

Enhancing Platforms at the Interface of Viruses and Directed Evolution

Samantha D. Levinson

A dissertation submitted to the Department of Chemistry
in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

May 2021

ENHANCING PLATFORMS AT THE INTERFACE OF VIRUSES AND DIRECTED EVOLUTION

by

Samantha D. Levinson

Advisor: Professor Abhishek Chatterjee, Ph.D.

Abstract

Directed evolution is a powerful technique to expand chemical space in biological systems. In particular, this method has been used to develop cellular machinery to enable genetic code expansion (GCE), the incorporation of unnatural amino acids (UAAs) into proteins during the translation process. GCE relies on evolving an aminoacyl tRNA synthetase (aaRS) and tRNA pair from a different domain of life to incorporate a UAA into proteins in their new host, as these evolutionarily distant pairs are less likely to be cross-reactive with host pairs. The aaRS and tRNA must meet a number of conditions to be useful for GCE: the pair must be orthogonal (non-cross-reactive) to the host's native aaRS/tRNA pairs in order to ensure site-specific UAA incorporation; the aaRS must have an active site suited to accept the shape of the UAA; and the tRNA must cooperate with the host ribosome, elongation and release factors, and other translational machinery to efficiently incorporate the UAA into the protein.

Numerous aaRS/tRNA pairs have been evolved to allow incorporation of diverse UAAs in bacteria due to the tractable nature of these organisms for directed evolution experiments. While an aaRS evolved in bacteria to charge a novel UAA can be used in eukaryotes, tRNAs cannot be evolved for GCE in bacteria and then used in eukaryotes

because they will not have evolved in the presence of the correct translational machinery. It is necessary to evolve tRNAs directly in their host cells. Unfortunately for researchers working on GCE in mammalian cells, it is difficult to perform directed evolution on small gene products in these hosts. Transformation efficiency in mammalian cells is poor, and transient transfection yields heterogeneous DNA distribution to target cells, making selection based on performance of individual library members impossible. Viruses are an ideal DNA delivery vector for mammalian cells, as production of recombinant viruses allows control over library member generation, and viruses can be delivered with exquisite copy number control. The Chatterjee lab recently developed a platform, Virus-Assisted Directed Evolution of tRNAs (VADER), using adeno-associated virus (AAV) to evolve tRNAs for GCE directly in mammalian cells.

While VADER is the first directed evolution platform that allows the evolution of small gene products in mammalian cells, its efficiency is limited by its continued reliance on transient transfection to deliver non-library DNA that is necessary for the production of rAAV. To overcome this limitation, baculovirus delivery vectors were developed to boost DNA delivery and AAV capsid production to improve virus production efficiency during selections. VADER allows the evolution of tRNAs to incorporate certain UAAs, but the technique relies on installing a UAA into the AAV capsid, which is sensitive to disruption caused by slight modifications in structure. To expand the scope of VADER to evolve tRNAs for UAAs that cannot be incorporated into the AAV capsid, an alternate selection handle (Assembly Activating Protein, or AAP) was deleted from the genome and provided *in trans* to incorporate 5-hydroxytryptophan (5HTP). Incorporating the UAA into this

flexible protein allows UAA-dependent production of AAV and expands the scope of tRNAs that can be evolved in mammalian cells.

Acknowledgements

I would like to thank my family for supporting me throughout my life, and especially over the last five years. A PhD may seem like an individual endeavor, but it is truly a group effort. Thank you to my mom, Laurie Lauchlan, for telling me at age 12 that I actually had to start doing my homework if I wanted to become a scientist and for working so hard so I could succeed. Thank you to my dad, Harry Levinson, for explaining the natural world to me in technical terms even when I was young and helping me know I could understand anything if I worked hard enough. Special thanks to my sister, Sarah Levinson, for supporting me without fail and showing me that being willing is worth it. Thank you to my Auntie Lynn and Uncle Sao for supporting me through my Master's, and to my Uncle Dave for encouraging me in my studies and in my swimming. Thank you to all of my grandparents and ancestors; I wish you could be here to read this thesis. I love you all.

Thank you to my advisor, Professor Abhishek Chatterjee, for your mentorship and guidance throughout these past five years. I would not be here if not for you.

Many thanks to my labmates. To Sarah Erickson, who blew my mind on my first day in lab by explaining how to tell what volume each pipette could hold, and who became the best labmate I could ask for and an excellent friend. To Rachel Kelemen, Xiaofu Cao, and Yunan Zheng, who taught me how to make viruses. To Rachel Huang and Lily Jewel for their technical assistance with tRNAs. And to Katie Grasso, Shu Wang, Elise Ficaretta, and Conor Loynd for their friendship.

Thank you to my friends. I would not have made it without your support. Thank you to Christine Ondris, who told me that if Britney Spears could make it through 2007, I could make it through 6 months of chemistry. It ended up being a little bit longer than 6

months, and your friendship and support got me through it. Thank you to Rachael and David Gelineau, my second family, for being there for me through all the trials and triumphs. Thank you to Olaf Haraldson for your support and for every single thing made out of stone. Thanks to Alyssa Greene and Andy Crow for friendship, baked goods, and many edits. Thank you to Professor (and friend) Cheryl Hackworth for sharing your love of nature both huge and tiny. Thank you to my coach, Craig Lewin, for being patient with me when I was in lab instead of in the water. Thank you many to the Northern Water Tribe: Fran O'Loughlin, Laurie Craigen, Amy Craigen, Chris Graefe, Karen Nazor, Rena Demeo, Natalie Lang, Stacie McHale, Elaine Howley, Jessica O'Herron, and everyone I've met through open water swimming. You all cheer me on, and I couldn't do it without you.

Thank you to Dr. Cheryl Cavanaugh for reminding me that I am not my grade.

Thank you so much to Daniela Klaz. We may have begun at the end, but your love has gotten me through the wall.

Finally, I would like to thank Professor Glenn Eldridge. This is all your fault.

Table of Contents

Acknowledgements.....	i
Table of Contents.....	iii
List of Figures.....	vi
List of Abbreviations.....	viii
Chapter 1: Introduction.....	1
1.1 Directed evolution.....	2
1.1.1 Methods to evolve gene targets.....	2
1.1.2 The state of the field of directed evolution in mammalian cells.....	3
1.2 Unnatural amino acid incorporation in proteins in mammalian cells.....	4
1.2.1 How to incorporate unnatural amino acids into proteins.....	4
1.2.2 UAA incorporation in mammalian cells.....	7
1.3 Adeno-associated virus as a tool for DNA delivery & directed evolution....	8
1.3.1 Viruses can be used to transduce mammalian cells.....	8
1.3.2 AAV as a facile tool for mammalian cell transduction.....	10
1.3.3 An AAV-based selection scheme for mammalian cells.....	11
1.4 Expanding the scope of viral vectors in the field of directed evolution.....	13
1.5 References.....	16
Chapter 2: Baculovirus vectors to amplify AAV during directed evolution.....	19
2.1 Introduction.....	20
2.1.1 Directed evolution in mammalian cells.....	20
2.1.2 Using viruses to transduce mammalian cells.....	21
2.1.3 VADER: an AAV-based directed evolution platform.....	22
2.1.4 DNA delivery by baculovirus.....	24
2.2 Results and Discussion.....	26
2.2.1 Determining the effects of addition of PEI on AAV infection.....	26
2.2.2 Comparing pAcBac vectors to original rAAV production plasmids..	28
2.2.3 Comparing pAcBac vectors to original rAAV production plasmids for stop codon suppression.....	35
2.2.4 The source of the improved AAV production.....	38
2.3 Conclusions.....	40
2.3.1 Ongoing work and future directions.....	41
2.4 Acknowledgements.....	41
2.5 Experimental Procedures.....	41
2.5.1. Strains and cell lines.....	41
2.5.2. Plasmids.....	41
2.5.3. Unnatural amino acids.....	42
2.5.4. Producing recombinant baculovirus.....	42
2.5.5. Preparing recombinant AAV.....	43
2.5.6. Determining the effects of addition of PEI on AAV infection.....	44

2.5.7. Comparing pAcBac shuttle plasmids to original rAAV production plasmids.....	44
2.5.8. Comparing transfection to infection with baculovirus.....	44
2.6 References.....	46
Chapter 3: Development of an alternate selection handle to evolve tRNAs in mammalian cells.....	48
3.1 Introduction.....	49
3.1.1 VADER as a solution for evolving small gene products.....	49
3.1.2 The limitations of VADER: the AAV capsid.....	50
3.1.3 The ideal selection handle: Assembly Activating Protein.....	52
3.1.4 A selection platform based on AAP.....	53
3.2 Results and Discussion.....	55
3.2.1 Producing infectious AAV with TAG-containing AAP.....	55
3.2.2 Determining the optimum site for a TAG mutation in AAP.....	57
3.2.3 Incorporating 5-hydroxytryptophan into AAP.....	59
3.2.4 Mock selections to determine if VADER is possible with AAP as a selection handle.....	61
3.3 Conclusion.....	63
3.3.1 Ongoing work and future directions.....	63
3.4 Acknowledgments.....	64
3.5 Experimental Procedures.....	64
3.5.1 Strains and cell lines.....	64
3.5.2 Plasmids.....	64
3.5.3 Unnatural amino acids.....	65
3.5.4 Preparing recombinant AAV.....	66
3.5.4 Producing infectious AAV with TAG-containing AAP.....	66
3.5.5 Determining the optimum site for a TAG mutation in AAP.....	66
3.5.6 Incorporating 5-hydroxytryptophan into AAP.....	67
3.5.7 Mock selections to determine if VADER is possible with AAP as a selection handle.....	67
3.6 References.....	69
Chapter 4: Development of an optimized baculovirus vector for incorporation of lysine analogs.....	70
4.1 Introduction.....	71
4.1.1 Different methods to deliver genetic code expansion machinery to mammalian cells.....	71
4.1.1a Transient transfection.....	71
4.1.1b AAV.....	72
4.1.1c Lentivirus.....	73
4.1.1d Baculovirus.....	74
4.1.2 A new baculovirus vector for the pyrrolysyl system.....	75
4.2 Results and Discussion.....	77
4.2.1 Comparing A2 PytR variant to wild type PytR.....	77
4.2.2 Designing an optimized PytR baculovirus vector.....	81

4.2.3 Incorporating other optimized synthetases.....	84
4.3 Conclusion.....	93
4.4 Acknowledgments.....	94
4.5 Materials and Methods.....	94
4.5.1. Strains and cell lines.....	94
4.5.2. Plasmids.....	94
4.5.3 Unnatural amino acids.....	97
4.5.4 Producing recombinant baculovirus.....	97
4.5.5 Comparing A2 PytR variant to wild type PytR.....	97
4.5.6 Designing an optimized PytR baculovirus vector.....	97
4.5.7 Incorporating other optimized synthetases.....	98
4.6 References.....	99
Chapter 5: A virus-based two-hybrid selection scheme to probe the mammalian interactome.....	101
5.1 Introduction.....	102
5.1.1 Two-hybrid selection.....	102
5.1.2 Yeast two-hybrid systems.....	103
5.1.3. Mammalian two-hybrid systems.....	104
5.1.4. Developing high throughput two-hybrid systems for mammalian cells.....	105
5.2 Results and Discussion.....	107
5.2.1 Producing GFP with the use of two-hybrid binding partners.....	107
5.2.2 Testing the effects of minimal promoters on GFP production with a two-hybrid system.....	110
5.2.3 Producing rAAV dependent on two-hybrid selection.....	112
5.3 Conclusion.....	114
5.3.1 Future directions.....	114
5.4. Materials and Methods.....	115
5.4.1 Strains and cell lines.....	115
5.3.2. Plasmids.....	115
5.3.3 Preparing recombinant AAV.....	116
5.3.4 Producing GFP with the use of two-hybrid binding partners.....	116
5.3.5 Testing the effects of different minimal promoters on GFP production with a two-hybrid system.....	117
5.3.6 Producing rAAV dependent on two-hybrid selection.....	117
5.5 References.....	119
Appendix I. Plasmid maps and sequences.....	120
Appendix II. Oligonucleotide primers.....	308

List of Figures

Chapter 2 Figures

Figure 2.1	Structure of polyethylenimine (PEI).....	24
Figure 2.2	The effects of PEI on transduction by rAAV.....	28
Figure 2.3	The effects of PEI on GFP production in infected cells.....	28
Figure 2.4	Cells packaging rAAV and cells infected with harvested lysate.....	29
Figure 2.5	GFP fluorescence of cells infected with lysate from cells packaging AAV.....	30
Figure 2.6	Determining the source of limited rAAV production.....	31
Figure 2.7	GFP fluorescence of cells to test the source of limited rAAV production...31	
Figure 2.8	Anti-AAV western blot to test the source of limited rAAV production.....32	
Figure 2.9	Anti-AAV western blot to demonstrate rescued rAAV production.....33	
Figure 2.10	Transfecting cells with optimized plasmids.....	33
Figure 2.11	GFP fluorescence of cells to test the optimized pAcBac shuttle plasmids..34	
Figure 2.12	Cells packaging rAAV and cells infected with rAAV-containing lysate...35	
Figure 2.13	mCherry fluorescence of cells infected with rAAV-containing lysate.....35	
Figure 2.14	Structures of BocK (1) and AzK (2).....	36
Figure 2.15	Cells packaging rAAV with either wild type or stop codon-containing capsids and cells infected with rAAV-containing lysate.....	37
Figure 2.16	Anti-AAV western to determine whether 454TAG capsid proteins are limiting.....	38
Figure 2.17	Flow cytometry analysis of cells producing rAAV-mCherry.....	39

Chapter 3 Figures

Figure 3.1	Structure of 5-hydroxytryptophan (5-HTP).....	51
Figure 3.2	Incorporation of 5HTP into the AAV2 capsid.....	51
Figure 3.3	Structure of the AAV genome.....	53
Figure 3.4	The structure of the AAPstop60 genome.....	54
Figure 3.5	Producing AAV dependent on supplemental AAP.....	56
Figure 3.6	GFP fluorescence of cells infected with AAV produced using the AAPstop60 system.....	57
Figure 3.7	Producing AAV using AAP with TAG mutations.....	58
Figure 3.8	GFP fluorescence of cells infected with AAV produced with AAP TAG variants.....	59
Figure 3.9	Infection of cells with virus produced using AAP with TGA mutations....60	
Figure 3.10	GFP fluorescence of cells infected with AAV produced with AAP TGA variants.....	61
Figure 3.11	Enrichment of WtR+mCherry over PytR+GFP in a mock selection.....	63

Chapter 4 Figures

Figure 4.1	Structure of O-methyltyrosine.....	75
Figure 4.2	Map of pAcBac3 baculovirus vector.....	76
Figure 4.3	mCherry fluorescence of cells infected with baculovirus.....	79
Figure 4.4	GFP fluorescence of cells infected with baculovirus.....	79
Figure 4.5	Comparison of GFP/mCherry ratios of A2 and wild type PytR.....	80
Figure 4.6	Comparison of GFP/mCherry ratios of HTS25 and wild type PytR.....	82
Figure 4.7	Structure of the photocaged lysine PCK.....	84
Figure 4.8	Comparison of NES-MbPylRS and wild type MbPylRS to charge PCK...	85
Figure 4.9	Comparison of GFP/mCherry ratios for NBK-RS and the wild type MbPylRS.....	85
Figure 4.10	Structure of DiazK.....	87
Figure 4.11	Comparison of DiazK-MbPylRS and wild type MbPylRS to charge DiazK.....	87
Figure 4.12	Comparison of GFP/mCherry ratios for DiazK-RS and the wild type MbPylRS.....	87
Figure 4.13	Structure of AcK.....	89
Figure 4.14	Comparison of AcK-MbPylRS and wild type MbPylRS to charge AcK.....	89
Figure 4.15	Comparison of GFP/mCherry ratios for AcK-RS and the wild type MbPylRS.....	89
Figure 4.16	Structure of cyclo-octyne lysine (SCOK).....	91
Figure 4.17	Comparison of BiotinK-MbPylRS and wild type MbPylRS to charge AzK and SCOK.....	92
Figure 4.18	Comparison of GFP/mCherry ratios for AcK-RS and the wild type MbPylRS.....	92

Chapter 5 Figures

Figure 5.1	Overview of a two-hybrid assay.....	102
Figure 5.2	Luminescence of cells testing the two-hybrid Mammalian Matchmaker Assay.....	108
Figure 5.3	GFP fluorescence of cells using the two-hybrid system.....	109
Figure 5.4	Comparison of GFP fluorescence using minimal promoters.....	111
Figure 5.5	mCherry fluorescence of cells infected with rAAV-containing lysate.....	112
Figure 5.6	Cells packaging rAAV and cells infected with rAAV-containing lysate..	113

List of Abbreviations

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

5-FOA	5-fluoroorotic acid
5HTP	5-hydroxy-L-tryptophan
A2	acceptor stem library hit 2
AAP	Assembly Activating Protein
aaRS	aminoacyl-tRNA synthetase
AAV	adeno-associated virus
AAV2	adeno-associated virus 2
AcK	N ^ε -acetyl-L-lysine
AcMNPV	<i>Autographa californica</i> multicapsid nucleopolyhedrovirus
Ad	Adenovirus
ATM	altered translational machinery
AzK	azidolysine
BocK	N ^ε -Boc-L-Lysine
CAPPIA	cell array protein-protein interaction assay
CMV	cytomegalovirus
CO ₂	carbon dioxide
CRISPR	clustered regularly interspaced short palindromic repeats
CSS	consensus splice site
DNA	deoxyribonucleic acid
DBCO	dibenzocyclooctyne
DiazK	diazirine lysine
DMEM	Dulbecco's Modified Eagle Medium
EcWRS	<i>Escherichia coli</i> tryptophanyl aminoacyl-tRNA synthetase
EcWtR	<i>E. coli</i> tryptophanyl tRNA
EcYtR	<i>E. coli</i> tyrosyl tRNA
eGFP	enhanced green fluorescent protein
GPCR	G protein-coupled receptor
GFP	green fluorescent protein
HEK-293T	human embryonic kidney 293 cells with a temperature-sensitive allele of the
SV40 T	antigen
HIV	human immunodeficiency virus
HSV	herpes simplex virus
HTS25	high-throughput sequencing hit 25
IPTG	Isopropyl β-d-1-thiogalactopyranoside
ITR	inverted terminal repeat
Kb	kilobase
LB	Luria broth
LDL-R	low-density lipoprotein receptor
MbPylRS	<i>Methanosarcina barkeri</i> pyrrolysyl aminoacyl-tRNA synthetase
MmPylR	<i>M. mazei</i> pyrrolysyl tRNA
mL	milligram

MOI	multiplicity of infection
MTH	metallothionein
NBK	norbornene lysine
NES	nuclear export sequence
ORF	open reading frame
P0	passage 0
P1	passage 1
PBS	phosphate-buffered saline
PCK	photocaged lysine
PEG	polyethylene glycol
PEI	polyethylenimine
PPI	protein-protein interaction
PytR	pyrrolysyl tRNA
rAAV	recombinant adeno-associated virus
RNA	ribonucleic acid
RC2	<i>RepCap2</i>
RFU	relative fluorescence units
SEAP	secreted alkaline phosphatase
SCOK	cyclo-octyne lysine
Sf9	<i>Spodoptera frugiperda</i> 9 cells
Sf21AE	<i>S. frugiperda</i> 21AE cells
snRNP	small nuclear ribonucleoprotein
tRNA	transfer ribonucleic acid
U	unit
UAA	unnatural amino acid
UbiC	Ubiquitin-C
VADER	Virus-Assisted Directed Evolution of tRNAs
VP	viral protein
VSV-G	vesicular stomatitis virus G glycoprotein
WRS	tryptophanyl aminoacyl-tRNA synthetase
WtR	tryptophanyl tRNA
Wt	wild-type
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
YtR	tyrosyl tRNA
μ g	microgram
μ L	microliter

Chapter 1

Introduction

1.1 Directed evolution

Directed evolution is a method to engineer gene products by a controlled process similar to artificial selection, in which a gene for the biomolecule of interest is modified over subsequent rounds of mutagenesis followed by selection for a desired feature.^{1,2} This method of engineering is useful in chemistry because it allows researchers to bypass rational catalyst design¹ to achieve a number of important goals. Through directed evolution, researchers have expanded chemical space in biological systems,³⁻⁵ developed novel methods to create therapeutics,⁶ and developed novel functions in biological catalysts.⁷⁻⁹ Directed evolution is a valuable technique to rapidly develop new chemistries in biological systems.

1.1.1 Methods to evolve gene targets

There are two major approaches to directed evolution: screening and selection. In both methods, the investigator generates a library of variants, usually via random mutagenesis of a particular region of interest in the target gene.^{2,10} When using a screening method to search for active variants, or “hits,” the result of an assay is directly tied to the target gene’s desired function. Screening thereby allows straightforward interpretation of the results of the experiment. However, because a researcher must perform the assay to test each library member individually, the size of the test library is restricted to the number of iterations it is feasible for the researcher to perform. This limitation may prevent investigators from testing and accessing the best possible hits. Unlike in a screening system, a directed evolution experiment using selection requires that the performance of each library member is tied to the survival of an organism or the propagation of a virus. It is necessary for a researcher to develop a selection scheme to provide this link between

propagation and the activity of the library member. While developing such a scheme can be a complex process, this technique allows for a much larger library size because it is not necessary to manually test the performance of each individual library member. Nonfunctional library members will not appear as outputs of each round of selection, and over multiple rounds, the best hits can become enriched in the output, linking desired function with evolutionary fitness.

1.1.2 The state of the field of directed evolution in mammalian cells

Directed evolution schemes exist to engineer target gene products in bacteria, yeast, and phages,^{11–16} but a variety of constraints prevents such schemes from being widely used directly in mammalian cells. Transformation efficiency in bacteria allows the creation of large libraries in bacterial cells, where each cell contains only one library member. Mammalian cells exhibit poor transformation efficiency,¹⁷ so transient transfection is often used to introduce foreign DNA into mammalian cells. Transient transfection using such reagents as polyethylenimine results in an extremely heterogeneous distribution of plasmid DNA into target cells, such that some cells receive multiple copies of the plasmid of interest and some cells receive none. Such a plasmid distribution cannot ensure that cells each receive only one library member, making directed evolution impossible because the cell's output phenotype could be the result of the activity of any of a number of library members. Unlike bacteria and yeast, mammalian cells do not amplify and maintain foreign DNA, so retrieving the sequences of successful library members is technically challenging.

Different methods have been developed to circumvent the impracticality of introducing a DNA library by transformation. Wang and Tsien developed a method of directed evolution in B-cells relying on somatic hypermutation.⁶ RNA viral vectors such

as Sindbis virus¹⁸ and human immunodeficiency virus (HIV)-1¹⁹ with naturally high genome replication error rates, as well as DNA viruses such as adenovirus with viral polymerases evolved to be more error prone,²⁰ have been used to deliver the DNA for directed evolution to the host cells. These methods successfully allowed the evolution of a variety of biomolecules of interest, such as a new fluorescent protein,⁶ GPCRs,¹⁸ and transcription factors.¹⁸⁻²⁰ However, the utility of the techniques using somatic hypermutation and the viral vectors previously described are limited to the evolution of large genes. The generation of library members in these techniques rely on error-prone DNA replication in either the genome of the host cell or the viral vectors themselves, such that the mutation frequency required to successfully target small genes would be lethal for the host cell or would prevent replication of the viral vector by damaging other areas of their genomes. In addition, such overall random mutagenesis may not yield improved hits as efficiently as a library mutated in a target area, because random mutagenesis cannot specifically focus on the active site of an enzyme to change its substrate specificity.¹³ These limitations have thus far inhibited the development of directed evolution schemes in mammalian cells.

1.2 Unnatural amino acid incorporation in proteins in mammalian cells

1.2.1 How to incorporate unnatural amino acids into proteins

Directed evolution is a useful technique to develop tools that permit the incorporation of unnatural amino acids (UAAs) into proteins via translation. Introducing UAAs site-specifically into proteins allows for expanded chemical space beyond the limited functionalities of the 20 canonical amino acids, so an important goal for research in chemical biology is to incorporate UAAs with high efficiency into proteins of interest.

Unfortunately, while cell-free protein production techniques such as solid-state synthesis allow tight control over the placement of a UAA, these methods are expensive and have low yields. Therefore, in order to efficiently incorporate UAAs into proteins, it is necessary to do so during the translation process. To achieve this goal, researchers must evolve translational machinery in the target organism to optimize UAA incorporation within the cell. Directed evolution of this machinery has allowed researchers to incorporate multiple types of UAAs into proteins in both bacteria and eukaryotic cells, including bio-orthogonal conjugation handles,²¹⁻²³ fluorescent probes,^{24,25} photo-crosslinkers,²⁶⁻²⁸ photocaged groups,^{29,30} and residues mimicking post-translational modifications.^{31,32}

Introducing a UAA into proteins via translation requires the host cell to have an aminoacyl-tRNA synthetase (aaRS) and tRNA pair that can work with the native translation machinery to incorporate the amino acid at a specific site in a protein. Such aaRS/tRNA pairs must be orthogonal to the native translational machinery: that is, they must not cross-react with any of the existing cellular machinery. The orthogonal aaRS must accept only the UAA, and none of the natural amino acids, whereas native synthetases must not charge the UAA, or else that UAA's incorporation into proteins will not be site-specific. Native synthetases should not be able to charge the orthogonal tRNA with any natural amino acids, nor should the orthogonal aaRS be able to charge a native tRNA with a UAA. The orthogonal tRNA must only incorporate the UAA in response to a blank codon, typically one of the stop codons (often the amber stop codon TAG) or a quadruplet codon.¹⁷ If all of these conditions are met, a cell should be able to incorporate a UAA in response to the appropriate blank codon.

Typically, aaRS/tRNA pairs are taken from a distantly related domain of life and introduced into the host organism to maximize the chance that they are orthogonal in the host species.¹⁷ aaRS/tRNA pairs from archaea and eukaryotes are typically orthogonal in bacteria, and pairs from bacteria are typically orthogonal in eukaryotes, due to the distant evolutionary relationship between these domains of organisms. The pyrrolysyl aaRS/tRNA pair from methanogenic archaea in the family Methanosarcinaceae naturally incorporates pyrrolysine in response to the TAG codon, and this pair is naturally orthogonal in both bacteria and eukaryotes because TAG is a stop codon for species in those domains.³³

When aaRS/tRNA pairs are incorporated into new hosts, they may not function optimally with the native translational machinery, or they may not incorporate a desired UAA into proteins, so it is necessary to evolve these pairs to optimize their performance and UAA incorporation. Evolution of either the tRNA or the aaRS can be done in bacteria via a two-step selection scheme.¹⁷ During the positive selection step, cells with a blank codon in a gene for antibiotic resistance are transformed with the aaRS/tRNA pair and grown in the presence of antibiotics and the UAA; only cells that receive a functional aaRS/tRNA pair can grow. In the negative selection step, bacteria with a blank codon in a toxic gene are transformed with the aaRS/tRNA pair but are not grown with the UAA. Cells that receive a library member that does not incorporate a natural amino acid in response to the blank codon survive; in this manner, aaRS/tRNA pairs that are active and orthogonal are identified. A similar two-step selection system is used in yeast, where the UAA must be incorporated in a transcriptional activator for necessary metabolic genes during the positive selection, and the negative selection uses a blank codon-containing *URA3* gene and 5-FOA to weed out library members that incorporate natural amino acids

in response to the blank codon.¹⁷ These two-step selection schemes have allowed the successful evolution of a number of aaRS/tRNA pairs for the incorporation of UAAs into proteins at the translational level.

1.2.2 UAA incorporation in mammalian cells

Many more UAAs have been incorporated in bacteria than in eukaryotes because it is easier to perform directed evolution in bacteria than in eukaryotes.^{17,34,35} Only a subset of those UAAs that have been incorporated in eukaryotes have been used in mammalian cells. Directed evolution techniques are not available for evolving aaRS/tRNA pairs in mammalian cells due to the above-described technical limitations of performing directed evolution in mammalian cells, so aaRS/tRNA pairs are often evolved in yeast and then used in mammalian cells.^{11,13} A recent platform developed by Italia et al.³⁵ allows the evolution of bacterial synthetases in *Escherichia coli* with Altered Translational Machinery (ATM) to select for new synthetase variants that can be imported into mammalian cells in order to incorporate UAAs which were previously unavailable in these systems. While the ATM platform has the potential to vastly expand the number of synthetases and UAAs that can be used in mammalian cells, it does not solve every problem for UAA incorporation in mammalian cells, as the synthetase is not the only piece of cellular machinery involved in translation.

In order to site-specifically incorporate a UAA, an evolved synthetase need only interact with its tRNA and the UAA, making evolution of synthetases for use in eukaryotes a relatively simple problem to solve using the ATM platform or by performing directed evolution in yeast. However, the tRNA in question must interact with the host ribosome and elongation factors to install its UAA in the protein containing the blank codon. The

translational machinery in mammals with which the tRNA must interact is highly divergent from that of bacteria, and is even divergent between yeast and mammalian cells, which are both eukaryotes. The activity of the orthogonal tRNA can be the limiting component in UAA incorporation in proteins in mammalian cells.³⁶ It is therefore necessary to evolve the tRNA directly in mammalian cells to improve its performance in incorporating UAAs into proteins in mammalian cells.

1.3 Adeno-associated virus as a tool for DNA delivery & directed evolution

1.3.1 Viruses can be used to transduce mammalian cells

In order to evolve tRNAs in mammalian cells to optimize UAA incorporation, researchers need robust selection schemes that can take place directly in mammalian cells. As discussed above, delivering DNA to mammalian cells is the major challenge in developing successful selection schemes for directed evolution. Transformation efficiency is poor in mammalian cells,¹⁷ and transient transfection provides a highly heterogeneous distribution of DNA to target cells so as to render impossible any selection scheme in which a cell's phenotype is based on the genotype of a single library member that it receives. An additional complication is that, unlike bacteria and viruses, mammalian cells do not maintain plasmids, so retrieving the library members after a round of selection is difficult. Any DNA delivery method for directed evolution in mammalian cells must overcome these obstacles.

One DNA delivery vector that circumvents each of these challenges is the virus. Because viruses have evolved to enter and transduce cells, any virus that can infect mammalian cells can overcome the problem of poor transformation efficiency in these cells. Mammalian viruses have also evolved to amplify and package their genetic material

within the host cell, in this case the mammalian cells that otherwise do not maintain or amplify foreign DNA. This feature allows researchers to retrieve library members after each round of selection and then store them in a relatively stable state; use them to infect new cells for subsequent rounds of selection; or characterize their sequences. Modern molecular cloning techniques make producing recombinant viruses relatively simple, so scientists can add the gene for their biomolecule of interest to the virus genome or replace one of its components with a library member. It is also a straightforward and precise process to titer a virus, so it is possible to have exquisite control over the copy number of virus delivered to cells. These features of viruses allow researchers to deliver at most one library member to host cells during a selection, and also allow them to retrieve and characterize DNA after the selection.

Multiple viruses have been used to deliver DNA to mammalian cells for chemical biology experiments. Adeno-associated virus (AAV) has been widely used as a DNA delivery vector because it is replication-deficient, which gives it a favorably biosafety profile for lab work.^{37,38} While it is easy to produce rAAV, its small genome capacity limits the sizes of the genes of the biomolecules of interest it can carry as cargo.³⁹⁻⁴² *Autographa californica* multicausid nucleopolyhedrovirus (AcMNPV, or baculovirus) is a virus that infects insect cells but has been modified to include the vesicular stomatitis virus G glycoprotein (VSV-G) in its envelope, which allows it to enter any cell that expresses the mammalian low-density lipoprotein receptor (LDL-R).⁴³⁻⁴⁶ Baculovirus has an essentially unlimited capacity for recombination in its double-stranded DNA genome due to its capsid assembly process, which makes it an attractive delivery vector for large genetic cargos.⁴⁷ Sindbis virus and HIV-1 are RNA viruses with relatively large genomes that, as previously

discussed, have been used for directed evolution experiments that rely on the error-prone polymerases of these RNA viruses. Adenovirus (Ad) also has a larger cargo capacity, as well as a stable double-stranded DNA genome. HIV, Sindbis virus, and Ad all require additional safety protocols for handling in the lab, as these viruses are all human pathogens.

1.3.2 AAV as a facile tool for mammalian cell transduction

One of the most popular viruses for use in chemical biology experiments is AAV. AAV is a small, non-enveloped virus from the family Parvoviridae, with a single-stranded DNA genome approximately 5 kb in length.³⁹ The genome consists of three open reading frames (ORFs) flanked by two inverted terminal repeats (ITRs) that act as origins of replication and packaging signals.^{48,49}

The ORF *Rep* codes for four non-structural proteins responsible for AAV DNA replication and packaging.⁵⁰⁻⁵³ One of these proteins, Rep78, induces apoptosis by activating caspase-3,⁵⁴ so the AAV genome cannot be stably integrated into mammalian cells during rAAV production. *Cap* codes for the three structural proteins of the capsid; VP1, VP2, and VP3 are present in the capsid in a 1:1:10 ratio.³⁷ The capsid proteins fit tightly together, restricting the size of the cargo in recombinant AAV (rAAV) to strictly 5 kb or smaller and limiting structural modifications to the capsid. The constraints on rAAV cargo size and AAV's genome toxicity result in the need for AAV genes to be delivered separately from the recombinant cargo during rAAV production.

AAV's third ORF codes for the nonstructural protein Assembly Activating Protein (AAP), which is required for capsid assembly, and whose gene sits within *Cap* in a different reading frame with the weak CTG start codon.⁴⁰ Its ORF occurring in the same mRNA as the Cap proteins facilitates the timely expression of all proteins required for the capsid

assembly process. AAP contains threonine/serine rich regions between these termini that allow these regions to be quite flexible,^{55,56} which may provide opportunity for manipulation.

AAV is a replication-deficient virus and requires a co-infection of a helper virus (typically adenovirus, but sometimes herpesvirus or baculovirus) to replicate within mammalian cells.^{38,57-59} The regions of adenovirus that are required for AAV replication are: VA, E2A, and E4.⁶⁰ Ad DNA-binding protein, which is coded for in the E2A gene, increases the processivity of AAV DNA replication.⁵⁹ Proteins from the E4 ORF and the VA RNA genes promote the second strand synthesis of AAV.³⁷ Without second strand synthesis, AAV cannot replicate. While AAV was discovered as a contaminant in Ad purifications,³⁸ adenovirus is not the only helper virus for AAV. Both herpes simplex virus 1 and 2 (HSV-1 and HSV-2) provide complete helper functions for AAV by allowing *Rep* gene expression and recruiting AAV DNA to replication centers.^{37,57} Baculovirus can also act as a helper virus for AAV, but the mechanisms by which it does so are unknown.³⁷

Because it is replication deficient, AAV is an attractive target for use in clinical research and gene therapies due to its inability to cause harm to researchers and patients.³⁷ Production of rAAV is also facile, as Rep proteins will package any genes into the capsid as long as they sit within the ITRs and are 5kb or less. AAV is therefore a useful tool to use in transducing mammalian cells.

1.3.3 An AAV-based selection scheme for mammalian cells

Virus-Assisted Directed Evolution of tRNAs (VADER), a new directed evolution technique developed by Kelemen et al.,⁶¹ uses rAAV with wild-type capsids to deliver a library of pyrrolysyl tRNA variants to mammalian cells. In that study, the authors infect

cells at a multiplicity of infection (MOI) of less than one virus per cell so that each cell receives at most one library member. They provide the other genes via transient transfection that AAV needs to replicate and produce a UAA-containing capsid (*AdHelper*, *Rep*, and a TAG-containing *Cap*), as well as the UAA azidolysine (AzK) in the cell media. Only cells that received an active tRNA are able to incorporate an amino acid into response to the TAG codon within *Cap*, and these tRNAs are enriched during this positive selection. The newly packaged capsids of viruses containing an orthogonal tRNA contain the azide functional group, but those containing cross-reactive tRNAs contain only canonical amino acids. The negative selection step in this scheme consists of a bio-orthogonal capture, in which the virus capsids are conjugated to DBCO-biotin and are then captured on a streptavidin resin and released by exposure to light. The hits resulting from this selection can be amplified, sequenced, and further characterized to determine their activity in relation to the wild-type tRNA. In that paper, the authors identify a tRNA, A2, that exhibits more than 300% improved activity over the wild-type pyrrolysyl tRNA via transient transfection.

VADER fills an important gap in the field of directed evolution, in that this technique makes it possible to perform selections directly in mammalian cells in order to optimize UAA-incorporating tRNA function in the intended host. This technique ensures that each cell receives at most one library member, which has been generated in a controlled fashion by molecular cloning. VADER does not rely on somatic hypermutation, which allows VADER to evolve small gene products such as tRNAs. As discussed previously, viruses such as AAV evolved to amplify themselves inside the host cell, so as long as the infected cell receives the other genes necessary for AAV replication *in trans*, the cell will

maintain and replicate the library member enough to retrieve and characterize hits from the selection.

However groundbreaking this technique is, there are still limitations to VADER. While hits can be retrieved from each round of selection using this method, it is necessary to re-amplify, re-clone, and re-package the output virus from each round of selection before proceeding to successive rounds of selection. Each of these steps is labor-intensive and takes considerable time. The negative selection from VADER also relies on capturing the output virus using a bio-orthogonal conjugation handle incorporated into the virus capsid, which consists of tightly-fitted proteins that may not tolerate perturbation by all UAAs of interest.⁶²⁻⁶⁴ Indeed, the AAV2 capsid does not tolerate the insertion of several UAAs of interest to researchers, including 5-hydroxytryptophan (unpublished data). It is necessary to modify the established VADER technique to improve the efficiency of the workflow, and so that it does not rely on UAA insertion into the AAV capsid in order to perform selections on tRNAs for such UAAs that the AAV capsid does not tolerate.

1.4 Expanding the scope of viral vectors in the field of directed evolution

In this thesis, I present work to enhance directed evolution platforms in mammalian cells. I demonstrate that it is possible to re-amplify rAAV using recombinant baculoviruses to streamline the workflow between rounds of selection. This modification to the VADER method allows researchers to proceed more quickly between rounds of selection, making it possible to arrive at an optimized evolutionary solution faster than if re-cloning the selection output was necessary after every round.

In addition to using these baculovirus vectors to streamline the current VADER process, I also demonstrate that it is possible to evolve tRNAs using only the positive

selection scheme from VADER without incorporating a UAA into the AAV capsid. Instead, I incorporate the UAA of interest into the essential protein AAP. The termini of AAP are highly structured and do not tolerate modifications, but the flexible threonine/serine rich region in the middle of the protein tolerates modifications and even large insertions.^{56,65} This region of AAP is therefore an ideal location to modify with UAAs that would not be tolerated in the virus capsid, and since AAP is essential to AAV replication,⁴⁰ I demonstrate that I can use it as a selection handle to evolve tRNAs. Although a negative selection is not possible with this system because there is no bio-conjugation handle in the AAV capsid here, I show here that the positive selection from VADER is sufficient to evolve a tRNA to incorporate 5-hydroxytryptophan into proteins in mammalian cells.

Additionally, I present multiple baculovirus vectors that make use of VADER-evolved pyrrolysyl tRNAs to improve the incorporation of diverse UAAs into proteins in mammalian cells. These vectors demonstrate that the tRNAs evolved using the VADER platform suppress TAG stop codons at higher levels than do the wild-type pyrrolysyl tRNA, to a higher degree than previously reported when comparing tRNA performances via transient transfection. Using baculovirus allows the delivery of tRNAs to cells at a controlled level, such that researchers can deliver far fewer copies of the tRNA per cell than is possible using transient transfection. These experiments demonstrate that the evolved tRNA's improvements are especially evident at low copy numbers. I have also incorporated different evolved synthetases into these baculovirus vectors to allow for the incorporation of diverse UAAs, and I have optimized the expression of the protein of

interest by changing the promoter for the aaRS³⁶ and adding a nuclear export sequence to the terminus of the aaRS to ensure that it localizes to the cytosol.

The use of selections in chemical biology research is not limited to evolving gene products. In this thesis, I demonstrate the utility of AAV-based selections to discover protein binding partners in the mammalian proteome. Using packaged rAAV as the output, I show here the development of a two-hybrid selection platform in which the association of bait and prey proteins is necessary to drive transcription of VP1, an AAV capsid protein that is essential to produce infectious virus.⁶⁶ The gene for the prey protein is inserted into the rAAV genome, and the genome will only be repackaged into an AAV capsid if the prey and bait proteins associate, leading to the transcription of the essential VP1 protein. Any virus which is produced will only be infectious if the bait and prey proteins associate, so this system can be used to select for prey-binding partners to any particular bait within the mammalian proteome.

In summary, I demonstrate in this thesis the utility of selection schemes in mammalian cells; I furthermore enhance and expand their applications within current systems. I have developed viral vectors that will be useful to the larger scientific community for stop codon suppression and UAA incorporation. The modifications and enhancements I have made to a protocol for a selection scheme in mammalian cells will make the process more efficient and generalizable. Finally, the two-hybrid system that I have developed based on infectious virus production will allow high-throughput scanning of the human proteome for protein binding partners. These developments expand the scope of directed evolution in mammalian cells.

1.5 References

- (1) Arnold, F. H. *Acc. Chem. Res.* **1998**, *31* (3), 125.
- (2) Sutherland, J. D. *Curr. Opin. Chem. Biol.* **2000**, *4* (3), 263.
- (3) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. *J. Am. Chem. Soc.* **2017**, *139* (34), 11670.
- (4) Usanov, D. L.; Chan, A. I.; Maianti, J. P.; Liu, D. R. *Nat. Chem.* **2018**, *10* (7), 704.
- (5) Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. *Science* **2018**, *360* (6384), 71.
- (6) Wang, L.; Tsien, R. Y. *Nat. Protoc.* **2006**, *1* (3), 1346.
- (7) Pedersen, Henrik; Hölder, Swen; Sutherlin, Daniel P.; Schwitter, Urs; King, David S.; Schultz, P. G. *Proc. Natl. Acad. Sci.* **1998**, *95* (18), 10523.
- (8) Arnold, F. H. *Trends Biotechnol.* **1990**, *8* (C), 244.
- (9) Umeno, D.; Arnold, F. H. *J. Bacteriol.* **2004**, *186* (5), 1531.
- (10) E. T. Farinas, T. Bulter, F. H. A. *Curr. Opin. Biotechnol.* **2001**, *12* (6), 545.
- (11) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. *Angew. Chemie Int. Ed.* **2009**, *48* (22), 4052.
- (12) Bessette, P. H.; Rice, J. J.; Daugherty, P. S. *Protein Eng. Des. Sel.* **2004**, *17* (10), 731.
- (13) Liu, W.; Brock, A.; Chen, S.; Chen, S.; Schultz, P. G. *Nat. Methods* **2007**, *4* (3), 239.
- (14) Boder, E. T.; Wittrup, K. D. *Nat. Biotechnol.* **1997**, *15* (6), 553.
- (15) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472* (7344), 499.
- (16) Clackson, T.; Hoogenboom, H. R.; Griffiths, A. D.; Winter, G. *Nature* **1991**, *352* (6336), 624.
- (17) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79* (1), 413.
- (18) English, J. G.; Olsen, R. H. J.; Lansu, K.; Strachan, R. T.; Wacker, D.; Correspondence, B. L. R. *Cell* **2019**, *178*, 748.
- (19) Das, A. T.; Zhou, X.; Vink, M.; Klaver, B.; Verhoef, K.; Marzio, G.; Berkhout, B. *J. Biol. Chem.* **2004**, *279* (18), 18776.
- (20) Berman, C. M.; Papa, L. J.; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D. *J. Am. Chem. Soc.* **2018**, *140* (51), 18093.
- (21) Dommerholt, J.; Schmidt, S.; Temming, R.; Hendriks, L. J. A.; Rutjes, F. P. J. T.; van Hest, J. C. M.; Lefeber, D. J.; Friedl, P.; van Delft, F. L. *Angew. Chemie Int. Ed.* **2010**, *49* (49), 9422.
- (22) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2009**, *131* (25), 8720.
- (23) Elliott, T. S.; Townsley, F. M.; Bianco, A.; Ernst, R. J.; Sachdeva, A.; Elsässer, S. J.; Davis, L.; Lang, K.; Pisa, R.; Greiss, S.; Lilley, K. S.; Chin, J. W. *Nat. Biotechnol.* **2014**, *32* (5), 465.
- (24) Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G. *J. Am. Chem. Soc.* **2013**, *135* (34), 12540.
- (25) Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. *J. Am. Chem. Soc.* **2014**, *136* (44), 15551.
- (26) Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S.

- Nat. Methods* **2005**, 2 (3), 201.
- (27) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, 124 (31), 9026.
- (28) Lin, S.; He, D.; Long, T.; Zhang, S.; Meng, R.; Chen, P. R. *J. Am. Chem. Soc.* **2014**, 136 (34), 11860.
- (29) Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2010**, 132 (12), 4086.
- (30) Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2012**, 134 (29), 11912.
- (31) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, 4 (4), 232.
- (32) Kim, C. H.; Kang, M.; Kim, H. J.; Chatterjee, A.; Schultz, P. G. *Angew. Chemie - Int. Ed.* **2012**, 51 (29), 7246.
- (33) Krzycki, J. A. *Curr. Opin. Microbiol.* **2005**, 8 (6), 706.
- (34) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chemical Science*. Royal Society of Chemistry January 1, 2015, pp 50–69.
- (35) Italia, J. S.; Addy, P. S.; Wrobel, C. J. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A. *Nat. Chem. Biol.* **2017**, 13 (4), 446.
- (36) Zheng, Y.; Lewis, T. L.; Igo, P.; Polleux, F.; Chatterjee, A. *ACS Synth. Biol.* **2017**, 6 (1), 13.
- (37) Samulski, R. J.; Muzyczka, N. *Annu. Rev. Virol.* **2014**, 1 (1), 427.
- (38) Atchison, R. W.; Casto, B. C.; Hammon, W. M. D. *Science (80-)*. **1965**, 149 (3685), 754.
- (39) Srivastava, A.; Lusby, E. W.; Berns, K. I. *J. Virol.* **1983**, 45 (2).
- (40) Sonntag, F.; Schmidt, K.; Kleinschmidt, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107 (22), 10220.
- (41) Nakai, H.; Storm, T. A.; Kay, M. A. *Nat. Biotechnol.* **2000**, 18 (5), 527.
- (42) Dyka, F. M.; Boye, S. L.; Chiodo, V. A.; Hauswirth, W. W.; Boye, S. E. *Hum. Gene Ther. Methods* **2014**, 25 (2), 166.
- (43) Jarvis, D. L. In *Methods in Enzymology*; Academic Press Inc., 2009; Vol. 463, pp 191–222.
- (44) Barsoum, J.; Brown, R.; McKee, M.; Boyce, F. M. *Hum. Gene Ther.* **1997**, 8 (17), 2011.
- (45) Kaikkonen, M. U.; Rätty, J. K.; Airene, K. J.; Wirth, T.; Heikura, T.; Ylä-Herttua, S. *Gene Ther.* **2006**, 13 (4), 304.
- (46) Kost, T. A.; Condreay, J. P.; Jarvis, D. L. *Nature Biotechnology*. Nature Publishing Group May 5, 2005, pp 567–575.
- (47) Kost, T. A.; Condreay, J. P. *Trends in Biotechnology*. Elsevier Current Trends April 1, 2002, pp 173–180.
- (48) McLaughlin, S. K.; Collis, P.; Hermonat, P. L.; Muzyczka, N. *J. Virol.* **1988**, 62 (6).
- (49) Hauswirth, W. W.; Berns, K. I. *Virology* **1977**, 78 (2), 488.
- (50) Pereira, D. J.; McCarty, D. M.; Muzyczka, N. *J. Virol.* **1997**, 71 (2), 1079.
- (51) Ni, T. H.; McDonald, W. F.; Zolotukhin, I.; Melendy, T.; Waga, S.; Stillman, B.; Muzyczka, N. *J. Virol.* **1998**, 72 (4), 2777.
- (52) Im, D. S.; Muzyczka, N. *Cell* **1990**, 61 (3), 447.
- (53) Im, D. S.; Muzyczka, N. *J. Virol.* **1992**, 66 (2).

- (54) Schmidt, M.; Afione, S.; Kotin, R. M. *J. Virol.* **2000**, *74* (20), 9441.
- (55) Tse, L. V.; Moller-Tank, S.; Meganck, R. M.; Asokan, A. *J. Virol.* **2018**, *92* (14).
- (56) Naumer, M.; Sonntag, F.; Schmidt, K.; Nieto, K.; Panke, C.; Davey, N. E.; Popa-Wagner, R.; Kleinschmidt, J. A. *J. Virol.* **2012**, *86* (23), 13038.
- (57) Buller, R. M. L.; Janik, J. E.; Sebring, E. D.; Rose, J. A. *J. Virol.* **1981**, *40* (1), 241.
- (58) Urabe, M.; Nakakura, T.; Xin, K.-Q.; Obara, Y.; Mizukami, H.; Kume, A.; Kotin, R. M.; Ozawa, K. *J. Virol.* **2006**, *80* (4), 1874.
- (59) Ward, P.; Dean, F. B.; O'Donnell, M. E.; Berns, K. I. *J. Virol.* **1998**, *72* (1), 420.
- (60) Matsushita, T.; Elliger, S.; Elliger, C.; Podsakoff, G.; Villarreal, L.; Kurtzman, G. J.; Iwaki, Y.; Colosi, P. *Gene Ther.* **1998**, *5* (7), 938.
- (61) Kelemen, Rachel E.; Jewel, Delilah; Pasha, Muhammad; Cao, Xiaofu; Chatterjee, A. *Submiss.*
- (62) Xie, Q.; Bu, W.; Bhatia, S.; Hare, J.; Somasundaram, T.; Azzi, A.; Chapman, M. *S. Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (16), 10405.
- (63) Agbandje-McKenna, M.; Kleinschmidt, J. *Methods Mol. Biol.* **2011**, *807*, 47.
- (64) Kelemen, R. E.; Mukherjee, R.; Cao, X.; Erickson, S. B.; Zheng, Y.; Chatterjee, A. *Angew. Chemie Int. Ed.* **2016**, *55* (36), 10645.
- (65) Tse, L. V.; Moller-Tank, S.; Meganck, R. M.; Asokan, A. *J. Virol.* **2018**, *92* (14).
- (66) Girod, A.; Wobus, C. E.; Zádori, Z.; Ried, M.; Leike, K.; Tijssen, P.; Kleinschmidt, J. A.; Hallek, M. *J. Gen. Virol.* **2002**, *83* (5), 973.

Chapter 2

Baculovirus vectors to amplify AAV during directed evolution

2.1 Introduction

The principles of directed evolution and the limitations of performing selections directly in mammalian cells are outlined in detail in Chapter 1. This chapter deals with the use of viruses instead of transient transfection to transduce mammalian cells to boost the amplification of library members during and after selections.

2.1.1 *Directed evolution in mammalian cells*

Directed evolution is a highly important tool in chemical biology that allows researchers to access new chemistries in biological contexts.¹⁻⁵ While this technique is highly developed in bacteria, yeast, and phages,⁶⁻¹² the controlled evolution of small gene products in mammalian cells was not possible until relatively recently, due to specific restrictions on working with mammalian cells. Because of poor transformation efficiency in mammalian cells,⁵ researchers usually introduce foreign DNA into mammalian cells via transient transfection. Unfortunately, this system does not permit the controlled introduction of one library member per cell due to the heterogeneous distribution of DNA uptake during transient transfection. Somatic hypermutation can be used to generate libraries within mammalian cells,¹³ but its reliance on random mutagenesis may not yield the kind of targeted libraries that researchers often prefer.⁶ This technique is also limited to large gene products because the mutational frequency necessary to generate a library via this method in a small gene product would kill the host cell.

It is relatively simple to transform yeast and bacteria, both of which also maintain and amplify plasmid DNA. Researchers may therefore generate a randomized library via established cloning methods and transform this library into either bacteria or yeast. These cells will have at most one library member per cell, and the researcher may proceed with

selections on this transformed library and retrieve the DNA afterwards for sequence characterization. Mammalian cells, on the other hand, do not maintain foreign DNA as readily, so it is challenging to retrieve and sequence the DNA from any selections undertaken in mammalian cells.

2.1.2 Using viruses to transduce mammalian cells

A solution that overcomes all of these challenges of performing selections in mammalian cells is to transduce the cells with viruses, rather than relying on transformation, transient transfection, or somatic hypermutation. While mammalian cells do not naturally maintain foreign DNA, viruses evolved to replicate themselves within their host cells, so mammalian viruses should be able to amplify their own genetic material once they infect the cell. Retrieving and characterizing the library members becomes simple. It is also possible to control the copy number of viruses delivered to the host cell, as one need only infect cells with a titered virus at a level of fewer than one virus per cell, to ensure that each cell has at most one library member when selection begins. Finally, since recombinant viruses are easy to produce via molecular cloning, it is possible to generate a library of small gene products that would be inaccessible to directed evolution via somatic hypermutation, or even continuous evolution techniques using viral vectors that depend upon error-prone polymerases.¹⁴⁻¹⁶ Viruses are the ideal vector for directed evolution of small gene products in mammalian cells.

Adeno-associated virus (AAV) is an excellent virus for use in mammalian cell-based directed evolution experiments. AAV is a small, non-enveloped virus from the family Parvoviridae with a 5kb single-stranded DNA genome.¹⁷ This genome consists of three open reading frames (ORFs). The ORF *Rep* codes for four nonstructural proteins

expressed from the p5 and p19 promoters that replicate and package the genome into the capsid.¹⁸⁻²¹ *Cap* is expressed from the constitutively active p40 promoter and codes for the three capsid proteins VP1, VP2, and VP3.^{22,23} Assembly Activating Protein (AAP) is also expressed from the p40 promoter, from a weak CTG start codon in a different reading frame from the capsid proteins.^{23,24} All of these genes are flanked by two inverted terminal repeats (ITRs) that act as origins of replication and packaging signals.^{17,25,26} AAV is a replication deficient virus that requires helper genes from adenovirus, baculovirus, or herpesvirus to replicate inside mammalian cells.²⁷⁻³⁰ These genes can be provided by transient transfection to eliminate the need for coinfection by a helper virus during AAV production.³¹

AAV is of particular interest to researchers for its potential uses for gene therapies due to its low immunogenicity and lack of cytotoxicity due to being replication deficient.^{23,32} This replication deficiency and low toxicity to human and animal subjects makes AAV an attractive vector to use in directed evolution applications as well. rAAV is simple to create: any genes up to 5kb can be packaged inside an AAV capsid as its single-stranded DNA genome as long as they sit within the ITRs. Since it is easy and relatively safe to manipulate rAAV, this virus was the vector of choice for Virus-Assisted Directed Evolution of tRNAs (VADER).³³

2.1.3 VADER: an AAV-based directed evolution platform

VADER is a technique that relies on AAV as a vector to deliver libraries of small gene products to cells.³³ In this work, a library of mutant pyrrolysyl tRNAs (PylR) was packaged along with a fluorescent reporter into wild type AAV capsids. Since it is simple to measure an infective titer for this virus by flow cytometry, it is possible to infect a

population of cells at a multiplicity of infection (MOI) less than 1, such that each cell will be infected by at most one virus. This level of control is necessary to ensure that one cell does not receive 2 or more library members. The authors provided *AdHelper*, helper genes from adenovirus that AAV needs to replicate,^{31,34} along with *Rep* and *Cap* with a T454TAG mutation. They also added the unnatural amino acid (UAA) azidolysine (AzK) into the cell media, so that any cell that received a virus containing an active PytR variant would produce AAV with AzK in the capsid. This exposed AzK allowed the authors to conjugate the virus to DBCO-biotin and perform a pulldown on a streptavidin resin to capture only virus that contained AzK in the capsid. The pulldown excluded all viruses that had cross-reactive PytR library members that incorporated canonical amino acids in response to the TAG codon in *Cap*. The active library members could be re-amplified out of the AAV genome and re-cloned into plasmids for sequencing, further characterization, and transient transfection into mammalian cells to produce AAV for subsequent rounds of selection.

Re-cloning the outputs of each round of selection is necessary because after the negative selection's streptavidin pulldown and photo-release, there are so few viral particles remaining that if one were to proceed directly to the next round of selection with those output viruses, there would be too few cells both infected and transfected to undergo another round of selection. Indeed, it is necessary to concentrate the output virus to even proceed with re-cloning.³³ While re-cloning the output virus is desirable for sequencing and testing hits in transient transfection experiments, it is a time-consuming and labor-intensive process that delays further selections. It is therefore necessary to develop a procedure to re-amplify output AAV between rounds of selection that does not depend on re-cloning the library.

While VADER is an important new technique in the field of directed evolution, there are limitations to the protocol. During the positive selection, cells are infected with a library of tRNA-containing rAAV, and then transfected with plasmids containing the other genes necessary to replicate and repackage rAAV. This system limits the amplification of rAAV in multiple ways. Not every cell which is infected by a library member-containing virus will receive the transfected plasmids because of the heterogeneous distribution of DNA during transient transfection.³⁵⁻³⁷ Transfection reagents like PEI (Figure 2.1) block the entry of viruses into cells, so it is possible that not every library member makes it into a cell to undergo selection, which creates the additional drawback of possibly missing highly active variants during selection. With more complex selections involving more genetic cargo, it might be necessary to use multiple plasmids to deliver all the necessary DNA, and transfection efficiency drops with each additional plasmid used. If researchers wish to optimize this selection system, it will be necessary to break its dependence on transient transfection.

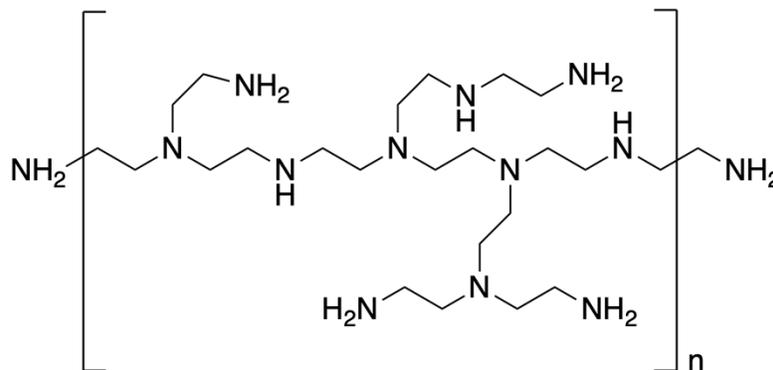


Figure 2.1. Structure of polyethylenimine (PEI)

2.1.4 DNA delivery by baculovirus

While AAV is a small virus with strictly limited cargo capacity, other viruses are not so limited in the amount of recombinant DNA they can package. Baculovirus is one such virus that has a much larger cargo capacity. The *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV), the most commonly used baculovirus, has a large double stranded circular genome almost 134kb in length.³⁸⁻⁴⁰ Since this virus infects insects, it will replicate in insect cells but not in mammalian cells, and therefore poses minimal risk to human researchers, a feature which makes this virus useful in laboratory settings.³⁹ Baculovirus has a large, potentially unlimited cargo capacity of at least 38kb of DNA; this large cargo capacity is likely due to its capsid assembly process, which seems to extend to accommodate extra DNA.⁴¹ This feature makes it an attractive vector to deliver DNA to mammalian cells.

Recombinant baculovirus is relatively simple to produce. Because the virus genome is so large, direct cloning into it is difficult; therefore, shuttle plasmids containing the genes of interest can be transformed into *Escherichia coli* cells which maintain a bacmid containing the rest of the baculovirus genome. The shuttle plasmid undergoes recombination with the bacmid, resulting in a recombinant bacmid that can be isolated and transfected into insect cells.^{38,42} Recombinant baculovirus are typically grown in Sf9 or Sf21AE cells derived from *Spodoptera frugiperda* ovarian tissue, which can be conveniently cultured either in suspension or adherent.³⁹ Even though baculovirus with a wild type envelope can infect mammalian cells, transduction efficiency is low. Barsoum et al. engineered a baculovirus vector to express the vesicular stomatitis virus G glycoprotein (VSV-G) in its envelope, which allows the virus to enter any cell that expresses the low density lipoprotein receptor (LDL-R).⁴³ This modification greatly improves baculovirus

transduction in mammalian cells by improving viral entry into cells by augmenting gp46-mediated endosomal release.^{44,45} Recombinant baculovirus with their large cargo capacities have become useful and facile vectors for research in mammalian cells.

It is possible to produce rAAV on a large scale in Sf9 cells using baculovirus to provide helper genes as well as the AAV genes.⁴⁶⁻⁴⁸ Sf9 cells can be co-infected with two recombinant baculoviruses: one containing the rAAV genome containing the genes of interest, and the other containing helper genes along with *Rep* and *Cap*.⁴⁸ Since Sf9 cells can be grown in suspension to scale up production, this system allows the production of high titer rAAV. This system provides a model for approaching rAAV re-amplification. The output of VADER selections cannot be re-amplified in Sf9 cells, as AAV does not infect insect cells efficiently, but it is possible to use recombinant baculovirus to re-amplify AAV in mammalian cells because of the VSV-G modification that allows baculovirus to infect mammalian cells.

Here I present a method to efficiently re-amplify rAAV in mammalian cells using baculovirus vectors to provide *AdHelper* genes and the AAV genes *Rep* and *Cap*. I demonstrate the utility of co-infection with baculovirus over transient transfection to produce rAAV during the positive selection process in VADER and to re-amplify rAAV in mammalian cells after a round of selection. In addition, I show that the expression of capsid proteins from their endogenous promoters expressed from baculovirus vectors limits the re-amplification of rAAV and provide an engineered baculovirus vector to overcome this limitation.

2.2 Results and Discussion

2.2.1 Determining the effects of addition of PEI on AAV infection

A drawback of the VADER protocols of infecting cells with rAAV containing the tRNA library and then providing the other genes necessary to make AAV via transient transfection is that the presence of PEI or other transfection reagents blocks AAV entry into the cell. This limitation prevents researchers from ensuring that all library members enter cells and are subject to selection, so the best hits from a randomized library could potentially be missed. Transfection with more plasmids is also less efficient; as such, optimizing complex selections using transient transfection would also be more difficult. In order to prevent these problems, dependence on transient transfection during selections must be reduced.

To quantify the effect of PEI on AAV infection efficiency, I infected HEK-293 cells with rAAV-GFP at either 1 or 5 MOI, and added PEI simultaneously, 4 hours later (to simulate VADER's transfection protocol), or not at all (Figures 2.2 and 2.3). Pictures of infected cells expressing the GFP carried as cargo in the rAAV genome show that the addition of PEI inhibits viral transduction of HEK cells (Figure 2.2). Delaying the addition of PEI improves infectivity, but it does not rescue it to the levels achieved when PEI is not added. Using GFP fluorescence as a proxy for infection efficiency, it appears that an MOI of 5, and adding the PEI 4 hours after infection, is necessary to achieve infection results similar to those when using an MOI of 1, without PEI (Figure 2.3). The addition of PEI clearly inhibits transduction of mammalian cells by AAV.

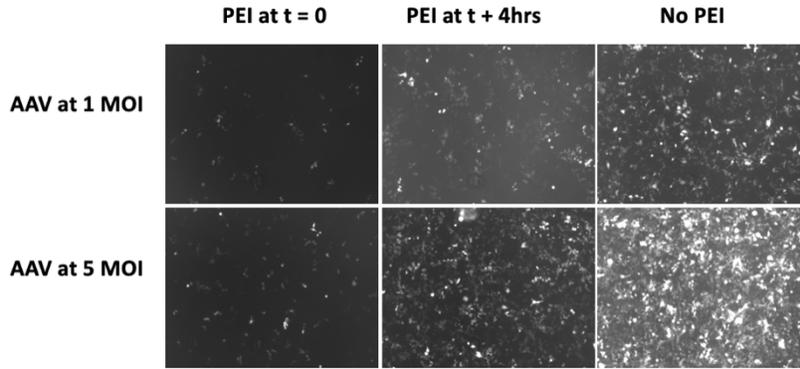


Figure 2.2. The effects of PEI on transduction by rAAV. HEK-293T cells were infected with rAAV-GFP at either 1 or 5 MOI. Cells also received no PEI, PEI at the time of infection ($t = 0$), or PEI 4 hours post-infection ($t + 4\text{hrs}$).

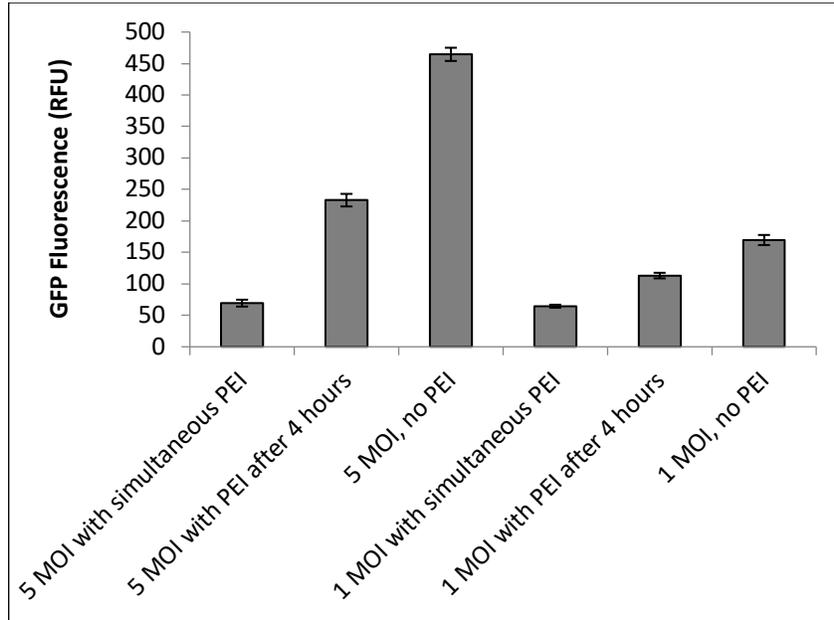


Figure 2.3. The effects of PEI on GFP production in infected cells. Cells infected with rAAV-GFP and treated with PEI were lysed, and GFP fluorescence was measured.

2.2.2 Comparing pAcBac vectors to original rAAV production plasmids

In order to improve the efficiency of rAAV transduction of mammalian cells during directed evolution experiments, it is necessary to provide the other genes (*AdHelper*, *Rep*, and *Cap*) necessary to make rAAV by a means other than transient transfection. Baculovirus vectors present a promising solution to this problem because of their

practically unlimited cargo capacity, their favorable biosafety profile, and the ease with which they can be manipulated to deliver cargo to mammalian cells.

I prepared baculovirus vector shuttle plasmids containing these genes and compared the production of infective rAAV via original infection with rAAV and transient transfection using the original pHelper and pIDT-MbPylRS-RC2wt plasmids versus the shuttle plasmids pAcBac-Helper and pAcBac-MbPylRS-RC2wt. The pAcBac shuttle plasmids contained the same cargo as the pIDT and pHelper plasmids, but in the plasmid backbone that can undergo recombination with a bacmid to produce recombinant baculovirus.

After 72 hours, the cells packaging rAAV show similar numbers of fluorescent cells, indicating that the cells were infected by similar amounts of input AAV, and that the input virus recombinant genome undergoes replication, transcription, and translation at similar rates (Figure 2.4). However, cells infected with AAV harvested from those packaging cells and produced using the pAcBac shuttle plasmids exhibit far less fluorescence than cells infected with the same volume of AAV produced with the original VADER plasmid system (Figure 2.5), and far fewer new cells seemed to be infected with that harvested AAV (Figure 2.4).

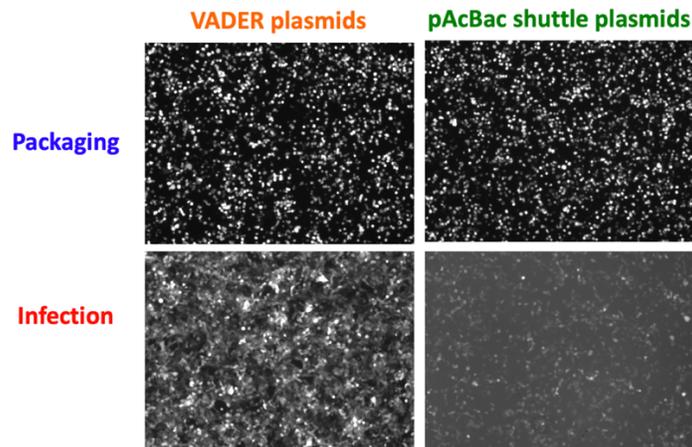


Figure 2.4. Cells packaging rAAV and cells infected with harvested lysate. Cells were imaged 72 hours after being infected with rAAV-GFP and transfected with either the original VADER plasmids or pAcBac shuttle plasmids. Packaging cells were lysed, and new cells were infected with that lysate and imaged after 48 hours.

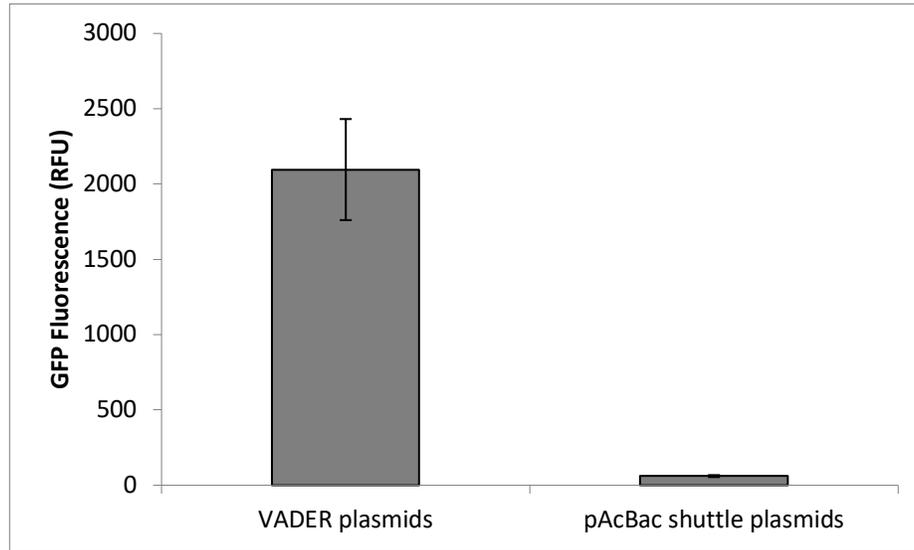


Figure 2.5. GFP fluorescence of cells infected with lysate from cells packaging AAV. New cells were infected with lysate from cells that were infected with rAAV-GFP and transfected with either the original VADER plasmids or with the pAcBac shuttle plasmids. Cells were imaged and lysed after 48 hours, and GFP fluorescence measurements were taken.

In order to determine which plasmid (pAcBac-Helper or pAcBac-MbPylRS-RC2wt) is at fault for reducing the amount of rAAV produced via transient transfection, I infected packaging cells with rAAV-GFP and then transfected with the following pairs of plasmids: pHelper and pIDT-MbPylRS-RC2wt, pAcBac1-Helper and pIDT-MbPylRS-RC2wt, pHelper and pAcBac-MbPylRS-RC2wt, and pAcBac1-Helper and pAcBac-MbPylRS-RC2wt (Figure 2.6). I compared the fluorescence of cells which were infected with lysate from the packaging cells (Figure 2.7). The packaging cells which were transfected with pAcBac-MbPylRS-RC2wt produced much less AAV, regardless of which helper plasmid they received, than cells transfected with pIDT-MbPylRS-RC2wt (based

on the results from cells infected with their lysate). An anti-AAV western blot of lysate from packaging cells shows that much less capsid protein is produced when the AAV genes are provided in the pAcBac backbone (Figure 2.8).

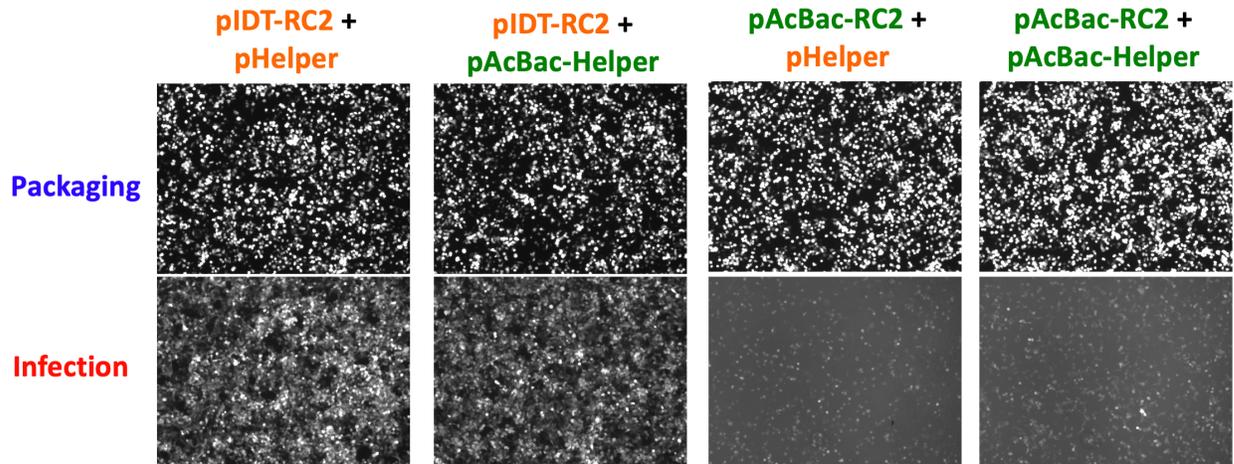


Figure 2.6. Determining the source of limited rAAV production. Cells were infected with rAAV-GFP and transfected with *AdHelper* and *RC2* from either one of the original VADER plasmids or one of the pAcBac shuttle plasmids. After 72 hours, cells were imaged and lysed, and new cells were infected with that lysate. These cells were imaged and lysed after 48 hours, and GFP fluorescence measurements were taken.

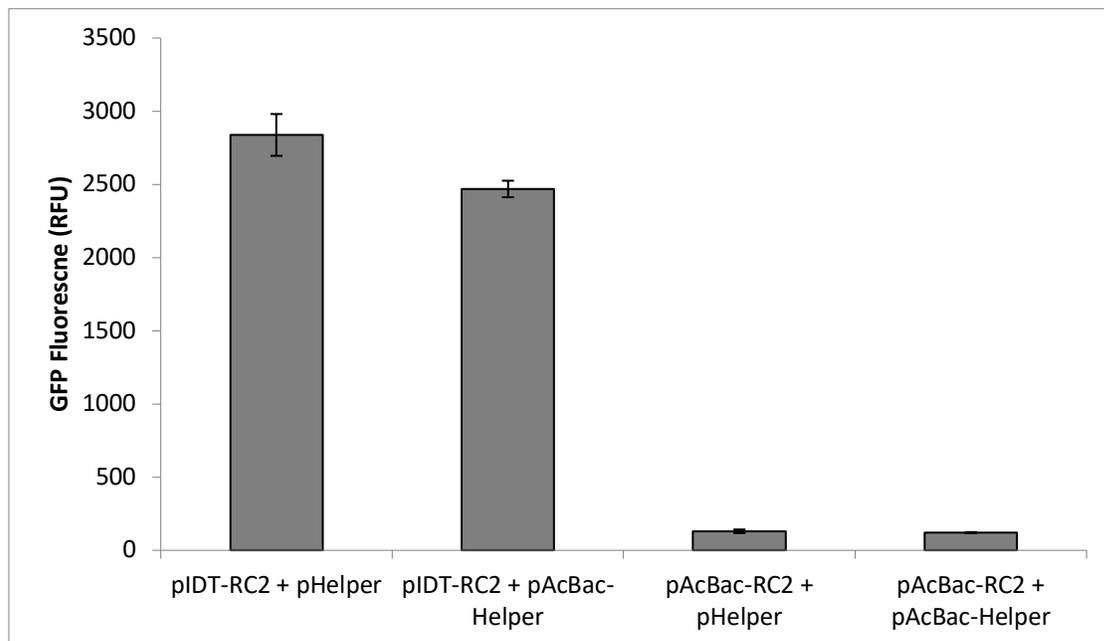


Figure 2.7. GFP fluorescence of cells to test the source of limited rAAV production. Cells were infected with lysate from packaging cells transfected with combinations of the original VADER plasmids and the pAcBac shuttle plasmids to determine the source of reduced rAAV production. After 48 hours, cells were imaged and lysed, and GFP fluorescence was measured.

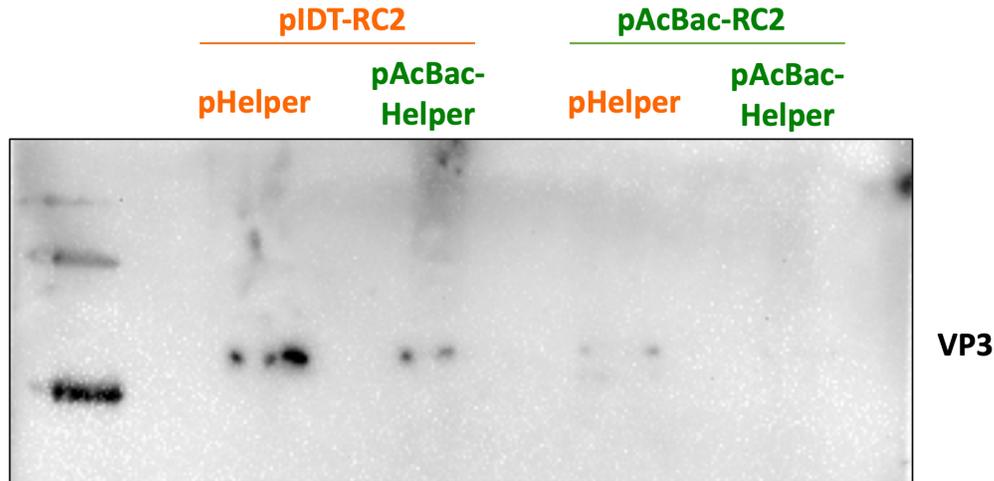


Figure 2.8. Anti-AAV western blot. An anti-VP western blot was performed on lysate from cells packaging rAAV.

To solve this problem of capsid protein limitation, I cloned out the *Cap* genes from AAV2 and placed them under the control of the strong constitutive CMV promoter in a pAcBac shuttle plasmid to maximize production of capsid proteins. I also added the U1 snRNP consensus binding splice site to improve the transcription efficiency of the *Cap* genes. I cloned *Rep* in separately, keeping these genes under the control of their endogenous p5 and p19 promoters to create the final plasmid pAcBac1-Rep-CMV-CapWT-CSS-CMV-MbPylRS (see Appendix for construct map and sequence). I found that overexpressing *Cap* genes from the CMV promoter rescued the production of capsid proteins (Figure 2.9) and also rescued the production of infectious virus (Figure 2.10, Figure 2.11).

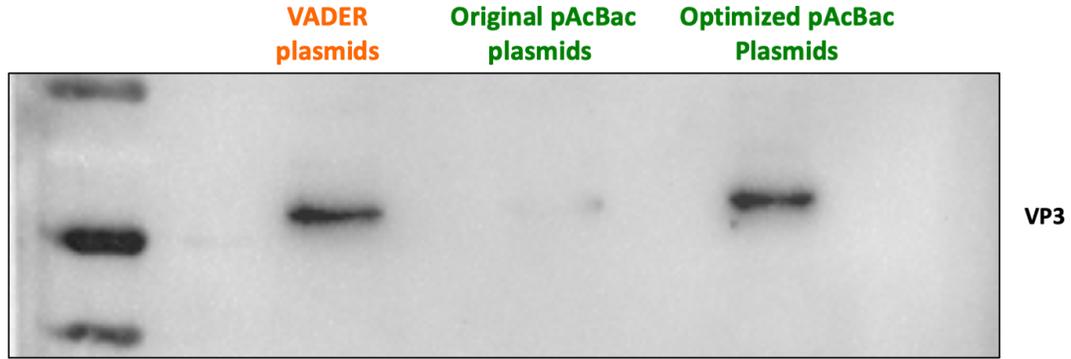


Figure 2.9. Anti-AAV western blot. An anti-AAV western blot was performed on cells packaging rAAV using the VADER plasmids, the original pAcBac shuttle plasmids, or the optimized pAcBac shuttle plasmids.

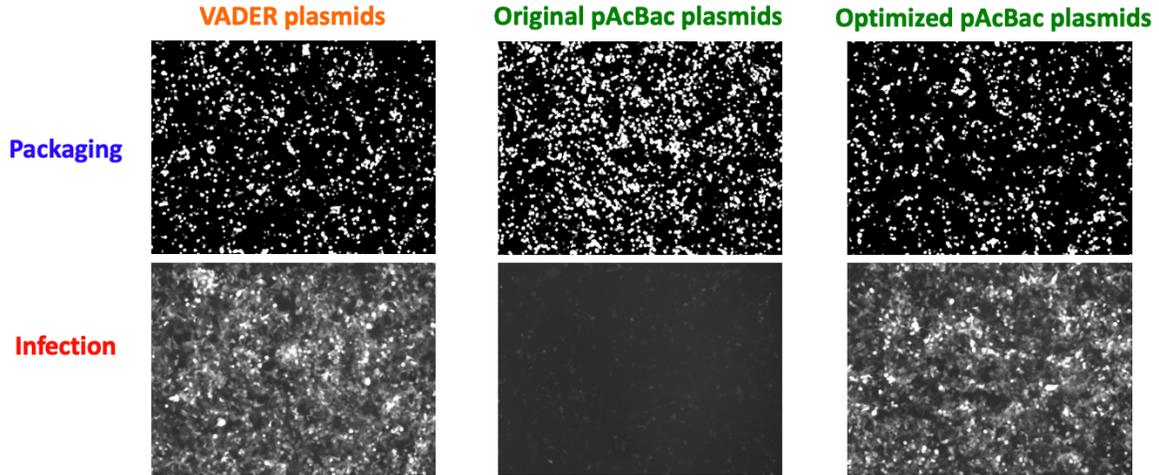


Figure 2.10. Transfecting cells with optimized plasmids. Cells were infected with rAAV and transfected with the VADER plasmids (pHelper and pIDT-RC2-MbPylRS), the original pAcBac shuttle plasmids (pAcBac-Helper and pAcBac1-RC2-MbPylRS), or the optimized pAcBac shuttle plasmids (pAcBac-Helper and pAcBac1-Rep-CMV-CapWT-CSS-CMV-MbPylRS). After 72 hours, the packaging cells were imaged and lysed, and new cells were infected with that lysate. These cells were imaged after 48 hours.

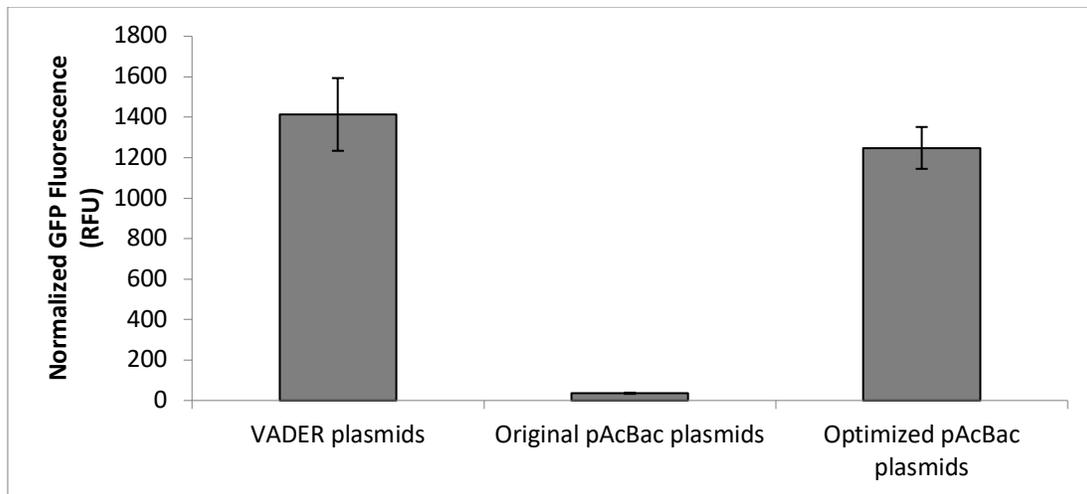


Figure 2.11. GFP fluorescence of cells to test the efficacy of the optimized pAcBac shuttle plasmids. New cells were infected with lysate from cells packaging AAV. After 48 hours, these cells were imaged and lysed, and GFP fluorescence was taken.

I then compared the production of rAAV produced by cells that were infected with rAAV-mCherry and which received *AdHelper*, *Rep*, and *Cap* via the following different methods: transfection with the original VADER plasmid system, transfection with my optimized pAcBac shuttle plasmids, or infection with P1 baculovirus produced from those optimized pAcBac plasmids. Based on images of cells infected with the same volume of lysate from the cells that packaged the rAAV (Figure 2.12), and based on mCherry fluorescence of the cells infected with that lysate (Figure 2.13), it appears that transfection with the optimized pAcBac plasmids produced a similar amount of AAV as transfection with the original VADER plasmids. However, infection with the optimized P1 baculoviruses vastly improved the amount of AAV produced, increasing mCherry fluorescence by 5-fold.

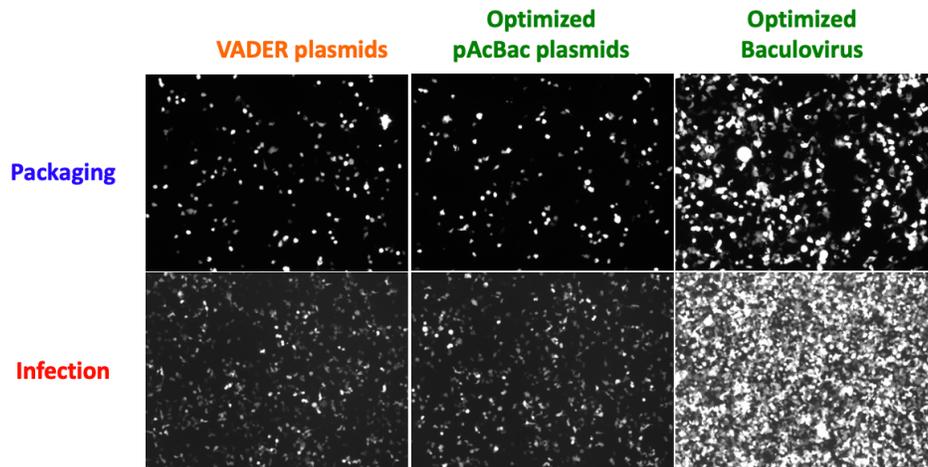


Figure 2.12. Cells packaging rAAV and cells infected with rAAV-containing lysate. Cells were infected with rAAV-mCherry and transfected with the original VADER plasmids, transfected with the optimized pAcBac shuttle plasmids, or infected with baculovirus made from the optimized pAcBac shuttle plasmids. After 72 hours, the packaging cells were imaged and lysed, and new cells were infected with the lysate. These cells were imaged after 48 hours.

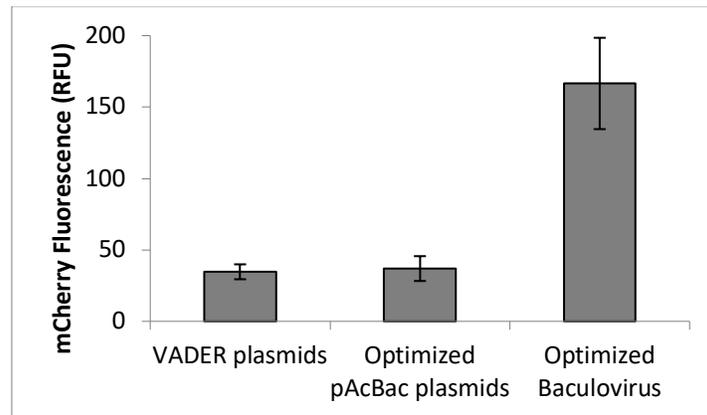


Figure 2.13. mCherry fluorescence of cells infected with rAAV-containing lysate. Cells were infected with rAAV produced by packaging cells to test the effectiveness of the optimized shuttle plasmids and the baculovirus produced from those shuttle plasmids. After 48 hours, these cells were imaged and lysed, and mCherry fluorescence was taken.

2.2.3 *Comparing pAcBac vectors to original VADER plasmids for stop codon suppression*

In the original VADER protocol, rAAV is packaged into wild type capsids, and it is only during the positive selection itself that viral capsids containing AzK are produced. It is therefore necessary to determine whether the baculovirus infection system will

outperform VADER's original transient transfection system in producing more output rAAV.

I compared the amount of output virus produced in cells infected by rAAV-mCherry-PytR and either infected with P1 baculovirus produced using the plasmids pAcBac1-Helper and pAcBac1-Rep-CMV-Cap454TAG-CSS-CMV-MbPylRS or transfected with pHelper and pIDT-MbPylRS-RC2-454TAG. Instead of AzK, which is an excellent selection handle for VADER but a suboptimal substrate for MbPylRS, I used BocK as the UAA for this experiment, as bio-conjugation was not required and as BocK is a superior substrate for the synthetase (Figure 2.14). Surprisingly, the 454TAG version of the optimized baculovirus did not lead to an increase in the amount of infectious AAV produced (Figure 2.15).

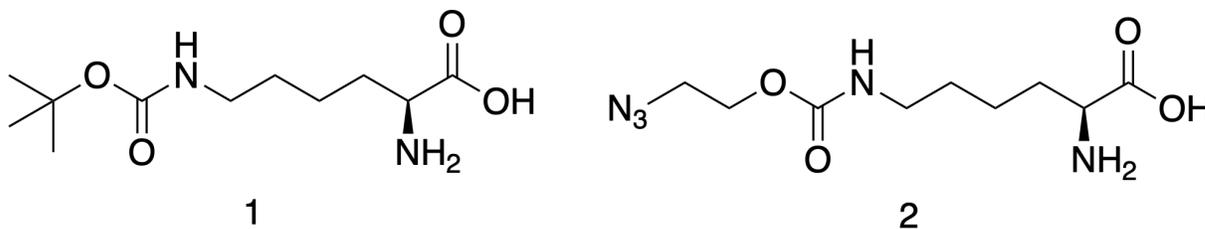


Figure 2.14. Structures of BocK (1) and AzK (2). BocK was used as the unnatural amino acid in all experiments in this chapter.

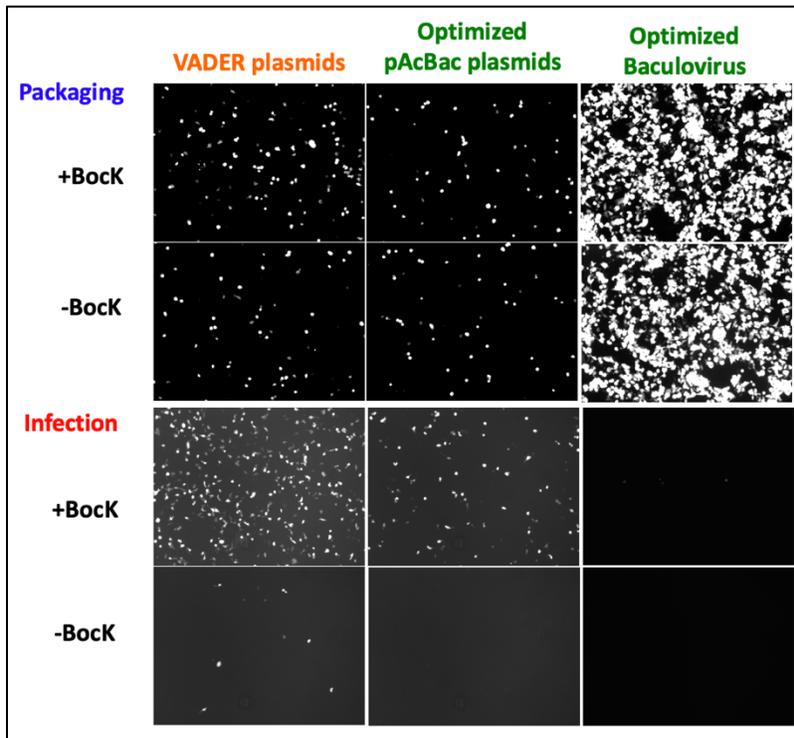


Figure 2.15. Cells packaging rAAV with either wild type or stop codon-containing capsids and cells infected with rAAV-containing lysate. Cells were infected with rAAV and either transfected with plasmids or infected with baculovirus. Top set of images: wild type capsids. Bottom set of images: T454TAG capsids.

An anti-AAV western blot on lysate from cells packaging new AAV after being infected with rAAV-mCherry-PytR and either infected with baculovirus or transfected with either set of plasmids shows that there is still insufficient capsid protein produced when the capsid proteins contain a TAG codon at position 454, even when those genes are being expressed from the highly active CMV promoter (Figure 2.16). It will be necessary to solve the problem of continued capsid protein limitation to move forward with baculovirus-assisted selections like this in the future.

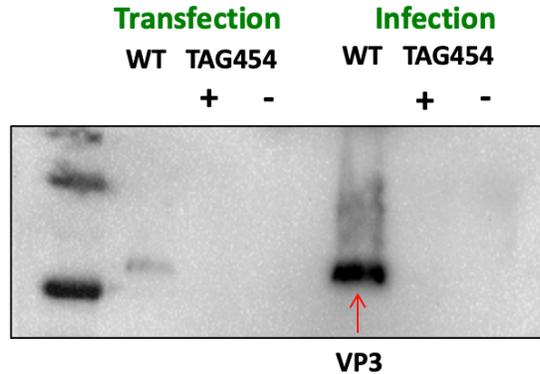


Figure 2.16. Anti-AAV western to determine whether 454TAG capsid proteins are limiting. An anti-AAV western blot was performed on cells infected with rAAV-mCherry and either transfected with the optimized pAcBac shuttle plasmids or infected with the optimized baculovirus.

2.2.4 The source of the improved AAV production

I developed these baculovirus delivery vectors to solve the problem of limited AAV production during the positive selection step in VADER, a problem which has two sources. First, PEI limits AAV entry into the cell. Second, transient transfection delivers DNA to cells in a highly heterogeneous fashion, so that most cells will not receive all the DNA they need to produce more rAAV. Providing the other necessary genes (*AdHelper*, *Rep*, and *Cap*) via baculovirus infection instead of by transient transfection should allow researchers to provide these genes in a much more homogenous manner; it should furthermore also allow improved AAV entry into the packaging cells. More host cells should therefore be involved in packaging AAV.

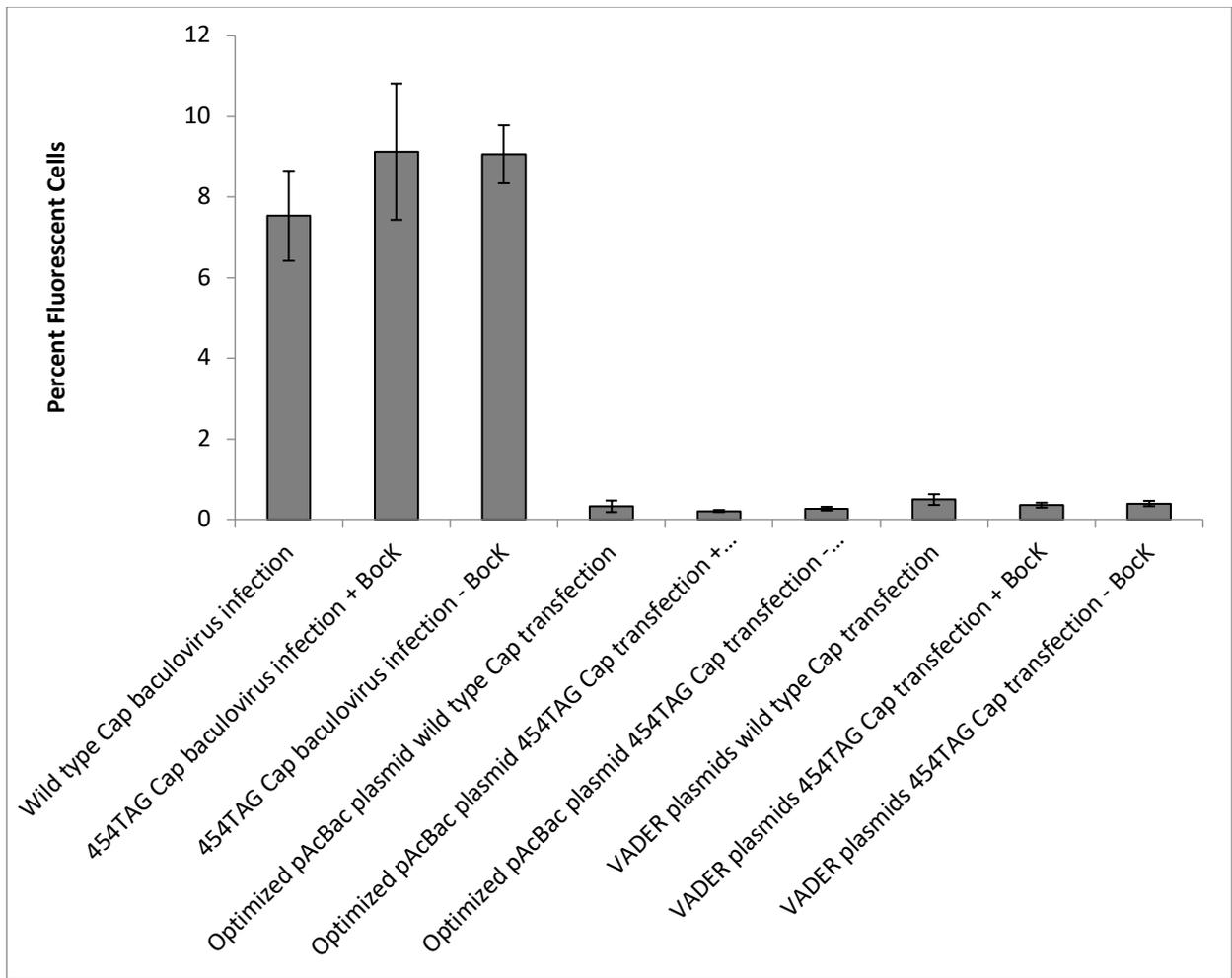


Figure 2.17. Flow cytometry analysis of cells producing rAAV-mCherry. Cells were infected with rAAV-mCherry and either infected with baculovirus or transfected with plasmids to deliver the other genes to make AAV. These cells were harvested and analyzed by FACS to determine the percent fluorescent cells.

To determine whether it is indeed the case that more host cells receive all of the necessary genes to package rAAV, I infected HEK-293T cells with rAAV-mCherry-PyIR at 0.01 MOI and provided *AdHelper*, *Rep*, and *Cap* in one of the following manners: infection of P1 baculovirus (pAcBac-Helper and either pAcBac-Rep-CMV-CapWT-CSS-MbPyIR or pAcBac-Rep-CMV-Cap454TAG-CSS-MbPyIR), transfection with baculovirus shuttle plasmids (with either wild type *Cap* or 454TAG *Cap*), or transfection with the original VADER plasmids (pHelper and either pIDT-MbPyIR-RC2wt or pIDT-

MbPyIRS-RC2-454TAG). I analyzed these packaging cells by FACS and found that there were at least 15x more fluorescent cells in treatments where the host cells had been infected by baculovirus, as opposed to having been transfected with plasmids and PEI (Figure 2.17). These cells are fluorescent because they received all the necessary genes for amplifying the rAAV genome: the mCherry cargo; *Rep*, which replicates the cargo between the ITRs, making it available for transcription to produce the fluorescent protein; and *AdHelper*, without which Rep proteins are not active. This result indicates that one of the sources of increased rAAV production is more cells receiving all of the necessary genes to amplify rAAV.

2.3 Conclusions

The results of these experiments demonstrate that infection with baculovirus is a much more efficient method to amplify rAAV during and between positive selections than is transient transfection. Not only does far less input AAV need to be added during the positive selection when co-infecting with baculovirus, but the other genes (*AdHelper*, *Rep*, and *Cap*) are provided much more homogeneously, so that more cells can be involved in packaging rAAV. This improvement will allow researchers to reamplify rAAV between rounds of selection and to ensure that each potential hit can be accessed during directed evolution. This technique promises to allow access to complex selection systems involving large genetic cargos (such as the CRISPR-Cas system) due to the vast capacity of baculovirus.

2.3.1 Ongoing work and future directions

Because there is not sufficient UAA-containing capsid protein produced using the CMV-Cap baculovirus vectors, it is still necessary to optimize the expression of stop

codon-containing capsid proteins with this baculovirus vector. Future work will also involve using baculovirus to reamplify AAV between rounds of selection to proceed with directed evolution experiments more quickly.

2.4 Acknowledgements

Dr. Rachel Kelemen trained me to make rAAV and gave me access to her plasmid and primer stocks. Dr. Yunan Zheng trained me to make recombinant baculovirus.

2.5 Experimental Procedures

2.5.1. Strains and cell lines

Clonings and plasmid propagations were performed in electrocompetent DH10B *E. coli* cells. Bacmid transposition was performed in chemically competent DH10Bac cells (Invitrogen).

Sf9 cells (Thermo Fisher) were maintained at 27°C in Sf-900 III serum free medium (Gibco), in suspension. HEK-293T (ATCC) cells were maintained at 37°C and 5% CO₂ in DMEM-high glyucose (HyClone) supplemented with 10% fetal bovine serum (Corning) and antibiotic-antimycotic (100 U/mL of penicillin, 100 µg/mL of streptomycin, and 0.25 µg/µL of Fungizone).

2.5.2. Plasmids

All plasmid maps and sequences can be found in the Appendix.

The plasmids pAAV-GFP+1xEcYtR, pAAV-mCherry+1xMmPytR, pHelper, pIDT-MbPylRS-RC2wt, pIDT-MbPylRS-RC2-454TAG, pAcBac1-Helper, pAcBac1-MbPylRS-RC2wt, and pAcBac1-MbPylRS-RC2-454TAG were propagated from Dr. Rachel Kelemen's stocks.

To construct the pAcBac1-Rep-CMV-CapWT-CSS-CMV-MbPylRS and pAcBac1-Rep-CMV-Cap454TAG-CSS-CMV-MbPylRS constructs, the consensus splice donor site (CSS) was added to the 5' end of the *Cap* genes from, downstream of the transcription start site, with the primers Consensus-splice-donor-site-sequence-F. This cassette was amplified with the primers NheI-p40-F and Cap-SalI-R, and this construct was cloned into the pIDT backbone (pIDTSMART-MbPylRS) using the NheI and SalI cut sites to produce the plasmid pIDTSMART-CMV-CAP-CSS. The cassette, along with the CMV promoter and Bgh polyA sequence, was cut with the SpeI and AvrII cut sites, and cloned into pAcBac-MbPylRS-RC2, which had been cut with AvrII. *Rep* was then amplified with the primers SbfI-Rep-F and NotI-Rep-R and cloned into both the wild type and the 454TAG constructs using the SbfI and NotI cut sites.

2.5.3. Unnatural amino acids

The N^ε-Boc-L-Lysine used in these experiments was purchased from Chem-Impex International (catalog number 00363).

2.5.4. Producing recombinant baculovirus

The pAcBac1 shuttle plasmids were chemically transformed into DH10Bac cells (Invitrogen) and grown for 48 hours on LB agar containing 50μg/mL kanamycin, 7μg/mL gentamicin, 10μg/mL tetracycline, 100μg/mL X-gal, and 40μg/mL IPTG. One white colony per construct was chosen and streaked out on fresh medium. After 48 hours of additional growth, one white colony per construct was cultured overnight in LB with kanamycin, gentamicin, and tetracycline. The bacmids were minipreped according to Invitrogen's Bac-to-Bac Expression protocol.

To prepare P0 recombinant baculovirus, 4 μ g of miniprepmed bacmid was transfected into adherent Sf9 cells using X-tremeGENE transfection reagent (Sigma-Aldrich). After 96 hours, the media containing P0 virus was harvested and stored at 4°C away from light. P1 recombinant baculovirus was prepared by infecting shaking flasks of Sf9 cells with P0 virus. After 72 hours, the cultures were spun to pellet out the Sf9 cells, and the supernatant was spun for 60 minutes at 39,800g to pellet the P1 virus. The virus was resuspended in 10% DMSO in PBS and flash-frozen for storage at -80°C. It was titered using the Clontech BacPAK™ Baculovirus Rapid Titer Kit.

2.5.5. Preparing recombinant AAV

To produce rAAV with wild-type capsids, the following plasmids were transfected into HEK-293T cells using PEI (Sigma): pHelper, pIDT-MbPylRS-RC2wt, and either pAAV-Cargo-GFP+1xEcYtR or pAAV-Cargo-mCherry+1xMmPytR. After 72 hours, cells were harvested by scraping and pelleted, saving the culture media. The cell pellets were resuspended in PBS with 300mM NaCl, freeze-thawed twice in a dry ice/ethanol bath and a 37°C water bath to lyse virus-containing cells, and then spun to pellet the cell fragments. The supernatant was then added back to the culture media, and this lysate was incubated at room temperature for 30 minutes with Pierce Universal Nuclease to degrade any unpackaged nucleic acids. After incubation with nuclease, PEG was added to the lysate to 11.5%, and the lysate was precipitated overnight at 4°C. The precipitated lysate was spun to pellet the virus, resuspended in PBS, and flash frozen for storage at -80°C. Infective titer for rAAV containing the fluorescent reporters GFP or mCherry was determined by FACS after infecting HEK-293T cells with rAAV, using sodium butyrate.

To harvest AAV produced on a small scale (in a 12-well plate), cells were lysed with CelLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease. The resultant lysate was flash-frozen and stored at -80°C.

2.5.6. Determining the effects of addition of PEI on AAV infection

HEK-293T cells were infected with either 1 or 5 MOI of rAAV-GFP. Cells also received PEI at the same time as infection with AAV, 4 hours after infection, or not at all. After 48 hours, cells were imaged and lysed, and GFP fluorescence was measured.

2.5.7. Comparing pAcBac shuttle plasmids to original rAAV production plasmids

HEK-293T cells were infected with 5 MOI rAAV-GFP. After 4 hours, cells were transfected with one of the following pairs of plasmids: pHelper and pIDT-MbPylRS-RC2; pAcBac1-Helper and pAcBac1-MbPylRS-RC2wt; or pAcBac1-Helper and pAcBac1-Rep-CMV-CapWT-CSS-CMV-MbPylRS. The infected and transfected cells were imaged and lysed 72 hours later. Western blots were performed on the lysate with the B1 anti-VP antibody which binds to the C terminus of the capsid proteins. 10% of the lysate was used to infect new HEK-293T cells. 48 hours later, infected cells were imaged and lysed, and GFP fluorescence was measured.

2.5.8. Comparing transfection to infection with baculovirus

HEK-293T cells were infected with 5 MOI rAAV-GFP. Cells that were infected with baculovirus were infected simultaneously with 100 MOI P1 virus prepared from the shuttle plasmids pAcBac1-Helper and either pAcBac1-Rep-CMV-CapWT-CSS-CMV-MbPylRS or pAcBac1-Rep-CMV-Cap454TAG-CSS-CMV-MbPylRS. Cells that were transfected were transfected 4 hours later with one of the following pairs of plasmids: pHelper and either pIDT-MbPylRS-RC2 or pIDT-MbPylRS-RC2-454TAG; pAcBac1-

Helper and either pAcBac1-Rep-CMV-CapWT-CSS-CMV-MbPylRS or pAcBac1-Rep-CMV-Cap454TAG-CSS-CMV-MbPylRS. To test the production of rAAV containing a UAA in the capsid, N_{α} -Boc-L-Lysine (BocK) was added to the cell media to a final concentration of 1.0 mM. 72 hours later, cells were imaged and lysed. 10% of that lysate was used to infect new HEK-293T cells, and an anti-VP western blot was also performed on the lysate. After 48 hours, the infected cells were imaged and lysed, and GFP fluorescence was measured.

2.6 References

- (1) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. *J. Am. Chem. Soc.* **2017**, *139* (34), 11670.
- (2) Usanov, D. L.; Chan, A. I.; Maianti, J. P.; Liu, D. R. *Nat. Chem.* **2018**, *10* (7), 704-714.
- (3) Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. *Science* **2018**, *360* (6384), 71.
- (4) Umeno, D.; Arnold, F. H. *J. Bacteriol.* **2004**, *186* (5), 1531.
- (5) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79* (1), 413.
- (6) Liu, W.; Brock, A.; Chen, S.; Chen, S.; Schultz, P. G. *Nat. Methods* **2007**, *4* (3), 239.
- (7) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. *Angew. Chemie Int. Ed.* **2009**, *48* (22), 4052.
- (8) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science (80-.)*. **2003**, *301* (5635), 964.
- (9) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472* (7344), 499.
- (10) Bessette, P. H.; Rice, J. J.; Daugherty, P. S. *Protein Eng. Des. Sel.* **2004**, *17* (10), 731.
- (11) Boder, E. T.; Wittrup, K. D. *Nat. Biotechnol.* **1997**, *15* (6), 553.
- (12) Clackson, T.; Hoogenboom, H. R.; Griffiths, A. D.; Winter, G. *Nature* **1991**, *352* (6336), 624.
- (13) Wang, L.; Tsien, R. Y. *Nat. Protoc.* **2006**, *1* (3), 1346.
- (14) English, J. G.; Olsen, R. H. J.; Lansu, K.; Strachan, R. T.; Wacker, D.; Correspondence, B. L. R. *Cell* **2019**, *178*, 748.
- (15) Berman, C. M.; Papa, L. J.; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D. *J. Am. Chem. Soc.* **2018**, *140* (51), 18093.
- (16) Das, A. T.; Zhou, X.; Vink, M.; Klaver, B.; Verhoef, K.; Marzio, G.; Berkhout, B. *J. Biol. Chem.* **2004**, *279* (18), 18776.
- (17) Srivastava, A.; Lusby, E. W.; Berns, K. I. *J. Virol.* **1983**, *45* (2).
- (18) Pereira, D. J.; McCarty, D. M.; Muzyczka, N. *J. Virol.* **1997**, *71* (2), 1079.
- (19) Ni, T.-H.; McDonald, W. F.; Zolotukhin, I.; Melendy, T.; Waga, S.; Stillman, B.; Muzyczka, N. *J. Virol.* **1998**, *72* (4), 2777.
- (20) Im, D. S.; Muzyczka, N. *Cell* **1990**, *61* (3), 447.
- (21) Im, D. S.; Muzyczka, N. *J. Virol.* **1992**, *66* (2).
- (22) Tratschin, J. D.; West, M. H.; Sandbank, T.; Carter, B. J. *Mol. Cell. Biol.* **1984**, *4* (10), 2072.
- (23) Samulski, R. J.; Muzyczka, N. *Annu. Rev. Virol.* **2014**, *1* (1), 427.
- (24) Sonntag, F.; Schmidt, K.; Kleinschmidt, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (22), 10220.
- (25) McLaughlin, S. K.; Collis, P.; Hermonat, P. L.; Muzyczka, N. *J. Virol.* **1988**, *62* (6).
- (26) Hauswirth, W. W.; Berns, K. I. *Virology* **1977**, *78* (2), 488.
- (27) Atchison, R. W.; Casto, B. C.; Hammon, W. M. D. *Science (80-.)*. **1965**, *149* (3685), 754.

- (28) Buller, R. M. L.; Janik, J. E.; Sebring, E. D.; Rose, J. A. *J. Virol.* **1981**, *40* (1), 241.
- (29) Urabe, M.; Nakakura, T.; Xin, K.-Q.; Obara, Y.; Mizukami, H.; Kume, A.; Kotin, R. M.; Ozawa, K. *J. Virol.* **2006**, *80* (4), 1874.
- (30) Ward, P.; Dean, F. B.; O'Donnell, M. E.; Berns, K. I. *J. Virol.* **1998**, *72* (1), 420.
- (31) Matsushita, T.; Elliger, S.; Elliger, C.; Podsakoff, G.; Villarreal, L.; Kurtzman, G. J.; Iwaki, Y.; Colosi, P. *Gene Ther.* **1998**, *5* (7), 938.
- (32) Wang, D.; Tai, P. W. L.; Gao, G. *Nature Reviews Drug Discovery*. Nature Publishing Group May 1, 2019, pp 358–378.
- (33) Kelemen, Rachel E.; Jewel, Delilah; Pasha, Muhammad; Cao, Xiaofu; Chatterjee, A. *Submiss.*
- (34) Xiao, X.; Li, J.; Samulski, R. J. *J. Virol.* **1998**, *72* (3), 2224.
- (35) Cohen, R. N.; van der Aa, M. A. E. M.; Macaraeg, N.; Lee, A. P.; Szoka, F. C. *J. Control. Release* **2009**, *135* (2), 166.
- (36) Carpentier, E.; Paris, S.; Kamen, A. A.; Durocher, Y. *J. Biotechnol.* **2007**, *128* (2), 268.
- (37) Glover, D. J.; Leyton, D. L.; Moseley, G. W.; Jans, D. A. *J. Gene Med.* **2010**, *12* (1), 77.
- (38) Jarvis, D. L. In *Methods in Enzymology*; Academic Press Inc., 2009; Vol. 463, pp 191–222.
- (39) Kost, T. A.; Condreay, J. P. *Trends in Biotechnology*. Elsevier Current Trends April 1, 2002, pp 173–180.
- (40) Ayres, M. D.; Howard, S. C.; Lopez-Ferber, M.; Lopez-Ferber, M.; Possee, R. D. *Virology* **1994**, *202* (2), 586.
- (41) Cheshenko, N.; Krougliak, N.; Eisensmith, R. C.; Krougliak, V. A. *Gene Ther.* **2001**, *8* (11), 846.
- (42) Kost, T. A.; Condreay, J. P.; Jarvis, D. L. *Nature Biotechnology*. Nature Publishing Group May 5, 2005, pp 567–575.
- (43) Barsoum, J.; Brown, R.; McKee, M.; Boyce, F. M. *Hum. Gene Ther.* **1997**, *8* (17), 2011.
- (44) Kaikkonen, M. U.; Rätty, J. K.; Airene, K. J.; Wirth, T.; Heikura, T.; Ylä-Herttuala, S. *Gene Ther.* **2006**, *13* (4), 304.
- (45) Chatterjee, A.; Xiao, H.; Bollong, M.; Ai, H. W.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (29), 11803.
- (46) Kotin, R. M. *Hum. Mol. Genet.* **2011**, *20* (R1), R2.
- (47) Urabe, M.; Ding, C.; Kotin, R. M. *Hum. Gene Ther.* **2002**, *13* (16), 1935.
- (48) Smith, R. H.; Levy, J. R.; Kotin, R. M. *Mol. Ther.* **2009**, *17* (11), 1888.

Chapter 3

Development of an alternate selection handle to evolve tRNAs in mammalian cells

3.1 Introduction

The concepts of directed evolution, especially the difficulties of performing selections directly in mammalian cells, have been outlined in detail in Chapter 1. Chapter 2 outlined the use of viruses as DNA delivery vectors to perform selections in mammalian cells. This chapter focuses on the use of an alternative selection handle in a viral protein to allow the evolution of tRNAs to incorporate UAAs that are not tolerated in the AAV capsid.

3.1.1 *VADER as a solution for evolving small gene products*

The development of directed evolution schemes in mammalian cells has historically been limited by the constraints of delivering DNA to mammalian cells. Several selection schemes have been developed recently for mammalian cells that use viruses to deliver library members to the host cells, but these techniques have relied on error-prone replication to generate variation in library members.¹⁻³ Somatic hypermutation is a technique that also solves the problem of delivering DNA to cells, as it does not require continued supplemental DNA delivery to cells for the selection process to occur.⁴ While these selection schemes work well for large gene products such as GPCRs and transcription factors, they are unsuitable for the evolution of small gene products because the random mutation rate necessary to drive variation in small genes such as tRNAs would be too high to permit the continued function of essential genes, as deleterious missense and nonsense mutations would accumulate in the viral or cellular genome. A new selection scheme is needed to evolve small gene products such as tRNAs.

VADER is a technique which solves the problem of delivering library members to mammalian host cells while not relying on random mutagenesis to generate variation in the library members.⁵ Since the selection scheme does not make use of random

mutagenesis, it can be used to perform directed evolution on tRNAs, which could not be accessed via somatic hypermutation or existing available virus-based selection systems. VADER is a powerful technique that has yielded multiple improved tRNAs that install lysine variants in response to TAG codons. These tRNAs improve the efficiency of UAA incorporation in mammalian cells through their enhanced activities and are excellent tools for the scientific community working on stop codon suppression.

3.1.2 The limitations of VADER: the AAV capsid

As it currently stands, the VADER scheme works well for evolving tRNAs to install UAAs that fit into the AAV capsid. However, the AAV capsid does not tolerate the incorporation of all UAAs. The AAV capsid is composed of 60 individual proteins that fit tightly and rigidly together. It is easy to perturb the complex manner in which the capsid proteins fit together, which results in unassembled capsids and no functional virions.^{6,7} It is also easy to modify the capsid protein structure in such a way that they can no longer perform the functions necessary for viral infectivity, such as by disrupting VP1's phospholipase activity that is necessary for endosomal escape or encapsulating the single-stranded DNA genome.⁸⁻¹¹ If a UAA does not fit into the capsid correctly, it could destroy the capsid's ability to assemble or the virion's ability to infect a cell. It is therefore not possible to use these UAAs as reactive handles in VADER for the bio-orthogonal pulldown, so researchers cannot perform directed evolution on the tRNAs to install these UAAs into proteins in mammalian cells.

5-hydroxy-L-tryptophan (5HTP, Figure 3.1) is one such UAA that is not tolerated in the AAV capsid (Figure 3.2), even at position 454, which is a highly permissive site for Bock and AzK incorporation.¹² It is likely that the shape of this UAA perturbs the assembly

of the capsid, even though residues in this position should be surface-exposed and not involved in the inter-protein junctions. Installing 5HTP at position 454 does not result in infective virions, so it is not possible to use this UAA in VADER. This situation is unfortunate, because 5HTP is a useful bioconjugation handle.¹³⁻¹⁵ The chemical biology community would benefit from the ability to evolve a tryptophanyl tRNA (WtR) to improve the incorporation of 5HTP in proteins in mammalian cells for the purposes of site-specific labeling. It is necessary to modify the VADER technique to access the WtR for the purposes of directed evolution.

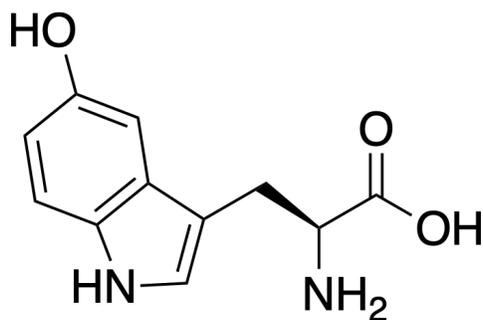


Figure 3.1. Structure of 5-hydroxytryptophan (5-HTP). 5HTP is one of the UAAs that is not tolerated in the AAV capsid.

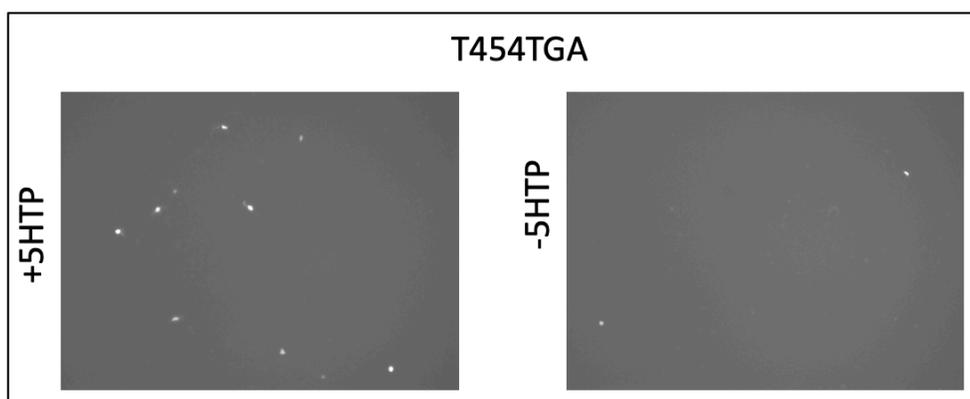


Figure 3.2. Incorporation of 5HTP into the AAV2 capsid. AAV2 does not tolerate 5HTP in its capsid, as evidenced by the lack of production of infectious virus even when provided with the UAA.

Delicate capsid assembly is not the only limitation to successful virion assembly during the VADER process. Capsid proteins are required at an extremely high level in order to produce enough virus to capture during selections. As I showed in Chapter 2 of this dissertation, production of capsid proteins containing UAAs is a limiting factor in being able to produce infectious virus in the VADER process. Being able to install a UAA into a protein selection handle which is required at much lower levels than are the capsid proteins would free researchers from the constraint of needing to produce such high levels of protein using tRNAs which may not yet be as active as native tRNAs. The ideal selection handle for UAAs such as 5HTP would be flexible, so in order to eliminate worries about structural perturbation, and it would be a catalytic protein, so that it would be required at much lower levels.

3.1.3 The ideal selection handle: Assembly Activating Protein

One protein in the AAV proteome that satisfies the requirements of being a structurally flexible catalyst is Assembly Activating Protein (AAP). The gene that codes for this protein is located within the *Cap* ORF, but within an alternate reading frame with a CTG start codon (Figure 3.3).¹⁶ During AAV2 replication, AAP is localized in the host nucleolus and targets newly-synthesized Cap proteins to that location, where capsid assembly takes place.¹⁶ AAP's highly structured N- and C-termini are relatively conserved between different AAV serotypes.^{17,18} The termini of AAP likely interact with the capsid proteins and act as scaffolds to promote capsid protein interaction along the axis of 2-fold symmetry.¹⁸

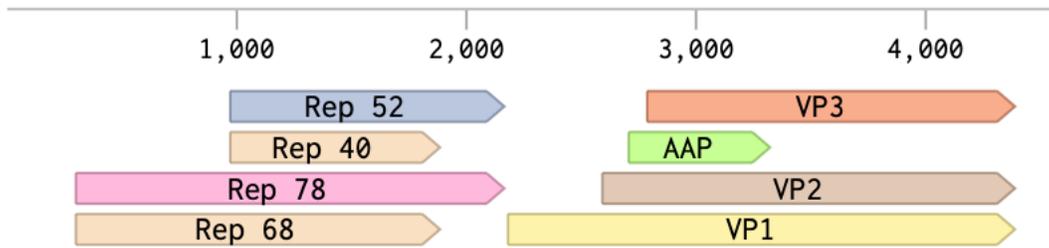


Figure 3.3. Structure of the AAV genome. The *Cap* gene in the AAV genome contains the genes for the three capsid proteins VP1, VP2, and VP3, as well as the gene for AAP in a different reading frame.

While AAP is required for AAV2 capsid assembly, it is likely needed at much lower levels than are the capsid proteins, as AAP acts catalytically to assemble proteins into the viral capsid.^{16,19} AAP's CTG start codon is relatively weak, which is another clue indicating that this protein is needed at lower levels. AAP interacts closely with the capsid proteins during the assembly process, but its termini are not as structurally rigid as the AAV capsid itself, and there are several threonine/serine rich regions in the protein that are quite flexible.¹⁸ This protein is therefore a strong candidate for the structurally flexible catalyst that would tolerate the incorporation of 5HTP.

3.1.4 A selection platform based on AAP

As described above, the gene that codes for AAP sits within the ORF for the capsid proteins, but in an alternative reading frame. This genetic organization permits a way to silence the expression of AAP without silencing the expression of the capsid proteins: the AAPstop60 platform.^{20,21} A mutation in residue 64 of AAP from TCA to TGA results in an S64TGA mutation in AAP (Figure 3.4). As AAP's gene is in a different reading frame than the genes for the capsid proteins, there is a silent GTC to GTG mutation at residue 239 of the VP proteins. Use of the AAPstop60 system has allowed researchers to silence AAP

without perturbing the expression of the capsid protein and to demonstrate that AAP is required to for the production of infectious AAV2 capsids.²⁰

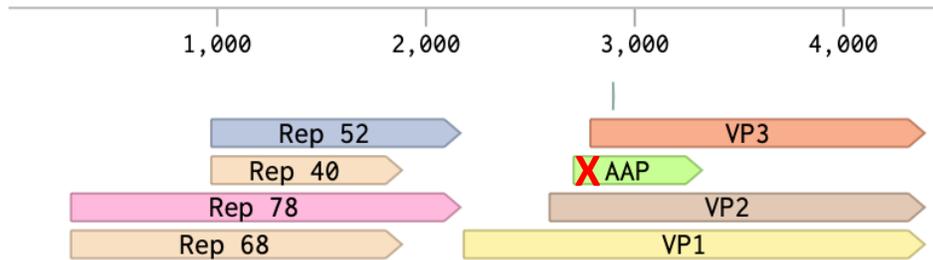


Figure 3.4. The structure of the AAPstop60 genome. A point mutation produces a S64TGA mutation in AAP while resulting in a silent mutation in the VP genes because these genes are in different reading frames within the same ORF.

This system will be useful for developing an AAP-based selection scheme. It is possible to silence the endogenous AAP in the AAV2 genome with this S64TGA mutation and then provide supplemental AAP *in trans* with a blank codon mutation of choice, as well as a tRNA that can suppress that blank codon. In this manner, it is possible to tie the production of infectious AAV2 to the tRNA's ability to incorporate a UAA into AAP during translation.

In order to make use of AAP in selection schemes to evolve tRNAs, it is necessary to find an optimum location to position a stop codon in the gene for AAP, where the UAA will be introduced. The structures of the termini of AAP are highly conserved between AAV serotypes,¹⁸ which indicates that these exact structures are essential for the protein's capsid assembly functions. Since these regions are essential, requiring the incorporation of a UAA in the N-terminus might prove to be a stringent selection, which could benefit the evolution of a highly active tRNA. If these regions are too structurally rigid to permit the incorporation of a UAA such as 5HTP, there are flexible regions downstream of the N-

terminus that are rich in threonine and serine residues, which tolerate large insertions which the AAV capsid does not.^{17,18} These T/S-rich regions might be ideal candidate regions to incorporate 5HTP into AAP.

In this chapter, I present a modified selection scheme in which I use AAP as a selection handle to evolve mutant WtRs to incorporate 5HTP in proteins in mammalian cells. I silence the expression of AAP from the AAV2 genome using the AAPstop60 method, and I provide wild-type AAP back *in trans* to demonstrate that it is possible to rescue production of infectious AAV using this supplemental AAP. I optimize the location of a TAG stop codon in AAP and show that UAA-containing AAP can still facilitate the production of infectious AAV2. By changing this TAG stop codon to TGA, I am able to incorporate 5HTP into AAP using the *Escherichia coli* tryptophanyl tRNA (EcWtR) with a UCA anticodon. Finally, I am able to use this platform to subject a library of EcWtR variants to positive selection and identify hits that incorporate 5HTP into proteins in mammalian cells at a highly improved level.

3.2 Results and Discussion

3.2.1 Producing infectious AAV with TAG-containing AAP

To determine whether it is possible to produce infectious AAV when there is a stop codon in AAP, I transfected cells with plasmids containing an AAV2 genome with an S64TGA mutation in AAP that is silent in the *Cap* genes and provided supplemental either wild type or L12TAG AAP *in trans* under the control of the strong CMV promoter. The AAV produced in this way has wild type capsids, but its production depends on UAA incorporation (in this case, BocK) into AAP. I compared these results to those achieved

with the VADER plasmid system, which relies on a T454TAG mutation in the capsid proteins (Figure 3.5).

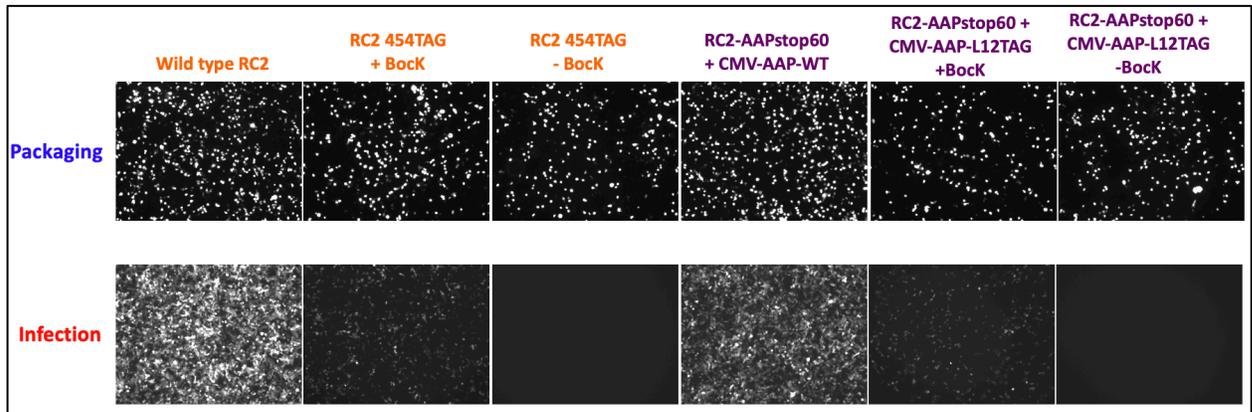


Figure 3.5. Producing AAV dependent on supplemental AAP. AAV was produced using the AAPstop60 system, providing supplemental AAP *in trans*. Production of AAV in this way was compared to production of AAV using the RC2-454TAG system from VADER. After 72 hours, the packaging cells were imaged and lysed, and new cells were infected with the lysate. These cells were imaged after 48 hours.

The plasmid system producing rAAV using *Cap* genes with the T454TAG mutation produced 7.44% of the infectious virus when supplemented with Bock as did the plasmid system using wild type *Cap* genes, as measured by fluorescence imaging and GFP fluorescence. When wild type AAP was provided *in trans* along with the AAPstop60 system, cells produced 57.5% of the infectious virus as made with wild type AAV2 genes (Figure 3.6). With the L12TAG in AAP, cells only produced 6.39% of the virus as they had with supplemental wild type AAP, which is not an improvement over the *Cap* T454TAG system. Even though AAP is likely used catalytically in capsid assembly, it appears that the low levels of UAA-containing AAP produced using this system were not sufficient to assemble infectious AAV capsids at levels approaching that when wild type AAP is provided.

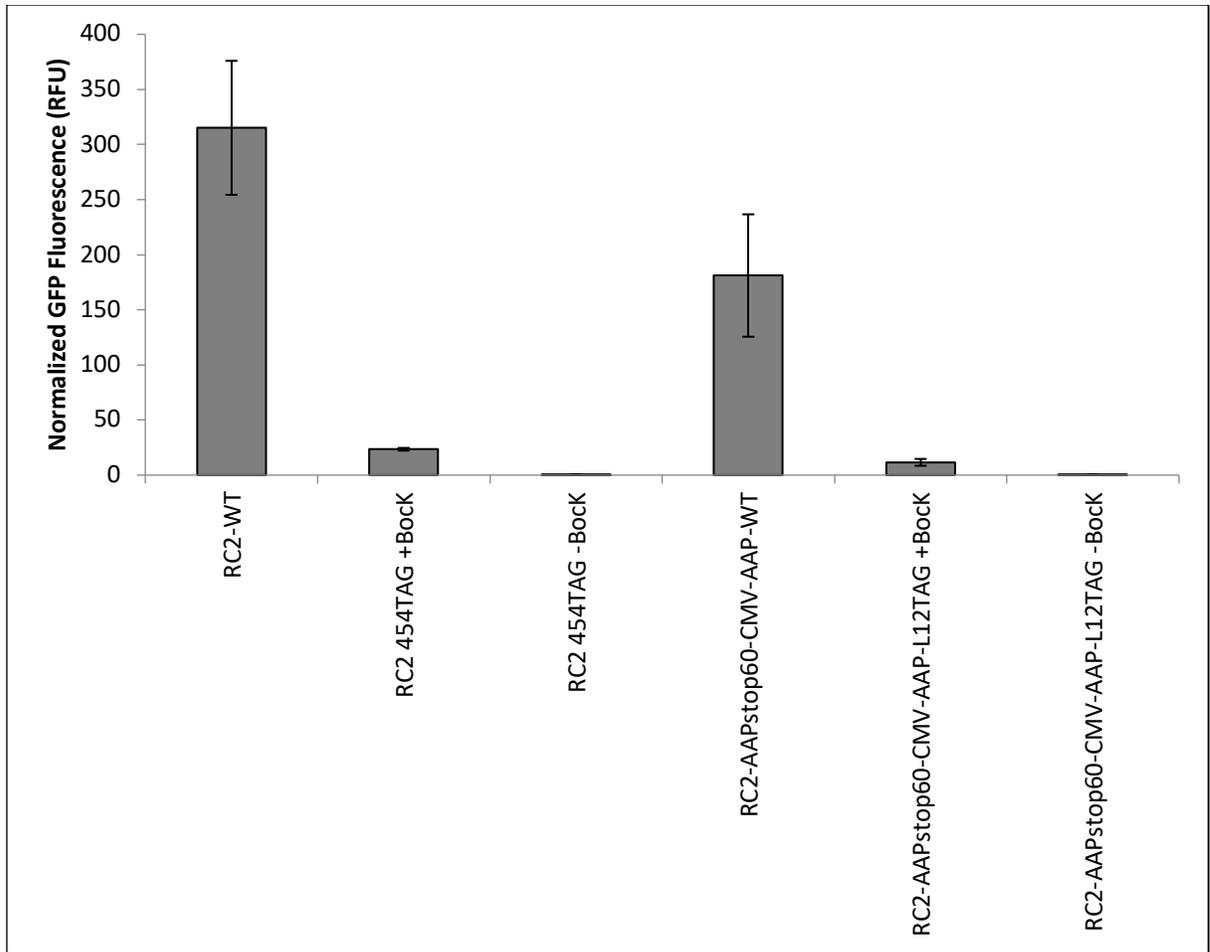


Figure 3.6. GFP fluorescence of cells infected with AAV produced using the AAPstop60 system. Cells were infected with rAAV produced by packaging cells to test the effectiveness of the AAPstop60 plus supplemental AAP to produce rAAV with wild type capsids. After 48 hours, these cells were imaged and lysed, and GFP fluorescence was taken.

3.2.2 Determining the optimum site for a TAG mutation in AAP

While it is possible to produce some infectious AAV by installing a UAA in the N-terminus of AAP, there may be other locations within the protein that are better suited to UAA incorporation. To that end, I created supplemental AAP variants with TAG mutations in the flexible T/S-rich regions in the protein, as these regions are highly tolerant of mutations and insertions,¹⁷ while the termini of the protein are much more conserved

between serotypes and are not mutation-tolerant.^{17,18} The mutations tested with supplemental AAP provided *in trans* were: T78TAG, T97TAG, T110TAG, T124TAG, and T177TAG (Figure 3.7).

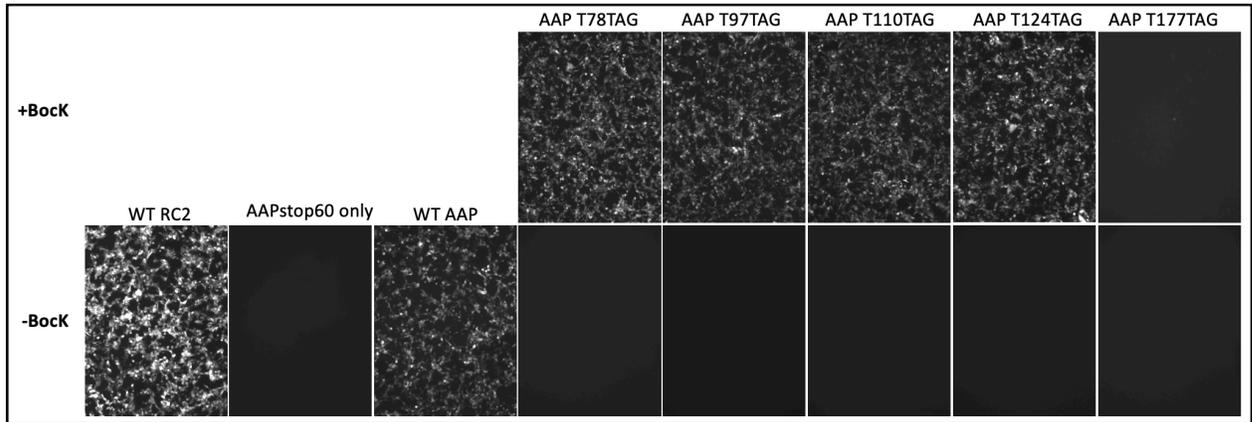


Figure 3.7. Producing AAV using AAP with TAG mutations. AAV was produced using the AAPstop60 system, providing supplemental AAP with TAG stop codons in different locations in the T/S rich regions. These cells were infected with lysate from cells that packaged rAAV and were imaged after 48 hours post-infection.

For the T78TAG, T97TAG, T110TAG, and T124TAG mutations, the amount of infectious AAV produced with the addition of BocK was approximately equal to that produced when wild type AAP was provided, as evidenced by fluorescence images and GFP fluorescence readings (Figure 3.8). AAP with the T177TAG mutation produced negligible amounts of infectious virus. These surprising and encouraging results indicate that it is possible to tie the production of infectious AAV dependent on stop codon suppression in AAP and still achieve reasonably high AAV titers.

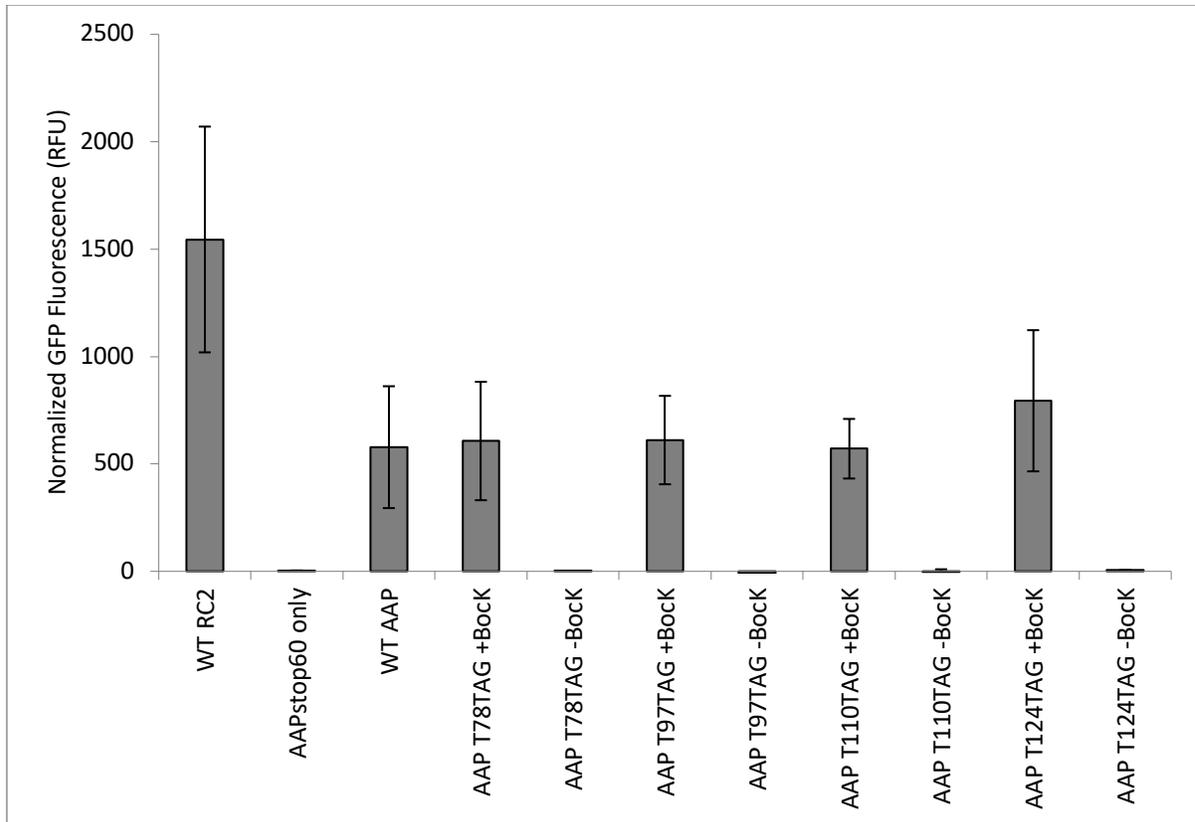


Figure 3.8. GFP fluorescence of cells infected with AAV produced with AAP TAG variants. Cells were infected with rAAV produced by packaging cells to determine the optimum site for a TAG mutation. After 48 hours, these cells were imaged, and GFP fluorescence was taken.

3.2.3 *Incorporating 5-hydroxytryptophan into AAP*

Having determined that it is possible to make the production of AAV dependent on stop codon suppression in AAP, I moved to test whether it was possible to make the production of AAV dependent on the incorporation of 5-hydroxytryptophan in AAP. The AAV capsid does not tolerate 5HTP incorporation, but because AAP tolerates mutations and insertions in its highly flexible T/S rich regions, it is likely that it would tolerate the mutation of one amino acid to 5HTP at one of these sites: T97TGA, T110TGA, and T124TGA (Figure 3.9).

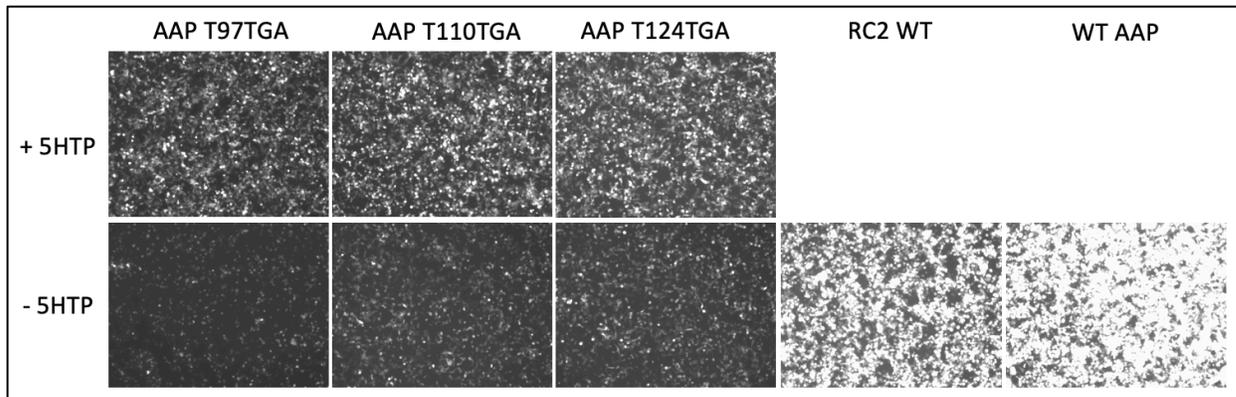


Figure 3.9. Infection of cells with virus produced using AAP with TGA mutations. AAV was produced using the AAPstop60 system, with supplemental AAP TGA mutants provided on the same plasmid for improved DNA delivery. The above cells were infected with lysate from cells that packaged rAAV and were imaged after 48 hours post-infection.

In this experiment, I provided supplemental AAP on the same plasmid as the mutant RC2-AAPstop60 AAV genome. Addition of wild type AAP rescued infective AAV production to wild type levels in this case, while it had not done so when provided on a different plasmid *in trans* in the previous experiments with AAPtag, likely due to the random delivery of DNA which occurs with transient transfection. The stop codon in the supplemental AAP was changed to TGA, as the EcWtR anticodon for stop codon suppression is UCA. Here, infective AAV production reached only approximately 10% of that achieved with wild type AAP, which is reasonable to expect because TGA suppression is not as efficient as TAG suppression. All three TGA mutants tested performed similarly in the presence of 5HTP, with little AAV production in the absence of 5HTP (Figure 3.10). Since all three mutants performed similarly well, I chose to move forward with the T97TGA mutant in selections for more efficient tryptophanyl tRNAs.

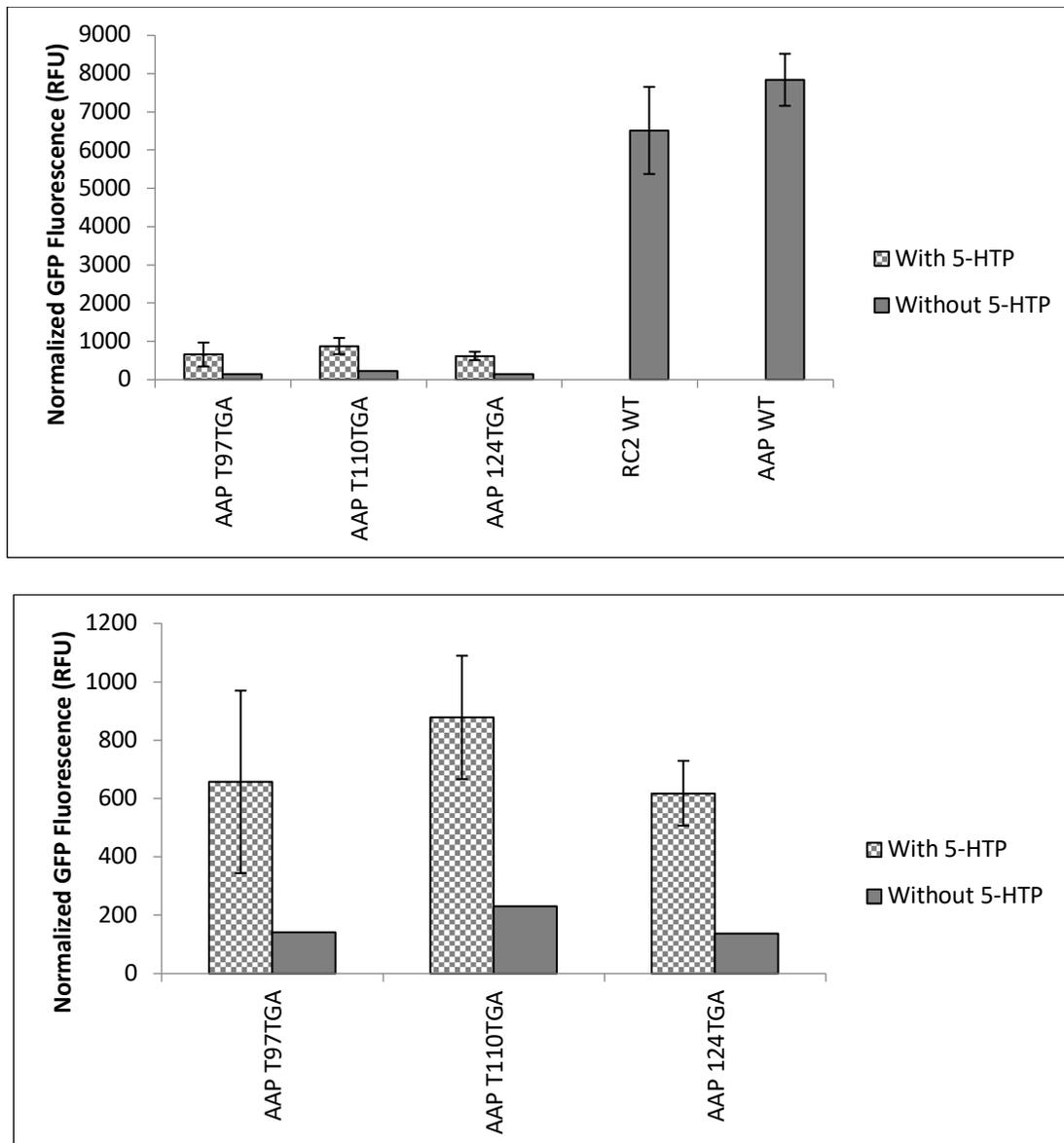


Figure 3.10. GFP fluorescence of cells infected with AAV produced with AAP TGA variants. Cells were infected with rAAV produced by packaging cells to determine whether AAP could tolerate insertion of 5HTP at various sites. After 48 hours, these cells were imaged, and GFP fluorescence was taken.

3.2.4 *Mock selections to determine if VADER is possible with AAP as a selection handle*

Before proceeding with selections on real libraries to evolve the *E. coli* tryptophanyl tRNA, it was necessary to test the selection parameters with a mock selection. To that end, I infected HEK-293T cells with rAAV containing as cargo GFP and the

MmPytR with a CUA anticodon at an MOI of 5 and rAAV containing as cargo mCherry and the EcWtR with a UCA anticodon at an MOI of 0.05 for a ratio of 100:1 PytR to WtR. I provided 5HTP in the cell media and the other genes necessary to make AAV via transient transfection, as established in the original VADER protocol, with the plasmids pHelper and pIDTSMART-RC2-AAPstop60-CMV-AAP97tga-CMV-WRS. The WRS should not charge the PytR, and the PytR should not incorporate any amino acid at the TGA stop codon in AAP. In the mock selection, the PytR mimics any inactive tRNA library member, while the WtR should be active.

The production of virus in each round of this selection scheme depends on the incorporation of 5HTP into AAP, but not into the capsid of the virus, as the structure of the viral capsid does not permit the disruption caused by the introduction of 5-HTP. Therefore, the viral capsids produced from this selection scheme are wild type capsids, and it is not possible to perform a pulldown to capture viruses that contain only active but not cross-reactive tRNAs. This version of the VADER selection scheme will rely only on positive selections to continuously enrich active tRNAs.

After harvesting the output virus from one round of the mock selection, I infected cells with the output virus as well as the original input virus and analyzed the ratio of green to red cells by flow cytometry. This ratio will indicate the ratio of PytR to WtR, as the genes for fluorescent reporter and tRNA are packaged together in the genome of the rAAV. The ratio of green to red cells from the input virus was 101:1, which demonstrates that the original calculations for infections were accurate. For the output virus, the ratio of green to red cells was 0.823:1, for a 121-fold enrichment of mCherry and WtR-containing rAAV (Figure 3.11). This result indicates that it is possible to achieve at least 100-fold enrichment

of active tRNAs after just one round of positive selection, which will allow researchers to evolve the previously un-evolvable WtR in mammalian cells.

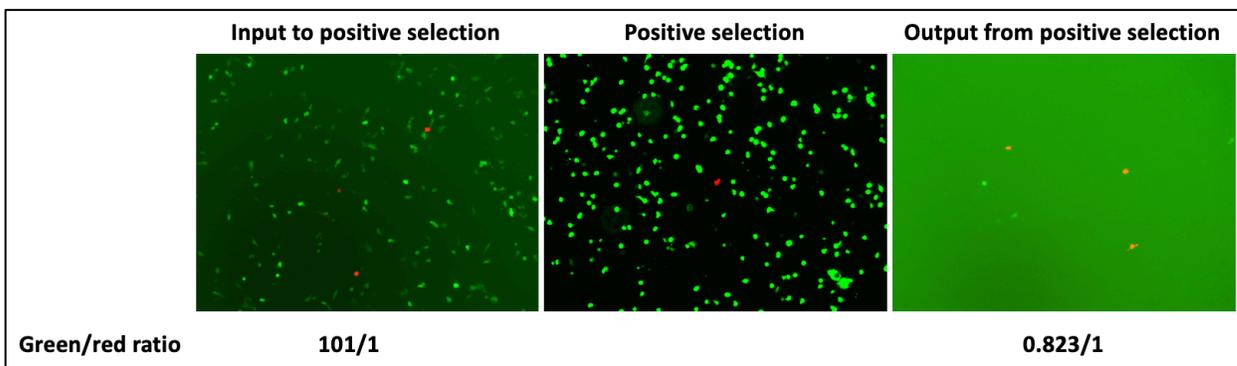


Figure 3.11. Enrichment of WtR+mCherry over PytR+GFP in a mock selection. An inactive tRNA (PytR) was provided to cells at an MOI 100 times that of the active tRNA (WtR). These tRNAs underwent a mock selection in HEK-293T cells. The resulting output virus was used to infect new cells. Cells infected with the input and output virus mixes were analyzed by FACS to determine the ratio of green (PytR) to red (WtR) viruses.

3.3 Conclusion

With this result in hand, it will be possible to evolve libraries of WtR variants to efficiently install 5HTP in proteins in mammalian cells. The selection scheme here will only involve positive selections, but with over 100-fold enrichments of active tRNAs over inactive ones, repeated rounds of positive selection should yield highly active WtR variants. This tool should enable researchers to achieve high yields of 5HTP-containing proteins, which aid the development of novel bioconjugation chemistries.

3.3.1 *Ongoing work and future directions*

A library of WtR variants with mutations in the acceptor stem has been developed and will be subject to selection via the method developed here. Because there is no negative selection step in this selection scheme, all selections will be repeated without adding 5HTP, so that if a tRNA variant is cross-reactive with a native aaRS and canonical amino acid, it has the opportunity to be charged with that amino acid as well. High-throughput

sequencing will make it possible to distinguish between variants that are active but not cross-reactive and variants that are active and cross-reactive.

AAP can also be used for a variety of future applications for selections in mammalian cells. For example, split halves of AAP can be fused to bait and potential prey proteins in a two-hybrid system. Since AAP can tolerate the insertions of large proteins in its Thr/Ser rich regions,¹⁷ it is likely that it will still perform its functions with the addition of the associated bait and prey proteins in this region. This two-hybrid selection can use the production of AAV as the reporter to screen the mammalian proteome for binding partners.

3.4 Acknowledgments

Dr. Rachel Kelemen and Dr. James Italia gave me access to their plasmid stocks. Dr. Kelemen also trained me to perform selections using rAAV. Rachel Huang provided invaluable assistance in constructing the WtR library that will be used for future selections.

3.5 Experimental Procedures

3.5.1. Strains and cell lines

Clonings and plasmid propagations were performed in electrocompetent DH10B *E. coli* cells.

HEK-293T (ATCC) cells were maintained as described in Chapter 2.

3.5.2. Plasmids

All plasmid maps and sequences can be found in the Appendix.

The plasmids pAAV-GFP+1xMmPytR, pAAV-GFP+1xEcWtRtga, pAAV-mCherry+1xEcWtRtga, pIDTSMART-8xPytR-ITR-GFP, pHelper, pIDTSMART-

MbPylRS-RC2wt, and pIDTSMART-MbPylRS-RC2-454TAG were propagated from Dr. Rachel Kelemen's stocks.

To construct pIDTSMART-MbPylRS-RC2-AAPstop60, I introduced the mutation into the *Cap* gene using the primers AAPstop60-F and AAPstop60-R. I amplified this *Cap* containing the mutant gene for AAP using the terminal primers RC2int-HindIII-F and RC2-SbfI-R and cloned it back into the vector using the HindIII and SbfI cut sites.

I introduced a Kozak sequence to the 5' end of the gene for AAP and amplified the construct using the primers NheI-Kozak-AAP-WT-F and XhoI-AAP-R. To generate pIDTSMART-CMV-AAPwt, I cloned the resulting cassette into a pIDT vector using the NheI and XhoI restriction sites. I used mutagenesis primers to generate the following variants for the pIDTSMART-CMV-AAP plasmid: L12TAG, T78TAG, T97TAG, T110TAG, T124TAG, T177TAG, T97TGA, T110TGA, and T124TGA.

To generate pIDTSMART-MbPylRS-RC2-AAPstop60-CMV-AAP (both wild type and T97TGA variants), I cut the plasmids pIDTSMART-CMV-AAP (either wild type AAP or T97TGA) and pIDTSMART-MbPylRS-RC2-AAPstop60 with NheI and SphI and ligated the AAP construct into the AAPstop60 vector.

The *E. coli* tryptophanyl aaRS was amplified out of pBK-MCS-EcWRS from Dr. James Italia's stock using the primers AvrII-CMV-WRS-F and SpeI-WRS-R. It was then cloned into the vector pIDTSMART-MbPylRS-RC2-AAPstop60-CMV-AAP-97TGA which had been cut with AvrII.

3.5.3. Unnatural amino acids

BocK was purchased from Chem-Impex International (catalog number 00363). 5HTP was purchased from Chem-Impex International (catalog number 00607).

3.5.4 Preparing recombinant AAV

For details on producing recombinant AAV, see Chapter 2 of this thesis.

3.5.5 Producing infectious AAV with TAG-containing AAP

To determine whether it is possible to produce infectious AAV whose packaging depends on stop codon suppression in AAP, HEK-293T cells were transfected with the following sets of plasmids: pHelper and pIDTSMART-MbPylRS-RC2wt; pHelper, pIDTSMART-MbPylRS-RC2-454TAG, and pAAV-mCherry+1xMmPylR; pHelper, pIDTSMART-MbPylRS-RC2-AAPstop60, and pIDTSMART-AAPwt; or pHelper, pIDTSMART-MbPylRS-RC2-AAPstop60, pIDTSMART-AAP-L12TAG, and pAAV-mCherry+1xMmPylR. N^ε-Boc-L-Lysine (BocK) was added to the cell media to test the UAA-dependence of AAP production. Cells were lysed with CellLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease. The resultant lysate was flash-frozen and stored at -80°C. Lysate was thawed and used to infect new HEK-293T cells with sodium butyrate to 1.0 mM. After 48 hours, cells were imaged and lysed, and GFP fluorescence was measured.

3.5.6 Determining the optimum site for a TAG mutation in AAP

To find an improved location for the TAG mutation in AAP, cells were transfected with the following sets of plasmids: pHelper and pIDTSMART-MbPylRS-RC2wt; pHelper and pIDTSMART-MbPylRS-RC2-AAPstop60; pHelper, pIDTSMART-MbPylRS-RC2-AAPstop60, and pIDTSMART-AAPwt; or pHelper, pAAV-Cargo-GFP+1xMmPylR, pIDTSMART-MbPylRS-RC2-AAPstop60, and pIDTSMART-AAPtag with one of the following mutations: T78TAG, T97TAG, T110TAG, T124TAG, or T177TAG. N_α-Boc-L-Lysine (BocK) was added to the cell media to test the UAA-dependence of AAP

production. Cells were lysed with CelLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease. The resultant lysate was flash-frozen and stored at -80°C. Lysate was thawed and used to infect new HEK-293T cells with sodium butyrate to 1.0 mM. After 48 hours, cells were imaged and lysed, and GFP fluorescence was measured.

3.5.7 Incorporating 5-hydroxytryptophan into AAP

Cells were transfected with the following sets of plasmids: pHelper and pIDTSMART-MbPylRS-RC2wt; pHelper, pIDTSMART-MbPylRS-RC2-AAPstop60, and pIDTSMART-AAPwt; pHelper, pAAV-GFP+1xExWtRtga, and pIDTSMART-RC2-AAPstop60-CMV-AAP-CMV-WRS with one of the following mutations in AAP: T97TGA, T110TGA, or T124TGA. 5-hydroxy-L-tryptophan (5HTP) was added to the cell media to test the UAA-dependence of AAP production. Cells were lysed with CelLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease. The resultant lysate was flash-frozen and stored at -80°C. Lysate was thawed and used to infect new HEK-293T cells with sodium butyrate to 1.0 mM. After 48 hours, cells were imaged and lysed, and GFP fluorescence was measured.

3.5.8 Mock selections to determine if VADER is possible with AAP as a selection handle

HEK-293T cells were simultaneously infected with 5 MOI of rAAV-GFP-PytR and 0.05 MOI of rAAV-mCherry-WtR. These cells were transfected with pHelper and pIDTSMART -RC2-AAPstop60-CMV-AAP-CMV-WRS, and 5HTP was added to the cell media to 1.0 mM. Output virus was harvested, PEG precipitated, and flash frozen for storage at -80°C.

Fresh HEK-293T cells were infected with the input virus (calculated to deliver rAAV-GFP-PytR and rAAV-mCherry-WtR to the selection cells at a ratio of 100:1) or the

virus harvested after selection, with sodium butyrate to 1.0 mM. These infected cells were analyzed using flow cytometry to determine the ratio of green to red cells, which would correspond to the ratio of cells infected by rAAV-GFP-PytR to cells infected by rAAV-mCherry-WtR.

3.6 References

- (1) English, J. G.; Olsen, R. H. J.; Lansu, K.; Strachan, R. T.; Wacker, D.; Roth, B. L. *Cell* **2019**, *178*, 748.
- (2) Berman, C. M.; Papa, L. J.; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; Hoeben, R. C.; Shoulders, M. D. *J. Am. Chem. Soc.* **2018**, *140* (51), 18093.
- (3) Das, A. T.; Zhou, X.; Vink, M.; Klaver, B.; Verhoef, K.; Marzio, G.; Berkhout, B. *J. Biol. Chem.* **2004**, *279* (18), 18776.
- (4) Wang, L.; Tsien, R. Y. *Nat. Protoc.* **2006**, *1* (3), 1346.
- (5) Kelemen, Rachel E.; Jewel, Delilah; Pasha, Muhammad; Cao, Xiaofu; Chatterjee, A. *Submiss.*
- (6) 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. *J. Virol.* **2003**, *77* (20), 11072.
- (7) Grieger, J. C.; Snowdy, S.; Samulski, R. J. *J. Virol.* **2006**, *80* (11), 5199.
- (8) Bleker, S.; Sonntag, F.; Kleinschmidt, J. A. *J. Virol.* **2005**, *79* (4), 2528.
- (9) Wu, P.; Xiao, W.; Conlon, T.; Hughes, J.; Agbandje-McKenna, M.; Ferkol, T.; Flotte, T.; Muzyczka, N. *J. Virol.* **2000**, *74* (18), 8635.
- (10) Sonntag, F.; Bleker, S.; Leuchs, B.; Fischer, R.; Kleinschmidt, J. A. *J. Virol.* **2006**, *80* (22), 11040.
- (11) Girod, A.; Wobus, C. E.; Zádori, Z.; Ried, M.; Leike, K.; Tijssen, P.; Kleinschmidt, J. A.; Hallek, M. *J. Gen. Virol.* **2002**, *83* (5), 973.
- (12) Kelemen, R. E.; Mukherjee, R.; Cao, X.; Erickson, S. B.; Zheng, Y.; Chatterjee, A. *Angew. Chemie Int. Ed.* **2016**, *55* (36), 10645.
- (13) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. *J. Am. Chem. Soc.* **2017**, *139* (34), 11670.
- (14) Sarathi Addy, P.; Italia, J. S.; Chatterjee, A. *ChemBioChem* **2018**, *19* (13), 1375.
- (15) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. In *Bioconjugation, Methods in Molecular Biology*; Humana Press Inc., 2019; Vol. 2033, pp 239–251.
- (16) Sonntag, F.; Schmidt, K.; Kleinschmidt, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (22), 10220.
- (17) Tse, L. V.; Moller-Tank, S.; Meganck, R. M.; Asokan, A. *J. Virol.* **2018**, *92* (14).
- (18) Naumer, M.; Sonntag, F.; Schmidt, K.; Nieto, K.; Panke, C.; Davey, N. E.; Popa-Wagner, R.; Kleinschmidt, J. A. *J. Virol.* **2012**, *86* (23), 13038.
- (19) Samulski, R. J.; Muzyczka, N. *Annu. Rev. Virol.* **2014**, *1* (1), 427.
- (20) Maurer, A. C.; Pacouret, S.; Karla, A.; Diaz, C.; Blake, J.; Andres-Mateos, E.; Vandenberghe, L. H. *Cell Reports* **2018**, *23*, 1817.
- (21) Maurer, A. C.; Cepeda Diaz, A. K.; Vandenberghe, L. H. *J. Virol.* **2019**, *93* (7), 2013.

Chapter 4

Development of an optimized baculovirus vector for incorporation of lysine analogs

4.1 Introduction

Chapter 2 of this thesis examined improvements to the VADER selection protocol using baculovirus vectors. This chapter will focus on the development of optimized baculovirus vectors to improve the incorporation of a variety of UAAs in mammalian cells, using tRNAs evolved using the VADER selection platform.

4.1.1 Different methods to deliver genetic code expansion machinery to mammalian cells

There are multiple vector types for expressing the machinery needed to install UAAs in proteins in mammalian cells. Each vector has benefits and drawbacks to its use, especially in the context of delivering small gene products such as tRNAs, which may be required at higher copy numbers to improve stop codon suppression.

4.1.1a Transient transfection

Because mammalian cells exhibit poor transformation efficiency,¹ the most commonly used method to deliver DNA to mammalian cells is transient transfection. Transfection is a simple procedure to perform in mammalian cells, as the only materials required are purified plasmids and transfection reagents. Researchers can deliver the orthogonal aaRS/tRNA pair and the gene of interest containing a stop codon, using one or two plasmids, and achieve relatively high transfection efficiencies. However, if the experiment requires more extensive genetic cargo, such as any work involving the CRISPR-Cas system, it may be necessary to use additional plasmids to deliver all of the cargo to the host cells. Since transfection efficiency drops with each additional plasmid used, and since DNA distribution to host cell nuclei during transient transfection is highly heterogeneous,²⁻⁴ using this method may not result in optimum UAA incorporation in the target protein.

4.1.1b AAV

Viral vectors are attractive delivery vehicles for transducing mammalian cells. It is possible to infect cells with precise copy numbers of viruses, making delivery of DNA much more homogenous than with transient transfection. Modern molecular biology techniques have rendered facile the production of high titers of recombinant viruses for use in complex experiments.

AAV is one of the most popular viral vectors for use in research and clinical applications. This virus exhibits low immunogenicity due to its small size, as well as low cytotoxicity because it cannot replicate without co-infection of a helper virus.⁵⁻⁷ AAV is therefore an excellent choice for a DNA delivery vector due to this favorable biosafety profile for researchers and patients. Its single-stranded DNA genome is relatively stable, but it is small at approximately 5 kb, in comparison to other viruses used to deliver DNA to mammalian cells.⁸ AAV's capsid is rigid and cannot tolerate inclusion of additional genetic material, so if an experiment requires the delivery of large amounts of genetic cargo, it will be necessary to use multiple rAAV vectors.^{9,10}

One additional drawback to using AAV as a vector is particularly problematic for genetic code expansion experiments. Replication of AAV's single-stranded DNA genome relies on Rep proteins binding to the ITRs.^{11,12} RNA sequences in the AAV genome can fold in on themselves to make hairpins, and Rep proteins can "mistake" these hairpins for the 5' ITR and truncate the genome during replication.^{13,14} These truncations can excise key components of the rAAV cargo, and are much more likely to occur when the cargo contains tRNAs for use in stop codon suppression or sgRNAs for use in CRSIPR experiments. While AAV is an excellent vector for delivering proteins of interest to

mammalian cells, its tendency towards truncation with RNA cargos limits its use for genetic code expansion applications.

4.1.1c Lentivirus

Lentiviral vectors have increased in popularity as DNA delivery vehicles due to their large cargo capacity and ability to integrate into the mammalian genome. Retroviruses use reverse transcriptase to replicate their RNA genomes, so the DNA daughter strands can integrate into genomic DNA to produce stable cell lines. This feature is attractive for researchers because stably integrated genes can offer more homogenous and robust expression than can transient transfection. Since lentiviruses such as HIV-1 are highly pathogenic to humans, it is essential for researcher safety that all vectors used be replication-deficient. Recombinant lentivirus vectors are typically pseudotyped with VSV-G instead of the viral envelope protein Env to expand the viral tropism.¹⁵ Third-generation lentivirus vectors split the replication-deficient HIV-1 genome into four plasmids to reduce the probability of recombination that could create replication-competent viruses.¹⁶ While this third-generation system is safer for human researchers and patients receiving these vectors for gene therapies, it faces the same limitation of decreased transfection efficiency yielding lower viral titers than systems that rely on fewer plasmids.

Lentivirus has been used for genetic code expansion applications in mammalian cells. Expression of UAA-containing proteins appears to improve when the lentiviral vectors are used to stably integrate the aaRS/tRNA pair into the genome than when the cargo is delivered via transient transfection.,¹⁷⁻¹⁹ but genomic integration can lead to disruption of essential host cell genes. Additionally, the risk of recombination in lentiviruses is still high with their single-stranded RNA genomes,²⁰ so it may be difficult

to reliably produce recombinant lentivirus containing tRNA cassettes to use for genetic code expansion.

4.1.1d Baculovirus

Baculovirus vectors solve the problems posed by transient transfection and by AAV and lentivirus vectors. As is the case with AAV and lentivirus, infection with baculovirus results in much more homogenous transgene expression than does transient transfection.²¹ Baculovirus is an insect virus that has been pseudotyped with VSV-G to allow it to infect mammalian cells,²²⁻²⁴ its genome is silent in human cells, so it is safe for researchers to use.²⁵ These features also make baculovirus a useful vector for *in vivo* applications.

Unlike AAV, baculovirus has a large genome with a potentially unlimited cargo capacity because of the mechanism of genome packaging involves continuous assembly of capsid proteins around the DNA until it is fully enclosed.^{26,27} This massive cargo capacity allows researchers to produce recombinant baculovirus containing all of the genes needed for complex experiments, including genetic code expansion. Baculovirus's large, double-stranded DNA genome is quite stable and can carry highly repetitive cassettes of tRNAs without any evidence of recombination.²¹ Because the virus genome itself is quite large (approximately 134 kb), it is not practical to clone transgenes directly into it.²⁶ Instead, transgenes can be cloned into shuttle plasmids which are transformed into DH10Bac cells to undergo recombination with a bacmid that contains the rest of the baculovirus genome.^{28,29} Researchers can choose to transfect the shuttle plasmids directly into host cells, which is a convenient procedure, or to produce infective baculovirus, which will transduce cells more homogeneously. These features (large and stable genome, safe use, and

flexibility of use) make baculovirus an ideal vector to deliver the genetic machinery needed for unnatural amino acid incorporation in mammalian cells.

4.1.2 *A new baculovirus vector for the pyrrolysyl system*

Previous work has produced baculovirus vectors for tyrosyl and pyrrolysyl systems to incorporate tyrosine and lysine derivatives in proteins in mammalian cells.^{21,24} The baculovirus vector pAcBac3 in particular optimizes incorporation of O-methyltyrosine (Figure 4.1) over the original pAcBac2 system, with 20 copies of the suppressor tyrosyl tRNA (YtR) and UbiC promoter for the aaRS (Figure 4.2).²¹ Since the pyrrolysyl system is popular for genetic code expansion applications in mammalian cells (due to the ease of evolving synthetases in bacteria that will be naturally orthogonal in eukaryotes),^{30,31} it would be beneficial for the scientific community to have an optimized baculovirus vector for the pyrrolysyl system.

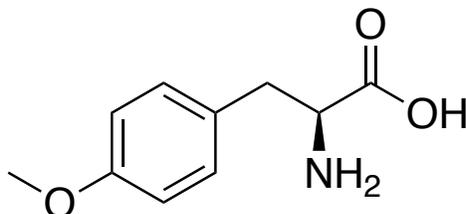


Figure 4.1. Structure of O-methyltyrosine. The baculovirus vector pAcBac3 was able to significantly improve the incorporation of O-methyltyrosine in proteins in mammalian cells.

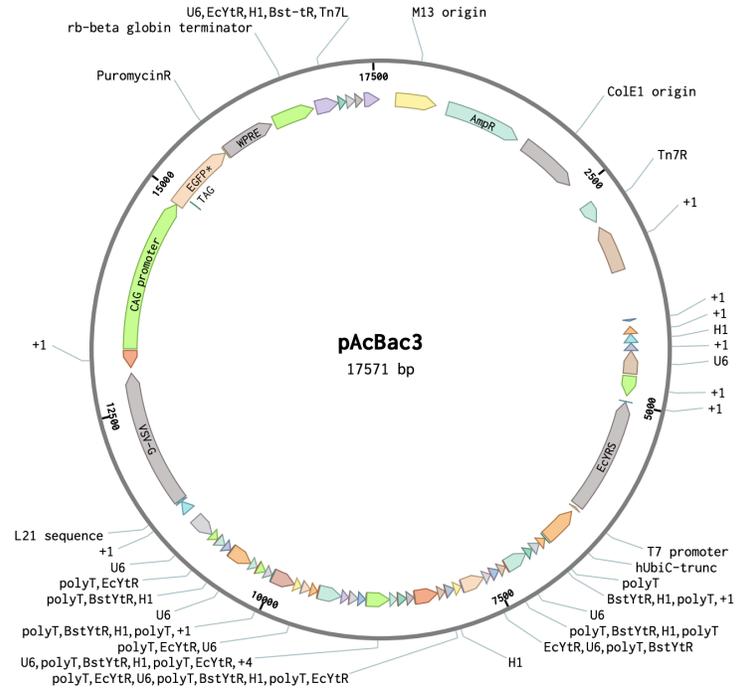


Figure 4.2. Map of pAcBac3 baculovirus vector.²¹ 20 copies of the tRNA cassette (distributed over three locations within the vector, one of which contains 16 tandem repeats of the tRNA) rescues incorporation of O-methyltyrosine in mammalian cells.

As previously discussed, the VADER selection scheme has yielded tRNAs that install lysine variants in response to TAG stop codons much more efficiently.³² Two of these variants, A2 and HTS25, have activities that improve the incorporation of AzK several-fold over that of the wild type PytR, in experiments where they are delivered via transient transfection (HTS25: unpublished data). Since infection by baculovirus results in much more homogenous host cell transduction than does transient transfection,²¹ a baculovirus vector carrying these improved PytR variants would improve stop codon-containing transgene expression both by increased DNA delivery to the nucleus and by enhanced tRNA activity. This enhanced baculovirus vector will be highly beneficial to the scientific community.

Even though baculovirus vectors have the potential to be the standard vector for mammalian genetic code expansion, their use has not been broadly adopted by researchers in this field. The development of a baculovirus vector that contains evolved PytR variants, as well as genes for an evolved synthetase and stop codon-containing protein of interest that can be easily exchanged via molecular cloning, has the potential to popularize the use of these versatile vectors.

In this chapter, I demonstrate that delivering tRNAs by infecting cells with baculovirus allows researchers to deliver low copy numbers of tRNAs to cells and to see improved activity in an evolved PytR variant that is not evident when the cells receive more copies of the tRNAs. In addition, I develop an enhanced baculovirus vector using another, more efficient PytR variant. I determine the optimum promoter to use with the MbPylRS and whether a nuclear export sequence improves UAA incorporation in this system. Finally, I demonstrate that this vector can be used with a number of different evolved MbPylRS variants to incorporate a variety of UAAs into proteins in mammalian cells.

4.2 Results and Discussion

4.2.1 Comparing A2 PytR variant to wild type PytR

Previously, Kelemen et al. evolved a PytR variant A2 that outperforms the wild type PytR by 300%, as demonstrated via transient transfection in HEK-293T cells.³² While this demonstrated improvement is considerable, it is possible that A2's actual improved efficiency would be even more evident were the tRNAs delivered at lower copy numbers. With transient transfection, cells that receive DNA may receive hundreds of copies of the tRNA, and the difference in performance between tRNAs may not be evident when there are so many copies of a less efficient tRNA. I generated baculovirus vectors to carry the

tRNAs so that I could precisely control the copy numbers delivered to cells, as well as ensure a homogenous distribution of infected cells.

While infective titer by FACS is more precise than by using the Clontech BacPAK™ Baculovirus Rapid Titer Kit, meaning that it is possible to deliver precise copy numbers of baculovirus to target cells, there is still variation in expression of the genes these viruses encode. To control for this variation in gene expression from the MTH-mCherry and tRNA containing viruses, I analyzed all GFP expression data by comparing it to the mCherry expression data for that particular treatment. Results are therefore given as normalized GFP and mCherry fluorescence in RFU, and also as GFP/mCherry ratios.

Expression of eGFP from the pAcBac-CMV-MbPylRS-CMV-GFPtag39 virus reached approximately 10% that of wild type eGFP expression from pAcBac-CMV-MbPylRS-CMV-GFPwt. For cells infected with baculovirus containing MTH-mCherry, mCherry expression increased approximately linearly as MOI increased, for both PytR and A2 virus (Figure 4.3). In cells infected with wild type PytR virus, GFP expression continued to increase with MOI up to the maximum MOI of 20 (Figure 4.4). For A2 virus, GFP expression plateaued at approximately 400 RFU with infection from 5-20 MOI, indicating that even at low copy numbers (20-80 copy numbers, with 4xtRNA per virus), this tRNA has maximized its ability to incorporate Bock at TAG codons in mammalian cells.

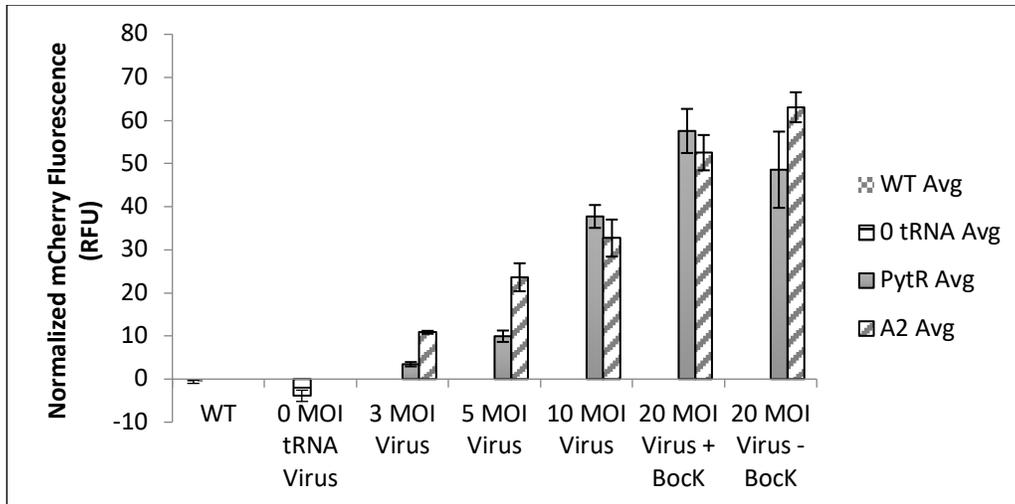
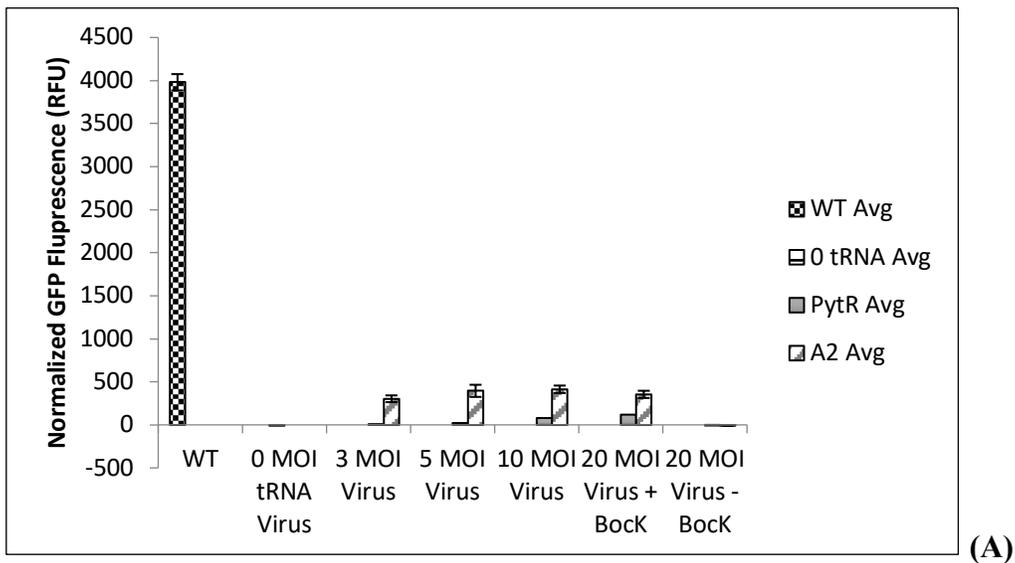


Figure 4.3. mCherry fluorescence of cells infected with baculovirus. HEK-293T cells were infected with 100 MOI of GFP-containing virus (either wild type GFP or GFP39tag) and varying MOIs of viruses containing 4 copies of a tRNA cassette (wild type PytR or A2) and mCherry expressed from the zinc metallothionine promoter. After 48 hours, cells were imaged and lysed, and mCherry and GFP fluorescence was taken.



(A)

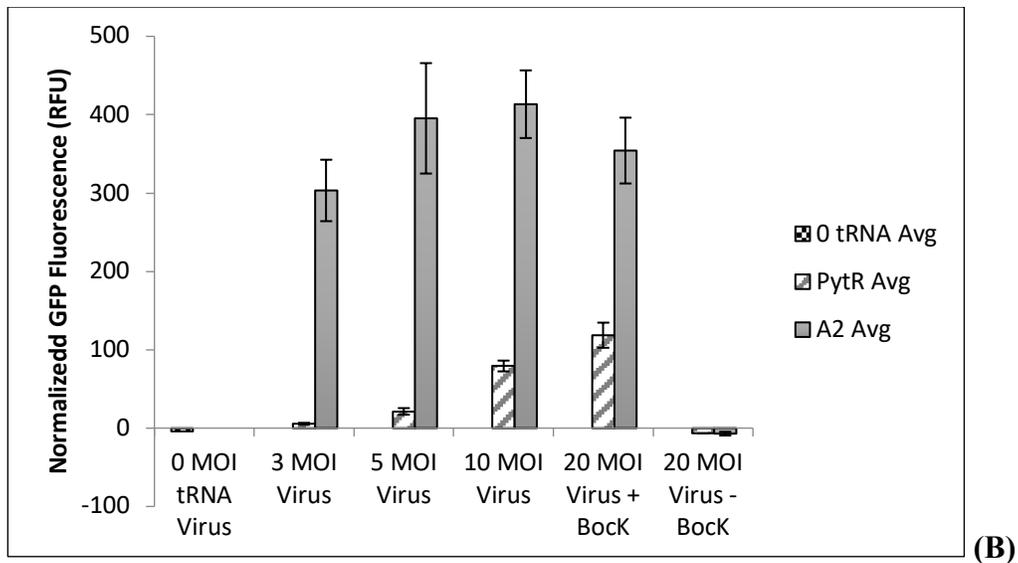


Figure 4.4. GFP fluorescence of cells infected with baculovirus. HEK-293T cells were infected with 100 MOI of GFP-containing virus (either wild type GFP or GFP39tag) and varying MOIs of viruses containing 4 copies of a tRNA cassette (wild type PytR or A2) and mCherry expressed from the zinc metallothionine promoter. After 48 hours, cells were imaged and lysed, and mCherry and GFP fluorescence was taken. (A) GFP fluorescence of all treatments. (B) GFP fluorescence of cells that received GFP39tag virus and tRNA viruses.

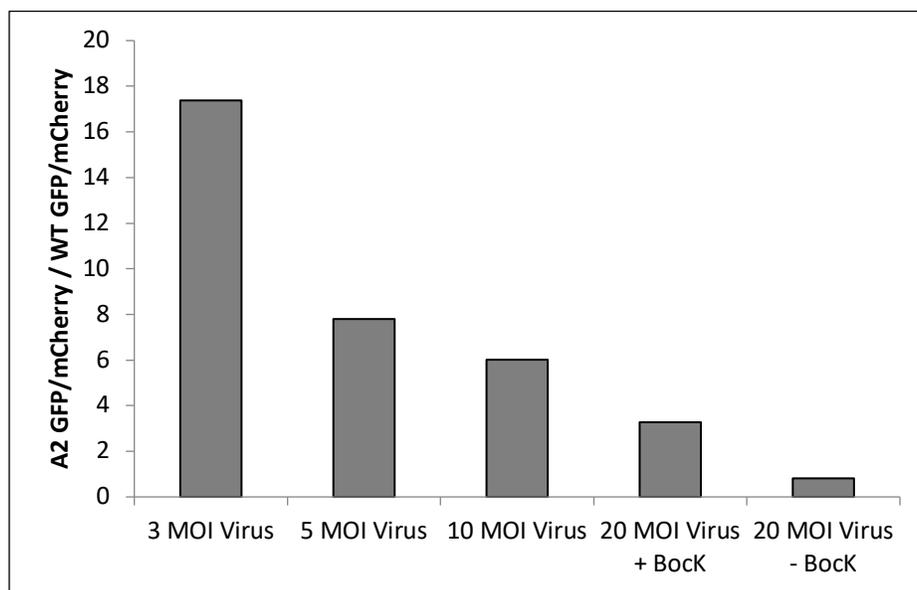


Figure 4.5. Comparison of GFP/mCherry ratios of A2 and wild type PytR. For each tRNA at each MOI, ratios were taken of average GFP values and average mCherry values. The ratio of these ratios (A2/wild type) at each MOI is provided here.

Comparing GFP/mCherry values for each virus at each MOI demonstrates the superiority of the A2 tRNA. The ratio between A2's and the wild type PytR's GFP/mCherry ratios is highest at MOI of 3 and decreases as MOI increases (Figure 4.5), which indicates that as copy number decreases, the heightened efficiency of this evolved tRNA becomes more important. At MOI of 3, the ratio between A2's and the wild type PytR's GFP/mCherry ratios is 17.4, which shows an improvement far beyond the 3-fold improvement previously suspected for A2 when the comparisons were made by delivering the tRNAs via transient transfection.

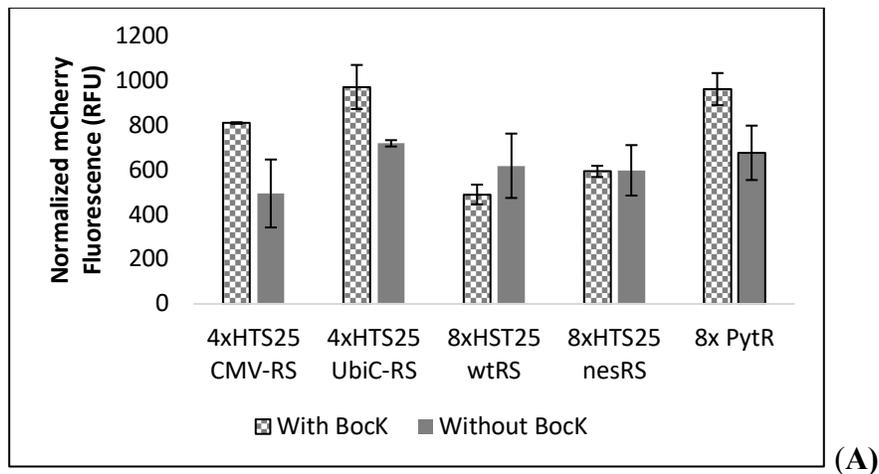
4.2.2 Designing an optimized PytR baculovirus vector

While infection via baculovirus allows for exquisite copy number control and homogenous DNA delivery, most applications for stop codon suppression in mammalian cells make use of transient transfection, as this method requires less time and labor than does preparing baculovirus vectors. Therefore, I set out to design a baculovirus vector that makes use of our highest-performing PytR, HTS25, that was developed using the VADER selection scheme. Each of these viral vectors contains an mCherry-TAG-GFP construct, which allows each vector to act as its own internal control for fluorescence expression.

In the pAcBac3 vector optimized for TAG suppression and O-methyltyrosine (OMeY) incorporation, putting the synthetase under the control of the weaker promoter UbiC (as opposed to the strong promoter CMV) predictably decreased synthetase expression, but substantially improved the expression of eGFPtag39.²¹ To determine which promoter would enable the expression of more GFP in this vector, I compared GFP/mCherry ratios in cells transfected with either pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25 or pAcBac-UbiC-MbPylRS-CAG-mCherry-TAG-GFP-

4xHTS25 (Figure 4.6). These vectors only contained one tRNA cassette due to cloning site restrictions after cloning in the different promoters. The GFP/mCherry ratio for the vector with the CMV promoter was 2.3x higher than that of the vector with the UbiC promoter, demonstrating that unlike with the tyrosyl vector, having the synthetase under the control of the stronger promoter in the pyrrolysyl vector improves stop codon suppression.

I cloned in an extra copy of the tRNA cassette into the vector with the CMV promoter, and made a similar construct using the wild type PytR, so that each vector had eight copies of its respective tRNA. The GFP/mCherry ratio for the 8xPytR vector was slightly lower than that for the 4xHTS25 vector, which demonstrates that even by transient transfection, which delivers high copy numbers of the DNA to cells, that cells needed twice as much of the wild type PytR to make as much GFP as the cells that received HTS25 did. The 8xHST25 vector performed 1.4x better than did the 8xPytR. It is likely that the differences between these vectors will be more extreme when they are packaged into baculoviruses and infected at low MOI.



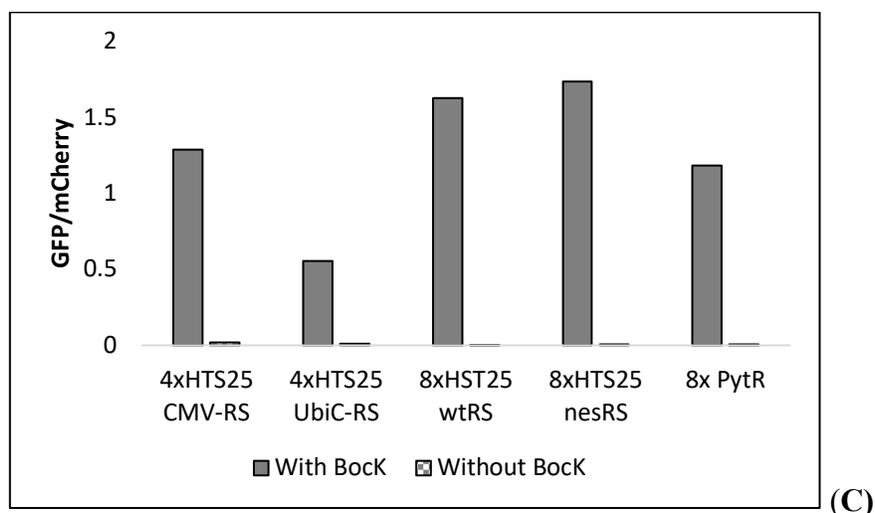
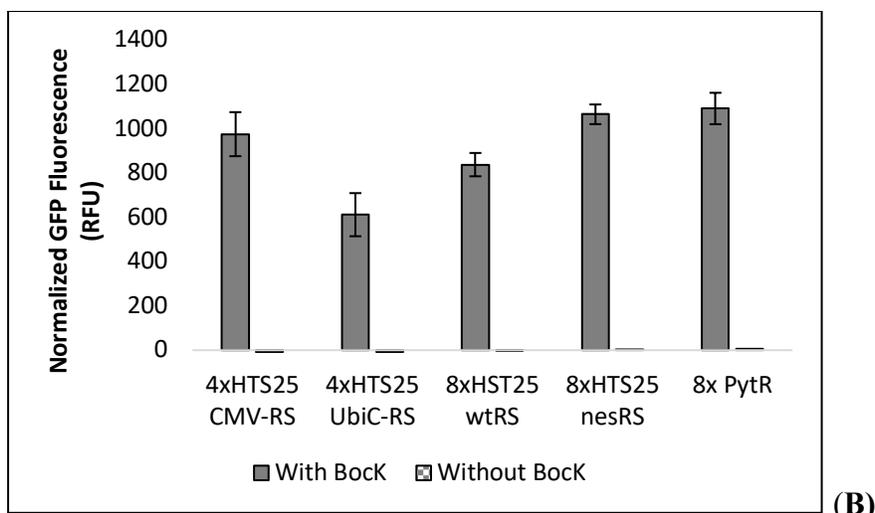


Figure 4.6. Comparison of GFP/mCherry ratios of HTS25 and wild type PytR. HEK-293T cells were transfected with pAcBac vectors containing the mCherry-TAG-GFP cassette and the MbPylRS under control of either the CMV or the UbiC promoter. These vectors contained 4 or 8 copies of the HTS25 tRNA, or 8 copies of the wild type PytR. One of the 8xHTS25 vectors also contained an MbPylRS with a nuclear export sequence (nes) on its N-terminus. (A) mCherry fluorescence of lysate of infected cells. (B) GFP fluorescence of infected cell lysate. (C) Comparison of mCherry/GFP ratios between cells infected with each vector.

While the differences between the wild type PytR and the HTS25 variant are evident even by transfection, the nuclear export sequence on the N-terminus of the MbPylRS did not make a substantial improvement to Bock incorporation when these vectors were delivered by transient transfection. Many small differences between

performance of these synthetases may have masked because of the high copy numbers delivered to each cell in transfection experiments, so it is possible that the NES may offer a substantial improvement to BocK incorporation once baculovirus is produced using these vectors and used to infect mammalian cells.

4.2.3 Incorporating other optimized synthetases

There have been a variety of pyrrolysyl synthetase variants that have been optimized to charge lysine variants of various sizes and characteristics for different uses. I made pAcBac vectors with some of these MbPylRS variants to demonstrate the versatility of these vectors for incorporating such diverse UAAs.

The NBK RS developed by the Schultz lab can be used to incorporate photocaged lysine variants. I replaced the wild type MbPylRS with the NBK-RS to produce pAcBac-CMV-NBK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 and compared the incorporation of the photocaged lysine PCK (Figure 4.7) between cells transfected with this plasmid and its wild type MbPylRS counterpart. The GFP/mCherry ratios were 4.4x higher with the NBK-RS vector than with the wild type RS, and there was negligible GFP expression from either vector in the absence of PCK (Figures 4.8, 4.9). This NBK-RS pAcBac vector with the evolved tRNA cassettes is therefore a highly useful tool for incorporating photocaged lysine UAAs in mammalian cells.

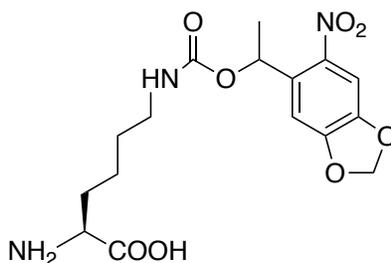


Figure 4.7. Structure of the photocaged lysine PCK.

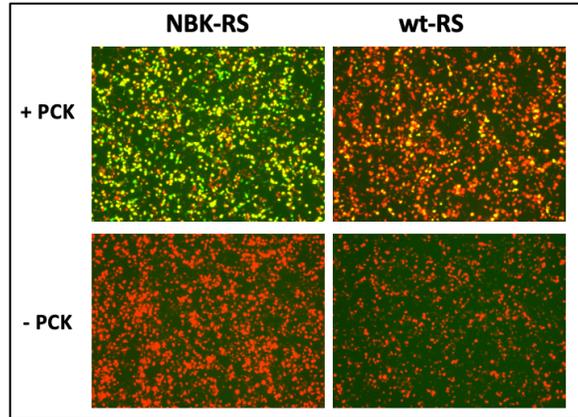
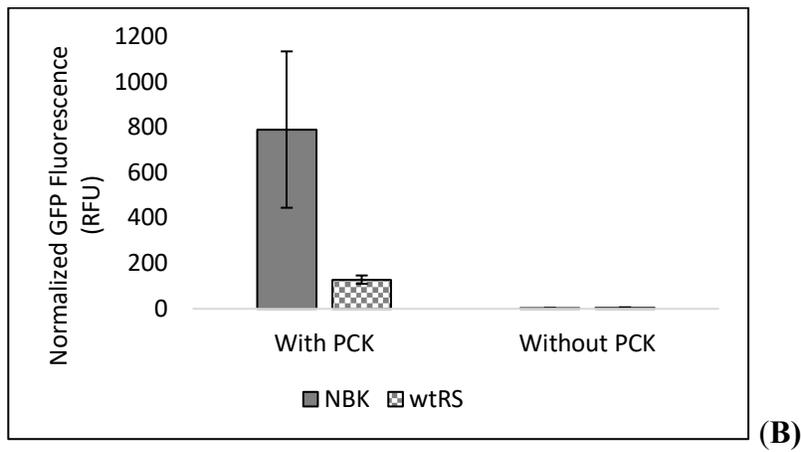
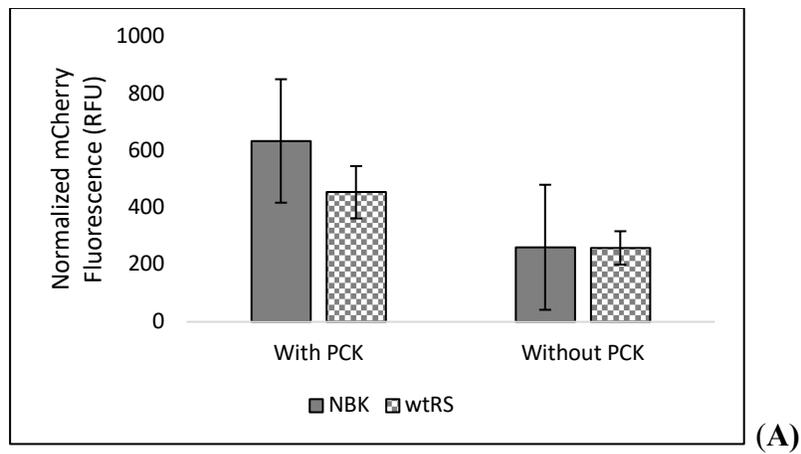


Figure 4.8. Comparison of NBK-MbPylRS and wild type MbPylRS to charge PCK. HEK-293T were transfected with pAcBac-CMV-NBK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 or pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25. Cells were imaged after 72 hours.



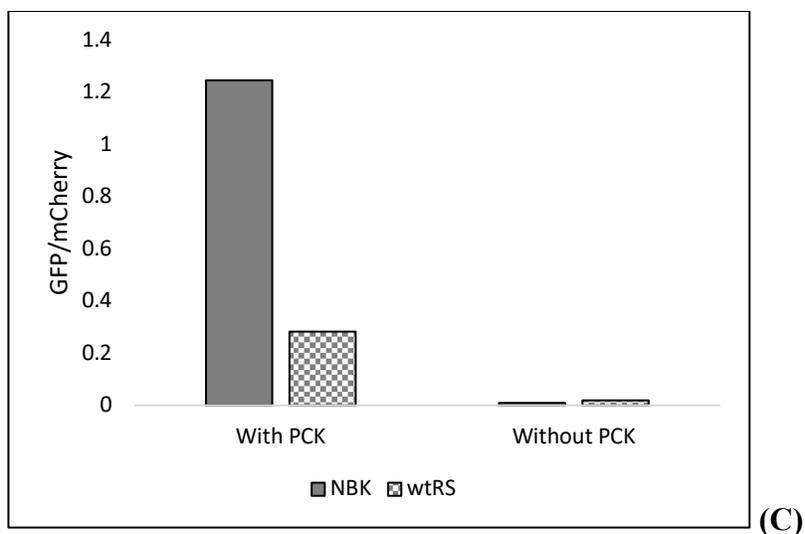


Figure 4.9. Comparison of GFP/mCherry ratios for NBK-RS and the wild type MbPylRS. HEK-293T cells were transfected with pAcBac vectors containing the mCherry-TAG-GFP cassette and either the wild type MbPylRS or the NBK-RS. Cells were imaged and lysed after 72 hours, then GFP and mCherry fluorescence were taken. (A) mCherry fluorescence of lysate of infected cells. (B) GFP fluorescence of infected cell lysate. (C) Comparison of mCherry/GFP ratios between cells infected with either the wild type synthetase or the NBK-RS.

Diazirine lysine (DiazK; Figure 4.10) is another photocaged amino acid for which an MbPylRS has been evolved. Unlike its NBK counterpart, pAcBac-CMV-DiazK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 did not result in higher GFP expression than the vector with the wild type MbPylRS (Figure 4.11). Rather, GFP/mCherry ratio for the wild type synthetase was 7.6x higher than that for the DiazK synthetase (Figure 4.12). It is possible that in the context of this vector, the DiazK synthetase was not expressed optimally for DiazK incorporation, and that using a different promoter for this gene would improve synthetase function.

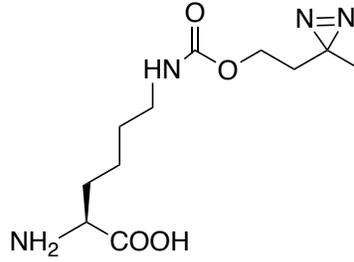


Figure 4.10. Structure of DiazK.

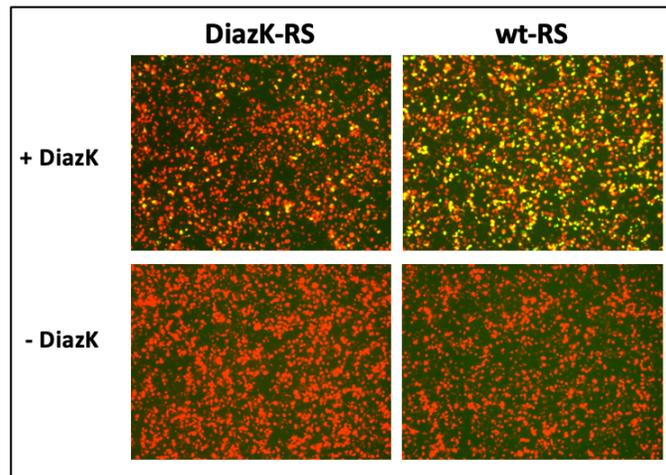
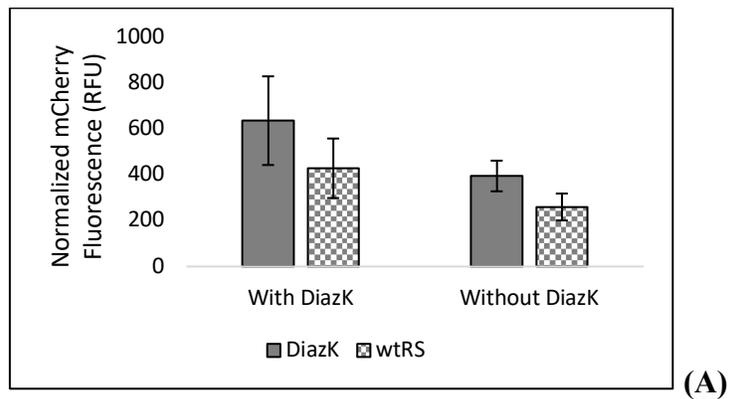


Figure 4.11. Comparison of DiazK-MbPylRS and wild type MbPylRS to charge DiazK. HEK-293T were transfected with pAcBac-CMV-DiazK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 or pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25. Cells were imaged after 72 hours.



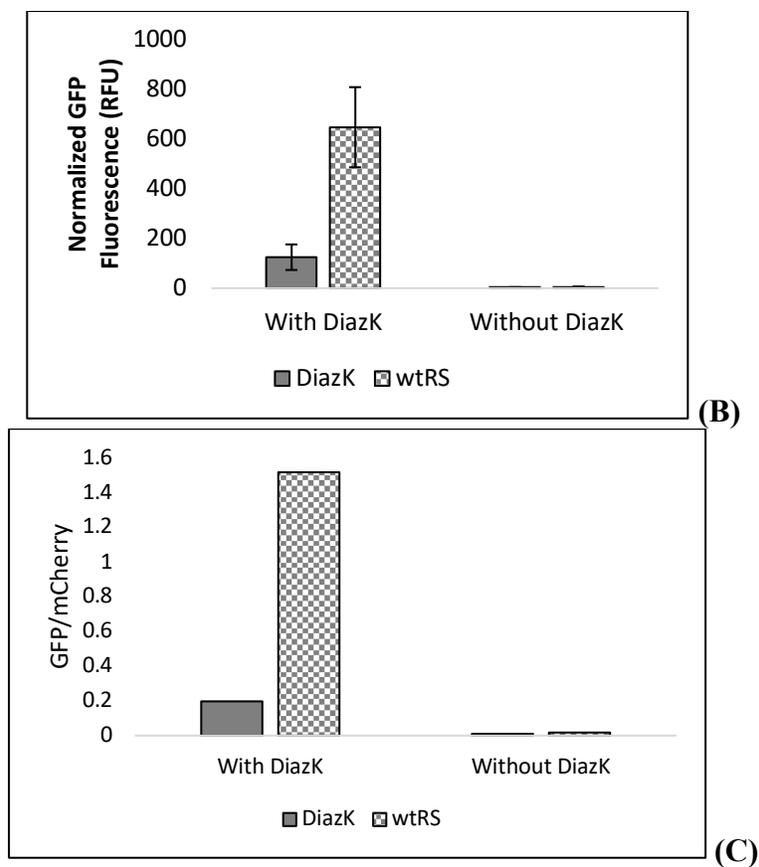


Figure 4.12. Comparison of GFP/mCherry ratios for DiazK-RS and the wild type MbPylRS. HEK-293T cells were transfected with pAcBac vectors containing the mCherry-TAG-GFP cassette and either the wild type MbPylRS or the DiazK-RS. Cells were imaged and lysed after 72 hours, then GFP and mCherry fluorescence were taken. (A) mCherry fluorescence of lysate of infected cells. (B) GFP fluorescence of infected cell lysate. (C) Comparison of mCherry/GFP ratios between cells infected with either the wild type synthetase or the DiazK-RS.

N^ϵ -acetyl-L-lysine (AcK) is a UAA whose incorporation in mammalian cells at the point of translation mimics the posttranslational acetylation of lysine (Figure 4.13). There is an MbPylRS variant selected for incorporation of AcK. In this backbone, however, it did not incorporate AcK at a level that improved GFP expression more than did the wild type synthetase (Figure 4.14). Both synthetases incorporated GFP at a negligible level (Figure 4.15). The context of this vector may also require further optimization for use with this

synthetase, such as using a different promoter for the synthetase as the tyrosyl RS requires in pAcBac3.²¹

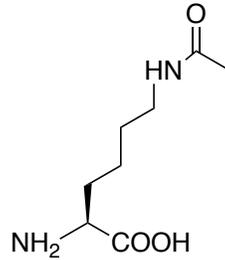


Figure 4.13. Structure of AcK.

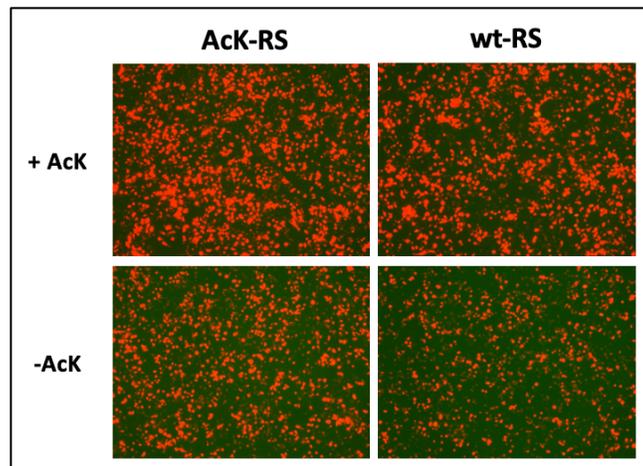
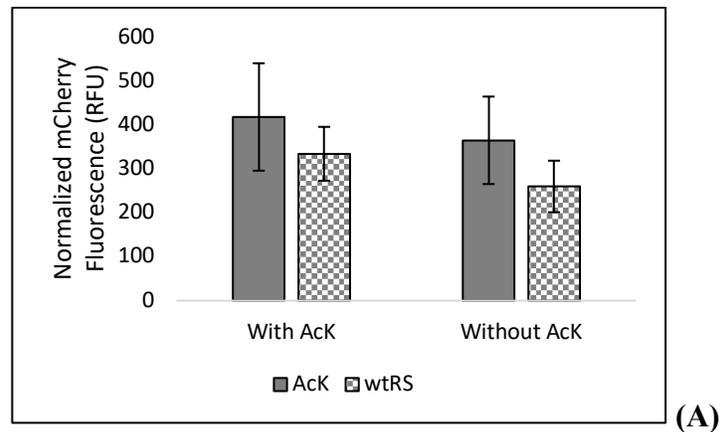


Figure 4.14. Comparison of AcK-MbPylRS and wild type MbPylRS to charge AcK. HEK-293T were transfected with pAcBac-CMV-AcK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 or pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25. Cells were imaged after 72 hours.



(A)

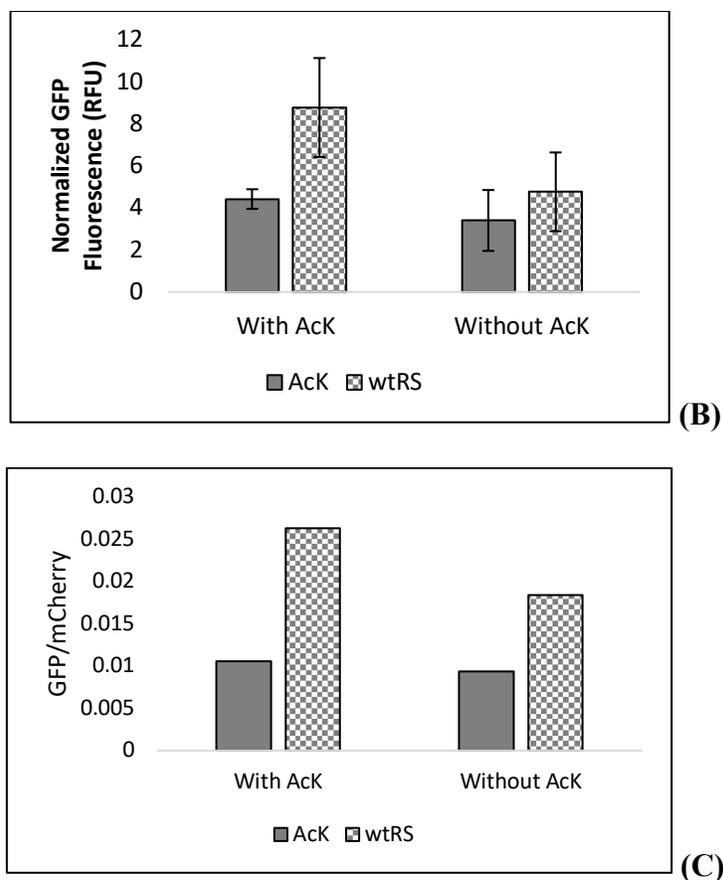


Figure 4.15. Comparison of GFP/mCherry ratios for AcK-RS and the wild type MbPylRS. HEK-293T cells were transfected with pAcBac vectors containing the mCherry-TAG-GFP cassette and either the wild type MbPylRS or the AcK-RS. Cells were imaged and lysed after 72 hours, then GFP and mCherry fluorescence were taken. (A) mCherry fluorescence of lysate of infected cells. (B) GFP fluorescence of infected cell lysate. (C) Comparison of mCherry/GFP ratios between cells infected with either the wild type synthetase or the AcK-RS.

There is a MbPylRS variant (the biotin-lysine variant, or Biotin-K) that is a polyspecific synthetase capable of charging a number of diverse lysine variants, including those too large for most synthetases. I compared the performance of this synthetase with that of the wild type MbPylRS at incorporating azidolysine (AzK; Figure 2.14) and cyclo-octyne lysine (SCOK; Figure 4.16). The wild type synthetase outperformed the Biotin-K synthetase at incorporating AzK into the mCherry-TAG-GFP construct (Figure 4.17), with a GFP/mCherry ratio 1.4x that of the Biotin-K's GFP/mCherry ratio. With SCOK,

however, the Biotin-K synthetase had a GFP/mCherry ratio an impressive 53x that of the wild type synthetase's GFP/mCherry ratio (Figure 4.18). The wild type synthetase did not charge SCOK well, as evidenced by the little GFP fluorescence in cells transfected with the wild type synthetase pAcBac construct and provided with SCOK. This wild type synthetase does not have an active site large enough to accommodate large amino acids such as SCOK, but the Biotin-K synthetase's evolved active site is. The Biotin-K synthetase did charge AzK, and did produce an appreciable amount of GFP, but its activity was not as high as the wild type synthetase's. This variation in performance is to be expected in a polyspecific synthetase, especially one with a substrate range as large as that of the Biotin-K synthetase, as it is not possible for a synthetase to optimally charge such different UAAs with equal efficiency.

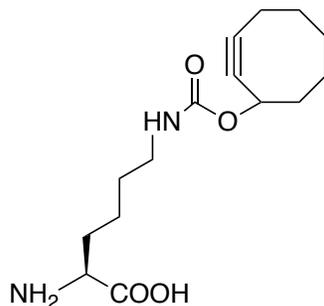


Figure 4.16. Structure of cyclo-octyne lysine (SCOK).

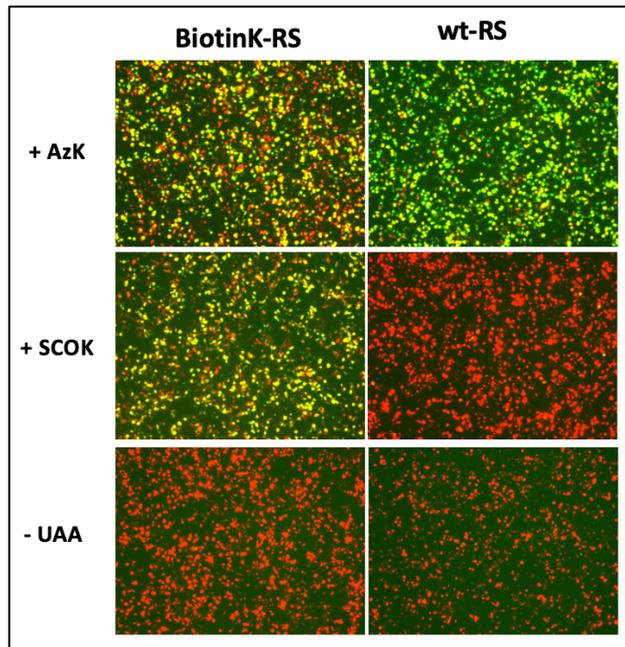
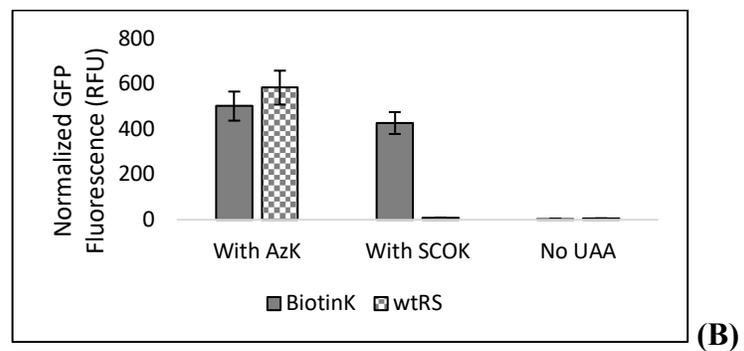
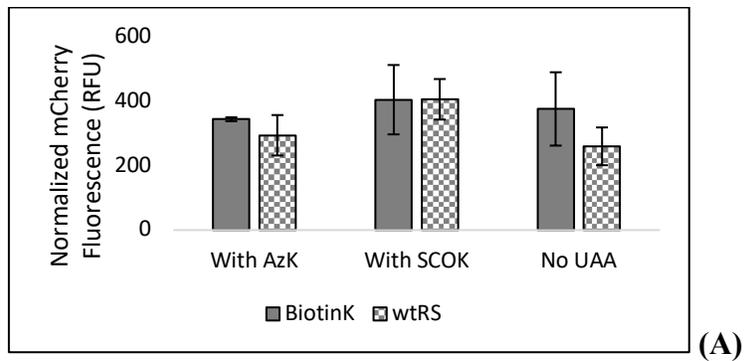


Figure 4.17. Comparison of BiotinK-MbPylRS and wild type MbPylRS to charge AzK and SCOK. HEK-293T were transfected with pAcBac-CMV-BiotinK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 or pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25. Cells were imaged after 72 hours.



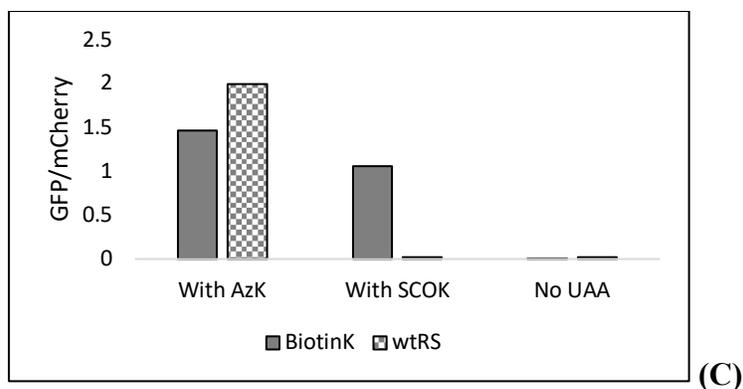


Figure 4.18. Comparison of GFP/mCherry ratios for AcK-RS and the wild type MbPylRS. HEK-293T cells were transfected with pAcBac vectors containing the mCherry-TAG-GFP cassette and either the wild type MbPylRS or the AcK-RS. Cells were imaged and lysed after 72 hours, then GFP and mCherry fluorescence were taken. (A) mCherry fluorescence of lysate of infected cells. (B) GFP fluorescence of infected cell lysate. (C) Comparison of mCherry/GFP ratios between cells infected with either the wild type synthetase or the BiotinK-RS, for each UAA.

4.3 Conclusion

These experiments demonstrate the power of baculovirus vectors for incorporating UAAs into proteins in mammalian cells. These vectors, when made into P1 baculovirus, can deliver aaRS/tRNA pairs to mammalian cells with exquisite control over copy number. Transduction by virus allows the homogeneous distribution of DNA across the population of cells, which is useful in *in vivo* experiments, as well as delivery of low copy numbers of virus to each cell, which here allows the true elucidation of the improved efficiency of an evolved tRNA. Even though the superiority of this evolved tRNA is most evident in infection experiments, these pAcBac vectors are also useful when they are delivered via transient transfection. The NES appended to the N-terminus of the synthetase unfortunately does not boost stop codon suppression by a measurable amount with transfection, when many small differences in efficiencies between constructs are masked. HTS25's superiority over the wild type PytR in suppressing TAG codons is evident even in these transfection experiments, as eight copies of the wild type PytR are required for the same amount of stop

codon suppression as provided by four copies of HTS25. These vectors can be used for a variety of experiments, as the synthetase in the vector can be easily swapped for one specifically suited to the experiment. Overall, these pAcBac vectors present a useful tool for the chemical biology community to use in experiments involving lysine variants and TAG suppression in mammalian cells.

4.4 Acknowledgments

Dr. Yunan Zheng gave me access to her plasmid stocks and trained me to prepare recombinant baculovirus. Dr. Sarah Erickson gave me access to her plasmid stocks and stores of UAAs. Dr. Rengeng Gu gave me access to the stock of plasmid containing the DiazK-RS.

4.5 Materials and Methods

4.5.1. Strains and cell lines

Plasmid propagations were performed as described in Chapter 2 of this thesis. Sf9 cells and HEK-293T cells were maintained as described in Chapter 2.

4.5.2. Plasmids

All plasmid maps and sequences can be found in the Appendix.

The plasmids pAcBac-CMV-MbPylRS-CMV-GFPwt and pAcBac-CMV-MbPylRS-CMV-GFP39tag were propagated from Dr. Rachel Kelemen's stocks.

To construct pAcBac-4xPytRwt-MTH-mCherry and pAcBac-4xA2-MTH-mCherry, pAcBac-4xPytRwt and pAcBac-4xA2 (both from Dr. Kelemen's stocks) were cut with SbfI and NotI. An MTH-mCherry g-block was amplified with the primers MTH-mCherry-SbfI-F and MTH-mCherry-NotI-R, and this cassette was cloned into the pAcBac plasmids.

The mCherry-TAG-GFP construct consists of a wild type mCherry protein, followed by a flexible linker containing a stop codon (Gly-Gly-Ser-TAG-Gly-Gly-Ser), and then a wild type EGFP protein. To produce this construct, the N-terminus of the flexible linker was added to wild type EGFP using the primer NheI-linker-EGFP-F, with the reverse terminal primer GFP-EcoRI-6xHis-R. This fragment was cloned into pAcBac1-GFPwt using the EcoRI and NheI sites to produce the intermediate plasmid pAcBac1-GFPwt+linker. The mCherry component of the construct was amplified with the primers NheI-Kozak-mCherry-F and mCherry-linker-AvrII-R, which added a Kozak sequence to the 5' end of the DNA fragment and the other half of the linker to the C-terminus of mCherry. The intermediate plasmid pAcBac1-GFPwt+linker and this second DNA fragment were digested with NheI and AvrII, and the mCherry half of the construct was cloned into these sites, destroying the NheI and AvrII cut sites in the linker and producing the plasmid pAcBac-mCherry-TAG-GFP. The mCherry-TAG-GFP construct can be cut, intact, from the plasmid using the NheI and EcoRI cut sites for ease of future clonings.

To produce pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25 and pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xPytRwt, a number of intermediate plasmids were generated before arriving at the final constructs. I digested pIDT-4xHTS25 and pIDT-4xPytRwt with AvrII and NheI. I digested pAcBac-CMV-MbPylRS-WPRE-CAG-GFPtag39AvrII~~del~~-WPRE with AvrII and cloned the 4xHTS25 and 4xPytRwt cassettes in using those cut sites to generate the intermediate plasmids pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-4xHTS25 and pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-4xPytRwt. To add a second tRNA cassette to bring the total tRNA copy number in each plasmid to 8, I digested the intermediate plasmids with SpeI, which is

compatible with both AvrII and NheI, allowing me to produce the plasmids pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-8xHTS25 and pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-8xPytRwt.

It was necessary to replace the GFPtag construct with the mCherry-TAG-GFP construct so that there would be an internal fluorescent control in all experiments using these plasmids. SfiI sites were added to either end of the mCherry-TAG-GFP-WPRE construct using the primers Gibson-mCherry-F and Gibson-WPRE-R. The intermediate plasmids pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-4xHTS25, pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-4xPytRwt, pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-8xHTS25, and pAcBac-CMV-MbPylRS-CAG-GFPtagAvrII~~del~~-8xPytRwt were all digested with SfiI to remove the GFPtag construct, and the mCherry-TAG-GFP-WPRE construct was added to each of these plasmids using Gibson assembly to produce pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25, pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xPytR, pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xHTS25, and pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xPytR.

pAcBac-UbiC-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25 were produced by digesting pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25 with SpeI and SacI, and using Gibson assembly with a UbiC fragment amplified with the primers Gibson-UbiC-F and Gibson-UbiC-R.

The nuclear export sequence was added to the N-terminus of the MbPylRS to produce NES-MbPylRS using the primers NheI-Lemke-NES-MbPylRS-F and MbPylRS-WPRE-XhoI-R. The point mutations to produce the Biotin-K MbPylRS were introduced

using the internal primers L274A-F, L274A-mut-R, V370R-F, V730R-mut-R, C313V-M315Y-F, and C313V-M315Y-mut-R; and the terminal primers MbPylRS-Nterm-F and MbPylRS-Cterm-R.

4.5.3 Unnatural amino acids

PCK was synthesized as previously described.³³ DiazK was purchased from SiChem (catalog number SC-8034). AcK was purchased from Thermo Fisher Scientific (catalog number AAJ613906). AzK was purchased from Iris Biotech (catalog number HAA1625). SCOK was purchased from SiChem (catalog number SC-8000).

4.5.4 Producing recombinant baculovirus

Recombinant baculoviruses were produced as described in detail in Chapter 2 of this thesis.

4.5.5 Comparing A2 PytR variant to wild type PytR

HEK-293T cells were infected with either pAcBac-CMV-MbPylRS-CMV-GFPwt or pAcBac-CMV-MbPylRS-CMV-GFPtag39 P1 virus at 100 MOI. Cells that received GFPtag39 virus received either pAcBac-4xPytRwt-MTH-mCherry or pAcBac-4xA2-MTH-mCherry P1 virus at the following MOIs: 0, 3, 5, 10, or 20. BocK was also added to the cell media of wells that received GFPtag39 and tRNA virus. No BocK controls wells received GFPtag39 virus and either tRNA virus at MOI 20.

4.5.6 Designing an optimized PytR baculovirus vector

To determine whether the CMV or UbiC promoter for the MbPylRS is optimum for enhancing the incorporation of lysine variants at TAG codons, pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25 was compared with pAcBac-UbiC-

MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25 via transient transfection, with and without the addition of BocK to the cell media.

To demonstrate the improvement of the HTS25 PytR over the wild type PytR, I compared pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xPytR to pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 via transient transfection, with and without the addition of BocK to the cell media. I also compared pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 to pAcBac-CMV-NES-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 via transient transfection to determine whether the Lemke nuclear export sequence improved the incorporation of BocK and therefore GFP expression.

4.5.7 Incorporating other optimized synthetases

I compared pAcBac-CMV-NBK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 to the same vector with the wild type synthetase via transient transfection with a photocaged lysine (PCK) as the substrate. I also tested pAcBac-CMV-DiazK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 with DiazK, pAcBac-CMV-AcK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 with AcK, and pAcBac-CMV-BiotinK-MbPylRS-CAG-mCherry-TAG-GFP-8xHST25 with azidolysine (AzK) and cyclo-octyne lysine (SCOK).

4.6 References

- (1) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79* (1), 413.
- (2) Cohen, R. N.; van der Aa, M. A. E. M.; Macaraeg, N.; Lee, A. P.; Szoka, F. C. *J. Control. Release* **2009**, *135* (2), 166.
- (3) Carpentier, E.; Paris, S.; Kamen, A. A.; Durocher, Y. *J. Biotechnol.* **2007**, *128* (2), 268.
- (4) Glover, D. J.; Leyton, D. L.; Moseley, G. W.; Jans, D. A. *J. Gene Med.* **2010**, *12* (1), 77.
- (5) Atchison, R. W.; Casto, B. C.; Hammon, W. M. D. *Science (80-.)*. **1965**, *149* (3685), 754.
- (6) Samulski, R. J.; Muzyczka, N. *Annu. Rev. Virol.* **2014**, *1* (1), 427.
- (7) Wang, D.; Tai, P. W. L.; Gao, G. *Nature Reviews Drug Discovery*. Nature Publishing Group May 1, 2019, pp 358–378.
- (8) Srivastava, A.; Lusby, E. W.; Berns, K. I. *J. Virol.* **1983**, *45* (2).
- (9) Nakai, H.; Storm, T. A.; Kay, M. A. *Nat. Biotechnol.* **2000**, *18* (5), 527.
- (10) Rendahl, K. G.; Leff, S. E.; Otten, G. R.; Spratt, S. K.; Bohl, D.; Van Roey, M.; Donahue, B. A.; Cohen, L. K.; Mandel, R. J.; Danos, O.; Snyder, R. O. *Nat. Biotechnol.* **1998**, *16* (8), 757.
- (11) Hauswirth, W. W.; Berns, K. I. *Virology* **1977**, *78* (2), 488.
- (12) Im, D. S.; Muzyczka, N. *J. Virol.* **1992**, *66* (2).
- (13) Xie, J.; Mao, Q.; Tai, P. W. L.; He, R.; Ai, J.; Su, Q.; Zhu, Y.; Ma, H.; Li, J.; Gong, S.; Wang, D.; Gao, Z.; Li, M.; Zhong, L.; Zhou, H.; Gao, G. *Mol. Ther.* **2017**, *25* (6), 1363.
- (14) Tran, N. T.; Heiner, C.; Weber, K.; Weiland, M.; Wilmot, D.; Xie, J.; Wang, D.; Brown, A.; Manokaran, S.; Su, Q.; Zapp, M. L.; Gao, G.; Tai, P. W. L. *Mol. Ther. - Methods Clin. Dev.* **2020**, *18*, 639.
- (15) Naldini, L.; Blömer, U.; Gallay, P.; Ory, D.; Mulligan, R.; Gage, F. H.; Verma, I. M.; Trono, D. *Science (80-.)*. **1996**, *272* (5259), 263.
- (16) Dull, T.; Zufferey, R.; Kelly, M.; Mandel, R. J.; Nguyen, M.; Trono, D.; Naldini, L. *J. Virol.* **1998**, *72* (11), 8463.
- (17) Shen, B.; Xiang, Z.; Miller, B.; Louie, G.; Wang, W.; Noel, J. P.; Gage, F. H.; Wang, L. *Stem Cells* **2011**, *29* (8), 1231.
- (18) Sakin, V.; Hanne, J.; Dunder, J.; Anders-Össwein, M.; Laketa, V.; Nikić, I.; Kräusslich, H. G.; Lemke, E. A.; Müller, B. *Cell Chem. Biol.* **2017**, *24* (5), 635.
- (19) Zheng, Y.; Yu, F.; Wu, Y.; Si, L.; Xu, H.; Zhang, C.; Xia, Q.; Xiao, S.; Wang, Q.; He, Q.; Chen, P.; Wang, J.; Taira, K.; Zhang, L.; Zhou, D. *Nucleic Acids Res.* **2015**, *43* (11), 73.
- (20) Chen, J.; Powell, D.; Hu, W.-S. *J. Virol.* **2006**, *80* (19), 9651.
- (21) Zheng, Y.; Lewis, T. L.; Igo, P.; Polleux, F.; Chatterjee, A. *ACS Synth. Biol.* **2017**, *6* (1), 13.
- (22) Barsoum, J.; Brown, R.; McKee, M.; Boyce, F. M. *Hum. Gene Ther.* **1997**, *8* (17), 2011.
- (23) Kaikkonen, M. U.; Rätty, J. K.; Airene, K. J.; Wirth, T.; Heikura, T.; Ylä-Herttua, S. *Gene Ther.* **2006**, *13* (4), 304.
- (24) Chatterjee, A.; Xiao, H.; Bollong, M.; Ai, H. W.; Schultz, P. G. *Proc. Natl. Acad.*

- Sci. U. S. A.* **2013**, *110* (29), 11803.
- (25) Kost, T. A.; Condreay, J. P. *Trends in Biotechnology*. Elsevier Current Trends April 1, 2002, pp 173–180.
- (26) Ayres, M. D.; Howard, S. C.; Lopez-Ferber, M.; Lopez-Ferber, M.; Possee, R. D. *Virology* **1994**, *202* (2), 586.
- (27) Cheshenko, N.; Krougliak, N.; Eisensmith, R. C.; Krougliak, V. A. *Gene Ther.* **2001**, *8* (11), 846.
- (28) Jarvis, D. L. In *Methods in Enzymology*; Academic Press Inc., 2009; Vol. 463, pp 191–222.
- (29) Kost, T. A.; Condreay, J. P.; Jarvis, D. L. *Nature Biotechnology*. Nature Publishing Group May 5, 2005, pp 567–575.
- (30) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chemical Science*. Royal Society of Chemistry January 1, 2015, pp 50–69.
- (31) Italia, J. S.; Addy, P. S.; Wrobel, C. J. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A. *Nat. Chem. Biol.* **2017**, *13* (4), 446.
- (32) Kelemen, Rachel E.; Jewel, Delilah; Pasha, Muhammad; Cao, Xiaofu; Chatterjee, A. *Submiss.*
- (33) Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2010**, *132* (12), 4086.

Chapter 5

A virus-based two-hybrid selection scheme to probe the mammalian interactome

5.1 Introduction

The prior chapters of this thesis focused on enhancing selection schemes for the directed evolution of tRNAs and optimizing viral vectors for genetic code expansion in mammalian cells. This chapter deals with using AAV-based selection schemes to develop a high-throughput two-hybrid assay for detecting binding partners in the mammalian proteome.

5.1.1 *Two-hybrid selection*

Many cellular processes are controlled by protein-protein interactions (PPIs). Typical cellular functions, such as growth and differentiation, or signal transduction, rely on PPIs.^{1,2} Parasites and pathogens also manipulate PPIs to gain access to cells and replicate.³ Understanding the interactome in mammalian cells can help researchers to better understand our physiology and how to treat diseases when the underlying causes involve interactions between proteins.

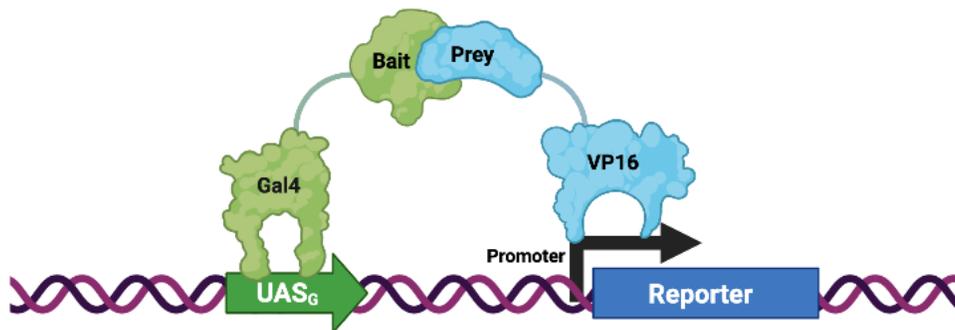


Figure 5.1. Overview of a two-hybrid assay. In a two-hybrid system, a DNA binding domain (typically Gal4) fused to a bait protein binds to the UAS_G region of the GAL1 promoter. The bait protein associates with the prey protein, which is fused to the VP16 transcription activator. VP16 binds to the promoter of the reporter gene, which is then transcribed. (Image designed using BioRender.)

In a two-hybrid experiment, cells are transformed with plasmids containing a DNA binding domain fused to a bait protein of interest, a transcription activator fused to a prey protein (or one member of a library of proteins), and a reporter gene that cannot be

expressed unless the DNA binding domain binds to an area upstream of the gene *and* the transcription activator is recruited. The DNA binding domain will bind the region upstream of the reporter gene, but the transcription activator will not be recruited unless the bait and prey proteins physically associate. Transcription of the reporter gene is thus tied to the interaction of the bait and prey proteins (Figure 5.1). Typically, the DNA binding domain used is the Gal4 DNA binding domain (the N-terminus of the GAL4 protein, which binds to the UAS_G in the GAL1 promoter).⁴ The *Herpesvirus* protein VP16 is a popular transcription activator.⁵

5.1.2 Yeast two-hybrid systems

The most commonly used two-hybrid system is the yeast two-hybrid system.^{6,7} Yeast cells exhibit a number of advantages over bacteria and mammalian cells for use in two-hybrid experiments,⁵ although two-hybrid protocols exist for each of those cell types.⁸⁻¹⁰ Transformation efficiency in yeast cells is good, and since yeast cells amplify and maintain plasmids, recovering plasmids from them after a screen or a selection is simple. Two-hybrid screens can be performed easily in yeast cells grown on X-gal with LacZ as the reporter gene, turning colonies blue if the bait and prey proteins associate. Selections are also straightforward using this system: growing an auxotrophic strain on media lacking a particular essential amino acid like histidine and using a gene such as *HIS3* as the reporter protein will only result in cell survival and colony growth if the bait and prey proteins interact. Since *Saccharomyces cerevisiae* is a eukaryote, mammalian bait and prey proteins are more likely to be translated and folded properly than if the selection process took place in bacteria. These features of the yeast two-hybrid system are attractive to researchers investigating the eukaryotic interactome.

While working with yeast is advantageous in many respects, yeast two-hybrid systems also present problems when working with the mammalian interactome. Even though yeast are eukaryotes, they are evolutionarily divergent from mammals, and many mammalian proteins will not fold correctly in yeast due to lack of appropriate chaperones. Additionally, certain PPIs may depend on post-translational modifications or other interactome-dependent changes that are not possible when the proteins are removed from their native context. Performing two-hybrid experiments directly in mammalian cells would resolve these challenges.

5.1.3. Mammalian two-hybrid systems

Mammalian two-hybrid systems exist that allow detection of protein binding partners in their native cellular contexts. In these systems, three plasmids are typically used to deliver the necessary genetic cargo to the cells for the assay: the Gal4-bait protein fusion, the VP16-prey protein fusion, and the reporter gene downstream of the UAS_G region of the GAL1 promoter.¹⁰ This system is limited by the efficiency of transient transfection, as not all cells will receive all three plasmids necessary to undergo the two-hybrid assay. One solution could be a two-plasmid system in which the reporter gene and the bait/Gal4 binding domain are on the same plasmid and the VP16/prey fusion are on a different plasmid. Unfortunately, it is not possible to control DNA delivery in transient transfection across a large population in a single dish, so this method renders impossible the delivery of library of prey proteins so that each cell only receives one library member. Without the ability to deliver a single prey library member per cell, it is not possible to scan the proteome for binding partners in a two-hybrid assay.

A newer method using microarrays addresses this problem of extremely low-throughput pairwise selections. Cell array protein-protein interaction assay (CAPPIA) makes use of robotically printed microarrays and an autofluorescent reporter to deliver DNA individually to cells deposited on a slide.¹¹ This system permits testing of a library of prey proteins using transient transfection, which would be impossible in a larger plate where the researcher cannot control the delivery of individual plasmids. While CAPPIA addresses this problem, its implementation is also limited. While the authors of this method claim that CAPPIA is high throughput, they only tested a library of 160 prey proteins. That library size may be an improvement upon what an individual researcher can test manually, but under 200 proteins does not approach the size and complexity of the mammalian proteome that a researcher may want to screen. To have a truly high throughput library, it will be necessary to test several orders of magnitude more proteins than the CAPPIA system has been used for thus far. Additionally, performing CAPPIA requires the use of a robot, which not every lab has the means to purchase. A mammalian two-hybrid system should be as high-throughput and accessible as its yeast counterparts.

5.1.4. Developing high throughput two-hybrid systems for mammalian cells

In order to develop a truly high throughput mammalian two-hybrid system, it will be necessary to deliver the library of prey proteins to the host cells in a manner that ensures that each host cell receives only one library member. The delivery technique that works best for mammalian cells that meets that criterion is viral infection. As previously discussed, recombinant viruses are easy to produce and titer, which allows researchers to exert exquisite control over the copy number of viruses delivered to their host cells. Doing

so allows a researcher to perform a selection or screen in a large plate of cells, much like a bacterial or yeast selection, without the need for expensive robotics.

Recombinant AAV possesses a number of features that make it amenable to being used for two-hybrid selections in mammalian cells. It is possible to make VP2/VP3-only or VP3-only AAV capsids, but VP1 is required to produce infective virus.^{12,13} It is also possible to knock out VP1 expression from the AAV2 genome and place transcription of VP1 under control of a separate promoter, which is a technique that has been used successfully to install UAAs into only VP1 and not VP2 or VP3 (unpublished data). Separating the genes for the proteins in this way can allow researchers to place VP1 transcription, and hence the production of infectious virus, under the control of a promoter in a two-hybrid system. VP1 and the packaging of infectious virus becomes the reporter for a prey library member binding to the bait protein. While AAV is a small virus with a limited cargo capacity,¹⁴ the large majority of the human proteome is less than 1300 amino acids in length,¹⁵ which is the upper limit for the size of a prey protein that can be fused to VP16 and have the gene for this fusion fit within the AAV capsid. Infectious virus will contain packaged DNA, which will include the gene for the prey protein whose association with the bait protein prompted transcription of VP1. The genetic cargo of these reporter viruses can be characterized to determine which library members are hits. rAAV can potentially be both an excellent delivery vector and a reporter for a high throughput mammalian two-hybrid system.

Screens such as CIPPIA and selections such as metabolics-based yeast two-hybrid systems are certainly useful for identifying protein binding partners, but this goal is not the only one that can be achieved with a selection system. Other selection schemes have

evolved promoters in bacteria and yeast to improve production of a protein of interest.^{16–19} If a consistent bait and prey protein pair can be maintained, the sequences of promoters can be randomized and evolved to maximize reporter gene transcription.

Here, I present a two-hybrid selection scheme for use in mammalian cells. I demonstrate that existing two-hybrid assays using secreted alkaline phosphatase (SEAP) can be modified to use a fluorescent reporter. I then demonstrate that it is possible to place VP1 under control of the two-hybrid system such that production of infectious virus becomes the reporter. This system will allow researchers to develop a high throughput selection using AAV as both a vector to deliver library members and a reporter from which the sequence of a hit can be retrieved. I also demonstrate that this AAV-based two-hybrid system produces different quantities of virus depending on the strength of the promoter upstream of VP1. This AAV two-hybrid system can also therefore be used in the future to evolve minimal promoters for general synthetic biology uses.

5.2 Results and Discussion

5.2.1 Producing GFP with the use of two-hybrid binding partners

To test the generalized applicability of Matchmaker Mammalian Assay (Takara) protocol, I determined whether I could use a two-hybrid assay to express a reporter protein of interest (in this case, SEAP and eGFP). The Gal4 binding domain binds to GAL1 binding sites upstream of DNA to be transcribed, and the herpesvirus protein VP16 AD interacts with host transcription factors to induce transcription. With the Matchmaker assay, a fusion of the DNA binding domain GAL4 and the mouse p53 protein can bind to the fusion of the herpesvirus protein VP16 AD and the SV40 large T-antigen through the association of p53 and the T-antigen. To provide a control, VP16 AD is fused to the polyoma virus coat

protein (CP), which does not bind to p53. In addition to this test, in which GAL4 and VP16 must associate in the cell, I also provided a GAL4-VP16 fusion protein to determine the maximum level at which eGFP can be expressed using this system (when these proteins are constantly bound).

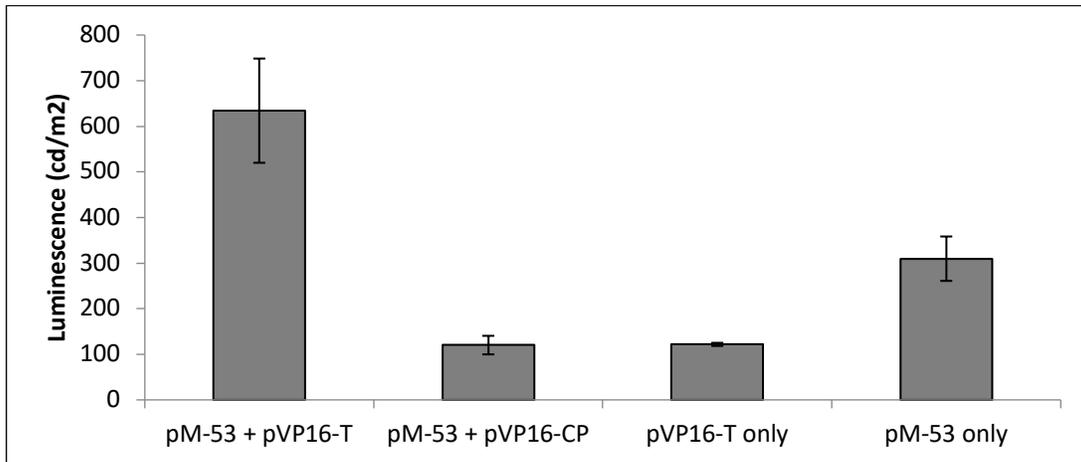


Figure 5.2. Luminescence of cells testing the two-hybrid Mammalian Matchmaker Assay. HEK-293T cells were transfected with plasmids to test this two-hybrid assay. After 72 hours, the SEAP detection assay was performed with CSPD as the substrate.

With the luminescence assay, cells produced much more SEAP when they received pM-53 and pVP16-T than when they received pM-53 and pVP16-CP or either pM-53 or pVP16-T alone (Figure 5.2). This result demonstrates that the two-hybrid system functions well in HEK-293T cells. While this system works well, luminescence assays are not as facile to perform as are fluorescence measurements, and so I replaced SEAP with GFP in the reporter plasmid to determine whether this system worked equally well for this reporter protein (Figure 5.3).

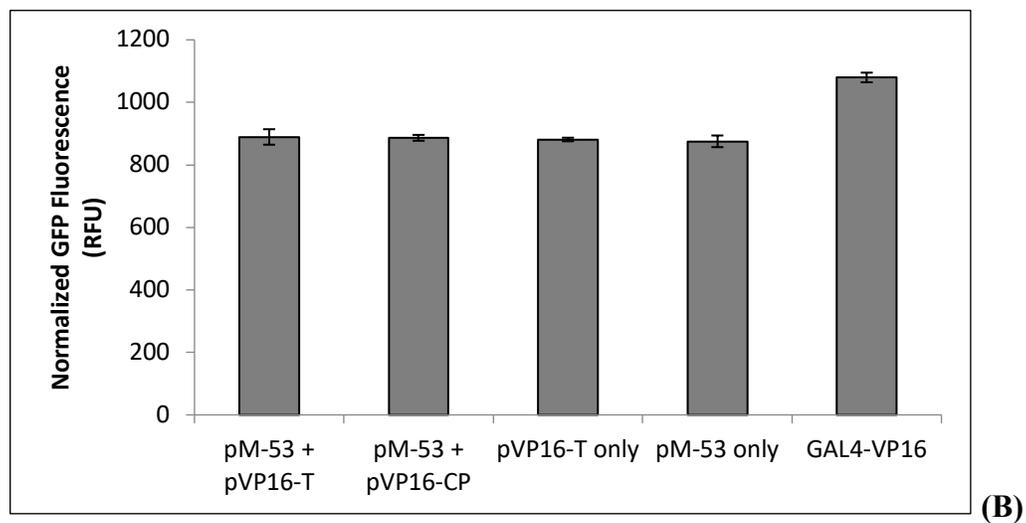
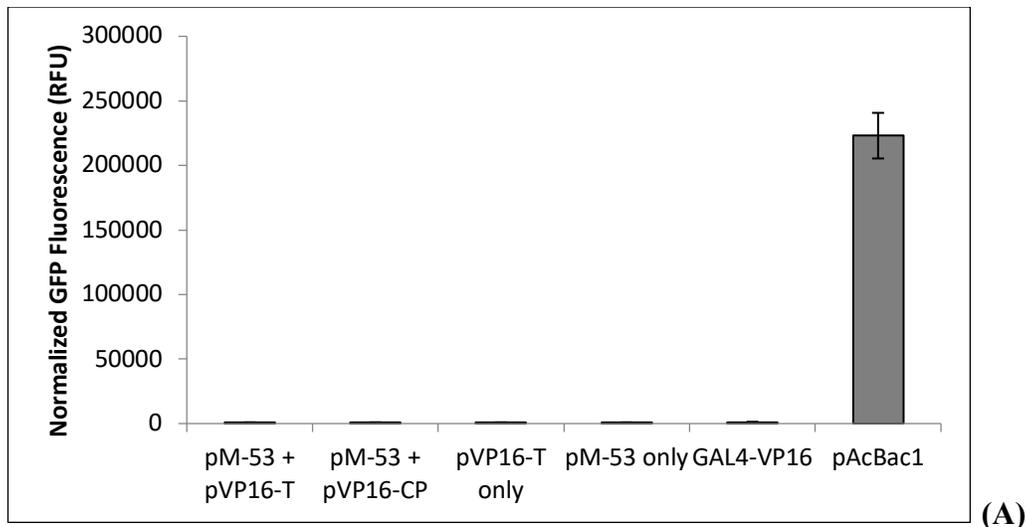


Figure 5.3. GFP fluorescence of cells using the two-hybrid system. SEAP was replaced with GFP in the Mammalian Matchmaker Assay, and this system was tested alongside a Gal4-VP16 fusion, with a constitutively expressed GFP as a control.

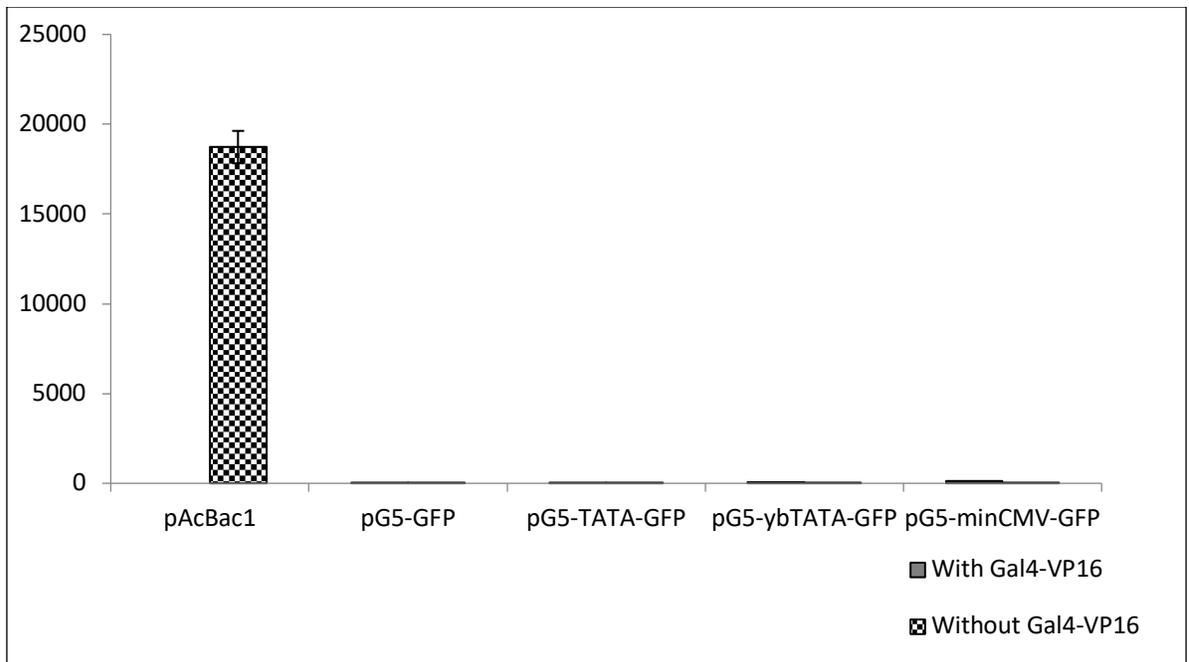
The vector pAcBac1-GFPwt acted as a control here, as transfection of this vector in mammalian cells results in high levels of expression of eGFP (Figure 5.3a). Transfection with the two-hybrid system of plasmids resulted in much lower levels of GFP production, as expected. Unlike the luminescence assay, there was not a large difference between the levels of GFP produced did not differ between treatments with pM-53 and pVP16-T, pM-53 and pVP16-CP, pM-53 alone, and pVP16-T alone (Figure 5.3b). However, the treatment

with the GAL4-VP16 fusion construct resulted in higher levels of GFP production, indicating that it is possible to differentiate between levels of association between the two-hybrid proteins based on this assay.

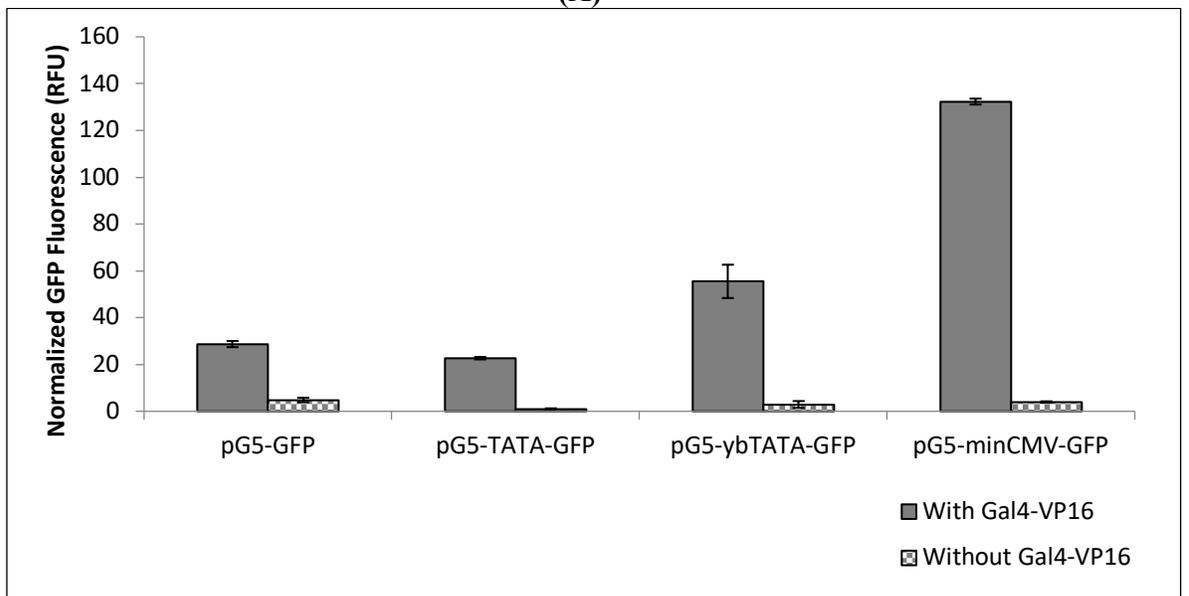
5.2.2 Testing the effects of minimal promoters on GFP production with a two-hybrid system

The original pG5 vector controls the transcription of the reporter protein (originally SEAP, which I swapped with GFP for convenience) with 5 consecutive GAL4 DNA binding domains and a Kozak sequence. To improve production of the reporter protein, I added different minimal promoters upstream of the Kozak sequence with the intention of improving transcription of the reporter protein gene. I tested 3 different minimal promoters: a TATA box, a YB-TATA box, and a minimal CMV synthetic promoter (see the appendix for promoter sequences). For these experiments, I used the GAL4-VP16 fusion construct to maximize transcription while testing which minimal promoter performed best.

In these experiments, the vector pAcBac1 acted as a control. Cells transfected with this plasmid produced much more GFP, which is expected because GFP is expressed constitutively from the CMV promoter in that vector. The YB-TATA and minCMV promoters performed better than the original construct, which lacked any promoter, as well as the TATA promoter (Figure 5.4). There was 4.6x as much fluorescence from cells which received GFP under the control of the minCMV promoter as from cells which received the original construct. Importantly, there was negligible fluorescence in the absence of the GAL4-VP16 fusion construct from all of the pG5-GFP plasmids tested. This system can be reliably used for two-hybrid selections, as even a small amount of leakiness from the promoter can cause false positives in a screen for protein binding partners.



(A)



(B)

Figure 5.4. Comparison of GFP fluorescence using minimal promoters. HEK-293T cells were transfected with Gal4-VP16 and GFP expressed from different minimal promoters. After 72 hours, cells were imaged and lysed, and GFP fluorescence was measured. (A) Including fluorescence from cells transfected with pAcBac1. (B) Only including fluorescence from cells using the two-hybrid system.

5.2.3 *Producing rAAV dependent on two-hybrid selection*

In order to create a system in which AAV can be used in a screen for mammalian protein binding partners, it is necessary for the production of AAV to be dependent on the association of Gal4 and VP16. To produce such a system, I used a construct in which VP1 is deleted from the genome of AAV and re-introduced under the control of the CMV promoter. In this case, I replaced GFP with VP1 in the pG5-minCV plasmid, so that transcription of VP1 was under the control of the five Gal4 binding domains and the synthetic minimal CMV promoter. In order for infectious AAV to be produced, the Gal4-VP16 construct must be present promote transcription of VP1, which is required for AAV to be infectious.

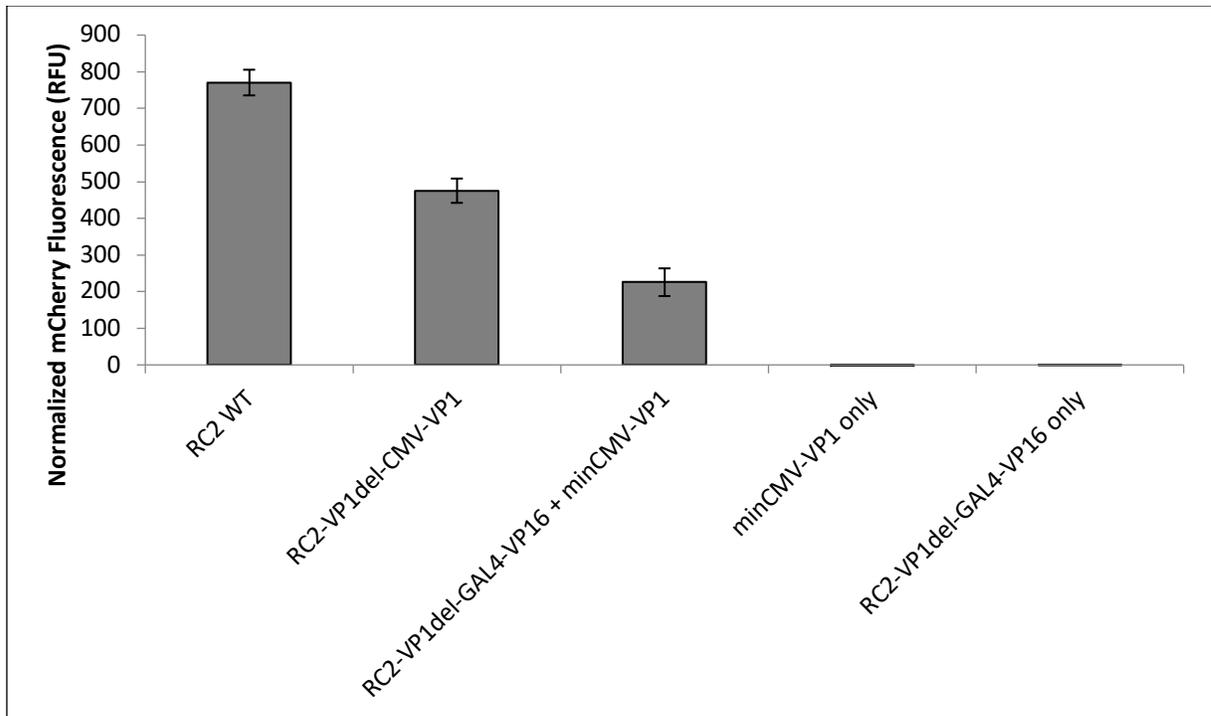


Figure 5.5. mCherry fluorescence of cells infected with rAAV-containing lysate. Cells were infected with rAAV produced by packaging cells to test the production of infectious virus when VP1 expression is controlled by Gal4-VP16. After 48 hours, these cells were imaged and lysed, and mCherry fluorescence was taken.

Production of AAV with the construct in which VP1 is deleted from the genome and then reintroduced on a CMV promoter (RC2-VP1del-CMV-VP1) was 61.7% of that produced from the wild type AAV2 genome (RC2wt), as measured by mCherry fluorescence of cells infected with lysate from HEK-293T cells packaging virus (Figure 5.5). When VP1 is provided on the pG5-minCMV-VP1 plasmid, and the rest of the AAV genome (RC2-VP1del) and GAL4-VP16 are provided on a different plasmid, mCherry fluorescence is 47.5% that from the RC2-VP1del-CMV-VP1 construct. This result indicates that a substantial quantity of output virus can be produced from this system.

There was negligible virus produced when only minCMV-VP1 was provided, as expected, since this treatment does not provide Rep, which is necessary for AAV DNA replication, or VP2 and VP3, which are necessary to produce the majority of the AAV capsid. Importantly, negligible infectious virus was produced when minCMV-VP1 was not provided, indicating that the production of VP1 from the minCMV promoter is indeed essential to produce infectious AAV and that the output of this virus can be used to determine whether GAL4 and VP16 are associated (Figure 5.6).

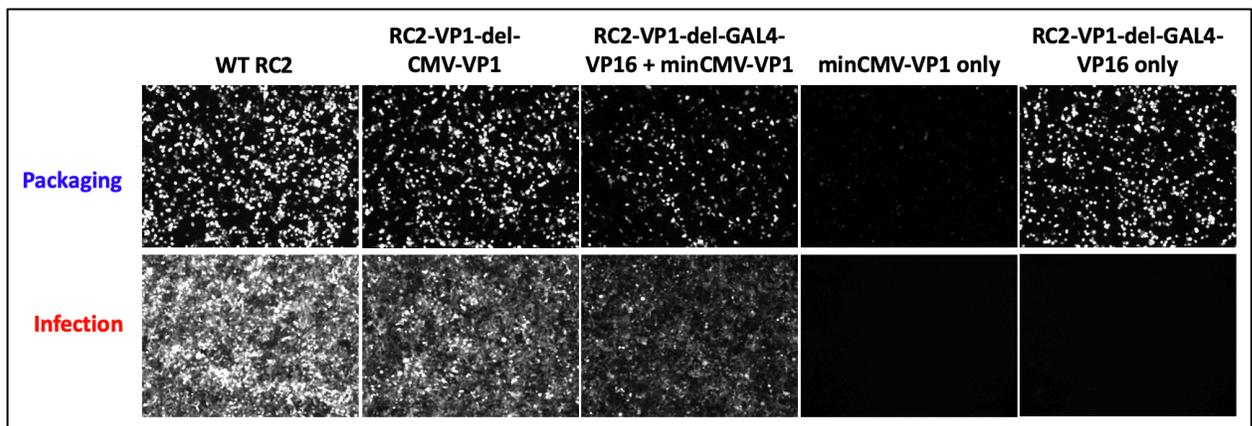


Figure 5.6. Cells packaging rAAV and cells infected with rAAV-containing lysate. Cells were infected with rAAV-mCherry and transfected plasmids to test the conditionality VP1 production dependent on Gal4-VP16 binding to the minCMV synthetic minimal

promoter. After 72 hours, the packaging cells were imaged and lysed, and new cells were infected with the lysate. These cells were imaged after 48 hours.

5.3 Conclusion

It is difficult to perform two-hybrid selections involving libraries of proteins in mammalian cells due to the technical challenge of being able to provide only one library member per cell. This virus-based system can circumvent that challenge and fill a gap in the available technology to enable high-throughput two-hybrid selections. Putting an essential viral protein, VP1, under the control of the GAL1 promoter enables infectious AAV to act as both a DNA delivery vector and the two-hybrid reporter. The platform presented here will be a valuable new application of directed evolution technology.

5.3.1 Future directions

This two-hybrid system, using VP1 as the output protein for the production of infectious AAV, has the potential to be a highly responsive selection system. One could provide a construct in which a bait protein is fused to GAL4 to cells via transient transfection, and also infect the cells with rAAV in which the genome is a library of proteins fused to VP16. With the two-hybrid selection system in which production of VP1 is tied to the association of GAL4 and VP16, infectious output virus would only be produced if the protein contained in the rAAV genome could bind to the bait protein. This system would allow researchers to scan the human proteome for binding partners of particular bait proteins of interest.

While transcription of VP1 from the minCMV promoter led to the production of a respectable quantity of AAV, the fluorescence output (and therefore likely infectious AAV production) was less than half that which was produced when VP1 was under the control of the larger CMV promoter. The difference between the activities of the typical CMV and

the minimal promoters was even more stark with GFP expressed constitutively or from the two-hybrid system. There appears to be a dynamic range in the activity of these promoters sufficient for the evolution of more active minimal promoters. The two-hybrid system, using infectious AAV as the reporter, would allow for this selection to be done with relative ease.

5.4. Materials and Methods

5.4.1 Strains and cell lines

Plasmid propagations were performed as described in Chapter 2 of this thesis. HEK-293T cells were maintained as described in Chapter 2.

5.4.2. Plasmids

All plasmid maps and sequences can be found in the Appendix.

pHelper, pAAV-mCherry+1xMmPytR, pIDTSMART-MbPylRS-RC2wt, pIDTSMART-MbPylRS-RC2-VP1del, and pIDTSMART-RC2-VP1del-CMV-VP1wt were propagated from Dr. Rachel Kelemen's stocks. pAcBac1-GFPwt was propagated from Dr. Yunan Zheng's stocks.

pG5-SEAP, pM53, pVP16-T, and pVP16-CP were part of the Matchmaker Mammalian Assay kit (Takara).

pECE-Gal4-VP16 was accessed from Addgene (plasmid #71728).

To produce pG5-GFP, eGFP was amplified from pAcBac1-GFPwt using the terminal primers XbaI-eGFP-R and SphI-Kozak-GFP-F, which also appended a Kozak sequence to the 5' end of the gene for GFP. pG5-SEAP was digested with XbaI and SphI, and GFP was cloned into the plasmid using these cut sites. The TATA sequence was appended to the 5' end of the GFP construct using the primer TATA-GFP-F. The YB

TATA sequence was added to the 5' end of the GFP construct using YB-TATA-GFP-F, and the minimal CMV promoter was added to the GFP construct using min-CMV-GFP-F. Each of these constructs was amplified using the reverse primer KpnI-GFP-R because pG5-GFP cannot be cut with XbaI due to DNA methylation that was not present in the context of pG5-SEAP. pG5-GFP was cut with KpnI and SphI, and the GFP constructs were cloned in to produce pG5-TATA-GFP, pG5-YBTATA-GFP, and pG5-minCMV-GFP.

The VP1-only construct was amplified out of pIDTSMART-RC2-VP1del-CMV-VP1wt using the terminal primers SphI-VP1-only-F and XbaI-VP1-only-R. The TATA, YB TATA, and minCMV sequences were appended to the 5' end of the VP1 construct using the respective forward primers SphI-TATA-VP1-only-F, SphI-YBTATA-VP1-only-F, and SphI-minCMV-VP1-only-F. These constructs were cloned into the pG5 backbone by digesting pG5-SEAP with SphI and XbaI and ligating in the digested VP1 constructs to produce pG5-VP1, pG5-TATA-VP1, pG5-YBTATA-VP1, and pG5-minCMV-VP1.

The Gal4-VP16 construct was amplified out of pECE-Gal4-VP16 using the terminal primers SbfI-Gal4-VP16-F and AvrII-Gal4-VP16-R. pIDTSMART-MbPyIRS-RC2-VP1del was digested with SbfI and AvrII, and the Gal4-VP16 construct was cloned in to produce pIDTSMART-RC2-VP1del-Gal4-VP16.

5.4.3 Preparing recombinant AAV

Recombinant AAV was produced as described in detail in Chapter 2 of this thesis.

5.4.4 Producing GFP with the use of two-hybrid binding partners

The following plasmids were transfected into HEK-293T cells to test the Matchmaker luminescence assay: pG5-SEAP with pM-53 and pVP-16T, pG5-SEAP with pM-53 and pVP-16CP, pG5-SEAP with pM-53 only, and pG5-SEAP with pVP16-T only.

Cells were lysed with CelLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease. Lysate luminescence was measured using the GreatEscAPE™ SEAP Chemiluminescence Detection Kit with CSPD as the substrate.

To determine whether it was possible to drive production of a protein of interest by relying on two-hybrid selection, the following plasmids were transfected into HEK-293T cells: pG5-GFP with pM-53 and pVP-16T, pG5-GFP with pM-53 and pVP-16CP, pG5-GFP with pM-53 only, pG5-GFP with pVP16-T only, pG5-GFP with pECE-GAL4-VP16, and pAcBac1-GFPwt as a control. Cells were lysed with CelLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease, and GFP fluorescence was measured.

5.4.5 Testing the effects of different minimal promoters on GFP production with a two-hybrid system

The following plasmids were transfected into HEK-293T cells to test whether certain minimal promoters could enhance the production of GFP while still maintaining dependence on this two-hybrid system: pG5-GFP alone, pG5-GFP with pECE-GAL4-VP16, pG5-TATA-GFP alone, pG5-TATA-GFP with pECE-GAL4-VP16, pG5-YBTATA-GFP alone, pG5-YBTATA-GFP with pECE-GAL4-VP16, pG5-minCMV-GFP alone, pG5-minCMV-GFP with pECE-GAL4-VP16, and pAcBac1-GFPwt as a control. Cells were lysed with CelLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease, and GFP fluorescence was measured.

5.4.6 Producing rAAV dependent on two-hybrid selection

To determine whether it is possible to tie AAV production to this two-hybrid system, these plasmids were transfected into HEK-293T cells, along with pHelper: pIDTSMART-MbPyIRS-RC2wt only, pIDTSMART-RC2-VP1del-CMV-VP1 only,

pIDTSMART-RC2-VP1del-GAL4-VP16 and pG5-minCMV-VP1, pG5-minCMV-VP1 only, and pIDTSMART-RC2-VP1del-GAL4-VP16 only. rAAV-mCherry-PytR was delivered via infection at 5 MOI, 4 hours before transfection. Cells were lysed with CellLytic™ M (Sigma-Aldrich) with 0.00001% Pierce Universal Nuclease. The resultant lysate was flash-frozen and stored at -80°C. Lysate was thawed and used to infect new HEK-293T cells with sodium butyrate to 1.0 mM. After 48 hours, cells were imaged and lysed, and mCherry fluorescence was measured.

5.5 References

- (1) Loregian, A.; Palù, G. *J. Cell. Physiol.* **2005**, *204* (3), 750.
- (2) Arkin, M. R.; Whitty, A. *Curr. Opin. Chem. Biol.* **2009**, *13* (3), 284.
- (3) Loregian, A.; Marsden, H. S.; Palù, G. *Rev. Med. Virol.* **2002**, *12* (4), 239.
- (4) Keegan, L.; Gill, G.; Ptashne, M. *Science (80-.)*. **1986**, *231* (4739), 699.
- (5) Fields, S.; Sternglanz, R. *Trends in Genetics*. Elsevier Current Trends August 1, 1994, pp 286–292.
- (6) Fields, S.; Song, O. K. *Nature* **1989**, *340* (6230), 245.
- (7) Fields, S. *Methods* **1993**, *5* (2), 116.
- (8) Karimova, G.; Pidoux, J.; Ullmann, A.; Ladant, D. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (10), 5752.
- (9) Finkel, T.; Duc, J.; Fearon, E. R.; Dang, C. V.; Tomaselli, G. F. *J. Biol. Chem.* **1993**, *268* (1), 5.
- (10) Patrício, D.; Fardilha, M. *Drug Discov. Today* **2020**, *25* (4), 764.
- (11) Fiebitz, A.; Nyarsik, L.; Haendler, B.; Hu, Y.-H.; Wagner, F.; Thamm, S.; Lehrach, H.; Janitz, M.; Vanhecke, D. *BMC Genomics* **2008**, *9*.
- (12) Girod, A.; Wobus, C. E.; Zádori, Z.; Ried, M.; Leike, K.; Tijssen, P.; Kleinschmidt, J. A.; Hallek, M. *J. Gen. Virol.* **2002**, *83* (5), 973.
- (13) Grieger, J. C.; Johnson, J. S.; Gurda-Whitaker, B.; Agbandje-McKenna, M.; Samulski, R. J. *J. Virol.* **2007**, *81* (15), 7833.
- (14) Srivastava, A.; Lusby, E. W.; Berns, K. I. *J. Virol.* **1983**, *45* (2).
- (15) Zhuang, Yonlong; Ma, Fei; Li-Ling, Jesse; Xu, Xiaofeng; Li, Y. *Mol. Biol. Evol.* **2003**, *20* (12), 1978.
- (16) Yona, A. H.; Alm, E. J.; Gore, J. *Nat. Commun.* **2018**, *9* (1), 1.
- (17) Yokobayashi, Y.; Weiss, R.; Arnold, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (26), 16587.
- (18) Tyo, K. E. J.; Nevoigt, E.; Stephanopoulos, G. Academic Press, 2011; Vol. 497, pp 135–155.
- (19) Alper, H.; Fischer, C.; Nevoigt, E.; Stephanopoulos, G. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (36), 12678.

Appendix I

Plasmid maps and sequences

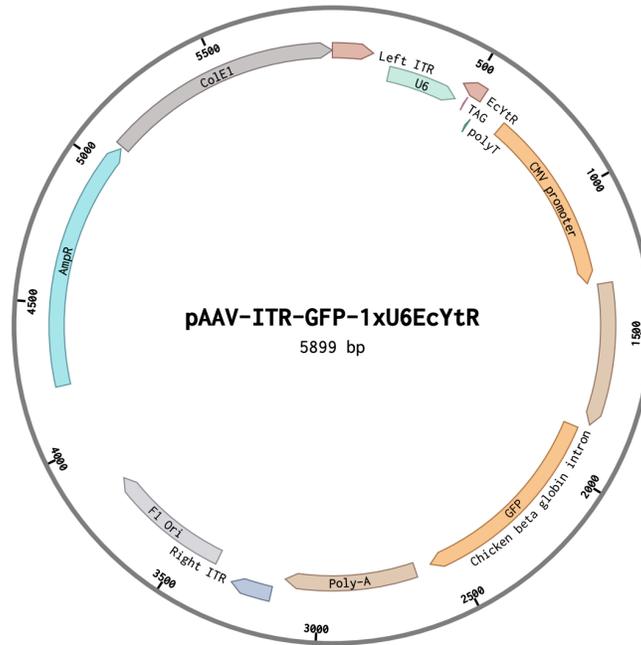
List of plasmid sequences

Chapter 2 Plasmids	122
pAAV-GFP+1xEcYtR.....	122
pAAV-mCherry+1xMmPytR.....	125
pHelper.....	128
pIDT-MbPylRS-RC2.....	133
pAcBac-Helper.....	138
pAcBac-MbPylRS-RC2.....	145
pIDTSMART-CMV-CAP-CSS.....	151
pAcBac1-Rep-CMV-Cap-CSS-CMV-MbPylRS.....	154
Chapter 3 Plasmids	161
pAAV-GFP+1xEcYtR.....	161
pAAV-GFP+1xMmPytR.....	162
pAAV-GFP+1xEcWtR.....	165
pIDTSMART-8xPytR-ITR-GFP.....	169
pHelper.....	173
pIDTSMART-MbPylRS-RC2.....	174
pIDTSMART-CMV-AAP.....	176
pIDTSMART-RC2-AAPstop60-CMV-AAP-CMV-WRS.....	181
pIDTSMART-RC2-AAPstop60-CMV-AAP.....	187
Chapter 4 plasmids	191
pAcBac-CMV-MbPylRS-CMV-GFP.....	191
pAcBac-4xPytR.....	198
pAcBac-4xPytR-MTH-mCherry.....	203
pAcBac1-GFPwt+linker.....	208
pAcBac1-mCherry-TAG-GFP.....	213
pIDTSMART-4xU6MmPytR.....	218
pAcBac-CMV-MbPylRS-WPRE-CAG-GFPtag39AvrIIIdel-WPRE.....	220
pAcBac-CMV-MbPylRS-CAG-GFPtagAvrIIIdel-4xPytR.....	226
pAcBac-CMV-MbPylRS-CAG-GFPtagAvrIIIdel-8xPytR.....	233
pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25.....	240
pAcBac-UbiC-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25.....	247
pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xPytR.....	254

Chapter 5 plasmids	264
pG5-SEAP.....	264
pG5-GFP.....	267
pG5-TATA-GFP.....	270
pG5-VP1.....	273
pG5-TATA-VP1.....	276
pG5-YBTATA-VP1.....	279
pECE-GAL4-VP16.....	282
pM-53.....	285
pVP16.....	289
pIDTSMART-MbPylRS-RC2.....	290
pIDTSMART-RC2-VP1del-CMV-VP1wt.....	292
pIDTSMART-RC2-VP1del-GAL4-VP16.....	292
pHelper.....	301
pAAV-GFP+1xMmPytR.....	302
pAcBac1-GFPwt.....	303

Chapter 2 Plasmids

pAAV-ITR-GFP+1xEcYtR



Sequence color-coding key

Feature	Color
Left ITR	text
U6 promoter	text
EcYtR	text
polyT	text
CMV promoter	text
Chicken beta globin intron	text
GFP	text
Poly-A	text
Right ITR	text
F1 Ori	text
Amp R	text
ColE1	text

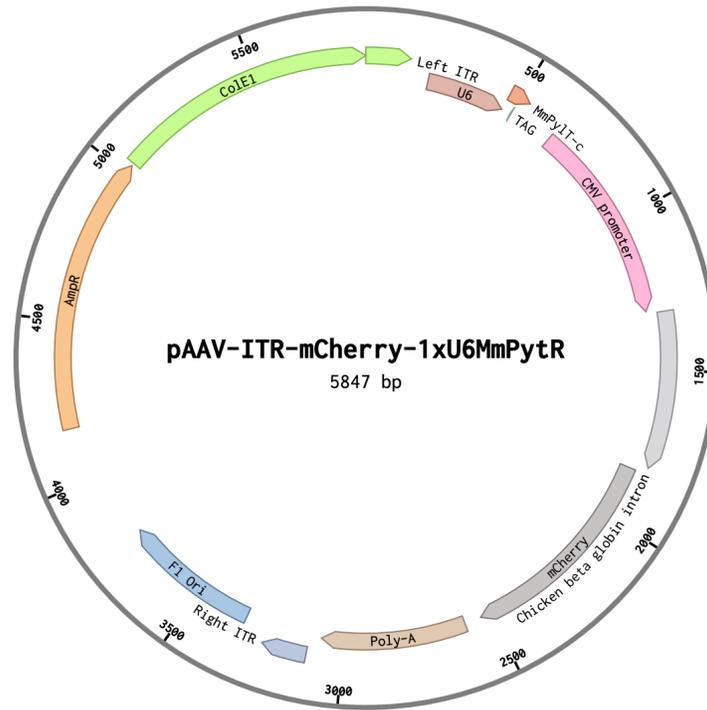
Sequence

```
cctgcaggcagctgcgcgctcgctcgctcactgaggccgccgggcaaagcccgggct
cgggcgacctttggtcgcccggcctcagtgagcgagcgagcgcgagagggagtggc
caactccatcactaggggttcctgcggccgcacgcgtctcgcggtccagtagtgatcga
cactgctcgatccgctcgccaccctaggtcggggcaggaagagggcctatttccatgat
```

tccttcatatTTTgcatatacGatacaaggctgTtagagagataattagaattaatttga
ctgtaaacaCAAagatattagTacaaaatacgtgacgtagaaagtaataatttcttggg
tagtttgcagtttttaaattatgTTTTaaatggactatcatatgcttaccgtaacttg
aaagtatttcgatttcttggctttatatatcttTgtgaaaggacgaaacaccggTggg
tccccgagcggcCAAagggagcagactcTaaatctgcccTcacagacttcgaaggTtcg
aatccttccccaccatTTTTgctagcggatcGacgagagcagcgcgactggatctgt
cgcccgtctCAAacgcaaccctccggcggTcgcatatcattcaggacgagcctcagact
ccagcgtAACacgcgtggagctagTtattaatagTaatcaattacggggTcattagTtc
atagcccatatatggagTtccgcgttacataactTaccgTaaatggcccgcctggctga
ccgcccAACgacccccgccattgacgtcaataatgacgtatgttcccatagTaaagTc
aatagggactttccattgacgtcaatgggtggagTattTaccgTaaactgccactTgg
cagtacatcaagtgtatcatatgccaagTaccccccctattgacgtcaatgacggTaaa
TggcccgcctggcattatgcccagTacatgaccttatgggactttcctactTggcagT
catctacgtattagTcatcgtattaccatggTgatgcggtttTggcagTacatcaatg
ggcgtggatagcggTttgactcaggggatttccaagTctccaccctattgacgtcaat
gggagTttgtttTgcacaaaatcaacgggactttccaaaatgtcgtAACaaactccgc
ccattgacgcaaatgggcggtaggcgtgTaccgTgggaggtctatataagcagagctcg
tttagTgaaccgtcagatcgcctggagacgccatccacgctgtttTgacctccatagaa
gacaccgggaccgatccagcctccgcggattcgaatccccggccgggAACggTgcattgg
aacgcggattccccgtgccaagagTgacgTaaagTaccgcctatagagTctataggccca
caaaaaatgctttcttcttttaataactTTTTgtttatcttatttctaaactttcc
ctaactctttctttcagggcaataatgatacaatgtatcatgcctctttgcaccattc
taaagaataacagTgataattctgggttaaggcaatagcaatattctgcatataaat
attctgcatataaattgTaaactgatgTaaagagTttcatattgctaaatagcagctaca
atccagctaccattctgcttttattttatggTtgggataaggctggattattctgagTc
caagctaggccctttTgctaactatgttcataaccttattcttccTccacagctcctg
ggcaacgtgctggTctgtgtgctggcccatcactttggcaagaattgggattcgaaca
tcgattgaattctgaatggTgagcaagggcgaggaggataaacatggccatcatcaagga
gttcatgcgcttcaaggTgcacatggagggctccgtgaacggccacgagTtcgagatcg
agggcgagggcgagggccgccctacgagggcaccagaccgccaagctgaaggTgacc
aagggtggccccctgcccttcgcctgggacatcctgtcccctcagTtcatgtacggctc
caaggcctacgtgaagcaccocgccgacatccccgactactTgaagctgtccttccccg
aggcttcaagTgggagcgcgtgatgaacttcgaggacggcggcgtggTgaccgtgacc
caggactcctccctgcaggacggcgagTtcatctacaaggtgaagctgcgoggcaacca
cttccccctccgacggccccgTaatgcagaagaagaccatgggctgggaggcctcctccg
agcggatgtaccccgaggacggcgcctgaagggcgagatcaagcagaggctgaagctg
aaggacggcggccactacgacgtgaggtcaagaccacctacaaggccaagaagcccgt
gcagctgcccggcctacaacgtcaacatcaagTtggacatcacctcccacaacgagg
actacaccatcgtggaacagTaccgaacgcgcgagggccgocactccaccggcggcatg
gacgagctgtacaagggatcctctagagTcgacctgcagaagcttgccctcgagcagcgc
tgctcgagagatctacgggtggcatccctgtgaccctccccagTgcctctcctggccc
TggaagTtggccactccagTgcccaccagcctTgtcctaataaaaattaaagTtgcatt
ttgtctgactaggtgtcctctataatattatggggTggagggggTggtatggagcaa
ggggcaagTtgggaagacaacctgtaggccctgcgggtctattgggaaccaagctgga
gtgcagTggcacaactcttggctcactgcaatctccgcctcctgggttcaagcattctc
ctgcctcagcctcccagTtggTgggattccaggcatgcatgaccaggctcagctaatt
tttgtttttTggtagagacggggTtccaccatattggccaggctggTctccaactcct
aatctcaggtgatctaccacctTggcctcccaaatTgctgggattacaggcgtgaacc

actgctcccttccctgtccttctgattttgtaggtaaccacgtgcggaccgagcggccg
caggaacccttagtgatggagttggccactccctctctgcgcgctcgctcactga
ggccggcgaccaaaaggtcgcccgacgccgggctttgccggggcgcctcagtgagcg
agcgagcgcgcagctgcctgcaggggcgccctgatgcggtatcttctccttacgcatctg
tgcggtatctcacaccgcatacgtcaaagcaaccatagtagcgcctgtagcggcgca
ttaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcct
agcgcctcctcttctccttcccttctccttctcgcacagttcgccggcttcccc
gtcaagctctaaatcgggggctcccttaggggtccgatttagtgcttacggcacctc
gaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagac
ggtttttcgcccttgacgttgagtcacggtctttaaatagtggactcttggtccaaa
ctggaacaacactcaacctatctcgggctattcttttgattataagggattttgccg
atttcggcctattggttaaaaaatgagctgatttaacaaaaatthaacgcgaatttta
caaatattaacgtttacaatttatggtgcactctcagtacaatctgctctgatgccg
catagttaagccagccccgacaccgccaacaccgctgacgcgcctgacgggcttgt
ctgctcccggcatccgcttacagacaagctgtgaccgctctccgggagctgcatgtgtca
gaggttttcaccgctcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctat
tttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcgg
gaaatgtgcgcggaaccctatttgttatttttctaaatacattcaaatatgtatcc
gctcatgagacaataaccctgataaatgcttcaataatattgaaaaggaagagtatga
gtattcaacatttccgtgtcgccttattcccttttttgccgcattttgcttccctgtt
tttgctcaccagaaacgctgggtgaaagtaaaagatgctgaagatcagttgggtgcacg
agtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccg
aagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcc
cgtattgacgcgggcaagagcaactcggctgcgcgatacactattctcagaatgactt
ggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaat
tatgcagtgctgccataaccatgagtgataaactgcggccaacttacttctgacaacg
atcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaactcg
ccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacacca
cgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttact
ctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccact
tctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagc
gtgggtctcgcggtatcattgcagcactggggccagatggttaagccctcccgtatcgta
gttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgtga
gataggtgcctcactgattaagcattggtaaactgtcagaccaagtttactcatatatac
tttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatcctttt
gataatctcatgacaaaatcccttaacgtgagttttcgttccactgagcgtcagacc
cgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgct
tgcaaacaaaaaaaccaccgctaccagcgggtggtttggttgccggatcaagagctacca
actctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttct
agtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcg
ctctgctaatectgttaccagtggtgctgctgccagtgggcgataagtcggtgcttaccggg
ttggactcaagacgatagttaccggataaggcgcagcggctcgggctgaacggggggttc
gtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtg
agctatgagaaagcgcacgcttccogaagggagaaagggcggacaggtatccggtaagc
ggcagggctcggaacaggagagcgcacgaggagcttccagggggaaacgcctggatct
ttatagtcctgtcgggttccgccacctctgacttgagcgtcgatttttgtgatgctcgt
cagggggggcggagcctatggaaaacgccagcaacgcggcctttttacggttcctggcc
ttttgctggccttttgcacatgt

pAAV- mCherry+1xMmPytR



Sequence color-coding key

Feature	Color
Left ITR	text
U6 promoter	text
MmPytR	text
polyT	text
CMV promoter	text
Chicken beta globin intron	text
mCherry	text
Poly-A	text
Right ITR	text
F1 Ori	text
Amp R	text
ColE1	text

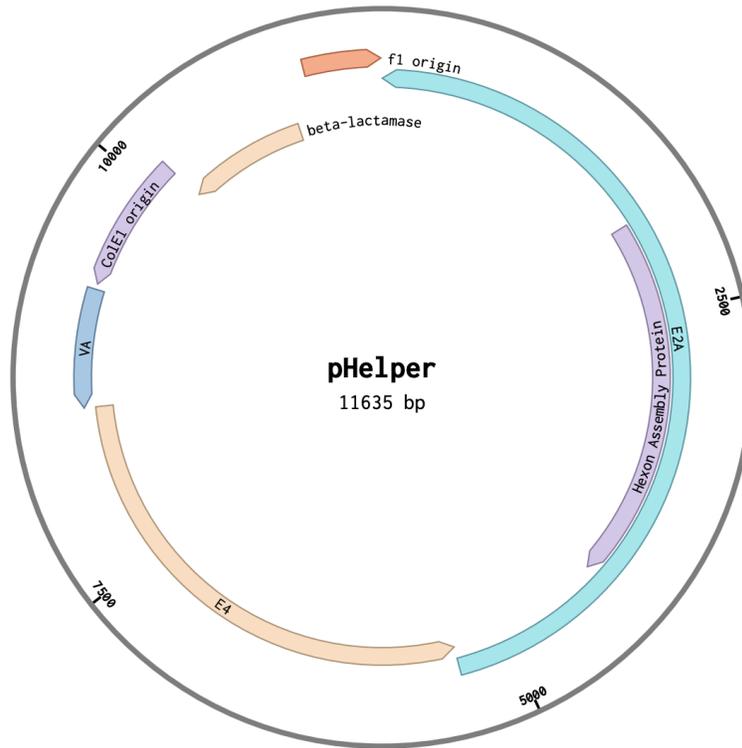
Sequence

cctgcaggcagctgcgcgctcgctcgctcactgaggccgccgggcaaagcccgggct
 cgggcgacctttggtcgcccgccctcagtgagcgagcgagcgcgagagaggagtggc
 caactccatcactaggggttcctgcggccgcacgcgtctcgcggtccagtagtgatcga
 cactgctcgatccgctcgcaccctaggtcgggcaggaagagggcctatttccatgat
 tccttcatatttgcatatacgatacaaggctgtagagagataattagaattaatttga
 ctgtaaacacaaagatattagtagcaaaatcgtgacgtagaaagtaataatttcttggg

tagtttgcagttttaaaattatgttttaaaatggactatcatatgottaccgtaacttg
aaagtatttcgatttcttggctttatataatcttgtggaaaggacgaaacaccggaaacc
tgatcatgtagatcgaacggactctaaatccgttcagccgggtagattcccggggtt
ccgttttttgctagcggatcgacgagagcagcgcgactggatctgtcgcctctcaaa
cgcaaccctccggcggtcgcataatcattcaggacgagcctcagactccagcgtaacacg
cgtggagctagttattaatagtaatacaattacgggggtcattagttcatagcccataat
ggagttccgcggttacataacttacggtaaatggcccgcctggctgaccgcccacgacc
cccgccattgacgtcaataatgacgtatgttcccatagtaacgtcaatagggactttc
cattgacgtcaatgggtggagtatttacggtaaactgccacttggcagtacatcaagt
gtatcatatgccaagtacgcccctattgacgtcaatgacggtaaatggcccgcctggc
attatgccagtacatgaccttatgggactttcctacttggcagtacatctacgtatta
gtcatcgtattaccatggatgctgggttttggcagtacatcaatgggcgtggatagcg
gtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgtttt
gcacaaaatcaacgggactttccaaaatgtcgtacaactccgccattgacgcaaa
tgggcggtaggcgtgtacgggtgggaggtctatataagcagagctcgtttagtgaaccgt
cagatcgcctggagacgccatccacgctgttttgacctccatagaagacaccgggaccg
atccagcctccgcggattcgaatcccgccgggaacgggtgcattggaacgcggattccc
cgtgccaaagagtacgtaagtaccgcctatagagtctataggcccacaaaaaatgcttt
cttcttttaataacttttttgtttatcttatttctaataactttccctaattctttct
ttcagggcaataatgatacaatgtatcatgocctctttgcaaccattctaagaataacag
tgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatata
aattgtaactgatgtaagaggttcattatgtaatagcagctacaatccagctaccat
tctgcttttattttatgggtgggataaggctggattattctgagtcgaagctaggccct
tttgctaatacatgttccatacctcttattcttccacagctcctgggcaacgtgctgg
tctgtgtgctggcccatacctttggcaaagaattgggattcgaacatcgattgaattct
gaatgggtgagcaagggcgaggagctgttaccgggggtgggtgccatcctgggtcgagctg
gacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccac
ctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgccctgacctggc
ccaccctcgtgaccaccctgacctacggcgtgcagtgcttccagccgctaccccgaccac
atgaagcagcagcacttcttcaagtccgccatgccgaaggctacgtccaggagcgcac
catcttcttcaaggacgacggcaactacaagaccgcgccgaggtgaagttcgagggcg
acaccctggtgaaccgcatcgagctgaagggtcagcttcaaggaggacggcaacatc
ctggggcacaagctggagtacaactacaacagccacaacgtctatataatgcccgacaa
gcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcg
tgcagctcgcgcgaccactaccagcagaacacccccatcggcgacggccccgtgctgctg
cccgacaaccactacctgagcaccagctccgccctgagcaaaagaccccacgagaagcg
cgatcacatgggtcctgctggagttcgtgaccgcccgggatcactctcggcatggacg
agctgtacaagtactcagatctcagagctcaagtagggatcctctagagtcgacctgcag
aagcttgctcgcagcagcgtgctcgcagagatctacgggtggcatccctgtgaccctc
cccagtgctctcctggccctggaagttgccactccagtgcccaccagccttctcctaa
taaaattaagttgcatcattttgtctgactaggtgtccttctataatattatggggtgg
aggggggtggatggagcaaggggcaagttgggaagacaacctgtagggcctgcccgggt
ctattgggaaccaagctggagtgcagtgccacaatcttggctcactgcaatctccgcct
cctgggttcaagcgattctcctgcctcagcctcccagttggtgggattccaggcatgc
atgaccaggctcagctaatttttgttttttggtagagacgggggtttccacatattggc
caggctggtctccaactcctaattctcaggtgatctaccaccttggcctcccacaaattgc
tgggattacaggcgtgaaccactgctcccttccctgtccttctgattttgtaggttaacc
acgtgcggaccgagcggccgcaggaacccttagtgatggagttggccactccctctctg

cgcgctcgctcgctcaactgaggccggcgaccaagggtcgcccgaagccgggctttgc
ccggggcgccctcagtgagcgagcgagcgcgagctgcctgcaggggcgccctgatgcgt
atcttctccttacgcacatctgtgcggatcttccacaccgcatacgtcaagcaaccatagt
acgcgcctgtagcggcgccattaagcgcggcgggtgtggtggttacgcgcagcgtgacc
gctacacttgccagcgccttagcgcgccgctcctttcgtttcttcccttcttctcgc
cacgttcgccggctttccccgtaagctctaaatcgggggctccctttagggttcogat
ttagtgccttacggcacctcgacccccaaaaacttgatttgggtgatggttcacgtagt
gggcatcgccctgatagacgggtttttcgccctttgacggttgagtcacggtctttaa
tagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttg
attataagggattttgccgatttcggcctattgggttaaaaaatgagctgatttaaca
aaatttaacgcgaattttaacaaaatattaacgtttacaatttatggtgactctcag
tacaatctgctctgatgcccatagtttaagccagccccgaccccccaaacaccgctg
acgcgcctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgct
tccgggagctgcatgtgtcagaggttttcaccgctcatcaccgaaacgcgcgagacgaa
ggcctcgtgatacgcctatcttttataggttaatgtcatgataataatggtttcttaga
cgtcagggtggcacttttcggggaaatgtgcgcggaaccctatctgtttatctttaa
atacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataata
ttgaaaaaggaagagtatgagatctcaacatttccggtgcgccttattcccttttttg
cggcattttgccttctgtttttgctcaccagaaacgctgggtgaaagtaaagatgct
gaagatcagttgggtgcacgagtggttacatcgaactggatctcaacagcggtaagat
ccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagtctgc
tatgtggcgcgggtattatcccgtattgacgcgggcaagagcaactcggtcgccgata
cactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacgga
tggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataaacactgcg
ccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcaaac
atgggggatcatgtaactgccttgatcgttggaaccggagctgaatgaagccatacc
aaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacggtgcgcaaacat
taactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcg
gataaagttgcaggaccacttctgcgctcggccttccggctggctggtttattgctga
taaactcggagccgggtgagcgtgggtctcgggtatcattgcagcactggggccagatg
gtaagccctcccgtatcgtagtattctacacgacggggagtcaggcaactatggatgaa
cgaatagacagatcgctgagataggtgcctcactgattaagcattggtaacctgtcaga
ccaagtttactcatatatacttttagattgatttaaaacttcatttttaatttaaagga
tctaggtgaagatcctttttgataatctcatgacaaaatccctaacgtgagttttcg
ttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttt
tctgcgcgtaatctgctgcttgcacacacacacacacacacacacacacacacacac
tgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcag
ataccaaataactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgt
agcaccgcctacatacctcgtctgctaatoctgttaccagtggtgctgctgccagtggcg
ataagtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcgg
tcgggctgaacggggggttcgtgcacacagcccagcttgagcgaacgacctacaccga
actgagatacctacagcgtgagctatgagaaagcgcacgcttcccgaagggagaaagg
cggacaggtatccggtaagcggcagggctcgaacaggagagcgcagaggagcttcca
gggggaaacgcctggatctttatagtcctgtcgggtttcggccaccttgacttgagcg
tcgatttttgatgctcgtcaggggggagcctatggaaaaacgcagcaacgcgg
cctttttacgggtcctggccttttgctggccttttgctcacatgt

pHelper



Sequence color-coding key

Feature	Color
E2A	text
E4	text
VA	text
ColE1 ori	text
Beta-lactamase	text
F1 ori	text

Sequence

```

ggtagcccaactccatgcttaaacagtcccaggtacagcccaccctgctcgcaaccagg
aacagctctacagcttccctggagcgccactcgcctacttccgcagccacagtgcgcag
attaggagcgccacttctttttgtcacttgaaaaacatgtaaaaataatgtactaggag
acactttcaataaaggcaaatgtttttatttgtacactctcgggtgattatttaccccc
cacccttgccgtctgcgccgtttaaaaatcaagggttctgccgogcatcgctatgcg
ccactggcagggacacgcttgcgatactggtgttagtgctccacttaaactcaggcaca
accatccgcggcagctcggtaggttttactccacaggctgcgcaccatcaccaacgc
gttagcaggtcgggcgccgatatcttgaagtcgcagttggggcctccgccttgcgcgc
gaggttgcgatacacagggttgcagcactggaacactatcagcgcgggtggtgcacg
ctggccagcacgctcttgtcggagatcagatccgcgtccaggtcctccgcgttgctcag
    
```

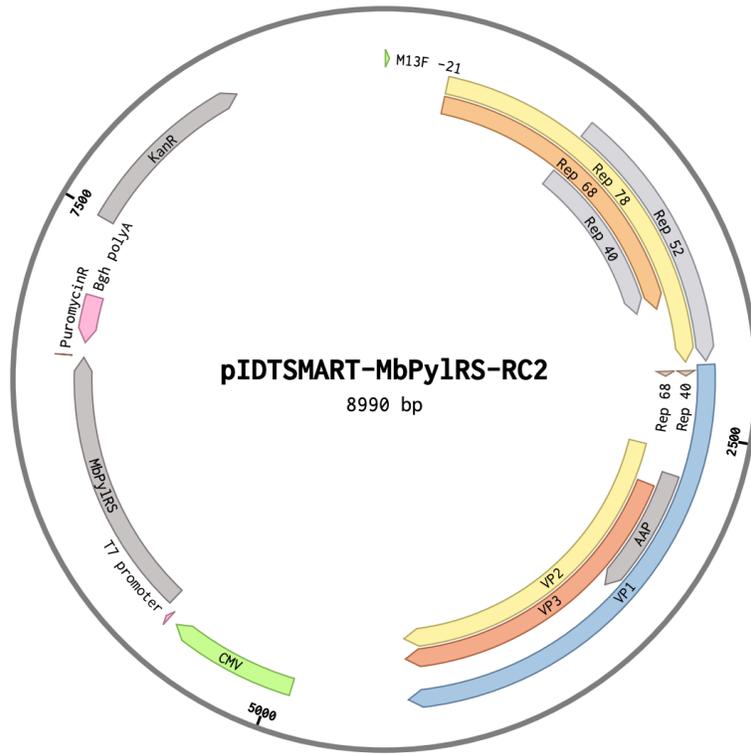
ggcgaacggagtcaactttggtagctgccttcccaaaaaggggtgcatgccaggtttg
agttgactcgcaccgtagtggcatcagaaggtgaccgtgcccggtctggcgtagga
tacagcgcctgcatgaaagccttgatctgcttaaaagccacctgagcctttgcgccttc
agagaagaacatgccgcaagacttgccggaaaactgattggccggacagggccgcgcat
gcacgcagcaccttgcgtcgggtgttggagatctgcaccacatttcggccccaccggttc
ttcacgatcttggccttgctagactgctccttcagcgcgcgctgcccgttttcgcctct
cacatccattttcaatcacgtgctccttatttatcataatgctcccgtgtagacacttaa
gctcgccttcgatctcagcgcagcgggtgcagccacaacgcgcagcccgtgggctcgtgg
tgctttaggttacctctgcaaagcactgcaggtacgcctgcaggaatcgcacctcat
cgtcacaaggtcttgggtgctggtgaaggtcagctgcaaccgcgggtgctcctcgttta
gccaggtcttgatacggccgagagcttccacttgggtcaggcagtagcttgaagttt
gccttttagatcgttatccacgtggtacttgtccatcaacgcgcgcgagcctccatgcc
cttctcccacgcagacacgatcggcaggctcagcgggtttatcacctgctttcacttt
ccgcttactggactcttcttttctcttgcgtccgcataccccgcgccactgggtcg
tcttcattcagccgcgcacogtgcgcttacctcccttgcogtgccttgattagcaccgg
tggttgcgtgaaaccaccattttagcgcacacatcttctcttcttctctcgcgtgcca
cgatcacctctggggatggcgggctcgggcttgggagaggggcgcttcttttcttt
ttggacgcaatggccaaatccgccgtcgaggtcgatggccgcgggctgggtgtgcgcgg
caccagcgcaccttgtgacgagtcttctcgtcctcggactcgagacgccgcctcagcc
gcttttttgggggcgcggggagggcgggcagggcgaggggacgacacgtcctcc
atggttgggtggacgtcgcgcgcacccgcgtccgcgtcgggggtgggttcgcgctgctc
ctctcccactggccatttcttctctctataggcagaaaaagatcatggagtgcgtcg
agaaggaggacagcctaaccgccccctttgagttcgcaccaccgcctccaccgatgcc
gccaacgcgcctaccaccttccccgtcgaggcaccoccgcttgaggaggaggaagtgat
tatcgagcaggaccaggttttgtaagcgaagacgacgaggatcgtcagtaccaacag
aggataaaaagcaagaccaggacgacgcagaggcaaacgaggaacaagtcggggcggggg
gaccaaaaggcatggcgactacctagatgtgggagacgacgtgctggtgaagcatctgca
gcgccagtgcgccattatctgcgacgcgttgcaagagcgcagcgcgatgtgccctcgcca
tagcggatgtcagccttgctacgaacgccacctgttctcaccgcgcgtacccccaaa
cgccaagaaaacggcacatgcgagcccaaccgcgcctcaacttctaccocgtatttgc
cgtgccagaggtgcttggcacctatcacatctttttccaaaactgcaagataccctat
cctgccgtgccaaccgcagccgagcggacaagcagctggccttgcggcagggcgctgct
atacctgatctgcctcgcctcgacgaagtgccaaaaatctttgagggcttggacgcga
cgagaaacgcgcggcaaacgctctgcaacaagaaaacagcgaataatgaaagtcactgtg
gagtgctggtggaacttgaggggtgacaacgcgcgcctagccgtgctgaaacgcagcatc
gaggtcaccactttgcctaccggcacttaacctaccccccaagggtatgagcacagt
catgagcgcgctgatcgtgcgccgtgcacgaccctggagagggatgcaaaccttgaag
aacaaccgaggagggctaccgcagttggcgatgagcagctggcgcgctggcttgag
acgcgcgagcctgccgacttggaggagcgaocgaagctaatgatggccgcagtgcttgt
taccgtggagcttgagtgcatgcagcggttcttctgctgaccggagatgcagcgaagc
tagaggaaacggtgcaactacactttcgccagggtacgtgcgccaggcctgcaaaatt
tccaacgtggagctctgcaacctggctcctaccttggaaattttgcacgaaaaccgcct
cgggcaaacgtgcttcatccacgctcaagggcgagggcgccgcgactacgtccgcg
actgcgttacttatttctgtgctacacctggcaaacggccatggcgctgtggcagcaa
tgcttggaggagcgaacctaaaggagctgcagaagctgctaaagcaaaaacttgaagga
cctatggacggccttcaacgagcgcctccgtggccgcgcacctggcggacattatcttcc
ccgaacgcctgcttaaaacctgcaacaggggtctgccagacttaccagtc aaagcatg
ttgcaaaaacttttaggaactttatcctagagcgttcaggaattctgcccgcacactgctg

tgcgcttcctagcgaactttgtgccattaagtaccgtgaatgccctccgcgctttggg
gtcactgctaccttctgcagctagccaactaccttgctaccactccgacatcatggaa
gacgtgagcgggtgacggcctactggagtgtcactgtcgtgcaacctatgcaccccgca
ccgctccctgggtctgcaattcgcaactgcttagcgaagtcaaattatcggtacctttg
agctgcaggggtccctcgcctgacgaaaagtccgcggctccgggggttgaaactcactccg
gggctgtggacgtcggctaccttcgcaatttgtacctgaggactaccacgcccacga
gattaggttctacgaagaccaatcccgcgcccaatgctggagcttaccgcctgcgtca
ttaccaggggccacatccttgccaattgcaagccatcaacaagcccgcgaagagttt
ctgctacgaaaggacgggggggtttacctggacccccagtcggcgaggagctcaacc
aatcccccgccgcgcagccctatcagcagccgcggggcccttgcttcccaggatggca
cccaaaaagaagctgcagctgccgcgcgccaccacggacgaggaggaatactggga
cagtcaggcagaggaggttttgacgaggaggaggatgatggaagactgggacagcc
tagacgaagcttccgaggccgaagaggtgtcagacgaaacaccgtcacctcggtcgca
ttcccctcgcggcgccccagaaattggcaaccgttcccagcatcgctacaacctccgc
tctcagggcgcggcggcactgctgttcgcccagccaaccgtagatgggacaccactg
gaaccaggggccgtaagtctaagcagccgcgcgcttagcccaagagcaacaacagcgc
caaggctaccgctcgtggcgcgggcacaagaacgccatagttgcttgcttgcaagactg
tgggggcaacatctccttcgcccgcgctttcttctctaccatcacggcgtggccttcc
cccgtaacatcctgcattactaccgtcatctctacagcccctactgcaccggcggcagc
ggcagcggcagcaacagcagcggtcacacagaagcaaggcgaccggatagcaagactc
tgacaaagcccaagaaatccacagcggcggcagcagcaggaggaggagcgtgcgtctg
gcccacaacgaaccgctatcagcccgcgagcttagaaataggattttcccactctgta
tgctatattcaacaagcaggggccaagaacaagagctgaaaataaaaaacaggtctc
tgcgctccctcaccgcagctgctgtatcacaagaagcgaagatcagcttcggcgcagc
ctggaagacgoggaggctctctcagcaataactgcgcgctgactcttaaggactagtt
tcgcccctttctcaaatttaagcgcgaaaactacgtcatctccagcggccacaccgg
cgccagcacctgtcgtcagcgcattatgagcaaggaaattcccacgccctacatgtgg
agttaccagccacaaatgggacttgccggctggagctgccaagactactcaaccgcaat
aaactacatgagcgcgggaccccacatgatatcccgggtcaacggaatccgcgcccacc
gaaaccgcaattctcctcgaacaggcggctattaccaccacacctcgtaataaccttaat
ccccgtagttggcccgtgccctgggtgtaccaggaaagtcccgtcccaccactgtggt
actcccagagacgcccaggccgaagttcagatgactaactcaggggagcagcttgccg
gcccgtttcgtcacaggggtgcggtcgcggggcgttttagggcggagtaacttgcatgt
attgggaattgtagtttttttaaaatgggaagtgacgtatcgtgggaaaacggaagtga
agatttgaggaagttgtgggttttttgctttctggttctgggcgtaggttcgcgtgcgg
ttttctgggtgtttttgtggactttaaccgttacgtcatttttttagtcctatatatac
tcgctctgtacttgccctttttacactgtgactgattgagctggtgccgtgtcagagt
gtgttttttaatataggttttttactggtaaggctgactgttatggctgccgctgtggaa
gcgctgtatgttgttctggagcgggagggtgctattttgcctaggcaggagggtttttc
aggtgtttatgtgtttttctctcctattaattttgttataacctcctatgggggctgtaa
tgttgtctctacgcctgcgggtatgtattccccgggctatctcggtcgttttttagca
ctgaccgatgttaaccaacctgatgtgtttaccgagtcctacattatgactccggacat
gaccgaggaactgtcgggtgggtgcttttttaacacgggtgaccagtttttttacggtcacg
ccggcatggcogtagtccgtcttatgcttataaggggtgtttttcctgttgtaagacag
gcttctaagtgttaaatgtttttttttgttattttttgtgtttaatgcaggaac
ccgcagacatgtttgagagaaaaatgggtgtctttttctgtgggtgggtccggaacttacc
tgcctttatctgcatgagcatgactacgatgtgcttgcttttttgccgcgaggctttgcc
tgattttttgagcagcaccttgcatttttatatcgccgcccatgcaacaagcttacatag

gggctacgctggtagcatagctccgagtatgctgtcataatcagtggtgggttctttt
gtcatgggtcctggcggggaagtggccgctggtccgtgcagacctgcacgattatgt
tcagctggccctgcgaaggacctacgggatcgcggtatTTTTgttaatgttccgctt
tgaatcttatacaggtctgtgaggaacctgaatTTTTgcaatcatgattcgtgcttga
ggctgaaggtggagggcgctctggagcagatTTTTacaatggccggacttaatatcgg
gatttgcttagagacatattgataaggtggcgagatgaaaattatttgggcatggttga
agtgctggaatgtttatagaggagattcaccctgaagggttagcctttacgtccact
tgacgtgagggcagtttgccttttggagccattgtgcaacatcttacaatgccatt
atctgttctttggctgtagagtttgaccacgccaccggaggggagcgcgttacttaat
agatcttcattttgaggttttggataatcttttggataaaaaaaaaaaaaaacatggttc
ttccagctcttcccgtcctcccgtgtgtgactcgcagaacgaatgtgtaggtggctg
gggtgtggcttattctgcggtgggtggatgttatcagggcagcggcgcaggaagggtta
catagaaccggaagccagggggcgcctggatgctttgagagagtgatataactacaact
actacacagagcagcgaagcagcagaccggagacgcagatctgtttgtcacgcccgc
acctggttttgcttcaggaatatgactacgtccggcgttccatttggcatgacactac
gaccaacacgatctcgggtgtctcggcgcactccgtacagtagggatcgcctacctcct
tttgagacagagaccgcgctaccatactggaggatcatccgctgctgcccgaatgtaa
cactttgacaatgcacaacgtgagttacgtgcgaggtcttccctgcagtggtggattta
cgctgattcaggaatgggttgttccctgggatatgggtctgacgcgggaggagcttcta
atcctgaggaagtgtatgcacgtgtgctgtgttgtgccaacattgatatacagcag
catgatgatccatggttacgagtcctgggctctccactgtcattgttccagtcccgggt
ccctgcagtgcatagccggcgggaggttttggccagctgggttaggatgggtgggtggat
ggcgcctatgtttaatcagaggtttatatgggtaccgggaggtgggtgaattacaacatgcc
aaaagaggtaatgtttatgtccagcgtgtttatgaggggtcgcacttaatctacctgc
gcttgtggatgatggccacgtgggttctgtggtccccgccatgagctttggatacagc
gccttgactgtgggattttgaacaatattgtgggtgctgtgctgcagttactgtgctga
tttaagtgagatcaggggtgcgctgctgtgcccgaggacaaggcgtctcatgctgcggg
cggtgcgaatcatcgtgaggagaccactgccatggttgtattcctgcaggacggagcgg
cggcggcagcagtttattcgcgcgctgctgcagcaccaccgccctatcctgatgcacga
ttatgactctacccccatgtaggcgtggacttccccctcgcgcggcgttgagcaaccgc
aagttggacagcagcctgtggctcagcagctggacagcgcacatgaacttaagcagcgtg
cccggggagtttattaatatcactgatgagcgtttggctcgcagaggaaaccgtgtggaa
tataacacctaagaatatgtctgttaccatgatgatgctttttaaggccagccggg
gagaaaggactgtgactctgtgtgttgggagggaggtggcaggttgaatactaggggt
ctgtgagtttgattaaggtacggtgatcaatataagctatgtgggtgggtgggctatact
actgaatgaaaaatgacttgaattttctgcaattgaaaaataaacacgttgaacata
acatgcaacaggttcacgattctttattcctgggcaatgtaggagaaggtgtaagagtt
ggtagcaaaaagtttcagtggtgtatTTTccactttcccaggacctgtaaaagacatag
agtaagtgttacctcgttagtttctgtggattcactagaatcgatgtaggatgttgcc
cctcctgacgcggtaggagaaggggaggtgccctgcatgtctgccgctgctcttgctc
ttgccgctgctgaggagggggcgcacatctgccgcagcaccggatgcatctgggaaaagc
aaaaaaggggctcgtccctgtttccggaggaatttgcaagcggggcttgcacgaggg
gaggaaccctcgttcgccgcagtcggccggcccgagactcgaaccgggggtcctgc
gactcaacccttggaaaataaccctccggctacagggagcagaccacttaatgctttcg
ctttccagcctaaccgcttacgccgcgcggccagtggccaaaaaagctagcgcagca
gccgccgcgcttggaggaagccaaaaggagcgcctccccgttgtctgacgtcgcacac
ctgggttcgacacgcggggcggttaaccgatggatcacggcggacggccggatccgggt
tcgaaccctggctcctccatgatacccttgcgaatttatccaccagaccacggaaga

gtgccgcttacaggctctccttttgacggcttagagcgtcaacgactgogcagcct
caccggccagagcgtcccaccatggagcactttttgccgctgcgcaacatctggaacc
ggtcccgactttccgcgcgctccaccaccgccggccatcacctggatgtccagg
tacatctacggattacgtcgacggttaaacatgatcagctcactcaaaggcggtaa
tacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccag
caaaaggccaggaaccgtaaaaaggccgcttgctggcgtttttccataggctccgcc
ccctgacgagcatcacaanaatcgacgctcaagtcagaggtggcgaaaccgcagggac
tataaagataaccaggcgtttcccctggaagctccctcgtgogctctcctggtccgacc
ctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctca
tagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtg
tgcacgaacccccggttcagcccagccgctgogccttatccggtaactatcgtcttgag
tccaaccggtaagacacgacttatcgccactggcagcagccactggtaacaggattag
cagagcgaggtatgtaggcggtgctacagagttcttgaagtggcctaactacggct
acactagaagaacagtatttggtatctgogctctgctgaagccagttaccttcggaaaa
agagttggtagctcttgatccggcaaaacaaaccaccgctggtagcgggtggttttttgt
ttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttt
ctacggggtctgacgctcagtggaacgaaaactcacgttaagggttttggtcatgaga
ttatcaaaaaggatcttcacctagatccttttaattaaaaatgaagtttaaatcaat
ctaaagtatatatgagtaaaacttgggtctgacagttaccaatgcttaactcagtgaggcac
ctatctcagcgatctgtctatttcgttcacatagttgctgactccccgctgctgtag
ataactacgatacgggagggcttaccatctggcccagtgctgcaatgataccgcgaga
cccacgctcaccggctccagatttatcagcaataaaccagccagccggaaggccgagc
gcagaagtggctcctgcaactttatccgcctccatccagctctattaattggtgccgggaa
gctagagtaagtagttcggccagttaatagtttgcgcaacggttggtgccattgctacagg
catcgtggtgtcacgctcgtcgttttggtatggcttcattcagctccggttcccaacgat
caaggcgagttacatgatccccatggtgtgcaaaaaagcgggttagctccttcggctcct
ccgatcgttgctcagaagtaagttggccgcagtggtatcactcactggttatggcagcact
gcataattctcttactgtcatgccatccgtaagatgcttttctgtgactgggtgagtact
caaccaagtcattctgagaatagtgatgcggcgaccgagttgctcttgccggcgctca
atcgggataataaccgcgccacatagcagaactttaaaagtgctcatcattggaaaacg
ttcttcggggcgaaaactctcaaggatcttaccgctggtgagatccagttcgatgtaac
ccactcgtgcaccaactgatcttcagcatcttttactttcaccagcgtttctgggtga
gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggcgacacggaaatggtg
aatactcactccttcttttcaatattattgaagcatttatcagggttattgtctca
tgagcggatacatatgtgatgtatttagaaaaataaacaataggggttccgcgcaca
tttccccgaaaagtgccacctaaattgtaagcgttaatattttgttaaaattcgcgta
aatttttgttaaatcagctcattttttaaccaataggccgaaatcggcaaaatccctta
taaatcaaaagaatagaccgagatagggttgagtgttggtccagtttggaacaagagtc
cactattaaagaacgtggactccaacgtcaaagggogaaaaaccgtctatcaggggogat
ggccactacgtgaaccatcacctaatcaagttttttggggctcagagtgccgtaaaagc
actaaatcggaaccctaaaggagccccgatttagagcttgacggggaaagccggcga
acgtggcgagaaaggaagggaagaaagcgaaaggagcggcgctagggcgctggcaagt
gtagcgggtcacgctgogcgttaaccaccacaccgcgcttaatgogccgctacaggg
cgcatgatcc

pIDT-MbPylRS-RC2



Sequence color-coding key

Feature	Color
Rep	text
Cap	text
CMV promoter	text
T7 promoter	text
MbPylRS	text
Bgh polyA	text
KanR	text

Sequence: Wild type RC2

```

cccggtgtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
gaaggtgagccagtgagttgattgcagtcacgctggagctctgaggctcgtcctg
aatgatatgacgaccgccggagggttgctttgagacggcgacagatccagtcgcgctg
ctctcgtcgatccgctagggcggccgctctagaactagtgatcccccggaagatcaga
agttcctattccgaagttcctattctctagaaagtataggaacttctgatctgcbcagc
cgccatgcccggggttttacgagattgtgattaaggtcccagcgaccttgacgagcattc
tgcccggcatttctgacagctttgtgaactgggtggccgagaaggaatgggagttgccg
ccagattctgacatggatctgaatctgattgagcaggcaccctgaccgtggccgagaa
gctgcagcgcgacttctgacggaatggcgcgctgtgagtaaggccccggaggcccttt
    
```

tctttgtgcaat ttgagaagggagagagctacttccacatgcacgtgctcgtggaacc
accggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaat ttaccgcggtatcgagccgactttgccaaactggttcgcggtcacaaga
ccagaaatggcgccggaggcggaacaaggtggtggatgagtgctacatccccattac
ttgctcccaaaaaccagcctgagctccagtgggcgtggactaatatggaacagtat tt
aagcgctgtttgaatctcacggagcgtaaacggttggtggcgcagcatctgacgcacg
tgtcgcagacgcaggagcagaacaagagaatcagaatcccaattctgatgcgccggtg
atcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggacaaggg
gattacctcggagaagcagtggtccaggaggaccaggcctcatacatctccttcaatg
cggcctccaactcgcggtcccaaatcaaggctgccttggaacatgoggaaagattatg
agcctgactaaaaccgccccgactacctggtgggcccagcagcccgtggaggacatttc
cagcaatcggatttataaaaat ttggaactaaacgggtacgatccccaatatgcggtt
ccgtctttctgggatgggcccagaaaaagttcggcaagaggaacaccatctggctggtt
gggctgcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgccctt
ctacgggtgcgtaaactggaccaatgagaactttccttcaacgactgtgtcgacaaga
tggtgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggccaaagcc
attctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagataga
cccgactcccgtgatcgtcacctccaacaccaacatgtgcgcctgattgacgggaact
caacgaccttcgaacaccagcagccgttgcaagaccggatggtcaaatttgaactcacc
cgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagactttttccg
gtgggcaaaggatcacgtggttgagggtggagcatgaattctacgtcaaaaagggtggag
ccaagaaaagaccgccccagtgacgcagatataagtgagcccaaacgggtgcgcgag
tcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggta
ccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttcctgcagacaat
gcgagagaatgaatcagaattcaaatatctgcttcaactcacggacagaaagactgttta
gagtgctttcccggtgcagaatctcaaccgtttctgtcgtcaaaaaggcgtatcagaa
actgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctgcgatc
tggtcaatgtggatttggtgactgcacttttgaacaataa atgatttaaatcagggtat
ggctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataagac
agtgggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggac
gacagcaggggtcttggtcttccctgggtacaagtacctcggacccttcaacggactcga
caagggagagccggtcaacgaggcagacgccgcggccctcgagcacgacaaaagcctacg
accggcagctcgacagcggagacaaccgctacctcaagtacaaccacgccgacgcggag
tttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagctct
ccaggcgaaaaagaggggtcttgaacctctgggctggttgaggaacctgttaagacgg
ctccgggaaaaaagaggccggttagagcactctcctgtggagccagactcctcctcggga
accggaaaggcgggcccagcagcctgcaagaaaaagattgaattttggtcagactggaga
cgcagactcagtaacctgacccccagcctctcggacagccaccagcagccccctctggtc
tgggaaactaatac gatggctacaggcagtgggcgcaccaatggcagacaataacgagggc
gccgacggagtggttaattcctcgggaaattggcattgcgattccacatggatgggcca
cagagtcataccaccagcaccgcaacctgggcccctgccacctacaacaaccacctct
acaacaaaat ttccagccaatcaggagcctcgaacgacaatcactactttggctacagc
acccttgggggtat ttgacttcaacagattccactgccacttttaccacgtgactg
gcaaagactcatcaacaactggggattccgacccaagagactcaacttcaagctct
ttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgccaataac
cttaccagcacggttcaggtggttactgactcggagtagcagctcccgtacgtcctcgg
ctcggcgcacaaaggtgcctcccgcggtccagcagacgtcttcatgggtgccacagt
atggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgc

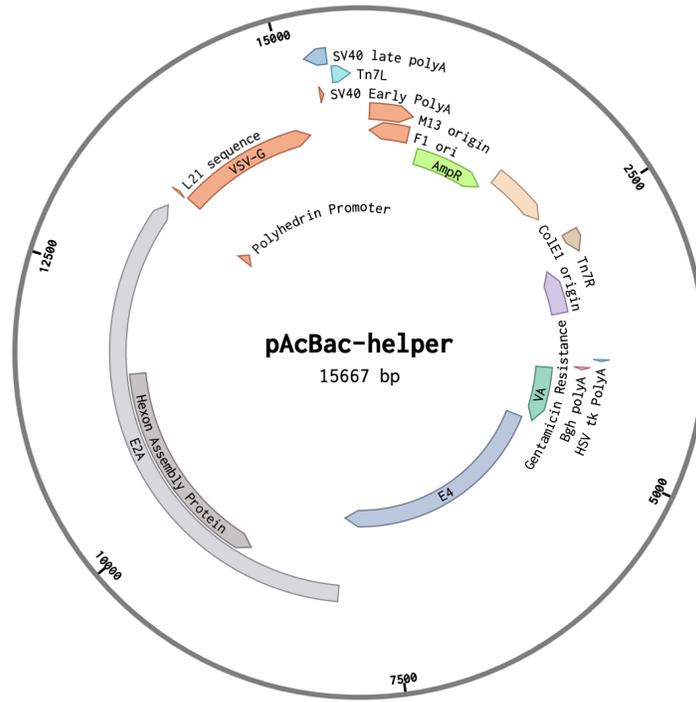
ctggagtactttccttctcagatgctgctaccggaacaactttaccttcagctacac
ttttgaggacgttcctttccacagcagctacgctcacagccagagtctggaccgtctca
tgaatcctctcatcgaccagtacctgtattacttgagcagaacaaactccaagtga
accaccacgcagtcaaggcttcagttttctcaggccggagcagtgacattcgggacca
gtctaggaactggcttcctggacctgttaccgccagcagcagtgatcaaagacatctg
cggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggc
agagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaa
gttttttctcagagcggggttctcatctttgggaagcaaggctcagagaaaaaaaatg
tggacattgaaaaggctcatgattacagacgaagaggaaatcaggacaaccaatcccgtg
gctacggagcagtggttctgtatctaccaacctccagagaggcaacagacaagcagc
taccgcagatgtcaacacacaaggcgttcttccaggcatggctctggcaggacagagatg
tgtacctcaggggcccctctgggcaagattccacacacggacggacattttcacccc
tctcccctcatgggtggattcggacttaaacacctcctccacagatttctcatcaagaa
caccocggtacctgcgaatccttcgaccaccttcagtgccgcaagtttgcttcttca
tcacacagtaactccacgggacaggtcagcgtggagatcgagtgggagctgcagaaggaa
aacagcaaacgctggaatcccgaattcagtaacttccaactacaacaagtctgttaa
tgtggactttactgtggacactaatggcgtgtattcagagcctcgccccattggcacca
gatacctgactcgtaatctgtaattgcttgtaatacaataaacggttaattcgtttca
gttgaactttggtctctgctatcttcttcttatctagtttccatggctacgtagataa
gtagcatggcgggttaatacattaactacagcccggcggttaaacagcgggaggagg
tggagtcgtgacgtgaattacgtcataggggttagggaggtcctgtattagaggtcacgt
gagtgttttgcgacattttgcgacaccatgtggtctcgctggggggggggggcccagtg
agcacgcagggtctccattttgaagcgggaggtttgaacgagcgtggcgcgctcactg
gccgtcgttttacaacgtcgtgactgggaaaacctggcgttaccacaacttaatcgcc
tgcagcacatcccccttctgcagctggcgtaatagcgaagaggcccgcaccgatcgc
cttcccctcatcgccgcaataacctgcaggatccgttttgcgctgcttcgcgatgta
cgggccagatatacgcgttgacattgattattgactagttattaatagtaataattac
ggggtcattagttcatagcccataataggagttccgcgttacataacttacggtaaatg
gcccgcctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgtt
cccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggta
aactgccacttggcagtacatcaagtgtatcatatgccaaagtacgccccctattgacg
tcaatgacggtaaatggcccgcctggcattatgccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcacgctattaccatgggtgatgcgggtttg
gcagtacatcaatggcggtgatagcgggttgactcacggggatttccaagtctccacc
ccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgt
cgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtcta
tataagcagagctctctggctaactagagaaccactgcttactggcttatcgaatta
atacgaactactataaggagaccaagctggctagcgcaccatggataaaaaaccatt
agatgttttaatatctgcgaccgggctctggatgtccaggactggcagcctccacaaaa
tcaagcaccatgaggtctcaagaagtaaaatatacattgaaatggcgtgtggagaccat
cttgttgtgaataattccaggagttgtagaacagccagagcattcagacatcataagta
cagaaaaacctgcaaacgatgtagggttccggacgaggatatcaataattttctcacia
gatcaaccgaaagcaaaaacagtgtgaaagttagggtagtttctgctccaaagggtcaaa
aaagctatgcgaaatcagtttcaagggtccgaagcctctggaaaattctgtttctgc
aaaggcatcaacgaacacatccagatctgtaccttcgcctgcaaaaatcaactccaaat
cgtctgttcccgcacatcggtcctgctccttcaactacaagaagccagcttgatagggtt
gaggctctcttaagtccagaggataaaaatttctctaaatatggcaagcctttcagggg
acttgagcctgaacttgtgacaagaagaaaaaacgattttcagcggctctataccaatg

atagagaagactacctcggtaaactcgaacgtgatattacgaaatTTTTcgtagaccgg
ggTTTTctggagataaagtctcctatccttattccggcggaatacgtggagagaatggg
tattaataatgatactgaactttcaaacagatcttccgggtggataaaaatctctgct
tgaggccaatgcttgccccgactctttacaactatctgcgaaaactcgataggatttta
ccaggccaataaaaatTTTcgaagtcggaccttgttaccggaaagagtctgacggcaa
agagcacctggaagaatttactatggtgaacttctgtcagatgggttcgggatgtactc
gggaaaatcttgaagctctcatcaagagtttctggactatctggaaatcgacttcgaa
atcgtaggagattcctgtatggtctttggggatactcttgatataatgcacggggacct
ggagctttcttcggcagtcgtcgggccagtttctcttgatagagaatgggggtattgaca
aacatggataggtgcaggTTTTggtcttgaacgcttgctcaaggttatgcacggctt
aaaacattaagagggcatcaaggtccgaatcttactataatgggatttcaaccaatct
gtaagaattcaacgcgtaagtcgactttaaactcgagtctagagggcccgTTTTaaacc
gctgatcagcctcgaactgtgcttctagttgccagccatctgttgTTTTgcccctcccc
gtgcttctcttgaccctggaaggtgccactcccactgtcctttcctaataaaaatgagga
aattgcatcgcattgtctgagtaggtgtcattctattctgggggggtgggggtggggcagg
acagcaagggggaggattgggaagacaatagcaggcattgctggggatgcggtgggctct
atggcttctgagggcggaaagaaccctaggggtgcgagcggatcgagcagtgctgatcac
tactggaccgcgagctgtgctgcgaccctgatcttacggcattatacgtatgatcggt
ccagatcagctagattatctagtcagcttgatgtcatagctgtttcctgaggctcaat
actgaccatttaaatcatacctgacctccatagcagaaagtcaaaagcctccgaccgga
ggcttttgacttgatcggcacgtaagaggttccaacttccaccataatgaaataagatc
actaccgggcgtatTTTTtgagttatcgagatTTTcaggagctaaggaagctaaaatga
gcatattcaacgggaaacgtcttgcttgaagccgcgattaaattccaacatggatgct
gatttataatgggtataaaatgggctcgcgataatgtcgggcaatcaggtgcgacaatcta
tcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcg
ttgccaatgatgttacagatgagatggtcaggctaaactggctgacggaatttatgct
cttccgaccatcaagcattttatccgtactcctgatgatgcattggttactcaccactgc
gatcccagggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaata
ttggtgatgcgctggcagtgttcctgcgccgggtgcattcgattcctgTTTTgtaattgt
ccttttaacggcgatcgcgtatTTTcgtctcgcctcaggcgcaatcacgaatgaataacgg
tttggttggtgcgagtgattttgatgacgagcgtaatggctggcctggtgaacaagtct
ggaaagaaatgcataaactcttgccattctcaccggattcagtcgtcactcatggatg
ttctcacttgataacctatTTTTgacgaggggaaattaataggttgattgatggtgg
acgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcgggtg
agttttctccttattacagaaacggctTTTTcaaaaataggtattgataatcctgat
atgaataaattgcagtttacttgatgctcgatgagtttttctaatgaggacctaataatg
taatcacctggctcaccttcgggtgggcctttctgcggttgctggcgTTTTtccataggc
tccgccccctgacgagcatcacaaaaatcgatgctcaagtcagaggtggcgaaaccgg
acaggactataaagataaccaggcgtttccccctggaagctccctcgtgcgctctcctgt
tccgacctgcccgttaccggatacctgtccgcctttctcccttcgggaagcgtggcgc
tttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctg
ggctgtgtgcacgaacccccggttcagcccagccgctgcgccttatccggtaactatcg
tcttgagtccaacccggtaagacacgacttatcgcactggcagcagccactggtaaca
ggattagcagagcaggtatgtaggcgggtgtacagagttcttgaagtgggtggcctaac
tacggctacactagaagaacagtatTTTggtatctgcgctctgctgaagccagttacctc
ggaaaaagagttggtagctcttgatccggcaaaacaaccaccgctggtagcgggtgggtt
TTTTgTTTgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttga
TTTTctaccgaagaaggcca

The sequence for RC2-454TAG is identical except the sequence for *Cap* is:

atggctgccgatggttatcttccagattggctcagggacactctctctgaaggaataag
acagtgggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataag
acgacagcaggggtcttgtgcttccctgggtacaagtacctcggacccttcaacggactc
gacaagggagagccgggtcaacgaggcagacgccgcggccctcgagcacgacaaagccta
cgaccggcagctcgacagcggagacaacccgtacctcaagtacaaccacgccgacgcgg
agtttcaggagcgccttaagaagatacgtcttttgggggcaacctcggacgagcagtc
ttccagggcaaaaagaggggttcttgaacctctgggcctggttgaggaacctgttaagac
ggctccgggaaaaaagaggccggtagagcactctcctgtggagccagactcctcctcgg
gaaccggaaaggcggggccagcagcctgcaagaaaaagattgaattttggtcagactgga
gacgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctgg
tctgggaactaatacgatggctacaggcagtgggcgcaccaatggcagacaataacgagg
gcgccgacggagtgggtaattcctcgggaaattggcattgogattccacatggatgggc
gacagagtcaccaccagcTAGcgaacctggggccctgccacctacaacaaccacct
ctacaacaaatattccagccaatcaggagcctcgaacgacaatcactactttggctaca
gcacccttgggggtattttgacttcaacagattccactgccacttttcaccacgtgac
tggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagct
ctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcattgccaata
accttaccagcagcgggttcagggtgtttactgactcggagtaccagctcccgtacgtcctc
ggctcggcgcacatcaaggatgcctcccgcggttcccagcagacgtcttcatgggtgccaca
gtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttact
gcctggagtactttccttctcagatgctgcgtaccggaaacaactttaccttcagctac
acttttgaggacgttcccttccacagcagctacgctcacagccagagctctggaccgtct
catgaatcctctcatcgaccagctacctgtattacttgagcagaacaaactccaagtg
gaaccaccacgcagtcagggttcagttttctcaggccggagcagtgacattcgggac
cagctctaggaactggcttctcggaccctgttaccgccagcagcagtgatcaaagacatc
tgcgataacaacaacagtgaaactcgtggactggagctaccaagtaccacctcaatg
gcagagactctctgggtgaatccgggcccggccatggcaagccacaaggacgatgaagaa
aagtttttctcagagcgggggttctcatctttgggaagcaaggctcagagaaaacaaa
tgtggacattgaaaaggatcatgattacagacgaagaggaaatcaggacaaccaatcccg
tggctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagca
gctaccgcagatgtcaacacacaaggcgttcttccaggcatggtctggcaggacagaga
tgtgtaccttcaggggcccacctctgggcaagattccacacacggacggacattttcacc
cctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaag
aacaccccggctacctgcgaatccttcgaccaccttcagtgccgcaaagtttgcttccct
catcacacagctactccacgggacaggtcagcgtggagatcgagtgaggagctgcagaagg
aaaacagcaaacgctggaatcccgaatcagctacacttccaactacaacaagtctgtt
aatgtggactttactgtggacactaatggcgtgtattcagagcctcgcgccattggcac
cagatacctgactcgtaatctgtaa

pAcBac-Helper



Sequence color-coding key

Feature	Color
M13 ori	text
AmpR	text
ColE1 ori	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
VA	text
E4	text
E2A	text
Polyhedrin promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
Tn7L	text

Sequence

ttctctgtcacagaatgaaaatttttctgtcatctcttcggtattaatgtttgaattg
actgaatatcaacgcttatttgcagcctgaatggcgaatgggacgcgccctgtagcggc

gcattaagcgcggcggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgc
cctagcgcgccgctcctttcgtttcttcccttcttctcgcacgctcgcggctttc
cccgtcaagctctaaatcggggctccctttagggttccgatttagtgctttacggcac
ctcgaaccccaaaaacttgattagggtgatggttcacgtagtgggccatcgcctgata
gacggtttttcgccctttgacggtggagtcacgcttcttaaatagtggaactcttgttcc
aaactggaacaactcaaccctatctcggctctattcttttgatttataagggattttg
ccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattt
taacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgvcgcgaa
ccctatttgttttatttttctaaatacattcaaataatgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccg
tgctgcccttattcccttttttgcggcattttgccttctgtttttgctcaccagaaa
cgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaa
ctggatctcaacagcggtaagatccttgagagtttcgccccgaagaacgttttccaat
gatgagcacttttaagttctgctatgtggcgcggtattatcccgtattgacgccgggc
aagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactacca
gtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtgccat
aacatgagtgataaactgcggccaacttacttctgacaacgatcggaggaccgaagg
agctaaccgcttttttgacaacatgggggatcatgtaactcgccttgatcgttgggaa
ccggagctgaatgaagccatacacaacgacgagcgtgacaccagatgctgtagcaat
ggcaacaacgttgcgcaaacattaactggogaactacttactctagcttcccggcaac
aattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggccctt
ccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcgggat
cattgcagcactggggccagatggtaagccctcccgtatcgtagtattctacacgacgg
ggagtcaggcaactatggatgaacgaaaatagacagatcgtgagataggtgcctcactg
attaagcattggtaactgtcagaccaagtttactcatatatacttttagattgatttaa
acttcatttttaatttaaaaggatctaggtgaagatcctttttgataatctcatgacca
aaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaa
ggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaacaaaaaaacc
accgctaccagcgggtggtttgtttgccggatcaagagctaccaactctttttccgaagg
taactggcttcagcagagcgcagatacacaataactgtccttctagtgtagccgtagtta
ggccaccacttcaagaactctgtagcaccgcctacatacctcgcctctgctaactcctgtt
accagtggctgctgccagtgggcgataagtcgtgtcttaccgggttgactcaagacgat
agttaccggataagggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaaagcgc
cacgcttcccgaagggagaaaaggcggacaggtatccggtaagcggcagggctcggaaacag
gagagcgcacgagggagcttccaggggaaacgcctggtatctttatagtcctgtcggg
tttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggaggagcct
atggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttg
ctcacatgttctttcctgcgttatccctgattctgtggataaccgtattaccgccttt
gagtgagctgataaccgctcggcgcagccgaacgaccgagcgcagcagtcagtgagcga
ggaagcggaaagagcgcctgatgcggtattttctccttacgcctctgtgcggtatttcac
accgcagaccagccgcgtaacctggcaaaatcggttacggttgagtaataaatggatgc
cctgcgtaagcgggtgtggggcggacaataaagtcttaaaactgaacaaaatagatctaaa
ctatgacaataaagtcttaaaactagacagaatagttgtaaaactgaaatcagtcagtta
tgctgtgaaaaagcatactggactttttgttatggctaaagcaaaccttcattttctga
agtgcaaatgcccgtcgtattaagagggggcgtggccaagggcatggtaaaagactata
ttcgcggcgttgtagacaatttaccgaacaactccgcggccgggaagccgatctcggcct
gaacgaattgttaggtggcggctacttgggtcgcgatatcaaaagtgcactcacttctcccg

atgccccactttgtatagagagccactgcgggatcgtcaccgtaaatctgcttgcaagta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtg
caatgccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcacta
cgcggtgctcaaacctgggcagAACGTAAGCCGCGAGAGCGCCAACAACCGCTTCTTG
gtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccaggtaaatcgg
agtccggtgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatca
agagcagcccgcgatggatgtgacttggtcagggccgagcctacatgtgccaatgatgcc
catacttgagccacctaactttgttttagggcgactgccctgctgcgtaaacatcgttgc
tgctgcgtaaacatcgttgctgctccataaacatcaaacatcgaccacggcgtaacgcgc
ttgctgcttgatgcccgaggcatagactgtacaaaaaacagtcataacaagccatga
aaaccgccactgcccgttaccaccgctgcgttcggtaaggttctggaccagttgcgt
gagcgcatacgtacttgcattacagttacgaaccgaacaggcttatgtcaactgggt
tcgtgccttcatccgtttccacggtgtgcgtcaccgcgcaaccttgggcagcagcgaag
tcgaggcatttctgtcctggctggcgaacgagcgaaggtttcggctctccacgcacgt
caggcattggcggccttgcgttcttctacggcaaggtgctgtgcacggatctgcctg
gcttcaggagatcggtagacctcggccgctcgcggcgcttgcgggtggtgctgacccgg
atgaagtggttcgcacctcgggttttctggaaggcgagcatcgtttgctcggccaggac
tctagctatagttctagtggttggctacgtaccgtaagtggtatggcagggcttgcgc
ttaatgcgccgctacagggcgctggggataccccctagagcccagctggttctttcc
gcctcagaagccatagagcccaccgcatccccagcatgcccgtaatccgtagatgtacct
ggacatccaggatgatgccggcgggcgggtggaggcgcgcggaaagtcgcggacgcggt
tccagatgttgcgagcggcaaaaagtgctccatggctcgggacgctctggccggtgagg
cgtgcgagtcggtgacgctctagaccgtgcaaaaggagagcctgtaagcgggcaactct
tccgtggtctggtggataaattcgaagggtatcatggcggacgaccggggttcgaacc
ccgatccggccgctccgcgctgatccatgcggttaccgcccgcgtgtcgaaccagggtg
tgcgacgtcagacaacgggggagcgtccttttggcttccctccaggcgcggcggctgc
tgcgctagcttttttggccactggccgcgcgcggcgtaagcggttaggctggaaagcga
aagcattaagtggctcgtccctgtagccggagggttattttccaagggttgagtcgca
ggacccccggttcagagtctcgggcccggcggactgcccgaacgggggtttgcctcccc
gtcatgcaagaccccgcttgcaaattcctccgaaacagggacgagccccctttttgct
tttcccagatgcatccggtgctgcccagatgcgccccctcctcagcagcggcaagag
caagagcagcggcagacatgcagggcaccctccccttctcctaccgcgtcaggaggggc
aacatcctacatcgattctagtgaaatccacagaaactagcagaggttaagcaacttactta
tgtcttttacatggtcctgggaaagtggaaaatacaccactgaaacttttgctaccaac
tcttacaccttctcctacattgccaggaataaagaatcgtgaacctggtgcatggtat
gtttcaacgtggttatttttcaattgcagaaaatttcaagtcatttttattcagtagt
atagccccaccaccacatagcttatattgatcaccgtaccttaatcaaacctcacagaac
cctagtattcaacctgccacctcctcccacacacagagtacacagtcctttctcccc
ggctggccttaaaaagcatcatatcatgggtaacagacatattcttaggtggtatatte
cacacggtttctgctcagcgaacgctcatcagtgatattaataaacctccccgggcag
ctcgtttaagttcatgtcgtgtccagctgctgagccacaggctgctgtccaacttgcg
ggtgctcaacgggcccgaagggaagtccacgcctacatgggggttagagtcataatcg
tgcacaggataggcgggtggtgctgcagcagcgcgcgaataaacctgctgcccgcgccc
ctccgtcctgcaggaataacaacatggcagtggtctcctcagcagatgattcgcaccgccc
gcagcatgagacgccttgcctccgggcacagcagcgcaccctgatctcacttaaatca
gcacagtaactgcagcacagcaccacaatattgttcaaaatcccacagtgcaaggcgt
gtatccaaagctcatggcggggaccacagaaccacgctggccatcataccacaagcgcga
ggtagattaagtggcgaccctcataaacacgctggacataaacattacctcttttggc

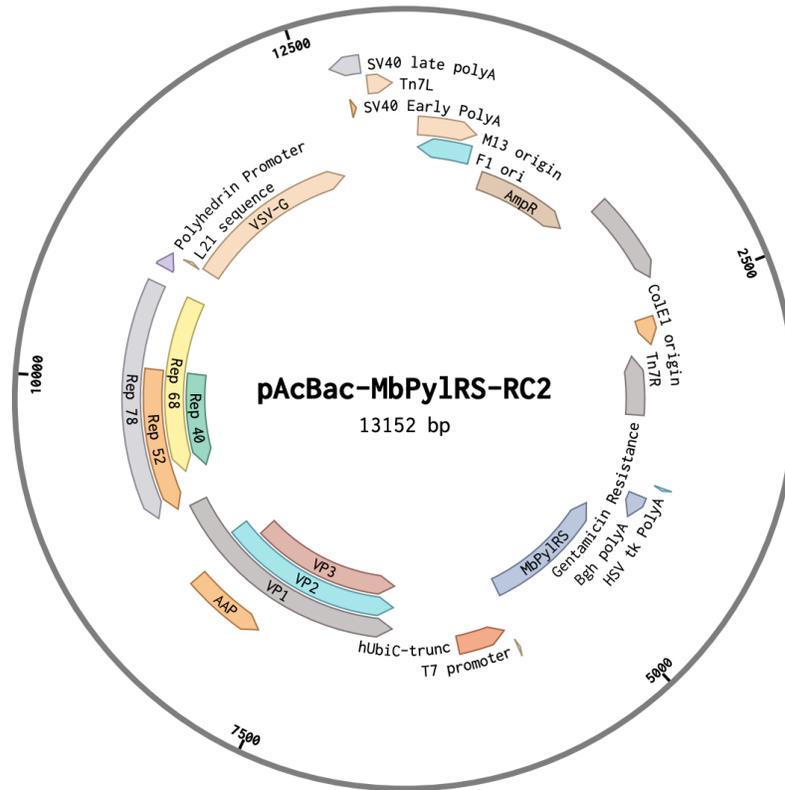
atggttgtaattcaccacctcccggtaccatataaacctctgattaaacatggcgccatc
caccaccatcctaaccagctggccaaaacctgcccgcggctatgcactgcagggaac
cgggactggaacaatgacagtgagagcccaggactcgtaaccatggatcatcatgctc
gtcatgatatcaatggttggcacaacacagggcacacgctgcatacacttcctcaggattac
aagctcctcccgcgctcagaacctatcccagggaaacaaccttcctgaatcagcgtaa
atcccacactgcagggagacctcgcacgtaactcacgttgtgcattgtcaaagtgtta
cattcggggcagcagcggatgatcctccagtatggttagcgcgggtctctgtctcaaaagg
aggtaggcgatccctactgtacggagtgcccgagacaaccgagatcgtggtggtcgta
gtgtcatgccaaatggaacgcccggacgtagtcatatctcctgaagcaaaaaccaggtgcg
ggcgtgacaaacagatctgcgtctccggctctcgtcgttagctcgtctgtgtagtagt
tgtagtatatccactctctcaaagcatccaggcgcacctggcttcgggttctatgtaa
actccttcatgcgcccgtgccctgataacatccaccaccgcagaataagccacaccag
ccaacctacacattcgttctgcgagtcacacacgggaggagcgggaagagctggaagaa
ccatgtttttttttttttattccaaaagattatccaaaacctcaaaatgaagatctatt
aagtgaacgcgctccctccgggtggcgtggtcaaacctctacagccaaagaacagataat
ggcatttgtaagatggtgcaaatggcttccaaaaggcaaacctgccctcacgtccaagt
ggacgtaaaggctaaacccttcagggtgaatctcctctataaacattccagcaccttca
accatgcccaataattttcatctcgcacccttatcaatatgtctctaagcaaatcccg
aatattaagtccggccattgtaaaaatctgctccagagcgcctccaccttcagcctca
agcagogaatcatgattgcaaaaatcagggtcctcacagacctgtataagattcaaaa
gcggaacattaacaaaaataaccgcatcccgtagggtcccttcgcagggccagctgaaca
taatcgtgcaggtctgcacggaccagcgcggccacttccccgccaggaacctgacaaa
agaaccacactgattatgacacgcatactcggagctatgctaaccagcgtagccccta
tgtaagcttggtgcatgggcggcgatataaaatgcaaggtgctgctcaaaaaatcaggg
aaagcctcgcgcaaaaaagcaagcacatcgtagtcatgctcatgcagataaaggcaggt
aagttccggaaccaccacagaaaaagacaccatttttctctcaaacatgtctgcgggtt
cctgcattaacacaaaaataaaaaatacaaaaaaaaaaacatttaaacattagaagcctg
tcttacaacaggaaaaaacaaccttataagcataagacggactacggccatgccggcgt
gaccgtaaaaaaactggtcaccgtgattaaaaagcaccacogacagttcctcggtcatg
tccggagtcataatgtaagactcggtaaacacatcaggttggttaacatcggtcagtg
taaaaagcgaccgaaatagcccgggggaatacataccgcagggcgtagagacaacatta
cagccccataggaggtataacaaaattaataggagagaaaaacacataaacacctgaa
aaacctcctgcctaggcaaaatagcacctcccgcctccagaacaacatacagcgttc
cacagoggcagccataacagtcagccttaccagtaaaaaaacctattaaaaaacaccac
tcgacacggcaccagctcaatcagtcacagtgtaaaaaggcccaagtacagagcagta
tatataggactaaaaaatgacgtaacggttaaagtccacaaaaaacaccagaaaaccg
cacgcgaacctacgcccagaaacgaaagccaaaaaacccacaacttcctcaaatcttca
cttccgttttcccacgatacgtcacttcccatttttaaaaaaactacaattcccaataca
tgcaagttactccgcctaaaacgcccgggogaccgcacctgtgacgaaagccgcccg
caagctgcgcccctgagttagtcactcgaacttcggcctgggcgtctctggaagtacc
acagtggtgggagcgggactttcctggtagaccagggcagcgggccaactacggggatt
aaggttattacgaggtgtggtggaatagccgcctgttcgaggagaattcgggttcgggt
gggcgoggattccggtgaccgggatcatgtgggggtcccgcgctcatgtagtttatt
cgggttgagtagtcttgggcagctccagccgcaagtcccatttgtggctggtaactcca
catgtaggcgtgggaatttccttgcataatggcgtgacgacaggtgctggcgccg
gggtgtggccgctggagatgacgtagttttcgcgcttaaatgagaaagggcgcgaaac
tagtccttaagagtcagcgcgcagtatgctgaagagagcctccgcgtctccagcgt
gcgccgaagctgatcttcgcttttgtgatacaggcagctgcgggtgagggagcgcagag

acctgttttttattttcagctcttgttcttggccctgctttgttgaaatatagcatac
agagtgggaaaaatcctattttctaagctcgcgggtcgatacgggttcgttgggcccag
acgcagcgcctcctcctcctgctgctgccgcgctgtggatttcttgggctttgtcagag
tcttgcctatccggtcgcctttgcttctgtgtgaccgctgctggtgctgccgctgccgct
gccgccggtgcagtagggctgtagagatgacggtagtaatgcaggatgttacggggga
aggccacgccgtgatggtagagaagaaagcggcggggaaggagatgttgcacccacag
tcttgaagcaagcaactatggcgcttcttgtgccgcgccacgagcggtagccttggcg
ctggttgttgccttgggctaacggcggcggctgcttagacttaccggccctggttccag
tggtgtcccatctacgggttgggtcggcgaacaggcagtgccggcggcgcctgaggagcg
gaggttgtagcgtgctgggaacgggttccaatttctggggcggcggcgaggggaatgc
gaccgaggtgacgggtgttctgtctgacacctcttcggcctcggaagcttctgtctagge
tgtcccagctctccatcatctcctcctcctcctcctcctcctcctcctcctcctcctc
cagtattcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc
catcctgggaagcaagggcccgcggctgctgatagggtcggcggcggggggattggg
ttgagctcctcgcggactgggggtccaggtaaaccctcctcctcctcctcctcctcctc
ctccttggcgggctttgttgatggcttgaattggccaaggatgtggccctgggtaatga
cgcaggcggtaagctccgcatttggcggcgggattggtcttcgtagaacctaatctcg
tggcggtggttagtccctcaggtacaaatttgcgaaggtaagccgacgtccacagccccg
agtgagtttcaaccocggagccgcggacttttctcagggcagggaccctgcagctcaa
aggtaccgataatttgcactttcgttaagcagttgcgaattgcagaccaggagcgggtgc
ggggtgcataggttgcagcgcagtgacactccagtaggcgctcaccgctcacgtcttc
catgatgtcggagtggttaggcaaggtagttggctagctgcagaaggtagcagtgaccc
aaagcggcggagggcattcacggtacttaatgggcacaaagtcgctaggaagcgcacag
caggtggcgggcagaattcctgaacgctctaggataaagttcctaaagttttgcaacat
gctttgactggtgaagtctggcagaccctgttgcagggttttaagcaggcgttcgggga
agataatgtccgccaggtgcgcggccacggagcgtcgttgaaggccgtccataggtcc
ttcaagttttgcttttagcagcttctgcagctcctttagggttgcgctcctccaggcattg
ctgccacacgcccatggccgcttggcagggtgtagcacagaaataagtaaaccgagtcgc
ggacgtagtcgcggcgcgctcgccttgagcgtggaatgaagcacgttttgcggcagg
cggttttcgtgcaaaattccaaggtaggagaccagggtgcagagctccacggttggaaat
tttgcaggcctggcgcacgtagccctggcgaagggtgtagtgcaacgtttcctctagct
tgcgctgcatctccgggtcagcaaaagaccgctgcatgactcaagctccacggtaaca
agcactgcccacattagcttgcgtcgtcctccaagtcggcaggctcgcgcgctctc
aagccagcgcgcagctgctcatcgcactgcgggtaggccctcctcgggtttgttctt
gcaagtttgcatccctctccagggtcgtgcacggcgcacgatcagctcgtcatgact
gtgctcataaccttgggggtaggttaagtgcgggtaggcaagtggtgacctcgat
gctgcgtttcagcacggctaggcgcgcgcttgcaccctcaagttccaccagcactccac
agtgactttcattttcgtggttttcttgttgcagagcgtttgcccgcgcttctcgtcg
cgtccaagaccctcaaagatttttggcacttcgctcagcgcagggcagatcaggtatgac
agcgcctcgcgcaaggccagctgcttgcgcgctcggctgcgggttggcacggcaggata
ggggtatcttgcagttttggaaaaagatgtgataggtggcaagcacctctggcacggca
aatacggggtagaagttgaggcgcgggttgggctcgcagtgccggtttcttggcggtt
ggggggtacgcgcgggtgagaacagggtggcgttctgtaggcaaggctgacatccgctatgg
cgaggggcacatcgtcgcctcttgaacgcgtcgcagataatggcgcactggcgcgtgc
agatgcttcaacagcacgtcgtctccacatctaggtagtcgcatgcttgggtccc
ccgcccgacttgttctcgtttgcctctgcgtcgtcctggcttctgctttttatcctctg
ttggtagtgagcgtcctcgtcgtcttctcgttataaaacctgggtcctgctcgataatc
actcctcctcctcaagcgggggtgcctcgcaggggaagggtgtaggcgcgcttggcggc

atcgggtggaggcgggtgggtggcgaactcaaagggggcgggttaggctgtcctccttctcga
ctgactccatgatcttttctgcctataggagaaggaaatggccagtcgggaagaggag
cagcgcgaaccacccccgagcgcggacgcgggtgcggcgcgacgtccaccaacctatgga
ggacgtgtcgtccccgtcgcggtcgcgcgcctccccgcgcgcccccaaaaagcggc
tgaggcggcgtctcagatccgaggacgaagaagactcgtcacaagatgcgctggtgccg
cgcacaccagcccgcggccatcgacctcgacggcggatttggccattgcgtccaaaaa
gaaaaagaagcgcacctctccaagcccgagcgcgccgcatccccagaggatgatcgtgg
acagcgaggaagaagagaagatgtggcgctacaaatggtgggtttcagcaaccaccg
gtgctaatacagcacggcaaggaggtgaagcgcacgggtgcggcggctgaatgaagacga
cccagtgggcgcggggtatgcggacgcaagaggaaaaggaagagtcagtgaaagcggaaa
gtgaaagcacgggtgataaacccgctgagcctgccgatcgtgtctgcgtgggagaagggc
atggaggctgcgcgcgcttgatggacaagtaccacgtggataacgatctaaaggcaaa
cttcaagctactgcctgaccaagtgaagctctggcggccgtagcaagacctggctaa
acgaggagcaccgcgggttcagctgacctcaccagcaacaagacctttgtgacgatg
atggggcgattcctgcaggcgtacctgcagtcgtttgcagaggtaacctacaagacca
cgagcccacgggctgcgcgcttggtggctgcaccgctgcgctgagatcgaaggcagctta
agtgtctacacgggagcattatgataaataaggagcacgtgatgaaatggatgtgacg
agcgaaaacgggcagcgcgcgctgaaggagcagcttagcaaggccaagatcgtgaagaa
ccgggtggggccgaaatgtggtgcagatctcaaacaccgacgcaagggtgctgcgtgatg
acgcgccctgtccggccaatcagttttccggcaagtcttgccgcatgttcttctctgaa
ggcgcaaaaggctcagggtggcttttaagcagatcaaggctttcatgcaggcgcgtgatcc
taacgccagaccgggcacggtcaccttctgatgccactacggtgcgagtgcaactcaa
agcctgggcatgcacccttttgggaaggcagctaccaagttgactccgttcgcctg
agcaacgcggaggacctggacgcggatctgatctccgacaagagcgtgctggccagcgt
gcaccacccggcgcgtgatagtgtccagtgctgcaacctgtgtatcgcaactcgcgcg
cgcagggcggaggccccaaactgcgacttcaagatatcggcgcccgacctgctaaacgcg
ttggtgatggtgcgcgacctgtggagtgaaaacttcaccgagctgccgcggatggttgt
gctgagtttaagtggagcactaacaccagtatcgcaacgtgtccctgccagtgggcg
atagcgatgcgcggcagaaccttttgatttttaaacggcgcagacggcaagggtgggg
ggtaaataatcaccgagagtgtaaaaataaaaacatttgcctttattgaaagtgtctc
ctagtagattattttacatgttttcaagtgacaaaaagaagtggcgtcctaattctg
cgcactgtggctgcggaagtggggcagtgggcgtccaggaagctgtagagctgttccct
ggttgcgacgcagggtgggctgtacctggggactgttaagcatggagtgggtaccgcg
gccgcgggtccgtatactccggaatattaatagatcatggagataattaaatgataacc
atctcgcaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacct
ataaatattccggattatcataccgtcccaccatcggggcgaactcctaaaaaacg
ccaccatgaagtgccttttgtacttagcctttttattcattgggggtgaattgcaagttc
accatagttttccacacaacccaaaaggaaactggaaaaatgttccttctaattacca
ttattgcccgtcaagctcagatttaaattggcataatgacttaataggcacagccttac
aagtcaaaatgcccagagtcacaaggctattcaagcagacgggttgatgtgtcatgct
tccaaatgggtcactacttgtgatttccgctggatggaccgaagtatataacacattc
catccgatccttactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaac
aaggaaactggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacg
gatgcogaagcagtgattgtccaggtgactcctcaccatgtgctgggttgatgaatacac
aggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgcccc
ctgtccataactctacaacctggcattctgactataagggtcaaagggtatgtgattct
aacctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctggg
aaaggagggcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcct

gcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgag
atggctgataaggatctctttgctgcagccagattccctgaatgccagaagggtaag
tatctctgctccatctcagacctcagtgatgtaagtctaattcaggacgttgagagga
tcttgattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatc
tctccagtgatctcagctatcttgctcctaaaaaccaggaaccggctctgctttcac
cataatcaatggtagccctaaaatactttgagaccagatacatcagagtcgatattgctg
ctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaaggggaactg
tgggatgactgggcacatataagacgtggaaattggaccaatggagttctgaggac
cagttcaggatataagtttctttatacatgattggacatggtatggttgactccgatc
ttcatcttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcg
caacttctgatgatgagagtttattttttgggtgatactgggctatccaaaaatccaat
cgagcttgtagaaggttggttcagtagttggaaaagctctattgcctctttttcttta
tcatagggttaatcattggactattcttggttctccgagttggtatccatctttgcatt
aaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttgg
aaagtgataaggccagggcccaagcttgctcgagaagtactagaggatcataatcagc
cataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaa
cctgaaacataaaatgaatgcaattggtggtgtaacttgtttattgcagcttataatg
gttacaataaagcaatagcatcacaatttcacaataaagcatttttttctactgcat
tctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtagctgatcact
gcttgagcctaggagatccgaaccagataagtgaaatctagttccaaactattttgta
ttttaattttcgtattagcttacgacgctacaccagttcccatctattttgtaactc
ttccctaataatccttaaaaactccatttccaccctcccagttccaactatttgt
ccgcccacagcggggcatttttcttctggtatggttttaataaacatcctgccaact
ccatgtgacaaaccgtcatcttcggtacttt

pAcBac-MbPylRS-RC2



Sequence color-coding key

Feature	Color
M13 ori	text
AmpR	text
ColE1 ori	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
MbPylRS	text
T7 promoter	text
hUbiC-trunc	text
Cap	text
Rep	text
Polyhedrin promoter	text
L21 sequence	text
VSV-G	text
SV40 late PolyA	text
Tn7L	text

Sequence: wild type RC2

ttctctgtcacagaatgaaaatTTTTctgtcatctcttcggttattaatgTTTgtaattg
actgaatatcaacgcttatttgcagcctgaatggcgaatgggacgCGCCctgtagCGGC
gcattaagcgcggcgggtgtggtggttacgCGcagcgtgaccgctacacttGCCagcgc
cctagcgcCCgctcctttcgtttcttcccttctctcgcCCacgTtcGCCggctttc
ccgTcaagctctaaatcgggggctcccttagggTtccgatttagtgctttacggcac
ctcGacCCcaaaaaacttgattagggTgatggTtcacgtagTgggCCatcGCcctgata
gacggTTTTcgcctttgacgTtgagTccacgTtctttaaTtagTggactcttgTtcc
aaactGgaacaacactcaaccctatctcggTctattcttttgatttataagggattttg
ccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatTTaacgcgaattt
taacaaaatattaacgTttacaatttcaggtggcacttttcggggaaatgtgcgcggaa
ccctatttgTTTatttttctaaatacattcaaataTgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagatgagTattcaacatttccg
TgtcgccttattccctTTTTtgCGgcattttgccttctgTTTTgctcaccagaaa
cgctggTgaaagTaaaagatgctgaagatcagTtgggtgcacgagTgggttacatcgaa
ctggatctcaacagcggtaagatccttgagagTtttcgccccgaagaacgTTTTccaat
gatgagcactTTTaaagTtctgctatgtggcgcggTattatcccgTattgacgcCGggc
aagagcaactcggTcgcgcatacactattctcagaatgactTggtTgagTactacca
gtcacagaaaagcatcttacggatggcatgacagTaaagaaTtatgcagTgctgcat
aacatgagTgataacactgcgGCCaacttacttctgacaacgatcggaggaccgaagg
agctaaccgctTTTTtgcaaacatgggggatcatgtaactcgccttgatcgtTgggaa
ccggagctgaatgaagccataccaaacgacgagcgtgacaccgatgctgtagcaat
ggcaacaacgTtgcgcaaacTattaactggcgaactacttactctagcttcccggcaac
aattaatagactggatggaggcggataaagTtgcaggaccacttctgcgctcggccctt
ccggctggctggTttattgctgataaatctggagccggtgagcgtgggtctcgcggTat
cattgcagcactggggccagatggtaagccctcccgtatcgtagTtatctacacgacgg
ggagTcaggcaactatggatgaacgaaatagacagatcgtgagataggtgctcactg
attaagcattggTaaactgtcagaccaagTttactcatatatactttagattgatttaa
acttcatTTTTtaatttaaaggatctaggtgaagatcTTTTtgataatctcatgacca
aaatcccttaacgTgagTtttcgTtccactgagcgtcagacCCcgtagaaaagatcaaa
ggatcttcttgagatcctTTTTctgcgcgTaatctgctgcttgcaaacaaaaaac
accgctaccagcggTggTttgTttgcCGgatcaagagctaccaactctTTTTccgaagg
taactggcttcagcagagcgcagataccaaatactgtccttctagTgtagccgtagTta
ggccaccacttcaagaactctgtagcaccgcctacatacctcgcTctgctaactcctgtt
accagTggctgctgccagTggcgataagTcgtgtcttaccgggttgactcaagacgat
agTtaccggataaggcgcagcggTcgggctgaacggggggttcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgc
cacgcttcccgaagggagaaaggcggacaggtatccgTaaagcggcagggTcggaaacg
gagagcgcacgagggagcttccaggggaaacgcctggtatctttatagTcctgtcggg
tttcGCCacctctgacttgagcgtcgatttttgTgatgctcgtcaggggggCGgagcct
atggaaaaacgCCagcaacCGggcTTTTtacggTtccTggccttttgctggccttttg
ctcacatgttcttctcgttatccctgattctgtggataaccgTattaccgccttt
gagTgagctgataccgctcGCCcagccgaacgaccgagcgcagcagTcagTgagcga
ggaagcggaaagagcgcctgatgcggTatTTTctccttacgcatctgtgcggTatttcac
accgCagaccagccgcgTaaactggcaaaatcggTtacggTtagTaaTaaatggatgc
cctgcgTaaagcggTgtgggCGgacaataaagTctTaaactgaacaaaatagatctaaa
ctatgacaataaagTctTaaactagacagaatagTtgTaaactgaaatcagTccagTta

tgctgtgaaaaagcatactggactttttgttatggctaaagcaaactcttcattttctga
agtgcaaattgcccgtcgtattaaagaggggctggccaagggcatggtaaagactata
ttcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggctt
gaacgaattgttaggtggcgggtacttgggtcgatatcaaagtgcacacttcttcccgt
atgcccactttgtatagagagccactgcgggatcgtcaccgtaaatctgcttgcaagta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtg
caatgcctgcctccgggtgctcgccggagactgcgagatcatagatatagatctcacta
cgcggtgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttg
gtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccaggtaatcgg
agtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatca
agagcagcccgcattgatttgccttggcagggccgagcctacatgtgcaatgatgcc
catacttgagccacctaactttgttttagggcactgccctgctgcgtaacatcgttgc
tgctgcgtaacatcgttgcctccataacatcaaacatcgaccacggcgtaacgcgc
ttgctgcttggatgcccaggcatagactgtacaaaaaacagtcataacaagccatga
aaaccgccactgcgcggttaccaccgctgcgttcggtaaggttctggaccagttgcgt
gagcgcatacgtacttgcattacagttacgaaccgaacaggcttatgtcaactgggt
tcgtgccttcatccgtttccacgggtgtgcgtcaccgcgcaaccttgggcagcagcgaag
tcgaggcatttctgtcctggctggcgaacgagcgaaggtttcggctcaccacgatcgt
caggcattggcggccttgcgttcttctacggcaaggtgctgtgcacggatctgcctg
gcttcaggagatcggtagacctcggccgctcgcggcgttgcgggtgggtgctgaccocgg
atgaagtggttcgcacctcgggttttctggaaggcagcagcgtttgttcgccaggac
tctagctatagttctagtggttggctacgtaccgtagtggctatggcagggcttgc
ttaatgcgccgctacagggcgcgtggggataccccctagagcccagctgggtctttcc
gcctcagaagccatagagcccaccgcatcccagcatgcctgctattgtcttcccacatc
ctcccccttgcgtcctgccccaccaccaccaccagaatagaatgacacctactcagac
aatgcgatgcaatttccctcattttattaggaaaggacagtgggagtgccacctccagg
gtcaaggaaggcagggggaggggcaacaacagatggctggcaactagaaggcacagt
cgaggctgatcagcgggtttaaacgggccctctagactcgagttaaagtcgacttaacg
cgttgaattcttacagattgggttgaatcccattatagtaagattcggaccttgatgcc
ctcttaatgtttttaagccgtgcataaccttgagcaagcgttcaagaccaaaacctgc
acctatccatgggttgtcaatacccattctctatcaagagaaactggcccagcagctg
ccgaagaaagctccagggtcccgtgcattatatcaagagtatcccaagaccatacag
gaatctcctacgatttccgaagtcgatttccagatagtcagaaaactctttgatgagagc
ttcaagatttcccaggtacatcccgaaccatctgacagaagttcccatagtaaatt
cttccagggtgctctttgcgctcagactcttccggtaacaaggctccgacttcgaaaatt
tttattgggcctggtaaaatcctatcgagtttccgagatagttgtaaagagtcggggc
aagcattggcctcaagcagagatttttatccaccgggaagatctgttttgaaagttcag
tatcattattaatacccattctctccacgtattccgcccgaataaggataggagacttt
atctccagaaaaccocggtctacgaaaaatctcgtaatatcacgttcgagtttacggag
gtagctctctctatcattgggtatagagccgctgaaaatcgttttttcttcttgcacaa
gttcaggctcaagttccctgaaaggccttggccatatttagagaaatttatcctctgga
cttaagagagcctcaaccctatcaagctggcttcttctaagtgaaaggagcaggagccga
tgccgggaacagacgaatttggagttgatttgcaggcgaaggtacagatctggatgtgt
tcggtgatgccttgcagaaacagaatttccagaggcttcggagoccttgaaactgat
ttcggcatagctttttgaccttggagcagaaactaccctaactttcacactgttttt
gcttccggttgatcttgtgagaaaattattgatatcctcgtccgaaaccctacatcgtt
tgcagggttttctgtacttatgatgtctgaatgctctggctgttctacaactcctggaa
ttattcacaacaagatggctctccacacgccatttcaatgtatattttacttcttgagac

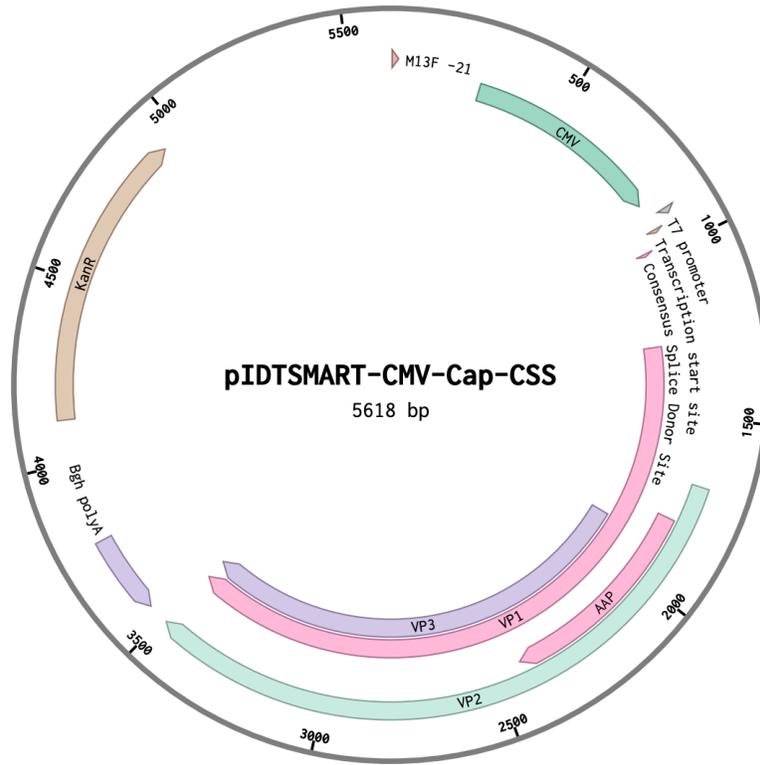
ctcatgggtgcttgatTTTTgtggagcgtgccagtcctggacatccagagcccggtcgcag
atattaaaacatcctaattggTTTTTtatccatgggtggcgttagccagcttgggtctccc
atagtgagtcgtattaatttcgataagccagtaagcagtggggtctcttagttagccaga
gagctctagaccaagtgacgatcacagcgcacacaaacaagaaccgcgacccaatcc
cggctgcgacggaactagctgtgccacaccggcgcgtccttataataatcatcggcgtt
caccgccccacggagatccctccgcagaatcgccgagaagggactacttttctcgcct
gttccgctctctggaagaaaaccagtgccttagagtcaccaagtcccgtcctaaat
gtccttctgctgatactggggttctaaggccgagtccttatgagcagcgggcccgtgtcc
tgagcgtccgggcggaaggatcaggacgctcgtcgcgcccttcgtctgacgtggcagcg
ctcgcctgtgaggagggggggcggccgaggggcggcggaggcccttcg
aacggccactagtcataatcaatgtcaacgcgtatatctggcccgtacatcgcgaagc
agcgcaaaacggatcctgcaggtatTTTgcccgcgatgcatgggaagggcgatcgggtgcg
ggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagtt
gggtaacgccagggtTTTcccagtcacgacgTTTgtaaaacgacggccagtgagcgcgcc
agcgtcgttcaaacctcccgttcaaaatggagaccctgcggtgctcactcgggcccc
ccccccagcgcgagaccacatgggtgctgcaaaatgctgcaaaacactcacgtgacctctaa
tacaggacctccctaaccctatgacgtaattcacgtcacgactccaccctccgcccgc
tgTTTTaaacgcccgggctgtagTTaatgattaaccgcctatgctacttatctacgtagc
catggaaactagataaagaagaatacgcagagaccaaagTTcaactgaaacgaattaa
acggTTTTattgattaacaagcaatttacagattacgagtcaggtatctgggtgccaatggg
gagaggctctgaatacacgccattagtgTccacagtaaaagTccacattaacagactTgt
TgtagTTggaagtgtactgaattcgggattccagcgtTTgctgTTTTccttctgcagc
Tccactcgatctccacgctgacctgtcccgtggagtactgtgtgatgaaggaagcaaa
ctTTgcccactgaaggTggTcgaaggattcgcaggtaccggggTgttcttgatgagaa
TctgtggaggaggTgtTTaagtccgaatccaccatgaggggagaggggTgaaaatgt
ccgtccgtgtgtggaatctTTgccagatgggcccctgaaggTcacacatctctgtcctg
ccagaccatgcctggaagaacgcctTgtgtgtTgacatctgcggtagctgctTgtctgt
TgcctctctggaggTtgTtagatacagaaccatactgctccgtagccacgggattggTt
gtcctgattTcctcttcgtctgtaatcatgacctTTTcaatgtccacattTgtTTTTctc
TgagcctTgctTcccaaagatgagaaccccgctctgaggaaaaaactTTTcttcatcgt
cctTgtggctTgccatggccgggcccggattcaccagagagTctctgccattgaggTg
TactTggtagctccagTccacgagTattcactgtTgtTgttatccgcagatgtctTTga
TactcgtgctggcggTaacaggTccaggaagccagTtccTtagactggTcccgaatgt
cactcgtcTccggcctgagaaaactgaagcctTgactgcgtggTggTtccactTggagTg
TTTgtTctgctcaagTaatacagTactggTcgatgagaggattcatgagacggTccag
actctggctgtgagcgtagctgctgtggaaggaacgTcctcaaaagtgtagctgaagg
TaaagTgtTccggTacgcagcatctgagaaggaaagTactccaggcagTaaaatgaa
gagcgtcctactgctgactcccgtTgtTcagggtgaggtatccatactgtggcaccat
gaagacgtctgctgggaacggcgggaggcatcctTgatgcgccgagccgaggacgtacg
ggagctggTactccgagTcagTaaacacctgaaccgtgctggTaaaggTtattggcaatc
gtcgtcgtaccgtcattctgcgtgacctctTTgactTgaatgtTaaagactTgaagTt
gagTctctTgggtcggaaTcccagTtTgtTgtTgatgagTctTTgccagTcacgtggTg
aaaagTggcagTggaatctgtTgaagTcaaaatcccccaaggggTgctgtagccaaag
tagTgattgtcgttcgaggctcctgattggctggaatTTTgtTgttagaggTggTtgt
gtaggTgggcagggccaggtTcgggtgctggTggTgatgactctgtcgcocatccatg
TggaatcgcaatgccaatTTcccgaggaattaccactccgtcggcgcctcgttattg
TctgccattggTgcgccactgctgtagccatcgtattagTtcccagaccagagggggc
TgctggTggctgtccgagaggctgggggtcaggtactgagTctgcgtctccagTctgac

caaaattcaatctttttcttgcaggctgctggccccgcctttccgggtccccgaggaggag
tctggctccacaggagagtgtctaccggcctctttttcccgagccgtcttaacagg
ttcctcaaccaggcccagagggtcaagaaccctctttttcgctggaagactgctcgtc
cgaggttgccccaaaagacgtatcttctttaaggcgctcctgaaactccgcgtcggcg
tggttgacttgaggtaggggtgtctccgctgctcagagctgccggctcgtaggctttgtc
gtgctcagagggccgcggcgtctgcctcgttgaccggctctcccttgctcagagtcggtga
agggtccgaggtagctgtaccaggaagcacaagaccctgctgtcgtccttatgcgcg
tctgcgggctttggtggtggtgggcccagggtttgagcttccaccactgtcttattccttc
agagagagtgtcctcgagccaatctggaagataaccatcggcagccat**acctgatttaa**
atcatttattggtcaaagatgcagtcacaaatccacattgaccagatcgcaggcagt
gcaagcgtctggcacctttcccatgatatgatgaatgtagcacagtttctgatacgcct
ttttgacgacagaaacgggttgagattctgacacgggaaagcactctaaacagtctttc
tgtccgtgagtgaagcagatatttgaattctgattcattctctcgcattgtctgcaggg
aaacagcatcagattcatgcccagtgacgagaacatttgttttggtagctgtctgcgt
agttgatcgaagcttccgcgtctgacgtcagatggctgcgcaactgactcgcgcaccct
ttgggctcacttatactgcgtcactggggggcgggtctttcttggtccacccttttt
gacgtagaattcatgctccacctcaaccacgtgatcctttgccaccggaaaaagtctt
tgacttctgcttggtgaccttccaaagtcatgatccagacggcgggtgagttcaaat
ttgaacatccggctttgcaacggctgctggtgttcgaaggctcgttgagttcccgtaac
cacggcgcacatggttggtgttgagggtgacgatcacgggagtcgggtctatctgggccc
aggacttgcatcttctggtccacgcgcaccttgcttccctccgagaatggctttggcgcac
tccacgaccttggcggctcatcttcccctcctcccaccagatcaccatcttctgcacaca
gtcgttgaagggaagttctcatgggtccagtttacgcaccctagaagggcacagtggt
gggctatggcctccgcgatggttggtcttcccggtagttgcaggcccaaacagccagatg
gtggtcctcttgccgaactttttcgtggcccatcccagaaagacggaagccgcataattg
gggatcgtacccttttagttccaaaattttataaatccgattgctggaatgtcctcca
cgggctgctggcccaccaggtagtcggggggcggtttttagtcaggctcataatctttccc
gcattgtccaaggcagccttgatttgggaccgcgagttggaggccgcattgaaggagat
gtatgaggcctggtcctcctggatccactgcttctccgaggtaatccccttgctccacga
gccaccgcaccagctccatgtacctggctgaagtttttgatctgatcaccggcgcataca
gaattgggattctgattctctttgttctgctcctgctcgtcgcacacgtgcgtcagatg
ctgcgccaccaaccgtttacgctccgtgagattcaaacaggcgttaaatactgttcca
tattagtccacgccactggagctcaggctgggttttggggagcaagtaattggggatg
tagcactcatccaccaccttggtcccgctccggcgcatttctggtctttgtgaccgc
gaaccagtttgcaaacgtcggctcagatcccgcggtaaattctctgaatcagttttcgc
gaatctgactcaggaaacgtcccaaacatggatttcaccccgggtggtttccacgagc
acgtgcatgtggaagtagctctctcccttctcaaatgcacaaagaaaaggcctccgg
ggccttactcacacggcgcattccgtcagaaagtcgcgctgcagcttctcggccacgg
tcaggggtgctgctcaatcagattcagatccatgtcagaatctggcggcaactccat
tccctctcggccaccagttcacaagctgtcagaaatgcggggcagatgctcgtcaag
gtcgtggtgggaccttaatcacaatctcgtaaaaccccgcat**ggcggctgcgcagatca**
gaagttcctatactttctagagaataggaacttcggaataggaacttctgatcttccgg
gggatccactagttctagagcggccgcgggtccgtatactccggaatattaatagatcat
ggagataattaaatgataaccatctcgcaataaataagtattttactgttttcgtaa
cagttttgtaataaaaaaacctataaatattccggattattcataaccgtcccaccatcg
ggcgcg**aactcctaaaaaacggcacc**atgaagtgccttttgacttagcctttttatt
cattgggggtgaattgcaagtccaccatagttttccacacaaccaaaaaggaaactgga
aaaatggtccttctaattaccattattgcccgtcaagctcagatttaattggcataat

gacttaatagggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaage
agacgggttgatgtgtcatgcttccaaatgggtcactacttgtgatttccgctggtatg
gaccgaagtataaacacattccatccgatccttcaactccatctgtagaacaatgcaag
gaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaag
ttgtggatatgcaactgtgacggatgccgaagcagtgattgtccagggtgactcctcacc
atgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaa
tgcagcaattacatatgccccactgtccataactctacaacctggcattctgactataa
ggtcaaagggctatgtgattctaacctcatttccatggacatcaccttcttctcagagg
acggagagctatcatccctgggaagggagggcacagggttcagaagtaactactttgct
tatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagact
cccatcagggtgtctgggtcgagatggctgataaggatctctttgctgcagccagattcc
ctgaatgcccagaaggggtcaagtatctctgctccatctcagacctcagtggtatgtaagt
ctaattcaggacggtgagaggatcttgattattccctctgccaagaaacctggagcaa
aatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcctaaaaacc
caggaaccggctcctgctttcaccataatcaatggtagccctaaaatactttgagaccaga
tacatcagagtcgatattgctgctccaatcctctcaagaatggtcggaatgatcagtg
aactaccacagaaaggggaactgtgggatgactgggcaccatataagacgtggaaattg
gaccaatggagttctgaggaccagttcaggatataagtttctttatacatgattgga
catgggatgttgactccgatcttcatcttagctcaaaggctcagggtgttogaacatcc
tcacattcaagacgctgcttgcgaacttctgatgatgagagtttattttttggtgata
ctgggctatccaaaaatccaatcgagcttgtagaagggtgggtcagtagttggaaaagc
tctattgctcttttttctttatcatagggttaatcattggactattcttggttctccg
agttggatccatctttgcattaattaagcacaccaagaaaagacagatttatacag
acatagagatgaaccgacttggaaagtgataaggccaggccggccaagcttgtcgagaa
gtactagaggatcataatcagccataccacatttgtagagggtttacttgcctttaaaaa
acctcccacacctccccctgaacctgaaacataaaatgaatgcaattggtggttaac
ttgtttattgcagcttataatggttacaataaagcaatagcatcacaatttcacaaa
taagcatttttttcaactgcattctagttgtggtttgtccaaactcatcaatgtatctt
atcatgtctggatctgatcactgcttgagcctaggagatccgaaccagataagtgaaat
ctagttccaaactattttgtcatttttaattttcgtattagcttaacgacgtacacca
gttcccactatatttgcactcttccctaataatccttaaaaactccatttccacccc
tcccagttcccaactattttgtcggcccacagcggggcatttttcttctggtatggtt
ttaatcaaacatcctgccaactccatgtgacaaaaccgtcatcttccggctacttt

The sequence for RC2-454TAG in this plasmid is identical except the sequence for *Cap* has the same substitution as in pIDT-MbPylRS-RC2-454TAG.

pIDTSMART-CMV-CAP-CSS



Sequence color-coding key

Feature	Color
CMV promoter	text
T7 promoter	text
Transcription start site	text
Consensus splice donor site	text
Cap	text
Bgh polyA	text
KanR	text

Sequence: wild type Cap

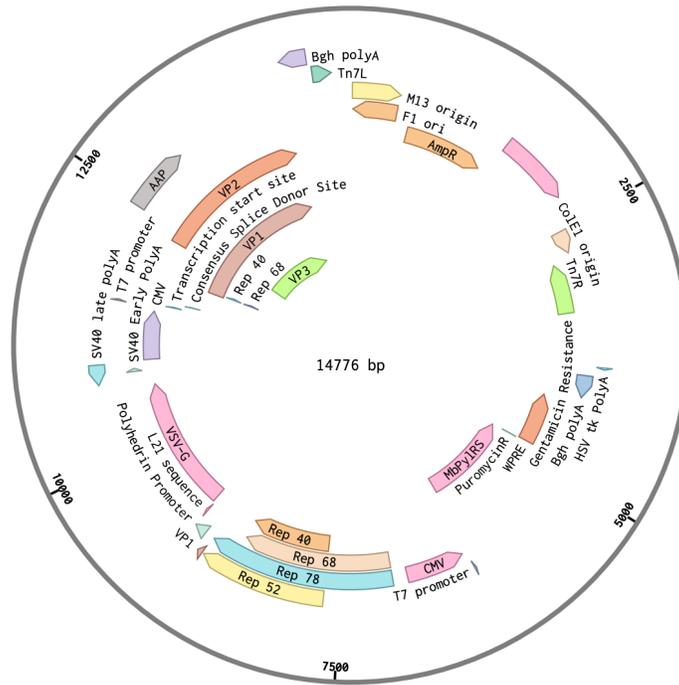
cccgtgtaaaacgacggccagttatctagtcagcttgattctagctgatcgtggaccg
 gaagtgagccagtgagttgattgcagtcagttacgctggagctctgaggctcgtcctg
 aatgatatgcgaccgcccggagggttgcgtttgagacgggcgacagatccagtcgcgctg
 ctctcgtcgateccgctagggcggccgcaaacctgcaggatccgttttgcgctgcttc
 gcgatgtacgggcccagatatagcgttgacattgattattgactagttattaatagtaa
 tcaattacggggtcattagttcatagccatatatggagttccgcgttacataacttac
 ggtaaattggcccgcctggctgaccgccaacgacccccgccattgacgtcaataatga
 cgtatggtcccatagtaacgccaatagggaacttccattgacgtcaatgggtggactat
 ttacggtaaactgccacttggcagtacatcaagtgtatcatatgccaaagtacgcccc

tattgacgtcaatgacggtaaattggcccgcctggcattatgccagtacatgaccttat
gggactttcctacttggcagtacatctacgtattagtcacgcctattaccatggtgatg
cggttttggcagtacatcaatggcggtggatagcggtttgactcacggggatttccaag
tctccacccattgacgtcaatgggagtttgttttggcaccaaatcaacgggactttc
caaatgtcgtacaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtg
gaggtctatataagcagagctctctggcctaactagagaaccactgcttactggcttat
cgaaattaatacagactcactatagggagacccaagctggctagcagttgcgagccatc
gacgtcagacgcggaagcttcgatcaactacgcagacaggttaagtaacaaatgttctc
gtcacgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaaatgaatcag
aattcaaatatctgcttcaactcagggacagaaagactgttttagagtgtttcccggtgc
agaatctcaaccgtttctgtcgtcaaaaaggcgtatcagaaactgtgctacattcatc
atatcatgggaaaggtgccagacgcttgactgcctgcgatctggtcaatgtggatttg
gatgactgcatctttgaacaataaatgatttaaatacaggtatggctgccgatggttatc
ttccagattggctcgaggacactctctctgaaggaataagacagtggtggaagctcaaa
cctggcccaccaccacaaagcccgcagagcggcataaggacgacagcaggggtcttgt
gcttctgggtacaagtacctcggacccttcaacggactcgacaagggagagccgggtca
acgagggcagacgcgcggccctcgcagcagacaagcctacgaccggcagctcgacagc
ggagacaaccgctacctcaagtaaccacgcccagcgcggagtttcaggagcgccttaa
agaagatacgtcttttgggggcaacctcggacgagcagcttccaggcgaaaaagagg
ttcttgaacctctgggctgggtgaggaacctgttaagacggctccgggaaaaagagg
ccggtagagcactctcctgtggagccagactcctcctcgggaaccggaaggcgggcca
gcagcctgcaagaaaaagattgaatthttggctcagactggagacgcagactcagtaactg
accccagcctctcggacagccaccagcagccccctctggtctgggaactaatacgatg
gctacaggcagtggcgcaccaatggcagacaataacgagggcgccgacggagtgggtaa
ttcctcgggaaattggcattgcgattccacatggatggcgacagagtcaccacca
gcaccggaacctgggcccctgccacctacaacaaccacctctacaacaaatthccagc
caatcaggagcctcgaacgacaatcactactttggctacagcacccttgggggtatth
tgacttcaacagattccactgccactthtaccacgtgactggcaaagactcatcaaca
acaactggggattccgacccaagagactcaacttcaagctctttaaacttcaagtcaaa
gaggtcacgcagaatgacgggtacgacgacgattgccaataaccttaccagcaggttca
gggtgttactgactcggagtagcagctcccgtacgtcctcggctcggcgcatcaaggat
gcctcccggcgttcccagcagacgtcttcatggtgccacagtatggatacctcaccctg
aacaacgggagtcaggcagtaggacgctcttcatthtactgctggagtagcttctctc
tcagatgctgctacggaaacaactttaccttcagctacactthtgaggacgttctct
tccacagcagctacgtcacagccagagctctggaccgtctcatgaatcctctcatcgac
cagtaactgtattacttgagcagaacaacactccaagtggaaaccaccacgcagtaag
gcttcagtttctcaggccggagcaggtgacattcgggaccagcttaggaactggcttc
ctggaccctgttaccgccagcagcaggtatcaaagacatctgcggataacaacaacagt
gaatactcgtggactggagctaccaagtaccacctcaatggcagagactctctggtgaa
tccgggcccggccatggcaagccacaaggacgatgaagaaaagthtttctcagagcg
gggttctcatctthgggaagcaaggctcagagaaaaacaatgtggacattgaaaaggtc
atgattacagacgaagaggaaatcaggacaaccaatcccgtggctacggagcagtatgg
ttctgtatctaccaacctccagagaggcaacagacaagcagctaccgcagatgtcaaca
cacaaggcgttcttccaggcatggtctggcaggacagagatgtgtaccttccaggggcc
atctgggcaagattccacacacggacggacattthtaccctctcccctcatgggtgg
attcggacttaaacacctcctccacagattctcatcaagaacaccccggtagctgcga
atccttcgaccaccttcagtgcggaagtttgcttcttcatcacacagtagctccacg
ggacaggtcagcgtggagatcgagtgggagctgcagaaggaaaacagcaaacgctggaa

tcccgaattcagtacacttccaactacaacaagtctgttaatgtggactttactgtgga
aactaatggcgtgtattcagagcctcgccccattggcaccagatacctgactcgtaat
ctgtaattgctgctgactttaactcgagctctagagggcccgtttaaacccgctgatcag
cctcgaactgtgaccttagttgcccagccatctgtttgtttgcccctccccgctgacctcc
ttgacctggaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatc
gcattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagg
gggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatggcttct
gaggcggaaagaaccctaggggtgagcggatcgagcagtgctcgatcactactggacc
gcgagctgtgctgagaccgctgatcttacggcattatacgtatgatcgggccacgatca
gctagattatctagtcagcttgatgtcatagctgtttcctgaggctcaatactgacct
ttaaatcatacctgacctccatagcagaaagtcaaaagcctccgaccggaggcttttga
cttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccggg
cgtattttttgagttatcgagattttcaggagcctaaggaagctaaaatgagccatattc
aacgggaaacgtcttgcttgaagccgcgattaaattccaacatggatgctgatttataat
gggtataaatgggctcgcgataatgtcggggcaatcaggtgogacaatctatcgattgta
tgggaaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatg
atgttacagatgagatggtcaggctaaactggctgacggaatttatgcctcttccgacc
atcaagcattttatccgactcctgatgatgcattgggtactcaccactgcgatcccagg
gaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatg
cgctggcagtgctcctgcccgggttgcattogattcctgtttgtaattgtccttttaac
ggcgatcgcgtatttctgctcgcctcagggcgaatcacgaatgaataacggtttggtgg
tgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaa
tgataaaactcttgccattctcaccggattcagtcgctcactcatgggtgatttctcactt
gataaccttatttttgacgaggggaaattaataggttgattgatggtggacgagtcgg
aatcgcagaccgataaccaggatcttgccatcctatggaactgcctcgggtgagtttctc
cttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaa
ttgcagtttcaacttgatgctcogatgagtttttcaaatgaggacctaaatgtaatcacct
ggctcaccttcgggtgggcctttctgcgcttgcctggcgtttttccataggctccgcccc
ctgacgagcatcacaanaatcgatgctcaagtcagaggtggcgaaaccgacaggacta
taaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgacct
gccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcata
gctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtg
cacgaacccccgttcagcccagccgctgcgccttatccggtaactatcgtcttgagtc
caaccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagca
gagcaggtatgtaggggtgctacagagttcttgaagtgggtggcctaactacggctac
actagaagaacagtatttggtatctgcgctctgctgaagccagttacctcggaaaaaga
gttggtagctcttgatccggcaaaacaaccaccgctggtagcgggtgggttttttggttg
caagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttctacc
gaagaaaggcca

The sequence for RC2-454TAG in this plasmid is identical except the sequence for *Cap* has the same substitution as in pIDT-MbPyIRS-RC2-454TAG.

pAcBac1-Rep-CMV-Cap-CSS-CMV-MbPylRS



Sequence color-coding key

Feature	Color
M13 ori	text
AmpR	text
ColE1 ori	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
CMV promoter	text
Rep	text
Polyhedrin promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
CMV promoter	text
T7 promoter	text
Transcription start site	text
Consensus Splice Donor Site	text
Cap	text

Bgh polyA	text
Tn7L	text

Sequence: wild type Cap

ctgaatggcgaatgggacgcgcctgtagcggcgcattaagcgcggcgggtgtgggtgt
tacgcgcagcgtgaccgctacacttgccagcgcctagcgcgcctcctttcgctttct
tcccttcccttctcgccacgttcgcggctttcccgtcaagctctaaatcgggggctc
ccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttg
agtccacgttctttaatagtggactcttggtccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggatgttgccgatttcggcctattggttaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaaggaagagtatgagtattcaacatttcggtgctgcccttattcccttttttgogg
cattttgccttctggtttttgctcaccagaaacgctggtgaaagtaaagatgctgaa
gatcagttgggtgacgagtggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctat
gtggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgccgcatacac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtataaactgcggcca
actacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgacacaacatg
gggatcatgtaactcgccttgatcgttggaaccggagctgaatgaagccatacacia
cgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaa
ctggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttcggctggctggtttattgctgataa
atctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagttaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggttaactgtcagacca
agtttactcatatataacttttagattgatttaaaacttcatttttaatttaaaggatct
aggtgaagatcctttttgataatctcatgaccaaatacccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttct
gcgcgtaatctgctgcttgcaaaaaaaaccaccgctaccagcgggtggtttgtttgc
cggatcaagagctaccaactccttttccgaaggtaactggcttcagcagagcgcagata
caaataactgtccttctagtgtagcctgtagttaggccaccactcaagaactctgtagc
accgcctacatacctcgtctgtctaactcgttaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctc
ggctgaacggggggttcgtgcacacagcccagcttgagcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggtaagcggcagggctcgaacaggagagcgcacgagggagcttccagg
gaaacgcctggatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg
atthttgtgatgctcgtcaggggggagcctatggaaaaacgccagcaacgcggcct
tttacggttcctggccttttgctggccttttgctcacatggtctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataaccgctcggcgcag
ccgaacgaccgagcgcagcagctcagtgagcaggaagcggaaagagcgcctgatgcgg
atthttccttacgcattctgtgcggtatthcacaccgcagaccagccgcgtaacctggc
aaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgtggggcgaca

ataaagtcttaactgaacaaaatagatctaaactatgacaataaagtcttaactaga
cagaatagttgtaaactgaaatcagtcagttatgctgtgaaaaagcactactggacttt
tggtatggctaaagcaactcttcattttctgaagtgcaaattgcccgctgtattaaag
agggcggtggccaagggcatggtaaagactatattcgcgcggtgtgacaatttaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtactt
gggtcgatatcaaagtgcatacacttcttcccgtatgcccaactttgtatagagagccac
tgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccgaagcgcgttgg
cctcatgcttgaggagattgatgagcgcggtggcaatgccctgcctccgggtgctcgcg
gagactgcgagatcatagatatagatctcactacgcggctgctcaaacctgggcagaac
gtaagccgcgagagcgcacaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatg
tcttactacggagcaagttcccggaggtaatcggagtcggctgatggtgggagtaggtg
gctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcattgatttgacttg
gtcagggccgagcctacatgtgcaatgatgccatacttgagccacctaactttgttt
tagggcgactgccctgctgcgtaacatcgttgcgctgcgtaacatcgttgcgctcca
taacatcaaacatcgaccacggcgtaacgcgcttgcgcttggatgcccgaggcatag
actgtacaaaaaacagtcataacaagccatgaaaaccgccactgcgcggttaccaccg
ctgcgctcgggtcaaggttctggaccagttgcgctgagcgcatacgcctacttgcaattacag
tttacgaaccgaacaggcttatgtcaactgggttcgctgccttcatccggttccacgggtg
tgcgtaaccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcg
aacgagcgaaggttctcggctccacgcatacgtcaggcattggcggccttgcgcttctt
ctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggc
cgtcgcggcgcttgcgggtgggtgctgaccccgatgaagtgggtcgcacctcgggttt
ctggaaggcgagcatcgtttgctcggccaggactctagctatagttctagtggttggct
acgtaccgtagtggcctatggcagggcttgcgcttaatgcgccgctacagggcgcggtg
ggataccccctagagccccagctggttcttcccgcctcagaagccatagagcccaccgc
atccccagcatgcctgctattgtcttcccacatcctcccccttgcgctcctgccccacc
cccccccagaatagaatgacacctactcagacaatgcgatgcaatttctcattttat
taggaaaggacagtgaggagtggcaccttccagggtcaaggaaggcacgggggaggggca
aacaacagatggctggcaactagaaggcacagtcgaggctgatcagcgggtttaaaccg
gccctctagactcagagttaaagtcgacgcggggaggcggcccaaggagatccgactc
gtctgagggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcggga
aggaaggtccgctggattgagggccgaaggacgtagcagaaggacgtcccgcgcagaa
tccaggtggcaacacaggcagcagccaaggaaaggacgatgatttccccgacaacacc
acggaattgtcagtgcccaacagccgagccctgtccagcagcgggcaaggcaggcggc
gatgagttccgccgtggcaatagggagggggaaagcgaagtcgccggaaggagctgac
aggtgggtggcaatgccccaccagtggggggttgcgctcagcaaacacagtgcacaccacg
ccacgttgcctgacaacgggccacaactcctcataaagagacagcaaccaggatttata
caaggaggagaaaatgaaagccatacgggaagcaatagcatgatataaaggcattaaag
cagcgtatccacatagcgtaaaaggagcaacatagtttaagaataccagtcfaatcttca
caaattttgtaatccagaggttgattgtcgcacttaacgcggttgaattcttacagattgg
ttgaaatcccattatagtaagattcggaccttgatgccctcttaattgtttttaagccg
tgcataaccttgagcaagcgttcaagaccaaaccctgcacctatccatggtttgtcaat
acccattctctatcaagagaaactggcccagcactgccgaagaaagctccaggtccc
cgtgcattatatcaagagtatccccaaagaccatacaggaatctcctacgatttccgaag
tcgatttccagatagtcagaaactctttagatgagagcttcaagattttcccaggtaca
tcccgaaccatctgacagaagttcaccatagtaaattcttccagggtgctctttgcccgt
cagactcttccggtaacaaggtccgacttcgaaaatttttatgggcctggtaaaatc
ctatcgagttttcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag

atTTTTatccaccggaagatctgTTTTgaaagttcagtatcattattaataccattc
tctccacgtattccgccggaataaggataggagactttatctccagaaaaccccggtc
acgaaaaatttcgtaatatcacgttcgagttaccgaggtagtcttctctatcattggt
atagagccgctgaaaatcgTTTTtcttctgtcacaagttcaggctcaagttccctga
aaggctttgccatatttagagaaattttatcctctggacttaagagagcctcaacccta
tcaagctggcttcttgtaagtgaaggagcaggagccgatgcggaacagacgaatttg
agttgattttgcaggcgaaggtacagatctggatgtgttcggtgatgcctttgcagaaa
cagaattttccagaggcttcggagcccttgaaactgatttcggcatagctTTTTtgacc
tttgagcagaaaactaccctaactttcacactgTTTTgcttcgggtgatcttgtgag
aaaattattgatatcctcgtccgaaaccctacatcgtttgcaggTTTTtctgtacttat
gatgtctgaatgctctggctgttctacaactcctggaattattcacaacaagatggtct
ccacacgccatttcaatgtatattttacttcttgagacctcatgggtgcttgattttgtg
gagcgtgccagtcctggacatccagagcccggtcgcagatattaaacatctaaggtt
TTTTatccatggtggcgtagccagcttgggtctcccTatagtgagtcgtattaatttc
gataagccagtaagcagtggttctctagttagccagaGagctctgcttatatagacct
cccaccgtacacgectaccgccatttgcgtaaatggggcggagttgttacgacatttt
ggaaagtcccgttgattttggtgccaaaacaaactcccattgacgtcaatggggtggag
acttgaaatcccgtgagtcacaccgctatccacgccattgatgtactgcaaaacc
gcatcaccatggtaatagcgtactaatacgtagatgtactgccaagtaggaaagtcc
cataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgtcaat
agggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttacgta
aatagtccaccattgacgtcaatggaaagtccctattggcgttactatgggaacatac
gtcattattgacgtcaatgggcgggggtcgttggcggtcagccaggcgggccatttac
cgtaagttatgtaacgcggaactccatataatgggctatgaactaatgaccccgtaattg
attactattaataactagtcaataatcaatgtcaacGcgtatatctggcccgtaacatc
cgaagcagcgaacacggatcctgcagggctctagaactagtggatcccccggaagatc
agaagttcctattccgaagttcctatttctctagaaagtataggaacttctgatctgcgc
agccgcCatgccggggttttacgagattgtgattaaggtcccagcgcacttgacgagc
atctgcccgcatTTctgacagctttgtgaactgggtggccgagaaggaatgggagttg
ccgccagattctgacatggatctgaatctgattgagcaggcaccctgacogtggccga
gaagctgcagcgcgacttctgacggaatggcgcgtgtgagtaaggccccggaggccc
TTTTcttTgtgcaatttgagaaggagagagctacttccacatgcacgtgctcgtggaa
accaccggggtgaaatccatggTTTTgggacgttTcctgagtcagattcgcgaaaact
gattcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaa
agaccagaaatggcgcggaggcgggaacaaggtgggtggatgagtgctacatccccaat
tacttgctccccaaaaccagcctgagctccagtgggcgtggactaatatggaacagta
tTaaagcctgTttgaaTctcacggagcgtaaacggTtggTggcgcagcatctgacgc
acgtgtcgcagacgcaggagcagaacaagagaatcagaatccaattctgatgcgccg
gtgatcagatcaaaaacttcagccaggtacatggagctggTcgggtggctcgtggacaa
ggggattacctcggagaagcagtggtatccaggaggaccaggcctcatacatctccttca
atgcccctccaactcgcggTcccaaatcaaggctgcctTggacaatgcccggaaagatt
atgagcctgactaaaaccgccccgactacctggTggccagcagcccgtggaggacat
tTccagcaatcggatttataaaattTtTggaactaaacgggtacgatccccaatatgccc
cttccgtctTtctgggatgggcccagaaaaagTtccggcaagaggaacaccatctggctg
ttTgggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgcc
cttctacgggtgcgTaaactggaccaatgagaactTtccctTcaacgactgtgtcgaca
agatggTgatctggTgggaggagggaagatgaccgccaaggtcgtggagtcggccaaa
gccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagTcctcggcccagat

agaccogactcccgtgatcgtcacctccaacaccaacatgtgcgccgtgattgacggga
actcaacgaccttcgaacaccagcagccgttgcaagaccggatggtcaaatttgaactc
accgcccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagactttt
ccggtgggcaaaggatcacgtggttgagggtgagcatgaattctacgtcaaaaagggtg
gagccaagaaaagaccgcgccccagtgacgcagatataagtgagcccaaagggtgcg
gagtcagttgocgagccatcgacgtcagacgcggaagcttogatcaactacgcagacag
gtaccaaacaatgttctcgtcacgtgggcatgaatctgatgctggttccctgcagac
aatgcgagagaatgaatcagaattcaaatatctgcttactcaccgacagaagactgt
ttagagtgtttcccggtgcagaatctcaaccggtttctgtcgtcaaaaaggcgtatca
gaaactgtgctacattcatcatatcatgggaaagggtccagacgcttgactgcctgcg
atctggatcaatgtggatttggatgactgcatctttgaacaataatgatttaaatcag
tatggctgccgatggttatcttccagattggctcagggacactctctctgagcggccg
ggctccgtatactccggaatattaatagatcatggagataattaaaatgataacctctc
gcaataaataagatattttactgttttcgtaacagttttgtaataaaaaaacctataaa
tattccggattattcataccgtcccaccatcgggcggaactcctaaaaaacgccacc
atgaagtgcctttgtacttagcctttttattcattgggggtgaattgcaagttcaccat
agttttccacacaacccaaaaggaaactggaaaaatgttccttctaattaccattatt
gccgctcaagctcagatttaaattggcataatgacttaataggcacagccttacaagtc
aaaatgccaagagtcacaaggctattcaagcagacgggttgatgtgtcatgcttcaa
atgggtcactacttgtgatttccgctggatggaccgaagtataaacacattccatcc
gatccttactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaacaagga
acttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgc
cgaagcagtgattgtccaggtgactcctcaccatgtgctgggtgatgaatacacaggag
aatgggttgattcacagttcatcaacggaaaatgcagcaattacatagcccactgtc
cataactctacaacctggcattctgactataagggtcaaagggtatgtgattcctaact
catttccatggacatcaccttctctcagaggacggagagctatcatccctgggaaagg
agggcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaa
atgcaatactgcaagcattggggagtcagactcccatcagggtgctggttcgagatggc
tgataaggatctctttgctgcagccagattccctgaatgccagaagggtcaagtatct
ctgctccatctcagacctcagtgatgtaagtcataattcaggacggtgagaggatcttg
gattattccctctgccaaagaaacctggagcaaaaatcagagcgggtcttccaatctctc
agtggatctcagctatcttgctcctaaaaaccaggaaccggctcctgctttcaccataa
tcaatggtagccctaaaatactttgagaccagatacatcagagtcgatattgctgctcca
atcctctcaagaatgggtcggaatgatcagtggaactaccacagaaagggaactgtggga
tgactgggcaccatataagacgtggaaattggacceaatggagttctgaggaccagtt
caggatataagtttcccttatacatgattggacatggtatggtggactccgatcttcat
cttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaact
tctgatgatgagagtttattttttgggtgatactgggctatccaaaaatccaatcgagc
ttgtagaagggtgggttcagtagttggaaaagctctattgcctctttttctttatcata
gggttaatcattggactattcttggttctccgagttggatccatctttgcattaaatt
aaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttgaaagt
gataaggccaggccggccaagcttgtcgagaagtactagaggatcataatcagccatac
cacattttagagaggttttacttggctttaaaaacctccacacctcccctgaacctga
aacataaaatgaatgcaattgttgttggtaacttgtttattgcagottataatggttac
aataaagcaatagcatcaaaattcacaataaagcatttttttactgcattctag
ttgtgggttgtccaaactcatcaatgtatcttatcatgtctgagatctgatcactgcttg
agcctagttattaatagtaaatcaattacggggctattagttcatagcccatataggg
ttccgcgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccg

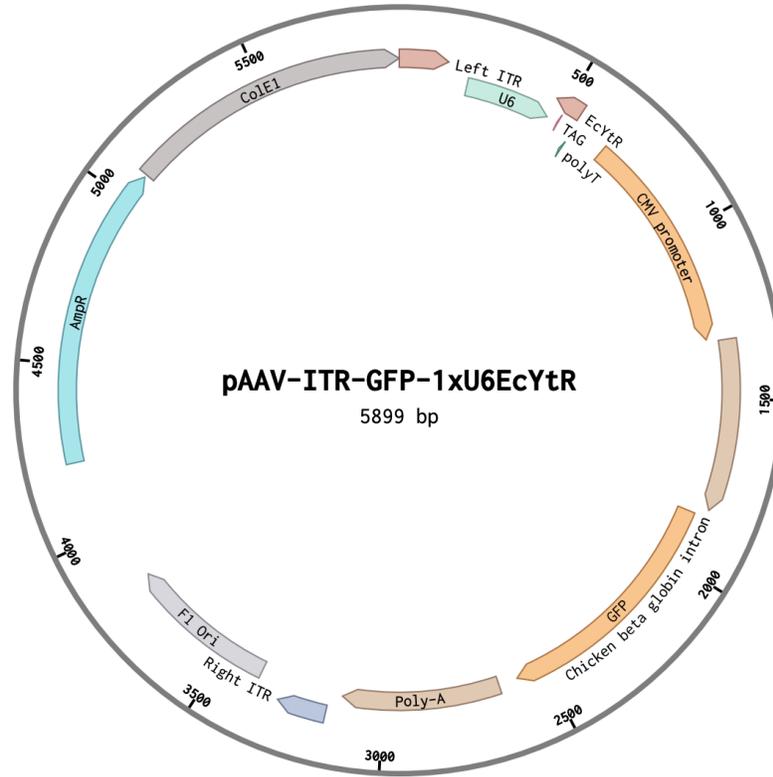
ccattgacgtcaataatgaogtatggtcccataagtaacgccaatagggaactttccatt
gacgtcaatgggtggactatttacggtaaactgccacttggcagtacatcaagtgtat
catatgccaagtacgcccctattgacgtcaatgacggtaaattggccgctggcatta
tgcccagtacatgacctatgggactttcctacttggcagtacatctacgtattagtca
tcgctattaccatgggtgatgoggttttggcagtacatcaatgggcgtggatagcggttt
gactcacggggatttccaagtctccaccattgacgtcaatgggagtttgttttggca
ccaaaatcaacgggactttccaaaatgtcgtacaactccgcccattgacgcaaatgg
cggtaggcgtgtacgggtgggaggtctatataagcagagctctctgggtaactagagaa
cccactgcttactggcttatcgaatlaatacgaactactatagggagaccaagctgg
ctagcagttgocgagccatcgacgtcagacgcggaagcttogatcaactacgcagacag
gtaagttaacaaatggttctcgtcacgtggcatgaatctgatgctgtttccctgcagac
aatgcgagagaatgaatcagaattcaaatatctgcttactcacggacagaaagactgt
ttagagtgtttcccgtgtcagaatctcaaccgtttctgtcgtcaaaaaggcgtatca
gaaactgtgtacattcatcatatcatgggaaaggtgccagacgcttgactgctgctgcg
atctggtcaatgtggatttggatgactgcatctttgaacaataaatgatttaaatcagg
tatggotgocgatggttatcttccagattggctcagaggacactctctctgaaggaaata
gacagtgtggaagctcaaacctggcccaccaccaccaagcccgcagagcggcataag
gacgacagcaggggtcttgtgcttctcgggtacaagtacctcggaccttcaacggact
cgacaaggagagocgggtcaacgaggcagacgcgcggcctcagagcagacaagcct
acgaccggcagctcagacagcggagacaaccogtacctcaagtacaaccacgocgacgg
gagtttcaggagcgccttaagaagatacgtcttttggggcaacctcggacgagcagt
cttccaggcgaanaagagggttcttgaacctctgggctggttgaggaacctgttaaga
cggctccgggaaaaagaggccggtagagcactctcctgtggagccagactcctcctcg
ggaaccggaaaggcggccagcagcctgcaagaaaaagattgaatthtggtcagactgg
agacgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctg
gtctgggaactaatacagatggctacaggcagtgccgcaccaatggcagacaataacgag
ggcgcgcagcggagtgggtaattcctcgggaaattggcattgcgattccacatggatggg
cgacagagtoatcaccaccagcaccgaaacctgggcccctgccacctacaacaaccacc
tctacaaaacaaatthccagocaaataggagcctogaacgacaataactactttggctac
agcacccttgggggtatthtgaacttcaacagattccactgccactthtccaccagtg
ctggcaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagc
tctthaacattcaagtcaagagggtcacgcagaatgacggtagcagcagcagattgccaat
aaccttaccagcagcgggtcagggtgttactgactcggagtaccagctccogtaogtctc
cggctcgggcacacaggatgcctcccgcggttcccagcagacgtcttcatggtgcccac
agtatggatacctcaccctgaacaacgggagtcagggcagtaggacgctcttcattht
tgctggagtactthctctcagatgctgogtaccggaaacaactthaccttcagcta
cactthtggagcgttctthtccacagcagctacgctcacagccagagctctggaccgct
tcatgaatcctctcatcagaccagctcgttacttggagcagacaacaactcacaagt
ggaaccaccacgcagtcgaagcctcagtttctcaggccggagcagtgacattcggga
ccagctotaggaactggcttctcggacctgttaccgcccagcagcagtagtatcaagacat
ctcgggataacaacaacagtgaaactcgtggactggagctaccaagtaccacctcaat
ggcagagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaaga
aaagthtthtctcagagcgggttctcatctthtgggaagcaaggctcagagaaaacaa
atgtggacattgaaaaggctcatgattacagacgaagaggaaatcaggacaaccaatccc
gtggctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagc
agctaccgcagatgtcaacacacaaggcgttcttccaggcatggtctggcaggacagag
atgtgtaccttccaggggcccactctgggcaagattccacacacggacggacattthtcc
ccctctcccctcatgggtggattcggacttaaacacctctcctccacagattctcatcaa

gaacaccccggtacctgogaatccttcgaccaccttcagtgccgcaagtttgcttct
tcacacacagtaactccacgggacaggtcagcgtggagatcgagtgggagctgcagaag
gaaaacagcaaacgctggaatcccgaattcagtaacttccaactacaacaagtctgt
taatgtggaacttaactgtggacactaatggcgtgtattcagagcctcgccccattggca
ccagatacctgaactgtaaatctgtaaattgctgtcgaactttaactcgagcttagagggcc
cgtttaaaccgctgatcagcctcgaactgtgccttctagttgccagccatctgtttgtt
gccccccccgctgccttcttgaccctggaaggtgccactcccactgtcctttcctaa
taaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctgggggggtg
ggggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatg
cgggggctctatggccttctgaggcggaaagaaccctaggagatccgaaccagataagt
gaaatctagttccaaactatgttgcatttttaattttcgtattagcttaocgacgtac
accagttcccattctatgttgcactcttccctaataatccttaaaaactccatttcc
accctcccagttcccaactatgttgcggcccacagcggggcatttttcttctggtta
tgtttttaatacaaacatcctgccaactccatgtgacaaaccgctcatcttcggctacttt
ttctctgtcacagaatgaaaatcttctgtcatctcttcggttattaatgtttgtaattg
actgaatatcaacgcttatttgcagc

The sequence for RC2-454TAG in this plasmid is identical except the sequence for *Cap* has the same substitution as in pIDT-MbPylRS-RC2-454TAG.

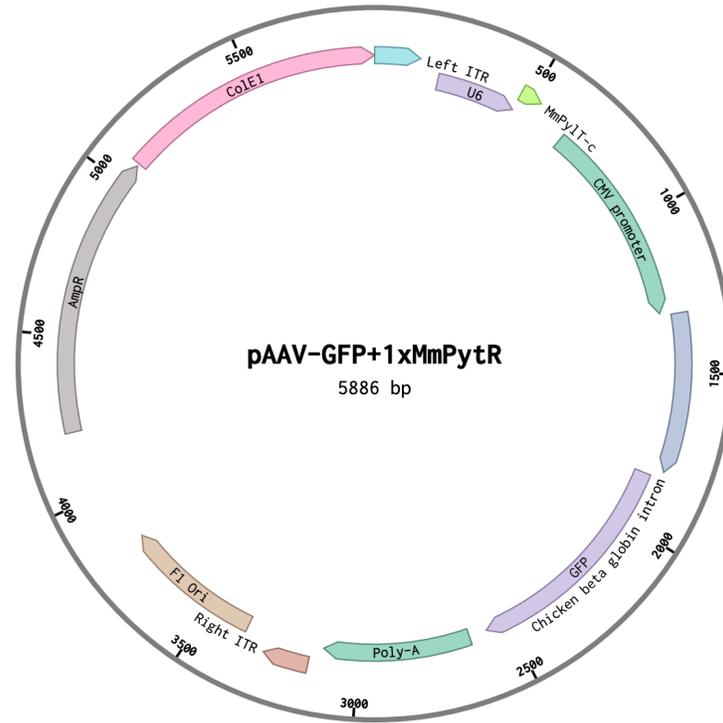
Chapter 3 Plasmids

pAAV-ITR-GFP+1xEcYtR



The sequence for this plasmid is identical to the pAAV-ITR-GFP-1xEcYtR found in the Chapter 2 section of this appendix.

pAAV-GFP+1xMmPytR



Sequence color-coding key

Feature	Color
Left ITR	text
U6 promoter	text
MmPytR	text
polyT	text
CMV promoter	text
Chicken beta globin intron	text
GFP	text
Poly-A	text
Right ITR	text
F1 Ori	text
Amp R	text
ColE1	text

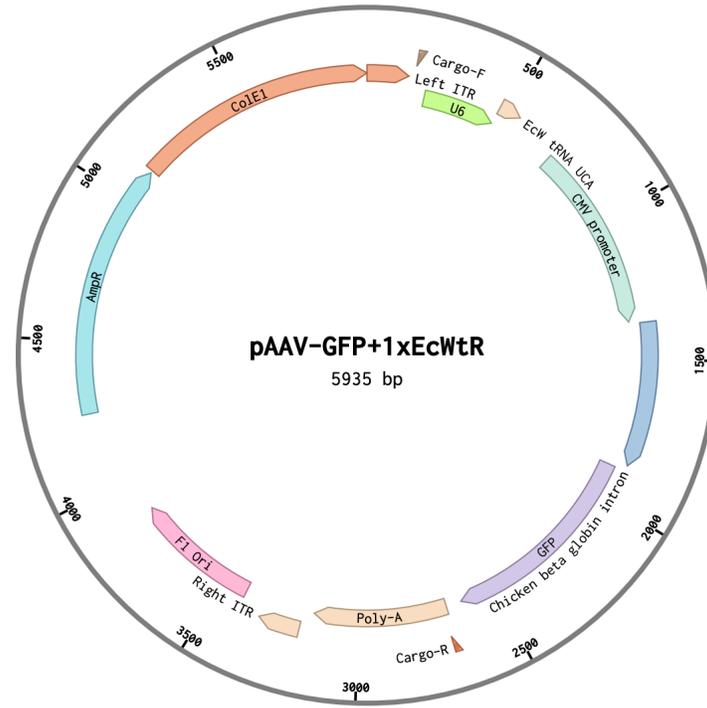
Sequence

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccggggt
 cgggcgacctttggtcgcccgctcagtgagcgagcgagcgcgagagggagtggc
 caactccatcactaggggttcctgcggccgcacgcgtctcgcggtccagtagtgatcga
 cactgctcgatccgctcgcaccctaggtcgggcaggaagagggcctatttcccatgat
 tccttcatatttgcatatacgatacaaggctgtagagagataattagaattaatttga

ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggg
tagtttgagttttaaattatgttttaaattggactatcatatgcttaccgtaacttg
aaagtatttcgatttcttggctttatatatcttgtggaaggacgaaacaccggaacc
tgatcatgtagatcgaacggactcctaaatccggtcagccgggttagattccccggggtt
ccgtttttggctagcggatcgacgagagcagcgcgactggatctgtcgcgccgtctcaaa
cgcaaccctccggcggtcgcatatcattcaggacgagcctcagactccagcgtaacacg
cgtggagctagttattaatagtaatacaattacgggggtcattagttcatagcccatatat
ggagttccgcggttacataacttacggtaaatggcccgcctggctgaccgccaacgacc
cccgccattgacgtcaataatgacgtatgttcccatagtaacgtcaatagggactttc
cattgacgtcaatgggtggagtatttacggtaactgccacttggcagtacatcaagt
gtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggc
attatgccagtacatgaccttatgggactttcctacttggcagtacatctacgtatta
gtcatcgctattaccatggtgatgcgggttttggcagtacatcaatggcgtggatagcg
gtttgactcacggggatttccaagtctccacccttggcagtcattgggagttgtttt
gcacaaaatcaacgggactttccaaaatgtcgtacaactccgcccttggcgtcaaa
tgggcggttaggcgtgtacgggtgggaggtctatataagcagagctcgtttagtgaaacct
cagatcgctggagacgccatccacgctgtttgacctccatagaagacaccgggaccg
atccagcctccgcggttcgaatcccggccgggaacgggtgcattggaacgcggattccc
cgtgccaaagagtgacgtaagtaccgcctatagagtctatagggccacaaaaaatgcttt
cttcttttaataacttttttgtttatcttatttctaatactttccctaattctctttct
ttcagggaataatgatacaatgtatcatgcctctttgcaccattctaagaataaacag
tgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatata
aattgtaactgatgtaagaggttcctatattgctaataagcagctacaatccagctaccat
tctgcttttattttatgggtgggataaggctggattatttctgagtccaagctaggccct
tttgctaatacatgttcatacctcttattcttccacagctcctgggcaacgtgctgg
tctgtgtgctggcccatcactttggcaaagaattgggattcgaacatcgattgaattct
gaatgggtgagcaagggcgaggagctgttcaccgggggtggtgccatcctggtcgagctg
gacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccac
ctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgccgctgccctggc
ccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccac
atgaagcagcagcacttctcaagtccgccatgccgaaggctacgtccaggagcgcac
catcttctcaaggacgacggcaactacaagaccgcgccgaggtgaagttcgagggcg
acaccctggtgaaccgcacgcagctgaagggcacgcacttcaaggaggacggcaacatc
ctggggcacaagctggagtacaactacaacagccacaacgtctatcatggccgacaa
gcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcg
tgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctg
cccgacaaccactacctgagcaccagctccgccctgagcaaaagaccccaacgagaagcg
cgatcacatggtcctgctggagttcgtgaccgcccgggatcactctcggcatggacg
agctgtacaagtactcagatctcgagctcaagtagggatcctctagagtcgacctgcag
aagcttgctcagcagcgtgctcagagatctacgggtggcaccctgtgaccctc
cccagtgctctcctggccctggaagttgccactccagtgcccaccagccttgtcctaa
taaaattaagttgcatcattttgtctgactaggtgtccttctataatattatgggggtg
aggggggtggtatggagcaaggggcaagttgggaagacaacctgtagggcctgcgggt
ctattgggaaccaagctggagtgcagtgccacaatcttggctcactgcaatctccgct
cctgggttcaagcgattctcctgcctcagcctcccagttggtgggattccaggcatgc
atgaccaggctcagctaatttttgttttttggtagagacgggggtttcccatattggc
caggctggtctccaactcctaattctcaggtgatctaccaccttggcctcccaattgc
tgggattacaggcgtgaaccactgctcccttccctgtccttctgattttgtaggttaacc

acgtgocggaccgagcggccgcaggaaccctagtgatggagttggccactccctctctg
cgcgctcgctcgctcactgaggccggcgaccaaggctcgcccgacgcccggccttgc
ccgggocggcctcagtgagcgcgagcgcgcagctgcctgcaggggocgctgatgcggt
atcttctccttacgcatctgtgcggtatctcacaccgcatacgtcaaagcaaccatagt
acgcgocctgtagcggcgcaattaagcgcggcgggtgtggtggttacgcgcagcgtgacc
gctacacttgccagcgcocctagcgcocgctcctttcgctttcttcccttctctcgc
cacgttcgcccggccttccccgtcaagctctaaatcgggggctccctttagggtccgat
ttagtgccttacggcaacctcgacccccaaaaacttgatttgggtgatggttcacgtagt
gggcatcgccctgatagacgggtttttcgccctttgacgttgagtcacgttctttaa
tagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttg
attataagggatcttgcgatttcggcctattggttaaaaaatgagctgatttaacaa
aaatttaacgcgaattttaacaaaatattaacgcttacaatttatggtgcactctcag
tacaatctgctctgatgcccgatagttaagccagccccgacaccgccaacaccgctg
acgcgocctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgctc
tccgggagctgcatgtgtcagaggttttcaccgctcatcaccgaaacgcgcgagacgaaa
gggocctcgtgatacgcctatcttttataggttaatgtcatgataataatggtttcttaga
cgtcaggtggcacttttcggggaaatgtgcgcggaaccctatctgttatttttctaa
atacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataata
ttgaaaaaggaagagtatgagtattcaacatttcggtgctgccttattcccttttttg
cggcattttgccttctgtttttgctcaccagaaacgctgggtgaaagtaaaagatgct
gaagatcagttgggtgcacgagtggttacatcgaactggatctcaacagcggtaagat
ccttgagagttttcgcccgaagaacgctttccaatgatgagcacttttaaagttctgc
tatgtggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgcccata
cactattctcagaatgacttgggtgagtagtaccagtcacagaaaagcatcttacgga
tggcatgacagtaagagaattatgcagtgctgcataaccatgagtgataaacactgcgg
ccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaac
atgggggatcatgtaactcgccctgatcgttgggaaccggagctgaatgaagccatacc
aaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgcttgcgcaactat
taactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcg
gataaagttgcaggaccacttctgcgctcggccttccggctggctggtttattgctga
taaactcggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatg
gtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaa
cgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaaactgtcaga
ccaagtttactcatatatacttttagattgatttaaaacttcatttttaatttaaaagga
tctaggtgaagatcctttttgataatctcatgacaaaatcccttaacgtgagttttcg
ttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttt
tctgcgcgtaatctgctgcttgaacaaaaaaaaccaccgctaccagcgggtggtttggt
tgccgatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcag
ataccaaataactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgt
agcaccgcctacatacctcgtctgctaactcctgttaccagtggtgctgctgccagtggcg
ataagtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcgg
tcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgacctacaccga
actgagatacctacagcgtgagctatgagaaagcgcacgcttcccgaagggagaaagg
cggacaggtatccggtaagcggcagggctcggaacaggagagcgcacgaggggagcttcca
gggggaaacgcctggatctttatagtcctgtcgggttccgccacctctgacttgagcg
tcgatttttgtgatgctcgtcagggggcggagcctatggaaaaacgccagcaacgcgg
cctttttacggttcctggccttttgccttttgccttttgcctacatgt

pAAV-GFP+1xEcWtR



Sequence color-coding key

Feature	Color
Left ITR	text
U6 promoter	text
EcWtR	text
polyT	text
CMV promoter	text
Chicken beta globin intron	text
GFP	text
Poly-A	text
Right ITR	text
F1 Ori	text
Amp R	text
ColE1	text

Sequence

cctgcaggcagctgcgcgctcgctcgctcactgaggccgccgggcaaagcccgggct
 cgggcgaccttggtcgccggcctcagtgagcgagcgagcgagagggagtggc
 caactccatcactaggggttcctgcggccgcacgcgtctcgcggtccagtagtgatcga
 cactgctcgatccgctcgcaccctaggtcggggcaggaagagggcctatttcccatgat
 tccttcatatttgcatatacgatacaaggctgtagagagataattagaattaatttga
 ctgtaaacacaaagatattagtagtaaaaacgtgacgtagaaagtaataatttcttggg

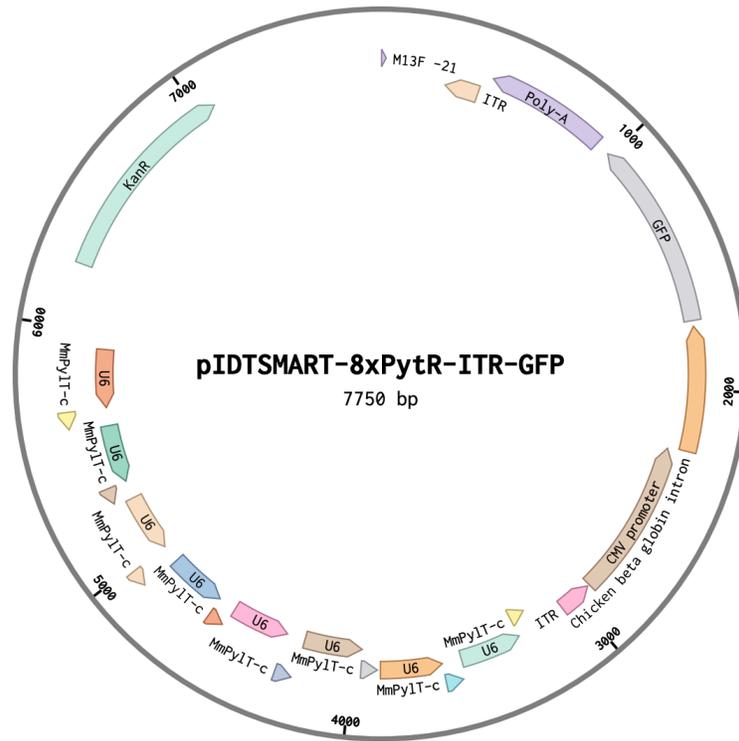
tagtttgagttttaaaattatgttttaaaatggactatcatatgottaccgtaacttg
aaagtatttcgatttcttggctttatatatcttgtggaaggacgaaacaccagggcg
tagttcaattggtagagcaccggtcttcaaaaccgggtggtgggagttcgagtctctcc
gcccctgccattttttgctagcggatcgacgagagcagcgcgactggatctgtcggccg
tctcaaacgcaacctccggcggtcgcatatcattcaggacgagcctcagactccagcg
taacacgcgtggagctagttattaatagtaatacaattacggggtcattagttcatagcc
catatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcc
aacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgtcaatagg
gactttccattgacgtcaatgggtggagatatttacggtaaaactgccacttggcagtac
atcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggccc
gcctggcattatgccagtacatgacctatgggactttcctacttggcagtacatcta
cgtattagtcacgcctattaccatgggtgatgcggttttggcagtacatcaatgggcgtg
gatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagt
ttgttttgcacaaaatcaacgggactttccaaaatgtcgtacaactccgccccattg
acgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctcgtttagt
gaaccgtcagatcgctggagacgccatccacgctgttttgacctccatagaagacacc
gggaccgatccagcctccgcggattcgaatcccggccgggaacgggtgattggaacgcg
gattccccgtgccaagagtacgtaagtaccgcctatagagtctataggccacaaaaa
atgctttcttctttaataactttttgtttatcttatttcttaataactttccctaatac
tctttcttcagggaataatgatacaatgtatcatgcctctttgcaccattctaaaga
ataacagtgataatttctgggttaaggcaatagcaatatttctgcatataaatatttct
gcatataaattgtaactgatgtaagaggttcatattgctaataagcagctacaatccag
ctaccattctgcttttattttatgggtgggataaggctggattattctgagtcgaagct
aggcccttttgctaatacatgttccataacctttatcttctcctcccacagctcctgggcaac
gtgctggctctgtgtgctggccatcactttggcaagaattgggatcgaacatcgatt
gaattctgaatgggtgagcaagggcgaggaggataacatggccatcatcaaggagttcat
gcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcg
agggcgagggccgcccctacgagggcaccagaccgccaagctgaaggtgaccaagggg
ggccccctgccttcgctgggacatcctgtcccctcagttcatgtacggctccaaggg
ctacgtgaagcaccocgcccacatccccgactacttgaagctgtccttccccgagggct
tcaagtgaggagcgcgtgatgaacttcgaggacggcggcgtgggtgaccgtgaccaggac
tctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccacttccc
ctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggg
tgtaccccgaggacggcgcctgaagggcgagatcaagcagaggctgaagctgaaggac
ggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagct
gcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacgaggactaca
ccatcgtggaacagtacgaacgcgcggaggggccgcccactccaccggcggcatggacgag
ctgtacaagggatcctctagagtcgacctgcagaagcttgcctcgagcagcgcgtgctcg
agagatctacgggtggcatccctgtgaccctccccagtgccctctcctggccctggaag
ttgccactccagtgcccaccagccttgtcctaataaaattaagttgcatcattttgtct
gactaggtgtccttctataatattatgggggtggaggggggtggtatggagcaaggggca
agttgggaagacaacctgtagggcctgcggggctctattgggaaccaagctggagtgcag
tggcacaatcctggctcactgcaatctccgcctcctgggttcaagcattctcctgcct
cagcctcccgagttggtgggattccaggcatgcatgaccaggctcagctaattttggtt
tttttggtagagacggggtttcaccatattggccaggctgggtctccaaactcctaactc
aggtgatctaccaccttggcctcccaaattgctgggattacaggcgtgaaccactgct
cccttccctgtccttctgattttgtaggttaaccacgtgcggaccgagcggccgagga
cccctagtgatggagttggccactccctctctgcgcgctcgcctcactgaggccgg

gcgaccaaaggctgcccagcggccggccttggccgggcccctcagtgagcgcgag
cgcgagctgcctgcaggggcgctgatgcggatattttctccttacgcatctgtgcgt
atcacaccgcatacgtcaaagcaaccatagtagcgccctgtagcggcgattaagc
gcgccgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcctagcgc
cgctcctttcgctttcttcccttctttctcgccacgttcgcccggctttcccgcgtaag
ctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgacccc
aaaaacttgattgggtgatggttcacgtagtgggccatcgccctgatagacggttt
tcgccctttgacggttgagtcacggttctttaatagtgactcttggtccaaactggaa
caactcaaccctatctcgggctattcttttgattataagggattttgcccatttcg
gcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat
attaacggtttacaatttatgggtgactctcagtacaatctgctctgatgccgcatagt
taagccagccccgacaccgccaacaccgctgacgcgccctgacgggcttgctgctc
ccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggt
ttcacgctcatcaccgaaacgcgcgagacgaaaggccctcgtgatacgcctatttttat
aggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaat
gtgcgcggaaccctatattgttttttctaaatacattcaaatatgtatccgctcat
gagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtatc
aacatttcgctgctgccttattcccttttttgccgcattttgcttctctgttttgct
caccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggg
ttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaac
gtttccaatgatgagcacttttaagttctgctatgtggcgcggtattatccgctatt
gacgcgggcaagagcaactcgctcgcgcatacactattctcagaatgacttggtga
gtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgca
gtgctgccataaccatgagtataaactgcccgaacttacttctgacaacgatcggg
ggaccgaaggagcctaaccgcttttttgcaaacatgggggatcatgtaactcgccttga
tcggtgggaaccggagctgaatgaagccatacacaacgacgagcgtgacaccagatgc
ctgtagcaatggcaacaacgttgcgcaaacatattaactggcgaacttacttctagct
tcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcg
ctcggcccttcggctggctggtttattgctgataaatctggagccggtgagcgtgggt
ctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatc
tacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgtgagatagg
tgctcactgattaagcattggtaactgctcagaccaagtttactcataataactttaga
ttgatttaaaacttcatttttaatttaaaaggatctaggtgaagatcctttttgataat
ctcatgacaaaaatcccttaacgtgagttttcgcttccactgagcgtcagaccccgtaga
aaagatcaaaggatcttcttgagatcctttttctgcgcgtaatctgctgcttgcaaa
caaaaaaccaccgctaccagcgggtggtttggttgccggatcaagagctaccaactctt
ttccgaaggtaactggcttcagcagagcgcagataccaataactgtccttctagtga
gccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgtctgc
taatcctgttaccagtggtgctgccagtgccgataagtcgtgtcttaccgggttgac
tcaagacgatagttaccggataaggcgcagcggctcgggctgaacggggggttcgtgcac
acagcccagcttgagcgaacgacctacaccgaactgagatacctacagcgtgagctat
gagaaagcggcagcttcccgaaggagaaaggcggacaggtatccggtaagcggcagg
gtcggaacaggagagcgcacgaggagcttccaggggaaacgcctggatctttatag
tctgtcgggtttcgccacctctgacttgagcgtcgatttttgatgctcgtcagggg
ggcggagcctatgaaaaacgccagcaacgcggcctttttacggttctggccttttgc
tggccttttgctcacatgt

The mCherry sequence for the mCherry-containing version of this plasmid is:

```
atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctgga
cggcgacgtaaaccggccacaagttcagcgtgtccggcgagggcgagggcgatgccacct
acggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggccc
accctcgtgaccaccctgacctacggcgtgcagtgttcagccgctaccccgaccacat
gaagcagcacgacttcttcaagtccgcatgcccgaaggctacgtccaggagcgcacca
tcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgac
accctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct
ggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagc
agaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtg
cagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcc
cgacaaccactacctgagcaccagtcgccctgagcaaagaccccaacgagaagcgcg
atcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgag
ctgtacaagtactcagatctcgagctcaagtag
```

pIDTSMART-8xPytR-ITR-GFP



Sequence color-coding key

Feature	Color
ITR	text
Poly-A	text
GFP	text
Chicken beta globin intron	text
CMV promoter	text
ITR	text
MmPylT-c	text
U6	text
KanR	text

Sequence

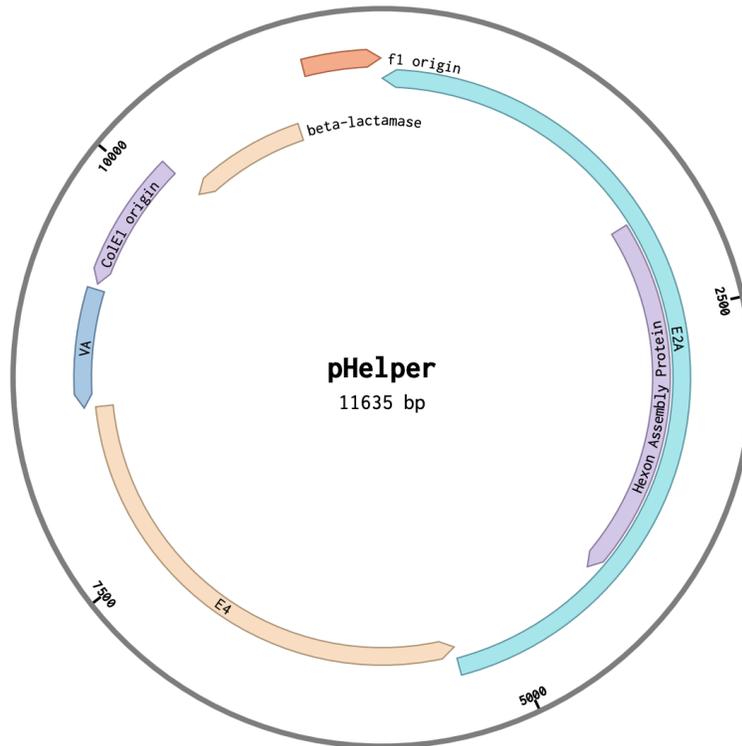
cccgtgtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
 gaaggtgagccagtgagttgattgcagtcacgctggagctctgaggctcgtcctg
 aatgatatgacgaccgcccggagggttgctttgagacgggcgacagatccagtcgcgctg
 ctctcgtcgatccgctaggactatggttgctttgacgtatgcggtgtgaaataaccgcac
 agatgcgtaaggagaaaataaccgcatcaggcgccctgcaggcagctgcgcgctcgcctc
 gctcactgaggccgcccgggcaaagcccgggcgctcgggcgacctttggtcgcccggcct
 cagtgagcgcgagcgcgcagagaggagtgcccaactccatcactaggggttcctgc
 ggccgctcgggtccgcacgtggttacctacaaaatcagaaggacaggggaagggagcagtg

gttacgcctgtaatcccagcaatthgggaggccaaggtgggtagatcacctgagatta
ggagttggagaccagcctggccaatatggtgaaaccccgtctctacaaaaaaacaaaa
attagctgagcctggatcatgcatgcctggaatccaacaactcgggaggctgaggcagg
agaatcgcttgaaccaggaggcggagattgcagtgagccaagattgtgccactgcact
ccagcttggttcccaatagaccccgcaggccctacagggttgtcttcccaacttgccct
tgctccataaccacccccctccaccccataatattatagaaggacacctagtcagacaaa
atgatgcaacttaatthttattaggacaaggctggtgggactggagtggaacttccag
ggcaggagaggcactggggagggggtcacagggatgccaccctagatctctcagagcag
cgctgctcagaggcaagcttctgcaggctgactctagaggatccctacttgagctcgaga
tctgagtaacttgtagcctcgtccatgccgagagtgatccggcgggcggtcacgaactc
cagcaggaccatgtgatcgcgcttctcgttggggcttctgctcagggcggactgggtgc
tcaggtagtgggtgtcgggcagcagcacggggcgtcgcgatgggggtgttctgctgg
tagtggtcggcgagctgcacgctgccgtcctcgatggtgtggcggatcctgaagttcac
cttgatgccgttcttctgcttgtcggccatgatatagacgcttgtggctgtttagtgg
actccagcttgtgccccaggatggtgccgtcctccttgaagtcgatgcccttcagctcg
atgcggttcaccagggtgtcgcctcgaacttcacctcggcgcggttctttagtggc
gtcgtccttgaagaagatggtgcgctcctggacgtagccttcgggcatggcgacttga
agaagtcgtgctgcttcatgtggtcggggtagcggctgaagcactgcacgccgtaggtc
agggtggtcacgagggtgggcccagggcacgggcagcttgcgggtggtgcagatgaact
cagggtcagcttgcgtaggtggcatcgcctcgcctcgcggacacgctgaacttgt
ggcgtttacgtcgcgctccagctcagaccaggatgggcaccaccccgtgaacagctcc
tcgcccttgtcaccattcagaattcaatcgatggtcgaatcccaattctttgccaag
tgatgggcccagcacacagaccagcacgttgcccaggagctgtgggaggaagataagagg
tatgaacatgattagcaaaaggcctagcttggactcagaataatccagccttatccca
accataaaaataaaagcagaatggttagctggattgtagctgctattagcaatatgaaac
tcttacatcagttacaatthtatatgcagaaatatttatatgcagaaatattgctattgc
cttaaccagaaattatcactgttattctttagaatggtgcaaagaggcatgatacatt
gtatcattattgccctgaaagaagagattagggaaagtattagaaataagataaacia
aaaagtatattaaaagaagaagcattthttgtgggctatagactctataggcggtag
ttacgtcactcttggcacggggaatccgcggttccaatgcaccggtcccggccgggattc
gaatccgcggaggctggatcgggtcccgtgtcttctatggagggtcaaaacagcgtggat
ggcgtctccaggcagctgtgacggttcaactaaacgagctctgcttatatagacctcccac
cgtacacgcctaccgcccattthgctcaatggggcggagttgttacgacattthggaaa
gtcccgttgattthggtgcaaaacaaactcccattgacgtcaatgggggtggagactthg
aaatccccgtgagtcaaacgcctatccacgcccattgatgtactgcaaaaaccgcatca
ccatggtaatagcagatgactaatacgtagatgtactgccaagtaggaaagtcccataag
gtcatgtactgggcataatgccaggcgggcccattthaccgtcattgacgtcaataggggg
cgtactthggcatatgatacacttgatgtactgccaagtgggcagttthaccgtaatact
ccaccattgacgtcaatggaaagtcctattgacgttactatgggaacatacgtcatt
attgacgtcaatgggcgggggtcgttgggcgggtcagccaggcgggcccattthaccgtaag
ttatgtaacgcggaactccatataatgggctatgaactaatgaccccgttaattgattact
attaataactagctccacgcgtgcgggccgcaggaacccctagtgatggagttggccact
ccctctctgcgcgctcgtcgtcactgaggccgggcgaccaaaggctgcocgacgccc
gggctthgcccgggcggcctcagtgagcagcagcagcagctgcctgcaggacatgt
gagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgcctggcgtthttc
cataggctccgccccctgacgagcatcacaataatcgacgctcaagtcagaggtggcg
aaaccgagctagcaaaaaacggaacccccgggaatctaaccggctgaacggattthag
agtccgttcgatctacatgatcaggtthccgggtgtthcgtcctthccacaagatatata

aagccaagaaatcgaataactttcaagttacggtaagcatatgatagtcatttttaaaa
cataatttttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttg
tactaataatctttgtgtttacagtcaaattaattctaattatctcttaacagccttgt
atcgtatatgcaaataatgaaggaatcatgggaaataggccctcttctgcccgaacctag
caaaaaacggaaaccccggaatctaaccggctgaacggatttagagtcggttcgatc
tacatgatcaggtttccgggtgtttcgtcctttccacaagatatataaagccaagaaatc
gaaatactttcaagttacggtaagcatatgatagtcatttttaaaacataatttttaaaa
ctgcaaactacccaagaaattattactttctacgtcacgtattttgtactaataatcttt
gtgtttacagtcaaattaattctaattatctcttaacagccttgtatcgtatatgcaa
atatgaaggaatcatgggaaataggccctcttctgcccgaacctagcaaaaaacggaaa
ccccgggaatctaaccggctgaacggatttagagtcggttcgatctacatgatcaggt
ttccgggtgtttcgtcctttccacaagatatataaagccaagaaatcgaataactttcaa
gttacggtaagcatatgatagtcatttttaaaacataatttttaaaactgcaaactacc
aagaaattattactttctacgtcacgtattttgtactaataatctttgtgtttacagtca
aattaattctaattatctcttaacagccttgtatcgtatatgcaaataatgaaggaatc
atgggaaataggccctcttctgcccgaacctagcaaaaaacggaaaccccggaatcta
accggctgaacggatttagagtcggttcgatctacatgatcaggtttccgggtgtttcg
tcctttccacaagatatataaagccaagaaatcgaataactttcaagttacggtaagca
tatgatagtcatttttaaaacataatttttaaaactgcaaactacccaagaaattattac
tttctacgtcacgtattttgtactaataatctttgtgtttacagtcaaattaattcta
tatctcttaacagccttgtatcgtatatgcaaataatgaaggaatcatgggaaataggc
cctcttctgcccgaacctagcaaaaaacggaaaccccggaatctaaccggctgaacg
gatttagagtcggttcgatctacatgatcaggtttccgggtgtttcgtcctttccacaag
atatataaagccaagaaatcgaataactttcaagttacggtaagcatatgatagtcatt
tttaaaacataatttttaaaactgcaaactacccaagaaattattactttctacgtcacg
tattttgtactaataatctttgtgtttacagtcaaattaattctaattatctcttaaca
gccttgtatcgtatatgcaaataatgaaggaatcatgggaaataggccctcttctgccc
gaacctagcaaaaaacggaaaccccggaatctaaccggctgaacggatttagagtcggt
ttcgatctacatgatcaggtttccgggtgtttcgtcctttccacaagatatataaagcca
agaaatcgaataactttcaagttacggtaagcatatgatagtcatttttaaaacataat
tttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgtactaa
tatctttgtgtttacagtcaaattaattctaattatctcttaacagccttgtatcgtat
tatgcaaataatgaaggaatcatgggaaataggccctcttctgcccgaacctagcaaaaa
acggaaaccccggaatctaaccggctgaacggatttagagtcggttcgatctacatg
atcaggtttccgggtgtttcgtcctttccacaagatatataaagccaagaaatcgaata
ctttcaagttacggtaagcatatgatagtcatttttaaaacataatttttaaaactgcaa
actacccaagaaattattactttctacgtcacgtattttgtactaataatctttgtgttt
acagtcaaattaattctaattatctcttaacagccttgtatcgtatatgcaaataatga
aggaatcatgggaaataggccctcttctgcccgaacctagcaaaaaacggaaaccccg
gaatctaaccggctgaacggatttagagtcggttcgatctacatgatcaggtttccgg
gtgtttcgtcctttccacaagatatataaagccaagaaatcgaataactttcaagttacg
gtaagcatatgatagtcatttttaaaacataatttttaaaactgcaaactacccaagaa
ttattactttctacgtcacgtattttgtactaataatctttgtgtttacagtcaaatta
ttctaattatctcttaacagccttgtatcgtatatgcaaataatgaaggaatcatggga
aataggccctcttctgcccgaacctaggggtgagcgagcggatcgagcagtgatcact
actggaccgcgagctgtgctgagaccgtgatcttacggcattatacgtatgatcgggtc
cacgatcagctagattatctagtcagcttgatgcatagctgtttcctgaggctcaata
ctgaccatttaaatcatacctgacctccatagcagaaagtcaaaagcctccgaccggag

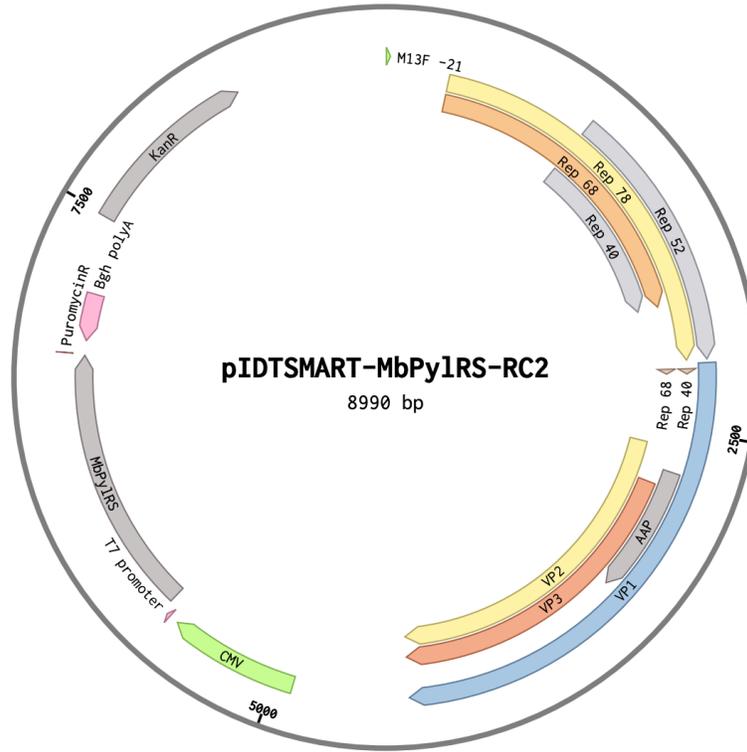
gcttttgacttgatcggcacgtaagaggttccaactttcaccataatgaaataagatca
ctaccgggctgattttttgagttatcgagattttcaggagctaaggaagctaaaatgag
ccatattcaacgggaaacgtcttgcttgaagccgcgattaaattccaacatggatgctg
atztatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcaaatctat
cgattgtatgggaagcccgatgcccagagttgtttctgaaacatggcaaaggtagcgt
tgccaatgatggtacagatgagatgggcaggctaaactggctgacggaatztatgcct
ttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcg
atcccagggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatat
tgttgatgctgctggcagtggtcctgcccgggtgcattcgattcctggttgtaattgct
cttttaacggcgatcgcgtatctcgtctcgtcaggcgcaatcacgaatgaataacggt
ttggttggtgctgagtgattttgatgacgagcgtaatggctggcctggtgaacaagtctg
gaaagaaatgcataaactcttgccattctcaccggattcagtcgtcactcatggtgatt
tctcacttgataaccttatttttgacgaggggaaattaataggttgattgatggtgga
cgagtcggaatcgcagaccgataaccaggatcttgccatcctatggaactgcctcgggtga
gttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgata
tgaataaattgcagtttctcattgatgctcgatgagtttttctaataaggacctaataatgt
aatcacctggctcaccttcgggtgggcttttctgcggtgctggcgtttttccataggct
ccgccccctgacgagcatcacaaaaatcgatgctcaagtcagaggtggcgaaaccgga
caggactataaagataaccaggcgtttccccctggaagctccctcgtgcgctctcctggt
ccgaccctgcccgttaccggatacctgtccgcctttctcccttcgggaagcgtggcgt
ttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgg
gctgtgtgcacgaacccccggtcagcccgaccgctgcgccttatccggtaactatcgt
cttgagtccaaccggtaagacacgacttatcgccactggcagcagccactggtaacag
gattagcagagcagaggtatgtaggcgggtgctacagagttcttgaagtggcctaact
acggctacactagaagaacagtatgggtatctgcgctctgctgaagccagttacctcg
gaaaaagagttggtagctcttgatccggcaaaaccaccgctggtagcgggtggtttt
tttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgat
tttctaccgaagaaggcca

pHelper



The sequence for this plasmid is identical to the pHelper found in the Chapter 2 section of this appendix.

pIDTSMART-MbPyIRS-RC2



The sequence for this plasmid is identical to the pIDTSMART-MbPyIRS-RC2wt and pIDTSMART-MbPyIRS-RC2-454TAG found in the Chapter 2 section of this appendix.

The sequence for the *Cap* gene for pIDTSMART-MbPyIRS-RC2-AAPstop60 is:

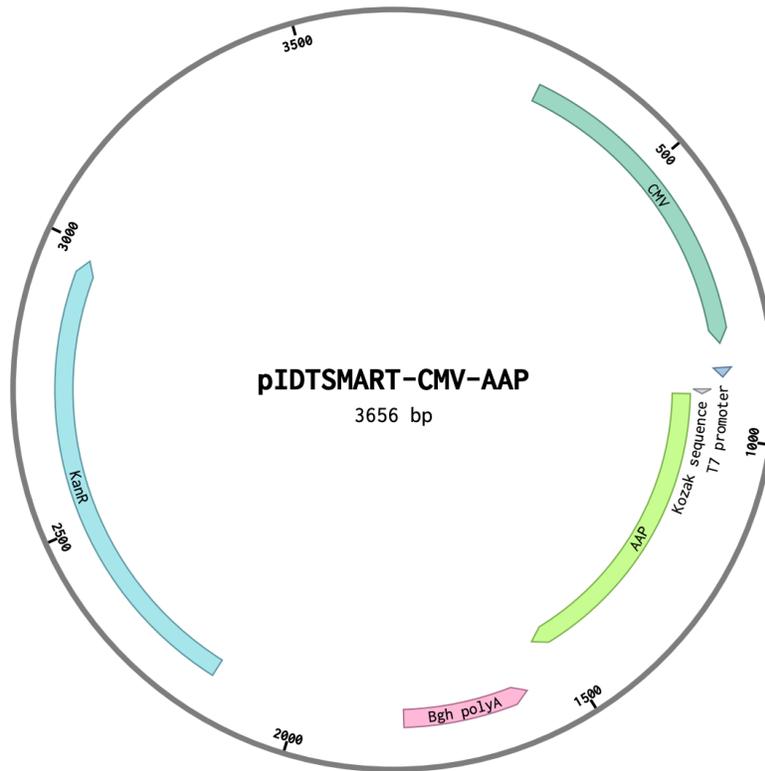
```

atggctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataag
acagtgggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataag
acgacagcaggggtccttgtgcttcttgggtacaagtacctcggacccttcaacggactc
gacaaggagagccgggtcaacgaggcagacgccgcggccctcgagcacgacaaagccta
cgaccggcagctcgacagcggagacaacctgacctcaagtacaaccacgccgacgcgg
agtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtc
ttccaggcgaaaaagagggttcttgaacctctgggcctggttgaggaacctgttaagac
ggctccgggaaaaaagaggccggtagagcactctcctgtggagccagactcctcctcgg
gaaccggaaaggcgggcccagcagcctgcaagaaaaagattgaattttggtcagactgga
gacgcagactcagtagctgacccccagcctctcggacagccaccagcagccccctctgg
tctgggaactaatacagatggctacaggcagtgggcgcaccaatggcagacaataacgagg
gcccgcagcggagtgggtaattcctcgggaaattggcattgcgattccacatggatgggc
gacagagtGatcaccaccagcaccgaacctgggcccctgccacctacaacaaccacct
ctacaacaatttccagccaatcaggagcctcgaacgacaatcactactttggctaca
gcacccttgggggtatthttgacttcaacagattccactgccacttttcaccacgtgac
tggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagct
cttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcagattgccaata
accttaccagcagcggttcagggtgttactgactcggagtaccagctcccgtacgtcctc
ggctcggcgcacatcaaggatgcctcccgcggttcccagcagacgtcttcatgggtgccaca
gtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcatthttact

```

gcctggagtactttccttctcagatgctgcgtaccggaacaactttaccttcagctac
acttttgaggacgttcctttccacagcagctacgctcacagccagagtctggaccgtct
catgaatcctctcatcgaccagctacgtgattacttgagcagaacaactccaagtg
gaaccaccacgcagtcaggcttcagttttctcaggccggagcgagtgacattcgggac
cagcttaggaactggcttctggaccctgttaccgccagcagcgagtatcaaagacatc
tgccgataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatg
gcagagactctctggtgaatccggggcccggccatggcaagccacaaggacgatgaagaa
aagttttttcctcagagcgggggttctcatctttgggaagcaaggctcagagaaaacaaa
tgtggacattgaaaaggatcatgattacagacgaagaggaaatcaggacaaccaatcccg
tggctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagca
gctaccgcagatgtcaacacacaaggcgttcttccaggcatggtctggcaggacagaga
tgtgtaccttcagggggccatctgggcaagattccacacacggacggacattttcacc
cctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaag
aacaccccggtagctgcgaatccttcgaccaccttcagtgccgcaaagtttgcttcctt
catcacacagtagtccacgggacaggtcagcgtggagatcgagtgaggagctgcagaagg
aaaacagcaaacgctggaatcccgaattcagtagacttccaactacaacaagtctgtt
aatgtggactttactgtggacactaatggcgtgtattcagagcctcgccccattggcac
cagatacctgactcgtaatctgtaa

pIDSMART-CMV-AAP



Sequence color-coding key

Feature	Color
CMV	text
T7 promoter	text
Kozak sequence	text
AAP	text
Bgh polyA	text
KanR	text

Sequence

cccgtgtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
 gaaggtgagccagtgagttgattgcagtcacggttacgctggagctctgaggctcgtcctg
 aatgatatgacgaccgccggagggttgcgttgagacggcgacagatccagtcgcgctg
 ctctcgtcgcgctagggcgccgcaaacctgcaggatccggtttgcgctgcttc
 gcgatgtacggccagatatacgcgttgacattgattattgactagttattaatagtat
 agttcatagccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
 gctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagta
 acgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgccca
 cttggcagtacatcaagtgtatcatatgccaagtacgcccctattgacgtcaatgacg
 gtaaattggcccgcctggcattatgccagtacatgaccttatgggactttcctacttgg

cagtaacatctacgtattagtcacgcgtattaccatggatgacgggttttggcagtaacat
caatgggctggatagcgggtttgactcacgggatttccaagtctccaccattgacg
tcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaaacaac
tccgccccattgacgcaaattgggcggtaggcgtgtacgggtgggaggtctatataagcag
agctctctggctaactagagaaccactgcttactggcttatcgaattaaatagcactc
actataggagacccaagctggctagcggcggcaccctggagacgcagactcagtaacct
gacccccagcctctcggacagccaccagcagccccctctgggtctgggaactaatacgat
ggctacaggcagtgggcgcaccaatggcagacaataacgagggcgccgacggagtggtta
attcctcgggaaattggcattgctgattccacatggatgggacgacagagtcacaccacc
agcaccogaacctgggacctgccacctacaacaaccacctctacaacaaatttccag
ccaatcaggagcctcgaacgacaatcactactttggctacagcacccttgggggtatt
ttgacttcaacagattccactgccacttttaccacgtgactggcaagactcatcaac
aacaactggggattccgacccaagagactcaacttcaagctctttaacattcaagtcaa
agaggtcacgcagaatgacggtagcagcagcagattgccaataaccttaccagcaggttc
aggtgttactgactcggagtaccagctcccgtagctcctcggctcggcgcacaaagga
tgctcccgcgctcccagcagacgtcttcatggtgccacagtatggatacctcaccct
gactcagagtcgagtcctagagggcccgtttaaaccgctgatcagcctcagctgtgcctt
ctagttgccagccatctgtttgtttgccctccccctgccttcttgaccctggaaggt
gccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcattgtctgagtag
gtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaag
acaatagcaggcatgctggggatgcgggtgggctctatggcttctgaggcggaaagaacc
ctaggggtgcgagcggatcagcagtgctgatcactactggaccgcgagctgtgctgcg
accgctgatcttacggcattatacgtatgatcggctccacgatcagctagattatctagt
cagcttgatgtcatagctgtttcctgaggctcaatactgaccatttaaatcatacctga
cctccatagcagaaagtcaaaagcctccgaccggaggcttttgacttgatcggcagcta
agaggttccaactttcaccataatgaaataagatcactaccgggcgtatTTTTTgagtt
atcgagattttcaggagcaaggaagctaaaatgagccatattcaacgggaaacgtctt
gcttgaagccgcgattaaattccaacatggatgctgatttataatgggtataaatgggct
cgcgataatgtcgggcaatcagggtgcgacaatctatcgattgtatgggaagccgatgc
gccagagttgtttctgaaacatggcaaggtagcgttgccaatgatgttacagatgaga
tggtcaggctaaactggctgacggaatttatgcctcttccgaccatcaagcattttatc
cgtactcctgatgatgcatggttactcaccactgcgatcccagggaaaacagcattcca
ggattagaagaatatcctgattcagggtgaaaatattgttgatgcgctggcagtggtcc
tgccgcgggtgcattogattcctgtttgtaattgtccttttaacggcgatcgcgattt
cgtctcgtcagggcgaatcacgaatgaataacggtttggttggtgcgagtgattttga
tgacgagcgtaatggctggcctggtgaacaagtctggaaagaaatgcataaactcttg
cattctcaccggattcagtcgtcactcatgggtgatttctcacttgataacctattttt
gacgaggggaaattaataggttgattgatggtggacgagtcggaatcgcagaccgata
ccaggatcttgccatcctatggaactgcctcgggtgagttttctccttattacagaaac
ggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttacttg
atgctcogatgagtttttctaataaggacctaataatgtaatcacctgggctcaccttcgggt
gggcctttctgcggttgcgctggcgtttttccataggctccgccccctgacgagcatcaca
aaaatcgatgctcaagtcagaggtggcgaaccgcagaggactataaagataaccaggcg
tttccccctggaagctccctcgtgcgctctcctgttccgacctgcccgttaccggata
cctgtccgcctttctcccttcgggaagcgtggcgttttctcatagctcacgctgtaggt
atctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgtt
cagcccagccgctgcgccttatccggtaactatcgtcttgagtccaaccgggaagaca
cgacttatcggcactggcagcagccactggtaacaggattagcagagcagaggtatgtag

gcggtgctacagagttcttgaagtgggtggcctaactacggctacactagaagaacagta
tttgggtatctgctgctgctgaagccagttacctcggaagagagttggtagctcttga
tccggcaaacaccaccgctgtagcgggtgtttttgtttgcaagcagcagattac
gcgagaaaaaaggatctcaagaagatcctttgattttctaccgaagaaggccca

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-L12TAG is:

ctggagacgcagactcagtacctgacccccagctagtcggacagccaccagcagcccc
tctgggtctgggaactaatacagatggctacaggcagtgggcgaccaatggcagacaataa
cgagggcgccgacggagtggttaattcctcgggaaattggcattgcgattccacatgga
tggcgacagagtcaccaccagcaccgaacctgggacctgcccactacaacaac
cacctctacaacaatttccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtattttgacttcaacagattccactgccacttttcaccac
gtgactggcaaagactcatcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgc
caataaccttaccagcaggttcaggtgtttactgactcggagtaccagctcccgtacg
tctcggctcggcgcatcaaggatgcctcccgcggttcccagcagacgtcttcatgggtg
ccacagtatggatacctcaccctga

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T78TAG is:

ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctgggtctgggaactaatacagatggctacaggcagtgggcgaccaatggcagacaataa
cgagggcgccgacggagtggttaattcctcgggaaattggcattgcgattccacatgga
tggcgacagagtcaccaccagcaccgaacctgggacctgcccactacaTAGac
cacctctacaacaatttccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtattttgacttcaacagattccactgccacttttcaccac
gtgactggcaaagactcatcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgc
caataaccttaccagcaggttcaggtgtttactgactcggagtaccagctcccgtacg
tctcggctcggcgcatcaaggatgcctcccgcggttcccagcagacgtcttcatgggtg
ccacagtatggatacctcaccctga

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T97TAG is:

ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctgggtctgggaactaatacagatggctacaggcagtgggcgaccaatggcagacaataa
cgagggcgccgacggagtggttaattcctcgggaaattggcattgcgattccacatgga
tggcgacagagtcaccaccagcaccgaacctgggacctgcccactacaacaac
cacctctacaacaatttccagccaatcaggagcctcgaacgacaatcactTAGttgg
ctacagcacccttgggggtattttgacttcaacagattccactgccacttttcaccac
gtgactggcaaagactcatcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgc
caataaccttaccagcaggttcaggtgtttactgactcggagtaccagctcccgtacg
tctcggctcggcgcatcaaggatgcctcccgcggttcccagcagacgtcttcatgggtg
ccacagtatggatacctcaccctga

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T110TAG is:

ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctgggtctgggaactaatacagatggctacaggcagtgggcgaccaatggcagacaataa
cgagggcgccgacggagtggttaattcctcgggaaattggcattgcgattccacatgga

tgggcgacagagtcaccaccagcaccggaacctgggcccctgccacctacaacaac
cacctctacaacaatattccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtatatttgacttcaTAGgattccactgccacttttcaccac
gtgactggcaaagactcatcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcattgc
caataaccttaccagcagcggttcaggtgtttactgactcggagtaccagctcccgtacg
tctcggctcggcgcacatcaaggatgcctcccgcggttcccagcagacgtcttcatggg
ccacagtatggatacctcaccctga

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T124TAG is:

ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctggctctgggaactaatacagatggctacaggcagtgggcgcaccaatggcagacaataa
cgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatgga
tgggcgacagagtcaccaccagcaccggaacctgggcccctgccacctacaacaac
cacctctacaacaatattccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtatatttgacttcaacagattccactgccacttttcaccac
gtgactggcaaagacTAGtcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcattgc
caataaccttaccagcagcggttcaggtgtttactgactcggagtaccagctcccgtacg
tctcggctcggcgcacatcaaggatgcctcccgcggttcccagcagacgtcttcatggg
ccacagtatggatacctcaccctga

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T177TAG is:

ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctggctctgggaactaatacagatggctacaggcagtgggcgcaccaatggcagacaataa
cgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatgga
tgggcgacagagtcaccaccagcaccggaacctgggcccctgccacctacaacaac
cacctctacaacaatattccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtatatttgacttcaacagattccactgccacttttcaccac
gtgactggcaaagactcatcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcattgc
caataaccttaccagcagcggttcaggtgtttactgactcggagtaccagctcccgtTAG
tctcggctcggcgcacatcaaggatgcctcccgcggttcccagcagacgtcttcatggg
ccacagtatggatacctcaccctga

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T97TGA is:

ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctggctctgggaactaatacagatggctacaggcagtgggcgcaccaatggcagacaataa
cgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatgga
tgggcgacagagtcaccaccagcaccggaacctgggcccctgccacctacaacaac
cacctctacaacaatattccagccaatcaggagcctcgaacgacaatcactTGAttgg
ctacagcacccttgggggtatatttgacttcaacagattccactgccacttttcaccac
gtgactggcaaagactcatcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcattgc
caataaccttaccagcagcggttcaggtgtttactgactcggagtaccagctcccgtacg
tctcggctcggcgcacatcaaggatgcctcccgcggttcccagcagacgtcttcatggg
ccacagtatggatacctcaccctga

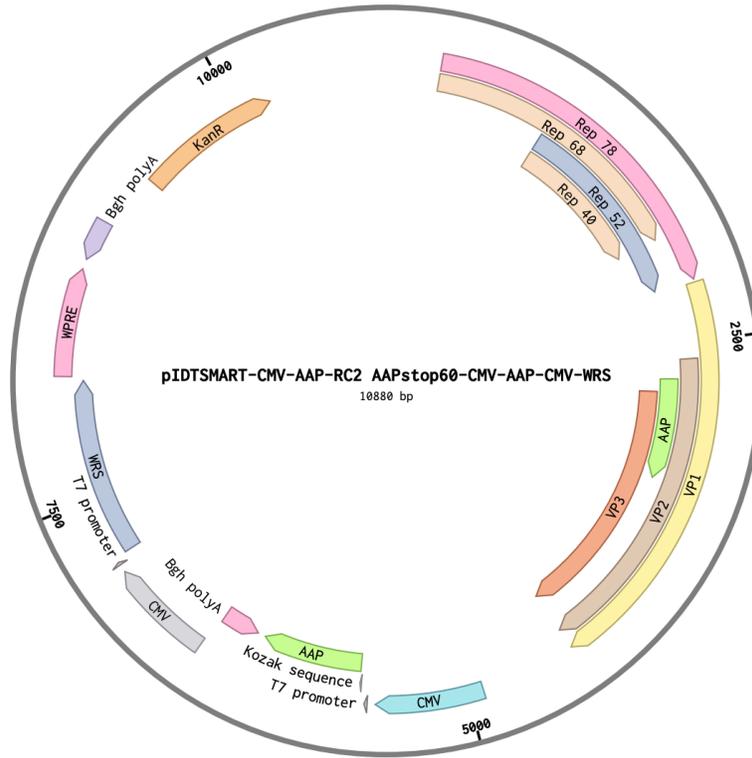
The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T110TGA is:

```
ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctgggtctgggaactaatac gatggctacaggcagtg ggcgcaccaatggcagacaataa
cgagggcgccgacggagtg ggttaattcctcgggaaattggcattg cgattccacatgga
tgggcgacagagtc atcaccaccagcacc cgaacctgggacctgccacctacaacaac
cacctctacaaaca aatttccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtat ttttgacttcaTGA gattccactgccacttttcaccac
gtgactggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttc
aagctctttaacattca agtcaaagaggtcacgcagaatgacgg tacgacgacgattgc
caataaccttaccagc acggttcagggtgtttactgactcggagtaccagctcccgtacg
tcctcggctcggcgcatca aggatgcctcccgcggttcccagcagacgtcttcatggtg
ccacagtatggatacctcaccctga
```

The sequence for the *AAP* gene for pIDTSMART-CMV-AAP-T124TGA is:

```
ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctgggtctgggaactaatac gatggctacaggcagtg ggcgcaccaatggcagacaataa
cgagggcgccgacggagtg ggttaattcctcgggaaattggcattg cgattccacatgga
tgggcgacagagtc atcaccaccagcacc cgaacctgggacctgccacctacaacaac
cacctctacaaaca aatttccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtat ttttgacttcaacagattccactgccacttttcaccac
gtgactggcaaagacTGA tcaacaacaactggggattccgacccaagagactcaacttc
aagctctttaacattca agtcaaagaggtcacgcagaatgacgg tacgacgacgattgc
caataaccttaccagc acggttcagggtgtttactgactcggagtaccagctcccgtacg
tcctcggctcggcgcatca aggatgcctcccgcggttcccagcagacgtcttcatggtg
ccacagtatggatacctcaccctga
```

pIDTSMART-RC2-AAPstop60-CMV-AAP-CMV-WRS



Sequence color-coding key

Feature	Color
Rep	text
Cap – AAPstop60	text
CMV	text
T7 promoter	text
Kozak sequence	text
AAP	text
Bgh polyA	text
CMV	text
T7 promoter	text
WRS	text
WPRE	text
Bgh polyA	text
KanR	text

Sequence for AAP-97TGA

cccgtgtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
 gaagtgagccagtgagttgattgcagtcacgctggagctctgaggctcgtcctg

aatgatatgCGaccgCCggagggttgcgtttgagacgggCGacagatccagtcgcgctg
ctctcgatccgctagggcgccgctctagaactagtggatccccggaagatcaga
agttcctattccgaagttcctattctctagaaagtataggaacttctgatctgcgcagc
cgccatgccccgggttttacgagattgtgattaaggtccccagcgaccttgacgagcatc
tccccggcatttctgacagctttgtgaactgggtggccgagaaggaatgggagttgccg
ccagattctgacatggatctgaatctgattgagcaggcaccctgaccgtggccgagaa
gctgcagcgcgactttctgacggaatggcgccgtgtgagtaaggccccggaggccctt
tctttgtgcaatttgagaagggagagagctacttccacatgcacgtgctcgtggaacc
accggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaatttaccgoggatcgagccgactttgccaaactggttcgCGgtcacaaga
ccagaaatggcgccggaggcggaacaaggtggtggatgagtgctacatcccccaattac
ttgctccccaaaaccagcctgagctccagtgggcggtgactaatatggaacagtattt
aagcgctgtttgaatctcacggagcgtaaacggttggtggcgagcatctgacgcagc
tgctcgacagcgcaggagcagaacaaagagaatcagaatcccaattctgatgcgcccgtg
atcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggacaaggg
gattacctcgagagaagcagtggtaccaggaggaccaggcctcatacatctccttcaatg
cgccctccaactcgcggtccccaaatcaaggctgccttggaacaatgcgggaaagattatg
agcctgactaaaaccgccccgactacctgggtggccagcagcccgtggaggacatttc
cagcaatcggatttataaaattttggaactaaacgggtacgatccccaatatgcggctt
ccgtctttctgggatgggcccagaaaaagttcggcaagaggaacaccatctggctgtt
ggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgcctt
ctacgggtgcgtaaacctggaccaatgagaactttcccttcaacgactgtgtcgacaaga
tggtgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggccaagcc
attctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtctcggcccagataga
cccgactcccgtgatcgtcacctccaacaccaacatgtgcgcccgtgattgacgggaact
caacgaccttcgaacaccagcagccgttgcaagaccggatggtcaaatttgaactcacc
cgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagactttttccg
gtgggcaaaggatcacgtggttgagggtggagcatgaattctacgtcaaaaagggtggag
ccaagaaaagaccgccccagtgacgcagatataagtgagccaaacgggtgcgcgag
tcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggta
ccaaaacaaatgttctcgtcacgtggcatgaatctgatgctgtttccctgcagacaat
gcgagagaatgaatcagaattcaaatatctgcttcaactcacggacagaaagactgttta
gagtgctttcccgtgtcagaatctcaaccgtttctgtcgtcaaaaaggcgtatcagaa
actgtgctacattcatatcatgggaaaggtgccagacgcttgcactgcctgcgatc
tggtcaatgtggatttgatgactgcatctttgaacaataaatgatttaaatcaggtat
ggctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataagac
agtgggtggaagctcaaaccctggcccaccaccaccaagcccgcagagcggcataaggac
gacagcaggggtcttgtgcttccctgggtacaagtacctcggacccttcaacggactcga
caagggagagccggtcaacgaggcagacgcgcggccctcgagcacgacaaagcctacg
accggcagctcgacagcggagacaaccctgacctcaagtacaaccacgcccagcgcggag
tttcaggagcgccttaagaagatacgtcttttgggggcaacctcggacgagcagctctt
ccaggcgaaaaagaggggtcttgaacctctggccctggttgaggaacctgtaagacgg
ctccgggaaaaagaggccggtagagcactctcctgtggagccagactcctcctcggga
accggaaaggcggccagcagcctgcaagaaaaagattgaattttggctcagactggaga
cgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctggtc
tgggaaactaatcagatggctacaggcagtgggcgaccaatggcagacaataacgagggc
gccgacggagtggttaattcctcgggaaattggcattgcgattccacatggatgggcga
cagagtgatcaccaccagcaccgaacctgggcccctgccacctacaacaaccacctct

acaacaaatttccagccaatcaggagcctogaacgacaatcactactttggctacagc
acccttgggggtatthttgacttcaacagattccactgccacttttcaccacgtgactg
gcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagctct
ttaacattcaagtcaaagaggtcacgcagaatgacggtacgacgacgattgccaataac
cttaccagcacggttcaggtgthttactgactcggagtagcagctcccgtacgtcctcgg
ctcggcgcacaaaggtgcctcccggcgttcccagcagacgtcttcatgggtgccacagt
atggatacctcacctgaacaacgggagtcaggcagtaggacgctcttcattttactgc
ctggagtagctttccttctcagatgctgcgtaccggaaacaactttacctcagctacac
ttttgaggacgttctttccacagcagctacgctcacagccagagctctggaccgtctca
tgaatcctctcatcgaccagtagctgtattacttgagcagaacaacaactccaagtgg
accaccacgcagtcaggcttcagttttctcaggccggagcagtgacattcgggacca
gtctaggaactggcttctggaccctgttaccggcagcagcagtagatcaagacatctg
cggataacaacaacagtgaaactcgtggactggagctaccaagtagcacctcaatggc
agagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaa
gttttttctcagagcggggttctcatctttgggaagcaaggctcagagaaaacaatg
tggacattgaaaaggtcatgattacagacgaagaggaaatcaggacaaccaatcccgtg
gctacggagcagtaggttctgtatctaccaacctccagagaggcaacagacaagcagc
taccgcagatgtcaacacacaaggcgttcttccaggcatgggtctggcaggacagagatg
tgtacctcaggggccatctgggcaaagattccacacacggacggacattttcacccc
tctcccctcatgggtggattcggacttaaacacctcctccacagattctcatcaagaa
caccccgtagctgcgaatccttcgaccacctcagtgccgcaaagtttgcttcttca
tcacacagtagctccacgggacaggtcagcgtggagatcgagtgaggagctgcagaaggaa
aacagcaaacgctggaatcccgaatcagtagacttccaactacaacaagtctgttaa
tgtggactttactgtggacactaatggcgtgtattcagagcctcgccttggcaccac
gatacctgactcgtaatctgtaattgcttgtaatacaataaacgtttaattcgtttca
gttgaactttgggtctctgcgtatthttcttcttatctagthttccatgggtacgtagata
gtagcatggcgggttaatacattaactacagcccgggctttaaacagcgggaggagg
tggagctcgtgacgtgaattacgtcatagggttagggaggtcctgtattagaggtcacgt
gagtgthttgogacattttgogacaccatgtggtctcgtcggggggggggggcccagtg
agcacgcagggtctccattttgaagcgggaggttgaaacgagcgtggcgcgctcactg
gccgtcgtthttacaacgtcgtgactgggaaaacctggcgttaccacacttaatcgct
tgcagcacatcccccttccagcgtggcgtaatagcgaagaggcccgcaccgatcgcc
cttcccacatgcacggccgcaataacctgcaggatccgthttgctgctgcttccgatgta
cgggcccagatatacgcgttgacattgattattgactagttattaatagtaatacaattac
ggggtcattagttcatagcccataatggagttccgcgttacataacttacggtaaatg
gcccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgtt
cccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggta
aactgccacttggcagtagcatcaagtgtatcatatgccaagtagcccccctattgacg
tcaatgacggtaaatggcccgcctggcattatgccagtagacatgacctatgggacttt
cctacttggcagtagcatctacgtattagtcacgctattaccatgggtgatgcggtthttg
gcagtagcatcaatgggctggatagcggthttgactcacggggatttccaagctccacc
ccattgacgtcaatgggagthttgthttggcaccaaaatcaacgggactttccaaaatgt
cgtaacaactccgccccattgacgcaaatgggaggtaggcgtgtacgggtgggaggtcta
tataagcagagctctctggctaactagagaacctgcttactggcttatcgaaatta
atagtagctcactatagggagaccaagctgggtagcgcggcaccctgggagacgcagac
tcagtagctgacccccagcctctcggacagccaccagcagccccctctgggtctgggaa
taatacagtaggctacaggcagtgccgcaccaatggcagacaataacgagggcggcagc
gagtgggtaattcctcgggaaatggcattgcgattccacatggatgggagcagagct

atcaccaccagcaccogaacctgggocctgccacctacaacaaccacctotacaaca
aatttccagccaatcaggagcctcgaacgacaatcactgattggctacagcacccttg
ggggtattttgacttcaacagattccactgccacttttcaccacgtgactggcaaagac
tcatcaacaacaactggggattccgaccaagagactcaacttcaagctctttaacatt
caagtcaaagaggtcacgcagaatgacggtagcagcagcattgccaataaccttaccag
cacggttcagggtgttactgactcggagtaccagctcccgtacgtcctcggctcggcgc
atcaaggatgcctcccgcggttcccagcagacgtcttcatgggtgccacagtatggatac
ctcaccctgactcagagtcagagtcagagggcccggtttaaacccgctgatcagcctcgac
tgtgccttctagttgccagccatctggtgtttgcccctccccgctgccttcttgacc
tggaagggtgccactccactgtcctttcctaataaaaatgaggaaattgcatcgcattgt
ctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggagga
ttgggaagacaatagcaggcatgctggggatgcgggtgggctctatggcctctgaggcgg
aaagaaccctagggttattaatagtaatacaattacggggtcattagttcatagccatata
atggagttccgcggttacataacttacggtaaatggcccgcctggctgaccgccaacga
ccccgcctttagcgtcaataatgacgtatggtcccatagtaacgccaatagggactt
tccattgacgtcaatgggtggactatttacggtaaacgcccacttggcagtacatcaa
gtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgcctg
gcattatgccagtacatgacctatgggactttcctacttggcagtacatctacgtat
tagtcatcgtattaccatgggtgatgcgggttttggcagtacatcaatgggctggatag
cggtttgactcacggggatttccaagctccaccccattgacgtcaatgggagtttgtt
ttggcaccaaaatcaacgggactttccaaaatgctcgtacaactccgccccattgacgc
aaatgggcggtaggcgtgtacgggtgggagggtctatataagcagagctctctggctaact
agagaaccactgcttactggcttatcgaatataatagactcactataggagacca
agctggctagcgcaccatgactaagccatcgtttttgctggcgcacagocctcagggt
gaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatga
ctaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcac
agaagctgcgtaaagcgcagcctggatacgcctggccttgatctggcttgtggatcgat
cctgagaaaagcaccattttgttcagtcacacgtgccggaacatgcacagttaggctg
ggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagata
aatctgcgcgttatgocgagaacatcaacgctggctctgtttgactatccggtgctgatg
gcagcggacatcctgctgtatcaaactaatctgggtccttgtgggtgaagaccagaaaca
gcacctcgaactgagccgcgatattgccagcgtttcaacgcgctgtatggcgagatct
ttaagggtgccggagccggtttattccgaaatctggcgcgcgctaatgctcgtgctggag
ccgaccaagaagatgtccaagctgacgataatcgaataaacgttatcggcctgctgga
agatccgaaatcggtagtgaagaaaatcaaacgtgcgggtcactgactccgacgagccgc
cggtagttcgtacgatgtgcagaacaaagcgggcggttccaacctggtggatatacctt
tcagcggtaacgggcccagagcatcccagaactggaaaaacagttcgaaggcaagatgta
tggtcatctgaaagggtgaagtggctgatgcggtttccggtatgctgactgaattgcagg
aacgctataccggtttccgcaacgatgaagoccttccotgcaacaggtgatgaaagatggc
gcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcgggtgtacgaagcgattgg
ttttgtggcgaagccgtaagaattcaacgcggttaagtcgacaaatcaacctctggattac
aaaatttgtgaaagattgactgggtattcttaactatggttgctccttttacgctatgtgg
atacgtgctttaatgcctttgtatcatgctattgcttcccgtatggcctttcattttct
cctccttgataaaatcctgggtgctgtctctttatgaggagttgtggcccgttgtcagg
caacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactgggtggggcattgc
caccacctgtcagctcctttccgggactttcgtttccccctccctattgccacggcgg
aactcatcgcgcctgccttgcccgcgtgctggacaggggctcggctgttgggcactgac
aattccgtgggtgttgcggggaatcatcgtcctttccttggctgctcgcctgtgttgc

cacctggattctgcgogggagctccttctgctacgtcccttcggccctcaatccagcgg
accttcctcccggcctgctgcccgtctgcccctctccgcgtcttcgcttcgc
cctcagacgagtcggatctccctttgggcccctccccgcgtcgactttaactcgagtc
tagagggcccgtttaaacccgctgatcagcctcgaactgtgccttctagttgccagccat
ctggttgttggccctccccgctgcttcccttgaccctggaaggtgccactcccactgct
ctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattct
ggggggtggggggggcaggacagcaagggggaggattgggaagacaatagcagggcatg
ctggggatgcccgtgggctctatggactaggggtgagcagcggatcgagcagtgctgatca
ctactggaccgagctgtgctgagaccgctgatcttacggcattatacgtatgatcgg
tccacgatcagctagattatctagtcagcttgatgtcatagctgtttcctgaggctcaa
tactgaccatttaaatcatacctgacctccatagcagaaagtcaaaagcctccgaccgg
aggcttttgacttgatcggcacgtaagaggttccaactttcaccataatgaaataagat
cactaccggcgtatTTTTTgagttatcgagatTTTcaggagctaaggaagctaaaatg
agccatattcaacgggaaacgtcttgccttgaagccgcgattaaattccaacatggatgc
tgatttataatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatct
atcgattgatgggaagcccgatgcgccagagttgttctgaaacatggcaaggtagc
gttgccaatgatgttacagatgagatggtcaggctaaactggctgacggaatttatgcc
tctccgaccatcaagcattttatccgactcctgatgatgatggttactcaccactg
cgatcccagggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaat
attggtgatgctgctggcagtgcttctgcccgggttgcattcogattcctgtttgtaattg
tccttttaacggcgatcgcgtatttctgctcgcctcagggcgaatcacgaatgaataacg
gtttgggtggtgagtgatTTTTgatgacgagcgaatggctggcctgttgaacaagtc
tgaaagaaatgcataaactcttgccattctcaccggattcagtcgctcactcatggtga
tttctcacttgataaccttatttttgacgaggggaaattaatagggttgattgatggtg
gacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcgg
gagttttctccttcattacagaaacggcTTTTTcaaaaatatggtattgataatcctga
tatgaataaattgcagtttacttgatgctcgatgagtttttctaataaggacctaata
gtaatcacctggctcaccttcgggtgggcttttctgcttgcgttgcggctttttccatagg
ctccgccccctgacgagcatcacaanaatcgatgctcaagtcagaggtggcgaaaccc
gacaggactataaagataccaggcgtttccccctggaagctccctcgtgctcctcctg
ttccgacctgccgcttaccggatacctgtccgccttttctcccttcgggaagcgtggcg
ctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagct
gggctgtgtgcacgaacccccggtcagcccagcctgctgccttatccggtaactatc
gtcttgagccaaccggtaagacagcacttatcgcactggcagcagccactggtaac
aggattagcagagcaggtatgtaggggtgctacagagttcttgaagtgggtggcctaa
ctacggctacactagaagaacagtatTTTggtatctgctcctgctgaagccagttacct
cggaaaaagagttggtagctcttgatccggcaaaacaaaccaccgctggtagcgggtggt
TTTTTggttgaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttg
atTTTctaccgaagaaaggcca

The sequence for the *AAP* gene for pIDTSMART-RC2-AAPstop60-CMV-AAP110TGA-CMV-WRS is:

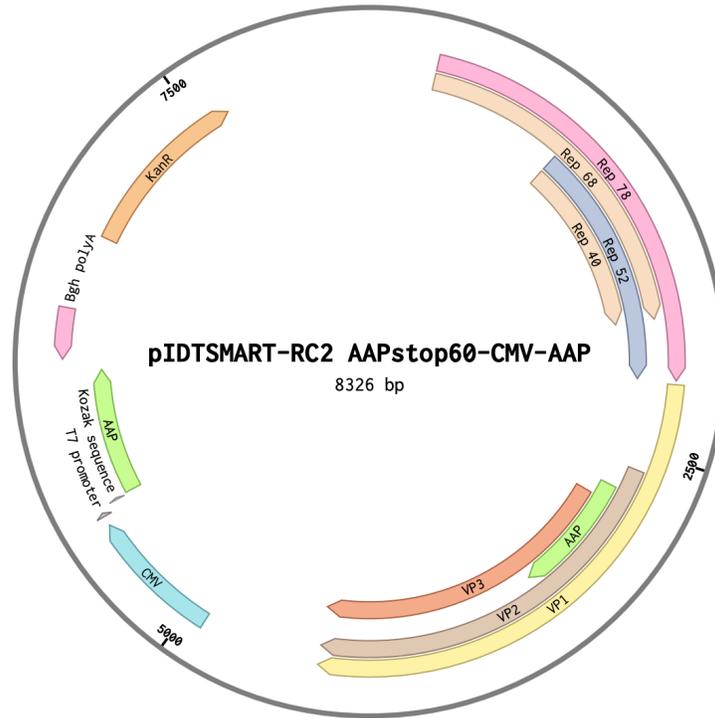
ctggagacgcagactcagtaacctgacccccagcctctcggacagccaccagcagcccc
tctgggtctgggaactaatacagatggctacaggcagtgggcgcaccaatggcagacaataa
cgagggcgccgacggagtggttaattcctcgggaaattggcattgctgattccacatgga
tgggcgacagagtcattaccaccagcaccggaacctgggcccctgccacctacaacaac
cacctctacaanaaatttccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtatTTTgacttcaTGAgttccactgccacttttccacc

gtgactggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgc
caataaccttaccagcacggttcaggtggttactgactcggagtaccagctcccgtacg
tcctcggctcggcgcacatcaaggatgcctcccgccgttcccagcagacgtcttcatggtg
ccacagtatggatacctcaccctga

The sequence for the *AAP* gene for pIDTSMART-RC2-AAPstop60-CMV-AAP124TGA-CMV-WRS is:

ctggagacgcagactcagtaacctgacccccagcctctcggacagccaccagcagcccc
tctggctctgggaactaatacagatggctacaggcagtgggcgcaccaatggcagacaataa
cgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatgga
tgggcgacagagtcaccaccagcaccggaacctgggccctgccacctacaacaac
cacctctacaacaatttccagccaatcaggagcctcgaacgacaatcactactttgg
ctacagcacccttgggggtatthttgacttcaacagattccactgccacttttccacc
gtgactggcaaagacTGAtcaacaacaactggggattccgacccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgc
caataaccttaccagcacggttcaggtggttactgactcggagtaccagctcccgtacg
tcctcggctcggcgcacatcaaggatgcctcccgccgttcccagcagacgtcttcatggtg
ccacagtatggatacctcaccctga

pIDTSMART-RC2-AAPstop60-CMV-AAP



Sequence color-coding key

Feature	Color
Rep	text
Cap – AAPstop60	text
CMV	text
T7 promoter	text
Kozak sequence	text
AAP	text
Bgh polyA	text
KanR	text

Sequence for wild type AAP

```

cccggtgtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
gaaggtgagccagtgagttgattgcagtcacgtaacgctggagctctgaggctcgtcctg
aatgatatgacgaccgccggagggttgcgtttgagacgggcgacagatccagtcgcgctg
ctctcgtcgatccgctagggcggccgctctagaactagtgatcccccggaagatcaga
agttcctattccgaagttcctattctctagaaagtataggaacttctgatctgctgcagc
cgccatgccggggttttacgagattgtgattaaggtcccagcgaccttgacgagcatc
tgcccggcatttctgacagctttgtgaactgggtggccgagaaggaatgggagttgccg
ccagattctgacatggatctgaatctgattgagcaggcaccctgaccgtggccgagaa
gctgcagcgcgactttctgacggaatggcgccgtgtgagtaaggccccggaggccctt
tctttgtgcaatttgagaaggagagagctactccacatgcacgtgctcgtggaacc
    
```

accggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaatttaccgcggtatcgagccgactttgccaaactggttcgcggtcacaaga
ccagaaatggcgccggaggcgggaacaaggtggtggatgagtgctacatccccaattac
ttgctcccaaaaccagcctgagctccagtggcggtgactaatatggaacagtat
aagcgctggttgaatctcacggagcgtaaacgggtggtggcgagcatctgacgcag
tgtcgcagacgcaggagcagaacaagagaatcagaatcccaattctgatgcccggg
atcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggacaagg
gattacctcggagaagcagtggtaccaggaggaccaggcctcatacatctccttcaatg
cggcctccaactcgcgggtcccaaatcaaggctgccttgacaatgcgggaaagattatg
agcctgactaaaaccgccccgactacctggtgggocagcagcccgtggaggacatttc
cagcaatcggattataaaaatttggaactaaacgggtacgatcccaatatgcccgtt
ccgtcttctgggatgggcccagaaaaagtctggcaagaggaacaccatctggctggtt
ggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgcctt
ctacgggtgcgtaaactggaccaatgagaactttccctcaacgactgtgtcgacaaga
tggtgatctggtgggaggagggaagatgaccgccaaggtcgtggagtcggccaaagcc
attctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagataga
cccgactcccgtgatcgtcacctccaacaccaacatgtgcgcccgtgatgacgggaact
caacgaccttcgaacaccagcagccgttgcaagaccggatggtcaaatttgaactcacc
cgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagacttttccg
gtgggcaaaggatcacgtggttgagggtgagcatgaattctacgtcaaaaagggtggag
ccaagaaaagaccgccccagtgacgcagatataagtgagccaaacgggtgcgcgag
tcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggt
ccaaaacaatggtctcgtcacgtgggatgaatctgatgctggttccctgcagacaat
gogagagaatgaatcagaattcaaatatctgcttactcacggacagaaagactgttta
gagtgcttcccggtgcagaatctcaaccgcttctgctcgtcaaaaaggcgtatcagaa
actgtgctacattcatatcatatgggaaaggtgccagacgcttgcaactgcctgcgatc
tggtcaatgtggatttgatgactgcatcttgaacaataaatgatttaaatcaggtat
ggctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataagac
agtgggtggaagctcaaacctggcccaccaccaccaagcccgcagagcggcataaggac
gacagcaggggtcttgtgcttctgggtacaagtacctcggacccttcaacggactcga
caagggagagccggtcaacgaggcagacgcgcggccctcgagcacgacaaaagcctacg
accggcagctcgacagcggagacaaccgctacctcaagtacaaccacgcccagcggag
tttcaggagcgccttaagaagatacgtcttttgggggcaacctcggacgagcagctct
ccaggcgaaaaagaggggtcttgaacctctgggctggttgaggaacctgttaagacgg
ctccgggaaaaaagagggcggtagagcactctcctgtggagccagactcctcctcggga
accggaaaggcgggcccagcagcctgcaagaaaaagattgaattttggtcagactggaga
cgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctggtc
tggaactaatacagatggctacaggcagtgggcgcaccaatggcagacaataacgagggc
gccgacggagtggttaattcctcgggaaattggcattgcgattccacatggatgggcca
cagagtgatcaccaccagcaccgaaacctgggcccctgccacctacaacaaccacctct
acaacaaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagc
acccttgggggtattttgacttcaacagattccactgccacttttaccacgtgactg
gcaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagctct
ttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgccaataac
cttaccagcaggttcaggtggttactgactcggagtagcagctcccgtacgtcctcgg
ctcggcgcacaaaggtgctcctccgcccgttcccagcagacgtcttcatgggtgccacagt
atggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgc
ctggagtagcttctctcagatgctgcgtaccggaaacaactttacctcagctacac

ttttgaggacggttcctttccacagcagctacgctcacagccagagtctggaccgtctca
tgaatcctctcatcgaccagtacctgtattacttgagcagaacaaactccaagtga
accaccacgcagtcaggcttcagttttctcaggccggagcagtgacattcgggacca
gtctaggaactggcttctcaggacctgttaccgccagcagcagtgatcaaagacatctg
cggataacaacaacagtgaaactcgtggactggagctaccaagtaccacctcaatggc
agagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaa
gttttttctcagagcggggttctcatctttgggaagcaaggctcagagaaaaaaaatg
tgacattgaaaaggctcatgattacagacgaagaggaaatcaggacaaccaatcccgtg
gctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagcagc
taccgcagatgtcaacacacaaggcgttcttccaggcatggctctggcaggacagagatg
tgtacctcaggggccatctgggcaagattccacacacggacggacattttcacccc
tctcccctcatgggtggattcggacttaaacacctcctccacagattctcatcaagaa
caccccgtacctgcgaatccttcgaccaccttcagtgcggaagtttgcttccttca
tcacacagtaactccacgggacaggtcagcgtggagatcgagtgggagctgcagaaggaa
aacagcaaacgctggaatcccgaattcagtaacttccaactacaacaagtctgttaa
tgtggactttactgtggacactaatggcgtgtattcagagcctcgcctcattggacca
gatacctgactcgtaatctgtaattgcttgtaatacaataaacggttaattcgtttca
gttgaactttgggtctctgcgtattttctttcttatctagtttccatggctacgtagataa
gtagcatggcgggttaatcattaactacagcccgggctttaaacagcgggaggagggg
tgagtcgtgacgtgaattacgtcatagggttagggaggtcctgtattagaggtcacgt
gagtgttttgcgacattttgcgacaccatgtggtctcgtggggggggggggccgagtg
agcacgcagggtctccattttgaagcgggaggtttgaacgagcgtggcgcgctcactg
gccgtcgttttacaacgctcgtgactgggaaaacctggcgttaccacttaatcgct
tgcagcacatccccctttcgcagctggcgtaatagcgaagaggcccgcaccgatcgcc
cttcccctcatcgccgcaataacctgcaggatccggttttgcgctgcttcgcgatgta
cgggcccagatatacgcgttgacattgattattgactagttattaatagtaatacaattac
ggggtcattagttcatagcccataatggagttccgcgttacataacttacggtaaatg
gcccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgtt
cccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggta
aactgccacttggcagtacatcaagtgtatcatatgccaaagtagccccctattgacg
tcaatgacggtaaatggcccgcctggcattatgccagtagatgacctatgggacttt
cctacttggcagtagatctacgtattagtcacgctattaccatgggtgatgcggttttg
gcagtagatcaatgggcgtggatagcgggttgactcacggggatttccaagtctccacc
ccattgacgtcaatgggagtttggttttggcaccaaaatcaacgggactttccaaaatgt
cgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtcta
tataagcagagctctctggctaactagagaaccactgcttactggcttatcgaaatta
atacgactcactataggagacccaagctgggctagcggcggcaccctggagacgcagac
tcagtagctgacccccagcctctcggacagccaccagcagccccctctgggtctgggaac
taatacagatggctacaggcagtggcgcaccaatggcagacaataacgagggcgccgacg
gagtggttaattcctcgggaaattggcattgcgattccacatggatgggcgacagagtc
atcaccaccagcaccogaacctgggcccctgccacctacaacaaccacctctacaaca
aatttcagccaatcaggagcctcgaacgacaatcactactttggctacagcaccctt
gggggtattttgacttcaacagattccactgccacttttaccacgtgactggcaaaga
ctcatcaacaacaactggggattccgacccaagagactcaactcaagctctttaacat
tcaagtcaaaagaggtcacgcagaatgacggtagcagcagattgccaataaccttacc
gcacgggtcaggtgtttactgactcggagtagaccagctcccgtacgtcctcggctcggcg
catcaaggatgcctcccggcgttcccagcagacgtcttcatgggtgccacagtatggata
cctcacctga

ctgtgccttctagttgccagccatctggtggttggccctccccgtgccttcttgacc
ctggaagtgccactcccactgtccttctctaataaaatgaggaaattgcatcgattg
tctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggagg
attgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcg
gaaagaaccctaggggtgcgagcggatcgagcagtgtcgatcactactggaccgcgagc
tgtgctgcgaccctgatcttacggcattatacgtatgatcggtccacgatcagctaga
ttatctagtcagcttgatgtcatagctgtttctgaggctcaatactgaccatttaaat
catacctgacctccatagcagaaagtcaaaagcctccgaccggaggcttttgacttgat
cggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggctatt
ttttgagttatcgagattttcaggagctaaggaagctaaaatgagccatattcaacggg
aaacgtcttgcttgaagccgcgattaaattccaacatggatgctgatttataatgggtat
aaatgggctcgcgataatgtcgggcaatcagggtgcgacaatctatcgattgtatgggaa
gcccgatgcccagagttgttctgaaacatggcaaaggtagcgttgccaatgatgta
cagatgagatggtcaggctaaactggctgacggaatttatgcctcttccgaccatcaag
cattttatccgtactcctgatgatgcattggtactcaccactgcgatcccagggaaaac
agcattccaggattagaagaatatacctgattcagggtgaaaatattggtgatgcgctgg
cagtgttcctgcgcccgggtgcattcagattcctgtttgaattgtccttttaacggcgat
cgcgtatttcgtctcgcctcaggcgcgaatcacgaatgaataacgggtttgggtggcgcgag
tgattttgatgacgagcgtaatggctggcctggtgaacaagtctggaaagaaatgcata
aacctttgccattctcaccggattcagtcgtcactcatgggtgatttctcacttgataac
cttatttttgacgaggggaaattaataggttgattgatggtggacgagtcggaatcgc
agaccgataaccaggatcttgccatcctatggaactgcctcgggtgagtttctccttcat
tacagaaacggctttttcaaaaataggtattgataatcctgatatgaataaattgcag
tttacttgatgctcgatgagtttttctaataaggacctaataatgtaatcacctggctca
ccttcgggtgggcttttctgcgttgctggcggtttttccataggctccgccccctgacg
agcatcacaanaatcgatgctcaagtcagaggtggcgaaaccgcagaggactataaaga
taccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgct
taccggatacctgtccgcctttctccctcgggaagcgtggcgcttttctcatagctcac
gctgtaggatctcagttcgggtgtaggtcgttcgcctccaagctgggctgtgtgcagaa
cccccgcttcagcccagcctgcgccttatccggtaactatcgtcttgagccaacc
ggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcga
ggatgtaggcggtgctacagagttcttgaagtgggtggcctaactacggctacactaga
agaacagtatttggtatctgcgctctgctgaagccagttacctcggaanaagagttgg
agctcttgatccggcaaaacaaccacgctggtagcgggtgggttttttgggttgcaagca
gcagattacgcgcagaaaaaaggatctcaagaagatcctttgattttctaccgaagaa
aggcca

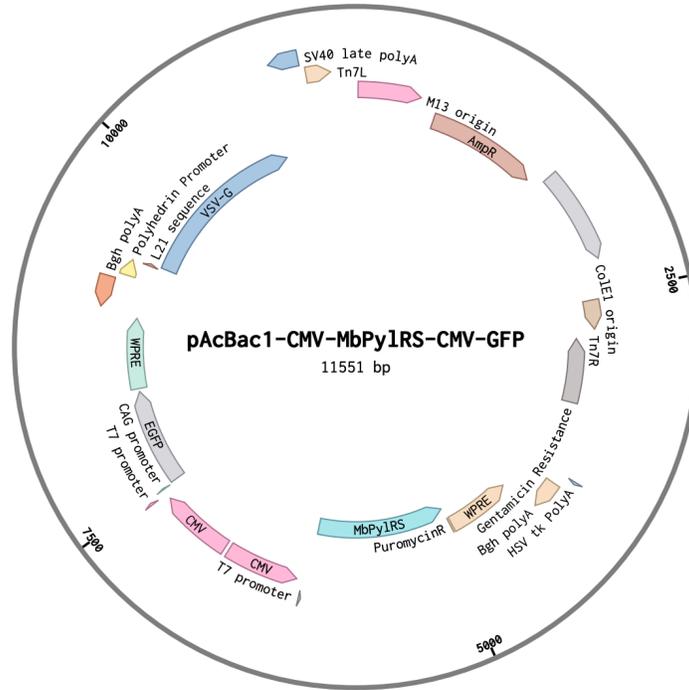
The sequence for the *AAP* gene for pIDTSMART-RC2-AAPstop60-CMV-AAP-97TGA is:

ctggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagcccc
tctggctctgggaactaatacgtatggctacaggcagtgggcgcaccaatggcagacaataa
cgagggcgccgacggagtggttaattcctcgggaaattggcattgcatccacatgga
tgggcgacagagtcaccaccagcaccgaacctggggcctgccacctacaacaac
cacctctacaanaaatttccagccaatcaggagcctcgaacgacaatcactTGAttgg
ctacagcacccttgggggtattttgacttcaacagattccactgccacttttccacc
gtgactggcaaagactcatcaacaacaactggggattccgaccaagagactcaacttc
aagctctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcattgc

caataaccttaccagcacggttcaggtgtttactgactcggagtaccagctcccgtacg
tcctcggctcggcgcacatcaaggatgcctcccgcggttcccagcagacgtcttcatggg
ccacagtatggatacctcaccctga

Chapter 4 Plasmids

pAcBac-CMV-MbPylRS-CMV-GFP



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
CMV promoter	text
CMV Promoter	text
T7 promoter	text
EGFP	text
WPRE	text
Bgh polyA	text
Polyhedrin Promoter	text

L21 sequence	text
VSV-G	text
SV40 late polyA	text
Tn7L	text

Sequence for wild type EGFP

ctgaatggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggt
tacgcgcagcgtgaccgctacacttgccagcgccttagcgcgccgctcctttcgctttct
tccttcctttctcgccacgttcgcccggctttcccgcgtaagctctaaatcggggctc
ccttagggttccgatttagtgctttacggcacctcgaccccaaaaacttgattagg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccccttgacgttg
agtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggattttgccgatttcggcctattgggttaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgtttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaaggaagagtatgagattcaacatttccgtgctcgcccttattcccttttttgccg
cattttgccttctggtttttgctcaccagaaacgctggtgaaagtaaaagatgctgaa
gatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgcccgaagaacgttttccaatgatgagcacttttaaagttctgctat
gtggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgcccatacac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtgataaacactgcccga
actacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatg
gggatcatgtaactcgcttgatcggtgggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgctgtagcaatggcaacaacgcttgcgcaaactattaa
ctggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcgcccttccggctggctggtttattgctgataa
atctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagacca
agtttactcatatatacttttagattgattttaaacttcatttttaatttaaaggatct
aggatgaagatcctttttgataatctcatgaccaaatacccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttct
gcgcgtaatctgctgcttgcaaaaaaaaccaccgctaccagcgggtggtttgtttgc
cggatcaagagctaccaactccttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagcgcgtagttaggccaccacttcaagaactctgtagc
accgctacatacctcgctctgctaatacctggtaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctc
ggctgaacggggggttcgtgcacacagcccagcttgagcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggtaagcggcagggctcggaacaggagagcgcacgaggagcttccaggg
ggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg
atttttgtgatgctcgtcagggggggcggagcctatggaaaaacgccagcaacgcggcct
ttttacggttcctggcctttttgctggcctttttgctcacatggtcttttctgcttatcc
cctgattctgtggataaccgtattaccgcttttgagtgagctgataccgctcgccgag
ccgaacgaccgagcgcagcaggtcagtgagcaggaagcgggaagagcgcctgatgcggt

at t t t t c t c t t a c g c a t c t g t g c g g t a t t t c a c a c c g c a g a c c a g c c g c g t a a c c t g g c
a a a a t c g g t t a c g g t t g a g t a a t a a t g g a t g c c c t g c g t a a g c g g g t g t g g g c g g a c a
a t a a a g t c t t a a a c t g a a c a a a a t a g a t c t a a a c t a t g a c a a t a a a g t c t t a a a c t a g a
c a g a a t a g t t g t a a a c t g a a a t c a g t c c a g t t a t g c t g t g a a a a g c a t a c t g g a c t t t
t g t t a t g g c t a a a g c a a a c t c t t c a t t t t c t g a a g t g c a a a t t g c c c g t c g t a t t a a a g
a g g g g c g t g g c c a a g g g c a t g g t a a a g a c t a t a t t c g c g g c g t t g t g a c a a t t a c c g a
a c a a c t c c g c g g c c g g g a a g c c g a t c t c g g c t t g a a c g a a t t g t t a g g t g g c g g t a c t t
g g g t c g a t a t c a a a g t g c a t c a c t t c t t c c c g t a t g c c c a a c t t t g t a t a g a g a g c c a c
t g c g g g a t c g t c a c c g t a a t c t g c t t g c a c g t a g a t c a c a t a a g c a c c a a g c g c g t t g g
c c t c a t g c t t g a g g a g a t t g a t g a g c g c g g t g g c a a t g c c c t g c c t c c g g t g c t c g c c g
g a g a c t g c g a g a t c a t a g a t a t a g a t c t c a c t a c g c g g t g c t c a a a c c t g g g c a g a a c
g t a a g c c g c g a g a g c g c c a a c a c c g c t t c t t g g t c g a a g g c a g c a a g c g c g a t g a a t g
t c t t a c t a c g g a g c a a g t t c c c g a g g t a a t c g g a g t c c g g c t g a t g t t g g g a g t a g g t g
g c t a c g t c t c c g a a c t c a c g a c c g a a a g a t c a a g a g c a g c c c c g c a t g g a t t t g a c t t g
g t c a g g g c c g a g c c t a c a t g t g c g a a t g a t g c c c a t a c t t g a g c c a c c t a a c t t t g t t
t a g g g c g a c t g c c c t g c t g c g t a a c a t c g t t g c t g c t g c g t a a c a t c g t t g c t g c t c c a
t a a c a t c a a a c a t c g a c c c a c g g c g t a a c g c g t t g c t g c t t g g a t g c c c g a g g c a t a g
a c t g t a c a a a a a a c a g t c a t a a c a a g c c a t g a a a c c g c c a c t g c g c c g t t a c c a c c g
c t g c g t t c g g t c a a g g t t c t g g a c c a g t t g c g t g a g c g c a t a c g t a c t t g c a t t a c a g
t t t a c g a a c c g a a c a g g c t t a t g t c a a c t g g g t t c g t g c t t c a t c c g t t t c c a c g g t g
t g c g t c a c c c g g c a a c c t t g g g c a g c a g c g a a g t c g a g g c a t t t c t g t c c t g g c t g g c g
a a c g a g c g c a a g g t t t c g g t c t c c a c g c a t c g t c a g g c a t t g g c g g c c t t g c t g t t c t t
c t a c g g c a a g g t g c t g t g c a c g g a t c t g c c c t g g c t t c a g g a g a t c g g t a g a c c t c g g c
c g t c g c g g c g c t t g c c g g t g g t g c t g a c c c g g a t g a a g t g g t t c g c a t c c t c g g t t t
c t g g a a g g c g a g c a t c g t t t g t t c g c c c a g g a c t c t a g c t a t a g t t c t a g t g g t t g g c t
a c g t a c c c g t a g t g g c t a t g g c a g g g c t t g c g t t a a t g c g c c g c t a c a g g g c g c g t g g
g g a t a c c c c t a g a g c c c c a g c t g g t t c t t t c c g c c t c a g a a g c c a t a g a g c c a c c g c
a t c c c c a g c a t g c c t g c t a t t g t c t t c c c a a t c c t c c c c t t g c t g t c c t g c c c a c c c
c a c c c c c a g a a t a g a a t g a c a c c t a c t c a g a c a a t g c g a t g c a a t t t c c t c a t t t t a t
t a g g a a g g a c a g t g g g a g t g g c a c c t t c c a g g g t c a a g g a a g g c a c g g g g g a g g g g c a
a a c a c a g a t g g c t g g c a a c t a g a a g g c a c a g t c g a g g c t g a t c a g c g g g t t t a a c g g
g c c c t c t a g a c t c g a g t t a a a g t c g t c g c g g g a g g c g g c c c a a a g g g a g a t c c g a c t c
g t c t g a g g g c g a a g g c g a a g a c g c g g a a g a g g c c g c a g a g c c g g c a g c a g g c c g c g g g a
a g g a a g g t c c g c t g g a t t g a g g g c c g a a g g g a c g t a g c a g a a g g a c g t c c c g c g c a g a a
t c c a g g t g g c a a c a c a g g c g a g c a g c c a a g g a a g g a c g a t g a t t t c c c c g a c a a c a c c
a c g g a a t t g t c a g t g c c c a a c a g c c g a g c c c c t g t c c a g c a g c g g g c a a g g c a g g c g g c
g a t g a g t t c c g c c g t g g c a a t a g g g a g g g g a a a g c g a a a g t c c c g g a a a g g a g c t g a c
a g g t g g t g g c a a t g c c c c a a c c a g t g g g g g t t g c g t c a g c a a a c a c a g t g c a c a c c a c g
c c a c g t t g c c t g a c a a c g g g c c a c a a c t c c t a t a a a g a g a c a g c a a c c a g g a t t t a t a
c a a g g a g g a g a a a t g a a a g c c a t a c g g g a a g c a a t a g c a t g a t a c a a a g g c a t t a a a g
c a g c g t a t c c a t a g c g t a a a a g g a g c a a c a t a g t t a a g a a t a c c a g t c a a t c t t t c a
c a a a t t t t g t a a t c c a g a g g t t g a t t g t c g t c t t a a c g c g t t g a a t t c t t a c a g a t t g g
t t g a a a t c c c a t t a t a g t a a g a t t c g g a c c t t g a t g c c c t c t t a a t g t t t t t a a a g c c g
t g c a t a a c c t t g a g c a a g c g t t c a a g a c c a a a a c c t g c a c c t a t c c a t g g t t t t g t c a a t
a c c c c a t t c t c t a t c a a g a g a a a c t g g c c c g a c g a c t g c c g a a g a a a g c t c c a g g t c c c
c g t g c a t t a t a t c a a g a g t a t c c c c a a a g a c c a t a c a g g a a t c t c c t a c g a t t t c g a a g
t c g a t t t c c a g a t a g t c c a g a a a c t c t t t g a t g a g a g c t t c a a g a t t t t c c c g a g t a c a
t c c c g a a c c c a t c t g a c a g a a g t t c a c c a t a g t a a a t t c t t c c a g g t g c t c t t t g c c g t

cagactctttccggtaacaaggtccgacttcgaaaatTTTTattgggctggtaaaatc
ctatcgagttttcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag
atTTTTatccaccggaagatctgTTTTgaaagttcagtatcattattaataccattc
tctccacgtattccgcccgaataaggataggagactttatctccagaaaaccccggtct
acgaaaaatTTcgtaatatcacgttcaggtttaccgaggtagtcttctctatcattggg
atagagccgctgaaaatcgTTTTtcttcttgtcacaagttcaggctcaagttccctga
aaggctttgccatatttagagaaatTTTatcctctggacttaagagagcctcaacccta
tcaagctggcttcttgtaagtgaaggagcaggagccgatgcgggaacagacgaatttg
agttgattttgcaggcgaaggtacagatctggatgtgttcggtgatgcctttgcagaaa
cagaatTTTccagaggcttcggagcccttgaaactgatttcggcatagctTTTTtgacc
tttgagcagaaaactaccctaactTTTcacactgTTTTgcttctcggttgatcttgtgag
aaaattattgatacctcgtccgaaaccctacatcgTTTgcaggTTTTtctgtacttat
gatgtctgaatgctctggctggttctacaactcctggaattattcacaacaagatggctc
ccacacgccatttcaatgtatatTTTacttcttgagacctcatgggtgcttgattttgtg
gagcgtgccagtcctggacatccagagcccggctcgcagatattaaaacatctaattggtt
TTTTatccatgggtggcgctagtcagcttgggtctcccTatagtgagtcgattaatttc
gataagccagtaagcagtggggttctctagtttagccagagagctctgcttatatagacct
cccaccgtacacgcctaccgccattttgctgcaatggggcggagttgttacgacatttt
ggaaagtcccgttgattttggtgccaaaacaaactcccattgacgtcaatgggggtggag
acttggaatccccgtgagtcaaaccgctatccacgccattgatgtactgccaaaacc
gcatcaccatggtaatagegatgactaatacgtagatgtactgccaagtaggaaagtcc
cataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgtcaat
agggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgta
aatagtcaccaccattgacgtcaatggaaagtccctattggcgttactatgggaacatac
gtcattattgacgtcaatgggcgggggtcgttgggoggctcagccaggcgggccatttac
cgtaagttatgtaacgcggaaactccatatatgggctatgaactaatgaccccgtaattg
attactattaataactagaaataatgctagcaataatcctgcaggattattctagttatt
aatagtaatcaattacggggctattagttcatagcccatatatggagttccgcgttaca
taacttacggtaaatggcccgcctggctgaccgcccacgacccccgccattgacgtc
aataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatggg
tgactatttacggtaaaactgccacttggcagtacatcaagtgtatcatatgccaagt
acgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtacat
gaccttatgggactttcctacttggcagtacatctacgtattagtcacgtattacca
tggtgatgcggTTTTggcagtacatcaatgggcgtggatagcggTTTTgactcacgggga
tttccaagctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacg
ggactttccaaaatgtcgtacaactccgccccattgacgcaaatgggcggtagggcgtg
tacgggtgggaggtctatataagcagagctgtcccgggaactagagaaccactgcttac
tggttatcgaaatTaatacgactcactatagggagacccggtaagctggatagcggccg
ccaccatggtgagcaagggcgaggagctgttcaccgggggtggtgccatcctggtcgag
ctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgc
cacttacggcaagctgaccctgaagttcatctgcaccaccggcaagctgccctgacct
ggcccacctcgtgaccacctgacctacggcgtgcagtgcttcagccgctaccccgac
cacatgaagcagcagacttcttcaagtccgccatgccgaaggctacgtccaggagcg
caccatcttctcaaggacgacggcaactacaagaccgcgccgaggtgaagttcgagg
gcgacacctgggtgaaccgcatcgagctgaaggcctcagcttcaaggaggacggcaac
atcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccga
caagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggca
gcgtgcagctcggcgaccactaccagcagaacacccccatcggcgacggccccgtgctg

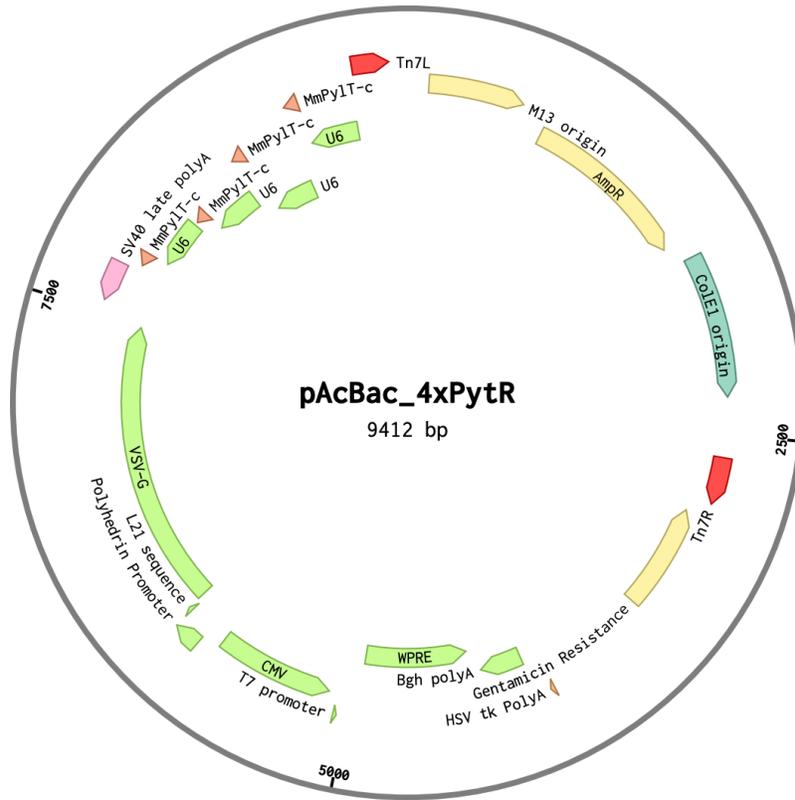
ctgcccgacaaccactacctgagcaccagtcgccctgagcaagaccccaacgagaa
gcgcatcacatggctctgctggagttcgtgaccgccgcccgggatcactctcggcatgg
acgagctgtacaagggcccttcgaacaaaaactcatctcagaagaggatctgaatatg
cataccggtcatcatcaccatcaccattgataaaaattcaacgcgttaagtgcacaatc
aacctctggattacaaaatttgtgaaagattgactggattcttaactatggtgctcct
tttacgctatgtggatagctgctttaaagcctttgtatcatgctattgcttcccgtat
ggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgt
ggcccgttgtcaggcaacgtggcgtgggtgtgactgtgtttgtgacgcaacccccact
ggttggggcattgccaccacctgtcagctcctttccgggactttcgtcttccccctccc
tattgccacggcggaactcatcgccgctgcttgcggctgctggacaggggctcggc
tgttgggactgacaattccgtgggtgttgcggggaaatcatcgtcctttccttggctg
ctcgcctgtgttggcaccctggattctgcgcgggacgtccttctgctacgtcccttcggc
cctcaatccagcggaccttccctcccggcctgctgccggctctgcggcctctccgc
gtcttcgcttcgccctcagacgagtcggatctcccttggggccgctccccgcgacga
ctttaactcgagtctagagggcccgtttaaacccgctgatcagcctcga**ctgtgcoctt**
tagttgccagccatctgttgtttgcccccccccgctgcocttcocttgaccctggaagggtg
ccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcatgtctgagtagg
tgctattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaaga
caatagcagccatgctggggatgggggtgggctctatgcattattgcccggcgggtccgt
aactccggaatattaatagatcatggagataaattaaaatgataaacatctcgcaata
aataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccg
gattattcataccgtcccaccatcggggcgg**aaactcctaataaaaaccgccacc**atgaagt
gccttttgtacttagcctttttattcattgggggtgaattgcaagttcaccatagttttt
ccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtc
aagctcagattttaaattggcataatgacttaataggcacagccttacaagtcaaaatgc
ccaagagtcacaaggctattcaagcagacgggtggatgtgtcatgcttccaaatgggtc
actacttgtgatttccgctggatggaccgaagtatataacacattocatccgatcctt
cactccatctgtagaacaatgcaaggaaagcattgaacaaacgaacaaggaacttggc
tgaatccaggcttccctcctcaaagttgtggatgcaactgtgacggatgccgaagca
gtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggt
tgattcacagttcatcaacggaaaatgcagcaattacatatgcccactgtccataact
ctacaacctggcattctgactataaggtcaaagggtatgtgattctaacctcattttcc
atggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggaggggcac
agggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaat
actgcaagcattggggagtcagactcccacaggtgtctgggtcagatggctgataag
gatctctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctcc
atctcagacctcagtgatgtaagtctaattcaggacggtgagaggatcttggattatt
ccctctgccaaagaacctggagcaaaatcagagcgggtcttccaatctctccagtggat
ctcagctatcttgcctctaaaaccaggaaaccggctcctgctttcaccataatcaatgg
taccctaaaatactttgagaccagatacatcagagtcgatattgctgctccaatcctct
caagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgg
gcaccatatgaagacgtggaaattggaccaatggagttctgaggaccagttcaggata
taagtttctttatacatgattggacatggatgttggactccgatcttcatcttagct
caaaggctcaggtgttcgaacatcctcacattcaagacgctgcttgcgaacttctgat
gatgagagtttattttttgggtgatactgggctatccaaaaatccaatcgagctttaga
aggttgggtcagtagttggaaaagctctattgcctcttttttctttatcatagggttaa
tcattggactattcttggttctccgagttggatccatctttgcattaaattaagcac
accaagaaaagacagatttatacagacatagagatgaaccgacttggaagtgataa**gg**

ccaggccggccaagccttgctcgagaagtactagaggatcataatcagccataccacattt
gtagagggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataa
aatgaatgcaattggtgtgtaacttgtttattgcagcttataatggttacaataaa
gcaatagcatcacaatttcacaataaagcatttttttcaactgcattctagttgtggt
ttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagcctag
gagatccgaaccagataagtgaaatctagttccaaactattttgtcatttttaattttc
gtattagcttacgacgctacaccagttcccacatctattttgcactcttccctaataa
tccttaaaaactccatttccaccctcccagttcccaactattttgtccgcccacagcg
gggcatttttcttctggtatggttttaatacaaacatcctgccaactccatgtgacaaa
ccgtcatcttccggtactttttctctgtcacagaatgaaaatttttctgtcatctcttc
gttattaatgtttgtaattgactgaatatcaacgcttatttgcage

The sequence for GFP-39tag is:

atggtgagcaagggcgaggagctgttcaccggggtggtgccatcctggtcgagctgga
cggcgacgtaaaccggccacaagttcagcgtgtccggcgagggcgagggcgatgccact
agggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggccc
accctcgtgaccacctgacctacggcgtgcagtgttcagccgctaccccgaccacat
gaagcagcacgacttcttcaagtccgccatgccgaaggctacgtccaggagcgcacca
tcttcttcaaggacgacggcaactacaagaccgcgccgaggtgaagttcgagggcgac
accctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct
ggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagc
agaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtg
cagctcgccgaccactaccagcagaacaccccacatcgggcgacggccccgtgctgctgcc
cgacaaccactacctgagcaccagtcgccctgagcaaagaccccacgagaagcgcg
atcacatggtcctgctggagttcgtgaccgcccgggatcactctcggcatggacgag
ctgtacaaggggccccttcgaacaaaaactcatctcagaagaggatctgaatatgcatac
cggtcacatcaccatcaccattgataa

pAcBac-4xPytR



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
T7 promoter	text
CMV promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
MmPytR	text
U6	text
Tn7L	text

Sequence of 4xPylR-A2 variant

ttctctgtcacagaatgaaaatTTTTctgtcatctcttctgattattaatgTTTgtaattg
actgaatatcaacgcttatttgcagcctgaatggcgaatgggacgCGCCctgtagCGGC
gcattaagcgcggcgggtgtggTggttacgCGcagcgtgaccgctacacttGCCagcgc
cctagcGCCgctcctttcgtttcttcccttctccttctCGccacgttcGCCggctttc
ccgTcaagctctaaatcgggggctccctttagggTtccgatttagtgctttacggcac
ctcgacccccaaaaacttgattagggTgatggTtcacgtagtgggcccacCGcctgata
gacggTTTTcgcctttgacgtTggagTccacgttctttaaTtagtggactcttgTtcc
aaactggaacaactcaaccctatctcggTctattcttttgatttataagggattttg
ccgatttcggcctattggTtaaaaaatgagctgatttaacaaaaatTTaacgcgaattt
taacaaaatattaacgTttacaatttcaggtggcacttttcggggaaatgtgCGcggaa
ccctatttgTTTatttttctaaatacattcaaataTgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatgagTattcaacatttccg
TgtcGCCcttattccctTTTTTgCGgcattttgCcttctgTTTTTgctcaccagaaa
cgctggTgaaagTaaaagatgctgaagatcagTtgggTgcacgagTgggTtacatcgaa
ctggatctcaacagcggTaaagatcctTgagagTtttcgccccgaagaacgttttccaat
gatgagcacttttaagTtctgctatgtggcgcggTattatcccgTattgacgcccggc
aagagcaactcggTcGCCcatacactattctcagaatgactTggTtgagtactacca
gtcacagaaaagcatcttacggatggcatgacagTaaagaaTtatgcagTgctGCCat
aacatgagTgataacactgCGGCCaacttacttctgacaacgatcggaggaccgaagg
agctaaccgctTTTTTgcacaacatgggggatcatgtaactcgcctTgatcgtTgggaa
ccggagctgaatgaagccataccaaacgacgagcgtgacaccgatgctgtagcaat
ggcaacaacgtTgcGcaaactattaactggcgaactacttactctagcttcccggcaac
aattaatagactggatggaggcggataaagTtgcaggaccacttctgcgctcggccctt
ccggctggctggTttattgctgataaatctggagccggTgagcgtgggTctcgcggTat
cattgcagcactggggccagatggTaaagccctcccgtatcgtagTtatctacacgacgg
ggagTcaggcaactatggatgaacgaaatagacagatcgtgagataggtgCctcactg
attaagcattggTaaactgtcagaccaagTttactcatatatacttttagattgatttaa
acttcatTTTTtaatttaaaggatctaggtgaagatcTTTTTgataatctcatgacca
aaatcccttaacgtgagTtttcgTtccactgagcgtcagacccgtagaaaagatcaaa
ggatcttcttgagatcctTTTTTctgCGcgtaatctgctgctTgcaacaaaaaaacc
accgctaccagcggTggTttgTttgCCggatcaagagctaccaactctTTTTTccgaagg
taactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagtta
ggccaccacttcaagaactctgtagcaccgcctacatacctcgcTctgctaactcctgtt
accagTggctgctGCCagTggcgataagTcgtgtcttaccgggTtgactcaagacgat
agTtaccggataaggcgcagcggTcgggctgaacgggggTtctgTgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgc
cacgcttcccgaagggagaaaggcggacaggtatccggTaaagcggcagggTcggaaacg
gagagcgcacgagggagcttccaggggaaacgcctggtatctttatagTcctgtcggg
tttcGCCacctctgactTgagcgtcgattTTTTTgtgatgctcgtcaggggggCGgagcct
atggaaaaacGCCagcaacCGggcTTTTTaccgTtccTggcctTTTTTgctggcctTTTTg
ctcacatgttcttctcgttatccctgattctgtggataaccgTattaccgCcttt
gagTgagctgataccgctcGCCcagccgaacgaccgagcgcagcagTcagTgagcga
ggaagcggaaagagcgcctgatgcggTattttctccttacgcatctgtgcggTatttcac
accgCagaccagccgcgTaaactggcaaaatcggTtacggTtgagTaaTaaatggatgc
cctgcgTaaagcggTgtgggCGgacaataaagTctTaaactgaacaaaatagatctaaa
ctatgacaataaagTctTaaactagacagaatagTtgTaaactgaaatcagTccagTta

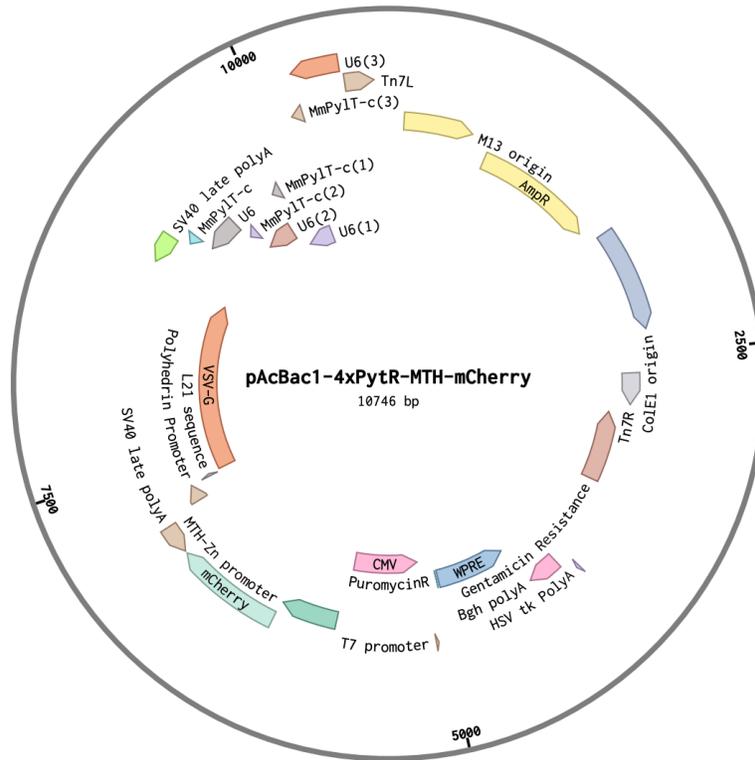
tgctgtgaaaaagcatactggactttttgttatggctaagcaaactcttcattttctga
agtgc aaattgccgctcgtattaaagaggggctggccaagggcatggtaaagactata
ttcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggctt
gaacgaattgttaggtggcgggtacttgggtcgatatcaaagtgcacacttcttcccgt
atgcccaactttgtatagagagccactgcgggatcgtcaccgtaaatctgcttgcaagta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtg
caatgccctgcctccgggtgctcgcgggagactgcgagatcatagatatagatctcacta
cgcggtgctcaaacctgggcagaacgtaagccgcgagagcgcgaacaaccgcttcttg
gtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccaggtaatcgg
agtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatca
agagcagcccgcattgattgacttggcaggccgagcctacatgtgcgaatgatgcc
catacttgagccacctaactttgttttagggcactgccctgctgcgtaacatcgttgc
tgctgcgtaacatcgttgcgctccataacatcaaacatcgaccacggcgtaacgcgc
ttgctgcttggatgcccaggcatagactgtacaaaaaacagtcataacaagccatga
aaaccgccactgcgcggttaccaccgctgcgttcgggtcaaggttctggaccagttgcgt
gagcgcatacgcctacttgcattacagtttacgaaccgaacaggcttatgtcaactgggt
tcgtgccttcatccgtttccacgggtgtgcgctaccgcgcaaccttgggcagcagcgaag
tcgaggcatttctgtcctggctggcgaacgagcgcgaaggttccggctcaccgcacatcgt
caggcattggcggccttgcgttcttctacggcaaggtgctgtgcacggatctgcctg
gcttcaggagatcggtagacctcggccgctcgcggcgttgcgggtgggtgctgaccocgg
atgaagtggttcgcacctcgggttttctggaaggcgcagcatcgtttgttcgccaggac
tctagctatagttctagtggttggctacgtaccgtagtggctatggcagggttgcgc
ttaatgcgcgctacagggcgcgtggggataccocctagagcccagctgggtcttctcc
gcctcagaagccatagagcccaccgcacccccagcatgcttgcattgtcttcccacatc
ctcccccttgcgtcctgccccaccocccccagaatagaatgacacctactcagac
aatgcgatgcaatttccctattttattaggaaaggacagtgagggtggcacctccagg
gtcaaggaaggcagggggaggggcaacaacagatggctggcaactagaaggcacagt
cgaggctgatcagcgggtttaaacgggccctctagactcgagttaaagtcgacgcgggg
aggcggcccaaaggagatccgactcgtctgagggcgaaggcgaagacgcggaagaggc
cgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaagggac
gtagcagaaggacgtcccgcgcagaatccagggtggcaacacaggcgcgagcgaaggaa
aggacgatgatttcccgcacaacaccacggaattgtcagtgcccaacagccgagcccct
gtccagcagcgggcaaggcaggcggcgcgatgagttccgcgctggcaatagggaggggaa
agcgaagtcocggaaaggagctgacaggtgggtggcaatgcccacaccagtggggggtg
cgtcagcaaacacagtgacaccacgccacgttgcctgacaacgggcccacaactcctca
taaagagacagcaaccaggatttatacaaggaggagaaaatgaaagccatacgggaagc
aatagcatgatacaaaggcattaaagcagcgtatccacatagcgtaaaaggagcaacat
agttaagaataccagtcacttccacaattttgtaatccagagggttgattgtcgcact
taacgcgttgaattccaccacactggactagtggtatccgagctcggtagcaagcttaag
tttaaacgctagccagcttgggtctcccctatagtgagtcgtattaatttcgataagcca
gtaagcagtggttctctagttagccagagagctctgcttataatagacctccaccgta
cacgcctaccgccatttgcgtcaatggggcggagttgttacgacattttggaaagtcc
cgttgattttgggtgccaaaacaaactcccattgacgtcaatgggggtggagacttgaaa
tcccgcgtgagtc aaocgctatccaogcccattgatgtactgcaaaaaccgcacacca
tggtaatagcgcgatgactaatacgtatagatgtactgccaagtaggaaagtcccataaggct
atgtactgggcataatgccaggcgggcccatttaccgctattgacgtcaatagggggct
acttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaaatagtcaca
ccattgacgtcaatggaaagtcctattggcgttactatgggaaacatacgtcattatt

gacgtcaatgggcgggggtcggtgggcggtcagccagggcgggccatcttacogtaagtta
tgtaacgcggaactccatatatgggctatgaactaatgaccccgtaattgattactatt
aataactagtcaataatcaatgtcaacgcgtatatctggcccgtacatcgcggaagcagc
gcaaaacgggatcctgcaggtatcttgcggcgcggtccgtataactccggaatattaatag
atcatggagataattaaaatgataaccatctcgcaaataaataagtatcttactgtttt
cgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccac
catcgggcggaactcctaaaaaacggccaccatgaagtgccttttgtacttagccttt
ttattcattgggggtgaattgcaagttcaccatagttttccacacaaccaaaaaggaaa
ctggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaattggc
ataatgacttaataggcacagccttacaagtcaaaatgccaagagtcacaaggctatt
caagcagacgggttgatgtgtcatgcttccaaatgggtcactacttgtgatttccgctg
gtatggaccgaagtataaacacattccatccgatccttcaactccatctgtagaacaat
gcaaggaaagcattgaacaaacgaaacaaggaacttggtgaatccaggcttccctct
caaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtccagggtgactcc
tcaccatgtgctgggttgatgaatacacaggagaatgggttgattcacagttcatcaacg
gaaaatgcagcaattacatatgcccactgtccataactctacaacctggcattctgac
tataagggtcaaagggtatgtgattctaacctcatttccatggacatcaccttcttctc
agaggacggagagctatcatccctgggaaggaggggcacagggttcagaagtaactact
ttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtc
agactcccatcagggtgtctgggtcgagatgggtgataaggatctctttgctgcagccag
attccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggtg
taagtctaattcaggacgttgagaggatcttgattattccctctgccaagaaacctgg
agcaaaatcagagcgggtcttccaatctctccagtggtatctcagctatcttgctcctaa
aaaccagggaaccggctctgcttccaccataatcaatggtaccctaaaataactttgaga
ccagatacatcagagtcgatattgctgctccaatcctctcaagaatggtcggaatgatc
agtggaaactaccacagaaagggaactgtgggatgactgggcaccatataagacgtgga
aattggaccaatggagttctgaggaccagttcaggatataagtttcccttatacatga
ttggacatggtatggtgactccgatcttcatcttagctcaaaggctcagggttccgaa
catcctcacattcaagacgctgcttgcgaacttctgatgatgagagtttattttttgg
tgatactgggtatccaaaaatccaatcgagcttgtagaagggtgggtcagtagttgga
aaagctctattgcctctttttctttatcatagggttaatcattggactattcttggt
ctccgagttggtatccatctttgcattaaattaagcacaccaagaaaagacagattta
tacagacatagagatgaaccgacttggaagtgataaggccaggccggccaagcttgtc
gagaagtactagaggatcataatcagccataccacattttagtagaggtttacttgctt
aaaaacctcccacacctcccctgaacctgaaacataaaatgaatgcaattggtggtg
ttaacttgtttattgcagcttataatggttacaataaagcaatagcatcacaatttc
acaataaagcatttttttcaactgcaattctagttgtggtttgtccaaactcatcaatgt
atcttatcatgtctggatctgatcactgcttgagccttagcaaaaaacgaggaccccggg
aatctaaccggctgaaccggatttagagtcogttcgatctacatgatcagggtcccgggt
gtttogtcccttccacaagatatataaagccaagaaatcgaaatacttcaagttacgg
taagcatatgatagtcattttaaaacataattttaaaactgcaactaccgaagaat
tattactttotacgtcaogtattttgtactaatatctttgtggttacagtc aaattaat
tetaattatctctotaacagccttgatogtatatgcaaatatgaaggaaatcatgggaa
ataggccctcttctgcccgaggtacccttagcaaaaaacgaggaccccgggaatctaa
cccggctgaaccgatttagagtcogttcgatctacatgatcagggtccccgggtgttctg
ccttccacaagatatataaagccaagaaatcgaaatacttcaagttacggtaagcat
atgatagtc aattttaaaacataattttaaaactgcaactaccgaagaattattact
ttotacgtcaogtattttgtactaatatctttgtggttacagtc aaattaattotaatt

atctcttaacagccttgatcgtatatgcaaatatgaaggaatcatgggaaataggcc
ctcttctgcccgaggtaccctagcaaaaaacgaggaccccggaatctaaccggct
gaacggatttagagtccgttcgatctacatgatcaggtccccgggtgtttcgctcttcc
acaagatatataagccaagaaatogaaatactttcaagttacggtaagcatatgatag
tccattttaaaacataattttaaaactgcaaaactaccaagaattattactttotaag
tcacgtatgtgactaataatctttgtgtttacagtcaaattaattctaattatctctc
taacagccttgatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttcc
tgcccgaggtaccctagcaaaaaacgaggaccccggaatctaaccggctgaacgga
tttagagtccgttcgatctacatgatcaggtccccgggtgtttcgctcttccacaagat
atataaagccaagaaatogaaatactttcaagttacggtaagcatatgatagtcattt
taaacataattttaaaactgcaaaactaccaagaattattactttctaogtcaogta
ttttgactaataatctttgtgtttacagtcaaattaattctaattatctctctaacagc
cttgatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttctgcccga
ggtagccttaggagatccgaaccagataagtgaaatctagttccaaactatgttgcatt
ttttaattttcgtattagcttacgaogctacaccagttcccatctatgttgcactct
tccetaaataatccttaaaaactccatttccaccctcccagttccaactatgttgc
cgcccacagcggggcatttttcttctggttatgttttaatacaaacatcctgccaactc
catgtgacaaaccgtcatcttccgctactt

The sequence for the wild type PytR is:
ggaaacctgatcatgtagatcgaacggactctaaatccgttcagccgggttagattccc
ggggttccg

pAcBac-4xPytR-MTH-mCherry



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
T7 promoter	text
CMV promoter	text
MTH-Zn promoter	text
mCherry	text
SV40 late polyA	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
MmPytR	text
U6	text

Tn7L	text
------	------

Sequence of 4xPylR-A2 variant

ttctctgtcacagaatgaaaatTTTTctgtcatctcttcgTTattaatgTTTgtaattg
actgaatatcaacgcttatttgcagcctgaatggcgaatgggacgcgccctgtagcggc
gcattaagcgcggcggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgc
cctagcgcgccgctcctttcgttttcttcccttccctttctcgcacgttcgcggctttc
cccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcac
ctcgaccccaaaaaacttgattagggtgatgggtcacgtagtggccatcgccctgata
gacggTTTTcgcctttgacgTTggagTccacgTtctttaatagtggactcttgTtcc
aaactggaacaacactcaaccctatctcggTctattcttttgatttataagggattttg
ccgatttccggcctattggTtaaaaaatgagctgatttaacaaaaatttaacgcgaattt
taacaaaaatattaacgTTtacaatttcaggTggcacttttcggggaaatgtgcgcggaa
cccctatttgTTtatttttctaaatacattcaaatatgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatgagTattcaacatttccg
tgTgcgcccttattccctTTTTtgcggcattttgccttctgTTTTgctcaccagaaa
cgctggtgaaagTaaaagatgctgaagatcagTtgggtgcacgagTgggttacatcgaa
ctggatctcaacagcggtaagatccttgagagTTTTcgcCCcgaagaacgTTTTccaat
gatgagcactTTtaagTtctgctatgtggcgcggTattatccgTattgacgccgggc
aagagcaactcgTcgcgcatacactattctcagaatgactTggtTgagTactcacca
gtcacagaaaagcatcttacggatggcatgacagTaaagaaattatgcagTgctgccat
aacatgagTgataacactgcccacaacttacttctgacaacgatcggaggaccgaagg
agctaaccgctTTTTtgcaaacatgggggatcatgtaactcgcctTgatcgtTgggaa
ccggagctgaatgaagccataccaaacgacgagcgtgacaccagatgcctgtagcaat
ggcaacaacgTtgcgcaaactattaactggcgaactacttactctagcttcccggcaac
aattaatagactggatggaggcggataaaagTtgcaggaccacttctgcgctcggccctt
ccggctggctggTttattgctgataaatctggagccggTgagcgtgggtctcgcggTat
cattgcagcactggggccagatggtaagccctcccgtatcgtagTtatctacacgacgg
ggagTcaggcaactatggatgaacgaaatagacagatcgtgagataggtgcctcactg
attaagcattggtaactgtcagaccaagTttactcatatatactttagattgatttaa
acttcatTTTTaatttaaaaggatctaggTgaagatcctTTTTgataatctcatgacca
aaatcccttaacgtgagTTTTcgttccactgagcgtcagaccccgtagaaaagatcaaa
ggatcttcttgagatcctTTTTtctgcgcgtaatctgctgctTgcaacaaaaaaacc
accgctaccagcggTggTttgTttgcccggatcaagagctaccaactctTTTTccgaagg
taactggctcagcagagcgcagataccaaatactgtccttctagtgtagccgtagtta
ggccaccacttcaagaactctgtagcaccgcctacatacctcgcTctgctaatcctgtt
accagTggctgctgccagTggcgataagTcgtgtcttaccgggtTggactcaagacgat
agTtaccggataagggcgcagcggTcgggctgaacgggggTtctgTgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgc
cacgcttcccgaagggagaaagggcgacaggtatccggtaagcggcagggTcggaaacag
gagagcgcacgagggagcttccaggggaaacgcctggTatctttatagtccTgtcggg
tttcgccacctctgactTgagcgtcgatTTTTgtgatgctcgtcaggggggcggagcct
atggaaaaacgccagcaacgcggcTTTTtacggTtccTggcTTTTgctggcTTTTg
ctcacatgTtcttctcgttatcccctgattctgtggataaccgTattaccgccttt
gagTgagctgataaccgctcgcgcagccgaacgaccgagcgcagcagTcagTgagcga
ggaagcggaaagagcgcctgatgcggTattttctccttacgcTctgtgcggTatttca
accgagaccagccgcgtaacctggcaaaatcggTtacggTtgagTaaatggatgc

cctgcgtaagcgggtgtgggoggacaataaagtcttaaactgaacaaaatagatctaaa
ctatgacaataaagtcttaaactagacagaatagttgtaaactgaaatcagtcagtta
tgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcattttctga
agtgcaaattgccgctcgtattaaagaggggctggccaagggcatggtaaagactata
ttcgcggcgcttgtagacaatttaccgaacaactccgcggccgggaagccgatctcggct
gaacgaattgttaggtggcggacttgggtcgatatcaaagtgcacacttcttcccgt
atgcccaactttgtatagagagccactgcgggatcgtcacccgtaatctgcttgacgta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcgggtg
caatgccctgcctccgggtgctcgccggagactgcgagatcatagatatagatctcacta
cgcggctgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttg
gtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccagagtaatcgg
agtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatca
agagcagcccgcattgattgacttggcagggccgagcctacatgtgcaatgatgcc
catacttgagccacctaactttgttttagggcgactgcctgctgcgtaacatcgttgc
tgcgtaacatcgttgctgctccataacatcaaacatcgaccacggcgtaacgcgc
ttgctgcttggatgcccgaggcatagactgtacaaaaaacagtcataacaagccatga
aaaccgccactgcgcgcttaccaccgctgcgctcggtaacaggttctggaccagttgct
gagcgcatacgtacttgcattacagtttacgaaccgaacaggcttatgtcaactgggt
tcgtgccttcatccgtttccacgggtgtgcgtcaccggcaaccttgggcagcagcgaag
tcgaggcatttctgtcctggctggcgaacgagcgaaggtttcggctctccacgcacgt
caggcattggcggccttctgttcttctacggcaaggtgctgtgcacggatctgcctg
gcttcaggagatcggtagacctcggccgctcgcggcgttgcgggtggtgctgacccgg
atgaagtggttcgcacctcgggtttctggaaggcgagcatcgtttgctcggccaggac
tctagctatagttctagtggttggctacgtaccgtagtggctatggcagggcttgc
ttaatgcgccgctacagggcgctggggataccccctagagcccagctggttctttcc
gcctcagaagccatagagcccaccgcatccccagcatgctgctattgtcttcccaatc
ctcccccttgcgtgctcctgccccaccccacccccagaatagaatgacacctactcagac
aatgcgatgcaatttctcattttattaggaaggacagtgagggtggcaccttccagg
gtcaaggaaggcaggggggaggggcaacaacagatggctggcaactagaaggcacagt
cgaggctgatcagcgggtttaaaggccctctagactcgagttaaagtcgacgcgggg
aggcggcccaaaggagatccgactcgtctgagggcgaaggcgaagacgcggaagaggc
cgcagagccggcagcaggccgcgggaagggaaggtccgctggattgagggccgaaggac
gtagcagaaggacgtcccgcgcagaatccaggtggcaacacaggcgcgagccaaggaa
aggacgatgatttccccgacaacaccacggaattgtcagtgcccaacagccgagcccc
gtccagcagcgggcaaggcaggcggcgatgagttccgcgctggcaataggagggggaa
agcgaagtcccggaaaggagctgacaggtgggtggcaatgccccaccagtggggggtg
cgtcagcaaacacagtgacaccacgccacgttgctgacaacgggcccacaactcctca
taaagagacagcaaccaggatttatacaaggaggagaaaatgaaagccatacgggaagc
aatagcatgatacaaaggcattaaagcagcgtatccacatagcgtaaaaggagcaacat
agttaagaataaccagtcattcttccacaattttgtaatccagagggttgattgtcgact
taacgcgttgaattccaccacactggactagtgatccgagctcggtagcaagcttaag
tttaaacgctagccagcttgggtctccctatagtgagtcgtattaatttcgataagcca
gtaagcagtggggtctctagttagccagagagctctgcttataatagacctcccacgta
cacgcctaccgccatttgcgtcaatggggoggagttgttacgacattttggaaagtc
cgttgattttgggtgccaaaacaaactcccattgacgtcaatgggggtggagacttggaa
tccccgtgagtcaaaccgctatccacgccattgatgtactgcaaaaaccgcatcaca
tggtaatagcagatgactaatacgtagatgtactgccaagtaggaaagtcccataaggtc
atgtaactgggcataatgccaggcggggccatttaccgctcattgacgtcaataggggcgt

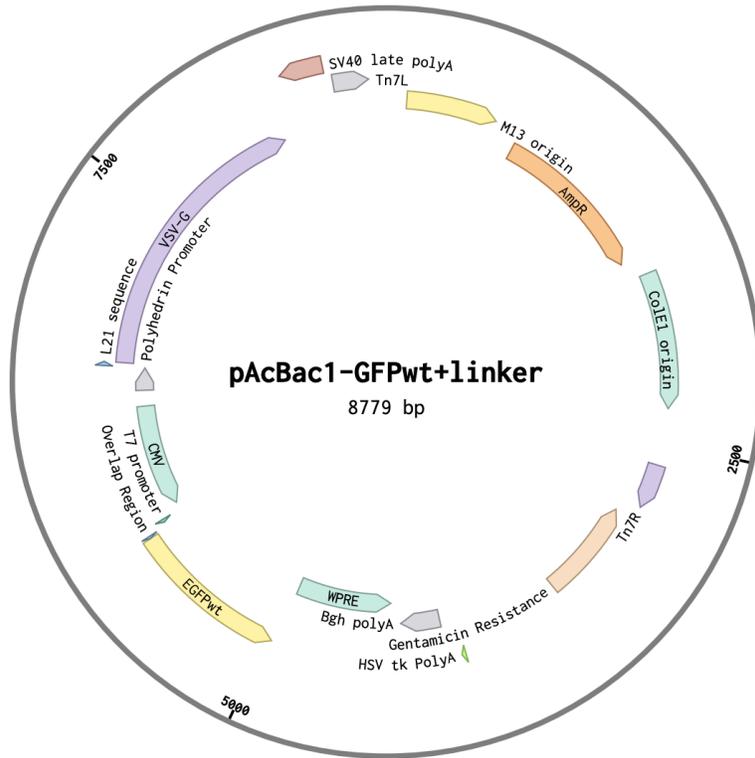
acttggcatatgatacacttgatgtactgccaagtgggcagtttacgtaaatagtcca
cccattgacgtcaatggaaagtccctattggcgttactatgggaacatacgtcattatt
gacgtcaatgggcgggggctcgttggcggtcagccaggcgggccatttacgtaagtta
tgtaacgcggaactccatataatgggctatgaactaatgaccccgtaattgattactatt
aataactagtcaataatcaatgtcaacgcgtatatctggcccgtacatcgccaagcagc
gcaaaacggatcctgcaggaccatggctcgagatcccgggtgatcaagtcttcgtcggg
cccgattattcataccgtcccaccatcgggcgcgatctcgagctcaagcttcgggaa
ctgtcccgggtccgagcgcagatcctttgcgctcagtcctggtggccgtgcccagagggg
gtgactcagagcggaggcgtgtgcaggcggatcccacgtgcaggaagagggacgcttg
cagctcgcgggacgtgtgcagctcgcggggcgatgcagggtggtgggcaacctgggcg
gacctttgcgctcgcocaaatccttccctctataaaggcagcgtctgcacgcttct
ctccatcacgccttctcctagcgtcgggtttctgtgtttcatcgtcctcagccgggtc
tggatctaccggccgcccaccatggtgagcaaggcggaggagataaacatggccatcatc
aaggagttcatgcgcttcaaggtgcacatggagggtccgtgaacggccacgagttcga
gatcgagggcgagggcgagggccgccctacgagggcaccagaccgccaagctgaagg
tgaccaaggggtggccccctgcccttcgctgggacatcctgtcccctcagttcatgtac
ggctccaaggcctacgtgaagcaccocgcccacatcccgcactacttgaagctgtcctt
ccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccg
tgaccagagactcctccctgcaagacggcgagttcatctacaaggtgaagctgcgggc
accaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctc
ctccgagcggatgtaccccgaggacggcgcctgaagggcgagatcaagcagaggctga
agctgaaggacggcggccactacgacgtgaggtcaagaccactacaaggccaagaag
cccgtgcagctgccggcgcctacaacgtcaacatcaagttggacatcacctcccaca
cgaggactacaccatcgtggaacagtacgaacgcgcggagggccgcccactccaccggc
gcatggacgagctgtacaagtaaTTTTacttgctttAAAAacctcccacacctcccc
tgaacctgaaacataaaatgaatgcaattgttggttaacttgtttattgcagcttat
aatggttacaataaagcaatagcatcacaatttcacaataaagcatttttttact
gcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgGCGgtatt
tgccggccggtccgtatactccggaatattaatagatcatggagataattaaatgat
aacatctcgcaataaataagtatTTTactgttttcgtaacagttttgtaataaaaa
acctataaatattccggattattcataccgtcccaccatcgggcgcaactcctaaaaa
accgccaccatgaagtgccttttgtacttagcctttttattcattgggggtgaattgca
gttcaccatagttttccacacaoccaaaaaggaaactggaaaaatgttcttotaatt
accattattgcccgtcaagctcagatttaaattggcataatgaottaataggaacagcc
ttacaagtcaaaatgcccagagtcacaaggctattcaagcagacgggttgatgtgtca
tgcttccaaatgggtcactacttgtgatttccgctggatggaccgaagtataaacac
attccatccgatccttcaactccatctgtagaacaatgaaggaagcattgaacaaacg
aaacaaggaacttggctgaatccaggttccctcctcaagttgtggatgtgaactgt
gacggatgcogaagcagtgattgtccaggtgaectcctcaccatgtgctgggtgatgat
acacaggagaatgggttgatccacagttcatcaacggaaaatgcagcaattacatatgc
cccactgtccataactctacaacctggcattctgactataaggtcaagggtatgtga
ttctaacctcatttccatggacatcacttctctcagaggacggagagctatcatccc
tgggaaaggagggcacagggttcagaagtaactaotttgottatgaaactggaggcaag
gocctgcaaaatgcaatactgcaagcattggggagtcagactcccacaggtgtctgggt
cgagatggctgataaggatctctttgctgcagccagatccctgaatgcccagaagggt
caagtatctctgctccatctcagacctcagtggtatgtaagtctaatcaggacgttgag
aggatcttgattattccctctgccaagaaacctggagcaaaatcagagcgggtcttcc
aatctctccagtggtctcagctatcttgotcctaaaaaaccaggaaccggctcctgctt

tcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatatt
gctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagga
actgtgggatgactggcaccatataagacgtggaaattggaccaatggagttctga
ggaccagttcaggatataagtttctttatacatgattggacatggtatgttggactcc
gatcttcatotttagotcaaaggotoagggtttogaacatctcacaattoagaagotgc
ttcgcaacttctgatgatgagagtttatttttgggtgatactgggctatccaaaaatc
caatcgagcttgtagaaggttggttcagtagttggaaaagctctattgectctttttc
tttatcatagggtaatcattggaactattcttggttctccgagttggatccatctttg
cattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaacggac
ttggaaagtgataaggccaggccggccaagcttgtcgagaagtactagaggatcataat
cagccataccacattttagtagaggttttacttgctttaaaaaacctcccacacctcccc
tgaacctgaaacataaaatgaatgcaattggttggtaacttgtttattgcagcttat
aatggttacaataaagcaatagcatcacaatttcacaaataaagcatttttttctact
gcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgat
cactgcttgagcctagcaaaaaacgaggacccccgggaatctaaccggctgaacggatt
tagagtccgttcgatctacatgatcaggtccccgggtgtttcgtcctttccacaagatat
ataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtccatttta
aacataattttaaaactgcaaactaccaagaaattattactttctacgtcacgtatt
ttgtactaatatctttgtgtttacagtcaaattaattctaattatctcttaacagcct
tgtatcgtatatgcaaatatgaaggaatcatgggaataggccctcttctgcccgagg
taccctagcaaaaaacgaggacccccgggaatctaaccggctgaacggatttagagtc
cggttcgatctacatgatcaggtccccgggtgtttcgtcctttccacaagatatataaagc
caagaaatcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacata
attttaaaactgcaaactaccaagaaattattactttctacgtcacgtattttgtact
aatatctttgtgtttacagtcaaattaattctaattatctcttaacagccttgatcgc
tatatgcaaatatgaaggaatcatgggaataggccctcttctgcccgaggtaccct
agcaaaaaacgaggacccccgggaatctaaccggctgaacggatttagagtcggttcga
tctacatgatcaggtccccgggtgtttcgtcctttccacaagatatataaagccaagaaa
tcgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataattttaa
aactgcaaactaccaagaaattattactttctacgtcacgtattttgtactaatatct
ttgtgtttacagtcaaattaattctaattatctcttaacagccttgatcgtatatgc
aaatatgaaggaatcatgggaataggccctcttctgcccgaggtaccctagcaaaa
aacgaggacccccgggaatctaaccggctgaacggatttagagtcggttcgatctacat
gatcaggtccccgggtgtttcgtcctttccacaagatatataaagccaagaaatcgaaat
actttcaagttacggtaagcatatgatagtccattttaaaacataattttaaaactgca
aactaccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgtt
tacagtcaaattaattctaattatctcttaacagccttgatcgtatatgcaaatatg
aaggaatcatgggaataggccctcttctgcccgaggtaccctaggagatccgaacc
agataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttac
gacgctacaccagttccatctattttgtcactcttccctaataatccttaaaaact
ccatttccaccctcccagttcccaactattttgtccgcccacagcggggcatttttct
tcctgttatgtttttaatcaaacatcctgccaaactccatgtgacaaaccgctcatcttcg
gctacttt

The sequence for the wild type PytR is:

ggaaacctgatcatgtagatcgaacggactctaaatccgttcagccgggttagattccc
ggggtttccg

pAcBac1-GFPwt+linker



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
EGFPwt	text
Overlap region	text
T7 promoter	text
CMV promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
Tn7L	text

Sequence

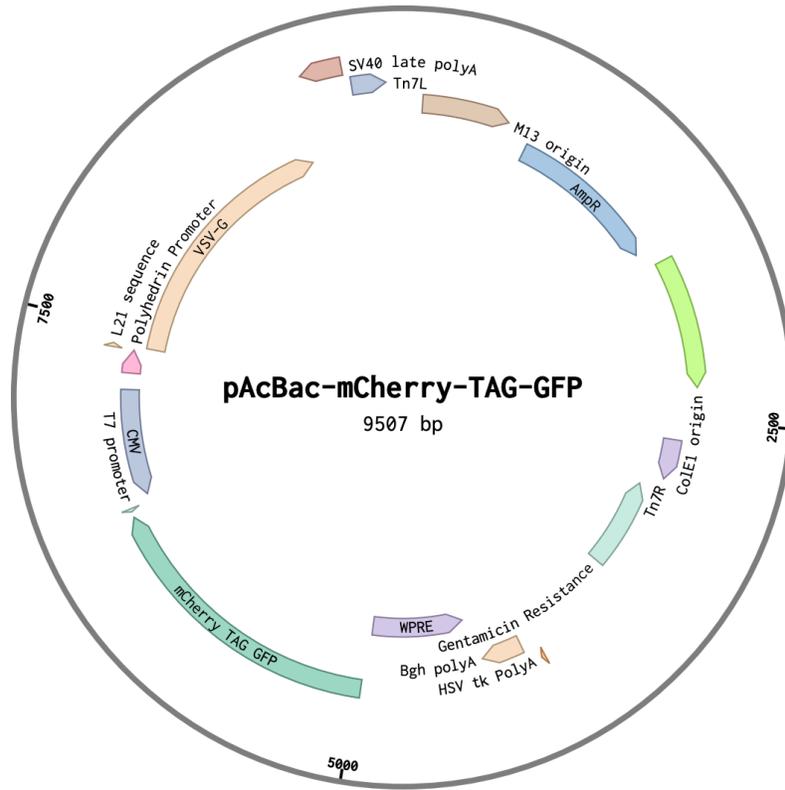
ttctctgtcacagaatgaaaatTTTTctgtcatctcttcgTTattaatgTTTgTaattg
actgaatatcaacgcttatttgcagcctgaatggcgaatgggacgCGcCctgtagCGgc
gcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgc
cctagcgcCCgctcctttcgtttcttcccttcttctcgcCCacgTtcgCCggctttc
cccgTcaagctctaaatcgggggctccctttagggTtccgatttagtgctttacggcac
ctcgacCCcaaaaaacttgattagggTgatggTtcacgtagTgggCCatcgccctgata
gacggTTTTtcgCCcttgacgTtgagTccacgTtctttaaTtagTggactcttgTtcc
aaactggaacaacactcaaccctatctcggTctattcttttgatttataagggattttg
ccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatTTaacgcgaattt
taacaaaaatattaacgTttacaatttcaggtggcacttttcggggaaatgtgcgcggaa
cccctatttgTTTatttttctaaatacattcaaataTgtatccgctcatgagacaataa
cccTgataaatgcttcaataatattgaaaaaggaagatgagTattcaacatttccg
TgtcgcCcttattccctTTTTtgCGgcattttgccttctgTTTTgctcaccagaaa
cgctggTgaaagTaaaagatgctgaagatcagTtgggTgcacgagTgggttacatcgaa
ctggatctcaacagcggTaagatccttgagagTtttcgCCCCgaagaacgTtttccaat
gatgagcacttttaaagTtctgctatgtggcgcggTattatcccgTattgacgCCgggc
aagagcaactcggTcgcgcatacactattctcagaatgactTggtTgagTactacca
gtcacagaaaagcatcttacggatggcatgacagTaaagaaTtatgcagTgctgCCat
aacatgagTgataaacactgcggccaacttacttctgacaacgatcggaggaccgaagg
agctaaccgctTTTTtgcaaacatgggggatcatgTaaTcgCcttgatcgTtgggaa
ccggagctgaatgaagccataccaaacgacgagcgtgacaccgatgCctgtagcaat
ggcaacaacgTtgcgcaaacTattaactggcgaactacttactctagcttcccggcaac
aattaatagactggatggaggcggataaagTtgcaggaccacttctgcgctcggccctt
ccggctggctggTttattgctgataaatctggagccggTgagcgtgggtctcgcggTat
cattgcagcactggggccagatggTaagccctcccgtatcgtagTtatctacacgacgg
ggagTcaggcaactatggatgaacgaaatagacagatcgctgagataggtgCctcactg
attaagcattggTaaactgTcagaccaagTttactcatatatactttagattgatttaa
acttcatTTTTtaatttaaaggatctaggtgaagatcTTTTtgataatctcatgacca
aaatcccttaacgTgagTtttcgTtccactgagcgtcagacCCcgtagaaaagatcaaa
ggatcttcttgagatcctTTTTctgcgcgTaatctgctgcttgcaaacaaaaaac
accgctaccagcggTggTttgTttgcCGgatcaagagctaccaactctTTTTccgaagg
taactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagTta
ggccaccacttcaagaactctgtagcaccgcctacatacctcgcTctgctaactcctgTt
accagTggctgctgccagTggcgataagTcgtgtcttaccgggttgactcaagacgat
agTtaccggataaggcgcagcggTcgggctgaacgggggTtctgTgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgc
cacgcttcccgaagggagaaaggcggacaggtatccgTaaagcggcagggTcggaaacg
gagagcgcacgagggagcttccaggggaaacgcctgTatctttatagTcctgTcggg
tttcgCCacctctgacttgagcgtcgatttttgTgatgctcgtcaggggggCGgagcct
atggaaaaacgCCagcaacCGggcTTTTtacggTtccTggcTTTTgctggcTTTTg
ctcacatgTtcttctcgtgTtatccctgattctgtggataaccgTattaccgcctt
gagTgagctgataccgctcgcgcagccgaacgaccgagcgcagcagTcagTgagcga
ggaagcggaaagagcgcctgatgcggTatttctccttacgcatctgtgcggTatttcac
accgCagaccagccgcgTaaactggcaaaatcggTtacggTtagTaaTaaatggatgc
cctgcgTaaagcggTgtgggCGgacaataaagTctTaaactgaacaaaatagatctaaa
ctatgacaataaagTctTaaactagacagaatagTtgTaaactgaaatcagTccagTta

tgctgtgaaaaagcatactggactttttgttatggctaagcaaactcttcattttctga
agtgcaaattgcccgtcgtattaaagaggggctggccaagggcatggtaaagactata
ttcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggctt
gaacgaattgttaggtggcgggtacttgggtcgatatcaaagtgcacacttcttcccgt
atgcccactttgtatagagagccactgcgggatcgtcaccgtaaatctgcttgcaagta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtg
caatgcctgcctccgggtgctcgccggagactgcgagatcatagatatagatctcacta
cgcggtgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttg
gtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccaggtaatcgg
agtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatca
agagcagcccgcattgattgacttggcagggccgagcctacatgtgcgaatgatgcc
catacttgagccacctaactttgttttagggcactgccctgctgcgtaacatcgttgc
tgctgcgtaacatcgttgcgctccataacatcaaacatcgaccacggcgtaacgcgc
ttgctgcttggatgcccaggcatagactgtacaaaaaacagtcataacaagccatga
aaaccgccactgcgcggttaccaccgctgcgttcgggtcaaggttctggaccagttgcgt
gagcgcatacgcctacttgcattacagttacgaaccgaacaggcttatgtcaactgggt
tcgtgccttcatccgtttccacgggtgtgcgtcaccgcgcaaccttgggcagcagcgaag
tcgaggcatttctgtcctggctggcgaacgagcgcgaaggttccggctcaccgcacatcgt
caggcattggcggccttgcgttcttctacggcaaggtgctgtgcacggatctgcctg
gcttcaggagatcggtagacctcggccgctcgcggcgttgcgggtgggtgctgaccocgg
atgaagtggttcgcacctcgggttttctggaaggcgcagcatcgtttgttcgccaggac
tctagctatagttctagtggttggcctacgtaccgcgtgtggctatggcaggccttgc
cttaatgcgccgctacagggcgcgtggggataccocctagagcccagctgggtcttctc
cgctcagaagccatagagcccaccgcaccccagcagcctgctattgtcttcccacat
cctcccccttgcgtgcctgccccaccccacccccagaatagaatgacacctactcaga
caatgcgatgcaatttctcattttattaggaaaggacagtgaggagtgccaccttccag
ggtaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaaggcacag
tcgaggctgatcagcgggtttaaaccgggcccctctagactcgcagttaaagtcgacgcggg
gaggcggcccaaaggagatccgactcgtctgagggcgaaggcgaagacgcgggaagagg
ccgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaaggga
cgtagcagaaggacgtcccgcgcagaatccaggtggcaacacaggcgcagcagccaagga
aaggacgatgatttccccgacaacaccacggaattgtcagtgcccaacagccgagcccc
tgtccagcagcgggcaaggcaggcggcgatgagttccgccgtggcaatagggaggggga
aagcgaaagtcccggaaaggagctgacaggtgggtggcaatgccccaacagtggggggt
gcgtcagcaaacacagtgacacaccgcacgcttgcctgacaacgggcccacaactcctc
ataaagagacagcaaccaggattatacaaggaggagaaaatgaaagccatacgggaag
caatagcatgatacaaaggcattaaagcagcgtatccacatagcgtaaaaggagcaaca
tagttaagaataccagtcattcttccacaaatttgtaatccagaggttgattgtcgcac
ttaacgcggttggaaattcctcattaatgggtgatgggtgatgaccggatgcatattca
gatcctcttctgagatgagttttgttcgaagggcccttgtacagctcgtccatgccg
agagtgatcccggcggcggtcacgaactccagcaggaccatgtgatcgcgcttctcgtt
ggggcttctgctcagggcggactgggtgctcaggtagtggttgcgggcagcagcacgg
ggccgctcgcgatgggggtgttctgctggtagtggtcggcgcagctgcacgctgccgtcc
tcgatgttgtggcggatcttgaagttcaccttgatgccgttcttctgcttgcggccat
gatatagacgttgtggctgttgtagttgactccagcttgtgcccaggatgttgcctt
cctccttgaagtcgatgcccttcagctcgcgtcgcgggttaccagggtgctgcctcgaac
ttcacctcggcgcgggtcttgtagttgccgctcgtccttgaagaagatgggtgcgctcctg
gacgtagccttcgggcgatggcggacttgaagaagtcgtgctgcttcatgtggctcgggg

agcggctgaagcactgcacgccgtaggtcaggggtggtcacgaggggtgggcccagggcagc
ggcagcttgccggtgggtgcagatgaacttcaggggtcagcttgccgtaggtggcatcgcc
ctcgccctcgccggacacgctgaacttggtggccggttacgctcgccggtccagctcgacca
ggatgggaccaccccggtgaacagctcctcgcccttgctcacggagccgcccctgctag
ccagcttgggtctccctatagtgagtcgtattaatttcgataagccagtaagcagtggtg
ttctctagtttagccagagagctctgcttatatagacctcccaccgtacacgcctaccgc
ccatttgctcaatggggcggagttgttacgacattttggaaagtcccgttgattttgg
tgccaaaacaaactcccattgacgtcaatgggggtggagacttgaaaatcccgtgagtc
aaaccgctatccagcccattgatgtactgccaaaaccgcatcaccatggtaatagcga
tgactaatacgtagatgtactgccaagtaggaaagtccataaggatcatgtactgggca
taatgccaggcgggcccatttaccgtcattgacgtcaatagggggcgtacttggcatatg
atacacttgatgtactgccaagtgggcagtttaccgtaaatagtcaccaccattgacgtc
aatggaaagtccctattggcgttactatgggaacatacgtcattattgacgtcaatggg
cgggggtcgttggcggtcagccaggcgggcccatttaccgtaagttatgtaacgcggaa
ctccatatatgggctatgaactaatgaccccgtaattgattactattaataactagtc
ataatcaatgtcaacgcgtatatctggcccgtacatcgcggaagcagcgcaaaaacggatc
ctgcaggtatttgcccgcgggtcccgtatactccggaatattaatagatcatggagata
attaaaatgataaccatctcgcaataaataagtatcttactgttttcgtaacagtttt
gtaataaaaaaacctataaatattccggattattcataccgtcccaccatcgggcccga
actcctaaaaaacggcccaccatgaagtgccttttgtacttagcctttttattcattggg
gtgaattgcaagttcaccatagttttccacacaacccaaaaaggaaactggaaaaatgt
tccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgacttaa
taggcacagccttacaagtcaaaatgccaagagtcacaaggctattcaagcagacggg
tggtgtgtcatgcttccaaatgggtcactacttgtgatttccgctgggatggaccgaa
gtatataacacattccatccgatccttccatcctgtagaacaatgcaaggaaagca
ttgaacaaacgaaacaaggaacttgggtgaatccaggcttccctcctcaaagttgtgga
tatgcaactgtgacggatgccgaagcagtgattgtccagggtgactcctcaccatgtgct
ggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagca
attacatatgcccactgtccataactctacaacctggcattctgactataaggtcaaa
gggctatgtgattctaacctcatttccatggacatcaccttcttctcagaggacggaga
gctatcatccctgggaaaggagggcacagggtcagaagtaactactttgcttatgaaa
ctggaggcaaggcctgcaaaaatgcaatactgcaagcattggggagtcagactcccatca
gggtgtctgggttcgagatgggtgataaggatctctttgctgcagccagattccctgaatg
cccagaagggtcaagtatctctgctccatctcagacctcagtggtgtaagtctaattc
aggacggtgagaggatcttggattattccctctgccaagaaacctggagcaaaaatcaga
gcccgtcttccaatctctccagtggtatctcagctatcttgcctctaaaaaccaggaac
cggctcctgctttcaccataatcaatgggtaccctaaaatactttgagaccagatacatca
gagtcgatattgctgctccaatcctctcaagaatgggtcggaatgatcagtggaactacc
acagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaattggaccocaa
tgaggttctgaggaccagttcaggatataagtttcccttatacatgattggacatggta
tggtggactccgatcttcatcttagctcaaaggctcaggtggtcgaacatcctcacatt
caagacgctgcttcgcaacttccctgatgatgagagtttattttttggtgatactgggct
atccaaaaatccaatcgagcttgtagaagggttggttcagtagttggaaaagctctattg
cctcttttttcttatcatagggtaaatcattggactattcttgggtctccogagttggt
atccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacataga
gatgaaccgacttgaaagtgataaggccaggccgggccaagcttgtcgagaagtagtag
aggatcataatcagccataccacattttagagaggttttacttgctttaaaaaacctccc
aacctcccctgaacctgaaacataaaaatgaatgcaattgttgttgaacttgttta

ttgcagcttataatgggttacaataaagcaatagcatcacaatttcacaaataaagca
ttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgt
ctg gatctgatcactgcttgagcctaggagatccgaaccagataagtgaaatctagttc
caactatcttgcatttttaattttcgtattagcttacgacgctacaccagttcca
totatcttgcactcttccctaaataatccttaaaaaactccatttccaccctcccagt
tcccaactatcttgcgcccacagcggggcatttttcttccctggttatgtttttaatca
aacatcctgccaactccatgtgacaaaccgctcatcttcggctacttt

pAcBac-mCherry-TAG-GFP



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
mCherry-TAG-GFP	text
T7 promoter	text
CMV promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
Tn7L	text

Sequence

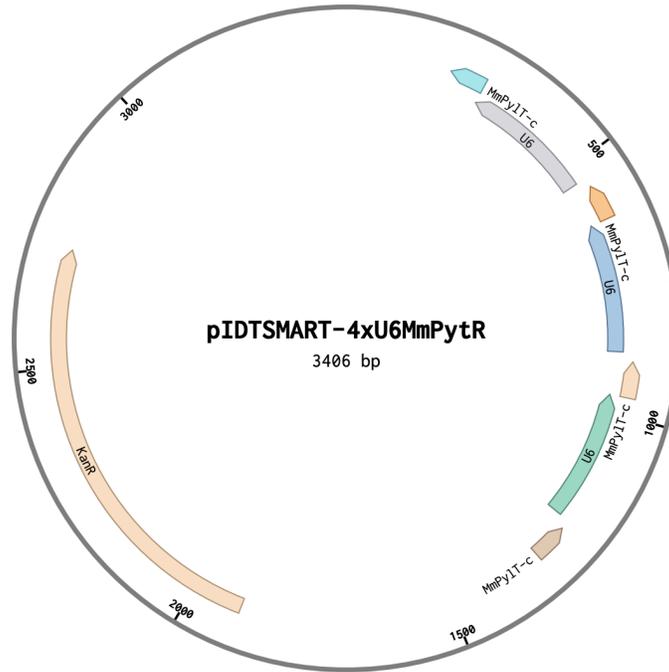
ttctctgtcacagaatgaaaatTTTTctgtcatctcttcgTTattaatgTTTgtaattg
actgaatatcaacgcttatttgcagcctgaatggcgaatgggacgCGcCCTgtagCGgc
gcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttGCCagcgc
cctagcgcCCgctcctttcgtttcttcccttcttctcGCCacgTtcGCCggctttc
ccgTcaagctctaaatcgggggctcccttagggTtcgatttagtgctttacggcac
ctcGacCCcaaaaaacttgattagggTgatggTtcacgtagTgggCCatcGCcctgata
gacggTTTTtcGCCctttgacgTtgagTccacgTtctttaaTtagTggactcttgTtcc
aaactggaacaacactcaaccctatctcggTctattcttttgatttataagggattttg
ccgatttcggcctattggTtaaaaaatgagctgatttaacaaaaatTTaacgcgaattt
taacaaaaatTTaacgTttacaatttcaggtggcacttttcggggaaatgtgcgcggaa
ccctatttgTTtatttttctaaatacattcaaataTgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagatagagTattcaacatttccg
TgtcGCCcttattccctTTTTtgCGgcattttgCcttctgTTTTgtcCaccagaaa
cgctggTgaaagTaaaagatgctgaagatcagTtgggTgcacgagTgggttacatcgaa
ctggatctcaacagcggtaagatccttgagagTtttcGccccgaagaacgTtttccaat
gatgagcacttttaaagTtctgctatgtggcgcggTattatccgTattgacgcCGggc
aagagcaactcggTcGCCcatacactattctcagaatgactTggTtgagTactacca
gtcacagaaaagcatcttacggatggcatgacagTaaagaaTtatgcagTgctGCCat
aacatgagTgataacactgcggccaacttacttctgacaacgatcggaggaccgaagg
agctaaccgctTTTTtgcaaacatgggggatcatgtaactcgccttgatcgtTgggaa
ccggagctgaatgaagccataccaaacgacgagcgtgacaccgatgCctgtagcaat
ggcaacaacgTtgcGcaaactattaactggogaactacttactctagcttcccggcaac
aattaatagactggatggaggcggataaagTtgcaggaccacttctgcgctcggccctt
ccggctggctggTttattgctgataaatctggagccggtgagcgtgggtctcgcggTat
cattgcagcactggggccagatggtaagccctcccgtatcgtagTtatctacacgacgg
ggagTcaggcaactatggatgaacgaaatagacagatcGctgagataggtgCctcactg
attaagcattggTaaactgtcagaccaagTttactcatatatactttagattgatttaa
acttcatTTTTtaatttaaaggatctaggtgaagatcTTTTtgataatctcatgacca
aaatcccttaacgTgagTtttcgTtccactgagcgtcagacCCcgtagaaaagatcaaa
ggatcttcttgagatcctTTTTctgCgcgtaatctgctgcttgcaaacaaaaaac
accgctaccagcggTggTttgTttgCCggatcaagagctaccaactctTTTTccgaagg
taactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagtta
ggccaccacttcaagaactctgtagcaccgcctacatacctcGctctgctaactcgtt
accagTggctgctGCCagTggcgataagTcgtgtcttaccgggttgactcaagacgat
agTtaccggataaggcgcagcggTcgggctgaacgggggTtcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgc
cacgcttcccgaagggagaaaggcggacaggtatccgTaaagcggcagggTcGgaacag
gagagcgcacgagggagcttccaggggaaacgcctgTatctttatagTcctgtcggg
tttcGCCacctctgacttgagcgtcgatttttTgtgatgctcgtcaggggggCGgagcct
atggaaaaacGCCagcaacCGggcTTTTtacggTtccTggccttttgctggccttttg
ctcacatgttcttctcgttatccctgattctgtggataaccgTattaccgccttt
gagTgagctgataccgctcGCCcagccgaacgaccgagcgcagcagTcagTgagcga
ggaagcGgaagagcgcctgatgcggTattttctccttacgcatctgtgcggTatttcac
accgCagaccagccgcgtaacctggcaaaatcggTtacggTtgagTaaTaatggatgc
cctgcgTaaagcggTgtgggCGgacaataaagTctTaaactgaacaaaatagatctaaa
ctatgacaataaagTctTaaactagacagaatagTtgTaaactgaaatcagTccagTta

tgctgtgaaaaagcatactggactttttgttatggctaagcaaaactcttcattttctga
agtgcaaattgcccgtcgtattaaagaggggctggccaagggcatggtaaagactata
ttcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggctt
gaacgaattgttaggtggcgggtacttgggtcgatatcaaagtgcacacttcttcccgt
atgcccaactttgtatagagagccactgcgggatcgtcaccgtaaatctgcttgcaagta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtg
caatgccctgcctccgggtgctcgccggagactgcgagatcatagatatagatctcacta
cgcggtgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttg
gtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccaggtaatcgg
agtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatca
agagcagcccgcattgattgacttggcagggccgagcctacatgtgcaatgatgcc
catacttgagccacctaactttgttttagggcactgccctgctgcgtaaacatcgttgc
tgctgcgtaaacatcgttgcgctccataacatcaaacatcgaccacggcgtaacgcgc
ttgctgcttggatgcccaggcatagactgtacaaaaaacagtcataacaagccatga
aaaccgccactgcgcggttaccaccgctgcgttcggtaaggttctggaccagttgcgt
gagcgcatacgcctacttgcattacagttacgaaccgaacaggcttatgtcaactgggt
tcgtgccttcatccgtttccacgggtgtgcgtcaccgcgcaaccttgggcagcagcgaag
tcgaggcatttctgtcctggctggcgaacgagcgaaggttccggctcaccgcacatcgt
caggcattggcggccttgcgttcttctacggcaaggtgctgtgcacggatctgcctg
gcttcaggagatcggtagacctcggccgctcgcggcgttgcgggtgggtgctgaccocgg
atgaagtggttcgcacctcgggttttctggaaggcagcagcgtttgttcgccaggac
tctagctatagttctagtggttggcctacgtaccgcgtgtggctatggcaggcttgcg
cttaatgcgccgctacagggcgcgtggggataccocctagagcccagctgggtcttctc
cgcctcagaagccatagagcccaccgcaccccagcagcagcagcagcagcagcagcagc
cctcccccttgcgtgcctgccccacccccacccccagaatagaatgacacactactcaga
caatgcgatgcaatttctcattttattaggaaaggacagtgaggagtgacaccttccag
ggtaaggaaggcacgggggaggggcaaacacagatggctggcaactagaaggcacag
tcgaggctgatcagcgggtttaaaccgggcccctctagactcagagttaaagtcgacgcggg
gagggcggccaaagggagatccgactcgtctgagggcgaaggcgaagacgcgggaagagg
ccgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaagggg
cgtagcagaaggacgtcccgcgcagaatccaggtggcaacacaggcgcagcagcagga
aaggacgatgatttccccgacaacaccacggaattgtcagtgcccaacagccgagcccc
tgtccagcagcgggcaaggcaggcggcgatgagttccgcctggcaatagggagggggg
aagcgaaagtcccggaaaggagctgacaggtgggtggcaatgccccaacagtggggggt
gcgtcagcaaacacagtgacacaccgcacgcttgcctgacaacgggcccacaactcctc
ataaagagacagcaaccaggattatacaaggaggagaaaatgaaagccatacgggaag
caatagcatgatacaaaggcattaaagcagcgtatccacatagcgtaaaaggagcaaca
tagttaagaataccagtcattctttcacaattttgtaatccagaggttgattgtcgac
ttaacgcggtgaattctcattaatgggtgatgggtgatgatgaccggatgcatattcaga
tcctcttctgagatgagttttgttcgaagggccccttgtacagctcgtccatgccgag
agtgatcccggcggcgggtcacgaactccagcaggaccatgtgatcgccttctcgttgg
ggcttttgcagggcggactgggtgctcaggtagtggttgcgggcagcagcagcgggg
ccgtcgcgatgggggtgttctgctggtagtggtcggcgagctgcacgctgccgtcctc
gatgttgtggcggatcttgaagttcaccttgatgcgcttcttctgcttgtcggccatga
tatagacgttgtggctgttgtagttgtactccagcttgtgccccaggatggtgccgtcc
tccttgaagtcgatgccctcagctcagatgcggttcaccagggtgtcgcctcgaactt
cacctcggcgcgggtcttgtagttgcgctcgtccttgaagaagatgggtgcgctcctgga
cgtagccttcgggcatggcggacttgaagaagtcgtgctgcttcatgtgggtcgggtag

cggtgaagcaactgcacgccgtaggtcaggggtggtcacgaggggtgggccagggcacggg
cagcttgccgggtggtgcagatgaacttcaggggtcagcttgccgtaggtggcatcgccct
cgccctcgccggacacgctgaacttgtggccgtttacgtcgccgtccagctcgaccagg
atgggcaccaccccggtgaacagctcctcgcccttgctcacgagggagccgcctaaga
gccgcccttgtagcagctcgtccatgcccgggtggagtggcggccctcggcgcgttcgt
actgttccacgatgggtgtagtccctcgttgtgggaggtgatgtccaacttgatggtgacg
ttgtaggcgccgggcagctgcacgggcttcttggccttgtaggtggcttgacctcagc
gtcgtagtggccgcccgtccttcagcttcagcctctgcttgatctcgcccttcagggcgc
cgtcctcggggtacatccgctcggaggaggcctcccagcccatggctcttcttctgcatt
acggggccgctcggaggggaagtgggtgccgcgcagcttcaccttgtagatgaactcgcc
gtcctgcagggaggagtcctgggtcacggtcaccacgccgcccgtcctcgaagtcatca
cgcgctcccacttgaagccctcggggaaggacagcttcaagtagtcggggatgctggcg
gggtgcttcacgtaggccttggagccgtacatgaactgaggggacaggatgtcccaggc
gaagggcagggggccacccttggtcacctcagcttggcgggtcgggtgccctcgtagg
ggcggccctcgcctcgcctcgcactcgaactcgtggccggttcacggagccctccatg
tgcaccttgaagcgcgatgaactccttgatgatggccatggtatcctcctcgccttgc
caccatgggtggcggcgttagccagcttgggtctccc**atagtgagtcgtatta**atctcg
ataagccagtaagcagtggggttctctagttagccaga**gagctctgcttatatagacctc**
ccaccgtacacgctaccgccatttgcgtcaatggggcggagttggttacgacattttg
gaaagtcccgttgattttgggtgccaaaacaaactccattgacgtcaatgggggtggaga
cttgaaatcccgtgagtcaaacgctatccagcccattgatgtactgcaaaaaccg
catcaccatggtaatagcgatgactaatacgtagatgtactgccaagtaggaaagtccc
ataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgtcaata
gggggogtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaa
atagtccaccattgacgtcaatggaaagtccctattggcgttactatgggaacatag
tcattattgacgtcaatgggcgggggtcgttgggcgggtcagccaggcgggccatttacc
gtaagttatgtaacgcggaactccatataatgggctatgaactaatgaccccgtaattga
ttactattaataactagtcaataatcaatgtcaacgcgtatatctggcccgtacatcgc
gaagcagcgcaaaacggatcctgcaggtatcttgcggccgcgggtccgtatactccggaat
attaatagatcatggagataaataaaatgataaccatctcgcaataaataagtatctt
actgttttcgtaacagttttgtaataaaaaaacctataaatatccggattatcatac
cgctcccaccatcgggcgcgaactcctaaaaaacgccaccatgaagtgccttttgtact
tagcctttttattcattgggggtgaattgcaagttcaccatagttttccacacaacca
aaaggaaactggaaaaatgttccttctaattaccattattgcccgtaagctcagattt
aaattggcataatgacttaataggcacagccttacaagtcaaaatgcccagagtcaca
aggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgtgat
ttccgctggtatggaccgaagtatataacacattccatccgatccttactccatctgt
agaacaatgcaaggaaagcattgaacaaacgaacaaggaacttggctgaatccaggct
tccctcctcaaagttgtggatgcaactgtgacggatgccgaagcagtgattgtccag
gtgactcctcaccatgtgctgggtgatgaatacacaggagaatgggttgattcacagtt
catcaacggaaaatgcagcaattacatatgcccactgtccataactctacaacctggc
attctgactataaggtcaagggtatgtgattctaacctcatttccatggacatcacc
ttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaag
taactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcatt
ggggagtcagactcccacaggtgtctggttcgagatggctgataaggatctctttgct
gcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctc
agtggatgtaagtctaattcaggacggtgagaggatcttggattattccctctgccaa
aacctggagcaaaatcagagcgggtcttccaatctctccagtggtatctcagctatctt

gctcctaaaaaccaggaaccggctcctgctttcaccataatcaatggtagccctaaaata
ctttgagaccagatacatcagagtcgatattgctgctccaatcctctcaagaatggtcg
gaatgatcagtggaactaccacagaaaggggaactgtgggatgactgggcacatgatgaa
gacgtggaattggaccaatggagttctgaggaccagttcaggatataagtttccttt
atacatgattggacatggtatgttggactccgatcttcatcttagctcaaaggctcagg
tgttcgaacatcctcacattcaagacgctgcttcgcaacttctgatgatgagagttta
tttttggtgatactgggctatccaaaaatccaatcgagcttgtagaagggttggttcag
tagttggaaaagctctattgcctcttttttctttatcatagggttaatcattggactat
tcttggttctccgagttggatccatctttgcattaaattaaagcacaccaagaaaaga
cagatttatacagacatagagatgaaccgacttggaagtgataaggccaggccggcca
agcttgtcagagaagtactagaggatcataatcagccataccacatttgtagaggtttta
cttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaaatgaatgcaat
tgttggtgtaacttgtttattgcagcttataatggttacaaataaagcaatagcatca
caaatttcacaaataaagcatttttttctactgcattctagttgtggttgtccaaactc
atcaatgtatcttatcatgtctgatctgatcactgcttgagcctaggagatccgaacc
agataagtgaaatctagttccaaactattttgtcatttttaattttogtattagettac
gacgctacaccagttccatctattttgtcactctccctaaataatccttaaaaact
ccatttccaccctccagttcccaactattttgtccgcccacagcggggcatttttct
tcctggtatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttcg
gctacttt

pIDTSMART-4xU6MmPytR



Sequence color-coding key

Feature	Color
MmPytR	text
U6	text
KanR	text

Sequence for HTS25 construct

```

cccgtagtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
gaagtgagccagtgagttgattgcagtcacgctggagctctgaggctcgtcctg
aatgatatgacgaccgcccggagggttgcgtttgagacgggcgacagatccagtcgcgctg
ctctcgtcgcgtagcaaaaaacgggcagcccgggaatctaaccggctgaacgga
tttagagtcggttcgatctacatgatcagccgcccgggtgtttcgtcctttccacaagat
atataaagccaagaaatcgaatactttcaagttacggtaagcatatgatagtccattt
taaaacataattttaaaactgcaaaactaccaagaaattattactttctacgtcacgta
tttgtactaataatctttgtgtttacagtcaaattaatttctaattatctctctaacagc
cttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttctctgcccga
ggtaccctagcaaaaaacgggcagcccgggaatctaaccggctgaacgatttagag
tccgttcgatctacatgatcagccgcccgggtgtttcgtcctttccacaagatatataa
gccaagaaatcgaatactttcaagttacggtaagcatatgatagtccattttaaaca
taattttaaaactgcaaaactaccaagaaattattactttctacgtcacgtattttgta
ctaataatctttgtgtttacagtcaaattaatttctaattatctctctaacagccttgat
cgtatatgcaaatatgaaggaatcatgggaaataggccctcttctctgcccgaggtacc
ctagcaaaaaacgggcagcccgggaatctaaccggctgaacgatttagagtcggttc
gatctacatgatcagccgcccgggtgtttcgtcctttccacaagatatataaagccaaga

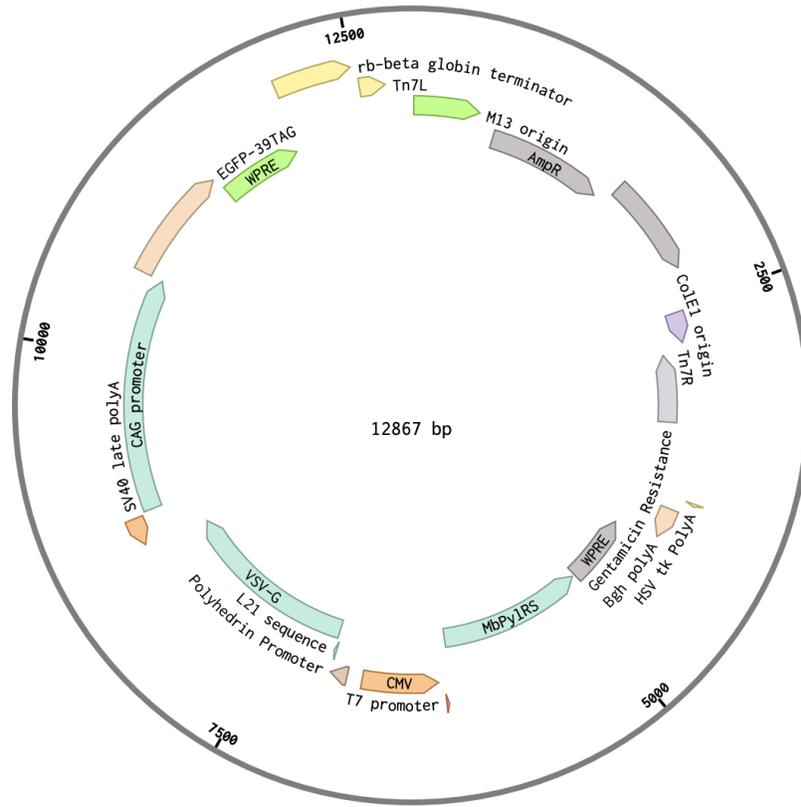
```

aatcgaataactttcaagttacggtaagcatatgatagtcatttttaaacataat
aaaactgcaaactaccaagaattattactttctacgtcacgtattttgtactaatat
ctttgtgtttacagtcaaattaattctaattatctcttaacagccttgatcgtatat
gcaaatatgaaggaatcatgggaataggccctcttctgcccgaggtacccttagcaa
aaa**cgggcagccgggaatctaaccggctgaacggatttagagtccttgcgatctac**
atgatcagccgccgggtgttctcgtcctttccacaagatatataagccaagaatcgaa
atactttcaagttacggtaagcatatgatagtcatttttaaacataattttaaaactg
caaactaccaagaattattactttctacgtcacgtattttgtactaatatctttgtg
tttacagtcaaattaattctaattatctcttaacagccttgatcgtatatgcaata
tgaaggaatcatgggaataggccctcttctgcccgaggtacccttaggggtgcgagc
ggatcgcgagcagtgatcactactggaccgcgagctgtgctgcgaccctgatcttac
ggcattatacgtatgatcgggtccacgatcagctagattatctagtcagcttgatgcat
agctgtttcctgaggctcaatactgaccatttaaatcatacctgacctccatagcagaa
agtcaaaagcctccgaccggaggcttttgacttgatcggcacgtaagaggttccaactt
tcaccataatgaaataagatcactaccgggogtattttttgagttatcgagattttcag
gagctaaggaagctaaa**atgagccatattcaacgggaaacgtcttgcttgaagccgcga**
ttaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgg
gcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcgccagagttgtttc
tgaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcaggctaac
tggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatga
tgcattggtactcaccactgcgatcccagggaaacagcattccaggtattagaagaat
atcctgattcaggtgaaaatattggtgatgcgctggcagtggtcctgcgccggttgc
tcgattcctggttgtaattgtcctttaacggcgatcgcgatttctcctcgcctcaggc
gcaatcacgaatgaataacgggttggtggtgcgagtgattttgatgacgagcgtaatg
gctggcctggtgaacaagtctggaaagaaatgcataaactcttgccattctcaccggat
tcagtcgtcactcatgggtgatttctcacttgataacctatttttgacgaggggaaat
aatagggtgtattgatgtggacgagtcggaatcgcagaccgataccaggatcttgcca
tcctatggaactgcctcggtagttttctccttcattacagaaacggctttttcaaaaa
tatgggtattgataatcctgatatgaataaattgcagtttcaactgatgctcagatgagt
tttctaattgaggacctaaatgtaatcacctggctcaccttcgggtgggcctttctgcgt
tgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgatgctca
agtacagaggtggcgaaacccgacaggactataaagataaccaggcgtttccccctggaag
ctccctcgtgcgctctcctggtccgaccctgccgcttaccggatacctgtccgctttc
tccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtg
taggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccagccgctg
cgccttatccggtaactatcgtcttgagtcacaaccggtaagacacgacttatcggcac
tggcagcagccactggtaacaggattagcagagcaggtatgtaggcgggtgctacagag
ttcttgaagtgggtggcctaactacggctacactagaagaacagatatttggtatctgcgc
tctgctgaagccagttacctcggaaaaagagttggtagctcttgatccggcaaacaaac
caccgctggtagcgggtggtttttttggttgaagcagcagattacgcgcagaaaaaaag
gatctcaagaagatcctttgattttctaccgaagaaggcca

The sequence for the wild type PytR is:

ggaaacctgatcatgtagatcgaacggactctaaatccggttcagccgggttagattccc
ggggtttccg

pAcBac-CMV-MbPylRS-WPRE-CAG-GFPtag39AvrIIdeI-WPRE



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
CMV promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
CAG promoter	text
EGFP-39tag	text
WPRE	text

rb-beta globin terminator	text
Tn7L	text

Sequence

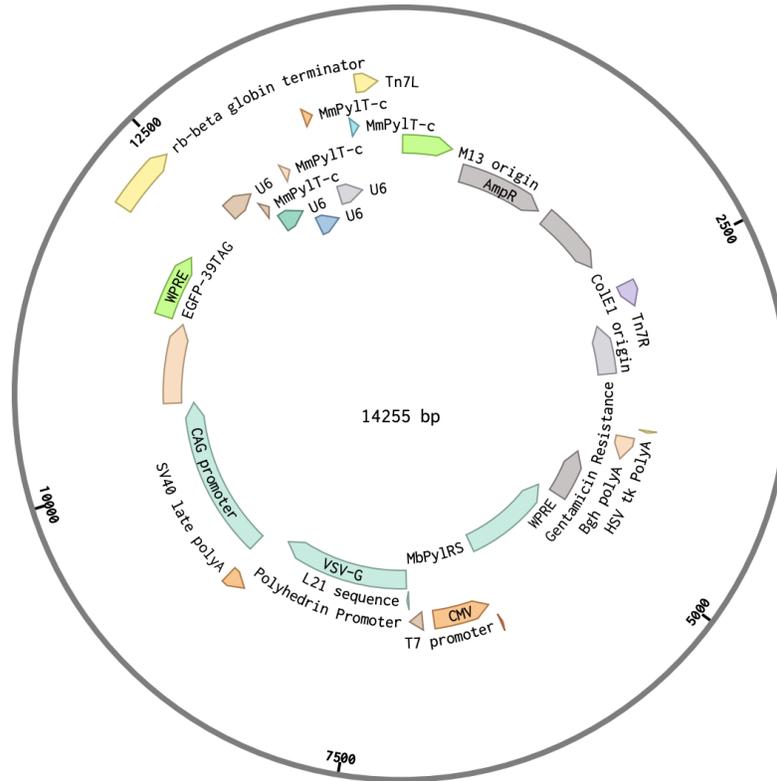
ctgaatggcgaatgggacgcgcctgtagcggcgcattaagcgcggcgggtgtgggtggt
tacgcgcagcgtgaccgctacacttgccagcgcctagcgcgcctcctttcgctttct
tcccttcccttctcgcacgttcgcggctttcccgtcaagctctaaatcgggggctc
ccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttg
agtccacgttctttaatagtggaactcttgttccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaggaagagtatgagtattcaacatttcggtgctgcccttattcccttttttgogg
cattttgccttctggtttttgctcaccagaaacgctggtgaaagtaaagatgctgaa
gatcagttgggtgcaagagtggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctat
gtggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgccgcatacac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtgataaacactgcggcca
actacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgacacaacatg
gggatcatgtaactcgccttgatcggtgggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaa
ctggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttcggctggctggtttattgctgataa
atctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggttaactgtcagacca
agtttactcatatataacttttagattgatttaaaacttcatttttaatttaaaggatct
aggatgaagatcctttttgataatctcatgaccaaatacccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttct
gcgcgtaatctgctgcttgcaaaaaaaaccaccgctaccagcgggtggtttgtttgc
cggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagata
caaataactgtccttctagtgtagcgttaggttagccaccactcaagaactctgtagc
accgcctacatacctcgtctgctaactctgttaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctc
ggctgaacggggggttcgtgcacacagcccagcttgagcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggtaagcggcagggctcggaaacaggagagcgcacgagggagcttccagg
gaaacgcctggatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg
atttttgtgatgctcgtcaggggggaggagcctatggaaaaacgccagcaacgcggcct
tttacggttcctggccttttgctggccttttgctcacatggtctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgcgcag
ccgaacgaccgagcgcagcagtgagtgagcaggaagcggaaagagcgcctgatgcgg
atcttctccttacgcacatctgtgcggtatctcacaccgcagaccagccgcgtaacctggc
aaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgtgggaggaca

ataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactaga
cagaatagttgtaaactgaaatcagtcagttatgctgtgaaaaagcactactggacttt
tggtatggctaaagcaactcttcattttctgaagtgcaaattgcccgctgtattaaag
agggcggtggccaagggcatggtaagactatattcgcgcggtgtgacaatttaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtactt
gggtcgatatcaaagtgcatacacttcttcccgtatgcccaactttgtatagagagccac
tgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccgaagcgcgttgg
cctcatgcttgaggagattgatgagcgcgggtggcaatgccctgcctccgggtgctcgcg
gagactgcgagatcatagatatagatctcactacgcggctgctcaaacctgggcagaac
gtaagccgcgagagcgcacaaccgcttcttggtcgaaggcagcaagcgcgatgaatg
tcttactacggagcaagttcccggaggtaatcggagtcggctgatggtgggagtaggtg
gctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcattgatttgacttg
gtcagggccgagcctacatgtgcaatgatgccatacttgagccacctaaactttgttt
tagggcgactgccctgctgcgtaacatcgttgcgctgcgtaacatcgttgcgctcca
taacatcaaacatcgaccacggcgtaacgcgcttgcgcttggatgcccgaggcatag
actgtacaaaaaacagtcataacaagccatgaaaaccgccactgcgcggttaccaccg
ctgcgctcgggtcaaggttctggaccagttgcgctgagcgcatacgcctacttgcaatacag
tttacgaaccgaacaggcttatgtcaactgggttcgctgccttcatccggttccacgggtg
tgcgtaaccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcg
aacgagcgaaggttccggtctccacgcatacgcagcattggcggccttgcgcttctt
ctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggc
cgctcgcggcgcttgcgggtgggtgctgaccccgatgaagtgggtcgcacctcggtttt
ctggaaggcgagcatcgtttgctcggccaggactctagctatagttctagtggttggct
acgtaccgtagtggcctatggcagggcttgcgcttaatgcgcccgtacagggcgcggtg
ggataccccctagagccccagctggttcttcccgcctcagaagccatagagcccaccgc
atccccagcatgcctgctattgtcttcccactcctcccccttgcgctcctgccccacc
ccccccagaatagaatgacacctactcagacaatgcgatgcaatttctcattttat
taggaaaggacagtgaggagtgccaccttccagggtcaaggaaggcacgggggaggggca
aacaacagatggctggcaactagaaggcacagtcgaggctgatcagcgggtttaaaccg
gccctctagactcagagttaaagtcgacgcggggaggcggcccaaggagatccgactc
gtctgagggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcggga
aggaaggtccgctggattgaggccgaaggacgtagcagaaggacgtcccgcgcagaa
tccaggtggcaacacaggcagcagccaaggaaaggacgatgatttccccgacaacacc
acggaattgtcagtgcccaacagccgagccctgtccagcagcgggcaaggcaggcggc
gatgagttccgccgtggcaatagggagggggaaagcgaagtcgccggaaggagctgac
aggtgggtggcaatgcccccaaccagtggggggttgcgctcagcaaacacagtgccaccacg
ccacgttgcctgacaacgggcccacaactcctcataaagagacagcaaccaggatttata
caaggaggagaaaatgaaagccatacgggaagcaatagcatgatataaaggcattaaag
cagcgtatccacatagcgtaaaaggagcaacatagtttaagaataccagtcfaatcttca
caaattttgtaatccagaggttgattgtcgcacttaacgcggttgaattcttacagattgg
ttgaaatcccattatagtaagattcggaccttgatgccctcttaattgtttttaagccg
tgcataaccttgagcaagcgttcaagacaaaacctgcacctatccatggtttgcacat
acccattctctatcaagagaaactggcccagcactgccgaagaaagctccaggtccc
cgtgcattatatcaagagtatccccaaagaccatacaggaatctcctacgatttccgaag
tcgatttccagatagtcagaaactctttgatgagagcttcaagattttcccaggtaca
tcccgaaccatctgacagaagttcaccatagtaaattcttccagggtgctctttgccgt
cagactcttccggtaacaaggtccgacttcgaaaatttttatgggcctggtaaaatc
ctatcgagttttcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag

at t t t t a t c c a c c c g g a a g a t c t g t t t t g a a g t t c a g t a t c a t t a t t a a t a c c c a t t c
t c t c c a c g t a t t c c g c c g g a a t a a g g a t a g g a g a c t t t a t c t c c a g a a a a c c c c g g t c t
a c g a a a a a t t t c g t a a t a t c a c g t t c g a g t t t a c c g a g g t a g t c t t c t c t a t c a t t g g t
a t a g a g c c g c t g a a a t c g t t t t t t c t t c t t g t c a c a a g t t c a g g c t c a a g t t c c c t g a
a a g g c t t t g c c a t a t t t t a g a g a a t t t t a t c c t c t g g a c t t a a g a g a g c c t c a a c c c t a
t c a a g c t g g c t t c t t g t a a g t g a a g g a g c a g g a g c c g a t g c g g g a a c a g a c g a a t t t g g
a g t t g a t t t t g c a g g c g a a g g t a c a g a t c t g g a t g t g t t c g t t g a t g c t t t t g c a g a a
c a g a t t t t c c a g a g g c t t c g g a g c c c t t g a a a c t g a t t t c g g c a t a g c t t t t t t g a c c
t t t g g a g c a g a a a c t a c c c t a a c t t t c a c a c t g t t t t t g c t t t c g g t t g a t c t t g t g a g
a a a t t a t t g a t a t c c t c g t c c g a a a c c c t a c a t c g t t t g c a g g t t t t t c t g t a c t t a t
g a t g t c t g a a t g c t c t g g c t g t t c t a c a a c t c c t g g a a t t a t t c a c a a c a a g a t g g t c t
c c a c a c g c c a t t t c a a t g t a t a t t t t a c t t c t t g a g a c c t c a t g g t g c t t g a t t t t g t g
g a g c g t g c c a g t c c t g g a c a t c c a g a g c c c g g t c g c a g a t a t t a a a c a t c t a a t g g t t
t t t t a t c c a t g g t g g c g t a g c c a g c t t g g g t c t c c c t a t a g t g a g t c g t a t t a
g a t a a g c c a g t a a g c a g t g g g t t c t c t a g t t a g c c a g a g a g c t c t g c t t a t a t a g a c c t
c c c a c c g t a c a c g c c t a c c g c c a t t t g c g t c a a t g g g g c g g a g t t g t t a c g a c a t t t t
g g a a a g t c c c g t t g a t t t t g g t g c c a a a a c a a a c t c c c a t t g a c g t c a a t g g g g t g g a g
a c t t g g a a a t c c c c g t g a g t c a a a c c g c t a t c c a c g c c a t t g a t g t a c t g c c a a a a c c
g c a t c a c c a t g g t a a t a g c g a t g a c t a a t a c g t a g a t g t a c t g c c a a g t a g g a a a g t c c
c a t a a g g t c a t g t a c t g g g c a t a a t g c c a g g c g g g c a t t t a c c g t c a t t g a c g t c a a t
a g g g g g c g t a c t t g g c a t a t g a t a c a c t t g a t g t a c t g c c a a g t g g g c a g t t t a c c g t a
a a t a g t c c a c c c a t t g a c g t c a a t g g a a a g t c c c t a t t g g c g t t a c t a t g g g a a c a t a c
g t c a t t a t t g a c g t c a a t g g g c g g g g t c g t t g g g c g g t c a g c c a g g c g g g c a t t t a c
c g t a a g t t a t g t a a c g c g g a a c t c c a t a t a t g g g c t a t g a a c t a a t g a c c c c g t a a t t g
a t t a c t a t t a a t a a c t a g t c a a t a a t c a a t g t c a a c g c g t a t a t c t g g c c c g t a c a t c g
c g a a g c a g c g c a a a a c g g a t c c t g c a g g t a t t t g c g g c c g c g g t c c g t a t a c t c c g g a a
t a t t a a t a g a t c a t g g a g a t a a t t a a a a t g a t a a c c a t c t c g a a a t a a a t a a g t a t t
t a c t g t t t t c g t a a c a g t t t t g t a a t a a a a a a c c t a t a a a t a t t c c g g a t t a t t c a t a
c c g t c c c a c c a t c g g g c g c g a a c t c c t a a a a a a c c g c c a c c a t g a a g t g c c t t t t g t a c
t t a g c c t t t t t a t t c a t t g g g g t g a a t t g c a a g t t c a c c a t a g t t t t t c c a c a c a c c a
a a a a g g a a a c t g g a a a a t g t t c c t t c t a a t t a c c a t t a t t g c c c g t c a a g c t c a g a t t
t a a a t t g g c a t a a t g a c t t a a t a g g c a c a g c c t t a c a a g t c a a a a t g c c c a a g a g t c a c
a a g g c t a t t c a a g c a g a c g g t t g g a t g t g t c a t g c t t c c a a a t g g g t c a c t a c t t g t g a
t t t c c g c t g g t a t g g a c c g a a g t a t a t a a c a c a t t c c a t c c g a t c c t t c a c t c c a t c t g
t a g a a c a a t g c a a g g a a g c a t t g a c a a a c g a a c a a g g a a c t t g g c t g a a t c c a g g c
t t c c c t c c t c a a a g t t g t g g a t a t g c a a c t g t g a c g g a t g c c g a a g c a g t g a t t g t c a
g g t g a c t c c t c a c c a t g t g c t g g t t g a t g a a t a c a c a g g a g a a t g g g t t g a t t c a c a g t
t c a t c a a c g g a a a t g c a g c a a t t a c a t a t g c c c c a c t g t c c a t a a c t c t a c a a c c t g g
c a t t c t g a c t a t a a g g t c a a a g g g c t a t g t g a t t c t a a c c t c a t t t c c a t g g a c a t c a c
c t t c t t c t c a g a g g a c g g a g a g c t a t c a t c c c t g g g a a g g a g g g c a c a g g g t c a g a a
g t a a c t a c t t t g c t t a t g a a a c t g g a g g c a a g g c c t g c a a a a t g c a a a c t g c a a g c a t
t g g g g a g t c a g a c t c c c a t c a g g t g t c t g g t t c g a g a t g g c t g a t a a g g a t c t c t t t g c
t g c a g c c a g a t t c c c t g a a t g c c c a g a a g g g t c a a g t a t c t c t g c t c c a t c t c a g a c c t
c a g t g g a t g t a a g t c t a a t t c a g g a c g t t g a g a g g a t c t t g g a t t a t t c c c t c t g c c a a
g a a c c t g g a g c a a a a t c a g a g c g g g t c t t c c a a t c t c t c c a g t g g a t c t c a g c t a t c t
t g c t c c t a a a a a c c c a g g a a c c g g t c c t g c t t t c a c c a t a a t c a a t g g t a c c c t a a a a t
a c t t t g a g a c c a g a t a c a t c a g a g t c g a t a t t g c t g c t c c a a t c c t c t c a a g a a t g g t c
g g a a t g a t c a g t g g a a c t a c c a c a g a a a g g g a a c t g t g g g a t g a c t g g g c a c c a t a t g a

aacagocacaacgtctatatacatggocgacaagcagaagaacggcatcaaggtgaactt
caagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcaga
acacccccatcgggcagggccccgtgctgctgccgacaaccactacctgagcaccag
tccgcctgagcaaagacccccaacgagaagcgcgatcacatggctcctgctggagtctgt
gaccgocgocgggatcactctcgccatggacgagctgtacaaggggccccttcgaacaaa
aactcatctcagaagaggatctgaatatgcataccggatcatcatcaccatcaccattga
gaattcaacgcggttaagtcgacaaatcaacctctggattacaaaatgtgaaagattga
ctggattcttaactatggtgctccttttacgctatgtggatacgtgctttaatgcct
ttgtatcatgctattgcttcccgtatggctttcattttctcctccttgataaaatcctg
gttgctgtctctttatgaggagtgtggcccggtgtcaggcaacgtggcgtggtgtgca
ctgtggttgctgacgaacccccactgggtggggcattgccaccacctgtcagctcctt
tccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgcct
tggccgctgctggacaggggctcggctggtgggactgacaattccgtgggtgtgtcgg
ggaaatcatcgctcctttccttggctgctcgccgtgtgttgccacctggattctgcgogg
acgtccttctgctacgtcccttcggccctcaatccagcggaccttccctcccgcggcct
gctgcggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatct
ccctttgggcccgcctccccggcggcctggccggccgcccagcacagtggctcgatcga
ccaa
tggcctggctcacaataaccactgagatctttttccctctgccaaaaattatggggaca
tcatgaagccccttgagcatctgacttctggctaataaaggaaatatttttattgca
atagtgtgttggaaatgtgtctctcactcggaggacatatgggagggcaaatca
tttaaacaatcagaatgagtatttggtttagagtttggcaacatagcccatatgctgg
ctgccatgaacaaagggtggctataaagaggatcagtatatgaaacagccccctgct
gtccattccttattccatagaaaagccttgacttgaggttagatttttttatattttg
ttttgtgttattttttctttaacatccctaaaattttcccttacatgtttactagcca
gatttttccctcctcctgactactcccagtcatagetgtccctcttctcttatggaga
tccctcgacctgccctaggagatccgaaccagataagtgaaatctagttccaaactatt
ttgtcatttttaattttcgtattagcttacgacgctacaccagttcccatctattttg
tcactcttccctaaataatccttaaaaactccatttccaccctcccagttcccaacta
ttttgtccgcccacagcggggcatttttcttccgttatgtttttaatcaaacatcctg
ccaactccatgtgacaaaccgtcatcttcggctactttttctctgtcacagaatgaaaa
ttttctgtcatctcttcggttattaatggttgtaattgactgaatatcaacgcttattt
gcagc

pAcBac-CMV-MbPylRS-CAG-GFPtagAvrIIdeI-4xPylR



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
UbiC promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
CAG promoter	text
GFP39tag	text
WPRE	text

rb-beta globin terminator	text
U6	text
MmPytR	text
Tn7L	text

Sequence for HTS25 construct

ctgaatggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtgggtggt
tacgcgcagcgtgaccgctacacttgccagcgccttagcgcgccgctcctttcgctttct
tccttcctttctcgccacgttcgcccggctttcccgcgtaagctctaaatcgggggctc
cctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggg
tgatggttcacgtagtgggcatcgccctgatagacggtttttcgccccttgacgttg
agtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggattttgccgatttcggcctattgggttaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgtttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaaggaagagtatgagattcaacatttccgtgctcgccttattcccttttttgccg
cattttgccttctggtttttgctcaccagaaacgctggtgaaagtaaaagatgctgaa
gatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgcccgaagaacgttttccaatgatgagcacttttaaagttctgctat
gtggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgcgcatacac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggcca
actacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatg
gggatcatgtaactcgccttgatcggtgggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgcttgcgcaaactattaa
ctggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataa
atctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agcctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgtgagataggtgcctcactgattaagcattggtaactgtcagacca
agtttactcatatatacttttagattgattttaaacttcatttttaatttaaaggatct
aggatgaagatcctttttgataatctcatgaccaaatacccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttct
gcgcgtaatctgctgcttgcaaaaaaaaccaccgctaccagcgggtggtttgtttgc
cggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagcgcgtagttaggccaccacttcaagaactctgtagc
accgcctacatacctcgctctgctaatacctggtaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctc
ggctgaacggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggtaagcggcagggctcggaacaggagagcgcacgaggagcttccaggg
ggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg
atthttgtgatgctcgtcagggggggcggagcctatggaaaaacgccagcaacgcggcct
ttttacggttcctggcctttttgctggcctttttgctcacatggtctttcctgcggtatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgcgcgag
ccgaacgaccgagcgcagcaggtcagtgagcgggaagcgggaagagcgcctgatgcggt

at t t t t c t c t t a c g c a t c t g t g c g g t a t t t c a c a c c g c a g a c c a g c c g c g t a a c c t g g c
a a a a t c g g t t a c g g t t g a g t a a t a a t g g a t g c c c t g c g t a a g c g g g t g t g g g c g g a c a
a t a a a g t c t t a a a c t g a a c a a a a t a g a t c t a a a c t a t g a c a a t a a a g t c t t a a a c t a g a
c a g a a t a g t t g t a a a c t g a a a t c a g t c c a g t t a t g c t g t g a a a a g c a t a c t g g a c t t t
t g t t a t g g c t a a a g c a a a c t c t t c a t t t t c t g a a g t g c a a a t t g c c c g t c g t a t t a a a g
a g g g g c g t g g c c a a g g g c a t g g t a a a g a c t a t a t t c g c g g c g t t g t g a c a a t t a c c g a
a c a a c t c c g c g g c c g g g a a g c c g a t c t c g g c t t g a a c g a a t t g t t a g g t g g c g g t a c t t
g g g t c g a t a t c a a a g t g c a t c a c t t c t t c c c g t a t g c c c a a c t t t g t a t a g a g a g c c a c
t g c g g g a t c g t c a c c g t a a t c t g c t t g c a c g t a g a t c a c a t a a g c a c c a a g c g c g t t g g
c c t c a t g c t t g a g g a g a t t g a t g a g c g c g g t g g c a a t g c c c t g c c t c c g g t g c t c g c c g
g a g a c t g c g a g a t c a t a g a t a t a g a t c t c a c t a c g c g g c t g c t c a a a c c t g g g c a g a a c
g t a a g c c g c g a g a g c g c c a a c a c c g c t t c t t g g t c g a a g g c a g c a a g c g c g a t g a a t g
t c t t a c t a c g g a g c a a g t t c c c g a g g t a a t c g g a g t c c g g c t g a t g t t g g g a g t a g g t g
g c t a c g t c t c c g a a c t c a c g a c c g a a a g a t c a a g a g c a g c c c g c a t g g a t t t g a c t t g
g t c a g g g c g a g c c t a c a t g t g c g a a t g a t g c c c a t a c t t g a g c c a c c t a a c t t t g t t t
t a g g g c g a c t g c c c t g c t g c g t a a c a t c g t t g c t g c t g c g t a a c a t c g t t g c t g c t c c a
t a a c a t c a a a c a t c g a c c c a c g g c g t a a c g c g t t g c t g c t t g g a t g c c c g a g g c a t a g
a c t g t a c a a a a a a c a g t c a t a a c a a g c c a t g a a a c c g c c a c t g c g c c g t t a c c a c c g
c t g c g t t c g g t c a a g g t t c t g g a c c a g t t g c g t g a g c g c a t a c g t a c t t g c a t t a c a g
t t t a c g a a c c g a a c a g g c t t a t g t c a a c t g g g t t c g t g c t t c a t c c g t t t c c a c g g t g
t g c g t c a c c c g g c a a c c t t g g g c a g c a g c g a a g t c g a g g c a t t t c t g t c c t g g c t g g c g
a a c g a g c g c a a g g t t t c g g t c t c a c g c a t c g t c a g g c a t t g g c g g c c t t g c t g t t c t t
c t a c g g c a a g g t g c t g t g c a c g g a t c t g c c c t g g c t t c a g g a g a t c g g t a g a c c t c g g c
c g t c g c g g c g c t t g c c g g t g g t g c t g a c c c g g a t g a a g t g g t t c g c a t c c t c g g t t t t
c t g g a a g g c g a g c a t c g t t t g t t c g c c c a g g a c t c t a g c t a t a g t t c t a g t g g t t g g c t
a c g t a c c c g t a g t g g c t a t g g c a g g g c t t g c g t t a a t g c g c c g c t a c a g g g c g c g t g g
g g a t a c c c c t a g a g c c c c a g c t g g t t c t t t c c g c c t c a g a a g c c a t a g a g c c a c c g c
a t c c c c a g c a t g c c t g c t a t t g t c t t c c c a a t c c t c c c c t t g c t g t c c t g c c c a c c c
c a c c c c c a g a a t a g a a t g a c a c c t a c t c a g a c a a t g c g a t g c a a t t t c c t c a t t t t a t
t a g g a a g g a c a g t g g g a g t g g c a c c t t c c a g g g t c a a g g a a g g c a c g g g g g a g g g g c a
a a c a c a g a t g g c t g g c a a c t a g a a g g c a c a g t c g a g g c t g a t c a g c g g g t t t a a c g g
g c c c t c t a g a c t c g a g t t a a a g t c g a c g c g g g g a g g c g g c c c a a a g g g a g a t c c g a c t c
g t c t g a g g g c g a a g g c g a a g a c g c g g a a g a g g c c g c a g a g c c g g c a g c a g g c c g c g g g a
a g g a a g g t c c g c t g g a t t g a g g g c c g a a g g g a c g t a g c a g a a g g a c g t c c c g c g c a g a a
t c c a g g t g g c a a c a c a g g c g a g c a g c c a a g g a a g g a c g a t g a t t t c c c c g a c a a c a c c
a c g g a a t t g t c a g t g c c c a a c a g c c g a g c c c c t g t c c a g c a g c g g g c a a g g c a g g c g g c
g a t g a g t t c c g c c g t g g c a a t a g g g a g g g g a a a g c g a a a g t c c c g g a a a g g a g c t g a c
a g g t g g t g g c a a t g c c c c a a c c a g t g g g g g t t g c g t c a g c a a a c a c a g t g c a c a c c a c g
c c a c g t t g c c t g a c a a c g g g c c a c a a c t c c t a t a a a g a g a c a g c a a c c a g g a t t t a t a
c a a g g a g g a g a a a t g a a a g c c a t a c g g g a a g c a a t a g c a t g a t a c a a a g g c a t t a a a g
c a g c g t a t c c a t a g c g t a a a a g g a g c a a c a t a g t t a a g a a t a c c a g t c a a t c t t t c a
c a a a t t t t g t a a t c c a g a g g t t g a t t g t c g a c t t a a c g c g t t g a a t t c t t a c a g a t t g g
t t g a a a t c c c a t t a t a g t a a g a t t c g g a c c t t g a t g c c c t c t t a a t g t t t t t a a a g c c g
t g c a t a a c c t t g a g c a a g c g t t c a a g a c c a a a a c c t g c a c c t a t c c a t g g t t t t g t c a a t
a c c c c a t t c t c t a t c a a g a g a a a c t g g c c c g a c g a c t g c c g a a g a a a g c t c c a g g t c c c
c g t g c a t t a t a t c a a g a g t a t c c c c a a a g a c c a t a c a g g a a t c t c c t a c g a t t t c g a a g
t c g a t t t c c a g a t a g t c c a g a a a c t c t t t g a t g a g a g c t t c a a g a t t t t c c c g a g t a c a
t c c c g a a c c c a t c t g a c a g a a g t t c a c c a t a g t a a a t t c t t c c a g g t g c t c t t t g c c g t

cagactctttccggtaacaagggtccgacttcgaaaatTTTTattgggcttggtaaaatc
ctatcgagttttcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag
atTTTTatccaccggaagatctgTTTTgaaagttcagtatcattattaataccattc
tctccacgtattccgcgcggaataaggataggagactttatctccagaaaaccccggtct
acgaaaaatTTcgtaatatcacgttcaggtttaccgaggtagtcttctctatcattggg
atagagccgctgaaaatcgTTTTtcttcttgtcacaagttcaggctcaagttccctga
aaggctttgccatatttagagaaatTTTatcctctggacttaagagagcctcaacccta
tcaagctggcttcttgtaagtgaaggagcaggagccgatgcgggaacagacgaatttg
agttgattttgcaggcgaaggtacagatctggatgtgttcgttgatgcctttgcagaaa
cagaatTTTccagaggcttcggagcccttgaaactgatttcggcatagctTTTTtgacc
tttgagcagaaaactaccctaactTTTcacactgTTTTgcttctcggttgatcttgtgag
aaaattattgatatcctcgtccgaaaccctacatcgTTTgcaggTTTTtctgtacttat
gatgtctgaatgctctggctgttctacaactcctggaattattcacaacaagatggctct
ccacacgccatttcaatgtatatTTTacttcttgagacctcatgggtgcttgattttgtg
gagcgtgccagtcctggacatccagagcccggctcgcagatattaaaacatctaattggtt
TTTTatccatgggtggcgctagccagcttgggtctcccTatagtgagtcgattaatttc
gataagccagtaagcagtggggttctctagtttagccagagagctctgcttatatagacct
cccaccgtacacgcctaccgccattttgctgcaatggggcggagttgttacgacatttt
ggaaagtcccgttgattttgggtgccaaaacaaactcccattgacgtcaatgggggtggag
acttgaaatccccgtgagtcaaaccgctatccacgccattgatgtactgccaaaacc
gcatcaccatggtaatagegatgactaatacgtagatgtactgccaagtaggaaagtcc
cataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgtcaat
agggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgta
aatagtcaccaccattgacgtcaatggaaagtccctattggcgttactatgggaacatac
gtcattattgacgtcaatgggcgggggtcgttgggoggctcagccaggcgggccatttac
cgtaagttatgtaaccgcggaactccatatatgggctatgaactaatgaccccgtaattg
attactattaataaCtagtcaataatcaatgtcaacgcgtatatctggcccgtacatcg
cgaagcagcgcaaaacggatcctgcaggatTTTgcgggcgcggctccgtatactccggaa
tattaatagatcatggagataattaaaatgataaccatctcgcaataaataagtattt
tactgTTTTcgtaacagttttgtaataaaaaaacctataaatattccggattattcata
ccgtcccaccatcgggcggaactcctaAAAAaccgccaccatgaagtgccttttgtac
ttagcctTTTTattcattgggggtgaattgcaagttcaccatagttttccacacaacca
aaaaggaaactggaaaaatgttcttctaattaccattattgcccgtaagctcagatt
taaattggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcac
aaggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgtga
tttccgctggatggaccgaagtatataacacattccatccgatccttactccatctg
tagaacaatgcaaggaaagcattgaacaacgaaacaaggaacttggctgaatccaggc
ttcctcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtcca
ggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagt
tcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaacctgg
cattctgactataagggtcaaagggtatgtgattctaacctcatttccatggacatcac
cttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaa
gtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcat
tggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctctttgc
tgcagccagattcctgaatgccagaagggtcaagtatctctgctccatctcagacct
cagtggtatgtaagtctaattcaggacgttgagaggatcttggattattccctctgcca
gaaacctggagcaaaatcagagcgggtcttccaatctctccagtggtatctcagctatct
tgctcctaaaaaccaggaaccggctcctgctttcaccataatcaatggtaccctaaat

actttgagaccagatacatcagagtgatattgctgctccaatcctctcaagaatggc
ggaatgatcagtggaactaccacagaaaggaactgtgggatgactgggcaccatgatga
agacgtgaaattggaccaatggagttctgaggaccagttcaggatataagtttcctt
tatacatgattggacatggtatgttggactccgatcttcatcttagctcaaaggctcag
gtgttcgaacatcctcacattcaagacgctgcttcgcaacttctgatgatgagagttt
atTTTTGGTGATACTGGGCTATCCAAAAATCCAATCGAGCTTGTAGAAGGTTGGTTCA
GTAGTTGGAAAAGCTCTATTGCCCTTTTTCTTTATCATAGGGTTAATCATTGGACTA
TTCTTGGTCTCCGAGTTGGTATCCATCTTGCATTAATAAAGCACACCAAGAAAAG
ACAGATTTATACAGACATAGAGATGAACCGACTTGGAAAGTGATAAGTGCAGAGAAGTACT
AGAGGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAACCTC
CCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGT
TATTGCAGCTTATAATGGTTACAATAAAGCAATAGCATCACAAATTCACAAATAAAG
CATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCAT
GTCTGGATCTGATCACTGCTTGAGCTAGTTATTAATAGTAATCAATTAAGGGGTCA
AGTTCATAGCCATATATGGAGTTCGCGTTACATAAETTAAGGTAATGGCCGCGCTG
GCTGACCGCCCAACGACCCCCGCCCATTGAOCTCAATAATGACGTATGTTCCCATAGTA
ACGCCAATAGGGACTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCA
CTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACG
GTAATGGCCGCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAOTTGG
CAGTACATCTAAGTATTAGTCATCGCTATTACCATGGTGGAGGTGAGCCCCAGTTCTG
CTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTT
AATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGGGG
GCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAATCAGAGCG
GCGCGCTCCGAAAGTTTCTTTTATGGCGAGGCGGGCGGGCGGGCGGGCCATATAAAAAG
CGAAGCGCGCGGGCGGGGAGTGGCTGGCTTGCCTTGCCTCGCGCGTCCCGCGTCCCGCG
CGCTCGCGCGCGCGCGCGCGCGCTGACTGACCGGTTACTCCACAGGTGAGCGGGC
GGGACGGCCCTTCTCCTCCGGGTGTAATTAGCGCTTGGTTAATGACGGCTCGTTTCT
TTCTGTGGGTGCGTGAAAGCCTTAAAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGGGA
GGGTCTGGGGGGTGGCTGGGTGTGTGTGTGGTGGGGAGCGCGCGTGGGGCCCGCG
TGCCCGGGGTGTGAGCGCTGGGGCGGGGCGGGGGCTTTGTGCGCTCCCGGTGTG
GCGAGGGGAGCGCGGCGGGGGCGGTGCCCGCGGTGGGGGGGGCTGCGAGGGGAACA
AAGGTGCGTGCGGGGTGTGTGCGTGGGGGGTGGAGCGGGGGTGTGGGCGCGGGGT
GGGTGTAACCCCCCTGCACCCCCCTCCCGAGTTGCTGAGCACGGCCCGGTCTGG
GTGGGGGTCCGTGGGGGCGTGGCGGGGGTGGCGGTGCGGGGGGGGGGGTGGGG
CAGGTGGGGGTGCGGGGCGGGGCGGGGCGCGCTCGGGCGGGGGAGGGCTCGGGGGAGGG
GCGCGGCGGCCCCGGAGCGCGCGGGCTGTGAGGCGCGGGGAGCGCGCAGCCATTGCT
TTATGGTAATCGTGCAGAGGGGCGCAGGGACTTCTTTGTCCAAATCTGGCGGAGCC
GAAATCTGGGAGGCGCGCGCACCCCCCTAGCGGGCGGGGGGAAGCGGTGGGGCG
CGGCAAGGAAGGAATGGGGGGGGAGGGCCCTCGTGGTGGCGCGCGCGCGTCCCTT
TCCATCTCAGCCTCGGGGTGCGCAGGGGGACGGGTGCTTGGGGGGGACGGGGCA
GGGCGGGGTTGCGCTTCTGGGTGTGACCGGCGGCTTAGAGCTCTGCTAACCATGTT
CATGCTTCTTCTTTTCTACAGCTCCTGGCAACGTGGTGGTTATTTGTGTCTCTCA
TCATTTTGGCAAAGAATTGGCCAAGGAGGCCACTGGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGTGGTGCCATCCTGGTGGAGCTGGACGGCGAAGTAAACGGCCACAGTTC
AGCGTGTCCGGCGAGGGCGAGGGCGATGCCACTTAGGGCAAGCTGACCCTGAAGTTCAT
CTGCACCACCGCAAGCTGCCCGTGCCTGGCCACCCTCGTGACCACCCTGACCTACG
GCGTGCAGTGTTCAGCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCC
GCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA

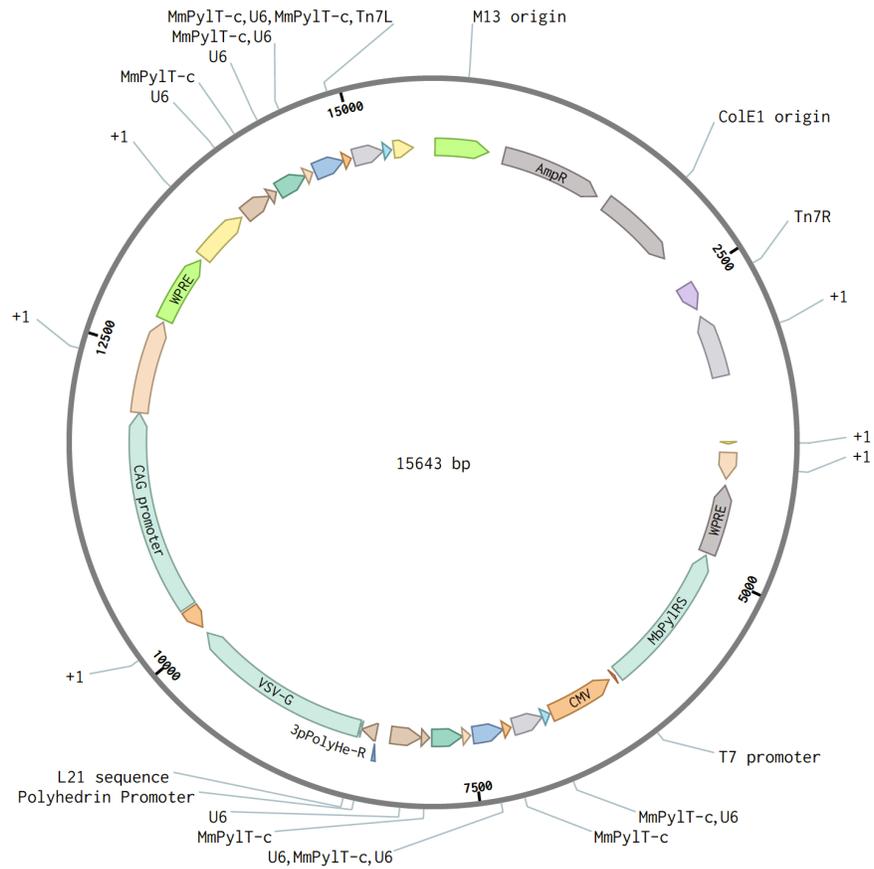
caagaccgcgcccagaggtgaagttcgagggogacaccctgggtgaaccgcatcgagctga
agggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactac
aacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaagggtgaactt
caagatccgcccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcaga
acacccccatcgggagcggccccctgctgctgcccgacaaccactacctgagcaccag
tccgcccctgagcaagaccccaacgagaagcgcgatcacatggtcctgctggagtctgt
gaccgcccggggatcactctcgccatggacgagctgtacaaggggccccttcgaacaaa
aactcatctcagaagaggatctgaatatgcataccgggtcatcatcaccatcaccattga
gaattcaacgcggttaagtgcacaatcaacctctggattacaaaatcttgtaaagattga
ctggattcttaactatgttgctccttttaocgctatgtggatacgtgctttaatgctt
ttgtatcatgctattgcttcccgtatggctttcattttctcctccttgataaatcctg
gttgctgtctctttatgaggagttgtggcccgttgctcaggcaacgtggcgtggtgtgca
ctgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctt
tccgggactttcgctttccccctccctattgccacggcggaaactcatcgccgctgctt
tgcccgtgctggacaggggctcggctggtgggactgacaattcogtgggtgtgtcgg
ggaaatcatcgctcctttccttggtgctcgcctgtgttgccacctggattctgcgagg
acgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggct
gctgcccgtctgcccctcttccgctcttcgccttcgcctcagacgagtcggatct
ccctttgggcccctccccggcctggccggccgcccagcacagtgggtcgatcgacaac
tgccctggctcacaataccactgagatctttttccctctgcaaaaaattatggggaca
tcatgaagccccttgagcatctgacttctggctaataaaggaaatctttttcattgca
atagtgtgttggaatctttttgtgtctctcactcgggaaggacatatgggagggcaaatca
tttaaacatcagaatgagtatttggttagagtttggaacatatgcccatatgctgg
ctgccatgaacaaagggttgctataaagaggtcatcagtatatgaaacagccccctgct
gtccattccttattccatagaaaagccttgacttgaggtttagatcttttttatatcttg
ttttgtgttattttttctttaacatccctaaaatcttcttacatgttttactagcca
gatttttctcctctcctgactactcccagtcocatagctgtccctcttctcttatatggaga
tccctcgacctgccctaggggtaccctcgggcaggaagaggggctatcttccatgattcc
ttcatatctgcatatacgatacaaggctgtagagagataattagaattaatttgactg
taaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttag
tttgagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaa
gtatttgcatttcttggtttatatactctgtggaaggacgaaacaccgggcccgtga
tcatgtagatcgaacggactctaaatccgttcagcccgggttagattcccgggctgcccg
ttttttgctaggggtaccctcgggcaggaagagggcctatcttccatgattccttcatat
ttgcatatacgatacaaggctgtagagagataattagaattaatttgactgtaaacac
aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttagttgcag
ttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtattc
gatttcttggtttatatactctgtggaaggacgaaacaccgggcccgtgatcatgta
gatcgaacggactctaaatccgttcagcccgggttagattcccgggctgcccgttttttg
ctaggggtaccctcgggcaggaagagggcctatcttccatgattccttcatatctgata
tacgatacaaggctgtagagagataattagaattaatttgactgtaaacacaaagata
ttagtacaaaatacgtgacgtagaaagtaataatttcttggttagttgcagttttaa
attatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttgcatttct
tggctttatatactctgtggaaggacgaaacaccgggcccgtgatcatgtagatcga
cggactctaaatccgttcagcccgggttagattcccgggctgcccgttttttgctaggg
taccctcgggcaggaagagggcctatcttccatgattccttcatatctgcatatacgata
caaggctgtagagagataattagaattaatttgactgtaaacacaaagatattagtac
aaaatacgtgacgtagaaagtaataatttcttggttagtttgcagttttaaaattatgt

tttaaagtggactatcatatgcttaccgtaacttgaaagtatttcgattttcttggett
atatacttgtggaaggacgaaacaccgggcggtgatcatgtagatcgaacggactc
taaaccggttcagccgggttagattcccgggctgcccgttttttgctaggagatccgaa
ccagataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagctt
acgacgctacaccagttcccactctattttgtcactcttccctaaataatccttaaaaa
ctccatttccaccctcccagttcccaactattttgtccgccaca gcggggcattttt
cttctgttatggttttaatacaacatcctgccaactccatgtgacaaaccgtcatctt
cggctactttttctctgtcacagaatgaaaatttttctgtcatctcttcggttattaatg
tttgaattgactgaatatcaacgcttatttgcagc

The sequence for the wild type PytR is:

ggaaacctgatcatgtagatcgaacggactctaaatccggttcagccgggttagattccc
ggggttccg

pAcBac-CMV-MbPylRS-CAG-GFPtagAvrIIIdel-8xPylR



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
UbiC promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
CAG promoter	text

GFP39tag	text
WPRE	text
rb-beta globin terminator	text
U6	text
MmPytR	text
Tn7L	text

Sequence for HTS25 construct

ctgaatggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtgggtgt
tacgcgcagcgtgaccgctacacttgccagcgccttagcggccgctcctttcgctttct
tcccttcccttctcgcacgcttcgcccgtttccccgtcaagctcctaaatcgggggctc
cctttagggttccgatttagtgctttacggcacctcgaccccaaaaacttgattaggg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccccttgacgttg
agtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggattttgccgatttcggcctattgggttaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgtttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaaggaagagtatgagattcaacatttccgtgctcgccttattcccttttttgcg
cattttgccttctggtttttgctcaccagaaacgctggtgaaagtaaaagatgctgaa
gatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgcccgaagaacgttttccaatgatgagcacttttaaagttctgctat
gtggcgcggtattatcccgtattgacgcccggcaagagcaactcggtcgcccatacac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtgataaacactgcgcca
acttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcaacaatg
gggatcatgtaactcgccttgatcgttggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaa
ctggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataa
atctggagccgggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagttaggcaactatggatgaacga
aatagacagatcgtgagataggtgacctactgattaagcattggtaactgtcagacca
agtttactcatatatacttttagattgattttaaacttcatttttaatttaaaggatct
aggtgaagatcctttttgataatctcatgaccaaatacccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttct
gcgcgtaatctgctgcttgcaacaaaaaaaccaccgctaccagcgggtggtttggttg
cggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagccgtagttaggccaccactcaagaactctgtagc
accgcctacatacctcgtctctgctaatacctggtaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctc
ggctgaacgggggggttcgtgcacacagcccagcttgaggcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggtaagcggcagggtcggaacaggagagcgcacgaggagcttccaggg
ggaacgcctggtatctttatagtcctgctcgggtttcgccacctctgacttgagcgtcg
atthttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcct

ttttacggttcctggccttttggctggccttttggctcacatggtctttcctgcgttatcc
cctgattctgtggataaccgtattaccgctttgagtgagctgataccgctcgccgag
ccgaacgaccgagcgcagcagtgagtgagcaggaagcggagagcgcctgatgcgt
atcttctccttacgcatctgtgcggtatctcacaccgagaccagccgctaacctggc
aaaatcggttacggttgagtaataaatggatgcctgagcgttaagcgggtgtggggcggaca
ataaagtcttaactgaacaaaatagatctaaactatgacaataaagtcttaactaga
cagaatagttgtaaactgaaatcagtcagttatgctgtgaaaaagcatactggacttt
tgttatggctaaagcaactcttcattttctgaagtgcaaattgcccgctcgattaaag
agggcggtggccaagggcatggtaagactatattcgcgcggttgtagacaatttaccga
acaactccgcgggcgggaagccgatctcggcttgaacgaattgttaggtggcggtaactt
gggtcgatatcaaagtgcatacacttcttcccgtatgcccaactttgtatagagagccac
tgcgggatcgtcaccgtaatctgcttgacgtagatcacataagcaccgaagcgcgttg
cctcatgcttgaggagattgatgagcgcgggtggcaatgccctgcctccggtgctcgccg
gagactgagagatcatagatatagatctcactacgcggctgctcaaactgggcagaac
gtaagccgagagagcgcacaaccgcttcttggtcgaaggcagcaagcgcgatgaatg
tctactacggagcaagttcccggaggttaatcggagtcgggctgatggtgggagtaggtg
gctacgtctccgaactcagcaccgaaaagatcaagagcagcccgcattgatttgacttg
gtcagggccgagcctacatgtgcaatgatgccatacttgagccacctaactttgttt
tagggcgactgcctgctgcgtaacatcgttggctgctgcgtaacatcgttgctgctcca
taacatcaaacatcgaccacggcgtaacgcgcttggctgcttggatgcccgaggcatag
actgtacaaaaaacagtcataacaagccatgaaaaccgccactgcccgttaccaccg
ctgcttcgggtcaagggtctggaccagttgctgagcgcatacgtacttgattacag
tttacgaaccgaacaggcttatgtcaactgggttcgctgccttcacggttccacgggtg
tgctcaccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcg
aacgagcgaaggtttcggctccacgcacgctcaggcattggcggccttggctgttctt
ctacggcaagggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggc
cgctcgcggcgttgccggtgggtgctgaccccggatgaagtgggttcgcacctcggtttt
ctggaaggcagcagctgtttgttcgcccaggactctagctatagttctagtggttggct
acgtaccgtagtgggctatggcagggcttggcgttaatgcgcccgtacagggcgcgtgg
ggataccccctagagccccagctggttctttccgctcagaagccatagagcccaccgc
atccccgatgcctgctattgtcttcccactcctcccccttggctgtcctgccccacc
ccccccagaatagaatgacacctactcagacaatgcgatgcaatttctcattttat
taggaaaggacagtgggagtggcaccttccagggcgaaggaaggcagggggaggggca
aacaacagatggctggcaactagaaggcacagtgcaggctgatcagcgggtttaaaccg
gccctctagactcgagttaaagtcgacgcggggaggcggcccaaggagatccgactc
gtctgagggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcggga
aggaaggctccgctggattgagggccgaaggacgtagcagaaggacgtcccgcgcagaa
tccaggtggcaacacaggcagcagccaaggaaaggacgatgatttccccgacaacacc
acggaattgtcagtgcccaacagccgagccctgtccagcagcgggcaaggcaggcggc
gatgagttccgcccgtggcaatagggagggggaaagcgaagtcgggaaaggagctgac
aggtgggtggcaatgccccaccagtggggggttgcgtcagcaaacacagtgcacaccacg
ccacgttgctgacaacgggcccacaactcctcataaagagacagcaaccaggatttata
caaggaggagaaaatgaaagccatacgggaagcaatagcatgatacaaaggcattaaag
cagcgtatccacatagcgtaaaaggagcaacatagtttaagaataccagtcfaatcttca
caaattttgtaatccagaggttgattgtcgaacttaacgcgttgaattcttacagattgg
ttgaaatcccattatagtaagatccggaccttgatgccctcttaatgtttttaagccg
tgcataaccttgagcaagcgttcaagacccaaacctgcacctatccatggtttgtaacat
acccattctctatcaagagaaactggcccagcactgccgaagaaagctccaggtccc

cgtgcattatatcaagagtatccccaaagaccatacaggaatctcctacgatttcgaag
tcgatttccagatagtcagaaaactctttgatgagagcttcaagatttccccgagtaca
tccgaaccatctgacagaagttcaccatagtaaattcttccagggtctcttggcgt
cagactcttccggtacaaggtccgacttcgaaaattttattgggcctggtaaaatc
ctatcgagtttccgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag
attttatccaccggaagatctgttttgaaagttcagtatcattattaataccattc
tctccacgtattccgccggaataaggataggagactttatctccagaaaaccccggtc
acgaaaaatttcgtaatatcacgttcgagtttaccgaggtagtcttctctatcattggg
atagagccgctgaaaatcgtttttcttcttctgtcacaagttcaggctcaagttccctga
aaggcttggccatatttagagaaattttatcctctggacttaagagagcctcaacccta
tcaagctggcttcttgtaagtgaaggagcaggagccgatgcgggaacagacgaatttgg
agttgatttgcaggcgaaggtacagatctggatgtgttcggtgatgccttgcagaaa
cagaatttccagaggcttcggagcccttgaaactgatttcggcatagcttttttgacc
tttggagcagaaactaccctaactttcacactgttttgccttcgggtgatcttgtgag
aaaattattgatatcctcgtccgaaaccctacatcgtttgcagggttttctgtacttat
gatgtctgaatgctctggctgttctacaactcctggaattattcacaacaagatggctc
ccacacgccatttcaatgtatattttacttcttgagacctcatgggtgcttgatttggg
gagcgtgccagtcctggacatccagagcccggctgcagatattaaacatctaattgggt
ttttatccatgggtggcgctagccagcttgggtctcccctatagtgagtcgtattat
gataagccagtaagcagtggttctctagttagccagagagctctgcttatatagacct
cccaccgtacacgectaccgccatttgcgtcaatggggcggagttgttacgacatttt
ggaaagtcccgttgatttgggtgcccacaaactcccattgacgtcaatgggggtggag
acttgaaatcccgtgagtcacaccgctatccacgccattgatgtactgcccacacc
gcatccatggtaatagcgtactaatacgtagatgtactgccaagtaggaaagtcc
cataaggtcatgtactgggcataatgccaggcgggocatttaccgtcattgacgtcaat
agggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgta
aatagtcacccattgacgtcaatggaaagtcctattggcggtactatgggaacatac
gtcattattgacgtcaatggggcggggctggttggcggtcagccaggcgggocatttac
cgtaagttatgtaacgcggaactccatataatgggctatgaactaatgaccccgtaattg
attactattaataactagcaaaaaacgggcagcccgggaatctaaccgggtgaacgga
tttagagtcctgctacatgatcagccgcccgggtgttctcctttccacaagat
atataagccaagaatcgaataactttcaagttacggtaagcatatgatagtcattt
taaacataattttaaaactgcaactacccaagaaattattactttctacgtcacgta
ttttgtactaataatcttctgtgtttacagtcacaaatattctaatatctctctaacagc
cttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttctcctgccga
ggtagccctagcaaaaaacgggcagcccgggaatctaaccgggtgaacggatttagag
tccgttcgatctacatgatcagccgcccgggtgttctcctttccacaagatatataaa
gccaagaatcgaataactttcaagttacggtaagcatatgatagtcattttaaaaca
taattttaaaactgcaactacccaagaaattattactttctacgtcacgtattttgta
ctaataatcttctgtgtttacagtcacaaatattctaatatctctctaacagccttgtat
cgtatatgcaaatatgaaggaatcatgggaaataggccctcttctcctgccgaggtacc
ctagcaaaaaacgggcagcccgggaatctaaccgggtgaacggatttagagtcctgct
gatctacatgatcagccgcccgggtgttctcctttccacaagatatataagccaaga
aatcgaataactttcaagttacggtaagcatatgatagtcattttaaaacataatttt
aaaactgcaactacccaagaaattattactttctacgtcacgtattttgtactaataat
cttctgtgtttacagtcacaaatattctaatatctctctaacagccttgtatcgtatat
gcaaatatgaaggaatcatgggaaataggccctcttctcctgccgaggtaccctagcaa
aaaaacgggcagcccgggaatctaaccgggtgaacggatttagagtcctgctac

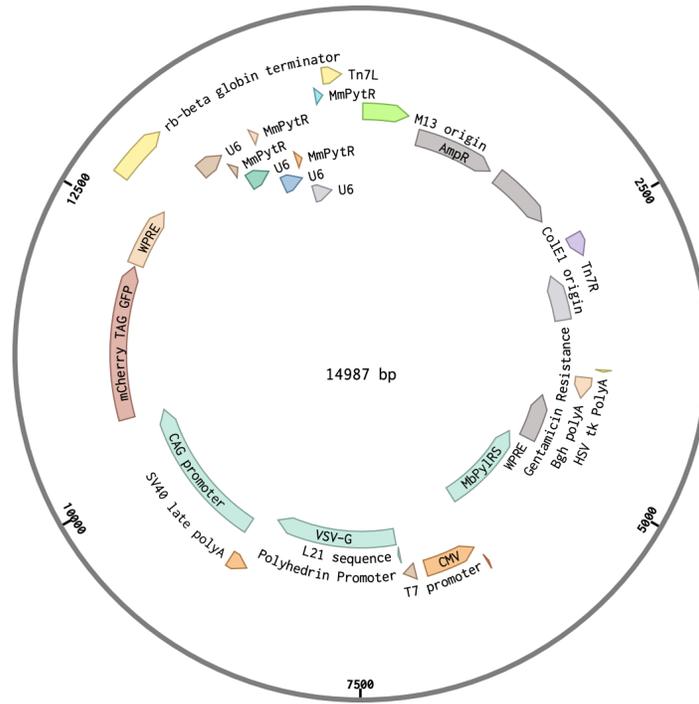
atgatcagccgccgggtgtttcgctcctttccacaagatatataaagccaagaaatcgaa
atactttcaagttacggtaagcatatgatagtcattttaaaacataattttaaaactg
caaactaccaagaaattattactttctacgtcacgtatgttactaataatctttgtg
tttacagtcaaattaattctaattatctctctaacagccttgatcgtatatgcaaata
tgaaggaatcatgggaaatagccctcttccctgccgaaggtagtcaataatca
atgtcaacgcgtatatctggcccgtacatcggaagcagcgcgaaaacggatcctgcagg
tatttgccggccgcgggtccgtatactccggaatattaatagatcatggagataattaa
tgataaccatctcgcaataaataagtattttactgttttcgtaacagttttgtaataa
aaaacctataaatattccggattattcataccgtcccaccatcggggcgaactccta
aaaaaccgccaccatgaagtgccttttctacttagcctttttattcattgggggtgaatt
gcaagttcaccatagtttttccacacaacaaaaaggaaactggaaaaatgttccttct
aattaccattattgccgtcaagctcagatttaaattggcataatgacttaataggcac
agccttacaagtcaaaatgccaagagtcacaaggctattcaagcagacgggttgatgt
gtcatgcttccaaatgggtcactacttgtgatttccgctggatggaccgaagtatata
acacattccatccgatccttccactccatctgtagaacaatgcaaggaaagcattgaaca
aacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaa
ctgtgacggatgccgaagcagtgattgtccagggtgactcctcaccatgtgctggttgat
gaatacacaggagaatgggttgattcacagttcatcaacggaaatgcagcaattacat
atgccccactgtccataactctacaacctggcattctgactataagggtcaaagggtat
gtgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatca
tccctgggaaaggaggccacaggggttcagaagtaactactttgcttatgaaactggagg
caaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcagggtgtct
gggtcgagatggctgataaggatctctttgctgcagccagattccctgaatgccagaa
gggtcaagtatctctgctccatctcagacctcagtggtatgtaagtctaattcaggacgt
tgagaggatccttgattattccctctgccaagaaacctggagcaaaatcagagcgggtc
ttccaatctctccagtggatctcagctatcttgcctctaaaaccagggaaccggctct
gctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcga
tattgctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaa
gggaactgtgggatgactgggcaccatatgaagacgtggaaattggaccaatggagtt
ctgaggaccagttcaggatataagtttctttatacatgattggacatggatgttgga
ctccgatcttcatcttagctcaaaggctcagggttccgaacatcctcacattcaagacg
ctgcttcgcaacttctgatgatgagagtttattttttgggtgatactgggctatccaaa
aatccaatcgagcttgtagaagggttggtcagtagttggaaaagctctattgcctcttt
tttctttatcatagggttaatcattggactattcttgggttctccgagttggatccatc
tttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaac
cgacttggaagtgataagtcgagaagtagtagaggatcataatcagccataaccacatt
tgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaacata
aatgaatgcaattggtgttggttaacttgtttattgcagcttataatggttacaataa
agcaatagcatcaciaaatttcaciaaataaagcatttttttccactgcattctagttgtg
tttgccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagccta
gttattaatagtaatecaattacgggggtcattagttcagccatataatggagttccgc
gttacataaactacggtaaatggcccgcctggctgaccgoccaaogacccccgoccat
gacgtcaataatgaogtatgttcccataagtaacgocaaatagggactttccattgacgtc
aatgggtggagtttaacggtaaaactgcccacttggcagtaacatcaagtgtatcatatg
ccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgocca
gtacatgaccttatgggactttcctacttggcagtaacatctacgtattagtcacgcta
ttaccatggctgaggtgagocccacgttctgcttccactctccccccctcc
ccacoccccaattttgtatbtatttttttaattattttgtgcagcagatggggggcggg

ggacatatgggagggcaaatacattttaaacaatcagaatgagatatttggttagagtttg
gcaacatatgcccatatgctggctgccatgaacaaaggttggctataaagaggtcatca
gtatatgaaacagccccctgctgtccattccttattccatagaaaagccttgacttgag
gtagatattttttatattttggtttgtgttattttttctttaacatccctaaaattt
tccttacatgttttaactagccagatttttcctcctctcctgactactcccagtcatagc
tgtccctcttctcttatggagatccctcgacctgccctaggggtaccctcgggcaggaag
agggcctatttcccatgattccttcatatttgcataacgatacaaggctgtagagag
ataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtag
aaagtaataatttcttgggtagtttgcagttttaaattatgttttaaattggactatc
atatgcttacogtaacttgaaagtatttgcatttcttggctttatatacttgtggaaa
ggacgaaacaccgggcggtgatcatgtagatcgaacggactctaaatccgttcagccg
ggtagattcccgggctgcccgttttttgctaggggtaccctcgggcaggaagagggcct
atttcccatgattccttcatatttgcataacgatacaaggctgtagagagataatta
gaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaa
taatttcttgggtagtttgcagttttaaattatgttttaaattggactatcatatgct
taccgtaacttgaaagtatttgcatttcttggctttatatacttgtggaaaggacgaa
acaccgggcggtgatcatgtagatcgaacggactctaaatccgttcagccgggtaga
ttcccgggctgcccgttttttgctaggggtaccctcgggcaggaagagggcctatttccc
atgattccttcatatttgcataacgatacaaggctgtagagagataattagaattaa
tttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataattt
ttgggtagtttgcagttttaaattatgttttaaattggactatcatatgcttaccgta
acttgaaagtatttgcatttcttggctttatatacttgtggaaaggacgaaacaccgg
gcggtgatcatgtagatcgaacggactctaaatccgttcagccgggtagattcccgg
gctgcccgttttttgctaggggtaccctcgggcaggaagagggcctatttcccatgattc
cttcatatttgcataacgatacaaggctgtagagagataattagaattaatttgact
gtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggt
gtttgcagttttaaattatgttttaaattggactatcatatgcttaccgtaacttgaa
agtatttgcatttcttggctttatatacttgtggaaaggacgaaacaccgggcggtg
atcatgtagatcgaacggactctaaatccgttcagccgggtagattcccgggctgccc
gttttttgctaggagatccgaaccagataagtgaatctagttccaaactatttgtca
ttttaattttcgtatttagcttacgacgtacaccagttcccatctatttgtcactc
ttccctaataatccttaaaaactccatttccaccctcccagttcccaactatttgt
ccgcccacagcggggcatttttcttctgttatgtttttaatcaaacatcctgccaact
ccatgtgacaaaccgtcatcttccggtactttttctctgtcacagaatgaaaattttc
tgtcatctcttcggttattaatgtttgtaattgactgaatatcaacgcttatttgcagc

The sequence for the wild type PytR is:

ggaaacctgatcatgtagatcgaacggactctaaatccgttcagccgggtagattccc
ggggtttccg

pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
CMV promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
CAG promoter	text
mCherry TAG GFP	text
WPRE	text
rb-beta globin terminator	text
U6	text

MmPytR	text
Tn7L	text

Sequence

ctgaatggcgaatgggacgcgcctgtagcggcgcattaagcgcggcgggtgtgggtgt
tacgcgcagcgtgaccgctacacttgccagcgcctagcgcgcctcctttcgctttct
tcctttcctttctcgccacgttcgcggctttcccgtcaagctctaatacgggggctc
ccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccccttgacgttg
agtccacgttctttaatagtggaactctgttccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggatgttgccgatttcggcctattggttaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaaggaagagtatgagtattcaacatttcggtgctgccttattcccttttttgogg
cattttgccttcctgtttttgctcaccagaaacgctggtgaaagtaaagatgctgaa
gatcagttgggtgcaagcagtggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgccccgaagaacgttttccaatgatgagcacttttaagttctgctat
gtggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgcccatacac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtataaacactgcggcca
actacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgacacaacatg
gggatcatgtaactcgcttgatcggtgggaaccggagctgaatgaagccatacacia
cgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgcttgcgcaaactattaa
ctggcgaactactactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttcggctggctggtttattgctgataa
atctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggttaactgtcagacca
agtttactcatatataacttttagattgatttaaaacttcatttttaatttaaaggatct
aggtgaagatcctttttgataatctcatgacaaaatcccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttct
gcgcgtaatctgctgcttgcaaaaaaaaccaccgctaccagcgggtggtttgtttgc
cggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagata
caaataactgtccttctagtgtagcctgtagttaggccaccactcaagaactctgtagc
accgcctacatacctcgctctgctaactctgttaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctc
ggctgaacggggggttcgtgcacacagcccagcttgagcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggtaagcggcagggctcggaaacaggagagcgcacgagggagcttccagg
gaaacgcctggatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg
atttttgtgatgctcgtcaggggggagcctatggaaaaacgccagcaacgcggcct
tttacggttcctggccttttgctggccttttgctcacatggtctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgcgcgag
ccgaacgaccgagcgcagcagtcagtgagcaggaagcggaaagagcgcctgatgcgg
atcttctccttacgcacatctgtgcggtatctcacaccgcagaccagccgcgtaacctgg
aaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgtgggaggaca

ataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactaga
cagaatagttgtaaactgaaatcagtcagttatgctgtgaaaaagcactactggacttt
tggtatggctaaagcaactcttcattttctgaagtgcaaattgcccgctgtattaaag
agggcggtggccaagggcatggtaagactatattcgcgcggtgtgacaatttaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtactt
gggtcgatatcaaagtgcatacacttcttcccgtatgcccaactttgtatagagagccac
tgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccgaagcgcgttgg
cctcatgcttgaggagattgatgagcgcgggtggcaatgccctgcctccgggtgctcgcg
gagactgcgagatcatagatatagatctcactacgcggctgctcaaacctgggcagaac
gtaagccgcgagagcgcacaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatg
tcttactacggagcaagttcccagggtaatcggagtcggctgatggtgggagtaggtg
gctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcgatggatttgacttg
gtcagggccgagcctacatgtgcaatgatgccatacttgagccacctaactttgttt
tagggcgactgccctgctgcgtaacatcgttgcgctgcgtaacatcgttgcgctcca
taacatcaaacatcgaccacggcgtaacgcgcttgcgcttggatgcccgaggcatag
actgtacaaaaaacagtcataacaagccatgaaaaccgccactgcgcggttaccaccg
ctgcgttcgggtcaagggtctggaccagttgcgtgagcgcatacgcctacttgcaattacag
tttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccacgggtg
tgcgtcaccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcg
aacgagcgaaggtttcggctccacgcacatcgtcaggcattggcggccttgcgcttctt
ctacggcaagggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggc
cgtcgcggcgcttgcgggtgggtgctgaccccgatgaagtgggtcgcacctcggtttt
ctggaaggcgagcatcgtttgctcggccaggactctagctatagttctagtggttggct
acgtaccgtagtggctatggcagggcttgcgcttaatgcgccgctacagggcgcggtg
ggataccccctagagccccagctggttcttcccgcctcagaagccatagagcccaccgc
atccccagcatgcctgctattgtcttcccacatcctcccccttgcgctcctgccccacc
cccccccagaatagaatgacacctactcagacaatgcgatgcaatttctcattttat
taggaaaggacagtgaggagtgccaccttccagggtcaaggaaggcacgggggaggggca
aacaacagatggctggcaactagaaggcacagtcgaggctgatcagcgggtttaaaccg
gccctctagactcgagttaaagtcgacgcggggaggcggcccaaggagatccgactc
gtctgagggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcggga
aggaagggtccgctggattgagggccgaaggacgtagcagaaggacgtcccgcgcagaa
tccagggtggcaacacaggcagcagccaaggaaaggacgatgatttccccgacaacacc
acggaattgtcagtgcccaacagccgagccctgtccagcagcgggcaaggcaggcggc
gatgagttccgccgtggcaatagggagggggaaagcgaagtcgccggaaggagctgac
agggtgggtggcaatgcccccaaccagtggggggttgcgtcagcaaacacagtgcacaccacg
ccacgttgcctgacaacgggccacaactcctcataaagagacagcaaccaggatttata
caaggaggagaaaatgaaagccatacgggaagcaatagcatgatataaaggcattaaag
cagcgtatccacatagcgtaaaaggagcaacatagtttaagaataccagtcfaatcttca
caaattttgtaatccagaggttgattgtcgcacttaacgcggttgaattcttacagattgg
ttgaaatcccattatagtaagattcggaccttgatgccctcttaatgtttttaaagccg
tgcataaccttgagcaagcgttcaagacaaaacctgcacctatccatggtttgcacat
acccattctctatcaagagaaactggcccagcactgccgaagaaagctccaggtccc
cgtgcattatatcaagagtatccccaaagaccatacaggaatctcctacgatttcgaag
tcgatttccagatagtcagaaactctttgatgagagcttcaagattttcccaggtaca
tcccgaaccatctgacagaagttcaccatagtaaattcttccagggtgctctttgccgt
cagactcttccggtaacaagggtccgacttcgaaaatttttatgggcctggtaaaatc
ctatcgagttttcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag

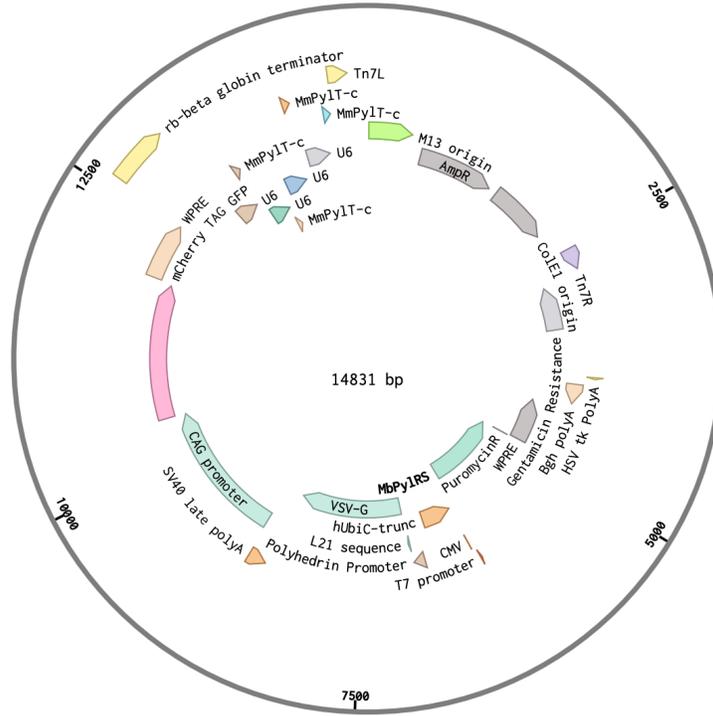
at t t t t a t c c a c c c g g a a g a t c t g t t t t g a a g t t c a g t a t c a t t a t t a a t a c c c a t t c
t c t c c a c g t a t t c c g c c g g a a t a a g g a t a g g a g a c t t t a t c t c c a g a a a a c c c c g g t c t
a c g a a a a a t t t c g t a a t a t c a c g t t c g a g t t t a c c g a g g t a g t c t t c t c t a t c a t t g g t
a t a g a g c c g c t g a a a a t c g t t t t t t c t t c t t g t c a c a a g t t c a g g c t c a a g t t c c c t g a
a a g g c t t t g c c a t a t t t t a g a g a a t t t t a t c c t c t g g a c t t a a g a g a g c c t c a a c c c t a
t c a a g c t g g c t t c t t g t a a g t g a a g g a g c a g g a g c c g a t g c g g g a a c a g a c g a a t t t g g
a g t t g a t t t t g c a g g c g a a g g t a c a g a t c t g g a t g t g t t c g t t g a t g c t t t t g c a g a a
c a g a a t t t t c c a g a g g c t t c g g a g c c c t t g a a a c t g a t t t c g g c a t a g c t t t t t t g a c c
t t t g g a g c a g a a a c t a c c c t a a c t t t c a c a c t g t t t t t g c t t t c g g t t g a t c t t g t g a g
a a a t t a t t g a t a t c c t c g t c c g a a a c c c t a c a t c g t t t g c a g g t t t t t c t g t a c t t a t
g a t g t c t g a a t g c t c t g g c t g t t c t a c a a c t c c t g g a a t t a t t c a c a a a a g a t g g t c t
c c a c a c g c c a t t t c a a t g t a t a t t t t a c t t c t t g a g a c c t c a t g g t g c t t g a t t t t g t g
g a g c g t g c c a g t c c t g g a c a t c c a g a g c c c g g t c g c a g a t a t t a a a c a t c t a a t g g t t
t t t t a t c c a t g g t g g c g t a g c c a g c t t g g g t c t c c c t a t a g t g a g t c g t a t t a t a g a c c t
g a t a a g c c a g t a a g c a g t g g g t t c t c t a g t t a g c c a g a g a g c t c t g c t t a t a t a g a c c t
c c c a c c g t a c a c g c c t a c c g c c a t t t g c g t c a a t g g g g c g g a g t t g t t a c g a c a t t t t
g g a a a g t c c c g t t g a t t t t g g t g c c a a a a c a a a c t c c c a t t g a c g t c a a t g g g g t g g a g
a c t t g g a a a t c c c c g t g a g t c a a a c c g c t a t c c a c g c c a t t g a t g t a c t g c c a a a a c c
g c a t c a c c a t g g t a a t a g c g a t g a c t a a t a c g t a g a t g t a c t g c c a a g t a g g a a a g t c c
c a t a a g g t c a t g t a c t g g g c a t a a t g c c a g g c g g g c a t t t a c c g t c a t t g a c g t c a a t
a g g g g g c g t a c t t g g c a t a t g a t a c a c t t g a t g t a c t g c c a a g t g g g c a g t t t a c c g t a
a a t a g t c c a c c c a t t g a c g t c a a t g g a a a g t c c c t a t t g g c g t t a c t a t g g g a a c a t a c
g t c a t t a t t g a c g t c a a t g g g c g g g g t c g t t g g g c g g t c a g c c a g g c g g g c a t t t a c
c g t a a g t t a t g t a a c g c g g a a c t c c a t a t a t g g g c t a t g a a c t a a t g a c c c c g t a a t t g
a t t a c t a t t a a t a a c t a g t c a a t a a t c a a t g t c a a c g c g t a t a t c t g g c c c g t a c a t c g
c g a a g c a g c g c a a a a c g g a t c c t g c a g g t a t t t g c g g c c g c g g t c c g t a t a c t c c g g a a
t a t t a a t a g a t c a t g g a g a t a a t t a a a a t g a t a a c c a t c t c g a a a t a a a t a a g t a t t
t a c t g t t t t c g t a a c a g t t t t g t a a t a a a a a a c c t a t a a a t a t t c c g g a t t a t t c a t a
c c g t c c c a c c a t c g g g c g c g a a c t c c t a a a a a a c c g c c a c c a t g a a g t g c c t t t t g t a c
t t a g c c t t t t t a t t c a t t g g g g t g a a t t g c a a g t t c a c c a t a g t t t t t c c a c a c a c c a
a a a a g g a a a c t g g a a a a t g t t c c t t c t a a t t a c c a t t a t t g c c c g t c a a g c t c a g a t t
t a a a t t g g c a t a a t g a c t t a a t a g g c a c a g c c t t a c a a g t c a a a a t g c c c a a g a g t c a c
a a g g c t a t t c a a g c a g a c g g t t g g a t g t g t c a t g c t t c c a a a t g g g t c a c t a c t t g t g a
t t t c c g c t g g t a t g g a c c g a a g t a t a t a a c a c a t t c c a t c c g a t c c t t c a c t c c a t c t g
t a g a a c a a t g c a a g g a a g c a t t g a c a a a c g a a a c a a g g a a c t t g g c t g a a t c c a g g c
t t c c c t c c t c a a a g t t g t g g a t a t g c a a c t g t g a c g g a t g c c g a a g c a g t g a t t g t c a
g g t g a c t c c t c a c c a t g t g c t g g t t g a t g a a t a c a c a g g a g a a t g g g t t g a t t c a c a g t
t c a t c a a c g g a a a t g c a g c a a t t a c a t a t g c c c c a c t g t c c a t a a c t c t a c a a c c t g g
c a t t c t g a c t a t a a g g t c a a a g g g c t a t g t g a t t c t a a c c t c a t t t c c a t g g a c a t c a c
c t t c t t c t c a g a g g a c g g a g a g c t a t c a t c c c t g g g a a g g a g g g c a c a g g g t c a g a a
g t a a c t a c t t t g c t t a t g a a a c t g g a g g c a a g g c c t g c a a a a t g c a a a c t g c a a g c a t
t g g g g a g t c a g a c t c c c a t c a g g t g t c t g g t t c g a g a t g g c t g a t a a g g a t c t c t t t g c
t g c a g c c a g a t t c c c t g a a t g c c c a g a a g g g t c a a g t a t c t c t g c t c c a t c t c a g a c c t
c a g t g g a t g t a a g t c t a a t t c a g g a c g t t g a g a g g a t c t t g g a t t a t t c c c t c t g c c a a
g a a c c t g g a g c a a a a t c a g a g c g g g t c t t c c a a t c t c t c c a g t g g a t c t c a g c t a t c t
t g t c c t a a a a a c c c a g g a a c c g g t c c t g c t t t c a c c a t a a t c a a t g g t a c c c t a a a a t
a c t t t g a g a c c a g a t a c a t c a g a g t c g a t a t t g c t g c t c c a a t c c t c t c a a g a a t g g t c
g g a a t g a t c a g t g g a a c t a c c a c a g a a a g g g a a c t g t g g g a t g a c t g g g c a c c a t a t g a

agacgtggaattggaccaatggagttctgaggaccagttcaggatataagtttcoct
tatacatgattggacatggtatggtggactccgatcttcatcttagctcaaaggctcag
gtgttcgaacatcctcacattcaagacgctgcttcgcaacttctgatgatgagagttt
atTTTTGGTgatactgggctatccaaaatccaatcgagcttgtagaaggttggttca
gtagttggaagctctattgcctctTTTTctttatcatagggttaatcattggacta
ttcttggttctccgagttggtatccatctttgcattaaattaagcacaccaagaaaag
acagatttatacagacatagagatgaaccgacttggaaagtgataaagtcgagaagtact
agaggatcataatcagccataccacattttagtagaggTTTTacttgcTTTAAAAacctc
ccacacctccccctgaacctgaaacataaaatgaatgcaattgTTTgTTgTTaacttgtt
tattgcagcttataatggttacaataaagcaatagcatcacaatttcacaaataaag
catttttttcaactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcat
gtctggatctgatcactgcttgagcctagttattaatagtaatcaattacggggtcatt
agtccatagccatataatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccacgacccccgccttgaagtcacaatgacgtatgTTccatagta
acgccaatagggaactttccattgaagtcacaatgggtggagtatttacggtaaaactgccc
cttggaagtcacatcaagtgtatcatatgccaagtaacgcccctattgaagtcacaatgacg
gtaaatggcccgcctggcattatgcccagtcacatgaccttattgggactttcctacttg
cagtacatctacgtattagtcacogctattaccatggctogaggtgagcccacgTTctg
cttcaactctccccctcccccccccccccccccaattttgtatttttttttttttt
aattttttgtgcagcagtggggggcccccccccccccccccccccccccccccccccccc
gccccccgagggggcc
gccccctccgaaagtttcttttatggcagggcgggcgggcgggcgggcctataaaaag
cgaagcgcggggggggagtcgctgcttgccttcgcccctgcccgcctccgccc
cgctccgcccgcgcccggcctgactgaccgcttactcccacaggtgaggggg
gggacggccttctcctccggcctgtaattagccttggTTTTaatgacggcctgTTTT
ttctgtggctgctgaaagccttaaaggcctccgggagggcctttgtgcccccccccccc
ggggctccgggggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg
tgcccggcggctgtgagcctgcc
ggaggggagcggggcc
aaggtgctgctgcc
gggctgtaaccc
gtgccccctccgtgcc
caggtgggggtgcc
gcc
tttatggtaatctgagagggggcc
gaaatctgggagggcc
cggcaggaagaaatgggcc
tccatctccagcctggggctgcc
gggggggggttccgttctggcgtgtgacggggggctctagagcctctgtaaccatggt
catgcctctctcttttttccacagctctgggcaacgtgctggttattgtgctgtctca
tcattttggcacaagaattggccaaggaggccaccatggtgagcaagggcgagggaggata
acatggccatcatcaaggagttcatgcttcaaggtgcacatggagggctccgtgaac
ggccacgagttcgagatcgagggcgagggcgagggcgcccctacgagggcaccagac
cgccaagctgaaggtgaccaaggtggccccctgcccctcgcctgggacatccctgtccc
ctcagttcatgtacggctccaagggcctacgtgaagcaccocgcccacatccccgactac
ttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcagaggacgg
cggcgtggtgaccgtgaccagggactcctcccctgcaggacggcgagttcatctacaagg
tgaagctgcggggcacaacttccccctccgacggcccccgtaatgcagaagaagaccatg

ggctgggaggcctcctccgagcggatgtaccccgaggacggcgcctgaagggcgagat
caagcagaggctgaagctgaaggacggcggccactacgacgctgaggtcaagaccact
acaaggccaagaagcccgtgcagctgccggcgcctacaacgtcaacatcaagttggac
atcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgcggaggccg
cactccaccggcggcatggacgagctgtacaagggcggctcttagggcggctccctcg
tgagcaagggcgaggagctgttcaccggggtggtgccatcctggctgagctggacggc
gacgtaaaccggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacgg
caagctgaccctgaagttcatctgcaccaccggcaagctgccctgcccggccacc
tcgtgaccaccctgacctacggcgtgcagtgttcagccgtaccccgaccacatgaag
cagcaogacttcttcaagtcgccatgcccgaaggctacgtccaggagcgcacctctt
cttcaaggacgacggcaactacaagaccgcgcggaggtgaagttcgagggcgacacc
tggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgggg
cacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaa
gaaccggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagc
tcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgccgac
aaccactacctgagcaccagtcgcacctgagcaaaagaccccaacgagaagcgcgatca
catggctcctgctggagttcgtgaccgccgcgggatcactctcggcatggacgagctgt
acaaggggccccttcgaacaaaaactcatctcagaagaggatctgaatatgcataccggt
catcatcaccatcaccattaatgaagaattcaacgcgtaagtcgacaatcaacctctgg
attacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgcta
tgtggatacgtgctttaatgcctttgtatcatgctattgcttcccgtatggccttcat
tttctcctccttgataaatcctggtgctgtctctttatgaggagttgtggcccgttg
tcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggc
attgccaccacctgtcagctcctttccgggacttctgcttccccctcctattgccc
ggcggaaactcatcgcgcctgacctgcccgtgctggacaggggctcggctgttgggca
ctgacaattccgtggtgttgcgggaaatcatcgtcctttccttggctgctcgcctgt
ggtgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatcc
agcggaccttccctcccgcgccctgctgcccgtctgcccgtcttccgcgtcttcgcc
ttcgcctcagacgagtcggatctccctttggggccgcctccccgcggcctggccggccg
ccagcacagtggctcgatcgaaccaatgccctggctcacaataccactgagatctttttc
cctctgccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaa
taaaggaaatttattttcattgcaatagtgtggttggaaatttttgtgtctctcactcgg
aaggacatatgggagggcaaatcatttaaacaatcagaatgagtatttggtttagagtt
tggcaacatatgcccatatgctggctgccatgaacaaaggttggctataaagaggtcat
cagtatatgaaacagccccctgctgtccattccttattccatagaaaagccttgacttg
aggttagatttttttatattttgttttgtgtattttttctttaacatccctaaaat
ttccttacatgtttactagccagattttctcctctcctgactactcccagtcata
gctgtccctcttctcttatggagatccctcgacctgccctaggggtaccctcgggcagga
agagggcctatttcccattgattccttcataatgtgcatatacagatacaaggctgtagag
agataattagaattaattgactgtaaacacaaagatattagtaaaaatacgtgacgt
agaaagtaataatcttgggtagttgacgttttaaaattatgttttaaaatggacta
tcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatacttgtgga
aaggacgaacaccgggcggtgatcatgtagatcgaacggactctaaatccgttcagc
cgggttagattcccgggctgcccgttttttctgtaggggtaccctcgggcaggaagagggc
ctatttccatgattccttcataatgtgcatatacagatacaaggctgtagagagataat
tagaattaatgtgactgtaaacacaaagatattagtaaaaatacgtgacgtagaaagt
aataatcttgggtagttgacgttttaaaattatgttttaaaatggactatcatatg
cttaccgtaacttgaaagtatttcgatttcttggctttatatacttgtggaagggacg

aaacaccgggoggctgatcatgtagatcgaacggactctaaatccgttcagccggggttagattcccgggctgccggttttttgctaggggtaccctcgggcaggaagagggcctatttccatgattccttcatatttgcataacgatacaaggctgtagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccgggoggctgatcatgtagatcgaacggactctaaatccgttcagccggggttagattcccgggctgccggttttttgctaggggtaccctcgggcaggaagagggcctatttccatgatccttcatatttgcataacgatacaaggctgtagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccgggoggctgatcatgtagatcgaacggactctaaatccgttcagccggggttagattcccgggctgccggttttttgctagggatccgaaccagataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttaogacgotacaoccagttcccatttttgtcaccttccctaataataccttaaaaactccatttccaccctcccagttcccaactattttgtccgcccacagcggggcatttttcttctggtatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttcggctactttttctctgtcacagaatgaaaattttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagc

pAcBac-UbiC-MbPylRS-CAG-mCherry-TAG-GFP-4xHTS25



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
UbiC promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
CAG promoter	text
mCherry TAG GFP	text
WPRE	text
rb-beta globin terminator	text
U6	text

MmPytR	text
Tn7L	text

Sequence

ctgaatggcgaatgggacgcgcctgtagcggcgcattaagcgcggcgggtgtgggtgt
tacgcgcagcgtgaccgctacacttgccagcgcctagcgcgcctcctttcgctttct
tcccttcccttctcgcacgttcgcggctttcccgtcaagctctaaatcgggggctc
ccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttg
agtccacgttctttaatagtggaactcttggtccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggatgttgccgatttcggcctattggttaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgtttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaaggaagagtatgagtattcaacatttcggtgctgcccttattcccttttttgogg
cattttgccttctggtttttgctcaccagaaacgctggtgaaagtaaagatgctgaa
gatcagttgggtgcacgagtggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgccccgaagaacgttttccaatgatgagcacttttaagttctgctat
gtggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgccgcatacac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtgataaacactgcggcca
actacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgacacaacatg
gggatcatgtaactcgccttgatcgttggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaa
ctggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttcggctggctggtttattgctgataa
atctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggttaactgtcagacca
agtttactcatatataacttttagattgatttaaaacttcatttttaatttaaaaggatct
aggtgaagatcctttttgataatctcatgaccaaatacccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttct
gcgcgtaatctgctgcttgcaaaaaaaaccaccgctaccagcgggtggtttgtttgc
cggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagcgttaggttagccaccactcaagaactctgtagc
accgcctacatacctcgtctgtctaactctgttaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggtcg
ggctgaacggggggttcgtgcacacagcccagcttgagcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggtaagcggcagggctcggaaacaggagagcgcacgagggagcttccaggg
gaaacgcctggatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg
atttttgtgatgctcgtcaggggggaggagcctatggaaaaacgccagcaacgcggcct
tttacggttcctggccttttgctggccttttgctcacatggtctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgcgcag
ccgaacgaccgagcgcagcagtcagtgagcaggaagcggaaagagcgcctgatgcggt
atcttctccttacgcacatctgtgcggtatctcacaccgcagaccagccgcgtaacctggc
aaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgtgggaggaca

ataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactaga
cagaatagttgtaaactgaaatcagtcagttatgctgtgaaaaagcactactggacttt
tggtatggctaaagcaactcttcattttctgaagtgcaaattgcccgctgtattaaag
agggcggtggccaagggcatggtaagactatattcgcgcggtgtgacaatttaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtactt
gggtcgatatcaaagtgcatacacttcttcccgtatgcccaactttgtatagagagccac
tgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccgaagcgcgttgg
cctcatgcttgaggagattgatgagcgcgggtggcaatgccctgcctccgggtgctcgcg
gagactgcgagatcatagatatagatctcactacgcggctgctcaaacctgggcagaac
gtaagccgcgagagcgcacaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatg
tcttactacggagcaagttcccggaggtaatcggagtcggctgatggtgggagtaggtg
gctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcgatggatttgacttg
gtcagggccgagcctacatgtgcaatgatgccatacttgagccacctaaactttgttt
tagggcgactgccctgctgcgtaacatcgttgcgctgcgtaacatcgttgcgctcca
taacatcaaacatcgaccacggcgtaacgcgcttgcgcttggatgcccgaggcatag
actgtacaaaaaacagtcataacaagccatgaaaaccgccactgcgcggttaccaccg
ctgcgcttcgggtcaaggttctggaccagttgcgctgagcgcatacgcctacttgcaatacag
tttacgaaccgaacaggcttatgtcaactgggttcgctgccttcatccggttccacgggtg
tgcgtaaccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcg
aacgagcgaaggttctcggctccacgcacatcgtcaggcattggcggccttgcgcttctt
ctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggc
cgctcgcggcgcttgcgggtgggtgctgaccccgatgaagtgggtcgcacctcgggttt
ctggaaggcgagcatcgtttgttcgccaggactctagctatagttctagtggttggct
acgtaccgtagtggcctatggcagggttgcgcttaatgcgccgctacagggcgcggtg
ggataccccctagagccccagctggttcttcccgcctcagaagccatagagcccaccgc
atccccagcatgcctgctattgtcttcccacatcctcccccttgcgctcctgccccacc
cccccccagaatagaatgacacctactcagacaatgcgatgcaatttctcattttat
taggaaaggacagtgaggagtgccaccttccagggtcaaggaaggcacgggggaggggca
aacaacagatggctggcaactagaaggcacagtcgaggctgatcagcgggtttaaaccg
gccctctagactcgagttaaagtcgacgcggggaggcggcccaaggagatccgactc
gtctgagggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcggga
aggaaggtccgctggattgagggccgaaggacgtagcagaaggacgtcccgcgcagaa
tccaggtggcaacacaggcagcagccaaggaaaggacgatgatttccccgacaacacc
acggaattgtcagtgcccaacagccgagccctgtccagcagcgggcaaggcaggcggc
gatgagttccgccgtggcaatagggagggggaaagcgaagtcgccggaaggagctgac
aggtgggtggcaatgcccccaaccagtggggggttgcgctcagcaaacacagtgcacaccacg
ccacgttgcctgacaacgggccacaactcctcataaagagacagcaaccaggatttata
caaggaggagaaaatgaaagccatacgggaagcaatagcatgatataaaaggcattaaag
cagcgtatccacatagcgtaaaaggagcaacatagtttaagaataccagtcfaatcttca
caaattttgtaatccagaggttgattgtcgcacttaacgcggttgaattcttacagattgg
ttgaaatcccattatagtaagattcggaccttgatgccctcttaattgtttttaagccg
tgcataaccttgagcaagcgttcaagaccaaaccctgcacctatccatggtttgcacat
acccattctctatcaagagaaactggcccagcactgccgaagaaagctccaggtccc
cgtgcattatatcaagagtatccccaaagaccatacaggaatctcctacgatttcgaag
tcgatttccagatagtcagaaactctttagatgagagcttcaagattttcccaggtaca
tcccgaaccatctgacagaagttcaccatagtaaattcttccagggtgctctttgcccgt
cagactcttccggtaacaaggtccgacttcgaaaatttttatgggcctggtaaaatc
ctatcgagttttcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag

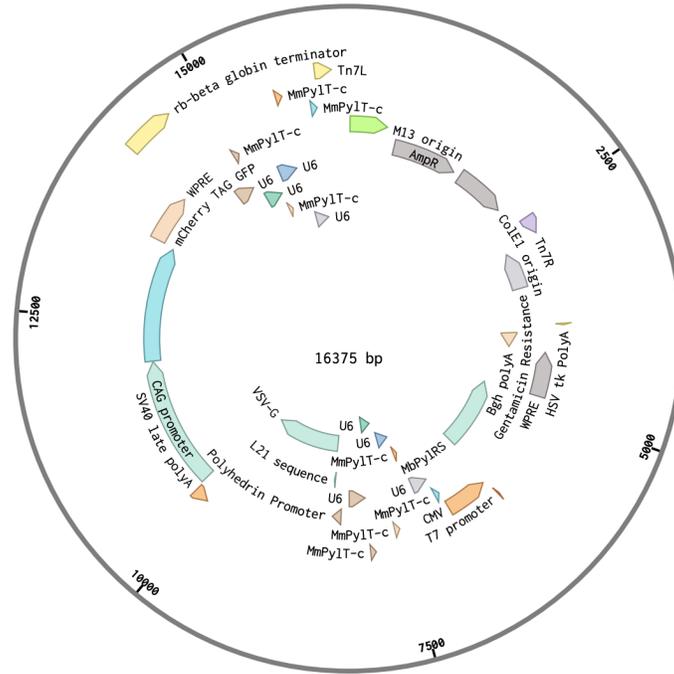
atTTTTatccaccggaagatctgTTTTgaaagttcagtatcattattaataccattc
tctccacgtattccgccggaataaggataggagactttatctccagaaaaccccggtc
acgaaaaatttcgtaatatcacgttcgagttaccgaggtagtcttctctatcattggt
atagagccgctgaaaatcgTTTTtcttctgtcacaagttcaggctcaagttccctga
aaggctttgccatatttagagaaattttatcctctggacttaagagagcctcaacccta
tcaagctggcttcttgtaagtgaaggagcaggagccgatgcgggaacagacgaatttg
agttgattttgcaggcgaaggtacagatctggatgtgttcggtgatgcctttgcagaaa
cagaattttccagaggcttcggagcccttgaaactgatttcggcatagctTTTTtgacc
tttgagcagaaactaccctaactttcacactgTTTTgctttcgggtgatcttgtgag
aaaattattgatatcctcgtccgaaaccctacatcgtttgacggtTTTTctgtacttat
gatgtctgaatgctctggctgttctacaactcctggaattattcacaacaagatggct
ccacacgccatttcaatgtatattttacttcttgagacctcatgggtgcttgattttgtg
gagcgtgccagtcctggacatccagagcccggtcgcagatattaaacatctaattggtt
TTTTatccatggtggcgtagccagcttgggtctcccTatagtgagtcgtattaatttc
gataagccagtaagcagtggttctctagtttagccagagcctagaccaagtgcgac
acagcgatccacaaacaagaaccgcgacccaaatcccggctgcgacggaactagctgtg
ccacaccggcgctccttatataatcatcggcggttcaccgccccacggagatccctcc
gcagaatcgccgagaagggactacttttctcctgctgttccgctctctggaagaaaac
cagtgccctagagtcaccaagtcccgctcctaaaatgtccttctgctgatactggggt
ctaaggccgagcttatgagcagcgggcccgtgtcctgagcgtccggggcgaaggatca
ggacgctcgtgcgacctcgtctgacgtggcagcgtcgcgctgaggagggggggcgcc
cgcgggaggcgccaaaaccggcgcgaggccactagtcaataatcaatgtcaacgcgt
atatctggcccgtacatcgcgaagcagcgcgaaaacggatcctgcaggtatttgccggccg
cggtcctataactccggaatattaatagatcatggagataaataaaatgataaccatct
cgcaataaataagtttttactgTTTTcgtaacagttttgtaataaaaaaacctataa
atattccggattattcataccgtcccaccatcggggcgcaactcctaaaaaacggccac
catgaagtgccttttgtacttagcctTTTTtattcattggggtgaattgcaagttcacca
tagTTTTccacacaacaaaaaggaaactggaaaaatgttcttctaattaccattat
tgcccgtcaagctcagatttaaattggcataatgacttaataggcacagccttacaagt
caaatgcccaagagtcacaaggctattcaagcagacgggttgatgtgtcatgcttcca
aatgggtcactacttgtgatttccgctggatggaccgaagtatataacacattccatc
cgatccttactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaacaagg
aacttggctgaatccaggcttccctcctcaaagttgtggatgcaactgtgacggatg
ccgaagcagtgattgtccagggtgactcctcaccatgtgctgggtgatgaatacacagga
gaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgcccactgt
ccataactctacaacctggcattctgactataagggtcaaagggtatgtgattctaacc
tcatttccatggacatcaccttcttctcagaggacggagagctatcatccctgggaaag
gagggcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaa
aatgcaatactgcaagcattggggagtcagactcccacaggtgtctgggttcgagatgg
ctgataaggatctctttgctgcagccagattccctgaatgccagaagggtcaagtatc
tctgctccatctcagacctcagtgatgtaagtctaattcaggacggtgagaggatctt
ggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctc
cagtggatctcagctatcttgcctcctaaaaccaggaaccggctcctgctttcaccata
atcaatggtaccctaaaataactttgagaccagatacatcagagtcgatattgctgctcc
aatcctctcaagaatggctcggaatgatcagtggaactaccacagaaagggaaactgtggg
atgactgggcaccatataagacgtggaaattggaccaatggagttctgaggaccagt
tcaggatataagtttctttatacatgattggacatgggatgttgactccgatcttca
tcttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaac

ttcctgatgatgagagtttatttttttgggtgatactgggctatccaaaaatccaatcgag
cttgtagaagggttggttcagtagttggaaaagctctattgcctcttttttctttatcat
agggttaatcattggactattcttgggttctccgagttggatccatctttgcattaat
taaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttggaaag
tgataa|gtc|gaga|agtactagaggatcataatcagccataccacattt|gtagagg|ttt
acttgcttt|aaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaa
ttg|ttgt|ttg|ttaact|tg|tttatt|gcag|cttataat|gg|ttaca|aataaag|caatag|catc
aca|aat|ttcaca|aataaag|catt|tttt|tact|gcatt|ctag|ttgt|gg|ttt|gtc|caaact
cat|caat|gtat|cttat|cat|gt|ct|ggat|ct|gat|cact|gct|tgag|**ct**ag|tt|atta|tag|t
aat|caat|ta|c|gg|gg|tcatt|ag|tt|ca|tag|ccc|ata|tat|gg|ag|tt|cc|gg|tt|ac|ata|act|t
ac|gg|ta|aat|gg|ccc|g|cct|gg|gt|gac|c|g|cc|ca|ac|g|a|cccc|c|g|cc|catt|gac|gt|ca|ata|at
gac|gt|at|gt|t|cc|cat|ag|ta|ac|g|cca|at|agg|g|act|tt|cc|att|gac|gt|ca|at|gg|gt|gg|agt
att|ac|gg|ta|aa|act|g|cc|c|act|t|gg|c|agt|ac|at|ca|agt|gt|at|ca|tat|g|cc|a|agt|ac|g|ccc
cct|att|gac|gt|ca|at|gac|gg|ta|aat|gg|ccc|g|cct|gg|c|att|at|g|cc|agt|ac|at|gac|ct|t
at|gg|g|act|tt|cc|ta|c|tt|gg|c|agt|ac|at|c|ta|c|gt|att|ag|tc|at|c|g|ct|att|acc|at|gg|tc|g
ag|gt|gac|ccc|ac|gt|tt|ct|g|ct|c|act|ct|cccc|at|ct|cccc|cccc|ct|cccc|cccc|ca|at|t
tt|gt|att|t|att|t|tt|tt|ta|att|tt|t|gt|gc|ag|c|gat|gg|gg|gg|gg|gg|gg|gg|gg|gg|gg|g|g|c
g|c|g|c|g|c|c|agg|c|gg|gg|c|gg|gg|c|gg|gg|c|g|ag|gg|c|gg|gg|c|g|ag|gg|c|gg|gg|g|ag|gg|t|g|c|g|
g|c|g|g|c|ag|cca|at|c|ag|ag|c|gg|c|g|c|g|ct|c|c|g|aa|ag|ttt|c|ct|tt|t|at|g|g|c|g|ag|gg|c|g|g|g|c|g|
g|c|g|g|c|gg|cc|tata|aaaa|ag|c|g|aa|g|c|g|c|g|g|g|c|gg|gg|c|g|g|g|ag|tc|g|ct|g|c|gt|t|g|c|ct|tc|g|c|
cc|gt|g|cccc|g|ct|cc|g|c|g|cc|c|ct|c|g|c|c|c|c|c|g|cc|c|c|c|g|c|c|c|t|c|t|g|act|gac|c|g|c|gt|ta
ct|ccc|ac|agg|tg|ag|c|gg|g|c|gg|g|ac|gg|c|ct|t|c|c|ct|c|c|gg|g|c|t|g|ta|at|tag|c|gt|t|gg|t|
tt|aat|gac|gg|ct|c|g|t|tt|ct|tt|ct|g|t|gg|ct|g|c|gt|g|aa|ag|c|ct|ta|a|agg|g|ct|c|c|gg|g|ag|g|
g|cc|ct|tt|gt|g|c|gg|gg|gg|g|ag|c|gg|ct|c|gg|gg|gg|t|g|c|gt|g|c|gt|g|t|g|t|g|t|g|c|gt|g|gg|g|ag|
c|g|c|g|c|gt|g|c|gg|ccc|c|g|c|gt|g|c|cc|gg|c|gg|c|t|g|t|g|ag|c|g|ct|g|c|gg|g|c|g|c|gg|c|g|c|gg|g|c|t|
tt|gt|g|c|gt|c|c|g|c|gt|g|t|g|c|g|c|g|ag|gg|g|ag|c|g|c|g|c|c|gg|gg|g|c|gg|t|g|cccc|c|g|gg|t|g|c|g|g|
gg|gg|g|ct|g|c|g|ag|gg|g|aa|ca|aa|agg|ct|g|c|gt|g|c|gg|gg|t|g|t|g|t|g|c|gt|g|gg|gg|gg|t|g|ag|c|agg|
gg|gt|g|t|gg|g|c|g|c|gg|g|t|c|gg|g|ct|g|ta|a|cccc|cccc|ct|g|c|a|cccc|cccc|ct|ccc|g|agt|t|g|ct|
g|ag|c|ag|c|gg|cc|c|g|gt|t|c|gg|gt|g|c|gg|gg|ct|c|c|gt|g|c|gg|gg|g|c|gt|gg|c|g|c|gg|gg|c|t|g|c|c|gt|
g|cc|gg|g|c|gg|gg|gt|gg|c|gg|c|agg|t|gg|gg|gt|g|c|c|gg|g|c|gg|gg|c|gg|gg|c|c|c|ct|c|gg|g|c|c|g|
gg|g|ag|gg|ct|c|gg|gg|g|ag|gg|g|c|g|c|gg|c|cccc|gg|ag|c|c|c|g|gg|c|t|g|tc|g|ag|g|c|g|c|g|
c|g|ag|ccc|g|ag|cc|att|g|c|ct|tt|at|gg|ta|at|c|gt|g|c|g|ag|ag|gg|c|g|c|ag|gg|act|t|c|ct|tt|g|
t|ccc|aa|at|ct|g|g|c|gg|ag|c|c|g|aa|at|ct|g|gg|ag|g|c|c|c|c|c|c|c|c|c|ct|c|t|ag|c|gg|g|c|g|c|
gg|g|g|aa|g|c|gg|t|g|c|gg|c|g|c|gg|c|ag|ga|ag|aa|at|gg|g|c|gg|g|ag|gg|c|ct|t|c|gt|g|c|gt|g|c|
c|c|g|c|g|c|c|c|g|t|cccc|t|ct|c|cat|ct|cc|ag|c|ct|c|gg|gg|ct|g|c|c|g|c|ag|gg|gg|ag|c|gg|ct|g|c|
ct|tc|gg|gg|gg|g|ac|gg|gg|c|ag|gg|c|gg|gg|tt|c|gg|ct|t|c|t|gg|c|gt|g|t|gac|c|gg|c|gg|ct|c|t|ag
ag|c|ct|c|t|g|ta|acc|at|g|tt|c|at|g|c|ct|t|ct|t|t|tt|ct|ac|ag|ct|c|ct|gg|g|ca|ac|gt|g|c|
tg|gt|t|att|gt|g|ct|g|t|c|at|c|at|tt|tt|gg|caa|aga|att|gg|cca|agg|agg|cc|acc|at|gg|gt|g|
ag|ca|agg|g|c|g|ag|g|ag|g|ata|aac|at|gg|c|at|c|at|ca|agg|agt|t|c|at|g|c|g|ct|t|ca|agg|t|g|ca
cat|gg|ag|gg|ct|cc|gt|ga|ac|gg|cc|ac|g|agt|t|c|g|ag|at|c|g|ag|gg|c|g|ag|gg|c|g|ag|gg|c|g|c|c|
c|ct|ac|g|ag|gg|c|acc|c|ag|acc|g|cca|ag|ct|ga|ag|gt|g|acca|agg|gt|gg|cccc|ct|g|c|c|ct|c|
g|c|ct|gg|g|ac|at|c|ct|gt|cccc|ct|c|agt|t|c|at|gt|ac|gg|ct|c|ca|agg|c|ct|ac|gt|ga|ag|c|acc|c|
c|g|cc|g|ac|at|cccc|g|act|act|t|ga|ag|ct|gt|c|ct|t|cccc|g|ag|gg|ct|t|ca|agt|gg|g|ag|c|g|c|g|
t|gat|ga|act|t|c|g|ag|g|ac|gg|c|gg|c|gt|gg|t|ga|oc|gt|g|ac|cc|agg|act|c|ct|c|c|ot|g|c|ag|g|ac|
gg|c|g|ag|tt|c|at|c|t|aca|ag|gt|ga|ag|ct|g|c|g|c|gg|c|acca|act|t|cccc|t|c|c|g|ac|gg|cccc|ct|
aat|g|c|aga|aga|ag|acc|at|gg|g|ct|gg|g|ag|g|c|ct|c|ct|c|c|g|ag|c|gg|at|gt|ac|ccc|g|ag|g|ac|g|
g|c|g|c|c|ct|ga|agg|g|c|g|ag|at|ca|ag|c|ag|ag|g|ct|ga|ag|ct|ga|agg|ac|gg|c|gg|c|c|act|ac|g|ac|
g|ct|g|ag|gt|ca|ag|acc|ac|t|aca|agg|cca|aga|ag|ccc|gt|g|c|ag|ct|g|cccc|gg|c|g|c|ct|aca|a|

cgteaacatcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagt
acgaacgcgccgagggccgcccactcccaccggcgccatggacgagctgtacaagggcggc
tcttagggcggctcccctcgtgagcaagggcgaggagctgttcaccggggtggtgccat
cctggtcgagctggacggcgacgtaaaccggccacaagttcagcgtgtccggcgagggcg
agggcgatgccacctacggcaagetgacctgaagttcatctgcaccaccggcaagctg
cccgtgccctggcccacctcgtgaccacctgacctacggcgtgcagtgettacgccg
ctaccccgaccacatgaagcagcagcacttcttcaagtccgccatgcccgaaggctacg
tccaggagcgcaccatcttcttcaaggacgacggcaactacaagaccgcgccgaggtg
aagttcgagggcgacaccctgggtaaccgcacatcgagctgaagggcactgcacttcaagga
ggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgctctata
tcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatc
gaggacggcagcgtgcagctcgcggaccactaccagcagaacacccccatcggcgacgg
ccccgtgctgctgcccgacaaccactacctgagcaccacagtcgccctgagcaaaagacc
ccaacgagaagcgcgatcacatggctcctgctggagttcgtgaccgcgccgggatcact
ctcggcatggacgagctgtacaaggggccccttcgaacaaaaactcatctcagaagagga
tctgaatatgcataccgggtcatcaccatcaccattaatgaatcaacgcggttaa
gtcgcacaaatcaacctctggattacaaaatttgtgaaagattgactgggtattcttaacta
tgttgctccttttacgctatgtggatacgtgctttaaagcctttgtatcatgctattg
cttcccgtatggctttcattttctcctccttgataaaatcctggttgctgtctctttat
gaggagttgtggcccgttgtcaggcaacgtggcgtgggtgtgactgtgtttgctgacgc
aacccccactgggtggggcattgccaccacctgtcagctcctttccgggactttcgctt
tccccctcctattgccacggcggaactcatcgccgctgccttggccgctgctggaca
ggggctcggctgttgggcaactgacaattccgtgggtgttgcgggaaatcatcgtcctt
tccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacg
tcccttccggccctcaatccagcggaccttccctccgcggcctgctgccggctctgcgg
cctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctttgggcccgcctc
cccgcggcctggccggccgcccagcacagtggtcgatcgaaccaatgccctggctcaca
aaatccactgagatctttttccctctgccccaaaattatggggacatcatgaagccccttga
gcatctgacttctggotaataaaggaaatttattttcattgcaatagtgtgttgaatt
ttttgtgtctctcactcgggaaggacatatgggagggcaaatcatttaaacatcagaat
gagtatttgggttagagtttggcaacatatgccatatgctggctgccatgaacaaagg
ttggctataaagaggtcatcagtatatgaaacagccccctgctgtccattccttattcc
atagaaaagccttgacttgaggttagatttttttatattttgttttggttatttttt
tctttaacatccctaaaattttccttacatgttttaactagccagatttttccctctc
ctgactactcccagtcatagetgtccctcttctcttatggagatccctcgacctgcct
aggggtaccctcgggcaggaagagggcctatttcccatgattccttcatatttgcatata
cgatacaaggctgtagagagataattagaattaatttgactgtaaacacaaagatatt
agtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgagttttaa
tatgttttaaataatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttg
gctttatatacttgtggaaggacgaaacaccgggcccgtgatcatgtagatcgaacg
gactctaaatccgttcagccgggttagattcccgggctgcccgttttttgctaggggta
cctcgggcaggaagagggcctatttcccatgattccttcatatttgcatatacgataca
aggctgtagagagataattagaattaatttgactgtaaacacaaagatatttagtaca
aatcgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaattatgttt
taaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttat
atatcttgtggaaggacgaaacaccgggcccgtgatcatgtagatcgaacggactcta
aatccgttcagccgggttagattcccgggctgcccgttttttgctaggggtaccctcggg
caggaagagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgt

tagagagataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgt
gacgtagaaagtaataatttcttgggtagtttgacgtttttaaattatgtttttaaagt
gactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatactt
gtggaaaggacgaaacaccgggcggtgatcatgtagatcgaacggactctaaatccgt
tcagccgggttagattcccgggctgcccgttttttgctaggggtaccctcgggcaggaag
aggcctatttcccatgattccttcatatttgcataacgatacaaggctgtagagag
ataattagaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtag
aaagtaataatttcttgggtagtttgacgtttttaaattatgtttttaaattggactatc
atatgcttaccgtaacttgaaagtatttcgatttcttggctttatatacttgtggaaa
ggacgaaacaccgggcggtgatcatgtagatcgaacggactctaaatccgttcagccg
ggtagattcccgggctgcccgttttttgctaggagatccgaaccagataagtgaatc
tagttccaaactattttgtcatttttaattttcgtattagcttacgacgctacaccag
ttcccatctattttgtcactcttccctaataatccttaaaaactccatttccaccct
cccagttcccaactattttgtccgccacagcggggcatttttcttctgttatgtttt
taatcaaacatcctgccaactccatgtgacaaaccgtcatcttcggctactttttctct
gtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaa
tatcaacgcttatttgcagc

pAcBac-CMV-MbPylRS-CAG-mCherry-TAG-GFP-8xPylR



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
MbPylRS	text
T7 promoter	text
CMV promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
CAG promoter	text
mCherry TAG GFP	text
WPRE	text
rb-beta globin terminator	text
U6	text
MmPylR	text

Sequence of HTS25 variant

ctgaatggcgaatgggacgcgcctgtagcggcgcattaagcgcggcgggtgtggtggt
tacgcgcagcgtgaccgctacacttgccagcgccttagcgcggcgcctcctttcgctttct
tccttcctttctcgcacgttcgcggctttcccgcgtaagctctaaatcgggggctc
ccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccccttgacgttg
agtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatc
tcggtctattcttttgatttataagggattttgccgatttcggcctattgggttaaaaa
tgagctgatttaacaaaaatttaacgcgaatttttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaaccctatttgtttatttttctaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg
aaaaaggaagagtatgagtattcaacatttccgtgctgccttattcccttttttgccg
cattttgccttctggtttttgctcaccagaaacgctggtgaaagtaaagatgctgaa
gatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcct
tgagagttttcgcccgaagaacgttttccaatgatgagcacttttaagttctgctat
gtggcgcggtattatcccgtattgacgcgggcaagagcaactcggctgcgcgcatcac
tattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtgataaacactgcccga
actacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgacacaacatg
gggatcatgtaactcgcttgatcggtgggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaacattaa
ctggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttccggctgggtttattgctgataa
atctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagacca
agtttactcatatatactttagattgatttaaaacttcatttttaatttaaaggatct
aggtgaagatcctttttgataatctcatgaccaaatacccttaacgtgagttttcgttc
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttct
gcgcgtaatctgctgcttgcaaaaaaaaccaccgctaccagcgggtggtttgtttgc
cggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagccgtagttaggccaccactcaagaactctgtagc
accgctacatacctcgctctgctaatacctggtaccagtggtgctgccagtgccgata
agtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctc
ggctgaacggggggttcgctgcacacagcccagcttgagcgaacgacctacaccgaact
gagatacctacagcgtgagcattgagaaagcgcacgcttcccgaaggagaaaggcgg
acaggtatccggaagcggcagggtcggaacaggagagcgcacgagggagcttccaggg
ggaaacgcctggtatctttatagtcctgctcgggtttcgccacctctgacttgagcgtcg
atttttgtgatgctcgtcagggggggcggagcctatggaaaaacgccagcaacgcggcct
ttttacggttcctggccttttgctggccttttgctcacatggtctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgcgcgag
ccgaacgaccgagcgcagcaggtcagtgagcaggaagcgggaagagcgcctgatgcgg
attttctccttacgcacatctgtgcggtatctcacaccgcagaccagccgcgtaacctggc
aaaatcggttacggttgagtaataaatggatgcctgcgtaagcgggtgtggggcggaca
ataaagtcttaactgaacaaaatagatctaaactatgacaataaagtcttaactaga

cagaatagttgtaaactgaaatcagtcagttatgctgtgaaaaagcatactggacttt
tggtatggctaaagcaaactcttcattttctgaagtgcaaattgcccgctgtattaaag
agggcggtggccaagggcatggtaaagactatattcgcggcggtgtgacaattaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtactt
gggtcgatatcaaagtgcatacacttcttcccgatgcccaactttgtatagagagccac
tgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccaagcgcgttgg
cctcatgcttgaggagattgatgagcgcggtggcaatgccctgcctccggtgctcggc
gagactgcgagatcatagatatagatctcactacgcggctgctcaaactgggcagaac
gtaagccgcgagagcgcacaaccgcttcttggtcgaaggcagcaagcgcgatgaatg
tcttactacggagcaagttcccgaggtaatcggagtccggctgatggtgggagtaggtg
gctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcattgacttg
gtcagggccgagcctacatgtgcgaatgatgccatacttgagccactaactttgttt
tagggcgactgccctgctgcgtaacatcgttgcgctgcgtaacatcgttgcgctcca
taacatcaaacatcgaccacggcgtaacgcgcttgcgcttggatgcccgaggcatag
actgtacaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttaccaccg
ctgcgctcgggtcaagggtctggaccagttgcgtgagcgcatacgtacttgcatcacg
tttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccacgggtg
tgcgtcacccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcg
aacgagcgaaggtttccggtctccacgcacgcagcagcattggcggccttgcgttctt
ctacggcaagggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggc
cgtcgcggcgttgcgggtgggtgctgaccccgatgaagtgggttcgcatcctcggttt
ctggaaggcagcatcgtttgttcgccaggactctagctatagttctagtggttggct
acgtaccgtagtggctatggcagggcttgcgcttaatgcgccgctacagggcgcgtgg
ggataccccctagagccccagctggttcttccgcctcagaagccatagagcccaccgc
atccccagcatgcctgctattgtcttcccaatcctcccccttgcgtgctcgtccccacc
cccccccagaatagaatgacacctactcagacaatgcgatgcaatttctcattttat
taggaaaggacagtgaggagtgacacctccagggtcaaggaaggcacgggggaggggca
aacaacagatggctggcaactagaaggcacagtcgaggctgatcagcgggtttaaaccg
gccctctagactcgagttaaagtcgacgccccggaggcggcccaaggagatccgactc
gtctgagggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcggga
aggaagggtccgctggattgagggccgaaggacgtagcagaaggacgtcccgcgcagaa
tccagggtggcaacacaggcagcagccaaggaaaggacgatgatttccccgacaacacc
acggaattgtcagtgcccaacagccgagcccctgtccagcagcgggcaaggcaggcggc
gatgagttccgccgtggcaatagggagggggaaagcgaagtcccggaaaggagctgac
aggtgggtggcaatgccccaccagtggggggttgcgtcagcaaacacagtgcacaccacg
ccacgttgcctgacaacgggcccacaactcctcataaagagacagcaaccaggatttata
caaggaggagaaaatgaaagccatacgggaagcaatagcatgatataaaggcattaaag
cagcgtatccacatagcgtaaaaggagcaacatagtttaagaataccagtcfaatcttca
caaattttgtaatccagaggttgattgtcgaacttaacgcgttgaattcttacagattgg
ttgaaatcccattatagtaagattcggaccttgatgccctcttaatgtttttaagccg
tgcataaccttgagcaagcgttcaagacccaaaacctgcacctatccatggtttgc
acccattctctatcaagagaaactggcccagcactgccgaagaaagctccagggtccc
cgtgcattatatacaagagtatccccaaagaccatacaggaatctcctacgatttcgaag
tcgatttccagatagtcagaaactctttagatgagagcttcaagattttcccgagtaca
tcccgaaccatctgacagaagttcaccatagtaaattcttccagggtgctctttgccgt
cagactcttccggtaacaagggtccgacttcgaaaatttttatgggcctggtaaaatc
ctatcgagttttcgcagatagttgtaaagagtcggggcaagcattggcctcaagcagag
atttttatccaccgggaagatctgttttgaaagttcagtatcattattaataccattc

tctccacgtattccgcccgaataaggataggagactttatctccagaaaaccccggctc
acgaaaaatctcgtaatatcacgttcgagtttaccgaggtagtcttctctatcattggt
atagagccgctgaaaatcgtttttcttctgtcacaagttcaggctcaagttccctga
aaggctttgccatatttagagaaatcttctctggacttaagagagcctcaacccta
tcaagctggcttctgtgaagtgaaggagcaggagccgatgcccgaacagacgaatttg
agttgattttgcaggcgaaggtacagatctggatgtgttcggtgatgcctttgcagaa
cagaatctccagaggcttcggagcccttgaactgatttcggcatagcttttttgacc
tttgagcagaaactaccctaactttcacactgtttttgctttcgggtgatcttgtgag
aaaattattgatatcctcgtccgaaaccctacatcgtttgaggtttttctgtacttat
gatgtctgaatgctctggctgttctacaactcctggaattattcacaacaagatggctc
ccacacgccatttcaatgtatattttacttcttgagacctcatgggtgcttgattttgtg
gagcgtgccagtcctggacatccagagcccggctgcagatattaaaacatctaattggt
tttatccatggtggcgctagccagcttgggtctccc**tatagtgagtcgtatta**aatttc
gataagccagtaagcagtggttctctagttagccagagagctctgcttatatagacct
cccaccgtacacgcctaccgcccatttgcgtaaatggggcggagttgttacgacatttt
ggaaagtcccgttgattttggtgcccacaaactcccattgacgtcaatgggggtggag
acttggaatcccgtgagtcacaccgctatccacgccattgatgtactgcccacacc
gcatcaccatggtaatagcgatgactaatacgtagatgtactgccaagtaggaagtc
cataaggtcatgtactgggcataatgccaggcgggcccatttaccgtcattgacgtcaat
agggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttacgta
aatagtcacaccattgacgtcaatggaaagtcctattggcgttactatgggaacatac
gtcattattgacgtcaatggggggggctggtggggcgtcagccaggcgggcccatttac
cgtaagttatgtaacgcccgaactccataatgggctatgaactaatgacccgtaattg
attactattaataactagcaaaaa**cgggcagcccgggaatctaaccggctgaacgga**
tttagagtcctgatctacatgatcagccgcccgggtgtttcgtcctttccacaagat
atataaagccaagaatacgaataactttcaagttacggtaagcatatgatagtccattt
taaacataattttaaaactgcaactacccaagaaattattactttctacgtcacgta
ttttgtactaataatctttgtgtttacagtcacaaattaattctaattatctcttaacagc
cttgtatcgtatatacgaatatgaaggaatcatgggaaataggccctcttctcctgcccga
ggtagccctagcaaaaa**cgggcagcccgggaatctaaccggctgaacggatttagag**
tccgttcgatctacatgatcagccgcccgggtgtttcgtcctttccacaagatataaaa
gccaagaatacgaataactttcaagttacggtaagcatatgatagtccattttaaaca
taattttaaaactgcaactacccaagaaattattactttctacgtcacgtaattttgta
ctaataatctttgtgtttacagtcacaaattaattctaattatctcttaacagccttgat
cgtatatacgaatatgaaggaatcatgggaaataggccctcttctcctgcccgaggtacc
ctagcaaaaa**cgggcagcccgggaatctaaccggctgaacggatttagagtcctg**
gatctacatgatcagccgcccgggtgtttcgtcctttccacaagatataaagccaaga
aatcgaataactttcaagttacggtaagcatatgatagtccattttaaacataatctt
aaaactgcaactacccaagaaattattactttctacgtcacgtaattttgtactaata
ctttgtgtttacagtcacaaattaattctaattatctcttaacagccttgatcgtat
gcaaatatgaaggaatcatgggaaataggccctcttctcctgcccgaggtaccctagcaa
aaaa**cgggcagcccgggaatctaaccggctgaacggatttagagtcctg**
atgatcagccgcccgggtgtttcgtcctttccacaagatataaagccaagaatacga
aactttcaagttacggtaagcatatgatagtccattttaaacataatctttaaactg
caactacccaagaaattattactttctacgtcacgtaattttgtactaataatctttgtg
tttacagtcacaaattaattctaattatctcttaacagccttgatcgtatatacgaata
tgaaggaatcatgggaaataggccctcttctcctgcccgaaggtaccctagtcataatca
atgtcaacgcgtatctgcccgtacatcgccaagcagcgaacacggatcctgcagg

tatttgccgcccgcgggtccgtataactccggaatattaatagatcatggagataaattaa
tgataaacatctcgcaataaataagtattttactgttttcgtaacagttttgtaataa
aaaaacctataaatattccggattattcataccgtcccaccatcgggcggaactccta
aaaaccgccaccatgaagtgccttttgtacttagcctttttattcattgggggtgaatt
gcaagttcaccatagtttttccacacaacaaaaaggaaactggaaaaatgttccttct
aattaccattattgcccgtcaagctcagattttaattggcataatgacttaataggcac
agcettacaagtcaaaatgcccaagagtcacaaggctattcaagcagacgggttgatgt
gtcatgcttccaaatgggtcactacttgtgatttccgctggatggaccgaagtatata
acacattccatccgatccttccactccatctgtagaacaatgcaaggaaagcattgaaca
aacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaa
ctgtgacggatgccgaagcagtgattgtccagggtgactcctcaccatgtgctggttgat
gaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacat
atgccccactgtccataactctacaacctggcattctgactataagggtcaaagggtat
gtgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatca
tccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaactggagg
caaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccacaggtgtct
ggttcgagatggctgataaggatctctttgctgcagccagattccctgaatgccagaa
gggtcaagtatctctgctccatctcagacctcagtgatgtaagtctaattcaggacgt
tgagaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtc
ttccaatctctccagtgatctcagctatcttggctcctaaaaaccaggaaccggctct
gctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcga
tattgctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaa
gggaactgtgggatgactgggcaccatatgaagacgtggaatggaccaatggagtt
ctgaggaccagttcaggatataagtttctttatacatgattggacatggtatggtgga
ctccgatcttcatcttagctcaaaggctcagggtgttcgaacatcctcacattcaagacg
ctgcttcgcaactcctgatgatgagagtttatttttggtgatactgggctatccaaa
aatccaatcgagcttgtagaaggttgggtcagtagttggaaaagctctattgcctctt
tttctttatcatagggttaatcattggactattcttggttctccgagttggatccatc
tttgcattaaattaagcacaccaagaaaagacagatttatacagacatagagatgaac
cgacttggaagtgataagtcgagaagtactagaggatcataatcagccataccacatt
tgtagaggttttacttggcttttaaaaaacctcccacacctccccctgaacctgaaacata
aatgaatgcaattggtggtgtaacttgtttatttgcagcttataatggttacaataa
agcaatagcatcacaatttcacaaataaagcatttttttactgcattctagttgtgg
tttgtccaaactcatcaatgtatcttatcatgtctg gatctgatcactgcttgagccta
gttattaatagtaatcaattacgggggtcattagttcatagcccataatggagttccgc
gttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgccatt
gacgtcaataatgacgtatgttcccataagtaacgccaatagggactttccattgacgtc
aatgggtggagtatttacggtaaaactgccacttggcagtaacataaagtgtatcatatg
ccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccc
gtacatgaccttatgggactttcctacttggcagtaacatacgtattagtcategcta
ttaccatggtcgaggtgagccccacgttctgcttccactctccccatctccccccccctcc
ccaccccccaattttgtatttatttatttttaattattttgtgcagogatggggggcggg
ggggggggggggcgcgcgcagggcggggcggggcggggcggggcggggcggggcggggcggg
cggagaggtgoggoggcagccaatcagagcgggcgcgcctccgaaagtttccctttatggc
gaggcggcggcggcggcggcctataaaaagcgaagcgcgcggcggcggggcgggagtcgctg
cgttgccttcgccccgtgccccgcctccgcgcgcctccgcgcgcgcgcgcgcgcgcgcgcgc
ctgaccgcgttactcccacaggtgagcggggcggggacggccctctcctccgggctgtaa
ttagoccttggtttaatgaoggetogtttottttotgtggctgogtgaaagcettaag

ggctcgggagggcccttggcgggggggagcggctcggggggtgctgogtgtgtgt
gtgctggggagcgcgcgtgcgcccgcgctgccggcgctgtgagcgtgcgggcg
cggcgcggggttggcgctccgctgtgcgcgaggggagcggccggggcggtgc
cccgggtgoggggggctgcgaggggaacaaaggctgogtgcgggggtgtgtgogtgg
gggtgagcaggggggtgtggcgogggcggtcgggtgtaacccccccctgcaccccc
ccccgagttgctgagcagcccggttcgggtgcggggtccgtgcggggctggcgc
gggctcgcctgcccgggggggggtggcgccaggtgggggtgccggggcgggcggggc
cgctcgggcccgggagggctcgggggagggcgcgccggccccggagcgcggcggt
gtcgagggcgggcgagccgcagccattgcottttatggtaatcgtgogagagggogcag
ggaecttccttggtoocaaatctggcgagccgaaatctgggagggcgccgcaccccc
ctagcggggcggggogaagcgtgogggcgccggcaggaaggaaatggcggggagggc
ctcgtgctgcccgcgcgcctcccttctccatctccagcctcggggctgcccgcag
gggacggctgccttcgggggggacggggcagggcggggttcggcttctggcgtgtgac
cggcgctctagagcctctgctaaccatgctcatgccttcttctttctacagctcc
tgggcaacgtgctgggttattgtgctcctcatttttggcaagaattggccaaggag
gccaccatgggtgagcaagggcgagggagataacatggccatcatcaaggagttcatg
cttcaagggtcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagg
gcgagggccgcccctacgagggcaccagaccgccaagctgaaggtagcaagggtggc
cccctgcccttcgcttgggacatcctgtcccctcagttcatgtacggctccaaggccta
cgtgaagcaccocgcgacatccccgactacttgaagctgtccttccccgagggcttca
agtgggagcgcgtgatgaacttcgaggacggcgcgctggtgaccgtgaccaggaactcc
tccctgcaggacggcgagttcatctacaaggtagagctgcgcgccaccaactcccctc
cgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgt
accccgaggacggcgccctgaaggggogagatcaagcagaggctgaagctgaaggacggc
ggccactacgacgctgaggtcaagaccactacaaggccaagaagcccgtgcagctgcc
cggcgccctacaacgtcaacatcaagttggacatcacctcccacaacgaggactacacca
tcgtggaacagtacgaacgcgcccaggggccccactccaccggcgggcatggacgagctg
tacaagggcggtctcttagggcggtccctcgtgagcaagggcgaggagctgttcaccgg
gggtggtgccatcctggctgagctggacggcgagcgtaaacggccacaagttcagcgtgt
ccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctgcacc
accggcaagctgccctgcccggccaccctcgtgaccacctgacctacggcgtgca
gtgcttcagccgctaccccgaccacatgaagcagcagcacttcttcaagtccgccatgc
ccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacc
cgcgcgaggtgaagttcgagggcgacaccctggtgaaccgcacgcagctgaagggcac
cgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagcc
acaacgtctatatcatggccgacaagcagaagaacggcatcaagggtgaacttcaagatc
cgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaaccccc
catcggcgacggccccgctgctgctgcccgacaaccactacctgagcaccagctccgccc
tgagcaaaagaccccaacgagaagcgcgatcacatggctcctgctggagttcgtgacggcc
gccgggatcactctcggcatggacgagctgtacaaggggcccctcgaacaaaaactcat
ctcagaagaggatctgaatatgcataccggctcatcatcaccatcaccattaatgagaa
tcaacgcggttaagtcgacaaatcaacctctggattacaaaatttgtgaaagattgactgg
tattcttaactatgttgcctcttttacgctatgtggatacgtgctttaaagcctttgt
atcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctgggtg
ctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgactgt
gtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccg
ggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttggc
cgctgctggacaggggctcggctgttggggactgacaattccgtgggtgttgcggggaa

atcatogtcctttccttggctgctcgctgtggtgccacctggattctgcgcgggacgt
ccttctgctacgtccctcggccctcaatccagcggaccttccctcccgcggcctgctg
ccggctctgcggcctcttccgcgtcttcgccttcgacctcagacgagtcggatctccct
ttgggcgcctccccgcggcctggccggccgacacagtggtcgatcgaaccaatgcc
ctggctcacaataaccactgagatctttttccctctgccaaaaattatggggacatcat
gaagccccttgagcatctgacttctggctaataaaggaaattttttcattgcaatag
tgtgttggaattttttgtgtctctcactcggaaggacatatgggagggcaaatcattta
aaacatcagaatgagtatttggtttagagtttggcaacatatgcccatatgctggctgc
catgaacaagggtggctataaagggtcatcagtatatgaaacagccccctgctgtcc
attccttattccatagaaaagccttgacttgaggttagatttttttataattttgtttt
gtgttattttttctttaacatccctaaaattttccttacatgttttactagccagatt
ttcctcctcctcactgactactccagtcatactgtccctcttctcttatggagatccc
tcgacctgccctaggggtaccctgggcaggaagaggcctatttcccatgattccttca
tatttgcatatacgatacaaggctgtagagagataattagaattaatttgactgtaaa
cacaagatatttagtacaataacgtgacgtagaaagtaataatttcttgggtagtttg
cagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagta
ttcgatttcttggctttatatacttgtggaaggacgaaacaccgggcggctgatcat
gtagatcgaacggactctaaatccgttcagccgggtagattcccgggctgcccgtttt
ttgctaggggtaccctgggcaggaagaggcctatttcccatgattccttcatatttgc
atatacgatacaaggctgtagagagataattagaattaatttgactgtaaacacaaag
atatttagtacaataacgtgacgtagaaagtaataatttcttgggtagtttgcagtttt
aaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtaattcgatt
tcttggctttatatacttgtggaaggacgaaacaccgggcggctgatcatgtagatc
gaacggactctaaatccgttcagccgggtagattcccgggctgcccgttttttgctag
gggtaccctgggcaggaagaggcctatttcccatgattccttcatatttgcatacag
atacaaggctgtagagagataattagaattaatttgactgtaaacacaaagatattag
tacaataacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaatta
tgttttaaaatggactatcatatgcttaccgtaacttgaaagtaatttgcatttcttggc
tttatatacttgtggaaggacgaaacaccgggcggctgatcatgtagatcgaacgga
ctctaaatccgttcagccgggtagattcccgggctgcccgttttttgctaggggtacc
tcgggcaggaagaggcctatttcccatgattccttcatatttgcatacagatacaag
gctgtagagagataattagaattaatttgactgtaaacacaaagatatttagtacaata
tacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgtttta
aaatggactatcatatgcttaccgtaacttgaaagtaatttgcatttcttggctttatata
atcttgtggaaggacgaaacaccgggcggctgatcatgtagatcgaacggactctaaa
tccgttcagccgggtagattcccgggctgcccgttttttgctaggagatccgaaccag
ataagtgaatctagttccaaactattttgtcatttttaattttcgtattagcttacga
cgctacaccagttcccatctattttgtcactcttccctaaataatccttaaaaactcc
atttccaccctcccagttcccaactattttgtccgccacaagcggggcattttttcttc
ctgttatgtttttaatcaaacatcctgccaaactccatgtgacaaaccgtcatcttcggc
tactttttctctgtcacagaatgaaaattttctgtcatctcttcggtattaatgtttg
taattgactgaatatcaacgcttatttgcagc

The sequence for the wild type PytR is:
ggaaacctgatcatgtagatcgaacggactctaaatccgttcagccgggtagattccc
ggggtttccg

The sequence for the NES-MbPyIRS is:

atggcctgccagtgccccttcagttgcctccgctggagagactgaccctcgacgataa
aaaaccattagatgttttaatatctgcgaccgggctctggatgtccaggactggcacgc
tccacaaaatcaagcaccatgaggtctcaagaagtaaaatatacattgaaatggcgtgt
ggagaccatcttggtgtgaataattccaggagttgtagaacagccagagcattcagaca
tcataagtacagaaaaacctgcaaacgatgtagggtttcggacgaggatatcaataatt
ttctcacaagatcaaccgaaagcaaaaacagtgtaaggttagggtagtttctgctcca
aaggtcaaaaagctatgccgaaatcagtttcaagggctccgaagcctctggaaaattc
tgtttctgcaaaggcatcaacgaacacatccagatctgtaccttcgcctgcaaaatcaa
ctccaaattcgtctgttcccgcacatcggtcctgctccttcaactacaagaagccagctt
gataggggtgaggctctcttaagtccagaggataaaatctctaaatatggcaaagcc
tttcagggaaacttgagcctgaacttgtgacaagaagaaaaaacgattttcagcggctct
ataccaatgatagagaagactacctcggtaaactcgaacgtgatattacgaaattttc
gtagaccggggtttctggagataaagtctcctatccttattccggcggaaatcgtgga
gagaatgggtattaataatgatactgaactttcaaacagatcttccgggtggataaaa
atctctgcttgaggccaatgcttgccccgactctttacaactatctgcaaaaactcgat
aggattttaccaggccaataaaaaatttcgaagtcggaccttgttaccggaaagagtc
tgacggcaagagcacctggaagaatttactatgggtgaacttctgtcagatgggttcgg
gatgtactcgggaaaatcctgaagctctcatcaagagtttctggactatctggaaatc
gacttcgaaatcgtaggagattcctgtatggcttttggggatactcttgatataatgca
cggggacctggagctttcttcggcagtcgtcgggccagtttctcttgatagagaatggg
gtattgacaaaccatggataggtgcaggttttggcttgaacgcttgctcaaggttatg
cacggctttaaaaacattaagagggcatcaaggtccgaatcttactataatgggatttc
aaccaatctgtaa

The sequence for NBK-RS is:

atggataaaaaaccattagatgttttaatatctgcgaccgggctctggatgtccaggac
tggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatacattgaaa
tggcgtgtggagaccatcttggtgtgaataattccaggagttgtagaacagccagagca
ttcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcggacgaggatat
caataattttctcacaagatcaaccgaaagcaaaaacagtgtaaggttagggtagttt
ctgctccaaaggtcaaaaagctatgccgaaatcagtttcaagggctccgaagcctctg
gaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgtaccttcgcctgc
aaaatcaactccaaattcgtctgttcccgcacatcggtcctgctccttcaactacaaga
gccagcttgataggggtgaggctctcttaagtccagaggataaaatctctaaatatg
gcaaagcctttcagggaaacttgagcctgaacttgtgacaagaagaaaaaacgattttca
gaggctctataccaatgatagagaagactacctcggtaaactcgaacgtgatattacga
aatttttcgtagaccggggtttctggagataaagtctcctatccttattccggcggaa
tacgtggagagatttgggtattaataatgatactgaactttcaaacagatcttccgggt
ggataaaaatctctgcttgaggccaatgcttagcccgactctttgcaactatatgcaaa
aactcgataggattttaccaggccaataaaaaatttcgaagtcggaccttgttaccgg
aaagagcttgacggcaagagcacctggaagaatttactatgggtgaacttctgtcagat
gggttcgggatgtactcgggaaaatcctgaagctctcatcaagagtttctggactatc
tggaatcgaacttcgaaatcgtaggagattcctgtatggcttttggggatactcttgat
ataatgcacggggacctggagctttcttcggcagtcgtcgggccagtttctcttgatag
agaatggggatttgacaaaccatggataggtgcaggttttggcttgaacgcttgctca
aggttatgcacggctttaaaaacattaagagggcatcaaggtccgaatcttactataat
gggattttcaaccaatctgtaa

The sequence for CouK-RS is:

```
atggataaaaaaccattagatggttttaatatctgcgaccgggctctggatgtccaggac
tggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatacattgaaa
tggcgtgtggagaccatcttggtgtgaataattccaggagttgtagaacagccagagca
ttcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcggacgaggatat
caataattttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagttt
ctgctccaaaggtcaaaaaagctatgccgaaatcagtttcaagggctccgaagcctctg
gaaaattctggtttctgcaaaggtatcaacgaacacatccagatctgtaccttcgcctgc
aaaatcaactccaaattcgtctgttcccgcacatcggtcctgctccttcaacttacaagaa
gccagcttgataggggtgaggctctcttaagtcagaggataaaaatttctctaaatag
gcaaagcctttcagggaaacttgagcctgaacttgtgacaagaagaaaaaacgattttca
gcggtctataccaatgatagagaagactacctcggtaaaactcgaacgtgatattacga
aattttctgtagaccggggttttctggagataaagtcctcctatccttattccggcggaa
tacgtggagagaatgggtattaataatgatactgaactttcaaacagatcttccgggt
ggataaaaatctctgcttgaggccaatgcttgccccgactcttgccaactatatgcgaa
aactcgataggattttaccaggccaataaaaattttcgaagtcggaccttggtaccgg
aaagagtctgacggcaagagcacctggaagaatttactatggtgaacttctgtcagat
gggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagtttctggactatc
tggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgat
ataatgcacggggacctggagctttcttcggcagtcgtcgggccagtttctcttgatag
agaatggggtattgacaaaccatggataggtgcagggttttgggtcttgaacgcttgctca
aggttatgcacggcttttaaaaacattaagagggcatcaaggtccgaatcttactataat
gggatttcaaccaatctgtaa
```

The sequence for AcK-RS is:

```
atggataaaaaaccattagatggttttaatatctgcgaccgggctctggatgtccaggac
tggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatacattgaaa
tggcgtgtggagaccatcttggtgtgaataattccaggagttgtagaacagccagagca
ttcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcgggtgaggatat
caataattttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagttt
ctgctccaaaggtcaaaaaagctatgccgaaatcagtttcaagggctccgaagcctctg
gaaaattctggtttctgcaaaggtatcaacgaacacatccagatctgtaccttcgcctgc
aaaatcaactccaaattcgtctgttcccgcacatcggtcctgctccttcaacttacaagaa
gccagcttgataggggtgaggctctcttaagtcagaggataaaaatttctctaaatag
gcaaagcctttcagggaaacttgagcctgaacttgtgacaagaagaaaaaacgattttca
gcggtctataccaatgatagagaagactacctcggtaaaactcgaacgtgatattacga
aattttctgtagaccggggttttctggagataaagtcctcctatccttattccggcggaa
tacgtggagagaatgggtattaataatgatactgaactttcaaacagatcttccgggt
ggataaaaatctctgcttgaggccaatggtggccccgactatttttaactatgcgcgaa
aactcgataggattttaccaggccaataaaaattttcgaagtcggaccttggtaccgg
aaagagtctgacggcaagagcacctggaagaatttactatggtgaacttcttccagat
gggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagtttctggactatc
tggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgat
ataatgcacggggacctggagctttcttcggcagtcgtcgggccagtttctcttgatag
agaatggggtattgacaaaccatggataggtgcagggttttgggtcttgaacgcttgctca
aggttatgcacggcttttaaaaacattaagagggcatcaaggtccgaatcttactataat
gggatttcaaccaatctgtaa
```

The sequence for BiotinK-RS is:

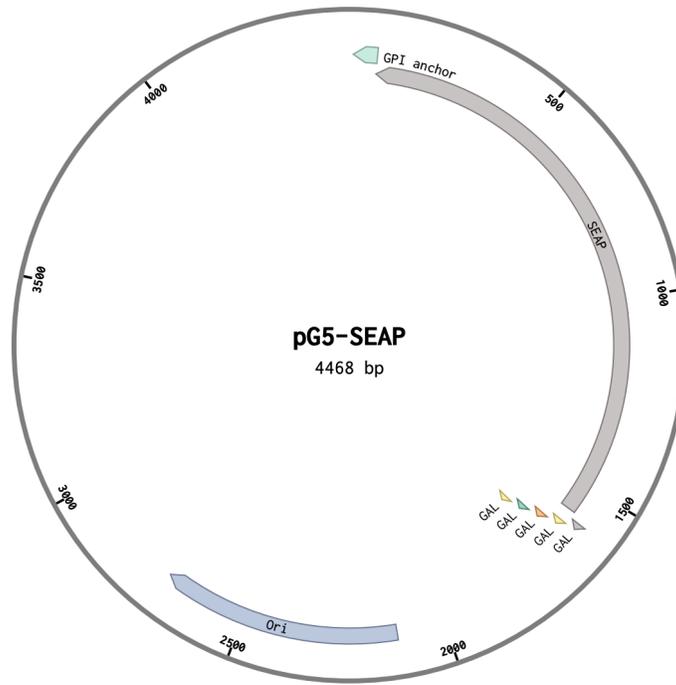
atggataaaaaaccattagatggttttaatatctgcgaccgggctctggatgtccaggac
tggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatacattgaaa
tggcgtgtggagaccatcttggtgtgaataattccaggagttgtagaacagccagagca
ttcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcggacgaggatat
caataattttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagttt
ctgctccaaaggtcaaaaaagctatgccgaaatcagtttcaagggtccgaagcctctg
gaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgtaccttcgcctgc
aaaatcaactccaaattcgtctgttcccgcacatcggtcctgctccttcacttacaagaa
gccagcttgataggggtgaggctctcttaagtccagaggataaaaatttctctaaatag
gcaaagcctttcaggaacttgagcctgaacttgtgacaagaagaaaaaacgattttca
gcggtctatataccaatgatagagaagactacctcggtaaaactcgaacgtgatattacga
aattttctgtagaccggggttttctggagataaagtctcctatccttattccggcggaa
tacgtggagagaatgggtattaataatgatactgaactttcaaacagatcttccgggt
ggataaaaatctctgcttgaggccaatgcttgccccgactcttgcaactatgtgcaaa
aactcgataggattttaccaggccaataaaaattttcgaagtcggaccttgttaccgg
aaagagtctgacggcaagagcacctggaagaatttactatggtgaacttcgctcagat
tggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagtttctggactatc
tggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgat
ataatgcacggggacctggagctttcttcggcagtcgctcgggccagttctcttgatag
agaatggggtattgacaaaccatggataggtgcaggttttgggtcttgaacgcttgctca
aggttatgcacggcttttaaaaacattaagagggcatcaaggtccgaatcttactataat
gggatttcaaccaatctgtaa

The sequence for DiazK-RS is:

atggataaaaaaccattagatggttttaatatctgcgaccgggctctggatgtccaggac
tggcacgctccacaaaatcaagcaccatgaggtctcaagaagtaaaatatacattgaaa
tggcgtgtggagaccatcttggtgtgaataattccaggagttgtagaacagccagagca
ttcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcggacgaggatat
caataattttctcacaagatcaaccgaaagcaaaaacagtgtgaaagttagggtagttt
ctgctccaaaggtcaaaaaagctatgccgaaatcagtttcaagggtccgaagcctctg
gaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgtaccttcgcctgc
aaaatcaactccaaattcgtctgttcccgcacatcggtcctgctccttcacttacaagaa
gccagcttgataggggtgaggctctcttaagtccagaggataaaaatttctctaaatag
gcaaagcctttcaggaacttgagcctgaacttgtgacaagaagaaaaaacgattttca
gcggtctatataccaatgatagagaagactacctcggtaaaactcgaacgtgatattacga
aattttctgtagaccggggttttctggagataaagtctcctatccttattccggcggaa
tacgtggagagaatgggtattaataatgatactgaactttcaaacagatcttccgggt
ggataaaaatctctgcttgaggccaatgcttgccccgactctgtataactatgtgcaaa
aactcgataggattttaccaggccaataaaaattttcgaagtcggaccttgttaccgg
aaagagtctgacggcaagagcacctggaagaatttactatggtgaacttcgctcagat
gggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagtttctggactatc
tggaaatcgacttcgaaatcgtaggagattcctgtatggtctttggggatactcttgat
ataatgcacggggacctggagctttcttcggcagtcgctcgggccagttctcttgatag
agaatggggtattgacaaaccatggataggtgcaggttttgggtcttgaacgcttgctca
aggttatgcacggcttttaaaaacattaagagggcatcaaggtccgaatcttactataat
gggatttcaaccaatctataa

Chapter 5 Plasmids

pG5-SEAP



Sequence color-coding key

Feature	Color
GPI anchor	text
SEAP	text
GAL	text
GAL	text
Ori	text

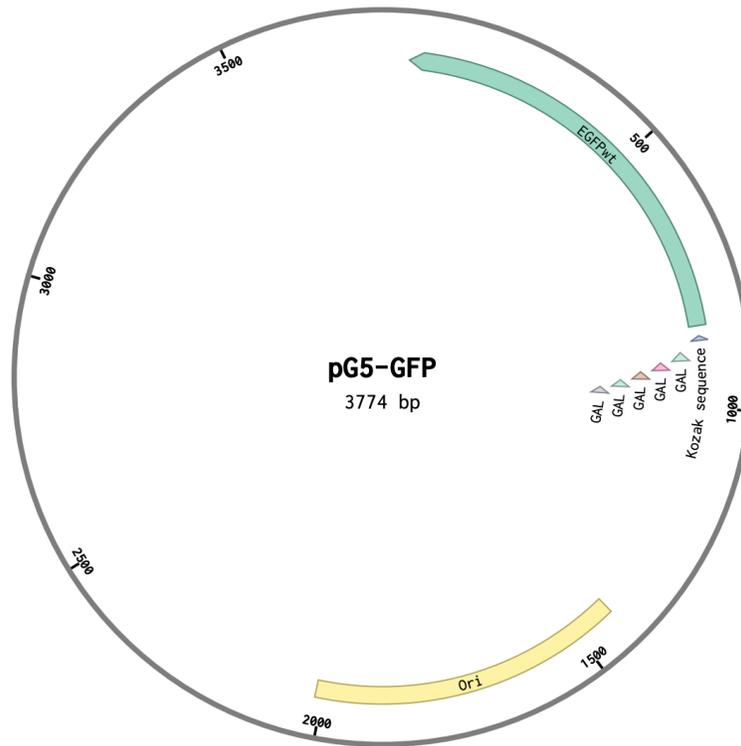
Sequence

```

caatgtatcttatcatgtctgctcgaagcggccggccgccccgactctagagtaacc
ggcgcgcggcgtcgggtggtgcccggggggcgccaggctcgcaggcgggtgtagggctc
agcaggcggcgaaggccatgacgtgcgctatgaaggctctgctcctgcacgccgtgaac
caggtgcgcctgcgggcccgcgcgaacaccgccacgtcctcgccctgcgtgggtctctt
cgtccaggggcaactgctgactgctgcccatactcggggctcccgctctcgcctcggta
acatccggccggggcgcgctccttgagcacatagcctggaccggtttccgatataggaggac
cgtgtaggccttccctgtcccgggccttgccagggggccagcccgaagatggagctccctc
gcaggggtagcctccgaaggagaagacgtgggagtggtcggcagtgacgaggctcagc
gtgtcctcctcgtggtgagctggcccgcctctcaatggcgtcgtcgaacatgatcgt
ctcagtcagtgcccggtaagccctgctttcatgatgaccatggtcgatgcgaccacct
    
```


tggtgagtactcaaccaagtcattctgagaatagtgatgogggcgaccgagttgctctt
gcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatc
attggaaaacggttcttcggggcgaaaactctcaaggatcttaccgctggtgagatccag
ttcgatgtaaccactcgtgcaccaactgatcttcagcatctttactttcaccagcg
tttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaagggaataagggcgaca
cggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcaggg
ttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaatatagggg
ttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatg
acattaacctataaaaataggcgtatcacgaggccctttcgtctcgcgcggtttcggtga
tgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgctgtaag
cggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtggtggcgggtgctcg
ggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccatagcgggtg
tgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattca
ggctgcgcaactggtgggaagggcgatcgggtcgggcctcttcgctattacgccagctg
gcaaggggggatgtgctgcaagcgattaagttgggtaacgccagggttttcccagtc
acgacggtgtaaacgacggccagtgaaatcgagctcggtagcttgttgtaacttgtt
tattgcagcttataatggttacaataaagcaatagcatcacaatttcacaaataaag
cattttttcactgcattctagttgtggtttgtccaaactcat

pG5-GFP



Sequence color-coding key

Feature	Color
eGFP	text
Kozak sequence	text
GAL	text
GAL	text
Ori	text

Sequence

caatgtatcttatcatgtctgctcgaagcggccggccgccccgactctagatcattaat
 ggtgatggtgatgatgaccggtatgcatattcagatcctcttctgagatgagttttgt
 tcgaagggccccttgtacagctcgtccatgocgagagtgatcccggcggcggtcaogaa
 ctccagcaggaccatgtgatcgcgcttctcgttggggcttttgctcagggcggactggg
 tgctcaggtagtggttgctcgggcagcagcacggggccgctcggcgatgggggtgttctgc
 tggtagtggtcggcgagctgcacgctgccgtcctcgatggtgtggcggatcctgaagtt
 caccttgatgocgcttcttctgcttgctcggccatgatatagacgcttggtggtggtgtagt
 tgtaactccagcttggtgccccaggatggtgccgtcctccttgaagtogatgcccttcagc
 tcgatgcggttcaccagggtgtcgcctcgaacttcacctcggcgcgggtccttgtagtt
 gccgtcgtccttgaagaagatggtgcgctcctggacgtagccttcgggcatggcggact
 tgaagaagtcgtgctgcttcatgtggtcgggtagcggctgaagcactgcacgccgtag
 gtcaggggtggtcacgaggggtgggccagggcacgggcagcttgccggtggtgcagatgaa

cttcagggtcagcttgccgtaggtggcatcgccctcgccctcgccggacacgctgaact
tgtggccggtttacgtcgccgtccagctcgaccaggatgggcaccaccccggtgaacagc
tcctcgcccttgctcaccatgggtggcggcgctagccagcgggtggcggcgcatgctagag
tcgacggatcggagtagtctcctccgagcggagtagtctcctcgagcggagtagtctc
tccgagcggagtagtctcctccgagcggagtagtctcctccgagcggatccccgggaat
tagcttggcgtaatcatggtcataagctgthttcctgtgtgaaattgthtatccgctcaca
ttccacacaacatacgagccggaagcataaagtgtaaagcctgggggtgcctaattgagtg
agctaactcacattaattgcggtgctgctcactgcccgctttccagtcgggaaacctgct
gtgccagctgcattaatgaatcgcccaacgcgcggggagaggcgggttgcgtattgggc
gctcttccgcttctcctcgtcactgactcgtcgcgctcggtcggtcggctgocggcgagcg
gtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcagg
aaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgc
tggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagt
cagaggtggcgaaccccgacaggactataaagataaccaggcgtttcccctggaagctc
cctcgtgcgctctcctggtccgaccctgccgcttaaccggataacctgtccgcctttctcc
cttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtag
gtcgttcgctccaagctgggctgtgtgcacgaacccccggttcagcccagccgctgocg
cttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatccgactgg
cagcagccactggtaacaggattagcagagcggaggtatgtaggcgggtgctacagagttc
ttgaagtgggtggcctaactacggctacactagaagaacagtatttggtatctcgcgctc
gctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaaacaacca
ccgctggttagcgggtgggttttttgtttgcaagcagcagattacgcgcagaaaaaaagga
tctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactc
acgttaagggtatttgggtcatgagattatcaaaaaggatcttcacctagatccttttaa
atataaaatgaagttttaaatacaatctaaagtataatgagtaaaccttgggtctgacagt
taccatgcttaatcagtgaggcacctatctcagcagatctgtctatctcgttcatccat
agttgcctgactccccgctcgtgtagataactacgatacgggagggcttaccatctggcc
ccagtgctgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaata
aaccagccagccggaaggccgagcgcagaagtggctcctgcaactttatccgcctccat
ccagttatataattggtgcccgggaagctagagtaagtagttccgagttaatagtttgc
gcaacggtgttgccattgctacaggcatcgtgggtgacgctcgtcgtttggtagggct
tcattcagctccgggtcccaacgatcaaggcaggttacatgatccccatggtgtgcaa
aaaagcgggttagctccttcgggtcctccgatcgtgtgcagaagtaagttggccgcagtg
tatcactcatgggtatggcagcactgcataattctcttactgtcatgccatccgtaaga
tgcttttctgtgactgggtgagtagtcaaccaagtcattctgagaatagtgtagcggcg
accgagttgctcttggccggcgtcaatacgggataataaccgcgccacatagcagaactt
taaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccg
ctggtgagatccagttcgatgtaaccactcgtgcaccaactgatcttcagcatcttt
tactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaag
gaataaggcgcacacggaaatggtgaatactcatactcttcctttttcaatattattga
agcatttatcagggttattgtctcatgagcggatacatatttgaatgtatthagaaaa
taacaaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaa
ccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtctc
gocggtttcgggtgatgacgggtgaaaacctctgacacatgcagctcccggagacggtcac
agcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtg
ttggcgggtgctcggggctggcttaactatgcggcatcagagcagattgtactgagagtg
caccatagcgggtgtgaaataccgcacagatgcgtaaggagaaaataccgcacagggc
ccattcgcattcaggctgcgcaactgttgggaaggccgatcgggtgocggcctcttcgc

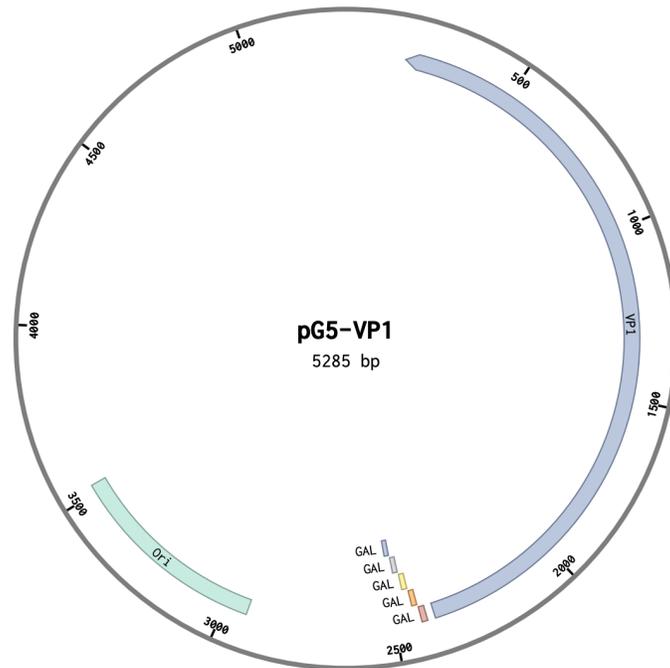
tattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgcca
gggttttcccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacctt
gttgttaacttgtttattgcagcttataatggttacaataaagcaatagcatcacaaa
tttcacaaataaagcatttttttctactgcattctagttgtggtttgtccaaactcat

actgtcctccgagcggagtactgtcctcagcggagtactgtcctccgagcggagtact
gtcctccgagcggagtactgtcctccgagcggatccccgggaattagcttggcgtaac
atggatcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatac
gagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacatta
attgcggtgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcatta
atgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttct
cgctcactgactcgctgcgctcggtcgctcggctgcggcgagcggatcagctcactca
aaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagc
aaaaggccagcaaaaaggccaggaaccgtaaaaaggccgcttgctggcgcttttccata
ggctccgccccctgacgagcatcacaanaatcgaogctcaagtcagaggtggcgaaac
ccgacaggactataaagataaccaggcgtttccccctggaagctccctcgtgcgctctcc
tgctccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtgg
cgctttctcatagctcacgctgtaggatctcagttcgggtgtaggtcgttcgctccaag
ctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcttataccggtaacta
tcgctttgagtcacaaccggtaagacacgaactatogccactggcagcagocactggta
acaggattagcagagcgggtatgtaggcgggtctacagagttcttgaagtgggtggct
aactacggctacactagaagaacagtatgttggtatctgcgctctgctgaagccagttac
cttcggaaaaagagttggtagctcttgatccggcaaaaaccaccgctggtagcgggtg
gtttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcct
ttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggtttt
ggatcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagt
ttaaatacaatctaaagtatatatgagtaaacttggctctgacagttaccaatgcttaac
agtgaggcacctatctcagcgatctgtctatctcgttccatccatagttgcctgactccc
cgtcgtgtagataactacgatacgggagggttaccatctggccccagtgctgcaatga
taccgcgagaccacgctcaccggctccagattatcagcaataaaccagccagccgga
agggccgagcgcgagaagtggctcctgcaactttatccgcctccatccagctctattaattg
ttgcccgggaagctagagtaagttagttcgccagttaatagtttgcgcaacggttgttgcca
ttgctacaggcatcgtgggtgtcacgctcgtcgtttggatggcttcattcagctccggg
tcccaacgatcaaggcgagttacatgatccccatggtgtgcaaaaaagcgggttagctc
cttcggctcctccgatcgttgtcagaagtaagttggccgagtggtatcactcatgggta
tggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgact
ggtagtactcaaccaagtcattctgagaatagtgatgcggcgaccgagttgctcttg
cccggcgtcaatacgggataataaccgcgccacatagcagaactttaaagtgtcatca
ttggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctggtgagatccagt
tcgatgtaaccactcgtgcaccaactgatcttcagcatcttttactttcaccagcgt
ttctgggtgagcaaaaacaggaaggcaaaaatgccgcaaaaaaggggaataagggcgacac
ggaaatggtgaatactcactcttctcttttcaatattattgaagcatttatcagggt
tattgtctcatgagcggatacatatgttgatgtattagaaaaataaacaataggggt
tccgcgcacatttccccgaaaagtgcacctgacgtctaagaaaccattattatcatga
cattaacctataaaaataggcgtatcacgaggccctttcgtctcgcgcgtttcgggtgat
gacgggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttgtctgtaagc
ggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggg
gctggcttaactatgcccgcacagagcagattgtactgagagtgaccatattgcgggtg
gaaataccgcacagatgcgtaaggagaaaataccgcacagggccattcggcattcag
gctgcgcaactggtgggaagggcgatcgggtgcgggctcttcgctattacgccagctgg
cgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggttttccagctca
cgacggtgtaaaacgacggccagtgaaatccagactcg

The sequence for the YBTATA minimal promoter is:
tctagagggatatataatgggggcca

The sequence for the minCMV minimal promoter is:
Gtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatc

pG5-VP1



Sequence color-coding key

Feature	Color
VP1	text
GAL	text
Ori	text

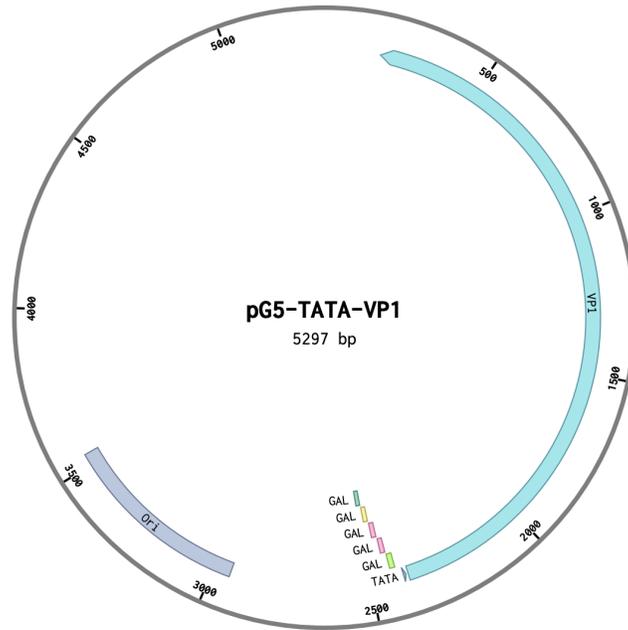
Sequence

caatgtatcttatcatgtctgctcgaagcggcggccgcccgactctagagctgtagt
 taatgattaaccgcatgctacttatctacgtagccatggaaactagataagaaagaa
 atacgcagagaccaaagttcaactgaaacgaattaaacggtttattgattaacaagcaa
 ttacagattacgagtcaggtatctggtgccaatggggcgaggctctgaatacacgccat
 tagtgccacagtaaagtccacattaacagacttgttgtagttggaagtgtactgaatt
 tcgggattccagcgtttgctgtttctctgcagctcccactcgatctccacgctgac
 ctgtcccgtggagtactgtgtgatgaaggaagcaaaactttgccgactgaagggtggtcg
 aaggattcgcaggtaccggggtggtcttgatgagaatctgtggaggagggtgtttaagt
 ccgaatccaccatgaggggagaggggtgaaaatgtccgtccgtgtgtggaatctttgc
 ccagatgggcccctgaaggtacacatctctgtcctgccagaccatgcctggaagaacgc
 ctgtgtgttgacatctgcggtagctgcttgctggttgccctctctggaggttggtagat
 acagaaccatactgctccgtagccacgggattggttgctctgatttctcttctgtctgt
 aatcatgacctttcaatgtccacatttgtttctctgagccttgcttcccaaagatga
 gaacccgctctgaggaaaaaactttcttcatcgtccttggtggcttgccatggccggg
 cccggattcaccagagagtctctgccattgaggtggtacttggtagctccagtcacga
 gtattcactgttggtgttatccgcagatgtctttgatactcgtgctggcggtaacagg
 gtccaggaagccagttcctagactggtcccgaatgtcactcgtcctccggcctgagaaaac

tgaagccttgactgcgtggtggttccacttggagtgtttgttctgctcaagtaatacag
gtactggctgatgagaggattcatgagacggccagactctggctgtgagcgtagctgc
tgtgaaaggaacgtcctcaaaagtgtagctgaaggtaaagttgttccggtagcagc
atctgagaaggaaagtactccaggcagtaaaatgaagagcgtcctactgctgactccc
gttgttcagggtgaggtatccatactgtggcaccatgaagacgtctgctgggaacggcg
ggaggcatccttgatgcgccgagccgaggacgtacgggagctggtactccgagtcagta
aacacctgaacctgctggttaaggttattggcaatcgtcgtcgtaccgtcattctgct
gacctcttgaacttgaatgttaaagagcttgaagttgagtctcttgggtcggaaatccc
agttgttgttgatgagctcttgcagtcacgtggtgaaaagtgccagtggaatctgttg
aagtcaaaatacccccaaggggtgctgtagccaaagtagtgattgtcgttcgaggtcc
tgattggctggaaatttgtttagaggtggttgttaggtgggcagggcccagggtc
gggtgctggtggtgatgactctgtcggcagccatgtggaatcgcaatgccaatccc
gaggaattaccactccgtcggcgcctcgttattgtctgcgagtggtgcgccactgcc
tgtagcaggggtattagttccgagaccagaggggctgctggtggctgtccgagaggct
gggggtcaggtactgagctctgcgtctccagctcgacaaaattcaatcttttcttgca
ggctgctggcccgccttccgggtcccagaggagctggtccacaggagagtgctc
taccggcctctttttccggagcggcttaaacagggtcctcaaccaggcccagagggt
caagaacctctttttgcctggaagactgctcgtccgaggttgccccaaaagacgta
tcttcttaagggcgtcctgaaactccgcgtcggcgtggttgtacttgaggtacgggt
gtctccgtgtcgagctgccggctgtaggcttctgtcgtcgtcgagggccgoggcgtctg
cctcgttgaccgctctcccttgcgagtcggtgaagggctccgaggtacttgtaacca
ggaagcacaagacctgctgtcgtccttatgccgctctgcgggcttgggtggtggtgg
gccaggttgagctccaccactgtcttattccttcagagagagtgctcctcgagccaat
ctggaagataaccatcggcagccatgcatgctagagtcgacggatcggagtagtctcct
ccgagcggagtagtctcctcgagcggagtagtctcctccgagcggagtagtctcctccg
agcggagtagtctcctccgagcggatccccgggaattagcttggcgtaatcatggtcat
agctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacagaccgga
agcataaagtgtaaagcctgggggtgcctaataagtgagtgagtaactcacattaattgcgt
gcgctcactgccgcttccagctcgggaacctgtcgtgccagctgcattaatgaatcg
gccaacgcgcggggagagggcgggttgcgtattgggcgtcttccgcttccctcgtcact
gactcgtcgcgtcggctcgttcggctgcggcgagcggatcagctcactcaaggcgg
aatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggcc
agcaaaaggccaggaaccgtaaaaaggccgctgctggcgtttttccataggctccgc
ccccctgacgagcatcacaataatcgacgtcaagtcagaggtggcgaaaccggacagg
actataaagataaccaggcgttccccctggaagctcccctcgtgcgctctcctgttccga
ccctgcccgttaccggatacctgtccgcctttctccctcgggaagcgtggcgctttct
catagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctg
tgtgcacgaacccccgttcagcccagcctgtgcgcttatccggtaactatcgtcttg
agtccaaccggtaagacacgacttatcgcactggcagcagccactggtaacaggatt
agcagagcaggtatgtaggcgggtgtacagagttcttgaagtgggtggcctaactacgg
ctacactagaagaacagtatttggtatctgcgctctgctgaagccagttacctcggaa
aaagagttggtagctcttgatccggcaaaacaaaccaccgctggtagcgggtgggttttt
gtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatctt
ttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttgggtcatga
gattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtttaaatca
atctaaagtatatatgagtaaaacttgggtctgacagttaccaatgcttaatcagtgaggc
acctatctcagcgtctgtctatctcgttcctcatcatagttgcctgactccccgctcgtg
agataactacgatacgggagggcctaccatctggccccagtgctgcaatgataccgcga

gaccacgctcaccggctccagatthtatcagcaataaaccagccagccggaagggccga
gcgcagaagtggctctgcaactttatccgcctccatccagtctattaattggtgccggg
aagctagagtaagtagttcgccagttaatagtttgcgcaacgttggtgccattgctaca
ggcatcgtgggtgacgctcgtcgtttggtatggcttcattcagctccggttcccaacg
atcaaggcgagttacatgatccccatggttgcaaaaaagcggttagctccttcggtc
ctccgatcgttgtcagaagtaagttggccgcagtggtatcactcatggttatggcagca
ctgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagta
ctcaaccaagtcattctgagaatagtgatgcgccgaccgagttgctcttgcccggcgt
caatacgggataataaccgcgccacatagcagaactttaaaagtgctcatcattgaaaa
cgttcttcggggcgaaaactctcaaggatcttaccgctggtgagatccagttcgatgta
accactcgtgcaccaactgatcttcagcatcttttactttcaccagcgtttctgggt
gagcaaaaacaggaaggcaaaaatgccgcaaaaaagggaataagggcgacacggaatgt
tgaatactcatactcttctttttcaatattattgaagcatttatcagggttattgtct
catgagcggatacatatthgatgtatthtagaaaaataaacaataggggttccgcgca
catttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacc
tataaaaataggcgtatcacgaggccctttcgtctcgcgcggttcgggtgatgacggtga
aaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccg
ggagcagacaagcccgtcagggcgcgtcagcgggtggtggcgggtgctcggggctggctt
aactatgcggcatcagagcagattgtactgagagtgcaccatatgcgggtgtaaatacc
gcacagatgcgtaaggagaaaataccgcatcagggcgcattcgcattcaggctgcgca
actgttgggaagggcgatcgggtcgggcctcttcgctattacgccagctggcgaaaggg
ggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacggtg
taaacgacggccagtgaaatcgagctcgggtaccttggtggttaactggtttattgcagc
ttataatggttacaataaagcaatagcatcacaatthcacaataaagcattthttt
cactgcattctagttgtggtttgtccaaactcat

pG5-TATA-VP1



Sequence color-coding key

Feature	Color
VP1	text
TATA box	text
GAL	text
Ori	text

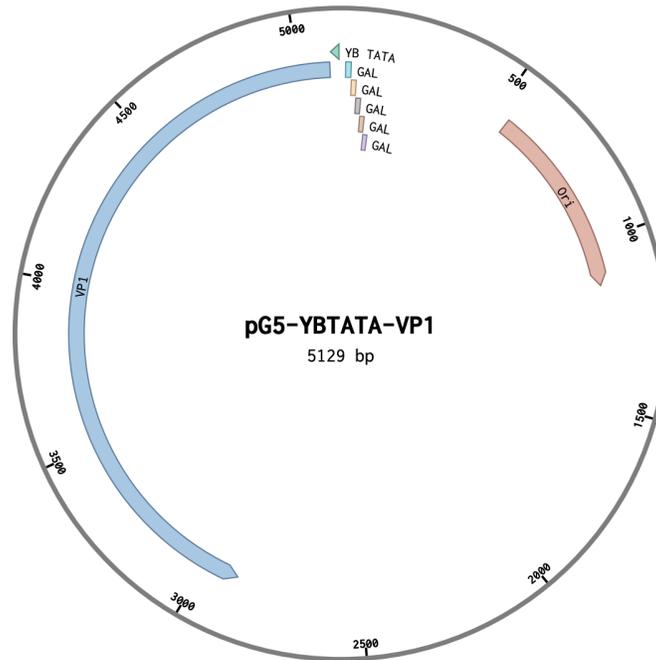
Sequence

caatgtatcttatcatgtctgctcgaagcggccggccgcccgactctagagctgtagt
 taatgattaaccgcgatgctacttatctacgtagccatggaaactagataagaaagaa
 atacgcagagaccaaagttcaactgaaacgaattaaacggtttattgattaacaagcaa
 ttacagattacgagtcaggtatctggtgccaatggggcgaggctctgaatacacgccat
 tagtgtccacagtaaagtccacattaacagacttggtgtagttggaagtgtactgaatt
 tcgggattccagcgtttgctggtttccttctgcagctcccactcgatctccacgctgac
 ctgtcccgtggagtactgtgtgatgaaggaagcaaaactttgccgactgaagggtgctg
 aaggattcgcaggtaccggggtgttcttgatgagaatctgtggaggagggtgtttaagt
 ccgaatccacccatgaggggagaggggtgaaaatgtccgtccgtgtgtggaatctttgc
 ccagatgggcccctgaaggtacacatctctgtcctgccagaccatgcctggaagaacgc
 cttgtgtgttgacatctgcggtagctgcttgtctggttgctctctggaggttggtagat
 acagaaccatactgctccgtagccacgggattggttgtcctgatttcctcttcgtctgt
 aatcatgacctttcaatgtccacatttgtttctctgagccttgcttcccaaagatga
 gaaccccgctctgaggaaaaaacttttcttcatcgtccttgtggcttgccatggccggg
 cccgattcaccagagagtctctgccattgaggtggtacttggttagctccagtcacga
 gtattcactgttggttattccgcagatgtctttgatactcgtgctggcggtaaacagg
 gtccaggaagccagttcctagactggtcccgaatgtcactcgtcctccggcctgagaaaac

tgaagccttgactgcgtggtggttccacttggagtgtttgttctgctcaagtaatacag
gtactggctgatgagaggattcatgagacggccagactctggctgtgagcgtagctgc
tgtggaaaggaacgtcctcaaaagtgtagctgaaggtaaagttgttccggtagcagc
atctgagaaggaaagtactccaggcagtaaaatgaagagcgtcctactgcctgactccc
gttgttcaggggtgaggtatccatactgtggcaccatgaagacgtctgctgggaacggcg
ggaggcatccttgatgcgccgagccgaggacgtacgggagctggtactccgagtcagta
aacacctgaacctgctggtaaggttattggcaatcgtcgtcgtaccgtcattctgcgt
gacctcttggacttgaatgttaaagagcttgaagttgagtctcttgggtcggaatcccc
agttgttgttgatgagtcttggccagtcacgtggtgaaaagtgccagtggaatctgttg
aagtcaaaatacccccaaggggtgctgtagccaaagtagtgattgtcgttcgaggtcc
tgattggctggaaatttgtttagaggtggttgttaggtgggcagggcccagggtc
gggtgctggtggtgatgactctgtcggcagccatgtggaatcgcaatgccaatcccc
gaggaattaccactccgtcggcgcctcgttattgtctgcgagtggtgcgccactgcc
tgtagcaggggtattagttccgagaccagagggggctgctggtggctgtccgagaggct
gggggtcaggtactgagtctgcgtctccagctcgacaaaattcaatcttttcttgca
ggctgctggcccgccttccgggtccccgaggaggagctggctccacaggagagtgctc
taccggcctcttttttccggagcggcttaaacagggtcctcaaccaggcccagagggt
caagaacctctttttgcctggaagactgctcgtccgaggttgccccaaaagacgta
tcttctttaaggcgtcctgaaactccgcgtcggcgtggttgtacttgaggtaggggt
gtctccgctgtcgagctgccggctgtaggcttctgtcgtcgtcgagggccgoggcgtctg
cctcgttgaccggtctccttctcgtcgtcgttgaaggggtccgaggtacttgtaacca
ggaagcacaagacctgctgtcgtccttatgccgctctgcgggcttgggtggtggtgg
gccaggttgagctccaccactgtcttattccttcagagagagtgctcctcgagccaat
ctggaagataaccatcggcagccatatttcttttatagcatgctagagtcgacggatcg
gagtagtctcctccgagcggagtagtctcctcgagcggagtagtctcctccgagcggag
tagtctcctccgagcggagtagtctcctccgagcggatccccgggaattagcttggcgt
aatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaac
atagcagccggaagcataaagtgtaaagcctgggggtgcctaatgagtgagctaactcac
attaattgcgttgcgtcactgcccgttccagtcgggaaacctgtcgtgccagctgc
attaatgaatcggccaacgcgcggggagagggcgttgcgtattgggcgtcttccgct
tctcgtcactgactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc
ctcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaagaacatgt
gagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgcgtggcgttttcc
cataggctccgccccctgacgagcatcaaaaaatcgacgctcaagtcagaggtggcg
aaaccgcagaggactataaagataccaggcgttccccctggaagctcctcgtgcgt
ctcctgttccgacctgcccgttaccggatacctgtccgccttctccttcgggaagc
gtggcgttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctc
caagctgggctgtgtgcacgaaccccccttcagcccagcgtgcgccttatccggta
actatcgtcttgagtccaaccggtaagacacgacttatcgcactggcagcagccact
ggtaacaggattagcagagcaggtatgtaggcgggtgctacagagttcttgaagtggg
gcctaactacggctacactagaagaacagatattgggtatctgcgtctgctgaagccag
ttaccttcggaaaaagagttggtagctcttgatccggcaacaaccaccgctggtagc
ggtggttttttgttgcagcagcagattacgcgcagaaaaaaggatctcaagaaga
tctttgatctttctacgggtctgacgctcagtggaacgaaaactcacgttaagga
ttttgggtcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatga
agtttaaatcaatctaaagtatatatgagtaaacttgggtctgacagttaccaatgctt
aatcagtgaggcacctatctcagcgtatctgtctatttctgttcacatagttgcctgac
tccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgca

atgataccgcgagacccacgctcacccggctccagatttatcagcaataaaccagccagc
cggaaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagctatta
attggtgcccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgtt
gccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctc
cggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggta
gctccttcggctcctccgatcgttgtcagaagtaagttggccgcagtggtatcactcatg
gttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgt
gactggtgagtactcaaccaagtattctgagaatagtgtatgcgggcagccgagttgct
cttggccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgtc
atcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctggtgagatc
cagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcacca
gcgtttctgggtgagcaaaaacaggaaggcaaaaatgccgcaaaaaaggggaataagggcg
acacggaaatggtgaatactcatactcttcctttttcaatattattgaagcatttatca
gggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaatag
gggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatc
atgacattaacctataaaaataggcgtatcacgaggccctttcgtctcgcgcggttcgg
tgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgt
aagcggatgccgggagcagacaagcccgtcagggcgcgctcagcgggtgttggcgggtgt
cggggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccatagcg
gtgtgaaataccgcacagatgcgtaaggagaaaataccgcacagggcgcattcgcacat
tcaggctgcgcaactggtgggaagggcgatcgggtgcgggcctcttcgctattacgccag
ctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggttttccca
gtcacgacgttgtaaacgacggccagtgaattcgagctcggtaaccttgttgtaactt
gtttattgcagcttataatgggttacaataaagcaatagcatcacaatttcacaaata
aagcatttttttactgcattctagttgtggtttgtccaaactcat

pG5-YBTATA-VP1



Sequence color-coding key

Feature	Color
GAL	text
Ori	text
VP1	text
YBTATA	text

Sequence

ctagagtcgacggatcgggagtactgtcctccgagcgggagtactgtcctcgagcggagta
ctgtcctccgagcgggagtactgtcctccgagcgggagtactgtcctccgagcgggatcccc
gggaattagcttggcgtaatcatgggtcatagctgtttcctgtgtgaaattgttatccgc
tcacaattccacacaacatacagagccggaagcataaagtgtaaagcctgggggtgcctaa
tgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaa
cctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgta
ttgggcgctcttccgcttctcctcgtcactgactcgtcgcctcggtcgctcgggctgcgg
cgagcgggtatcagctcactcaaaggcggtaatacgggtatccacagaatcaggggataa
cgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccg
cgttgctggcgtttttccataggctccgccccctgacgagcatcacaanaatcgacgc
tcaagtcagaggtggcgaaaccgcagaggactataaagataaccaggcgtttccccctgg
aagctccctcgtgcgctctcctgttccgaccctgcgcgttaccggatacctgtccgct
ttctcccttcgggaagcgtggcgttttctcatagctcacgctgtaggtatctcagttcg
gtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccg
ctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgc
cactggcagcagccactggtaacaggattagcagagcagaggtatgtaggcgggtgctaca

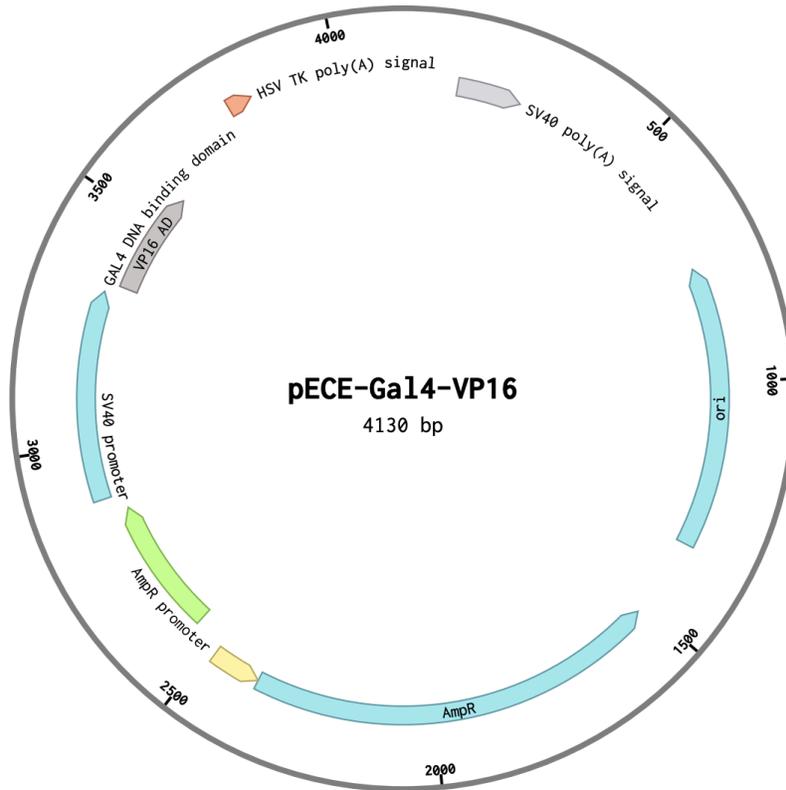
gagttcttgaagtgggtggcctaactacggctacactagaagaacagtatgttggatctg
cgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaac
aaaccaccgctggtagcgggtgttttttggttgcaagcagcagattacgcgcagaaaa
aaaggatctcaa gaagatcctttgatcttttctacggggtctgacgctcagtggaaacga
aaactcacgttaagggatgttggatgagattatcaaaaaggatcttcacctagatcc
ttttaaatataaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggct
gacagttaccaatgcttaatcagtgagggcacctatctcagcgatctgtctatttcgctc
atccatagttgctgactccccgctcgtgtagataactacgatacgggagggttaccat
ctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagatttatca
gcaataaaccagccagccggaagggccgagcgcagaagtggctctgcaactttatccgc
ctccatccagctctattaattggtgccgggaagctagagtaagtagttcggcagttaata
gtttgcgcaacggtgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttgg
atggcttcattcagctccggttccaacgatcaaggcgagttacatgatccccatggt
gtgcaaaaagcgggttagctccttcggctcctccgatcgttgtcagaagtaagttggccg
cagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatcc
gtaagatgcttttctgtgactggtgagtagtcaaccaagtcattctgagaatagtgat
gcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagca
gaactttaaaagtgctcatcattggaaaacggttcttcggggcgaaaactctcaaggatc
ttaccgctggtgagatccagttcgatgtaaccactcgtgcaccaactgatcttcagc
atcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaatgccgcaa
aaaagggaaataagggcgacacggaaatggtgaatactcatactcttctttttcaatat
tattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtattta
gaaaaataaacaataggggttccgcgcacatttccccgaaaagtgccacctgacgtct
aagaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctt
cgtctcgcgcggttcgggtgatgacggtgaaaacctctgacacatgcagctcccggagac
ggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgctcag
cgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactg
agagtgcaccatatgcgggtgtgaaataccgcacagatgcgtaaggagaaaataccgcat
cagggccattcggcattcaggctgcgcaactgttgggaagggcgatcgggtcggggcct
cttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggta
acgccaggggttttcccagtcacgacggtgtaaaacgacggccagtgaattcgtctgtagt
taatgattaaccgccatgctacttatctacgtagccatggaaactagataagaaagaa
atacgcagagaccaaagttcaactgaaacgaattaaacggtttattgattaacaagcaa
ttacagattacgagtcaggtatctgggtgccaatggggcgaggctctgaatacacgccat
tagtgtccacagtaaagttccacattaacagacttgttgtagttggaagtgtactgaatt
tcgggattccagcgtttgctgttttcttctgcagctcccactcgatctccacgctgac
ctgtcccgtggagtagtctgtgtgatgaaggaagcaaacctttgccgcactgaaggtggctc
aaggattcgcaggtaccgggggtgttcttgatgagaatctgtggaggagggtgtttaagt
ccgaatccaccatgagggggagaggggtgaaaatgtccgtccgtgtgtggaatctttgc
ccagatggggcccctgaaggtacacatctctgtcctgccagaccatgcctggaagaacgc
cttgtgtgttgacatctgcggtagctgcttgtctgttgctctctggagggttggtagat
acagaaccatactgctccgtagccacgggattgggtgtcctgatttctctctcgtctgt
aatcatgaccttttcaatgtccacatttgttttctctgagccttgccttcccaaagatga
gaaccccgctctgaggaaaaaacttttcttcatcgtccttgtggcttgccatggccggg
cccggattcaccagagagctctctgccattgaggtggtagcttggtagctccagttccacga
gtattcactggtgttattccgcagatgtctttgatactcgtctgctggcggtaacagg
gtccaggaagccagttcctagactgggtcccgaatgtcactcgtcggcctgagaaaac
tgaagccttgactgcgtgggtgggtccacttggagtggttctgctcaagtaatacag

gtactggctcgatgagaggattcatgagacgggtccagactctggctgtgagcgtagctgc
tgtggaaaggaacgtcctcaaaagtgtagctgaaggtaaagttgtttccgggtacgcagc
atctgagaaggaaagtactccaggcagtaaaatgaagagcgtcctactgcctgactccc
gttgttcaggggtgaggtatccatactgtggcaccatgaagacgtctgctgggaacggcg
ggagggatccttgatgcgccgagccgaggacgtacgggagctggtactccgagtcagta
aacacctgaaccgtgctggtaaggttattggcaatcgctcgtcgtaccgtcattctgcgt
gacctctttgacttgaatgttaaagagcttgaagttgagtctcttgggtcgggaatcccc
agttgttgttgatgagtccttgccagtcacgtggtgaaaagtgccagtggaatctggtg
aagtcaaaatacccccaaggggtgctgtagccaaagtagtgattgtcgttcgaggctcc
tgattggctggaaatttgtttgtagaggtggttgttgtaggtgggcagggcccagggtc
gggtgctggtggtgatgactctgctgccgagccatgtggaatcgcaatgccaatctcc
gaggaattaccactccgctcggcgccctcgttattgtctgcgagtggtgcgccactgcc
tgtagcgaggggtattagttccgagaccagagggggctgctggtggctgtccgagaggct
gggggtcaggtactgagtcctgcgtctccagtcctgaccaaattcaatctttttcttgc
ggctgctggcccgcctttccgggtcccgaggaggagctggctccacaggagagtgctc
taccggcctcttttttcccgagcgggtctaacagggtcctcaaccaggcccagagggt
caagaaccctctttttcgcctggaagactgctcgtccgaggttgccccaaaagacgta
tcttctttaagggcgtcctgaaactccgctcggcgtggttgtacttgaggtacgggtt
gtctccgctgtcagctgccggctcgtaggctttgtcgtgctcagggcccgggcgtctg
cctcgttgaccggctctcccttgctcagctcgttgaagggtccgaggtacttgtacca
ggaagcacaagaccctgctgctcgtccttatgccgctctgcgggctttggtggtggtgg
gccaggtttgagcttccaccactgtcttattccttcagagagagtgctcctcagccaat
ctggaagataaccatcggcagccattggccccattatataccctctagagcatg

The sequence for the minCMV minimal promoter is:

gatctgacggttcactaaacgagctctgcttatatagacctcccaccgtacacgcctac

pECE-Gal4-VP16



Sequence color-coding key

Feature	Color
SV40 poly(A) signal	text
Ori	text
AmpR	text
AmpR promoter	text
SV40 promoter	text
GAL4 DNA binding domain	text
VP16 AD	text
HSV TK poly(A) signal	text

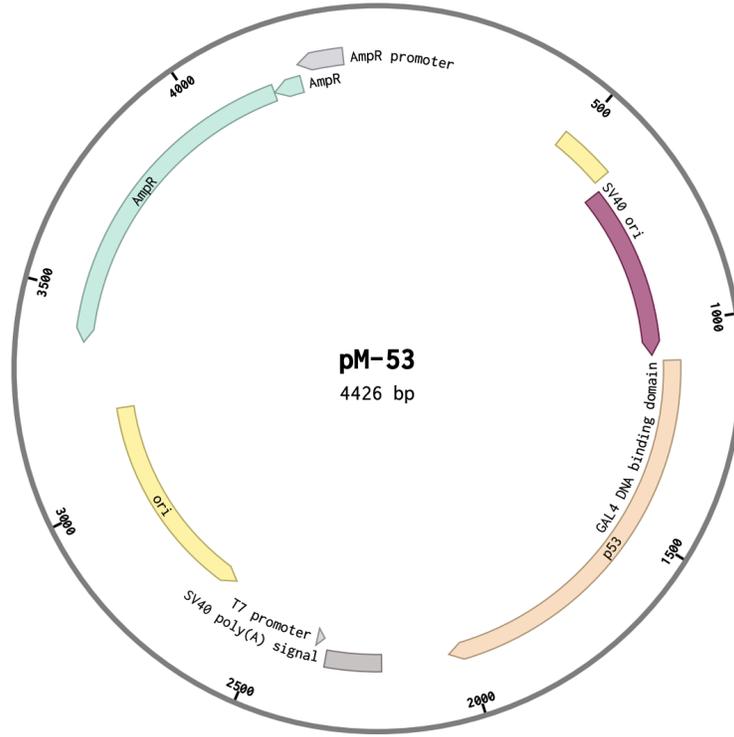
Sequence

taagtaatgatcataatcagccatatcacatctgtagaggttttacttgctttaaaaa
cctcccacacctccccctgaacctgaaacataaaatgaatgcaattggtggtggt**aact**
tgtttattgcagcttataatggttacaataaagcaatagcatcacaatttcacaat
aaagcatttttttactgcattctagttgtggtttgtccaaactcatcaatgtatctta
tcatgtctggatcctctacgccggacgcacatcggtggccggcatcacccggcggcacagggtg
cggttgctggcgcctatatcgccgacatcacccgatggggaagatcgggctcggcacttc
gggctcatgagcgttgtttcggcgtgggtatggtggcaggccccgtggccgggggact
gttgggcgccatctccttgcacaccattccttgcggcggcggtgctcaacggcctca
acctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtgcggtgtga

aataccgcacagatgcgtaaggagaaaataaccgcatcaggcgctcttccgcttcctcgc
tcaactgactcgctgcgctcggctcggtcggctgcgggagcggtatcagctcaactcaag
gcggttaatacgggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaa
aggccagcaaaaaggccaggaaccgtaaaaaggccgcttgctggcggttttccataggg
tccgccccctgacgagcatcacaaaaatcgacgctcaagtcagagggtggcgaaacccg
acaggactataaagataaccaggcgtttccccctggaagctccctcgtgcgctctcctgt
tccgacctgcccgttaccggatacctgtccgccttttcccttcgggaagcgtggcgc
tttctcaatgctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctg
ggctgtgtgcacgaacccccggttcagcccagcctgctgccttatccggtaactatcg
tcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaaca
ggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggcctaac
tacggctacactagaaggacagtatattggtatctgcgctctgctgaagccagttacct
cgaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtt
tttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttg
atcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggttttgg
catgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttta
aatcaatctaaaagtatatatgagtaaaacttggtctgacagttaccaatgcttaacag
tgaggcacctatctcagcgatctgtctatcttctcgttcacatagttgctgactccccg
tcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgata
ccgcgagaccacgctcaccggctccagatttatcagcaataaaccagccagccggaag
ggccgagcgcagaagtggcctgcaactttatccgcctccatccagctattaattggt
gccgggaagctagagtaagttagtccgacgtaataagtttgcaacggttggtgccatt
gctgcaggcatcgtgggtgcacgctcgtcgtttggatggcttcattcagctccgggtc
ccaacgatcaaggcgagttacatgatccccatggttgcaaaaaagcgggttagctcct
tcggtcctccgatcgttgctcagaagtaagttggccgcagtggtatcactcatggttatg
gcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactgg
tgagtactcaaccaagtcattctgagaatagtgatgcgggcagccaggtgctcttgcc
cggcgtcaacacgggataataccgcgccacatagcagaactttaaaagtgtcatcatt
ggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctggtgagatccagttc
gatgtaaccactcgtgcaccaactgatcttcagcatcttttactttcaccagcgttt
ctgggtgagcaaaaacaggaaggcaaaaatgccgcaaaaaagggaataagggcgacacgg
aaatggtgaatactcactcttcttcttttcaatattattgaagcatttatcagggtta
ttgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaatagggggtc
cgcgcacatttccccgaaaagtgccacctgcttctgaggcggaaagaaccagctgtgga
atgtgtgtcagtttagggtgtggaagtcgccaggtcccccagcaggcagaagtatgcaa
agcatgcatctcaattagtcagcaaccagggtgtggaaagtccccaggtccccagcagg
cagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcgcccccctaactc
cgcccatcccgccctaactccgccagttccgccattctccgcccatgggtgacta
atTTTTTTTatTTTatGcagaggccgaggccgctcggcctctgagctattccagaagta
gtgaagaggctTTTTTggaggagatctaagcttgaagcaagcctcctgaaagatgaagc
tactgtcttctatcgaacaagcatgcatatttgccgacttaaaaagctcaagtgtcc
aaagaaaaaccgaagtgcgccaagtgtctgaagaacaactgggagtgctgctactctcc
caaaacaaaaggctcctccgctgactagggcacatctgacagaagtggaatcaaggctag
aaagactggaacagctatttctactgatttttctcogagaagaccttgacatgattttg
aaaatggattctttacaggatataaaagcattgttaacaggattatttgtacaagataa
tgtgaataaagatgcccgtcacagatagattggcttcagtgagactgatatgcctctaa
cattgagacagcatagaataagtgcgacatcatcatcggaagagagtagtaacaaaggt
caaagacagttgactgtatgcgggaattccccgggatctggcccccccgaccgatgt

cagcctgggggacgagctccacttagacggcgaggacgtggcgatggcgcatgccgacg
cgctagacgatttcgatctggacatggtgggggacggggattccccgggtccgggatt
acccccacgactccgccccctacggcgctctggatatggccgacttcgagtttgagca
gatgtttaccgatgcccttgaattgacgagtacggtagggtagggggcgcgaccggacc
cgcatccccgtctgggttttccccctcccggtcccggttcgtatccacaataaacacga
gcacatacattacaaaaacctgcggttgctgctgattatgggtggatccgggggaga
tgggggaggctaactgaaacacggaaggagacaataccggaaggaaccgcgctatgac
ggcaataaaaagacagaataaaacgcacgggtggtgggtcgtttgttc ataaacgcggg
gttcggtcccagggtggcactctgtcgataccccaccgagaccccattggggccaata
cgcccggtttcttcttttccccaccccccaagttcgggtgaaggcccagggc
tcgcagccaacgtcggggcggcaggccctgccatagccactggccccgtgggttaggga
cgggggtccccatggggaatggtttatggttcgtgggggttattatgggcggttgcg
tggggtcaggtccacgaccaagcttgtcgacggtacccccgggaattcgagctctaga

pM-53



Sequence color-coding key

Feature	Color
SV40 ori	text
GAL4 DNA binding domain	text
p53	text
SV40 poly(A) signal	text
T7 promoter	text
Ori	text
AmpR	text
AmpR promoter	text

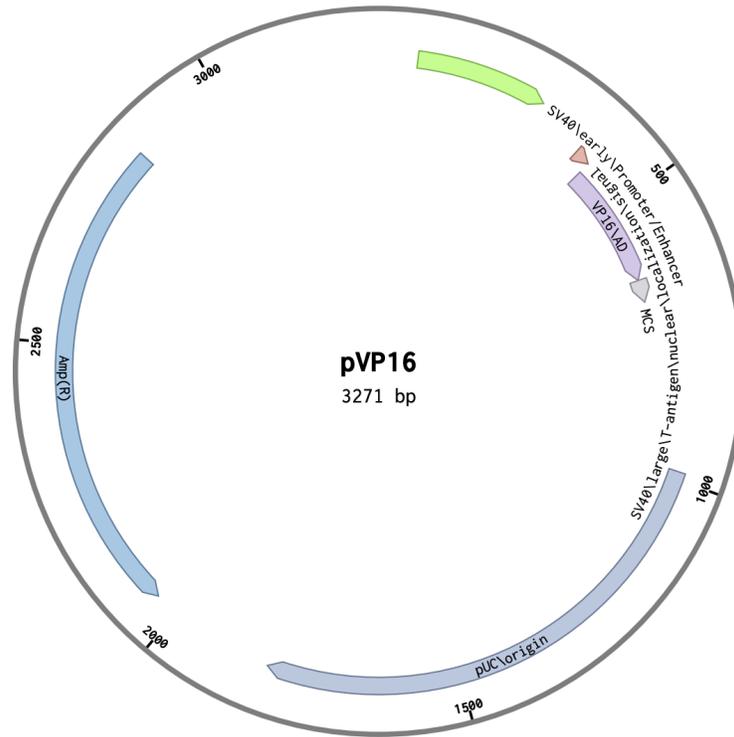
Sequence

tcacgagggccctttcgtctcgcgcggtttcggatgatgacggtgaaaacctctgacacatg
cagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagccc
tcagggcgcgtcagcgggtggtggcgggtgtcgggctggcttaactatgcggcacag
agcagattgtactgagagtgcaccatatggacatattgtcgttagaacgcggctacaat
taatacataaccttatgtatcatcacacacacgatttaggtgacactatagaactcgatg
tggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtat
gcaaagcatgcatctcaattagtcagcaaggaaagtccccaggctccccagcaggcaga
agtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgc
catcccgcccctaactccgcccagttccgcccatttccgcccattggctgactaattt
tttttatttatgcagagggccgaggccgcctcggcctctgagctattccagaagtagtga

agaggcttttttggaggagatctaagctagcctcctgaaagatgaagctactgtcttct
atcgaacaagcatgcgatatttgccgacttaaaaagctcaagtgctccaaagaaaaacc
gaagtgcgccaagtgtctgaagaacaactgggagtgtcgctactctcccaaaacaaaa
ggtctccgctgactagggcacatctgacagaagtggaaatcaaggctagaaagactggaa
cagctatcttctactgatttttctcgcgagaagaccttgacatgattttgaaaatggattc
tttacaggatataaaagcattgttaacaggattatgttacaagataatgtgaataaag
atgccgtcacagatagattggcttcagtgaggactgatatgcctctaacattgagacag
catagaataagtgcgacatcatcatcgggaagagagtagtaacaaaggtcaaagacagtt
gactgtatcgccggaattccctgtcaccgagaccctgggcccagtggcccctgccccag
ccactccatggcccctgtcatcttttgtcccttctcaaaaaacttaccagggcaactat
ggcttccacctgggcttctgcagctctgggacagccaagtctgttatgtgcacgtactc
tctccccctcaataagctattctgccagctggcgaagacgtgccctgtgcagttgtggg
tcagcgccacacctccagctgggagccgtgtccgcgccatggccatctacaagaagtca
cagcacatgacggaggtcgtgagacgctgccccaccatgagcgtgctccgatggtg
tgacctggctcctccccagcatcttatccgggtggaaggaaatttgatcccgagatc
tggaaagacaggcagacttttcgccacagcgtgggtgtaccttatgagccaccgaggcc
ggctctgagtataccaccatccactacaagtacatgtgtaatagctcctgcatggggg
catgaaccgcccagctatccttaccatcatcacactggaagactccagtggggaacctt
tgggacgggacagctttgaggttcgtgtttgtgcctgccctgggagagaccgccgtaca
gaagaagaaaatctccgcaaaaaggagtcctttgccctgaactgccccagggagcgc
aaagagagcgtgccacctgcacaagcgcctctccccgcaaaaagaaaaaccacttg
atggagagtatttcacctcaagatccgcgggcgtaaacgcttcgagatgttccgggag
ctgaatgaggccttagagttaaaggatgccatgctacagaggagtctggagacagcag
ggctcactccagctacctgaagaccaagaaggccagctacttcccgccataaaaaaa
caatggtaagaaagtggggcctgactcagactgactgcctctgcatcccggatccgtc
gacgcgtctgcagaagcttctagataagtaatgatcataatcagccatatacatctgt
agaggttttacttgctttaaanaaacctcccacacctccccctgaacctgaaacataaaa
tgaatgcaattgttggtgtaacttgtttattgcagcttataatgggttacaataaagc
aatagcatcacaatctcacaataaagcatttttttctactgcattctagttgtggttt
gtccaaactcatcaatgtatcttatcatgtctggatctgccggtctcctatagtgagt
cgtattaatttcgataagccaggttaacctgcattaatgaatcggccaacgcgcgggga
gaggcggtttgctattgggcgctcttccgcttctcctcgtcactgactcgtgcgctcg
gtcgttccgctgcggcgagcggatcagctcactcaaaggcggtaatacggttatccac
agaatcaggggataacgcaggaaagaacatgtgagcaaaaaggccagcaaaaaggccagga
accgtaaaaaggccgcgttgcgtggcgttttccataggctccgccccctgacgagcat
cacaanaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatacca
ggcgtttccccctggaagctccctcgtgcgctctcctgttccgacctgcccgttaccg
gatacctgtccgcttttctcccttcgggaagcgtggcgcttttctcatagctcacgctgt
aggatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcaogaaccccc
cgttcagcccagcgtgcgccttatccggtaactatcgtcttgagccaaccggtaa
gacacgacttatcgccactggcagcagccactggtaacaggattagcagagcaggtat
gtaggcgggtgctacagagttcttgaagtgggtggcctaactacggctacactagaagaac
agtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagct
cttgatccggcaaaaaccaccgctggtagcgggtgggtttttttgtttgcaagcagcag
attacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacgggggtctga
cgctcagtggaacgaaaactcacgttaagggattttggatcatgagattatcaaaaagga
tcttcacctagatccttttaaatataaaaatgaagttttaaatcaatctaagtatatat
gagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcga

ctgtctatthtoggttcatccatagttgctgactccccgtcgtgtagataactacgatac
gggagggccttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccg
gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggctc
tgcaactttatccgcctccatccagtctattaattgttgcccgggaagctagagtaagta
gttcgccagttaatagtttgccgaacgttgttgccattgctacaggcatcgtggtgtca
cgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggctcctccgatcgttgta
gaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctctt
actgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcatt
ctgagaatagtgtatgcccgcaccgagttgctcttgcccggcgtcaatacgggataata
ccgcgccacatagcagaactttaaaagtgtcatcattggaaaacgttcttcggggcga
aaactctcaaggatcttaccgctgttgagatccagttcgatgtaaccactcgtgcacc
caactgatcttcagcatctttactttcaccagcgtttctgggtgagcaaaaacaggaa
ggcaaaatgccgcaaaaaaggaataagggcgacacggaaatgttgaatactcatactc
ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacat
atthgatgtatthtagaaaaataaacaataggggttccgcgcacatttccccgaaaag
tgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgt
a

pVP16



Sequence color-coding key

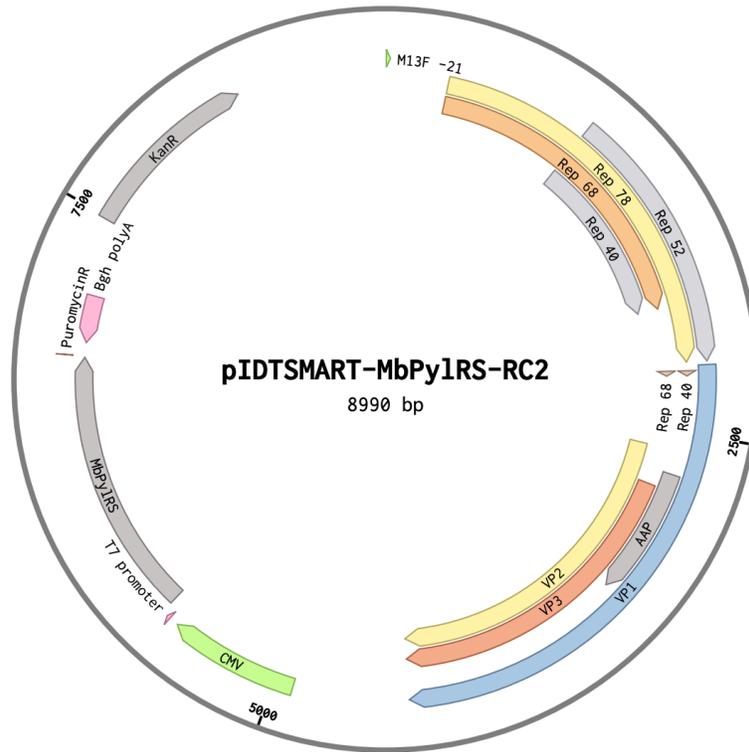
Feature	Color
SV40 early Promoter/Enhancer	text
SV40 large T-antigen nuclear localization signal	text
VP16-AD	text
Multiple cloning site	text
pUC Ori	text
AmpR	text

Sequence

tatgtatcatcacatacagatttaggtgacactatagaactcgatgtggaatgtgtgtc
 agttagggtgtgaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcat
 ctcaattagtcagcaaggaaagtccccaggctccccagcaggcagaagtatgcaaagca
 tgcattctcaattagtcagcaaccatagtcctcgcccctaactccgcccattccgccccta
 actccgcccagttccgcccattctccgcccattggtgactaattttttttatattatgc
 agaggccgaggccgctcggcctctgagctattccagaagtagtgaagaggcttttttg
 gaggatctaaagctagcgcgccaccatggggcctaaaagaagcgtaaagtcgcccc
 ccgaccgatgtcagcctgggggacgagctccacttagacggcgaggacgtggcgatgg
 cgcattgccgacgcgctagacgatttcgatctggacatggtgggggacggggattccccg
 ggtccgggatttaccocccacgactccgcccctacggcgctctggatattggccgactt
 cgagtttgagcagatggttaccgatgccttggaattgacgagtagcgggggaattcc
 cggggatccgctcagcgcgtctgcagaagcttctagataagtaatgatcataatcagcca

tatcacatctgtagagggttttacttgctttaaanaaacctcccacacctccccctgaacc
tgaaacataaaaatgaatgcaattggtggtgtaacttgtttattgcagcttataatggt
tacaataaagcaatagcatcacaatttcacaaataaagcattttttcactgcattc
tagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgccggtctc
cctatagtgagtcgtattaatttcgataagccagggttaacctgcattaatgaatcggcc
aacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttccctcgtcactgac
tcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaat
acggttatccacagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagc
aaaaggccaggaaccgtaaaaaggccgcgttgctggcggtttttccataggtccgcccc
cctgacgagcatcacaanaatcgacgctcaagtcagaggtggcgaaaccgcagaggact
ataaagataaccaggcgtttccccctggaagctccctcgtgcgctctcctggtccgacc
tgccgcttaccgataacctgtccgcctttctcccttcgggaagcgtggcgctttctcat
agctcacgctgtaggtatctcagttcgggtgtaggtcgctccagctgggctgtgt
gcacgaacccccggttcagcccagccgctgcgccttatccggtaactatcgtcttgagt
ccaaccggtaagacacgacttatcgcactggcagcagccactggtaacaggattagc
agagcgaggtatgtaggcggtgctacagagttcttgaagtgggtggcctaactacggcta
cactagaagaacagtatattgggtatctgcgctctgctgaagccagttaccttcggaaaa
gagttggtagctcttgatccggcaaacacaaccaccgctggtagcgggtggttttttgtt
tgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttttc
tacggggtctgacgctcagtggaacgaaaactcacgtaagggattttgggtcatgagat
tatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaataaatc
taaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacc
tatctcagcgatctgtctatcttgcgttcacccatagttgcctgactccccgctcgtgtaga
taactacgatacggggagggttaccatctggccccagtgctgcaatgataccgcgagac
ccagctcacoggctccagatttatcagcaataaaccagccagccggaagggccgagcg
cagaagtggctcgaactttatccgcctccatccagcttattaattggtgccgggaag
ctagagtaagtagttccagcagtaataagtttgcgcaacgttggtgcccattgctacaggc
atcgtgggtgacgctcgtcgtttgggtatggcttcattcagctccgggtcccaacgatc
aaggcgagttacatgatccccatggttggtgcaaaaaagcgggttagctccttcgggtcctc
cgatcgttgtcagaagtaagttggcgcagtggtatcactcatggttatggcagcactg
cataattctcttactgcatgccatccgtaagatgcttttctgtgactggtgagtaactc
aaccaagtcatcttgagaatagtgtatgcggcgaccgagttgctcttgcccggcgctcaa
tacgggataataccgcgccacatagcagaactttaaagtgtcatcattggaaaacgt
tcttcggggcgaaaactctcaaggatcttaaccgctggtgagatccagttcgatgtaacc
cactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgag
caaaaacaggaaggcaaatgcccgaanaaagggaataagggcgacacggaaatggtga
atactcatactcttcccccttttcaatattattgaagcatttatcagggttattgtctcat
gagcggatacatatttgaatgtatttagaaaaataaacaataaggggtccgcgcacat
ttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctat
aaaaataggcgtatcacgaggccctttcgtctcgcgcggtttcggtgatgacgggtgaaaa
cctctgacacatgcagctcccggagacggtcacagcttgctgtaagcggatgccggga
gcagacaagcccgtcagggcgctcagcgggtggtggcgggtgctcggggctggcttaac
tatgcggcatcagagcagattgtactgagagtgaccatattggacatattgtcgttaga
acgcggctacaattaatacataacct

pIDT-MbPylRS-RC2



The sequence for the wild type RC2 sequence can be found in the Chapter 2 Plasmids section of this appendix.

The sequence for the VP1del *Cap* gene is:

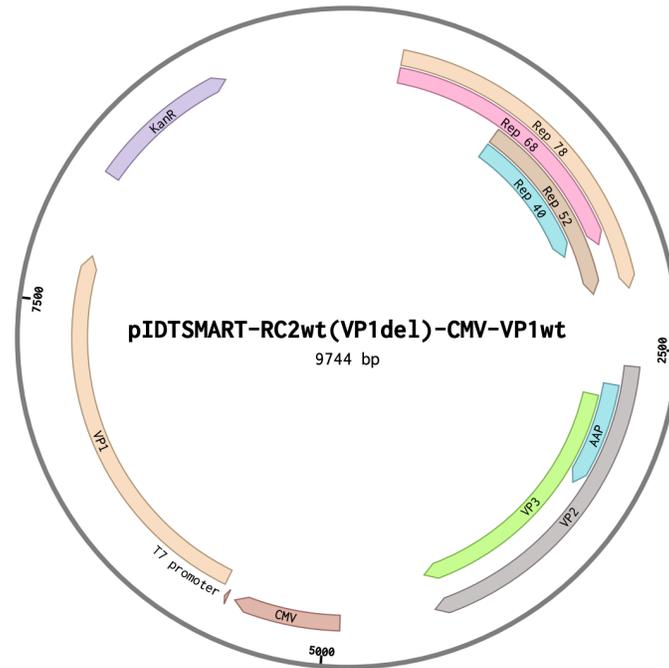
```

CTCgctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataag
acagtgggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataagg
acgacagcaggggtcttgtgcttctctgggtacaagtacctcggacccttcaacggactc
gacaagggagagccgggtcaacgaggcagacgccgcggccctcgagcacgacaaagccta
cgaccggcagctcgacagcggagacaaccttacctcaagtacaaccacgcccagcgcgg
agtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtc
ttccaggcgaaaaagaggggttcttgaacctctgggcctgggtgaggaacctgttaagac
ggctccgggaaaaaagaggccggtagagcactctctctgtggagccagactcctcctcgg
gaaccggaaaggcgggcccagcagcctgcaagaaaaagattgaattttggtcagactgga
gacgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctgg
tctgggaactaatac gatggctacaggcagtgggcgcaccaatggcagacaataacgagg
gcccgcagcggagtgggtaattcctcgggaaattggcattgcgattccacatggatgggc
gacagagtcaccaccagcaccggaacctgggcctgccacctacaacaaccacct
ctacaacaaatctccagccaatcaggagcctcgaacgacaatcactactttggctaca
gcacccttgggggtatcttgacttcaacagattccactgccacttttcaccacgtgac
tggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagct
ctttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagcagattgccaata
accttaccagcagcgggttcagggtgttactgactcggagtagcagcctcccgtagctcctc

```

ggctcggcgcacatcaaggatgcctcccgcggtcccagcagacgtcttcatgggtgccaca
gtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttact
gcctggagtactttccttctcagatgctgcgtaccggaacaactttaccttcagctac
acttttgaggacgttcctttccacagcagctacgctcacagccagagtctggaccgtct
catgaatcctctcatcgaccagtacctgtattacttgagcagaacaaactccaagtg
gaaccaccacgcagtcaggcttcagttttctcaggccggagcagtgacattcgggac
cagtcctaggaactggcttccctggaccctgttaccgccagcagcagtgatcaaagacatc
tgcggataacaacaacagtgaaactcgtggactggagctaccaagtaccacctcaatg
gcagagactctctgggtgaatccgggcccggccatggcaagccacaaggacgatgaagaa
aagttttttcctcagagcgggggttctcatctttgggaagcaaggctcagagaaaacaaa
tgtggacattgaaaaggatcatgattacagacgaagaggaaatcaggacaaccaatcccg
tggctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagca
gctaccgcagatgtcaacacacaaggcgttcttccaggcatggctctggcaggacagaga
tgtgtaccttcaggggcccattctgggcaaagattccacacacggacggacattttcacc
cctctcccctcatgggtggattcggacttaaaccacctcctccacagattctcatcaag
aacaccccggctacctgcgaatccttcgaccaccttcagtgccgcaaagtttgcttcct
catcacacagtactccacgggacaggtcagcgtggagatcgagtgaggagctgcagaagg
aaaacagcaaacgctggaatcccgaatcagtacacttccaactacaacaagtctgtt
aatgtggactttactgtggacactaatggcgtgtattcagagcctcgccccattggcac
cagatacctgactcgtaactctgtaa

pIDTSMART-RC2-VP1del-CMV-VP1wt



Sequence color-coding key

Feature	Color
Rep	text
Cap (VP1del)	text
CMV promoter	text
T7 promoter	text
VP1 wild type	text
KanR	text

Sequence

```

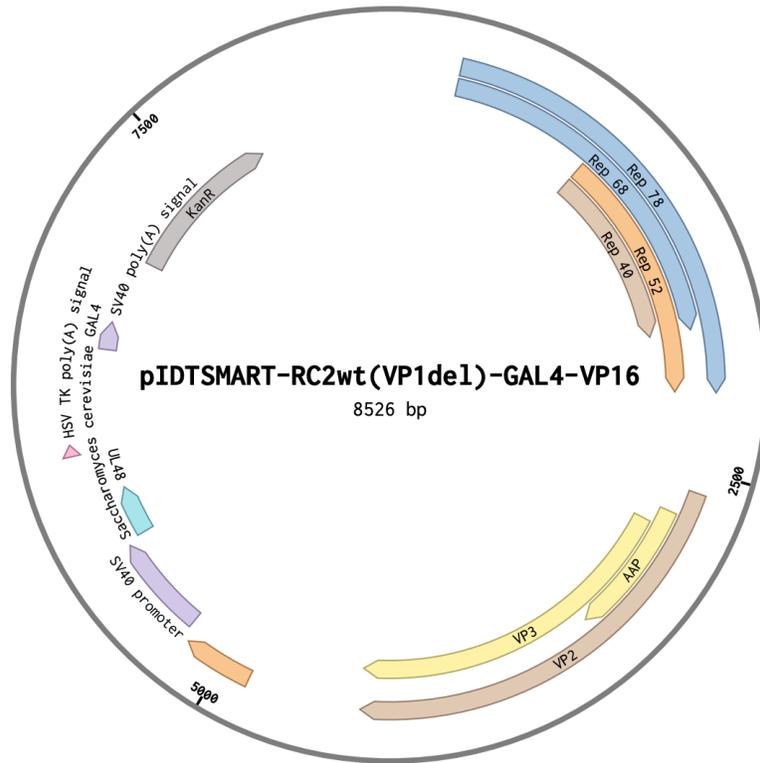
cccggtgtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
gaaggtgagccagtgagttgattgcagtcacggttacgctggagctctgaggctcgtcctg
aatgatatgacgaccgcccggagggttgcgtttgagacgggcgacagatccagtcgcgctg
ctctcgatcgatccgctagggcggccgctctagaactagtgatcccccggaagatcaga
agttcctattccgaagttcctattctctagaaagtataggaacttctgatctgcccagc
cgccatgcccggggttttacgagattgtgattaaggtcccagcgaacctgacgagcattc
tgcccggcatttctgacagctttgtgaactgggtggccgagaaggaatgggagttgccg
ccagattctgacatggatctgaatctgattgagcaggcaccctgaccgtggccgagaa
gctgcagcgcgacttctgacggaatggcgcgctgtgagtaaggccccggaggcccttt
tctttgtgcaatttgagaagggagagagctactccacatgcacgtgctcgtggaacc
accgggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaatttaccgcccggatcgagccgactttgccaaactggttcgcccgtcacaaga
ccagaaatggcgcggaggcgggaacaaggtggtggatgagtgctacatccccaattac
ttgctcccaaaaccagcctgagctccagtgggcgtggactaatatggaacagatttt
    
```

aagcgctgtttgaatctcacggagcgtaaacggttggtggcgcagcatctgacgcacg
tgtcgcagacgcaggagcagaacaaagagaatcagaatcccaattctgatgcccgggtg
atcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggacaagg
gattacctcggagaagcagtggtaccaggaggaccaggcctcatacatctccttcaatg
cggcctccaactcgcgggtcccaaatcaaggctgccttggacaatgcccggaaagattatg
agcctgactaaaaccgccccgactacctggtggggccagcagcccgtggaggacatttc
cagcaatcggatttataaaattttggaactaaacgggtacgatccccaatatgcccgtt
ccgtctttctgggatgggcccacgaaaaagttcggcaagaggaacaccatctggctgttt
ggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgccctt
ctacgggtgcgtaaactggaccaatgagaactttcccttcaacgactgtgtcgacaaga
tggtgatctggtgggaggagggaagatgaccgccaaggtcgtggagtcggccaaagcc
attctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagataga
cccgactcccgtgatcgtcacctccaacaccaacatgtgcgccgtgattgacgggaact
caacgaccttcgaacaccagcagccgttgcaagaccggatggtcaaatttgaactcacc
cgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagactttttccg
gtgggcaaaggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtggag
ccaagaaaagaccgccccagtgacgcagatataagtgagccaaaacgggtgcgcgag
tcagttgcgcagccatcgcagtcagacgcggaagcttcgatcaactacgcagacaggta
ccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttccctgcagacaat
gcgagagaatgaatcagaattcaaatatctgcttcaactcagggacagaaagactgttta
gagtgtttcccggtgcagaatctcaaccgtttctgtcgtcaaaaaggcgtatcagaa
actgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctgcgatc
tggtcaatgtggatttggtgactgcatctttgaacaataaagatttaaatcaggctc
cgctgcccgatggttatcttccagattggctcgaggacactctctctgaaggaataagac
agtgggtggaagctcaaacctggcccaccaccacaaagcccgcagagcggcataaggac
gacagcaggggtcttgtgcttccctgggtacaagtacctcggacccttcaacggactcga
caagggagagccggtcaacgaggcagacgcgcggccctcgagcacgacaaaagcctacg
accggcagctcgacagcggagacaaccctgacctcaagtacaaccacgcccagcggag
tttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagctctt
ccaggcgaaaaagaggggtcttgaacctctgggctggttgaggaacctgttaagacgg
ctccgggaaaaaagaggccggttagagcactctcctgtggagccagactcctcctcggga
accggaaggcgggcccagcagcctgcaagaaaaagattgaattttggtcagactggaga
cgcagactcagtacctgaccccagcctctcggacagccaccagcagccccctctggtc
tgggaaactaatcagatggctacaggcagtgggcaccatggcagacaataacgagggc
gccgacggagtgggtaattcctcgggaaattggcattgcgattccacatggatgggcga
cagagtcaccaccagcaccggaacctgggcccctgccacctacaacaaccacctt
acaacaaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagc
acccttgggggtattttgacttcaacagattccactgccacttttcaccacgtgactg
gcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagctct
ttaacattcaagtcaaagaggtcacgcagaatgacggtacgacgacgattgccaataac
cttaccagcacggttcaggtgttactgactcggagtaccagctcccgtacgtcctcgg
ctcggcgcacaaaggatgcctcccgcggttccagcagacgtcttcatgggtgccacagt
atggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgc
ctggagtactttccttctcagatgctgcgtaccggaaacaactttaccttcagctacac
ttttgaggacgttctttccacagcagctacgctcacagccagagctctggaccgtctca
tgaatcctctcatcgaccagctacctgtattacttgagcagaacaaacactccaagtgga
accaccagcagtcaggcttcagttttctcaggccggagcagtgacattcgggacca
gtctaggaactggcttccctggaccctgttaccgccagcagcagtgatcaaagacatctg

cgataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggc
agagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaa
gttttttctcagagcggggttctcatctttggaagcaaggctcagagaaaaaaatg
tggacattgaaaaggatgattacagacgaagaggaaatcaggacaaccaatcccgtg
gctacggagcagatggttctgtatctaccaacctccagagaggcaacagacaagcagc
taccgcagatgtcaacacacaaggcgttcttccaggcatggctctggcaggacagagatg
tgtacctcaggggcccattctgggcaagattccacacacggacggacattttcacc
tctcccctcatgggtggattcggacttaaacacctcctccacagatttctcatcaagaa
caccgcgtacctgcgaatccttcgaccacctcagtgcggaagtttgcttcttca
tcacacagtaactccacgggacaggtcagcgtggagatcgagtgggagctgcagaaggaa
aacagcaaacgctggaatcccgaattcagtaacttccaactacaacaagtctgttaa
tgtggactttactgtggacactaatggcgtgtattcagagcctcgcaccattggacca
gatactgactcgtaatctgtaattgcttgttaatcaataaacggtttaattcgtttca
gttgaactttggtctctgcgtatttcttcttatctagtttccatggctacgtagataa
gtagcatggcgggttaataactacagcccggcgtttaaacagcgggaggagg
tgagctcgtgacgtgaattacgtcataggggttagggaggtcctgtattagaggtcacgt
gagtgttttgcgacattttgcgacacctatgtggtctcgcctggggggggggggccgagtg
agcacgcagggtctccattttgaagcgggaggtttgaacgagcgtggcgcgctcactg
gccgtcgttttacaacgtcgtgactgggaaaacctggcgttaccacttaatacgcct
tgcagcacatcccccttccgagctggcgtaatagcgaagaggcccgaccgatcgc
cttcccatgcatcggccgcaataacctgcaggatccgttttgcgctgcttcgcgatgta
cgggccagatatacgcggtgacattgattattgactagttattaatagtaataaattac
ggggtcattagttcatagcccataatggagttccgcgttacataacttacggtaaatg
gcccgcctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgtt
cccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggta
aactgcccacttggcagtacatcaagtgtatcatatgccaaagtacgccccctattgacg
tcaatgacggtaaatggcccgcctggcattatgccagtacatgacctatgggacttt
cctacttggcagtacatctacgtattagtcacgcctattaccatgggtgatgcggttttg
gcagtacatcaatgggctggatagcgggttgactcacggggatttccaagtctccacc
ccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgt
cgtaacaactccgccccattgacgcaaatgggctgtaggcgtgtacgggtgggaggtcta
tataagcagagctctctggctaactagagaacctgcttactggcttatcgaaat
atacgaactactataggagacccaagctggctagcatggtgccgatggttatcttcc
agattggctcagggacactctctctgaaggaataagacagtggtggaagctcaaacctg
gccaccaccaccaaagcccgcagagcggcataaggacgacagcaggggtcttctgctt
cctgggtacaagtacctcggaccttcaacggactcagacaagggagagccgggtcaacga
ggcagacgcccgcggccctcagacacgacaaaagcctacgaccggcagctcagacagcggag
acaaccctgacctcaagtacaaccacgcccagcgggagtttcaggagcgccttaagaa
gatacgtcttttgggggcaacctcggacgagcagctcttccaggcgaaaaagagggttct
tgaacctctgggctggttgaggaaacctgttaagaccgctccgggaaaaaagaggccgg
tagagcactctcctgtggagccagactcctcctcgggaaccggaaggcgggcccagcag
cctgcaagaaaaagattgaattttggtcagactggagacgcagactcagtaacctgacc
ccagcctctcggacagccaccagcagccccctctggtctcggactaataacctcgccta
caggcagtgggcaccactcgcagacaataacgagggcgccgacggagtggttaattcc
tcgggaaattggcattgcgattccacatggctcggcgacagagtcataccaccagcac
ccgaacctgggcccctgccacctacaacaacctctacaacaaaatttccagccaat
caggagcctcgaacgacaatcactactttggctacagcacccttgggggtattttgac
ttcaacagattccactgccacttttaccacgtgactggcaagactcatcaacaaca

acgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacg
aacccccgttcagcccgaccgctgcgcttatccggttaactatcgtcttgagtccaac
ccggttaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagc
gaggtatgtaggcggtgctacagagttcttgaagtggcctaactacggctacacta
gaagaacagtatcttgggtatctgcgctctgctgaagccagttacctcggaaaaagagttg
gtagctcttgatccggcaaaaaaccaccgctggtagcgggtggtttttttggtttgaag
cagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgattttctaccgaag
aaaggcca

pIDTSMART-RC2-VP1del-GAL4-VP16



Sequence color-coding key

Feature	Color
Rep	text
Cap (VP1del)	text
SV40 promoter	text
Gal4	text
UL48 (VP16)	text
HSV TK poly(A) signal	text
SV40 poly(A) signal	text
KanR	text

Sequence

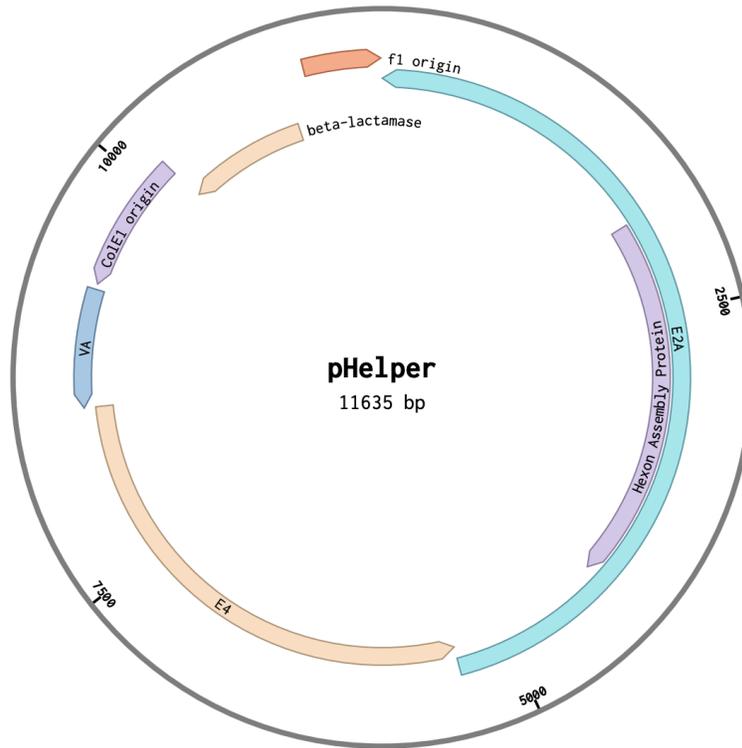
cccgtgtaaaacgacggccagtttatctagtcagcttgattctagctgatcgtggaccg
 gaagtgagccagtgagttgattgcagtcacgttacgctggagctctgaggctcgtcctg
 aatgatatgacgaccgocggagggttgcgtttgagacgggcgacagatccagtcgcgctg
 ctctcgtcgatccgctagggcgccgctctagaactagtgatccccggaagatcaga
 agttcctattccgaagttcctattctctagaaagtataggaacttctgatctgcgcagc
 cgccatgcccggggttttacgagattgtgattaaggtcccagcgaccttgacgagcatc
 tggccggcattttctgacagctttgtgaactgggtggccgagaaggaatgggagttgccg
 ccagattctgacatggatctgaatctgattgagcaggcaccctgaccgtggccgagaa
 gctgcagcgcgactttctgacggaatggcgccgtgtgagtaaggccccggaggccctt

tctttgtgcaat ttgagaagggagagagctacttccacatgcacgtgctcgtggaacc
accggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaat ttaccgcggtatcgagccgactttgccaaactggttcgcggtcacaaga
ccagaaatggcgccggaggcggaacaaggtggtggatgagtgctacatccccattac
ttgctccccaaaaccagcctgagctccagtgggcgtggactaatatggaacagtat tt
aagcgctgtttgaatctcacggagcgtaaacggttggtggcgcagcatctgacgcacg
tgctgcagacgcaggagcagaacaaagagaatcagaatcccaattctgatgcgccggtg
atcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggacaaggg
gattacctcggagaagcagtggtccaggaggaccaggcctcatacatctcctcaatg
cggcctccaactcgcggtcccaaatcaaggctgccttgacaatgoggaaagattatg
agcctgactaaaaccgccccgactacctggtgggcccagcagcccgtggaggacatttc
cagcaatcggatttataaaaat ttggaactaaacgggtacgatccccaatatgcggtt
ccgtctttctgggatgggccacgaaaaagttcggcaagaggaacaccatctggctgttt
gggctgcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgccctt
ctacgggtgcgtaaactggaccaatgagaactttccttcaacgactgtgtcgacaaga
tggtgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggccaaagcc
attctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagataga
cccgactcccgtgatcgtcacctccaacaccaacatgtgcgcctgattgacgggaact
caacgaccttcgaacaccagcagccgttgcaagaccggatggtcaaatttgaactcacc
cgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagactttttccg
gtgggcaaaggatcacgtggttgagggtggagcatgaattctacgtcaaaaagggtggag
ccaagaaaagaccgccccagtgacgcagatataagtgagcccaaacgggtgcgcgag
tcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggta
ccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttcctgcagacaat
gcgagagaatgaatcagaattcaaatatctgcttcaactcacggacagaaagactgttta
gagtgctttcccgtgtcagaatctcaaccgtttctgtcgtcaaaaaggcgtatcagaa
actgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctgcgatc
tggtcaatgtggatttgatgactgcatctttgaacaataa atgatttaaatcagggtc
cgctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataagac
agtgggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggac
gacagcaggggtcttggtcttccctgggtacaagtacctcggacccttcaacggactcga
caagggagagccggtcaacgaggcagacgccgcggccctcgagcacgacaaagcctacg
accggcagctcgacagcggagacaaccgctacctcaagtacaaccacgccgacgcggag
tttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagctct
ccaggcgaaaaagaggggtcttgaacctctgggctggttgaggaacctgttaagacgg
ctccgggaaaaaagaggccggtagagcactctcctgtggagccagactcctcctcggga
accggaaaggcgggcccagcagcctgcaagaaaaagattgaattttggtcagactggaga
cgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctggtc
tgggaaactaatcagatggctacaggcagtgggcgcaccaatggcagacaataacgagggc
gccgacggagtggttaattcctcgggaaattggcattgcgattccacatggatgggcca
cagagtcataccaccagcaccgaaacctgggcccctgccacctacaacaaccacctct
acaacaaaat ttccagccaatcaggagcctcgaacgacaatcactactttggctacagc
acccttgggggtat ttgacttcaacagattccactgccacttttaccacgtgactg
gcaaagactcatcaacaactggggattccgacccaagagactcaacttcaagctct
ttaacattcaagtcaaagaggtcacgcagaatgacggtagcagcagattgccaataac
cttaccagcacggttcaggtggttactgactcggagtagcagctcccgtacgtcctcgg
ctcggcgcacaaaggtgcctcccgcggtccagcagacgtcttcatgggtgccacagt
atggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgc

ctggagtactttccttctcagatgctgctaccggaacaactttaccttcagctacac
ttttgaggacggttcctttccacagcagctacgctcacagccagagtctggaccgtctca
tgaatcctctcatcgaccagtacctgtattacttgagcagaacaaactccaagtga
accaccacgcagtcaaggcttcagttttctcaggccggagcagtgacattcgggacca
gtctaggaactggcttcctggacctgttaccgccagcagcagtgatcaaagacatctg
cggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggc
agagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaa
gttttttctcagagcggggttctcatctttgggaagcaaggctcagagaaaaaaaatg
tggacattgaaaaggctcatgattacagacgaagaggaaatcaggacaaccaatcccgtg
gctacggagcagtgatggttctgtatctaccaacctccagagaggcaacagacaagcagc
taccgcagatgtcaacacacaaggcgttcttccaggcatggctctggcaggacagagatg
tgtacctcaggggcccctctgggcaagattccacacacggacggacattttcacccc
tctcccctcatgggtggattcggacttaaacacctcctccacagatttctcatcaagaa
caccocggtacctgcgaatccttcgaccaccttcagtgcggaagtttgcttcttca
tcacacagtaactccacgggacaggtcagcgtggagatcagtgaggagctgcagaaggaa
aacagcaaacgctggaatcccgaattcagtaacttccaactacaacaagtctgttaa
tgtggactttactgtggacactaatggcgtgtattcagagcctcgcaccttgccacca
gatacctgactcgtaatctgtaattgcttggttaataataaacctttaaattcgtttca
gttgaactttggtctctgctgattttcttcttatctagtttccatggctacgtagataa
gtagcatggcgggttaatacattaactacagcccggcgtttaaacagcgggaggggg
tggagtcgtgacgtgaattacgtcataggggttagggaggtcctgtattagaggtcacgt
gagtgttttgcgacattttgcgacaccatgtggtctcgctggggggggggggcccagtg
agcacgcagggtctccattttgaagcgggaggtttgaacgagcgtggcgcgctcactg
gccgtcgttttacaacgtcgtgactgggaaaacctggcgttaccaacttaatcgct
tgcagcacatcccccttctgcagctggcgtaatagcgaagaggcccgcaccgatcgc
cttcccctcatcgccgcaataacctgcagggtgtggaaagtccccagggtccccag
caggcagaagtatgcaagcatgcatctcaattagtcagcaaccagggtgtggaaagtcc
ccagggtccccagcaggcagaagtatgcaagcatgcatctcaattagtcagcaaccat
agtcccgccttaactccgcccctccgccccttaactccgcccagttccgcccattctc
cgcccctatggctgactaattttttttatttatgcagaggccgaggccgctcggcctct
gagctattccagaagtagtgaagaggcttttttggaggagatctaagcttgaagcaagc
ctcctgaaagatgaagctactgtcttctatcgaacaagcatgcatatttgcccactta
aaaagctcaagtgtccaagaaaaaccgaagtgcgccaagtgtctgaagaacaactgg
gagtgctcgtactctccaaaacaaaaggctctccgctgactagggcacatctgacaga
agtggaatcaaggctagaagactggaacagctatcttactgatttttctcgcagaag
accttgacatgatthtgaatggattctttacaggatataaaagcattgttaacagga
ttatttgtaacaagataatgtgaataaagatgccgtcacagatagattggcttcagtgga
gactgatatgcctctaacttgagacagcatagaataagtgcgacatcatcatcggaag
agagtagtaacaaagggtcaaagacagttgactgtatcgccggaattccccgggatctgg
gccccccgaccgatgtcagcctgggggacgagctccacttagacggcgaggacgtggc
gatggcgcatgccgacgcgctagacgatttcgatctggacatggtgggggacggggatt
ccccgggtccgggatttccccccacgactccgcccctacggcgctctggatattggc
gacttcgagtttgagcagatgtttaccgatgcccttggattgacgagtagcgtgggta
ggggggcgcgaccggaccgcacatccccgtctgggttttcccctcccggtcccggttcgt
atccacaataaacacgagcacatacattacaaaacctgcgggtgtcgtctgattatt
gggtggatccgggggagatgggggaggctaactgaaacacggaaggagacaataaccgga
ggaaccgcgctatgacggcaataaaaagacagaataaaacgcacgggtgttgggtcgt
ttgttdataaacgcggggttcgggtcccagggtggcactctgtcgataccccaccgaga

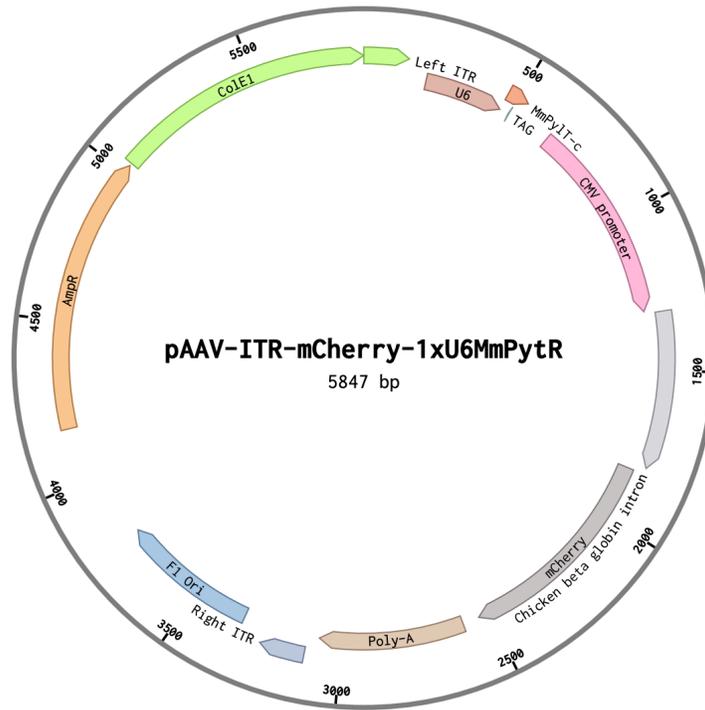
ccccattggggccaatacgcgccggtttcttccctttccccaccccaccccccaagttc
gggtgaaggcccagggctcgcagccaacgtcggggcggcaggccctgccatagccactg
gccccgtgggttagggacgggggtcccccatggggaatggtttatgggtcgtgggggta
ttatthttgggcggttgcgtgggggtcaggtccacgacccaagcttgtcgcaggtaccccg
ggaattcgagctctagattaagtaatgatcataatcagccatatcacatctgtagaggt
ttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatg
caattggtggttgaacttggttattgcagcttataatggttacaataaagcaatagc
atcacaatthtcacaataaagcatttttttactgcattctagttgtgggtttgtccaa
actcatcaatgtatcttatcatgtctggatccctaggggtgcgagcggatcgagcagtg
tcgatcactactggaccgagctgtgctgogaccogtgcattacggcattatacgt
tgatcgggtccacgatcagctagattatctagtcagcttgatgtcatagctgthtccctga
ggctcaatactgaccatttaaatcatacctgacctccatagcagaaagtcaaaagcctc
cgaccggaggctthtgacttgatcggcagctaagaggttccaactthcaccataatgaa
ataagatcactaccgggctgattttttgagttatcgagattthtcaggagctaaggaagc
taaatgagccatattcaacgggaaacgtcttgcttgaagccgagattaaattccaaca
tggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcg
acaatctatcgattgtatgggaagcccgatgcgccagagttgthtctgaaacatggcaa
aggtagcgttgccaatgatgttacagatgagatggtcaggctaaactggctgacggaat
ttatgcctcttccgaccatcaagcattthtatccgactcctgatgatgcatggthactc
accactgcgatcccagggaaaacagcattccaggtattagaagaatatcctgattcagg
tgaanaatattggtgatgcgctggcagtgthcctgcgccggttgcatcctgattcctgtht
gtaattgtcctthtaacggcgatcgcgattthcgtctcgcctcaggcgcaatcacgaatg
aataacggthtggtgggtgcgagtgattthgatgacgagcgtaatggctggcctgttg
acaagtctggaaagaaatgcataaactcttgccattctcaccggattcagtcgctcactc
atggtgattthctcacttgataacctattthtgacgaggggaaattaataggttgatt
gatgthggacgagtcggaatcgcagaccgataaccaggatcttgccatcctatggaactg
cctcggtgagthtctccttcattacagaaacggctthtcaaaaatatggtattgata
atcctgatatgaanaaatgcagthtctacttgatgctcgcgatgagthtctaatgagga
cctaaatgtaatcacctggctcacctcgggtgggcctthtctgcgthtggcgttht
ccataggctccgccccctgacgagcatcacaanaatcgatgctcaagtcagaggtggc
gaaaccgacaggaactataaagataaccaggcgtthccccctggaagctccctcgtgcgc
tctcctgthccgacctgcccgttaccggatacctgtccgcctthtctccttcgggaag
cgtggcgctthtctcatagctcacgctgtaggtatctcagthtccggtgtaggtcgttcgct
ccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcttatccggt
aactatcgtcttgagthcaaccggtaagacacgacttatcgcactggcagcagccac
tggtaacaggattagcagagcaggtatgtagggcgtgctacagagthcttgaaagtgg
ggcctaactacggctacactagaagaacagtatthggtatctgcgctctgctgaagcca
gthacctcggaaaagagthggtagctcttgatccggcaaacacaaccaccgctggtagc
gggtggthtthtggthtgcaagcagcagattacgcgcagaaaaaaggatctcaagaaga
tcctthgattthtctaccgaagaaaggcca

pHelper



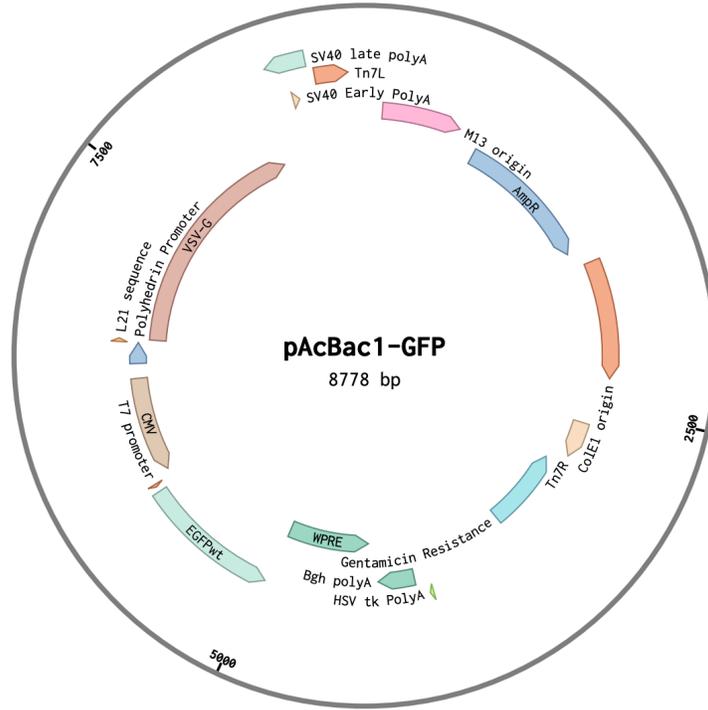
The sequence for this plasmid is the same as the pHelper sequence found in the Chapter 2 Plasmids section of this appendix.

pAAV-mCherry+1xMmPytR



The sequence for this plasmid is the same as the pAAV-mCherry+1xMmPytR sequence found in the Chapter 2 Plasmids section of this appendix.

pAcBac1-GFPwt



Sequence color-coding key

Feature	Color
M13 origin	text
AmpR	text
ColE1 origin	text
Tn7R	text
GentamicinR	text
HSV tk PolyA	text
Bgh polyA	text
WPRE	text
EGFPwt	text
T7 promoter	text
CMV promoter	text
Polyhedrin Promoter	text
L21 sequence	text
VSV-G	text
SV40 late polyA	text
Tn7L	text

Sequence

ttctctgtcacagaatgaaaattttctgtcatctcttcggttattaatgtttgtaattg
 actgaatatcaacgcttatttgcagcctgaatggcgaatgggacgcgccctgtagcggc

gcattaagcgcggcggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgc
cctagcgcgccgctcctttcgtttcttcccttcttctcgcacgctcgcggctttc
cccgtcaagctctaaatcggggctccctttagggttccgatttagtgctttacggcac
ctcgaaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgcctgata
gacggtttttcgccctttgacggtggagtcacgcttctttaatagtgactcttgttcc
aaactggaacaactcaaccctatctcggctattcttttgatttataagggattttg
ccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaatt
taacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgvcgcgaa
ccctatttgttttatttttctaaatacattcaaataatgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccg
tgctgcccttattcccttttttgcggcattttgccttctgtttttgctcaccagaaa
cgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaa
ctggatctcaacagcggtaagatccttgagagtttcgccccgaagaacgttttccaat
gatgagcacttttaagttctgctatgtggcgcggtattatcccgtattgacgccgggc
aagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactacca
gtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtgccat
aacatgagtgataaactgcggccaacttacttctgacaacgatcggaggaccgaagg
agctaaccgcttttttgacaacatgggggatcatgtaactcgccttgatcgttgggaa
ccggagctgaatgaagccatacacaacgacgagcgtgacaccagatgctgtagcaat
ggcaacaacgttgcgcaaacattaactggogaactacttactctagcttcccggcaac
aattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggccct
ccggtggtggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtat
cattgcagcactggggccagatggtaagccctcccgtatcgtagtattctacacgacgg
ggagtccaggcaactatggatgaacgaaaatagacagatcgtgagataggtgctcactg
attaagcattggttaactgtcagaccaagtttactcatatatacttttagattgatttaa
acttcatttttaatttaaaaggatctaggtgaagatcctttttgataatctcatgacca
aaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaa
ggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaacaaaaaaacc
accgctaccagcgggtggtttgtttgccggatcaagagctaccaactctttttccgaagg
taactggcttcagcagagcgcagatacacaataactgtccttctagtgtagccgtagtta
ggccaccacttcaagaactctgtagcaccgcctacatacctcgtctgctaactcctgtt
accagtggctgctgccagtgggcgataagtcgtgtcttaccgggttgactcaagacgat
agttaccggataagggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaaagcgc
cacgcttcccgaaggagaaaggcggacaggtatccggtaagcggcagggctcggaaacag
gagagcgcacgagggagcttccaggggaaacgcctggtatctttatagtcctgtcggg
tttcgccacctctgacttgagcgtcagatttttgatgctcgtcaggggggaggagcct
atggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttg
ctcacatgttctttcctgcgttatccctgattctgtggataaccgtattaccgccttt
gagtgagctgataaccgctcggcgcagccgaacgaccgagcgcagcagtcagtgagcga
ggaagcggaaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcac
accgcagaccagccgcgtaacctggcaaaatcggttacggttgagtaataaatggatgc
cctgcgtaagcgggtgtggggcggacaataaagtcttaaaactgaacaaaatagatctaaa
ctatgacaataaagtcttaaaactagacagaatagttgtaaaactgaaatcagtcagtta
tgctgtgaaaaagcatactggactttttgttatggctaaagcaaaccttcattttctga
agtgcaaatgcccgtcgtattaagagggggcgtggccaagggcatggtaaaagactata
ttcgcggcgttggtgacaatttaccgaacaactccgcggccgggaagccgatctcggcct
gaacgaattgttaggtggcggtaacttgggtcagatatcaaaagtgcatacacttcttcccg

atgcccactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcaagta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtg
caatgcctgctccgggtgctcgccggagactgcgagatcatagatatagatctcacta
cgcggtgctcaaacctgggcagAACGTAAGCCGCGAGAGCGCCAACAACCGCTTCTT
gtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccagaggaatcgg
agtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatca
agagcagcccgcgatggatttgacttggtcagggccgagcctacatgtgccaatgatgcc
catacttgagccacctaactttgttttagggcgactgccctgctgcgtaaacatcgttgc
tgctgcgtaaacatcgttgcctcacaacatcaaacatcgaccacggcgtaacgcgc
ttgctgcttggatgcccagggcatagactgtacaaaaaacagtcataacaagccatga
aaaccgccactgcccgttaccaccgctgcgttcggtaaacggttctggaccagttgcgt
gagcgcatacgtacttgcattacagtttacgaaccgaacaggcttatgtcaactgggt
tcgtgccttcatccgtttccacgggtgctgcgtcaccggcaaccttgggcagcagcgaag
tcgaggcatttctgtcctggctggcgaacgagcgaaggtttcggctctccacgcacgt
caggcattggcggccttgcgttcttctacggcaaggtgctgtgcacggatctgcctg
gcttcaggagatcggtagacctcggccgtcgcggcgcttgcgggtgggtgctgacccgg
atgaagtggttcgcacctcgggttttctggaaggcgagcactcgtttgttcgccaggac
tctagctatagttctagtggttggcctacgtaccgcgtatgtggctatggcagggcttgc
ctaatgcgcgcgtacagggcgcgtggggataccccctagagcccagctgggtctttc
cgctcagaagccatagagcccaccgcacccccagcatgctgctattgtcttcccaat
cctcccccttgcgtgctcgtccccacccccacccccagaatagaatgacacactactcaga
caatgcgatgcaatttctcattttattaggaaaggacagtgaggagtgccaccttccag
ggtaaggaaggcacgggggaggggcaaacacagatggctggcaactagaaggcacag
tcgaggctgatcagcgggttaaaccgggcccctctagactcgagttaaagtgcagcggg
gagggcggccaaaggagatccgactcgtctgagggcgaaggcgaagacgcggaagagg
ccgcagagccggcagcaggccgcgggaagggaaggtccgctggattgagggccgaaggga
cgtagcagaaggacgtcccgcgcagaatccaggtggcaacacaggcgagcagccaagga
aaggacgatgatttccccgacaacaccacggaattgtcagtgcccaacagccgagcccc
tgtccagcagcgggcaaggcagggcggcgatgagttccgccgtggcaatagggaggggga
aagcgaaagtcccggaaaggagctgacaggtgggtggcaatgccccaccagtggggggt
gctcagcaaacacagtgacaccacgccacgttgcctgacaacgggcccacaactcctc
ataaagagacagcaaccaggatttatacaaggaggagaaaatgaaagccatacgggaag
caatagcatgatacaaaggcattaaagcagcgtatccacatagcgtaaaaggagcaaca
tagttaagaataccagtcattctttcacaattttgtaatccagaggttgattgtcgac
ttaacgcgttgaattctcattaatgggtgatgggtgatgatgaccggatgcatattcaga
tcctcttctgagatgagttttgttcgaagggccccttgtacagctcgtccatgccgag
agtgatcccggcggcgggtcacgaactccagcaggaccatgtgatcgcgcttctcgttgg
ggcttttgcctcagggcggactgggtgctcaggtagtggttgtcgggcagcagcagggg
ccgtcgcgatgggggtgttctgctggtagtggtcggcgagctgcacgctgccgtcctc
gatgttgtggcggatcttgaagttcaccttgatgccgttcttctgcttgtcggccatga
tatagacgttgtggctgttgttagttgtactccagcttgtgcccaggatggtgcccgtcc
tccttgaagtgcgatgccctcagctcgatgcggttcaccagggtgtcgcctcgaactt
cacctcggcgcgggtcttgtagttgccgtcgtccttgaagaagatgggtgcgctcctgga
cgtagccttcgggcatggcggacttgaagaagtcgtgctgcttcatgtggtcggggtag
cggctgaagcactgcacgccgtaggtcaggggtggtcacgaggggtgggcccagggcacggg
cagcttgcgggtgggtgcagatgaacttcaggggtcagcttgcgtaggtggcatcgcct
cgccctcgcggacacgctgaacttgtggcggtttacgtcgcgctccagctcgaccagg
atgggcaccacccgggtgaacagctcctcgccttgcctcaccatgggtggcggcgctagc

cagcttgggtctccc**tatagtgagtcgtatta**atctcgataagccagtaagcagtgggg
tctctagttagccagagagctctgcttatatagacctcccaccgtacacgcctaccgcc
catttgcgtcaatggggcggagttggttacgacattttggaaagtcctcggtgattttggt
gcaaaaactcccattgacgtcaatgggggagacttgaaatcccctgagtc
aacgctatccacgccattgatgtactgcaaaaccgcatcaccatggtaatagcgt
gactaatagctagatgtactgccaagtaggaaagtcccataagggtcatgtactgggcat
aatgccaggcgggccattaccgtcattgacgtcaatagggggcgtacttgccatatga
tacacttgatgtactgccaagtgggcagtttacccgtaaatagtccaccattgacgtca
atggaaagtcctattggcgttactatgggaacatacgtcattattgacgtcaatgggc
gggggctggtggggcggtcagccaggcgggccatttacccgtaagttatgtaacgcggaac
tccatataatgggctatgaactaatgaccccgtaattgattactattaataactagtcaa
taatcaatgtcaac**gcgtatatctggccgtacatcgcgaagcagcgc**caaaacggatcc
tgcaggatatttgcggcgggtccgtatactccggaatattaatag**atcatggagataa**
ttaaaatgataaccatctcgcaaaataaataagtatcttactgtttcgtaacagttttg
taataaaaaaacctataaatattccggattattcataaccgtcccaccatcggggcgc**aa**
ctcctaaaaaacgccaccatgaagtgccttttgtacttagcctttttattcattgggg
tgaattgcaagttcaccatagttttccacacaaccaaaaaggaaactggaaaaatggt
ccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgacttaat
aggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacgggt
ggatgtgtcatgcttccaaatgggtcactacttgtgatttccgctggatggaccgaag
tatataacacattccatccgatccttccactccatctgtagaacaatgaaggaaagcat
tgaacaaacgaaacaaggaacttggtgaatccaggcttccctcctcaaagttgtggat
atgcaactgtgacggatgccgaagcagtgattgtccagggtgactcctcaccatgtgctg
ggtgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaa
ttacatatgcccactgtccataactctacaacctggcattctgactataagggtcaaa
ggctatgtgattctaacctcattccatggacatcaccttcttctcagaggacggagag
ctatcatccctgggaaaggaggccacagggttcagaagtaactactttgcttatgaaac
tgagggaaggcctgcaaatgcaatactgcaagcattggggagtcagactcccatcag
gtgtctggttcgagatggctgataaggatctctttgctgcagccagattccctgaatgc
ccagaagggtcaagtatctctgctccatctcagacctcagtggtatgtaagtctaattca
ggagttgagaggatcttgattattccctctgccaagaaacctggagcaaaatcagag
cgggtcttccaatctctccagtggtctcagctatcttgctcctaaaaaccaggaacc
ggctctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcag
agtcgatattgctgctccaatcctctcaagaatggtcggaatgatcagtggaactacca
cagaaagggaactgtgggatgactgggcaccatataagacgtggaaattggaccaat
ggagttctgaggaccagttcaggatataagtttctttatacatgattggacatgggat
ggtggactccgatcttcatcttagctcaagggtcaggtgttcgaacatcctcacattc
aagacgctgcttcgcaacttctgatgatgagagtttattttttggtgatactgggcta
tcaaaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgc
ctctttttctttatcatagggttaatcattggactattcttggttctccgagttggt
tccatctttgcattaataaagcacaccaagaaaagacagatttatacagacatagag
atgaaccgacttgaaagtgataa**ggccaggccggccaagcttgctcgagaagtactaga**
ggatcataatcagccataccacatttgtagaggtttacttgcttt**aaaaaacctocca**
cacctcccctgaacotgaaacataaaaatgaatgcaattggtggttaacttgttat
tgcagcttataatggttacaataaagcaatagcatcacaatttcacaaataaagcat
tttttctactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgct
tgatctgatcactgcttgagcctaggagatccg**aaccagataagtgaaatctagttcc**
aaactatcttctcatttttaattttcgtattagcttacgacgtacacccagttccat

ctat tttgtc actottcc otaaataatc otaaaaaactcc atttcc acccctccc agtt
ccaactat tttgtccgccaca gcggggcatttttctt cctgttatg tttttaatcaa
acatcctgccaactccatgtgacaaaaccg tcatcttcggctacttt

Appendix II

Oligonucleotide primers

Chapter 2 primers

Primer name	Oligonucleotide sequence (5' → 3')
Consensus-splice-donor-site-sequence-F	CGCGGAAGCTTCGATCAACTACGCAGACAGGTAAGTAAACAAATGT TCTCGTCACGTGGG
NheI-p40-F	ATTATTAGCTAGCAGTTGCGCAGCCATCGACGTCAGACCGGAAGCT TCGATCAACTACG
Cap-SalI-R	ATTATTAGTCGACTTACAGATTACGAGTCAGGTATCTGGCAATGG
SbfI-Rep-F	AATTAATTCCCTGCAGGCTCTAGAAGTAGTGGATCCCCCGGAAGATC
NotI-Rep-R	AATTAATTGCGCCGCTCAGAGAGAGTGTCCCTCGAGCC AATCTGG

Chapter 3 primers

Primer name	Oligonucleotide sequence (5' → 3')
AAPstop60-F	GGATGGGCGACAGAGTGATCACCACCAGCACC
AAPstop60-R	GGTGCTGGTGGTGATCACTCTGTGCGCCATCC
RC2int-HindIII-F	CGTCAGACGCGGAAGCTTCGATCAAC
RC2-SbfI-R	GCAAATACCTGCAGGATCCGTTTTGCGCTG
NheI-Kozak-AAP-WT-F	ATTATTGCTAGCGCCGCCACCATGGCCTAGCTGGAGACGCAGAC TCAGTACCTGACC
XhoI-AAP-R	ATTATTCTCGAGTCAGGGTGAGGTATCCATACTGTGGCACC
AAP-L12TAG-F	CCTGACCCCGAGCTAGTCGGACAGCC
AAP-L12TAG-R	GGTGGCTGTCCGACTAGCTGGGGGTCAGG
AAP-T78TAG-F	GGGCCCTGCCACCTACATAGACCACCTCTACAAACAAATTTCC
AAP-T78TAG-R	GTTTGTAGAGGTGGTCTATGTAGGTGGGCAGGGCCAGGTTCCGG G
AAP-T124TAG-F	CCACGTGACTGGCAAAGACTAGTCAACAACAACCTGGGG
AAP-T124TAG-R	CCAGTTGTTGTTGACTAGTCTTTGCCAGTCACGTGGTG
AAP-T177TAG-F	GTACCAGCTCCCGTTAGTCCTCGGCTCGGCGCA
AAP-T177TAG-R	CCGAGCCGAGGACTAACGGGAGCTGGTACTCCG
AAP-T97TAG-F	CGAACGACAATCACTTAGTTGGCTACAGCACCCCTTGG
AAP-T97TAG-R	GGGGTGCTGTAGCCAATAAGTGATTGTCGTTGAGG
AAP-T110TAG-F	GGGTATTTTGACTTCATAGGATTCCACTGCCACTTTTC
AAP-T110TAG-R	GTGGCAGTGGAATCCTATGAAGTCAAATACCCC
AAP-T78TGA-F	GGGCCCTGCCACCTACATGAACCACCTCTACAAACAAATTTCC
AAP-T78TGA-R	GTTTGTAGAGGTGGTTCATGTAGGTGGGCAGGGCCAGGTTCCGG G
AAP-T97TGA-F	CGAACGACAATCACTTGATTGGTAACAGCACCCCTTGG
AAP-T97TGA-R	GGGGTGCTGTAGCCAATCAAGTGATTGTCGTTGAGG
AAP-T110TGA-R	GTGGCAGTGGAATCTCATGAAGTCAAATACCCC
AAP-T110TGA-F	GGGTATTTTGACTTAATGAGATTCCACTGCCACTTTTC
AAP-T124TGA-F	CCACGTGACTGGCAAAGACTGATCAACAACAACCTGGGG
AAP-T124TGA-R	CCAGTTGTTGTTGATCAGTCTTTGCCAGTCACGTGGTG
AvrII-CMV-WRS-F	ATTATTCCTAGGTTATTAATAGTAATCAATTACGGGGTCATTAG TTCATAGCCC
SpeI-WRS-R	AATAATACTAGTCCATAGAGCCCACCGCATCCCCAGC

Chapter 4 primers

Primer name	Oligonucleotide sequence (5' → 3')
MTH-mCherry-SbfI-F	ATTATTCCTGCAGGACCATGGCTCGAGATCCCGGGTGATC
MTH-mCherry-NotI-R	AATAATGCGGCCAGACATGATAAGATACATTGATGAGTTTG GACAAACC
NheI-linker-EGFP-F	TAATAATGCTAGCAGGGCGGCTCCGTGAGCAAGGGCGAGGA GCTGTTCCACCGGG
GFP-EcoRI-6xHis-R	ATTATTGAATTCCTCATTAATGGTGATGGTGATGATGACCG G
NheI-Kozak-mCherry-F	ATTATTGCTAGCGCCGCCACCATGGTGAGCAAGGGCGAGGA GG
mCherry-linker-AvrII-R	TAATAATCCTAGGAGCCGCCCTTGTACAGCTCGTCCATGCC GCCGGTGGAGTGG
Gibson-mCherry-F	GTTATTGTGCTGTCTCATCATTTTTGGCAAAGAATTGGCCAA GGAGGCCACCATGGTGAGC
Gibson-WPRE-R	GTCGATCGACCACTGTGCTGGCGGCCGGCCAGGCCGCGGGG AGG
Gibson-UbiC-F	CCAGTAAGCAGTGGGTTCCTCTAGTTAGCCAGAGAGCTCTAG ACCAAGTGACGATCACAGC
Gibson-UbiC-R	GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAG TGGCCTCCGCGCCGG
NheI-Lemke-NES-MbPylRS-F	ATTATTGCTAGCGCCACCATGGCCTGCCAGTGCCCTTCA GTTGCCTCCGCTGGAGAGACTGACCCTCGACGATAAAAAAC CATTAGATGTTTTAATATCTGCGACCGG
MbPylRS-WPRE-XhoI-R	AATAATCTCGAGTTAAAGTCGACGCGGGG
L274A-F	GCTTGCCCCGACTCTTGCCAACATATGTGCGAAAACCTCGATA GGATTTTACCAGGCC
L274A-mut-R	CCATGGTTTGTCAATACCCCATTTCTCTATCAAGAGACCGTG GCCCGACGACTGC
V370R-F	CTTCGGCAGTCGTCGGGCCACGGTCTCTTGATAGAGAATGG GGTATTGACAAACCATG
V730R-mut-R	CCATGGTTTGTCAATACCCCATTTCTCTATCAAGAGACCGTG GCCCGACGACTGC
C313V-M315Y-F	GAGCACCTGGAAGAATTTACTATGGTGAACCTTCGTGCAGTA CGGTTCCGGATGTACTCGGGAAAATCTTG
C313V-M315Y-mut-R	CAAGATTTTCCCGAGTACATCCCGAACCGTACTGCACGAAG TTCACCATAGTAAATTC
MbPylRS-Nterm-F	ATTATTGCTAGCGCCACCAATGGATAAAAAACCATTAG
MbPylRS-Cterm-R	AATAATGAATTCCTTACAGATTGGTTGAAATCCCATTTATAGT AAGATTCGGACC

Chapter 5 primers

Primer name	Oligonucleotide sequence (5' → 3')
XbaI-eGFP-R	AATAATTCTAGATCATTAATGGTGATGGTGATGGTGATGATGACCGG
SphI-Kozak-GFP-F	ATTATTGCATGCGCCGCCACCGCTGGCTAGCGCCGCCACCATG GTGAGCAAGGGCG
TATA-GFP-F	ATTATTGCATGCATTTATAAAAGAAATTAATACGACTCACTAT AGGGAGACCCAAGCTGG
YB-TATA-GFP-F	ATTATTGCATGCATTTCTAGAGGGTATATAATGGGGGCCAAAA TTAATACGACTCACTATAGGGAGACCCAAGCTGGC
min-CMV-GFP-F	ATTATTGCATGCATTTGTAGGCGTGTACGGTGGGAGGTGTATAT AAGCAGAGCTCGTTTAGTGAACCGTCAGATCAAATTAATACGA CTCACTATAGGGACCAAGCTGG
KpnI-GFP-R	AATAATGGTACCTCAATGGTGATGGTGATGATGACCGGTATGC
SphI-VP1-only-F	ATTATTGCATGCATGGCTGCCGATGGTTATCTTCCAGATTG
XbaI-VP1-only-R	AATAATTCTAGATGTAGTTAATGATTAACCCGCCATGCTACTT ATCTACG
SphI-TATA-VP1-only-F	ATTATTGCATGCTATAAAAGAAATATGGCTGCCGATGGTTATC TTCCAGATTG
SphI-YBTATA-VP1-only-F	TATTGCATGCTCTAGAGGGTATATAATGGGGGCCAATGGCTGC CGATGGTTATCTTCCAG
SphI-minCMV-VP1-only-F	ATTATTGCATGCGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTCAGATCATGGCTGCCGATGGT TATCTTCCAGATT
SbfI-Gal4-VP16-F	ATTATTCCCTGCAGGGGTGTGGAAAGTCCCCAGGCTCCCC
AvrII-Gal4-VP16-R	AATAATCCTAGGGATCCAGACATGATAAGATACATTGATGAGT TTGGACAAACC