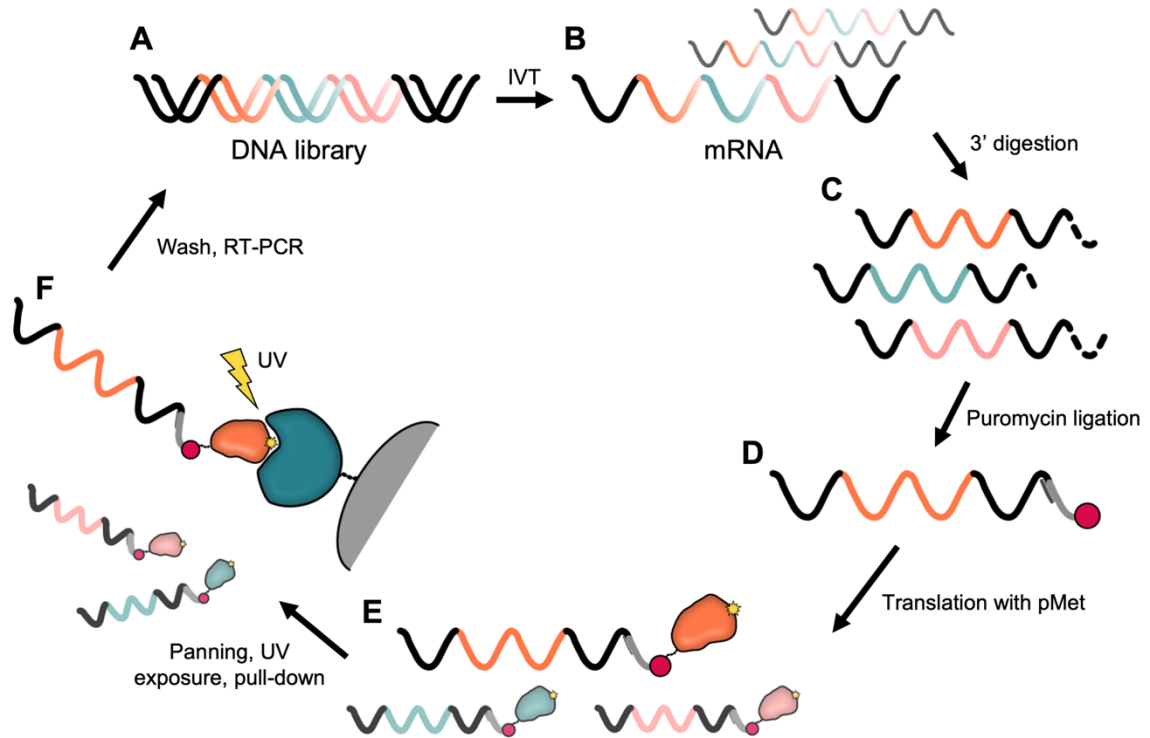


Figure 3-3. Workflow for a SPACE display selection round.

(A) DNA library is *in vitro* transcribed to (B) mRNA with variable 3'-termini, which then undergoes (C) 3'-terminal digestion to create uniform mRNA template and (D) ligation to a puromycin-modified primer. (E) The mRNA is translated with a photoactivable amino acid (pMet) and panned against a target protein of interest. (F) Exposure to UV covalently crosslinks bound library members to the target, then the binder-target complexes can be pulled out of solution by affinity pull-down, (F) washed extensively to remove nonspecific background, and analyzed by on-bead RT-PCR.

### 3.4. Results

#### 3.4.1. Summary of previous contributions

The following subsections (3.3.1.1 and 3.3.1.2) briefly describe the conceptual and experimental contributions by Dr. Igor Dodevski for the nanobody library and photocrosslinked mRNA display protocol as they apply to the SPACE library selection campaigns.

##### 3.4.1.1. Synthetic nanobody library design

The nanobody library was designed using sequence and structural analysis of previously characterized nanobodies and/or nanobody-target complexes, found as crystal structures in the Protein Data Bank (PDB) and previously published synthetic nanobody libraries.<sup>134,139</sup> Key features of the nanobody library include: (1) stable conserved framework regions, guided by consensus design of crystal structures, (2) maximal structural diversity in the CDR3 binding paratope, by including cloning strategies to produce loops of varying lengths (“Long CDR3s” with 13 - 16 amino acids, and “Short CDR3s” with 7 - 10 amino acids) which mimic the diversity found in natural nanobodies (similarly described by Chen *et al.*<sup>133</sup>), and (3) a methionine-free framework that allows for strategic, rational incorporation of methionine positions in the binding paratope, for subsequent crosslinking to targets using pMet. To this end, one framework Met was mutated to Leu (M83L), and single Met (ATG) positions were incorporated at six possible locations, at one of three residues within CDR2 or within CDR3. Two different versions were designed; the first iteration of the library had a fixed CDR3 length of 16 amino acids with 9 variable positions (encoded by NNK), while the second-generation library contained variable CDR3 lengths (7 - 16 amino acids).

#### 3.4.1.2.mRNA particle preparation

The mRNA display particle preparation primarily follows that of Barendt *et al.*<sup>69</sup> with three main changes: (1) exonuclease digestion of the 3' terminus of the mRNA prior to splinted ligation, (2) “split-and-combine” approach for introducing pMet, and (3) group separation for purification. First, following *in vitro* transcription (IVT) and mRNA purification by a series of salt- and ethanol-precipitation steps, the 3' terminus of the mRNA was trimmed to create uniform termini by Exonuclease T (Exo T), a single-stranded RNA/DNA specific nuclease that cuts with varying efficiency to different nucleotides (Figure 3-4A).<sup>141</sup>



Transcription products synthesized with T7 RNA polymerase are usually heterogeneous at the 3' end<sup>142</sup> and the splint-enabled ligation that bridge the 3' end of the mRNA and the 5' end of the puromycin linker to the puromycin linker require the ends to be in perfect register for efficient ligation.<sup>143</sup> By creating uniform 3'-termini for all mRNA transcripts via Exo T digestion, we can eliminate mismatches and gaps created by run-off transcription and increase the efficiency of puromycin ligation. Exo T exhibits unusual base specificity, and its nuclease activity is largely defined by the last 4 residues on a single-stranded template (although it can bind up to 10 nucleotides in its active site), with >100-fold reduced activity at a single 3'-C residue and loss of activity at 2 consecutive terminal C residues.<sup>141</sup> Based on a previous study comparing nuclease activity between templates with CC-containing 4-nucleotide termini, we chose 'AACC' as it showed near-complete loss of enzymatic activity.<sup>141</sup> Using an analytical Cy5-labeled primer (performed in parallel with puromycin-primer), we detected 25-40% ligation efficiency of the primer when splinted with the mRNA transcript (Figure 3-4B), which is notably higher than typical splinted ligation efficiencies which may be  $\leq 10\%$ .<sup>144,145</sup>

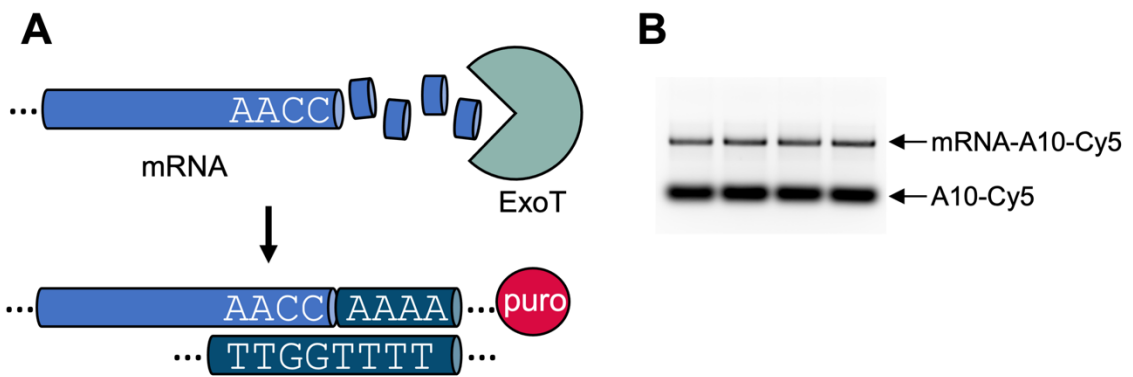


Figure 3-4. Additional mRNA preparation prior to selection.

(A) The heterogeneous 3' terminus is made uniform by Exo T digestion to 5'-AACC-3' and ligated to a polyA-puromycin primer (polyA-puro) with a ssDNA splint and T4 DNA ligase. (B) Analytical gel of polyA-Cy5 ligation shows an average of ~40% ligation efficiency (as analyzed by ImageJ) of primer for four different biological reactions (different libraries used in this experiment).

Next, for translation of the mRNA-puromycin template, we modified the PURExpress cell-free expression system with our “split-and-combine” protocol to accommodate the pMet addition for our photocrosslinking approach. Briefly, the translation mix was split into two; one containing the 19-amino acids (minus Met) and the other with pMet and recombinantly expressed MRS5m, an engineered tRNA synthetase that catalyzes aminoacylation of pMet (with an enlarged diazirine group) to a methionyl tRNA.<sup>146</sup> To favor pMet incorporation, these two mixtures were incubated for 20 min at 37 °C before combining and adding the mRNA-puromycin template for translation to proceed. Following translation, the mixture was put in a high Mg(OAc)<sub>2</sub> buffer to stabilize the ternary ribosome-mRNA-protein complex to allow for fusion of the nascent polypeptide chain to the 3'-puromycin. Finally, to cleave the start codon (initiator Met, which would also be a pMet since no other Met is present in the reaction), protease cleavage was performed such that the UV-induced covalent linkage only occurred at the binding interface.

Third, to purify the translation and fusion mixture, we utilized group separation purification to remove excess pMet/other small molecules and simultaneously perform buffer exchange. While many approaches also include polyacrylamide gel electrophoresis (PAGE) or dT-based purifications to remove non-fused proteins and mRNA<sup>30</sup>, with the inclusion of the photo-crosslinking step, we reasoned that free mRNA and protein that were not covalently linked should not be recovered in excess post-selection.

#### *3.4.2. Development and optimization of library-target photocrosslinking conditions*

A key feature of our mRNA display approach is the ability of the translated protein to covalently crosslink to its target. Photocrosslinking via UV exposure, however, is nontrivial as too little exposure may not allow sufficient crosslinking if the photoactivatable amino acid is buried within the protein-protein interface, but too much exposure may result

in overheating of the panning mixture, which can denature the proteins and/or adversely affect ligand-target interactions. In addition, prolonged UV exposure can also damage the genotype (RNA) by causing it to self-cleave by UV light-induced oxidation or inducing photochemical modification.<sup>147,148</sup> Thus, we aimed to establish experimental conditions which maximized photocrosslinking efficiency while minimizing overall UV exposure. For these studies, we optimized conditions using a model interaction between the designed ankyrin repeat protein (DARPin) Off7. Substitutions were made for native internal Met positions based on published data<sup>149</sup> (M34L) and modeled from the Off7-MBP co-crystal structure (M109S and M114L)<sup>14</sup>, and a single internal substitution (D112M) at the binding interface was introduced for pMet incorporation.

For crosslinking analysis, we used sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel-shift assays to identify crosslinked ligand-target complexes after UV irradiation.<sup>123</sup> Since the PURExpress system used for pMet-incorporated translation contains many other unrelated proteins, we labeled Off7<sub>D112M</sub> ligand with Alexa Fluor 647 for imaging of only the ligand and/or ligand-target complex in the Alexa Fluor 647 channel. We panned against MBP in solution and UV irradiated the panning mixture before running the sample on SDS-PAGE. For the UV irradiation, we tested two different approaches: light exposure via (1) the DAPI filter cube on a Cytation plate reader (better control of light intensity and light wavelength cutoffs, but weaker energy) and (2) a traditional UV lamp used for curing polymers (higher energy, but broader exposure and less control of light intensity). After repeated troubleshooting, we tested mild UV exposure conditions – exposure for 2, 4, or 6 min with the DAPI channel and two 15- or 30-second intervals with the UV lamp – and observed varying levels of crosslinking for both UV irradiation approaches. In the SDS-PAGE gel analysis, the Off7<sub>D112M</sub> protein is visible in both unexposed (negative control) and UV-exposed lanes (Figure 3-5). In

conditions with higher UV exposure (longer DAPI channel or UV lamp exposure), we observed a higher molecular weight band corresponding to MBP-Off7 crosslinked complex in UV-exposed samples. Interestingly, the gel showed two distinct target-ligand complex bands; a strong band at a higher molecular weight, and a fainter band with a slightly faster migration, which may correspond to a complex in which the target was crosslinked to the initial starting Met on the ligand (for these analytical studies, the C3 protease cleavage to remove the starting Met was not performed). Notably, we noticed that samples with longer DAPI exposure (6 - 10 min) became warm and bubbly, which may be a sign of protein denaturation from prolonged exposure. Since our results showed sufficient crosslinking with the 15-second / UV lamp exposure condition, we used this condition with the selection experiments and single clone screens moving forward.

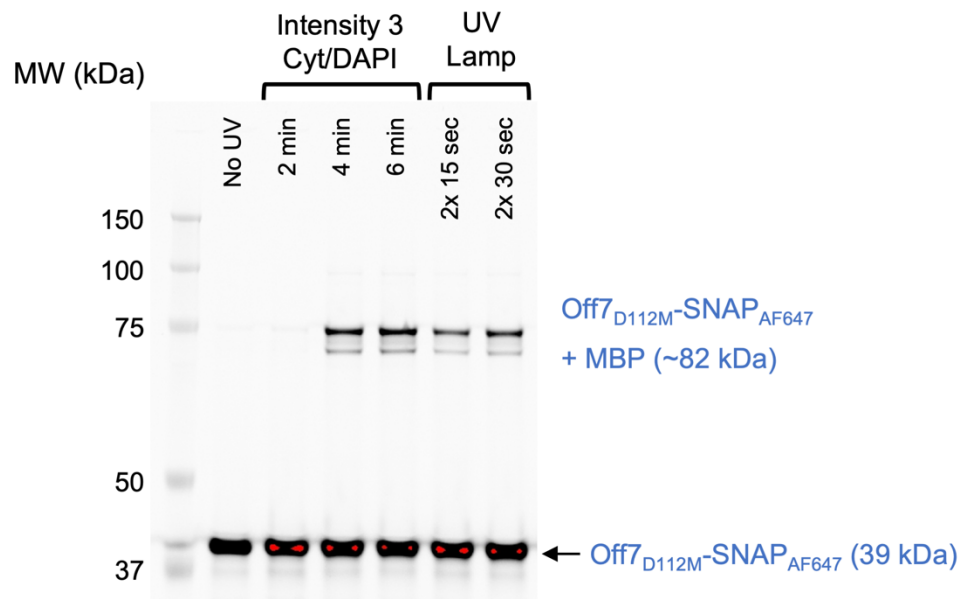


Figure 3-5. Comparison of photocrosslinking conditions.

Reaction conditions were tested with the Off7-MBP model ligand-target interaction imaged on SDS-PAGE with Alexa Fluor 647 channel (image is overexposed to show detail). Similar levels of crosslinking were achieved with 2x 15- or 30-second interval UV exposure (with 2-minute interval in between) using a UV lamp and long UV exposure using the DAPI channel. Crosslinked product migrates slightly faster than predicted molecular weight; however, this may be due to intermolecular crosslinking, leading to a different hydrodynamic radius.

### *3.4.3. Initial affinity selections and single clone screens of the nanobody library against model protein MBP*

For initial nanobody library selections, we panned against the model protein MBP using a library with a CDR3 loop length of 9 amino acids with six randomized single Met positions (L9MX) alongside a control library with no internal Met positions (L9M0). The MBP target was expressed as an N-terminal fusion to a SNAP-tag, a self-labeling protein that binds to benzylguanine (BG) derivatives using a covalent thioether bond via reactive cysteine residue, to biotinylate the binder using BG-biotin. This biotinylation allowed the entire library-target complex to be pulled down onto magnetic streptavidin beads to perform harsh washing protocols to reduce nonspecific background. Although biotin-avidin binding is technically a non-covalent interaction, the binding is the strongest known non-covalent interaction ( $K_D = 10^{-15}$  M) that can withstand high pH and temperature, and mild levels of denaturing agents like urea and SDS.<sup>150-152</sup> Thus, we used stringent washing buffers including 1% SDS or 6 M urea with heated (30 °C) incubations (see Methods). After 2 rounds, the selections with the Met-containing L9MX library showed strong enrichment for UV-exposed samples compared to non-Met containing control library L9M0. Analysis of 18 in-frame single clones showed high diversity retained at all CDRs, particularly in CDR3 (Figure 3-6A). Following the same “split-and-combine” approach for pMet incorporation using cell-free protein expression, the selected clones were subject to subsequent crosslinking gel-shift assays with MBP, which showed that 12 clones had some level of crosslinking to the target (Figure 3-6B and 3-7). Compared to the positive control (Off7<sub>D112M</sub>), many crosslinked bands were fainter, which suggested weaker binding and/or decreased ability to crosslink to the target.

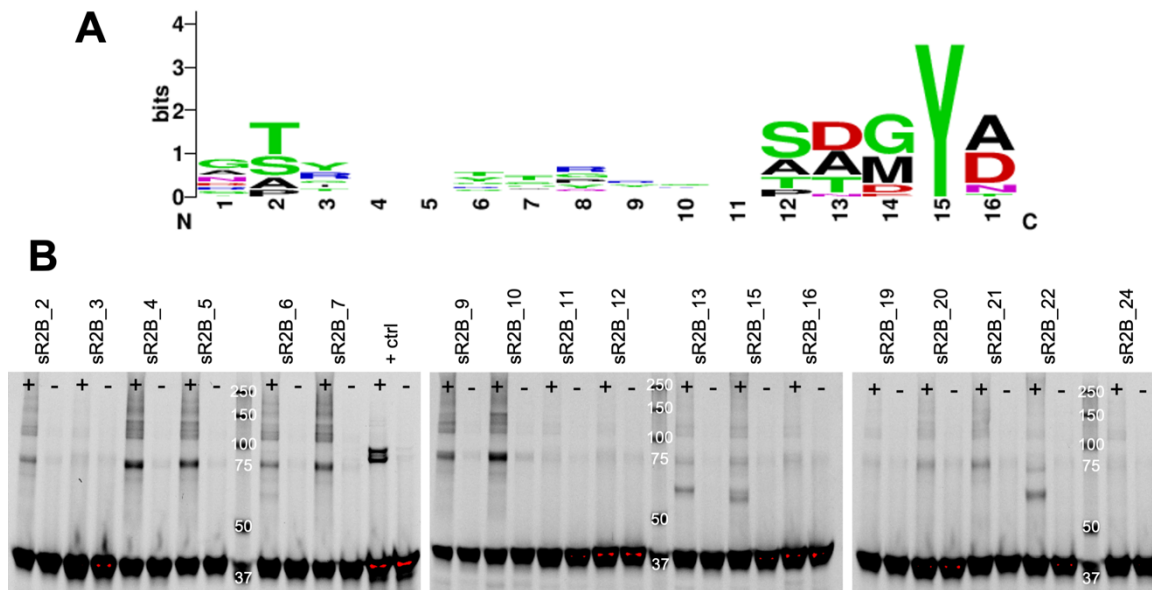


Figure 3-6. Single clone screens of binders recovered from library selection. (A) WebLogo<sup>153</sup> of amino acid distribution at CDR3 shows diverse amino acid composition. Tyrosine at the 15<sup>th</sup> position (within CDR3) was not randomized. Positions 4, 5, and 11 were blank due to lack of sequence conservation. (B) Photocrosslinking assay of selected binders (nomenclature for 'sR2B' from selected Round 2 against MBP) show varying level of crosslinking to MBP upon UV exposure (+) compared to non-exposed (-). Nanobody-SNAP is ~37.5 kDa, and nanobody-SNAP crosslinked to MBP is ~80.4 kDa. Positive control is Off7<sub>D112M</sub>, as used in Figure 3-5. (Gel image is overexposed to show detail.)

To further assess functional MBP-binding of the recovered clones, we expressed the nanobodies as SNAP fusions via PURExpress (standard reaction, without pMet) and immobilized the binders onto BG-functionalized polystyrene beads. We then allowed the immobilized nanobodies to bind with Alexa Fluor 647-labeled MBP in solution and analyzed the fluorescence using flow cytometry. Using 1  $\mu$ M of target MBP and 10  $\mu$ M competitor (free MBP), only 4 clones showed marginally higher signal above competed samples (Figure 3-7). The 4 clones were then panned against 10  $\mu$ M target and 100  $\mu$ M competitor; in these experiments, the uncompleted vs. competed were either equal (showing no specific binding) or had reverse trends. Repeated experimentation showed

that the trends remained the same, indicating that none of the recovered clones were true binders to MBP.

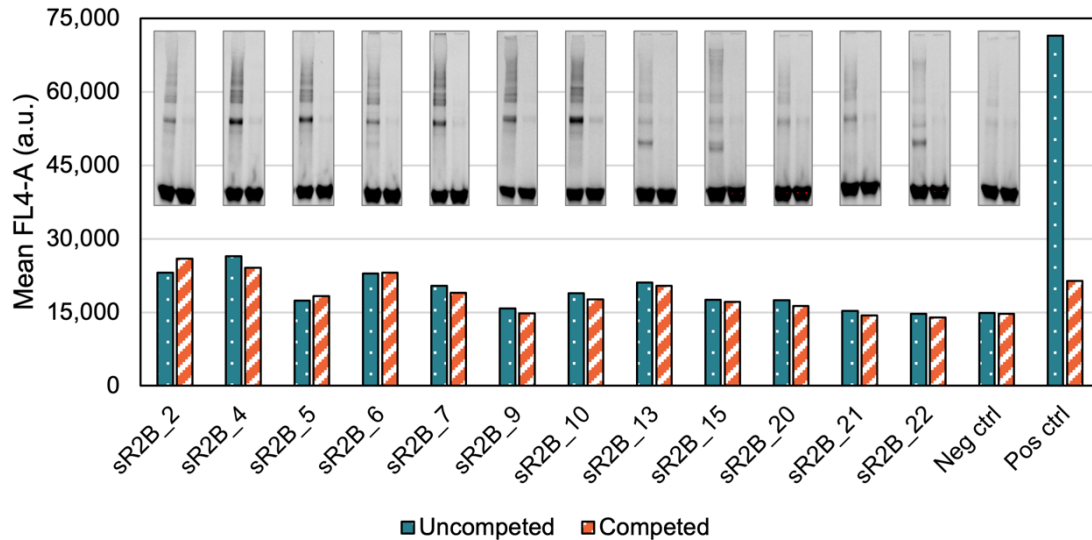


Figure 3-7. Binding assay with selected clones.

Flow cytometry binding assay of selected binders with 1  $\mu$ M binder and 10  $\mu$ M competitor (free MBP in solution) show minimal binding activity compared to positive control (anti-MBP2 nanobody). Negative control is a recovered clone with no internal Met and no crosslinking activity. Comparison to gel-shift assay, as noted toward the top of the graph, shows that crosslinking ability (i.e., intensity of gel-shifted band) does not seem to correlate with binding ability, which suggests that the crosslinking and binding are nonspecific. Further studies confirmed lack of binding even with higher concentration of target protein.

#### 3.4.4. Efforts toward additional selections against MBP and other targets with second-generation nanobody library

As nanobody binding paratopes can vary with CDR3 loop lengths, we reasoned that perhaps the fixed CDR3 length (9 amino acids) in the first iteration of the library did not provide sufficient adaptability in the binding interface to bind to MBP. Thus, we performed additional selections with a second-generation nanobody library in which the CDR3 loops were more varied between 7 and 16 amino acids (LNMX). Target proteins were also extended to include biotinylated sfGFP, hHER2 ECD, human angiotensin converting enzyme 2 (hACE2), and Spike receptor binding domain (S RBD) in addition to

MBP. For sfGFP, strong enrichment for UV-exposed samples was observed after 3 rounds of selection (Figure 3-8A). However, analysis of 26 in-frame single clones for sfGFP showed that the library had nearly converged to a single sequence with only a single Met position (M56, out of six possible locations) represented, and crosslinking gel-shift assays showed either no crosslinking or crosslinking against non-target proteins. The selections against MBP showed modest enrichment for UV-exposed samples (Figure 3-8B), and analysis of 7 (out of 15) in-frame single clone screens showed more variable CDR regions; however, gel-shift assays once again showed minimal or no cross-linking. The selections against hHER2, hACE2, and S RBD all showed no enrichment for UV-exposed panning solutions, with some showing opposite results after 2 selection rounds (Figure 3-8C). Thus far, these results demonstrated failure of the nanobody library selection campaign.

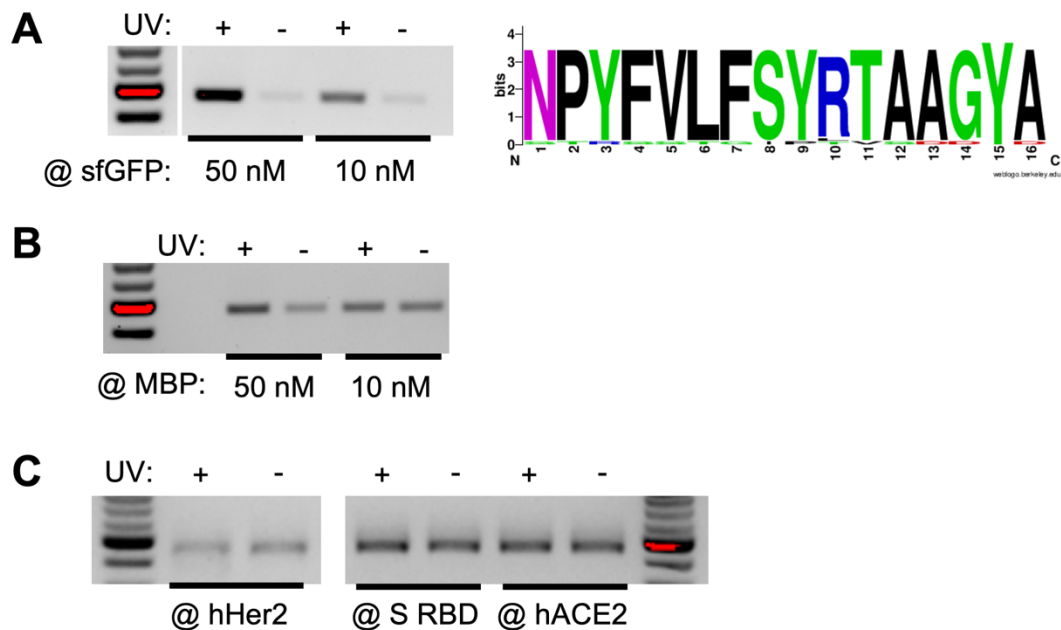


Figure 3-8. RT-PCR and single clone data from second library selection.

(A) Strong enrichment for selections against sfGFP was observed after 2 rounds in RT-PCR analysis, but analysis of 26 single clones (from selections against both 50 nM and 10 nM target) showed near convergence (CDR3 shown, in Web Logo format) with many aromatic residues. Subsequent crosslinking studies (data not shown) demonstrated no target-specific binding. (B) Modest enrichment for MBP selections was also observed after 2 rounds, but no specific crosslinking was observed in further studies. (C) Selections against 50 nM hHER2, S



RBD, and hACE2 target antigens showed no UV-specific enrichment after 2 selection rounds. (RT-PCR data from 3-8A and 3-8B are courtesy of Dr. Igor Dodevski.)

#### *3.4.5. Proof-of-principle experiments demonstrate successful crosslinking and recovery of mRNA displayed particle to target*

Pivoting away from library selections, we performed proof-of-concept experiments to demonstrate the feasibility of our photocrosslinking-mRNA display method using the ligand-target interaction between Off7 and MBP which we had previously used for establishing photocrosslinking assays. In addition to the Off7<sub>D112M</sub> mutant with one internal Met position at the binding interface, we included a wild-type (WT) control with no internal Met position. Following the same mRNA terminal modification, translation, and affinity selection protocols as established previously, we successfully observed a much higher signal-to-noise (SNR) difference for the UV-exposed over non-UV-exposed sample for the Off7<sub>D112M</sub> mutant (Figure 3-9A and 3-9B). Low signal in the control samples (Off7<sub>WT</sub>, panning against sfGFP instead of MBP) showed target and UV-specific binding, which was in contrast to our nanobody library selections. Finally, to test the ability of the displayed binders to crosslink specifically in a complex environment, we also performed the same control selection in the presence of cell lysate and saw that the high SNR was retained. The background was slightly higher, however, which we reasoned may be due to nonspecific interactions with the cell lysate in solution. In sum, the proof-of-concept data show experimental validation of our photocrosslinking mRNA display approach.

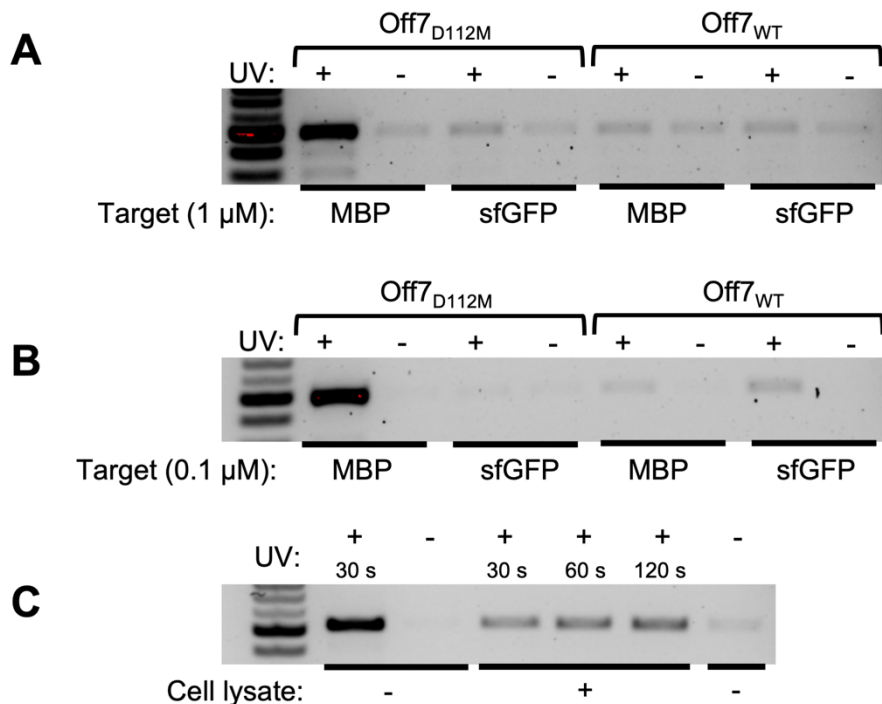


Figure 3-9. Proof-of-concept demonstration of crosslinking-enriched binder recovery.

(A) Selections against 1 μM target show UV- and target-specific binding, and (B) selections against 100 nM target shows improved signal-to-noise ratio. (C) Repeated panning selection (with Off7<sub>D112M</sub> and 1 μM MBP target) with cell lysate in solution shows decreased SNR possibly due to increased crosslinking to other proteins in the complex panning mixture but retention of crosslinking-specific enrichment.

### 3.5. Discussion

Synthetic library selections have led to successful binder discovery campaigns<sup>134,139</sup>; however, without prior antigen-driven challenge of the library as in an immunized (with immunoglobulin genes taken from immunized animals) or semi-synthetic (created by diversification of known binder) library, massively large libraries must be screened to find specific binders.<sup>154</sup> Considering that the number of non-functional proteins correlates with increasing library size, specific binders that are weak or lowly presented may be outcompeted and lost in early selection rounds. In addition, the required rounds of amplification can lead to over-enrichment of single clones, and the increasing selection pressure may only allow clones with the highest affinity (as governed by slowest

dissociation) to be recovered, which can heavily reduce the diversity of the selected binder pool. With these challenges in mind, we sought to develop an *in vitro* mRNA display platform that allows streamlined discovery of specific and diverse nanobody binders through binder-target crosslinking.

Therapeutic interest in nanobody binders has grown substantially since the discovery of this scaffold in 1993, with many nanobodies (in monovalent and multivalent formats) in the clinical approval pipeline.<sup>155,156</sup> Its small size, ability to acquire different binding paratopes via varying CDR3 domains, and stability make the nanobody a highly useful protein scaffold with many applications. In our display approach, we engineered single Met substitutions throughout CDR2 or CDR3 within the nanobody to allow for photocrosslinking to the target via pMet. While our selection campaigns ultimately failed, photocrosslinking between selection display particle and target protein has been previously shown in literature. In a recent study, Chen *et al.*<sup>128</sup> demonstrated crosslinking of phage-displayed single-chain variable fragment (scFv) libraries to soluble target antigens expressed with p-benzoyl-L-phenylalanine (pBpa), a photoactivatable noncanonical amino acid (ncAA). In contrast to our study, the photoactivatable amino acid was engineered within the target antigen, as the goal of this study was to direct scFv binding towards a specific epitope on the antigen. Despite their success, the single point ncAA-antigen engineering required detailed structural information and epitope-specific positive and negative controls to identify optimal pBpa mutations and produced only modest, albeit specific, binders after multiple rounds of selection, alluding to the complexity of this photocrosslinking selection approach. However, the concept of engineering the photocrosslinking mutations on the target antigen instead of the binder library, as we did in this study, may be a viable path forward with our approach. In addition to recombinantly expressed soluble targets, application of amber suppression systems

can also enable incorporation of photoreactive amino acids and other non-canonical amino acids in cellular expression systems for cell-surface panning.<sup>157–159</sup> Finally, other recent efforts toward engineering display systems with target-binder crosslinking ability<sup>127,129,160</sup> require pre-labeling of target antigens or placing photoactivatable molecules away from binders (on the genotype), which can limit the generalizability of the approach and decrease target-specific enrichment, respectively.

In contrast to our selection campaign, our proof-of-concept experiment with Off7<sub>D112M</sub> and MBP showed target-specific and UV-crosslinking-enabled binding using our modified mRNA display approach. The modest crosslinking ability of nearly all recovered binders from the library selection suggest nonspecific crosslinking (i.e., not target-binding mediated), which is the opposite of our control Off7<sub>D112M</sub>-MBP experiment, in which the pMet placed precisely in a slightly flexible region within the binding interface, adjacent to residues that interact with MBP.<sup>14</sup> The added flexibility may allow the UV-activated pMet to extend and crosslink to nearby residues more easily, although even our crosslinking optimization experiments with purified protein showed minimal crosslinking (Figure 3-5), which suggests the overall difficulty of our approach. Given the potentially sub-optimal pMet positioning, a possible modification to our strategy may be to decrease the CDR3 loop length diversity such that pMet can be placed in a more favorable position within the binding interface, much like the D112M position in Off7. Trimming the 3'-terminus of mRNA with Exo T digestion produces uniform termini which increases the efficiency of puromycin ligation; although we used a splinted ligation approach for puromycin linkage, our digestion protocol is broadly applicable to other mRNA terminus ligations like Y-splinted ligation.<sup>143</sup> In addition, the UV irradiation condition we established and optimized led to specific photocrosslinking with no notable damage to the RNA as suggested by our non-binding Off7<sub>WT</sub> control (RT-PCR recovery is the same for both UV and non-UV

exposed samples). Since psoralen-mediated UV-crosslinking (~15 min exposure) is another common method used for puromycin crosslinking in mRNA display particle synthesis, it is possible that RNA is not adversely affected by mild UV exposure.<sup>30,31</sup> Notably, we tested other Off7 point mutants around the Off7-MBP binding interface for pMet incorporation and found varying degrees of crosslinking efficiency; similarly, additional photocrosslinking screens using a nanobody MBP binder (e.g., the positive control used in the single clone screens) can inform possible locations for better pMet substitutions on a nanobody scaffold.

The wash cycles post-panning and pulldown for removing nonspecific binders involved harsh reagents such as 0.5% w/v Tween-20 (10-fold higher concentration than in ribosome display washes), 1.5% SDS, 10 mM EDTA, 1 mM DTT, and 6 M urea. In other display methods that are non-covalent (e.g., ribosome display) or cell-based, such reagents cannot be used due to particle instability. The ability to perform harsh wash cycles is unique to our crosslinking mRNA display approach, even though the biotin-avidin target pulldown, despite its  $K_D = 10^{-15}$  M, is technically non-covalent.<sup>150</sup> Using a true covalent pulldown method through protein-fusions like the SNAP-tag<sup>131</sup> or SpyTag<sup>161</sup> may further improve binder recovery as it is subjected to our washing methodology.

Similar to our work in Chapter 2, one of the goals of our photocrosslink mRNA display approach was to perform selections directly on cell surfaces, a highly complex panning environment which may greatly benefit from covalent binder-target linkage and harsh washing to remove non-specific membranes, proteins, and other biomolecules. To this end, C-terminally tagged (via SNAP-tag) cell receptor targets were prepared using transfection of HEK293 cells and SNAP-tag-enabled pulldown protocols were established using agarose beads (see Appendix); however, the nanobody selections did not proceed to cell-surface selections due to difficulties with selections against soluble antigens.

Despite the challenges with the nanobody selections, eventual application of this technology to direct cell-surface selections could be highly useful, since the crosslinking allows binder-target capture regardless of target affinity, which may be lower for a buried epitope on a surface receptor, or concentration, which could also be lower for a cell surface receptor since overexpression can lead to cell toxicity.<sup>160</sup> In a separate study, we also noted that reverse-transcribed mRNA/cDNA heteroduplex templates have low nonspecific adsorption to cell surfaces much like dsDNA (using analytical methods noted in Chapter 2 and Chapter 4), which supports the potential feasibility of a cell-surface selection approach with mRNA display.

### **3.6. Conclusion**

In summary, the photocrosslinking mRNA display method established in this study is a significant technical advancement to the emerging field of crosslinking display techniques. Crosslinking enables higher specificity of recovered binder pools in complex environments such as within large synthetic libraries full of nonfunctional binders, or within cell lysates/surfaces. Although nanobody library selections against multiple targets were ultimately unsuccessful in our study, proof-of-concept experiments show specific and UV-enabled crosslinking and enrichment of mRNA display particle to its target. The contributions in optimizing mRNA display particle preparation, cell-free pMet incorporation, and establishing harsh washing protocols are not only useful to our crosslinking display approach but also to the broader field of *in vitro* display technologies.

### **3.7. Materials and Methods**

#### *3.7.1. Expression and purification of engineered MRS5m tRNA synthetase*

The expression plasmid was synthesized by digestion-ligation reactions in pET-22b(+) plasmids (Novagen) with amino acid sequence from Lee *et al.*<sup>146</sup> and verified by Sanger sequencing. *E. coli* BL21 (DE3) cells were transformed with pET-22b(+)-MRS5m plasmid and grown overnight on LB/carbenicillin (carb) agar plates at 37 °C. Individual colonies were picked and grown in 5 mL of LB/carb in a microbial shaker set at 200 rpm at 37 °C overnight. The next day, the culture was diluted in 50 mL of 2xYT media (BD Difco) to an OD<sub>600</sub> = ~0.05 and grown for 2 hours at 37 °C at 185 rpm in 250 mL culture flasks. Expression was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD<sub>600</sub> = 0.6. The cells were grown for an additional 4 hours at 25 °C at 185 rpm. Cells were harvested at 4,000 x g for 20 min, and lysed with Lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.8, 300 mM NaCl, 5 mM imidazole) supplemented with lysozyme (1 mg/mL, Calbiochem), 6 U DNase I (New England Biolabs), beta-mercaptoethanol (14.3 mM, Millipore-Sigma), and EDTA-free protease inhibitor cocktail (Roche) and shaken gently at 4 °C for 30 min. The mixture was probe sonicated for 2 min total (6x 10 seconds on/10 seconds off) (Sonic Dismembrator Model 100, Fisher Scientific) on ice. The lysate was clarified by centrifugation at 20,000 x g for 30 min, and syringe filtered with a 0.45 μm filter (Millipore-Sigma). The supernatant was loaded onto a gravity column with Ni-NTA resin (Qiagen) and washed with 20 column volumes (CVs) of Wash buffer (Lysis buffer with 10 mM imidazole). The protein was eluted with Elution buffer (Lysis buffer with 250 mM imidazole) and rebuffed using 30 kDa MWCO diafiltration columns (Amicon) into a storage buffer (50 mM HEPES pH 7.5, 50 mM KCl, 1 mM TCEP) and stored at 4 °C.

### 3.7.2. Expression and purification of mRNA display targets MBP-SNAP and sfGFP-SNAP

MBP-SNAP and sfGFP-SNAP were cloned, expressed, harvested, and lysed as above. The clarified lysate was loaded onto a gravity column with Ni-NTA resin and washed with 15 CVs of Wash buffer, followed by 2.5 CVs of Wash buffer with 800 mM

NaCl, then 2.5 CVs with Wash buffer. The protein was eluted and rebuffed as above into a storage buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM TCEP) and stored at 4 °C.

### *3.7.3. Biotinylation of purified MBP-SNAP and sfGFP-SNAP targets*

For biotinylation of recombinantly expressed MBP-SNAP and sfGFP-SNAP described in 3.6.2, 20 µM of SNAP-fusion protein was mixed with 22 µM BG-biotin (stored as 500 µM aliquots in -80 °C) (NEB) with TBS-DTT (TBS with 1 mM DTT) up to 50 µL. The reaction was incubated for 1 hour at 37 °C, then stored at 4 °C.

### *3.7.4. Nanobody library construction*

The nanobody library was cloned by Dr. Igor Dodevski using sequential primer extension, Type IIS restriction digestion, and ligation steps using Ultramer (IDT) and/or trinucleotide (TRIM) oligos. Additional ligation and overlap PCR steps were used to stitch together framework and CDR segments.

### *3.7.5. DARPin cloning for proof-of-concept experiments*

Genes for control selections were purchased as gBlocks (IDT) and subcloned into pET-22b(+) with restriction enzymes XbaI and XhoI. Gene features including ToIA-ExoT, C3 cleavage, and methionine-free framework are highlighted in Table 3-4.

### *3.7.6. In vitro transcription*

Linear PCR products (for nanobody selections) or pET-22b(+) plasmids were amplified by PCR using primers listed in Table 3-2 and Phusion Hot Start Flex (NEB). For library selections, cycling was kept to 10 cycles to preserve library diversity, and multiple PCRs were performed to retain high input into IVT. The IVT reaction was performed with homemade T7 buffer (5X)<sup>162</sup>, rNTP mix (NEB), RNasin Plus ribonuclease inhibitor (Promega), and T7 RNA polymerase (NEB) and assembled as follows:



Reagent (concentration)	Volume or mass
PCR product	1000 - 1500 ng
T7 buffer (5X)	20 $\mu$ L
rNTP mix (25 mM each)	6 $\mu$ L
RNasin Plus	2 $\mu$ L
T7 RNA polymerase	4 $\mu$ L
Nuclease free water	Up to 100 $\mu$ L
Total	100 $\mu$ L

The reaction was incubated at 37 °C for 2 - 3 hours, and 4 U DNase I (RNase-free, NEB) was added to digest the starting DNA template for an additional 20 - 30 min at 37 °C. The template was purified using salt/ethanol precipitation steps as described previously.<sup>162</sup> mRNA was quantified by absorbance measurement at 260 nm (BioTek), aliquoted into smaller volumes, flash frozen in liquid nitrogen, and stored at -80 °C. All aliquots were subject to no more than two freeze-thaw cycles.

### 3.7.7. Preparation of mRNA for mRNA display

mRNA mass and reaction volumes were variable depending on the yield and corresponding mRNA concentration from previous steps; approximate ranges used are noted in this section. For Exonuclease T (Exo T) digestion of the 3' terminus, mRNA (mass ranging from 70 - 90  $\mu$ g) was first denatured by mixing with DEPC-treated water (IBI Scientific) to ~150  $\mu$ L, and incubating the solution at 65 °C for 10 min before snap-chilling on ice for 5 min. For digestion with Exo T (NEB), the solution was mixed with NEBuffer 4 (NEB) (1X in reaction) and 0.075 units of Exo T per pmol of RNA, to a final reaction volume of 150 - 175  $\mu$ L. The solution was incubated for 30 minutes at 30 °C and purified with a modified RNeasy column purification protocol using buffers from the RNeasy kit (Buffers RLT, RW1, and RPE) (Qiagen). Briefly, 150 - 175  $\mu$ L of Exo T digest reaction was mixed with 4X reaction volume of Buffer RLT and 5X reaction volume of 70% ethanol and loaded onto a single RNeasy column. The column was washed 1x with 500  $\mu$ L Buffer RW1, 2x with 500  $\mu$ L Buffer RPE, then eluted with 60  $\mu$ L of RNase-free water, loaded 3x onto the

column membrane for maximum elution. mRNA was quantified by absorbance measurement at 260 nm as before.

Next, 200 - 250 pmol of Exo T-digested mRNA was mixed with 1.4 - 1.6-fold molar excess polyA-Cy5 primer (for analytical experiments) or polyA-puromycin primer (for preparative experiments) (IDT, listed in Table 3-2) and 1.3 - 1.5-fold molar excess splint DNA primer. DEPC-treated water was added such that the Exo T-digested mRNA was 3  $\mu$ M in solution (typically 70 - 85  $\mu$ L). The solution was heated to 94 °C for 1 min, then cooled on ice for 5 min. The cooled solution was mixed with 10X T4 DNA ligase buffer (1X in reaction) and 10 units of T4 DNA ligase per pmol of mRNA. Finally, DEPC-water was added so that the Exo T-digested mRNA was 1.5  $\mu$ M in solution. The reaction was incubated for 1.5 hours at 20 °C before being subject to the modified RNeasy column purification. Cy5 analytical samples were run on an 1% agarose nucleic acid gel and analyzed by Cy5 fluorescence (ChemiDoc MP Imager, Bio-Rad).

#### *3.7.8. Modified PURExpress reactions*

mRNA display particles were prepared using in vitro expression with the PURExpress  $\Delta$  (aa, tRNA) kit (NEB) with several modifications. In this “split-and combine” approach, two pre-translation mixtures (“Pre-19” and “Pre-pMet”) were prepared to favor efficient incorporation of pMet; the 19 amino acids (minus Met) (Promega) and the pMet (Thermo Fisher Scientific) were split such that the aminoacylation (charging) of methionine tRNA (tRNA<sup>Met</sup>) by MRS5m favored pMet incorporation. The reaction was prepared with the following pipetting scheme (4X of a 20  $\mu$ L reaction is shown for reference, but reaction volumes may be adjusted):

Pre-19 (4X):

Reagent	Volume ( $\mu\text{L}$ )
$\Delta$ Solution A + tRNA*	5
$\Delta$ Solution B	5
RNasin Plus	0.66
19 amino acid mix (1 mM each)	5
Subtotal	15.66

Pre-pMet (4X):

Reagent	Volume ( $\mu\text{L}$ )
$\Delta$ Solution A + tRNA*	15
$\Delta$ Solution B	15
RNasin Plus	2
pMet (12 mM)	2
MRS5m (120 $\mu\text{M}$ )	3.4
DEPC-free water	4.6
Subtotal	42

\*  $\Delta$  Solution A and tRNA from the PURExpress  $\Delta$  kit are pre-mixed

Each tRNA aminoacylation reaction was incubated for 20 min at 37 °C, and the “Pre-19” mixture was added to “Pre-pMet” with an additional 15  $\mu\text{L}$  of 19 amino acid mix (for 4X, as shown above). Meanwhile, the puromycin-ligated mRNA was denatured by heating to 94 °C for 1 min, then cooling on ice for 10 min. To 18  $\mu\text{L}$  (for 1X) of the PURExpress mix, 2  $\mu\text{L}$  of denatured mRNA (corresponding to 4.5 - 6  $\mu\text{g}$  of mRNA) was added, and incubated for 30 min at 37 °C.

For C3-mediated cleavage of the initiator pMet, to 20  $\mu\text{L}$  of translation reaction, 2  $\mu\text{L}$  of 100 mM  $\text{Mg}(\text{OAc})_2$  (~9 mM final concentration) and 2 units of PreScission protease (Cytiva) were added for 1 hour at room temperature. To complete the protein-puromycin fusion reaction by stabilizing the ternary complex, an additional 2.1  $\mu\text{L}$  of 500 mM  $\text{Mg}(\text{OAc})_2$  (~50 mM final concentration) was added for an additional 1 hour at room temperature. Next, to remove insoluble material, the mixture was diluted 2-fold (to 50  $\mu\text{L}$ ) with 0.5X TBS (25 mM Tris-HCl pH 8, 75 mM NaCl), transferred to a 1.5 mL tube, and centrifuged for 10 min at 20,000  $\times g$  at 20 °C.

Finally, excess pMet removal and buffer exchange were performed using Micro Bio-Spin P-30 gel columns (Bio-Rad). The columns were prepared by following manufacturer's protocol. Then, 16  $\mu$ L of 5X TBS and 16  $\mu$ L of DEPC-treated water were added to the collection tube. Next, 40  $\mu$ L of the centrifuged fusion reaction was applied directly to the center of the column and centrifuged for 4 min at 1,000  $\times$  g at 20 °C. DEPC-treated water was added to column flow-through (to ~80  $\mu$ L total volume). The final solution was aliquoted to smaller volumes, flash-frozen in liquid nitrogen, and stored at -80 °C.

### *3.7.9. Affinity selection for nanobody library*

The following section describes conditions for affinity selections against MBP-biotin, sfGFP-biotin, hHER2 Fc Chimera Avi-tag (R&D Biosystems), Spike RBD-Avitag (Acro Biosystems), and hACE2-Avitag (Acro Biosystems) which were reconstituted in 1X DPBS, pH 7.4 except for hACE2, which was stored in 1X TBS pH 7.5 with 25% glycerol, and aliquots stored in -80 °C. Panning was performed with 4  $\mu$ L RNasin Plus (Promega), 0.2% w/v Tween-20 (Pierce), 1  $\mu$ M bovine serum albumin (BSA) (Millipore-Sigma), 50 nM target protein, 40  $\mu$ L of displayed library, and 1X TBS up to 400  $\mu$ L total volume for 20 min at RT in a 1.5 mL centrifuge tube (pre-blocked with 1 mL SynBlock for 20 min at room temperature, and washed 3x with 1 mL TBS prior to panning). The panning solution (200  $\mu$ L, split into 4x 50  $\mu$ L aliquots in 0.2 mL PCR tubes) was exposed for 15 sec (2x, with 2 min in between exposures) on a UV lamp (ELC-4001 UV Flood Curing System with ELC-2542 Power Supply), while the other 200  $\mu$ L was shielded from light at RT. Then, 200  $\mu$ L of UV-exposed / non-UV-exposed mix was incubated with 5  $\mu$ L of 10 mg/mL MyOne Streptavidin C1 beads (Thermo Fisher Scientific) (pre-washed 3x with TBS-T (TBS with 0.5% w/v Tween-20)) for 1 hour at RT on a tube rotator. Then, an additional 5  $\mu$ L of streptavidin beads was added and incubated for another 1 hour at RT on a tube rotator.

For all washing steps, beads were first centrifuged on a tabletop centrifuge (Thermo Fisher Scientific) at 10,000  $\times g$  for 20 sec 3x, rotating the tube 180° in between spins. Then, beads were captured with a handheld magnet (Amazon) at the bottom of the tube. With the magnet still at the bottom, the supernatant was vacuum aspirated to ~30  $\mu\text{L}$ . Next, the magnet was slowly dragged to the side of the tube above the liquid line (perpendicular to the side of the tube). Quickly, the rest of the liquid was aspirated, and the beads were resuspended in 500  $\mu\text{L}$  of the next wash liquid and mixed by pipetting and light vortexing.

For the first wash, the solution was transferred to a 1.5 mL tube with 500  $\mu\text{L}$  of TBS-T and resuspended by gentle vortexing for 5 sec. After aspiration as described above, the beads were washed 3x with 500  $\mu\text{L}$  TBST-T05 (TBS with 0.5% Tween-20) with 1 mM DTT. Next, the beads were washed 2x in 500  $\mu\text{L}$  TBS with 1.5% SDS (Fisher Scientific), 10 mM EDTA, and 1 mM DTT with 30 min incubation at 30 °C in between washes. Then, the beads were washed 2x in 500  $\mu\text{L}$  TBS with 6 M urea, 0.2% w/v Tween-20, and 1 mM DTT with 30 min incubation at 30 °C in between washes. Next, beads were washed 1x with 500  $\mu\text{L}$  DEPC-T05 (DEPC-water with 0.5% w/v Tween-20 and 1 mM DTT). Finally, the beads were resuspended in 30  $\mu\text{L}$  of DEPC-T05. Approximately 19  $\mu\text{L}$  were taken directly to RT, while the rest were flash frozen in liquid nitrogen and stored at -80 °C.

#### *3.7.10. RT-PCR and library reformatting*

The resuspended beads (19  $\mu\text{L}$ ) were mixed with 1  $\mu\text{L}$  of RT primer (50  $\mu\text{M}$ , Table 3-2) and incubated for 20 min at 65 °C for primer annealing. The reaction was cooled on the benchtop for 5 min. To the mixture, 3  $\mu\text{L}$  each of 10X AffinityScript RT buffer (Agilent), dNTPs (10 mM each) (NEB), 100 mM DTT (Agilent), and 1  $\mu\text{L}$  of AffinityScript RT (Agilent) were added and incubated for 60 min at 55 °C.

To 5  $\mu$ L of RT sample (beads should be resuspended), 10  $\mu$ L 5X HF Buffer (NEB), 1  $\mu$ L dNTPs (10 mM each), forward and reverse primers (500 nM final), 0.5  $\mu$ L Phusion Hot Start Flex DNA polymerase (NEB), and nuclease-free water (IDT) up to 50  $\mu$ L were added. Sample aliquots (5  $\mu$ L) were taken out at various cycles (noted in figure legends), mixed with 1  $\mu$ L 6X Orange DNA loading dye (Thermo Fisher Scientific), and analyzed by gel electrophoresis.

For reformatting of libraries to subsequent rounds, the library was first amplified with L9MX\_MRGS\_NdeI\_fo and L9MX\_PostRT\_inner\_re for 20-25 cycles and purified using QIAquick PCR Purification (Qiagen). Next, a low-cycle PCR (~10 cycles) was performed with L9MX\_MRGS\_NdeI\_fo and L9.MX.F4-SSGN\_Bpi\_re to introduce the 3'-Bpil site and purified as the previous reaction. Next, the 5' end of the nanobody library was digested with NcoI-HF (NEB) following manufacturer's protocol (with ~3-fold overdigest) and column purified. Then, the 3' end of the nanobody library was digested with Bpil (Thermo Fisher Scientific) following manufacturer's protocol (with ~3-fold overdigest) and column purified. The T7 module was synthesized by PCR with pPUR\_seq\_for and L9MX\_NcoI\_re (selected single clone plasmids were typically used as template) and column purified. The TolA-ExoT module was synthesized by PCR with TolA\_Bpi\_fo and TolA\_ExoT\_re (selected single clone plasmids were typically used as template) and column purified. Both T7 and TolA-ExoT modules were digested with NcoI-HF and Bpil, respectively. Three-piece linear ligation was performed (with ~1.2 - 3-fold excess of the modules to the library) for 2-4 hours at 20 °C. The desired product (T7-Nb library-TolA-ExoT) was gel purified and amplified by PCR (~5-8 cycles) using Lib\_T7pro\_hiTm\_fo and TolA\_ExoT\_re. Phusion Hot Start Flex (NEB) was used for all PCRs following manufacturer's protocol.

#### *3.7.11. Single clone binding assay with MBP*

Selected clones, subcloned and PCR amplified as SNAP fusions, were expressed in PURExpress following manufacturer's protocol, scaled to 5.5  $\mu\text{L}$  reaction volumes and expressed for 2 hours at 37 °C. BG-functionalized polystyrene beads were mixed with 4.5  $\mu\text{L}$  of the PURExpress reaction ( $\sim 1.3 \times 10^5$  beads in reaction) in TBS-T (TBS with 0.2% v/v Tween-20) in a 10  $\mu\text{L}$  reaction volume and incubated for 1 hour at 37 °C on a plate shaker at 950 rpm. The beads were washed 4x with 0.2 mL TBS-T, and resuspended in 83  $\mu\text{L}$  of TBS-T and stored at 4 °C.

Recombinantly expressed SNAP-MBP was labeled with BG-Alexa Fluor 647 (NEB) in a reaction with 5  $\mu\text{M}$  SNAP-MBP, 10  $\mu\text{M}$  BG-Alexa647, and TBS up to 200  $\mu\text{L}$ . The reaction was incubated at 37 °C for 1 - 3 hours and purified via group separation (described in Section 3.7.8). The binding reaction was performed by combining 5  $\mu\text{L}$  binder-immobilized beads, 20  $\mu\text{L}$  MBP-Alexa647, and 10  $\mu\text{L}$  TBS or competitor (unlabeled MBP) for 1 hour at 37 °C on a plate shaker at 950 rpm. The reaction was diluted with 0.25 mL ice-cold TBS immediately before running the measurement on an Accuri C6 flow cytometer (BD).

### 3.7.12. Control affinity selections with Off7<sub>D112M/WT</sub> and MBP

For control selections with Off7<sub>D112M/WT</sub> and MBP, panning volumes and conditions varied slightly from library selections. Panning was performed with 0.5  $\mu\text{L}$  RNasin Plus (Promega), 0.08% w/v Tween-20 (Pierce), 100 nM or 1  $\mu\text{M}$  target protein, 12  $\mu\text{L}$  of displayed library, and 1X TBS up to 34.6  $\mu\text{L}$  total volume for 40 min at RT in a 0.2 mL PCR tube centrifuge tube. An aliquot of the panning solution (17.3  $\mu\text{L}$ ) was UV-exposed as before, while the rest was shielded from light at RT. The mixture was pulled down as before, except with 2  $\mu\text{L}$  beads for the second incubation. All washing steps were kept the same as the library selections except the final resuspension, which was in 40  $\mu\text{L}$  DEPC-T05.

For RT, 12  $\mu$ L of the resuspended beads were mixed with 1  $\mu$ L of RT primer (50  $\mu$ M, Table 3-2) and incubated for 15 min at 65 °C for primer annealing. The reaction was cooled on the benchtop for 5 min. To the mixture, 2  $\mu$ L each of 10X AffinityScript RT buffer (Agilent), dNTPs (10 mM each) (NEB), 100 mM DTT (Agilent), and 1  $\mu$ L of AffinityScript RT (Agilent) were added and incubated for 60 min at 54 °C. The subsequent PCR setup was kept the same as the library selections, and primers Off7\_Bsal\_for and Off7\_Bsal\_re were used.

#### *3.7.13. Preparation of cell lysate for panning*

HEK293 cells (ATCC) were grown to 80% confluency in DMEM (Gibco) supplemented with 10% fetal bovine serum (Gibco) at 37 °C with 5% CO<sub>2</sub>. Cells were collected with trypsin-EDTA (0.25%) (Gibco) and washed twice with PBS (Gibco). Cells were transferred to a 2 mL tube, pelleted, and resuspended in DEPC-water with protease inhibitor. The tube was flash frozen in liquid nitrogen for 1 min, thawed to 25 °C, and rotated for 10 min at 4 °C. The lysed cell suspension was supplemented with Tris-HCl, pH 8 (50 mM) and NaCl (150 mM) and kept on ice.

The SPACE display protocol involving HEK293 cell lysate was performed as described in 3.6.11 except for the panning step, in which cell lysate was included in the biopanning step. Here, 6  $\mu$ L mRNA-displayed protein was mixed with 10  $\mu$ L cell lysate (5E4 cells) and 1  $\mu$ M biotinylated MBP-SNAP target protein. The reaction was incubated for 40 min at 20 °C and all subsequent steps were performed as described above.

#### *3.7.14. SDS-PAGE gel-shift crosslinking assay with Off7<sub>D112M</sub> and MBP*

All internal Met positions in Off7 and SNAP except for D112M (in Off7) were substituted (Off7: M35L, M109S and M114L; SNAP: M7L, M60V). Off7<sub>D112M</sub>-SNAP PCR product was translated using the “split-and-combine” PURExpress reaction for 2 hours at 37 °C. The translated protein was labeled with benzylguanine (BG)-Alexa647 (5  $\mu$ M)



(NEB) for 1 hour at 37 °C. To prevent labeling of the target protein, MBP-SNAP (15 µM) was blocked with SNAP-Surface Block (20 µM) (NEB) for 1 hour at 37 °C. For the binding reaction, the Off7<sub>D112M</sub>-Alexa647 was incubated with blocked MBP (10 µM) for 30 min at 20 °C, then placed on ice. Meanwhile, a 384-well plate (Greiner Bio-One) was blocked with 50 µL SynBlock (Bio-Rad) for 30 min at 20 °C, then washed with 50 µL TBS (50 mM Tris-HCl, pH 8, 150 mM NaCl). The binding reaction was split into two wells (15 µL each). One well was UV-exposed for 10 min using the Cytation DAPI channel (Intensity 10, z-height 3499 µm), while the other was covered with foil to prevent UV exposure. The UV lamp-exposed samples were exposed 2x for 15 sec or 30 sec with a 2-min break in between. Both exposed (+ UV) and unexposed (- UV) samples were diluted with LDS (1X, Thermo Fisher Scientific) and DTT (50 mM), and run on a 4-12% Bis-Tris SDS-PAGE (Thermo Fisher Scientific) gel with 1X MES buffer (Thermo Fisher Scientific). The gel was imaged on Alexa647 blot channel (ChemiDoc MP, Bio-Rad).

Table 3-1. Annotated nanobody library composition/diversity used for mRNA display.

Nanobody region name	Description	Amino acid sequence (N → C)
N'-terminal regulatory	MRGS-His6, GS linker, and C3 cleavage site that precedes nanobody library	<sup>1</sup> MRGSHHHHHHGGGSGGGSLVLFQGPS <sup>27</sup>
FW1	Nanobody FW1 (annotation restarted at 1 for simplicity)	<sup>1</sup> QVQLVESGGGLVQAGGSLRLSCAASG <sup>26</sup>
CDR1	Diversity encoded into CDR1 (vertical columns indicate possible amino acids); AA34 is originally a Met, but substituted to be L, I, or V	RTFSSYSL <sup>34</sup> S T AI N PV D V R G
FW2		GWFRQAPGKEREFVA <sup>49</sup>
CDR2	Diversity encoded into CDR2 (vertical columns indicate possible amino acids)	AISWSGGTTY <sup>58</sup> T TST AS D S R TI A V Q SN R G L
FW3		YADSVKGRFTISRDNKNTVYQLQNSLKPEDTAVYYCAA <sup>125</sup>
CDR3	Diversity encoded into CDR3 (vertical columns indicate possible amino acids); * region can be 5 to 9 amino acids (9 shown)	GA*****ATEYD <sup>136-141</sup> RP PAD N ST TNG T DS SD A T A N
FW4		YWGQGTQVTVS <sup>147-152</sup>

\*First iteration of the library had CDR3 diversity encoded with NNK, not as shown above

Table 3-2. List of primers used for mRNA display.

Primer name	Description	Sequence (5' → 3')
A15-Puro	Puromycin primer for ligation	/5phos/AAAAAAAAAA*A*A*A*A/iSp18/iSp18/ACC/3Puro/
A10-Cy5	Analytical primer for ligation	/5Phos/AAAAAAAAAA/3Cy5Sp/
ExoT-TolA-splint_12	Splint ssDNA for Puro ligation	TTTTTTTTTTGGTTTGAGTTAC
Off7_Bsal_for	Forward primer for RT-PCR of control Off7-MBP selections	CCAGGGGTCTCGCCATGGATCTGGGCAGAAAAC
Off7_Bsal_re	RT and reverse primer for RT-PCR of control Off7-MBP selections	GGCCAGGTCTCCGCCAAGCTTCTGCAGAATTTTC
L9MX_MRGS_Ndel_fo	Forward primer for RT-PCR for Nb selections	GGAGATATACATATGAGAGGATCGCACC
L9MX_PostRT_inner_re	Reverse primer for RT-PCR for Nb selections	CACTACTCACGGTTACTTGCGTACC
Nb_FW4_RT_16nt	Primer for RT step in Nb selections	CACTACTCACGGTTAC
Lib_T7pro_hiTm_fo	Forward primer for PCR of nanobody library (LNMx, LNM0)	CGAAAAGTGCTAGTGGTGCTAGCCCCG
L9.MX.F4-SSGN_Bpi_re	Reverse primer for PCR of nanobody library (LNMx, LNM0)	CCATGAAGACCCGCTATTACCACTACTCACGGTTACTTGCGTACCCTGG
TolA_Bpi_fo	Forward primer for PCR of TolA-ExoT module	CCATGAAGACGCTAGCGGTGGTGGTGGCGCTAAGGCTGTAGAA GAAGCAGCTA
TolA_R2_Bpi_fo	Optional forward primer for PCR of TolA-ExoT module	CCATGAAGACGCTAGCGGTGGTGGTGGCGCTAAG
TolA_ExoT_re	Reverse primer for PCR of TolA-ExoT module	TTTATATTATTTATTTTGGTTTGAGTTACTTCTACTGCAGCTGCTTCTACCG

TolA_ExoT_ultra_re	Reverse primer for PCR of full T7-Nb-TolA-ExoT library for mRNA display (to be used with Lib_T7pro_hiTmfo)	TTTTTTATATTATTATTTTGGTTTGGATTACTTCTACTGCAGCTGCTTCTACCG
pPUR_seq_for	Forward primer for PCR of T7 module	CGAAAAGTGCTAGTGGTGTC
L9MX_Ncol_re	Reverse primer for PCR of T7 module	GCCGCCATGGTGGTGATG

Table 3-3. List of amino acid sequences for proteins used in library and control selections.

Construct name	Description	Amino acid sequence
Off7 <sub>D112M</sub> -SNAP	Off7 mutant used for establishing photocrosslinking assays (*SNAP tag has internal Met residues substituted)	MRGSEVLFQGPDLGRKLLAARAGQDDEVRI LLANGADVNAADNTGTTPLHLAAYSGLHLEIVEVLLKHGADVNDASDVFGYTPHLAAYWGHLEIVEVLLKNGADVNASDSMGLTPLHLAAKWGYLEIVEVLLKHGADVNAQDKFGKTAFDISIDNGNEDLAEILQKLGSGGGSGGGDKDCELRRTTLDSPGKLELSGCEQGLHEIIFLGKGTSAADAVEVPAPAAVLGGPEPLVQATAWLNAYFHQPEAIEEFPPALHHPVFQQESFTRQVLWKLKLVKVFGEVISYSHLAALAGNPAATAAVKLTALSGNPVPI LIPCHR VVQGDLDVGGYEGGLAVKEWLLAHEGHRLGKPGLG
MRS5m	Engineered methionyl transferase for pMet aminoacylation	MRGSHHHHHMTQVAKKILVTCGSPYANGSIHLGHMLEHIQADVWVRYQMRGHEVNFICADDAHGTPIMLKAQQLGITPEQMIGEMSQEHQTDFAGFNISYDNYHSTHSEENRQLSELIYTRLKENGFIKNRTISQLYDPEKGMFLPDRFVKGTCPKCKSPDQYGDNCEVCGATYSPTELIEPKSVVSGATPVMRDESHFFFDLPSFSEMLQAWTRSGALQEQQVANKMQEWFESGLQQWDISRDAYFGEFIPNAPGKYFYVWLDAPIGFMGSFKNLCKDRGDSVSDFEYWKDSTAELYHFYIGKDVVYFLSLFWPAMLEGSNFRKPTNLFVHGYVTVNGAKMSKSRGTFIKASTWLNHFADSLRYYYTAKLSSRIDDI DLNLEDFVQRVNADIVNKVVNLASRNAGFINKRFDGVLASELADPQLYKFTDAAEVI GEAWESREFGKAVREIMALADLANRYVDEQAPWVAKQEGRDADLQAICSMGINLFRVLMTYLKPVLPKLTERAEAFNLNTELTWDGIQQPLLGHKVNPFKALYNRIDMKQVEALVEASKE
SNAP-MBP	Target for nanobody library and Off7 control selections	MRGSHHHHHMDKCEMKRTTLDSPGKLELSGCEQGLHEIIFLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPPALHHPVFQQESFTRQVLWKLKLVKVFGEVISYSHLAALAGNPAATAAVKLTALSGNPVPI LIPCHR VVQGDLDVGGYEGGLAVKEWLLAHEGHRLGKPGLGSGGGSGGGSGSMGKTEEGKLVIIWINGDKYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDII FWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDVAVRYNGKLIAYPIAVEALS LIYNKDLLP NPKTWEEI PALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKDYIKDVGVNAGAKAGLTFVLVDLIK NKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS YEEEE LAKDPRI AATMENA QKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

sfGFP-SNAP	Target for nanobody library and Off7 control selections	MRGSHHHHHHLEVLFGPGSMSKGEELFTGVVPIIIVELDGD VNGHKFSVRGEGEGDATNGKLTLLKFICTTGKLPVPWPPTLVT TLTYGVQCFSRYPDHMKRHDFFKSAPEGYVQERTISFKDD GTYKTRAEVKFEGDTLVNRIELKIDFKEDGNI LGHKLEYN FNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQN TPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMLLEFVTA AGITHGMDELYKLLGGGGSGGGSGGGSGGGSGGGSGMDKCEM KRITLDSPLGKLELSGCEQGLHEIIFLGKGTSAADAVEVPA PAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVF QQESFTRQVLWKLKLVVVFGEVVISYSHLAALAGNPAATAAV KTALSGNFPVILI PCHRVVQGDLDVGGYEGGLAVKEWLLAH EGHRLGKPGLG
------------	---	---

Table 3-4. List of DNA sequences for Off7 control selections.

Construct name	Description	DNA sequence
Off7 <sub>D112M</sub> -TolA-ExoT	Off7 mutant used for control selections with MBP	ATGAGAGGATCGCTGGAAGTTCTGTTCCAGGGACCTGATCT GGGCAGAAAACCTTCTGGAAGCGGCTCGCGCCGGTCAGGATG ATGAAGTGCGCATTCTTTTGGCGAACGGCGCTGACGTTAAT GCCGCTGACAATACTGGTACTACCCCGCTTACCTGGCAGC GTATAGTGGCCACCTGGAGATCGTTGAAGTGCTGCTTAAAC ACGGTGCCGATGTTGACGCATCTGACGTTTTTGGCTATACG CCGCTGCATCTGGCTGCATACTGGGGTCATTTGGAAATTGT TGAGGTTCTGCTGAAGAACGGTGCAGACGTGAACCGCTCGG ATTCTATGGGCTTGACTCCACTGCACCTTGCGGGTAAGTGG GGTACCTTGAAATCGTGGAAGTGTACTGAAACATGGCGC GGATGTTAATGCTCAGGATAAATTCGGCAAACCGCGTTCCG ATATCTCTATTGACAACGGCAACGAAGATCTGGCCGAAATT CTGCAGAAGCTTAGTGGAATAGCGGTGGTGGTGGCGCTAA GGCTGTAGAAGAAGCAGCTAAGAAAGCGGCTGTAGACGCTA AGAAAAAGCTGAGGTAGAAGCCGCTAAGGCCGAGTAGAA GCGCAGAAAAAGTAGAGGCAGCCGCTGCGGCAGTGAAGAA GAAAGCGGAAGCGGTAGAAGCAGCTGCAGTAGAAGTAATC AAACCAAAATAAATAATATAAACTCGAG
Off7 <sub>WT</sub> -TolA-ExoT	Off7 WT used for control selections (negative control) with MBP	ATGAGAGGATCGCTGGAAGTTCTGTTCCAGGGACCTGATCT GGGCAGAAAACCTTCTGGAAGCGGCTCGCGCCGGTCAGGATG ATGAAGTGCGCATTCTTTTGGCGAACGGCGCTGACGTTAAT GCCGCTGACAATACTGGTACTACCCCGCTTACCTGGCAGC GTATAGTGGCCACCTGGAGATCGTTGAAGTGCTGCTTAAAC ACGGTGCCGATGTTGACGCATCTGACGTTTTTGGCTATACG CCGCTGCATCTGGCTGCATACTGGGGTCATTTGGAAATTGT TGAGGTTCTGCTGAAGAACGGTGCAGACGTGAACCGCTCGG ATTCTGATGGCTTGACTCCACTGCACCTTGCGGGTAAGTGG GGTACCTTGAAATCGTGGAAGTGTACTGAAACATGGCGC GGATGTTAATGCTCAGGATAAATTCGGCAAACCGCGTTCCG ATATCTCTATTGACAACGGCAACGAAGATCTGGCCGAAATT CTGCAGAAGCTTAGATCCAGTGGTAATAGCGGTGGTGGTGG CGCTAAGGCTGTAGAAGAAGCAGCTAAGAAAGCGGCTGTAG ACGCTAAGAAAAAGCTGAGGTAGAAGCCGCTAAGGCCGCA GTAGAAGCGCAGAAAAAGTAGAGGCAGCCGCTGCGGCAGT GAAGAAGAAAGCGGAAGCGGTAGAAGCAGCTGCAGTAGAAG TAAC TCAAACAAAATAAATAATATAAACTCGAG

## Chapter 4: Multivalent in vitro display and selection (MIDAS)

---

Adapted with permission from “Ohoka, A. and Sarkar, C.A. Facile display of homomultivalent proteins for *in vitro* selections. *ACS Synth. Biol.* **2023**, Article ASAP. DOI: 10.1021/acssynbio.2c00563.” Copyright 2023 American Chemical Society.

### 4.1. Summary

Low-affinity protein binders are emerging as valuable domains for therapeutic applications due to their higher specificity when presented in multivalent ligands that increase the overall strength and selectivity of receptor binding. *De novo* discovery of low-affinity binders would be enhanced by the large library sizes attainable with *in vitro* selection systems, but these platforms generally maximize recovery of high-affinity monovalent binders. In existing multivalent selection systems like cell-based phage display<sup>163</sup> or *in vitro* DNA display<sup>26</sup>, the valency and arrangement of the displayed library members cannot be simultaneously and easily controlled. Thus, there is a need for a precise, easy, and high-throughput multivalent selection method for low-affinity binders.

In this study, we present a facile technology that uses rolling circle amplification (RCA) to create homomultivalent libraries. We show proof of principle of this approach in ribosome display with off-rate selections of a bivalent ligand against monovalent and bivalent targets, demonstrating high enrichment (up to 166-fold) against a low-affinity target that is bivalent but not monovalent. This approach to homomultivalent library construction can be applied to any binder tolerant of N- and C-terminal fusions and provides a platform for performing *in vitro* display selections with controlled protein valency and orientation. We also employed RCA to create mock bivalent DNA display particles to perform biopanning experiments and demonstrate improved recovery of a weak binder in a bivalent presentation.

## 4.2. Introduction

*In vitro* display systems enable high-throughput selection of large protein libraries, up to  $10^{14}$  molecules, in protein engineering campaigns.<sup>164</sup> However, current affinity and off-rate selections using these methods generally only allow for the strongest binders – primarily dictated by having the slowest dissociation – to be recovered and further evolved, with specificity to the target characterized subsequently.<sup>165</sup> Thus, recovery of low-affinity binders is difficult, despite the fact that these proteins may be of great interest in creating homo- or hetero-multivalent constructs to achieve highly specific and/or selective binding with low off-target effects given their low monovalent affinities.<sup>166–168</sup> For example, chimeric antigen receptor T-cells (CAR T-cells) with bivalent, low affinity antigens have shown higher specificity and AND-gate function over their monovalent, high affinity counterpart.<sup>78</sup> Moreover, low affinity binders can be useful for achieving better binding profiles; for instance, *in vivo* studies have shown that high affinity monoclonal antibodies (mAbs) suffer from localization and low tumor penetration in cancer applications.<sup>169</sup> Decreasing the affinity and potency for high expression targets may allow for higher dosing, which can lead to increased homogenous distribution and reduced toxicity by lowering target-mediated uptake in healthy tissue. Beyond achieving the desired affinity to the target antigen, there are additional barriers to biologics development, such as intrinsic immunogenicity, self-association, and high viscosity.<sup>170</sup> Therefore, retaining sequence diversity (i.e., both low- and high-affinity binders) may be important for developability downstream of an affinity-based selection.

In existing multivalent selection systems, the valency and arrangement of the displayed library members cannot be simultaneously and easily controlled. Increased valency achieved through cell- and virus-based methods are simple to implement but lack the control over valency and spatial arrangement. For example, phage display has three

distinct display strategies (i.e., fusion to pIII, pVIII, and pIX) that allow for some phage-enabled focused ligand arrangement and can be tuned to “low valency” or “high valency” with the addition of helper phages<sup>171</sup>, but these still do not allow the user to prescribe the exact valency and spatial arrangement of the displayed library.<sup>172</sup> These *in vivo* methods also decrease the screening throughput by orders of magnitude due to inefficiencies in transformation. While current *in vitro* multivalent methods like polysome display maintain large library sizes, they still lack precise binder presentation, which limits their utility since matching the interdomain topology of a binder to that of its target can optimize avidity enhancement.<sup>168,173</sup> Controlled spatial arrangement of multivalent binders has demonstrated enhanced receptor engagement and super-selectivity for an array of therapeutic applications including cell signaling<sup>167</sup> and viral sensing.<sup>168</sup> Finally, approaches that allow user-defined spatial presentation like SNAP-dendrimer display rely on technically challenging techniques such as *in vitro* compartmentalization.<sup>174,175</sup> Thus, there is a need for a precise, easy, and high-throughput multivalent selection method for low-affinity binders.

Rolling circle amplification (RCA) is an isothermal method for amplifying circular DNA or RNA molecules; it utilizes cyclized DNA or RNA (e.g., splinted ssDNA or circularized dsDNA) as a template and a polymerase with high processivity and strand displacement such as phi29 to amplify DNA or RNA from annealed primer(s) to create long ssDNA templates with repeated units complementary to the template.<sup>176</sup> RCA is often used to amplify genetic material for diagnostic applications although there are studies that have used RCA to create concatemeric DNA aptamer libraries for nanoparticle selection and long hairpin RNA libraries from DNA for gene silencing in plants.<sup>176–178</sup> As an enabling technology for protein display applications, RCA can allow the user to not only multivalently encode a protein binder but also control the valency and spatial presentation



to match that of their desired target for optimal multivalent engagement.

In our study, we developed an experimental method for creating multivalent binder libraries for *in vitro* display systems using RCA. We first applied this technology to create mock DNA display particles conjugated to monovalent or bivalent designed ankyrin repeat protein (DARPin) binders against human epidermal growth factor receptor 2 (HER2)<sup>179</sup>, a cell surface receptor, and performed biopanning experiments directly on a live cell surface. Building on the DNA-biopanning protocol developed in Chapter 2, we first adapted the particle synthesis and purification methods for a DNA template conjugated to a protein binder, and also modified to a suspension cell approach to allow for increased target-ligand interaction. With similar motivations as discussed earlier in Chapters 2 and 3, we initially applied our multivalent *in vitro* display approach toward panning experiments directly on the surface of live cells. As cell-surface interactions are typically low affinity ( $\mu\text{M}$  range)<sup>180</sup> and membrane proteins often organize into oligomeric architectures as driven by the lipid composition in the membrane<sup>181–184</sup>, we reasoned that our multivalent approach would uniquely suit the challenges and opportunities reflective of the receptor presentation within the cell membrane.

Due to limitations in our capability to create water-in-oil emulsions necessary for performing DNA display selections, however, we then adapted the experimental method to ribosome display (RD) selections and showed increased recovery of bivalently displayed RD particles in control selections with soluble targets and made efforts toward library selections. In these later studies, to motivate our multivalent designs, we applied MVsim, a graphical user interface-based simulator for predicting multivalent binding kinetics based on a comprehensive mechanistic model that was previously developed in the lab.<sup>62,63</sup> Briefly, the model treats receptor-ligand interactions as a dynamic network of microstate configurations controlled by affinity, linkage, and valency. The affinity of the

interaction is determined by monovalent rate constants that describe association ( $k_{on}$ ) and dissociation ( $k_{off}$ ) of receptor-ligand binding. Valency dictates the total possible microstates of the receptor-ligand interaction to calculate effective concentrations of all the interactions that may be sampled in a multivalent network to give the time-dependent evolution of each microstate configuration during association and dissociation phases. Lastly, the linkage, or the topological constraints of the interdomain linker, is used in a probability density function (PDF) to approximate the receptors and ligands as rods with a given distance between interfacial points and calculate the effective concentrations for the various microstates. The output of the model shows association and dissociation kinetics of the modeled multivalent interaction as simulated surface plasmon resonance (SPR) sensorgrams, which provide both quantitative and mechanistic characterizations with distinct behaviors between ligands with different valency.

Beyond our applications, this novel RCA-enabled multivalent library is also modular in that it may be applied to any binder library (i.e., other scaffold proteins) for other *in vitro* display methods such as mRNA or DNA display, which greatly expands the repertoire of potential binders recovered in these selection systems (Figure 4-1).

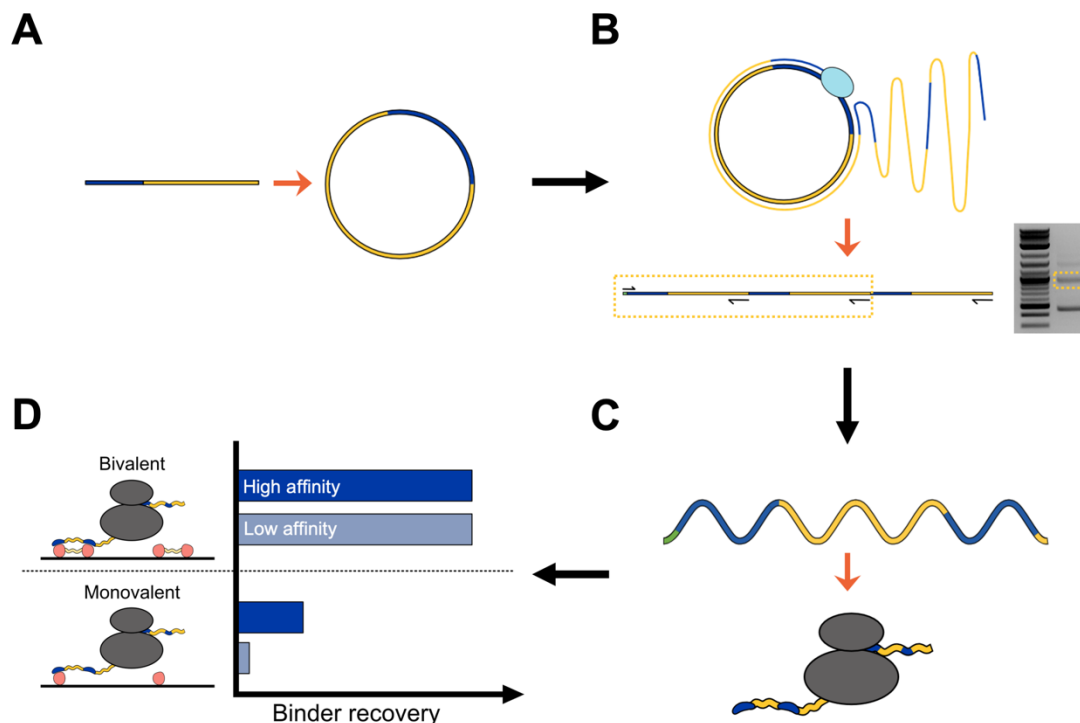


Figure 4-1. Workflow for preparing a multivalent *in vitro* display particle.

(A) DNA template encoding a binder library is fused to a gene corresponding to the desired interdomain linker and is circularized using a ligase. The circular template is (B) amplified using rolling circle amplification to create concatemeric repeats of a binder-linker, and PCR amplified to create templates with varying valency. (C) The user can select the desired valency and reformat to create multivalent *in vitro* display particles to (D) achieve improved binder recovery for weak binders.

### 4.3. Results

#### 4.3.1. Strategies for rolling circle amplification of dsDNA template and mock DNA display particle synthesis and purification

For our initial demonstration of RCA-enabled multivalency, we used G3, a DARPin that binds HER2<sup>19</sup>, fused N-terminally to a flexible (G<sub>4</sub>S)<sub>4</sub> linker to amplify via PCR with outer primers containing Type IIP restriction sites. We used both the wild-type (WT) ( $k_{\text{off}} = 1 \times 10^{-4} \text{ s}^{-1}$ ,  $K_{\text{D}} = 0.091 \text{ nM}$ ) and the AVD mutant ( $k_{\text{off}} = 4.4 \times 10^{-3} \text{ s}^{-1}$ ,  $K_{\text{D}} = 10.2 \text{ nM}$ ) G3 DARPin binders to compare variants with different affinities and dissociation kinetics.<sup>179</sup>

For self-circularization of the double-digested PCR product, we considered the probability of intramolecular base matching to maximize the efficiency of self-ligation. We accounted for the rigid nature of DNA and ensured that the total length of the digested dsDNA template was a multiple of the helical repeat of  $\sim 10.5$  bp per turn such that the DNA ends matched the bending fluctuations of the helix.<sup>185</sup> In addition, the concentrations of the T4 DNA ligase and the dsDNA were kept low ( $1.5$  ng/ $\mu$ L of reaction) to prevent intermolecular ligation and encourage intramolecular ligation. Further, we evaluated the free energy ( $\Delta G$ ) of ligation for different sticky ends created by common Type IIP enzymes using NP-Sticky, a web-based linear ligation calculator that was developed previously in the lab.<sup>186</sup> To encourage correct sticky-end ligation, we aimed to both decrease (i.e., make more favorable) the free energy ( $\Delta G$ ) of sticky-end ligations and increase  $\Delta G$  of sticky-end mismatches and nucleotide gaps for both the upper and lower strands as calculated using NP-Sticky. Notably, we tested the restriction enzymes MfeI and BamHI (with sticky end  $\Delta G$  of  $-9.78$  and  $-11.98$  kcal/mol, respectively) and found that while self-circularization occurred for MfeI-digested templates, the BamHI-digested templates only formed dimeric linear ligation products (Figure 4-2). This suggests that contrary to linear ligation in which a lower  $\Delta G$  of sticky ends may be preferred for higher efficiency of ligation, a higher  $\Delta G$  of sticky ends may allow for better intramolecular ligation.

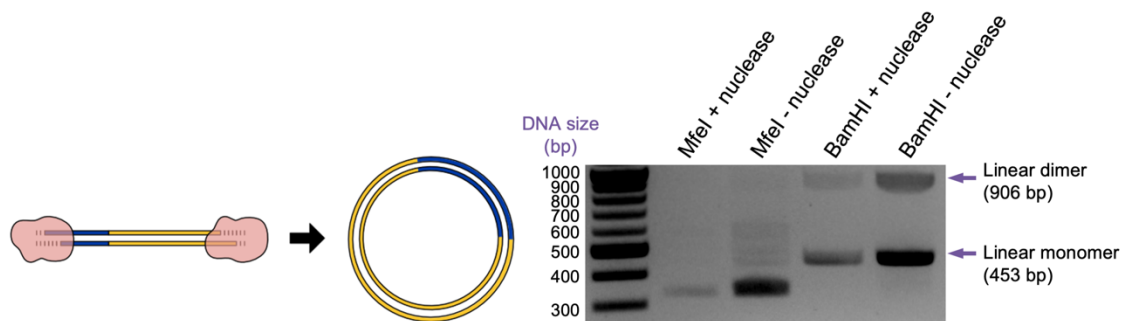


Figure 4-2. Circularization of dsDNA.

Agarose gel image of circularized dsDNA product, created by digesting both 5' and 3' ends with restriction enzymes and performing a ligation reaction with T4 DNA ligase. Ligation reaction of BamHI-digested product shows only linear monomer and dimer products, which is

in contrast to the ligation reaction of MfeI-digested product. Unexpectedly, the circularized product migrates faster than the linear product, which may be due to some secondary structure. (Samples with “+ nuclease” indicate treatment with a linear DNA nuclease which was not used for the rest of the experiments, but the whole gel is shown for singular image with ladder.)

Following circularization, templates were purified and rebuffed with standard nucleic acid purification columns. The input templates undergo a primer annealing step prior to isothermal amplification using a polymerase such as phi29 to create concatemeric ssDNA. After RCA, the ssDNA product is put into a standard PCR to recreate the dsDNA template, resulting in a mixed product with templates of different valency. The product with the desired valency can be easily isolated by gel extraction to be reformatted for different applications such as vector subcloning or RD. Notably, further amplification of the gel-extracted bivalent template using outer primers resulted in only monovalent product, most likely due to amplification bias of shorter templates.

For preparation of DNA display particles, the RCA-PCR amplified product was subcloned into an expression vector as an N-terminal fusion to SNAP, an affinity tag that covalently binds to benzylguanine (BG), a synthetic derivative of guanine, through a thiol-bond (Figure 4-3A).<sup>131</sup> While *in vitro* compartmentalization (IVC) of the DNA template and *in vitro* transcription translation (IVTT) reaction mixture would be performed for a library selection, we used recombinantly expressed proteins that were subsequently conjugated to DNA templates for the control experiments. To this end, proteins were expressed and purified as described previously (Section 3.7.1), and the n5T template (Section 2.3.2) was used for all binders instead of the gene-encoding template to mitigate any amplification bias for qPCR analysis. A BG molecule was also introduced to the 5'-terminus to covalently conjugate the protein to the n5T DNA template. To accomplish this, we utilized amine-reactive crosslinking with a 5'-amino-modified primer and a bifunctional BG-N-hydroxysuccinimide (BG-GLA-NHS) molecule (Figure 4-3B). Understanding the

limitations with amine-NHS reactivity as noted in Chapter 2, the reaction was repeated twice for higher yield and analytically monitored using NHS-Alexa555, which showed increased product yield after two couplings. The modified 5'-BG primer was incorporated to the n5T template via standard PCR, and the purified PCR product was coupled to the protein, G3<sub>WT / AVD</sub> (monovalent or bivalent)-SNAP, by incubating them together at 37 °C (Figure 4-3C).

After protein conjugation, column purification was not used due to the presence of guanidine hydrochloride (GdnHCl), a strong denaturant<sup>187</sup>, in the DNA purification buffer. Thus, we again applied the purification protocol developed in Chapter 2 (section 2.3.2) using gel electrophoresis to separate reacted and unreacted products, followed by agarase digestion, column separation, and rebuffing using diafiltration. Initial attempts at agarase digestion with the 50 °C incubation led to aggregation of the SNAP tags, so the incubation temperature was lowered to 42 °C. Finally, since quantification using absorbance measurements at 260 nm was not possible due to presence of the conjugated protein, the template was quantified using gel imaging with a reference DNA template. Notable contamination of unconjugated template was visible for some preparations, but since DNA does not significantly adhere to cell surfaces (as observed in Chapter 2), we reasoned that these templates should only be recovered minimally after panning experiments (Figure 4-3D).

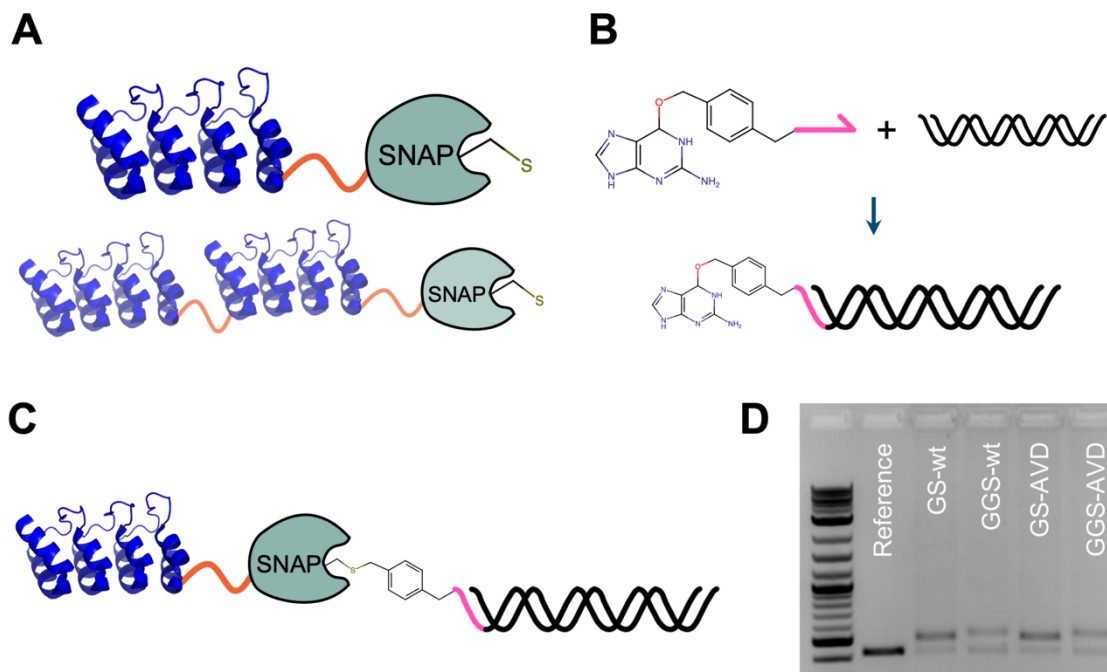


Figure 4-3. Summary of steps for bioconjugation of protein to control n5T DNA template. (A) Monovalent or bivalent G3 binders (PDB: 2JAB)<sup>179</sup> are subcloned as SNAP fusions and recombinantly expressed and purified. (B) 5'-terminus of DNA template is modified with a BG-primer that is incorporated through PCR. (C) Co-incubation of products in (A) and (B) leads to covalent thiol-linkage between BG (on the DNA) and SNAP (on the protein). (D) Gel analysis of purified products G3<sup>WT</sup> monovalent-SNAP (GS-wt), G3-G3<sup>WT</sup> bivalent-SNAP (GGS-wt), G3<sup>AVD</sup> monovalent-SNAP (GS-AVD), and G3-G3<sup>AVD</sup> bivalent-SNAP (GGS-AVD), shows notable contamination for some templates (~30 - 40% for the bivalent G3 constructs, calculated by ImageJ), but the unmodified DNA templates should not be recovered post-panning.

#### 4.3.2. Biopanning and analysis of binder-conjugated DNA templates

Building on the adherent cell panning strategies developed in Chapter 2, we adapted the protocol to perform panning with cells in suspension to achieve a higher target concentration by not only having more cells in solution but also having the whole cell surface accessible for binding. For the cell panning experiments, we used the epithelial ovarian cancer cell line, SK-OV-3, which overexpresses HER2 (~350,000 receptors per cell<sup>188</sup>). The cells, which are grown adherently, were collected by mild trypsinization, washed, and resuspended in DPBS with 10% BSA to reduce nonspecific binding.

Approximately  $2.8 \times 10^{10}$  display particles were mixed with the cell suspension containing 200,000 cells and panned for 75 min at 16 °C with shaking to prevent cells from settling. Competed samples included 200x molar excess of the respective protein during panning. To reduce the number of free particles in solution, washing steps were performed by a series of light centrifugation and buffer resuspension steps performed at a combination of RT and 4 °C temperatures; since most steps were conducted quickly (~5 min at RT, with the rest at 4 °C), possible internalization at the elevated temperatures was generally not considered. After the last wash, the remaining cell suspension was counted and placed directly into qPCR. The number of particles per cell was calculated as described previously in Chapter 2; unlike the previous study, however, the particles were recovered via ligand-receptor (i.e., G3-HER2) interactions of the G3-conjugated DNA template instead of nonspecific adsorption.

Analysis of the qPCR data showed significant differences in recovery of DNA templates with monovalent G3<sub>WT</sub> and G3<sub>AVD</sub>, the faster dissociating mutant, as expected (Figure 4-4). Comparing the monovalent G3<sub>AVD</sub> and bivalent G3<sub>AVD</sub> constructs, the 32-fold difference strongly supports our hypothesis that bivalency in selection particles can rescue weak binders in panning experiments. Interestingly, the reduced particle recovery for the competed bivalent G3<sub>WT</sub> sample over the competed monovalent G3<sub>WT</sub> sample suggests that the free bivalent G3<sub>WT</sub> binds so strongly to the dimeric HER2 receptor such that it displaces and/or does not allow binding of the bulkier DNA-conjugated bivalent G3<sub>WT</sub>. Overall, the control samples competed with excess free protein also suggest specific binding.



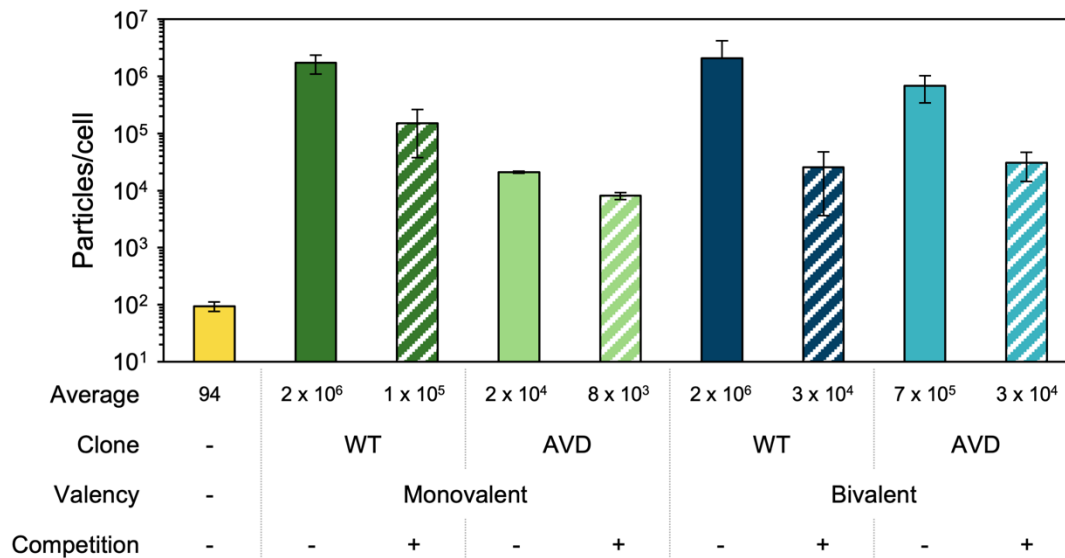


Figure 4-4. Comparison of particle recovery per cell for G3-conjugated DNA templates. G3-coupled DNA show significantly higher binding to the HER2-overexpressing SK-OV-3 cells over unconjugated DNA. G3<sub>WT</sub> monovalent and bivalent binding are similar, while the weaker, faster dissociating G3<sub>AVD</sub> benefits from bivalent presentation. Lower background (i.e., competed samples) of bivalent G3<sub>WT</sub> demonstrate increased binding strength and lack of dissociation compared to monovalent G3<sub>WT</sub>, while higher background of bivalent G3<sub>AVD</sub> suggest potentially higher stickiness of the G3<sub>AVD</sub> bivalent construct.

#### 4.3.3. Modeling of bivalent binding kinetics using MVsim

As mentioned previously, despite the positive results from the mock DNA display-biopanning experiments, we were unable to create water-in-oil emulsions necessary for performing library selections with DNA display, and thus adapted the experimental approach to ribosome display (RD). Prior to performing library selections, we applied MVsim toward predictions of multivalent binding to simulate and motivate our protein designs (Figure 4-5A and 4-5B). Even with differences in the panning format, by modeling the dissociation kinetics of two ligand-target pairs with significantly different monovalent dissociation rate constants, we can gain insights into expected differences in the dissociation timescales between monovalent and bivalent presentation formats.

For library selections and proof-of-concept control experiments, we used a model

interaction between a shortened peptide motif from the SLP-76 scaffold protein and two cognate targets, the SH3 C-terminal domains (SH3C) of the adapter proteins Grb2 and Gads. Despite binding to the same ligand, Grb2 and Gads have markedly different dissociation kinetics from SLP-76 ( $k_{\text{off}} = 68.5 \text{ s}^{-1}$  and  $0.5 \text{ s}^{-1}$ , respectively),<sup>189</sup> making these ideal targets to understand how multivalent presentation of the SLP-76 ligand affects particle recovery in an *in vitro* selection system. To promote separation between each ligand-target binding interaction, we introduced a mixed-species interdomain linker in which flexible regions flanked a rigid domain (Table 4-2).

For Grb2, MVsim predicts the near-complete loss of monovalent signal on the timescale of seconds due to the large  $k_{\text{off}}$ , whereas the bivalent presentation is significantly retained (Figure 4-5A). In contrast, the monovalent Gads-SLP76 interaction is not fully dissociated for minutes (Figure 4-5B), with the corresponding bivalent construct effectively undissociated over the course of the entire simulation (3600 s). From these simulations, we sought to test various wash conditions in RD to quantify particle recovery at different timepoints. While the qualitative trends were expected to hold in experiments, differences in the experimental conditions compared to the MVsim simulations, especially the lower temperature required to maintain selection particle stability, likely decrease dissociation rate constants compared to the room-temperature values that were available in the literature to use in MVsim, so we chose to test commensurately longer washing times.

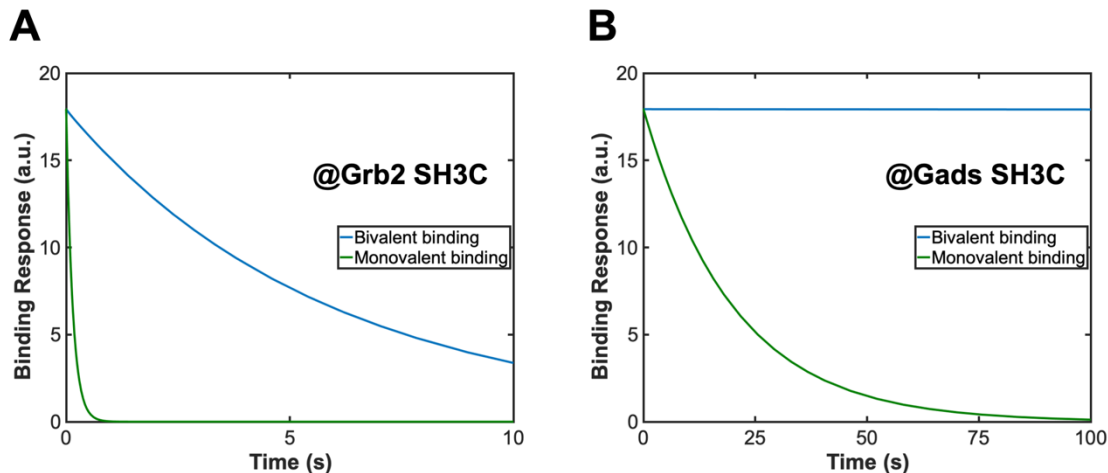


Figure 4-5. Dissociation kinetics from MVsim simulations.

Monovalent-monovalent or bivalent-bivalent ligand-target binding of shortened SLP-76 11-mer peptide ( $A_{232}PSIDRSTKPA$ ) to (A) Grb2 SH3C or (B) Gads SH3C were modeled, with kinetic binding parameters taken from Seet et al.<sup>189</sup> The  $k_{on}$  value for both is estimated at  $8.5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ , although the  $k_{off}$  values are very different at  $0.5 \text{ s}^{-1}$  for Gads SH3C-SLP76 ( $K_D = 59 \text{ nM}$ ) and  $68.85 \text{ s}^{-1}$  for Grb2 SH3C-SLP76 ( $K_D = 8.1 \text{ }\mu\text{M}$ ). For bivalent binding, Gads dissociation is much slower than Grb2, and the signal is nearly fully retained for the timescale shown. Despite its more appreciable dissociation, the simulations suggest that it is also possible to recover a monovalent binder with micromolar affinity (i.e., SLP76 to Grb2) through bivalent representation of the ligand and target.

#### 4.3.4. Initial proof-of-principle panning experiments with SLP76 against Grb2 and Gads

We sought to validate our homomultivalent designs in proof-of-principle RD experiments using the SLP-76 peptide against the Grb2 and Gads targets. In control experiments, the ligand (panned as RD particles) was fixed to be bivalent-SLP76 with mixed-species interdomain linkers to minimize signal differences that could potentially arise from varying efficiencies during RD particle synthesis and reverse transcription (RT) steps. Bivalent-SLP76 was translated and panned against immobilized monovalent or bivalent SH3C Gads or Grb2 domains. Following *in situ* RT, the cDNA was recovered and quantified using qPCR.

For panning against the faster-dissociating Grb2 SH3C domains (monovalent  $K_D = 8 \text{ }\mu\text{M}$ ,  $k_{off} = 68.5 \text{ s}^{-1}$ ), we observed a 9-fold difference on average in the number of

recovered mRNA (i.e., particles) (Figure 4-6A).<sup>37</sup> In the control selection against Gads SH3C, we tested an array of off-rate selection washes and saw much higher enrichment (166-fold on average) with varying degrees per condition tested (Figure 4-6B). The relatively consistent signal for the bivalent panning condition is also in good agreement with the MVsim simulations, in which the bivalent-Gads binding signal was retained throughout the duration of the simulation. In additional control experiments with Gads in which we pre-incubated the translated bivalent-SLP76 RD complexes with monovalent or bivalent free Gads prior to panning and a 1-hour competitive wash, our results suggest not only a specific interaction but also the slower dissociation behavior of the bivalent-bivalent interaction (Figure 4-6C).

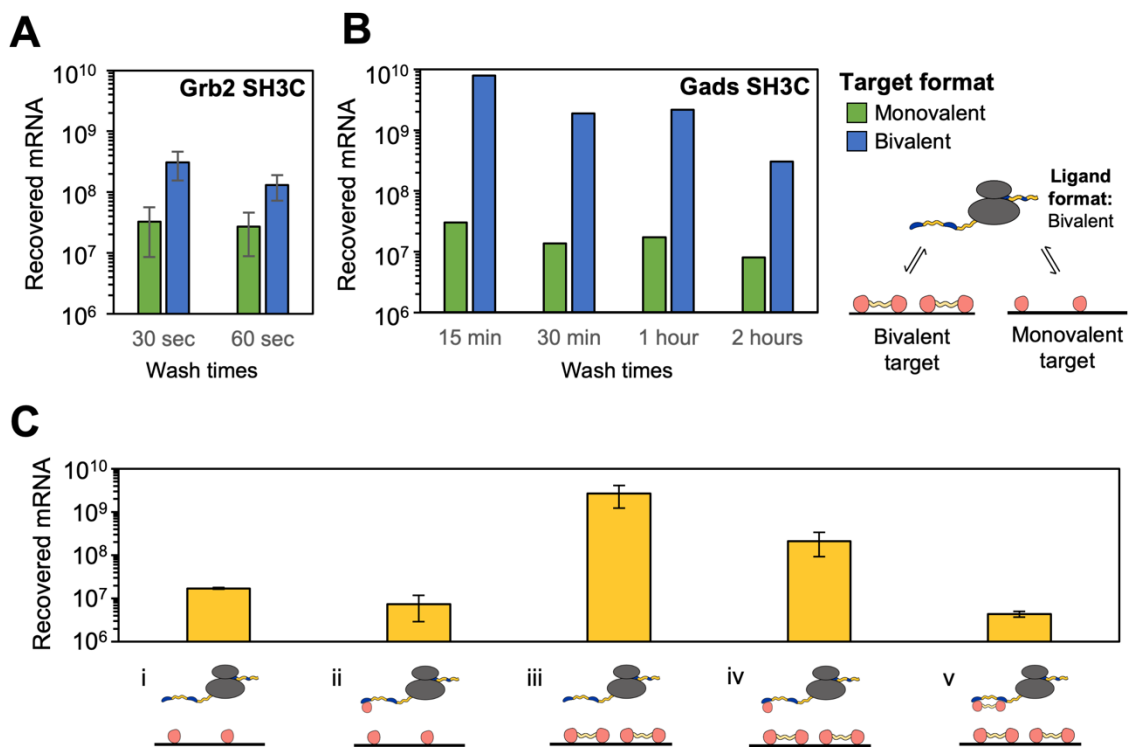


Figure 4-6. RD-qPCR data from control selections against monovalent or bivalent target. (A) Comparison of recovered mRNA molecules from panning bivalent SLP76 RD particles against monovalent or bivalent Grb2 SH3C with 30- or 60-sec washes in a mock off-rate selection. Total signal is relatively low due to faster dissociation kinetics; however, ~9-fold enrichment is possible (60-sec wash) even for the micromolar interaction. (B) Same experiment as (A), but against monovalent or bivalent Gads SH3C. Approx. ~166-fold

enrichment is achieved for the Gads SH3C interaction with a 60-min wash condition. Background signal for panning experiments is  $\sim 4 \times 10^6$  mRNA molecules (from Fig. 4-6C(v)). (C) Comparison of qPCR signals as seen in control conditions against Gads SH3C domains with 60-min wash steps. Monovalent targets were immobilized in (i) and (ii), and bivalent targets in (iii-v). RD particles were subject to pre-panning in solution with no target in (i) and (iii), monovalent target in (ii) and (iv), or bivalent target in (v). Comparing (i) and (iii), the bivalent target presentation increases recovery by 156-fold and comparing samples (ii) and (v) to (iv) shows differences in the ability of monovalent and bivalent target to compete ligand-target engagement.

#### 4.3.5. Modified RCA-enabled multivalent library synthesis and efforts toward affinity selections against monovalent or bivalent target

For library selections, we began with the shortened 9-amino acid SLP-76 peptide motif used in the previous study, PSIDRSTKP, and randomized the inner 7 amino acid positions using NNK codons to create a library (PXXXXXXXXP, theoretical size of  $3.4 \times 10^{10}$ ).<sup>189</sup> Our approach was to perform panning selections with the library against monovalent or bivalent targets (Gads or Grb2 SH3C domains) to quantitatively track enrichment of the specific binding motif, PXXXXRXXKP, over several selection rounds and using next generation sequencing (NGS) of the recovered pools.<sup>189</sup> Again, we used RCA to encode the library spaced by the mixed-species interdomain linkers. The flexible region of the linker between the soluble targets were kept slightly shorter than that of the library to allow for sufficient spatial sampling by the library. Targets were cloned using linear digestion-ligations and recombinantly expressed and purified.

In contrast to the dsDNA circularization performed previously, we used a ssDNA input as the template since our library was short enough to be encoded by essentially a long primer (i.e., Ultramer), and the ssDNA circularization reaction enabled higher input molecules, which is ideal for an *in vitro* library selection. For this approach, we utilized CircLigase, a commercially available enzyme that circularizes 5'-phosphorylated ssDNA templates. We achieved  $\sim 90\%$  ligation with a high DNA input of  $5 \times 10^{12}$  molecules per

reaction, which is important when considering a large, diverse library (Figure 4-7A). Next, RCA, PCR, and gel extraction for the bivalent template were performed as before, and the bivalent library was double-digested using *Typell*S restriction enzymes and ligated to a 5'-T7 promoter and 3'-ToIA spacer for RD (Figure 4-7B).<sup>37</sup> In addition, the sequences between the library-linker and the ToIA spacer were optimized to introduce more unique sequences (originally glycine-serine rich region) and allow for improved primer annealing during reverse transcription. Sanger sequencing of this mixed pool showed good NNK distribution and the library in-frame, flanking the interdomain linker (Figure 4-7C).

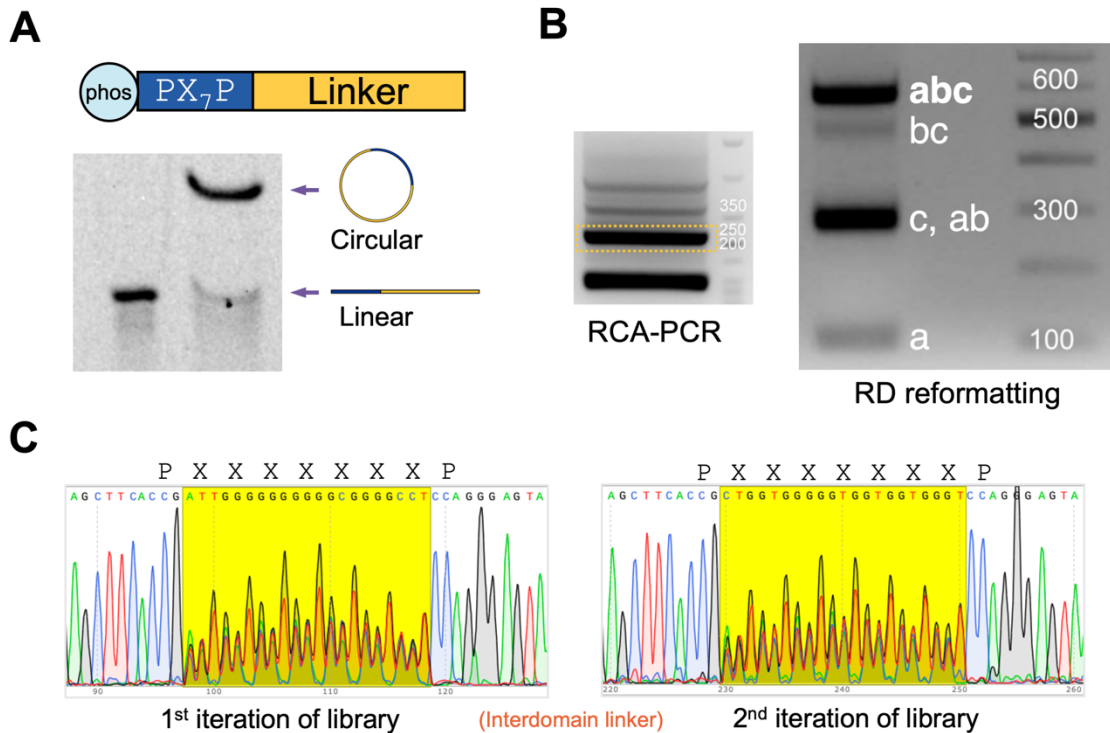


Figure 4-7. Circularization of ssDNA and overview of peptide library cloning.

(A) 15% acrylamide-urea denaturing gel of starting material (5'-phosphorylated ssDNA, left) and circularized material (right) using the *CirLigase* enzyme. Circularized product migrates slower than linear ssDNA. (B) Gel extraction of bivalent template (b) and 3-piece linear ligation to T7 promoter (a) and ToIA spacer (c) to create full RD template (abc). (C) Sequencing data from mixed pool Sanger sequencing of library shows libraries in-frame with good NNK distribution (highlighted).

RD library selections were performed against monovalent or bivalent Grb2 or Gads

SH3C targets immobilized on well plates. Target input was kept 10-fold lower than the library input to minimize crosslinking across two separate targets, and the ribosome input was also slightly reduced (90% of mRNA input) to minimize polysome formation. The washes were kept to instant washes for the first round given the diversity of the input library, but longer washes, the same as those in the control off-rate panning experiments, were used to increase the selection stringency over successive rounds.

While we were able to see slight differences in enrichment between selections against monovalent or bivalent target for Gads after 4 rounds in our RT-PCR data, we did not observe any signal differences for selections against monovalent and bivalent Grb2 (Figure 4-8A). Unexpectedly, sequencing of single clones of the selected pools from selections against Gads (Round 4) showed many frameshift mutations (insertions and deletions) for both monovalent and bivalent pools. More troublingly, the first and second iteration of the library flanking the interdomain linker were not the same, which strongly undermined our bivalent binder library strategy. To investigate further, we sequenced single clones from previous rounds against Gads (Rounds 1 - 3) as well as the naïve library and found that the errors accumulated over each round, which suggested major issues with amplification between rounds, most likely from the repetitive nature of the interdomain linker.

Since precise library presentation, in which the first and second iterations of the library are presented in-frame adjacent to the interdomain linker, was still observed in ~29% of the naïve library, we performed a single round of selection with stringent wash conditions (30 or 60 sec for Grb2, 30 or 60 min for Gads) to see if we could capture differences in enrichment between selections against monovalent and bivalent targets. Unfortunately, no signal differences were observed across different targets and wash conditions in both RT-PCR and qPCR data. Altogether, these results showed that the

bivalent library selections failed due to technical challenges with amplification.

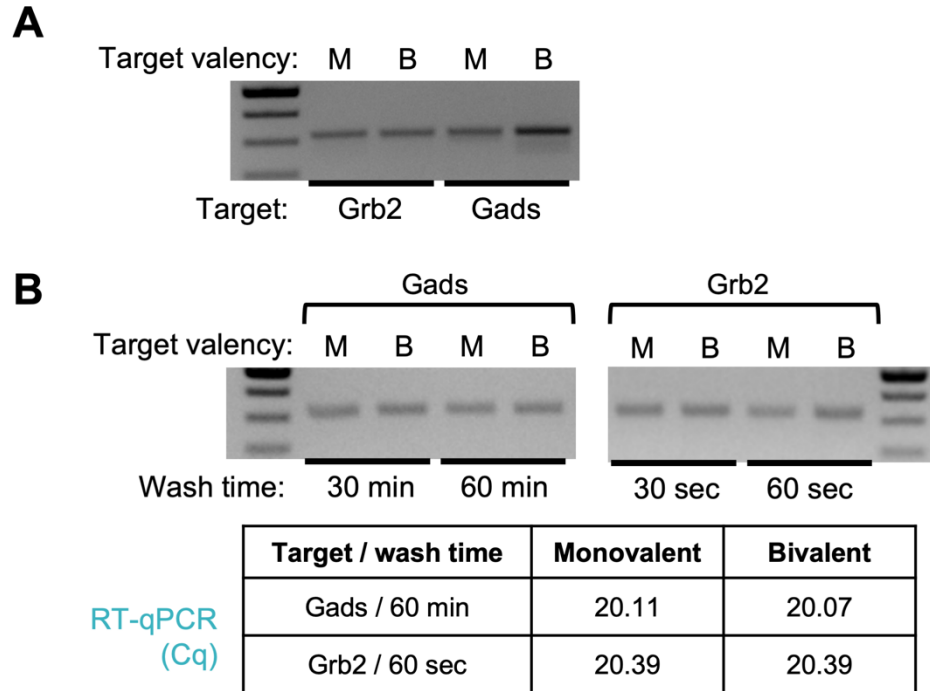


Figure 4-8. RT-PCR and RT-qPCR data from library selections.

(A) RT-PCR results from 4 rounds of selection against monovalent (M) or bivalent (B) Grb2 SH3C or Gads SH3C. Washes for initial rounds (rounds 1 and 2) were mostly instant, while later rounds used more stringent wash conditions (see Materials and Methods). Some signal difference is observed for selections against monovalent or bivalent Gads SH3C, but sequencing showed unsuccessful amplification between rounds. (B) RT-PCR results for a single library selection round, this time with varied wash conditions. Data showed lack of signal difference between all conditions (target format, wash duration) in both RT-PCR and qPCR.

#### 4.3.6 Proof-of-principle validation of bivalent RD particle binding through control selection

Shifting away from library selections, we expanded on our initial control off-rate selections and performed proof-of-concept selections with mock libraries to demonstrate improved particle recovery in a bivalent RD system. Using previously characterized single alanine substitutions of SLP-76 with varying dissociation rates, we created a panel of bivalent and monovalent templates to pan against bivalent Gads-SH3C targets in RD (Figure 4-9A). For two of these mutants, P241A and L243A, we combined them at two different molar ratios (1:1 and 10:1 WT-monovalent:Ala-mutant) (Figure 4-9B-E). We



observed that at the same 10:1 ratio against WT-monovalent, the bivalent ligands produced 97-fold (P241A) and 527-fold (L243A) higher recoveries over their monovalent counterparts (Figure 4-9B and 4-9C). Surprisingly, L243A, the putatively slower-dissociating mutant, showed a greater difference in particle recovery than did P241A; however, our single-clone panning data suggest that L243A may be a weaker binder than P241A in our experimental system, which would explain its greater signal difference.

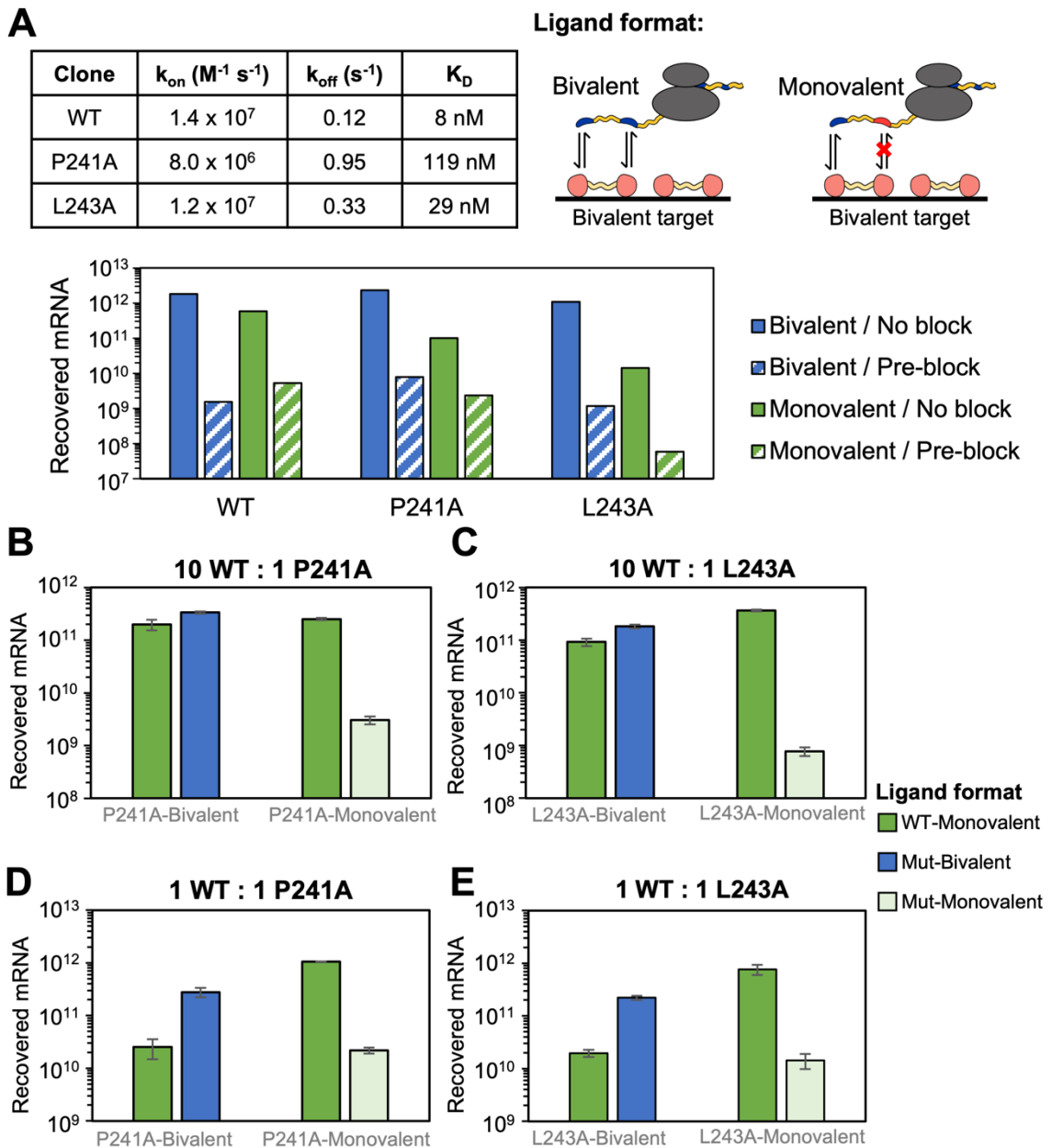


Figure 4-9. Test selections of SLP76 using our bivalent RD system.

(A) Table shows published kinetic rate constants for WT and the single-alanine substituted mutants P241A and L243A,<sup>63,189</sup> and the schematic shows bivalent and monovalent ligand-target binding, with binding-ablating mutations on the second peptide in the monovalent template to minimize other changes (such as translation efficiency or nonspecific binding) that could occur if the peptide length was changed. Bivalent or monovalent templates of WT, P241A, or L243A clones were translated and panned following incubation with WBT (“No block”) or WBT with 20-fold excess free bivalent Gads SH3C (“Pre-block”). The bivalent and monovalent unblocked WT templates have minimal signal difference (3-fold), while the faster dissociating mutants P241A and L243A have greater signal differences between their bivalent and monovalent unblocked templates (23-fold and 76-fold, respectively). (B) Two binary libraries containing mRNA templates for WT and P241A (in bivalent or monovalent formats) at a ratio of 10:1 were subjected panning against bivalent Gads-SH3C with 60-minute wash and analyzed by qPCR. The left plot shows mock library selection data of 10 WT monovalent : 1 P241A bivalent, and the right plot shows data of 10 WT monovalent : 1 P241A monovalent. Recovered mRNA shown are of each respective gene post-selection. (C) Same experiment as in (B), but with WT and L243A (also in monovalent and bivalent formats). (D) Binary libraries containing mRNA templates for WT and P241A at equimolar ratio in the same panning protocol as (B) and (C) showed 42-fold enrichment of bivalent mutant templates for P241A. (E) Same experiment as (D) showed 39-fold enrichment for bivalent L243A templates than monovalent WT templates.

## 4.4 Discussion

### 4.4.1 RCA-enabled synthesis of *in vitro* multivalent libraries

Multivalency is a powerful tool for effective targeting and discovery of protein therapeutics by leveraging increased avidity and higher specificity from proximity-induced increases in effective concentrations. Naturally occurring biological systems harness this effect, as seen in bivalent mAb binding, cell-surface binding via glycoproteins, transcription factors binding to DNA, ubiquitin-mediated cell signaling events, and viral entry, in which clusters of low affinity viral coat proteins engage multiple cell surface receptors to gain entry.<sup>190–192</sup> While some *in vitro* display methods leverage multivalency to engage multiple targets, the valency achieved is often imprecise<sup>174,193</sup>, which does not fully exploit increased avidity effects, or requires *in vitro* compartmentalization, which limits library diversity and can be technically challenging.<sup>26,175</sup> Thus, we implemented an RCA-enabled *in vitro* library construction approach for facile display of homomultivalent ligands.

A key first step in creating a homomultivalent library for RCA is designing the interdomain linker. For our initial biopanning experiments with bivalent HER2-binding DARPins, we used a simple and flexible glycine-serine linker that was sufficiently long to span the distance between homodimeric HER2 receptors. Although we were able to show successful bivalent engagement of our constructs, the receptor distribution and mobility on the cell surface is an important criterion to consider for multivalent biopanning experiments.<sup>184,194</sup> For example, localization microscopy studies showed that in the SK-BR-3 breast cancer cell line, HER2 clusters within the membrane and that the size of the clusters follows a normal distribution.<sup>195</sup> Thus, receptor organization, including binder-driven clustering events, must be considered for designing interdomain linkers.

In our biopanning experiments, we observed low cell-surface adsorption of unmodified DNA (~100 copies per cell). While our extended washing protocol likely contributed to decreased nonspecific binding, the lack of adsorption suggests that for biopanning against overexpressed targets, DNA stealthing approaches as described in Chapter 2 may not be necessary for initial selection rounds. However, later selection rounds may still benefit from stealthing modifications to ensure that potentially nonspecific, sticky particles do not outcompete true binders.

Library selections using our RCA-enabled multivalent libraries were ultimately unsuccessful. Single-clone sequencing data performed in between successive selection rounds indicated single nucleotide frameshifts, most likely due to the repetitive linker sequence. The relatively GC-rich (59%) mixed-species interdomain linker, in part due to the high glycine content in the flexible regions, has a low energy of folding ( $\Delta G = -17.55$  kcal/mol, calculated by UNAFold<sup>196</sup> under standard conditions) which may result in stable secondary structure even at the elevated 72 °C extension steps. This may lead to a myriad of issues with polymerase extension, such as the polymerase slipping/skipping or falling

off altogether, which can subsequently cause frameshift mutations and partial products.<sup>197</sup> In addition, we observed DNA template recombination, in which two different templates anneal and presumably prime one another prior to completion of extension, resulting in mixing of the first and second iterations of the library on a single template. This effect has also been observed in PCR amplification of repeat proteins, in which partially extended products act as “mega primers” that randomly prime other regions of the repeat protein.<sup>198</sup> Previous works aimed at optimizing PCR conditions include addition of DMSO to minimize secondary structure of DNA, using a high-fidelity DNA polymerase such as Phusion (which we have used here, in addition to Q5, an even higher-fidelity DNA polymerase), addition of single-strand binding proteins (SSBs) to stabilize primer annealing, changing salt conditions in the buffer, and changing annealing and denaturation temperatures.<sup>198–200</sup> Unfortunately, many of these improvements were unsuccessful at completely removing undesired side products, and were generally template- and primer-dependent; implementation of these optimizations will require careful testing of each parameter.

To address the issues with template recombination post-denaturation and insufficient polymerase extension, an isothermal amplification method, much like the one used for RCA, may be useful. While RCA does not allow exponential amplification, in a recent study, Gavrilov *et al.* developed an engineered helicase that unwinds DNA with high speed and processivity and when combined with SSB and standard PCR reagents, enables exponential isothermal amplification.<sup>201</sup> Eliminating high-temperature denaturation steps, in addition to optimization of the linker composition by incorporating more unique amino acid sequences and applying combinatorial codon scrambling<sup>202</sup> for less repetitive nucleotide sequences, could allow for more efficient and faithful extension during amplification and may be an alternative approach for multivalent library synthesis. In addition to improvements in amplification, a non-repetitive linker may also increase the

stability and monomeric ratio (vs. aggregated forms) of the translated binder.<sup>203,204</sup>

While we were not able to show enrichment in a library system, our proof-of-concept selections demonstrated that precise multivalent presentation is a useful strategy for rescuing low-affinity binders in RD, an *in vitro* selection method. Notably, panning experiments against Grb2 showed that this approach can be applied to micromolar binders, even in conditions in which the binder may be lost within the noise of the experiment. RCA can also be used to quickly create homomultivalent constructs for analytical RD applications to characterize fast-dissociating binders that are otherwise difficult to study in monovalent formats. Compared to traditional ribosome display, the additional input processing steps outlined here are estimated to reduce the functional number of input library members by only ~5-fold (Table 5-4), thereby maintaining the advantage of cell-free library size compared to cell-based methods. In addition to the technical optimizations proposed above, incorporation of in-solution panning and subsequent pull-down to reduce crosslinking may also improve the current methodology for successful implementation of this technology in full library selections. Finally, beyond achieving specific enrichment of bivalent ligand-target interactions over monovalent interactions, both our biopanning and RD data support that monovalent binders can be competed away by bivalent binders due to their slower dissociation, which suggests that multivalency may be a useful approach for engineering therapeutics to compete with endogenous agonists to partially suppress receptor signaling.<sup>47</sup>

#### *4.4.2 Extended commentary and perspective: General design principles and considerations for multivalent therapeutics*

Within the last decade, multivalent and multispecific therapeutics such as bi- and trivalent nanobodies and bispecific antibodies have achieved modest clinical success, with

more in the clinical development pipeline.<sup>156,205</sup> Multivalent polymers and nanoparticles, which achieve higher valency from conjugation to functionalized polymers and/or association within lipid bilayers in micelles or liposomes have also been used for effective targeting and drug delivery applications.<sup>206,207</sup> Despite mounting interest in this field, prior experimental applications of this phenomenon for therapeutic design and binder discovery often do not exploit multivalency effects to their fullest potential.

As mentioned previously, three important criteria to consider in multivalent design are affinity, valency, and linkage.<sup>62</sup> Affinity (of the monovalent interaction) is a highly useful parameter that can be tuned by the user to produce their desired biological effect. Previous studies have successfully demonstrated ultrahigh-affinity of multimeric constructs, often synthesized by linking high-affinity ligands to leverage avidity effects.<sup>208,209</sup> For example, an extended repertoire of binders developed against the Spike (S) protein of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) utilize multivalent designs to take advantage of multi-domain binding to the trimeric S protein (Table 4-1).

<b>Binder</b>	<b>Monovalent affinity (K<sub>D</sub>)</b>	<b>Valency</b>	<b>Linkage</b>	<b>Affinity against trimeric S (K<sub>D</sub>)</b>	<b>Ref.</b>
Nanobody	38.6 nM	3	N-term, PEG linker	2.31 nM	210,211
Recombinant ACE2 binding domain	N.D.	~5	Self-assembly through coiled-coil protein domain	0.62 nM	212
IgG	9.45 nM (IgG)	10	Engineered into pentameric IgM	< 0.001 nM	213
Diabody	0.16 nM	4	Connect IgG and diabodies through constant regions (CH2-CH3)	< 0.001 nM	214

Recombinant ACE2 binding domain (fused onto IgG)	3.4 nM (IgG)	12	Engineered into hexameric IgM	0.9 nM	215
Monobody	0.42 nM	2	Tandem dimer with (G <sub>4</sub> S) <sub>3</sub> linker	< 0.001 nM	216
Minibinders (3 separate mutants)*	0.034 - 0.35 nM	3 (heterovalent)	Pro <sub>12</sub> , Pro <sub>24</sub> , or Gly <sub>10</sub>	0.005 ~0.007 nM**	173

\*3 different mutants were linked together in heterotrivalent format, with varying linkers in between the domains

\*\*Affinity varies slightly based on binder and linker combination used for heterovalent constructs

Table 4-1. Selected recently published multivalent therapeutic binders against trimeric Spike protein.

Binders can be linked biologically, through self-assembly or genetically encoded interdomain linkers, or synthetically, by conjugation to polymeric scaffolds like multifunctional PEG (see Table 4-1). Other works have closely matched the specific geometries of the S receptor binding domain (RBD) to engage the target most optimally.<sup>173,217</sup> Here, application of multivalent designs is highly effective, as the increased avidity effects are useful for retaining binding activity and minimizing mutational escape even for evolved variants despite decreased monovalent affinities due to novel mutations.

In other contexts, in which qualities such as specificity or selectivity are desired beyond achieving increased affinity, combining low-affinity binders into multivalent constructs can be a useful approach to reduce off-target binding. For example, tumor cells are often distinguishable from healthy cells by overexpression of cell surface receptors like select GPCRs and epidermal growth factor receptors (EGFRs).<sup>218</sup> By linking low-affinity binders into multivalent formats, selective tumor targeting can be achieved as the weak individual interactions require avidity, conferred by the high surface expression on cancer cells, to bind effectively.<sup>219–221</sup> This effect has been exploited by chimeric antigen

receptor (CAR) T cell targeting<sup>222</sup>, in which reduced binding to tumor-associated antigens (TAAs) like EGFR and HER2 allow T cells to distinguish tumor cells<sup>223</sup> and even enable avidity-dependent AND-logic gates<sup>166</sup>. The monovalent affinity of each domain within a homo- or heteromultivalent protein can be fine-tuned to achieve more avid, selective, or logic-gated binding profiles.

Like affinity, valency is also a highly modular parameter within multivalent design. Immunoglobulin M (IgM), a naturally occurring macromolecular assembly, can achieve valency of up to 10 (made of 5 subunits, rarely 6)<sup>224</sup>; while individual IgMs display low antigen affinity, pentameric IgM achieve sufficient binding avidity for antigen recognition and opsonization.<sup>225</sup> As a result, IgMs are particularly effective in providing protection against microbes with closely spaced and highly expressed surface epitopes<sup>226</sup>, even demonstrating differences in complement activity towards distinct surface curvatures.<sup>227</sup> In systematic studies, tuning the valency of multidomain proteins or ligands on the surface of nanoparticles has demonstrated varying therapeutic responses and modes of cell internalization. In one study, Romero *et al.* created multivalent anti-CD99 antibody fragment (Fab) with 2, 3, or 4 subunits via engineering of tetravalent streptavidin and showed that while an increased valency beyond 2 had no effect on binding, a valency of 3 is necessary and sufficient for inducing cytotoxicity.<sup>228</sup>

In a nanoparticle-based study using trans-activating transcription (TAT) peptide-functionalized nanoparticles, Dalal and Jana showed that low TAT peptide valency (TAT<sub>10</sub>) led to different uptake mechanisms compared to those with high TAT valency (TAT<sub>75</sub>); high TAT valency led to strong interactions with the cell membrane, which produced large vesicular traps and led to microtubule-dependent exocytosis after 12 hours.<sup>229</sup> Contrastingly, low TAT valency led to smaller vesicle formation and microtubule-dependent trafficking toward the Golgi apparatus and the nucleus.<sup>229</sup> Finally, in another



study, Wang *et al.* compared cellular uptake of nanoparticles functionalized with varying valency of HER2-targeting peptides across breast cancer cell lines with different levels of HER2 expression. Here, they showed that higher peptide valency is most beneficial (i.e., higher uptake) only when the receptor density is also high, and suggested that this effect can be exploited to design multivalent nanoparticle therapies to match the level of HER2 expression for cancer patients.<sup>230</sup> Thus, valency of a multivalent therapeutic should closely mimic the targeting protein or receptor, and can be tailored to achieve specific uptake profiles.

Lastly, linkage between monomeric domains, closely associated with the valency, remains an underexplored parameter; increased efficacies of multivalent therapeutics are often underwhelming due to suboptimal linker designs between ligands. In many cases, binders are made bivalent or trivalent by simply adding an arbitrary flexible linker.<sup>216,231</sup> These linkers are not always optimized for target engagement *in vivo*; thus, a drastic decrease in binding dissociation constants *in vitro* do not necessarily result in an increase in efficacy to the same magnitude. Flexible linkers, often encoded by the amino acids glycine and serine, are commonly used as they maintain the ability to explore conformation space after binding.<sup>232</sup> Despite their apparent interchangeability, differences in the composition of glycine and serine residues within a linker leads to differences in spacing and separation.<sup>233</sup> Serine has a longer persistence length and is more suited for bridging long distances as it can take on more conformations. Interestingly, one study showed that increasing glycine-serine subunits within a linker does not linearly increase its hydrodynamic radius as the entropic costs to span and occupy the full length is too high.<sup>234</sup> Another practical consideration for linker composition and design is that computational<sup>235</sup> and experimental studies<sup>236</sup> have also suggested that too much flexibility of an interdomain linker can destabilize protein fusions, and that some restrictions on ligand flexibility may

be beneficial for ligand-receptor binding.<sup>237</sup> With these considerations, a mixed-species linker with both rigid and flexible regions may be beneficial for achieving separation and solubility while maintaining sufficient exploration of the conformation space.

A defined spatial separation, which can be achieved by matching the geometry of the ligand-receptor binding interface and adding rigidity, can also be useful in some contexts for inducing cell signaling events. In an elegant study, Mohan *et al.* showed that the topology of a dimeric erythropoietin receptor (EpoR)-binding DARPin ligand can be precisely controlled to tune the strength of cytokine signaling, from full, biased, and partial agonism of EpoR signaling in cells.<sup>238</sup> Structure-guided approaches for linker design have also been applied to bivalent ligands for GPCRs<sup>239,240</sup>; one study showed that longer distances (20 or 30 Å) between ligands induced dimerization with unique signaling responses, while a non-dimerization-inducing ligand (10 Å interdomain distance) also produced a different response due to allosteric effects.<sup>239</sup> In a comprehensive study, Hunt *et al.* performed high-throughput screening of structurally-guided multivalent minibinders against S RBD, and compared various linker compositions and lengths to develop binders that match specific conformations (open vs. closed) of the trimeric S protein (also listed in Table 4-1).<sup>173</sup> Interestingly, cryo-EM structures comparing a short, flexible interdomain linker with a longer, rigid interdomain linker showed intrinsically different modes of binding but similar neutralization potencies. Given sufficient structural information, distinct separation of receptor-binding domains can be useful for capturing specific conformations and subsequently inducing desired cellular responses.

Altogether, affinity, valency, and linkage are critical parameters that should be carefully considered for optimal multivalent design. Beyond simple avidity effects, structural-<sup>241</sup> and computationally-guided<sup>63</sup> designs for precise matching of ligand-target geometries can allow the user to rationally tune designs to therapeutically relevant

arrangements to achieve better selectivity, specificity, and/or cell signaling.

#### **4.5 Conclusion**

In summary, our results demonstrate specific enrichment of bivalent ligand-target interactions over monovalent interactions in RD, an ultra-high-capacity *in vitro* selection system. As we observed in our peptide library selections, successful implementation of this technology in full library selections will require further technical improvements for high-fidelity amplification such as optimized linker composition and innovative polymerase extension steps. While precise spatial matching of ligand to target is most desirable, as indicated by MVsim, recombinantly produced single-chain multivalent targets may be substituted with other methods such as streptavidin-mediated multivalent presentation of biotinylated targets; this should ideally be accompanied by a suitably spaced bivalent RD presentation to mirror the target display. Finally, in addition to its utility in engineering new proteins, RCA can be used analytically to quickly create homomultivalent constructs for RD applications to characterize fast-dissociating binders that are otherwise difficult to study in monovalent formats.

#### **4.6 Materials and Methods**

All oligonucleotides (Ultramers, gBlocks, and primers) were purchased from Integrated DNA Technologies (IDT). DNA purification was performed with the QIAquick PCR purification, nucleotide removal, or gel extraction kits (Qiagen). All plasmids were purified using the QIAprep spin miniprep kit (Qiagen). Restriction enzymes, T4 DNA ligase, phi29 DNA polymerase, Phusion HotStart DNA polymerase, and Q5 polymerase were purchased from New England Biolabs (NEB). CircLigase II enzyme was purchased from

Lucigen. Concentrations of nucleic acids (DNA and RNA) and proteins were determined by absorbance readings at 260 nm and 280 nm, respectively, using Cytation3 (Biotek).

#### *4.6.1 Circularization and rolling circle amplification*

For circularization of ssDNA, a 5'-phosphorylated Ultramer was used as input and circularized using CirLigase II for 2 hours at 60 °C according to the manufacturer's protocol (20 ng/μL of reaction). Ligation efficiency was estimated using ImageJ. For circularization of dsDNA, a PCR-amplified template was digested at 37 °C using either MfeI-HF or BamHI-HF and column purified. Although this concentration is lower than that of the ssDNA ligation, 1.5 ng/μL in a 100 μL reaction volume still accommodates an input of  $\sim 2.8 \times 10^{11}$  library members for a 500 bp template, which is sufficient for  $\sim 94\%$  probability of capturing a  $10^{11}$  library diversity. The purified product was incubated with T4 DNA ligase and purified by gel extraction. Both ssDNA and dsDNA circularized templates were amplified via rolling circle amplification (RCA) with phi29 DNA polymerase for 30 min at 30 °C. The RCA products were purified by column purification and amplified by PCR, and this mixed-valency RCA-PCR product was purified by gel extraction to extract the band corresponding to the desired valency.

#### *4.6.2 Plasmid construction and protein expression*

The sequences of the SH3 C-terminal domain of the human adapter proteins Gads and Grb2 were obtained from UniProt and subcloned into the pET22b vector (Novagen) in monovalent or bivalent (with an interdomain linker) formats and expressed in *E. coli* BL21(DE3) (Agilent). Similarly, all G3 (WT and AVD, in monovalent or bivalent) constructs synthesized by RCA-PCR or PCR were subcloned into pET22b constructs as N-terminal fusions to SNAP (described in 3.7.2). Proteins were purified by Ni<sup>2+</sup> immobilized metal ion affinity chromatography (Qiagen), rebuffed, and stored at -80 °C.

#### 4.6.3 Preparation of mock DNA display particles

Coupling of BG-GLA-NHS (NEB) to an amino-modified primer was performed as follows:

Reagent	Final concentration
BG-GLA-NHS	300 $\mu$ M
5'-amino primer (in water)	80 $\mu$ M
Phosphate buffer, pH 7.8	50 mM
Nuclease free water	Up to 50 $\mu$ L

The mixture was incubated for 2 hours at RT and purified using the nucleotide removal kit. The incubation and purification were repeated 1x. The BG-functionalized primer was used as the forward primer in a standard PCR reaction for n5T template synthesis (20 cycles). For covalent coupling of the BG-template to SNAP-fusion proteins, the incubation was performed as follows:

Reagent	Final concentration
BG-DNA template	0.18 $\mu$ M (~600 ng)
SNAP-fusion protein	4.5 $\mu$ M
TCEP, pH 7	1 mM
NaCl	150 mM
Tris-HCl, pH 7.5	50 mM
Nuclease free water	Up to 12 $\mu$ L

The reaction was incubated at 37 °C for 4.5 hours, then at 4 °C overnight. Multiple reactions (~4x) were performed to achieve higher input. To purify, 2.4  $\mu$ L of 6X loading dye (no SDS)<sup>242</sup> was added to the 12  $\mu$ L reaction and run on a 0.5% low melting point agarose gel (Millipore-Sigma) for 85 min at 110V in the cold room. The gel slice containing the protein-conjugated template was extracted, placed in a Freeze 'n Squeeze filter cup (Bio-Rad), and macerated. The spin column with the filter cup was placed at 4 °C for 45 min, and NaCl (150 mM), TCEP, pH 7 (1 mM) and 1X agarase buffer (NEB) was added. The reaction (in the filter cup inside the spin column) was incubated in a water bath at 42 °C for 50 min. Then, 2.4  $\mu$ L of  $\beta$ -agarase I (10 U/ $\mu$ L) (NEB) was added, pipetted to mix using a positive-displacement pipet (Gilson), and incubated at 42 °C for 45 min. The spin

column was then centrifuged at 13,000 x g for 3 min at RT (spinning at lower temperature will cause the gel to solidify). Finally, the eluate was transferred to a 30 kDa MWCO diafiltration column (Amicon) and rebuffered using Buffer EB with repeated centrifugation steps at 7,500 x g. The remaining dead volume was adjusted by adding concentrated NaCl (5 M), TCEP (250 mM), and Tris-HCl, pH 7.5 (500 mM) to final concentrations of 150 mM, 1 mM, and 50 mM, respectively. Purified protein-conjugated templates were kept at 4 °C.

#### 4.6.4 Biopanning

SK-OV-3 cells were grown at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub> in McCoy's 5A medium (Gibco) containing 10% (v/v) fetal bovine serum (FBS) (Gibco). For all biopanning experiments, the cells were harvested at 80-90% confluency by adding 0.25% trypsin-EDTA (Gibco) and incubating at 37 °C for 3 min. To terminate trypsinization, the trypsin was diluted with 7 mL of PBS-B (1X DPBS with 10% BSA (Millipore-Sigma)) and transferred to a 15 mL conical tube. Cells were washed 3X with PBS-B and diluted to 1.3 x 10<sup>7</sup> cells/mL. To diluted protein-conjugated particles (2.8 x 10<sup>10</sup> copies in 9 µL of PBS-B), 15.5 µL of the cell suspension (200,000 cells) and 0.5 µL of DTT (1 mM) were added in 0.2 mL PCR tubes and pipetted to mix. For competed samples, 200x molar excess of free protein (corresponding to the template-conjugated protein) was added to the panning mixture. Next, the PCR tubes containing the panning mixture were taped to a vortex mixer (knob 4, Fisher Scientific) and panned for 75 min at 16 °C (inside microbial shaker set at 16 °C).

For washes, PBS (RT) was added to 257.5 µL total volume, pipetted to mix, and incubated for 5 min at RT. Then, the PCR tubes were centrifuged at 325 x g for 6 min at 10 °C and aspirated until ~20 µL remained. Finally, 12.5 µL of remaining buffer were manually pipetted out such that ~7.5 µL remained in the tube. This wash cycle was repeated 5x. After the last wash, each sample was adjusted to 20 µL, and 10 µL was taken

directly into qPCR, while the remaining 10  $\mu$ L was used for cell counting. The same qPCR protocol as described in 2.6.7 was used, but without template correction.

#### *4.6.5 Library and control template cloning*

The bivalent library was prepared by circularizing the ssDNA Ultramer as described in 4.6.1. The gel-extracted and purified bivalent library was double digested using BbsI-HF and column purified. The T7 promoter and TolA genes were PCR amplified from pRDV2 Off7<sup>243</sup>, digested with BbsI-HF, and column purified. The purified products were ligated at a ratio of 2 : 1 : 2 (T7 : library : TolA)<sup>243</sup> for 13 hours at 16 °C, then 3 hours at RT. The ligated product was run on an 1.5% agarose gel, extracted, and purified. Low-cycle (7 - 10 cycles) PCR was performed using gel extracted product for higher input. IVT was performed as described in 3.7.6.

For the control selections, bivalent SLP76 was subcloned into pRDV2,<sup>69</sup> transcribed, and purified as described in 3.7.6 or using RNeasy purification kits (Qiagen). Monovalent templates used in model selections were made with affinity-ablating mutations in the second peptide<sup>63</sup> (to keep the construct length and composition largely intact) and assembled by linear digestion-ligation reactions (Table 5-2).

#### *4.6.6 Ribosome display*

Target protein was added to each well of a Maxisorp (Nunc) strip and immobilized overnight at 4 °C, then blocked for 2 - 4 hours with SynBlock (Bio-Rad) at 4 °C. PURExpress (NEB) was used to prepare the translation mix and stopped to generate mRNA-ribosome-protein complexes as described previously.<sup>244</sup> For the pre-blocked control experiments, the translation mix was incubated with WBT (50 mM Tris-HCl, pH 8, 150 mM NaCl, 0.05% v/v Tween-20, 50 mM MgOAc) or target protein for 30 min at 4 °C prior to panning for 1 hour. The molar ratio of target : ribosome : mRNA in the selection was kept to 0.1 : 0.9 : 1 (1 being  $6 \times 10^{12}$  molecules) for both library selections and

Grb2/Gads monovalent and bivalent off-rate panning experiments to prevent crosslinking and polysome formation. For the test selections with P241A and L243A mutants, the target input was increased from  $6 \times 10^{11}$  to  $3 \times 10^{12}$  as the second peptide was non-binding. Following panning, the wells were subjected to six instant washes with WBT and one competitive wash (WBT with 20-fold excess free target) for 1 hour unless stated otherwise. For library selections, the washes for selections against Grb2 SH3C were as follows: Round 1 (R1) 7x instant; R2 7x instant; R3 3x instant, 3x 2 min; R4 6x 2 min. Similarly, washes for library selections against Gads SH3C were as follows: Round 1 (R1) 7x instant; R2 5x instant, 2x 1 min; R3 5x instant, 1x 3 hours (competitive, with 100-fold molar excess free bivalent Gads); R4 5x instant, 1x 5 hours (competitive, with 50-fold molar excess free bivalent Gads).

#### 4.6.7 RT-qPCR

*In situ* reverse transcription was performed using AffinityScript reverse transcriptase (Agilent) as described previously.<sup>69</sup> The RT product was collected into 0.2 mL tubes, and 1  $\mu$ L of the product was put into a 10 - 20  $\mu$ L qPCR reaction using PowerUp SYBR Green master mix (Thermo Fisher Scientific) (Table 5-3). The total mRNA copy number was calculated using a standard curve from reverse transcribed purified mRNA and the total volume recovered from the RT reaction.



Table 4-2. List of amino acid sequences for proteins used in library and control selections.

Construct name	Description	Amino acid sequence
G3 <sub>WT</sub> -SNAP	Monovalent G3 <sub>WT</sub> for biopanning (GS and KL are BamHI and HindIII site used for subcloning, respectively)	MRGSHHHHHHLEVLVLFQGGPSDLGKKLLEAARAGQDDEVRI MANGADVNAKDEYGLTPLYLATAHGHLEIVEVLLKNGADV AVDAIGFTPLHLAAFIGHLEIAEVLLKHGADVNAQDKFGKT AFDISIGNGNEDLAEILQKLGSGGGSGGGSGGGSGGGSM DKDCEMKRRTTLDSPGKLELSGCEQGLHEIIFLKGKTSAA AVEVPAPAAVGGPEPLMQATAWLNAYFHQPEAIEEFPVPA LHHPVFQQESFTRQVLWKLKLVVVFGEVISYSHLAAALAGNP AATAAVKTAALSGNPVPIIPCHRUVVQGDLDVGGYEGGLAVK EWLLAHEGHRLGKPLG
G3 <sub>AVD</sub>	Monovalent G3 <sub>AVD</sub> for biopanning (gene between GS and KL restriction sites shown)	DLGKKLLEAARAGQDDEVRI MANGADVNAKDEYGLTPLYL AAAHGHLEIVEVLLKNGADVNAVDAIGFTPLHLAAFIGHLE IVEVLLKHGADVNAQDKFGKTAFDISIDNGNEDLAEILQ
G3 <sub>WT</sub> -G3 <sub>WT</sub>	Bivalent G3 <sub>WT</sub> for biopanning (gene between GS and KL restriction sites shown)	MDLGKKLLEAARAGQDDEVRI MANGADVNAKDEYGLTPLY LATAHGHLEIVEVLLKNGADVNAVDAIGFTPLHLAAFIGHL EIAEVLLKHGADVNAQDKFGKTAFDISIGNGNEDLAEILQK LGSGGGSGGGSGGGSGGGSGGGSMKDEQLMDLGKKLLEAARA GQDDEVRI MANGADVNAKDEYGLTPLYLATAHGHLEIVEV LLKNGADVNAVDAIGFTPLHLAAFIGHLEIAEVLLKHGADV NAQDKFGKTAFDISIGNGNEDLAEILQ
G3 <sub>AVD</sub> -G3 <sub>AVD</sub>	Bivalent G3 <sub>AVD</sub> for biopanning (gene between GS and KL restriction sites shown)	MDLGKKLLEAARAGQDDEVRI MANGADVNAKDEYGLTPLY LAAAHGHLEIVEVLLKNGADVNAVDAIGFTPLHLAAFIGHL EIVEVLLKHGADVNAQDKFGKTAFDISIDNGNEDLAEILQK LGSGGGSGGGSGGGSGGGSGGGSMKDEGSDLGKKLLEAARAG QDDEVRI MANGADVNAKDEYGLTPLYLAAAHGHLEIVEV LLKNGADVNAVDAIGFTPLHLAAFIGHLEIVEVLLKHGADV NAQDKFGKTAFDISIDNGNEDLAEILQ
Monovalent Gads	Monovalent human Gads SH3C target for all RD experiments	MRGSHHHHHHLEVLVLFQGGPSGSGSTGSRWARALYDFEAL EDELGFHSGEVVEVLDSSNPSWWTGRLHNKLGFLPANYVAP MTR
Monovalent Grb2	Monovalent human Grb2 SH3C target for all RD experiments	MRGSHHHHHHLEVLVLFQGGPSGSGSTGGPTYVQALFDFDPQ EDELGFRRGDFIHVMDNSDPNWWKGACHGQTGMFPRNYVTP VNR
Bivalent Gads	Bivalent human Gads SH3C target for all RD experiments	MRGSHHHHHHLEVLVLFQGGPSGSGSTGSRWARALYDFEAL EDELGFHSGEVVEVLDSSNPSWWTGRLHNKLGFLPANYVAP MTRSSGSGEAAAKEAAAKEAAKGTGSGSRWARALYDFEA LEDELGFHSGEVVEVLDSSNPSWWTGRLHNKLGFLPANYV APMTR
Bivalent Grb2	Bivalent human Grb2 SH3C target for all RD experiments	MRGSHHHHHHLEVLVLFQGGPSGSGSTGGPTYVQALFDFDPQ EDELGFRRGDFIHVMDNSDPNWWKGACHGQTGMFPRNYVTP VNRSSGSGEAAAKEAAAKEAAKGTGSGGPTYVQALFDFDP QEDGELGFRRGDFIHVMDNSDPNWWKGACHGQTGMFPRNYV TPVNR
Linker 1	Interdomain linker for bivalent targets	SSGSGEAAAKEAAAKEAAKGTGSG
Linker 2	Interdomain linker for bivalent ligand in Fig. 4-6A-C	SGSSGSGEAAAKEAAAKEAAKGTGSGGS
Linker 3	Interdomain linker for bivalent library	GSNSGTSSSGEAAAKEAAAKEAAKGGSGSGTGAS
Linker 4	Interdomain linker for bivalent ligand in Fig. 4-9A-E	GSGSNGSAEAAAKEAAAKEAAKATGSGGTS

SLP-76 peptide 1	Peptide ligand encoded in pRDV2 for Fig. 4-6A-C	PSIDRSTKP
SLP-76 peptide 2	Peptide ligand encoded in pRDV2 for Fig. 4-9A-E	A <sup>232</sup> PSIDRSTKPPLDR
Non-binding SLP-76 peptide	Non-binding peptide ligand encoded in pRDV2 for Fig. 4-9A-E	A <sup>232</sup> PSID <u>A</u> ST <u>A</u> APPLDR
P241A SLP-76 peptide	Weaker-affinity peptide ligand encoded in pRDV2 for Fig. 4-9A, B, D	A <sup>232</sup> PSIDRSTK <u>A</u> PLDR
L243A SLP-76 peptide	Weaker-affinity peptide ligand encoded in pRDV2 for Fig. 4-9A, C, E	A <sup>232</sup> PSIDRSTKPP <u>A</u> DR

Table 4-3. List of primers and gBlocks used in this study.

Primer name	Description	Sequence (5' → 3')
G3_MfeI_for	Forward primer (MfeI) for PCR of G3 for circularization	CGCCAATTGATGGACCTGGGTAAAAAACTG
G3_MfeI_rev	Reverse primer (MfeI) for PCR of G3 for circularization	GGCCAATTGCTCATCTTTGTCCATGGAACC
G3_BamHI_for	Forward primer (BamHI) for PCR of G3 for circularization	CGGGGATCCATGGACCTGGGTAAAAAACTG
G3_BamHI_rev	Reverse primer (BamHI) for PCR of G3 for circularization	GGCGGATCCCTCATCTTTGTCCATGGAACC
SLP76-linker-ultramer	Ultramer for SLP-76 ssDNA circularization (used for initial reaction parameter testing)	/5Phos/GGCCCGAGTATTGACCGCTCAACGAAGCCTAGCGGCAGTAGCGGATCCGGAGAGGCAGCGGCAAAAAGAGCTGCAGCGAAAAGAAGCTGCAGCGAAAAGGTACCGGTGTCAGGTGGCTCT
PXXXXXXXP-ultramer	Ultramer for binder library-linker ssDNA circularization	/5Phos/GGCACAGGAGCTTCCCGNNKNNKNNKNNKNNKNNKCCAGGGAGTAATCAGGAACGTCGAGTAGCGGAGAGGCAGCGCAAAAAGAAGCTGCAGCGAAAAGAAGCTGCAGCGAAAAGGTGGCTCTGCTCT
RCA_BbsI_library_rev	Reverse primer for RCA and RCA-PCR	GTAGAAGACATTTTCGTTCCCTGAGTTACTCCCTGG

RCA-PCR_Bbsl_for_v2	Forward primer for RCA-PCR	CGCTTAGTCCTGAAGACATGCTTGGC
RT_upstr_TolA_rev	Primer for RT step in library selections	CTTGCTTCTGAACACCATTTC
HiTm_T7temp_amp_for	Forward primer for RT-PCR	GCAGCTTAATACGACTCACTATAGGGACACC
Lib_Bbsl_re	Reverse primer for RT-PCR	CAGCTTGCTTCTGAAgACCATTCGTTCTCTG
RD-qPCR_T7_for	Forward primer for qPCR of library (Fig. 4-8B)	CTCACTATAGGGACACCACAAC
RD-qPCR_T7_rev	Reverse primer for qPCR of library (Fig. 4-8B)	TCTCCTTCTTTCCGGTGTGTTTC
Newupstr_T7prom_noBsa_for	Forward primer for PCR of T7 module from pRDV2	GCAGCTTAATACGACTCACTATAGGGACACCACAACGG
T7_newBbsl_rev	Reverse primer for PCR of T7 module from pRDV2	GATGAAGACCCAAGCATGGATATATCTCCTTCTTTTCGGTGTGTTCAAATTATTTTC
Bbsl_TolA_for	Forward primer for PCR of TolA module from pRDV2	GAAGAAGACTACGAATGGTGTTCAGAAGCAAGCTGAGGAG
T7prom_pRDV_noBsal_for	Forward primer for PCR of template from pRDV2 for <i>in vitro</i> transcription (IVT)	ATACGAAATTAATACGACTCACTATAGGGACACCACAACGG
tolAk	Reverse primer for PCR of TolA module and template from pRDV2 for IVT and reverse transcription (RT)	CCGCACACCAGTAAGGTGTGCGGTTTCAGTTGCCGCTTCTTTCT
RD_qPCR1.F	Forward primer for qPCR of all templates in experiments for Fig. 4-6	TTCCGGTGGCCAGAAG
RD_mid_rev	Reverse primer for qPCR of all templates in experiments for Fig. 4-6	GCATCTACCTCAGCCTTAGC
T7_RD-qPCR_v2_for	Forward primer for qPCR of all templates in experiments for Fig. 4-9	GAAGACACCGAAAGAAGGAGGAA

RD-qPCR_WT-M_rev	Reverse primer for qPCR of monovalent WT template in experiments for Fig. 4-9	GCGATCCAGTGGCGGTTTTG
RD-qPCR_P241A-M_rev	Reverse primer for qPCR of monovalent P241A template in experiments for Fig. 4-9A, B, D	ACGATCAAGCGGCGCTTTTC
RD-qPCR_L243A-M_rev	Reverse primer for qPCR of monovalent L243A template in experiments for Fig. 4-9A, C, E	GCGATCAGCCGGAGGTTTTTC
RD-qPCR_WT-B_rev	Reverse primer for qPCR of bivalent WT template in experiments for Fig. 4-9A	GCGGTCCAGAGGTGGTTTTC
RD-qPCR_P241A-B_rev	Reverse primer for qPCR of bivalent P241A template in experiments for Fig. 4-9A, B, D	GCGGTCCAGAGGTGCTTTTC
RD-qPCR_L243A-B_rev	Reverse primer for qPCR of bivalent L243A template in experiments for Fig. 4-9A, C, E	GCGGTCTGCAGGTGGTTTTC
gBlock_Grb2SH3C-UnivLink-GadsSH3C	gBlock used for cloning Grb2 and Gads	GGCCCAACCTATGTTTCAGGCATTATTTGACTTTGATC CACAAGAAGATGGTGAATTGGGATTCCGCCGTGGCGA CTTTATCCACGTCATGGACAACCTCCGATCCTAATTGG TGGAAGGGAGCATGTCACGGACAGACCGGCATGTTCC CACGCAACTATGTCACGCCGGTCAATCGCAGTAGCGG TTCCGGAGAGGCAGCGGCAAAAGAAGCTGCAGCGAAA GAAGCTGCAGCGAAAGGTACCGGGTCAGGTAGTGTGC GCTGGGCACGTGCGCTTTATGACTTCGAGGCATTGGA GGATGACGAGTTGGGCTTCCATTCTGGTGAGGTAGTC GAGGTACTGGATAGCAGCAATCCCTCGTGGTGGACGG GTCGCCTTCATAACAAACTTGGTCTGTTCCCGGCAAA TTATGTTGCACCGATGACGCGC

Table 4-4. Comparison of estimated yields of DNA input for traditional ribosome display and homobivalent ribosome display.

Step	Traditional Ribosome Display		Homobivalent Ribosome Display	
	<i>Yield</i>	<i>Molecules</i>	<i>Yield</i>	<i>Molecules</i>
Starting material		1.8 x 10 <sup>13</sup>		1.8 x 10 <sup>13</sup>
Circularization	–	–	0.90	1.6 x 10 <sup>13</sup>
Circularization purification	–	–	0.74	1.2 x 10 <sup>13</sup>
RCA	–	–	1 <sup>a</sup>	1.2 x 10 <sup>13</sup>
RCA purification	–	–	0.75 <sup>b</sup>	9.0 x 10 <sup>12</sup>
RCA-PCR purification	–	–	0.95 <sup>b</sup>	8.6 x 10 <sup>12</sup>
Bivalent template gel extraction purification	–	–	0.4 <sup>b</sup>	3.4 x 10 <sup>12</sup>
Template double digest purification	0.83	1.5 x 10 <sup>13</sup>	0.83	2.8 x 10 <sup>12</sup>
Three-piece linear ligation and gel extraction	0.30	4.5 x 10 <sup>12</sup>	0.30	8.5 x 10 <sup>11</sup>
<i>Final yield relative to traditional ribosome display</i>	<i>1</i>		<i>0.19</i>	

<sup>a</sup> Every input template is assumed to be amplified in RCA, which is feasible with low input and/or multiple reactions.

<sup>b</sup> Estimates are based on column yields according to manufacturer and general experimental protocols.

## Chapter 5: TurboA $\beta$ : Proximity biotinylation approaches for identifying amyloid beta interactomes at the blood-brain barrier

---

### 5.1 Acknowledgement and contribution

Study conceptualization, cell work (blood-brain barrier model, cell-based assays, and imaging), and selected *in vitro* assays were designed and performed by Paulina Eberts. I designed, cloned, expressed, and purified all TurboID and miniTurboID constructs and performed related protein characterization assays. Experimental design and analysis of data for both *in vitro* and cell assays were performed with Paulina Eberts.

### 5.2 Summary

Novel therapeutic targets are discovered through various techniques such as genome-wide association studies (GWAS) and identifying molecules within cellular signaling pathways via proteomic studies. Protein-protein interactions (PPIs) (such as those between a ligand and its therapeutic target) are further investigated through methods like co-immunoprecipitation (co-IP) and subsequently identified and validated using analytical tools such as mass spectrometry (MS) and western blots (WBs). Despite its utility, co-IP is challenging for identifying cell-surface receptor interactions, as they require solubilization conditions that may disrupt PPIs. In addition, as co-IP requires the PPI to stay intact, it may not be able to capture transient interactions (i.e., fast dissociating) and bias recovered protein-protein interactions towards those with higher affinity. In contrast, proximity labeling uses promiscuous labeling enzymes targeted to certain proteins or receptors by genetic fusions, which allow spatially restricted covalent tagging of PPIs and neighboring proteins. This approach can allow the user to label transient interactions and perform recovery of protein interaction partners without the need to keep

the PPIs intact, which may be useful for membrane-bound targets like cell-surface receptors.

In this study, we employed TurboID, a promiscuous proximity-labeling enzyme that covalently tags proximal lysine residues with biotin, fused to the amyloid beta (A $\beta$ ) peptide and various peptide binders for identification of cell-surface receptors and other membrane proteins that interact with the peptides in a blood-brain barrier (BBB) model. Our approach aims to identify new A $\beta$ -interacting receptors, quantify levels of new and existing A $\beta$ -protein interactions, and compare differences in interactomes across various primary cell models with distinct protein isoforms known to bind A $\beta$  differentially. *In vitro* assays demonstrated functional biotinylation activity of recombinant TurboID-fusion constructs; in contrast, initial cell-surface labeling and additional *in vitro* assays showed lack of peptide-specific biotinylation. Continued work and effort is focused on demonstration of peptide-driven specific binding with the BBB model.

### 5.3 Introduction

Protein-protein interactions (PPIs) define all cellular signaling and biological processes.<sup>70</sup> The total number of PPIs in humans is estimated to be 650,000,<sup>245</sup> and furthering our study of PPIs is central to our understanding of basic biology and disease. This fundamental research is vital for identifying new targets for therapeutic modulation (i.e., inhibition or activation of a protein and its downstream cell signaling pathway)<sup>246</sup>, and finding new therapeutic targets is a critical component of all drug discovery campaigns. As discussed in earlier chapters, cell-surface receptors and other membrane proteins are often implicated in various signaling and regulatory pathways, and mapping the whole interactome of a ligand (i.e., the set of all the molecules and proteins it interacts with) at the cell surface is important not only for identification of single therapeutic targets, but also

to understand the biological network of molecular interactions at a systems level.<sup>247</sup> Indeed, the “Zauberkegel (magic bullet)” concept of “one gene, one drug, one disease” as described by Paul Ehrlich in 1907<sup>248</sup> has been challenged in recent years to suggest that simultaneous modulation of multiple protein targets may be a more effective approach within drug development for certain diseases.<sup>249</sup>

Despite difficulties in membrane protein handling due to their hydrophobic membrane-bound structures (previously described in Chapters 2 and 3), PPI characterization directly at the living cell membrane is particularly desirable, as it allows the surface receptors and other membrane-associated proteins/biomolecules to be presented in biologically or disease-relevant environments (e.g., with co-factors, post-translational modifications, and accurate spatial organization). Current efforts toward characterizing cell-based PPIs within the field of ligand-based receptor capture (LRC) technology typically employ ligands that are modified with heterofunctional small molecules that can photo- or chemically-crosslink ligands to target receptors *in vivo* (on cells) or on tissues and identifies interactions using MS.<sup>250–255</sup> Though ligand-receptor interactions have been successfully identified with this strategy, these approaches require advanced techniques like multi-step chemical synthesis of complex small molecules<sup>253</sup> or rely on N-glycosylation of receptors for crosslinking<sup>250</sup> which limit the utility of the approach.

Yeast-two-hybrid (Y2H) screening is another commonly used approach that allows some detection of membrane-bound PPIs, but this method requires transformations of ligands and gene libraries, which also limits the versatility of this technique.<sup>256</sup> Co-IP followed by PPI characterization using MS (e.g., affinity purification MS) or WB is a more broadly applicable and simple method for identifying protein interaction partners; however, its application to identifying membrane-associated interactions is challenging due to solubilization requirements that may disrupt the structure of membrane-bound proteins.



Co-IP methods are also limited to interactions with relatively high affinities for recovery<sup>257</sup>, which may not be suited for studying physiological ligand-receptor interactions which are typically low affinity ( $K_D$  in the  $\mu\text{M}$  to  $\text{mM}$  range).<sup>180</sup>

In contrast to the previously mentioned techniques, enzyme-catalyzed proximity labeling (PL) is a transient labeling approach that allows for spatial study of PPIs.<sup>68</sup> In PL methods, a promiscuous labeling enzyme is targeted to subcellular compartments or specific proteins via a genetic fusion to a protein or peptide ligand to spatially restrict labeling.<sup>68</sup> Enzymes that are used in PL typically convert a substrate into a short-lived reactive species – like a radical<sup>258</sup> or an activated ester<sup>70</sup> – that covalently tags neighboring proteins with molecules like biotin, which can subsequently be pulled out of solution on affinity-based solid supports for identification and quantification using MS.<sup>68</sup> The radius for labeling activity is defined by the concentration of quenchers (e.g., amines) in the environment as well as the half-life of the reactive species.<sup>68</sup> As an alternative to IP and biochemical fractionation techniques, PL does not rely on high-affinity interactions between the ligand and target, and allows recovery of weaker or more transient interactions that may otherwise be lost. In addition, unlike LRC which requires complex chemical synthesis of multi-functional crosslinkers, the labeling enzymes are readily expressed and functional, with successful applications *in vitro* and *in vivo*.<sup>70,259</sup>

Examples of enzymes used in PL include BioID (based on BirA, a biotin protein ligase in *E. coli*, with a single mutation (R118G)), APEX (engineered ascorbate peroxidase), and TurboID (Table 5-1).

Enzyme	Size (kDa)	Labeling substrate	Tagging residues	Labeling time	Ref.
APEX / APEX2	28	Biotin-phenol + $\text{H}_2\text{O}_2$	Cysteine, Histidine, Tryptophan, Tyrosine	1 min	258,260,261
BioID	35	Biotin	Lysine	18 h	262

BioID2	27	Biotin	Lysine	18 h	263
TurboID	35	Biotin + ATP	Lysine	10 min	70
miniTurboID	28	Biotin + ATP	Lysine	10 min - 1 h	70

Table 5-1. Comparison of different proximity labeling enzymes.

As described in the table, the proximity ligases APEX and APEX2 require treatment of cells with H<sub>2</sub>O<sub>2</sub> in the presence of biotin-phenol (BP)<sup>258</sup>; although the incubation is short (~1 min), H<sub>2</sub>O<sub>2</sub> can still be toxic to cells. In addition, APEX labeling still requires incubation of the APEX-fusion with BP for 30 min before initiation of biotinylation with the addition of H<sub>2</sub>O<sub>2</sub>, which may not be suitable for cell-surface labeling as biomolecules are readily endocytosed at 37 °C<sup>264</sup> and the fusion construct may be internalized. Furthermore, BioID requires even longer labeling times of 18 hours at 37 °C, which will likely also result in internalization of the BioID-fusion protein before sufficient biotin labeling is allowed to occur.

In contrast, TurboID, a yeast display-engineered BirA mutant, uses biotin-adenosine monophosphate (AMP) which is nontoxic and has a fast labeling time of 10 min at 37 °C.<sup>70</sup> Upon addition of biotin and adenosine triphosphate (ATP), covalent tagging of proteins occurs within ~10 nm of the enzyme.<sup>70</sup> Its N-terminal domain truncated version, miniTurboID, is slightly smaller (28 kDa, compared to 35 kDa) and has less background labeling, but labels slower (~1 hour) and exhibits 1.5 - 2-fold less activity than TurboID.<sup>70</sup> With its superior biocompatibility over other biotin ligases, we used TurboID fusions in our study for proximity labeling at the cell surface.

Here, our goal was to use a physiologically relevant cell model to label all the receptors and membrane-associated proteins that the targeting ligand interacts with. While TurboID and other biotin ligases have been used extensively in intracellular applications and to study PPIs, to the best of our knowledge, we have observed only one

proof-of-concept study for extracellular cell-surface labeling with known ligand-receptor interactions.<sup>265</sup> In our study, we fused our proximity ligases to the amyloid beta 1-42 peptide (A $\beta$ ) for labeling in a blood-brain barrier (BBB) model. A $\beta$  is derived from the single-pass transmembrane amyloid precursor protein (APP), a highly expressed protein in the brain.<sup>266</sup> APP is cleaved by beta-secretase on the extracellular side, then by gamma-secretase on the intracellular/membrane side to produce a 40-42 amino acid peptide.<sup>71</sup> Although A $\beta$ 40/42 are the most common amyloid species, the APP, depending on the isoform, can be cleaved at several different sites, giving rise to different A $\beta$  peptide lengths.<sup>71</sup> The A $\beta$  peptide is primarily secreted by neurons and rapidly cleared into the bloodstream via the BBB<sup>267</sup>, although exact mechanisms by which A $\beta$  is cleared from the brain are not entirely understood.<sup>268</sup> While present in normal physiology, dysregulation of its clearance and the subsequent buildup of large amyloid plaques is one of the defining pathophysiological characteristics of Alzheimer's disease (AD), a neurodegenerative disease that affects 47 million people worldwide.<sup>269</sup> An increased level of A $\beta$  (100- to 200-fold higher in AD patients<sup>267</sup>) makes it susceptible to aggregation and fibrillar formation, and a dense core formed by amyloid fibrils is a primary protein component of extracellular amyloid plaques.<sup>270</sup> Plaque formation is often associated with degeneration of neurites, astrocytes, and astrocytic processes.<sup>270</sup> Given its prevalence, most of the 441 drugs in development for the treatment of AD are aimed at A $\beta$ <sup>269</sup>, although recent works have challenged the amyloid hypothesis, which states that accumulation of A $\beta$  is the main cause of AD.<sup>271</sup> Further, A $\beta$ -related drug discovery for the treatment of AD remains unfruitful, including recent controversy over the efficacy of aducanumab, an FDA-approved drug for AD that targets aggregated A $\beta$ .<sup>272</sup> Nevertheless, A $\beta$  remains a highly interesting therapeutic target given its physiological importance in disease.

In this study, our aim was to fuse A $\beta$  to TurboID to pan against the cell surface of a BBB cell model for biotin-labeling and perform subsequent pull-down onto streptavidin beads to characterize and quantify the A $\beta$  cell-surface interactome using MS (Figure 5-1). Identification of proteins that interact with A $\beta$  could not only provide insight into potential novel drug targets but also help deepen our understanding of the role of A $\beta$  and A $\beta$  dysregulation in AD.

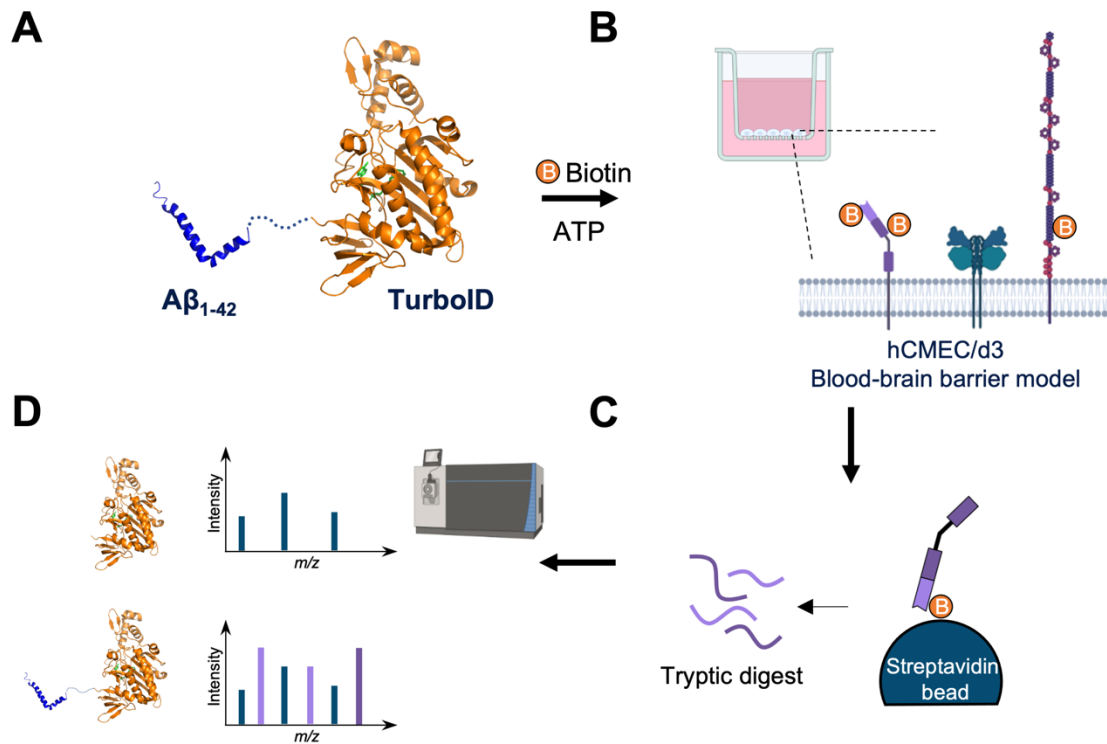


Figure 5-1. Workflow for a TurboAb panning assay.

(A) TurboID construct fused to targeting peptide, A $\beta$ , is added (negative control has no fusion) to (B) BBB transwell model with biotin and ATP. (C) Subsequent lysis and pull-down onto streptavidin beads separates labeled proteins, followed by on-bead tryptic digest, and (D) analysis by MS and comparison to negative control allows identification of A $\beta$ -specific interacting receptors and other membrane-associated proteins.

## 5.4 Results

### 5.4.1 Design of A $\beta$ peptide-TurboID fusion constructs

Initial design considerations for the A $\beta$ -TurboID fusions (TurboA $\beta$ ) involved substitution mutations of A $\beta$  to create a more soluble peptide given its propensity to

aggregate in aqueous environments. Recombinant expression is highly challenging as A $\beta$  begins to aggregate once it is refolded from insoluble fractions and/or cleaved from solubility tags.<sup>273,274</sup> Similar to previous studies that have utilized fusions to highly soluble proteins like maltose-binding protein (MBP)<sup>275</sup> and intestinal fatty acid binding protein (IFABP)<sup>273</sup> to express A $\beta$ , TurboID, engineered from a well-expressing, native *E. coli* biotin ligase protein, may also act as a solubility tag for A $\beta$ . Here, our goal was to make a non-aggregating (i.e., less amyloidogenic) variant of A $\beta$  to accommodate higher solubility during not only expression and purification but also longer incubation times at elevated (37 °C) temperatures for performing cell labeling assays. Notably, in familial AD (FAD), different single-point mutations lead to drastically different disease phenotypes and/or neuropathological effects (e.g., Dutch E22Q mutation, which increases aggregation and causes severe cerebral amyloid angiopathy); thus, it is possible that any mutations we introduce may also affect subsequent functional activity like cell-membrane binding.<sup>276</sup> However, given potential challenges with solubility, mutations may still be necessary to minimize aggregation and enable a functional assay. To this end, we sought to mutate residues within <sup>17</sup>LVFFA<sup>21</sup> known as the central hydrophobic cluster, which determines the rate of monomeric assembly and forms beta sheet hairpins in fibrils.<sup>277</sup> This cluster, with an additional two amino acids on either side (KK-<sup>17</sup>LVFFA<sup>21</sup>-ED), was sufficient to form insoluble amyloid-like structures.<sup>278</sup> Previous engineering approaches for soluble expression and characterization of recombinantly expressed A $\beta$  suggested that mutating the hydrophobic positions, particularly <sup>19</sup>FF<sup>20</sup> (which can act as nucleation sites due to pi-pi stacking), reduced aggregation.<sup>279–281</sup> Based on the study by Park *et al.*,<sup>281</sup> we chose a single point mutation, F20P, which demonstrated reduced aggregation *in vivo*, for our A $\beta$  fusion constructs (Figure 5-2A).

For the interdomain linker design, previous work with A $\beta$  or TurboID fusions used a range of amino acid lengths, from a single restriction site (corresponding to two amino acids)<sup>274</sup> to 14 amino acids.<sup>282</sup> In a seminal study by Branon *et al.*,<sup>70</sup> glycine-serine linkers of 4 amino acids or 18 amino acids were used to connect targeting proteins to TurboID. Based on these works, we reasoned that a short, flexible linker should suffice, as too much separation could cause aggregation of the A $\beta$  peptides within the fusion constructs, and the TurboID must remain in close proximity to achieve better spatial control of labeling. Through oligonucleotide annealing and overlap PCR, A $\beta_{WT}$  and A $\beta_{F20P}$  were cloned, and subsequently digested and ligated into plasmids containing C-terminal TurboID-6xHis fusions (Figure 5-2C).

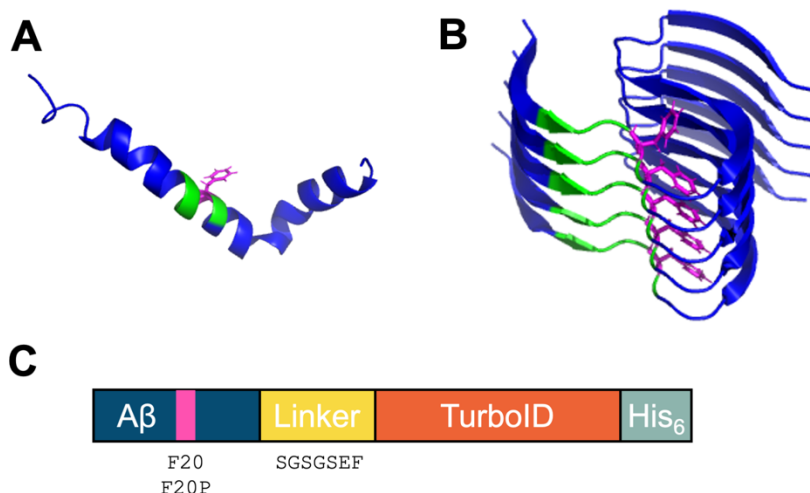


Figure 5-2. Amyloid beta structures.

(A) Crystal structure of monomeric amyloid beta, with hydrophobic cluster in a different color and Phe20 shown in stick form (PDB: 1IYT).<sup>283</sup> (B) Angled top-down view of amyloid beta fibrillar organization, driven by hydrophobic interactions with Phe20 central to the hydrophobic cluster (PDB: 2MXU)<sup>284</sup>. (C) Schematic of our recombinant protein design, including A $\beta$  peptide, linker, TurboID, and 6xHis tag for purification. “EF” within the linker corresponds to an EcoRI restriction site.

#### 5.4.2 Recombinant expression and purification of TurboA $\beta_{WT}$ and TurboA $\beta_{F20P}$ fusion constructs

Mild expression conditions (0.5 mM IPTG, growth at 15 °C overnight, TB culture media) were used for the TurboA $\beta$  constructs to minimize aggregation and limit formation of inclusion bodies.<sup>285</sup> For lysis, we initially used solution-based methods with lysozyme, an enzyme that breaks down bacterial cell walls by peptidoglycan cleavage,<sup>286</sup> and B-PER, a commercially available detergent-based bacterial protein extraction reagent, to avoid potential protein denaturation and aggregation which can occur from localized heat that may be generated during physical lysis. However, SDS-PAGE analysis of soluble and insoluble fractions showed most of the protein in the insoluble fraction (Figure 5-3A). In contrast, intermittent sonication with a probe sonicator on ice led to efficient lysis and more protein in the soluble fraction; thus, we used this lysis method moving forward.

TurboA $\beta$  proteins were purified using nickel-nitrilotriacetic acid (Ni<sup>2+</sup>-NTA) immobilized metal affinity chromatography (IMAC) under gravity flow, a commonly used method for purifying proteins with polyhistidine tags.<sup>287</sup> All steps were performed at 4 °C to help increase the stability of the protein and minimize aggregation. Buffer compositions were adapted from literature<sup>288</sup>, with the addition of 10% glycerol in later experiments to further minimize protein aggregation and help retain the enzymatic activity of TurboID. SDS-PAGE and ImageJ analysis showed a relatively clean elution fraction (72%, 79%, and 93% purity for TurboA $\beta$ <sub>WT</sub>, TurboA $\beta$ <sub>F20P</sub>, and no-fusion TurboID, respectively, calculated from elution fraction 2) (Figure 5-3B). The proteins were rebuffed using diafiltration centrifugal concentrators to remove the imidazole present in the elution buffer. As expected, small precipitates were visible during the centrifugation steps for all constructs, most notably for TurboA $\beta$ <sub>WT</sub>; however, absorbance reads at 280 nm showed sufficient protein yields within the soluble fraction.

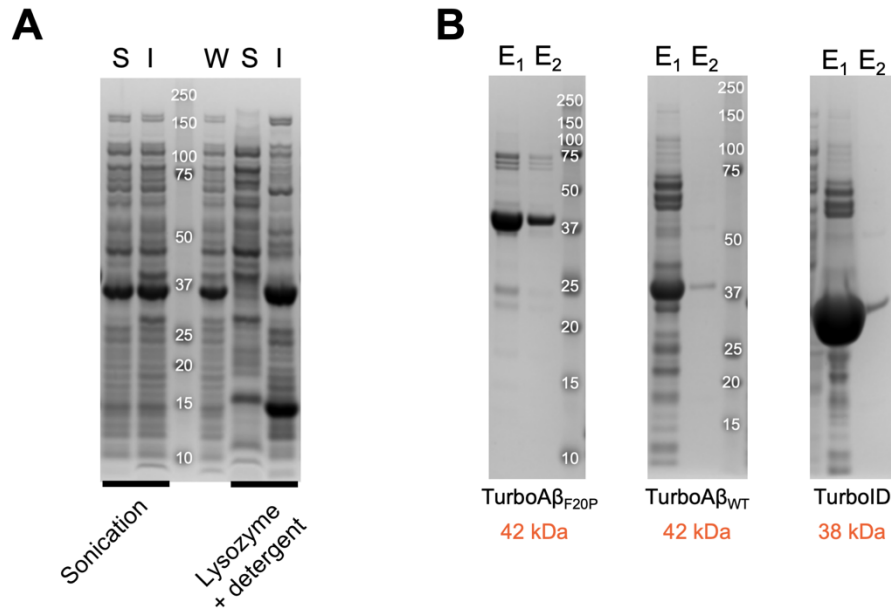


Figure 5-3. TurboA $\beta$  expression and purifications.

(A) Comparison of solution-based and physical lysis techniques for protein extraction. Detergent/lysozyme lysis (right) showed much of the protein in the insoluble fraction (I), while probe sonication (left) was much more efficient at extracting protein to the soluble fraction (S) (TurboA $\beta$ <sub>WT</sub> shown, comparison to whole-cell sample post-induction (W)). (B) Comparison of protein elution fractions 1 and 2 (E<sub>1</sub> and E<sub>2</sub>) of different constructs by SDS-PAGE. TurboID is run on same gel as TurboA $\beta$ <sub>WT</sub> (E<sub>1</sub> was overloaded due to unexpectedly high yield compared to fusions). ImageJ was used to estimate purity in elution fraction 2.

#### 5.4.3 Protein characterization assays show functional activity of TurboID and some aggregation of TurboA $\beta$ constructs

To assess the functionality (i.e., biotinylation activity) of TurboID within the TurboA $\beta$  construct, we incubated the TurboA $\beta$  construct with a non-target protein, sfGFP, with or without biotin and ATP for 10 min at 37 °C. Following incubation, the biotinylation activity was analyzed using WB with streptavidin labeled with Alexa Fluor 555. As expected, negative controls with no biotin/ATP had negligible labeling for both TurboA $\beta$  constructs, while conditions with biotin/ATP had strong signals (Figure 5-4). Notably, much of the biotinylation came from self-biotinylation, although faint sfGFP bands indicated that there was some non-specific intramolecular labeling activity. As self-labeling can be



distinguished from target labeling for MS analysis, we reasoned that self-biotinylation may not be a concern in our assay. Comparison of labeling activity between TurboA $\beta_{WT}$  and TurboA $\beta_{F20P}$  showed increased self-labeling for TurboA $\beta_{F20P}$  which may indicate decreased aggregation compared to TurboA $\beta_{WT}$ . However, the level of sfGFP labeling was similar for both TurboA $\beta$  constructs.

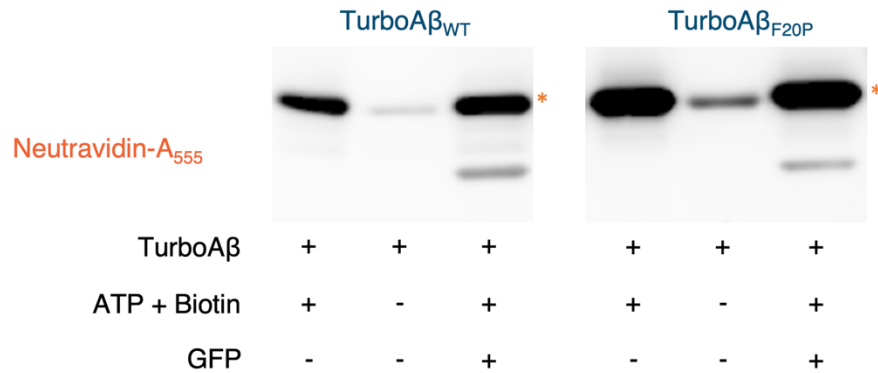


Figure 5-4. In vitro biotinylation assay and analysis by WB. Constructs were allowed to interact for 2 hours at 4 °C, then labeled for 10 min at 37 °C with the addition of ATP and biotin. Self-labeling is indicated with an asterisk (\*), and accounts for much of the biotinylation signal.

#### 5.4.4 Native PAGE assay shows aggregation of TurboA $\beta_{WT}$ and TurboA $\beta_{F20P}$ after 37 °C incubation

To better understand the aggregation propensity of the TurboA $\beta$  constructs, we incubated constructs at 37 °C for 0 - 3 hours and assessed the migration pattern by native PAGE. Unlike SDS-PAGE in which proteins are typically denatured into elongated polypeptide chains, native PAGE allows protein complexes to remain in oligomeric states; aggregates will migrate slower than monomers, and the difference can be easily visualized on a gel.<sup>289,290</sup> Compared to TurboA $\beta_{F20P}$ , TurboA $\beta_{WT}$  seemed to aggregate more quickly (< 1 hour) at 37 °C. TurboA $\beta_{F20P}$  still aggregated, but at a much slower rate (> 2 - 3 hours) based on the weaker band intensity of the monomeric protein at longer incubation times (Figure 5-5). As expected, the control (non-fused TurboID) in the middle lane did not

aggregate after 3 hours at 37 °C. The results of this assay suggested that the F20P mutant introduced into the A $\beta$  peptide substantially reduces aggregation of the A $\beta$  peptide.

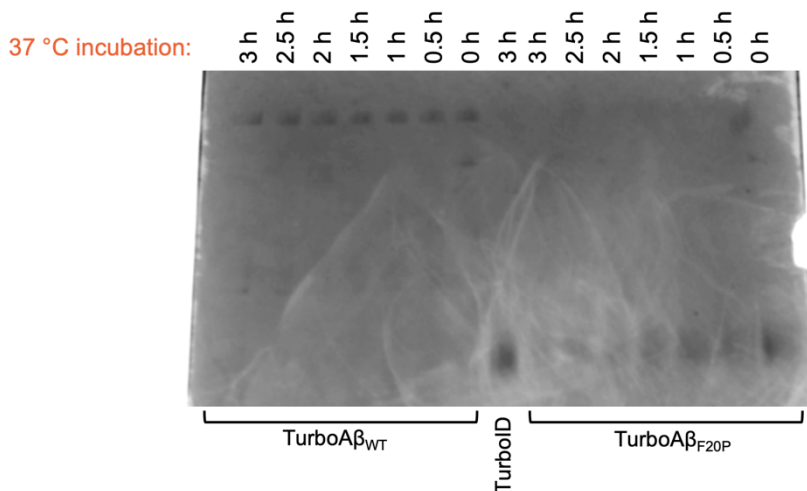


Figure 5-5. Native PAGE analysis of protein aggregates. TurboA $\beta_{WT}$  (left) incubations show most of the product in higher-order oligomers compared to TurboA $\beta_{F20P}$  (right). TurboID (not fused, center) remains monomeric even after 3 hours at 37 °C.

#### 5.4.5 Characterization of known A $\beta$ binders in hCMEC/d3 model system and panning of TurboA $\beta$ constructs

The hCMEC/d3 brain microvascular endothelial cell line has been widely used as a model for human blood-brain function.<sup>291</sup> Further, the model has also been used to study the effects of A $\beta$  on brain microvasculature, particularly its effect on basolateral to apical transport.<sup>292</sup> Thus, we chose to use the hCMEC/d3 BBB model for cell-surface panning experiments with TurboA $\beta$ . Prior to performing cell panning, we confirmed expression of known A $\beta$ -binding receptors, low-density lipoprotein receptor-related protein-1 (LRP1)<sup>293</sup> and receptor for advanced glycation end products (RAGE)<sup>294</sup>, on the cell surface. LRP1 and RAGE, two of many receptors implicated in A $\beta$  transport, play opposite roles: LRP1 mediates A $\beta$  clearance from the brain into the bloodstream, while RAGE mediates A $\beta$  transport from the bloodstream into the brain.<sup>294</sup> For our cell model, hCMEC/d3 cells were

cultured on cell culture inserts and immunostained for RAGE and LRP1 over several days and imaged to assess the availability of the receptors on the cell surface. Confocal z-stack images confirmed expression of LRP1 and RAGE on the luminal surface which would be accessible to panned constructs (Figure 5-6). As expected, longer incubations resulted in multiple cell layers as hCMEC/d3 cells are not contact inhibited<sup>291</sup>; thus, we moved forward with low density seeding and growing the cells to confluence (4 - 5 days) for cell panning experiments.

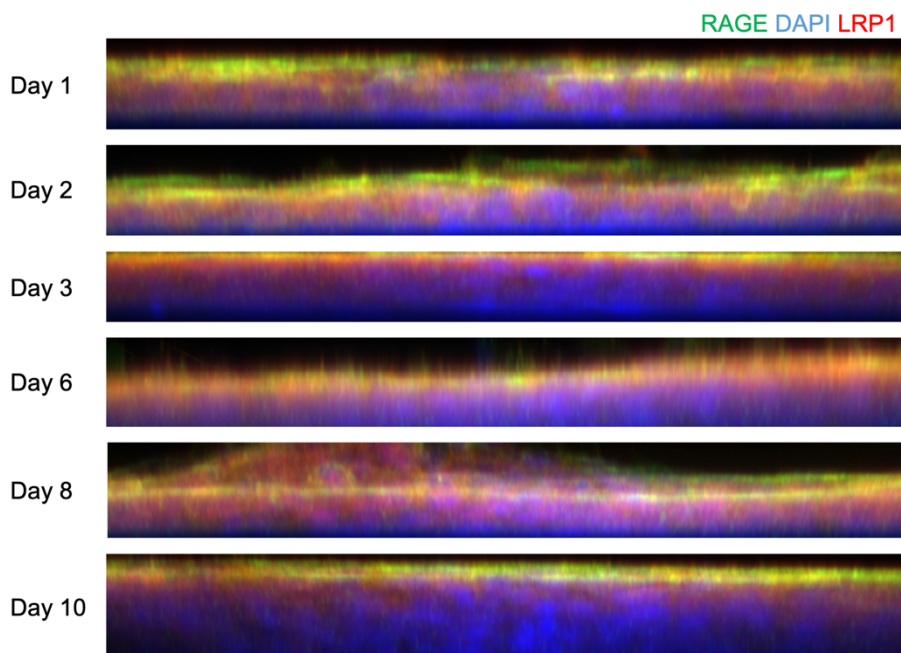


Figure 5-6. Z-stack confocal images for surface receptor characterization on BBB model. Cells were stained for RAGE, LRP1, and the nuclear marker DAPI over several days to assess the formation of a uniform cell barrier and maximal availability of target receptors on the cell surface (as determined by homogeneous RAGE and LRP1 expression, seen around Day 6) prior to overconfluence, at which point multiple cell layers had formed (as given by DAPI staining at multiple layers and/or at the same layer as target, as observed on Days 6, 8, and 10). Based on this study, we moved forward with conditions slightly before Day 6, at Day 5. (Images courtesy of Paulina Eberts)

With the validated BBB cell model, we then performed cell surface panning experiments with the TurboA $\beta$  constructs. For our initial approach, we pre-incubated TurboA $\beta$  with the cells at 4 °C to allow the TurboA $\beta$ -receptor binding to reach equilibrium,

then replaced the medium with prewarmed medium containing ATP, biotin, and additional TurboA $\beta$  (at the same concentration). Unfortunately, cell imaging did not show significant biotinylation by either TurboA $\beta$  or TurboID constructs even at prolonged incubation times (Figure 5-7). Some extended incubations (e.g., 18 hours) showed significant internalization as indicated by large puncti (Figure 5-7, '18 hours'). We also observed moderate levels of endogenous biotin in cell imaging experiments (Figure 5-7, '10 min').

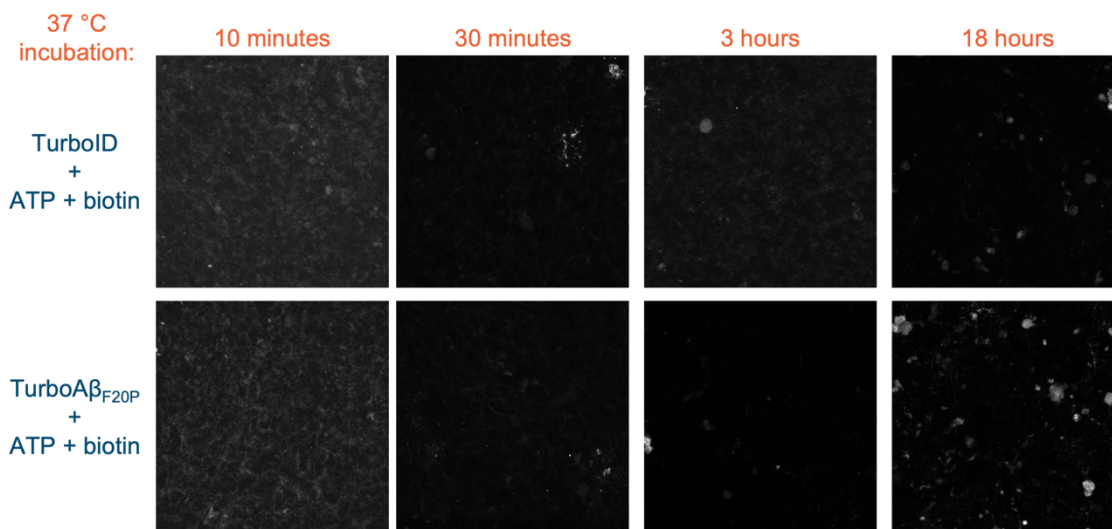


Figure 5-7. Cell images showing treatment of hCMEC/d3 with Turbo constructs. TurboA $\beta$  (F20P shown here, but WT yields similar results) and staining with Neutravidin-Alexa 555 with various incubation times at 37 °C. Comparison to non-targeting (no A $\beta$  fusion) control shows lack of binding with TurboA $\beta$ . (Images courtesy of Paulina Eberts)

#### 5.4.6 In vitro assays for assessing TurboA $\beta$ functionality show lack of A $\beta$ -specific labeling

As cell-surface panning assays showed no specific TurboA $\beta$  labeling, we further investigated TurboA $\beta$  binding in a cleaner, cell-free system to identify target vs. non-target binding. For this *in vitro* assay, the target, RAGE-Fc chimera, or no target (for negative controls) was immobilized on wells overnight, blocked, and construct was added at 4 °C. Similar to the cell assay, the wells were pre-incubated with TurboA $\beta$ <sub>WT</sub>, TurboA $\beta$ <sub>F20P</sub>, and TurboID at 4 °C to allow the binding to reach equilibrium, and the solution was replaced with the same pre-warmed panning solution at 37 °C before being analyzed using

streptavidin-horseradish peroxidase (HRP) and enhanced chemiluminescence (ECL) detection. Comparing the TurboA $\beta$ <sub>WT</sub>, TurboA $\beta$ <sub>F20P</sub>, TurboID, and control panning against RAGE-immobilized wells, we initially observed what appeared to be A $\beta$ -RAGE driven, concentration-dependent biotinylation (Figure 5-8A). However, upon repeating the experiment with a negative control in which RAGE-Fc chimera was not immobilized, biotinylation was present in all samples and increased with higher concentrations of TurboA $\beta$ , indicating lack of A $\beta$ -specific labeling (Figure 5-8B). Repeated experiments with LRP1 showed similar trends with nonspecific binding, suggesting failure of this approach to detect specific A $\beta$ -driven labeling.

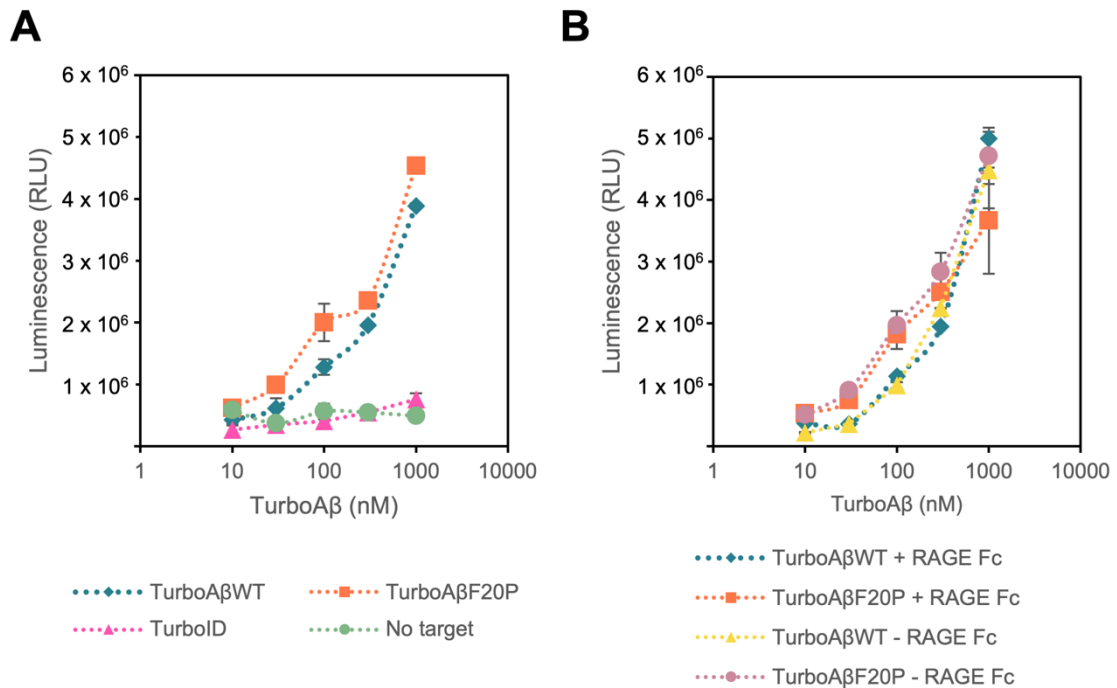


Figure 5-8. Biotin labeling activity of TurboID constructs in *in vitro* assays. RAGE-Fc immobilized well-panning initially suggested (A) A $\beta$ -RAGE driven, TurboA $\beta$  concentration-dependent biotinylation, but (B) an additional negative control experiment showed lack of A $\beta$ -specific binding and biotinylation. (Data courtesy of Paulina Eberts)

#### 5.4.7 Current efforts toward creating new targeting peptide fusions

Moving away from A $\beta$ -targeted binding and biotinylation, we designed and synthesized new peptides for fusion to TurboID. With this new repertoire of peptides, our goal was to target receptors that were highly expressed at the BBB in hCMEC/d3 cells with peptides that were less prone to aggregation than A $\beta$ . To accomplish this, we cloned peptides that target transferrin receptor (TfR) and integrins, which have been shown to have higher surface expression than the known A $\beta$ -binding receptors LRP1 and RAGE (Table 5-2).<sup>295</sup>

Peptide	Target	AA sequence	Ref.
Angiopep-2 (AP-2)	LRP1	TFYFGGSRGKRNNFKTEEY	296
GRGDSP	Integrin	GRGDSP	297
Trivalent GRGDSP	Integrin	(See Table 5-3)	297
T7	Transferrin	HAIYPRH	298
Bivalent T7	Transferrin	(See Table 5-3)	298
T12	Transferrin	THRPPMWSPVWP	298
A $\beta$ <sub>1-28</sub>	LRP1, RAGE, etc.	DAEFRHDSGYEVHHQKLVFFAEDVGSNK	71,299

Table 5-2. Repertoire of new peptide-TurboID fusions.

We also included a LRP-1 targeting peptide which could easily be tested using our established *in vitro* biotinylation assay as well as a truncated version of A $\beta$ <sub>WT</sub> (A $\beta$ <sub>1-28</sub>). The membrane-bound portion of A $\beta$  (within the APP, prior to cleavage by gamma-secretase) is known to be highly toxic and form fibrillar aggregates, and truncation of C-terminal residues has been shown to improve monomeric state of A $\beta$ , which motivated our choice of the truncated A $\beta$  peptide.<sup>71,300</sup> The new peptide-TurboID fusion designs also included a slightly longer interdomain linker based on the long interdomain linker (14 amino acids) used in the original TurboID work (Table 5-3).<sup>70</sup> A C-terminal HA tag was also added to allow for triple-staining of the target, biotin, and TurboID fusion construct; during immunostaining, the colocalization of target, biotin, and HA signal would indicate specific target-ligand binding and subsequent biotinylation activity. Finally, multivalent (bivalent or

trivalent) constructs were also created for peptides with low monovalent affinity to utilize higher avidity effects achieved by collective binding (as described in Chapter 4).<sup>301</sup>

The constructs were recombinantly expressed and purified as before, and subject to the well-based *in vitro* biotinylation assay as described in the previous section (subsection 5.4.6). Since the biotinylation signal in the previous TurboA $\beta$  binding assays predominantly came from non-target binding, we tested three common blocking agents (Ultrablock, Synblock, and 1% BSA) to solely test the background signal (i.e., adsorption to the well surface due to nonspecific binding). To this end, we blocked the wells without any target immobilization, and detected luminescence signal after incubation with the panning solution (peptide-TurboID, biotin, ATP). Unfortunately, overall background signal remained high for many of the constructs, although some were considerably higher than others (e.g., AP-2 vs. GRGDSP) (Figure 5-9). Surprisingly, blocking with 1% BSA solution resulted in the lowest background signal, despite the presence of many lysine residues in BSA and presumably bulkier molecular weight. Continued work will focus on optimizing other reaction conditions to increase reaction rate and lower nonspecific binding, as discussed in the following section.

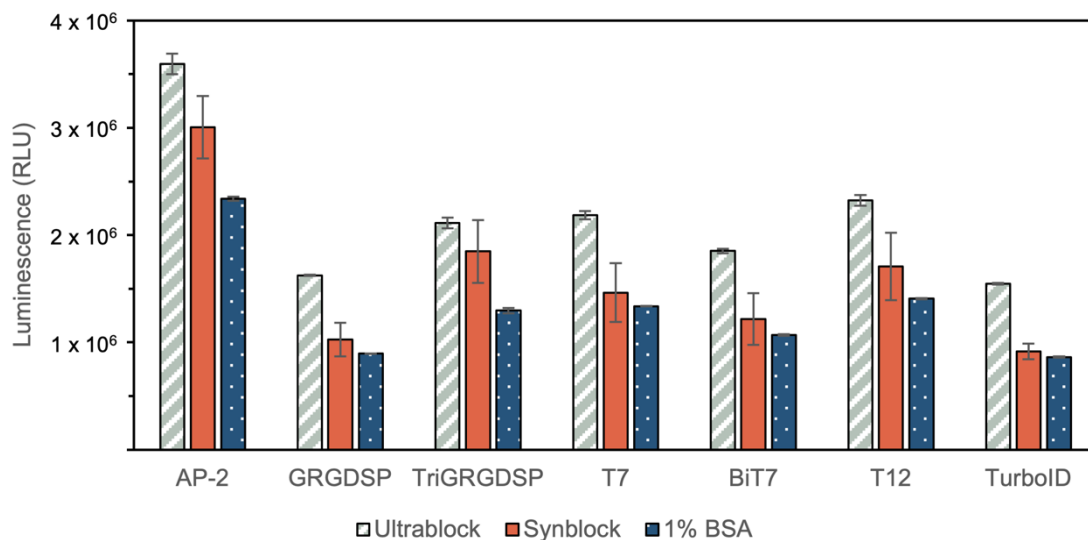


Figure 5-9. *In vitro* assay to assess nonspecific binding across different blocking conditions.

High nonspecific biotinylation observed with new peptide-TurboID fusion constructs in the *in vitro* assay. As no target is immobilized, all signal is due to non-specific binding and subsequent biotinylation. ( $A\beta_{1-28}$ , the newest peptide-fusion construct, was not included in this assay)

## 5.5 Discussion

Identification and quantification of protein-protein interactions at the cell surface is critical to discovery of new therapeutic targets and furthering our understanding of comprehensive cell-surface interactions given the essential roles that ligand-receptor interactions play in regulating processes like transport of biomolecules and responding to external stimuli in a highly dynamic environment.<sup>302</sup> Classical affinity purification-based techniques for isolating PPIs are unsuitable for cell-surface proteins due to dissociation of protein architecture outside of the membrane and overall low affinity of cell-surface interactions.<sup>257</sup> Considering these challenges, our goal in this study was to adapt a peptide-targeted proximity biotinylation for covalent tagging of membrane proteins directly on the cell surface in a disease-relevant model.

Peroxidases and biotin ligases are useful tools for covalent tagging in PL applications and have been used in a wide range of applications.<sup>68,303</sup> While we used TurboID, a biotin ligase, in our study due to nontoxicity of the required biotin derivative, it is possible that the longer labeling times required by TurboID lead to less control, resulting in higher nonspecific binding as observed in our second peptide-TurboID *in vitro* assay. High concentrations of TurboID fusions have also demonstrated higher nonspecific binding *in vivo*.<sup>304</sup> “Split” versions of PL enzymes like Split-BioID<sup>305</sup> and Split-TurboID<sup>306</sup> aim to address these shortcomings by dividing the enzymes into two fragments and requiring the complex formation of the enzyme to activate the labeling activity for better spatiotemporal control. Although the nonspecific adsorption of  $A\beta$  still remains a challenge, a split-biotin ligase could be a possible alternative to our approach with the peptide-



TurboID assays, especially for achieving higher specificity within a complex environment such as the cell surface.

In addition, TurboID is a much larger protein (~35 kDa) compared to the fusion partners that we used in this study (~1 - 4 kDa). Indeed, larger proximity ligases have been shown to increase off-target labeling in some applications; this limitation has led to additional engineering of proximity ligases, prompting the development of BioID2<sup>263</sup>, miniTurboID<sup>70</sup>, microID<sup>307</sup>, and ultraID<sup>307</sup> which are slightly smaller in size compared to their predecessors. In our study, we initially created A $\beta$  fusions with miniTurboID for testing in parallel with TurboID. Due to the slower biotinylation activity of miniTurboID compared to TurboID (1 hour vs. 10 min), we opted to move forward with TurboID over concerns of internalization at prolonged incubation times especially at 37 °C. However, since we were unable to achieve labeling at 10 min and most likely require longer incubation times for surface labeling in our cell model, fusion to smaller biotin ligases may allow for better spatial control and faster diffusion for enhanced labeling activity.

Data from our cell panning assay unexpectedly showed high levels of endogenous biotin, which can confound the interactions identified by TurboID biotinylation. Approaches for separating membrane and cytosolic fractions using Mem-PER, a commercially available mammalian protein extraction reagent, were unsuccessful as it combined plasmid and organelle membranes and we were unable to see signal difference between membrane and cytosolic fractions. Endogenously biotinylated enzymes like pyruvate carboxylase, which resides in the mitochondrial membrane<sup>308</sup>, can also be recovered, although MS analysis should distinguish endogenous biotinylation more accurately over methods like WB. Also, PL using APEX and APEX2 with biotin-phenol have shown that this biotin-derivative is generally membrane-impermeable and could be an alternative to TurboID for surface-specific labeling.<sup>261</sup>

Despite engineering of the A $\beta$  peptide for improved solubility, A $\beta$ -targeted biotinylation was not successful in our approach. As our goal was to identify A $\beta$  binders in a BBB model, we initially chose to introduce a single mutation such that it closely resembled the native A $\beta_{1-42}$  peptide. We also avoided introducing substitutions at known mutated positions in familial subtypes of AD to prevent bias toward certain types of FAD.<sup>309</sup> The F20P mutation, supported by *in vivo* data for reducing A $\beta$  plaque deposits<sup>281</sup> was used, as proline is known to be a potent breaker of beta-sheet structures, which form the core of A $\beta$  oligomers and fibrils.<sup>310</sup> A possible explanation for lack of specific binding with our TurboA $\beta$  constructs is that the TurboA $\beta_{WT}$  aggregates too fast (as supported by the native PAGE data) to meaningfully participate in A $\beta$ -targeted binding, while the TurboA $\beta_{F20P}$ , while soluble, is not able to bind functionally due to the F20P mutation. Further analytical techniques like size exclusion chromatography (SEC), which separates protein by size, can be useful for assessing the aggregation and monomeric state of TurboA $\beta$  constructs.<sup>311,312</sup> Altogether, the data suggest that recombinantly expressed A $\beta_{1-42}$  may not be a suitable choice for use in proximity labeling due to poor solubility and tendency to adsorb nonspecifically, and significant engineering efforts may be required to retain A $\beta$ -mediated targeting ability.

Due to the challenges with A $\beta$ -specific binding, new peptide-TurboID fusion constructs were created to demonstrate feasibility of the TurboID-BBB model labeling approach with well-behaved peptides and highly expressed targets. Again, the results showed high nonspecific biotinylation and most recent data (not shown) demonstrated lack of target-specific signal for the AP-2-TurboID target construct tested thus far. Although the blocking agents used (e.g., Synblock and Ultrablock) are suitable for traditional WB applications and claim to be protein-free, possible alternatives for better background reduction may be to use guaranteed biotin- and serum-free alternatives such

as SuperBlock and StartingBlock (Thermo Fisher Scientific). The lack of binding may also be due to low affinity and/or specificity, which is common for peptides as they are intrinsically disordered and can adopt more diverse conformations compared to proteins.<sup>313</sup> While larger protein constructs could afford us increased affinity and specificity, a recent study by Alwash and Gariépy<sup>265</sup> in which they fused affibodies to BirA\* (i.e., BioID) for cell-surface labeling stated that conjugation of BirA\* to larger proteins like antibodies and nanobodies did not retain binding nor biotinylation activity. Here, they hypothesized that the close proximity and poor solubility caused insufficient labeling and aggregation, respectively. Notably, this recent work, to the best of our knowledge, is the only publication that has shown recombinant expression of a biotin ligase fusion and subsequent cell-surface labeling, which may allude to the complexity of our approach.

Finally, an initial goal of our TurboA $\beta$  cell-surface labeling strategy was to pan against engineered primary cell models with different genotypes, particularly that of apolipoprotein E (APOE), to compare the binding profile of A $\beta$  in various FAD models. While difficulties with A $\beta$  expression have challenged our project goals, continued work in characterizing PPIs at the cell surface with other disease-relevant binders like neurotransmitters and tau<sup>269</sup> may greatly contribute to the development of new therapeutics and our understanding of biological networks in AD.

## **5.6 Conclusion**

In conclusion, the TurboID-fusion cell surface labeling method that we are currently developing has the potential to contribute to the challenging yet highly important field of protein-protein identification at the cell membrane. Covalent, ligand-driven tagging of receptors and other membrane-bound proteins can be a more accessible, modular strategy to PPI characterization compared to previous crosslinking approaches. While

TurboA $\beta$  surface labeling on our hCMEC/d3 BBB model surface ultimately failed, the progress made toward establishing TurboID fusion design along with expression and purification, *in vitro* and *in vivo* assays for assessing peptide-specific binding, and experimental optimizations for reaction conditions are useful for continued work in the field of PL at the cell surface.

## 5.7 Materials and Methods

### 5.7.1 Molecular cloning of TurboID-fusion constructs

All oligos and gBlocks were purchased from IDT, and TurboID-His6\_pET21a (#107177) and miniTurboID-His6\_pET21a (#107178) plasmids were purchased from Addgene.

Amyloid beta peptide constructs were cloned using oligo annealing-extension steps and overlap PCR. ssDNA oligos (primers) were designed to split the 42-amino acid peptide into ~21 amino acid halves with 5'-NdeI and 3'-BamHI sites for subcloning and a short interdomain glycine-serine linker between the N-terminal A $\beta$  peptide and C-terminal TurboID (Table 5-3). For oligo annealing, approximately 0.75  $\mu$ g of a 5'-forward and 3'-reverse primer with ~18-20 overlap (reverse complementary) regions were mixed with 1  $\mu$ L dNTPs (10 mM each) (NEB), 8  $\mu$ L 5X HF Buffer (NEB), and nuclease-free water (IDT) up to 40  $\mu$ L. In a thermal cycler (Bio-Rad), the solution was heated to 98  $^{\circ}$ C for 5 min, then cooled down to 80  $^{\circ}$ C for 5 min; halfway through the 80  $^{\circ}$ C incubation, a solution with 0.5  $\mu$ L Phusion Hot Start Flex polymerase (NEB), 2  $\mu$ L 5X HF Buffer, and 7.5  $\mu$ L nuclease-free water was added to the tube. Next, the thermal cycler was cooled to 72  $^{\circ}$ C to allow extension of the primers to create the short dsDNA oligo. To increase the yield, this reaction was cycled 3 - 5 times. To stitch the short dsDNA oligos together, 2.5  $\mu$ L of the primer extensions were mixed together with 10  $\mu$ L 5X HF Buffer, 1  $\mu$ L dNTPs, 0.5  $\mu$ L

Phusion Hot Start Flex polymerase, and 33.5  $\mu$ L nuclease-free water. The cycling parameters were as follows: 1x: 98 °C for 60 sec, 18x: 98 °C for 15 sec, 63 °C for 20 sec, 72 °C for 20 sec, then 1x: 72 °C for 5 min, then hold indefinitely at 4 °C. Reactions were purified using QIAquick PCR purification columns.

For the shorter peptides A $\beta$ <sub>1-28</sub>, T7, T12, GRGDSP, and AP-2 (Table 5-2), oligo annealing-extension steps were used (no overlap PCR). Additional peptide-encoding genes were purified using the nucleotide removal kit (for products shorter than 40 bp) or the PCR purification kit (for products greater than 70 bp). In addition, the TurboID genes within the pET21a vector were modified (amplified with outer primers and subcloned back into pET21a) to include a longer 5'-flexible linker (upstream of TurboID, downstream of restriction site) and 3'-HA tag.

All peptide-encoding genes and gBlocks were double digested using the TypeIIIP restriction enzymes NdeI and BamHI (NEB), column purified, and ligated into pET21a vectors as N-terminal fusions to TurboID or miniTurboID. Ligation reactions were transformed into XL1-Blue *E. coli* (Agilent), and individual colonies were picked, plasmid prepped using the QIAprep spin miniprep kit (Qiagen), and sequence verified using Sanger sequencing.

#### *5.7.2 Expression and purification of TurboID/miniTurboID fusion constructs*

*E. coli* BL21 (DE3) cells were transformed with plasmids synthesized in Section 5.7.1 and grown overnight on LB/carbenicillin (carb) agar plates at 37 °C. Individual colonies were picked and grown in 25 mL of LB/carb in a microbial shaker set at 135 rpm at 37 °C overnight. The next day, the culture was diluted in 100 - 200 mL of Terrific Broth (TB) (BD Difco) with carb to an OD<sub>600</sub> = ~ 0.05 and grown for 2 - 3 hours at 37 °C at 135 rpm in 500 - 1000 mL culture flasks. At OD<sub>600</sub> = 0.5, flasks containing culture were transferred to the cold room / on ice for 10 min before expression was induced with 0.5

mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). The cells were grown overnight (15 - 17 hours) at 15 °C at 135 rpm. Cells were harvested at 4,000 x g for 25 min at 4 °C, and pellets were resuspended in 10 - 15 mL total of lysis buffer (20 mM Tris-HCl, pH 8, 200 mM NaCl, 1 mM imidazole, 10 mM beta-mercaptoethanol (BME)) supplemented with EDTA-free protease inhibitor cocktail (Roche). The mixture was probe sonicated for 2 min total (5 sec on/5 sec off for 60 sec total, 60 sec off, then repeat) (Sonic Dismembrator Model 100, Fisher Scientific) on ice. The lysate was clarified by centrifugation at 20,000 x g for 35 min at 4 °C, and syringe filtered with a 0.45  $\mu$ m filter (Millipore-Sigma).

For IMAC purification, the supernatant was loaded onto a gravity column with 0.5 - 1 mL Ni-NTA resin (Qiagen) and washed with 15 - 20 column volumes (CVs) of Wash buffer (lysis buffer with 15 mM imidazole). Initially, 0.9 CV of Elution buffer (lysis buffer with 250 mM imidazole) was applied to the column and let drip; protein was then collected with an additional 1.5 CVs of Elution buffer. The eluate was diluted 20-fold in overnight storage buffer (20 mM Tris-HCl, pH 8, 200 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), pH 7) and stored at 4 °C overnight. The next day, the protein was rebuffed using 30 kDa MWCO diafiltration columns (Amicon or Pierce) into a storage buffer (overnight storage buffer with 10% glycerol). After diluting the original column eluate at least ~500-fold into the storage buffer, the dead volume was transferred to a 1.5 mL tube and centrifuged at 15,000 x g for 10 min at 4 °C. The supernatant (some precipitants are typically visible) was transferred to a new 1.5 mL tube and quantified with absorbance at 280 nm (Bio-Tek). Proteins were diluted to 100 - 200  $\mu$ M in storage buffer, aliquoted in 1.5 mL tubes, flash-frozen in liquid nitrogen, and stored in -80 °C as single-use aliquots.

For analytical SDS-PAGE, whole-cell and column eluates were prepared with 1X LDS (Thermo Fisher Scientific), 50 mM DTT, and Milli-Q (MQ) water. Insoluble fractions were prepared with an additional 1% SDS for better solubilization. Reactions were

incubated at 70 °C for 30 - 90 min before being run on 4-12% NuPAGE Bis-Tris gel (Thermo Fisher Scientific) with 1X MOPS or 1X MES (Thermo Fisher Scientific), stained with SimplyBlue SafeStain (Thermo Fisher Scientific), and imaged on a Gel Doc EZ system (Bio-Rad).

#### 5.7.3 In vitro biotinylation test for assessing intramolecular labeling activity

For initial tests of labeling activity, 2.4  $\mu\text{M}$  TurboA $\beta_{\text{WT}}$  or TurboA $\beta_{\text{F20P}}$  in 1X DPBS (Gibco) was incubated with or without recombinant sfGFP (expressed and purified as described in 3.7.2) on ice for 2 hours in 50  $\mu\text{L}$  in 0.2 mL PCR tubes. Tubes were moved to RT for 10 min, and 50  $\mu\text{M}$  biotin (Millipore Sigma) and 1 mM ATP (Lucigen) were added to experimental samples for incubation at 37 °C for 10 min. Next, 10  $\mu\text{L}$  were removed for SDS-PAGE samples and the gel was run as described in 5.7.2.

NeutrAvidin-Alexa Fluor 647 conjugates for staining were made by the following pipetting scheme with NeutrAvidin (Thermo Fisher Scientific), sodium bicarbonate (Millipore Sigma), and NHS-Alexa555 (Thermo Fisher Scientific):

Reagent	Final concentration
NeutrAvidin	37.9 $\mu\text{M}$
Sodium bicarbonate	0.15 mM
NHS-Alexa555	150 $\mu\text{M}$
DPBS	Up to 200 $\mu\text{L}$
Total	200 $\mu\text{L}$

The reaction was incubated in the dark for 3 hours, then purified using an Amicon Ultra-0.5 filter unit (EMD Millipore) following the manufacturer's protocol and stored at 4 °C. WB was performed as described previously<sup>314</sup> and visualized on the Alexa555 blot channel (ChemiDoc MP, Bio-Rad)

#### 5.7.4 Native PAGE assay

TurboA $\beta_{\text{WT}}$ , TurboA $\beta_{\text{F20P}}$ , or TurboID (1.9 - 2.3  $\mu\text{M}$ ) in 1X DPBS (Gibco) was incubated on ice for 2 hours in 200  $\mu\text{L}$  in 0.2 mL PCR tubes. Aliquots (30  $\mu\text{L}$ ) were taken

out sequentially and incubated at 37 °C on a thermal cycler for 3, 2.5, 2, 1.5, 1, and 0.5 hours. Native PAGE samples were prepared and run on a 4-16% native PAGE gel (Thermo Fisher Scientific) according to manufacturer's protocol. Staining and de-staining was performed using the Coomassie Brilliant Blue R-250 Staining Solution kit (Bio-Rad) according to manufacturer's protocol.

#### *5.7.5 Additional experiments performed by Paulina Eberts*

The following sections describe a selection of cell-based and *in vitro* experiments performed by Paulina Eberts that are presented in this thesis.

##### *5.7.5.1 Immunostaining for confirming LRP1 and RAGE expression*

hCMEC/d3 cells were cultured in EndoGRO TM-MV Complete Media Kit (EMD Millipore) supplemented with 1 ng/mL FGF-2 (Peprotech) on Collagen Type I, Rat Tail (EMD Millipore). Transwell inserts (12-well polyester membrane; Corning) were coated with Collagen Type I, Rat Tail and hCMEC/d3 cells were seeded at a density of 50,000 cells/cm<sup>2</sup>. Medium was changed every 2 days and the cells were grown to confluence for ~4 - 5 days. Cells were stained for the ZO-1, LRP1, and RAGE as described elsewhere.<sup>315</sup> Cells were washed 1x with PBS and fixed for 15 minutes using 4% formaldehyde. After fixation, cells were washed 3x with PBS (five minutes per wash) and blocked using 5% normal goat serum and 0.3% Triton X-100 in PBS for 1 hour at RT. Primary antibodies were prepared in PBS containing 1% BSA and 0.3% Triton X-100 using the following dilutions: ZO-1 (1:50, Santa Cruz Biotechnology), LRP1 (5 µg/ml, Millipore-Sigma), and RAGE (4.5 µg/ml, Bio-Techne). Cells were incubated with the primary antibodies overnight at 4 °C. The next day, cells were washed 3x with PBS (five minutes per wash). Secondary antibodies were prepared in PBS containing 1% BSA and 0.3% Triton X-100 using the following dilutions: Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647 (1:200, Invitrogen), Goat anti-Mouse IgG (H+L) Cross-Adsorbed



Secondary Antibody, Alexa Fluor™ 488 (1:200, Invitrogen), Goat anti-Rat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 555 (1:200, Invitrogen). Cells were incubated with the secondary antibodies for 1 hour at RT. Cells were washed 3x with PBS (5 min per wash) and counterstained with DAPI (1:4000, Invitrogen) for 5 min at RT. Then, cells were washed 1x with PBS before the Transwell membranes were cut out of inserts and mounted on slides using 90% glycerol in PBS. Samples were then imaged using a confocal microscope (FluoView FV1000 BX2 Upright Confocal, Olympus).

#### *5.7.5.2 TurboA $\beta$ validation on hCMECs via immunofluorescence*

For labeling experiments with TurboA $\beta$  fusion constructs, cells were grown as described above. Prior to incubation, the cells were cooled to 4 °C for 30 min. Then, 1  $\mu$ M of TurboA $\beta$  constructs was diluted in 1 mL growth medium, added to cells, and incubated for 2 hours at 4 °C to allow the construct-cell surface receptor binding to reach equilibrium without internalization. The medium was then removed and replaced with prewarmed (37 °C) medium containing ATP (10 mM), biotin (50  $\mu$ M), and TurboA $\beta$  constructs (1  $\mu$ M). Cells were incubated for 10 min, 30 min, 3 hours, or 18 hours at 37 °C on a rocker. Finally, cells were placed back on ice and washed with ice-cold PBS 5x (five minutes per wash) to remove free construct in solution. Cells were stained with the NeutrAvidin-Alexa Fluor 555 using the same immunostaining protocol as above.

#### *5.7.5.3 Cell-free assay for assessing Ab-TurboID fusion construct activity*

Target RAGE Fc Chimera (R&D Biosystems) in PBS were immobilized overnight (buffer only in negative control wells) on Maxisorp plates (Thermo Fisher Scientific). The next day, the wells were washed with 0.05% Tween-20 in PBS and blocked with Synblock (Bio-Rad) for 24 hours at 4 °C. The wells were washed and panned with TurboA $\beta$  constructs for 2 hours at 4 °C. Then the wells were washed and incubated with prewarmed labeling solution (1  $\mu$ M TurboA $\beta$  constructs, 1 mM ATP, 50  $\mu$ M biotin) for 1 hour at 37 °C.

For detecting biotinylation, the wells were washed and incubated with streptavidin-HRP (Invitrogen) for 30 min at RT before being washed again and incubated with ECL solution (Thermo Fisher Scientific) for 5 min. Luminescence was detected on Synergy plate reader (Bio-Tek).

#### *5.7.5.4 Optimized cell-free assay for assessing activity of additional peptide-TurboID fusion constructs*

Maxisorp plates (Thermo Fisher Scientific) were blocked with Ultrablock (Bio-Rad), Synblock (Bio-Rad), or 1% w/v bovine serum albumin (Sigma-Aldrich) in PBS overnight at 4 °C. The wells were washed 4x with PBST (PBS with 0.05% Tween-20 (Sigma-Aldrich)) and incubated with panning solution (1 µM peptide-TurboID fusion constructs, Tris buffer KCl (TBK, 50 mM Tris, 100 mM KCl, pH 8.5), 10 mM ATP, 10 mM MgOAc, 50 µM biotin) for 3 hours at 37 °C on a rocker. For biotinylation detection, the wells were washed 4x with PBST, incubated with 1:200 dilution of streptavidin-HRP for 1 hour at RT, washed 4x with PBST again, and labeled with 1:1 ECL solution for 3 min at RT. Luminescence was detected on Synergy plate reader (Bio-Tek).

Table 5-3. Extended list of amino acid sequences used for creating TurboID fusion constructs.

Construct name	Description	Amino acid sequence
TurboID	Biotin ligase used in this assay	MKDNTVPLKLIALLANGFHSGEQLGETLGMSRAAIN KHIQTLRDWGVDFVTVPGKGYSLPEPIPLLNAKQILG QLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEY QQAGRGSRGRKWFSPFGANLYLSMFWRKRGPAIIGL GPVIGIVMAEALRKLKLGADKVRVWPNLDLYLQDRKLAG ILVELAGITGDAQIVIGAGINVAMRRVEESVNVQGW ITLQEAGINLDRNTLAATLIRELRAALELFEQEG LAP YLPWEKLDNFINRPVKLIIGDKEIFGISRGIDKQGA LLEQDGVIKPVMGGEISLRSAEK
miniTurboID	Biotin ligase used for initial expressions, but not used in experimental assays	MIPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIG ELKSGDACIAEYQQAGRGSRGRKWFSPFGANLYLSMF WRKRGPAIIGLGPVIGIVMAEALRKLKLGADKVRVWPN NDLYLQDRKLAGILVELAGITGDAQIVIGAGINVAM RRVEESVNVQGWITLQEAGINLDRNTLAAMLIRELRA ALELFEQEG LAPYLSRWEKLDNFINRPVKLIIGDKEI FGISRGIDKQGALLEQDGVIKPVMGGEISLRSAEK
Interdomain linker ver. 1	Linker between A $\beta$ <sub>1-42</sub> constructs and TurboID (EF = EcoRI site from Addgene plasmid)	SGSGSEF
Interdomain linker ver. 2	Linker between non-A $\beta$ (except A $\beta$ <sub>1-28</sub> ) constructs and TurboID	GSGSNGSTGS
A $\beta$ <sub>WT</sub>	42 amino acid A $\beta$ peptide	MDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMV GGVIA
A $\beta$ <sub>F20P</sub>	42 amino acid A $\beta$ peptide with single point mutation at Phe20 (bolded, underlined)	MDAEFRHDSGYEVHHQKLVF <b><u>P</u></b> AEDVGSNKGAIIGLMV GGVIA
Bivalent T7	Bivalent T7 peptide with mixed species interdomain linker (underlined)	<u>HAIYPRHGGTSGGDNSGTGGGAEAAKASGSTGGNSG</u> <u>GNSGGHAIYPRH</u>
Trivalent GRGDSP	Trivalent GRGDSP peptide with mixed species interdomain linker (underlined)	GRGDSP <u>PAGSGTSGGAGRGDSPAGSGTGSGAGRGDSP</u>

Table 5-4. List of primers and gBlocks used for oligo-annealing, overlap-extension PCR, and subcloning.

Primer name	Description	Sequence (5' → 3')
Ab_oligo-1_for	Forward primer for building first half of A $\beta$	GACCATATGGACGCCGAGTTTCGCCATGATTCGGGCTACGA AGTTCACCATCAG

Ab_wt_oligo-1-2_rev	Reverse primer for building first half of A $\beta$ <sub>WT</sub>	GTTGGAACCCACATCCTCAGCAAAGAAGACTAACTTCTGAT GGTGAACCTTCGTAGC
Ab_oligo-2-3_for	Forward primer for building second half of A $\beta$	GCTGAGGATGTGGGTCCAACAAGGGCGCGATCATCGGTCT GATGGTCGGAGGCGTC
Ab_oligo-3-4_rev	Reverse primer for building second half of A $\beta$	GCAGGATCCACTTCCGGACGCAATAACGACGCCCTCCGACCA TCAG
Ab_F20P_oligo-1-2_rev	Reverse primer for building first half of A $\beta$ <sub>F20P</sub>	GTTGGAACCCACATCCTCAGCCGGGAAGACTAACTTCTGAT GGTGAACCTTCGTAGC
Link-TurboID_for_v2	Forward primer for adding 5'-linker extension for interdomain linker ver. 2	GGATCCGGCAGTAATGGTTCTACCGGGAGCATGAAAGACAA TACTGTGCC
TurboID-HA_XhoI_rev	Reverse primer for adding 3'-HA tag to TurboID	CAATCTCGAGAGCGTAATCTGGAACATCGTATGGGTAAAGC TTCTTTTCGGCAGAC
AP-2_NdeI_for	Forward primer for cloning AP-2 peptide	GTACATATGACCTTTTTCTATGGTGGCTCTCGTGGTAAACG CAATAACTTCA
AP-2_BamHI_rev	Reverse primer for cloning AP-2 peptide	AGAACCATTACTGCCGGATCCATATTCTTCGGTCTTGAAGT TATTGCGTTTACC
GRGDSP_for	Forward primer for cloning GRGDSP peptide	GTACATATGGGTCTGGCGATAGCCCGGGATCCGGCAGTAA TGGTTCT
GRGDSP_rev	Reverse primer for cloning GRGDSP peptide	AGAACCATTACTGCCGGATCCCGGGCTATCGCCACGACCCA TATGTAC
HAIYPRH_for	Forward primer for cloning T7 peptide	GTACATATGCATGCGATTTATCCGCGTCACGGATCCGGCAG TAATGGTTCT
HAIYPRH_rev	Reverse primer for cloning T7 peptide	AGAACCATTACTGCCGGATCCGTGACGCGGATAAATCGCAT GCATATGTAC
Tf-pep_NdeI_for	Forward primer for cloning T12 peptide	GTACATATGACCCATCGTCCGCCGATGTGGAGCCCAGTGTG
Tf-pep_BamHI_rev	Reverse primer for cloning T12 peptide	AGAACCATTACTGCCGGATCCCGGCCACACTGGGCTCCACA TC

gBlock_T7linkT7	gBlock for subcloning bivalent T7 peptide	<p>GTCTGGTTCCTGGTGTTACATATGCACGCAATCTACCCGCG  CCATGGTGGCACGTCTGGCGGGGATAATAGCGGTACCGGCG  GTGGCGCCGAAGCGGCAGCTAAAGCTAGCGGATCTACCGGT  GGTAACAGTGGCGGCAATTCAGGCGGGCATGCGATTTATCC  TCGCCATGGATCCGTATTGCTGATCTGTTTCCAG</p>
gBlock_TriRGD	gBlock for subcloning trivalent GRGDSP peptide	<p>GTCTGGTTCCTGGTGTTACATATGGGTCGTGGCGATTCTCC  AGCAGGCAGTGGTACCTCTGGAGGCGCCGGTCGCGGTGATA  GTCCAGCTGGCTCAGGTACCGTTCTGGCGCAGGCCGTGGC  GATAGCCCTGGATCCGTATTGCTGATCTGTTTCCAG</p>

## Chapter 6: Conclusion, future directions, and outlook

---

The molecular engineering approaches and experimental protocols developed in this thesis improve the signal-to-noise ratio for the specific enrichment of protein binders in an array of biologics discovery platforms. Towards minimizing noise from nonspecific adsorption, we developed PEGylation strategies for stealthing DNA templates for biopanning applications in DNA display (Chapter 2). Further, to reduce noise from nonfunctional library members and nonspecific binding to enrich recovery of true binders, we introduced methods to incorporate photocrosslinking of the binder library to its target in mRNA display (Chapter 3). To improve signal (i.e., achieve higher binder recovery), we developed a cloning method using rolling circle amplification to create *in vitro* homomultivalent protein libraries (Chapter 4). Finally, in our protein engineering work, we made efforts toward engineering biotin ligase fusions to identify new therapeutic targets and understand disease biology with proximity ligation using a blood-brain barrier cell model (Chapter 5).

While the methods developed will require further optimization for implementation in full library selections and PPI identification, additional protocols and experimental techniques were introduced to improve and expand the capabilities of current selection platforms. In Chapters 2 - 4, we presented new protocols for performing off-rate wash cycles, a critical component of affinity-based selections. In Chapter 2, we presented an adherent biopanning approach with automated washing methods. The washing cycles, performed on 96-well plates with a plate washer and optimized for cell panning applications, can be adapted for high-throughput biopanning experiments to perform large binder screens for cell surface binding. We modified the biopanning approach in Chapter 4 to cells in suspension; here, the wash cycles we established showed minimal nonspecific

adsorption (~100 copies per cell) and sufficient cell surface binding. In Chapter 3, we introduced highly stringent wash buffers containing strong denaturants like urea and SDS. Although these specific wash conditions require covalent linkage or extremely high affinity interactions (such as the biotin-streptavidin interaction we used) between the target and the selection particle, the buffers used may be adapted to milder conditions to further select for binders with other desirable biophysical qualities, such as stronger protein folding and higher stability. Previous studies have shown that streptavidin can retain tetrameric structure in solutions with up to 1% SDS<sup>150</sup> and 6 M urea<sup>151</sup> although systematic demonstration of how biotin-streptavidin binding is affected by these denaturants has not been shown. Besides the automated plate washer, these protocols also do not require special equipment and can be easily implemented in other biopanning experiments with different display methods.

The probe qPCR approach used in Chapters 2 and 4 allowed for precise quantitative analyses even with a high cellular background (~10<sup>4</sup> cells per sample). Generally, the incorporation of qPCR within the *in vitro* display and selection workflow is a highly useful tool to quantitatively assess approximate enrichments and supplement RT-PCR steps, which have historically been qualitative. Due to the highly repetitive nature of the TolA spacer, some optimization was necessary for TolA-based amplifications. However, we have also successfully used the T7 promoter for designing qPCR amplicons, which is broadly applicable to qPCR analysis for DNA, mRNA, and ribosome displays.

The various technical improvements for RD introduced in Chapter 4 build on optimizations that were previously developed in our lab. Beginning with the T7 promoter, the sequences upstream were modified to encourage higher transcription efficiency<sup>316</sup>, and the sequences immediately surrounding the ribosome binding site (RBS) were also edited to decrease secondary structure in the region to increase translation efficiency.

Restriction sites such as NcoI and HindIII in pRDV2<sup>69</sup> were changed to reduce GC-content and charged amino acids, respectively. The short linker between the binder library and the ToIA spacer was also modified to improve primer annealing in the RT step by reducing secondary structure and lowering the melting temperature. For the 3-piece ligation used to synthesize the full T7-library-ToIA template for RD, we adapted Type IIS restriction sites and ligation conditions first described by Ng and Sarkar<sup>243</sup> for our template and achieved good ligation efficiency. Lastly, we compared PCR methods preceding IVT and achieved higher yield with the DNA polymerases Q5 (NEB) and SuperFi II (Thermo Fisher Scientific) compared to Phusion Hot Start Flex (NEB), which we had used in most of our cloning experiments; thus, we switched to Q5-based amplifications for the latest RD experiments. Many of these changes are reflective of better engineered enzymes and higher quality reagents that have been released throughout the duration of this thesis work. As cloning is an essential component of all directed evolution experiments, it may be helpful to continuously test new reagents and enzymes. Overall, the technical optimizations presented throughout this work may be highly useful for future library selections.

In addition to the proximity labeling project, each research chapter in this thesis demonstrates work with different *in vitro* display methods. Reflecting on this unique opportunity, my vision is that a combination of molecular engineering approaches in each chapter may be useful in building a better cell-surface selection platform. As discussed throughout the chapters, live-cell panning approaches are still extremely challenging, and many approaches require knowing the target *a priori*, which can greatly limit the target space. In a multi-faceted biopanning approach, one can begin with a homomultivalent binder library (Chapter 4) to allow for better enrichment and higher specificity, which is important for selection in a complex cellular environment. For the display method, mRNA display – but as a reverse-transcribed mRNA/cDNA heteroduplex – can be used, as it



generally does not adsorb to cell surfaces much like dsDNA, which we demonstrated in Chapter 4. In addition, the library can be encoded with a C-terminal AviTag to be biotinylated *in vitro* by BirA in solution, the precursor to the TurboID biotin ligase used in Chapter 5. This not only allows for ligand-based recovery post-selection, but also pre-selects the library to binders that are in-frame and functionally expressed in full, which were challenges that we aimed to address in Chapter 3. The particles can then be panned directly on live-cell surfaces following protocols developed in Chapter 2, for adherent cells, or Chapter 4, for cell suspensions. Next, the binders can be chemically crosslinked to surface binders using bifunctional cross-linkers; in contrast to some of the difficulties presented in Chapter 3 with photocrosslinking, the chemical crosslinking does not require engineering of binder libraries to accommodate photoactivatable amino acids. The biotinylated ligands can then be isolated by streptavidin beads, and the wash protocols developed in Chapter 3 can be applied for stringent washing. By placing the biotin tag on the ligand (i.e., binder library), the user can select against targets without knowing the target *a priori*; as such, the target space is not limited to tagged transmembrane proteins that can be transfected successfully or soluble biotinylated targets that can be expressed recombinantly or purchased. Finally, the binders can be recovered by RT-PCR methods described in Chapters 3 and 4. In addition, mass spectrometry analysis can be utilized to identify targets, and NGS can be performed to compare selection pools (e.g., conditions with additional washes pre-crosslinking for reduced background) to analyze true binders vs. non-specifically bound binders. Despite shortcomings we faced with library selections, combining elements of our methodology from individual *in vitro* library approaches may address many of the challenges we faced in each project.

Finally, in addition to developing platform technologies to perform cell surface selections and expanding the capabilities of current selection display methods, a grander

challenge in drug discovery is designing biologics that not only bind to the target but also produce desired therapeutic effects. The extended commentary in Section 4.4.2 present design considerations for multivalent proteins, which have successfully demonstrated superior selectivity, potency/efficacy, and specificity over their monovalent counterparts in an array of therapeutic applications. With the rise of computational models for *de novo* protein design<sup>317,318</sup> and ease of protein structure determination<sup>319,320</sup>, binder discovery and subsequent structure-guided multivalent protein design will likely experience a renaissance in the next decade. The molecular engineering approaches developed in this thesis work, as a complement to these technological advancements, contribute to the broader molecular biology toolbox for the discovery of new biologics.

## Bibliography

---

1. Makurvet, F. D. Biologics vs. small molecules: Drug costs and patient access. *Medicine in Drug Discovery* **9**, 100075 (2021).
2. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proceedings of the National Academy of Sciences* **70**, 3240–3244 (1973).
3. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).
4. Mullard, A. 2022 FDA approvals. *Nature Reviews Drug Discovery* (2023) doi:10.1038/d41573-023-00001-3.
5. Schellekens, H. Biosimilar therapeutics—what do we need to consider? *NDT Plus* **2**, i27–i36 (2009).
6. Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA* **323**, 844–853 (2020).
7. Mullard, A. 2017 FDA drug approvals. *Nature Reviews Drug Discovery* **17**, 81–85 (2018).
8. Mullard, A. 2018 FDA drug approvals. *Nature Reviews Drug Discovery* **18**, 85–89 (2019).
9. Mullard, A. 2019 FDA drug approvals. *Nature Reviews Drug Discovery* **19**, 79–84 (2020).
10. Mullard, A. 2020 FDA drug approvals. *Nature Reviews Drug Discovery* **20**, 85–90 (2021).
11. Mullard, A. 2021 FDA approvals. *Nature Reviews Drug Discovery* **21**, 83–88 (2022).
12. Nord, K. *et al.* Binding proteins selected from combinatorial libraries of an  $\alpha$ -helical bacterial receptor domain. *Nat Biotechnol* **15**, 772–777 (1997).
13. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of molecular biology* **332**, 489–503 (2003).
14. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnology* **22**, 575–582 (2004).
15. Gebauer, M. & Skerra, A. Engineered Protein Scaffolds as Next-Generation Therapeutics. *Annu Rev Pharmacol Toxicol* **60**, 391–415 (2020).
16. Owens, B. Faster, deeper, smaller—the rise of antibody-like scaffolds. *Nature Biotechnology* **35**, 602–603 (2017).
17. McLure, R. J., Radford, S. E. & Brockwell, D. J. High-throughput directed evolution: a golden era for protein science. *TRECHEM* **4**, 378–391 (2022).
18. Lutz, S. Beyond directed evolution - semi-rational protein engineering and design. *Curr Opin Biotechnol* **21**, 734–743 (2010).
19. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat Rev Genet* **16**, 379–394 (2015).
20. Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **228**, 1315–1317 (1985).
21. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **15**, 553–557 (1997).
22. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences* **94**, 12297–12302 (1997).
23. Hanes, J. & Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4937–42 (1997).
24. Almagro, J. C., Pedraza-Escalona, M., Arrieta, H. I. & Pérez-Tapia, S. M. Phage Display Libraries for Antibody Therapeutic Discovery and Development. *Antibodies* **8**, 44 (2019).
25. Könning, D. & Kolmar, H. Beyond antibody engineering: directed evolution of alternative binding scaffolds and enzymes using yeast surface display. *Microbial Cell Factories* **17**, 32 (2018).
26. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T. & Yanagawa, H. DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Res.* **31**, e118 (2003).
27. Du, Y. *et al.* mRNA display with library of even-distribution reveals cellular interactors of influenza virus NS1. *Nat Commun* **11**, 2449 (2020).
28. Stafford, R. L. *et al.* In vitro Fab display: a cell-free system for IgG discovery. *Protein Engineering, Design and Selection* **27**, 97–109 (2014).

29. Bidlingmaier, S. & Liu, B. Construction and Application of a Yeast Surface-displayed Human cDNA Library to Identify Post-translational Modification-dependent Protein-Protein Interactions \*. *Molecular & Cellular Proteomics* **5**, 533–540 (2006).
30. Seelig, B. mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nat Protoc* **6**, 540–552 (2011).
31. Kurz, M., Gu, K. & Lohse, P. A. Psoralen photo-crosslinked mRNA–puromycin conjugates: a novel template for the rapid and facile preparation of mRNA–protein fusions. *Nucleic Acids Res* **28**, e83 (2000).
32. Aviner, R. The science of puromycin: From studies of ribosome function to applications in biotechnology. *Comput Struct Biotechnol J* **18**, 1074–1083 (2020).
33. Wang, H. & Liu, R. Advantages of mRNA display selections over other selection techniques for investigation of protein–protein interactions. *Expert Rev Proteomics* **8**, 335–346 (2014).
34. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448**, 828–831 (2007).
35. Takahashi, T. T., Austin, R. J. & Roberts, R. W. mRNA display: ligand discovery, interaction analysis and beyond. *Trends in Biochemical Sciences* **28**, 159–165 (2003).
36. Shuwei Li, Steven Millward, and & Roberts\*, R. In Vitro Selection of mRNA Display Libraries Containing an Unnatural Amino Acid. *ACS Publications* <https://pubs.acs.org/doi/pdf/10.1021/ja026789q> doi:10.1021/ja026789q.
37. Zahnd, C., Amstutz, P. & Plückthun, A. Ribosome display: Selecting and evolving proteins in vitro that specifically bind to a target. *Nature Methods* **4**, 269–279 (2007).
38. Ueda, T., Kanamori, T. & Ohashi, H. Ribosome display with the PURE technology. *Methods Mol Biol* **607**, 219–225 (2010).
39. Matsuura, T., Yanagida, H., Ushioda, J., Urabe, I. & Yomo, T. Nascent chain, mRNA, and ribosome complexes generated by a pure translation system. *Biochemical and Biophysical Research Communications* **352**, 372–377 (2007).
40. Matsuura, T. & Plückthun, A. Selection based on the folding properties of proteins with ribosome display. *FEBS Letters* **539**, 24–28 (2003).
41. Markou, G. C. & Sarkar, C. A. A cell-free approach to identify binding hotspots in plant immune receptors. *Sci Rep* **12**, 501 (2022).
42. Kaltenbach, M. & Hollfelder, F. SNAP Display: In Vitro Protein Evolution in Microdroplets. in *Ribosome Display and Related Technologies: Methods and Protocols* (eds. Douthwaite, J. A. & Jackson, R. H.) 101–111 (Springer, 2012). doi:10.1007/978-1-61779-379-0\_7.
43. Doi, N. *et al.* DNA Display Selection of Peptide Ligands for a Full-Length Human G Protein-Coupled Receptor on CHO-K1 Cells. *PLOS ONE* **7**, e30084 (2012).
44. MacConnell, A. B. & Paegel, B. M. Poisson Statistics of Combinatorial Library Sampling Predict False Discovery Rates of Screening. *ACS Comb. Sci.* **19**, 524–532 (2017).
45. Diamante, L., Gatti-Lafranconi, P., Schaerli, Y. & Hollfelder, F. In vitro affinity screening of protein and peptide binders by megavalent bead surface display. *Protein engineering, design & selection : PEDS* **26**, 713–724 (2013).
46. Valldorf, B. *et al.* Antibody display technologies: selecting the cream of the crop. *Biological Chemistry* **403**, 455–477 (2022).
47. Saxton, R. A., Glassman, C. R. & Garcia, K. C. Emerging principles of cytokine pharmacology and therapeutics. *Nat Rev Drug Discov* **22**, 21–37 (2023).
48. Hashimoto, Y. *et al.* Engineered membrane protein antigens successfully induce antibodies against extracellular regions of claudin-5. *Sci Rep* **8**, 8383 (2018).
49. Jones, M. L. *et al.* Targeting membrane proteins for antibody discovery using phage display. *Scientific reports* **6**, 26240 (2016).
50. Panagides, N. *et al.* Evaluation of Phage Display Biopanning Strategies for the Selection of Anti-Cell Surface Receptor Antibodies. *Int J Mol Sci* **23**, 8470 (2022).
51. Lim, C. C., Woo, P. C. Y. & Lim, T. S. Development of a Phage Display Panning Strategy Utilizing Crude Antigens: Isolation of MERS-CoV Nucleoprotein human antibodies. *Sci Rep* **9**, 6088 (2019).
52. Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nat Biotechnol* **23**, 1105–1116 (2005).
53. Dominik, P. K. *et al.* Conformational Chaperones for Structural Studies of Membrane Proteins Using Antibody Phage Display with Nanodiscs. *Structure* **24**, 300–309 (2016).
54. Pavlidou, M., Hänel, K., Möckel, L. & Willbold, D. Nanodiscs Allow Phage Display Selection for Ligands to Non-Linear Epitopes on Membrane Proteins. *PLoS One* **8**, e72272 (2013).

55. Hötzel, I. *et al.* Efficient production of antibodies against a mammalian integral membrane protein by phage display. *Protein Engineering, Design and Selection* **24**, 679–689 (2011).
56. Thompson, A. A. *et al.* GPCR stabilization using the bicelle-like architecture of mixed sterol-detergent micelles. *Methods* **55**, 310–317 (2011).
57. Dai, Z. *et al.* Ultra-high-diversity factorizable libraries for efficient therapeutic discovery. *Genome Res.* **32**, 1787–1794 (2022).
58. Ponsel, D., Neugebauer, J., Ladetzki-Baehs, K. & Tissot, K. High Affinity, Developability and Functional Size: The Holy Grail of Combinatorial Antibody Library Generation. *Molecules* **16**, 3675–3700 (2011).
59. de Brito, P. M., Saruga, A., Cardoso, M. & Goncalves, J. Methods and cell-based strategies to produce antibody libraries: current state. *Appl Microbiol Biotechnol* **105**, 7215–7224 (2021).
60. Corzo, J. Time, the forgotten dimension of ligand binding teaching. *Biochemistry and Molecular Biology Education* **34**, 413–416 (2006).
61. Vauquelin, G. & Charlton, S. J. Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *Br J Pharmacol* **168**, 1771–1785 (2013).
62. Errington, W. J., Bruncsics, B. & Sarkar, C. A. Mechanisms of noncanonical binding dynamics in multivalent protein–protein interactions. *Proceedings of the National Academy of Sciences* **116**, 25659–25667 (2019).
63. Bruncsics, B., Errington, W. J. & Sarkar, C. A. MVsim is a toolset for quantifying and designing multivalent interactions. *Nature Communications* **13**, 5029 (2022).
64. Schmidt, M. M. & Wittrup, K. D. A modeling analysis of the effects of molecular size and binding affinity on tumor targeting. *Molecular Cancer Therapeutics* **8**, 2861–2871 (2009).
65. Cao, C. & Moulton, J. GWAS and drug targets. *BMC Genomics* **15**, S5 (2014).
66. Morris, J. H. *et al.* Affinity purification–mass spectrometry and network analysis to understand protein–protein interactions. *Nat Protoc* **9**, 2539–2554 (2014).
67. Bosch, J. A., Chen, C.-L. & Perrimon, N. Proximity-dependent labeling methods for proteomic profiling in living cells: an update. *Wiley interdisciplinary reviews. Developmental biology* **10**, e392 (2021).
68. Qin, W., Cho, K. F. & Cavanagh, P. E. Deciphering molecular interactions by proximity labeling. *Nature Methods* doi:10.1038/s41592-020-01010-5.
69. Barendt, P. A., Ng, D. T. W., McQuade, C. N. & Sarkar, C. A. Streamlined protocol for mRNA display. *ACS Combinatorial Science* **15**, 77–81 (2013).
70. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880–887 (2018).
71. Chen, G. *et al.* Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacol Sin* **38**, 1205–1235 (2017).
72. Amstutz, P., Forrer, P., Zahnd, C. & Plückthun, A. In vitro display technologies: novel developments and applications. *Current opinion in biotechnology* **12**, 400–405 (2001).
73. Wada, A. Development of Next-Generation Peptide Binders Using In vitro Display Technologies and Their Potential Applications. *Frontiers in Immunology* **4**, (2013).
74. Bradbury, A. R. M., Sidhu, S., Dübel, S. & McCafferty, J. Beyond natural antibodies: the power of in vitro display technologies. *Nat Biotechnol* **29**, 245–254 (2011).
75. Rawlings, A. E. Membrane proteins: always an insoluble problem? *Biochemical Society Transactions* **44**, 790–795 (2016).
76. Yin, H. & Flynn, A. D. Drugging Membrane Protein Interactions. *Annual Review of Biomedical Engineering* **18**, 51–76 (2016).
77. Cournia, Z. *et al.* Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. *J Membrane Biol* **248**, 611–640 (2015).
78. To'a Salazar, G., Huang, Z., Zhang, N., Zhang, X. G. & An, Z. Antibody Therapies Targeting Complex Membrane Proteins. *Engineering* **7**, 1541–1551 (2021).
79. Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol Pharmacol* **93**, 251–258 (2018).
80. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* **18**, 581–586 (2008).
81. Pandey, A., Shin, K., Patterson, R. E., Liu, X.-Q. & Rainey, J. K. Current strategies for protein production and purification enabling membrane protein structural biology. *Biochem. Cell Biol.* **94**, 507–527 (2016).

82. Wiseman, D. N. *et al.* Expression and purification of recombinant G protein-coupled receptors: A review. *Protein Expr Purif* **167**, 105524 (2020).
83. Stern, L. A. *et al.* Geometry and expression enhance enrichment of functional yeast-displayed ligands via cell panning. *Biotechnology and Bioengineering* **113**, 2328–2341 (2016).
84. McNaughton, B. R., Cronican, J. J., Thompson, D. B. & Liu, D. R. Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proceedings of the National Academy of Sciences* **106**, 6111–6116 (2009).
85. Wu, M. & Yuan, F. Membrane Binding of Plasmid DNA and Endocytic Pathways Are Involved in Electrotransfection of Mammalian Cells. *PLoS One* **6**, e20923 (2011).
86. Panja, S., Aich, P., Jana, B. & Basu, T. How does plasmid DNA penetrate cell membranes in artificial transformation process of *Escherichia coli*? *Molecular Membrane Biology* **25**, 411–422 (2008).
87. Neumann, E., Kakorin, S., Tsoneva, I., Nikolova, B. & Tomov, T. Calcium-mediated DNA adsorption to yeast cells and kinetics of cell transformation by electroporation. *Biophys J* **71**, 868–877 (1996).
88. Li, S.-D. & Huang, L. Stealth Nanoparticles: High Density but Sheddable PEG is a Key for Tumor Targeting. *J Control Release* **145**, 178–181 (2010).
89. Andersson, D., Svec, D., Pedersen, C., Henriksen, J. R. & Ståhlberg, A. Preamplification with dUTP and Cod UNG Enables Elimination of Contaminating Amplicons. *Int J Mol Sci* **19**, 3185 (2018).
90. Fantoni, N. Z., El-Sagheer, A. H. & Brown, T. A Hitchhiker's Guide to Click-Chemistry with Nucleic Acids. *Chemical Reviews* **121**, 7122–7154 (2021).
91. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).
92. Dommerholt, J. *et al.* Highly accelerated inverse electron-demand cycloaddition of electron-deficient azides with aliphatic cyclooctynes. *Nat Commun* **5**, 5378 (2014).
93. Leunissen, E. H. P. *et al.* Copper-Free Click Reactions with Polar Bicyclononyne Derivatives for Modulation of Cellular Imaging. *ChemBioChem* **15**, 1446–1451 (2014).
94. Yang, B. *et al.* Evaluation of Dibenzocyclooctyne and Bicyclononyne Click Reaction on Azido-Functionalized Antifouling Polymer Brushes via Microspotting. *Advanced Materials Interfaces* **9**, 2102325 (2022).
95. Presolski, S. I., Hong, V. P. & Finn, M. G. Copper-catalyzed azide-alkyne click chemistry for bioconjugation. *Current protocols in chemical biology* **3**, 153–162 (2011).
96. Vandeventer, P. E., Mejia, J., Nadim, A., Johal, M. S. & Niemz, A. DNA Adsorption to and Elution from Silica Surfaces: Influence of Amino Acid Buffers. *J Phys Chem B* **117**, 10742–10749 (2013).
97. Alfaleh, M. A., Jones, M. L., Howard, C. B. & Mahler, S. M. Strategies for Selecting Membrane Protein-Specific Antibodies using Phage Display with Cell-Based Panning. *Antibodies (Basel)* **6**, 10 (2017).
98. Yoon, H. *et al.* An efficient strategy for cell-based antibody library selection using an integrated vector system. *BMC Biotechnology* **12**, 62 (2012).
99. Watters, J. M., Telleman, P. & Junghans, R. P. An optimized method for cell-based phage display panning. *Immunotechnology* **3**, 21–29 (1997).
100. Stern, L. A., Lown, P. S. & Hackel, B. J. Ligand Engineering via Yeast Surface Display and Adherent Cell Panning. in *Genotype Phenotype Coupling: Methods and Protocols* (eds. Zielonka, S. & Krah, S.) 303–320 (Springer US, 2020). doi:10.1007/978-1-4939-9853-1\_17.
101. McGuire, M. J., Li, S. & Brown, K. C. Biopanning of phage displayed peptide libraries for the isolation of cell-specific ligands. *Methods in molecular biology (Clifton, N.J.)* **504**, 291–321 (2009).
102. Umlauf, B. J. *et al.* Identification of variable lymphocyte receptors that can target therapeutics to pathologically exposed brain extracellular matrix. *Science Advances* **5**, eaau4245 (2019).
103. Sefah, K., Shangguan, D., Xiong, X., O'Donoghue, M. B. & Tan, W. Development of DNA aptamers using Cell-SELEX. *Nat Protoc* **5**, 1169–1185 (2010).
104. Philpott, D. N. *et al.* Rapid On-Cell Selection of High-Performance Human Antibodies. *ACS Cent. Sci.* **8**, 102–109 (2022).
105. Wang, J. *et al.* Selection of phage-displayed peptides on live adherent cells in microfluidic channels. *Proceedings of the National Academy of Sciences* **108**, 6909–6914 (2011).
106. Goryunova, M. S., Arzhanik, V. K., Zavriev, S. K. & Ryazantsev, D. Y. Rolling circle amplification with fluorescently labeled dUTP—balancing the yield and degree of labeling. *Anal Bioanal Chem* **413**, 3737–3748 (2021).
107. Smolina, I. V. *et al.* High-density fluorescently labeled rolling-circle amplicons for DNA diagnostics. *Analytical Biochemistry* **347**, 152–155 (2005).

108. Paul, N. & Yee, J. PCR incorporation of modified dNTPs: the substrate properties of biotinylated dNTPs. *BioTechniques* **48**, 333–334 (2010).
109. Mariconti, M., Morel, M., Baigl, D. & Rudiuk, S. Enzymatically Active DNA-Protein Nanogels with Tunable Cross-Linking Density. *Biomacromolecules* **22**, 3431–3439 (2021).
110. Tasara, T. *et al.* Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA. *Nucleic Acids Res* **31**, 2636–2646 (2003).
111. Li, M. *et al.* Brush Conformation of Polyethylene Glycol Determines the Stealth Effect of Nanocarriers in the Low Protein Adsorption Regime. *Nano Lett* **21**, 1591–1598 (2021).
112. Galindo-Murillo, R. & Cheatham, T. E., III. Ethidium bromide interactions with DNA: an exploration of a classic DNA–ligand complex with unbiased molecular dynamics simulations. *Nucleic Acids Research* **49**, 3735–3747 (2021).
113. Pozzi, D. *et al.* Effect of polyethyleneglycol (PEG) chain length on the bio–nano-interactions between PEGylated lipid nanoparticles and biological fluids: from nanostructure to uptake in cancer cells. *Nanoscale* **6**, 2782–2792 (2014).
114. Amoozgar, Z. & Yeo, Y. Recent advances in stealth coating of nanoparticle drug delivery systems. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **4**, 219–233 (2012).
115. Nicholas, A. R., Scott, M. J., Kennedy, N. I. & Jones, M. N. Effect of grafted polyethylene glycol (PEG) on the size, encapsulation efficiency and permeability of vesicles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1463**, 167–178 (2000).
116. Alberts, B. *et al.* The Structure and Function of DNA. *Molecular Biology of the Cell. 4th edition* (2002).
117. Perry, J. L. *et al.* PEGylated PRINT Nanoparticles: The Impact of PEG Density on Protein Binding, Macrophage Association, Biodistribution, and Pharmacokinetics. *Nano Lett* **12**, 5304–5310 (2012).
118. Klug, A. & Lutter, L. C. The helical periodicity of DNA on the nucleosome. *Nucleic Acids Res* **9**, 4267–4283 (1981).
119. Quinn, J. R., Zimmerman, S. C., Del Bene, J. E. & Shavitt, I. Does the A·T or G·C Base-Pair Possess Enhanced Stability? Quantifying the Effects of CH···O Interactions and Secondary Interactions on Base-Pair Stability Using a Phenomenological Analysis and ab Initio Calculations. *J. Am. Chem. Soc.* **129**, 934–941 (2007).
120. Costantini, D. L., Bateman, K., McLarty, K., Vallis, K. A. & Reilly, R. M. Trastuzumab-Resistant Breast Cancer Cells Remain Sensitive to the Auger Electron-Emitting Radiotherapeutic Agent <sup>111</sup>In-NLS-Trastuzumab and Are Radiosensitized by Methotrexate. *J Nucl Med* **49**, 1498–1505 (2008).
121. Atwood, B. K., Lopez, J., Wager-Miller, J., Mackie, K. & Straiker, A. Expression of G protein-coupled receptors and related proteins in HEK293, AtT20, BV2, and N18 cell lines as revealed by microarray analysis. *BMC Genomics* **12**, 14 (2011).
122. Kaur, H. *et al.* Single-cell profiling reveals heterogeneity and functional patterning of GPCR expression in the vascular system. *Nat Commun* **8**, 15700 (2017).
123. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* **2**, 261–268 (2005).
124. Newton, M. S., Cabezas-Perusse, Y., Tong, C. L. & Seelig, B. *In Vitro* Selection of Peptides and Proteins—Advantages of mRNA Display. *ACS Synth. Biol.* **9**, 181–190 (2020).
125. Huang, Y., Wiedmann, M. M. & Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* **119**, 10360–10391 (2019).
126. Kamalinia, G., Grindel, B. J., Takahashi, T. T., Millward, S. W. & Roberts, R. W. Directing evolution of novel ligands by mRNA display. *Chem. Soc. Rev.* **50**, 9055–9103 (2021).
127. Terai, T., Koike, T. & Nemoto, N. Photocrosslinking of cDNA Display Molecules with Their Target Proteins as a New Strategy for Peptide Selection. *Molecules* **25**, 1472 (2020).
128. Chen, L. *et al.* Epitope-directed antibody selection by site-specific photocrosslinking. *Science Advances* **6**, eaaz7825 (2020).
129. Golden, M. C., Collins, B. D., Willis, M. C. & Koch, T. H. Diagnostic potential of PhotoSELEX-evolved ssDNA aptamers. *Journal of Biotechnology* **81**, 167–178 (2000).
130. Iacobucci, C. *et al.* Carboxyl-Photo-Reactive MS-Cleavable Cross-Linkers: Unveiling a Hidden Aspect of Diazirine-Based Reagents. *Anal. Chem.* **90**, 2805–2809 (2018).
131. Cole, N. B. Site-Specific Protein Labeling with SNAP-Tags. *Curr Protoc Protein Sci* **73**, 30.1.1–30.1.16 (2013).
132. Bannas, P., Hambach, J. & Koch-Nolte, F. Nanobodies and Nanobody-Based Human Heavy Chain Antibodies As Antitumor Therapeutics. *Front. Immunol.* **8**, 1603 (2017).
133. Chen, X., Gentili, M., Hacohen, N. & Regev, A. A cell-free nanobody engineering platform rapidly generates SARS-CoV-2 neutralizing nanobodies. *Nat Commun* **12**, 5506 (2021).

134. Zimmermann, I. *et al.* Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife* **7**, e34317 (2018).
135. Morrison, C. Nanobody approval gives domain antibodies a boost. *Nature Reviews Drug Discovery* **18**, 485–487 (2019).
136. Li, D. *et al.* Directed evolution of a picomolar-affinity, high-specificity antibody targeting phosphorylated tau. *Journal of Biological Chemistry* **293**, 12081–12094 (2018).
137. Yang, E. Y. & Shah, K. Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics. *Frontiers in Oncology* **10**, (2020).
138. De Genst, E. *et al.* Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proceedings of the National Academy of Sciences* **103**, 4586–4591 (2006).
139. McMahon, C. *et al.* Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat Struct Mol Biol* **25**, 289–296 (2018).
140. Mitchell, L. S. & Colwell, L. J. Comparative analysis of nanobody sequence and structure data. *Proteins* **86**, 697–706 (2018).
141. Zuo, Y. & Deutscher, M. P. The physiological role of RNase T can be explained by its unusual substrate specificity. *The Journal of biological chemistry* **277**, 29654–29661 (2002).
142. Liu, R., Barrick, J. E., Szostak, J. W. & Roberts, R. W. [19] Optimized synthesis of RNA-protein fusions for in vitro protein selection. in *Methods in Enzymology* vol. 318 268–293 (Academic Press, 2000).
143. Tabuchi, I. *et al.* An Efficient Ligation Method in the Making of an in vitro Virus for in vitro Protein Evolution. *Biol Proced Online* **4**, 49–54 (2002).
144. Nemoto, N., Miyamoto-Sato, E., Husimi, Y. & Yanagawa, H. In vitro virus: Bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Letters* **414**, 405–408 (1997).
145. Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T. & Husimi, Y. Y-ligation: An efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase. *Mol Divers* **4**, 187–190 (1998).
146. Lee, Y., Jeong, J., Lee, G., Moon, J. H. & Lee, M. K. Covalent and Oriented Surface Immobilization of Antibody Using Photoactivatable Antibody Fc-Binding Protein Expressed in *Escherichia coli*. *Anal. Chem.* **88**, 9503–9509 (2016).
147. Wurtmann, E. J. & Wolin, S. L. RNA under attack: Cellular handling of RNA damage. *Crit Rev Biochem Mol Biol* **44**, 34–49 (2009).
148. Ariza-Mateos, A. *et al.* RNA self-cleavage activated by ultraviolet light-induced oxidation. *Nucleic Acids Research* **40**, 1748 (2012).
149. Simon, M., Zangemeister-Wittke, U. & Plückthun, A. Facile Double-Functionalization of Designed Ankyrin Repeat Proteins using Click and Thiol Chemistries. *Bioconjugate Chem.* **23**, 279–286 (2012).
150. Waner, M. J., Navrotskaya, I., Bain, A., Oldham, E. D. & Mascotti, D. P. Thermal and Sodium Dodecylsulfate Induced Transitions of Streptavidin. *Biophys J* **87**, 2701–2713 (2004).
151. Kurzban, G. P., Bayer, E. A., Wilchek, M. & Horowitz, P. M. The quaternary structure of streptavidin in urea. *Journal of Biological Chemistry* **266**, 14470–14477 (1991).
152. Rosenberg, M. *et al.* Denaturing cross-linking immunoprecipitation to identify footprints for RNA-binding proteins. *STAR Protocols* **2**, 100819 (2021).
153. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res* **14**, 1188–1190 (2004).
154. Muyldermans, S. A guide to: generation and design of nanobodies. *The FEBS Journal* **288**, 2084–2102 (2021).
155. Valdés-Tresanco, M. S., Molina-Zapata, A., Pose, A. G. & Moreno, E. Structural Insights into the Design of Synthetic Nanobody Libraries. *Molecules* **27**, 2198 (2022).
156. Wang, J. *et al.* Research Progress and Applications of Multivalent, Multispecific and Modified Nanobodies for Disease Treatment. *Frontiers in Immunology* **12**, (2022).
157. Hino, N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2**, 201–206 (2005).
158. Ovaa, H. & wals, kim. Unnatural amino acid incorporation in *E. coli*: current and future applications in the design of therapeutic proteins. *Frontiers in Chemistry* **2**, (2014).
159. Dumas, A., Lercher, L., Spicer, C. & G. Davis, B. Designing logical codon reassignment – Expanding the chemistry in biology. *Chemical Science* **6**, 50–69 (2015).
160. Cai, B. *et al.* Selection of DNA-Encoded Libraries to Protein Targets within and on Living Cells. *J. Am. Chem. Soc.* **141**, 17057–17061 (2019).



161. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences* **109**, E690–E697 (2012).
162. Zahnd, C., Amstutz, P. & Plückthun, A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nat Methods* **4**, 269–279 (2007).
163. Newton, M. S., Cabezas-Perusse, Y., Tong, C. L. & Seelig, B. *In Vitro* Selection of Peptides and Proteins—Advantages of mRNA Display. *ACS Synth. Biol.* **9**, 181–190 (2020).
164. Dreier, B. & Plückthun, A. *Ribosome display: a technology for selecting and evolving proteins from large libraries. Methods in Molecular Biology* vol. 687 (Humana Press, 2011).
165. Zahnd, C., Sarkar, C. A. & Plückthun, A. Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution. *Protein Engineering, Design and Selection* **23**, 175–184 (2010).
166. Salzer, B. *et al.* Engineering AvidCARs for combinatorial antigen recognition and reversible control of CAR function. *Nature Communications* **11**, 4166 (2020).
167. Schmidt, N. W. *et al.* Liquid-crystalline ordering of antimicrobial peptide–DNA complexes controls TLR9 activation. *Nature Materials* **14**, 696–700 (2015).
168. Kwon, P. S. *et al.* Designer DNA architecture offers precise and multivalent spatial pattern-recognition for viral sensing and inhibition. *Nature Chemistry* **12**, 26–35 (2020).
169. Thurber, G. M., Schmidt, M. M. & Wittrup, K. D. Factors determining antibody distribution in tumors. *Trends in Pharmacological Sciences* **29**, 57–61 (2008).
170. Raybould, M. I. J. *et al.* Five computational developability guidelines for therapeutic antibody profiling. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4025–4030 (2019).
171. Rondot, S., Koch, J., Breitling, F. & Dübel, S. A helper phage to improve single-chain antibody presentation in phage display. *Nat Biotechnol* **19**, 75–78 (2001).
172. Høydahl, L. S. *et al.* Multivalent pIX phage display selects for distinct and improved antibody properties. *Scientific Reports* **6**, 1–13 (2016).
173. Hunt, A. C. *et al.* Multivalent designed proteins neutralize SARS-CoV-2 variants of concern and confer protection against infection in mice. *Science Translational Medicine* **14**, (2022).
174. Mattheakis, L. C., Bhatt, R. R. & Dower, W. J. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proceedings of the National Academy of Sciences* **91**, 9022–9026 (1994).
175. Kaltenbach, M., Stein, V. & Hollfelder, F. SNAP Dendrimers: Multivalent Protein Display on Dendrimer-Like DNA for Directed Evolution. *ChemBioChem* **12**, 2208–2216 (2011).
176. Ali, M. M. *et al.* Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chemical Society Reviews* **43**, 3324–3341 (2014).
177. Steiner, J. M. *et al.* DeNA<sup>o</sup>: Selectable deoxyribonucleic acid nanoparticle libraries. *Journal of Biotechnology* **145**, 330–333 (2010).
178. Wang, L. *et al.* Rolling circle amplification-mediated hairpin RNA (RMHR) library construction in plants. *Nucleic Acids Research* **36**, e149 (2008).
179. Zahnd, C. *et al.* A Designed Ankyrin Repeat Protein Evolved to Picomolar Affinity to Her2. *Journal of Molecular Biology* **369**, 1015–1028 (2007).
180. Wright, G. J. Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol. BioSyst.* **5**, 1405–1412 (2009).
181. Simons, K. & Sampaio, J. L. Membrane Organization and Lipid Rafts. *Cold Spring Harb Perspect Biol* **3**, a004697 (2011).
182. Mondal, K. *et al.* Thermodynamic and kinetic characterization of hydroxyethylamine  $\beta$ -secretase-1 inhibitors. *Biochemical and biophysical research communications* **441**, 291–6 (2013).
183. Mansoor, S., Kayık, G., Durdagi, S. & Sensoy, O. Mechanistic insight into the impact of a bivalent ligand on the structure and dynamics of a GPCR oligomer. *Comput Struct Biotechnol J* **20**, 925–936 (2022).
184. Morzy, D. & Bastings, M. Significance of Receptor Mobility in Multivalent Binding on Lipid Membranes. *Angewandte Chemie International Edition* **61**, e202114167 (2022).
185. Du, Q., Smith, C., Shiffeldrim, N., Vologodskaya, M. & Vologodskii, A. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proceedings of the National Academy of Sciences* **102**, 5397–5402 (2005).
186. Ng, D. T. W. & Sarkar, C. A. NP-Sticky: A Web Server for Optimizing DNA Ligation with Non-Palindromic Sticky Ends. *Journal of Molecular Biology* **426**, 1861–1869 (2014).

187. Monera, O. D., Kay, C. M. & Hodges, R. S. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci* **3**, 1984–1991 (1994).
188. Li, J. Y. *et al.* A Biparatopic HER2-Targeting Antibody-Drug Conjugate Induces Tumor Regression in Primary Models Refractory to or Ineligible for HER2-Targeted Therapy. *Cancer Cell* **29**, 117–129 (2016).
189. Seet, B. T. *et al.* Efficient T-cell receptor signaling requires a high-affinity interaction between the Gads C-SH3 domain and the SLP-76 RxxK motif. *EMBO Journal* **26**, 678–689 (2007).
190. Liu, F. & Walters, K. J. Multitasking with ubiquitin through multivalent interactions. *Trends in Biochemical Sciences* **35**, 352–360 (2010).
191. Lees, W. J., Spaltenstein, A., Kingery-Wood, J. E. & Whitesides, G. M. Polyacrylamides Bearing Pendant  $\alpha$ -Sialoside Groups Strongly Inhibit Agglutination of Erythrocytes by Influenza A Virus: Multivalency and Steric Stabilization of Particulate Biological Systems. *Journal of Medicinal Chemistry* **37**, 3419–3433 (1994).
192. Mammen, M., Choi, S. K. & Whitesides, G. M. Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angewandte Chemie - International Edition* **37**, 2754–2794 (1998).
193. Mankowska, S. A. *et al.* A Shorter Route to Antibody Binders via Quantitative in vitro Bead-Display Screening and Consensus Analysis. *Scientific Reports* **6**, 1–11 (2016).
194. Jacobson, K., Liu, P. & Lagerholm, B. C. The Lateral Organization and Mobility of Plasma Membrane Components. *Cell* **177**, 806–819 (2019).
195. Kaufmann, R., Müller, P., Hildenbrand, G., Hausmann, M. & Cremer, C. Analysis of Her2/neu membrane protein clusters in different types of breast cancer cells using localization microscopy. *Journal of Microscopy* **242**, 46–54 (2011).
196. Markham, N. R. & Zuker, M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* **453**, 3–31 (2008).
197. Viguera, E., Canceill, D. & Ehrlich, S. D. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J* **20**, 2587–2595 (2001).
198. Hommelsheim, C. M., Frantzeskakis, L., Huang, M. & Ülker, B. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci Rep* **4**, 5052 (2014).
199. Sahdev, S., Saini, S., Tiwari, P., Saxena, S. & Singh Saini, K. Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. *Molecular and Cellular Probes* **21**, 303–307 (2007).
200. Riet, J., Ramos, L. R. V., Lewis, R. V. & Marins, L. F. Improving the PCR protocol to amplify a repetitive DNA sequence. *Genet Mol Res* **16**, (2017).
201. Gavrilov, M. *et al.* Engineered helicase replaces thermocycler in DNA amplification while retaining desired PCR characteristics. *Nat Commun* **13**, 6312 (2022).
202. Tang, N. C. & Chilkoti, A. Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nature Mater* **15**, 419–424 (2016).
203. Arslan, M. *et al.* Effect of non-repetitive linker on in vitro and in vivo properties of an anti-VEGF scFv. *Scientific Reports* **2022 12:1** **12**, 1–8 (2022).
204. Zhao, H. L. *et al.* Increasing the homogeneity, stability and activity of human serum albumin and interferon- $\alpha$ 2b fusion protein by linker engineering. *Protein Expression and Purification* **61**, 73–77 (2008).
205. Ma, J. *et al.* Bispecific Antibodies: From Research to Clinical Application. *Frontiers in Immunology* **12**, (2021).
206. Hong, S. *et al.* The Binding Avidity of a Nanoparticle-Based Multivalent Targeted Drug Delivery Platform. *Chemistry & Biology* **14**, 107–115 (2007).
207. van Dongen, M. A., Dougherty, C. A. & Banaszak Holl, M. M. Multivalent Polymers for Drug Delivery and Imaging: The Challenges of Conjugation. *Biomacromolecules* **15**, 3215–3234 (2014).
208. Strauch, E.-M. *et al.* Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site. *Nat Biotechnol* **35**, 667–671 (2017).
209. Silverman, J. *et al.* Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nat Biotechnol* **23**, 1556–1561 (2005).
210. Wrapp, D. *et al.* Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell* **181**, 1004–1015.e15 (2020).
211. Van Fossen, E. M. *et al.* Nanobody assemblies with fully flexible topology enabled by genetically encoded tetrazine amino acids. *Science Advances* **8**, eabm6909 (2022).

212. Britton, D. *et al.* Engineered multivalent self-assembled binder protein against SARS-CoV-2 RBD. *Biochem Eng J* **187**, 108596 (2022).
213. Hale, M. *et al.* IgM antibodies derived from memory B cells are potent cross-variant neutralizers of SARS-CoV-2. *Journal of Experimental Medicine* **219**, e20220849 (2022).
214. Miersch, S. *et al.* Ultrapotent and broad neutralization of SARS-CoV-2 variants by modular, tetravalent, bi-paratopic antibodies. *Cell Reports* **39**, (2022).
215. Svilenov, H. L. *et al.* Multimeric ACE2-IgM fusions as broadly active antivirals that potently neutralize SARS-CoV-2 variants. *Commun Biol* **5**, 1–7 (2022).
216. Kondo, T. *et al.* Antibody-like proteins that capture and neutralize SARS-CoV-2. *Science Advances* **6**, 3916–3930 (2020).
217. Chonira, V. *et al.* A potent and broad neutralization of SARS-CoV-2 variants of concern by DARPins. *Nat Chem Biol* 1–8 (2022) doi:10.1038/s41589-022-01193-2.
218. Akhtar, M. J., Ahamed, M., Alhadlaq, H. A., Alrokayan, S. A. & Kumar, S. Targeted anticancer therapy: Overexpressed receptors and nanotechnology. *Clinica Chimica Acta* **436**, 78–92 (2014).
219. Zuckier, L. S. *et al.* Influence of Affinity and Antigen Density on Antibody Localization in a Modifiable Tumor Targeting Model. *Cancer Research* **60**, 7008–7013 (2000).
220. Calson, C. B., Mowery, P., Owen, R. M., Dykhuizen, E. C. & Kiessling, L. L. Selective tumor cell targeting using low-affinity, multivalent interactions. *ACS Chemical Biology* **2**, 119–127 (2007).
221. Csizmar, C. M. *et al.* Multivalent Ligand Binding to Cell Membrane Antigens: Defining the Interplay of Affinity, Valency, and Expression Density. *J. Am. Chem. Soc.* **141**, 251–261 (2019).
222. Flugel, C. L. *et al.* Overcoming on-target, off-tumour toxicity of CAR T cell therapy for solid tumours. *Nat Rev Clin Oncol* **20**, 49–62 (2023).
223. Caruso, H. G. *et al.* Tuning Sensitivity of CAR to EGFR Density Limits Recognition of Normal Tissue While Maintaining Potent Antitumor Activity. *Cancer Research* **75**, 3505–3518 (2015).
224. Nezlin, R. CHAPTER 1 - General Characteristics of Immunoglobulin Molecules. in *The Immunoglobulins* (ed. Nezlin, R.) 3–73 (Academic Press, 1998). doi:10.1016/B978-012517970-6/50001-1.
225. Mahon, E. & Barboiu, M. Organic & Biomolecular Chemistry Synthetic multivalency for biological applications. *Organic & Biomolecular Chemistry* **13**, 10590 (2015).
226. Oostindie, S. C., Lazar, G. A., Schuurman, J. & Parren, P. W. H. I. Avidity in antibody effector functions and biotherapeutic drug design. *Nat Rev Drug Discov* **21**, 715–735 (2022).
227. Pedersen, M. B. *et al.* Curvature of Synthetic and Natural Surfaces Is an Important Target Feature in Classical Pathway Complement Activation. *The Journal of Immunology* **184**, 1931–1945 (2010).
228. Romero, L. A. *et al.* High-valency Anti-CD99 Antibodies Toward the Treatment of T Cell Acute Lymphoblastic Leukemia. *Journal of Molecular Biology* **434**, 167402 (2022).
229. Dalal, C. & Jana, N. R. Multivalency Effect of TAT-Peptide-Functionalized Nanoparticle in Cellular Endocytosis and Subcellular Trafficking. *J. Phys. Chem. B* **121**, 2942–2951 (2017).
230. Wang, J., Min, J., Egtesadi, S. A., Kane, R. S. & Chilkoti, A. Quantitative Study of the Interaction of Multivalent Ligand-Modified Nanoparticles with Breast Cancer Cells with Tunable Receptor Density. *ACS Nano* **14**, 372–383 (2020).
231. Glassman, P. M. *et al.* Molecularly Engineered Nanobodies for Tunable Pharmacokinetics and Drug Delivery. *Bioconjugate Chemistry* **31**, 1144–1155 (2020).
232. Chen, X., Zaro, J. & Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv Drug Deliv Rev* **65**, 1357–1369 (2013).
233. van Rosmalen, M., Krom, M. & Merkx, M. Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. *Biochemistry* **56**, 6565–6574 (2017).
234. Klein, J. S. *et al.* Design and characterization of structured protein linkers with differing flexibilities. *Protein Engineering, Design and Selection* **27**, 325–330 (2014).
235. Arviv, O. & Levy, Y. Folding of multidomain proteins: Biophysical consequences of tethering even in apparently independent folding. *Proteins: Structure, Function, and Bioinformatics* **80**, 2780–2798 (2012).
236. Sokolovski, M., Bhattacharjee, A., Kessler, N., Levy, Y. & Horovitz, A. Thermodynamic Protein Destabilization by GFP Tagging: A Case of Interdomain Allostery. *Biophysical Journal* **109**, 1157–1162 (2015).
237. Tjandra, K. C. & Thordarson, P. Multivalency in Drug Delivery—When Is It Too Much of a Good Thing? *Bioconjugate Chem.* **30**, 503–514 (2019).
238. Mohan, K. *et al.* Topological control of cytokine receptor signaling induces differential effects in hematopoiesis. *Science* **364**, eaav7532 (2019).

239. Romantini, N. *et al.* Exploring the signaling space of a GPCR using bivalent ligands with a rigid oligoproline backbone. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2108776118 (2021).
240. Hübner, H. *et al.* Structure-guided development of heterodimer-selective GPCR ligands. *Nat Commun* **7**, 12298 (2016).
241. Koenig, P.-A. *et al.* Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape. *Science* **eabe6230**, (2021).
242. DNA loading buffer (6X). *Cold Spring Harb Protoc* **2007**, pdb.rec11045 (2007).
243. Ng, D. T. W. & Sarkar, C. A. Model-guided ligation strategy for optimal assembly of DNA libraries. *Protein Engineering Design and Selection* **25**, 669–678 (2012).
244. Markou, G. C., Ohoka, A. & Sarkar, C. A. Engineering a Minimal Leucine-rich Repeat IgG-binding Module. *Applied Biochemistry and Biotechnology* **194**, 1636–1644 (2022).
245. Stumpf, M. P. H. *et al.* Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6959–6964 (2008).
246. Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of early drug discovery. *Br J Pharmacol* **162**, 1239–1249 (2011).
247. Plewczyński, D. & Ginalski, K. The interactome: Predicting the protein-protein interactions in cells. *Cell Mol Biol Lett* **14**, 1–22 (2008).
248. Strebhardt, K. & Ullrich, A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat Rev Cancer* **8**, 473–480 (2008).
249. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* **4**, 682–690 (2008).
250. Frei, A. P., Moest, H., Novy, K. & Wollscheid, B. Ligand-based receptor identification on living cells and tissues using TRICEPS. *Nat Protoc* **8**, 1321–1336 (2013).
251. Frei, A. P. *et al.* Direct identification of ligand-receptor interactions on living cells and tissues. *Nat Biotechnol* **30**, 997–1001 (2012).
252. Müller, M. *et al.* Light-mediated discovery of surfaceome nanoscale organization and intercellular receptor interaction networks. *Nat Commun* **12**, 7036 (2021).
253. Sobotzki, N. *et al.* HATRIC-based identification of receptors for orphan ligands. *Nat Commun* **9**, 1519 (2018).
254. Tremblay, T.-L. & Hill, J. J. Biotin-transfer from a trifunctional crosslinker for identification of cell surface receptors of soluble protein ligands. *Sci Rep* **7**, 46574 (2017).
255. Müskens, F. M. *et al.* Design, Synthesis, and Evaluation of a Diazirine Photoaffinity Probe for Ligand-Based Receptor Capture Targeting G Protein-Coupled Receptors. *Mol Pharmacol* **95**, 196–209 (2019).
256. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *Int J Mol Sci* **10**, 2763–2788 (2009).
257. Slavoff, S. A. & Saghatelian, A. Discovering ligand-receptor interactions. *Nat Biotechnol* **30**, 959–961 (2012).
258. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**, 51–54 (2015).
259. Mathew, B., Bathla, S., Williams, K. R. & Nairn, A. C. Deciphering Spatial Protein-Protein Interactions in Brain Using Proximity Labeling. *Mol Cell Proteomics* **21**, 100422 (2022).
260. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol* **30**, 1143–1148 (2012).
261. Rhee, H.-W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* **339**, 1328–1331 (2013).
262. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology* **196**, 801–810 (2012).
263. Kim, D. I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *MBoC* **27**, 1188–1196 (2016).
264. Silverstein, S. C., Steinman, R. M. & Cohn, Z. A. Endocytosis. *Annu Rev Biochem* **46**, 669–722 (1977).
265. Alwash, M. & Gariépy, J. Labeling Cell Surface Receptors with Ligand.BirA\* Bispecifics. *ACS Pharmacol. Transl. Sci.* **5**, 62–69 (2022).
266. Ma, C., Hong, F. & Yang, S. Amyloidosis in Alzheimer's Disease: Pathogeny, Etiology, and Related Therapeutic Directions. *Molecules* **27**, 1210 (2022).
267. Bateman, R. J. *et al.* Human amyloid- $\beta$  synthesis and clearance rates as measured in cerebrospinal fluid in vivo. *Nat Med* **12**, 856–861 (2006).

268. Shokri-Kojori, E. *et al.*  $\beta$ -Amyloid accumulation in the human brain after one night of sleep deprivation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4483–4488 (2018).
269. van Bokhoven, P. *et al.* The Alzheimer's disease drug development landscape. *Alz Res Therapy* **13**, 1–9 (2021).
270. Bharadwaj, P. R., Dubey, A. K., Masters, C. L., Martins, R. N. & Macreadie, I. G. A $\beta$  aggregation and possible implications in Alzheimer's disease pathogenesis. *J Cell Mol Med* **13**, 412–421 (2009).
271. Makin, S. The amyloid hypothesis on trial. *Nature* **559**, S4–S7 (2018).
272. Mullard, A. Controversial Alzheimer's drug approval could affect other diseases. *Nature* **595**, 162–163 (2021).
273. Garai, K., Crick, S. L., Mustafi, S. M. & Frieden, C. Expression and purification of amyloid- $\beta$  peptides from *Escherichia coli*. *Protein Expression and Purification* **66**, 107–112 (2009).
274. Zhang, L. *et al.* Expression, purification, and characterization of recombinant human  $\beta$ -amyloid42 peptide in *Escherichia coli*. *Protein Expression and Purification* **64**, 55–62 (2009).
275. Jia, L. *et al.* Highly efficient soluble expression, purification and characterization of recombinant A $\beta$ 42 from *Escherichia coli*. *RSC Adv.* **8**, 18434–18441 (2018).
276. Hatami, A., Monjazeb, S., Milton, S. & Glabe, C. G. Familial Alzheimer's Disease Mutations within the Amyloid Precursor Protein Alter the Aggregation and Conformation of the Amyloid- $\beta$  Peptide. *J Biol Chem* **292**, 3172–3185 (2017).
277. Esler, W. P. *et al.* Point substitution in the central hydrophobic cluster of a human beta-amyloid congener disrupts peptide folding and abolishes plaque competence. *Biochemistry* **35**, 13914–13921 (1996).
278. Wood, S. J., Wetzel, R., Martin, J. D. & Hurler, M. R. Prolines and Amyloidogenicity in Fragments of the Alzheimer's Peptide .beta./A4. *Biochemistry* **34**, 724–730 (1995).
279. Hilbich, C., Kisters-Woike, B., Reed, J., Masters, C. L. & Beyreuther, K. Substitutions of hydrophobic amino acids reduce the amyloidogenicity of Alzheimer's disease  $\beta$ A4 peptides. *Journal of Molecular Biology* **228**, 460–473 (1992).
280. Wurth, C., Guimard, N. K. & Hecht, M. H. Mutations that Reduce Aggregation of the Alzheimer's A $\beta$ 42 Peptide: an Unbiased Search for the Sequence Determinants of A $\beta$  Amyloidogenesis. *Journal of Molecular Biology* **319**, 1279–1290 (2002).
281. Park, K.-W. *et al.* Gene therapy using A $\beta$  variants for amyloid reduction. *Molecular Therapy* **29**, 2294–2307 (2021).
282. Ochiishi, T. *et al.* Development of new fusion proteins for visualizing amyloid- $\beta$  oligomers in vivo. *Scientific Reports* **6**, 1–15 (2016).
283. Crescenzi, O. *et al.* Solution structure of the Alzheimer amyloid  $\beta$ -peptide (1–42) in an apolar microenvironment. *European Journal of Biochemistry* **269**, 5642–5648 (2002).
284. Xiao, Y. *et al.* A $\beta$ (1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat Struct Mol Biol* **22**, 499–505 (2015).
285. Palmer, I. & Wingfield, P. T. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from *Escherichia coli*. *Curr Protoc Protein Sci* Unit-6.3 (2004) doi:10.1002/0471140864.ps0603s38.
286. Primo, E. D., Otero, L. H., Ruiz, F., Klinke, S. & Giordano, W. The disruptive effect of lysozyme on the bacterial cell wall explored by an in-silico structural outlook. *Biochemistry and Molecular Biology Education* **46**, 83–90 (2018).
287. Bornhorst, J. A. & Falke, J. J. [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol* **326**, 245–254 (2000).
288. Vagenende, V., Yap, M. G. S. & Trout, B. L. Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. *Biochemistry* **48**, 11084–11096 (2009).
289. Wittig, I., Braun, H. P. & Schägger, H. Blue native PAGE. *Nature Protocols* **2006 1:1**, 418–428 (2006).
290. Ma, J. & Xia, D. The use of blue native PAGE in the evaluation of membrane protein aggregation states for crystallization. *J Appl Crystallogr* **41**, 1150–1160 (2008).
291. Weksler, B., Romero, I. A. & Couraud, P.-O. The hCMEC/D3 cell line as a model of the human blood brain barrier. *Fluids and Barriers of the CNS* **10**, 16 (2013).
292. Biemans, E. A. L. M., Jäkel, L., de Waal, R. M. W., Kuiperij, H. B. & Verbeek, M. M. Limitations of the hCMEC/D3 cell line as a model for A $\beta$  clearance by the human blood-brain barrier. *Journal of Neuroscience Research* **95**, 1513–1522 (2017).
293. Van Gool, B. *et al.* LRP1 Has a Predominant Role in Production over Clearance of A $\beta$  in a Mouse Model of Alzheimer's Disease. *Mol Neurobiol* **56**, 7234–7245 (2019).

294. Donahue, J. E. *et al.* RAGE, LRP-1, and amyloid-beta protein in Alzheimer's disease. *Acta Neuropathol* **112**, 405–415 (2006).
295. Urich, E., Lazic, S. E., Molnos, J., Wells, I. & Freskgård, P.-O. Transcriptional Profiling of Human Brain Endothelial Cells Reveals Key Properties Crucial for Predictive In Vitro Blood-Brain Barrier Models. *PLoS One* **7**, e38149 (2012).
296. Demeule, M. *et al.* Identification and Design of Peptides as a New Drug Delivery System for the Brain. *J Pharmacol Exp Ther* **324**, 1064–1072 (2008).
297. Kapp, T. G. *et al.* A Comprehensive Evaluation of the Activity and Selectivity Profile of Ligands for RGD-binding Integrins. *Sci Rep* **7**, 39805 (2017).
298. Lee, J. H., Engler, J. A., Collawn, J. F. & Moore, B. A. Receptor mediated uptake of peptides that bind the human transferrin receptor. *European Journal of Biochemistry* **268**, 2004–2012 (2001).
299. Eker, F., Griebenow, K. & Schweitzer-Stenner, R. A $\beta$ 1-28 Fragment of the Amyloid Peptide Predominantly Adopts a Polyproline II Conformation in an Acidic Solution. *Biochemistry* **43**, 6893–6898 (2004).
300. Hou, L. *et al.* Solution NMR Studies of the A $\beta$ (1–40) and A $\beta$ (1–42) Peptides Establish that the Met35 Oxidation State Affects the Mechanism of Amyloid Formation. *J. Am. Chem. Soc.* **126**, 1992–2005 (2004).
301. Krishnamurthy, V. M., Estroff, L. A. & Whitesides, G. M. Multivalency in Ligand Design. in *Fragment-based approaches in Drug Discovery* 11–53 (Wiley-Blackwell, 2006). doi:10.1002/3527608761.ch2.
302. Wojtowicz, W. M. *et al.* A Human IgSF Cell-Surface Interactome Reveals a Complex Network of Protein-Protein Interactions. *Cell* **182**, 1027-1043.e17 (2020).
303. Mair, A. & Bergmann, D. C. Advances in enzyme-mediated proximity labeling and its potential for plant research. *Plant Physiology* **188**, 756–768 (2022).
304. Xiong, Z. *et al.* In vivo proteomic mapping through GFP-directed proximity-dependent biotin labelling in zebrafish. *eLife* **10**, e64631 (2021).
305. Schopp, I. M. *et al.* Split-BiolD a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nat Commun* **8**, 15690 (2017).
306. Cho, K. F. *et al.* Split-TurboID enables contact-dependent proximity labeling in cells. *Proceedings of the National Academy of Sciences* **117**, 12143–12154 (2020).
307. Kubitz, L. *et al.* Engineering of ultraID, a compact and hyperactive enzyme for proximity-dependent biotinylation in living cells. *Commun Biol* **5**, 1–14 (2022).
308. Minde, D.-P., Ramakrishna, M. & Lilley, K. S. Biotin proximity tagging favours unfolded proteins and enables the study of intrinsically disordered regions. *Commun Biol* **3**, 1–13 (2020).
309. Attar, A., Meral, D., Urbanc, B. & Bitan, G. Chapter 38 - Assembly of Amyloid  $\beta$ -Protein Variants Containing Familial Alzheimer's Disease-Linked Amino Acid Substitutions. in *Bio-nanoimaging* (eds. Uversky, V. N. & Lyubchenko, Y. L.) 429–442 (Academic Press, 2014). doi:10.1016/B978-0-12-394431-3.00038-9.
310. Li, S. C., Goto, N. K., Williams, K. A. & Deber, C. M. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proceedings of the National Academy of Sciences* **93**, 6676–6681 (1996).
311. Hong, P., Koza, S. & Bouvier, E. S. P. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol* **35**, 2923–2950 (2012).
312. Li, Y. *et al.* A Unique SEC Method Overcomes the Multiple-Monomeric-Peak Profile of the Broadly Neutralizing HIV-1 Antibody 10E8. *Anal Chem* **90**, 12390–12394 (2018).
313. Parker, B. W. *et al.* Mapping low-affinity/high-specificity peptide–protein interactions using ligand-footprinting mass spectrometry. *Proceedings of the National Academy of Sciences* **116**, 21001–21011 (2019).
314. Western blot protocol | Abcam. <https://www.abcam.com/protocols/general-western-blot-protocol>.
315. Immunofluorescence Protocol with Formaldehyde Fixation. *Cell Signaling Technology* <https://www.cellsignal.com/learn-and-support/protocols/protocol-if>.
316. Baklanov, M. M., Golikova, L. N. & Malygin, E. G. Effect on DNA transcription of nucleotide sequences upstream to T7 promoter. *Nucleic Acids Res* **24**, 3659–3660 (1996).
317. Cao, L. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
318. Chidyausiku, T. M. *et al.* De novo design of immunoglobulin-like domains. *Nat Commun* **13**, 5661 (2022).
319. Renaud, J.-P. *et al.* Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat Rev Drug Discov* **17**, 471–492 (2018).

320. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
321. Liu, H., Schittny, V. & Nash, M. A. Removal of a Conserved Disulfide Bond Does Not Compromise Mechanical Stability of a VHH Antibody Complex. *Nano Letters* **19**, 5524–5529 (2019).
322. Xu, L., Song, X. & Jia, L. A camelid nanobody against EGFR was easily obtained through refolding of inclusion body expressed in *Escherichia coli*. *Biotechnology and Applied Biochemistry* **64**, 895–901 (2017).
323. Vallejo, L. F. & Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories* **3**, 11 (2004).
324. Yamaguchi, H. & Miyazaki, M. Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules* **4**, 235–251 (2014).
325. de Marco, A. Recombinant expression of nanobodies and nanobody-derived immunoreagents. *Protein Expression and Purification* **172**, 105645 (2020).
326. Singh, P. *et al.* Effect of Signal Peptide on Stability and Folding of *Escherichia coli* Thioredoxin. *PLoS ONE* **8**, 63442 (2013).
327. Schlegel, S. *et al.* Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). *J Mol Biol* **423**, 648–659 (2012).
328. Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial Cell Factories* **4**, 1 (2005).
329. Kiefhaber, T., Rudolph, R., Kohler, H.-H. & Buchner, J. Protein Aggregation in vitro and in vivo: A Quantitative Model of the Kinetic Competition between Folding and Aggregation. *Nat Biotechnol* **9**, 825–829 (1991).
330. Veggiani, G. & de Marco, A. Improved quantitative and qualitative production of single-domain intrabodies mediated by the co-expression of Erv1p sulfhydryl oxidase. *Protein Expr Purif* **79**, 111–114 (2011).
331. Lobstein, J. *et al.* SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories* **11**, 1–16 (2012).
332. Goldman, E. R., Liu, J. L., Zabetakis, D. & Anderson, G. P. Enhancing Stability of Camelid and Shark Single Domain Antibodies: An Overview. *Front Immunol* **8**, 865 (2017).
333. Kanthala, S., Mill, C. P., Riese, D. J., Jaiswal, M. & Jois, S. Expression and purification of HER2 extracellular domain proteins in Schneider2 insect cells. *Protein Expr Purif* **125**, 26–33 (2016).
334. Roland, J. *et al.* Role of the intracellular domains of CXCR4 in SDF-1-mediated signaling. *Blood* **101**, 399–406 (2003).
335. Chenal, A. *et al.* Does fusion of domains from unrelated proteins affect their folding pathways and the structural changes involved in their function? A case study with the diphtheria toxin T domain. *Protein Engineering, Design and Selection* **15**, 383–391 (2002).
336. Collinet, B. *et al.* Functionally Accepted Insertions of Proteins within Protein Domains. *Journal of Biological Chemistry* **275**, 17428–17433 (2000).
337. Keeble, A. H. *et al.* Evolving Accelerated Amidation by SpyTag/SpyCatcher to Analyze Membrane Dynamics. *Angewandte Chemie International Edition* **56**, 16521–16525 (2017).

## Appendix: Additional experimental data for Chapter 3

---

This appendix details additional experiments and protocols developed for the mRNA display library selections in Chapter 3. As the nanobody selections did not progress beyond soluble targets, these methods were ultimately not used; however, they are detailed here for possible future applications.

### A.1 Results and discussion

#### *A.1.1 Efforts toward nanobody expression optimization prior to library selection*

In anticipation of subsequent single clone screening of recovered binders, we considered various approaches for large-scale recombinant expression of nanobodies. Unlike some protein scaffolds such as DARPin, nanobodies have two conserved cysteine residues that create a disulfide bond within the VHH framework that is critical for proper folding and thermal stability.<sup>321</sup> With *E. coli* expression, the cytoplasm is a naturally reducing environment that does not allow disulfide bonds to form.<sup>322</sup> Special requirements are needed for subsequent refolding of disulfide bond-containing proteins to prevent mismatched disulfides<sup>323</sup>; in addition, post-expression refolding using methods such as urea gradients is generally a time-consuming, trial-and-error process.<sup>324</sup> Thus, we tested two different approaches to express the nanobody, primarily focusing on proper disulfide bond formation during expression.

In addition to establishing a working protocol for nanobody expression, we also sought to evaluate the expression, stability, and functionality for nanobody scaffolds in which the internal Methionine position was substituted (M83L) to enable single pMet positions at strategic positions around the binding interface for photocrosslinking. For these test expressions, we used a wild-type (WT) MBP-binding nanobody found in the literature (anti-MBP2)<sup>134</sup>, along with a variant in which we removed an internal methionine



position (M83L). We also included another mutant (M83L and V112I) to investigate a potential cooperative effect with the M83L mutation. These clones are here forth referred to as anti-MBP2 wildtype (aMBP2 WT), aMBP2 removed methionine (aMBP2 RM), and aMBP2 removed methionine with additional isoleucine mutation (aMBP2 RMI), respectively.

As an initial approach, we turned to a conventional method for recombinant nanobody expression by secreting the protein into the *E. coli* periplasm. The periplasm is an oxidizing/non-reducing environment that favors disulfide bond formation, and may be suitable for nanobody expression.<sup>325</sup> To this end, we N-terminally fused a pelB leader sequence to the nanobody, which directs the protein out of the cytoplasm and into the bacterial periplasm before being removed by a signal peptidase in the periplasm.<sup>326</sup> We first tested a tunable expression system using the Lemo21(DE3) expression strain, an engineered BL21(DE3) strain containing a pLysS plasmid under the control of a rhamnose promoter, which allows for lysozyme-controlled inhibition of T7 RNA polymerase.<sup>327</sup> With this system, the addition of L-Rhamnose suppresses the transcription of the protein of interest, thereby allowing the user to decrease the metabolic strain from recombinant expression of a difficult-to-fold protein. We reasoned that reduced induction levels may allow for higher amounts of soluble protein due to lower cellular protein concentration, which can favor folding.<sup>328</sup> Following protein expression and cell harvest, we performed a hypertonic step using a buffer with ethylenediaminetetraacetic acid (EDTA) and sucrose to collect the periplasmic fraction by weakening and destabilizing the outer membrane.<sup>325</sup> Then, we performed a hypotonic step with a MgSO<sub>4</sub> buffer to break the outer membrane and collect the cytoplasmic fraction.

We tested various concentrations (0, 50, 100, 200, 400  $\mu$ M) of L-Rhamnose and two expression temperatures (20 °C and 30 °C) to test different levels of suppression.

Generally, aggregation is favored at higher temperatures and reduced growth conditions (e.g., at decreased temperature) can result in increased yield of soluble protein.<sup>329</sup> Surprisingly, the condition with no L-Rhamnose produced the most amount of protein in whole-cell SDS-PAGE, which suggested that the growth suppression from decreased temperature is sufficient for expression and/or nanobody expression is more tolerated than we assumed previously (Figure A-1A). Indeed, recent literature has demonstrated high expression and functional folding of nanobodies even in the cytoplasm<sup>325</sup>, some with the aid of solubility tags or co-expression of helper proteins like oxidases.<sup>330</sup> While the expression yielded some protein, the purified fraction had very low yield, incomplete pelB leader sequence cleavage, and several contaminant bands (Figure A-1B). Since the periplasmic space of *E. coli* is limited in comparison to the cytoplasm, the expression quantity may be too low for high yield requirements.<sup>322</sup> Thus, we turned to a different strain to try to achieve higher expression yields.

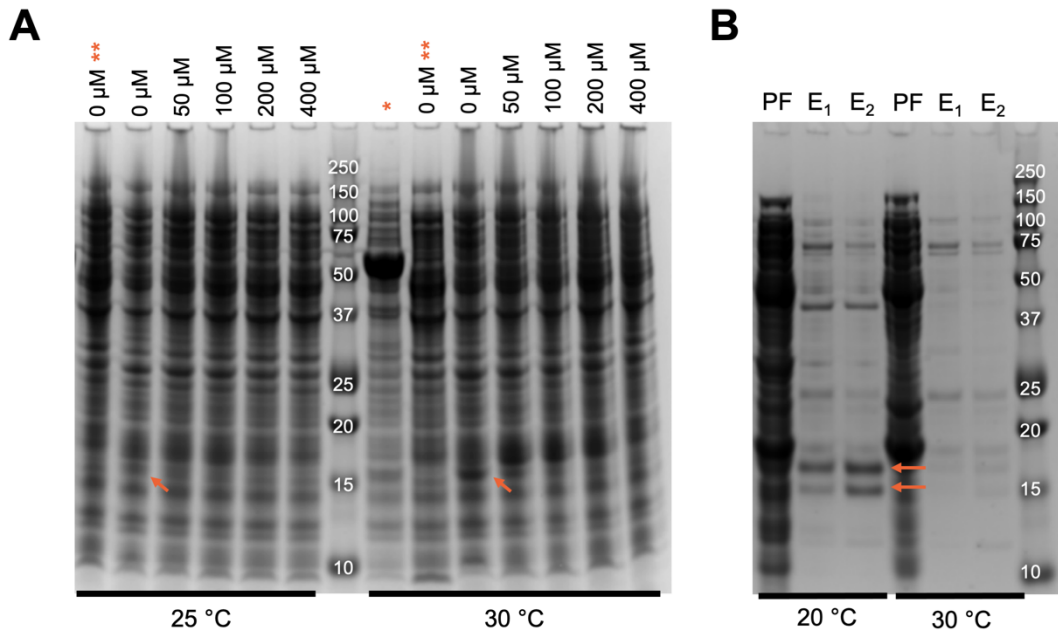


Figure A-1. Comparison of nanobody expression conditions.

(A) Whole-cell SDS-PAGE of Lemo21 expression at 25 °C (left) or 30 °C (right) with increasing L-Rhamnose concentration (listed at top of gel) for aMBP2 WT nanobody. Double asterisk (\*\*) is uninduced condition, and single asterisk (\*) is an induced control (MBP-aMBP2 construct). Arrow indicates protein expression in induced sample with 0 μM L-Rhamnose. (B) Periplasmic

fraction (PF) and elution fractions (E<sub>1</sub> and E<sub>2</sub>) of periplasmic expression and purification of aMBP2 WT showed low protein yield in 20 °C expression condition, with almost no expression when expressed at 30 °C. Eluate fraction also shows two bands (noted with two arrows), one corresponding to nanobody with pelB cleavage, and the other with incomplete pelB cleavage.

We tested another engineered strain, T7 SHuffle (from New England Biolabs) which contains mutations that minimize the cytoplasmic reductive pathways in *E. coli*, allowing formation of disulfide bonds in the cytoplasm.<sup>331</sup> Compared to previous approaches, the whole cell SDS-PAGE indicated higher yield for all three clones (WT, RM, RMI). The cytoplasmic fraction was collected and purified using IMAC purification. While contaminant bands were still present, much of the elution contained recombinant protein. The overall yield was 10.9 mg/L, 11.3 mg/L and 3.8 mg/L for WT, RM, and RMI, respectively (determined by absorbance at 280 nm); most importantly, the WT and RM mutants had similar yields, which suggested that removing a conserved internal Methionine within the framework did not reduce nanobody expression (Figure A-2A). The 3-fold reduction in yield for RMI (RM with V112I) suggested that the native V112 position may be necessary for higher expression.

We tested the functionality of the expressed proteins by performing a bead pull-down experiment. For these experiments, polystyrene beads were functionalized with MBP, allowed to incubate with the T7 SHuffle-expressed proteins, and pelleted. The supernatant fractions of the pelleted beads were run on an SDS-PAGE for band analysis. Comparison to negative controls in which the beads were not functionalized with MBP showed MBP-specific pulldown of all aMBP2 nanobodies, as indicated by a decrease in protein in the supernatant (Figure A-2B). In a follow-up experiment, we also tested the stability of the nanobody by first incubating the protein at higher temperatures (37 °C and 45 °C) for 0-6 hours and subsequently performing the same pulldown experiment. The solution incubated at 37 °C still retained much of its binding even after 6 hours, while the

solution at 45 °C had diminished binding (i.e., functionality) even after 1 hour, decreasing with longer incubations, as expected (Figure A-2C). Altogether, our results suggest that the SHuffle-expressed proteins have proper disulfide formation, which enabled functional binding activity. As expected, the nanobody remains stable and functional at an elevated temperature (37 °C), which is a desirable feature for physiological applications. Beyond the scope of our study, additional protein engineering strategies like the addition of non-canonical disulfide bonds and point mutations have successfully demonstrated improved biophysical properties like increased melting temperature and refolding ability, alluding to the robustness and utility of the nanobody scaffold.<sup>332</sup>

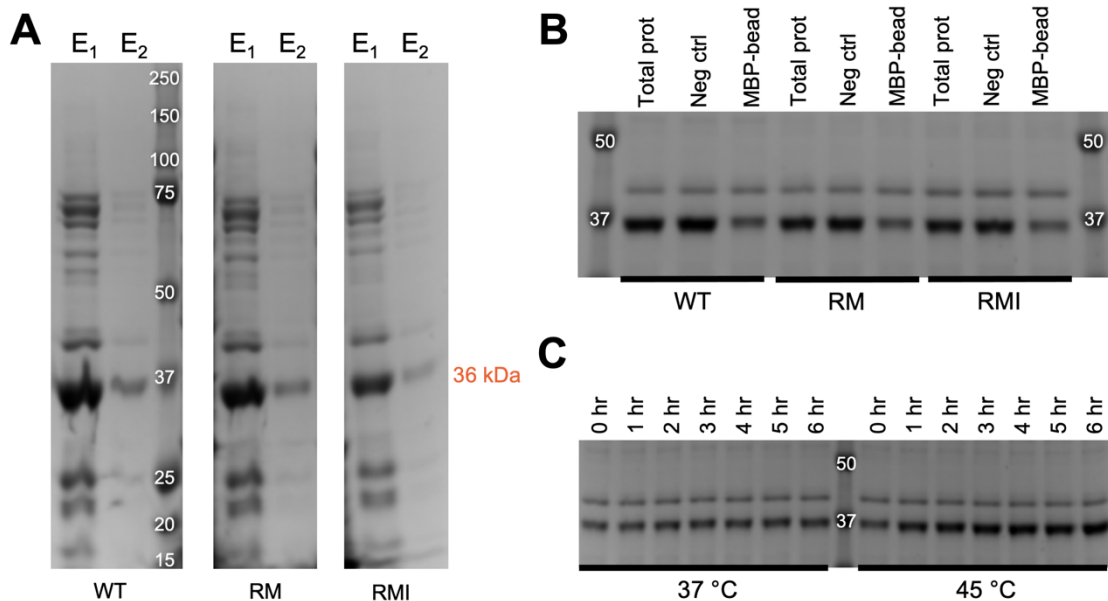


Figure A-2. Nanobody expression with T7 SHuffle and binding assay.

(A) Comparison of eluted fractions (E<sub>1</sub> and E<sub>2</sub>) for aMBP2 WT, RM, and RMI mutants expressed in T7 SHuffle *E. coli*. Contaminant bands are visible, but protein yield is much higher than previous Lemo21 expression. (B) Pull-down assay with MBP-immobilized bead (MBP-bead) or non-target immobilized bead (Neg ctrl), with total protein loaded into the pull-down reaction as reference (Total prot). All variants show MBP-specific pull-down, suggesting functional activity. (C) Similar assay to (B) for aMBP2 WT, but with pre-incubation of the aMBP2 nanobody at elevated temperatures. Incubation at 37 °C seems to retain binding activity even after a 6-hour incubation, while incubation at 45 °C considerably decreases functional binding.

*A.1.2 Preparation of cell surface targets for biopanning and protocol development for C-terminal tag pulldown onto solid support*

For subsequent selection against therapeutically relevant cell-surface targets, HEK293 cells were transfected with C-terminally tagged fusion proteins. To this end, we used two model surface proteins with varying epitope accessibility, HER2, a single-pass transmembrane receptor, and chemokine receptor 4 (CXCR4), a seven-pass-transmembrane GPCR, fused to SNAP for covalent pulldown onto BG-beads. While HER2 has a large extracellular domain<sup>333</sup> that may confer higher photocrosslinking efficiency, CXCR4 has a more minimal extracellular domain<sup>334</sup> that may better challenge the ability of our photocrosslinking selection approach. Cells were transfected with the transfection reagent, Lipofectamine, before being harvested with a non-enzymatic solution (Versene), and the expression was evaluated using immunostaining and flow cytometry. Both transfected targets showed functional expression of cell surface proteins (Figure A-3).

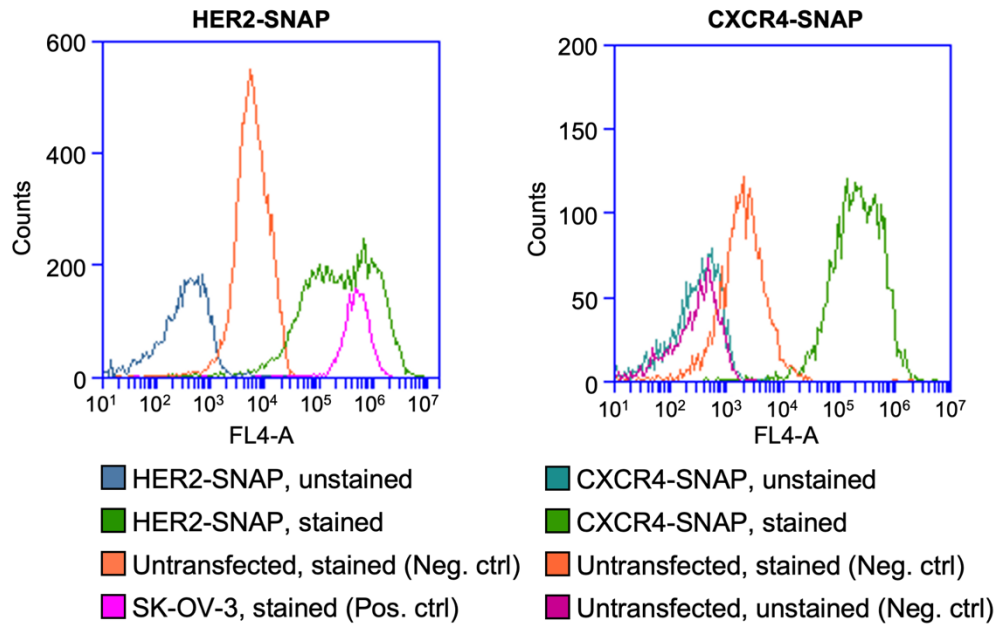


Figure A-3. Validation of cell-surface target expression by flow cytometry. Both HER2 and CXCR4 transfections show surface expression even with a C-terminal fusion to a SNAP-tag. Comparison of HER2 expression to SK-OV-3, a HER2-overexpressing cell, shows a bimodal distribution of HER2 expression for HEK293 transfected cells, which may be

due to reduced expression efficiencies as caused by the C-terminal SNAP tag (e.g., SNAP-tag forming intramolecular disulfide due to thiol group). CXCR4 expression was lower overall, which is likely due to the complexity of the protein compared to HER2.

After confirming successful membrane expression of the transfected targets, the membrane was enriched using a commercially available membrane protein extraction kit (Mem-PER) and pulled down onto BG-functionalized polystyrene beads via the C-terminal-SNAP tag. The supernatant from the pulldown incubation was then labeled using BG-Alexa Fluor 647 to identify remaining target proteins in solution, and the solution was run on SDS-PAGE and imaged using the Alexa Fluor 647 channel. Band analysis indicated decent enrichment of the target proteins in the membrane fraction, with minimal protein present in the cytosolic fraction (Figure A-4). Further, the experimental samples showed successful pulldown of the targets within the membrane fraction, as suggested by a decreased band intensity. Interestingly, the HER2 samples exhibited cleaner migration within the SDS-PAGE gel while the CXCR4 samples were more smeary and had less distinct bands corresponding to the predicted molecular weight, which aligns with the structural complexity of the proteins. Although we were able to retain binding activity with a protein tag in this experiment, possible unfolding of the N-terminal membrane receptor upon membrane isolation may also affect the binding ability of the C-terminal affinity tag, as protein unfolding of one fusion partner may destabilize the other.<sup>335</sup> However, this effect is most likely highly protein dependent (some proteins can readily accept insertions without reduced function<sup>336</sup>) and may be mitigated by using peptide affinity tags which are intrinsically not structured (e.g., SpyTag<sup>337</sup>).

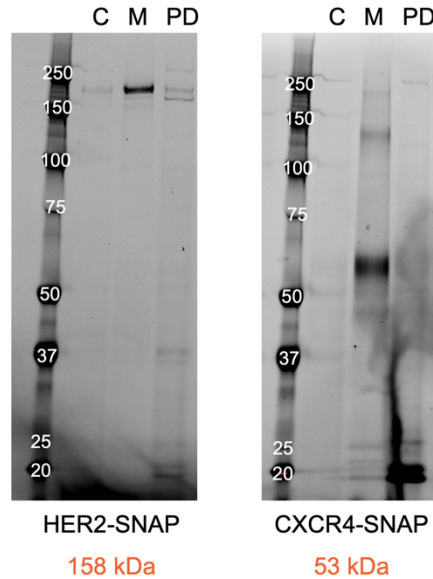


Figure A-4. Pull-down assays with enriched membrane fractions.

Comparison of cytoplasmic (C) and membrane (M) fractions from Mem-PER protocol shows most of the Alexa647-labeled HER2 and CXCR4 in the membrane fraction. Pull-down (PD) experiment demonstrates efficient pull-down of target proteins with the addition of BG-beads.

## A.2 Materials and Methods

### A.2.1. Plasmids for cell transformations and transfections

Plasmids for both *E. coli* transformations and HEK293 cell transfections were made by restriction digest and ligation reactions. The mammalian transfection plasmids were purchased from Addgene (#98942 and #39321). Plasmids were purified using the QIAprep Spin Miniprep and Plasmid Maxi Kits (Qiagen) and sequence-verified using Sanger sequencing.

### A.2.2. *Lemo21* expression test

*E. coli* Lemo21(DE3) (NEB) cells were transformed with plasmids and grown overnight on LB/carb/chloramphenicol agar plates at 37 °C. Individual colonies were picked and grown in 5 mL of LB/carb/chloramphenicol overnight at 37 °C, then diluted in Terrific Broth (TB) (Apex BioResearch) with carb and chloramphenicol, and with or without L-Rhamnose (Millipore Sigma) to an  $OD_{600} = 0.05$  and grown for 2 hours at 37 °C in a

microbial shaker. At  $OD_{600} = 0.6$ , cultures were placed on ice for 5 min before expression was induced with 0.4 mM IPTG. Cells were grown for an additional 6 hours at 25 °C or 30 °C before samples were taken for whole-cell SDS-PAGE (described in Section 5.7.2).

### *A.2.3. Lemo21 expression, periplasmic extraction, and purification*

For preparative protein expression, methods described above (Section A.2.2) were used with a few modifications. First, cells were grown in 150 mL TB with carb/chloramphenicol (no L-Rhamnose) and grown overnight (14 - 16 hours) at 25 °C. Cells were harvested at  $4,500 \times g$  for 20 min at 4 °C and pellets were resuspended in 8 mL of TBS (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM TCEP). Next, the cell suspension was split into four 2 mL tubes and centrifuged at  $10,000 \times g$  for 10 min at 4 °C. The supernatant was removed, and the pellet was resuspended in 0.75 mL of ice-cold resuspension buffer (RB) (30 mM Tris-HCl, pH 8.0). Then, 1 mL of ice-cold 2X periplasmic extraction buffer (PEB) (30 mM Tris-HCl, pH 8.0, 40% w/v sucrose, 2 mM EDTA) was added and mixed by pipetting. The suspension was incubated for 1 hour at 4 °C on a tube rotator. To pellet, 0.5 mL of water was added, and the tubes were centrifuged at  $10,000 \times g$  for 15 min at 4 °C, rotating the tube 180° every 5 min. The supernatant was removed, and 1.5 mL of ice-cold 5 mM  $MgSO_4$  (Millipore Sigma) was added and mixed by pipetting. The suspension was incubated for 30 min at 4 °C on a tube rotator before pelleting at  $10,000 \times g$  for 10 min at 4 °C. The supernatant, containing the protein, was rebuffed with concentrated Tris-HCl, pH 7.5 and NaCl to 50 mM and 150 mM final concentration, respectively.

For IMAC purification, the supernatant was diluted 10-fold with TBS and Triton-X (Millipore Sigma) was added to 0.1% v/v final concentration. Then, it was loaded onto a gravity column with 0.5 mL Ni-NTA resin (Qiagen) equilibrated with 10 column volumes (CVs) of W-T150 (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole, and 1 mM



DTT). After loading the supernatant, the column was washed with 10 CV of W-T150, 5 CV of W-T800 (W-T150 with 800 mM NaCl), and 5 CVs of W-T150. To elute, 0.5 CVs of E-T150 (W-T150 with 250 mM imidazole) was added (not collected), and two elution fractions were collected with 1 CV. Fractions were analyzed by SDS-PAGE (described previously in Section 5.7.2).

#### A.2.4. T7 SHuffle expression and purification

SHuffle T7 Express Competent *E. coli* (NEB) cells were transformed with plasmids and grown overnight following manufacturer's protocol. Individual colonies were picked and grown in 10 mL of TB/carb overnight at 30 °C, then diluted in 50 mL TB/carb to an  $OD_{600} = 0.05$  and grown for 4 hours at 30 °C in a microbial shaker. At  $OD_{600} = 0.6$ , cultures were placed on ice for 5 min before expression was induced with 0.4 mM IPTG. Cells were grown overnight at 20 °C before being harvested at 4,000 x *g* for 20 min at 4 °C. The supernatant was removed, and the pellet was flash frozen in liquid nitrogen and stored at -80 °C. Purification was performed using methods detailed in Section 3.7.1, and proteins were rebuffed into TBS.

#### A.2.5. Pulldown of SHuffle-expressed proteins onto MBP-beads

MBP-immobilized beads were made by a two-step reaction. First, recombinantly expressed SNAP-MBP was biotinylated via BG-biotin as follows:

Reagent	Final concentration
DTT	1 mM
SNAP-MBP	5 μM
BG-biotin (NEB)	10 μM
TBS	Up to 500 μL

The reaction was incubated at 37 °C for 30 min, and excess BG-biotin was removed by diafiltration using 10 kDa MWCO diafiltration columns (Amimcon).

Next 50 μL streptavidin Sepharose beads (50% slurry) (GE Healthcare) was washed 3x in 0.5 mL TBS, resuspended in 154 μL TBS, and mixed with 200 μL biotinylated

MBP. The reaction was incubated at RT for 1 hour on a tube rotator, washed 3x in TBS-T (TBS with 0.2% w/v Tween-20), and resuspended in 77  $\mu$ L TBS-T. Negative controls did not have target immobilized, but were washed and resuspended with the same protocol.

For the aMBP2 nanobody pulldown, the prepared beads were mixed with purified protein as follows:

Reagent	Final concentration
Beads	6.25% w/v
aMBP2 nanobody	4 $\mu$ M
TBS-T	Up to 20 $\mu$ L

The reaction was incubated at RT for 1 hour on a tube rotator and centrifuged on a tabletop minifuge to pellet the beads (20 sec, 3x). The supernatant was taken out for SDS-PAGE analysis.

The stability studies with aMBP2 WT were performed by incubating the protein at 37 °C or 45 °C for 1 - 6 hours (hourly samples were collected and stored on ice), then repeating the same pulldown experiment described above, with the addition of 1 mM DTT.

#### *A.2.6. Transfection procedure*

HEK293 cells (cultured as described in Section 3.7.13) were plated at a density of  $8 \times 10^5$  cells/well in a 6-well plate by diluting a suspension to  $4 \times 10^5$  cells/mL of complete media (DMEM + 10% v/v FBS), adding 1 mL, and swirling to cover the well, and adding another 1 mL dropwise in the center of the well. The plate was left at RT for 15 min to let the cells settle before carefully moving the plate in the incubator. After 21 hours of incubation, the cells were gently washed with 2 mL DPBS and replaced with 2 mL of complete media before being placed in the incubator for 1 hour. Cells were transfected with Lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's protocol with the following modifications: after addition of the DNA-lipid complex, the cells were

incubated for 5 hours at 37 °C. The media was aspirated and replaced with complete media, then incubated for an additional 48 hours.

For flow cytometry, the cells were rinsed once with pre-warmed PBS, then harvested by incubating with a non-enzymatic cell dissociation media (1X Versene, Thermo Fisher Scientific) to preserve cell surface receptors for 5 min at 37 °C. Cells were strained using a 35 µm strainer (Gibco), pelleted, and resuspended in FACS buffer (0.1% BSA, 2 mM EDTA, 0.05% sodium azide in PBS). Cells were labeled by antiCXCR4-Alexa Fluor 647 or antiHER2-Alexa Fluor 647 conjugate (R&D Biosystems) for 1 hour at 4 °C, washed 3x in ice-cold FACS buffer, and strained again through a 35 µm mesh strainer (Gibco). Fluorescence was analyzed by flow cytometry (Accuri C6, BD Biosciences).

#### *A.2.7. Pulldown of target proteins in the membrane fraction on BG-beads*

To enrich the membrane fraction, transfected HEK293 cells were collected using Versene, washed, and counted. Approximately  $3 \times 10^6$  cells were aliquoted into 15 mL tube, and membrane extraction was performed using Mem-PER following manufacturer's protocol. Meanwhile, 68 µL of SNAP-Capture magnetic beads (NEB) were washed 3x with 1 mL TBS. Approximately 50 µL of Mem-PER extraction (equivalent to  $8 \times 10^5$  cells) were added to the washed beads and incubated for 1 hour at RT on a tube rotator. Next, the beads were pelleted by centrifuging for 1 min at 11,000  $\times g$ , and 5 - 10 µL of supernatant was incubated with 10 µM of BG-Alexa647 and water up to 15 µL, and incubated for an additional 1 hour at RT. The mixture was analyzed with SDS-PAGE and imaged on the Alexa Fluor 647 channel (ChemiDoc MP, Bio-Rad).

Table A-1. List of amino acid sequences used in this study.

Construct name	Description	Amino acid sequence
aMBP2 WT	Anti-MBP2 nanobody used in expression tests (pulldown assays are fused to SNAP; mutated positions are bolded and underlined)	MKYLLPTAAAGLLLLLAAQPAMAMDIGINSDPLEVLFQ GPSQVQLVESGGSVQAGGSLRLSCAASGQIEHIGYL GWFRRQAPGKEREGVAALITYTGHYTYADSVKGRFTVS LDNAKNTVYLO <b>M</b> <u>NS</u> SLKPEDTALYCAAAEWGSQSPLT QWFYRYWGQGTQ <b>V</b> <u>T</u> VSALHHHHHH
HER2-SNAP	ORF of HER2-SNAP fusion used for HEK293 transfections	MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLLPA SPETHLDMLRHLYQGCQVVQGNLELTYLPTNASLSFL QDIQEVQGYVLIHQNVRQVPLQRLRIVRGTTQLFEDN YALAVLDNGDPLNNTTPVTGASPGGLRELQLRSLTEI LKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLID TNRSRACHPCSPMCKGSRWGESSEDCQSLTRTVACG GCARCKGPLPTDCHEQCAAGCTGPKHSDCLACLHFN HSGICELHCPALVYNTDTFESMPNPEGRYTFGASCV TACPYNLSTDVGSCTLVCPHNLQEVTAEDGTQRCEK CSKPCARVCYGLGMEHLREVRVTSANIQEFAGCKKI FGSALFLPESFDGDPASNTAPLQPEQLQVFETLEEIT GYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLT LQGLGISWLGRLRLRELGLSGLALIHNNHLCFVHTVP WDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARG HCWGPPTQCVNCSQFLRGQECVEECRVLQGLPREYV NARHCLPCHPECQPNQSVTCFGPEADQCVACAHYKD PPFVVARCPGSKPDLSPYMPIWKFPEEGACQPCPIN CTHSCVDLDDKGCPAEQRASPLTSII SAVVGILLVVV LGVVFGLIKRRQQKIRKYTMRRLLQETELVEPLTPS GAMPNQAQMRIKTELRLKVKVLGSGAFGTVYKGIWI PDGENVKIPVAIKVLENTSPKANKEILDEAYVMAGV GSPYVSRLGLICTSTVQLVTLQMPYGLLDHVRNRN RLGSDLLNWCMIKAGMSYLEDVRLVHRDLAARNV LVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKW MALESILRRRFTHQSDVWSYGVTWELMTFGAKPYDG IPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMID SECRPRFRELVSEFSRMARDPQRFVVIQNEIDLGPASP LDSTFYRSLLEDDDMGDLVDAEYLVPPQGGFFCPDPA PGAGGMVHHRHRSSTRSGGDLTLGLEPSEEEAPRS PLAPSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLO RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPDPVRP QPPSPREGPLPAARPAATLERPKTLPSPGKNGVVKDV FAFGAVENPEYLTPOGGAAPQHPPPAFSPAFDONLY YWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPTVGS GSGSGGGKLMKDCCKMRTTLDSPGLKLELSGCEQG LHEIIFLGKGTSAADAVEVPAPAAVLGGPEPLMQATA WLNAYFHQPEAIEEFVVPALHHPVFQQESFTRQVLWK LLKVVKFGEVISYSHLAALAGNPAATAAVKTAALSGNP VPILIPCHRVRVQGDLDVGGYEGGLAVKEWLLAHEGHR LGKPGLG
CXCR4-SNAP	ORF of CXCR4-SNAP fusion used for HEK293 transfections	MEGISITYSDNYTEEMSGDYDSMKEPCFREANANFN KIFLPTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTD KYRLHLSVADLLFVITLFPFVAVDANWYFGNFLCKA VHVITYTNLYSSVLI LAFISLDRYLAI VHATNSQRPR KLLAEKVVVYGVWIPALLLTIPDFIFANVSEADDRYI CDRFYPNDLWVVVFQFHIMVGLILPGIVILSCYCI I ISKLSHSGKHQKRKALKTTVILILAFFACWLPHYI GI SIDSFILLEI IKQCEFENTVHKWISITEALAFFHCC LNPILYAF LGAKFKTSAQHALTSVSRGSSLKILSKGK RGGHSSVSTESESSFHSSGSGSKLMKDCCKMRTT

		LDSP LGKLELSGCEQGLHEIIFLGKGTSAADAVEVPA PAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALH HPVFQQESFTRQVLWKLKVVKFGEVISYSHLAALAG NPAATAAVKTALSGNPVPIIPCHRVVQGDLDVGGYE GGLAVKEWLLAHEGHR LGK PGLG
--	--	--