Biological Parts (Protein expression and plasmids)

Clay Wright
BSE3984

Plasmids, the basic unit of recombinant DNA

Plasmids are:

- small, extrachromosomal DNA molecules within a cell
- can replicate independently
- are most commonly found as small circular, double-stranded DNA molecules in bacteria, but also exist in other forms and in archaea and eukaryotes

Molecular cloning = assembly of recombinant DNA molecules typically in a plasmid to facilitate their replication within different host organisms

<u>A note on terminology:</u>

Plasmids are sometimes called vectors because they carry (vector) your gene of interest.

Vectors can also be used to integrate into the genome. Once in the genome, origin of replication and selection marker are no longer needed.

Construct is also a more general term for any piece of recombinant DNA.

Construct>Vector>Plasmid

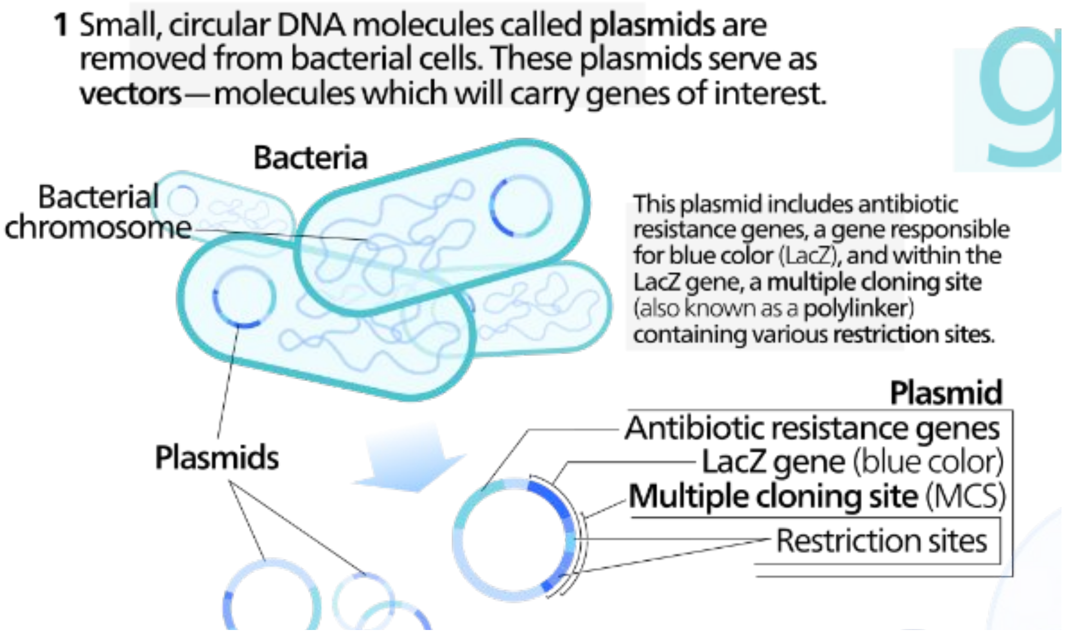
Necessary parts of a plasmid

A plasmid must contain:

- selection marker (antibiotic resistance)
- origin of replication (ORI)

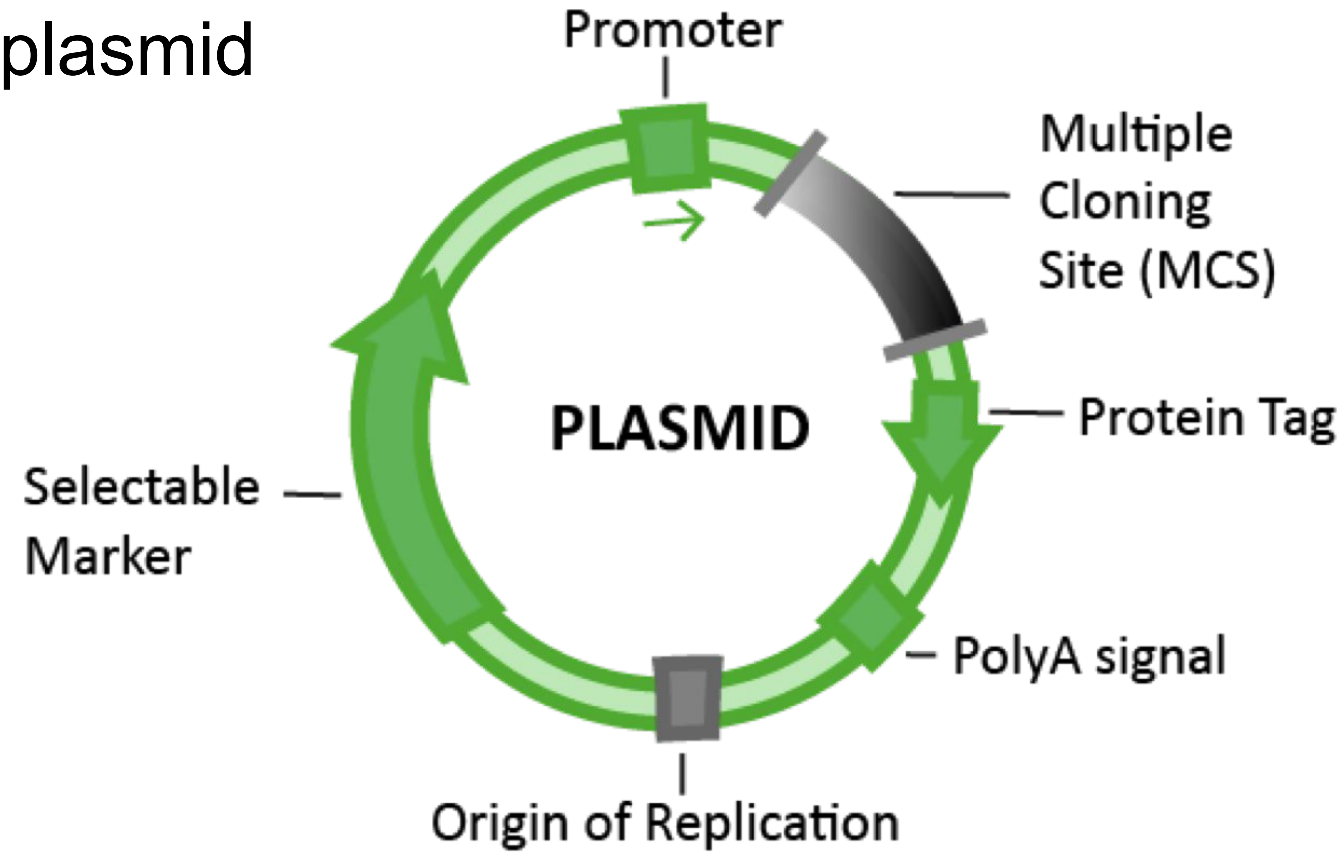
Plasmids can contain:

- Your gene of interest
- Expression cassettes
  - Promoter
  - Coding Sequence (CDS)
  - Terminator
- Restriction sites
- Any other useful DNA sequence

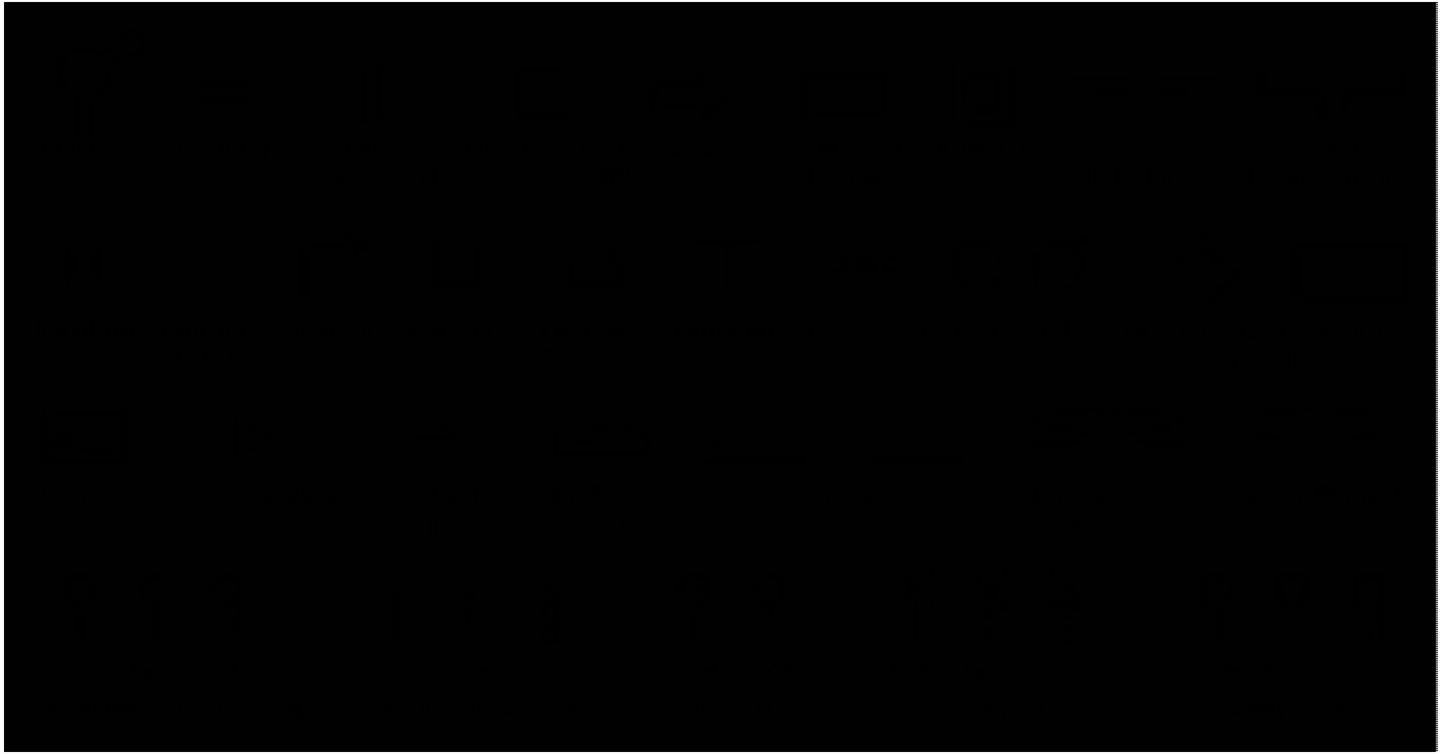


https://commons.wikimedia.org/wiki/File:Gene_cloning.svg

A typical plasmid



Synthetic Biology Open Language (SBOL) provides common, consistent symbols (Glyphs)



Origin of Replication - must have

- Allows a DNA polymerase to replicate the plasmid
- Can be high, medium, or low copy number
- High copy number makes more protein than low
- Origins of replication can be difficult for PCR polymerases to process through

Category	Typical number of copies per bacterial cell
Low copy (e.g. pBR322 and derivatives)	15–20 copies per cell
Medium copy	20–100 copies per cell
High copy (e.g. pUC18, pUC19 vectors)	500–700 copies per cell

https://bitesizebio.com/22824/how-to-manipulate-plasmid-copy-number/

Growth Selections

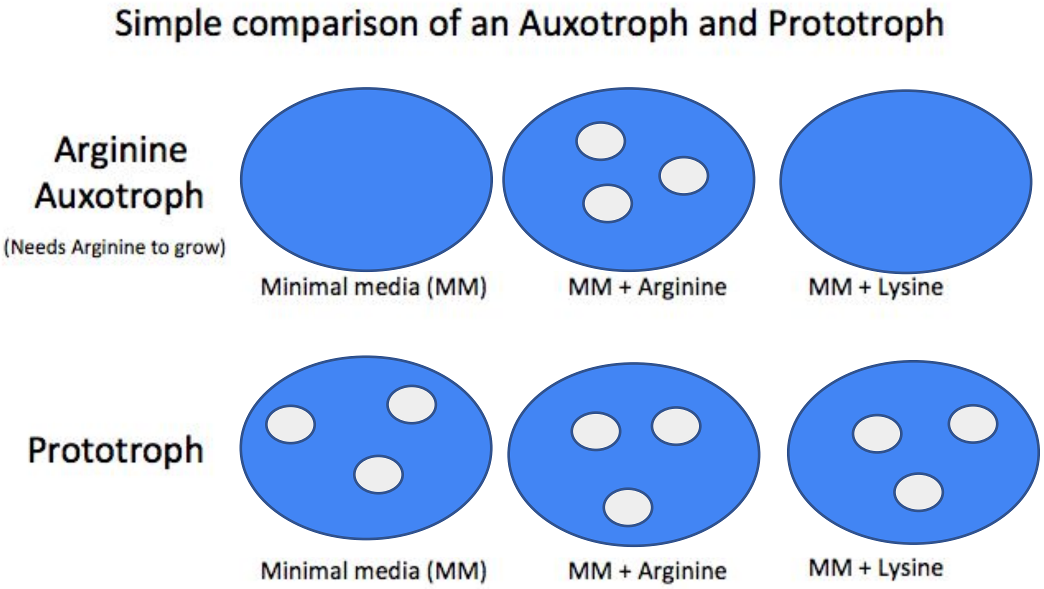
Without something forcing the cells to maintain the plasmid it may get lost when cells divide.

- Toxin Resistance
  - Antibiotic resistance
    - Used to maintain plasmids (extrachromosomal DNA) in bacteria (and sometimes yeast)
- Auxotroph/Prototroph
  - Auxotroph = a strain that has lost the ability to produce a vital nutrient, but that can grow on media supplemented with that nutrient
  - Prototroph = an auxotrophic strain that has been supplemented with the gene to produce the lacking nutrient, allowing it to grow on media lacking that nutrient
    - Used to maintain plasmids in yeast and bacteria

Auxotrophy-Prototrophy

Arginine biosynthesis Knock Out

Complement With your gene of interest + Arginine biosynthesis



Expression cassettes

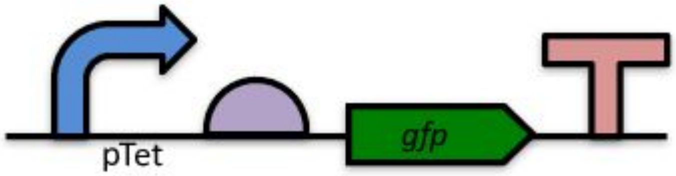
(Cassette is sometimes used to mean group of DNA elements, an operon is another group of DNA elements, both are protein coding sequences and regions regulating them)

Must haves:

- Promoter (pTet)
- Ribosome entry/binding site or Kozak sequence (often in promoter)
- Coding sequence (CDS)

Ideal to have but not *totally* necessary

- Terminator



Promoters 

Constitutive or Inducible?

- Constitutive means it is constantly produced/expressed
  - However “constant” is almost never truth, and promoters change expression with the growth conditions and metabolism of the organism
  - Can have very different levels (high, medium, and low, spanning several orders of magnitude in RNA/protein concentration)
- Inducible means you can increase expression by changing conditions
  - Adding some chemical
    - Sugars - lactose, galactose, sucrose, arabinose
    - Other chemicals - ethanol, copper, antibiotics (tetracycline)
    - Changing conditions - temperature, pH
- Promoters contain
  - Minimal promoter - required for entry of RNA polymerase
  - Cis-elements - are what transcription factors bind to recruit RNA polymerase
  - Intervening sequence - may not have a known function, may space out transcription factors

Ribosome entry/binding sites and Kozak sequences 

- Differ between prokaryotes and eukaryotes (and species)
- All help to assemble the ribosome and get it to begin translating at the right start codon
- Sometimes RBS sites are difficult to identify
  - they vary quite a bit,
  - the perfect RBS is 5'-AGGAGG-3' (literally the complement of the ribosomal RNA)
- In eukaryotes the 5’ cap of mRNA is where the ribosome binds
- the Kozak sequence is immediately surrounding the start codon and determines where translation starts

These sequences can have as much effect on protein expression as promoter

Coding Sequences (CDS)

- Coding sequences code for proteins
- They contain an Open Reading Frame (ORF)
  - ORF = a series of codons going from a start codon to a stop codon
  - Software like ApE (a plasmid editor), vector editor, benchling, etc, can find ORFs for you and translate them
- Can be translated using a codon table into the amino acid sequence for the protein they make

Advanced Coding sequences 

Often times proteins and their functions are what we want to measure as biological engineers

How do we measure proteins in cells?

Reporters proteins provide some output that we can measure most often via spectroscopy of some sort (UV/Vis/Fluorescent/Luminescent)

We can create “fusion” or “hybrid” proteins by removing the stop codon from one CDS and adding another CDS after it

Typically a “linker” will be added in between, coding for 5-15 flexible (small, hydrophilic) amino acids such as GGSGG

Fluorescent proteins are often fused to proteins to determine where that protein is in the cell, how much of that protein is in the cell

Introns

In eukaryotes introns can regulate RNA processing and protein synthesis

Positively or **Negatively**

In plants an intron, such as the Ubi1 intron, even placed before the coding sequence in an mRNA can increase expression 10-fold

Often times introns will be removed, particularly when moving genes from one organism to another as intron processing (RNA splicing) may be different

Terminators 

Terminate transcription

Not a lot is know about them, but they can dramatically change expression levels, perhaps through changing RNA stability

It’s typical to use ~500 bp downstream of a common coding sequence in the organism you are trying to engineer

The organism you are trying to engineer is called the “host” or “chassis”