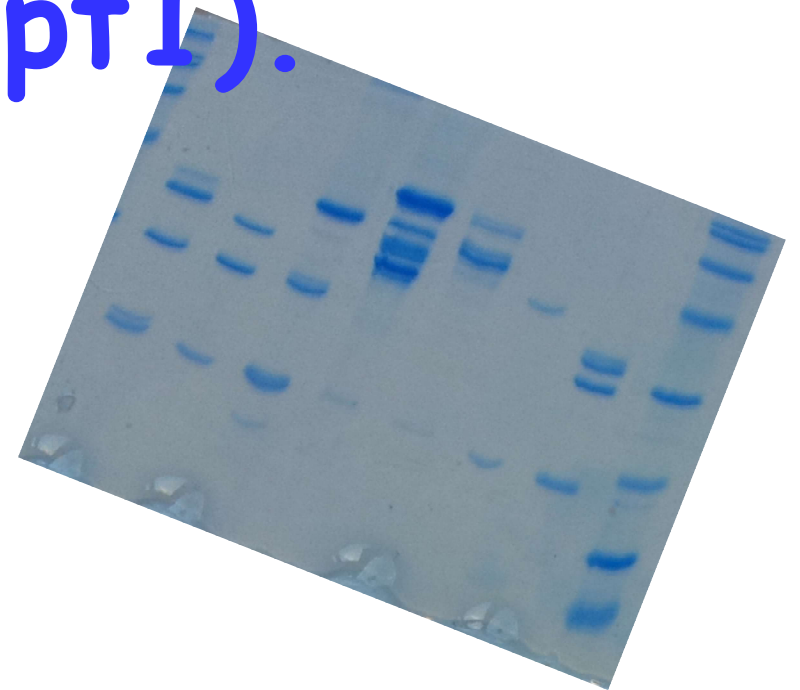


**Proteine ricombinanti
fantastiche e come
ottenerle (pt1).**



Protein structural studies

- Structural studies provide the basis for understanding forms and functions of proteins.
- X-ray crystallography, NMR and cryoEM can guide the drug discovery.
- Sample preparation is a key step.
- Natural sources or protein engineering (recombinant proteins; parallel approaches and automation)

Structural techniques

- **X-ray crystallography**: High purity, stability, monodisperse, highly concentrated.
- **NMR**: High purity, stability, monodisperse, highly concentrated, isotopic labelling.
- **EM e CryoEM**: High purity, stability, monodisperse.
- **SAXS**: High purity, stability, monodisperse.

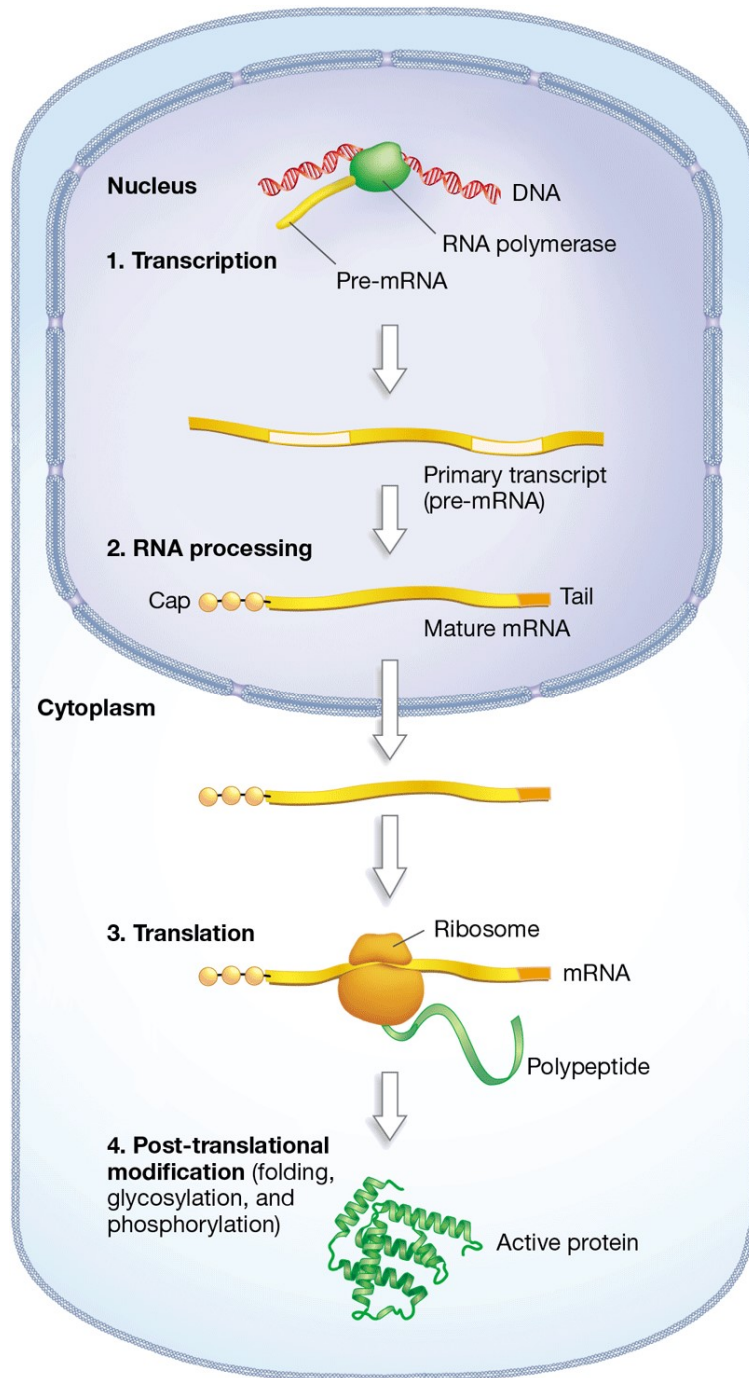
Protein sources:

Natural vs recombinant

Even if natural sources are still exploited especially for the purification of large complexes, most of the proteins for structural biology are currently "home-made" as recombinant proteins.

- Regulate the over-expression in a variety of different cell systems
(*e.g. E. coli*, yeast, insect cells, mammalian cells)
- Easier to handle than large animals
- Easier to change specie
- Easier purification (tags)
- Allows to modify the protein
- Possible to co-express more than one protein to make a complex

Protein synthesis in cells



1. Transcription: a tight regulated process. RNA polymerase synthesize a new filament of messenger RNA (pre mRNA)

2. RNA processing: RNA maturation and splicing, capping and transport outside the nucleus

3. Translation: Ribosomes bind the mRNA, tRNA and new polypeptide synthesis.

4. Post-translational modification: folding, glycosylation, phosphorylation, active protein

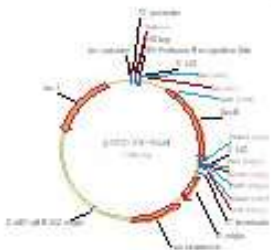
Pipeline



- ★ Selection gene of interest, sequence analysis system expression and DNA source



- ★ Primer design and PCR reaction



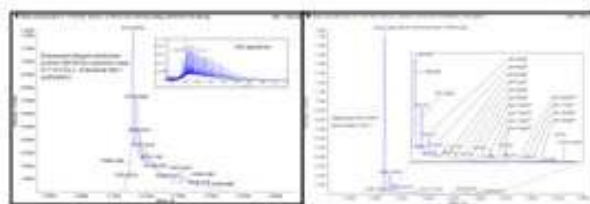
- ★ Cloning and transformation/transfection



- ★ Test expression and purification



- ★ Scale up



- ★ Protein characterization and quality control

Literature and preliminary knowledge, what do I know?

- Literature research: PubMed, Uniprot (organism, cellular localization, function, post-translational modification, substrates, interaction partners).

The screenshot shows the UniProtKB website interface. At the top, there is a search bar with the UniProt logo and a navigation menu with links like BLAST, Align, Retrieve/ID mapping, Peptide search, and SPARQL. A banner below the search bar reads "The new UniProt website is here! Take me to UniProt BETA". The main heading is "UniProtKB - Q96RI1 (NR1H4_HUMAN)". Below this, there is a "Display" section with tabs for Entry, Publications, Feature viewer, and Feature table. The "Entry" tab is selected, showing the protein's name "Bile acid receptor", gene name "NR1H4", organism "Homo sapiens (Human)", and status "Reviewed - Annotation score: 5.00 - Experimental evidence at protein level". The "Function" section is expanded, showing a detailed description of the protein's role as a ligand-activated transcription factor, its involvement in bile acid homeostasis, and its interaction with various receptors and corepressors. The "Function" section also includes a list of related proteins and their functions.

UniProtKB - Q96RI1 (NR1H4_HUMAN)

Display [Help video](#) [BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#) [Add a publication](#) [Feedback](#)

Entry [Publications](#) [Feature viewer](#) [Feature table](#)

Protein **Bile acid receptor**
Gene **NR1H4**
Organism *Homo sapiens (Human)*
Status [Reviewed](#) - Annotation score: 5.00 - Experimental evidence at protein levelⁱ

Functionⁱ

Ligand-activated transcription factor. Receptor for bile acids (BAs) such as chenodeoxycholic acid (CDCA), lithocholic acid, deoxycholic acid (DCA) and allocholic acid (ACA). Plays an essential role in BA homeostasis through the regulation of genes involved in BA synthesis, conjugation and enterohepatic circulation. Also regulates lipid and glucose homeostasis and is involved in innate immune response (PubMed:10334992, PubMed:10334993, PubMed:21383957, PubMed:22820415).

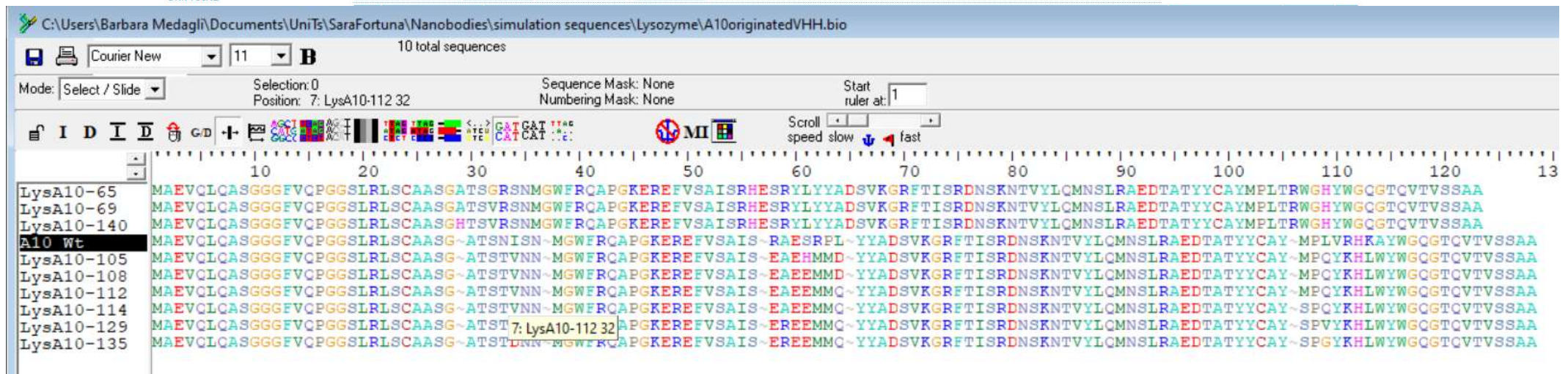
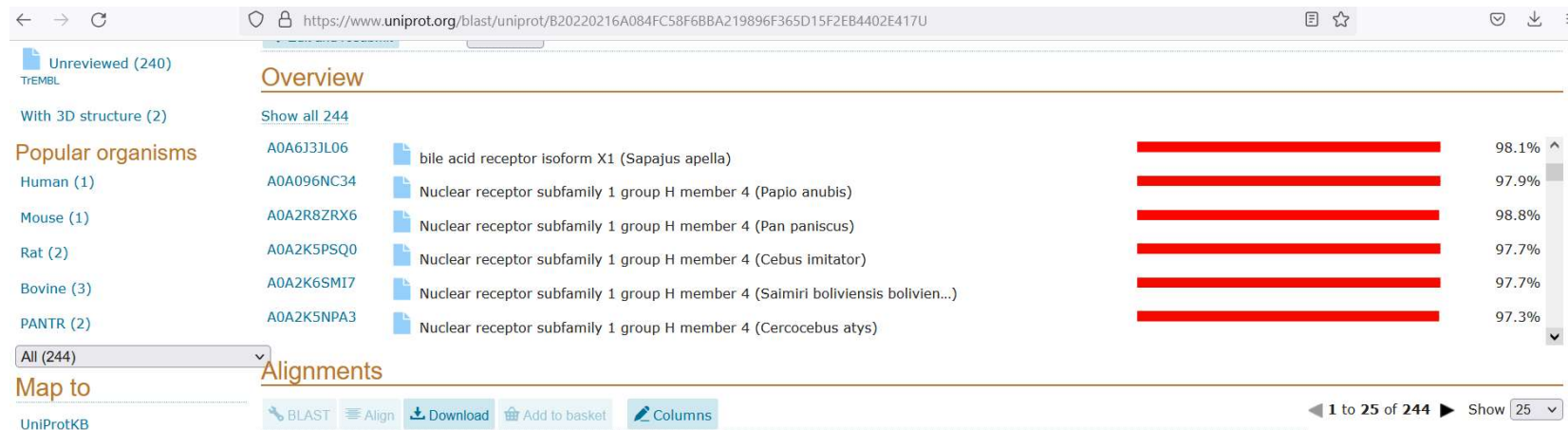
The FXR-RXR heterodimer binds predominantly to farnesoid X receptor response elements (FXREs) containing two inverted repeats of the consensus sequence 5'-AGGTCA-3' in which the monomers are spaced by 1 nucleotide (IR-1) but also to tandem repeat DR1 sites with lower affinity, and can be activated by either FXR or RXR-specific ligands. It is proposed that monomeric nuclear receptors such as NR5A2/LRH-1 bound to coregulatory nuclear responsive element (NRE) halfsites located in close proximity to FXREs modulate transcriptional activity (By similarity).

In the liver activates transcription of the corepressor NR0B2 thereby indirectly inhibiting CYP7A1 and CYP8B1 (involved in BA synthesis) implicating at least in part histone demethylase KDM1A resulting in epigenomic repression, and SLC10A1/NTCP (involved in hepatic uptake of conjugated BAs). Activates transcription of the repressor MAFG (involved in regulation of BA synthesis) (By similarity).

Activates transcription of SLC27A5/BACS and BAAT (involved in BA conjugation), ABCB11/BSEP (involved in bile salt export) by directly recruiting histone methyltransferase CARM1, and ABCC2/MRP2 (involved in secretion of conjugated BAs) and ABCB4 (involved in secretion of phosphatidylcholine in the small intestine) (PubMed:12754200, PubMed:15471871, PubMed:17895379).

In silico analysis

- Analysis of the peptide sequence. BLAST, Expasy, Expresso, Phyre, Protein Data Bank (homologous search, sequence analysis, **sequence alignment**, secondary structure prediction, structure)



Constructs design

Definition of domain boundaries: to increase the chance of crystallization, all the flexible parts of protein are removed, as well as the transmembrane domain for membrane proteins

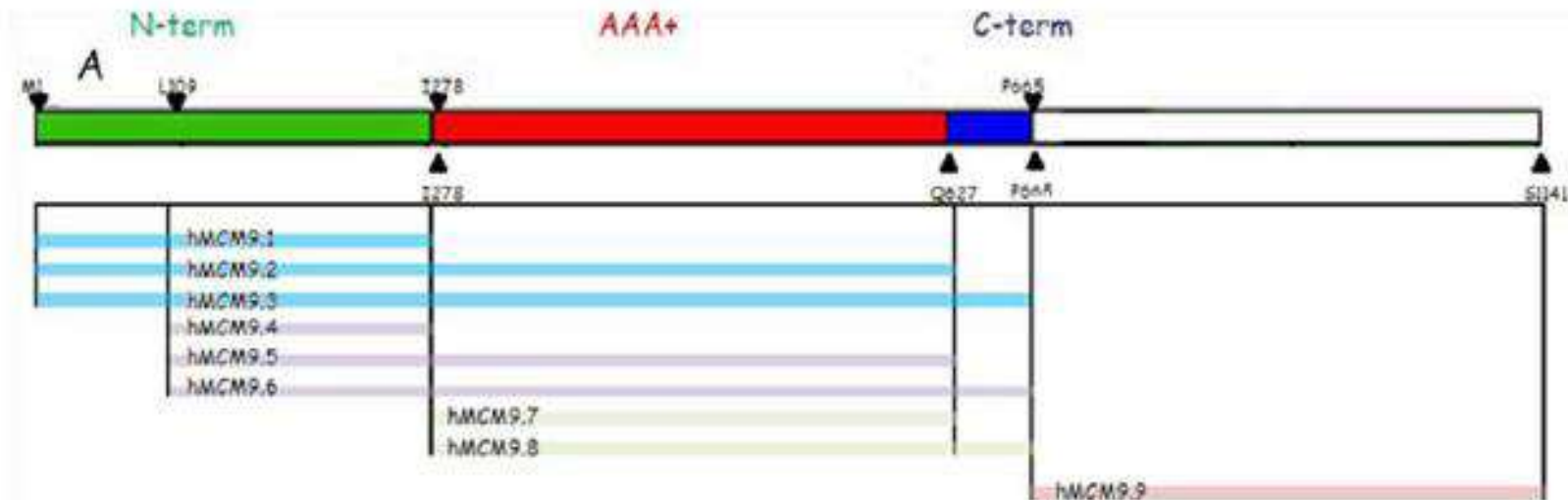
Full-length protein

Empirical design

- Deletion mutants
- Single domains
- Domains combination

Experimental design

- Limited proteolysis
- N-terminal sequencing
- Mass spectroscopy



Constructs design

Removal of glycosylation sites: Mutation of residues involved in post-translational modification enhances homogeneity of the sample.

Thermostabilization: Mutation of residues to improve stability of the protein.

Removal of phosphorylation sites: mutation of the residues involved in the modification to increase homogeneity in the sample.

Genome, cDNA, synthetic gene

- Bacterial, Archaeal, yeast proteins: cloning from the genome.
- Eukariotic protein: cloning from mRNA, cDNA.
- Synthetic gene (codon code optimization and DNA sequence manipulation, signal peptide insertion)

Cloning and expression

The DNA corresponding to a gene is amplified by **polymerase chain reaction** (PCR) or can have synthetic origin.

The fragment is inserted into a small circular DNA molecule called **plasmid** or **vector**

The vector is used to insert the gene into a **host cell** (transformation or transfection)

The vector is designed in such a way that the **expression** of the gene of interest can be controlled

The expression system can be prokaryotic (*E. coli*) or eukaryotic (insect or mammalian cells)

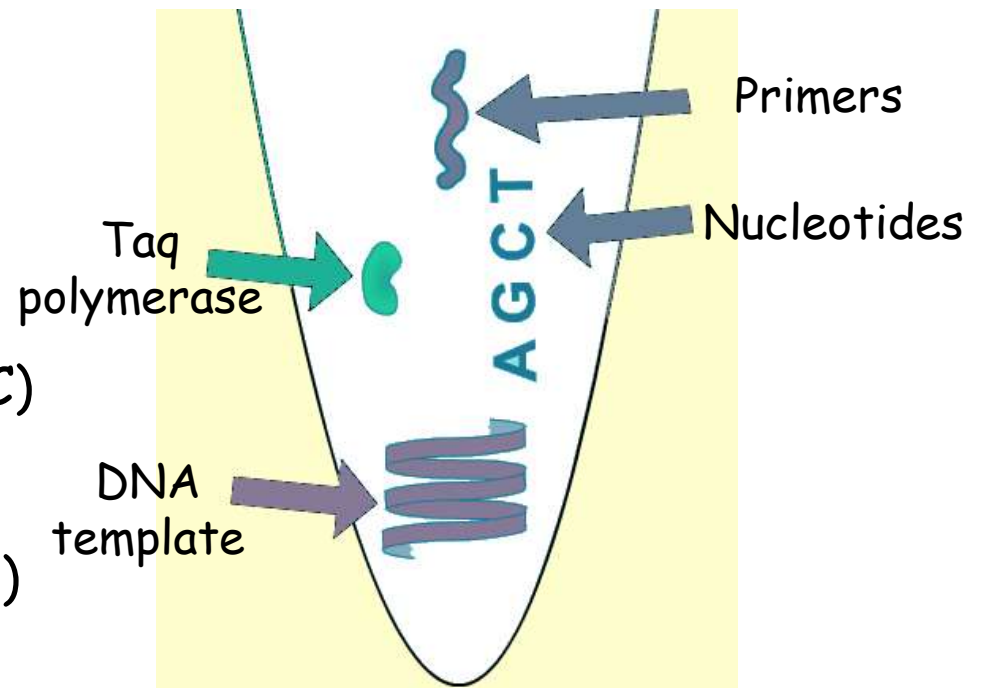
PCR

(recipe and steps)

PCR is an easy method for amplifying a fragment of DNA (**gene of interest, GoI**), once the sequences of the beginning and the end of the **fragment** are known.

A PCR reaction comprises three steps repeated n cycles ($n=20-30$ cycles):

- Denaturation to separate double stranded DNA template (95°C)
- Annealing of the **primers** to the complementary regions on DNA ($45-65^{\circ}\text{C}$)
- Elongation of the primers by a thermostable DNA polymerase ($68-72^{\circ}\text{C}$)

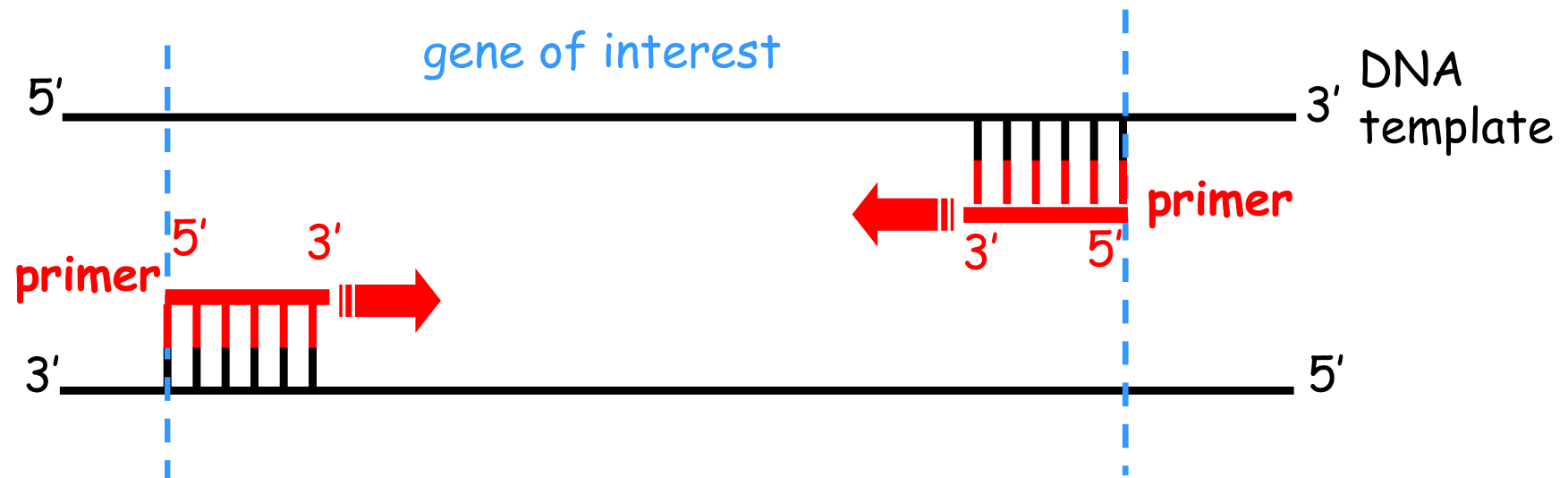


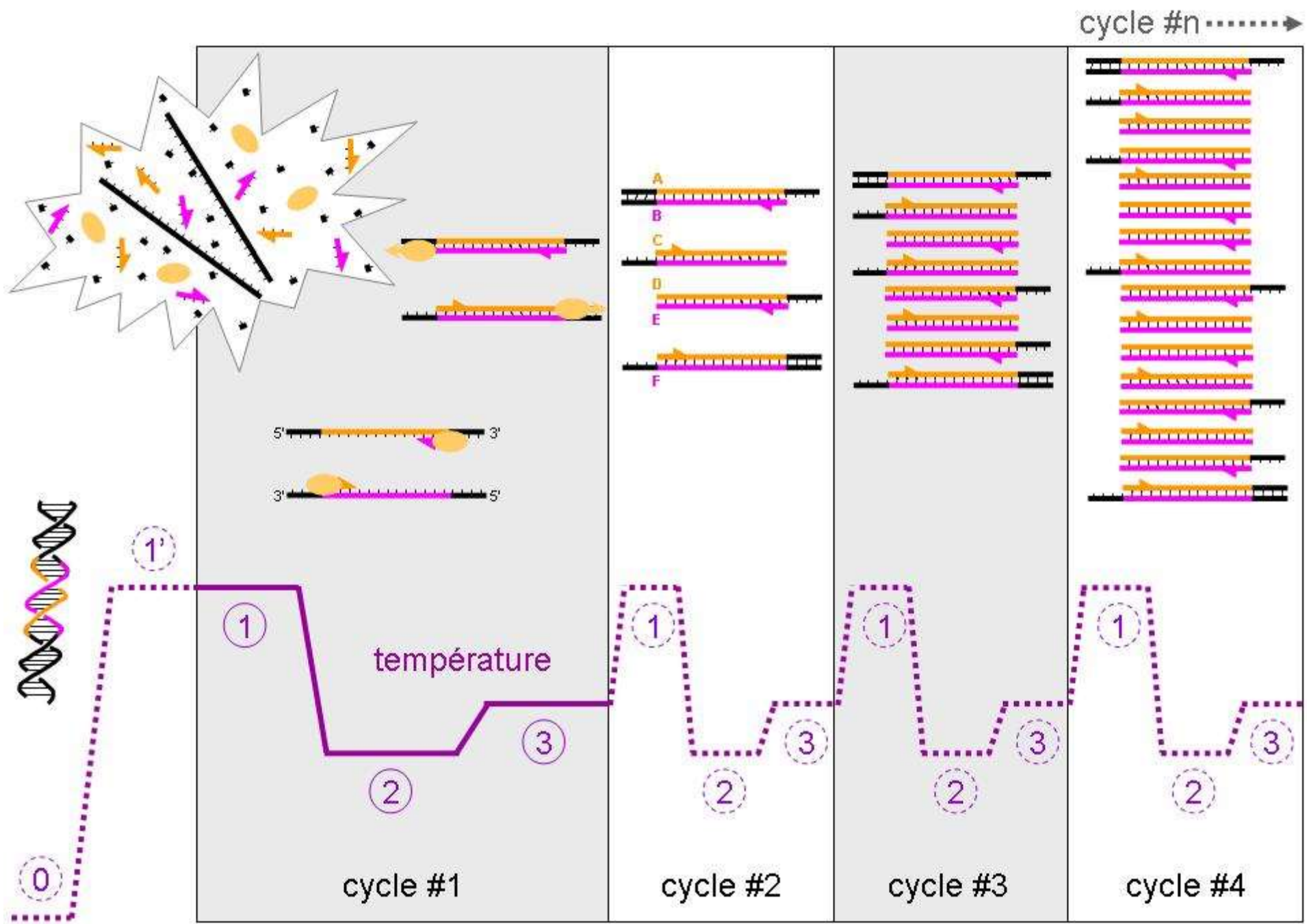
The PCR is a powerful technique that amplifies DNA by a factor of $10^{20}-10^{30}$ and therefore can be used as a diagnostic tool to detect up to a single DNA molecule in a sample!

PCR (reaction)

Primers:

- oligonucleotides range: **21-50 bases**;
- melting temperature (T_m). above **60°C**;
- the GC between **40 and 60%**.

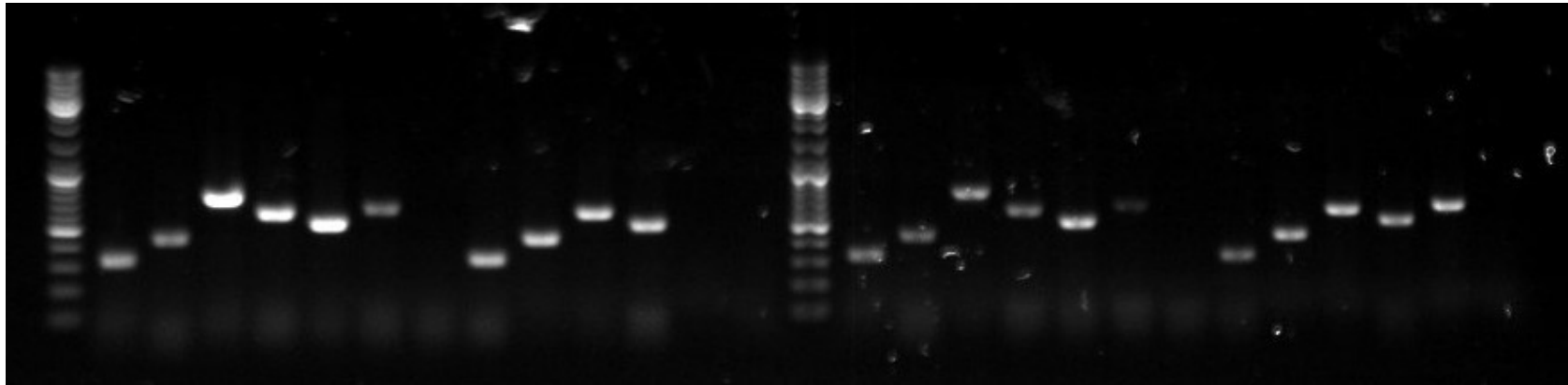




PCR (results)

1 % DMSO

3 % DMSO



1 % Agarose gel
of a Touch down PCR

95°C 5'

95°C 30"
65°C 45"
72°C 1'/1kb
Repeat 5 cycles

95°C 30"
60°C 45"
72°C 1'/1kb
Repeat 5 cycles

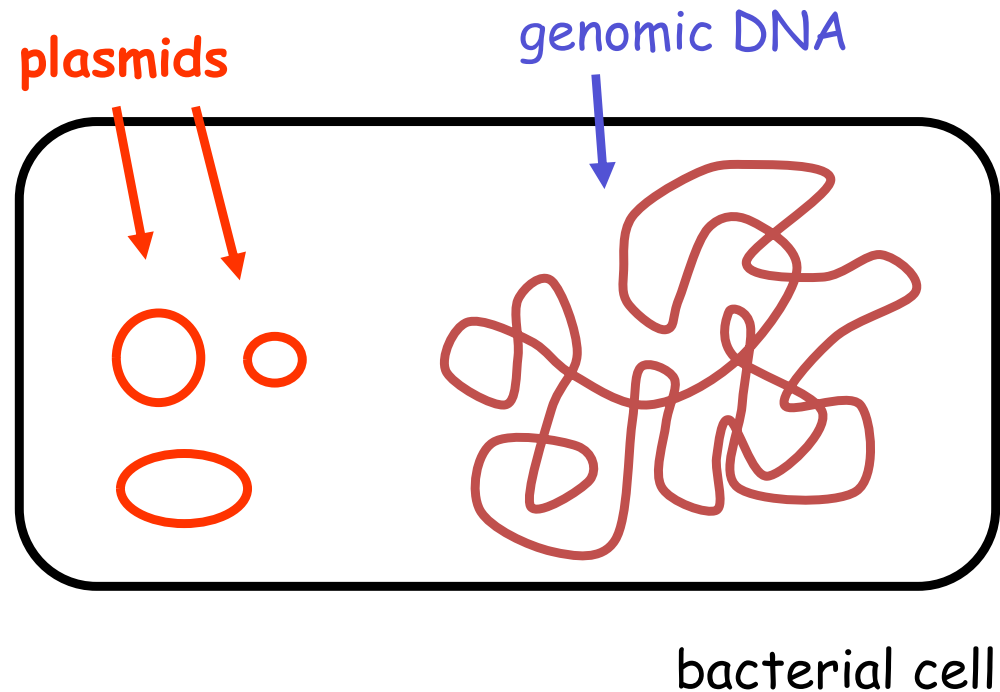
95°C 30"
55°C 45"
72°C 1'/1kb
Repeat 5 cycles

95°C 30"
50°C 45"
72°C 1'/1kb
Repeat 15 cycles

Plasmids and vectors

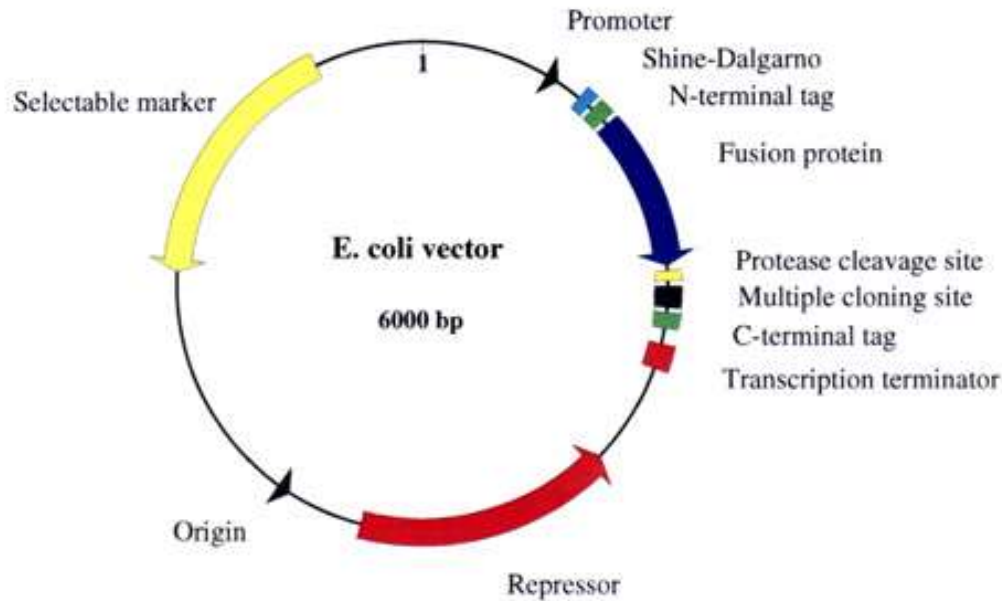
Plasmid = small circular molecule of double stranded DNA, which can be replicated and is independent from genomic DNA.

- small (5-10 kb)
- able to replicate
- confers some advantage to the cell (i.e. antibiotic resistance)



Vector = plasmid artificially modified to carry a gene

Expression vector

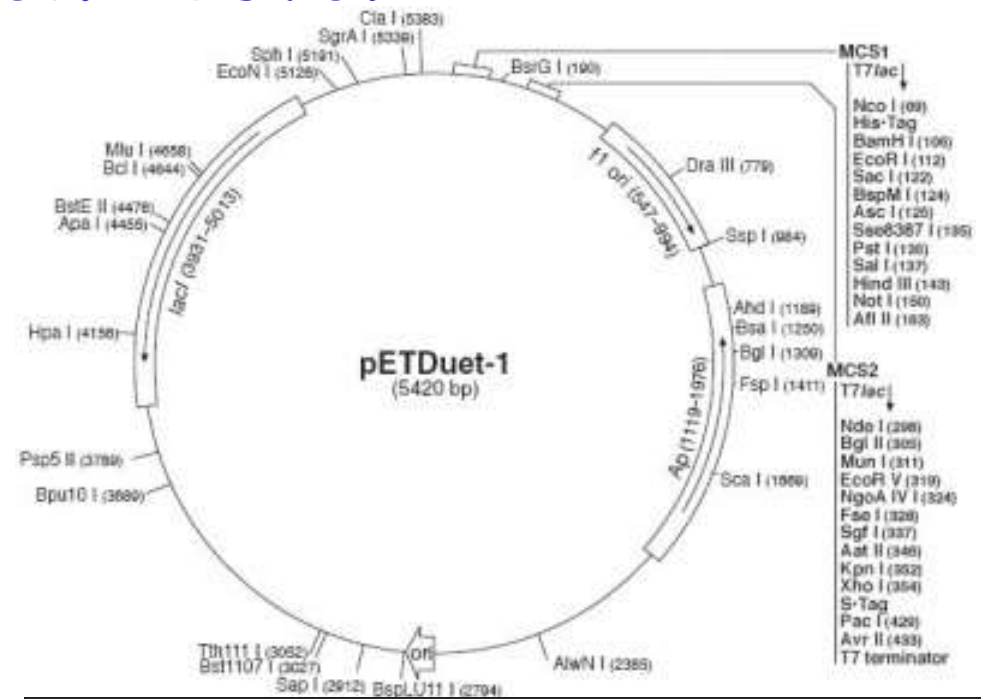


- **Origin of replication** (controls the plasmid copy number),
- **Selectable marker**
- **Multiple cloning site** (or polylinker).
- **Strong inducible promoter**
- **Tags and fusion proteins** [short peptides (tags) or to other proteins (fusion partners)]

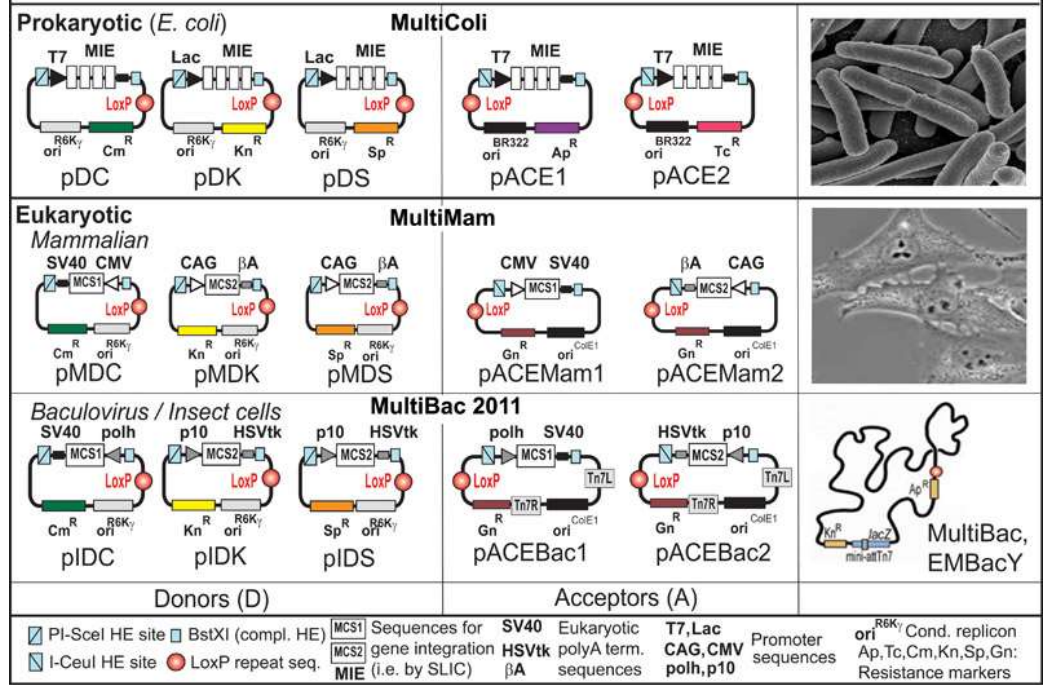
- The construct is inserted into *E. coli* cells (transformation)
- Positive and/or negative colonies selection

Expression vector

- Plasmid (3-10 kbp)
- Bacmid (up to 300 kbp)
- Bicistronic/Multi-cistronic
- Multi hosts



ACEMBL Systems for Multiprotein Complex Expression



Positive or negative selection

Positive selection: Cells that gain a specific gene survive

In positive selection, only cells that contain a specific gene survive.

Antibiotic selection

The inserted DNA includes an antibiotic resistance gene. After transformation, cells are grown on media with the corresponding antibiotic. Only the cells with our vector will survive.

Auxotrophy selection (Auxotrophy: inability to synthesize compounds needed to grow).

Auxotrophic strain and a plasmid that complements the deficiency (i.e., synthesizes the critical amino acid). When plated on media lacking the essential compound, the cells will only grow if they acquired the transformed plasmid.

Negative selection: Cells that have lost a specific gene survive

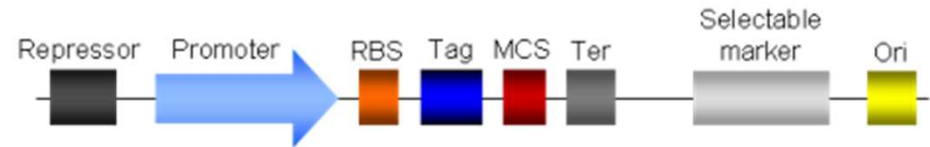
In negative selection, selecting for the loss of a toxic gene product

SacB selection

The SacB gene, comes *Bacillus subtilis* and encodes the that converts sucrose into a toxic metabolite. Plating on (5%) sucrose medium will select for cells that have lost the *sacB* gene. Exchange of the *sacB* gene for the gene of interest allows the bacteria to grow on sucrose.

Tags and fusion proteins

- Improved expression.
- Improved solubility.
- Improved detection.
- Improved purification (affinity)
- Protease cleavage site



Vector for N-terminal fusion



Vector for C-terminal fusion

Vectors for N & C-terminal fusion of affinity tag

Common protease used:

Thrombin: cleaves the sequence: LVPR/GS,

TEV (Tobacco etch virus): cleaves the sequence: ENLYFQ/G,

3c/prescission (rhinovirus protease) cleaves the sequence: LEVLFQ/GP,

Sumo protease: recognise the ternary sumo structure and cleaves after GG/*

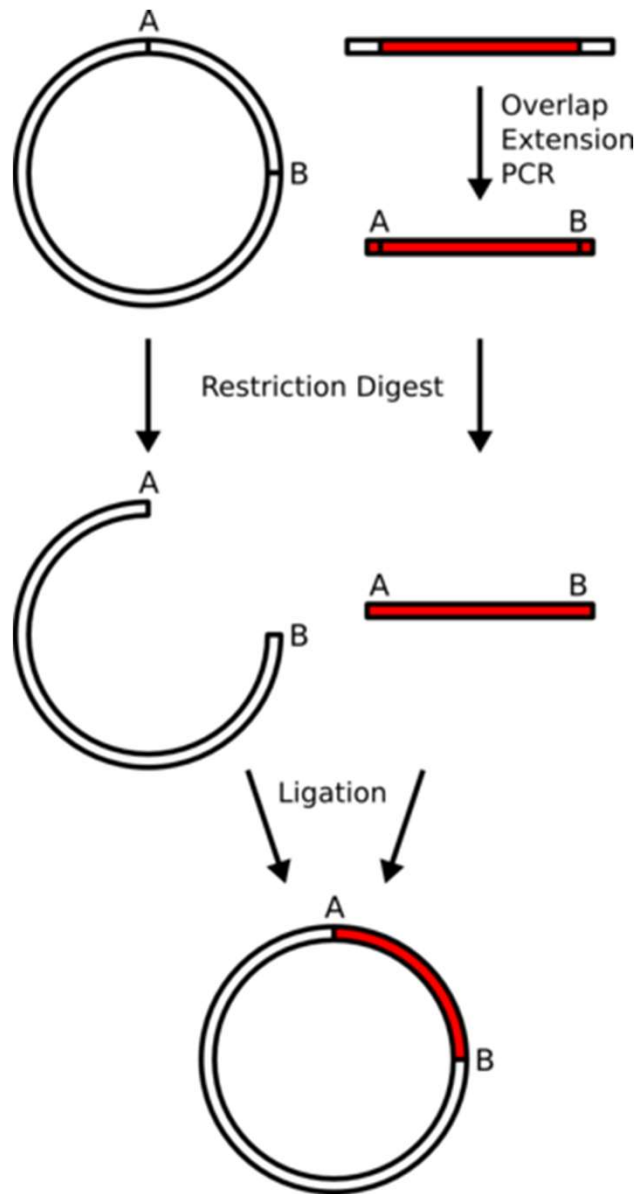
Tags and fusion proteins

Tag	Fusion partner	size [kDa]
N/C-terminal His ₆ -tag		
N/C-terminal His ₆ -tag	Flag-tag (Flag-tag)	12
N-terminal His ₆ -tag	ZZ-tag (ZZ-tag)	10.2
N-terminal His ₆ -tag	thioredoxin (TrxA)	14.3
N-terminal His ₆ -tag	Glutathione-S-transferase (GST)	28.9
N-terminal His ₆ -tag	Maltose binding protein (MBP)	42.2
N-terminal His ₆ -tag	N-utilization substance protein A (NusA)	57.3
N-terminal His ₆ -tag	small ubiquitin-like modifier (Sumo)	13
N/C-terminal His ₆ -tag	Green Fluorescent Protein (GFP)	27

Cloning Methods

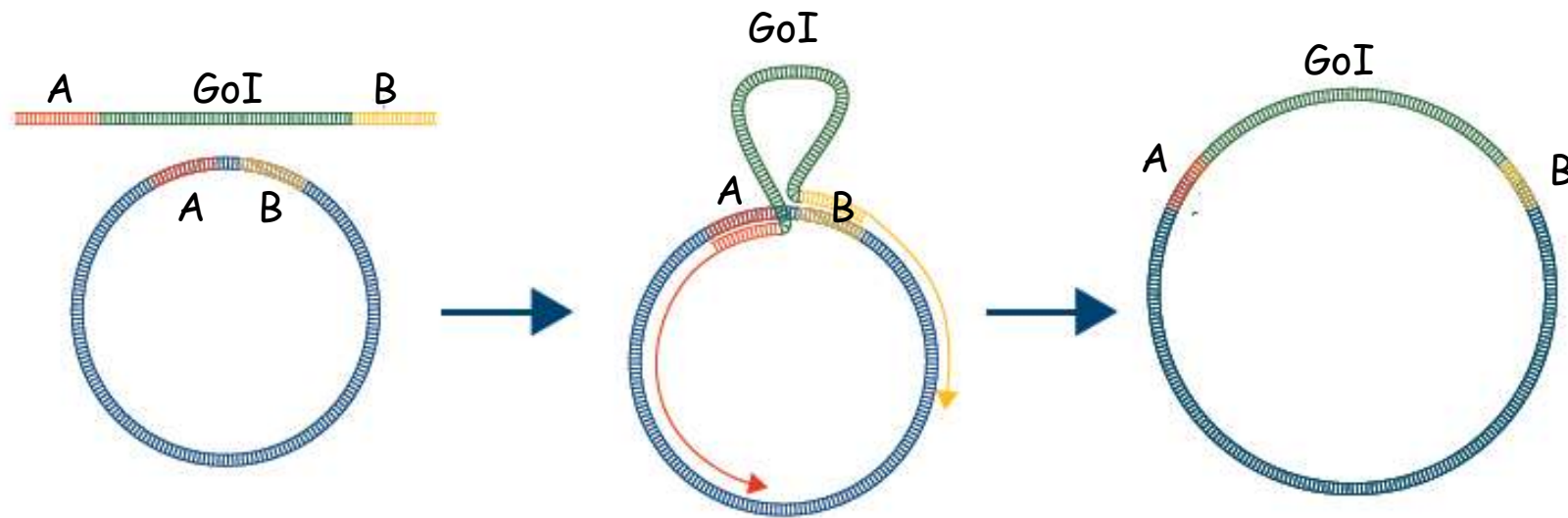
- Cloning using restriction enzymes
- TA cloning and TOPO TA cloning
- Restriction Free cloning (RF)
- The Polymerase Incomplete Primer Extension (PIPE)
- GATEWAY Cloning Technology (Invitrogen)
- Ligation independent cloning (LIC)
- BioBrick

Restriction enzyme cloning



- Primer (21-25 bases): restriction site (6-8 b) + GoI (15-18 b)
- Plasmid and Insert treatment with same restriction enzymes (A and B)
- DNA Ligase joins the ends and allows the circularization.
- The restriction enzyme must be unique in the polylinker and not present in the GoI.

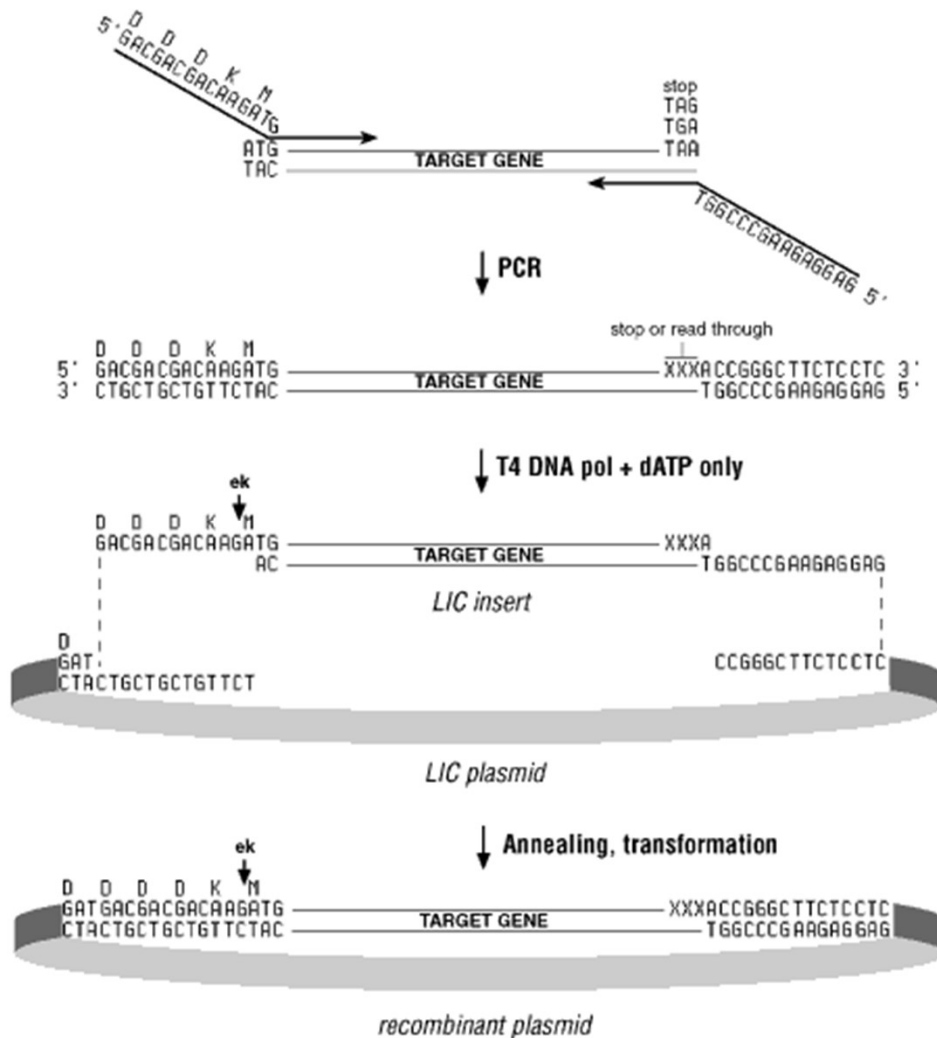
Restriction Free cloning



- Primer (45-50 bases): vector(21-25 b) + GoI (21-25 b)
- Second PCR with Megaprimer which is the GoI, (linear PCR)
- Dpn I treatment (digestion of parental plasmid).

Transfert PCR (TPCR) allows cloning in a single PCR reaction tuning the Primers concentration

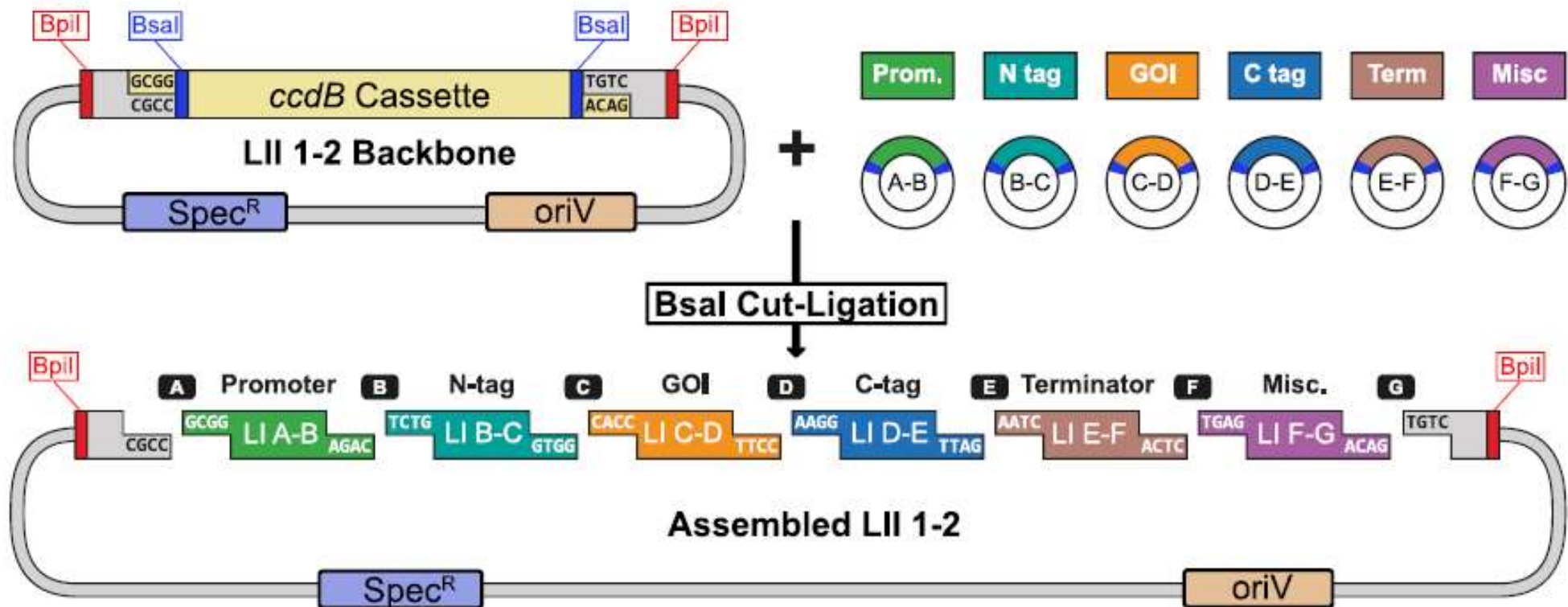
Ligation independent cloning



- Primer (35-40 bases): vector (12-15 b) + GoI (21-25 b)
- Plasmid linearization with restriction enzyme
- T4 Dna Polymerase treatment
- Annealing
- Suitable for high-throughput system

BioBrick cloning

- Complete custom vector

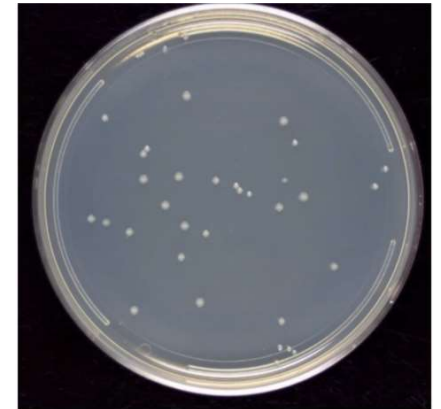


Finally the construct(s)

Transformation in DH5 α , Top10, XL1blue, XL10gold, DH10B

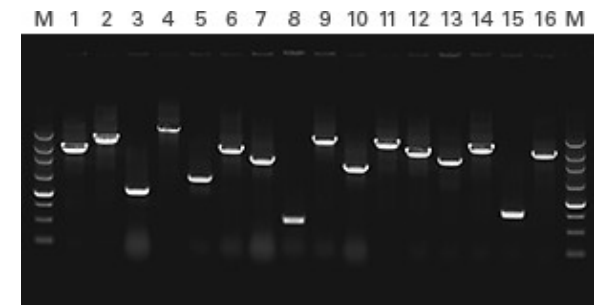
E.coli cells with mutations that avoid:

- ✓ Plasmid recombination with the genome (Δ RecA1)
- ✓ Plasmid degradation (Δ EndA1)
- ✓ Mutation in Lac operon e galactose metabolism (lacZ Δ M15)

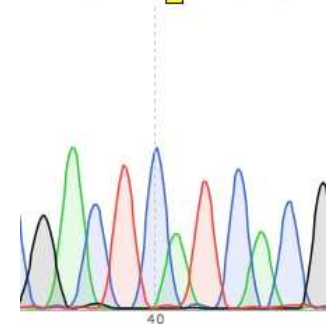


Selection of the positive clones by:

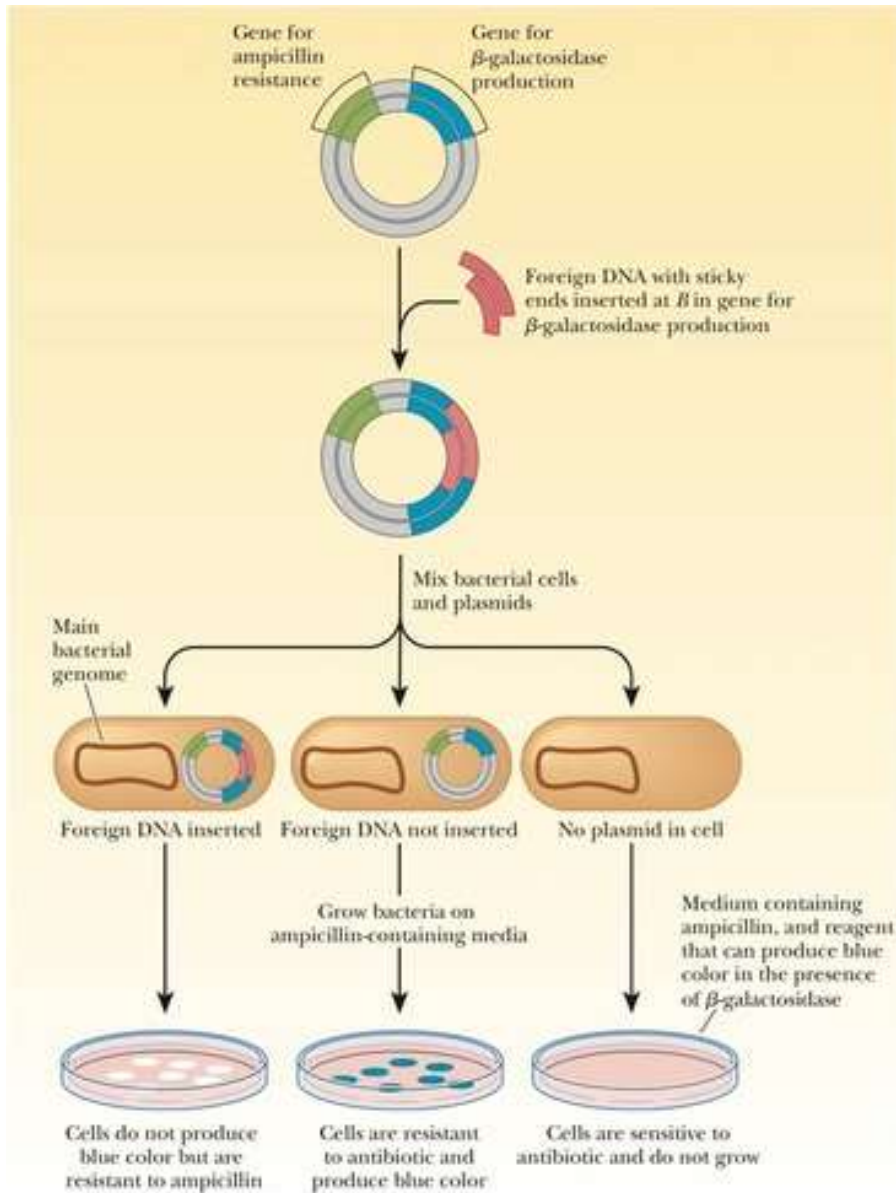
- ✓ Colony PCR
- ✓ Dna extraction and sequencing
- ✓ Diagnostic digestion



G A C T C N T C A C G



Finally the construct(s): Blue-white screening



LacZ: encode for β -galactosidase.

β -galactosidase convert X-Gal in a blue product.

E. Coli grown in presence of X-Gal and antibiotic

Selection of positive colonies by colour:

White +

Blue -



Expression system

Cell-based systems

Bacteria systems

Escherichia coli

Corynebacterium

Pseudomonas fluorescens

Yeast systems

Saccharomyces cerevisiae

Pichia pastoris

Insect cell systems

Baculovirus transfected **insect cells** (*Sf1* e *Sf9* *Spodoptera frugiperda*)

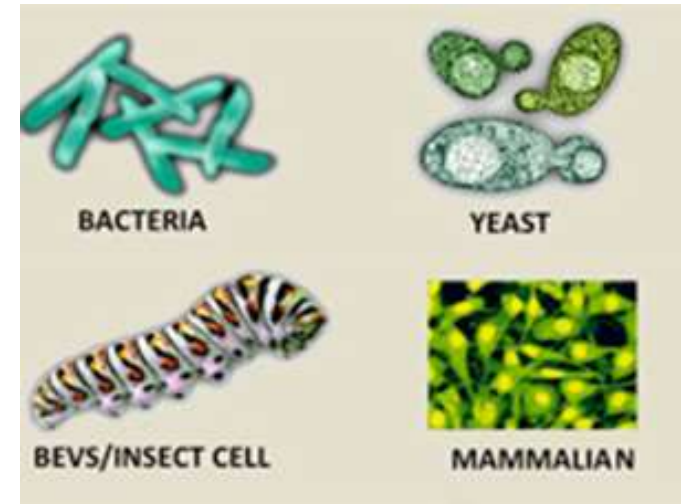
Eukaryotic systems

Virus transfected **Mammalians** (stable or transient, *Hek 293*, *CHO*)

Plants

Cell-free systems

Commercially available (various extract e.g.: *E.coli*, Human)



Expression systems: pros and cons

Bacteria systems: Cheap and quick growing (48 hs max)
Antibiotic selection
Good expression level
No post-translational modification

Yeast systems: Cheap and quick growing
Autotrophic selection
Post-translational modification

Insect cell &

Eukaryotic systems: Post-translational modification
Time-consuming (4-7 days)
No selection
Expensive

Cell-free systems: Post-translational modification
Fast
Expensive

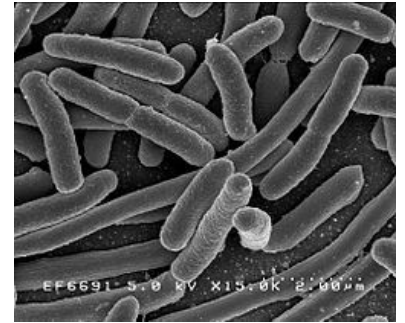
Bacterial expression system: E.coli

Gram-negative bacteria, engineered to not be pathogenic.

Modified genome to reduce protein proteolytic events.

Selection, quick growth, inducible expression.

Protein expression in cytosol, periplasm, inclusion bodies.



BL21

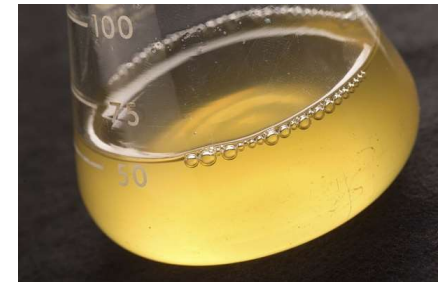
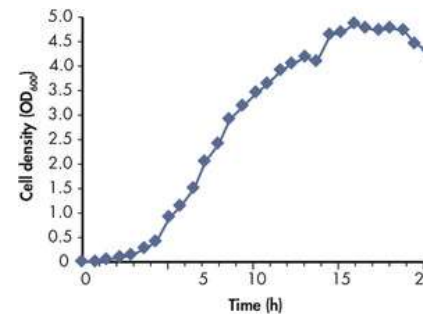
Rosetta/BL21 C+: codon rare

Origami: sulphite bridges

C41 and Lemo21: membrane protein

B834: methionine auxotroph

Artic: low temperature expression



DE3: T7 pol under Lactose control, control of the protein expression.

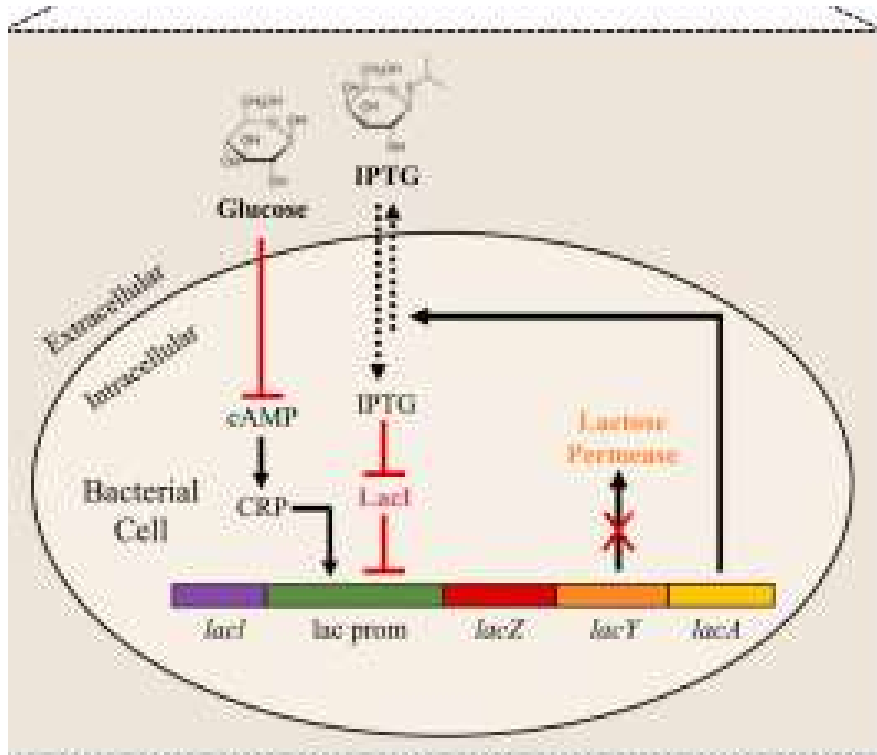
pLys: lysozyme for the T7 pol inhibition, recommended for toxic proteins.

Co-expression with chaperones.

Lac operon

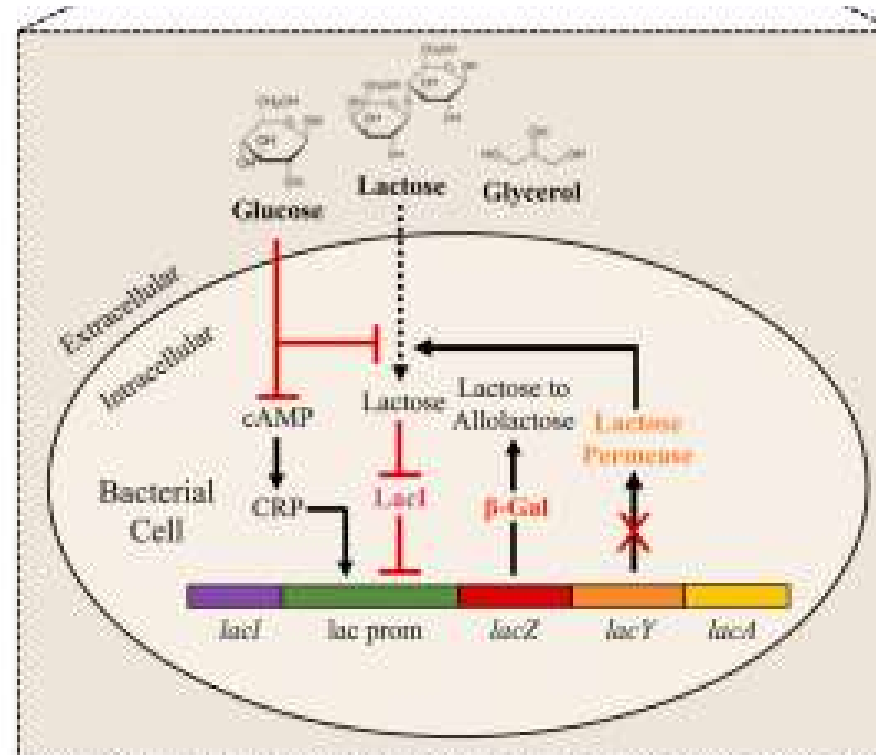
- LacI protein (repressor) bound to the operator impairs the mRNA synthesis.
- Glucose inhibits cAMP production and CRP activation

IPTG induction



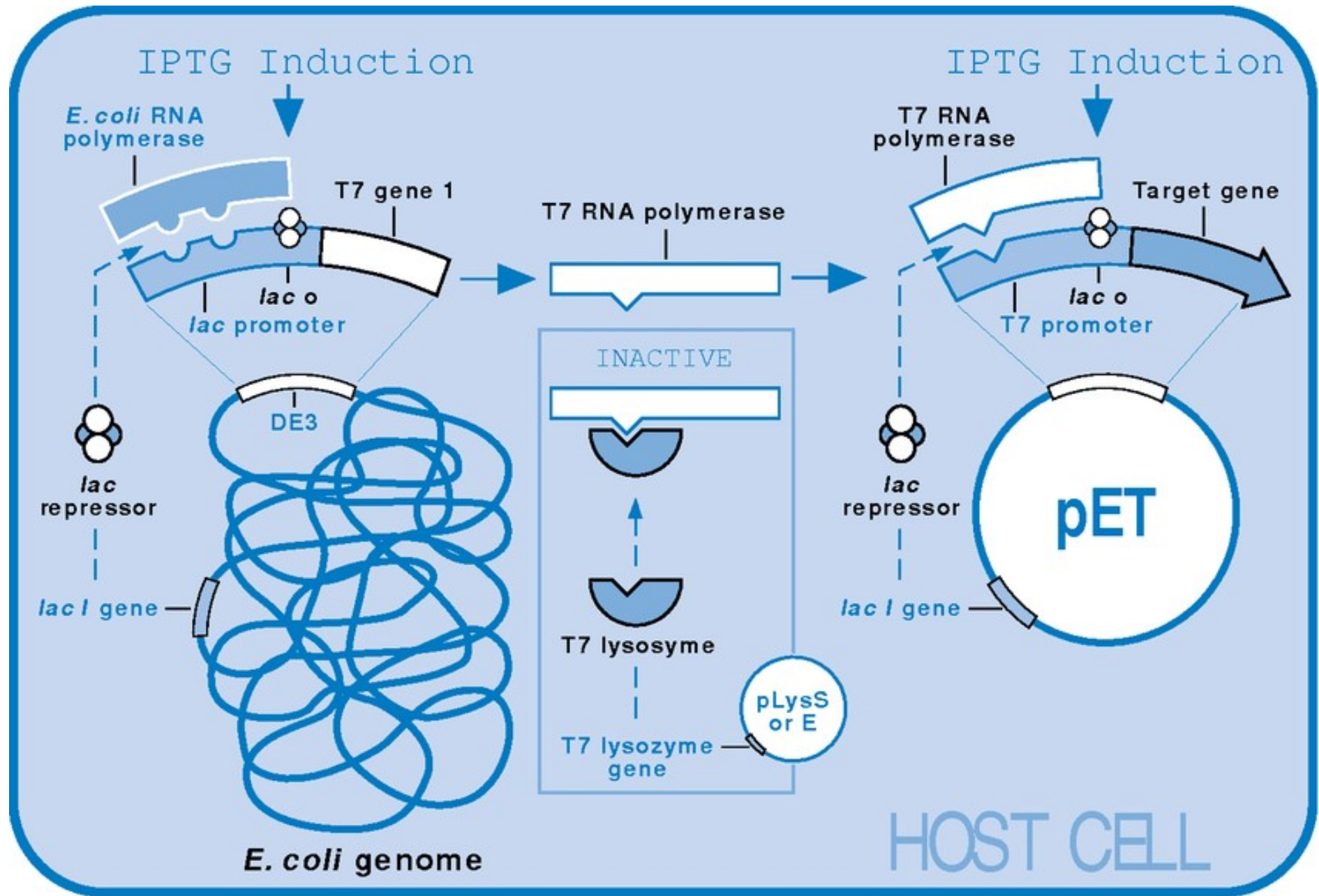
IPTG binds the repressor, operon released and RNA Pol II can start the mRNA production.

Autoinduction



Based on media that induces protein expression in *E. coli* when cells reach saturation. Auto-induction regulated by adjusting glucose/lactose levels in media

E. coli inducible expression



Expression and solubility test

E.Coli strain	Temperature	media
BL 21	37°C	2YT
Ros	37°C	2YT
Ori	37°C	SB
C41	37°C	TB
BL 21	25°C	TB
Ros	25°C	SB
Ori	25°C	TB
C41	25°C	2YT
BL 21	17°C	SB
Ros	17°C	TB
Ori	17°C	2YT
C41	17°C	SB

- **Media** (TB, LB, SB, ZY): carbon source (sugars); aminoacids, salts and trace metals

- **Induction System:** Autoinduction (glucose/lactose ratio), IPTG (lactose analog)

- **Temperature and timing:** from 15-37°C, timing can reach upto 48 hs

Eukaryotic expression system: yeast

Simple Eukarya (*S.cerevisiae*; *P.pastoris*).

Known genome, efficient homologous recombination.

Selection, quick growth, inducible expression.

Protein expression in cytosol, secreted.

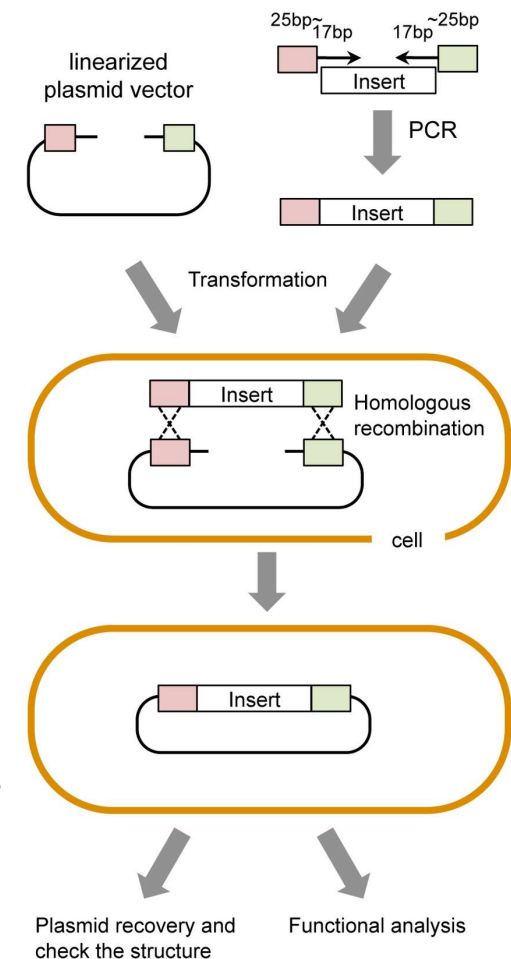
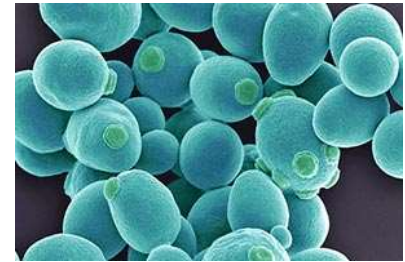
Post-translational modification (glycoprotein).

S.cerevisiae: Baker's yeast (not-Methylotrophic)

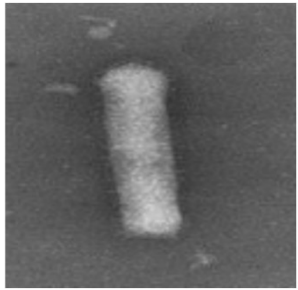
Autotrophy selection, Galactose induction,
easy to culture, transient clones.

P.pastoris: (Methylotrophic)

Methanol induction, stable clones, lipid composition
similar to higher Eukarya.



Eukaryotic expression system: Baculovirus/Insect cells (BEVS)



Baculovirus: AcMNPV (Autographa californica, multiple nuclear polyhedrosis virus), large, circular double-stranded DNA genome (about 130 kb).



Spodoptera frugiperda Sf9 or Sf21 insect cells or Trichoplusia ni Hi-5 insect cells

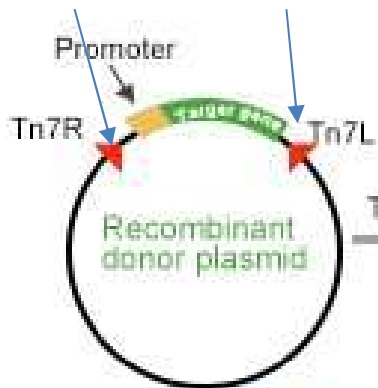
Originally recombinant baculoviruses produced by homologous recombination between the circular viral genome and a transfer vector.

Bacmid (Bacterial artificial vector) is amplified and modified in E.coli, a site-directed transposition occurs between the bacmid and a donor vector, inserting the gene of interest disrupting the lacZ sequence. Recombinant bacmid is identified, based on antibiotic resistance and blue/white selection on agar plates, it is amplified, isolated, and introduced into insect cells for BV production

Bac to Bac system

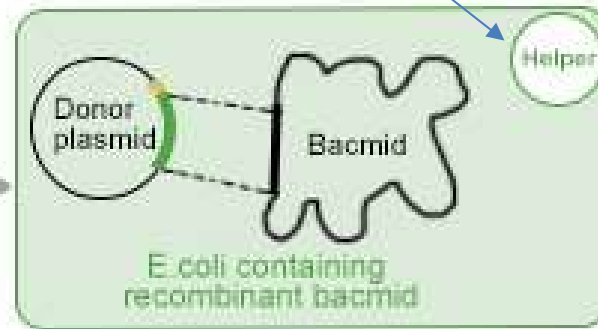
Protein expression in cytosol or secreted, post-translational modification (glycosylation is different from mammalian)

Tn7 transposase flanking region



Transformation

Plasmid encoding for transposase

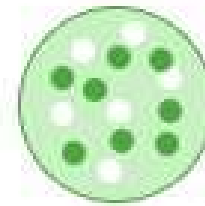


Transformation
Antibiotic selection

Blue/white selection and
PCR colonies screening

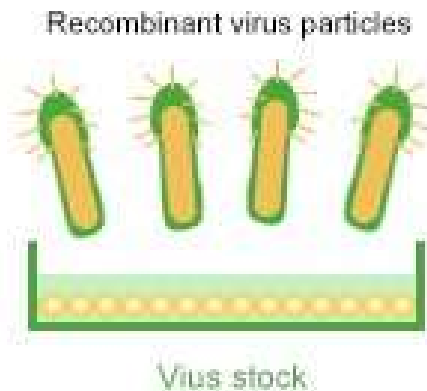


Extraction of
recombinant
bacmid DNA



Recombinant bacmid DNA

Transfection of insect cells
with recombinant bacmid DNA



10^6 cells/mL culture
P0 virus generation
4-7 days

Infection of
insect cells



10^6 cells/mL culture
P1 virus generation or expression
4-7 days

Determine viral titer
via plaque assay



Eukaryotic expression system: Mammalian cells

CHO (Chinese hamster ovary) or HEK 293 (Human embryonic kidney) cell lines.

Stable cell line: GoI integrated (by recombination) in the cell genome in a random/specific site; expression levels strongly dependent on where the transgene integrates. Clonal selection by antibiotics, expression inducible or constitutive and needs to be tested. Usually, recombinant protein secreted in the media. Long and expensive process.

Transient cell line: transfection of plasmid DNA containing GoI, quick and suitable method for structural studies. Transfection to be performed each time and variability in the preparation.

Promoter (deriving from viruses) used for expression depends on the toxicity of the protein.

Transfection in mammalian cells

Biological transfection: (usually for stable cell line)

Inactivated Viruses are used to infect cells: lentivirus or adenovirus **virus-like particles (VLPs)**. Viruses can self propagate and are risky for operators!!!

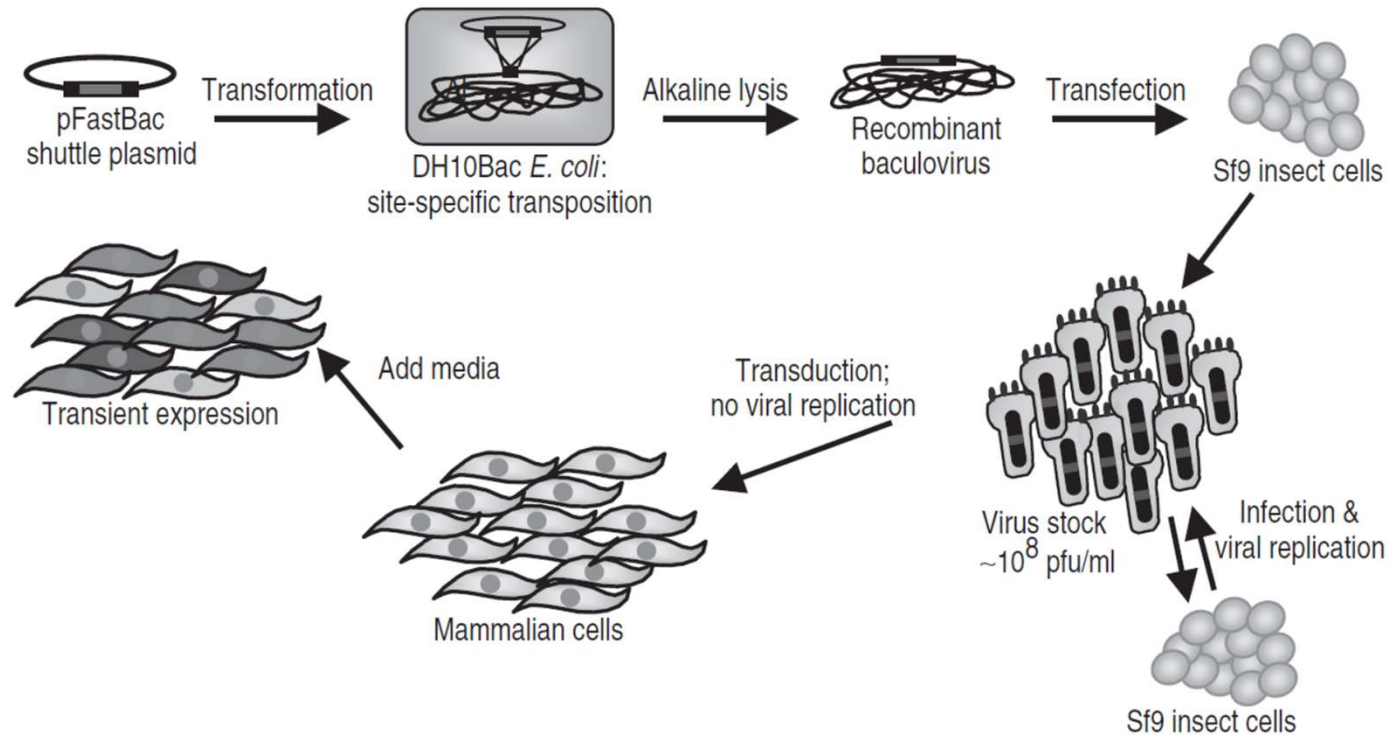
Baculovirus based system: Baculovirus-mediated expression using a mammalian promoter and a reporter gene (considered safe).

Chemical transfection: (transient transfection)

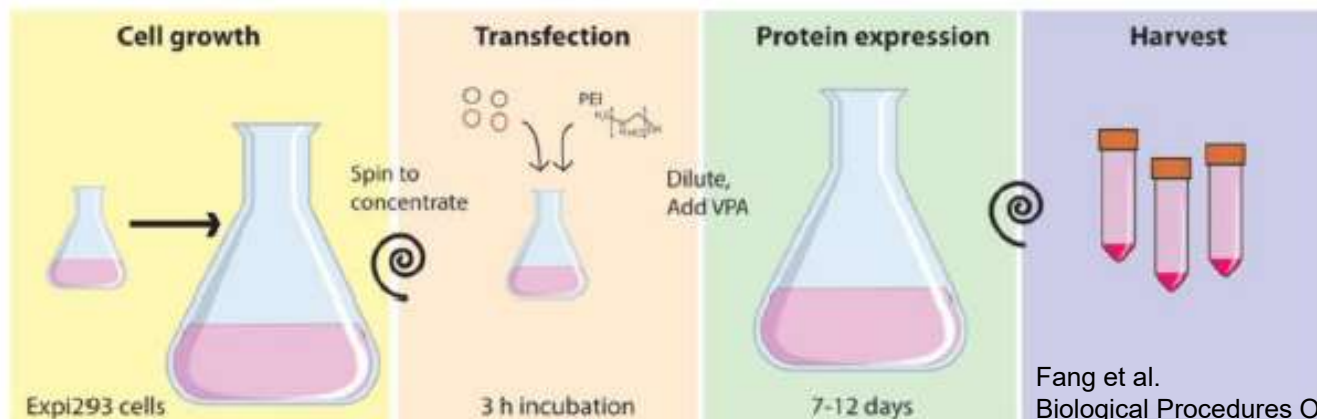
cationic polymers (polyethylenimine PEI), calcium phosphate, or cationic lipid, the formation of complexed with DNA (RNA), entering the cell by either endocytosis or phagocytosis

Reporter gene: gene expressed with the GoI, confers to cells features easy to follow (eg: GFP protein).

BacMam system



Chemical transfection



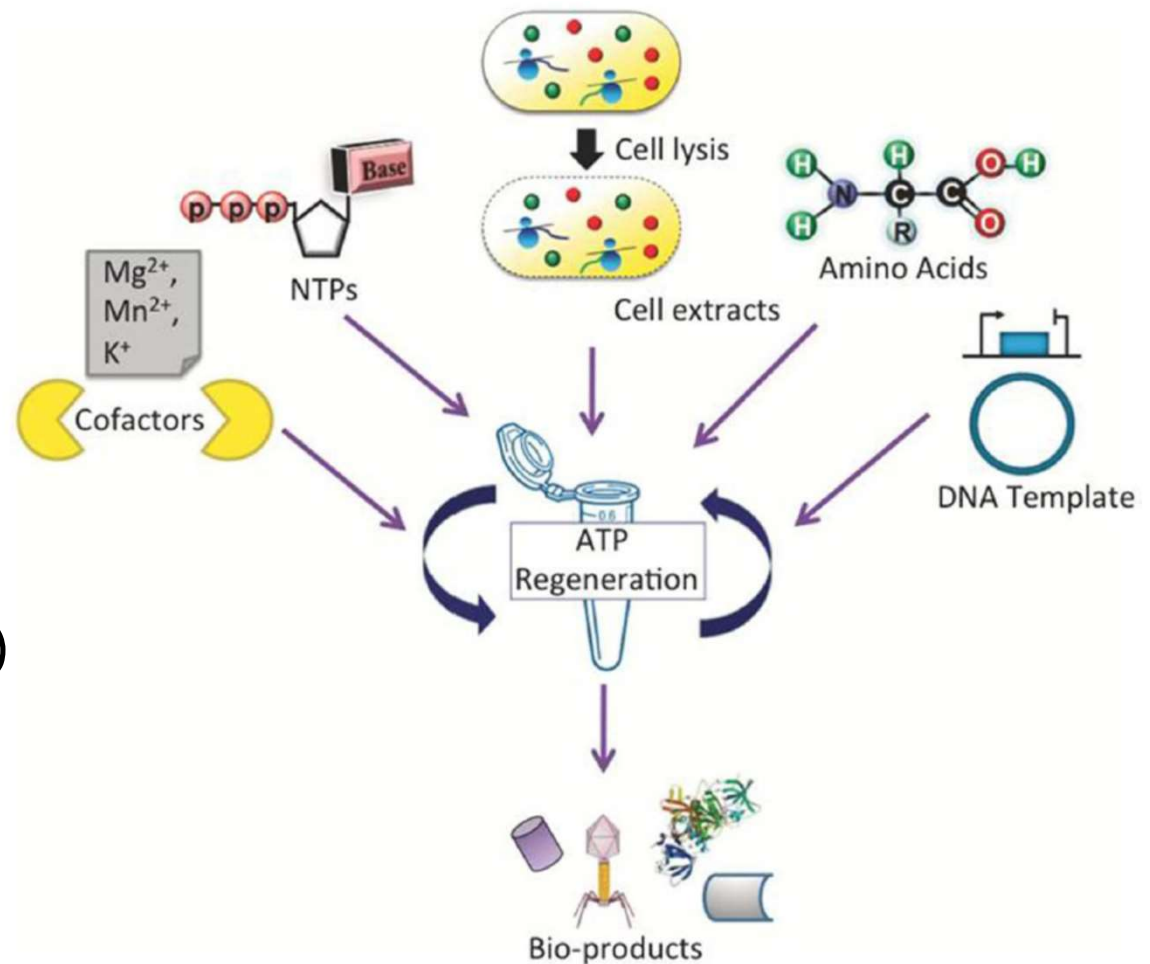
Fang et al.
Biological Procedures Online (2017)

Cell-free expression system

Components from crude cellular lysates of microorganisms, plants, or animals.

Commercially available:

- E. coli
- Rabbit reticulocytes,
- Wheat germ
- Insect cells
- Systems of purified recombinant elements (PURE)



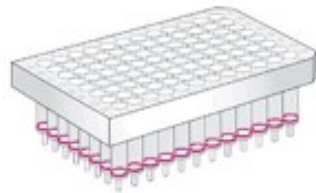
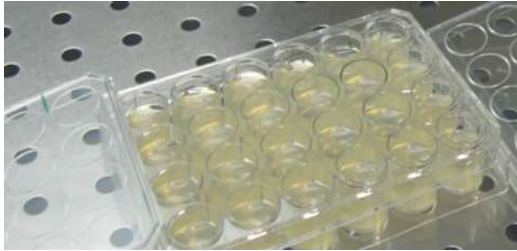
- **E.Coli:** + High yield and easy. Low cost. Can express membrane proteins.
- Limitation of prokaryotic systems. Post-translational folding.

Mammalian system:

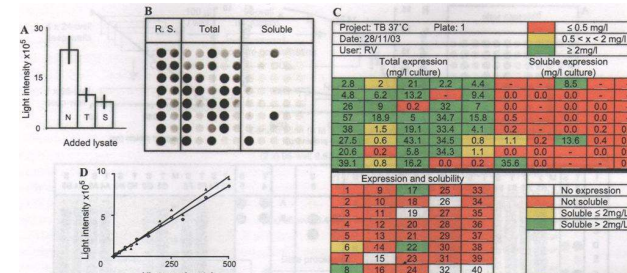
- **Rabbit Reticulocyte:** + Co-translational folding. Can express membrane proteins.
- No protein glycosylation. Low yield. Expensive.
- **Wheat germ:** + High yield. Low cost. Can be stored dried (1.5 yrs). Co-translational folding. Suitable for large proteins.
- No mammalian post-translational modifications. No membrane proteins
- **Insect:** + Suitable for large proteins. Partial glycosylation.
- Not complete mammalian post-translational modifications.
- **Human:** + Optimal co-and post-translational modifications for human proteins. Co-translational folding. Production of completely functional proteins.
- Low yield. Expensive.

High throughput approach

...by Hand



...by Robot



Dot-blot

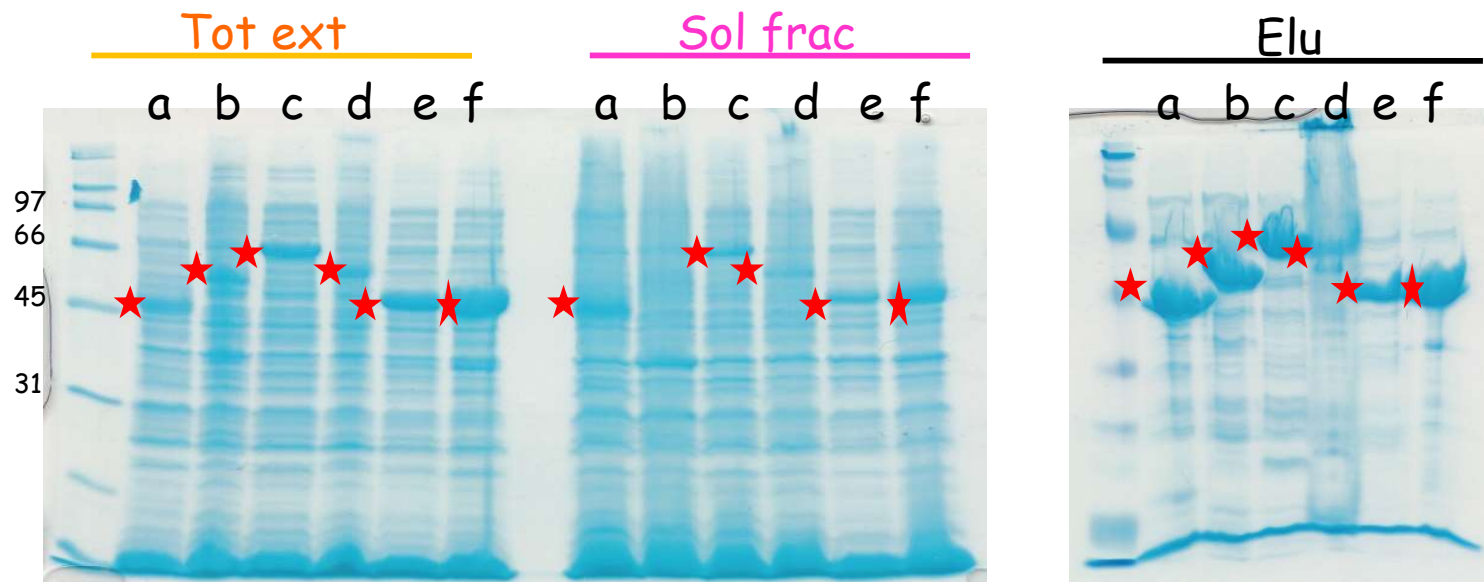


Caliper
"Virtual
Gel"

Small scale test

(a protocol, lucky case)

- 4 ml LB autoinduction media, 4hs 37°C + 17°C 18hs in Rosetta 2 pLysS
- Collect cells and resuspension (buffer:according to 1 purification step)
 - Cell lysis: lysozyme and freeze/thaw cycles (-80°C/37°C)
 - Recovering soluble fraction
 - Binding to affinity resin
 - Wash and Elution



References

Expression systems and vectors: Addgene, "Plasmids 101: A Desktop Resource", October 2015 (2nd Edition), www.addgene.org

Cloning: DNA Cloning and Assembly Methods, Editors: Valla S., Lale R.

Yeast cell: Baghban R et al., Yeast Expression Systems: Overview and Recent Advances. Mol Biotechnol. 2019 May;61(5):365-384. doi: 10.1007/s12033-019-00164-8.

Insect cell: Chambers, A. C., et al Overview of the baculovirus expression system. *Current Protocols in Protein Science*, 91, 5.4.1-5.4.6. (2018). doi: 10.1002/cpps.47; Bac-to-Bac™ Baculovirus Expression System (user guide)

Mammalian cell: Fang, X.T., Sehlin, D., Lannfelt, L. et al. Efficient and inexpensive transient expression of multispecific multivalent antibodies in Expi293 cells. Biol Proced Online 19, 11 (2017). <https://doi.org/10.1186/s12575-017-0060-7>; Boudjelal M et al. The application of BacMam technology in nuclear receptor drug discovery. Biotechnol Annu Rev. 2005;11:101-25. doi: 10.1016/S1387-2656(05)11003-5.

Cell-free: Khambhat K, et al Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. Front. Bioeng. Biotechnol. 7:248 (2019) . doi: 10.3389/fbioe.2019.00248