Using Unnatural Amino Acid Incorporation to Modify and Manipulate Adeno-Associated Virus

Sarah B. Erickson

A dissertation

submitted to the Faculty of

the department of Chemistry

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

July 2020

Using Unnatural Amino Acid Incorporation to Modify and Manipulate Adeno-Associated Virus

Sarah B. Erickson

Advisor: Abhishek Chatterjee, Ph.D.

Abstract

Adeno-Associated Virus (AAV) has been developed into a powerful therapeutic tool- in the last ten years it has acted as a gene-delivery vehicle in several approved therapeutics and many more therapeutics on trial. Despite extensive research, gaps in our understanding of AAV's infectious cycle still exist, and further development is needed for the creation of improved gene therapy vectors. Technology to incorporate Unnatural Amino Acids (UAAs) into the AAV capsid has recently been developed, and could aid in both furthering our understanding of AAV's biology and in the therapeutic advancement of AAV. In this work, we demonstrate how the functionalization of the AAV capsid using UAA incorporation can advance our control over the AAV capsid and aid in probing and manipulating AAV biology.

We describe our use UAA incorporation to place a bio-orthogonal reactive handle into AAV's capsid followed by functionalization with a targeting moiety and demonstrate the unprecedented amount of control that UAA incorporation provides in the creation of a functional virus conjugate. We are able to control both the precise placement and the stoichiometry of the targeting moiety on the AAV capsid, providing a platform that, for the first time, can undergo rigorous optimization analogous to that which medicinal chemists put small molecules through. We also describe the creation of a new platform to

site-specifically modify the AAV capsid using cysteine incorporation, a technique that retains the ability to site-specifically modify the capsid as UAA incorporation does, but does not require the excess machinery that UAA incorporation requires.

Next we discuss the incorporation of a photocaging amino acid, NBK, into the AAV capsid. Using NBK, we were able to effectively block AAV's primary binding interaction with Heparan Sulfate Proteoglycan (HSPG) and control the timing of AAV infection using light to chemically remove the photo-protecting group. While photocaging the HSPG interaction is only a proof of concept, it demonstrates the remarkable amount of control that UAA incorporation affords, and lends insight to what could be accomplished using the functionalities that can be placed on the AAV capsid with UAAs.

Acknowledgements

I would be remiss if I did not first thank Abhishek, who provided guidance and encouragement throughout my PhD. Thank you for your patience and assistance in problem solving. I would also like to thank my committee, Professor Gao and Professor Weerapana for advice, support, and encouragement. Dale, Lynne, Ian, Bob, and Juan- thank you for your excellent work as well as your support.

The entire Chatterjee lab has been a huge support to me- both predecessors and those who have joined the lab after me. I would like to particularly thank Yunan and Rachel for their guidance and helping me troubleshoot whenever I needed it. I would also like to thank Team AAV- Rachel, Xiaofu, Sam, Lily, Rachet, and Quan- I admire the hard work and dedication you all exhibit and am excited to see what you guys can accomplish. I am also grateful for the help I received from the undergrads I worked with, Xiaofu, Zai, and Stephanie. I frequently sought advice from Dan Bak and Jennifer Peeler from the Weerapana lab, and Samantha Cambray from the Gao lab. I'm grateful for their support and guidance!

On a personal note, I am deeply grateful to my parents, who taught me the importance of an education and demonstrated hard work through my entire life. Their support has been imperative to working through this program. Finally I would like to thank both Trevor and Parker. I am grateful for Trevor's support and willingness to go through a Ph.D. program at the same time as me- as well as start down the road of having a family during this already crazy time. And Parker- you have no idea how much joy you have brought us. Just keep being perfect.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1 Tables

Chapter 2 Tables

Chapter 3 Tables

Chapter 6 Tables

LIST OF FIGURES

Chapter 1 Figures

Chapter 2 Figures

Chapter 3 Figures

Chapter 4 Figures

Chapter 5 Figures

Chapter 6 Figures

LIST OF ABBREVIATIONS

Standard one-letter and three-letter abbreviations are used for the 20 natural amino acids.

Abbreviation	Full name
5AzW	5-Azido-Tryptophan
5HTP	5-Hydroxy-Tryptophan
5PrW	5-Propargyl-Tryptophan
AAP	Adeno-Associated Virus Assmebly Activating Protein
aaRS	Amionoacyl-tRNA synthetase
AAV	Adeno-associated Virus
AAV#	Adeno-associated Virus serotype #
AcK	Acetyl Lysine
ATM	Altered Translation Machinery
ATP	Adenosine Triphosphate
AzK	Azido-lysine
BCNK	Bicyclo [6.1.0] nonyne-Lysine
Cap	AAV *Cap* gene encoding the capsid proteins
CMV	cytomegalovirus intermediate-early promoter
CRACR	Chemoselective Rapid Azo-Coupling Reaction
CuAAC	Copper-catalyzed Azide-Alkyne Cycloadditions
DBCO	Dibenzylcyclooctyne
EGFP	Enhanced Green Fluorsescent Protein
FACS	Fluorescence Activated Cell Sorting
FGE	Formylglycine-generating enzyme
FKBP	FK-binding protein
FRET	Forster Resonance Energy Transfer
GC	Genome Copies (viral titer)
GCE	Genetic Code Expansion
GFP	Green Fluorescent Protein
Grb2	Growth factor receptor bound protein
HEK293T	A Human cell line derived from embryonic kidney cells
hSert	Human Serotonin Transporter
HSPG	Heparan Sulfate Proteglycan

HSV	Herpes Simplex Virus
IEDDA	Inverse electron-demand Diels-Alder cycloadditions
IMAPP	In-situ cleaveage and MS-label transfer Afer Protein Photocrosslinking
ITR	Inverted Terminal Repeats
MbPylRS	*M. barkeri* pyrrolysyl-tRNA synthetase
MOI	Multiplicity of Infection
OMeY	O-Methyl-Tyrsoine
ONBK	N^{ε}-cyclopentyloxycarbonyl-L-Lysine
ORF	Open Reading Frame
pAcF	p-Acetyl-phenylalanine
pAzF	p-Azido-phenylalanine
pBpA	p-benzoylphenylalanine
PCK	Photocaged Lysine
PEG	Polyethylene Glycol
PEI	Polyethyleneimine
qPCR	Quantitaive Polymerase Chain Reaction
Rep	AAV *Rep* gene encoding the replication proteins
SANH	succinimidyl 4-hydrazinonicotinate acetone hydrazine
ScoK	Strained Cyclooctyne-Lysine
sfGFP	Superfolder Green Fluorescent Protein
SK-BR-3	A human breast cancer cell line
SK-OV-3	A human ovarian cancer cell line
SPAAC	Strain-Promoted Azide-Alkyne Cycloaddition
tREX	tRNA Expansion
tRNA	Transfer RNA
UAA	Unnatural Amino Acid
UAA-AAV	AAV containing one or more unnatural amino acid in its capsid

INTRODUCTION

Adeno-Associated Virus has been developed into a powerful therapeutic tool- in the last ten years it has acted as a gene-delivery vehicle in several approved therapeutics and many more therapeutics on trial. Despite extensive research, gaps in our understanding of AAV's infectious cycle still exist, and further development is needed for the creation improved gene therapy vectors. Technology to incorporate Unnatural Amino Acids (UAAs) into the AAV capsid has recently been developed, and could aid in both furthering our understanding of AAV's biology and in the therapeutic advancement of AAV. In this work, we demonstrate how the functionalization of the AAV capsid using UAA incorporation can advance our control over the AAV capsid and aid in probing and manipulating AAV biology.

We describe our use UAA incorporation to place a bio-orthogonal reactive handle into AAV's capsid followed by functionalization with a targeting moiety and demonstrate the unprecedented amount of control that UAA incorporation provides in the creation of a functional virus conjugate. We are able to control both the precise placement and the stoichiometry of the targeting moiety on the AAV capsid, providing a platform that, for the first time, can undergo rigorous optimization analogous to that which medicinal chemists put small molecules through. We also describe the creation of a new platform to site-specifically modify the AAV capsid

using cysteine incorporation, a technique that retains the ability to site-specifically modify the capsid as UAA incorporation does, but does not need the excess machinery that UAA incorporation requires.

Next we discuss the incorporation of a photocaging amino acid, NBK, into the AAV capsid. Using NBK, we were able to effectively block AAV's primary binding interaction with HSPG and control the timing of AAV infection using light to chemically remove the photo-protecting group.

1.0 INTRODUCTION

My research has primarily been focused on the incorporation of Unnatural Amino Acids (UAAs) into the Adeno-Associated Virus (AAV) capsid to alter its activity. Functionalities provided by UAA incorporation, which will be further discussed, enable the creation of AAV vectors that that can be used to develop AAV both as a therapeutic tool and to further the understanding of AAV's infection cycle. A large portion of this thesis discuses the optimization of a functionalized, retargeted AAV capsid. It also describes the development of cysteine incorporation into the capsid to be used as a new functionalization strategy, and several other UAAs that could be used as tools to modify AAV's behavior.

1.1 ADENO-ASSOCIATED VIRUS

Adeno-Associated Virus (AAV) is a small, non-pathogenic member of the parvovirus family. Since it was discovered as a contaminant in Adenovirus purifications, AAV has received considerable attention for its potential usefulness as a gene-therapy vector. This thesis provides an overview of what is currently known

about AAV biology, as well as how the AAV capsid has been engineered for gene therapy purposes.

1.1.1 Adeno-Associated Virus Genome and Capsid Structure

The AAV genome is a roughly 4.7kb single stranded DNA genome (Figure 1). It consists of three open reading frames (ORFs) that code for functional proteins, enclosed in Inverted Terminal Repeats (ITRs, Figure 1A)[1,2]. The ITR's are T-shaped, base-paired hairpin structures that contain *cis*-elements required for replication and packaging. Two of the AAV ORFs code for the Rep proteins, which are non-structural proteins required for the replication of AAV. The p5 promoter initiates the translation of the genes *rep78* and *rep68.* Those transcripts are then translated into proteins Rep 78 and Rep 68, which have site-specific, single stranded endonuclease activity, DNA helicase activity, and ATPase activity[3,4]. Rep 68 differs from Rep 78 in that it is a spliced variant of Rep 78. The p19 promoter initiates the translation of the *rep52* and *rep40* genes, which are then translated into Rep 52 and Rep 40. Of these two proteins, Rep 40 is spliced. Rep 52 and Rep40 contain helicase activity that is essential for packaging DNA into formed capsids[5,6]. The *Cap* ORF encodes four proteins, all initiated from the p40 promoter. VP1, VP2, and VP3 are structural proteins that make up the viral capsid, and the Assembly Activating Protein (AAP) aids in capsid assembly but is not found in the assembled capsid[7–9]. VP1 is the largest of the three structural proteins, and arises from an un-spliced mRNA transcript. VP2, VP3, and AAP all arise from a spliced mRNA transcript. In addition to being alternatively spliced, the VP proteins and AAP also use alternative

start codons. VP1 uses ATG, VP2 uses AUG, VP3 uses ATG, and AAP uses a nonconventional CTG start codon that is found in an alternate reading frame[8,10]. VP1, VP2, and VP3 are present in the capsid in a roughly 1:1:10 ratio, with VP3 being the main structural protein (Figure 1C). VP1 and VP2 both have extended N-termini to VP3. VP1 contains a Phospholipase A_2 domain and three nuclear localization signals that are essential for infectivity[11–16]. VP2 contains one nuclear localization signal, but is not strictly necessary for viral infectivity (Figure 1B). Although the N-termini of VP1 and VP2 could not be visualized in the crystal structure, cryo-electron microscopy and image reconstruction data suggest that these tails are located inside the AAV capsid, through the five-fold pore[17–19].

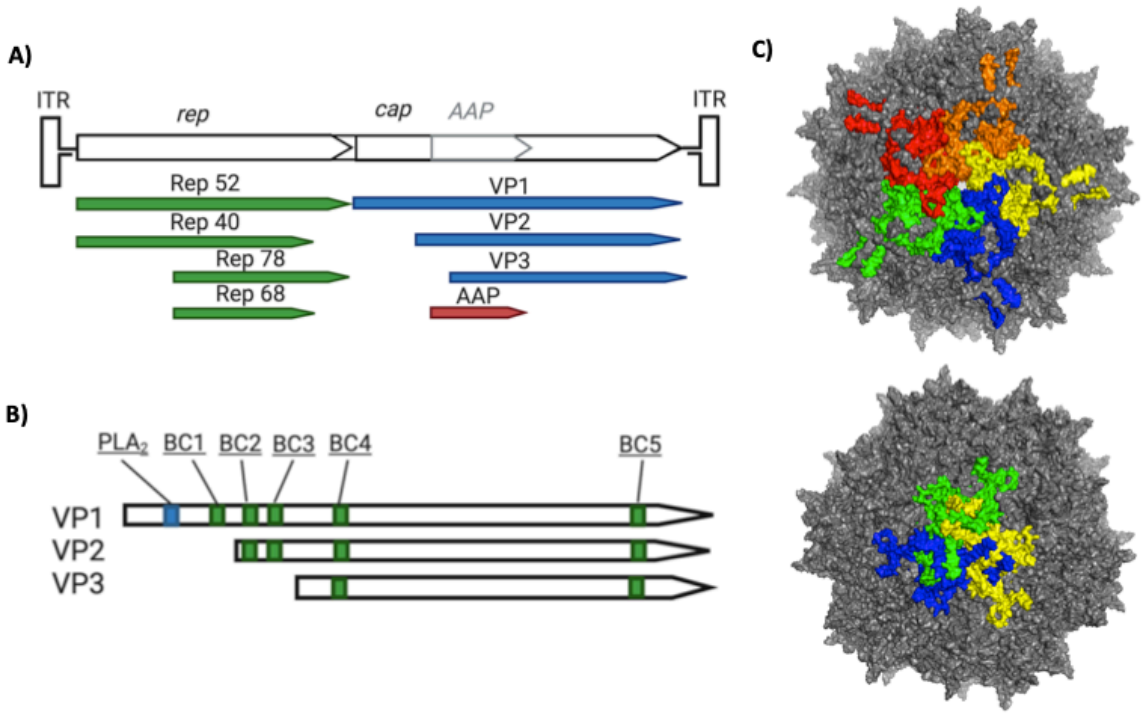


Figure 1.1. The AAV genome and capsid structure. A) The AAV genome consists of three orf's enclosed in ITRs. The *rep* orf encodes for four essential replication proteins; Rep 52, Rep 40, Rep 78, and Rep 68. The *cap* orf encodes for three structural proteins, VP1, VP2, and VP3, and the Assembly Activating Protein (AAP). B) VP1, VP2, and VP3 all share the VP3 C-terminal region. The extended N-terminus of VP1 contains a PLA_2 domain as well as three BC regions. The extended VP2 N-terminus contains two BC regions. C) The AAV capsid with either the 5-fold axis or the 3-fold axis of symmetry highlighted. Each color is one VP protein.

1.1.2 The AAV Infection Cycle

While AAV is a relatively simple virus, its infection cycle is complex and many details of viral infection remain poorly understood. Additionally, the details of infection differ from one serotype of AAV to another. As the large majority of studies have been performed on AAV2, this work outlines what is understood about the AAV2 infection cycle (Figure 2). Infection begins with AAV's primary binding interaction with Heparan Sulfate Proteoglycan (HSPG). HSPG is a long, negatively charged proteoglycan on the cell surface that interacts with five residues on the AAV capsid: R585, R588, R484, R487, and K532. Mutational studies have revealed that R585 and R588 are essential for infectivity, and R484, R487, and K532 contribute to the HSPG interaction but are not essential for infectivity[20,21]. Interaction with a secondary receptor is also required. Several co-receptors that have been implicated to be involved in AAV2 endocytosis include FGFR1, the $\alpha_V\beta_5$ integrin, HGFR, and integrins 37/67 kDa LamR[22–25]; however, AAVR has recently been discovered as the secondary receptor for several AAV serotypes[26].

After primary and secondary binding, the AAV is endocytosed. Many studies suggest that AAV is endocytosed through clathrin coated pits; however one study suggest that successful infection occurs through the Clathrin-Independent Carriers/GPI Enriched (CLIC/GEEC) pathway and that clathrin coated pits are not essential for AAV infectivity[27–31]. As AAV travels through the endosome, a conformational change occurs in the capsid structure that allows the N-terminal regions of VP1 and VP2 to come out of the capsid, exposing the PLA_2 domain on VP1[7,13]. While this conformational change is thought to be due to the acidification of

the endosome destabilizing the capsid, one study suggested that the acidification alone was not enough[7]. Another study indicated that endosomal cathepsins B and L are necessary for infection and implied that their role could be to aid in endosomal escape. The details implicating endosomal cathepsins B and L in endosomal escape are not yet clear[32].

While the steps following endosomal escape remain unclear, reports suggest that the AAV2 capsid travels through the cytoplasm with exposed VP1 and VP2 N-terminals[7]. In addition to the PLA_2 domain, The VP1 and VP2 N-terminals bear nuclear localization signals. Nuclear localization signals are concentrated in basic residues, and are therefore referred to as Basic Clusters (BCs). Three BC regions have been found to be important in directing the AAV capsid into the nucleus (Figure 1B)[7,15,33,34]. BC1 is found exclusively in the VP1 N-terminal region while BC2 and BC3 are found in both VP1 and VP2's N-terminal regions.

Once AAV has entered the nucleus, it is rapidly shuttled into the nucleosome, where the fate of the infection depends on the presence or absence of helper virus, such as Adenovirus or herpes simplex virus. In the absence of helper virus, AAV will integrate into the host chromosome on human chromosome 19[35–37]. In the absence of both helper virus genes and *Rep* genes, the AAV genome will be stored as an episome[2,35–38]. In the presence of helper virus, AAV will continue into the lytic cycle. In the nucleosome, the single-stranded genome is unpackaged and *rep* genes are transcribed and translated using the genome's ITR's as self-primers in a process that relies on cellular machinery and helper proteins[39]. The replication of the AAV

genome then takes place through a mechanism known as the "rolling hairpin mechanism", which requires extensive use of the Rep proteins.

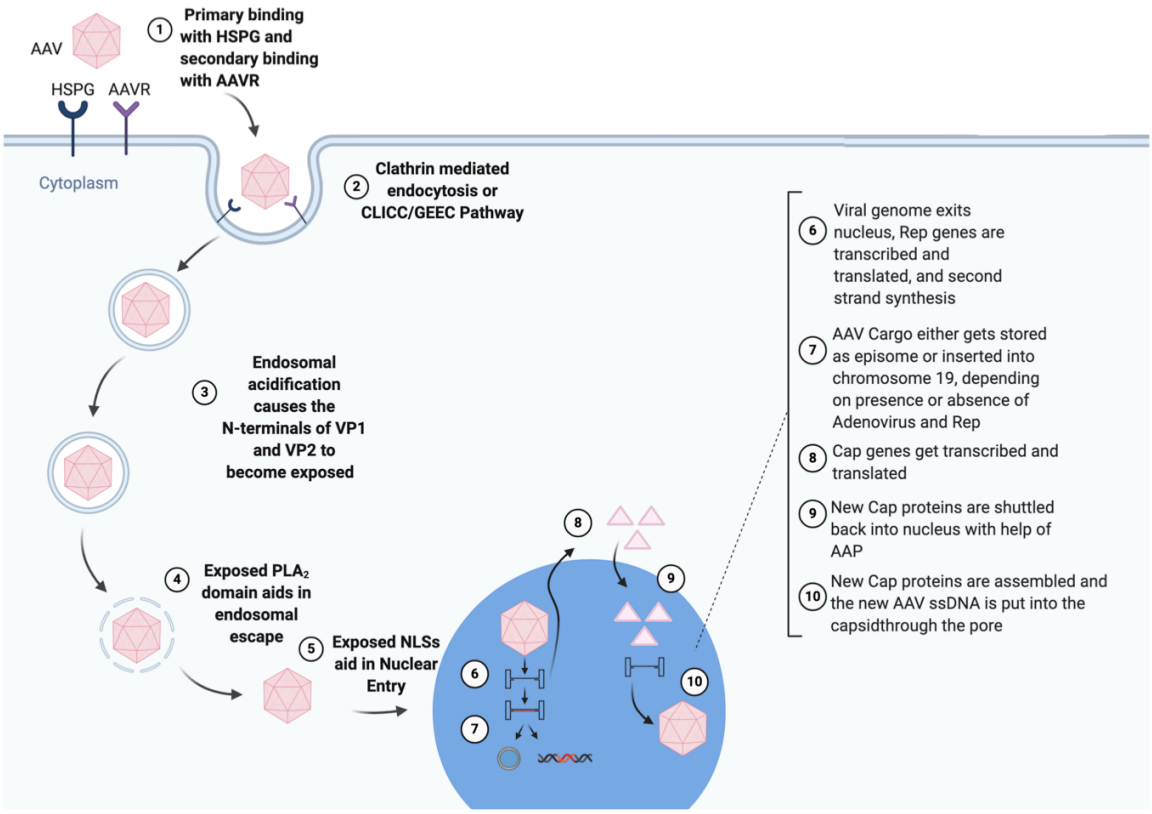


Figure 1.2. The AAV infection cycle

Once the AAV genome is replicated, Cap proteins are synthesized in the cytoplasm and transported back into the nucleus[40]. Two BC regions have been identified as important in shuttling free VP1, VP2, and VP3 proteins into the nucleus prior to capsid assembly, present in all three VP proteins. BC4 is found in the VP3 N-terminal region and BC5 is found in the C-terminal region of all three proteins (Figure 1B)[34]. In addition to these BC regions, the Assembly Activating Protein (AAP) shuttles the VP proteins into the nucleus and aids in assembly[8,9,41]. In the absence of AAP, the VP proteins rapidly undergo ubiquitination and degradation[42]. It has also been proposed that AAP acts as a scaffolding protein during capsid

formation[9]. Once the capsids are assembled and the AAV genome is replicated, the genome is packaged into the AAV capsid inside of the nucleus. Packaging occurs once the single stranded genome is fully replicated. The single stranded genome is translocated into the capsid in a 3' to 5' direction by the helicase/ATPase activity of the Rep52 or Rep40 proteins through the pore at the 5-fold axis of the virus capsid[43]. This process is dependent on proteins expressed from the helper virus genes.

1.2 RETARGETING ADENO-ASSOCIATED VIRUS FOR USE IN GENE THERAPY

1.2.1 AAV as a Gene Therapy Vector

AAV has emerged as a highly attractive gene therapy vehicle due to its replication deficient nature, the low immunogenic response it elicits, its lack of known pathogenicity, its ability to infect both dividing and non-dividing cells, and its ability to provide stable, long-term expression of the delivered gene without integrating into the host chromosome[44,45]. AAV also has a variety of different serotypes with specific tropisms that can be used to target different cell types (Chart 1).

To date, there have been three approved drug therapies that use AAV to deliver a functional copy of a gene to replace a non-functional copy. Alipogene tiparvovec (marketed as Glybera) was the first gene therapy treatment to be approved in Europe in 2012. Glybera treats Lipoporotein Lipase (LPL) deficiency, a

Table 1.1. A list of proteins encoded by the AAV genome and their roles.

Protein	ORF	Role(s)	Domain(s)
Rep78	*rep*	Genome replication, packaging, chromosomal integration	Endonuclease, helicase
Rep 68	*rep*	Genome replication, packaging, chromosomal integration	Endonuclease, helicase
Rep52	*rep*	Genome replication, packaging	Helicase
Rep40	*rep*	Genome replication, packaging	Helicase
VP1	*cap*	Structural, infection	Phospholipase A2, Nuclear localization sequence, receptor binding
VP2	*cap*	Structural, infection	Nuclear localization sequence, receptor binding
VP3	*cap*	Major Structural, infection	Receptor binding
AAP	*aap*	Assembly	Nuclear Localization signal

rare inherited disorder that is caused by a mutation in the LPL gene and can lead to severe pancreatitis[46,47]. Glybera used AAV1 to deliver a functional copy of the LPL gene to muscle cells. Unfortunately, due to the high cost and low demand for the treatment, Glybera was removed from the market in 2017. Vortigene Neparvovec (Luxturna) was approved in the US in 2017 and treats Leber congenital amaurosis, a rare form of early onset blindness. Luxturna uses a subretinal injection of AAV2 to deliver a functional copy of the *RPE65* gene, which encodes for all-trans retinyl ester isomerase[48]. The third AAV-based gene therapy drug to be approved is Onasemnogene abeparvovec (Zolgensma), which is used to treat spinal muscular atrophy in children under the age of 2. It delivers a functional copy of the Survival Motor Neuron 1 (SMN1) gene. The SMN protein is critical for the function of neurons, and a mutated/deleted copy of the SMN1 gene can lead to serious and fatal

muscle weakness. AAV9 was used for this treatment, as it can cross the blood brain barrier and transduce the Central Nervous System (CNS)[49,50]. In addition to these drugs that have been FDA approved, there are additional AAV-based treatments in clinical trials for a variety of diseases[45].

Despite AAV's success as a gene therapy vehicle, several limitations remain. Although there is only a small immune response to rAAV, overcoming this response is imperative for any treatments that would require multiple doses. Other challenges include expanding AAV's small cargo capacity, and overcoming the need to use AAV's naturally occurring serotypes as they do not address many of the diseases that we would like to be able to cure and have broad tropisms leading to off-target effects[44]. While there are efforts to overcome all of these limitations, this work focuses on efforts made to create custom viral vectors with refined specificity that deliver cargo to a destination of interest, rather than to its natural host. Four main methods have been used to achieve this: loop insertion, N-terminal extension, directed evolution or capsid shuffling, and unnatural amino acid (UAA) incorporation to attach targeting ligands or otherwise modify capsid specificity (Figure 3).

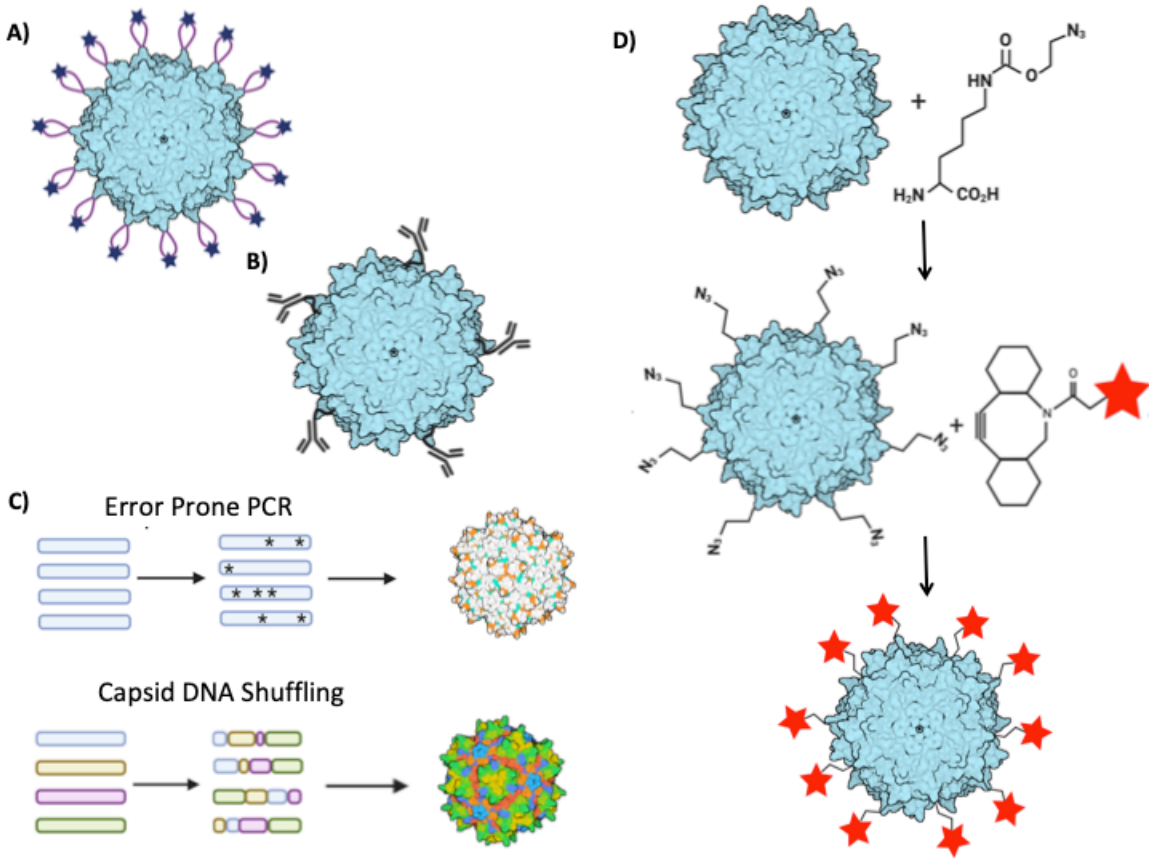


Figure 1.3. The Commonly used methods of modifying the AAV capsid for altered tropism. A) Peptide loop insertion into allowed regions of the VP3 region introducing a new functionality B) Genetic fusion of a new functionality to the VP2 N-terminus. C) Directed evolution or capsid shuffling to alter capsid tropism. D) Unnatural amino acid incorporation into the AAV capsid followed by chemical ligation of a targeting moiety to the unnatural amino acid.

1.2.2 Engineered AAV vectors

The first attempt to re-direct AAV tropism was done using a natural AAV capsid and a bispecific antibody in which one arm recognized the AAV capsid and the other arm recognized the $\alpha_{IIb}\beta_3$ integrin receptor that is over-expressed in megakaryocetes and platelet specific cells[51]. Since then, a variety of different methods have been used to modify the AAV capsid. The most straightforward way has been to modify the naturally occurring arginine and lysine residues on the capsid. Both biotin and streptavidin have been attached to exposed primary amines

on the capsid surface using NHS esters to create universal retargeting platforms[52,53]. Additionally, the exposed arginine residues have been modified with a α-dicarbonyl compound, methyl glyoxal, resulting in charge neutral hydroimidazolones. This masked the arginine residues that interact with HSPG, resulting in AAV that was no longer targeted to HSPG. This capsid was found to retain the ability to infect neurons and was redirected from liver to skeletal and cardiac muscle[54]. While these studies achieved some degree of success, the modification of surface exposed residues on the AAV capsid lacks the ability to control for stoichiometry or placement of the modification and therefore has not been a widely used method.

Perhaps the most commonly used method to create targeted AAV vectors has been to incorporate a peptide loop into the AAV capsid (Figure 3A)[55,56,65,66,57–64]. This method genetically incorporates short peptide motifs, usually around 15 amino acids long into different positions in the capsid. However as the AAV capsid is a very intricate structure assembled from 60 different subunits, these peptide insertions usually cause poorly formed capsids and low titers. While many positions on the capsid have been tested for the ability to accept a peptide insertion, only two spots have been found to accept peptide loops; the "spike" region around T454, and the positions near the heparan binding amino acids, R585-R588[55,62]. Additionally, the success of this method, even at the accepted sites, seems to depend on the amino acid sequence that is being incorporated[60]. While the peptide loops that have been incorporated often include a targeting moiety, such as the RGD motif or the NGR tumor targeting motifs, other motifs have also been incorporated that allow for further modification of the capsid[55,56]. One example of this is the incorporation of a

biotin acceptor peptide (BAP) that can be later modified with BirA, a biotin ligase that catalyzes the transfer of biotin to the epsilon amino group of a specified lysine residue[62,63,67]. One study used the surface exposed biotin to retarget to rat glioma BT4c cells that were engineered to express scavidin, an artificial biotin acceptor, while another study used a ketone isotere of biotin followed by further modification with hydrazides to attach fluorescent probes or an RGD targeting motif. Another interesting technique used was accomplished by incorporating an amino acid sequence containing a cysteine that was modified using Formylglycine Generating Enzyme (FGE) to produce an exposed aldehyde tag on the capsid surface. A succinimidyl-4-hydrazinonicotinate acetone hydrazone (SANH) crosslinker was that used to create a free hydrazide that allowed conjugation via an aldehyde tag. This method was used to attach gold nanoparticles, a cRGD motif, and an antibody onto the capsid surface[65].

Another commonly used method to modify the AAV capsid is to genetically fuse a sequence to the N-terminal region of the VP2 protein (Figure 3B). Any fusion to either the VP1 or VP3 N-terminus results in no virus production; the VP1 amino terminus has the essential PLA_2 domain and the NLS sequences and fusions interfere with these interactions, and modifications to the VP3 protein, which is the major structural protein, interferes with capsid assembly[68,69]. While fusions to the VP2 N-terminus result in functional virus, it is thought that the VP1 and VP2 N-termini are folded into the capsid structure through the pore, so addition of sequences or proteins onto the VP2 N-termini could greatly disrupt capsid assembly [17–19]. The earliest study that attempted to modify the N-termini of the VP proteins

genetically fused the variable region of a single-chain antibody against human CD34 molecules to the N-termini of VP1 or VP2. However, because viral titers were so poor, the study resorted to expressing the wild type VP proteins in addition to the modified proteins in order to get usable titers[70]. Since this study, others have more successfully created VP2 fusions with DARPins or affibodies[71–73], and one study fused the human FK-binding protein (FKBP) to the VP2 N-terminus to create a switch for controlling viral infectivity[74].

Directed evolution relies on the creation of a library followed by a stringent selection scheme to evolve a desired functionality (Figure 3C). Directed evolution has been performed on the AAV capsid to both evolve new binding specificities as well as engineer capsids that evade neutralizing antibodies[75–78]. These studies used error-prone PCR to create libraries of different capsids. Capsid shuffling is a similar method, which involves shuffling the DNA from multiple serotypes of AAV to create large chimeric *cap* libraries. Capsid shuffling has been used to create AAV capsids with modified receptor binding or the ability to escape the immune response[79–83]. AAV-DJ was derived from shuffling the DNA of eight different AAV serotypes. It is made of a mixture of AAV2, AAV8, and AAV9, but the final product is composed of mostly AAV2. AAV-DJ displays high transduction efficiency as well as high infectivity across a broad range of cells[80]. In addition to shuffling current AAV capsids, several studies have also been performed that predict ancestral AAV capsids and identify and modify their tropisms[84,85].

The most recent technique developed to create designer AAV capsids is the incorporation of Unnatural Amino Acids (UAAs) using a technology called genetic

code expansion (GCE)(Figure3D). GCE uses an engineered aminoacyl-tRNA synthetase/tRNA pair that is orthogonal to the host system to encode an unnatural amino acid in response to a stop codon[86]. Production of UAA-containing AAV requires that cells be supplied with both necessary UAA machinery as well as AAV production machinery. Required UAA machinery includes the tRNA/synthetase pair, and AAV machinery includes the helper plasmid, the viral genome enclosed in the ITR's, and the viral *Rep* and *Cap* genes.

Lysines with azide functionalities have been incorporated into the AAV capsid into a variety of sites followed by conjugation with copper free click chemistry to fluorophores, retargeting moieties, PEG groups, and 'lipid based cloaks'[31,87–89]. Additionally, a photocaged lysine has been incorporated into the AAV capsid at the sites involved in HSPG binding to create a photo-removable protecting group hindering viral infection[90]. While UAA incorporation into the AAV capsid requires the delivery of additional machinery to the cell, UAA incorporation has several advantages over other technologies. UAA incorporation is one of the smallest modifications that you can make to a protein, so this technique is less likely to perturb the capsid structure; UAA incorporation works at a variety of sites, which allows for site-specific modification of the AAV capsid; and the ability to incorporate bio-conjugation handles allows for the subsequent attachment of a wide variety of modifications to the capsid surface.

1.3 GENETIC CODE EXPANSION

The ability to create proteins with custom functionalities enables us to create tools to better understand the complex biology that occurs in cells, as well as manipulate biology to make bio-therapeutics. While proteins can be modified in a variety of ways to impart desired functionalities, such as the fusion of protein or peptide tags, or the modification of exposed residues like lysine or arginine, these methods often disturb the protein's function or alter its behavior. As UAA incorporation is such a small modification, it is less likely to impair the protein structure or function than the incorporation of a tag or fusion to another protein. GCE relies on a specifically evolved aminoacyl-tRNA Synthetase/tRNA pair to incorporate a UAA of desired functionality into a protein at a specified nonsense or frameshift codon[86,91] (Figure 4). The synthetase/tRNA pair must be orthogonal to the host's machinery; meaning that it cannot interact with the host synthetases or tRNAs to incorporate natural amino acids [86,91]. Once incorporated, UAAs provide an entirely new toolbox to work with. UAA's have functionalities including bioconjugation handles, post-translational modifications, and light activated amino acids such as photocaged and photocrosslinking amino acids[92]. This thesis will discuss the current synthetase/tRNA pairs that are in use, as well as some of the notable amino acids that have been incorporated into proteins.

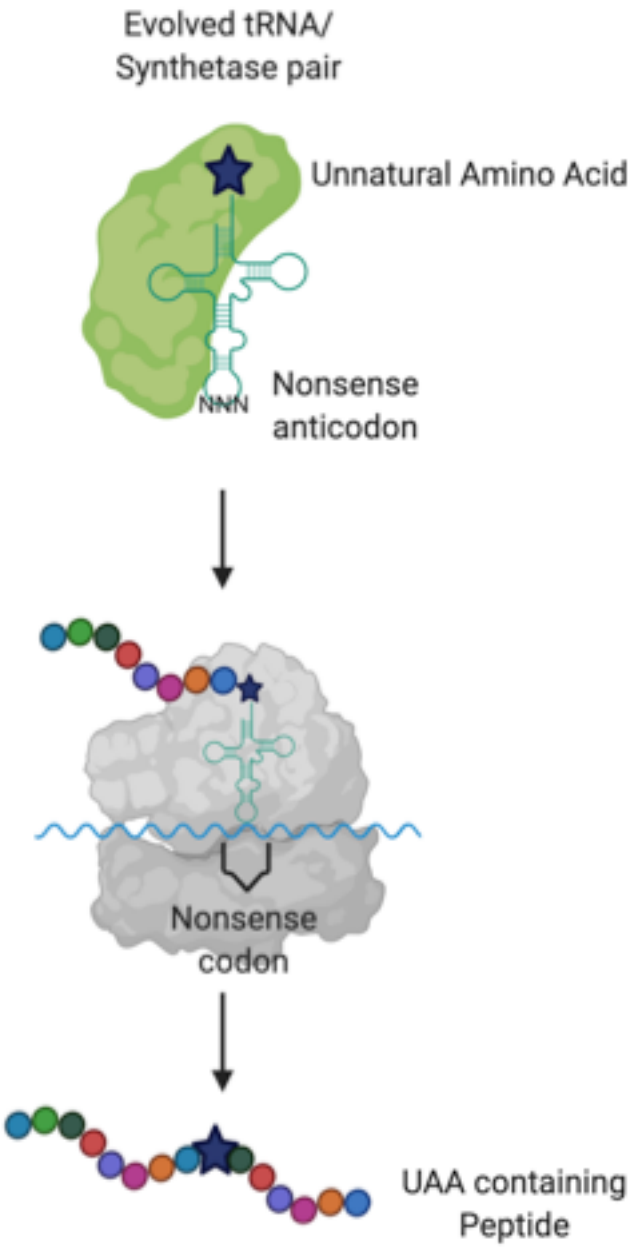


Figure 1.4. Unnatural amino acids incorporation. UAAs are incorporated using a specially evolved tRNA/tRNA synthetase pair that incorporates an unnatural amino acid in response to a nonsense codon.

1.3.1 Synthetase/tRNA pairs

There are currently four main synthetase/tRNA pairs that can be used to incorporate UAAs into cells: the *Methanococcus Jannaschii* tyrosyl (*Mj*TyrRS)/$tRNA_{CUA}$ pair that is orthogonal in *E. coli* but not in eukaryotes, the *E. coli* tyrosyl-tRNA synthetase (*Ec*TyrRS)/$tRNA_{CUA}$ pair and the *E. coli* leucyl-tRNA synthetase (*Ec*LeuRS)/$tRNA_{CUA}$ pair that are orthogonal in eukaryotes but not *E. coli*, and the pyrrolysyl-tRNA synthetase (PylRS)/$tRNA_{CUA}$ pair from either *Methanosarcina barkeri* or *Methanosarcina jannaschii,* which are orthogonal in both eukaryotes and bacteria[86]. Because the synthetase/tRNA pair must be orthogonal to the host cell machinery, these pairs are taken from a different domain of life than the host cell and evolved in

the host cell to aminoacylate the UAA of interest. In addition to these four systems, both the *E. coli* tryptophanyl-tRNA synthetase (TrpRS)/$tRNA^{Trp}$ pair and the *E. coli* tyrosyl-tRNA synthetase/tRNA pair have been evolved to be an orthogonal pair in both an engineered strain of *E. coli* and mammalian cells[93].

1.3.2 The Evolution of Orthogonal Synthetase/tRNA Pairs

The first synthetase/tRNA pair to be developed was the *Methanococcus Jannaschii* tyrosyl pair used in *Escherichia coli*[94]. The *Mj*TyrRS/tRNA pair was chosen for GCE for three main reasons: the tRNA has different identity elements to that of the *E. coli* $tRNA^{tyr}$, it does not aminoacylate *E. coli* tRNAs at a high level, and the synthetase has no editing mechanism and therefore will not deacylate the tRNA if it is mischarged. The tRNA was first optimized to reduce its recognition by *E. coli* synthetases. A tRNA library was created by randomizing 11 nucleotides and was passed through several rounds of negative and positive selections. The negative selection used suppression of amber stop codons in the barnase gene. If the host cell machinery charged the *M. jannaschii* tRNA, then the cells would translate the toxic barnase gene and die. The positive round of selection used the suppression of amber mutations in the ß-lactamase gene in the presence of Ampicillin. If the tRNA was aminoacylated, then the ß-lactamase gene would be transcribed and the cell would live in the presence of ampicillin. Next the synthetase was optimized to alter its amino acid specificity. A library was created which randomized 5 active-site amino acids. Mutants were selected based off of several rounds of positive and negative selections. The positive selection used an amber suppression placed in the

chloramphenicol acetyltransferase (CAT) gene and cells were grown in the presence of chloramphenicol and the desired amino acid. Cells that grew in the presence of chloramphenicol demonstrated the ability to charge the desired amino acid, TryRS genes from these cells were recombined and placed into the negative selection, which used the same conditions as the positive selection but replica plated the cells in both the presence and absence of UAA. The TyrRS genes from cells that grew only in the presence but not the absence of UAA were isolated, recombined, and put through more rounds of selection with increasing amounts of chloramphenicol. Using this scheme, the *Mj*TyrRS/$tRNA_{CUA}$ pair was evolved to incorporate O-methyl-tyrosine at the amber codon. This evolution scheme became a widely used way to evolve synthetases to accept specific amino acids in *E. coli*.

A similar evolution scheme was developed to develop synthtase/tRNA pairs in yeast[95–97]. The *E. coli* TyrRS/$tRNA_{CUA}$ and LeuRS/$tRNA_{CUA}$ pairs were developed using this system[98]. A library of synthetase mutants is generated and transformed along with its heterologous tRNA into *Saccharomyces cerevisea* strain that is uracil auxotrophic and contains both the *Ura3* and *HIS3* genes with stop codons in them, as well as the *lacZ* gene for easy detection under the control of the GAL4 transcriptional activator. For the positive selection, the cells are grown in the absence of uracil and the presence of the UAA. Clones that survive are able to charge the stop codon and produce uracil and survive. The surviving cells are then put through the negative selection, which grew the cells in the presence of 5- fluorootic acid (5-FOA) and uracil, and the absence of the UAA. If an endogenous amino acid

was charged at the amber stop codon, then 5-FOA was converted to a toxic product that killed the cells. Surviving clones are put through further rounds of selection.

The pyrrolysyl pairs are taken from archaeal organisms, and are therefore orthogonal in both *E. coli* and eukaryotic systems. Because of this, they can be evolved in *E. coli* and then moved into eukaryotic systems[99–101]. This, and the fact that the PylRS/$tRNA_{CUA}$ pair naturally accepts a wide range of UAAs has made it the most popular platform for UAA incorporation in both bacterial and mammalian systems[92]. The PylRS has several unique features that make its widespread use possible. The PylRS has no editing mechanism, so "mistakes" will not be proofread and fixed. Additionally, it has low selectivity towards the pyrrolysyl α-amine, which means that it will accept a variety of other substrates[102,103]. Lastly, the PylRS displays low selectivity towards the tRNA anticodon, allowing for the use of different codons to code for the UAA, such as opal UGA codons, ochre UAA codons, or 4-base codons instead of the native amber UAG codon[104–107]. The ability to use an alternative stop codon opens the possibility of incorporating multiple UAAs, which will be further discussed.

The endogenous *E. coli* tryptophanyl-tRNA synthetase (TrpRS)/$tRNA^{Trp}$ pair has been repurposed in a way that it can serve as an orthogonal pair in both *E. coli* and mammalian cells[93]. A strain was created in which an *E. coli* optimized *S. cerevisiae* TrpRS/$tRNA^{ScTrp}$ pair was used to replace the endogenous TrpRS/$tRNA^{Trp}$ pair, creating the Altered Translation Machinery (ATM) strain. The removed *Ec*TrpRS/$tRNA^{EcTrp}$ pair could then be placed back into the ATM strain as an orthogonal pair and evolved to charge UAA's at specified nonsense codons. Because

it is a bacterial pair, the evolved *Ec*TrpRS/tRNA$^{Ec\text{Trp}}$ pair is orthogonal in mammalian cells and can be moved directly into mammalian cells. The same method was used to create an ATMY strain in which the *E. coli* TyrRS/tRNA$^{Ec\text{Tyr}}$ pair was evolved to incorporate amino acids such as O-methyl-tyrosine, O-propargyl-tyrosine, and p-borono-phenylalanine[108]. While a synthetase had previously been developed that incorporates some of these UAA's, the resurrected *Ec*TyrRS/tRNA$^{Ec\text{Tyr}}$ pair showed improved incorporation[98].

The Chin lab recently came up with a strategy in which they discover five orthogonal pairs and evolve two of those pairs for UAA incorporation[109]. Their strategy uses a computational scoring system to identify tRNAs from bacteria, archaea, chloroplasts, and bacteriophages that could be orthogonal to the endogenous *E. coli* machinery. The identified tRNAs are screened for orthogonality using tRNA Expansion (tREX), a method developed specifically to measure tRNA aminoacylation that is independent of the anticodon, the nature of the UAA, and is scalable for large-scale investigation. Cognate synthetases that aminoacylated the orthogonal tRNAs were then screened for orthogonality, and pairs that were found to be orthogonal were evolved to be more even more orthogonal and specific in *E. coli.* Two of these pairs were then further evolved to incorporate UAA's at amber codons.

1.3.3 Functionalities that can be incorporated using Genetic Code Expansion

Unnatural amino acid incorporation has enabled us to use a variety of powerful new functionalities to study interactions that occur biologically. The ability

to incorporate fluorescent probes, for example, has provided a means to track a protein's activity *in vivo* or *in vitro*. UAA incorporation allows for site-specific placement of a small fluorescent probe without the addition of a genetically fused fluorescent protein or peptide tag. Inherently fluorescent amino acids that have been encoded into proteins include dansylalanine, coumarin lysine, and Anap (Figure 5, 1-3). These probes have been used for applications such as tracking proteins movement through the cell and FRET[97,110–114]. For example, dansylalanine was used to monitor the unfolding of superoxide dismutase, coumarin lysine has been used to visualize proteins such as GroEL or FitsZ, and Anap has been used for both FRET and other cellular imaging[97,110,111,114,115]. While very useful, one of the downsides of incorporating inherently fluorescent amino acids into proteins is that doing so requires evolving a specific synthetase for each fluorescent amino acid. Additionally, while the smaller probes to fit into synthetase active sites, most fluorescent probes are too large to fit into the active site and cannot be directly incorporated.

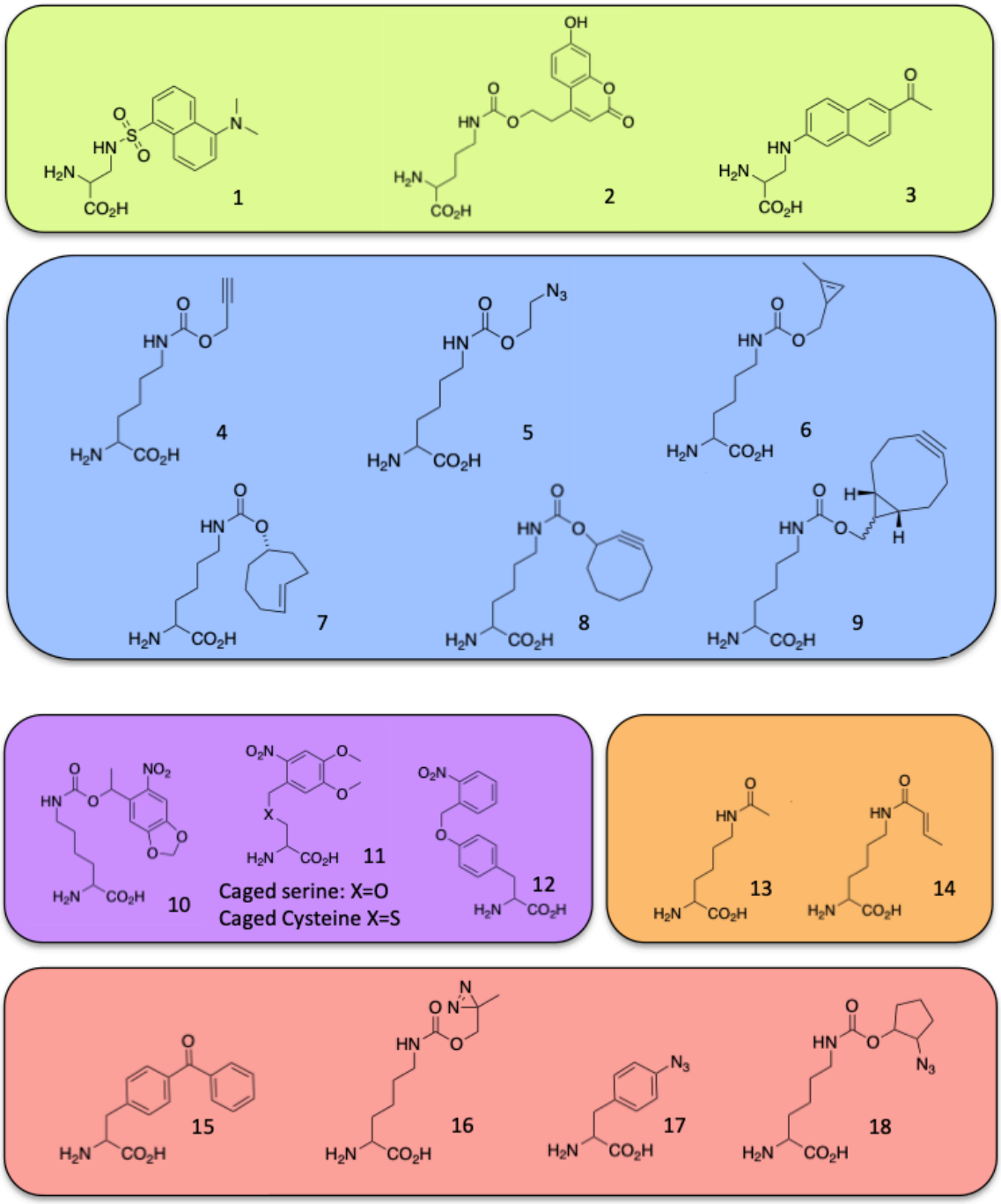


Figure 1.5. Examples of unnatural amino acids that can be genetically encoded. 1-3 are inherently fluorescent unnatural amino acids, 4-9 are bio-conjugation handles, 10-12 are photocaged amino acids, 13 and 14 are post-translational amino acids, and 15-18 are photocrosslinking amino acids.

Another approach to incorporating a fluorescent probe is to incorporate a bioconjugation handle onto which a probe can be attached. A large number of bioconjugation handles have been incorporated into cells and subsequently

modified with biophysical probes (Figure 5, 4-9). These amino acids allow for chemoselective bio-orthogonal conjugation reactions to site-specifically alter the protein of interest. Handles include functionalities such as azides, alkynes, strained alkenes, and strained alkynes that can be used in bioorthogonal reactions such as copper-catalyzed azide-alkyne cycloadditions (CuAAC) or strain-promoted azide-alkye cycloadditions (SPAAC), the Staudinger ligation, inverse electron-demand Diels-Alder cycloadditions (IEDDA), and "photoclick chemistry" between tetrazoles and alkenes. In addition to enabling modification with fluorescent probes for cellular tracking[116–120], bioconjugation handles have been used to create antibody-drug conjugates and other therapeutically relevant proteins[121], and modify virus capsids[31,87,88,90,122,123], among other things.

Optogenetic probes include photocaged and photocrosslinking amino acids. Photocaged amino acids have protection groups that are light removable, resulting in the exposure of the corresponding native amino acid (Figure 5, 10-12). This approach offers control over specific protein functions. Derivatives of tyrosine, cysteine, lysine, and serine have been incorporated into both prokaryotes and eukaryotes. These UAAs have been used to gain optochemical control of the activity of ion channels, kinases, protein localization through a caged nuclear localization signal, and control of gene expression through caged nucleases and polymerases[96,105,132,124–131]. Photocrosslinking amino acids allow for the study of protein-protein or nucleic acid-protein interactions. These amino acids respond to light stimuli to form rapid covalent bonds to organic molecules they are interacting with in close proximity. This method allows identification of weak, transient, or pH-

dependent interactions that would not get identified without the covalent interaction and can be identified using a pull down. Photocrosslinking amino acids such as diazirine lysine, p-azidophenylalanine, and p-benzoyl-L-phenylalanine (Figure 5, 15-18) have been incorporated and used to study binding interactions with proteins such as the Growth factor receptor bound protein (Grb2)[133,134], the enteric bacterial pathogens' acid stress chaperone HdeA[135], and the human serotonin transporter (hSERT)[136]. Additional studies worth noting include the incorporation of DiZHSeC, a photocrosslinker-DiZPK that has a selenium atom in the place of the γ-carbon, which allows for the oxidative cleavage of the photocrosslinker, leaving a mass spectrometry-identifiable label on the bait. This design was termed In-situ cleavage and MS-label transfer After Protein Photocrosslinking (IMAPP), and facilitated in both the pre-bait separation and the downstream target identification[137]. A second study that stands out used the incorporation of a post-translational modification, acetyl lysine (AcK), into histone 3 to lure the bait. This transient interaction was captured using the photocrosslinker p-benzoylphenylalanine (pBpA), showing the effectiveness of incorporating multiple amino acids[138].

Most proteins undergo some form of post-translational modification on their amino acid side chains; however studying these modifications has proven difficult due to challenges in attaining homogenous preparations of these proteins. The ability to encode amino acids which harbor post-translational modifications has allowed the preparation of large, homogenous preps of specifically modified proteins, with post-translationally modified amino acids such as acetyl-lysine,

crotonyl-Lysine, hydroxyisobutyryl-lysine, sulfotyrosine, and phosphoserine (Figure 5, 13-14)[99,139,148,140–147]. For example the interactions made with histones have been extensively studied using acetyl-lysine, crotonyl-lysine, dimethyl-lysine, and hydroxyisobutyryl-lysine[138,142–145,149]. Another study incorporated both Acetyl-lysine and phosphoserine into malate dehydrogenase, found in the citric acid cycle, to study the effects of multiple post-translational modifications in a protein[141].

The vast majority of studies incorporating UAAs have been done using a single UAA. While this greatly facilitates our ability to study and understand protein activities, incorporating multiple UAAs is advantageous as it enables multiple new functionalities to be incorporated. For example, one could incorporate a photocaged amino acid as well as a chemical handle to be used for photocrosslinking and the subsequent pull-down. In order to incorporate two distinct UAAs, two distinct and orthogonal systems have to be used. These systems have to be orthogonal both to the host machinery as well as to each other. This has accomplished in *E. coli* using a variety of approaches, and more recently, a method was developed to incorporate up to three distinct UAAs into mammalian cells[105,107,141,150–154].

1.4 REFERENCES

(1) Samulski, R. J.; Muzyczka, N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annu. Rev. Virol.* **2014**, *1* (1), 427–451. https://doi.org/10.1146/annurev-virology-031413-085355.

(2) Linden, M. D. W. and R. M. Adeno-Associated Virus Biology. *Gene Ther.* **2011**, *807*, 119–140. https://doi.org/10.1007/978-1-61779-370-7.

(3) Im, D. S.; Muzyczka, N. The AAV Origin Binding Protein Rep68 Is an ATP-Dependent Site-Specific Endonuclease with DNA Helicase Activity. *Cell* **1990**,

61 (3), 447–457. https://doi.org/10.1016/0092-8674(90)90526-K.
(4) Im, D. S.; Muzyczka, N. Partial Purification of Adeno-Associated Virus Rep78, Rep52, and Rep40 and Their Biochemical Characterization. *J. Virol.* **1992**, *66* (2), 1119–1128. https://doi.org/10.1128/jvi.66.2.1119-1128.1992.
(5) King, J. A.; Dubielzig, R.; Grimm, D.; Kleinschmidt, J. A. DNA Helicase-Mediated Packaging of Adeno-Associated Virus Type 2 Genomes into Preformed Capsids. *EMBO J.* **2001**, *20* (12), 3282–3291. https://doi.org/10.1093/emboj/20.12.3282.
(6) Smith, R. H.; Kotin, R. M. The Rep52 Gene Product of Adeno-Associated Virus Is a DNA Helicase with 3′-to-5′ Polarity. *J. Virol.* **1998**, *72* (6), 4874–4881. https://doi.org/10.1128/jvi.72.6.4874-4881.1998.
(7) Sonntag, F.; Bleker, S.; Leuchs, B.; Fischer, R.; Kleinschmidt, J. A. Adeno-Associated Virus Type 2 Capsids with Externalized VP1/VP2 Trafficking Domains Are Generated Prior to Passage through the Cytoplasm and Are Maintained until Uncoating Occurs in the Nucleus. *J. Virol.* **2006**, *80* (22), 11040–11054. https://doi.org/10.1128/jvi.01056-06.
(8) Sonntag, F.; Schmidt, K.; Kleinschmidt, J. A. A Viral Assembly Factor Promotes AAV2 Capsid Formation in the Nucleolus. *Proc. Natl. Acad. Sci.* **2010**, *107* (22), 10220–10225. https://doi.org/10.1073/pnas.1001673107.
(9) Naumer, M.; Sonntag, F.; Schmidt, K.; Nieto, K.; Panke, C.; Davey, N. E.; Popa-Wagner, R.; Kleinschmidt, J. A. Properties of the Adeno-Associated Virus Assembly-Activating Protein. *J. Virol.* **2012**, *86* (23), 13038–13048. https://doi.org/10.1128/JVI.01675-12.
(10) Becerra, S. P.; Koczot, F.; Fabisch, P.; Rose, J. A. Synthesis of Adeno-Associated Virus Structural Proteins Requires Both Alternative MRNA Splicing and Alternative Initiations from a Single Transcript. *J. Virol.* **1988**, *62* (8), 2745–2754. https://doi.org/10.1128/jvi.62.8.2745-2754.1988.
(11) Zádori, Z.; Szelei, J.; Lacoste, M. C.; Li, Y.; Gariépy, S.; Raymond, P.; Allaire, M.; Nabi, I. R.; Tijssen, P. A Viral Phospholipase A2 Is Required for Parvovirus Infectivity. *Dev. Cell* **2001**, *1* (2), 291–302. https://doi.org/10.1016/S1534-5807(01)00031-4.
(12) Stahnke, S.; Lux, K.; Uhrig, S.; Kreppel, F.; Hösel, M.; Coutelle, O.; Ogris, M.; Hallek, M.; Büning, H. Intrinsic Phospholipase A2 Activity of Adeno-Associated Virus Is Involved in Endosomal Escape of Incoming Particles. *Virology* **2011**, *409* (1), 77–83. https://doi.org/10.1016/j.virol.2010.09.025.
(13) Kronenberg, S.; Bottcher, B.; von der Lieth, C. W.; Bleker, S.; Kleinschmidt, J. A. A Conformational Change in the Adeno-Associated Virus Type 2 Capsid Leads to the Exposure of Hidden VP1 N Termini. *J. Virol.* **2005**, *79* (9), 5296–5303. https://doi.org/10.1128/jvi.79.9.5296-5303.2005.
(14) Earley, L. F.; Kawano, Y.; Adachi, K.; Sun, X.-X.; Dai, M.-S.; Nakai, H. Identification and Characterization of Nuclear and Nucleolar Localization Signals in the Adeno-Associated Virus Serotype 2 Assembly-Activating Protein. *J. Virol.* **2015**, *89* (6), 3038–3048. https://doi.org/10.1128/jvi.03125-14.
(15) Grieger, J. C.; Snowdy, S.; Samulski, R. J. Separate Basic Region Motifs within the Adeno-Associated Virus Capsid Proteins Are Essential for Infectivity and

Assembly. *J. Virol.* **2006**, *80* (11), 5199–5210. https://doi.org/10.1128/jvi.02723-05.
(16) Johnson, J. S.; Li, C.; DiPrimio, N.; Weinberg, M. S.; McCown, T. J.; Samulski, R. J. Mutagenesis of Adeno-Associated Virus Type 2 Capsid Protein VP1 Uncovers New Roles for Basic Amino Acids in Trafficking and Cell-Specific Transduction. *J. Virol.* **2010**, *84* (17), 8888–8902. https://doi.org/10.1128/jvi.00687-10.
(17) Xie, Q.; Bu, W.; Bhatia, S.; Hare, J.; Somasundaram, T.; Azzi, A.; Chapman, M. S. The Atomic Structure of Adeno-Associated Virus (AAV-2), a Vector for Human Gene Therapy. *Proc. Natl. Acad. Sci.* **2002**, *99* (16), 10405–10410. https://doi.org/10.1073/pnas.162250899.
(18) Nam, H.-J.; Lane, M. D.; Padron, E.; Gurda, B.; McKenna, R.; Kohlbrenner, E.; Aslanidi, G.; Byrne, B.; Muzyczka, N.; Zolotukhin, S.; et al. Structure of Adeno-Associated Virus Serotype 8, a Gene Therapy Vector. *J. Virol.* **2007**, *81* (22), 12260–12271. https://doi.org/10.1128/jvi.01304-07.
(19) Padron, E.; Bowman, V.; Kaludov, N.; Govindasamy, L.; Levy, H.; Nick, P.; McKenna, R.; Muzyczka, N.; Chiorini, J. A.; Baker, T. S.; et al. Structure of Adeno-Associated Virus Type 4. *J. Virol.* **2005**, *79* (8), 5047–5058. https://doi.org/10.1128/jvi.79.8.5047-5058.2005.
(20) Kern, A.; Schmidt, K.; Leder, C.; Müller, O. J.; Wobus, C. E.; Bettinger, K.; Von der Lieth, C. W.; King, J. A.; Kleinschmidt, J. A. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. *J. Virol.* **2003**, *77* (20), 11072–11081.
(21) Opie, S. R.; Warrington, K. H.; Agbandje-McKenna, M.; Zolotukhin, S.; Muzyczka, N. Identification of Amino Acid Residues in the Capsid Proteins of Adeno-Associated Virus Type 2 That Contribute to Heparan Sulfate Proteoglycan Binding. *J. Virol.* **2003**, *77* (12), 6995–7006.
(22) Ummerford, C. A. S.; Artlett, J. E. S. B.; Amulski, R. I. J. U. D. E. S. α V β 5 Integrin : A Co-Receptor for Adeno-Associated Virus Type 2 Infection. **1999**, *5* (1), 1–5.
(23) Qing, K.; Mah, C.; Hansen, J.; Zhou, S.; Dwarki, V.; Srivastava, A. Human Fibroblast Growth Factor Receptor 1 Is a Co-Receptor for Infection by Adeno-Associated Virus 2. *Nat. Med.* **1999**, *5* (1), 71–77. https://doi.org/10.1038/4758.
(24) Akache, B.; Grimm, D.; Pandey, K.; Yant, S. R.; Xu, H.; Kay, M. A. The 37/67-Kilodalton Laminin Receptor Is a Receptor for Adeno-Associated Virus Serotypes 8, 2, 3, and 9. *J. Virol.* **2006**, *80* (19), 9831–9836. https://doi.org/10.1128/jvi.00878-06.
(25) Kashiwakura, Y.; Tamayose, K.; Iwabuchi, K.; Hirai, Y.; Shimada, T.; Matsumoto, K.; Nakamura, T.; Watanabe, M.; Oshimi, K.; Daida, H. Hepatocyte Growth Factor Receptor Is a Coreceptor for Adeno-Associated Virus Type 2 Infection. *J. Virol.* **2005**, *79* (1), 609–614. https://doi.org/10.1128/jvi.79.1.609-614.2005.
(26) Pillay, S.; Meyer, N. L.; Puschnik, A. S.; Davulcu, O.; Diep, J.; Ishikawa, Y.; Jae, L. T.; Wosen, J. E.; Nagamine, C. M.; Chapman, M. S.; et al. An Essential Receptor for Adeno-Associated Virus Infection. *Nature* **2016**, *530* (7588), 108–112.

https://doi.org/10.1038/nature16465.
(27) Nonnenmacher, M.; Weber, T. Adeno-Associated Virus 2 Infection Requires Endocytosis through the CLIC/GEEC Pathway. *Cell Host Microbe* **2011**, *10* (6), 563–576. https://doi.org/10.1016/j.chom.2011.10.014.
(28) Nonnenmacher, M.; Weber, T. Intracellular Transport of Recombinant Adeno-Associated Virus Vectors. *Physiol. Behav.* **2011**, *176* (1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040.
(29) Sanlioglu, S.; Benson, P. K.; Yang, J.; Atkinson, E. M.; Reynolds, T.; Engelhardt, J. F. Endocytosis and Nuclear Trafficking of Adeno-Associated Virus Type 2 Are Controlled by Rac1 and Phosphatidylinositol-3 Kinase Activation. *J. Virol.* **2000**, *74* (19), 9184–9196. https://doi.org/10.1128/jvi.74.19.9184-9196.2000.
(30) Bartlett, J. S.; Wilcher, R.; Samulski, R. J. Infectious Entry Pathway of Adeno-Associated Virus and Adeno-Associated Virus Vectors. *J. Virol.* **2000**, *74* (6), 2777–2785. https://doi.org/10.1128/JVI.74.6.2777-2785.2000.
(31) Zhang, C.; Zhou, X.; Yao, T.; Tian, Z.; Zhou, D. Precision Fluorescent Labeling of an Adeno-Associated Virus Vector to Monitor the Viral Infection Pathway. *Biotechnol. J.* **2018**, *1700374*, 1–11. https://doi.org/10.1002/biot.201700374.
(32) Akache, B.; Grimm, D.; Shen, X.; Fuess, S.; Yant, S. R.; Glazer, D. S.; Park, J.; Kay, M. A. A Two-Hybrid Screen Identifies Cathepsins B and L as Uncoating Factors for Adeno-Associated Virus 2 and 8. *Mol. Ther.* **2007**, *15* (2), 330–339. https://doi.org/10.1038/sj.mt.6300053.
(33) Hoque, M.; Ishizu, K.; Matsumoto, A.; Han, S.-I.; Arisaka, F.; Takayama, M.; Suzuki, K.; Kato, K.; Kanda, T.; Watanabe, H.; et al. Nuclear Transport of the Major Capsid Protein Is Essential for Adeno-Associated Virus Capsid Formation. *J. Virol.* **1999**, *73* (9), 7912–7915. https://doi.org/10.1128/jvi.73.9.7912-7915.1999.
(34) Popa-Wagner, R.; Sonntag, F.; Schmidt, K.; King, J.; Kleinschmidt, J. A. Nuclear Translocation of Adeno-Associated Virus Type 2 Capsid Proteins for Virion Assembly. *J. Gen. Virol.* **2012**, *93* (PART 9), 1887–1898. https://doi.org/10.1099/vir.0.043232-0.
(35) Weitzman, M. D.; Kyosti, S. R. M.; Kotin, R. M.; Owens, R. A. Adeno-Associated Virus (AAV) Rep Proteins Mediate Complex Formation between AAV DNA and Its Integration Site in Human DNA. **1994**, *91* (June), 5808–5812.
(36) Samulski, R. J.; Zhu, X.; Xiao, X.; Brook, J. D.; Housman, D. E.; Epstein, N.; Hunter, L. A. Targeted Integration of Adeno-Associated Virus (AAV) into Human Chromosome 19. *EMBO J.* **1991**, *10* (12), 3941–3950. https://doi.org/10.1002/j.1460-2075.1991.tb04964.x.
(37) Linden, R. M.; Ward, P.; Giraud, C.; Winocourt, E.; Berns, K. I. Site-Specific Integration by Adeno-Associated Virus. **1996**, *93* (October), 11288–11294.
(38) Surosky, R. T.; Urabe, M.; Godwin, S. G.; McQuiston, S. A.; Kurtzman, G. J.; Ozawa, K.; Natsoulis, G. Adeno-Associated Virus Rep Proteins Target DNA Sequences to a Unique Locus in the Human Genome. *J. Virol.* **1997**, *71* (10), 7951–7959. https://doi.org/10.1128/jvi.71.10.7951-7959.1997.
(39) Johnson, J. S.; Samulski, R. J. Enhancement of Adeno-Associated Virus Infection by Mobilizing Capsids into and Out of the Nucleolus. *J. Virol.* **2009**, *83* (6),

2632–2644. https://doi.org/10.1128/jvi.02309-08.
(40) Bennett, A.; Mietzsch, M.; Agbandje-Mckenna, M. Understanding Capsid Assembly and Genome Packaging for Adeno-Associated Viruses. *Future Virol.* **2017**, *12* (6), 283–297. https://doi.org/10.2217/fvl-2017-0011.
(41) Sonntag, F.; Kother, K.; Schmidt, K.; Weghofer, M.; Raupp, C.; Nieto, K.; Kuck, A.; Gerlach, B.; Bottcher, B.; Muller, O. J.; et al. The Assembly-Activating Protein Promotes Capsid Assembly of Different Adeno-Associated Virus Serotypes. *J. Virol.* **2011**, *85* (23), 12686–12697. https://doi.org/10.1128/jvi.05359-11.
(42) Grosse, S.; Penaud-Budloo, M.; Herrmann, A.-K.; Börner, K.; Fakhiri, J.; Laketa, V.; Krämer, C.; Wiedtke, E.; Gunkel, M.; Ménard, L.; et al. Relevance of Assembly-Activating Protein for Adeno-Associated Virus Vector Production and Capsid Protein Stability in Mammalian and Insect Cells. *J. Virol.* **2017**, *91* (20), 1–30. https://doi.org/10.1128/jvi.01198-17.
(43) Agbandje-McKenna, M.; Kleinschmidt, J. *AAV Capsid Structure and Cell Interactions*; 2011; Vol. 807. https://doi.org/10.1007/978-1-61779-370-7.
(44) Li, C.; Samulski, R. J. Engineering Adeno-Associated Virus Vectors for Gene Therapy. *Nat. Rev. Genet.* **2020**, *21* (April). https://doi.org/10.1038/s41576-019-0205-4.
(45) Keeler, A. M.; Flotte, T. R. Recombinant Adeno-Associated Virus Gene Therapy in Light of Luxturna (and Zolgensma and Glybera): Where Are We, and How Did We Get Here? *Annu. Rev. Virol.* **2019**, *6* (1), 601–621. https://doi.org/10.1146/annurev-virology-092818-015530.
(46) Bryant, L. M.; Christopher, D. M.; Giles, A. R.; Hinderer, C.; Rodriguez, J. L.; Smith, J. B.; Traxler, E. A.; Tycko, J.; Wojno, A. P.; Wilson, J. M. Lessons Learned from the Clinical Development and Market Authorization of Glybera. *Hum. Gene Ther. Clin. Dev.* **2013**, *24* (2), 55–64. https://doi.org/10.1089/humc.2013.087.
(47) Wierzbicki, A. S.; Viljoen, A. Alipogene Tiparvovec: Gene Therapy for Lipoprotein Lipase Deficiency. *Expert Opin. Biol. Ther.* **2012**, *13* (1), 7–10. https://doi.org/10.1517/14712598.2013.738663.
(48) Russell, P. S.; Bennett, P. J.; Wellman, J. A.; Daniel, C.; Yu, Z.; Tillman, A.; Wittes, J.; Pappas, J.; Mccague, S.; Cross, D.; et al. Efficacy and Safety of Voretigene Neparvovec (AAV2-HRPE65v2) in Patients with RVE65-Mediated Inherited Retinal Dystrophy: A Randomised, Controlled, Open-Label, Phase 3 Trial. **2018**, *390* (10097), 849–860. https://doi.org/10.1016/S0140-6736(17)31868-8.Efficacy.
(49) Waldrop, M. A.; Kolb, S. J. Current Treatment Options in Neurology—SMA Therapeutics. *Curr. Treat. Options Neurol.* **2019**, *21* (6). https://doi.org/10.1007/s11940-019-0568-z.
(50) Dabbous, O.; Sproule, D. M.; Feltner, D. E.; Droege, M.; Khan, F.; Arjunji, R. The Value of Onasemnogene Abeparvovec (AVXS-101) Gene-Replacement Therapy for Spinal Muscular Atrophy Type 1. *J. Neurol. Sci.* **2019**, *405*, 59. https://doi.org/10.1016/j.jns.2019.10.327.
(51) Bartlett, J. S.; Kleinschmidt, J.; Boucher, R. C.; Samulski, R. J. Targeted Adeno-Associated Virus Vector Transduction of Nonpermissive Cells Mediated by a Bispecific f(Ab'γ)2 Antibody. *Nat. Biotechnol.* **1999**, *17* (2), 181–186.

https://doi.org/10.1038/6185.
(52) Ponnazhagan, S.; Mahendra, G.; Kumar, S.; Thompson, J. A.; Castillas, M. Conjugate-Based Targeting of Recombinant Adeno-Associated Virus Type 2 Vectors by Using Avidin-Linked Ligands. *J. Virol.* **2002**, *76* (24), 12900–12907. https://doi.org/10.1128/jvi.76.24.12900-12907.2002.
(53) Lee, S.; Ahn, H. J. Anti-EpCAM-Conjugated Adeno-Associated Virus Serotype 2 for Systemic Delivery of EGFR ShRNA: Its Retargeting and Antitumor Effects on OVCAR3 Ovarian Cancer in Vivo. *Acta Biomater.* **2019**, *91*, 258–269. https://doi.org/10.1016/j.actbio.2019.04.044.
(54) Horowitz, E. D.; Weinberg, M. S.; Asokan, A. Glycated AAV Vectors: Chemical Redirection of Viral Tissue Tropism. *Bioconjug. Chem.* **2011**, *22* (4), 529–532. https://doi.org/10.1021/bc100477g.
(55) Leinschmidt, J. Ü. K.; Eléage, G. I. D.; Allek, M. I. H. Genetic Capsid Modifications Allow Efficient Re-Targeting of Adeno-Associated Virus Type 2. **1999**, *5* (9).
(56) Grifman, M.; Trepel, M.; Speece, P.; Gilbert, L. B.; Arap, W.; Pasqualini, R.; Weitzman, M. D. Incorporation of Tumor-Targeting Peptides into Recombinant Adeno-Associated Virus Capsids. *Mol. Ther.* **2001**, *3* (6), 964–975. https://doi.org/10.1006/mthe.2001.0345.
(57) Girod, A.; Buning, H.; Hallek, M.; Leike, K.; Ried, M. U. Adeno-Associated Virus Capsids Displaying Immunoglobulin-Binding Domains Permit Antibody-Mediated Vector Retargeting to Specific Cell Surface Receptors. *J. Virol.* **2002**, *76* (9), 4559–4566. https://doi.org/10.1128/jvi.76.9.4559-4566.2002.
(58) Müller, O. J.; Kaul, F.; Weitzman, M. D.; Pasqualini, R.; Arap, W.; Kleinschmidt, J. A.; Trepel, M. Random Peptide Libraries Displayed on Adeno-Associated Virus to Select for Targeted Gene Therapy Vectors. *Nat. Biotechnol.* **2003**, *21* (9), 1040–1046. https://doi.org/10.1038/nbt856.
(59) Shi, W.; Bartlett, J. S. RGD Inclusion in VP3 Provides Adeno-Associated Virus Type 2 (AAV2)-Based Vectors with a Heparan Sulfate-Independent Cell Entry Mechanism. *Mol. Ther.* **2003**, *7* (4), 515–525. https://doi.org/10.1016/S1525-0016(03)00042-X.
(60) Douar, A. M.; Poulard, K.; Danos, O. Deleterious Effect of Peptide Insertions in a Permissive Site of the AAV2 Capsid. *Virology* **2003**, *309* (2), 203–208. https://doi.org/10.1016/S0042-6822(03)00186-7.
(61) Shi, W.; Hemminki, A.; Bartlett, J. S. Capsid Modifications Overcome Low Heterogeneous Expression of Heparan Sulfate Proteoglycan That Limits AAV2-Mediated Gene Transfer and Therapeutic Efficacy in Human Ovarian Carcinoma. *Gynecol. Oncol.* **2006**, *103* (3), 1054–1062. https://doi.org/10.1016/j.ygyno.2006.06.020.
(62) Arnold, G. S.; Sasser, A. K.; Stachler, M. D.; Bartlett, J. S. Metabolic Biotinylation Provides a Unique Platform for the Purification and Targeting of Multiple AAV Vector Serotypes. *Mol. Ther.* **2006**, *14* (1), 97–106. https://doi.org/10.1016/j.ymthe.2006.02.014.
(63) Stachler, M. D.; Chen, I.; Ting, A. Y.; Bartlett, J. S. Site-Specific Modification of AAV Vector Particles with Biophysical Probes and Targeting Ligands Using Biotin Ligase. *Mol. Ther.* **2008**, *16* (8), 1467–1473. https://doi.org/10.1038/mt.2008.129.

(64) Boucas, J.; Lux, K.; Huber, A.; Schievenbusch, S.; von Freyend, M. J.; Perabo, L.; Quadt-Humme, S.; Odenthal, M.; Hallek, M.; BÃ¼ning, H. Engineering Adeno-Associated Virus Serotype 2-Based Targeting Vectors Using a New Insertion Site-Position 453-and Single Point Mutations. *J. Gene Med.* **2009**, *11* (12), 1103–1113. https://doi.org/10.1002/jgm.1392.

(65) Liu, Y.; Fang, Y.; Zhou, Y.; Zandi, E.; Lee, C. L.; Joo, K. Il; Wang, P. Site-Specific Modification of Adeno-Associated Viruses via a Genetically Engineered Aldehyde Tag. *Small* **2013**, *9* (3), 421–429. https://doi.org/10.1002/smll.201201661.

(66) Zdechlik, A. C.; He, Y.; Aird, E. J.; Gordon, W. R.; Schmidt, D. Programmable Assembly of Adeno-Associated Virus-Antibody Composites for Receptor-Mediated Gene Delivery. *Bioconjug. Chem.* **2019**. https://doi.org/10.1021/acs.bioconjchem.9b00790.

(67) Beckett, D.; Kovaleva, E.; Petter, S.; Schatz, P. J. A Minimal Peptide Substrate in Biotin Holoenzyme Synthetase-Catalyzed Biotinylation. *Protien Sci.* **1999**, *8*, 921–929.

(68) Wu, P.; Xiao, W.; Conlon, T.; Hughes, J.; Agbandje-McKenna, M.; Ferkol, T.; Flotte, T.; Muzyczka, N. Mutational Analysis of the Adeno-Associated Virus Type 2 (AAV2) Capsid Gene and Construction of AAV2 Vectors with Altered Tropism. *J. Virol.* **2000**, *74* (18), 8635–8647. https://doi.org/10.1128/JVI.74.18.8635-8647.2000.

(69) Jr, K. H. W.; Gorbatyuk, O. S.; Jeffrey, K.; Opie, S. R.; Zolotukhin, S.; Warrington, K. H.; Harrison, J. K.; Muzyczka, N. Adeno-Associated Virus Type 2 VP2 Capsid Protein Is Nonessential and Can Tolerate Large Peptide Insertions at Its N Terminus. *J. Virol.* **2004**, *78* (12), 6595–6609. https://doi.org/10.1128/JVI.78.12.6595.

(70) Yang, Q.; Mamounas, M.; Yu, G.; Kennedy, S.; Leaker, B.; Mersn, J.; Yu, F. W. M. A. N. G.; Barber, J. R.; Al, Y. E. T. Development of Novel Cell Surface CD34-Targeted Recombinant Adenoassociated Virus for Gene Therapy. **1998**, *1937*, 1929–1937.

(71) Münch, R. C.; Janicki, H.; Völker, I.; Rasbach, A.; Hallek, M.; Büning, H.; Buchholz, C. J. Displaying High-Affinity Ligands on Adeno-Associated Viral Vectors Enables Tumor Cell-Specific and Safe Gene Transfer. *Mol. Ther.* **2013**, *21* (1), 109–118. https://doi.org/10.1038/mt.2012.186.

(72) Hagen, S.; Baumann, T.; Wagner, H. J.; Morath, V.; Kaufmann, B.; Fischer, A.; Bergmann, S.; Schindler, P.; Arndt, K. M.; Müller, K. M. Modular Adeno-Associated Virus (RAAV) Vectors Used for Cellular Virus-Directed Enzyme Prodrug Therapy. *Sci. Rep.* **2014**, *4*, 1–11. https://doi.org/10.1038/srep03759.

(73) Münch, R. C.; Muth, A.; Muik, A.; Friedel, T.; Schmatz, J.; Dreier, B.; Trkola, A.; Plückthun, A.; Büning, H.; Buchholz, C. J. Off-Target-Free Gene Delivery by Affinity-Purified Receptor-Targeted Viral Vectors. *Nat. Commun.* **2015**, *6*, 2–10. https://doi.org/10.1038/ncomms7246.

(74) Hörner, M.; Kaufmann, B.; Cotugno, G.; Wiedtke, E.; Büning, H.; Grimm, D.; Weber, W. A Chemical Switch for Controlling Viral Infectivity. *Chem. Commun.* **2014**, *50* (71), 10319–10322. https://doi.org/10.1039/c4cc03292f.

(75) Bartel, M. A.; Weinstein, J. R.; Schaffer, D. V. Directed Evolution of Novel Adeno-Associated Viruses for Therapeutic Gene Delivery. *Gene Ther.* **2012**, *19* (6), 694–700. https://doi.org/10.1038/gt.2012.20.
(76) Perabo, L.; Endell, J.; King, S.; Lux, K.; Goldnau, D.; Hallek, M.; Büning, H. Combinatorial Engineering of a Gene Therapy Vector: Directed Evolution of Adeno-Associated Virus. *J. Gene Med.* **2006**, *8* (2), 155–162. https://doi.org/10.1002/jgm.849.
(77) Maheshri, N.; Koerber, J. T.; Kaspar, B. K.; Schaffer, D. V. Directed Evolution of Adeno-Associated Virus Yields Enhanced Gene Delivery Vectors. *Nat. Biotechnol.* **2006**, *24* (2), 198–204. https://doi.org/10.1038/nbt1182.
(78) Asuri, P.; Bartel, M. A.; Vazin, T.; Jang, J. H.; Wong, T. B.; Schaffer, D. V. Directed Evolution of Adeno-Associated Virus for Enhanced Gene Delivery and Gene Targeting in Human Pluripotent Stem Cells. *Mol. Ther.* **2012**, *20* (2), 329–338. https://doi.org/10.1038/mt.2011.255.
(79) Li, W.; Asokan, A.; Wu, Z.; Van Dyke, T.; DiPrimio, N.; Jarrod, J. S.; Govindaswamy, L.; Agbandje-McKenna, M.; Leichtle, S.; Redmond, E. D.; et al. Engineering and Selection of Shuffled AAV Genomes: A New Strategy for Producing Targeted Biological Nanoparticles. *Mol. Ther.* **2008**, *16* (7), 1252–1260. https://doi.org/10.1038/mt.2008.100.
(80) Grimm, D.; Lee, J. S.; Wang, L.; Desai, T.; Akache, B.; Storm, T. A.; Kay, M. A. In Vitro and In Vivo Gene Therapy Vector Evolution via Multispecies Interbreeding and Retargeting of Adeno-Associated Viruses. *J. Virol.* **2008**, *82* (12), 5887–5911. https://doi.org/10.1128/jvi.00254-08.
(81) Lisowski, L.; Dane, A. P.; Chu, K.; Zhang, Y.; Cunningham, S. C.; Wilson, E. M.; Nygaard, S.; Grompe, M.; Alexander, I. E.; Kay, M. A. Selection and Evaluation of Clinically Relevant AAV Variants in a Xenograft Liver Model. *Nature* **2014**, *506* (7488), 382–386. https://doi.org/10.1038/nature12875.
(82) Pekrun, K.; De Alencastro, G.; Luo, Q. J.; Liu, J.; Kim, Y.; Nygaard, S.; Galivo, F.; Zhang, F.; Song, R.; Tiffany, M. R.; et al. Using a Barcoded AAV Capsid Library to Select for Clinically Relevant Gene Therapy Vectors. *JCI Insight* **2019**, *4* (22). https://doi.org/10.1172/jci.insight.131610.
(83) Cabanes-Creus, M.; Ginn, S. L.; Amaya, A. K.; Liao, S. H. Y.; Westhaus, A.; Hallwirth, C. V.; Wilmott, P.; Ward, J.; Dilworth, K. L.; Santilli, G.; et al. Codon-Optimization of Wild-Type Adeno-Associated Virus Capsid Sequences Enhances DNA Family Shuffling While Conserving Functionality. *Mol. Ther. - Methods Clin. Dev.* **2019**, *12* (March), 71–84. https://doi.org/10.1016/j.omtm.2018.10.016.
(84) Zinn, E.; Pacouret, S.; Khaychuk, V.; Turunen, H. T.; Carvalho, L. S.; Andres-Mateos, E.; Shah, S.; Shelke, R.; Maurer, A. C.; Maurer, E.; et al. In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep.* **2015**, *12* (6), 1056–1068. https://doi.org/10.1016/j.celrep.2015.07.019.
(85) Santiago-Ortiz, J.; Ojala, D. S.; Westesson, O.; Weinstein, J. R.; Wong, S. Y.; Steinsapir, A.; Kumar, S.; Holmes, I.; Schaffer, D. V. AAV Ancestral Reconstruction Library Enables Selection of Broadly Infectious Viral Variants. *Gene Ther.* **2015**, *22* (12), 934–946. https://doi.org/10.1038/gt.2015.74.

(86) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79* (1), 413–444. https://doi.org/10.1146/annurev.biochem.052308.105824.
(87) Zhang, C.; Yao, T.; Zheng, Y.; Li, Z.; Zhang, Q.; Zhang, L.; Zhou, D. Development of next Generation Adeno-Associated Viral Vectors Capable of Selective Tropism and Efficient Gene Delivery. *Biomaterials* **2016**, *80*, 134–145. https://doi.org/10.1016/j.biomaterials.2015.11.066.
(88) Kelemen, R. E.; Mukherjee, R.; Cao, X.; Erickson, S. B.; Zheng, Y.; Chatterjee, A. A Precise Chemical Strategy To Alter the Receptor Specificity of the Adeno-Associated Virus. *Angew. Chemie - Int. Ed.* **2016**, *55* (36), 10645–10649. https://doi.org/10.1002/anie.201604067.
(89) Katrekar, D.; Moreno, A. M.; Chen, G.; Worlikar, A.; Mali, P. Oligonucleotide Conjugated Multi-Functional Adeno-Associated Viruses. *Sci. Rep.* **2018**, *8* (1), 3589. https://doi.org/10.1038/s41598-018-21742-x.
(90) Erickson, S. B.; Mukherjee, R.; Kelemen, R. E.; Wrobel, C. J. J.; Cao, X.; Chatterjee, A. Precise Photoremovable Perturbation of a Virus–Host Interaction. *Angew. Chemie - Int. Ed.* **2017**, *56* (15), 4234–4237. https://doi.org/10.1002/anie.201700683.
(91) Xiao, H.; Schultz, P. G. At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (9). https://doi.org/10.1101/cshperspect.a023945.
(92) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing Logical Codon Reassignment-Expanding the Chemistry in Biology. *Chem. Sci.* **2015**, *6* (1), 50–69. https://doi.org/10.1039/c4sc01534g.
(93) 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. *Nat. Chem. Biol.* **2017**, *13* (4), 446–450. https://doi.org/10.1038/nchembio.2312.
(94) Fan, C.; Ip, K.; Söll, D. Expanding the Genetic Code of Escherichia Coli. *FEBS Lett.* **2001**, 3040–3047. https://doi.org/10.1002/1873-3468.12333.
(95) Chin, J. W.; Cropp, T. A.; Chu, S.; Meggers, E.; Schultz, P. G. Progress Toward an Expanded Eukaryotic Genetic Code. *Chem. Biol.* **2003**, *10* (6), 511–519. https://doi.org/10.1016/S1074-5521(03)00123-6.
(96) Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. A Genetically Encoded Photocaged Amino Acid. *J. Am. Chem. Soc.* **2004**, *126* (44), 14306–14307. https://doi.org/10.1021/ja040175z.
(97) Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G. A Genetically Encoded Fluorescent Probe in Mammalian Cells. *J. Am. Chem. Soc.* **2013**, *135* (34), 12540–12543. https://doi.org/10.1021/ja4059553.
(98) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. An Expanded Eukaryotic Genetic Code. **2003**, *301* (August), 964–968.
(99) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically Encoding Nε-Acetyllysine in Recombinant Proteins. *Nat. Chem. Biol.* **2008**, *4* (4), 232–234. https://doi.org/10.1038/nchembio.73.
(100) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-TRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys.*

Acta - Proteins Proteomics **2014**, *1844* (6), 1059–1070. https://doi.org/10.1016/j.bbapap.2014.03.002.
(101) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. Adding L-Lysine Derivatives to the Genetic Code of Mammalian Cells with Engineered Pyrrolysyl-TRNA Synthetases. *Biochem. Biophys. Res. Commun.* **2008**, *371* (4), 818–822. https://doi.org/10.1016/j.bbrc.2008.04.164.
(102) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Crystallographic Studies on Multiple Conformational States of Active-Site Loops in Pyrrolysyl-TRNA Synthetase. *J. Mol. Biol.* **2008**, *378* (3), 634–652. https://doi.org/10.1016/j.jmb.2008.02.045.
(103) Kavran, J. M.; Gundllapalli, S.; Donoghue, P. O.; Englert, M.; So, D.; Steitz, T. A. Structure of Pyrrolysyl-TRNA Synthetase , an Archaeal Enzyme for Genetic Code Innovation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11268–11273.
(104) Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Araiso, Y.; Ishitani, R.; Umehara, T.; Söll, D.; Nureki, O. Pyrrolysyl-TRNA Synthetase-TRNAPyl Structure Reveals the Molecular Basis of Orthogonality. *Nature* **2009**, *457* (7233), 1163–1167. https://doi.org/10.1038/nature07611.
(105) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Bernhard, H.; Schultz, P. G. A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells. **2010**, *48* (22), 4052–4055. https://doi.org/10.1002/anie.200900683.A.
(106) Odoi, K. A.; Huang, Y.; Rezenom, Y. H.; Liu, W. R. Nonsense and Sense Suppression Abilities of Original and Derivative Methanosarcina Mazei Pyrrolysyl-TRNA Synthetase-TRNAPyl Pairs in the Escherichia Coli BL21(DE3) Cell Strain. *PLoS One* **2013**, *8* (3). https://doi.org/10.1371/journal.pone.0057035.
(107) Niu, W.; Schultz, P. G.; Guo, J. An Expanded Genetic Code in Mammalian Cells with a Functional Quadruplet Codon. *ACS Chem. Biol.* **2013**, *8* (7), 1640–1645. https://doi.org/10.1021/cb4001662.
(108) Italia, J. S.; Latour, C.; Wrobel, C. J. J.; Chatterjee, A. Resurrecting the Bacterial Tyrosyl-TRNA Synthetase/TRNA Pair for Expanding the Genetic Code of Both E. Coli and Eukaryotes. *Cell Chem. Biol.* **2018**, *25* (10), 1304-1312.e5. https://doi.org/10.1016/j.chembiol.2018.07.002.
(109) Cervettini, D.; Tang, S.; Fried, S. D.; Willis, J. C. W.; Funke, L. F. H.; Colwell, L. J.; Chin, J. W. Rapid Discovery and Evolution of Orthogonal Aminoacyl-TRNA Synthetase–TRNA Pairs. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0479-2.
(110) Charbon, G.; Wang, J.; Brustad, E.; Schultz, P. G.; Horwich, A. L.; Jacobs-Wagner, C.; Chapman, E. Localization of GroEL Determined by in Vivo Incorporation of a Fluorescent Amino Acid. *Bioorganic Med. Chem. Lett.* **2011**, *21* (20), 6067–6070. https://doi.org/10.1016/j.bmcl.2011.08.057.
(111) Charbon, G.; Brustad, E.; Scott, K. A.; Wang, J.; Løbner-Olesen, A.; Schultz, P. G.; Jacobs-Wagner, C.; Chapman, E. Subcellular Protein Localization by Using a Genetically Encoded Fluorescent Amino Acid. *ChemBioChem* **2011**, *12* (12), 1818–1821. https://doi.org/10.1002/cbic.201100282.
(112) Wang, J.; Xie, J.; Schultz, P. G. A Genetically Encoded Fluorescent Amino Acid. *J. Am. Chem. Soc.* **2006**, *128* (27), 8738–8739.

https://doi.org/10.1021/ja062666k.
(113) Hyun, S. L.; Guo, J.; Lemke, E. A.; Dimla, R. D.; Schultz, P. G. Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in Saccharomyces Cerevisiae. *J. Am. Chem. Soc.* **2009**, *131* (36), 12921–12923. https://doi.org/10.1021/ja904896s.
(114) Mitchell, A. L.; Addy, P. S.; Chin, M. A.; Chatterjee, A. A Unique Genetically Encoded FRET Pair in Mammalian Cells. *ChemBioChem* **2017**, *18* (6), 511–514. https://doi.org/10.1002/cbic.201600668.
(115) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. A Genetically Encoded Fluorescent Amino Acid. *Proc. Natl. Acad. Sci.* **2006**, *103* (26), 9785–9789.
(116) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114* (9), 4764–4806. https://doi.org/10.1021/cr400355w.
(117) Kim, C. H.; Axup, J. Y.; Schultz, P. G. Protein Conjugation with Genetically Encoded Unnatural Amino Acids. *Current Opinion in Chemical Biology*. 2013, pp 412–419. https://doi.org/10.1016/j.cbpa.2013.04.017.
(118) Nikic, I.; Kang, J. H.; Girona, G. E.; Aramburu, I. V.; Lemke, E. A. Labeling Proteins on Live Mammalian Cells Using Click Chemistry. *Nat. Protoc.* **2015**, *10* (5), 780–791. https://doi.org/10.1038/nprot.2015.045.
(119) Uttamapinant, C.; Howe, J. D.; Lang, K.; Beránek, V.; Davis, L.; Mahesh, M.; Barry, N. P.; Chin, J. W. Genetic Code Expansion Enables Live-Cell and Super-Resolution Imaging of Site-Specifically Labeled Cellular Proteins. *J. Am. Chem. Soc.* **2015**, *137* (14), 4602–4605. https://doi.org/10.1021/ja512838z.
(120) Peng, T.; Hang, H. C. Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* **2016**, *138* (43), 14423–14433. https://doi.org/10.1021/jacs.6b08733.
(121) Sun, S. B.; Schultz, P. G.; Kim, C. H. Therapeutic Applications of an Expanded Genetic Code. *ChemBioChem* **2014**, *15* (12), 1721–1729. https://doi.org/10.1002/cbic.201402154.
(122) Zheng, Y.; Yu, F.; Wu, Y.; Si, L.; Xu, H.; Zhang, C.; Xia, Q.; Xiao, S.; Wang, Q.; He, Q.; et al. Broadening the Versatility of Lentiviral Vectors as a Tool in Nucleic Acid Research via Genetic Code Expansion. *Nucleic Acids Res.* **2015**, *43* (11), e73. https://doi.org/10.1093/nar/gkv202.
(123) Yao, T.; Zhou, X.; Zhang, C.; Yu, X.; Tian, Z.; Zhang, L.; Zhou, D. Site-Specific PEGylated Adeno-Associated Viruses with Increased Serum Stability and Reduced Immunogenicity. **2017**. https://doi.org/10.3390/molecules22071155.
(124) Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. Control of Protein Phosphorylation with a Genetically Encoded Photocaged Amino Acid. *Nat. Chem. Biol.* **2007**, *3* (12), 769–772. https://doi.org/10.1038/nchembio.2007.44.
(125) Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W. Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells. *J. Am. Chem. Soc.* **2010**, *132* (12), 4086–4088. https://doi.org/10.1021/ja910688s.
(126) Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W. Photocontrol of Tyrosine

Phosphorylation in Mammalian Cells via Genetic Encoding of Photocaged Tyrosine. *J. Am. Chem. Soc.* **2012**, *134* (29), 11912–11915. https://doi.org/10.1021/ja3046958.
(127) Kang, J. Y.; Kawaguchi, D.; Coin, I.; Xiang, Z.; O'Leary, D. D. M.; Slesinger, P. A.; Wang, L. In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids. *Neuron* **2013**, *80* (2), 358–370. https://doi.org/10.1016/j.neuron.2013.08.016.
(128) Uprety, R.; Luo, J.; Liu, J.; Naro, Y.; Samanta, S.; Deiters, A. Genetic Encoding of Caged Cysteine and Caged Homocysteine in Bacterial and Mammalian Cells. *ChemBioChem* **2014**, *15* (12), 1793–1799. https://doi.org/10.1002/cbic.201400073.
(129) Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W. Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *J. Am. Chem. Soc.* **2014**, *136* (6), 2240–2243. https://doi.org/10.1021/ja412191m.
(130) Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. Genetically Encoded Optochemical Probes for Simultaneous Fluorescence Reporting and Light Activation of Protein Function with Two-Photon Excitation. *J. Am. Chem. Soc.* **2014**, *136* (44), 15551–15558. https://doi.org/10.1021/ja5055862.
(131) Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A. Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. *ChemBioChem* **2017**, *18* (14), 1442–1447. https://doi.org/10.1002/cbic.201700147.
(132) Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G. A Genetically Encoded Photocaged Tyrosine. *Angew. Chemie - Int. Ed.* **2006**, *45* (17), 2728–2731. https://doi.org/10.1002/anie.200600264.
(133) Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. Protein Photo-Cross-Linking in Mammalian Cells by Site-Specific Incorporation of a Photoreactive Amino Acid. *Nat. Methods* **2005**, *2* (3), 201–206. https://doi.org/10.1038/nmeth739.
(134) Hino, N.; Oyama, M.; Sato, A.; Mukai, T.; Iraha, F.; Hayashi, A.; Kozuka-Hata, H.; Yamamoto, T.; Yokoyama, S.; Sakamoto, K. Genetic Incorporation of a Photo-Crosslinkable Amino Acid Reveals Novel Protein Complexes with GRB2 in Mammalian Cells. *J. Mol. Biol.* **2011**, *406* (2), 343–353. https://doi.org/10.1016/j.jmb.2010.12.022.
(135) Lin, S.; Zhang, Z.; Xu, H.; Li, L.; Chen, S.; Li, J.; Hao, Z.; Chen, P. R. Site-Specific Incorporation of Photo-Cross-Linker and Bioorthogonal Amino Acids into Enteric Bacterial Pathogens. *J. Am. Chem. Soc.* **2011**, *133* (50), 20581–20587. https://doi.org/10.1021/ja209008w.
(136) Rannversson, H.; Andersen, J.; Sørensen, L.; Bang-Andersen, B.; Park, M.; Huber, T.; Sakmar, T. P.; Strømgaard, K. Genetically Encoded Photocrosslinkers Locate the High-Affinity Binding Site of Antidepressant Drugs in the Human Serotonin Transporter. *Nat. Commun.* **2016**, *7*. https://doi.org/10.1038/ncomms11261.
(137) Yang, Y.; Song, H.; He, D.; Zhang, S.; Dai, S.; Lin, S.; Meng, R.; Wang, C.; Chen, P.

R. Genetically Encoded Protein Photocrosslinker with a Transferable Mass Spectrometry-Identifiable Label. *Nat. Commun.* **2016**, *7* (1), 12299. https://doi.org/10.1038/ncomms12299.
(138) Zheng, Y.; Gilgenast, M. J.; Hauc, S.; Chatterjee, A. Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis. *ACS Chem. Biol.* **2018**, *13* (5), 1137–1141. https://doi.org/10.1021/acschembio.8b00021.
(139) Liu, C. C.; Schultz, P. G. Recombinant Expression of Selectively Sulfated Proteins in Escherichia Coli. *Nat. Biotechnol.* **2006**, *24* (11), 1436–1440. https://doi.org/10.1038/nbt1254.
(140) Neumann, H.; Hazen, J. L.; Weinstein, J.; Mehl, R. A.; Chin, J. W. Genetically Encoding Protein Oxidative Damage. *J. Am. Chem. Soc.* **2008**, *130* (12), 4028–4033. https://doi.org/10.1021/ja710100d.
(141) Venkat, S.; Sturges, J.; Stahman, A.; Gregory, C.; Gan, Q.; Fan, C. Genetically Incorporating Two Distinct Post-Translational Modifications into One Protein Simultaneously. *ACS Synth. Biol.* **2018**, *7* (2), 689–695. https://doi.org/10.1021/acssynbio.7b00408.
(142) Neumann, H.; Hancock, S. M.; Buning, R.; Routh, A.; Chapman, L.; Somers, J.; Owen-Hughes, T.; van Noort, J.; Rhodes, D.; Chin, J. W. A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Mol. Cell* **2009**, *36* (1), 153–163. https://doi.org/10.1016/j.molcel.2009.07.027.
(143) Kim, C. H.; Kang, M.; Kim, H. J.; Chatterjee, A.; Schultz, P. G. Site-Specific Incorporation of μ-N-Crotonyllysine into Histones. *Angew. Chemie - Int. Ed.* **2012**, *51* (29), 7246–7249. https://doi.org/10.1002/anie.201203349.
(144) Xiao, H.; Xuan, W.; Shao, S.; Liu, T.; Schultz, P. G. Genetic Incorporation of ε-N-2-Hydroxyisobutyryl-Lysine into Recombinant Histones. *ACS Chem. Biol.* **2015**, *10* (7), 1599–1603. https://doi.org/10.1021/cb501055h.
(145) Elsässer, S. J.; Ernst, R. J.; Walker, O. S.; Chin, J. W. Genetic Code Expansion in Stable Cell Lines Enables Encoded Chromatin Modification. *Nat. Methods* **2016**, *13* (2), 158–164. https://doi.org/10.1038/nmeth.3701.
(146) Park, H. S.; Hohn, M. J.; Umehara, T.; Guo, L.-T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Söll, D. Expanding the Genetic Code of Escherichia Coli with Phosphotyrosine. *FEBS Lett.* **2016**, *333* (August), 3040–3047. https://doi.org/10.1002/1873-3468.12333.
(147) Beránek, V.; Reinkemeier, C. D.; Zhang, M. S.; Liang, A. D.; Kym, G.; Chin, J. W. Genetically Encoded Protein Phosphorylation in Mammalian Cells. *Cell Chem. Biol.* **2018**, *25* (9), 1067-1074.e5. https://doi.org/10.1016/j.chembiol.2018.05.013.
(148) Italia, J. S.; Peeler, J. C.; Hillenbrand, C. M.; Latour, C.; Weerapana, E.; Chatterjee, A. Genetically Encoded Protein Sulfation in Mammalian Cells. *Nat. Chem. Biol.* **2020**, *16* (4), 379–382. https://doi.org/10.1038/s41589-020-0493-1.
(149) Nguyen, D. P.; Alai, M. M. G.; Virdee, S.; Chin, J. W. Genetically Directing ε-N, N-Dimethyl-l-Lysine in Recombinant Histones. *Chem. Biol.* **2010**, *17* (10), 1072–1076. https://doi.org/10.1016/j.chembiol.2010.07.013.

(150) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* **2010**, *464* (7287), 441–444. https://doi.org/10.1038/nature08817.
(151) Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. Catalyst-Free and Site-Specific One-Pot Dual-Labeling of a Protein Directed by Two Genetically Incorporated Noncanonical Amino Acids. *ChemBioChem* **2012**, *13* (10), 1405–1408. https://doi.org/10.1002/cbic.201200281.
(152) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia Coli. *Biochemistry* **2013**, *52* (10), 1828–1837. https://doi.org/10.1021/bi4000244.
(153) Xiao, H.; Chatterjee, A.; Choi, S.; Bajjuri, K. M.; Sinha, S. C.; Schultz, P. G. Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells. *Angew. Chemie* **2013**, *125* (52), 14330–14333. https://doi.org/10.1002/ange.201308137.
(154) Italia, J. S.; Addy, P. S.; Erickson, S. B.; Peeler, J. C.; Weerapana, E.; Chatterjee, A. Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J. Am. Chem. Soc.* **2019**, jacs.8b12954. https://doi.org/10.1021/jacs.8b12954.
(155) Lochrie, M. A.; Tatsuno, G. P.; Christie, B.; McDonnell, J. W.; Zhou, S.; Surosky, R.; Pierce, G. F.; Colosi, P. Mutations on the External Surfaces of Adeno-Associated Virus Type 2 Capsids That Affect Transduction and Neutralization. *J. Virol.* **2006**, *80* (2), 821–834. https://doi.org/10.1128/JVI.80.2.821-834.2006.
(156) Shi, W.; Arnold, G. S.; Bartlett, J. S. Insertional Mutagenesis of the Adeno-Associated VIrus Type 2 (AAV2) Capsid Gene and Generation of AAV2 Vectors Targeted to Alternative Cell-Surface Receptors. **2001**, *12*.
(157) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-TRNA Synthetase/TRNACUA Pair and Click Chemistry. *J. Am. Chem. Soc.* **2009**, *131* (25), 8720–8721. https://doi.org/10.1021/ja900553w.
(158) Atchison, R. W.; Casto, B. C.; Hammon, William, M. Adenovirus-Associated Defective Virus Particles. **1965**, *149* (3685), 754–756.
(159) Hoggan, M.; Blacklow, N.; Rowe, W. Studies of Small Dna Viruses Found in Various Adenovirus Preparations: Physical, Biological, and Immunological Characteristics. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *55* (6), 1467–1474.
(160) Samulski, R. J.; Berns, K. I.; Tan, M.; Muzyczka, N. Cloning of Adeno-Associated Virus into PBR322: Rescue of Intact Virus from the Recombinant Plasmid in Human Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79* (6 I), 2077–2081. https://doi.org/10.1073/pnas.79.6.2077.
(161) McLaughlin, S. K.; Collis, P.; Hermonat, P. L.; Muzyczka, N. Adeno-Associated Virus General Transduction Vectors: Analysis of Proviral Structures. *J. Virol.* **1988**, *62* (6), 1963–1973. https://doi.org/10.1128/jvi.62.6.1963-1973.1988.
(162) Matsushita, T.; Elliger, S.; Elliger, C.; Podsakoff, G.; Villarreal, L.; Kurtzman, G. J.; Iwaki, Y.; Colosi, P. Adeno-Associated Virus Vectors Can Be Efficiently

Produced without Helper Virus. *Gene Ther.* **1998**, *5* (7), 938–945. https://doi.org/10.1038/sj.gt.3300680.
(163) Grieger, J. C.; Samulski, R. J. *Adeno-Associated Virus Vectorology, Manufacturing, and Clinical Applications*, 1st ed.; Elsevier Inc., 2012; Vol. 507. https://doi.org/10.1016/B978-0-12-386509-0.00012-0.
(164) Ingham, K. C. Protein Precipitation with Polyethylene Glycol. *Methods Enzymol.* **1984**, *104* (C), 351–356. https://doi.org/10.1016/S0076-6879(84)04101-X.
(165) Burova, E.; Ioffe, E. Chromatographic Purification of Recombinant Adenoviral and Adeno-Associated Viral Vectors: Methods and Implications. *Gene Ther.* **2005**, *12*, S5–S17. https://doi.org/10.1038/sj.gt.3302611.
(166) Auricchio, A.; Hildinger, M.; O'Connor, E.; Gao, G. P.; Wilson, J. M. Isolation of Highly Infectious and Pure Adeno-Associated Virus Type 2 Vectors with a Single-Step Gravity-Flow Column. *Hum. Gene Ther.* **2001**, *12* (1), 71–76. https://doi.org/10.1089/104303401450988.
(167) Nass, S. A.; Mattingly, M. A.; Woodcock, D. A.; Burnham, B. L.; Ardinger, J. A.; Osmond, S. E.; Frederick, A. M.; Scaria, A.; Cheng, S. H.; O'Riordan, C. R. Universal Method for the Purification of Recombinant AAV Vectors of Differing Serotypes. *Mol. Ther. - Methods Clin. Dev.* **2018**, *9* (June), 33–46. https://doi.org/10.1016/j.omtm.2017.12.004.
(168) Wang, Q.; Lock, M.; Prongay, A. J.; Alvira, M. R.; Petkov, B.; Wilson, J. M. Identification of an Adeno-Associated Virus Binding Epitope for AVB Sepharose Affinity Resin. *Mol. Ther. - Methods Clin. Dev.* **2015**, *2*, 15040. https://doi.org/10.1038/mtm.2015.40.
(169) Kapp, T. G.; Rechenmacher, F.; Neubauer, S.; Maltsev, O. V.; Cavalcanti-Adam, E. A.; Zarka, R.; Reuning, U.; Notni, J.; Wester, H. J.; Mas-Moruno, C.; et al. A Comprehensive Evaluation of the Activity and Selectivity Profile of Ligands for RGD-Binding Integrins. *Sci. Rep.* **2017**, *7* (November 2016), 1–13. https://doi.org/10.1038/srep39805.
(170) Wei, F.; McConnell, K. I.; Yu, T. K.; Suh, J. Conjugation of Paclitaxel on Adeno-Associated Virus (AAV) Nanoparticles for Co-Delivery of Genes and Drugs. *Eur. J. Pharm. Sci.* **2012**, *46* (3), 167–172. https://doi.org/10.1016/j.ejps.2012.02.022.
(171) Girod, A.; Wobus, C. E.; Zádori, Z.; Ried, M.; Leike, K.; Tijssen, P.; Kleinschmidt, J. A.; Hallek, M. The VP1 Capsid Protein of Adeno-Associated Virus Type 2 Is Carrying a Phospholipase A2 Domain Required for Virus Infectivity. *J. Gen. Virol.* **2002**, *83* (5), 973–978. https://doi.org/10.1099/0022-1317-83-5-973.
(172) Lux, K.; Goerlitz, N.; Schlemminger, S.; Perabo, L.; Goldnau, D.; Endell, J.; Leike, K.; Kofler, D. M.; Finke, S.; Hallek, M.; et al. Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking. *J. Virol.* **2005**, *79* (18), 11776–11787. https://doi.org/10.1128/jvi.79.18.11776-11787.2005.
(173) Asokan, A.; Johnson, J. S.; Li, C.; Samulski, R. J. Bioluminescent Virion Shells: New Tools for Quantitation of AAV Vector Dynamics in Cells and Live Animals. *Gene Ther.* **2008**, *15* (24), 1618–1622. https://doi.org/10.1038/gt.2008.127.
(174) Muik, A.; Reul, J.; Friedel, T.; Muth, A.; Hartmann, K. P.; Schneider, I. C.; Münch,

R. C.; Buchholz, C. J. Covalent Coupling of High-Affinity Ligands to the Surface of Viral Vector Particles by Protein Trans-Splicing Mediates Cell Type-Specific Gene Transfer. *Biomaterials* **2017**, *144*, 84–94. https://doi.org/10.1016/j.biomaterials.2017.07.032.

(175) Thadani, N. N.; Yang, J.; Moyo, B.; Lee, C. M.; Chen, M. Y.; Bao, G.; Suh, J. Site-Specific Post-Translational Surface Modification of Adeno-Associated Virus Vectors Using Leucine Zippers. *ACS Synth. Biol.* **2020**, acssynbio.9b00341. https://doi.org/10.1021/acssynbio.9b00341.

(176) Snijder, J.; Van De Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-Mckenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (20), 7295–7299. https://doi.org/10.1021/ja502616y.

(177) Gunnoo, S. B.; Madder, A. Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17* (7), 529–553. https://doi.org/10.1002/cbic.201500667.

(178) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. - An Asian J.* **2009**, *4* (5), 630–640. https://doi.org/10.1002/asia.200800427.

(179) Pulicherla, N.; Kota, P.; Dokholyan, N. V.; Asokan, A. Intra- and Inter-Subunit Disulfide Bond Formation Is Nonessential in Adeno-Associated Viral Capsids. *PLoS One* **2012**, *7* (2), 1–6. https://doi.org/10.1371/journal.pone.0032163.

(180) Chandran, J. S.; Sharp, P. S.; Karyka, E.; Aves-Cruzeiro, J. M. da C.; Coldicott, I.; Castelli, L.; Hautbergue, G.; Collins, M. O.; Azzouz, M. Site Specific Modification of Adeno-Associated Virus Enables Both Fluorescent Imaging of Viral Particles and Characterization of the Capsid Interactome. *Sci. Rep.* **2017**, *7* (1), 14766. https://doi.org/10.1038/s41598-017-15255-2.

(181) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *124* (21), 6063–6076. https://doi.org/10.1021/ja017687n.

(182) Seo, J. W.; Ingham, E. S.; Mahakian, L.; Tumbale, S.; Wu, B.; Aghevlian, S.; Shams, S.; Baikoghli, M.; Jain, P.; Ding, X.; et al. Positron Emission Tomography Imaging of Novel AAV Capsids Maps Rapid Brain Accumulation. *Nat. Commun.* **2020**, *11* (1), 1–13. https://doi.org/10.1038/s41467-020-15818-4.

(183) Hauf, M.; Richter, F.; Schneider, T.; Faidt, T.; Martins, B. M.; Baumann, T.; Durkin, P.; Dobbek, H.; Jacobs, K.; Möglich, A.; et al. Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *ChemBioChem* **2017**, *18* (18), 1819–1823. https://doi.org/10.1002/cbic.201700327.

(184) Peeler, J. C.; Falco, J. A.; Kelemen, R. E.; Abo, M.; Chartier, B. V; Edinger, L. C.; Chen, J.; Chatterjee, A.; Weerapana, E. Generation of Recombinant Mammalian Selenoproteins through Genetic Code Expansion with Photocaged Selenocysteine. *ACS Chem. Biol.* **2020**, 1–6. https://doi.org/10.1021/acschembio.0c00147.

(185) Welegedara, A. P.; Adams, L. A.; Huber, T.; Graham, B.; Otting, G. Site-Specific Incorporation of Selenocysteine by Genetic Encoding as a Photocaged Unnatural Amino Acid. *Bioconjug. Chem.* **2018**, *29* (7), 2257–2264. https://doi.org/10.1021/acs.bioconjchem.8b00254.
(186) Hemphill, J.; Borchardt, E. K.; Brown, K.; Asokan, A.; Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J. Am. Chem. Soc.* **2015**, *137* (17), 5642–5645. https://doi.org/10.1021/ja512664v.
(187) Hemphill, J.; Chou, C.; Chin, J. W.; Deiters, A. Genetically Encoded Light-Activated Transcription for Spatiotemporal Control of Gene Expression and Gene Silencing in Mammalian Cells. *J. Am. Chem. Soc.* **2013**, *135* (36), 13433–13439. https://doi.org/10.1021/ja4051026.
(188) Gautier, A.; Deiters, A.; Chin, J. W. Light-Activated Kinases Enable Temporal Dissection of Signaling Networks in Living Cells. *J. Am. Chem. Soc.* **2011**, *133* (7), 2124–2127. https://doi.org/10.1021/ja1109979.
(189) Liaunardy-Jopeace, A.; Murton, B. L.; Mahesh, M.; Chin, J. W.; James, J. R. Encoding Optical Control in LCK Kinase to Quantitatively Investigate Its Activity in Live Cells. *Nat. Struct. Mol. Biol.* **2017**, *24* (12), 1155–1163. https://doi.org/10.1038/nsmb.3492.
(190) Luo, J.; Arbely, E.; Zhang, J.; Chou, C.; Uprety, R.; Chin, J. W.; Deiters, A. Genetically Encoded Optical Activation of DNA Recombination in Human Cells. *Chem. Commun.* **2016**, *52* (55), 8529–8532. https://doi.org/10.1039/c6cc03934k.
(191) Engelke, H.; Chou, C.; Uprety, R.; Jess, P.; Deiters, A. Control of Protein Function through Optochemical Translocation. *ACS Synth. Biol.* **2014**, *3* (10), 731–736. https://doi.org/10.1021/sb400192a.
(192) Davis, L.; Radman, I.; Goutou, A.; Tynan, A.; Baxter, K.; Xi, Z.; Chin, J. W.; Greiss, S. Optically Splitting Symmetric Neuron Pairs in C . Elegans. **2020**.
(193) Rahman, S. T.; Zhou, W.; Deiters, A.; Haugh, J. M. Optical Control of MAP Kinase Kinase 6 (MKK6) Reveals That It Has Divergent Roles in pro-Apoptotic and Anti-Proliferative Signaling. **2020**, *6*. https://doi.org/10.1074/jbc.RA119.012079.
(194) Zhou, W.; Hankinson, C. P.; Deiters, A. Optical Control of Cellular ATP Levels with a Photocaged Adenylate Kinase. *ChemBioChem* **2020**, 1–6. https://doi.org/10.1002/cbic.201900757.
(195) Liu, J.; Hemphill, J.; Samanta, S.; Tsang, M.; Deiters, A. Genetic Code Expansion in Zebrafish Embryos and Its Application to Optical Control of Cell Signaling. *J. Am. Chem. Soc.* **2017**, *139* (27), 9100–9103. https://doi.org/10.1021/jacs.7b02145.
(196) Courtney, T. M.; Deiters, A. Optical Control of Protein Phosphatase Function. *Nat. Commun.* **2019**, *10* (1), 1–10. https://doi.org/10.1038/s41467-019-12260-z.
(197) Luo, J.; Kong, M.; Liu, L.; Samanta, S.; Van Houten, B.; Deiters, A. Optical Control of DNA Helicase Function through Genetic Code Expansion. *ChemBioChem* **2017**, *18* (5), 466–469. https://doi.org/10.1002/cbic.201600624.
(198) Opie, S. R.; Warrington, K. H.; Agbandje-McKenna, M.; Zolotukhin, S.; Muzyczka, N. Identification of Amino Acid Residues in the Capsid Proteins of

Adeno-Associated Virus Type 2 That Contribute to Heparan Sulfate Proteoglycan Binding. *J. Virol.* **2003**, *77* (12), 6995–7006.
(199) Kern, A.; Schmidt, K.; Leder, C.; Müller, O. J.; Wobus, C. E.; Bettinger, K.; Von der Lieth, C. W.; King, J. A.; Kleinschmidt, J. A. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. *J. Virol.* **2003**, *77* (20), 11072–11081.
(200) Italia, J. S.; Zheng, Y.; Kelemen, R. E.; Erickson, S. B.; Addy, P. S.; Chatterjee, A. Expanding the Genetic Code of Mammalian Cells. *Biochem. Soc. Trans.* **2017**, *45* (2), 555–562. https://doi.org/10.1042/BST20160336.
(201) Dommerholt, J.; Van Rooijen, O.; Borrmann, A.; Guerra, C. F.; Bickelhaupt, F. M.; Van Delft, F. L. Highly Accelerated Inverse Electron-Demand Cycloaddition of Electron-Deficient Azides with Aliphatic Cyclooctynes. *Nat. Commun.* **2014**, *5*, 1–7. https://doi.org/10.1038/ncomms6378.
(202) Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A. Genetically Encoded Copper-Free Click Chemistry. *Angew. Chemie - Int. Ed.* **2011**, *50* (17), 3878–3881. https://doi.org/10.1002/anie.201008178.
(203) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. Genetic Encoding of Bicyclononynes and Trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels-Alder Reactions. *J. Am. Chem. Soc.* **2012**, *134* (25), 10317–10320. https://doi.org/10.1021/ja302832g.
(204) McCraw, D. M.; O'Donnell, J. K.; Taylor, K. A.; Stagg, S. M.; Chapman, M. S. Structure of Adeno-Associated Virus-2 in Complex with Neutralizing Monoclonal Antibody A20. *Virology* **2012**, *431* (1–2), 40–49. https://doi.org/10.1016/j.virol.2012.05.004.
(205) Ai, H.; Shen, W.; Sagi, A.; Chen, P. R.; Schultz, P. G. Probing Protein-Protein Interactions with a Genetically Encoded Photo-Crosslinking Amino Acid. *ChemBioChem* **2011**, *12* (12), 1854–1857. https://doi.org/10.1002/cbic.201100194.
(206) Hallam, T. J.; Wold, E.; Wahl, A.; Smider, V. V. Antibody Conjugates with Unnatural Amino Acids. *Mol. Pharm.* **2015**, *12* (6), 1848–1862. https://doi.org/10.1021/acs.molpharmaceut.5b00082.
(207) Agarwal, P.; Bertozzi, C. R. Site-Specific Antibody-Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjug. Chem.* **2015**, *26* (2), 176–192. https://doi.org/10.1021/bc5004982.
(208) Hutchins, B. M.; Kazane, S. A.; Staflin, K.; Forsyth, J. S.; Felding-Habermann, B.; Schultz, P. G.; Smider, V. V. Site-Specific Coupling and Sterically Controlled Formation of Multimeric Antibody Fab Fragments with Unnatural Amino Acids. *J. Mol. Biol.* **2011**, *406* (4), 595–603. https://doi.org/10.1016/j.jmb.2011.01.011.
(209) Kim, C. H.; Axup, J. Y.; Dubrovska, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G. Synthesis of Bispecific Antibodies Using Genetically Encoded Unnatural Amino Acids. *J. Am. Chem. Soc.* **2012**, *134* (24), 9918–9921. https://doi.org/10.1021/ja303904e.
(210) Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.;

Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; et al. Synthesis of Site-Specific Antibody-Drug Conjugates Using Unnatural Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (40), 16101–16106. https://doi.org/10.1073/pnas.1211023109.
(211) Lu, H.; Wang, D.; Kazane, S.; Javahishvili, T.; Tian, F.; Song, F.; Sellers, A.; Barnett, B.; Schultz, P. G. Site-Specific Antibody-Polymer Conjugates for SiRNA Delivery. *J. Am. Chem. Soc.* **2013**, *135* (37), 13885–13891. https://doi.org/10.1021/ja4059525.
(212) Kularatne, S. A.; Deshmukh, V.; Ma, J.; Tardif, V.; Lim, R. K. V.; Pugh, H. M.; Sun, Y.; Manibusan, A.; Sellers, A. J.; Barnett, R. S.; et al. A CXCR4-Targeted Site-Specific Antibody-Drug Conjugate. *Angew. Chemie - Int. Ed.* **2014**, *53* (44), 11863–11867. https://doi.org/10.1002/anie.201408103.
(213) Zimmerman, E. S.; Heibeck, T. H.; Gill, A.; Li, X.; Murray, C. J.; Madlansacay, M. R.; Tran, C.; Uter, N. T.; Yin, G.; Rivers, P. J.; et al. Production of Site-Specific Antibody-Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System. *Bioconjug. Chem.* **2014**, *25* (2), 351–361. https://doi.org/10.1021/bc400490z.
(214) Oza, J. P.; Aerni, H. R.; Pirman, N. L.; Barber, K. W.; Ter Haar, C. M.; Rogulina, S.; Amrofell, M. B.; Isaacs, F. J.; Rinehart, J.; Jewett, M. C. Robust Production of Recombinant Phosphoproteins Using Cell-Free Protein Synthesis. *Nat. Commun.* **2015**, *6*. https://doi.org/10.1038/ncomms9168.
(215) Dondapati, S. K.; Stech, M.; Zemella, A.; Kubick, S. Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. *BioDrugs* **2020**, *34* (3), 327–348. https://doi.org/10.1007/s40259-020-00417-y.
(216) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **2017**, *139* (34), 11670–11673. https://doi.org/10.1021/jacs.7b05125.
(217) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An Evolved Aminoacyl-TRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **2011**, *50* (11), 1894–1900. https://doi.org/10.1021/bi101929e.
(218) Kotterman, M. A.; Schaffer, D. V. Engineering Adeno-Associated Viruses for Clinical Gene Therapy. *Nat. Rev. Genet.* **2014**, *15* (7), 445–451. https://doi.org/10.1038/nrg3742.
(219) Pishesha, N.; Ingram, J. R.; Ploegh, H. L. Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol.* **2018**, *34* (1), 163–188. https://doi.org/10.1146/annurev-cellbio-100617-062527.
(220) Theile, C. S.; Witte, M. D.; Blom, A. E. M.; Kundrat, L.; Ploegh, H. L.; Guimaraes, C. P. Site-Specific N-Terminal Labeling of Proteins Using Sortase-Mediated Reactions. *Nat. Protoc.* **2013**, *8* (9), 1800–1807. https://doi.org/10.1038/nprot.2013.102.
(221) Chen, I.; Dorr, B. M.; Liu, D. R. A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11399–11404. https://doi.org/10.1073/pnas.1101046108.
(222) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; et al. HaloTag: A

Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3* (6), 373–382. https://doi.org/10.1021/cb800025k.
(223) England, C. G.; Luo, H.; Cai, W. HaloTag Technology: A Versatile Platform for Biomedical Applications. *Bioconjug. Chem.* **2015**, *26* (6), 975–986. https://doi.org/10.1021/acs.bioconjchem.5b00191.

2.0 PROBING THE EFFECTS OF SITE OF CHEMICAL RETARGETING ON ADENO-ASSOCIATED VIRUS

The three AAV-based gene therapy drugs - Glybera, Luxturna, and Zolgensma - each use a native serotype of AAV. In developing further gene therapies, using the native serotypes of AAV presents two main problems: these serotypes do not bind to receptors necessary to reach many desirable therapeutic targets, and the broad tropism of AAV leads to off-target affects. The ability to alter the cell-surface receptor that AAV binds to would expand its potential for use in gene therapy. The incorporation of a bioorthogonal handle into the AAV capsid followed by chemical conjugation to a targeting moiety is one of the methods that have been used to re-target AAV[87,88]. As UAA's have a small footprint, they are well tolerated in the capsid and can be incorporated site-specifically. Just as small-molecule drugs go through stringent optimization in medicinal chemistry, the development of a re-targeted AAV capsid for gene therapy should be similarly stringent. By systematically optimizing the site of the targeting moiety, I was able to demonstrate the importance of the placement of the moiety on the capsid, an important step to creating a functional virus-conjugate for gene therapy.

2.1 INTRODUCTION

For a more complete discussion on the methods that have previously been used to alter AAV's surface receptors, see chapter 1. This section focuses specifically on work that has been done to identify where the AAV capsid will tolerate modifications

2.1.1 An overview of previous work performed establishing where in the viral capsid modifications are tolerated

The AAV capsid is an intricately structured compilation of 60 capsid proteins. As such, modification of the capsid can have disastrous effects on capsid assembly or viral infectivity. Prior to the solving of AAV's crystal structure[17], several studies were performed that made various mutations in the capsid structure to identify regions of the capsid that were involved in the various steps of the AAV life-cycle[155,156]. As AAV has become more popular as a gene-therapy vector, the ability to retarget the AAV capsid has become a sought-after ability and other studies have probed the capsid for locations tolerant of peptide insertions[55,58,59,61,64,68,156]. Insertion of a peptide with a given targeting motif has become one of the most popular methods of retargeting AAV; however there are few sites in the capsid that are tolerant of such peptide insertions. Regions that generally accept peptide insertions include the VP2 N-terminal region; the highly surface exposed "spike" around residues 453-454; and the heparan-binding region, near residues R585 or R588. However, even at these tolerant sites, capsid stability and targeting capability

is still dependent on the sequence of the inserted peptide[156]. Besides UAA incorporation, peptide loop insertion is the method that allows for the most control over where a targeting motif is placed.

While VP1 and VP3 N-termini do not tolerate extensions, the VP2 capsid protein is non-essential and can be extended to accommodate targeting motifs or proteins[69]. Examples of this include the genetic fusion of receptor ligands, an anti-CD34 antibody, Darpins, affibodies, and a human FK-binding protein (FKBP) to the N-terminus of the VP2 capsid protein[68,70–74]. However, as it is thought that the VP1 and VP2 N-terminals reside on the inside of the capsid in the pore, additions to the VP2 N-terminals almost certainly destabilize the capsid integrity. A third method that has been used to attach a novel functionality to the AAV capsid has been to modify existing lysine or arginine residues on the capsid surface. For example, biotin and an anti EpCAM antibody have been conjugated to the capsid using an NHS-ester, and the capsid has been glycosylated using a α-dicarbonyl compound that modifies arginine residues[52–54]. Both the modification of the VP2 N-terminus and the modification of existing lysine and arginine residues completely lack the ability to control where the modification is being placed onto the capsid surface.

In contrast to the above-mentioned methods, UAA incorporation is one of the smallest modifications that can be made to a protein, so incorporation minimally affects the capsid folding. Additionally, UAAs can be site-specifically incorporated in place of another amino acid, allowing for precise placement of a new functional group of interest. UAA incorporation requires that plasmids harboring an evolved aminoacyl-tRNA synthetase/tRNA pair that is specific to the UAA of interest and is

orthogonal to the host machinery be transfected in with the AAV production plasmids. For a more in-depth discussion on UAA incorporation and its history, refer to chapter 1.3.

2.1.2: An overview of UAA incorporation into the AAV capsid

UAA incorporation into the AAV capsid is the most recent method that has been used to develop specialized AAV vectors, with the first papers appearing in 2016. This technology has been used to PEGylate AAV to increase serum stability and reduce immunogenicity[123], to track AAV through the viral infection pathway[31], to shield the AAV capsid using lipid-based cloaks, to create a vector that is light activated[90], and to create re-targeted capsids for the purpose of gene therapy[87,88]. This section focuses on UAA incorporation to create re-targeted capsids.

Incorporation of azide bearing lysine residues into the AAV capsid followed by conjugation to a cRGD moiety has thus far been performed in two cases[87,88]. Zhang et. al. incorporated N^{ε}-2-azidoethyloxycarbonyl-L-lysine (NAEK) using the NAEK-specific tRNA/aaRS pair[157]. NAEK was incorporated into 14 sites on the capsid surface and resulting AAV was quantified using qPCR (Chart 1). Infectivity was then assayed using constant titer infection. Based on the results from the qPCR and infectivity experiments, sites R447, G453, S578, N587, N587+1, and S662 were chosen to conjugate cRGD-DIBO to the azide moiety of NAEK. The resulting viruses were used to infect U87 cells, and the most effective of these viruses at retargeting was site 587+1.

Table 2.1 Sites in the AAV capsid that UAAs have been incorporated into for subsequent functionalization with a re-targeting moiety.

Site of Incorporation	Author	UAA incorporated	Approximate Titer as % of WT	Retargeting capability compared to WT
S261	Zhou	NAEK	5%	NT
D327	Chatterjee	AzK	25%	NT
N381	Zhou	NAEK	5%	NT
Y444	Zhou	NAEK	5%	NT
R447	Zhou	NAEK	50%	No Improvement
G453	Chatterjee, Zhou	AzK, NAEK	15%, 20%	Small decrease
T454	Chatterjee	AzK	22%	NT
S458	Zhou	NAEK	5%	NT
S492	Zhou	NAEK	5%	NT
Y500	Zhou	NAEK	5%	NT
T573	Zhou	NAEK	5%	NT
S578	Zhou	NAEK	70%	Big improvement
N587	Chatterjee, Zhou	AzK, NAEK	10%, 20%	Big decrease
N587+1	Zhou	NAEK	20%	Big improvement
N588	Chatterjee	AzK	15%	NT
S662	Zhou	NAEK	70%	Big decrease

Kelemen et. al. developed a method of incorporating the UAA Azido-Lysine (AzK) into the AAV capsid using the *Methanosarcina barker* pyrrolysyl-tRNA-synthetase (*Mb*PylRS) and *Methanosarcina mazei* pyrrolysyl $tRNA_{CUA}$ ($tRNA^{MmPyl}_{CUA}$). Kelemen et. al. incorporated AzK into the AAV capsid at representative positions D327, G453, T454, N587, and R588 (Figure 1). D327 was chosen for its proximity to the 5-fold pore, G453 and T454 were chosen because they are highly surface exposed, and N587and R588 were chosen because they are in the HSPG binding site.

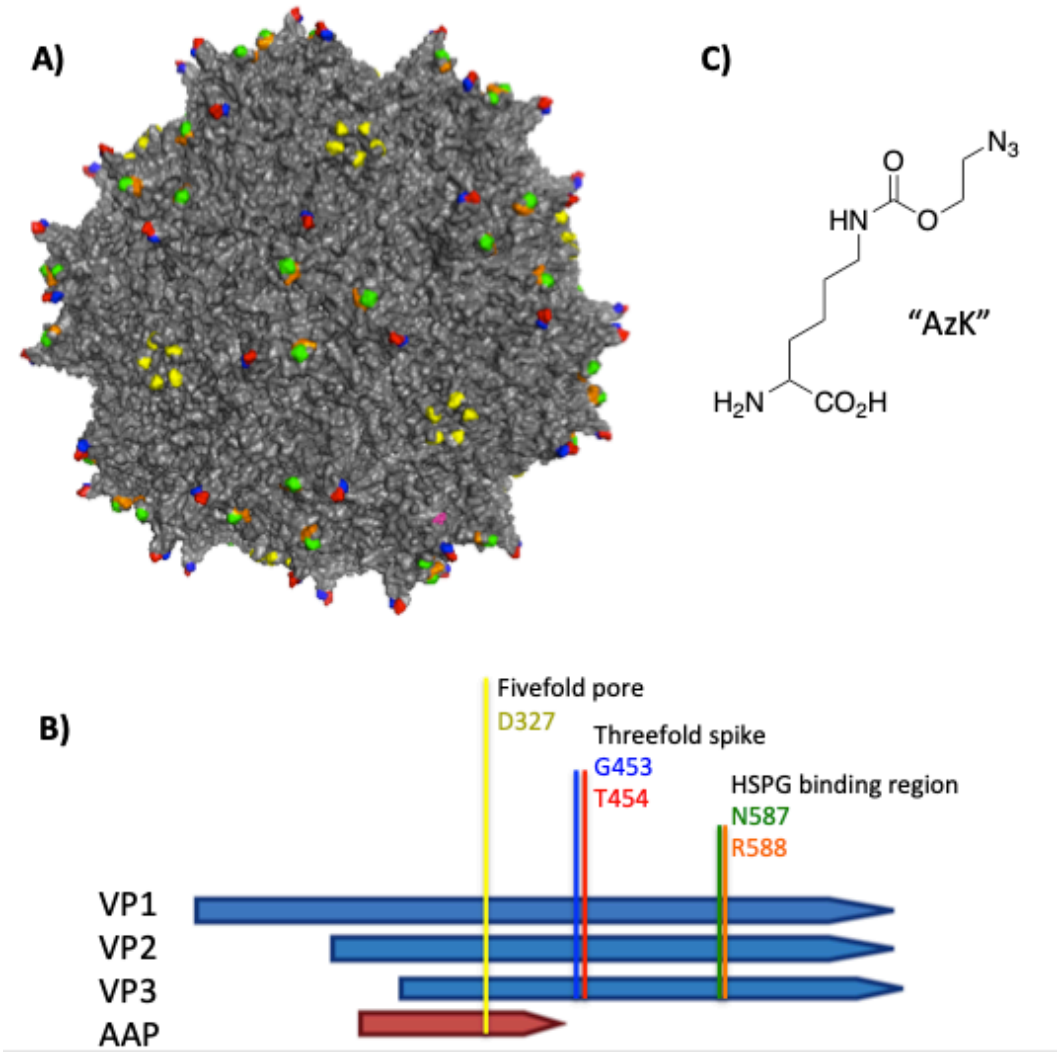


Figure 2.1. AzK incorporation into the AAV2 capsid as performed by Kelemen et. al. A) The sites that were chosen for AzK incorporation. D327 (yellow) was chosen because it sits at the five-fold pore, G453 and T454 (blue and red) were chosen because they sit on the threefold spike and are the two most surface exposed residues. N587 and R588 (green and orange) were chosen because they are in the HSPG binding region. Previous studies have shown that the threefold spike and the HSPG binding region tolerate modifications. B) The location of each of the stop codon insertions in the *cap* gene. C) The structure of AzK.

AzK was incorporated into each of the se sites and the resulting capsids maintained 10-25% of wild-type production, as quantified by qPCR (Figure 2A, B). HEK293T cell infection with constant multiplicity of infection (MOI) showed that virus that had AzK incorporated maintained wild type-like infectivity; with the exception of R588AzK as the HSPG binding interaction was ablated (Figure 2C). The

presence of AzK in VP1, VP2, and VP3 was confirmed by conjugating a fluorescent DBCO-cy5 probe to the capsid and viewing fluorescence by SDS-Page (Figure 2D, E).

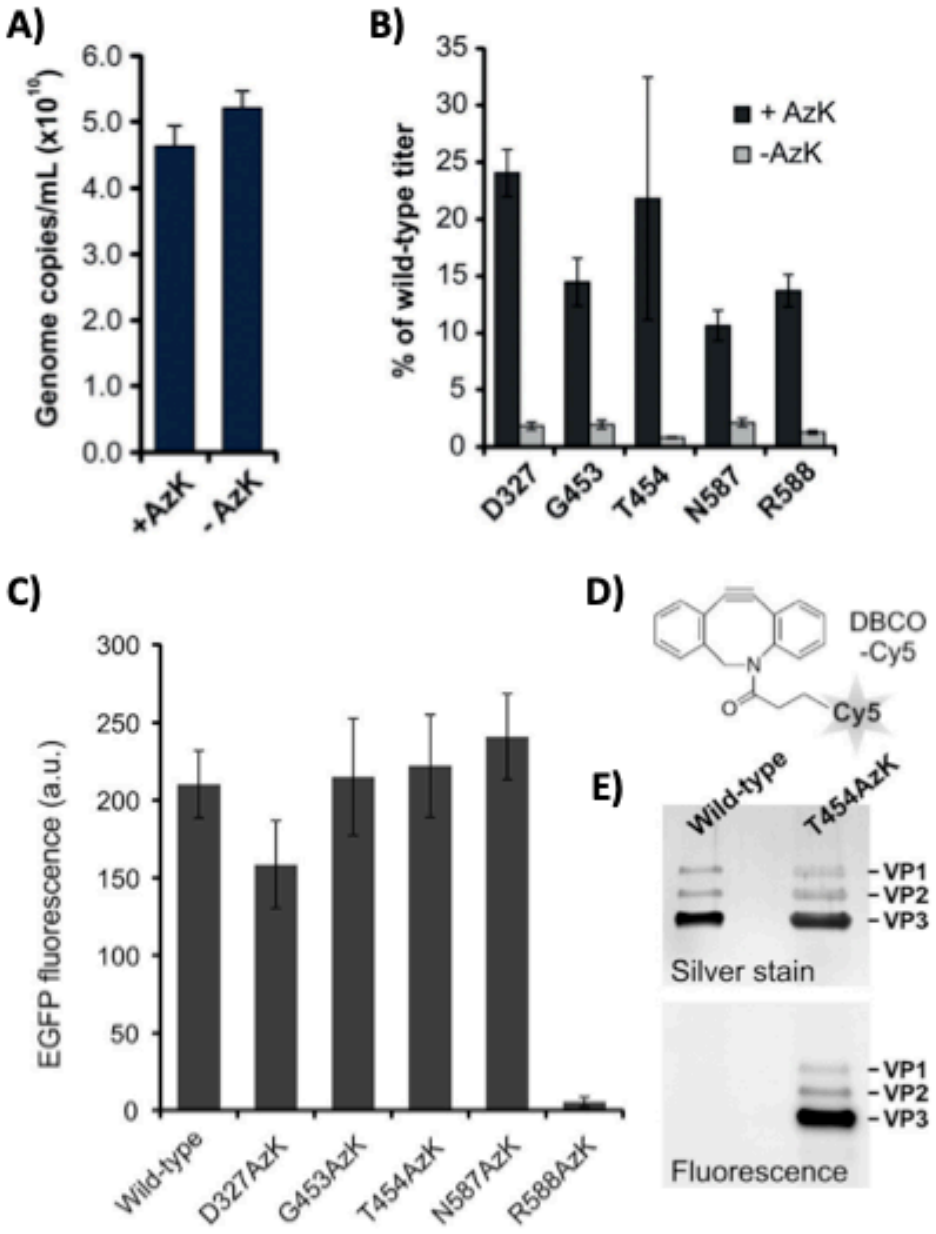


Figure 2.2. Characterization of AzK containing virus performed by R. Kelemen. A) Titers of wild type AAV made in the presence or absence of AzK as determined by qPCR. B) The production of site-specific UAA viruses as determined by qPCR made in the presence or absence of AzK. C) Infectivity of wild type or AzK viruses determined by EGFP fluorescence in the lysate of cells infected with a constant MOI of each virus. D) Structure of DBCO-Cy5. E) Silver stain and corresponding Cy5-fluorescence imaging analysis of SDS-PAGE resolved protein of purified wild type or T454AzK AAV2 following treatment with 5μM DBCO-Cy5. These images are taken with permission from reference [88]**. © 2017 Wiley-VCH Verlag GmbH &Co. KGaA, Weinheim**

The retargeting capability of these engineered capsids was tested by conjugating a DCBO-cRGDFC moiety, which binds to the $\alpha_v\beta_3$ integrin receptor that is overexpressed in certain types of cancer cells, onto the T454AzK and R588AzK in the capsid surface (Figure 3A). The resulting virus was used to infect SK-OV-3 cells (high $\alpha_v\beta_3$ expression) and HEK293T cells (low $\alpha_v\beta_3$ expression). While T454-cRGDFC AAV failed to infect either HEK293T cells or SK-OV-3 cells, R588-cRGDFC AAV lost infectivity towards HEK293T cells but showed about 80% of wild-type infectivity in SK-OV-3 cells; this experiment demonstrates that this strategy is an effective way to produce re-targeted AAV capsids (Figure 3B, C). Additionally, Kelemen et. al. confirmed that the R588-cRGDFC AAV was entering the cell by an $\alpha_V\beta_3$ dependent mechanism by repeating the infection using either free RGDS peptide, which would block binding to the $\alpha_V\beta_3$ integrin receptor, or free heparin sulfate, which would block viral HSPG binding. In these conditions, the modified virus was inhibited by free RGDS peptide but not by free heparin sulfate (Figure 3D).

This chapter is an extension of the work performed by Kelemen et. al. Using the platform created, this work explores the effect that the precise placement of the cRGDFC motif has on retargeting capability.

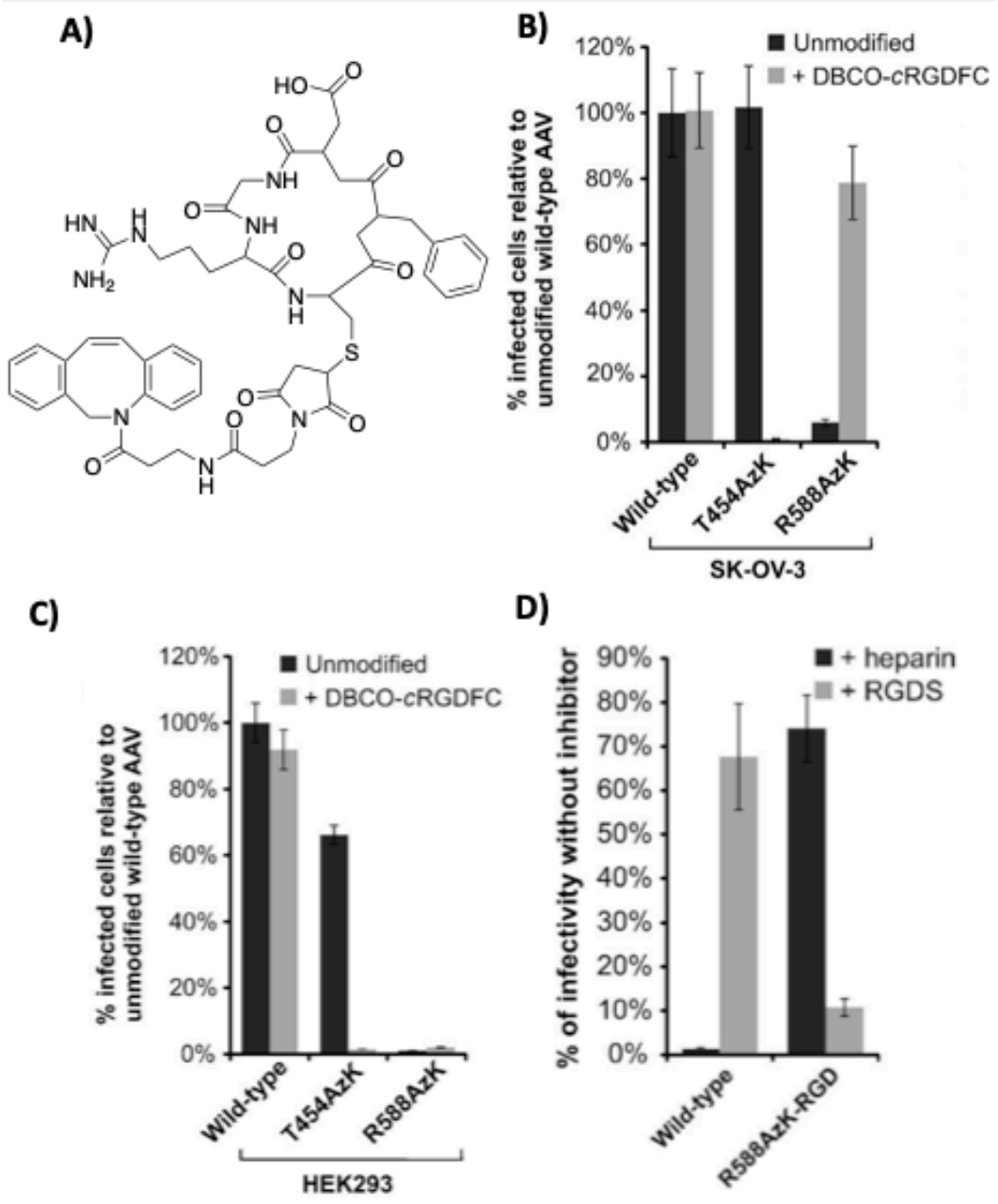


Figure 2.3 Functionalization of AzK-AAV with DBCO-cRGDFC and its retargeting efficiency. A) The Structure of DBCO-cRGDFC. B, C) Infectivity of wild type, T454AzK, and R588AzK AAV towards B) SK-OV-3 and C) HEK293T cells before and after functionalization with A). D) Infectivity of wild type and DBCO-cRGDFC-modified R588AzK AAV2 towards SK-OV-3 cells in the presence or either RGDS peptide or heparin. These images are taken with permission from reference 88[88]**. © 2017 Wiley-VCH Verlag GmbH &Co. KGaA, Weinheim**

2.1.3 AAV production

AAV was discovered as a contaminant of Adenovirus preparations[158,159]. It is replication deficient on its own and requires the aid of a "helper virus", Adenovirus

or Herpes Simplex Virus (HSV). Production of AAV was first achieved using infection with Adenovirus; however, this method resulted in Adenovirus production that could not be completely purified out of AAV stocks[160,161]. The essential Ad genes for AAV production were determined to be the E1a, E1b, E2a, E4orf6, and VA RNA genes, and a system was developed to produce AAV using co-transfection with a plasmid composed of Ad helper genes in order to avoid production of a secondary virus[162].

The AAV genome is comprised of the *Rep, Cap,* and *AAP* genes flanked by Inverted Terminal Repeats (ITRs). As the ITRs are the only cis-acting elements necessary for genome replication, integration, and packaging into the capsid, it is possible to express the *Rep, Cap,* and *AAP* genes in trans, while packaging the capsid with a gene of interest, such as a fluorescent reporter or a therapeutic protein (Figure 4A)[163]. Production of AAV is typically done using a triple-plasmid transfection, with Ad-helper genes on one plasmid, the AAV-ITR encased genome on a second plasmid, and the *Rep* and *Cap* genes on a third plasmid (Figure 4B).

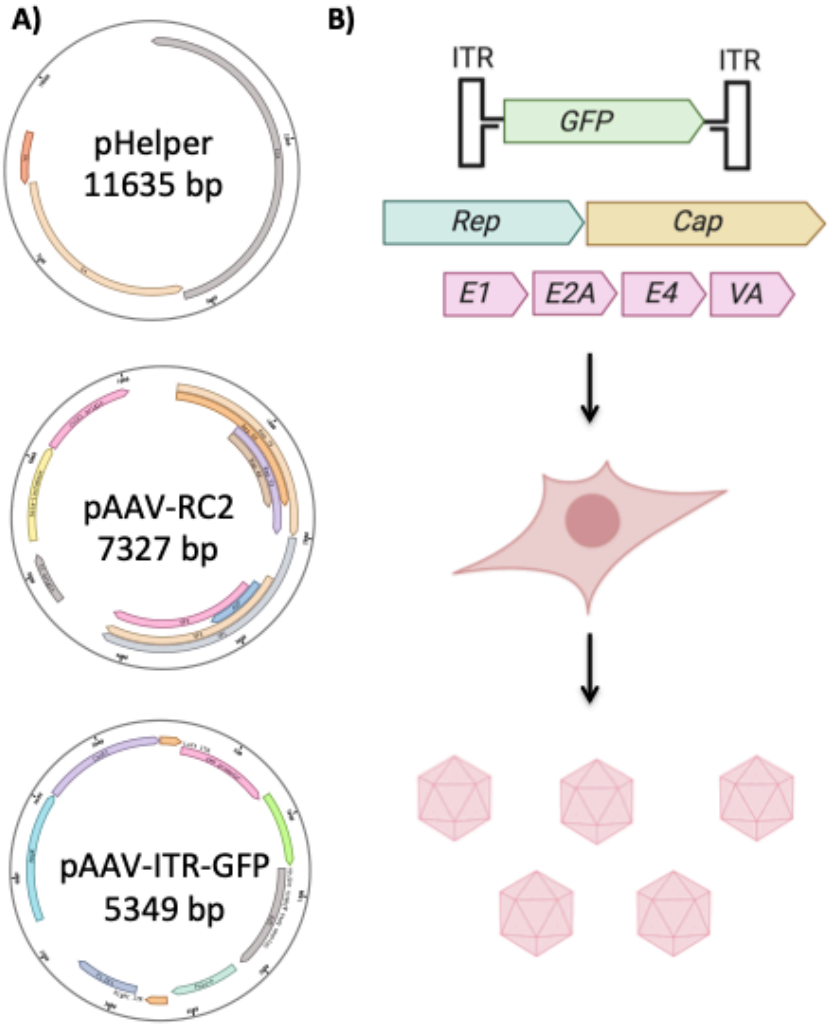


Figure 2.4. Plasmids to produce recombinant wild type AAV. A) An example of the plasmids that would be used to make wild type AAV2 by transfection in HEK293T cells. B) AAV2 is made by providing HEK293T cells with the necessary AAV2 genes by transfection. Necessary genes include the AAV genome packaged in ITRs, *Rep* and *Cap* genes, and the helper genes from adenovirus.

UAA incorporation into the AAV capsids requires a specifically evolved synthetase/tRNA pair in addition to the components required to make AAV. AzK-bearing AAV was made using a triple plasmid transfection (Figure 5A, B). pHelper contained all of the helper genes from Adenovirus necessary to make AAV, pIDTSmart-RC2-MbPylRS contained the *Mb*PylRS driven by a CMV promoter and the viral *Rep* and *Cap* genes, and pIDTSmart-8xPytR-ITR-GFP-ITR contained 8 copies of the $tRNA^{MmPyl}_{CUA}$ each driven by a U6 promoter and EGFP packed inside the AAV

ITRs to act as a florescent reporter. Transient transfection in HEK293T was performed with PEI and AAV was harvested by lysing the cells 3 days post transfection. Post-harvesting, infectivity was assayed using crude lysate; however further purification would be required for most applications.

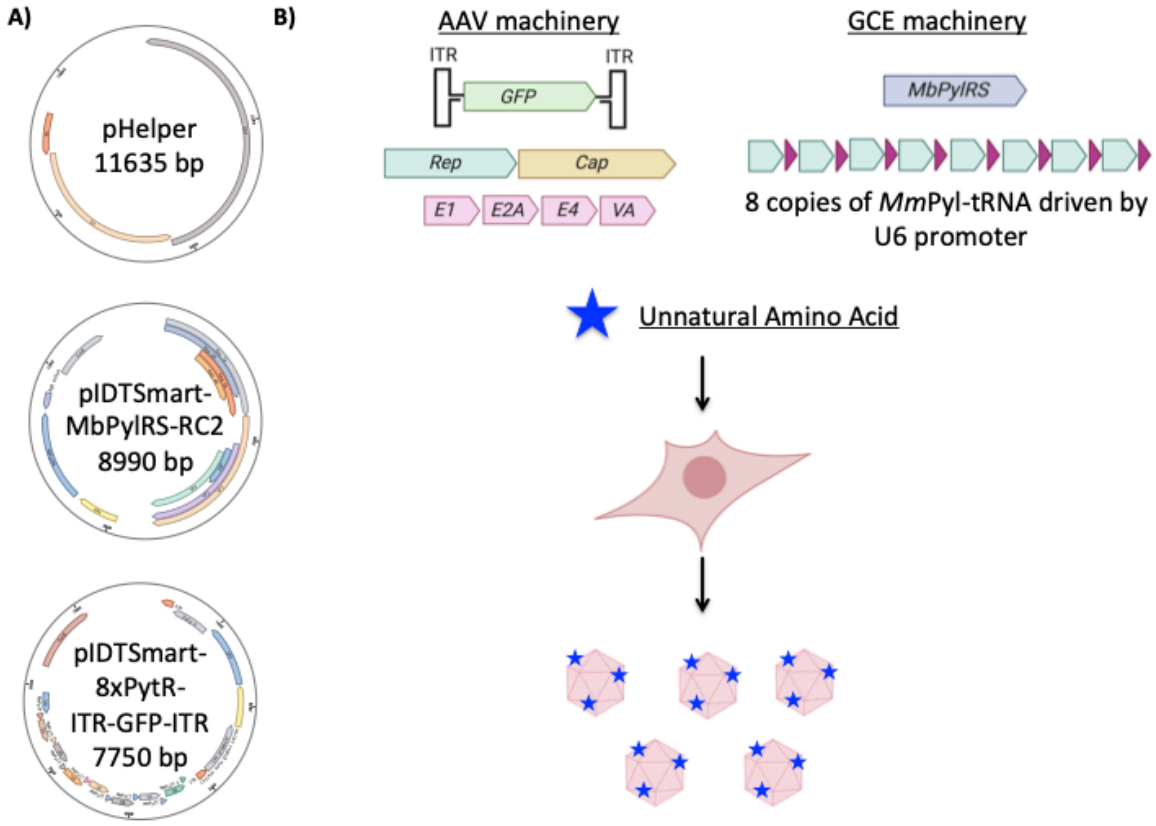


Figure 2.5. Plasmids to produce AAV2 with UAA incorporation. A) The plasmid system created by R. Kelemen to manufacture UAA-containing AAV2 capsids. B) Making UAA-containing AAV requires that HEK293T cells be transfected with genes to make AAV as well as genes for UAA incorporation and that the UAA being incorporated be supplied.

2.1.4 AAV purification

Because AAV infectivity can be assessed using crude lysate on HEK 293T cells using a fluorescent reporter, much of Kelemen et. al.'s work was done using semi-purified AAV. Small-scale batches of AAV were harvested using the commercially available AAVpro extraction solution by TAKARA, and large-scale batches of AAV

were semi-purified by Polyethylene Glycol (PEG) precipitation. PEG is a non-denaturing, water-soluble polymer that will precipitate protein without denaturing or otherwise interacting with proteins and is often used for the concentration of large complexes[164].

There are a variety of methods that can be used to produce pure rAAV preparations such as ultracentrifugation, affinity purification, ion exchange chromatography, or analytical HPLC[165]. Most of these methods produce preparations of rAAV with a certain amount of contamination and must be combined with a secondary purification step to produce rAAV that can be used in animals. Ultracentrifugation in CsCl is one of the most popular methods as it is a simple technique that yields highly pure rAAV samples. Disadvantages of CsCl ultracentrifugation include the need for extensive dialysis to remove toxic CsCl, and the tendency for rAAV particles to aggregate in CsCl, leading to a higher number of inactive capsids. An iodixanol gradient can also be used to purify rAAV, which is less toxic and prevents the aggregation of rAAV particles. However, while two rounds of CsCl ultracentrifugation will produce highly pure virus preparations, it takes a different secondary method (such as HPLC) to produce highly pure rAAV preps when using iodixanol purification. HPLC techniques include ion-exchange chromatography and HPLC using a heparin column. Purification by HPLC is most often used following iodixanol ultracentrifugation or with at least two HPLC steps.

Both ultracentrifugation and HPLC require the use of special equipment; however there are several methods that do not require the use of any special equipment. Affinity purification methods are available which produce highly pure

rAAV, but most affinity purification resins are serotype-specific as they are based on the capsid's binding patterns. For example, purification of AAV2 can be achieved using a single-step heparin agarose gravity column, but this technique cannot be used for any serotype that does not bind the HSPG motif[166]. Another disadvantage to heparin agarose is that there are several cellular proteins that bind to heparin and co-purify with the AAV. AVB resin is an affinity resin that is based on single-domain antibody fragments from the Camelidae family that will purify multiple serotypes of AAV with varying degrees of success[167,168]. AVB resin binds well to AAV1, AAV2, AAV5, and AAV6, but does not bind to AAV9 for example. An additional downside to affinity purification is that it does not discriminate against packaged vs. empty capsids, so both are present in the preparations.

As our lab is not equipped with an ultracentrifuge, or the equipment to perform the subsequent HPLC purifications, we used affinity purifications by gravity flow. All purifications performed by R. Kelemen were done using heparin affinity agarose; however because we began to work with AAV that did not bind to heparin, we also optimized a protocol for purification using gravity flow with AVB resin.

2.1.5 AAV titering for packaging and infectivity

AAV titer can be determined using quantitative PCR (qPCR). qPCR is a PCR reaction that uses a fluorescent reporter to monitor the amplification of a targeted DNA molecule in real time. Virus titers can be calculated using a series of positive controls that also go through the PCR reaction. There are two main methods of qPCR; the first is to use a non-specific dye that binds specifically to double stranded

DNA (most commonly SYBR Green), or to use a specific oligonucleotide that has both a quencher and a reporter that permits detection after the PCR reaction with a specific Taq DNA polymerase. The second method is often referred to as “TaqMan”. Because SYBR Green is a non-specific dye that binds by intercalation, SYBR Green can have more background fluorescence, but if primers are designed correctly, either method can produce high quality qPCR.

In either case, the general preparation to do qPCR on AAV remains the same. The first step in preparing AAV samples for qPCR is to incubate the virus sample with DNase I. DNase I degrades any DNA other than viral DNA, which is protected by the AAV capsid. The DNAse I is inactivated by heat and the viral capsids are lysed using a combination of heat and a lysis buffer to release viral DNA. Samples are then diluted to the optimal concentration for a PCR reaction and mixed into a master-mix that contains all necessary components for the qPCR reaction. A standard curve is created using positive controls of known concentrations that the qPCR primers are specific for. Using the standard curve and the amplification data for each sample, it is possible to calculate the titer of each sample. qPCR data obtained by Kelemen was performed using the commercially available AAVPro Titration kit made by Takara. While details of this kit are proprietary, we know that this kit uses SYBR green to detect amplification, and that the primers sit in the ITRs. I also optimized another qPCR protocol without the use of a proprietary kit. This method also used SYBR green, with primers that sat in the CMV (Primer sequences can be found in the experimental procedures).

2.2 RESULTS AND DISCUSSION

2.2.1 Capsid scanning

In order to further probe the ability of the AAV capsid to accept UAAs, a panel of surface-exposed sites on the AAV capsid was chosen for UAA incorporation (Figure 6). Ten of these sites (comprising residues 450-459) make up the capsid "spike", which is a highly surface-exposed area. Six of these residues (H271, D463, D494, T503, H509, and E530) are found in the capsid "dead zone[155]", which is an area adjacent to the HSPG binding site. In alanine scanning experiments, alanine mutants in the dead zone have been shown to be non-infectious, while maintaining the ability to bind to HSPG. An additional six sites were chosen for UAA incorporation from around the capsid for their surface exposed nature (Q263, T330, D327, T659, K706, K717).

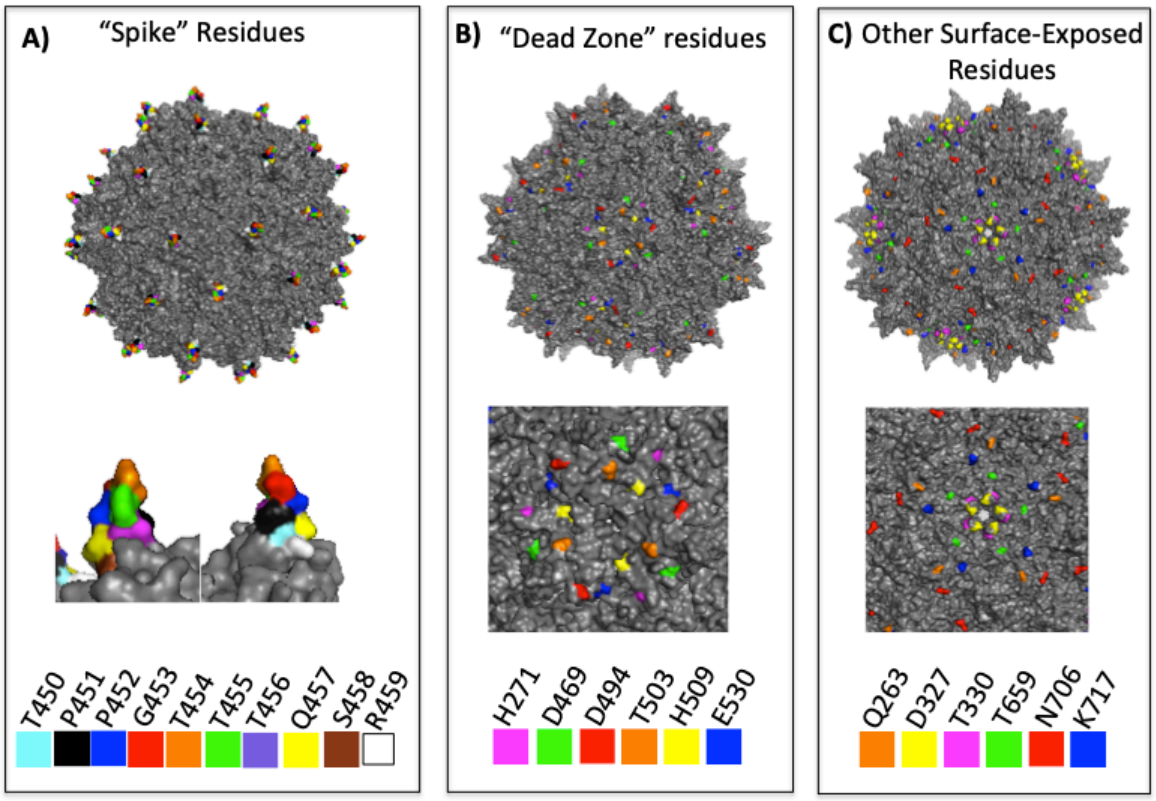


Figure 2.6. Sites chosen for UAA incorporation in the AAV2 capsid. A), B), C) Sites chosen for UAA incorporation in the AAV2 capsid were chosen from three different regions, A) the capsid "spike", B) the "Dead zone", or C) other surface exposed residues on the capsid surface.

While the genetic components used to make these viruses were the same as that used by Kelemen et. al., a new set of plasmids was generated that separated the synthetase and RC2 AAV genes, to facilitate the incorporation of different UAAs in the future. pHelper was kept the same, but *Mb*PylRS was placed onto a plasmid containing the EGFP reporter packaged in the AAV ITRs and the 8x copies of $tRNA^{MmPyl}_{CUA}$, and RC2 was left on its own plasmid (Figure 7). Using this system, the proper synthetase did not have to be cloned into every plasmid containing a stop codon at a different location in the AAV capsid. A TAG stop codon was placed at each of these above mentioned sites using overlap PCR and BocK was incorporated to assess infectivity of resulting capsids using the *Mb*PylRS/$tRNA^{MmPyl}_{CUA}$ pair[88].

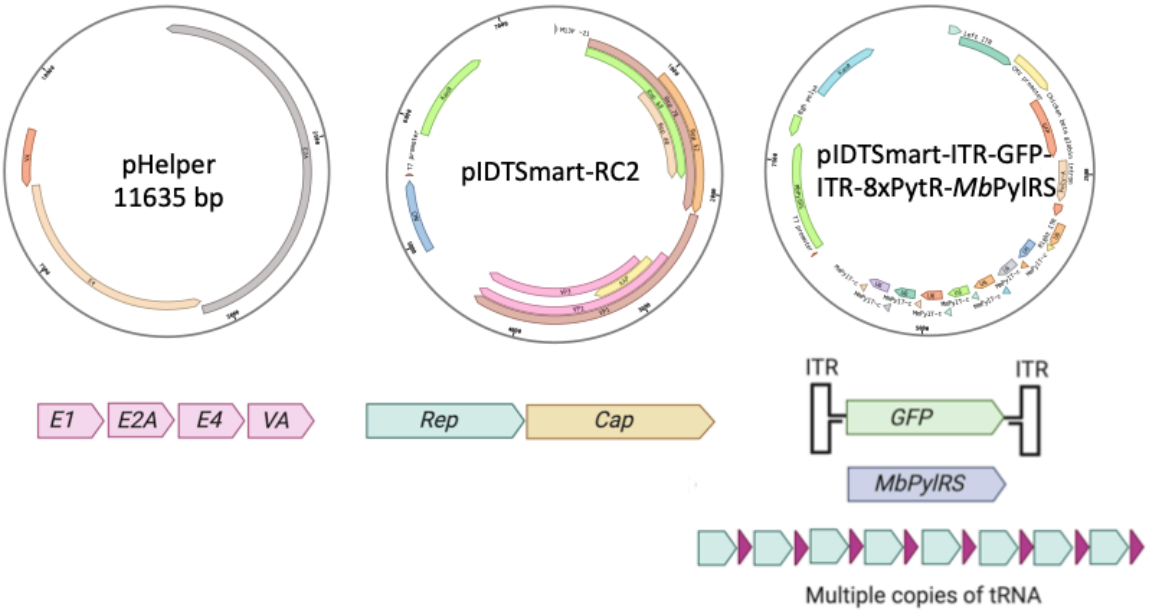


Figure 2.7. Updated plasmid system to produce AAV2 with UAA incorporation. pHelper contains the genes from Adenovirus that are essential to making AAV. pIDTSmart-RC2 contains the AAV *Rep* and *Cap* genes. pIDTSmart-ITR-GFP-ITR-8xPytR-*Mb*PylRS contains the EGFP reporter enclosed in the AAV ITRs, the *Mb*PylRS driven by the CMV promoter, and 8 copies of the tRNA^{MmPyl}.

All of the sites that make up the capsid spike were infectious with the exception of site P451TAG. As prolines are often important structural residues, it makes sense that altering this residue would have negative consequences. Similar to the alanine scanning results, with the exception of residue D469, each of the capsids with a mutated site in the 'dead zone' was rendered non-infectious. The remaining sites displayed varying degrees of infectivity. The "Other" residues displayed varying levels of infectivity (Figure 8).

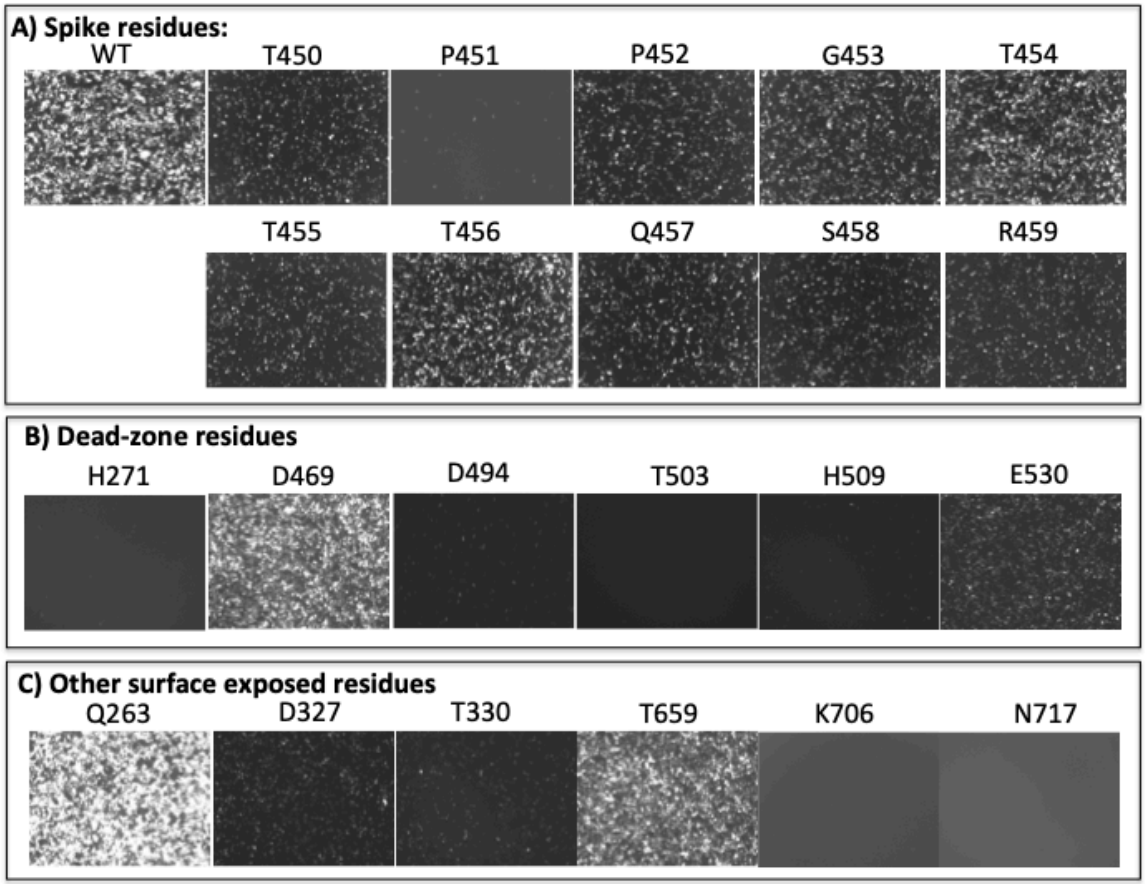


Figure 2.8. Constant volume pictures of UAA incorporation into the AAV capsid in A) the "Spike" residues, B) the "Dead-zone" residues, or C) the other surface exposed residues that were chosen.

2.2.2 Assaying the infectivity of the spike mutants containing AzK

Due to the highly surface exposed nature of the spike residues as well as the fact that almost all of these residues displayed infectivity with UAA incorporation, we decided to probe these sites for their ability to re-target AAV. AzK was incorporated into each of these sites and resulting virus titers were determined by qPCR (Figure 9A). Infectivity was assessed by infection of HEK 293T cells at a constant MOI of 50gc/cell, using EGFP in the viral cargo as a reporter. Infectivity was measured 48 hours post infection by EGFP fluorescence that was present in the lysate and microscopy (Figure 9B, C).

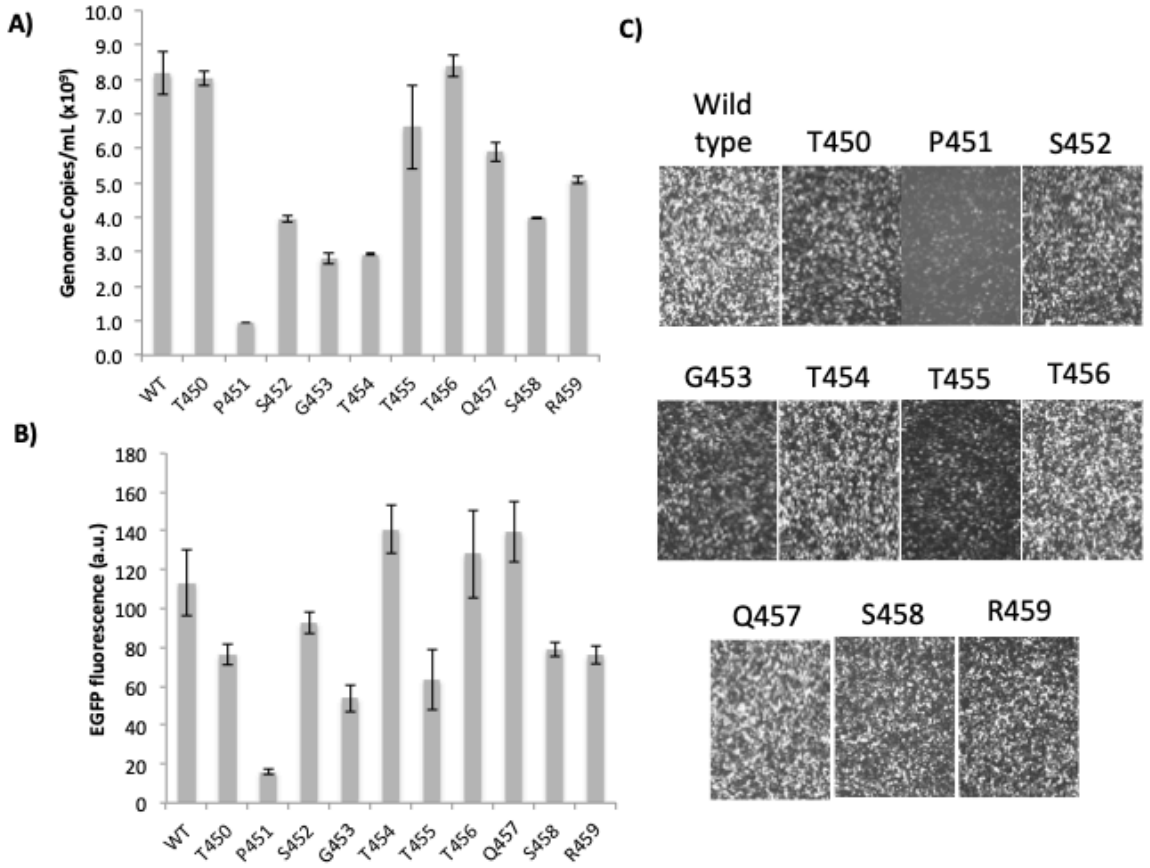


Figure 2.9. Production and infection efficiency of the spike mutants containing AzK. A) qPCR data showing production efficiency of AAV containing AzK. B) Constant titer infection of 50gc/cell on HEK293T cells showing infectivity of AAV from A. Analyzed using EGFP fluorescence of lysed cells. C) Fluorescence images corresponding to B.

2.2.3 cRGDFC synthesis

The RGD peptide was chosen as a retargeting motif by Kelemen et. al. because it has been used as a targeting motif in numerous studies that showed some degree of success. Additionally, because there are so many unknowns in the AAV infection cycle, Kelemen et. al. did not want to choose a motif that would make a primary interaction but prohibit necessary downstream interactions from being made. At this point, AAVR had not been discovered, so moving into an untested cell

line carried a lot of risk. Cyclic RGD was chosen because it's more stable and it has been shown that cyclic peptides bind the $\alpha_V\beta_3$ receptor with much higher affinity[169].

cRGDFC-DBCO was synthesized as previously described[88] using two commercially available starting materials. The cysteine residue on cRGDFC was conjugated to DBCO-maleimide. The reaction progress was monitored using LCMS and the product's reactivity was confirmed by labeling sfGFP-151-AzK (Figure 10).

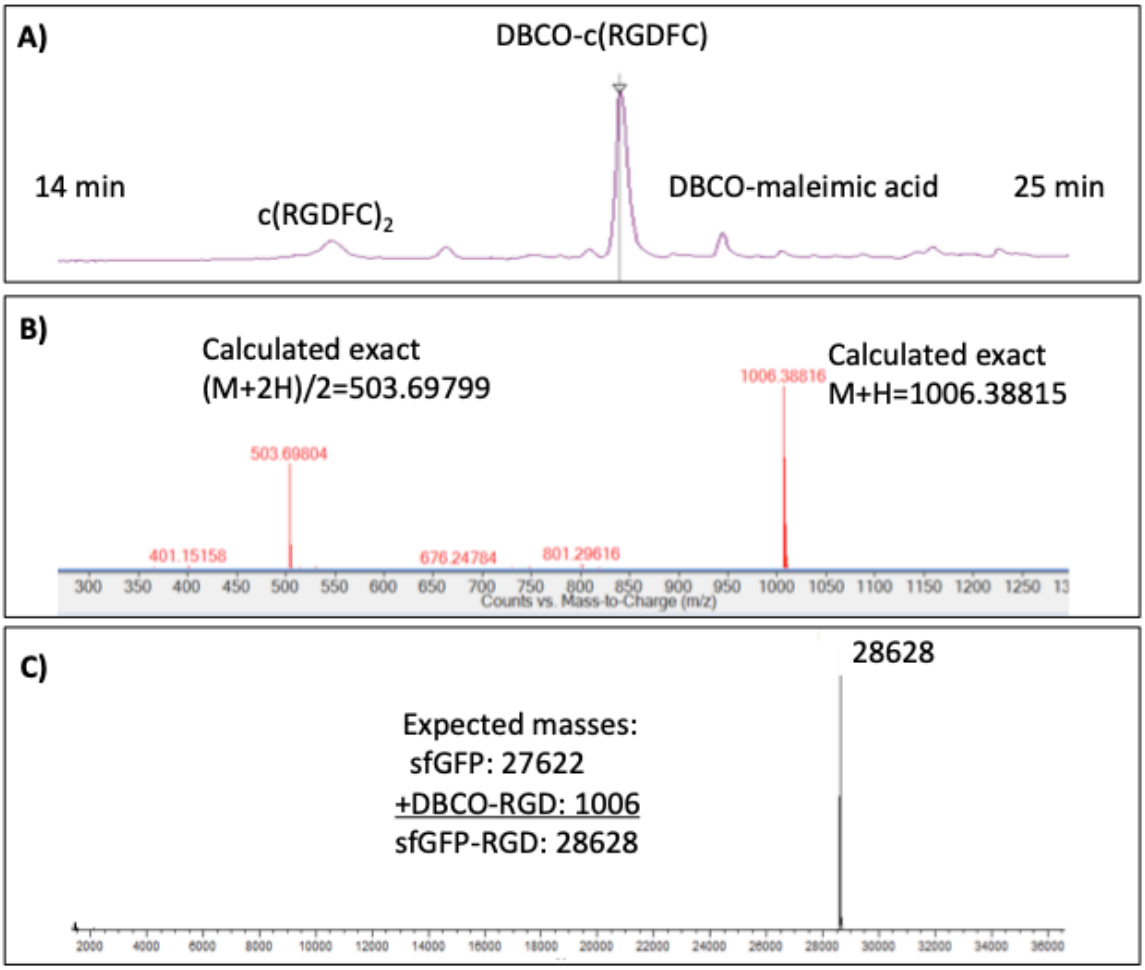


Figure 2.10 cRGDFC characterization. A) HPLC characterization of DBCO-cRGDFC. B) TOF MS characterization of DBCO-cRGDFC. C) AzF-sfGFP labeling with DBCO-cRGDFC. 20uM DBCO-cRGDFC overnight at room temperature.

2.2.4 Assaying the infectivity of AAV-cRGDFC conjugates

Next, cRGDFC was conjugated to these AAV mutants to assess the ability of these virus conjugates to re-target to SK-OV-3 cells. cRGDFC was conjugated to AzK-AAV by incubation at room temperature for 2 hours using a previously optimized method[88]. The resulting virus-conjugate was used to infect either HEK 293T cells at an MOI of 50gc/cell or SK-OV-3 cells at an MOI of 2,500gc/cell, as SK-OV-3 cells are naturally poor AAV targets. Addition of the cRGDFC moiety at each of these sites rendered the AAV non-infectious towards both HEK 293T cells and SK-OV-3 cells (Figure 11A, B).

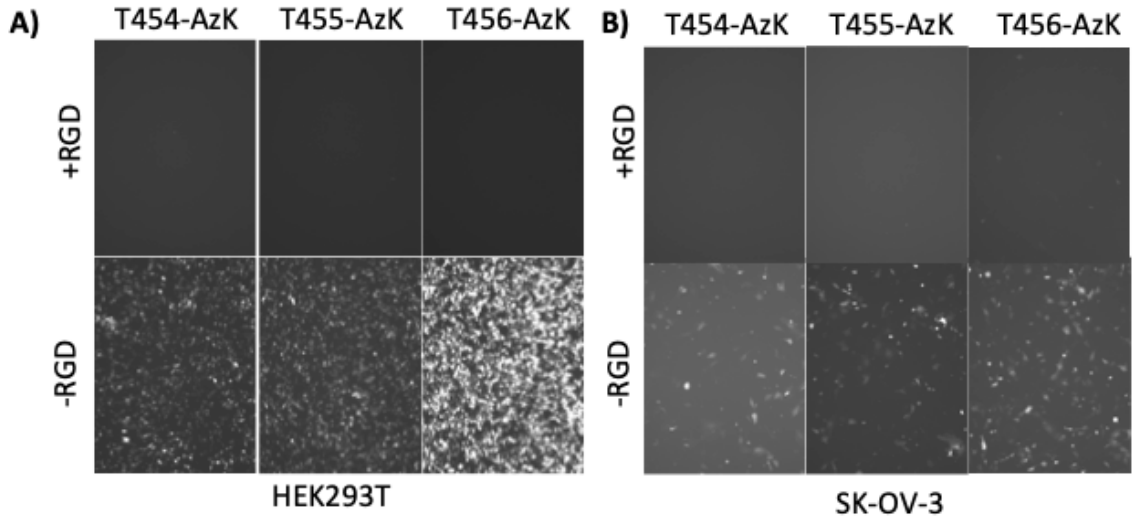


Figure 2.11. Retargeting of AAV-AzK with cRGDFC conjugation. Representative images of a constant MOI of AAV-AzK with or without cRGDFC conjugation. A) Constant MOI of 50gc/cell on HEK293T cells. B) Constant MOI of 2500gc/cell on SK-OV-3 cells.

This is likely due to the fact that these virus-conjugates now had the capability to bind both HSPG as well as the $\alpha_V\beta_3$ integrin receptor, two primary binding interactions. The combination of these two interactions likely occluded necessary downstream interactions from occurring, rendering the virus inactive. As seen in previous reports[64,71,73], this problem can be remedied by abolishing the HSPG

interaction, so R585 and R588 were mutated to alanine in each construct. Constructs with the R585A-R588A mutations will be denoted with AA from now on (Characterization in Figure 12A, B).

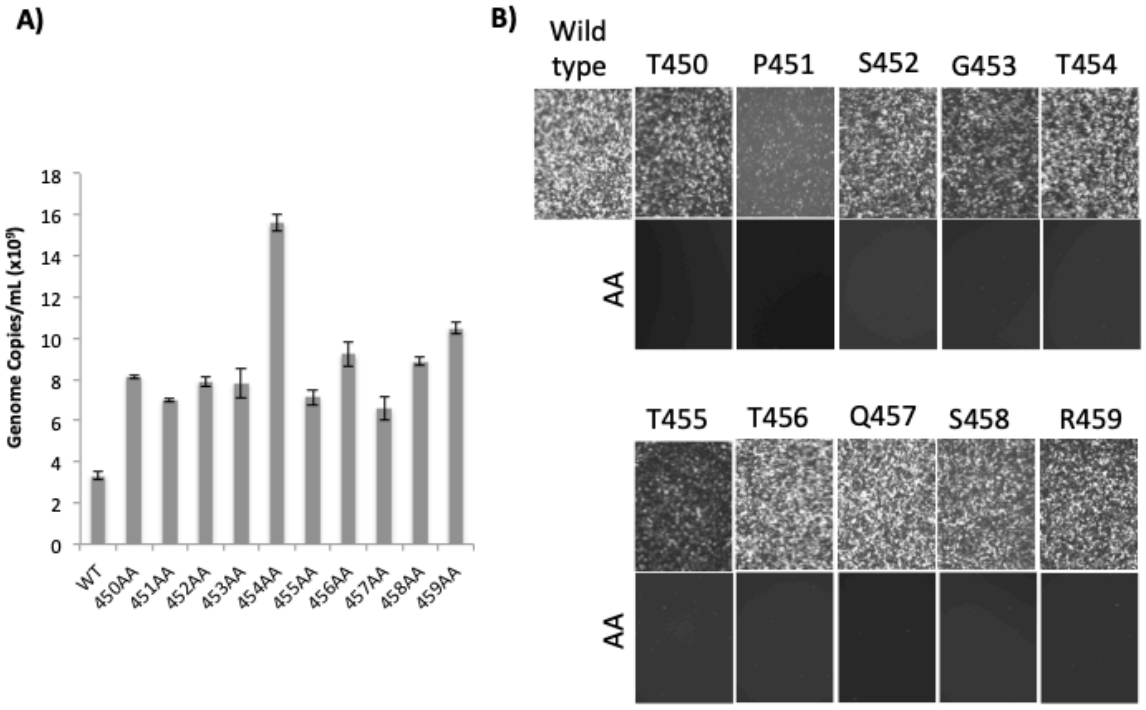


Figure 2.12 Characterization of HSPG-detargeted spike mutants. A) qPCR titers of AAV-AzK-AA virus made from 10cm dishes. B) Constant MOI infection of 50gc/cell on HEK23T cells of AAV-AzK and corresponding AAV-AzK-AA virus.

HSPG de-targeted cRGDFC virus conjugates were used to infect both HEK 293T cells and SK-OV-3 cells at a constant MOI of 50gc/cell and 2,500gc/cell, respectively. As predicted, eradicating the HSPG binding interaction caused the cRGDFC-conjugated AAV to gain infectivity towards SK-OV-3 cells (Figure 13A with corresponding images in Figure 14). The degree of infectivity from one site of cRGDFC conjugation to another varied greatly, from T456AzK-AA displaying 50% of WT infectivity to the least infective sites displaying around 5% of WT infectivity.

The ability to precisely place a targeting moiety on the AAV capsid to this degree has never before been achieved. While generally, the more surface exposed a residue was, the more efficient it was at retargeting, surface exposure does not appear to be the only factor that influenced retargeting capability (Figure 13B). As the AAV capsid and the interactions that it makes with cell receptors are highly complicated, we cannot know all the reasons why some residues were better at retargeting than others.

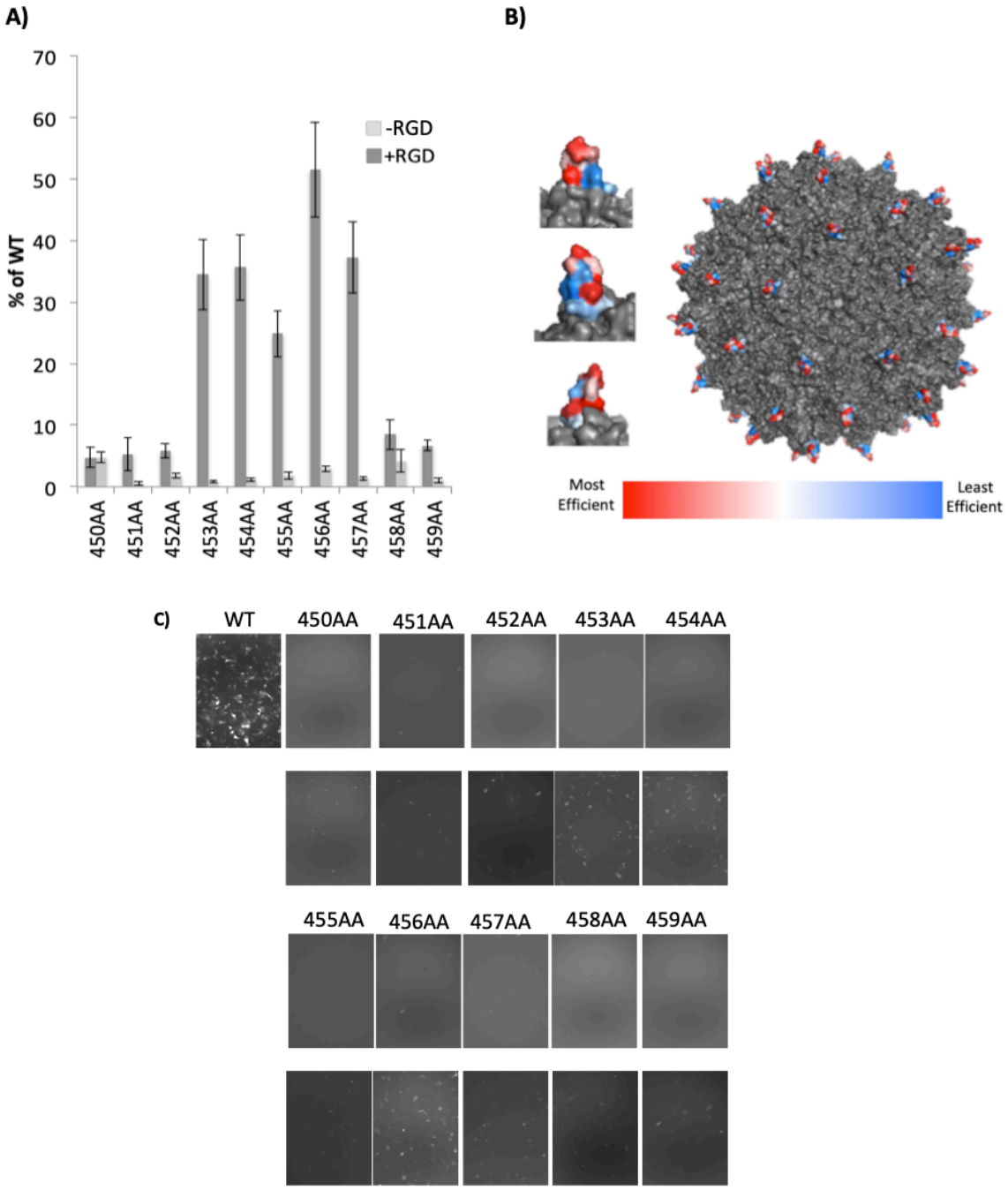


Figure 2.13. Retargeting efficiency of detargeted AAV-cRGDFC conjugates A) Infectivity of re-targeted viruses towards SK-OV-3 cells, infected at 2,500gc/cell. B) Map of virus showing the most infective re-targeting sites in red, and the least infective re-targeting sites in blue. C) Representative images corresponding to Figure 2.13A.

2.2.5 Inhibition assay

In order to insure that it was indeed the interaction with the primary binding receptor that had been re-directed, an inhibition assay was performed. Either wild

type or T456TAG-R585A-R588A virus was used to infect SK-OV-3 cells at a constant MOI of 2,500gc/cell in the presence of either free heparin sulfate or free cRGDS. Free heparin sulfate should inhibit the interaction between the R585-R588 residues and HSPG, whereas free cRGDS should have no effect. Conversely, the addition of free heparin sulfate should have no effect on the de-targeted T456TAG-R585A-R588A virus, which should be inhibited by free cRGDS. Indeed, this is exactly what was observed, suggesting that successful re-targeting is attributed to the de-targeted virus with the addition of cRGDFC (Figure 14A, B).

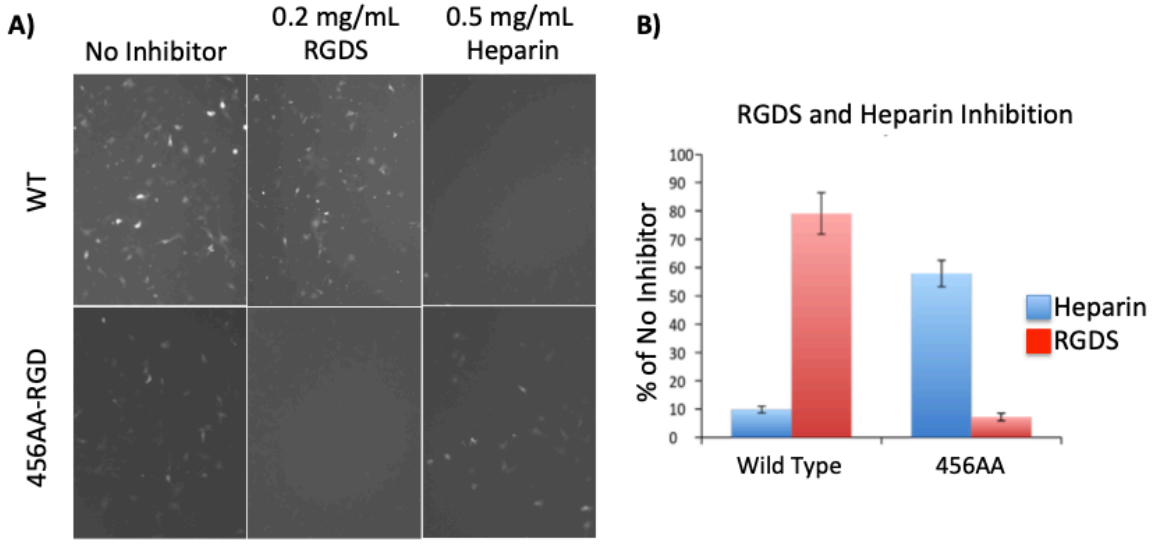


Figure 2.14 A), B) Inhibitor assay showing the % infectivity of no inhibitor as analyzed by A) microscopy and B) EGFP fluorescence of lysed cells that were infected with a constant MOI of cRGDFC modified AAV in the presence of free RGDS peptide or free Heparin.

2.2.6 The kinetics of capsid labeling

Optimization of the protocol for conjugating DBCO-cRGDFC to AzK containing AAV was done using T454AzK virus, which is perhaps the most surface exposed residue on the entire capsid. While there are 60 available azides on the surface of AAV containing AzK, it is important to know to what degree the capsid should be labeled for "optimal" infectivity. To understand this, T454AzK-AA was incubated

with 20μM cRGDFC and the reaction was quenched with free AzK at various time-points up to 24 hours after the start of the reaction. The capability that the resulting T454-cRGDFC-AA had to re-target was observed by infecting SK-OV-3 cells with a constant MOI. The resulting infectivity was observed using both microscopy and FACS analysis, and showed that at the beginning of the time-course, there was very little infectivity. However infectivity rapidly increased over the course of two hours until an "optimal" level of labeling was achieved. After two hours, increasing the number of modifications to the capsid surface resulted in a decrease in infectivity until the T454-cRGDFC-AA was rendered completely non-infectious (Figure 15).

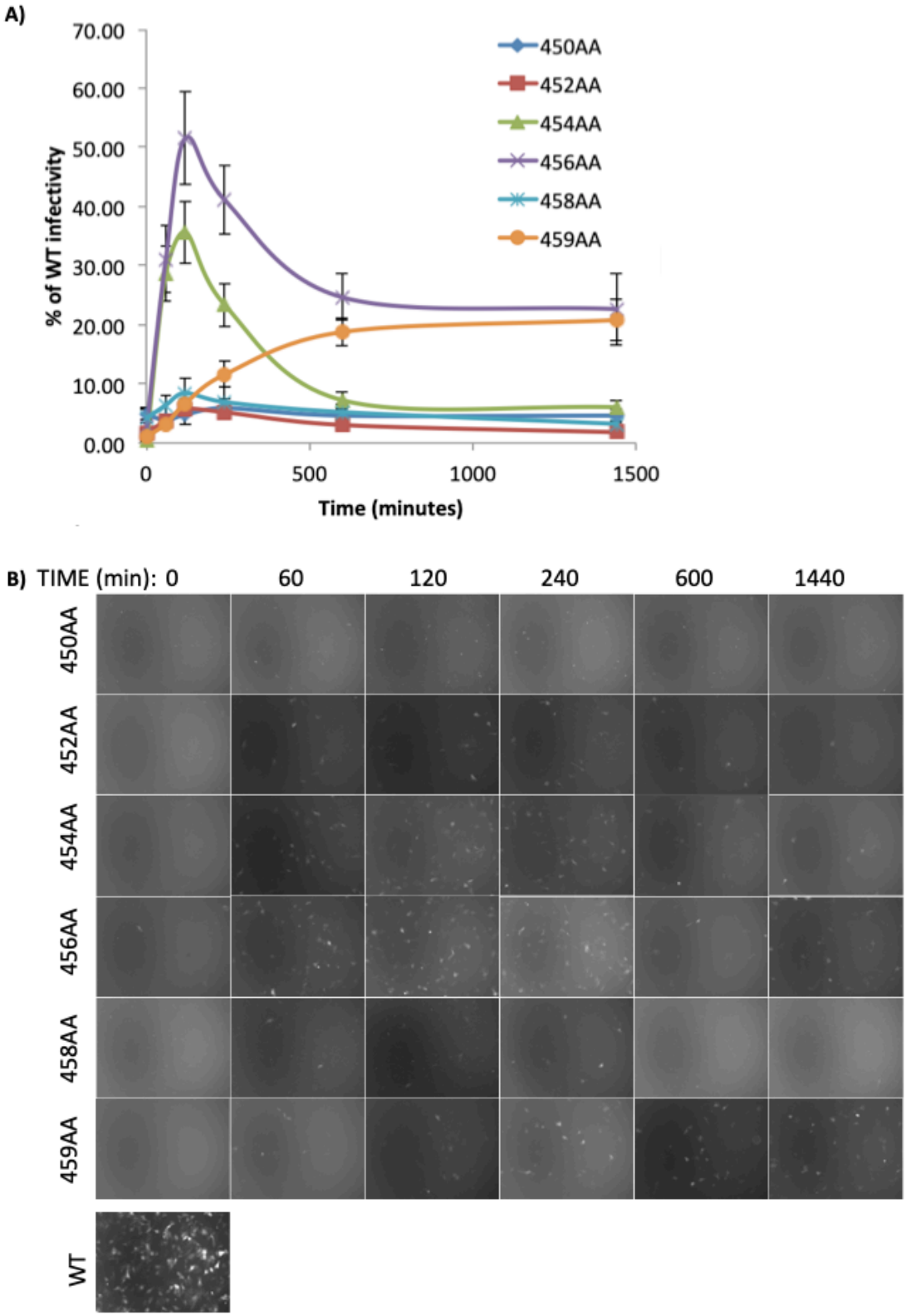


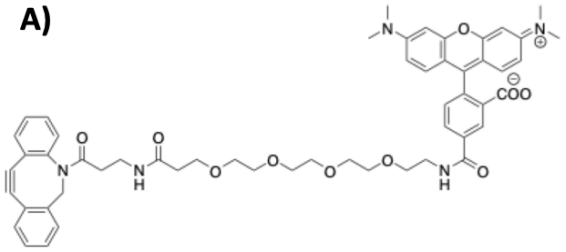
Figure 2.15. Retargeting efficiency of AAV-AzK with varying degrees of cRGDFC conjugation. A) % of wild type infectivity of virus-cRGDFC conjugates as analyzed by FACS. B) Representative images corresponding to A.

This same protocol was performed to test the retargeting capability of each virus mutant, regardless of the site of AzK incorporation. The question was raised whether the AzK's at each site were being modified at the same rate, or if the retargeting capability was simply better at some sites than others because they were being modified more due to faster labeling kinetics. To address this, six representative AzK locations (T450AzK-AA, S452AzK-AA, T454AzK-AA, T456AzK, S458AzK-AA, and R459AzK-AA) were chosen to perform additional optimization with. AzK-AAV mutants were incubated with 20uM DBCO-cRGDFC for varying amounts of time, quenching the reaction with 1mM free AzK over the course of 24 hours. SK-OV-3 cells were infected with the resulting virus and infectivity was analyzed. It was found that the majority of the sites followed similar labeling kinetics to that of T454, with optimal retargeting occurring after the reaction was allowed to proceed for two hours. The exception to this was R459AzK-AA, where infectivity continues to increase over the course of the 24 hours incubation (Figure 16). Although it is not clear why R459AzK-AA-cRGDFC follows different labeling kinetics, we hypothesize that the azide is more hidden in this spot than in the other spots, resulting in slower kinetics. However even after 24 hours of labeling at R459AzK-AA, T456AzK-AA-cRGDFC remains the most efficient retargeting virus conjugate.

2.2.7 Estimating the number of modifications displayed on the AAV capsid at the optimal labeling for retargeting

To gain insight into how many of the AzK's on the capsid surface were labeled under optimal labeling conditions, DBCO-Tamra (Figure 16A) was used in place of DBCO-cRGDFC to label the capsid surface for varying amounts of time, and the reaction was quenched with 1mM free AzK at various time points. The amount of labeling was then measured by comparing fluorescence intensity on an SDS-Page gel, which revealed that about 15%, or about 9 of the 60 available azides displayed on the capsid are conjugated to DBCO-cRGDFC at the optimal level of infectivity (Figure 16B).

A)



DBCO-TAMRA

B)

Reaction Time:	0 Hours	2 Hours	24 Hours	24 Hours
DBCO-TAMRA Concentration	20μM	20μM	20μM	200μM

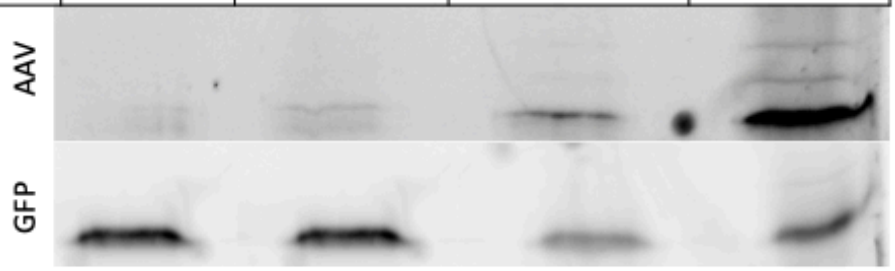


Figure 2.16. Estimation of the number of cRGDFC modifications on the AAV Capsid at the "optimal" time point for retargeting. A) The structure of DBCO-TAMRA. B) Fluorescent images of SDS-Page gel showing the labeling progress of T454AzK-AA virus. GFP was pre-labeled with DBCO-TAMRA and used as a loading control.

2.3 CONCLUSIONS

In this work, we have demonstrated the flexibility that UAA incorporation into the AAV capsid imparts to creating a functional virus conjugate. Additionally, we have used this technology to demonstrate the importance of the optimization of the placement of the re-targeting moiety, which has been impossible to do with such precision using previous methods.

The next step is the optimization of the stoichiometry of the cRGDFC moiety on the capsid surface. In this chapter, we demonstrated that re-targeting works best after 2 hours of labeling the AAV-AzK-AA capsid with 20μM cRGDFC, so there is clearly an optimal number of targeting moieties that can be displayed on the capsid surface and over-labeling the capsid results in a decrease of infectivity. We wanted to create a system in which the AAV capsid could be modified with a controlled number of cRGDFC molecules in a more homogenous manner. The way that we approached and achieved this is described in Chapter 3.

2.4 ACKNOWLEDGEMENTS

Thanks to Rachel Kelemen for doing the foundational work for this project and teaching me the techniques required to complete this work. Thanks also to the rest of Team AAV, including Xiaofu Cao, who assisted in cloning. Additional thanks to the Weerapana lab, particularly Tyler Bechtel for SK-OV-3 cells, and the Cam lab for help with qPCR.

2.5 EXPERIMENTAL PROCEDURES

2.5.1 Cell Culture

Hek293T and SK-OV-3 cells were obtained and cultured as previously described[1]

2.5.2 Cloning and Plasmids

Starting plasmids were obtained as previously described. pIDTSmart-RC2 was made by digesting pIDTSmart-MbPylRS-RC2 with AvrII and NheI and removing the *Mb*PylRS region. As AvrII and NheI are compatible cohesive ends, the digested plasmid was gel-purified and ligated back together without the *Mb*PylRS.

TAG mutations were introduced as previously described using primers found in Appendix 2. pIDTSmart-ITR-EGFP-ITR-8xPytR-*Mb*PylRS was made by digesting the previously described pIDTSmart-8xPytR$_{CUA}$ with AvrII and NheI to isolate and gel purify the 8xtRNA which was then inserted into the pIDTSmart-*Mb*PylRS plasmid which was linearized with SpeI. ITR-CMV-EGFP-ITR was inserted into pIDTSmart-8xPytR-*Mb*PylRS using an SbfI site.

pHelper was obtained as previously described[1].

2.5.3 Unnatural Amino Acids

Azido lysine (H-L-Lys(EO-N3)-OH) was purchased from Iris biotech.

2.5.4 Production of UAA-containing AAV

UAA-containing AAV was made as previously described[88]

2.5.5 qPCR

Experiments performed with virus that was harvested from a 12 well dish were titered using the AAVPro titration kit (Takara). Virus from 10cm dishes was first harvested by freezing in a dry-ice/ethanol bath and thawing in a 37˚C water bath two times, followed by the removal of cellular debris by centrifugation at 5,000xG for 5 minutes. DNAse I digestion was used to remove any DNA from the lysate, with the AAV capsid protecting the AAV genome. DNAseI was denatured using heat denaturation at 90˚C for ten minutes, and the capsid was further opened using Takara Bio's lysis buffer and incubation at 70˚C. qPCR was performed using the PowerUP SYBR Green Mastermix (Thermo Fisher) using the qPCR procedure recommended by the TAKARA bio kit. The sequence of the CMV-specific primers are:

CMV qPCR Fwd Primer: 5'-CATCTACGTATTAGTCATCGCTATTACCT-3'
CMV qPCR Rev Primer: 5'-GAAATCCCCGTGAGTCAAACC-3'.

2.5.6 AVB Purification of AAV

AVB resin was purchased from GE Healthcare. One 10cm dish was seeded with 8 million cells. 24 hours later, the cells were transfected as described. Two days later, cells were scraped off the 10cm dish and pelleted using centrifugation at 4,500xg for 10 minutes. The pelleted cells were re-suspended in 1xPBS and lysed by freezing in a dry-ice/ethanol bath and thawing in the 37˚C water bath two times. Cellular debris was pelleted by centrifugation at 5,000xg for 5 minutes. 250µL of settled resin was used per 10cm dish. Resin was washed with 10CV of PBS and lysate was applied to equilibrated resin. Lysate was allowed to flow through the column two times, and the resin bed was washed with 25 column volumes of PBS. Virus was eluted using 6 column volumes of 0.1M glycine pH 2.8, and each elution fraction was immediately

neutralized using 10% 1M Tris pH 8. The virus was precipitated using 11% Polyethylene-glycol (PEG-8000, Fisher Bioreagents, BP233) at 4˚C overnight. PEG-precipitated virus was pelleted by centrifugation at 1200xg for 30 minutes and re-suspended in 25µL 10% glycerol in PBS.

2.5.7 DBCO-TAMRA labeling and SDS Page

5µM sfGFP was mixed with 200µM DBCO-Tamra (Click chemistry tools, A131) and allowed to react with gentle agitation for 3 hours. Excess DCBO-TAMRA was then removed by dialysis (Thermo Scientific Slide-A-Lyzer MINI Dialysis Device, PI88402) and dialyzed against PBS at 4˚C overnight to remove excess DBCO-Tamra. The resulting GFP-Tamra was mixed with AAV and either 20 or 100µM DBCO-Tamra. The reaction was conducted in the dark at room temperature. At each time point, the reaction was quenched with 1mM AzK and then the mixture was dialyzed against PBS overnight to remove excess DBCO-Tamra and AzK. 1x SDS loading buffer was mixed with the samples after dialysis and samples were boiled at 95˚C for 5 minutes and then run on a 10% SDS Page gel.

2.5.8 DBCO-cRGDFC synthesis

DBCO-cRGDFC was synthesized as previously described[88].

2.5.9 Re-targeting and Flow Cytometry

20uM DBCO-cRGDFC was used to label AAV-AzK mutants for varying amounts of time at room temperature in the dark. A final concentration of 1mM AzK was used to quench the reactions. Virus was used to infect cells once the final time point was quenched. 0.2 million SK-OV-3 cells were seeded per 12-well plate about 24 hours prior to infection. 2,500gc/cell of virus was added to SK-OV-3 cells with 5mM sodium butyrate (Sigma-Aldrich) to enhance expression of AAV encoded transgenes. 48 hours post infection, infectivity was visualized by EGFP expression using a Zeiss Axio Observer fluorescence microscope with an XCite Series 120Q light source and Zeiss filter 44 (excitation 475/40nm, beamsplitter 500nm, emission 530/50nm). To prep cells for FACS, 200uL of warm 0.25% trypsin-EDTA solution was added to each well. Plates were incubated at 37˚C until cells began to detach from the plate, about 2 minutes. 500uL of ice-cold DMEM+10% FBS was added to each well to quench the trypsin. Cells were re-suspended by gentle pipetting and transferred to a microcentrifuge tube on ice. An additional 500uL of DMEM+10% FBS was used to rinse each well and the rinse was combined with the re-suspended cells. Cells were pelleted by centrifugation at 2,500xg for five minutes. The supernatant was discarded and cells were gently re-suspended in 500uL of ice-cold PBS and passed through a 100um filter. FACS analysis was performed using a Bio-Rad S3e cell sorter to quantify EGFP fluorescence.

2.6 REFERENCES

(1) Kelemen, R. E.; Mukherjee, R.; Cao, X.; Erickson, S. B.; Zheng, Y.; Chatterjee, A. A Precise Chemical Strategy To Alter the Receptor Specificity of the Adeno-Associated Virus. *Angew. Chemie - Int. Ed.* **2016**, *55* (36), 10645–10649. https://doi.org/10.1002/anie.201604067.

(2) Zhang, C.; Yao, T.; Zheng, Y.; Li, Z.; Zhang, Q.; Zhang, L.; Zhou, D. Development of next Generation Adeno-Associated Viral Vectors Capable of Selective Tropism and Efficient Gene Delivery. *Biomaterials* **2016**, *80*, 134–145. https://doi.org/10.1016/j.biomaterials.2015.11.066.

(3) Xie, Q.; Bu, W.; Bhatia, S.; Hare, J.; Somasundaram, T.; Azzi, A.; Chapman, M. S. The Atomic Structure of Adeno-Associated Virus (AAV-2), a Vector for Human Gene Therapy. *Proc. Natl. Acad. Sci.* **2002**, *99* (16), 10405–10410. https://doi.org/10.1073/pnas.162250899.

(4) Lochrie, M. A.; Tatsuno, G. P.; Christie, B.; McDonnell, J. W.; Zhou, S.; Surosky, R.; Pierce, G. F.; Colosi, P. Mutations on the External Surfaces of Adeno-Associated Virus Type 2 Capsids That Affect Transduction and Neutralization. *J. Virol.* **2006**, *80* (2), 821–834. https://doi.org/10.1128/JVI.80.2.821-834.2006.

(5) Shi, W.; Arnold, G. S.; Bartlett, J. S. Insertional Mutagenesis of the Adeno-Associated VIrus Type 2 (AAV2) Capsid Gene and Generation of AAV2 Vectors Targeted to Alternative Cell-Surface Receptors. **2001**, *12*.

(6) Wu, P.; Xiao, W.; Conlon, T.; Hughes, J.; Agbandje-McKenna, M.; Ferkol, T.; Flotte, T.; Muzyczka, N. Mutational Analysis of the Adeno-Associated Virus Type 2 (AAV2) Capsid Gene and Construction of AAV2 Vectors with Altered Tropism. *J. Virol.* **2000**, *74* (18), 8635–8647. https://doi.org/10.1128/JVI.74.18.8635-8647.2000.

(7) Leinschmidt, J. Ü. K.; Eléage, G. I. D.; Allek, M. I. H. Genetic Capsid Modifications Allow Efficient Re-Targeting of Adeno-Associated Virus Type 2. **1999**, *5* (9).

(8) Müller, O. J.; Kaul, F.; Weitzman, M. D.; Pasqualini, R.; Arap, W.; Kleinschmidt, J. A.; Trepel, M. Random Peptide Libraries Displayed on Adeno-Associated Virus to Select for Targeted Gene Therapy Vectors. *Nat. Biotechnol.* **2003**, *21* (9), 1040–1046. https://doi.org/10.1038/nbt856.

(9) Boucas, J.; Lux, K.; Huber, A.; Schievenbusch, S.; von Freyend, M. J.; Perabo, L.; Quadt-Humme, S.; Odenthal, M.; Hallek, M.; BÃ¼ning, H. Engineering Adeno-Associated Virus Serotype 2-Based Targeting Vectors Using a New Insertion Site-Position 453-and Single Point Mutations. *J. Gene Med.* **2009**, *11* (12), 1103–1113. https://doi.org/10.1002/jgm.1392.

(10) Shi, W.; Bartlett, J. S. RGD Inclusion in VP3 Provides Adeno-Associated Virus Type 2 (AAV2)-Based Vectors with a Heparan Sulfate-Independent Cell Entry Mechanism. *Mol. Ther.* **2003**, *7* (4), 515–525. https://doi.org/10.1016/S1525-0016(03)00042-X.

(11) Shi, W.; Hemminki, A.; Bartlett, J. S. Capsid Modifications Overcome Low Heterogeneous Expression of Heparan Sulfate Proteoglycan That Limits

AAV2-Mediated Gene Transfer and Therapeutic Efficacy in Human Ovarian Carcinoma. *Gynecol. Oncol.* **2006**, *103* (3), 1054–1062. https://doi.org/10.1016/j.ygyno.2006.06.020.

(12) Jr, K. H. W.; Gorbatyuk, O. S.; Jeffrey, K.; Opie, S. R.; Zolotukhin, S.; Warrington, K. H.; Harrison, J. K.; Muzyczka, N. Adeno-Associated Virus Type 2 VP2 Capsid Protein Is Nonessential and Can Tolerate Large Peptide Insertions at Its N Terminus. *J. Virol.* **2004**, *78* (12), 6595–6609. https://doi.org/10.1128/JVI.78.12.6595.

(13) Yang, Q.; Mamounas, M.; Yu, G.; Kennedy, S.; Leaker, B.; Mersn, J.; Yu, F. W. M. A. N. G.; Barber, J. R.; Al, Y. E. T. Development of Novel Cell Surface CD34-Targeted Recombinant Adenoassociated Virus for Gene Therapy. **1998**, *1937*, 1929–1937.

(14) Münch, R. C.; Janicki, H.; Völker, I.; Rasbach, A.; Hallek, M.; Büning, H.; Buchholz, C. J. Displaying High-Affinity Ligands on Adeno-Associated Viral Vectors Enables Tumor Cell-Specific and Safe Gene Transfer. *Mol. Ther.* **2013**, *21* (1), 109–118. https://doi.org/10.1038/mt.2012.186.

(15) Hagen, S.; Baumann, T.; Wagner, H. J.; Morath, V.; Kaufmann, B.; Fischer, A.; Bergmann, S.; Schindler, P.; Arndt, K. M.; Müller, K. M. Modular Adeno-Associated Virus (RAAV) Vectors Used for Cellular Virus-Directed Enzyme Prodrug Therapy. *Sci. Rep.* **2014**, *4*, 1–11. https://doi.org/10.1038/srep03759.

(16) Hörner, M.; Kaufmann, B.; Cotugno, G.; Wiedtke, E.; Büning, H.; Grimm, D.; Weber, W. A Chemical Switch for Controlling Viral Infectivity. *Chem. Commun.* **2014**, *50* (71), 10319–10322. https://doi.org/10.1039/c4cc03292f.

(17) Münch, R. C.; Muth, A.; Muik, A.; Friedel, T.; Schmatz, J.; Dreier, B.; Trkola, A.; Plückthun, A.; Büning, H.; Buchholz, C. J. Off-Target-Free Gene Delivery by Affinity-Purified Receptor-Targeted Viral Vectors. *Nat. Commun.* **2015**, *6*, 2–10. https://doi.org/10.1038/ncomms7246.

(18) Ponnazhagan, S.; Mahendra, G.; Kumar, S.; Thompson, J. A.; Castillas, M. Conjugate-Based Targeting of Recombinant Adeno-Associated Virus Type 2 Vectors by Using Avidin-Linked Ligands. *J. Virol.* **2002**, *76* (24), 12900–12907. https://doi.org/10.1128/jvi.76.24.12900-12907.2002.

(19) Lee, S.; Ahn, H. J. Anti-EpCAM-Conjugated Adeno-Associated Virus Serotype 2 for Systemic Delivery of EGFR ShRNA: Its Retargeting and Antitumor Effects on OVCAR3 Ovarian Cancer in Vivo. *Acta Biomater.* **2019**, *91*, 258–269. https://doi.org/10.1016/j.actbio.2019.04.044.

(20) Horowitz, E. D.; Weinberg, M. S.; Asokan, A. Glycated AAV Vectors: Chemical Redirection of Viral Tissue Tropism. *Bioconjug. Chem.* **2011**, *22* (4), 529–532. https://doi.org/10.1021/bc100477g.

(21) Yao, T.; Zhou, X.; Zhang, C.; Yu, X.; Tian, Z.; Zhang, L.; Zhou, D. Site-Specific PEGylated Adeno-Associated Viruses with Increased Serum Stability and Reduced Immunogenicity. **2017**. https://doi.org/10.3390/molecules22071155.

(22) Zhang, C.; Zhou, X.; Yao, T.; Tian, Z.; Zhou, D. Precision Fluorescent Labeling of an Adeno-Associated Virus Vector to Monitor the Viral Infection Pathway. *Biotechnol. J.* **2018**, *1700374*, 1–11. https://doi.org/10.1002/biot.201700374.

(23) Erickson, S. B.; Mukherjee, R.; Kelemen, R. E.; Wrobel, C. J. J.; Cao, X.; Chatterjee, A. Precise Photoremovable Perturbation of a Virus–Host Interaction. *Angew. Chemie - Int. Ed.* **2017**, *56* (15), 4234–4237. https://doi.org/10.1002/anie.201700683.
(24) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-TRNA Synthetase/TRNACUA Pair and Click Chemistry. *J. Am. Chem. Soc.* **2009**, *131* (25), 8720–8721. https://doi.org/10.1021/ja900553w.
(25) Atchison, R. W.; Casto, B. C.; Hammon, William, M. Adenovirus-Associated Defective Virus Particles. **1965**, *149* (3685), 754–756.
(26) Hoggan, M.; Blacklow, N.; Rowe, W. Studies of Small Dna Viruses Found in Various Adenovirus Preparations: Physical, Biological, and Immunological Characteristics. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *55* (6), 1467–1474.
(27) Samulski, R. J.; Berns, K. I.; Tan, M.; Muzyczka, N. Cloning of Adeno-Associated Virus into PBR322: Rescue of Intact Virus from the Recombinant Plasmid in Human Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79* (6 I), 2077–2081. https://doi.org/10.1073/pnas.79.6.2077.
(28) McLaughlin, S. K.; Collis, P.; Hermonat, P. L.; Muzyczka, N. Adeno-Associated Virus General Transduction Vectors: Analysis of Proviral Structures. *J. Virol.* **1988**, *62* (6), 1963–1973. https://doi.org/10.1128/jvi.62.6.1963-1973.1988.
(29) Matsushita, T.; Elliger, S.; Elliger, C.; Podsakoff, G.; Villarreal, L.; Kurtzman, G. J.; Iwaki, Y.; Colosi, P. Adeno-Associated Virus Vectors Can Be Efficiently Produced without Helper Virus. *Gene Ther.* **1998**, *5* (7), 938–945. https://doi.org/10.1038/sj.gt.3300680.
(30) Grieger, J. C.; Samulski, R. J. *Adeno-Associated Virus Vectorology, Manufacturing, and Clinical Applications*, 1st ed.; Elsevier Inc., 2012; Vol. 507. https://doi.org/10.1016/B978-0-12-386509-0.00012-0.
(31) Ingham, K. C. Protein Precipitation with Polyethylene Glycol. *Methods Enzymol.* **1984**, *104* (C), 351–356. https://doi.org/10.1016/S0076-6879(84)04101-X.
(32) Burova, E.; Ioffe, E. Chromatographic Purification of Recombinant Adenoviral and Adeno-Associated Viral Vectors: Methods and Implications. *Gene Ther.* **2005**, *12*, S5–S17. https://doi.org/10.1038/sj.gt.3302611.
(33) Auricchio, A.; Hildinger, M.; O'Connor, E.; Gao, G. P.; Wilson, J. M. Isolation of Highly Infectious and Pure Adeno-Associated Virus Type 2 Vectors with a Single-Step Gravity-Flow Column. *Hum. Gene Ther.* **2001**, *12* (1), 71–76. https://doi.org/10.1089/104303401450988.
(34) Nass, S. A.; Mattingly, M. A.; Woodcock, D. A.; Burnham, B. L.; Ardinger, J. A.; Osmond, S. E.; Frederick, A. M.; Scaria, A.; Cheng, S. H.; O'Riordan, C. R. Universal Method for the Purification of Recombinant AAV Vectors of Differing Serotypes. *Mol. Ther. - Methods Clin. Dev.* **2018**, *9* (June), 33–46. https://doi.org/10.1016/j.omtm.2017.12.004.
(35) Wang, Q.; Lock, M.; Prongay, A. J.; Alvira, M. R.; Petkov, B.; Wilson, J. M. Identification of an Adeno-Associated Virus Binding Epitope for AVB Sepharose Affinity Resin. *Mol. Ther. - Methods Clin. Dev.* **2015**, *2*, 15040.

https://doi.org/10.1038/mtm.2015.40.
(36) Kapp, T. G.; Rechenmacher, F.; Neubauer, S.; Maltsev, O. V.; Cavalcanti-Adam, E. A.; Zarka, R.; Reuning, U.; Notni, J.; Wester, H. J.; Mas-Moruno, C.; et al. A Comprehensive Evaluation of the Activity and Selectivity Profile of Ligands for RGD-Binding Integrins. *Sci. Rep.* **2017**, *7* (November 2016), 1–13. https://doi.org/10.1038/srep39805.

3.0 PROBING THE EFFECTS OF THE STOICHIOMETRY OF CRGDFC TARGETING MOIETIES DISPLAYED ON THE AAV CAPSID SURFACE IN RETARGETING

As discussed in chapter 2, there is a great need for customizable AAV vectors to be used as gene therapy vehicles. Many strategies, which have been extensively highlighted, have been created in an attempt to alter and hone AAV's tropism; however these strategies all fall short because they cannot undergo the necessary optimization to create a functional virus conjugate. As discussed in chapter 2, the incorporation of a UAA into the AAV capsid followed by chemical conjugation to a targeting moiety allows for much more extensive optimization of the virus conjugate, particularly in the placement of the targeting moiety. Chapter 2 also briefly discusses the importance of the number of targeting moieties placed on the capsid. This chapter delves further into the importance of the number of targeting moieties placed on the capsid surface. While the number of targeting moieties placed on the capsid surface can be controlled by chemical means- by stopping the chemical reaction after a specified time- we wanted to create a strategy that produced a uniformly-labeled population of AAV capsids. We did this by creating constructs where the UAA would be selectively incorporated into specific VP proteins. Incorporation into only VP1 or only VP2 results in the addition of roughly 5 modifications on the capsid surface, or incorporation into VP1 and VP2 results in the

incorporation of roughly 10 modifications. These constructs allow for the creation of a homogenous population of virus conjugate for which we can better control the number and distribution of modifications on the capsid surface.

3.1 INTRODUCTION

The background information presented in this chapter highlights previous work performed to control the number of modifications that can be placed on the AAV capsid. While no studies to date have focused solely on this optimization, several studies have created constructs similar to the constructs we create and present in this chapter. Much of this work is also discussed in chapters 1 and 2.

3.1.1 Previously used methods to control stoichiometry

While there has been an extensive amount of research invested into determining which sites on the AAV capsid will accept peptide insertions, there has not been nearly as much effort put into the optimization of the number of targeting moieties displayed on the capsid surface. This is in part due to the fact that the a site that will tolerate incorporation must be identified before any optimization can be done, as well as the fact that there are not currently very many ways to control the number of modifications placed on the capsid surface. However, there have been a number of studies that have been vital to understanding how the ratio of capsid modifications can be controlled, which guided many of our decisions in the following work. These publications will be highlighted in this section.

One common way to modify proteins is to modify exposed residues such as lysine or arginine. This can and has been done on the AAV capsid surface[52–54,170], however in addition to not having control over the location of the modification, it is impossible to control the amount of labeling that occurs. Chemical reactions can be stopped through means of quenching, but that is the extent to which it can be controlled.

A second method of controlling the number of modifications on a capsid is to modify specific VP proteins rather than modifying all three. The AAV capsid is composed of 60 capsid proteins, made up of VP1, VP2, and VP3 in a 1:1:10 ratio. Modification of all 60 capsid proteins can be achieved in a straightforward manner by genetically modifying the *Cap* gene, which will alter all three VP proteins. However it is also possible to uncouple the expression of the VP proteins from each other and achieve modifications in specific VP proteins. Incorporation into VP1 or VP2 results in roughly 5 modifications per capsid, whereas modification of both VP1 and VP2 results in about 10 modifications per capsid.

The earliest study to uncouple the VP proteins genetically fused a CD34-targeted single chain antibody fragment to the N-terminus of VP1, VP2, or VP3 and attempted to produce retargeted capsids out of the resulting fusions[70]. However they were unable to produce capsids from any combination of either all three VP fusion proteins or from one fusion protein and two unmodified VP proteins. This work resorted to including all three unmodified VP proteins into the packaging along with the addition of a single sFv fusion protein, which led to an undefined mixture of unmodified to modified capsid proteins. Only inclusion of the VP2 fusion

protein produced AAV particles capable of retargeting the capsids to the CD34 molecule, while inclusion of the VP1 or VP3 proteins had no effect on the retargeting capability.

Another study introduced the serpin ligand (FVFLI) in both the VP1 N-terminus as well as at the VP2 N-terminus; however, they did not uncouple VP2 expression from VP1 and VP3, so this modification appeared in both VP1 and VP2. They showed that capsids formed after insertion of the serpin ligand was capable of retargeting at both of these sites[68]. A study that inserted three different proteins, the chemokine binding domain of rat fractalkine, the human hormone leptin, and GFP at the N-terminus of each VP protein with the remaining two capsid proteins expressed separately as unmodified proteins demonstrated that only the VP2 N-terminus is capable of tolerating modifications and forming infectious particles[69].

These results have been further built upon to reveal that the N-terminus of VP1 contains essential nuclear localization signals and the PLA_2 domain, rendering it necessary for infectivity[13,15,171]. Additionally, VP3 is the main structural capsid protein and capsids will not assemble if its N-terminus is modified (for a more complete discussion on the work that has been performed on re-targeting AAV, refer to chapter 1). This work has been the foundation for further studies and modifications such as genetic fusions to GFP, darpins, affibodies, antibodies, and luciferase to the VP2 N-terminus have been successfully created[71–73,172–174]. Additionally, several studies have uncoupled the VP proteins' expression from one another in order to incorporate motifs into distinct capsid proteins in regions such as the spike or the HSPG binding site[66,175].

It is clear from previous work that modifications are only tolerated in certain parts of the AAV capsid. All of these studies have been performed in an attempt to better understand where modifications can be accepted, and not to better understand how the number of modifications on the capsid surface will affect the capsid's ability to re-target. In chapter 2, I demonstrated the importance of the stoichiometry of the number of modifications on the capsid surface. In that work, AzK's were displayed on all 60 capsid proteins, so the population of virus-conjugate that we created was a heterogeneous population- the number and placement of cRGDFC motifs are displayed on the AAV capsid differently form conjugate to conjugate. In this chapter, we create a platform where to produce a homogenous population of virus conjugates with controlled stoichiometry to better understand the importance of the stoichiometry of labeling.

We created a number of AAV constructs that incorporate AzK into single VP proteins. To express AzK into only VP1, we created a construct in which the start codon for VP1 is deleted, and VP1 is expressed *in trans* after a CMV promoter. Similar constructs were created to incorporate AzK into only VP2, or both VP1 and VP2, using the same methods (Figure 1). While we rely heavily on the estimation that the VP proteins are present in a 1:1:10 ratio of VP1:VP2:VP3, some work suggests that this ratio is incorrect. The Heck lab uses a modified orbitrap mass spectrometer to determine the VP ratio and their work suggests that the VP1:VP2:VP3 ratio is much less defined with only 0-2 copies of VP1, 8-11 copies of VP2, and 48-51 copies of VP3 in an individual capsid[176]. Even in this case, the populations of AAV-cRGDFC that we create in this work still produce a much more

homogenous population of virus conjugate than when there are 60 azides available, as well as a considerable increase over the control of the number of modifications that can be placed on the AAV capsid.

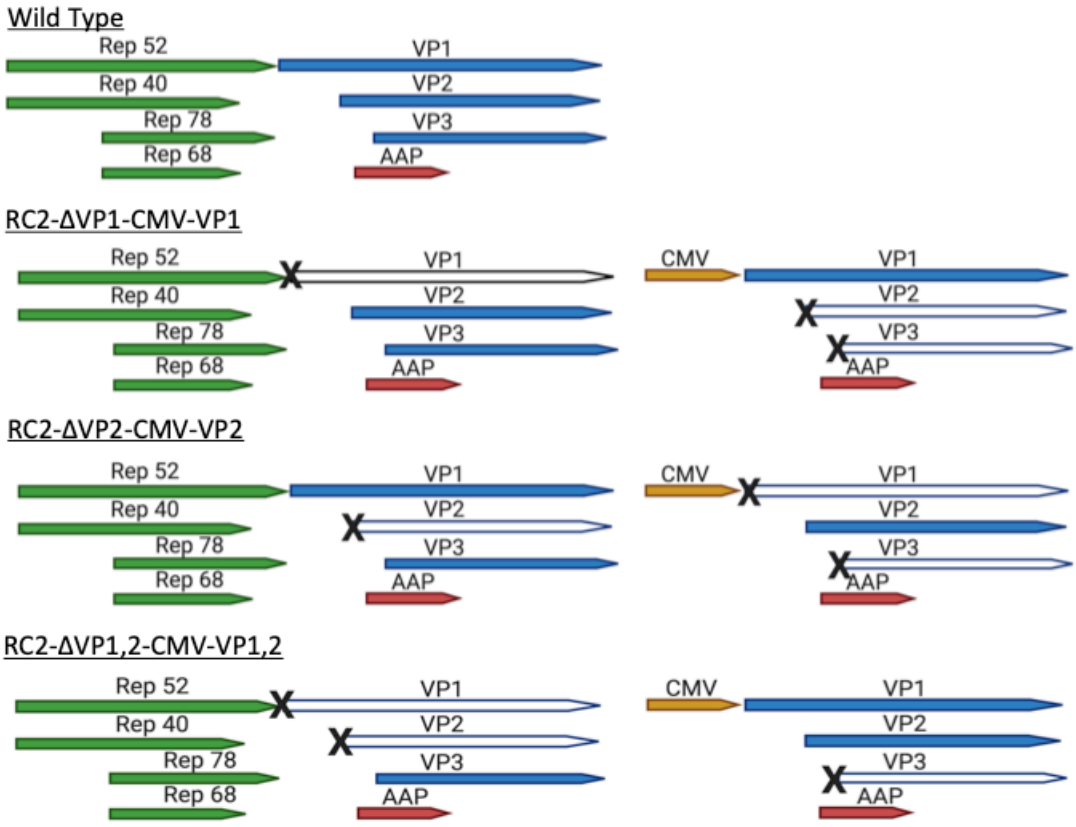


Figure 3.1 Schematic depicting the single VP protein constructs. In wild type AAV, VP1, VP2, and VP3 are all expressed from the same open reading frame. Their N-termini differ through the use of minor start codons and alternative splicing. In order to separate the expression of VP1 from that of VP2 and VP3 in the RC2-ΔVP1-CMV-VP1 construct, VP1's start codon was deleted. The *cap* genes were then placed after a start codon with a functional start codon for VP1 but mutated start codons to suppress expression of VP2 and VP3. In this way, a stop codon can be placed into VP1 only, making it so that UAA incorporation only occurs in VP1. RC2-ΔVP2-CMV-VP2 and RC2-ΔVP1,2-CMV-VP1,2 were created in a similar fashion.

3.1.2 Plasmid design and cloning

The first step in uncoupling the VP proteins from each other was abolishing their expression with the other two VP proteins. This was done by mutating the individual VP proteins start codons using overlap PCR. To delete VP1's expression, the M1 start codon was mutated to leucine. A second construct was made where a

modified *Cap* gene followed the CMV promoter where VP1 was expressed from its native promoter, but VP2 and VP3's expression were deleted. Deleting VP2's start codon was achieved by making a silent mutation of ACG to ACC (T138T). VP3 was deleted by mutating 5 potential start codons- M203L (ATG→CTC), M211L (ATG→CTC), M235L (ATG→CTC), T197L (ACG→ACC) and L202L (CTG→CTC). It has previously been reported that it is necessary to mutate M211 and M235 to eliminate VP3 expression, and T197 and L202 are two potential non-canonical start codons that could initiate translation of VP3[43]. The six mutations that were made to abolish VP2 and VP3's expression were generated by overlap extension PCR, creating pIDTSmart-CMV-VP1. The CMV-VP1 fragment was digested out of pIDTSmart-CMV-VP1 and ligated in place of the *Mb*PylRS in pIDTSmart-RC2-ΔVPI-*Mb*PylRS to make pIDTSmart-RC2-ΔVP1-CMV-VP1.

Similar steps were followed to create pIDTSmart-RC2-ΔVP2-CMV-VP2. pIDTSmart-RC2-VP2Δ-*Mb*PylRS, an intermediate plasmid, was generated by making a silent mutation on VP2's start codon, T138T (ACG→ACG). To build pIDTSmart-CMV-VP2, the VP2 start codon, T138 was mutated to ATG, a stronger start codon. To avoid any expression of VP1 as well as retain the VP3 deletions previously made, a PCR reaction that amplified DNA starting from the VP2 start codon from pIDTSmart-CMV-VP1 was performed. This product was ligated into the pIDTSmart-CMV backbone. pIDTSmart-RC2-ΔVP2 and pIDTSmart-RC2-ΔVP2-CMV-VP2 were created following the same scheme as the VP1 version.

Finally, pIDTSmart-RC2-ΔVP1, VP2 was created by making both of the aforementioned M1L and T138T mutations to abolish VP1 and VP2 expression in

the pIDTSmart-RC2 plasmid. Intermediate plasmid pIDTSmart-CMV-VP1-VP2(ATG) was created by making the T138M mutation by overlap PCR in pIDTSmart-RC2-ΔVP3, and PCR amplifying out the entire VP region. pIDTSmart-RC2-ΔVP1,2-*Mb*PylRS was then digested with NheI and AvrII and the *Mb*PylRS region was replaced with the VP1-VP2(ATG)-ΔVP3 region that was amplified out. All mutations can be seen in the Table 1.

Table 3.1 A comprehensive chart showing all mutations made to delete start codons in creating the single VP protein constructs. In addition to the start codons deleted in this chart, it should be noted that the VP2 start codon was mutated from ACG to ATG when it was introduced back in after the CMV promoter.

Construct	Deleted Start Codons in RepCap	Deleted start codons in *Cap* following CMV
pIDTSmart-RC2-ΔVP1	M1L (ATG→CTC)	
pIDTSmart-RC2-ΔVP2	T138T (ACG→ACC)	
pIDTSmart-RC2-ΔVP3	M203L (ATG→CTC), M211L (ATG→CTC), M235L (ATG→CTC), T197L (ACG→ACC), L202L (CTG→CTC)	
pIDTSmart-RC2-ΔVP1-CMV-VP1	M1L (ATG→CTC)	T138T (ACG→ACC), T197T (ACG→ACC), L202L (CTG→CTC), M203L (ATG→CTC), M211L (ATG→CTC), M235L (ATG→CTC)
pIDTSmart-RC2-ΔVP2-CMV-VP2	T138T (ACG→ACC)	M1L(ATG→CTC), T197T (ACG→ACC), L202L (CTG→CTC), M203L (ATG→CTC), M211L (ATG→CTC), M235L (ATG→CTC)
pIDTSmart-RC2-ΔVP1,VP2-CMV-VP1,VP2	M1L (ATG→CTC), T197T (ACG→ACC)	T197T (ACG→ACC), L202L (CTG→CTC), M203L (ATG→CTC), M211L (ATG→CTC), M235L (ATG→CTC)

From these plasmids, stop codons were incorporated by performing overlap extension PCR on pIDTSmart-RC2 plasmids using the appropriate terminal primers (Appendix 2) and ligated into the corresponding intermediate plasmid.

3.2 RESULTS AND DISCUSSION

3.2.1 AAV production and quantification

We first decided to look into the constructs that would produce capsids with UAA incorporated into only VP1 or only VP2. rAAV was produced as described in chapter 2 with a triple transfection (Figure 2). Western blotting with the anti-AAV-B1 IgG primary antibody, which recognizes each individual capsid protein, was used to ensure that the designed constructs were producing the intended VP proteins (Figure 3). The wild type AAV shows all three VP bands, and the RC2-ΔVP1 and RC2-ΔVP2 show the disappearance of VP1 and VP2, respectively. We show that VP1 and VP2 can be produced alone, with only a minimal amount of VP3 being expressed. This small amount is likely from leaky VP3 translation initiated by additional non-canonical start codons. Lastly, the western blot shows that VP1 and VP2 can be provided *in trans* along with the other VP proteins so that all three VP proteins are present. qPCR was used to quantify the production of each virus. It was discovered that each virus had titers similar to the wild-type titer and, in most cases, the production of mutant virus was better than that of wild type. When AzK was introduced, titers remained high in both the presence and absence of AzK. In the absence of AzK, the stop codon should result in the truncation of the protein, which will not be assembled into the capsid structure. However previous studies have shown that non-infectious virus-like particles can form from VP3 only[69], which accounts for the high titers in the absence of AzK.

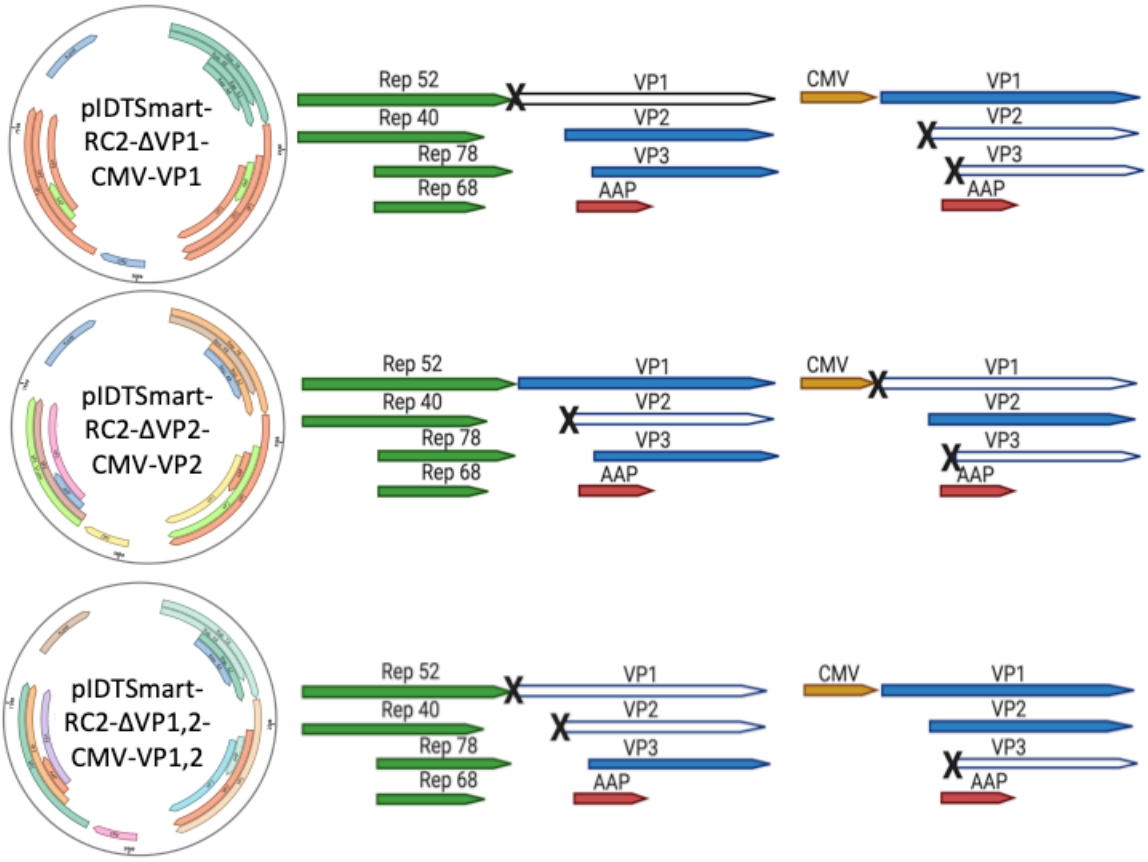


Figure 3.2. Plasmids used to make AAV with uncoupled VP protein expression and accompanying diagrams. Single protein constructs were expressed from pIDTSmart plasmids.

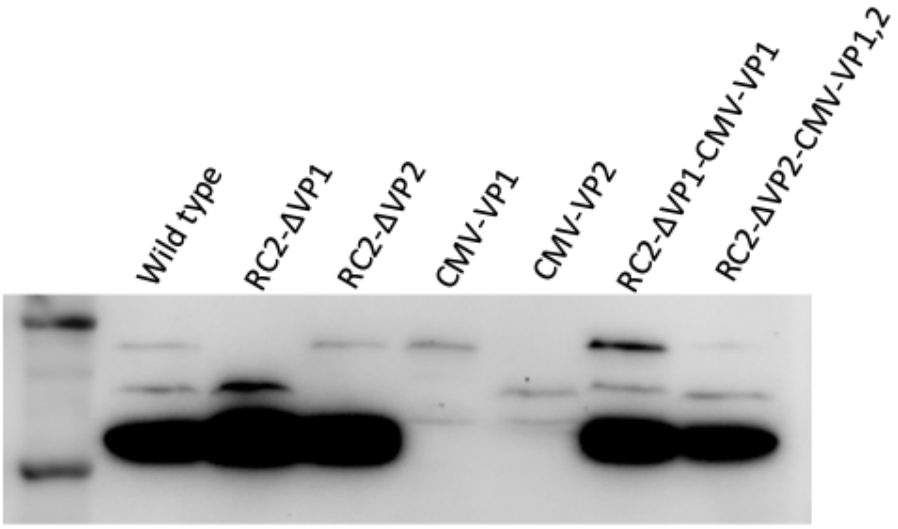


Figure 3.3. Western blot of AAV2 capsid protein production confirming the presence of the expected VP proteins in mutant constructs in HEK293T cells. Equal volumes of lysate were resolved by 10% SDS-PAGE and analyzed by western blotting using the B1 antibody. This gel was made and run by Xiaofu Cao.

HEK-293T cells were infected with a constant MOI of 50gc/cell and infectivity was assayed using both fluorescence microscopy and quantification using the fluorescence of EGFP in the lysate. Infectivity showed that deleting VP1 rendered non-infectious virus (Figure 4). Deleting VP2 also caused a large decease in infectivity, but not as much as the deletion of VP1. Expressing these proteins after the CMV promoter restored infectivity to wild type levels. Additionally, infectivity results support the hypothesis that virus constructs containing a stop codon that are made in the absence of AzK are non-infectious virus-like particles: in the presence of AzK, infectivity is near that of wild type but drastically decreases in the absence of AzK (Figure 4).

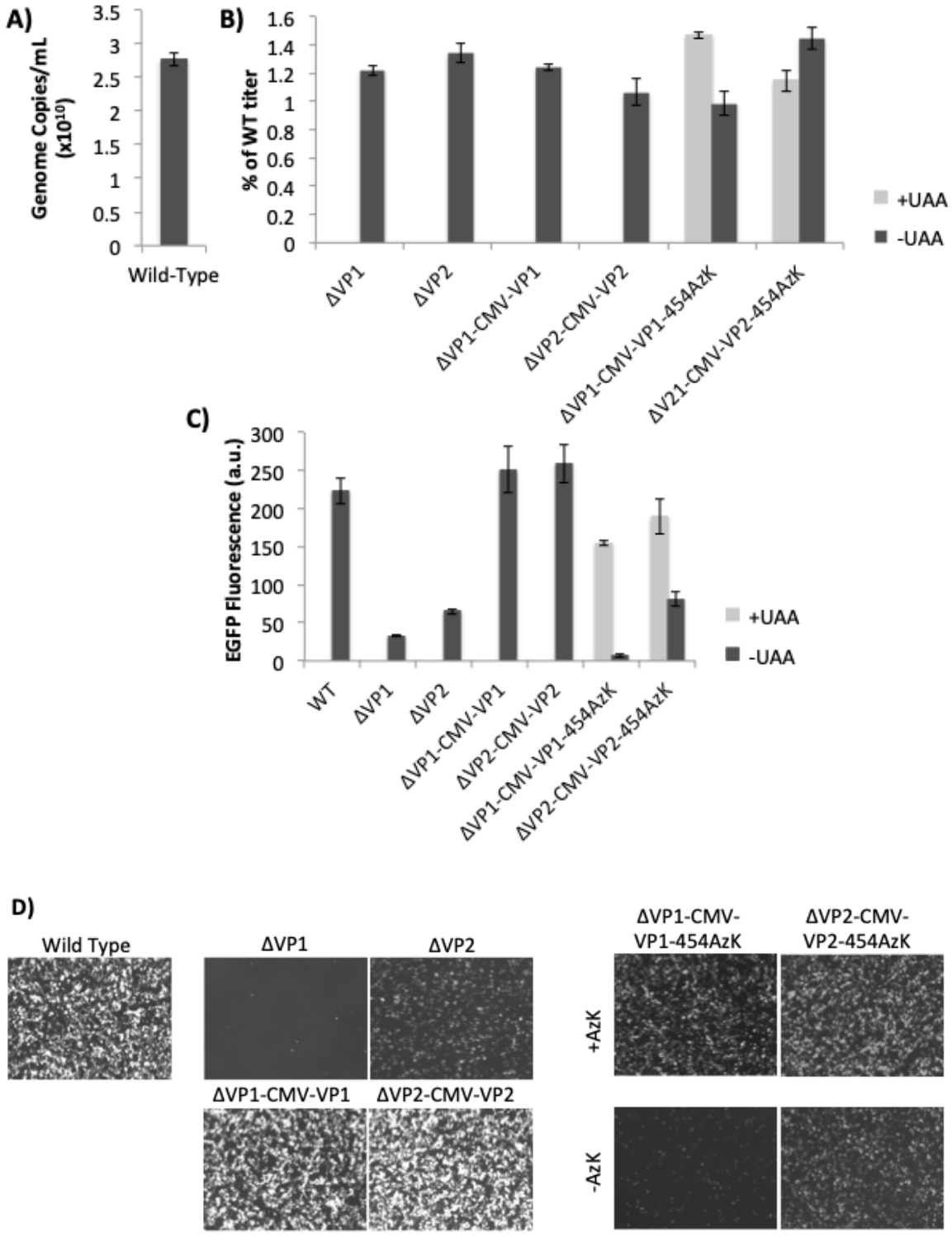


Figure 3.4. Production efficiency of AAV with uncoupled VP protein expression. A) Production efficiency of wild type AAV as analyzed by qPCR. B) Production efficiency as a percent of wild type of constructs with VP1 or VP2 removed, Expressed in trans under the control of a CMV promoter, or expressed in trans under the control of a CMV promoter with the incorporation of AzK. C) EGFP expression analyzed by lysed cells expressing EGFP from the AAV cargo of the constructs from A and B. D) Images corresponding to the fluorescence values from C. These experiments were performed with Xiaofu Cao.

At this point, we made the decision to look into the incorporation of AzK into VP1 and VP2 as well, so these experiments were done with those constructs as well (Figure 5). The presence of AzK in the capsid was confirmed by purifying virus using either heparin affinity resin or AVB agarose and conjugating the purified capsids to either DBCO-Cy5 or DBCO-TAMRA using copper-free click chemistry and analyzing by SDS-page (Figure 6).

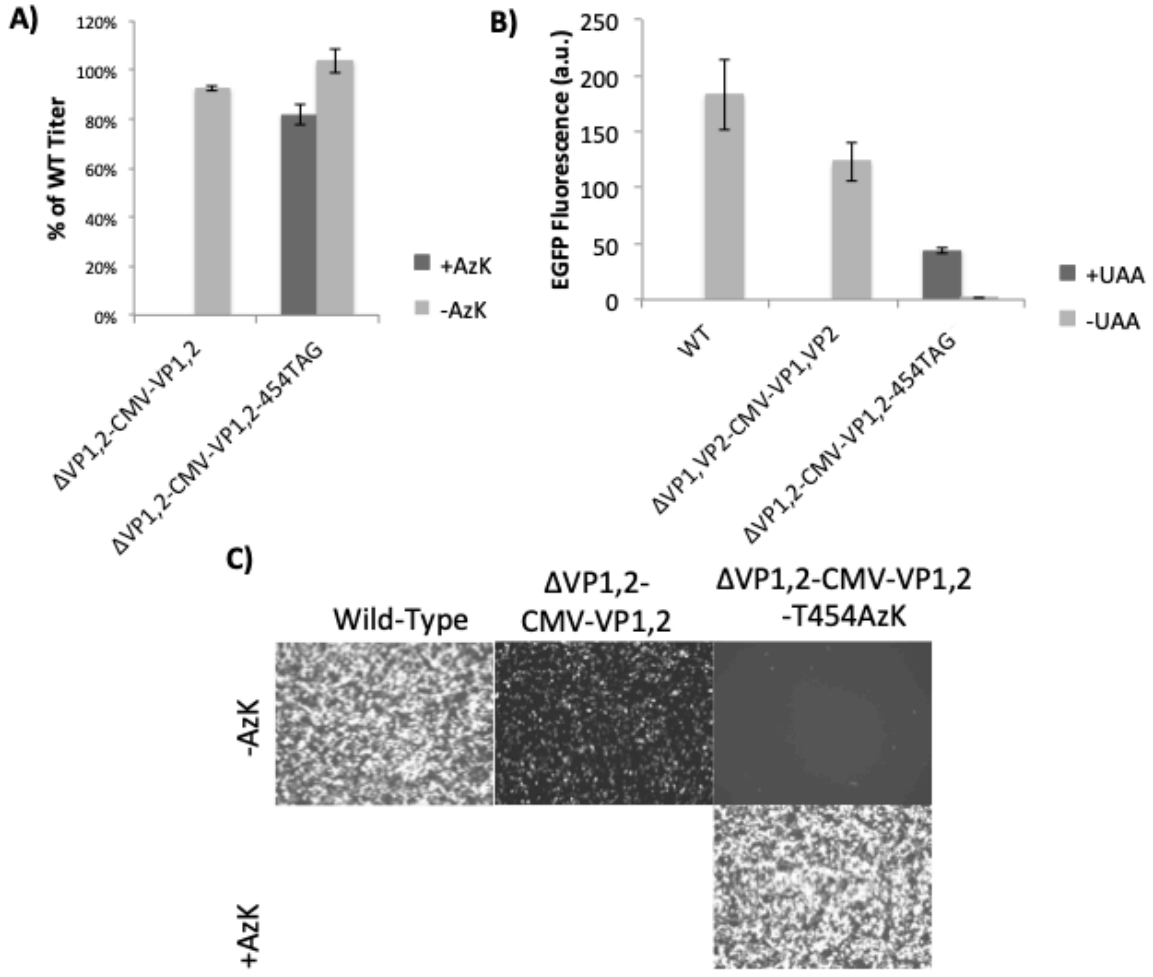


Figure 3.5. Production efficiency of AAV with uncoupled VP1 and VP2 protein expression. A) Production efficiency as a percent of wild type of constructs with both VP1 and VP2 expression uncoupled and expressed in trans under the control of a CMV promoter with or without the incorporation of AzK. B) EGFP expression analyzed by lysed cells expressing EGFP from the AAV cargo of the constructs from B. C) Images corresponding to the fluorescence values from B.

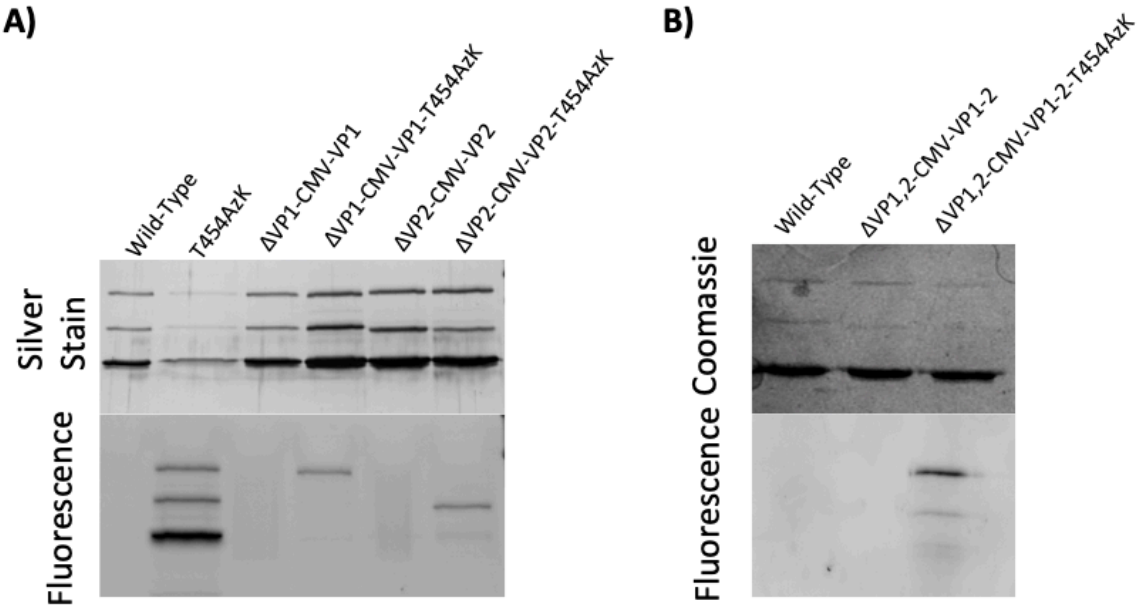


Figure 3.6. SDS-Page analysis of virus with AzK incorporated into VP1, VP2, and VP3, only VP1, or only VP2 (A), or VP1 and VP2 (B). Virus was labeled with DBCO-Cy5, run on an SDS-Page gel, and imaged with both fluorescence imaging or by silver stain (A) or coomassie (B). Image 6A was made by Xiaofu Cao.

3.2.2 Retargeting using selective UAA incorporation constructs

Constructs with AzK incorporated into specific VP proteins, containing roughly five or ten AzK's were assayed for their ability to re-target to SK-OV-3 cells. Retargeting experiments were performed using virus that had AzK incorporated at site T454, as it is the most surface exposed residue on the capsid surface. Additionally, all constructs had R585A-R588A mutations in all VP proteins to abolish the HSPG interaction. As previously done, 20uM cRGDFC-DBCO was incubated with virus preps and the reaction was quenched at different time points over the course of 24 hours. The resulting virus-conjugates were used to infect SK-OV-3 cells at a constant MOI, and infectivity was assayed by FACS analysis (Figure 7). Unlike the virus conjugates with 60 available azides, which rapidly gained in infectivity until it peaked at about 30% of WT infectivity and then became overly

modified, virus that had 5 available azides (ΔVP1-CMV-VP1-454TAG-AA and ΔVP2-CMV-VP2-454TAG-AA slowly gained in infectivity until infectivity plateaued at about 10% of WT infectivity, and never lost infectivity. Additionally, the plateau of infectivity also suggests that the available azides are being completely modified. As 5 cRGDFC moieties on the capsid surface resulted in approximately 10% of WT infectivity we concluded that 5 targeting moieties on the capsid surface was not enough. Interestingly, incorporation into VP2 only resulted in slightly higher retargeting efficiency than incorporation into VP1 only, supporting the idea that VP2 might be present at a higher number in the capsid than VP1[176]. The ΔVP1, VP2-AA-CMV-VP1, VP2-454TAG-AA construct that had roughly 10 available azides followed a similar trend, with a more rapid increase in infectivity and a plateau at about 25% of WT infectivity. Again, the infectivity plateaued and a loss in infectivity was never seen, which suggests that the fully conjugated virus did not have enough targeting moieties to reach its most infective state. However, it was much closer to the efficiency displayed when there were 60 azides available on the capsid surface.

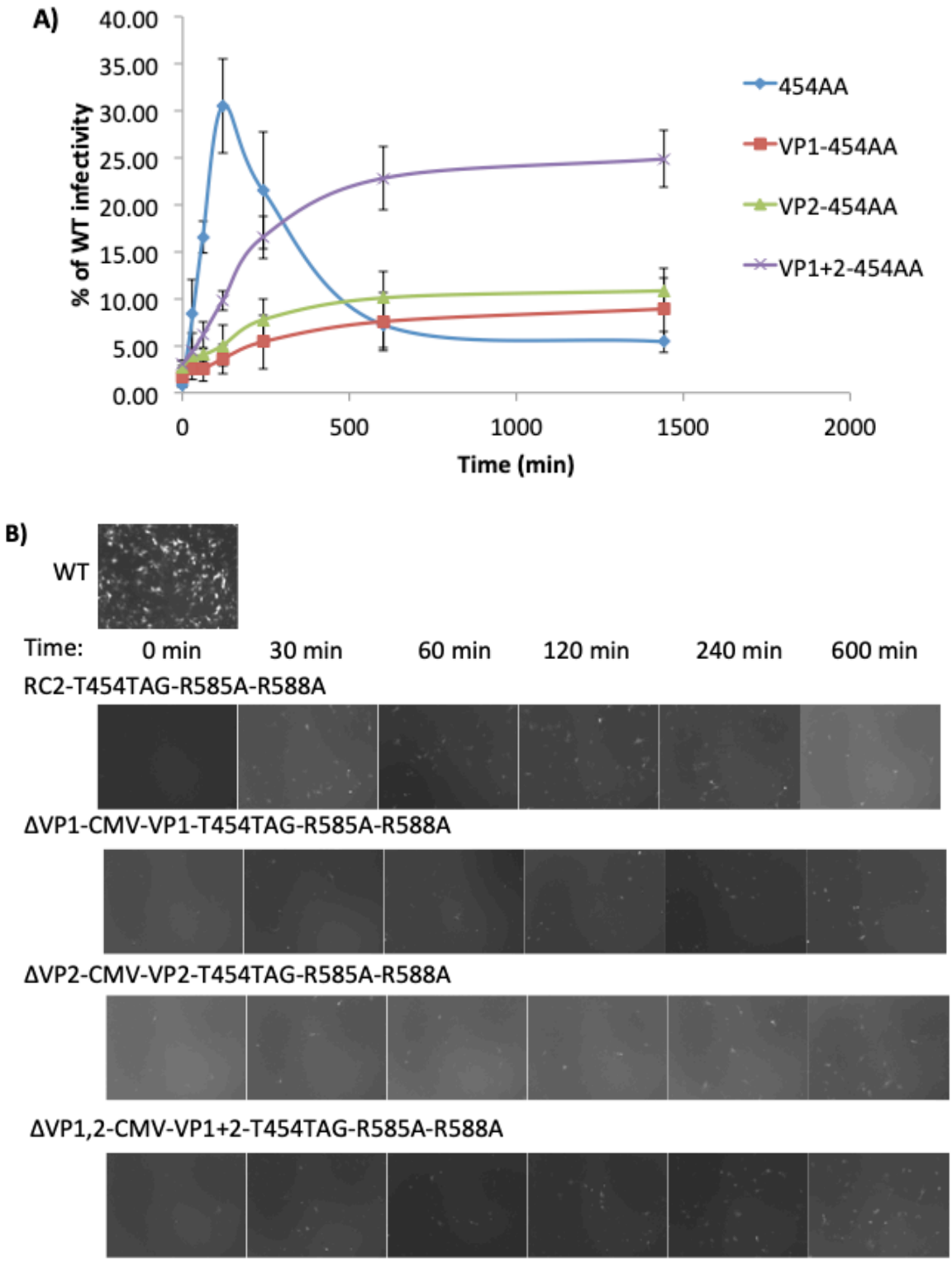


Figure 3.7. Retargeting efficiency of AAV-cRGDFC conjugates. A) Retargeting efficiency as a percent of wild type infectivity of AAV-cRGDFC conjugates displayed on all 60 capsid proteins (VP1, VP2, and VP3), on about 5 capsid proteins (VP1 or VP2 only), or on about 10 capsid proteins (VP1 and VP2) as analyzed using FACS analysis. B) images accompanying the FACS analysis from A.

AAV2 virus with 5 or 10 available azides displayed a slow gain of infectivity over time. During this gain of infectivity, we expected to see a trend in infectivity that followed pseudo first order kinetics- one cRGDFC would react with one cell receptor and result in infectivity. However upon inspection of the curve, two different rates are being observed (Figure 8). There is a period of time where no gain in infectivity is observed from 0 to about 50 minutes, and after 50 minutes, infectivity increases until it plateaus, suggesting that all available azides have been modified. This suggests that there must be multiple interactions between the AAV capsid and the $\alpha_V\beta_3$ integrin receptors for cellular entry to occur. While this was initially surprising, it is similar to the interactions made between AAV2's natural primary binding with HSPG. R585 and R588 appear in clusters around the three-fold axis and appear to interact with HSPG in these clusters (Figure 8B). Future experiments could be performed that design virus conjugates with designed multivalent interactions for further optimization.

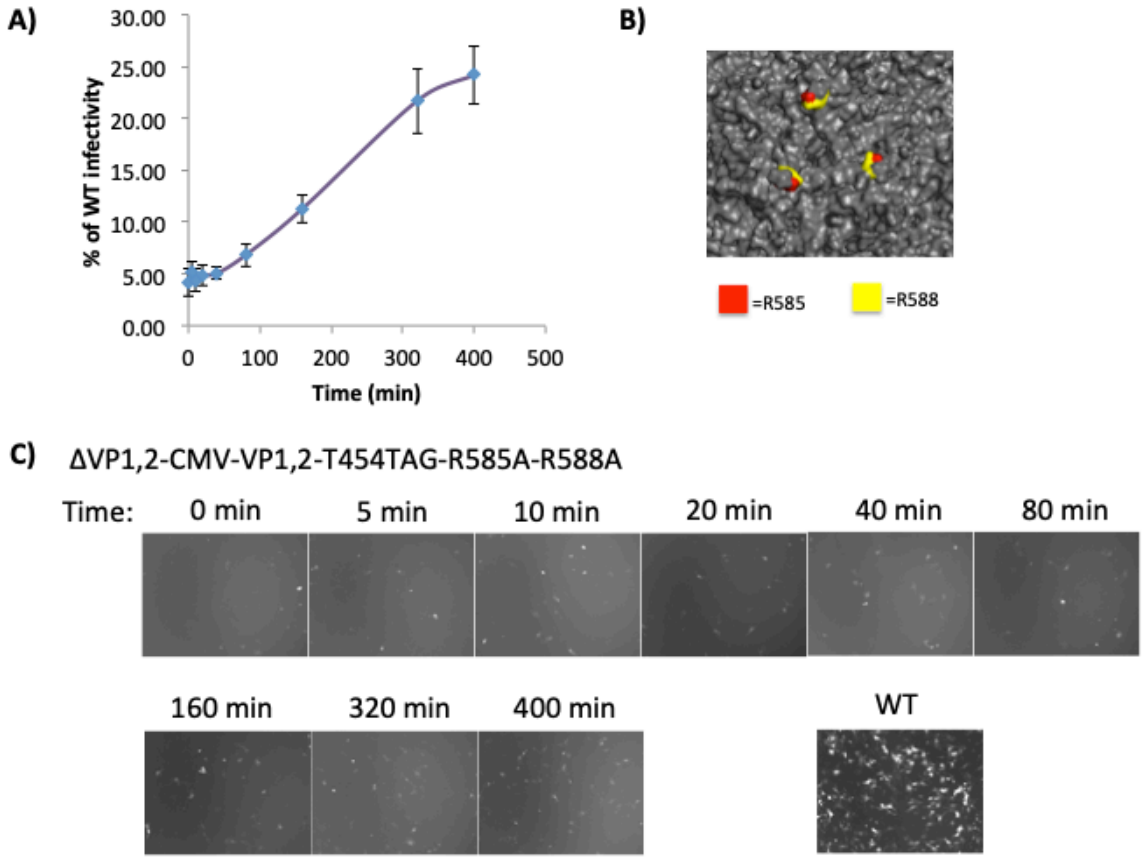


Figure 3.8 .Retargeting efficiency of AAV-VP1, VP2-cRGDFC with varying degrees of cRGDFC modifications. The infectivity of ΔVP1,VP2-VP1-VP2-454AzK AAV over the course of time while being labeled with DBCO-cRGDFC as analyzed by FACS. B) Crystal structure highlighting residues R585 and R588, the two residues that are most important in the HSPG-AAV interaction. C) Fluorescence images corresponding to the graph in A.

3.3 CONCLUSIONS AND FUTURE WORK

In this chapter we have further optimized a functional virus conjugate. We created constructs to selectively incorporate AzK into only VP1, only VP2, or both VP1 and VP2, which allows for the production of a more homogenous preparation of modified AAV capsids. Using these constructs, we show that displaying modifications on VP1 or VP2 only gives an insufficient number of modifications and re-targeting is efficiency is lower than when modifications can be placed on more

than 5 of the VP proteins. Displaying modifications on VP1 and VP2 produces virus with a retargeting efficiency more similar to the efficiency when there are 60 azides available for modification. Lastly, we hypothesize that, based on the kinetics of infectivity gain, more than one interaction between the AAV capsid and the cell must be made in order for successful infection to occur. Although these constructs do not improve upon the efficiency of retargeting, they are an important step towards the optimization of a functional virus conjugate. Future experiments could include the optimization of a linker that enables placing multiple targeting on the capsid, or to incorporate the UAA into multiple sites on the AAV capsid to increase the number of available modification sites. One could also perform experiments to further understand the importance of multiple interactions occurring for successful infection- by placing multiple targeting motifs close in close proximity.

The constructs built in this chapter are useful for more than creating retargeting vehicles. We have since shown that a number of UAAs are poorly tolerated in the AAV capsid (discussed further in chapter 6). Incorporation of a smaller number of these UAAs by selective incorporation could increase virus production greatly. This has been shown with the incorporation of an exposed cysteine in the capsid in chapter 4, which is not a UAA but is poorly tolerated when placed into all 60 capsid proteins, as well as 5-hydroxytryptophan in chapter 6. Future experiments are planned to test how selective incorporation affects virus production with additional UAAs. Additionally, we plan to build constructs that can be used to incorporate multiple distinct UAAs into the capsid by incorporation into

VP1 or VP2 only, for example, one UAA would be expressed only in VP1 while another would be expressed only in VP2.

3.4 ACKNOWLEDGEMENTS

Thanks to the members of Team AAV who worked on this project with me. Rachel Kelemen developed the many of the methods and taught me, Xiaofu Cao helped with some of the cloning, and Zianab Kiyam helped with transfections. Thanks also to Tyler Bechtel and the Weerapana lab for providing SK-OV-3 cells.

3.5 EXPERIMENTAL PROCEDURES

3.5.1 Cell Culture Hek293T cells and SK-OV-3 cells were obtained and maintained as previously described[88].

3.5.2 Cloning and Plasmids pHelper was obtained as previously described[88]. pIDTSmart-ITR-EGFP-ITR-8xPytR-*Mb*PylRS was generated by digesting the previously described pIDTSmart-8xPytR-*Mb*PylRS with restriction enzymes AvrII and NheI. The fragment containing the 8xPytR was isolated and purified by gel purification and inserted into the pIDTSmart-*Mb*PylRS plasmid at the SpeI site. ITR-EGFP-ITR was then inserted into the resulting plasmid using an SbfI site. pIDTSmart-RC2 was made by digesting pIDTSmart-*Mb*PylRS-RC2 with AvrII and NheI and ligating the plasmid back together, removing the *Mb*PylRS region.

pIDTSmart-RC2-ΔVP1-CMV-VP1 was generated as follows. pIDTSmart-RC2-MbPylRS was used as the starting plasmid. To abolish VP1 expression, overlap PCR was performed to mutate the VP1 start codon, M1, to Leucine. Restriction sites HindIII and SbfI were used to replace the modified C*ap* with the M1L version, which resulted in the intermediate plasmid pIDTSmart-RC2-ΔVP1-*Mb*PylRS. A second intermediate plasmid, pIDTSmart-CMV-VP1 was built by halting VP2 and VP3 expression by mutating their start codons. Starting from pIDTSmart-MbPylRS-RC2,

VP2's start codon ACG was mutated to ACC (T138T). To abolish VP3 expression, VP3's start codon was mutated to Leucine, M203L (ATG→CTC). Based on a previous report, eliminating VP3's ATG start codon is not enough to eliminate VP3 expression, so two additional downstream methionines, M211 and M235, were also mutated to Leucine (M211L (ATG→CTC), and M235L (ATG→CTC))[43]. Lastly, two additional non-canonical start codons upstream of M203 were mutated to account for the possibility that VP3 translation might be initiated: T197L (ACG→ACC) and L202L (CTG→CTC). A total of five mutations were made to completely abolish VP3 expression. All mutations were generated by overlap extension PCR, with PCR products digested with NheI and AvrII and ligated into the pIDTSmart-CMV backbone, which was created by digesting pIDTSmart-*Mb*PylRS with NheI and AvrII to remove the *Mb*PylRS region. pIDTSmart-CMV-VP1 was digested with NheI and AvrII and the CMV-VP1 fragment was isolated and incorporated into the pIDTSmart-RC2-ΔVPI-*Mb*PylRS using AvrII and NheI to make pIDTSmart-RC2-ΔVP1-CMV-VP1. Plasmid pIDTSmart-RC2-ΔVP1, which only encodes VP2 and VP3 from the *Cap* ORF was obtained at this step by directly re-ligating the vector after digestion with NheI and AvrII, as they are compatible sticky ends.

pIDTSmart-RC2-ΔVP2-CMV-VP2 was generated by first creating the intermediate plasmid pIDTSmart-RC2-VP2Δ-*Mb*PylRS by making a silent mutation on VP2's start codon, T138T (ACG→ACG). To build pIDTSmart-CMV-VP2, the VP2 start codon, T138 was mutated to ATG, a stranger start codon, and a PCR that amplified DNA starting from the VP2 start codon from pIDTSmart-CMV-VP1 was performed, gel purified, and ligated into the pIDTSmart-CMV backbone. This ensured that the VP1 expression was eliminated and that the VP3 deletions were retained as in pIDTSmart-CMV-VP1. pIDTSmart-RC2-ΔVP2 and pIDTSmart-RC2-ΔVP2-CMV-VP2 were created following the same scheme as the VP1 version.

pIDTSmart-RC2-ΔVP1,VP2 was created by making aforementioned M1L and T138T mutations to abolish VP1 and VP2 expression in the pIDTSmart-RC2 plasmid. Intermediate plasmid pIDTSmart-CMV-VP1-VP2(ATG) was created by making the T138M mutation by overlap PCR in pIDTSmart-RC2-ΔVP3, and PCR amplifying out the entire VP region. pIDTSmart-RC2-ΔVP1,2-*Mb*PylRS was then digested with NheI and AvrII and the *Mb*PylRS region was replaced with the VP1-VP2(ATG)-ΔVP3 region that was amplified out. All mutations can be seen in Chart 1.

From these plasmids, stop codons were incorporated into the final plasmids by performing overlap extension PCR on pIDTSmart-RC2 plasmids using the appropriate terminal primers (SI Table 1) and ligated into the corresponding intermediate plasmid.

3.5.3 Production of UAA containing AAV AAV was produced by transfeting HEK293T at 70% confluency in either a 12-well or 10cm dish with plasmids containing AAV genes, necessary helper genes, and the pyrrolysyl tRNA/synthetase pair in the presence of 1mM AzK (H-L-Lys(EO-N3)-OH, Iris biotech GMBH). AAV was harvested about 72 hours after transfection. The Takara AAVPro Extraction solution kis was used to harvest virus from 12-well plates. To harvest virus from 10cm dishes, cells underwent two rounds of freeze/thaws with a dry-ice ethanol bath and a 37˚C water bath followed by centrifugation at 5,000xg for 5 minutes to pellet cellular debris.

This was followed by Polyethylene Glycol (PEG 8000, Fisher Bioreagents) precipitation overnight. AAV from 10cm dishes was then pelleted by centrifugation at 12,000xg for 30 minutes and resuspended in 10% glycerol in PBS and stored at -80˚C for use.

3.5.4 Purification of AAV2 by AVB resin Each 10cm dish was seeded with 8 million HEK293T cells. Cells were transfected as described at 75% confluency, and three days after transfection cells freeze-thawed as described above, followed by treatment with universal nuclease for 10 minutes at room temperature. 250µL of settled AVB resin (GE healthcare) was used per 10cm dish. Resin was equilibrated with 10 column volumes of PBS and lysate was applied to the equilibrated resin. Lysate was allowed to flow through the column two times, and the resin bed was washed with 25 column volumes of PBS. Virus was eluted using 6 column volumes of 0.1M glycine pH 2.8, and each elution fraction was immediately neutralized using 10% 1M Tris pH 8. The virus was precipitated using 11% PEG-8000 at 4˚C overnight. PEG-precipitated virus was pelleted by centrifugation at 1200xg for 30 minutes and re-suspended in 25µL 10% glycerol in PBS.

3.5.5 Purification of AAV2 by Heparin agarose Heparin purifications were performed as previously described[88].

3.5.6 AAV Titration by qPCR Experiments performed with virus that was harvested from a 12 well dish were titered using the AAVPro titration kit (Takara). Virus from 10cm dishes was harvested by first performing a DNAse I digestion where any DNA from lysate was digested. Viral DNA remained protected in the capsid. Viral particles were then opened by heat denaturation at 90˚C for ten minutes. qPCR was performed using the PowerUP SYBR Green Mastermix (Thermo Fisher) using the following primers:
CMV qPCR Fwd Primer: 5'-CATCTACGTATTAGTCATCGCTATTACCT-3'
CMV qPCR Rev Primer: 5'-GAAATCCCCGTGAGTCAAACC-3'.

3.5.7 Assaying the infectivity of UAA-AAV Infectivity was assayed as previously described[88].

3.5.8 Cy5 and TAMRA labeling, SDS Page, and Western Blotting Approximately 10^9 genome copies of purified virus (by AVB resin or heparin sulfate affinity resin) were incubated 20uM DBCO-Cy5 or 50uM DBCO-TAMRA (Sigma-Aldrich) at room temperature for 30 minutes. Excess dye was quenched with free Azido-Lysine and the reaction was dialyzed against PBS overnight at 4˚C to remove excess dye. Viral capsid proteins were then resolved by SDS-PAGE gel. Cy5 and TAMRA fluorescence were imaged with the Cy5 and rhodamine settings on the ChemiDoc MP imaging system (BioRad). For visualizing with Coomassie, then the SDS-PAGE gel was then stained overnight with staining buffer and destained the following morning. For visualizing with silver stain, then gel was shaken in fixation solution 1 (30% ethanol, 10% acetic acid, in ddH_2O) for 1 hour in a new container and replaced with fresh fixation solution1 overnight. The following day, the gel was rinsed 3 times for 5

minutes in fixation solution II (30% ethanol in ddH_2O). The gel was rehydrated in ddH_2O at room temperature for 20 minutes followed by incubation in 2.5mM sodium thiosulfate at room temperature with gentle agitation for two minutes. Sodium thiosulfate was removed and the gel was washed twice for ten minutes with ddH_2O. Next, the gel was incubated in staining solution (0.2%$AgNO_3$, 28uL of 37% formaldehyde stock in ddH_2O to 35mL) at room temperature with gentle agitation for 30 minutes. The gel was then rinsed with ddH_2O for one minute and revelation solution (1 g sodium carbonate, 17uL of 37% stock, and 2.5mM sodium thiosulfate in ddH_2O to 35mL) was added with agitation until bands were visible when glacial acetic acid was added to stop development. Resulting gel was rinsed with water and imaged on the ChemiDoc MP imaging system using silver stain settings.

3.5.9 DBCO-cRGDFC synthesis DBCO-RGDFC was synthesized as previously described[88]. The reaction completion was verified using LCMS (Supplemental Figure 1).

3.5.10 Re-targeting and Flow Cytometry

Re-targeting a Flow cytometry were performed as described in Chapter 2.

3.6 REFERENCES

(1) Ponnazhagan, S.; Mahendra, G.; Kumar, S.; Thompson, J. A.; Castillas, M. Conjugate-Based Targeting of Recombinant Adeno-Associated Virus Type 2 Vectors by Using Avidin-Linked Ligands. *J. Virol.* **2002**, *76* (24), 12900–12907. https://doi.org/10.1128/jvi.76.24.12900-12907.2002.

(2) Horowitz, E. D.; Weinberg, M. S.; Asokan, A. Glycated AAV Vectors: Chemical Redirection of Viral Tissue Tropism. *Bioconjug. Chem.* **2011**, *22* (4), 529–532. https://doi.org/10.1021/bc100477g.

(3) Lee, S.; Ahn, H. J. Anti-EpCAM-Conjugated Adeno-Associated Virus Serotype 2 for Systemic Delivery of EGFR ShRNA: Its Retargeting and Antitumor Effects on OVCAR3 Ovarian Cancer in Vivo. *Acta Biomater.* **2019**, *91*, 258–269. https://doi.org/10.1016/j.actbio.2019.04.044.

(4) Wei, F.; McConnell, K. I.; Yu, T. K.; Suh, J. Conjugation of Paclitaxel on Adeno-Associated Virus (AAV) Nanoparticles for Co-Delivery of Genes and Drugs. *Eur. J. Pharm. Sci.* **2012**, *46* (3), 167–172. https://doi.org/10.1016/j.ejps.2012.02.022.

(5) Yang, Q.; Mamounas, M.; Yu, G.; Kennedy, S.; Leaker, B.; Mersn, J.; Yu, F. W. M. A. N. G.; Barber, J. R.; Al, Y. E. T. Development of Novel Cell Surface CD34-Targeted Recombinant Adenoassociated Virus for Gene Therapy. **1998**, *1937*, 1929–1937.

(6) Wu, P.; Xiao, W.; Conlon, T.; Hughes, J.; Agbandje-McKenna, M.; Ferkol, T.;

Flotte, T.; Muzyczka, N. Mutational Analysis of the Adeno-Associated Virus Type 2 (AAV2) Capsid Gene and Construction of AAV2 Vectors with Altered Tropism. *J. Virol.* **2000**, *74* (18), 8635–8647. https://doi.org/10.1128/JVI.74.18.8635-8647.2000.

(7) Jr, K. H. W.; Gorbatyuk, O. S.; Jeffrey, K.; Opie, S. R.; Zolotukhin, S.; Warrington, K. H.; Harrison, J. K.; Muzyczka, N. Adeno-Associated Virus Type 2 VP2 Capsid Protein Is Nonessential and Can Tolerate Large Peptide Insertions at Its N Terminus. *J. Virol.* **2004**, *78* (12), 6595–6609. https://doi.org/10.1128/JVI.78.12.6595.

(8) Girod, A.; Wobus, C. E.; Zádori, Z.; Ried, M.; Leike, K.; Tijssen, P.; Kleinschmidt, J. A.; Hallek, M. The VP1 Capsid Protein of Adeno-Associated Virus Type 2 Is Carrying a Phospholipase A2 Domain Required for Virus Infectivity. *J. Gen. Virol.* **2002**, *83* (5), 973–978. https://doi.org/10.1099/0022-1317-83-5-973.

(9) Kronenberg, S.; Bottcher, B.; von der Lieth, C. W.; Bleker, S.; Kleinschmidt, J. A. A Conformational Change in the Adeno-Associated Virus Type 2 Capsid Leads to the Exposure of Hidden VP1 N Termini. *J. Virol.* **2005**, *79* (9), 5296–5303. https://doi.org/10.1128/jvi.79.9.5296-5303.2005.

(10) Grieger, J. C.; Snowdy, S.; Samulski, R. J. Separate Basic Region Motifs within the Adeno-Associated Virus Capsid Proteins Are Essential for Infectivity and Assembly. *J. Virol.* **2006**, *80* (11), 5199–5210. https://doi.org/10.1128/jvi.02723-05.

(11) Lux, K.; Goerlitz, N.; Schlemminger, S.; Perabo, L.; Goldnau, D.; Endell, J.; Leike, K.; Kofler, D. M.; Finke, S.; Hallek, M.; et al. Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking. *J. Virol.* **2005**, *79* (18), 11776–11787. https://doi.org/10.1128/jvi.79.18.11776-11787.2005.

(12) Münch, R. C.; Janicki, H.; Völker, I.; Rasbach, A.; Hallek, M.; Büning, H.; Buchholz, C. J. Displaying High-Affinity Ligands on Adeno-Associated Viral Vectors Enables Tumor Cell-Specific and Safe Gene Transfer. *Mol. Ther.* **2013**, *21* (1), 109–118. https://doi.org/10.1038/mt.2012.186.

(13) Münch, R. C.; Muth, A.; Muik, A.; Friedel, T.; Schmatz, J.; Dreier, B.; Trkola, A.; Plückthun, A.; Büning, H.; Buchholz, C. J. Off-Target-Free Gene Delivery by Affinity-Purified Receptor-Targeted Viral Vectors. *Nat. Commun.* **2015**, *6*, 2–10. https://doi.org/10.1038/ncomms7246.

(14) Hagen, S.; Baumann, T.; Wagner, H. J.; Morath, V.; Kaufmann, B.; Fischer, A.; Bergmann, S.; Schindler, P.; Arndt, K. M.; Müller, K. M. Modular Adeno-Associated Virus (RAAV) Vectors Used for Cellular Virus-Directed Enzyme Prodrug Therapy. *Sci. Rep.* **2014**, *4*, 1–11. https://doi.org/10.1038/srep03759.

(15) Asokan, A.; Johnson, J. S.; Li, C.; Samulski, R. J. Bioluminescent Virion Shells: New Tools for Quantitation of AAV Vector Dynamics in Cells and Live Animals. *Gene Ther.* **2008**, *15* (24), 1618–1622. https://doi.org/10.1038/gt.2008.127.

(16) Muik, A.; Reul, J.; Friedel, T.; Muth, A.; Hartmann, K. P.; Schneider, I. C.; Münch, R. C.; Buchholz, C. J. Covalent Coupling of High-Affinity Ligands to the Surface of Viral Vector Particles by Protein Trans-Splicing Mediates Cell Type-Specific Gene Transfer. *Biomaterials* **2017**, *144*, 84–94.

https://doi.org/10.1016/j.biomaterials.2017.07.032.
(17) Thadani, N. N.; Yang, J.; Moyo, B.; Lee, C. M.; Chen, M. Y.; Bao, G.; Suh, J. Site-Specific Post-Translational Surface Modification of Adeno-Associated Virus Vectors Using Leucine Zippers. *ACS Synth. Biol.* **2020**, acssynbio.9b00341. https://doi.org/10.1021/acssynbio.9b00341.
(18) Zdechlik, A. C.; He, Y.; Aird, E. J.; Gordon, W. R.; Schmidt, D. Programmable Assembly of Adeno-Associated Virus-Antibody Composites for Receptor-Mediated Gene Delivery. *Bioconjug. Chem.* **2019**. https://doi.org/10.1021/acs.bioconjchem.9b00790.
(19) Snijder, J.; Van De Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-Mckenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (20), 7295–7299. https://doi.org/10.1021/ja502616y.
(20) Agbandje-McKenna, M.; Kleinschmidt, J. *AAV Capsid Structure and Cell Interactions*; 2011; Vol. 807. https://doi.org/10.1007/978-1-61779-370-7.
(21) Kelemen, R. E.; Mukherjee, R.; Cao, X.; Erickson, S. B.; Zheng, Y.; Chatterjee, A. A Precise Chemical Strategy To Alter the Receptor Specificity of the Adeno-Associated Virus. *Angew. Chemie - Int. Ed.* **2016**, *55* (36), 10645–10649. https://doi.org/10.1002/anie.201604067

4.0 DISPLAYING CYSTEINE ON THE AAV CAPSID SURFACE AS A NEW CONJUGATION STRATEGY

This thesis has discussed the many ways that modifications have been made on the AAV capsid, and highlighted their strengths and weaknesses. Besides UAA incorporation, all methods used to date lack the ability to site specifically modify the capsid. Cysteine conjugation is a popular biochemical method of labeling proteins that, like UAA incorporation, can be used to site-specifically label proteins. However only recently have any studies attempted to use cysteine to modify the AAV capsid. We show that the incorporation of a surface exposed cysteine into all three VP proteins is poorly tolerated and does produce sufficient infectious virus. We then incorporate cysteine into exposed sites in only VP1 or only VP2 and show, that by reducing the number of cysteine residues placed into the capsid, we can produce near wild type titers of AAV that can be modified using cysteine-labeling strategies.

4.1 INTRODUCTION

4.1.1 Cysteine conjugation methods

AAV capsid modification strategies, discussed in chapter 1, are all routinely used to modify less complex proteins. However, not all common protein modification strategies have proven amenable to modifying AAV. For example, the modification of surface exposed cysteine residues is one of the most popular ways to make Antibody Drug Conjugates (ADCs). To date, there are only a small number of studies that have used cysteine as a way to modify the AAV capsid. In this chapter, I discuss work that I have done to create a platform for modifying the AAV capsid with a surface exposed cysteine residue.

Labeling strategies exist for other natural amino acids such as lysine, however, because lysine appears in proteins so frequently it is difficult to modify a single site in a homogeneous manner. Cysteine residues occur much less frequently, so it's more likely that there are no accessible cysteine residues in a protein. Because of this, incorporation of cysteine is one way to achieve homogenous labeling at a single site. While homogeneous labeling can also be achieved using UAA incorporation, no excess machinery is needed to incorporate cysteine, making it a more facile conjugation strategy.

Protein modification strategies using cysteine have been developed extensively[177,178]. Two of the most commonly used methods in biological applications are the use of iodoacetamides and maleimides. Iodoacetamides react with cysteine residues in a substitution reaction with iodide (Figure 1A). There is

the potential cross-reactivity with other nucleophilic side chains. Maleimides are Michael acceptors (Figure 1B) and form thiosuccinimide bonds with the cysteine. Maleimide labeling is advantageous as it has high selectivity towards thiol groups, there is a lack of side products, and it can be performed in aqueous media without catalysts. One disadvantage of maleimide chemistry is that retro-michael addition can take place, which means that this reaction has some reversibility.

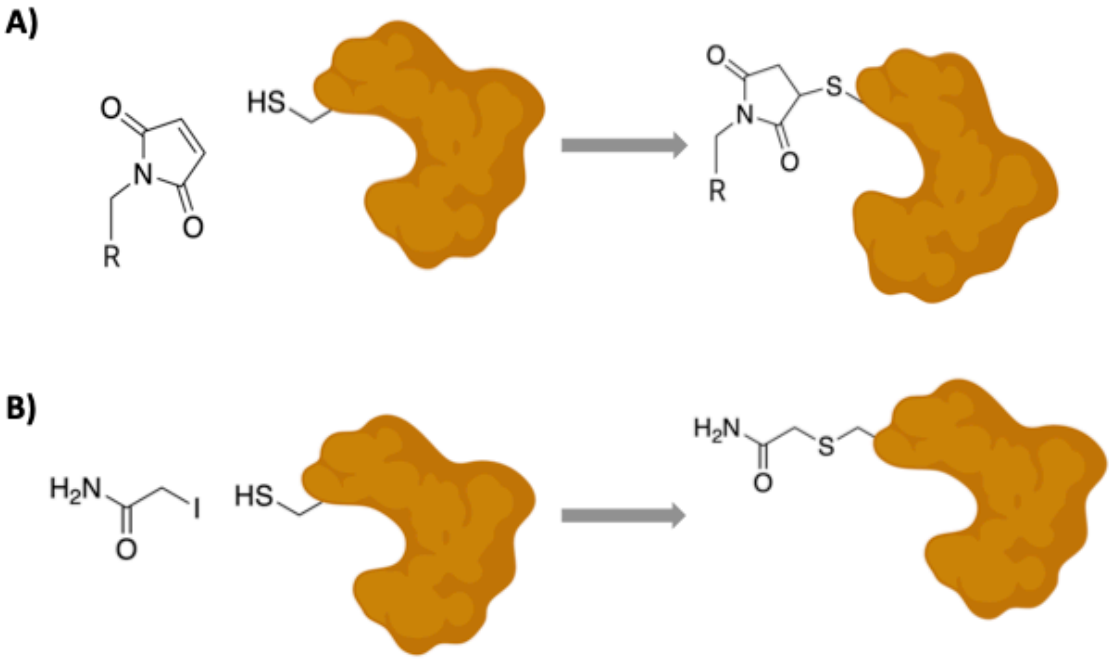


Figure 4.1. The reactions between cysteine residues and maleimides (A) or iodoacetamide (B).

The ability to incorporate surface-exposed cysteine residues on the AAV capsid would introduce an attractive new method to site-specifically alter the AAV2 capsid. While it would maintain some of the advantages that UAA incorporation has, such as the ability to site-specifically modify the capsid, it would have the added advantage that no additional machinery would need to be used to produce the capsid. Despite these advantages, there have been very few studies that have successfully used cysteine to modify the AAV capsid.

4.1.2 Natural cysteine residues in the AAV capsid

The AAV capsid has 5 highly conserved cysteine residues in its capsid, C230, C289, C361, C394, and C482. None of these cysteine residues are surface exposed (Figure 2) and cysteine to serine or valine mutations suggest that none of these residues form disulfide bonds[179]. Because none of these cysteine residues are exposed, the addition of an accessible cysteine is one way to site-specifically label a single residue in a VP protein on the AAV capsid.

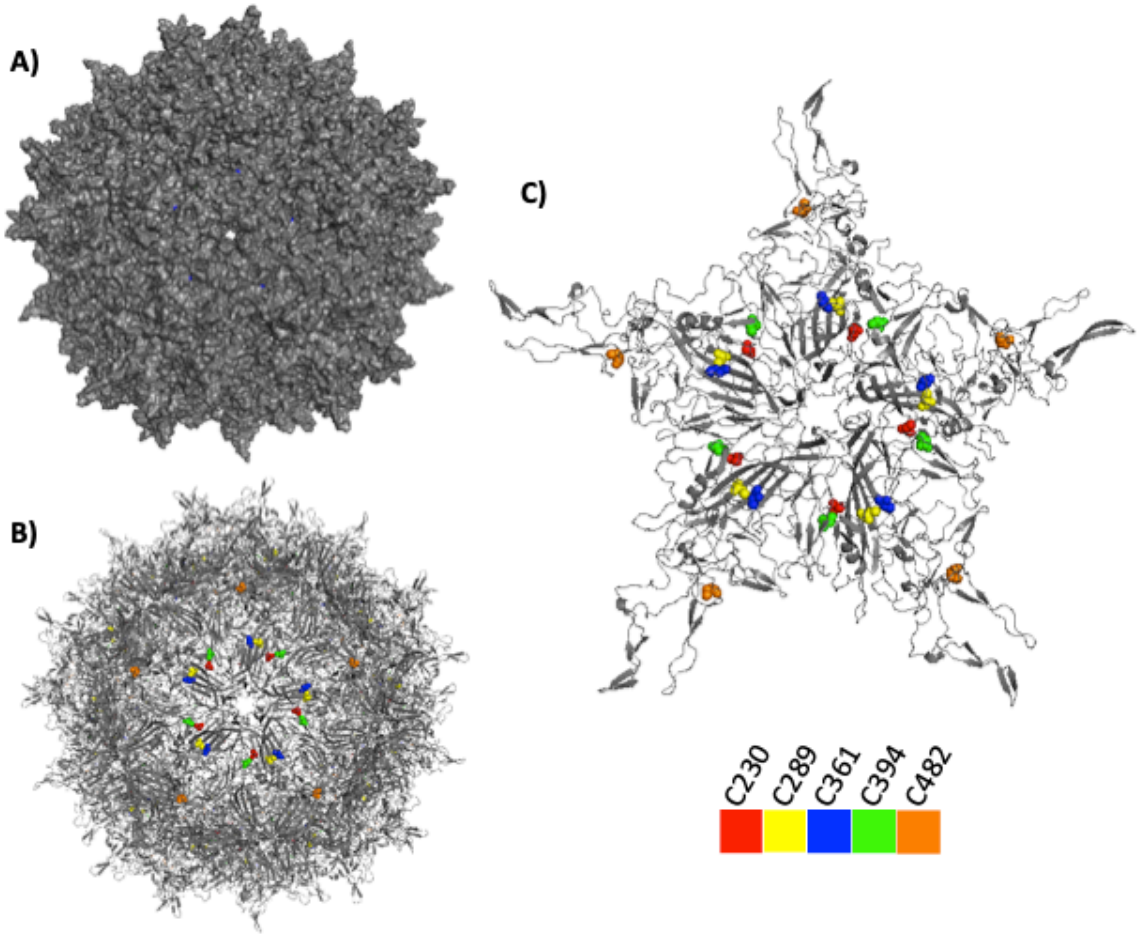


Figure 4.2 Crystal structure of the AAV2 capsid with all cysteine residues highlighted. A) Surface view, showing the surface exposed cysteine residues. B) Cartoon view of entire capsid showing orientation of the cysteine residues. C) Cartoon view of the five fold pore, showing orientation of cysteine residues.

4.1.3 An overview of studies that have successfully incorporated cysteine into the AAV Capsid

There are studies such as those that introduce a targeting motif into the VP3 region that use cysteine as a way to close the peptide "loop" by adding cysteine residues to the base of the loop that will form disulfide bonds[59], however these do not use the cysteine residues for further modification. Another study genetically inserted a 13-amino acid sequence (LCTPSRAALLTGR) into the AAV2 capsid following residue 587 in VP1, VP2, and VP3[65]. Production of AAV containing this peptide was evaluated using qPCR and was found to be comparable to wild type capsids. The functional titer, which was monitored using infection of HEK293T cells with a GFP reporter gene, was also found to be similar to that of a wild type capsid. Following insertion of the peptide, they used formylglycine-generating enzyme (FGE) to convert the cysteine residue to an aldehyde-bearing formylglycine residue. They then conjugated gold particles, an antibody, and a cRGD-targeting moiety to the capsid using linkers with a terminal amine group that they converted to a hydrazide using succinimidyl 4-hydrazinonicotinate acetone hydrazine (SANH). A third study inserted a 12 amino acid sequence that contained a tetra-cysteine motif (HRWCCPGCCKTF) into the AAV9 capsid[180]. Roger Tsien described this sequence for use with FlASH/ReAsH compounds, however rather than using FlASH/ReAsH compounds, the tetra-cysteine motif was reduced and then conjugated using a maleimide compound[181]. Chandran et. al. attempted to introduce this sequence into four locations on the capsid surface; into VP1 only at residue 34, into the disordered region between VP1 and VP2 at residue 138, and in all three VP proteins at residues

583 or 589. Infectious virus was only formed when the sequence was inserted at residue 138. AAV containing the tetra-cysteine motif was conjugated to a fluorescent probe and used to label and image the brain in mice. The then biotinylated the capsid and used that to probe AAV's interactome. The tetracysteine motif described has been used in one more study[8], however cysteine incorporation into the AAV capsid has not been widely developed. Additionally, in each of the studies mentioned above, the cysteine is added in a peptide, which does not allow for precise control over the site of modification.

We have developed a strategy to site specifically incorporate a cysteine residue into the AAV capsid. This is an exciting new way to modify the AAV capsid using cysteine-labeling chemistry that is already well established. Additionally, there is no cross-reactivity with the click-chemistry reactions that we have previously used to modify the capsid, and would be good way to place multiple distinct modifications on the AAV capsid.

4.2 RESULTS AND DISCUSSION

4.2.1 Production of cysteine-AAV

The cysteine residue was placed at site T454 in either all three VP proteins, only VP1, or only VP2 (Figure 3). AAV-Cys was produced using a triple transfection in HEK293T cells. Each transfection used pHeler, pAAV-ITR-GFP (or, in some cases

pAAV-ITR-mCherry), and either pIDTSmart-RC2-T454C, pIDTSmart-RC2-ΔVP1-CMV-VP1-T454C, or pIDTSmart-RC2-ΔVP1-CMV-VP1-T454C.

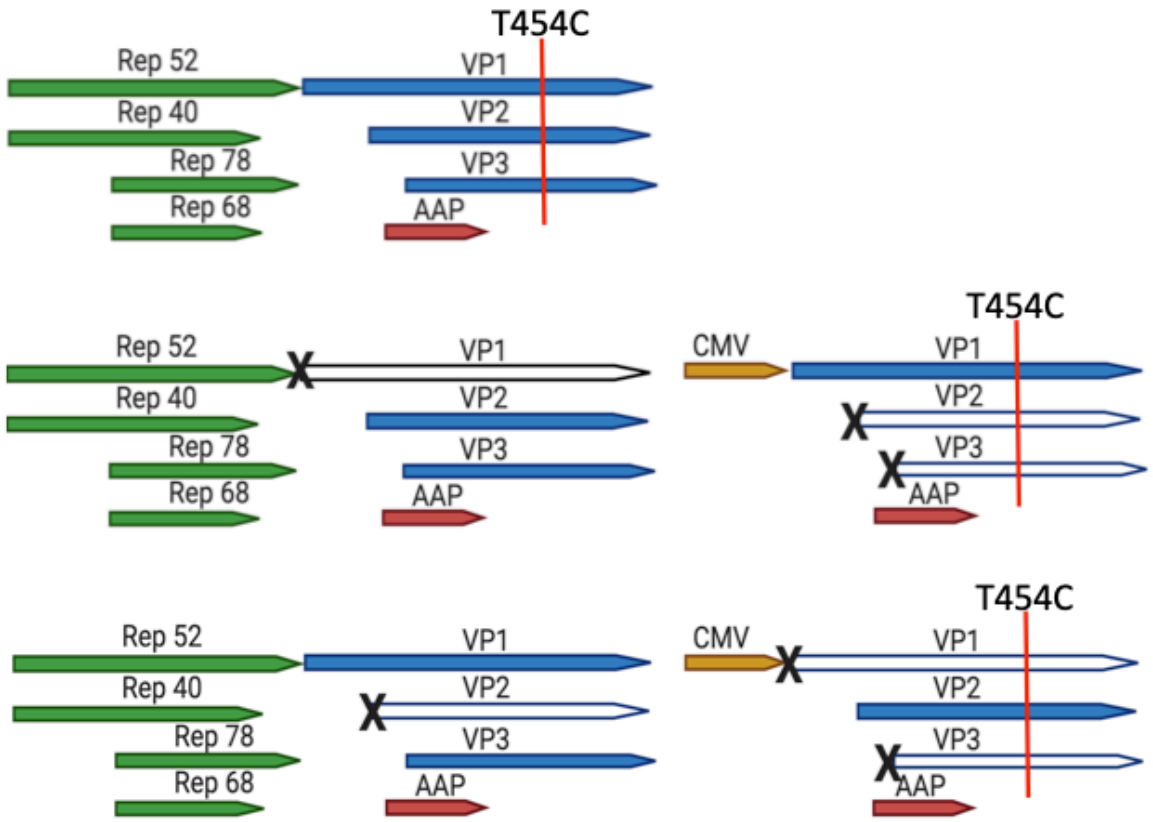


Figure 4.3 Constructs where T454C is expressed in all three VP proteins, or selectively in only VP1 or only VP2.

4.2.2 Characterization of cysteine-AAV

Production of AAV-Cys was assayed using qPCR (Figure 4). AAV with cysteine incorporated into all three capsid proteins resulted in a 5-fold hit in production as assayed by qPCR, however only a small portion of this proved to be infectious. Introduction of cysteine into only VP1 or VP2 resulted in a two-fold or less hit in viral production.

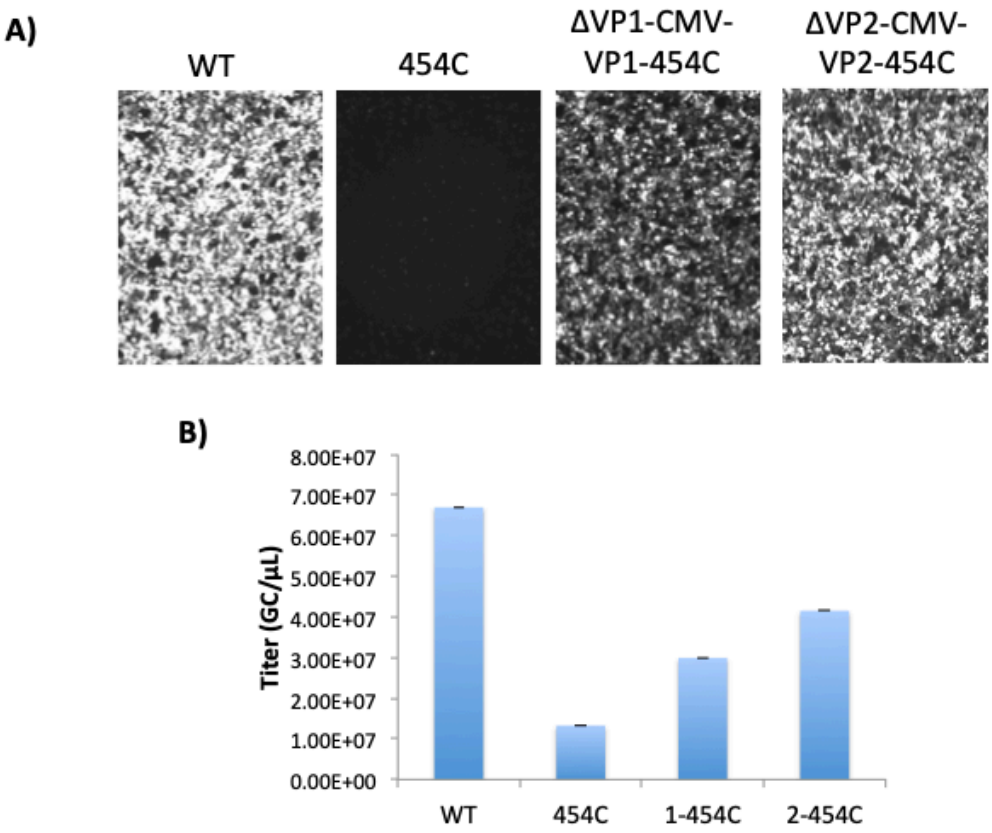


Figure 4.4. Production efficiency of AAV with a surface exposed cysteine residue. A) Constant volume infection using AAV containing cysteine at T454 in various capsid proteins. B) qPCR of AAV from A.

Infection with a constant MOI followed by FACS analysis revealed that this virus maintained wild type like infectivity (Figure 5). The results revealing that incorporation of a cysteine into all three VP proteins produces very little infectious virus lends insight as to why cysteine has not been widely used to modify the AAV capsid. While it is not totally clear why cysteine incorporation causes such a hit to the production of infectious AAV, we hypothesize that the formation of disulfide bonds between VP proteins could interfere with capsid formation. We also hypothesized that disulfide bonds could be forming between capsid proteins, which could be hindering infection, however reduction of AAV-cysteine prior to infection did not improve infection (Figure 6).

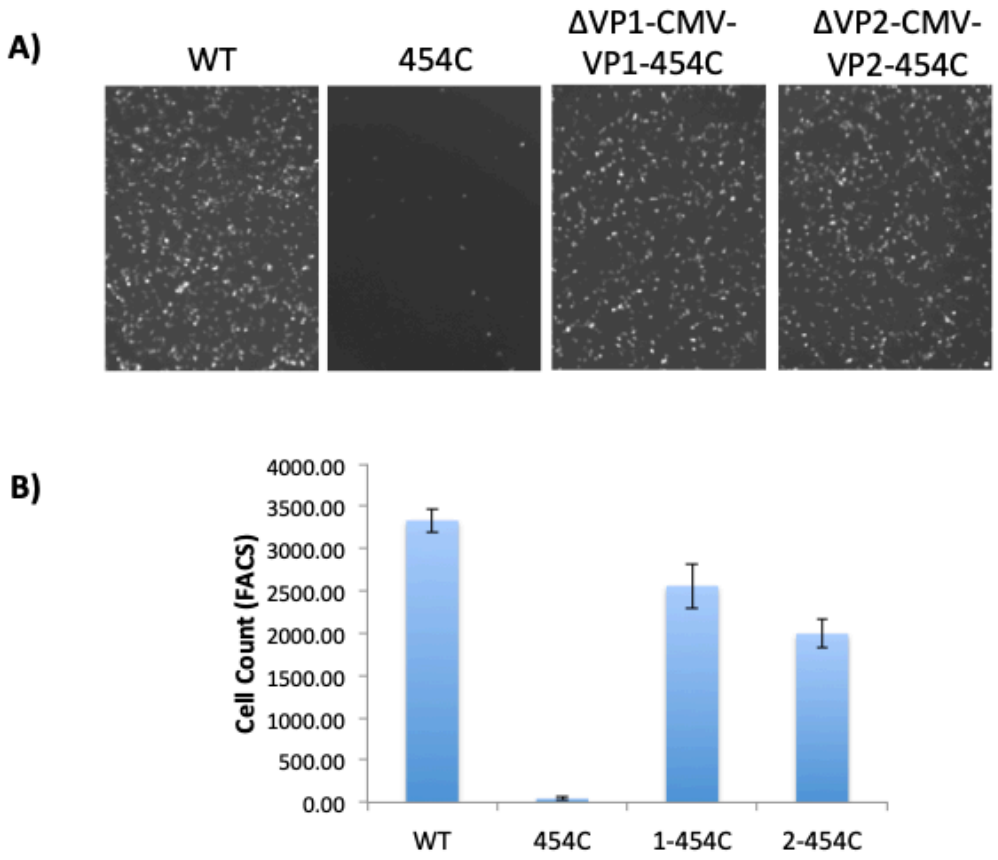


Figure 4.5. Infectivity of AAV with a surface exposed cysteine residue. A) Constant titer infection (20gc/cell) of AAV-T454C from qPCR in figure 3 analyzed by A) microscopy or B) FACS analysis.

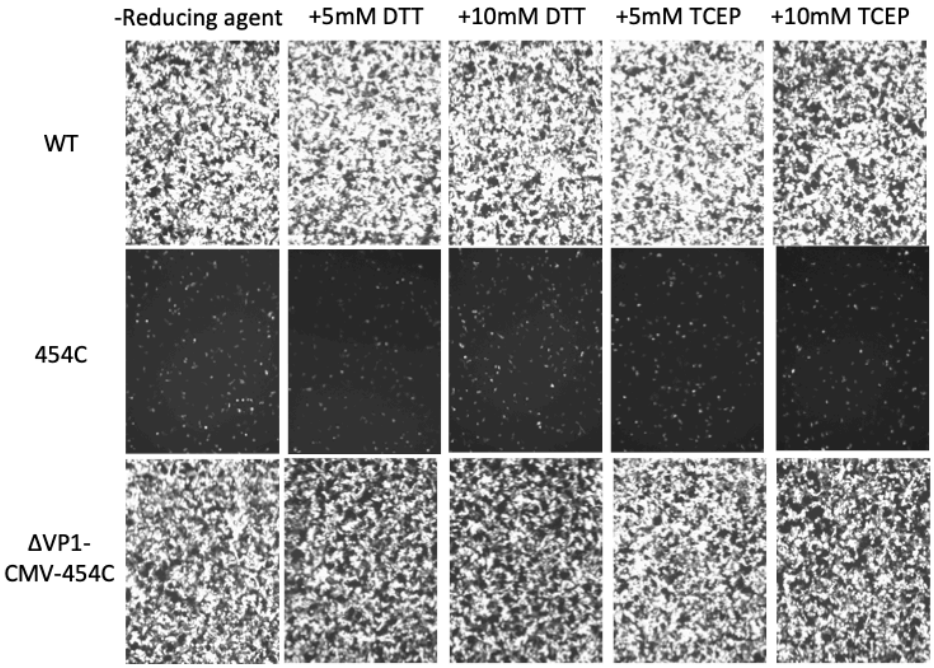


Figure 4.6. The effect of reducing agents on AAV-cys infectivity. Constant volume infection of AAV-T454C with either DTT or TCEP to see if reducing the surface exposed cysteine would increase infectivity.

4.2.3 Labeling cysteine-AAV

To confirm that the cysteine was incorporated into the capsid in the correct VP protein and that the VP proteins were present at the correct levels, both wild type and VP1-T454C AAV were labeled with Cy5.5-maleimide and analyzed by SDS-Page. Fluorescence imaging showed that only VP1 was labeled and imaging using SYPRO Red showed that bands are present in a 1:1:10 ratio (Figure 7).

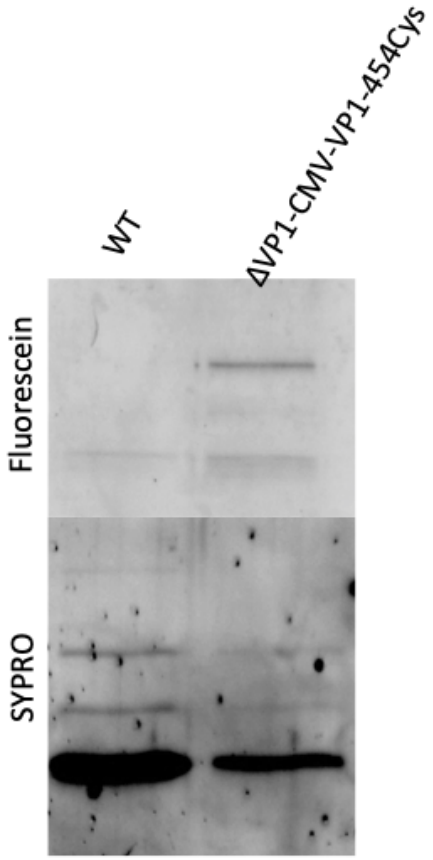


Figure 4.7. SDS-PAGE analysis of wild type or ΔVP1-CMV-VP1-454Cys AAV with Cy5.5-maleimide conjugation, visualized with either Fluorescein or SYPRO orange.

4.3 CONCLUSIONS AND FUTRUE WORK

We have demonstrated that incorporation of a surface exposed cysteine residue into all 60 capsid proteins results in very low titers of non-infectious virus. However we also demonstrated that placing a surface exposed cysteine residue into only VP1 or only VP2 results in wild type-like production of AAV-Cys. We also showed that AAV-Cys retains wild type like levels of infectivity and that the cysteine can act as a chemical handle to modify the AAV capsid.

One potential future direction for this project would be scanning the capsid to see which sites exposed cysteine residues are tolerated at. Additionally, some sort of proof of concept experiment is needed. A possible proof of concept would be to use this as a way to attach a protein to the outside of the AAV capsid. This is a direction that I was working on, but was not able to finish (discussed in Chapter 6).

As mentioned, a cysteine residue would also be useful for conjugating multiple different modifications to the AAV capsid. For example, UAA incorporation could be used to incorporate a photo-crosslinking amino acid, and the cysteine could be used to attach a handle for subsequent pull down. While this could be accomplished using the incorporation of multiple UAAs, the use of a cysteine residue decreases the amount of additional machinery that has to be transfected into the cell. Using a single UAA and a Cysteine residue rather than 2 UAAs would mean that there wouldn't need to be two synthetase/tRNA pairs transformed into the cells.

4.4 ACKNOWLEDGEMENTS

Thanks to Dan Bak and Jenny Peeler from the Weerapana lab for helping me optimize the cysteine purification and labeling protocol.

4.5 EXPERIMENTAL PROCEURES

Cell Culture

Hek293T and SK-OV-3 cells were obtained and cultured as previously described[1]

Cloning and Plasmids

Starting plasmids were obtained as previously described.

qPCR

Experiments were performed with virus that was harvested from a 10cm dish and were titered using the AAVPro titration kit (Takara). Briefly, Hek293T cells were harvested and lysed using a lysis buffer. Lysate was clarified and DNAse I digestion was used to remove any DNA from the lysate, with the AAV capsid protecting the AAV genome. DNAseI was denatured using heat denaturation at 90˚C for ten minutes, and the capsid was further opened using Takara Bio's lysis buffer and incubation at 70˚C. The qPCR reaction was set up as directed using the procedure recommended by the TAKARA bio kit.

AVB Purification of AAV

AVB resin was purchased from GE Healthcare. One 10cm dish was seeded with 8 million cells. 24 hours later, the cells were transfected with PEI. Two days later, cells were scraped off the 10cm dish and pelleted using centrifugation at 4,500xg for 10 minutes. The pelleted cells were either flash frozen and stored in the -80˚C freezer or immediately re-suspended in 1xPBS and lysed by freezing in a dry-ice/ethanol bath and thawing in the 37˚C water bath two times. Cellular debris was pelleted by centrifugation at 5,000xg for 5 minutes. Lysate was then treated with Universal nuclease for 10 minutes followed by incubation with 5mM DTT for 45 minutes at room temperature with gentle agitation. 250µL of settled resin was used per 10cm dish. Resin was washed with 10CV of PBS+1mM DTT and lysate was applied to equilibrated resin. Lysate was allowed to flow through the column two times, and the resin bed was washed with 25 column volumes of PBS+1mM DTT. Virus was eluted using 6 column volumes of 0.1M glycine pH 2.8+1mM DTT, and

each elution fraction was immediately neutralized using 10% 1M Tris pH 8. The virus was precipitated using 11% Poly ethylene-glycol (PEG-8000, Fisher Bioreagents, BP233) at 4˚C overnight. PEG-precipitated virus was pelleted by centrifugation at 1200xg for 30 minutes and re-suspended in 25µL 10% glycerol in PBS.

Cy5.5-maleimde labeling and SDS Page

Purified AAV-Cys was incubated with 500mM DTT for 30 minutes. DTT was removed using a P-30 spin column that had been buffer exchanged to PBS. Immediately after desalting, 20µM Cy5.5-maleimide was added to virus. The reaction proceeded for 30 minutes with gentle agitation, and was quenched with 100mM DTT. The mixture was placed into a dialysis cup and dialyzed against PBS overnight to remove excess Cy5.5-maleimide and DTT. 1x SDS loading buffer was mixed with the samples after dialysis and samples were boiled at 95˚C for 5 minutes and then run on a 10% SDS Page gel.

Re-targeting and Flow Cytometry

Confluent HEK293T cells were seeded with a constant MOI of 20gc/cell of wild type or AAV-Cys virus with 5mM sodium butyrate (Sigma-Aldrich) to enhance expression of AAV encoded transgenes. 48 hours post infection, infectivity was visualized by EGFP expression using a Zeiss Axio Observer fluorescence microscope with an XCite Series 120Q light source and Zeiss filter 44 (excitation 475/40nm, beamsplitter 500nm, emission 530/50nm). Cells were prepped and FACS was performed as previously described in Chapter 2.

4.6 REFERENCES

(1) Gunnoo, S. B.; Madder, A. Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17* (7), 529–553. https://doi.org/10.1002/cbic.201500667.

(2) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. - An Asian J.* **2009**, *4* (5), 630–640. https://doi.org/10.1002/asia.200800427.

(3) Pulicherla, N.; Kota, P.; Dokholyan, N. V.; Asokan, A. Intra- and Inter-Subunit Disulfide Bond Formation Is Nonessential in Adeno-Associated Viral Capsids. *PLoS One* **2012**, *7* (2), 1–6. https://doi.org/10.1371/journal.pone.0032163.

(4) Shi, W.; Bartlett, J. S. RGD Inclusion in VP3 Provides Adeno-Associated Virus Type 2 (AAV2)-Based Vectors with a Heparan Sulfate-Independent Cell Entry Mechanism. *Mol. Ther.* **2003**, *7* (4), 515–525. https://doi.org/10.1016/S1525-0016(03)00042-X.

(5) Liu, Y.; Fang, Y.; Zhou, Y.; Zandi, E.; Lee, C. L.; Joo, K. Il; Wang, P. Site-Specific Modification of Adeno-Associated Viruses via a Genetically Engineered

Aldehyde Tag. *Small* **2013**, *9* (3), 421–429. https://doi.org/10.1002/smll.201201661.
(6) Chandran, J. S.; Sharp, P. S.; Karyka, E.; Aves-Cruzeiro, J. M. da C.; Coldicott, I.; Castelli, L.; Hautbergue, G.; Collins, M. O.; Azzouz, M. Site Specific Modification of Adeno-Associated Virus Enables Both Fluorescent Imaging of Viral Particles and Characterization of the Capsid Interactome. *Sci. Rep.* **2017**, *7* (1), 14766. https://doi.org/10.1038/s41598-017-15255-2.
(7) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *124* (21), 6063–6076. https://doi.org/10.1021/ja017687n.
(8) Seo, J. W.; Ingham, E. S.; Mahakian, L.; Tumbale, S.; Wu, B.; Aghevlian, S.; Shams, S.; Baikoghli, M.; Jain, P.; Ding, X.; et al. Positron Emission Tomography Imaging of Novel AAV Capsids Maps Rapid Brain Accumulation. *Nat. Commun.* **2020**, *11* (1), 1–13. https://doi.org/10.1038/s41467-020-15818-4.

5.0 THE INTRODUCTION OF PHOTOCAGED-LYSINE RESIDUES INTO THE AAV CAPSID TO OBTAIN OPTOCHEMICAL CONTROL OVER CELLULAR INFECTION

Incorporation of photocaged-UAAs into the AAV capsid provides the ability to streamline a biological process. This could be particularly useful for studying events in the AAV infection cycle. For example, photocaging the nuclear localization signals present on the VP proteins, would allow control over whether or not the VP proteins were transported into the nucleus. Another example could be obstructing the interaction between the capsid and AAP so that these events could be observed at a controlled pace, rather than trying to observe each capsid making the interaction at a different time. We chose to incorporate a photocaged-lysine into the AAV2 capsid at residues that block AAV's primary binding interaction with HSPG, rendering AAV non-infectious. Upon shining light on the AAV capsid, the bulky group blocking the interaction is removed and AAV regains infectivity (Figure 1).

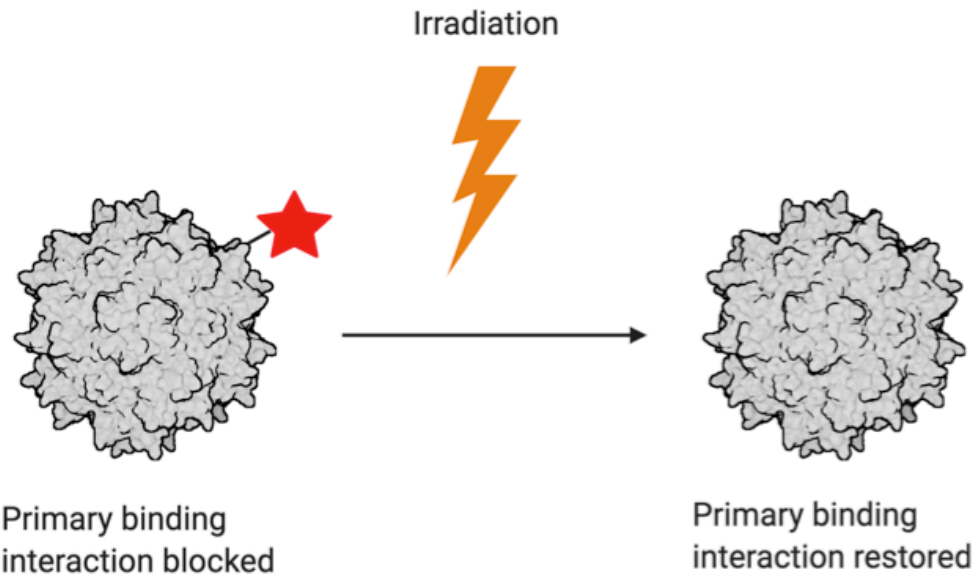


Figure 5.1. Cartoon depicting AAV with a photocaged primary binding interaction. The incorporation of a photocaged-UAA can block AAV's primary binding interaction, rendering it non-infectious. Upon light irradiation, the perturbation is removed, restoring infectivity.

5.1 INTRODUCTION

5.1.1 Photocaged amino acids that can have been incorporated into proteins

As previously discussed in section (1.3.3), one class of UAAs that can be incorporated using genetic code expansion is photocaged amino acids. These amino acids have a caging group that, when exposed to light of a specific wavelength, undergo a chemical reaction that leaves the native amino acid and the removal of the caging group. Photocaged analogs of cysteine[1–4], tyrosine[5–7], serine[8], and lysine[9–11] have been incorporated into *E. coli* or mammalian cells. Recently, genetic code expansion machinery has also been evolved for photocaged amino acids such as homocysteine[3], DOPA[12], and selenocysteine[13,14] as well (Figure 2).

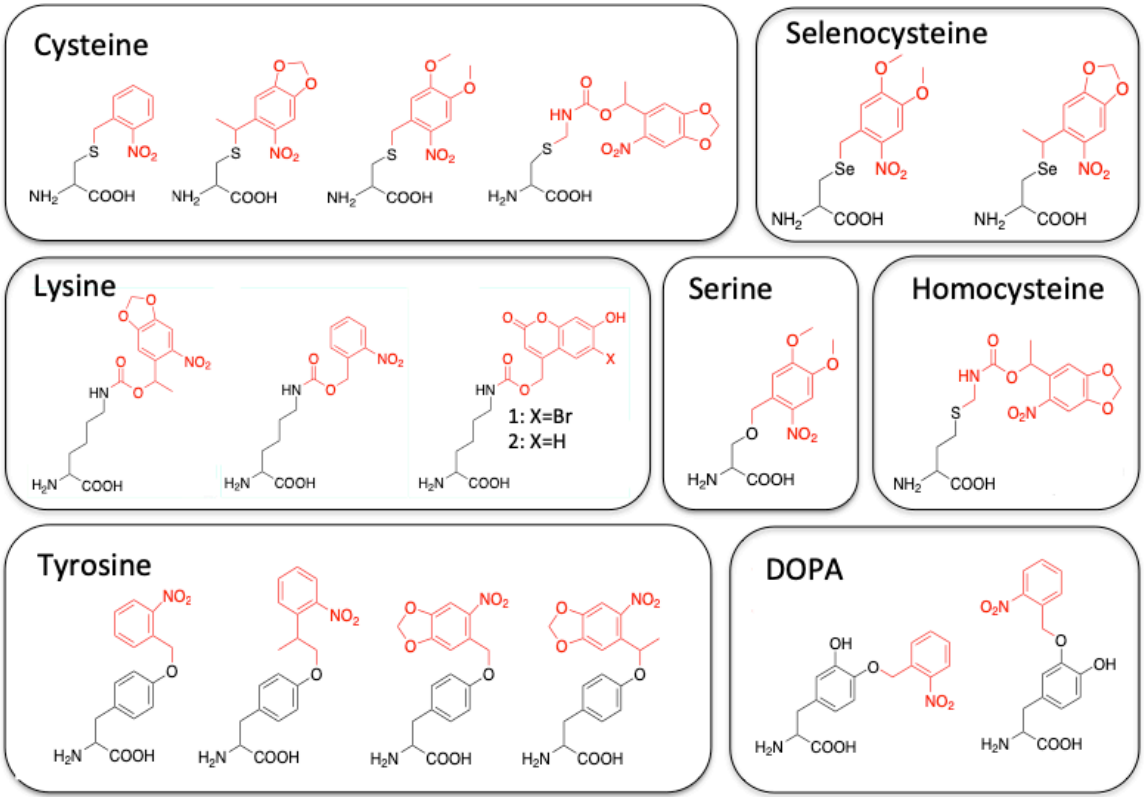


Figure 5.2. Photocaged analogues of amino acids that have can be incorporated into proteins using genetic code expansion. The photocaged portions that are removed upon irradiation are highlighted in red.

5.1.2 Photocaged lysine analogs that have been incorporated into proteins

Several different photocaged lysine residues have been incorporated using genetic code expansion (Figure 2). The *Mm*PylRS/tRNA^{Pyl} pair was evolved to incorporate N$^{\varepsilon}$-cyclopentyloxycarbonyl-L-Lysine (ONBK) in both *E. coli* and mammalian cells. Machinery for a second photocaged lysine, PCK, was evolved at around the same time. PCK (designated as NBK in this work) is advantageous over ONBK for two reasons; it can be decaged using nonphototoxic light, which prevents cellular damage such as nucleic acid damage, and the photolysis product generates a ketone byproduct that does not undergo condensation reactions with the ε-amino group of lysine that the byproduct of ONBK does. PCK can be incorporated using

PCKRS, an evolved *Mb*PylRS/tRNA^{CUA} pair. Two additional coumarin lysine residues have also been incorporated and have the advantage of containing a benzopyrone core that can act as a fluorescent probe in addition to having the ability to function as a photocaged amino acid.

In this work, we chose to incorporate two different photocaged lysine residues into the AAV capsid: NBK and one of the Coumarin lysine analogues. The Chin lab evolved the *Mb*PylRS/tRNA$_{CUA}$ pair for NBK incorporation in 2010. Five positions in *Mb*PylRS, M241, A267, Y271, L274, and C313, were randomized and after three rounds of positive and negative selections in *E. coli*, the most active synthetase contained the mutations M241F, A267S, Y271C, and L274M. NBK was then incorporated into GFP to verify that it was incorporating correctly, as well as a nuclear localization signal (NLS) of the molecular chaperone nucleoplasmin. The photocaged NLS was used it to control the nuclear import of p53, a tumor suppressor. NBK has been used in a variety of studies since its GCE machinery was developed, including incorporation into systems such as additional NLS sequences, CRISPR/Cas9, kinases, the T7 RNA polymerase, and Cre recombinase enzymes[9,15–22].

Several coumarin lysine analogues have had GCE machinery evolved for their incorporation, and two of these analogues have the distinct advantage that they can act both as a fluorescent probe as well as a photocaging amino acid. The first coumarin lysine analogue to be incorporated could only act as a fluorescent probe[23], but by altering the structure, it was possible to create an analogue that decages with UV light, and one that decages using near IR excitation. While these analogues are structurally similar, modification of the carbamate linker by removing a carbon

atom enables photocaging with UV (this analog is termed HCK), and the further placement of a bromine at the 6-position of the benzopyrone core (termed BHCK) enables decaging using near IR[11]. The *Mb*PylRS synthetase was evolved for coumarin lysine incorporation, and the final synthetase, termed BhcKRS, incorporates all three analogues of coumarin lysine and contains only two mutations, Y271A and L274M. Using this synthetase, Coumarin lysine was placed into an N-terminal NLS that delivered EGFP to the nucleus of HEK cells using optical control, and into the chromophore of EGFP to control its fluorescence. Coumarin lysine has since been used in applications such as controlling cellular ATP levels using a photocaged adenylate kinase, the optical control of protein phosphatase function, the optical control of DNA helicase, and has been incorporated into zebrafish embryos to track MEK activation at different stages in development[24–27]. In this work, we sought to incorporate NBK and coumarin lysine into the AAV capsid to control its primary binding interaction with HSPG and, by extension, its ability to infect cells.

5.2 RESULTS AND DISCUSSION

5.2.1 NBK incorporation into GFP

Xiaofu Cao, an undergrad in the Chatterjee lab at the time, made the previously described mutations for the NBKRS, M241, A267, Y271, L274, and C313, in the *Mb*PylRS using overlap PCR. Additionally, Raja Mukherjee and Chester Wrobel

synthesized NBK as previously described[9]. We decided to first test the incorporation of NBK in EGFP, before incorporating into the AAV capsid. Plasmids bearing NBKRS, 8x copies of the *Mm*PytR, and EGFP-TAG39 were transfected into HEK293T cells in the presence and absence of NBK. Incorporation efficiency was observed using microscopy (Figure 3A). The cells were harvested two days post transfection and EGFP-Y39NBK was purified using nickel affinity resin. Purified EGFP-Y39NBK was verified by LCMS. Next, the decaging efficiency of EGFP-Y39NBK was tested by irradiation using 365nm light for 30s and the resulting protein was evaluated by LCMS. LCMS results showed the expected mass for purified EGFP-Y39NBK and revealed that NBK was completely decaged after 30s of irradiation (Figure 3B).

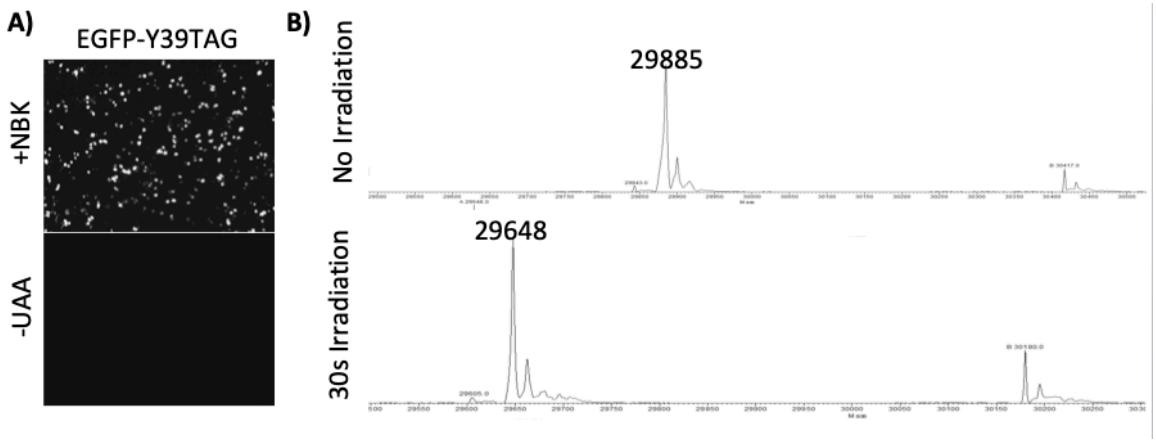


Figure 5.3. EGFP production with NBK. A) EGFP-Y39TAG production in the presence or absence of NBK. B) LCMS trace of EGFP-NBK verifying the presence of the photocaging group in the absence of irradiation and the absence of the photocaging group after exposure to irradiation with 365nm light for 30s.

5.2.2 Verification that lysine can functionally replace arginine at positions R585 and R588

As previously described in chapter (1.1.3), AAV's primary binding interaction is made between heparan sulfate proteoglycan (HSPG, Figure 4C) and a cluster of arginine and lysine residues on the AAV capsid. Using alanine scanning, R585 and

R588 have been shown to be essential in this interaction, and R484, R487, and K532 have been shown to be involved in the interaction but not essential (Figure 4, A and B)[28,29]. Because of the essential nature of R585 and R588 in AAV2's infectivity, we chose to incorporate NBK at these positions, to create an AAV capsid whose infectivity could be optically controlled. Because no photocaged arginine residues have been genetically encoded into eukaryotes, we first investigated the possibility of using a photocaged lysine at these positions, as mutation to lysine conserves the positive charge. Previous studies report that substitution of R585 or R588 to lysine maintains the interaction[30]. We individually mutated R585 and R588 to lysine residues using overlap PCR and evaluated production efficiency using qPCR, which showed that both mutant viruses were produced at wild type levels (Figure 4D). Infectivity of R585K or R588K AAV was evaluated by infecting HEK293T cells at a constant MOI of 50gc/cell. Three days after infection, infectivity was evaluated by fluorescence microscopy as well as by FACS using the EGFP reporter packaged in the rAAV cargo. R585K showed 30% infectivity, and R588K 70% infectivity relative to wild type infectivity (Figure 4E), agreeing with previous results that mutations of R585 or R588 to lysine does not significantly alter the viral entry into cells.

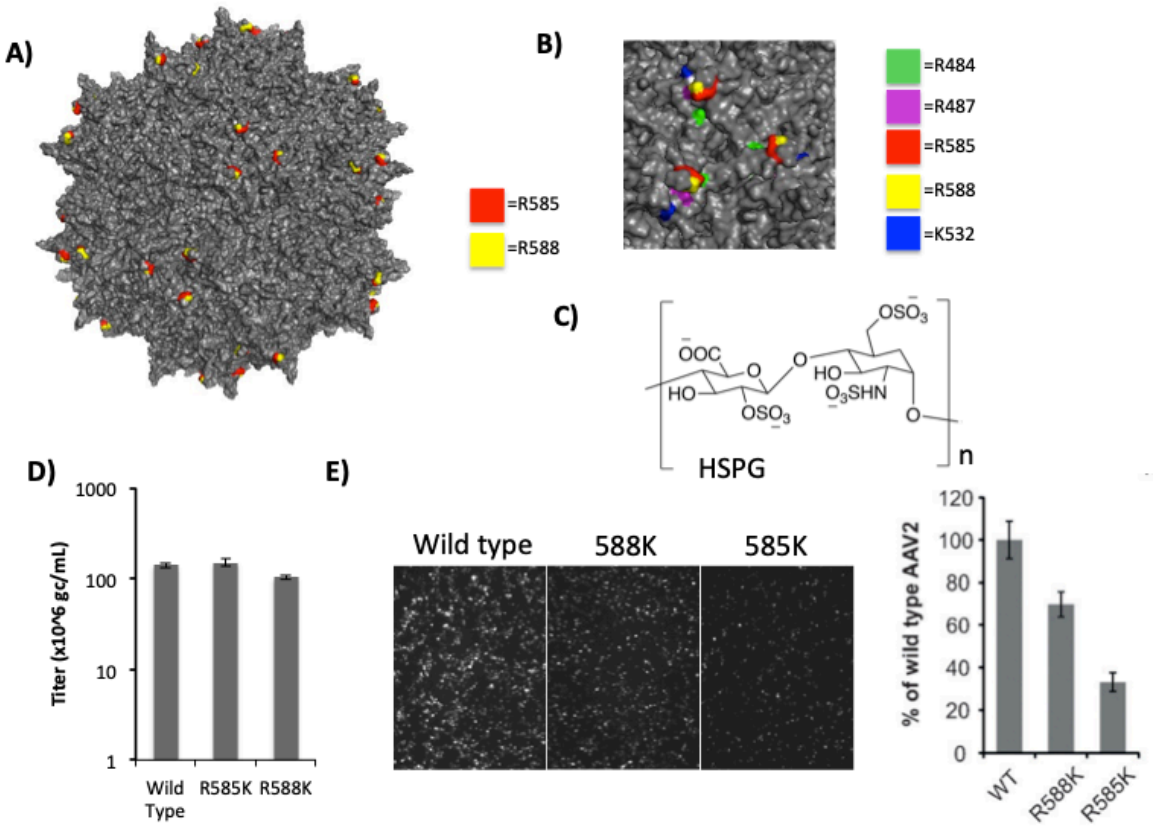


Figure 5.4. AAV's primary binding interaction with HSPG is conserved with lysine residues replacing R585 and R588. A) Crystal structure of AAV highlighting residues R585 and R 588, which are primarily responsible in the interaction with HSPG. B) Crystal structure of AAV highlighting all residues involved in HSPG binding. C) Structure of Heparan Sulfate Proteoglycan. D) qPCR titers of wild type, R585K, and R588K virus. E) Constant titer infection analyzed by microscopy and EGFP fluorescence of wild type AAV and the R588K and R585K mutants. Figure 5.4 E is taken with permission from reference 31[31]. © 2017 Wiley-VCH Verlag GmbH &Co. KGaA, Weinheim

5.2.3 NBK incorporation into AAV

R585TAG and R588TAG constructs were made by overlap PCR and AAV was made by transient transfection in HEK293T cells in both the presence and absence of NBK. A western blot of the resulting virus showed that capsid protein expression is considerably higher in the presence of NBK than in the absence (Figure 5A). qPCR was also used to confirm the production of packaged AAV capsids containing NBK (Figure 5B). Infection of HEK293T cells at a constant MOI of 50gc/cell of R585NBK or R588NBK without irradiation showed very little infection compared to wild type

or the lysine analogues. However irradiation at 365 nm of R585NBK and R588NBK virus for 60s prior to infection restored infectivity (Figure 5C, D). R588NBK infectivity was restored to about 80% of the wild-type infectivity, whereas R585NBK infectivity was only restored to about 30% of wild type infectivity. While it is not totally clear why R585NBK infectivity was restored at a lower level than R588NBK upon irradiation, it is possible that the qPCR assay slightly underestimated the number of packaged genome copies due to the low production efficiency. Alternatively, there might be a subtle reason why infectivity is restored to a lower level due to the way that R585NBK is packaged.

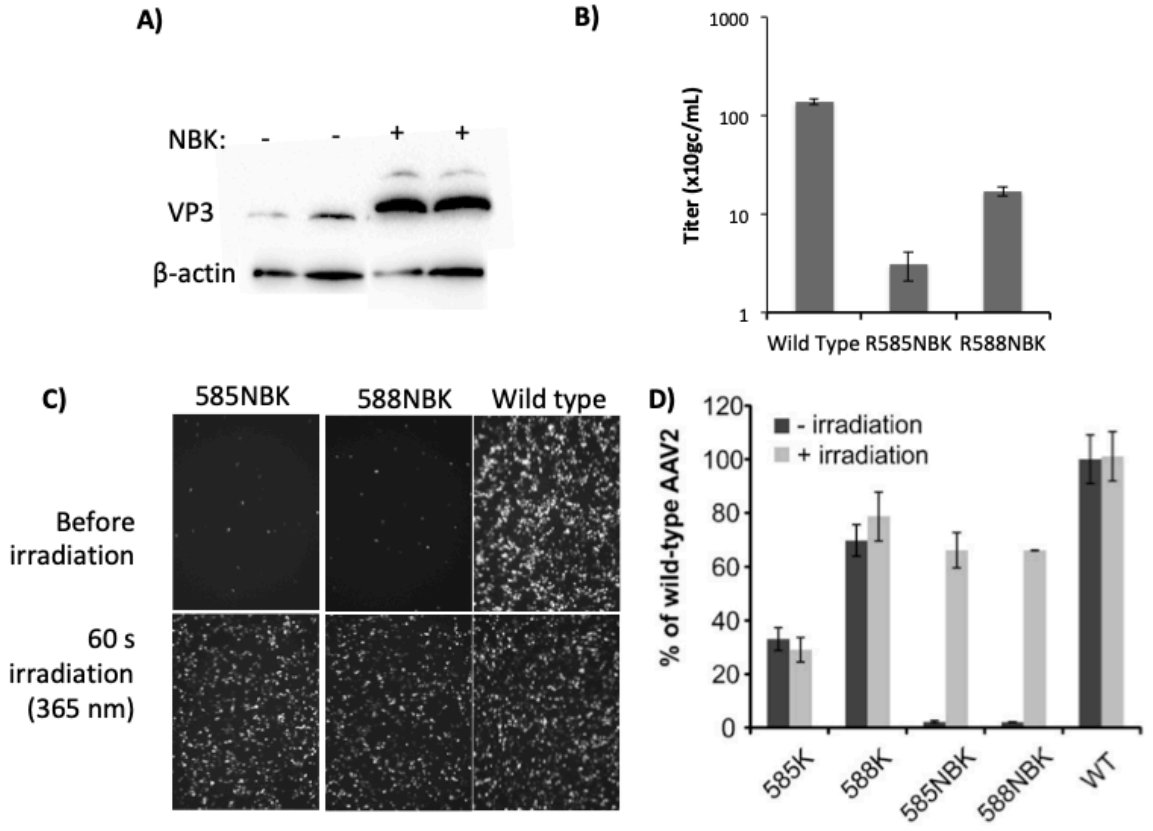


Figure 5.5. AAV's primary binding interaction can be controlled using NBK. A) Western blot showing AAV VP capsid protein production in the presence and absence of NBK. B) qPCR titers showing the production efficiency of wild type or AAV-NBK mutants. C) Constant titer infection of wild type or AAV-NBK virus with or without irradiation. D) FACS analysis corresponding to images in C, as well as the R585K and R588K mutants. Figure 5.5 is taken with permission from reference 31[31]. © 2017 Wiley-VCH Verlag GmbH &Co. KGaA, Weinheim

To further demonstrate that the incorporation of NBK was interfering with the interaction between AAV and HSPG, AAV preparations were passed through a column of heparin agarose. The amount of virus that was retained by the column was measured by performing qPCR on samples before and after passing them through the column (Figure 6). 94% of wild type AAV2 was found to be retained by the column, whereas the R585A-R588A double mutant, which won't bind to the column, had 26% retention. Prior to irradiation, 22% of the R588NBK mutant was retained, similar to the R585A-R588A double mutant. Upon irradiation with 365nm light for 60 seconds, the R588NBK was retained to 80%, showing that the interaction was perturbed in the presence of the caging group but restored upon the removal of the caging group.

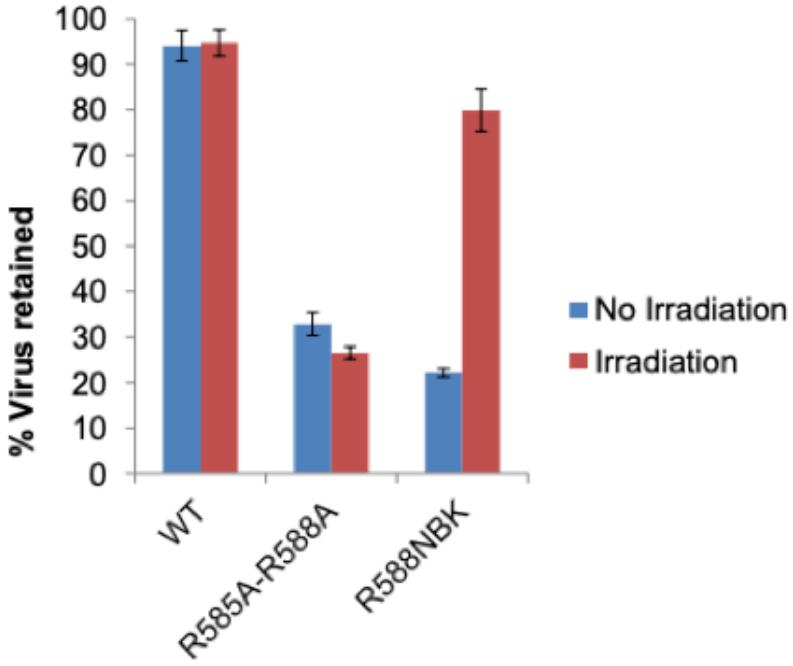


Figure 5.6. Incorporation of NBK perturbs HSPG binding, as demonstrated using heparin agarose. The % of virus retained by a heparin agarose column, demonstrating each virus's capability to bind to heparin. Figure 5.6 is taken with permission from reference 31[31]. © 2017 Wiley-VCH Verlag GmbH &Co. KGaA, Weinheim

5.2.4 Conclusions

Overall, this work demonstrates the feasibility of regulating viral entry into human cells using a genetically encoded photocaged amino acid. We first demonstrated that lysine can functionally replace arginine, retaining the interaction between AAV and HSPG. We next incorporated NBK into the AAV capsid at R585 and R588 and showed that NBK at these positions can be used to control the timing of viral entry into cells. Lastly we showed that the AAV-NBK interaction with Heparin can also be demonstrated using heparin agarose.

5.3 COUMARIN-LYSINE INCORPORATION INTO THE AAV CAPSID

5.3.1 Coumarin-lysine incorporation into GFP

Zianab Kiyam, an undergraduate in the Chatterjee lab, and Partha Addy, a post-doc in the Chatterjee lab, synthesized the HCK analogue of coumarin-lysine (referred to as coumarinK for the rest of this chapter) and used overlap PCR to put the previously described mutations into the *Mb*PylRS (Y271A and L274M) for its incorporation. As with NBK, coumarin lysine incorporation was first tested by incorporating it into EGFP-Y39TAG in HEK293T cells. HEK293T cells were transfected with plasmids bearing the BhcKRS, the *Mm*PytR, and EGFP-Y39TAG in the presence or absence of coumarinK. As a control, a transfection with the NBKRS in the presence and absence of NBK was also set up. Two days post transfection, EGFP-Y39-coumarinK and EGFP-Y39NBK production was analyzed using

fluorescence microscopy (Figure 7A). EGFP-Y39coumarinK was purified nickel affinity resin and its mass was analyzed by LCMS. CoumarinK was decaged using irradiation with 405nm light for 60s, and the efficiency of decaging was monitored by LCMS. The caging group on coumarinK in EGFP-Y39coumarinK was completely removed after 60s of irradiation (Figure 7B).

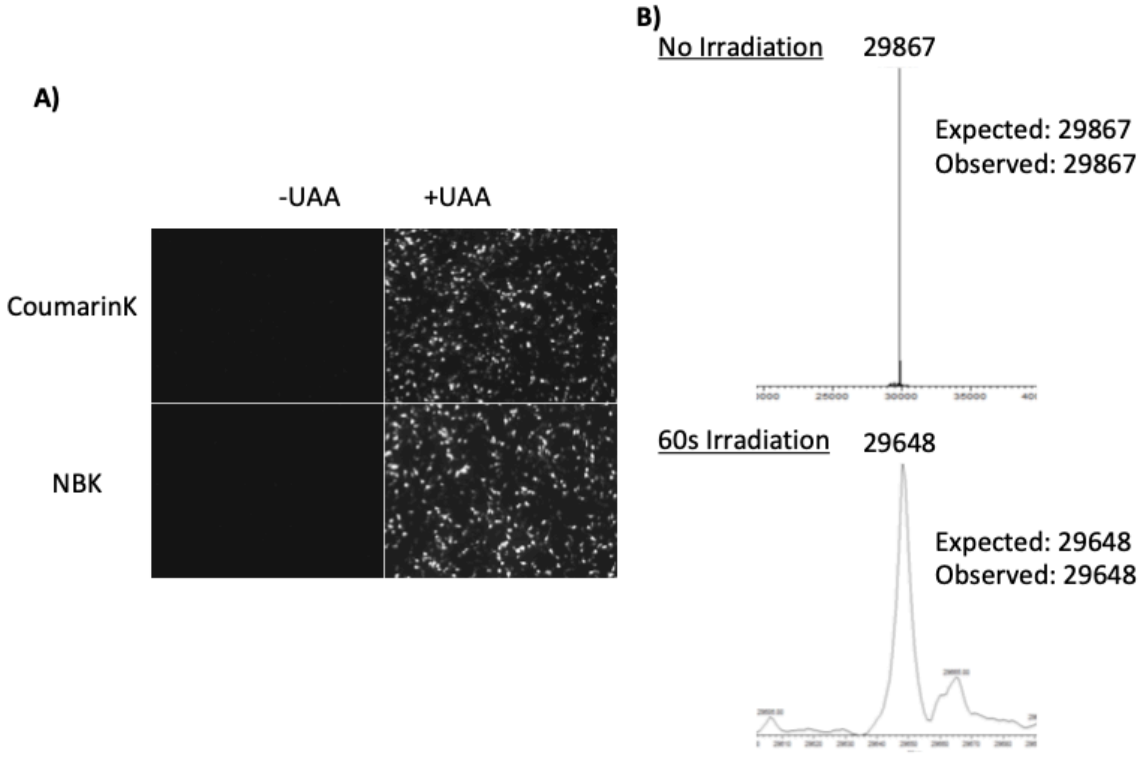


Figure 5.7. EGFP production with CoumarinK. A) Microscopic images of EGFP-Y39TAG transfection in the presence or absence of UAA. B) LCMS analysis of purified EGFP-Y39-coumarinK with or without irradiation at 405nm, confirming the presence of the photocaging without irradiation, and the removal of the photocaging group upon irradiation.

The BhcKRS was moved into the plasmid that also bears the AAV2 RC2 to test incorporation of coumarinK into the AAV2 capsid at position R585. R585-coumarinK AAV was harvested three days post transfection and HEK293T cells were subsequently infected with the harvested virus that had undergone no irradiation, 30s irradiation, or 60s irradiation (Figure 8). As expected, there was no

infectious virus produced in the absence of coumarinK, with or without irradiation. Interestingly, there was very little production of infectious virus made in the presence of coumarinK with irradiation as well. Additionally, the addition of increasing amounts of coumarinK did not affect virus production. Because the AAV capsid is such a complicated assembly of proteins, it is hard to know why NBK was efficiently incorporated but coumarinK was not, especially because both UAA's have a similar size. Further experiments could be performed in order to understand why AAV-R588coumarinK was not produced, such as a western blot to analyze capsid protein production, and qPCR to evaluate whether or not capsids that were capable of protecting the viral genome were produced, however these experiments have not yet been performed. Several more UAA's that we tried to incorporate but did not produce virus will also be discussed in chapter 6.

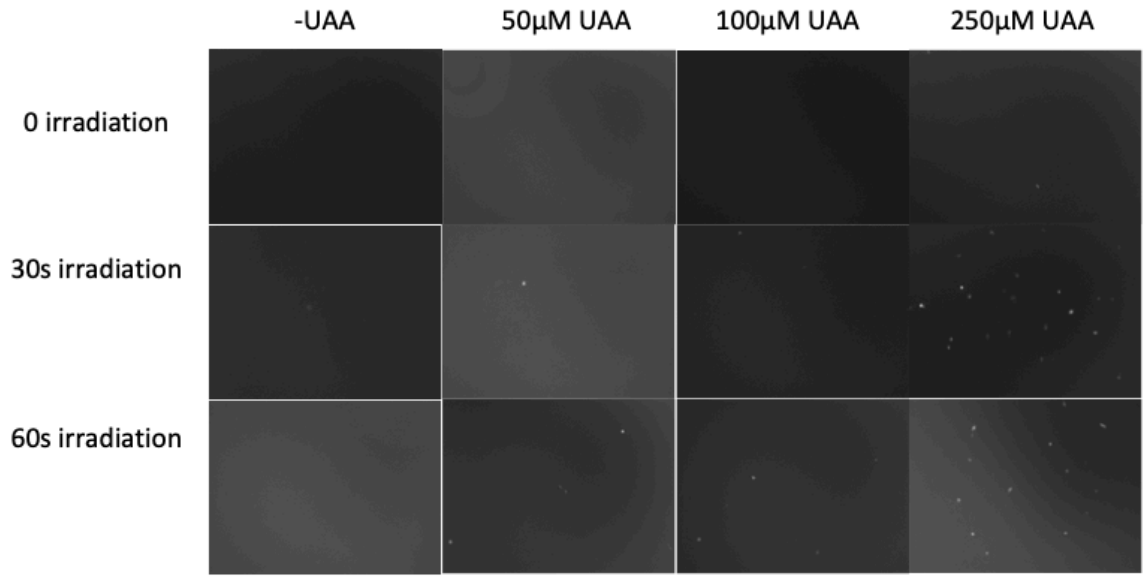


Figure 5.8. AAV-CoumarinK infectivity. Microscopy images of HEK293T cells infected with AAV-R585 made in the presence or absence of coumarinK with or without irradiation with 405nm light.

5.3.2 Conclusions

After the success we had with the incorporation of NBK, we attempted to incorporate a second photocaged lysine, coumarin lysine. Unlike NBK, coumarinK incorporation failed to produce infectious virus, with or without irradiation. Because these UAAs are structurally similar, we expected them to function with similar efficiency. Understanding why NBK is efficiently incorporated while coumarinK is not would advantageous when predicting which UAAs would produce functional AAV.

Photocaged amino acids could be used to control additional interactions in the AAV infection cycle, such as the interaction between AAV and AAVR, AAV's secondary binding receptor, or it could be used to control the nuclear transport of the VP proteins among other uses. These applications have not yet been explored, but incorporation of a photocrosslinking amino acid rather than a photocaged amino acid was explored and is further discussed in chapter 6.

5.4 ACKNOWLEDGEMENTS

Thanks to Rachel Kelemen for the guidance she provided through this project and for teaching me to work with AAV. Thanks to Chester Wrobel and Raja Mukherjee for UAA synthesis, Xiaofu Cao for cloning help, and thanks to Zainab Kiyam and Partha Addy for UAA synthesis and help with coumarinK.

5.5 EXPERIMENTAL PROCEDURES

5.4.1 Cell Culture

Cells were cultured as described in Chapter 1.

5.4.2 Cloning and Plasmids

Plasmids were obtained as described in Chapter 1.

TAG and lysine mutations were introduced into the Cap reading frame using site directed mutagenesis with the primers listed below. After PCR amplification, the vector plasmid and insert were digested with SbfI and HindIII. They were then ligated together and transformed. Primers sequences used for cloning are in Appendix 2. DNA oligomers were purchased form Integrated DNA Technologies (IDT), and PCR reactions were performed on a C1000 Touch thermocycler (Bio-Rad) using Phusion Hot Start II polymerase (Thermo) according to the manufacturer's instructions. Restriction enzymes were from New England Biolabs, and digested DNA was purified using the Spin Smart gel extraction kit (Machery-Nagel). Eton Biosciences was used to sequence the DNA.

5.4.3 Unnatural Amino Acids

NBK was synthesized as previously described by Raja Mukherjee and Chester Wrobel. Coumarin Lysine was synthesized by Zianab Kiyam and Partha Addy.

5.4.4 Production of NBK modified AAV2

AAV was produced as described in Chapter 1 by transfection using 0.48 µg of pIDTSmart-MbPylRS-RC2, 0.53µg of pIDTSmart-8xMmPytR-ITR-GFP, and 0.38µg of pHelper (1.4µg total DNA) in 24.5µL of DMEM, 4.2 µL of 1mg/mL PEI (sigma), and 1mM UAA per well. Cells were harvested after 72 hours using the TAKARA AAVpro kit (ClonTech).

To remove the photocaging group from virus, NBK-virus was irradiated with 360nm light for 60 seconds and coumarinK-virus was irradiated with 405nm light for 30-60 seconds.

5.4.5 qPCR

qPCR was performed using the TAKARA AAVpro kit, performed following the manufacturer's instructions and as described in Chapter 1.

5.4.6 Flow Cytometry

HEK293T cells were infected with 50 GC/cell of virus. 48 hours later, cells were detached from plate using 200 µL of warm 0.25% Trypsin (HyClone) for two minutes at 37 ˚C. Trypsin activity was quenched with 1 mL ice cold DMEM+FBS. Cells were pelleted by centrifugation at 2000xg for five minutes and resuspended in

500 µL ice cold PBS. HEK293T cells were then analyzed for EGFP fluorescence by flow cytometry (Bio-Rad) using the S3e cell sorter.

5.6 REFERENCES

(1) Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. A Genetically Encoded Photocaged Amino Acid. *J. Am. Chem. Soc.* **2004**, *126* (44), 14306–14307. https://doi.org/10.1021/ja040175z.

(2) Kang, J. Y.; Kawaguchi, D.; Coin, I.; Xiang, Z.; O'Leary, D. D. M.; Slesinger, P. A.; Wang, L. In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids. *Neuron* **2013**, *80* (2), 358–370. https://doi.org/10.1016/j.neuron.2013.08.016.

(3) Uprety, R.; Luo, J.; Liu, J.; Naro, Y.; Samanta, S.; Deiters, A. Genetic Encoding of Caged Cysteine and Caged Homocysteine in Bacterial and Mammalian Cells. *ChemBioChem* **2014**, *15* (12), 1793–1799. https://doi.org/10.1002/cbic.201400073.

(4) Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W. Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *J. Am. Chem. Soc.* **2014**, *136* (6), 2240–2243. https://doi.org/10.1021/ja412191m.

(5) Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G. A Genetically Encoded Photocaged Tyrosine. *Angew. Chemie - Int. Ed.* **2006**, *45* (17), 2728–2731. https://doi.org/10.1002/anie.200600264.

(6) Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W. Photocontrol of Tyrosine Phosphorylation in Mammalian Cells via Genetic Encoding of Photocaged Tyrosine. *J. Am. Chem. Soc.* **2012**, *134* (29), 11912–11915. https://doi.org/10.1021/ja3046958.

(7) Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A. Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. *ChemBioChem* **2017**, *18* (14), 1442–1447. https://doi.org/10.1002/cbic.201700147.

(8) Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. Control of Protein Phosphorylation with a Genetically Encoded Photocaged Amino Acid. *Nat. Chem. Biol.* **2007**, *3* (12), 769–772. https://doi.org/10.1038/nchembio.2007.44.

(9) Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W. Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells. *J. Am. Chem. Soc.* **2010**, *132* (12), 4086–4088. https://doi.org/10.1021/ja910688s.

(10) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Bernhard, H.; Schultz, P. G. A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells. **2010**, *48* (22), 4052–4055. https://doi.org/10.1002/anie.200900683.A.

(11) Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. Genetically Encoded Optochemical Probes for Simultaneous Fluorescence Reporting and Light Activation of Protein Function with Two-Photon Excitation. *J. Am. Chem. Soc.* **2014**, *136* (44), 15551–15558. https://doi.org/10.1021/ja5055862.
(12) Hauf, M.; Richter, F.; Schneider, T.; Faidt, T.; Martins, B. M.; Baumann, T.; Durkin, P.; Dobbek, H.; Jacobs, K.; Möglich, A.; et al. Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *ChemBioChem* **2017**, *18* (18), 1819–1823. https://doi.org/10.1002/cbic.201700327.
(13) Peeler, J. C.; Falco, J. A.; Kelemen, R. E.; Abo, M.; Chartier, B. V; Edinger, L. C.; Chen, J.; Chatterjee, A.; Weerapana, E. Generation of Recombinant Mammalian Selenoproteins through Genetic Code Expansion with Photocaged Selenocysteine. *ACS Chem. Biol.* **2020**, 1–6. https://doi.org/10.1021/acschembio.0c00147.
(14) Welegedara, A. P.; Adams, L. A.; Huber, T.; Graham, B.; Otting, G. Site-Specific Incorporation of Selenocysteine by Genetic Encoding as a Photocaged Unnatural Amino Acid. *Bioconjug. Chem.* **2018**, *29* (7), 2257–2264. https://doi.org/10.1021/acs.bioconjchem.8b00254.
(15) Hemphill, J.; Borchardt, E. K.; Brown, K.; Asokan, A.; Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J. Am. Chem. Soc.* **2015**, *137* (17), 5642–5645. https://doi.org/10.1021/ja512664v.
(16) Hemphill, J.; Chou, C.; Chin, J. W.; Deiters, A. Genetically Encoded Light-Activated Transcription for Spatiotemporal Control of Gene Expression and Gene Silencing in Mammalian Cells. *J. Am. Chem. Soc.* **2013**, *135* (36), 13433–13439. https://doi.org/10.1021/ja4051026.
(17) Gautier, A.; Deiters, A.; Chin, J. W. Light-Activated Kinases Enable Temporal Dissection of Signaling Networks in Living Cells. *J. Am. Chem. Soc.* **2011**, *133* (7), 2124–2127. https://doi.org/10.1021/ja1109979.
(18) Liaunardy-Jopeace, A.; Murton, B. L.; Mahesh, M.; Chin, J. W.; James, J. R. Encoding Optical Control in LCK Kinase to Quantitatively Investigate Its Activity in Live Cells. *Nat. Struct. Mol. Biol.* **2017**, *24* (12), 1155–1163. https://doi.org/10.1038/nsmb.3492.
(19) Luo, J.; Arbely, E.; Zhang, J.; Chou, C.; Uprety, R.; Chin, J. W.; Deiters, A. Genetically Encoded Optical Activation of DNA Recombination in Human Cells. *Chem. Commun.* **2016**, *52* (55), 8529–8532. https://doi.org/10.1039/c6cc03934k.
(20) Engelke, H.; Chou, C.; Uprety, R.; Jess, P.; Deiters, A. Control of Protein Function through Optochemical Translocation. *ACS Synth. Biol.* **2014**, *3* (10), 731–736. https://doi.org/10.1021/sb400192a.
(21) Davis, L.; Radman, I.; Goutou, A.; Tynan, A.; Baxter, K.; Xi, Z.; Chin, J. W.; Greiss, S. Optically Splitting Symmetric Neuron Pairs in C . Elegans. **2020**.
(22) Rahman, S. T.; Zhou, W.; Deiters, A.; Haugh, J. M. Optical Control of MAP Kinase Kinase 6 (MKK6) Reveals That It Has Divergent Roles in pro-Apoptotic and Anti-Proliferative Signaling. **2020**, *6*. https://doi.org/10.1074/jbc.RA119.012079.

(23) Wang, J.; Xie, J.; Schultz, P. G. A Genetically Encoded Fluorescent Amino Acid. *J. Am. Chem. Soc.* **2006**, *128* (27), 8738–8739. https://doi.org/10.1021/ja062666k.
(24) Zhou, W.; Hankinson, C. P.; Deiters, A. Optical Control of Cellular ATP Levels with a Photocaged Adenylate Kinase. *ChemBioChem* **2020**, 1–6. https://doi.org/10.1002/cbic.201900757.
(25) Liu, J.; Hemphill, J.; Samanta, S.; Tsang, M.; Deiters, A. Genetic Code Expansion in Zebrafish Embryos and Its Application to Optical Control of Cell Signaling. *J. Am. Chem. Soc.* **2017**, *139* (27), 9100–9103. https://doi.org/10.1021/jacs.7b02145.
(26) Courtney, T. M.; Deiters, A. Optical Control of Protein Phosphatase Function. *Nat. Commun.* **2019**, *10* (1), 1–10. https://doi.org/10.1038/s41467-019-12260-z.
(27) Luo, J.; Kong, M.; Liu, L.; Samanta, S.; Van Houten, B.; Deiters, A. Optical Control of DNA Helicase Function through Genetic Code Expansion. *ChemBioChem* **2017**, *18* (5), 466–469. https://doi.org/10.1002/cbic.201600624.
(28) Opie, S. R.; Warrington, K. H.; Agbandje-McKenna, M.; Zolotukhin, S.; Muzyczka, N. Identification of Amino Acid Residues in the Capsid Proteins of Adeno-Associated Virus Type 2 That Contribute to Heparan Sulfate Proteoglycan Binding. *J. Virol.* **2003**, *77* (12), 6995–7006.
(29) Kern, A.; Schmidt, K.; Leder, C.; Müller, O. J.; Wobus, C. E.; Bettinger, K.; Von der Lieth, C. W.; King, J. A.; Kleinschmidt, J. A. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. *J. Virol.* **2003**, *77* (20), 11072–11081.
(30) Lochrie, M. A.; Tatsuno, G. P.; Christie, B.; McDonnell, J. W.; Zhou, S.; Surosky, R.; Pierce, G. F.; Colosi, P. Mutations on the External Surfaces of Adeno-Associated Virus Type 2 Capsids That Affect Transduction and Neutralization. *J. Virol.* **2006**, *80* (2), 821–834. https://doi.org/10.1128/JVI.80.2.821-834.2006.

6.0 SHORT STORIES

This chapter focuses on the work I did that doesn't yet form a complete story. Most of this work isn't published, but is important for future students to know about. I first discuss the various UAAs that we attempted to incorporate into the AAV capsid, followed by the optimization of a Fab expression system, and finally the work directed at linking proteins to the outside of the AAV capsid.

6.1 INCORPORATION OF OTHER UAA'S INTO THE AAV CAPSID

6.1.1 Introduction

As discussed in Chapter 1, UAAs with a wide variety of functionalities can be incorporated into proteins. Many of these functionalities would be useful if they could be incorporated into the AAV capsid- for example a fluorescent UAA to track the AAV capsid, a photocrosslinker UAA to probe the interactions made by AAV, or additional chemical handles for faster click reactions or the incorporation of multiple handles onto the AAV capsid. We attempted to incorporate several of these UAAs into the capsid with mixed results. In addition to the UAAs discussed below, Coumarin lysine was discussed in chapter 5. The incorporation of these UAAs into

the AAV capsid often produced little to no infectious AAV particles. Understanding why certain UAA's are incorporated into AAV and others are not is an important problem, however, we haven't yet invested the necessary time to understand why it happens. We hypothesize that many of these UAA's get incorporated into the AAV VP proteins, but there are problems with VP protein folding or capsid assembly, as these UAA's have been incorporated into other proteins in mammalian cells. Experiments that could address this include qPCR to test whether or not intact and packaged capsids are made, and western blots, which would show whether or not the capsid proteins are made.

6.1.2 Other UAAs with the pyrrolysyl pair

The wild type pyrrolysyl pair notoriously accepts a wide range of UAA's and has been extensively evolved to accept additional UAA's[1]. We have used it to try to incorporate several UAAs in addition to AzK and NBK.

One disadvantage of the strain-promoted alkyne-azide cycloaddition (SPAAC) reaction, that we have used extensively with AzK, is that it is relatively slow[2]. Strain-promoted inverse-electron-demand Diels-Alder Cycloaddition (SPIEDAC) reactions have a faster reaction rate. SPAAC and SPIEDAC reactions are bioorthogonal, and so could be used to label the AAV capsid with two different functionalities. Two UAAs that can participate in either SPAAC or SPIEDAC are Bicyclo [6.1.0] nonyne-Lysine (BCNK) and strained cyclooctyne-lysine (SCOK) (Figure 1 A, B). Both of these UAAs can be incorporated using evolved versions of the *Methanosarcina mazei* Pyrrolysyl tRNA/synthetase pair[3,4]. BCNK has been

incorporated using two different synthetases, one evolved by Jason Chin and one evolved by Edward Lemke. We initially decided to incorporate BCNK using Jason Chin's synthetase, which contains three amino acid mutations; Y271M, L274G, and C313A. We created a construct bearing the BCNRS as described by Jason Chin, but ultimately decided to move forward with SCOK instead of BCNK and created the SCOKRS, which is the *Mm*PylRS bearing Y306A and Y384F mutations. AAV was produced and harvested by transient transfection, however upon infection, no infectious virus was detected (Figure 1C).

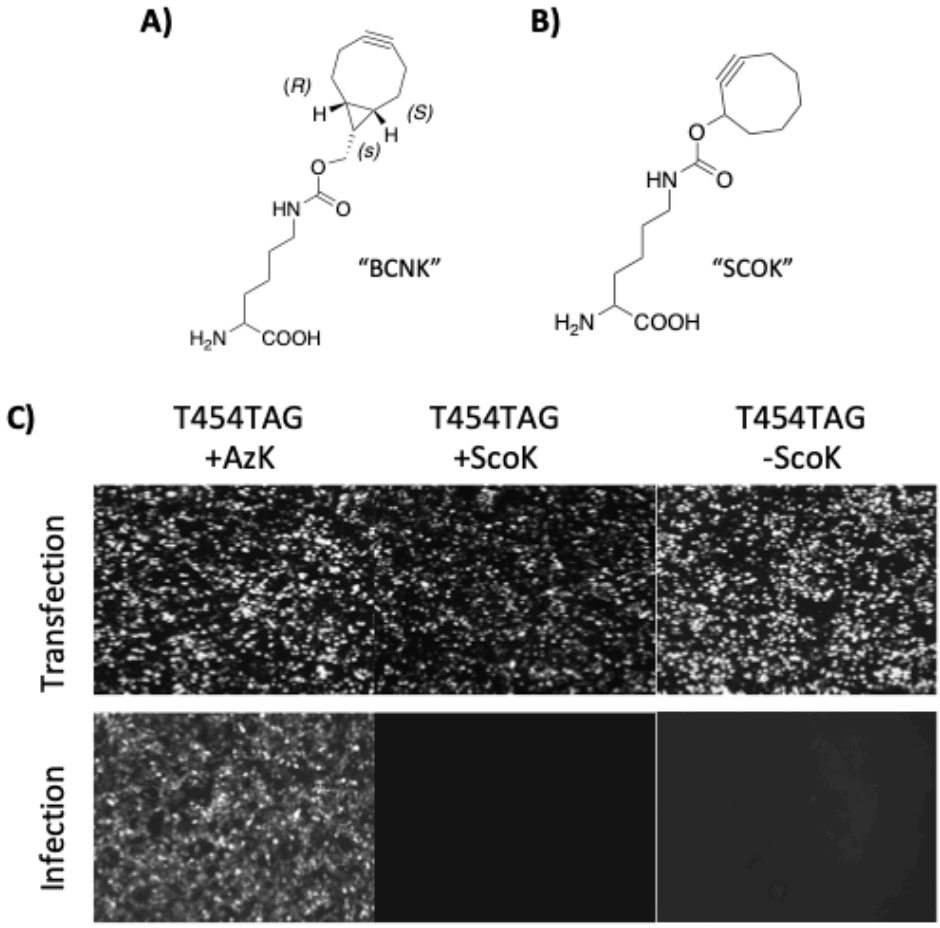


Figure 6.1. ScoK incorporation into the AAV capsid. A) Structure of BCN-Lysine. B) Structure of SCO-Lysine. C) Fluorescence images showing the transfection of AAV with SCOK incorporated. Beneath the transfection pictures are the subsequent infection pictures. The transfection, infection, and imaging of these pictures was performed by Stephanie Smith.

Diazirine lysine (diazK) is a photocrosslinking amino acid that can be incorporated using the wild type *Mb*PylRS/tRNA pair (Figure 2A). The incorporation of a photocrosslinker into the AAV capsid would allow us to probe the poorly understood interactions that AAV makes with the cell. As a proof of concept, we decided to incorporate diazK into residues on the AAV capsid that are known to bind to the AAV A20 antibody, which recognizes formed capsids, and photocrosslink the capsid to the A20 antibody. Seven residues, T251, S468, S662, K665, T671, K706, and K716 were chosen based off of the crystal structure of the AAV2 capsid bound to the A20 antibody (Figure 2B)[5]. T454 was also included in this study as a positive control. These residues were chosen because they appear to be in close proximity to the A20 antibody, but would not directly interfere with binding (Figure 2C).

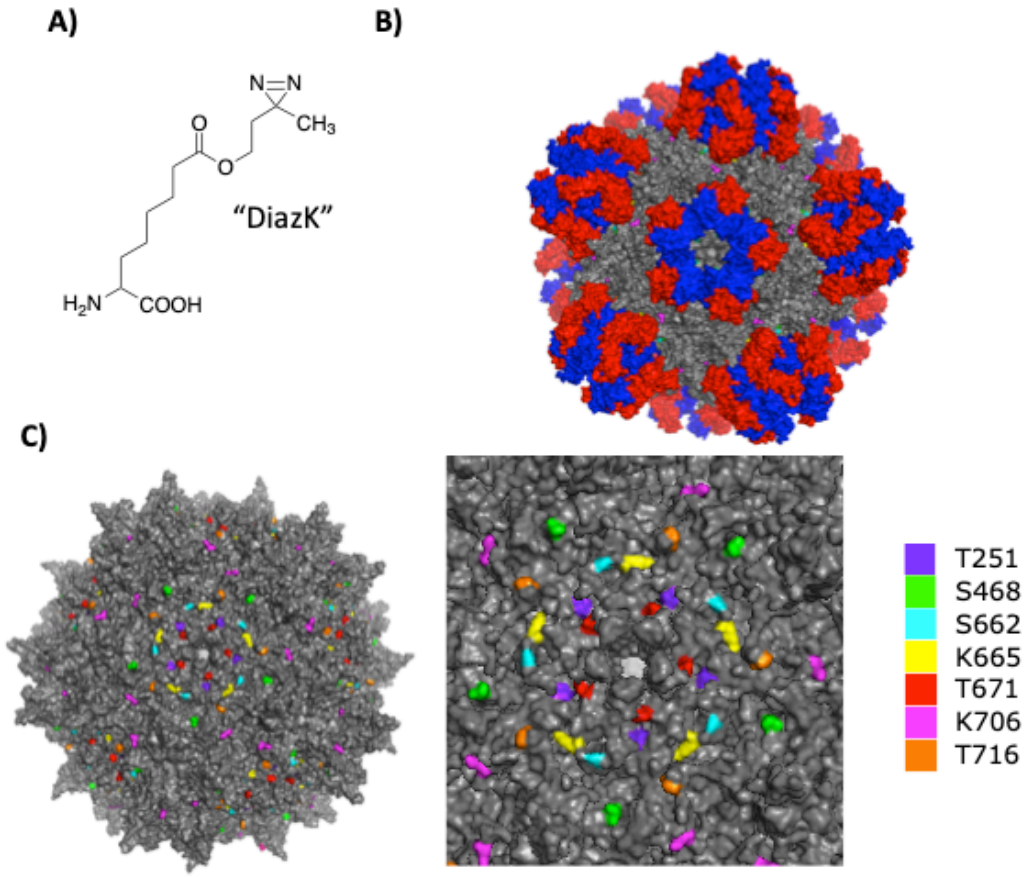


Figure 6.2. Crystal structures highlighting residues on the AAV capsid selected for DiazK incorporation. A) Structure of diazirine lysine. B) crystal structure of the αAAV A20 antibody (red and blue) bound to the AAV2 capsid (grey). C) Residues that sit near the αAAV A20 antibody that were chosen for diazirine lysine incorporation.

AAV was made using either diazK or AzK as a positive control. Virus made with AzK was infectious to varying degrees of success, with some sites being highly infectious and some sites yielding non-infectious virus. AAV made with diazK incorporated into the capsid produced much less infectious virus, with four of the eight sites producing no infectious virus and the remaining four sites producing a small amount of infectious virus (Figure 3A). No infectious virus was detected for any of the sites in the absence of UAA. AzK and diazK were then both incorporated into EGFP to assess incorporation efficiency in a less complicated protein structure (Figure 3B). EGFP was produced to a lower efficiency was lower using diazK than

AzK. Lastly, to further probe why incorporation efficiency of diazK was so poor, a western blot using the AAV B1 antibody was performed to see protein production at site T454. The western blot showed a visible VP3 band when AzK was incorporated, but not when diazK was incorporated, meaning that the protein was likely being made at a much lower level (Figure 3C). At this point we decided to move on from this project because of the low production of AAV. While it was an option to try to incorporate diazK into a select VP protein, we chose not to because we did not think that a limited number of diazirine lysines being displayed would be enough to achieve successful crosslinking.

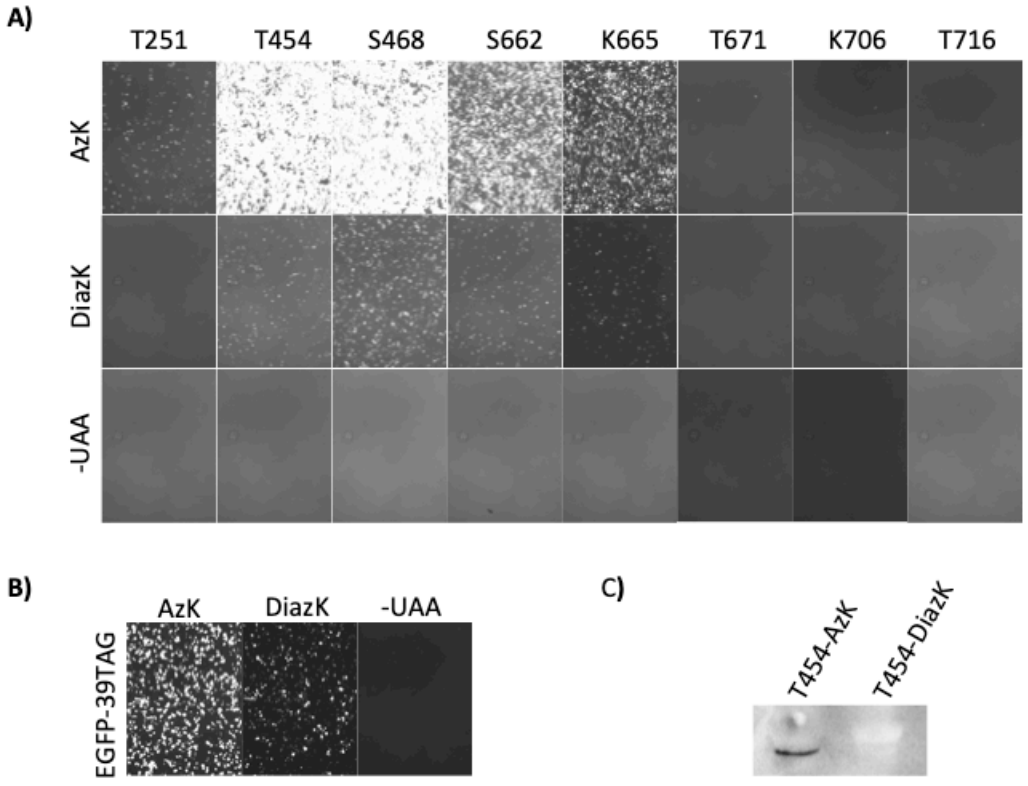


Figure 6.3. Infectivity of AAV-DiazK. A) Images of constant volume infection of AAV produced with AzK, DiazK, or in the absence of UAA. B) EGFP production with the incorporation of either AzK, DiazK, or in the absence of UAA. C) αAAV-B20 western blot showing production of AAV-T454AzK or AAV-T454DiazK.

There are several steps that can be taken to improve diazK incorporation. For example, rather than using the wild type *Mb*PylRS, we could try to use the diazirine lysine-specific AbKRS, which was evolved by the Schultz lab specifically to incorporate diazK into cells[6]. Directed evolution on the wild type MbPylRS was used to make the AbKRS, and it contains three mutations: L274M, C313A, and Y349F. Renpeng Gu, a post-doc in the Chatterjee lab, has had success using this synthetase for the incorporation of diazK into bacterial cells. Additionally, it has been shown that one of the major limiting factors in genetic code expansion is the tRNA, which is why often provide 8 copies of a tRNA. Rachel Kelemen and Delilah Jewel recently evolved an improved Pyrrolysyl tRNA that exhibits 6x better incorporation than wild type *Mm*PyltR. Using the improved synthetase and the improved tRNA could greatly improve the incorporation of diazK into the AAV capsid and is a potential future direction.

6.1.3 The E. coli tyrosyl pair

We attempted to incorporate three UAA's into the AAV capsid using the *E. coli* tyrosyl OMeYRS/EcYtr pair, O-methyl-tyrosine (OMeY), p-Acetyl-phenylalanince (*p*AcF), and , *p*-Azido-phenylalnine (pAzF) (Figure 4A). While OMeY and pAcF showed robust incorporation, pAzF did not produce infectious AAV (Figure 4B). OMeY and pAcF incorporation were also tested at several different concentrations (Figure 4C). 0.5mM, 1mM, and 2.5mM UAA were all found to efficiently produce infectious AAV, but higher concentrations decreased the amount of virus produced.

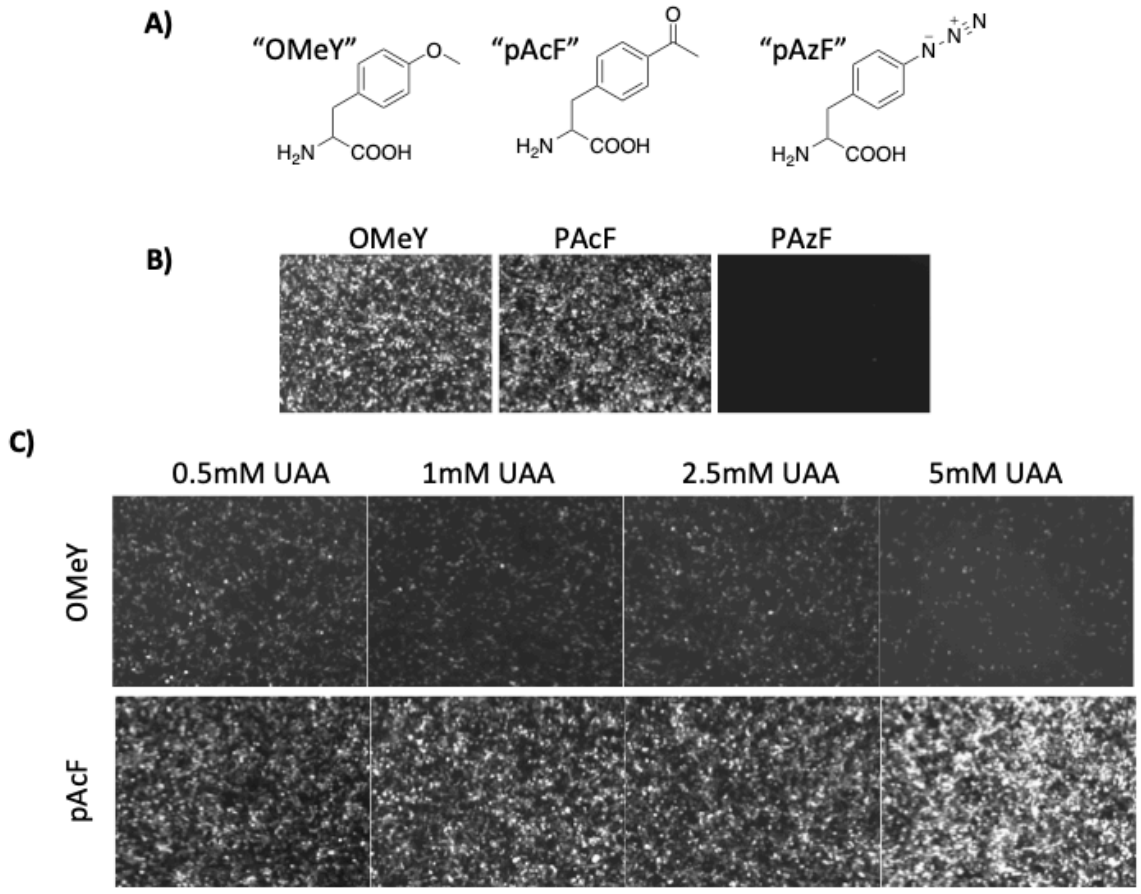


Figure 6.4. Incorporation of OMeY, pAcF, and pAzF into the AAV capsid. A) Structures of O-Methyl-Tyrosine ("OMeY"), p-Acetyl-Phenylalanine ("pAcF"), and p-Azido-Phenylalanine ("pAzF"). B) Images of constant volume infection with AAV produced with either OMeY, pAcF, of pAzF. C) Images of constant volume infection with AAV produced with increasing concentrations of either OMeY of pAcF.

6.1.4 The resurrected E. coli tryptophan pair

As discussed in Chapters 1, the Chatterjee lab recently developed a repurposed *E. coli* tryptophanyl synthetase that is orthogonal in both mammalian cells and certain strains of *E. coli*. This synthetase can incorporate several amino acids, such as 5-propargyl tryptophan (5PrW), 5-Hydroxy tryptophan (5HTP), and 5-Azido-tryptophan (5AzW), which are all bio-orthogonal handles (Figure 5A). 5PrW and 5HTP also undergo click reactions that are orthogonal to the SPAAC reaction that AzK undergoes. A construct was made which would deliver the

*Ec*WRS/tRNA hit 14 pair in addition to the AAV2 machinery to mammalian cells using transient transfection. As these constructs were evolved to suppress a TGA stop codon rather than a TAG stop codon, residue T454 was mutated to TGA. Incorporation into T454TGA in all three capsid proteins resulted in the production of a very small amount of infectious AAV, so we decided to see if incorporation into VP1 or VP2 only would improve efficiency (Figure 5B). Incorporation into VP2 only (2-454TGA) was tested first, however, due to the combination of the leakiness of the synthetase and the fact that infectious virus can be produced in the absence of VP2, no difference was seen in production efficiency in the presence and absence of 5HTP[7]. Next we tried to incorporate it into VP1 only (1-454TGA), which is necessary to create infectious virus. Infectious virus was detected in both the presence and absence of 5HTP, but there was an increase of infectivity with the incorporation of 5HTP (Figure 5C), which means that it was likely incorporated. 5PrW was also tested in VP1 only, and was incorporated to a similar efficiency as 5HTP(Figure 5D). We also tried to incorporate AzW, however the addition of AzW caused the HEK293T cells to die, likely because the stock had degraded.

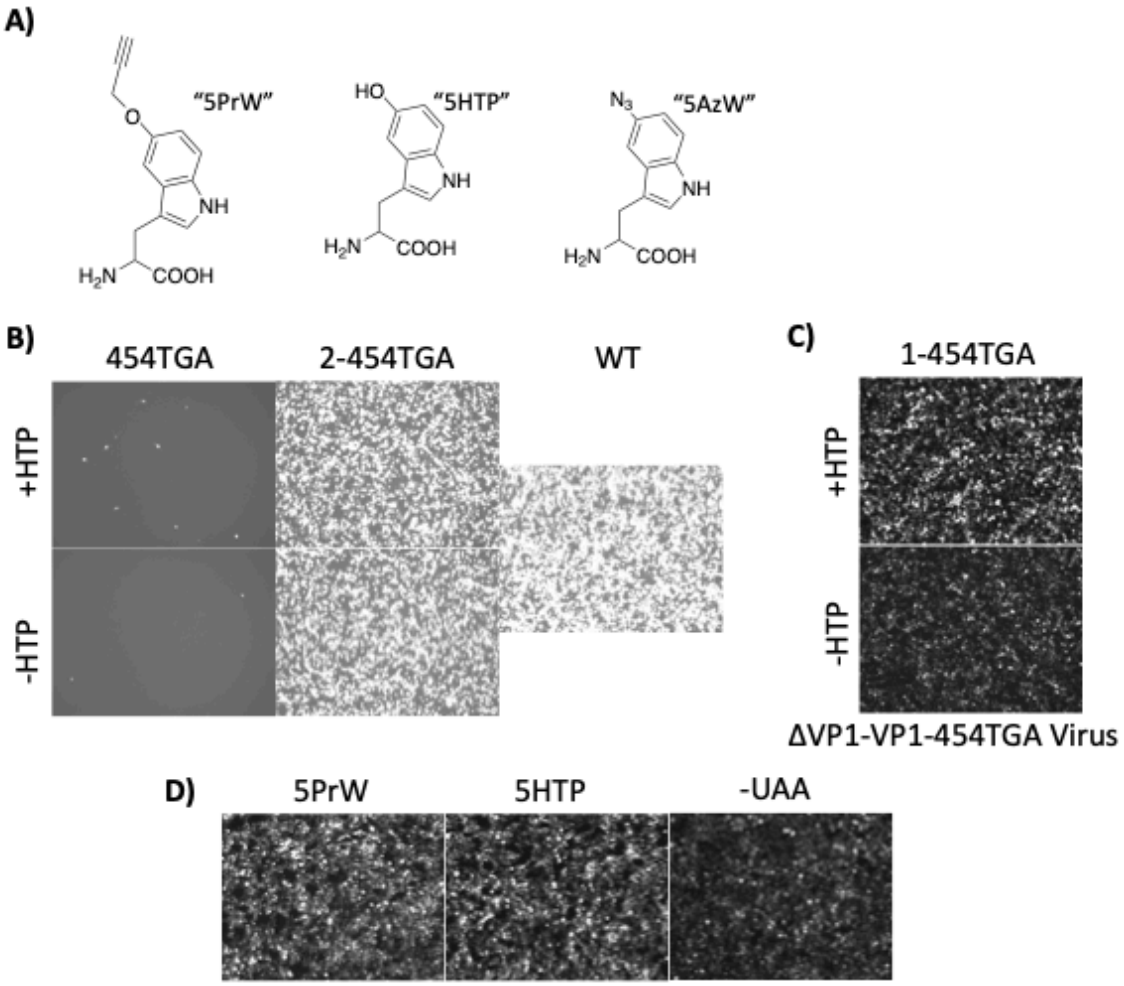


Figure 6.5. Incorporation of 5PrW, 5HTP, or 5AzW. A) Structure of 5-Propargyl-Tryptophan ("5PrW"), 5-Hydroxy-Tryptophan ("5HTP"), or 5-Azido-Tryptophan ("5AzW"). B) Images of constant volume infection of RC2-454TGA, RC2-VP2Δ-CMV-VP2-454TGA, made in the presence or absence of 5HTP. C) Images of constant volume infection of RC2-VP1Δ-CMV-VP1-454TGA, made in the presence or absence of 5HTP. D) Images of constant volume infection of RC2-VP1Δ-CMV-VP1-454TGA, made in the presence of 5PrW, 5HTP, or in the absence of UAA.

6.1.5 Conclusions

While none of the UAAs discussed produced infectious virus to a satisfactory degree, this work opened several avenues for future work. Further experiments, such as qPCR and western blotting, could be performed to better understand why these UAA's are not being incorporated. It is important to know if the capsids are being formed and packaged, which could be assessed by qPCR, and then whether or not capsid proteins are even being produced, which could be assessed by western

blotting using the B1 antibody. Additionally, as discussed, there are several avenues for improving the incorporation of several of these UAAs. The incorporation of diazirine lysine remains highly desirable to probe interactions made between the AAV capsid and the cell, and the ability to incorporate additional photocrosslinkers is always desirable to be able to perform different reactions or multiple reactions on the AAV capsid.

6.2 THE OPTIMIZATION OF A FAB EXPRESSION SYSTEM

6.2.1 Introduction

While most of the work in this thesis has been directed towards creating functional AAV-conjugates using UAA's, this section focuses on the optimization of an antibody Fab expression system. While the incorporation of UAA's into the AAV capsid has proven successful for attaching small molecules to the capsid surface, much of the optimization performed with cRGD would have to be re-optimized for each small molecule attached. Additionally, there are numerous receptors that can't be targeted using small molecules. We thought that one way to create a more universal re-targeting platform would be to use antibody Fabs. The work that has been directed towards achieving this specific goal will be addressed in section (6.3.2), but the Fab that was also created to be used to demonstrate both the use of a new bio-orthogonal reaction as well as our ability to use this new bio-orthogonal reaction for the incorporation of multiple UAAs.

Antibodies are commonly used both as detection tools (ELISA, Western blots, flow cytometry, fluorescence microscopy, etc.) and in biopharmaceutical drugs. In either case, the ability to efficiently label antibodies-with a fluorescent probe or a toxic drug, for example- is a challenge that has been extensively worked on but still needs improvement. Among the most commonly used therapeutics are Antibody Drug Conjugates (ADCs), which are drugs designed to target and kill cancer cells. ADC's are made up of three components; the antibody that targets the cellular receptor, a toxic drug, and the linker that joins them. Four main methods have been used to modify antibodies; modification of the inherent cysteine residues, modification of native glycans, the use of peptide tags, or through the incorporation of UAAs[8,9]. An extensive discussion of cysteine modification strategies can be found in Chapter 4. Both enzymatic and chemical modification strategies exist to modify native glycans on antibodies. Antibodies are glycosylated on the Fc region, far away from the antigen-binding site; however, achieving homogenous preparations of modified antibodies using this method is difficult because glycosylation itself is a heterogeneous post-translational modification. A second downside of modifying glycosylations is the lack of control over the placement of the modification. Several methods have been optimized involving the introduction of a peptide tag followed by enzymatic modification of the tag. Transglutaminase and sortase are two enzymes that can be used to catalyze bond formation between proteins and small molecules, and tags such as the SNAP tag, TMP tag, or Aldehyde tags have commonly been used.

The most flexible technique for conjugating modifications onto antibodies is using the incorporation of a bioorthogonal functional group through genetic code expansion. UAA incorporation offers control over the site of the modification as well as the ability to produce homogenous preparations of modified antibodies. The Schultz lab has primarily driven UAA incorporation into antibodies, focusing on the αHer2 Fab[10–14]. Besides work from the Schultz lab, cell free methods of producing UAA-antibodies have been established, which have the advantage of being much faster, but suffer from low production yields[15–17].

A number of challenges exist in incorporating UAAs into antibodies. One of these challenges is that, in order to create a more potent ADC, the Drug to Antibody Ratio (DAR) must be optimized, which means incorporating more UAAs into the antibody. The Schultz lab has demonstrated the ability to incorporate two different UAA's into a full-length antibody[18]. In order to incorporate multiple distinct UAAs into a protein, multiple distinct synthetase/tRNA pairs that are orthogonal to both the host system and each other must be used, in addition to multiple distinct nonsense codons. The Schultz lab was able to incorporate both pAcF and AzK into a full length antibody. pAcF was incorporated using the OMeYRS and a TAG stop codon in the heavy chain, and AzK was incorporated using the MbPylRS and the TAA stop codon in the light chain. In selecting UAAs to be incorporated, the additional challenge of selecting amino acids that can be used for two distinct bio-orthogonal reactions that are not cross-reactive with each other also exists. There are a variety of chemoselective reactions that exist, which have been discussed in more detail in Chapter 1, however these reactions suffer from slow kinetics, the necessity for toxic

catalysts, or the lack of compatibility with other chemistries to allow for multiple reactions to occur at the same time.

The work done in the Chatterjee lab over the last five years has addressed both the issue of incorporating multiple distinct UAAs as well as the development of new chemoselective reactions to be used in biological systems. First, the Chatterjee lab developed the ATMW1 strain and re-purposed the *Ec*TrpRS. The EcTrpRS/tRNA pair is a new orthogonal pair that can be used in combination with the other pairs that have been developed[19]. Additionally, the Chatterjee lab has optimized a new chemoselective reaction using 5HTP[20], which is efficiently incorporated using the *Ec*TrpRS. The Chemoselective Rapid Azo-Coupling Reaction (CRACR) is an azo-coupling reaction between an aromatic diazonium ion and 5HTP. The azo-coupleing reaction between aromatic diazonioums and aromatic amino acid residues has previously been used, but the CRACR reaction improves upon this by using 5HTP to make it a selective reaction.

6.2.2 FAB production and verification with AzF

The sequence for the α-Her2-Fab used by the Schultz lab was obtained. Two gblocks were purchased from IDT, which were used to build the plasmid encoding the α-Her2-Fab. The gblocks were cloned into a pEvol plasmid under the control of the arabinose operon. As the correct conditions are crucial for disulfide bond formation in antibodies, when Fabs are expressed in *E. coli*, they must be exported into the periplasmic space for proper folding. To achieve this, an STII leader sequence is placed in front of both the light chain and the heavy chain. Wild type α-

Her2-Fab was expressed and purified as previously described[10,11]. The *M. jannashii* pyrrolysyl pair that was evolved to incorporate Cyano-phenylalanine was used to incorporate AzF[21]. Two sites were chosen for AzF incorporation based on the optimization done by the Schultz lab, K169 and S202 (Figure 6A). K169 and S202 were mutated to TAG stop codons using overlap PCR. Plasmids bearing the α-Her2-Fab and the *Mj*PylRS-CNF/tRNA pair were co-transformed into *E.coli* Top10 cells and used to express AzF in the α-Her2-Fab. Wild type and AzF Fab was purified and validated using both SDS-PAGE and LCMS (Figure 6B, C), yields can be found in Table 1. In order to ensure that the Fab binding to Her2 was not diminished with the incorporation of AzF, DBCO-Carboxyrhodamine 110 was conjugated to the AzF and the resulting Fabs were used to label SK-BR-3 cells, which overexpress the Her2 receptor. The conjugation of the DBCO-carboxyrhodamine 110 was confirmed by LCMS and the Fab binding to cells was confirmed with FACS analysis (Figure 7).

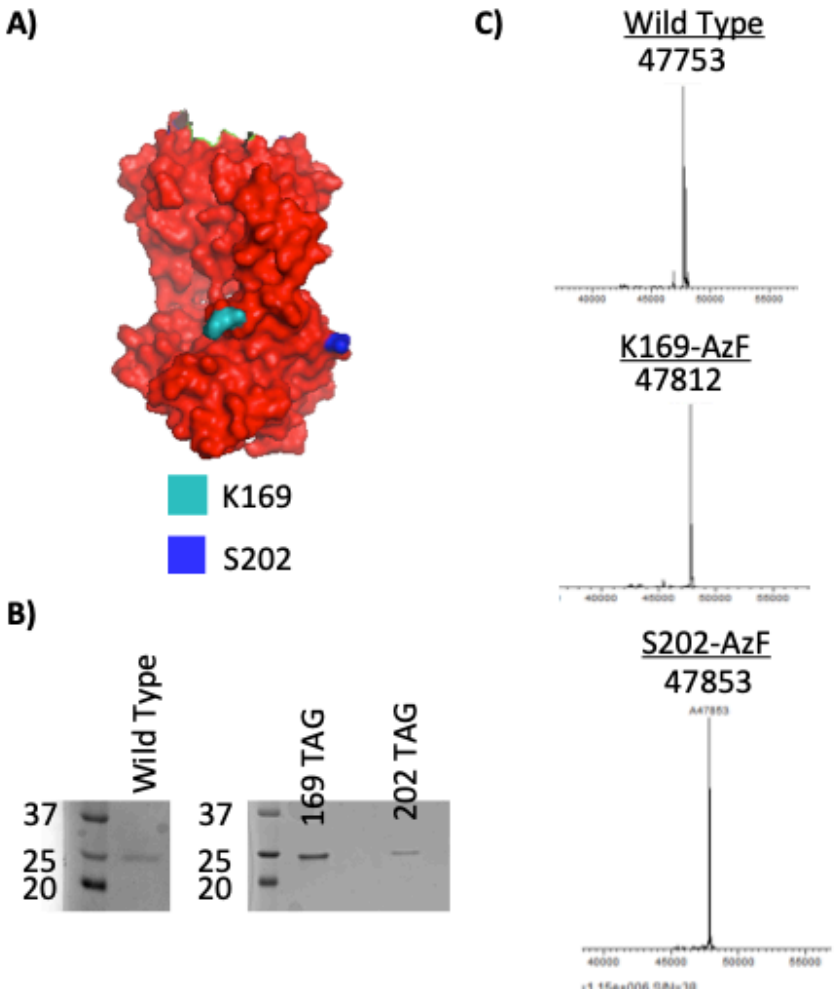


Figure 6.6. UAA incorporation into the αHer2 Fab. A) Crystal structure of the αHer2-Fab with residues K169 and S202 highlighted, where AzF was incorporated. B) SDS Page analysis of purified Fab protein. C) LCMS analysis of purified Fab protein.

Table 6.1. Different variations of the α-Her2-Fab that were made, which UAA was incorporated, the Strain they were produced in, purification yield, and the expected mass in kDa.

Mutant	UAA incorporated	Strain	Yield	Expected mass (kDa)
Wild type	None	Top10	1.5mg/L	47753
K169TAG	AzF	Top10	0.8mg/L	47813
S202TAG	AzF	Top10		47854
Wild type	None	ATMW1	0.35mg/L	47753
K169TGA	5HTP	ATMW1	0.17mg/L	47826
K169TGA-S202TAG	5HTP, AzF	ATMW1	0.025mg/L	47927
K169TGA-S163TAG	5HTP, AzF	ATMW1	-	47927
K169TGA-T198TAG	5HTP, AzF	ATMW1	-	47913
K169TGA-K213TAG	5HTP, AzF	ATMW1	-	47886

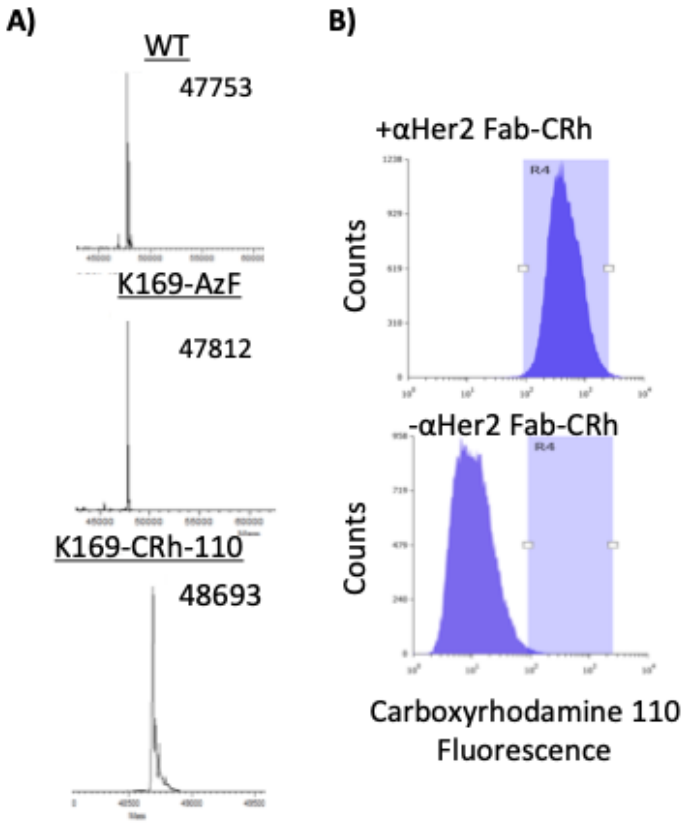


Figure 6.7. αHer2-Fab-AzK efficiently binds to the Her2 Receptor. A) LCMS analysis showing wild type, K169-AzF, and K169-AzF conjugated to DBCO-Carboxyrhodamine 110. B) FACS analysis of SK-BR-3 cells in the presence or absence of the αHer2-Fab-Carboxyrhodamine-110

6.2.3 5HTP incorporation

In order to incorporate 5HTP, several changes to the previously described plasmid system had to be made. The α-Her2-Fab was moved to a pBK plasmid in order to make it compatible with the previously made pEvoltac plasmid harboring the *Ec*TrpRS/tRNA pair and, because incorporation with the TGA stop codon rather than the TAG stop codon displayed a lower level of cross reactivity, the K169TAG was mutated to K169TGA using overlap PCR[19]. The plasmids, pBK-aHer2-Fab-169TGA and pEvoltac-EcW-TGA-h14, were co-transformed into the *E. coli* ATMW1 strain, which was used to produce α-Her2-Fab-K169-5HTP. Unfortunately, expression in the ATMW1 strain greatly diminished yields (Chart 1), but production

of both wild type α-Her2-Fab and 5HTP containing Fab were successfully expressed and purified, as validated by SDS-PAGE and LCMS (Figure 8A, B). Fluorescein-diazo was then conjugated to α-Her2-Fab-169TGA, which was also validated by SDS-PAGE and LCMS. The ability of α-Her2-Fab-169-fluorescein to bind to the Her2 receptor was then validated by incubation with SK-BR-3 cells followed by FACS analysis (Figure 8B, C).

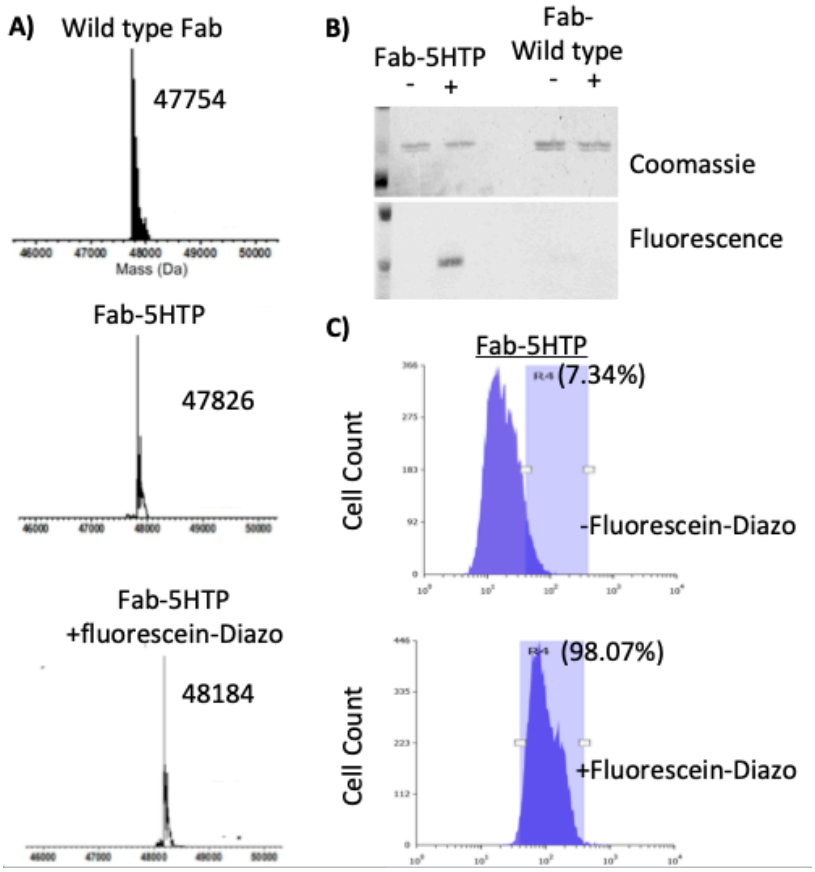


Figure 6.8. Incorporation of 5HTP into the αHer2-Fab. A) LCMS analysis of wild type, K169-5HTP, and K169-5HTP+fluorescein-diazo αHer2 Fab produced in the ATMW1 strain. B) SDS-Page analysis of wild type or K169-5HTP-αHer2 Fab in the presence or absence of the fluorescein-diazo probe. Both fluorescence and coomassie imaging are shown. C) FACS analysis demonstrating that K169-5HTP-αHer2 Fab successfully conjugates to the fluorescein-diazo probe and will then bind SK-BR-3 cells. These experiments were performed with Partha Addy.

6.2.4 Double suppression using 5HTP and AzF

The ability to incorporate multiple distinct UAAs into proteins would be highly advantageous, as it would enable for multiple different functional groups to be incorporated into the protein of interest. Two of the limiting factors for incorporating multiple bioorthogonal handles are the number of orthogonal synthetase/tRNA pairs and the number of bioorthogonal reactions that are not cross reactive. The development of the repurposed *Ec*TrpRS/tRNA pair and the CRACR reaction addresses both of these limitations, as they offer both a new orthogonal pair and a new, highly selective bioorthogonal reaction. We wanted to demonstrate this using the α-Her2-Fab. We decided to keep the incorporation of 5HTP at K169 and chose four new sites on the α-Her2-Fab for the incorporation of AzF. Sites S202 in the light chain and S163, T198, and K213 in the heavy chain were mutated to TAG using overlap PCR. pBK-α-Her2-Fab with both the K169-TGA mutation and the one of the four TAG mutations was co-transformed into the ATMW1 strain with pEvoltac-EcW-H14-*Mj*Y-CNF.TAG (Figure 9).

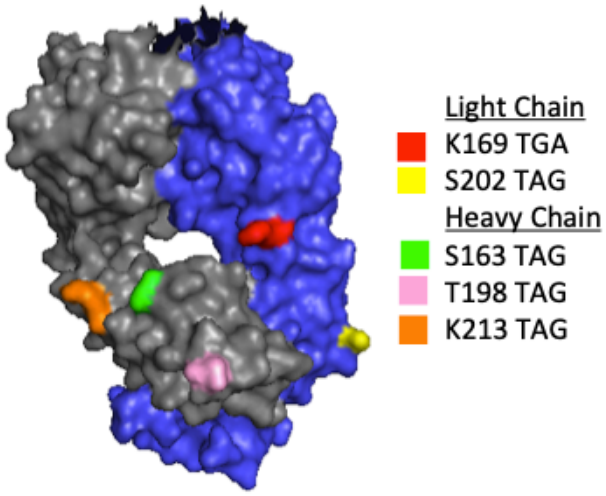


Figure 6.9. Crystal structure highlighting sites chosen for double suppression in the αHer-Fab with mutations in the light chain (blue) and heavy chain (grey) highlighted. Highlighted sites were were tested in double suppression using 5HTP and AzF.

While all four of these constructs were expressed, purified, and validated by SDS-PAGE, we were only able to achieve validation by LCMS of the α-Her2-Fab-K169TGA-S202TAG construct (Figure 10A). Because of this, we chose to move forward with the α-Her2-Fab-K169-TGA-S202TAG mutant. We next showed that it was possible to label the α-Her2-Fab-K169-5HTP-S202-AzF with DBCO-TAMRA or Fluorescein-Diazo (Cdz) both individually as well as using a "one pot" method. These reactions were validated by both SDS Page and LCMS, while the wild type Fab remained unchanged (Figure 10A, B).

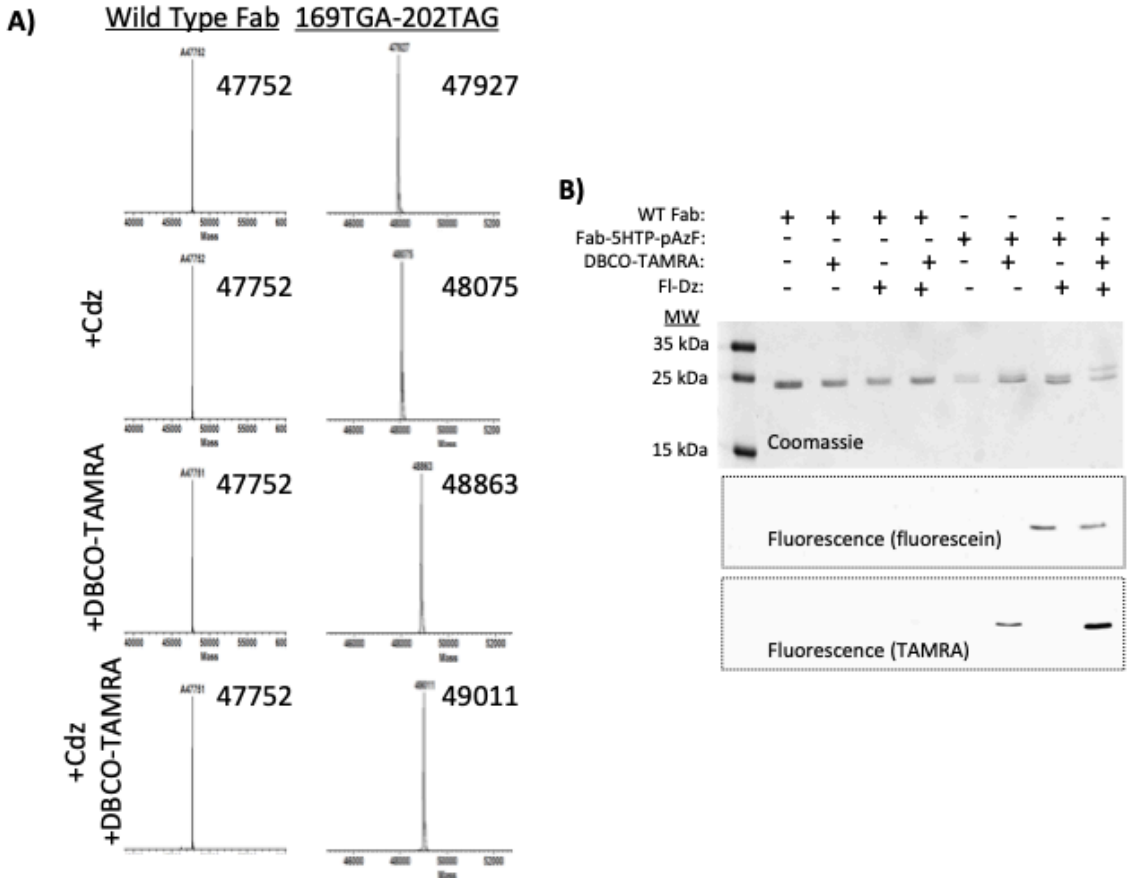


Figure 6.10. Double suppression in the α-Her2 Fab. A) LCMS analysis of wild type or αHer2-Fab-169-5HTP-202-AzF in the presence or absence of either DBCO-Tamra, Cdz, or both. B) SDS-page analysis showing selective reactions between or αHer2-Fab-169-5HTP-202-AzF with either DBCO-TAMRA, Cdz, or both. These experiments were performed with Partha Addy.

6.2.5 Conclusions

We have shown the successful optimization of the αHer2-Fab, with AzF incorporation at both residues K169 and S202. We then conjugated a Carboxyrhodamine 110-DBCO to the AzF and showed that the Fab binds specifically to cells that overexpress the Her2 receptor, namely SK-BR-3 cells. Additionally, we transitioned this machinery to work in the ATM1 strain to incorporate 5HTP and successfully labeled the 5HTP using our CRACR reaction. Lastly, we used the

ATMW1 system to incorporate both 5HTP in combination with AzF and accomplished labeling using two bio-orthogonal probes.

The ability to incorporate two different and orthogonal UAA's into an antibody Fab enable for the production of uniform preparations of modified antibodies that could be used to prepare antibody drug conjugates. The next step would be to move this to full-length antibodies, which Conor Lyond is currently working on.

6.3 LINKING PROTEINS TO THE CAPSID SURFACE

6.3.1 Introduction

One of the downsides that are discussed in great lengths in this thesis of using AAV as a gene therapy vector is its broad tropism and our inability to direct AAV to desired targets. However, there are several other drawbacks, including AAV's limited cargo capacity. Popular systems such as the CRISPR/Cas9 system are too large to be packaged into a single AAV. Dual AAV systems have been created, however they are dependent on a cell receiving both viruses. Expanding the AAV cargo capacity is a challenging problem that, if solved, would enable the delivery of systems such as CRISPR/Cas9, Zinc-finger nucleases, transcription activator-like effector nucleases, or even just the delivery of genes that exceed the cargo capacity[22].

One strategy that we envisioned is to attach a desired protein to the outside of the AAV capsid and have the capsid deliver the protein into the cell while maintaining the ability to also deliver a gene packaged in the AAV genome. Several

studies have been performed that have attached proteins such as darpins, antibodies, affibodies, GFP, or luciferase among others to the capsid and successfully been incorporated into the cell for retargeting or tracking purposes, however to my knowledge, this strategy has not been developed to deliver machinery into the cell[23–27]. When working towards this goal, we envisioned three key pieces that needed to be optimized: the linker between the AAV capsid, the strategy used to attach the linker to the AAV capsid, and the strategy used to attach the linker to the protein of interest (Figure 11A). While there are many additional strategies that could be used to achieve linking proteins to the AAV capsid, I focused on three different strategies; the incorporation of UAA's into both the AAV capsid and the protein of interest with a chemical linker using click chemistry to link the protein to the AAV capsid, the use of sortase to link a UAA containing AAV capsid to a protein that is genetically fused to the LPETG motif that is recognized by sortase, and lastly the use of Halotag to link the AAV capsid to a protein of interest (Figure 11B-D).

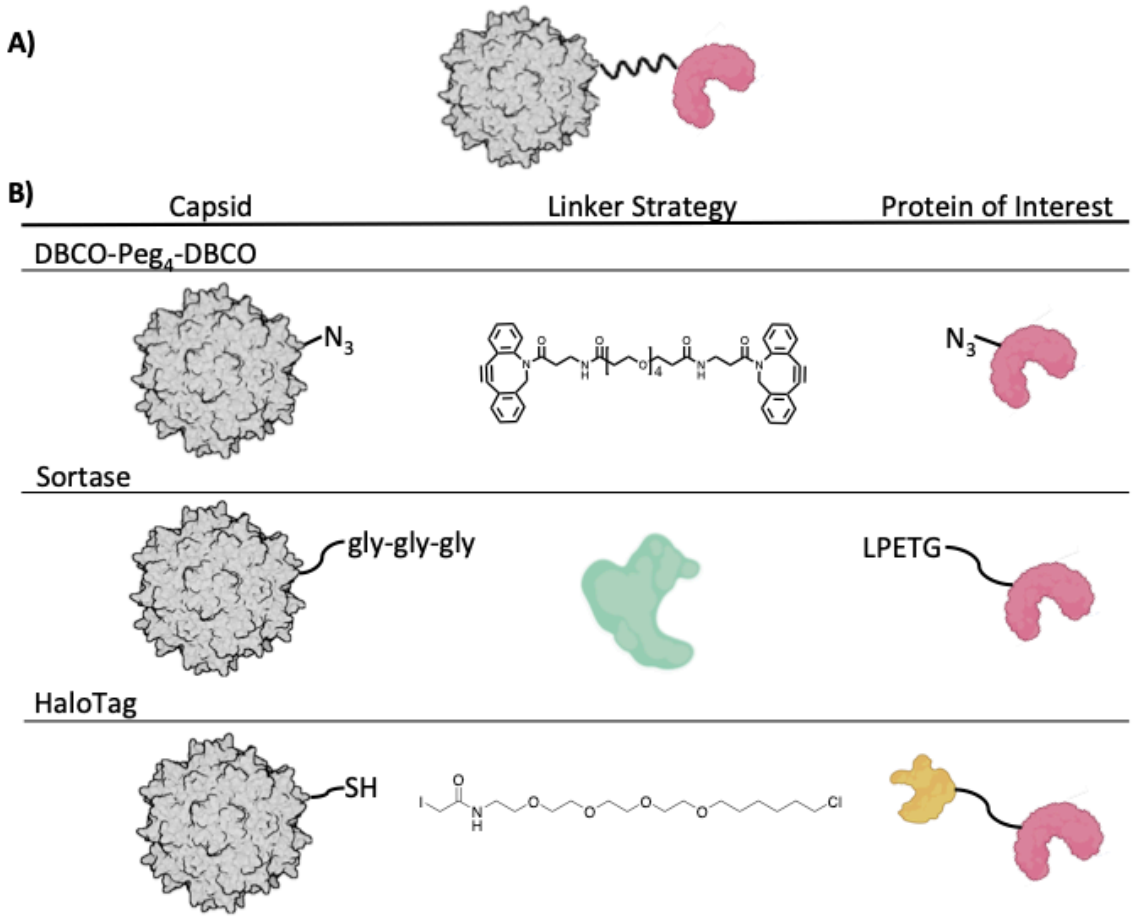


Figure 6.11. Linking a protein to the AAV capsid. A) Linking the AAV capsid to a protein of interest. B) The three strategies that we attempted were to attach the AAV capsid to the protein of interest by incorporating azide-bearing UAA's into both the capsid and the protein of interest, and linking them using a DBCO-Peg$_4$-DBCO linker, through the use of sortase chemistry, and through the use of the HaloTag technology.

6.3.2 Using a DBCO-Peg$_4$-DBCO Peg Linker to connect the AAV capsid to a protein

The first strategy that we used to attempt to link a protein to the AAV capsid revolved around the incorporation of an azide containing UAA into both the AAV capsid and the αHer2 antibody Fab, and linking them together with a DBCO-Peg$_4$-DBCO molecule. AzK was incorporated into the AAV2 capsid and AzF into the αHer2 Fab. This would be achieved by first incubating the αHer2 Fab with the DBCO-Peg$_4$-DBCO linker until the αHer2 Fab was completely labeled, and excess linker would be

removed by dialysis against PBS. Next, the αHer2 Fab-Peg_4-DBCO would be conjugated to AAV-AzK. Excess αHer2 Fab-Peg_4-DBCO would be removed by quenching the reaction with biotin-Peg_4-azide followed by removal with streptavidin resin (Figure 12). Prior to performing the completed experiment, there was a good amount of optimization that was performed, which all used sfGFP-AzF rather than the αHer2-Fab due to ease of production. The linker chemistry between sfGFP and the DBCO-Peg_4-DBCO linker had to be optimized and validated by LCMS and extensive efforts were put into showing that the sfGFP-Peg_4-DBCO was being successfully linked to the AAV capsid. Unfortunately, I was never able to definitively prove that the linker chemistry was successful before we decided to move on to a different direction. Here I discuss the efforts that I put into making this project work.

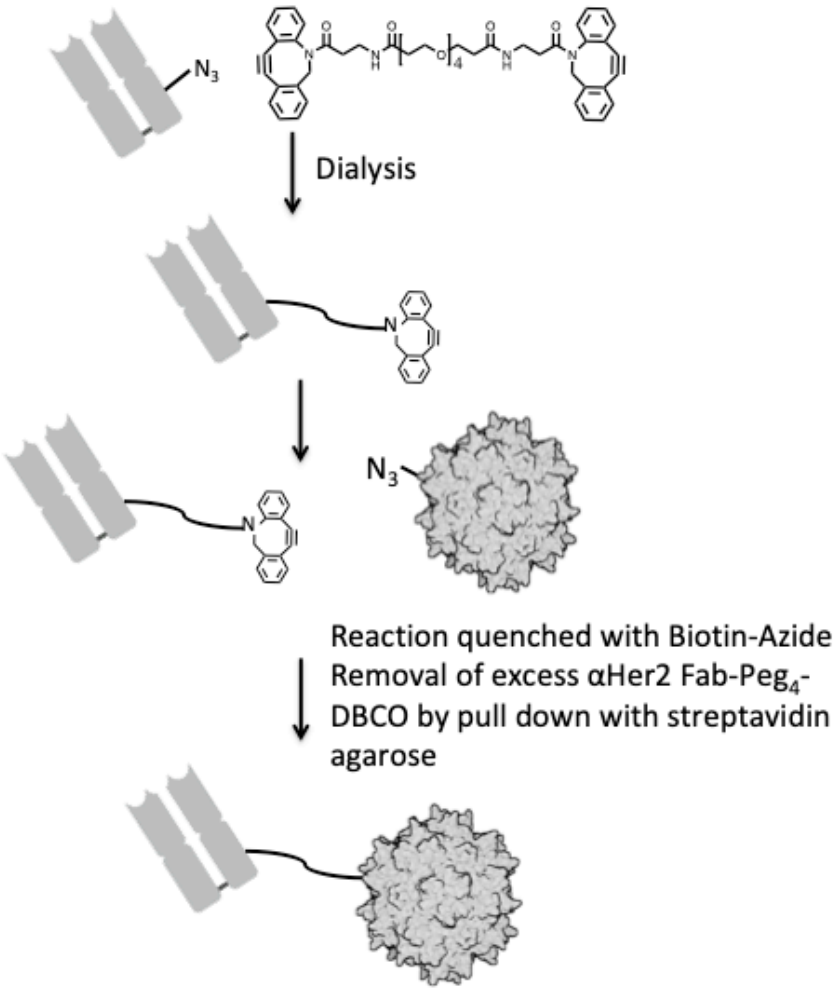


Figure 6.12. An overview of linking the αHer2-Fab to the AAV capsid. The DBCO-Peg$_4$-DBCO linker would first be attached to the αHer2-Fab using the DBCO-azide reaction. Excess DBCO-Peg$_4$-DBCO would be removed by dialysis and the αHer2-Fab-DBCO would be mixed with azide-bearing AAV capsid. This reaction would be quenched using biotin-azide and excess αHer2-Fab-DBCO would be removed using streptavidin agarose.

When linking DBCO-Peg$_4$-DBCO to an azide containing protein, there are three possible outcomes. The first is that, by adding an excess of the azide containing protein, not all of the protein will be labeled with DBCO-Peg$_4$-DBCO. The second outcome is if there is not enough of an excess of DBCO-Peg$_4$-DBCO, then some of the Protein-Peg$_4$-DBCO will react with azide containing protein linking two proteins together. The final and desired outcome is to use a large enough excess of DBCO-Peg$_4$-DBCO that the azide containing protein would be uniformly labeled and

none of the protein would be linked together. It was possible to demonstrate all three of these scenarios by varying the ratios of the sfGFP-151AzF to DBCO-Peg_4-DBCO and analyzing the final mixture by LCMS (Figure 13). A 1:5 ratio of the DBCO-Peg_4-DBCO linker: sfGFP-AzF resulted in a majority of unmodified sfGFP-151AzF, with a small amount of sfGFP-Peg_4-DBCO, a 1:1 ratio resulted in a mixture of sfGFP-Peg_4-DBCO and sfGFP-Peg_4-sfGFP, and a ratio 2:1 showed full conversion to sfGFP-Peg_4-DBCO with no linked sfGFP. Moving forward, I used a 5:1 ratio of linker to sfGFP to decrease the amount of time the reaction needed to reach completion.

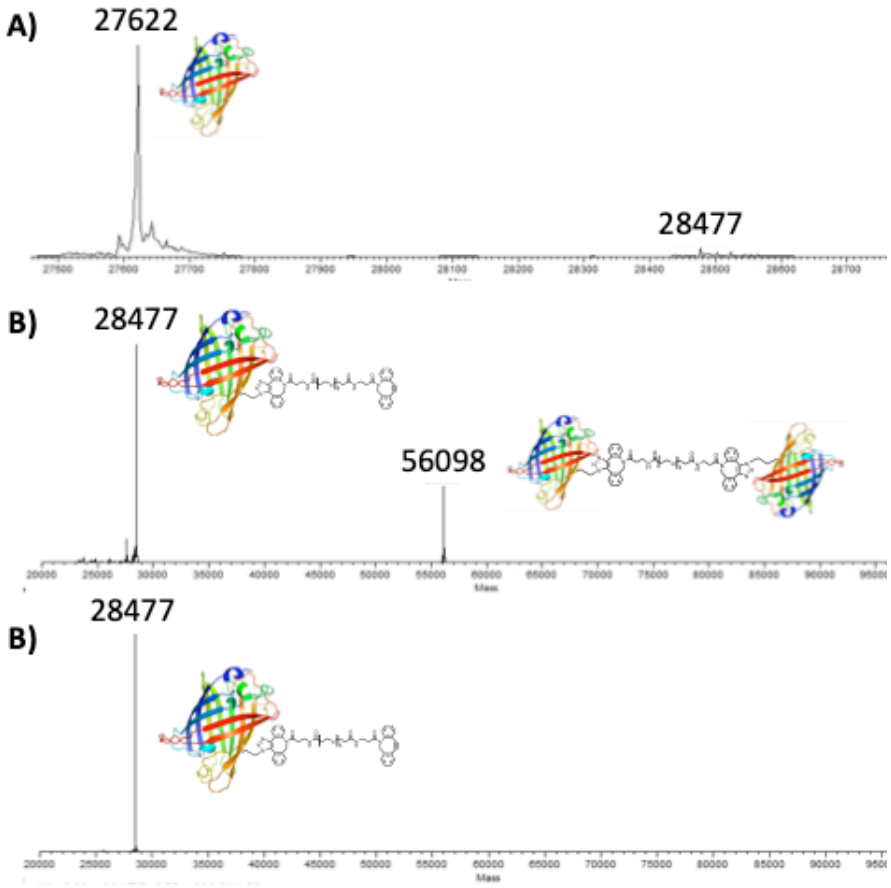


Figure 6.13. The optimization of the reaction between sfGFP-151-AzF and the DBCO-Peg_4-DBCO linker as verified by LCMS. LCMS showing the extent of conversion when mixed at a ratio of A) 1μM diDBCO:5μM sfGFP, B) 5μM diDBCO:5μM sfGFP, or C) 10μM diDBCO:5μM sfGFP

Because we were using semi-pure AAV, we decided to monitor the reaction between the sfGPF-Peg_4-DBCO and the AAV-454AzK using western blotting. We decided to use the AAV B1 antibody, which recognizes unfolded VP1, VP2, and VP3, followed by stripping the western blot membrane and the secondary use of the αGFP antibody. The most successful western blot that I achieved shows the disappearance of the AAV VP1, VP2, and VP3 bands at 90, 70, and 60 kDa, and the appearance of much higher bands (Figure 14). Additionally, the αGFP antibody shows the disappearance of the GFP band with increasing concentrations of GFP-Peg_4-DBCO. However, there are several concerns about this western blot. The first is that the αGFP western blot is messy, showing a lot of bands that do not correlate with purified GFP. The second is that the appearance of the new band on the AAV-B1 blot occurs much higher than expected. VP3 is 60 kDa, and sfGFP is 28, meaning that the expected AAV-sfGFP conjugate would appear at around 90kDa, right around VP1.

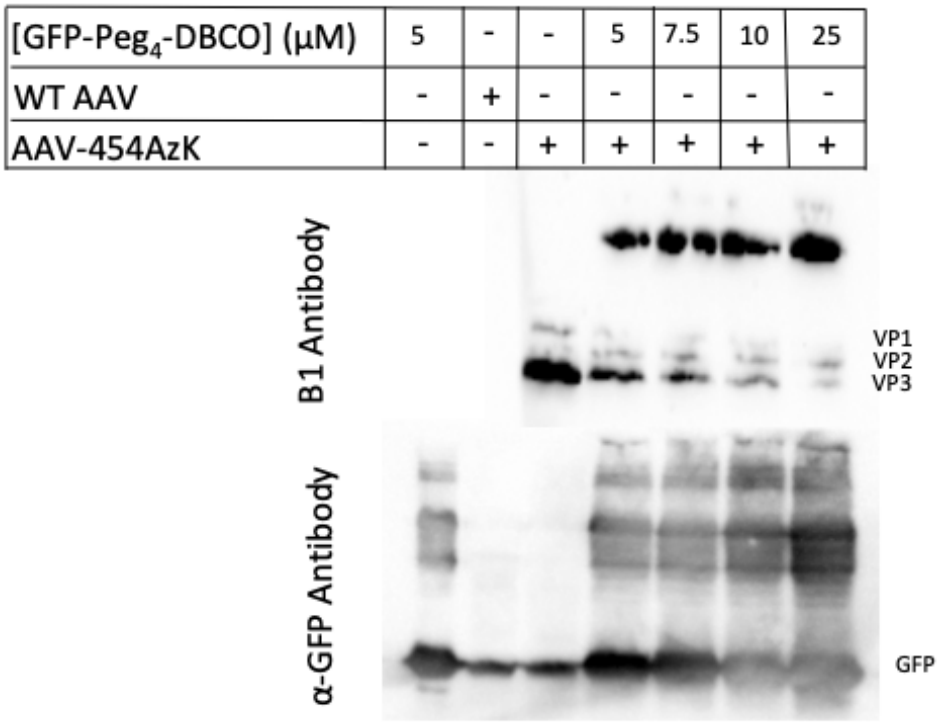


Figure 6.14. Western blot showing AAV linking to DBCO-Peg$_4$-sfGFP. A) α-AAV B1 antibody and B) α-GFP western blot showing the reaction of increasing concentrations of sfGFP-Peg$_4$-DBCO with a fixed amount of AAV-454AzK.

Lastly, despite the absence of verification that the conjugation was successfully occurring, I attempted to go through the complete protocol as outlined in Figure 12 and the experimental procedures multiple times, infecting SK-BR-3 cells with AAV-454AzK that should have been linked to sfGPF-Peg$_4$-DBCO. Unfortunately, no infectivity was detected.

There are several ways that this could now be done better. At the time that most of this optimization was being performed, we did not have the means in place to completely purify the AAV-454AzK. The ability to purify the AAV would enable us to monitor the reaction by Coomassie or silver stain rather than western blot, and would remove the lysate that could be interfering with the reaction. Additionally,

the ability to produce large amounts of AAV would be highly advantageous. The ability to increase production would be necessary to the success of this project.

6.3.3 The optimization of the sortase reaction

Sortase A is a bacterial transpeptidase that recognizes the conserved LPXTG sequence. In bacteria, it recognizes the LPXTG motif on proteins and uses it to anchor the protein to a peptidoglycan cell wall. Structural biologists have taken advantage of the sortase reaction to fuse proteins together. It is possible to link two proteins together by adding the LPXTG motif to one protein and a string of at least three glycines to another[28,29]. We decided to take advantage of the sortase reaction in our attempts to link the αHer2-Fab to the AAV capsid.

A pet29 plasmid bearing the Sortase A pentamutant developed by David Liu was purchased and Sortase A was expressed, purified, and confirmed to have the correct mass using LCMS (Figure 15A). The Liu lab evolved the Sortase A enzyme and removed the transmembrane domain, residues 1-59. It has an additional 5 mutations that improve kinetics and activity[30]. An AALPETG-6xHis tag motif was cloned onto the C-terminus of both the αHer2-Fab and sfGFP, for optimization of the sortase reaction. Both of these proteins were successfully expressed and purified, and their masses were also confirmed using LCMS (Figure 15B,C).

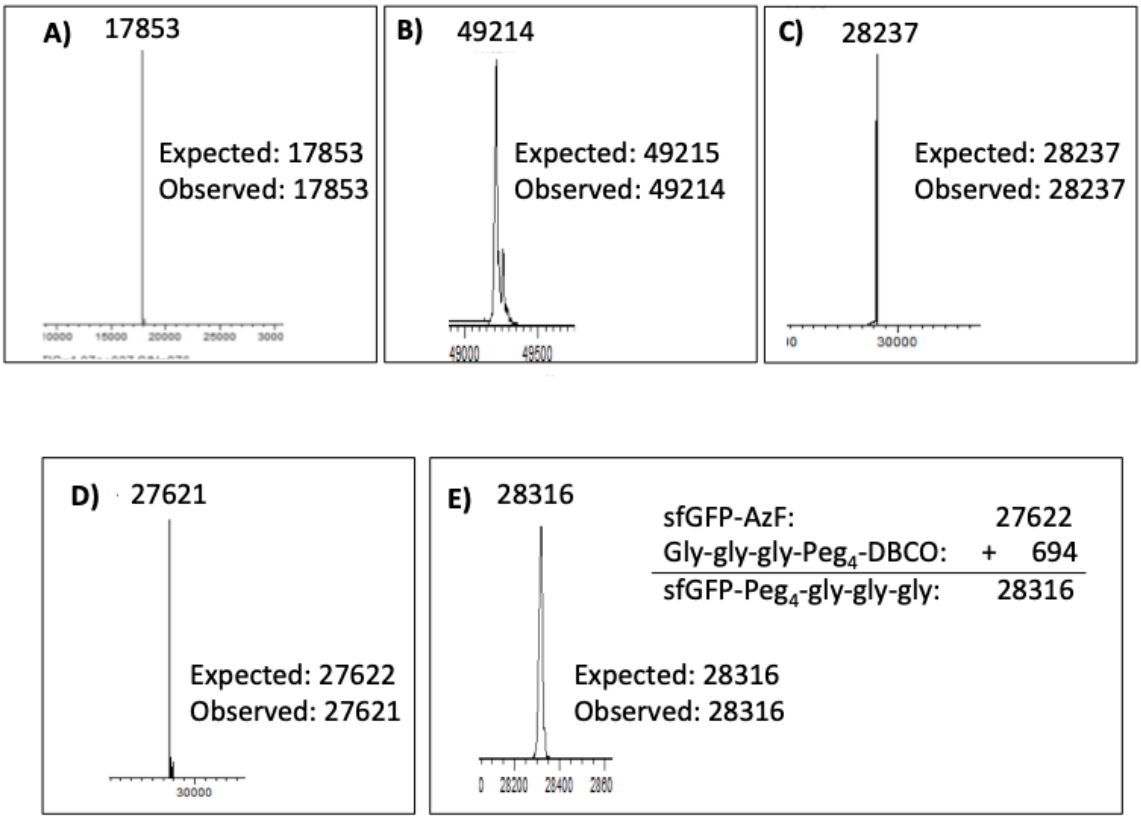
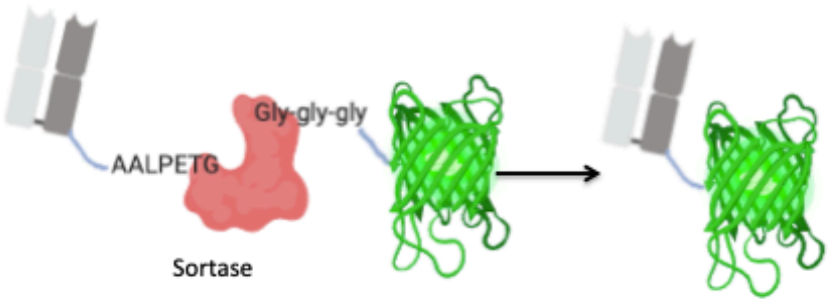


Figure 6.15. LCMS analysis of purified proteins used in the sortase reaction. A) LCMS analysis of purified Sortase protein. B) LCMS analysis of purified αHer2-Fab-AALPETG-6xHis tag C) LCMS of purified sfGFP-AALPETG-6xHis tag D) LCMS of purified AzF-sfGFP. E) LCMS analysis showing complete conversion of AzF-sfGFP to sfGFP-gly-gly-gly

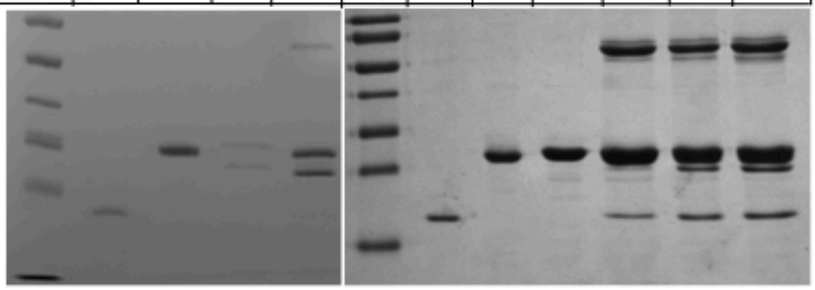
The gly-gly-gly motif was chemically fused to sfGFP-AzF using the commercially available gly-gly-gly-Peg$_4$-DBCO. sfGFP-AzF was fused to gly-gly-gly-Peg$_4$-DBCO, confirmed by LCMS (Figure 15D,E), and excess gly-gly-gly-Peg$_4$-DBCO was removed by dialysis in PBS. The sortase reaction was the optimized using either αHer2-Fab-LPETG or sfGFP-LPETG and gly-gly-gly-Peg$_4$-sfGFP, and monitored by SDS-Page. Conditions, including the reaction time and the optimal buffer (previous protocols reported the reaction in both TRIS and PBS), and whether or not spiking in excess sortase part way through the reaction were optimized (Figure 16). Unfortunately, we found that we could not achieve a satisfactory level of conversion to linked protein to move forward with this strategy.

A)



B)

Sortase (1μM)	+	-	-	+		+	-	-	+	++	+++
[sfGFP-Gly_3] (μM)	-	50	-	50		-	-	12.5	25	25	25
[αHer2-LPETG] (μM)	-	-	25	25		-	25	-	12.5	12.5	12.5
Time (hr)	-	-	-	1		-	-	-	1	2	3
% conversion	-	-	-	30							



C)

Sortase (1μM)	-	-	+	+	+	+	+	+		-	-	+	+	+	+	+	+
[sfGFP-Gly_3] (12.5μM)	-	+	-	+	+	+	+	+		-	+	-	+	+	+	+	+
[GFP-LPETG] (25μM)	+	-	-	+	+	+	+	+		+	-	-	+	+	+	+	+
Time (hr)	-	-	-	1	2	3	4	5		-	-	-	1	2	3	4	5

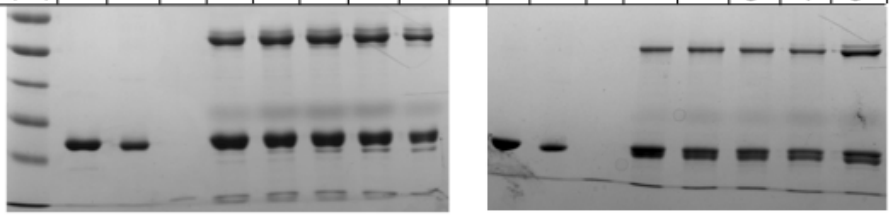


Figure 6.16. Sortase optimization. A) Depiction of linking two proteins together using sortase. B) SDS-Page analysis of sortase reaction linking αHer2-AALPETG-6xHis tag and sfGFP-$(gly)_3$ for one hour (left) and the sortase reaction between sfGFP-AALPETG-6xHis tag and sfGFP-$(gly)_3$. The sortase concentration was increased by 1μM every hour for three hours (right). C) SDS-Page analysis of sortase reactions over the course of 5 hours in either TRIS buffer (left) or PBS buffer (right).

6.3.4 The use of HaloTag technology to link a protein to the AAV capsid

HaloTag technology was developed for the purification and labeling of proteins[31,32]. It involves the use of the HaloTag protein, a 33kDa protein that is

typically genetically fused to the protein of interest, and the HaloTag ligand, a synthetic reactive chloroalkane linker that is attached to the functional group of choice. The HaloTag protein is based on a bacterial haloalkane dehalogenase enzyme from *Rhodococcus rhodochrous*. The reaction proceeds via an alkyl-enzyme intermediate during nucleophilic displacement of the linker chloride with the HaloTag protein's Asp 106. Whereas in the dehalogenase reaction, an active sit Histidine would function as a base to catalyze hydrolysis to release the enzyme, the HaloTag protein instead has a Phenylalanine, which will not catalyze hydrolysis and instead covalently links the HaloTag ligand to the HaloTag protein, connecting the protein of interest and the functional group on the HaloTag ligand.

While the HaloTag technology is most often used to link small probes to proteins of interest, we envisioned using the HaloTag technology to link a protein to the AAV capsid. We anticipated combining the HaloTag technology with the technology we developed to incorporate a Cysteine into the AAV capsid (discussed in Chapter 4). Once a reactive cysteine was present on the capsid surface, the Halotag ligand would be attached to the capsid surface via an iodoacetamide linker. The HaloTag protein would be genetically fused to the protein of interest. These two components could then be combined to covalently fuse the AAV capsid to the protein of interest (Figure 17A).

Of all the projects I explored, this was the least completed. Constructs to express sfGFP-HaloTag and NanoLuc-HaloTag were cloned, and sfGFP-HaloTag was purified with its mass confirmed by LCMS (Figure 17B). Additionally, an sfGFP-151C construct was cloned so that the chemistry between cysteine and the maleimide or

iodoacetamide linker could be optimized. sfGFP-151C was expressed, purified, and its mass was confirmed via LCMS (Figure 17C). It was possible to optimize the chemistry between sfGFP-151C and Cy5.5-maleimide (Figure 17D), and Roy Singha, a post-doc in the Chatterjee lab, synthesized an iodoacetamide-HaloTag linker, however the optimization with this linker was not as straight forward and the reaction was never confirmed or optimized.

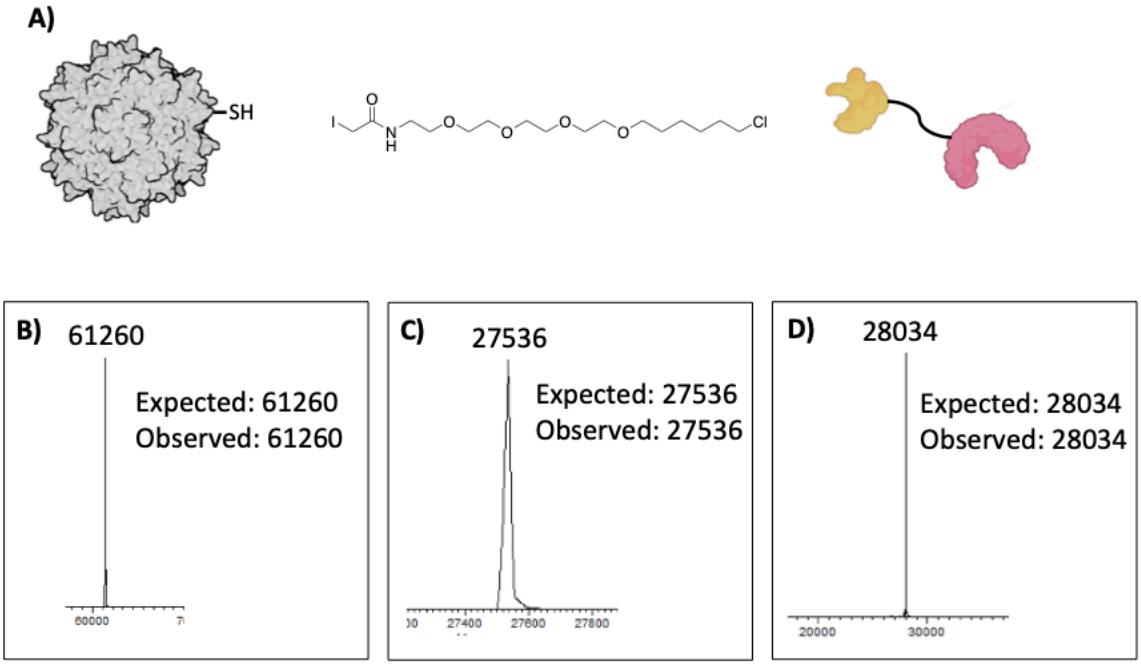


Figure 6.17. LCMS anlysis of purified proteins to be used with HaloTag technology. A) Schematic depicting HaloTag technology envisioned linking the AAV capsid to the protein of interest. B) LCMS analysis of purified sfGFP-151C. C) LCMS analysis of sfGFP-151C modified with Cy5.5-maleimide. D) LCMS analysis of sfGFP-Halotag fusion protein.

Moving forward, the chemistry between the HaloTag linker iodoacetamide and the cysteine of interest would first need to be optimized, followed by optimization of the reaction between the HaloTag-protein conjugate and the HaloTag-ligand-protein conjugate, monitored by SDS-PAGE analysis. Once this was accomplished, the next step would be to optimize the reaction using AAV conjugated to the HaloTag ligand.

6.3.4 Conclusions

I have shown the beginnings of three different methods that could be used to attach a protein to the outside of the capsid and discussed why they were unsuccessful and how they could be improved. Quan Pham, who is now in the Chatterjee lab, will continue this work, likely using the HaloTag technology. Previous work by other labs demonstrates that this is possible, however the proper tools, such as the ability to produce large-scale preparations of purified virus are necessary for this project to successfully move forward.

6.4 ACKNOWLEDGEMENTS

Thanks to Rachel Kelemen and James Italia for guidance with UAA incorporation into the AAV capsid. Partha Addy and James Italia helped tremendously with 5HTP incorporation and CRACR labeling on the αHer2-Fab.

6.5 EXPERIMENTAL PROCEDURES

6.4.1 *Plasmids:*

Primer and Plasmid sequences can be found in the appendix.

pIDTSmart-BCNKRS and pIDTSmart-SCOKRS were cloned from the starting plasmid pIDTSmart-*Mm*PylRS. AAV made with SCOK was made using pIDTSmart-RC2-TAG454-SCOKRS, pIDTSmart-8x*Mm*PytR-ITR-GFP, and pHelper.

Diazirine Lysine incorporation was achieved using a pIDTSmart-*Mb*PylRS plasmid along with pIDTSmart-RC2-TAG454, pIDTSmart-8xPytR-ITR-GFP-ITR-MbPylRS, and pHelper.

5HTP, AzW, and 5PrW were incorporated into the AAV capsid using pIDT-ITR-GFP-4xwtR-wRS, pHelper, and pIDTSmart-RC2-454TGA, pIDTSmart-RC2-ΔVP1-CMV-VP1-454TGA or pIDTSmart-RC2-ΔVP2-CMV-VP2-454TGA.

The αHer2-Fab was expressed using pEvol-α-Her2-Fab and pUltra-*Mj*Y-CNF for either wild type expression or with AzF incorporation. For 5HTP incorporation, pBK-α-Her2-Fab and pEvoltac-EcW-TGA.h14 were used, and for incorporation of both 5HTP and AzF, pBK-α-Her2-Fab and pEvoltac-EcW-H14-*Mj*Y-CNF.TAG

Pet29-eSrtA, pENTR4-Halotag, and pUAS-NanoLuc were purchased from addgene (Plasmid #75144, 29644, and 87696, respectively). pET22b-T5-NanoLuc-Halotag and pET22b-T5-sfGFP-Halotag were cloned using overlap PCR. Primer sequences are in the appendix.

6.4.2 *Antibody expression and purification*

The α-Her2-Fab sequence was obtained and cloned into a pBK vector using gBlocks (IDT). TAG or TGA stop codons were introduced into the Fab using overlap PCR. For AzF incorporation, pEvol-α-Her2-Fab and pUltra-*Mj*Y-CNF were co-transformed into electro-competent Top10 cells made in house. For 5HTP incorporation, pBK-α-Her2-Fab and pEvoltac-EcW-TGA.h14 were co-transformed into *E. coli* strain ATMW1, and for incorporation of both 5HTP and AzF, pBK-α-Her2-Fab and pEvoltac-EcW-H14-*Mj*Y-CNF.TAG were co-transformed into *E. coli* strain ATMW1. A single colony was picked and a starter culture was grown and used to inoculate a larger culture. Cells were grown in LB supplemented with 35μg/mL chloramphenicol and (100)μg/mL spectinomycin. Protein expression was induced with 0.02% arabinose and 1mM IPTG upon reaching an OD_{600} of 0.6 and the culture was also supplemented with 1mM AzF. Cells were then allowed to grow for 16 hours at 30˚C in shake flasks, harvested by centrifugation, and resuspended in a periplasmic lysis buffer (20% sucrose, 30mM Tris, pH 8, 1mM EDTA, 0.2 mg/mL lysozyme, and Halt protease inhibitor) for 30 minutes at 37˚C. Lysate was then diluted 1:1 with binding buffer (50mM NaOAc, pH 5.2) and clarified by centrifugation at 1700 rpm for 30 minutes. α-Her2-Fab was then purified using Pierce Protein G resin according to the manufacturer's instruction. Following purification, elutions that contained protein were identified using Bradford reagent. These elutions were pooled and dialzed against PBS overnight at 4˚C. The following day, antibody was concentrated.

6.4.3 *Antibody labeling*

Fluorescein-diazonium compound was freshly prepared by dissolving 10mM 6-aminofluorescein in 10mM HCl. A 60mM sodium nitrite solution was prepared and both solutions were cooled on ice. Next, 20μL $NaNo_2$ (60mM) was added to 100 μL of fluorescein, vortexed for 10s, and cooled on ice. This made an 8.3mM fluorescein-dazonium solution, which was further diluted to 1mM with water immediately prior to the labeling studies.

4μM α-Her2-Fab was incubated with 80μM fluorescein-diazonium compound for 15 minutes on ice. This reaction was quenched using 0.5mM 5HTP. Labeled antibody was dialyzed against PBS overnight at 4˚C.

6.4.4 *Cell Culture*

Adherent SK-BR-3 cells were purchased from ATCC. They were grown in DMEM-High glucose (HyClone) supplemented with 100 U/mL penicillin/streptomycin (HyClone) and 10% fetal bovine serum (FBS; Corning) and maintained at 37˚C (humidified) and 5% CO_2. HEK293T cells were maintained as described in chapter 2.

6.4.5 *Association of fluorescein-labeled α-Her2-Fab to SK-BR-3 cells*

SK-BR-3 cells were plated onto a 12-well dish. Upon reaching confluency, cells were detached from the plate using warm 0.25% trypsin (Hyclone) for about 2 minutes at 37˚C. Trypsin was quenched with ice-cold DMEM+10% FBS and cells were transferred to a microcentrifuge tube. The DMEM/FBS/Tryspin mix was removed by centrifuging cells at 2,000xg for 5 minutes, removing the liquid, and re-suspending in ice-cold PBS. Fluorescein-labeled α-Her2-Fab was added to SK-BR-3 cells at 4˚C with gentle agitation for one hour. Cells were then washed once with ice cold PBS and re-suspended in ice-cold PBS. They were then analyzed by flow cytometry using a Bio-Rad S3e cell sorter. All flow cytometry plots were generated in ProSort (BioRad).

6.4.6 *Unnatural amino acids*

SCO-L-Lysine was purchased from Sirius fine chemicals (SIChem, SC-8001). Diazirine Lysine was synthesized by Raja Mukherjee. 5HTP was purchased from Chem-Impex Int'L (00607), and 5PrW and 5AzW were synthesized in house.

6.4.7 *Protein Affinity Purifications*

All variations of sfGFP, and sortase were all purified using a 6x Histidine tag and Nickel affinity resin. Purifications were performed as per the manufacturer's protocol.

αHer2-LPETG was purified using protein G resin, as previously described[20].

6.4.8 *DBCO-Peg$_4$-DBCO Click reaction and streptavidin agarose pull down*

For complete labeling of sfGFP-AzF with DBCO-Peg$_4$-DBCO, 25μM DBCO-Peg$_4$-DBCO was combined with 5μM sfGFP-AzF in PBS. The reaction proceeded for 2 hours in the dark, with gentle agitation. To remove excess DBCO-Peg$_4$-DBCO, the reaction was dialzed against PBS overnight at 4˚C. The final product was analyzed on LCMS.

sfGFP-Peg$_4$-DBCO was then combined with a constant volume of PEG precipitated AAV-454C. The reaction was allowed to incubate in the dark with gentle agitation for varying amounts of time and quenched with a 10-fold excess of biotin azide. To reduce the amount of streptavidin agarose needed, the excess biotin azide was removed by another overnight dialysis against PBS at 4˚C. Excess sfGFP-Peg$_4$-DBCO was removed by doing a pull down with the streptavidin agarose. The amount of agarose was calculated using the manufacturer's binding capacity. Streptavidin agarose was pipetted into a microcentrifuge tube and washed with PBS 3 times. Between each step the beads were pelleted by centrifugation at 500xg for 1 minute and the excess buffer was removed by careful pipetting. The AAV/sfGFP-Peg$_4$-DBCO mixture was then added to the streptavidin agarose and mixed with gentle agitation at 4˚C for one hour. The beads were pelleted by centrifugation at 500xg for 1 minute and the supernatant was pipetted off and added to confluent SK-BR-3 cells with sodium butyrate.

6.4.9 *Sortase reaction conditions*

The optimized sortase reaction used 1μM pure eSortA, 25μM sfGFP-LPETG (or αHer2-LPETG), and 50μM GFP-gly$_3$, in 300mM Tris pH 7.5, 150mM NaCl, and 5mM $CaCl_2$ for 1 hour. After 1 hour, 5μL was removed and the reaction was quenched with SDS Loading buffer and heated to 100˚C prior to analysis on a 12% SDS-Page gel.

6.6 REFERENCES

(1) Italia, J. S.; Zheng, Y.; Kelemen, R. E.; Erickson, S. B.; Addy, P. S.; Chatterjee, A. Expanding the Genetic Code of Mammalian Cells. *Biochem. Soc. Trans.* **2017**, *45* (2), 555–562. https://doi.org/10.1042/BST20160336.

(2) Dommerholt, J.; Van Rooijen, O.; Borrmann, A.; Guerra, C. F.; Bickelhaupt, F. M.; Van Delft, F. L. Highly Accelerated Inverse Electron-Demand Cycloaddition of Electron-Deficient Azides with Aliphatic Cyclooctynes. *Nat. Commun.* **2014**, *5*, 1–7. https://doi.org/10.1038/ncomms6378.

(3) Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A. Genetically Encoded

Copper-Free Click Chemistry. *Angew. Chemie - Int. Ed.* **2011**, *50* (17), 3878–3881. https://doi.org/10.1002/anie.201008178.
(4) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. Genetic Encoding of Bicyclononynes and Trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels-Alder Reactions. *J. Am. Chem. Soc.* **2012**, *134* (25), 10317–10320. https://doi.org/10.1021/ja302832g.
(5) McCraw, D. M.; O'Donnell, J. K.; Taylor, K. A.; Stagg, S. M.; Chapman, M. S. Structure of Adeno-Associated Virus-2 in Complex with Neutralizing Monoclonal Antibody A20. *Virology* **2012**, *431* (1–2), 40–49. https://doi.org/10.1016/j.virol.2012.05.004.
(6) Ai, H.; Shen, W.; Sagi, A.; Chen, P. R.; Schultz, P. G. Probing Protein-Protein Interactions with a Genetically Encoded Photo-Crosslinking Amino Acid. *ChemBioChem* **2011**, *12* (12), 1854–1857. https://doi.org/10.1002/cbic.201100194.
(7) Jr, K. H. W.; Gorbatyuk, O. S.; Jeffrey, K.; Opie, S. R.; Zolotukhin, S.; Warrington, K. H.; Harrison, J. K.; Muzyczka, N. Adeno-Associated Virus Type 2 VP2 Capsid Protein Is Nonessential and Can Tolerate Large Peptide Insertions at Its N Terminus. *J. Virol.* **2004**, *78* (12), 6595–6609. https://doi.org/10.1128/JVI.78.12.6595.
(8) Hallam, T. J.; Wold, E.; Wahl, A.; Smider, V. V. Antibody Conjugates with Unnatural Amino Acids. *Mol. Pharm.* **2015**, *12* (6), 1848–1862. https://doi.org/10.1021/acs.molpharmaceut.5b00082.
(9) Agarwal, P.; Bertozzi, C. R. Site-Specific Antibody-Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjug. Chem.* **2015**, *26* (2), 176–192. https://doi.org/10.1021/bc5004982.
(10) Hutchins, B. M.; Kazane, S. A.; Staflin, K.; Forsyth, J. S.; Felding-Habermann, B.; Schultz, P. G.; Smider, V. V. Site-Specific Coupling and Sterically Controlled Formation of Multimeric Antibody Fab Fragments with Unnatural Amino Acids. *J. Mol. Biol.* **2011**, *406* (4), 595–603. https://doi.org/10.1016/j.jmb.2011.01.011.
(11) Kim, C. H.; Axup, J. Y.; Dubrovska, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G. Synthesis of Bispecific Antibodies Using Genetically Encoded Unnatural Amino Acids. *J. Am. Chem. Soc.* **2012**, *134* (24), 9918–9921. https://doi.org/10.1021/ja303904e.
(12) Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; et al. Synthesis of Site-Specific Antibody-Drug Conjugates Using Unnatural Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (40), 16101–16106. https://doi.org/10.1073/pnas.1211023109.
(13) Lu, H.; Wang, D.; Kazane, S.; Javahishvili, T.; Tian, F.; Song, F.; Sellers, A.; Barnett, B.; Schultz, P. G. Site-Specific Antibody-Polymer Conjugates for SiRNA Delivery. *J. Am. Chem. Soc.* **2013**, *135* (37), 13885–13891. https://doi.org/10.1021/ja4059525.
(14) Kularatne, S. A.; Deshmukh, V.; Ma, J.; Tardif, V.; Lim, R. K. V.; Pugh, H. M.; Sun,

Y.; Manibusan, A.; Sellers, A. J.; Barnett, R. S.; et al. A CXCR4-Targeted Site-Specific Antibody-Drug Conjugate. *Angew. Chemie - Int. Ed.* **2014**, *53* (44), 11863–11867. https://doi.org/10.1002/anie.201408103.

(15) Zimmerman, E. S.; Heibeck, T. H.; Gill, A.; Li, X.; Murray, C. J.; Madlansacay, M. R.; Tran, C.; Uter, N. T.; Yin, G.; Rivers, P. J.; et al. Production of Site-Specific Antibody-Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System. *Bioconjug. Chem.* **2014**, *25* (2), 351–361. https://doi.org/10.1021/bc400490z.

(16) Oza, J. P.; Aerni, H. R.; Pirman, N. L.; Barber, K. W.; Ter Haar, C. M.; Rogulina, S.; Amrofell, M. B.; Isaacs, F. J.; Rinehart, J.; Jewett, M. C. Robust Production of Recombinant Phosphoproteins Using Cell-Free Protein Synthesis. *Nat. Commun.* **2015**, *6*. https://doi.org/10.1038/ncomms9168.

(17) Dondapati, S. K.; Stech, M.; Zemella, A.; Kubick, S. Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. *BioDrugs* **2020**, *34* (3), 327–348. https://doi.org/10.1007/s40259-020-00417-y.

(18) Xiao, H.; Chatterjee, A.; Choi, S.; Bajjuri, K. M.; Sinha, S. C.; Schultz, P. G. Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells. *Angew. Chemie* **2013**, *125* (52), 14330–14333. https://doi.org/10.1002/ange.201308137.

(19) 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. *Nat. Chem. Biol.* **2017**, *13* (4), 446–450. https://doi.org/10.1038/nchembio.2312.

(20) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **2017**, *139* (34), 11670–11673. https://doi.org/10.1021/jacs.7b05125.

(21) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An Evolved Aminoacyl-TRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **2011**, *50* (11), 1894–1900. https://doi.org/10.1021/bi101929e.

(22) Kotterman, M. A.; Schaffer, D. V. Engineering Adeno-Associated Viruses for Clinical Gene Therapy. *Nat. Rev. Genet.* **2014**, *15* (7), 445–451. https://doi.org/10.1038/nrg3742.

(23) Münch, R. C.; Janicki, H.; Völker, I.; Rasbach, A.; Hallek, M.; Büning, H.; Buchholz, C. J. Displaying High-Affinity Ligands on Adeno-Associated Viral Vectors Enables Tumor Cell-Specific and Safe Gene Transfer. *Mol. Ther.* **2013**, *21* (1), 109–118. https://doi.org/10.1038/mt.2012.186.

(24) Münch, R. C.; Muth, A.; Muik, A.; Friedel, T.; Schmatz, J.; Dreier, B.; Trkola, A.; Plückthun, A.; Büning, H.; Buchholz, C. J. Off-Target-Free Gene Delivery by Affinity-Purified Receptor-Targeted Viral Vectors. *Nat. Commun.* **2015**, *6*, 2–10. https://doi.org/10.1038/ncomms7246.

(25) Yang, Q.; Mamounas, M.; Yu, G.; Kennedy, S.; Leaker, B.; Mersn, J.; Yu, F. W. M. A. N. G.; Barber, J. R.; Al, Y. E. T. Development of Novel Cell Surface CD34-Targeted Recombinant Adenoassociated Virus for Gene Therapy. **1998**, *1937*, 1929–1937.

(26) Hagen, S.; Baumann, T.; Wagner, H. J.; Morath, V.; Kaufmann, B.; Fischer, A.; Bergmann, S.; Schindler, P.; Arndt, K. M.; Müller, K. M. Modular Adeno-

Associated Virus (RAAV) Vectors Used for Cellular Virus-Directed Enzyme Prodrug Therapy. *Sci. Rep.* **2014**, *4*, 1–11. https://doi.org/10.1038/srep03759.
(27) Asokan, A.; Johnson, J. S.; Li, C.; Samulski, R. J. Bioluminescent Virion Shells: New Tools for Quantitation of AAV Vector Dynamics in Cells and Live Animals. *Gene Ther.* **2008**, *15* (24), 1618–1622. https://doi.org/10.1038/gt.2008.127.
(28) Pishesha, N.; Ingram, J. R.; Ploegh, H. L. Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol.* **2018**, *34* (1), 163–188. https://doi.org/10.1146/annurev-cellbio-100617-062527.
(29) Theile, C. S.; Witte, M. D.; Blom, A. E. M.; Kundrat, L.; Ploegh, H. L.; Guimaraes, C. P. Site-Specific N-Terminal Labeling of Proteins Using Sortase-Mediated Reactions. *Nat. Protoc.* **2013**, *8* (9), 1800–1807. https://doi.org/10.1038/nprot.2013.102.
(30) Chen, I.; Dorr, B. M.; Liu, D. R. A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11399–11404. https://doi.org/10.1073/pnas.1101046108.
(31) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; et al. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3* (6), 373–382. https://doi.org/10.1021/cb800025k.
(32) England, C. G.; Luo, H.; Cai, W. HaloTag Technology: A Versatile Platform for Biomedical Applications. *Bioconjug. Chem.* **2015**, *26* (6), 975–986. https://doi.org/10.1021/acs.bioconjchem.5b00191.

Appendix 1

Plasmid maps and sequences

List of Plasmids

Chapter 2 Plasmids

pIDTSmart-8xU6MmPytR-ITR-GFP-MbPylRS

Orange=MbPylRS
Yellow=MmPytR
Red=CMV Promoter

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGT
GGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCT
GAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGG
CGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGGCGGCCGCAAATA
CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCC
CGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC
AGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTGCGGCCGCACGCGTG
GAGCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATAT
ATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC
CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC
GTCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGA
CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAT
GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG
GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA
GTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATA
GAAGACACCGGGACCGATCCAGCCTCCGCGGATTCGAATCCCGGCCGGGAAC
GGTGCATTGGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTACCGCCTA
TAGAGTCTATAGGCCCACAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTT
TATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATAC
AATGTATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGG
GTTAAGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTA
ACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATT
CTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAG
GCCCTTTTGCTAATCATGTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGC
AACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTGGGATTCGA
ACATCGATTGAATTCTGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG
TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAA
GTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA
CCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCA
GCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCG
AGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG
AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACA

ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAA
GATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACC
TGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAG
CTGTACAAGTACTCAGATCTCGAGCTCAAGTAGGGATCCTCTAGAGTCGACC
TGCAGAAGCTTGCCTCGAGCAGCGCTGCTCGAGAGATCTACGGGTGGCATCC
CTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGC
CCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTG
TCCTTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAG
TTGGGAAGACAACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAG
TGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGA
TTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATGCATGACCAGG
CTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCAG
GCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAAT
TGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGT
AGGTAACCACGTGCGGACCGAGCGGCCGCAGGAACCCCTAGTGATGGAGTT
GGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAG
GTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGC
GCAGCTGCCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATA
TACGCGTTGACATTGATTATTGACTAGGTCGGGCAGGAAGAGGGCCTATTTC
CCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATT
AGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTA
GAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATG
GACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATAT
ATCTTGTGGAAAGGACGAAACACCGGAAACCTGATCATGTAGATCGAACGGA
CTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGTTTTTTGCTAGG
TCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGA
TACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGA
TATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCA
GTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA
GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGAA
ACCTGATCATGTAGATCGAACGGACTCTAAATCCGTTCAGCCGGGTTAGATTC
CCGGGGTTTCCGTTTTTTGCTAGGTCGGGCAGGAAGAGGGCCTATTTCCCATG
ATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAAT
TAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAG
TAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTA
TCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTT
GTGGAAAGGACGAAACACCGGAAACCTGATCATGTAGATCGAACGGACTCT
AAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGTTTTTTGCTAGGTCGG
GCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA
AGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATT
AGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTT
TAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGAAACCT
GATCATGTAGATCGAACGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCG

GGGTTTCCGTTTTTTGCTAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT
CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAA
TTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAA
TAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCAT
ATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGG
AAAGGACGAAACACCGGAAACCTGATCATGTAGATCGAACGGACTCTAAATC
CGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGTTTTTTGCTAGGTCGGGCAGG
AAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCT
GTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTAC
AAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAA
ATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
ATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGAAACCTGATC
ATGTAGATCGAACGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGT
TTCCGTTTTTTGCTAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTT
CATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTG
ACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAAT
TTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATG
CTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAA
GGACGAAACACCGGAAACCTGATCATGTAGATCGAACGGACTCTAAATCCGT
TCAGCCGGGTTAGATTCCCGGGGTTTCCGTTTTTTGCTAGGTCGGGCAGGAAG
AGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTT
AGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAA
ATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTA
TGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTT
CTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGAAACCTGATCATG
TAGATCGAACGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCC
GTTTTTTGCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCC
CATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGA
CCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAA
ACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTAT
TGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACC
TTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC
CATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAC
TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTG
GCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTG
ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTC
TCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACT
CACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGATAAAAAACCATTA
GATGTTTTAATATCTGCGACCGGGCTCTGGATGTCCAGGACTGGCACGCTCCA
CAAAATCAAGCACCATGAGGTCTCAAGAAGTAAAATATACATTGAAATGGCG
TGTGGAGACCATCTTGTTGTGAATAATTCCAGGAGTTGTAGAACAGCCAGAG
CATTCAGACATCATAAGTACAGAAAAACCTGCAAACGATGTAGGGTTTCGGA
CGAGGATATCAATAATTTTCTCACAAGATCAACCGAAAGCAAAAACAGTGTG
AAAGTTAGGGTAGTTTCTGCTCCAAAGGTCAAAAAAGCTATGCCGAAATCAG
TTTCAAGGGCTCCGAAGCCTCTGGAAAATTCTGTTTCTGCAAAGGCATCAACG

AACACATCCAGATCTGTACCTTCGCCTGCAAAATCAACTCCAAATTCGTCTGT
TCCCGCATCGGCTCCTGCTCCTTCACTTACAAGAAGCCAGCTTGATAGGGTTG
AGGCTCTCTTAAGTCCAGAGGATAAAATTTCTCTAAATATGGCAAAGCCTTTC
AGGGAACTTGAGCCTGAACTTGTGACAAGAAGAAAAAACGATTTTCAGCGGC
TCTATACCAATGATAGAGAAGACTACCTCGGTAAACTCGAACGTGATATTAC
GAAATTTTTCGTAGACCGGGGTTTTCTGGAGATAAAGTCTCCTATCCTTATTC
CGGCGGAATACGTGGAGAGAATGGGTATTAATAATGATACTGAACTTTCAAA
ACAGATCTTCCGGGTGGATAAAAATCTCTGCTTGAGGCCAATGCTTGCCCCG
ACTCTTTACAACTATCTGCGAAAACTCGATAGGATTTTACCAGGCCCAATAAA
AATTTTCGAAGTCGGACCTTGTTACCGGAAAGAGTCTGACGGCAAAGAGCAC
CTGGAAGAATTTACTATGGTGAACTTCTGTCAGATGGGTTCGGGATGTACTCG
GGAAAATCTTGAAGCTCTCATCAAAGAGTTTCTGGACTATCTGGAAATCGAC
TTCGAAATCGTAGGAGATTCCTGTATGGTCTTTGGGGATACTCTTGATATAAT
GCACGGGGACCTGGAGCTTTCTTCGGCAGTCGTCGGGCCAGTTTCTCTTGATA
GAGAATGGGGTATTGACAAACCATGGATAGGTGCAGGTTTTGGTCTTGAACG
CTTGCTCAAGGTTATGCACGGCTTTAAAAACATTAAGAGGGCATCAAGGTCC
GAATCTTACTATAATGGGATTTCAACCAATCTGTAAGAATTCAACGCGTTAAG
TCGACTTTAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGA
CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCT
TGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATT
GCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG
CGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCCTAGGGGTGCGAGCGGA
TCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACCCGTGATC
TTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGC
TTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCATACCT
GACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTTGAT
CGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGG
GCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGAG
CCATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAACATG
GATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAG
GTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTG
AAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGGC
TAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGT
ACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACAGCAT
TCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTG
GCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAAC
GGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTT
GGTTGGTGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAA
GTCTGGAAAGAAATGCATAAACTCTTGCCATTCTCACCGGATTCAGTCGTCAC
TCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAG
GTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGC
CATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTT
TTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCACTTG
ATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCTCACCT
TCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTG

ACGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAG
GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAG
CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC
GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAA
CAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTT
TGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
TTTTCTACCGAAGAAAGGCCCA

pIDTSmart-RC2

Orange=*Rep*
Blue=*Cap*
Green=AAP

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGT
GGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCT
GAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGG
CGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGGCGGCCGCTCTAG
AACTAGTGGATCCCCCGGAAGATCAGAAGTTCCTATTCCGAAGTTCCTATTCT
CTAGAAAGTATAGGAACTTCTGATCTGCGCAGCCGCCATGCCGGGGTTTTAC
GAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCGGCATTTC
TGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGAT
TCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGA
AGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGA
GGCCCTTTTCTTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACG
TGCTCGTGGAAACCACCGGGGTGAAATCCATGGTTTTGGGACGTTTCCTGAGT
CAGATTCGCGAAAAACTGATTCAGAGAATTTACCGCGGGATCGAGCCGACTT
TGCCAAACTGGTTCGCGGTCACAAAGACCAGAAATGGCGCCGGAGGCGGGA
ACAAGGTGGTGGATGAGTGCTACATCCCCAATTACTTGCTCCCCAAAACCCA
GCCTGAGCTCCAGTGGGCGTGGACTAATATGGAACAGTATTTAAGCGCCTGT
TTGAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGT
CGCAGACGCAGGAGCAGAACAAAGAGAATCAGAATCCCAATTCTGATGCGC
CGGTGATCAGATCAAAAACTTCAGCCAGGTACATGGAGCTGGTCGGGTGGCT
CGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAGGC
CTCATACATCTCCTTCAATGCGGCCTCCAACTCGCGGTCCCAAATCAAGGCTG
CCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCCGACTA
CCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAA
ATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGG
ATGGGCCACGAAAAAGTTCGGCAAGAGGAACACCATCTGGCTGTTTGGGCCT
GCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCT
TCTACGGGTGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTC

GACAAGATGGTGATCTGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCGTG
GAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGGTGCGCGTGGACCAGAAA
TGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTGATCGTCACCTCCAACA
CCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAACACCAGCA
GCCGTTGCAAGACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATG
ACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAA
GGATCACGTGGTTGAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCC
AAGAAAAGACCCGCCCCCAGTGACGCAGATATAAGTGAGCCCAAACGGGTG
CGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCGGAAGCTTCGATCAACT
ACGCAGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGAT
GCTGTTTCCCTGCAGACAATGCGAGAGAATGAATCAGAATTCAAATATCTGC
TTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCCCGTGTCAGAATCTCA
ACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATA
TCATGGGAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGA
TTTGGATGACTGCATCTTTGAACAATAAATGATTTAAATCAGGTATGGCTGCC
GATGGTTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAAGAC
AGTGGTGGAAGCTCAAACCTGGCCCACCACCACCAAAGCCCGCAGAGCGGC
ATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTACAAGTACCTCGGACC
CTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGC
CCTCGAGCACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCC
GTACCTCAAGTACAACCACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAA
GATACGTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGA
GGGTTCTTGAACCTCTGGGCCTGGTTGAGGAACCTGTTAAGACGGCTCCGGG
AAAAAAGAGGCCGGTAGAGCACTCTCCTGTGGAGCCAGACTCCTCCTCGGGA
ACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAG
ACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCCACCAG
CAGCCCCCTCTGGTCTGGGAACTAATACGATGGCTACAGGCAGTGGCGCACC
AATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAA
TTGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACC
CGAACCTGGGCCCTGCCCACCTACAACAACCACCTCTACAAACAAATTTCCA
GCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTACAGCACCCCTTG
GGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACCACGTGACTGGC
AAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGAGACTCAACTTCAA
GCTCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGACGACG
ATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCA
GCTCCCGTACGTCCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAG
CAGACGTCTTCATGGTGCCACAGTATGGATACCTCACCCTGAACAACGGGAG
TCAGGCAGTAGGACGCTCTTCATTTTACTGCCTGGAGTACTTTCCTTCTCAGA
TGCTGCGTACCGGAAACAACTTTACCTTCAGCTACACTTTTGAGGACGTTCCT
TTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCCTCT
CATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACTCCAAGTGGAACC
ACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCGGG
ACCAGTCTAGGAACTGGCTTCCTGGACCCTGTTACCGCCAGCAGCGAGTATC
AAAGACATCTGCGGATAACAACAACAGTGAATACTCGTGGACTGGAGCTACC
AAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATGG
CAAGCCACAAGGACGATGAAGAAAAGTTTTTTCCTCAGAGCGGGGTTCTCAT

CTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTGAAAAGGTCAT
GATTACAGACGAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGGAGCA
GTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACC
GCAGATGTCAACACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGAG
ATGTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACA
TTTTCACCCCTCTCCCCTCATGGGTGGATTCGGACTTAAACACCCTCCTCCAC
AGATTCTCATCAAGAACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGT
GCGGCAAAGTTTGCTTCCTTCATCACACAGTACTCCACGGGACAGGTCAGCG
TGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAACGCTGGAATCCCG
AAATTCAGTACACTTCCAACTACAACAAGTCTGTTAATGTGGACTTTACTGTG
GACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAGATACCTGA
CTCGTAATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTTG
AACTTTGGTCTCTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATA
AGTAGCATGGCGGGTTAATCATTAACTACAGCCCGGGCGTTTAAACAGCGGG
CGGAGGGGTGGAGTCGTGACGTGAATTACGTCATAGGGTTAGGGAGGTCCTG
TATTAGAGGTCACGTGAGTGTTTTGCGACATTTTGCGACACCATGTGGTCTCG
CTGGGGGGGGGGGCCCGAGTGAGCACGCAGGGTCTCCATTTTGAAGCGGGA
GGTTTGAACGAGCGCTGGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGAC
TGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTT
CGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCATGCA
TCGGCCGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGC
CAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTA
CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCA
TATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGG
CATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC
GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC
GTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA
GGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGG
CTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGGGGT
GCGAGCGGATCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGA
CCCGTGATCTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTAT
CTAGTCAGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAA
ATCATACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTT
TGACTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGAT
CACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCT
AAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAATT
CCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGG
GCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAG
TTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGA
TGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCAT
TTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAA

AACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTT
GATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTG
TCCTTTTAACGGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGA
ATAACGGTTTGGTTGGTGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCT
GTTGAACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCTCACCGGATT
CAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGG
AAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACC
AGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAG
AAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCA
GTTTCACTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACC
TGGCTCACCTTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCC
GCCCCCCTGACGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTG
CGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCT
TCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGT
GTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG
GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
ACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGG
AAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pHelper

Red=E2A
Blue=E4
Green=VA

*E2A, E4, and VA are Adenoviral Helper genes

GGTACCCAACTCCATGCTTAACAGTCCCCAGGTACAGCCCACCCTGCGTCGC
AACCAGGAACAGCTCTACAGCTTCCTGGAGCGCCACTCGCCCTACTTCCGCA
GCCACAGTGCGCAGATTAGGAGCGCCACTTCTTTTTGTCACTTGAAAAACATG
TAAAAATAATGTACTAGGAGACACTTTCAATAAAGGCAAATGTTTTTATTTGT
ACACTCTCGGGTGATTATTTACCCCCCACCCTTGCCGTCTGCGCCGTTTAAAA
ATCAAAGGGGTTCTGCCGCGCATCGCTATGCGCCACTGGCAGGGACACGTTG
CGATACTGGTGTTTAGTGCTCCACTTAAACTCAGGCACAACCATCCGCGGCA
GCTCGGTGAAGTTTTCACTCCACAGGCTGCGCACCATCACCAACGCGTTTAGC
AGGTCGGGCGCCGATATCTTGAAGTCGCAGTTGGGGCCTCCGCCCTGCGCGC
GCGAGTTGCGATACACAGGGTTGCAGCACTGGAACACTATCAGCGCCGGGTG
GTGCACGCTGGCCAGCACGCTCTTGTCGGAGATCAGATCCGCGTCCAGGTCC
TCCGCGTTGCTCAGGGCGAACGGAGTCAACTTTGGTAGCTGCCTTCCCAAAA
AGGGTGCATGCCCAGGCTTTGAGTTGCACTCGCACCGTAGTGGCATCAGAAG
GTGACCGTGCCCGGTCTGGGCGTTAGGATACAGCGCCTGCATGAAAGCCTTG

ATCTGCTTAAAAGCCACCTGAGCCTTTGCGCCTTCAGAGAAGAACATGCCGC
AAGACTTGCCGGAAAACTGATTGGCCGGACAGGCCGCGTCATGCACGCAGCA
CCTTGCGTCGGTGTTGGAGATCTGCACCACATTTCGGCCCCACCGGTTCTTCA
CGATCTTGGCCTTGCTAGACTGCTCCTTCAGCGCGCGCTGCCCGTTTTCGCTC
GTCACATCCATTTCAATCACGTGCTCCTTATTTATCATAATGCTCCCGTGTAG
ACACTTAAGCTCGCCTTCGATCTCAGCGCAGCGGTGCAGCCACAACGCGCAG
CCCGTGGGCTCGTGGTGCTTGTAGGTTACCTCTGCAAACGACTGCAGGTACGC
CTGCAGGAATCGCCCCATCATCGTCACAAAGGTCTTGTTGCTGGTGAAGGTC
AGCTGCAACCCGCGGTGCTCCTCGTTTAGCCAGGTCTTGCATACGGCCGCCA
GAGCTTCCACTTGGTCAGGCAGTAGCTTGAAGTTTGCCTTTAGATCGTTATCC
ACGTGGTACTTGTCCATCAACGCGCGCGCAGCCTCCATGCCCTTCTCCCACGC
AGACACGATCGGCAGGCTCAGCGGGTTTATCACCGTGCTTTCACTTTCCGCTT
CACTGGACTCTTCCTTTTCCTCTTGCGTCCGCATACCCCGCGCCACTGGGTCG
TCTTCATTCAGCCGCCGCACCGTGCGCTTACCTCCCTTGCCGTGCTTGATTAG
CACCGGTGGGTTGCTGAAACCCACCATTTGTAGCGCCACATCTTCTCTTTCTT
CCTCGCTGTCCACGATCACCTCTGGGGATGGCGGGCGCTCGGGCTTGGGAGA
GGGGCGCTTCTTTTTCTTTTTGGACGCAATGGCCAAATCCGCCGTCGAGGTCG
ATGGCCGCGGGCTGGGTGTGCGCGGCACCAGCGCATCTTGTGACGAGTCTTC
TTCGTCCTCGGACTCGAGACGCCGCCTCAGCCGCTTTTTTGGGGGCGCGCGGG
GAGGCGGCGGCGACGGCGACGGGGACGACACGTCCTCCATGGTTGGTGGAC
GTCGCGCCGCACCGCGTCCGCGCTCGGGGGTGGTTTCGCGCTGCTCCTCTTCC
CGACTGGCCATTTCCTTCTCCTATAGGCAGAAAAAGATCATGGAGTCAGTCG
AGAAGGAGGACAGCCTAACCGCCCCCTTTGAGTTCGCCACCACCGCCTCCAC
CGATGCCGCCAACGCGCCTACCACCTTCCCCGTCGAGGCACCCCCGCTTGAG
GAGGAGGAAGTGATTATCGAGCAGGACCCAGGTTTTGTAAGCGAAGACGAC
GAGGATCGCTCAGTACCAACAGAGGATAAAAAGCAAGACCAGGACGACGCA
GAGGCAAACGAGGAACAAGTCGGGCGGGGGGACCAAAGGCATGGCGACTAC
CTAGATGTGGGAGACGACGTGCTGTTGAAGCATCTGCAGCGCCAGTGCGCCA
TTATCTGCGACGCGTTGCAAGAGCGCAGCGATGTGCCCCTCGCCATAGCGGA
TGTCAGCCTTGCCTACGAACGCCACCTGTTCTCACCGCGCGTACCCCCCAAAC
GCCAAGAAAACGGCACATGCGAGCCCAACCCGCGCCTCAACTTCTACCCCGT
ATTTGCCGTGCCAGAGGTGCTTGCCACCTATCACATCTTTTTCCAAAACTGCA
AGATACCCCTATCCTGCCGTGCCAACCGCAGCCGAGCGGACAAGCAGCTGGC
CTTGCGGCAGGGCGCTGTCATACCTGATATCGCCTCGCTCGACGAAGTGCCA
AAAATCTTTGAGGGTCTTGGACGCGACGAGAAACGCGCGGCAAACGCTCTGC
AACAAGAAAACAGCGAAAATGAAAGTCACTGTGGAGTGCTGGTGGAACTTG
AGGGTGACAACGCGCGCCTAGCCGTGCTGAAACGCAGCATCGAGGTCACCCA
CTTTGCCTACCCGGCACTTAACCTACCCCCCAAGGTTATGAGCACAGTCATGA
GCGAGCTGATCGTGCGCCGTGCACGACCCCTGGAGAGGGATGCAAACTTGCA
AGAACAAACCGAGGAGGGCCTACCCGCAGTTGGCGATGAGCAGCTGGCGCG
CTGGCTTGAGACGCGCGAGCCTGCCGACTTGGAGGAGCGACGCAAGCTAATG
ATGGCCGCAGTGCTTGTTACCGTGGAGCTTGAGTGCATGCAGCGGTTCTTTGC
TGACCCGGAGATGCAGCGCAAGCTAGAGGAAACGTTGCACTACACCTTTCGC
CAGGGCTACGTGCGCCAGGCCTGCAAAATTTCCAACGTGGAGCTCTGCAACC
TGGTCTCCTACCTTGGAATTTTGCACGAAAACCGCCTCGGGCAAAACGTGCTT
CATTCCACGCTCAAGGGCGAGGCGCGCCGCGACTACGTCCGCGACTGCGTTT

ACTTATTTCTGTGCTACACCTGGCAAACGGCCATGGGCGTGTGGCAGCAATG
CCTGGAGGAGCGCAACCTAAAGGAGCTGCAGAAGCTGCTAAAGCAAAACTT
GAAGGACCTATGGACGGCCTTCAACGAGCGCTCCGTGGCCGCGCACCTGGCG
GACATTATCTTCCCCGAACGCCTGCTTAAAACCCTGCAACAGGGTCTGCCAG
ACTTCACCAGTCAAAGCATGTTGCAAAACTTTAGGAACTTTATCCTAGAGCGT
TCAGGAATTCTGCCCGCCACCTGCTGTGCGCTTCCTAGCGACTTTGTGCCCAT
TAAGTACCGTGAATGCCCTCCGCCGCTTTGGGGTCACTGCTACCTTCTGCAGC
TAGCCAACTACCTTGCCTACCACTCCGACATCATGGAAGACGTGAGCGGTGA
CGGCCTACTGGAGTGTCACTGTCGCTGCAACCTATGCACCCCGCACCGCTCCC
TGGTCTGCAATTCGCAACTGCTTAGCGAAAGTCAAATTATCGGTACCTTTGAG
CTGCAGGGTCCCTCGCCTGACGAAAAGTCCGCGGCTCCGGGGTTGAAACTCA
CTCCGGGGCTGTGGACGTCGGCTTACCTTCGCAAATTTGTACCTGAGGACTAC
CACGCCCACGAGATTAGGTTCTACGAAGACCAATCCCGCCCGCCAAATGCGG
AGCTTACCGCCTGCGTCATTACCCAGGGCCACATCCTTGGCCAATTGCAAGCC
ATCAACAAAGCCCGCCAAGAGTTTCTGCTACGAAAGGGACGGGGGGTTTACC
TGGACCCCCAGTCCGGCGAGGAGCTCAACCCAATCCCCCCGCCGCCGCAGCC
CTATCAGCAGCCGCGGGCCCTTGCTTCCCAGGATGGCACCCAAAAAGAAGCT
GCAGCTGCCGCCGCCGCCACCCACGGACGAGGAGGAATACTGGGACAGTCA
GGCAGAGGAGGTTTTGGACGAGGAGGAGGAGATGATGGAAGACTGGGACAG
CCTAGACGAAGCTTCCGAGGCCGAAGAGGTGTCAGACGAAACACCGTCACCC
TCGGTCGCATTCCCCTCGCCGGCGCCCCAGAAATTGGCAACCGTTCCCAGCAT
CGCTACAACCTCCGCTCCTCAGGCGCCGCCGGCACTGCCTGTTCGCCGACCCA
ACCGTAGATGGGACACCACTGGAACCAGGGCCGGTAAGTCTAAGCAGCCGC
CGCCGTTAGCCCAAGAGCAACAACAGCGCCAAGGCTACCGCTCGTGGCGCGG
GCACAAGAACGCCATAGTTGCTTGCTTGCAAGACTGTGGGGGCAACATCTCC
TTCGCCCGCCGCTTTCTTCTCTACCATCACGGCGTGGCCTTCCCCCGTAACAT
CCTGCATTACTACCGTCATCTCTACAGCCCCTACTGCACCGGCGGCAGCGGCA
GCGGCAGCAACAGCAGCGGTCACACAGAAGCAAAGGCGACCGGATAGCAAG
ACTCTGACAAAGCCCAAGAAATCCACAGCGGCGGCAGCAGCAGGAGGAGGA
GCGCTGCGTCTGGCGCCCAACGAACCCGTATCGACCCGCGAGCTTAGAAATA
GGATTTTTCCCACTCTGTATGCTATATTTCAACAAAGCAGGGGCCAAGAACA
AGAGCTGAAAATAAAAAACAGGTCTCTGCGCTCCCTCACCCGCAGCTGCCTG
TATCACAAAAGCGAAGATCAGCTTCGGCGCACGCTGGAAGACGCGGAGGCT
CTCTTCAGCAAATACTGCGCGCTGACTCTTAAGGACTAGTTTCGCGCCCTTTC
TCAAATTTAAGCGCGAAAACTACGTCATCTCCAGCGGCCACACCCGGCGCCA
GCACCTGTCGTCAGCGCCATTATGAGCAAGGAAATTCCCACGCCCTACATGT
GGAGTTACCAGCCACAAATGGGACTTGCGGCTGGAGCTGCCCAAGACTACTC
AACCCGAATAAACTACATGAGCGCGGGACCCCACATGATATCCCGGGTCAAC
GGAATCCGCGCCCACCGAAACCGAATTCTCCTCGAACAGGCGGCTATTACCA
CCACACCTCGTAATAACCTTAATCCCCGTAGTTGGCCCGCTGCCCTGGTGTAC
CAGGAAAGTCCCGCTCCCACCACTGTGGTACTTCCCAGAGACGCCCAGGCCG
AAGTTCAGATGACTAACTCAGGGGCGCAGCTTGCGGGCGGCTTTCGTCACAG
GGTGCGGTCGCCCGGGCGTTTTAGGGCGGAGTAACTTGCATGTATTGGGAAT
TGTAGTTTTTTTAAAATGGGAAGTGACGTATCGTGGGAAAACGGAAGTGAAG
ATTTGAGGAAGTTGTGGGTTTTTTGGCTTTCGTTTCTGGGCGTAGGTTCGCGT
GCGGTTTTCTGGGTGTTTTTTGTGGACTTTAACCGTTACGTCATTTTTTAGTCC

TATATATACTCGCTCTGTACTTGGCCCTTTTTACACTGTGACTGATTGAGCTGG
TGCCGTGTCGAGTGGTGTTTTTTAATAGGTTTTTTTACTGGTAAGGCTGACTGT
TATGGCTGCCGCTGTGGAAGCGCTGTATGTTGTTCTGGAGCGGGAGGGTGCT
ATTTTGCCTAGGCAGGAGGGTTTTTCAGGTGTTTATGTGTTTTTCTCTCCTATT
AATTTTGTTATACCTCCTATGGGGGCTGTAATGTTGTCTCTACGCCTGCGGGT
ATGTATTCCCCCGGGCTATTTCGGTCGCTTTTTAGCACTGACCGATGTTAACC
AACCTGATGTGTTTACCGAGTCTTACATTATGACTCCGGACATGACCGAGGA
ACTGTCGGTGGTGCTTTTTAATCACGGTGACCAGTTTTTTTACGGTCACGCCG
GCATGGCCGTAGTCCGTCTTATGCTTATAAGGGTTGTTTTTCCTGTTGTAAGA
CAGGCTTCTAATGTTTAAATGTTTTTTTTTTTGTTATTTTATTTTGTGTTTAATG
CAGGAACCCGCAGACATGTTTGAGAGAAAAATGGTGTCTTTTTCTGTGGTGG
TTCCGGAACTTACCTGCCTTTATCTGCATGAGCATGACTACGATGTGCTTGCT
TTTTTGCGCGAGGCTTTGCCTGATTTTTTGAGCAGCACCTTGCATTTTATATCG
CCGCCCATGCAACAAGCTTACATAGGGGCTACGCTGGTTAGCATAGCTCCGA
GTATGCGTGTCATAATCAGTGTGGGTTCTTTTGTCATGGTTCCTGGCGGGGAA
GTGGCCGCGCTGGTCCGTGCAGACCTGCACGATTATGTTCAGCTGGCCCTGC
GAAGGGACCTACGGGATCGCGGTATTTTTGTTAATGTTCCGCTTTTGAATCTT
ATACAGGTCTGTGAGGAACCTGAATTTTTGCAATCATGATTCGCTGCTTGAGG
CTGAAGGTGGAGGGCGCTCTGGAGCAGATTTTTACAATGGCCGGACTTAATA
TTCGGGATTTGCTTAGAGACATATTGATAAGGTGGCGAGATGAAAATTATTT
GGGCATGGTTGAAGGTGCTGGAATGTTTATAGAGGAGATTCACCCTGAAGGG
TTTAGCCTTTACGTCCACTTGGACGTGAGGGCAGTTTGCCTTTTGGAAGCCAT
TGTGCAACATCTTACAAATGCCATTATCTGTTCTTTGGCTGTAGAGTTTGACC
ACGCCACCGGAGGGGAGCGCGTTCACTTAATAGATCTTCATTTTGAGGTTTTG
GATAATCTTTTGGAATAAAAAAAAAAAAACATGGTTCTTCCAGCTCTTCCCGC
TCCTCCCGTGTGTGACTCGCAGAACGAATGTGTAGGTTGGCTGGGTGTGGCTT
ATTCTGCGGTGGTGGATGTTATCAGGGCAGCGGCGCATGAAGGAGTTTACAT
AGAACCCGAAGCCAGGGGGCGCCTGGATGCTTTGAGAGAGTGGATATACTAC
AACTACTACACAGAGCGAGCTAAGCGACGAGACCGGAGACGCAGATCTGTTT
GTCACGCCCGCACCTGGTTTTGCTTCAGGAAATATGACTACGTCCGGCGTTCC
ATTTGGCATGACACTACGACCAACACGATCTCGGTTGTCTCGGCGCACTCCGT
ACAGTAGGGATCGCCTACCTCCTTTTGAGACAGAGACCCGCGCTACCATACT
GGAGGATCATCCGCTGCTGCCCGAATGTAACACTTTGACAATGCACAACGTG
AGTTACGTGCGAGGTCTTCCCTGCAGTGTGGGATTTACGCTGATTCAGGAATG
GGTTGTTCCCTGGGATATGGTTCTGACGCGGGAGGAGCTTGTAATCCTGAGG
AAGTGTATGCACGTGTGCCTGTGTTGTGCCAACATTGATATCATGACGAGCAT
GATGATCCATGGTTACGAGTCCTGGGCTCTCCACTGTCATTGTTCCAGTCCCG
GTTCCCTGCAGTGCATAGCCGGCGGGCAGGTTTTGGCCAGCTGGTTTAGGAT
GGTGGTGGATGGCGCCATGTTTAATCAGAGGTTTATATGGTACCGGGAGGTG
GTGAATTACAACATGCCAAAAGAGGTAATGTTTATGTCCAGCGTGTTTATGA
GGGGTCGCCACTTAATCTACCTGCGCTTGTGGTATGATGGCCACGTGGGTTCT
GTGGTCCCCGCCATGAGCTTTGGATACAGCGCCTTGCACTGTGGGATTTTGAA
CAATATTGTGGTGCTGTGCTGCAGTTACTGTGCTGATTTAAGTGAGATCAGGG
TGCGCTGCTGTGCCCGGAGGACAAGGCGTCTCATGCTGCGGGCGGTGCGAAT
CATCGCTGAGGAGACCACTGCCATGTTGTATTCCTGCAGGACGGAGCGGCGG
CGGCAGCAGTTTATTCGCGCGCTGCTGCAGCACCACCGCCCTATCCTGATGCA

CGATTATGACTCTACCCCCATGTAGGCGTGGACTTCCCCTTCGCCGCCCGTTG
AGCAACCGCAAGTTGGACAGCAGCCTGTGGCTCAGCAGCTGGACAGCGACAT
GAACTTAAGCGAGCTGCCCGGGGAGTTTATTAATATCACTGATGAGCGTTTG
GCTCGACAGGAAACCGTGTGGAATATAACACCTAAGAATATGTCTGTTACCC
ATGATATGATGCTTTTTAAGGCCAGCCGGGGAGAAAGGACTGTGTACTCTGT
GTGTTGGGAGGGAGGTGGCAGGTTGAATACTAGGGTTCTGTGAGTTTGATTA
AGGTACGGTGATCAATATAAGCTATGTGGTGGTGGGGCTATACTACTGAATG
AAAAATGACTTGAAATTTTCTGCAATTGAAAAATAAACACGTTGAAACATAA
CATGCAACAGGTTCACGATTCTTTATTCCTGGGCAATGTAGGAGAAGGTGTA
AGAGTTGGTAGCAAAAGTTTCAGTGGTGTATTTTCCACTTTCCCAGGACCATG
TAAAAGACATAGAGTAAGTGCTTACCTCGCTAGTTTCTGTGGATTCACTAGAA
TCGATGTAGGATGTTGCCCCTCCTGACGCGGTAGGAGAAGGGGAGGGTGCCC
TGCATGTCTGCCGCTGCTCTTGCTCTTGCCGCTGCTGAGGAGGGGGGCGCATC
TGCCGCAGCACCGGATGCATCTGGGAAAAGCAAAAAAGGGGCTCGTCCCTGT
TTCCGGAGGAATTTGCAAGCGGGGTCTTGCATGACGGGGAGGCAAACCCCCG
TTCGCCGCAGTCCGGCCGGCCCGAGACTCGAACCGGGGGTCCTGCGACTCAA
CCCTTGGAAAATAACCCTCCGGCTACAGGGAGCGAGCCACTTAATGCTTTCG
CTTTCCAGCCTAACCGCTTACGCCGCGCGCGGCCAGTGGCCAAAAAAGCTAG
CGCAGCAGCCGCCGCGCCTGGAAGGAAGCCAAAAGGAGCGCTCCCCCGTTGT
CTGACGTCGCACACCTGGGTTCGACACGCGGGCGGTAACCGCATGGATCACG
GCGGACGGCCGGATCCGGGGTTCGAACCCCGGTCGTCCGCCATGATACCCTT
GCGAATTTATCCACCAGACCACGGAAGAGTGCCCGCTTACAGGCTCTCCTTTT
GCACGGTCTAGAGCGTCAACGACTGCGCACGCCTCACCGGCCAGAGCGTCCC
GACCATGGAGCACTTTTTGCCGCTGCGCAACATCTGGAACCGCGTCCGCGAC
TTTCCGCGCGCCTCCACCACCGCCGCCGGCATCACCTGGATGTCCAGGTACAT
CTACGGATTACGTCGACGTTTAAACCATATGATCAGCTCACTCAAAGGCGGT
AATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGC
AAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT
TTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTG
TCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAG
GTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAAC
CCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCC
AACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA
TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCC
TAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAG
CCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA
CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT
GGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT
CTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTA
TATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCT
ATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGA
TACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCC

AGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATC
CAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCG
TTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTG
TCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCAT
AATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTA
CTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCT
CATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGT
TGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCT
TTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCG
CAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCT
TTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCC
CGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTA
AATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAAT
CCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTT
TGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA
AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAA
GTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAG
CCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA
AGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGT
CACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGC
GCGATGGATCC

Chapter 3 Plasmids

pIDTSmart-RC2-ΔVP1-CMV-VP1

Red=*Rep*
Light Blue=*Cap*-ΔVP1
Dark Blue=*Cap*-VP1 only
Orange=CMV
Green=AAP

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGTGGACCGG
AAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCTGAGGCTCGTCCTGAA
TGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATCCAGTCGCGCTGCTC
TCGTCGATCCGCTAGGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGAAGATCAGAAGTT
CCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTGATCTGCGCAGCCGCCA
TGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCG
GCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGAT
TCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAG
CGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCCTTTTCTTTGTG

CAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACGTGCTCGTGGAAACCACCGGGGTG
AAATCCATGGTTTTGGGACGTTTCCTGAGTCAGATTCGCGAAAAACTGATTCAGAGAAT
TTACCGCGGGATCGAGCCGACTTTGCCAAACTGGTTCGCGGTCACAAAGACCAGAAATGG
CGCCGGAGGCGGGAACAAGGTGGTGGATGAGTGCTACATCCCCAATTACTTGCTCCCCAA
AACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGGAACAGTATTTAAGCGCCTGTTT
GAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGTCGCAGACGCA
GGAGCAGAACAAAGAGAATCAGAATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAA
CTTCAGCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGA
AGCAGTGGATCCAGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGC
GGTCCCAAATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACC
GCCCCCGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTAT
AAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGG
GCCACGAAAAAGTTCGGCAAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGG
AAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCTTCTACGGGTGCGTAAACTGG
ACCAATGAGAACTTTCCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGTGGGAGGA
GGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGG
TGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTGATCGTCA
CCTCCAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAACACCAGC
AGCCGTTGCAAGACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTTG
GGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGTT
GAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCCCCCAG
TGACGCAGATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTC
AGACGCGGAAGCTTCGATCAACTACGCAGACAGGTACCAAAACAAATGTTCTCGTCACGT
GGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGCGAGAGAATGAATCAGAATTCAA
ATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCCCGTGTCAGAATCTC
AACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCATGG
GAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCA
TCTTTGAACAATAAATGATTTAAATCAGGTCTCGCTGCCGATGGTTATCTTCCAGATTGG
CTCGAGGACACTCTCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCCACCA
CCACCAAAGCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTAC
AAGTACCTCGGACCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCC
GCGGCCCTCGAGCACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCCGTAC
CTCAAGTACAACCACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAAGATACGTCTTTT
GGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGAGGGTTCTTGAACCTCTGGGC
CTGGTTGAGGAACCTGTTAAGACGGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCT
GTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGA
TTGAATTTTGGTCAGACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAG
CCACCAGCAGCCCCCTCTGGTCTGGGAACTAATACGATGGCTACAGGCAGTGGCGCACCA
ATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAATTGGCATTG
CGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCC
CACCTACAACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAA
TCACTACTTTGGCTACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCA
CTTTTCACCACGTGACTGGCAAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGAG
ACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGAC
GACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCT
CCCGTACGTCCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAGCAGACGTCTTC

ATGGTGCCACAGTATGGATACCTCACCCTGAACAACGGGAGTCAGGCAGTAGGACGCTCT
TCATTTTACTGCCTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAACTTTACC
TTCAGCTACACTTTTGAGGACGTTCCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTG
GACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACT
CCAAGTGGAACCACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATT
CGGGACCAGTCTAGGAACTGGCTTCCTGGACCCTGTTACCGCCAGCAGCGAGTATCAAAG
ACATCTGCGGATAACAACAACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCTC
AATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCACAAGGACGATGAA
GAAAAGTTTTTTCCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAAC
AAATGTGGACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATC
CCGTGGCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAG
CAGCTACCGCAGATGTCAACACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGAG
ATGTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACATTTTCACC
CCTCTCCCCTCATGGGTGGATTCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAGA
ACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCA
TCACACAGTACTCCACGGGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAA
AACAGCAAACGCTGGAATCCCGAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAAT
GTGGACTTTACTGTGGACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAGA
TACCTGACTCGTAATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTT
GAACTTTGGTCTCTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATAAGTAG
CATGGCGGGTTAATCATTAACTACAGCCCGGGCGTTTAAACAGCGGGCGGAGGGGTGGA
GTCGTGACGTGAATTACGTCATAGGGTTAGGGAGGTCCTGTATTAGAGGTCACGTGAGT
GTTTTGCGACATTTTGCGACACCATGTGGTCTCGCTGGGGGGGGGGGCCCGAGTGAGCAC
GCAGGGTCTCCATTTTGAAGCGGGAGGTTTGAACGAGCGCTGGCGCGCTCACTGGCCGTC
GTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA
CATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCATG
CATCGGCCGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGAT
ATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACG
CCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTG
GCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAA
TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGG
CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGG
AGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCA
TTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTG
GCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGA
GACCCAAGCTGGCTAGCATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACACTC
TCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCCACCACCACCAAAGCCCG
CAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTACAAGTACCTCGGAC
CCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGC
ACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCCGTACCTCAAGTACAACC
ACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAAGATACGTCTTTTGGGGGCAACCTCG
GACGAGCAGTCTTCCAGGCGAAAAAGAGGGTTCTTGAACCTCTGGGCCTGGTTGAGGAAC
CTGTTAAGACCGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCTGTGGAGCCAGACT

CCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTC
AGACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCC
CCTCTGGTCTCGGAACTAATACCCTCGCTACAGGCAGTGGCGCACCACTCGCAGACAATA
ACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAATTGGCATTGCGATTCCACATGGC
TCGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCCACCTACAACAACC
ACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCT
ACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACCACGTG
ACTGGCAAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGAGACTCAACTTCAAGC
TCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGACGACGATTGCCAAT
AACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACGTCCTC
GGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAGCAGACGTCTTCATGGTGCCACAG
TATGGATACCTCACCCTGAACAACGGGAGTCAGGCAGTAGGACGCTCTTCATTTTACTGC
CTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAACTTTACCTTCAGCTACACT
TTTGAGGACGTTCCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATG
AATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACTCCAAGTGGAACC
ACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCGGGACCAGTCT
AGGAACTGGCTTCCTGGACCCTGTTACCGCCAGCAGCGAGTATCAAAGACATCTGCGGAT
AACAACAACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGAC
TCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCACAAGGACGATGAAGAAAAGTTTTTT
CCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACAT
TGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGG
AGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACCGCAG
ATGTCAACACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTC
AGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACATTTTCACCCCTCTCCCCTCA
TGGGTGGATTCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAGAACACCCCGGTAC
CTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCATCACACAGTACT
CCACGGGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAACGC
TGGAATCCCGAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAATGTGGACTTTACT
GTGGACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAGATACCTGACTCGT
AATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTTGAACTTTGGTCT
CTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATAAGTAGCATGGCGGGTTA
ATCATTAACTACAGCCCTAGGGGTGCGAGCGGATCGAGCAGTGTCGATCACTACTGGACC
GCGAGCTGTGCTGCGACCCGTGATCTTACGGCATTATACGTATGATCGGTCCACGATCAG
CTAGATTATCTAGTCAGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTT
AAATCATACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTT
GATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTA
TTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGAGCCATATTCAACG
GGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGT
ATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGG
AAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTT
ACAGATGAGATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAG
CATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACA
GCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGC
AGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCG
CGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGA
TTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAAC

TCTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTA
TTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGAC
CGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAG
AAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCA
CTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCTCACCTTCG
GGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
CAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGA
CTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG
AAAAAAAGGATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pIDTSmart-RC2-ΔVP2-CMV-VP2

Red=*Rep*
Light Blue=*Cap*-ΔVP1
Dark Blue=*Cap*-VP1 only
Orange=CMV
Green=AAP

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGTGGACCGG
AAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCTGAGGCTCGTCCTGAA
TGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATCCAGTCGCGCTGCTC
TCGTCGATCCGCTAGGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGAAGATCAGAAGTT
CCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTGATCTGCGCAGCCGCCA
TGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCG
GCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGAT
TCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAG
CGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCCTTTTCTTTGTG
CAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACGTGCTCGTGGAAACCACCGGGGTG
AAATCCATGGTTTTGGGACGTTTCCTGAGTCAGATTCGCGAAAAACTGATTCAGAGAAT
TTACCGCGGGATCGAGCCGACTTTGCCAAACTGGTTCGCGGTCACAAAGACCAGAAATGG
CGCCGGAGGCGGGAACAAGGTGGTGGATGAGTGCTACATCCCCAATTACTTGCTCCCCAA
AACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGGAACAGTATTTAAGCGCCTGTTT
GAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGTCGCAGACGCA
GGAGCAGAACAAAGAGAATCAGAATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAA
CTTCAGCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGA
AGCAGTGGATCCAGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGC
GGTCCCAAATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACC
GCCCCCGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTAT

AAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGG
GCCACGAAAAAGTTCGGCAAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGG
AAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCTTCTACGGGTGCGTAAACTGG
ACCAATGAGAACTTTCCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGTGGGAGGA
GGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGG
TGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTGATCGTCA
CCTCCAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAACACCAGC
AGCCGTTGCAAGACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTTG
GGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGTT
GAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCCCCCAG
TGACGCAGATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTC
AGACGCGGAAGCTTCGATCAACTACGCAGACAGGTACCAAAACAAATGTTCTCGTCACGT
GGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGCGAGAGAATGAATCAGAATTCAA
ATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCCCGTGTCAGAATCTC
AACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCATGG
GAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCA
TCTTTGAACAATAAATGATTTAAATCAGGTATGGCTGCCGATGGTTATCTTCCAGATTG
GCTCGAGGACACTCTCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCCACC
ACCACCAAAGCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTA
CAAGTACCTCGGACCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGC
CGCGGCCCTCGAGCACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCCGTA
CCTCAAGTACAACCACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAAGATACGTCTTT
TGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGAGGGTTCTTGAACCTCTGGG
CCTGGTTGAGGAACCTGTTAAGACCGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCC
TGTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAG
ATTGAATTTTGGTCAGACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACA
GCCACCAGCAGCCCCCTCTGGTCTGGGAACTAATACGATGGCTACAGGCAGTGGCGCACC
AATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAATTGGCATT
GCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGC
CCACCTACAACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACA
ATCACTACTTTGGCTACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCC
ACTTTTCACCACGTGACTGGCAAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGA
GACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGA
CGACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGC
TCCCGTACGTCCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAGCAGACGTCTT
CATGGTGCCACAGTATGGATACCTCACCCTGAACAACGGGAGTCAGGCAGTAGGACGCTC
TTCATTTTACTGCCTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAACTTTAC
CTTCAGCTACACTTTTGAGGACGTTCCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCT
GGACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACAC
TCCAAGTGGAACCACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACAT
TCGGGACCAGTCTAGGAACTGGCTTCCTGGACCCTGTTACCGCCAGCAGCGAGTATCAAA
GACATCTGCGGATAACAACAACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCT
CAATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCACAAGGACGATGA
AGAAAAGTTTTTTCCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAA
CAAATGTGGACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAAT
CCCGTGGCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAA

GCAGCTACCGCAGATGTCAACACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGA
GATGTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACATTTTCAC
CCCTCTCCCCTCATGGGTGGATTCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAG
AACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTC
ATCACACAGTACTCCACGGGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGA
AAACAGCAAACGCTGGAATCCCGAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAA
TGTGGACTTTACTGTGGACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAG
ATACCTGACTCGTAATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGT
TGAACTTTGGTCTCTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATAAGTA
GCATGGCGGGTTAATCATTAACTACAGCCCGGGCGTTTAAACAGCGGGCGGAGGGGTGG
AGTCGTGACGTGAATTACGTCATAGGGTTAGGGAGGTCCTGTATTAGAGGTCACGTGAG
TGTTTTGCGACATTTTGCGACACCATGTGGTCTCGCTGGGGGGGGGGGCCCGAGTGAGCA
CGCAGGGTCTCCATTTTGAAGCGGGAGGTTTGAACGAGCGCTGGCGCGCTCACTGGCCGT
CGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGC
ACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCA
TGCATCGGCCGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCAG
ATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCAT
TAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAA
CGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTA
AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGT
ACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTC
TGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGG
GAGACCCAAGCTGGCTAGCATGGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCTGT
GGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGATT
GAATTTTGGTCAGACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCC
ACCAGCAGCCCCCTCTGGTCTCGGAACTAATACCCTCGCTACAGGCAGTGGCGCACCACT
CGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAATTGGCATTGCG
ATTCCACATGGCTCGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCCA
CCTACAACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATC
ACTACTTTGGCTACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACT
TTTCACCACGTGACTGGCAAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGAGAC
TCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGACGA
CGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCC
CGTACGTCCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAGCAGACGTCTTCAT
GGTGCCACAGTATGGATACCTCACCCTGAACAACGGGAGTCAGGCAGTAGGACGCTCTTC
ATTTTACTGCCTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAACTTTACCTT
CAGCTACACTTTTGAGGACGTTCCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGA
CCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACTCC
AAGTGGAACCACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCG
GGACCAGTCTAGGAACTGGCTTCCTGGACCCTGTTACCGCCAGCAGCGAGTATCAAAGAC
ATCTGCGGATAACAACAACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCTCAA

TGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCACAAGGACGATGAAGA
AAAGTTTTTTCCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAA
ATGTGGACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCC
GTGGCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCA
GCTACCGCAGATGTCAACACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGAGAT
GTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACATTTTCACCCC
TCTCCCCTCATGGGTGGATTCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAGAAC
ACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCATC
ACACAGTACTCCACGGGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAA
CAGCAAACGCTGGAATCCCGAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAATGT
GGACTTTACTGTGGACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAGATA
CCTGACTCGTAATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTTGA
ACTTTGGTCTCTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATAAGTAGCA
TGGCGGGTTAATCATTAACTACAGCCCTAGGGGTGCGAGCGGATCGAGCAGTGTCGATCA
CTACTGGACCGCGAGCTGTGCTGCGACCCGTGATCTTACGGCATTATACGTATGATCGGT
CCACGATCAGCTAGATTATCTAGTCAGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATA
CTGACCATTTAAATCATACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGG
CTTTTGACTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACT
ACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGAGCC
ATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAACATGGATGCTGATT
TATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGA
TTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCC
AATGATGTTACAGATGAGATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCC
GACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCC
AGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTG
ATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTA
ACGGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTG
GTGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAA
ATGCATAAACTCTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTT
GATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGG
AATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCC
TTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAT
TGCAGTTTCACTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGG
CTCACCTTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGA
CGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT
TACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGC
TGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCC
CCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTA
AGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT
GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCT
TGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATT
ACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pIDTSmart-RC2-ΔVP1,2-CMV-VP1,2
Red=*Rep*
Light Blue=*Cap*-ΔVP1
Dark Blue=*Cap*-VP1 only
Orange=CMV
Green=AAP

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGTGGACCGG
AAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCTGAGGCTCGTCCTGAA
TGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATCCAGTCGCGCTGCTC
TCGTCGATCCGCTAGGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGAAGATCAGAAGTT
CCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTGATCTGCGCAGCCGCCA
TGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCG
GCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGAT
TCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAG
CGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCCTTTTCTTTGTG
CAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACGTGCTCGTGGAAACCACCGGGGTG
AAATCCATGGTTTTGGGACGTTTCCTGAGTCAGATTCGCGAAAAACTGATTCAGAGAAT
TTACCGCGGGATCGAGCCGACTTTGCCAAACTGGTTCGCGGTCACAAAGACCAGAAATGG
CGCCGGAGGCGGGAACAAGGTGGTGGATGAGTGCTACATCCCCAATTACTTGCTCCCCAA
AACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGGAACAGTATTTAAGCGCCTGTTT
GAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGTCGCAGACGCA
GGAGCAGAACAAAGAGAATCAGAATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAA
CTTCAGCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGA
AGCAGTGGATCCAGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGC
GGTCCCAAATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACC
GCCCCCGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTAT
AAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGG
GCCACGAAAAAGTTCGGCAAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGG
AAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCTTCTACGGGTGCGTAAACTGG
ACCAATGAGAACTTTCCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGTGGGAGGA
GGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGG
TGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTGATCGTCA
CCTCCAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAACACCAGC
AGCCGTTGCAAGACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTTG
GGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGTT
GAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCCCCCAG
TGACGCAGATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTC
AGACGCGGAAGCTTCGATCAACTACGCAGACAGGTACCAAAACAAATGTTCTCGTCACGT
GGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGCGAGAGAATGAATCAGAATTCAA
ATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCCCGTGTCAGAATCTC
AACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCATGG
GAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCA
TCTTTGAACAATAAATGATTTAAATCAGGTCTCGCTGCCGATGGTTATCTTCCAGATTGG
CTCGAGGACACTCTCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCCACCA
CCACCAAAGCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTAC

AAGTACCTCGGACCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCC
GCGGCCCTCGAGCACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCCGTAC
CTCAAGTACAACCACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAAGATACGTCTTTT
GGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGAGGGTTCTTGAACCTCTGGGC
CTGGTTGAGGAACCTGTTAAGACCGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCT
GTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGA
TTGAATTTTGGTCAGACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAG
CCACCAGCAGCCCCCTCTGGTCTGGGAACTAATACGATGGCTACAGGCAGTGGCGCACCA
ATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAATTGGCATTG
CGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCC
CACCTACAACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAA
TCACTACTTTGGCTACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCA
CTTTTCACCACGTGACTGGCAAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGAG
ACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGAC
GACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCT
CCCGTACGTCCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAGCAGACGTCTTC
ATGGTGCCACAGTATGGATACCTCACCCTGAACAACGGGAGTCAGGCAGTAGGACGCTCT
TCATTTTACTGCCTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAACTTTACC
TTCAGCTACACTTTTGAGGACGTTCCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTG
GACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACT
CCAAGTGGAACCACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATT
CGGGACCAGTCTAGGAACTGGCTTCCTGGACCCTGTTACCGCCAGCAGCGAGTATCAAAG
ACATCTGCGGATAACAACAACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCTC
AATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCACAAGGACGATGAA
GAAAAGTTTTTTCCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAAC
AAATGTGGACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATC
CCGTGGCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAG
CAGCTACCGCAGATGTCAACACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGAG
ATGTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACATTTTCACC
CCTCTCCCCTCATGGGTGGATTCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAGA
ACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCA
TCACACAGTACTCCACGGGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAA
AACAGCAAACGCTGGAATCCCGAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAAT
GTGGACTTTACTGTGGACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAGA
TACCTGACTCGTAATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTT
GAACTTTGGTCTCTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATAAGTAG
CATGGCGGGTTAATCATTAACTACAGCCCGGGCGTTTAAACAGCGGGCGGAGGGGTGGA
GTCGTGACGTGAATTACGTCATAGGGTTAGGGAGGTCCTGTATTAGAGGTCACGTGAGT
GTTTTGCGACATTTTGCGACACCATGTGGTCTCGCTGGGGGGGGGGCCCGAGTGAGCAC
GCAGGGTCTCCATTTTGAAGCGGGAGGTTTGAACGAGCGCTGGCGCGCTCACTGGCCGTC
GTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA
CATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCATG
CATCGGCCGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGAT
ATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACG

CCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTG
GCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAA
TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGG
CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGG
AGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCA
TTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTG
GCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGA
GACCCAAGCTGGCTAGCATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACACTC
TCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCCACCACCACCAAAGCCCG
CAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTACAAGTACCTCGGAC
CCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGC
ACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCCGTACCTCAAGTACAACC
ACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAAGATACGTCTTTTGGGGGCAACCTCG
GACGAGCAGTCTTCCAGGCGAAAAAGAGGGTTCTTGAACCTCTGGGCCTGGTTGAGGAAC
CTGTTAAGATGGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCTGTGGAGCCAGACT
CCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTC
AGACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCC
CCTCTGGTCTCGGAACTAATACCCTCGCTACAGGCAGTGGCGCACCACTCGCAGACAATA
ACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAATTGGCATTGCGATTCCACATGGC
TCGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCCACCTACAACAACC
ACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCT
ACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACCACGTG
ACTGGCAAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGAGACTCAACTTCAAGC
TCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGACGACGATTGCCAAT
AACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACGTCCTC
GGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAGCAGACGTCTTCATGGTGCCACAG
TATGGATACCTCACCCTGAACAACGGGAGTCAGGCAGTAGGACGCTCTTCATTTTACTGC
CTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAACTTTACCTTCAGCTACACT
TTTGAGGACGTTCCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATG
AATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACTCCAAGTGGAACC
ACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCGGGACCAGTCT
AGGAACTGGCTTCCTGGACCCTGTTACCGCCAGCAGCGAGTATCAAAGACATCTGCGGAT
AACAACAACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGAC
TCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCACAAGGACGATGAAGAAAAGTTTTTT
CCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACAT
TGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGG
AGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACCGCAG
ATGTCAACACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTC
AGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACATTTTCACCCCTCTCCCCTCA
TGGGTGGATTCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAGAACACCCCGGTAC
CTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCATCACACAGTACT
CCACGGGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAACGC
TGGAATCCCGAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAATGTGGACTTTACT
GTGGACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAGATACCTGACTCGT
AATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTTGAACTTTGGTCT

CTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATAAGTAGCATGGCGGGTTA
ATCATTAACTACAGCCCTAGGGGTGCGAGCGGATCGAGCAGTGTCGATCACTACTGGACC
GCGAGCTGTGCTGCGACCCGTGATCTTACGGCATTATACGTATGATCGGTCCACGATCAG
CTAGATTATCTAGTCAGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTT
AAATCATACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTT
GATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTA
TTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGAGCCATATTCAACG
GGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGT
ATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGG
AAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTT
ACAGATGAGATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAG
CATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACA
GCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGC
AGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCG
CGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGA
TTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAAC
TCTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTA
TTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGAC
CGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAG
AAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCA
CTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCTCACCTTCG
GGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
CAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGA
CTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG
AAAAAAAGGATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

Chapter 4 Plasmids

pAAV-ITR-GFP

Green= GFP
Pink=ITRs

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCC
CGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC
AGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTGCGGCCGCACGCGTG
GAGCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATAT
ATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC

CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC
GTCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGA
CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAT
GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG
GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA
GTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATA
GAAGACACCGGGACCGATCCAGCCTCCGCGGATTCGAATCCCGGCCGGGAAC
GGTGCATTGGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTACCGCCTA
TAGAGTCTATAGGCCCACAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTT
TATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATAC
AATGTATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGG
GTTAAGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTA
ACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATT
CTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAG
GCCCTTTTGCTAATCATGTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGC
AACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTGGGATTCGA
ACATCGATTGAATTCTGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG
TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAA
GTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA
CCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCA
GCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCG
AGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG
AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACA
ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAA
GATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACC
TGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAG
CTGTACAAGTACTCAGATCTCGAGCTCAAGTAGGGATCCTCTAGAGTCGACC
TGCAGAAGCTTGCCTCGAGCAGCGCTGCTCGAGAGATCTACGGGTGGCATCC
CTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGC
CCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTG
TCCTTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAG
TTGGGAAGACAACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAG
TGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGA
TTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATGCATGACCAGG
CTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCAG
GCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAAT
TGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGT
AGGTAACCACGTGCGGACCGAGCGGCCGCAGGAACCCCTAGTGATGGAGTT

GGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAG
GTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGC
GCAGCTGCCTGCAGGGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC
GGTATTTCACACCGCATACGTCAAAGCAACCATAGTACGCGCCCTGTAGCGG
CGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACG
TTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCG
ATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGATGGTT
CACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAG
TCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCC
TATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTG
GTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATA
TTAACGTTTACAATTTTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGC
ATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGG
GCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAG
CTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAG
GGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTC
TTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTT
TATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGA
TAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCC
GTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCC
AGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGT
GGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC
GGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACAC
TATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTAC
GGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGAT
AACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAA
CCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAA
CCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCT
GTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTC
TAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGG
ACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTG
GAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGG
TAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATG
GATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATT
GGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTT
CATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGAC
CAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAA
AGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGC
AAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCT
ACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAAT
ACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGA

GCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAG
CGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAG
GGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA
TCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTG
ATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTT
TTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT

pAAV-ITR-mCherry

Pink=ITRs
Red=mCherry

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCC
CGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC
AGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTGCGGCCGCACGCGTG
GAGCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATAT
ATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC
CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC
GTCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGA
CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAT
GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG
GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA
GTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATA
GAAGACACCGGGACCGATCCAGCCTCCGCGGATTCGAATCCCGGCCGGGAAC
GGTGCATTGGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTACCGCCTA
TAGAGTCTATAGGCCCACAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTT
TATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATAC
AATGTATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGG
GTTAAGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTA
ACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATT
CTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAG
GCCCTTTTGCTAATCATGTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGC
AACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTGGGATTCGA
ACATCGATTGAATTCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCA
TCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCA
CGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCA
GACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGAC
ATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGC
CGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAG
CGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCT
CCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTT
CCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC

TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAG
AGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACC
TACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCA
AGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGA
ACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTA
ATCTAGAGTCGACCTGCAGAAGCTTGCCTCGAGCAGCGCTGCTCGAGAGATC
TACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGT
TGCCACTCCAGTGCCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTT
TGTCTGACTAGGTGTCCTTCTATAATATTATGGGGTGGAGGGGGGTGGTATGG
AGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGCCTGCGGGGTCTATTGG
GAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCCT
CCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAG
GCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTT
CACCATATTGGCCAGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCAC
CTTGGCCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCT
GTCCTTCTGATTTTGTAGGTAACCACGTGCGGACCGAGCGGCCGCAGGAACC
CCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGG
CCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGT
GAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCCTGATGCGGTATTTTCTC
CTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGTAC
GCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTT
CCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTC
CCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGA
TTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC
CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA
ACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCC
GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCG
AATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTACAAT
CTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCT
GACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTG
TGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAA
CGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCAT
GATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGC
GGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATG
AGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGA
GTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTC
CTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATC
CTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGT
TCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTC
GGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCAC
AGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCC
ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAG
GACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCG
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGT

GACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTG
GCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGC
GGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTA
TTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGC
ACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG
AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCT
CACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAG
ATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTT
TGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT
CAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTT
GCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGA
GCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT
CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG
TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACG
ATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACA
CAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATC
CGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGG
GAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT

Chapter 5 Plasmids

pIDTSmart-MbPylRS-RC2

Red=*Rep*
Blue = *Cap*
Green=AAP
Orange = MbPylRS

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGTGGAC
CGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCTGAGGCTCGT
CCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATCCAGTC
GCGCTGCTCTCGTCGATCCGCTAGGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGA
AGATCAGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTGAT
CTGCGCAGCCGCCATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCT
TGACGAGCATCTGCCCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGA
ATGGGAGTTGCCGCCAGATTCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCT
GACCGTGGCCGAGAAGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTA
AGGCCCCGGAGGCCCTTTTCTTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCCACA
TGCACGTGCTCGTGGAAACCACCGGGGTGAAATCCATGGTTTTGGGACGTTTCCTGA
GTCAGATTCGCGAAAAACTGATTCAGAGAATTTACCGCGGGATCGAGCCGACTTTGC
CAAACTGGTTCGCGGTCACAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGGTG
GTGGATGAGTGCTACATCCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAG
TGGGCGTGGACTAATATGGAACAGTATTTAAGCGCCTGTTTGAATCTCACGGAGCGT
AAACGGTTGGTGGCGCAGCATCTGACGCACGTGTCGCAGACGCAGGAGCAGAACAA

AGAGAATCAGAATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAACTTCAGCCAG
GTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGT
GGATCCAGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGCGGT
CCCAAATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACC
GCCCCCGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATT
TATAAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGG
GATGGGCCACGAAAAAGTTCGGCAAGAGGAACACCATCTGGCTGTTTGGGCCTGCA
ACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCTTCTACGGG
TGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTCGACAAGATGGTG
ATCTGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCAT
TCTCGGAGGAAGCAAGGTGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAG
ACCCGACTCCCGTGATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGACGGGA
ACTCAACGACCTTCGAACACCAGCAGCCGTTGCAAGACCGGATGTTCAAATTTGAAC
TCACCCGCCGTCTGGATCATGACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGACT
TTTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAGCATGAATTCTACGTCAAAA
AGGGTGGAGCCAAGAAAAGACCCGCCCCCAGTGACGCAGATATAAGTGAGCCCAAA
CGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCGGAAGCTTCGATCAAC
TACGCAGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATGCTG
TTTCCCTGCAGACAATGCGAGAGAATGAATCAGAATTCAAATATCTGCTTCACTCAC
GGACAGAAAGACTGTTTAGAGTGCTTTCCCGTGTCAGAATCTCAACCCGTTTCTGTCG
TCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCATGGGAAAGGTGCCAG
ACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCATCTTTGAACA
ATAAATGATTTAAATCAGGTATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGG
ACACTCTCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCCACCACCAC
CAAAGCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTAC
AAGTACCTCGGACCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGA
CGCCGCGGCCCTCGAGCACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACA
ACCCGTACCTCAAGTACAACCACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAA
GATACGTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGAGGGTT
CTTGAACCTCTGGGCCTGGTTGAGGAACCTGTTAAGACGGCTCCGGGAAAAAAGAG
GCCGGTAGAGCACTCTCCTGTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGG
GCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGAGACGCAGACTCA
GTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAACT
AATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGA
CGGAGTGGGTAATTCCTCGGGAAATTGGCATTGCGATTCCACATGGATGGGCGACAG
AGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCCACCTACAACAACCACCTCTA
CAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTACAG
CACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACCACGTGAC
TGGCAAAGACTCATCAACAACAACTGGGGATTCCGACCCAAGAGACTCAACTTCAA
GCTCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGACGACGATTGC
CAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTA
CGTCCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCCAGCAGACGTCTTCAT
GGTGCCACAGTATGGATACCTCACCCTGAACAACGGGAGTCAGGCAGTAGGACGCT
CTTCATTTTACTGCCTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAACTT
TACCTTCAGCTACACTTTTGAGGACGTTCCTTTCCACAGCAGCTACGCTCACAGCCAG
AGTCTGGACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAA
CAAACACTCCAAGTGGAACCACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAG
CGAGTGACATTCGGGACCAGTCTAGGAACTGGCTTCCTGGACCCTGTTACCGCCAGC
AGCGAGTATCAAAGACATCTGCGGATAACAACAACAGTGAATACTCGTGGACTGGA
GCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATG
GCAAGCCACAAGGACGATGAAGAAAAGTTTTTTCCTCAGAGCGGGGTTCTCATCTTT

GGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTGAAAAGGTCATGATTACAGA
CGAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAGTATGGTTCTGTATC
TACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACCGCAGATGTCAACACACAAG
GCGTTCTTCCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCT
GGGCAAAGATTCCACACACGGACGGACATTTTCACCCCTCTCCCCTCATGGGTGGAT
TCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGA
ATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCATCACACAGTACTCCAC
GGGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAACGCT
GGAATCCCGAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAATGTGGACTTTA
CTGTGGACACTAATGGCGTGTATTCAGAGCCTCGCCCCATTGGCACCAGATACCTGA
CTCGTAATCTGTAATTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTTGAACT
TTGGTCTCTGCGTATTTCTTTCTTATCTAGTTTCCATGGCTACGTAGATAAGTAGCATG
GCGGGTTAATCATTAACTACAGCCCGGGCGTTTAAACAGCGGGCGGAGGGGTGGAG
TCGTGACGTGAATTACGTCATAGGGTTAGGGAGGTCCTGTATTAGAGGTCACGTGAG
TGTTTTGCGACATTTTGCGACACCATGTGGTCTCGCTGGGGGGGGGGGCCCGAGTGA
GCACGCAGGGTCTCCATTTTGAAGCGGGAGGTTTGAACGAGCGCTGGCGCGCTCACT
GGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCG
CCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGA
TCGCCCTTCCCATGCATCGGCCGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAA
TCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTA
CGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAA
TGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
ACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTA
CGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACA
TGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC
CATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT
CAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGT
AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCC
ACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGC
TAGCGCCACCATGGATAAAAAACCATTAGATGTTTTAATATCTGCGACCGGGCTCTG
GATGTCCAGGACTGGCACGCTCCACAAAATCAAGCACCATGAGGTCTCAAGAAGTA
AAATATACATTGAAATGGCGTGTGGAGACCATCTTGTTGTGAATAATTCCAGGAGTT
GTAGAACAGCCAGAGCATTCAGACATCATAAGTACAGAAAAACCTGCAAACGATGT
AGGGTTTCGGACGAGGATATCAATAATTTTCTCACAAGATCAACCGAAAGCAAAAAC
AGTGTGAAAGTTAGGGTAGTTTCTGCTCCAAAGGTCAAAAAAGCTATGCCGAAATCA
GTTTCAAGGGCTCCGAAGCCTCTGGAAAATTCTGTTTCTGCAAAGGCATCAACGAAC
ACATCCAGATCTGTACCTTCGCCTGCAAAATCAACTCCAAATTCGTCTGTTCCCGCAT
CGGCTCCTGCTCCTTCACTTACAAGAAGCCAGCTTGATAGGGTTGAGGCTCTCTTAAG
TCCAGAGGATAAAATTTCTCTAAATATGGCAAAGCCTTTCAGGGAACTTGAGCCTGA
ACTTGTGACAAGAAGAAAAAACGATTTTCAGCGGCTCTATACCAATGATAGAGAAG
ACTACCTCGGTAAACTCGAACGTGATATTACGAAATTTTTCGTAGACCGGGGTTTTCT
GGAGATAAAGTCTCCTATCCTTATTCCGGCGGAATACGTGGAGAGAATGGGTATTAA
TAATGATACTGAACTTTCAAAACAGATCTTCCGGGTGGATAAAAATCTCTGCTTGAG
GCCAATGCTTGCCCCGACTCTTTACAACTATCTGCGAAAACTCGATAGGATTTTACCA
GGCCCAATAAAAATTTTCGAAGTCGGACCTTGTTACCGGAAAGAGTCTGACGGCAAA
GAGCACCTGGAAGAATTTACTATGGTGAACTTCTGTCAGATGGGTTCGGGATGTACT
CGGGAAAATCTTGAAGCTCTCATCAAAGAGTTTCTGGACTATCTGGAAATCGACTTC
GAAATCGTAGGAGATTCCTGTATGGTCTTTGGGGATACTCTTGATATAATGCACGGG
GACCTGGAGCTTTCTTCGGCAGTCGTCGGGCCAGTTTCTCTTGATAGAGAATGGGGT

ATTGACAAACCATGGATAGGTGCAGGTTTTGGTCTTGAACGCTTGCTCAAGGTTATG
CACGGCTTTAAAAACATTAAGAGGGCATCAAGGTCCGAATCTTACTATAATGGGATT
TCAACCAATCTGTAAGAATTCAACGCGTTAAGTCGACTTTAACTCGAGTCTAGAGGG
CCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC
TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGG
GGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
CTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCCTAGGGGTGCGA
GCGGATCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACCCGTGATC
TTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTTGAT
GTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCATACCTGACCTCCATA
GCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTTGATCGGCACGTAAGAGGT
TCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGA
GATTTTCAGGAGCTAAGGAAGCTAAAATGAGCCATATTCAACGGGAAACGTCTTGCT
TGAAGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTC
GCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATG
CGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATG
AGATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATT
TTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACAG
CATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGG
CAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACGGCGA
TCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGC
GAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAAT
GCATAAACTCTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTT
GATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTC
GGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTT
TCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGA
ATAAATTGCAGTTTCACTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAA
TCACCTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTC
CGCCCCCCTGACGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCC
GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTG
GCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCA
AGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTA
ACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGT
GGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGA
AGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGC
TGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATC
TCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pIDTSmart-8xU6MmPytR-ITR-GFP
Yellow=*Mm*PytR
Pink=ITR
Green=GFP

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGTGGAC
CGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCTGAGGCTCGT
CCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATCCAGTC
GCGCTGCTCTCGTCGATCCGCTAGGACTATGGTTGCTTTGACGTATGCGGTGTGAAAT
ACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCCCTGCAGGCAGCTGC

GCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTG
GTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC
ACTAGGGGTTCCTGCGGCCGCTCGGTCCGCACGTGGTTACCTACAAAATCAGAAGGA
CAGGGAAGGGAGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGT
GGGTAGATCACCTGAGATTAGGAGTTGGAGACCAGCCTGGCCAATATGGTGAAACC
CCGTCTCTACCAAAAAAACAAAAATTAGCTGAGCCTGGTCATGCATGCCTGGAATCC
CAACAACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGATTG
CAGTGAGCCAAGATTGTGCCACTGCACTCCAGCTTGGTTCCCAATAGACCCCGCAGG
CCCTACAGGTTGTCTTCCCAACTTGCCCCTTGCTCCATACCACCCCCCTCCACCCCAT
AATATTATAGAAGGACACCTAGTCAGACAAAATGATGCAACTTAATTTTATTAGGAC
AAGGCTGGTGGGCACTGGAGTGGCAACTTCCAGGGCCAGGAGAGGCACTGGGGAGG
GGTCACAGGGATGCCACCCGTAGATCTCTCGAGCAGCGCTGCTCGAGGCAAGCTTCT
GCAGGTCGACTCTAGAGGATCCCTACTTGAGCTCGAGATCTGAGTACTTGTACAGCT
CGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGAT
CGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTC
GGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGA
GCTGCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTT
CTTCTGCTTGTCGGCCATGATATAGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTG
TGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTTCA
CCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTT
GAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGT
CGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGGG
TGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTC
AGGGTCAGCTTGCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTG
TGGCCGTTTACGTCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGC
TCCTCGCCCTTGCTCACCATTCAGAATTCAATCGATGTTCGAATCCCAATTCTTTGCC
AAAGTGATGGGCCAGCACACAGACCAGCACGTTGCCCAGGAGCTGTGGGAGGAAGA
TAAGAGGTATGAACATGATTAGCAAAAGGGCCTAGCTTGGACTCAGAATAATCCAG
CCTTATCCCAACCATAAAATAAAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAG
CAATATGAAACCTCTTACATCAGTTACAATTTATATGCAGAAATATTTATATGCAGA
AATATTGCTATTGCCTTAACCCAGAAATTATCACTGTTATTCTTTAGAATGGTGCAAA
GAGGCATGATACATTGTATCATTATTGCCCTGAAAGAAAGAGATTAGGGAAAGTATT
AGAAATAAGATAAACAAAAAAGTATATTAAAAGAAGAAAGCATTTTTTGTGGGCCT
ATAGACTCTATAGGCGGTACTTACGTCACTCTTGGCACGGGGAATCCGCGTTCCAAT
GCACCGTTCCCGGCCGGGATTCGAATCCGCGGAGGCTGGATCGGTCCCGGTGTCTTC
TATGGAGGTCAAAACAGCGTGGATGGCGTCTCCAGGCGATCTGACGGTTCACTAAAC
GAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCGTCAATGG
GGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTTGGTGCAAAACAAACTC
CCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCA
CGCCCATTGATGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGACTAATACG
TAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAG
GCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACT
TGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGG
AAAGTCCCTATTGACGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGG
GGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCGGAAC
TCCATATATGGGCTATGAACTAATGACCCCGTAATTGATTACTATTAATAACTAGCTC
CACGCGTGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCG
CTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCC
GGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGACATGTGAGCAAAA
GGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA

ACCCGAGCTAGCAAAAAACGGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTA
GAGTCCGTTCGATCTACATGATCAGGTTTCCGGTGTTTCGTCCTTTCCACAAGATATA
TAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTT
AAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGT
ATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAG
CCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCC
CGACCTAGCAAAAAACGGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAG
TCCGTTCGATCTACATGATCAGGTTTCCGGTGTTTCGTCCTTTCCACAAGATATATAA
AGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAA
ACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTT
TGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTT
GTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAC
CTAGCAAAAAACGGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCG
TTCGATCTACATGATCAGGTTTCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCC
AAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACAT
AATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTA
CTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTAT
CGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAG
CAAAAAACGGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCGTTCG
ATCTACATGATCAGGTTTCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGA
AATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTT
TAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAAT
ATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTAT
ATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAA
AACGGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCGTTCGATCTAC
ATGATCAGGTTTCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCG
AAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAA
ACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCT
TTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGC
AAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAAC
GGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCGTTCGATCTACATG
ATCAGGTTTCC GGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAA
TACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTG
CAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAA
TATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAACCCCG
GGAATCTAACCCGGCTGAACGGATTTAGAGTCCGTTCGATCTACATGATCAGGTTTC
C
GGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGT
T ACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAA
ACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTT
ACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATG
AAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAACGGAAACC
CCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCGTTCGATCTACATGATCAGGT
TTCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCA
AGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTA
CCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAG
TCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGG
AATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGGGGTGCGAGCGGATCGAGCA
GTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACCCGTGATCTTACGGCATTAT
ACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTTGATGTCATAGCTGTT

TCCTGAGGCTCAATACTGACCATTTAAATCATACCTGACCTCCATAGCAGAAAGTCA
AAAGCCTCCGACCGGAGGCTTTTGACTTGATCGGCACGTAAGAGGTTCCAACTTTCA
CCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGG
AGCTAAGGAAGCTAAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCGC
GATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATG
TCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGT
TGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCA
GGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTAC
TCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACAGCATTCCAGGT
ATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCT
GCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCGCGTATTT
CGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGATTTT
GATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACT
CTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTT
ATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCA
GACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCAT
TACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGC
AGTTTCACTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGC
TCACCTTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCT
GACGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACT
ATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTC
ATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTT
GAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATTTTCTACCGAAGAAAGGCCCA

Chapter 6 Plasmids

pIDTSmart-MmPylRS-SCOKRS

Green=CMV
Blue=SOCKRS

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGT
GGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCT
GAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGG
CGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGGCGGCCGCAAATA
CCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGC
CTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGT
TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATT

TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAG
TACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT
CGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAG
CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG
TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCG
CCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG
CAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTA
ATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGATAAAA
AGCCTCTGAACACTCTGATTTCTGCGACCGGTCTGTGGATGTCCCGCACCGGC
ACCATCCACAAAATCAAACACCATGAAGTTAGCCGTTCCAAAATCTACATTG
AAATGGCTTGCGGCGATCACCTGGTTGTCAACAACTCCCGTTCTTCTCGTACC
GCTCGCGCACTGCGCCACCACAAATATCGCAAAACCTGCAAACGTTGCCGTG
TTAGCGATGAAGATCTGAACAAATTCCTGACCAAAGCTAACGAGGATCAGAC
CTCCGTAAAAGTGAAGGTAGTAAGCGCTCCGACCCGTACTAAAAAGGCTATG
CCAAAAAGCGTGGCCCGTGCCCCGAAACCTCTGGAAAACACCGAGGCGGCTC
AGGCTCAACCATCCGGTTCTAAATTTTCTCCGGCGATCCCAGTGTCCACCCAA
GAATCTGTTTCCGTACCAGCAAGCGTGTCTACCAGCATTAGCAGCATTTCTAC
CGGTGCTACCGCTTCTGCGCTGGTAAAAGGTAACACTAACCCGATTACTAGC
ATGTCTGCACCGGTACAGGCAAGCGCCCCAGCTCTGACTAAATCCCAGACGG
ACCGTCTGGAGGTGCTGCTGAACCCAAAGGATGAAATCTCTCTGAACAGCGG
CAAGCCTTTCCGTGAGCTGGAAAGCGAGCTGCTGTCTCGTCGTAAAAAGGAT
CTGCAACAGATCTACGCTGAGGAACGCGAGAACTATCTGGGTAAGCTGGAGC
GCGAAATTACTCGCTTCTTCGTGGATCGCGGTTTCCTGGAGATCAAATCTCCG
ATTCTGATTCCGCTGGAATACATTGAACGTATGGGCATCGATAATGATACCG
AACTGTCTAAACAGATCTTCCGTGTGGATAAAAACTTCTGTCTGCGTCCGATG
CTGGCCCCGAACCTGGCCAACTATCTGCGTAAACTGGACCGTGCCCTGCCGG
ACCCGATCAAAATTTTCGAGATCGGTCCTTGCTACCGTAAAGAGTCCGACGG
TAAAGAGCACCTGGAAGAATTCACCATGCTGAACTTCTGCCAGATGGGTAGC
GGTTGCACGCGTGAAAACCTGGAATCCATTATCACCGACTTCCTGAATCACCT
GGGTATCGATTTCAAAATTGTTGGTGACAGCTGTATGGTGTTCGGCGATACGC
TGGATGTTATGCACGGCGATCTGGAGCTGTCTTCCGCAGTAGTGGGCCCAATC
CCGCTGGATCGTGAGTGGGGTATCGACAAACCTTGGATCGGTGCGGGTTTTG
GTCTGGAGCGTCTGCTGAAAGTAAAACACGACTTCAAGAACATCAAACGTGC
TGCACGTTCCGAGTCCTATTACAATGGTATTTCTACTAACCTGTAAGAATTCA
ACGCGTTAAGTCGACTTTAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTG
ATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCC
CGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAA
ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCCTAGGGG
TGCGAGCGGATCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCG
ACCCGTGATCTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTA
TCTAGTCAGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTA
AATCATACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTT
TTGACTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGA

TCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGC
TAAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAAT
TCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCG
GGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGA
GTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAG
ATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGC
ATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGG
AAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTG
TTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATT
GTCCTTTTAACGGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATG
AATAACGGTTTGGTTGGTGCGAGTGATTTTGATGACGAGCGTAATGGCTGGC
CTGTTGAACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCTCACCGGA
TTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGG
GGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATA
CCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTAC
AGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTG
CAGTTTCACTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCA
CCTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCT
CCGCCCCCCTGACGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGA
AACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG
TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCC
CTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCG
GTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG
GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
ACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGG
AAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pEvol-aHer2-Fab

Green=AraC
Yellow=araBAD
Blue=StII leader sequence
Purple=Light Chain
Orange=Heavy Chain

TCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCAT
TATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAA
AAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTC
TGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTG
ATGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGT
GGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGC

GTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCT
TACTATGTTGGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAG
AAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCC
GCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGA
AATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTT
AACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCC
CCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCG
ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCT
CTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTT
TGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCT
GGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTA
ACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGC
AGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTA
AGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTA
CCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCA
AGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATC
TCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAAT
TTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAA
TTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAATTATGACAAC
TTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTCGG
GCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCA
AAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAA
GCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATC
CCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACA
TGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGAT
GTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCG
TTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAG
CAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACA
GGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATT
GGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGGACGA
AAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGT
GATGAATCTCTCCTGGCGGGAACAGCAAAATATCACTCGGTCGGCAAACAAA
TTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAATGGTGAGATTGAGAA
TATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTT
GGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCC
GGCAGCAGGGGATCATTTTACTTTTCATACTCCCGCCATTCAGAGAAGAAAC
CAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTT
CTCGCTAACCAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGC
GGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGC
AGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGC
ATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTC
TACTGTTTCTCCATACCCGTTTTTTTGGGCTAAAGAAATAATTTTGTTTAACTT
TAAGAAGGAGAATACATCAACTAGTACGCAAGTTCACGTAAAAAGGGTATCT
AGAGGTTGAGGTGATTTTATGAAAAAGAATATCGCATTTCTTCTTGCTAGCAT
GTTCGTTTTTTCTATTGCTACAAACGCATACGCTGACATCCAGATGACCCAGT
CTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGG

GCAAGTCAGGATGTGAATACCGCGGTCGCATGGTATCAGCAGAAACCAGGG
AAAGCCCCTAAGCTCCTGATCTATTCTGCATCCTTCTTGTATAGTGGGGTCCC
ATCAAGGTTCAGTGGCAGTAGATCTGGGACAGATTTCACTCTCACCATCAGC
AGTCTGCAACCTGAAGATTTTGCAACTTACTACTGTCAACAGCATTACACTAC
CCCTCCGACGTTCGGCCAAGGTACCAAGCTTGAGATCAAACGAACTGTGGCT
GCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAC
TGCCTCTGTCGTGTGCCTGCTGAATAACTTCTATCCCAGAGAGGCCAAAGTAC
AGTGGAAGGTGGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAGTGTCAC
AGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCTGACGCT
GAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCA
TCAGGGCCTGTCCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTTAA
GCCTGCAGGGGATCCTCTAGAGGTTGAGGTGATTTTATGAAAAAGAATATCG
CATTTCTTCTTGCATCTATGTTCGTTTTTTCTATTGCTACAAACGCGTACGCTG
AGGTGCAGCTGGTGGAGTCTGGAGGAGGCTTGGTCCAGCCTGGGGGGTCCCT
GAGACTCTCCTGTGCAGCCTCTGGGTTCAATATTAAGGACACTTACATCCACT
GGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCGCACGTATTTATCC
TACCAATGGTTACACACGCTACGCAGACTCCGTGAAGGGCCGATTCACCATC
TCCGCAGACACTTCCAAGAACACGGCGTATCTTCAAATGAACAGCCTGAGAG
CCGAGGACACGGCCGTGTATTACTGTTCGAGATGGGGCGGTGACGGCTTCTA
TGCCATGGACTACTGGGGCCAAGGAACCCTGGTCACCGTCTCCTCAGCCTCC
ACCAAGGGCCCATCGGTCTTCCCCCTGGCACCCTCCTCCAAGAGCACCTCTGG
GGGCACAGCGGCCCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAACCGGTG
ACGGTGTCGTGGAACTCAGGCGCCCTGACCAGCGGCGTGCACACCTTCCCGG
CTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACTGTGCCC
TCTAGCAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCA
GCAACACCAAGGTGGACAAGAAAGTTGAGCCCAAATCTTGTGACAAAACTCA
CACATAAACCGATGCGGCCGCTTGAGAGTCAGCTCCTTCCGGTGGGCGTGCC
TGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTG
AAACGCCGTAGCGCC
GATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAA
ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTGTCGACCGAATTTCTGC
CATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGG
CACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGT
ACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGC
ATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAAT
ATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTT
TAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATA
TTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCAC
ATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCC
AGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTG
AACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCG
GATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTT
GTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGG
TCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTA
CGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCAT
TTTAGCTTCCTTAGC

pULTRA-Mj-CNF

TCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATA
AACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGACTGCGGCG
GGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGG
ACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAG
GCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTTCCGGTGCATCTG
TGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACGGGCGAAACCCGAC
AGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCC
TGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGG
AAGTGTGGCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGT
CGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGC
TGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTA
AAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTT
TAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTG
GAAGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAG
CAGTTCCCCAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTT
TCGTTTACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCA
AATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTA
GTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGG
GCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGC
GTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATT
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGG
GTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGC
CTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGG
CGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTC
GGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCAGCCCGGAC
TCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCA
TCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCG
GACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCG
AGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTT
AATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCT
CCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGG
TGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCT
TCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCAC
TGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCC
GCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGA
GATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGG
TGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCG
GTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTT
CGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAG
ACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCAC
CCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGC
GCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCAT

TAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGA
GCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGGACGAATTTGAAATG
ATAAAGAGAAACACATCTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGTT
TTAAAAAAAGATGAAAAATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAA
TACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCT
GGATTTGATATAATTATAGTTTTGGCTGATTTACATGCCTATTTAAACCAGAA
AGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATAACAAAAAAGTTTTT
GAAGCAATGGGGTTAAAGGCAAAATATGTTTATGGAAGTGAATGGATGCTTG
ATAAGGATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAA
AGAGCAAGAAGGAGTATGGAACTTATAGCAAGAGAGGATGAAAATCCAAAG
GTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGGTGCTCATTATCTTGG
CGTTGATGTTGCAGTTGGGGGGATGGAGCAGAGAAAAATACACATGTTAGCA
AGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTCACAACCCTGTCTTAACGGG
TTTGGATGGAGAAGGAAAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTT
GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCATACTGCCCAG
CTGGAGTTGTTGAAGGAAATCCAATAATGGAGATAGCTAAATACTTCCTTGA
ATATCCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTTGACAGTT
AATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAA
TGGATTTAAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAAT
TAGAAAGAGATTATAAGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGTTTT
GGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAA
GCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACC
TGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTG
GGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAA
GGCTCAGTCGAAAGACTGGGCCTTGTTTGTGAGCTCCCGGTCATCAATCATCC
CCATAATCCTTGTTAGCCTGCAGGTAATTCCGCTTCGCAACATGTGAGCACCG
GTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGC
CAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTGC
ACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCT
ATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCC
GGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCATGGCAGGG
GTTCAAATCCCCTCCGCCGGACCAAATTCGAAAAGCCTGCTCAACGAGCAGG
CTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCCATGGCGGCCACCAG
GTACCACCGGCGCCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGG
TAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACC
CTGCCCTGAACCGACGACCGGGTCATCGTGGCCGGATCTTGCGGCCCCTCGG
CTTGAACGAATTGTTAGACATTATTTGCCGACTACCTTGGTGATCTCGCCTTT
CACGTAGTGGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCT
TCTTCTTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTG
GGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATT
TTGCCGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTC
GCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTT
AGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCG
CTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCC
AGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGC
GCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATG

ATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTC
TCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTT
GTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGG
CTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTC
GGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATA
CTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAGCA
TTTA

pBK_Her2_Fab

Green=AraC
Yellow=araBAD
Blue=StII leader sequence
Purple=Light Chain
Orange=Heavy Chain

GAATTCCTGCTTTTCTTCGCGAATTAATTCCGCTTCGCACATGTGAGCAAAAG
GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCA
TAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGG
TGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC
GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACT
ACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT
TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCT
GGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAG
GATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAC
GAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTAC
ATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTT
GCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAA
ATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTAT
GGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTG
CCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTAT
GCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTAC
TCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCC
TGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGC
ATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCG
CTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGA
TGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAG
CTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGAT
AACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGT
CGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGT

GAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAA
TCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATC
AGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGG
GACGGCGGCTTTGTTGAATAAATCGCCTAGGCGATATAAGTTGTAATTCTCAT
GTTTGACAGCTTATCATCGATAAGCTTGGTACCCAATTATGACAACTTGACGG
CTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCC
CCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAA
CATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTT
CGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACT
GCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTG
CGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTG
ACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATC
GCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCT
CCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCG
CTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAA
ATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGGACGAAAGTA
AACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGA
ATCTCTCCTGGCGGGAACAGCAAAATATCACTCGGTCGGCAAACAAATTCTC
GTCCCTGATTTTTCACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAA
CCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTC
AATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGC
AGGGGATCATTTTGCGCTTCAGCCATACTTTTCATACTCCCGCCATTCAGAGA
AGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACT
GGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAA
CAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAAT
CACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGC
CATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGC
AACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAAGAAATAATTTTGT
TTAACTTTAAGAAGGAGAATACATCAACTAGTACGCAAGTTCACGTAAAAAG
GGTATCTAGAGGTTGAGGTGATTTTATGAAAAAGAATATCGCATTTCTTCTTG
CTAGCATGTTCGTTTTTTCTATTGCTACAAACGCATACGCTGACATCCAGATG
ACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCAC
TTGCCGGGCAAGTCAGGATGTGAATACCGCGGTCGCATGGTATCAGCAGAAA
CCAGGGAAAGCCCCTAAGCTCCTGATCTATTCTGCATCCTTCTTGTATAGTGG
GGTCCCATCAAGGTTCAGTGGCAGTAGATCTGGGACAGATTTCACTCTCACC
ATCAGCAGTCTGCAACCTGAAGATTTTGCAACTTACTACTGTCAACAGCATTA
CACTACCCCTCCGACGTTCGGCCAAGGTACCAAGCTTGAGATCAAACGAACT
GTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATC
TGGAACTGCCTCTGTCGTGTGCCTGCTGAATAACTTCTATCCCAGAGAGGCCA
AAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAG
TGTCACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCT
GACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGT
CACCCATCAGGGCCTGTCCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAG
TGTTAAGCCTGCAGGGGATCCTCTAGAGGTTGAGGTGATTTTATGAAAAAGA
ATATCGCATTTCTTCTTGCATCTATGTTCGTTTTTTCTATTGCTACAAACGCGT
ACGCTGAGGTGCAGCTGGTGGAGTCTGGAGGAGGCTTGGTCCAGCCTGGGGG

GTCCCTGAGACTCTCCTGTGCAGCCTCTGGGTTCAATATTAAGGACACTTACA
TCCACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCGCACGTAT
TTATCCTACCAATGGTTACACACGCTACGCAGACTCCGTGAAGGGCCGATTC
ACCATCTCCGCAGACACTTCCAAGAACACGGCGTATCTTCAAATGAACAGCC
TGAGAGCCGAGGACACGGCCGTGTATTACTGTTCGAGATGGGGCGGTGACGG
CTTCTATGCCATGGACTACTGGGGCCAAGGAACCCTGGTCACCGTCTCCTCAG
CCTCCACCAAGGGCCCATCGGTCTTCCCCCTGGCACCCTCCTCCAAGAGCACC
TCTGGGGGCACAGCGGCCCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAAC
CGGTGACGGTGTCGTGGAACTCAGGCGCCCTGACCAGCGGCGTGCACACCTT
CCCGGCTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACTG
TGCCCTCTAGCAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAA
GCCCAGCAACACCAAGGTGGACAAGAAAGTTGAGCCCAAATCTTGTGACAA
AACTCACACATAAACCGATGCGGCCGCTTGAGAGTCAGCTCCTTCCGGTGGG
CGTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCA
GAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAG
TAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGG
GCCTTGTCGACCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCG
TAGCAACCAGGCGTTTAAGGGCAC

pEvoltac-EcW-TGA.h14

Green=LacI
Orange=TacI promoter
Blue=EcTrpRS-h14
Pink=proK promoter
Teal=EcTrp-tRNA UCA

GGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTG
TATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGTGCTTCATG
TGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAG
GATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGC
GGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCA
GGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCAT
AGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTC
CCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTT
ATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTT
CGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCG
CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGC
ACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTC
ATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCT
CCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAA
AACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAG
ACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAG
ATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACG
ATATAAGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGCAATT

TATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAAT
TCTCATGTTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAG
GCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTA
CATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGC
CAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCC
CTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGC
CCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATG
AGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGC
AGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGG
CAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGT
TGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAAT
TTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAG
ACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGA
CCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACT
GTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTG
CAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGA
TCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGC
TTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGAT
CGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAG
ACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGT
GCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTC
CCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTC
TGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCAC
ATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAA
AGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGA
CTCCTGCATTAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGG
AATTGTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGACTAAGC
CCATCGTTTTTAGTGGCGCACAGCCCTCAGGTGAATTGACCATTGGTAACTAC
ATGGGTGCGCTGCGTCAGTGGGTAAACATGCAGGATGACTACCATTGCATTT
ACTGTATCGTTGACCAACACGCGATCACCGTGCGCCAGGATGCACAGAAGCT
GCGTAAAGCGACGCTGGATACGCTGGCCTTGTATCTGGCTTGTGGTATCGATC
CTGAGAAAAGCACCATTTTTGTTCAGTCCCACGTGCCGGAACATGCACAGTT
AGGCTGGGCACTGAACTGCTATACCTACTTCGGCGAACTGAGTCGCATGACG
CAGTTTAAAGATAAATCTGCGCGTTATGCCGAGAACATCAACGCTGGTCTGTT
TGACTATCCGGTGCTGATGGCAGCGGACATCCTGCTGTATCAAACTAATCTGG
TACCGGTGGGTGAAGACCAGAAACAGCACCTCGAACTGAGCCGCGATATTGC
CCAGCGTTTCAACGCGCTGTATGGCGAGATCTTTAAGGTGCCGGAGCCGTTTA
TTCCGAAATCTGGCGCGCGCGTAATGTCGCTGCTGGAGCCGACCAAGAAGAT
GTCCAAGTCTGACGATAATCGCAATAACGTTATCGGCCTGCTGGAAGATCCG
AAATCGGTAGTGAAGAAAATCAAACGTGCGGTCACTGACTCCGACGAGCCGC
CGGTAGTTCGCTACGATGTGCAGAACAAAGCGGGCGTTTCCAACCTGTTGGA
TATCCTTTCAGCGGTAACGGGCCAGAGCATCCCAGAACTGGAAAAACAGTTC
GAAGGCAAGATGTATGGTCATCTGAAAGGTGAAGTGGCTGATGCCGTTTCCG
GTATGCTGACTGAATTGCAGGAACGCTATCACCGTTTCCGCAACGATGAAGC
CTTCCTGCAACAGGTGATGAAAGATGGCGCGGAAAAAGCCAGCGCGCACGC

TTCCCGTACGCTAAAAGCGGTGTACGAAGCGATTGGTTTTGTGGCGAAGCCG
TAAGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCGGATGAGAG
AAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAA
AACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCC
GAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCAT
GCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAA
AGACTGGGCCTTGTTTGTGAGCTCCCGGTCATCAATCATCCCCATAATCCTTG
TTAGCCTGCAGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCC
CCGTGGAGGTAATAATTGACGATATGATCAGTGCACGGCTAACTAAGCGGCC
TGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGG
GCGTATCTGCGCAGTAAGATGCGCCCCGCATTTAGGGGCGTAGTTCAATTGG
TAGAGCACCGGTCTTCAAAACCGGGTGTTGGGAGTTCGAGTCTCTCCGCCCCT
GCCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCA
GCTCAGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACT
TATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAA
AATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCAT
TCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGC
GGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGG
GGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTC
ACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGA
AATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAG
AAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCA
GTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCA
GCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGG
GCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGT
CTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGA
GCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATC
AACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGA
AAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGT
GAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTG
GCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGA
AGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCT
GCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTT
CTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGA

pIDT-ITR-GFP-4xwtR-wRS

Pink=ITRs
Green=GFP
Blue=U6 promoter
Red=wtR
Yellow=CMV promoter
Orange=wRS

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGT
GGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCT
GAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGG
CGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGACTATGGTTGCTTT
GACGTATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCAT
CAGGCGCCCCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGG
GCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGC
GAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTGCGGCCG
CTCGGTCCGCACGTGGTTACCTACAAAATCAGAAGGACAGGGAAGGGAGCA
GTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCA
CCTGAGATTAGGAGTTGGAGACCAGCCTGGCCAATATGGTGAAACCCCGTCT
CTACCAAAAAAACAAAAATTAGCTGAGCCTGGTCATGCATGCCTGGAATCCC
AACAACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAG
ATTGCAGTGAGCCAAGATTGTGCCACTGCACTCCAGCTTGGTTCCCAATAGAC
CCCGCAGGCCCTACAGGTTGTCTTCCCAACTTGCCCCTTGCTCCATACCACCC
CCCTCCACCCCATAATATTATAGAAGGACACCTAGTCAGACAAAATGATGCA
ACTTAATTTTATTAGGACAAGGCTGGTGGGCACTGGAGTGGCAACTTCCAGG
GCCAGGAGAGGCACTGGGGAGGGGTCACAGGGATGCCACCCGTAGATCTCT
CGAGCAGCGCTGCTCGAGGCAAGCTTCTGCAGGTCGACTCTAGAGGATCCCT
ACTTGAGCTCGAGATCTGAGTACTTGTACAGCTCGTCCATGCCGAGAGTGATC
CCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGG
GGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCAG
CACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGC
ACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTT
CTTCTGCTTGTCGGCCATGATATAGACGTTGTGGCTGTTGTAGTTGTACTCCA
GCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCG
ATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTA
GTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGC
ATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGA
AGCACTGCACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGCACGG
GCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGTAGGTGGC
ATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCCGT
CCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCT
CACCATTCAGAATTCAATCGATGTTCGAATCCCAATTCTTTGCCAAAGTGATG
GGCCAGCACACAGACCAGCACGTTGCCCAGGAGCTGTGGGAGGAAGATAAG
AGGTATGAACATGATTAGCAAAAGGGCCTAGCTTGGACTCAGAATAATCCAG
CCTTATCCCAACCATAAAATAAAAGCAGAATGGTAGCTGGATTGTAGCTGCT
ATTAGCAATATGAAACCTCTTACATCAGTTACAATTTATATGCAGAAATATTT
ATATGCAGAAATATTGCTATTGCCTTAACCCAGAAATTATCACTGTTATTCTT
TAGAATGGTGCAAAGAGGCATGATACATTGTATCATTATTGCCCTGAAAGAA
AGAGATTAGGGAAAGTATTAGAAATAAGATAAACAAAAAAGTATATTAAAA
GAAGAAAGCATTTTTTGTGGGCCTATAGACTCTATAGGCGGTACTTACGTCAC
TCTTGGCACGGGGAATCCGCGTTCCAATGCACCGTTCCCGGCCGGGATTCGA
ATCCGCGGAGGCTGGATCGGTCCCGGTGTCTTCTATGGAGGTCAAAACAGCG
TGGATGGCGTCTCCAGGCGATCTGACGGTTCACTAAACGAGCTCTGCTTATAT
AGACCTCCCACCGTACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAGTT

GTTACGACATTTTGGAAAGTCCCGTTGATTTTGGTGCAAAACAAACTCCCATT
GACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCA
CGCCCATTGATGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGACTA
ATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGC
ATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTT
GGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTC
CACCCATTGACGTCAATGGAAAGTCCCTATTGACGTTACTATGGGAACATAC
GTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGC
CATTTACCGTAAGTTATGTAACGCGGAACTCCATATATGGGCTATGAACTAAT
GACCCCGTAATTGATTACTATTAATAACTAGCTCCACGCGTGCGGCCGCAGG
AACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACT
GAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCT
CAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG
CGAAACCCGAGCTAGCAAAAAATGGCAGGGGCGGAGAGACTCGAACTCCCA
ACACCCGGTTTTGAAGACCGGTGCTCTACCAATTGAACTACGCCCCTGGTGTT
TCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTA
CGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTAC
CCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTT
ACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAA
ATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAA
ATGGCAGGGGCGGAGAGACTCGAACTCCCAACACCCGGTTTTGAAGACCGGT
GCTCTACCAATTGAACTACGCCCCTGGTGTTTCGTCCTTTCCACAAGATATAT
AAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCA
TTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTA
CGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATT
ATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAA
ATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATGGCAGGGGCGGAGAGACT
CGAACTCCCAACACCCGGTTTTGAAGACCGGTGCTCTACCAATTGAACTACG
CCCCTGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATA
CTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAA
CTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAAT
ATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTAT
CGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGA
CCTAGCAAAAAATGGCAGGGGCGGAGAGACTCGAACTCCCAACACCCGGTTT
TGAAGACCGGTGCTCTACCAATTGAACTACGCCCCTGGTGTTTCGTCCTTTCC
ACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCAT
ATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATT
ATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAAT
TAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGA
ATCATGGGAAATAGGCCCTCTTCCTGCCCGACTTATTAATAGTAATCAATTAC
GGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGG
TAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAAT
AATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT
GGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCAT

ATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC
ATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTAC
GTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGG
GCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG
TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGT
AACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGG
TCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTT
ATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACC
ATGACTAAGCCCATCGTTTTTGCTGGCGCACAGCCCTCAGGTGAATTGACCAT
TGGTAACTACATGGGTGCGCTGCGTCAGTGGGTAAACATGCAGGATGACTAC
CATTGCATTTACTGTATCGTTGACCAACACGCGATCACCGTGCGCCAGGATGC
ACAGAAGCTGCGTAAAGCGACGCTGGATACGCTGGCCTTGTATCTGGCTTGT
GGTATCGATCCTGAGAAAAGCACCATTTTTGTTCAGTCCCACGTGCCGGAAC
ATGCACAGTTAGGCTGGGCACTGAACTGCTATACCTACTTCGGCGAACTGAG
TCGCATGACGCAGTTTAAAGATAAATCTGCGCGTTATGCCGAGAACATCAAC
GCTGGTCTGTTTGACTATCCGGTGCTGATGGCAGCGGACATCCTGCTGTATCA
AACTAATCTGGGTCCTTGTGGTGAAGACCAGAAACAGCACCTCGAACTGAGC
CGCGATATTGCCCAGCGTTTCAACGCGCTGTATGGCGAGATCTTTAAGGTGCC
GGAGCCGTTTATTCCGAAATCTGGCGCGCGCGTAATGTCGCTGCTGGAGCCG
ACCAAGAAGATGTCCAAGTCTGACGATAATCGCAATAACGTTATCGGCCTGC
TGGAAGATCCGAAATCGGTAGTGAAGAAAATCAAACGTGCGGTCACTGACTC
CGACGAGCCGCCGGTAGTTCGCTACGATGTGCAGAACAAAGCGGGCGTTTCC
AACCTGTTGGATATCCTTTCAGCGGTAACGGGCCAGAGCATCCCAGAACTGG
AAAAACAGTTCGAAGGCAAGATGTATGGTCATCTGAAAGGTGAAGTGGCTGA
TGCCGTTTCCGGTATGCTGACTGAATTGCAGGAACGCTATCACCGTTTCCGCA
ACGATGAAGCCTTCCTGCAACAGGTGATGAAAGATGGCGCGGAAAAAGCCA
GCGCGCACGCTTCCCGTACGCTAAAAGCGGTGTACGAAGCGATTGGTTTTGT
GGCGAAGCCGTAAGAATTCAACGCGTTAAGTCGACAATCAACCTCTGGATTA
CAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGC
TATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGG
CTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGT
TGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCA
ACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTT
CGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCC
GCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTC
GGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTC
TGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTT
CCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCG
CCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCGTCGACTTTAAC
TCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCT
AGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATT
GTCTGAGTAGGTGTCATTCTATTCTGGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCAC
GCGCCCTGTAGCGGCGCATTAAGCGCAAGCCCTGCCATAGCCACTACGGGTA

CGTAGCCAACCACTAGAACTATAGCTAGAGTCCTGGGCGAACAAACGATGCT
CGCCTTCCAGAAAACCGAGGATGCGAACCACCCTAGGCTAGGGGTGCGAGC
GGATCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACCCGTG
ATCTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTC
AGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCATA
CCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTT
GATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTAC
CGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAAT
GAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAAC
ATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAAT
CAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTT
TCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTC
AGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTAT
CCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACA
GCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGC
GCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTT
TAACGGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAAC
GGTTTGGTTGGTGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGA
ACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCTCACCGGATTCAGTC
GTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATT
AATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGAT
CTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACG
GCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTC
ACTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCT
CACCTTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCG
ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAG
GTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACC
GCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGA
CTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT
GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAA
AGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGA
TCCTTTGATTTTCTACCGAAGAAAGGCCCA

pEvoltac-EcW-H14-MjY-CNF.TAG

Red=TacI promoter
Orange=MjYRS-CNF
Teal=proK promoter
Green=tRNA
Purple=LacI
Light green=EcWRS

Pink=EcWtR TGA

GGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGAGCTG
TTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACA
ATTTCACAAAGGAGGTGCGGCCGCATGGACGAATTTGAAATGATAAAGAGA
AACACATCTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAA
GATGAAAAATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAATACATTTAG
GGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTGGATTTGA
TATAATTATAGTTTTGGCTGATTTACATGCCTATTTAAACCAGAAAGGAGAGT
TGGATGAGATTAGAAAAATAGGAGATTATAACAAAAAAGTTTTTGAAGCAAT
GGGGTTAAAGGCAAAATATGTTTATGGAAGTGAATGGATGCTTGATAAGGAT
TATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGAGCAA
GAAGGAGTATGGAACTTATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTG
AAGTTATCTATCCAATAATGCAGGTTAATGGTGCTCATTATCTTGGCGTTGAT
GTTGCAGTTGGGGGGATGGAGCAGAGAAAAATACACATGTTAGCAAGGGAG
CTTTTACCAAAAAAGGTTGTTTGTATTCACAACCCTGTCTTAACGGGTTTGGA
TGGAGAAGGAAAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGAC
TCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCATACTGCCCAGCTGGAG
TTGTTGAAGGAAATCCAATAATGGAGATAGCTAAATACTTCCTTGAATATCCT
TTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCT
ATGAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT
AAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAG
AGATTATAAGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCGGAT
GAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCT
GATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCC
ATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTC
CCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAG
TCGAAAGACTGGGCCTTGTTTGTGAGCTCCCGGTCATCAATCATCCCCATAAT
CCTTGTTAGCCTGCAGGATCCGGCCATGGCGGCCACCAGGTACCCACCGGTT
TATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCA
GTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTGCAC
GGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTAT
TAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCCGG
CGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCATGGCAGGGGTT
CAAATCCCCTCCGCCGGACCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTT
TTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGGGTCCTAGCGGAGTGTATA
CTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGC
AGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATA
TATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCG
AGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAA
GATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGC
TCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCG
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTC
GTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGG
CCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCT
CCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTT

ATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCA
CTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGC
GCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAA
GCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACC
GCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCA
AAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTC
AGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATA
AGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGCAATTTATCTC
TTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCAT
GTTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAA
TTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTG
CATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCC
AGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCAC
CGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGC
AGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGT
CTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCAGCCC
GGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACC
AGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAA
ACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGAT
TGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAG
AACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAG
ATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTG
ATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGG
CAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAG
CCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCG
ACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGC
GCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTG
GAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCA
CGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGC
GTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGAT
AAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTC
ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGT
TTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCC
TGCATTAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATT
GTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGACTAAGCCCAT
CGTTTTTAGTGGCGCACAGCCCTCAGGTGAATTGACCATTGGTAACTACATGG
GTGCGCTGCGTCAGTGGGTAAACATGCAGGATGACTACCATTGCATTTACTGT
ATCGTTGACCAACACGCGATCACCGTGCGCCAGGATGCACAGAAGCTGCGTA
AAGCGACGCTGGATACGCTGGCCTTGTATCTGGCTTGTGGTATCGATCCTGAG
AAAAGCACCATTTTTGTTCAGTCCCACGTGCCGGAACATGCACAGTTAGGCT
GGGCACTGAACTGCTATACCTACTTCGGCGAACTGAGTCGCATGACGCAGTT
TAAAGATAAATCTGCGCGTTATGCCGAGAACATCAACGCTGGTCTGTTTGACT
ATCCGGTGCTGATGGCAGCGGACATCCTGCTGTATCAAACTAATCTGGTACC
GGTGGGTGAAGACCAGAAACAGCACCTCGAACTGAGCCGCGATATTGCCCA
GCGTTTCAACGCGCTGTATGGCGAGATCTTTAAGGTGCCGGAGCCGTTTATTC

CGAAATCTGGCGCGCGCGTAATGTCGCTGCTGGAGCCGACCAAGAAGATGTC
CAAGTCTGACGATAATCGCAATAACGTTATCGGCCTGCTGGAAGATCCGAAA
TCGGTAGTGAAGAAAATCAAACGTGCGGTCACTGACTCCGACGAGCCGCCGG
TAGTTCGCTACGATGTGCAGAACAAAGCGGGCGTTTCCAACCTGTTGGATAT
CCTTTCAGCGGTAACGGGCCAGAGCATCCCAGAACTGGAAAAACAGTTCGAA
GGCAAGATGTATGGTCATCTGAAAGGTGAAGTGGCTGATGCCGTTTCCGGTA
TGCTGACTGAATTGCAGGAACGCTATCACCGTTTCCGCAACGATGAAGCCTTC
CTGCAACAGGTGATGAAAGATGGCGCGGAAAAAGCCAGCGCGCACGCTTCC
CGTACGCTAAAAGCGGTGTACGAAGCGATTGGTTTTGTGGCGAAGCCGTAAG
CGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGA
TTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACA
GAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAC
TCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGA
GAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGAC
TGGGCCTTGTTTGTGAGCTCCCGGTCATCAATCATCCCCATAATCCTTGTTAG
CCTGCAGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGT
GGAGGTAATAATTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCT
GACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCG
TATCTGCGCAGTAAGATGCGCCCCGCATTTAGGGGCGTAGTTCAATTGGTAG
AGCACCGGTCTTCAAAACCGGGTGTTGGGAGTTCGAGTCTCTCCGCCCCTGCC
AAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCT
CAGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTAT
TCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAAT
TACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCT
GCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGG
CATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGG
CGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCAC
CCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAA
TAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAA
ACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGT
TTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGC
TCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGC
AAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCT
TTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGC
AACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAA
CGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAA
ATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGA
AAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGC
CCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGT
GATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCC
AACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTG
TTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACG

pET29b-eSRTA
Purple=eSRTA
Red=T7

Orange=LacI
Blue=LacI promoter

CACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCG
ATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCC
GGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGA
CCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTA
GGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAA
TGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACC
GAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGC
TTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCC
CGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTG
GGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGC
GGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGC
AAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAA
ATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGA
CAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCT
GACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGA
GTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGG
AAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGC
GGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCA
ACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTC
CACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGC
GGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTC
CGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCC
ATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCA
TTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCC
GCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGAC
GCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTG
GTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGG
GAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACG
CCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAG
CGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACC
GCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCT
GGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGC
GCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGC
CCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCC
GCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCAC
GCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAAC
GTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCA
TGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGC
TCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCC
GTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAA
CAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTC
ATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATAT
AGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCC

GGCGTAGAGGATCGAGATCGATCTCGATCCCGCGAAATTAATACGACTCACT
ATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAA
CTTTAAGAAGGAGATATACATATGCAAGCTAAACCTCAAATTCCGAAAGATA
AATCAAAAGTGGCAGGCTATATTGAAATTCCAGATGCTGATATTAAAGAACC
AGTATATCCAGGACCAGCAACACGCGAACAATTAAATAGAGGTGTAAGCTTT
GCAGAAGAAAATGAATCACTAGATGATCAAAATATTTCAATTGCAGGACACA
CTTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAAAAAA
GGTAGTATGGTGTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGA
CAAGTATAAGAAACGTTAAGCCAACAGCTGTAGAAGTTCTAGATGAACAAAA
AGGTAAAGATAAACAATTAACATTAATTACTTGTGATGATTACAATGAAGAG
ACAGGCGTTTGGGAAACACGTAAAATCTTTGTAGCTACAGAAGTCAAACTCG
AGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGG
AAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGG
GGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCC
GGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGT
GGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCT
CCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAA
GCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCT
CGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCC
TGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGG
ACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTG
ATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGT
GGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAAT
ACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGC
ATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTG
AAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGG
ATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACA
ACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCAT
GAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCA
GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCA
ACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT
CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGA
ACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAAT
ACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATC
AGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGC
CAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCC
ATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTG
TCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCA
GCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAAT
ATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTG
TTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCC
CGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT
GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGA
TCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
ATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA

CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTT
ACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCC
CAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTA
TGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTA
AGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAAC
GCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG
ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAAC
GCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTT
CCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC
TGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGA
GGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGT
ATTTCACACCGCAATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAG
TTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCC
CGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGG
CATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAG
GTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCA
GCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTC
GTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGT
TAAGGGCGGTTTTTTCCTGTTTGGT

pUAS-NanoLuc

Green=NanoLuciferase

AACTGGCCGGTACCTGAGCTCGCTAGCCTCGAGGATATCAAGATCTGGCCTC
GGCGGCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGT
ACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGC
GGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGTATAAATAGA
GGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACG
TCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAAC
AATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAACCAGCAAC
CAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATA
CAAGAAGAGAACTCTGAATAGGGAATTGGGAATTCGTTAACAGATCCGATAT
CCGCCACCATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGAC
AGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTG
TTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGG
TGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTG
AGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTG
TGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGAC
GGGGTTACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCG
CCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAA
CAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGA
GTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGT
AATAATTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGATA
CATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTT

ATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAAT
AAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGG
AGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAAT
CGATAAGGATCCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTC
CTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCT
TTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACT
CAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGA
ACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGT
TGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGA
CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
TTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACC
GGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTC
ACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTG
TGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGT
CTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAA
GTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATC
AAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAA
TCTAAAGTATATATGAGTAAACTTGGTCTGACAGCGGCCGCAAATGCTAAAC
CACTGCAGTGGTTACCAGTGCTTGATCAGTGAGGCACCGATCTCAGCGATCT
GCCTATTTCGTTCGTCCATAGTGGCCTGACTCCCCGTCGTGTAGATCACTACG
ATTCGTGAGGGCTTACCATCAGGCCCCAGCGCAGCAATGATGCCGCGAGAGC
CGCGTTCACCGGCCCCCGATTTGTCAGCAATGAACCAGCCAGCAGGGAGGGC
CGAGCGAAGAAGTGGTCCTGCTACTTTGTCCGCCTCCATCCAGTCTATGAGCT
GCTGTCGTGATGCTAGAGTAAGAAGTTCGCCAGTGAGTAGTTTCCGAAGAGT
TGTGGCCATTGCTACTGGCATCGTGGTATCACGCTCGTCGTTCGGTATGGCTT
CGTTCAACTCTGGTTCCCAGCGGTCAAGCCGGGTCACATGATCACCCATATTA
TGAAGAAATGCAGTCAGCTCCTTAGGGCCTCCGATCGTTGTCAGAAGTAAGT
TGGCCGCGGTGTTGTCGCTCATGGTAATGGCAGCACTACACAATTCTCTTACC
GTCATGCCATCCGTAAGATGCTTTTCCGTGACCGGCGAGTACTCAACCAAGTC
GTTTTGTGAGTAGTGTATACGGCGACCAAGCTGCTCTTGCCCGGCGTCTATAC
GGGACAACACCGCGCCACATAGCAGTACTTTGAAAGTGCTCATCATCGGGAA
TCGTTCTTCGGGGCGGAAAGACTCAAGGATCTTGCCGCTATTGAGATCCAGTT
CGATATAGCCCACTCTTGCACCCAGTTGATCTTCAGCATCTTTTACTTTCACCA
GCGTTTCGGGGTGTGCAAAAACAGGCAAGCAAAATGCCGCAAAGAAGGGAA
TGAGTGCGACACGAAAATGTTGGATGCTCATACTCGTCCTTTTTCAATATTAT
TGAAGCATTTATCAGGGTTACTAGTACGTCTCTCAAGGATAAGTAAGTAATAT
TAAGGTACGGGAGGTATTGGACAGGCCGCAATAAAATATCTTTATTTTCATTA
CATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAACATACGCTCTCCAT
CAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCA
AGTGCAGGTGCCAGAACATTTCTCTGGCCT

PENTR4-Halotag

Green=Halotag
Red=TEV site

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCTAGCATGG
ATCTCGGGGACGTCTAACTACTAAGCGAGAGTAGGGAACTGCCAGGCATCAA
ATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATCTGTTGTTT
GTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAAC
GTTGTGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACT
GCCAGGCATCAAACTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGT
TTCTACAAACTCTTCCTGTTAGTTAGTTACTTAAGCTCGGGCCCCAAATAATG
ATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATG
CTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTCCACCATGGCAGAA
ATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGGAAGTCCTGGGCGAGCG
CATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCCTGC
ACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCA
CCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACA
AACCAGACCTGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTC
ATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCACGACTGGGGCT
CCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGTCAAAGGTAT
TGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAAT
TTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGCT
GATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCC
GCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCC
TGTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGT
GAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGGACTGGCTGCACC
AGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCA
CCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGG
ACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGG
CAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCCGGCGAGCCAACC
ACTGAGGATCTGTACTTTCAGAGCGATAACGCGAATTCAGTCGACTGGATCC
GGTACCGAATTCGCGGCCGCACTCGAGATATCTAGACCCAGCTTTCTTGTACA
AAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGT
CACTATCAGTCAAAATAAAATCATTATTTGCCATCCAGCTGCAGCTCTGGCCC
GTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATG
AACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGC
CATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTG
ATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGAC
AATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATG
GCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTG
GCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTG
GTGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGT
ATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTG
TTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGAT

CGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGA
TGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGG
AAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGG
TGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTA
TTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCT
ATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAA
AATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTC
GATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACATTATTCAGATTG
GGCCCCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCT
TGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACC
GCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA
AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTA
GCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC
GCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCT
TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGC
TGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCG
AACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGG
GAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCG
CACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGG
TTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCG
GAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGTT

pET22b-T5-NanoLuc-Halotag

Green= Nanoluc
Orange=GlyGlySer overlap
Blue=Halotag
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTT
CGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCT
AAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACC
CCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTAT
AAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAA
AAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCAC
TTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTC
AAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTT
TTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGT
AAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT
CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAAT
GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACG
CCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGT
TGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA

GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTAC
TTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAT
GGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACG
TTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATT
AATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC
CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTC
TCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTA
GTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAG
ATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAG
TTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGA
TCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAG
GTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT
CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC
GAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAG
CGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTA
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTAC
GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC
CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCT
CCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAG
GCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCT
GCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTG
GCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTG
ATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAA
ACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTA
CTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAG
AGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTG
TTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGT
GCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAG
ACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCA
CGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCC
AGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGC
CGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGA

CCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGC
GACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGT
CATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACT
GGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGA
GCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAG
CTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACG
CTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGA
TATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCA
CCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCT
GATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTG
CATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTA
TCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGA
CCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGG
AACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA
TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCG
CTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCA
CCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGT
GCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGC
CAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTT
CCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGG
GAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTA
CTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCC
ATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTC
CCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
GAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAG
TCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATG
AGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGG
CGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGC
GTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGG
AATTGTGAGCGGATAACAATTCCCCTCTAGAGTTTGACAGCATTATCATCGAT
CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAAT
AGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAG
AAATTACATATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGA
CAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTT
GTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCG
GTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCT
GAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCT
GTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGA
CGGGGTTACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATC
GCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCA
ACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCG

AGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCG
GGTGGCAGCGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGG
AAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCAC
CCCTGTGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACA
TCATCCCGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGT
ATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGACCACGTCC
GCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTC
ATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAG
AGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGG
GACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCG
ACGTCGGCCGCAAGCTGATCATCGATCAGAACGTTTTTATCGAGGGTACGCT
GCCGATGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGC
GAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTTCCCAAACG
AGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATA
CATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCC
CAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCC
TAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGAC
AACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGA
TTTCCGGCCACCACCACCACCACCACTAAAAGCTTAATTAGCTGAGCTTGGA
CTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAG
AACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTG
GCGGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTA
ACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACT
AGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAG
GAGGAACTATATCCGGAT

pET22b-T5-sfGFP-Halotag
Green=sfGFP
Orange=GlyGlySer linker
Blue=Halotag

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTT
CGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCT
AAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACC
CCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTAT
AAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAA
AAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCAC
TTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTC
AAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTT
TTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGT
AAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT
CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAAT

GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACG
CCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGT
TGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA
GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTAC
TTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAT
GGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACG
TTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATT
AATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC
CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTC
TCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTA
GTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAG
ATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAG
TTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGA
TCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAG
GTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT
CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC
GAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAG
CGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTA
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTAC
GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC
CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCT
CCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAG
GCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCT
GCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTG
GCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTG
ATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAA
ACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTA
CTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAG
AGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTG
TTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGT
GCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAG
ACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCA

CGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCC
AGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGC
CGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGA
CCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGC
GACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGT
CATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACT
GGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGA
GCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAG
CTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACG
CTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGA
TATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCA
CCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCT
GATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTG
CATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTA
TCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGA
CCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGG
AACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA
TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCG
CTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCA
CCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGT
GCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGC
CAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTT
CCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGG
GAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTA
CTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCC
ATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTC
CCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
GAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAG
TCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATG
AGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGG
CGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGC
GTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGG
AATTGTGAGCGGATAACAATTCCCCTCTAGAGTTTGACAGCATTATCATCGAT
CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAAT
AGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAG
AAATTACATATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTC
TTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAG
GGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTA
CTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGT
GTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAA
GAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGAT

GACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTG
TTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTT
GGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAG
ACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGA
AGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGC
GATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCT
TTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTA
ACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCGGTG
GCAGCGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGGAAGTC
CTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTG
TGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATC
CCGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGG
CAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGACCACGTCCGCTTCA
TGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCAC
GACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCG
TCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAA
TGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCG
GCCGCAAGCTGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGAT
GGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCG
TTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCC
AATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGGAC
TGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGT
TCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGC
AAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGG
ACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCCGG
CCACCACCACCACCACCACTAAAAGCTTAATTAGCTGAGCTTGGACTCCTGTT
GATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCG
GTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGGCGGC
CGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCC
CGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAAC
CCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAAC
TATATCCGGAT

pET22b-T5-sfGFP-AALPETG-6xHIS

Purple=AALPETG-6xHIS
Orange=T7 Promoter
Green=sfGFP
Red=T5/lac O/lac O

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTT
CGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCT
AAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACC
CCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA

GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTAT
AAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAA
AAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCAC
TTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTC
AAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTT
TTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGT
AAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT
CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAAT
GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACG
CCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGT
TGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA
GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTAC
TTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAT
GGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACG
TTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATT
AATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC
CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTC
TCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTA
GTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAG
ATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAG
TTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGA
TCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAG
GTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT
CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC
GAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAG
CGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTA
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTAC
GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC
CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCT
CCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAG

GCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCT
GCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTG
GCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTG
ATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAA
ACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTA
CTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAG
AGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTG
TTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGT
GCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAG
ACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCA
CGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCC
AGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGC
CGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGA
CCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGC
GACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGT
CATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACT
GGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGA
GCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAG
CTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACG
CTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGA
TATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCA
CCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCT
GATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTG
CATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTA
TCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGA
CCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGG
AACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA
TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCG
CTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCA
CCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGT
GCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGC
CAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTT
CCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGG
GAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTA
CTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCC
ATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTC
CCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
GAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAG
TCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATG
AGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGG
CGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGC

GTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGG
AATTGTGAGCGGATAACAATTCCCCTCTAGAGTTTGACAGCATTATCATCGAT
CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAAT
AGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAG
AAATTACATATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTC
TTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAG
GGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTA
CTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGT
GTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAA
GAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGAT
GACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTG
TTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTT
GGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAG
ACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGA
AGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGC
GATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCT
TTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTA
ACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCGCCG
CTCTGCCGGAGACTGGTCATCACCACCATCACCATTAATAAAAGCTTAATTAG
CTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTG
GATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCC
AAGCTAGCTTGGCGGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGA
TCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCT
GAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT
TTTTGCTGAAAGGAGGAACTATATCCGGAT

pET22b-T5-sfGFP

Orange=T7 Promoter
Green=sfGFP
Red=T5/lac O/lac O

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTT
CGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCT
AAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACC
CCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTAT
AAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAA
AAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCAC
TTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTC
AAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTT
TTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGT

AAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT
CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAAT
GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACG
CCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGT
TGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA
GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTAC
TTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAT
GGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACG
TTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATT
AATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC
CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTC
TCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTA
GTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAG
ATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAG
TTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGA
TCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAG
GTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT
CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC
GAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAG
CGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTA
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTAC
GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCC
CTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCT
CCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAG
GCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCT
GCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTG
GCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTG
ATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAA
ACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTA
CTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAG
AGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTG
TTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGT

GCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAG
ACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCA
CGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCC
AGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGC
CGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGA
CCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGC
GACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGT
CATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACT
GGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGA
GCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAG
CTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACG
CTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGA
TATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCA
CCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCT
GATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTG
CATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTA
TCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGA
CCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGG
AACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA
TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCG
CTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCA
CCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGT
GCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGC
CAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTT
CCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGG
GAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTA
CTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCC
ATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTC
CCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
GAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAG
TCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATG
AGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGG
CGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGC
GTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGG
AATTGTGAGCGGATAACAATTCCCCTCTAGAGTTTGACAGCATTATCATCGAT
CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAAT
AGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAG
AAATTACATATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTC
TTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAG
GGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTA
CTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGT

GTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAA
GAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGAT
GACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTG
TTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTT
GGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAG
ACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGA
AGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGC
GATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCT
TTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTA
ACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCCACC
ACCACCACCACCACTAAAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAG
ATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTGC
CGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGGCGGCCGCA
CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAA
AGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCT
TGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATA
TCCGGAT

Appendix 2
Oligonucleotide Primers

Chapter 2 Oligos

Primers	Sequence
RC2-HindIII_F:	CGTCAGACGCGGAAGCTTCGATCAAC
RC2-SbfI_R:	GATCCTGCAGGTATTTGCGGCCGATG
Q263TAG-F:	CAAACAAATTTCCAGCTAGTCAGGAGCCTCGAACGACAA TCAC
Q263TAG-R:	CGAGGCTCCTGACTAGCTGGAAATTTGTTTGTAGAGGTG GTTGTTG
H271TAG-F:	CTCGAACGACAATTAGTACTTTGGCTACAGCACCCCTTGG G
H271TAG-R:	CTGTAGCCAAAGTACTAATTGTCGTTCGAGGCTCCTGATT GG
T330TAG-F:	GAATGACGGTACGTAGACGATTGCCAATAACCTTACCAG CACGG
T330TAG-R:	GGCAATCGTCTACGTACCGTCATTCTGCGTGACCTCTTTG
T450TAG-F:	GAGCAGAACAAACTAGCCAAGTGGAACCACCACGCAGTC AAG
T450TAG-R:	GTGGTTCCACTTGGCTAGTTTGTTCTGCTCAAGTAATACA GGTACTGGTC
T451TAG-F:	GAGCAGAACAAACACTTAGAGTGGAACCACCACGCAGTC AAG
P451TAG-R:	GTGGTTCCACTCTAAGTGTTTGTTCTGCTCAAGTAATACA GGTACTGG
S452TAG-F:	GAACAAACACTCCATAGGGAACCACCACGCAGTCAAGGC
S452TAG-R:	GTGGTGGTTCCCTATGGAGTGTTTGTTCTGCTCAAGTAAT ACAGGTAC
T455TAG-F:	GTGGAACCTAGACGCAGTCAAGGCTTCAGTTTTCTCAG
T455TAG-R:	GAAGCCTTGACTGCGTCTAGGTTGCACTTGGAGTGTTTGT TCTGCTC
T456TAG-F:	GTGGAACCACCTAGCAGTCAAGGCTTCAGTTTTCTCAGGC CG
T456TAG-R:	GAAGCCTTGACTGCTAGGTGGTTCCACTTGGAGTGTTTGT TC
Q457TAG-F:	GTGGAACCACCACGTAGTCAAGGCTTCAGTTTTCTCAGGC CG
Q457TAG-R:	GAAGCCTTGACTACGTGGTGGTTCCACTTGGAGTGTTTG
S458TAG-F:	CACCACGCAGTAGAGGCTTCAGTTTTCTCAGGCCGGAG
S458TAG-R:	CTGAAGCCTCTACTGCGTGGTGGTTCCACTTGGAGTG
R459TAG-F:	CACGCAGTCATAGCTTCAGTTTTCTCAGGCCGGAGCG
R459TAG-R:	GAGAAAACTGAAGCTATGACTGCGTGGTGGTTCCACTTG G

D469TAG-F:	CGGAGCGAGTTAGATTCGGGACCAGTCTAGGAACTGGC
D469TAG-R:	GTCCCGAATCTAACTCGCTCCGGCCTGAGAAAACTG
I494TAG-F:	CGAGTATCAAAGACTAGTGCGGATAACAACAACAGTGAA TACTCGTGG
I494TAG-R:	GTTGTTGTTATCCGCACTAGTCTTTGATACTCGCTGCTGG CG
T503TAG-F:	GAATACTCGTGGTAGGGAGCTACCAAGTACCACCTCAAT GGC
T503TAG-R:	CTTGGTAGCTCCCTACCACGAGTATTCACTGTTGTTGTTA TCCGC
H509TAG-F:	GCTACCAAGTACTAGCTCAATGGCAGAGACTCTCTGGTG AATCCGG
H509TAG-R:	CTGCCATTGAGCTAGTACTTGGTAGCTCCAGTCCACGAGT ATTC
E530TAG-F:	GCCACAAGGACGATTAGGAAAAGTTTTTTCCTCAGACGC CCCTTCTC
E530TAG-R:	GAGGAAAAAACTTTTCCTAATCGTCCTTGTGGCTTGCCAT GG
T659TAG-F:	GCGAATCCTTCGTAGACCTTCAGTGCGGCAAAGTTTGCTT C
T659TAG-R:	GCCGCACTGAAGGTCTACGAAGGCTTCGCAGGTACCGGG
K706TAG-F:	CCAACTACAACTAGTCTGTTAATGTGGACTTTACTGTGG
K706TAG-R:	GTCCACATTAACAGACTAGTTGTAGTTGGAAGTGTACTG AATTTC
N717TAG-F:	CTGTGGACACTTAGGGCGTGTATTCAGAGC
N717TAG-R:	GAATACACGCCCTAAGTGTCCACAGTAAAGTCCACATTA AC
585A/588A-F:	CCTCCAGGCCGGCAACGCCCAAGCAGCTACCGCAGATGT CAACACAC
585A/588A-R:	GCTTGGGCGTTGCCGGCCTGGAGGTTGGTAGATACAGAA CCATACTGCTC

Chapter 3 Oligos

RC2-int-HindIII-F:	CGTCAGACGCGGAAGCTTCGATCAAC
RC2-SbfI-R:	GATCCTGCAGGTATTTGCGGCCGATG
VP1-NheI-F:	TTATTTAGCTAGCATGGCTGCCGATGGTTAT CTTCCAGATTGGC
VP2-NheI-T138M-F:	TTATTTAGCTAGCATGGCTCCGGGAAAAAA GAGGCCGG
RC2-AvrII-R:	TTATTTACCTAGGGCTGTAGTTAATGATTAA CCCGCCATGCTACTTATC
VP1-M1L-F:	GATTTAAATCAGGTCTCGCTGCCGATGGTT ATCTTCCAGATTGGC
VP1-M1L-R:	CCATCGGCAGCGAGACCTGATTTAAATCAT

	TTATTGTTCAAAGATGCAGTCATCC
VP2-T138T-F:	GAGGAACCTGTTAAGACCGCTCCGGGAAAA AAGAGGCCGGTAGAG
VP2-T138T-R:	CCCGGAGCGGTCTTAACAGGTTCCTCAACC AGGCCC
VP3- L202L/M203L/211L-F:	CTAATACCCTCGCTACAGGCAGTGGCGCAC CACTCGCAGACAATAACGAGGG
VP3- T197T/L202L/M203L -R:	GCGCCACTGCCTGTAGCGAGGGTATTAGTT CCGAGACCAGAGGGGGCTGCTG
T454TAG-F:	CAAACACTCCAAGTGGATAGACCACGCA
T454TAG-R:	CTGCGTGGTCTATCCACTTGGAGTGTTTTG
R585A/R588A-F:	CCTCCAGGCCGGCAACGCCCAAGCAGCTAC CGCAGATGTCAACACAC
R585A/R588A-R:	GCTTGGGCGTTGCCGGCCTGGAGGTTGGTA GATACAGAACCATACTGCTC

Chapter 4 Oligos

T454Cys F:	CTCCAAGTGGATGCACCACGCAGTCAAGGC TTCAG
T454Cys R:	CTGCGTGGTGCATCCACTTGGAGTGTTTGTT CTGCTCAAG
RC2-int-HindIII-F:	CGTCAGACGCGGAAGCTTCGATCAAC
RC2-SbfI-R:	GATCCTGCAGGTATTTGCGGCCGATG
	GTTTGTCTGCCGTGATGCATACATTGTGTGA
sfGFP_151C_F:	GTTAAAGTTGTACTCGAGTTTGTGTCC
	GTTTGTCTGCCGTGATGCATACATTGTGTGA
sfGFP_151C_R:	GTTAAAGTTGTACTCGAGTTTGTGTCC
	GAGAAATTACATATGAGCAAAGGAGAAGA
sfGFP-NdeI-F:	ACTTTTCACTGGAGTTG
	AATAATAATTAAGCTTTTAGTGGTGGTGGTG
sfGFP-HindIII-R:	GTGG

Chapter 5 Oligos

RC2-HindIII_F:	CGTCAGACGCGGAAGCTTCGATCAAC
RC2-SbfI_R:	GATCCTGCAGGTATTTGCGGCCGATG
	CCAACCTCCAGAAGAGGCAACAGACAAGCAGCTACCGCA
R585K-F:	GATG
	CTTGTCTGTTGCCCTTCTGGAGGTTGGTAGATACAGAACCA
R585K-R:	TACTGCTC
588K-F:	GAGGCAACAAGCAAGCAGCTACCGCAGATGTCAACACAC
	GCGGTAGCTGCTTGCTTGTTGCCTCTCTGGAGGTTGGTACA
588K-R:	TACAG
	CCAACCTCCAGTAGAGGCAACAGACAAGCAGCTACCGCAG
R585TAG-F:	ATG

R585TAG-R:	CTTGTCTGTTGCCCTACTGGAGGTTGGTAGATACAGAACCA TACTGCTC
R588TAG-F:	CTGCGGTAGCTGCTTGCTAGTTGCCTCTCTGGAGGTTGGTA GATACAG
R588TAG-R:	CCAGAGAGGCAACTAGCAAGCAGCTACCGCAGATGTCAAC ACAC
585K/588K R:	CCAACCTCCAGAAGGGCAACAAGCAAGCAGCTACCGCAG ATGTCAACAC
585K/588K F:	GTAGCTGCTTGCTTGTTGCCCTTCTGGAGGTTGGTAGATAC AGAACCATACTGCTC

Chapter 6 Oligos

mmPylRS Terminal primer, NheI F:	AATTAAGCTAGCATGGATAAAAAG CCTCTGAACACTCTGATTTCTGCG
mmPylRS Terminal primer, XhoI, R:	ATTTAACTCGAGTTACAGGTTAGT AGAAATACCATTGTAATAGGACTC GGAAC
mmPylRS WT Y306A F:	GGCCCCGAACCTGGCCAACTATCT GCGTAAACTGGACCGTGCCC
mmPylRS WT Y306A R:	CGCAGATAGTTGGCCAGGTTCGGG GCCAGCATCGG
mmPylRS WT Y384 F:	GCTGTATGGTGTTCGGCGATACGC TGGATGTTATGCACG
mmPylRS WT Y384 R:	CGTATCGCCGAACACCATACAGCT GTCACCAACAATTTTGAAATC
T251TAG_F:	CCCTGCCCTAGTACAACAACCACC TCTACAAACAAATTTCCAGCCAAT CAGG
T251TAG_R:	GTTGTTGTACTAGGGCAGGGCCCA GGTTCGGGTGCTGG
S468TAG_F:	CCGGAGCGTAGGACATTCGGGACC AGTCTAGGAACTGGCTTCCTG
S468TAG_R:	CCGAATGTCCTACGCTCCGGCCTG AGAAAACTGAAGCCTTGACTGCG
T671TAG_F:	GCTTCCTTCATCTAGCAGTACTCCA CGGGACAGGTCAGCGTGGAGATCG AG
T671TAG_R:	CCGTGGAGTACTGCTAGATGAAGG AAGCAAACTTTGCCGCACTGAAGG TGG
K665TAG_F:	CAGTGCGGCATAGTTTGCTTCCTTC ATCACACAGTACTCCACGGGACAG
K665TAG_R:	GAAGGAAGCAAACTATGCCGCACT GAAGGTGGTCGAAGGATTCGCAGG
S662TAG_F:	GACCACCTTCTAGGCGGCAAAGTT

S662TAG_R:	TGCTTCCTTCATCACACAGTACTCCACG CTTTGCCGCCTAGAAGGTGGTCGAAGGATTCGCAGGTACCGGGG
K716TAG_F:	GACTTTACTGTGGACTAGAATGGCGTGTATTCAGAGCCTCGCCCCATTGGC
K716TAG_R:	GAATACACGCCATTCTAGTCCACAGTAAAGTCCACATTAACAGACTTGTTGTAGTTGG
BamHI_pEvol_FAB F:	CCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGC
BamHI_pEvol FAB R:	TCTAGAGGATCCCCTGCAGGCTTAACACTC
SbfI_pEvol_FAB F:	GTTAAGCCTGCAGGGGATCCTCTAGAGGTTG
SalI_pEvol_FAB R:	AAATTCGGTCGACAAGGCCCAGTCTTTCGACTG
Her2_K169TAG_F:	GCAGGACAGCTAGGACAGCACCTACAGCCTCAGCAGCACCCTG
Her2-_K169TAG_R:	GGTGCTGTCCTAGCTGTCCGTCTCTGTGACACTCTCCTGGGAGTTACC
Her2_S202_F:	CAGGGCCTGTAGTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTG
Her2_S202_R:	GACGGGCGACTACAGGCCCTGATGGGTCACTTCGCAGGCGTAG
Her2_K169TGA_F:	GCAGGACAGCTGAGACAGCACCTACAGCCTCAGCAGCACCCTG
Her2_K169TGA_R:	GGTGCTGTCTCAGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACC
Pevol_Her2_F_AvrII:	AATATTCCTAGGCGATATAAGTTGTAATTCTCATGTTTGACAGC
pEvol_Her2_R_EcoRI:	ATTTAAGAATTCGTGCCCTTAAACGCCTGGTTGCTAC
aHer2-HC-S163TAG_F:	CGTGGAACTAGGGCGCCCTGACCAGCGGCG
aHer2-HC-S163TAG_R:	CAGGGCGCCCTAGTTCCACGACACCGTCACCGGTTCG
aHer2-HC-T198TAG_F:	CAGCTTGGGCTAGCAGACCTACATCTGCAACGTGAATCACAAGCCC
aHer2-HC-T198TAG_R:	GATGTAGGTCTGCTAGCCCAAGCTGCTAGAGGGCACAG
aHer2-HC-K213TAG_F:	CCAGCAACACCTAGGTGGACAAGAAAGTTGAGCCCAAATCTTGTGAC
aHer2-HC-K213TAG_R:	CTTTCTTGTCCACCTAGGTGTTGCTGGGCTTGTGATTCACG

Ec.WRS1 Bsu631_F:	AATTAACCTGAGGGTGGTTCGCATCCTCGGTTTTCTGGAAGGCGAGCATCGTTTG
Ec.WRS1 Bsu631_R:	TTATAACCTCAGGCGGAGTATACGGACCGCGGCCGCAAATACCTGCAGG
T456TGA_F:	GTGGAACCACCTGACAGTCAAGGCTTCAGTTTTCTCAGGCCG
T456TGA_R:	GAAGCCTTGACTGTCAGGTGGTTCCACTTGGAGTGTTTGTTC
EcWRS1_AvrII_F:	AATTAACCTAGGGTGGTTCGCATCCTCGGTTTTCTGGAAGGCGAGCATCGTTTG
Fab_LPETGHHHHHHNot1:	TTTAATGCGGCCGCTTATTAATGGTGATGGTGGTGATGACCAGTCTCCGGCAGAGCGGCTGTGTGAGTTTTGTCACAAGATTTGGGCTCAAC
GFP-LPETG-R:	TAATTAAAGCTTTTATTAATGGTGATGGTGGTGATGACCAGTCTCCGGCAGAGCGGCGGATCCTTTGTAGAGCTCATCCATGCCATGTG
Halotag_NdeI_R:	CGCGAATTCAGTCGACTGGATCCGGCTGCCACCGCTGCCACCCATATGTTAATT
sfGFP_Halotag overlap_F:	TTTCTGCGCTGCCACCGGATCCTTTGTAGAGCTCATCCATGCC
sfGFP_Halotag overlap_R:	GGATCCGGTGGCAGCGCAGAAATCGGTACTGGCTTTCCATTC
Halotag_Histag_R:	AATAATAATTAAGCTTTTAGTGGTGGTGGTGGTGGTGGCCGGAAATCTCGAGCGTCGA
NdeI-NanoLuc_F:	TTAATTCATATGGTCTTCACACTCGAAGATTTCGTTGGG
NanoLuc_glyglyser_Halotag_overlap_F:	GATTTCTGCGCTGCCACCCGCCAGAATGCGTTCGCACAG
NanoLuc_glyglyser_Halotag_overlap_R:	CTGGCGGGTGGCAGCGCAGAAATCGGTACTGGCTTTCCATTCGAC
HindIII_6xHis_Halotag_R:	AATAATAAGCTTTTAGTGGTGGTGGTGGTGGTGGCCGGAAATCTCGAGCGTCGAC