UNIVERSITY OF CAMBRIDGE

Protein expression systems

Day 1: Monday 20th March

Proteins are polymers of amino acids

- Protein sequence is encoded in genes
- Three nucleotides encode each amino acid

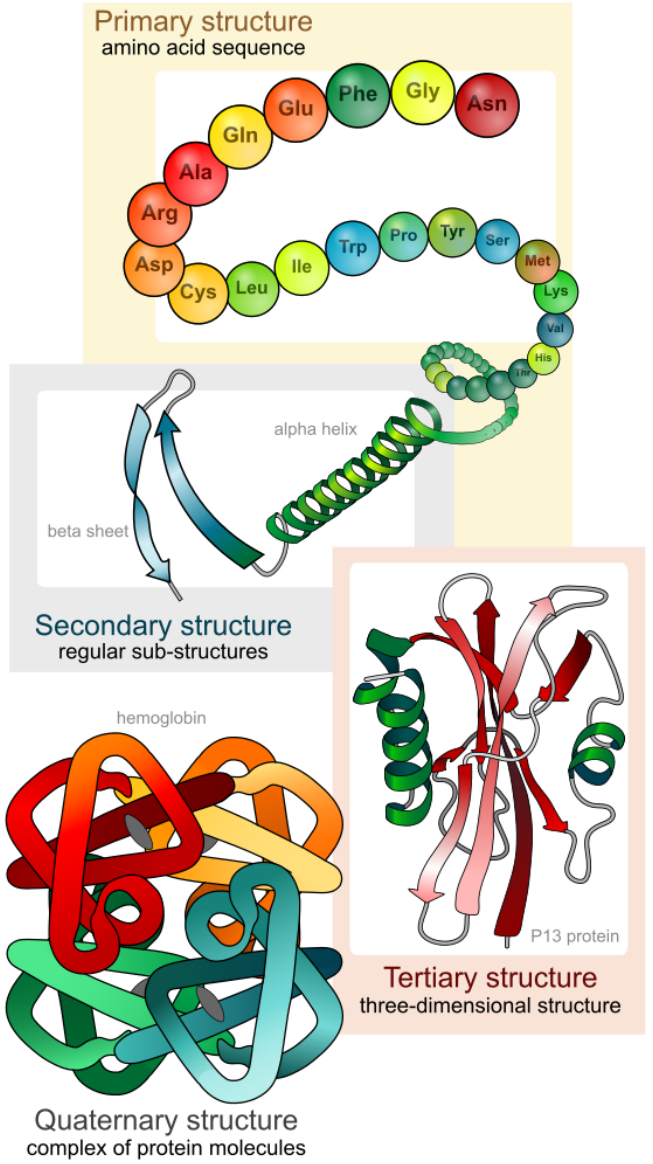
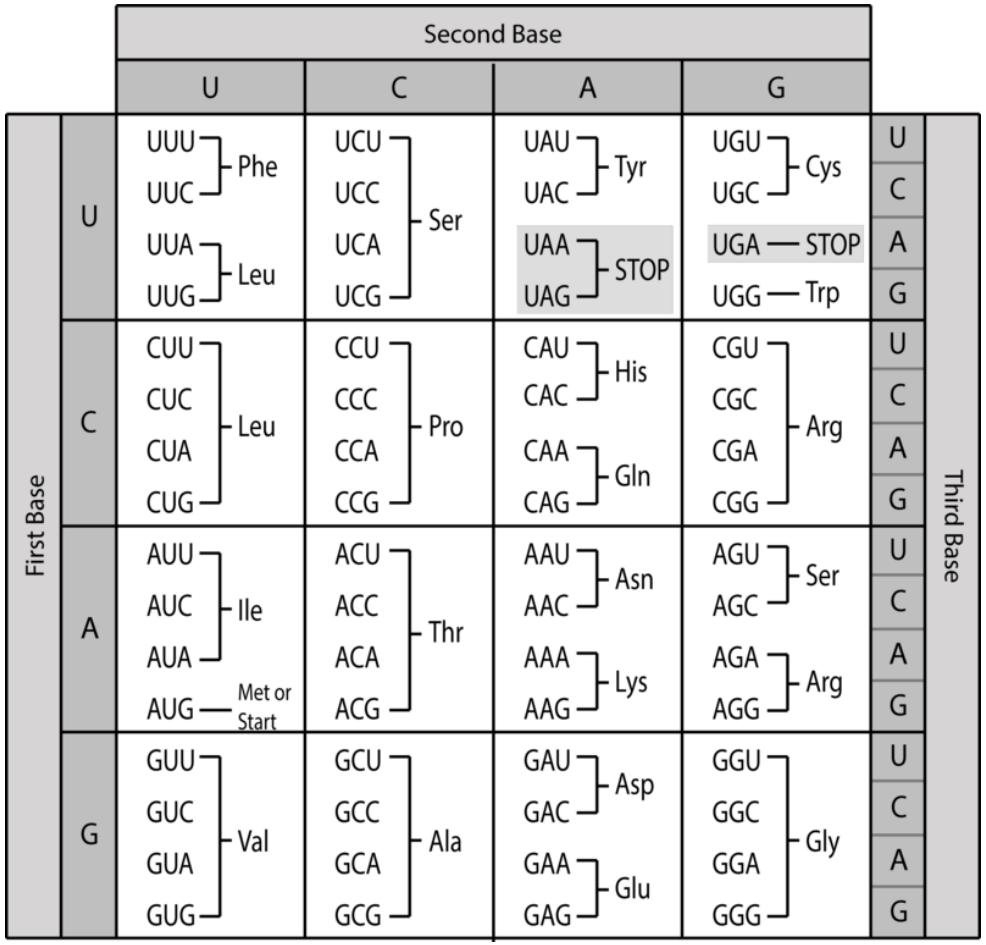


Image: LadyofHats (Wikimedia)

Proteins are polymers of amino acids

- Protein sequence is encoded in genes
- Three nucleotides encode each amino acid
- Genetic code is degenerate:
  20 naturally occurring amino acids but 64 triplex codons
  - Genetic code identical in many organisms but not all

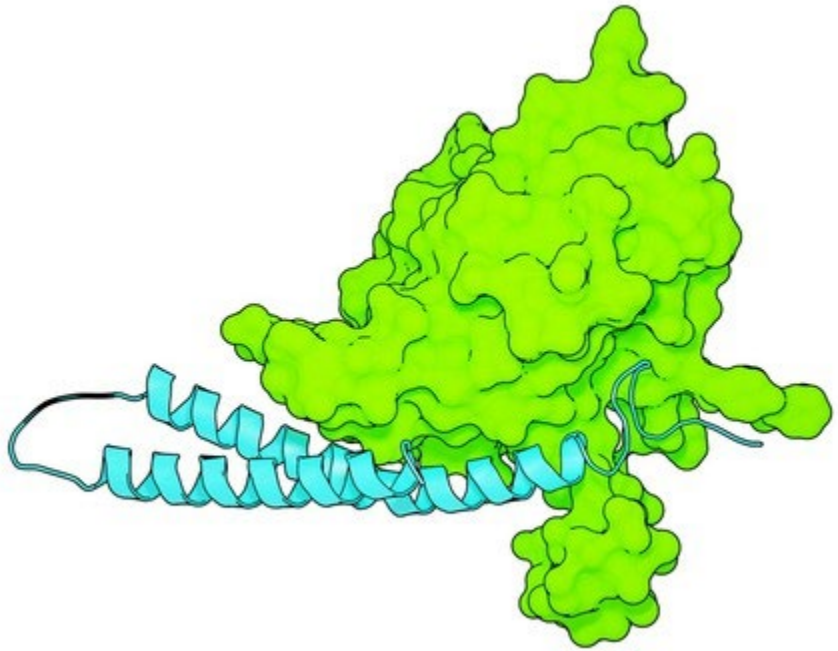


First Base	Second Base: U	Second Base: C	Second Base: A	Second Base: G	Third Base
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA STOP	UGA STOP	A
	UUG Leu	UCG Ser	UAG STOP	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met or Start	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Image: Sarah Greenwood (Wikimedia https://commons.wikimedia.org/wiki/File:Genetic_Code.png)

Why study proteins?

- Proteins are the engines of life, responsible for many of the interactions and chemical reactions that sustain cells and organisms
- Understanding the structure and interactions of proteins can tell us about their functions
- Biochemical studies of function require highly purified components (proteins)
  - Many proteins are expressed at low abundance in their native context



This talk

- Expression vectors and the basics of molecular cloning
- Expression systems:
  - Bacteria
  - Mammalian cells
  - Others (Yeast, insect cells, plants, cell-free)

Expression vectors

- Expression vectors contain a gene of interest (**GOI**) that encodes your protein of interest (**POI**)
- The vector also contains regulatory elements (e.g. promoters) that ensure your **GOI** is expressed within the expression organism
- Vectors often have a selectable markers (e.g. antibiotic resistance) so you can select for expression organisms containing the vector
- The most common expression vector is the **plasmid**

Plasmids

- Small circular segments of DNA that can self-replicate within organisms, independently of the genome
- Bacterial plasmids are formed of double-stranded DNA
- In nature, plasmids enable the horizontal transfer of genetic material (including antibiotic resistance)

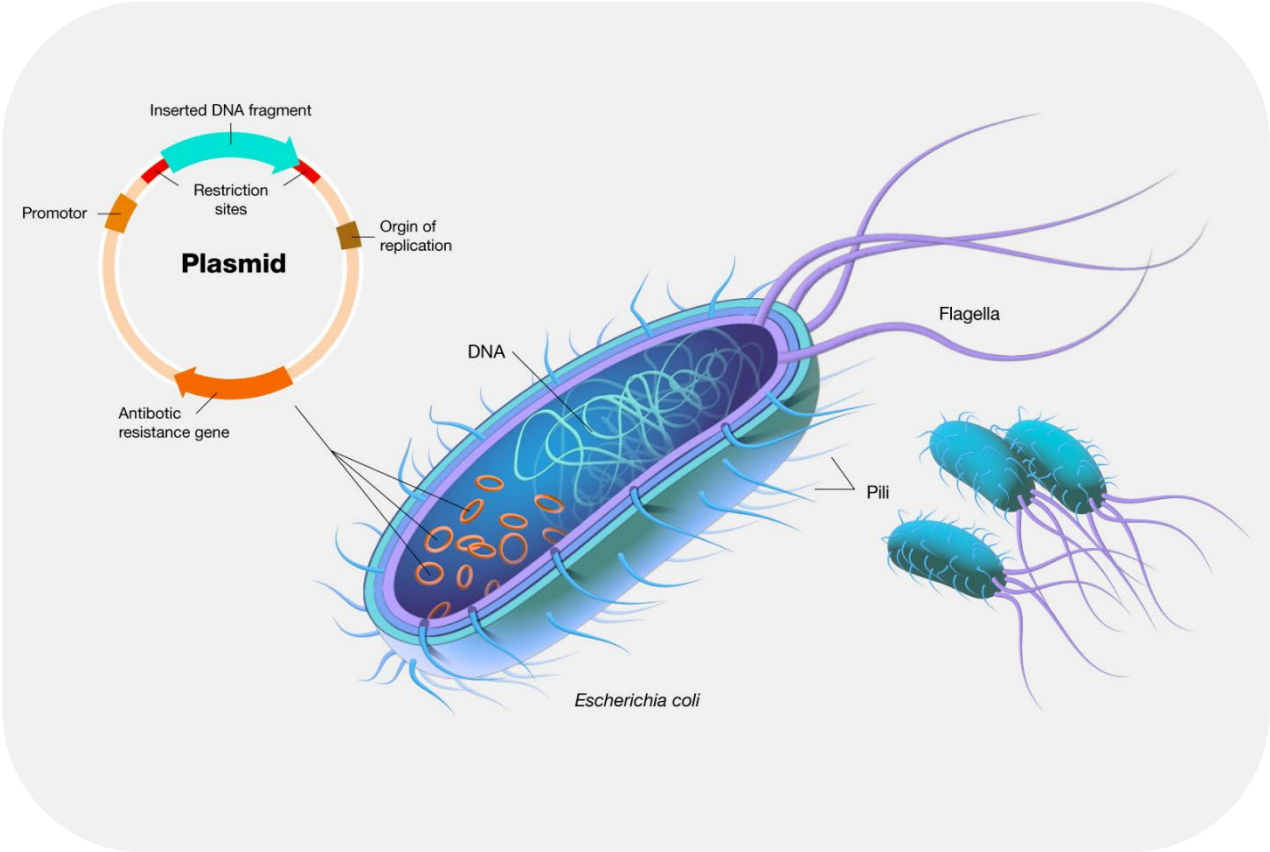


Image: NIH National Human Genome Research Institute

Plasmids

- Plasmids replicate within bacteria
  - Requires an origin of replication
- Generally for molecular biology we use *Escherichia coli* to amplify our plasmids
- Plasmids can be purified from away from *E. coli* chromosomal DNA using simple chemistry
  - Silica resins = midiprep kits
- Introducing a plasmid into a bacteria is called *transformation*

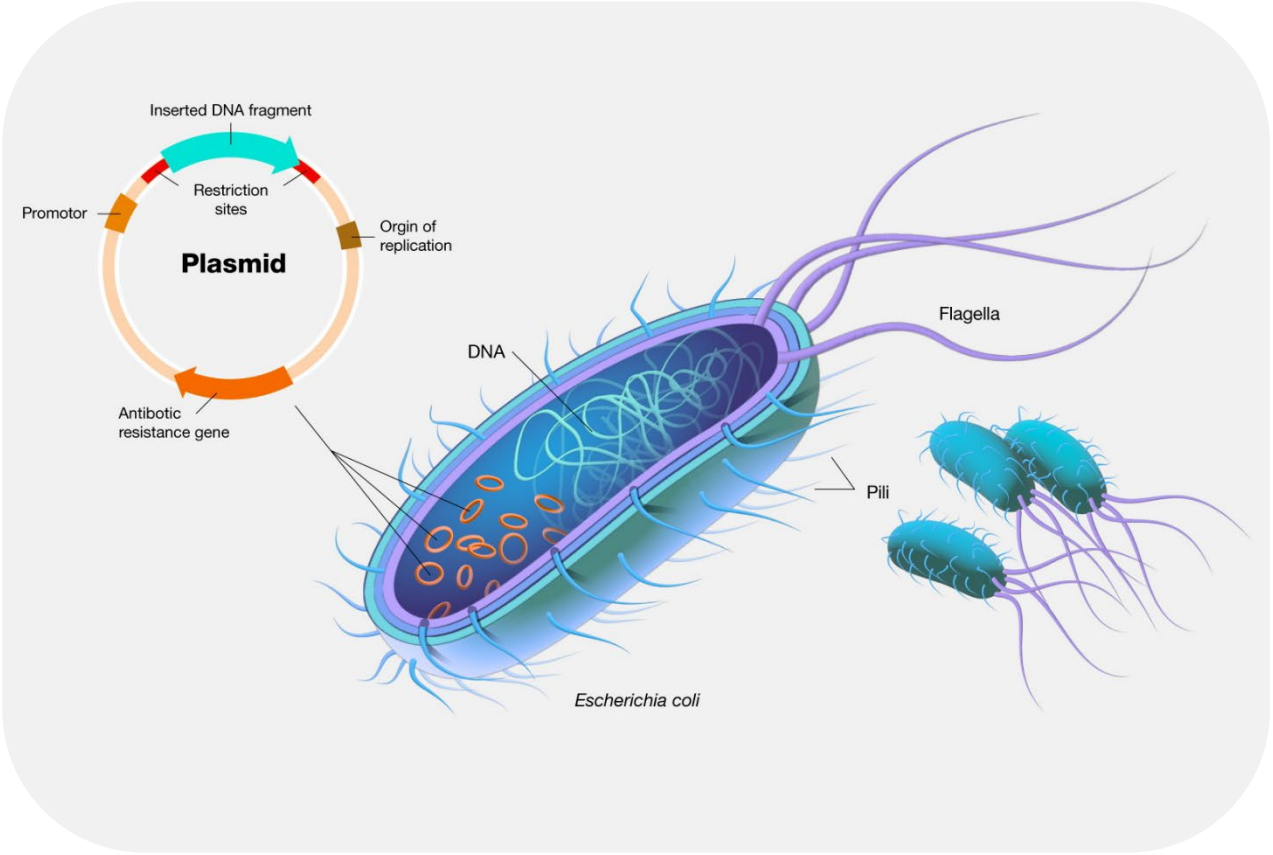
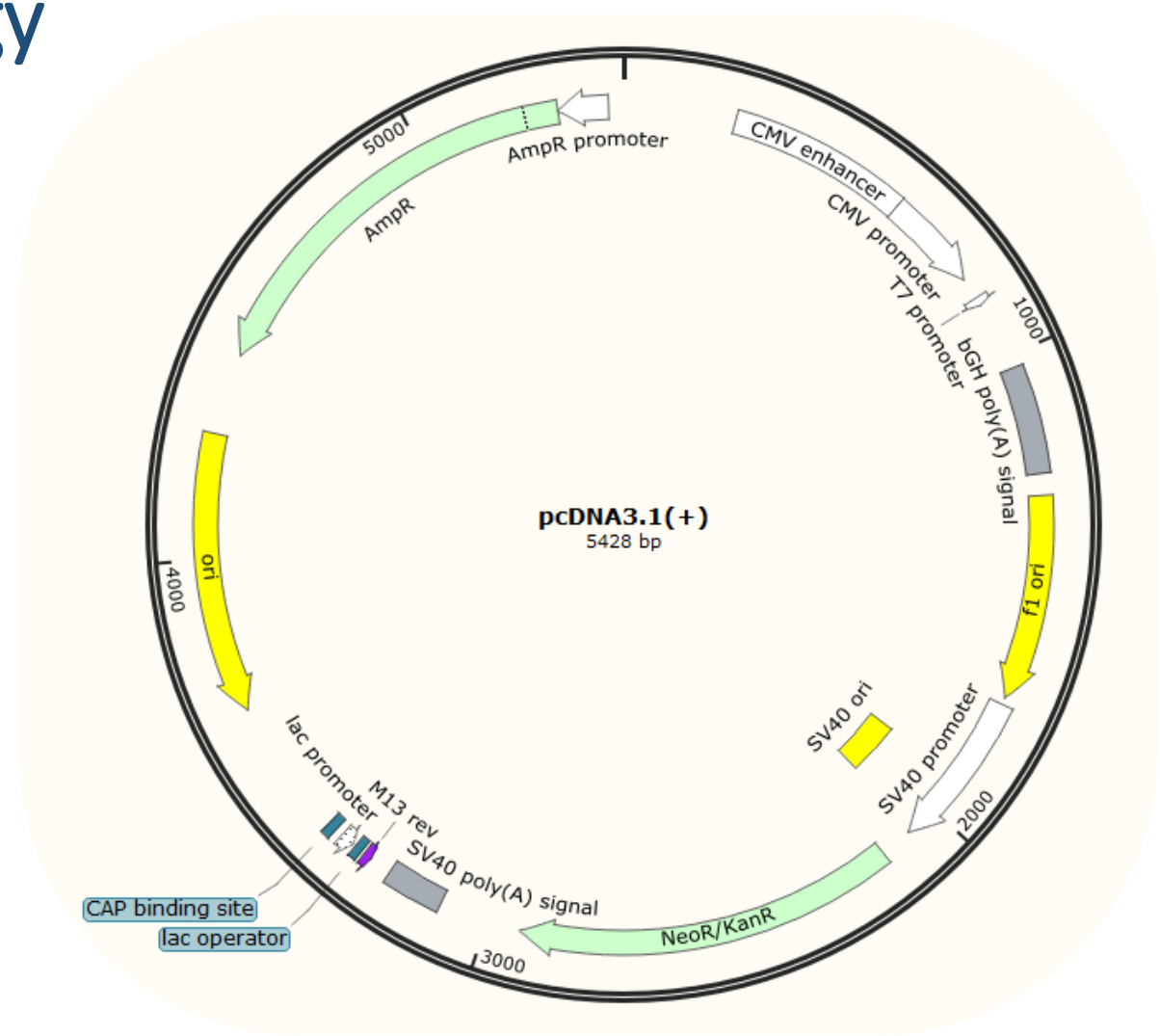


Image: NIH National Human Genome Research Institute

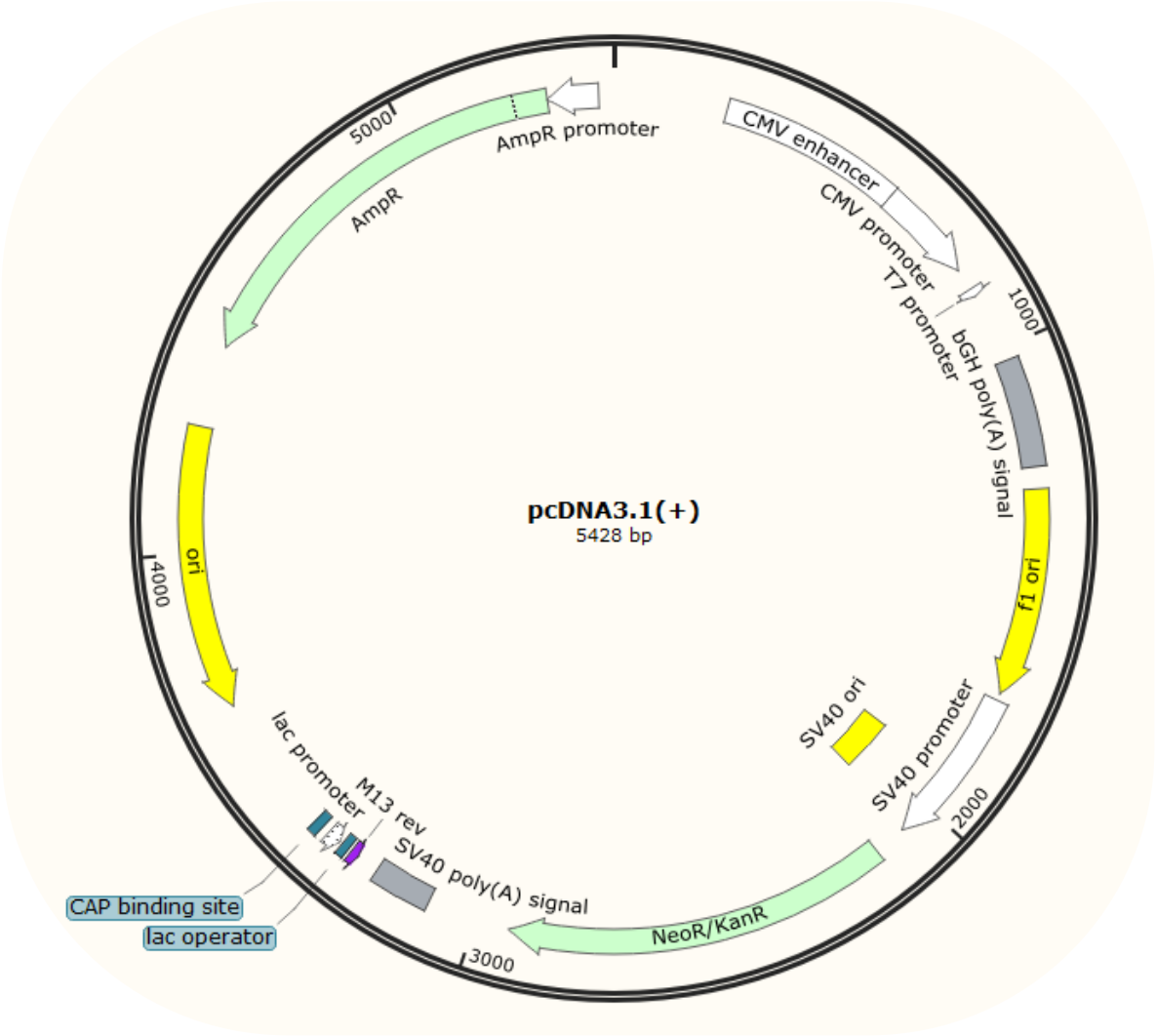
Plasmids for molecular biology

- Components:
  - Origin of replication
  - Selection marker(s)
  - Promoter(s)
  - Other DNA regulatory elements
  - Multiple cloning site(s)
  - Tags



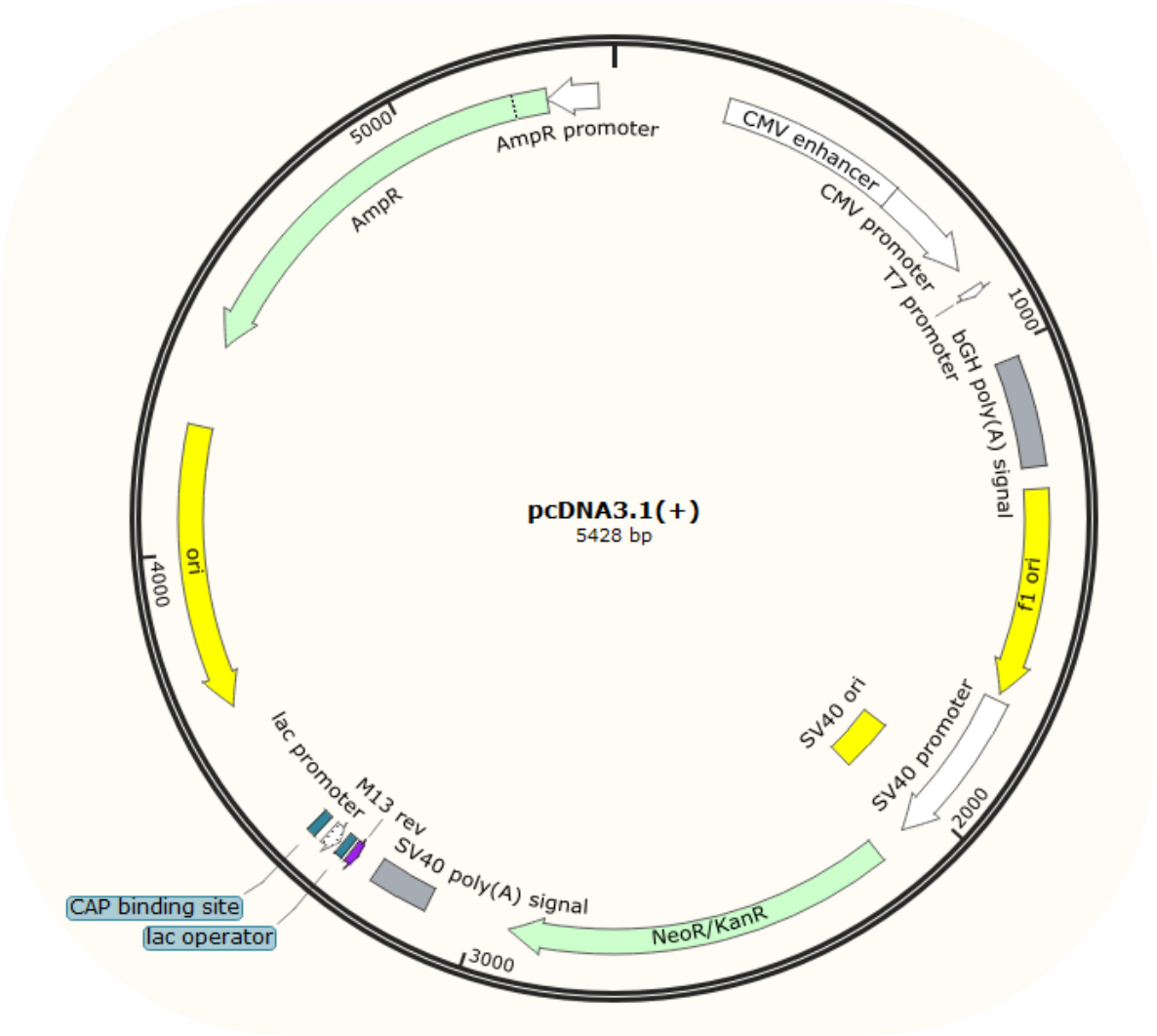
Origin of replication (*ori*)

- Promotes binding of the DNA replication machinery to copy the plasmid (replicon)
  - Proteins or RNA
- Bacterial origins of replication control the *copy number* of the plasmids
  - How many plasmid molecules per bacteria, from <10 to >700!
- The pcDNA3.1 vector has the **pUC** *ori*, which is *high copy number*



Origin of replication (*ori*)

- pcDNA3.1 also has the *ori* of the polyomavirus SV40
  - Can be replicated as an ***episome*** in the nucleus of mammalian cells transformed with the SV40 large T-antigen (e.g. HEK293T cells)
- And pcDNA3.1 has f1 *ori* for replicating single stranded DNA that can be packaged into phage particles (e.g. M13)



Origin of replication (*ori*)

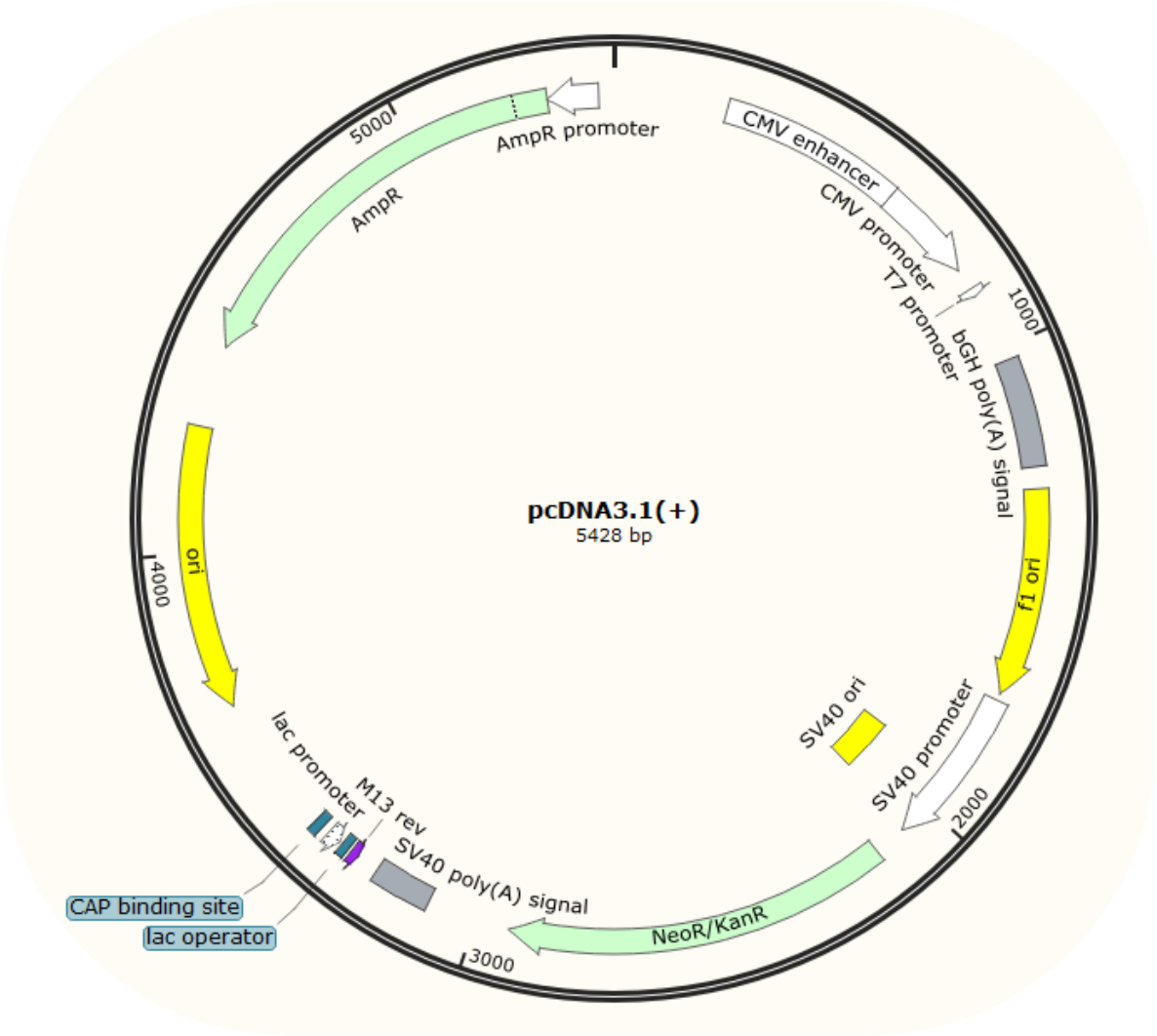
- You can co-transform multiple different plasmids into the one *E. coli*
- But bacterial origins of replication can ***compete*** for the same regulatory factors, leading to plasmid instability
  - Need to choose plasmids from different compatibility groups

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*	C	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

Table: Addgene (https://blog.addgene.org/plasmid-101-origin-of-replication)

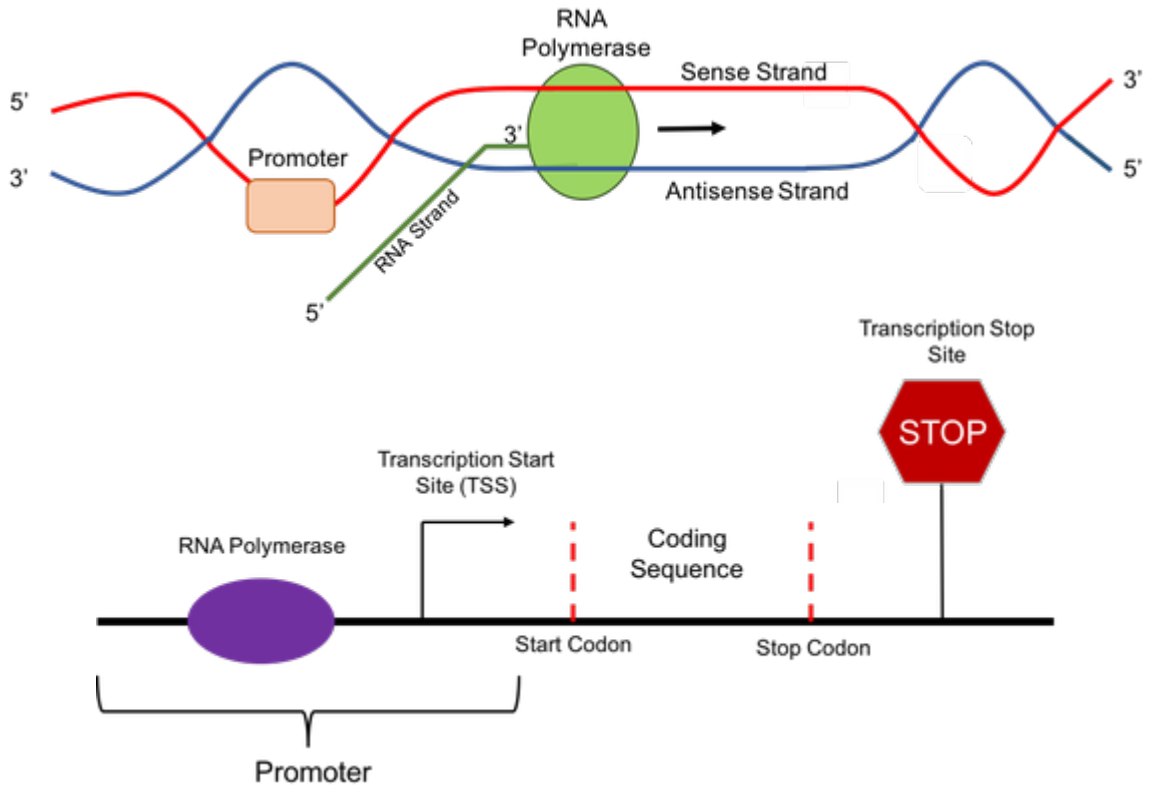
Selection markers

- Allow cells with/without the plasmid to be distinguished
- For bacteria, often genes that encode antibiotic resistance
  - Antibiotic will kill all bacteria that lacking the plasmid (*lethal selection*)
- pcDNA3.1 has the β-lactamase (*bla*) gene for resistance to ampicillin
  - Other common selection antibiotics are chloramphenicol and kanamycin



Promoters

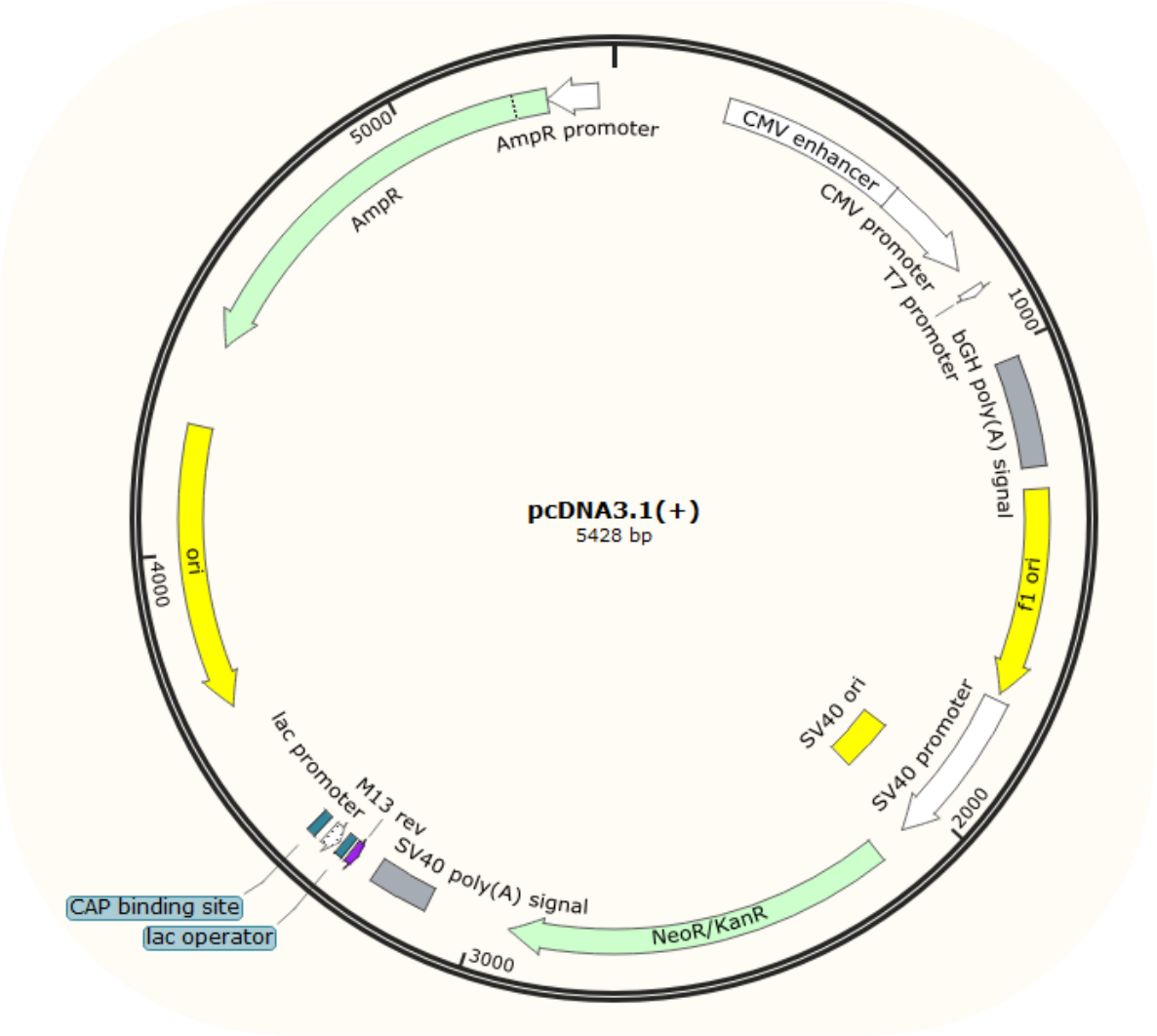
- Promoters are the regions that control the expression of downstream (3’) genes
  - *Cis* regulatory elements
- Represent the binding site for transcription factors and the RNA polymerase complex
  - These differ between organisms



Images: Addgene (https://www.addgene.org/mol-bio-reference/promoters/)

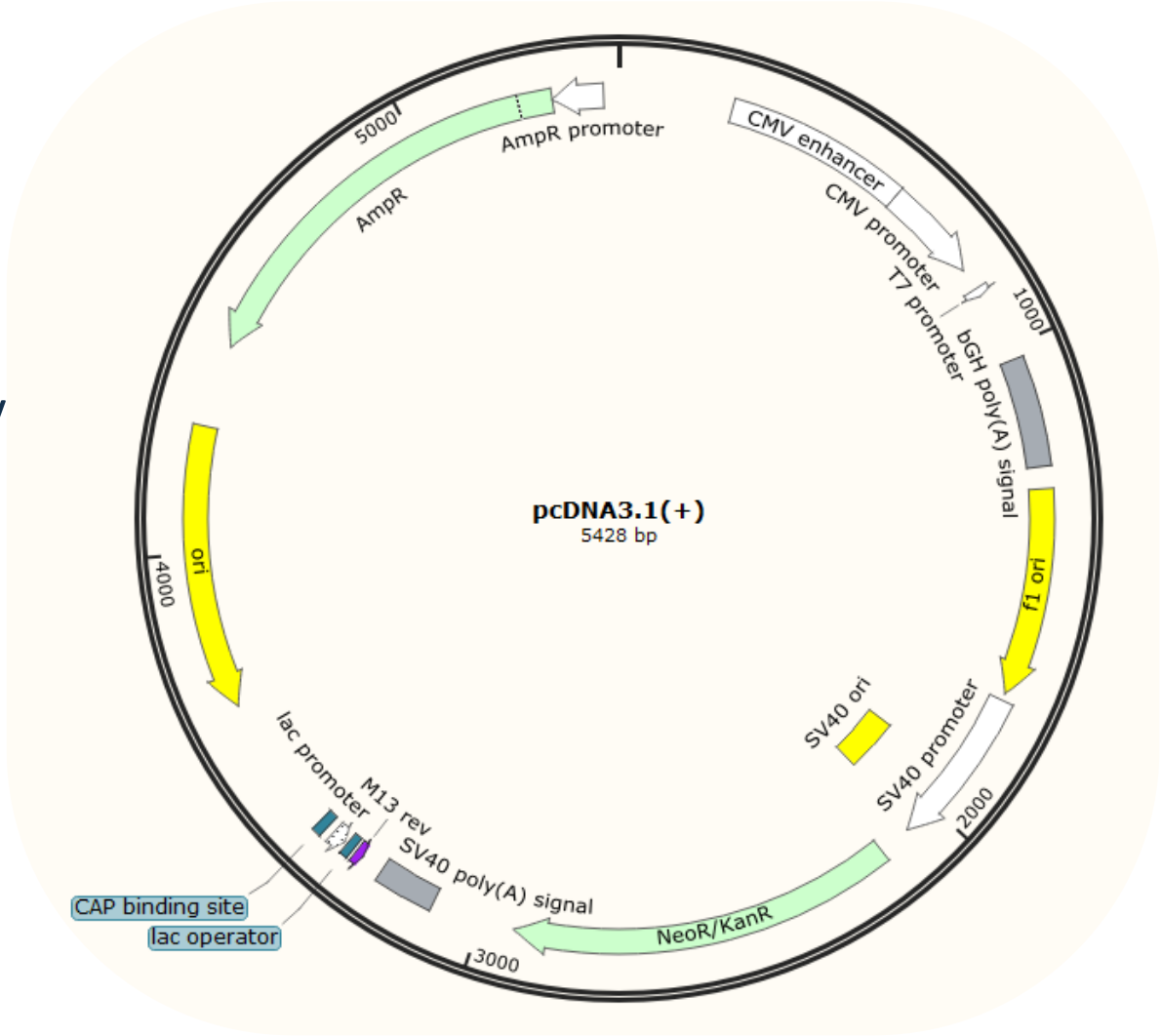
Promoters

- Promoters are the regions that control the expression of downstream (3’) genes
  - *Cis* regulatory elements
- Represent the binding site for transcription factors and the RNA polymerase complex
  - These differ between organisms
- pcDNA3.1 has 5 different promoters!



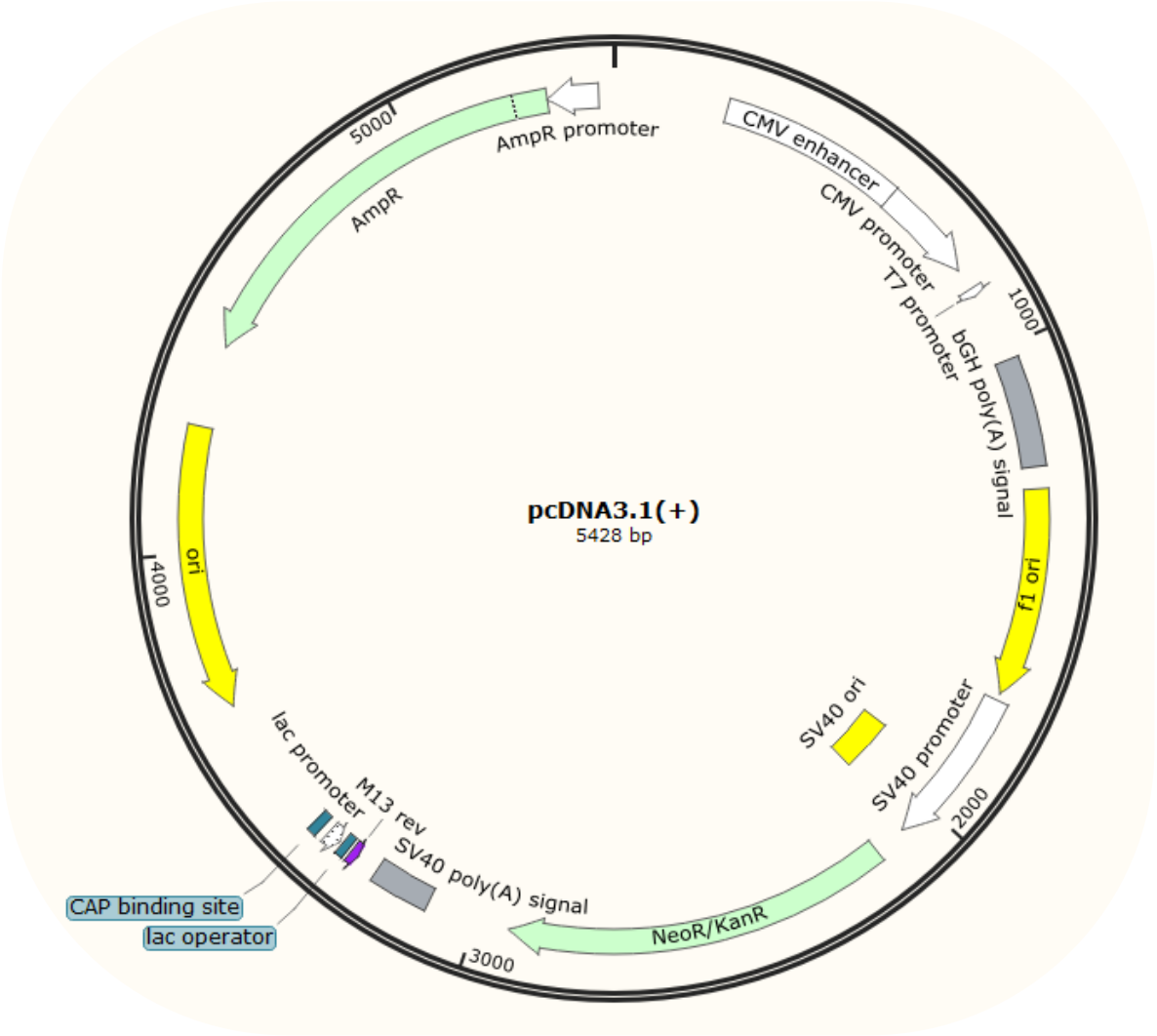
Promoters

- AmpR promoter (bacterial)
  - Low level constitutive expression of β-lactamase (ampicillin resistance)
- CMV promoter (mammalian)
  - Human cytomegalovirus immediate-early promoter, high level expression
- SV40 promoter (mammalian)
  - Constitutive expression
- T7 promoter (bacterial)
  - Constitutive expression in presence of T7 polymerase
- *lac* promoter (bacterial)
  - Constitutive expression in absence of the *lac* repressor (not used in pcDNA3.1)



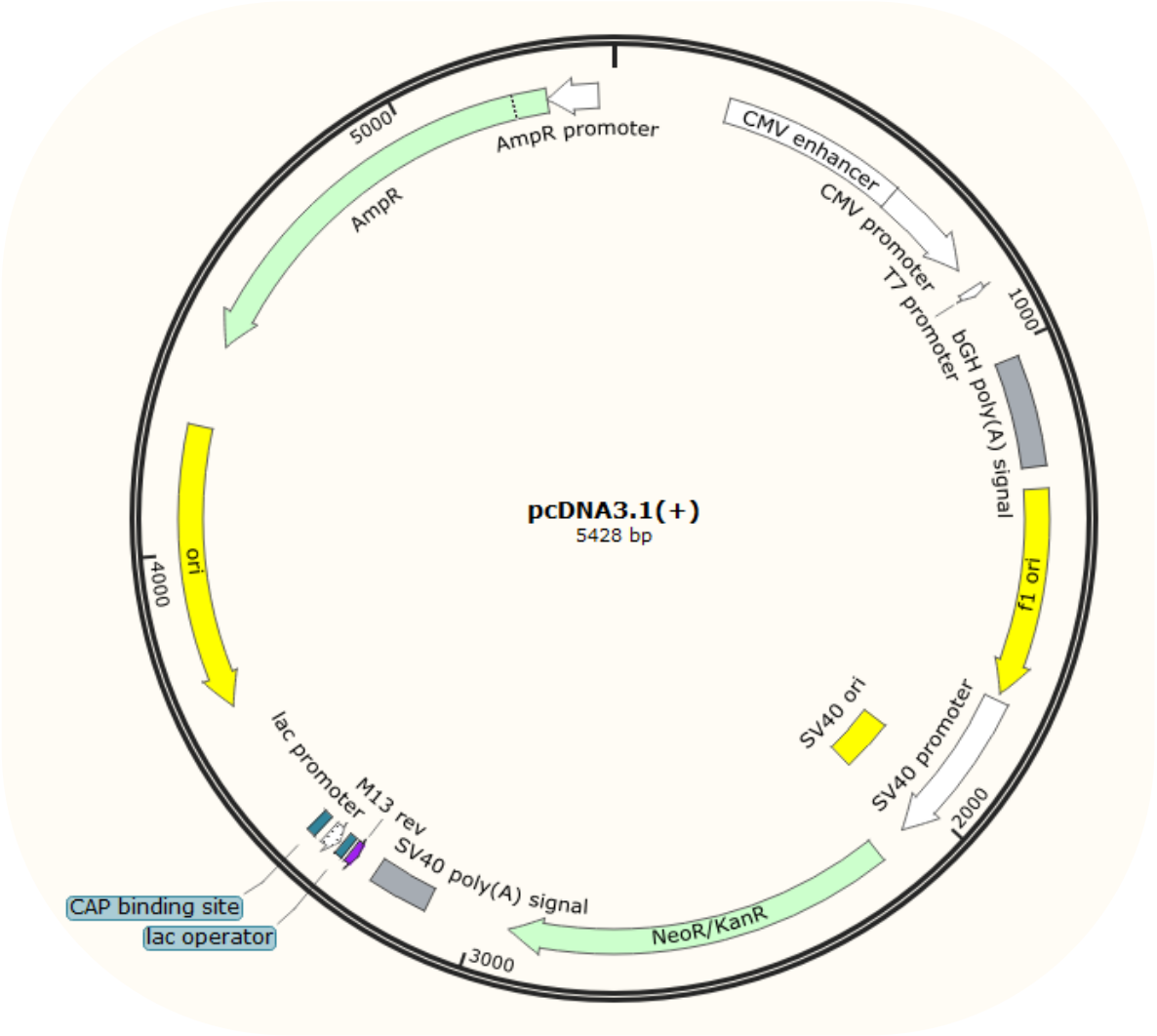
Promoters

- Different promoters determine *when* (which organism) genes will be expressed
  - AmpR promoter ensures constant production of β-lactamase in *E. coli*
  - CMV yields high levels of **GOI** transcription
  - SV40 promoter allows expression of neomycin resistance gene in mammalian cells (**selection marker**)
  - T7 promoter can be used with purified T7 RNA polymerase to make high yields of RNA for biochemistry



DNA regulatory elements

- Bacterial RNA polymerases will transcribe templates until they reach a **terminator** sequence
  - The same is true of the eukaryotic RNA polymerase II that makes mRNA
- Mammalian **poly(A) signals** include a sequence motif that promotes termination and poly-adenylation of the transcript
- pcDNA3.1 poly(A) signals:
  - SV40 after SV40 promoter
  - Bovine growth hormone (bGH) after CMV promoter



Transcriptional enhancers/repressors

- pcDNA3.1 encodes a CMV enhancer, that acts in *trans* to promote recruitment of RNA polymerase
- Other mammalian vectors include **splice sites** or the **woodchuck hepatitis virus posttranscriptional regulatory element** (WPRE) to promote mRNA stability and export from the nucleus

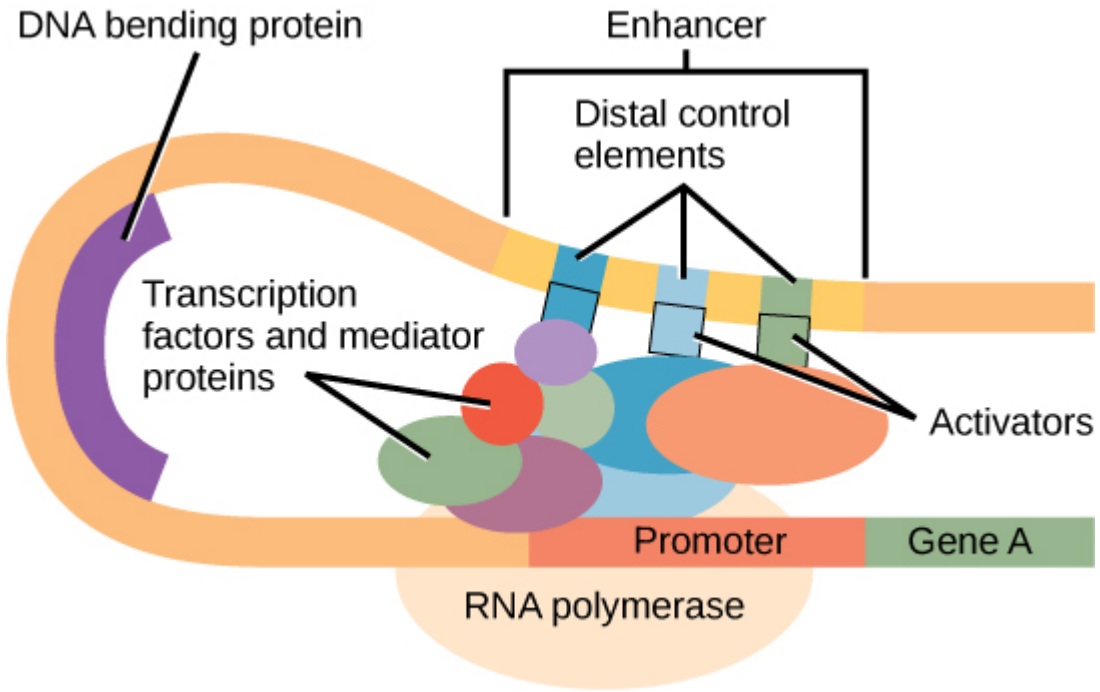
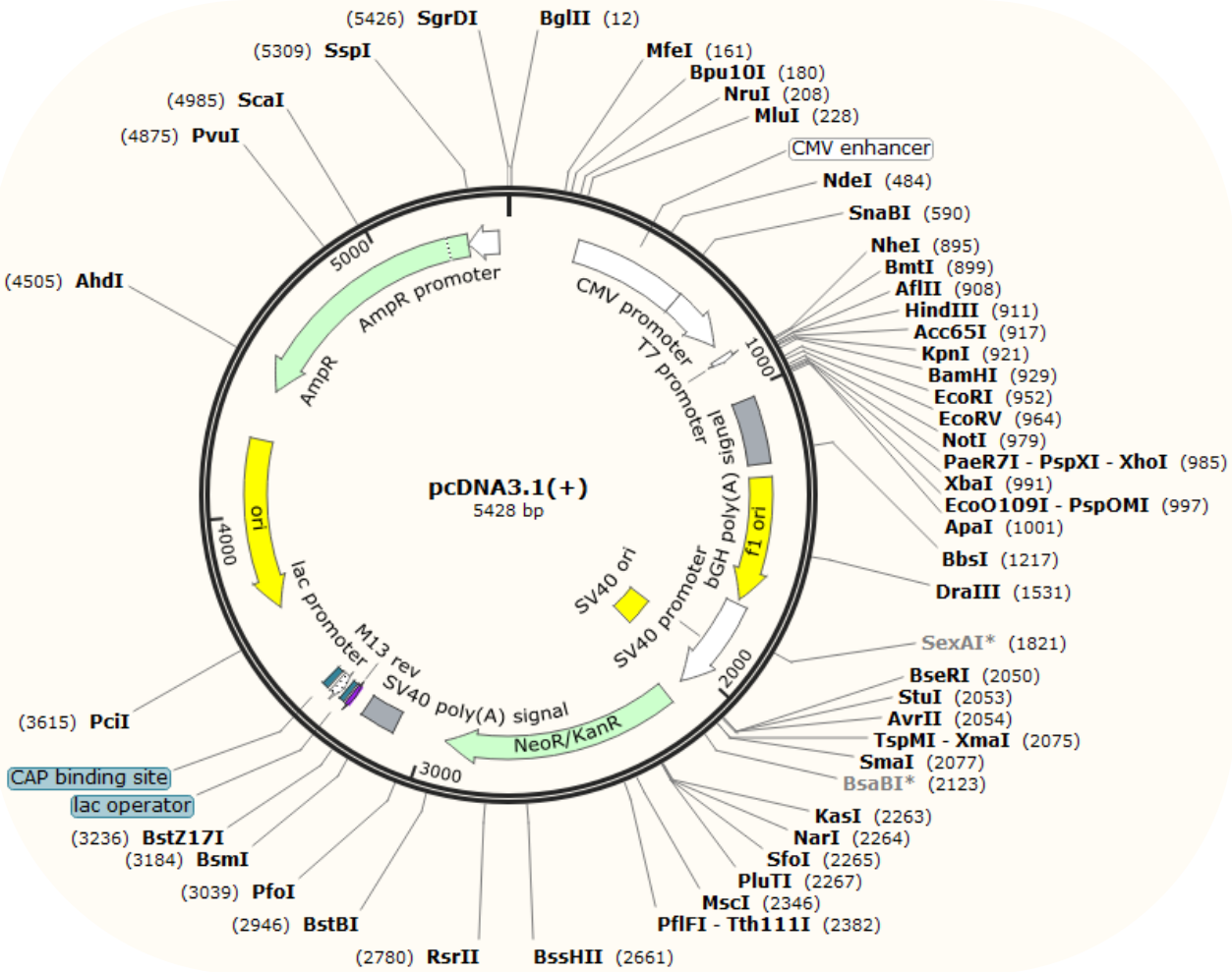


Image: LibreTexts Biology https://bio.libretexts.org/@go/page/13361

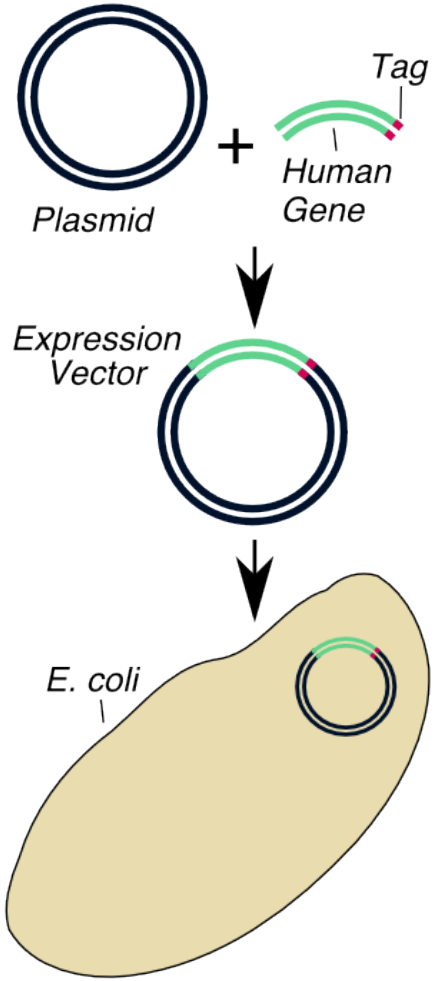
Multiple cloning site (MCS)

- This is the region of the plasmid where the **GOI** is inserted
- Will have lots of restriction sites present only once in vector (*Unique cutters*)
- In between the transcriptional start site (promoter) and end site (terminator or poly(A) signal)



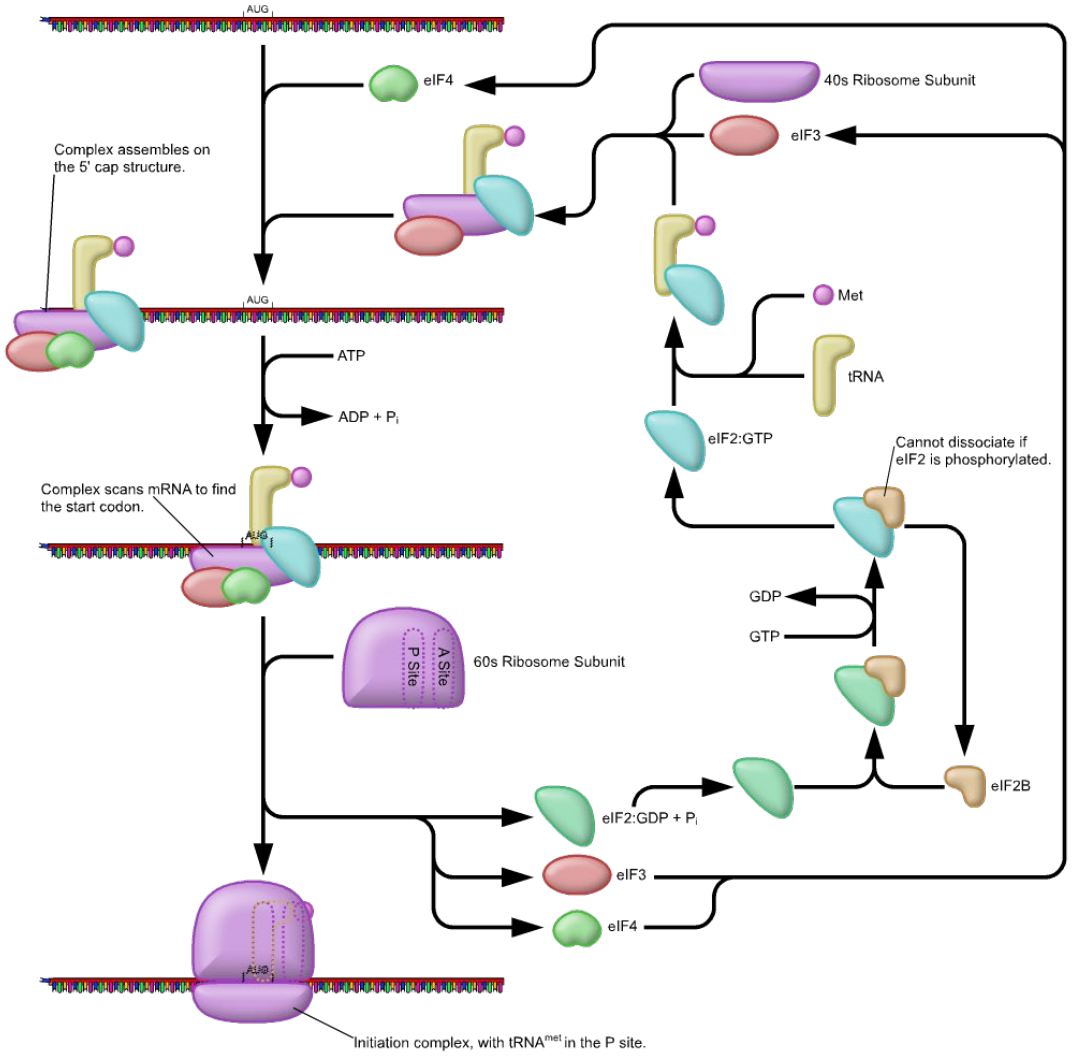
Cloning your GOI into the expression vector MCS

- For protein expression we usually clone the complementary DNA (cDNA), without introns, into the vector
- Need to have:
  - A ribosome binding site and translational start site
  - A stop codon
  - All protein elements *in frame* with your GOI



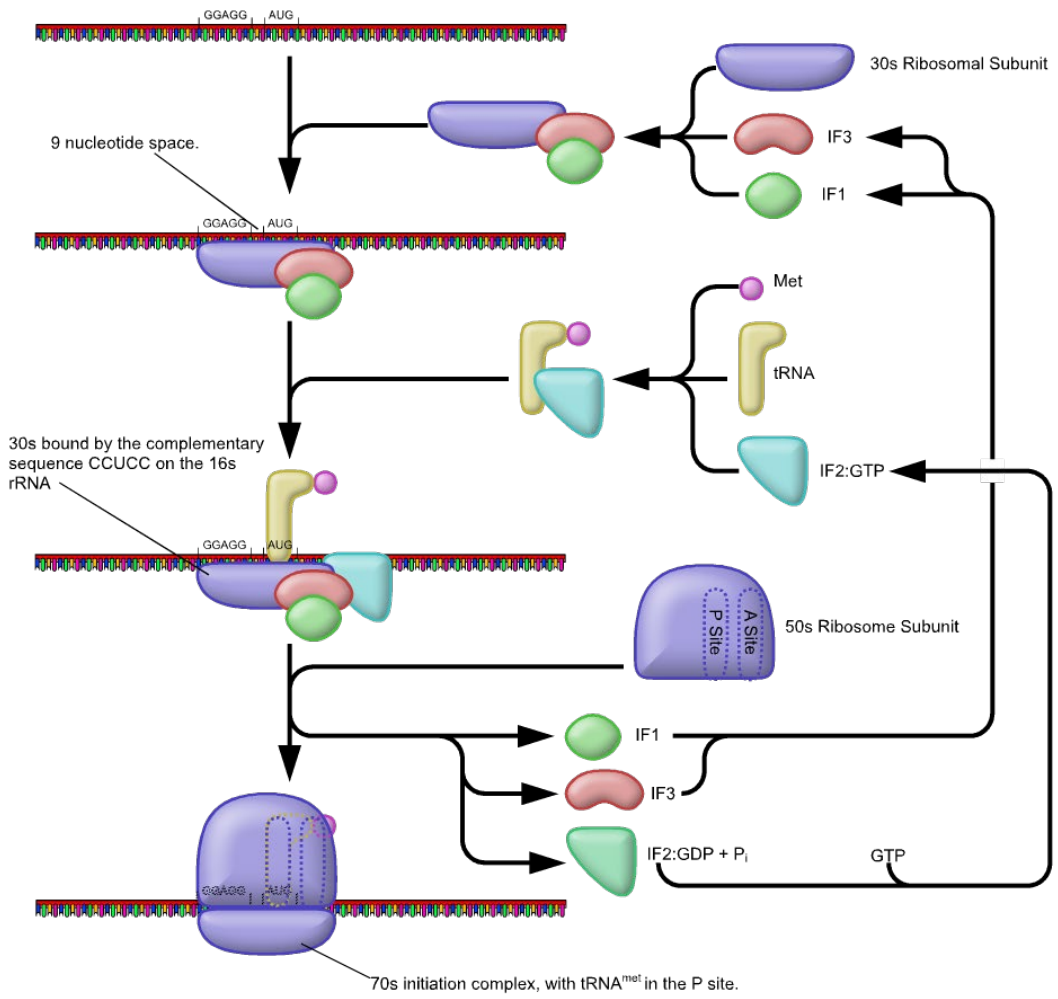
Ribosome binding sites and initiator methionine

- Protein translation is (usually) initiated at the Methionine codon (AUG)
- Initiation frequency depends upon the context of the AUG
- For efficient translation, need a ribosome-binding site
  - Mammalian expression: Kozak consensus sequence: ACC**AUG**G



Ribosome binding sites and initiator methionine

- Protein translation is (usually) initiated at the Methionine codon (AUG)
- Initiation frequency depends upon the context of the AUG
- For efficient translation, need a ribosome-binding site
  - Mammalian expression: Kozak consensus sequence: ACC**AUG**G
  - Bacterial expression: Shine/Dalgarno (SD) sequence: AGGAGGN$_{(5\text{-}9)}$**AUG**



Co-expression of multiple proteins from one vector

- For bacterial expression, you can encode multiple proteins by having multiple Shine/Dalgarno sequences and AUGs within one transcript (**polycistronic message**)

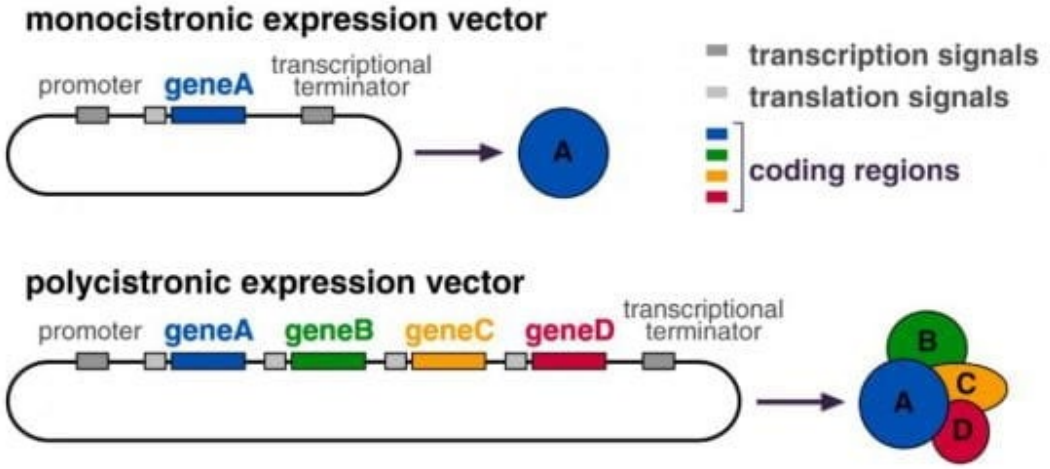
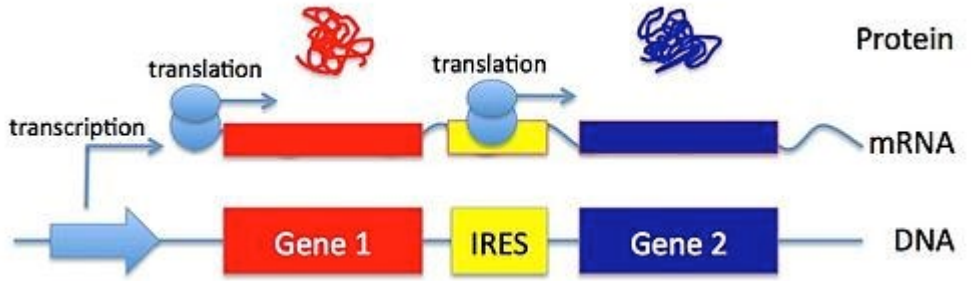
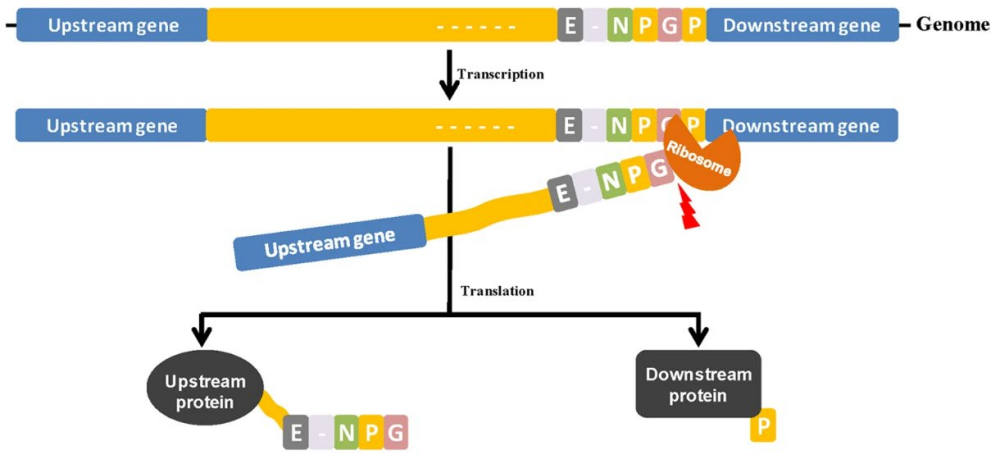


Image: Song Tan lab (https://sites.psu.edu/tanlab/polycistronic-expression/)

Co-expression of multiple proteins from one vector

- For bacterial expression, you can encode multiple proteins by having multiple Shine/Dalgarno sequences and AUGs within one transcript (**polycistronic message**)
- For mammalian expression, you can use viral internal ribosome entry sites (IRES) or translational stop-go sites (e.g. porcine teschovirus 2A) to encode multiple proteins off one transcript



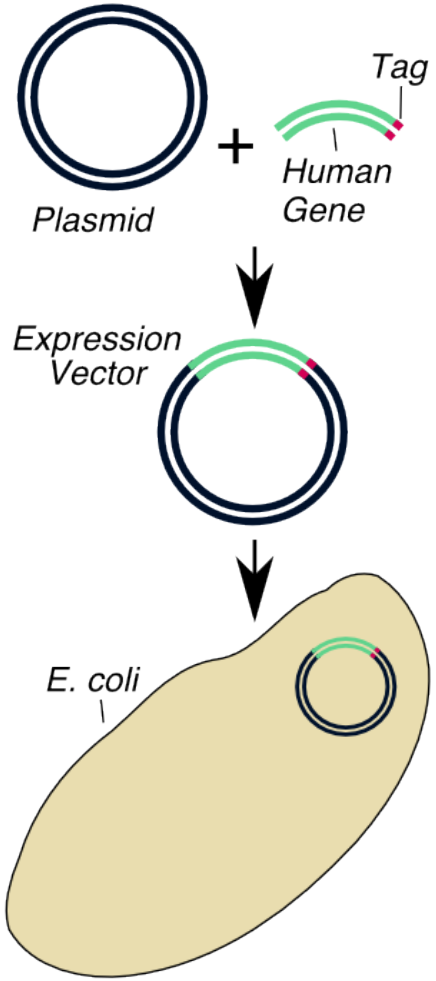


Top Image: Addgene (https://blog.addgene.org/plasmids-101-multicistronic-vectors)

Bottom Image: Wang et al. (2015) *Scientific Reports* 5, 16273

Cloning your GOI into the vector

- *Many* ways to clone your GOI into the plasmid of choice:
  - Restriction cloning
  - Gateway cloning
  - Gibson assembly
- Choice will be guided by:
  - Availability of restriction sites
  - Complexity of the cloning (simple cut-and-paste or complex assembly)
  - Available starting material



Restriction endonucleases

- Bacterial enzymes that defend against foreign DNA (e.g. bacteriophage) by degrading it
- Bind specific DNA sequences (restriction sites or recognition sequences)
- Usually cut both strands of DNA
- Often dimeric, cutting pallandromic sequences (*Type IIP*)
- Will often leave single-strand overhangs (**sticky ends**)

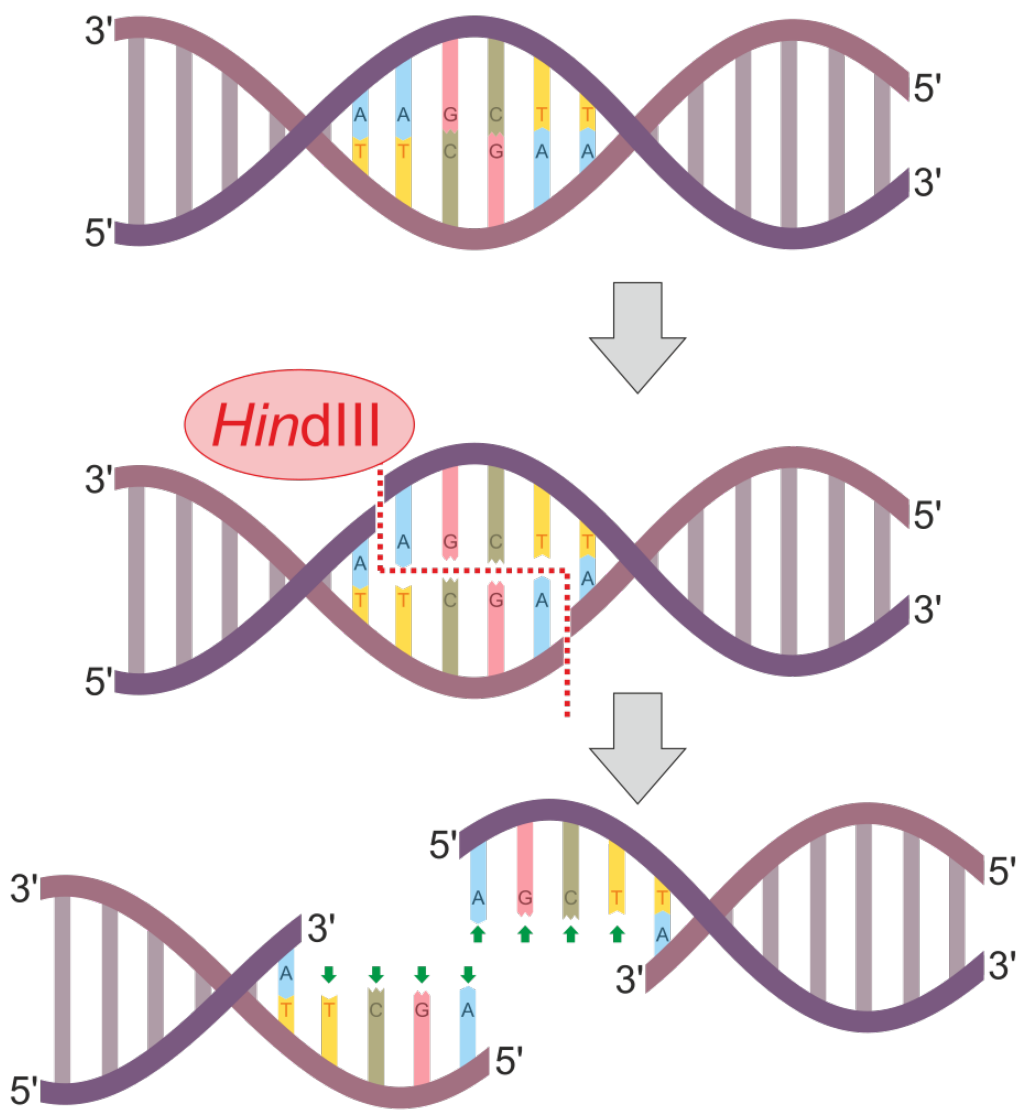
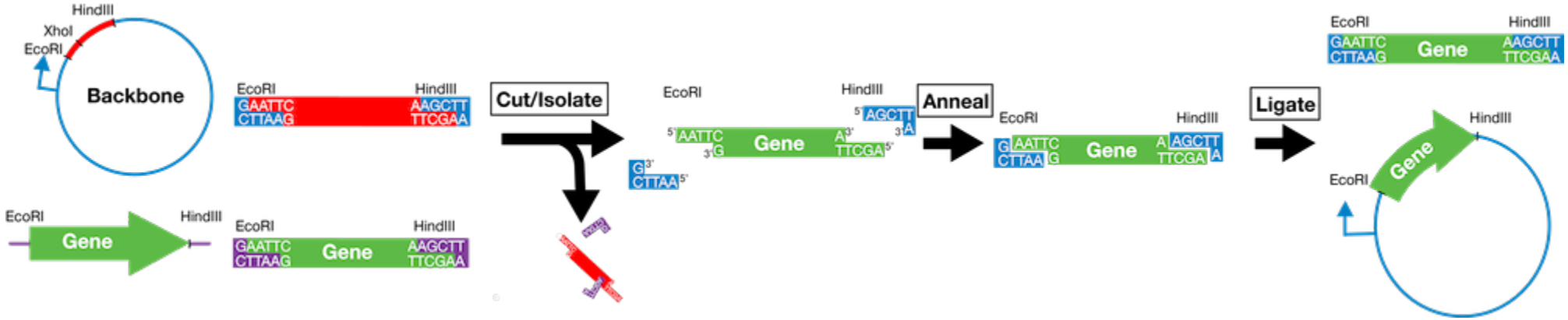


Image: https://bioclimate.commons.gc.cuny.edu/analyzing-dna/restriction-enzymes/

Restriction cloning



1. Uses *restriction endonucleases* to digest **insert** (containing GOI) and **backbone** (plasmid), leaves compatible **sticky ends**
2. Digested insert and backbone are isolated (Gels or PCR clean-up)
3. Insert and backbone are mixed, **ligated** using a DNA ligase (e.g. T4 bacteriophage DNA ligase), and used to **transform** bacteria

Image: Addgene (https://blog.addgene.org/plasmids-101-restriction-cloning)

Gateway cloning

- Proprietary cloning system developed by Invitrogen
- Exploits lambda bacteriophage DNA integration/excision reactions and a toxic gene (cccB) to achieve very high cloning efficiency
- Use the BP reaction to insert GOI into an ‘entry’ vector
- Use the LR reaction to transfer GOI into a ‘destination’ vector

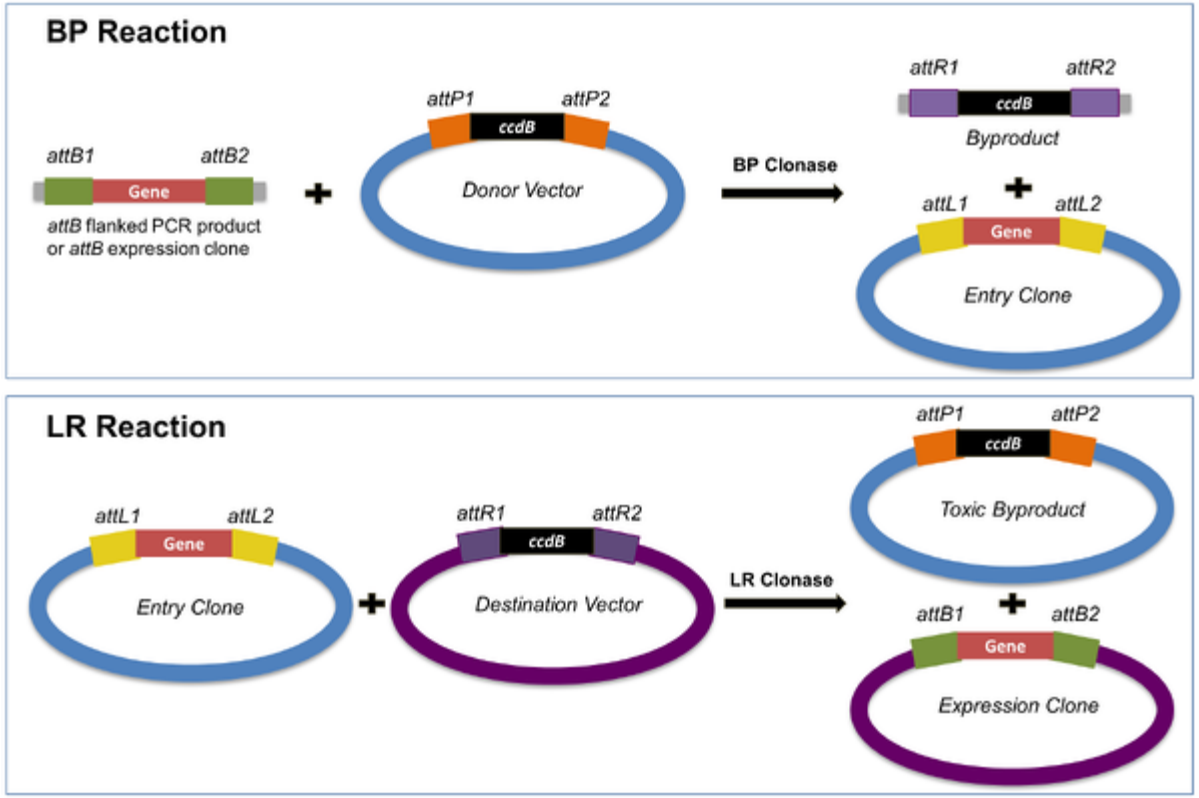


Image: Addgene (https://blog.addgene.org/plasmids-101-gateway-cloning)

Gibson assembly

- Mixture of three enzymes:
  - 5’ exonuclease to generate long single-strand overhands
  - Polymerase to fill in single-strand ‘gaps’ when DNA strands anneal
  - Ligase to repair nicks in the DNA backbone
- Reactions are isothermic (single temperature)

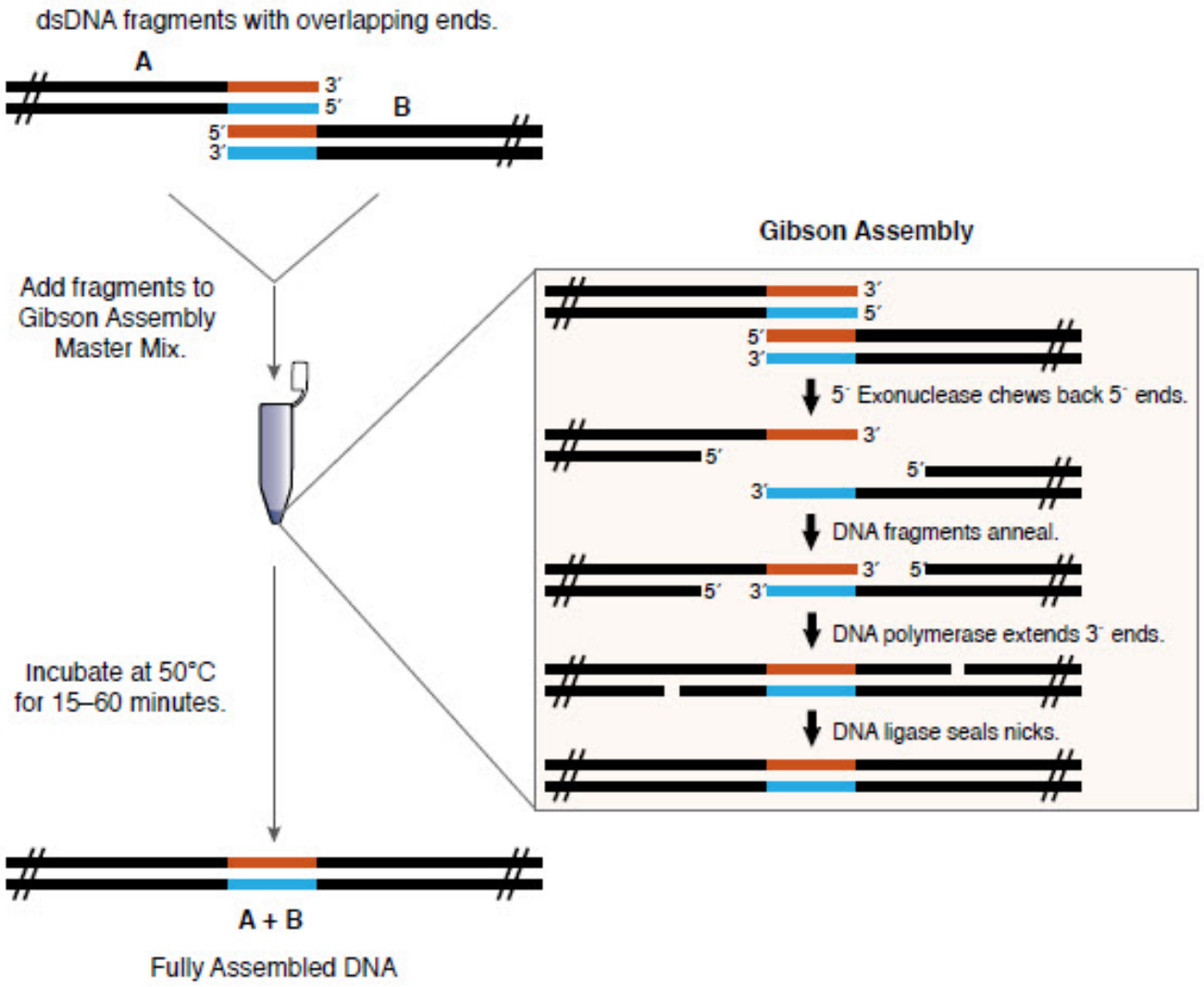


Image: New England Biolabs

Gibson assembly

- Mixture of three enzymes:
  - 5’ exonuclease to generate long single-strand overhands
  - Polymerase to fill in single-strand ‘gaps’ when DNA strands anneal
  - Ligase to repair nicks in the DNA backbone
- Reactions are isothermic (single temperature)
- Can use for large and complex DNA assembly

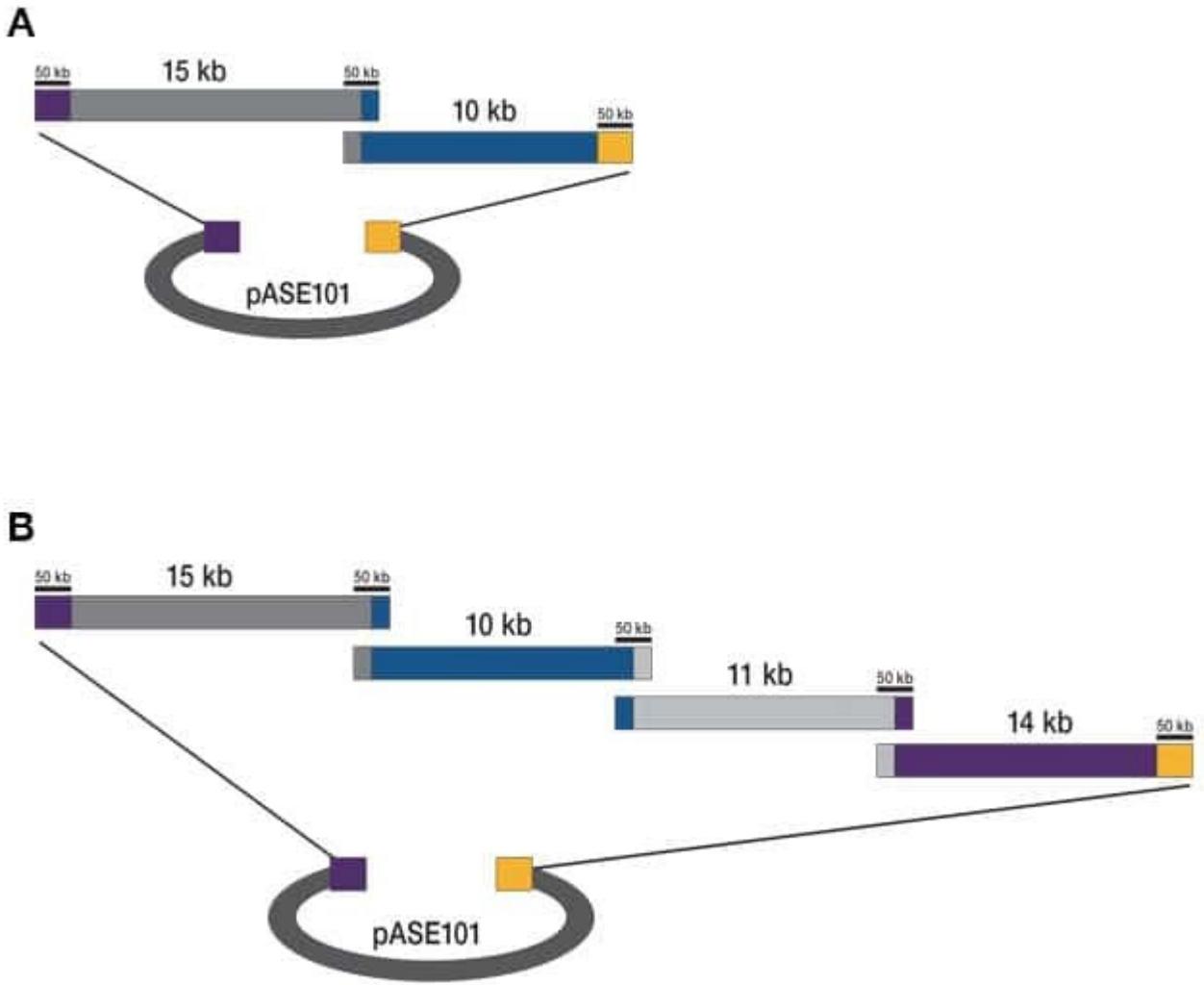
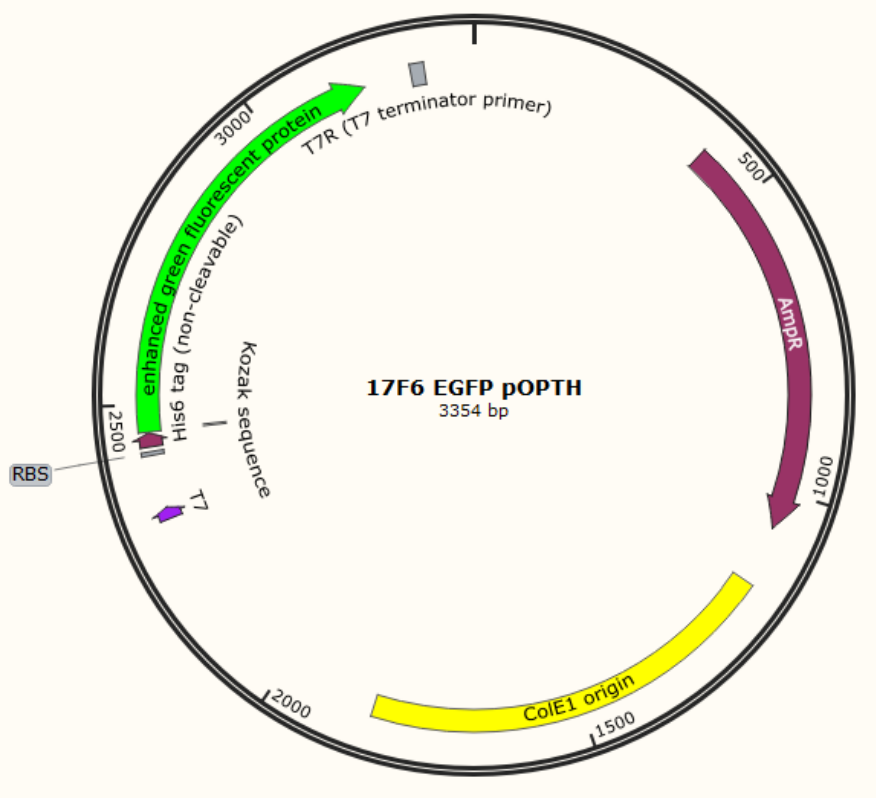
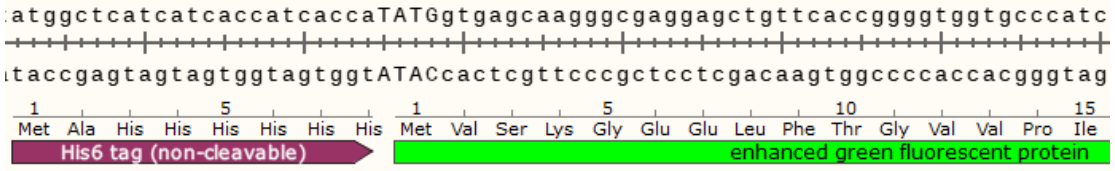


Image: ThermoFisher Scientific

Tags

- Adding tags can help you purify and detect your POI
- Common protein purification tags: His_6, glutathione S-transferase (GST), Strep-II, maltose binding protein, protein A
  - *We'll talk more about these in lecture 3*
- Many vectors encode tags
  - Need to clone GOI **in frame** with tag



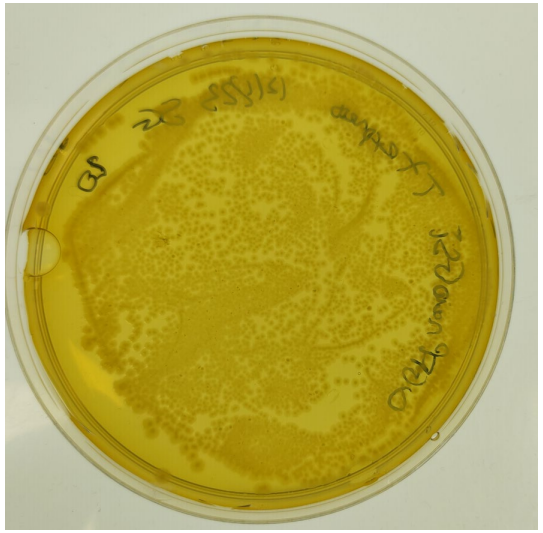


Which plasmid is right for me?

- Copy number
  - High if doing DNA manipulation in *E. coli* for downstream expression in another vector
  - Lower if gene product(s) will be expressed in *E. coli* or might be toxic
- Selection markers
  - Unique (co-transformed plasmids or genome of host don't confer resistance to same antibiotic)
- Origin of replication
  - Ensure it's compatible if co-transforming plasmids
- Expression host
  - Promoter and terminator correct for your expression host organism
- Tags

Bacterial expression

- The cheapest and simplest laboratory-based system for recombinant protein expression
- Can give very high yields of recombinant protein
  - 100s of mg per L of culture
- GOI generally encoded on an expression plasmid
- Doesn’t support most post-translational modifications



T7 polymerase and the pET system

- Developed in the 1980s by Bill Studier and colleagues
- Exploits the highly active RNA polymerase of T7 bacteriophage
  - Make RNA >8-times faster than *E. coli* RNA polymerase
  - Recognises a specific promoter
- pET (and derivative) systems use T7 polymerase to transcribe the GOI

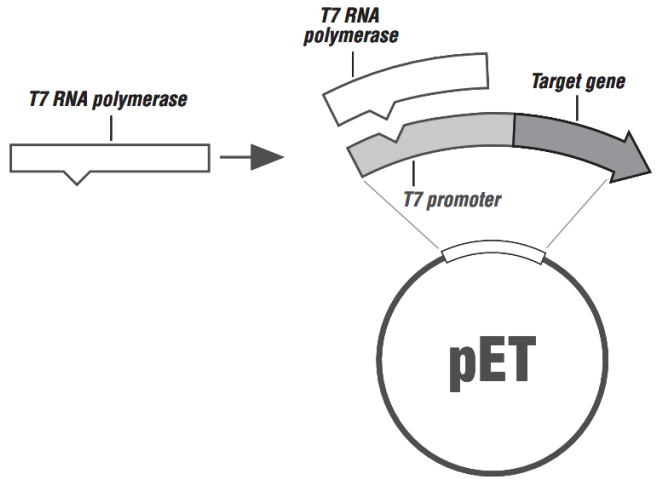


Image: pET System Manual, Novagen (2002)

T7 polymerase and the pET system

- T7 polymerase is encoded by the λDE3 lysogen, which is chromosomally integrated in DE3 bacteria like BL21(DE3)
- T7 polymerase expression is under the control of the *lac* promoter
- Expression from ***lac* promoter** is suppressed by ***lac* repressor** (encoded by bacterial *lacI* gene) that binds the ***lac* operator** (*lac o*)
- Isopropyl-β-D-thiogalactopyranoside (**IPTG**) induces expression by preventing the *lac* repressor from binding the *lac* operator

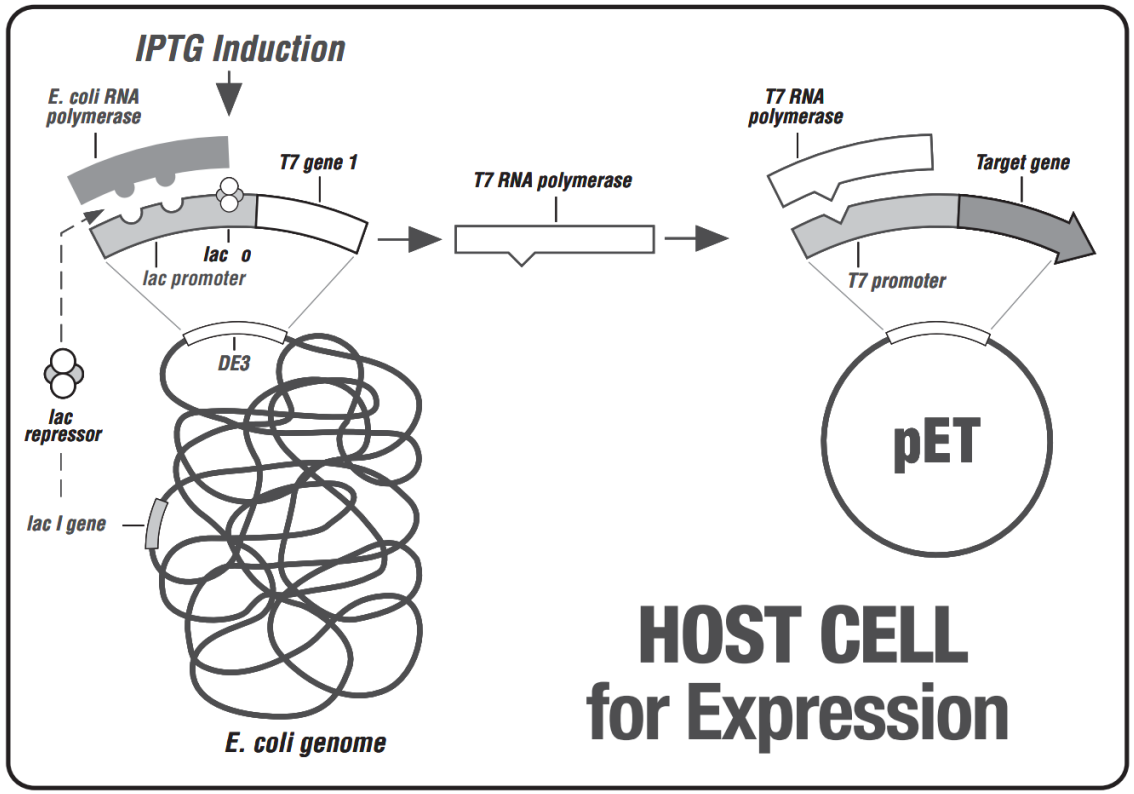


Image: pET System Manual, Novagen (2002)

T7 polymerase and the pET system

- There is low level leaky expression of T7 polymerase via the lac promoter

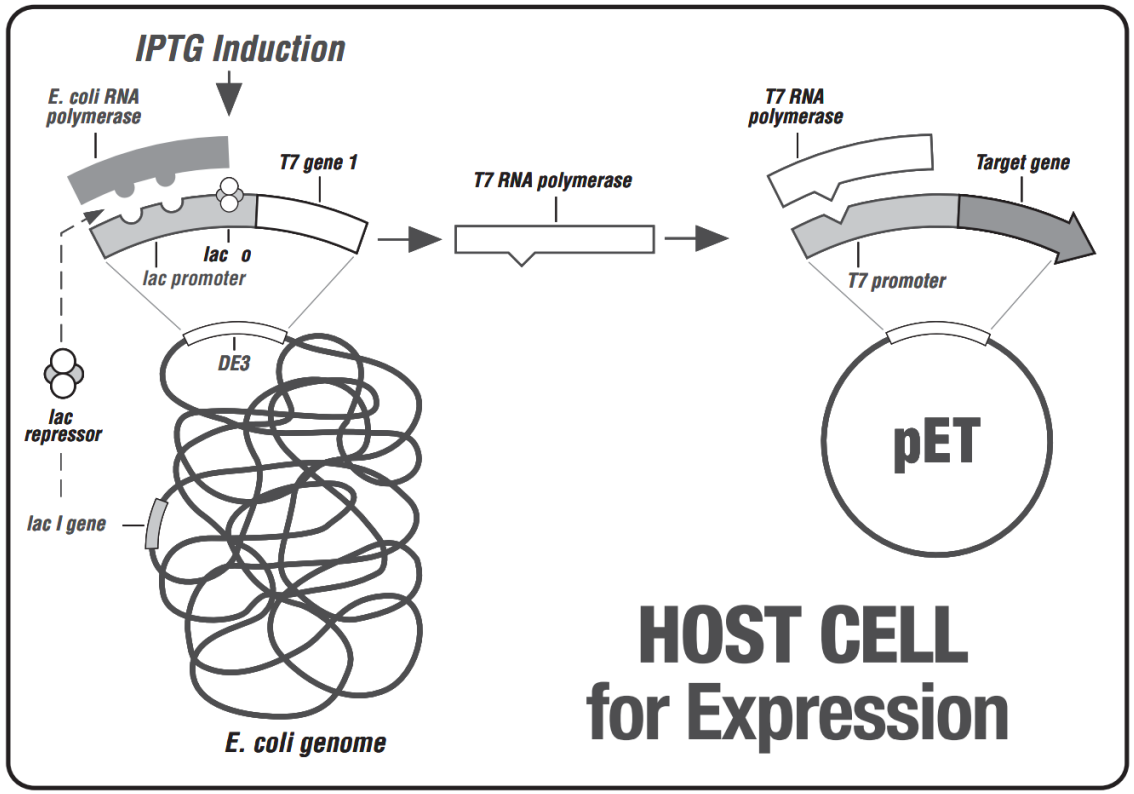


Image: pET System Manual, Novagen (2002)

T7 polymerase and the pET system

- There is low level leaky expression of T7 polymerase via the lac promoter
- The T7 lysozyme binds to T7 polymerase and inhibits its activity,
  - Encoded on plasmid pLysS (weak expression), pLysE (strong expression), or on bacterial chromosome (LysY)
- T7 lysozyme prevents the polymerase expressing GOI before IPTG induction

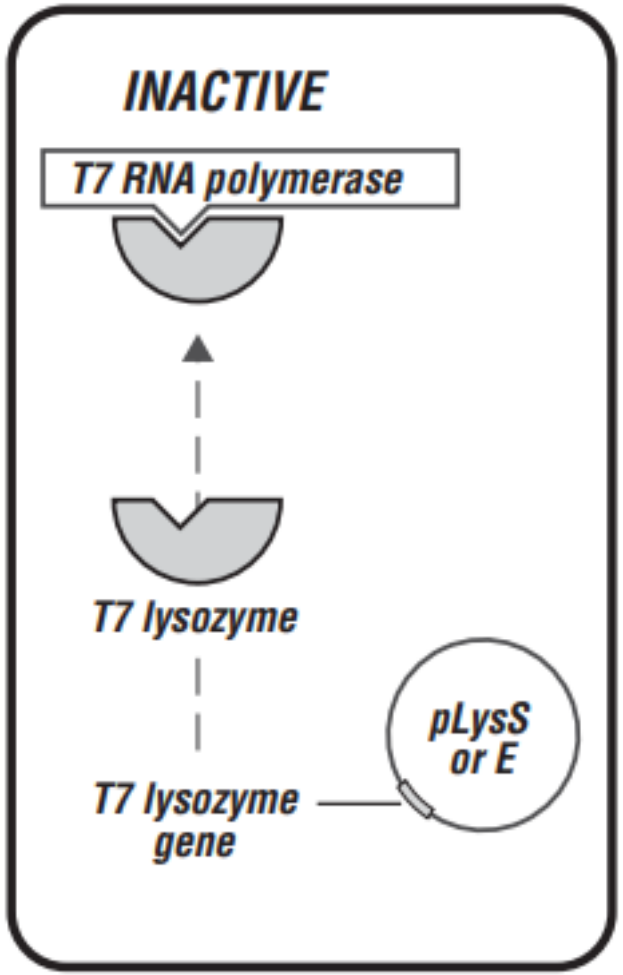


Image: pET System Manual, Novagen (2002)

Expression from T7 promoter is specific

- No expression of GOI in ‘normal’ bacterial cells that don’t contain DE3 lysogen expressing T7 polymerase

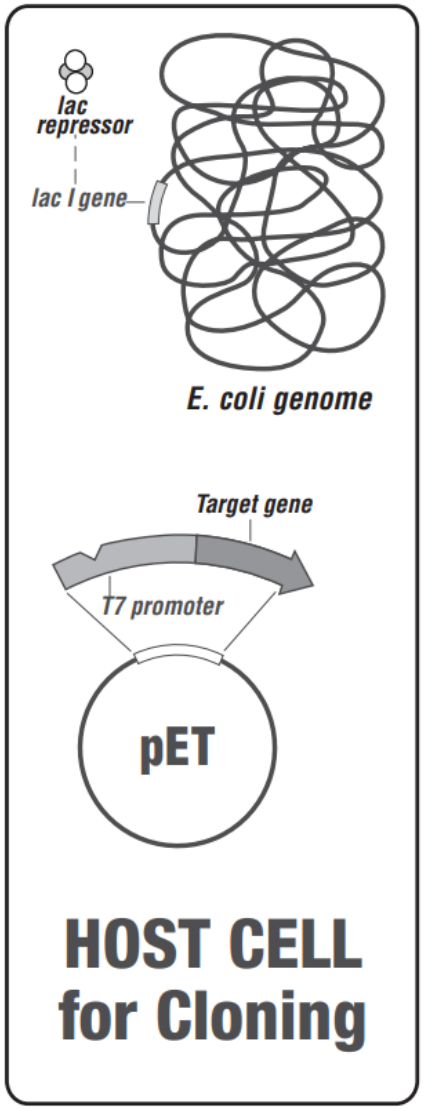
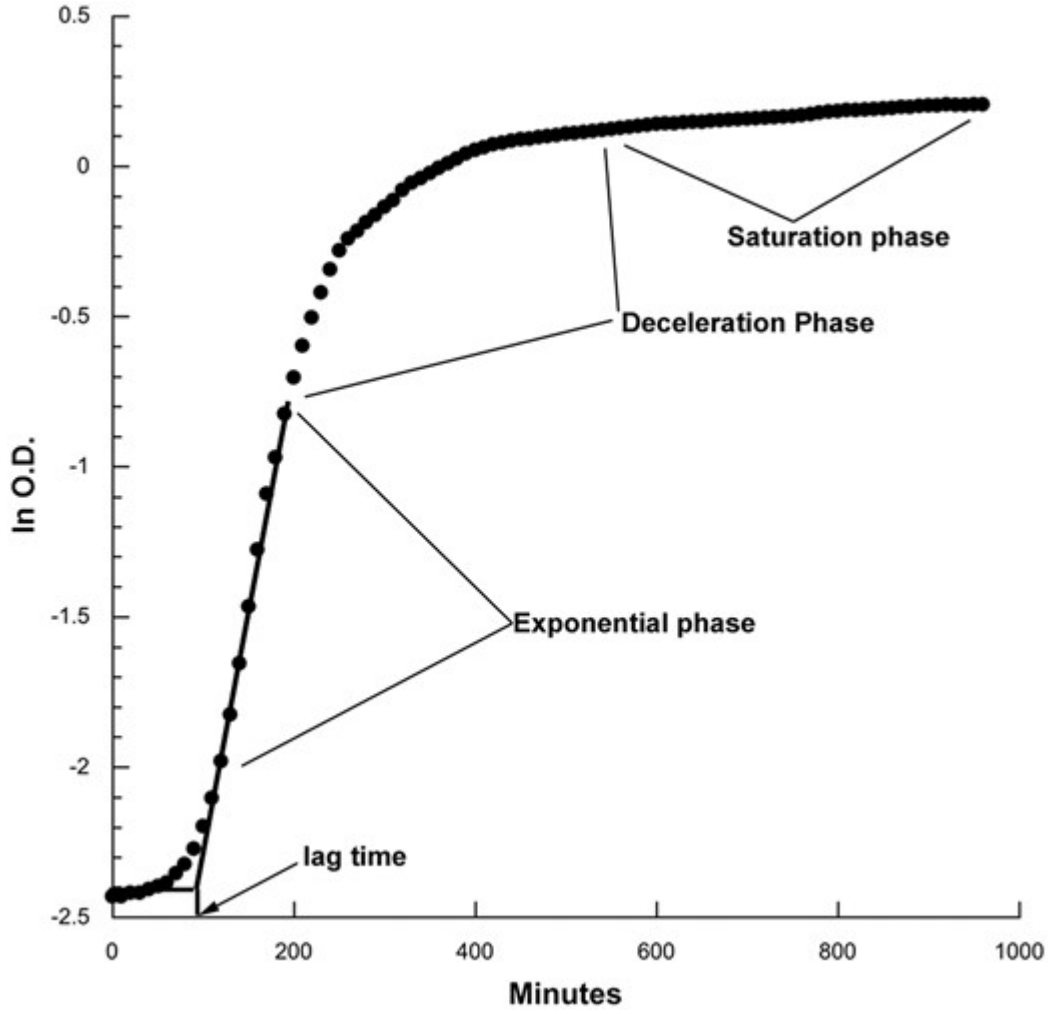


Image: pET System Manual, Novagen (2002)

Induce expression during exponential (log) phase growth

- *E. coli* divide by fission
  - One mother cell divides into two daughter cells
- *E. coli* grows in culture exponentially until all nutrients are exhausted
  - Divide every ~20-40 min
  - Most metabolically active during the exponential (log) phase
- Best to induce expression with IPTG in late log phase
  - pET system is so powerful, after induction bacteria make few proteins except POI

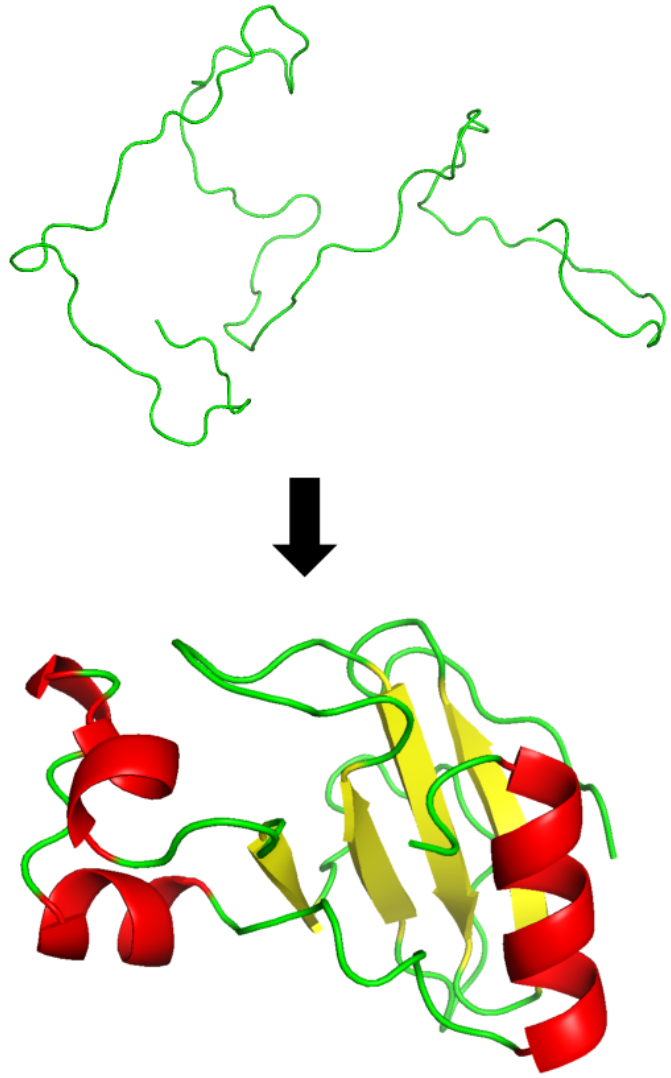


Timetable of ‘typical’ bacterial protein expression

- Day 1:
  - Transform DE3 bacterial with plasmid and incubate on selection plates overnight at 37°C
- Day 2:
  - Inoculate one or more small-scale (5 mL) starter cultures with single colonies from the plate and grow at 37°C overnight
- Day 3:
  - Inoculate large-scale (100 mL to 8 L) cultures with 1:1000 dilution of starter culture and grown at 37°C to an OD_{600} of ~0.8–1.2
  - Induce with 0.2–1.0 mM IPTG
  - Grow for a further 2–16 hours before harvesting by centrifugation

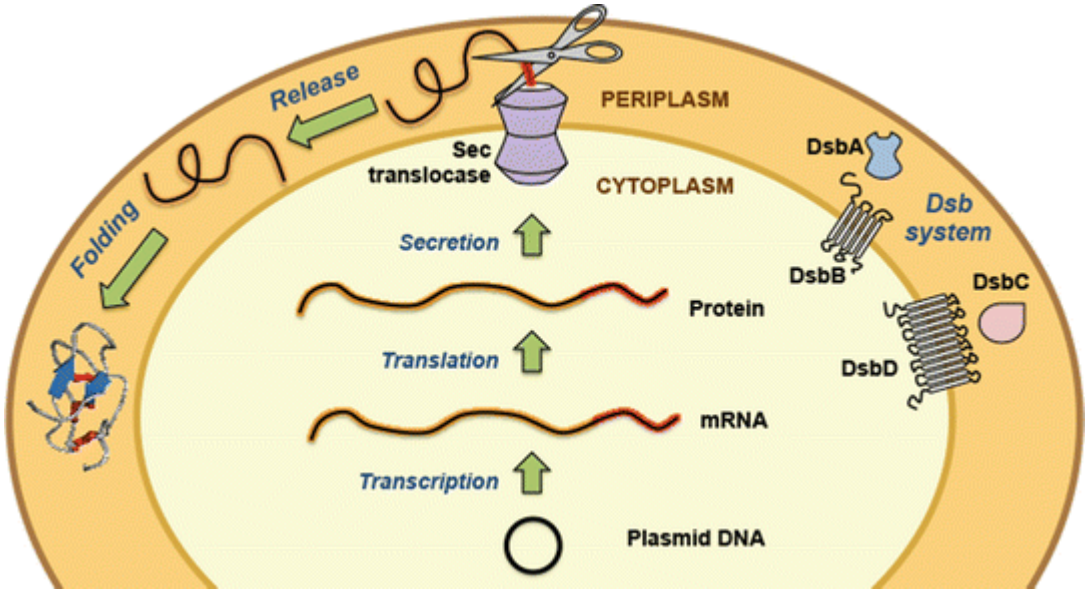
Optimising protein folding

- Many mammalian proteins fold poorly in bacteria
  - Can be improved by dropping growth temperature to 22°C before induction
- Can also try co-expressing chaperones (e.g. GroEL and GroES) or inducing chaperone expression (add 1% ethanol to growth medium)
- Some bacterial cell lines are optimised to help protein folding
  - ActicExpress – express proteins at low temperatures (12°C)
  - Origami cells – mutations in redox proteins to enhance disulfide bond formation



Bacterial expression of disulfide-bonded proteins

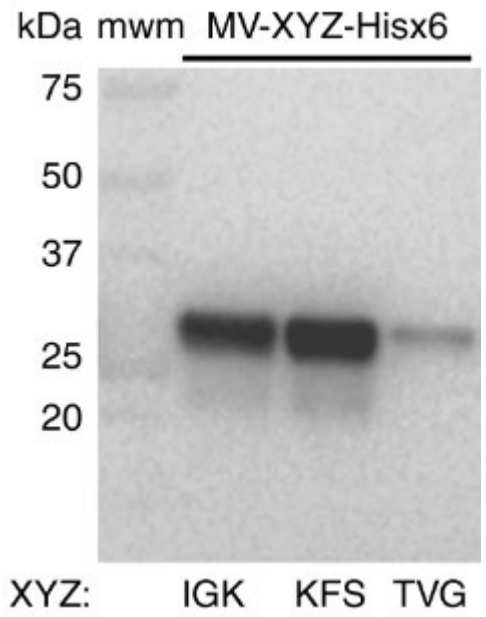
- The *E. coli* cytoplasm is a reducing environment
- The periplasm is an oxidising environment that contains disulfide isomerases (Dsb proteins)
- You can add the PelB leader sequence to your GOI and it will be targeted to the *E. coli* periplasm, allowing disulfide bond formation
  - Optimal for nanobody expression



Saez et al. (2017) *Heterologous Gene Expression in E.coli* pp 155–180

Optimising protein expression

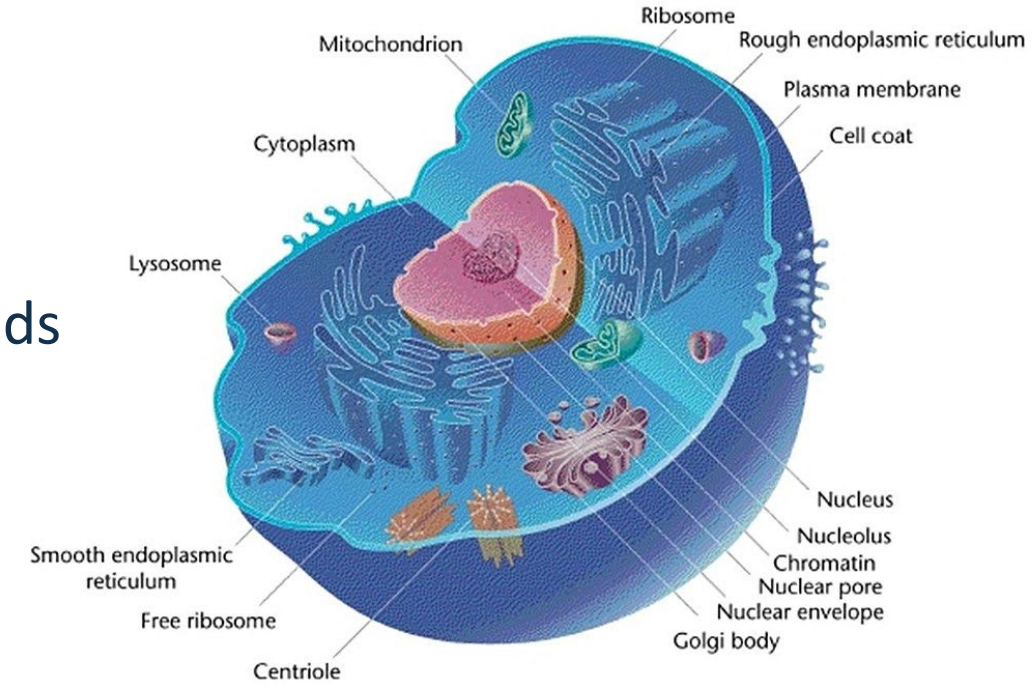
- Can use bacterial strains that encode tRNAs for ‘rare’ codons (common in mammals but not *E.coli*)
  - pRIL/pRARE or Rosetta/Rosetta2 cells
- Recent studies show nucleotide sequence positions 7–15 (amino acids 3–5) have a major effect upon recombinant *E. coli* protein expression
  - Translation initiation is a dominant factor in level of protein expression
  - Can boost protein expression by optimising first few amino acids of the sequence



Verma et al. (2019) *Nature Communications* 18, 5774

Mammalian expression

- Many mammalian proteins have post-translational modifications (PTMs): Phosphorylation, glycosylation, …
  - Especially true for important secreted mammalian proteins such as antibodies, that have glycosylation and multiple disulfide bonds
- Lots of large, multi-domain proteins fold poorly in bacterial cells
- Can use large-scale mammalian tissue culture to express cells
- Can express both secreted and cytoplasmic proteins



Mammalian expression systems

- Adherent cells
  - HEK293T
  - CHO
- Suspension cells
  - Freestyle 293F
  - Expi293
- Both used extensively in industry and academia
- Suspension cells allow greater cell density
  - Better for cytoplasmic proteins

Expression systems for mammalian cells

- Transient transfection
  - Protein encoded on a plasmid
  - Cells are transfected and express protein for short period of time
- Stable cell lines
  - Expression cassette is stably incorporated into eukaryotic genome
  - POI can be expressed constitutively or under control of inducer
- Vector needs to have eukaryotic transcription/translation signals!

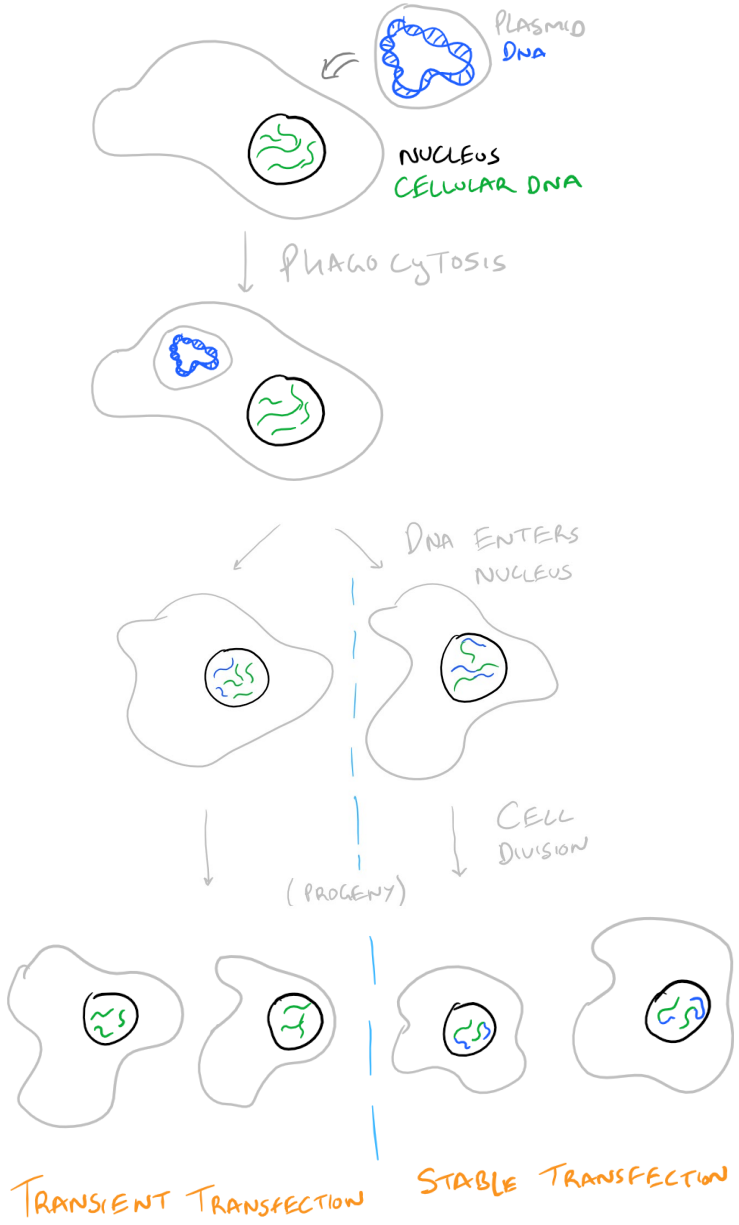


Image: Karthik Raman, Scigine (https://scigine.com/blog/calcium-phosphate-transient-transfection/)

Large-scale transient transfection

- Many cell biologists uses lipid-based transfection reagents to introduce plasmids into cells
  - E.g. FuGene, TransIT-LT1, Lipofectamine
- These can be prohibitively expensive for large-scale transfection
- High transfection efficiency can be obtained using branched polyethylenimine (PEI) in some cells lines like HEK293T
  - PEI is very cheap!

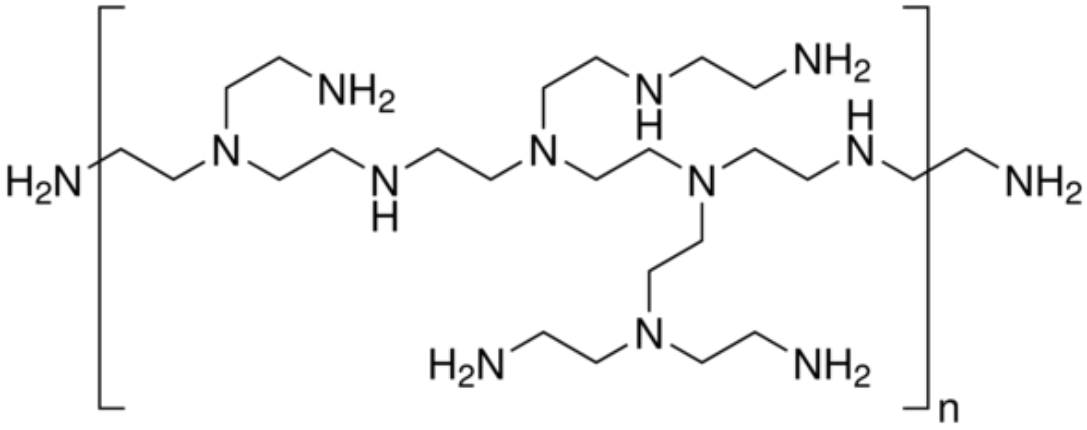
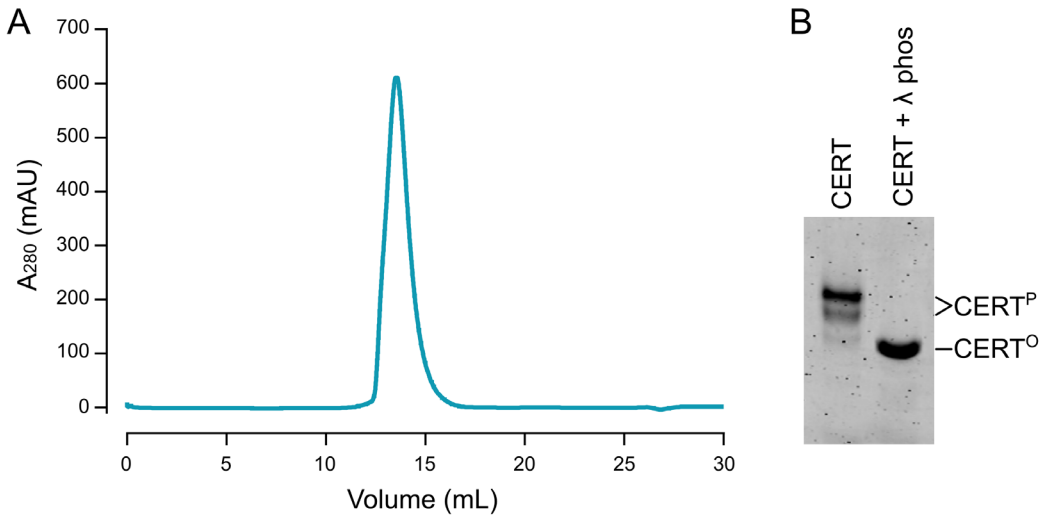


Image: Merck

Expressing cytoplasmic proteins with PTMs: CERT

- We needed to express a phosphorylated protein (CERT) for an assay
  - Measure dephosphorylation via a viral protein complex
- Used transient transfection of Freestyle 293F cells
  - Transfected 100 mL of cells with DNA+PEI, incubated for 72 hours before harvesting
- Purified cytoplasmic protein in presence of phosphatase inhibitors
- Very high efficiency purification of phosphorylated protein (mg/L of culture)



Benedyk et al. (2021) *PLoS Pathogens* 17: e1009824

Making stable cell lines

- Stable plasmid transfection
  - Transfect with plasmid encoding GOI as-per transient transfection
  - Select for stable (genomic) incorporation of plasmid using selection marker:
    - E.g. Geneticin (G-418), hygromycin, puromycin
  - Random incorporation and expression levels
    - Can get very high expression via clonal selection

Image: BioCat GmbH

Making stable cell lines

- Stable plasmid transfection
  - Transfect with plasmid encoding GOI as-per transient transfection
  - Select for stable (genomic) incorporation of plasmid using selection marker:
    - E.g. Geneticin (G-418), hygromycin, puromycin
  - Random incorporation and expression levels
    - Can get very high expression via clonal selection
- Transposase-based genomic incorporation
  - Transfect with plasmid encoding GOI plus selection marker, flanked by transposase inverted terminal repeats
  - Co-transfect with helper plasmid(s) that encode transposase
  - Select using similar markers as above
  - High efficiency genomic incorporation and consistent high-level expression

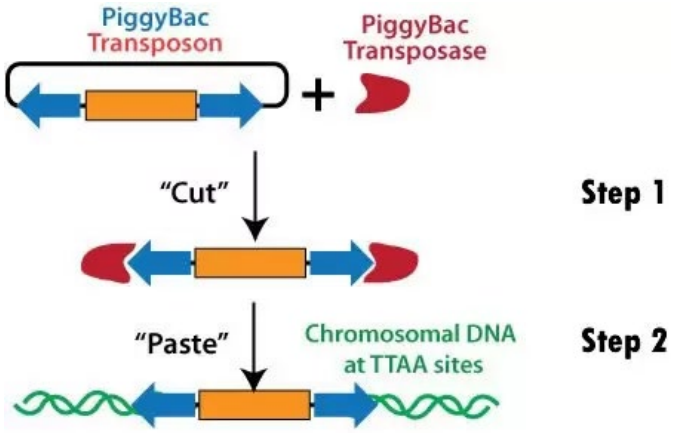


Image: BioCat GmbH

Inducible mammalian expression using doxycycline

- Controls protein expression using the tetracycline repressor from Gram-negative bacteria
  - TetR binds tetracycline or derivatives (including **doxycycline**) and the tet responsive element
- Two different types of regulation:
  - Tet-Off: Protein expressed in the **absence** of doxycycline
  - Tet-On: Protein expressed in the **presence** of doxycycline
- Tet-On systems are popular for protein expression
  - Tight regulation of expression (low leaky expression)
  - Low levels of doxycycline required for induction
  - Need constitutive expression of Tet-On protein

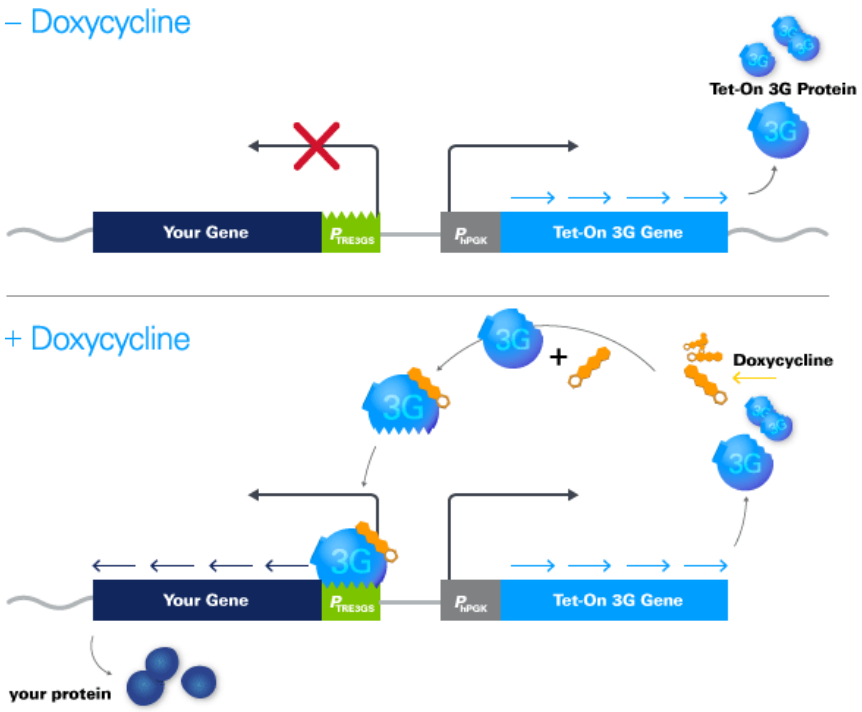


Image: Takara Bio

Piggybac transposase system

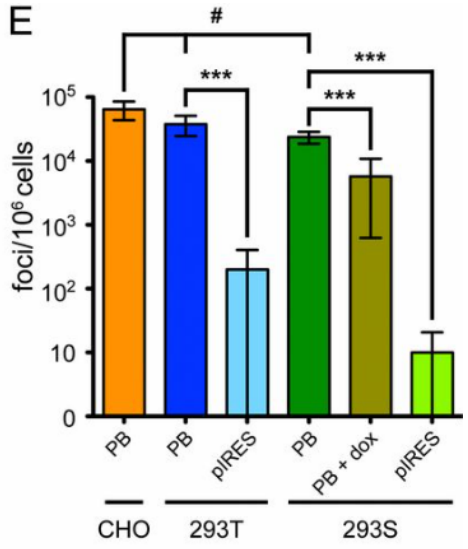
- Allows consistent high-level protein expression in suspension mammalian cell culture
- Transfect cells with:
  - Plasmid encoding GOI flanked by transposase inverted repeats
  - Plasmid encoding reverse tetracycline trans-activator (rtTA) for Tet-On expression
  - Plasmid encoding piggybac transposase

Li et al. (2013) *PNAS* 110:5004-9

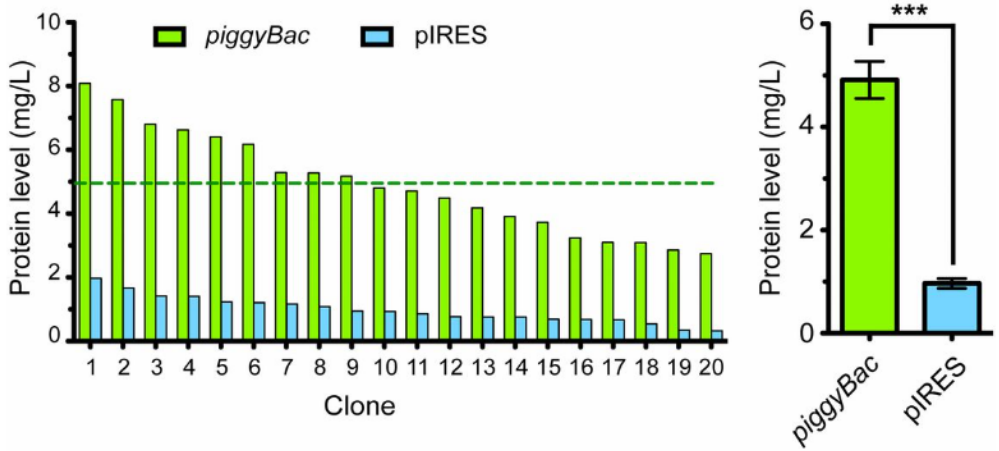
Piggybac transposase system

- Allows consistent high-level protein expression in suspension mammalian cell culture
- Transfect cells with:
  - Plasmid encoding GOI flanked by transposase inverted repeats
  - Plasmid encoding reverse tetracycline trans-activator (rtTA) for Tet-On expression
  - Plasmid encoding piggybac transposase
- Select stable cells within two weeks
  - Very consistent genome incorporation
- About 8-fold higher expression than transient transfection with PEI (personal observation)

Efficiency of generating stable cells



Expression levels (individual clones)



Li et al. (2013) *PNAS* 110:5004-9

Stable versus transient transfection

System	Pros	Cons
Transient	• Quick • Can express many different proteins from one maintenance flask • Easy to try lots of different constructs (start of project)	• Need to make DNA for each transfection • Expression levels can be variable
Stable	• Can get very high levels of expression (g/L culture) • Reproducible expression	• Slow to establish lines • Need to maintain separate cell lines for each different protein you're expressing

Expressing secreted proteins

- Many important mammalian proteins are extracellular
  - Either secreted or have extracellular (luminal) domains
- Can express in mammalian cells by encoding a **secretion signal** in the expression plasmid
  - At amino terminus of the protein
- Purify expressed protein from the culture medium
  - High purity!

Signal Peptide	Amino acid sequence
Secrecon	MWWRLWWLLLLLLLLWPMVWA
Mouse IgKVIII	METDTLLLWVLLLWVPGSTG
Human IgKVIII	MDMRVPAQLLGLLLLWLRGARC
CD33	MPLLLLLPLLWAGALA
tPA	MDAMKRGLCCVLLLCGAVFVSPS
Consensus	MLLLLLLLLLLALALA
Native	MLLLLLLLGLRLQLSLG

doi:10.1371/journal.pone.0155340.t001

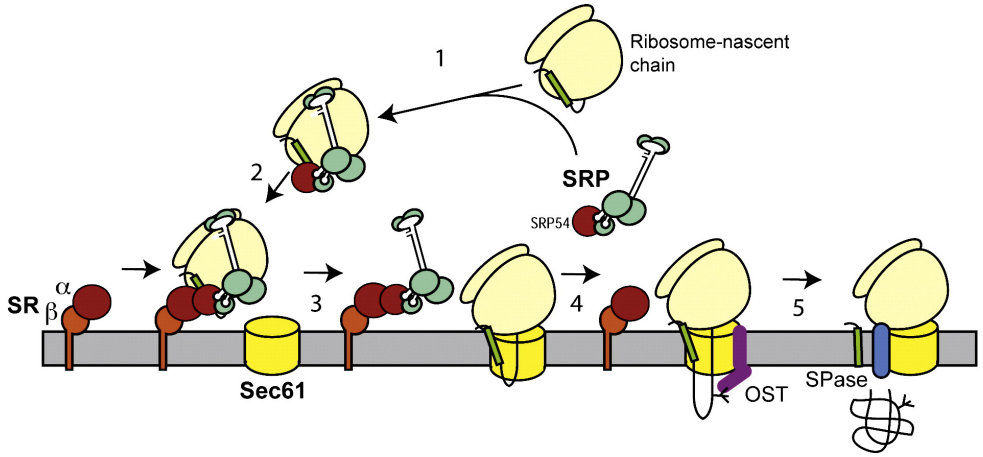


Table: Güler-Gane et al. (2016) *PLoS ONE* 11: e0155340

Image: Nyathi et al. (2013) *BBA – Molecular Cell Research* 1833, 2392-2402

Controlling N-linked glycosylation

- Secreted proteins can undergo extensive, complex glycosylation
- Glycans are flexible and heterogenous
  - Bad for structural studies (crystallisation)

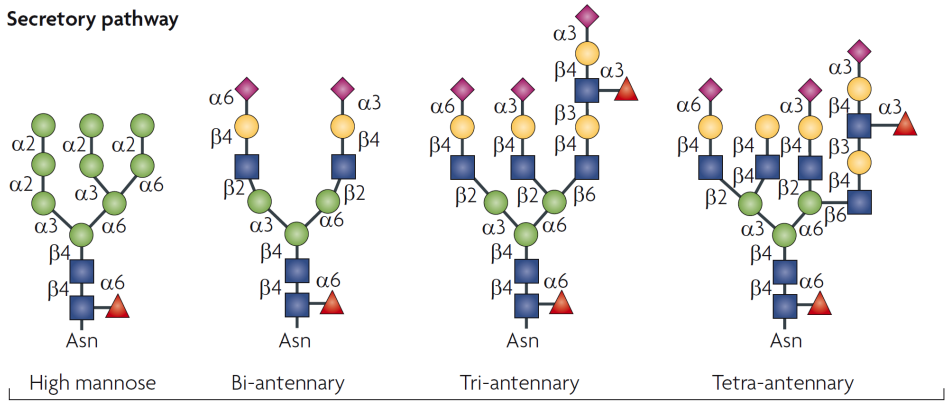


Image: Grewal (2008) *Nat Rev Immunol* 8, 874–887

Controlling N-linked glycosylation

- Secreted proteins can undergo extensive, complex glycosylation
- Glycans are flexible and heterogenous
  - Bad for structural studies (crystallisation)
- Endoglycosidase (Endo)H enzyme efficiently removes high-mannose glycans
- Can block complex processing of high-mannose to more complex glycans
  - Glycosylation deficient cell lines: HEK293S GnTI$^-$ or CHO-Lec1 cells
  - Drug treatment: Kifunensine

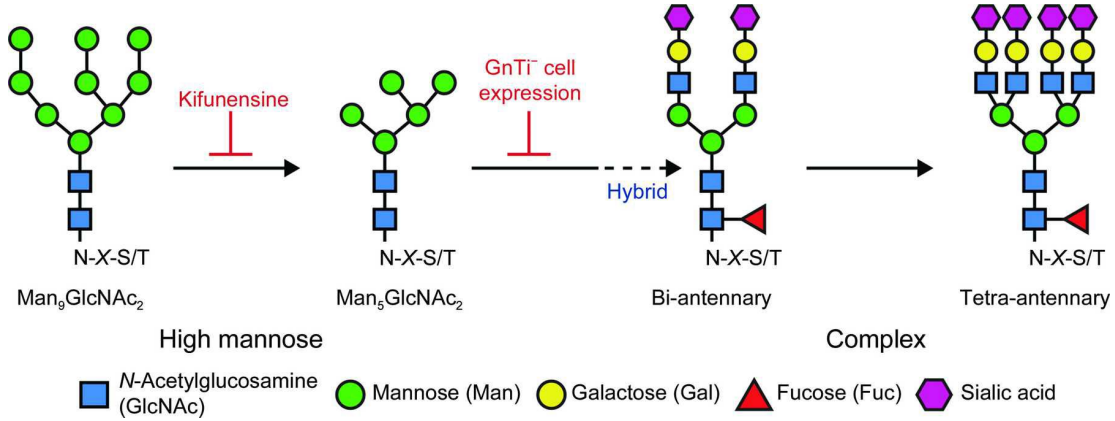


Image: Gristick (2017) *Acta Cryst D*73, 822–828

Other expression systems

- **Yeast**
  - Eukaryotic expression system that is cheap and scalable, with relatively easy genetic modification. Allows some post-translational modifications although glycosylation is different to mammalian cells. Cells lysis can be challenging.
- **Insect cells**
  - Excellent for expression of large multi-component complexes. Proteolysis can be an issue and generating baculovirus (vector) stocks is time consuming. Different glycosylation patterns compared to mammalian cells
- **Plants and algae**
  - Highly scalable and low cost, but technology isn't very mature and accessible
- **Cell-free**
  - Uses enriched lysates from higher eukaryotes (like wheat)
  - Very quick (reactions take hours) and allows production of highly toxic proteins, but very expensive and not very scalable.

Which expression system should I use?

- If you don’t need PTMs, always worth trying **bacteria** first because they are so cheap and quick
- If PTMs are required try **mammalian cell** culture or **yeast**
- If purifying secreted proteins that need native glycosylation, use **mammalian cell** culture
- If expressing cytosolic multi-component complexes, try **insect cell expression** or mammalian cells
- If only a tiny amount of protein is required, could try **cell-free** expression

This talk - Recap

- Expression vectors and the basics of molecular cloning
- Expression systems:
  - Bacteria
  - Mammalian cells
  - Others (Yeast, insect cells, plants, cell-free)

Tomorrow: What can we learn about a protein from its sequence?