Plasmids 101

July 2023
4th Edition

addgene

Available as part of our 101 collection.

Plasmids 101

Viral Vectors 101

Fluorescent Proteins 101

CRISPR 101

Science Career Guide

Find us

Addgene.org

blog.addgene.org

facebook.com/addgene

twitter.com/addgene

instragram.com/addgene

youtube.com/@addgene

linkedin.com/company/addgene

INTRODUCTION TO PLASMIDS 101

By the Addgene Team

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, 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.

CONTENTS

CONTENTS

CHAPTER 1

What is a Plasmid?

A Brief History of Plasmids

By Marcy Patrick | 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.

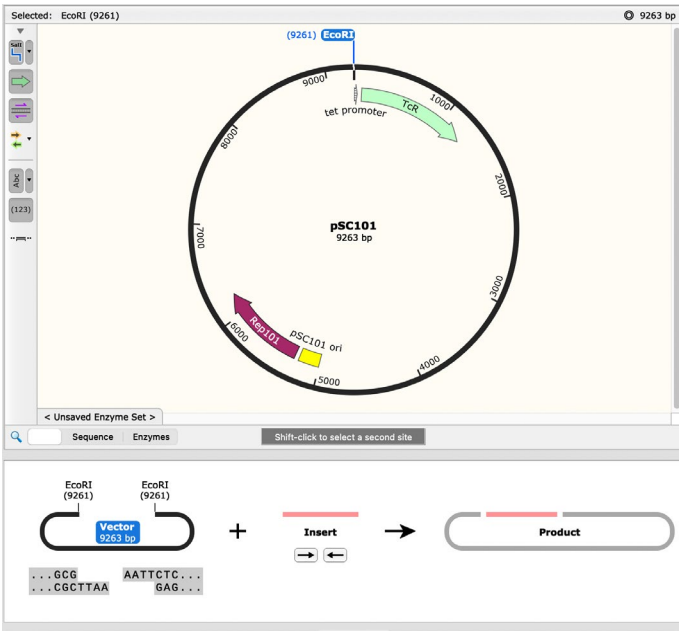
SO HOW DID PLASMIDS GET THEIR NAME?

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 extrachromsomal 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.

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 a 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

CSHL Meeting: Plasmids: History & Biology

The Joshua Lederberg Papers

Joshua Lederberg's Personal Perspective: Plasmid (1952-1997)

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.

Cell Genetics and Hereditary Symbiosis Lederberg, Joshua. Physiological Reviews 32.4 (1952) 403-430.

Sex Compatibility in Escherichia Coli. Lederberg, Joshua, Luigi L. Cavalli, and Esther M. Lederberg. Genetics 37.6 (1952): 720–730. PubMed PMID: 17247418.

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.

Transmissible Drug Resistance in an Epidemic Strain of Salmonella Typhimurium. Datta, Naomi. The Journal of Hygiene 60.3 (1962): 301–310. PubMed PMID: 14025218.

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.

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?

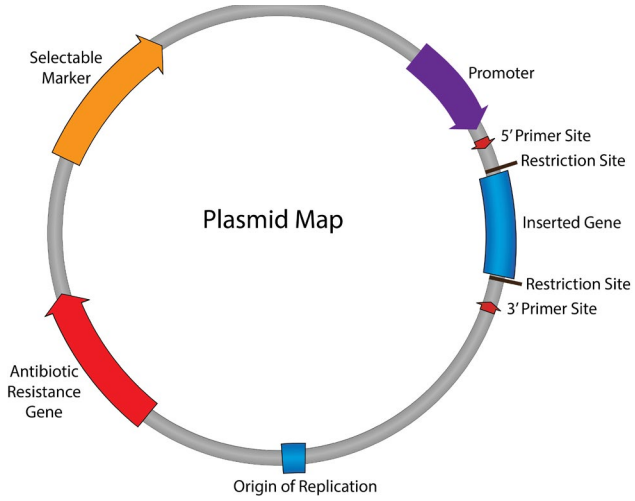
By Margo R. Monroe and Marcy Patrick | April 2020

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.

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, etc). The cloning method is ultimately chosen based on the plasmid you want to clone into. 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.



Importantly, because the bacteria from which plasmids are isolated grow quickly and make more of the plasmids as they grow, scientists can easily make large amounts of plasmid to manipulate and use in later work.

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 knockdown plasmids, reporter plasmids, viral plasmids, and genome engineering plasmids.

Some of the many things that plasmids can be used to do include:

- Produce large amounts of a protein so that scientists can purify and study it in a controlled setting
- Produce proteins that glow so that scientists can track their location or quantity inside a cell
- Monitor the level of a chemical in a particular environment
- Produce enzymes that will make specific, controlled changes to an organism's genome (genome engineering)
- Produce synthetic viruses that can be used in research or for therapeutics

Addgene has compiled various educational resources to facilitate plasmid use in the lab. Addgene's Molecular Biology Reference iincludes 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. ■

Table 1.1 – Plasmid Vector Elements

Vector Element	Description
Origin of Replication (ORI)	DNA sequence that allows initiation of replication within a plasmid by recruiting transcriptional machinery proteins.
Antibiotic Resistance	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.

Origin of Replication

By Kendall Morgan and Marcy Patrick | February 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? Ask yourself these questions:

- 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.

Table 1.2 – Origin of Replication

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. .
**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.

CHOOSING 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.

Table 1.2 highlights common cloning vectors, their copy number, ORI, and incompatibility group. Note the A-C compatibility grouping is an arbitrary designation, and plasmids from the same incompatibility group should not be co-transformed.

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 end A- *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!

.

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

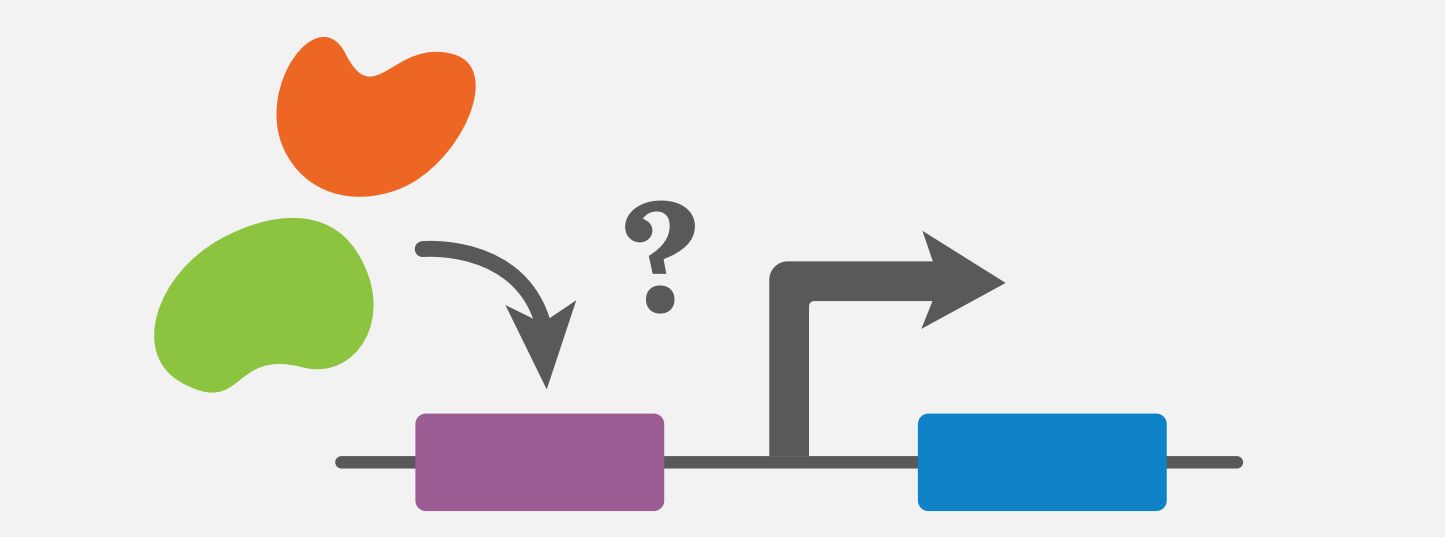
By Kendall Morgan, A. Max Juchheim, and Marcy Patrick | April 2014

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.

THE RNA POLYMERASES

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



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

promoters are commonly employed in lentiviral and retroviral constructs. These are covered in our viral vectors resources.

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.

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

Tables 1.3 and 1.4 list 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 1.3 - Eukarotyic 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	Constitutive	May include an enhancer
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 be-ta-globin splice acceptor.

Table 1.3 - Eukarotyic Promoters

Promoter	Primarily Used For	RNA Transcript	Description	Expression	Additional Considerations
TRE	General expression	mRNA	Tetracycline response element promoter	Inducible with Tetracyline 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
Polyhedrin	General expression	mRNA	Strong insect promoter from baculovirus	Constitutive	Commonly used in expression systems for insect cells
CaMKIIa	General expression	mRNA	Ca2+/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.

Table 1.3 - Eukarotyic Promoters

Promoter	Primarily Used For	RNA Transcript	Description	Expression	Additional Considerations
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-phosphage dehydrogenase	Specific	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.
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.

Table 1.4 - Prokaryotic Promoters

Promoter	Primarily Used For	Description	Expression	Additional Considerations
T7	In vitro transcription/ general expression	Promoter from T7 bacteriophage	Constitutive, but requires T7 RNA polymerase.	Can be used for in vitro 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	In vitro transcription/ general expression	Promoter from Sp6 bacteriophage	Constitutive, but requires SP6 RNA polymerase.	Can be used for in vitro transcription only if 2 different phage promoters are present in opposite orientations to gene.
trp	High levels of gene expression	Promoter from E. coli tryptophan operon	Repressible	Gets turned off with high levels of cellular tryptophan.
lac	General expression	Promoter from lac operon	Constitutive in the absense 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.
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.

Table 1.4 - Prokaryotic Promoters

Promoter	Primarily Used For	Description	Expression	Additional Considerations
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.

Terminators and Poly(a) Signals

By Julian Taylor-Parker | May 2016

Plasmids designed to express genes in a given host cell type are generally broken down into two broad categories, prokaryotic or eukaryotic, based on the functional elements they contain. Plasmid DNA in both prokaryotic and eukaryotic systems must be transcribed into RNA. Transcription occurs in three phases: initiation, elongation, and termination. We previously discussed the promoter's role in the initiation step of gene transcription; here we'll provide an overview on how transcription stops, or termination.

WHAT IS TERMINATION AND POLYADENYLATION?

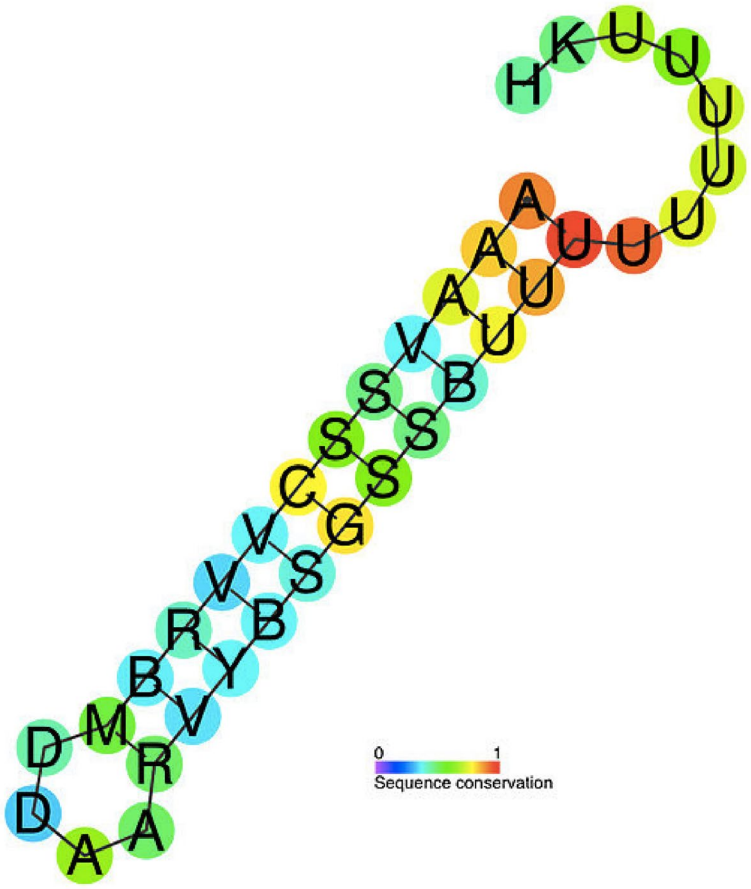
The role of the terminator, which is a sequence-based element, is to define the end of a transcriptional unit (such as a gene) and initiate the process of releasing the newly synthesized RNA from the transcription machinery. Terminators are found downstream of the gene to be transcribed, and typically occur directly after any 3' regulatory elements, such as the polyadenylation or poly(A) signal. While many studies focus on promoter strength as a determinant of gene expression levels, the terminator also plays an important role in RNA processing and contributes to variability in RNA half-life, and ultimately gene expression.

Polyadenylation, as the name implies, is the post-transcriptional additional of multiple adenine (A) nucleotides to the tail of a messenger RNA transcript. The purpose and mechanism of polyadenylation vary among cell types, but polyadenylation generally serves to promote transcript longevity in eukaryotes and promote transcript degradation in prokaryotes.

PROKARYOTIC TERMINATION

Prokaryotic termination mechanisms fall under two general categories: rho-dependent and rho-independent. Rho factor is a helicase which assists RNA polymerase in the termination of the transcript. Rho-dependent terminators are not usually employed in plasmid-based expression systems, so these will not be detailed here, but additional references are provided at the end.

Nearly all common **bacterial expression plasmids** use Rho-independent terminators, which include naturally occurring terminators, such as T7 and rrnB, as well as engineered high-efficiency terminators such as T0. Rho-independent termination is also known as intrinsic termination, and relies on the formation of a GC-rich hairpin in the RNA transcript followed by a weakly bound poly-uracil tract as shown in the figure to the right. The tertiary structure of the hairpin-DNA complex is thought to destabilize the transcription complex, initiating cleavage of the transcript.



A predicted conserved secondary structure and sequence conserved Rho-independent termination annotation for 90 bacterial elements. By Ppgardne at English Wikipedia, *CC BY-SA 3.0*

No terminator is 100% efficient at halting transcription of the template and initiating the desired cleavage event, although some engineered terminators come close (>95%). For most purposes, however, any common 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 multicistronic 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.

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 misfolded 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 polyadenlyation.

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 (Schek et. al., 1992; Gil et. al, 1987).

As alluded to above, termination and polyadenylation of Polymerase II transcripts (and therefore mRNAs) are 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 degredation, and is integral to the nuclear export and translation processes as well. ■

Further Reading

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.

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..

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.

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.

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.

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 Marcy Patrick | June 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?

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. It turns out that restriction enzymes are one half of naturally occuring 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 recognizes 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 protomoting 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

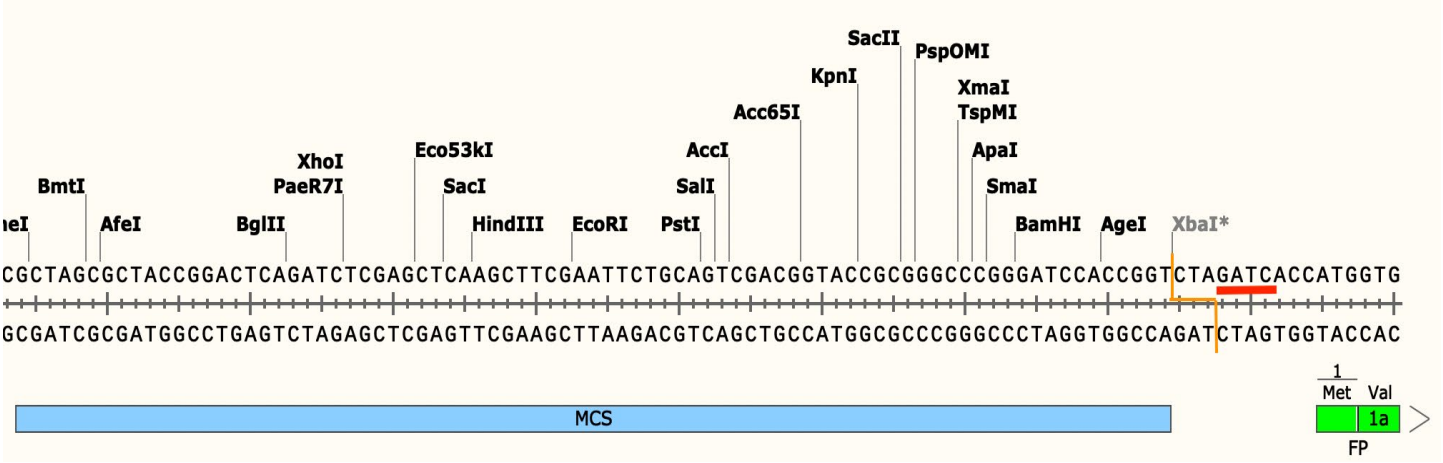


Figure 1

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 Figure 1, 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 preceeded 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 methyltransferases in the PCR reaction). The template plasmid, on the other hand, should be derived from a dam+ 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?

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 methyltransferases in the PCR reaction). The template plasmid, on the other hand, should be derived from a dam+ 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.

CONTROLLING METHYLATION

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.

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–/dcm– 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-/dcm- strains may have an increased rate of mutation (as these would also be deficient in mismatch repair functions of Dam), so these strains should not be used for long term storage. ■

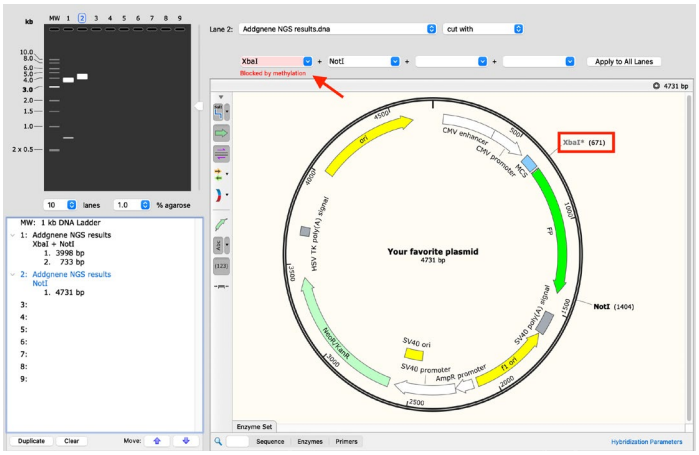


Image created using Snapgene(R).

Further Reading

Review of Restriction Modification Systems

NEB's Resource Describing Dam and Dcm Methylases

Discovery of DpnI

Review describing dam-/dcm- E. coli

Table 1.5 – Dam Methylation Enzymes

Enzyme	Dam methylation	Dcm methylation	EcoKI methylation
ApaI	Not affected	Blocked by overlapping methylation	Not affected
BsaI	Not affected	Blocked by overlapping methylation	Not affected
ClaI	Blocked by overlapping methylation	Not affected	Not affected
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
MboI	Not affected	Blocked by overlapping methylatio	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

CHAPTER 2

Bacteria

E. Coli Strains for Protein Expression

By Julian Taylor-Parker | February 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?

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 2.1).

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 Δlon) similarly reduce proteolysis of the expressed proteins.
- **hsdSB (rB- mB-):** These strains have an inactivated native restriction/methylation system. This means the strain can neither restrict nor methylate DNA.

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,"

Table 2.1 – *E. coli* Expression Strains

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

Table 2.1 – *E. coli* Expression Strains

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	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	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo (CamR)	Expression of toxic, insoluble, or membrane proteins
T7	Chloramphenicol (pRARE)	IPTG-inducible expression of T7 RNAP from the genome; does not restrict methylated DNA	fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10-	General protein expression
m15** pREP4*	Kanamycin (pREP4)	Cis-repression of the *E. coli* T5 promoter (found on vectors such as pQE or similar) inducible under IPTG (lac repressor on the pREP4 plasmid). Expression vector cannot have p15A origin of replication	F-, Φ80ΔlacM15, thi,	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	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL F'[lac+ lacIq pro] (DE3) gor522::Tn10 trxB (StrR, TetR)	Expression of insoluble proteins

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

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 timeframe. 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.

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. ■

Common Lab *E. coli* Strains

By Matthew Ferenc | November 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 2.2 outlines common lab strains of *E. coli*. ■

Table 2.2 – Common Gene Mutations Found in E. coli Strains

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

Table 2.2 – Common Gene Mutations Found in E. coli Strains

Genes	Recognition Site	Functional Consequence
mrr, Δ(mcrC-mrr)	Mutation in galactose metabolism pathway	Cells cannot grow on just galactose
mrr, Δ(mcrC-mrr)	Inactivation of pathway that cleaves methylated adenine or cytosine DNA	Allows for uptake of foreign (methylated) DNA
recA, recA1, re-cA13-mrr)	Mutation in a DNA-dependent ATPase that is essential for recombination and general DNA repair	Reduces plasmid recombination, increases plasmid stability
recBCD	Exonuclease V activity abolished	Increased plasmid stability, reduced recombination
relA or relA1	Relaxed phenotype, mutation eliminating stringent factor	Allows RNA synthesis in absence of protein synthesis
recBCD	Exonuclease V activity abolished	Increased plasmid stability, reduced recombination
ptrc-ccdA	N/A	ptrc-ccdA
hte	N/A	high transformation efficiency

Genes	Recognition Site	Functional Consequence
deoR	constitutive expression of genes for deoxyribose synthesis	Allows uptake of large plasmids
hee	N/A	High electroporation efficiency
supE44 (glnV44)	N/A	Suppression of the amber (UAG) stop codon by inserting glutamine
supF (tyrT)	N/A	Suppression of the amber (UAG) stop codon by inserting tyrosine
λ-thi-1 or thi1	Mutation in thiamine metabolism	Requires exogenous thiamine for growth
ara	Disruption of arabinose metabolism pathway	Inability to utilize arabinose as a carbon source
leuB	β-isopropyl malate dehydrogenase, inactivated	Requires exogenous leucine source for growth
proAB	mutation in proline biosynthesis pathway	Requires exogenous proline source for growth
rpsL	Mutation in subunit S12 of 30S ribosome	Confers resistance to streptomycin

Environmental Plasmids

By Jessica Welch | January 2017

Here at Addgene, we often refer to plasmids as lab or experimental tools. They certainly are very handy in research, but where did these tools come from and why do they exist in nature? In this section, we'll talk about environmental plasmids and how they've helped us develop molecular biology tools for the lab.

NATURAL PLASMIDS

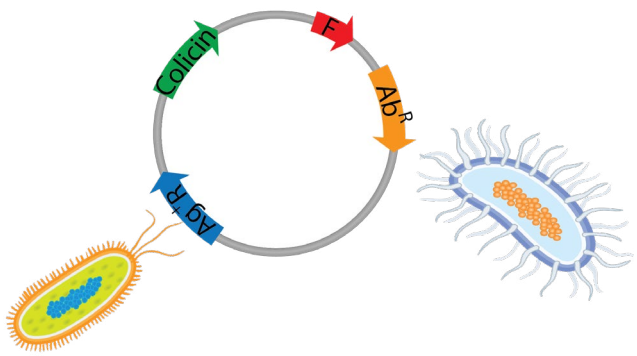
Plasmids found in nature often give their hosts beneficial traits that allow them to survive in competitive environments. Plasmids derived directly from the environment are sometimes called 'natural' plasmids, to distinguish them from the modified versions we usually work with in the lab. Horizontal (or 'lateral') gene transfer occurs when genetic material is passed between organisms by mechanisms other than reproduction (vertical gene transfer). Small, mobilizable genetic elements like natural plasmids enable horizontal gene transfer, allowing beneficial traits encoded by these plasmids to spread rapidly throughout the environment. Indeed, plasmid transfer between organisms is one of the most common methods of horizontal gene transfer in nature. These 'extra' pieces of DNA can be extremely useful for survival in the crowded and competitive natural environments outside the petri dish.

Although we usually think of bacteria when we think of plasmids, both Archaea and Eukaryota can also harbor these circular, extrachromosomal, double-stranded DNA passengers. For example, a variety of high-copy number yeast vectors have been derived from the natural yeast 2-micron plasmid.

PLASMID TRANSFER OUTSIDE THE LAB

In the lab, we use chemical transformation, electroporation, and transfection to put plasmids into cells. In nature, the two main routes of plasmid transfer are conjugation and natural transformation. Conjugation occurs when a sex pilus extends between two bacteria, allowing a copy of plasmid DNA to travel from one cell to another.

DNA can also be passed between bacteria by bacteriophages, the viruses that infect bacteria. Bacteriophage genomes can exist in bacterial cells as extrachromosomal plasmids known as



prophages, or they can integrate into the host chromosome using DNA recombination systems. Some of these systems have been modified for use in genome engineering, including the Cre-lox system from the P1 phage of *E. coli*.

Fun fact: There is evidence that lightning can also promote gene transfer in soil bacteria!

BENEFITS OF NATURAL PLASMIDS

What actually constitutes a "beneficial trait" encoded by a plasmid? In addition to the genes required for transfer and integration, natural plasmids can also convey a variety of functions that allow their host organisms to thrive in competitive environments. The genes encoding these functions form the basis of many molecular biology.

RESISTANCE GENES AND MICROBIAL WARFARE

One way to triumph over your neighbors is to eliminate the competition via the production of toxins and antibiotics. Fungi such as Penicillium produce bacteria-zapping compounds like penicillin, and bacteria in turn possess plasmid-borne toxins like colicins and subtilisins to kill other bacteria in the area. Of course, the next logical step in this microbial arms race is to develop resistance to these chemical weapons.

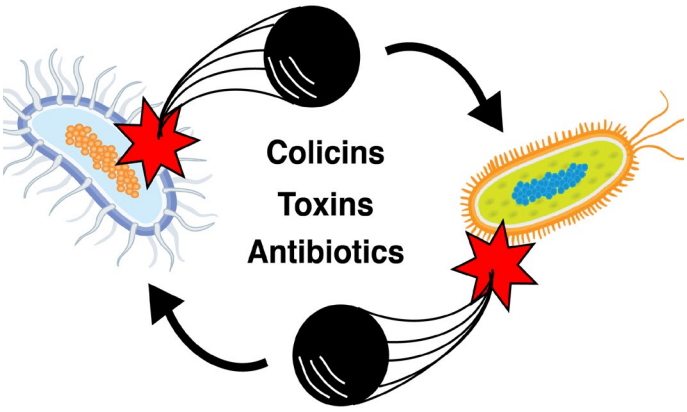
Conveniently, plasmids carrying genes for the production of bacterial toxins like colicin also often carry the necessary resistance genes so that the producer is immune. Resistance to antibiotic compounds produced by other species (like those pesky Penicillium) can be passed around on plasmids like a "get out of jail free card." These plasmids were historically called R (resistance) plasmids. While antibiotic resistance-mediated selection is indispensable for cloning work, the transfer of plasmids carrying these genes between bacteria can also have serious implications for public health.

DEGRADATIVE ENZYMES

There are less aggressive ways to survive in competitive environments, however, such as taking up an unoccupied niche.

Some organisms manage this by utilizing recalcitrant or unusual substrates that other organisms can't use. Others develop resistance to toxic substances such as heavy metals or pesticides, enabling them to survive where others can't.

To conserve energy, these specialized degradative capabilities are often tightly regulated to be active only in the presence of the target compound. These systems are a gold mine of parts for



synthetic biology, especially when combined with natural inducible expression systems that allow tuning of metabolic pathways.

VIRULENCE PLASMIDS

Similarly, conditionally-expressed plasmid genes can give bacteria the ability to invade environments found on and in living organisms. For example, there are plasmids that allow Salmonella and *E. coli* to infect the human intestine, and other plasmids that allow Agrobacteria to infect plants. These are known as virulence plasmids, as they help make an otherwise harmless bacterium pathogenic. While these plasmids can be dangerous to the health of the host organism in a natural setting, they also provide researchers with new means to deliver DNA and other molecules to host organisms. For instance, vectors derived from the Agrobacterium Ti virulence plasmid are now used routinely in biotechnology to transform plant cells.

We're only just breaking the surface of the environmental plasmid world in this post. It is the great diversity of natural plasmids and their functions that has enabled researchers to develop and deposit the many plasmid tools available in the Addgene repository. To learn more about some other common plasmid features such as origin of replication and promoters, keep reading our Plasmids 101 e-book and keep an eye on the Addgene blog as scientists continue to mine environmental plasmids for new and exciting molecular technologies! ■

REFERENCES

1. Demaneche, S., et al., Laboratory-scale evidence for lightning-mediated gene transfer in soil. Appl Environ Mi- crobiol (2001) 67: 3440-3444. PubMed PMID: 11472916. PubMed Central PMCID: PMC93040

Transformation, Transduction, Conjugation, and Transfection

By Alyssa Cecchetelli | June 2019

Horizontal gene transfer (HGT) is the movement of genetic material between organisms. It plays a key role in bacterial evolution and is the primary mechanism by which bacteria have gained antibiotic resistance and virulence. Scientists have studied how HGT occurs in nature and have learned how to introduce genetic materials into cells in the lab.

The introduction of foreign DNA or RNA into bacteria or eukaryotic cells is a common technique in molecular biology and scientific research. There are multiple ways foreign DNA can be introduced into cells including transformation, transduction, conjugation, and transfection. Transformation, transduction, and conjugation occur in nature as forms of HGT, but transfection is unique to the lab.

TRANSFORMATION

Various factors promote natural transformation in different bacteria such as the growth phase of the cells (Baltrus and Guillemin, 2006) or the presence of specific substances (Meibom et al., 2005).

Though not all bacteria are naturally competent to take up DNA, they can be made competent through chemical manipulation in the lab. This is commonly done using calcium chloride, which permeabilizes the cell membrane so the bacteria can easily uptake your plasmid of interest. Scientists can also use electroporation, the application of an electrical charge to cells, to increase cell membrane permeability and thus transformation efficiency. Check out Addgene's blog to learn about making your own competent cells and our protocols page to learn about bacterial transformation in the lab.

TRANSDUCTION

Transduction occurs when foreign DNA or RNA is introduced into bacterial or eukaryotic cells via a virus or viral vector. One example are bacteriophages that attach to bacterial membranes and inject their genetic material into the cell. Once inside, phages can follow one of two different life cycles: lytic or lysogenic. Lytic phages hijack the machinery of their bacterial hosts to make more viral particles. Eventually the cell lyses releasing the newly formed viral particles that can infect other bacteria. In the lysogenic cycle, the phage's genetic material is incorporated into the host's genome at a particular integration site. The integrated phage remains dormant until it is triggered to enter the lytic cycle.

During both of these life cycles, bacterial DNA can be accidentally packaged into the newly created phages. Transfer of this DNA to another cell is referred to as transduction. Once inside

the infected bacterium, transferred DNA can either exist as transient extrachromosomal DNA, like a plasmid, or it can integrate into the host bacterium's genome through homologous or site-directed recombination.

Transduction is a common tool used by scientists to introduce different DNA sequences of interest into a bacterial cell or a host's genome. To do this scientists commonly use phagemids, a DNA cloning vector that contains both bacteriophage and plasmid properties. The **phagemids** are packaged into replication-incompetent phage particles with assistance from a 'helper' phage prior to transduction.

Scientists also use transduction to introduce foreign DNA into eukaryotic cells, like mammalian cell lines. This can be done with **lentivirus** and **Adeno-Associated Viruses (AAV)**. They can be used to create both transient cell lines, where a gene of interest is expressed but not integrated into the genome, and **stable cell lines**, where foreign DNA is incorporated into the cell's genome and is thus passed down through cell division. You can find all kinds of different lentiviral and AAV plasmids as well as **ready-to-use viral preparations** at Addgene. For more information on viral vectors, including transduction, download our Viral Vectors 101 eBook.

BACTERIAL CONJUGATION

Conjugation was the first extensively studied method of gene transfer. It was discovered in 1946 by Joshua Lederberg and Edward Tatum when they observed genetic recombination between two nutritional deficient *E. coli* strains that resulted in a wild type *E. coli* (Griffiths et al., 2000).

During conjugation, genetic material is transferred from a donor bacterium to a recipient bacterium through direct contact. The donor bacterium contains a DNA sequence called the Fertility factor (F-factor). The F-factor is found on an episome, a piece of DNA that can replicate on its own or be integrated within a bacterial chromosome and allows the donor bacterium to make a small "bridge" or sex pilus that attaches to the recipient cell drawing it close. Once in contact the donor can transfer genetic material to the recipient bacterium. The genetic material transferred is commonly a plasmid and can infer genetic advantages such as antibiotic resistance.

TRANSFECTION

Unlike the last three methods which can be used in prokaryotes, transfection is only done in eukaryotic cells. Transfection is the process by which foreign DNA is deliberately introduced into a eukaryotic cell through non- viral methods including both chemical and physical methods in the lab. Chemicals like calcium phosphate and diethylaminoethyl (DEAE)-dextra neutralize or even impart an overall positive charge on DNA molecules so that it can more easily cross the negatively charged cell membrane. Physical methods such as electroporation or microinjection actually pokes holes in the cell membrane so DNA can be introduced directly into the cell.

Microinjection requires the use of a fine needle to deliver nucleic acids to individual cells. Electroporation, on the other hand, uses electrical pulses to create transient pores in the cell membrane that genetic material can pass through.

If you are starting any molecular biology experiment check out Addgene's **protocol page**, which has both protocols and videos for techniques in basic molecular biology, plasmid cloning, and viruses. ■

REFERENCES

Baltrus, David A., and Karen Guillemin. "Multiple phases of competence occur during the Helicobacter pylori growth cycle." FEMS microbiology letters 255.1 (2006): 148-155. PubMed PMID: 16436074.

Griffiths AJF, Miller JH, Suzuki DT, et al. Bacterial Conjugation. In An Introduction to Genetic Analysis. 7th edition (2000).

Meibom, Karin L., et al. "Chitin induces natural competence in Vibrio cholerae." Science 310.5755 (2005): 1824- 1827. PubMed PMID: 16357262.

CHAPTER 3

Selection & Screening

Antibiotic Resistance Genes

By Marcy Patrick | January 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! 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.

TIPS AND TRICKS FROM THE BENCH:

- Use fresh stocks. Most antibiotics are stable in powder form, but quickly break down in solution. 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**.

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**. ■

Table 3.1 – Common Antibiotics

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

**In prokaryotes. **Dissolve in dH2O and sterile filter unless otherwise specified*

Blue-White Screening

By Jessica Welch | June 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!

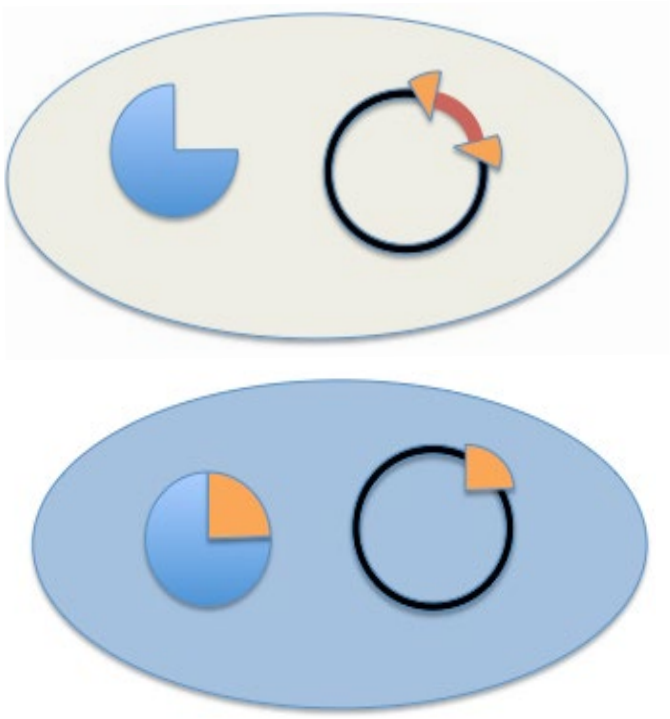
Let's begin at the beginning. The well-characterized bacterial lac operon contains a gene called lacZ that encodes for the enzyme β-galactosidase. Expression of the lac operon is induced by lactose, and also by a lactose analogue, IPTG (isopropyl β-D-1-thiogalactopyranoside). To be completely accurate,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

When expressed, the β-galactosidase enzyme can break down a dye-linked substrate called x-gal (5-bromo-4- chloro-3-indolyl-β-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 lacZΔM15) creates a non- functional β-galactosidase enzyme.

Providing DNA encoding this section of amino acids (called the α-peptide) to a lacZΔM15-mutant bacterial cell in trans complements the mutation allowing for a functional enzyme. This process is called α-complementation.



Blue-White screening. A) A white cell in which cloning was successful and the β-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 β-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.

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

Colonies with an insert-containing plasmid have a non-functional β-galactosidase, and remain the lovely whitish-cream color of standard *E. coli*. On the other hand, intact β-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 β-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 α-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 β-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 lacZΔM15 mutation)
XL1-Blue, DH5α, DH10B, JM109, STBL4, JM110, and Top10 are a few examples.

Make sure you use a proper plasmid (i.e. contains the α-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 α-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).

OTHER TYPES OF SCREENING METHOD

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 α-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.

OTHER TYPES OF SCREENING METHOD

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. ■

MORE RESOURCES

Addgene's curated list of Bacterial Expression Systems.

Positive and Negative Selection

By Jennifer Tsang | August 2019

You've worked hard to purify your gene of interest, get it into your plasmid backbone, and zap the mixture of DNA into cells. Unfortunately, not every cell successfully takes up plasmid DNA. Among those that do, some now have plasmids that contain your gene of interest, but others will uptake plasmid backbones that re-ligated back on themselves.

Therefore, your cloning strategy needs to identify cells containing the plasmid construct you're seeking. Fortunately, there are many ways to do this involving positive selection, negative selection, and/or screening. Here, we're focusing on positive and negative selection.

POSITIVE SELECTION: CELLS THAT HAVE GAINED A SPECIFIC GENE SURVIVE

In positive selection, only cells that contain a specific gene survive. The rest do not grow. While there are many ways to design your positive selection strategy, here are a few to consider.

Antibiotic selection

Selection is straightforward in cases where the inserted DNA includes an antibiotic resistance marker. After transformation, plate the cells on medium containing the corresponding antibiotic. For example, if the DNA you're inserting into the plasmid backbone contains a kanamycin resistance cassette, you would plate the cells on kanamycin-containing agar. Only cells carrying plasmids that contain a kanamycin resistance cassette will grow.

For example, if the DNA you're inserting into the plasmid backbone contains a kanamycin resistance cassette, you would plate the cells on kanamycin-containing agar. In these cases, antibiotic selection will select for cells that have been transformed with the plasmid backbone (with or without your inserted DNA) and further screening will help you determine if the plasmid backbone contains your gene of interest.

Auxotrophy

Besides antibiotic selection, you can select for a successful uptake of your DNA of interest by taking advantage of bacterial characteristics such as prototrophy, the ability to synthetize all compounds needed for growth, or auxotrophy, the inability to synthesize a compound needed to grow.

Auxotrophy and prototrophy come into play during cloning if you've designed your experiment using an auxotrophic strain of bacteria (one that doesn't synthesize an essential amino acid), and a plasmid that complements the auxotrophy (i.e., synthesizes the critical amino acid).).

Many common *E. coli* strains used for cloning are auxotrophic for specific amino acids. When plated on media lacking the essential amino acid, the bacteria will only grow if it acquired the plasmid containing the gene necessary to synthesize the missing amino acid.

Auxotrophic selection is often used for cloning in yeast since antibiotic resistance spontaneously arises in yeast. For example, Markus Ralser's lab created a yeast prototrophy kit containing 23 plasmids to complement auxotrophy.

NEGATIVE SELECTION: CELLS THAT HAVE LOST A SPECIFIC GENE SURVIVE

Unlike positive selection, negative selection means you're selecting for the loss of a gene product—usually something toxic. This gene is found on the original plasmid and either the insertion of a DNA fragment within the gene or loss of the gene alleviates its toxic effect. Negative selection is sometimes called counterselection.

Toxin-antitoxin systems

One negative selection strategy uses the toxin ccdB. Gateway cloning takes advantage of this property by including ccdB on its destination plasmids. In the destination plasmid, the ccdB gene is flanked by recombination sites. To introduce a gene of interest into the destination plasmid, recombination sites must first be added to flank the gene. Successful recombination and integration of the gene of interest at these sites in the plasmid excises the *ccdB* gene and the cells will only grow if they contain the gene of interest.

To introduce a gene of interest into the destination plasmid, recombination sites must first be added to flank the gene. Successful recombination and integration of the gene of interest at these sites in the plasmid excises the *ccdB* gene and the cells will only grow if they contain the gene of interest.

SacB counter-selection

SacB counter-selection relies on the toxic product produced by the *SacB* gene. SacB comes from the gram-positive bacteria *Bacillus subtilis* and encodes the enzyme levansucrase that converts sucrose into a toxic metabolite in gram-negative bacteria. Plating on sucrose medium will select for cells that contain constructs that have lost the sacB gene.

SacB can be used similarly to ccdB.. For example, in the gene expression plasmid pGTvL1-SGC from Nicola Burgess-Brown's lab, the sacB gene is used for negative selection on 5% sucrose plates. Exchange of the sacB gene for the gene of interest allows the bacteria to grow on sucrose.

Plasmid cloning: Creativity with naturally occurring genes

At this point in the blog post, you may have realized that positive and negative selection strategies adopt naturally occurring genes for cloning purposes (antibiotic resistance genes, toxin genes, and the sacB gene). By tapping into nature's toolbox, scientists are finding more and more ways to creatively perform selections in plasmid cloning. ■

Colony PCR

By Beth Kenkel | May 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 blog post discusses some of the key things to consider when performing colony PCR.

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.

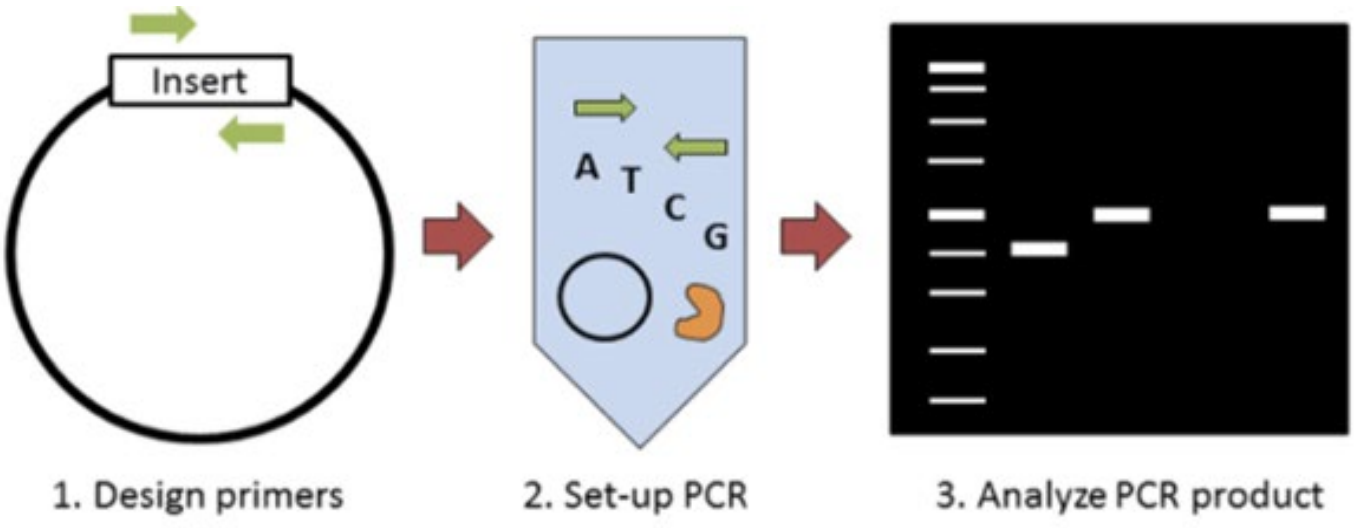


Figure 1. The first and perhaps most important step to colony PCR is designing primers. There are 3 strategies for primer design: 1) insert-specific primers, 2) backbone-specific primers, and 3) orientation-specific primers.

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 then 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.

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.

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.

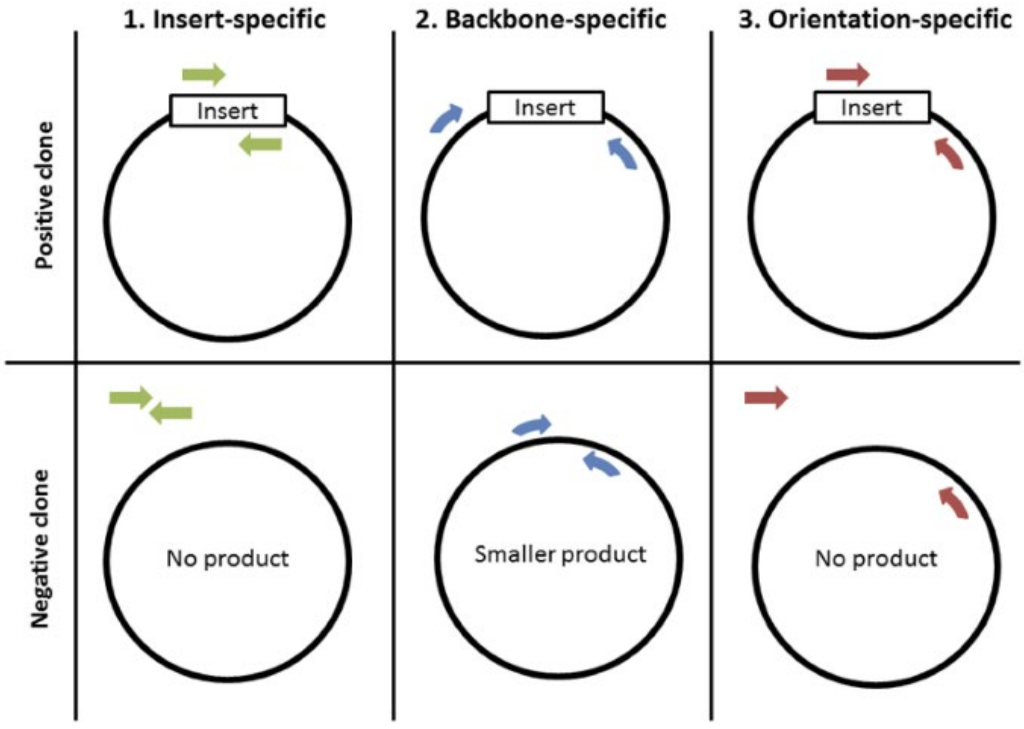


Figure 2

Table 3.2 - Primer Design Options

Primer Design	Pros	Cons
Insert Specific	• Simple "yes or no" test result	• Insert orientation not determined • Requires new primer pair for each insert
Backbone Specific	• Provides information about the size of the insert and if it's in the plasmid • Can be used to screen clones that were created with the same backbone	• Insert orientation not determined
Orientation Specific	• Tells you the orientation of the insert • Tells you if the insert is in the plasmid • Great for blunt end cloning	• Requires new primer pair for each

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.

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.

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-prep the positive ones. Regardless of which method you choose, make sure to use the appropriate antibiotic for selection.

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.

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

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. Figure 3 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.

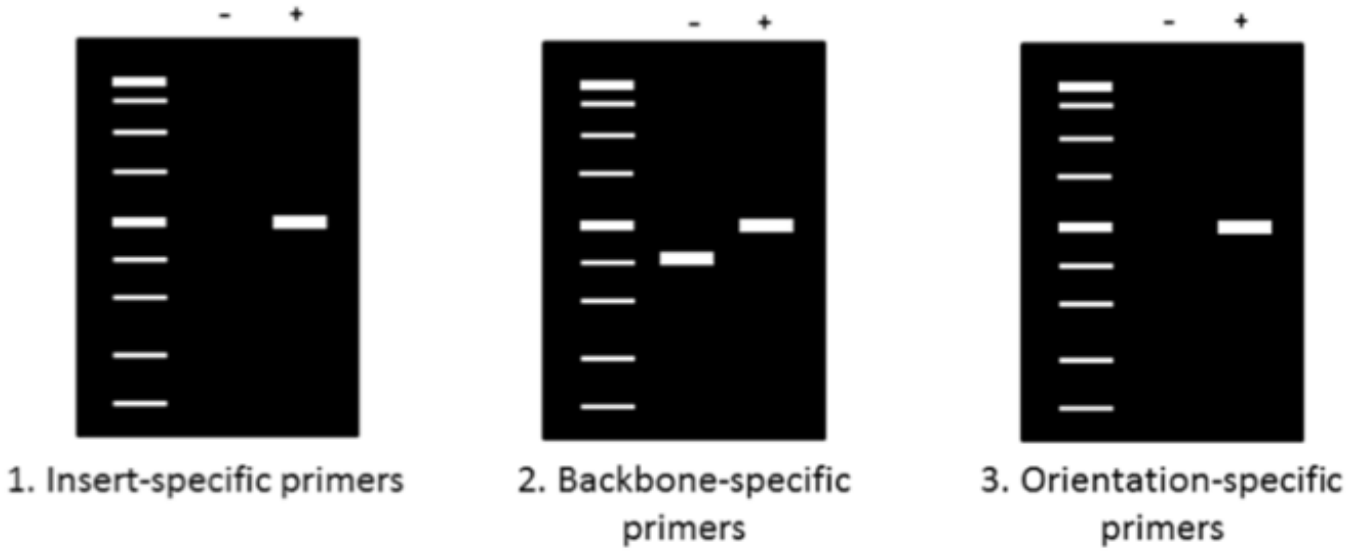


Figure 3 summarizes generalized expected results for the three primers previously described.

VERIFYING THE INSERT SEQUENCE WITH SANGER SEQUENCING

After identifying a few positive clones, the last step is to mini-prep 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

Promega Overview of
Screening for Recombinants

New England Biolabs
Overview of Colony PCR

CHAPTER 4

Controlling Expression

Protein Expression

By Alyssa Cecchetelli | June 2018

The central dogma in molecular biology is DNA>RNA>Protein. To synthesize a particular protein, DNA must first be transcribed into messenger RNA (mRNA). mRNA can then be translated at the ribosome into polypeptide chains that make up the primary structure of proteins. Most proteins are then modified via an array of post-translational modifications including protein folding, formation of disulfide bridges, glycosylation, and acetylation to create functional, stable proteins. Protein expression refers to the second step of this process: the synthesis of proteins from mRNA and the addition of post-translational modifications.

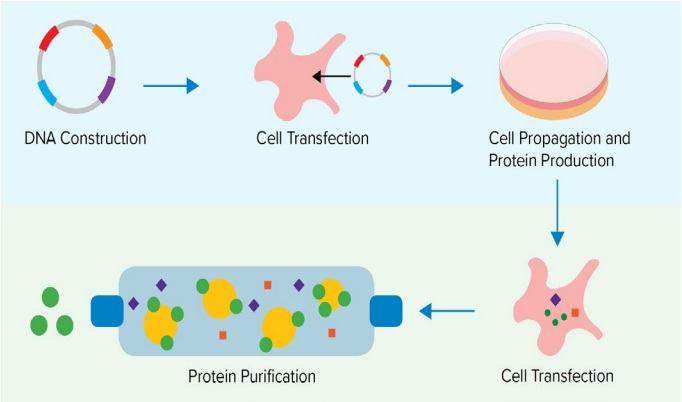
Researchers use various techniques to control protein expression for experimental, biotechnological, and medical applications. Researchers can visualize proteins in vivo by tagging them with **fluorescent proteins** to study localization or purify proteins to study their structures, interactions, and functions. Proteins can also be

purified for use in molecular biology research (eg. polymerases and other enzymes might be purified and used to manipulate DNA), or in medicine (e.g. insulin).

Proteins, unlike DNA, must be produced using complex mixtures derived from cells or using live cells. There are several types of expression systems used for protein production and purification. These include mammalian, insect, bacterial, plant, yeast, and cell free expression systems.

The general strategy for protein expression consists of transfecting cells with your DNA template of choice and allowing these cells to transcribe, translate, and modify your protein of interest. Modified proteins can then be extracted from lysed cells through the use of protein tags and separated from contaminants using a variety of purification methods. Deciding which expression system to use depends on several factors:

(1) The protein you are trying to express
(2) How much protein you need
(3) Your plans for downstream applications



Here, we will summarize some of the more common expression systems, including their advantages and caveats to keep in mind before choosing a system.

MAMMALIAN EXPRESSION SYSTEMS

Mammalian cells are an ideal system for the expression of mammalian proteins that require multiple post- translation modifications for proper protein function. Most DNA constructs designed for mammalian expression utilize viral promoters (SV40, CMV, and RSV) for robust expression post-transfection. Mammalian systems can express proteins both transiently or through stable cell lines. Both methods produce high protein yields if transfection is successful.

Some mammalian systems also allow for control over when a protein is expressed through the use of constitutive and inducible promoters. Inducible promoters are extremely useful if a desired protein product is toxic to cells at high concentrations. Despite their advantages, mammalian expression systems do require demanding cell culture conditions compared to other systems.

INSECT EXPRESSION SYSTEMS

Insect cells can also be used to produce complex eukaryotic proteins with the correct post-translational modifications. There are two types of insect expression systems: baculovirus infected and non-lytic insect cells.

Baculovirus expression systems are very powerful for high level recombinant protein expression. These systems enable high expression of very complex, glycosylated proteins that cannot be produced in *E. coli* or yeast cells. However, in baculovirus systems is that the infected host cell is eventually lysed. Cell lysis halts protein production, but there are non-lytic insect cell expression systems (sf9, Sf21, Hi-5 cells) that allow for continuous expression of genes integrated into the insect cell genome. Both of these types of insect expression systems can be scaled up for production of large amounts of protein.

Keep in mind that virus production in insect cells can be quite time consuming and they also require demanding culture conditions similar to mammalian expression systems.

BACTERIAL EXPRESSION SYSTEMS

When one wants to produce vast quantities of protein rapidly and cheaply, a bacterial host cell is almost always the answer. E. coli is definitely one of the most popular hosts for protein expression with several strains that are specialized for protein expression. Protein expression in bacteria is quite simple. DNA coding for your protein of interest is inserted into a plasmid expression vector that is then **transformed into a bacterial cell**. Transformed cells propagate, are induced to produce your protein of interest, and then lysed. Protein can then be purified from the cellular debris.

There are several popular DNA vectors that can be used to produce large amounts of protein in bacterial cells, like the pET, pRSET, Gateway pDEST, and pBAD vectors. Protein expression from each of these vectors is controlled by a different promoter resulting in different levels of expression from each vector; lower expression may be required if your protein is toxic to E. coli. Of all the vectors, pET, under the control of the T7 lac promoter and induced by lactose, provides the highest level of protein expression.

Despite their ease of use, it is important to note that bacteria usually cannot produce functional multi-domain mammalian proteins as bacterial cells are not equipped to add appropriate post-translational modifications.

In addition, many proteins produced by bacteria become insoluble, forming inclusion bodies that are difficult to extract without harsh reagents and patience.

PLANT EXPRESSION SYSTEMS

Plants provide a cheap and low-tech means of mass expression of recombinant proteins. Many cells from various types of plants such as maize, tobacco, rice, sugarcane, and even tubers of potatoes have been used for protein expression.

Plant systems share many of the same features and processing requirements as mammalian cell expression systems, including the majority

of complex post-translational modifications. Extraction and purification of recombinant proteins from plants however can be expensive and time consuming as plant tissues themselves are biochemically complex.

To circumvent these issues, scientists have taken advantage of the natural secretion of biochemicals and proteins through plant roots. Tagging recombinant proteins with a naturally secreted plant peptide allows for easier access and purification of a desired protein. Despite being a rather nascent technology, plant cells have been used to express a wide range of proteins including antibodies and pharmaceuticals, specifically interleukins.

YEAST EXPRESSION SYSTEMS

Yeast are a great expression system to generate large quantities of recombinant eukaryotic proteins. Although many species of yeast can be used for protein expression, S. cerevisiae, is the most reliable and frequently used species due to its use as a model organism in genetics and biochemistry.

When using *S. cerevisiae*, researchers often place recombinant proteins under the control of the galactose inducible promoter (GAL). Other commonly used promoters include the phosphate and copper inducible PHO5 and CUP1 promoters, respectively. Yeast cells are grown in well-defined media, though optimization is still often needed, and can be easily adapted to fermentation, allowing for large-scale, stable production of proteins.

Yeast expression systems are generally easier and cheaper to work with than mammalian cells, and are often capable of modifying complex proteins, unlike bacterial systems. Yeast cells, however, have a slower growth rate than bacterial cells and are also known for hyper glycosylating proteins, which may be an issue depending on your protein of choice.

CELL-FREE EXPRESSION SYSTEMS

In cell-free expression systems, proteins are assembled in vitro using purified components of the transcription and translation machinery. These include ribosomes, RNA polymerase, tRNAs, ribonucleotides, and amino acids. Cell-free expression systems are ideal for fast assembly of more than one protein in one reaction. A major advantage of these systems is their ability to assemble proteins with labeled or modified amino acids that are useful in different downstream applications. Cell-free expression systems ,however, are expensive and very technically challenging to use. ■

Inducible Promoters

By Mary Gearing | January 2018

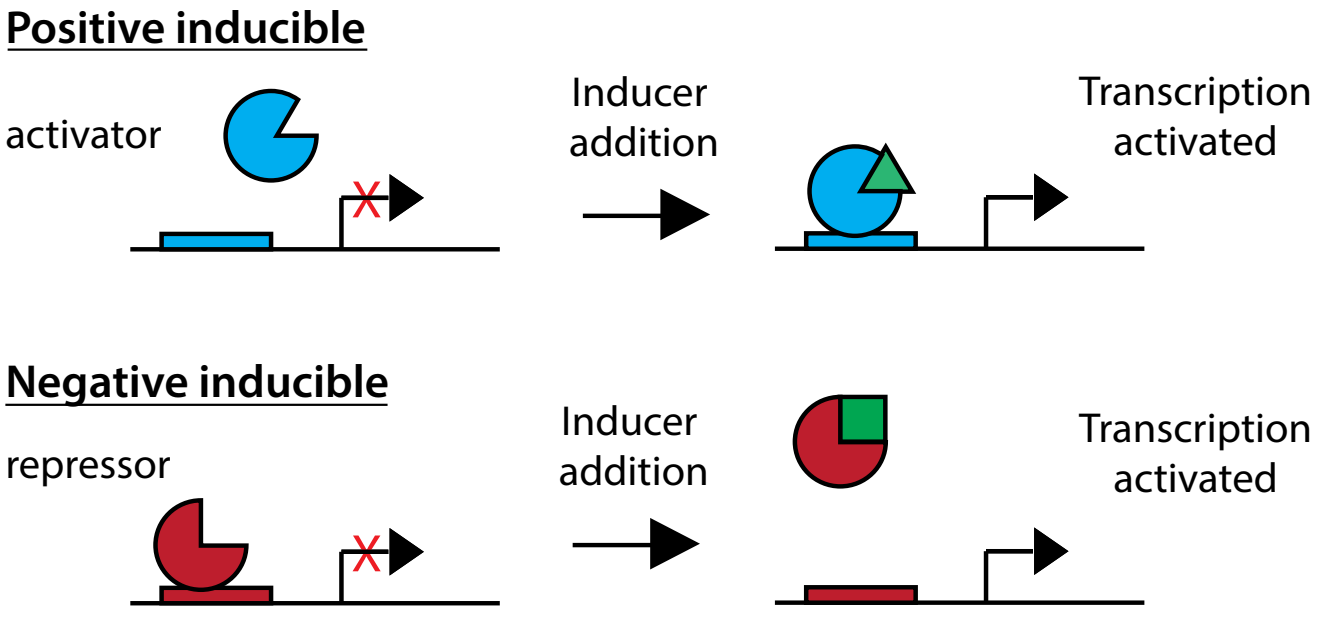
Promoters control the binding of RNA polymerase and transcription factors. Since the promoter region drives transcription of a target gene, it therefore determines the timing of gene expression and largely defines the amount of recombinant protein that will be produced. Many common promoters. like CMV, EF1A, and SV40 promoters, are always active and thus referred to as constitutive promoters. Others are only active under specific circumstances. Here, we'll discuss inducible promoters, which can be switched from an OFF to an ON state, and how you might use these in your research.

HOW ARE INDUCIBLE PROMOTERS REGULATED?

Inducible promoters can be regulated by positive or negative control.

Positive inducible

In the OFF state, the promoter is inactive because the activator protein cannot bind. After an inducer binds to the activator protein, the activator protein can bind to the promoter, turning it ON and initiating transcription.



Negative inducible

In the OFF state, the promoter is inactive because a bound repressor protein actively prevents transcription. Once an inducer binds the repressor protein, the repressor protein is removed from the DNA. With the repressor protein absent, transcription is turned ON.

TYPES OF INDUCIBLE PROMOTERS

Chemical agents, temperature, and light are all examples of factors that can lead to the induction of a promoter. Below, you'll find a short description of these three types of inducible promoters, and examples of each type. Many of these promoter systems are available at Addgene!

Chemically inducible promoters

Chemically regulated promoters are among the most common inducible promoters. The positive inducible tetracycline ON (**Tet-On**) system, a versatile tool developed for use in prokaryotes and eukaryotes, works via direct activation. In this system, the activator rtTA (reverse tetracycline-controlled transactivator) is normally inactive and cannot bind the tetracycline response elements (TRE) in a promoter. Tetracycline and its derivatives serve as inducing agents to allow promoter activation.

The cumate system, developed by Massie et. al, is functionally almost identical to the TET system and is available as an inducible or repressible promoter. It functions well in mammalian cells and has been shown to work in other organisms such as sphingomonads and bacillus.

One of the most commonly used prokaryotic promoters is the negative inducible pLac promoter. This promoter requires removal of the lac repressor (lacI protein) for transcription to be activated. In the presence of lactose or lactose analog IPTG, the lac repressor undergoes a conformational change that removes it from lacO sites within the promoter and ceases repression of the target gene. A simplified lac inducible system is found in many bacterial expression vectors.

Negative inducible promoter pBad is another popular prokaryotic promoter often used for bacterial protein purification. When arabinose is absent, regulatory protein AraC binds O and I1 sites upstream of pBad, blocking transcription. The addition of arabinose causes AraC to bind I1 and I2 sites, allowing transcription to begin. In addition to arabinose, cAMP complexed with cAMP activator protein (CAP) can also stimulate AraC binding to I1 and I2 sites. Supplementing cell growth media with glucose decreases cAMP and represses pBad, decreasing promoter leakiness.
Other examples of chemically induced promoters include positive inducible alcohol and steroid regulated promoters commonly used in plant research.

Temperature inducible promoters

Temperature sensitive expression systems are typically less leaky than chemically induced

promoters; they show near-zero expression at regular temperatures but can be induced by heat or cold exposure. Examples include the heat shock-inducible Hsp70 or Hsp90-derived promoters, in which a gene of choice is only expressed following exposure to a brief heat shock. In the case of Hsp70, the heat shock releases heat shock factor 1 (HSF-1), which subsequently binds to heat shock elements in the promoter, thereby activating transcription. Addgene depositors have developed heat shock-inducible Cre and Cas9 for easy genome engineering in species like C. elegans and Drosophila.

Light inducible promoters

Light is another way to activate gene expression, and two-component systems used in synthetic biology use light to regulate transcription. The plasmid pDawn contains the blue-light sensing protein YFI. When light is absent, YFI phosphorylates FixJ, which binds to the FixK2 promoter to induce transcription of the phage repressor cI. Repressor cI inhibits transcription from phage promoter pR, preventing expression of a reporter gene. When light is present, YFI is inactive, preventing repressor cI synthesis and allowing reporter gene transcription to take place. Addgene also has the light-regulated two component systems designed by the Tabor lab.

WHICH INDUCIBLE PROMOTER IS RIGHT FOR ME?

There are several important variables to consider when choosing an inducible promoter.

Experimental system

Because transcription machinery differs between cell types or organisms, promoters are 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. The induction mechanism must also be compatible with your experimental system.

Table 4.1

Promoter Subtype	Promoter Example	Activator	Addgene Plasmid/Kit
Alcohol-regulated	AlcA promoter	AlcR	pGGA008 (AlcA) and pGGC011 (AlcR) from the GreenGate kit
Steroid-regulated	LexA promoter	XVE (synthetic)	pFZ19 (red flame)

Leakiness

If you're expressing a gene that may be toxic, it's important that your inducible promoter not be too leaky. For inducible prokaryotic promoters, pLac is known to be slightly leaky, and pBad is likely a better option since ex- pression can be repressed by glucose. Temperature inducible promoters are also known for their low leakiness.

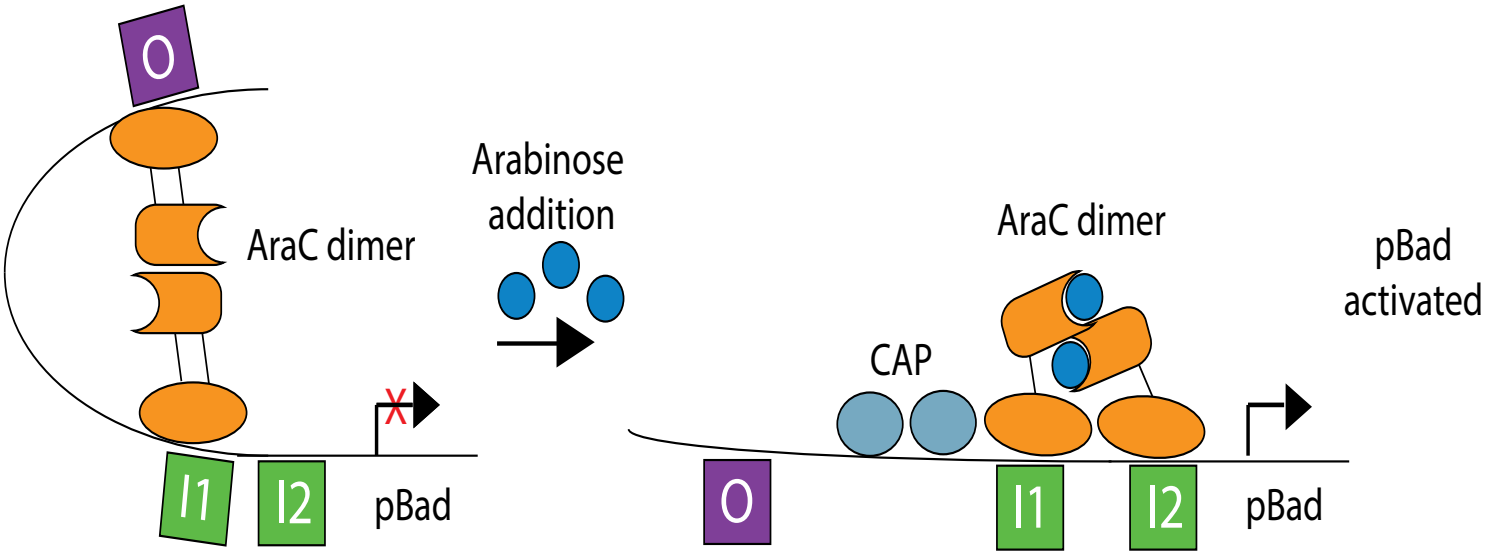
Inducibility level

How much do you want to induce transcription of your target gene? The strength of inducible promoters can vary a lot. If you're looking for high inducibility, the Tet-On system may be a good choice, as it's documented to induce transcription >1,000-fold when activated.

Lag time

The time needed to induce a given promoter varies. Promoters that require only the addition of an external inducer, like Tet systems, may be activated very rapidly. In contrast, promoter systems that require transcription of a repressor/ inducer will have higher lag time due to the time needed for transcription/translation.

Many thanks to Nicole Zurcher who contributed to the writing of this article. ■



CAP can also stimulate AraC to bind I1 and I2

Repressible Promoters

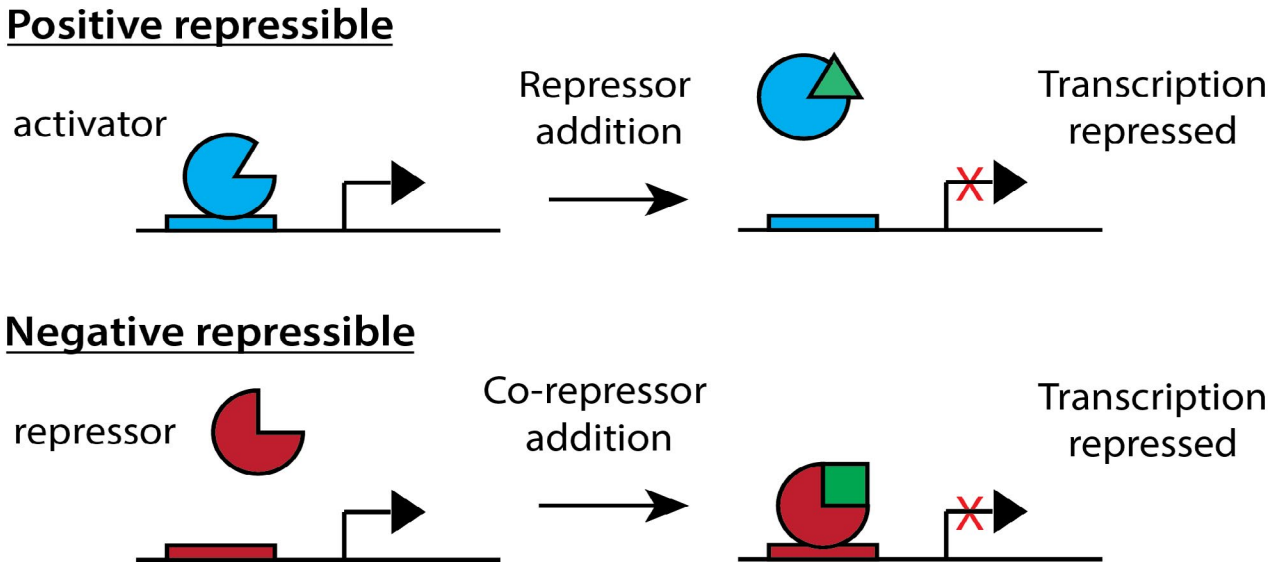
By Hannah Dotson and Mary Gearing | December 2022

Promoters control the binding of RNA polymerase and transcription factors. Since the promoter region drives transcription of a target gene, it therefore determines the timing of gene expression and largely defines the amount of recombinant protein that will be produced. Many commonly used promoters, such as T7, CMV, EF1A, and SV40, are always active and thus referred to as constitutive promoters. Others are only active under specific circumstances. In a previous post, we discussed inducible promoters, which can be switched from a default OFF to an ON state, and how you might use these in your research. Today, we'll look at repressible promoters, which can be switched from a default ON to an OFF state, as well as repressible binary systems. In recent years, a number of more complex systems have also been developed which don't fit in either category, but those are a subject for a future post!

HOW ARE INDUCIBLE PROMOTERS REGULATED?

Like inducible promoters, repressible promoters can be regulated via positive or negative control.

Positive Repressible: An activator protein is bound, turning transcription ON. When a repressor binds the activator protein, the activator protein cannot bind the promoter sequence anymore and transcription is turned OFF.



Negative Repressible: A repressor protein is present, but cannot bind the promoter sequence and transcription is ON. Once a co-repressor protein binds the repressor protein, the repressor protein can bind to the operator. The bound repressor then prevents transcription from occurring, which means that transcription is now OFF.

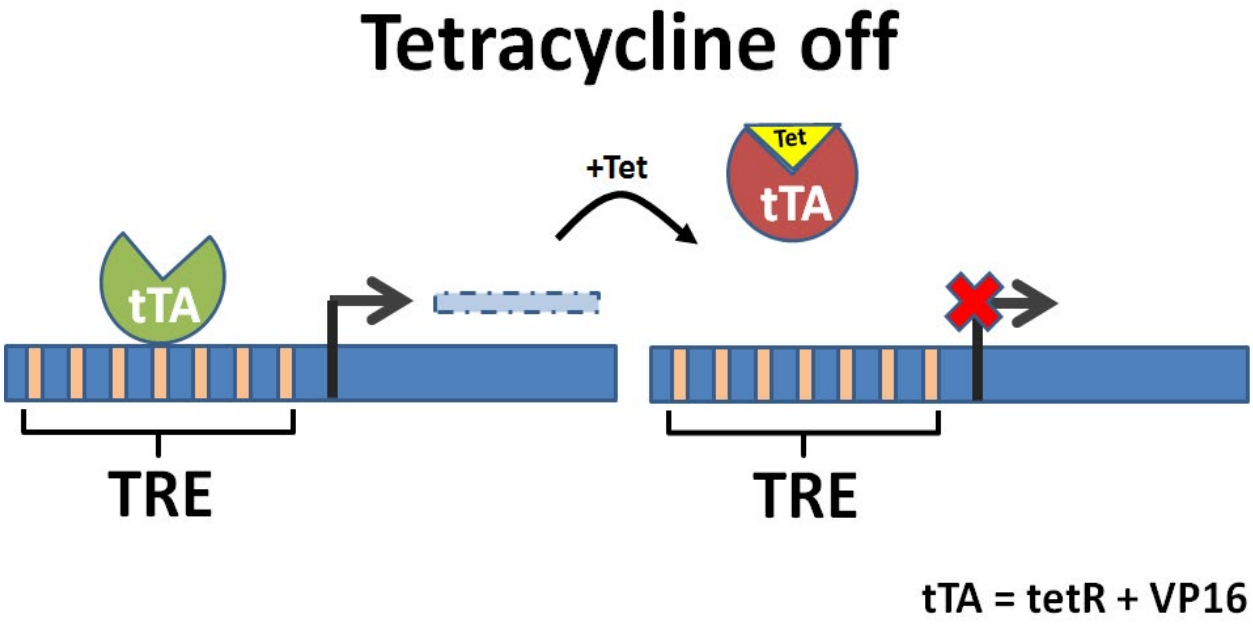
CHEMICALLY REPRESSIBLE PROMOTERS

Tet-Off

The Tet-Off system, a positive repressible promoter, was engineered from the bacterial tet operon. In its native context, the tetracycline repressor (TetR) can bind to the tetracycline operator sequences (TetO), preventing transcription. In the presence of tetracycline (Tet), TetR preferentially binds Tet over the TetO elements, allowing transcription to proceed (pJKR-H-tetR from the Church lab uses this system to control expression of GFP.)

To turn this inducible system into a repressible system, Gossen and Bujard created the tetracycline-controlled transactivator (tTA). They fused TetR with the transcriptional activation domain VP16 taken from the herpes sim- plex virus. tTA binds to promoters containing TetO elements (often linked in groups of seven as a Tet Response Element (TRE)), allowing transcription to proceed. When tetracycline or one of its derivatives is added, it binds tTA, resulting in a confirmation change that prevents binding to the promoter, thus turning transcription OFF.

Despite their bacterial origins, Tet systems function well in mammalian cells, and TRE-containing



promoters can be used in the repressible manner described above, as well as the inducible manner detailed in our previous post.

In the past 30 years since the original Tet-Off system was described by Gossen and Bujard, both the promoter and transactivator components of the system have been engineered and improved. Takara's "Tet-One systems" and "Tet-On 3G tetracycline-inducible expression systems" are the most recent variants in Tet technology.

Cumate switch

Massie et al. developed the cumate-inducible gene expression system, which functions in a similar manner to the Tet system. . The transactivator configuration of the system is a positive repressible promoter, and was engineered from the cmt and cym bacterial operons.

A chimeric transactivator, cTA, was created by fusing CymR to the activation domain VP16. cTA binds to promoters containing putative operator sequences (CuO) (linked in groups of 6), allowing transcription to proceed. When cumate is added, it binds cTA, resulting in a confirmation change that prevents binding to the promoter and turning transcription OFF.

The cumate switch system functions well in mammalian cells, and can be used in the repressible manner de- scribed above, as well as the inducible manner detailed in our previous post. While originally developed for use in mammalian cells, this system has also been implemented in a variety of other organisms, such as sphingomonads and bacillus.

ADH1

The ADH1 negative repressible promoter is commonly used in yeast. In the first 24 hours of culture, when glucose is abundant and ethanol concentrations are low, pADH1 displays low promoter activity (~20% activity of the strong yeast promoter pTEF). As ethanol accumulates, it binds to the repressor, enabling it to bind the promoter and repress pADH1 activity. Researchers have engineered pADH1 variants that are less sensitive to ethanol, including the "middle" ADH1 promoter pADH1m, which maintains activity during the late ethanol consumption phase of yeast culture.

REPRESSIBLE BINARY SYSTEMS

GAL4/UAS

In Drosophila or development studies, you may hear some transcription factor and promoter pairs referred to as binary systems. Binary systems are genetic tools that consist of two parts, both of which must be present for a gene of interest to be expressed. These permit exquisite control of gene expression and tracing of gene expression across development.. One common binary system is the GAL4/UAS system isolated from yeast. In this system, UAS basal promoter expression is low but is activated by GAL4 binding to UAS.

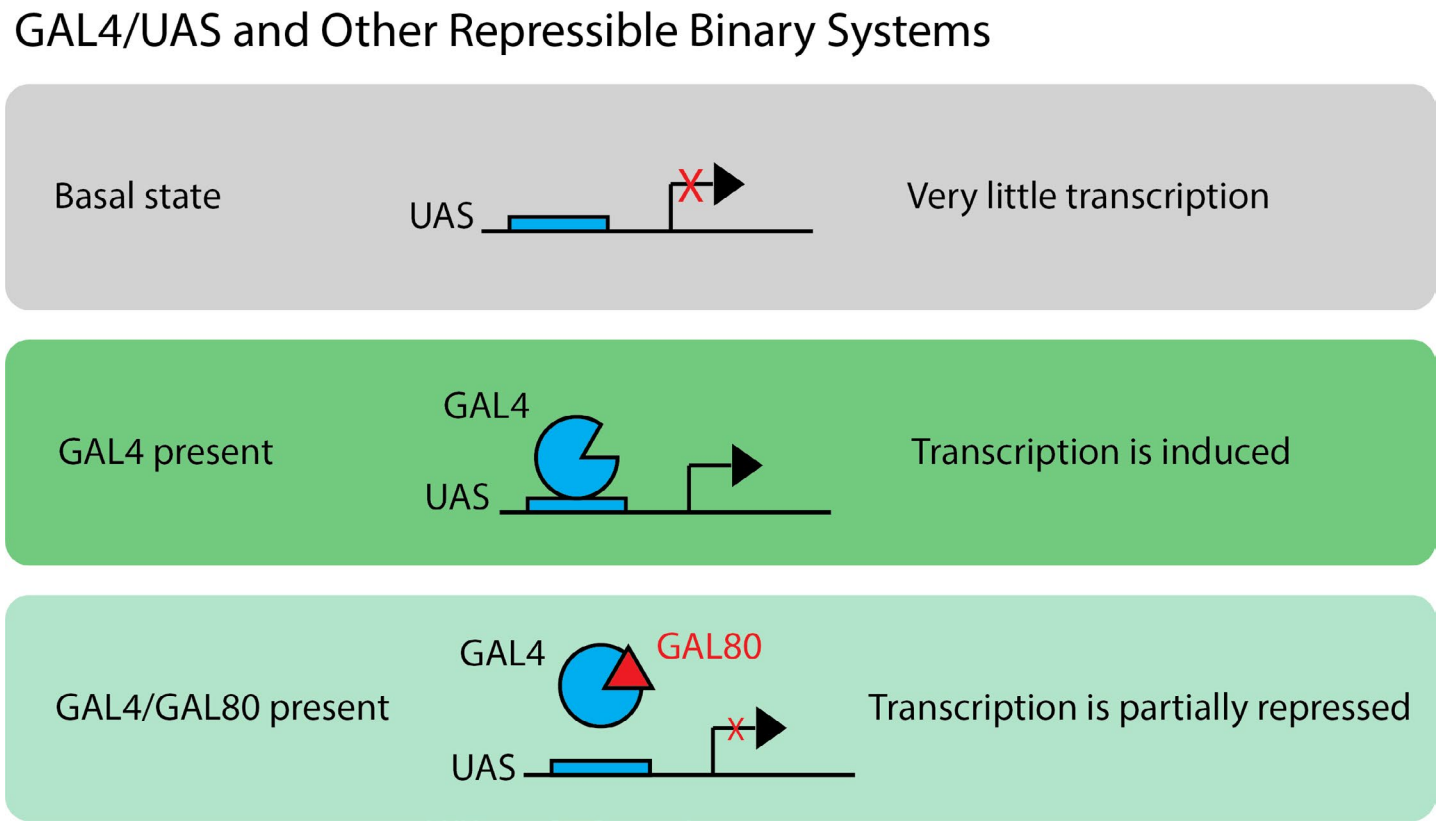
If you place the GAL4 gene downstream of a tissue- or developmental stage-specific promoter and design a UAS reporter construct, the reporter

will only be expressed where and/or when that promoter is active. In this way, you can interrogate the activity of uncharacterized promoters. Similarly, placing UAS upstream of a transgene permits directed expression of that gene in cells that also express GAL4. Usage of this system revolutionized genetic studies in flies.

An additional component, the GAL80 repressor, adds a repressible element to this binary system (making this a repressible binary system). The GAL80 repressor can bind to GAL4, partially inhibiting the binding of GAL4 to UAS. One can therefore place GAL4 and GAL80 under different promoters and create sophisticated patterns of UAS-driven gene expression. To further increase the complexity of the system, researchers have used split GAL4 architecture or combined the system with Flp/FRT, as reviewed by del Valle Rodriguez et al.

LexA/lexAop

A second binary system is LexA/lexAop, developed by Lai and Lee. LexA/lexAop is a complementary system to GAL4/UAS that functions in essentially the same manner, with LexA binding to the lexA operator (lexAop). Fusing the LexA DNA-binding domain (DBD) to GAL4 or VP16 activation domain results in chimeric GAL80-repressible and GAL80-insensitive activators of lexAop.. This system is commonly used with GAL4/UAS to examine the



expression of reporter genes, or to combine reporter expression from one promoter with transgene or siRNA expression from the other.

The Q system

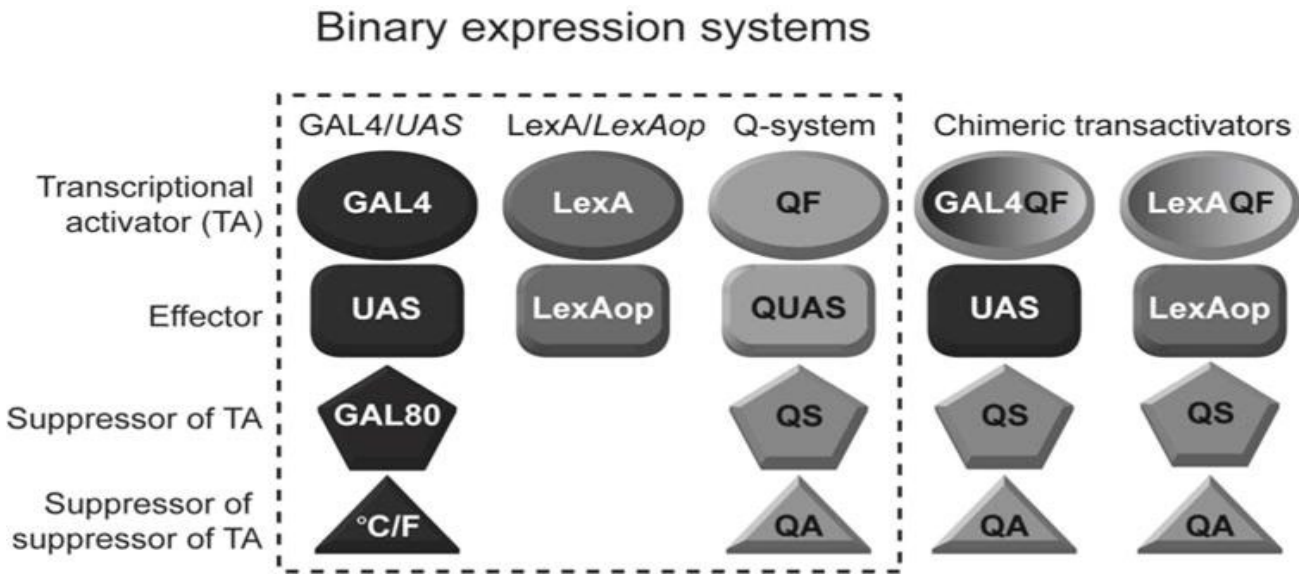
Potter et al. from Liqun Luo's lab developed the Q system from a gene cluster in the fungus Neurospora crassa. The QUAS promoter is less leaky than UAS in the basal state, and co-expression of the inducer QF can increase QUAS driven expression by ~3,300-fold in *Drosophila* cells and ~24,000-fold in human HeLa cells as compared to basal expression. Co-expression of repressor QS along with QF leads to intermediate expression from QUAS, as seen with GAL4/GAL80 and UAS. QF-mediated repression is reversible by the addition of quinic acid to Dro- sophila cells. Work by Wei et al. from Kang Shen's lab has also shown that the Q system is also functional in C. elegans.

Subsequent work by the Luo and Potter labs has refined the Q system and introduced the second generation QF2 activators. These activators solve the problem of broad QF expression-induced lethality in flies. Riabinina et al. also developed chimeric transactivators GALQF and LexAQF, which activate their respective UAS and lexAop sequences but are also regulated by QS and quinic acid.

USING BINARY SYSTEMS TOGETHER

You may be wondering, "why is it necessary to have all 3 of the different binary systems detailed above if they all behave very similarly?" The simple answer is that sometimes it is necessary or desirable to control or look at multiple genes at the same time.

The 3 binary systems listed above are orthogonal, meaning that they do not affect each other. For example, the GAL4 transcription factor cannot drive transcription from the QUAS promoter, and the QF transcription factor cannot drive transcription



from the UAS promoter. This is important because it allows them to be used together for complex genetic analysis in Drosophila, or for precise control over multiple genes in synthetic systems. Recent work has even used all 3 in a single modular vector for ease of use!

CONCLUSION

This post contains only a brief overview of some of the most common repressible promoter systems, but there are certainly more out there! The repressible promoter systems detailed in this post can also be combined with other systems, such as inducible promoter systems, chemogenetic systems, and optogenetic systems, allowing for more complex control over desired gene expression. Take a look at all of the options and decide what might work best for your experiments! ■

REFERENCES

Gossen M, Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A. 1992 Jun 15;89(12):5547-51. doi: 10.1073/pnas.89.12.5547. PMID: 1319065; PMCID: PMC49329.

Triezenberg SJ, Kingsbury RC, McKnight SL. Functional dissection of VP16, the trans-activator of herpes sim- plex virus immediate early gene expression. Genes Dev. 1988 Jun;2(6):718-29. doi: 10.1101/gad.2.6.718. PMID: 2843425.

Takara Bio. Tetracycline inducible expression—Tet-One systems. https://www.takarabio.com/products/gene-function/tet-inducible-expression-systems/tet-one-systems

Mullick, A., Xu, Y., Warren, R. et al. The cumate gene-switch: a system for regulated expression in mammalian cells. BMC Biotechnol 6, 43 (2006). https://doi.org/10.1186/1472-6750-6-43

Kaczmarczyk A, Vorholt JA, Francez-Charlot A. Cumate-inducible gene expression system for sphingomonads and other Alphaproteobacteria. Appl Environ Microbiol. 2013 Nov;79(21):6795-802. doi: 10.1128/AEM.02296-13. Epub 2013 Aug 30. PMID: 23995928; PMCID: PMC3811519.

Seo SO, Schmidt-Dannert C. Development of a synthetic cumate-inducible gene expression system for Bacillus. Appl Microbiol Biotechnol. 2019 Jan;103(1):303-313. doi: 10.1007/s00253-018-9485-4. Epub 2018 Nov 3. PMID: 30392122.

Pfeiffer, Barret D., et al. "Refinement of Tools for Targeted Gene Expression in Drosophila." Genetics 186(2) (2010): 735-55. PubMed PMID: 20697123. PubMed Central PMCID: PMC2942869.

del Valle Rodriguez, Alberto, Dominic Didiano & Claude Desplan. "Power tools for gene expression and clonal analysis in Drosophila." Nat Methods 9(1) (2011): 47-55. PubMed PMID: 22205518. PubMed Central PMCID: PMC3574576.

Lai, Sen-Li & Tzumin Lee. "Genetic mosaic with dual binary transcription systems in Drosophila." Nat Neurosci 9(5) (2006): 703-9. PubMed PMID: 16582903.

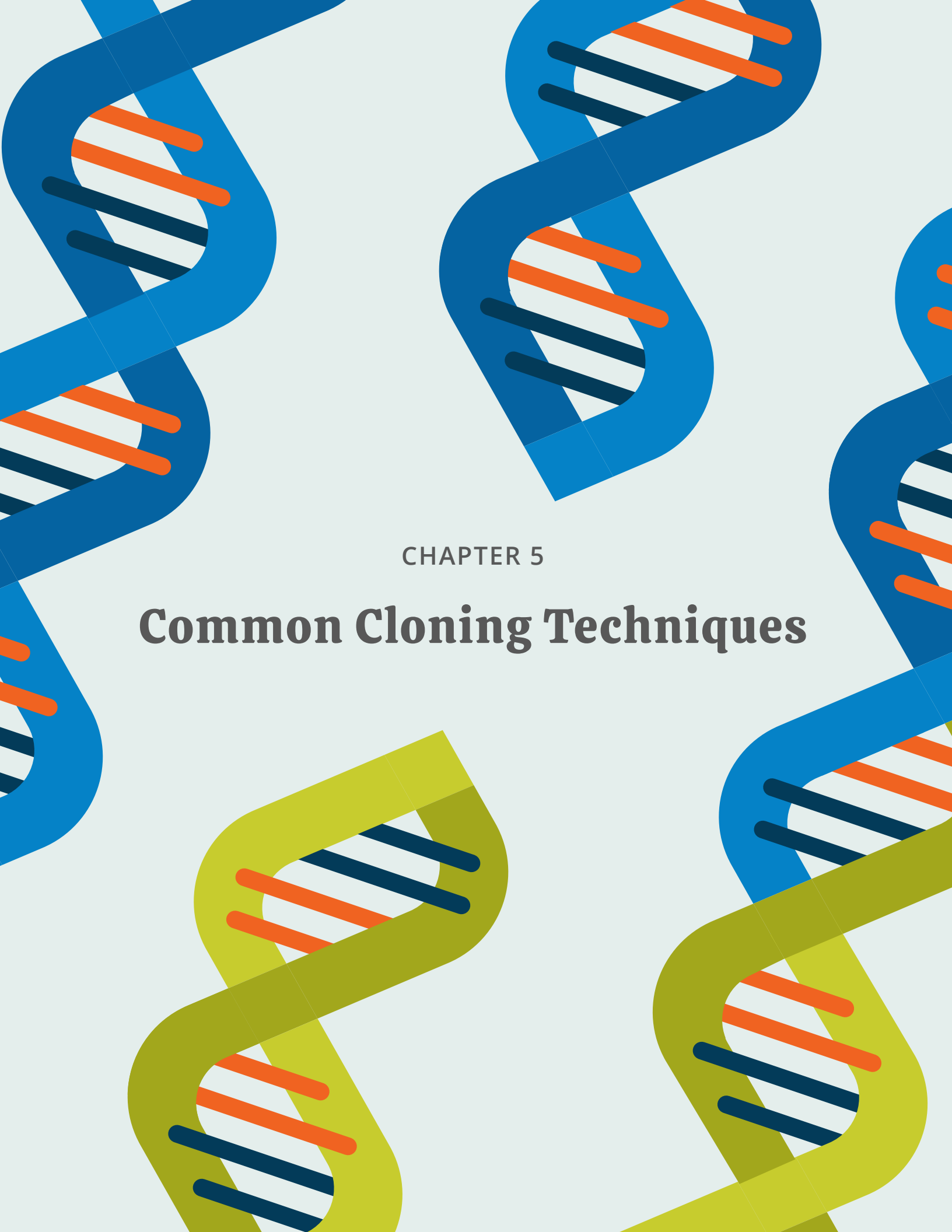
Potter, Christopher J., et al. "The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis." Cell 141(3) (2010): 536-48. PubMed PMID: 20434990. PubMed Central PMCID: PMC2883883.

Wei, Xing, Christopher J. Potter, Liqin Luo & Kang Shen. "Controlling gene expression with the Q repressible bi- nary expression system in Caenorhabditis elegans." Nat Methods 9(4) (2012): 391-5. PubMed PMID: 22406855. PubMed Central PMCID: PMC3846601.

Riabinina, et al. "Improved and expanded Q-system reagents for genetic manipulations." Nat Methods 12(3) (2015): 219-222. PubMed PMID: 25581800. PubMed Central PMCID: PMC4344399.

Riabinina O, Potter CJ. The Q-System: A Versatile Expression System for Drosophila. Methods Mol Biol. 2016;1478:53-78. doi: 10.1007/978-1-4939-6371-3_3. PMID: 27730575; PMCID: PMC5270762.

Wendler, F., Park, S., Hill, C. et al. A LexAop > UAS > QUAS trimeric plasmid to generate inducible and intercon- vertible Drosophila overexpression transgenes. Sci Rep 12, 3835 (2022). https://doi.org/10.1038/s41598-022- 07852-7



CHAPTER 5

Common Cloning Techniques

Five Factors to Help You Choose the Right Cloning Method

By Michael G. Lemieux | August 2018

You've spent days and weeks thinking of an amazing project. You've written your protocols, designed your experiments, and prepared your reagents. You're going to engineer the best thing since CRISPR; you are ready to clone! But...how?

Cloning is no longer limited to restriction enzymes - although many labs still love using them! With so many technologies currently available, you now face the challenge of trying to figure out what system might work best for your particular purposes. We're here to help with that!

Four important factors to consider for any experiment are:

(1) Speed
(2) Cost
(3) Difficulty
(4) Efficiency

For cloning in particular, another important consideration is the number of inserts, as some methods are more amenable to combining multiple DNA fragments than others.

Here, we will focus on a high-level comparison of popular cloning methodologies, with a special emphasis on these five categories. In the rest of this chapter, we'll describe many of the following techniques in their own sections!

SPEED

Your grant deadline is approaching or your advisor is demanding results. If you're in a hurry, Golden Gate cloning might be for you. This methodology relies on versatile Type IIS restriction enzymes. These cut outside their restriction sites allowing you to easily create customized overhangs to assemble up to 6 different fragments in a prescribed order. These fragments can be pooled in one tube and ligated in as little as 30 minutes!

Alternatively, you might consider Gibson cloning. In this technique, you first perform a round of PCR to gener- ate DNA fragments with 15-40 bp of homology at their ends. These fragments are then combined with an exo- nuclease, a polymerase, and a DNA ligase to carry out a pooled set of ligations in an hour or less. The ligations combine the fragments in an order determined by their homology.

You might also consider FastCloning, first described by Chaokun Li and colleagues. Here you simply PCR-am- plify your vector and your insert, DpnI digest away the templates, and transform the PCR products into E. coli. Ligases within the *E. coli* then fuse together the PCR products via overhangs left over from the PCR process (the precise mechanism isn't well understood).

Finally, TOPO cloning is another good option to consider when time is of the essence. This method requires a vector backbone containing a topoisomerase I cut site, toposisomerase I, and a taq polymerase-generated insert. When mixed, topoisomerase I provides both the endonuclease and ligase function while 5' adenosine overhangs left by Taq perfectly anneal with the topoisomerase I-generated overhangs in the backbone. Once combined, the cut and ligation processes happen in a matter of minutes.

COST

Funding is usually limited - what cloning approach should you use if the budget is particularly tight? Most cloning processes will require a PCR reaction or two to generate templates for the cloning process. Time spent actually doing the experiments are also generally comparable so the best way to compare costs is to look at the prices of the additional enzymes required.

FastCloning may be the least expensive of all the options discussed as it only requires PCR and the inexpensive DpnI enzyme. Otherwise you let your *E. coli* do the work for you.

Traditional restriction enzyme and ligase-based cloning is a relatively inexpensive way to get the job done, provided that you've already got the right enzymes laying around the lab. If not, you'll be looking at about $50-$60 per enzyme. Once you've got a few of these common enzymes, however, your cost per cloning reaction drops dramatically, as one tube of enzyme should be enough for at least 50 cloning reactions. The ligase cost is comparable to that of the restriction enzymes and overall you're looking at a price of a few dollars per reaction for traditional restriction cloning. Sequence and Ligation Independent Cloning (SLIC) could also be a great choice. The T4 polymerase needed for this method is relatively inexpensive and can be obtained for about $70, resulting in a price of a few dollars per reaction.

DIFFICULTY

If you're an experienced cloner, you're likely not intimidated by having to pick the right enzymes, optimize reagent concentration, or purify DNA fragments from gels. If you're relatively new to the cloning game, however, or simply looking for a more streamlined approach, you might want to give Gateway a try. This method, developed by Invitrogen and based on lambda phage integration sequences, relies on site-specific recombination reactions catalyzed by BP Clonase and LR Clonase enzyme mixes. After generating the appropriate inserts by PCR, there is no need for restriction digests, gel purifications, or ligations. The general Gateway cloning protocol is fairly flexible and adaptable to cloning different types of DNA, such as cDNA, PCR fragments, or genomic DNA.

Also remember that commercially available cloning kits can often make the process much more streamlined, because they will come with comprehensive and tested protocols, as well as premixed and ready-to-use reagents. Thermo Fisher Scientific Inc., Origene, and NEB® are all good places to look for some helpful kits!

EFFICIENCY

Wouldn't it be nice if after all the hard work involved with cloning, you would at least be guaranteed a successful result? While there are never any guarantees, some cloning techniques do tip the scales in your favor. For starters, Golden Gate cloning is extremely efficient due to the use of the versatile type II restriction enzymes. In Golden Gate cloning, the desired end construct does not contain the type IIS recognition sites, allowing you to keep the type IIS enzymes in the reaction to redigest vector reseals and other incorrect products. This feature biases the reaction in favor of the desired assembly to make the process close to 100% efficient.

Gateway cloning also offers a very high chance of success (>95% for a single insert), mostly due to the use of toxic ccdB to prevent any non-recombinant vectors from propagating. The ligation-independent methods like Gibson and SLIC are also quite efficient; although efficiency often declines as the number of fragments you're hoping to assemble increases.

Multiple inserts

There are some cloning methods that make working with multiple inserts a breeze. Among them are some of the homology-based methods, like Gibson or SLIC, since one can pool the components and define their final order through sequence homology. Similarly, Golden Gate assembly allows one to simultaneously complete an ordered set of scarless ligations following digestion with Type II restriction enzymes. All of these options could be good choices if you have a particularly large number of inserts to assemble. You could also consider using Gateway if your insert number is a bit smaller, let's say four or less.You could also consider using Gateway if your insert number is a bit smaller, let's say four or less. You could also consider using Gateway if your insert number is a bit smaller, let's say 4 or less.

THE LIST GOES ON...

I hope we've given you some tips for your next project. But keep in mind that there are many other cloning methods - including some that rely solely on PCR, and a host of commercially available cloning kits that one could choose from to perform these experiments. In most cases, there will be more than one right answer, so don't be afraid to explore multiple approaches to find out which one works best for you. Happy cloning (and be sure to deposit your finished constructs at Addgene!). ■

REFERENCES

Li, Chaokun, et al. "FastCloning: a highly simplified, purification-free, sequence-and ligation-independent PCR cloning method." BMC biotechnology 11.1 (2011): 92. PubMed PMID: 21992524.

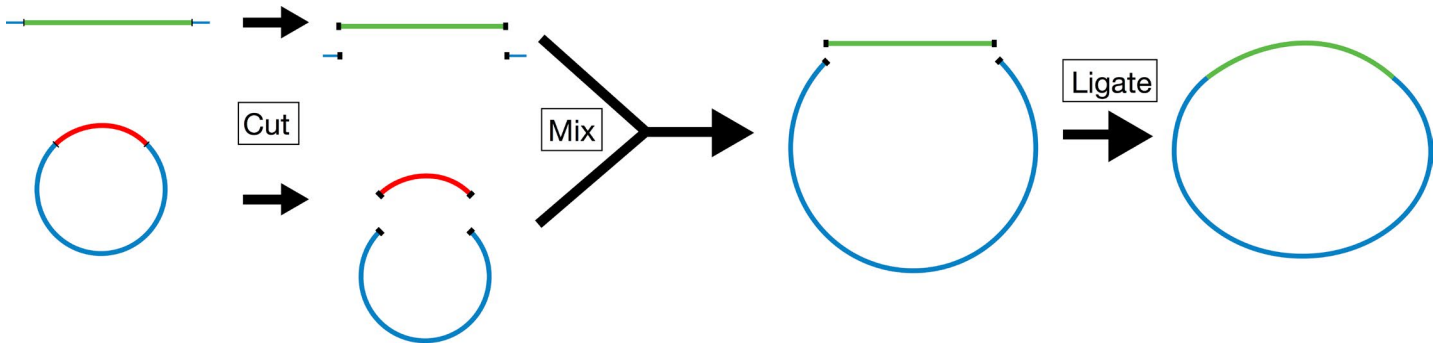
Restriction Cloning

By Tyler Ford | August 2015

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.

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,



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.

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.

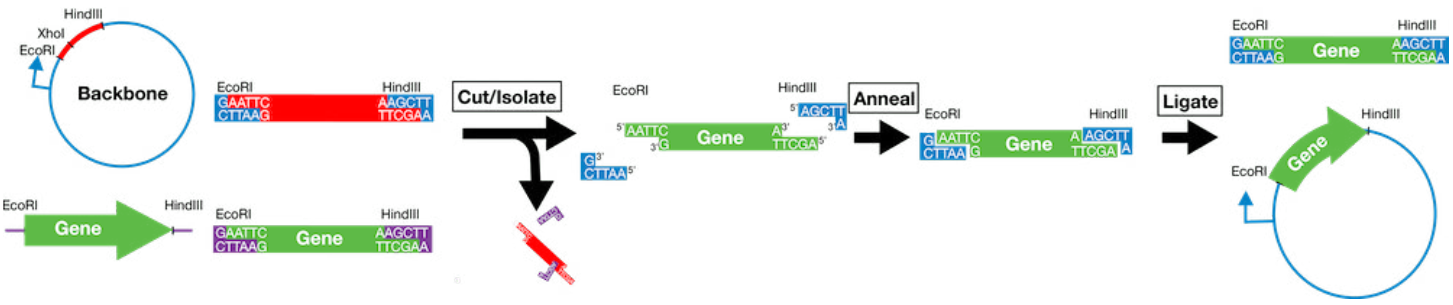
Alternatively, this whole process can be completed using a single enzyme if your insert is flanked on both sides by that enzyme's restriction sites, but the insert can then anneal to the backbone in either a forward or reverse orientation so you'll need some way to verify that the insert ended up in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

DIGESTION

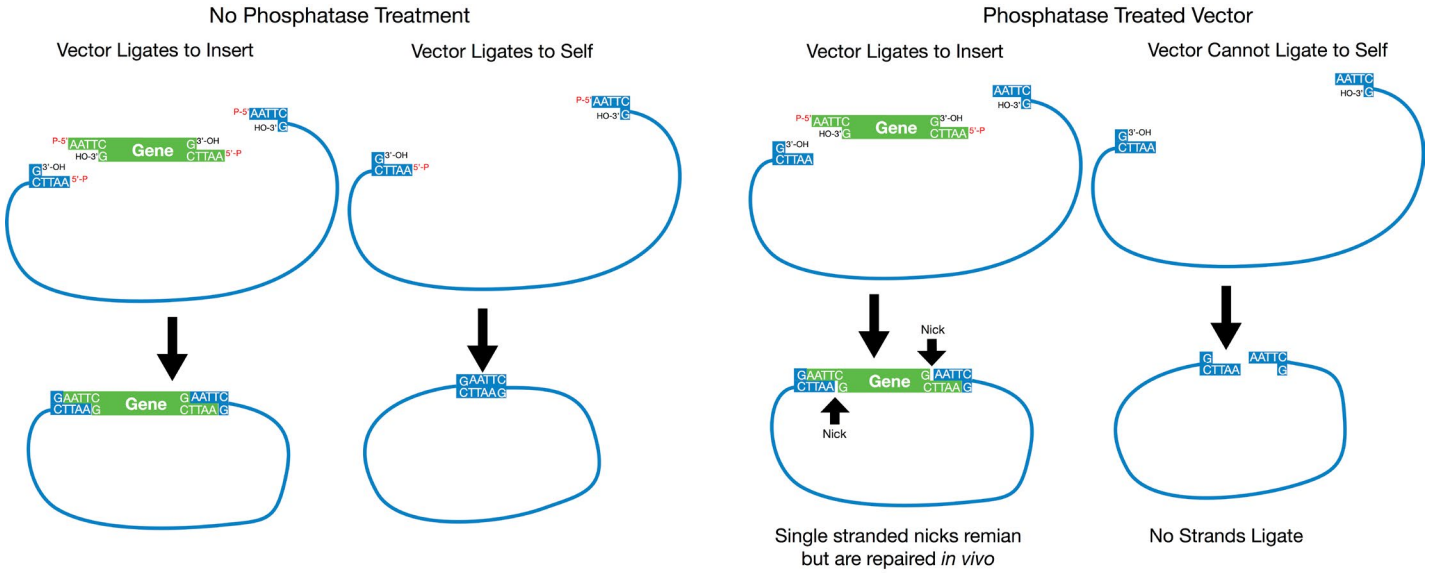
Set up restriction digests for your insert (or donor plasmid) and plasmid backbone. Because you lose some DNA during the gel purification step, it is important to digest plenty of starting material. We recommend 1.5—2 µg of insert and 1 µg of plasmid backbone. It is also critical that as much of the 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.

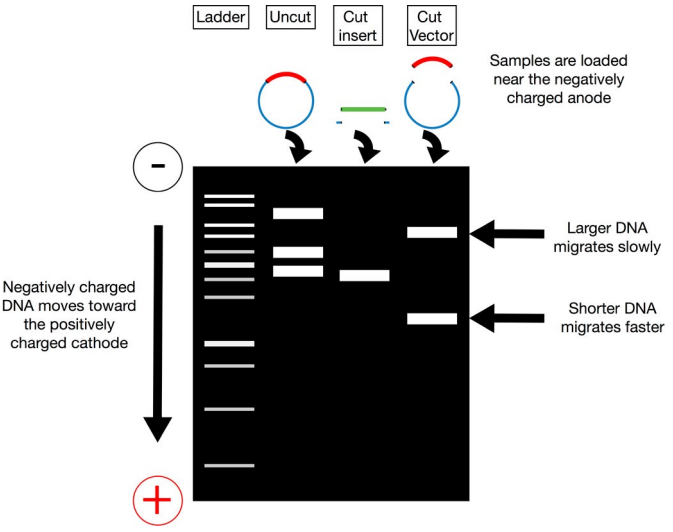


Detailed overview of the restriction cloning process. See text for details.



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 anode (- charge). When the voltage is applied across the gel, the DNA migrates toward the cathode (+ 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 (see Table 5.1).



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. Ethidium bromide can also be a mutagen.

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.

Table 5.1 – Common Gel Stains

Stain	Pre or Post Run?	Visualization	Sensitivity (ng)
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 (~500 nm) 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

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 (Figure1-Ligase).

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.

TRANSFORMATION

First, **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 µl 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.

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 in Figures 1, 2, and 3. 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 (Fig. 2) 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 lligase. If the

Figure 1-Ligase

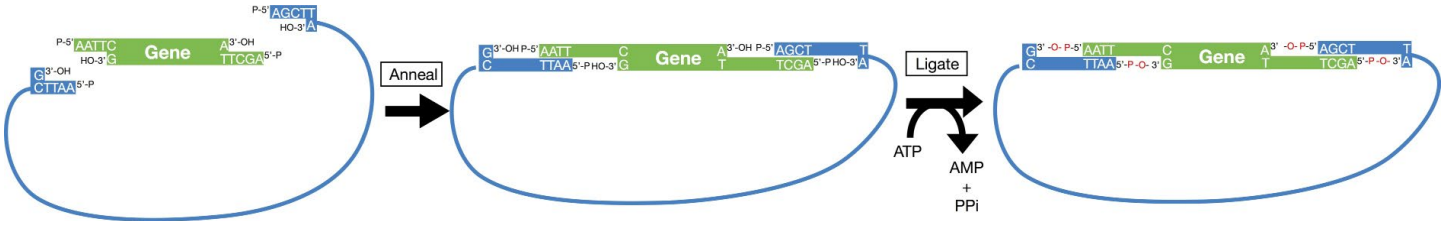


Figure 2-Transformation

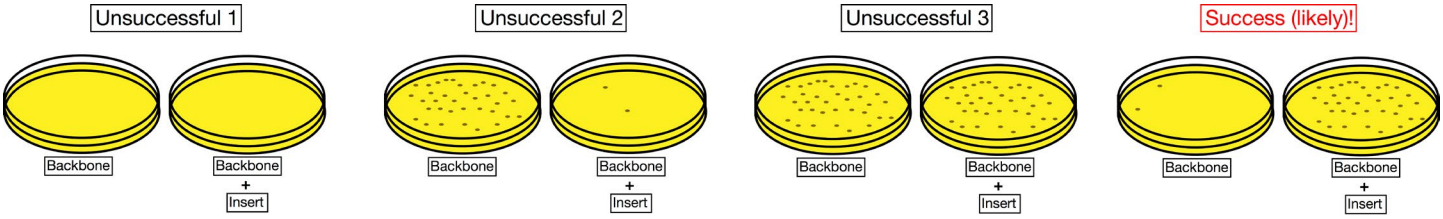
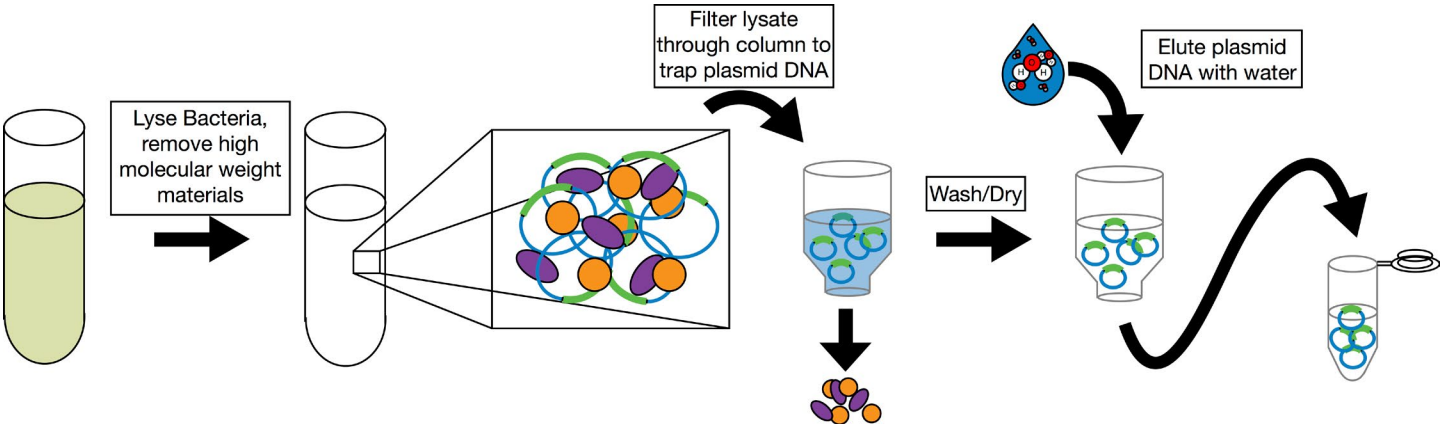


Figure 3-Purification



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.

PURIFY THE FINISHED PLASMID

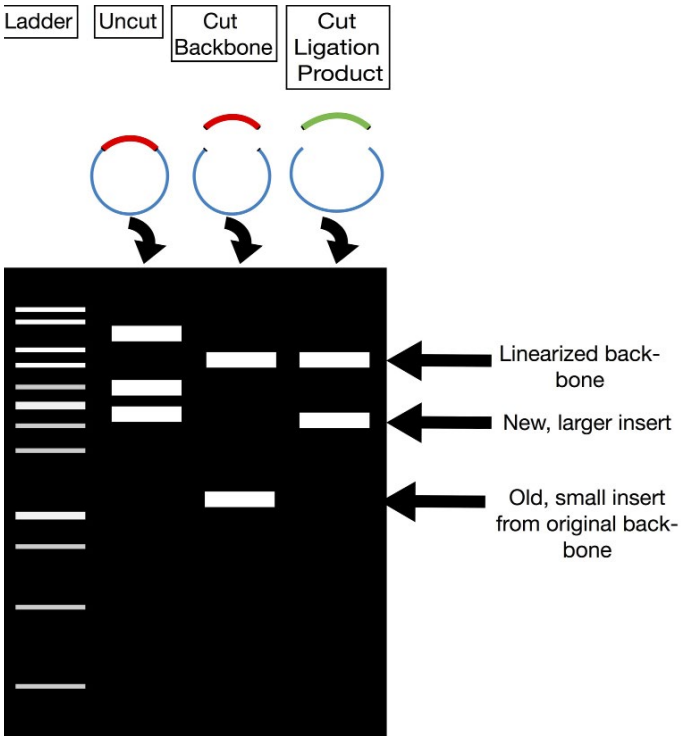
Once it looks like your ligation has worked, you will need to pick individual bacterial colonies and check them for successful ligation. Pick 3-10 colonies depending on the number of background colonies on your control plate (the more background, the more colonies you will need to pick) and grow overnight 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.

VERIFY THE FINISHED PLASMID

After **purifying the DNA**, conduct a **diagnostic restriction digest** of 100-300ng 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.

Once you know that your plasmid has an appropriately sized insert, you will need to verify its sequencing using a number of different methods - but that's a topic for another chapter! ■



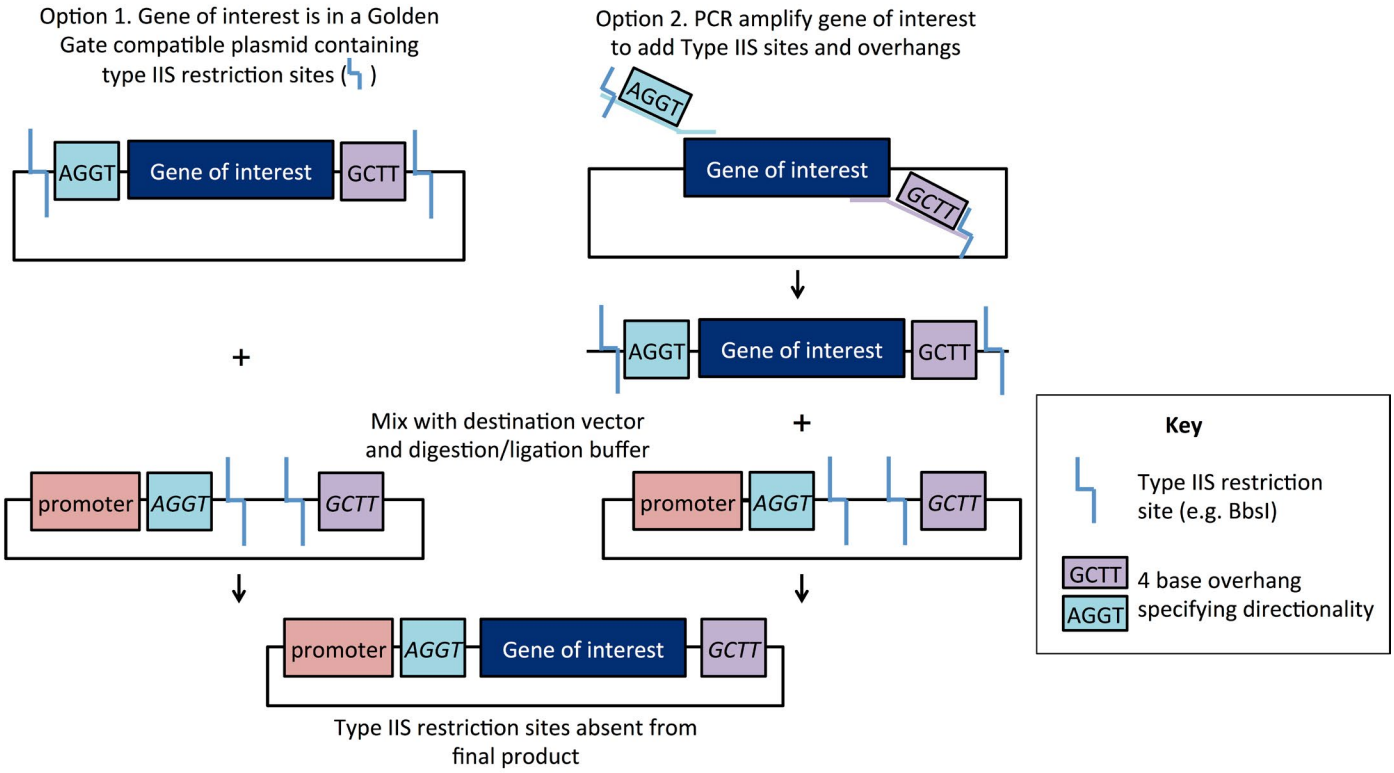
Golden Gate Cloning

By Mary Gearing | August 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

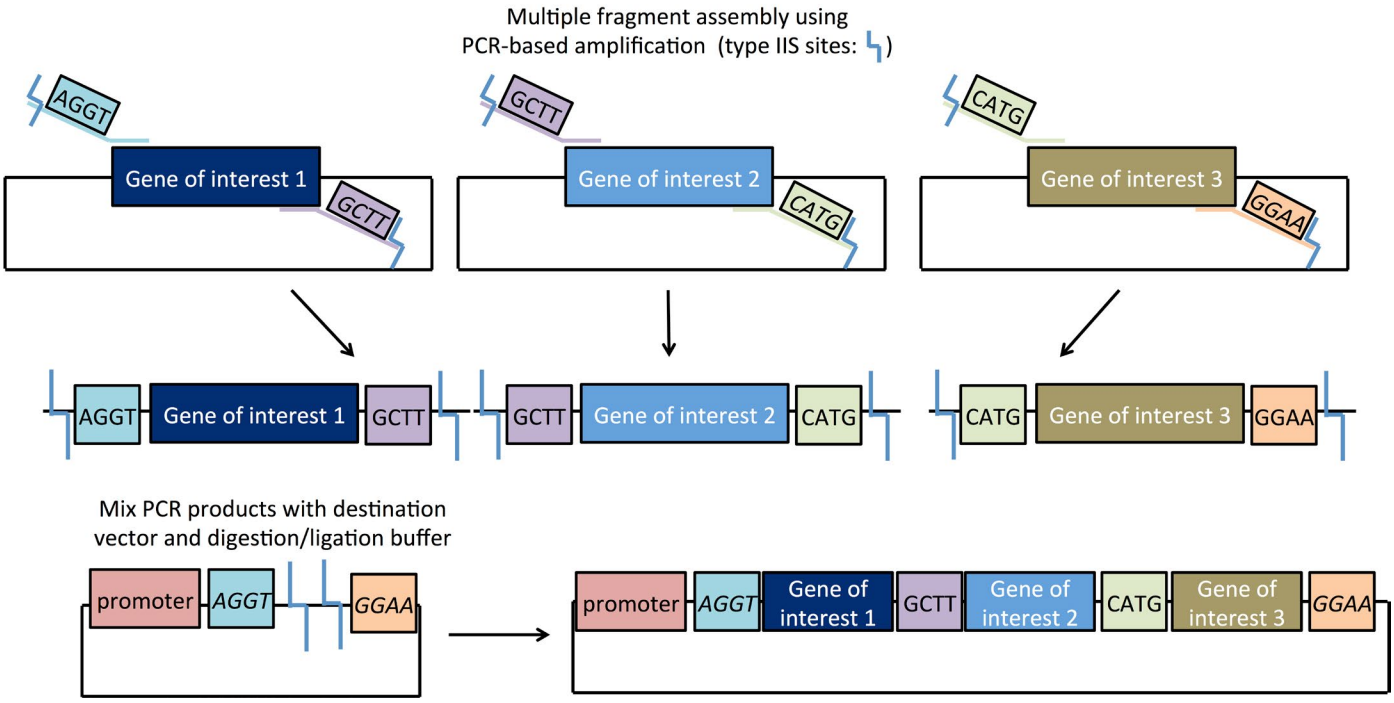


appear in the final construct. The destination vector contains sites with complementary overhangs that direct assembly of the final ligation product. As shown below, a fragment with 5′ overhang TGGA and 3′ overhang TCCG can be ligated into a vector containing those overhangs. Entry DNA overhangs may be present in the original plasmid (Option 1) or added using PCR-based amplification (Option 2).

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 AND SYNTHETIC BIOLOGY

Synthetic biologists have leveraged the power of Golden Gate cloning into a modular cloning strategy. Some- times 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

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 adapated 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. ■

Further Reading

Lee, Jae H., et al. "Sequential amplification of cloned DNA as tandem multimers using class-IIS restriction enzymes." Genetic analysis: biomolecular engineering 13.6 (1996): 139-145. PubMed PMID: 9117889.

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.

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.

Engler, Carola, and Sylvestre Marillonnet. "Golden gate cloning." DNA Cloning and Assembly Methods (2014): 119-131. PubMed MID: 24395361

Topo Cloning

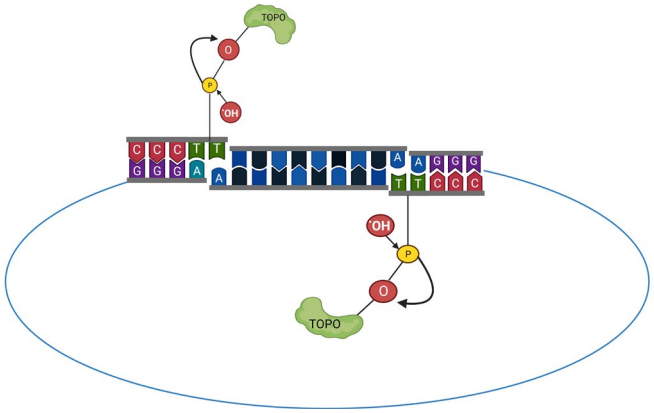
By Lianna Swanson | October 2016

Toposiomerase 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 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 endonuclease 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 (Shuman, S. 1991). 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 (Shuman, S. 1994).

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 ºC overnight.

4. Transform TOPO Cloning Reaction into Competent Cells: You can use your standard lab protocol for this; 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.

TIPS AND TRICKS FROM THE BENCH

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 ºC 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.
- 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. ■

REFERENCES

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.

Shuman S. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. J Biol Chem. 1994 Dec 23;269(51):32678-84. PubMed PMID: 7798275.

Sequence and Ligation-Independent Cloning (SLIC)

By Mary Gearing | 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.

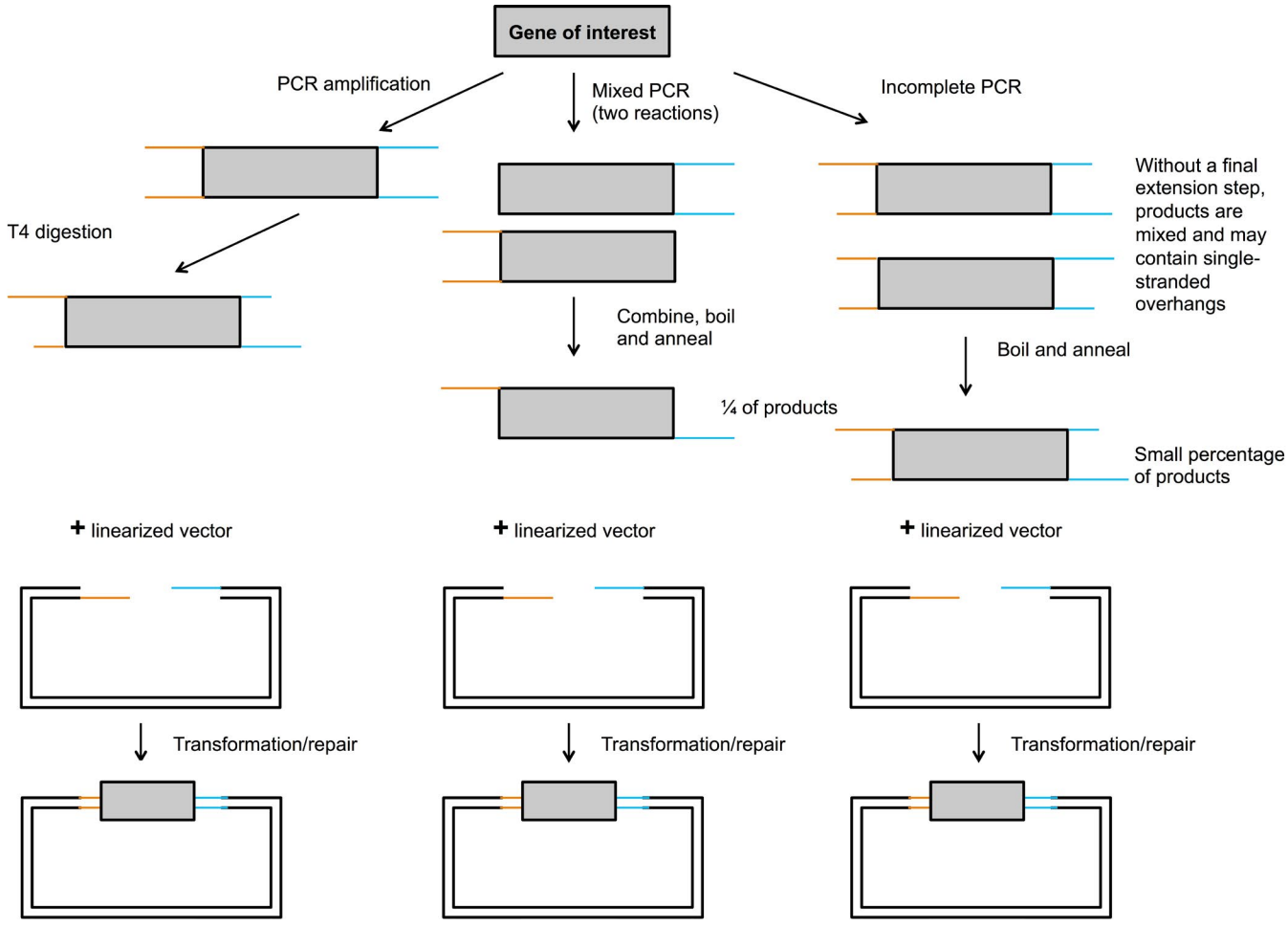
A SLIC MAKEOVER: SEQUENCE AND LIGATION-INDEPENDENT CLONING

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, 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



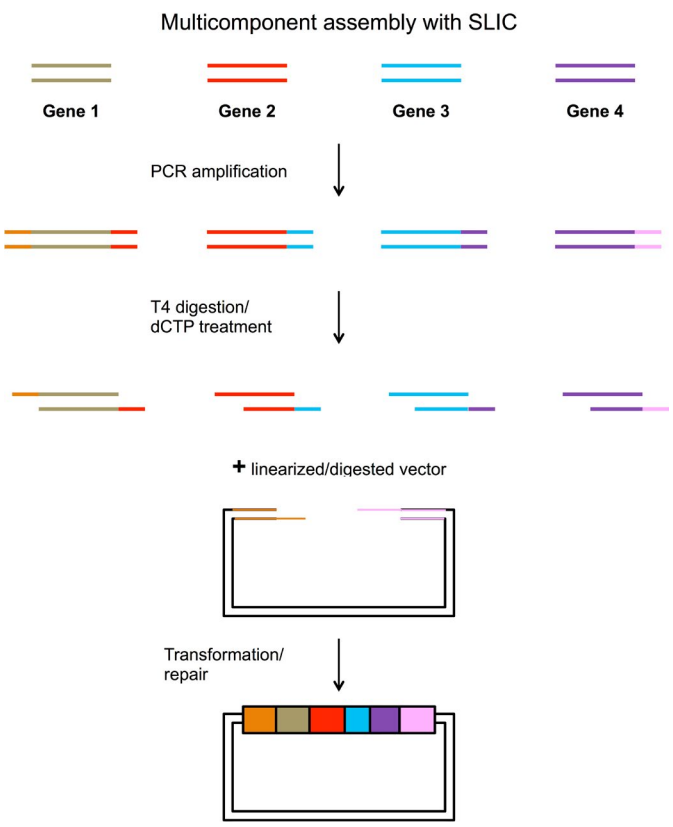
create an insert. Two PCR products are used to generate a target gene with a 5′ or 3′ overhang. When the products are mixed and annealed, 25% of the resulting DNA will have two single-stranded overhangs that can robustly stimulate recombination.

SLIC is ideal for multicomponent assembly, 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

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.

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.

ccdB: The Toxic Key to Efficient Cloning

By Michael G. Lemieux | December 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 GatewayTM 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.

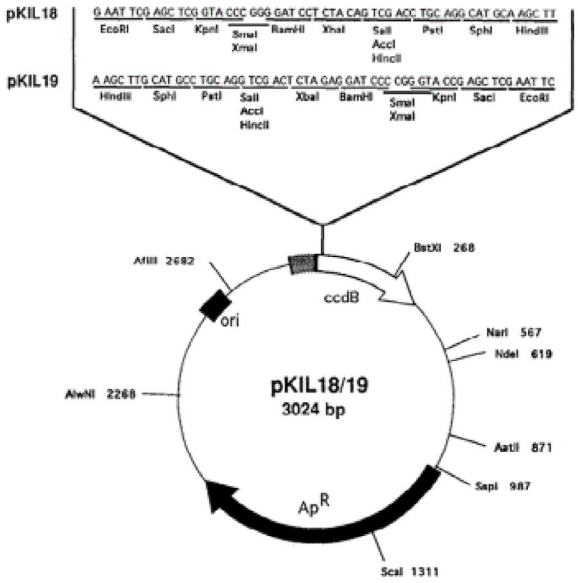


Image from Bernard, P., et al. showing the pKIL18/19 vectors and demonstrating the concept of restriction enzyme-mediated disruptions of ccdB, leading to positive identification of desired clones.

BECOMES A POWERFUL CLONING TOOL

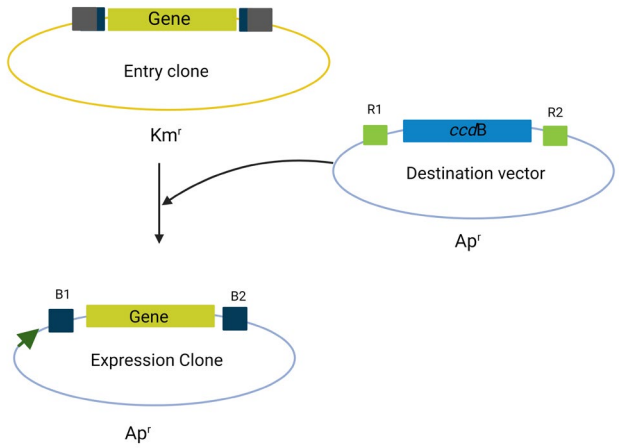
Molecular biologists first saw the potential of this system for enhancing cloning efficiency about 20 years ago and developed cloning vectors to harness it. These vectors, called pKIL18 and pKIL19, contained the ccdB gene inframe with an MCS. E. coli that were transformed with the empty vectors expressed 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 chapter on Gateway cloning. Focusing on the ccdB aspect, 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.

CCDB RESISTANT *E. COLI* STRAINS COMPLETE THE SYSTEML

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™ from InvitrogenTM. Using either DB3.1 or ccdB SurvivalTM, 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™ 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-); however, there are a few popular lab strains, such as NEB Stable, JM109, XL1 Blue or XL10 Gold, that are F'. These strains could possibly be used to propagate and prep your ccdB-containing empty backbones but should never be used when selecting for recombinant plasmids. Check out our post on common lab *E. coli* Strains for more strain info. ■



Further Reading

Bernard, P. Positive Selection of Recombinant DNA by CcdB. Biotechniques. 1996 Aug;21(2):320-3. PubMed PMID: 8862819.

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.

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.

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.

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.

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 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.

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

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.

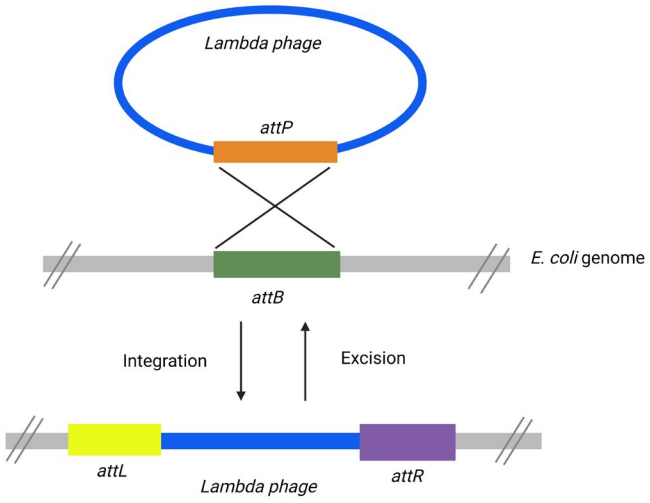


Figure 1: Lambda phage integration and excision reactions. Recombination of attP and attB sites creates attL and attR sites.

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 diffaerent 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. DH5α, 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

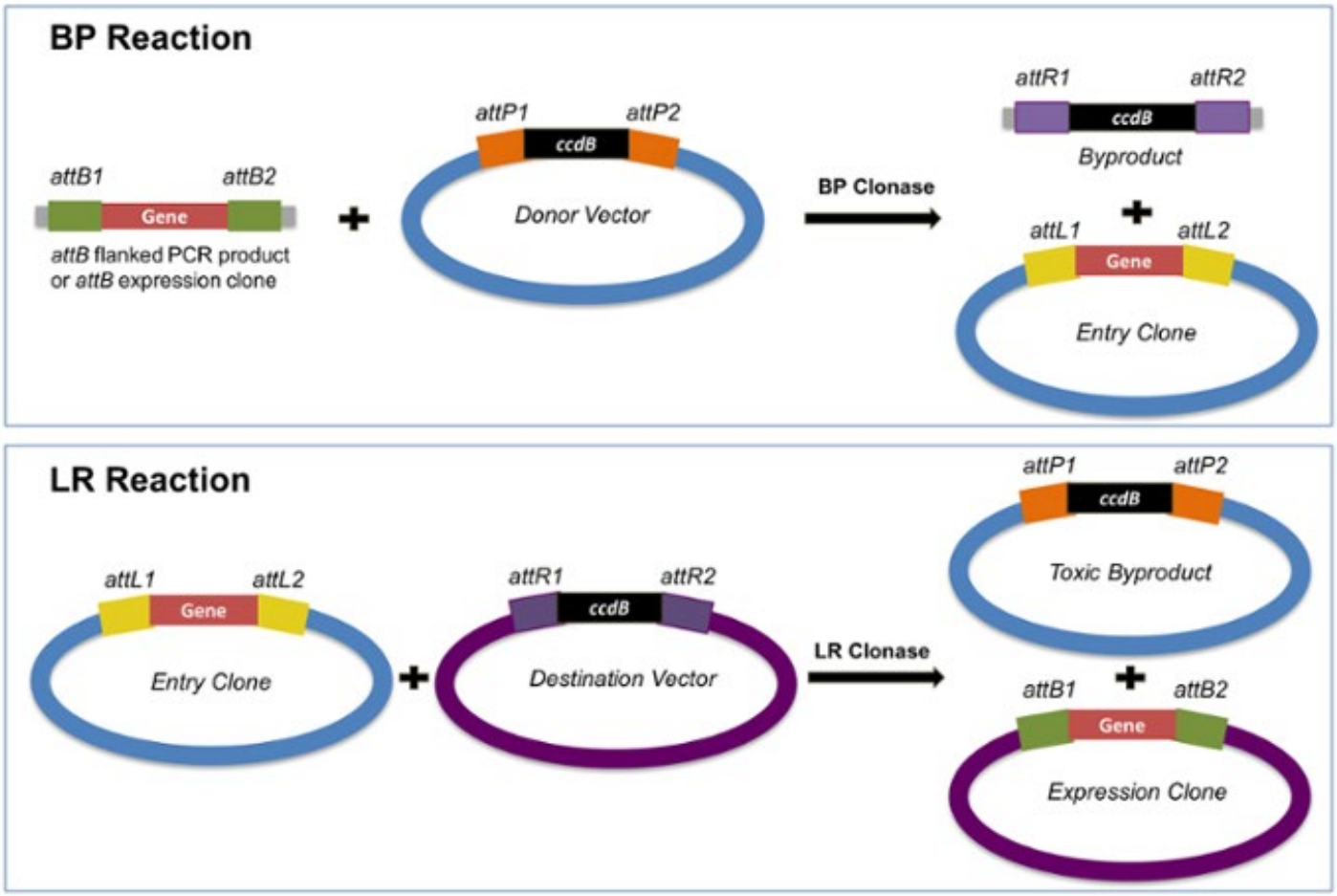


Figure 2: The Gateway system adapts phage integration into the BP and LR reactions. The BP reaction creates an attL-flanked entry clone. The LR reaction creates an expression clone with all of the components necessary for gene expression.

E. coli strains, most colonies should contain the desired, recombined construct. Read the previous section on CcdB for more information.

HOW TO CLONE USING GATEWAY TECHNOLOGY

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

Method A (Figure 3): 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.

Method B (Figure 4): 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.

Method C (Figure 5): 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.

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!

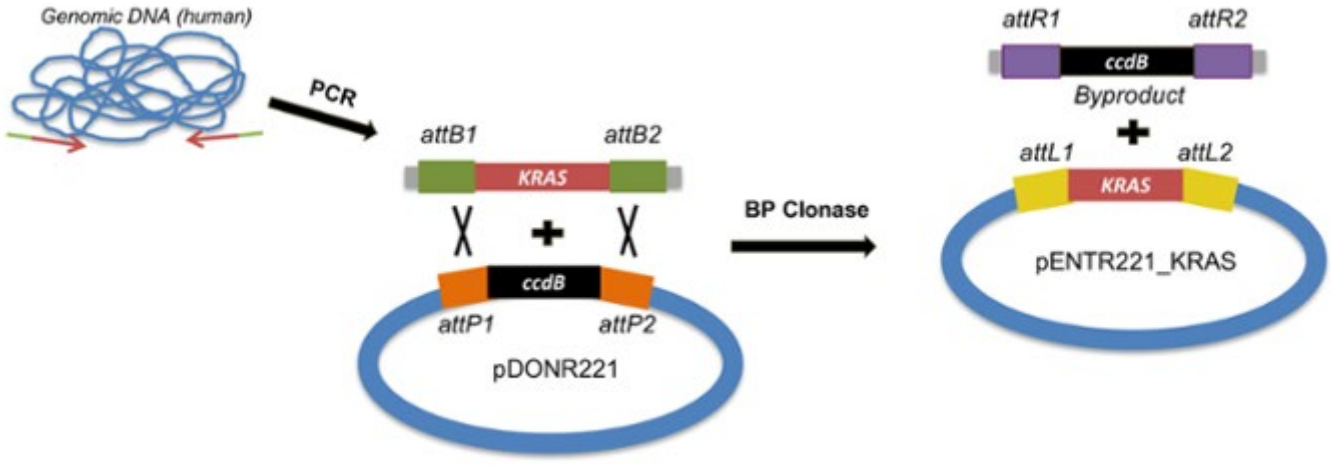


Figure 3: Method A to create an entry clone: recombination of an attB-flanked PCR product with an attP-containing donor vector.

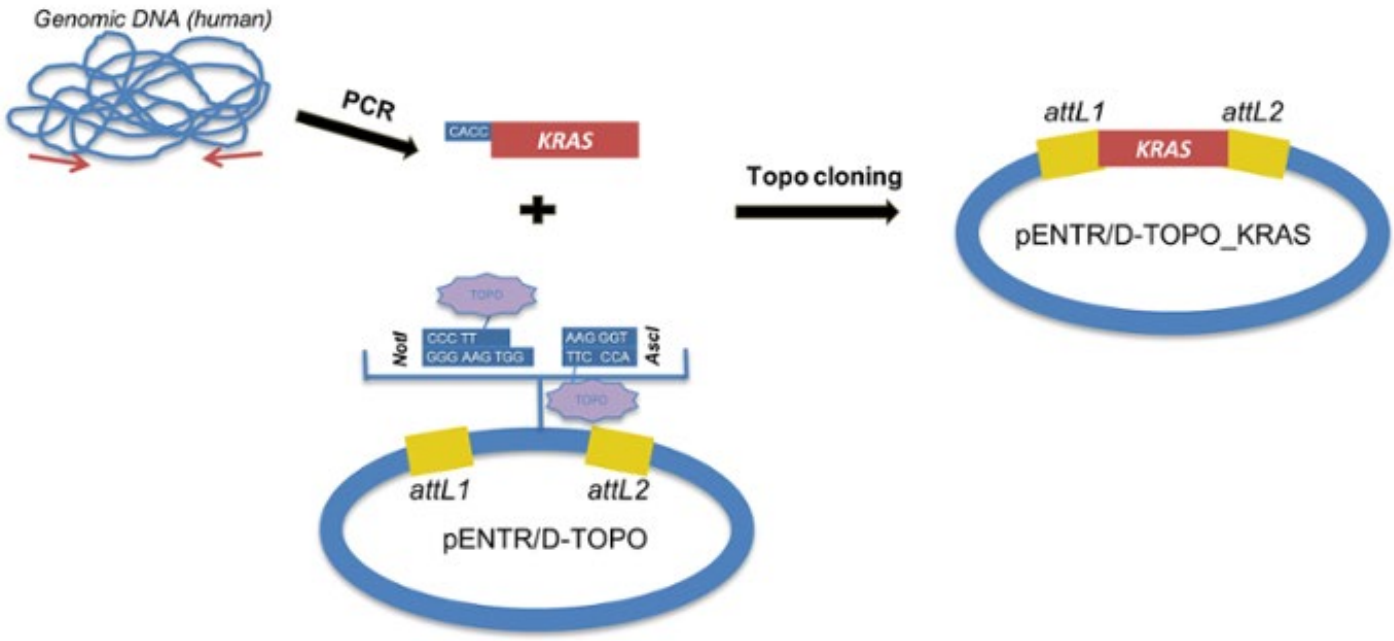


Figure 4: Method B to create an entry clone: TOPO cloning the insert into an attL-containing entry vector

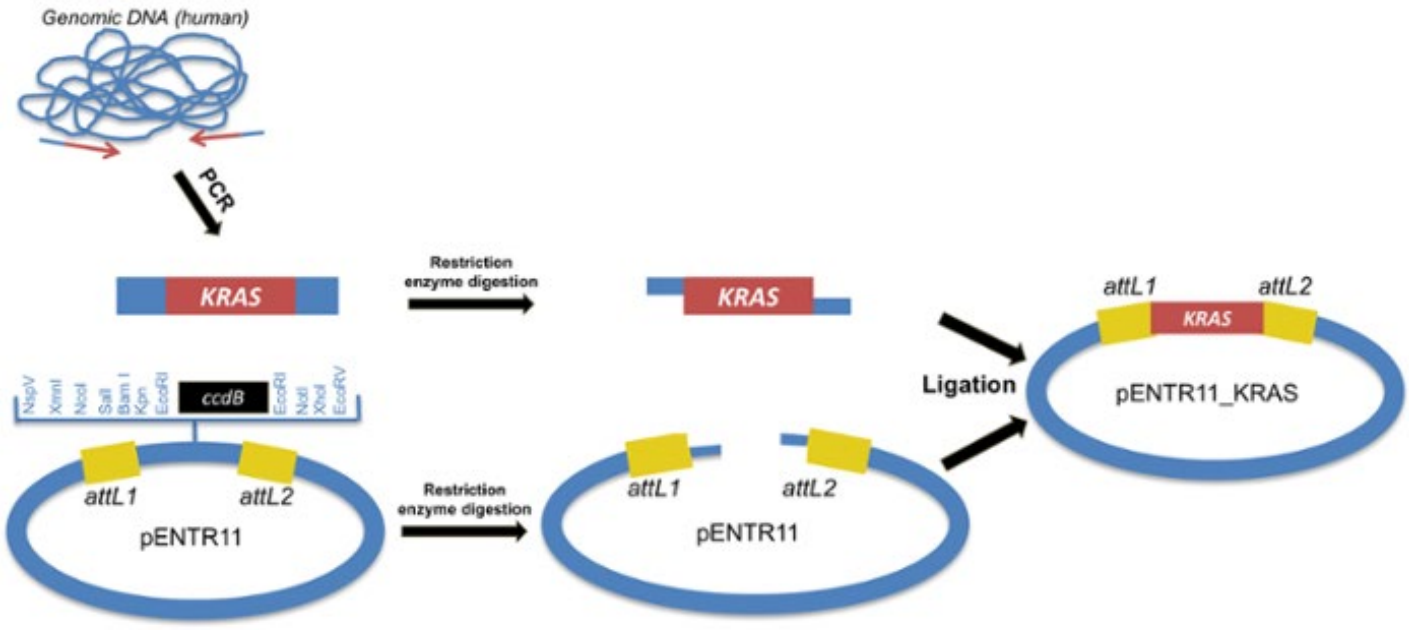


Figure 5: Method C to create an entry clone: Restriction clone the insert into an attL-containing entry vector.

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.

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.

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

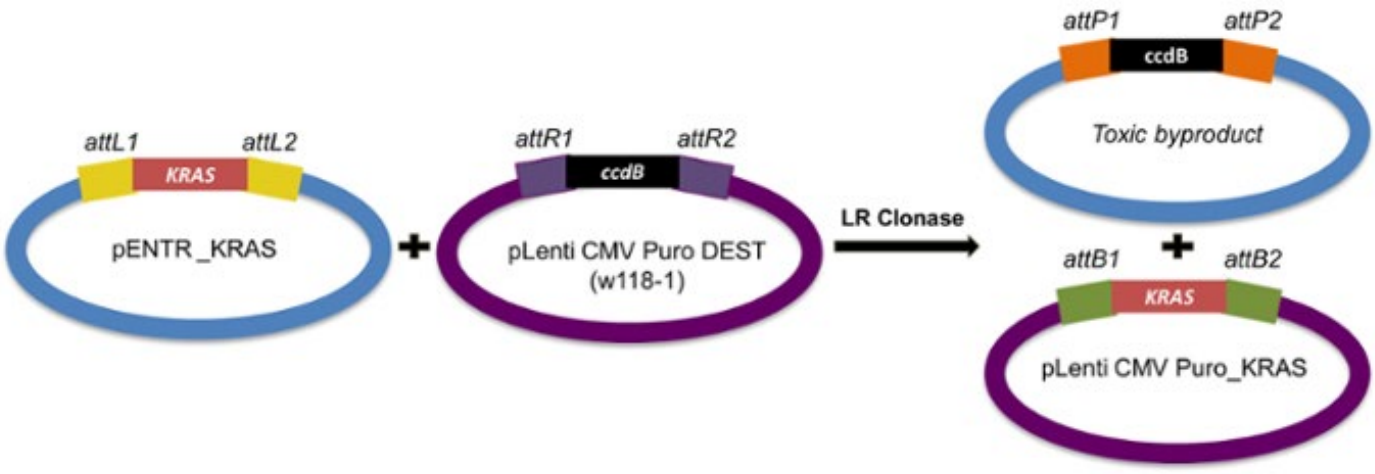


Figure 6: Generating an expression clone. The reaction between the entry clone and destination vector creates two products: the desired expression clone and a byproduct containing the ccdB gene. Since the ccdB product is toxic to the cell, Gateway cloning efficiency can reach >99%.

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 vectorsat 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 plasmids for plant expression
Gateway plasmids for worm expression
Gateway plasmids for insect expression

Table 5.2 – 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 entry clones
Entry vector	attL sites for recombination	Used to generate entry clones by TOPO cloning or by Restriction Cloning
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 expression clone

Further Reading

Chee JY, Chin CF (2015) Gateway Cloning Technology: Advantages and Drawbacks. Clon Transgen 4:138. doi:10.4172/2168-9849.1000138

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.

Ptashne, M. (1992). A Genetic Switch: Phage (Lambda) and Higher Organisms (Cambridge, MA: Cell Press).

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.

Gibson Assembly

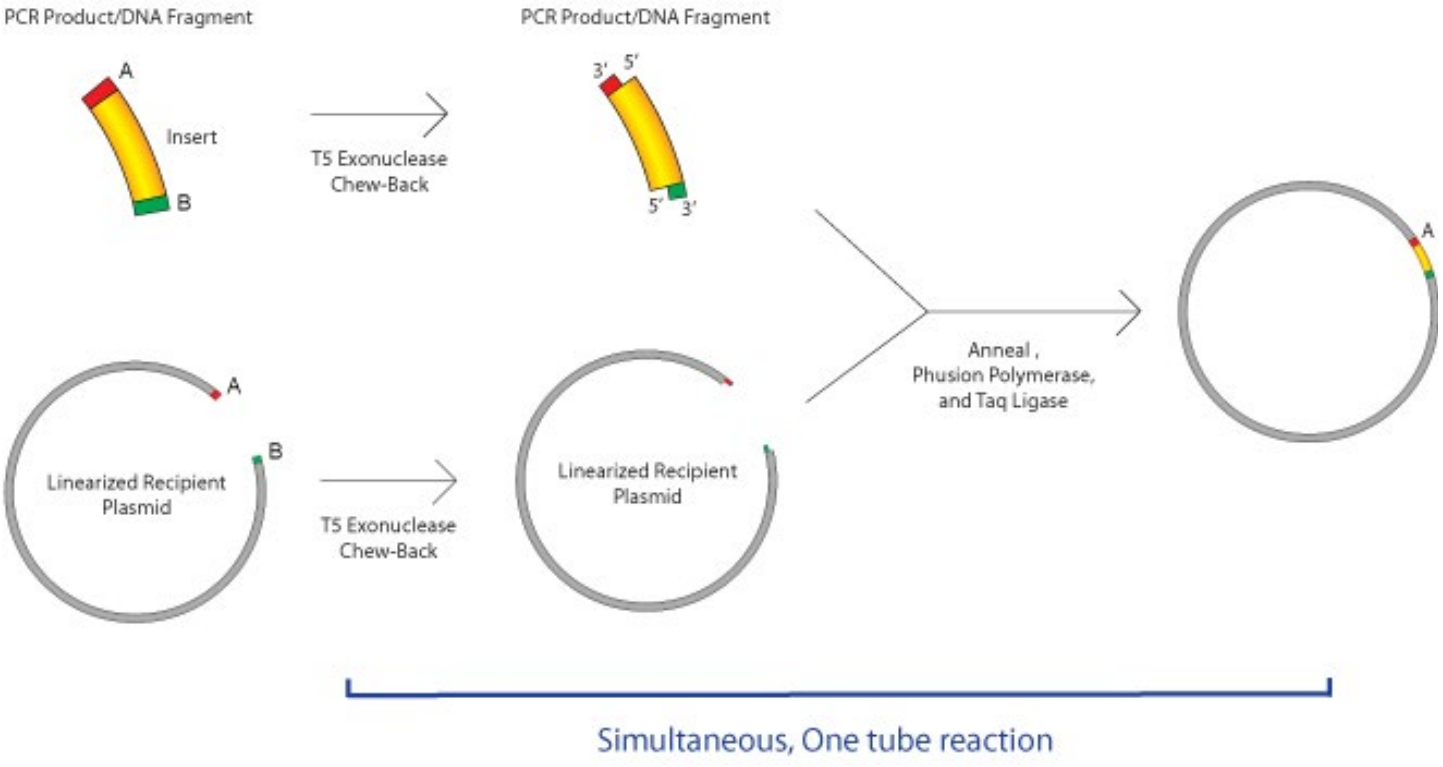
By Brook Pyhtila | March 2016

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.

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



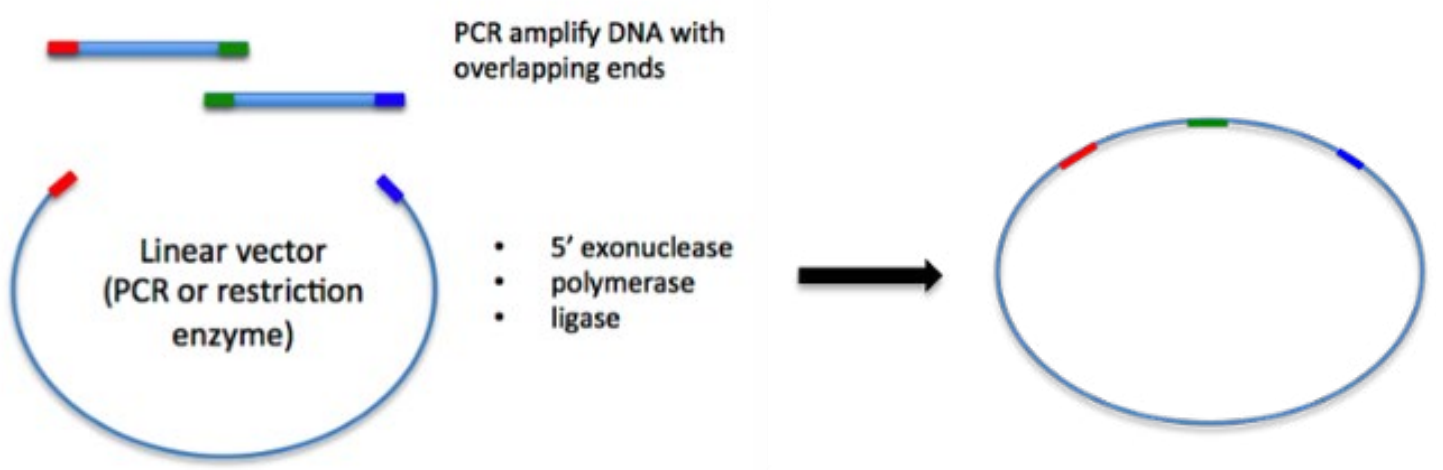
assembly of multiple linear DNA fragments, but can also be used in basic cloning of an insert into your vector of choice as shown in the figure below. To start, you need to have DNA fragments with regions of homology at their ends, which are typically created by PCR. Then, the fragments are incubated 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

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 ºC. 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 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 chapter. 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)

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 λ 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

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.

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.

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.

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.

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 6

Eukaryotic Expression Vectors

Mammalian Vectors

By Marcy Patrick | March 2014

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 discernible 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 transfection. 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 time. 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 hours 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. 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 are listed 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 hours post transfection when creating stable cell lines. ■

Table 6.1 – Commonly Used Mammalian Selection Markers

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

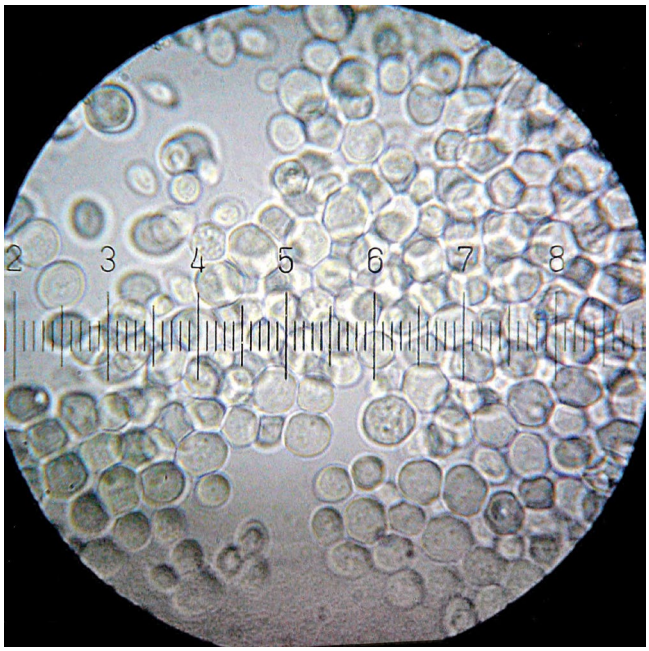
Looking for mammalian expression plasmids?

Check Out Addgene's Empty Backbone Resource.

Yeast Vectors

By Marcy Patrick | February 2014

Why do scientists use yeast vectors? 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 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 often 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.

Table 6.2 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. ■

Table 6.2 – 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?	Additional considerations
HIS3	L-hisitidine	No	S.cerevisiae	No	
URA3	pyrimidine (uracil)	Yes - Grow with 5-FOA.5-FOA.	S.cerevisiae	Yes - This can complement ura4S. pombe, but the complementation is weak.	
LYS2	L-lysine	Yes - Grow on plates containing alpha-aminoadipate in the absence of a nitrogen source.	S.cerevisiae	Yes - This can complement leu1- S. pombe, but mulitple copies are required.	
LEU2	L-leucine	HeLa, 293HEK, Jurkat T cells	S.cerevisiae	No	
TRP1*	L-tryptophan	yes - Grow with 5-FAA.	S.cerevisiae	No	TRP1 alters some yeast phenotypes. This marker should not be used in gene disruption experiments
MET15**	L-methionine and overproduces hydrosulfide ions	yes - Grow with methylmercury.	S.cerevisiae	No	Can be used for color and growth selection if divalent lead ions are used in the growth media.
ura4+***	pyrimidine (uracil)	yes - Grow with 5-FOA (see note to right).	S.pombe	No	FOA in fission yeast induces mutation in the ura5+ gene in addition to ura4+
leu1+	L-leucine		S.pombe	No	
ade6+	pyrimidine (uracil)		S.pombe	No	

**Additional considerations for* TRP1: TRP1 alters some yeast phenotypes. This marker should not be used in gene disruption experiments
** *Additional considerations for* MET15: Can be used for color and growth selection if divalent lead ions are used in the growth media.
****Additional considerations for ura4+*: FOA in fission yeast induces mutation in the ura5+ gene in addition to ura4+.

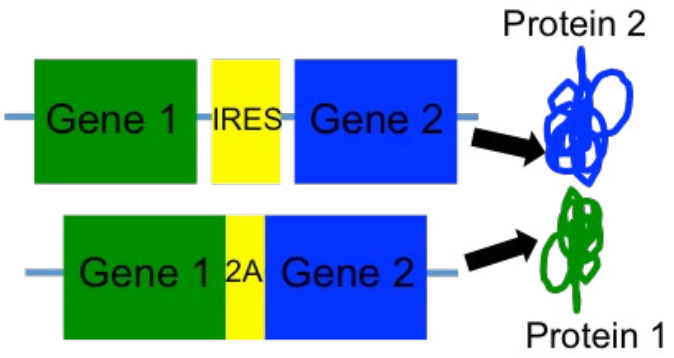
Multicistronic Vectors

By Melina Fan and Marcy Patrick | September 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. 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 above, 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 and 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.3

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.4

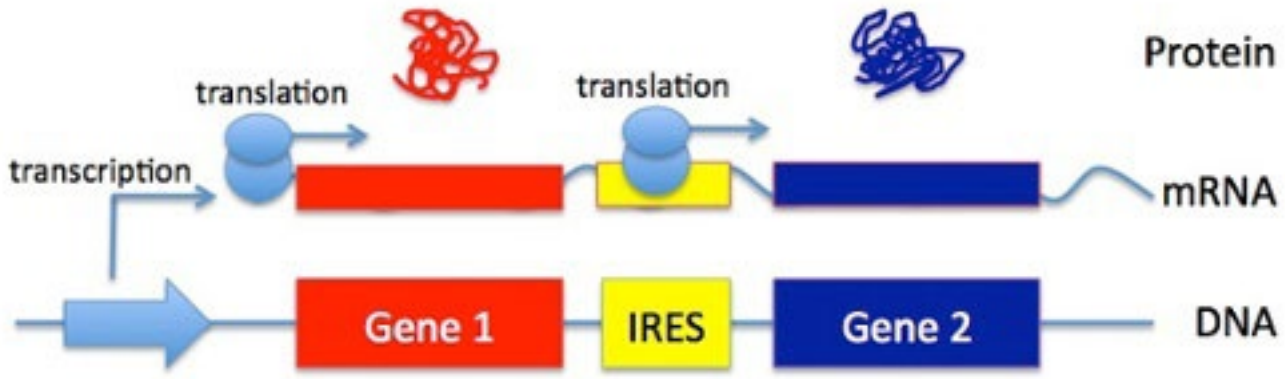


Table 6.3 – Common 2A Peptides

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.*

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 above lists the 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.

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

Table 6.3 – Some Multicistronic Vectors at Addgene

Plasmid Name	Multicistronic Element	Expression Type
MSCV-IRES-EGFP	IRES	Retroviral
pMSCV-pBabeMCS-IRES-RFP	IRES	Retroviral
pMSCV-IRES-YFP IIE2A(MCS)	IRES	Retroviral
pCMMP-MCS-IRES-Puro	IRES	Retroviral
pEF1a-IRES-Neo	IRES	Mammalian
MSCV-IRES-Luciferase	IRES	Retroviral
pWPI	IRES	Lentiviral
AmCyan-P2A-mCherry	P2A	Mammalian
pUltra	P2A and T2A	Lentiviral
Ac5-STABLE2-neo	T2A	Insect

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.

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. ■

Further Reading

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.

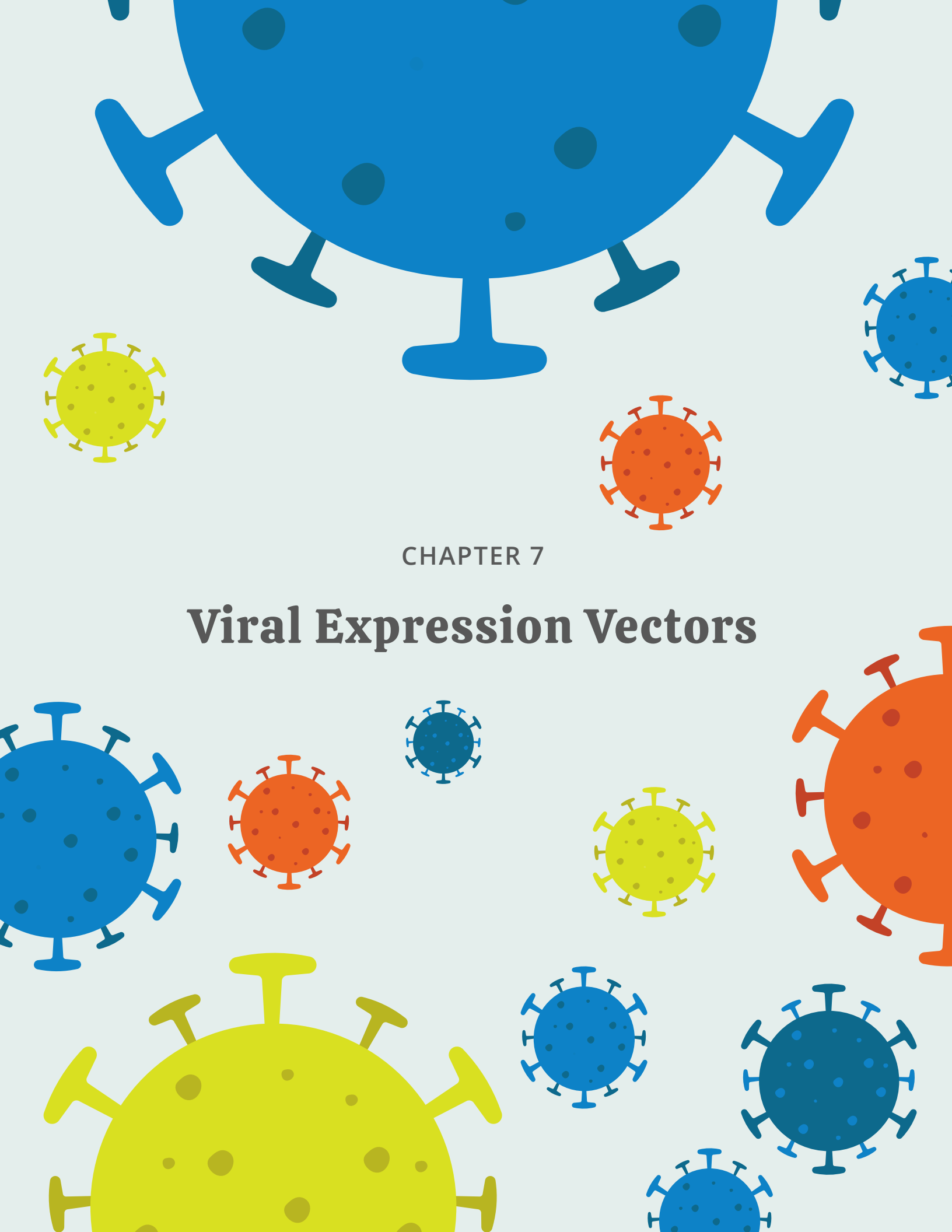
A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. Jang et al. J Virol. 1988 Aug;62(8):2636-43. PubMed PMID: 2839690.

Highly Efficient Multicistronic Lentiviral Vectorswith Peptide 2A Sequences. Ibrahimi et al. Hum Gene Ther. 2009 Aug;20(8):845-60. doi: 10.1089/hum.2008.188. PubMed PMID: 19419274.

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.

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.

Browse Addgene's Plasmid Guide for more molecular biology and cloning information.



CHAPTER 7

Viral Expression Vectors

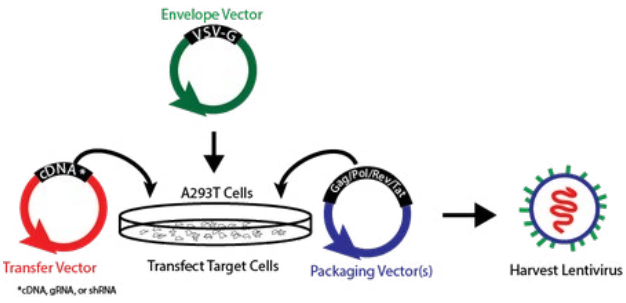
Viral Vectors: An Introduction

By Mary Gearing | August 2015

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 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 viruses. 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.

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. ■

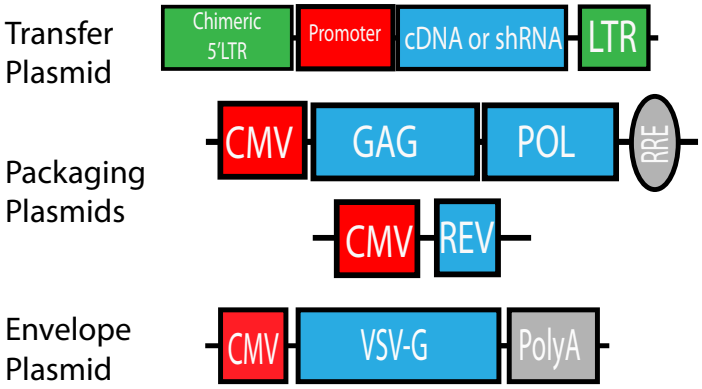
Table 7.1 - Viral Vectors

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

Viral Vector Elements

By Marcy Patrick | July 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 & 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 below list some components most commonly found in viral vector systems. ■

Table 7.2 - AAV Elements

AAV Elements	Plasmid Type	Provided?	Purpose
ITR	Cloning	*in cis*	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	*in trans*	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.
Cap	Packaging	*in trans*	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 7.3 - Gamma-retroviral and Lentiviral Elements

Elements	Plasmid Type	Provided?	Purpose
LTR	Transfer	*in cis*	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)	Packaging	*in cis*	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	*in cis*	Repeat region found within both the 5'and 3' LTRs of retro/lentiviral vectors. Tat binds to this region.
U5 (subelement of LTR)	Transfer	*in cis*	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	*in cis*	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)	*in cis*	Trans-activating response element; located in the R region of the LTR and acts as a binding site for Tat.
3' LTR	Transfer	*in cis*	Terminates transcription started by 5' LTR by the addition of a poly A tract just after the R sequence.
cPPT	Transfer	*in cis*	Central polypurine tract; recognition site for proviral DNA synthesis. Increases transduction efficiency and transgene expression.
Psi (ψ)	Transfer	*in cis*	RNA target site for packaging by Nucleocapsid

Further Reading

CHECK OUT VIRAL VECTORS AT ADDGENE:

Addgene's most popular lentiviral plasmids
Addgene's most popular retroviral plasmids
Addgene's collection of AAV plasmids

LEARN MORE:

Overview of Lentiviral Packaging Plasmids
Lentiviral Protocols and Resources
More Lentiviral FAQs Answered by Addgene
Didier Trono Lab: Lentivectors Toolbox

CHAPTER 8

Plasmids That Glow

History of Fluorescent Proteins

By A. Max Juchheim | 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.

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. 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. But the first time anyone truly examined what caused this glowing phenomenon wasn't until the 1960s, when Osamu Shimomura studied the bioluminescent properties observed in the crystal jellyfish Aequorea victoria 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 photo organs; 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! ■

Which Fluorescent Protein Should I Choose?

By Gal Haimovich of greenfluorescentblog | May 2014

Fluorescent protines(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.*

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 a 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.

Table 8.1 - Fluorescent Protein Colors

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

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

Fluorescent sensors

These FPs change their excitation/emission behavior upon environmental changes (e.g. pH, Ca2+ flux, etc). The most commonly used are GECIs - genetically encoded calcium indicators (e.g. GCaMP). Others include: pHluorin & pHTomato (pH sensors), HyPer (H2O2 sensor), ArcLight (voltage sensor), and iGluSnFr (glutamate sensor). More examples of these biosensors can be found at Addgene.

EIGHT POINTS TO KEEP IN MIND WHEN CHOOSING AN FP

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 488 nm and 561 nm, 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).

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).

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 environment. Recently, a new GFP isolated from the Unagi eel was shown to mature independently of oxygen, making it suitable for use in anaerobic conditions.

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.

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 GFPS65T is better suited for yeast studies (24—30 °C).

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.

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 100 ms (EBFP) or as long as 1 hour (mAmetrine1.2). However, for most FPs it is a few seconds to a few minutes.

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).

Keep this list handy to help you 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 green fluorescent blog. ■

Further Reading

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.

Fluorescent proteins and their applications in imaging living cells and tissues. Chudakov et. al. (2010). Physiol. Rev. 90:1103. PubMed PMID: 20664080.

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.
FP guide at Addgene

Fluorescence Spectrum Viewer from BD

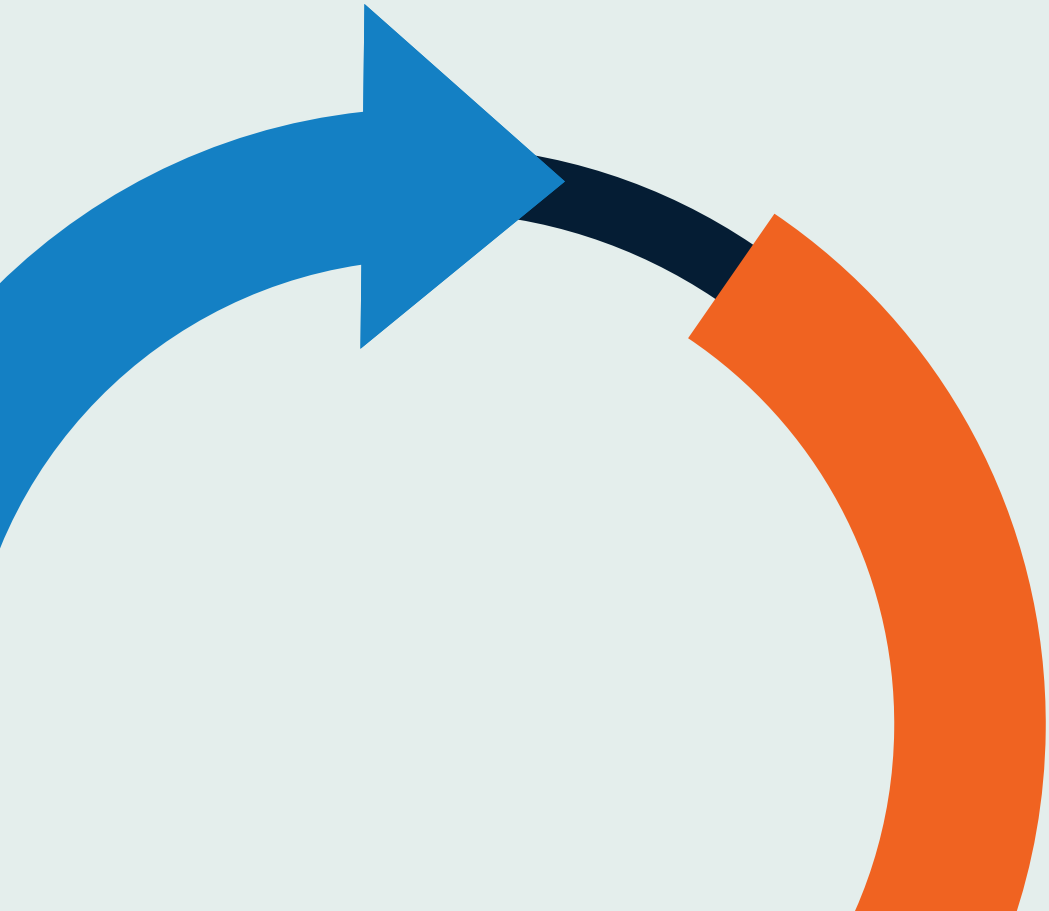
Bioscience Fluorescence SpectraViewer at Invitrogen (Life Technologies) site Evrogen

Spectra Viewer at Evrogen.

ilovegfp – a site with very comprehensive data sheets on many FP variants

CHAPTER 9

Plasmid Tags



Protein Tags

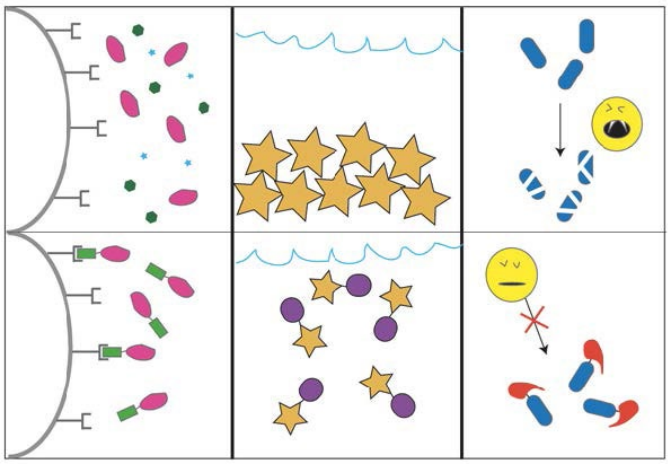
By Eric J. Perkins | December 2014

Protein tags are usually smallish peptides incorporated into a translated protein. They have a multitude of uses including (but not limited to) purification, detection, solubilization, localization, and 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 number of molecular tool box tags and fusion proteins to aid in the expression and purification of recombinant proteins.

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 being spent to make something the cell doesn't need to make. Eukaryotes and some bacteria deploy proteosomes to degrade what the cell might consider junk protein. Though there are a number of chemical and peptide-based proteosome 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.



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

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 your protein 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, immunoprecipitation, or purifying that protein may be to fuse an affinity tag to it. The FLAG, hemaglutinin 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 Ni2+ 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

Table 9.1 – Common Protein Tags

Tag	Epitope	Mass (kDa)	Function	Notes
CBP	KRRWKKNFIAVS AANRFKKISSSGAL	4	Affinity and Purification	Binding and elution steps use very moderate buffer conditions
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 solubililty
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)
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
S-tag	KETAAAKFERQHMDS	1.8	Solubility and Affinity	Abundance of charged and polar residues improves solubility; good for antibody-based detection

Table 9.1 – Common Protein Tags (cont.)

Tag	Epitope	Mass (kDa)	Function	Notes
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 AKKLNDAQAP KVDNKFNKEQ QNAFYEILHL PNLNEEQRNA FIQSLKDDPS QSANLLAEAK KLNDAQAPKV 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	GKPIPNPLLGLDST	1.4	Affinity and Purification	Good for antibody-based purification

the length of a polyhistidine tract does not greatly alter the size of the tag.

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 tags than protein), 2-3 tags is increasingly common. Tandem affinity purification (TAP) once referred specifically to a combo tag

composed 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 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^X, 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 9.2 - 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 categoriezed to help you better assess their function. More detailed information and some protocols can be found in the references provided. ■

Further Reading

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.

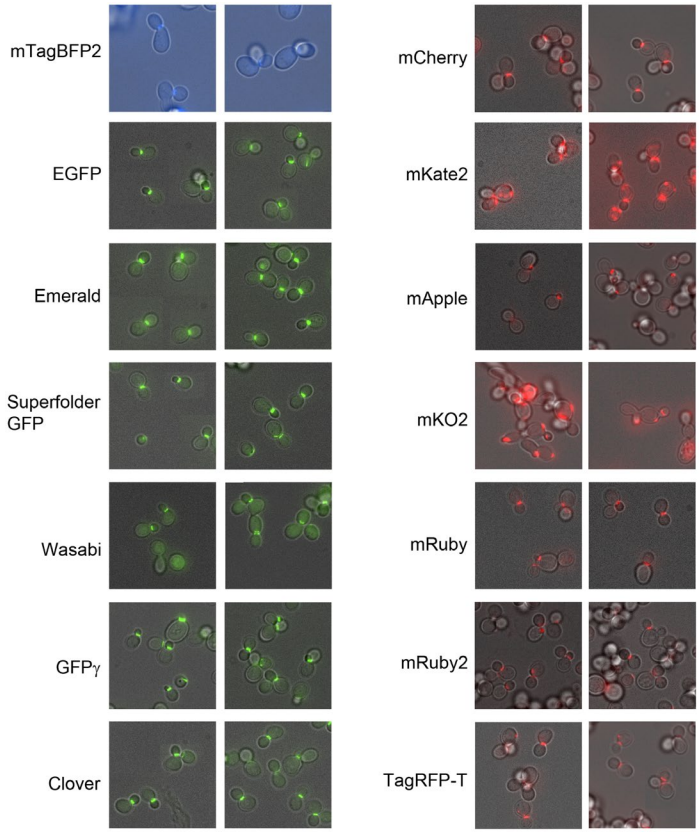
A complete list of free handbooks for protein purification is provided by Cytiva Life Sciences.

Tag Your Favorite Yeast Genes With Ease

By Julian Taylor-Parker | November 2013

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!



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

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 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.

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.

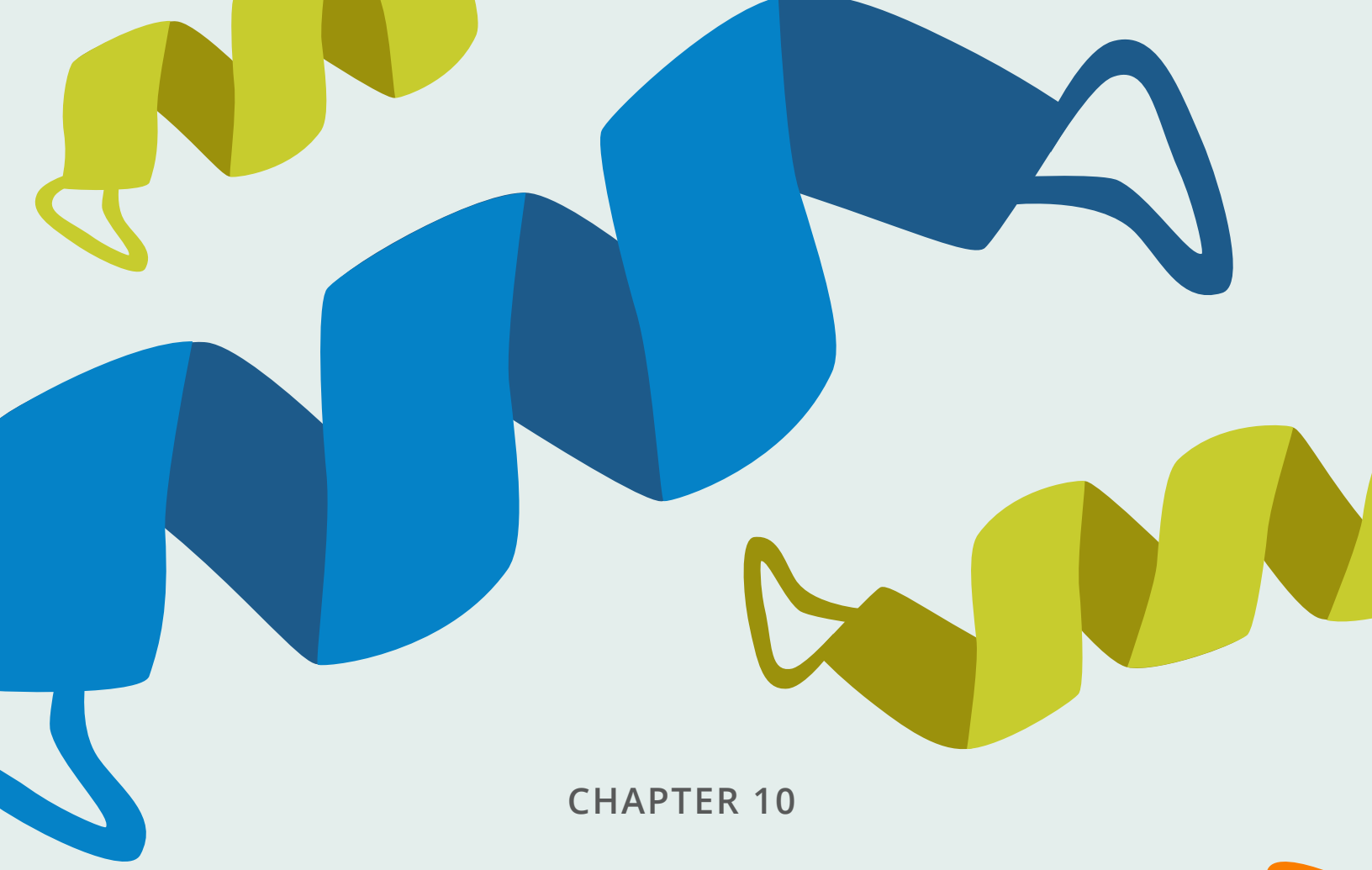
In addition to the collections featured above, many other modular yeast tagging systems have been developed in the labs of **Anne Robinson**, **Eishi Noguchi**, and **Melissa Moore**, to name a few. ■

Further Reading

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. PubMedPMID: 23844123.

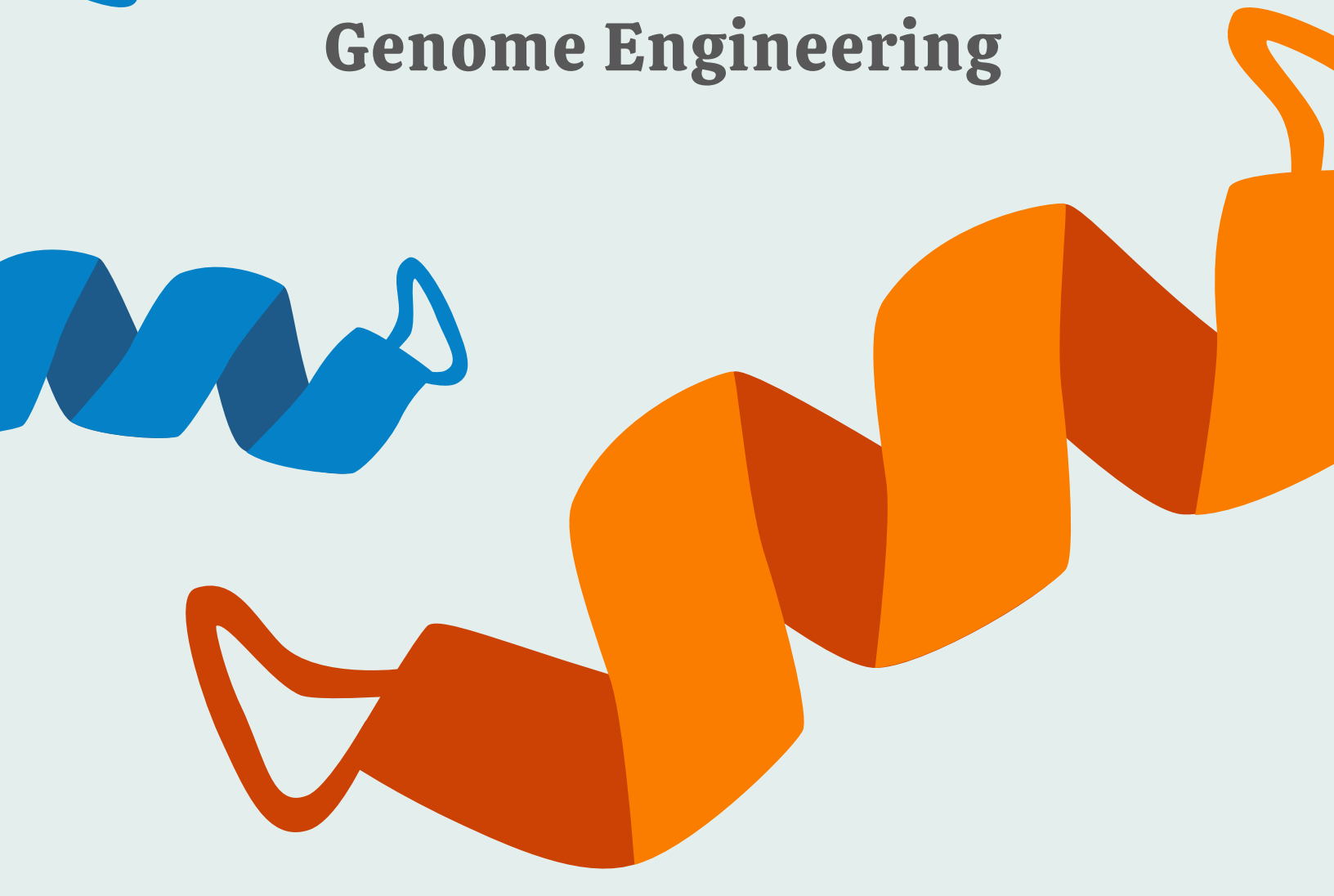
Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Bähler J, Wu JQ, Longtine MS, Shah NG, McKenzie A 3rd, Steever AB, Wach A, Philippsen P, Pringle JR. Yeast. 1998 Jul;14(10):943-51. PubMed PMID: 9717240.

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, Brachat A, Philippsen P, Pringle JR. Yeast. 1998 Jul;14(10):953-6. PubMed PMID: 9717241.



CHAPTER 10

Genome Engineering



Introduction to Genome Engineering

By Mary Gearing | August 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.

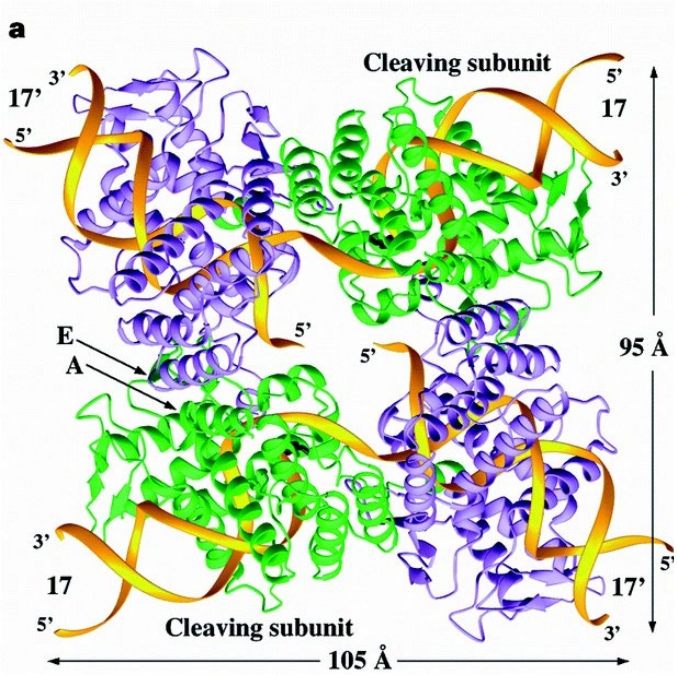
Genome engineering was born in the late 1970s, when multiple groups showed that exogenous DNA could be taken up by yeast or bacteria and randomly integrated into the genome. Subsequent work showed that this process could also occur in a targeted fashion. Addgene depositor Mario Capecchi realized that 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-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.



Cre-recombinase complexed to DNA Guo et al. 199

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 non modular 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. TALEN technology was rapidly adopted by the research community. 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.

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.

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

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

Addgene's Genome Engineering Guide

Cre-Lox

By A Max Juchheim | January 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 life cycle, and researchers have adapted the components for use in genome manipulation.

Cre recombinase, originally named because it "causes recombination" (although later referred to as the "cyclization recombinase"), 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.

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 sequence on a single loxP site is recognized and bound by a Cre protein, forming a dimer. The two loxP sites then align in a parallel orientation,

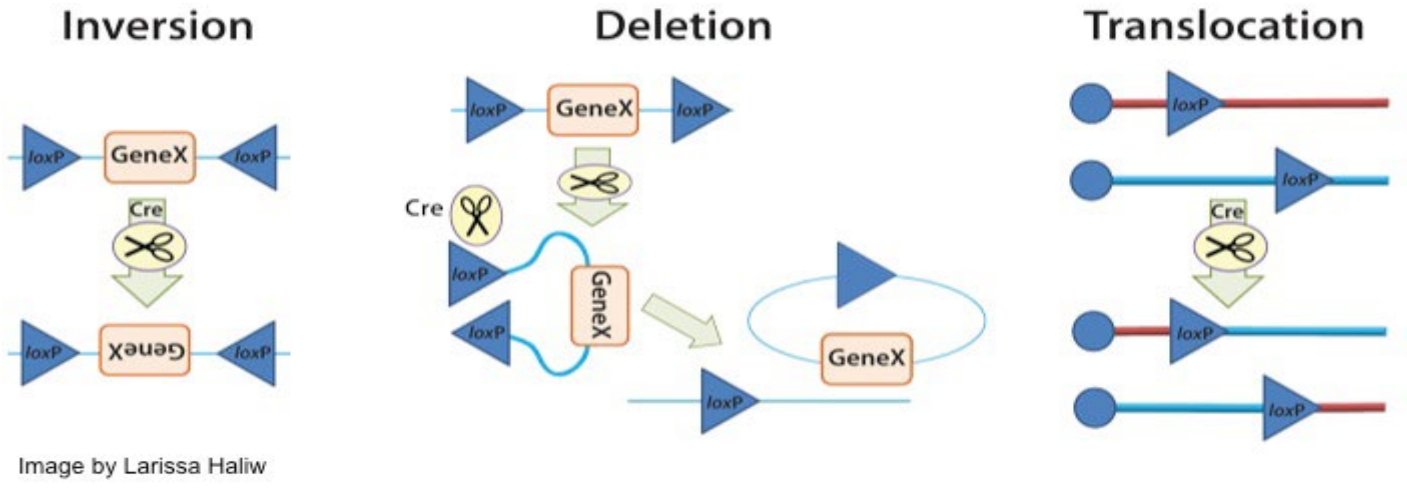


Image by Larissa Haliw

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.

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

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 "flanked by loxP"), will permit gene expression until Cre is present, at which time the gene will be disrupted or deleted.

Selection marker removal

In conventional mouse targeting, targeted clones are selected for using a resistance 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 specific 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.

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. ■

Further Reading

Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. Sternberg, N. and Hamilton,
D. 1981. PubMed PMID: 6276558.

Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Sauer, B. and
Henderson, N. 1988. PubMed PMID: 2839833.

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.

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.

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.

Addgene's Cre-Lox Guide

Knockout/Knock-In Plasmids

By Benoit Gique | December 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 a homologous sequence using a specially designed vector that contains sequence homology to the

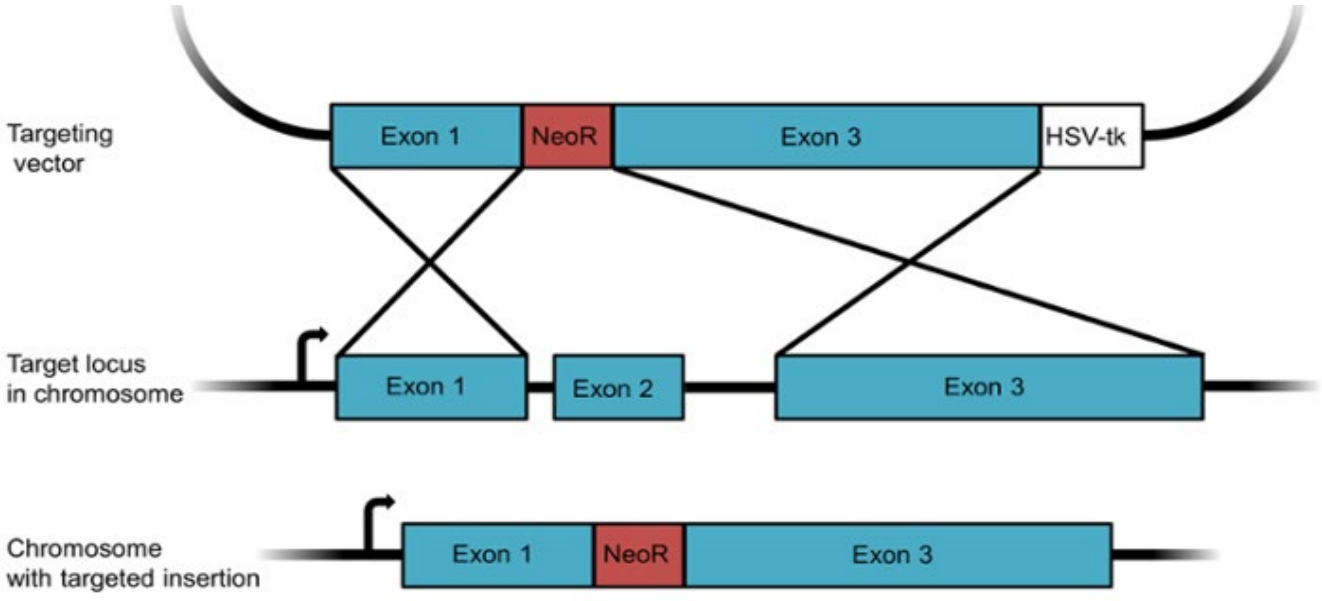


Figure 1: A knockout targeting vector designed to insert a resistance gene. The vector contains a neomycin resistance gene (NeoR) flanked by homology arms. The negative selection marker HSV-tk is used to select against random recombinants.

locus of interest. To give you an idea of the process, we'll walk through an experiment designed to knock exon 2 out of a given gene.

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 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).

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 in Figure 3, 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

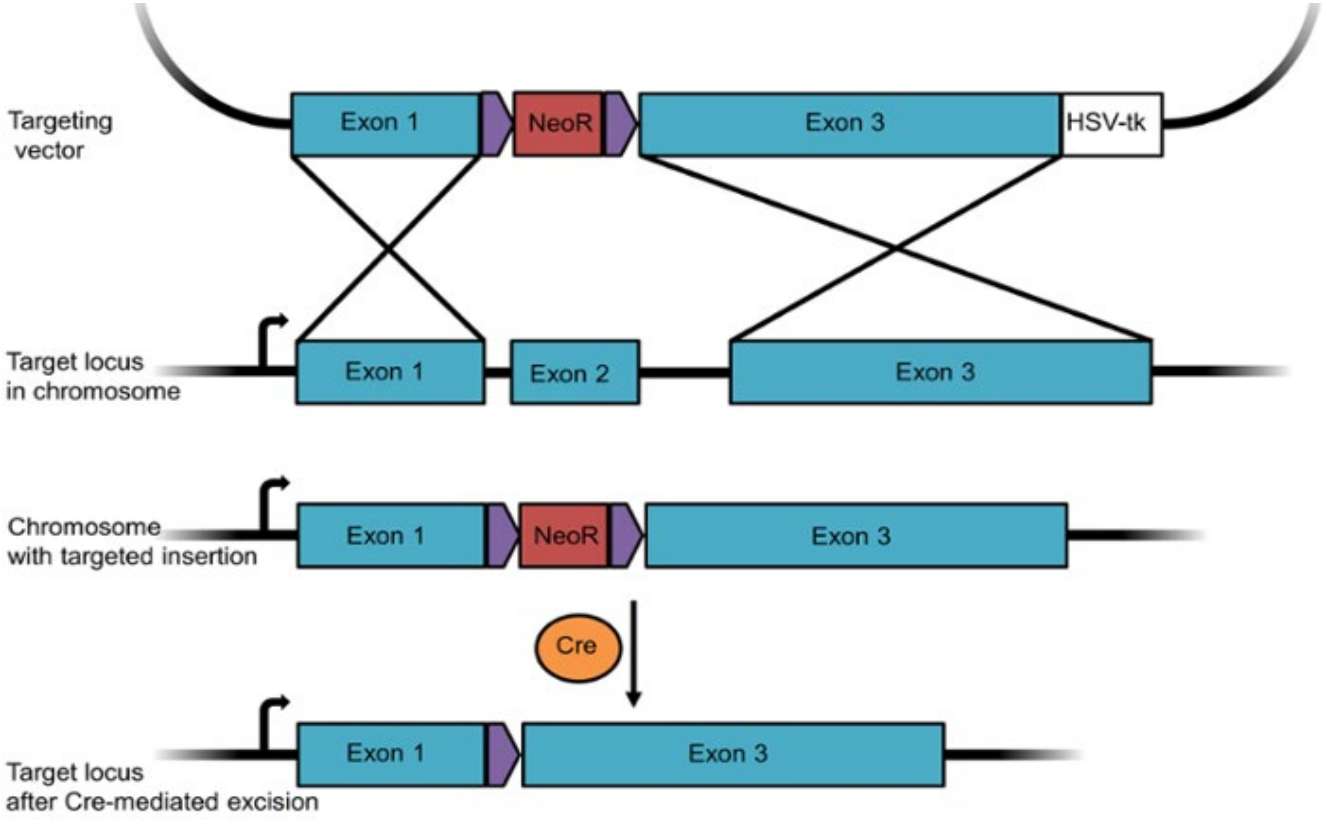


Figure 2: A knockout targeting vector designed to insert and later remove a resistance gene. In this construct, the neomycin resistance gene is floxed so that it can be removed post-selection through Cre-mediated recombination.

GFP is expressed (knocked in). As seen in Figure 3, you can remove a floxed resistance gene using Cre recombinase.If GFP is under control of an endogenous promoter, you can use expression GFP 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.

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.

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 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 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

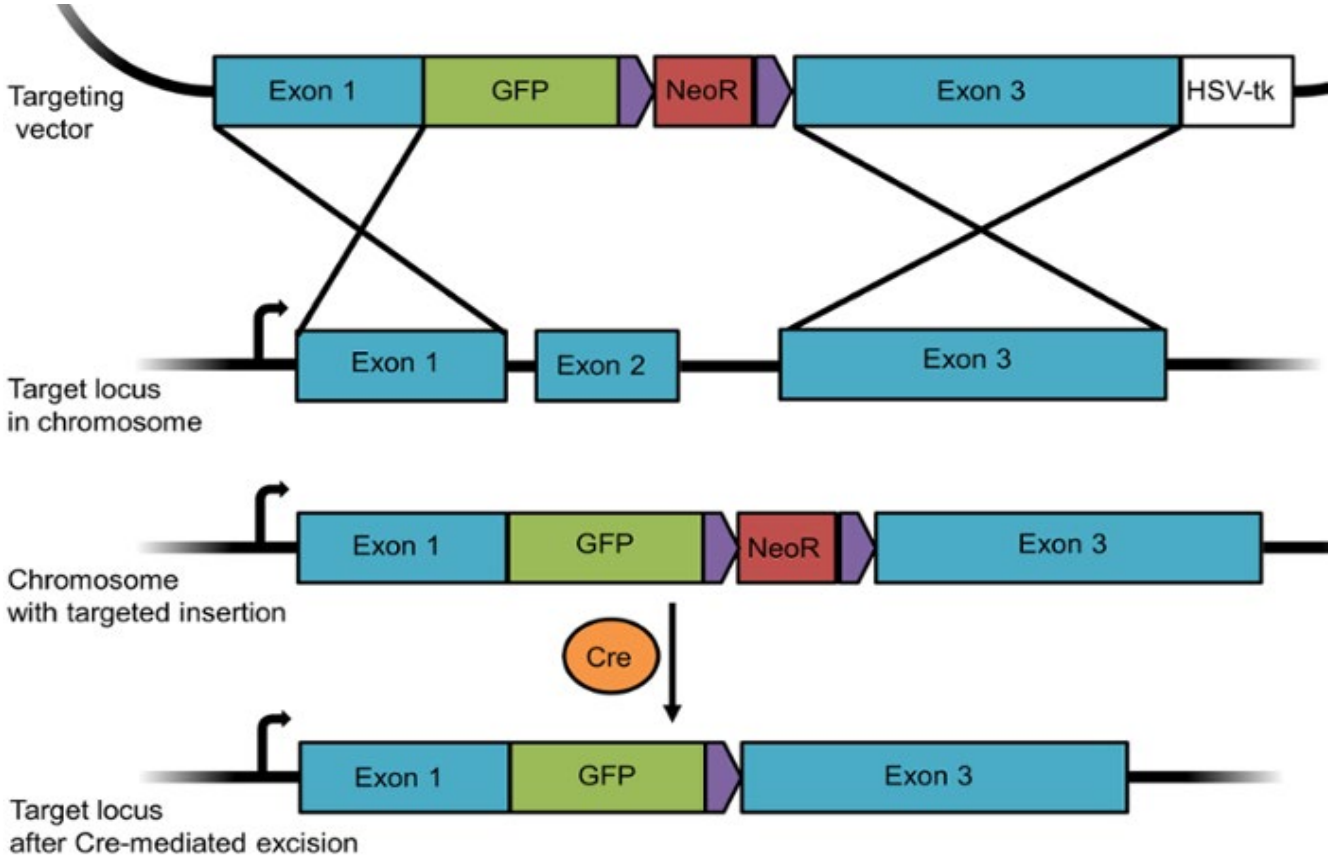


Figure 3: A knockout/knock-in targeting vector designed to insert GFP into a given locus.

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

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.

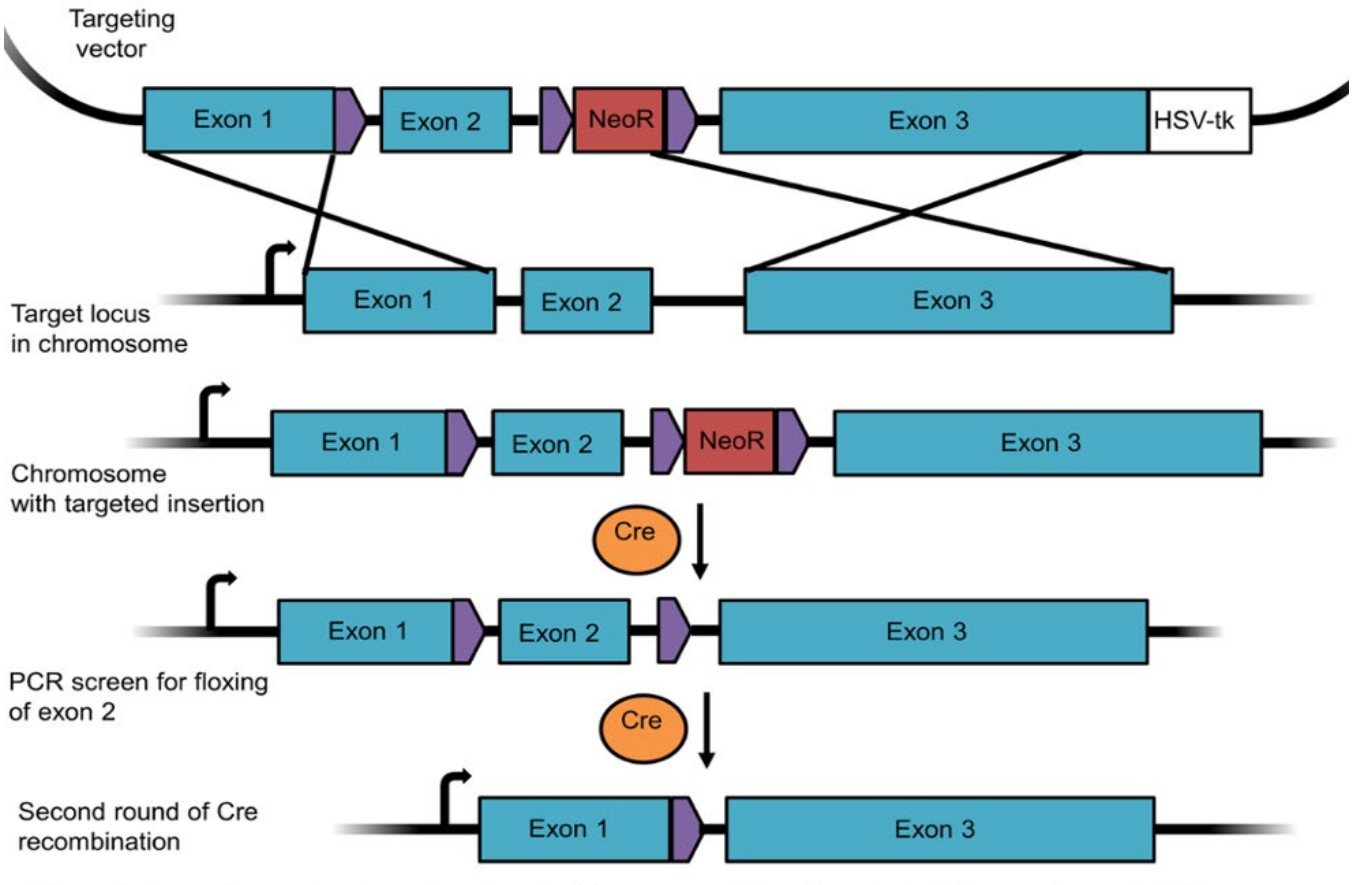


Figure 4: A targeting vector to create a floxed allele for a conditional knockout. This experimental design requires two rounds of re-combination - the first removes the selection marker, the second the floxed exon.

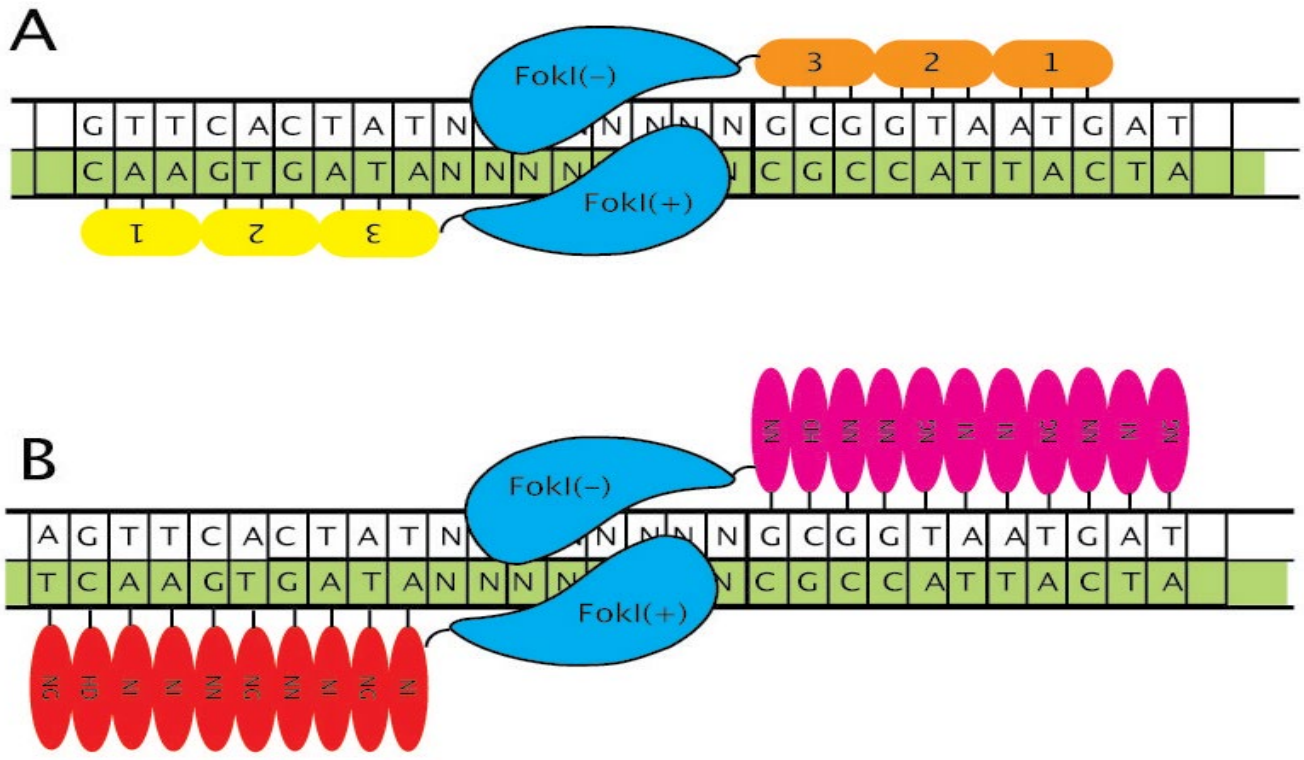
Overview of TALEN Technology

By Eric J. Perkins | December 2011

Zinc fingers were once the go-to technology for targeting enzymes and other useful protein domains to a specific DNA sequence, but the arrival of TAL effector and CRISPR based technologies permanetly changed the genome-editing game.

In 2009, groups led by Jens Boch at the Martin-Luther-University Halle-Wittenberg, and Adam Bogdanove at the 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. Ultimately, it is this RVD that determines which single nucleotide the TAL effector (TALe) will recognize. In Figure 1B, you can see that 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 35 bp.

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 Golden Gate cloning 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 Voytas/Bogdanove TALEN kit came online in the summer of 2011. 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

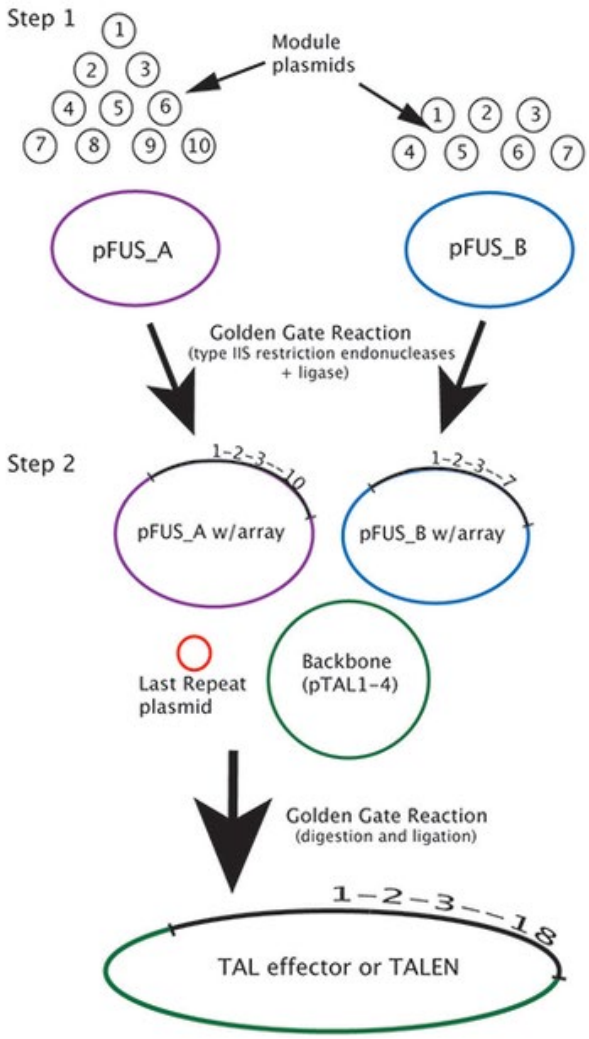


Figure 2: Simplified representation of the Voytas/Bogdanove Golden Gate TALEN kit. In this example, the final array is composed of 18 TAL effector repeats. The kit can be used to make arrays from 12-31 repeats.

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.

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 rapidly became an important resource for TALEN 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.■

Further Reading

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.

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.

TAL effectors: customizable proteins for DNA targeting. Bogdanove AJ, Voytas DF. Science 2011 Sep 30;333(6051):1843-6. PubMed PMID: 21960622.

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.

TALEs for the masses. Rusk N. Nat Methods 2011 Mar;8(3):197. PubMed PMID: 21473016.

Move over ZFNs. Nat Biotechnol. 2011 Aug 5;29(8):681-4. PubMed PMID: 21822235.

See the Q&A with Dr. Adam Bogdanove and his thoughts on the TALe technology here.

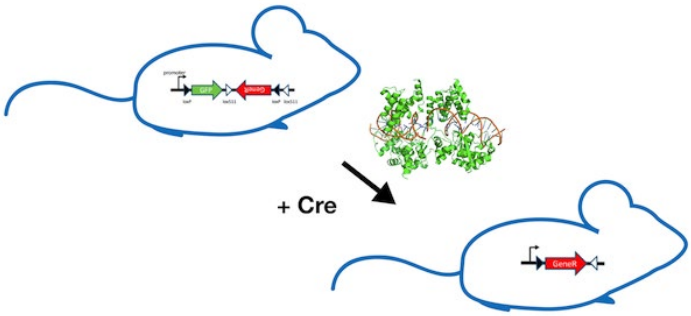
Flex Vectors

By Michelle Cronin | April 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. Today's post 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 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.

Once the SSR of choice is introduced, recombination can proceed either by first utilizing the loxP sites or the lox551 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 FLEX USEFUL?

The power of the FLEx system lies in the ability to conditionally turn off one gene and activate another. In the field of mouse genetics, FLEx is

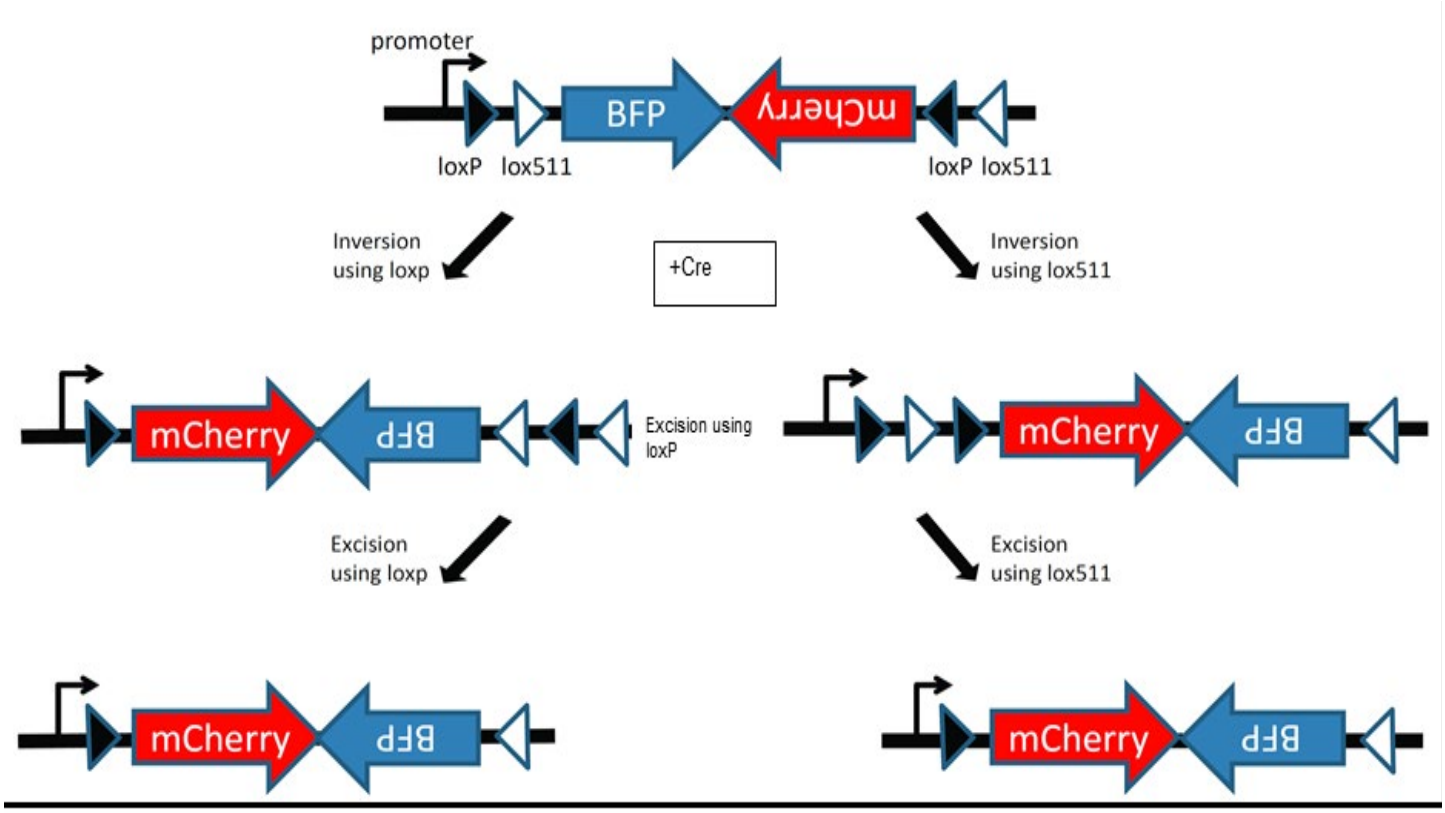


Figure 1 (modified from Schnutgen et al. Nat Biotech 2003): Basic FLEx switch will express BFP in the absence of, or mCherry in the presence of Cre recombinase.

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 FLEx 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 FLEx switch conditional knock-in cassette. FLEx switch using Cre to activate gene expressionAs 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 rescue 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.

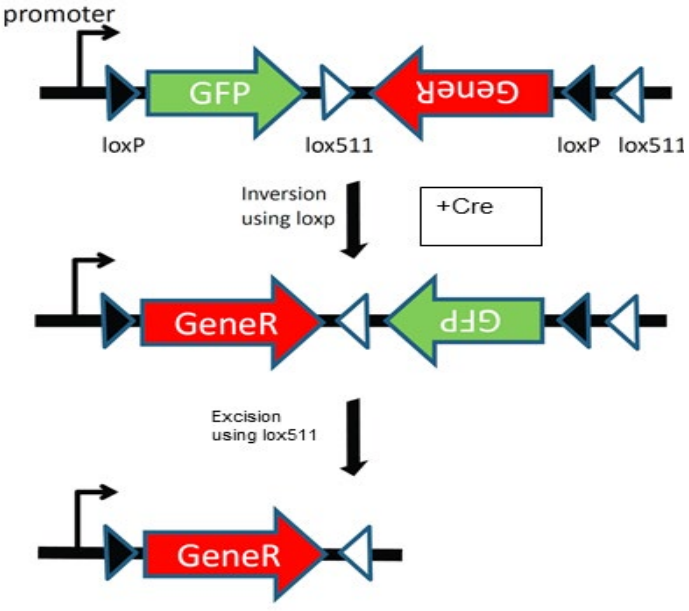


Figure 2. In the presence of Cre, GFP expression is lost, allowing for the rescued expression of GeneR.

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 cells. ■

Further Reading

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.

Branda, CS and Dymecki, SM. Talking about a Revolution: The Impact of Site-Specific Recombinases on Genetic Analyses in Mice. 2004. PubMed PMID: 14723844.

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

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 Michelle Cronin | April 2016

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 (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.

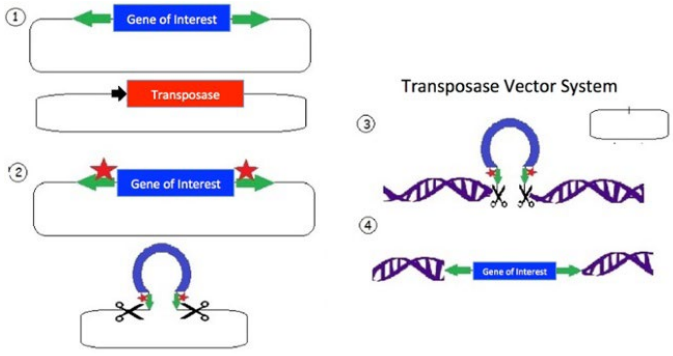


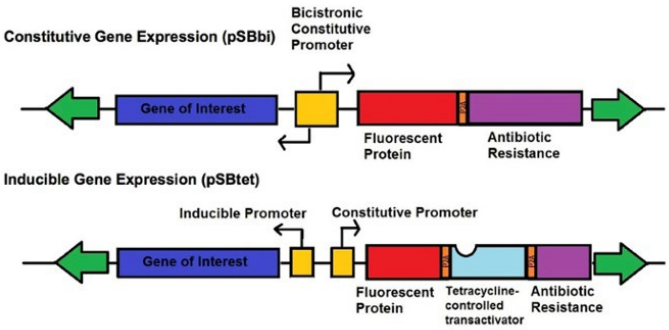
Figure 1 – (1) The transposon consists of a mirrored set of inverted repeats (green) flanking the gene of interest (blue) in a plasmid backbone. A separate plasmid contains the transposase gene (red) for expressing the transposase enzyme. (2) The transposase enzyme is expressed (red stars) and binds 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 composed 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 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 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.

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

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.

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.

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 11

Verifying Your Plasmid

How to Verify Your Plasmid

By Lianna Swanson | August 2014

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.

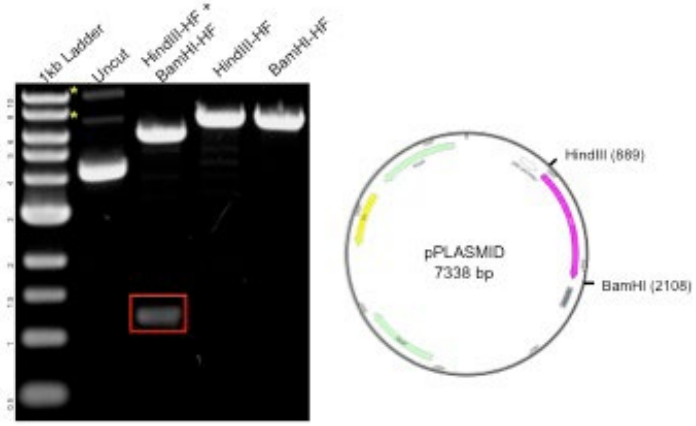
DIAGNOSTIC RESTRICTION DIGEST

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.

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:

Use unique enzymes

Enzymes that only cut once allow you to more easily and accurately visualize the full size of your construct.

Buffer and temperature compatibility

When digesting with more than one enzyme, make sure they are compatible in their temperature and buffer requirements. Consult the manufacturer's manual for the optimal working conditions for each enzyme.

Methylation

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:

Ethidium bromide (EtBr)

Add EtBr to your gel before pouring it (while it is still liquid). 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!

Loading buffer

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 you can estimate how far down the gel your bands have traveled 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.

Run controls

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.

Gel running speed

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

Visit Addgene's resource for plasmid verification to find additional tips and detailed protocols

Learn how to set up your digests

Learn how to pour and run a DNA gel

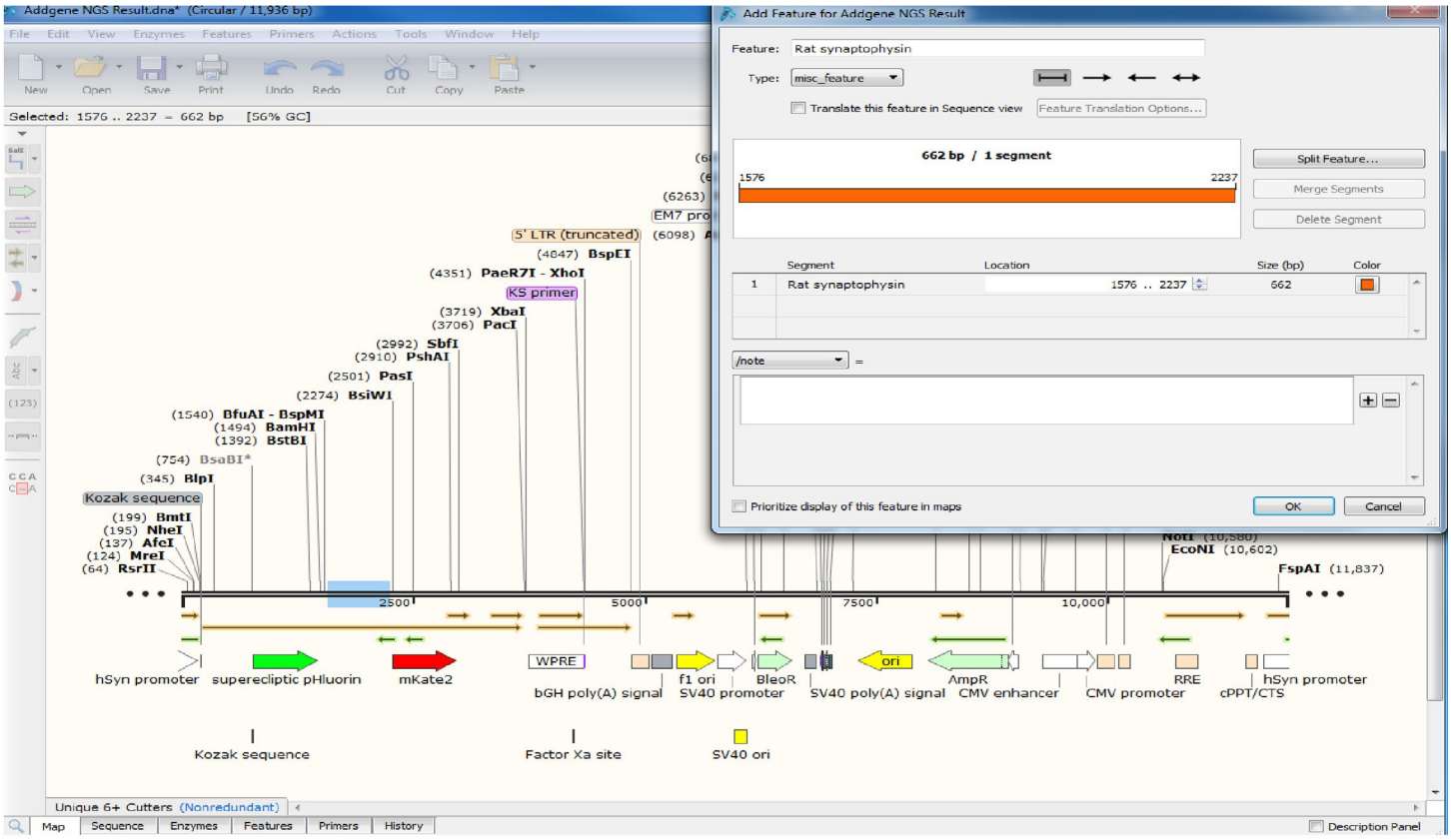
Using SnapGene

By Amanda Hazen | February 2018

Here at Addgene, we regularly use Snapgene for our quality control process because of its expansive feature library and useful tools. In this section, we'll walk you through how Addgenies use SnapGene to confirm the sequence of a plasmid and we'll highlight some of the new features available on our website through our SnapGene powered maps and sequence analysis tools In this section, we'll walk you through how Addgenies use Snapgene to confirm the sequence of a plasmid and we'll highlight some of the new features available on our website through our Snapgene powered maps and sequence analysis tools.

The main ways we use SnapGene for plasmid quality control are:

(1) To identify important plasmid features.
(2) To align next-generation sequencing (NGS) results with reference sequences.
(3) To design experiments for further plasmid verification.



Adding a custom feature in Snapgene to annotate the rat synaptophysin gene within the insert in Plasmid 70111

IDENTIFYING PLASMID FEATURES

While our quality control process for incoming plasmids can vary based on the contents of a plasmid, the first step is always to send prepped plasmid DNA for NGS. We then copy the contents of the resulting FASTA file into the desktop version of Snapgene to visualize the sequence and look for features. The first thing that the SnapGene software will confirm for us is whether it is a linear or circular sequence. Next, we use the "Detect Common Features" function to find and annotate promoters, selectable markers, tags, origins of replication, and fluorescent proteins that are already included in SnapGene's extensive database.

SnapGene also provides easy visualization of open reading frames, which is often our first clue about where the gene of interest might be located. By easily using BLAST to check the translated amino acid sequence from within SnapGene, we can confirm the identity of the gene and check for any mutations.

Once we identify the gene, we can easily annotate it by making a custom feature in SnapGene.This saves us time, as we are often analyzing deposits with many plasmids that contain similar features and it allows us to more quickly identify the same feature in other plasmids. This information is used to **generate the plasmid maps** you see on our plasmid pages.

ALIGNING SEQUENCES WITH SNAPGENE

Once we've identified the features within a plasmid, we next verify that **our sequencing results match sequence information** contributed by the depositor, published backbone sequences, and any known sequences for the insert. The SnapGene desktop software allows us to quickly align multiple sequences at the same time. This means we can easily compare and analyze any differences between our sequence results and the reference sequences.

USING SNAPGENE FOR ADDITIONAL VERIFICATION EXPERIMENTS: PRIMER AND RESTRICTION DIGEST DESIGN

Some regions of plasmids are difficult to sequence by NGS. These can include GC-rich and repetitive regions. These regions require additional verification experiments that can be partially designed using the SnapGene desktop software.

If a plasmid contains a region with a high amount of guanine and cytosine nucleotides, we might need to send it for Sanger sequencing to verify the sequence within that region. Here at Addgene, we have designed many primers for Sanger sequencing. By using SnapGene's "Import primers from a list" function, we can detect all of the primers we have designed that will anneal to the plasmid sequence. We can then quickly choose which primers we should use to further verify the difficult regions and send the plasmid off for Sanger sequencing.

If a repetitive region is present in the plasmid, we might need to perform a restriction digest to confirm the size of the plasmid and the insert. By using the "Simulate agarose gel" option with the desktop version of SnapGene, we can predict the

resulting band sizes when digesting the plasmid sequence with one or more enzymes.

We've worked with SnapGene to update our website and enable you to analyze sequencing results for plasmids in the repository at a glance. Our SnapGene-powered plasmid maps and sequences use the same feature detection software as the desktop version of SnapGene.

When you download a SnapGene file directly from our website, you can open it with either the SnapGene viewer or SnapGene Desktop programs to analyze the sequence data in depth. You will see the same features annotated in our plasmid maps, including our most popular primers and certain restriction enzymes, but you will also be able to create custom annotations and design additional experiments like restriction digests. ■

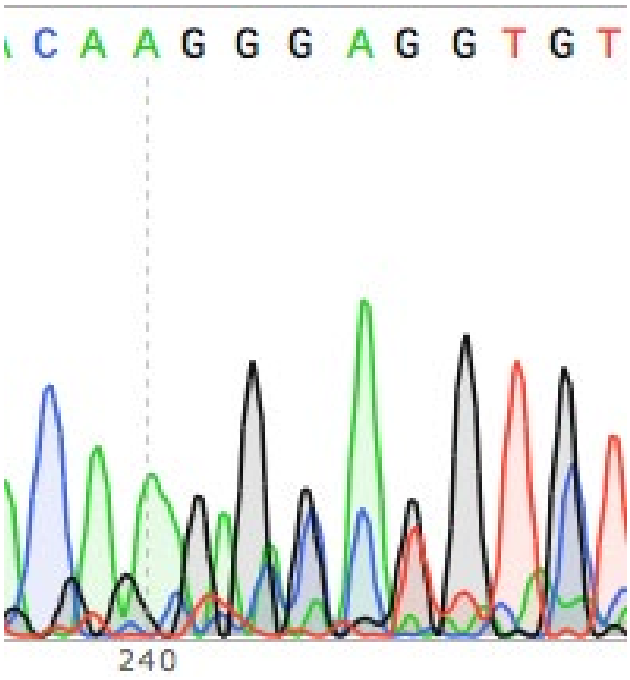
Six Tips for Analyzing and Troubleshooting DNA Sequencing Results

Originally published on BitesizeBio. By Lianna Swanson | June 2014

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.



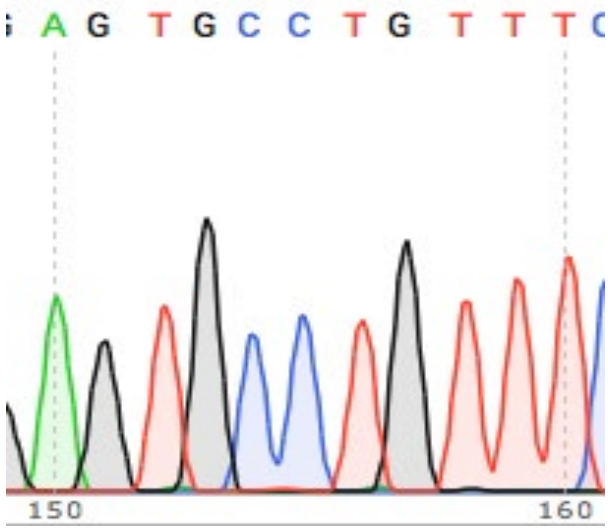
TROUBLESHOOTING AND ANALYSIS

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

1. 4Peaks (Mac)
2. SnapGene Viewer (Mac/PC)
3. FinchTV (Mac/PC)
4. Sequence Scanner (PC)
5. Chromas (PC)

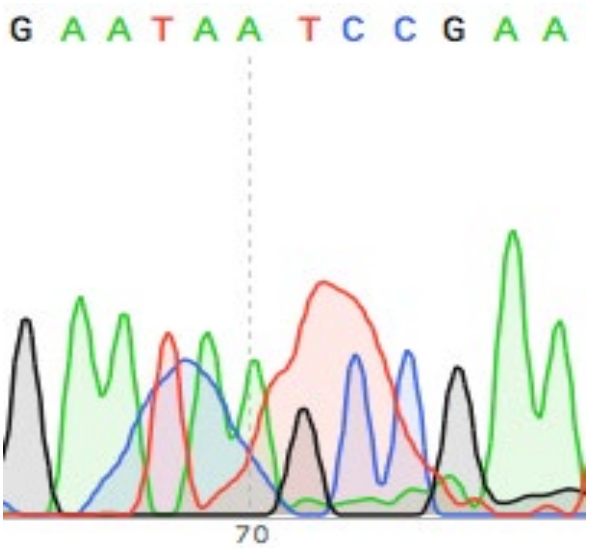
You should see individual, sharp and evenly spaced peaks

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



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.

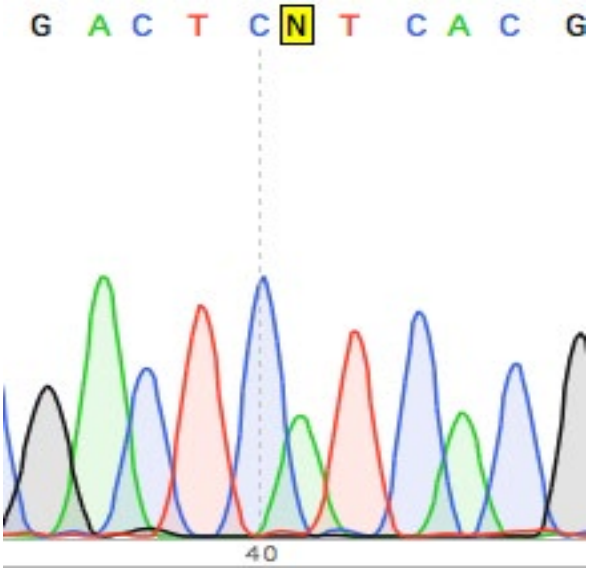


Never trust the first 20-30 bases of a DNA sequencing read

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

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

IIf your sequencing facility requires you to perform your own BigDye 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 to the left), so I would strongly recommend using a silica spin column. They can be pricey, but well worth it.

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 to troubleshoot any problems that come up. Good luck analyzing your sequences! ■

Further Reading

Tips for Using BLAST to Verify Plasmid

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.

NGS Plasmid Quality Control at Addgene

By Will Arnold and Amanda Hazen | April 2021

Our high-throughput DNA isolation process that yields high-quality isolated DNA samples of sufficient quantity for sequencing. This process is completed in a plate format yielding anywhere from two to six plates of 96 samples per week. This is where our NGS process begins. Partnering with seqWell, we are using the plexWell technology to easily and quickly create Illumina sequencing libraries.

From start to finish the library preparation process only takes about one day, even for six plates (576 individual plasmids!) These libraries are then QC'd, pooled, and prepared for Illumina sequencing on our newly donated MiSeq.

For Complete Plasmid Sequencing we perform a 2x251 run on our MiSeq that takes about two days to complete. After the run completes, we

begin our assembly process. Again, thanks to our partners at seqWell we make use of a pipeline that takes the raw data from each sample in our pool and individually assembles the reads into a single FASTA sequence that our QC scientists can easily analyze (see below for what we look out for while analyzing the plasmid sequences).

A BROADER VIEW OF THE WHOLE PLASMID

To confirm the sequence of the plasmid, we examine three things:

- Aligning the NGS result to a reference sequence to confirm backbone elements.
- Confirming the gene/insert by aligning to NCBI entry or using BLAST.
- Confirming tags and fusion proteins.

Aligning the sequence

Finally, we confirm promoters, tags, fusion proteins, and selectable markers by detecting common features using SnapGene. We're not necessarily expecting a perfect match; we will often find a few mismatches in the origin of replication or other common backbone elements. Since we've successfully grown the plasmid in culture to prepare it for sequencing, we feel confident that these few minor mismatches usually don't affect the function of the plasmid.

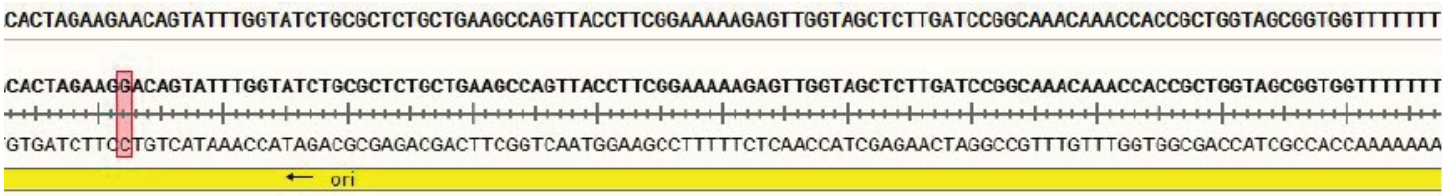


Figure 1. Mutations in the ori/backbone do not compromise plasmid function.The NGS sequence (top) was aligned to a known backbone sequence (bottom). This sequence alignment shows a mismatch, highlighted in red, in the backbone's origin of replication (labeled in yellow). Image created using Snapgene and pT2-cryR;snx5CE1-P1Egfp (Addgene #90152) from Nathan Lawson's lab.

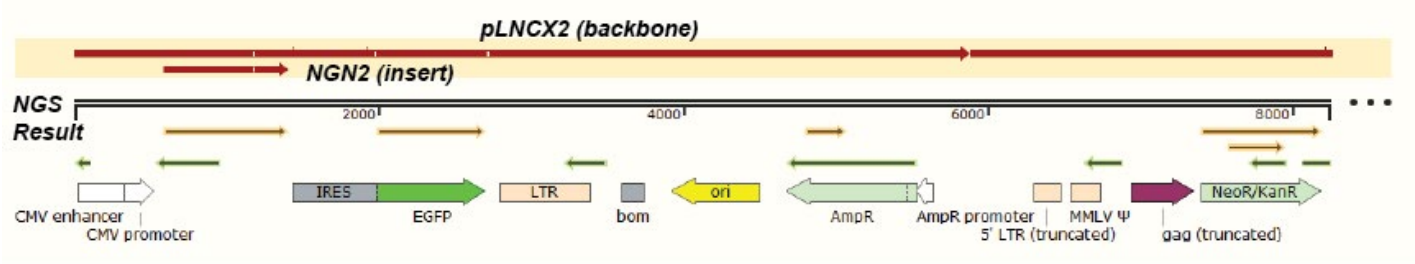


Figure 2. Confirming the insert. NGS sequence is shown as a double horizontal line with features annotated by Snapgene shown underneath. This sequence is aligned to the backbone sequence (top red arrow) and an NCBI reference sequence for the NGN2 gene (bottom red arrow). The alignment shows that the expected backbone elements and expected gene insert (NGN2-IRES-GFP) are present. Image created using Snapgene and pLNCX2-NGN2-IRES-GFP (Addgene #90212) from Chun-Li Zhang's lab.

Confirming the insert

We usually confirm the insert through **BLAST** or by direct alignment to an NCBI reference sequence or a sequence provided by the depositor. Depositing labs often provide insert sequences or annotated Genbank files that are useful for more complicated plasmids, like those that contain synthesized regions that have no publicly available reference sequence or plasmids containing genes with many modifications. We look for point mutations, truncations, and insertions that could compromise function. When we do find mutations, we check to see if they affect the translated amino acid. We also confirm that the species of the gene matches the data associated with the plasmid.

Confirming tags and fusion proteins

Finally, we confirm promoters, tags, fusion proteins, and selectable markers by detecting common features using Snapgene. If we find information that differs from what we would expect given the data provided by the depositing lab, we call these "quality control (QC) issues." We then ask the depositing laboratory to review the discrepancies. If the depositing laboratory confirms that these differences are expected and do not affect plasmid function, we will update information on the plasmid's page.

Sometimes we're missing a piece of the plasmid

There are regions of some plasmids that are particularly difficult to sequence and assemble, including GC rich regions. For example, the CAG promoter, a hybrid promoter consisting of a CMV enhancer, chicken beta actin promoter, and rabbit beta-globin splice acceptor, contains sequence that is over 80% GC. Some IRES sequences also contain regions that are GC rich. Since NGS assembles a larger sequence from many smaller sequences, regions with repeats may not assemble correctly. Due to these issues, NGS for some plasmids will not result in one complete, circular assembly. Most of the time, these plasmids are returned as one or two partial assemblies. If we obtain an NGS result that is useful, but not 100% complete, we will still make this data available as a partial sequence with the heading "Addgene Partial NGS Result." In some cases, there will be more information about our sequencing results in the Depositor Comments section at the bottom of the plasmid page. ■

Tips for Using BLAST to Verify Plasmids

By Jason Niehaus | June 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 40,000 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.

OPTIMIZING BLASTN SEARCHES

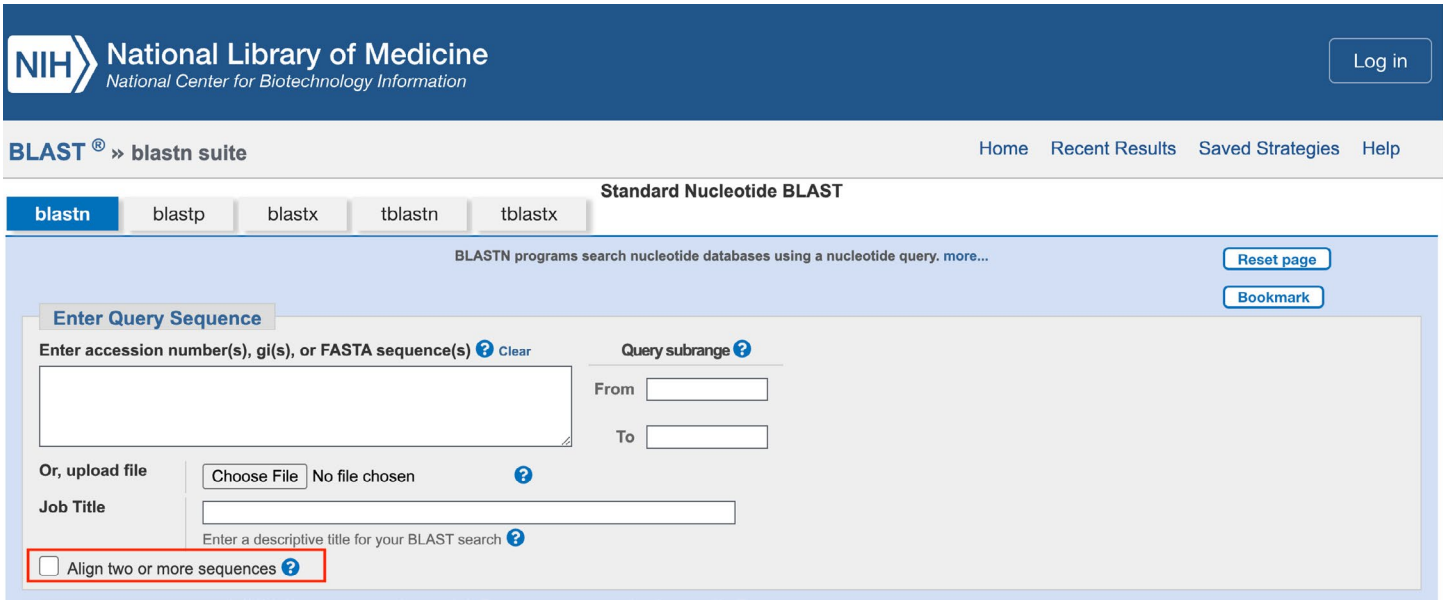
On the Standard Nucleotide BLAST page, the first decision we make is whether to compare our sequencing result to a single known reference sequence or to a BLAST sequence database. If you know the expected nucleotide sequence, check the "Align two or more sequences" checkbox and paste your reference sequence into the Subject Sequence box that appears. Aligning two nucleotide

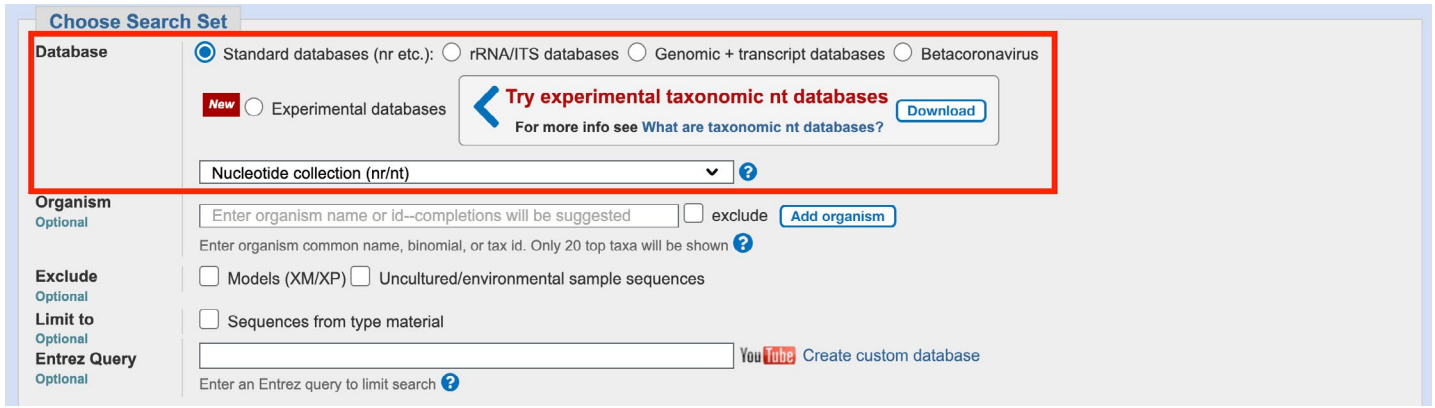
sequences is probably the fastest BLAST search to perform and will save you time compared to other types of BLAST searches.

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.

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 ≥ 95% (Our QC process would be trouble-free and much faster if





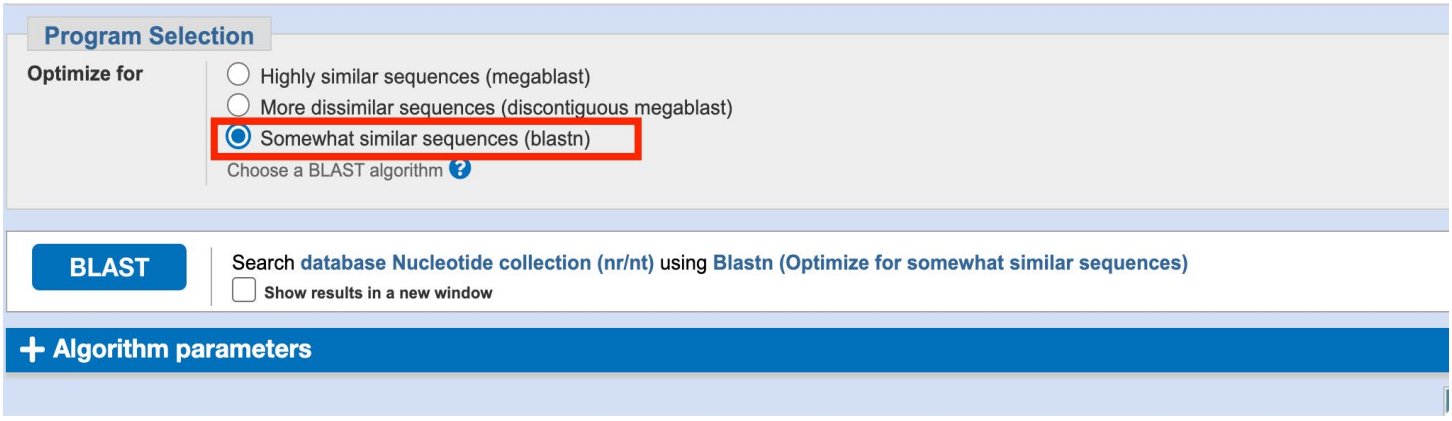
95% of our results were always correct!).

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)"

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.

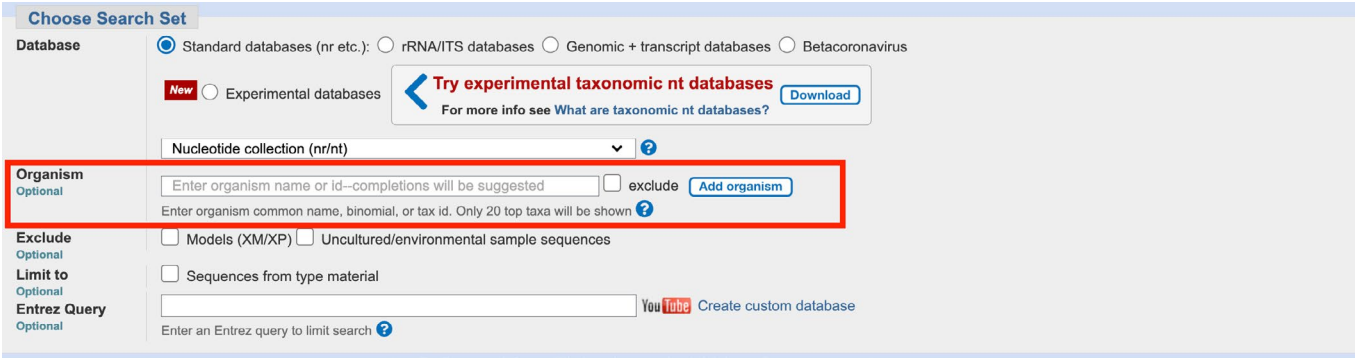
OPTIMIZING BLASTX SEARCHES

Once we have used blastn to determine the reliable portion of a sequencing result and



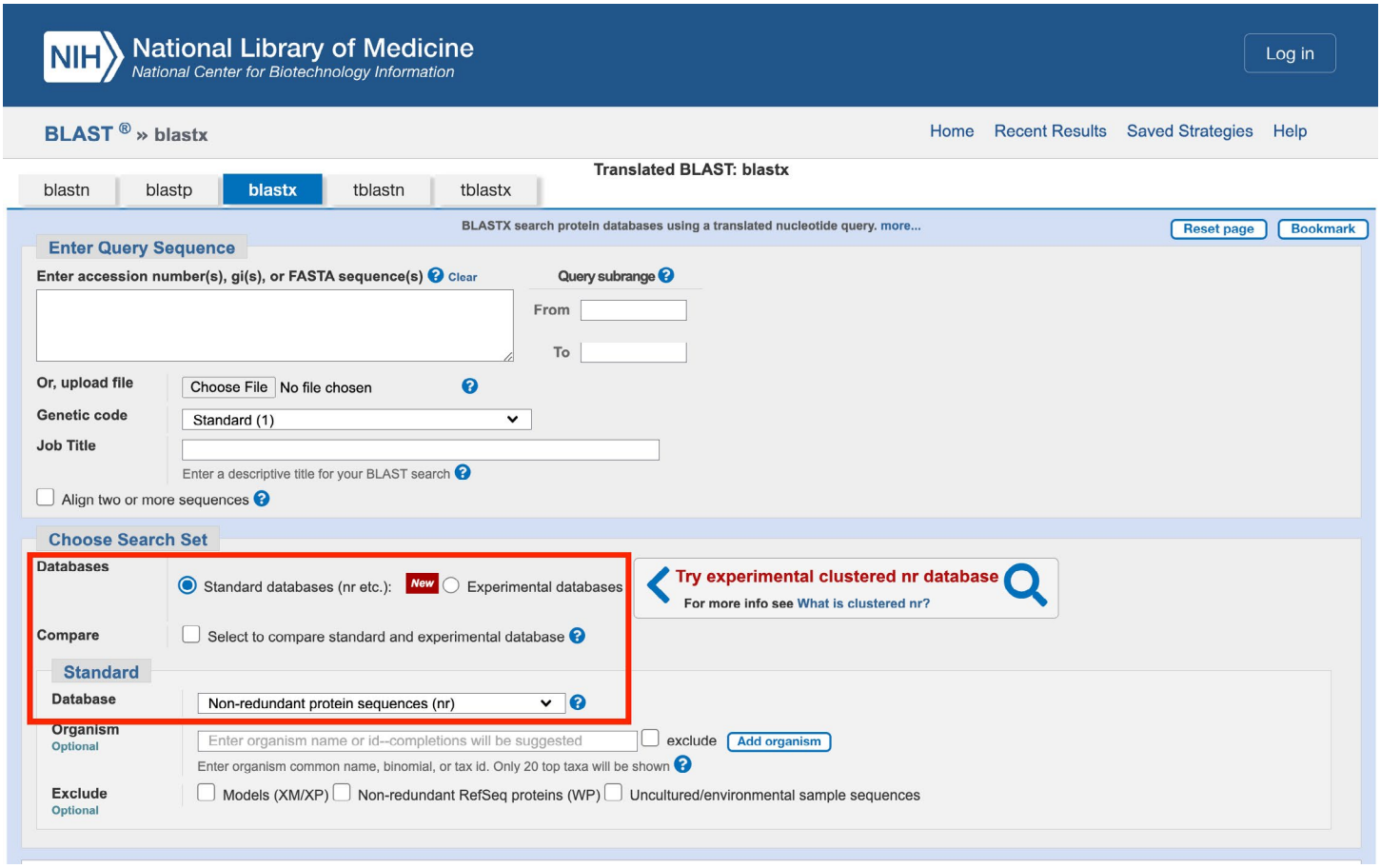
Pro Tip!

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!



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.

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



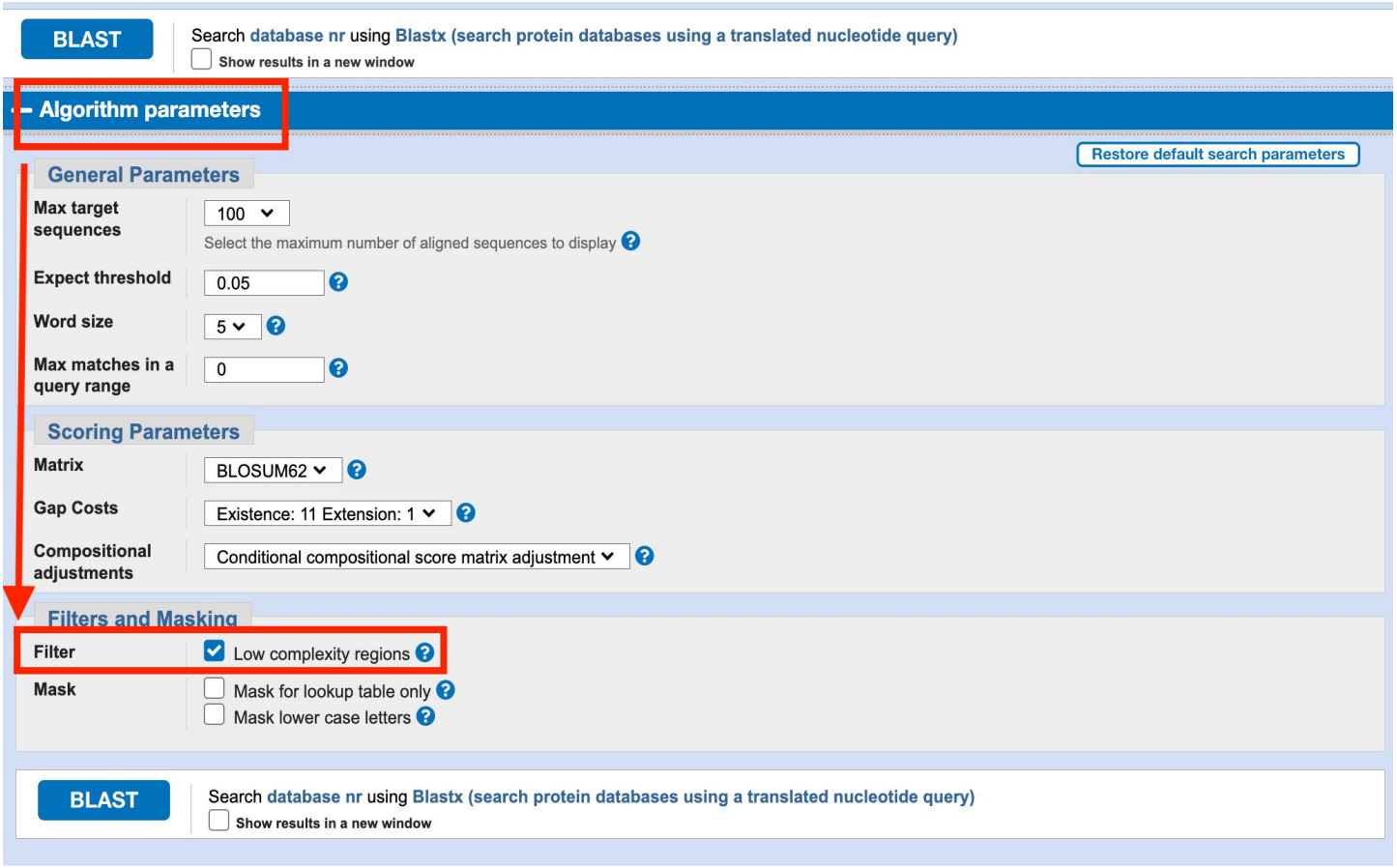
not infallible, we have found it saves analysis time to remove this default setting.

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

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.

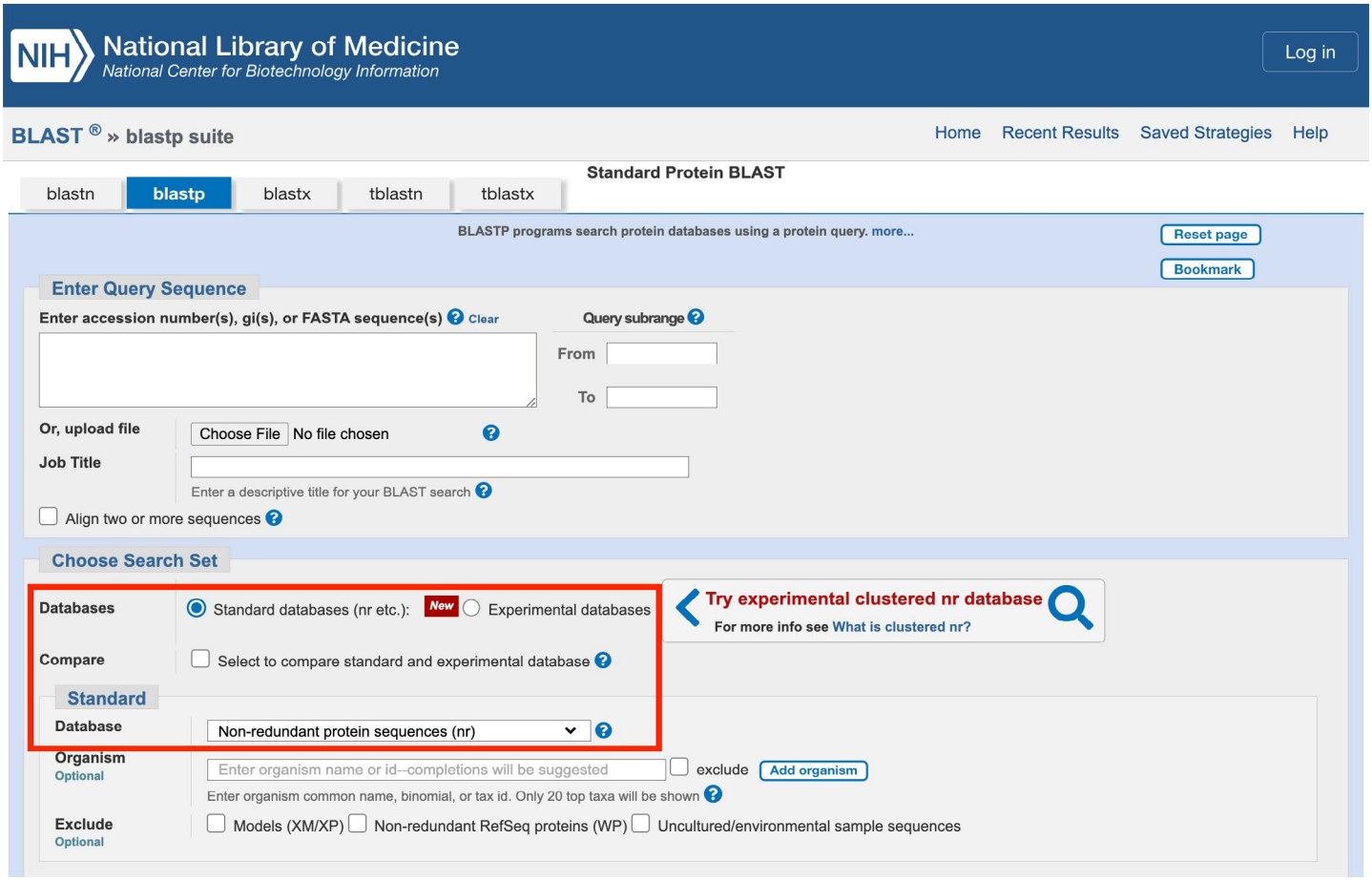


blastx searches are inherently slower than blastn or blastp, due to the additional searches involved in translating the nucleotide sequence into all six possible reading frames. If you know the expected protein sequence, use the "Align two or more sequences" option to drastically reduce waiting time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

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.

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.



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
- **EMBL-EBI** - http://www.ebi.ac.uk/services

- http://www.ebi.ac.uk/Tools/webservices/#multiple_sequence_alignment_msa

- http://www.ebi.ac.uk/Tools/webservices/#pairwise_sequence_alignment_psa

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.

Dimers and Multimers

By Matt Ouellette | January 203

When performing restriction digests on plasmids at Addgene, we sometimes observe something odd in our uncut DNA control: a band or two appear on an agarose gel at notably higher molecular weights than expected, given the size of the plasmid. This is seen only in the uncut DNA; the rest of the digest appears normal and produces the expected fragments for that plasmid. This phenomenon indicates the presence of a multimer.

Thanks to increasing affordability and accessibility, more researchers have been utilizing full plasmid sequencing services to verify the plasmids, including long-read sequencing techniques, such as Oxford Nanopore sequencing or PacBio SMRT sequencing, (which differ from the short-read Illumina MiSeq sequencing that Addgene uses.) However, when some researchers received the sequencing results for their plasmids, they found the plasmids were much longer than expected, and the sequence was repeated two or more times in tandem. Such plasmids are known as multimers, and they are the result of multiple plasmid copies combining in a process known as plasmid multimerization.

PLASMID MULTIMERIZATION

Early studies on plasmid topology recorded plasmid multimerization in the naturally occurring E. coli plasmid ColE1 (Goebel & Helinski, 1968; Bazaral & Helinski, 1968), in which the plasmid was observed to exist in various multimeric forms (e.g.

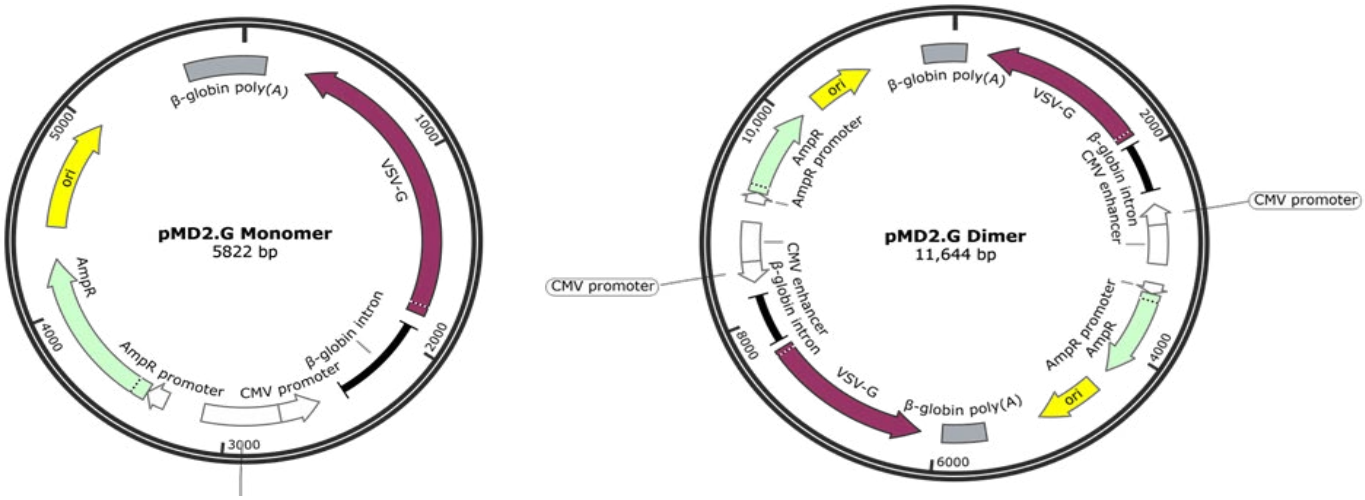


Fig. 1: The monomer of Plasmid 12259: pMD2.G (depicted on the left) is ~5.8 kb in length. The plasmid can exist on as a dimer, depicted on the right, with the entire monomeric sequence duplicated in tandem, resulting in a plasmid that is twice the length of the monomer (~11.6 kb). Image generated by SnapGene software.

concatemers, dimers, trimers, tetramers). Since then, multimerization has been observed to occur in many different kinds of plasmids, taking many forms (Levene, 2009; Higgins & Vologodskii, 2016). Plasmid multimerization occurs via homologous recombination between two or more copies of the same plasmid (Bedbrook & Ausubel, 1976). This recombination event produces a multimeric form of the plasmid with repeating sequences. In some specific cases, multimerization has been observed to occur more frequently in plasmids with large inserts (>7 kb)(Berg et al., 1989), direct repeats (Ribeiro et al., 2009), and at high copy numbers (Williams, Carnes, and Hodgson, 2009). More research is ongoing to determine plasmid features that affect the frequency of multimerization.

Theoretically, as long as all the plasmid features remain intact, plasmid multimers function just as well as monomers. However, due to their larger size, multimers may exhibit reduced transformation and transfection efficiency compared to the monomeric version of the plasmid (Maucksch et al., 2009). Since multimers have additional origins of replication compared to monomers, they are also able to replicate at a higher frequency (Summers & Sherratt, 1984; Summers et al., 1993). These larger plasmids are also maintained at lower copy numbers within bacterial cells, and bacterial cells with multimers grow more slowly compared to those containing the monomeric version of the plasmid (Summers & Sherratt, 1984; Summers et al., 1993; Summers, 1998). Multimeric plasmids are also considered to be less stable and more prone to being lost in bacterial cells compared to monomers, although this can be prevented via antibiotic selection (Crozat, 2014).

HOW TO DETECT PLASMID MULTIMERS

Run an undigested plasmid on an agarose gel

Running undigested DNA on a gel as a control when performing a diagnostic digest is a simple but effective method for detecting plasmid multimers. Digests themselves cannot be used to detect multimers, since restriction sites in multimers usually repeat in tandem along with the repeated plasmid sequence, giving greater concentration of cut DNA without changing the size. As a result, a restriction digest of multimers will produce the same fragment sizes and banding patterns on an agarose gel as the monomeric version. However,

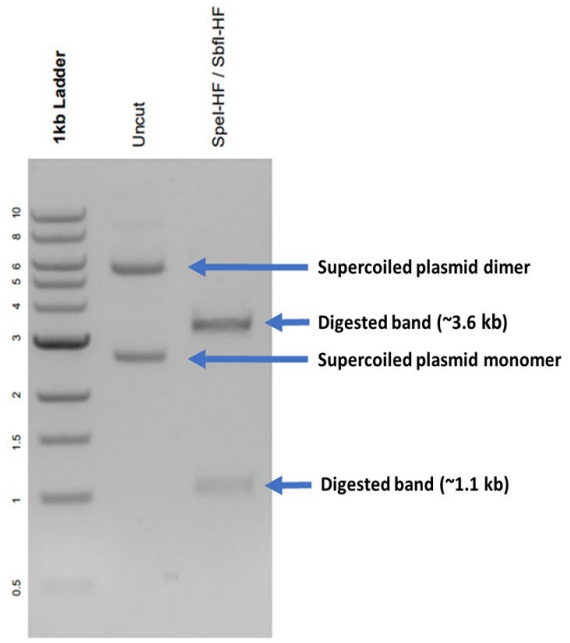


Fig. 2: Gel image of Plasmid 61564: pBAMD1-2 (~4.7 kb) digested with SpeI and SbfI with expected bands at ~3.6 kb and ~1.1 kb (lane 3). Uncut DNA (lane 2) has a band at <3 kb, which corresponds to the supercoiled plasmid monomer, and at <6 kb, consistent with the size of a supercoiled plasmid dimer. Please note that it is common to see other plasmid confirmations like linear and nicked plasmids, as described here.

undigested multimers are larger than monomers and will therefore run higher on an agarose gel (Bedbrook & Ausubel, 1976). In fact, if your plasmid sample has a mix of different-sized multimers, you may even see different sets of bands on the gel corresponding to each type of multimer present in your sample. In these instances, please note that the supercoiled dimer band can sometimes overlap with the band corresponding to the nicked or relaxed conformation of the plasmid monomer (see this article from BiteSize Bio for more information on how different plasmid conformations run on an agarose gel). If the presence of a multimer in your sample remains unclear after visualizing it on a gel, one of the other methods below should help in multimer identification.

Long-read sequencing

Addgene uses Illumina MiSeq NGS to perform whole plasmid sequencing of our plasmid samples during our quality control process. Since MiSeq involves sequence assembly from short reads (<300 bp) which cannot distinguish between large repeat sequences within a plasmid, it cannot be used to detect plasmid multimers. Some companies provide NGS whole plasmid sequencing services using long-read sequencing (1-25 kb), such as Oxford Nanopore sequencing or PacBio SMRT sequencing. These longer reads are able to detect large repeat sequences within a plasmid and, therefore, allow for the identification of plasmid multimers.

Capillary Gel Electrophoresis (CGE) analysis

Another technique that can detect multimers is capillary gel electrophoresis (CGE), which has been used in plasmid vaccine and gene therapy development to examine plasmid topologies (e.g. supercoiled, open circular, linearized, etc.) (Schleef et al., 2006; Holovics et al., 2010; Cook et al., 2020). CGE uses laser-induced fluorescence (CGE-LIF) for detection and can provide data on the quantity of various topologies and multimers present within a plasmid sample.

TROUBLESHOOTING TIPS

Grow plasmids in recombinase-deficient strains

Since plasmid multimerization is caused by homologous recombination, growing a plasmid in a bacterial strain that is deficient in recombinase protein RecA can reduce the frequency of multimerization, and in some cases completely prevent the formation of plasmid multimers (Bedbrook & Ausubel, 1976; Fishel et al., 1981). E. coli strains such as DH5α, HB101, NEB Stable, Stbl3, etc., are recA- mutant strains, and therefore should have reduced frequencies of plasmid multimerization. One study by Bacolla et al. (2011) indicated that strain HB101 produced the highest yield of monomeric plasmid compared to DH5α. However, more research is needed to determine which recA- strains are most effective at reducing multimerization and define recA-independent mechanisms.

Avoid overgrowing the bacterial culture

Another way to prevent plasmid multimerization is to avoid growing the transformed bacterial culture into the late stationary phase. Growing the culture

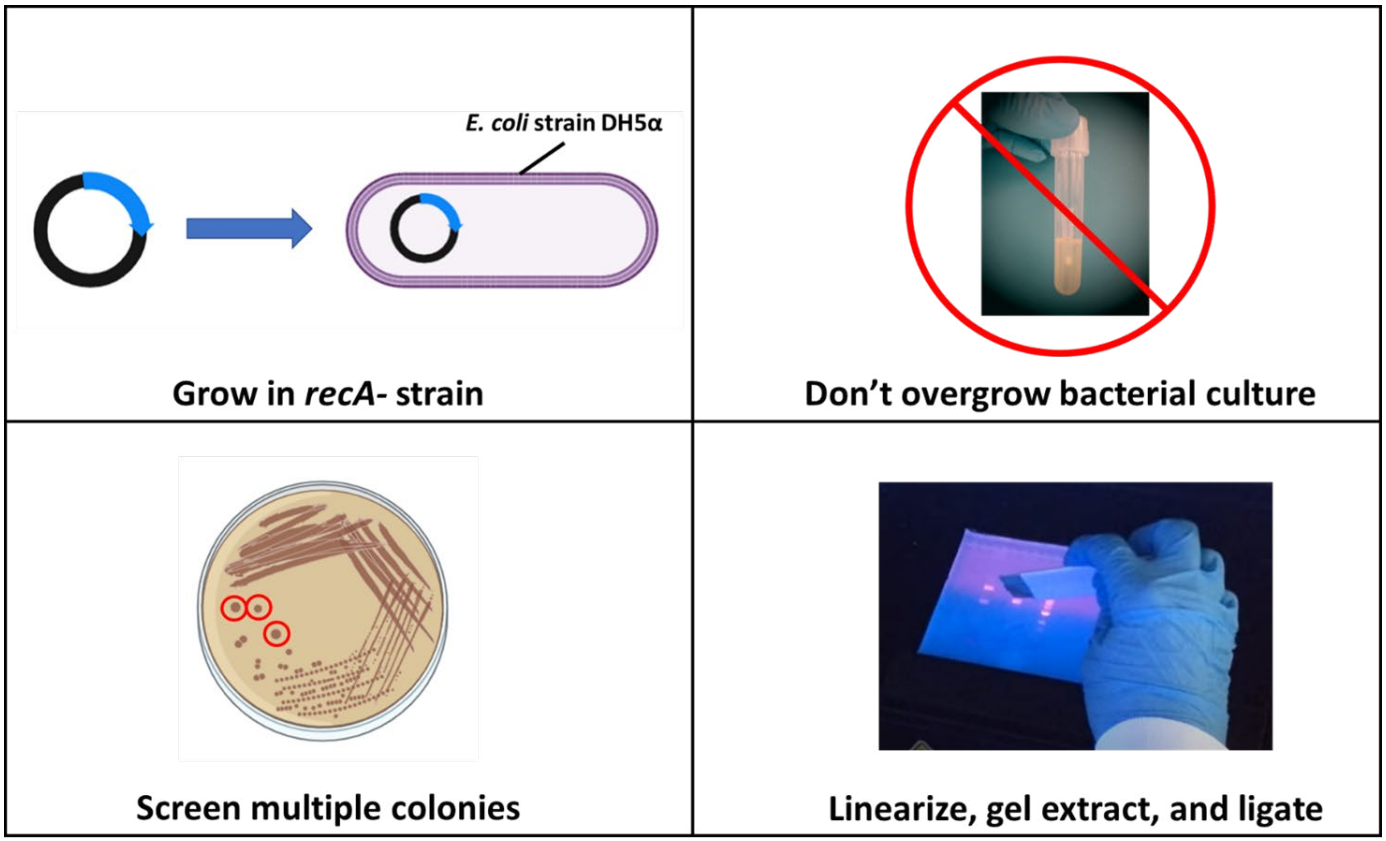
at a lower temperature such as 30°C or optimizing the incubation time can help maintain cultures in log phase growth. In a study by Williams et al. (2009), pUC multimerization was reduced by growing seed stocks at 30°C compared to 42°C.

Screen multiple colonies to isolate the monomer

Sometimes a plasmid can exist in a bacterial population as a mixture of monomers and various forms of multimers (Summers & Sharratt, 1984; Summers et al., 1993). In these cases, screening multiple bacterial colonies using one of the detection methods described above can help ensure the isolation of the monomeric version of the plasmid.

Linearize, gel extract, and ligate the plasmid

If you are having difficulty isolating a plasmid monomer from a plasmid sample, or if no monomers exist in your sample, you could try linearizing the plasmid via restriction digestion and then run the linearized plasmid on an agarose gel. You can then gel extract the linearized band, ligate the plasmid, and re-transform the ligated plasmid to recover the monomeric form of the plasmid. ■



REFERENCES

Bacolla A, Wang G, Jain A, Chuzhanova NA, Cer RZ, Collins JR, Cooper DN, Bohr VA, Vasquez KM. Non-B DNA-forming sequences and WRN deficiency independently increase the frequency of base substitution in human cells. J Biol Chem. 2011 Mar 25;286(12):10017-26. Doi: 10.1074/jbc.M110.176636. Epub 2011 Feb 1. PMID: 21285356; PMCID: PMC3060453.

Bazaral M, Helinski DR. Characterization of multiple circular DNA forms of colicinogenic factor E-1 from Proteus mirabilis. Biochemistry. 1968 Oct;7(10):3513-20. doi: 10.1021/bi00850a028. PMID: 4878700.

Bedbrook JR, Ausubel FM. Recombination between bacterial plasmids leading to the formation of plasmid multimers. Cell. 1976 Dec;9(4 PT 2):707-16. doi: 10.1016/0092-8674(76)90134-3. PMID: 797459.

Berg CM, Liu L, Coon M, Strausbaugh LD, Gray P, Vartak NB, Brown M, Talbot D, Berg DE. pBR322-derived multicopy plasmids harboring large inserts are often dimers in Escherichia coli K-12. Plasmid. 1989 Mar;21(2):138-41. doi: 10.1016/0147-619x(89)90057-7. PMID: 2544913.

Cook KS, Luo J, Guttman A, Thompson L. Vaccine Plasmid Topology Monitoring by Capillary Gel Electrophoresis.

Curr Mol Med. 2020;20(10):798-805. doi: 10.2174/1566524020066620042711045 2. PMID: 32338220.

Crozat E, Fournes F, Cornet F, Hallet B, Rousseau P. Resolution of Multimeric Forms of Circular Plasmids and Chromosomes. Microbiol Spectr. 2014 Oct;2(5). Doi: 10.1128/microbiolspec.PLAS-0025-2014. PMID: 26104344.

Fishel RA, James AA, Kolodner R. recA-independent general genetic recombination of plasmids. Nature. 1981 Nov 12;294(5837):184-6. doi: 10.1038/294184a0. PMID: 7029308.

Goebel W, Helinski DR. Generation of higher multiple circular DNA forms in bacteria. Proc Natl Acad Sci U S A. 1968 Dec;61(4):1406-13. doi: 10.1073/pnas.61.4.1406. PMID: 4884687; PMCID: PMC225270.

Higgins NP, Vologodskii AV. Topological Behavior of Plasmid DNA. Microbiol Spectr. 2015 Apr;3(2):10.1128/microbiolspec.PLAS-0036-2014. doi: 10.1128/microbiolspec.PLAS-0036-2014. PMID: 26104708; PMCID: PMC4480603.

Holovics HJ, He Y, Lacher NA, Ruesch MN. Capillary gel electrophoresis with laser-induced fluorescence of plasmid DNA in untreated capillary. Electrophoresis. 2010 Jul;31(14):2436-41. doi: 10.1002/elps.201000061. PMID: 20589859.

Levene SD. Analysis of DNA topoisomers, knots, and catenanes by agarose gel electrophoresis. Methods Mol Biol. 2009;582:11-25. doi: 10.1007/978-1-60761-340-4_2. PMID: 19763938.

Maucksch C, Bohla A, Hoffmann F, Schleef M, Aneja MK, Elfinger M, Hartl D, Rudolph C. Transgene expression of transfected supercoiled plasmid DNA concatemers in mammalian cells. J Gene Med. 2009 May;11(5):444-53. doi: 10.1002/jgm.1310. PMID: 19263463.

Ribeiro SC, Oliveira PH, Prazeres DM, Monteiro GA. High frequency plasmid recombination mediated by 28 bp direct repeats. Mol Biotechnol. 2008 Nov;40(3):252-60. Doi: 10.1007/s12033-008-9082-3. Epub 2008 Jul 8. PMID: 18607781.

CHAPTER 12

You've Made a Plasmid...Now What?

How to Name Your Plasmid in Three Easy Steps

By Matthew Ferenc, | February 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.

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

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), giving the format

pBACKBONE-hGene.

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 3: ADD YOUR TAGS

Add any tags or fusions that are on your insert. Typically you would list any tag or fusion protein 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.

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.

Pro Tip!

For a catalog of published and commercially available empty backbones, visit Vector Database.

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.

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. ■

Plasmid Incompatibility

By Leah Schwiesow | March 2020

Plasmid incompatibility is defined as the inability of different plasmids to be maintained in one bacterial cell. Here, we'll cover why this happens, how it might affect your work, and how understanding it can be used for good.

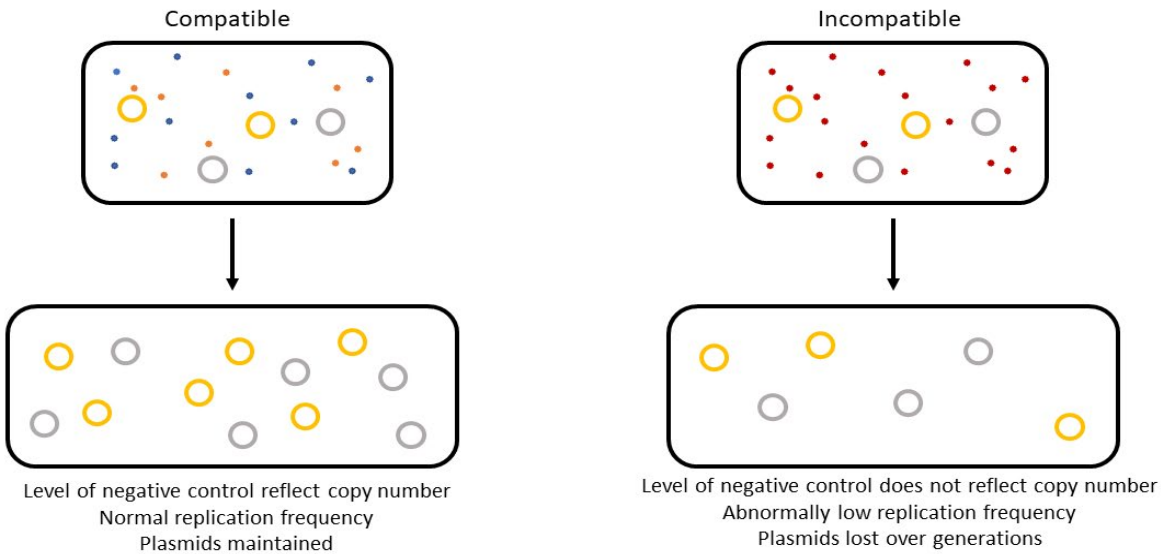
First, why are plasmids incompatible? Plasmid incompatibility occurs when multiple plasmids within one cell have the same replicon and/or partitioning system. Let's start with the replicon- the part of the plasmid that contains the origin of replication and the replication control machinery.

THE REPLICATION PROBLEM

Plasmids with the same replicon are incompatible because they compete for the same replication control machinery within the cell. Most plasmids encode a negative regulation system involving antisense RNAs or iterons that inhibits replication when copy number in the cell is high, but allows replication when copy number in the cell falls too low (Novcik, 1987).

Antisense RNAs encoded on the plasmid either indirectly inhibit replication through inhibiting translation of replication machinery proteins, or

Figure 1: The Replication Problem



directly inhibit replication by binding to the origin of replication and blocking replication machinery (del Solar, 1998). As the copy number of the plasmid increases, increased amounts of these antisense RNAs inhibit plasmid replication.

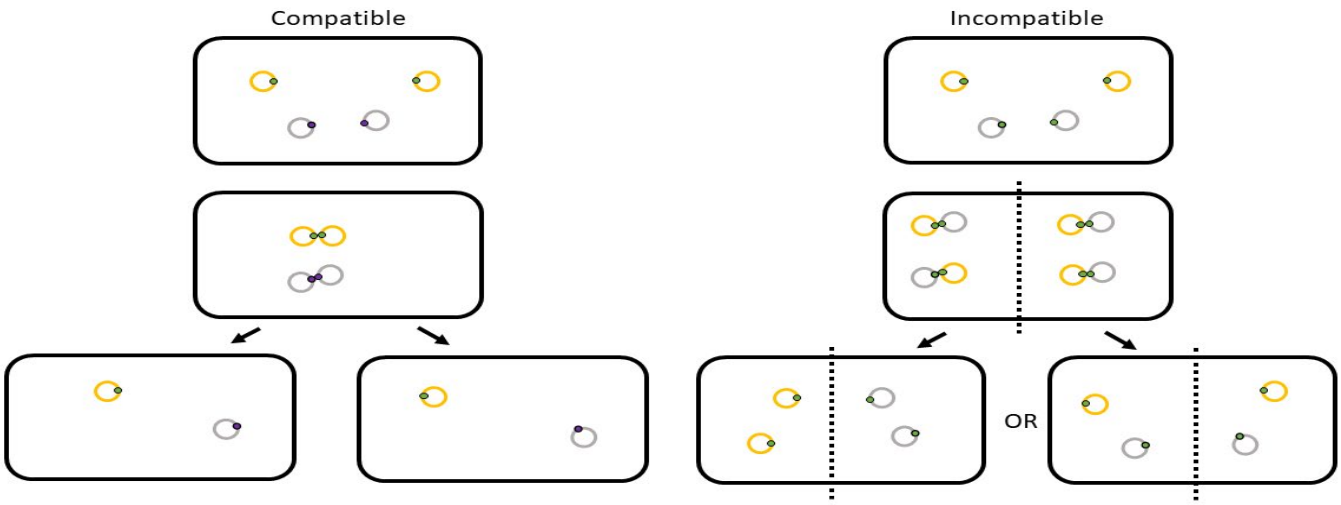
Iterons are repeated sequences within the origin of replication and are required for replication in plasmids that contain them (del Solar et al., 1998). There are two models for iteron-based plasmid replication control that are not necessarily mutually exclusive. In the first, the replication initiation factor RepA binds to the iteron to initiation replication. However, in iteron-containing plasmids, RepA can also bind to its own promoter and inhibit its own transcription. RepA molecules, forming a "handcuff" structure that physically blocks replication machinery when plasmid copy number is high (del Solar et al., 1998). In the second model, termed the "handcuff" model, iterons of two separate plasmids bind to the same RepA mole- cules, forming a "handcuff" structure that physically blocks replication machinery when plasmid copy number is high (del Solar et al., 1998).

In cases of compatible plasmids, different negative regulation systems control plasmid replication of each plasmid using unique replication machinery. However, when plasmids have the same origin of replication, the negative regulation system cannot distinguish between the different plasmids. The concentration of negative regulation system elements in the cell is artificially high for both plasmids, so each plasmid "thinks" it has a higher copy number than it actually does, and neither plasmid is maintained.

THE PARTITIONING PROBLEM

Plasmid incompatibility also occurs because of plasmid partitioning (Schumacher, 2012). This concept is a little more complex than simply competing for replication control factors and tends

Figure 2: The Partitioning Problem (*Adopted from Bouet et al, 2007*)



to be more of a concern with low copy plasmids, though the partitioning problem does occur in high copy plasmids (Diaz et al., 2015). Plasmids contain a system to partition themselves to each daughter cell during cell division. They don't simply rely on chance (or random diffusion) to make this happen.

There are several different versions of the plasmid partitioning system in the bacteria, but generally it consists of a centromere-like region on the plasmid, a centromere-like region binding protein (the CBP), and a partition NTPase (Schumacher, 2012). During cell division, CBPs bind to the centromere-like region on each plasmid and pair the plasmids together (akin to sister chromatids coming together in eukaryotic cells). The partition NTPase is recruited and "walks" each plasmid to a separate daughter cell. When plasmids are compatible, different CBPs bind to each plasmid type, and different NTPases separate the plasmid pairs into the new daughter cells.

For high copy plasmids, incompatibility due to partitioning is similar to incompatibility due to having the same replication machinery. High copy plasmids with the same centromere-like binding region compete for the same CBPs and NTPase to correctly partition plasmids to each daughter cell. When there are not enough CBPs and NTPase to go around, the plasmids are randomly positioned, leading to plasmid loss (Diaz et al., 2015).

For low copy plasmids, the main theory of plasmid incompatibility states it stems from **an identification issue:** the CBPs in the cell can't tell the difference between plasmids if they have the same centromere-like region, and the NTPases end up "walking" the same plasmids to one daughter cell, instead of partitioning each type of plasmid into a separate daughter cells (Figure 2) (Ebersbach et al., 2005). However, more recent studies suggest that the late replication that occurs for some types of low copy plasmids during cell division simply does not allow for enough time for correct partitioning to occur (Diaz et al., 2015).

It is also worth noting that plasmids can be symmetrically incompatible, meaning that both plasmids are lost from the cell with the same probability, or asymmetrical, where one plasmid is lost from the cell at a higher rate than the coresident plasmid. Asymmetric plasmid loss occurs for several reasons, including one plasmid blocking replication of another and one plasmid outcompeting the other plasmid for replication or partitioning machinery (Novick, 1987). In cases where one plasmid outcompetes the other, usually a smaller, higher copy plasmid can more effectively titrate replication or partitioning machinery away from a larger, low copy plasmid and the larger, low copy plasmid is more likely to be lost over time.

PLASMID INCOMPATIBILITY IN THE LAB

If you want to clone two plasmids within the same bacterial cell, think about whether plasmid incompatibility will affect your plans. This situation may arise for a microbiologist, where cloning two plasmids into one bacterium is not uncommon, or perhaps a synthetic biologist who wants to develop a plasmid-based tool that is easy to use with other plasmids.

How do you know if your plasmids are incompatible? Scientists have developed a system of bacterial incom- patibility groups based on similarity of replication and partitioning systems. Scientists have developed a system of bacterial incompatibility groups based on similarity of replication and partitioning systems. Currently, there are 27 incompatibility (Inc) groups in the Enterobacteriaceae family alone, and this is expanding as we obtain more genomic data about plasmids (Shintani et al., 2015). Now, bioinformatics tools help researchers determine how compatible their plasmids will be based on sequence data. However, if these do not provide clarity, a wet lab transformation experiment will ultimately inform whether your plasmids are compatible or not.

USING PLASMID INCOMPATIBILITY FOR THERAPEUTIC ADVANTAGE

Although it may seem like plasmid incompatibility may be standing in the way of bacterial cloning glory, scientists have applied these concepts to deal with the more sinister plasmids of the microbiological world. Scientists have designed small, high copy incompatible plasmids that lead to asymmetrical plasmid loss of the usually large, low copy virulence plasmids. This strategy has been successful in displacing, or "curing," the virulence plasmids from the bacterial pathogens Yersinia pestis, Agrobacterium tumefaciens, and Bacillus anthracis (Lui et al., 2012, Ni et al., 2008, Uraji et al., 2002). Scientists are beginning to use plasmid incompatibility to combat antimicrobial resistance. Using novel plasmids based partly on plasmid incompatibility with antibiotic-resistance plasmids, scientists have expelled antibiotic resistant plasmids from Enterobacteriaceae family members within the mouse gut (Kamruzzaman et al., 2017). ■

REFERENCES

del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R (1998) Replication and Control of Circular Bacterial Plasmids. Microbiology and Molecular Biology Reviews 62(2): 434–464. Pubmed PMID: 9618448

Diaz R, Rech J, Bouet J-Y (2015) Imaging centromere-based incompatibilities: Insights into the mechanism of incompatibility mediated by low-copy number plasmids. Plasmid 80:54–62 . https://doi.org/10.1016/j.plas- mid.2015.03.007 Pubmed PMID: 25889267

Diaz R, Rech J, Bouet J-Y (2015) Imaging centromere-based incompatibilities: Insights into the mechanism of incompatibility mediated by low-copy number plasmids. Plasmid 80:54–62. https://doi.org/10.1016/j.plasmid.2015.03.007 Pubmed PMID: 25889267

Kamruzzaman M, Shoma S, Thomas CM, Partridge SR, Iredell JR (2017) Plasmid interference for curing an- tibiotic resistance plasmids in vivo. PLOS ONE 12:e0172913 . https://doi.org/10.1371/journal.pone.0172913 Pubmed PMID: 28245276

Liu X, Wang D, Wang H, Feng E, Zhu L, Wang H (2012) Curing of Plasmid pXO1 from Bacillus anthracis Using Plasmid Incompatibility. PLoS ONE 7:e29875. https://doi.org/10.1371/journal.pone.0029875 Pubmed PMID: 22253811

Ni B, Du Z, Guo Z, Zhang Y, Yang R (2008) Curing of four different plasmids in Yersinia pestis using plasmid incompatibility. Letters in Applied Microbiology 47:235–240. https://doi.org/10.1111/j.1472-765x.2008.02426.x Pubmed PMID: 19241516

Novick RP (1987) Plasmid incompatibility. Microbiology Reviews. 1987;51(4):381–395. Pubmed PMID: 3325793

Schumacher MA (2012) Bacterial plasmid partition machinery: a minimalist approach to survival. Current Opinion in Structural Biology 22:72–79. https://doi.org/10.1016/j.sbi.2011.11.001 Pubmed PMID: 22153351

Shintani M, Sanchez ZK, Kimbara K (2015) Genomics of microbial plasmids: classification and identifcation based on replication and transfer systems and host taxonomy. Frontiers in Microbiology. https://doi. org/10.3389/fmicb.2015.00242 Pubmed PMID: 25873913

Uraji M, Suzuki K, Yoshida K (2002) A novel plasmid curing method using incompatibility of plant pathogenic Ti plasmids in Agrobacterium tumefaciens. Genes & Genetic Systems 77:1–9. https://doi.org/10.1266/ggs.77.1 Pubmed PMID: 12036099

Codon Usage Bias

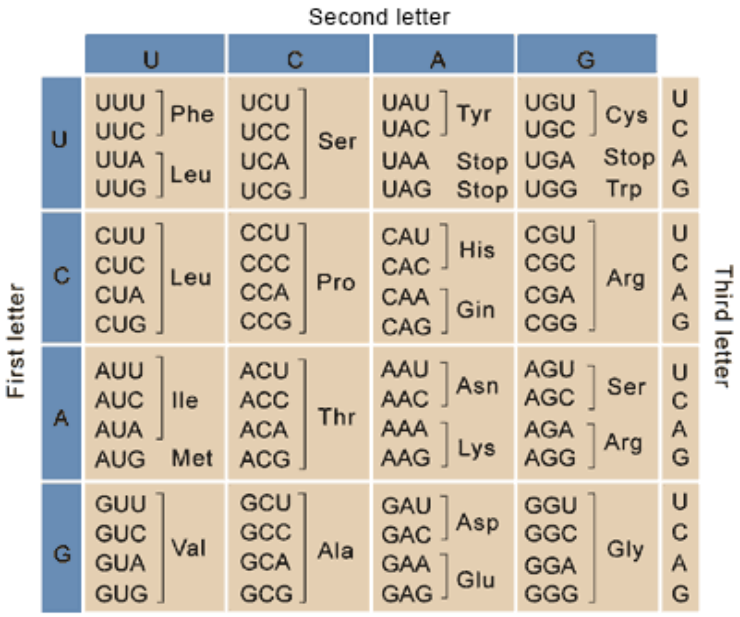
By Tyler Ford | September 2018

A similar genetic code is used by most organisms on Earth, but different organisms have different preferences for the codons they use to encode specifc amino acids. There are therefore 64 possible codons but only 20 amino acids and 3 stop codons to encode leaving 41 codons unaccounted for. The result is redundancy; multiple codons encode single amino acids. Evolutionary constraints have molded which codons are used preferentially in which organisms - organisms have codon usage bias.

You can fnd many codon tables showing which codons encode which amino acids.. With such simple rules, you might think it's easy to come up with a workable DNA sequence to encode your peptide of interest and produce that peptide in your organism of choice. Unfortunately, codon preferences make it so you cannot choose among the possible codons at random and expect your sequence to express well in any organism.

WHY DO ORGANISMS HAVE DIFFERENT CODON USAGE BIASES?

The reasons for varied codon preferences among organisms aren't completely understood, but some possible reasons include:



Metabolic pressures

It takes cellular resources to produce tRNAs that recognize different codons, modify the tRNAs correctly, and charge the tRNAs with the appropriate amino acids. If an organism uses only a subset of codons, it only needs to produce a subset of charged tRNAs and therefore may need fewer resources for the entire translation process. For example, during high growth rate conditions, E. coli preferentially upregulates production of tRNAs that recognize codons found in highly expressed genes (Emilsson and Kurland, 1990).

Controlling gene expression through gene sequence

Proteins that are encoded by codons with low abundance or poorly charged tRNAs may be produced at a lower rate than proteins encoded by highly abundant, charged tRNAs. For example

Tuller et al. found that translation efficiency is well correlated with codon bias in both *E. coli* and *S. cerevisiae*.

Protein folding

If a protein is encoded by a mixture of codons with highly and poorly charged tRNAs, different regions of the protein may be translated at different rates. The ribosome will move quickly along regions calling for abundant, charged tRNAs but will stall at regions calling for low abundance, poorly charged tRNAs. When the ribosome stalls, this may give the swiftly translated regions a chance to fold properly. For example, Pechmann and Frydman found that tracts of non-optimal codons are associated with specific secondary structures in 10 closely related yeast strains.

Adaptation to changing conditions

Organisms often need to express genes at different levels under different conditions. With varied codon usage, an organism can change which proteins are highly expressed and which are poorly expressed by producing and charging specific tRNA pools. For example, tRNAs used in genes encoding amino acid biosynthetic enzymes may be preferentially charged during amino acid starvation thus resulting in higher production of amino acid biosynthetic enzymes (Dittmar et al., 2005).

HOW DOES CODON USAGE BIAS AFFECT MY EXPERIMENTS?

While codon preferences can be very useful for organisms, they can be problematic for researchers trying to express proteins in heterologous hosts. If you simply amplify a gene of interest from the human genome, for instance, it may not express at all in *E. coli* (you can fnd a variety of databases showing various organisms' codon preferences online). Even if the gene is translated, it may not function properly. This is the result of a mismatch between human and *E. coli* codon preference. Some codons commonly used in humans are not at all common in *E. coli* and vice versa. When translating these codons, the ribosome may therefore stall at inappropriate locations or fail to make it through the entire transcript resulting in the production of nonfunctional proteins and protein fragments respectively.

SOLVING THE PROBLEM OF CODON USAGE BIAS

Codon optimization

With low cost DNA synthesis, one of the primary ways researchers solve the problem of codon choice is to resynthesize genes in such a way that their codons are more appropriate for the desired expression host. This is known as "codon optimization." Though simple in theory, this is not as easy as it sounds. Even for relatively short peptides, there can be many possible ways to encode them and what constitutes the "appropriate" codon is not necessarily obvious.

You might think, "Nonsense! I should just choose the codon with the most abundant pool of charged tRNAs in my host organism for every amino

acid I'd like to encode," but, as described above, not every region of a protein should necessarily be translated rapidly to produce a protein that functions properly.

You might then think, "Okay, I'll just make sure the abundances of the codons I choose for the host match the abundances of codons used in the native organism." This is possibly a better idea and has been used successfully in the past (**Angov et al., 2008**), but there are still many more features to consider when designing a full gene. A non-exhaustive list includes:

- Codon abundance relative to cognate tRNA abundance repetitive sequences
- Restriction sites
- Sequences prone to create secondary structures in RNA transcripts
- Effects on transcription (remember, it's not all about translation - e.g. codon choice may interrupt transcription actor binding sites)

It's not easy for humans to balance all of these factors on their own. Luckily, many researchers have created codon optimization algorithms and DNA synthesis companies such as IDT and GenScript host online codon optimization tools. Keep in mind that, just because you optimize a gene with one of these tools, it doesn't necessarily mean the gene is going to express well. If you do get good expression, you should also functionally analyze the protein produced to ensure that it has folded properly.

You may be able to avoid getting your genes of interest codon optimized by ordering plasmids containing them from Addgene. If a plasmid at Addgene contains a gene that's been codon optimized for a particular organism, this will sometimes (but not always) be noted in in the "mutation" field on the plasmid page (**see plasmid 87904** for example.) As many plasmids available from Addgene now have full sequence data, we recommend directly analyzing gene sequences for codon optimization and suitability for your expression host before using them in your experiments.

EXPRESSION OF ALTERNATIVE tRNAS

If you don't have the time or the funds to synthesize a codon optimized version of your gene of interest, it's possible to overexpress low abundance tRNAs in your expression host and thereby increase their abundance. For example, the commercial **Rosetta** *E. coli* strains express a variety of tRNAs that are normally found at low abundance in *E. coli*.

The advantage of producing additional tRNAs is that you can use the same expression system for many different genes without having to create new constructs. However, due to problems such as mismatched translation rates and potential effects on cell growth, even hosts producing alternative tRNAs may not express sufficient amounts of your protein of interest.

Regardless of which method you choose to overcome the problems surrounding codon choice, you should have some method to make sure the proteins you produce function properly. Overexpression can result in the production of

insoluble, nonfunctional globs of protein known as inclusion bodies that will generally segregate with the cell pellet during purification procedures. Even if you produce a large amount of protein in your expression host of choice, you should perform a functional assay to make sure your protein isn't forming inclusion bodies and is folding properly.

For a review of codon usage bias, we suggest:

Quax, Tessa EF, et al. "Codon bias as a means to fine-tune gene expression." Molecular cell 59.2 (2015): 149-161. PubMed PMID: 26186290. PubMed Central PMCID: PMC4794256 ■

REFERENCES

Angov, Evelina, et al. "Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host." PloS one 3.5 (2008): e2189. PubMed PMID: 18478103. PubMed Central PMCID: PMC2364656.

Dittmar, Kimberly A., et al. "Selective charging of tRNA isoacceptors induced by amino–acid starvation." EMBO reports 6.2 (2005): 151-157. PubMed PMID: 15678157. PubMed Central PMCID: PMC1299251.

Emilsson, Valur, and Charles G. Kurland. "Growth rate dependence of transfer RNA abundance in Escherichia coli." The EMBO journal 9.13 (1990): 4359-4366. PubMed PMID: 2265611. PubMed Central PMCID: PMC552224.

Gustafsson, Claes, Sridhar Govindarajan, and Jeremy Minshull. "Codon bias and heterologous protein expression." Trends in biotechnology 22.7 (2004): 346-353. PubMed PMID: 15245907.

Pechmann, Sebastian, and Judith Frydman. "Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding." Nature structural & molecular biology20.2 (2013): 237. PubMed PMID: 23262490. PubMed Central PMCID: PMC3565066.

Quax, Tessa EF, et al. "Codon bias as a means to fine-tune gene expression." Molecular cell 59.2 (2015): 149-161. PubMed PMID: 26186290. PubMed Central PMCID: PMC4794256.

Tuller, Tamir, et al. "Translation efficiency is determined by both codon bias and folding energy." Proceedings of the National Academy of Sciences 107.8 (2010): 3645-3650. PubMed PMID: 20133581. PubMed Central PM- CID: PMC2840511.

Optimizing Plasmid Yield

By Tyler Ford | September 2018

Most of the time, plasmid prepping is a breeze. You get your stab from Addgene, streak for single colonies, subculture, 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/µL, 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. Here, we will outline many of the variables that could afect DNA yields and suggest steps to supercharge 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 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 propage unstable plasmids, but are also endA+. 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 purifcation 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 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. 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.

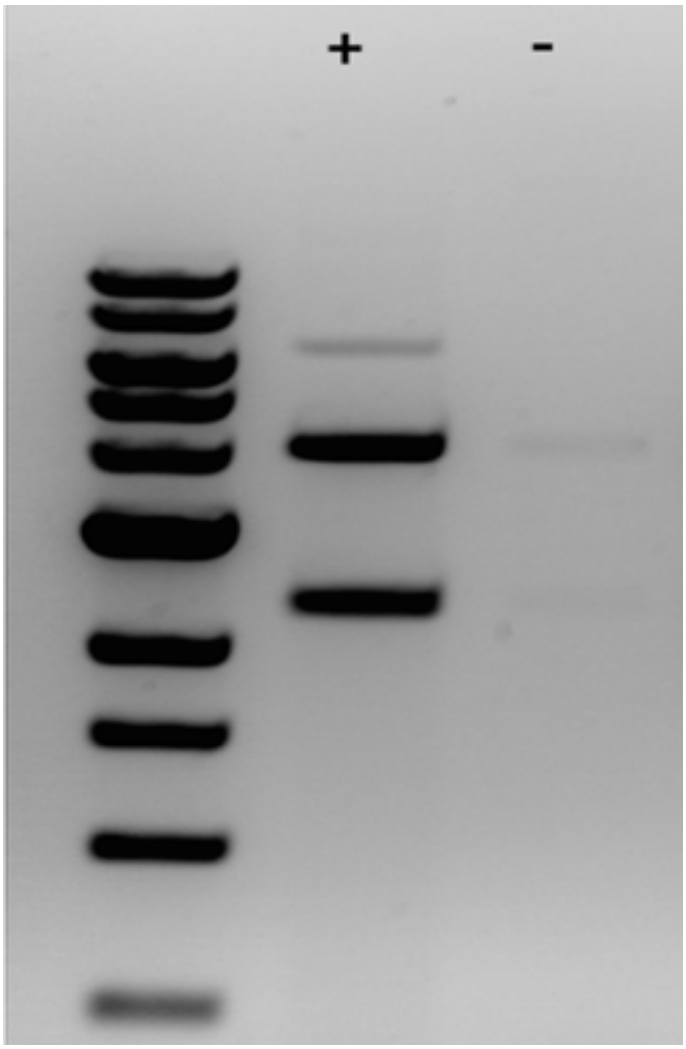


Figure 1

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 Terrifc Broth, 2xYT, or a home-brew optimized for plasmid yields. 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/mini prepping at various time points.

Generally, large cultures in fasks 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).

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

Sigma Microbial Media Guide

BitesizeBio: How to Get Better Plasmid Midiprep Yields

BitesizeBio: Better Plasmid Midipreps Part II: What Causes Low Yield

Control Plasmids

By Chari Cortez, | April 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.

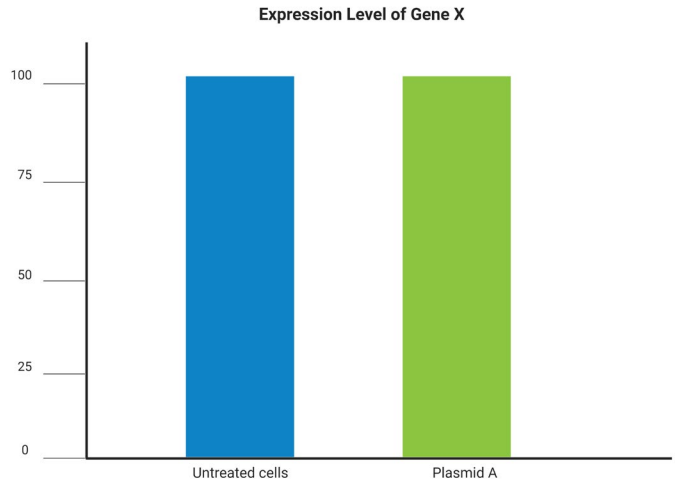


Figure 1: Experiment – attempt to knock down the expression of Gene X using shRNA expressed from plasmid A

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.

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 was delivered into the

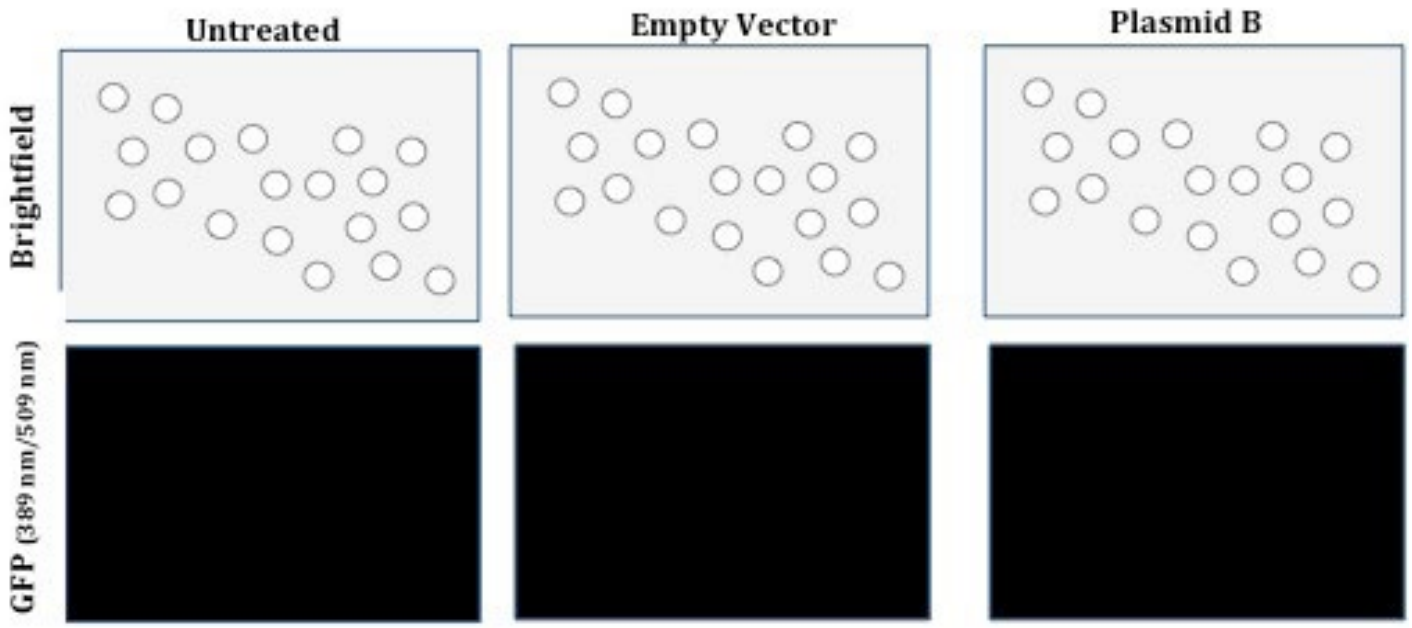


Figure 2: shRNA knock down experiment transfection control

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 controls

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

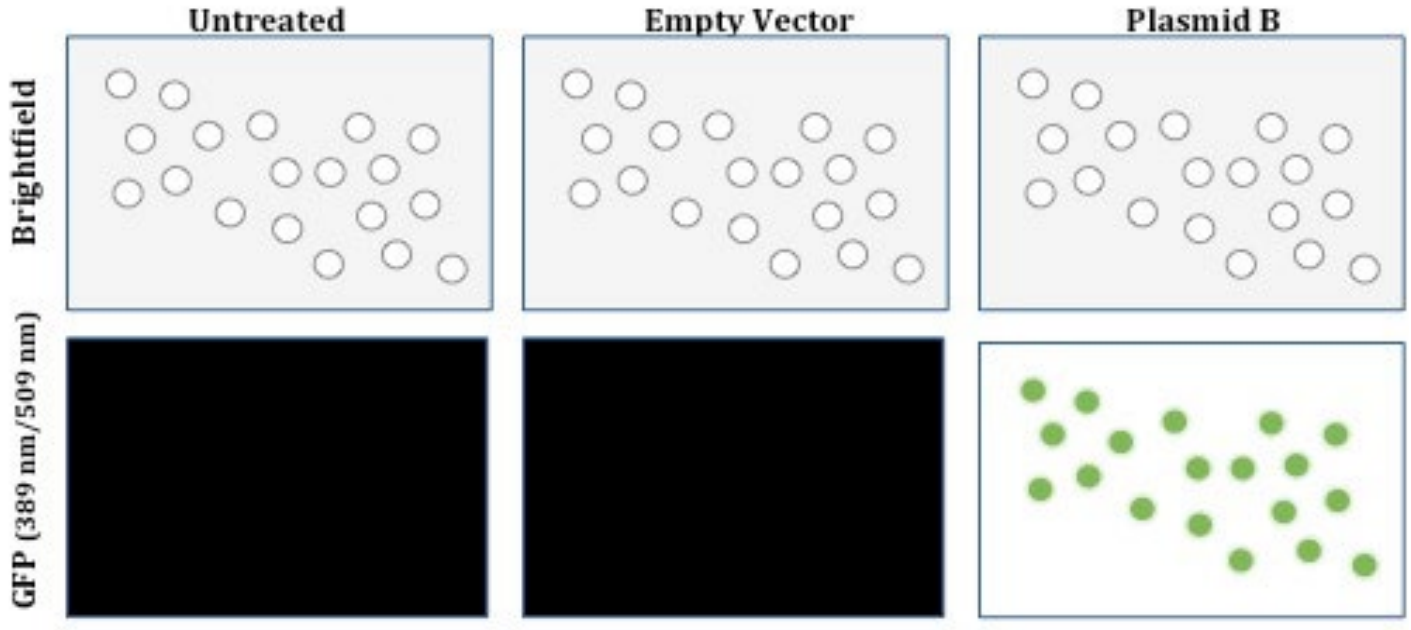


Figure 3: Successful transfection control

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, fuorescence microscopy images resulting from our experiment that includes the aforementioned internal control and is consistent with the result in Figure 1 could look 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 controls

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.

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.

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

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.

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

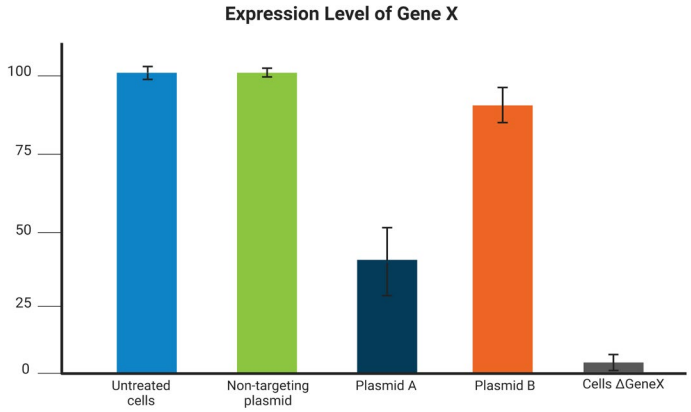


Figure 4: Measuring the expression level of gene X with all appropriate controls

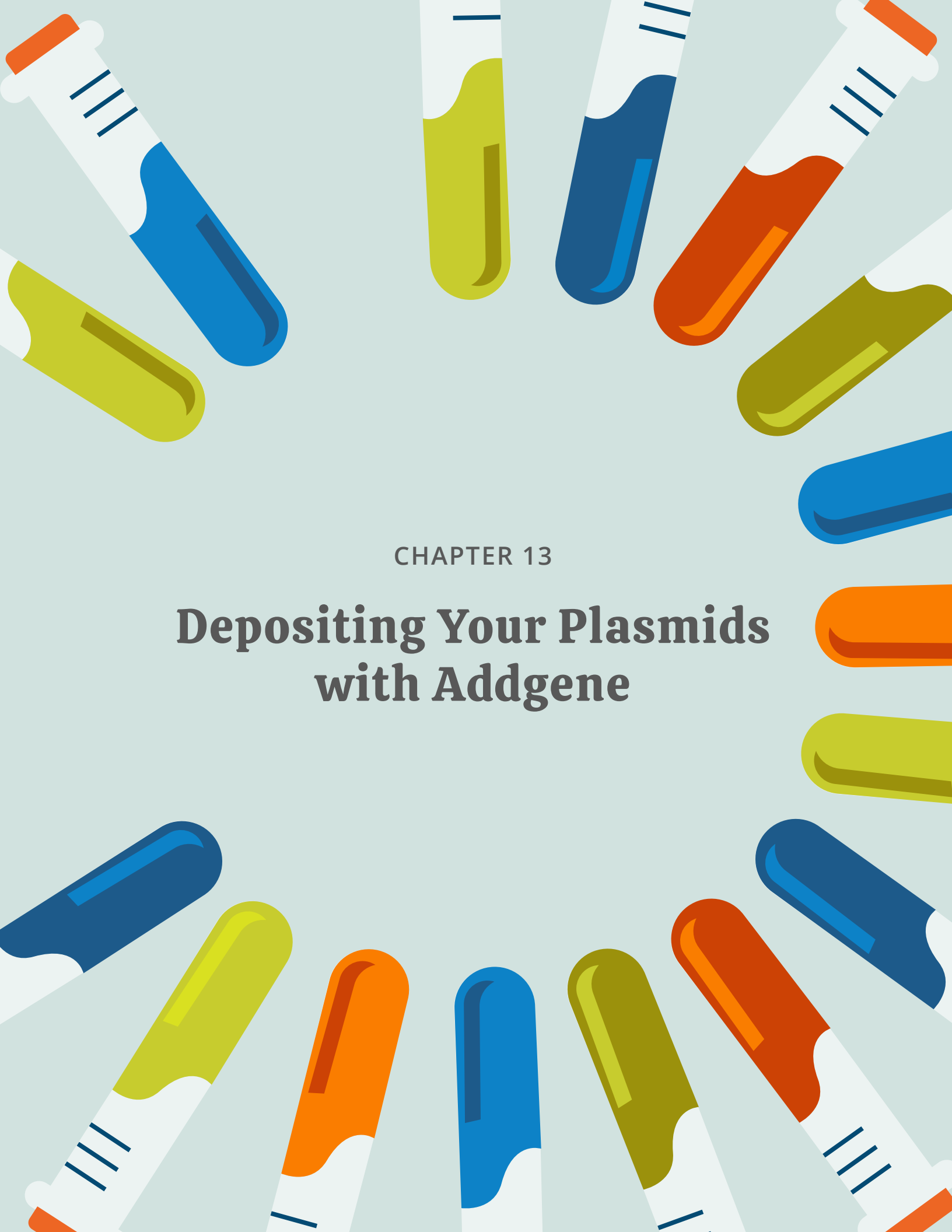
conditions. Now that we have successfully transfected our cells, we can continue with our experiment, incorporating additional positive and negative controls, performing multiple replicates, and ultimately getting interpretable results (Fig. 3).

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

The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? Lazic SE. BMC Neurosci. Jan 14;11:5. (2010). PubMed PMID: 20074371.

Replicates and repeats – what is the difference and is it significant? Vaux DL, et al. EMBO Rep. Apr; 13(4): 291-296. (2012). PubMed PMID: 22421999.



CHAPTER 13

Depositing Your Plasmids with Addgene

A Brief History of Addgene

Melina Fan | 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 (PIs), 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 2,500 labs from around the world have contributed over 42,000 plasmids for use in a wide variety of model organisms and research applications. In turn, we've helped those labs fulfill over 500,000 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

Benefits of Depositing

Various Addgenies | January 2023

SAVE TIME AND MONEY

Depositing plasmids at Addgene is like getting a free lab manager. There is no charge to deposit, and you will no longer have to find stocks, package materials, ship packages, and deal with MTAs. Plus, as your plasmids are requested from Addgene, you will earn Rewards Points that can be redeemed for free plasmids.

Addgene creates a website for each depositing lab, where you can direct plasmid requests. Addgene processes and ships your plasmids for you, and our knowledgeable and friendly support team answers questions about your deposited plasmids.

Addgene's one-time deposit agreement means that once your plasmids are online, we'll facilitate agreements (like MTAs!) between your organization's tech transfer office and that of your requesters so that you don't have to.

STAY ORGANIZED

How many times have you tried in vain to find a plasmid created by a former member of the lab? To remember the cloning sites used to create a plasmid? Or worried about losing your stocks from a freezer meltdown? Deposit plasmids at Addgene and relax. We will preserve and curate your samples and associated data for easy online access. Addgene stores samples in barcoded tubes on-site in Watertown, MA and off-site at a facility in Rockville, MD. You can easily download a CSV file with a complete list of the scientists who have requested your plasmids through your Addgene account.

SHARE WITH YOUR COMMUNITY

You want your research to make a difference. Sharing your plasmids allows other scientists to build upon your work. This is why NIH grants now require a resource sharing plan and most journals require all published materials be shared with other scientists. Addgene is recommended by over 20 journals and is actively used by many top scientists.

After depositing to Addgene, your materials will be searchable and accessible to other scientists. Addgene groups plasmids by article and by gene. Scientists who use your plasmids in a publication are asked to acknowledge you and cite your article, increasing the visibility of your research.

IT'S EASY TO DEPOSIT!

Depositing plasmids at Addgene is like answering a plasmid request, only better. Once available through Addgene, you will never have to send that plasmid out again. Addgene accepts either DNA (15 µL is plenty) or a bacterial streak of your sample. If you would like assistance with your deposit, just email deposit@addgene. org and

ask to be paired with an Addgene scientist. We can help you find plasmid information, enter data, and answer any questions.

You can start the deposit process on our website at any time! ■

The Deposit Spreadsheet

Various Addgenies | January 2023

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.

Plasmid Name	Plasmid Type	Purpose	PubMed ID	Gene or insert Name	Alternative Gene/Insert Name 1	Alternative Gene/Insert Name 2	Entrez Gene ID
		What does this plasmid do? Please enter a brief description (<200 characters).					

THE DATA ENTRY PROCESS

On the Addgene home page, if you scroll over the "Deposit" tab at the top of the page, you will see two options for data entry: **Online Submission** and **Spreadsheet Submission**. Selecting 'Online Submission' will direct you to submit plasmids from a published article or pre-publication/ unpublished plasmids. The online data entry process will walk you through that submission process step by step.

If you wish to deposit 10 or more similar plasmids, the easiest option for submitting your plasmids is using our deposit spreadsheet. Simply download the spreadsheet and copy and paste your plasmid data directly into our file, one plasmid per row, as outlined by the column headers in the spreadsheet. Required columns are shown in red on the spreadsheet and are described in Table 13.1 Some cells have drop-down options for you to choose from, while others need to be filled with short answer text. Providing sequences (either in the spreadsheet or as separate GenBank files) will help reduce Addgene's QC time and may make plasmids available sooner.

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 75 plasmids or have additional questions, please contact help@addgene.org.

When you have 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 with any associated plasmid maps and sequence files in any format as well as your 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. ■

Table 13.1

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
Species of Gene or Insert	Choose from *H. sapiens* (human), *M. musculus* (mouse), *R. norvegicus* (rat), *G. gallus* (chicken), *B. taurus* (bovine), *O. cuniculus* (rabbit), *X. laevis* (frog), *D. rerio* (zebrafish), *D. melanogaster* (fly), *C. elegans* (nematode), *S. cerevisiae* (budding yeast), *S. pombe* (fission yeast), *A. thaliana* (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 wildtype (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
Bacterial Resistance	Enter bacterial antibiotic resistance encoded in the plasmid choose from: ampicillin, apramycin, blasticidin, bleocin(zeocin), chloramphenicol, erythromycin, gentamycin, hygromycin, kanamycin, nourseothricin(clonNat), spectinomycin, streptomycin, tetracycline, 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 30 °C, 37 °C, or room temperature.
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 licenses that could prohibit Addgene from distributing the plasmid, select “yes” from the drop-down and contact Addgene.

ACKNOWLEDGEMENTS AND FINAL WORDS

If you have any questions, comments, or suggestions about how Addgene can improve its educational content, please contact us at news@addgene.org.

~ The Addgene Team

Find us

Addgene.org

blog.addgene.org

facebook.com/addgene

twitter.com/addgene

instragram.com/addgene

youtube.com/@addgene

linkedin.com/company/addgene

Special thanks to our guest author:

Gal Haimovich, greenfluorescentblog

Addgene Contributors

AUTHORS

Alyssa Cecchetelli
Chari Cortez
Michelle Cronin
Hannah Dotson
Melina Fan
Matthew Ferenc
Tyler J. Ford
Mary Gearing
Benoit Giquel
Larissa Haliw
Amanda Hazen
Jane Hannon
A. Max Juchheim
Beth Kenkel
Caroline LaManna
Michael G. Lemieux
Rachel Leeson
Emma Markham
Margo R. Monroe
Kendall Morgan
Jason Niehaus
Matt Ouellette
Marcy Patrick
Eric J. Perkins
Brook Pyhtila
Leah Schwiesow
Maria Soriano
Lianna Swanson
Julian Taylor-Parker
Jennifer Tsang
Nicole Waxmonsky
Jessica Welch

EBOOK EDITORS

Rachel Leeson
Jennifer Tsang
Caroline LaManna

ADDITIONAL THANKS TO

Steven DiPrizio
Jo Dwyer
Sarah Gordan
Angela Holmes
Christina Mork
Mike Lacy

DESIGN AND LAYOUT

Jason Snair
Eileen Wu

addgene

Addgene's eBook Collection

Plasmids 101

July 2023

4th Edition