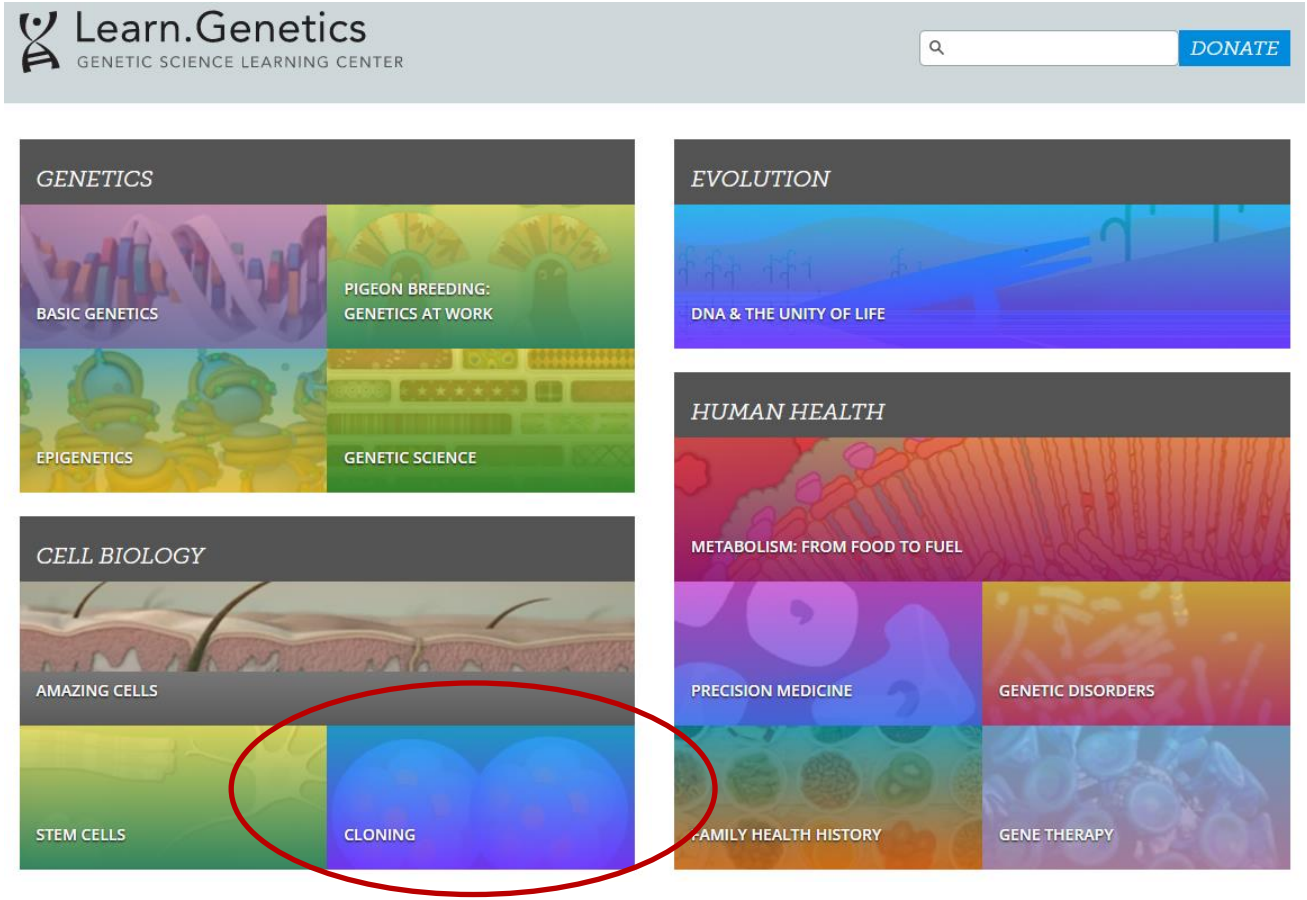
Biotechnology of drugs – Basics of genetic engineering I.

Doc. RNDr. Jan Hošek, Ph.D.
hosekj@pharm.muni.cz

Department of Molecular Pharmacy
FaF MU

Interesting site

https://learn.genetics.utah.edu/



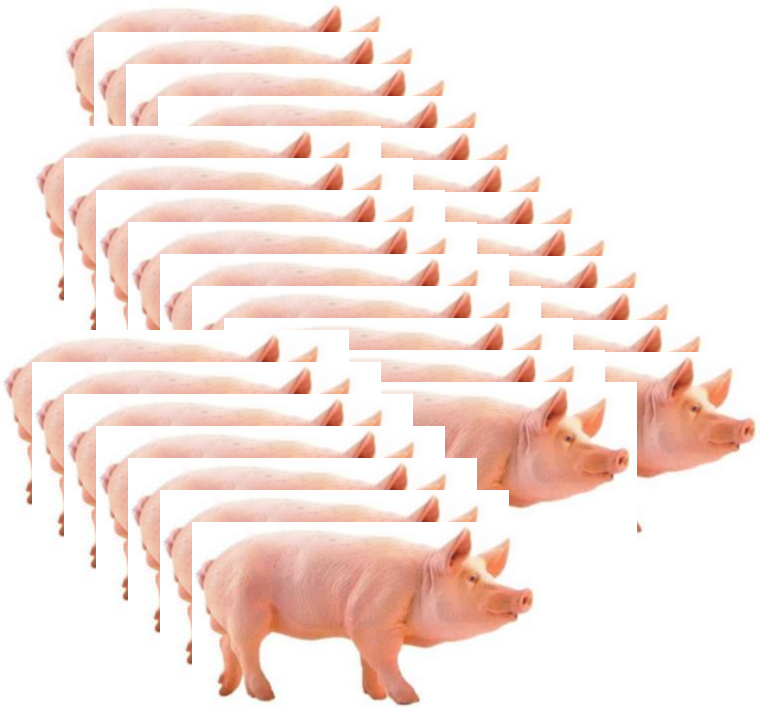
Biotechnological product can be created:

1. **By classical biotechnological procedure**
   - interbreed
   - mutagenesis of a population of cells
   - selection of cells/organisms with suitable properties
   - **they cannot force the organism to produce a protein that is not its own**

2. **By genetic engineering - recombinant DNA technology**
   - gene cloning → recombinant DNA technology
   - genetic manipulation *in vitro*
   - **breaks down barriers between species → heterologous systems**

The goal of recombinant technologies

The effort to genetically encode the synthesis of a commercially advantageous product into an organism that will produce it cheaply and with high yield.



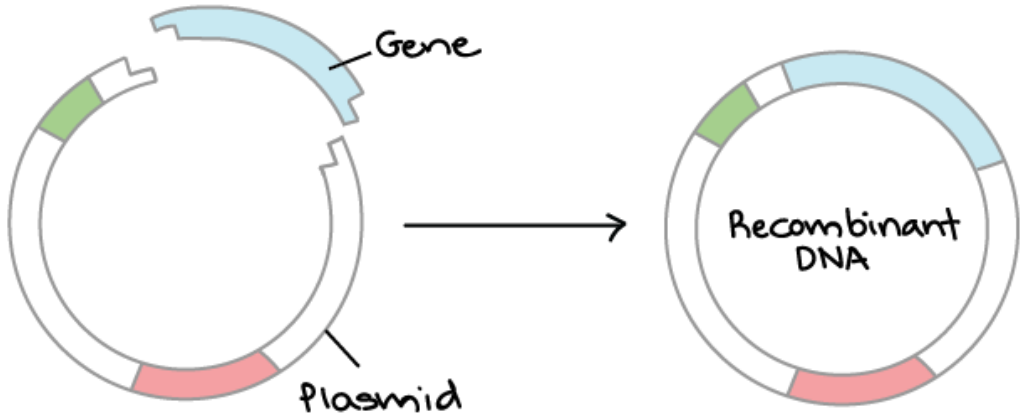
2000 kg of pancreas

200 mL of insulin

Basic principles of recombinant technologies

Recombinant technology - gene cloning - needs the incorporation of foreign DNA into the host cell

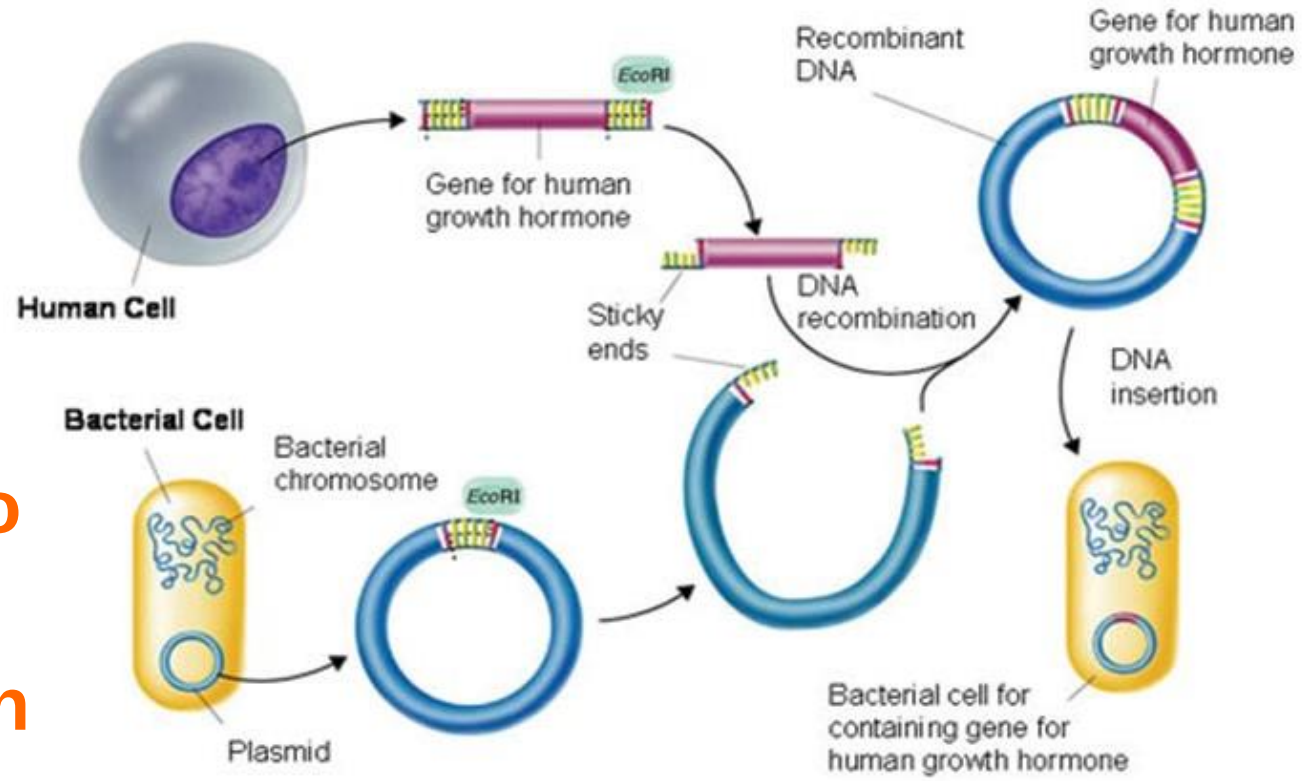
- **introduction of DNA (transformation) into the cell**
- **ensuring the survival of forein DNA**
- **ability to replicate**
- **assurance of expression**



What is needed for gene cloning

- own gene – DNA fragment
- vector – plasmid, phage, cosmid
- host – recipient of recombinant DNA

- insert = gene incorporated into the vector
- recombinant DNA = vector with insert



https://microbenotes.com/gene-cloning-requirements-principle-steps-applications/

Consequence of gene cloning = transformation

A permanent heritable change in the genetic material of a cell caused by the acceptance and incorporation of foreign genetic information

- **ability to replicate**
- **ability to express**

Sources of foreign genetic information

animal

vegetal

microbial

synthetic

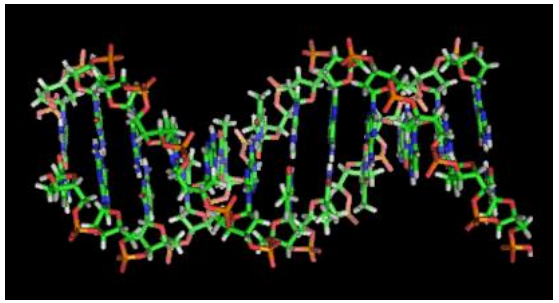
- The recombinant protein can be:
  - **Final product** (e.g. hormone, antibody)
  - **Tool for further synthesis** (e.g. various enzymes)

How to get a gene or its fragment?

Isolate DNA in its native state (see Methods of mol. biol.)

Process enzymes

Perform PCR



Direct synthesis

Clone only

And how to proceed further?

1) Did I get a nucleic acid?

YES ↓ NO →

2) Do I have it in sufficient quantity and quality?

YES ↓ NO →

a) Process enzymes (see Methods of mol. biol.)
b) Perform PCR (see Methods of mol. biol.)

Factors influencing the expression of cloned genes

Regulatory sequences for gene expression

1. Transcription
   a. The power of the promoter
   b. Transcription terminator
   c. Stability of mRNA
2. Translation
   a. Structure of the ribosome binding site
   b. Codon usage
3. Transport of proteins
   a. Character of the signal sequence

Properties of vectors

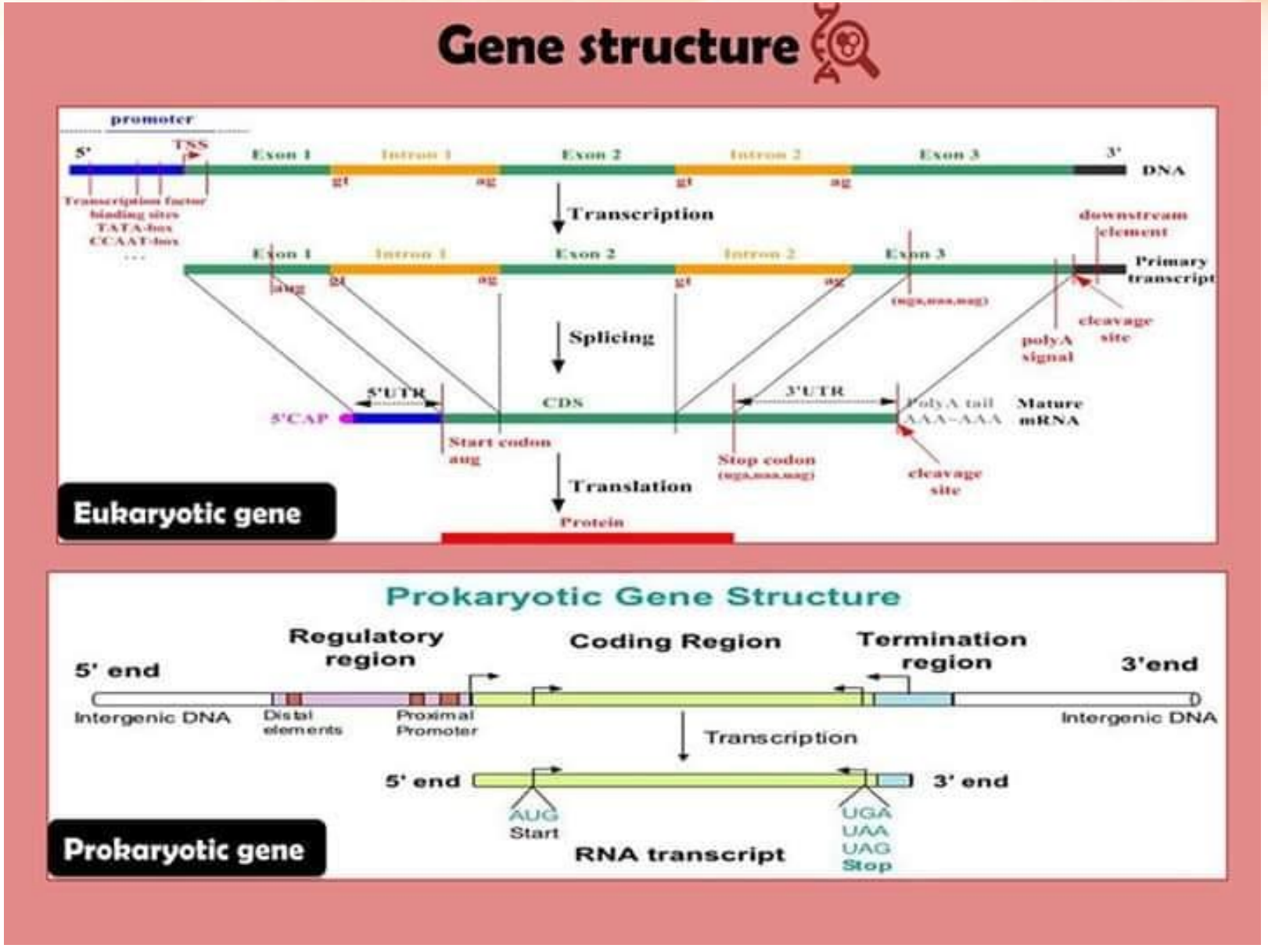
1. The number of copies of the vector in the cell
2. Vector stability

Physiology of the host cell

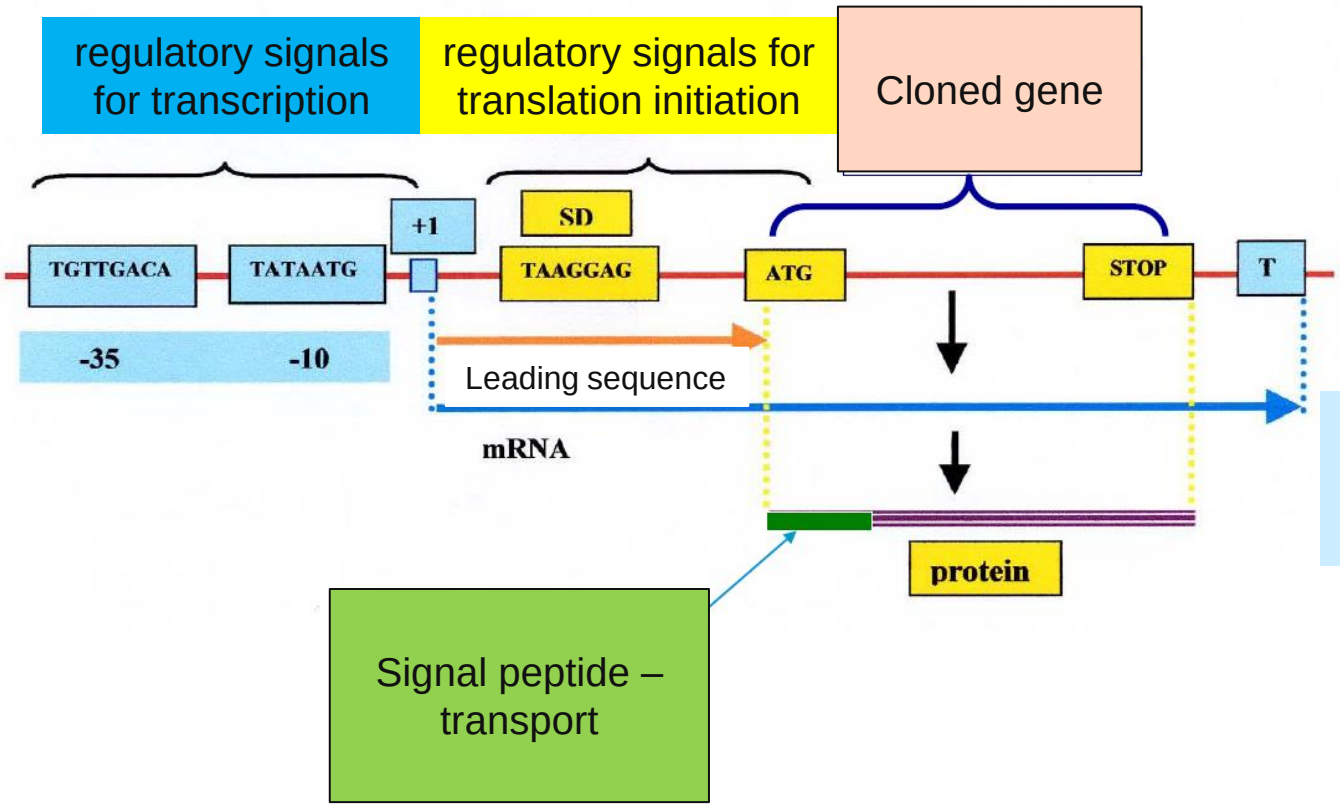
1. Growing conditions
2. The enzyme apparatus of the host cell

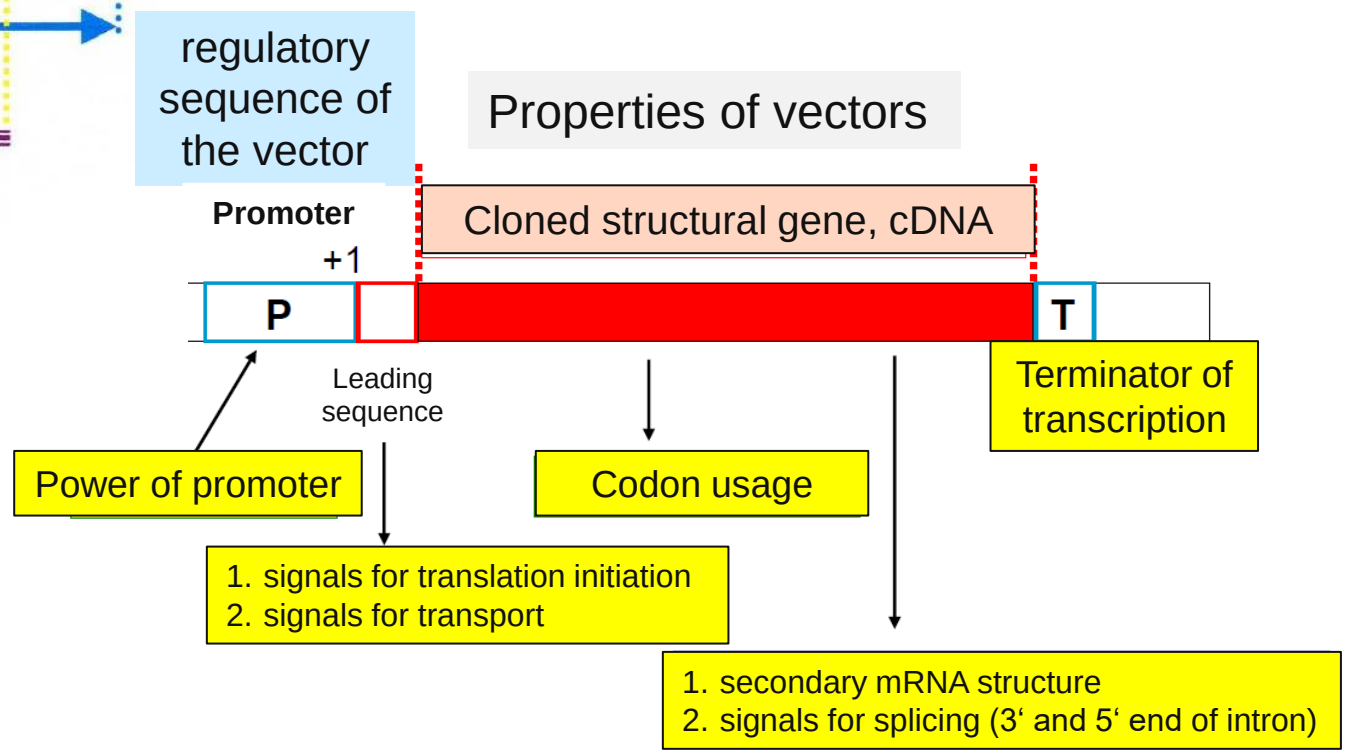
Insert properties

- It contains a coding sequence
- Contains the start of translation (ATG start codon)
- Contains the end of translation (stop codon)
- It carries the appropriate codons for specific amino acids



Signals influencing prokaryotic structural gene transcription and translation





The effect of the distance between the promoter and the start of translation on the secondary structure of mRNA and on the amount of expressed protein

Distance variability

Transcription from *lac* promoter

Recognition sequence for endonuclease *Bam*HI

Translational start signal for *cro* gene

SD_{lac}

SD_{cro}

AATTGTGAGCGGATAACAATTTCACACAGGAAACAGGATCCGGGACTATTTTACCTATGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATG

TTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCCTAGGCCCTGATAAAATGGATACCGCCACTATTACCAACGTACATGATTCCTCCAACATAC

Deletions of varying extent

	Relative amount of CRO protein produced
pTR213	1.63
pTR199	0.085
pTR214	1.00
pTR188	0.65
pTR194	0.51
pTR210	0.85
pTR182	0.0012
pTR190	<0.0006

Relative amount of CRO protein produced

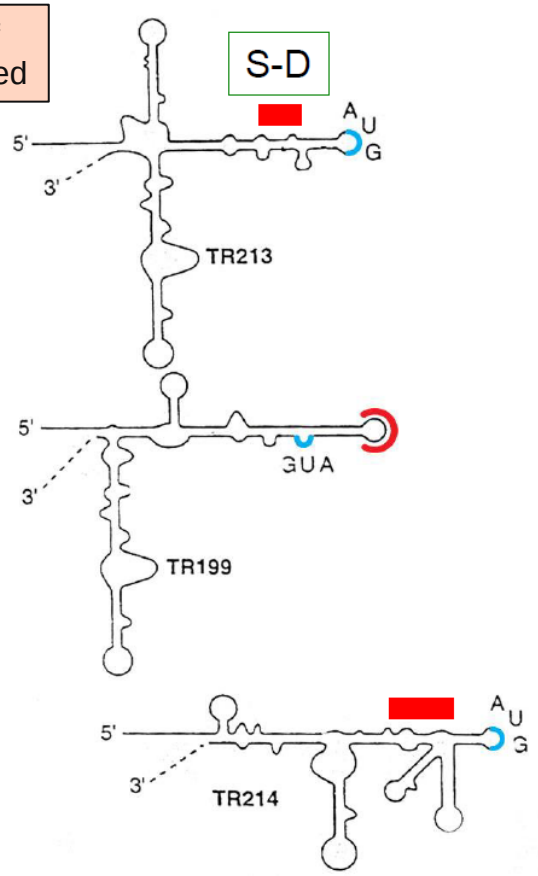
Secondary structure of *cro*-mRNA

Relative amount of CRO protein produced

163%

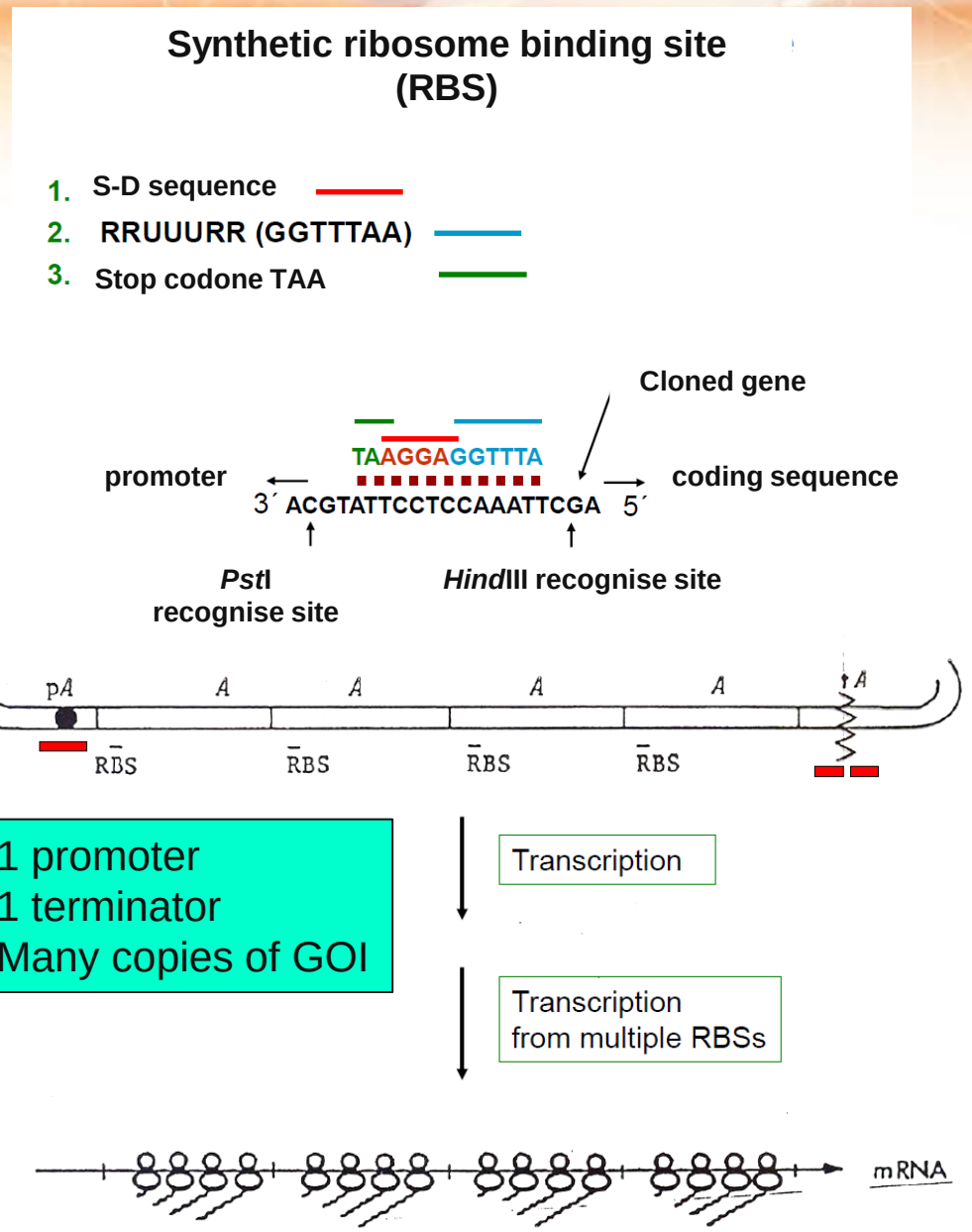
8%

100%



Possibilities of ensuring high proteosynthesis

- The rate of proteosynthesis depends on the amount of mRNA in the cell
- Modifications in the 3'-UTR or 5'-UTR can increase the mRNA half-life
- The **PuPuUUUPuPu** sequence near the Shine-Dalgarn sequence is essential for translation of eukaryotic genes in *E. coli*
- Construction of homopolycistronic expression cassettes
- Proteins in the periplasm have a longer half-life than in the cytoplasm
- Targeted inhibition of specific proteases
- Formation of fusion proteins – protein from the host organism (stabilizing partner) + protein of interest



Effect of codon usage

- Different species use codons with different frequencies → related to the amount of tRNA
- Solution:
  - Cloning rare tRNAs together with a gene of interest
  - Substitution of rare codons for common ones by site-directed mutagenesis

CODON USAGE IN *E. COLI* GENES[1]

	Codon	Amino acid[2]	%[3]	Ratio[4]	Codon	Amino acid	%	Ratio	Codon	Amino acid	%	Ratio	Codon	Amino acid	%	Ratio	
U	UUU	Phe (F)	1.9	0.51	UCU	Ser (S)	1.1	0.19	UAU	Tyr (Y)	1.6	0.53	UGU	Cys (C)	0.4	0.43	U
	UUC	Phe (F)	1.8	0.49	UCC	Ser (S)	1.0	0.17	UAC	Tyr (Y)	1.4	0.47	UGC	Cys (C)	0.6	0.57	C
	UUA	Leu (L)	1.0	0.11	UCA	Ser (S)	0.7	0.12	UAA	STOP	0.2	0.62	UGA	STOP	0.1	0.30	A
	UUG	Leu (L)	1.1	0.11	UCG	Ser (S)	0.8	0.13	UAG	STOP	0.03	0.09	UGG	Trp (W)	1.4	1.00	G
C	CUU	Leu (L)	1.0	0.10	CCU	Pro (P)	0.7	0.16	CAU	His (H)	1.2	0.52	CGU	Arg (R)	2.4	0.42	U
	CUC	Leu (L)	0.9	0.10	CCC	Pro (P)	0.4	0.10	CAC	His (H)	1.1	0.48	CGC	Arg (R)	2.2	0.37	C
	CUA	Leu (L)	0.3	0.03	CCA	Pro (P)	0.8	0.20	CAA	Gln (Q)	1.3	0.31	CGA	Arg (R)	0.3	0.05	A
	CUG	Leu (L)	5.2	0.55	CCG	Pro (P)	2.4	0.55	CAG	Gln (Q)	2.9	0.69	CGG	Arg (R)	0.5	0.08	G
A	AUU	Ile (I)	2.7	0.47	ACU	Thr (T)	1.2	0.21	AAU	Asn (N)	1.6	0.39	AGU	Ser (S)	0.7	0.13	U
	AUC	Ile (I)	2.7	0.46	ACC	Thr (T)	2.4	0.43	AAC	Asn (N)	2.6	0.61	AGC	Ser (S)	1.5	0.27	C
	AUA	Ile (I)	0.4	0.07	ACA	Thr (T)	0.1	0.30	AAA	Lys (K)	3.8	0.76	AGA	Arg (R)	0.2	0.04	A
	AUG	Met (M)	2.6	1.00	ACG	Thr (T)	1.3	0.23	AAG	Lys (K)	1.2	0.24	AGG	Arg (R)	0.2	0.03	G
G	GUU	Val (V)	2.0	0.29	GCU	Ala (A)	1.8	0.19	GAU	Asp (D)	3.3	0.59	GGU	Gly (G)	2.8	0.38	U
	GUC	Val (V)	1.4	0.20	GCC	Ala (A)	2.3	0.25	GAC	Asp (D)	2.3	0.41	GGC	Gly (G)	3.0	0.40	C
	GUA	Val (V)	1.2	0.17	GCA	Ala (A)	2.1	0.22	GAA	Glu (E)	4.4	0.70	GGA	Gly (G)	0.7	0.09	A
	GUG	Val (V)	2.4	0.34	GCG	Ala (A)	3.2	0.34	GAG	Glu (E)	1.9	0.30	GGG	Gly (G)	0.9	0.13	G
		U				C				A				G			

[1] The data shown in this table is from the Arabidopsis Research Companion on the World Wide Web (//weeds/mgh.harvard.edu). Codon frequencies for many other bacteria can be found at http://morgan.angis.su.oz.au/Angis/Tables.html.

[2] The letter in parenthesis represents the one-letter code for the amino acid.

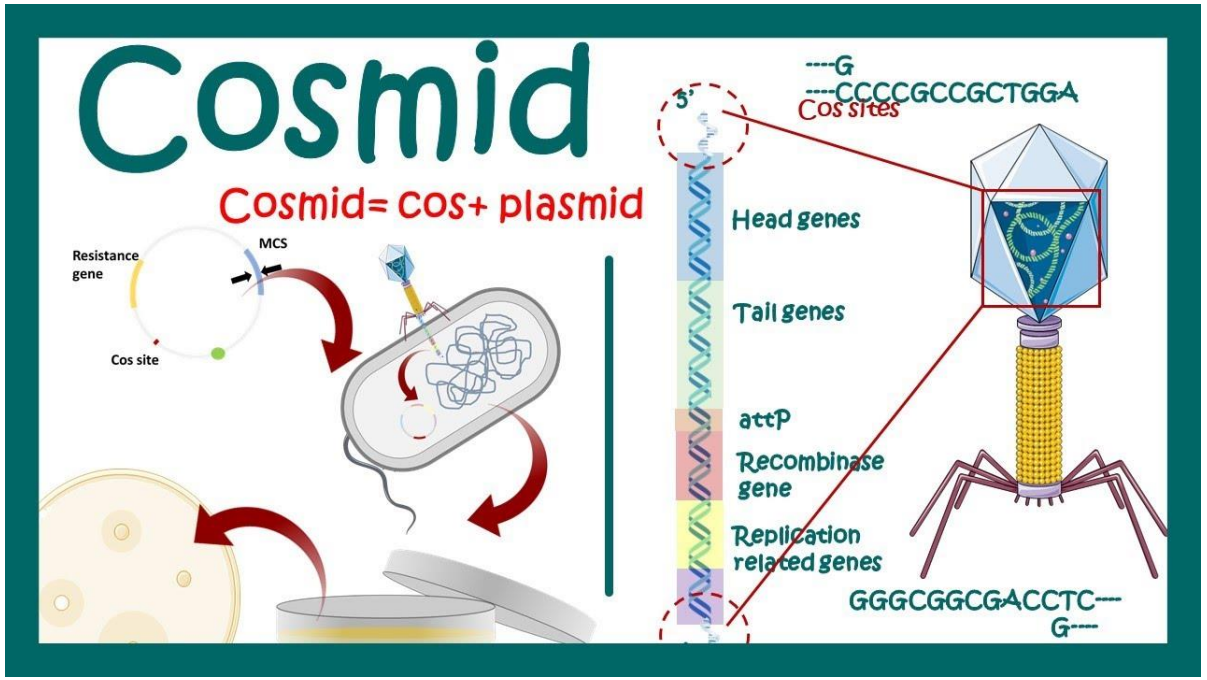
[3] % represents the average frequency this codon is used per 100 codons.

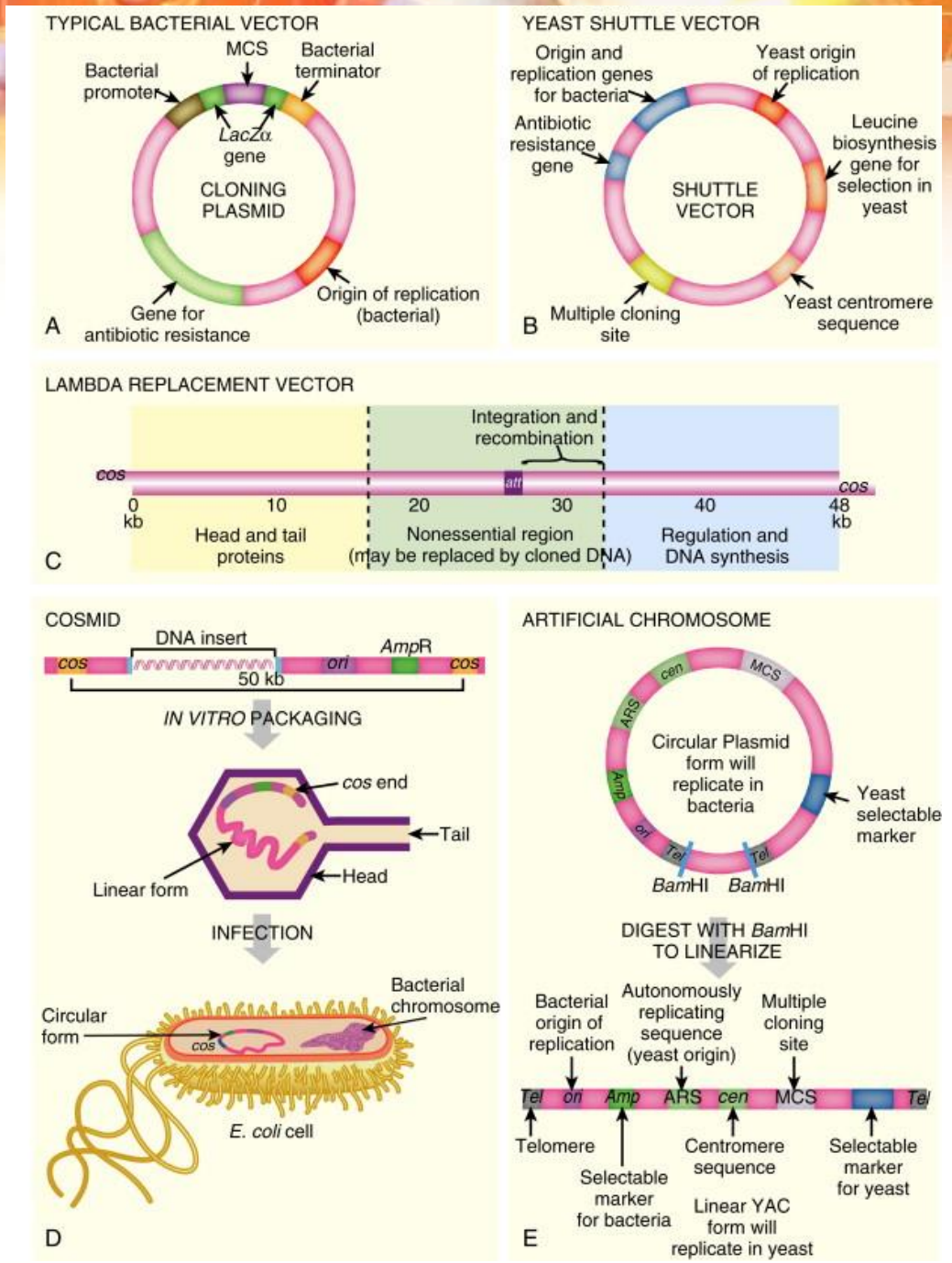
[4] Ratio represents the abundance of that codon relative to all of the codons for that particular amino acid.

http://2014.igem.org/Team:Penn_State/CodonOptimization

Vectors

- Plasmids
- Bacteriophages
- Cosmids
- Artificial chromosomes
  - BAC (bacterial artificial chromosomes)
  - YAC (yeast artificial chromosomes)

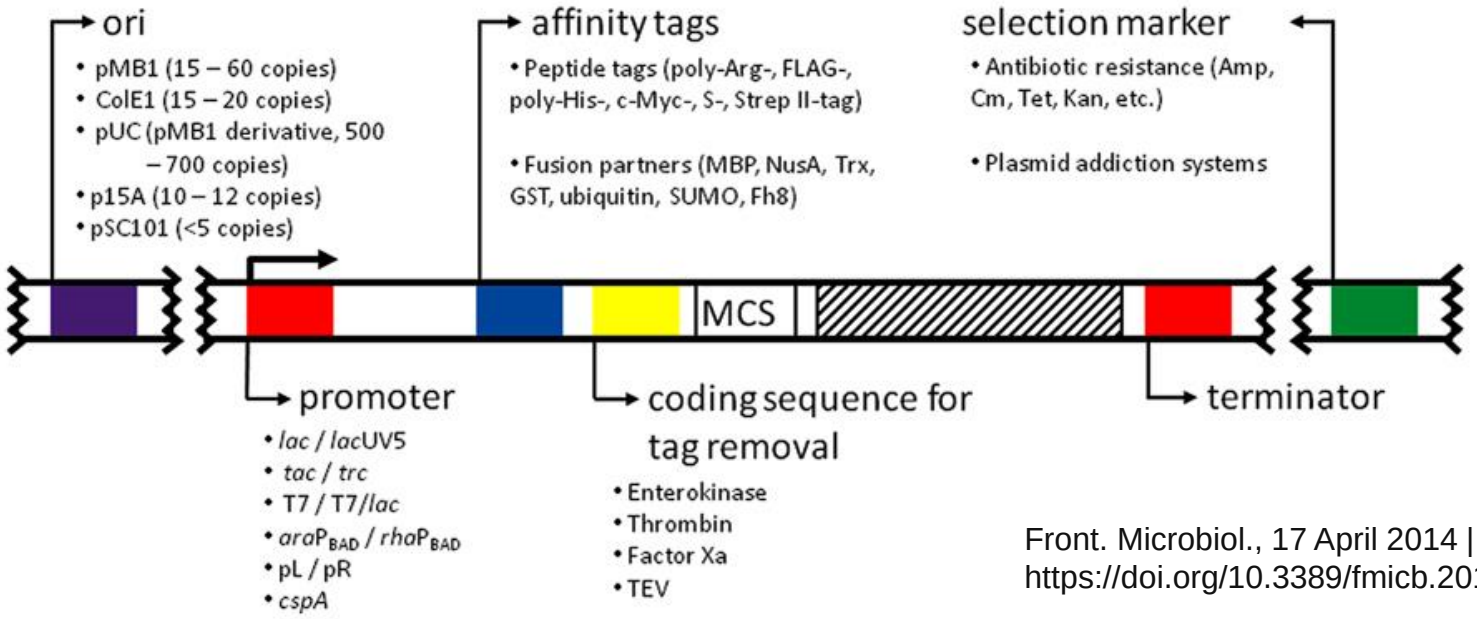




https://www.sciencedirect.com/topics/nursing-and-health-professions/artificial-chromosome

Anatomy of an expression vector

1) **origin of replication (ori), which is a condition for the production of new copies**
2) **inducible promoter, which will allow to regulate the expression of the desired protein**
3) **selection marker ensuring the preferential growth of transformed bacteria**
4) **multi-cloning site (MSC), allowing a foreign DNA fragment to be inserted into the plasmid**



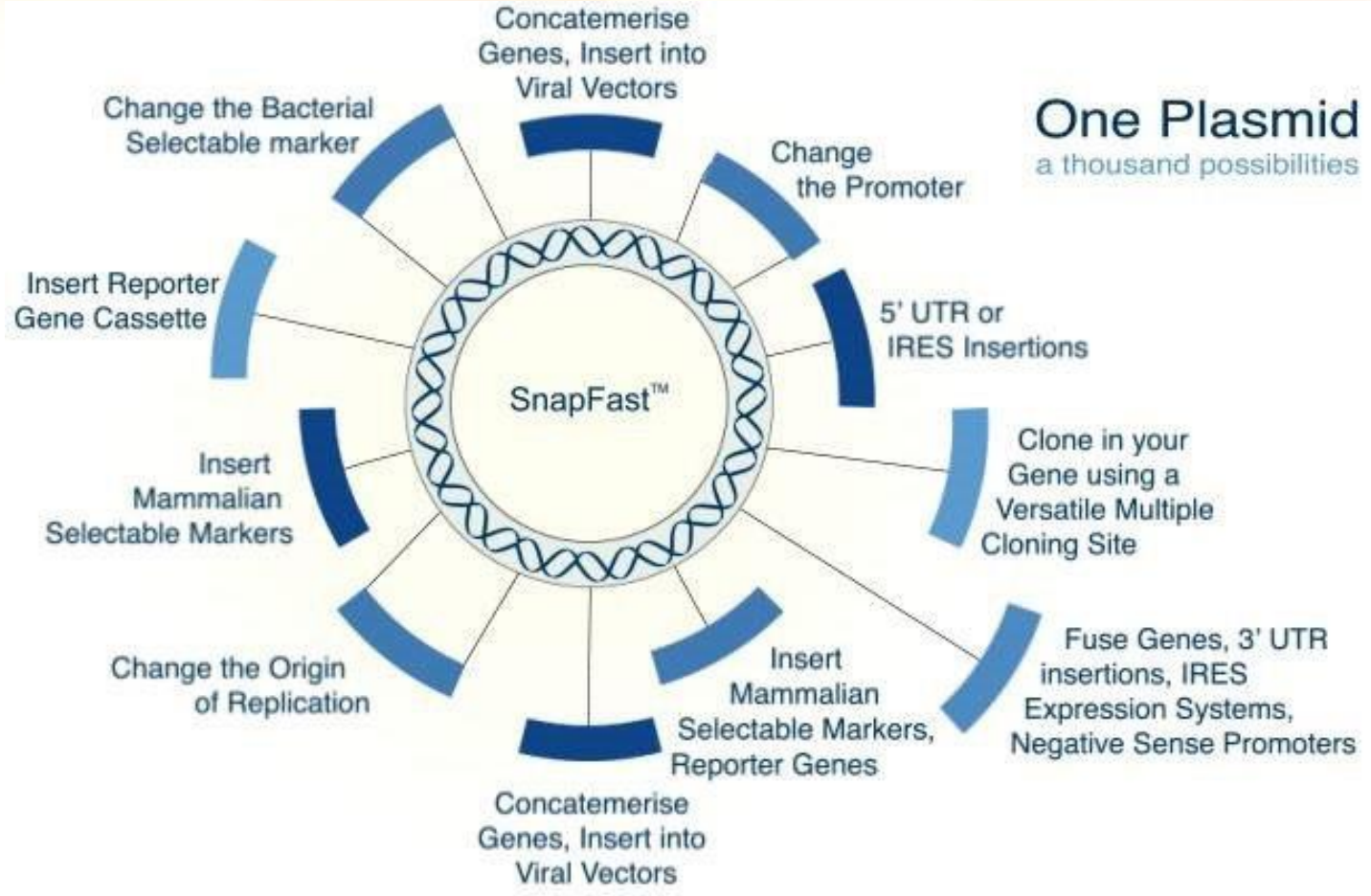
Front. Microbiol., 17 April 2014 | https://doi.org/10.3389/fmicb.2014.00172

Examples of regulatable promoters in expression vectors

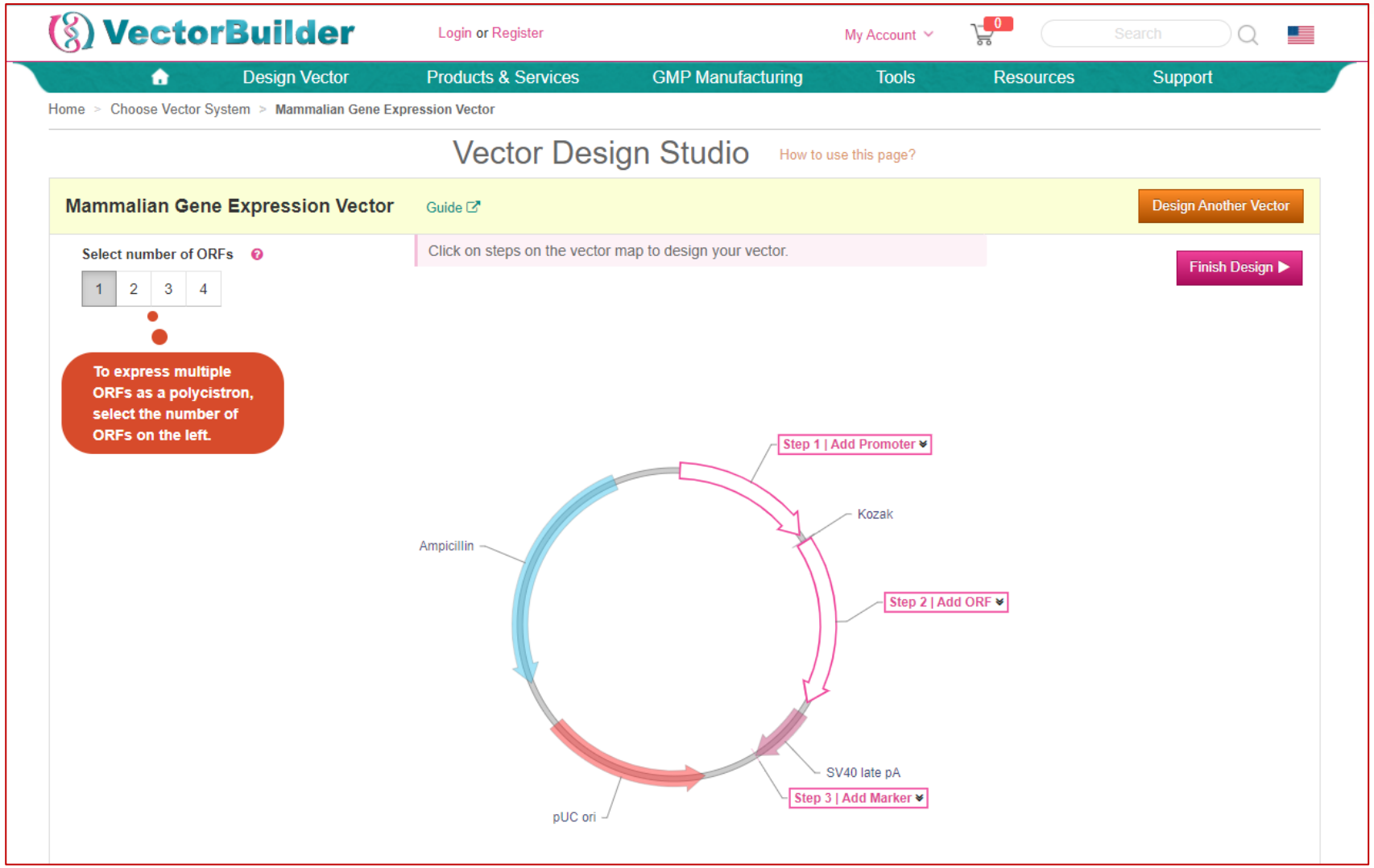
	Promoter	Origin	Way of regulation	
			Off	On
E. coli	λpL, λpR	Leftward and rightward early promoters of λ	30°C	>37°C (in cI_{857} host)
lac-UV5 (independence from catabolic repression)	*lac*	*E. coli lac* operon	—	IPTG in medium
	trp	*E. coli trp* operon	Tryptophan in medium	Indoleacetic acid in medium
	tac	*trp*-35 region *lac*-10 region hybrid	—	IPTG in medium
	phoA	*E. coli* alkaline phosphatase operon	Excess phosphate in medium	Phosphate-limited medium
	recA	*E. coli recA* gene	—	Mitomycin C in medium
utilization in eukaryotes	*tet*	Tn*10* tetracycline-resistance gene	—	Tetracyclines in medium

Plasmids as vectors

- extrachromosomal circular dsDNA
- occurrence in many bacterial species
- 1,000 to 200,000 bp in size
- carry only genes encoding secondary traits (resistance to antibiotics)
- autonomous replication
- **insert size = up to 25 kbp**



Recent trend →
plasmid synthes on demand



Plasmid suitability criteria

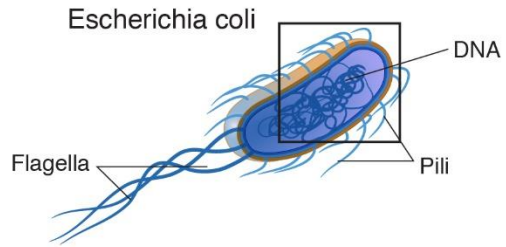
1) small size = ability to transform

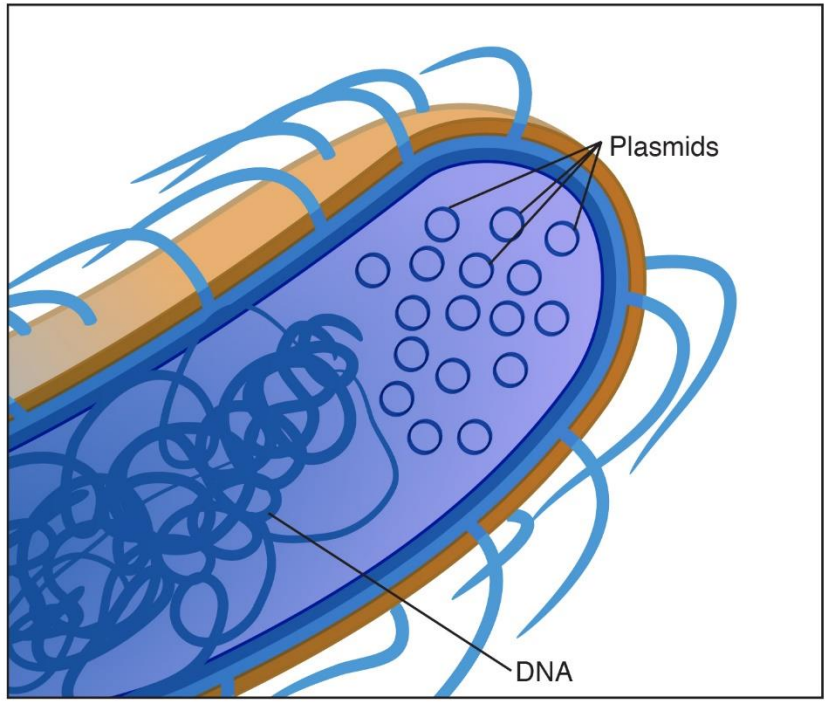
2) plasmid stability

3) high number of copies in the cell = yield

4) easy handling

5) "shuttle" vectors = work in multiple host species (e.g. *E. coli* + mammalian cells)

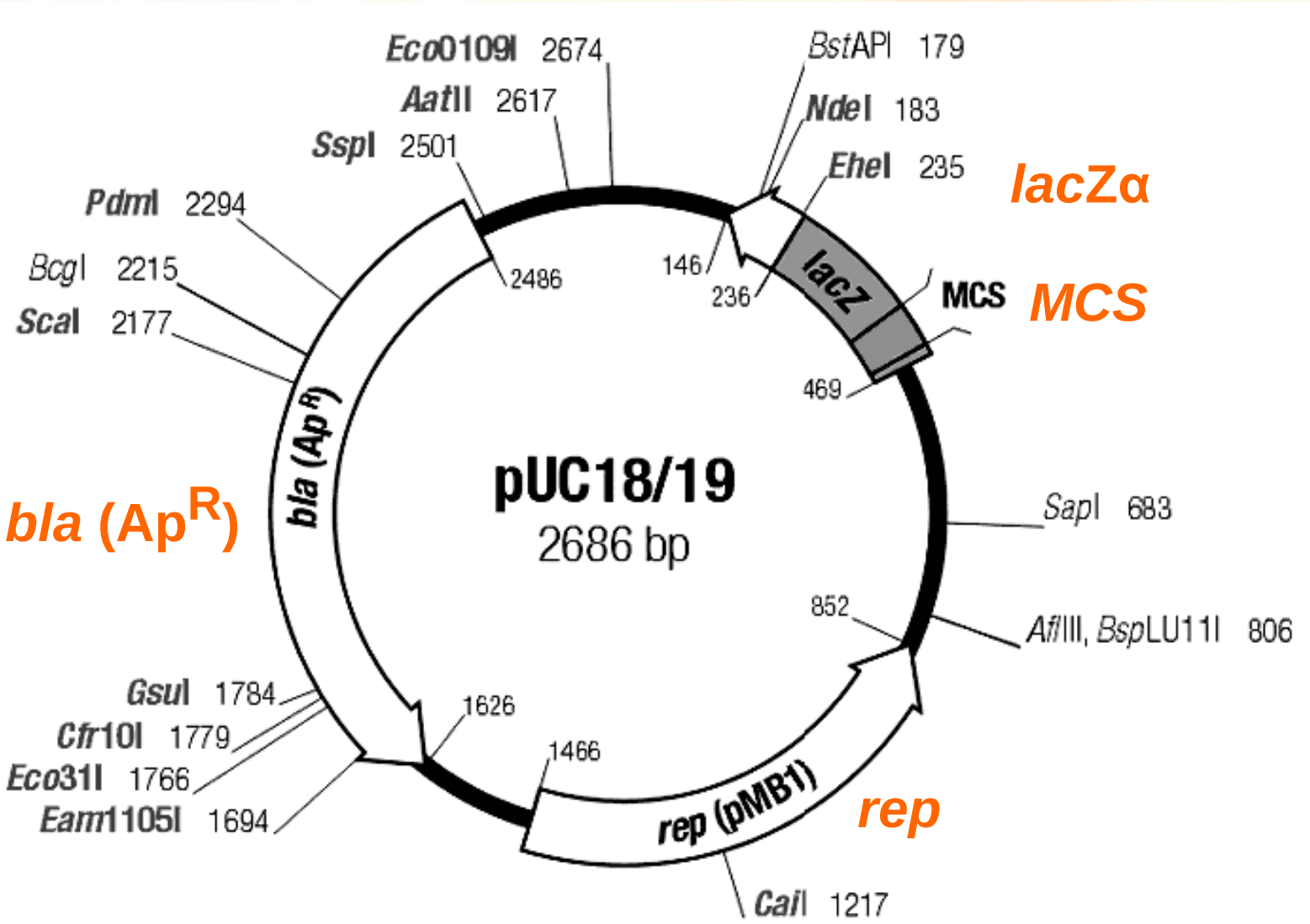




Escherichia coli

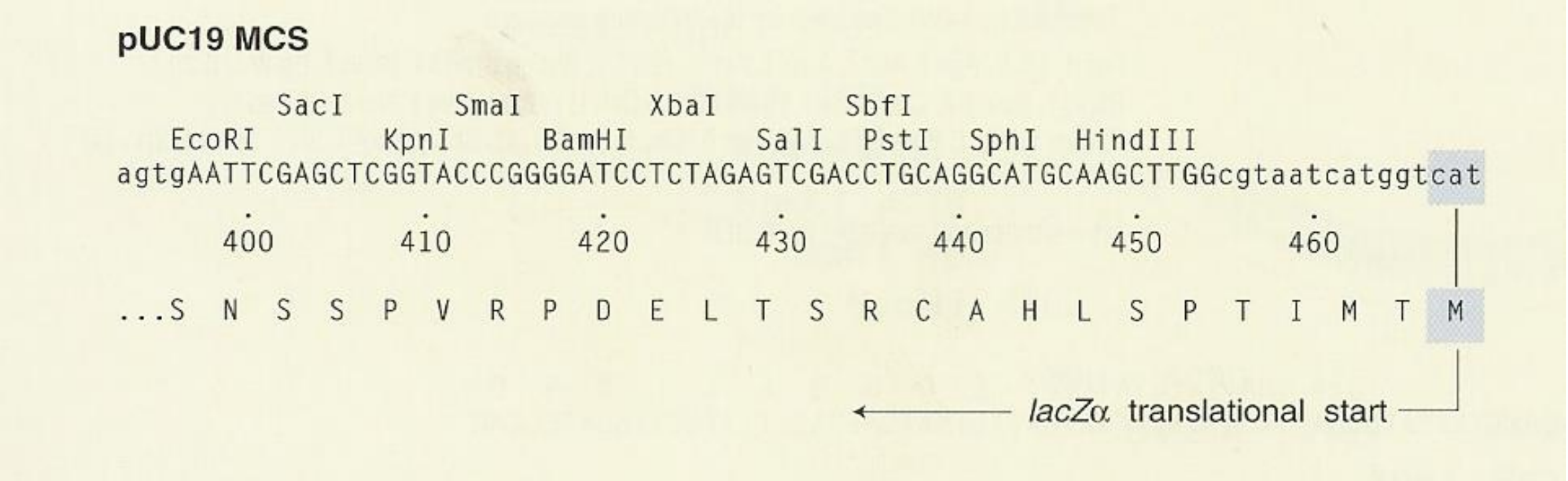
https://www.genome.gov/genetics-glossary/Plasmid

Plasmids pUC18 and pUC19

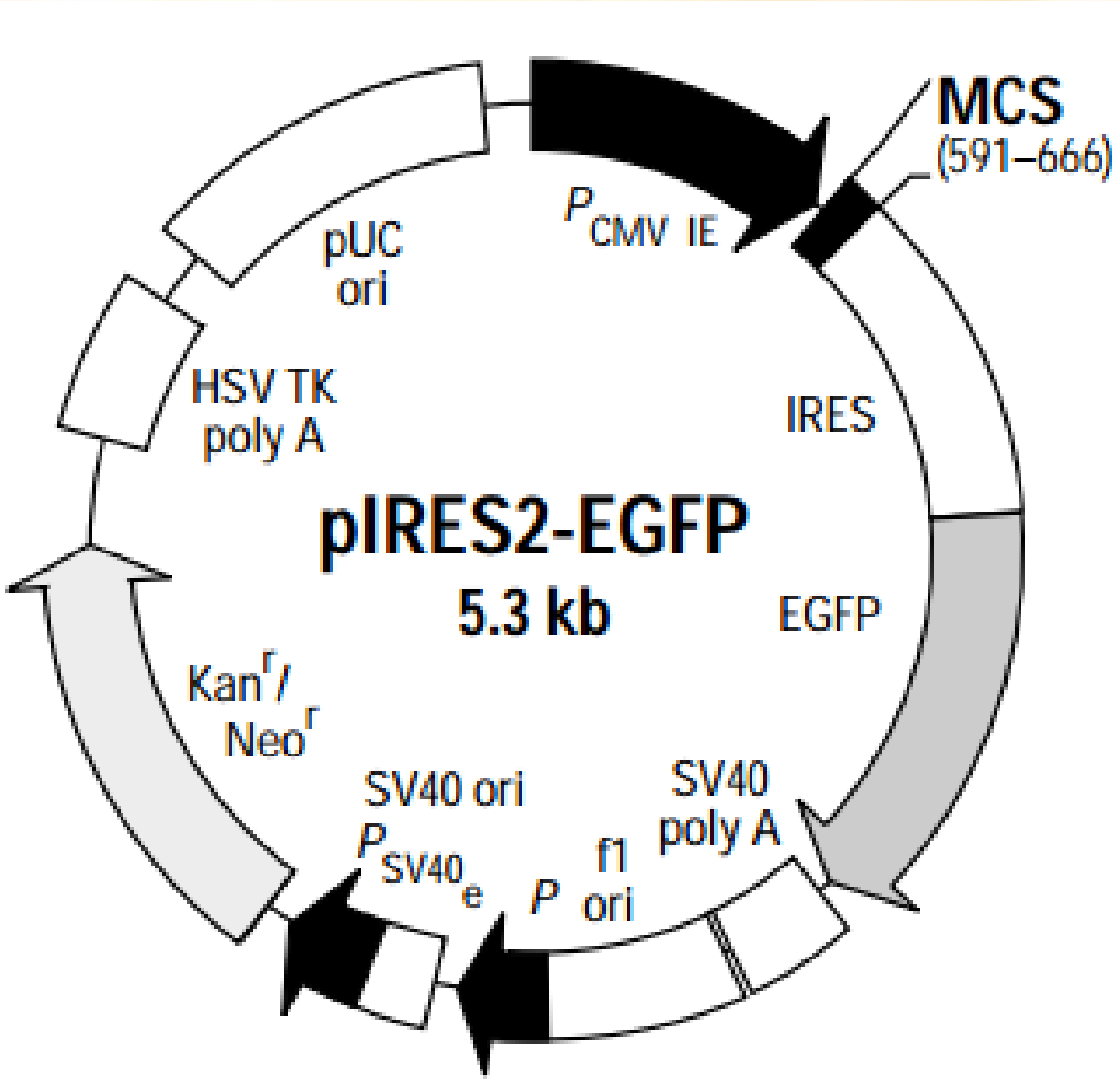


Polylinker pUC18 and pUC19

MCS = multi-cloning site

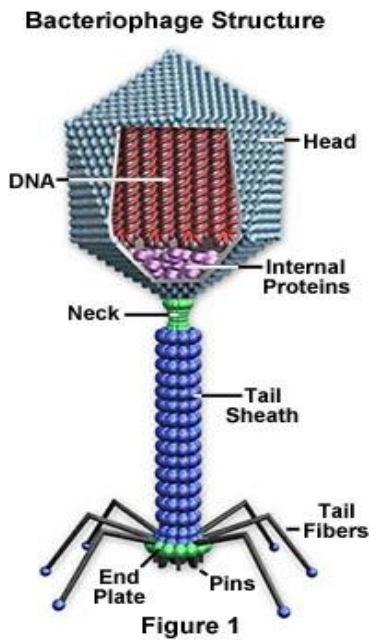


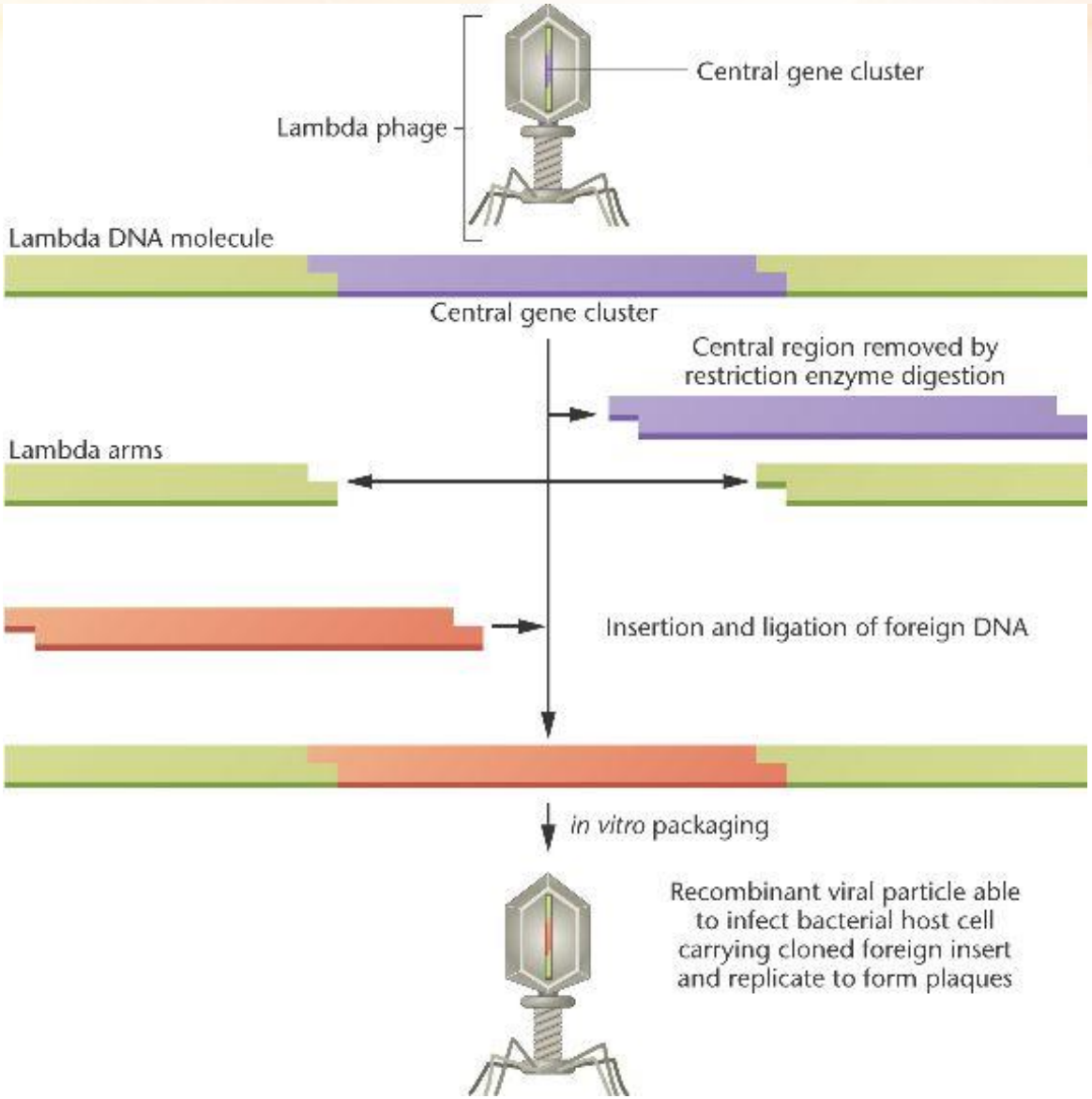
Expression plasmid pIRES2-eGFP



Bacteriophage λ-based vectors

- hey replace plasmids when longer DNA fragments need to be cloned
- **Insertion vectors 8 – 10 kbp**
- **Replacement vectors 8 – 24 kbp**

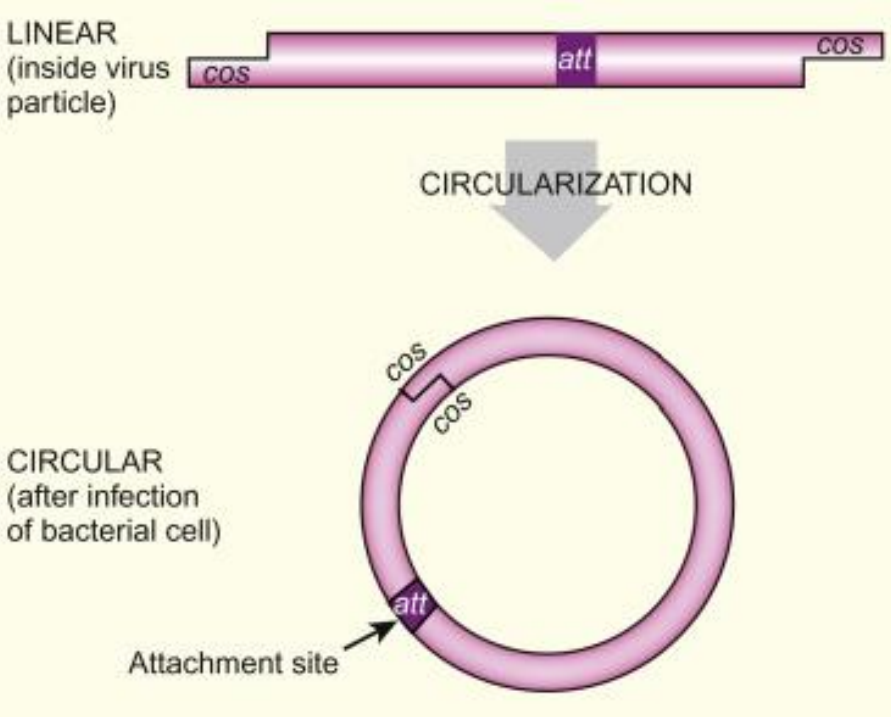


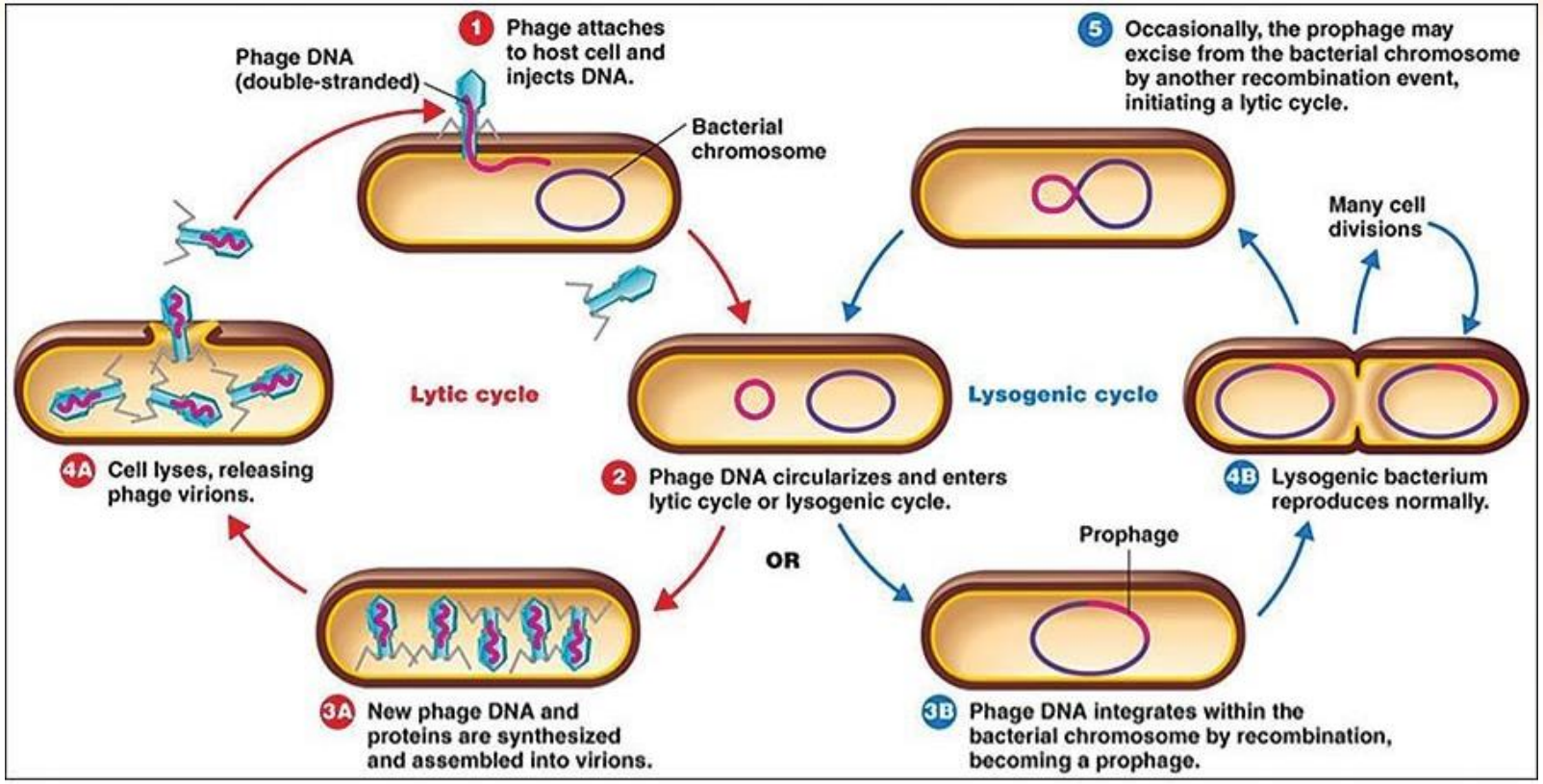


https://bio3400.nicerweb.com/Locked/media/ch19/lambda-vector.html

Bakteriophage λ

- 50 kbp dsDNA
- linear and circular form
- *cos* sites





http://hotcore.info/babki/bacteriophage-lysogenic-cycle.html

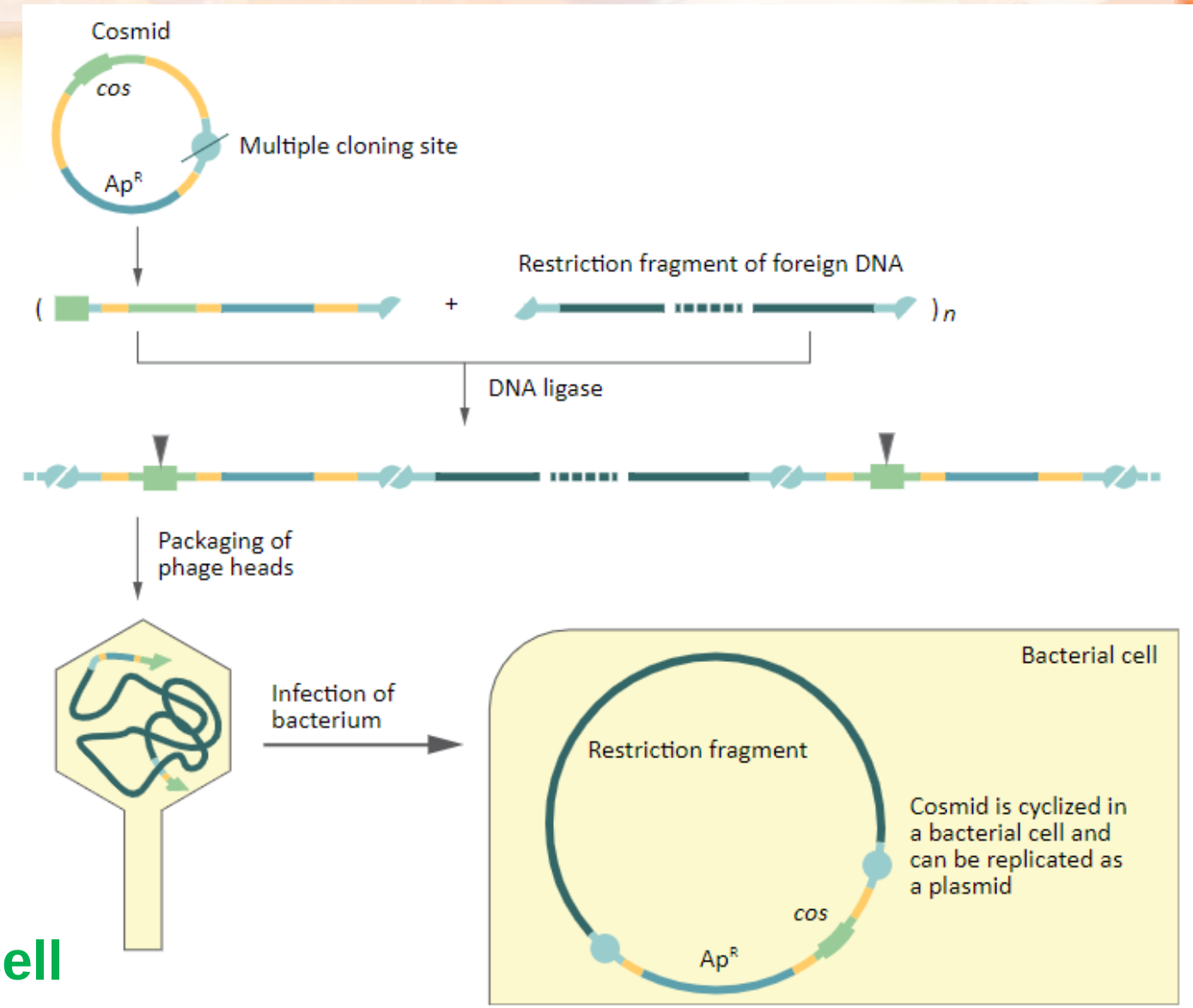
https://doi.org/10.1016/C2009-0-01986-2

Cosmids

- combination of plasmid and phage
- prokaryotic origin of replication oriV
- selection marker
- cloning site

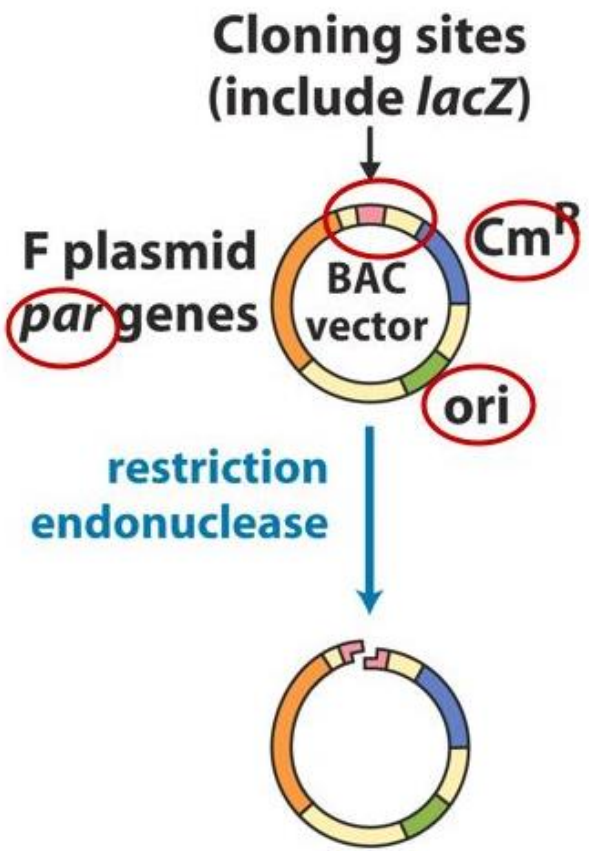
- capacity 37 - 52 kbp
- cos sites of bacteriophage λ
- for packaging, packaging proteins must be added

- enters as a phage
- it behaves like a plasmid in the cell



Bacterial Artificial Chromosome

- BAC = bacterial artificial chromosome
- Derived from plasmid F'
- Designed for cloning into bacterial cells
- They occur in the number of 1-2 copies per cell
- Cloned DNA is highly stable
- **Cloning capacity up to 300 kbp (maybe more)**



Artificial Chromosomes allow for cloning of large pieces of DNA

Bacterial Artificial Chromosome

1. *ori* allows for replication in bacteria,
2. *par* helps segregate BAC evenly between daughter cells,
3. *lac Z* allows for detection of insert,
4. Cm^R allows for selection of transformed cells.

https://library.uams.edu/assets/COM/BioChem/MolecularTools/MolecularToolsSDL12.html

- **Used in the HUGO project**
- **Today, they are replaced by the methods of whole-genome sequencing, next generation sequencing and third generation sequencing**

Other variants of vectors

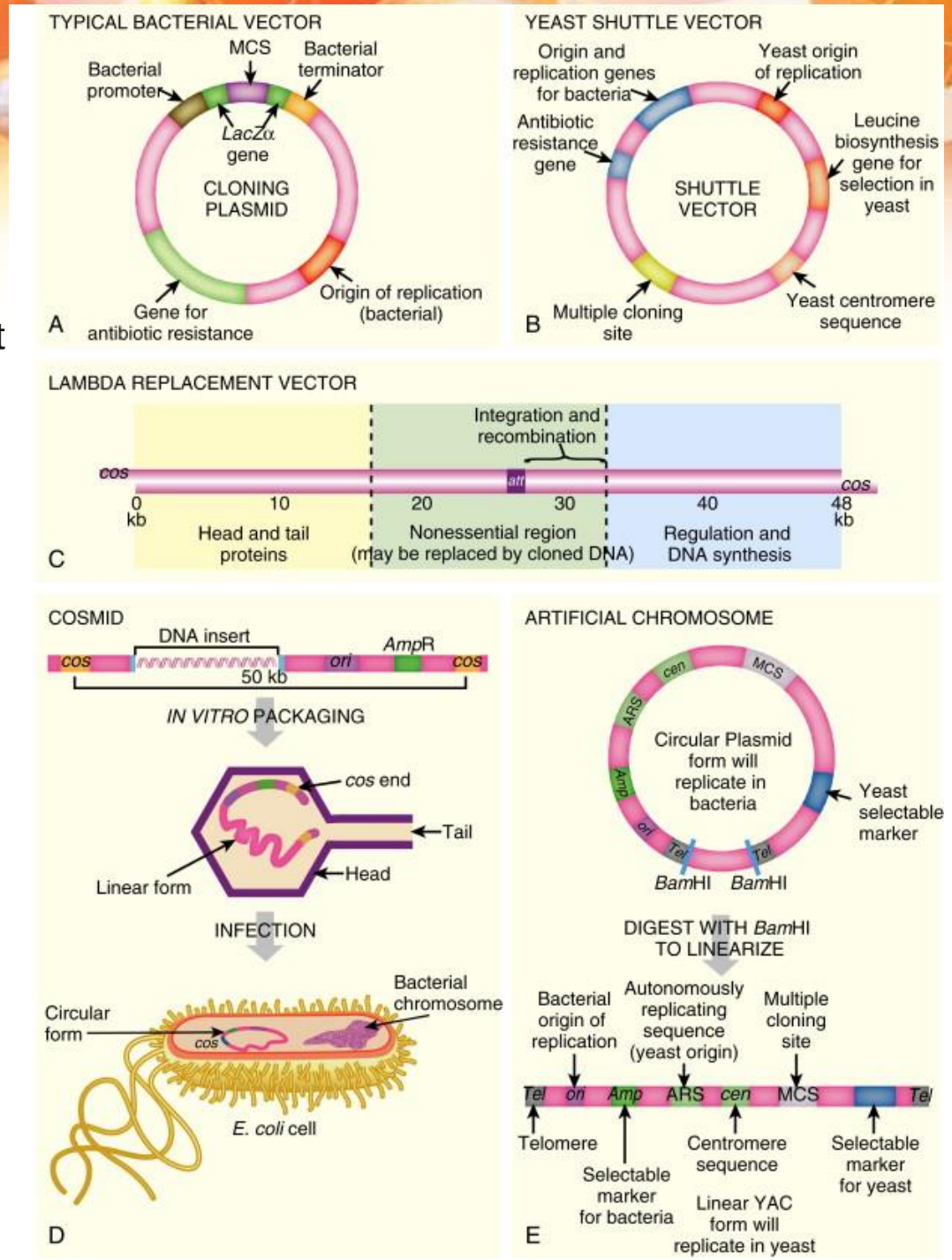
Phosmid

- Similar to a cosmid, but based on a bacterial F-plasmid
- The host (*E. coli*) may contain only a single molecule
- Cloning capacity up to 40 kbp
- Suitable for constructing stable libraries from complex genomes
- Highly stable; capable of maintaining human DNA for over 100 generations

Phagemid

- Plasmid containing origin of replication of phage M13
- It is used to prepare ssDNA.
- The best known examples are the pBluescript series of cloning vectors.

(A) Typical bacterial cloning vector. This vector has bacterial sequences to initiate replication and transcription. In addition, it has a multiple cloning site embedded within the *lacZ α* gene so that the insert can be identified by alpha-complementation. The antibiotic resistance gene allows the researcher to identify any *E. coli* cells that have the plasmid. (B) Yeast shuttle vector. This vector can survive in either bacteria or yeast because it has both yeast and bacterial origin of replication, a yeast centromere, and selectable markers for yeast and bacteria. As with most cloning vectors, there is a polylinker. (C) Lambda replacement vectors. Because lambda phage is easy to grow and manipulate, its genome has been modified to accept foreign DNA inserts. The region of the genome shown in green is nonessential for lambda growth and packaging. This region can be replaced with large inserts of foreign DNA (up to about 23 kb). (D) Cosmids. Cosmids are small multicopy plasmids that carry *cos* sites. They are linearized and cut so that each half has a *cos* site (not shown). Next, foreign DNA is inserted to relink the two halves of the cosmid DNA. This construct is packaged into lambda virus heads and used to infect *E. coli.* (E) Artificial chromosomes. Yeast artificial chromosomes have two forms: a circular form for growing in bacteria and a linear form for growing in yeast. The circular form is maintained like any other plasmid in bacteria, but the linear form must have telomere sequences to be maintained in yeast. The linear form can hold up to 2000 kb of cloned DNA and is very useful for genomics research.



https://doi.org/10.1016/C2009-0-64257-4

Hosts = recipients of recombinant DNA

Regardless of source type

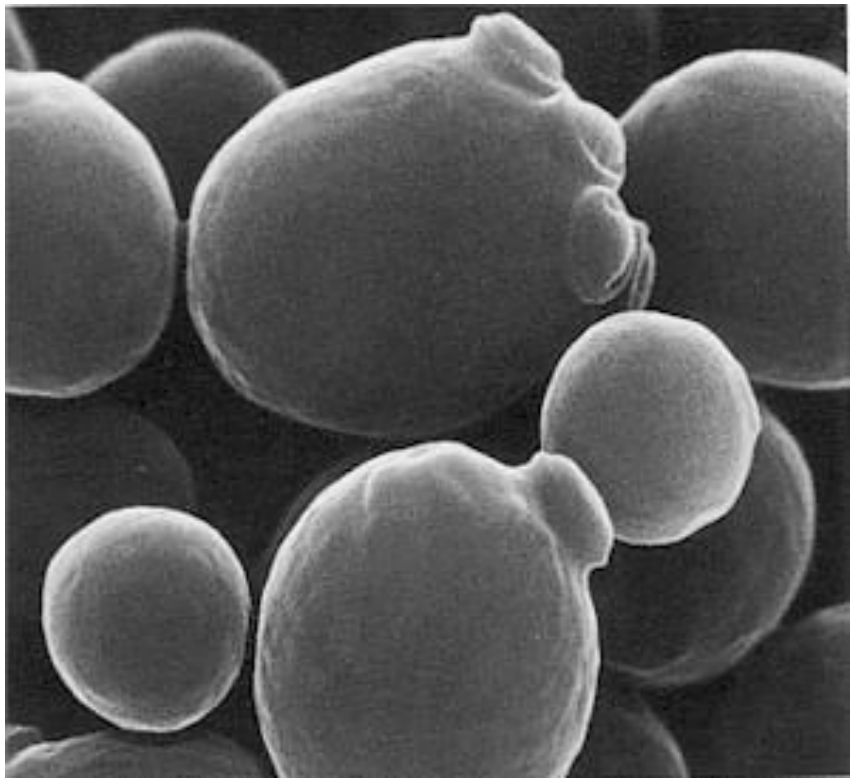
- **bacterial cells**
- **yeasts and molds**
- **plant and animal cells**
- **whole plant or animal**

Escherichia coli

- G- bacteria, circular chromosome 3×10^6 bp
- amount of usable plasmids
- generation time 20 min. → rapid biomass formation
- undemanding and cheap cultivation
- stationary phase 2×10^9 cells/mL
- a number of mutants (DH5α, HB101, BL21,…)
- **disadvantages** – significantly different post-translational modifications compared to eukaryotes

Saccharomyces cerevisiae

- linear chromosomes
- approximately 13×10^6 bp
- about 6,275 genes

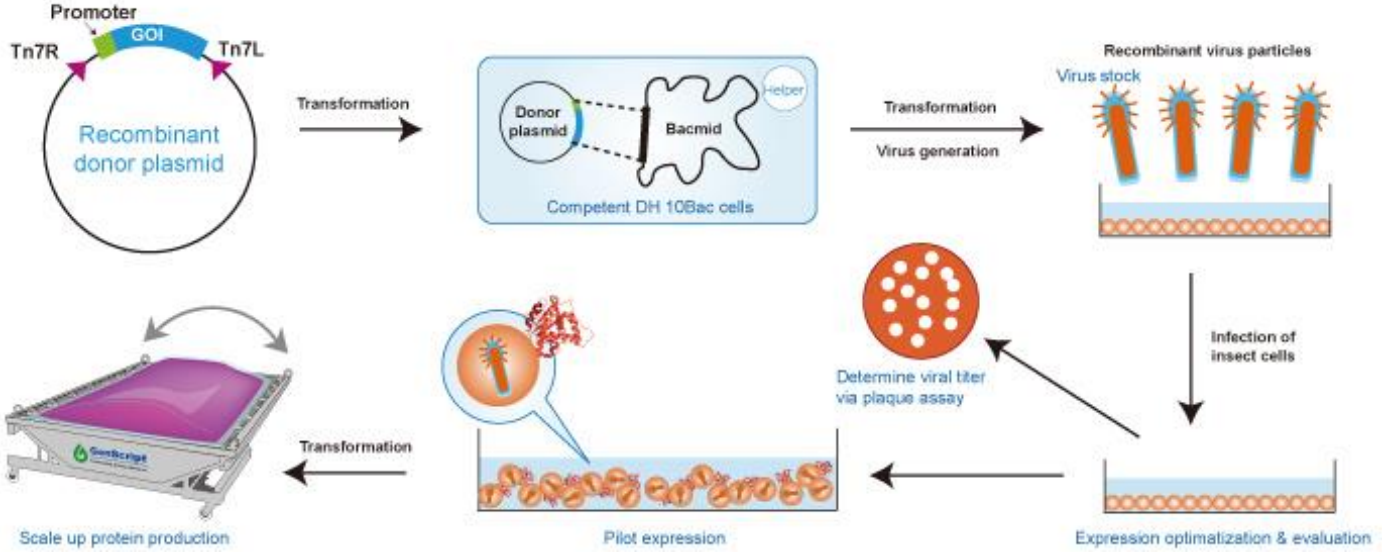


- the simplest eukaryotic organism
- identical transcription and translation apparatus with other eukaryotes
- differences in post-translational processes, **e.g. mannose hyperglycosylation**

- ***Schizosaccharomyces pombe***
- ***Pichia pastoris***

Insect cells and baculoviruses

- similar selection principles as in lower eukaryotes or bacteria
- use of binary "shuttle" vectors (bacteria + insects)
- replacement of the polyhedrin sequence of the virus with a recombinant gene
- the most commonly used Sf9 cell culture derived from *Spodoptera frugiperda*
- **baculoviruses only attack insect cells**



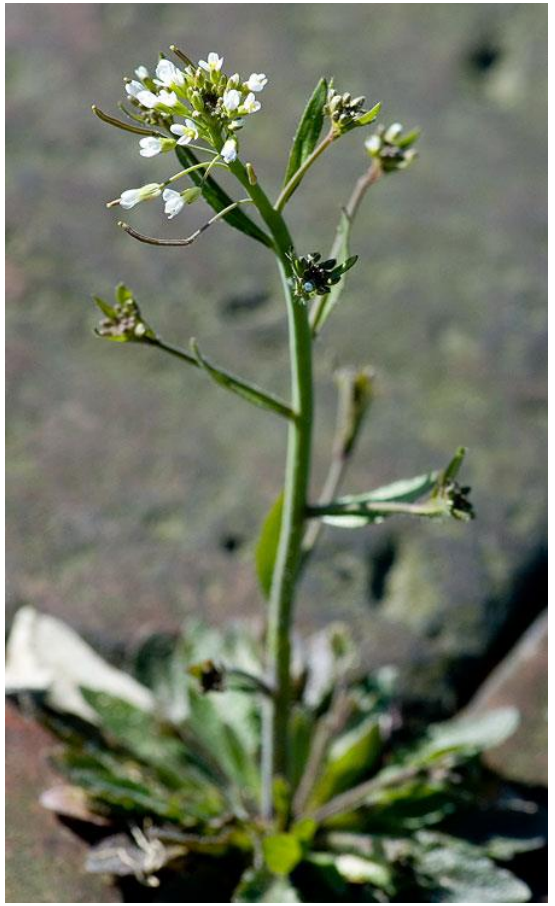
https://www.genscript.com/insect-customized-expression-package.html

Plant cells

- **vectors are typically bacterial plasmids that contain plant expression cassettes**

- **direct transformation**
- **transformation using Agrobacterium**
- **transformation by viral vector**

- **host cells – *Nicotiana tabacum*, *Arabidopsis thaliana*, …**



Relatively safe technology

Mammalian cells and their viruses

- systems closest to man
- the most common producer is mammalian CHO cells (Chinesse hamster ovaries)
- differential interspecies glycosylation
- as vectors serve adenoviruses, retroviruses, herpesviruses

Technology security issues !

Creation of recombinant proteins in different organisms

Parameter	Bacteria	Yeast	Mammalian cell culture	Transgenic plants
Glycosylation	None	Incorrect	Correct	Generally correct
Assembles multimeric proteins	Limited	Limited	Limited	Yes
Production costs	Medium	Medium	High	Low
Protein-folding accuracy	Low	Medium	High	High
Protein yield	High	High	Medium	Medium
Scale-up costs	High	High	High	Low
Time required	Low	Low	High	Medium
Skill level required for growth	Medium	Medium	High	Low