

**High-Throughput Particle Display Screening of RNA-Protein  
Interactions**

---

**Prime Editor-Mediated Programmable Insertion of UAAs into  
Endogenous Proteins**

Cristina M. Cheng

A thesis  
submitted to the Faculty of  
the department of chemistry  
in partial fulfillment  
of the requirements for the degree of  
Master of Science

Boston College  
Morrissey College of Arts and Sciences  
Graduate School

December 2021



## Abstract

### High-Throughput Particle Display Screening of RNA-Protein Interactions

Cristina M. Cheng

Thesis Advisor: Professor Jia Niu, Ph.D.

RNA-protein binding interactions have essential roles in many biological processes including transcriptional and translational control; thus, it is important to quantify the binding affinities of these biological complexes through functional binding assays. Although conventional binding assays have provided significant insight to these dynamic networks, they generally provide a relatively low throughput for a limited number of samples. To overcome the limitations of these conventional binding assays to study RNA-protein binding interactions, we propose to develop an *in vitro*, high-throughput particle display-based for RNA aptamer screening of RNA-protein complexes for the subsequent identification and characterization of novel RNA aptamers that influence protein binding. With this technique, we will be able to profile large numbers of binding events based on binding-induced fluorescence-enhancement for a more holistic understanding of the corresponding RNA-protein network. So far, we have confirmed that this particle display-based technique can be used to estimate the binding affinity of the well-characterized MS2-MCP model system, and plan to advance this technique to screen a library of MS2 variants for mutational analysis.

## Abstract

### **Prime editor-mediated programmable insertion of UAAs into endogenous proteins**

Cristina M. Cheng

Thesis Advisor: Professor Jia Niu, Ph.D.

The introduction of unnatural amino acids (UAAs) to endogenous cell surface proteins for site-specific bioconjugation reactions allows for the incorporation of clickable, fluorescent handles *in vivo*; however, the transient expression of proteins harboring UAAs is limited by its transfection efficiency. Thus, we propose to employ prime editors and tRNA/aminoacyl-tRNA synthetase technologies to introduce an UAA to endogenous proteins for downstream bioconjugation applications. Briefly, we propose to stably incorporate a stop codon into mammalian cells by prime editing which will be confirmed with a reporter system, such that this stop codon can mediate the introduction of an UAA through the associated tRNA/aminoacyl-tRNA synthetase technology. By permanently introducing a bioorthogonal, clickable handle onto an endogenous protein, its cellular signaling and localization patterns can be monitored *in vivo* for further classification of the behaviors of these proteins. So far, we identified a promising fluorescent reporter construct to validate the introduction of a stop codon into the mammalian genome by prime editing.

## Table of Contents

Acknowledgements.....	iii
List of Abbreviations.....	iv
List of Figures.....	vi
List of Tables.....	vii
<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1. RNA-protein complexes.....	1
1.2. <i>In vitro</i> RNA-RBP binding assays .....	2
1.2.1. Physical separation-based methods.....	3
1.2.2. Solution-based methods.....	5
1.3. RNA aptamer screening.....	7
1.3.1. SELEX.....	8
1.3.2. Particle display.....	8
1.4. References.....	9
<b>Chapter 2. High-Throughput Particle Display Screening of RNA-Protein Interactions.....</b>	<b>12</b>
2.1. Introduction.....	12
2.2. Results/Discussion.....	14
2.2.1. Aim 1. To generate aptamer particles expressing RNA aptamer of interest.....	14
2.2.2. Aim 2. To purify and label cognate RNA-binding protein with fluorescent dye.....	20
2.2.3. Aim 3. To perform binding reactions and measure the fluorescence emission profile to quantify the binding affinity.....	25
2.3. Future directions.....	29
2.4. Materials and methods.....	31
2.5. References.....	41
<b>Chapter 3. Introduction.....</b>	<b>44</b>
3.1. CRISPR-Cas systems.....	44

3.2. CRISPR-Cas technologies.....	46
3.2.1. Base editing.....	47
3.2.2. Prime editing.....	47
3.3 References.....	48
<b>Chapter 4. Prime editor-mediated programmable insertion of UAAs into endogenous proteins.....</b>	<b>50</b>
4.1. Introduction.....	50
4.2. Results/Discussion.....	52
4.2.1. Aim 1. To validate the insertion of an amber stop codon by expressing a fluorescent reporter in a stable HEK293T cell line.....	52
4.3. Future directions.....	57
4.4. Materials and methods.....	57
4.5. References.....	59
<b>Appendices.....</b>	<b>60</b>
Appendix 1. Supporting information for high-throughput particle display screening of RNA-protein interactions.....	60
Appendix 2. Supporting information for prime editor-mediated programmable insertion of UAAs into endogenous proteins.....	64

## Acknowledgements

First, I would like to sincerely thank my research advisor, Dr. Jia Niu, for welcoming me into his research lab and introducing me to molecular biology. His support for his students and excitement for our research is inspiring, and I am so grateful for his guidance throughout my graduate career.

I would like to also thank all the lab members for their support, both inside and outside the lab. It was a privilege to work with such kind and motivated individuals, and I wish you the best throughout your graduate and post-graduate career. I would especially like to thank Qiwen for being a wonderful mentor and friend.

Additionally, I would like to thank the Chemistry department for everything they do for their graduate students. I appreciate your kindness and support and the sense of community that Merkert Chemistry Center has.

Finally, I would like to thank my family for always being there for me. I would not be where I am today without your unwavering support and encouragement.

## List of Abbreviations

RBP – RNA-binding protein  
RNP – Ribonucleoprotein  
CLIP – Crosslinking and immunoprecipitation  
L – Ligand  
R – Receptor  
L · R – Ligand-receptor complex  
 $k_{on}$  – Association rate constant  
 $k_{off}$  – Dissociation rate constant  
 $K_a$  – Association constant  
 $K_d$  – Dissociation constant  
 $B_{max}$  – Specific binding maximum  
EMSA – Electrophoretic mobility shift assay  
FBA – Filter binding assay  
FP – Fluorescence polarization  
SPR – Surface plasmon resonance  
BLI – Biolayer interferometry  
SELEX – Systematic evolution of ligand by exponential enrichment  
PCR – Polymerase chain reaction  
GRAP – Gene-linked aptamer particle  
MCP – MS2 coat protein  
nt – Nucleotide  
dT – Deoxythymidine  
EDC – 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride  
dA – Deoxyadenosine  
AF647 – Alexa Fluor 647  
APC – Allophycocyanin  
bp – Base pair  
MW – Molecular weight  
kDa – Kilodalton  
His-tag – Polyhistidine tag  
IPTG – Isopropyl  $\beta$ -D-1-thiogalactopyranoside  
Ni-NTA – Nickel-nitrilotriacetic acid  
SEC – Size-exclusion chromatography  
PAGE – Polyacrylamide gel electrophoresis  
LC/MS – Liquid chromatography-mass spectrometry  
SDS-PAGE – Sodium dodecyl sulphate-polyacrylamide gel electrophoresis  
DOL – Degree of labeling  
FP – Forward primer  
RP – Reverse primer  
NaOH – Sodium hydroxide  
NaCl – Sodium chloride  
DMSO – Dimethyl sulfoxide

rpm – Rotations per minute  
MES – 2-(n-morpholino)ethanesulfonic acid  
HCl – Hydrochloride  
EDTA – Ethylenediaminetetraacetic acid  
dNTP – Deoxynucleoside triphosphate  
kb – kilobase  
TAE – Tris-acetate-EDTA  
EtBr – Ethidium bromide  
MgCl<sub>2</sub> –Magnesium chloride  
DTT – Dithiothreitol  
rNTP – Ribonucleoside triphosphate  
NaOAc – Sodium acetate  
EtOH – Ethanol  
TE – Tris-EDTA  
TBE – Tris-borate-EDTA  
LB – Luria broth  
NaHPO<sub>4</sub> – Sodium phosphate  
MWCO – Molecular weight cut-off  
NaHCO<sub>3</sub> – Sodium bicarbonate  
CRISPR – Clustered regularly interspaced short palindromic repeats  
Cas – CRISPR-associated proteins  
crRNA – CRISPR RNA  
PAM – Protospacer adjacent motif  
tracrRNA – Trans-activating CRISPR RNA  
DSB – Double-stranded break  
sgRNA – single-guide RNA  
NHEJ – Non-homologous end joining  
HDR – Homology directed repair  
MMEJ – Microhomology-mediated end joining  
Indel – Insertion or deletion  
nCas9 – Nickase Cas9  
dCas9 – Catalytically-dead Cas9  
CBEs – Cytosine base editors  
ABEs – Adenine base editors  
PE – Prime editors  
PE-pegRNA – Prime editing complex  
RT – Reverse transcriptase  
pegRNA – Prime editing guide RNA  
PBS – Primer binding site  
UAA – Unnatural amino acid  
PTM – Post-translational modification  
PEI – polyethylenimine

## List of Figures

- Figure 1. Physical separation-based methods used to analyze RNA-protein interactions.
- Figure 2. Solution-based methods used to analyze RNA-protein interaction.
- Figure 3. Gene-linked aptamer particle (GRAP) display method for RNA aptamer library screening of RNA-protein binding interactions.
- Figure 4. Bacteriophage MS2 viral genome diagram and crystal structure of MS2-MCP system with MS2 hairpin structure.
- Figure 5. Conversion of carboxylic acid beads to oligo(dT)<sub>25</sub>-modified beads annealed to an AF647-labeled complementary oligo(dA)<sub>25</sub> probe.
- Figure 6. Primer extension and *in vitro* transcription product analysis of CC\_MS2apt\_polytail by gel electrophoresis.
- Figure 7. Conversion of oligo(dT)<sub>25</sub>-modified beads expressing the MS2 aptamer annealed to an AF647-labeled oligo that is complementary to the MS2 aptamer region.
- Figure 8. SDS-PAGE analysis of MCP-SoxS-his protein and AF647-labeled MCP-SoxS-his protein.
- Figure 9. SDS-PAGE analysis of SoxS-his protein and AF647-labeled SoxS-his protein.
- Figure 10. Binding experiment of MS2-MCP system.
- Figure 11. Negative control experiment for MS2-MCP system.
- Figure 12. Negative control experiment for MS2-MCP system at an increased labeled protein concentration range.
- Figure 13. Synthesis of gene-linked aptamer particles for GRAP display-based technique for RNA aptamer library screening of MS2-MCP system.
- Figure 14. Schematic overview of CRISPR-Cas systems for genome editing.
- Figure 15. Schematic overview of base editing for genome editing.
- Figure 16. Schematic overview of prime editing for genome editing.
- Figure 17. Schematic overview of the installation of an UAA into an endogenous protein using prime editing and tRNA-aminoacyl-tRNA synthetase technologies.
- Figure 18. Prime editing-mediated turn-on and turn-off fluorescent reporter system.
- Figure 19. Transfection efficiency of turn-off fluorescent reporter.
- Figure 20. Transfection efficiency of turn-on fluorescent reporter.

## **List of Tables**

Table 1. Oligonucleotides used for particle display-based experiments, and their corresponding sequences.

Table 2. Reporter plasmids constructed and tested for prime editing project, and their corresponding transfection results.

## **Chapter 1. Introduction for high-throughput particle display screening of RNA-protein interactions**

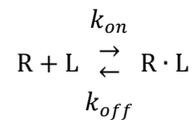
### **1.1. RNA-protein complexes**

RNA-protein interactions play essential roles in many biological functions, including gene regulation through transcriptional control, RNA processing and transport, and translational control.<sup>1,2,3</sup> RNA-binding proteins (RBPs) are modular structures that form ribonucleoprotein (RNP) complexes by recognizing and binding to RNA structural motifs with conserved structures and limited sequence variations.<sup>1</sup> These individual RNA-protein interactions often aggregate to create complex RNA-protein networks that influence many cellular processes, such that defects or regulatory inhibitors of these networks may cause damage to the cell.<sup>1</sup> Although it is vital to understand the biogenesis, proliferation, and fate of these RNA-protein networks, it is difficult to measure the dynamics of the individual or aggregated RNA-protein interactions *in vivo*. Current methods to measure the dynamics of RNA-protein interactions *in vivo* typically rely on RNA-protein crosslinking and immunoprecipitation (CLIP) to interrogate binding interaction patterns at various biological timescales that span different stages of cellular development.<sup>2,3</sup> Due to the technical challenges of determining the kinetics of RNA-protein interactions *in vivo*, several methods to analyze these interactions *in vitro* have been developed. Even though *in vitro* assays separate the biological complex of interest from its native setting, thereby removing external factors that affect the accessibility of RNA-binding domain such as the native concentrations of both binding partners or other competitive inhibitory molecules, it is important to understand the molecular behind each

RNA-protein interaction for insight into its influence on gene expression within cellular contexts.<sup>3</sup>

## 1.2. *In vitro* RNA-RBP binding assays

To understand how RNA-protein interactions contribute to these biological processes, it is important to quantify the binding interactions through binding assays. Identifying the binding affinities and rate constants of RNA-protein complexes through functional binding assays can help elucidate RNA-protein recognition mechanisms and determine strong binding partners, which is crucial to study the dynamics of these networks and their influence on biological processes.<sup>3</sup> When analyzing binding parameters of RNA-protein interactions, it is important to understand the fundamentals of the kinetics and thermodynamics of receptor-ligand interactions.<sup>4,5</sup> The association and disassociation reactions of the ligand (L) to the receptor (R) molecules are summarized below.<sup>5</sup>



Here, the forward reaction describes the complexation of the ligand to the receptor mediated by the on-rate constant,  $k_{on}$ , and the reverse reaction describes the dissociation of the receptor · ligand complex into its individual components mediated by the off-rate constant,  $k_{off}$ .<sup>5</sup> At equilibrium, the ratio of the concentration of the receptor · ligand complex is equal to the product of the concentrations of the free receptor and free ligand and is characterized as the association constant,  $K_a$ .<sup>5</sup> Likewise, the ratio of the concentrations of the free receptor and free ligand to the concentration of the receptor · ligand complex is characterized as the dissociation constant,  $K_d$ .<sup>5</sup> These relationships are shown in the equations below.

$$K_a = \frac{[L \cdot R]}{[L][R]} = \frac{k_{on}}{k_{off}}$$

$$K_d = \frac{1}{K_a} = \frac{[L][R]}{[L \cdot R]} = \frac{k_{off}}{k_{on}}$$

When the concentration of the ligand is equal to the dissociation constant, this means that half the receptors are occupied by a ligand at equilibrium.<sup>4,5</sup> If the receptors have high specificity and affinity to the ligand, the equilibrium dissociation constant will be low since it will take a lower ligand concentration to bind half the receptors.

To determine the equilibrium dissociation constant of RNA-protein binding interactions, saturation binding experiments are widely performed.<sup>4,5</sup> In these experiments, the concentration of the fluorescent, radiolabeled, or tagged ligand is varied while the number of receptors available for binding is kept constant, allowing the researcher to measure the binding interactions at equilibrium to determine the ligand concentration at which half the receptor sites are occupied ( $K_d$ ) and the ligand concentration at which all the receptor sites are occupied ( $B_{max}$ ) in the same experiment.<sup>4,5</sup> The techniques used to determine the binding affinities and other binding parameters of RNA-protein complexes are generally categorized into two groups, physical separation-based methods and solution-based methods.<sup>3</sup>

### 1.2.1. Physical separation-based methods

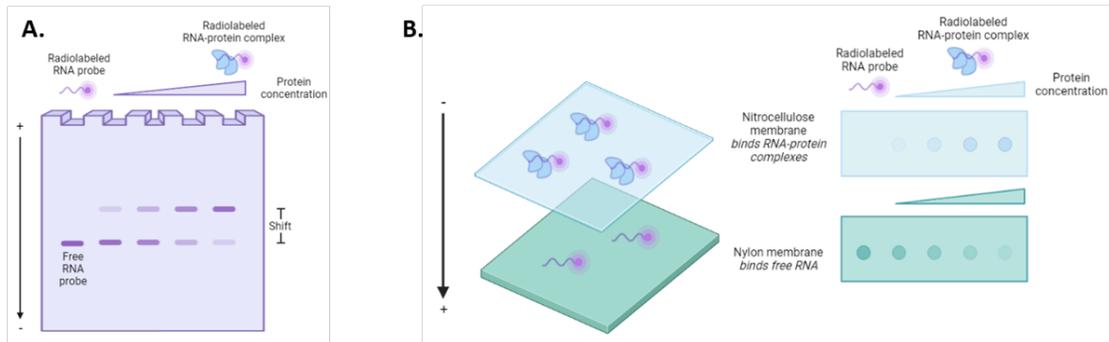
Physical separation-based methods monitor the kinetics of RNA-protein interactions by physically separating the RNA-protein complex from its free components and monitoring the changes of bound versus free states.<sup>3</sup> Electrophoretic mobility shift assays (EMSAs) and filter binding assays (FBAs) are examples of physical separation-based methods as the RNA-protein complex is physically separated from its free components (**Figure 1**).<sup>3</sup>

For a conventional EMSA protocol, the RNA of interest is radiolabeled and is incubated

with the protein sample at a range of concentrations in a binding reaction before it is subjected to a non-denaturing gel for analysis (**Figure 1A**).<sup>6,7</sup> The RNA probe is expected to have greater mobility than the RNA bound to the protein, proportional to the molecular weight, causing the sample to have slower mobility during electrophoresis and producing a migration shift relative to the free RNA probe.<sup>6,7</sup> This shift can be analyzed to quantify the signal of the bound RNA to the unbound RNA, which is then plotted against the protein concentration to determine the dissociation constant of the system.<sup>6,7</sup> This method has high sensitivity due to the radiolabeled probes and low sample consumption due to the small volumes used; however, the samples are not run at chemical equilibrium during electrophoresis, which may cause dissociation and prevent detection of complexes.<sup>6,7</sup> Further, the throughput of this method is very low, typically accommodating a limited number of samples for 10-20 gel wells.<sup>6,7</sup>

The filter binding assay is another example of a physical separation-based method in which the RNA of interest is radiolabeled and incubated with the protein sample at a range of concentrations in a binding reaction before it is subjected to a slot blot apparatus containing two membranes: a negatively charged nitrocellulose membrane on the top and a positively charged nylon membrane on the bottom (**Figure 1B**).<sup>8,9</sup> Since most proteins are positively charged, the proteins and RNA-protein complexes will bind to the negatively charged nitrocellulose membrane, whereas the negatively charged RNA probes will bind to the positively charged nylon membrane.<sup>8,9</sup> After washing and drying the membranes, they are exposed to phosphor-imaging screens to quantify the signal of the bound RNA to the unbound RNA, which is plotted against the protein concentration to determine the dissociation constant.<sup>8,9</sup> By using a slot blot procedure, this method has a

higher throughput than conventional EMSA as it can run through many samples at different protein concentrations in one reaction, but there is a possibility of dissociation during the filtering and washing steps that may interfere with the complexes.<sup>8,9</sup>



**Figure 1.** Physical separation-based methods used to analyze RNA-protein interactions. An overview of two conventional methods, **A.** electrophoretic mobility shift assay (EMSA) and **B.** filter-binding assay (FBA), are depicted above.

### 1.2.2. Solution-based methods

Solution-based methods monitor the kinetics of RNA-protein interactions by measuring a specific physical property of the RNA-protein complex that is sufficiently different from the free RNA and the free protein, so there is no need to separate the complexes.<sup>3</sup>

Fluorescence polarization (FP), surface plasmon resonance (SPR), and biolayer interferometry (BLI) are popular examples of solution-based methods (**Figure 2**).

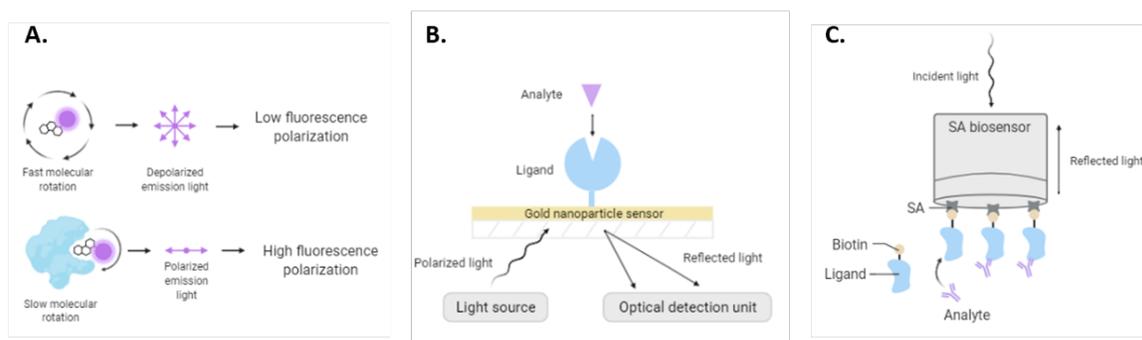
Fluorescence polarization, also known as fluorescence anisotropy, measures the change in the polarized light of fluorophores.<sup>10,11</sup> Fluorophores are typically small molecules that emit a depolarized emission light to produce a low fluorescence polarization. In these assays, the RNA is typically labeled with a fluorophore to produce a low fluorescence polarization when in its unbound state, such that once the binding protein is incubated with this fluorescently labeled RNA, the RNA-protein complex slows the molecular rotation due to its increased size.<sup>10,11</sup> When subjected to polarized light, the sample will

produce polarized emission light and high fluorescence polarization which can be analyzed with respect to the free RNA probe (**Figure 2A**).<sup>10,11</sup> The binding interaction can be calculated by taking the difference in anisotropy of the partially bound, unbound, and fully bound states by titrating the two binding partners.<sup>10,11</sup> This method has a relatively high-throughput and can directly measure the  $k_{on}$  but requires high fluorophore concentrations for signal detection and cannot easily measure  $k_{off}$ .<sup>10,11</sup>

Surface plasmon resonance is another solution-based method to monitor the kinetics of RNA-protein interactions that is label-free and allows for real-time binding analysis.<sup>12,13</sup> In typical SPR experiments, the ligand of interest is immobilized to a sensor chip, such that when the analyte passes through the flow cell during the association phase, it binds to the ligand immobilized on the sensor chip surface (**Figure 2B**).<sup>12,13</sup> The binding of the analyte to the ligand will cause the signal to increase until equilibrium is reached, and then the bound analyte is washed from the ligand receptors during the dissociation phase until equilibrium conditions are reached again.<sup>12,13</sup> The gold nanoparticles absorb the emitted light to create a baseline absorbance peak, such that when binding events or changes in the refractive index occur, the center of the absorbance peak will shift relative to the baseline.<sup>12,13</sup> At different analyte concentrations, a set of binding curves can be collected and fit with binding models to measure the  $k_{on}$ ,  $k_{off}$ , and  $K_d$  constants.<sup>12,13</sup> This technique is widely used since it is label-free and provides real-time measurements, but its low-throughput limits its translation to large-scale binding experiments.<sup>12,13</sup>

Another solution-based method to detect and analyze RNA-protein interactions is bilayer interferometry. Bilayer interferometry analyzes the interference pattern of depolarized light reflected from the layer of immobilized protein on the biosensor tip and

an internal reference layer.<sup>14,15</sup> In a typical BLI experiment, the binding between a ligand immobilized on the biosensor tip surface and an analyte in solution produces an increase in optical thickness at the biosensor tip, which results in a wavelength shift (**Figure 2C**).<sup>14,15</sup> Having a much higher throughput than SPR and allowing for real-time measurements, BLI is becoming more popular, but its high sample consumption is a significant drawback.<sup>14,15</sup>



**Figure 2.** Solution-based methods used to analyze RNA-protein interactions. An overview of three conventional methods, **A.** fluorescence polarization (FP), **B.** surface plasmon resonance (SPR), and **C.** biolayer interferometry (BLI), are depicted above.

### 1.3. RNA aptamer screening

Aptamers are short, single-stranded DNA or RNA oligonucleotides that non-covalently bind to target biomolecules, including proteins, peptides, carbohydrates, and small molecules, with high affinity and specificity.<sup>16</sup> The target recognition and binding capabilities of aptamers are influenced by the three-dimensional conformation of the aptamer, along with its hydrophobic interactions, intercalation, and base-stacking between the aptamer and its cognate binding partner.<sup>16</sup> Systematic evolution of ligand by exponential enrichment (SELEX) is the conventional method for aptamer discovery, but other methods have been recently developed to overcome some of the challenges SELEX-based methods encounter.

### **1.3.1. SELEX**

Systematic evolution of ligand by exponential enrichment (SELEX) is a combinatorial chemistry technique that produces single-stranded aptamers that specifically bind to their corresponding ligand.<sup>17</sup> Although there are many variations of SELEX for selection, the general procedure involves the synthesis of a large aptamer library consisting of randomized sequences flanked by constant regions for amplification.<sup>17,18,19</sup> The library is then exposed by a target ligand, such that the sequences that do not bind to the ligand of interest are removed via affinity chromatography or bead capture, and the sequences that bind to the ligand of interest are eluted from the purification method and amplified by polymerase chain reaction (PCR) in preparation for subsequent rounds of selection with increased stringency.<sup>17,18,19</sup> SELEX libraries typically have  $10^{12}$  sequences of randomized sequences from 20 – 50 nucleotides, and the selection cycle is repeated until aptamers with the desired affinity are identified.<sup>17,18,19</sup> Although SELEX has identified novel aptamers with high binding affinities to their cognate binding partners, the process typically requires 8 – 15 rounds of selection followed by PCR-based amplification, which can introduce undesired biases in aptamer discovery, such as the loss of rare sequences, PCR biases, and inadvertent amplification of low-affinity or non-specific sequences.<sup>17,18,19</sup> Further, the efficiency of SELEX-style methods is constrained by the limited enrichment achieved in one round.<sup>17,18,19</sup>

### **1.3.2. Particle display**

Particle display is a screening-based aptamer discovery technique that quantitatively measures the relative affinities of a library of DNA-based aptamer particles and isolates those demonstrating the highest affinities in a high-throughput manner.<sup>20,21,22</sup> This work

was further advanced to convert the DNA libraries to monoclonal “gene-linked aptamer particles” (GRAPs), which display multiple copies of a single RNA aptamer on monodisperse polymer beads.<sup>23</sup> Particle display overcomes some of the challenges faced by conventional SELEX-style methods for selection, as it can screen more than 100 million aptamers at a throughput rate of 5000 particles per second.<sup>23</sup> In this process, a randomized DNA library is subjected to emulsion PCR to create monoclonal aptamer particles that display a single DNA sequence.<sup>23</sup> Next, the immobilized DNA sequences are converted to RNA by emulsion transcription to create gene-linked aptamer particles.<sup>23</sup> The monoclonal GRAPs are then incubated with a target dye, such that the particles emit a fluorescent signal upon binding with the dye.<sup>23</sup> The fluorescent particles can be sorted from those that do not produce fluorescence and will undergo PCR amplification for (1) further rounds of screening with increased selection stringency or (2) for high-throughput sequencing to identify the corresponding RNA aptamer.<sup>23</sup> Not only can this *in vitro* technique for RNA aptamer discovery efficiently screen large libraries of GRAPs in fewer rounds than conventional aptamer screening methods, but it also enables users to isolate RNA aptamers that meet a target affinity and specificity based directly on their fluorescence-enhancing properties.<sup>23</sup>

#### 1.4. References

- [1] Hentze, M.; Castello, A.; Schwarzl, T., A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **2018**, *19*, 327–341.
- [2] Ramanathan, M.; Porter, D. F.; Khavari, P. A., Methods to study RNA–protein interactions. *Nature Methods* **2019**, *16* (3), 225-234.
- [3] Licatalosi, D. D.; Ye, X.; Jankowsky, E., Approaches for measuring the dynamics of RNA–protein interactions. *WIREs RNA* **2020**, *11* (1), e1565.
- [4] Pollard, T. D., A guide to simple and informative binding assays. *Molecular Biology of the Cell* **2010**, *21* (23), 4061-4067.

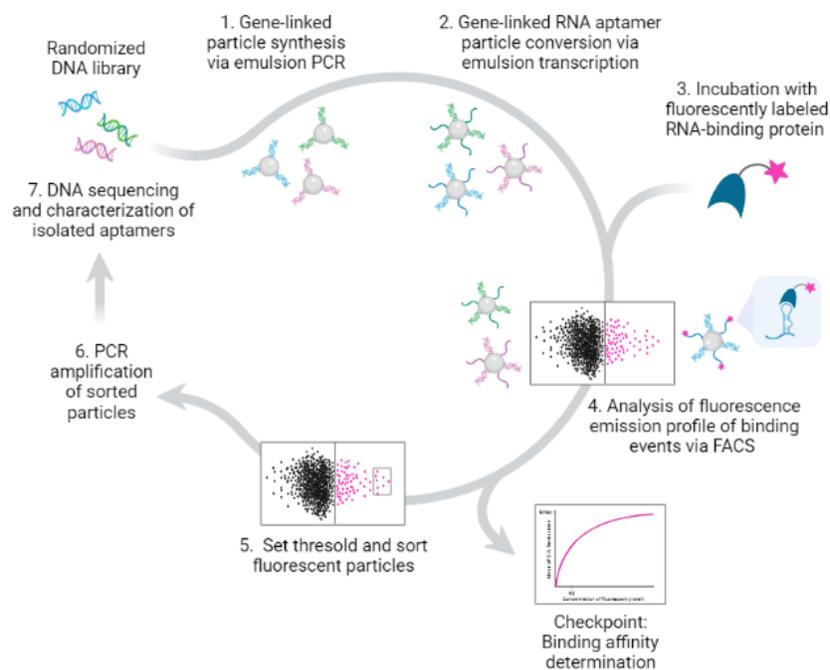
- [5] Motulsky, Harvey., The GraphPad guide to analyzing radioligand binding data. <http://www3.uah.es/farmamol/Public/GraphPad/radiolig.htm> (accessed September, 2021).
- [6] Rio, D. C., Electrophoretic mobility shift assays for RNA–protein complexes. *Cold Spring Harbor Protocols* **2014**, 2014 (4), pdb.prot080721.
- [7] Hellman, L.; Fried, M., Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat Protoc* **2007**, 2, 1849–1861.
- [8] Rio, D. C., Filter-binding assay for analysis of RNA–protein interactions. *Cold Spring Harbor Protocols* **2012**, 2012 (10), pdb.prot071449.
- [9] Hall, K. B.; Kranz, J. K., Nitrocellulose filter binding for determination of dissociation constants. *Methods in molecular biology (Clifton, N.J.)* **1999**, 118, 105–114.
- [10] Moerke, N. J., Fluorescence polarization (FP) assays for monitoring peptide-protein or nucleic acid-protein binding. *Current Protocols in Chemical Biology* **2009**, 1 (1), 1-15.
- [11] Rossi, A.; Taylor, C., Analysis of protein-ligand interactions by fluorescence polarization. *Nat Protoc* **2011** 6, 365–387.
- [12] Sultana, A.; Lee, J. E., Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Current Protocols in Protein Science* **2015**, 79 (1), 19.25.1-19.25.26.
- [13] Martin, S. R.; Ramos, A.; Masino, L., Biolayer interferometry: protein–RNA interactions. In *Protein-Ligand Interactions: Methods and Applications*, Daviter, T.; Johnson, C. M.; McLaughlin, S. H.; Williams, M. A., Eds. Springer US: New York, NY, 2021; pp 351-368.
- [14] Wang, S.; Poon, G. M. K.; Wilson, W. D., Quantitative investigation of protein–nucleic acid interactions by biosensor surface plasmon resonance. In *DNA-Protein Interactions: Principles and Protocols*, Leblanc, B. P.; Rodrigue, S., Eds. Springer New York: New York, NY, 2015; pp 313-332.
- [15] Nicoya. Resources. <https://nicoyalife.com/nicoya-surface-plasmon-resonance-resources> (accessed September, 2021).
- [16] Dunn, M. R.; Jimenez, R. M.; Chaput, J. C., Analysis of aptamer discovery and technology. *Nature Reviews Chemistry* **2017**, 1 (10), 0076.
- [17] Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, 249 (4968), 505-510.
- [18] Ruscito, A.; DeRosa, M. C., Small-molecule binding aptamers: selection strategies, characterization, and applications. *Frontiers in Chemistry* **2016**, 4 (14).

- [19] Zhuo, Z.; Yu, Y.; Wang, M.; Li, J.; Zhang, Z.; Liu, J.; Wu, X.; Lu, A.; Zhang, G.; Zhang, B., Recent advances in SELEX technology and aptamer applications in biomedicine. *International Journal of Molecular Sciences* **2017**, *18* (10).
- [20] Wang, J.; Gong, Q.; Maheshwari, N.; Eisenstein, M.; Arcila, M. L.; Kosik, K. S.; Soh, H. T., Particle display: a quantitative screening method for generating high-affinity aptamers. *Angewandte Chemie International Edition* **2014**, *53* (19), 4796-4801.
- [21] Wang, J.; Yu, J.; Yang, Q.; McDermott, J.; Scott, A.; Vukovich, M.; Lagrois, R.; Gong, Q.; Greenleaf, W. Eisenstein, M.; Ferguson, B. S.; Soh, H. T., Multiparameter particle display (MPPD): a quantitative screening method for the discovery of highly specific aptamers. *Angewandte Chemie International Edition* **2017**, *56* (3), 744-747.
- [22] Gordon, C. K. L.; Eisenstein, M.; Soh, H. T., Direct selection strategy for isolating aptamers with pH-sensitive binding activity. *ACS Sensors* **2018**, *3* (12), 2574-2580.
- [23] Gotrik, M.; Sekhon, G.; Saurabh, S.; Nakamoto, M.; Eisenstein, M.; Soh, H. T., Direct selection of fluorescence-enhancing RNA aptamers. *Journal of the American Chemical Society* **2018**, *140* (10), 3583-3591.

## Chapter 2. High-throughput particle display screening of RNA-protein interactions

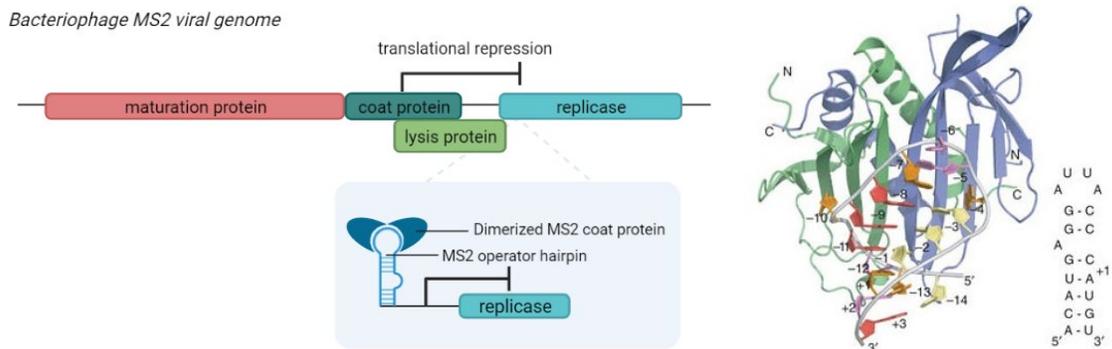
### 2.1. Introduction

To overcome the limitations of the conventional binding assays to study RNA-protein binding interactions, we propose to develop an *in vitro* particle display screening method to profile the binding affinities of RNA-protein binding interactions in a high-throughput manner. Using flow cytometry, we will be able to interrogate large numbers of binding events based directly on fluorescence. We envision using this particle display screening method to express an RNA library on monoclonal GRAPs to isolate RNA sequences whose binding-induced fluorescence enhancement surpass a user-defined gate. These isolated GRAPs will be subjected to subsequent identification and characterization to analyze if sequence diversity and convergence influence protein binding. This process is depicted in **Figure 3**.



**Figure 3.** Gene-linked aptamer particle (GRAP) display method for RNA aptamer library screening of RNA-protein binding interactions.

One well-characterized RNA-protein interaction is the MS2-MCP system (**Figure 4**).<sup>24,25,26,27</sup> Derived from the MS2 bacteriophage, this ribonucleoprotein complex is comprised of the bacteriophage MS2 RNA, which binds to the MS2 coat protein (MCP).<sup>24,25,26</sup> To infect bacterial cells, the maturation protein of the MS2 bacteriophage is attached to the cell surface and delivers the viral genome inside the host cell.<sup>24,25,26</sup> Once the MCP is translated and enough protein is produced, the MCP dimerizes and binds to the MS2 operator hairpin that is 19-nucleotide (nt) in length and is located upstream the replicase cistron.<sup>24,25,26</sup> The binding phenomenon of the MCP to the MS2 operator hairpin prevents ribosomes from binding to the initiation site of the viral genome, thereby repressing the translation of the replicase cistron and activating virion assembly, maturation, and release.<sup>24,25,26</sup> Because of its strong binding affinity (<10 nM), the MS2-MCP system has been widely studied for cellular imaging and kinetic studies.<sup>28,29,30</sup> Thus, we have focused on the MS2-MCP system for our preliminary experiments to confirm the binding affinity of this system using our particle display-based technique. For the proof-of-principle experiments that subject this technique to library and mutational studies, we will also continue to use the MS2-MCP system.



**Figure 4.** Bacteriophage MS2 viral genome diagram (left) and crystal structure of MS2-MCP system with MS2 hairpin structure (right).<sup>27</sup>

## 2.2. Results/Discussion

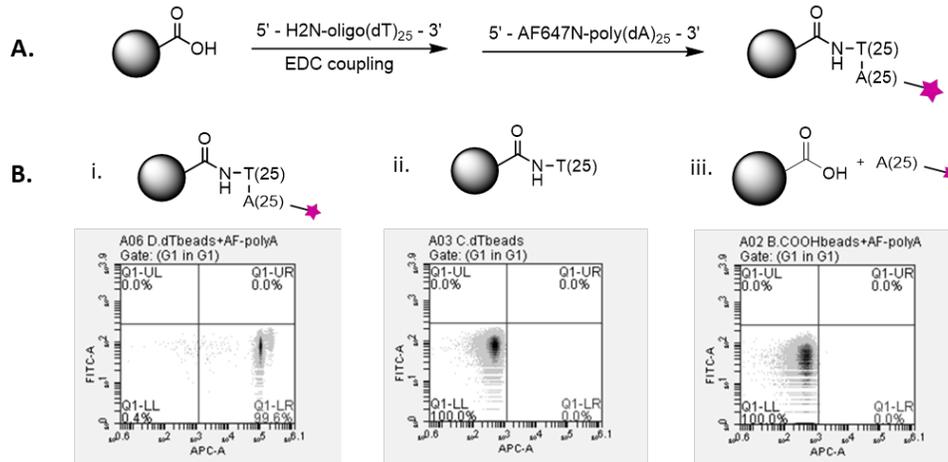
*Aim 1. To generate aptamer particles expressing RNA aptamer of interest.*

*Aim 1A. To perform carbodiimide chemistry to couple amino-modified oligo(dT)25 to carboxylic acid beads.*

The first step to generate aptamer particles expressing the MS2 aptamer is to perform carbodiimide chemistry to couple amino-modified oligo(deoxythymidine)25 (oligo(dT)25) to 1  $\mu\text{m}$  Dynabeads<sup>®</sup> MyOne<sup>™</sup> carboxylic acid-functionalized beads.<sup>31,32,33,34</sup> These mono-sized beads are composed of highly crosslinked polystyrene beads encapsulating superparamagnetic iron oxide and coated with a hydrophilic glycidyl ether layer with branching carboxylic acid groups.<sup>33</sup> To activate the beads for ligand coating, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) reacts with the surface carboxylic acid groups allowing covalent amide bond formation to the primary amino-modified oligonucleotide.<sup>31,32,33,34</sup> Specifically, the EDC reacts with the surface carboxylic acid groups to form an activated *O*-acylisourea intermediate that undergoes a nucleophilic attack with the 5'-amino-modified oligo(dT)25.<sup>31,32,33,34</sup> EDC crosslinking reactions must be performed quickly in acidic conditions devoid of carboxyls and amines as the intermediate *O*-acylisourea is labile and hydrolyzes in aqueous conditions.<sup>31,32,33,34</sup> The primary amine of the 5'-amino-modified oligo(dT)25 forms an amide bond to displace the original carboxyl group on the bead surface, resulting in the bead expressing 25 deoxythymidines on its surface.

*Aim 1B. To test the particle coating of oligo(dT)25-modified beads with AF647-labeled oligo(dA)25 probe.*

To validate the EDC crosslinking chemistry of the amino-modified oligo(dT)25 to the carboxylic acid groups on the bead surface, the coating of the oligo(dT)25-modified beads are tested with an Alexa Fluor 647-labeled complementary oligo(deoxyadenosine)25 (oligo(dA)25) probe. The Alexa Fluor 647 (AF647) dye is a far-red fluorescent dye that embodies high fluorescence quantum yield and high photostability for the detection of biological structures with high sensitivity.<sup>35</sup> If the modified beads properly expressed the oligo(dT)25 on each bead surface, then the oligo(dT)25 will hybridize to the AF647-labeled oligo(dA)25 probe via A-T hybridization and express fluorescence. For this experiment, the oligo(dT)25-modified beads were incubated with 1  $\mu$ M AF647-labeled oligo(dA)25, washed once to remove any unbound fluorescent probes, and suspended in buffer to be analyzed by flow cytometry (**Figure 5A**). **Figure 5B** shows the fluorescence emission profile of this control experiment, indicating that the oligo(dT)25 is correctly conjugated to the beads. Note that the vertical and horizontal gates on each plot were positioned based on the control oligo(dT)25-modified beads (**Figure 5B. ii.**), which does not express any fluorescence since these beads are not incubated with the fluorescent complementary probe. **Figure 5B. i.** indicates that 99.6% of the beads expressed fluorescence on the APC-A channel, which shows that the oligo(dT)25 on the bead surface hybridized to the AF647-labeled oligo(dA)25 probe and indicates an efficient functionalization of the beads. To cross-check the conjugation with a negative control, **Figure 5B. iii.** analyzes the fluorescent response when the unmodified carboxylic acid beads are incubated with the fluorescent oligo(dA)25 probe. The beads do not express fluorescence, which is expected since the oligo(dT)25 is not present to facilitate A-T hybridization.



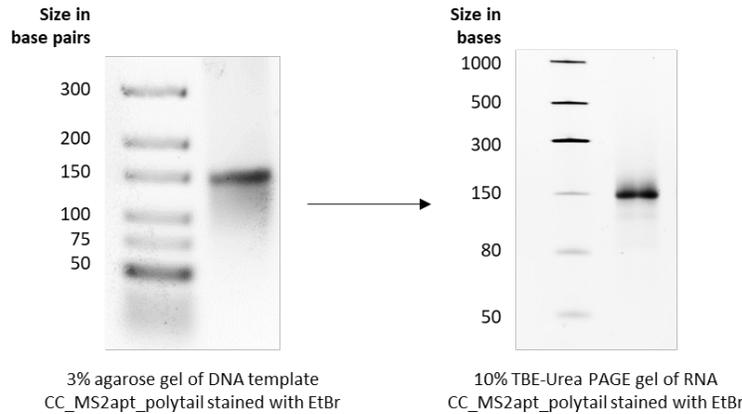
**Figure 5.** Conversion of carboxylic acid beads to oligo(dT)<sub>25</sub>-modified beads annealed to an AF647-labeled complementary oligo(dA)<sub>25</sub> probe. **A.** Schematic overview of conversion of carboxylic acid beads to oligo(dT)<sub>25</sub>-modified beads. **B.** Fluorescence emission profile of i. oligo(dT)<sub>25</sub>-modified beads hybridized to AF647-labeled oligo(dA)<sub>25</sub> probe, ii. oligo(dT)<sub>25</sub> beads, iii. carboxylic acid beads incubated with AF647-labeled oligo(dA)<sub>25</sub> probe.

*Aim 1C. To synthesize MS2 aptamer with a polyA tail by in vitro transcription.*

To express the MS2 aptamer on the oligo(dT)<sub>25</sub>-modified beads, the MS2 aptamer was extended with a 3'-polyA tail such that the transcript will hybridize to the oligo(dT)<sub>25</sub>-modified beads via the polyA(25) tail. The double-stranded DNA template, CC\_MS2apt\_polytail (149 bp), consisting of the T7 promoter region, MS2 aptamer region, and the polyA(25) tail, was generated by primer extension. To do so, two oligonucleotides were designed to span the entire desired DNA template with a 20-base pair (bp) overlap and were then subjected to a PCR reaction to anneal and extend the two oligonucleotides together.<sup>36</sup> The gel image on the left of **Figure 6** shows the DNA template is ~150 bp which is what was expected. The DNA template was purified and isolated from the reaction mixture using an agarose gel extraction protocol for direct use in *in vitro* transcription applications, as the quality of the DNA template greatly affects *in vitro* transcription efficiency and subsequent RNA integrity.

Once the DNA template was synthesized and purified, 0.3-0.5  $\mu\text{g}$  DNA template was subjected to a standard *in vitro* transcription reaction protocol to produce the single-stranded run-off transcript after the T7 promoter sequence (125 bases).<sup>37,38</sup> The RNA was purified and isolated from the reaction mixture using (1) DNase I treatment, (2) acidic phenol-chloroform extraction, and (3) ethanol precipitation, following standardized protocols.<sup>39,40</sup> To remove the DNA template from the *in vitro* transcription reaction, the endonuclease DNase I is added to cleave double-stranded DNA non-specifically to produce di-, tri-, and oligonucleotide products with 5'-phosphorylated and 3'-hydroxylated ends.<sup>39</sup> A phenol-chloroform extraction is then performed to remove the enzymes since most proteins are soluble in phenol and become denatured in chloroform, allowing the enzymes to migrate to the lower organic phase for removal.<sup>40</sup> The leftover DNA template can also be separated from the RNA solution if the phenol-chloroform extraction is performed under acidic conditions (pH 4-6).<sup>40</sup> Since the phase partitioning of nucleic acids is dependent on the pH of the solution, DNA becomes neutralized and moves to the organic phase and interface at pH 4-6 while the RNA will remain in the aqueous phase. Lastly, the RNA solution is purified by ethanol precipitation to remove impurities and concentrate the RNA.<sup>40</sup> The addition of sodium acetate and ethanol to the RNA solution makes the RNA more hydrophobic and allows for precipitation from any salts or water-soluble proteins. The gel image on the right of **Figure 6** is the purified RNA product slightly below the 150 base marker. Although the RNA transcript is expected to be 125 bases since the start of transcription is immediately downstream of the 24-nucleotide T7 promoter sequence, the polyA tail may have been extended past 25-nucleotides and produced a longer transcript than expected. As long as (1) the MS2

aptamer is fully transcribed and can properly fold into its native hairpin structure and (2) the polyA tail can efficiently hybridize to the oligo(dT)25 on the functionalized beads, which is assessed in Aim 1D, a slightly longer transcript is not a concern for the experiment.



**Figure 6.** Primer extension (left) and *in vitro* transcription (right) product analysis of CC\_MS2apt\_polytail by gel electrophoresis.

*Aim 1D.* To anneal the MS2 aptamer with a polyA tail to the oligo(dT)25-modified beads and analyze the RNA capture efficiency with an AF647-labeled oligo that has a complementary sequence to the RNA.

To anneal the RNA with the MS2 aptamer and polyA(25) tail to the oligo(dT)25-modified beads, oligo(dT)25-modified beads were incubated in 1  $\mu$ M RNA. This reaction mixture was subjected to the annealing reaction, and the beads were washed once with buffer. The RNA capture efficiency of the oligo(dT)25-modified beads was tested with an AF647-labeled probe complementary to the MS2 aptamer sequence. If the oligo(dT)25-modified beads hybridized to the RNA expressing the MS2 aptamer via A-T hybridization with the 3'-polyA tail, the beads would express fluorescence after incubation with the AF647-labeled MS2 aptamer complement (**Figure 7A**). **Figure 7B** shows the fluorescence emission profile of this control experiment, indicating that the

RNA expressing the MS2 aptamer properly annealed to the oligo(dT)25-modified beads.

Note that the vertical and horizontal gates on each plot were positioned based on the control oligo(dT)25-modified beads (**Figure 7B. ii.**), which do not express any

fluorescence since these beads are not incubated with the fluorescent complementary

probe. **Figure 7B. i.** indicates that 99.2% of the beads expressed fluorescence on the

APC-A channel demonstrating that the RNA had annealed to the oligo(dT)25 branching

from the bead surface via its 3'-polyA tail on the large majority of the beads.

To cross-check the conjugation with a negative control, **Figure 7B. iii.** analyzes the

fluorescent response when oligo(dT)25-modified beads are incubated with the fluorescent

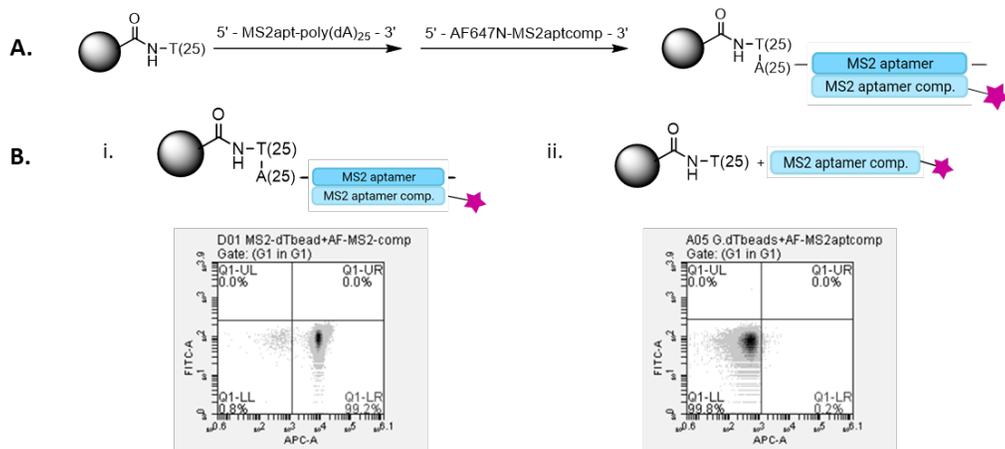
probe with a complementary sequence to the MS2 aptamer. The majority of the beads do

not express fluorescence, which is expected since the RNA harboring the MS2 aptamer is

not present to facilitate RNA-RNA hybridization. The 0.2% fluorescent events in the

lower right quadrant are most likely due to an insufficient wash step to remove the

residual fluorescent probe from the buffer.



**Figure 7.** Conversion of oligo(dT)25-modified beads to oligo(dT)25-modified beads expressing the MS2 aptamer annealed to an AF647-labeled oligo that is complementary to the MS2 aptamer region. **A.** Schematic overview of conversion of oligo(dT)25-modified beads to oligo(dT)25-modified beads expressing MS2 aptamer. **B.** Fluorescence emission profile of **i.** oligo(dT)25-modified beads expressing the MS2 aptamer hybridized to AF647-labeled MS2 aptamer complementary probe, **ii.** oligo(dT)25 beads expressing the MS2 aptamer incubated with AF647-labeled MS2 aptamer complementary probe.

*Aim 2. To purify and label cognate RNA-binding protein with fluorescent dye.*

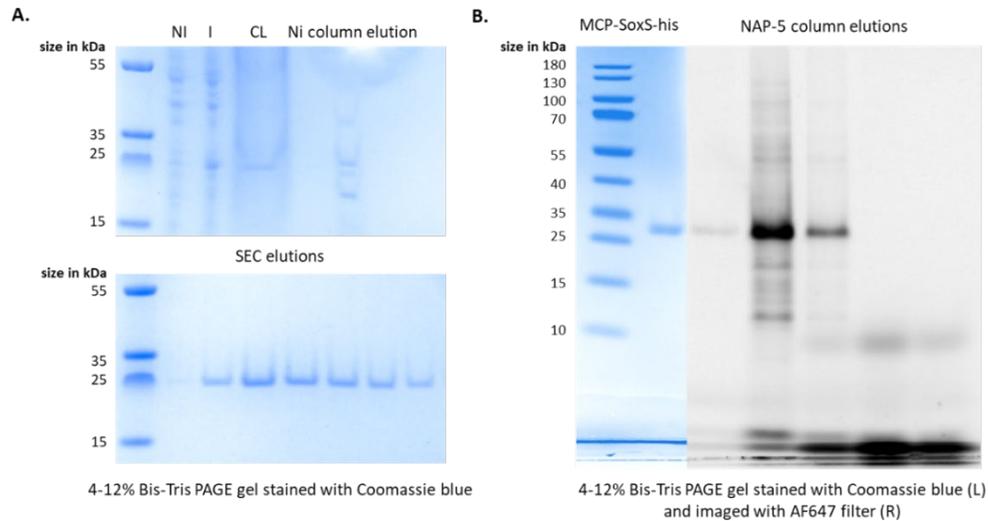
*Aim 2A. To label and purify MCP-SoxS-his protein with an AF647 dye to bind to the MS2 aptamer adhered to oligo(dT)25-modified beads.*

The first step to label and purify MCP-SoxS-his protein with an Alexa Fluor 647 dye is to express and purify MCP-SoxS-his protein (MW = 26.7 kDa). Derived from a previous construct, the protein of interest contains the MCP fused to a SoxS transcriptional activator and a polyhistidine tag (6xHis-tag). The fusion protein is then cloned into the pTYB21 expression plasmid designed for recombinant protein expression and purification.<sup>41</sup> The vector contains the medium copy number ColE1 replication origin and an inducible *lacI* gene downstream the T7 promoter sequence, such that the expression of the MCP-SoxS-his fusion protein can be induced with isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) in *Escherichia coli*.<sup>41</sup> The expression plasmid with the MCP-SoxS-his protein insert is shown in **Appendix 1**. After protein expression was induced, the bacteria cells were lysed via sonication, and the cell lysis was purified with nickel-nitrilotriacetic acid (Ni-NTA) column chromatography since the fusion protein expresses C-terminal 6xHis-tag.<sup>42</sup> **Figure 8A** shows the non-induced fraction, induced fraction, cell lysate fraction, and the Ni-NTA column elution from the leftmost column to the rightmost column, respectively. The Ni-NTA column elution sample contains unwanted impurities at higher and lower molecular weights than the expected protein size; therefore, the fractions eluted from the Ni-NTA column were concentrated, dialyzed, and further purified with the HiPrep 16/40 Sephacryl S-100 high-resolution size exclusion chromatography (SEC) column.<sup>43</sup> This SEC column consists of a cross-linked allyl dextran and N, N-methylenebisacrylamide matrix with a separation range of 1  $\times$

$10^3 - 1 \times 10^5$ .<sup>43</sup> The fractions eluted from the SEC column were monitored for protein by UV spectroscopy at Abs<sub>280</sub> and further analyzed with gel electrophoresis using a 4-12% Bis-Tris polyacrylamide gel electrophoresis (PAGE) gel, as shown in the bottom gel in **Figure 8A**. From the gel analysis, the MCP-SoxS-his protein has a size of ~25 kDa, which is close to the expected size. To further confirm the purity of the MCP-SoxS-his protein, LC/MS characterization was used to determine its size (MW = 26.7 kDa), which is consistent with what is expected (MW = 26.7 kDa) (**Appendix 1**).

To label the MCP-SoxS-his protein, the Alexa Fluor 647 protein labeling kit from Molecular Probes was purchased as it can quickly and efficiently label primary amines of proteins using an Alexa Fluor 647 dye with a succinimidyl ester reactive group.<sup>44</sup> Thus, the *N*-hydroxysuccinimide of the Alexa Fluor 647 succinimidyl ester acts as a leaving group to encourage the formation of an amide bond, linking the fluorescent dye to the primary amine group on the amine group on the MS2 coat protein.<sup>44,45</sup> The labeling and purification protocol was followed according to the manufacturer's directions; however, the included Bio-Rad BioGel P-30 fine size exclusion purification resin did not efficiently remove the free dye. Thus, the labeled protein was purified by the Cytiva illustra NAP-5 column Sephadex G-25 which significantly reduced the free dye present in the purified AF647-labeled MCP-SoxS-his protein samples, as indicated by gel electrophoresis. The BioGel P-30 resin has a fractionation range of 25000 – 40000, whereas the Sephadex G-25 resin has a fractionation range of 1000 – 5000.<sup>46,47</sup> When using the Sephadex G-25, the free dye will remain in the column, and the labeled protein will elute with the running buffer. **Figure 8B** shows the purified MCP-SoxS-his protein on the left and the fractions after the labeling reaction and NAP-5 column purification.

The first three fractions eluted from the NAP-5 column show the strongest fluorescent product between 25 and 35 kDa, whereas the last two fractions eluted were mainly comprised of free dye at < 10 kDa. The second fraction eluted from the NAP-5 column had the strongest presence of the expected labeled protein and the least free dye, so this fraction was used for the following experiments. The protein concentration was determined using a standardized Bradford assay, and the labeling efficiency was analyzed with the SDS-PAGE gel. The SDS-PAGE gel analysis of the second fraction shows higher and lower molecular weight impurities, likely due to the inconsistent labeling of the protein. Since the MCP-SoxS-his protein has ten lysines in its amino acid sequence, each protein molecule has ten different labeling positions. These positions are influenced by many factors, such as accessibility and stereochemical interactions, possibly resulting in the protein molecules varying in their degree of labeling. Although excessive under-labeling and over-labeling of the protein with the fluorescent dye would inhibit the binding abilities of the labeled protein to the MS2 aptamer, a slight variation of the number of dye molecules per protein molecule or position of dye molecules on each protein molecule should not affect the downstream applications. This is because the high-throughput particle display-based method designed for to determine the binding affinity of RNA-protein complexes profiles an entire population of RNA-protein complexes by analyzing and compiling hundreds of thousands of individual RNA-protein complexes.



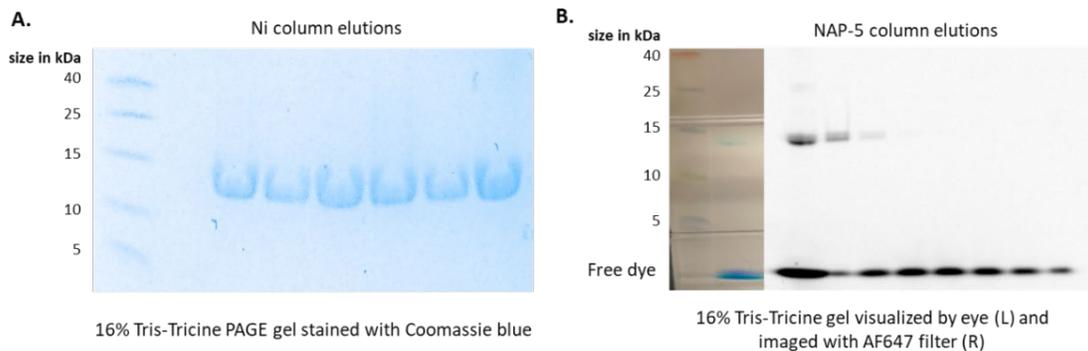
**Figure 8.** SDS-PAGE analysis of MCP-SoxS-his protein and AF647-labeled MCP-SoxS-his protein. SDS-PAGE analysis of **A.** non-induced (NI), induced (I), cell lysate (CL), Ni column elution, (top gel image) and SEC column elutions (bottom gel image) of MCP-SoxS-his protein, and **B.** NAP-5 column elutions of AF647-labeled MCP-SoxS-his protein.

*Aim 2B.* To label and purify SoxS-his protein with an AF647 dye as the negative control.

The next step is to label and purify the negative control protein that does not contain the MS2 coat protein, SoxS-his protein (MW = 13.6 kDa), with an Alexa Fluor 647 dye. The expression plasmid with the SoxS-his protein insert is shown in **Appendix 1**. Following the protocol to express and purify the MCP-SoxS-his protein, the SoxS-his protein expression was induced, the bacteria cells were lysed via sonication, and the cell lysis was purified with Ni-NTA column chromatography since the fusion protein expresses C-terminal 6xHis-tag. The fractions eluted from the Ni-NTA column were monitored for protein by UV spectroscopy at Abs<sub>280</sub> and further analyzed with gel electrophoresis using a 16% Tricine PAGE gel, as shown in **Figure 9A**. From the gel analysis, the SoxS-his protein has a size between 10 – 15 kDa, consistent with the expected size. The purified SoxS-his protein was also characterized by LCMS to determine its size (MW = 13.5 kDa) (**Appendix 1**). The slight discrepancy between the expected and resulting size of the protein is likely due to a loss of a histidine residue from the 6xHis-tag.

The same reaction conditions to label the positive control protein with the Alexa Fluor 647 dye were repeated for the negative control protein. The labeling and purification protocol was followed according to the manufacturer's directions; however, the labeling reaction was directly purified with the Cytiva illustra NAP-5 column Sephadex G-25, as described above. This purification significantly reduced the free dye present in the purified AF647-labeled SoxS-his protein samples, as indicated by gel electrophoresis.

**Figure 9B** shows the purified SoxS-his protein on the left and the fractions after the labeling reaction and NAP-5 column purification. The second fraction eluted from the NAP-5 column shows the strongest fluorescent product between 10 and 15 kDa, whereas the last six fractions eluted were mainly comprised of free dye at < 10 kDa. The second fraction eluted from the NAP-5 column had the strongest presence of the expected labeled protein and the least free dye, so this fraction was used for the following experiments. Since the SoxS-his protein has four lysines in its amino acid sequence, there is still a possibility that each protein molecule varies in its degree of labeling, although the sample analyzed by the SDS-PAGE gel appears to be quite pure.

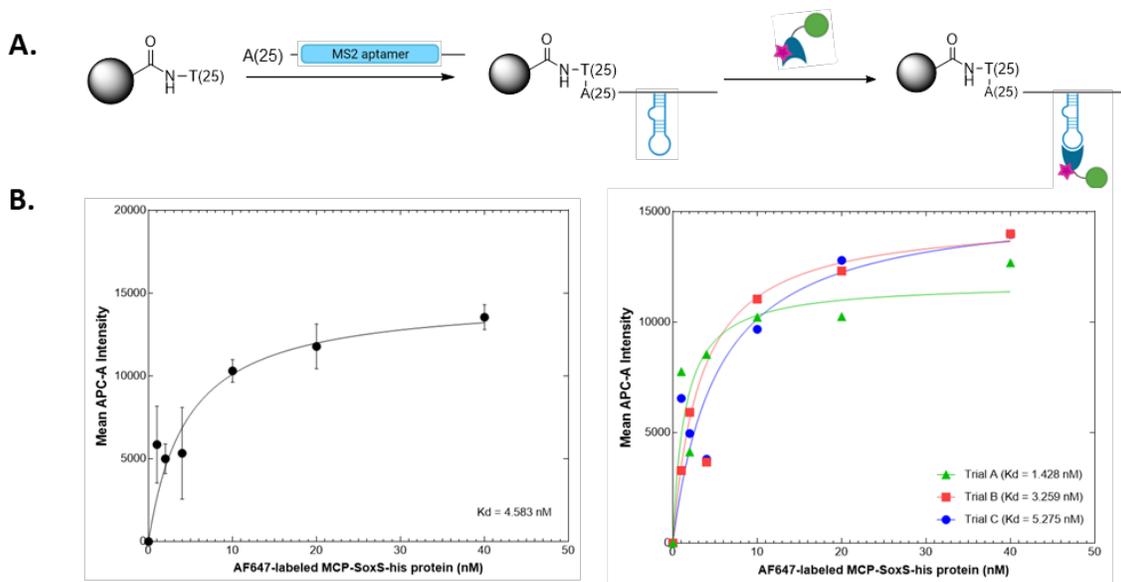


**Figure 9.** SDS-PAGE analysis of SoxS-his protein and AF647-labeled SoxS-his protein. SDS-PAGE analysis of **A.** Ni column elutions of SoxS-his protein and **B.** NAP-5 column elutions of AF647-labeled SoxS-his protein.

*Aim 3. To perform binding reactions and measure the fluorescence emission profile to quantify the binding affinity.*

*Aim 3A. MS2-MCP binding affinity determination*

To determine the MS2-MCP binding affinity, several binding reactions of the MS2 aptamer to the AF647-labeled MCP-SoxS-his protein were performed, and the resulting fluorescence emission spectrum was measured. For the general workflow of the binding experiment that is depicted in **Figure 10A**, the RNA expressing the MS2 aptamer was annealed to the oligo(dT)25-modified beads which were then incubated with a range of AF647-labeled MCP-SoxS-his protein (0, 1, 2, 4, 10, 20, 40 nM) for one hour at room temperature with rotation. The beads were washed once and suspended to be analyzed by flow cytometry. For each sample, the mean APC-A intensity of the lower-right quadrant for a consistent number of binding events for  $n = 3$  repetitive experiments were measured and plotted against the AF647-labeled MCP-SoxS-his protein concentration. The data were fit using the nonlinear regression program for receptor binding experiments (saturation binding: one site: specific binding) in GraphPad to calculate the  $K_d$  of this system (**Figure 10B**). The binding curve is shown in **Figure 10B** on the left that was fit according to the GraphPad program computed the  $K_d$  value of this MS2-MCP system to be 4.853 nM, which is consistent with the literature value. The binding curves of the individual trials are shown in **Figure 10B** on the right in which the green triangle points represent the first trial, the red square points represent the second trial, and the blue circle points represent the third trial, showing an increase of  $K_d$  values from 1.428 nM to 3.259 nM to 5.275 nM, respectively.

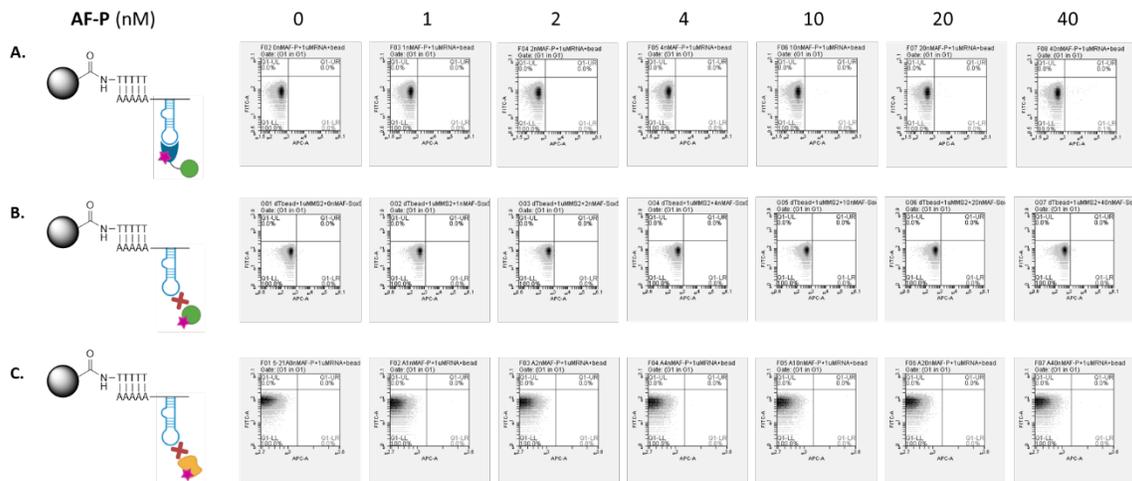


**Figure 10.** Binding experiment of MS2-MCP system. **A.** General workflow of particle display-based technique to determine the binding affinity of the MS2-MCP system. **B.** Binding curve of MS2-MCP system for the average (left) and individual (right) rounds. Values reported are the mean  $\pm$  SD of  $n = 3$  repetitive experiments which were fit using GraphPad (nonlinear regression: saturation binding; one site: specific binding).

*Aim 3B.* To assess AF647-labeled SoxS-his protein and AF647-labeled BSA protein as negative controls against the MS2-MCP system.

To confirm the specificity of the MS2-MCP system, the two negative control proteins, AF647-labeled SoxS-his protein and AF647-labeled BSA protein<sup>49</sup>, were subjected to the binding reactions and compared to that of the AF647-labeled MCP-SoxS-his protein. Following the same protocol as Aim 3A, the oligo(dT)25-modified beads expressing the MS2 aptamer were incubated with a range of AF647-labeled protein (0, 1, 2, 4, 10, 20, 40 nM) for one hour at room temperature with rotation. After the beads were washed and suspended in buffer, the mean APC-A intensity was measured for 100,000 events in G1 for  $n = 3$  repetitive experiments and demonstrated 0.0% fluorescent events in the lower right quadrant for all systems with an AF647-labeled protein (AF-P) concentration from 0 – 20 nM (**Figure 11A-C**). At AF-P concentration of 40 nM, the AF647-labeled MCP-

SoxS-his protein demonstrates 0.1% fluorescent events in Q1-LR, whereas the AF647-labeled SoxS-his and AF647-BSA protein demonstrates 0.0% fluorescent events (**Figure 11A-C**). Although there is an increase in fluorescence events for the positive control, there is not a significant visual difference until 40 nM. Since each sample undergoes a quality control check with the AF647-labeled MS2 complementary probe (see Aim 1D.), it is assumed that the majority of the beads are expressing the MS2 aptamer. This means that the likely reason that the large majority of the beads do not express fluorescence is due to the binding of the AF647-labeled protein to the MS2 aptamer.



**Figure 11.** Negative control experiment for MS2-MCP system. Fluorescence emission profile of 100,000 events in G1 of oligo(dT)25-modified beads expressing the MS2 aptamer incubated with **A.** AF647-labeled MCP-SoxS-his protein, **B.** AF647-labeled SoxS-his protein, and **C.** AF647-labeled BSA protein at a range of labeled protein concentrations (0, 1, 2, 4, 10, 20, 40 nM).

To see if there would be an increase in fluorescence with an increase in fluorescent protein concentration, the oligo(dT)25-modified beads expressing the MS2 aptamer were incubated with a higher range of AF647-labeled protein (0, 100, 200, 400 nM) for one hour at room temperature with rotation. After the beads were washed and suspended in buffer, the mean APC-A intensity was measured for 100,000 events in G1 for  $n = 3$  repetitive experiments (**Figure 12A-C**). The increase of fluorescent events for the MS2-

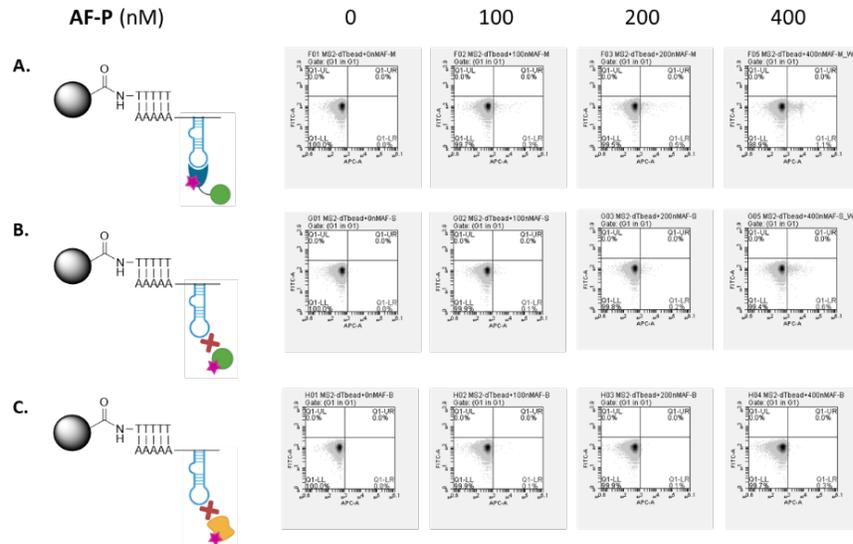
AF647-MCP-SoxS-his system compared to the MS2-AF647-SoxS-his and MS2-AF647-BSA systems validates that the protein expressing the MS2 coat protein specifically and selectively binds to the MS2 aptamer.

*Aim 3C. To further validate estimated  $K_d$  value of MS2-MCP system.*

One experiment to further cross-check and validate this binding assay would be to run an experiment parallel to Aim 3A with the two negative controls. Following the protocol described in Aim 3A, the RNA expressing the MS2 aptamer would be annealed to the oligo(dT)25-modified beads, and the beads would then be incubated with a range of AF647-labeled SoxS-his and AF647-labeled BSA protein (0, 1, 2, 4, 10, 20, 40 nM) for one hour at room temperature with rotation. The beads would then be washed once and suspended to be analyzed by flow cytometry. For each sample, the mean APC-A intensity of the lower-right quadrant for the same number of binding events measured for the positive control for  $n = 3$  repetitive experiments would be measured, if applicable.

Ideally, there should not be any beads that express fluorescence since the MS2 aptamer does not bind to the AF647-labeled SoxS-his and AF647-labeled BSA proteins; however, this is difficult to achieve due to non-specific binding and insufficient washing steps. I hypothesize that there may be fluorescent events in each sample, but it would require a larger sample volume, supporting that the fluorescent events for the MS2-AF647-MCP-SoxS-his system are from the binding interaction of the RNA-protein complex.

It is also essential to perform a second binding assay, like FP, SPR, or BLI, to further confirm the estimated  $K_d$  value of this particle display-based binding assay.



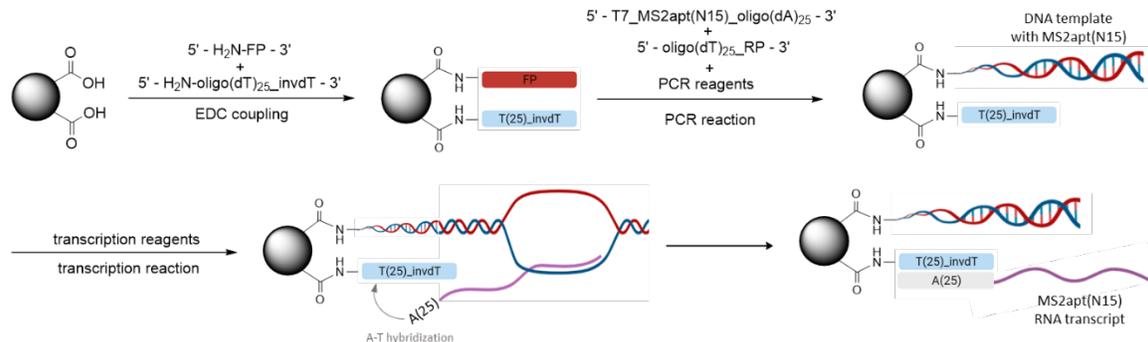
**Figure 12.** Negative control experiment for MS2-MCP system at an increased labeled protein concentration range. Fluorescence emission profile of 100,000 events in G1 of oligo(dT)25-modified beads expressing the MS2 aptamer incubated with **A.** AF647-labeled MCP-SoxS-his protein, **B.** AF647-labeled SoxS-his protein, and **C.** AF647-labeled BSA protein at a range of labeled protein concentrations (0, 100, 200, 400 nM).

### 2.3. Future directions

Since we have confirmed that this particle display screening method can estimate the binding affinity of a well-characterized RNA-protein interaction, we plan to use this technique to screen a library of RNA variants for mutational analysis. Given the high-throughput abilities of this method to measure the binding-dependent fluorescence emission profile of whole populations, we can interrogate a library of RNA variants to isolate, sequence, and characterize specific RNA sequences that result in strong (or weak) binding interactions with its cognate RNA-binding protein. As novel mutations in the MS2 aptamer region have been discovered to influence the binding interaction with the MS2 coat protein, we hope to use our high-throughput particle display screening method to rapidly screen a library of MS2 variants to expand the mutational arsenal.<sup>50,51,52</sup> As for our proposed proof-of-principle method, monoclonal RNA aptamer particles expressing an RNA library harboring sequence variations in the MS2 aptamer

region ( $N_{15} = 4 \times 10^9$  library variants) that binds to MCP will be produced, as detailed in **Figure 13**. To synthesize GRAPs expressing an MS2 aptamer library, carbodiimide conjugation chemistry will first convert carboxylic acid beads to forward primer (FP) beads expressing a 5'-amino-modified forward primer and 5'-amino-modified oligo(dT)<sub>25</sub> with a 3'-inverted deoxythymidine.<sup>23</sup> The resulting beads will then be emulsified with the reverse primer (RP), DNA template encoding the MS2 aptamer library, and other PCR reagents for emulsion PCR.<sup>23</sup> The emulsions will then be broken, resulting in particles expressing the DNA template encoding MS2 aptamer library and an oligo(dT)<sub>25</sub> with a 3'-inverted dT.<sup>23</sup> The beads will then undergo an emulsion transcription reaction, producing single-stranded RNA fragments which hybridize to the immobilized oligo(dT) on the beads via A-T hybridization.<sup>23</sup> The emulsions will then be broken, and the excess reagents will be removed to produce purified GRAPs expressing the MS2 aptamer library. Then, I will incubate this monoclonal RNA aptamer particle library with the fluorescently-labeled MCP and analyze the fluorescence emission profile of the binding events via FACS instrumentation as shown in **Figure 3**.<sup>23</sup> We envision that we will observe a shift in fluorescence based on the frequency and strength of the RNA-RBP binding interaction. From this, we can sort the fluorescent particles within a user-defined threshold for further DNA sequencing and characterization of specific mutation(s) in the RNA aptamer to isolate RNA aptamers that meet a target affinity and specificity based directly on their fluorescence-enhancing properties. Going forward, this method can be used for other RNA-protein interactions such as other RNA-RBP binding interactions (e.g., PP7-PCP system, com-Com system) and RNA-binding transcriptional factor systems for the identification of novel aptamer sequences that

influence the cognate binding interaction to further elucidate RNA-protein recognition mechanisms.



**Figure 13.** Synthesis of gene-linked aptamer particles for GRAP display-based technique for RNA aptamer library screening of MS2-MCP system.

## 2.4. Materials and Methods

### General

Flow cytometry data were obtained with the BD Accuri C6 flow cytometer. Gels were imaged with the BioRad ChemiDoc imaging system. LC/MS data were obtained on the 1260 Agilent Infinity Series HPLC/6230 Agilent TOF mass spectrometer. Absorbances were taken using the ThermoFisher Scientific NanoDrop One Micro-UV/Vis spectrophotometer.

**Table 1.** Oligonucleotides used for particle display-based experiments, and their corresponding sequences.

Use	Name	Sequence (5' – 3')
High-throughput particle display screening method	Aim 1A	AmM-polyT(25) /5AmMC6/ - TTT TTT TTT T TT TTT TTT TTT TTT T
	Aim 1B	647-polyA(25) /5Alex647N/ - AAA AAA AAA AAA AAA AAA AAA AAA A
	Aim 1C	CC_MS2apt_polytail_FP CCA AGT AAT ACG ACT CAC TAT AGG GGG TTT TAG AGC TAG AAA TAG CAA GTT AAA ATA AGG CTA GTC CGT TAT CAA CTT GAA AAA G
	Aim 1C	CC_MS2apt_polytail_RP TTT TTT TTT TTT TTT TTT TTT TTT TGG GCC CAA GCT TCG CAC ATG GGT GAT CCT CAT GTG CGC ACT TTT TCA AGT TGA TAA CGG A
	Aim 1D	CC_MS2-647 rGrCrA rCrArU rGrGrG rUrGrA rUrCrC rUrCrA rUrGrU rGrCrG rC /3AlexF647N/

*Aim 1. To generate aptamer particles expressing RNA aptamer of interest.*

*Aim 1A. To perform carbodiimide chemistry to couple amino-modified oligo(dT)25 to carboxylic acid beads.*

Following previously developed protocols, 500  $\mu\text{L}$  1  $\mu\text{m}$  Invitrogen Dynabeads® MyOne™ carboxylic acid beads were activated by washing once with 500  $\mu\text{L}$  0.01 M NaOH and five times with 500  $\mu\text{L}$  nuclease-free water using a DynaMag magnet. The activated carboxylic acid beads were incubated in a 500  $\mu\text{L}$  solution containing 1X conjugation buffer (200 mM NaCl, 100 nM imidazole), 20  $\mu\text{M}$  amino-modified polyT(25), and 250 mM EDC in 50% v/v DMSO. The reaction was incubated overnight at room temperature with rotation on a rotator at 10 rpm. The beads were washed twice with 500  $\mu\text{L}$  MES buffer (100 mM MES pH 4.7) for 10 minutes and twice with TT buffer (0.05% Tween-20, 50 mM Tris pH 7.5) for five minutes at room temperature at 10 rpm. The oligo(dT)25-modified beads were suspended in 500  $\mu\text{L}$  TNaTE buffer (10 mM Tris-HCl, 200 mM NaCl, 0.01% Tween-20, 0.1 mM EDTA) and stored at 4 °C.

*Aim 1B. To test the particle coating of oligo(dT)25-modified beads with AF647-labeled oligo(dA)25 probe.*

To test the conjugation efficiency of the EDC reaction, 0.2  $\mu\text{L}$  oligo(dT)25-modified beads were incubated with 1  $\mu\text{M}$  Alexa Fluor 647-polyA(25) in 100  $\mu\text{L}$  TNaTE buffer for five minutes at room temperature at 10 rpm and protected from light. After incubation, the beads were washed with TNaTE buffer, suspended in 100  $\mu\text{L}$  TNaTE buffer, and analyzed using the BD Accuri C6 flow cytometer.

*Aim 1C. To synthesize MS2 aptamer with a polyA tail by in vitro transcription.*

The DNA template for the *in vitro* transcription reaction was generated by primer extension, which followed a previously developed protocol.<sup>36</sup> Two oligonucleotides with a 20 bp overlap were annealed and extended to generate the DNA template of interest with a modified PCR reaction. A 50  $\mu$ L reaction of 1X Q5 reaction buffer (NEB), 10 mM dNTPs (NEB), 25  $\mu$ M CC\_MS2apt\_polytail\_FP (Genewiz), 25  $\mu$ M CC\_MS2apt\_polytail\_RP (Genewiz), and 1 U Q5 polymerase (NEB) was subjected to the thermal cycler with the following program: [94 °C for 5 min] + [94 °C for 20 s + annealing temperature for 30 s + 72 °C for 30 s] \* 3 cycles + [72 °C for 5 min] + [4 °C hold]. The amplified product was purified by agarose gel extraction using the QIAGEN QIAquick gel extraction kit according to the manufacturer's protocol for direct use in *in vitro* transcription applications. The resulting DNA template concentration was quantified by UV spectroscopy at Abs<sub>260</sub>. DNA integrity was analyzed by agarose gel electrophoresis with a 3% agarose gel, alongside the GeneRuler 1 kb Plus DNA ladder (ThermoFisher Scientific) in 1X TAE buffer at 120 V for 45 minutes stained with 0.33  $\mu$ g/mL EtBr solution.

The purified DNA template was subjected to an *in vitro* transcription reaction to transcribe the corresponding RNA of interest following previously developed protocols.<sup>37,38</sup> A 40  $\mu$ L reaction of 10X transcription buffer (150 mM MgCl<sub>2</sub>, 20 mM spermidine, 500 mM Tris-HCl pH 7.5, 50 mM DTT), 20 mM rNTPs (NEB), 0.3-0.5  $\mu$ g purified DNA template, and 25 U/ $\mu$ L T7 RNA polymerase (NEB) was subjected to the thermal cycler to incubate at 37 °C for three hours. The transcribed product was purified by (1) DNase I treatment, (2) acidic phenol-chloroform extraction, and (3) ethanol precipitation following previously developed protocols.<sup>39,40</sup> Per 100  $\mu$ L reaction, 10-20

$\mu\text{g}$  RNA solution was incubated with 1X DNase I reaction buffer (NEB) and 2 U DNase I endonuclease and subjected to the thermal cycler with the following program: [37 °C for 1 hr] + [75 °C for 10 min] + [4 °C hold]. After the DNase I treatment, the samples were combined and purified by an acidic phenol-chloroform extraction. The 180  $\mu\text{L}$  RNA solution was combined with 20  $\mu\text{L}$  3 M NaOAc (pH 5.2) and extracted with 200  $\mu\text{L}$  1:1 phenol-chloroform mixture (pH 4.3). The aqueous phase was extracted twice with 200  $\mu\text{L}$  chloroform to remove trace phenol residues. The 200  $\mu\text{L}$  RNA solution was then combined with 20  $\mu\text{L}$  3 M NaOAc (pH 5.2) and 600  $\mu\text{L}$  100% EtOH and incubated at -80 °C for at least one hour. After two washes with 500  $\mu\text{L}$  70% EtOH, the RNA was recovered by centrifugation at 20,000 x g for 20 minutes at 4 °C, and the purified RNA was suspended in TE buffer. The RNA concentration was quantified by UV spectroscopy at Abs<sub>260</sub>. RNA integrity was analyzed by denaturing polyacrylamide gel electrophoresis with a 10% TBE-Urea PAGE gel, alongside the Low Range ssRNA ladder (NEB), in 1X TBE buffer at 180 V for one hour stained with 1  $\mu\text{g}/\text{mL}$  EtBr solution for 30 minutes at room temperature.

*Aim 1D. To anneal the MS2 aptamer with a polyA tail to the oligo(dT)25-modified beads and analyze the RNA capture efficiency with an AF647-labeled oligo that has a complementary sequence to the RNA.*

To analyze the RNA capture efficiency on the oligo(dT)25-modified beads, 0.2  $\mu\text{L}$  oligo(dT)25-modified beads were incubated with 1  $\mu\text{M}$  RNA in 100  $\mu\text{L}$  TNaTE buffer and was subjected to the annealing reaction with the following program: [95 °C for 3 min] then gradually cool to [4 °C hold] at 0.1 °C/s. After the annealing reaction, the beads were washed once with 100  $\mu\text{L}$  TNaTE buffer and incubated with 1  $\mu\text{M}$  Alexa

Fluor 647-MS2 in 100  $\mu$ L TNaTE buffer for one hour at room temperature at 10 rpm and protected from light. After the second incubation step, the beads were washed once with TNaTE buffer, suspended in 100  $\mu$ L TNaTE buffer, and analyzed using the BD Accuri C6 flow cytometer.

*Aim 2. To purify and label cognate RNA-binding protein with fluorescent dye.*

*Aim 2A. To label and purify MCP-SoxS-his protein with an AF647 dye to bind to the MS2 aptamer adhered to oligo(dT)25-modified beads.*

Under the control of an inducible *lacI* promoter, the expression of the MS2 coat protein fused to the SoxS transcriptional activator and a 6xHis-tag was induced and expressed in *Escherichia coli* cells. Briefly, pQS17 was transformed in T7 express LysY cells and inoculated a 20 mL starter culture which grew overnight at 37 °C on an incubator shaker at 225 rpm. The starter culture was then diluted 1:50 in a 1 L solution of LB-Ampicillin and grew to  $OD_{600} = 0.5$  at 37 °C and 225 rpm. Protein expression was then induced at 37 °C and 225 rpm for three hours with the addition of 500 IPTG to obtain a final concentration of 0.5 mM IPTG. Cells were harvested with centrifugation at 6500 x rpm for 15 minutes at 4 °C. The cell pellets were suspended in 12 mL solubilization buffer (6 M guanidine-HCl, 20 mM Tris pH 7, 300 mM NaCl, 5% v/v glycerol, 5 mM imidazole), and cells were lysed using sonication at 4 °C for five cycles of 20 short pulses. The cell debris was separated from the cell lysate by centrifugation at 14000 x g for 30 minutes at 4 °C. The cell lysate was purified with a Ni-NTA column, and the protein of interest was eluted with fractions of 800  $\mu$ L elution buffer (4 M guanidine-HCl, 20 mM NaHPO<sub>4</sub> pH 5.5, 300 mM NaCl, 300 mM imidazole). The fractions eluted from the Ni-NTA column were concentrated to 0.5 mL with the Amicon centrifugal unit (3K MWCO) and dialyzed

against SEC column buffer (4 M guanidine-HCl, 20 mM NaHPO<sub>4</sub> pH 5.5, 300 mM NaCl) using the BioRad Micro Bio-Spin 6 chromatography column according to the manufacturer's instructions. The protein sample was then further purified with the HiPrep 16/40 Sephacryl S-100 high-resolution SEC column. The fractions eluted from the column were monitored for protein at Abs<sub>280</sub>. The fractions that contained protein were separated and desalted against dialysis buffer (20 mM NaHPO<sub>4</sub> pH 5.5, 300 mM NaCl, 40% glycerol) using the BioRad Micro Bio-Spin 6 chromatography column according to the manufacturer's instructions. Every other fraction that contained protein was denatured in SDS at 90 °C for 10 minutes and analyzed by gel electrophoresis with a 4-12% Bis-Tris PAGE gel, alongside the PageRuler Plus protein ladder (ThermoFisher Scientific), in 1X MES running buffer at 200 V for 50 minutes stained with Coomassie blue for one hour at room temperature. The purified MCP-SoxS-his protein was also characterized by LC/MS to determine its size (MW = 26.7 kDa), consistent with the expected size (MW = 26.7 kDa). The MCP-SoxS-his protein concentration was quantified by UV spectroscopy at Abs<sub>280</sub>.

To label the MS2 coat protein, the Alexa Fluor 647 protein labeling kit from Molecular Probes was followed according to the manufacturer's directions.<sup>44</sup> The purified MCP-SoxS-his protein was diluted to 2 mg/mL in SEC column buffer, and 50 µL 1 M NaHCO<sub>3</sub> was added to the protein solution to adjust the pH. The protein solution was then transferred to a vial of reactive dye and stirred at 1 hour at room temperature, protected from light. The labeled protein was purified by the Cytiva illustra NAP-5 column Sephadex G-25 equilibrated with SEC column buffer to remove the free dye. The eluted labeled protein fractions were denatured in SDS at 90 °C for 10 minutes and analyzed by

gel electrophoresis with a 4-12% Bis-Tris PAGE gel, alongside the PageRuler Plus protein ladder (ThermoFisher Scientific), in 1X MES running buffer at 200 V for 50 minutes stained with Coomassie blue for one hour at room temperature. The AF647-labeled MCP-SoxS-his protein concentration was quantified by UV spectroscopy at Abs<sub>280</sub>.

*Aim 2B. To label and purify SoxS-his protein with an AF647 dye as the negative control.*

Under the control of an inducible *lacI* promoter, the expression of the negative control protein that does not express the MS2 coat was induced and expressed in *Escherichia coli* cells following the protocol parallel to the one described in Aim 2A. Briefly, pQS18 was transformed in T7 express LysY cells and inoculated a starter culture. The SoxS-his protein expression via IPTG induction, cell lysis via sonication, and Ni-NTA column purification were followed according to the protocol described in Aim 2A. The fractions eluted from the Ni-NTA column were concentrated to 0.5 mL with the Amicon centrifugal unit (3K MWCO) and dialyzed against dialysis buffer using the BioRad Micro Bio-Spin 6 chromatography column according to the manufacturer's instructions. The protein fractions were denatured in SDS at 90 °C for 10 minutes and analyzed by gel electrophoresis with a 16% Tris-Tricine PAGE gel, alongside the Spectra Low Range protein ladder (ThermoFisher Scientific), in 1X tricine running buffer at 125 V for 1.5 hours stained with Coomassie blue for one hour at room temperature. The purified SoxS-his protein was characterized by LC/MS to determine its size (MW = 13.5 kDa). The SoxS-his protein concentration was quantified by UV spectroscopy at Abs<sub>280</sub>. To label the negative control SoxS-his protein, the Alexa Fluor 647 protein labeling protocol described in Aim 2A was followed according to the manufacturer's directions.<sup>44</sup>

The purified SoxS-his protein was diluted to 2 mg/mL in SEC column buffer, and 50  $\mu$ L 1 M NaHCO<sub>3</sub> was added to the protein solution to adjust the pH. The protein solution was then transferred to a vial of reactive dye and stirred at 1 hour at room temperature, protected from light. The labeled protein was purified by the Cytiva illustra NAP-5 column Sephadex G-25 equilibrated with SEC column buffer to remove the free dye. The eluted labeled protein fractions were denatured in SDS at 90 °C for 10 minutes and analyzed by gel electrophoresis with a 16% Tris-Tricine PAGE gel, alongside the Spectra Low Range protein ladder (ThermoFisher Scientific), in 1X tricine running buffer at 125 V for 1.5 hours stained with Coomassie blue for one hour at room temperature. The AF647-labeled SoxS-his protein concentration was quantified by UV spectroscopy at Abs<sub>280</sub>.

*Aim 3. To perform binding reactions and measure the fluorescence emission profile to quantify the binding affinity.*

*Aim 3A. MS2-MCP binding affinity determination*

To determine the MS2-MCP binding affinity, several binding reactions of 1  $\mu$ M RNA annealed to oligo(dT)25-modified beads at a range of AF647-labeled MCP-SoxS-protein concentrations were performed, and the resulting fluorescence emission spectra were measured. Per sample, 0.2  $\mu$ L oligo(dT)25-modified beads were incubated with 1  $\mu$ M RNA in 100  $\mu$ L TNaTE buffer and was subjected to the annealing reaction with the following program: [95 °C for 3 min] then gradually cool to [4 °C hold] at 0.1 °C/s. 1  $\mu$ M RNA was annealed to the oligo(dT)25-modified beads, and the beads were washed with TNaTE buffer. After the annealing reaction, the beads were washed with 100  $\mu$ L TNaTE buffer and incubated with a range of AF647-labeled MCP-SoxS-his protein (0, 1, 2, 4,

10, 20, 40 nM) in 100  $\mu$ L TNaTE buffer for one hour at room temperature at 10 rpm and protected from light. After the binding reaction, the beads were washed with TNaTE buffer, suspended in 100  $\mu$ L TNaTE buffer, and analyzed using the BD Accuri C6 flow cytometer. For each sample, the mean APC-A intensity of the modified beads was measured for 100,000 events in the first gate for  $n = 3$  repetitive experiments and for a consistent number of events in the lower right quadrant, which was then plotted against the protein concentration. The binding experiment was performed in triplicate. The data were fit using GraphPad software using a nonlinear regression program for receptor binding experiments (saturation binding: one site: specific binding), which calculated the  $K_d$  of this MS2-MCP system to be 4.853 nM.

*Aim 3B. To assess AF647-labeled SoxS-his protein and AF647-labeled BSA protein as negative controls against the MS2-MCP system.*

To confirm the specificity of the MS2-MCP system, AF647-labeled SoxS-his protein was subjected to the same protocol as the AF647-labeled MCP-SoxS-his protein described in Aim 3A. Per sample, 0.2  $\mu$ L oligo(dT)25-modified beads were incubated with 1  $\mu$ M RNA in 100  $\mu$ L TNaTE buffer and was subjected to the annealing reaction with the following program: [95  $^{\circ}$ C for 3 min] then gradually cool to [4  $^{\circ}$ C hold] at 0.1  $^{\circ}$ C/s. 1  $\mu$ M RNA was annealed to the oligo(dT)25-modified beads, and the beads were washed with TNaTE buffer. After the annealing reaction, the beads were washed with 100  $\mu$ L TNaTE buffer and incubated with a range of AF647-labeled SoxS-his protein (0, 1, 2, 4, 10, 20, 40 nM) in 100  $\mu$ L TNaTE buffer for one hour at room temperature at 10 rpm and protected from light. After the binding reaction, the beads were washed with TNaTE buffer, suspended in 100  $\mu$ L TNaTE buffer, and analyzed using the BD Accuri C6 flow

cytometer. For each sample, the mean APC-A intensity of the modified beads was measured for 100,000 events in the first gate for  $n = 3$  repetitive experiments.

To further confirm the specificity of the MS2-MCP system, the AF647-labeled BSA protein was subjected to the same protocol as the AF647-labeled MCP-SoxS-his protein and AF647-labeled SoxS-his protein described previously. Per sample, 0.2  $\mu\text{L}$  oligo(dT)25-modified beads were incubated with 1  $\mu\text{M}$  RNA in 100  $\mu\text{L}$  TNaTE buffer and was subjected to the annealing reaction with the following program: [95  $^{\circ}\text{C}$  for 3 min] then gradually cool to [4  $^{\circ}\text{C}$  hold] at 0.1  $^{\circ}\text{C}/\text{s}$ . 1  $\mu\text{M}$  RNA was annealed to the oligo(dT)25-modified beads, and the beads were washed once with TNaTE buffer. After the annealing reaction, the beads were washed once with 100  $\mu\text{L}$  TNaTE buffer and incubated with a range of AF647-labeled BSA protein (0, 1, 2, 4, 10, 20, 40 nM) in 100  $\mu\text{L}$  TNaTE buffer for one hour at room temperature at 10 rpm and protected from light. After the binding reaction, the beads were washed once with TNaTE buffer, suspended in 100  $\mu\text{L}$  TNaTE buffer, and analyzed using the BD Accuri C6 flow cytometer. For each sample, the mean APC-A intensity of the modified beads was measured for 100,000 events in the first gate for  $n = 3$  repetitive experiments.

To cross-check the binding specificity of the MS2-MCP system at higher concentrations of the AF647-labeled protein, the concentrations of the AF647-labeled MCP-SoxS-his, the AF647-labeled SoxS-his, and the AF647-labeled BSA proteins were increased.

Following the same protocols described, 1  $\mu\text{M}$  oligo(dT)25-modified beads expressing the RNA were incubated with a range of the AF647-labeled protein (0, 100, 200, and 400 nM) for one hour at room temperature at 10 rpm and protected from light. After each

sample was washed and suspended in buffer, the mean APC-A intensity of the modified beads was measured for 100,000 events in the first gate.

## 2.5. References

[24] Witherell, G. W.; Gott, J. M.; Uhlenbeck, O. C., Specific interaction between RNA phage coat proteins and RNA. In *Progress in Nucleic Acid Research and Molecular Biology*, Cohn, W. E.; Moldave, K., Eds. Academic Press: 1991; Vol. 40, pp 185-220.

[25] Johansson, H. E.; Liljas, L.; Uhlenbeck, O. C., RNA recognition by the MS2 phage coat protein. *Seminars in Virology* **1997**, *8* (3), 176-185.

[26] Stockley, P. G.; Stonehouse, N. J.; Murry, J. B.; Goodman, S. T. S.; Talbot, S. J.; Adams, C. J.; Liljas, L.; Valegå, K., Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Research* **1995**, *23* (13), 2512-2518.

[27] Chao, J. A.; Patskovsky, Y.; Almo, S. C.; Singer, R. H., Structural basis for the coevolution of a viral RNA–protein complex. *Nature Structural & Molecular Biology* **2008**, *15* (1), 103-105.

[28] Yoon, J.-H.; Srikantan, S.; Gorospe, M., MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. *Methods* **2012**, *58* (2), 81-87.

[29] Yoon, J.-H.; Gorospe, M., Identification of mRNA-interacting factors by MS2-TRAP (MS2-Tagged RNA Affinity Purification). In *RNA-Protein Complexes and Interactions: Methods and Protocols*, Lin, R.-J., Ed. Springer New York: New York, NY, 2016; pp 15-22.

[30] Tutucci, E.; Vera, M.; Biswas, J.; Garcia, J.; Parker, R.; Singer, R. H., An improved MS2 system for accurate reporting of the mRNA life cycle. *Nature Methods* **2018**, *15* (1), 81-89.

[31] Kolarova, H.; Hengerer, B., Preparation of magnetic oligo (dT) particles. *BioTechniques* **1996**, *20* (2), 196–198.

[32] Sehgal, D.; Vijay, I. K., A method for the high efficiency of water-soluble carbodiimide-mediated amidation. *Anal Biochem.* **1994**, *218* (1), 87-91.

[33] ThermoFisher Scientific. Dynabeads® MyOne™ Carboxylic Acid. <https://www.thermofisher.com/us/en/home/references/protocols/proteins-expression-isolation-and-analysis/protein-isolation-protocol/dynabeads-myone-carboxylic-acid.html> (accessed September, 2021).

[34] ThermoFisher Scientific. Carbodiimide Crosslinker Chemistry. <https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology->

learning-center/protein-biology-resource-library/pierce-protein-methods/carbodiimide-crosslinker-chemistry.html (accessed September 2021).

[35] ThermoFisher Scientific. Invitrogen™ Alexa Fluor™ 647 NHS Ester (Succinimidyl Ester). <https://www.thermofisher.com/order/catalog/product/A20006#/A20006> (accessed September, 2021).

[36] Stemmer, W. P. C.; Cramer, A.; Ha, K. D.; Brennan, T.; Heyneker, H. L., Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxynucleotides. *Gene* **1995**. *164* (1). 49-53.

[37] Milligan, J. F.; Uhlenbeck, O. C., Synthesis of small RNAs using T7 RNA polymerase. *Methods in Enzymology* **1989**. *180*, 51-62.

[38] NEB. *In vitro* transcription of short templates (50-300 nt). <https://www.neb.com/protocols/0001/01/01/2-in-vitro-transcription-of-short-templates-50-300-nt-e2030>. (accessed September, 2021).

[39] NEB. DNase I. <https://www.neb.com/products/m0303-dnase-i-rnase-free#Product%20Information> (accessed September, 2021).

[40] NEB. Purification of synthesized RNA. <https://www.neb.com/protocols/0001/01/01/purification-of-synthesized-rna-e2040> (accessed September, 2021).

[41] NEB. pTYB21 Vector. <https://www.neb.com/products/n6709-ptyb21-vector#Product%20Information>. (accessed September, 2021).

[42] ThermoFisher Scientific. Thermo Scientific™ HisPur™ Ni-NTA Resin. <https://www.thermofisher.com/order/catalog/product/88221#/88221> (accessed September, 2021).

[43] Cytiva Life Sciences. HiPrep Sephacryl S-100 HR. <https://www.cytivalifesciences.com/en/us/shop/chromatography/prepacked-columns/size-exclusion/hiprep-sephacryl-s-100-hr-p-05799> (accessed September, 2021).

[44] ThermoFisher Scientific. Invitrogen™ Alexa Fluor™ 647 Protein Labeling Kit. <https://www.thermofisher.com/order/catalog/product/A20173#/A20173> (accessed September, 2021).

[45] AAT Bioquest. Protocol for Labeling Antibodies and Proteins with AF647 NHS Ester. <https://www.aatbio.com/resources/protocols/protocol-for-labeling-antibodies-and-proteins-with-af647-nhs-ester-alexa-fluor-647-nhs-ester#selProtein> (accessed September, 2021).

[46] BioRad. Bio-Gel P-30 gel. <https://www.bio-rad.com/en-us/sku/1504150-bio-gel-p-30-gel?ID=1504150> (accessed September, 2021).

[47] Cytiva Life Sciences. NAP DNA Purification Columns. <https://www.cytivalifesciences.com/en/us/shop/molecular-biology/purification/gel-filtration-columns/illustra-nap-columns-p-00045> (accessed September, 2021).

[48] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; *Protein Identification and Analysis Tools on the ExPASy Server*; (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press (2005). pp. 571-607

[49] ThermoFisher Scientific. Invitrogen™ Albumin from Bosine Serum (BSA) Alexa Fluor™ 647 conjugate.

<https://www.thermofisher.com/order/catalog/product/A34785#/A34785> (accessed September, 2021).

[50] Valegård, K.; Murray, J. B.; Stonehouse, N. J.; van den Worm, S.; Stockley, P. G.; Liljas, L., The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *Journal of Molecular Biology* **1997**, 270 (5), 724-738.

[51] Convery, M. A.; Rowsell, S.; Storehouse, N. J.; Ellington, A. D.; Hirao, I.; Murray, J. B.; Peabody, D. S.; Phillips, S. E. V.; Stockley, P. G., Crystal structure of an RNA aptamer–protein complex at 2.8 Å resolution. *Nature Structural Biology* **1998**, 5 (2), 133-139.

[52] Rowsell, S.; Stonehouse, N. J.; Convery, M. A.; Adams, C. J.; Ellington, A. D.; Hirao, I.; Peabody, D. S.; Stockley, P. G.; Phillips, S. E. V., Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nature Structural Biology* **1998**, 5 (11), 970-975.

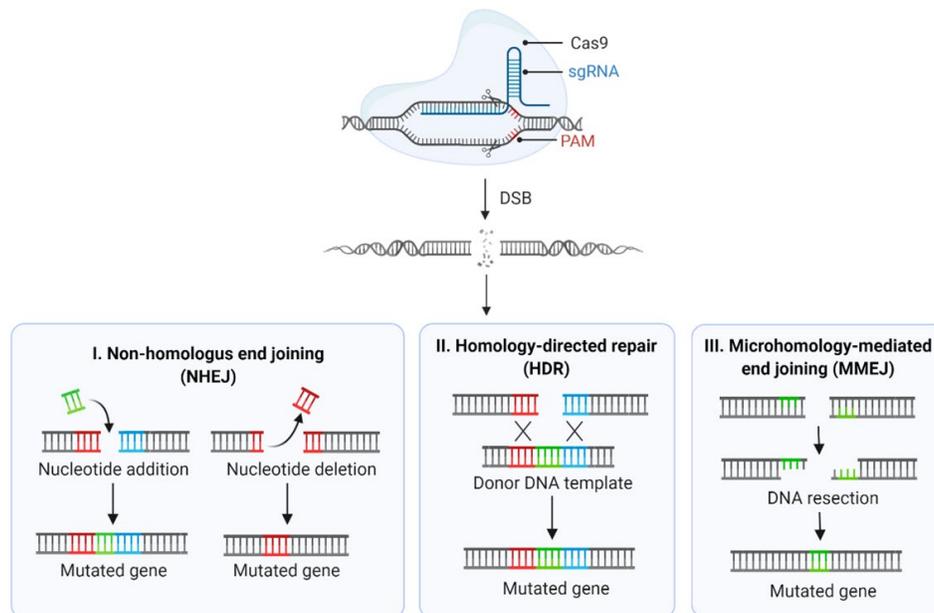
## **Chapter 3. Introduction for prime editor-mediated programmable insertion of UAAs into endogenous proteins**

### **3.1. CRISPR-Cas systems**

The adaption of CRISPR-Cas systems for programmable and site-directed mutagenesis has transformed the field of synthetic biology. Native to many bacterial species, clustered regularly interspaced short palindromic repeats (CRISPR) systems utilize CRISPR-associated proteins (Cas) as an adaptive immunity system to recognize and combat bacteriophage reinfection.<sup>53,54</sup> When a bacteriophage infects a bacterial cell, the virus will attach to the cell surface and inject its viral genome into the cell, initiating the *de novo* synthesis of viral proteins and genome using the host cell's machinery.<sup>53,54</sup> The viral proteins and newly replicated viral genome are assembled into virions for viral release and infection of other cells.<sup>53,54</sup> The CRISPR-Cas mechanism for adaptive bacterial immunity involves three stages: (1) spacer acquisition, (2) CRISPR-RNA (crRNA) processing, and (3) interference. Following bacteriophage infection, (1) the bacteria cell cleaves the viral genome and incorporates small fragments of the foreign DNA, called spacers, between the repeated sequences of the CRISPR locus.<sup>53,54</sup> The bacteria cell machinery specifically selects the fragment of the foreign DNA that is adjacent to the protospacer adjacent motif (PAM) sequence, such that it can systematically recognize this sequence in viral genomes to add to the CRISPR locus.<sup>53,54</sup> (2) The spacers are then transcribed and parsed into crRNAs to (3) guide the associated Cas nucleases to make a crRNA-Cas complex to recognize and destroy invading foreign DNA with complementary sequences.<sup>53,54</sup>

In the CRISPR-Cas family, three types of CRISPR-Cas systems have been identified (Type I-III) for adaptive microbial immunity.<sup>53,54</sup> Type II CRISPR-Cas systems are one of the best characterized, and thereby, have been widely reconstituted for genome engineering purposes.<sup>53,54,55,56</sup> In these systems, trans-activating CRISPR RNA (tracrRNA) are bound to the repeated sequences of the CRISPR locus, which forms double-stranded RNA to trigger cleavage by the Cas9 and RNase III enzymes to produce RNA fragments containing the CRISPR repeat, viral genome spacer, and tracrRNA sequences.<sup>53,54</sup> When a bacteriophage returns to infect a bacterial cell, the Cas9 enzyme will recognize the PAM sequence, and the spacer will recognize the viral genome DNA.<sup>53,54</sup> The Cas9 enzyme will undergo a double-stranded break (DSB), mediated by the HNH and RuvC/RNaseH-like endonuclease domains, to cleave the viral genome for the prevention of reinfection.<sup>53,54</sup> To exploit Type II CRISPR-Cas systems for genome engineering, a chimeric single-guide RNA (sgRNA), consisting of the crRNA and tracrRNA, is designed to direct the Cas nuclease and induce a DSB at a targeted site that is followed by the appropriate PAM sequence.<sup>53,54,55,56</sup> The DSB undergoes cellular repair mechanisms, such as non-homologous end joining (NHEJ), homology-directed repair (HDR), or microhomology-mediated end joining (MMEJ), to mediate insertions, deletions, translocations, or other DNA arrangements at the DSB site (**Figure 14**).<sup>53,54,55,56</sup> However, NHEJ-, HDR-, and MMEJ-resolved DSBs frequently introduce non-specific insertions or deletions (indels) that result in frameshift mutations that may inhibit protein synthesis and result in semi- or non-functional proteins. Nevertheless, these CRISPR-based systems and their derivatives that generate more precise mutagenesis on

single-nucleotide levels have been explored and applied across several domains of life as a biotechnological tool for the development of therapeutics and diagnostic tools.



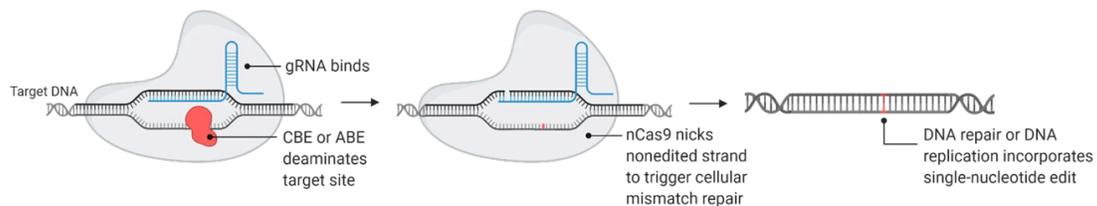
**Figure 14.** Schematic overview of CRISPR-Cas systems for genome editing.

### 3.2. CRISPR-Cas technologies

The promising future of CRISPR-Cas for genome engineering has fiercely inspired scientists to develop CRISPR-Cas-based technologies to advance and expand its applications in synthetic biology and biomedical sciences.<sup>57</sup> Not only have many researchers launched biotechnological companies for CRISPR-Cas development, but pharmaceutical companies and clinically-based research laboratories have sought to translate CRISPR-Cas systems from benchtop to bedside by researching genome editing-based therapeutics for life-threatening genetic diseases. Recently, the development of CRISPR-guided base editors and prime editors has led to more programmable genome editing with a reduced number of off-target effects, showing potential for its translation to clinical gene therapy.

### 3.2.1. Base editing

Developed by David Liu and coworkers, base editors are composed of fusions of nickase Cas9 (nCas9) or catalytically-dead Cas9 (dCas9) and a cytosine or adenine deaminase.<sup>58,59</sup> CRISPR-based cytosine base editors (CBEs) and CRISPR-based adenine base editors (ABEs) use a programmed sgRNA to guide the nickase or catalytically-dead Cas9 enzyme to the PAM sequence to unwind the complementary DNA for enzyme-mediated base conversion (**Figure 15**).<sup>58,59</sup> CBEs mediate C:G > T:A substitutions by converting cytosine to uracil, which converts to thymine through DNA replication or repair, whereas ABEs mediate A:T > G:C substitutions by converting adenine to inosine, which is read as guanosine by polymerases.<sup>58,59</sup> Base editors can execute site-specific editing of nucleobases at the targeted gene without introducing DSBs to perform all four possible transition mutations while maintaining a low indel rate; however, improvements for expansion in targeting scope, reduction of off-target activity, and safe delivery in cell lines and primary cells are still needed.<sup>58,59,60</sup>

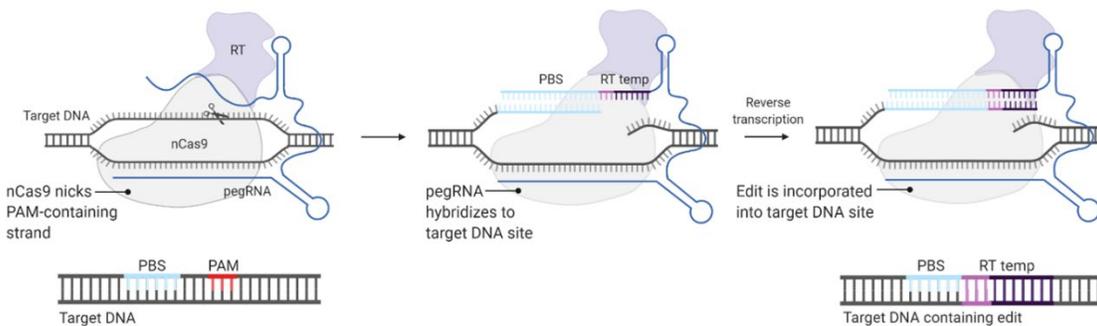


**Figure 15.** Schematic overview of base editing for genome editing.

### 3.2.2. Prime editing

More recently developed by Liu and coworkers, prime editors are genetic editing tools that perform site-directed transversions, insertions, and deletions without the requirement of double-stranded breaks or donor DNA templates.<sup>62</sup> Illustrated in **Figure 16**, the general construct of prime editors includes a prime editing complex (PE-pegRNA) containing a reverse transcriptase (RT) domain, prime editing guide RNA (pegRNA), and

nickase Cas9 (nCas9).<sup>62</sup> Specifically, the pegRNA will contain the primer binding site (PBS), which binds to the target DNA, and an RT template that encodes the edit.<sup>62</sup> Once activated, the pegRNA will site-specifically target and bind to the DNA, encouraging nCas9 to nick the PAM-containing strand at the programmed pegRNA nick site.<sup>62</sup> This nick will result in a 3'-hydroxyl group that will initiate the reverse transcription of the edit-encoded pegRNA extension, allowing the edited DNA strand to be incorporated into the target site after the equilibration of the two ssDNA flaps.<sup>62</sup> Notably, prime editing can mediate all 12 types of point mutations, insertions up to 44 bp, deletions up to 80 bp, and combinations of these.<sup>62</sup> Although prime editing shows less off-target activity than CRISPR-Cas systems, it has indicated a higher indel frequency compared to base editing. Nevertheless, the versatility of prime editing has advanced this tool to the forefront of genome editing technologies.<sup>60,62</sup>



**Figure 16.** Schematic overview of prime editing for genome editing.

### 3.3. References

- [53] Jinek M.; Chylinski K.; Fonfara I.; Hauer M.; Doudna J. A.; Charpentier, E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337* (6096), 816-821.
- [54] Ran, F.; Hsu, P.; Wright, J., Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **2013**, *8* (11), 2281–2308.
- [55] Cong, L.; Ran, F.A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P.D.; Wu, X.; Jiang, W.; Marraffini, L.A.; et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, *339*, 819–823.

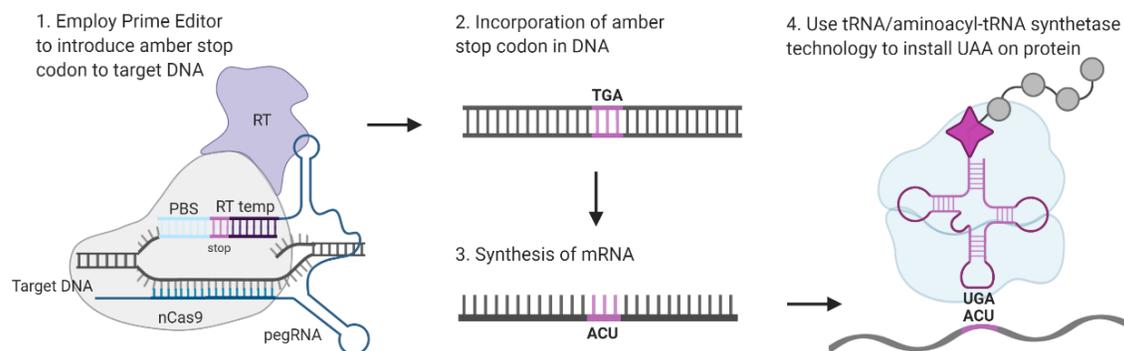
- [56] Mali, P.; Yang, L.; Esvelt, K.M.; Aach, J.; Guell, M.; DiCarlo, J.E.; Norville, J.E.; Church, G.M., RNA-guided human genome engineering via Cas9. *Science* **2013**, *339*, 823–826.
- [57] Pickar-Oliver, A.; Gersbach, C. A., The next generation of CRISPR–Cas technologies and applications. *Nature Reviews Molecular Cell Biology* **2019**, *20* (8), 490–507.
- [58] Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016**. 533, 420-424.
- [59] Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H.; Bryson, D. I.; and Liu, D. R., *Nature* **2017**. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *551*, 464-471.
- [60] Porto, E.M.; Komor, A.C.; Slaymaker, I.M., Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov* **2020**. *19*, 839–859.
- [61] Katti, A.; Foronda, M.; Zimmerman, J.; Diaz, B.; Zafra, M. P.; Goswami, S.; Dow, L. E., GO: a functional reporter system to identify and enrich base editing activity. *Nucleic Acids Research* **2020**, *48* (6), 2841-2852.
- [62] Anzalone, A.V.; Randolph, P.B.; Davis, J.R., Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **2019**. *576*, 149–157.

## Chapter 4. Prime editor-mediated programmable insertion of UAAs into endogenous proteins

### 4.1. Introduction

The use of unnatural amino acids (UAAs) for site-specific bioconjugation onto transiently expressed proteins for genetic code expansion has been extensively studied in cells and animals to manipulate and monitor biological processes.<sup>63,64,65,66</sup> With the discovery of orthogonal aminoacyl-tRNA synthetase/tRNA pairs, many biologically important UAAs have been incorporated into proteins to facilitate photocrosslinking and embed fluorescent probe motifs to proteins for *in vivo* protein signaling, transport, localization, and conformational studies.<sup>63</sup> However, the transient expression of the target proteins is limited by the transfection efficiency, and further, it is challenging to genetically direct post-translational modifications (PTMs) to these modified amino acids.<sup>63</sup> To combat these challenges, we hypothesize that prime editing technology can seamlessly and site-specifically insert a stop codon to incorporate UAAs. By combining the aforementioned tRNA/aminoacyl-tRNA synthetase techniques with prime editing, we propose employing prime editors to introduce an UAA to an endogenous protein for *in vivo* cellular signaling and trafficking patterns, dramatically expanding the toolbox of prime editing applications. Briefly, we propose to introduce a stop codon into mammalian cells by prime editing for confirmation with a reporter system, in which this stop codon can mediate the introduction of a functionalized UAA onto a model protein through tRNA/aminoacyl-tRNA synthetase technology for downstream imaging applications and proteomic analyses (**Figure 17**). Specifically, the pegRNA of the PE-pegRNA complex comprising of the nCas9 domain fused to an RT domain will direct the fusion protein to

the target DNA of the protein of interest.<sup>62</sup> The pegRNA will consist of the PBS that will hybridize to the target DNA and the RT template that encodes the desired TGA stop codon.<sup>62</sup> The nCas9 will nick the PAM-containing strand at the pegRNA nick site of the target DNA, allowing for hybridization of the PBS with the downstream RT template encoding the edit to the PAM-containing strand and priming the reverse transcription of the edit-encoding extension into the target DNA.<sup>62</sup> This will trigger cellular repair mechanisms to stably incorporate the TGA stop codon into the genome, later transcribed into the ACU codon. By using tRNA/aminoacyl-tRNA synthetase technology, a modified tRNA synthetase will recognize its orthogonal UAA introduced in the cell and aminoacylate its orthogonal tRNA.<sup>63</sup> Thus, once the UAA-charged tRNA complex is activated by the orthogonal aminoacyl-tRNA synthetase, the ribosome can read and transfer the UAA from the tRNA onto the growing polypeptide chain, systematically incorporating an UAA onto the protein of interest.<sup>63</sup> Using prime editing and tRNA/aminoacyl-tRNA synthetase technology to incorporate an UAA on endogenous proteins will allow for downstream imaging and proteomic analyses via bioconjugation and fluorescent-based tagging and enrichment studies.

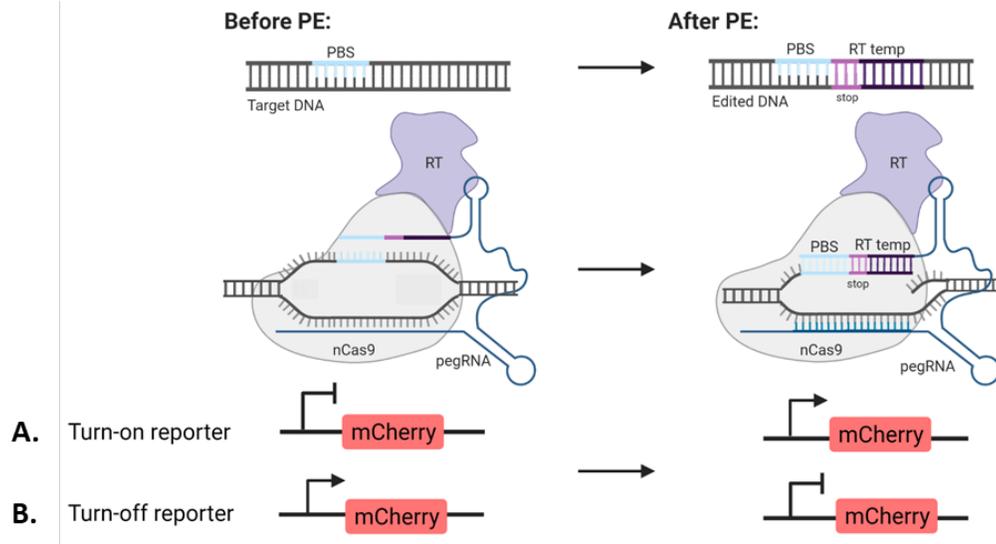


**Figure 17.** Schematic overview of the installation of an UAA into an endogenous protein using prime editing and tRNA-aminoacyl-tRNA synthetase technologies.

## 4.2. Results/Discussion

*Aim 1. To validate the insertion of a stop codon by expressing a fluorescent reporter in a stable HEK293T cell line with prime editing.*

*Aim 1A. To transiently transfect HEK293T cells with reporter plasmid to test if the PE linker is functional.*



**Figure 18.** Prime editing-mediated turn-on and turn-off fluorescent reporter system. Expected mCherry gene expression before and after prime editing of **A.** turn on and **B.** turn off fluorescent reporter system.

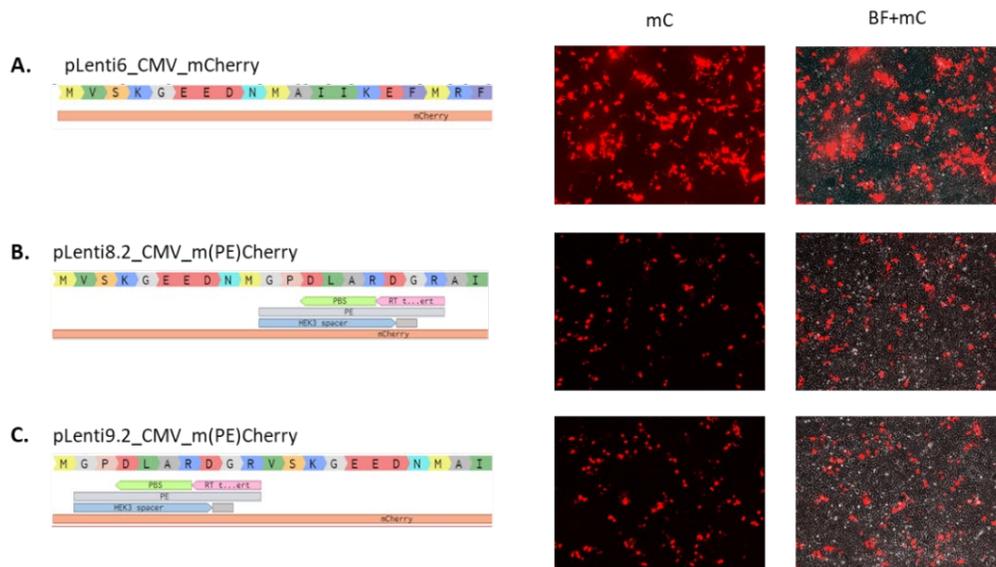
The reporter plasmid design was inspired by previous research in which the editing efficiency can be monitored with a fluorescent reporter system.<sup>61,62</sup> There will be a “turn on” reporter plasmid and a “turn off” reporter plasmid, such that it will either activate or deactivate mCherry expression after prime editing (**Figure 18**). The “turn on” reporter plasmids will incorporate the TGA stop codon within the PE linker, so it will only be able to express mCherry if prime editing deletes the stop codon upstream of the fluorescent protein. The “turn off” reporter plasmids do not incorporate a stop codon within the PE linker, so it will only be able to stop the expression of mCherry if prime editing inserts a stop codon upstream the fluorescent protein. Many different variations of this reporter plasmid duo were tested, and the results are described herein.

The first series of reporter plasmids tested was pLenti4\_CMV\_(PE)mCherry, pLenti5.1\_CMV\_m(PE\*)Cherry, and pLenti5.2\_CMV\_m(PE)Cherry (**Appendix 2**). pLenti4 has the full-length PE linker upstream of the mCherry sequence, pLenti5.1 has the truncated PE linker with the stop codon within the mCherry sequence, and pLenti5.2 has the truncated PE linker with the CTT insert within the mCherry sequence. For each construct, HEK293T cells were transfected with 5 µg of the reporter plasmid at approximately 70% confluency with 1 µg/µL polyethylenimine (PEI). Cell viability was monitored every 24 hours with the bright-field microscope, and the transfection efficiency was analyzed after 72 hours with the fluorescent microscope. If the PE linker is functional for the pLenti4 construct, it will not express mCherry due to the stop codon; however, many of the cells imaged expressed mCherry, which indicates that the PE linker was likely not functional, possibly due to the multiple translational sites of the fluorescent protein. The cells transfected with the pLenti5.1 construct did not show fluorescence, which was expected since the truncated PE linker has the stop codon. To confirm that the absence of mCherry was because of the stop codon within the PE linker of pLenti5.1, the pLenti5.2 construct was also tested. Here, if the PE linker is functional and its insertion does not disrupt the protein folding, there would be mCherry expression post-transfection; however, the cells did not express the fluorescent protein. This suggests that the PE linker disrupts the fluorescent protein folding, challenging the functionality of the PE linker in the pLenti5.1 construct.

The second series of reporter plasmids tested was pLenti6\_CMV\_mCherry and pLenti7.2\_CMV\_m(PE)Cherry (**Appendix 2**). pLenti6 acts as the positive control for the reporter plasmid experiments since it does not have the PE linker with a stop codon to

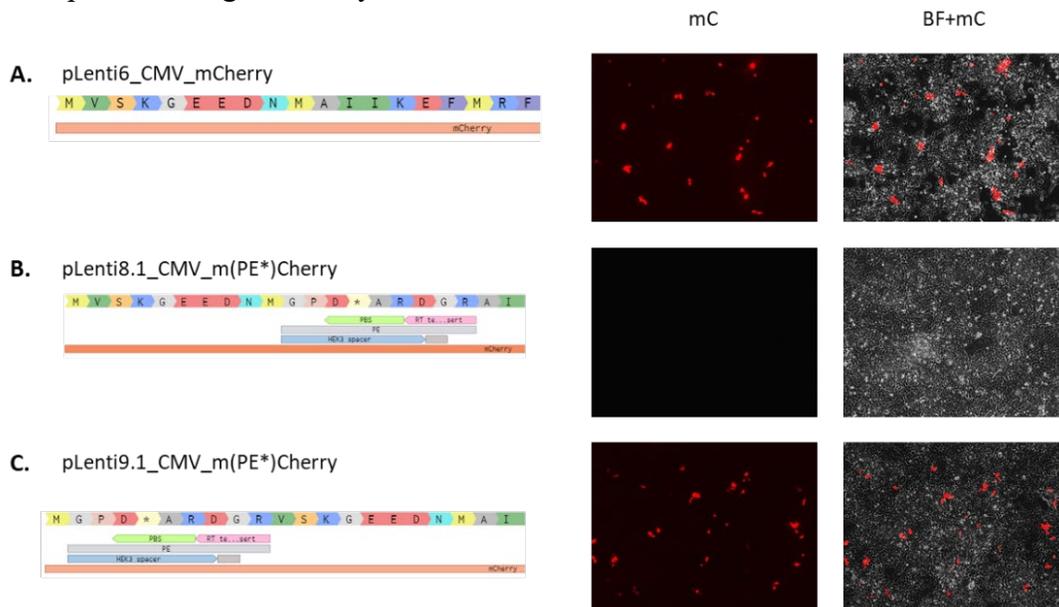
halt mCherry expression or potentially alter the folding of mCherry. The pLenti7.2 construct has the truncated PE linker with the CTT insert after the fourth methionine, which should express mCherry post-transfection; however, this construct did not, which is likely due to improper protein folding since the PE linker is positioned on the beta-barrel of the mCherry.

The third series of reporter plasmids tested was pLenti6\_CMV\_mCherry, pLenti8.2\_CMV\_m(PE)Cherry, and pLenti9.2\_CMV\_m(PE)Cherry (**Appendix 2**). The pLenti6 construct is a positive control, pLenti8.2 has the truncated PE linker with the CTT insert after the second methionine in the mCherry sequence, and pLenti9.2 has the truncated PE linker with the CTT insert after the first methionine in the mCherry sequence. The linker with the CTT insert was introduced after the first two methionines since it is positioned on the alpha helix, which will hopefully combat any protein folding alterations. After 72 hours, the HEK293T cells transfected with pLenti8.2 and pLenti9.2 expressed mCherry (**Figure 19B, C**), along with the pLenti6 positive control (**Figure 19A**), indicating that the position of the PE linker did not alter the folding of the fluorescent protein.



**Figure 19.** Transfection efficiency of turn-off fluorescent reporter. Images from left to right show the amino acid sequence of mCherry region with or without PE linker, HEK293T cells expressing reporter plasmid imaged with mCherry filter and brightfield + mCherry filter 72 hours post-transfection of **A.** pLenti6\_CMV\_mCherry, **B.** pLenti8.2\_CMV\_m(PE)Cherry, and **C.** pLenti9.2\_CMV\_m(PE)Cherry.

Since pLenti8.2 and pLenti9.2 expressed mCherry as expected, we could test the next construct with the PE linker with the stop codon. Thus, the fourth series of reporter plasmids tested was pLenti6\_CMV\_mCherry, pLenti8.1\_CMV\_m(PE\*)Cherry, and pLenti9.1\_CMV\_m(PE\*)Cherry (**Appendix 2**). The pLenti6 construct is a positive control, pLenti8.1 has the truncated PE linker with the stop codon after the second methionine in the mCherry sequence, and pLenti9.1 has the truncated PE linker with the stop codon after the first methionine in the mCherry sequence. After 72 hours, the HEK293T cells transfected with pLenti8.1 did not express mCherry (**Figure 20B**), and those transfected with pLenti6 and pLenti9.1 expressed mCherry (**Figure 20A, C**). The pLenti8.1 construct that has the PE linker with the stop codon positioned after the second methionine was consistent with what was expected, but the pLenti9.1 construct was not. This suggests that the pLenti8.1 and pLenti8.2 constructs would be the most promising to monitor prime editing efficiency in mammalian cells.



**Figure 20.** Transfection efficiency of turn-on fluorescent reporter. Images from left to right show the amino acid sequence of mCherry region with or without PE linker, HEK293T cells expressing reporter plasmid imaged with mCherry filter and brightfield + mCherry filter 72 hours post-transfection of **A.** pLenti6\_CMV\_mCherry, **B.** pLenti8.1\_CMV\_m(PE\*)Cherry, and **C.** pLenti9.1\_CMV\_m(PE\*)Cherry.

**Table 2.** Reporter plasmids constructed and tested for prime editing project, and their corresponding transfection results.

Use	Name	Location of PE linker	Transfection
List of reporter plasmids for stable cell line generation	pLenti4	Non-truncated PE linker with TGA stop codon before mCherry	p4 expressed mCherry
	pLenti5	Truncated PE linker after 5 <sup>th</sup> Met; p5.1 has TGA stop codon; p5.2 has CTT insert	p5.1 did not express mCherry; p5.2 expressed mCherry
	pLenti6	No PE linker	P6 expressed mCherry (positive control)
	pLenti7	Truncated PE linker after 4 <sup>th</sup> Met; p7.1 has TGA stop codon; p7.2 has CTT insert	p7.2 did not express mCherry due to improper protein folding
	pLenti8	Truncated PE linker after 2 <sup>nd</sup> Met; p8.1 has TGA stop codon; p8.2 has CTT insert	p8.1 did not express mCherry; p8.2 expressed mCherry
	pLenti9	Truncated PE linker after 1 <sup>st</sup> Met; p9.1 has TGA stop codon; p9.2 has CTT insert	p9.1 expressed mCherry; p9.2 expressed mCherry

*Aim 1B. To generate a stable HEK293T cell line to express reporter plasmid.*

To achieve this stable cell line, the desired reporter plasmid construct will be incorporated into the mammalian genome following a standardized 3<sup>rd</sup> generation lentivirus production protocol for generating a stable cell line. Briefly, HEK293T producer cells will be transfected with four plasmids: two packaging plasmids, envelope plasmid, and the lentiviral transfer plasmid encoding the insert of interest. The virus will assemble and be harvested for direct infection of HEK293T cells to stably incorporate the transgene into the mammalian genome.

*Aim 1C. To test prime editing efficiency.*

The stable HEK293T cell line expressing the fluorescent reporter transgene will then be transfected with one plasmid containing the wild-type engineered Moloney murine leukemia virus (M-MLV) reverse transcriptase, tethered by a flexible linker to the C-

terminus of nCas9, and a second plasmid containing the pegRNA encoding the PBS and RT template with the edit. To validate the insertion or deletion of the stop codon *in vivo* by prime editing and determine the prime editing efficiency of the constructs, the mCherry fluorescence levels will be approximated and monitored with the flow cytometer and further confirmed with high-throughput sequencing.

### **4.3. Future directions**

By employing prime editors to guide UAA introduction, we can insert stop codons on specific endogenous proteins to permanently introduce biorthogonal, clickable handles onto proteins to be used to study its cellular signaling and localization patterns *in vivo*. These enhanced functionalities will allow for further enrichment by affinity capture and future proteomic analyses. For example, we can use this technology to study low copy number endogenous proteins or PTM isoforms that are otherwise difficult to be transiently expressed and analyze their connections to various types of diseases.

### **4.4. Materials and Methods**

#### *General*

Gels were imaged with the BioRad ChemiDoc imaging system. Absorbances were taken using the Thermo Fisher Scientific NanoDrop One Micro-UV/Vis Spectrophotometer. Cells were monitored for viability with EVOS XL Core imaging system from Thermo Fisher Scientific and imaged with Zeiss Axio Observer A1 X-Cite Series 120 Q combination microscope.

*Aim 1. To validate the insertion of an amber stop codon by expressing a fluorescent reporter in a stable HEK293T cell line.*

To generate a HEK293T cell line expressing a fluorescent reporter to monitor prime editing activity, a series of potential reporter plasmids were first constructed using Gibson assembly. To clone the plasmids of interest using Gibson assembly, the backbones and inserts were amplified with PCR. A 50  $\mu$ L reaction of 1X Q5 reaction buffer (NEB), 200  $\mu$ M dNTPs (NEB), 0.5  $\mu$ M forward primer (Genewiz), 0.5  $\mu$ M reverse primer (Genewiz), and 0.02 U/ $\mu$ L Q5 polymerase (NEB) was subjected to the thermal cycler with the following program: [98 °C for 30 s] + [98 °C for 15 s + annealing temperature for 30 s + 72 °C for (# k length + 20 s)] \* 30 + [72 °C for 2 min] + [4 °C hold]. The amplified product was purified by PCR purification using the QIAGEN QIAquick PCR purification kit according to the manufacturer's protocol, and the resulting DNA template concentration was quantified by UV spectroscopy at Abs<sub>260</sub>. DNA integrity was analyzed by agarose gel electrophoresis with a 1% agarose gel, alongside the GeneRuler 1 kb Plus DNA ladder (ThermoFisher Scientific) in 1X TAE buffer at 120 V for 45 minutes stained with 0.33  $\mu$ g/mL EtBr solution. Once the PCR products were purified and quantified, an assembly reaction comprising 1X Gibson assembly master mix (NEB), 0.03 pmol vector, and 0.06 pmol insert(s) was diluted to 10  $\mu$ L with nuclease-free water. The reaction was subjected to the thermal cycler with the following program: [50 °C for 15 min] + [4 °C hold]. After the assembly reaction, 1  $\mu$ L the product was transformed into NEB Turbo electrocompetent *E. coli* cells using a benchtop electroporator, and the transformed cells were incubated in 1 mL SOC media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 20 mM glucose) for recovery at 37 °C for 1 hour. After recovery, 100  $\mu$ L recovered cells were cultured on Luria broth agar plates containing 100  $\mu$ g/mL ampicillin and incubated at 37 °C for 12 – 18 hours. The

following day, several single colonies were inoculated in Luria broth supplemented with 100 µg/mL ampicillin and grown at 37 °C for 12 – 18 hours at 250 rpm. The overnight cultures were then isolated and purified using the GeneJET Plasmid Miniprep kit (ThermoFisher Scientific), and the resulting plasmid concentration was quantified by UV spectroscopy at Abs<sub>260</sub>. To confirm that the isolated plasmid was correct, restriction enzyme digests were performed following NEB's standardized protocols, and the samples were sent out for Sanger sequencing (Genewiz) to confirm the DNA sequence. Next, the transfection efficiency of each isolated plasmid was measured to determine which reporter construct would be best for the generation of the stable HEK293T cell line. HEK293T cells were treated in 100-mm culture dishes (Corning). Cells were transfected at approximately 70% confluency with 1 µg/µL PEI (Sigma Aldrich) and 5 µg of the purified plasmid of interest. The cell viability was measured every 24 hours post-transfection using the bright-field microscope, and fluorescence was visualized with the fluorescent microscope after 72 hours to visualize mCherry gene expression levels.

#### 4.5 References

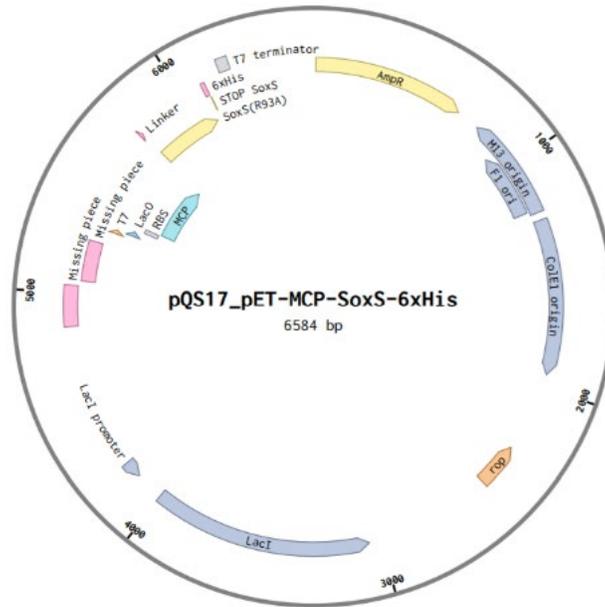
- [63] Chin, J. W., Expanding and reprogramming the genetic code of cells and animals. *Annual Review of Biochemistry* **2014**, *83* (1), 379-408.
- [64] Italia, J. S.; Addy, P. S.; Wrobel, C. J. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A., An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nature Chemical Biology* **2017**, *13* (4), 446-450.
- [65] Meineke, B.; Heimgärtner, J.; Eirich, J.; Landreh, M.; Elsässer, S. J., Site-specific incorporation of two ncAAs for two-color bioorthogonal labeling and crosslinking of proteins on live mammalian cells. *Cell Reports* **2020**, *31* (12), 107811.
- [66] Suzuki, T.; Asami, M.; Patel, S. G.; Luk, L. Y. P.; Tsai, Y.-H.; Perry, A. C. F., Switchable genome editing via genetic code expansion. *Scientific Reports* **2018**, *8* (1), 10051.

## Appendices

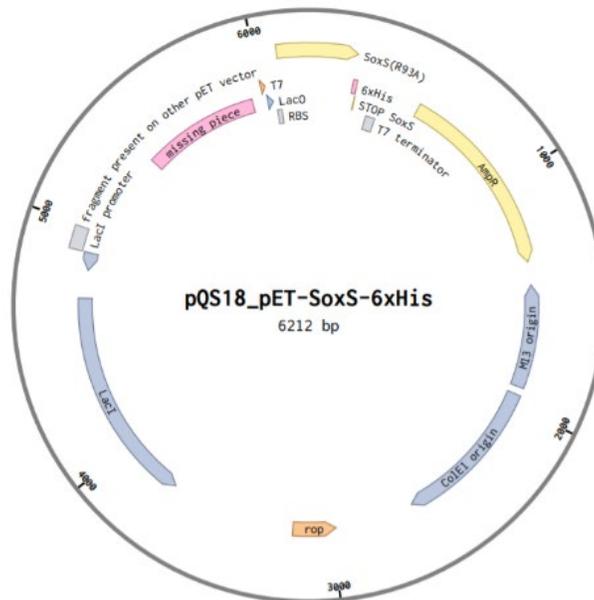
### Appendix 1. Supporting information for high-throughput particle display screening of RNA-protein interactions

#### Plasmid Maps

pQS17



pQS18



*Gene sequences*

MCP-SoxS-his

ATGGGGCCCGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTGCG  
CCCCAAGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGT  
AACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGG  
CGCCTGGCGTTCGTACTIONAAATATGGAACCTAACCATTCGAATTTTCGCCACGAATTCGACTGCGAGCTTA  
TTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAACCCGATTCCCTCAGCAATCGCAGCAAACCTCCGG  
CATCTACGGTGGCGGAGGTAGCATGTCCCATCAGAAAATTATTCAGGATCTTATCGCATGGATTGACGAG  
CATATTGACCAGCCGCTTAACATTGATGTAGTCGCAAAAAAATCAGGCTATTCAAAGTGGTACTTGCAAC  
GAATGTTCCGCACGGTGACGCATCAGACGCTTGGCGATTACATTCGCCAACGCCGCTGTTACTGGCCGC  
CGTTGAGTTGCGCACCACCGAGCGTCCGATTTTTGATATCGCAATGGACCTGGGTTATGTCTCGCAGCAG  
ACCTTCTCCCGCGTTTTTCGCGCGGCAGTTTGATCGCACTCCAGCGATTATCGCCACCGCCTGCATCACCA  
CCACCATCAC

MS2 coat protein

GGGS linker

SoxS transcriptional activator

6x histidine tag

SoxS-his

ATGTCCCATCAGAAAATTATTCAGGATCTTATCGCATGGATTGACGAGCATATTGACCAGCCGCTTAACAT  
TGATGTAGTCGCAAAAAAATCAGGCTATTCAAAGTGGTACTTGCAACGAATGTTCCGCACGGTGACGCAT  
CAGACGCTTGGCGATTACATTCGCCAACGCCGCTGTTACTGGCCGCCGTTGAGTTGCGCACCACCGAGC  
GTCCGATTTTTGATATCGCAATGGACCTGGGTTATGTCTCGCAGCAGACCTTCTCCCGCGTTTTTCGCGCGG  
CAGTTTGATCGCACTCCAGCGATTATCGCCACCGCCTGCATCACACCACCATCAC

SoxS transcriptional activator

6x histidine tag

CC\_gRNA\_tracrRNA\_MS2apt\_polytail

DNA sequence after primer extension (5' – 3'):

CCAAGTAATACGACTCACTATAGGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT

ATCAACTTGAAAAAGTGCGCACATGAGGATCACCCATGTGCGAAGCTTGGGCCCAAAAAAAAAAAAAAAAAA

AAAAAAAAAA

T7 promoter

GG extension

MS2 aptamer

poly A tail

RNA sequence after *in vitro* transcription (5' – 3'):

GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGC

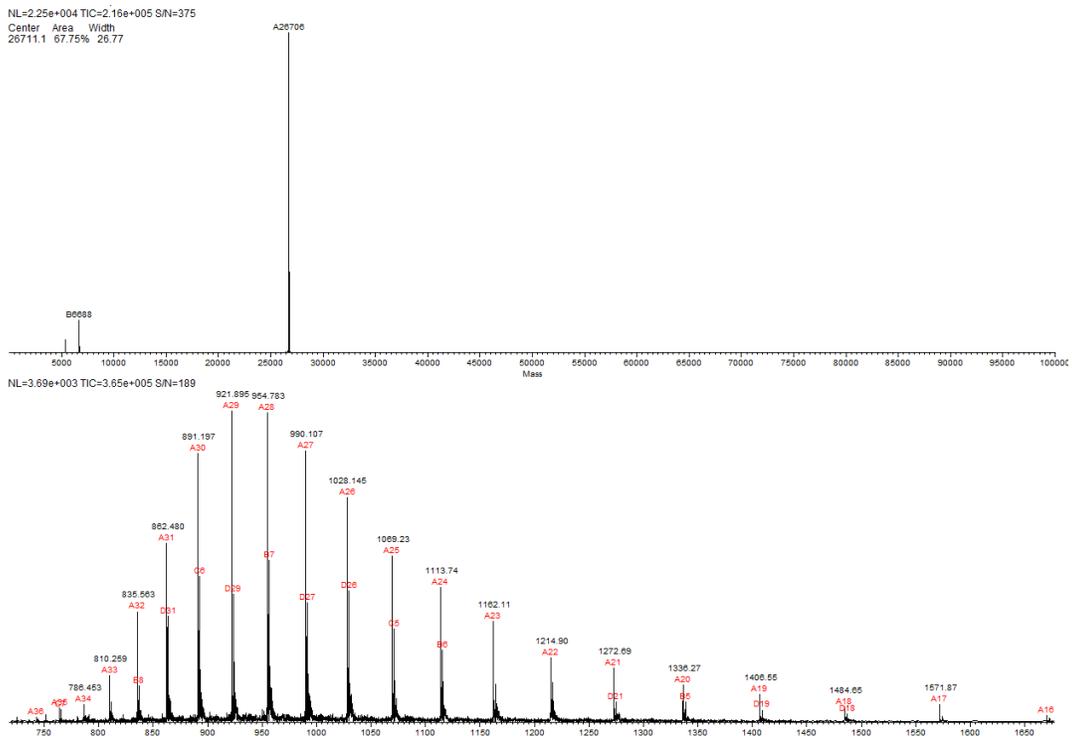
CAUGAGGAUCACCCAUGUGC GAAGCUUGGCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

MS2 aptamer

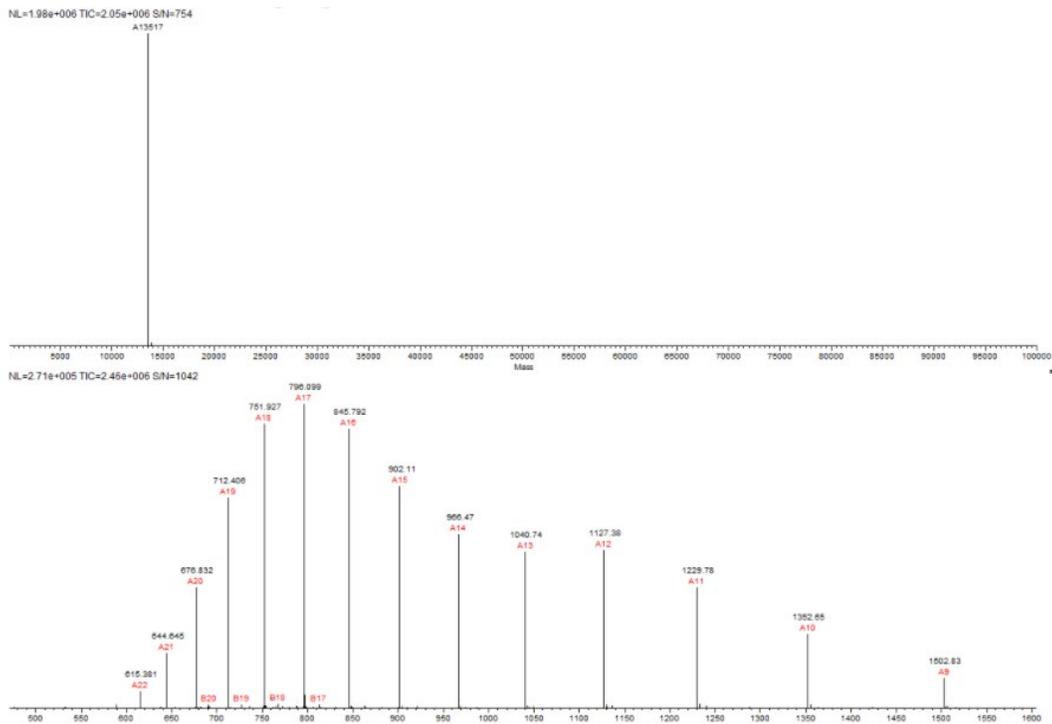
poly A tail

# LC/MS

## MCP-SoxS-his (Expected mass: 26708 Da; Experimental mass: 26706 Da)



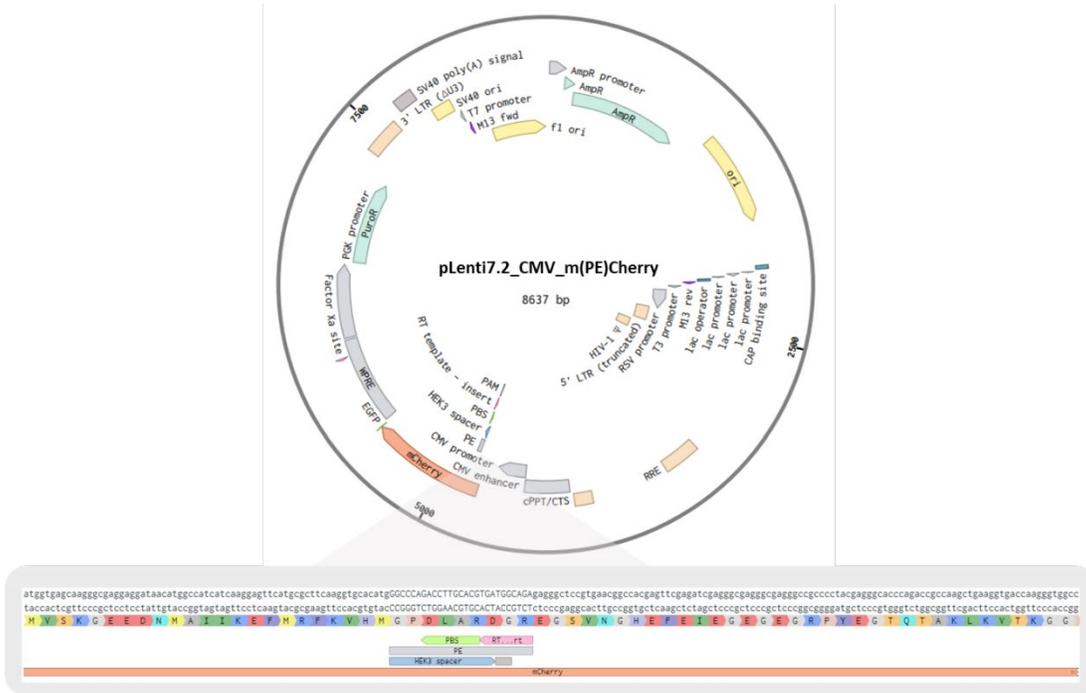
## SoxS-his (Expected mass: 13648 Da; Experimental mass: 13517 Da)



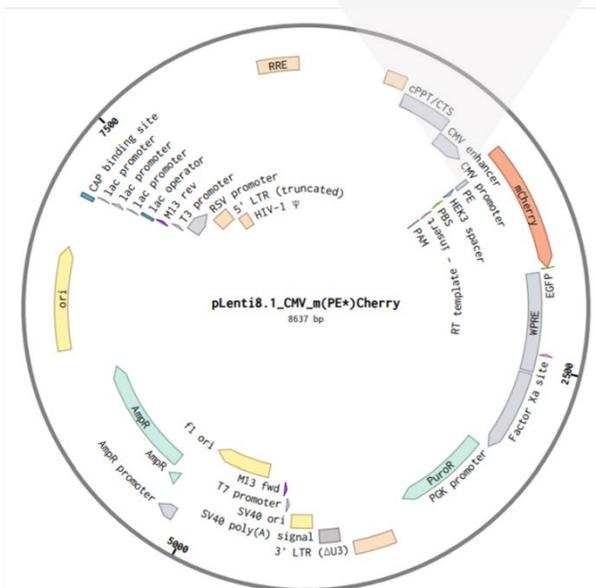




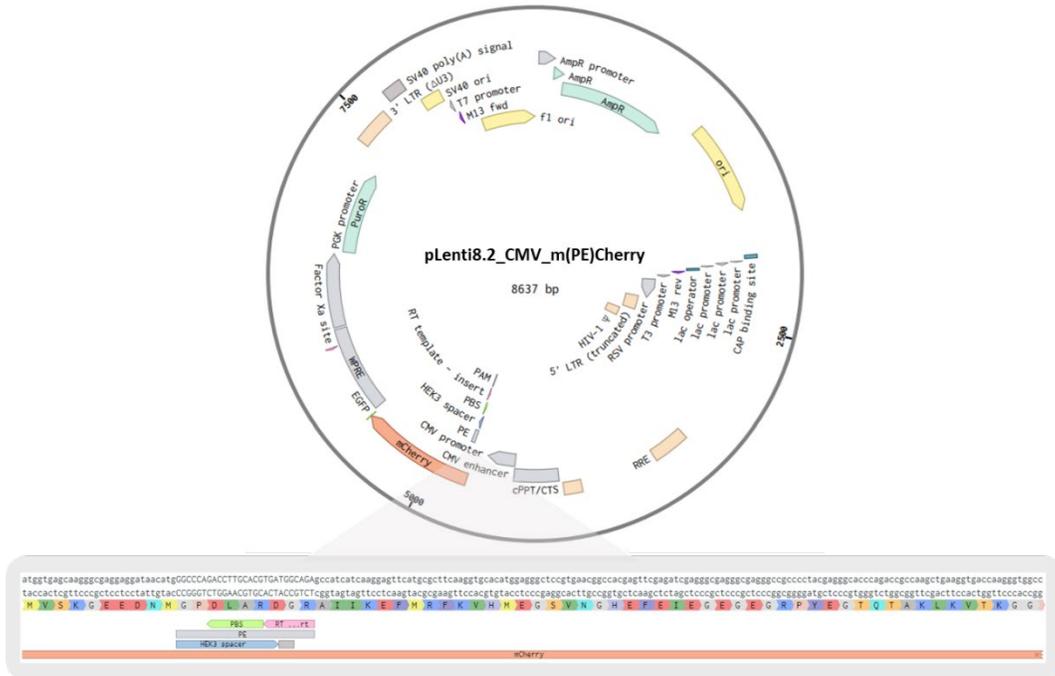
pLenti7.2\_CMV\_m(PE)Cherry



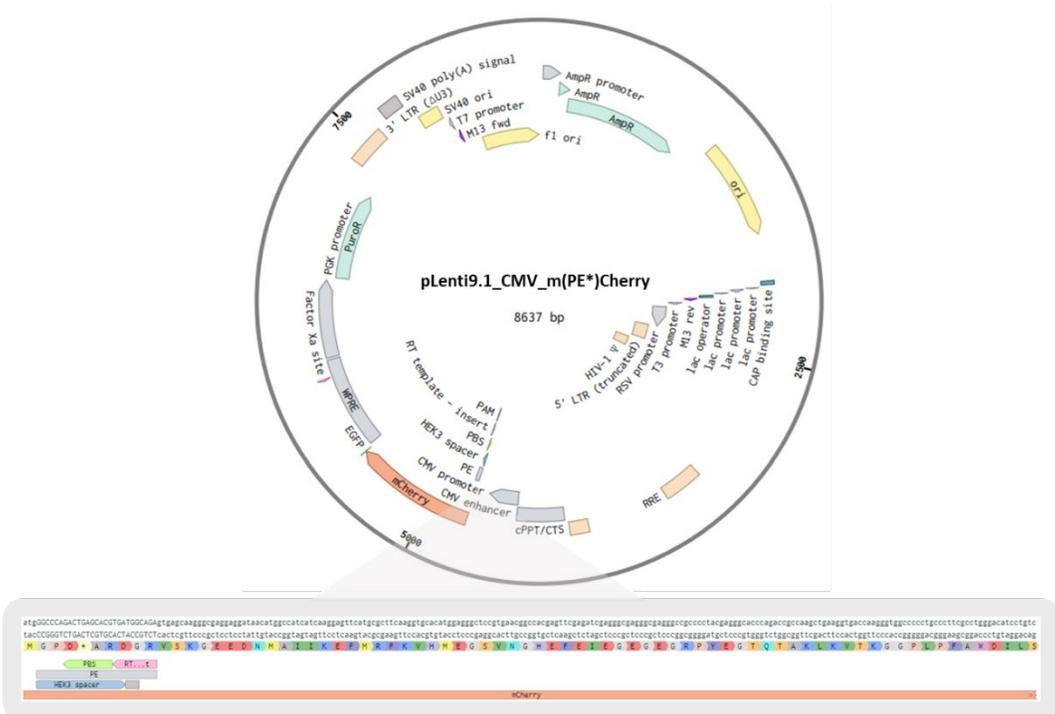
pLenti8.1\_CMV\_m(PE\*)Cherry



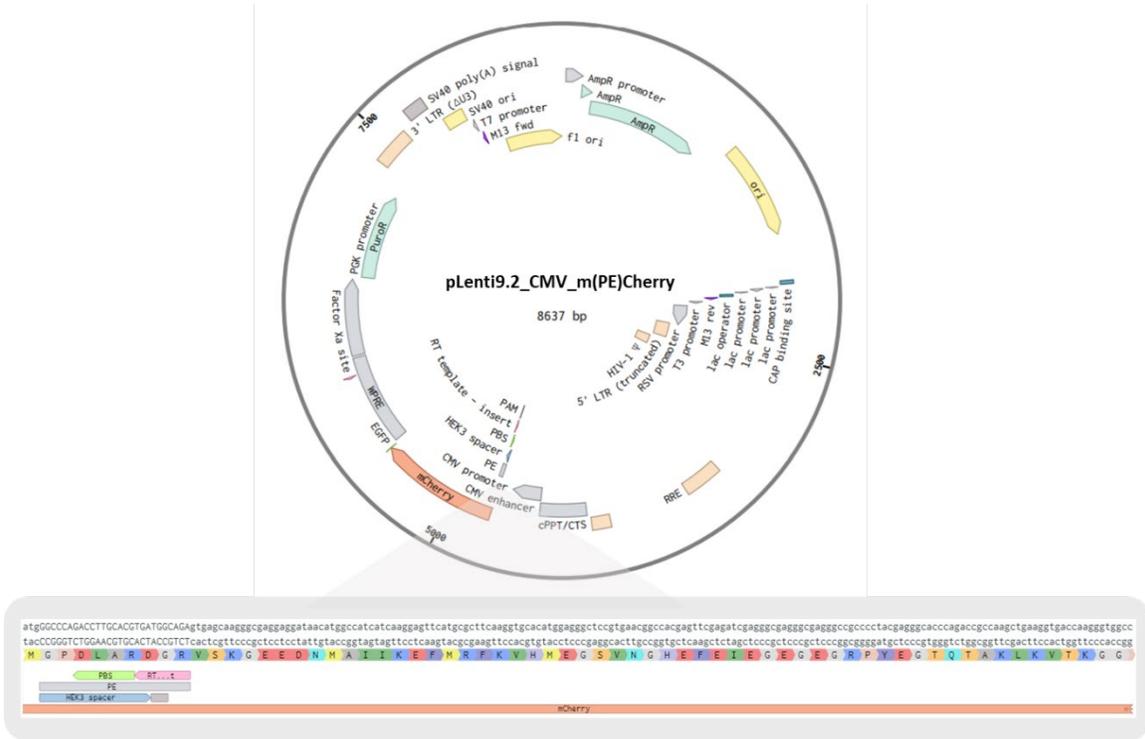
pLenti8.2\_CMV\_m(PE)Cherry



pLenti9.1\_CMV\_m(PE\*)Cherry



pLenti9.2\_CMV\_m(PE)Cherry



*Mammalian Cell Culture*

HEK293T cells were cultured in DMEM media. Cells were maintained at 37 °C with 5% CO<sub>2</sub>. Media and supplements were purchased from Fisher Scientific.