



# Plasmids 101

A Desktop Resource

Created and Compiled by Addgene  
March 2017 (3rd Edition)

[www.addgene.org](http://www.addgene.org)

# INTRODUCTION TO PLASMIDS 101

By The Addgene Team | March, 2017

Any newcomer who joins a molecular biology lab will undoubtedly be asked to design, modify, or construct a plasmid. Although the newcomer likely knows that a plasmid is a small circular piece of DNA often found in bacterial cells, additional guidance may be required to understand the specific components that make up a plasmid and why each is important.

Our mission with this eBook, Plasmids 101: A Desktop Resource, is to curate a onestop reference guide for plasmids. This resource is designed to educate all levels of scientists and plasmid lovers. It serves as an introduction to plasmids, allowing you to spend less time researching basic plasmid features and more time developing the clever experiments and innovative solutions necessary for advancing your field.

~ The Addgene Team



[addgene.org](http://addgene.org)



[blog.addgene.org](http://blog.addgene.org)



[facebook.com/addgene](https://facebook.com/addgene)



[twitter.com/addgene](https://twitter.com/addgene)



[linkedin.com/company/addgene](https://linkedin.com/company/addgene)



[youtube.com/addgene](https://youtube.com/addgene)



[blog.addgene.org/topic/podcast](http://blog.addgene.org/topic/podcast)

# TABLE OF CONTENTS

Page	Section
2	<a href="#">Introduction to Plasmids 101</a>
7	<b><a href="#">Chapter 1: What is a Plasmid?</a></b>
8	<a href="#">A Brief History of Plasmids</a>
10	<a href="#">What is a Plasmid?</a>
12	<a href="#">Antibiotic Resistance Genes</a>
14	<a href="#">Common Antibiotics Table</a>
15	<a href="#">Origin of Replication</a>
18	<a href="#">The Promoter Region</a>
25	<a href="#">Terminators and PolyA Signals</a>
28	<a href="#">Methylation and Restriction Enzymes</a>
31	<a href="#">Blue-White Screening</a>
34	<a href="#">Common Lab E. coli Strains</a>
39	<a href="#">E. coli Strains for Protein Expression</a>
44	<b><a href="#">Chapter 2: Common Cloning Techniques</a></b>
45	<a href="#">Restriction Cloning</a>
51	<a href="#">Golden Gate Cloning</a>
55	<a href="#">TOPO Cloning</a>
58	<a href="#">Sequence and Ligation Independent Cloning (SLIC)</a>
61	<a href="#">CcdB - The Toxic Key to Efficient Cloning</a>

## TABLE OF CONTENTS (CONT'D)

Page	Section
64	<a href="#">Gateway Cloning</a>
70	<a href="#">Gibson Cloning</a>
<b>74</b>	<b><a href="#">Chapter 3: Eukaryotic Expression Vectors</a></b>
75	<a href="#">Mammalian Vectors</a>
78	<a href="#">Yeast Vectors</a>
81	<a href="#">Multicistronic Vectors</a>
<b>85</b>	<b><a href="#">Chapter 4: Viral Expression Vectors</a></b>
86	<a href="#">Viral Vectors – An Introduction</a>
88	<a href="#">Viral Vector Elements</a>
92	<a href="#">Your Lentiviral Plasmid FAQs Answered</a>
95	<a href="#">AAV: A Versatile Tool for Gene Expression in Mammals</a>
<b>100</b>	<b><a href="#">Chapter 5: Plasmids That Glow</a></b>
101	<a href="#">History of Fluorescent Proteins</a>
102	<a href="#">Green Fluorescent Protein (GFP)</a>
106	<a href="#">Which Fluorescent Protein Should I Use?</a>
109	<a href="#">Choosing Your Fluorescent Proteins for Multi-Color Imaging</a>
112	<a href="#">Tips for Using FRET in Your Experiments</a>
117	<a href="#">Luciferase</a>
<b>120</b>	<b><a href="#">Chapter 6: Plasmid Tags</a></b>

## TABLE OF CONTENTS (CONT'D)

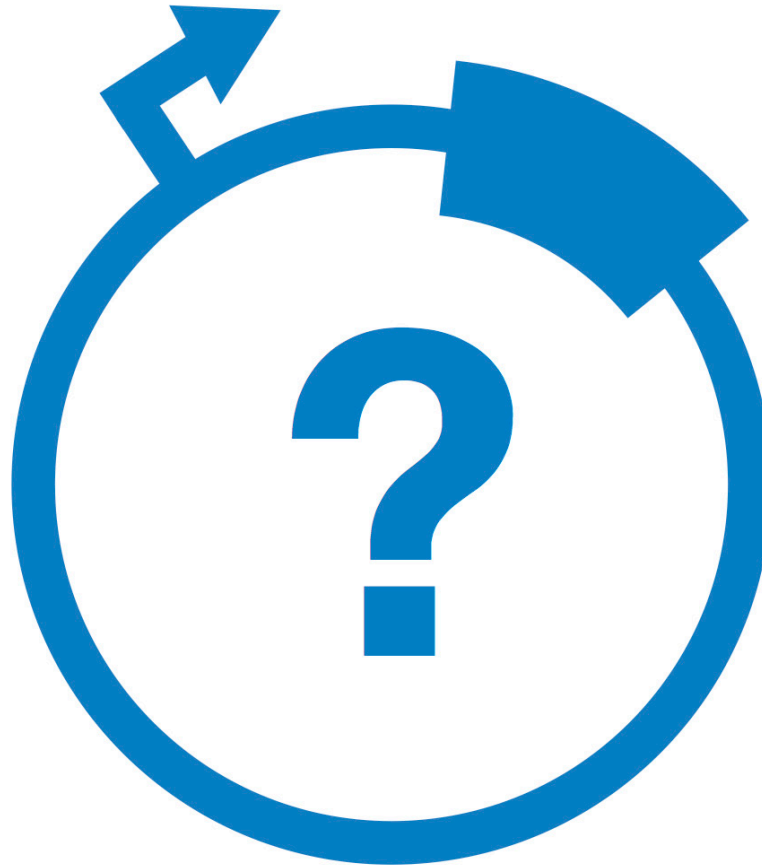
Page	Section
121	<a href="#">Protein Tags</a>
126	<a href="#">Tag Your Favorite Yeast Genes with Ease</a>
<b>128</b>	<b><a href="#">Chapter 7: Genome Engineering</a></b>
129	<a href="#">Introduction to Genome Engineering</a>
133	<a href="#">Cre-Lox</a>
137	<a href="#">Knockout/Knock-In Plasmids</a>
142	<a href="#">Overview of TALEN Technology</a>
145	<a href="#">Overview of CRISPR Technology</a>
148	<a href="#">FLEX Vectors</a>
152	<a href="#">Sleeping Beauty Awakens for Genome Engineering</a>
<b>155</b>	<b><a href="#">Chapter 8: You've Made a Plasmid ... Now What?</a></b>
156	<a href="#">How to Name Your Plasmid in 3 Easy Steps</a>
158	<a href="#">How to Verify Your Plasmid</a>
161	<a href="#">Colony PCR</a>
165	<a href="#">6 Tips for Analyzing and Troubleshooting Sanger Sequencing Results</a>
167	<a href="#">Tips for Using BLAST to Verify Plasmids</a>
174	<a href="#">Optimizing Plasmid Yields</a>
177	<a href="#">Control Plasmids</a>

## INTRODUCTION TO GENOME ENGINEERING (CONT'D)

---

Page	Section
182	<a href="#">Chapter 9: Depositing Your Plasmids with Addgene</a>
183	<a href="#">A Brief History of Addgene</a>
184	<a href="#">Benefits of Depositing</a>
185	<a href="#">The Deposit Spreadsheet</a>
189	<a href="#">Acknowledgements and Final Words</a>

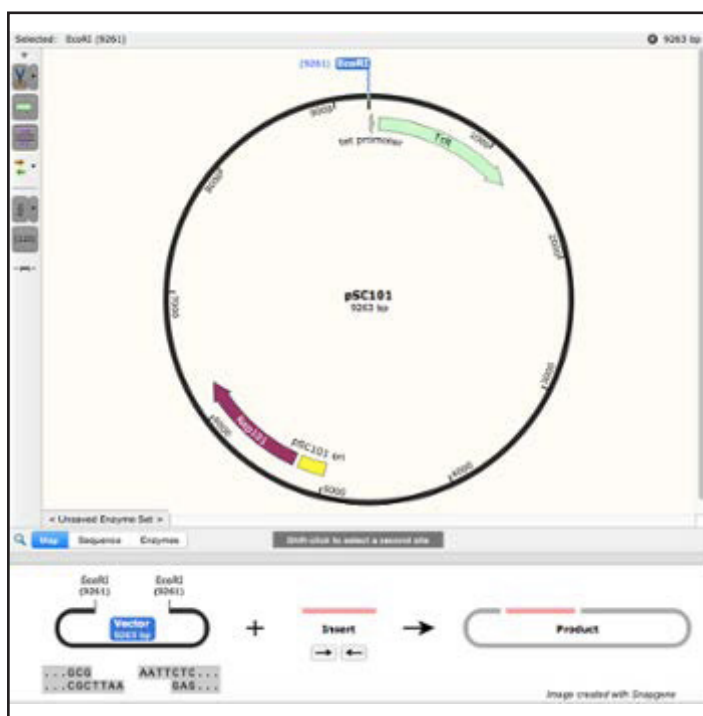
# CHAPTER 1: WHAT IS A PLASMID?



# A BRIEF HISTORY OF PLASMIDS

By Marcy Patrick, Addgene | October, 2015

Bioblasts? Plasmagenes? In the 1940s and 50s, scientists were working to understand genetic cytoplasmic factors that could be transferred between cells. At the time, these extranuclear agents of heredity were thought of as everything from parasites, to symbionts, to genes and the labels applied to them were vague or contradictory, owing in part to the fact that very little was known about the role these factors played within an organism.



Example plasmid.

## So How Did Plasmids Get Their Name?

In 1952, Joshua Lederberg set out to clarify the classification of these cytoplasmic inheritance factors. He proposed the catch-all term “plasmid” derived as a hybrid of “cytoplasm” and “id” (Latin for ‘it’), as “a generic term for any extrachromosomal hereditary determinant”. His proposal, however, was basically ignored. A separate term, “episome”, defined as “a non-essential genetic element which could exist either autonomously or integrated into the chromosome” was proposed a few years later by Élie Jacob and François Wollman and became the widely adopted name for these elements. At the time, the use of episome seemed appropriate, especially since the Fertility, or F-factor discovered by Ester Lederberg in 1952 was noted to integrate into the *E. coli* chromosome in some cases. This terminology held until the 1960s when scientists began to study other extrachromosomal particles, particularly Resistance or R-factors. Like F-factors, R-factors could be transferred between bacteria via cell-to-cell contact; however, scientists noted that, unlike F-factors, the evidence did not support the idea that

R-factors could integrate into the chromosome. Thus the term “episome” was eventually dropped and we’ve been using “plasmid” ever since!

## From Napkins to Notebooks

Although discovered in the early 1950s, it took until the 1970s for plasmids to gain prominence in the scientific community. Prior to this, bacteriophage, especially lambda, was the tool of choice for molecular biologists wanting to study bacterial genetics. This all changed thanks, in part, to a collaboration initiated at a Hawaiian deli in 1972. Using a deli napkin for paper, a small group of scientists including [Stanley Falkow](#), Stanley Cohen, Herbert Boyer, and Charles Brinton concocted a wild idea of using the newly discovered EcoRI enzyme (and its predictable cut site) to develop the first plasmid “cloning” experiment. Dr. Cohen and colleagues treated a tetracycline resistant plasmid, pSC101, and a newly developed kanamycin resistant plasmid, pSC102, with EcoRI and selected for *E. coli* transformants that were resistant to both. When this proved successful, pSC101 became the first plasmid cloning vector and molecular biology was never the same.



## A BRIEF HISTORY OF PLASMIDS (CONT'D)

---

Over the next few years genes from different bacterial (and eventually mammalian) species were cloned into plasmids and new cloning vectors such as pBR322, pACYC, and [pUC](#) were developed to provide higher copy number vectors that could be used in these cloning experiments.

Although plasmids started as a somewhat niche area of research, they are now seen as an ubiquitous tool that can be diversely applied to many different experiments. Addgene was founded in order to store, QC, curate, and distribute them all in the name of making it a little bit easier for scientists to conduct their research! Since their discovery in the 1950s, plasmids have impacted many areas of molecular biology and have been key in advancing our knowledge in areas such as bacterial conjugation and recombination, replication and topology, and cloning and gene expression.

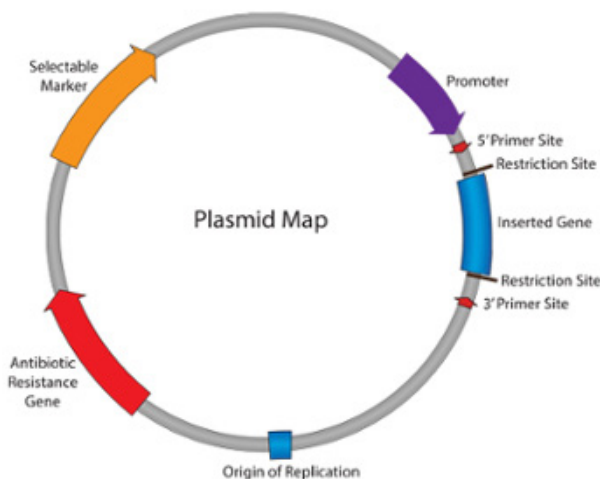
### Further Reading

1. [CSHL Meeting: Plasmids: History & Biology](#)
2. [The Joshua Lederberg Papers](#)
3. [Joshua Lederberg's Personal Perspective: Plasmid \(1952-1997\)](#)
4. DNA Cloning: A Personal View after 40 Years. Cohen, Stanley N. Proceedings of the National Academy of Sciences of the United States of America. 110.39 (2013): 15521–15529 [PubMed PMID: 24043817](#).
5. Cell Genetics and Hereditary Symbiosis. Lederberg, Joshua. Physiological Reviews 32.4 (1952) 403-430. [Link](#).
6. Sex Compatibility in Escherichia Coli. Lederberg, Joshua, Luigi L. Cavalli, and Esther M. Lederberg. Genetics 37.6 (1952): 720–730. [PubMed PMID: 17247418](#).
7. EPISOME-MEDIATED TRANSFER OF DRUG RESISTANCE IN ENTEROBACTERIACEAE VIII.: Six-Drug Resistance R Factor. Watanabe, Tsutomu, Chizuko Ogata, and Sachiko Sato. Journal of Bacteriology 88.4 (1964): 922–928. [PubMed PMID: 14219055](#).
8. Transmissible Drug Resistance in an Epidemic Strain of Salmonella Typhimurium. Datta, Naomi. The Journal of Hygiene 60.3 (1962): 301–310. [PubMed PMID: 14025218](#).
9. Construction of Biologically Functional Bacterial Plasmids In Vitro. Cohen, Stanley N. et al. Proceedings of the National Academy of Sciences of the United States of America 70.11 (1973): 3240–3244. [PubMed PMID: 1422013](#).
10. Uniform Nomenclature for Bacterial Plasmids: A Proposal. Novick, R P et al. Bacteriological Reviews 40.1 (1976): 168–189. [PubMed PMID: 16350226](#).

# WHAT IS A PLASMID?

By Margo R. Monroe, Addgene with contributions from Marcy Patrick, Addgene | Jan 14, 2014

**At their most basic level, plasmids are small circular pieces of DNA that replicate independently from the host's chromosomal DNA.** They are mainly found in bacteria, but also exist naturally in archaea and eukaryotes such as yeast and plants. In nature, plasmids provide one or more functional benefits to the host such as resistance to antibiotics, degradative functions, and/or virulence. All natural plasmids contain an origin of replication (which controls the host range and copy number of the plasmid) and typically include a gene that is advantageous for survival, such as an [antibiotic resistance gene](#). In contrast, plasmids utilized in the lab are usually artificial and designed to introduce foreign DNA into another cell. Minimally, lab-created plasmids have an origin of replication, selection marker, and cloning site. The ease of modifying plasmids and the ability of plasmids to self-replicate within a cell make them attractive tools for the life scientist or bioengineer.



Vector Element	Description
Origin of Replication (ORI)	DNA sequence that allows initiation of replication within a plasmid by recruiting transcriptional machinery proteins.
Antibiotic Resistance Gene	Allows for selection of plasmid-containing bacteria.
Multiple Cloning Site (MCS)	Short segment of DNA which contains several restriction sites allowing for the easy insertion of DNA. In expression plasmids, the MCS is often downstream from a promoter.
Insert	Gene, promoter, or other DNA fragment cloned into the MCS for further study.
Promoter Region	Drives transcription of the target gene. Vital component for expression vectors: determines which cell types the gene is expressed in and the amount of recombinant protein produced.
Selectable Marker	The antibiotic resistance gene allows for selection in bacteria. However, many plasmids also have selectable markers for use in other cell types.
Primer Binding Site	A short single-stranded DNA sequence used as an initiation point for PCR amplification or sequencing. Primers can be exploited for sequence verification of plasmids.

## CHAPTER 1: WHAT IS A PLASMID? (CONT'D)

---

### How is a Plasmid Constructed in the Lab?

Due to their artificial nature, lab plasmids are commonly referred to as “vectors” or “constructs”. To insert a gene of interest into a vector, scientists may utilize one of a variety of [cloning methods](#) (restriction enzyme, ligation independent, Gateway, Gibson, and more). The cloning method is ultimately chosen based on the plasmid backbone you choose. Regardless, once the cloning steps are complete, the vector containing the newly inserted gene is transformed into bacterial cells and selectively grown on antibiotic plates.

Addgene has compiled various educational resources to facilitate plasmid use in the lab. See Chapter 2 of this eBook for an introduction to some of the more popular plasmid cloning techniques. In addition, [Addgene's Online Plasmid Guide](#) includes information about molecular cloning, how to choose a plasmid vector, molecular biology tools and references, and how to maintain your plasmid stocks. The guide also contains multiple protocols and troubleshooting tips to make plasmid usage as simple and straightforward as possible.

### How Do Scientists Use Plasmids?

Generally, scientists use plasmids to manipulate gene expression in target cells. Characteristics such as flexibility, versatility, safety, and cost-effectiveness enable molecular biologists to broadly utilize plasmids across a wide range of applications. Some common plasmid types include: cloning plasmids, expression plasmids, gene knock-down plasmids, reporter plasmids, [viral plasmids](#), and [genome engineering plasmids](#).

To date, scientists around the world are extensively using these vectors for experiments encompassing [fluorescent imaging](#), recombinant DNA technology, mass protein production, disease modeling, drug discovery, and genome editing (just to name a few).

### Where Can I Find Additional Resources for Using Plasmids?

In this chapter, we will cover the basics of various plasmid elements, including the antibiotic resistance gene, origin of replication, promoter, and more. We also have tables and charts for you to use as references at the lab bench and practical tips for your experiments.

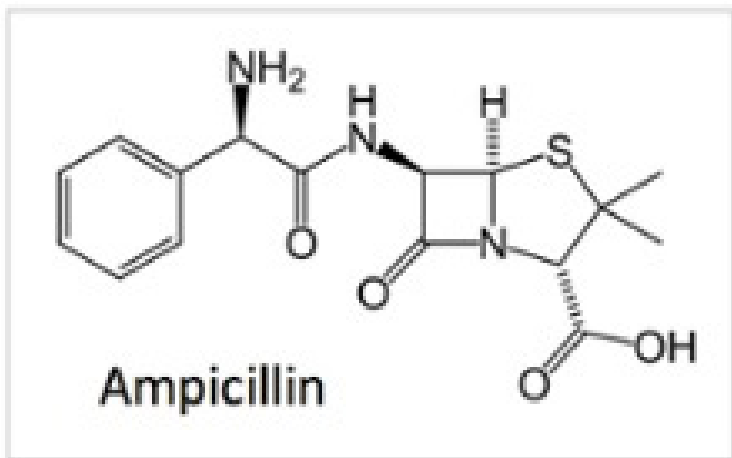
In addition to this eBook, Addgene has compiled more details on the history, importance, and types of plasmids in the [Addgene Molecular Biology Plasmid Reference Guide](#) on our website.

# ANTIBIOTIC RESISTANCE GENES

By Marcy Patrick, Addgene | Jan 30, 2014

**Antibiotic resistance genes are widely used tools in molecular biology, yet scientists rarely stop to think about how much easier they makes our lives.** Plasmid transformation into *E. coli* is a fairly inefficient process— just 1 out of 10,000 cells on average! Without some means of quickly determining which cells successfully received the correct plasmid, scientists would spend hours to days trying find their correct clones.

Additionally, the presence of a plasmid is disadvantageous from the bacterium's perspective – a plasmid-containing cell must replicate the plasmid in addition to its own chromosomal DNA, costing additional resources to maintain the plasmid. Adding an antibiotic resistance gene to the plasmid solves both problems at once; it allows a scientist to easily detect plasmid-containing bacteria when the cells are grown on selective media, and provides those bacteria with a pressure to keep the plasmid. Viva la (bacterial) resistance!



## What are Antibiotics?

Antibiotics are generally defined as agents that kill bacteria, or inhibit their growth. Although originally sourced from natural products, many common antibiotics used in labs today are semi-synthetic or fully synthetic compounds. Antibiotics can be categorized based on whether they directly kill bacteria (bactericidal) or slow growth/prevent cell division (bacteriostatic); however, the distinction between the two categories may be a bit of a gray area as some bacteriostatic reagents can kill bacteria when used at high concentrations (and vice versa). Looking around the lab, you'll likely find many of the antibiotics listed in the table below. Note, in this chapter we'll focus primarily on antibiotics against Gram negative bacteria. In future chapters, we'll detail selection in non-bacterial cells such as yeast or mammalian cells.

## How Else Can Antibiotics Be Used in the Lab?

Historically, antibiotics have also been used to disrupt genes at the chromosomal level. Scientists introduce an antibiotic resistance cassette within the coding region of the gene they are trying to disrupt or delete, which both inactivates the gene and acts as a marker for the mutation. When designing these types of experiments it is best practice not to use the same resistance cassette for the mutation and for plasmid selection. Additionally, scientists can use the loss of resistance as a marker for successful cloning. In these instances, the cloning vector typically has two separate resistance cassettes and the gene of interest is cloned into/inactivates or completely removes (in the case of [Gateway cloning](#)) one cassette. Counter selection allows the scientist to select bacteria that are only resistant to the antibiotic that remains intact.

## ANTIBIOTIC RESISTANCE GENES (CONT'D)

---

### Tips and Tricks from the Bench:

- Use fresh stocks. Most antibiotics are stable in powder form, but quickly break down in solution. Storing aliquots at -20°C and avoiding repeated freeze/thaw cycles will keep most antibiotics viable for at least 6 months.
- Ampicillin breaks down especially fast and plates should be used within 1 month for optimal efficiency. Beware of satellite colonies!
- Carbenicillin is more stable than Ampicillin and can be used in place of Ampicillin in most applications.
- Antibiotics vary in their sensitivity to heat and/or light – do not add them to media hotter than about 55°C and store plates/stocks wrapped in foil if a light-sensitive antibiotic like Tetracycline is used.
- Keep in mind that some *E. coli* strains have natural antibiotic resistances, so make sure your plasmid and *E. coli* strain are compatible! Check out this [list of common \*E. coli\* genotypes](#) and their natural resistances online from [OpenWetWare](#).

# COMMON ANTIBIOTICS

The table below lists antibiotics commonly found in the lab, their mechanism for killing bacteria, and their general working concentrations. For instructions on how to prepare antibiotic stocks, see [Addgene's Online Reference Page](#).

Name	Class	Mode of Action*	Bactericidal or Bacteriostatic	Working Concentration**
<a href="#">Ampicillin</a>	beta-lactam	Inhibits cell wall synthesis	Bactericidal	100-200 µg/mL
<a href="#">Bleomycin</a>	glycopeptide	Induces DNA breaks	Bactericidal	5-100 µg/mL
<a href="#">Carbenicillin</a>	beta-lactam	Inhibits cell wall synthesis	Bactericidal	100 µg/mL
<a href="#">Chloramphenicol</a>	N/A	Binds 50S ribosomal subunit; inhibits peptidyl translocation	Bacteriostatic	5-25 µg/mL in EtOH
<a href="#">Erythromycin</a>	macrolide	Blocks 50S ribosomal subunit; inhibits aminoacyl translocation	Bacteriostatic	50-100 µg/mL in EtOH
<a href="#">Kanamycin</a>	aminoglycoside	Binds 30S ribosomal subunit; causes mis-translation	Bactericidal	50-100 µg/mL
<a href="#">Polymyxin B</a>	polypeptide	Alters outer membrane permeability	Bactericidal	10-100 µg/mL
<a href="#">Spectinomycin</a>	aminoglycoside	Binds 30S ribosomal subunit; interrupts protein synthesis	Bactericidal	7.5-50 µg/mL
<a href="#">Streptomycin</a>	aminoglycoside	Inhibits initiation of protein synthesis	Bactericidal	25-100 µg/mL
<a href="#">Tetracycline</a>	tetracyclin	Binds 30S ribosomal subunit; inhibits protein synthesis (elongation step)	Bacteriostatic	10 µg/mL

\*In prokaryotes. \*\*Dissolve in dH<sub>2</sub>O and sterile filter unless otherwise specified

# ORIGIN OF REPLICATION

By Kendall Morgan, Addgene with contributions from Marcy Patrick, Addgene | Feb 6, 2014

---

**Let's consider another basic element of any plasmid: the origin of replication or "replicon."** The replicon is comprised of the origin of replication (ORI) and all of its control elements. The ORI is the place where DNA replication begins, enabling a plasmid to reproduce itself as it must to survive within cells.

The replicons of plasmids are generally different from those used to replicate the host's chromosomal DNA, but they still rely on the host machinery to make additional copies. ORI sequences are generally high in As and Ts; A-T base pairs are held together with two hydrogen bonds instead of three as G-C pairs are. As a result, stretches of DNA that are rich in A-T pairs melt more readily at lower temperatures. When DNA melts, it gives the replication machinery room to come in and get busy making copies.

## So Many Origins, So Little Time

There are lots of ORIs out there so, for now, we've ignored those used in eukaryotic cells and viruses and focused only on those found in bacteria. Some common ones you might see include ColE1, pMB1 (which comes in a few slightly different but well known derivatives), pSC101, R6K, and 15A. Not all origins of replication are created equal. Some will produce many plasmid copies and others produce just a few copies depending on how they are regulated. Generally, control of replication is referred to as "relaxed" or "stringent" depending on whether the ORI is positively regulated by RNA or proteins, respectively. A plasmid's copy number has to do with the balance between positive and negative regulation and can be manipulated with mutations in the replicon. For example, the pMB1 ORI maintains about 20 copies per cell, while pUC – which differs by only two mutations – will produce as many as 700 copies per cell.

So, how do you choose? Addgene Senior Scientist Marcy Patrick says researchers can ask themselves a few questions to get started: Will the plasmid be used exclusively in *E. coli*? Gram negative bacteria in general? Both Gram negatives and Gram positives? Will you have only one plasmid type in your cells at a time? Do you want to make a lot of your plasmid? Is the gene toxic in high amounts? It is always good to keep in mind that plasmids with low to medium copy numbers can still express massive amounts of protein given the proper promoter (stay tuned!) and growth conditions.

## Choose Your Origin of Replication Wisely

The best choice of ORI depends on how many plasmid copies you want to maintain, which host or hosts you intend to use, and whether or not you need to consider your plasmid's compatibility with one or more other plasmids. Generally speaking, plasmids with the same ORIs are incompatible because they will compete for the same machinery, creating an unstable and unpredictable environment. As a rule, plasmids from the same group should not be co-transformed, so if you require two plasmids for an experiment, make sure they have "compatible" ORIs. See the table below for more details.

The table below highlights common cloning vectors, their copy number, ORI, and incompatibility groups. Note the A-C compatibility grouping is an arbitrary designation, and plasmids from the same incompatibility group should not be co-transformed.

## ORIGIN OF REPLICATION (CONT'D)

Common Vectors	Copy Number*	ORI	Incompatibility Group	Control
pUC	~500-700	pMB1 (derivative)	A	Relaxed
pBR322	~15-20	pMB1	A	Relaxed
pET	~15-20	pBR322	A	Relaxed
pGEX	~15-20	pBR322	A	Relaxed
pColE1	~15-20	ColE1	A	Relaxed
pR6K	~15-20	R6K**	B	Stringent
pACYC	~10	P15A	B	Relaxed
pSC101	~5	pSC101	C	Stringent
pBluescript	~300-500	ColE1 (derivative) and F1***	A	Relaxed
pGEM	~300-500	pUC and F1	A	Relaxed

\*Actual copy number varies. See below for additional considerations.

\*\*Requires *pir* gene for replication ([reference](#)).

\*\*\*F1 is a phage-derived ORI that allows for the replication and packaging of ssDNA into phage particles.

Plasmids with phage-derived ORIs are referred to as [phagemids](#).

## Other Factors that Affect Copy Number

Although the sequence and regulation of the ORI dramatically affect the copy number of a plasmid, other external factors contribute as well. These considerations are especially useful to keep in mind if you are planning to purify your plasmid DNA:

### **The insert:**

- Bacteria tend to maintain fewer copies of plasmids if they contain large inserts or genes that create a toxic product.

### **The E. coli strain:**

- Most *E. coli* strains can be used to propagate plasmids, but endA- *E. coli* are best for high yields of plasmids.

### **Growth conditions:**

- The amount of aeration, temperature, culture volume, antibiotic, and medium can all affect copy number. Some ORIs are temperature sensitive; other ORIs can be “tricked” into amplifying more copies with the addition of Chloramphenicol – make sure your growth conditions aren’t working against you!



## ORIGIN OF REPLICATION (CONT'D)

---

### ***The culture inoculum:***

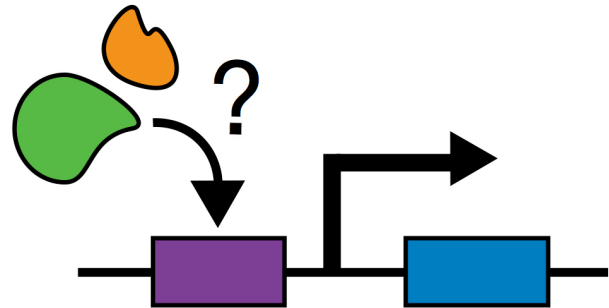
- Freshly streaked bacteria have higher copy numbers – for optimal results always pick a single colony and do not subculture directly from glycerol stocks, agar stabs, or liquid cultures. Incubation for 12-16 hours tends to give higher copy numbers since the bacteria have just reached stationary phase, but the cells have not started to die.

# THE PROMOTER REGION

By Kendall Morgan, Addgene with contributions from A. Max Juchheim, Addgene and Marcy Patrick, Addgene | Apr 3, 2014

Now that we can replicate our plasmid and make sure cells maintain it, the next step is getting the plasmid to express our gene of interest. Enter the promoter - the element responsible for initiating transcription of your insert.

In practice, the term “promoter” describes the combination of the promoter (RNA polymerase binding site) and operators (response elements). Promoters are about 100 to 1000 base pairs long and found upstream of their target genes. The sequence of the promoter region controls the binding of the RNA polymerase and transcription factors, therefore promoters play a large role in determining where and when your gene of interest will be expressed.



*Transcription factors binding to the promoter region (purple) of a gene (blue).*

## The RNA Polymerase(s)

RNA is transcribed from DNA using an RNA polymerase (RNAP). In bacteria, this is done by a single enzyme; however, eukaryotes have multiple polymerases which are each responsible for a specific subset of RNAs. To gain this specificity, the eukaryotic RNAP can recognize and bind to specific promoter elements. This means that the promoter present in your [plasmid backbone](#) must be compatible with the type of RNA that needs to be made: if you want mRNA (for gene expression) you need to use an RNAP II promoter, whereas small RNAs (such as shRNA) are transcribed from the RNAP III promoters. This section features promoters for general RNAP II and RNAP III transcription; however, viral LTRs such as RNAP II promoters are commonly employed in lentiviral and retroviral constructs and we will discuss these in the viral vectors chapter.

## Promoter Specificity

Aside from choosing a promoter based on the type of RNA transcript, you will also need to make sure your plasmid has a promoter suited to working in your host organism. Because transcription machinery differs between cell types or organisms, promoters must be similarly variable. Bacterial promoters only work in prokaryotic cells and typically only in the same or closely related species from which they were derived. Similarly, the various eukaryotic cell types (mammalian, yeast, plants, etc.) require unique promoters and there is very little crossover. Generally speaking, promoters in bacteria are less diverse and complex, having fewer parts than those in eukaryotic cells. Some promoters are constitutively active and on all the time, while others are more carefully controlled.

## THE PROMOTER REGION (CONT'D)

Regulated promoters might act only in certain tissues or at certain times in development or there may be ways to turn them on or off at will with a chemical, heat, or light. In the cell, promoters themselves are controlled by still other regulatory factors: enhancers, boundary elements, insulators, and silencers; however, some “leaky” transcription may occur. This is normally not a big issue for cells, but it may confound research results or even kill your cells if your gene of interest is toxic. To combat this, scientists have created synthetic promoters, which typically include some combination of other promoter elements, and tend to be more tightly regulated.

### Common Promoters for Eukaryotes and Prokaryotes

See the two reference tables below listing some of the most common bacterial and mammalian promoters. These lists are by no means exhaustive, but should be a good place to start when trying to pick your perfect promoter.

**Table 1: Eukaryotic Promoters\***



Promoter	Primarily Used For	RNA Transcript	Description	Expression	Additional Considerations
CMV	General expression	mRNA	Strong mammalian expression promoter from the human cytomegalovirus	Constitutive	May contain an enhancer region. Can be silenced in some cell types
EF1a	General expression	mRNA	Strong mammalian expression from human elongation factor 1 alpha	Constitutive	Tends to give consistent expression regardless of cell type or physiology
SV40	General expression	mRNA	Mammalian expression promoter from the simian vacuolating virus 40	Constitutive	May include an enhancer

## THE PROMOTER REGION - LET'S GO! (CONT'D)

**Table 1: Eukaryotic Promoters (cont'd)**

Promoter	Primarily Used For	RNA Transcript	Description	Expression	Additional Considerations
PGK1 (human or mouse)	General expression	mRNA	Mammalian promoter phosphoglycerate kinase gene.	Constitutive	Widespread expression, but may vary by cell type. Tends to resist promoter down regulation due to methylation or deacetylation.
Ubc	General expression	mRNA	Mammalian promoter from the human ubiquitin C gene	Constitutive	As the name implies, this promoter is ubiquitous.
human beta actin	General expression	mRNA	Mammalian promoter from beta actin gene	Constitutive	Ubiquitous. Chicken version is commonly used in promoter hybrids.
CAG	General expression	mRNA	Strong hybrid mammalian promoter	Constitutive	Contains CMV enhancer, chicken beta actin promoter, and rabbit beta-globin splice acceptor.
TRE	General expression	mRNA	Tetracycline response element promoter	Inducible with Tetracycline or its derivatives.	Typically contains a minimal promoter with low basal activity and several tetracycline operators. Transcription can be turned on or off depending on what tet transactivator is used.
UAS	General expression	mRNA	Drosophila promoter containing Gal4 binding sites	Specific	Requires the presence of Gal4 gene to activate promoter.
Ac5	General expression	mRNA	Strong insect promoter from Drosophila Actin 5c gene	Constitutive	Commonly used in expression systems for Drosophila.

## THE PROMOTER REGION - LET'S GO! (CONT'D)

**Table 1: Eukaryotic Promoters (cont'd)**

Promoter	Primarily Used For	RNA Transcript	Description	Expression	Additional Considerations
Polyhedrin	General expression	mRNA	Strong insect promoter from baculovirus	Constitutive	Commonly used in expression systems for insect cells.
CaMKIIa	Gene expression for optogenetics	mRNA	Ca <sup>2+</sup> /calmodulin-dependent protein kinase II promoter	Specific	Used for neuronal/CNS expression. Modulated by calcium and calmodulin.
GAL1, 10	General expression	mRNA	Adjacent, divergently transcribed promoters from yeast	Inducible with galactose; repressible with glucose	Can be used independently or together. Regulated by GAL4 and GAL 80.
TEF1	General expression	mRNA	Yeast transcription elongation factor promoter	Constitutive	Analogous to mammalian EF1a promoter.
GDS	General expression	mRNA	Strong yeast promoter from glyceraldehyde 3-phosphate dehydrogenase	Constitutive	Very strong, also called TDH3 or GAPDH.
ADH1	General expression	mRNA	Yeast promoter for alcohol dehydrogenase I	Repressed by ethanol	Full length version is strong with high expression. Truncated promoters are constitutive with lower expression.
CaMV35S	General expression	mRNA	Strong plant promoter from the Cauliflower Mosaic Virus	Constitutive	Active in dicots, less active in monocots, with some activity in animal cells.
Ubi	General expression	mRNA	Plant promoter from maize ubiquitin gene	Constitutive	Gives high expression in plants.

## THE PROMOTER REGION - LET'S GO! (CONT'D)

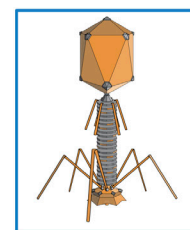
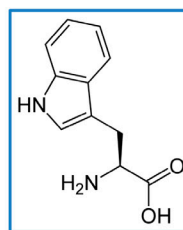
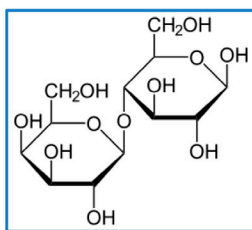
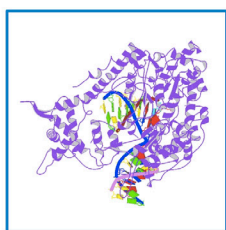
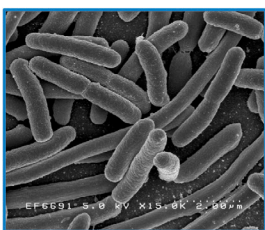
**Table 1: Eukaryotic Promoters (cont'd)**

Promoter	Primarily Used For	RNA Transcript	Description	Expression	Additional Considerations
H1	Small RNA expression	shRNA	From the human polymerase III RNA promoter	Constitutive	May have slightly lower expression than U6. May have better expression in neuronal cells.
U6	Small RNA expression	shRNA	From the human U6 small nuclear promoter	Constitutive	Murine U6 is also used, but may be less efficient.

*\*Image Attribution: Maize (second image from left) – [Wikipedia user Spedona](#); Drosophila (center) – [Wikipedia user Bbski](#); Mouse (second image from right) – [Wikipedia user Rama](#); HeLa cells (furthest right) - Fraunhofer Institute for Biomedical Engineering, St. Ingbert, Paul Anastasiadis, Eike White*

## THE PROMOTER REGION - LET'S GO! (CONT'D)

Table 2: Prokaryotic Promoters\*



Promoter	Primarily Used For	Description	Expression	Additional Considerations
T7	<i>In vitro</i> transcription/ general expression	Promoter from T7 bacteriophage	Constitutive, but requires T7 RNA polymerase.	Can be used for <i>in vitro</i> transcription only if 2 different phage promoters are present in opposite orientations to gene.
T7lac	High levels of gene expression	Promoter from T7 bacteriophage plus lac operators	Negligible basal expression when not induced. Requires T7 RNA polymerase, which is also controlled by lac operator. Can be induced by IPTG.	Commonly found in pET vectors. Very tightly regulated by the lac operators. Good for modulating gene expression through varied inducer concentrations.
Sp6	<i>In vitro</i> transcription/ general expression	Promoter from Sp6 bacteriophage	Constitutive, but requires SP6 RNA polymerase.	Can be used for <i>in vitro</i> transcription only if 2 different phage promoters are present in opposite orientations to gene.
trp	High levels of gene expression	Promoter from <i>E. coli</i> tryptophan operon	Repressible	Gets turned off with high levels of cellular tryptophan.
lac	General expression	Promoter from lac operon	Constitutive in the absence of lac repressor (lacI or lacIq). Can be induced by IPTG or lactose.	Leaky promoter with somewhat weak expression. lacIq mutation increases expression of the repressor 10x, thus tightening regulation of lac promoter. Good for modulating gene expression through varied inducer concentrations.

## THE PROMOTER REGION - LET'S GO! (CONT'D)

**Table 2: Prokaryotic Promoters (cont'd)**

Promoter	Primarily Used For	Description	Expression	Additional Considerations
Ptac	General expression	Hybrid promoter of lac and trp	Regulated like the lac promoter	Contains -35 region from trpB and -10 region from lac. Very tight regulation. Good for modulating gene expression through varied inducer concentrations. Generally better expression than lac alone.
pL	High levels of gene expression	Promoter from bacteriophage lambda	Can be temperature regulatable	Often paired with the temperature sensitive cI857 repressor.
araBAD	General expression	Promoter of the arabinose metabolic operon	Inducible by arabinose and repressed via catabolite repression in the presence of glucose or by competitive binding of the anti-inducer fucose	Weaker. Commonly found in pBAD vectors. Good for rapid regulation and low basal expression; however, not well-suited for modulating gene expression through varied inducer concentrations.

\*Image Attribution: T7 Polymerase (second image from left) - Temiakov, Dmitry, et al. "Structural basis for substrate selection by T7 RNA polymerase." *Cell* 116.3 (2004): 381-391; Bacteriophage (first image on right) - [Wikipedia user Adenosine](#)





# TERMINATORS AND POLYA SIGNALS (CONT'D)

terminator will suffice. Many commercial expression vectors use double terminators to reduce unwanted translation of downstream elements. A [high affinity terminator](#) may be desired for multi-cistronic constructs where high termination efficiency is necessary to minimize transcriptional read-through. Chris Voigt's lab has characterized a set of prokaryotic terminators and [deposited several with Addgene](#) (1).

## Prokaryotic Polyadenylation

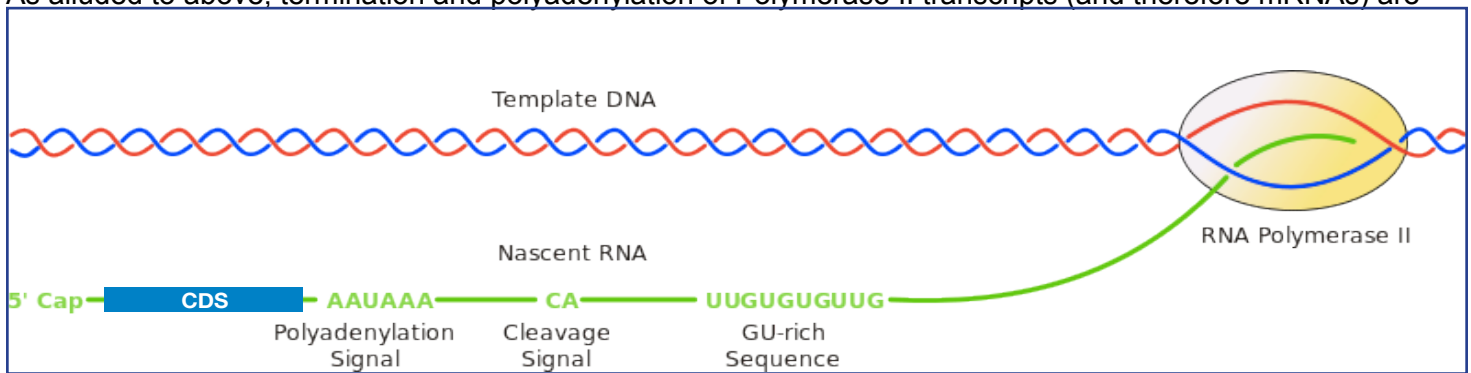
Although mostly thought of as a eukaryotic-specific process, prokaryotes also add poly(A) tails to certain RNAs. Unlike the eukaryotic mechanism which requires a consensus sequence for the addition of a poly(A) tail, the addition of a poly(A) tail on a prokaryotic transcript is non-specific and can be added to any accessible 3' end. The presence of the poly(A) tail targets the RNA to the degradosome, which contains enzymes that cut RNA not protected by secondary structure. Because it lacks specificity, it is thought that poly(A)s are used to control the cellular concentration of regulatory RNAs and may additionally act as a quality control mechanism to rid the cell of mis-folded RNAs.

## Eukaryotic Termination and Polyadenylation

Unlike prokaryotes that have a single RNA polymerase for transcription, eukaryotes have three RNA polymerases (Polymerases I, II, and III), each responsible for transcribing different types of RNA: Polymerase I is responsible for ribosomal RNA, Polymerase II is responsible for mRNA and miRNAs, and Polymerase III transcribes tRNA and other short RNAs. Although not as well studied as prokaryotic termination, the basic processes for eukaryotic termination are understood and it has been noted that each eukaryotic RNA polymerase terminates differently. Polymerase III, for example, relies on a specific sequence and RNA secondary structure to induce transcript cleavage, similar to the Rho-independent termination found in prokaryotes. This is different than Polymerases I and II, which both require binding of termination factors. Although both are termination factor dependent, Polymerases I and II employ different mechanisms to terminate transcription. Polymerase I uses a process similar to the prokaryotic Rho-dependent mechanism, whereas Polymerase II termination is more complex and involves two RNA polymerase-associated proteins, CPSF and CstF, which are responsible for recruiting the cleavage and polyadenylation enzymes, in a process that seems to couple termination with polyadenylation.

[Mammalian expression plasmids](#) are primarily used to create mRNA and the commonly used mammalian terminators (SV40, hGH, BGH, and rbGlob) include the sequence motif AAUAAA which promotes both polyadenylation and termination. Out of those listed, the SV40 late polyA and rbGlob polyA are thought to be more efficient in terminating transcription due to the presence of additional helper sequences (2-3).

As alluded to above, termination and polyadenylation of Polymerase II transcripts (and therefore mRNAs) are



*The conserved eukaryotic polyadenylation signal directs cleavage at the cleavage signal and addition of a poly-A tail to the mRNA transcript. By arunreginald at en.wikipedia, CC BY-SA 3.0. [https://en.wikipedia.org/wiki/Post-transcriptional\\_modification](https://en.wikipedia.org/wiki/Post-transcriptional_modification)*

# TERMINATORS AND POLYA SIGNALS (CONT'D)

coordinated processes. Cleavage between the consensus motif and a downstream GU-rich region (shown in the figure below) releases the mRNA from the polymerase and creates a free 3' end which is now available for polyadenylation. The addition of the poly(A) tail is important for stability of the mRNA, protection from degradation, and is integral to the nuclear export and translation processes as well.

## Further Reading

1. Peters JM, Vangeloff AD, Landick R. Bacterial Transcription Terminators: The RNA 3'-End Chronicles. *Journal of molecular biology*. 2011. PubMed [PMID: 21439297](#). PubMed Central [PMCID: PMC3622210](#).
2. Schek N, Cooke C, Alwine JC. Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. *Molecular and Cellular Biology*. 1992. PubMed [PMID: 1333042](#). PubMed Central [PMCID: PMC360476](#).
3. Gil A, Proudfoot NJ. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell*. 1987. PubMed [PMID: 3568131](#).
4. Hager S, Frame FM, Collins AT, Burns JE, Maitland NJ. An Internal Polyadenylation Signal Substantially Increases Expression Levels of Lentivirus-Delivered Transgenes but Has the Potential to Reduce Viral Titer in a Promoter-Dependent Manner. *Hum Gene Ther*. 2008. PubMed [PMID: 18627247](#).
5. Zufferey R, Donello JE, Trono D, Hope TJ. Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element Enhances Expression of Transgenes Delivered by Retroviral Vectors. *Journal of Virology*. 1999. PubMed [PMID: 10074136](#). PubMed Central [PMCID: PMC104046](#).
6. Wodrich H, Schambach A, Kräusslich H-G. Multiple copies of the Mason–Pfizer monkey virus constitutive RNA transport element lead to enhanced HIV-1 Gag expression in a context-dependent manner. *Nucleic Acids Research*. 2000. PubMed [PMID:10648781](#). PubMed Central [PMCID: PMC102582](#).

# METHYLATION AND RESTRICTION ENZYMES

By March Patrick | June 30, 2016

---

Have you ever tried digesting with XbaI or ClaI restriction enzymes and gotten unusual or unexpected results? Or considered why DpnI will degrade your template DNA from a [PCR reaction](#) but not the newly synthesized product from a site-directed mutagenesis experiment? The answer to both questions is the same--methylation! Read on to learn about how DNA methylation may affect your restriction digests.

## Why Methylate?

Restriction endonucleases are so ubiquitous in the lab that it is easy to forget that these enzymes naturally occur in bacteria for purposes other than [cloning](#) or [confirming plasmids](#). It turns out that restriction enzymes are one half of naturally occurring restriction modification systems that prokaryotes use to protect themselves from foreign DNA. The other component of these systems, methyltransferases, methylate DNA at particular sequences to prevent them from being degraded by restriction endonucleases. A given prokaryote typically has genes encoding one or a few restriction modification systems containing methyltransferases that add methyl groups to specific DNA sequences and companion endonucleases that recognize and cleave the same DNA sequence if not methylated (if it comes from a phage for instance).

Aside from restriction modification systems, DNA methylation also plays an integral role in regulating genome replication, repairing mismatched basepairs or small indels that occur during DNA synthesis, and promoting or repressing protein expression. Methylases involved with these processes (for example Dam and Dcm methylases) are independent from the restriction modification systems, yet can still affect whether certain restriction enzymes can effectively cleave DNA.

Common lab *E. coli* K12 strains such as DH5alpha contain 3 methylases that recognize and methylate different stretches of DNA:

- Dam methylase adds a methyl group to the adenine of GATC stretches of DNA
- Dcm methylase adds a methyl group to the second cytosine of CCWGG
- EcoKI methylase adds a methyl group to the adenine in AACNNNNNNGTGC or GCACNNNNNNGTT

## Impacts on Cloning and Digests

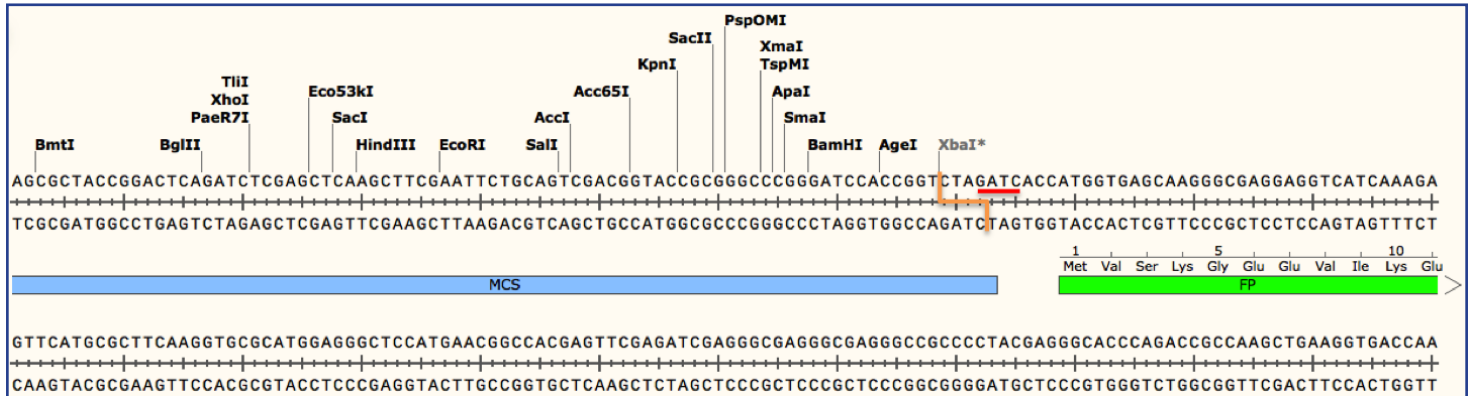
Although not all prokaryotic DNA is methylated to the same level, the potential for methylation should be considered when digesting DNA. Why? Well, even though Dam methylation sites are not specifically associated with any restriction modification systems, their sequences may overlap with restriction sites, inhibiting enzymes such as ClaI or XbaI, or, conversely, activating enzymes such as DpnI.

For example, cleavage by XbaI may be blocked due to methylation if the enzyme's recognition site (TCTAGA) is preceded by GA or followed by TC. As shown in the figure below, a Dam methylase recognition site (underlined in red) overlaps with the XbaI cut site (depicted as an orange line) because the restriction site is followed by TC. Addition of the methyl group at this site would block the enzyme from cutting the DNA there, although other restriction sites (including other XbaI sites not preceded by GA or followed by TC) would be cut as normal. If not taken into account, this blocked site could potentially give you results that would either be difficult to interpret or lead you to conclude your plasmid is not correct!

Conversely, enzymes such as DpnI require methylation at their recognition sites in order to efficiently cleave DNA. DpnI is often used for site directed mutagenesis. During this process, incorporation of a desired mutation into your plasmid of interest by PCR generates mutated plasmids with no methylation (there are no

## METHYLATION AND RESTRICTION ENZYMES (CONT'D)

methyltransferases in the PCR reaction). The template plasmid, on the other hand, should be derived from a *dam*<sup>+</sup> *E. coli* strain and will therefore have methylated adenines in any GATC sequences found in the plasmid. When the PCR products are digested with DpnI, only the non-mutated and methylated template is destroyed leaving behind a pool of mutated plasmids which can later be verified by Sanger sequencing.



### How Can I Tell if My Enzyme Will Cut?

Whether performing a digest for cloning purposes or for diagnostics, we suggest double checking to make sure your results will not be affected by methylation. Conveniently, the majority of restriction enzymes commonly used in the lab do not have recognition sites that could overlap with a methylation site. The quick-reference table below lists 10 common enzymes that may be affected by methylation. This table is by no means exhaustive, so you may want to consult the [REBASE](#) database for more detailed information.

Enzyme	Dam methylation	Dcm methylation	EcoKI methylation
Apal	Not affected	Blocked by overlapping methylation	Not affected
Bsal	Not affected	Not affected	Not affected
ClaI	Blocked by overlapping methylation	Mutation in a DNA-dependent ATPase that is essential for recombination and general DNA repair	Reduces plasmid recombination, increases plasmid stability
DraI	Not affected	Not affected	Blocked by overlapping methylation
HpaI	Not affected	Not affected	Blocked by overlapping methylation
MboI	Blocked by overlapping methylation	Not affected	Not affected
MscI	Not affected	Blocked by overlapping methylation	Not affected
PmeI	Not affected	Not affected	Blocked by overlapping methylation
XbaI	Blocked by overlapping methylation	Not affected	Not affected
DpnI	Requires methylation for activity	Not affected	Not affected

# METHYLATION AND RESTRICTION ENZYMES (CONT'D)

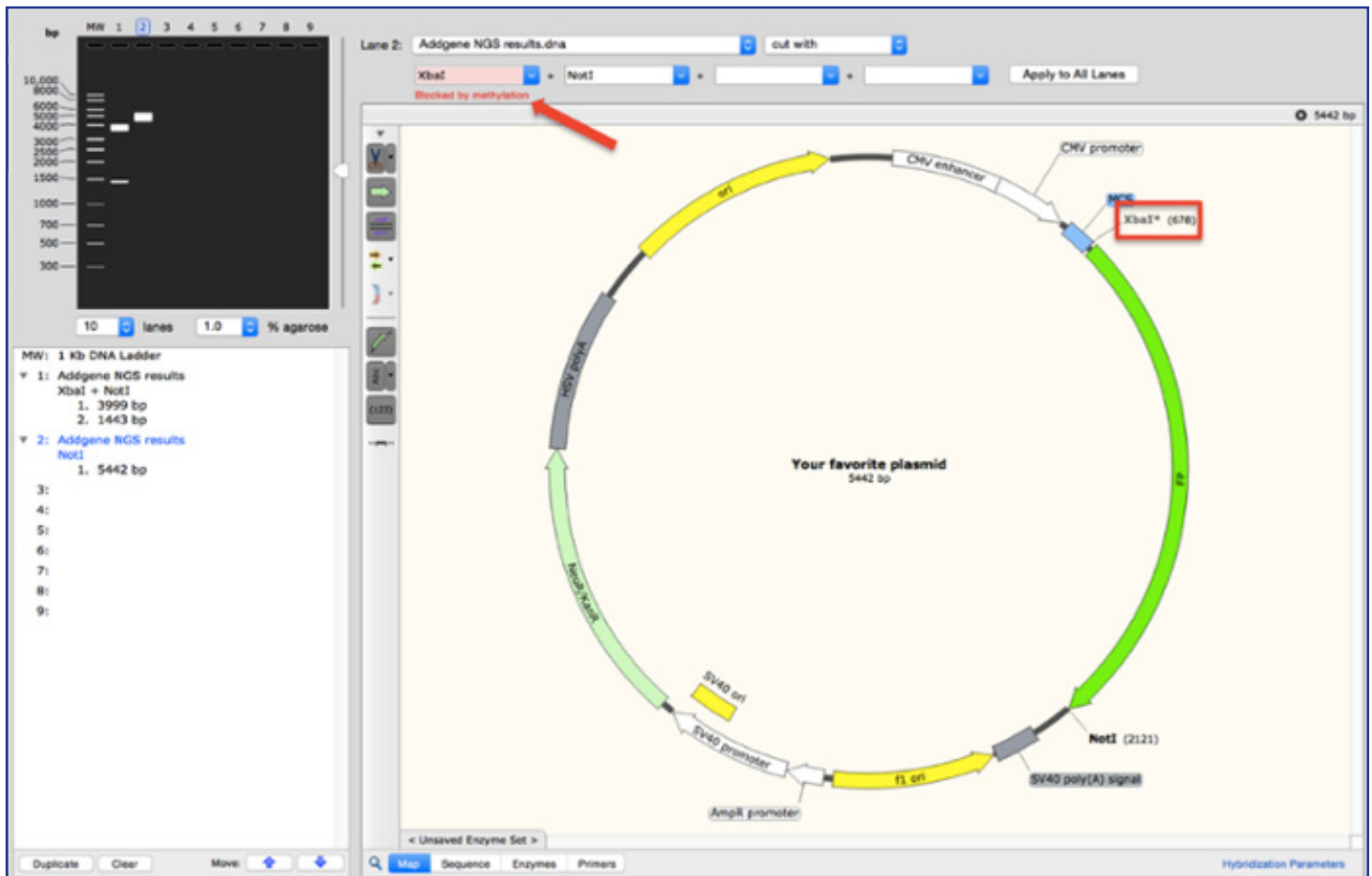


Image created using SnapGene(R)

## Controlling Methylation

Finally, you can control methylation by altering your choice of bacteria. For example, if you must use a restriction site that will be blocked by Dam or Dcm methylation, you can ensure this site remains unblocked by first cloning your DNA into a *dam*<sup>-</sup>/*dcm*<sup>-</sup> strain of *E. coli* such as [JM110](#) and re-purifying it. These specialized *E. coli* strains have been specifically engineered to be Dam and Dcm methylase-deficient, and, as such, produce DNA that is unmethylated at those sites. Please keep in mind that *dam*<sup>-</sup>/*dcm*<sup>-</sup> strains may have an increased rate of mutation (as these would also be deficient in mis-match repair functions of Dam), so these strains should not be used for long term storage.

## Further Reading

- Review of [Restriction Modification Systems](#)
- [NEB's Resource Describing Dam and Dcm Methylases](#)
- [Paper on the Discovery of DpnI](#)
- [Review describing \*dam\*<sup>-</sup>/\*dcm\*<sup>-</sup> \*E. coli\*](#)

# BLUE-WHITE SCREENING

By Jessica Welch, Addgene | Jun 4, 2015

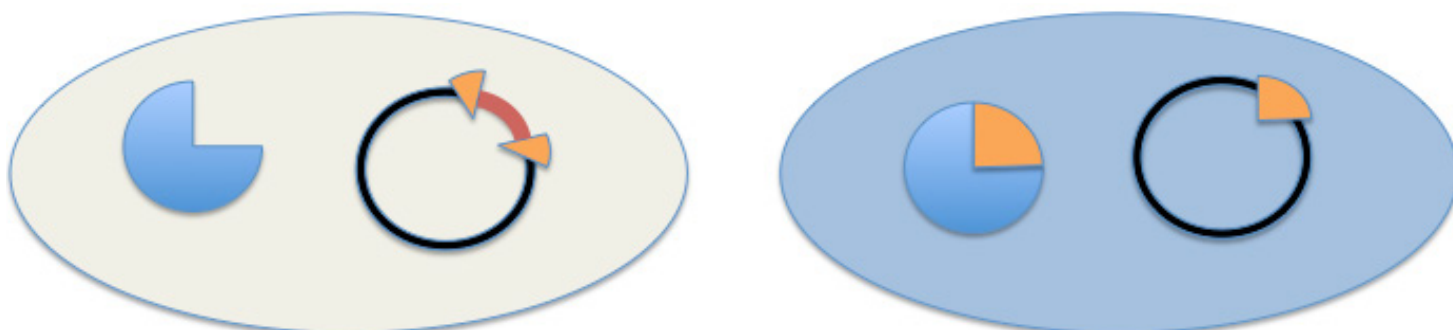
Now that we have covered [antibiotic selection](#), we can talk about an even more specific method of screening your cloning reaction. Being able to select for colonies that contain your plasmid is a great start when cloning, but how about being able to choose those that contain a plasmid with an insert? Blue-white selection is a widely used method to do just that!

Let's begin at the beginning. The well-characterized bacterial *lac* operon contains a gene called *lacZ* that encodes for the enzyme  $\beta$ -galactosidase. Expression of the *lac* operon is induced by lactose, and also by a lactose analogue, IPTG (isopropyl  $\beta$ -D-1-thiogalactopyranoside). To be completely accurate, IPTG binds and inactivates the *lac* operon repressor, thereby allowing *lac* expression.

When expressed, the  $\beta$ -galactosidase enzyme can break down a dye-linked substrate called x-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galacto-pyranoside) into galactose and an insoluble blue pigment (4-chloro-3-brom-indigo). So far, so good, but how does this help us screen plasmids?

## Blue-White Screening in the Lab

Scientists discovered that deleting a section from the *lacZ* gene (a mutation called *lacZ* $\Delta$ M15) creates a non-functional  $\beta$ -galactosidase enzyme. Providing DNA encoding this section of amino acids (called the  $\alpha$ -peptide) to a *lacZ* $\Delta$ M15-mutant bacterial cell *in trans* complements the mutation allowing for a functional enzyme. This process is called  $\alpha$ -complementation.



*Blue-White screening. A) A white cell in which cloning was successful and the  $\beta$ -galactosidase gene (*lacZ*) was disrupted (split wedge on black plasmid) preventing the cell from turning blue. B) Blue cell in which cloning was unsuccessful and the  $\beta$ -galactosidase gene is retained in the plasmid (un-split wedge) turning the cell blue. The pie charts in each cell show whether or not the genomic mutation in *lacZ* is complemented (B) creating a full pie, or not (A) leaving a wedge out of the pie.*



Image source: Wikimedia Commons; Accessed 6/3/15. Author: Stefan Walkowski

The system described above was put to practical use in the following way. Scientists engineered a multiple cloning site (MCS) into the  $\alpha$ -peptide (represented as an orange wedge in the figure above) and inserted it into a plasmid, creating an  $\alpha$ -complementation cloning vector. When a cloning reaction goes to plan and your DNA (red) is cloned into this MCS, the  $\alpha$ -peptide gets interrupted as shown in (A), and thus will not complement the a  $\beta$ -galactosidase mutation in the host cell. An unsuccessful cloning reaction leaves the  $\alpha$ -peptide intact, and therefore the cell will have a functional  $\beta$ -galactosidase enzyme through  $\alpha$ -complementation (cell B).

## BLUE-WHITE SCREENING (CONT'D)

---

As shown on the representative plate on the previous page, colonies with an insert-containing plasmid have a non-functional  $\beta$ -galactosidase and remain the lovely whitish-cream color of standard *E. coli*. On the other hand, intact  $\beta$ -galactosidase produces pigment from x-gal (included in the transformation plate medium), turning the bacterial colony blue. We should note that IPTG is also included in the medium to ensure transcription of the *lac* operon.

### Tips for Blue-White Screening

- **Use a good control:** Transform the backbone plasmid without insert. All colonies on this plate should be blue, indicating that your IPTG and x-gal are working as they should be.
- **Don't rush the process:** It is important to give your plates enough time for any intact  $\beta$ -galactosidase to be expressed and process x-gal into blue pigment (16-20 hours). A plate with only white colonies is very suspicious!
- **Refrigerate your plates:** Placing plates at 4° C for a few hours after the initial overnight incubation increases pigment precipitation, enhances the blue color of negative colonies, and allows for better differentiation between blue and white colonies.
- **Take care in making your plates:** X-gal is light and temperature sensitive and needs to be added to media after autoclaving. If spread on top of pre-made plates, make sure it is evenly distributed and allow sufficient drying time before use.
- **Beware of false positives:** Blue-white screening only indicates the presence of AN insert, not necessarily YOUR insert. Any cloning artifact that disrupts the  $\alpha$ -peptide DNA will also lead to a white colony.
- **And also false negatives:** These are rare, but if a small fragment is inserted in-frame, read-through can lead to a functional  $\beta$ -galactosidase enzyme and a blue colony. Blue-white screening is a good way to narrow down candidates for more specific analysis, like PCR or restriction digest.
- **Make sure you use a proper [E. coli strain](#)** (i.e. contains the lacZ $\Delta$ M15 mutation): XL1-Blue, DH5 $\alpha$ , DH10B, [JM109](#), STBL4, [JM110](#), and Top10 are a few examples.
- **Make sure you use a proper plasmid** (i.e. contains the  $\alpha$ -complementation cloning MCS): pGEM-T, [pUC18](#) and [pUC19](#), and pBluescript are a few common vectors.

Blue-white screening is just that – a screening process. It does not select only those cells that have taken up a plasmid and thus should be used in conjunction with selection methods. Combining selection and screening ensures that the white colonies you see are white due to successful cloning and not because the cell failed to take up the  $\alpha$ -complementation plasmid. This way you can quickly and easily identify colonies that not only have your plasmid (antibiotic resistance), but also confirm those plasmids have your insert (blue-white screening).



## BLUE-WHITE SCREENING (CONT'D)

---

### Other Types of Screening Methods

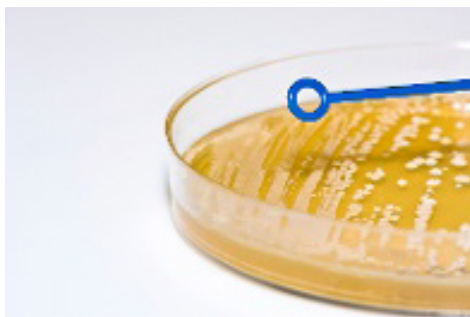
Although blue-white screening is probably the most widespread way to select for plasmids containing an insert, there are other methods. Positive selection vectors encode a gene which, when expressed, is lethal to the cell. Cloning fragments are inserted into an MCS in the center of this gene, disrupting the lethality. This is similar to  $\alpha$ -peptide DNA disruption in the blue-white screen. Antibiotic selection is also used in conjunction with this method to ensure that positive colonies do have the plasmid containing the lethal gene. An example is the [ccdB](#) gene system.

There are also methods to select for plasmid-containing cells without using antibiotic resistance. These rely on cell lines susceptible to or dependent on certain media components, and are rescued by genes supplied on the transformed plasmid. Such plasmids may contain genes that allow for the use of a particular substrate in defined medium, without which the cells cannot reproduce to form colonies (used to complement auxotrophic cell lines), or genes which rescue a lethal phenotype.

For more resources, browse Addgene's curated list of [Bacterial Expression Systems](#).

# COMMON LAB *E. COLI* STRAINS

By Matthew Ferenc, Addgene | Nov 7, 2014



You've worked hard designing your plasmid – you carefully selected the optimal promoter for your gene of interest, you painstakingly cloned into the perfect empty backbone, you made sure to add the right tags and a nuclear localization signature (NLS) to your gene, you put a fluorescent protein downstream, separated by an IRES element. You did a lot of work! But let's take a moment to recognize the prokaryotic minions that carried out the labor-intensive process of replicating your new plasmid: the *Escherichia coli* bacterium.

It's hard to count the number of different commercial strains of *E. coli* currently available – a quick Google search suggests there are hundreds.

This only includes general lab strains designed for subcloning or protein expression. If you were to include customized strains, the number is probably in the thousands! The goal of this section is to provide enough background for you to distinguish the features of any common lab strain and determine whether it is appropriate for propagating your plasmid or carrying out your experiment.

## History of *E. coli* Strains

*E. coli* are gram-negative, rod shaped bacteria that were named after Dr. Theodor Escherich, the scientist who first described them in 1885. *E. coli* are mainly found in the intestinal tract of animals. There are many different naturally occurring strains of *E. coli*, some of which are deadly to humans. The majority of all common, commercial lab strains of *E. coli* used today are descended from two individual isolates, the K-12 strain and the B strain. K-12 was isolated from a patient in 1920 and eventually led to the common lab strain MG1655, which led to DH5alpha and DH10b (also known as TOP10). The history of the B strain is a bit more convoluted due to researchers sharing and renaming samples throughout history. It was likely isolated in 1918 but was first referred to as "B strain" in 1942. The BL21 strain, (and derivatives) are the most common examples of the *E. coli* B strain.

## Common *E. coli* Strains Used in the Lab

Most of the commercial strains you find today are marketed for a specific purpose: fast growth, high-throughput cloning, routine cloning, cloning unstable DNA, preparing unmethylated DNA, and more. Many mutations that make these features possible are present in most commercial strains, especially mutations that make major improvements such as those that increase plasmid yield and/or DNA quality. Table 1 below outlines a few of the more common genetic changes found in *E. coli* strains and Table 2 outlines common lab strains of *E. coli*.

## COMMON LAB *E. COLI* STRAINS (CONT'D)

### Table 1: Common Gene Mutations Found in *E. coli* Strains

Genes	Description	Functional Consequence
<i>dam</i>	DNA Adenine methylase mutation (GATC)	Preparing unmethylated DNA, important when trying to cut with certain restriction enzymes (ex: ClaI or XbaI)
<i>dcm</i>	DNA Cytosine methylase mutation (CCWGG)	Preparing unmethylated DNA, important when trying to cut with certain restriction enzymes that are <a href="#">methylation sensitive</a> .
<i>dnaJ</i>	Mutation in a chaperonin gene	Increases the stability of certain expressed proteins
<i>endA, endA1</i>	Endonuclease I (nonspecific cleavage of dsDNA ) mutation	Improves plasmid yield
<i>F</i>	Host does (F <sup>+</sup> ) or does not (F <sup>-</sup> ) contain the fertility plasmid.	A low copy-number plasmid, encodes proteins needed for bacterial conjugation. Genes listed on F <sup>+</sup> are wild-type unless indicated otherwise
<i>fhuA (formerly tonA)</i>	ferric hydroxamate uptake, iron uptake receptor mutation.	T1/T5 Phage resistance
<i>gal</i>	Mutation in galactose metabolism pathway	cells cannot grow on just galactose
<i>gyrA, gyrA96</i>	DNA gyrase mutation	Confers resistance to nalidixic acid
<i>hsdRMS</i>	<i>hsdR(rk-, mk+)</i>	unmethylated DNA not degraded, cell still can methylate DNA
	<i>hsdS(rk-, mk-)</i>	unmethylated DNA not degraded, cell cannot methylate DNA
<i>lac</i>	Lac Operon Mutations	<a href="#">blue/white screening</a> of clones
	<i>lacIq</i>	lac repressor overproduced, expression from pLac repressed more
	<i>lacZ</i>	β-galactosidase activity abolished
	<i>lacY</i>	Lactose permease inactivated, lactose cannot be taken up by cell

## COMMON LAB *E. COLI* STRAINS (CONT'D)

**Table 1: Common Gene Mutations Found in *E. coli* Strains (cont.)**

Genes	Description	Functional Consequence
<i>mrcA, mcrBC</i>	Inactivation of pathway that cleaves methylated cytosine DNA	Allows for uptake of foreign (methylated) DNA
<i>mrr, Δ(mcrC-mrr)</i>	Inactivation of pathway that cleaves methylated adenine or cytosine DNA	Allows for uptake of foreign (methylated) DNA
<i>recA, recA1, recA13</i>	Mutation in a DNA-dependent ATPase that is essential for recombination and general DNA repair	Reduces plasmid recombination, increases plasmid stability
<i>recBCD</i>	Exonuclease V activity abolished	Increased plasmid stability, reduced recombination
<i>relA or relA1</i>	Relaxed phenotype, mutation eliminating stringent factor	Allows RNA synthesis in absence of protein synthesis
<i>ptrc-ccdA</i>		Propagation of <i>ccdB</i> -containing plasmids
<i>hte</i>		high transformation efficiency
<i>deoR</i>	constitutive expression of genes for deoxyribose synthesis	Allows uptake of large plasmids
<i>hee</i>		high electroporation efficiency
<i>supE44 (glnV44)</i>		Suppression of the amber (UAG) stop codon by inserting glutamine
<i>supF (tyrT)</i>		Suppression of the amber (UAG) stop codon by inserting tyrosine
<i>λ-thi-1 or thi1</i>	Mutation in thiamine metabolism	requires exogenous thiamine for growth
<i>ara</i>	disruption of arabinose metabolism pathway	inability to utilize arabinose as a carbon source
<i>leuB</i>	β-isopropyl malate dehydrogenase inactivated	requires exogenous leucine source for growth
<i>proAB</i>	mutation in proline biosynthesis pathway	requires exogenous proline source for growth
<i>rpsL</i>	Mutation in subunit S12 of 30S ribosome	Confers resistance to streptomycin

## COMMON LAB *E. COLI* STRAINS (CONT'D)

**Table 2: Lab Strains of *E. coli***

Strain	Natural Resistance	Primary Use	Genotype
ccdB Survival 2	N/A	For propagating plasmids expressing the ccdB gene (important in <a href="#">Gateway cloning</a> ).	<i>F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (Str<sup>R</sup>) endA1 nupG fhuA::IS2</i>
DB3.1	Streptomycin	For propagating plasmids expressing the ccdB gene (important in Gateway cloning).	<i>F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>) ara14 galK2 lacY1 proA2 rpsL20(Sm<sup>r</sup>) xyl5 Δleu mtl1</i>
DH10B	Streptomycin	MC1061 derivative. General cloning and storage, blue/white screening, leucine auxotroph.	<i>F endA1 deoR<sup>+</sup> recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) Str<sup>R</sup> λ<sup>-</sup></i>
DH5α	N/A	General cloning and storage of common plasmids, blue/white screening.	<i>F endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 Φ80dlacZΔM15 Δ(lacZYA-argF) U169, hsdR17(r<sub>K</sub><sup>-</sup> m<sub>K</sub><sup>+</sup>), λ<sup>-</sup></i>
HB101	Streptomycin	Hybrid of <i>E. coli</i> K12 and <i>E. coli</i> B (mostly K12, though), common lab strain for cloning and storage of pBR322 plasmids.	<i>F mcrB mrr hsdS20(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Sm<sup>R</sup>) glnV44 λ<sup>-</sup></i>
JM109	N/A	General cloning and plasmid maintenance, blue/white screening, partly restriction-deficient; good strain for cloning repetitive DNA.	<i>endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB<sup>+</sup> Δ(lac-proAB) e14- [F' traD36 proAB<sup>+</sup> lacI<sup>q</sup> lacZΔM15] hsdR17(r<sub>K</sub> m<sub>K</sub><sup>+</sup>)</i>
JM110	Streptomycin	For storing plasmids that should not be methylated, allows for methylation sensitive restriction enzymes to cut the plasmid after preparation.	<i>rpsL thr leu thi lacY galK galT ara tonA tsx dam dcm glnV44 Δ(lac-proAB) e14- [F' traD36 proAB<sup>+</sup> lacI<sup>q</sup> lacZΔM15] hsdR17(r<sub>K</sub> m<sub>K</sub><sup>+</sup>)</i>
MC1061	Streptomycin	Parent of DH10B/TOP10 and derived strains, common lab cloning and storage strain.	<i>F Δ(ara-leu)7697 [araD139]<sub>B/r</sub> Δ(codB-lacI)3 galK16 galE15 λ<sup>-</sup> e14- mcrA0 relA1 rpsL150(Str<sup>R</sup>) spoT1 mcrB1 hsdR2(rm<sup>+</sup>)</i>
MG1655	N/A	"Wild type" K-12 strain.	<i>F λ<sup>-</sup> ilvG- rfb-50 rph-1</i>

## COMMON LAB *E. COLI* STRAINS (CONT'D)

**Table 2: Lab Strains of *E. coli* (cont'd)**

Strain	Natural Resistance	Primary Use	Genotype
Pir1	N/A	For cloning and maintenance of a plasmids with R6Ky origin; contains a mutant allele of the pir gene that maintains the plasmids at ~250 copies per cell.	<i>F- Δlac169 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMluI)::pir-116</i>
Stbl2	N/A	JM109-derived. For storage of plasmids that have the potential to recombine. Example, the LTRs in lenti- and retro-viral plasmids.	<i>F- endA1 glnV44 thi-1 recA1 gyrA96 relA1 Δ(lac-proAB) mcrA Δ(mcrBC-hsdRMS-mrr) λ<sup>-</sup></i>
Stbl3	Streptomycin	Derived from HB101. For storage of plasmids that have the potential to recombine. Example, the LTRs in lenti- and retro-viral plasmids, endA <sup>+</sup> , use care in preparing DNA from this strain.	<i>F- glnV44 recA13 mcrB mrr hsdS20(rB<sup>-</sup>, mB<sup>-</sup>) ara-14 galK2 lacY1 proA2 rpsL20 xyl-5 leu mtl-1</i>
Top10	Streptomycin	MC1061 derivative. General cloning and storage, blue/white screening.	<i>F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str<sup>R</sup>) endA1 λ<sup>-</sup></i>
XL1 Blue	Tetracycline	Blue/white screening and routine cloning, nalidixic acid resistant.	<i>endA1 gyrA96(nal<sup>R</sup>) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB<sup>+</sup> lacI<sup>q</sup> Δ(lacZ)M15] hsdR17(r<sub>K</sub><sup>-</sup> m<sub>K</sub><sup>+</sup>)</i>
XL10 Gold	Tetracycline and Chloramphenicol	Cloning and propagation of large plasmids, high competency, nalidixic acid resistant.	<i>endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet<sup>R</sup> F'[proAB lacI<sup>q</sup>ZΔM15 Tn10(Tet<sup>R</sup> Amy Cm<sup>R</sup>)]</i>

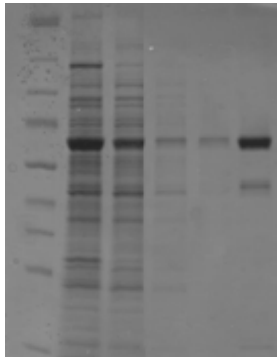
*Note: Generally inactivating mutations in specific genes are signified with a minus sign (-) as is typically standard; just having the gene listed indicates it is non-functional. If a gene is deleted that is normally noted with a Greek delta (Δ).*

# E. COLI STRAINS FOR PROTEIN EXPRESSION

By Julian Taylor-Parker, Addgene | Feb 10, 2015

We previously reviewed the salient features of several popular strains of [E. coli for DNA propagation](#). While great for cloning purposes, these *E. coli* strains are not usually well suited for recombinant protein expression. Many challenges can arise when over-expressing a foreign protein in *E. coli*. We will review the potential pitfalls of recombinant protein expression and some of the most popular commercial strains designed to avoid them.

## Why Do I Need an Expression Strain?



Coomassie stain showing protein expression and purification from *E. coli*

Protein expression from [high-copy number plasmids](#) and [powerful promoters](#) will greatly exceed that of any native host protein, using up valuable resources in the cell thus leading to slowed growth. Additionally, some protein products may be toxic to the host when expressed, particularly those that are insoluble, act on DNA, or are enzymatically active. For this reason, recombinant proteins are typically expressed in *E. coli* engineered to accommodate high protein loads using inducible promoter systems (which will be discussed later). In addition to the basic genotypes outlined below, certain specialized strains are available to confer greater transcriptional control, assist with proper protein folding, and deal with sub-optimal codon usage (Table 1).

A few mutations are common to all or most expression strains to accommodate high protein levels including:

- **ompT**: Strains harboring this mutation are deficient in outer membrane protease VII, which reduces proteolysis of the expressed recombinant proteins.
- **lon protease**: Strains where this is completely deleted (designated lon or  $\Delta$ lon) similarly reduce proteolysis of the expressed proteins.
- **hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>)**: These strains have an inactivated native restriction/methylation system. This means the strain can neither restrict nor methylate DNA.
- **dcm**: Similarly, strains with this mutation are unable to methylate cytosine within a particular sequence.

## E. COLI STRAINS FOR PROTEIN EXPRESSION (CONT'D)

### Table 1: *E. coli* Expression Strains

Note: All strains are derived from the *E. coli* B strain, except those marked with \*\* which are K12. \* indicates that strain component is a plasmid.

Strain	Resistance	Key Features	Genotype	Use
BL21(DE3)	N/A	Basic IPTG-inducible strain containing T7 RNAP (DE3)	<i>F- ompT lon hsdSB(rB-mB-) gal dcm (DE3)</i>	General protein expression
BL21(DE3) pLysS*	Chloramphenicol (pLysS)	pLysS expresses T7 lysozyme to reduce basal expression levels; expression vector cannot have p15A origin of replication	<i>F- ompT lon hsdSB(rB-mB-) gal dcm(DE3) pLysS (CamR)</i>	Expression of toxic proteins
BL21(DE3) pLysE*	Chloramphenicol (pLysE)	pLysE has higher T7 lysozyme expression than pLysS; expression vector cannot have p15A origin of replication	<i>F- ompT lon hsdSB (rB- mB-) gal dcm(DE3) pLysE (CamR)</i>	Expression of toxic proteins
BL21 star (DE3)	N/A	Lacks functional RNaseE which results in longer transcript half-life	<i>F- ompT lon hsdSB(rB-mB-) gal dcm rne131 (DE3)</i>	General expression; not recommended for toxic proteins
BL21-A1	Tetracycline	Arabinose-inducible expression of T7 RNAP; IPTG may still be required for expression	<i>F- ompT lon hsdSB(rB-mB-) gal dcm araB::T7RNAP-tetA [malB+]K-12(ΔS)</i>	General protein expression
BLR (DE3)	Tetracycline	RecA-deficient; best for plasmids with repetitive sequences.	<i>F- ompT lon hsdSB(rB-mB-) gal dcm(DE3) Δ(srl-recA)306::Tn10 (TetR)</i>	Expression of unstable proteins
HMS174 (DE3)**	Rifampicin	RecA-deficient; allows for cloning and expression in same strain	<i>F- recA1 hsdR(rK12-mK12+) (DE3) (Rif<sup>R</sup>)</i>	Expression of unstable proteins
Tuner (DE3)	N/A	Contains mutated lac permease which allows for linear control of expression	<i>F- ompT lon hsdSB(rB-mB-) gal dcm lacY1(DE3)</i>	Expression of toxic or insoluble proteins



## E. COLI STRAINS FOR PROTEIN EXPRESSION (CONT'D)

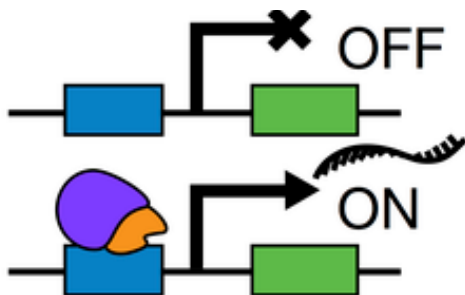
**Table 1: *E. coli* Expression Strains (cont'd)**

Note: All strains are derived from the *E. coli* B strain, except those marked with \*\* which are K12. \* indicates that strain component is a plasmid.

Strain	Resistance	Key Features	Genotype	Use
Rosetta2 (DE3)*	Chloramphenicol (pRARE)	Good for “universal” translation; contains 7 additional tRNAs for rare codons not normally used in <i>E. coli</i> . Expression vector cannot have p15A origin of replication	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)	Expression of eukaryotic proteins
Lemo21 (DE3)*	Chloramphenicol (pLemo)	Rhamnose-tunable T7 RNAP expression alleviates inclusion body formation. Expression vector cannot have p15A origin of replication	<i>fhuA2 [lon] ompT gal</i> ( $\lambda$ DE3) [ <i>dcm</i> ] $\Delta$ <i>hsdS/ pLemo</i> (CamR)	Expression of toxic, insoluble, or membrane proteins
T7 Express		IPTG-inducible expression of T7 RNAP from the genome; does not restrict methylated DNA	<i>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 <math>\Delta</math>(mcrC-mrr)114::IS10</i>	General protein expression
m15** pREP4*	Kanamycin (pREP4)	Cis-repression of the <i>E. coli</i> T5 promoter (found on vectors such as pQE or <a href="#">similar</a> ), inducible under IPTG (lac repressor on the pREP4 plasmid). Expression vector cannot have p15A origin of replication	F-, $\Phi$ 80 $\Delta$ <i>lacM15, thi, lac-, mtl-, recA+, KmR</i>	Expression of toxic proteins
Origami2 (DE3)**	Streptomycin and Tetracycline	Contains highly active thioredoxin reductase and glutathione reductase to facilitate proper folding; may increase multimer formation	$\Delta$ ( <i>ara-leu</i> )7697 $\Delta$ <i>lacX74 <math>\Delta</math>phoA PvuII phoR araD139 ahpC galE galK rpsL F'[lac+ lacIq pro]</i> (DE3) <i>gor522::Tn10 trxB</i> (StrR, TetR)	Expression of insoluble proteins

## E. COLI STRAINS FOR PROTEIN EXPRESSION (CONT'D)

### How Does Inducible Expression Work?



As mentioned above, many expression plasmids utilize inducible promoters, which are ‘inactive’ until an inducer such as IPTG is added to the growth medium. Induction timing is important, as you typically want to make sure your cells have first reached an appropriate density. Cells in the exponential growth phase are alive and healthy, which makes them ideal for protein expression. If you wait too long to induce, your culture will start collecting dead cells, and, conversely, you cannot induce too early as there are not enough cells in the culture to make protein.

The DE3 lysogen/T7 promoter combination is the most popular induction system. The DE3 lysogen expresses T7 RNA polymerase (RNAP) from the bacterial genome under control of the lac repressor, which is inducible by the addition of IPTG. T7 RNAP is then available to transcribe the gene of interest from a T7 promoter on the plasmid. Many commercial strains carry the DE3 lysogen, as indicated by the name of the strain. Conversely, other strains such as M15(pREP4) use a lac repressor to act directly on the expression plasmid in order to repress transcription from a hybrid promoter.

Although the DE3/T7 RNAP system works well for most experiments, the lac promoter can “leak,” meaning that a low level of expression exists even without the addition of IPTG. This is mostly a problem for toxic protein products, which can prevent the culture from reaching the desired density within a reasonable time-frame. For these cases, some strains carry an additional measure of control such as the pLys plasmid, which suppresses basal T7 expression. The pLys plasmid contains a chloramphenicol resistance cassette for positive selection and a p15A origin of replication, making it incompatible with other p15A plasmids. pLys comes in two flavors—pLysS and pLysE—the difference being that the latter provides tighter control of basal expression.

### What if I Don't See Protein Overexpression?

The strains described above should generate sufficient expression levels for most purposes, but what do you do when you've tried a common strain and don't get the desired level (or any) protein expression? Low expression outcomes can result from variety of sources, so fear not—there are a few simple troubleshooting measures that can help get you back on track:

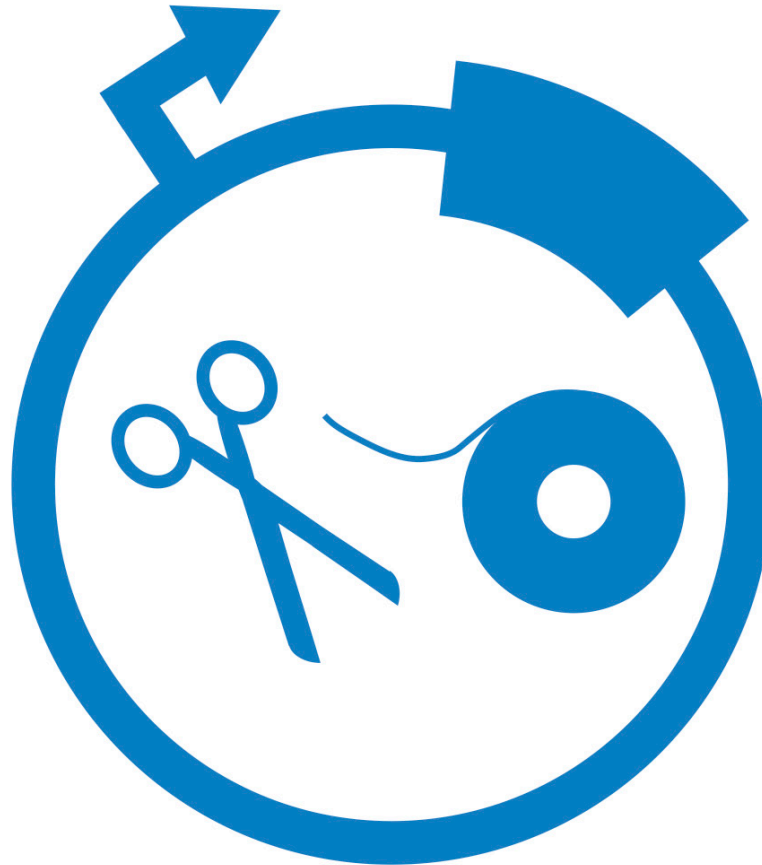
- **Compatibility:** Double-check your plasmid backbone and expression strain to make sure they are compatible. An arabinose-inducible plasmid will not express in an IPTG induction strain for example, nor will a p15 plasmid be compatible with a pLys strain. Your strain may require additional antibiotic selection or a special growth media, or if your plasmid is low-copy, consider reducing the antibiotic concentration.
- **Growth Temperature:** Analyze your expression conditions by setting up a small-scale expression experiment to test variables such as temperature, time, and media conditions. Many recombinant proteins express better at 30°C or room-temperature, which is accomplished by growing your culture to the desired density at 37°C and reducing the temperature or moving it to a bench-top shaker 10-20 minutes before adding the inducer.

## E. COLI STRAINS FOR PROTEIN EXPRESSION (CONT'D)

---

- **Growth Media:** Changing media is tricky, because there can be a trade-off between growth rate and protein quality. For many proteins, a rich media such as TB or 2XYT is optimal because of the high cell-density they support; however, minimal media supplemented with M9 salts may be preferable if the protein product is secreted to the medium or if slow expression is required due to solubility concerns.
- **Insoluble and Secreted Proteins:** The most common purification protocols are designed for soluble, cytosolic protein products, but this is not always achievable. Proteins which contain hydrophobic regions or multiple disulfide bonds may aggregate and become insoluble. These insoluble globs of misfolded protein are known as inclusion bodies, and can be recovered and purified using a [special protocol](#). Alternatively, reducing the concentration of inducer or adding an affinity tag such as GST may help with solubility issues.

# CHAPTER 2: COMMON CLONING TECHNIQUES

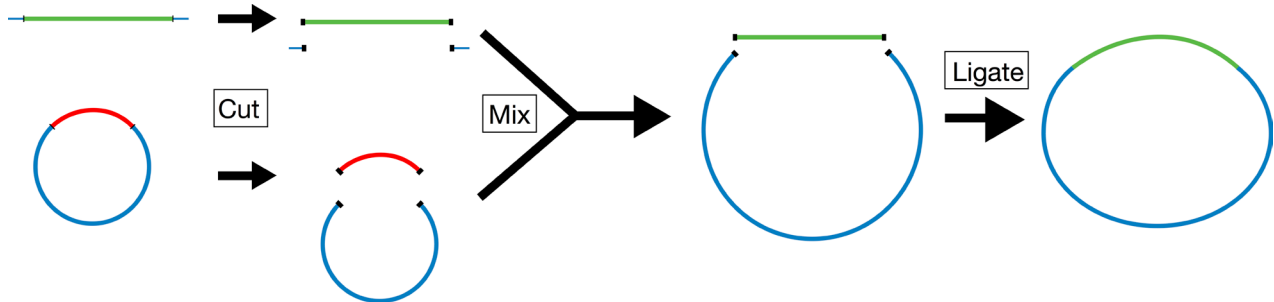


# RESTRICTION CLONING

By Various Addgenies | August 28, 2015

## Overview

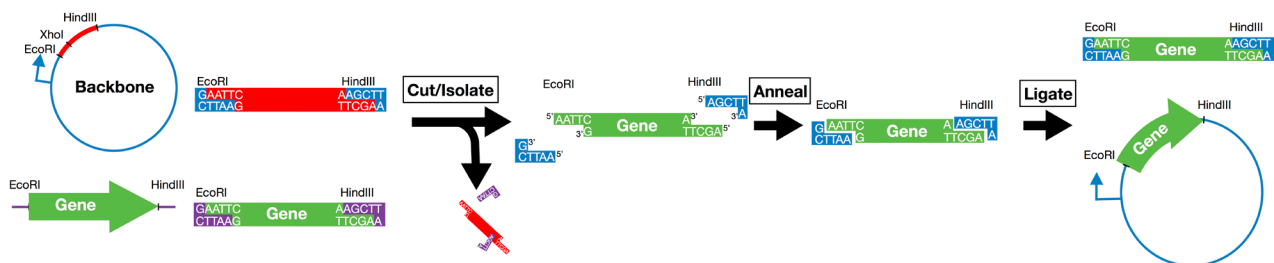
When cloning by restriction digest and ligation, you use restriction enzymes to cut open a plasmid (backbone) and insert a linear fragment of DNA (insert) that has been cut by compatible restriction enzymes. An enzyme, DNA ligase, then covalently binds the plasmid to the new fragment thereby generating a complete, circular plasmid that can be easily maintained in a variety of biological systems.



*Overview of the restriction cloning process. Both the plasmid (blue, backbone) and the DNA sequence of interest (green, insert) are cut with restriction enzymes to generate compatible overhangs that allow them to bind together. Ligase is used to make bonds between the insert and backbone covalent.*

Before beginning the restriction digest and ligation process, you should carefully choose your backbone and insert - these both must have compatible cut sites for restriction enzymes that allow your insert to be placed into the backbone in the proper orientation. For instance, if you were cloning a gene into an expression vector, you would want the start of the gene to end up just downstream of the promoter found in the backbone. Ideally, the backbone will contain a variety of restriction enzyme cut sites (restriction sites) downstream of the promoter as part of a multiple cloning site (MCS). Having multiple sites allows you to easily orient your gene insert with respect to the promoter.

For example, let's say your plasmid backbone looks like the one found on the left side of the image below. It has a promoter (blue arrow) followed by the restriction sites EcoRI, XhoI, and HindIII. To place your gene in the proper orientation downstream of the promoter, you can add an EcoRI site just 5' of the start of the gene and a HindIII site just 3' of the end of the gene. This way you can then cut the plasmid backbone as well as the insert with EcoRI and HindIII and, when you mix the cut products together, the two EcoRI digested ends will anneal and the two HindIII digested ends will anneal leaving the 5' end of your gene just downstream of the promoter and placing the gene in the proper orientation. You then add ligase to the mixture to covalently link the backbone and insert and, PRESTO, you have a full plasmid ready to be used in your experiments.



*Detailed overview of the restriction cloning process. See text for details.*

# RESTRICTION CLONING (CONT'D)

Alternatively, this whole process can be completed using a single enzyme if your insert is flanked on both sides by that enzyme's restriction sites, but the insert can then anneal to the backbone in either a forward or reverse orientation so you'll need some way to verify that the insert ended up in the direction you want - usually by Sanger sequencing or further restriction digests.

Of course there's much more detail and verification required for the process to work well, so let's go over the details step-by-step.

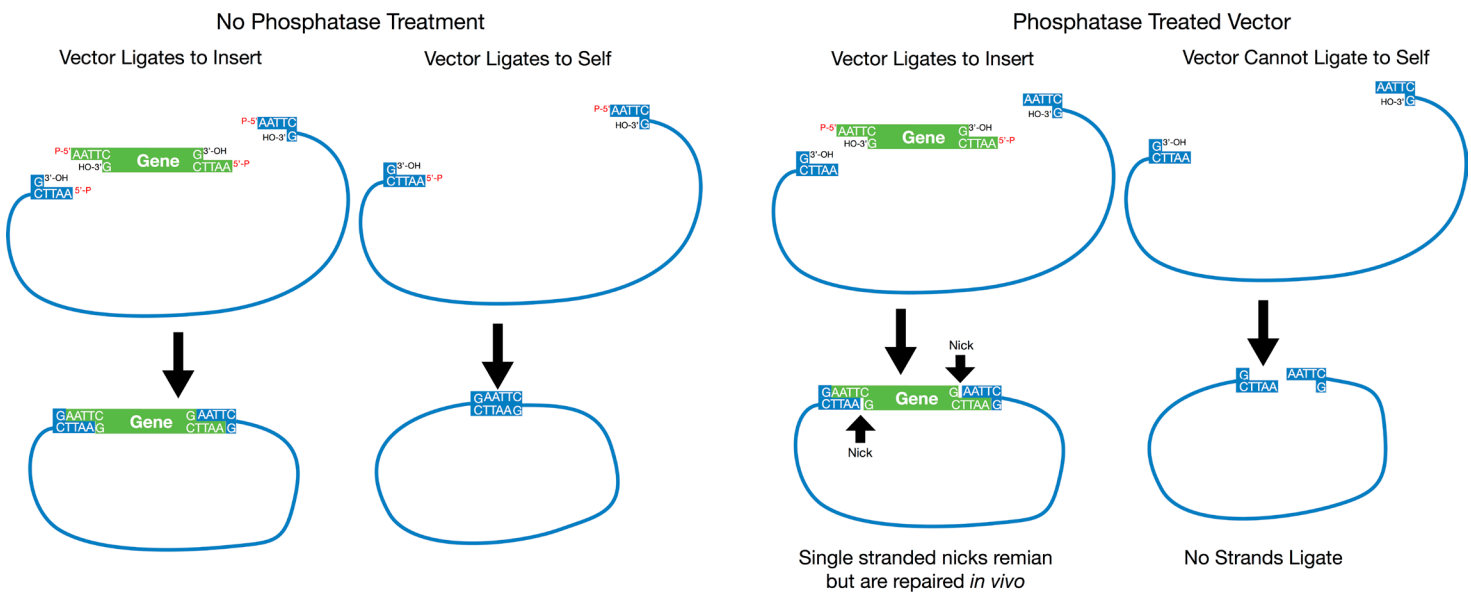
## 1. Digestion

Set up [restriction digests](#) for your insert (or donor plasmid) and plasmid backbone. Because you lose some DNA during the gel purification step, it is important to digest plenty of starting material. We recommend 1.5-2 µg of insert and 1 µg of plasmid backbone. It is also critical that as much of the backbone plasmid as possible be cut with both enzymes, and therefore it is important that the digest go until completion. The time required for complete digestion varies for different enzymes. Many companies now sell fast digest enzymes that can digest large amounts of DNA in as little as 10 minutes, but check with your enzyme's manufacturer to ensure that you're cutting for the proper duration and using the proper conditions.



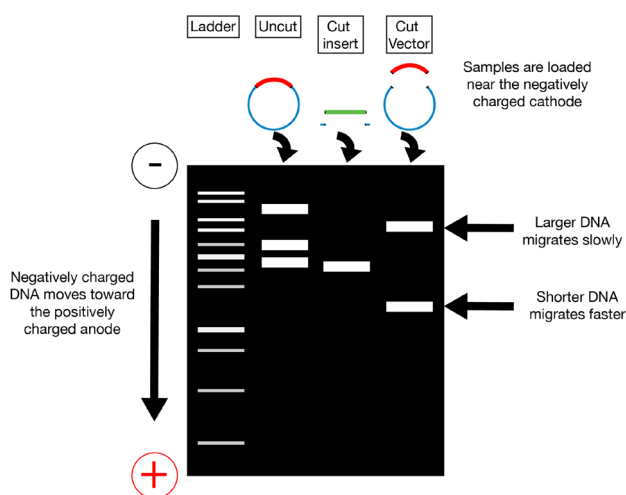
### Pro Tip!

If you are going to use only one restriction enzyme, or enzymes that have compatible overhangs or no overhangs after digestion, you will need to use a phosphatase to prevent re-circularization of the backbone plasmid (see below). You should treat your digested backbone plasmid with a phosphatase prior to the ligation step or prior to the gel purification step, depending on the phosphatase you choose. CIP (calf alkaline phosphatase) or SAP (shrimp alkaline phosphatase) are commonly used. Follow the manufacturer's instructions.



## RESTRICTION CLONING (CONT'D)

### 2. Isolate Your Insert and Vector by Gel Purification



Now that you've cut your insert and vector, unfortunately you can't just throw the digestion mixtures together. You need to isolate your insert and backbone from the enzymes used to digest them as well as any pieces cut out or off of them. An easy way to do this is gel purification. In gel purification, you use a voltage difference across a gel matrix (usually agarose) to pull your negatively charged DNA through the gel. As indicated in the figure on the left, your digested DNA (and undigested controls) are loaded at the top of the gel in wells positioned toward the cathode (- charge). When the voltage is applied across the gel, the DNA migrates toward the anode (+ charge). Larger fragments of linearized DNA migrate slower than smaller linearized fragments. You can separate your backbone away from

any inserts cut out of it and your new insert from any overhangs cut off of it via their different migration speeds; after running the gel for some duration of time, these differently sized pieces will be in different locations and can be cut out of the gel individually.

There are a variety of ways to visualize the DNA in your gel.

**Table 1: Common Gel Stains**

Stain	Pre or Post Run?	Visualization	Sensitivity (ng DNA)
SYBR Safe	Pre and Post	Blue or UV Light	0.5
GelRed	Pre and Post (post recommended)	UV Light	0.1
GelGreen	Pre and Post (post recommended)	UV or Green (~500nm) Light	0.1
Crystal Violet	Pre and Post	Visible Light	100
Methylene Blue	Post	Visible Light	100
Ethidium Bromide	Pre and Post	UV Light	0.5

## RESTRICTION CLONING (CONT'D)

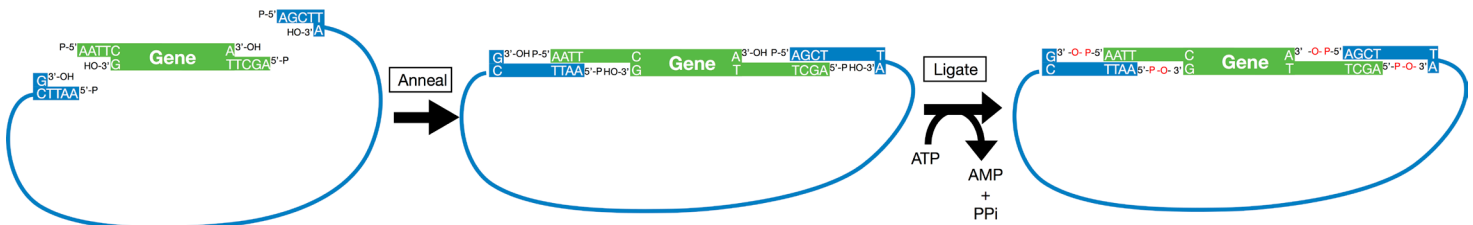
These stains require you to either stain your gel after you run your samples or add the stain as the gel is being made (post or pre run in the table above, respectively). Some of the above stains require you to visualize your DNA using UV light – please note that UV light can damage DNA and that proper personal protective equipment should be worn when visualizing using UV as it can cause damage to the eyes and skin.

When running a gel for purification purposes it is important to have nice crisp bands and to have space to cut out the bands. Because of this we recommend that you use a wide gel comb, run the gel on the slower side, and skip lanes between samples. In addition to a DNA ladder standard, it is also a good idea to run an uncut sample of each plasmid to help with troubleshooting if your digests don't look as expected.

Once you have cut out and purified your insert and recipient plasmid backbone bands away from the gel via your favorite [gel purification](#) method, it is important to [determine the concentration of recovered DNA](#) as this will be useful for the ligation step.

### 3. Ligate Your Insert into Your Vector

In the ligation step, you mix your purified, cut backbone and insert in a single tube allowing the compatible overhangs generated by restriction digestion to anneal to one another and form a complete, circular plasmid. You then add DNA ligase to covalently link the fragments together at the expense of ATP (see below).



We recommend around 100ng of total DNA in a standard ligation reaction. You ideally want a “recipient plasmid : insert ratio” of approximately 1:3. Since the number of base pairs for each varies, it is difficult to calculate this based on DNA concentration alone. One method is to conduct 2 ligations for each plasmid you are trying to create, with varying ratios of recipient plasmid to insert.

It is critical to set up a negative control ligation reaction with no insert DNA added. This will allow you to determine how many colonies you should expect in the transformation due to background re-circularization and contamination from uncut plasmid.

### 4. Transformation

[Transform](#) your ligation reaction into your bacterial strain of choice. Follow the manufacturer's instructions for your competent cells.

For most standard cloning, you can transform 1-2  $\mu$ l of your ligation reaction into competent cells such as DH5alpha or TOP10. If using much less total DNA (<1 ng) or if you are having trouble getting colonies, you might want to use higher competency cells. Additionally, if your final product is going to be very large (>10kb) you might want to use electro-competent cells instead of the more common chemically-competent cells.



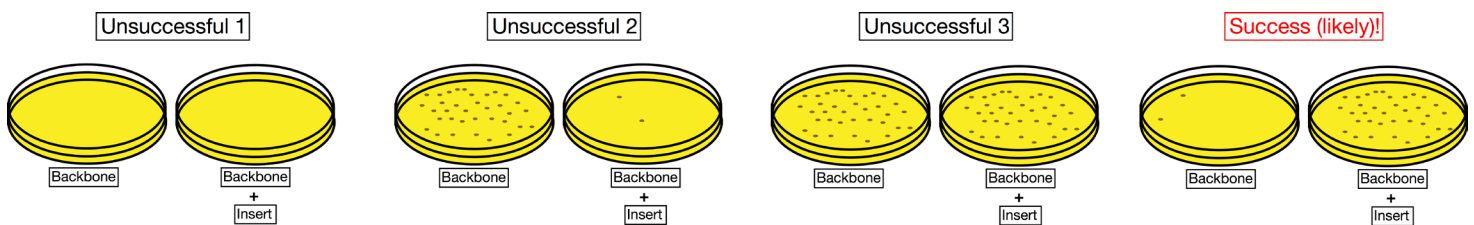
## RESTRICTION CLONING (CONT'D)

You should perform, at minimum, two transformations after a ligation:

1. Control transformation containing the ligation mixture with backbone alone;
2. Transformation containing ligation mixture with insert and backbone.

Sample results indicative of successful and unsuccessful ligations are indicated below. A successful ligation will have few colonies on the backbone alone plate and many colonies on the backbone + insert plate (or at least more colonies than the backbone alone plate). Unsuccessful ligations will usually result in few colonies on both plates (unsuccessful 1), in a vector alone plate with many more colonies than the vector + insert plate (unsuccessful 2), or roughly equivalent numbers of colonies on each plate (unsuccessful 3).

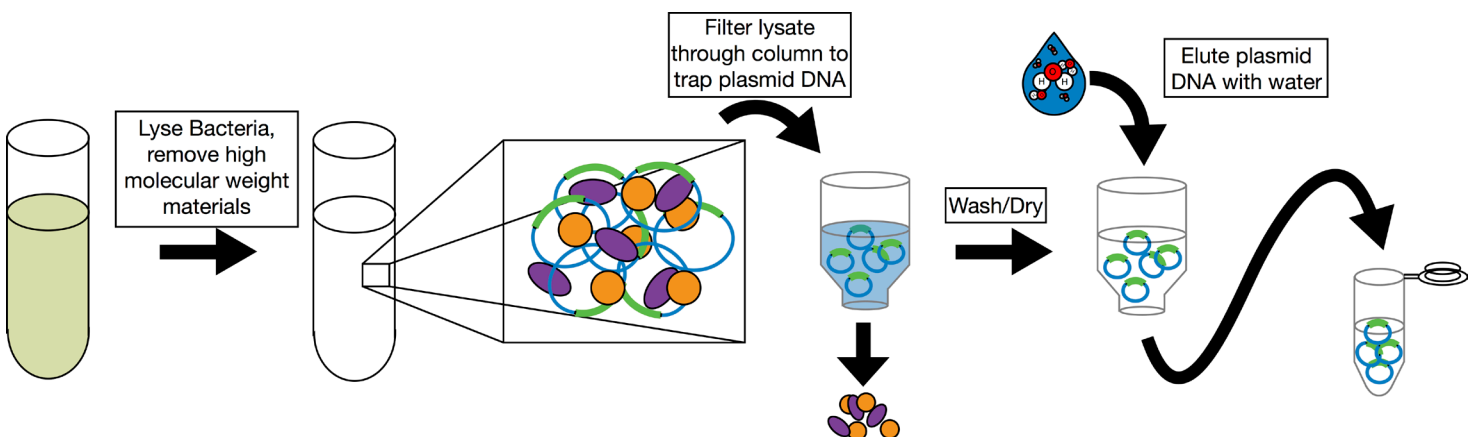
If you have a high number of colonies on your backbone plate (greater than or equivalent to backbone + insert,



unsuccessful 2 and 3 above), you can try ligating the recipient plasmid alone in the presence and absence of ligase. If the colonies are a result of uncut empty plasmid, you will still have colonies when you do not add ligase. If the colonies are a result of recipient plasmid self-ligation, you will see significantly more colonies when you add ligase.

If you do not see any colonies, you should conduct a positive control to ensure that your transformation worked. You should also verify that you are plating on the appropriate antibiotic and try varying the “recipient plasmid : insert ratio” in the ligation reaction.

## 5. Purify the Finished Plasmid



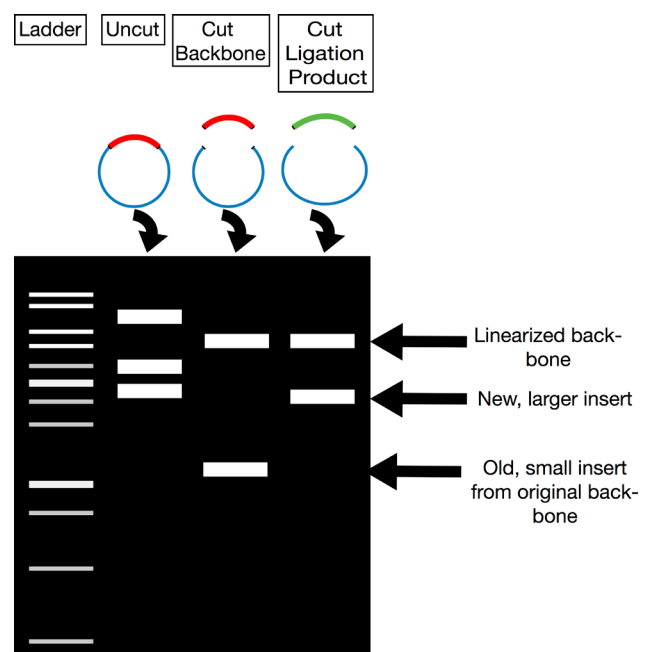
## RESTRICTION CLONING (CONT'D)

Once it looks like your ligation has worked, you will need to pick individual bacterial colonies and check them for successful ligation. Pick 3-10 colonies depending on the number of background colonies on your control plate (the more background, the more colonies you will need to pick) and grow overnight cultures for DNA purification. The simplest purification you can do is a miniprep, but if you need larger quantities of DNA, you'll need to do a midiprep or a maxiprep. In these purifications, you generally lyse the bacteria; add chemicals to precipitate out the high molecular weight genomic DNA; filter the remaining plasmid DNA through a column that binds the plasmid DNA and lets other materials pass through; and, finally, selectively elute the plasmid DNA from the column using a particular buffer or water. See column manufacturers for more detail. Columnless purification protocols can be found [here](#).

### 6. Verify the Finished Plasmid

After [purifying the DNA](#), conduct a [diagnostic restriction digest](#) of 100-300 ng of your purified DNA with the enzymes you used for cloning. Run your digest on an [agarose gel](#). You should see two bands, one the size of your backbone and one the size of your new insert (see right). If you used only one enzyme or used enzymes with compatible overhangs for your ligation, then you will need to verify the orientation of your insert. You may want to design a diagnostic digest for this purpose.

Ideally, once you know that your plasmid has an appropriately sized insert, you should send it for sanger sequencing using primers that will allow you to read over the insert. For more information on how to verify your plasmid see Chapter 9 below.



Once your complete plasmid has been verified, you're ready to get experimenting!

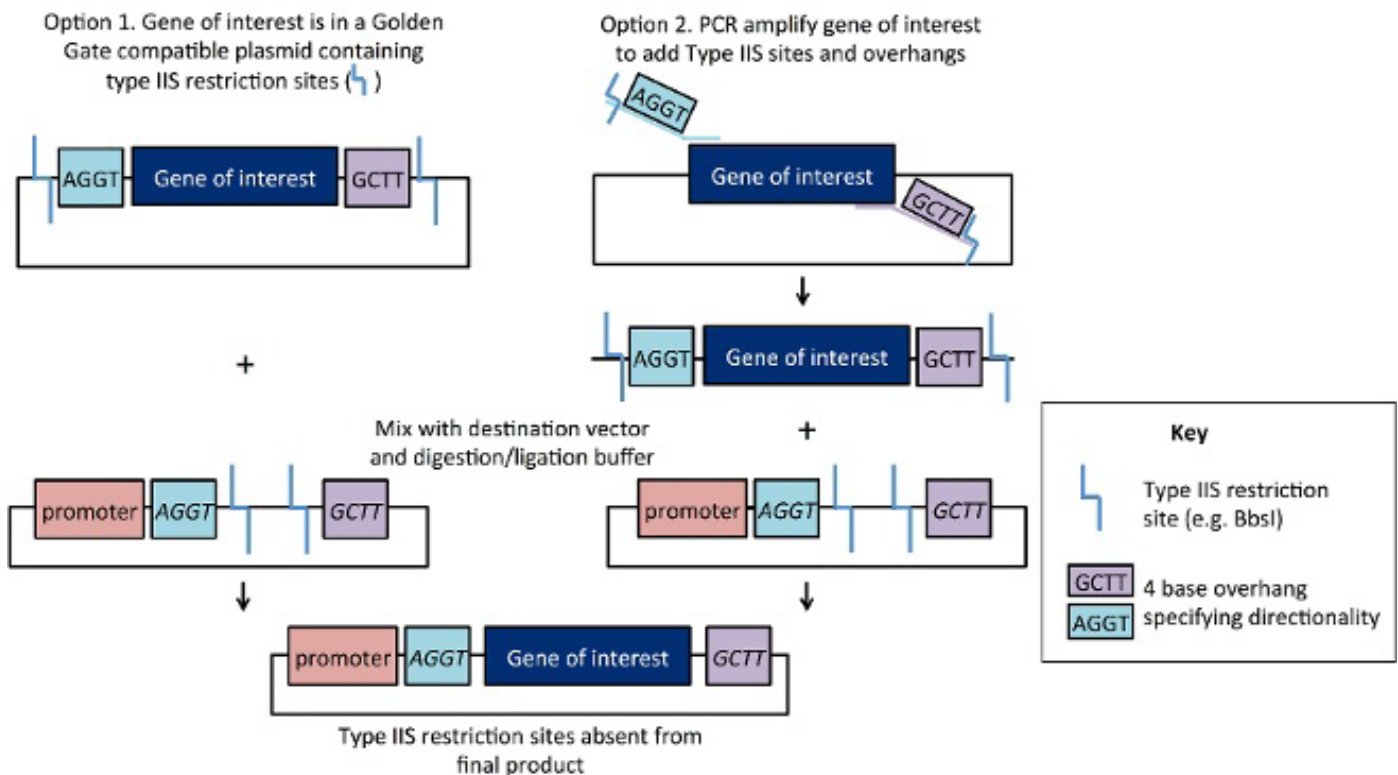
# GOLDEN GATE CLONING

By Mary Gearing, Addgene | August 27, 2015

Addgene's plasmids are used with a wide variety of restriction enzyme-based cloning methods. Each method has its own pluses and minuses, but Golden Gate cloning has been especially useful within both the [synthetic biology](#) and [genome engineering](#) fields. We'll walk you through how to apply this precise and easy-to-use system to your cloning efforts.

Golden Gate cloning technology relies on Type IIS restriction enzymes, first discovered in 1996. Type IIS restriction enzymes are unique from "traditional" restriction enzymes in that they cleave outside of their recognition sequence, creating four base flanking overhangs. Since these overhangs are not part of the recognition sequence, they can be customized to direct assembly of DNA fragments. When designed correctly, the recognition sites do not appear in the final construct, allowing for precise, scarless cloning.

The cloning scheme is as follows: the gene of interest is designed with Type IIS sites (such as BsaI or BbsI) that are located on the outside of the cleavage site. As a result, these sites are eliminated by digestion/ligation and do not appear in the final construct. The destination vector contains sites with complementary overhangs that direct assembly of the final ligation product. As shown below, a fragment with 5' overhang TGGGA and 3' overhang TCCG can be ligated into a vector containing those overhangs. Entry DNA overhangs may be present in the original plasmid (Option 1) or added using PCR-based amplification (Option 2).

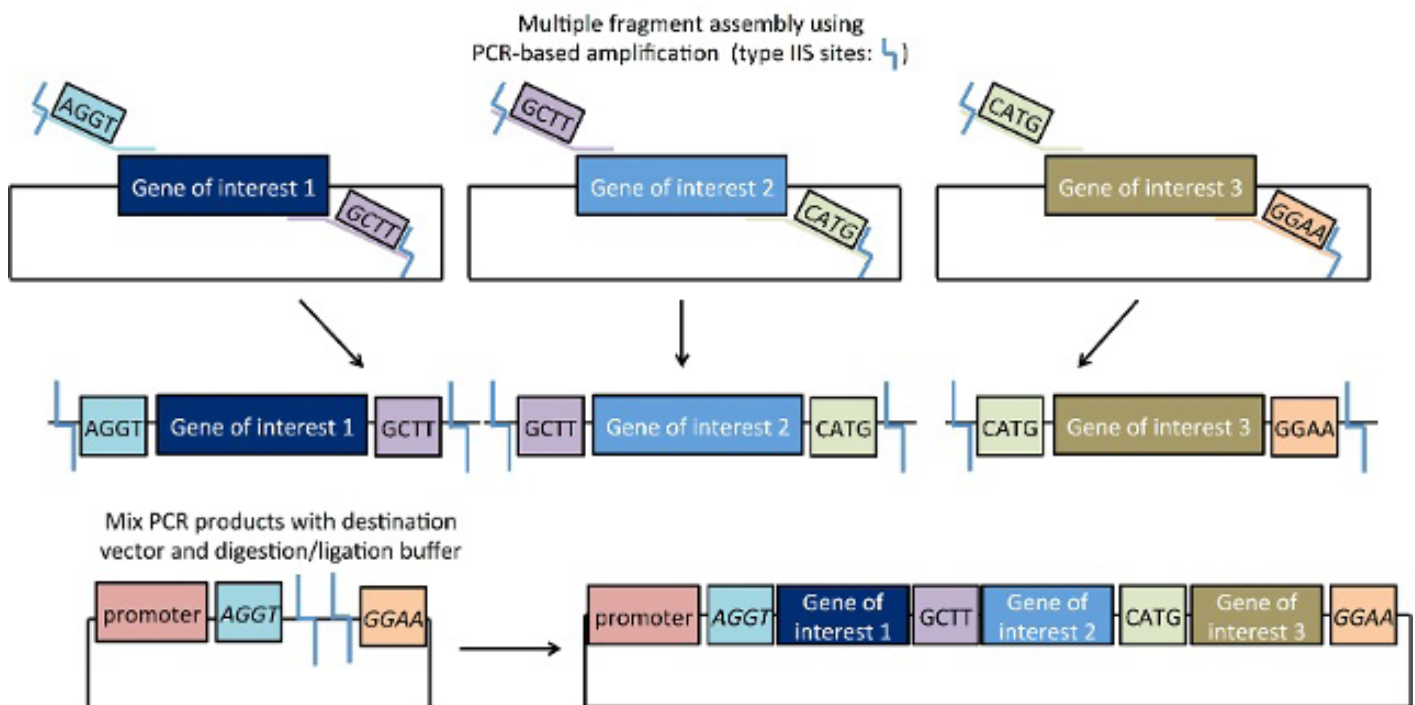


# GOLDEN GATE CLONING (CONT'D)

## Advantages of Golden Gate Cloning

Golden Gate cloning is one of the easiest cloning methods in terms of hands-on time, as digestion and ligation can be done in one 30-minute reaction. The destination vector and entry vector(s) are placed in a single tube containing the Type IIS enzyme and ligase. Although the original destination vector + insert may spontaneously re-ligate, this transient construct retains functional Type IIS sites and will be re-digested. In contrast, formation of the desired ligation product is irreversible because this construct does not retain the enzyme recognition sites. As a result, the ligation process is close to 100% efficient. Another strength of Golden Gate cloning is its scalability. Unique 4 base overhangs can be used to assemble multiple fragments - up to 10 fragments are commonly assembled in a single reaction! These overhangs specify the desired order of fragments, and the loss of enzyme recognition sites after ligation favors formation of the construct of interest. Although efficiency may decrease with an increased number of fragments, or the ligation of very small/very large fragments, these problems can be overcome by screening a higher number of potential clones.

Golden Gate assembly has a few advantages over [other cloning methods](#). Exonuclease-based methods like [Gibson assembly](#) require 20-40 bp of homology at the ends of DNA fragments to specify assembly order, so fragments with 5' or 3' sequence homology cannot be assembled using this method, but can be assembled with Golden Gate. The popular [Gateway cloning system](#) produces constructs with an attB recombination scar encoding eight amino acids, but Golden Gate assembly can be designed to be scarless. Golden Gate assembly is also less expensive than many commercial cloning methods.



## GOLDEN GATE CLONING (CONT'D)

---

### Golden Gate and Synthetic Biology

Synthetic biologists have leveraged the power of Golden Gate cloning into a modular cloning strategy. Sometimes referred to as MoClo, this strategy uses the Type IIS restriction enzymes BsaI and BpiI/BbsI to efficiently assemble up to six DNA fragments at a time. As with all Golden Gate-based methods, this system exploits the ability of Type IIS enzymes to cut outside their recognition site and permits DNA fragments with compatible overhangs to be efficiently assembled. Scientists can engineer unique enzyme recognition sites that flank their DNA fragment in an inverse orientation. This allows for multiple DNA components (promoters, genes, terminators, etc.) to be assembled in the correct order in a single reaction. For detailed Golden Gate protocols, complete with helpful tips and tricks, see [The Sainsbury Lab website](#) or [Engler & Marillonet](#).

### Golden Gate and Genome Engineering

In early 2011, the Bogdanove and Voytas groups described a new [Golden Gate-based technology](#) for genome editing which allowed for the ordered assembly of multiple DNA fragments to create TAL effector nucleases. These plasmids were designed to utilize the BsaI and BsmBI Type IIS sites such that custom TAL arrays could be constructed quickly and efficiently in just a few steps. More recently, CRISPR technology has adapted Golden Gate cloning for inserting the appropriate oligonucleotides specifying a gRNA target sequence into a Cas9-containing plasmid such as [pX330](#). This cloning strategy not only makes it easy to create a single gRNA-expressing plasmid, but it can also be adapted to express multiple gRNAs. Addgene has [two Golden Gate-based gRNA assembly methods](#) available, which allow you to efficiently clone up to 7 gRNAs into one destination vector, making multiplexing easy.

### Disadvantages of Golden Gate Cloning

Golden Gate cloning is not 100% sequence-independent: to avoid undesired digestion, the Type IIS site used must not be present within the fragments you seek to assemble. One way to work around this is to “domesticate” your fragment: PCR-based amplification can be used to create silent point mutations at internal recognition sites thus eliminating these from your gene of interest. PCR products are then digested with the Type IIS enzyme, and the mixture is ligated following a heat inactivation step. If your genes of interest or destination vector contain multiple internal restriction sites that may not be amenable to “domestication”, you might want to consider using an alternative method like Gateway cloning or Gibson assembly. Another important consideration is the design of flanking overhangs. Although there are theoretically 256 distinct flanking sequences, sequences that differ by only one base may result in unintended ligation products.

Whether you are using the Golden Gate method to create CRISPR/Cas9 constructs, assemble standard plasmids parts in different combinations, or other new and exciting applications, this system is an incredibly powerful tool for cloning complicated constructs in a single, high-efficiency step.

## GOLDEN GATE CLONING (CONT'D)

---

### Further Reading

1. Lee, Jae H., et al. "Sequential amplification of cloned DNA as tandem multimers using class-II restriction enzymes." *Genetic analysis: biomolecular engineering* 13.6 (1996): 139-145. PubMed [PMID: 9117889](#).
2. Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet. "A one pot, one step, precision cloning method with high throughput capability." *PloS one* 3.11 (2008): e3647. PubMed [PMID: 18985154](#).
3. Cermak, Tomas, et al. "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting." *Nucleic acids research* (2011): gkr218. PubMed [PMID: 21493687](#).
4. Engler, Carola, and Sylvestre Marillonnet. "Golden gate cloning." *DNA Cloning and Assembly Methods* (2014): 119-131. PubMed [PMID: 24395361](#).

# TOPO CLONING

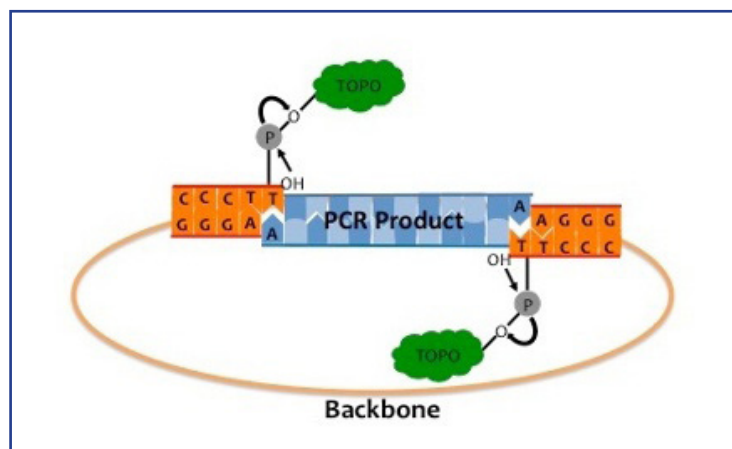
By Lianna Swanson | October 27, 2016

Topoisomerase based cloning (TOPO cloning) is a [DNA cloning](#) method that does not use [restriction enzymes](#) or [ligase](#), and requires no post-PCR procedures. Sounds easy right? The technique relies on the basic ability of complementary basepairs adenine (A) and thymine (T) to hybridize and form hydrogen bonds. This post focuses on “sticky end” TOPO (also called TOPO-TA) cloning; however, the TOPO cloning technique has also been adapted for blunt end cloning.

## So How Does TOPO Cloning Work?

As illustrated in the Figure below, the “A” overhang on the blue PCR product insert comes from using Taq polymerase for the amplification step since Taq polymerase leaves a single deoxyadenosine (A) at the 3' ends of PCR products. The complimentary “T” in the pair comes from a topoisomerase I-linearized backbone. DNA topoisomerase I (depicted as a green cloud) functions both as a restriction enonuclease and as a ligase by cleaving and rejoining supercoiled DNA ends to facilitate replication.

The TOPO technique specifically uses Vaccinia virus-isolated topoisomerase I as this enzyme recognizes the DNA sequence 5'-(C/T)CCTT-3' and digests double stranded DNA at this sequence. The energy from this breakage is stored as a covalent bond between the cleaved 3' DNA strand and a tyrosyl residue of topoisomerase I (1). If a 5' hydroxyl group from a different DNA strand comes along, it can attack this covalent bond thus joining the two DNA strands and releasing topoisomerase (2).



These days commercially available TOPO kits provide vectors or cloning arms with overhanging 3'

deoxythymidine (T) residues that are covalently linked to topoisomerase. Vectors in these kits also often have the topoisomerase site inserted into a beta-galactosidase cassette allowing a researcher to perform [blue-white screening](#) after [transformation](#) - self joining of the vector ends results in the production of blue colonies that do not need to be picked and [sequenced](#) for potential positive clones. Once you introduce your 3'-end “A” overhang insert, the magic of TOPO cloning happens.

## Basic Procedure

Let's break down the steps needed for TOPO cloning:

- 1. Create Your PCR Product:** Design standard primers (no need to add unique restriction sites on the ends) and amplify your sequence of interest with Taq polymerase using your favorite PCR protocol.
- 2. Set Up the TOPO Cloning Reaction:** Mix together the PCR product and TOPO Vector.
- 3. Incubate 5 Minutes at Room Temperature:** You can place your reaction on ice if you are planning to transform right away OR you can store the reaction at -20 °C overnight.
- 4. Transform TOPO Cloning Reaction into Competent Cells:** You can use your standard lab protocol for this;

## TOPO CLONING (CONT'D)

---

however, you should reduce your incubation time on ice to 5 minutes (incubating the full 30 minutes will not significantly improve the transformation efficiency).

**5. Select and Analyze 10 White or Light Blue Colonies:** You can confirm the presence of your insert by PCR, restriction digest, or sequencing.



### Pro Tip!

- Do not add 5' phosphates to your PCR primers; you need that free hydroxyl group!
- You may want to include extra extension time after the last cycle of PCR to make sure that the “A” gets added to all PCR products.
- Keep in mind that Taq polymerase has an error rate of about 1 in 3,500 bases. Typically polymerases with proofreading functionality are used in place of Taq to reduce error rates; however, proofreading polymerases will also remove all unpaired 3' ends in your PCR product. If you need to decrease error rate, please use one of these methods to ensure your insert retains the 3' A overhang:
  - Use a mixture of proofreading enzyme and Taq, with Taq used in an excess ratio of 10:1.
  - [Gel purify your PCR product](#) and incubate it with buffer, Taq and dATPs at 72 °C for 10-15 min.
- When mixing the PCR product with the TOPO vector, you may want to add extra salt to your reaction:

Topoisomerase I is released from the vector when the PCR product and vector ligate; however, it can potentially rebind and nick the newly ligated DNA. Salt helps prevent topoisomerase I from rebinding, which results in more intact molecules. (Note that the amount of salt you add will depend on whether you are planning on transforming your reaction into chemically or electro-competent *E. coli* - excess salt causes arcing during electroporation which would cause the electroporation to fail).
- When incubating at room temperature, it is not recommended that you exceed the 5 minute time limit (lower transformation efficiencies have been reported with longer incubation); however, you may need to incubate for 20-30 minutes if your PCR product is at a low concentration or you are cloning an extremely large insert.
- Since the standard ligation reaction is fairly quick, make sure you stay organized and prepare everything you need for the next step before proceeding.
- Pre-warming your antibiotic-containing plate prior to plating your transformation may allow you to see colonies within 8 hours.



## TOPO CLONING (CONT'D)

---

### Further Reading

1. Shuman S. "Recombination mediated by vaccinia virus DNA topoisomerase I in Escherichia coli is sequence specific." Proc Natl Acad Sci U S A. 1991 Nov 15;88(22):10104-8. PubMed [PMID: 1658796](#). PubMed Central [PMCID: PMC52876](#).
2. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. Shuman S. J Biol Chem. 1994 Dec 23;269(51):32678-84. PubMed [PMID: 7798275](#).

# SEQUENCE AND LIGATION-INDEPENDENT CLONING (SLIC)

By Mary Gearing, Addgene | September, 2015

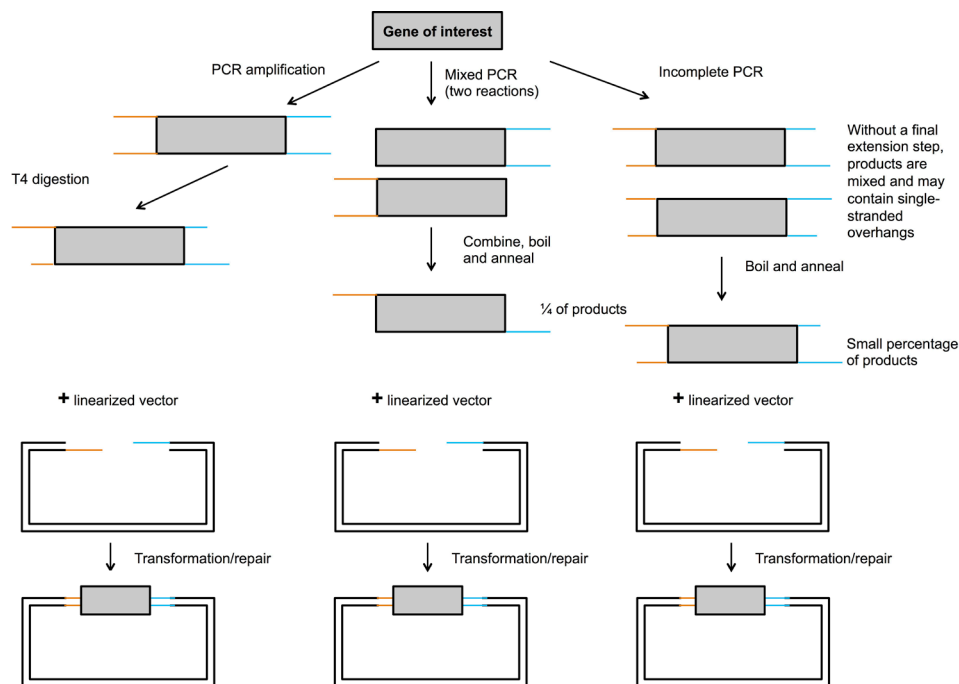
If cloning methods had personalities, SLIC (sequence- and ligation-independent cloning) would be a true rebel. Not only does this system not use site-specific recombination, it also doesn't require a ligation step! Based on the robust system of homologous recombination found in *E. coli*, SLIC is a cheap, standardized, and rapid multi-part DNA assembly method - read on to learn how to use it in your research.

## The First Step: Ligation-Independent Cloning

[Ligation-independent cloning \(LIC\)](#) was first developed in the 1990s. While traditional restriction enzyme cloning used short sticky ends, LIC employed the exonuclease activity of T4 DNA polymerase to create longer, "chewed-back" overhangs of about 10-12 bases. Only one type of dNTP would be present in the reaction mix, limiting the exonuclease activity to the first occurrence of that nucleotide. At that position, T4 would perform the favored polymerase reaction and subsequently stall due to the absence of other dNTPs. Once digested separately, the vector and insert could be annealed, forming a circular product with four nicks easily repairable by the bacteria after transformation.

LIC is a reliable cloning method, but it is limited by its sequence constraints. The 10-12 base overhangs must not contain the dNTP present in the reaction, or polymerization will occur at that position, preventing T4 from chewing back the entire 10-12 bases. As such, the use of LIC is often limited to specially-designed plasmids.

## A SLIC Makeover: Sequence and Ligation-Independent Cloning



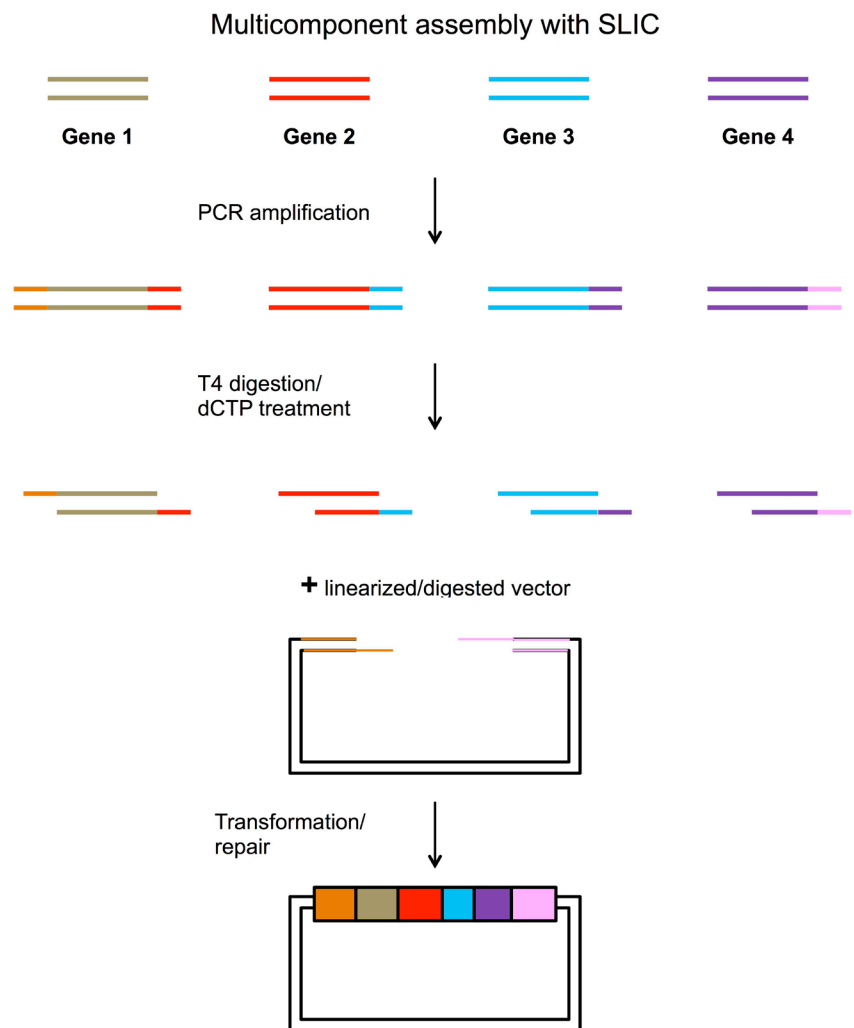
Means of using SLIC

## SEQUENCE AND LIGATION-INDEPENDENT CLONING (SLIC) (CONT'D)

In 2007, LIC received an important update, courtesy of Addgene depositor [Stephen Elledge](#). His new method, named sequence and ligation-independent cloning (SLIC), eliminates many of LIC's constraints. Key to SLIC is the power of homologous recombination. In *E. coli*, a robust homologous recombination system allows for the repair of gaps and overhangs based on regions of sequence homology. This process can occur through one of two pathways: RecA-mediated recombination or RecA-independent single-stranded annealing. Elledge realized that he could introduce imperfect "recombination intermediates" generated through PCR and imprecise T4 exonuclease activity, rather than the carefully designed DNA fragments used in LIC (see figure on the previous page). As long as there was enough sequence homology (20-60 bp) to organize the fragments and hold them together, *E. coli* would be able to "repair" the plasmid, generating recombinant DNA. Adding purified RecA to the pre-transformation incubation enhances the repair process, allowing SLIC to be used with very small amounts of DNA (e.g. 3 ng). With larger amounts of DNA (~100 ng), RecA is not required.

### The Practicalities of Being SLIC

To start the SLIC cloning process (see figure on the previous page), a fragment of interest is amplified using PCR to add in the specified 5' and 3' homology regions. This fragment and linearized plasmid are then partially digested using T4 polymerase in the absence of dNTPs. In the second step, the addition of a single dNTP stops the exonuclease reaction. These products are then combined, annealed, and transformed into *E. coli*. To make matters even easier, SLIC is also compatible with incompletely synthesized (iPCR) fragments. If PCR does not include a final extension step, many of the products will have single-stranded overhangs due to incomplete extension, and these fragments can induce recombination. Mixed PCR can also be used to create an insert. Two PCR products are used to generate a target gene with a 5' or 3' overhang. When the products are mixed and annealed, 25% of the resulting DNA will have two single-stranded overhangs that can robustly stimulate recombination.



## SEQUENCE AND LIGATION-INDEPENDENT CLONING (SLIC) (CONT'D)

---

SLIC is ideal for multicomponent assembly (see figure on previous page), as overlapping sequence homology specifies the order of multiple fragments, and the assembly is scarless. With 40 bp homology regions, a five piece assembly reaction is highly efficient (~80%). Ten-fragment assembly can also be successful, but at a lower efficiency (~20%).

### How Does SLIC Compare to Other Cloning Methods?

SLIC's limitations arise from its dependence on single-stranded overhangs. These overhangs must be accessible to allow for complementary base pairing, so SLIC can't be used if the overhangs would have stable ssDNA secondary structure. One common example is the stem-loop structure of transcriptional terminators. Another potential issue is sequence similarity. If fragments in a multicomponent assembly have 5' or 3' sequence homology to each other, they may be assembled incorrectly. To overcome this limitation, one option is to perform a hierarchical assembly, assembling fragments in multiple steps to avoid using multiple fragments that share homology in the same reaction. In these cases, the use of another cloning method, such as Golden Gate assembly, may also be beneficial.

SLIC is most often compared to Gibson assembly, another cloning method based on homologous recombination. Instead of using T4 DNA polymerase, Gibson assembly requires T5 exonuclease in combination with Phusion polymerase and DNA ligase. This reaction takes place in one step rather than two steps required for SLIC, and ligase may improve the efficiency of multipart assembly. The higher temperature at which Gibson assembly takes place may also limit formation of secondary structures at the ends of fragments. The major advantage of SLIC over Gibson assembly is cost, as T4 polymerase is much less expensive than the enzymes required for Gibson assembly.

SLIC is a standardized method for multi-fragment DNA assembly, and its low cost makes it ideal for researchers doing large amounts of cloning. Assembly is scarless, unlike Gateway cloning, and the method's flexibility allows it to be used with different types of PCR-generated inserts. By harnessing the power of DNA repair in *E. coli*, you can assemble multiple fragments without the need for specific restriction sites or DNA ligase!

### Further Reading

1. Li, Mamie Z., and Stephen J. Elledge. "Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC." *Nature methods* 4.3 (2007): 251-256. PubMed [PMID: 17293868](https://pubmed.ncbi.nlm.nih.gov/17293868/).
2. Li, Mamie Z., and Stephen J. Elledge. "SLIC: a method for sequence-and ligation-independent cloning." *Gene Synthesis: Methods and Protocols* (2012): 51-59. PubMed [PMID: 22328425](https://pubmed.ncbi.nlm.nih.gov/22328425/).

# CCDB-THE TOXIC KEY TO EFFICIENT CLONING

By Michael G. Lemieux | December 8, 2016

If you're into cloning, you're probably aware that there are several methodologies currently available for approaching it. These include the traditional [restriction enzyme/ligase-mediated method](#), the more recently developed [Gibson Assembly Cloning](#) and [Gateway® cloning](#) technologies, as well as several others. Each method is unique and relies on specific components that are key to the cloning reaction. Understanding the specific components is essential for choosing the correct cloning method for your own experiments, and here we will focus on a unique gene that makes the popular Gateway™ method possible: *ccdB*. But what is *ccdB*, what role does it play in modern cloning, and why should you learn more about it? Read on to find out how *ccdB* can make your cloning experiments a little easier.

One of the most time-consuming aspects of traditional cloning is the identification of clones that actually contain your insert of interest. Simply stated, *ccdB* makes cloning easier by selecting against vectors that did not take up your insert. But exactly how does *ccdB* accomplish this? Let's start with a brief history of the gene and how molecular biologists have harnessed it to evolve cloning technology.

## A Potent Toxin...

The *ccdB* gene, located on the F sex factor plasmid of *E. coli*, is part of a toxin-antitoxin system encoded by the *ccd* operon, which is responsible for plasmid maintenance during cell division. *ccdB* codes for the toxic protein (CcdB) that acts as a DNA gyrase poison, locking up DNA gyrase with broken double stranded DNA and ultimately causing cell death. *ccdA*, another gene found in the *ccd* operon, codes for the antitoxin protein (CcdA) that protects the cell against the toxic CcdB. Cells that lose *ccdA* through the loss of the F plasmid, succumb to the toxicity of CcdB.

## Becomes a Powerful Cloning Tool

Molecular biologists first saw the potential of this system for enhancing cloning efficiency about 20 years ago and developed cloning vectors to harness it. These vectors, called pKIL18 and pKIL19, contained the *ccdB* gene in frame with an MCS. *E. coli* that were transformed with the empty vectors expressed the *ccdB* gene and were therefore unable to propagate because CcdA wasn't available to counteract the toxin. If, however, an investigator were to successfully clone an insert into the vector, the *ccdB* reading frame would be disrupted allowing cells expressing the recombinant plasmid to propagate. Any cells that contained non-recombinant vectors (re-ligated empty vectors, for instance) would still express *ccdB* and therefore would die. This procedure dramatically reduces the number of clones that do not contain the recombinant plasmid and therefore makes the cloning process much more efficient, as one does not have to thoroughly screen colonies for the insert.

Gateway® technology (developed by [Invitrogen™](#)) is essentially a more modern version of this older system, with added advantages that are discussed in the section on [Gateway cloning](#). Focusing on the *ccdB* aspect,

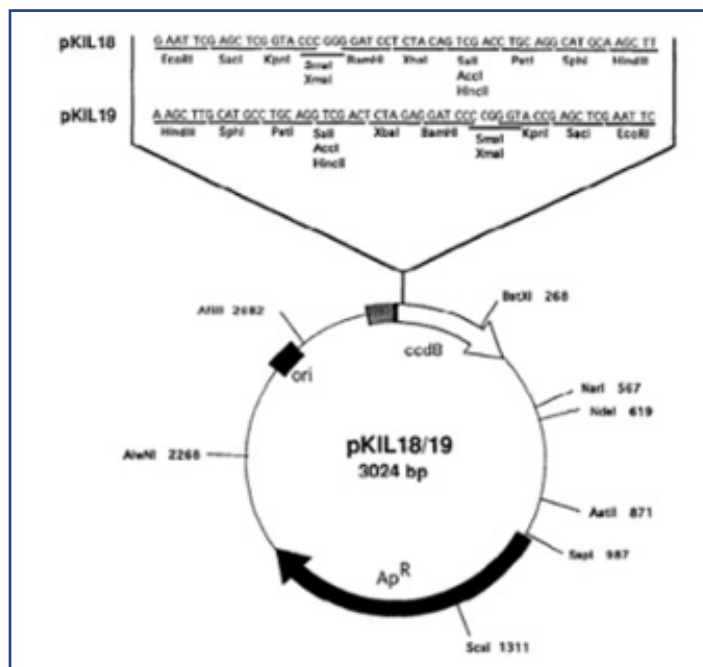


Image from Bernard, P., et al. showing the pKIL18/19 vectors and demonstrating the concept of restriction enzyme-mediated disruptions of *ccdB*, leading to positive identification of desired clones.

## CCDB-THE TOXIC KEY TO EFFICIENT CLONING (CONT'D)

Gateway takes advantage of the same principle that cells will not propagate while expressing the gene. Briefly, the vector “backbone” in this system contains *ccdB*. A successful insertion will completely replace *ccdB* with the investigator’s insert of interest. Hence correct clones are identified much more efficiently, as those that do not contain the desired insert should not grow.

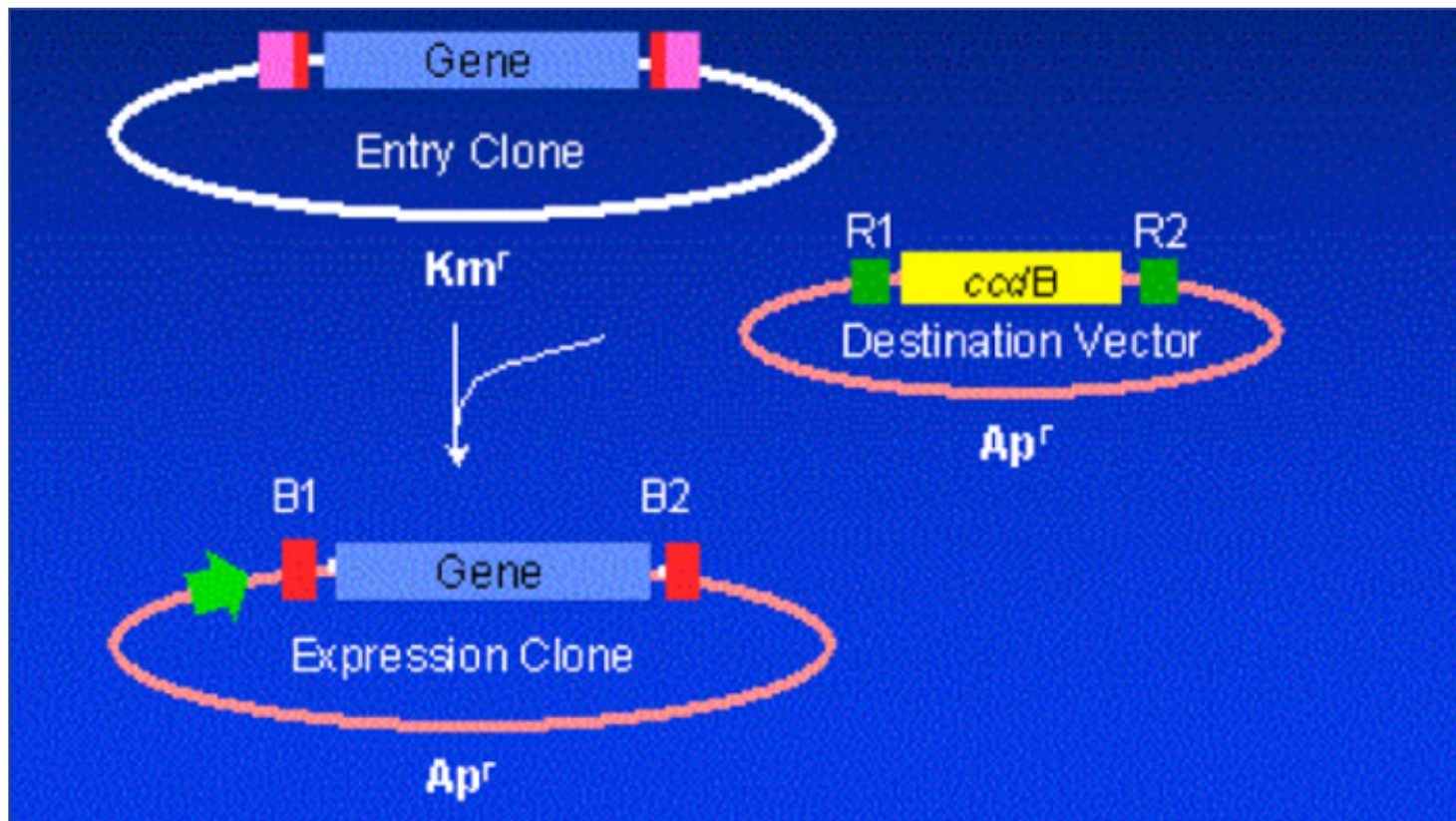


Image from <http://teachline.ls.huji.ac.il/72682/tutorials/gateway/introduction.html> demonstrating the principle behind the “swapping out” of *ccdB* with the investigator’s gene of interest.

### CcdB Resistant *E. coli* Strains Complete the System

But *ccdB* alone isn’t the only key to the system – how can one work with a plasmid/gene that kills the cells expressing it? The answer lies in special strains of *E. coli* that tolerate the expression of the toxin gene. One such strain is [DB3.1](#), which contains a mutant version of DNA gyrase (*gyrA462*) that is resistant to the toxic effects of CcdB. Another commercially available, CcdB-resistant strain is [ccdB Survival<sup>TM</sup>](#) from Invitrogen<sup>TM</sup>. Using either DB3.1 or *ccdB* Survival<sup>TM</sup>, one can efficiently propagate and prep plasmids containing the *ccdB* gene, which can then be used for downstream cloning applications. While these two strains ultimately perform the same function, there is some evidence that *ccdB* Survival<sup>TM</sup> can be more difficult to transform, depending upon the specific plasmids being used.

We would also like to note that while DB3.1 and *ccdB* Survival strains have been developed specifically with *ccdB*-containing vectors in mind, any plasmid that contains the F plasmid (F’ strains) will also be resistant to CcdB, as the native *ccdA* will be present. Most common cloning strains of *E. coli* do not contain the F plasmid (and are considered F<sup>-</sup>); however, there are a few popular lab strains, such as NEB Stable, JM109, XL1 Blue or XL10 Gold, that are F<sup>+</sup>. These strains could possibly be used to propagate and prep your *ccdB*-containing empty

## CCDB-THE TOXIC KEY TO EFFICIENT CLONING (CONT'D)

---

backbones but should never be used when selecting for recombinant plasmids. Check out the chapter on [Common Lab \*E. coli\* Strains](#) for more strain info.

### Additional Cloning Resources

As cloning continues to evolve, understanding all of the players becomes ever more important to ensure that you choose the most appropriate method for your own experiments. Please check out Addgene's [Choosing a Molecular Cloning Technique](#) page and the [Addgene Blog](#) for a variety of useful resources and protocols! We invite feedback from all members of the research community to help us develop these resources further. Feel free to offer your comments and suggestions, and even to [submit a guest blog entry of your own!](#)

### Further Reading

1. Bernard, P. "Positive Selection of Recombinant DNA by CcdB." *Biotechniques*. 1996 Aug;21(2):320-3. PubMed [PMID: 8862819](#).
2. Bahassi, EM., et al. "F plasmid CcdB killer protein: ccdB gene mutants coding for non-cytotoxic proteins which retain their regulatory functions." *Mol Microbiol*. 1995 Mar;15(6):1031-7. PubMed [PMID: 7623659](#).
3. Bernard, P., et al. "Positive-selection vectors using the F plasmid ccdB killer gene". *Gene*. 1994 Oct 11;148(1):71-4. PubMed [PMID: 7926841](#).
4. Bernard, P., et al. "The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase." *J Mol Biol*. 1993 Dec 5;234(3):534-41. PubMed [PMID: 8254658](#).
5. Bernard, P. and Couturier, M. "Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes." *J Mol Biol*. 1992 Aug 5;226(3):735-45. PubMed [PMID: 1324324](#).
6. Tam, JE. and Kline, BC. "Control of the ccd operon in plasmid F." *J Bacteriol*. 1989 May;171(5):2353-60. PubMed [PMID: 2651399](#). PubMed Central [PMCID: PMC209908](#).

# GATEWAY CLONING

By Maria Soriano | January 12, 2017

When facing a cloning project, scientists are no longer limited to traditional [restriction enzyme cloning](#). Instead, they can choose a molecular cloning technique that will work well with a given set of resources, time, and experimental needs. Since its invention in the late 1990s, Gateway cloning technology has become very popular as a rapid and highly efficient way to move DNA sequences into multiple vector systems. With the appropriate entry and destination vectors, one can use Gateway to clone a gene of interest into a variety of expression systems. Keep reading to learn more about the Gateway cloning method and its advantages.

## An Introduction to Gateway Technology

The Gateway cloning method, developed by [Invitrogen](#), is an *in vitro* version of the integration and excision recombination reactions that take place when lambda phage infects bacteria. *In vivo*, these recombination reactions are facilitated by the recombination of attachment sites from the phage (attP) and the bacteria (attB). As a result of recombination between the attP and attB sites, the phage integrates into the bacterial genome flanked by two new recombination sites (attL-left- and attR-right-, Figure 1). Under certain conditions, the attL and attR sites can recombine, leading to the excision of the phage from the bacterial chromosome and the regeneration of attP and attB sites.

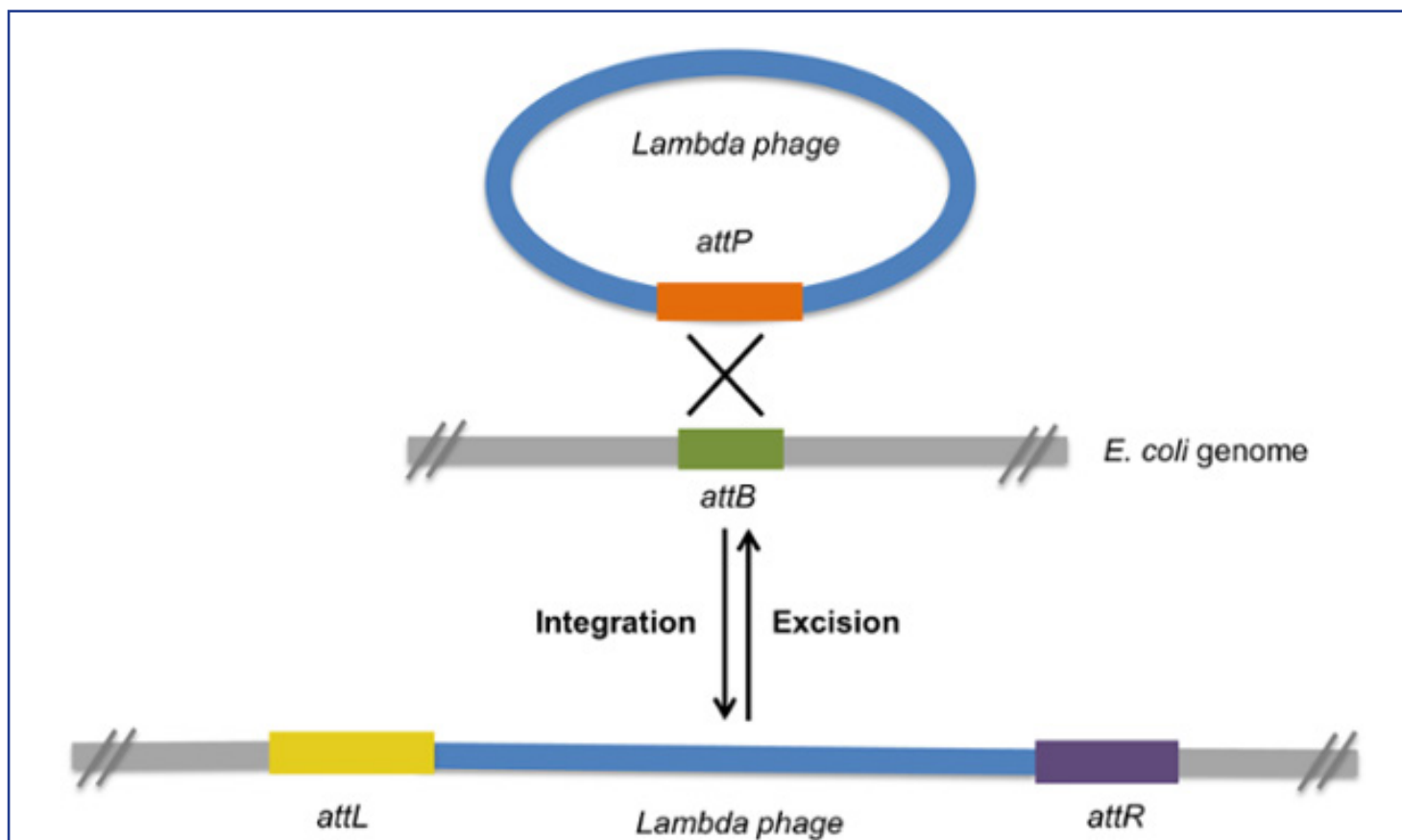


Figure 1: Lambda phage integration and excision reactions. Recombination of *attP* and *attB* sites creates *attL* and *attR* sites.



## GATEWAY CLONING (CONT'D)

Gateway vectors contain modified versions of the att sites so that scientists can easily clone in their desired DNA sequences. Gateway technology relies on the two reactions described below:

The BP Reaction takes place between the attB sites flanking the insert and the attP sites of the donor vector. This reaction is catalyzed by the BP Clonase enzyme mix and generates the entry clone containing the DNA of interest flanked by attL sites. As a byproduct of the reaction, the [ccdB gene is excised from the donor vector](#).

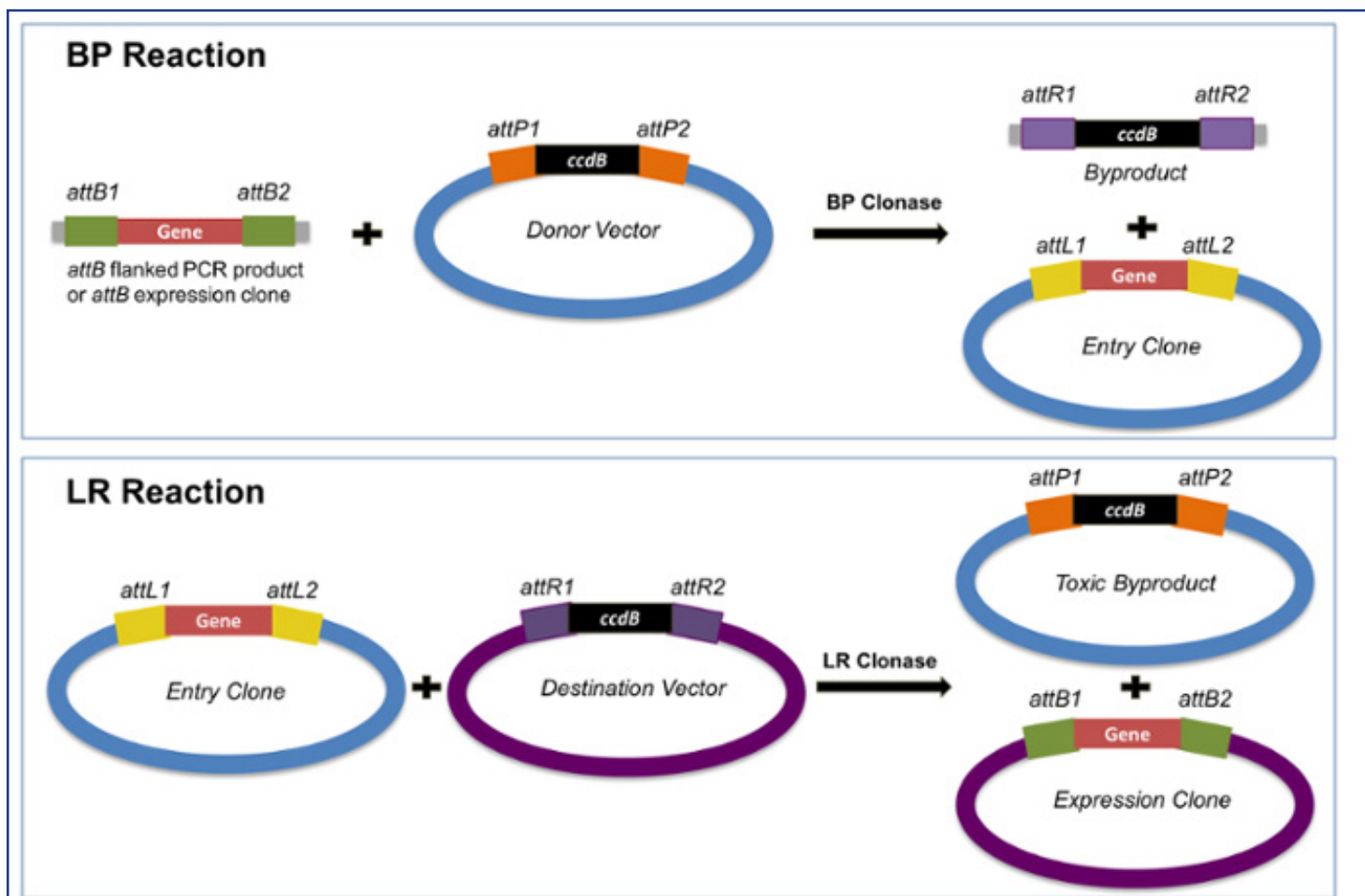


Figure 2: The Gateway system adapts phage integration into the BP and LR reactions. The BP reaction creates an attL-flanked entry clone. The LR reaction creates an expression clone with all of the components necessary for gene expression.

The LR Reaction takes place between the attL sites of the generated entry clone and the attR sites of the destination vector. This reaction is catalyzed by the LR Clonase enzyme mix. As a result, an expression clone with the DNA of interest flanked by attB sites is generated. As in the BP reaction, a DNA fragment containing the *ccdB* gene is excised from the destination vector.

Once the BP and/or LR reactions are performed, the next step is to transform competent *E. coli* cells and select the positive clones. The entry clone and destination vector carry different antibiotic resistance markers (indicated here by plasmid color), allowing you to easily select for the expression clone. You will also need to use a *E. coli* strain sensitive to CcdB (e.g. DH5 $\alpha$ , TOP10, Mach1). The *ccdB* gene is present in the donor vectors and the destination vectors prior to recombination, and it is exchanged with the gene of interest during the BP or LR reactions. Since the CcdB protein inhibits the growth of CcdB sensitive *E. coli* strains, most colonies should

## GATEWAY CLONING (CONT'D)

contain the desired, recombined construct. Read the previous section on [CcdB](#) for more information.

### How to Clone Using Gateway Technology

To better understand the process, we'll walk through an example experiment where we might use Gateway cloning to generate our desired constructs: lentiviral expression of the human KRAS gene in mammalian cells.

#### STEP 1: Generate an Entry Clone

There are a few different ways to generate our desired entry clone - human KRAS flanked by attL sites.

Method A: recombination of an attB-PCR product or plasmid with an attP donor vector. In this case, we would use PCR to add attB sites to either end of the KRAS coding sequence. If you choose this strategy, it's important to include the proper protein expression elements (ribosome recognition sequences, start codon, stop codons, reading frame considerations, etc). [This video](#) demonstrates how to use the [Snapgene program](#) to design Gateway plasmids.

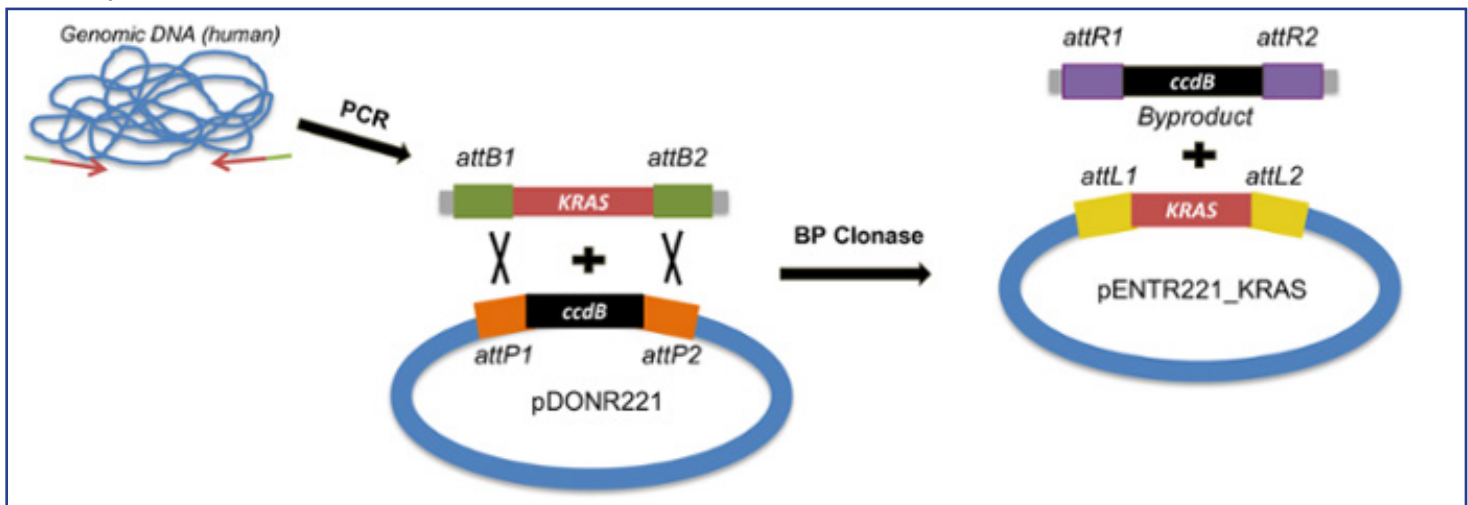


Figure 3: Method A to create an entry clone: recombination of an attB-flanked PCR product with an attP-containing donor vector.

Method B: [TOPO-cloning](#) of the desired insert into an attL-entry-TOPO vector. TOPO cloning adds short end(s) to facilitate cloning into an attL-containing entry vector.

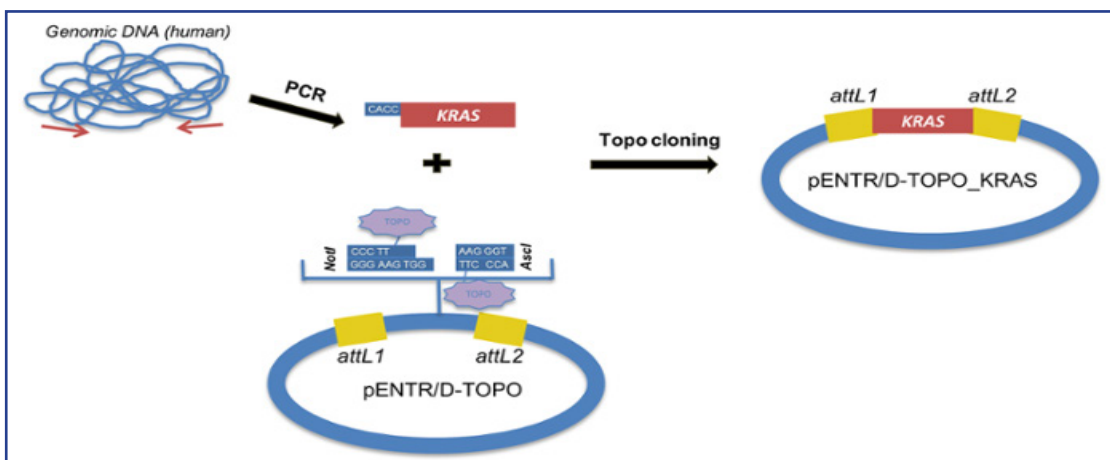


Figure 4: Method B to create an entry clone: TOPO cloning the insert into an attL-containing entry vector

## GATEWAY CLONING (CONT'D)

Method C: [Restriction cloning](#) of a restriction enzyme fragment containing the DNA of interest and an attL-entry vector. This fragment is inserted in a multiple cloning site (MCS) of an attL-containing entry vector.

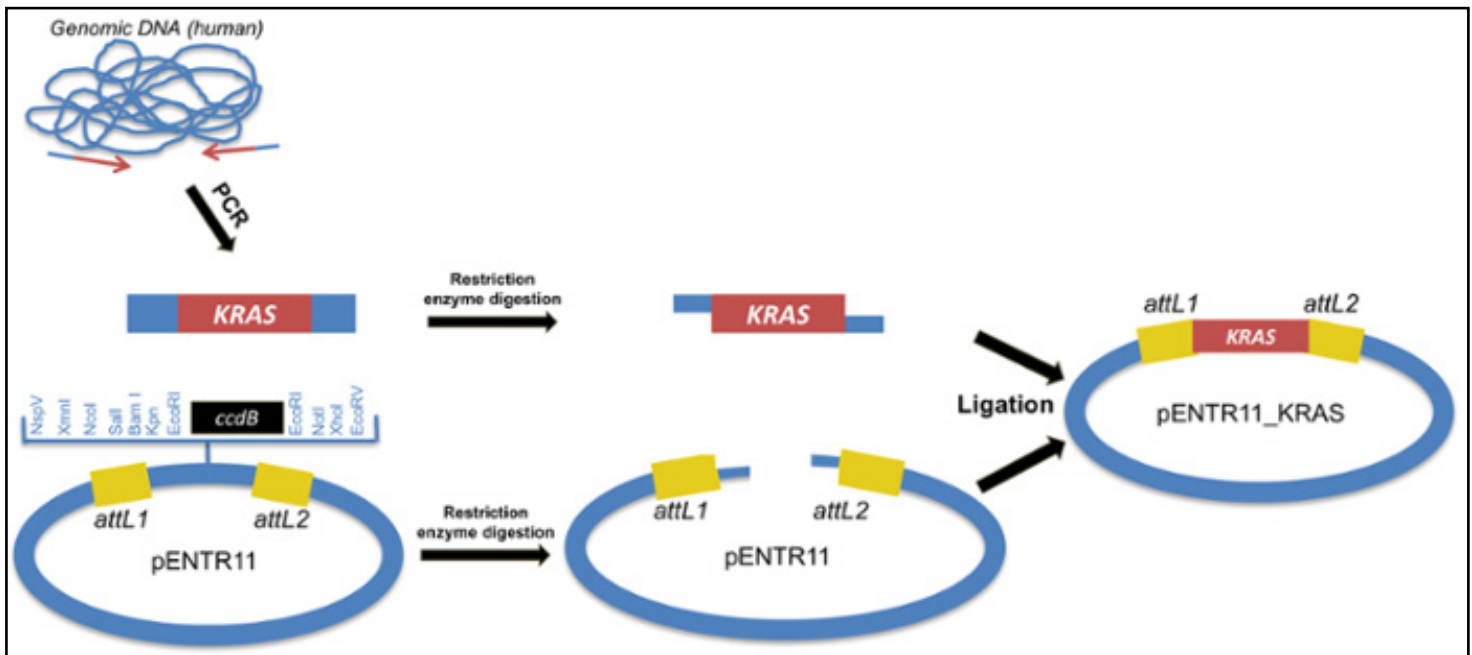


Figure 5: Method C to create an entry clone: Restriction clone the insert into an attL-containing entry vector.



**Pro Tip:** Addgene also has ready-made entry clones available for many popular genes, including [Hs.KRAS4a](#). Use our website to search for your favorite gene!

### STEP 2: Generate an Expression Clone

When making the expression clone, it is important to choose the destination vector that best fits your experiment. This choice will depend on a number of factors, like your organism, desired expression level, and experimental purpose. For mammalian lentiviral expression, we could use a vector like [pLenti CMV Puro DEST \(w118-1\)](#) or the doxycycline-inducible [pLIX\\_402](#). The chosen attR destination vector will recombine with the attL-entry clone to create the expression clone.

### STEP 3: Express your Gene of Interest!

Be sure to verify the integrity of your expression clone [via sequencing or restriction digest!](#) Then, you can transform or transfect the cells that you want to use for your experiments and verify that your construct is functional.

## Advantages of the Gateway Cloning Method

**Compatibility and flexibility:** Once you generate the entry clone with your DNA sequence of interest, you can move this DNA fragment across any expression system in just one recombination step. Addgene's ready-made entry clones can be used with a large variety of plasmids.

## GATEWAY CLONING (CONT'D)

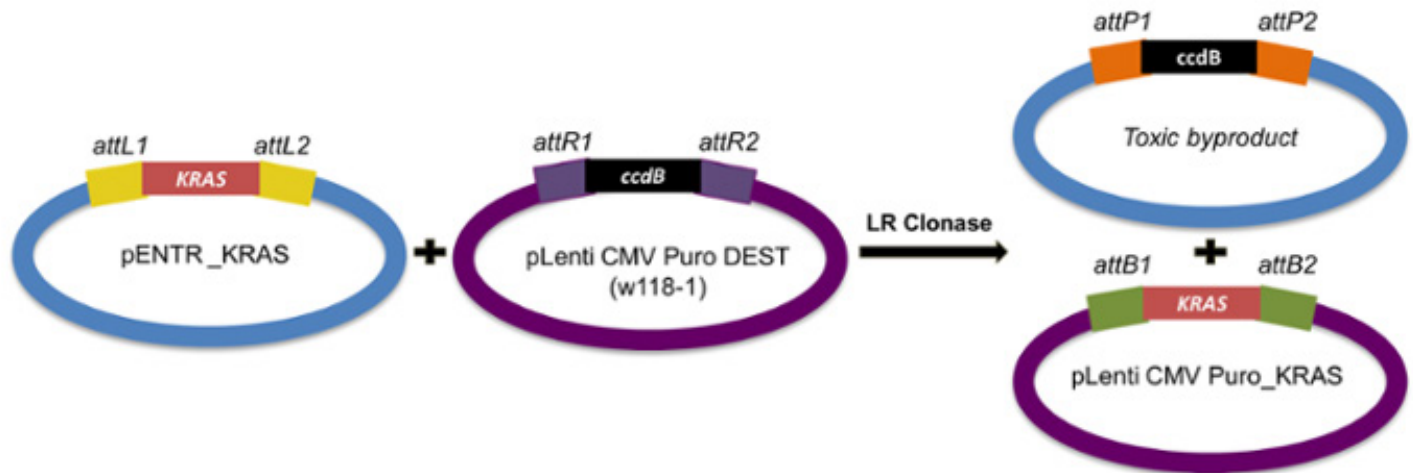


Figure 6: Generating an expression clone. The reaction between the entry clone and destination vector creates two products: the desired expression clone and a byproduct containing the *ccdB* gene. Since the *ccdB* product is toxic to the cell, Gateway cloning efficiency can reach >99%.

**Speed:** The Gateway system enables the generation of the expression construct in only 1 day, as opposed to 2+ days with traditional restriction and ligation cloning. It is also possible to set up the BP and LR reactions in the same tube, speeding up the cloning of the attB-PCR products directly into destination vectors. The cloning process is simple - no restriction, ligation or gel purification steps are required!

**Multiple fragment cloning:** You can use Gateway cloning to insert multiple DNA fragments into many vectors at once in the same tube. You can clone up to 4 DNA fragments, in a specific order and orientation, in one tube, into one Gateway vector to produce the desired expression clone. This is possible thanks to the Gateway vectors' design. They have modified versions of the attB, P, L and R sites that recombine very specifically and directionally: attB1 sites react only with attP1 sites; attB2 only with attP2, attL1 only with attR1; attL2 only with attR2, and so on. Take a look at some of the Gateway Multisite plasmids available at Addgene, including the Frew Lab [Multiple Lentiviral Expression Systems \(MuLE\) Kit](#), the [MultiSite Gateway cloning kit](#), and [MultiSite Gateway plasmids](#).

**Constant reading frame:** When you move a DNA fragment from one Gateway vector to another, the inserted DNA fragment stays in frame.

**High efficiency:** The positive (antibiotic) and negative (*CcdB*) selection markers used for Gateway Cloning can increase cloning efficiency to >99%.

**Universality:** All types of DNA fragments may be cloned: PCR fragments, cDNA or Genomic DNA and is available for all kind of organisms from mammals to *E. coli*.

**Ready to try out Gateway cloning?** Many scientists around the world have generated and deposited their own Gateway-compatible plasmids with Addgene. These can be used to express genes in a variety of model organisms. Use the links below to find Gateway plasmids for your organisms of interest:

- [Gateway plasmids for mammalian expression](#)
- [Gateway plasmids for bacterial expression](#)
- [Gateway plasmids for yeast expression](#)

## GATEWAY CLONING (CONT'D)

- [Gateway plasmids for plant expression](#)
- [Gateway plasmids for worm expression](#)
- [Gateway plasmids for insect expression](#)

### Glossary of Gateway Cloning Vectors

Vector Type	Vector Features	Purpose
Donor vector	attP sites for recombination; ccdB gene for negative selection	Used to clone attB-flanked genes of interest to generate <b>entry clones</b>
Entry vector	attL sites for recombination	Used to generate <b>entry clones</b> by <a href="#">TOPO cloning</a> or by <a href="#">Restriction Cloning</a>
Destination vector	attR sites for recombination; ccdB gene for negative selection; elements to express the gene of interest in the appropriate system	Recombines with the entry clone to generate an <b>expression clone</b>

### Further Reading

1. Chee JY, Chin CF (2015) Gateway Cloning Technology: Advantages and Drawbacks. Clon Transgen 4:138. doi:10.4172/2168-9849.1000138
2. Hartley JL. Use of the Gateway System for Protein Expression in Multiple Hosts. Curr Protoc Protein Sci. 2003 Feb;Chapter 5:Unit 5.17. PubMed [PMID:18429245](#).
3. Ptashne, M. (1992). [A Genetic Switch: Phage \(Lambda\) and Higher Organisms](#) (Cambridge, MA: Cell Press).

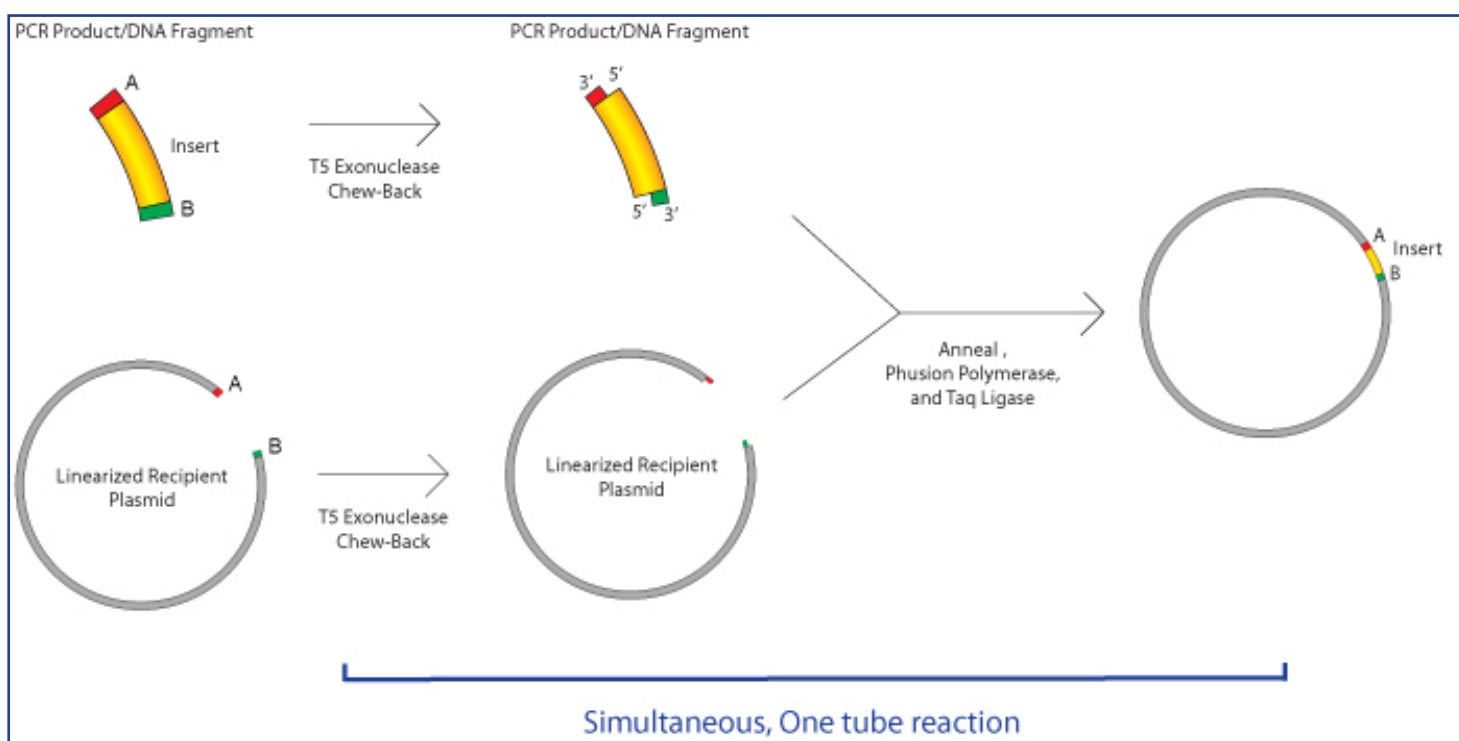
# GIBSON ASSEMBLY

By Brook Pyhtila | March 1, 2016

Over the past decade, scientists have developed and fine tuned many different ways to clone DNA fragments which have provided appealing alternatives to [restriction enzyme cloning](#). These newer technologies have become more and more common, and for good reason. They offer many advantages over the traditional restriction enzyme cloning we once relied exclusively on. In this blog post, I will go over some advantages, disadvantages, and examples of how scientists are using Gibson assembly to put together DNA fragments.

As a fun way to start, I highly recommend watching this [entertaining video](#) created by our friends on the [Cambridge 2010 iGEM](#) team that describes the basics of Gibson assembly as a parody of “Breakfast at Tiffany’s.”

## Overview of Gibson Assembly



The Gibson assembly technique was first described by Dr. Daniel Gibson and colleagues at the J. Craig Venter Institute in 2009. Here at Addgene we added this option to our drop down menu of common cloning options in the [deposit process](#) in 2014 because of its gain in popularity. Gibson assembly is well known for allowing easy assembly of multiple linear DNA fragments, but can also be used in basic cloning of an insert into your vector of choice as shown in the figure below. To start, you need to have DNA fragments with regions of homology at their ends, which are typically created by PCR. Then, the fragments are incubated together with an enzyme master mix, which contains three different enzymes:

- an exonuclease, which chews back the 5' ends of the fragment, generating long overhangs which allows the single stranded regions with homology to anneal
- a polymerase, to fill in the gaps
- a DNA ligase, to seal the nicks of the annealed and filled-in gaps

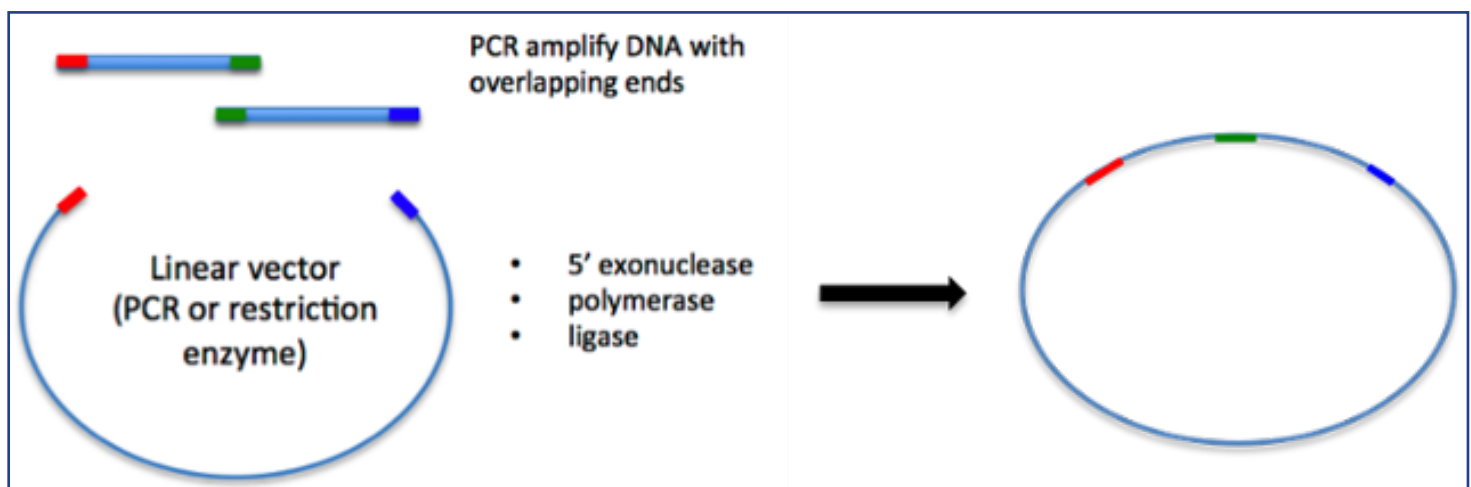
## GIBSON ASSEMBLY (CONT'D)

The great part about this mix of enzymes is that they can all work at the same temperature, so the entire reaction takes an hour or less to complete at 50 °C. After an hour or so, the sample is immediately ready to transform into competent cells. The master mix of enzymes can be purchased from a company (e.g. NEB or SGI-DNA), or can be mixed yourself (e.g. see the [Miller Lab Protocol](#)). The Gibson assembly process can be used to assemble up to 6 fragments in one step, resulting in scar-free assembly that does not require the presence of specific restriction sites (or lack thereof) nor a serious time commitment. Another advantage is that this process makes it easy to generate wild type and mutant constructs at the same time, rather than sequentially.

### How It Works

The required homology between neighboring fragments can be created via PCR amplification with primers that contain the appropriate homologous sequences. [NEB recommends](#) an overlap of 15-40 bp, with a primer melting temperature greater than 48 °C. Both Snapgene and NEB have tools that help you design primers for PCR amplification of fragments to incorporate such regions of homology. [This video](#) gives a helpful demonstration of how to use Snapgene's program to design primers for Gibson Assembly.

For a simple example of using Gibson assembly, imagine that you want to insert your gene of interest into a vector with a large [tag](#) at the N-terminus, but you don't have the tag already included in the vector you want to use. If you inserted these two pieces of DNA by restriction enzyme cloning, you would have to do it in two steps, and a scar would likely remain between the two fragments. However, by using Gibson assembly, you can insert both the gene of interest and the tag sequences into the vector in one step without scars as depicted below. First, you need to design primers to amplify the two fragments while also including regions of homology to the vector or neighboring fragment. Then you would amplify the fragments and vector by PCR, verify that you have bands of the correct size, and purify the DNA fragments. Finally, you simply incubate these three fragments together along with the Gibson assembly master mix for 1 hour, and then [transform into competent cells](#). The success rate of this reaction is typically fairly high, so there is usually no need to screen a large amount of colonies. Excluding the time it takes to get primers, you can have your construct made in 5 days.



One drawback to the Gibson assembly technique is that the process works best with fragments over 200 nucleotides. This may be because the exonuclease could chew through an entire fragment shorter than 200 nucleotides before the annealing and polymerization steps could occur. Secondly, it does not work well if the ends of the fragments have stable single stranded DNA secondary structure, such as a hairpin or a stem loop (as might be anticipated to occur within a terminator sequence), since this would directly compete with the required single-stranded annealing and priming of neighboring assembly fragments.

## GIBSON ASSEMBLY (CONT'D)

---

Gibson assembly is commonly used in synthetic biology, mainly because of the ease in assembling multiple fragments in one step with no scar sequences remaining in the final product. The long overlapping region between fragments also better ensures correct assembly order of the fragments as compared to the smaller overlapping sequences left with a restriction digest. A study in 2013 found that Gibson assembly was one of the most commonly used assembly methods ([Kahl 2013](#)). However, Gibson assembly isn't ideal for synthetic biology standards, which rely up on the re-use of parts between experiments. In Gibson assembly, long primers for each fragment must be designed and ordered, and are specific for each fragment as well the fragment you want next to it, so this doesn't allow for mix and match of many different fragments. One way around this is to use a combination of standard sequences with overlap regions, such as in the case of [MODAL](#) (modular overlap-directed assembly with linkers) which brings modularity to overlap-directed methods ([Casini 2014](#)). In this case linker sequences are added before the homology regions, which allows for the mixing and matching of parts.

### Gibson Assembly Meets CRISPR

Gibson can be adapted to more complicated cloning schemes, such as those where the vector that you want to use is very large, has a high GC content, contains a lot of repeats--any of which could make the PCR step difficult--OR there is no convenient restriction site for linearization. This is a perfect case for the use of Gibson assembly combined with the popular [CRISPR technology](#) and is described in a recent publication by the Lockey lab ([Wang, et al. 2015](#)). In this case, rather than using a restriction enzyme or PCR to make linearized vector, the Cas9 enzyme along with a specific gRNA was used to cleave the 22kb vector. When followed by the standard Gibson assembly techniques described above, this resulted in the direct and seamless cloning into a vector where no other methods were available. A second example of using Gibson Assembly along with CRISPR was also described recently ([Jiang et al. 2015](#)), in which very large fragments of the bacterial chromosome (up to 100 kb) were specifically cut out via CRISPR and then assembled into a vector using Gibson assembly. In this case the vector was PCR amplified to contain regions of homology with the bacterial chromosome fragments. CRISPR cutting was used to circumvent the need to PCR amplify the fragments of the chromosome, which was technically challenging.

### Other Homology Based Technologies

We've described [Sequence and Ligation Independent Cloning \(SLIC\)](#) in a previous section. Although SLIC may be more cost effective, Gibson assembly improves on two aspects of the SLIC methods. First, it uses a dedicated 5' exonuclease instead of using the exonuclease feature of T4 DNA polymerase, which must be controlled by the presence or absence of dNTPs. Secondly, in Gibson assembly a ligase is added to repair the nicks *in vitro*, whereas in SLIC these constructs are repaired *in vivo*, which ends up being much less efficient.

In addition to SLIC and Gibson, there are yet more homology based assembly methods that have been described--CPEC (circular polymerase extension cloning) and SLiCE (Seamless Ligation Cloning Extract) to name two more. Likewise, there are several store-bought cloning kits available which are all based on long-overlapping regions, have no requirements for restriction enzymes, and no scar sequences between fragments. Some of these products include:

- [GeneArt® Seamless Cloning and Assembly \(Thermo Fisher Scientific\)](#)
- [NEBuilder® HiFi DNA Assembly \(NEB\)](#)
- [Cold Fusion Cloning \(System Biosciences\)](#)
- [In-fusion Cloning \(Clontech\)](#)



## GIBSON ASSEMBLY (CONT'D)

---

Although these kits may come at a high price, their manufacturers tout great efficiency and low time commitments. These kits also come with specific protocols, suggestions for ratios of product to insert, and tools for primer design, so it's always best to check in with the instructions from the particular manufacturer of the kit you will be using. Some of these products may offer advantages that Gibson assembly does not such as increased efficiency, shorter incubation times, or the ability to accommodate smaller fragments.

### Know Your Cloning Methods

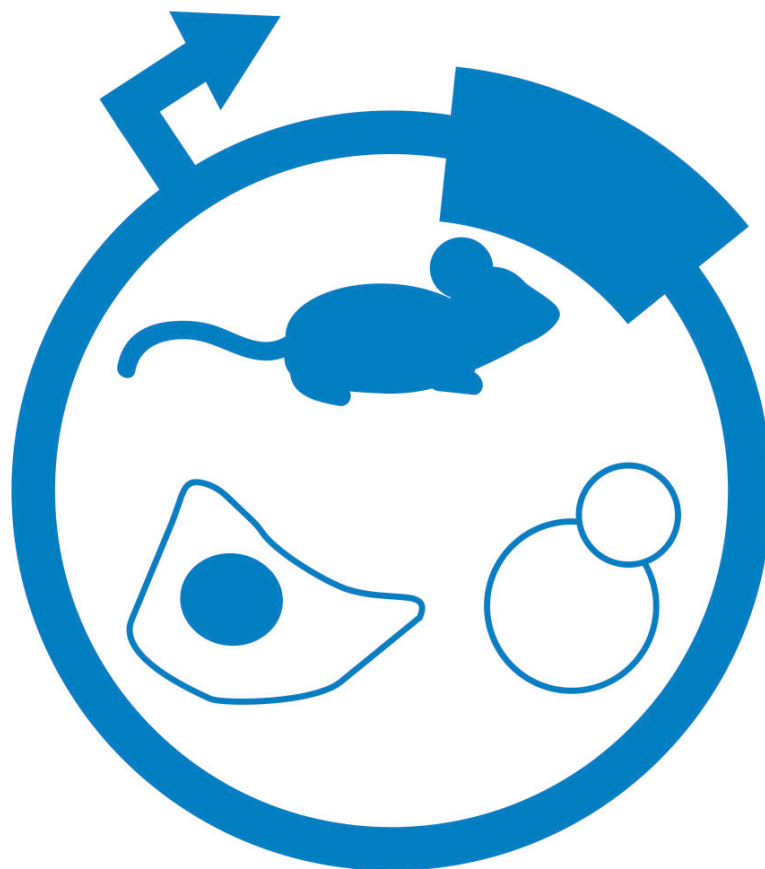
There are two other common cloning methods that are easy to get mixed up with Gibson assembly (they all start with a G!) but actually work by significantly different methods. [Golden Gate cloning](#) does result in the seamless joining of fragments, but uses site specific restriction sites (Type IIS restriction endonucleases) to cleave DNA outside of the recognition sequence. This requires that the vectors and DNA fragments contain these sites at the correct location and NOT in the middle of your insert. Gateway cloning utilizes  $\lambda$  integrase to catalyze directional cloning of DNA parts that are flanked by orthogonal versions of the attB and attP sites recognized by the integrase. This method requires specialized vectors containing these integration site and leaves scars between fragments, but allows for easy movement of a DNA fragment from one vector to another.

There are lots of different ways to clone these days. Gibson and the other long-homology based cloning methods are useful alternatives to the standard restriction/ligation, Gateway, or Golden Gate cloning methods. Whether it is for routine cloning, assembly of multiple fragments, or synthetic biology, you should consider giving it a try!

### Further Reading

1. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 2009; 6, 343-345. PubMed [PMID: 19363495](#).
2. Wang JW, Wang A, Li K, Wang B, Jin S, Reiser M, Lockey RF. CRISPR/Cas9 nuclease cleavage combined with Gibson assembly for seamless cloning. *BioTechniques* 2015; 58:161-170. PubMed [PMID: 25861928](#).
3. Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun.* 2015; Sep 1;6:8101. PubMed [PMID: 26323354](#). PubMed Central [PMCID: PMC4569707](#).
4. Kahl, L. J. & Endy, D. A survey of enabling technologies in synthetic biology. *J. Biol. Eng.* 2013; 7, 13. PubMed [PMID: 23663447](#). PubMed Central [PMCID: PMC3684516](#).
5. Casini, Arturo, et al. "One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy." *Nucleic acids research* 42.1 (2014): e7-e7. PubMed [PMID: 24153110](#). PubMed Central [PMCID: PMC3874208](#).

# CHAPTER 3: EUKARYOTIC EXPRESSION VECTORS



# MAMMALIAN VECTORS

By Marcy Patrick, Addgene | Mar 25, 2014

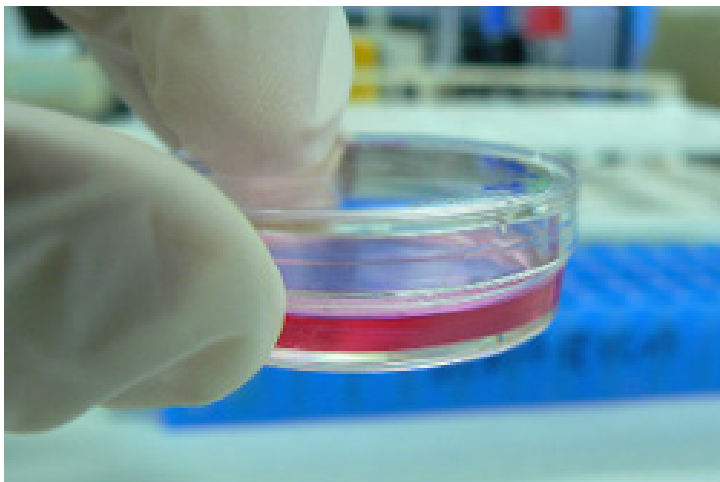


Image Source: Wikipedia, accessed March 14, 2014.  
Photo Author: kaibara87.

**Although plasmids do not naturally exist in mammals, scientists can still reap the benefits of plasmid-based research using synthetic vectors and cultured mammalian cells.** Of course, these mammalian vectors must be compatible with the cell type they are transfected into – a bacterial origin of replication (ORI) will not allow for plasmid replication in mammalian cells, for example, and a toxin that kills bacteria may not have any discernable effect on mammalian cells. So how do mammalian plasmids differ from their bacterial counterparts? How does replication occur? Is selection necessary for transfected cells?

## What is Transfection?

Before getting into the mammalian plasmid components, it may be useful to describe the means of introducing genetic material ([such as plasmids](#)) into mammalian cells, a process called [transfection](#). Transfection is somewhat comparable to bacterial transformation (the introduction of DNA into bacterial cells); however, the techniques and reagents vary. Plasmid transfection into mammalian cells is fairly straightforward and the resultant cells can either express the plasmid DNA transiently (similar to bacteria) or incorporate the genetic material directly into the genome to form a stable integration. Unlike bacterial transformation, scientists do not “select” for cells that have taken up the plasmid in the same way. Selection methods, described below, are typically employed when creating stable cell lines and are not used for general plasmid selection. Instead, reporter genes are often employed to easily monitor transfection efficiencies and expression levels in the cells. Ideally, the chosen reporter is unique to the cell, is expressed from the plasmid, and can be assayed conveniently. A direct test for your gene of interest may be another method to assess transfection success. [GFP is often used as a reporter](#). Click [here](#) to skip ahead in this eBook to fluorescent tagging.

## Transient Transfection and the Elusive “Mammalian ORI”

For many experiments, it is sufficient for the transfected plasmid to be expressed transiently. Since the DNA introduced in the transfection process is not integrated into the nuclear genome, in the absence of plasmid replication, the foreign DNA will be degraded or diluted over subsequent generations. This, however, may not be a problem depending on the duration or other parameters of your experiment. Mammalian cells double at a much slower rate than that of bacteria (~24 h vs 20 min, respectively). Therefore, it is not always mission critical to make sure the plasmid replicates in the cell, as many of these experiments are completed within 48 h of transfection.

Of course, it is possible that you may not want the plasmid depleted, but still want to use transient transfection methods. Since there are no “natural” mammalian ORIs, scientists have usurped viral-based ORIs to fill the void. These ORIs, however, require additional components expressed *in trans* within the cell for effective replication.

## MAMMALIAN VECTORS (CONT'D)

---

Cell lines expressing the Epstein–Barr virus (EBV) nuclear antigen 1 (EBNA1) or the SV40 large-T antigen (293E or 293T cells), allow for episomal amplification of plasmids containing the viral EBV or SV40 ORIs, respectively. The presence of these viral components greatly reduces the rate of plasmid dilution but does not guarantee 100% transfection efficiency.

### Stable Transfection

A stable transfection is used to create a population of cells that have fully and successfully incorporated foreign genetic material into their genomes. Unlike plasmids used for expression in yeast and bacteria, plasmids used for stable transfections rarely contain an ORI since the integrated DNA will be replicated as part of the genome. Because the foreign DNA becomes a permanent addition to the host genome, the cells will continually express the genetic traits of the foreign material and will subsequently pass it on to future generations. Stably transfected cells may be considered an entirely new cell line from that of the original parental cells.

### Positive Selection in Mammalian Cells

To achieve stable transfection, there should be a selective pressure to force cells to incorporate the plasmid DNA into the genome. For the purposes of this post, we will define positive selection as the means of picking up positive traits (i.e. the plasmid contains a cassette that will make cells resistant to a toxin), whereas negative selection would be the picking up of a negative trait (i.e. the plasmid contains a cassette that will make cells sensitive to a toxin). In the table below we focus on positive selection; however, negative selection techniques can be used in conjunction with positive selection to ensure your gene gets targeted to a specific location within the genome.

Positive selection in mammalian cells works similarly to that in bacteria and a table of the most commonly used selection markers can be found on the next page.

### Keep These Tips in Mind:

- There is not one recommended concentration for selection in mammalian cells. Before doing a transfection experiment, it is important to determine the proper concentration required for efficient selection. This is usually achieved by performing a “kill curve” (basically growing cells in various concentrations of the selection reagent). Cells should die within 3-5 days and resistant colonies appear in about 10-14 days depending on how quickly your cells divide.
- Gentamicin is often used as a supplement in mammalian cell culture to suppress bacterial growth, and is not appropriate for mammalian selection – do not confuse this with G418 (aka Geneticin).
- Neomycin should not be used for mammalian expression – instead use G418. This can be confusing since the neo/kan gene confers G418 resistance; however, like gentamicin, neomycin is typically used to suppress bacterial growth.
- The selection agent should not be added to culture media until 24-48 h post transfection when creating stable cell lines.

## MAMMALIAN VECTORS (CONT'D)

Positive selection in mammalian cells works similarly to that in bacteria and a table of the most commonly used selection markers is shown below:

**Table 1: Commonly Used Mammalian Selection Markers**

Name	Gene Conferring Resistance	Cell Types*	Mode of Action**	Working Concentration***
Blasticidin	<i>bsd</i>	HeLa, NIH3T3, CHO, COS-1, 293HEK	Inhibits termination step of translation	2-10 ug/mL
G418/Geneticin	<i>neo</i>	HeLa, NIH3T3, CHO, 293HEK, Jurkat T cells	Blocks polypeptide synthesis at 80S; inhibits chain elongation	100-800 ug/mL
Hygromycin B	<i>hygB</i>	HeLa, NIH3T3, CHO, Jurkat T cells	Blocks polypeptide synthesis at 80S; inhibits chain elongation	50-500 ug/mL
Puromycin	<i>pac</i>	HeLa, 293HEK, Jurkat T cells	Inhibits protein synthesis; premature chain termination	1-10 ug/mL
Zeocin	<i>Sh bla</i>	HeLa, NIH3T3, CHO, COS-1, 293HEK, Jurkat T cells	Complexes with DNA; causes strand scissions	100-400 ug/mL

\*Not comprehensive.

\*\* In eukaryotes.

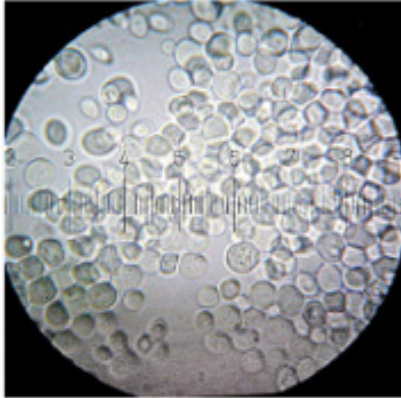
\*\*\*The concentration used for selection is typically more (double) than that used for maintenance of a transfected cell line.

[Looking for Mammalian Expression Plasmids? Check Out Addgene's Empty Backbone Resource.](#)

# YEAST VECTORS

By Marcy Patrick, Addgene | Feb 25, 2014

## Why Do Scientists Use Yeast Vectors?



Source: Wikipedia, accessed 2/23/2014.  
Author: Bob Blaylock

**Yeast are eukaryotes and thus contain complex internal cell structures similar to those of plants and animals.** Unlike bacteria, yeast can post-translationally modify proteins yet they still share many of the same technical advantages that come with working with prokaryotes. This includes but is not limited to: rapid growth, ease of replica plating and mutant isolation, a well-defined genetic system, and a highly versatile DNA transformation system. Unlike most other microorganisms, yeast have both a stable haploid and diploid state which is useful for genetic analysis, as well as an efficient mechanism of homologous recombination to facilitate simple gene replacement/mutation. Yeast expression plasmids used in the lab typically contain all the necessary components to allow shuttling between *E. coli* and yeast cells. To be useful in the lab, the vectors must contain a yeast-specific origin of replication (ORI) and a means of selection in yeast cells, in addition to the bacterial ORI and antibiotic selection markers.

## The Yeast Origin of Replication

**Please note:** This section primarily pertains to ORIs in budding yeast, *Saccharomyces cerevisiae*; however, we've also noted some features required for vector replication in fission yeast, *Schizosaccharomyces pombe*, at the end.

We have already covered how the regulation of bacterial ORIs determines plasmid copy number within the bacterial cell. Similarly, the specific ORI elements included within a [yeast vector](#) determine how the plasmid is replicated and maintained within the yeast cell. These elements control not only the number of plasmids found in each cell, but also whether the plasmid gets integrated into the host DNA or is independently replicated as an episome.

## The Four Main Types of Yeast Plasmids

- **Yeast Integrating plasmids (YIp):** These plasmids lack an ORI and must be integrated directly into the host chromosome via homologous recombination.
- **Yeast Replicating plasmids (YRp):** These vectors contain an Autonomously Replicating Sequence (ARS) derived from the yeast chromosome. As the name suggests, these vectors can replicate independently of the yeast chromosome; however, they tend to be unstable and may be lost during budding.
- **Yeast Centromere plasmids (YCp):** These are considered low copy vectors and incorporate part of an ARS along with part of a centromere sequence (CEN). These vectors replicate as though they are small independent chromosomes and are thus typically found as a single copy. Unlike the ARS vectors, CEN vectors are stable without integration.

## YEAST VECTORS (CONT'D)

---

- **Yeast Episomal plasmids (YEp):** These are most similar to bacterial plasmids and are considered “high copy.” A fragment from the 2 micron circle (a natural yeast plasmid) allows for 50+ copies to stably propagate per cell. The copy number of these vectors can also be controlled if specific regulatable elements are included ([reviewed here](#)).

Plasmids for use in *S. pombe*, on the other hand, do not require a well-defined ORI. Instead, the size and A-T content of the DNA (apparently independent of a known specific sequence) dictate the replication of these vectors. *S. pombe* plasmids oftentimes utilize an ARS to aid in high transformation efficiency; however, this region does not necessarily promote replication.

### Selection Markers for Yeast

Historically, scientists have utilized auxotrophic selection rather than antibiotic selection when working with yeast, due to high rates of spontaneously occurring resistant mutants and the insensitivity of yeast strains to some antibiotics. Auxotrophy is defined as the inability of an organism to synthesize a particular organic compound required for its growth. Many auxotrophic strains of yeast exist which can be easily maintained when grown on media containing the missing nutrient. Scientists can exploit these host mutations by including a copy of a functional gene which complements the host's auxotrophy. When grown on media NOT containing the nutrient, the host cells will die unless they have incorporated the plasmid carrying the required gene.

The table on the next page lists some of the most commonly used selection markers in yeast and provides the element needed to overcome the auxotrophy as well as additional uses for said element. This [link provides a more extensive list of yeast auxotrophic markers](#) and includes the associated references.

### Considerations When Using Auxotrophic Selection

There are some drawbacks to using auxotrophic markers as a means of selection:

- A specific selection marker needs to be used with a yeast strain deficient in that compound. Therefore known auxotrophic strain/selection element pairs must be utilized or a new combination needs to be created in advance of the experiment.
- The marker provided by the plasmid may be expressed at higher than normal physiological levels due to high copy numbers. This creates a potential metabolic burden on the yeast cells.
- Some phenotypes may be altered due to the presence of the selection marker at non-physiological levels.

Scientists have tried varied approaches to combat these issues. One method to reduce the amount of marker gene expression is to use a partially defective promoter to drive expression of the selection marker. This reduces the amount of gene product present in the cell, thus allowing the yeast to maintain higher copy numbers. Additionally, improvements in antibiotic selection have made utilizing the more traditional drug selection methods feasible in yeast as a complement or alternative to using auxotrophic markers. The [MX series of antibiotic resistance cassettes](#) is most commonly employed, with the [KanMX](#) being the most prevalent due to its versatility and ability to be used in bacterial, yeast, and mammalian cells.

## YEAST VECTORS (CONT'D)

### Table 1: Common Yeast Auxotrophic Markers

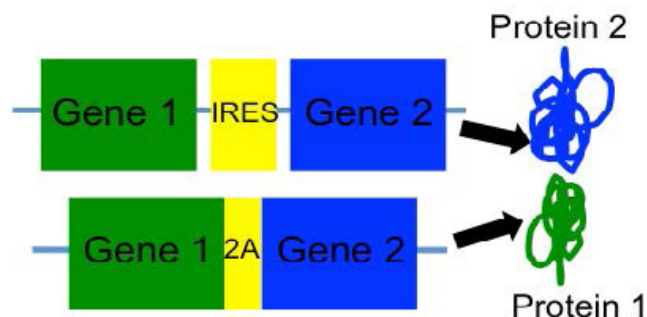
Yeast selection marker	What does the marker help synthesize?	Counterselection? (growth-based positive selection for the loss of the marker gene)	For use in	Can this be used in other yeast species?	Can this be used for auxotrophic selection in <i>E. coli</i> ?	Additional considerations
<i>HIS3</i>	L-histidine	no	<i>S. cerevisiae</i>	no	yes	
<i>URA3</i>	pyrimidine (uracil)	yes - Grow with 5-FOA.	<i>S. cerevisiae</i>	yes - This can complement <i>ura4-S. pombe</i> , but the complementation is weak.	yes	
<i>LYS2</i>	L-lysine	yes - Grow on plates containing alpha-aminoadipate in the absence of a nitrogen source.	<i>S. cerevisiae</i>	no	no	
<i>LEU2</i>	L-leucine	no	<i>S. cerevisiae</i>	yes - This can complement <i>leu1-S. pombe</i> , but multiple copies are required.	yes	
<i>TRP1</i>	L-tryptophan	yes - Grow with 5-FAA.	<i>S. cerevisiae</i>	no	no	<i>TRP1</i> alters some yeast phenotypes. This marker should not be used in gene disruption experiments.
<i>MET15</i>	L-methionine and overproduces hydrosulfide ions	yes - Grow with methylmercury.	<i>S. cerevisiae</i>	no	no	Can be used for color and growth selection if divalent lead ions are used in the growth media.
<i>ura4+</i>	pyrimidine (uracil)	yes - Grow with 5-FOA (see note to right).	<i>S. pombe</i>	no	no	FOA in fission yeast induces mutation in the <i>ura5+</i> gene in addition to <i>ura4+</i> .
<i>leu1+</i>	L-leucine		<i>S. pombe</i>	no	no	
<i>ade6+</i>	purine (adenine)		<i>S. pombe</i>	no	no	



# MULTICISTRONIC VECTORS

By Melina Fan, Addgene with contributions from Marcy Patrick, Addgene | Sep 8, 2014

**Co-expression of multiple genes is valuable in many experimental settings.** To achieve this, scientists use a multitude of techniques including co-transfection of two or more plasmids, the use of multiple or bidirectional promoters, or the creation of bicistronic or multicistronic vectors. Unlike promoters which will create unique mRNA transcripts for each gene that is expressed, multicistronic vectors simultaneously express two or more separate proteins from the same mRNA. We discussed eukaryotic [promoters](#) in chapter 1, so in this section we'll cover basics of multicistronic vectors: why they are useful, how they work, and how to get started with them.



## Why Use Multicistronic Vectors?

Detecting cells that are expressing your gene, especially if you are studying a novel gene, is not always a straightforward process. Rather than try to directly detect a gene of interest, scientists have instead developed novel methods to co-express a gene along with a reporter, such as a fluorescence gene or a resistance gene. These reporters allow you to easily screen or select for cells that are expressing your gene of interest at high levels. Unlike vectors that express screenable or selectable markers from a unique promoter, multicistronic plasmids ensure that any cells that are positive for your marker should also be expressing your gene as they are both derived from the same transcript.

Of course multicistronic vectors do not have to exclusively be used as a means of detection; they are useful almost anytime you want to express multiple genes in the same cell. Although it is possible to drive co-expression by using a plasmid with multiple, individual expression cassettes, having the genes expressed from the same cassette is sometimes advantageous, particularly when only a portion of the plasmid is packaged for [viral delivery](#), or the relative expression levels between two or more genes is important.

## How Do Multicistronic Vectors Work?

Scientists have “borrowed” some tricks discovered in positive single-stranded RNA viruses to allow for the efficient translation of multiple genes from a single transcript. The two strategies most widely incorporated into plasmids for research purposes are described below.

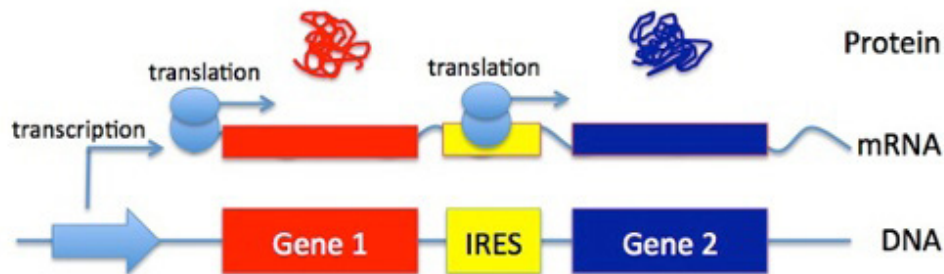
- **IRES Elements**

Translation in eukaryotes usually begins at the 5' cap so that only a single translation event occurs for each mRNA. However, some bicistronic vectors take advantage of an element called an Internal Ribosome Entry Site (IRES) to allow for initiation of translation from an internal region of the mRNA.

In the figure below, you can see that the IRES element acts as another ribosome recruitment site, thereby resulting in co-expression of two proteins from a single mRNA.

The IRES was originally discovered in poliovirus RNA, where it promotes translation of the viral genome in eukaryotic cells. Since then, a variety of IRES sequences have been discovered - many from viruses, but also

## MULTICISTRONIC VECTORS (CONT'D)



some from cellular mRNAs. What they have in common is the ability to spark translation initiation independent of the 5' cap.

IRES elements are very useful and commonly found in bicistronic vectors; however, they do have some disadvantages. These elements are quite large (500-600 bp) and may take up precious space in viral transfer vectors with limited packaging capacity. Additionally, it may not be feasible to express more than two genes at a time using IRES elements. Further, scientists have reported lower expression of the downstream cistron due to factors such as the experimental cell type and the specific genes cloned into the vector.<sup>3</sup>

### • 2A Peptides

To overcome some of the disadvantages of the IRES element, scientists have adapted “self-cleaving” 2A peptides into their multicistronic vectors. These peptides, first discovered in picornaviruses, are short (about 20 amino acids) and produce equimolar levels of multiple genes from the same mRNA. The term “self-cleaving” is not entirely accurate, as these peptides are thought to function by making the ribosome skip the synthesis of a peptide bond at the C-terminus of a 2A element, leading to separation between the end of the 2A sequence and the next peptide downstream.<sup>4</sup> The “cleavage” occurs between the Glycine and Proline residues found on the C-terminus meaning the upstream cistron will have a few additional residues added to the end, while the downstream cistron will start with the Proline.

The table below lists four common 2A peptides employed by scientists. 2A cleavage is universal in eukaryotic cells, and, although some scientists report close to 100% cleavage with some of these peptides, no consensus has been reached on which peptide works best. Likely the choice of specific 2A peptide will ultimately depend on a number of factors such as cell type or experimental conditions.

**Table 1: Common 2A Peptides**

Peptide	Amino Acid Sequence*
T2A:	(GSG) E G R G S L L T C G D V E E N P G P
P2A:	(GSG) A T N F S L L K Q A G D V E E N P G P
E2A:	(GSG) Q C T N Y A L L K L A G D V E S N P G P
F2A:	(GSG) V K Q T L N F D L L K L A G D V E S N P G P

\* (GSG) residues can be added to the 5' end of the peptide to improve cleavage efficiency.

## MULTICISTRONIC VECTORS (CONT'D)

### How Do I Get Started?

If you are looking to co-express your gene of interest along with a fluorescent protein or selectable marker, it is easiest to start with a plasmid that already has the multicistronic element and reporter cloned in. In these plasmids, you would simply clone your gene of interest into the multiple cloning site up or down stream of the IRES or 2A element (depending on the placement of the reporter gene).

Addgene's collection offers a variety of plasmids to express two or more genes, some of which are listed below. We should note that these vectors are primarily designed for bicistronic expression; however, many could be easily manipulated to express more than two genes.

**Table 2: Some Multicistronic Vectors at Addgene**

Plasmid Name	Multicistronic Element	Expression Type
<a href="#">MSCV-IRES-EGFP</a>	IRES	Retroviral
<a href="#">pMSCV-pBabeMCS-IRES-RFP</a>	IRES	Retroviral
<a href="#">pMSCV-IRES-YFP II</a>	IRES	Retroviral
<a href="#">pCMMP-MCS-IRES-Puro</a>	IRES	Retroviral
<a href="#">pEF1a-IRES-Neo</a>	IRES	Mammalian
<a href="#">MSCV-IRES-Luciferase</a>	IRES	Retroviral
<a href="#">pWPI</a>	IRES	Lentiviral
<a href="#">AmCyan-P2A-mCherry</a>	P2A	Mammalian
<a href="#">pC5Kan-P2A</a>	P2A	Insect
<a href="#">pUltra</a>	P2A and T2A	Lentiviral
<a href="#">Ac5-STABLE2-neo</a>	T2A	Insect

For the co-expression of multiple unique genes, you can start with a plasmid that has multiple cloning sites flanking the multicistronic element(s), or you could replace one of the reporter genes above with your gene or genes of interest. Some of the plasmids listed in the table above are designed to have one or more of the genes replaced.

Additionally, 2A peptides could be PCR-cloned between your genes of interest and you can then insert the whole multicistronic cassette into a backbone as a single unit. Although it is recommended to use the 2A peptides instead of an IRES when stoichiometrically equivalent levels of expression are required, we should also note that IRES and 2A peptides are not mutually exclusive elements. Labs have successfully utilized 2A and IRES elements within the same multicistronic vector, effectively making a construct that expresses multiple unique genes at equivalent levels upstream of an IRES fluorescent reporter for easy detection.<sup>5</sup>

### Further Reading

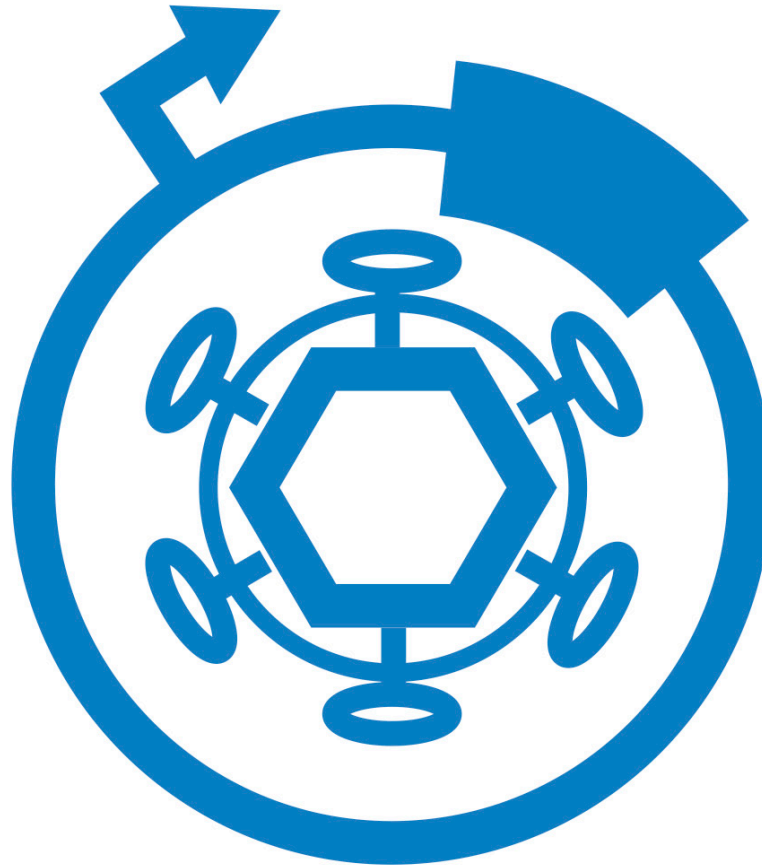
1. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Pelletier et al Nature. 1988 Jul 28;334(6180):320-5. [PMID: 2839775](#).
2. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. Jang et al. J Virol. 1988 Aug;62(8):2636-43. PubMed [PMID: 2839690](#).

## MULTICISTRONIC VECTORS (CONT'D)

---

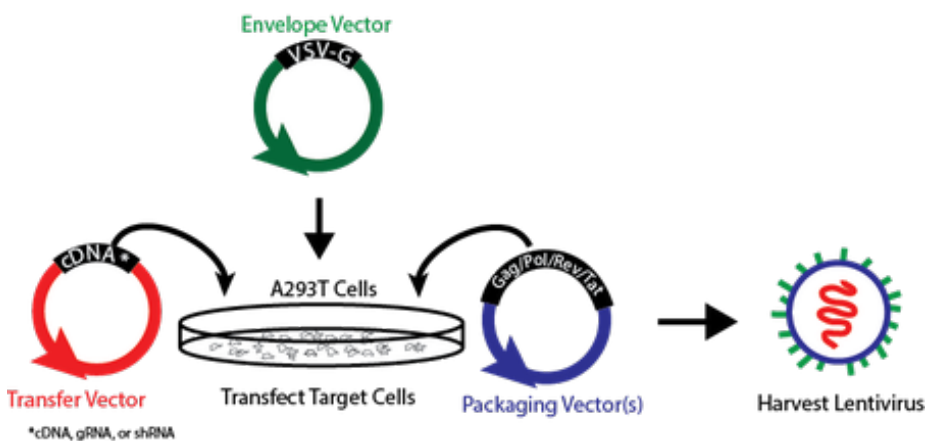
3. Highly Efficient Multicistronic Lentiviral Vectors with Peptide 2A Sequences. Ibrahim et al. Hum Gene Ther. 2009 Aug;20(8):845-60. doi: 10.1089/hum.2008.188. PubMed [PMID: 19419274](#).
4. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. Kim et al PLoS One. 2011;6(4):e18556. doi: 10.1371/journal.pone.0018556. Epub 2011 Apr 29. PubMed [PMID: 21602908](#).
5. Scalable signaling mediated by T cell antigen receptor-CD3 ITAMs ensures effective negative selection and prevents autoimmunity. Holst et al. Nat Immunol. 2008 Jun;9(6):658-66. doi: 10.1038/ni.1611. Epub 2008 May 11. PubMed [PubMed PMID: 18469818](#).
6. [Browse Addgene's Plasmid Guide](#) for more molecular biology and cloning information.

# CHAPTER 4: VIRAL EXPRESSION VECTORS



# VIRAL VECTORS – AN INTRODUCTION

By Mary Gearing, Addgene | August 31, 2015



*Schematic for the generation of lentivirus containing a gene of interest.  
The process for producing other types of viruses is similar.*

First isolated in the late 19th century, [viruses](#) are potent carriers of genetic information. Their ability to enter a cell and hijack its machinery to produce viral proteins makes them an ideal system for cellular transduction. Viral vectors are one of the most efficient ways to get genetic material into cells, but their use in research requires a balancing act of multiple parameters, including insert size, target cells, expression time, and biosafety. This chapter of Plasmids 101 will introduce you to four classes of common viral vectors and discuss key points of each system.

Viral vector systems based on modified retroviruses were first introduced in 1983. These two-piece systems were composed of packaging cell lines and transfer vectors. Packaging cell lines contained a mutated retrovirus (provirus) that produced retroviral proteins but could not package itself. A transfer vector expressing a gene of interest could be generated by cloning a transgene into a partial retrovirus genome. Once [transfected](#) into the packaging line, the transfer vector was efficiently packaged into viral particles, which were subsequently released into the media and could be collected for experimental use. Importantly, these particles were replication-incompetent because the packaging information was not included in the transfer vector but rather expressed *in trans* by the packaging cells. This type of system, called a helper-free system, represents the first method of creating viral vectors without the use of a wild-type virus, and the first step towards creating user-friendly viral vectors.

The 1990s marked the introduction of [adenoviral](#) and [lentiviral vector systems](#), designed using similar principles to those detailed above. A few years later, [adeno-associated viral \(AAV\) vector systems](#) that split transfer vector and packaging genes were introduced. The major caveat associated with early viral vector systems is safety. In some systems, just one recombination event between transfer vector and helper genome could produce replication competent virus, a true biosafety hazard. Subsequently, safer systems have split packaging genes across multiple plasmids, thereby increasing the number of recombination events necessary to generate replication competent virus. In the case of lentiviral systems, which are based on the HIV genome, many unnecessary proteins have been deleted to increase the safety of these vectors.

Once efficient viral packaging methods had been identified, attention quickly turned to customizing these vectors. Viruses of the same subtype share certain characteristics, like capsid or envelope proteins, but the specific proteins found in each virus alter viral infectivity and tropism, the types of cells a virus can infect. To alter tropism and improve infection rates, scientists began pseudotyping vectors by combining genes from multiple viruses. One common example is the VSV-G envelope, still used today with retroviral and lentiviral vectors to convey wide tropism and high infectivity. Conversely, pseudotyping may also be used *in vivo* to precisely define a set of cell types a vector can infect, as is the case with AAV-derived vectors.

## VIRAL VECTORS – AN INTRODUCTION (CONT'D)

Each of the four commonly used viral vector systems has its own advantages and disadvantages. Retroviral vectors have a large packaging capacity, but unlike the other vectors they can only transduce dividing cells. Lentiviral vectors integrate into the genome, potentially causing cellular transformation. Adenoviral vectors very efficiently transduce a gene of interest, but expression is transient and accompanied by a substantial immune response. AAV vectors provide long-lasting gene expression, but their packaging capacity is limited to about 4.5 kb.

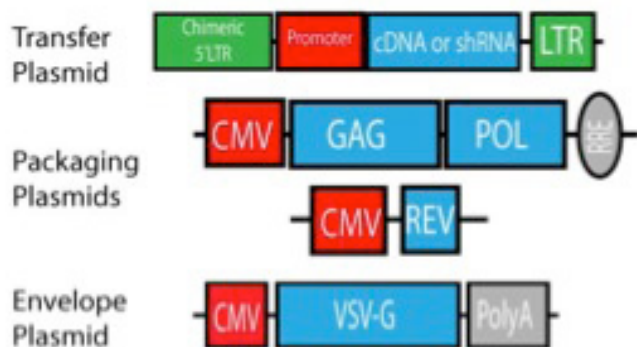
Current work focuses on optimizing these vector systems for both research and clinical uses. To decrease the immune response, some groups have used vectors with most of the viral genes removed or hybrid vectors. Improved targeting is another important goal, with work on pseudotyping and the use of antibody-conjugated vectors to preferentially target certain cell types ongoing. Increasing packaging capacity, especially with regard to AAV, would broaden the applications of these vector systems.

The optimal viral vector would have a large genetic capacity, low immune response, and high infectivity. Precise control over tropism and duration/level of transgene expression are also desirable characteristics. Although no viral vector system currently fits all of these criteria, viral vectors remain among the most useful tools for manipulating gene expression, and future research will likely further improve their utility. For help thinking about what type of virus you should use for a particular application, consult the following table:

Virus	Expression	Packaging Capacity	Cells Infected	Target Cell Genome Integration	Immune Response
<a href="#">Lentivirus</a>	Stable	<8 Kb	Dividing/Non-dividing	Yes	Low
<a href="#">AAV</a>	Transient or Stable	~ 4.5 Kb	Dividing/Non-dividing	No*	Very Low
<a href="#">Adenovirus</a>	Transient	>8 Kb	Dividing/Non-dividing	No	High
<a href="#">γ-Retrovirus</a>	Stable	<8 Kb	Dividing	Yes	Moderate

# VIRAL VECTOR ELEMENTS

By Marcy Patrick, Addgene | Jul 17, 2014



The use of viral vectors in research is beneficial for a number of reasons, including but not limited to: helping to get difficult-to-deliver DNA into mammalian cells, increasing the efficiency of gene transduction, allowing for control over which cells are infected through viral pseudotyping, and ease of [vector cloning](#) and modification. At the most basic level, viral vectors consist of a viral genome that has been adapted into a plasmid-based technology and modified for safety through the removal of many essential genes and the separation of the viral components. Of the many viruses out there, only a subset are commonly used in the lab: gamma-retrovirus, lentivirus, and adeno-associated virus.

## Adeno-associated Virus (AAV)

[Adeno-associated virus or AAV](#) is a small parvovirus that infects humans and some other primate species, but is not currently known to cause disease. If AAV is exclusively used to infect human cells, it stably integrates at a specific chromosomal site as part of its lysogenic life cycle; however, the addition of a helper virus such as adenovirus or a helper plasmid containing the specific viral proteins E1A, E1B, E2A, E4, and VA will cause AAV to enter a lytic cycle and replicate. There are at least eleven different serotypes of AAV, with more likely to be discovered. The most studied and experimentally used serotype is AAV2; however, many factors including species and cell type may make other serotypes more desirable.

## Gamma-retrovirus and Lentivirus

A retrovirus is an RNA virus that uses reverse transcriptase to make a DNA provirus which can be incorporated into the host's genome. There are two common genera of retrovirus used by scientists: [Gamma-retrovirus](#) (many times shortened to just "retrovirus") and [lentivirus](#). From these two genera, the most studied types are murine leukemia virus (MLV) and human immunodeficiency virus-1 (HIV-1), respectively. Since these viruses are closely related, their life cycle and the required components are basically the same, although differences do exist at the sequence level. This means that plasmids containing virus-specific elements such as long terminal repeats (LTRs) or structural proteins are not interchangeable; however, other, more general, viral components such as a heterologous envelope or a post-transcriptional regulatory element can be used across either system. The most notable difference between gamma-retrovirus and lentivirus is the fact that lentivirus can infect both dividing and non-dividing cells, whereas gamma-retrovirus is restricted to dividing cells only.

## Common Viral Vector Elements

When developing viral vectors, scientists strive for a number of features: low risk, high expression, large payload capacity, ability to infect target cells, no immune response from the host, and easy to develop/use in the lab. In virtually all viral expression systems employed by scientists, non-essential components are stripped away and the remaining native genes are spread over multiple plasmids to ensure safety. The choice for how the elements are divided up is dependent on whether the component needs to be provided *in cis* (on the same plasmid) or *in trans* (on a separate plasmid) as your insert. The tables on the following pages list some components most commonly found in viral vector systems.



## VIRAL VECTOR ELEMENTS (CONT'D)

### Table 1: AAV Elements

Element	Plasmid Type	Provided?	Purpose
ITR	Cloning	<i>in cis</i>	Inverted terminal repeat; 145 bases each. Symmetry of ITRs is required for efficient multiplication of the AAV genome. Forms a T-shaped hairpin that serves as the origin of viral DNA replication. Contains D region required for packaging. Cloning capacity between the ITRs is ~4kb.
Rep	Packaging	<i>in trans</i>	Packaging proteins with four possible variants: Rep78, Rep68, Rep52, and Rep40; Required for genome replication and necessary for integration. Rep proteins from most serotypes can be interchangeably used with any ITR serotype.*
Cap	Packaging	<i>in trans</i>	Structural capsid proteins with three variants: VP1, VP2, and VP3; VP1 possesses phospholipase A2 activity, which is likely necessary to release the AAV particles from late endosomes. VP2 and VP3 are crucial for correct virion assembly. Determines the serotype/viral tropism.

\*Exception is AAV5, which requires AAV5 Rep and AAV5 ITRs for packaging.

### Table 2: Gamma-retroviral and Lentiviral Elements

Element	Plasmid Type	Provided?	Purpose
LTR	Transfer	<i>in cis</i>	Long terminal repeats; U3-R-U5 regions found on either side of a retroviral provirus (see below). Cloning capacity between the LTRs is ~8.5kb, but inserts bigger than ~3kb are packaged less efficiently.
U3 (subelement of LTR)	Transfer	<i>in cis</i>	Unique 3'; region at the 3' end of viral genomic RNA (but found at both the 5' and 3' ends of the provirus). Contains sequences necessary for activation of viral genomic RNA transcription.
R (subelement of LTR)	Transfer	<i>in cis</i>	Repeat region found within both the 5' and 3' LTRs of retro/lentiviral vectors. Tat binds to this region.
U5 (subelement of LTR)	Transfer	<i>in cis</i>	Unique 5'; region at the 5' end of the viral genomic RNA (but found at both the 5' and 3' ends of the provirus).
5' LTR	Transfer	<i>in cis</i>	Acts as an RNA pol II promoter. The transcript begins, by definition, at the beginning of R, is capped, and proceeds through U5 and the rest of the provirus. Third generation vectors use a hybrid 5'LTR with a constitutive promoter such as CMV or RSV.
TAR	Transfer (2nd generation only)	<i>in cis</i>	Trans-activating response element; located in the R region of the LTR and acts as a binding site for Tat.
3' LTR	Transfer	<i>in cis</i>	Terminates transcription started by 5' LTR by the addition of a poly A tract just after the R sequence.
cPPT	Transfer	<i>in cis</i>	Central polypurine tract; recognition site for proviral DNA synthesis. Increases transduction efficiency and transgene expression.
Psi ( $\psi$ )	Transfer	<i>in cis</i>	RNA target site for packaging by nucleocapsid.

## VIRAL VECTOR ELEMENTS (CONT'D)

Element	Plasmid Type	Provided?	Purpose
RRE	Transfer	<i>in cis</i>	Rev Response Element; sequence to which the Rev protein binds.
WPRE	Transfer	<i>in cis</i>	Woodchuck hepatitis virus post-transcriptional regulatory element; sequence that stimulates the expression of transgenes via increased nuclear export.
Gag	Packaging	<i>in trans</i>	Precursor structural protein of the lentiviral particle containing Matrix, Capsid, and Nucleocapsid components.
Pol	Packaging	<i>in trans</i>	Precursor protein containing Reverse Transcriptase and Integrase components.
Rev	Packaging (on separate plasmid from Gag/Pol in third generation systems)	<i>in trans</i>	Binds to the Rev Response Element (RRE) within unspliced and partially spliced transcripts to facilitate nuclear export.
Tat	Packaging (second generation only)	<i>in trans</i>	Trans-activator; binds TAR to activate transcription from the LTR promoter.
VSVG	Envelope	<i>in trans</i>	Vesicular stomatitis virus G glycoprotein; Broad tropism envelope protein used to psuedotype most lentiviral vectors.

## VIRAL VECTOR ELEMENTS (CONT'D)

---

### Looking for Viral Vectors? Check out Viral Vectors at Addgene:

- [Addgene's most popular lentiviral plasmids](#)
- [Addgene's most popular retroviral plasmids](#)
- [Addgene's collection of AAV plasmids](#)
- [Check Out Our Viral Service](#)

### Learn More:

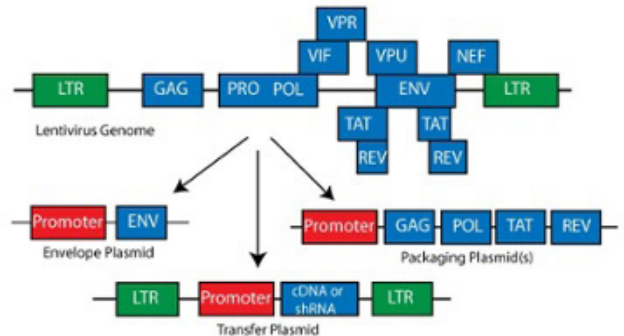
- [Overview of Lentiviral Packaging Plasmids](#)
- [Lentiviral Protocols and Resources](#)
- [More Lentiviral FAQs Answered by Addgene](#)
- [Didier Trono Lab: Lentivectors Toolbox](#)

# LENTIVIRAL PLASMID FAQs ANSWERED

By Kendall Morgan, Addgene | April 23, 2014

**Lentiviruses are useful and efficient tools to introduce your gene of interest into cells.** Unlike gamma-retroviruses that can only infect dividing cells, lentiviruses can infect dividing and non-dividing cells.

Addgene has an [extensive collection of lentiviral plasmids](#) created for a variety of applications including cDNA expression, shRNA-mediated knockdown, Tet and Cre-regulated expression, [CRISPR genome editing](#), and more. Not surprisingly, Addgene receives many questions from scientists all over the world looking for some additional information or clarification on lentiviral vectors.



## Frequently Asked Questions

### How are lentiviral particles produced?

[This useful resource](#) from Addgene Depositor (and lentiviral expert) [Didier Trono](#)'s lab webpage explains how lentiviral particles are produced by co-transfecting three essential ingredients: the lentiviral packaging vector, a transfer vector including the gene of interest, and the plasmid coding for an envelope. Three generations of lentiviral packaging systems have been developed over the years based on HIV-1; however, the 1st-generation vectors are never used as they pose too much of a biosafety risk to scientists. Second-generation vectors are likely sufficient for most experiments; however, the 3rd-generation packaging systems offer the maximal biosafety available for the technology. The minor limitation in using the 3rd-generation system is that it involves transfecting four different plasmids into the producer cells (two packaging plasmids, an envelope plasmid, and the lentiviral transfer vector), rather than the three used for 2nd-generation systems.

### Are retroviruses and lentiviruses the same thing? Can I package lentivirus with retroviral packaging plasmids (and vice versa)?

Lentiviruses are part of the retroviral family. What scientists often refer to as “retroviruses” are technically gamma-retroviruses - another, separate member of the retroviral family. While both lentiviruses and gamma-retroviruses use the same genes for packaging (that's gag, pol, and env), the isoforms of these proteins, as well as the viral long terminal repeats (LTRs) are different. As a result, lentiviral and retroviral packaging vectors are not interchangeable.

### Can I use lentiviral transfer vectors for transient transfections?

It is not recommended to use viral plasmids transiently. This is especially true if the transfer vector does not contain an independent promoter and instead exclusively uses the viral LTR promoter to drive expression, since the LTR is not strong enough to drive expression of the insert transiently. Virus must be produced and your target cells infected in order to see adequate expression.

## LENTIVIRAL PLASMID FAQs ANSWERED (CONT'D)

---

### 2nd vs. 3rd Generation Lentiviral Plasmids

#### *How can you tell 2nd- and 3rd-generation transfer vectors apart?*

Addgene defines 2nd- or 3rd-generation transfer vectors based on whether or not they have a chimeric 5'LTR. Second-generation transfer plasmids have a wildtype 5'LTR, which requires the presence of the Tat protein to work, while 3rd-generation transfer plasmids have a chimeric 5'LTR that includes a CMV or RSV promoter as well as a portion of the 5'LTR. Including a chimeric 5'LTR removes the requirement for the HIV Tat protein, thus decreasing the probability of creating replication-competent lentivirus in your target cells.

#### *How do the packaging systems differ between 2nd- and 3rd-generation vectors? What does this mean for transfer vectors?*

The [lentiviral packaging systems](#) differ in two ways: 1) Second-generation systems consist of two plasmids while 3rd-generation systems require three plasmids. 2) Second-generation systems express HIV Tat; 3rd-generation systems do not.

Second-generation transfer plasmids MUST be packaged with a 2nd-generation system because the wildtype 5'LTR promoter requires Tat to function. Third-generation plasmids can be packaged with either system. Please note that the generation used to package the virus does not change the generation of the transfer vector.

#### *What envelope glycoprotein will be expressed on virus particles?*

This depends on the envelope plasmid you use; however, the choice of envelope dictates the tropism of the virus. The VSV-G envelope is very common due to its broad host range, but viruses can be easily pseudotyped by utilizing an alternative envelope plasmid as described [here](#).

#### *Are transfer plasmids replication competent? What is SIN?*

Most (if not all) Addgene transfer plasmids are replication-deficient meaning they can be used to create virus that is capable of infecting target cells, but cannot produce any new viral particles after the initial infection.

SIN is shorthand for self-inactivating, which is achieved by deleting a large portion of the 3'LTR. This results in short, inactivated 5' and 3' LTRs after integration and further reduces the possibility of creating replication-competent virus. Both 2nd and 3rd generation transfer vectors can be SIN.

#### *Which genes are deleted or modified to result in replication deficiency? What percent of the vector components are HIV-based?*

No single plasmid contains all the components necessary to produce viral particles. The components are divided as follows:

- Transfer vectors contain minimal *cis*-acting HIV components: LTRs, PPT, RRE, and psi packaging signal. Viral components typically total <1.5kb, which is rarely more than 30% of the transfer plasmid. Many transfer vectors are SIN.

## LENTIVIRAL PLASMID FAQs ANSWERED (CONT'D)

- Packaging plasmid(s) contain the minimal number of HIV genes required for virus production (3 or 4). Third-generation vectors contain gag, pol, and rev. Second-generation vectors contain those three genes plus tat.
- The envelope plasmid provides a heterologous envelope for pseudotyping and is not HIV-derived.

### Concerning the Safety of Lentiviral Transduction

*You mentioned biosafety. What are the major risks when it comes to lentiviruses in terms of biosafety?*

According to the American Biological Safety Association, “the two major risks of lentiviral vectors are: 1) the potential generation of replication competent virus [usually HIV-1]; and 2) the potential for oncogenesis through insertional mutagenesis. These risks are largely based upon the vector system used and the transgene insert encoded by the vector.”

Biosafety should always be considered with respect to the specific experiments being performed and you should understand the guidelines for using such reagents as outlined by your institution or country.

The table below breaks down the biosafety concerns associated with various aspects of lentiviral transduction along with suggestions on how to lower the risk.

Biosafety Concern	High Risk	Lower Risk
Vectors	Two plasmids (1st gen) used to produce virus; expression of viral genes	Three (2nd gen) or four (3rd gen) plasmids used to produce virus; removal of viral genes
Insert/Transgene	Oncogenic, apoptotic, or toxic	Non-oncogenic, nonapoptotic, non-toxic
Envelope/Host	Pantropic/amphotrophic	Ecotropic/non-permissive
Propagation	Large scale	Lab scale (less than 100mL)

More biosafety resources can be found on the [Addgene Biosafety Page](#).

## AAV: A VERSATILE VIRAL TOOL FOR GENE EXPRESSION IN MAMMALS

By Didem Goz Ayturk | July 26, 2016

[Adeno-associated virus \(AAV\)](#) has emerged as a favorite viral tool for both research and clinical applications. AAV can be used to transiently express a gene of interest in a variety of cell types. It was first described about 50 years ago as a contaminant of adenoviral preparations, hence the name ([Atchison et al., 1965](#)). AAV is a single stranded, DNA virus belonging to the family Parvoviridae. It has a “simple” genome packaged in an icosahedral capsid. It does not have a lipid coat, also called an envelope, and thus cannot support the addition of a glycoprotein, such as VSV-G, to its surface. In research applications, the genome is typically gutted so that precious cargo space is opened for gene delivery, and for safety. You can easily complement the virus in a tissue culture setting, in other words “in trans”, by supplying the genes that encode the replicase functions and capsid proteins. This gives researchers the ability to produce more virus in a controlled setting. Even though AAV is isolated from a wide range of organisms, it has not been associated with disease, and it is considered a biosafety level 1 (BSL1) viral agent.

### 4 Considerations for the Perfect Recombinant AAV (rAAV) Tool

AAV can be used in a wide variety of applications including transient gene expression in a particular cell type, [CRISPR genome editing](#), [optogenetics](#), and [chemogenetics](#) experiments. If you are new to utilizing rAAVs as gene delivery tools, here are some things you should consider before getting started:

- 1. Cargo capacity:** Even the gutted version of rAAV has a cargo capacity of ~ 4.7 kb, which is one of the key limitations of the virus for gene delivery. However, if your gene(s) of interest is small enough, you can design single rAAV vectors with two genes packaged in them, and use elements such as [IRES or 2A](#), to co-express them from one promoter. Co-infection rates for different rAAVs are also quite high if the titers are high, so if you cannot fit both your genes in one vector, you can coinfect (although this may not always work efficiently *in vivo*).
- 2. Specificity:** There are a few different rAAV components that can drive gene expression in specific cells/tissues; these include the promoter, Flp or Cre dependent gene switches, and the serotype (serotype will be discussed more below). If you are aiming for broad expression, a broadly active promoter, such as CAG (chicken beta actin promoter with CMV immediate early enhancer), is a good choice. If you have a specific target cell type, you may want to try a different promoter, e.g. CamkIIa for a neuron-specific one. rAAV vectors with Cre- or Flp-dependent gene expression, on the other hand, can be injected into animals with Cre or Flp expressed in specific cell types resulting in gene expression only in these cells.
- 3. Serotype:** Capsid proteins are very important rAAV vector components and drive the biology of these vectors. Although multiple studies have shown that different serotypes differ in their ability to infect different cell types, a recent study showed that most (or all) serotypes use the same receptor (AAVR) ([Pillay et al., 2016](#)). The observed tropisms may be due to other factors, such as the attachment of the virions to the cell surface, or perhaps steps subsequent to entry, such as uncoating.

AAV nomenclature can be confusing. You will likely see designations such as AAV2/2 or AAV2/8, but what do these numbers actually mean? The first number stands for the inverted terminal repeat (ITR) type. ITRs are short DNA sequences that flank the AAV genome and allow it to form concatemers in host cells. They, along with the Rep protein, facilitate integration into the human genome at the AAVS1 site on chromosome 19, something that is only observed with wild type AAV – not the vector forms. Almost all vectors contain type 2 ITRs. Type 2 ITRs can be packaged with a variety of capsid types. The type of capsid used to package an rAAV vector, i.e. the serotype - is denoted by the second number. For example, if an rAAV vector has an ITR of type 2 and a capsid of type 8, it will be denoted AAV2/8.

## AAV: A VERSATILE VIRAL TOOL FOR GENE EXPRESSION IN MAMMALS (CONT'D)

**4. Genome:** Wild type AAV are single-stranded DNA viruses. After uncoating of the DNA, the virus relies on the host cell replication machinery to synthesize the complementary DNA strand. This step is considered to be a limiting factor in the transduction efficiency of rAAV. To overcome this, (McCarty et al) engineered a dimeric, or self-complementary, AAV (termed an scAAV) by mutating one of the ITRs. scAAV vectors display rapid onset (days) and higher level transgene expression than ssAAV. However their packaging capacity (<2.5Kb) is half that of ssAAV (<4.8Kb), restricting the number of genes and regulatory elements that can successfully be packaged.

scAAV may be valuable if you require rapid transgene expression. A lower titer may be able to achieve desired transgene expression levels, minimizing the chance of seeing toxicity or immunogenicity due to high virion concentrations.

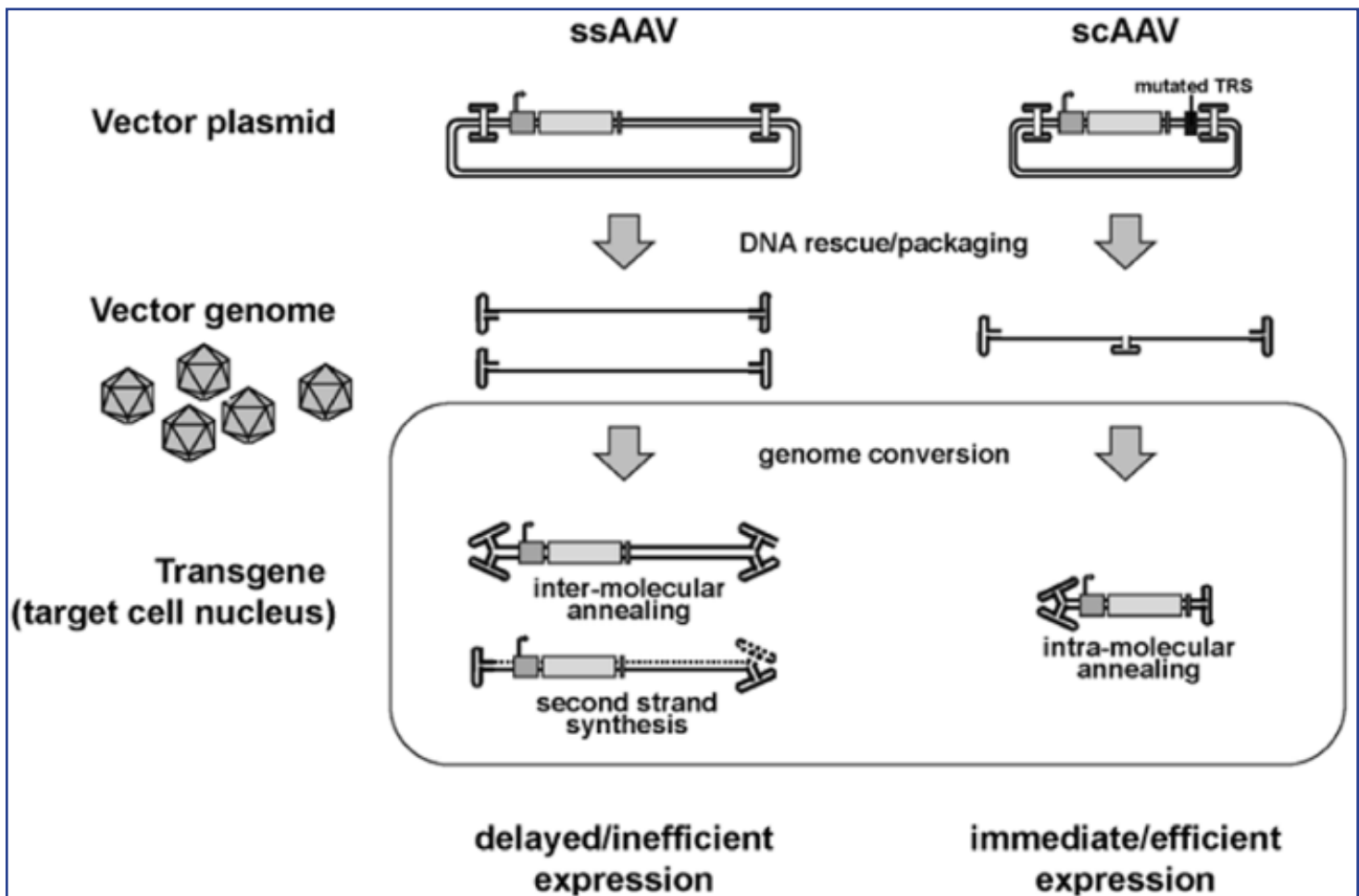


Figure 1: Comparison of ssAAV and scAAV genomes, packaging, and transduction. From Takashi Okada (2013). *Efficient AAV Vector Production System: Towards Gene Therapy for Duchenne Muscular Dystrophy*, Gene Therapy - Tools and Potential Applications, Dr. Francisco Martin (Ed.), InTech, DOI: 10.5772/53023

## Producing Your rAAV

There are several facilities nationwide where you can obtain excellent quality, high titer rAAV ([including Addgene!](#)), but you can also generate rAAVs in your own lab with standard molecular biology tools and tissue culture experience.



## AAV: A VERSATILE VIRAL TOOL FOR GENE EXPRESSION IN MAMMALS (CONT'D)

Briefly, you start by transfecting HEK293T cells with an rAAV vector carrying your gene of interest, an adenoviral helper plasmid, and a plasmid containing Rep and Cap (Figure 2, commonly referred to as “triple plasmid transfection”). After 2-3 days, you collect the supernatant, (or in some cases make a cell extract as some rAAV serotypes are released into the culture media while others are not) ([Vandenberghe et al](#)) from which you precipitate the virions using PEG. You then purify the virions further by density gradient centrifugation using a high-speed ultracentrifuge. The virions will form a band due to their high density, and you can collect this band from the gradient. You then remove the density gradient material by dialysis/buffer exchange and can further concentrate the virions if needed.

The virion prep can be titered by qPCR using primers targeting the viral genome, and/or protein gels ([Veldwijk et al](#)). These titers are for physical particles, many of which are not infectious. The ratio of physical particles to infectious particles can vary greatly, from 1 to >100.

AAV virions are quite stable. They will withstand freeze-thaw cycles and dehydration, which makes contamination from previous preps that might be on your bench, in your centrifuge, or in your incubator, possible. To minimize this, you should treat all disposable items and surfaces that come into contact with the AAV with a disinfectant, such as Coverage Plus NPD. Also important is to apply common cell culture practices diligently.

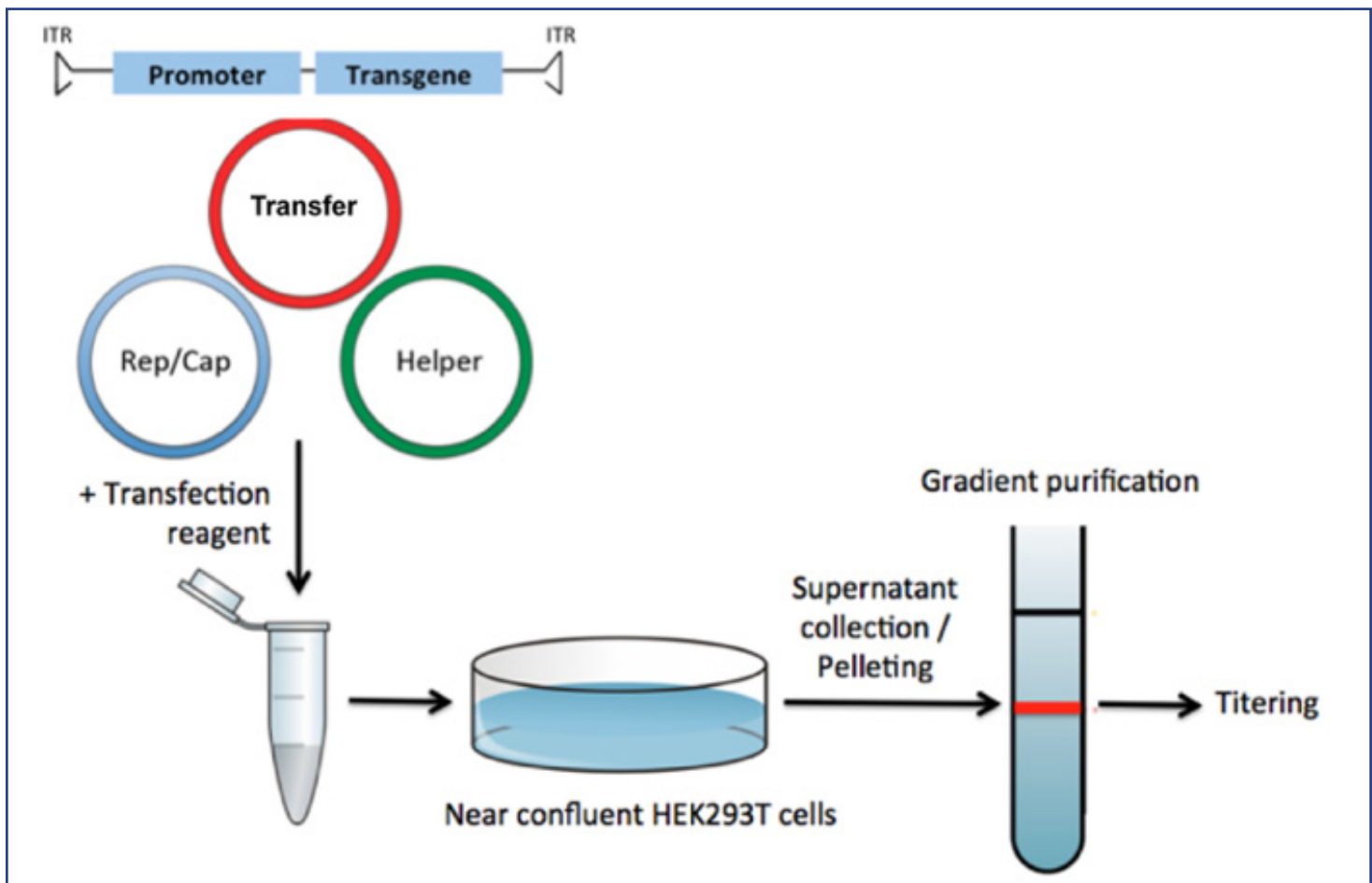


Figure 2: Recombinant AAV (rAAV) transfer vectors typically carry a transgene driven by a promoter of choice between ITRs, but do not encode Rep/Cap. Rep/Cap ORFs encode proteins essential for viral genome replication, virion assembly, and the 60 mer viral capsid. In this case, Rep/Cap are encoded by plasmids that are co-transfected with the transgene plasmid in tissue culture. Also needed for replication is an adenoviral helper plasmid, which induces the lytic life cycle for the AAV, and its release.

## AAV: A VERSATILE VIRAL TOOL FOR GENE EXPRESSION IN MAMMALS (CONT'D)

### *In vivo* rAAV Delivery

There are several factors concerning AAV delivery *in vivo* that vary, depending upon your biological system. These factors greatly influence the level of success of your infection. Whatever your application might be, listed below are a few parameters you may want to consider before proceeding with your injections:

**Injection Titer:** The number of viral particles you can deliver to the tissue will be determined by the titer of your viral prep and the maximum volume that can be delivered, and is of great importance. Unfortunately, as mentioned above, the titer of the physical particles cannot be directly translated to what you will observe *in vivo*, although it is a starting point. You can start by scouring the literature for advice on particular tissues, but, often, the optimal dose will need to be determined empirically *in vivo* by trying a series of dilutions.

**Age of the Animal:** First of all, it is important to remember that terminally differentiated cells (i.e. postmitotic cells) will have long term AAV expression, but, in dividing cells, expression will be lost. The age of the injected animal will determine what types of cells you can infect successfully (i.e. at an early developmental time point, cells that are not yet born, even if their mitotic progenitor cells are present, cannot be infected by AAV) ([Xiong and Cepko](#)).

**Visualization of the Infection:** You should also think about how you will visualize your rAAV infection before you start the delivery process. Have you included a reporter gene in your rAAV construct, such as GFP or mCherry? These will make it easy to directly observe where expression is occurring via fluorescence measurements. If not, you can consider co-injecting an AAV that carries a reporter gene under the same promoter driving your gene of interest and measuring reporter gene expression as a proxy for your gene of interest, which may not be easy to detect. If co-expression or co-infection are just not possible, detecting your protein by immunohistochemistry or *in-situ* hybridization might be your best bet.

**Time after infection:** An adequate amount of time needs to pass between injection and tissue processing to detect AAV-mediated gene expression. Timing will highly depend on the capsid type and on the tissue you're infecting. Waiting ~2 weeks is a good starting point for many tissues. AAV-mediated gene expression has been reported to be quite stable, lasting for several years in human clinical trials and in dogs ([Wonjo et al](#)).

Before Design and Growth	At and After Delivery
Cargo Capacity	Administration Titer/Volume
Promoter (Specific v. Broad)	Age of the Animal
Serotype - Tissue/Cell Tropism	Visualization of the Infection
Production (DIY or Purchase?)	Time after Infection

AAV is an extremely versatile tool for viral gene delivery. The demands of your particular research application will determine the appropriate parameters of AAV design and delivery. Although your final serotype/titer/age determination will heavily rely on pilot experiments, you can look to the ever-growing literature for design recommendations appropriate to many different biological systems. With more than a hundred clinical trials involving AAV started in the last two decades, it seems that this tool will remain popular in both therapeutic and

## AAV: A VERSATILE VIRAL TOOL FOR GENE EXPRESSION IN MAMMALS (CONT'D)

---

basic research settings in the years to come.

### Further Reading

1. Atchison, Robert W., Bruce C. Casto, and William McD Hammon. "Adenovirus-associated defective virus particles." *Science* 149.3685 (1965): 754-755. PubMed [PMID: 14325163](#).
2. Pillay, S., et al. "An essential receptor for adeno-associated virus infection." *Nature* 4:520(7588):108-12 (2016). PubMed [PMID: 26814968](#).
3. McCarty, D. M., Paul E. Monahan, and Richard J. Samulski. "Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis." *Gene therapy* 8.16 (2001). PubMed [PMID: 11509958](#).
4. Vandenberghe, Luk H., et al. "Efficient serotype-dependent release of functional vector into the culture medium during adeno-associated virus manufacturing." *Human gene therapy* 21.10 (2010): 1251-1257. PubMed [PMID: 20649475](#). PubMed Central [PMCID: PMC2957237](#).
5. Veldwijk, Marlon R., et al. "Development and optimization of a real-time quantitative PCR-based method for the titration of AAV-2 vector stocks." *Molecular Therapy* 6.2 (2002): 272. PubMed [PMID: 12349826](#).
6. Xiong, Wenjun, and Constance Cepko. "Distinct expression patterns of AAV8 vectors with broadly active promoters from subretinal injections of neonatal mouse eyes at two different ages." *Retinal Degenerative Diseases*. Springer International Publishing, 2016. 501-507. PubMed [PMID: 26427452](#).
7. Okada, Takashi. *Efficient AAV Vector Production System: Towards Gene Therapy For Duchenne Muscular Dystrophy*. INTECH Open Access Publisher, 2013. [Intech Webpage](#).
8. Wojno, Adam P., Eric A. Pierce, and Jean Bennett. "Seeing the light." *Science translational medicine* 5.175 (2013): 175fs8-175fs8. PubMed [PMID: 23467559](#).

# CHAPTER 5: PLASMIDS THAT GLOW



# HISTORY OF FLUORESCENT PROTEINS

By A. Max Juchheim, Addgene | September, 2015

Luminescent (or “light-emitting”) molecules are a very useful class of tools to scientists because light is something that we can easily detect and measure. Proteins that give off light include [fluorescent proteins](#), like Green Fluorescent Protein (GFP), as well as chemiluminescent ones, like [luciferases](#). These molecules occur naturally in bioluminescent organisms, but their real power lies in the clever adaptations scientists have made allowing their use in the laboratory.



*Aequorea victoria*. Attribution: Sierra Blakely

The story of how GFP (as well as other fluorescent proteins derived from it) came to be used as a research tool is particularly interesting. Although humans have known about bioluminescent organisms, and their characteristic radiance, for some time – there are records describing them dating back to the first century A.D. – the first time anyone truly examined what caused this glowing phenomenon wasn’t until much later. In the 1960s, Osamu Shimomura studied the bioluminescent properties observed in the crystal jellyfish *Aequorea victoria* (left) in molecular detail.

On the underside of the “umbrella” of *A. victoria* there is a ring that glows with a faint green light. To investigate this green glimmer, Shimomura collected many, many jellyfish specimens from Puget Sound in the Pacific Ocean, off of the coast of Washington state. Using these samples, he was able to isolate two proteins from the jellyfish’s photoorgans; the first, which he called aequorin, gives off a faint blue light when it binds calcium ions, and the second, which we now call GFP, absorbs that blue light and glows green.

In the late 1980s, another researcher, Douglas Prasher, got the idea that this new glowing green protein could be used to measure gene transcription as well as to track protein localization. And as it turns out, he was right! He began to study the *A. victoria* gene responsible for encoding GFP, and, in 1992, he reported its sequence. Soon after that, in 1994, Prasher’s collaborator Martin Chalfie expressed GFP in exogenous organisms (*E. coli*, and later *C. elegans*) for the first time. It was then that scientists began to truly realize the potential of GFP as a research tool, and the number of people studying it grew. In particular, studies in the laboratory of UCSD biochemist [Roger Tsien](#) resulted in much of our understanding of the mechanism of GFP’s function. His lab described the protein’s structure and also discovered many mutations that alter and improve its function.

The importance of GFP was recognized in 2008 when the Nobel Prize in chemistry was awarded to Shimomura, Chalfie, and Tsien for their work on the protein. This great honor serves to highlight how useful luminescent proteins like GFP really are. In this chapter, we’ll take a closer look at how they’re used, and give some tips on deciding which one is right for you, so I invite you to read on!

# GREEN FLUORESCENT PROTEIN

By Marcy Patrick, Addgene with contributions from A. Max Juchheim, Addgene | May 15, 2014

Bioluminescence and fluorescence from proteins such as Green Fluorescent Protein (GFP) has likely existed in creatures such as jellyfish for millions of years; however, it took until the 1960s for scientists to begin to study GFP and deduce its biochemical properties. Now GFP and its fluorescent derivatives are a staple in the lab. GFP is used in research across a vast array of biological disciplines and scientists employ GFP for a wide number of functions, including: tagging genes for elucidating their expression or localization profiles, acting as a biosensor or cell marker, studying protein-protein interactions, visualizing promoter activity, and much more.

## Why Green Fluorescent Protein?



Image source: [Wikimedia Commons](#), accessed: 5/15/2014.  
Photo author: [Ingrid Moen, et al.](#)

GFP is a ~27 kDa protein consisting of 238 amino acids derived from the crystal jellyfish *Aequorea victoria*. It has a fluorescent emission wavelength in the green portion of the visible spectrum (hence the name), which is due to a chromophore formed from a maturation reaction of three specific amino acids at the center of the protein (Ser65, Tyr66, and Gly67). When first discovered, one of the most surprising aspects of GFP was the fact that the chromophore forms spontaneously and without additional co-factors, substrates, or enzymatic activity – it only requires the presence of oxygen during maturation. This meant that the protein could be taken directly from *A. victoria* and expressed in any organism while still maintaining fluorescence.

The protein structure, first reported in 1996, is an eleven  $\beta$ -sheet-containing “barrel” shape, with the chromophore concealed at the center of the structure, shielded from quenching by aqueous solvent. This tightly-packed structure explains the importance of the entire GFP protein, which is almost completely required to maintain fluorescent activity; very little truncation is tolerated, however, point mutations are acceptable. GFP’s main advantage over conventional fluorescent dyes of the time was the fact that it was non-toxic and could be expressed in living cells, enabling the study of dynamic, physiological processes.

## Re-engineering GFP to Increase its Range of Color and Application

Almost as soon as its sequence was elucidated, scientists began engineering new versions of GFP through mutagenesis in order to improve its physical and biochemical properties. In 1995, Roger Y. Tsien described an S65T point mutation that increased the fluorescence intensity and photostability of GFP. This also shifted its major excitation peak from 395 nm to 488 nm, effectively ameliorating the deficiencies found in the wildtype protein and facilitating its widespread use in research. Many other mutations have since been introduced to GFP and new iterations of fluorophores are constantly being engineered.

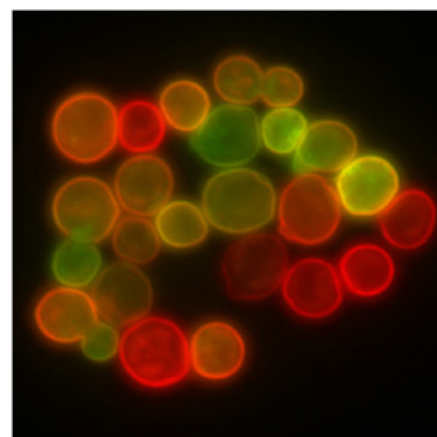


Image source: [Wikimedia Commons](#), accessed: 5/15/2014. Photo author: [Masur](#).

## GREEN FLUORESCENT PROTEIN (CONT'D)

Table 1 below lists a few common fluorescent proteins and their mutations relative to wildtype GFP. Although not listed here, many permutations within each color also exist with only slight variations separating them.

**Table 1: The Specific Mutations Comprising Common Fluorophores**

Fluorescent Protein	Mutations Relative to Wildtype GFP
EGFP	F64L; S65T
EYFP	S65G; V68L; S72A; T203Y
mYFP	S65G; V68L; Q69K; S72A; T203Y; A206K
Citrine	S65G; V68L; Q69K; S72A; T203Y
ECFP	F64L; S65T, Y66W; N149I; M153T; V163A
mCFP	F64L; S65T, Y66W; N149I; M153T; V163A; A206K
Cerulean	F64L, S65T, Y66W, S72A, Y145A, H148D, N149I, M153T, V163A
EBFP	F64L, S65T, Y66H, Y145F

**Please note** that many fluorescent proteins found on the red side of the spectrum are not GFP derivatives, but are instead related to the dsRed protein isolated from *Discosoma sp.* Similar work has been done to expand the red-fluorescent protein repertoire; however, these proteins are unique from GFP and the mutation definitions found in Table 2 below may not apply.

**Table 2: Functional Role of Specific Mutations in GFP Derivatives**

Mutation	Known Function
S65T	Increased fluorescence, photostability, and a shift of the major excitation peak to 488 nm
F64L	Increased folding efficiency at 37C
Y66W	Causes the chromophore to form with an indole rather than phenol component (cyan derivatives)
Y66H	Blue-shifts the wavelength (blue derivatives)
Y145F	Increases quantum yield for BFP
Y145A and H148D	Stabilizes the structure of Cerulean derivatives
F99S, M153T, V163A	Improves folding at 37C, reduces aggregation at high concentrations
T203Y	Red-shifts the wavelength (yellow derivatives)
A206K	Interferes with dimer interface (monomeric derivatives)
K26R, Q80R, N146H, H231L, (and probably others)	Neutral mutations

# GREEN FLUORESCENT PROTEIN (CONT'D)

---

## A Multitude of Applications

Due to its size and ease of use, GFP and other fluorescent proteins have become a mainstay in molecular biology. Scientists can easily utilize GFP-containing plasmids as a means to many functional ends. We've listed our favorites below, but many other uses currently exist, and new GFP technology is constantly being developed!

- **Fusion Tagging:** One of the most common uses, GFP can be fused to the N- or C-terminus of a protein, which allows the scientist to visualize when and where the gene is expressed. [Click here to view Addgene's collection of empty backbones for constructing fluorescent fusions.](#)
- **Transcription Reporter:** Placing GFP under the control of a promoter of interest can be used to effectively monitor gene expression from that promoter in a given cell type. This type of transcription reporting was among the earliest uses of GFP.
- **Förster Resonance Energy Transfer (FRET):** This is used to study the interactions between two proteins or between two domains of a protein that undergoes conformational change. Typically, two fluorescent proteins with overlapping excitation/emission spectra are used; one fused to each protein or domain being tested. [Find FRET plasmids here.](#)
- **Split EGFP:** An alternative to FRET, split enhanced GFP (EGFP) has also been used to study protein-protein interactions. In this case, two portions of EGFP are fused to the proteins of interest, and when they come into close proximity, the two halves of EGFP undergo folding, maturation, and fluorescence.
- **Biosensors:** A wide array of GFP-based fluorescent biosensors has been designed to detect a variety of intracellular conditions, including ion (such as Ca<sup>2+</sup>) concentrations and pH, using a range of strategies such as FRET, calmodulin, and others. [Review Addgene's collection of fluorescent biosensors here.](#)
- **Optogenetics:** Scientists can use light to detect, measure, and control molecular signals, cells, and groups of cells in order to understand their activity and visualize the effects of alterations to this activity. [Learn more about optogenetics at OpenOptogenetics](#) and [find optogenetic actuators and sensors at Addgene.](#)
- **Cell Marking/Selection:** Expression constructs like plasmids often include GFP as a marker to help identify which cells have successfully taken up the plasmid. This can serve as an alternative to [antibiotic selection](#). Plasmids of this type may have the GFP under the control of an additional promoter from that of the gene of interest, or expressed from the same transcript as the gene of interest, but after an [internal ribosome entry site \(IRES\)](#). This is oftentimes used in conjunction with FACS (see below).
- **Fluorescence-Activated Cell Sorting (FACS):** This is a type of flow cytometry that separates mixtures of cells into distinct populations based on fluorescent signal. Thus, FACS can be used to separate cells expressing GFP from cells that are not.
- **Developmental/Transgenic Uses:** Because of its stability, GFP can be used in lineage tracking capacities in cell fate studies. It can also be used, when put under control of promoters of interest, to visualize the developmental stage at which these promoters are active. Further, GFP can label transgenically modified ES cells, which can then be used for implantation and generation of transgenic mice.



## GREEN FLUORESCENT PROTEIN (CONT'D)

---

- **Purification:** GFP can be used as a general epitope tag for protein purification and a number of commercial antibodies to GFP are available.
- **Others:** We've really just scratched the surface of the potential applications for GFP. It has also been used to identify particular cell populations in drug screens, to visualize micrometastases in nude mice in cancer studies, to act as a reporter for DNA double strand break repair, and to label pathogenic intracellular microbes to visualize host/pathogen interactions. New uses for GFP are being realized all the time – maybe you'll even invent your own!

# WHICH FLUORESCENT PROTEIN SHOULD I CHOOSE?

By Gal Haimovich of [greenfluorescentblog](http://greenfluorescentblog.com) | May 20, 2014

## FPs Classified by the Emission Color (Emission Wavelength Range)

FPs are usually classified by emission color as outlined below (or emission wavelength range). By mutating GFP, the variants **blue FP** (BFP), **cyan FP** (CFP), and **yellow FP** (YFP) were derived. Many of the **orange FPs** and the **Red FPs** from which they were derived come from the coral *Discosoma sp.*

Color	Wavelength of Light
Blue	424 - 467 nm
Cyan	474 - 492 nm
Green	499 - 519 nm
Yellow	524 - 538 nm
Orange	559 - 572 nm
Red	574 - 610 nm
Far-Red	625 - 659 nm
Infra-Red	≥ 670 nm

## Unique Categories of FPs

In addition to emission wavelength range, there are other traits that need to be considered when choosing an FP:

- **Photoactivatable/Photoconvertible:** These proteins can switch their color when activated by a specific excitation wavelength. This means that the emission wavelength can change. In a few cases, the initial state of the protein is non-fluorescent, thus allowing very low background level of fluorescence. Examples for such [photoactivatable or photoconvertible proteins](#) are PA-GFP, Dendra2, and the mEOS proteins. Some proteins are reversibly switchable (e.g. rsEGFP, Dreiklang).
- **Fluorescent Timers (FT):** These proteins change their color over time. Therefore, these can be [used as “timers” for cellular processes](#) following their activation. The four main FTs are called Slow-FT, Medium-FT, Fast-FT, and mK-GO.
- **Large Stokes Shift (LSS):** Stokes shift (named after [George G. Stokes](#)) is the shift in wavelength from excitation to emission. For most FPs, Stokes shift is less than 50 nm (often much less). For LSS proteins, the Stokes shift is ≥ 100 nm. Specifically, these proteins are excited by UV light or blue light and their emission is green or red light. For example, [T-Sapphire](#), [LSSmOrange](#), and [LSSmKate](#).
- **Fluorescent Sensors:** These FPs change their excitation/emission behavior upon environmental changes (e.g. pH, Ca<sup>2+</sup> flux, etc). The most commonly used are GECIs - genetically encoded calcium indicators (e.g. GCaMP). Others include: pHluorin & pHTomato (pH sensors), HyPer (H<sub>2</sub>O<sub>2</sub> sensor), ArcLight (voltage sensor), and iGluSnFr (glutamate sensor). More examples of these [biosensors can be found at Addgene](#).

## WHICH FLUORESCENT PROTEIN SHOULD I CHOOSE? (CONT'D)

---

### 8 Points to Keep in Mind When Choosing an FP

#### 1. Excitation & Emission (ex/em):

- Each FP has its unique ex/em peak. Therefore, choose FPs that your system can excite, and detect the emission. For example, if your microscope has only two lasers, at 488nm and 561nm, you will not be able to use far-red FPs. If you do not have a filter that will pass blue light to the detector/camera, then BFPs are of no use to you.
- When using more than one FP, make sure their emission light does not overlap in wavelength. In many microscopes the filters are not narrow enough to distinguish between closely related colors. Furthermore, most FPs have a broad range of emission which will be detected by longer-wavelength filters (e.g. GFP also emits yellow light).

**2. Oligomerization:** The first generations of FPs were prone to oligomerize, which may affect the biological function of the FP-fusion protein. Therefore, it is recommended to use monomeric FPs (usually denoted by an “m” as the first letter in the protein name, e.g. mCherry).

**3. Oxygen:** The maturation of the chromophore on many FPs (particularly those derived from GFP) requires oxygen. Therefore, these FPs cannot be used in oxygen-deprived environments. Recently, a new GFP isolated from the Unagi eel was shown to mature independently of oxygen, making it suitable for use in anaerobic conditions.

**4. Maturation Time:** Maturation time is the time it takes the FP to correctly fold and create the chromophore. This can take anywhere from a few minutes after it is translated to a few hours. For example, superfolder GFP (sfGFP) and mNeonGFP can fold in <10 min at 37°C, mCherry takes ~15 min, TagRFP ~100 min and DsRed ~10 hours.

**5. Temperature:** FPs maturation times and fluorescent intensity can be affected by the temperature. For instance, enhanced GFP (EGFP) was optimized for 37°C, and is therefore most suited for mammalian or bacteria studies, whereas GFP<sup>S65T</sup> is better suited for yeast studies (24-30°C).

**6. Brightness:** Brightness is calculated as the product of extinction coefficient and quantum yield of the protein, divided by 1000. In many cases the brightness is compared to that of EGFP which is set as 1. Some proteins are very dim (e.g. TagRFP657, which has a brightness of 0.1) and this should be taken into account.

**7. Photostability:** Fluorescent molecules get bleached (i.e. lose the ability to emit light) after prolonged exposure to excitation light. Photostability can be as short as 100ms (EBFP) or as long as 1 hour (mAmetrine1.2). However, for most FPs it is a few seconds to a few minutes.

**8. pH Stability:** This parameter is important if you are planning to express the FP in acidic environments (e.g. yeast cytosol, which is slightly acidic, or synaptic vesicles). Some FPs have different ex/em spectra (e.g. mKeima) or change fluorescent intensity upon pH changes (e.g. pHluorin, pHTomato).

## WHICH FLUORESCENT PROTEIN SHOULD I CHOOSE? (CONT'D)

---

Keep this list handy to plan your next experiment or to share with the next labmate who asks you, “Which fluorescent protein should I use?” And if you’re looking for more fluorescence microscopy tools and techniques to aid your work, head over to [greenfluorescentblog](#).

### Further Reading

1. Fluorescent proteins as biomarkers and Biosensors: Throwing color lights on molecular and cellular processes. Stepaneko et. al. (2008). *Curr. Protein. Pept. Sci.* 9(4):338. PubMed [PMID: 18691124](#).
2. Fluorescent proteins and their applications in imaging living cells and tissues. Chudakov et. al. (2010). *Physiol. Rev.* 90:1103. PubMed [PMID: 20664080](#).
3. Modern fluorescent proteins and imaging technologies to study gene expression, nuclear localization, and dynamics. Wu et. al. (2011). *Curr. Opin. Cell. Biol.* 23:310. PubMed [PMID: 21242078](#).
4. [FP guide](#) at Addgene
5. [Fluorescence Spectrum Viewer](#) from BD Bioscience
6. [Fluorescence SpectraViewer](#) at Invitrogen (Life Technologies) site
7. [Evrogen Spectra Viewer](#) at Evrogen
8. [ilovegfp](#) – a site with very comprehensive data sheets on many FP variants

## CHOOSING YOUR FLUORESCENT PROTEINS FOR MULTI-COLOR IMAGING

By Kurt Thorn of the Nikon Imaging Center at UCSF | Oct 9, 2014

---

A common requirement for live cell imaging experiments is the ability to follow multiple fluorescently tagged species simultaneously. To do so with fluorescent protein labels requires [multiple fluorescent proteins](#) whose excitation and emission spectra differ sufficiently for them to be imaged in distinct fluorescent channels on the microscope. With the proliferation of fluorescent proteins in recent years, there are many fluorescent protein combinations that can be imaged together, but this also means that the choice of fluorescent proteins requires some thought.

The first step in choosing fluorescent proteins for your multi-color imaging experiment is to be aware of what fluorescent proteins are available. With new fluorescent proteins being published every month, deciding on the best protein for a given application is a challenge. To help keep you abreast of the latest fluorescent proteins, I maintain an [interactive graph and table of the best fluorescent proteins currently available](#).

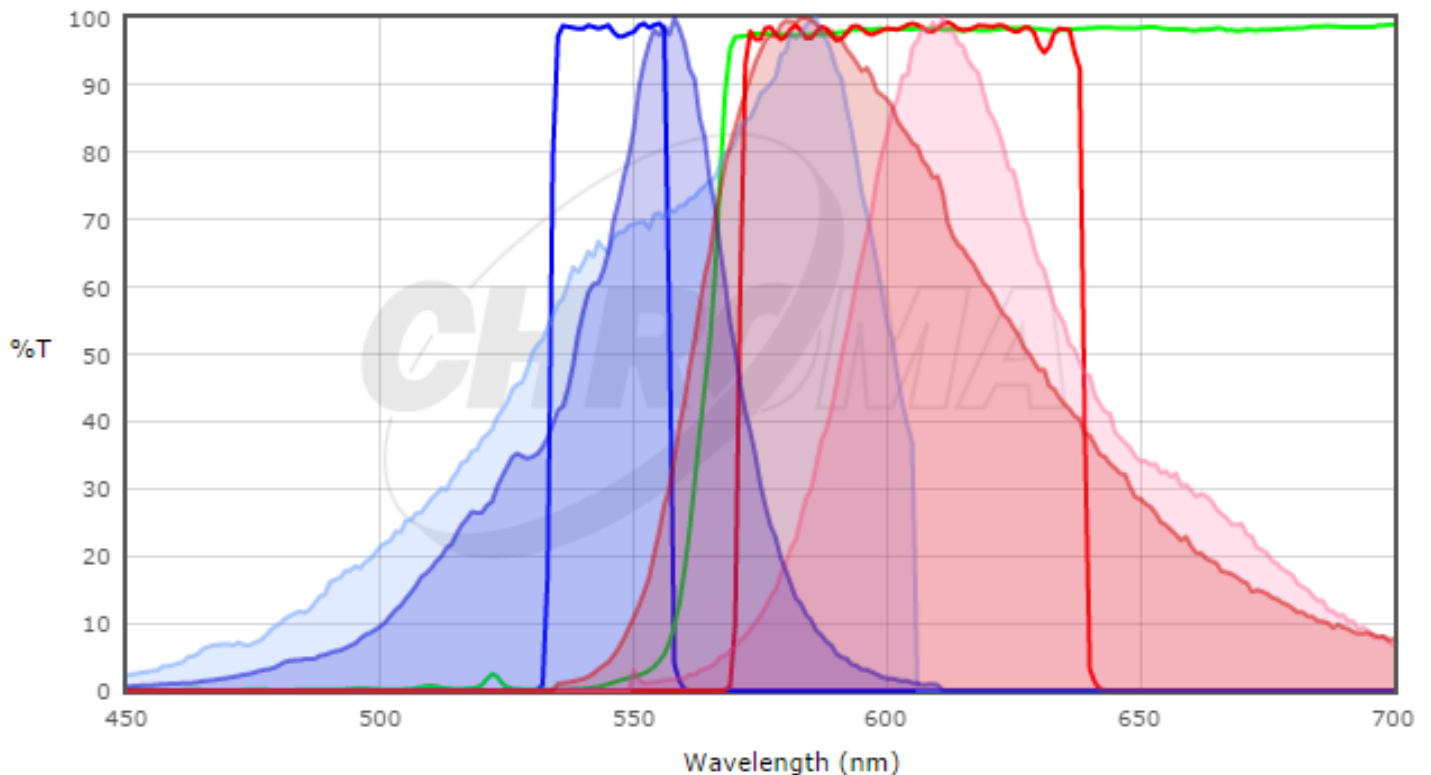
### Choosing Compatible Fluorescent Proteins

To choose a set of fluorescent proteins to be imaged together, you will need to consider the same factors as when choosing an individual fluorescent protein (brightness, photostability, and so on; see the [previous section](#) for more discussion of these factors). In addition, you will also need to choose fluorescent proteins that can be distinguished from one another and that can be imaged with the optics on the microscope(s) you intend to use. An accurate determination of whether two fluorescent proteins can be separated from each other requires knowledge of their excitation and emission spectra, but **a good rule of thumb is that both the peak excitation wavelength and peak emission wavelength of the two proteins should be separated by 50-60 nm**. For example, CFP (ex 430 nm / em 474 nm) and YFP (ex 514 nm / em 527 nm) can be imaged together but CFP and GFP (ex 488 nm / em 507 nm) show some crosstalk between the two fluorescent proteins. If you must image fluorescent proteins whose spectra overlap, there are techniques, like [spectral unmixing](#), which can be used to separate the fluorescent proteins, but these are beyond the scope of this eBook.

### Are Your Fluorescent Proteins Compatible with Your Microscope's Optics?

To determine if the fluorescent proteins you are interested in are compatible with your microscope optics, you will want to compare the excitation and emission spectra of your protein with the filter sets or lasers on your microscope. Ideally, you would like to have substantial overlap between the excitation and emission filters and the excitation and emission spectra of the protein, so that the protein is well excited by your microscope and the fluorescence emission of the protein is efficiently collected by the microscope. To compare the match between a fluorescent protein and a filter set, many filter set vendors provide tools to plot the fluorescence spectra of proteins and dyes and their filters (see [Chroma's](#), [Semrock's](#), or [Omega's](#)). While these don't contain all fluorescent proteins in common use (particularly not the most recently published ones), they can be a good starting point. In many cases it is sufficient to use a spectrum for a closely related protein, if you know that your protein of interest has a similar spectrum. For example, here's a screenshot from the Chroma Spectra Viewer comparing a standard Cy3 or Rhodamine filter set (Chroma #49004) to the spectra of both mCherry and TagRFP.

## CHOOSING YOUR FLUORESCENT PROTEINS FOR MULTI-COLOR IMAGING (CONT'D)



Here, the TagRFP spectrum is shown in the darker colors and the mCherry spectrum is shown in the lighter colors; excitation spectra are blue and emission spectra are red. Neither is a perfect match to the filter set but the excitation filter excites more of the peak of the TagRFP and the emission filter collects a larger fraction of the TagRFP emission than the mCherry emission. For this filter set, we would expect TagRFP to give a brighter signal than mCherry. In general, filter sets designed for Rhodamine / Cy3 will work better with shorter wavelength red fluorescent proteins like TagRFP or mRuby2 than longer wavelength proteins like mCherry. For background on fluorescence and filter sets, see the [Introduction to Fluorescence Microscopy](#) lecture at [iBiology](#).

### Commonly Used Filter Sets and Relevant Fluorescent Proteins

Commonly used filter sets for multicolor imaging include ones designed for CFP, YFP, and RFP or the Sedat Quad filter set, designed for DAPI / Fluorescein / Rhodamine / Cy5 (e.g. [Semrock's](#)) and the similar 4-laser combination on a confocal (405 / 488 / 561 / 640 nm). In our hands the best fluorescent proteins for imaging with this set are mTagBFP2, EGFP or one of the improved GFP variants, mRuby2 or TagRFP-T, and an infrared fluorescent protein such as iFP1.4 or iFP2.0. Beware that these infrared fluorescent proteins require biliverdin as a cofactor and so you may need to supplement your cells with biliverdin for maximal brightness. In mammalian cells, one of the improved folding variants of EGFP like mEmerald or Clover is probably best; mNeonGreen is an even newer green fluorescent protein that is supposed to be extremely bright. In *S. cerevisiae*, we've tested a number of green and red fluorescent proteins with this filter set and have [reported brightness measurements](#). Here, EGFP outperforms the improved folding variants, presumably due to the lower growth temperature.

## CHOOSING YOUR FLUORESCENT PROTEINS FOR MULTI-COLOR IMAGING (CONT'D)

---

This also suggests, however, that there is no single fluorescent protein optimal for all organisms and that if you want the brightest signal, you may need to try several proteins in your system of interest. Finally, in this set of proteins, the green and red proteins are generally the most detectable and so should be used to tag your least abundant proteins, with the blue and infrared channels used for more abundant proteins or marking subcellular compartments.

I hope this sheds some light on multicolor imaging with fluorescent proteins. With the right microscope and the right choice of fluorescent proteins, imaging four colors simultaneously should be pretty straightforward.

### Further Reading

1. Spectral imaging and its applications in live cell microscopy. Zimmermann, Timo, Jens Rietdorf, and Rainer Pepperkok. *FEBS letters* 546.1 (2003): 87-92. PubMed [PMID: 12829241](#).
2. Improved blue, green, and red fluorescent protein tagging vectors for *S. cerevisiae*. Lee, Sidae, Wendell A. Lim, and Kurt S. Thorn. *PloS one* 8.7 (2013): e67902. PubMed [PMID: 23844123](#).
3. Addgene's [Fluorescent Protein Guide](#)

# TIPS FOR USING FRET IN YOUR EXPERIMENTS

By Benoit Giquel, Addgene | Nov 5, 2014

---

The first time I heard about FRET during a journal club, my guitarist brain automatically thought about the raised element found on the neck of my guitar ... not really useful for a biologist you would say. The student was of course talking about the now well-known technique called [FRET, or Fluorescence \(Förster\) Resonance Energy Transfer](#), which allows the detection of molecules' interactions, modifications or dissociations *in situ*. Used since the mid-90s, this technique has revolutionized the way we apprehend molecular complexes and is still a very useful tool.

Like a guitar hero (that I'm not), FRET loves playing "live". Indeed, FRET was one of the first techniques that enabled the measurement of single molecule interactions in living cells using a microscope. Historically, molecular interactions were detected by indirect means often using probes with the potential to target several molecules. By analogy, it was like pointing out a group of students in a university hall but not knowing if these students knew or interacted with each other. FRET enhanced our perception about molecular interactions.

## What is FRET?

In their [2003 JCB paper](#), Sekar and Periasamy defined FRET as "a distance-dependent physical process by which energy is transferred non-radiatively from an excited molecular fluorophore (the donor) to another fluorophore (the acceptor) by means of intermolecular long-range dipole-dipole coupling." The emission of the acceptor fluorophore can be measured using microscopy techniques. FRET measurement sensitivity makes it suitable for studying interactions within living cells. By coupling fluorophores to proteins, pioneers of this technique were able to directly detect protein/protein interactions in living cells. Since then, FRET has also been used to measure conformational changes, cleavage activity using [FRET-based biosensors](#), and interactions between DNA and proteins. FRET, in theory, is an easy technique, but it is very important to follow some simple rules to avoid common pitfalls and to use it at its best.

## What Are the Parameters that Affect FRET?

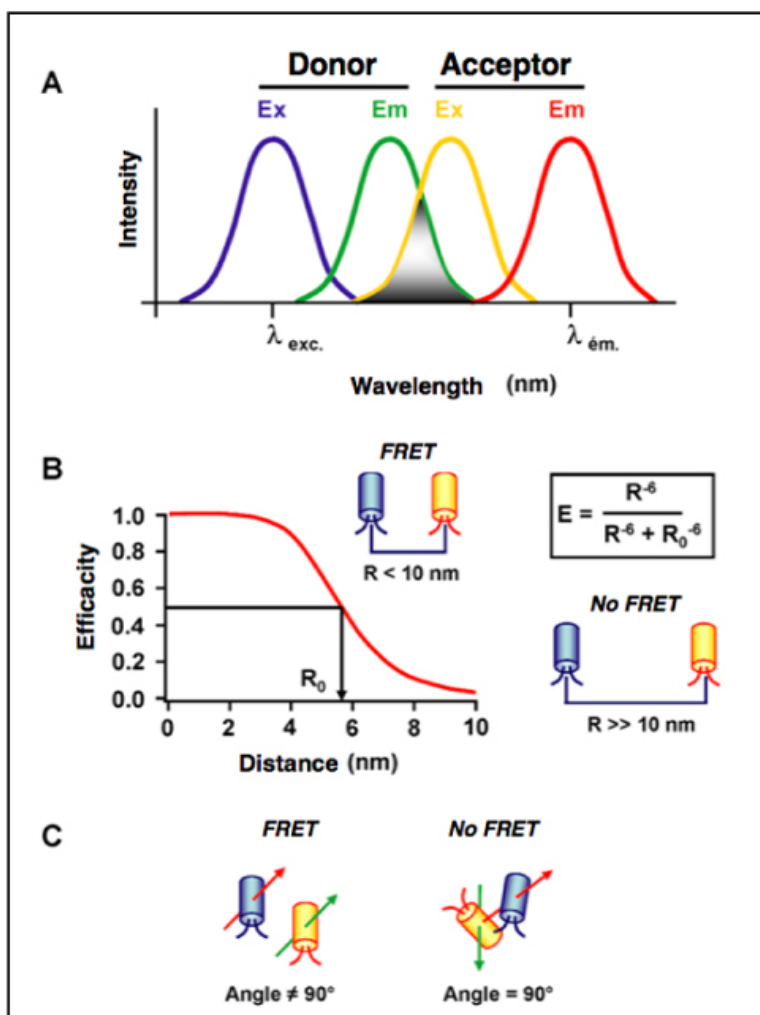
FRET occurs when the two fluorophores used are in close vicinity. Thus, the **distance between the two fluorophores and their orientation to one another** can affect FRET. When it comes to studying an unknown interaction between two proteins, these parameters are difficult to overcome but they have to be considered when analyzing data from your FRET experiments. You may have to design several different constructs to find one that could be used for the purpose of your studies. These parameters are less important when designing a biosensor, as the distance between the two fluorophores and their spatial orientation will be fixed.



## TIPS FOR USING FRET IN YOUR EXPERIMENTS (CONT'D)

The **quantum yield of the donor** (number of emitted photons per absorbed photons) and the **extinction coefficient of the acceptor** (linking the quantity of absorbed light at a given wavelength to the concentration of fluorophore in solution) are two additional parameters that can affect FRET efficiency. These issues can be overcome by choosing a complementary [pair of fluorophores](#). To maximize the FRET signal, you should choose the highest quantum yield donor, the highest absorbing acceptor, and fluorophores with significant overlap in their spectra. The pair CFP-YFP was the first to be used to study protein-protein interactions and several other pairs have been used since including: [mCerulean/mVenus](#), [mCerulean/Amber](#), [mCerulean/SYFP2A](#), and [mTurquoise/mVenus](#) among others. CFP-YFP is still one of the best and most used pairs to measure FRET.

The table below lists plasmids that can be used to create your choice of fluorescent fusion protein with your gene of interest:



Modified from Wikipedia, original author Maurel Damien. Accessed 11/4/2014.

**Table 1: Plasmids that Can Be Used to Make Fluorescent Protein Fusions**

Plasmid	Color	Expression	Description
<a href="#">pPROEX Aqua</a>	Cyan	Bacterial	Expresses Aquamarine with N-terminal His tag
<a href="#">pAquaN1</a>	Cyan	Mammalian	Expresses mammalian optimized Aquamarine
<a href="#">mCerulean N1</a>	Cyan	Mammalian	Expresses a gene of interest fused to the N-terminus of monomeric Cerulean
<a href="#">mCerulean C1</a>	Cyan	Mammalian	Expresses a gene of interest fused to the C-terminus of monomeric Cerulean
<a href="#">mTurquoise2</a>	Cyan	Mammalian	Constructs to target mTurquoise2 to various subcellular compartments
<a href="#">pCEP4CyPet-MAMM</a>	Cyan	Mammalian	Expresses mammalian optimized CyPet

## TIPS FOR USING FRET IN YOUR EXPERIMENTS (CONT'D)

**Table 1: Plasmids that Can Be Used to Make Fluorescent Protein Fusions (cont'd)**

Plasmid	Color	Expression	Description
<a href="#">pCyPet-His</a>	Cyan	Bacterial	Expresses CyPet with C-terminal His tag
<a href="#">SCFP3A</a>	Cyan	Mammalian	Expresses a gene of interest fused to the C-terminus of SCFP3A
<a href="#">Amber N1</a>	Yellow	Mammalian	Expresses a gene of interest fused to the N-terminus of Amber
<a href="#">Amber C1</a>	Yellow	Mammalian	Expresses a gene of interest fused to the C-terminus of Amber
<a href="#">mVenus N1</a>	Yellow	Mammalian	Expresses a gene of interest fused to the N-terminus of monomeric Venus
<a href="#">mVenus C1</a>	Yellow	Mammalian	Expresses a gene of interest fused to the C-terminus of monomeric Venus
<a href="#">pCEP4YPet-MAMM</a>	Yellow	Mammalian	Expresses mammalian optimized YPet
<a href="#">pYPet-His</a>	Yellow	Bacterial	Expresses YPet with C-terminal His tag
<a href="#">SYFP2</a>	Yellow	Mammalian	Expresses a gene of interest fused to the C-terminus of SYFP2
<a href="#">Clover</a>	Green	Mammalian	Expresses Clover (a GFP variant) commonly used with mRuby2
<a href="#">pLSSmOrange-N1</a>	Orange	Mammalian	Expresses a gene of interest fused to the N-terminus of LSSmOrange
<a href="#">pLSSmOrange-C1</a>	Orange	Mammalian	Expresses a gene of interest fused to the C-terminus of LSSmOrange
<a href="#">mRuby2</a>	Red	Mammalian	Expresses mRuby2 (a RFP variant) commonly used with Clover
<a href="#">pGWF1</a>	Cyan & Yellow	Bacterial	Gateway-compatible vector to express a gene of interest fused between ECFP and Venus

## TIPS FOR USING FRET IN YOUR EXPERIMENTS (CONT'D)

---

**Other issues that can affect FRET measurements include:** the brightness of a fluorophore pair, donor:acceptor stoichiometry, and cross-talk between the two fluorophore colours. Experts in FRET recommend using two fluorophores that have a similar brightness - even though in the most popular pair, CFP has five-fold less brightness than YFP. The stoichiometry of donor:acceptor is difficult to address, especially when you are investigating unknown molecular complexes, but it is a parameter that you have to keep in mind when analyzing FRET results. The cross-talk between the two fluorophores is linked to their excitation spectrum overlap. If you choose a pair too close to each other in the spectrum, you can easily directly excite the acceptor with the laser used to excite the donor. In that case the cross-talk is high and your background signal may be higher than the signal generated by the energy transfer. On the contrary, if you choose a pair whose members are too far from each other, you don't have any cross-talk but you also don't have any resonance transfer. You have to find a balance between FRET efficiency and cross-talk (see figure above).

### Methods to Measure FRET for Cell Biology Studies

Several methods have been used over the past 20 years to measure FRET and there is not one that is better than another. It is often recommended by FRET experts to use as many measurement methods as feasible when first beginning to establish the FRET methodology for a given experiment. You can then choose the most efficient approach for your own system.

The simplest and the most popular one is the sensitized emission method, where the donor is excited by a specific wavelength of light and the signal is collected by using emission filters chosen for the donor fluorescence and the acceptor fluorescence. Additionally, this method could be the best option if there is no cross-talk between FRET pairs. Unfortunately cross-talk between fluorophores does exist in the real world and corrective approaches and appropriate [controls](#) are required to make this method useful for dynamic experiments in which FRET changes are large.

**Two other methods are commonly used to measure FRET:** the acceptor photobleaching method and the fluorescence lifetime imaging microscopy (FLIM) method.

The acceptor photobleaching method is simple but limited to a single measurement. This method is based on the fact that the donor is quenched when FRET occurs. By photobleaching the acceptor, you release the donor's quenching and the fluorescence of the donor is increased. This method is straightforward and quantitative, but it is destructive and cannot be used for dynamic measurements. Extra care should be taken so as not to destroy the donor molecule.

FLIM has been developed more recently and is the most rigorous method for measuring FRET. FLIM measures the fluorescence decay time of the donor. When FRET occurs between the pairs, donor fluorescence is quenched and the fluorescence decay time of the donor is shortened, allowing FLIM to give an unambiguous value of FRET efficiency. As you don't measure acceptor fluorescence, this method is also less sensitive to direct acceptor excitation artifacts and it is possible to use a non-fluorescent acceptor. It should be noted that FLIM is a slower imaging method, limiting its use in many FRET experiments. In addition, other environmental factors, such as pH or autofluorescence background, can change the fluorescence decay time and have to be taken into account when interpreting data.

## TIPS FOR USING FRET IN YOUR EXPERIMENTS (CONT'D)

---

FRET is 20 years old but as you can imagine not old-fashioned at all. With the development of new fluorescent pairs and the emergence of faster measurement systems, this technique still has many good days and good “gigs” ahead...No offense to BRET (Bioluminescence resonance energy transfer), CRET (Chemiluminescence resonance energy transfer), and other RET techniques which have never outclassed the Master.

### Further Reading

1. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. Sekar, Rajesh Babu, and Ammasi Periasamy. *The Journal of cell biology* 160.5 (2003): 629-633. [PubMed PMID: 12615908](#).
2. Addgene's [FRET pages](#)

# LUCIFERASE

By Jason Niehaus, Addgene | Jun 24, 2014



**Luciferases** are a class of enzymes capable of catalyzing chemical reactions in living organisms resulting in the emission of photons. The most familiar bioluminescent organism for most people is the firefly (*Photinus pyralis*) and perhaps not surprisingly it is also the most commonly used bioluminescent reporter. This beetle emits a yellow-green light with a peak emission at 560 nm. Shortly after the initial article describing the cloning of firefly luciferase was published in 1985, several studies utilized luciferase as a genetic reporter in plant and mammalian cells. Luciferase assays have since become a gold standard in gene expression analysis and a luciferase gene (one of many available to choose from) is now a common feature in reporter plasmids.

## Reporter Gene Assays

A reporter gene, such as luciferase, usually serves as an indicator of transcription within cells, where detection of the reporter protein or its enzymatic activity is measured. The effect of promoters or enhancer regions on gene expression can be determined by detection of the reporter in a specific assay, which ideally would have low background signal, high sensitivity and, of course, be quick, accurate, and safe. In the case of a luciferase assay, one measures photon emission resulting from the catalysis of a chemical reaction requiring luciferin, ATP, and oxygen as substrates. Production of photons by this bioluminescent reporter occurs slower than fluorescent-based methods, such as excitation of Green Fluorescent Protein (GFP) because of the nature of the chemical reaction compared to using a high-intensity laser to rapidly excite GFP. As a result of the different mechanisms to produce photons, chemiluminescent reporters are generally less bright than fluorescent proteins, but have the advantage of lower background levels and improved signal sensitivity since photons are simply measured – they are not required to initiate the reaction.

## Biological Sources of Luciferase

Although luciferase isolated from the firefly beetle (*Photinus pyralis*) is the most frequently used bioluminescent reporter, other luciferases have been identified in various species with differing kinetics, substrate requirements, and photon emission wavelengths. *Renilla* luciferase from the sea pansy (*Renilla reniformis*) emits a blue light with a peak emission at 480 nm upon catalyzing a chemical reaction with the substrate coelenterazine and is likely the second most commonly used luciferase. Other organisms from which luciferases have been cloned include click beetles (*Pyrophorus plagiophthalmus*), railroad worm (*Phrixothrix hirtus*), and a variety of marine crustaceans. Many of these luciferases have been subjected to further engineering to improve upon the properties of the native enzyme. Table 1 below compares some general advantages and disadvantages of luminescence versus fluorescence while table 2 describes some of the important details to consider when using various luciferases.

## LUCIFERASE (CONT'D)

### Table 1: Properties of Luminescence Versus Fluorescence

Property	Luminescence	Fluorescence
Source of Emitted Photons	Chemical reaction	High-energy photons
Kinetics of Photon Generation	Slower	Faster
Cofactors/Substrates	Required	Not Required
Signal Strength	Lower	Higher
Sensitivity	Higher	Lower
Background	Lower	Higher
Post-translational Modification	Not Required	Required
Photobleaching/Phototoxicity	Not susceptible	Susceptible
Subcellular Imaging	Improving	Well-established
High-throughput assays	Improving	Well-established

### Table 2: Properties of Native Luciferases

Organism	Peak Emission (nm)	Size (kDa)	Substrate	Cofactor	Secreted
Luminous shrimp ( <i>Oplophorus gracilirostris</i> )	455	19	Coelenterazine	None	Yes
Luminous Japanese ostracod ( <i>Cypridina noctiluca</i> ) Sea firefly ( <i>Cypridina hilgendorffii</i> )	465	62	<i>Cypridina</i> ( <i>Vargulin</i> ) luciferin	None	Yes
Sea pansy ( <i>Renilla reniformis</i> )	480	36	Coelenterazine	None	No
Deep sea coceopod ( <i>Gaussia princeps</i> )	480	20	Coelenterazine	None	Yes
Maine planktonic copepod ( <i>Metridia longa</i> )	480	24	Coelenterazine	None	Yes
North American firefly ( <i>Photinus pyralis</i> )	560	61	Luciferin	ATP	No
Japanese firefly ( <i>Luciola cruciate/lateralis</i> ) Italian firefly ( <i>Luciola italica</i> ) Southern Russian firefly ( <i>Luciola mingrelica</i> )	562	64	Luciferin	ATP	No
Jamaican click beetle ( <i>Pyrophorus plagiophthalmus</i> )	544-593	64	Luciferin	ATP	No
Railroad worms ( <i>Phrixothrix hirtus</i> )	630	64	Luciferin	ATP	No

# LUCIFERASE (CONT'D)

---

## Common Uses for Plasmids Expressing Luciferase

### *Luciferase Reporter Assay*

The typical luciferase-containing plasmid is most often used to investigate the effect of regulatory elements, such as promoters, enhancers and untranslated regions, or the effect of mutations of these regulatory elements on gene expression. The regulatory element (wild-type or mutant) is cloned upstream of the luciferase gene in the plasmid, which is then introduced into the desired [mammalian](#), [plant](#), or [bacterial](#) cells. Expression of luciferase as the reporter gene is measured via its ability to produce light, often with a luminometer or scintillation counter, and the activity measurement can then be used to quantify the effect of the regulatory element. Depending on the specific luciferase used, a compatible substrate must be provided for the enzyme to catalyze the light-producing reaction. Many luciferase assays require cell lysis as the most efficient means to disrupt the cell membrane and deliver the substrate; however, secreted luciferase or the use of alternative substrate reagents can permit measurements of luminescence from live cells.

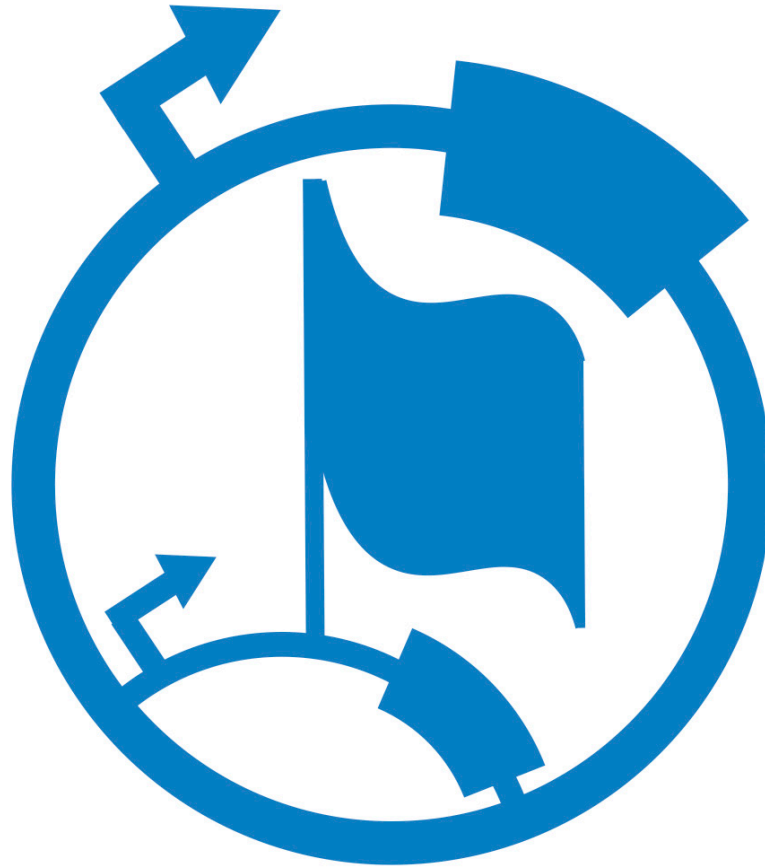
### *Monitor miRNA Knockdown*

Expression of luciferase can be monitored as a downstream measurement of miRNA efficiency on a target sequence. The gene of interest is [cloned into a plasmid](#) downstream of luciferase containing a stop codon, which is then transfected into the desired cells. A hybrid mRNA transcript containing luciferase and the gene of interest permits translation of functional luciferase, until the miRNA targets and degrades the mRNA sequence, resulting in a decrease in luciferase activity. This experimental protocol can readily screen miRNAs for effectiveness.

### *Analyze Cell Signaling Pathways*

Known cellular signaling pathways can be investigated in a high-throughput screen using luciferase activity as a measure of signaling activity. Multiple response elements are cloned into a plasmid upstream of a minimal promoter linked to luciferase. Activation of the pathway will result in luciferase expression and more complex cell treatments can be designed to further explore the pathway.

# CHAPTER 6: PLASMID TAGS

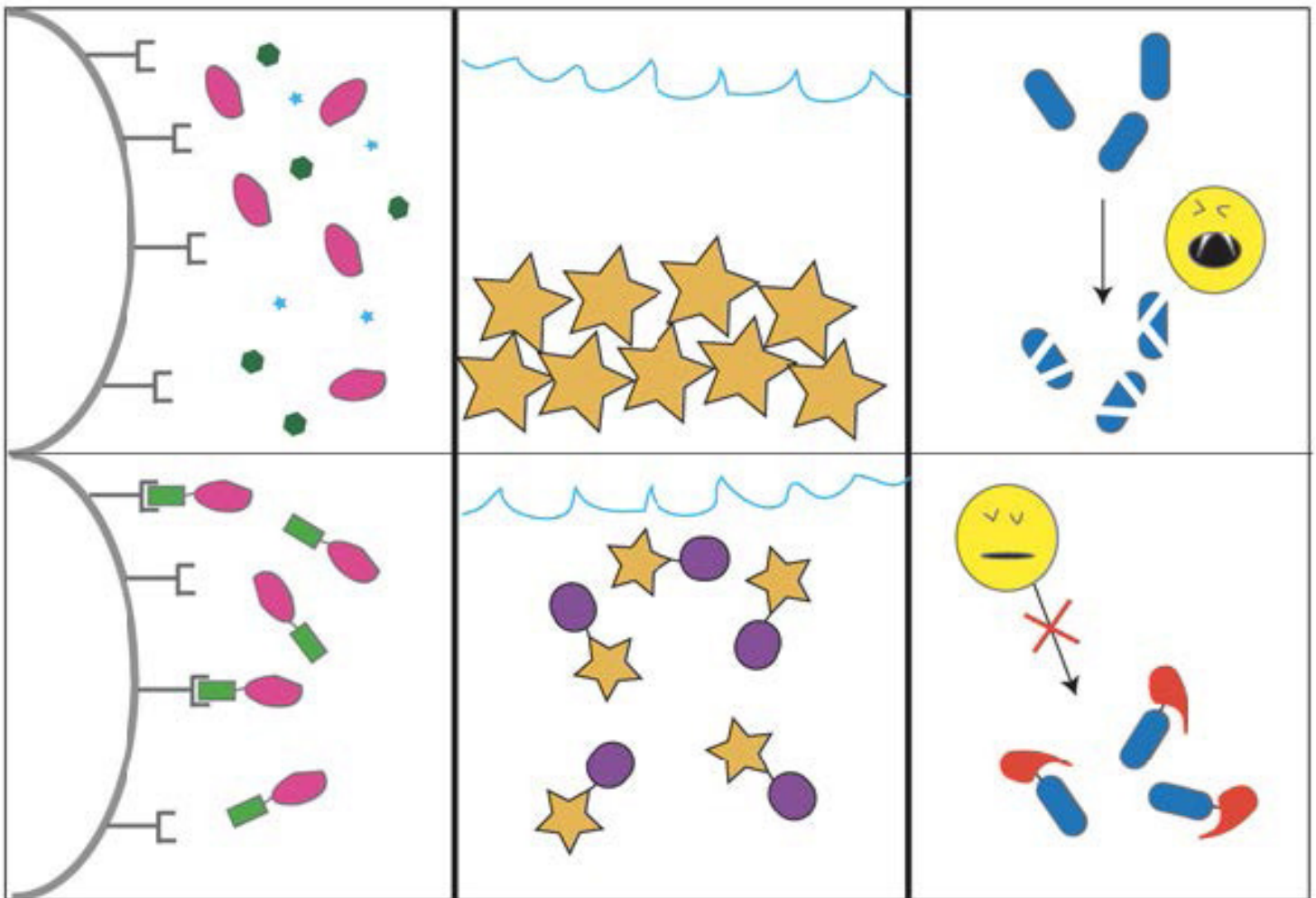




# PROTEIN TAGS

By Eric J. Perkins, Addgene | Dec 11, 2014

**Protein tags are usually smallish peptides incorporated into a translated protein.** As depicted in the accompanying cartoon, they have a multitude of uses including (but not limited to) purification, detection, solubilization, localization, or protease protection. We've already covered [GFP and its related fluorescent proteins](#), which are sometimes used as tags for detection; however, those constitute just one (admittedly large) class of common fusion protein tags. Biochemists and molecular biologists who need to overexpress and purify proteins can face any number of technical challenges depending on their protein of interest. After several decades of trying to address these challenges, researchers have amassed a considerable molecular tool box of tags and fusion proteins to aid in the expression and purification of recombinant proteins.



Different Uses for Protein tags. Left: Affinity tags. Middle: Solubility tags. Right: Cleavage tags. Image by Eric Perkins.

## PROTEIN TAGS (CONT'D)

### Tags for Stability and Solubility

What are some of the hurdles to overcome in order to overexpress a recombinant protein? It is not generally in a cell's best interest to overexpress a protein; energy and cellular resources are spent to make something the cell doesn't need to make. Eukaryotes and some bacteria deploy proteasomes to degrade what the cell might consider junk protein. Though there are a number of chemical and peptide-based proteasome inhibitors, glutathione S-transferase (GST), which can be fused to recombinant proteins for one-step purification with glutathione, can also protect against proteolysis.

That's one form of instability. Prokaryotes can also have a hard time folding eukaryotic proteins. You can get your bacteria to produce massive amounts of protein, but if it's not folded correctly, there's no point in crystallizing it or testing its function. Small ubiquitin-related modifier ([SUMO](#)) can help with folding and stabilization, as can maltose-binding protein ([MBP](#)). Overexpression can also lead to insolubility, and aggregated protein is not useful protein. MBP tags can help with solubility issues, but scientists may also choose to add smaller proteins, such as [Thioredoxin A \(TrxA\)](#) that improve disulfide bond formation in order to help keep a protein of interest soluble.

### Tags for Affinity and Purification

An affinity tag, generally a relatively small sequence of amino acids, is basically a molecular leash for your protein. If you're working with an uncharacterized protein, or a protein for which a good antibody has not been developed (and just because your protein has a commercially available antibody, that doesn't mean it's a *good* one), then your first step towards detecting, immunoprecipitating, or purifying that protein may be to fuse an affinity tag to it. The [FLAG](#), hemagglutinin antigen ([HA](#)), and [c-myc](#) tags have been the workhorses of the affinity tag world for years, and deciding on which one to use will depend on your application (see table below). The antibodies available for these tags really *are* good and can be used for western blots, IP, and affinity purification.

Arguably the simplest affinity tag is the polyhistidine ([His](#)) tag. Small and unlikely to affect function, His-tagged proteins can be purified using metal-affinity chromatography, usually using a Ni<sup>2+</sup> column. Like other affinity tags, a His tag can be fused to either the N- or C-terminus of a protein. Unlike other epitope tags – which when doubled or tripled increase the tag size quickly – modifying the length a polyhistidine tract does not greatly alter the size of the tag.

**Table 1 – Common Protein Tags**

Tag	Epitope	Mass (kDa)	Function	Notes
CBP	KRRWKKNFIAVSAANRFKKISSSGAL	4	Affinity and Purification	Binding and elution steps use very moderate buffer conditions

## PROTEIN TAGS (CONT'D)

**Table 1 – Common Protein Tags (CONT'D)**

Tag	Epitope	Mass (kDa)	Function	Notes
FLAG	DYKDDDD or DYKDDDDK or DYKDDDK	1	Affinity and Purification	Good for antibody-based purification; has inherent enterokinase cleavage site
GST	Large Protein	26	Purification and Stability	Good for purification with glutathione; protects against proteolysis, but may reduce solubility
HA	YPYDVPDYA or YAYDVPDYA or YDVPDYASL	1.1	Affinity	Frequently used for western blots, IP, co-IP, IF, flow-cytometry; can occasionally interfere with protein folding
HBH	HHHHHHAGKA GEGEIPAPLA GTVSKILVKE GDTVKAGQTV LVLEAMKMET EINAPTDGKV EKVLVKERDA VQGGQGLIKI GVHHHHHH	9	Combo	Consists of a bacterially-derived in-vivo biotinylation signaling peptide (Bio), flanked by hexahistidine motifs (6xHis)
MBP	Large Protein	40	Solubility and Purification	Can improve solubility and folding of eukaryotic proteins in prokaryotes; single step purification with amylose, but wicked huge
Myc	EQKLISEEDL	1.2	Affinity	Frequently used for western blots, IP, co-IP, IF, flow-cytometry, but rarely used for purification as elution requires low pH
poly His	HHHHHH	0.8	Affinity and Purification	Very small size, rarely affects function

## PROTEIN TAGS (CONT'D)

**Table 1 – Common Protein Tags (Cont.)**

Tag	Epitope	Mass (kDa)	Function	Notes
S-tag	KETAAAKFERQHMS	1.8	Solubility and Affinity	Abundance of charged and polar residues improves solubility; good for antibody-based detection
SUMO	~100 amino acid protein	12	Stability	At N-terminus, promotes folding and structural integrity; cleavable. Not great for purification; too cleavable in eukaryotes
TAP	GRRIPGLINP WKRRWKKNFI AVSAANRFKK ISSSGALDYD IPTTASENLY FQGEFGLAQH DEAVDNKFNK EQQNAFYEIL HLPNLNEEQR NAFIQSLKDD PSQSANLLAE AKKLNDAAQAP KVDNKFNKEQ QNAFYEILHL PNLNEEQRNA FIQSLKDDPS QSANLLAEAK KLNDAAQAPKV DANHQ	21	Combo	See text
TRX	MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QGKLTVAKLN IDQNPGTAPK YGIRGIPTLL LFKNGEVAAT KVGALSKGQL KEFLDANLAG SGSGHMHHHH HHSSGLVPRG	12	Solubility	Assists in proper folding
V5	GKPIPPLLGLDST	1.4	Affinity and Purification	Good for antibody-based purification

### Combo and Cleavage Tags

Frequently, a single tag is not enough. What if you need one tag to increase solubility and one tag for purification? Or you want to combine a fluorophore with a tag that localizes your protein to the nucleus? Or you want multiple rounds of purification to get your protein as pure as possible? Vectors that offer different combinations of [tags are readily available](#), and though adding too many tags and fusion proteins to your protein of interest will eventually get ridiculous (you generally don't want more tag than protein), 2-3 tags is increasingly common. Tandem affinity purification (TAP) once referred specifically to a combo tag comprised of a calmodulin binding peptide (CBP), a TEV cleavage site (more on that in a moment), and 2 ProtA IgG-binding domains. TAP has since come to encompass several other tag combinations, though frequently those combinations still include at least one element from the original TAP tag. The terms dual-labeling and dual-tagging are also used. Due

## PROTEIN TAGS (CONT'D)

to their small size and the ease with which they can be added to a purification scheme, His tags are frequently combined with other tags for dual-labeling.

The problem with all these tags is that many of them serve a one-time purpose, and you don't necessarily want them to stick around after that purpose has been served. At this point, proteases can be your friend rather than your enemy. Two common tags (SUMO and FLAG) are cleaved by specific proteases without requiring the addition of an independent cleavage recognition site. In fact, SUMO cannot be used in eukaryotes because there is already too much SUMO protease around, but it is convenient when used with purified protein since the enzyme cleaves the SUMO tag in the same manner as it would have in the context of a cell. FLAG tags can be cleaved by enterokinase, which recognizes DDDDK<sup>X</sup>, cleaving after the lysine. The efficiency of this cleavage depends on the identity of X.

A number of other proteases are available, but scientists would need to incorporate their recognition sites into their protein tag in order to use them effectively. One of the best optimized is the tobacco etch virus (TEV) protease. A TEV protease cleavage site is frequently placed between two tags being used for two rounds of purification, with the cleavage reaction taking place between column runs. The TEV protease itself, with various mutations used to increase its stability activity, can be readily purified using plasmids [found in this paper](#) (available at Addgene).

**Table 2 – Protease Recognition Sites Commonly Used with Tags**

Protease	Recognition Site	Notes
TEV	ENLYFQS	Cleaves between the Gln and Ser residues
Thrombin	LVPRGS	Cleaves between Arg and Gly residues
PreScission	LEVLFQGP	Cleaves between the Gln and Gly residues

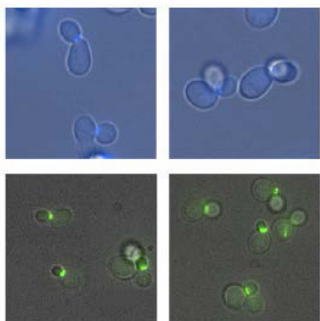
This section is not a comprehensive guide to all tags, but rather a quick overview of why scientists use tags, with a few time-tested tags and fusion proteins as examples. The tables list more common tags than are described in the post, but have been categorized to help you better assess their function. More detailed information and some protocols can be found in the references provided.

### Further Reading

1. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Young CL, Britton ZT, Robinson AS. *Recom Biotechnology Journal* 2012 , (7): 620-634. PubMed [PMID: 22442034](#).
2. A complete list of free handbooks for protein purification is provided by [GE Healthcare Life Sciences](#).

# TAG YOUR FAVORITE YEAST GENES WITH EASE

By Julian Taylor-Parker, Addgene with contributions from Marcy Patrick, Addgene | Nov 19, 2013



Modified from Lee S, et al. (2013) *PLoS ONE* 8(7): e67902.

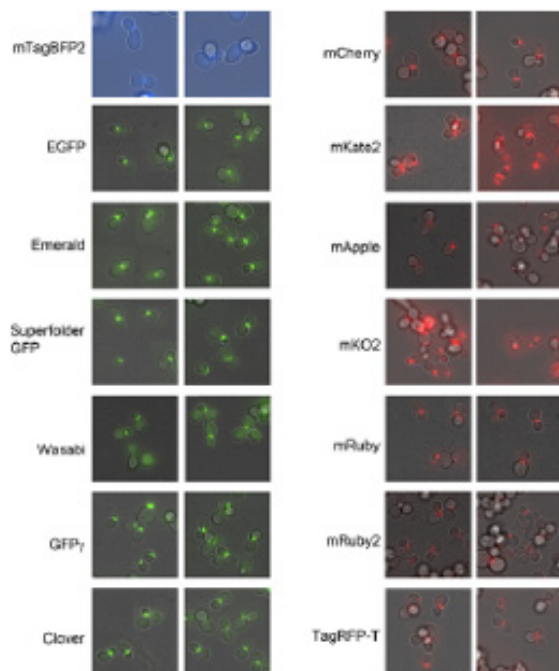
**Homologous recombination is the process by which nearly all domains of life repair genomic damage, specifically double strand breaks.** Researchers have long taken advantage of this natural process to integrate protein tags into the genomes of *S. cerevisiae* and *S. pombe*. The protocol is surprisingly simple, requiring only a PCR product containing the modifying sequence flanked by approximately 50 base pairs of sequence homologous to the chromosomal site of insertion. The linear PCR product is introduced into the cell by direct transformation. A given insert will typically contain both a protein modification sequence and a selectable gene product for isolation of successful transformants.

Addgene distributes several ready-to-use, modular plasmids, combining fluorescent tags, epitope tags, protease sites, and selection markers. These are especially useful in protein complex studies where tagging of multiple protein products is desired, as multiple selection markers can ensure that all desired tags have been integrated. Simply design your amplification primers with the desired targeting homology—in frame, of course—and start tagging!

## Yeast-Optimized Fluorophores for Imaging

Many imaging studies rely on direct fusion of [fluorescent proteins](#) (FPs) to a yeast gene of interest. These fluorescently tagged genes are expressed under native conditions and allow scientists to not only track the abundance, movement, and localization of individual proteins, but also investigate protein-protein interactions via [FRET](#). Sidae Lee, [Wendell Lim](#), and [Kurt Thorn](#) at UCSF have recently developed a series of blue, green, and red FPs that are codon optimized specifically for expression in yeast. These tagging vectors are based on previously described pFA6a-link vectors and include a Kan, SpHIS5, or CaURA3 selection marker. Lee *et al.* assessed many of these fluorescent tags in *S. cerevisiae*, looking at their performance in categories such as brightness, stability, and disruption of the tagged protein. Based on their findings, the authors recommend optimal FP combinations for use in yeast imaging, categorized by specific filter sets and experimental output requirements. Select from these [yeast-optimized fluorophore tagging vectors](#) for your single- or multi-color imaging experiments.

If you're looking for a great resource about imaging techniques, check out [Kurt Thorn's microscopy blog](#).



A variety of fluorescent proteins expressed in yeast. Adapted from Lee S, et al. (2013). *PLoS ONE* 8(7): e67902

## TAG YOUR FAVORITE YEAST GENES WITH EASE (CONT'D)

---

### Interested in Epitope Tags?

Others may be interested in **attaching epitope tags** to their genes of interest, allowing for easy capture and detection of proteins and complexes, without the artifacts sometimes associated with plasmid-based overexpression. [Tim Formosa](#), at the University of Utah, has built a complete [collection of yeast tagging modules](#) with each possible combination of protease site (TEV or PreScission), epitope tag (12xHis, 2xStrep, 3xFlag, Protein A, or V5), and selection marker (KanMX, HphMX, or His3MX). Each PCR product from this collection will yield an insert with the format: (protease site)-6xGly linker-(epitope tag)-ADH1 terminator-(selection marker). Additionally, Dr. Formosa has deposited six plasmids with a multiple cloning site in place of the epitope tag for creation of your own unique protein fusions. This collection is ideal for **tandem affinity purification** of protein complexes.

[John Pringle](#) and [Jürg Bähler](#) have deposited a large collection of plasmids with Addgene for genome modification in yeast. These were developed by Dr. Pringle's former lab at UNC Chapel Hill. Bähler *et al.* describe a modular collection of [plasmids for a wide variety of genome modifications in S. Pombe](#), including full and partial gene **deletion**, **overexpression** (by promoter substitution), and tagging at either the N- or C-terminus (3xHA, 13xMyc, GST, or GFP). Longtine *et al.* describe a complementary set of [plasmids for use in S. cerevisiae](#), with the additional benefit of multiple selection markers for combining modifications within a single strain.

In addition to the collections featured above, many other modular yeast tagging systems have been developed in the labs of [Anne Robinson](#), [Eishi Noguchi](#), and [Melissa Moore](#), to name a few.

### Further Reading

1. Improved blue, green, and red fluorescent protein tagging vectors for *S. cerevisiae*. Lee S, Lim WA, Thorn KS. *PLoS One*. 2013 Jul 2;8(7):e67902. PubMed [PMID: 23844123](#).
2. Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. Bähler J, Wu JQ, Longtine MS, Shah NG, McKenzie A 3rd, Steever AB, Wach A, Philippsen P, Pringle JR. *Yeast*. 1998 Jul;14(10):943-51. PubMed [PMID: 9717240](#).
3. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. Longtine MS, McKenzie A 3rd, Demarini DJ, Shah NG, Wach A, Brachet A, Philippsen P, Pringle JR. *Yeast*. 1998 Jul;14(10):953-6. PubMed [PMID: 9717241](#).

# CHAPTER 7: GENOME ENGINEERING



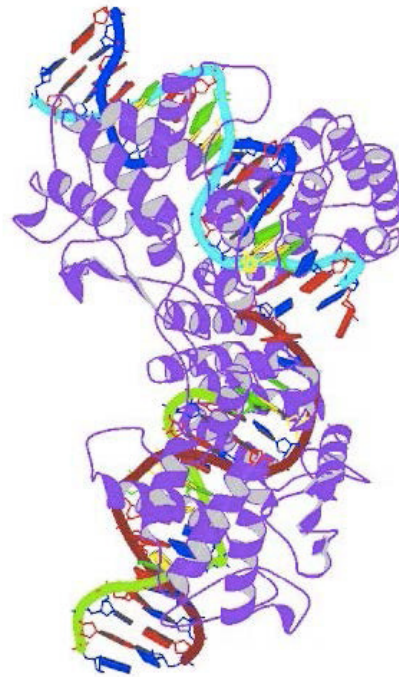


# INTRODUCTION TO GENOME ENGINEERING

By Mary Gearing, Addgene | August 26, 2015

Genome engineering has made it possible to not only dissect complex gene interactions, but also to build new pathways through [synthetic biology](#). The past few decades have seen tremendous advances in both the number and feasibility of genome engineering techniques, many of which are available from Addgene.

Genome engineering was born in the late 1970s, when multiple groups showed that exogenous DNA could be taken up by yeast or bacteria and randomly integrated into the genome. Subsequent work showed that this process could also occur in a targeted fashion. Addgene depositor [Mario Capecchi](#) realized that DNA microinjection into a cell's nucleus would stimulate cellular homologous recombination, permitting targeted genome modification. In 1989, he, Martin Evans and Oliver Smithies created the first knockout mouse, a watershed moment for genome engineering.



*Cre-recombinase complexed to DNA Guo et al. 1997*

## Cre-Lox

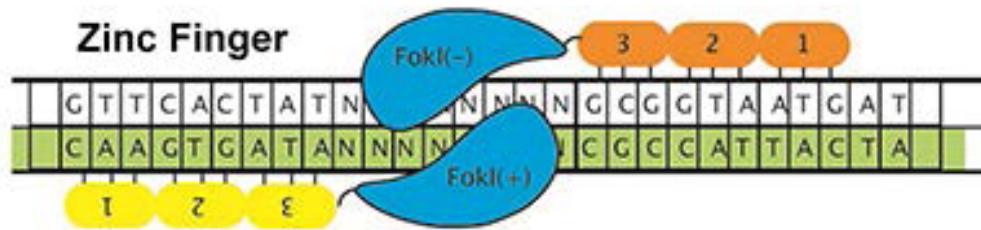
The late 1980s also marked the introduction of Cre-lox recombination, a system derived from P1 bacteriophage now widely used to control gene expression. Today, Cre recombinase under the control of various promoters, or in its inducible form, provides sophisticated spatiotemporal control of gene expression, especially in mouse transgenics.

## Homologous Recombination and the Journey Towards CRISPR/Cas9

Homologous recombination is a cornerstone of genome engineering, but with the caveat that it occurs at low frequencies, limiting editing efficiency. To improve editing rates, researchers hijacked the function of endonucleases, which create difficult-to-repair DNA double-stranded breaks. Targeting these enzymes to a given locus results in DNA cleavage and forces the cell to undergo either non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ occurs if no DNA repair template is provided, and its error-prone nature often leads to inactivating mutations. In the case of HDR, a repair template specifies desired genomic modifications, enabling precise editing. In addition to introducing point mutations or recombination sites, repair templates can also be used to introduce a gene of interest into a given locus.

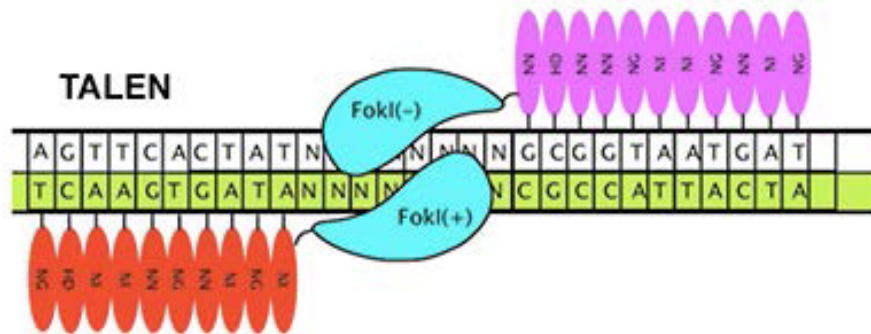
# INTRODUCTION TO GENOME ENGINEERING (CONT'D)

## Zinc Finger Nucleases



[Zinc finger nucleases \(ZFNs\)](#) represent the first step towards efficient, targeted nucleases. To create ZFNs, a series of zinc fingers is designed to bind to a specific genomic locus, and subsequently fused to FokI nuclease. Paired ZFNs recognizing two adjacent sites cleave DNA, initiating HDR. The utility of ZFNs is limited by their long synthesis time and nonmodular assembly process. Although computational tools helped improve targeting, it's not possible to design suitable ZFN pairs for every genomic locus.

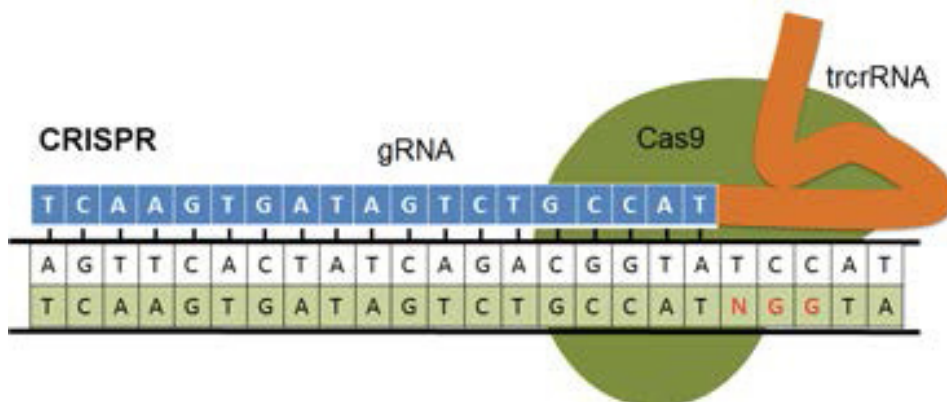
## TALENs



First reported in 2011, [TALENs](#) represented a huge step forward for genome engineering. This modular system is based on TAL effector DNA binding proteins, isolated from *Xanthomonas spp*, fused to FokI endonuclease. Whereas it might take an experienced scientist six weeks to make a ZFN, a newbie could make a TALEN in just a few days! TALEN technology was rapidly adopted by the research community, with the [Golden Gate TALEN kit](#) becoming Addgene's most popular kit. The customizable DNA-binding properties of TALENs also enabled the design of custom transcription factors to modulate gene expression.

# INTRODUCTION TO GENOME ENGINEERING (CONT'D)

## CRISPR



Just when we thought genome engineering couldn't get any better, along came the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 system. CRISPR is an important component of the bacterial immune system that allows bacteria to remember and destroy phages. In genome engineering applications, Cas9 endonuclease is targeted by guide RNA (gRNA) sequence homology to a given locus, where it induces a double stranded break. Like ZFNs and TALENs, CRISPR/Cas9 employs HDR, but the use of RNA to specify editing makes the system much less expensive and time-consuming, as well as more precise and scalable. For this reason, CRISPR/Cas9 has proved to be incredibly valuable for high-throughput genome engineering. CRISPR/Cas9 can also target multiple loci in one organism, and like TALENs, the system has also been adapted for other functions. CRISPRs are even more accessible to the research community than TALENs, with new papers using this technology published every week.

[Check out the CRISPR 101 eBook for More Information on CRISPR and Its Applications](#)

## Other Important Genome Engineering Tools

With the success of CRISPR, it's easy to forget about other genome engineering methods...but you shouldn't! Another HDR-based method, [recombineering](#), is commonly used in *E. coli* to make edits to the genome or a bacterial artificial chromosome (BAC). 50 bp homology arms flanking the repair template specify the site of recombination, catalyzed by phage recombinases. Since repair templates can be quickly generated using PCR, recombineering is easily scalable just like CRISPR. In addition to its applications in bacterial genome engineering, recombineering is also useful for creating BAC-based repair templates for other HDR genome engineering methods.

Outside of HDR, an exciting new gene transfer method is the [Sleeping Beauty transposon](#), reconstructed from fossils of ancient fish. Transposons are mobile DNA elements, and are thus ideal for gene transfer. This system is divided across two plasmids, one containing the gene of interest (GOI) flanked by inverted repeats, the other carrying the transposase. Following cotransfection, the transposase cleaves the GOI from the plasmid and facilitates a double strand break to allow the GOI to integrate into the genome. Sleeping Beauty was named 2009's Molecule of the Year, and it represents a robust alternative to viral vector-mediated gene transfer.

# INTRODUCTION TO GENOME ENGINEERING (CONT'D)

---

## Conclusion

With multiple robust and efficient genome engineering methods at our fingertips, we have entered a Golden Age of genome engineering. Current work focuses on refining these techniques to ensure high specificity and activity, whatever the desired target locus (or loci) may be, with the hope that these methods will be useful clinically. What we at Addgene find most exciting is the democratization of genome engineering, which has and will continue to allow researchers all over the world to use these tools in their research.

## Further Reading

1. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Guo, Feng, Deshmukh N. Gopaul, and Gregory D. Van Duyne. “.” *Nature* 389.6646 (1997): 40-46. PubMed [PMID: 9288963](#).
2. Addgene's [Genome Engineering Guide](#)

# CRE-LOX

By A Max Juchheim, Addgene | Jan 13, 2015

Previously, we examined a number of important plasmid elements – [promoters](#), [origins of replication](#), [protein tags](#), and [antibiotic resistance markers](#) (just to name a few). Here, we're going to take a look at a very interesting tool that can be used for **creating** (excuse the pun) specific, targeted DNA modifications in transgenic animals, embryonic stem cells, and/or tissue-specific cell types: Cre-lox recombination.

## What is Cre-Lox?

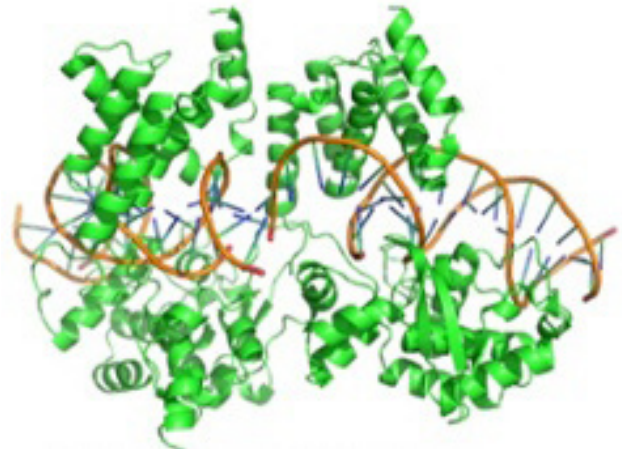
The Cre-lox system is a technology that can be used to induce site-specific recombination events. The system consists of two components derived from the P1 bacteriophage: the Cre recombinase and a loxP recognition site. The P1 bacteriophage uses these components as part of its natural viral lifecycle, and researchers have adapted the components for use in genome manipulation.

Cre recombinase, originally named because it “causes **re**combination” (although later referred to as the “**c**yclization **re**combinase”), is a 38 kDa protein responsible for intra- and inter-molecular recombination at loxP recognition sites. A key advantage of the system is that Cre acts independently of any other accessory proteins or co-factors, thus allowing for broad applications in a variety of experiments.

LoxP (**l**ocus **o**f **X**(cross)-over in **P**1) sites are 34-base-pair long recognition sequences consisting of two 13-bp long palindromic repeats separated by an 8-bp long asymmetric core spacer sequence. The asymmetry in the core sequence gives the loxP site directionality, and the canonical loxP sequence is ATAACTTCGTATAGCATACAT-TATACGAAGTTAT. The loxP sequence does not occur naturally in any known genome other than P1 phage, and is long enough that there is virtually no chance of it occurring randomly. Therefore, inserting loxP sites at deliberate locations in a DNA sequence allows for very specific manipulations as discussed below.

## How Does it Work?

As alluded to above, the Cre recombinase catalyzes a site specific recombination event between two loxP sites, which can be located either on the same or on separate pieces of DNA. Each 13 bp repeat sequences on a single loxP site are recognized and bound by a Cre protein, forming a dimer. The two loxP sites then align in a parallel orientation, allowing the four Cre proteins to form a tetramer. A double-strand DNA break occurs within the core spacer of each loxP site and the two strands are ligated, resulting in the reciprocal crossover event.



*Cre recombinase dimer bound to DNA. Image released to public domain. Source: Wikipedia accessed 17/2015. Author: Williamog1990.*

## CRE-LOX (CONT'D)

This event can have three general outcomes based on the location and orientation of the loxP sites:

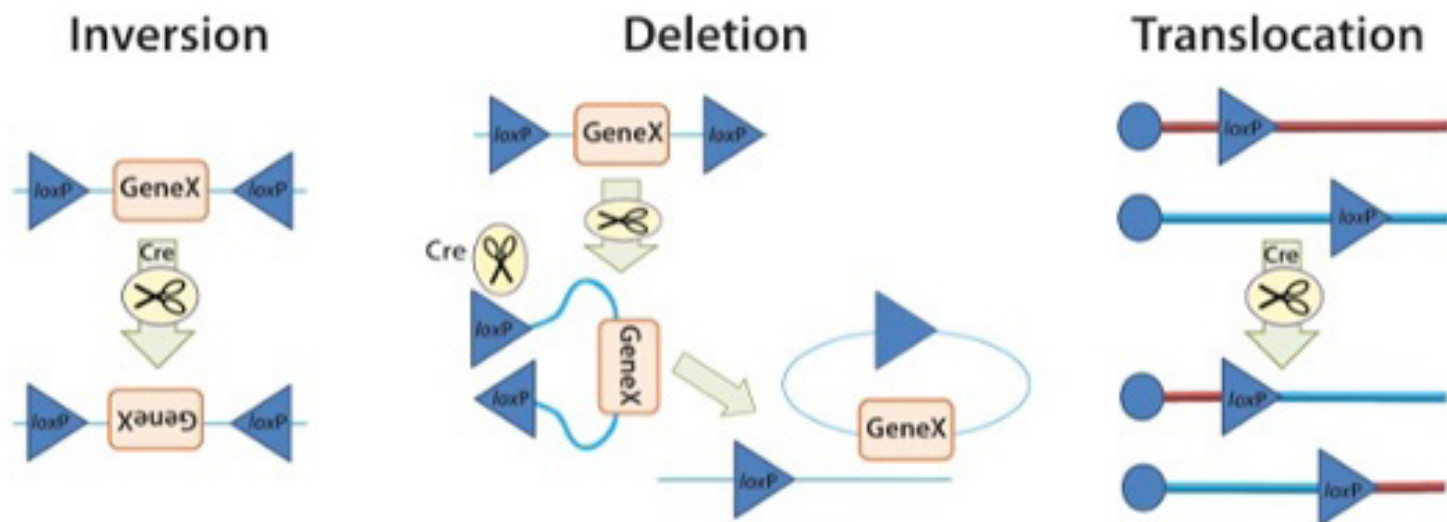


Image by Larissa Haliw

**Inversion:** If the loxP sites are on the same DNA strand and are in opposite orientations, recombination results in an inversion and the region of DNA between the loxP sites is reversed.

**Deletion:** If the sites face in the same direction, the sequence between the loxP sites is excised as a circular piece of DNA (and is not maintained).

**Translocation:** If the sites are on separate DNA molecules, a translocation event is generated at the loxP sites.

### How Can I Use Cre-Lox?

The Cre/lox system is a well-established research tool, especially in the field of mouse transgenics. A few of the most common uses are listed below.

- **Cre-dependent gene expression** – placing a stop codon with loxP sites on either side (often called a “lox-stop-lox” or “LSL” cassette) upstream of a gene of interest will prevent gene expression in the absence of Cre. In the presence of Cre, the stop codon is excised, and gene expression proceeds. One popular lox-stop-lox plasmid is from Tyler Jack’s lab: [Lox-Stop-Lox TOPO](#).
- **Cre-dependent gene knockout** – conversely, putting the loxP sites on either side of a gene (called “floxing”, for “**f**lanked by **loxP**”), will permit gene expression until Cre is present, at which time the gene will be disrupted or deleted.

## CRE-LOX (CONT'D)

---

- **Selection marker removal** – in conventional mouse targeting, targeted clones are selected for using a selection marker; however, it is often desirable to remove the marker after the initial selection process. By floxing the selection marker, Cre can be used to easily perform this eviction.
- **Regulated Cre expression** – by placing Cre downstream of promoters that are active only in certain cell or tissue types, during certain stages of development, or by making the Cre inducible (such as with tamoxifen or doxycycline), the Cre recombinase can be expressed only in specified cells or at specified times. Combining this with some of the loxP methods described above, a genetic modification can be restricted based on experimental constraints. This has been used for a wide range of purposes, including activating an oncogene only in a particular organ, or bypassing embryonic lethality.

To facilitate the use of Cre-lox technology, transgenic mice have been constructed that express Cre under a variety of ubiquitous and regulated promoters, and many loxP-containing alleles have also been constructed. Additionally, Cre-containing adenovirus (Ad-Cre) or AAV ([AAV-pgk-Cre](#)) has been used to successfully introduce Cre into cells of interest.

### What If I Need Two Separate Recombination Events?

One potential limitation of using loxP sites is the inability to tightly control which loxP sites recombine if more than two are present; intramolecular events happen with greater frequency than intermolecular events, but any two sites can potentially recombine. To account for this, alternate mutant versions of the loxP site have been created, which contain a unique asymmetric spacer “NNNTANNN”, where “N” indicates which bases may vary from the canonical sequence. Among these are loxN (GtATACcT), lox2272 (GgATACtT), and lox511 (GtATACAT). These variant lox sites undergo recombination with other sites of the same type, but are not cross-compatible. Using different lox site variants allows Cre to catalyze more than one specific recombination event in a single system.

The *Saccharomyces cerevisiae* FLP-FRT recombination system is another site-directed recombination technology very conceptually similar to Cre-lox, with flippase (Flp) and the short flippase recognition target (FRT) site being analogous to Cre and loxP, respectively. The FLP-FRT technology can be an effective alternative to Cre-lox, and has also been used in conjunction with it, allowing for two separate recombination events to be controlled in parallel.

## CRE-LOX (CONT'D)

---

### Further Reading

1. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. Sternberg, N. and Hamilton, D. 1981. PubMed [PMID: 6276558](#).
2. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Sauer, B. and Henderson, N. 1988. PubMed [PMID: 2839833](#).
3. Tissue- and site-specific recombination in transgenic mice. Orban, P.C., Chui, D., and Marth, J.D. 1992. PubMed [PMID: 1495975](#). PubMed Central [PMCID: PMC49604](#).
4. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Gu, H., Zou, Y.R., and Rajewsky, K. 1993. PubMed [PMID: 8513499](#).
5. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Gu, H., Zou, Y.R., and Rajewsky, K. 1993. PubMed [PMID: 8513499](#).
6. Addgene's [Cre-Lox Guide](#)



# KNOCKOUT/KNOCK-IN PLASMIDS

By Benoit Giquel | December 1, 2016

One of the most powerful strategies to investigate a gene's function is to inactivate, or "knockout", the gene by replacing it or disrupting it with a piece of DNA designed in the lab. Specially constructed plasmids can be used to replace genes in yeast, mice, or *Drosophila* through homologous recombination. The concept is simple: deliver a template with a modified version of the targeted sequence to the cell which will recombine the template with the endogenous gene. Here, we'll describe the techniques and the plasmids used to inactivate specific genes in mammalian cells. Despite the popularity of CRISPR-based [knockout/knock-in](#) systems, these systems remain valuable, especially in cases where CRISPR cannot be used (e.g. there are no suitable PAM sequences nearby or your gene of interest is difficult to target specifically with a gRNA). Be sure to keep these techniques in mind when choosing a knockout strategy!

## Knockout Plasmids

Homologous recombination is a mechanism to accurately repair harmful double stranded breaks, in which nucleotide sequences are exchanged between two similar or identical molecules of DNA. Gene targeting takes advantage of this natural process to replace a targeted genetic locus with homologous sequence using a specially designed vector that contains sequence homology to the locus of interest. To give you an idea of the process, we'll walk through an experiment designed to knock out exon 2 of a given gene.

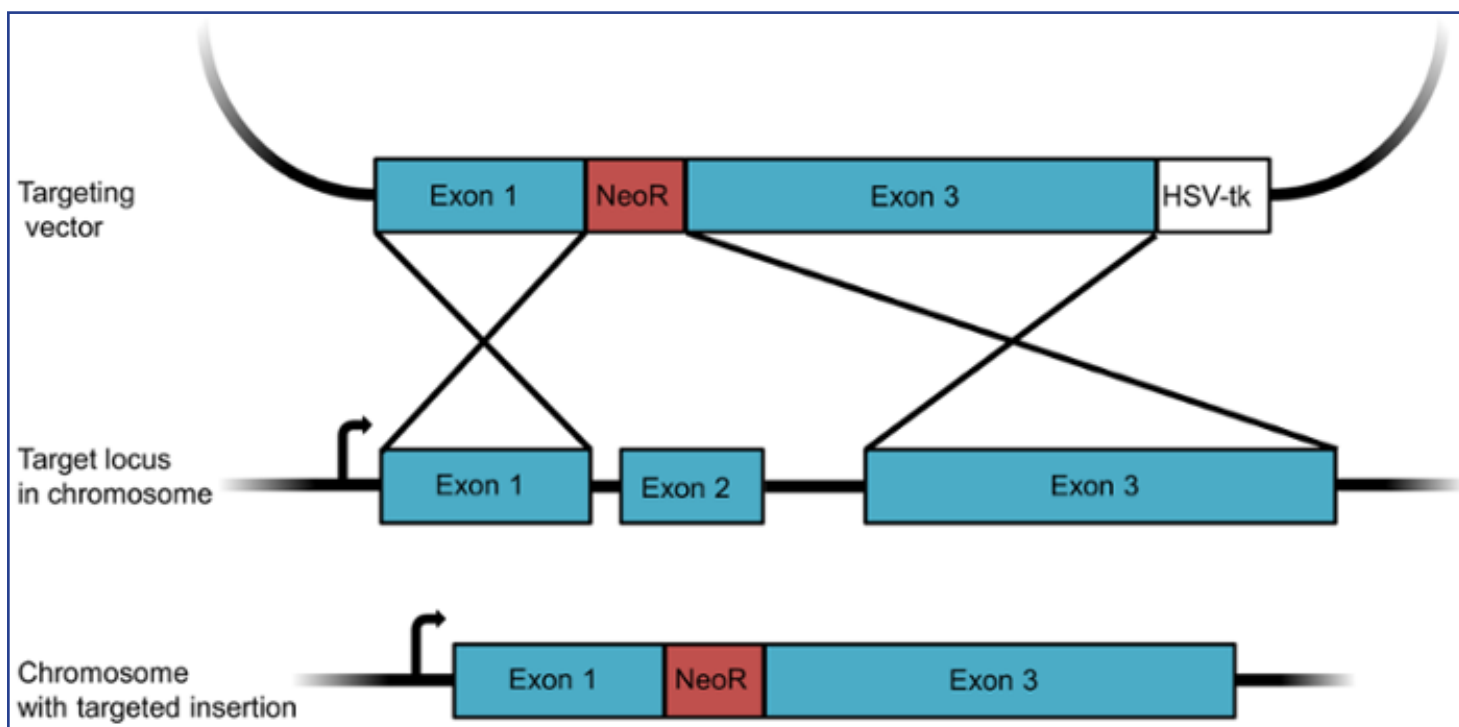


Figure 1: A knockout targeting vector designed to insert a resistance gene. The vector contains a neomycin resistance gene (*NeoR*) flanked by homology arms. The negative selection marker *HSV-tk* is used to select against random recombinants.

1. Design your targeting construct. For recombination to occur in a cell, a minimum of 2 kb of sequence homology is required, but 6 to 14 kb of homology is typical for targeting constructs. In the example shown in Figure 1, large sequences corresponding to exons 1 and 3 of the targeted gene have been cloned into the vector on either side of an [antibiotic resistance gene](#). To avoid selecting for cells in which the construct has randomly integrated into the genome, a negative selection marker like the HSV thymidine kinase (*HSV-tk*) is included just

## KNOCKOUT/KNOCK-IN PLASMIDS (CONT'D)

outside one of the homology arms. When we select cells later, we will first perform a positive selection for the antibiotic resistance gene and then a counter selection for the negative selection marker - this latter step will kill off many of the cells that have randomly integrated all or large portions of the plasmid.

2. Deliver your construct to your target cells. After recombination, exon 2 of the targeted gene will be removed from the chromosome and replaced by the resistance gene. The gene is thus disrupted, or knocked out.

3. Use positive and negative selection to find correctly recombined cells. Recombination is a rare event, so you must select as opposed to screen for cells where recombination has occurred. Neomycin, puromycin, and hygromycin resistance genes are commonly used for positive selection. While the positive selection marker selects for recombination, the negative selection marker selects against improper, random recombination into a different locus. Correctly recombined cells will not contain the negative selection marker, but cells with random recombination may incorporate the negative selection marker into the genome. This final product of non-homologous recombination can survive positive selection, but it will be sensitive to negative selection.

4. Remove the positive selection marker. Since an antibiotic resistance gene can affect a cell's phenotype, researchers usually remove it after selection using the [Cre/Lox recombination system](#). After "floxing" the resistance gene through the insertion of flanking LoxP sequences, you can remove the gene through the addition of Cre recombinase (Figure 2).

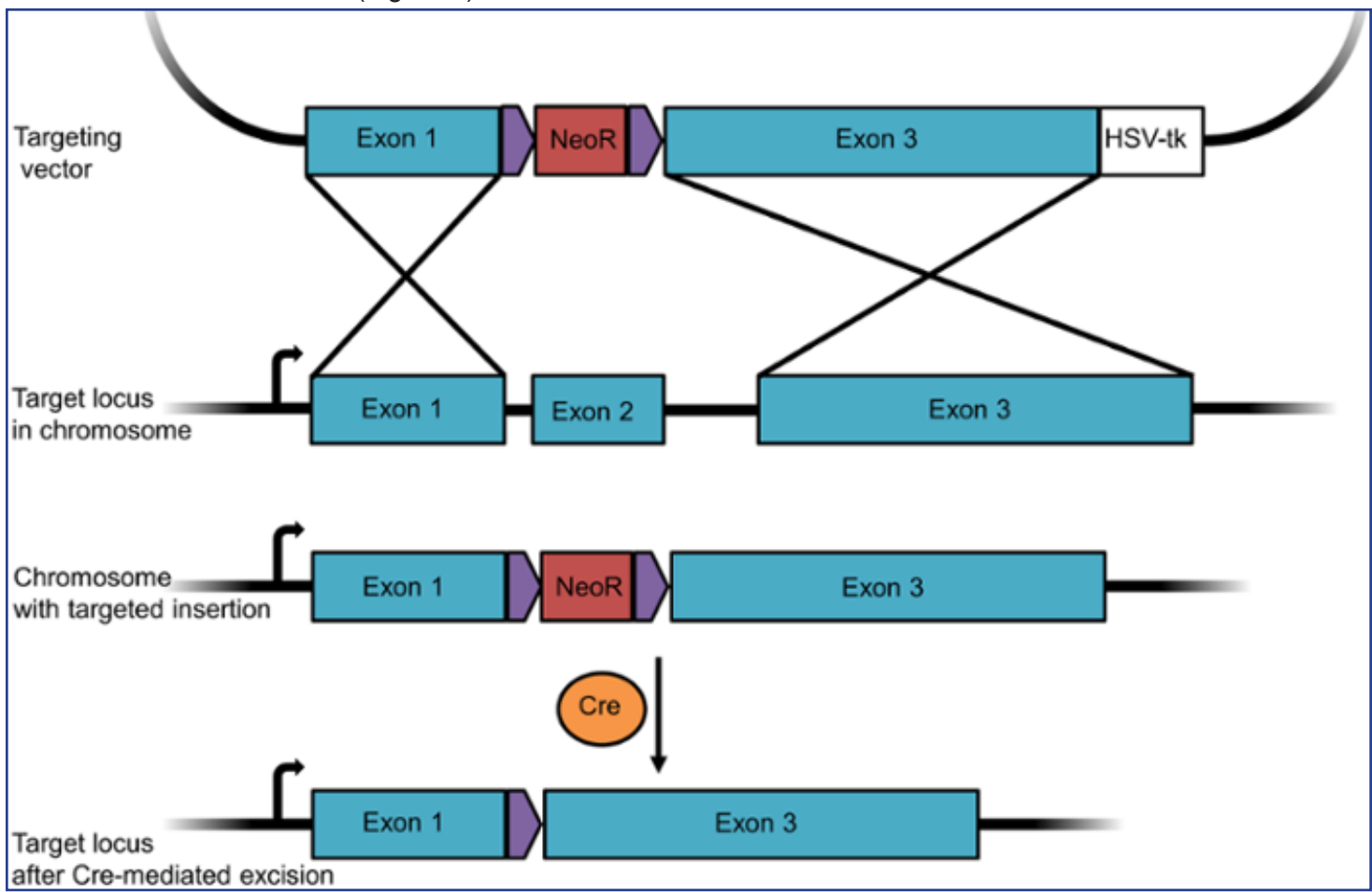


Figure 2: A knockout targeting vector designed to insert and later remove a resistance gene. In this construct, the neomycin resistance gene is floxed so that it can be removed post-selection through Cre-mediated recombination.

## KNOCKOUT/KNOCK-IN PLASMIDS (CONT'D)

### Knockout/Knock-In Plasmids

Gene targeting methods also make it possible to insert, or knock-in, any gene, tag, or mutated exon into the genome. For this purpose, the sequence to be inserted is cloned into the vector between the homologous sequences together with the positive selection marker. In order to both knockout a given gene and insert GFP into the genome, we'd create a plasmid similar to the one shown below, where the sequence of GFP is cloned together with the Neomycin resistance (NeoR) gene between exon 1 and 3 of the targeted gene. Upon recombination, the GFP/NeoR cassette is inserted in place of exon 2. Thus, the targeted gene is disrupted (knocked out) but the inserted GFP is expressed (knocked in). As seen in the above example, you can remove a floxed resistance gene using [Cre recombinase](#). If GFP is under control of an endogenous promoter, you can use GFP expression to track cells participating in development or other physiopathological events to which the chosen promoter responds. You can also use this method to tag an endogenous protein with GFP, as seen in blue flame plasmid [OCT4-eGFP-PGK-Puro](#) from the [Jaenisch lab](#).

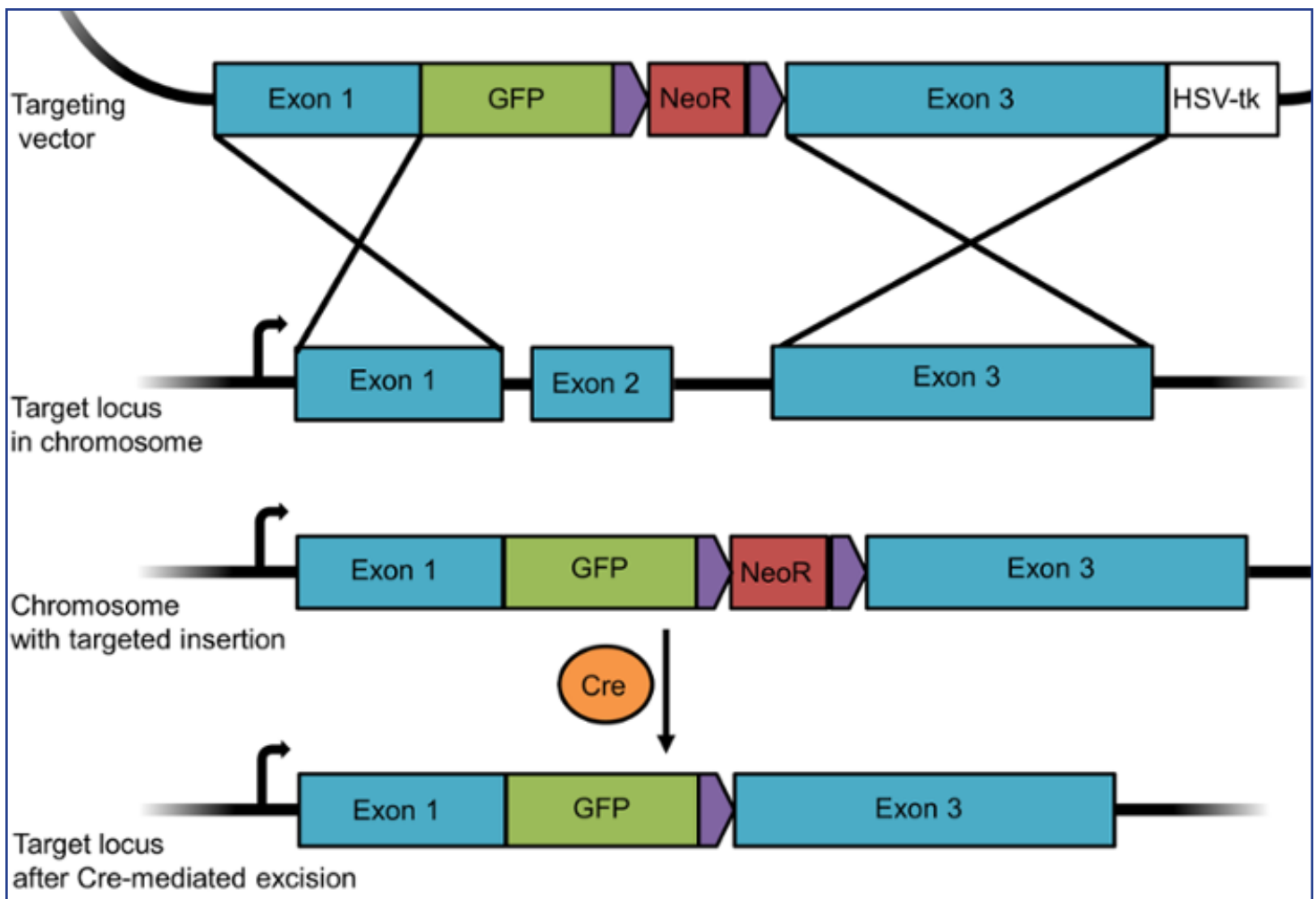


Figure 3: A knockout/knock-in targeting vector designed to insert GFP into a given locus.

# KNOCKOUT/KNOCK-IN PLASMIDS (CONT'D)

## Conditional Knockouts

The methods and plasmids described in the preceding sections are simple ways of knocking out a non-essential gene of interest. However, if your gene of interest is essential, a true knockout can be lethal, and you'd instead want to create a conditional knockout.

To make a conditional knockout, researchers often use the Cre/Lox system described earlier. In this case, you design your targeting vector such that a set of three LoxP sites flank the resistance gene and the targeted exon in the gene of interest (Figure 4). When recombination occurs, the gene still functions normally because one of its exons has simply been replaced with the same sequence flanked by LoxP sites while the resistance cassette has been placed into an intron.

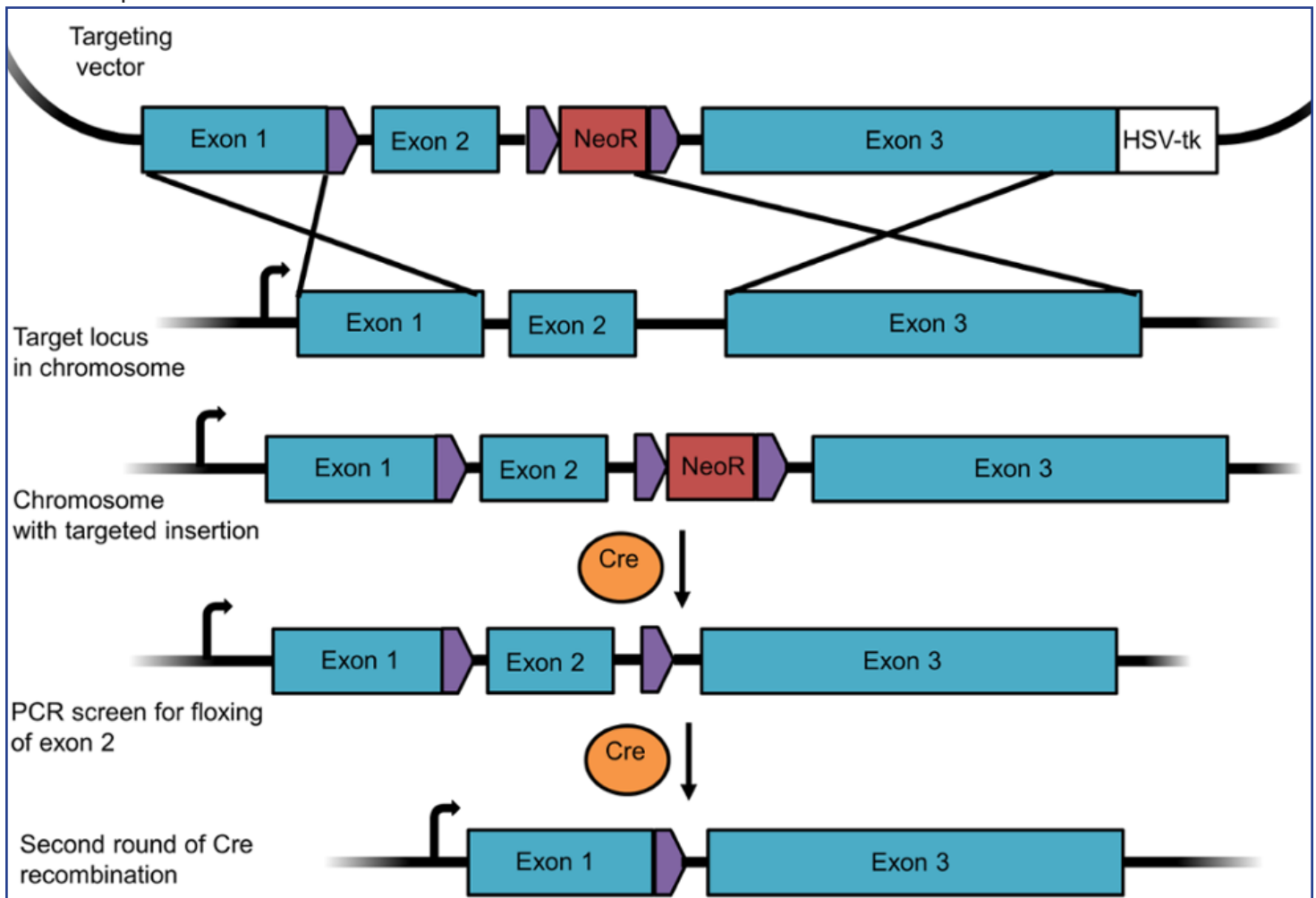


Figure 4: A targeting vector to create a floxed allele for a conditional knockout. This experimental design requires two rounds of recombination - the first removes the selection marker, the second the floxed exon.

After recombination has occurred, you'll first remove the resistance marker using Cre recombinase. Since there are 3 loxP sites, recombination can occur in multiple ways. The desired recombination event will remove NeoR only and leave exon 2 floxed, as seen in line 4 of Figure 4. Since the loxP sites are located in intronic regions, this gene will still be expressed. You'd first screen for this specific recombination outcome using PCR and then generate a monoclonal cell line with the floxed exon. You can then conditionally remove this exon (and hence

## KNOCKOUT/KNOCK-IN PLASMIDS (CONT'D)

---

knock out the gene) through a second round of Cre recombination.

### The Future of Knockout Plasmids

Although these methods have been used to create many knockout cell lines and animal models, their efficiency is very low, ranging from undetectable to 0.1%. In contrast, new genome editing techniques such as CRISPR are easier to use and are more efficient at inactivating genes. CRISPR can target a genomic sequence and create a break that can be repaired by homologous recombination using a repair template. These templates can include loxP sites to create [conditional floxed alleles](#). Although CRISPR is very good at making knockouts, knocking in large sections of DNA can be more difficult. Addgene depositors have developed new CRISPR-based methods for various knock-ins, discussed in our [CRISPR 101 eBook](#).

It's hard to believe that the first knockout mouse was created in 1989, less than thirty years ago. As traditional knockout and new CRISPR tools continue to be perfected, the production of cells and mouse knockout lines should increase. New genome engineering tools also offer hope for the creation of new knockout animal models in species that were previously difficult to engineer, like rats. If you have tools for creating knockout animals or cell lines, please consider sharing them by depositing with Addgene!

### Further Reading

1. Hall B, A Limaye, & AB Kulkarni. (2009) "Overview: Generation of Gene Knockout Mice". *Curr Protoc Cell Biol.* 19: 1217. PubMed [PMID: 19731224](#). PubMed Central [PMCID: PMC2782548](#).

# OVERVIEW OF TALEN TECHNOLOGY

By Eric J. Perkins, Addgene | 2011

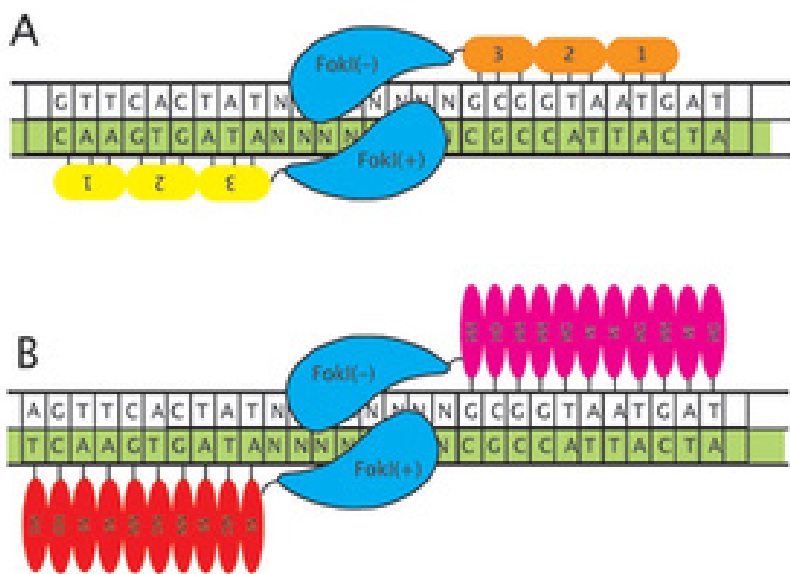


Figure 1: (A) Representation of a zinc finger (ZF) nuclease. Zinc fingers recognize triplets and the FokI nuclease operates as a dimer, cutting in the spacer region between two distinct ZF target sites. (B) A TAL effector nuclease (TALEN). Similar in principle to the ZF nuclease, the components of the array recognize individual nucleotides.

Zinc fingers were once the go-to technology for targeting enzymes and other useful protein domains to a specific DNA sequence but TAL effector and [CRISPR](#) based technologies have recently gained much greater prominence.

In 2009, groups led by Jens Boch at the Martin-Luther-University Halle-Wittenberg and [Adam Bogdanove](#) at Iowa State University published the nucleotide recognition code of the TAL effectors, which were isolated from the plant bacterial pathogen *Xanthomonas*. The central TAL targeting domain is composed of 33-35 amino acid repeats. These repeats only differ from each other by two amino acids, their repeat-variable di-residue (RVD). Ultimately, it is this RVD that determines which single nucleotide the TAL effector (TALE) will recognize: (**Figure 1B**) HD targets cytosine, NI targets adenine, NG targets thymine, and NN targets guanine (though NN can also bind adenine with lower specificity).

With the template for studying this type of technology already laid by the zinc finger community ([see interview with Adam Bogdanove](#)), progress on TALE research has been swift. The incentives are high. Since ZF targets are confined to sequences composed of triplets with corresponding zinc fingers, potential targetable sites in your average genome are every 500 bp. TAL effectors have some restrictions (for example, the target must start with a T), but they still have potential targetable sites approximately every 35bp.

Perhaps one of the most appealing features of the TAL effector arrays is the ease with which they can be made. Intuitively, one would expect that assembling relatively small, repeat-laden DNA regions into a single construct would be technically challenging. Thanks to efforts by the Bogdanove group and Daniel Voytas's group from the University of Minnesota, arrays can be assembled in a matter of days. In an open access [Nucleic Acids Research](#) article that came out in the spring of 2011, the groups describe a set of customized plasmids that can be used with the recently developed Golden Gate cloning method to assemble "multiple DNA fragments in an ordered fashion in a single reaction." Using these plasmids, which are available as a [kit from Addgene](#), custom arrays consisting of 12-31 repeats can be assembled and inserted into a variety of backbones in just a few steps (**Figure 2**). The Bogdanove group also hosts [web-based software](#) to help scientists find potential TAL effector targets.

## OVERVIEW OF TALEN TECHNOLOGY (CONT'D)

The [Voytas/Bogdanove TALEN kit](#) came online in the summer of 2011 and is one of the best-selling kits in Addgene's history. Dr. Keith Joung's lab at Massachusetts General Hospital also released [a TALEN kit](#) through Addgene. This kit, first described in *Nature Biotechnology* in August 2011, uses a serial ligation protocol to assemble arrays. Dr. Joung, co-founder of the [Zinc Finger Consortium](#) with Dr. Voytas, has considerable experience in the DNA targeting field. A third kit, the [TALe Toolbox](#), from Dr. Feng Zhang's group at the Broad Institute, is also available at Addgene. Dr. Zhang's lab published a paper on TAL effectors in *Nature Biotechnology* early in 2011 and they described their kit for assembling TAL effector arrays in a 2012 issue of *Nature Protocols*. The Zhang lab also hosts a helpful [TAL effectors website](#) that provides several TAL effector resources.

Whether you work in a *C. elegans* lab and have been struggling to mutate a specific gene or you're anxiously investigating gene therapy options for a rare disease in humans, TAL effectors could become a key tool in your plasmid toolbox. Addgene has rapidly become an important resource for this new technology, and we hope to continue to have a strong scientific partnership with both the labs that hone these tools and the researchers who are learning to wield them. Learn more about Addgene's TAL effector and TALEN kits [here](#).

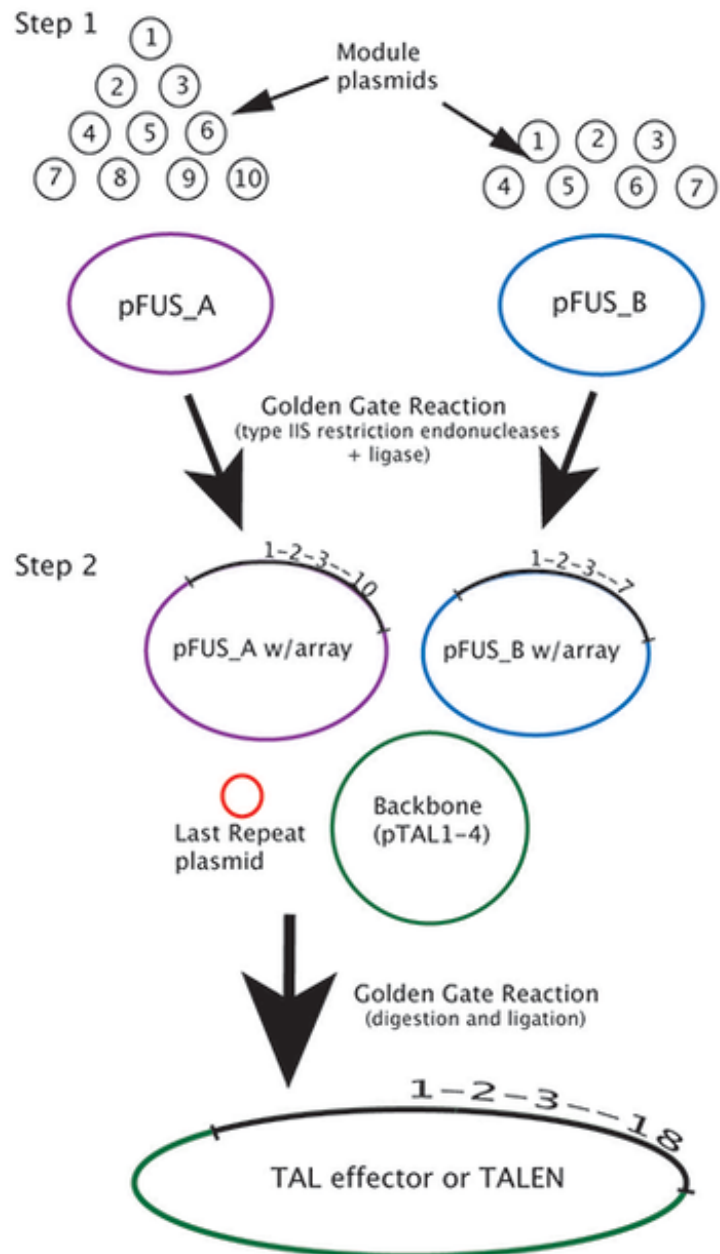


Figure 2: Simplified representation of the Voytas/Bogdanove Golden Gate TALEN kit. In this example, the final array is composed of 18 TAL effector repeats. The kit can be used to make arrays from 12-31 repeats.

## OVERVIEW OF TALEN TECHNOLOGY (CONT'D)

---

### Further Reading

1. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. *Nucleic Acids Res.* 2011 Jul;39(12):e82. Epub 2011 Apr 14. PubMed [PMID: 21493687](#).
2. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, Yeh JR. *Nat Biotechnol.* 2011 Aug 5;29(8):697-8. PubMed [PMID: 21822241](#).
3. TAL effectors: customizable proteins for DNA targeting. Bogdanove AJ, Voytas DF. *Science* 2011 Sep 30;333(6051):1843-6. PubMed [PMID: 21960622](#).
4. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M., & Arlotta, P. *Nat Biotechnol.* 2011 Feb;29(2):149-53. PubMed [PMID: 21248753](#).
5. TALEs for the masses. Rusk N. *Nat Methods* 2011 Mar;8(3):197. PubMed [PMID: 21473016](#).
6. Move over ZFNs. *Nat Biotechnol.* 2011 Aug 5;29(8):681-4. PubMed [PMID: 21822235](#).
7. See the Q&A with Dr. Adam Bogdanove and his thoughts on the TALE technology [here](#).



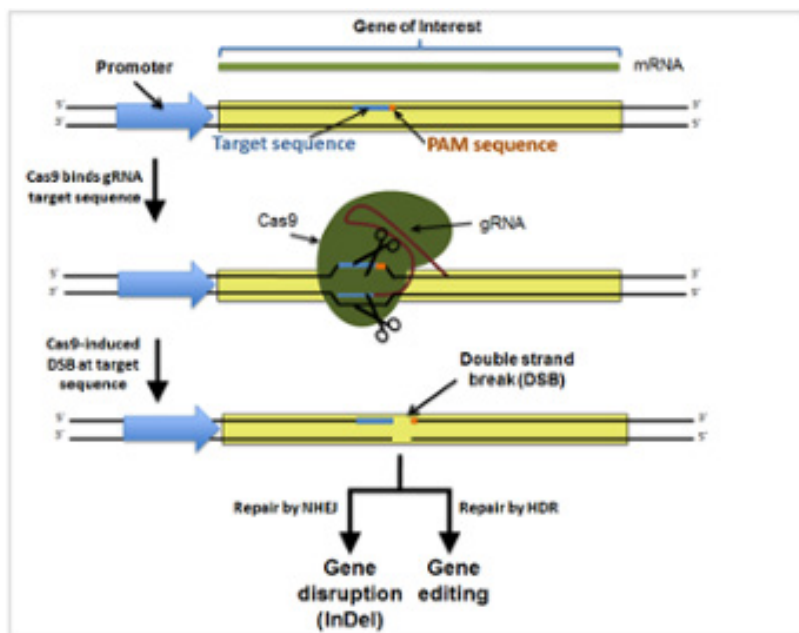
# OVERVIEW OF CRISPR TECHNOLOGY

By Marcy Patrick, Addgene | September 3, 2015

Although the **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats or CRISPR sequences were initially discovered in *E. coli* in 1987, the concept that these sequences operated as a safeguard against bacteriophages did not come to light until twenty years later. Initial experiments exposed *S. thermophilus* to predatory phages to test if the exogenous DNA would be incorporated into the bacterial genome. The results of these studies led scientists to hypothesize that prokaryotes developed an adaptive immune system that utilized various CRISPR-associated, or Cas genes, not only to store a record of invading phages, but to also destroy the phage upon re-exposure. More specifically, the Cas proteins snip the foreign DNA into small fragments of ~30 bp in length and insert it into the CRISPR sequence. RNA from these CRISPR loci is transcribed and used to guide a Cas nuclease to the specified exogenous DNA complementary to the target. Once the CRISPR/Cas complex binds to this foreign DNA, a cut is made to destroy the invader.

[Check out the CRISPR 101 eBook for More Information on CRISPR and Its Applications](#)

CRISPR/Cas-inspired genome editing was initially described by Cong *et al.* and Mali *et al.* in January 2013 and is currently the most commonly used technology for genome engineering. There are two distinct components



to this system: (1) a guide RNA and (2) an endonuclease, in this case Cas9 (see image below). The guide RNA combines a unique 20-nucleotide target sequence and a Cas9-recruiting tracrRNA sequence into a single chimeric guide RNA (gRNA) transcript.

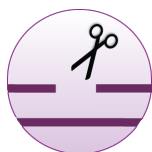
When expressed in the cell, the gRNA/Cas9 complex is recruited to the target sequence by base pairing between the gRNA target sequence and its complement in the genomic DNA. For Cas9 to successfully bind to DNA, the target sequence in the genomic DNA not only must be complementary to the gRNA sequence, but must also be immediately followed by the correct protospacer adjacent motif (PAM) sequence. Note, the PAM sequence is present in the DNA target sequence but not in the gRNA sequence, and any DNA sequence with the correct target sequence followed by the PAM sequence will be bound by Cas9. The functional consequences of Cas9 binding to the DNA are dependent on which Cas9 is expressed.

## OVERVIEW OF CRISPR TECHNOLOGY (CONT'D)

### The Following Cas9 Uses Have Been Developed



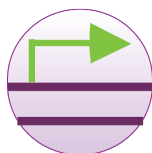
**Cut:** Wild type Cas9 efficiently generates double strand breaks at sequences homologous to a co-expressed gRNA. These double strand breaks are normally repaired by non-homologous end joining (NHEJ) which can result in gene disruptions. If a repair template containing homology to the cut region is provided, the cut can be repaired by homology directed repair (HDR) resulting in specific, targeted mutations or gene insertions (gene editing) (see image above).



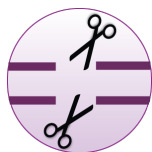
**Nick:** A mutated “nickase” version of the Cas9 enzyme generates a single-strand DNA break (Nick), instead of a double-strand DNA break (Cut). When targeted to two unique, proximal sites in the genome, nickases can generate double stranded breaks with higher specificity than wt Cas9 due to the requirement for two isolated, proximal cutting events.



**Interfere:** A catalytically inactive Cas9 (dCas9) can knockdown gene expression by interfering with transcription. The dCas9 can sometimes be fused to an additional repressor peptide potentially enhancing gene knockdown.



**Activate:** A catalytically inactive Cas9 (dCas9) fused to an activator peptide can activate or increase gene expression when targeted to promoter or other regulatory regions.



**dCas9-FokI:** A catalytically inactive Cas9 (dCas9) fused to FokI nuclease to generate double strand breaks (Cut) at sequences homologous to two co-expressed gRNAs.



**Purify:** Isolate specific genomic regions of interest using a catalytically inactive Cas9 (dCas9) fused to an epitope tag(s).



**Visualize:** Visualize specific genomic regions of interest using a catalytically inactive Cas9 (dCas9) fused to a fluorescent protein.

The above-listed Cas9 variants are all based on the *S. pyogenes* version of the nuclease; however, Cas9 nucleases from different species such as *N. meningitidis*, *S. aureus*, and others are currently being investigated for their genome editing capabilities. In late 2015, researchers from [Feng Zhang's lab](#) additionally discovered a new Cas9 homolog, [Cpf1](#), which, in addition to requiring only 1 RNA molecule for gene targeting, generates staggered cuts at target sites with a TTN PAM sequence. These staggered cuts may make Cpf1 better suited for HDR mediated gene editing than Cas9 and its variants.

# OVERVIEW OF CRISPR TECHNOLOGY (CONT'D)

---

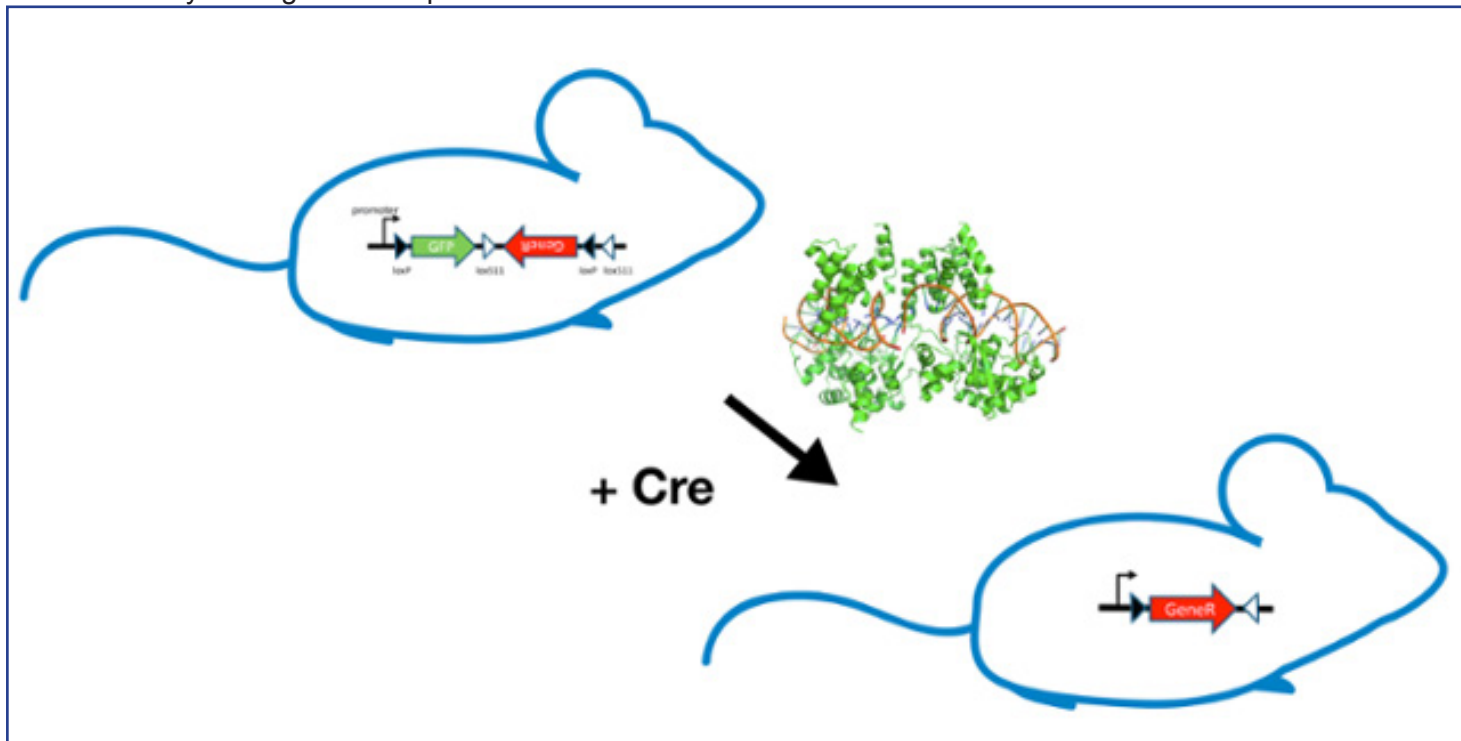
## Further Reading

1. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Jinek, Martin et al. Science 337.6096 (2012): 816-821. PubMed [PMID: 22745249](#).
2. Multiplex Genome Engineering Using CRISPR/Cas Systems. Cong, Le et al. Science 339.6121 (2013): 819–823. PubMed [PMID: 23287718](#).
3. RNA-Guided Human Genome Engineering via Cas9. Mali, Prashant et al. Science 339.6121 (2013): 823–826. PubMed [PMID: 23287722](#).
4. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Zetsche, Bernd, et al. Cell (2015). PubMed [PMID: 26422227](#).

# FLEX VECTORS

By Michelle Cronin | April 28, 2016

In a previous post from our Plasmids 101 series, we learned how the [Cre-loxP recombination system](#) can be used to induce site-specific recombination events, and that the orientation of the flanking loxP sites directs the Cre recombinase to invert, translocate, or excise a DNA fragment. The availability of both wild-type and mutant loxP sites has allowed scientists to leverage this system in new, creative ways. This section will focus on one such strategy--[the FLEx switch](#)--which utilizes recombination elements to turn off expression of one gene, while simultaneously turning on the expression of another!



## What Are FLEx Switches

FLEx (or “flip-excision”) switches were designed as a genetic tool for researchers to conditionally manipulate gene expression *in vivo* using site-specific recombination. The FLEx switch takes advantage of the orientation specificity of the site-specific recombinases (SSRs) Cre and FLP. SSRs bind DNA at target sites to induce site specific recombination events: Cre recombinase binds loxP sites, while FLP binds FRT sites. When a DNA sequence is flanked by target sites (floxed) in opposing orientations, an SSR will invert the DNA sequence between the sites. If a DNA sequence is floxed in the same orientation, the SSR will excise the DNA fragment. By manipulating the number, orientation, and type of target sites that flox your genes of interest, a powerful FLEx switch can be created for your specific experimental needs.

## How Does the FLEx Switch Work?

Let’s say you want to design a genetic FLEx switch that turns BFP expression off, while turning on mCherry expression. For this FLEx switch to successfully work, the cassette would need to contain the BFP coding sequence in the sense orientation, followed by the mCherry coding sequence in the antisense orientation (Figure 1, top). The entire DNA cassette would be flanked by two pairs of target sites- one wild-type pair (loxP, black arrowheads) and one mutated pair (for example- lox511, white arrowheads). It is necessary to use two different pairs of target sites for this strategy to work effectively. Both loxP and lox511 are recognized by Cre but lox511

## FLEX VECTORS (CONT'D)

sites can only recombine with other lox511 sites, not with loxP sites. Alternatively, you could use a pair of loxP sites and a pair of FRT sites and include both Cre and FLP recombinases.

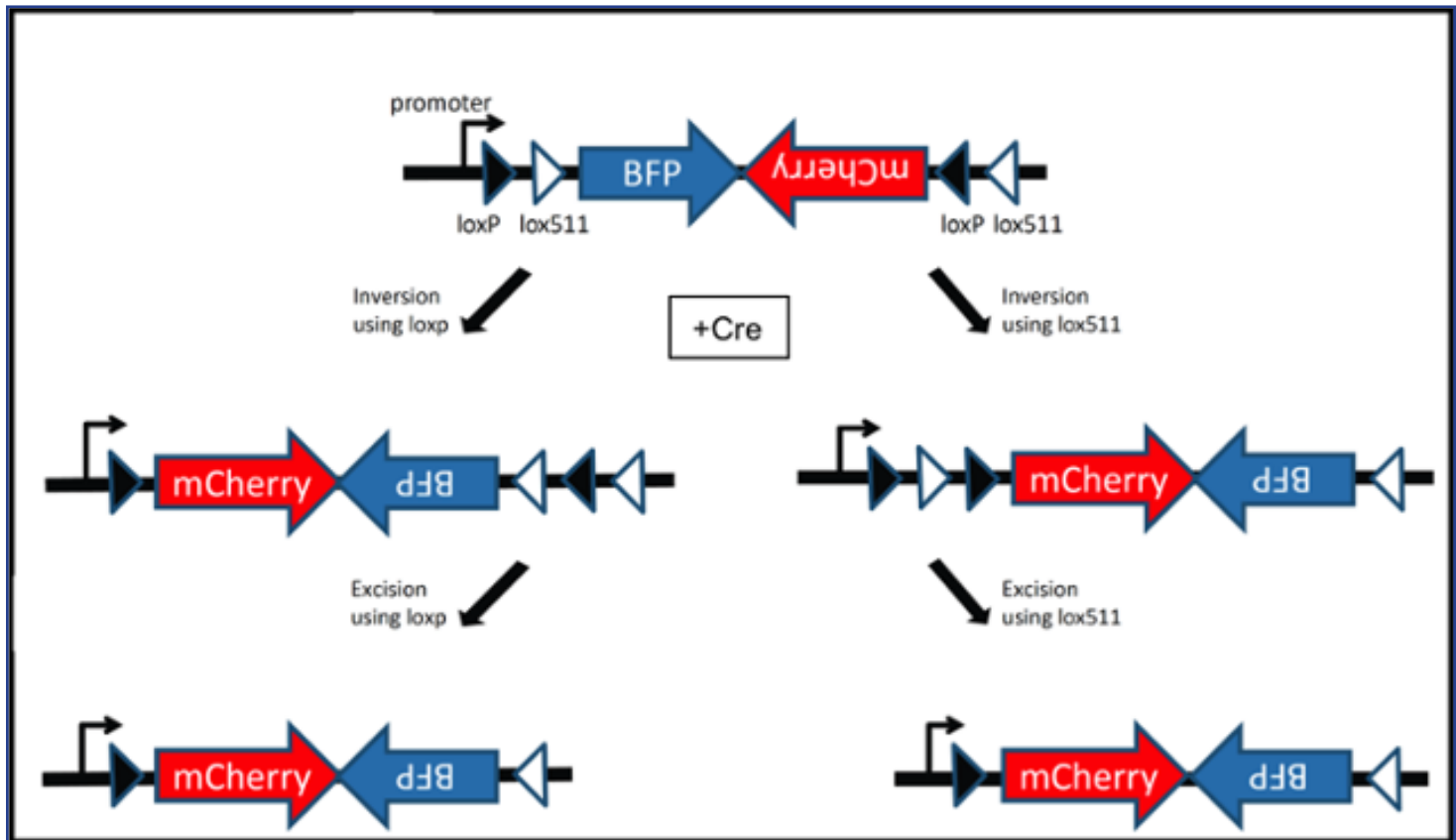


Figure 1. This basic FLEEx switch will express BFP in the absence of Cre recombinase or mCherry in the presence of Cre recombinase (adapted from Schnutgen et al. Nat Biotech 2003).

Once the SSR of choice is introduced, recombination can proceed either by first utilizing the loxP sites or the lox511 sites. Regardless, the first recombination step will invert the intervening DNA fragment using either loxP or lox511 sites, leaving two identical sites on one end of the DNA fragment (Figure 1, middle). A second recombination event then excises the DNA between the identical loxP or lox511 sites (Figure 1, bottom). Since only one loxP and lox511 site will remain on either side of the DNA fragment, any additional recombination events are impossible even in the presence of Cre recombinase. This plasmid now specifically drives expression of mCherry instead of BFP.

### Beyond Switching Fluorophores, Why Is FLEEx Useful?

The power of the FLEEx system lies in the ability to conditionally turn off one gene and activate another. In the field of mouse genetics, FLEEx is currently being used to study a wide variety of *in vivo* conditional gene rescue experiments, as well as for improvement of optogenetic [adeno-associated virus \(AAV\) vectors](#). Listed below are a few additional uses for FLEEx switches:

- **Conditional Rescue for a Gene Knockout:** To test if the phenotype you are observing is really due to a single gene deleted in your knockout mouse, you can design a FLEEx switch conditional knock-in cassette that uses Cre to activate gene expression as shown in Figure 2. The cassette should contain a reporter DNA sequence (such as GFP) in the sense orientation, followed by coding sequence of the gene you would like to res-

## FLEX VECTORS (CONT'D)

cue with in the antisense orientation (let's call the gene "GeneR"). The key to this experiment is the placement of the two pairs of target sites. By flanking the reporter with one loxP site and one lox511 site in the same orientation, and placing the additional loxP and lox511 sites after the inverted GeneR sequence in the opposing orientation, you have created a cassette that will express GeneR only in the presence of Cre. In this example, Cre would first invert the reporter and GeneR fragment. Then the second recombination event would excise the reporter from the cassette, allowing for expression of GeneR. This allows a researcher to test whether or not expressing GeneR alone rescues the phenotype being studied.

- **Introduction of a Conditional Point Mutation or Functional Mutation *in vivo*:** Similar to a conditional rescue experiment with a wildtype copy of a gene, you can also design FLEx switches to conditionally test the function of various point mutations and truncations in the absence of the endogenous gene.

- **Spatio-temporal Expression of Optogenetic reporters:** Since most optogenetic reporters are delivered using AAV injection, the majority of the cells surrounding the injection site will become infected and would have the potential to express the transduced reporter gene. Generating a FLEx switch to control expression of an optogenetic reporter would ensure that the reporter remains silent until a cell or tissue-specific Cre is provided. Since the FLEx switch ensures Cre-dependent reporter expression, the potential for background expression of the reporter will be greatly reduced, allowing for clearer characterization of the location, morphology, and circuit mapping of a targeted neuronal population.

Sometimes also referred to as DiO (Double-floxed inverse Orientation), FLEx systems are generally considered "Cre-On" meaning that your gene of interest starts in the inverse/antisense "off" position and is flipped to the sense (on) orientation in the presence of Cre. A similar strategy can be employed for "Cre-off" (i.e. your gene of interest starts in the sense orientation and Cre is used to flip it to "off"). These Cre-off switches may be referred to as DO (Double-floxed Orientation) and could be used to study the physiology or behavior of certain cell populations.

### Further Reading

1. Schnütgen F, Doerflinger N, Calléja C, Wendling O, Chambon P, and Ghyselinck NB. A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. 2003 PubMed [PMID: 12665802](#).
2. Branda, CS and Dymecki, SM. Talking about a Revolution: The Impact of Site-Specific Recombinases on Genetic Analyses in Mice. 2004. PubMed [PMID: 14723844](#).
3. Atasoy D, Aponte Y, Su HH, and Sternson SM. 2008. A FLEx switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. PubMed [PMID: 18614669](#). PubMed Central [PMCID: PMC2593125](#).

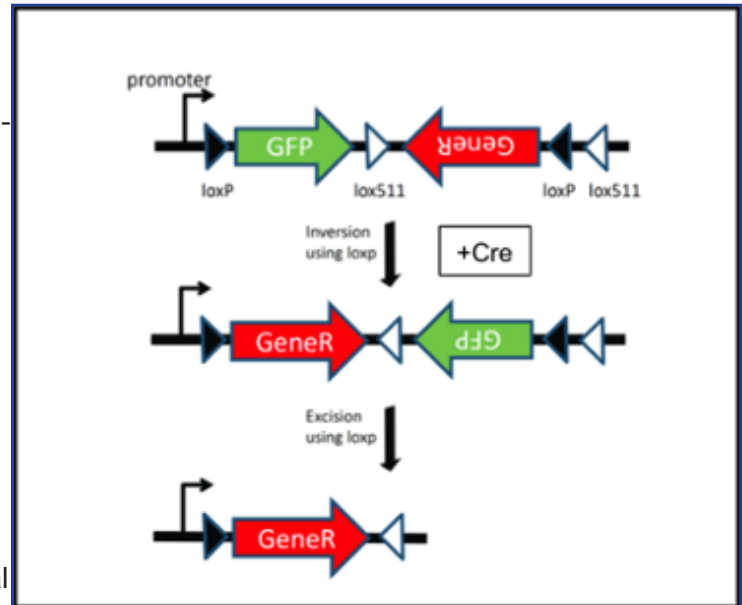


Figure 2. In the presence of Cre, GFP expression is lost, allowing for rescued expression of GeneR.

## FLEX VECTORS (CONT'D)

---

4. Saunders A, Johnson CA, Sabatini BL. 2012. Novel recombinant adeno-associated viruses for Cre activated and inactivated transgene expression in neurons. PubMed [PMID: 2286602](#). PubMed Central [PMCID: PMC3406316](#).

### Find FLEx Plasmids from Addgene Depositors

- [Ed Boyden](#)
- [Bernardo Sabatini](#)
- [Wenbiao Chen](#)
- [Liquin Luo](#)
- [Massimo Scanziani](#)
- [Scott Sternson](#)
- [Roger Tsien](#)
- [Naoshige Uchida](#)
- [Ian Wickersham](#)
- [Hongkui Zeng](#)

## SLEEPING BEAUTY AWAKENS FOR GENOME ENGINEERING

By Emma Markham, Addgene | June 30, 2015

Transposons are sequences of DNA that can move around in a genome. In a laboratory setting, transposons can be used to both introduce genes into an organism's genome (see figure 1) and to disrupt endogenous genes at the site of insertion. In both of these cases, transposons combine the advantages of [viruses](#) and naked DNA while eliminating some of the drawbacks. Specifically, viruses are able to infect and replicate in host cells, but they are susceptible to cells' defense mechanisms. The use of non-viral vectors, like transposons, avoids many, though not all, of these defenses. For some applications of genome engineering - such as certain forms of gene therapy - avoiding the use of viruses is also important for social and regulatory reasons.

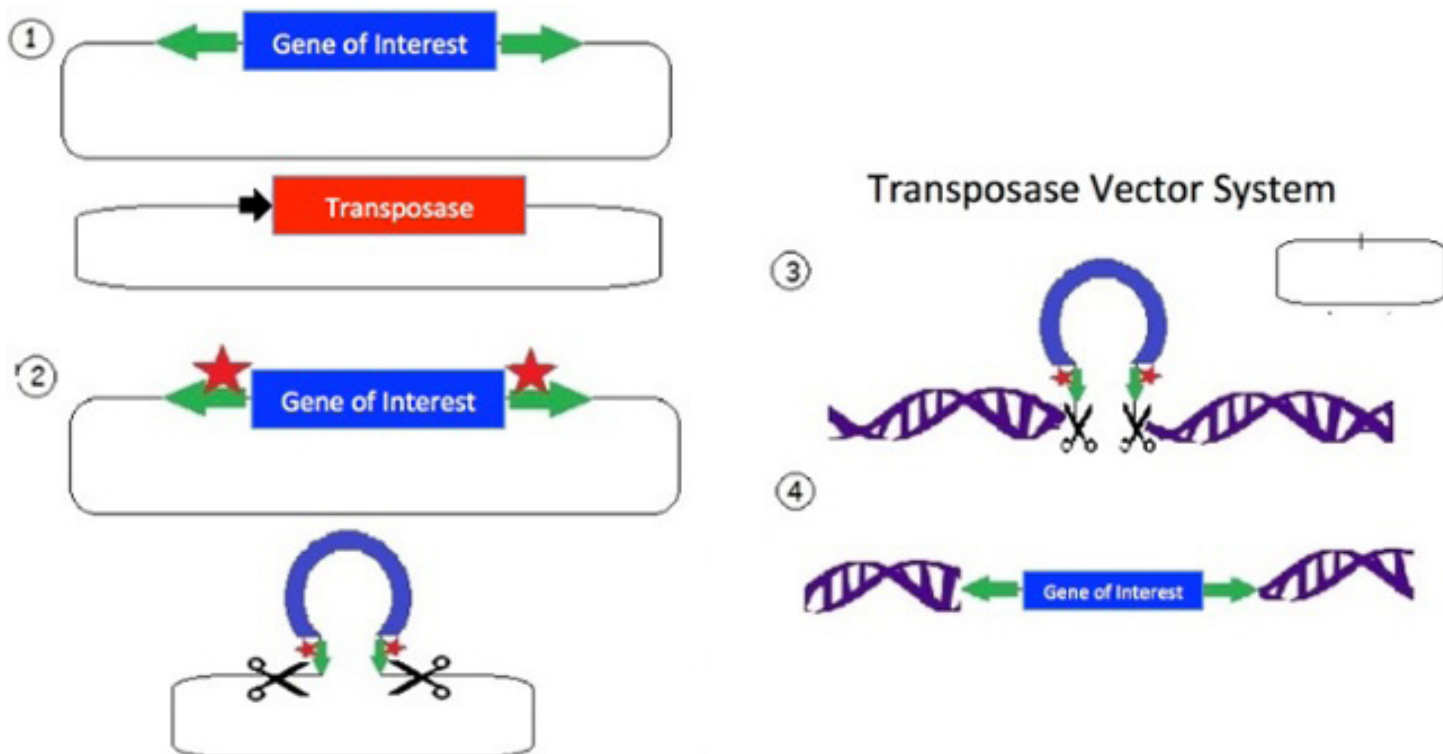


Figure 1 – (1) The transposon consists of a mirrored set of inverted repeats (green) flanking the gene of interest (blue) in a plasmid backbone. A separate plasmid contains the transposase gene (red) for expressing the transposase enzyme. (2) The transposase enzyme is expressed (red stars) and binds the inverted repeats (green); and an endonuclease reaction occurs which cuts the DNA. (3) The released transposon can now bind a strand of DNA with a TA dinucleotide (there are many such sites in the human genome). The original plasmid is empty following the removal of the transposon; the plasmid is then degraded by the cell. (4) The transposase creates a double strand break in the DNA and allows the transposon to integrate. Additionally, the TA sequence is duplicated near the gene of interest insertion site.

### The Sleeping Beauty Transposon System

The Sleeping Beauty (SB) transposon system is one such system which has been designed to perform gene transfer in vertebrates. The system is comprised of 2 components: 1) an SB transposase, the enzyme required for catalysis of transposition; and 2) a transposon containing the gene-expression cassette that can translocate within the genome. The Sleeping Beauty transposase gains its name from the fact that it was recreated from a defective transposase found in fish. This transposase has been found in the genomes of many different fish species in a variety of non-functional forms and is believed to be millions of years old. The first functional SB system was synthesized by fusing and modifying various extinct transposases. Improvements have been made

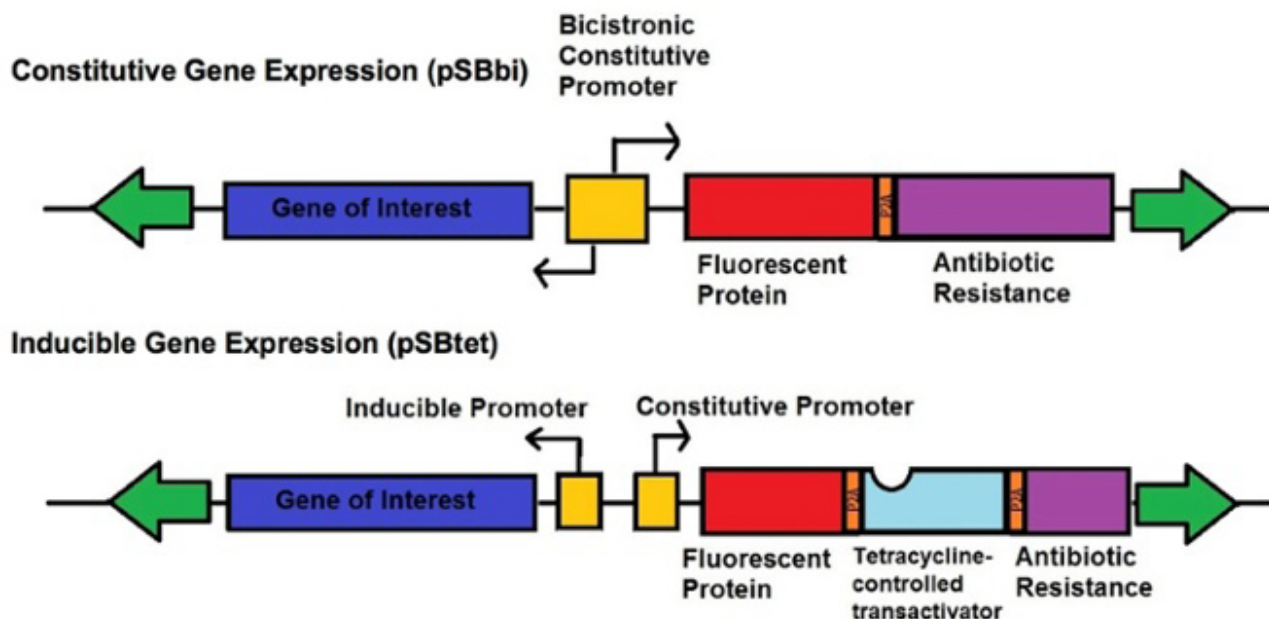


## SLEEPING BEAUTY AWAKENS FOR GENOME ENGINEERING (CONT'D)

since then by increasing the consensus between a greater number of extinct transposon sequences and then testing various combinations of changes.

The basic functional version of the Sleeping Beauty transposon (SB10), developed in 1997, continues to be altered in order to improve its gene insertion efficiency in primary cells. In 2009, [Zsuzsanna Izsvak's lab](#) screened many SB transposase variants for hyperactivity in mammalian cells. Their efforts resulted in a new and improved SB transposase, [SB100X](#), that is approximately 100-fold more efficient than the first-generation transposase (as determined by transposition assays of antibiotic-resistance genes conducted in HeLa cells). In their [Nature Genetics](#) paper, the lab showed that efficiencies of stable gene transfer using SB100x were comparable to stable transduction efficiencies of integrating viral vectors.

### New Sleeping Beauty Transposase Vectors



Even more recently, [Eric Kowarz's lab](#) developed two new types of vectors for the SB transposase system - one set of vectors for robust constitutive expression (pSBbi derivatives) and one set of vectors for inducible expression (pSBtet derivatives). The constitutive and inducible expression vectors were developed into 16 variants with different selection markers (blasticidin, neomycin, puromycin, hygromycin) and fluorescent protein reporters (GFP, BFP, RFP, no FP). The vectors each contain two [SfiI](#) restriction sites with different overhangs depending upon whether they are on the 3' or 5' end of the gene of interest allowing for insertion in the correct orientation. These plasmids use a novel synthetic RPBSA promoter, which is made up of a fragment of the RPL13a promoter fused to a region of the RPL41 gene, to drive the selection/reporter parts of each construct. This [new tool kit](#) allows the rapid generation of stable transgenic cell lines which can robustly express your gene-of-interest, either constitutively or regulated by doxycycline.

## SLEEPING BEAUTY AWAKENS FOR GENOME ENGINEERING (CONT'D)

---

The Sleeping Beauty transposase is proving an exciting tool for producing specific genetic mutations and gene disruptions without the use of viral vectors. In 2009, it was named Molecule of the Year. Additionally, in 2014 researchers at MD Anderson Cancer Center presented results from a clinical trial using the Sleeping Beauty transposase as a nonviral vector for CD19-directed chimeric antigen receptor (CAR) therapy. SB is a valuable tool and an alternative to CRISPRs for genome engineering whose use is only likely to grow in the coming decades.

### Further Reading

1. Optimized Sleeping Beauty Transposons rapidly generate stable transgenic lines. Kowarz E, et al. *Biotechnol J.* 2015 Feb 4. doi: 10.1002/biot.201400821. PubMed [PMID: 25650551](#).
2. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Mates L, et al. *Nat Genet.* 2009 Jun;41(6):753-61. Epub 2009 May 3. PubMed [PMID: 19412179](#).
3. A new approach to gene therapy using Sleeping Beauty to genetically modify clinical-grade T cells to target CD19. Singh H, et al. *Immunol Rev.* 2014 Jan;257(1):181-90. doi: 10.1111/imr.12137. PubMed [PMID: 24329797](#).

# CHAPTER 8: YOU'VE MADE A PLASMID....NOW WHAT?



# HOW TO NAME YOUR PLASMID IN 3 EASY STEPS

By Matthew Ferenc, Addgene | Feb 13, 2014



There are no universal rules for naming plasmids but here are some good guidelines to follow in order to ensure that people can quickly and easily identify what your plasmid contains and other important information.



## Pro Tip!

A lowercase “p” is often used as the first letter of a plasmid name and simply denotes that the object is a ‘plasmid’. “p” is for plasmid.

*pXXXXX-XXXX*

## Step 1: Backbone Name

Include the empty backbone name in your plasmid name. This simple piece of information can often convey many important details. Once you know the backbone a plasmid is based on, you can usually derive: a) the bacterial antibiotic resistance, b) the promoter that drives the insert, and c) whether the plasmid has any selection markers for use in other cell types (e.g. eukaryotic cells).

*pBACKBONE-XXXXX*



## Pro Tip!

For a catalog of published and commercially available empty backbones, visit the Addgene [Vector Database](#).

## Step 2: Insert Name

Include information about the insert in your plasmid name. This is often a 3-6 letter representation of the gene (or DNA sequence).

*pBACKBONE-hGene*



## Pro Tip!

Often researchers will add a lower case letter to the beginning of their insert abbreviation to specify what species it is. Example: ‘h’ is for Human (*homo sapiens*), ‘m’ is for mouse (*mus musculus*), ‘r’ is for rat (*rattus rattus* or *rattus norvegicus*), etc.

## Step 3: Add Your Tags

Add any tags or fusions that are on your insert. Typically you would list any tags or fusion proteins in the order they appear in the plasmid and their relative position to the insert. For example, if you have a Flag tag on the N-terminus of your insert, you would list it first.

*pBACKBONE-Flag-hGene*

If there was also an EGFP fused to the C-terminus of your insert you would list it after the insert.

## HOW TO NAME YOUR PLASMID IN THREE EASY STEPS (CONT'D)

---

*pBACKBONE-Flag-hGene-EGFP*

### Other Considerations:

If your insert contains a mutation or modification, this should be included in the plasmid name. Mutations are generally listed as the amino acid change and not a nucleotide change. The proper way to denote an amino acid mutation is to list the one letter abbreviation of the wild type amino acid immediately followed by its position (number) relative to the start Methionine (Met) followed by the one letter abbreviation of the mutated amino acid currently at that position.

In contrast, unmutated or Wild Type (wt) versions of the insert are often denoted by “wt” either directly before or directly after the insert name.

**Example:** If the Glutamine at position 295 was mutated to an Alanine, Q295A.

**Mutant version:** *pBACKBONE-Flag-hGene(Q295A)-EGFP*

**Wildtype version:** *pBACKBONE-Flag-hGene-EGFP* or *pBACKBONE-Flag-hGene(wt)-EGFP*

**These simple rules allow any scientist** to know what is in a plasmid and often how it can be used just by reading the name.

# HOW TO VERIFY YOUR PLASMID

By Lianna Swanson, Addgene | Aug 28, 2014



Image is released to public domain.  
Source: Wikipedia, accessed 8/27/2014.  
Author Giac 83.

**Congratulations, you have a plasmid expressing your gene of interest (YGOI) and are ready to dive into your functional experiments!** Whether you've cloned the plasmid yourself or obtained it from a colleague down the hall, it is always a good idea to take some time to confirm that you are working with the correct construct, and verify that the plasmid you received matches the expected sequence. Here at Addgene, we process all of the plasmids we distribute for quality control purposes in order to confirm the integrity of the DNA.

Below we discuss Addgene's two recommended methods for plasmid DNA verification: sequencing and diagnostic restriction digest.

## Sequencing

Sequencing determines the precise order of nucleotides within the DNA molecule, in this case a plasmid. To get started, you will first need to design and synthesize primers that perfectly complement your plasmid sequence. We recommend starting with a backbone-specific primer that will sequence over the Multiple Cloning Site (MCS) and into YGOI. This way you can avoid designing multiple primers to verify unique genes inserted into the same backbone. Addgene has curated a comprehensive [vector database](#) that will help you find reference sequence for many commonly used backbones, as well as the specific primers used to confirm their integrity. You can also find a list of our most commonly used sequencing primers [here](#). It usually takes a couple of days to receive results after submitting your sample to a sequencing core (depending on the core facility and services available at your institution); however, it will save you time in the long run knowing that you are working with the correct plasmid.

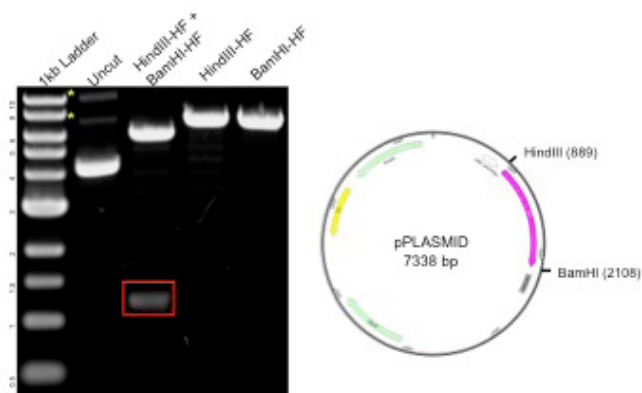
## Diagnostic Restriction Digest

[Diagnostic digests](#) can be used to confirm the relative structure of the plasmid based on the predicted sizes and organization of different features within the plasmid. One benefit of restriction analysis is that it can be used successfully without actually having full plasmid sequence available to you. This method is relatively quick and can be done right in your lab in less than a day (as long as you have purified DNA). Diagnostic restriction digests are comprised of 2 separate steps: 1) incubating your DNA with the selected endonucleases which cleave the DNA molecules at specific sites; and 2) running the reaction on an agarose gel to determine the relative sizes of the resulting DNA fragments.

The most common way of utilizing a restriction digest is to confirm the presence of an insert in a particular vector by excising it from the backbone. This is accomplished by using a combination of specific endonucleases that flank the insert. You will need to know both the approximate size of the [vector backbone](#) as well as the predicted size of the insert. You can search [NCBI](#) for YGOI to find the particular reference sequence if necessary.

## HOW TO VERIFY YOUR PLASMID (CONT'D)

The example plasmid on the right has a total size of 7.3kb, with the insert comprising 1.2kb of the total. The plasmid was digested with 2 unique enzymes (HindIII and BamHI) and run on an agarose gel. The resulting gel image includes a 1kb ladder (lane 1) that has bands ranging from about 500 bp to 10 kb, with the 3.0 kb fragment having increased intensity to serve as a reference band. The uncut DNA (lane 2) shows 3 possible plasmid conformations, with relaxed and nicked marked with asterisks (\*). The digested samples in the last 3 lanes include HindIII and BamHI alone resulting in a single band of the full size of the plasmid, ~7.3 kb, and one double digest with HindIII and BamHI together, matching the backbone size of about 6kb and the released insert at about 1.2 kb (red box). The results on the gel match the predicted sizes inferred from the plasmid information.



### [Check out Our Diagnostic Restriction Digest Video](#)

## Restriction Digest Tips and Tricks:

The following tips for your digest and for your gel will make it easier for you to obtain a useful and informative diagnostic restriction digest.

### *For your digest:*

- **Try choosing unique enzymes.** Enzymes that only cut once allow you to more easily and accurately visualize the full size of your construct.
- **Consider [buffer and temperature compatibility](#) when digesting with more than one enzyme.** Consult the manufacturer's manual for the optimal working conditions for each enzyme.
- **Watch out for methylation issues.** Enzymes like XbaI and ClaI are sensitive to methylation and their activity may be blocked. If you have to use these enzymes for your digest, you will need to purify your DNA from a dcm or dam methylation-deficient bacterial strain such as JM110 or INV110.
- **Avoid star activity.** Some endonucleases (for example BamHI) are capable of cleaving sequences which are similar, but not identical, to their defined recognition sequence. Most enzyme manufacturers make High Fidelity versions of the endonucleases and/or supply custom buffers as means to avoid this issue.

### *For your gel:*

- **Add ethidium bromide (EtBr) to your gel before pouring it.** EtBr binds to the DNA and allows you to visualize the DNA under ultraviolet (UV) light, and adding it to your gel will save time!
- **Don't forget to add loading buffer to your digest reactions before loading them.** The glycerol in the buffer will make sure your sample settles in the gel well and the dyes provide a visual reference point so you can easily assess how far the gel has run. *Bonus:* The dyes also [run at predicted sizes](#) so

## HOW TO VERIFY YOUR PLASMID (CONT'D)

---

you can estimate how far down the gel your bands have travelled based on the dye!

- **Always run a ladder.** Ladders allow you to interpret the bands that you get in your sample lanes. Choose your ladder based on the expected band sizes.
- **Always run control uncut DNA to ensure your enzymes are working.** When uncut plasmid DNA is isolated and run on an agarose gel, you are likely to see 3 bands. This is due to the fact that the circular DNA takes on several conformations the most abundant being: supercoiled, relaxed and nicked. If your digest lanes look like your uncut lane then there is something wrong!
- **Quantify your DNA.** Loading too much DNA will make it difficult to obtain crisp bands and analyze the results. *Bonus:* knowing how much DNA you have loaded in each well will allow you to [approximate the DNA mass](#) of comparably intense samples of similar size.
- **Run the gel at 80-150V until you have good separation between your bands.** Stopping the gel when the bromophenol blue dye line is approximately 75-80% of the way down the gel will ensure you keep smaller bands from running off; however, you may need to run the gel for longer to achieve good separation of larger DNA fragments.

### Further Reading

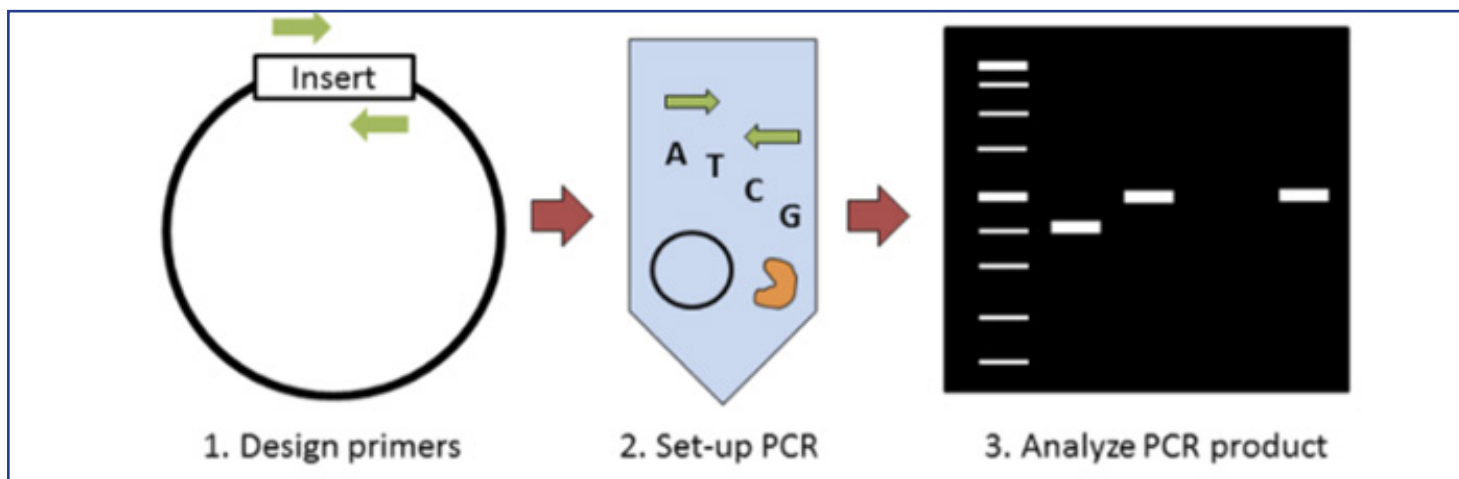
1. Learn [how to set up your digests](#)
2. Learn how to [pour and run a DNA gel](#)



# COLONY PCR

By Beth Kenkel | May 12, 2016

[Molecular cloning](#) requires some method of screening colonies for the presence of an insert. Traditionally this has been done with [restriction enzyme digest](#); however colony PCR can accomplish the same thing in less time and for less money. The key steps to colony PCR are: 1) design primers to detect the presence of your insert; 2) set up a standard PCR reaction (primers, dNTPs, polymerase) using the supernatant of lysed bacteria as template; and 3) [run your PCR product on a gel to analyze product size](#). This section discusses some of the key things to consider when performing colony PCR.



## Designing Colony PCR Primers

The first and perhaps most important step to colony PCR is designing primers. There are 3 strategies for primer design: 1) insert-specific primers, 2) backbone-specific primers, and 3) orientation-specific primers.

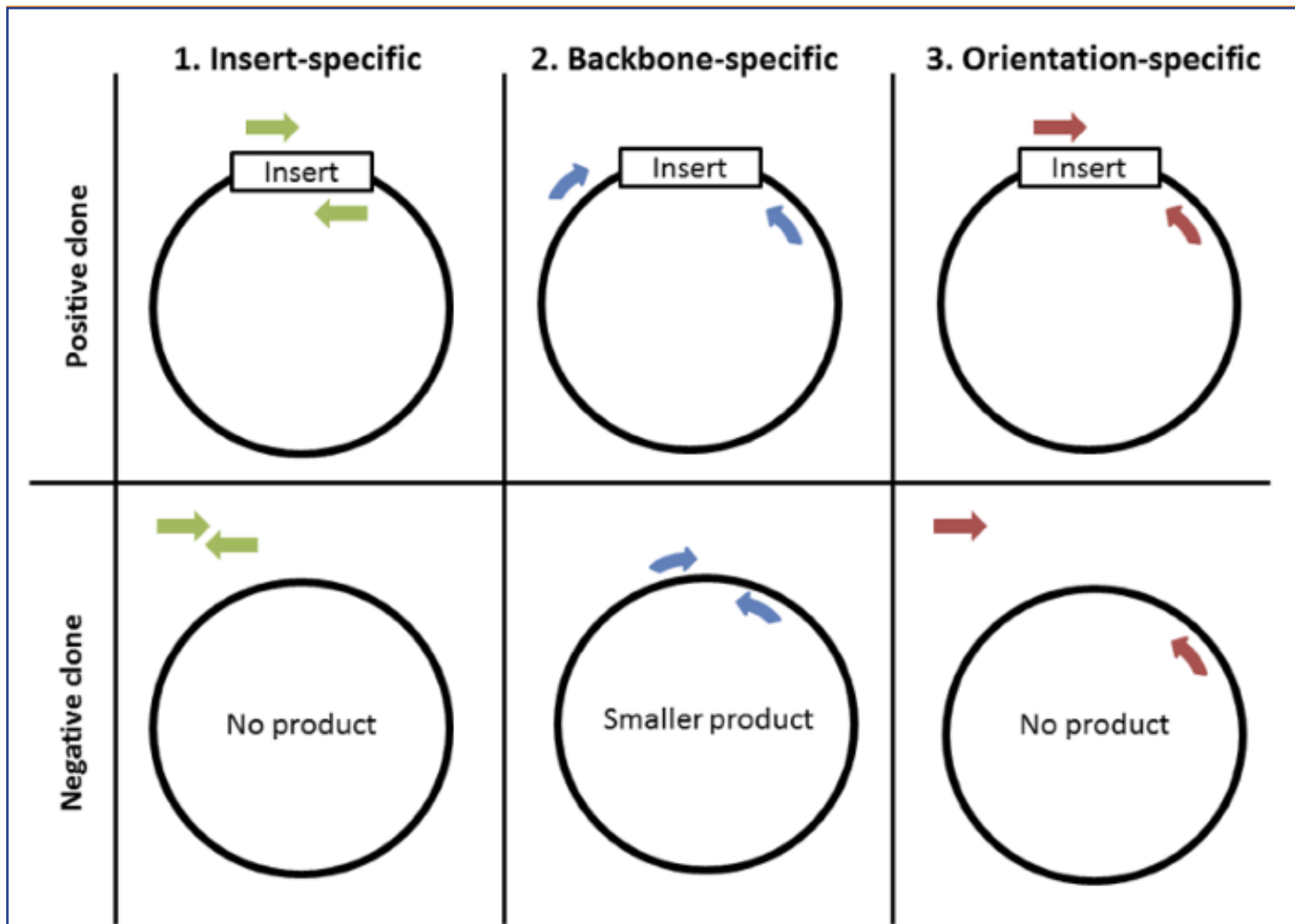
## Designing Colony PCR Primers

**1. Insert-specific primers:** Insert-specific primers are designed to anneal to an insert-specific sequence. This is a “yes or no” kind of test, with a positive clone amplifying a product and a negative clone resulting in no product. Additionally, this type of primer only tells you if your specific insert is present but not if it’s in the correct orientation or even if it’s in your plasmid backbone.

**2. Backbone-specific primers:** A second option is to design backbone-specific primers. These primers are designed to anneal to sites that flank the insert site. A positive clone will produce a larger size product than a negative clone without the insert. This type of primer pair can tell you if the insert is the correct size and whether it’s within your backbone. This type of primer pair is also great for screening clones created with the same backbone but that contain different inserts. When you design primers to anneal outside the cloning site, it doesn’t matter what the sequence of the insert is, allowing you to use the same primer pair to screen for the presence of many different inserts. The downside: this type of primer does not provide information about the orientation of the insert.

**3. Orientation-specific primers:** If you need information about insert orientation, then you might consider designing orientation-specific primers. Blunt end cloning is an example of when you might want to know the orientation of the insert. One member of this type of primer pair anneals to a sequence flanking the insert and one primer anneals to the insert. A simple way to create this type of primer pair is to mix-and-match insert-specific and back-bone specific primers.

# COLONY PCR (CONT'D)



Each approach has advantages and drawbacks, which are summarized in the table below. The type of primers you use depends on your preferences. Either way, make sure to test your colony PCR primers before using them to screen colonies. The best way to do this is by using your vector with and without an insert to verify that the primers amplify the expected size PCR products.

Primer Design	Pros	Cons
Insert Specific	<ul style="list-style-type: none"> <li>Simple “yes or no” test result</li> </ul>	<ul style="list-style-type: none"> <li>Insert orientation not determined</li> <li>Requires new primer pair for each insert</li> </ul>
Backbone Specific	<ul style="list-style-type: none"> <li>Provides information about the size of the insert and if it's in the plasmid</li> <li>Can be used to screen clones that were created with the same backbone</li> </ul>	<ul style="list-style-type: none"> <li>Insert orientation not determined</li> </ul>
Orientation Specific	<ul style="list-style-type: none"> <li>Tells you the orientation of the insert</li> <li>Tells you if the insert is in the plasmid</li> <li>Great for blunt end cloning</li> </ul>	<ul style="list-style-type: none"> <li>Requires new primer pair for each insert</li> </ul>

# COLONY PCR (CONT'D)

## PCR Set-Up

Setting up colony PCR reactions is nearly identical to preparing a standard [PCR reaction](#): combine template, primers, polymerase, and dNTPs and then incubate with a standard PCR thermocycling program. One key difference is the plasmid DNA must be released from the bacteria in order to serve as PCR template. Dealing with this and a few other colony PCR tips are highlighted below.

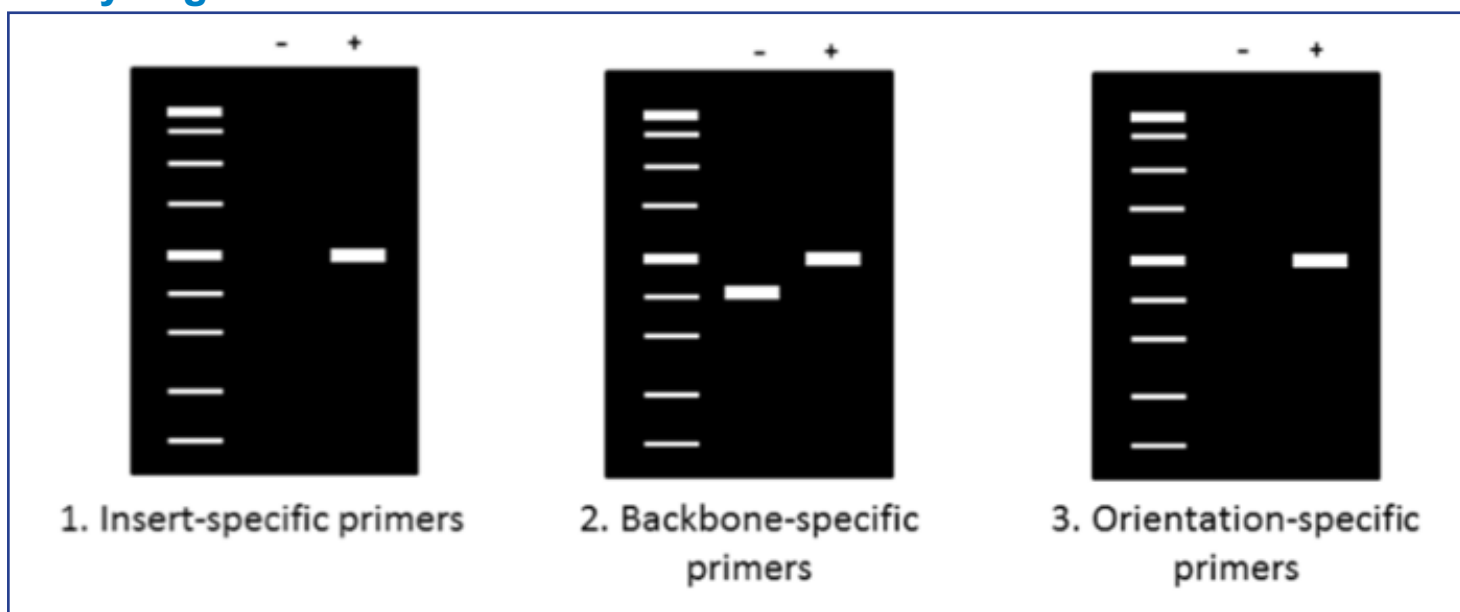
**1. Preparing the Template:** Pick a single colony with a sterile flat toothpick or pipette tip and swirl in a small amount of sterile water. Pick 3-10 colonies in total to test, depending on the number of background colonies on your no ligation control plate. The more background, the more colonies you will need to screen.

**2. Saving clones for later culture:** At this point, you will want to hang onto your clones for later use. There are a few ways you can do this. If you are going to complete your colony PCR analysis in the same day, you can save the leftover bacteria-water suspension and use them to start cultures of your positive clones. If you want to store your clones longer term, just streak the colonies on an LB plate. You can use this plate to start liquid cultures. Lastly, you can start small overnight liquid cultures with the clones you pick and only mini-prepare the positive ones. Regardless of which method you choose, make sure to use the appropriate [antibiotic for selection](#).

**3. Lysing bacteria and setting up PCR reactions:** The remaining bacteria-water suspension will serve as the template for your PCR reaction. You just need to lyse the bacteria to release the plasmid DNA by either briefly boiling the sample before use or by directly adding a small volume of the sample to the PCR reaction. The bacteria will be lysed during the initial heating step of the PCR reaction. A standard Taq polymerase is sufficient.

**4. Controls:** [Controls can make or break an experiment](#). The best controls for a colony PCR are the same ones used to verify if the colony PCR primers work in the first place: the backbone vector with and without an insert. These controls are quick references you can use when you run your PCR products out on a gel to determine if the colonies contain an insert. They also serve as controls for your PCR reaction. Running a no template control PCR reaction for detecting contamination is also a good idea.

## Analyzing PCR Product Size on a Gel



## COLONY PCR (CONT'D)

---

Now that your PCR is complete, it's time to run the products on an agarose gel to determine their size. Make sure to use an appropriate molecular weight standard for reference and to add a loading dye with glycerol to your samples before pipetting them onto the gel. The figure above summarizes generalized expected results for the three primers previously described. When using insert-specific primers (1), positive clones (+) will give a band, while a negative clone (-) will not. Backbone-specific primers (2) give larger sized products for positive clones (+) compared to negative clones. Finally, orientation-specific primers (3) give the same band (+) or no band (-) result as insert-specific primers but also tell you whether the insert has the correct directionality.

### Verifying the Insert Sequence with Sanger Sequencing

After identifying a few positive clones, the last step is to mini-prepare these clones and submit the plasmids for [Sanger sequencing](#). Sequencing allows you to confirm the sequence of the insert, insert orientation, and the sequences of the junctions between the plasmid and insert DNA. Colony PCR will greatly reduce the number of clones you'll need to send for sequencing, but won't tell you if your products have any mutations.

There are lots of different [cloning strategies](#), but regardless of which is your favorite, colony PCR is a useful tool to have in your molecular biology tool kit. Consider giving it a try next time you're screening for positive clones!

### Tips and Tricks from the Bench

- **Don't pick too large of a colony.** Too many bacteria can inhibit your PCR reaction or cause non-specific products to show up on your gel.
- **Beware of false positives.** Just because you get the expected sized PCR product doesn't mean there aren't mutations in your insert. Make sure to submit multiple positive clones for sequencing to verify the insert sequence before proceeding with your experiment.
- **Shorter amplicons tend to be better.** Shorter amplicons make for shorter PCR programs and are more likely to work in a PCR reaction that has bacterial debris.
- **Use a positive control.** A good positive control is bacteria transformed with the same backbone plasmid. If this control doesn't amplify a product, then you know there could be something wrong with the PCR setup and/or the primer design.
- **Use a negative control strain.** A good negative control strain is an untransformed culture of the same strain of bacteria you used for cloning. This type of control is especially important for insert-specific primers. If your negative control amplifies a product of the expected size, you know the genome of your bacteria already contains your target sequence.

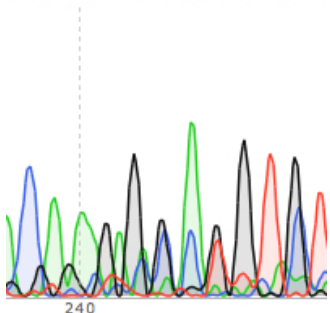
### Further Reading

1. Promega Overview of [Screening for Recombinants](#)
2. New England Biolabs [Overview of Colony PCR](#)

## 6 TIPS FOR ANALYZING SANGER SEQUENCING RESULTS

Originally published as blog on [BitesizeBio](#). By Lianna Swanson, Addgene | Jun 26, 2014

C A A G G G A G G T G T



**Addgene scientists analyze hundreds of sequencing reactions a week.** Here are some good habits that we wanted to pass on to you to make sure you are getting the most out of the data you get back from your sequencing runs.

When you run a restriction digest on a gel you always include proper controls like uncut DNA and the proper ladder. These controls help you properly visualize your results. The most important control for sequencing is to always look closely at the trace file (or chromatogram) of the sequencing results you get back from your favorite sequencing facility.

When it comes to DNA sequencing the chromatogram is your visual control. And, like all controls, leaving it out is a big mistake.

### An Example of Where the Chromatogram Can Come to Your Rescue for Analyzing and Troubleshooting Your DNA Sequencing Results

Above (top left) is an example of a seemingly clean DNA sequence (no Ns in sight). If you never looked at the trace you would be happy.

But look closer, the overlapping peaks in the chromatogram suggest the results are not as certain as the sequence may suggest. In fact this is so ambiguous that the DNA sequencing reaction should be repeated.

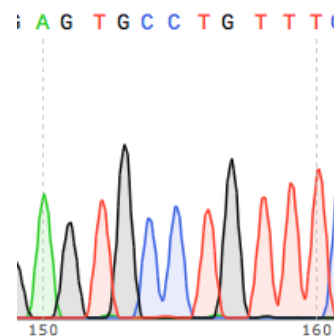
### Here Are a Few Guidelines to Help with DNA Sequencing Troubleshooting and Analysis

#### 1. You can use any of the following programs to view your .ab1 chromatogram file

- [4Peaks](#) (Mac)
- [SnapGene Viewer](#) (Mac/PC)
- [FinchTV](#) (Mac/PC)
- [Sequence Scanner](#) (PC)
- [Chromas](#) (PC)

#### 2. You should see individual, sharp and evenly spaced peaks

...like the peaks in the image to the right....



#### 3. Expect to get 500-700 bases of clean reliable DNA sequence

Anything less and you might suspect contamination in your sample or consider asking your sequencing facility to apply a special protocol for a difficult template. Anything more and you're venturing into the uncertain terrain.

#### 4. Never trust the first 20-30 bases of a DNA sequencing read

The peaks here are usually unresolved and small, so I suggest designing your primer at least 50bp upstream of the sequence of interest.

## 6 TIPS FOR ANALYZING SANGER SEQUENCING RESULTS (CONT'D)

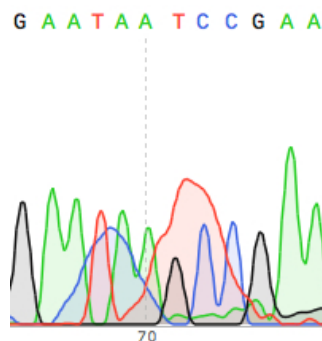
### 5. Use a silica spin column for purification of the samples you send for DNA sequencing

If your sequencing facility requires you to perform your own Big Dye PCR amplification reaction (as opposed to using the all-inclusive service some companies offer), you can purify the product either via the Sodium Acetate/isopropanol precipitation method or using a silica spin column available from several vendors. The precipitation method has an unfortunate side effect of messing up the reaction around base 70-75 of the read (see image below), so I would strongly recommend using a silica spin column. They can be pricey, but well worth it.

### 6. Edit your DNA sequence

Finally, when you do see a miscalled peak, don't be shy. Feel free to edit it. Most chromatogram viewing programs (even the free ones) allow you to edit the sequence.

We hope these tips will help you get the most out of your DNA sequencing results and troubleshoot any problems that come up. Good luck analyzing your sequences!

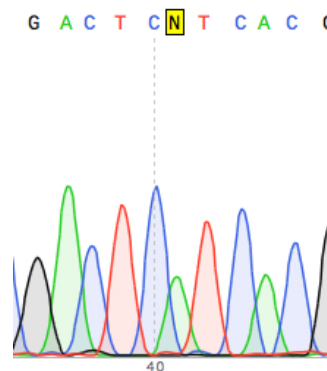


### More DNA Sequencing Resources:

- [Tips for Using BLAST to Verify Plasmids](#)

### Acknowledgements:

- Thank you [BitesizeBio](#) for originally publishing this article and allowing us to share it with our readers!
- The chromatograms in this article were created with [SnapGene](#).



# TIPS FOR USING BLAST TO VERIFY PLASMIDS

By Jason Niehaus, Addgene | June 24, 2014



**At Addgene, we continually use the Basic Local Alignment Search Tool (BLAST) provided by NCBI.** BLAST helps us compare the sequencing results of the plasmids in our repository with known reference sequences, such as full plasmid sequences provided by the laboratories that deposit their plasmids with us or other entries in NCBI's numerous databases.

As our repository has grown over the years (we now have over 50,000 plasmids!), the number of sequencing results we analyze as part of our quality control process has steadily grown. On a busy week, we may need to analyze more than 100 sequences each day as part of our quality control process. Consequently our team has refined our use of the BLAST

web browser interface as efficiently as possible. If you find yourself frequently on the [BLAST website to verify plasmids](#) or validating your new clones, try these tips to make the most of your time and sequence!

## Choosing a BLAST Program

Of the five BLAST programs available, we primarily use Standard Nucleotide BLAST (blastn), Standard Protein BLAST (blastp), and Translated BLAST (blastx), depending on the plasmid region sequenced. NCBI has a terrific [getting-started guide for BLAST](#), which includes a simple explanation of the different BLAST programs, databases, and elements of the BLAST search pages.

At Addgene, we use blastn to compare our sequencing results with empty vector or gene sequences to first determine how much of the result is trustworthy. Then we identify any discrepancies in the nucleotide sequence, such as mismatches, deletions, or insertions. We use blastp or blastx to compare our sequencing results to protein sequences to check open reading frames (ORFs) and determine the potential effect of any nucleotide discrepancies. The blastp and blastx programs are optimized differently and you may want to select one (or both) depending on the information you want to verify. We will delve into these differences below.

## Optimizing *blastn* Searches

On the [Standard Nucleotide BLAST](#) page, the first decision we make is whether to compare our sequencing result to a single known reference sequence or to a BLAST sequence database. If you know the expected nucleotide sequence, check the "Align two or more sequences" checkbox and paste your reference sequence into the Subject Sequence box that appears. Aligning two nucleotide sequences is probably the fastest BLAST search to perform and will save you time compared to other types of BLAST searches.

## TIPS FOR USING BLAST TO VERIFY PLASMIDS (CONT'D)

The screenshot shows the NCBI BLAST Standard Nucleotide BLAST interface. At the top, there are navigation tabs for Home, Recent Results, Saved Strategies, and Help. Below this, the page title is 'Standard Nucleotide BLAST'. The main form area includes a text input for 'Enter Query Sequence', a 'Clear' button, and a 'Query subrange' section with 'From' and 'To' fields. There is also an 'Or, upload file' section with a 'Browse...' button and a 'Job Title' field. At the bottom of the form, the checkbox 'Align two or more sequences' is highlighted with a red box.

If you do not know the exact reference sequence for your result, choose one of the BLAST sequence databases from the dropdown menu. Typically, we use the default nucleotide database “Nucleotide collection (nr/nt)” as it contains a composite of GenBank, EMBL, DDBJ, and PDB sequences and may be the most comprehensive for searching.

The screenshot shows the 'Choose Search Set' section of the BLAST interface. It includes a 'Database' section with radio buttons for 'Human genomic + transcript', 'Mouse genomic + transcript', and 'Others (nr etc.):'. The 'Others (nr etc.):' option is selected, and a dropdown menu below it shows 'Nucleotide collection (nr/nt)'. There is also an 'Organism' section with a text input field and an 'Exclude' checkbox. Below that, there are 'Exclude' and 'Entrez Query' sections with checkboxes and a text input field.



### Pro Tip!

If you know the species that your sequencing result should match, enter the common or scientific name into the Organism box. This small piece of information can significantly reduce your wait time for blastn, blastp, and blastx searches!

The screenshot shows the 'Choose Search Set' section of the BLAST interface, similar to the previous one. In this version, the 'Organism' text input field is highlighted with a red box. The 'Database' dropdown is still set to 'Nucleotide collection (nr/nt)'. The 'Organism' section also includes an 'Exclude' checkbox and a '+' button.



## TIPS FOR USING BLAST TO VERIFY PLASMIDS (CONT'D)

### Before You Click the BLAST Button...

Consider the Program Selection parameter, as this will affect the amount of time to perform the search as well as the overall alignment results. The default setting is “Optimize for Highly similar sequences (megablast)”, which is very fast and works best when the identity between your sequence and the reference/database sequence is  $\geq 95\%$  (Our QC process would be trouble-free and much faster if 95% of our results were always correct!).

Since sequencing reactions are imperfect and sequence near the beginning or end of a reaction is often unreliable, we routinely select the “Somewhat similar sequences (blastn)” program for blastn so that we can extract practically every single, reliable basepair from our results.

Program Selection

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST

Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

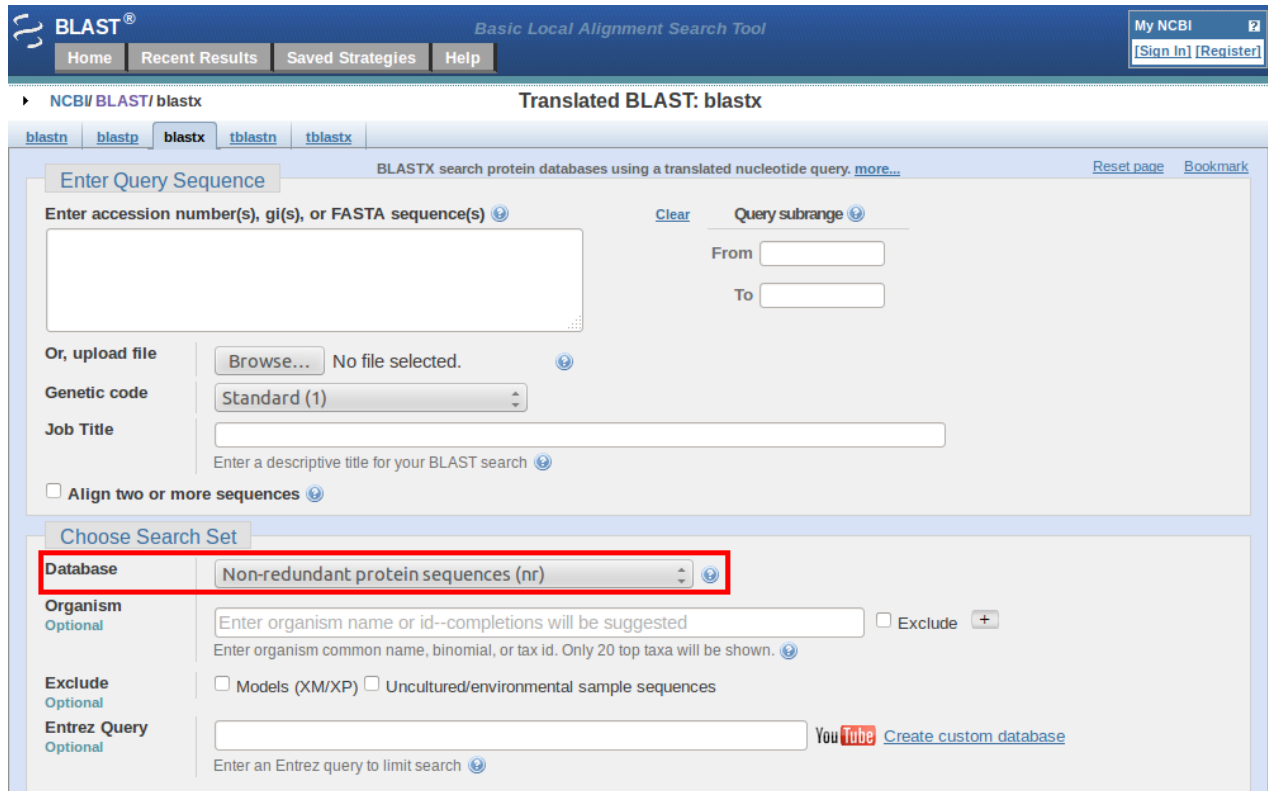
+ Algorithm parameters

This option is not as fast as megablast, but can return longer alignments to compare with your sequencing trace file. Unlike megablast, the regular blastn program uses a smaller word size and lower scoring penalties for mismatches and gaps in the alignment. If you are curious about the differences in the blastn programs, check out the [BLAST Help webpage](#).

### Optimizing *blastx* Searches

Once we have used blastn to determine the reliable portion of a sequencing result and noted any potential mismatches or gaps in the nucleotide sequence, we typically run a [Translated BLAST \(blastx\)](#) search to check for expected ORFs, mutations or truncations. A primary advantage of blastx is that you do not have to decide on a reading frame for your sequencing result – blastx checks all six possible frames against the database. Another benefit is that a frame shift mutation present in the ORF is readily apparent when viewing blastx results.

## TIPS FOR USING BLAST TO VERIFY PLASMIDS (CONT'D)



The screenshot displays the NCBI BLAST interface for a Translated BLAST (blastx) search. The page title is "Translated BLAST: blastx". The main section is titled "Enter Query Sequence" and includes a text input field for "Enter accession number(s), gi(s), or FASTA sequence(s)", a "Clear" button, and a "Query subrange" section with "From" and "To" input fields. Below this is an "Or, upload file" section with a "Browse..." button and "No file selected." text. The "Genetic code" is set to "Standard (1)". There is a "Job Title" input field with the placeholder text "Enter a descriptive title for your BLAST search". A checkbox for "Align two or more sequences" is present. The "Choose Search Set" section is highlighted with a red box around the "Database" dropdown menu, which is currently set to "Non-redundant protein sequences (nr)". Other options in this section include "Organism" (with a text input field and "Exclude" checkbox), "Exclude" (with checkboxes for "Models (XM/XP)" and "Uncultured/environmental sample sequences"), and "Entrez Query" (with a text input field and "Create custom database" link).

**When using blastx** at Addgene, we use the default “Non-redundant protein sequences (nr)” database as it contains the largest number of protein sequences. Just below the BLAST button, you may have noticed the “Algorithm parameters” link. Click on this link to view advanced BLAST options and for our suggested blastx customization. Similar to nucleotide sequences, proteins often have repeated or highly homologous regions, which by default are ignored in a standard blastx search. An alignment omitting repeated regions can be confusing, such as when you attempt to verify the starting methionine of a gene but the blastx results start the alignment at a more distal amino acid. We consistently run our blastx searches with the “Low complexity regions” filter unchecked so that these regions are included in the search to maximize the alignment length. While this recommendation is not infallible, we have found it saves analysis time to remove this default setting.

## TIPS FOR USING BLAST TO VERIFY PLASMIDS (CONT'D)

**BLAST** Search database **Non-redundant protein sequences (nr)** using **Blastx** (search protein databases using a translated nucleotide query)  
 Show results in a new window

**Algorithm parameters** Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

**General Parameters**

Max target sequences: 100  
 Select the maximum number of aligned sequences to display ⓘ

Expect threshold: 10 ⓘ

Word size: 3 ⓘ

Max matches in a query range: 0 ⓘ

**Scoring Parameters**

Matrix: BLOSUM62 ⓘ

Gap Costs: Existence: 11 Extension: 1 ⓘ

Compositional adjustments: Conditional compositional score matrix adjustment ⓘ

**Filters and Masking**

Filter:  Low complexity regions ⓘ

Mask:  Mask for lookup table only ⓘ  
 Mask lower case letters ⓘ



### Pro Tip!

blastx searches are inherently slower than blastn or blastp, due to the additional searches involved in translating the nucleotide sequence into all six possible reading frames. If you know the expected protein sequence, use the “Align two or more sequences” option to drastically reduce waiting time for search results.

## Optimizing *blastp* Searches

Depending on the sequencing result, we often choose between a [Standard Protein BLAST \(blastp\)](#) and blastx search to verify expected protein sequence in a plasmid. If you know which reading frame to choose for your sequencing result and can easily translate it, we recommend using blastp over blastx. The primary advantage is time savings but an added benefit is that blastp searches do not filter low complexity regions by default, meaning that you do not have to remember to adjust any blastp algorithm parameters. We use the default scoring matrix BLOSUM62, but you may want to check the description of the [other matrices](#) to see if another would be more advantageous for your search.

## TIPS FOR USING BLAST TO VERIFY PLASMIDS (CONT'D)

The screenshot shows the NCBI BLAST Standard Protein BLAST interface. The 'Database' dropdown menu is highlighted with a red box and contains the text 'Non-redundant protein sequences (nr)'. Other visible elements include the 'Enter Query Sequence' section with a text input field, 'Or, upload file' with a 'Browse...' button, 'Job Title' input, 'Align two or more sequences' checkbox, and the 'Choose Search Set' section with 'Organism', 'Exclude', and 'Entrez Query' options.

### Pro Tip!



Note that protein databases available are unlikely to have an exact entry for your favorite gene fused to an epitope tag or fusion protein. If your sequencing primer was chosen to confirm a tag or fusion protein is in-frame, we recommend using blastx with the “Align two or more sequences” option and pasting your expected protein sequence into the Subject Sequence box.

## BLAST Alternatives

Depending on your sequencing result and desired analysis, BLAST may not always be your optimal choice. For difficult sequence alignments that BLAST is unable to handle, [Clustal](#) is our frequent choice for pairwise or multiple sequence alignments of nucleotide or protein sequences. We also use [COBALT](#) for aligning multiple protein sequences, particularly for comparing different isoforms. In addition to our favorites, there are a number of sequence alignment tools available.

Try these resources for lists of alternatives to BLAST:

- ExPASy - [http://www.expasy.org/genomics/sequence\\_alignment](http://www.expasy.org/genomics/sequence_alignment)
- EMBL-EBI - <http://www.ebi.ac.uk/services>
  - o [http://www.ebi.ac.uk/Tools/webservices/#multiple\\_sequence\\_alignment\\_msa](http://www.ebi.ac.uk/Tools/webservices/#multiple_sequence_alignment_msa)
  - o [http://www.ebi.ac.uk/Tools/webservices/#pairwise\\_sequence\\_alignment\\_psa](http://www.ebi.ac.uk/Tools/webservices/#pairwise_sequence_alignment_psa)

## TIPS FOR USING BLAST TO VERIFY PLASMIDS (CONT'D)

### More Resources:

- [Sequence Analysis of your Addgene Plasmid](#)
- [Addgene Plasmid Guide](#)
- [Addgene Sequence Analyzer](#)

### Images:

- All BLAST images in this section are modified screen shots from the [NCBI BLAST website](#)

# OPTIMIZING PLASMID YIELDS

By Julian Taylor-Parker | May 26, 2016

Most of the time, plasmid prepping is a breeze. You get your stab from Addgene, streak for single colonies, sub-culture, and prep with one of the many commercially available DNA prep kits or your lab's favorite [in-house protocol](#). DNA yields for this procedure are typically in excess of 100 ng/ul, more than enough DNA to proceed with most applications, such as PCR, [cloning](#), [transfection](#), or long-term storage. But what about those pesky situations where your plasmid yield is sub-optimal? If you have already purified your plasmid, you can try to concentrate the DNA using a speed-vac,

[ethanol precipitation](#), or other chromatographic methods. But wouldn't it be nice to avoid an extra concentration step? If you are consistently getting sub-optimal plasmid yields from your prep, you may want to consider optimizing your growth conditions. In this section, we will outline many of the variables that could affect DNA yields and suggest steps to super-charge your plasmid preps.



## Copy Number

The copy number of a given plasmid refers to the number of copies supported within each bacterial cell of your culture. Copy number is ultimately limited by the plasmid's [origin of replication](#), but will also be affected, not only by the size and nature of the insert, but also the [propagation strain](#), growth conditions, and inoculum (see below). To further complicate matters, plasmid copy numbers are often ambiguous at best and many times should be viewed more as a guideline than a rule since they are typically based on the copy number of the empty plasmid backbone or a researcher's personal observations. If you are experiencing low yields from a plasmid whose copy number is unknown, it is entirely possible that you are working with a low copy plasmid.

When working with low copy plasmids, the first step to improving yield is often to start with more culture volume and many common plasmid prep kits explicitly state this in their protocols. A higher volume of lysate can be passed over the binding matrix, improving yield in the eluate; however, you do need to be cautious to stay within the recommended range of the filter. Overloading the column with too much lysate could clog the matrix and inadvertently lower the amount of DNA you purify! If you've reached your lysate volume limits, the copy number of some some origins of replication (e.g. p15A, ColE1), can be increased by adding [chloramphenicol to the culture medium](#).

## Bacterial Strain

There are many commercially available strains for the propagation and purification of plasmid DNA, and often it is difficult to know which will work best in your particular experiment. For example, certain *E. coli* strains such as HB101 and its derivatives may be appropriate for long term storage, but these also produce carbohydrates, which could reduce lysis efficacy and plasmid yield. Other strains, such as Stbl3, are engineered to reduce recombination and therefore better propagate unstable plasmids, but are also endA+. endA encodes a thermostable periplasmic endonuclease. If not properly removed during the wash steps of your purification protocol, EndA can co-purify with your plasmid DNA, and may shear the plasmid causing a smear when run on a gel.

DH5alpha has consistently been shown to give good, high quality preps due to its *endA1*, *recA1*, *relA1* genotype, as these mutations improve plasmid stability and yield, and this strain (or one of its derivatives) is recommended

## OPTIMIZING PLASMID YIELDS (CONT'D)

whenever possible. While choice of strain will ultimately depend on the particular features of your plasmid, it is advisable to check out the genotype of your *E. coli* to ensure it is suitable for your application. [This publication](#) looked at the DNA quality of plasmids isolated from different *E. coli* hosts. Additionally, we have previously reviewed many [common lab \*E. coli\* strains](#) on the blog, and also recommend this handy guide from [OpenWetWare](#) for a more extensive listing of common *E. coli* strains.

### Culture Media, Antibiotics, and Inoculum

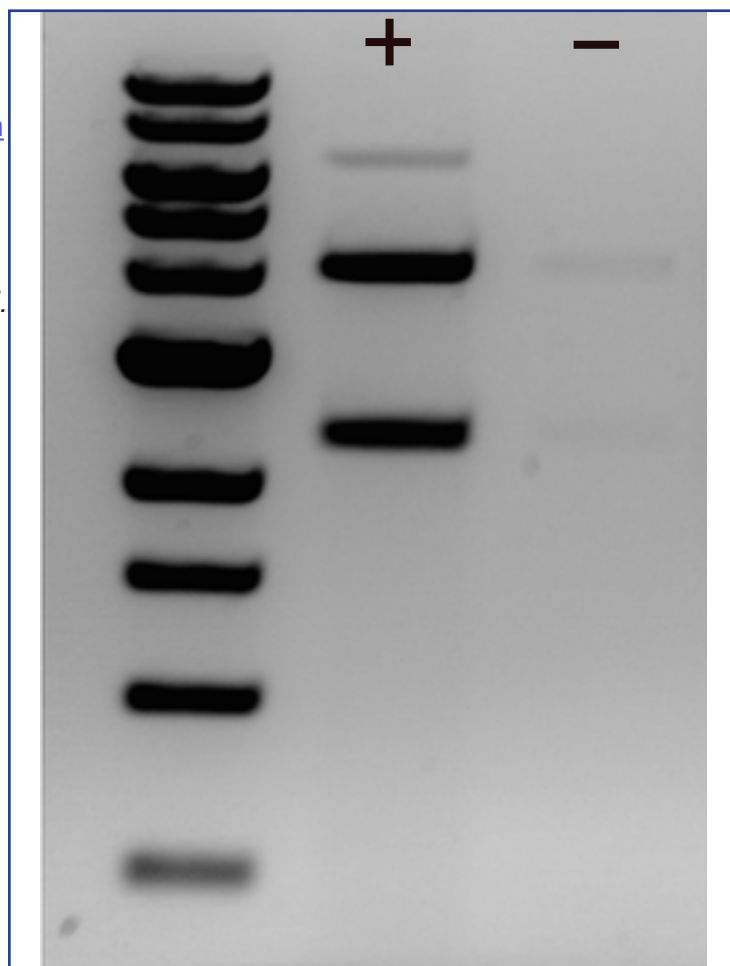
Culture conditions play a critical role in plasmid DNA yield as is demonstrated in the figure on the right. Different plasmids and strains will vary in their optimal growth conditions. For many high copy plasmids, standard LB broth and antibiotic concentrations will work just fine; however, for low copy plasmids or slower growing strains, a change in media may be advantageous. Low copy plasmids produce fewer transcripts of their antibiotic resistance genes, and can therefore be cultured in media containing half the normally [recommended antibiotic concentration](#).

Additionally, you can achieve higher cell density using a nutrient-rich broth such as Terrific Broth, 2xYT, or a home-brew optimized for plasmid yields (e.g. [H15](#)). Spiking your media with supplements such as magnesium salts, buffering agents, and/or supplying additional carbon sources such as glycerol or glucose (in moderation) may also serve to increase cell density and plasmid yield. For best results and highest DNA yields, it is always recommended to start from a single, fresh colony. Sub-culturing directly from a frozen glycerol stock or agar stab may lead to loss of the plasmid, and using older plates could increase plasmid loss or mutation.

### Time, Temperature, and Oxygen

Lastly, you can optimize your growth time, temperature, and shaking speed to maximize cell density and plasmid yield as these factors are generally, but not perfectly, correlated. Typical growth times for high copy plasmids in standard growth strains range from 12-16 hours, but cultures with lower copy plasmids often need to be grown for 20 hours or more to achieve maximum plasmid yield. The optimal growth time should be determined for each plasmid/strain combination individually, either by measuring OD or, ideally, by harvesting/miniprepping at various timepoints.

Additionally, certain plasmids and strains are best suited for growth at temperatures other than the standard 37° C. This should be indicated in the plasmid information or growth strain instructions from the manufacturer.



*Growing bacteria on agar with (+) and without (-) antibiotic prior to overnight growth can affect plasmid yield - fainter bands in the (-) lane show that much less plasmid was isolated from this sample. Image courtesy of Amanda Hazen.*

## OPTIMIZING PLASMID YIELDS (CONT'D)

---

Finally, you will want to make sure your culture is getting the right amount of oxygen for optimal growth as insufficient gas exchange will prevent your cultures from reaching the desired density. The volume of your flask or culture tube should be at least 4x greater than your total culture volume and you should ensure the shaker speed is fast enough for sufficient gas exchange in the overnight culture. Generally, large cultures in flasks may be shaken around 220 RPM, but smaller cultures, especially those in deep-well plates, require a faster shaking speed for proper aeration (260-300 RPM).

### Further Reading

1. Sigma [Microbial Media Guide](#)
2. BitesizeBio: [How to Get Better Plasmid Midiprep Yields](#)
3. BitesizeBio: [Better Plasmid Midipreps Part II: What Causes Low Yields](#)



# CONTROL PLASMIDS

By Chari Cortez, Addgene | Apr 9, 2015

There are many, many different types of experiments carried out by scientists everyday. Although the designs and outcomes may vary, one thing should be present in every experiment-based investigation of a hypothesis: proper controls!

For every experiment, an investigator needs a standard against which the results can be compared; results from an experiment lacking the proper controls are invariably inconclusive and unreliable. Proper controls provide the constant variables that enable the correct interpretation of the effect of the independent variable you are testing. Importantly, they demonstrate the functionality of your experimental system and help identify opportunities for troubleshooting or optimization within your experiment.

Read on to learn more about the various controls that can be used for plasmid-based experiments.



## What Are Control Plasmids?

Generally, control plasmids help to ensure that the observed phenomenon in your experiment is specifically associated with the independent variable you are testing and not some other unintended factor. Control plasmids are used to minimize any effects of the non-independent variables.

To illustrate the importance of control plasmids, we will take a step beyond discussion in the abstract and follow an example of a plasmid-based experiment, delineate its requisite control plasmids, and discuss why these control plasmids are critical for the correct design of the experiment.



Figure 1: Experiment – attempt to knock down the expression of Gene X using shRNA expressed from plasmid A

## CONTROL PLASMIDS (CONT'D)

---

The results shown above are from a single experiment wherein Plasmid A (encoding shRNA against human Gene X in backbone Y) was transfected into human cells.

From this result one could simply conclude that the shRNA didn't work, as the expression level of Gene X in cells treated with Plasmid A is similar to the expression level of Gene X in the untreated cells—but can we be certain? Do we know whether the plasmid delivered into the cells successfully? Is the shRNA target correct? Is the purified plasmid DNA viable? Or cytotoxic? How and where did things go wrong? The use of proper control plasmids would address and answer these questions.

### Types of Control Plasmids

Part of planning your experiment includes determining what factors need to be controlled for in order to eliminate any alternative interpretation of the results. Typically, plasmid-based experiments employ transfection, negative, positive, and replicate controls.

#### *Transfection Controls: Empty Vector and Internal Control*

A transfection control measures transfection efficiency and enables observation of any effects of the transfection itself (i.e., the vector used in [transfection](#), transfection reagent, or transfection process) may have on the target cells. One transfection control is an [empty vector](#) control; specifically, the plasmid without the independent variable. Referring back to the experiment associated with Figure 1, the independent variable is the shRNA. Therefore the empty control vector would be Plasmid A sans shRNA, or backbone Y alone. The empty vector control allows you to examine if the transfection reagents or the transfection process itself has any cytotoxic effects on the target cells.

Another type of transfection control is an internal control vector, which measures transfection efficiency. An internal control may be a plasmid that constitutively expresses a reporter protein (e.g., [GFP](#) or [luciferase](#)) that is either co-transfected with the test plasmid or transfected into a separate well of your cells. Regardless, the amount of reporter protein activity correlates to both the amount of DNA transfected into the cells and the ability of the cells to express the protein.

In analysis of the result in Figure 1, an internal control, such as the GFP-expressing Plasmid B, could demonstrate whether the cells were transfected successfully and expressing the protein. For example, fluorescence microscopy images resulting from our experiment that includes the aforementioned internal control and is consistent with the result in Figure 1 could look like those found in figure 2.

This result indicates that the transfection was not successful due to the absence of GFP fluorescence from the living, viable cells. This result presents an opportunity for troubleshooting and optimization of the transfection reagents and process. It is important to note that optimal experimental conditions, including how much plasmid DNA to use for any individual or co- transfection, should be determined empirically.

#### *Negative Controls: Untreated Cells, Empty Vector Control, and Non-Targeting Control*

Negative control conditions and plasmids should produce a null effect (i.e., no phenomenon is observed). In any plasmid-based experiment, untreated cells should be included as these provide the baseline/standard against which other samples can be compared.

## CONTROL PLASMIDS (CONT'D)

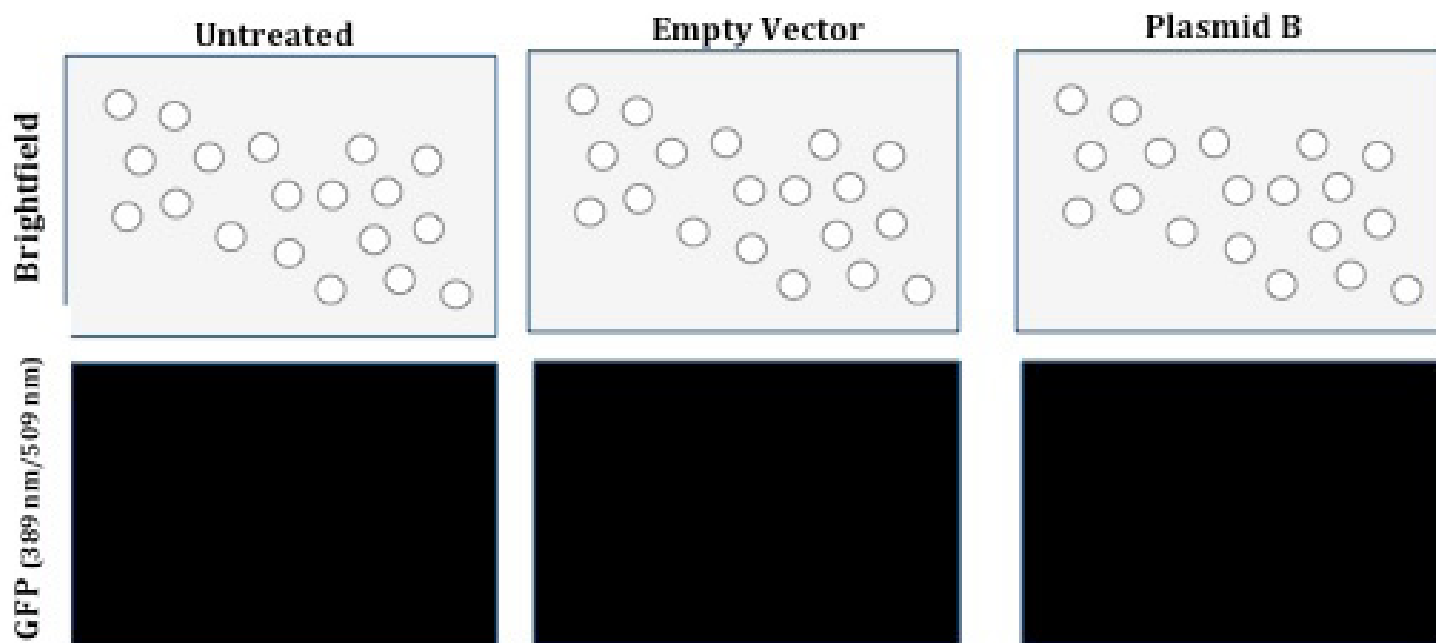


Figure 2: shRNA knock down experiment transfection control

The Empty Vector Control (mentioned above) could also serve as an important negative control. In this context, the empty vector control shows any effect of the vector/backbone itself on gene expression in your target cells.

In experiments employing gene targeting or genome editing technologies, such as RNAi or CRISPR, non-targeting controls may be appropriate as they allow you to assess the specificity of your observed result. [Non-targeting](#) controls are negative controls that produce a similar product, but do not target an endogenous gene in your experimental cells. For example, in the experiment above, Plasmid A contains an shRNA that targets human Gene X and human cells are being transfected. A proper non-targeting control in this experiment could be an shRNA—in the same backbone as Plasmid A—that does not target any mammalian gene. This non-targeting control is critical to the correct interpretation of the results because it provides an important reference point when analyzing the specificity of the shRNA targeting human Gene X. The non-targeting control also assesses any effects of the general introduction of shRNA into your target cells.

### Positive Controls

Positive control plasmids should produce the expected phenomenon. One example of a positive control, the internal control vector, was described earlier. Once you are sure your conditions are conducive to the successful delivery of plasmid DNA into the cells, the internal control vector then serves as a positive control for transfection because it produces the expected effect, which is green fluorescent cells (Figure 3).

Other positive controls are specific to the experiment and should be designed accordingly. If you are trying to activate a gene, you should design a control that shows maximal activation. Likewise if you are trying to repress a gene, your control might be a system where expression of that gene is knocked out completely. These controls may or may not be plasmid-based depending on the experimental needs. Using our experiment as an example, the expected result of the shRNA targeting human Gene X in Plasmid A is the decreased expression of Gene X. Ergo the positive control(s) should decrease the expression of Gene X.

## CONTROL PLASMIDS (CONT'D)

Since the key feature of a control plasmid is to minimize the effects of the non-independent variables within an experiment, both the design and selection of the positive control plasmids should be highly specific to the experiment and the interrogation of the independent variable.

### *Replicate Controls: Technical and Biological*

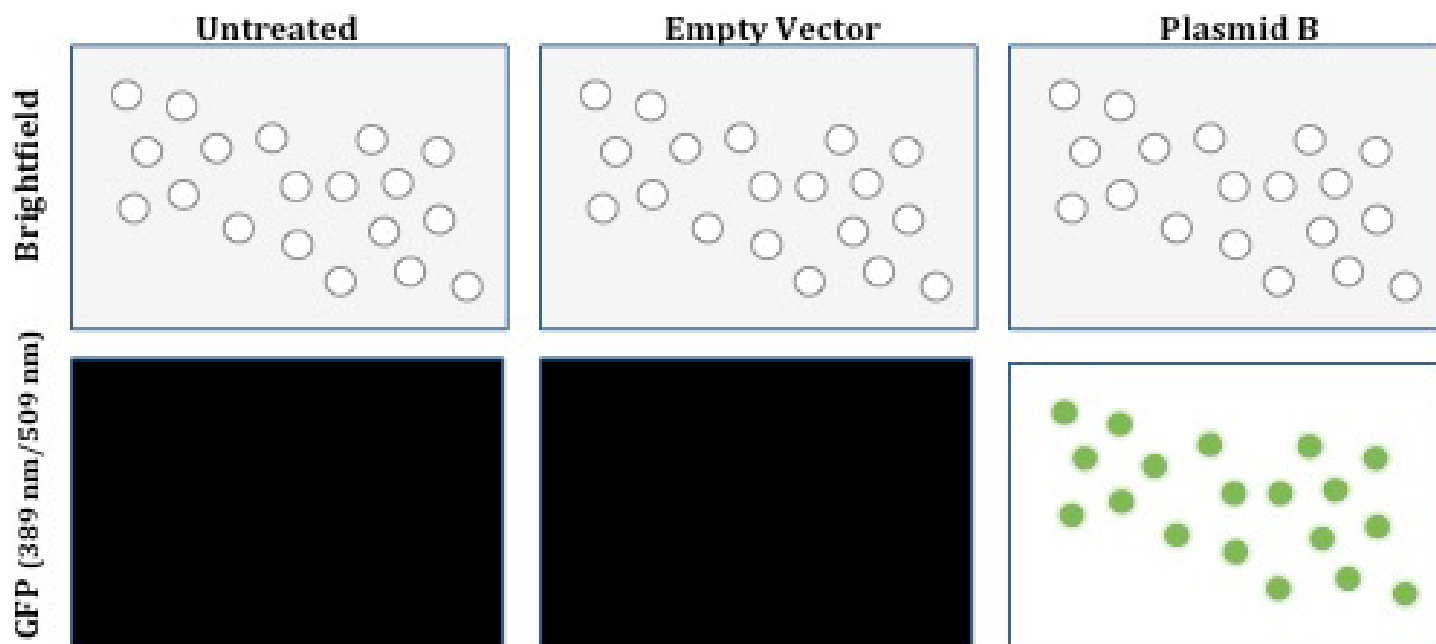


Figure 3: Successful transfection control

Reliable results are reproducible from properly controlled experiments. The observed phenotype associated with the independent variable should be consistent over time and from different preparations of the purified test plasmid expressing the independent variable. There are two types of replicate controls: technical and biological.

In general, technical replicates can be thought of as “plate controls”. They are NOT independent and are typically derived from one source. Conversely, biological replicates ARE independent and can be thought of as “reproducibility controls”. Biological replicates are what makes up your sample size (aka your n value) and should come from multiple, independent sources. We’ll describe these a bit more below, but the papers cited in the reference section provide more in depth information.

An example of a technical control is the transfection of multiple separate wells (within the same plate) with purified plasmid from the same aliquot/preparation. The replicate control isn’t measuring or assessing reproducibility of the effect of the independent variable because the purified plasmid, cells, and media used in each of the wells are not independent: they were derived from the same source and were incubated on the same plate at the same time. Although the replicate control is important, it speaks to the consistency of the reagents and hands performing the experiment, not the reproducibility or consistency of the observed phenotype associated with the independent variable.

## CONTROL PLASMIDS (CONT'D)

In any given set of experiments, the biological controls are more time consuming and ultimately more important. Ideally, each biological replicate should use fresh media, test independent aliquots of cells and plasmids, be performed on different days, etc, but outside constraints may impose some limitations and you should be cognizant of these when interpreting your results. The key to biological controls is independence: at the very least you should repeat your whole experiment from start to finish multiple times and on different days. Using our example above, we would test different preparations of Plasmid A in different aliquots of cells on several different days. Biological replicates control for the reproducibility and consistency of the observed phenotype associated with the independent variable and helps ensure the phenotype isn't singular or associated with only one aliquot/preparation of the test plasmid.

Let us now revisit our experiment. In Figure 1, it appeared as though the shRNA did not knock down expression of Gene X but, as shown in Figure 2, this was likely due to the original transfection conditions. Now that we have successfully transfected our cells (Figure 3), we can continue with our experiment, incorporating additional positive and negative controls, performing multiple replicates, and ultimately getting interpretable results:

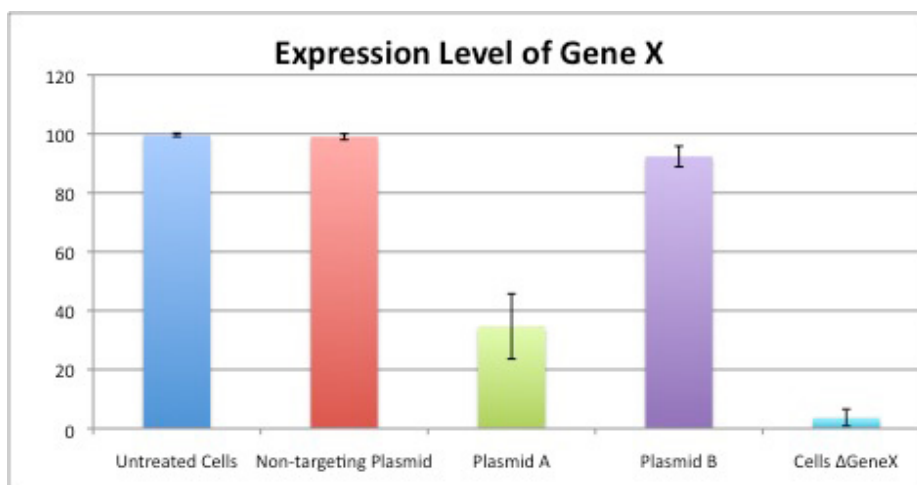


Figure 4: Measuring the expression level of gene X with all appropriate controls

The design and selection of proper experimental controls is not a trivial endeavor, as biological systems have many variables. However daunting designing and executing these steps may be, proper controls are a basic tenet of responsible scientific inquiry and investigation.

### Further Reading

1. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? Lazic SE. *BMC Neurosci.* Jan 14;11:5. (2010). [PubMed PMID: 20074371](#).
2. BitesizeBio: [How to Get Better Plasmid Midiprep Yields](#)
3. Replicates and repeats – what is the differences and is it significant? Vaux DL, et al. *EMBO Rep.* Apr; 13(4): 291-296. (2012). [PubMed PMID: 22421999](#).

## CHAPTER 9: DEPOSITING YOUR PLASMIDS WITH ADDGENE



# A BRIEF HISTORY OF ADDGENE

By Melina Fan, Addgene Co-founder | August, 2015

---

I founded Addgene as a nonprofit resource for scientists to share plasmids. The idea for Addgene came from difficulties that I encountered when requesting materials from other labs during my PhD thesis. After speaking to several Principal Investigators, it became clear that the reason labs were not sending materials was because of logistical hurdles rather than philosophical ones. Turnover of lab members made it difficult to locate plasmids and cloning data, and managing shipments took time away from busy research schedules.

When Addgene opened its (freezer) doors in 2004, the scientific community embraced the resource and began sharing materials. Over 3,000 labs from around the world have contributed over 52,000 plasmids for use in a wide variety of model organisms and research applications. In turn, we've helped those labs fulfill over 750,000 plasmid requests. From Nobel Prize winners to scientists who are just starting out, I've been inspired by the community's dedication to open science.

I encourage you to join today's scientific leaders by making your materials globally available through Addgene.

Melina Fan, PhD  
Cofounder and Chief Scientific Officer of Addgene



*Melina is a cofounder and Chief Scientific Officer of Addgene. She received her BS from MIT and her PhD from Harvard University where she studied diabetes and obesity. She also cofounded LabLife, a lab management software company. She loves that her job at Addgene connects her with scientists around the world who are working on everything from cancer to biofuels.*

# BENEFITS OF DEPOSITING

Various Addgenies | August, 2015

---

[Addgene](#) is a global, nonprofit plasmid repository dedicated to making it easier for scientists to share plasmids. Scientists from labs like yours deposit plasmids to Addgene's collection. Deposited plasmids are stored and distributed for use in research and discovery worldwide.

## Stay Organized...

Addgene manages archived samples for your lab, storing multiple copies at our site in Cambridge, MA and a copy at a partner facility offsite. All of your lab's deposited plasmids and cloning information are curated online for easy access.

## Save Time...

You can easily direct requests for your plasmids to your personalized Addgene page. When your plasmids are requested, Addgene processes and ships your plasmids for you.

## Increase Citations...

Scientists who request your plasmids cite your publication and acknowledge your lab in their published work.

## Track Reagents...

Receiving monthly updates on how many times your plasmids have been requested and who requests them.

## Improve Community Resources...

Contributed materials are available to the global research community.

## Increase Visibility...

Addgene's online plasmid database helps get your plasmids and your lab get exposure beyond your immediate field.

## Earn Free Plasmids...

As your plasmids are requested from Addgene, you earn reward points to use toward free plasmids.

## Sign Fewer MTAs...

Addgene's one-time deposit agreement means that once your plasmids are online, we'll facilitate agreements between your organization's tech transfer office and that of your requesters so that you don't have to.

## Manage Support Queries...

Addgene's knowledgeable and friendly support team answers questions about your deposited plasmids.



# THE DEPOSIT SPREADSHEET

By Tyler J. Ford | February, 2017

So you've made a plasmid (or plasmids) and would like to share them with the scientific community but don't know how to start. At Addgene, we try to make the [plasmid deposit process](#) as easy as possible. Our process can be broken down into three simple steps:

1. Send Plasmid Information to Addgene
2. Send Physical Plasmids to Addgene
3. Quality Control and Materials Transfer Agreement approval (MTA)

Steps 2 and 3 are usually very easy - once we have all of your plasmid data we'll send you prepaid shipping materials (i.e. a [deposit kit](#)) with instructions on how to send liquid DNA or bacterial streaks of your plasmid back to us. For step 3, our scientists will contact you if any issues arise during the QC process and our tech transfer team will communicate with your university directly to make sure the MTA is taken care of. As the depositing scientist, you will have the most involvement with step 1 and, while data entry can be a chore, we've made it easier than ever to send us your plasmid information with the Deposit Spreadsheet.

## The Data Entry Process

On the Addgene home page, if you scroll over the "Deposit Plasmids" tab at the top of the page and click "[Start Deposit](#)," you'll see that you have three options for data entry (see image on the left). The first two options will bring you to our online data entry process. A thorough guide to this process can be found [here](#).

The third option - the Deposit Spreadsheet - is relatively new but is gaining popularity and for good reason. If you're depositing 15 or more plasmids, this is the quickest option. Simply download the spreadsheet and each plasmid in your deposit gets one row where you'll input all necessary plasmid information as outlined by the column headers and (required columns are shown in red in the spreadsheet and described in the table below). Some cells have drop-down options for you to choose from while others need to be filled with short answer text. You can also directly paste sequence data into the appropriate full and partial sequence cells.


The spreadsheet makes it very easy to enter information for a group of plasmids with the same backbone - you just need to copy and paste any repeated information into the appropriate rows.

If you ever have a question about what should go into a cell, hover over the cell and instructions will pop up. If you plan on submitting more than 50 plasmids or have additional questions, please contact [help@addgene.org](mailto:help@addgene.org).

1

**Submit Plasmids from a Published Article**


- Look up your article as you would search on PubMed (i.e. title, author, or PMID).
  - No PubMed ID yet? Please use the pre-publication/unpublished submission option below.
- Addgene will provide links from your plasmids to this article.
- Scientists will be asked to cite this article in future publications.
- Must be [logged in](#) to start deposit.



2

**Submit Pre-Publication or Unpublished Plasmids**


- Enter a descriptive title (e.g. Smith lab MAPK plasmids).
- [To hold plasmids for publication](#): enter HOLD + your descriptive title (e.g. *HOLD--Smith lab MAPK plasmids*).
  - We can easily update the article information when the article becomes available.
- Addgene encourages submitting [pre-publication plasmids](#) so they can be available online when the paper is published.
- Must be [logged in](#) to start deposit.




3

**Submit Plasmids Using a Spreadsheet**

- Recommended for depositing 15 or more very similar plasmids (e.g. gRNA plasmids in the same backbone).
- Copy and paste your plasmid data directly into our file.
- Email the spreadsheet back to us at [deposit@addgene.org](mailto:deposit@addgene.org) along with your plasmid sequences, maps, or GenBank files, and shipping address.
  - We can accept sequence files in **any** format.



 **addgene**  
A better way to share plasmids

185 | Page

# THE DEPOSIT SPREADSHEET (CONT'D)

Plasmid Name	Plasmid Type	Purpose	PubMed ID	Gene or insert Name	Alternative Gene/Insert Name 1	Alternative Gene/Insert Name 2	Entrez Gene ID

Please enter a brief description (<200 characters) of the plasmid's purpose to answer the question "What does this plasmid do?"

When you're finished filling out the spreadsheet, save it in the format:

labname\_addgene\_batch\_upload.xls

Replace labname with the name of your PI and send the spreadsheet to [deposit@addgene.org](mailto:deposit@addgene.org) with any associated plasmid maps and sequence files in any format as well as your preferred shipping address. From there, we'll use the information in the spreadsheet to generate plasmid pages that we'll send you for review. We'll also mail you a deposit kit you can use to send the physical plasmids back to us.

We hope the deposit spreadsheet will greatly expedite your data entry process and make it easier than ever to share your plasmids with the scientific community.

Column Title	Description
Plasmid Name	The name of the plasmid as it appears in its publication or as it is commonly known in your lab
Plasmid Type	Choose from: Encodes one insert, encodes gRNA/shRNA, or empty backbone
Purpose	Answer "What does this plasmid do?" in <200 characters
Species of Gene or Insert	Choose from <i>H. sapiens</i> (human), <i>M. musculus</i> (mouse), <i>R. norvegicus</i> (rat), <i>G. gallus</i> (chicken), <i>B. taurus</i> (bovine), <i>X. laevis</i> (frog), <i>D. rerio</i> (zebrafish), <i>D. melanogaster</i> (fly), <i>C. elegans</i> (nematode), <i>S. cerevisiae</i> (budding yeast), <i>S. pombe</i> (fission yeast), <i>A. thaliana</i> (mustard weed), synthetic, or other. If other, the species can be indicated in a separate column.
Relevant Mutations	Enter any mutations in the gene/insert. If WT, leave this empty.
Primary Vector Type	Choose from mammalian expression, bacterial expression, lentiviral, retroviral, AAV, RNAi, luciferase, cre/lox, yeast expression, worm expression, plant expression, mouse targeting, CRISPR, TALEN, synthetic biology, unspecified, or other
Cloning Method	Choose from restriction enzyme, TOPO cloning, Gateway cloning, ligation independent cloning, Gibson cloning, or unknown

## THE DEPOSIT SPREADSHEET (CONT'D)

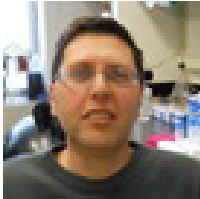
Column Title	Description
Bacterial Resistance	Enter bacterial antibiotic resistance encoded in the plasmid choose from: ampicillin, kanamycin, chloramphenicol, tetracycline, spectinomycin, hygromycin, bleocin (zeocin), gentamycin, streptomycin, nourseothricin (clonNat), combinations of the above, or other
High or Low Copy	Choose from high or low copy. Choose “high copy” if a sufficient quantity of DNA is produced from a miniprep and “low copy” if special growth conditions are required or the plasmid is difficult to grow.
Growth Temp	Choose from 30C, 37C, or room temp.
Growth Strain	Choose the strain addgene should use to distribute this plasmid. Choose from: DH5alpha, NEB Stable, ccdB Survival, or other
Hazardous	If the plasmid will produce anything dangerous or toxic to humans in bacteria, select “yes” in the drop-down and contact Addgene
Patents or Licenses	If there are any patents or licences that could prohibit Addgene from distributing the plasmid, select “yes” from the drop-down and contact Addgene.

# ACKNOWLEDGEMENTS AND FINAL WORDS

---

## Special thanks to our Guest Contributors!

Gal Haimovich



*Gal Haimovich, Ph.D., is a research fellow in the lab of Prof. Robert Singer at Albert Einstein College of Medicine. He is interested in everything related to gene expression, particularly at the RNA level. He maintains [greenfluorescentblog.wordpress.com](http://greenfluorescentblog.wordpress.com).*

Kurt Thorn



*Kurt Thorn is an associate professor at UCSF, where he directs the Nikon Imaging Center. He received his Ph.D. in biophysics from UCSF in the laboratory of Ronald Vale, after which he was a fellow at the Bauer Center for Genomics Research at Harvard University. Learn more at [his lab web page](#) or [his microscopy blog](#).*

Didem Goz



*Didem Goz Ayturk is a Postdoctoral Fellow in Connie Cepko's Lab investigating neuronal circuits of the retina using viral tools.*

Contributions to this eBook have been made by the following Addgenies past and present:

Michelle Cronin



*Michelle Cronin is an Addgene Senior Scientist, with a background in neurobiology, molecular biology, and biochemistry.*

Chari Cortez



*Chari was previously a Scientist at Addgene. Her primary research experiences are in Immunology, infectious disease, and host cell receptor signaling in the context of initiating inflammation and host-pathogen interactions. She loves Science, T cells, signaling pathways, and writing about Science!*

## ACKNOWLEDGEMENTS AND FINAL WORDS

---

Melina Fan



*Melina Fan is a cofounder and Chief Scientific Officer at Addgene. She also cofounded LabLife, a lab management software company. She loves that her job at Addgene connects her with scientists around the world who are working on everything from cancer to agriculture.*

Matthew Ferenc



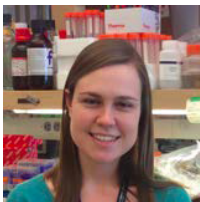
*Matt was formerly a Senior Scientist at Addgene and received his Ph.D. from Boston University. He is currently a Manager I, Publications and Analytics at Takeda.*

Tyler J. Ford



*Tyler J. Ford is an Outreach Scientist at Addgene. His professional duties include helping maintain the Addgene blog ([blog.addgene.org](http://blog.addgene.org)), talking to people about Addgene, and improving Addgene's services. His non-professional duties include running, biking, drawing, hiking, playing tennis, reading, and writing. Follow him on Twitter [@TyFordFever](https://twitter.com/TyFordFever).*

Mary Gearing



*Mary Gearing is a Scientist at Addgene. She got her start as a Science Communications Intern writing for the Addgene blog and website. As a full-time Addgenie, she still enjoys blogging about CRISPR and other cool plasmids! You can follow Mary on Twitter at [@megearing](https://twitter.com/megearing).*

Benoit Giquel



*Benoit Giquel is an Outreach Scientist at Addgene. Based in Europe, he is helping European scientists and institutions share their plasmids via Addgene. When he is not travelling for Addgene, Benoit likes playing the guitar, running, watching sports, and cooking for his family. Follow him on twitter [@bengiquel](https://twitter.com/bengiquel).*

Larissa Haliw



*Larissa Haliw was previously an Outreach and Marketing Associate at Addgene and now works as a Marketing Specialist at Abcam.*

## ACKNOWLEDGEMENTS AND FINAL WORDS

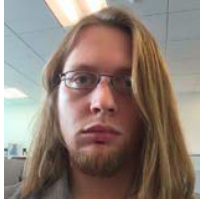
---

Jane Hannon



*As the Marketing & Outreach Specialist at Addgene, Jane's primary duties include spreading the word about her favorite non-profit plasmid repository and encouraging everyone to follow [@Addgene](#) on Twitter (nudge nudge...). Outside of work, you'll typically find her running, baking, or brushing up on the latest neuroscience news.*

A. Max Juchheim



*Max is a molecular biologist who does quality control sequence analysis for Addgene. He did his undergraduate studies at MIT where he worked on yeast genetics, and his graduate work at Harvard Medical School, where he studied mouse models of colon cancer. He loves science, but usually prefers to work "behind the scenes", so he's not quite sure how he was roped into writing for Addgene's blog!*

Beth Kenkel



*Beth Kenkel is currently a research scientist in the Department of Laboratory Medicine at the University of Washington. She is particularly interested in science communication and in vitro diagnostics.*

Caroline LaManna



*Caroline is currently the Associate Director of Scientific Outreach at Addgene. She received her Ph.D. in Biomedical Engineering from Boston University. Her goal is to make it easier for scientists to share reagents, communicate their research to others, and teach science in the community.*

Michael G. Lemieux



*Michael G. Lemieux is a Scientist at Addgene with interests in education and scientific writing.*

Diana Mai



*Diana Mai is an Outreach Assistant at Addgene. She is interested in layout, design, and organizing data. You can reach her at [diana@addgene.org](mailto:diana@addgene.org)*

## ACKNOWLEDGEMENTS AND FINAL WORDS

---

Emma Markham



*Emma Markham is an Operations Assistant at Addgene. She helps coordinate a variety of efforts across teams at Addgene and participates in European outreach.*

Margo Monroe



*Margo Monroe was previously a science blogger at Addgene and now works as a Staff Scientist at Hall & Steward LLP.*

Kendall Morgan



*Kendall K. Morgan, Ph.D., is a science writer based in North Carolina. She writes about science, health, and medicine for many organizations and publications with the goal to help researchers share their work with each other and the world.*

Jason Niehaus



*Jason L. Niehaus is the Associate Director of Biology at Addgene and enjoys the intersection of molecular biology and computing.*

Marcy Patrick



*Marcy is a Senior Scientist at Addgene. She received her Ph.D. in Microbiology and Immunology from the University of Michigan. She loves that her job allows her to interact with scientists from all over the world and help them easily share their reagents.*

Eric Perkins



*Eric J. Perkins is Addgene's Senior Scientific Project Lead. He used to say his background was in DNA repair, but now that's he's been at Addgene longer than he was in graduate school, he can now say his background is in a little bit of everything.*

## ACKNOWLEDGEMENTS AND FINAL WORDS

---

Brook Pyhtila



*Brook Pyhtila is a Scientific Curator at Addgene. She received her PhD at Stanford and did her post-doctoral work at Duke. She enjoys contributing to Addgene's database and web content so that scientists can more easily do their research.*

Maria Soriano



*Maria Soriano-Carot is an Outreach Scientist at Addgene and is particularly interested in Mentoring and Career Development in Science. You can follow her on Twitter @[SorianoCarotM](#).*

Lianna Swanson



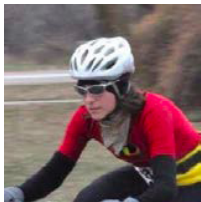
*Lianna graduated from Technion Israel Institute of Technology with a BA in Biology and went on to pursue her Ph.D. at Northwestern University in Evanston, IL. During her graduate school she worked with frogs, yeast, and (mainly) flies. She has extensive molecular biology knowledge and a fair amount of biochemistry experience (bacterial expression, protein purification and western blotting). Lianna is the Director of Biology at Addgene and has been with the organization since 2008.*

Julian Taylor-Parker



*Julian Taylor-Parker is a Scientist at Addgene with a background in molecular medicine and microbiology. He is currently working on earning his MBA with a focus on the health and biotechnology sectors.*

Nicole Waxmonsky



*Nicole is an Outreach Scientist and Curator at Addgene. When not riding her bike or playing broomball, Nicole enjoys using the web to increase scientific sharing and to improve access to basic science information.*

Jessica Welch



*Jessica Welch is an Outreach Scientist at Addgene. She moved to Boston from Australia in 2011 and her interests include microbiology, science education, and helping scientists all over the world share resources.*



## ACKNOWLEDGEMENTS AND FINAL WORDS

---

If you have any questions, comments, or suggestions about how Addgene can improve its educational content, please contact us at [blog@addgene.org](mailto:blog@addgene.org).

Looking for more information about Addgene?  
Connect with us!



[addgene.org](http://addgene.org)



[blog.addgene.org](http://blog.addgene.org)



[facebook.com/addgene](https://facebook.com/addgene)



[twitter.com/addgene](https://twitter.com/addgene)



[linkedin.com/company/addgene](https://linkedin.com/company/addgene)



[youtube.com/addgene](https://youtube.com/addgene)



[blog.addgene.org/topic/podcast](http://blog.addgene.org/topic/podcast)