Quality control for Biology (QC4Bio), 4th-8th April 2022, Institut Pasteur, France (Pasteur-PFBMI)

Protein production: from gene to test tube

Membrane Proteins and Membrane Systems Laboratory

Institute for Integrative Biology of the Cell (I²BC)

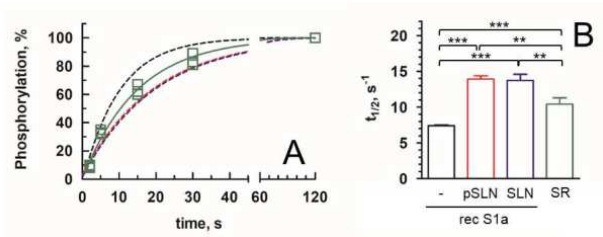
cedric.montigny@i2bc.paris-saclay.fr

Why do you need to isolate a protein?

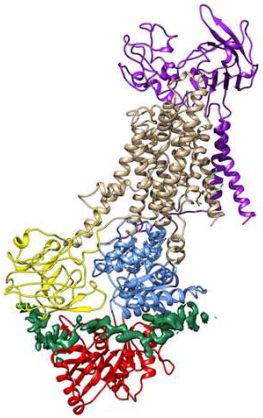
- Low availability from native sources
- Influence of the cellular environment
  - Interaction with other cellular components
  - Low accessibility within the cell
- Need for highly pure compounds (e.g. pharmacology)

Aims

- Biochemical studies, from µg to mg



- Structural studies, a few mg



- Industry, kg
  - Growth hormone (childhood growth defect ; dead body → yeast[1])
  - Insulin (diabetes ; porcine/bovine extract → yeast[2])
  - Factor VIII (coagulation defect, haemophilia A ; blood → yeast[3])
  - Monoclonal antibodies (mammalian cells[4])

1. Biotechnol Appl Biochem. 2009 Nov 13;54(4):197-205. doi: 10.1042/BA20090179.
2. Microb Cell Fact. 2014; 13: 141. DOI: 10.1186/s12934-014-0141-0
3. DOI: 10.1002/9783527669417.ch19
4. J Microbiol Biotechnol. 2021 Mar 28;31(3):349-357. doi: 10.4014/jmb.1911.11066.

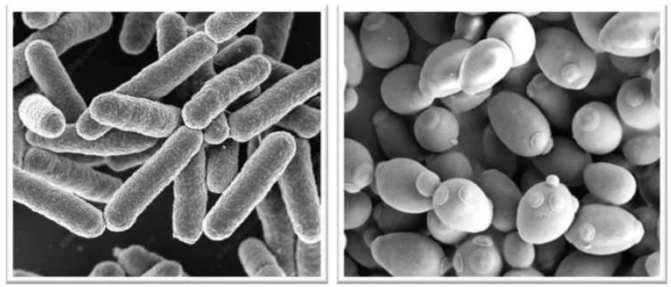
Various approaches for protein production

Native sources

Native context (PTM, addressing and maturation)

Low availability

Contamination (Growth hormone
or contaminated blood scandal in the 1980s)

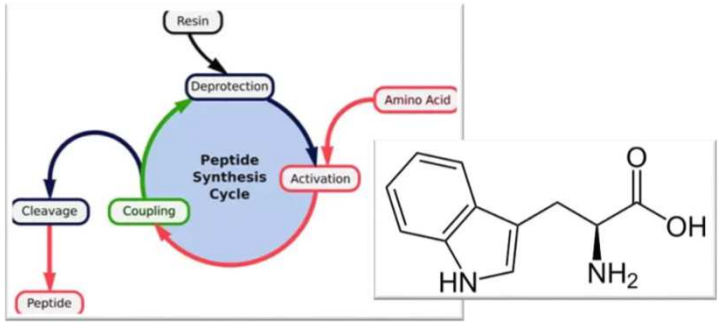


Recombinant protein expression

Different type of cells available (PTM, addressing and maturation)

High level of expression reachable

Purification can be a long-term task

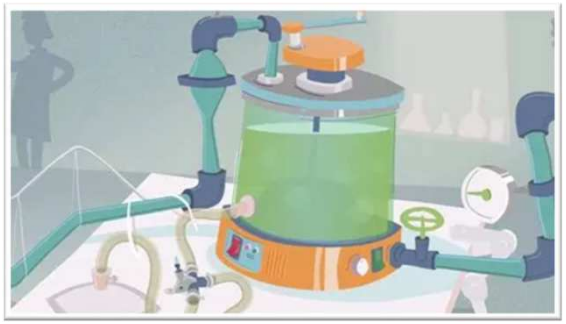


Peptidic synthesis

Mg amount available, highly pure

Addition of non-natural amino-acid, PTM

Limited to about 250 aa
or even only a few dozen for hydrophobic amino acids



Cell-free expression

Addition of non-natural amino-acid

Simplified purification protocol

Low yield of production,
Expensive, no PTM

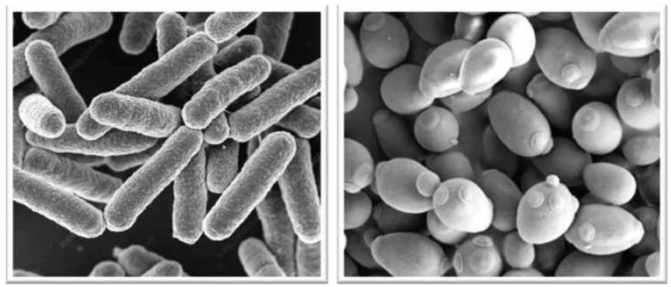
Various approaches for protein production

Native sources

Native context (PTM, addressing and maturation)

Low availability

Contamination (Growth hormone
or contaminated blood scandal in the 1980s)

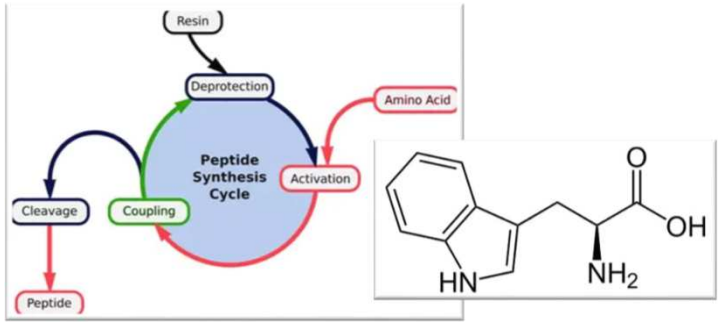


Recombinant protein expression

Different type of cells available (PTM, addressing and maturation)

High level of expression reachable

Purification can be a long-term task

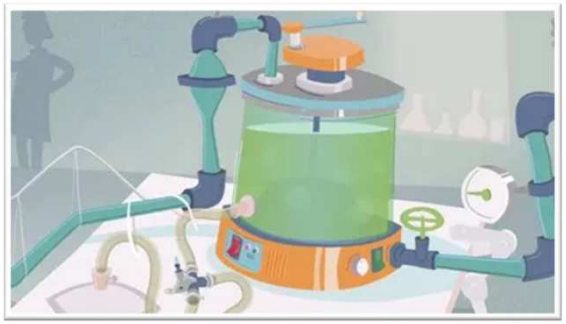


Peptidic synthesis

Mg amount available, highly pure

Addition of non-natural amino-acid, PTM

Limited to about 250 aa
or even only a few dozen for hydrophobic amino acids



Cell-free expression

Addition of non-natural amino-acid

Simplified purification protocol

Low yield of production,
Expensive, no PTM

How to produce recombinant proteins?

1. Cloning

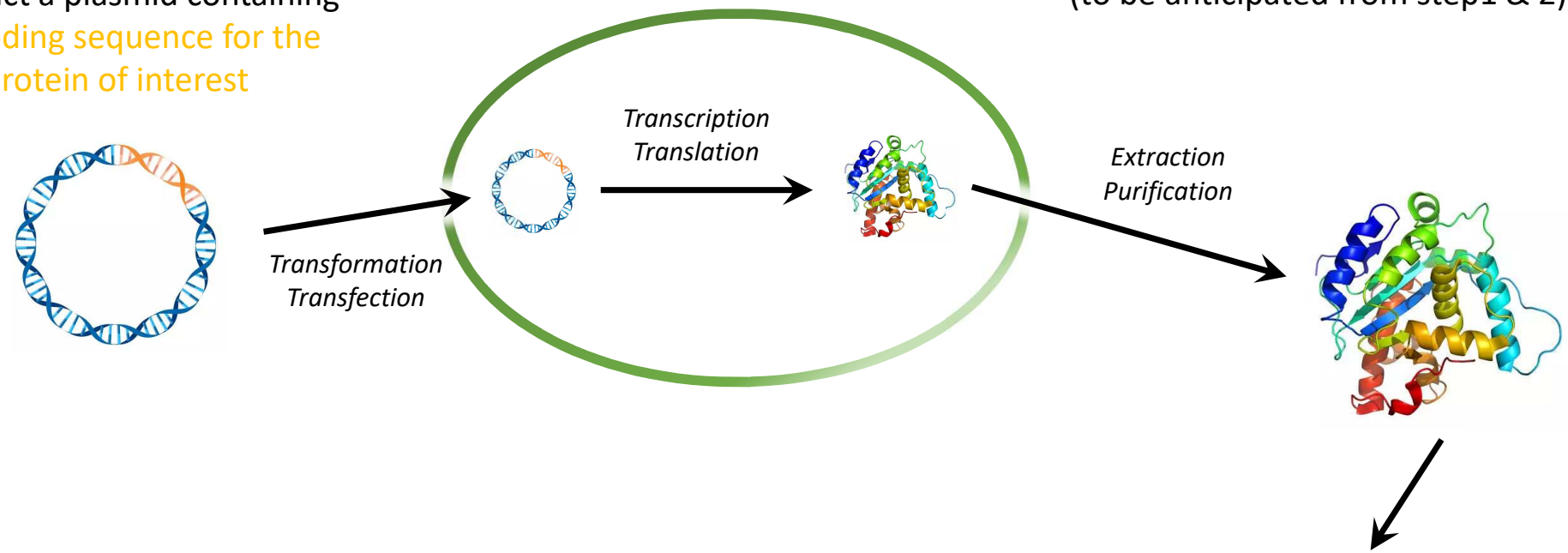
Molecular biology tools to construct a plasmid containing the coding sequence for the protein of interest

2. Cell growth & Expression

Choice of the host cell

3. Purification

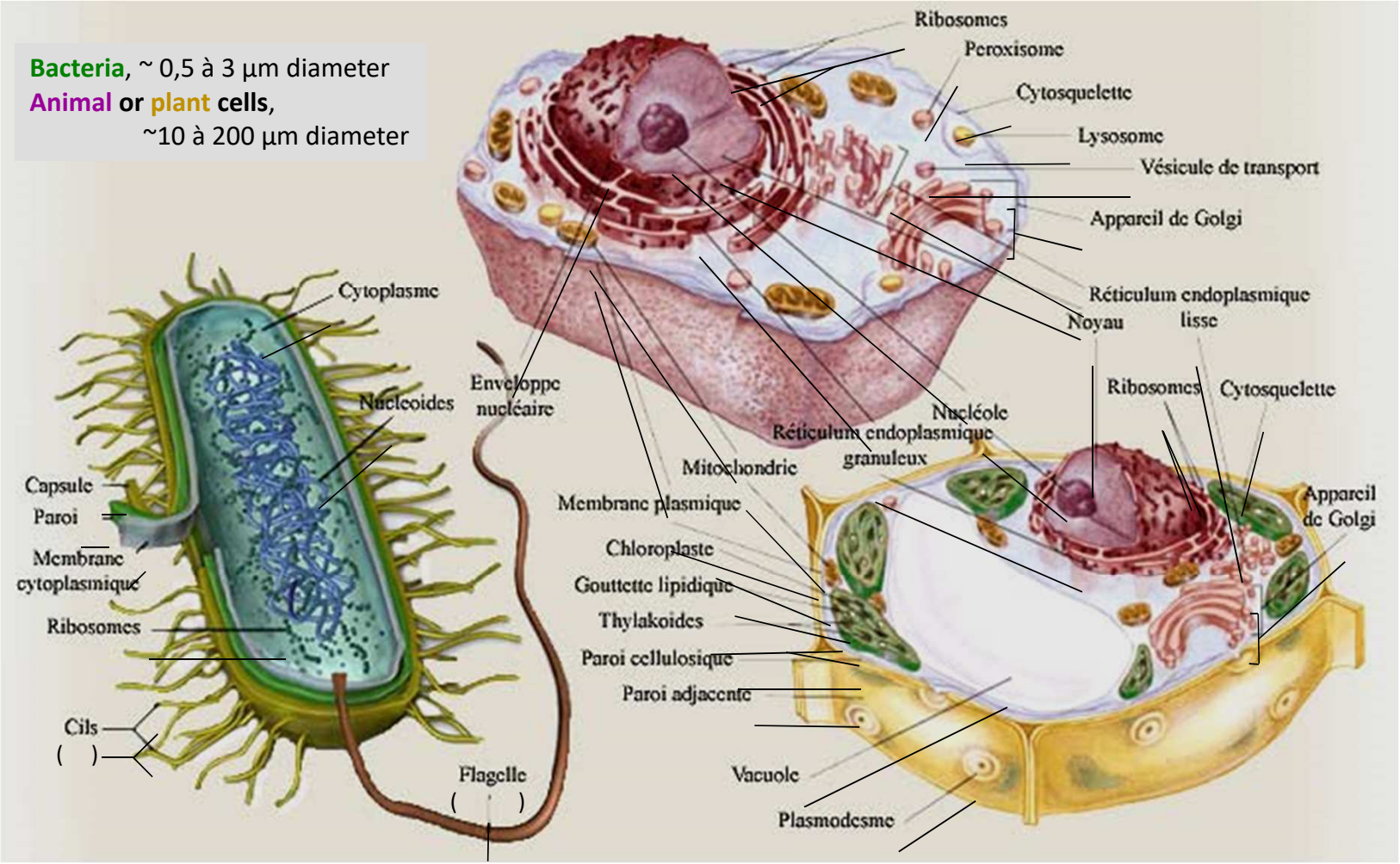
Choice of the method
(to be anticipated from step1 & 2)



4. Functional and structural studies

Cell organization

Bacteria, ~ 0,5 à 3 µm diameter
Animal or plant cells,
~10 à 200 µm diameter



Choice of the host cell

Characteristics of the protein that need to be taken into account

- Origin (Prok/Euk, codon bias)
- Molecular weight
- Fold and possible quaternary structure
- Post-translational modifications (PTM)*
- Secreted or not
- Toxicity
- Membrane protein*

	Expression System	Pros	Cons
E. coli	Plasmid	Inexpensive Simple culture requirements Rapid construct generation	* Problems folding IMPs Low yields * No glycosylation/post-translational modification
Yeast	Plasmid	Inexpensive Simple culture requirements Rapid construct generation Extensive genetics Eukaryotic processing *Pichia* can be grown to a high density Three structures solved (K^+-channel; Ca^{2+}-ATPase; SoPIP2)	* Non-native glycosylation/hyperglycosylation(*S. cerevisiae*) * Non-native lipid environment Poor expression of some proteins
Insect cells	General	* More native lipid environment than yeast Relatively robust culture requirements Good safety profile Good track record of functional expression	Cost * Non-native glycosylation * Non-native lipid environment
	Baculovirus	Well-established protocols Viruses are easy to titrate Multi-cistronic constructs available Structure of spinach aquaporin	Time required to generate virus Cell lysis
	Transient transfection	Rapid construct generation	Cost Amount of DNA required
	Bacmam	Cross-over	Requirement for two different kinds of cells Lysis in insect cells
Mammalian cells	General	* Native lipid environment * Native secretory/post-translational pathways Good track record of functional expression	Cost Technical requirements
	Stable integration	Consistent expression levels No need to generate vector	Time required to establish Instability of integration
	Semliki Forest Virus (SFV)	Success in expression-screening studies Efficient infection	Requirement for helper RNA Technically demanding
	Transient transfection	Speed Structure of recombinant rhodopsin	Amount of DNA required Cost of reagents at large scale
	Bacmam	Lack of lysis in mammalian cells Cross-over	Requirement for two different kinds of cells

J Struct Biol. 2007 160(3):265-74. doi: 10.1016/j.jsb.2007.07.001.

Choice of the host cell

Characteristics of the protein that need to be taken into account

- Origin (Prok/Euk, codon bias)
- Molecular weight
- Fold and possible quaternary structure
- Post-translational modifications (PTM)*
- Secreted or not
- Toxicity, solubility
- Membrane protein*

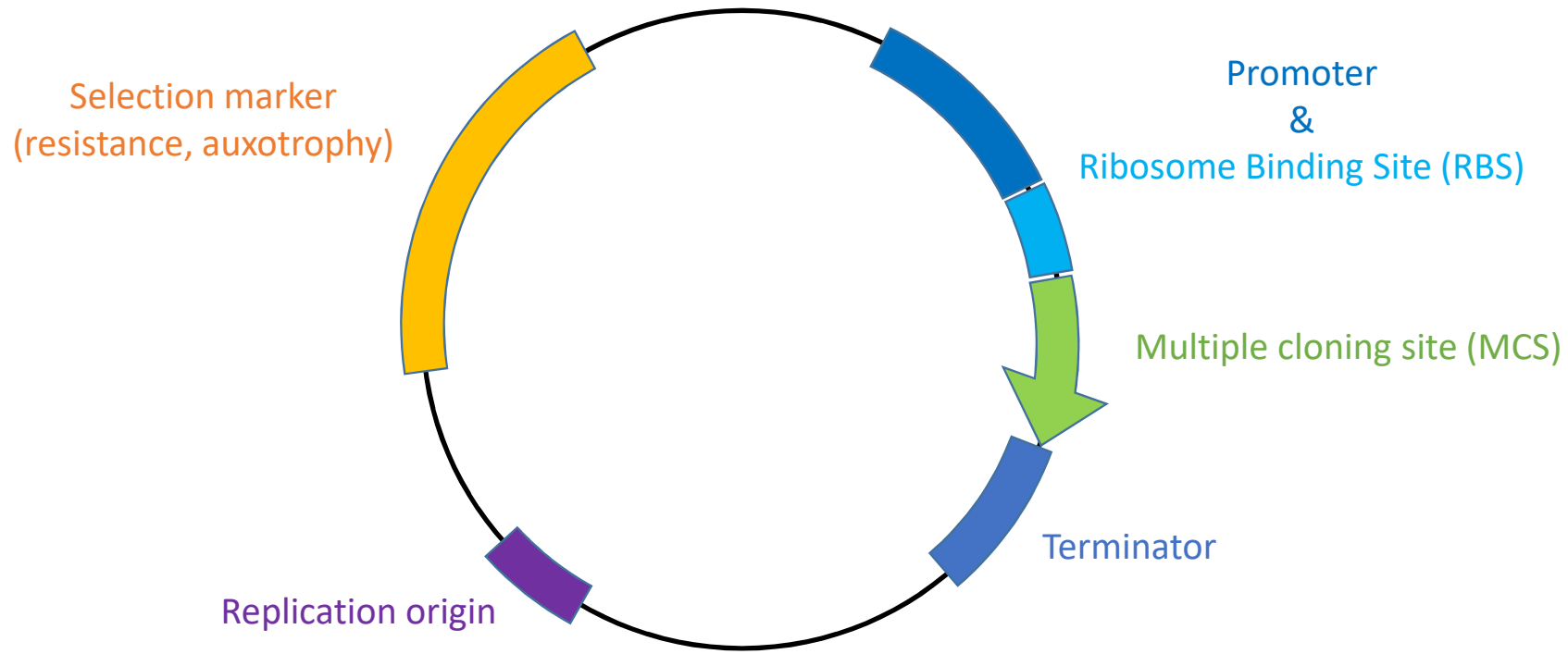
the more complex the system, the more expensive it is and the longer it will take to implement

	Expression System	Pros	Cons
E. coli	Plasmid	Inexpensive Simple culture requirements Rapid construct generation	* Problems folding IMPs Low yields * No glycosylation/post-translational modification
Yeast	Plasmid	Inexpensive Simple culture requirements Rapid construct generation Extensive genetics Eukaryotic processing *Pichia* can be grown to a high density Three structures solved (K^+-channel; Ca^{2+}-ATPase; SoPIP2)	* Non-native glycosylation/hyperglycosylation(*S. cerevisiae*) * Non-native lipid environment Poor expression of some proteins
Insect cells	General	* More native lipid environment than yeast Relatively robust culture requirements Good safety profile Good track record of functional expression	Cost * Non-native glycosylation * Non-native lipid environment
	Baculovirus	Well-established protocols Viruses are easy to titrate Multi-cistronic constructs available Structure of spinach aquaporin	Time required to generate virus Cell lysis
	Transient transfection	Rapid construct generation	Cost Amount of DNA required
	Bacmam	Cross-over	Requirement for two different kinds of cells Lysis in insect cells
Mammalian cells	General	* Native lipid environment * Native secretory/post-translational pathways Good track record of functional expression	Cost Technical requirements
	Stable integration	Consistent expression levels No need to generate vector	Time required to establish Instability of integration
	Semliki Forest Virus (SFV)	Success in expression-screening studies Efficient infection	Requirement for helper RNA Technically demanding
	Transient transfection	Speed Structure of recombinant rhodopsin	Amount of DNA required Cost of reagents at large scale
	Bacmam	Lack of lysis in mammalian cells Cross-over	Requirement for two different kinds of cells

J Struct Biol. 2007 160(3):265-74. doi: 10.1016/j.jsb.2007.07.001.

Expression in *Escherichia coli*

Choice of the plasmid: strategy for expression



Replication origin: cloning basics

- Sequence for the initiation of the DNA replication
- Cloning = maintain the plasmid and the sequence of interest through numerous cell divisions
- Higher is the number of copies, higher should be the level of expression

Replication origin	Plasmid type	Copies/cells
pMB1	pBR322	15-20
pMB1 muté	pUC	500-700
p15A	pACYC	10-12
pSC101	pSC101	~5
ColE1	ColE1	10-15

Two different plasmids can only co-exist in a cell if their origins belong to different compatibility groups

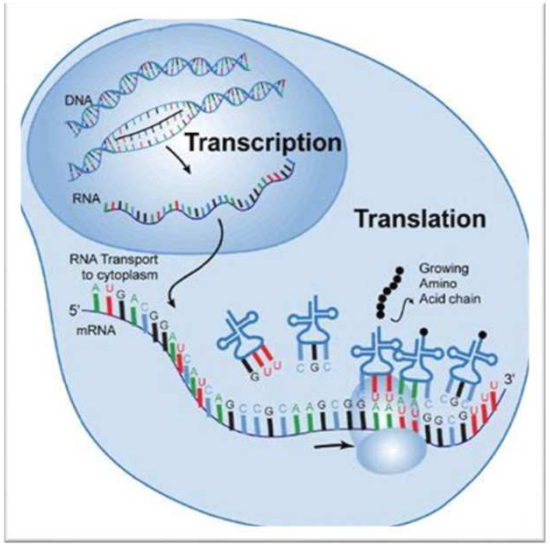
✓ Co-expression strategies for protein complexes expression (or use polycistronic vectors).

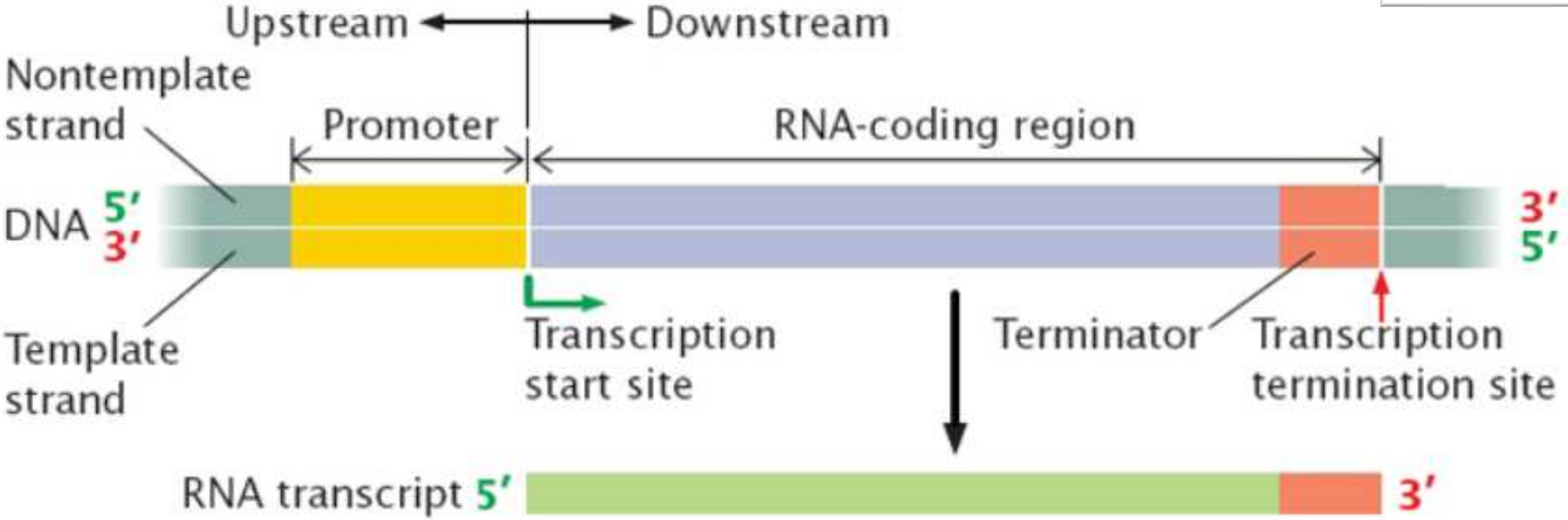
Selection markers confer resistance to antibiotics

- Clone selection: bacteria should keep the plasmid in order to survive

	Target	Location	Resistance
Ampicilline	Cell-wall synthesis	Périplasme	ß- lactamase
Chloramphénicol	Protein synthesis	Cytoplasme	Chloramphenicol acétyl transferase
Kanamycine	Protein synthesis	Cytoplasme	Neomycine phosphotransférase
others			

From gene to protein:
1. Transcription





Promoter: regulation of the transcription

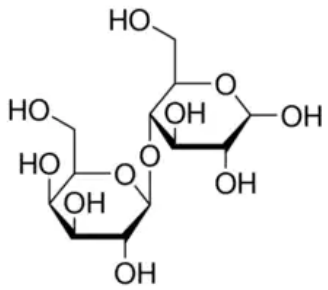
- Site for the binding of the RNA polymerase for mRNA synthesis
- Weak to strong promoters = yield of transcription depending both on the sequence and on the RNA pol.
- Two classes of promoter:

1. **Constitutive promoter** = expression takes place throughout the life of the cell
   - ☹ Most recombinant protein expression strategies involve a preliminary biomass increase phase before inducing expression of the protein of interest (higher yield, limited toxicity)

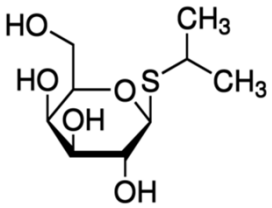
2. Can be tightly associated with regulating sequences = **Inducible promoter**
   - ☹ Check possible leaks
   - ☺ While some systems offer an all-or-nothing approach (e.g pET), others allow you to modulate the level of expression by controlling the amount of inducer (e.g pBAD)

The *lac* operon

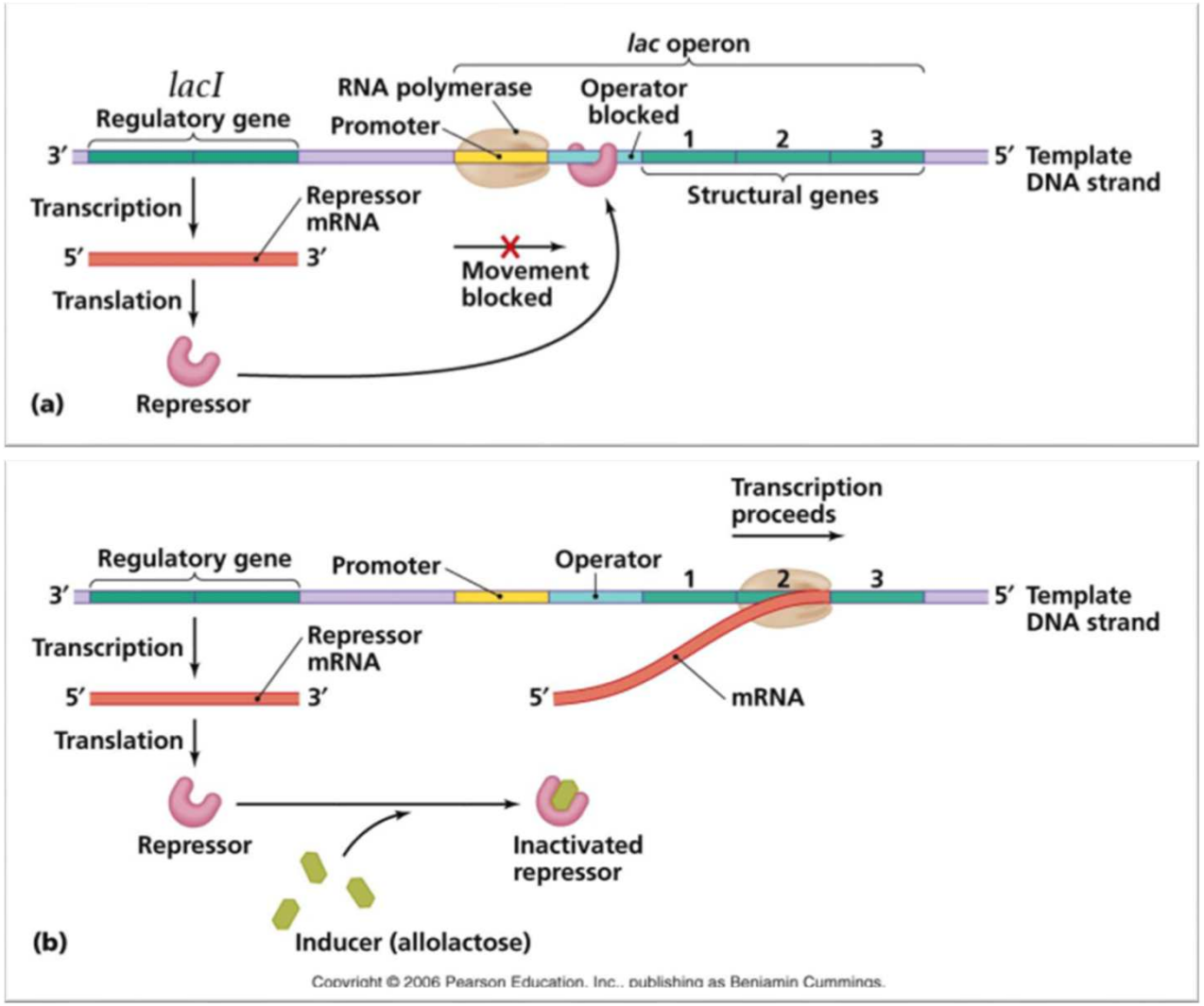
- expression of three key enzymes for lactose metabolism: β-galactosidase, β-galactoside permease and thiogalactoside transacétylase
- In absence of lactose, operon is tightly repressed (a) while in its presence (or analog), expression is relieved (b)



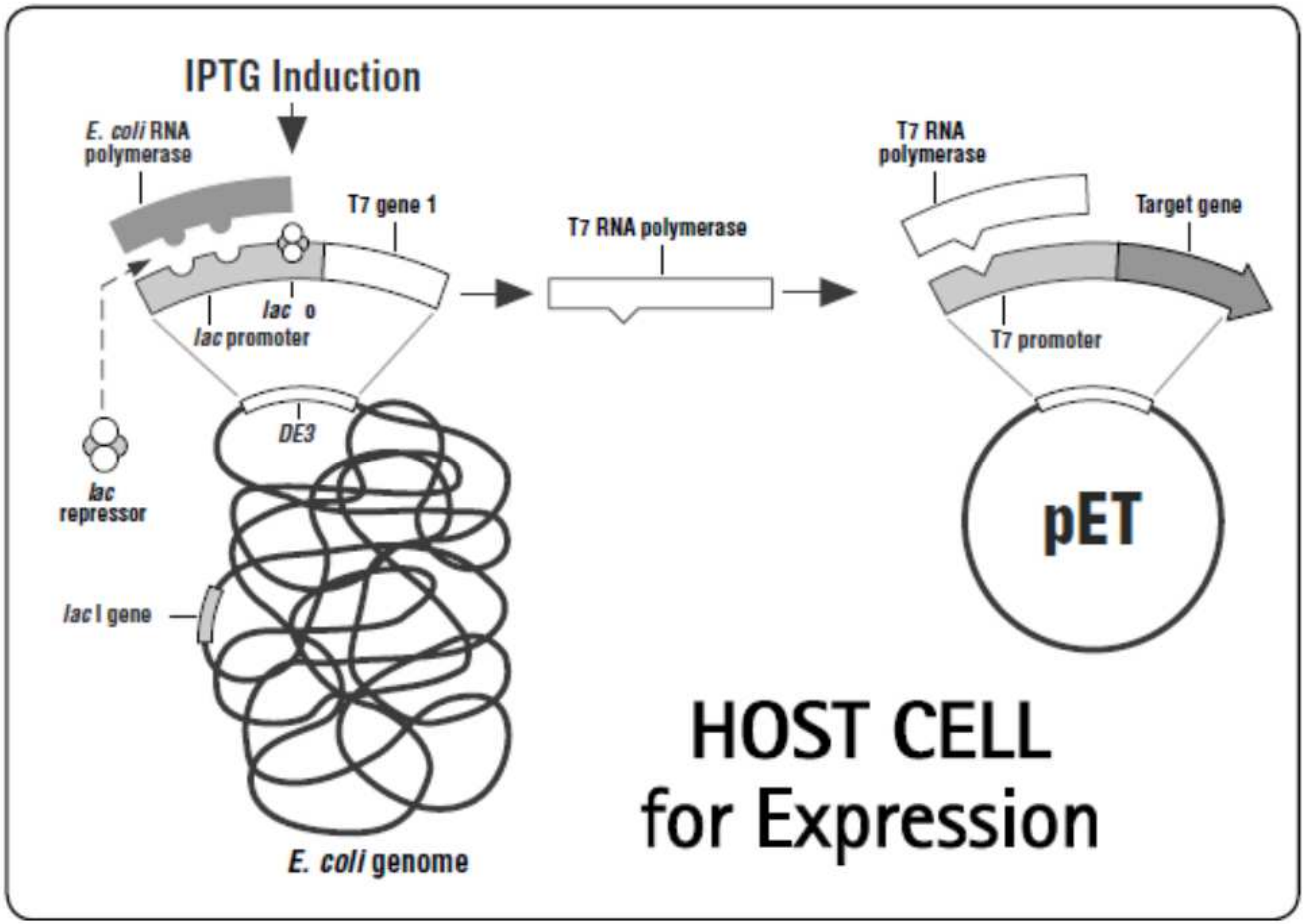
Lactose



Isopropyl β-D-thiogalactoside (IPTG)



The pET system from Novagen



- In vectors from the pET series, your target gene is under the control of the **very strong T7 promoter**
- **T7 RNA polymerase** is integrated in the bacteria genome, under the control of **the strong inducible *lac* promoter**

In presence of IPTG, repression is relieved, T7 RNA pol is overexpressed and then activates transcription of the target gene

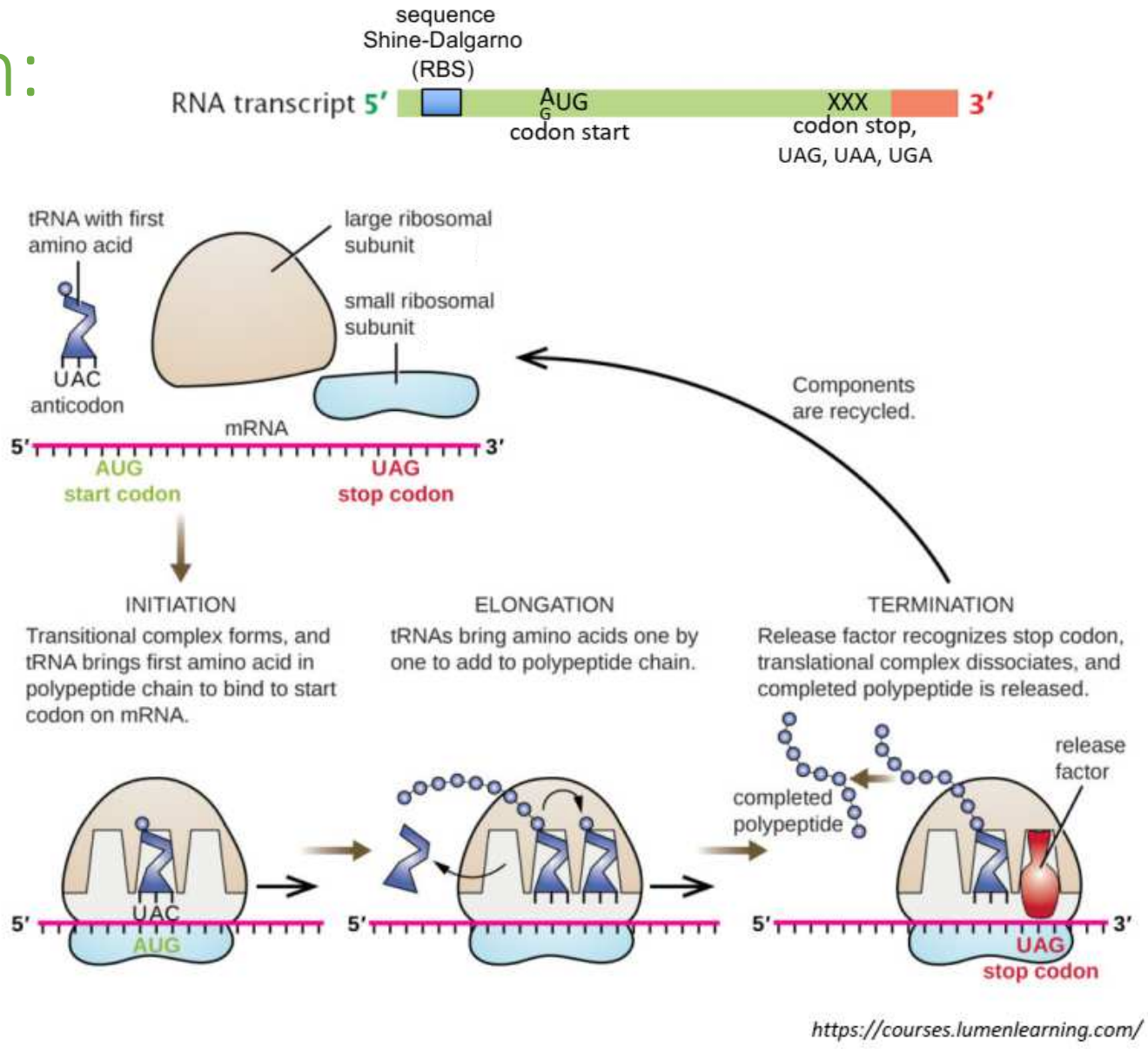
This highly efficient system mobilizes all the ribosomes in the cell = ***the speed of transcription increases from about 60 nucleotides per second to over 300 nucleotides per second***

From gene to protein: 2. Translation

- RBS corresponds to a 3 to 9 base canonical sequence acting as a binding site for the small subunit of the ribosome
- Start codon (AUG) must be located 4 to 12 nucleotide downstream of the RBS
- Stop codon are UAA, UAG or UGA.
- The nucleotide following the STOP codon is important for the efficiency of translation termination. UAUU is the best signal in *E. coli.*

➢ ***Codon bias & optimization of the gene***

➢ ***Folding strongly depends on the coupling between transcription and translation***



https://courses.lumenlearning.com/

The codon bias problem

- The genetic code is universal with 20 amino-acid encoded by 61 codons + 3 stop codons

BUT

- The frequency of usage in the genome and the frequency of each corresponding tRNA in the cell vary from one organism to the other
- some tRNAs can be limiting factors for translation

CCT	P	0.31	13.6
CCC	P	0.15	6.8
CCA	P	0.41	18.2
CCG	P	0.12	5.3

Saccharomyces cerevisiae

CCT	P	0.28	17.3
CCC	P	0.33	20.0
CCA	P	0.27	16.7
CCG	P	0.11	7.0

Homo sapiens

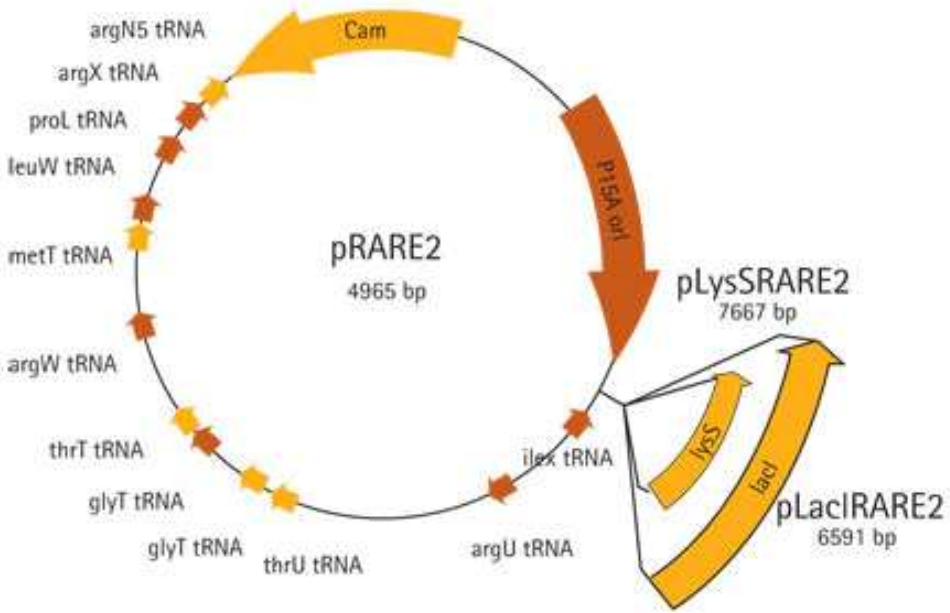
Triplet	Amino acid	Fraction	Frequency/ Thousand	Triplet	Amino acid	Fraction	Frequency/ Thousand
TTT	F	0.58	22.1	TCT	S	0.17	10.4
TTC	F	0.42	16.0	TCC	S	0.15	9.1
TTA	L	0.14	14.3	TCA	S	0.14	8.9
TTG	L	0.13	13.0	TCG	S	0.14	8.5
TAT	Y	0.59	17.5	TGT	C	0.46	5.2
TAC	Y	0.41	12.2	TGC	C	0.54	6.1
TAA	*	0.61	2.0	TGA	*	0.30	1.0
TAG	*	0.09	0.3	TGG	W	1.00	13.9
CTT	L	0.12	11.9	CCT	P	0.18	7.5
CTC	L	0.10	10.2	CCC	P	0.13	5.4
CTA	L	0.04	4.2	CCA	P	0.20	8.6
CTG	L	0.47	48.4	CCG	P	0.49	20.9
CAT	H	0.57	12.5	CGT	R	0.36	20.0
CAC	H	0.43	9.3	CGC	R	0.36	19.7
CAA	Q	0.34	14.6	CGA	R	0.07	3.8
CAG	Q	0.66	28.4	CGG	R	0.11	5.9
ATT	I	0.49	29.8	ACT	T	0.19	10.3
ATC	I	0.39	23.7	ACC	T	0.40	22.0
ATA	I	0.11	6.8	ACA	T	0.17	9.3
ATG	M	1.00	26.4	ACG	T	0.25	13.7
AAT	N	0.49	20.6	AGT	S	0.16	9.9
AAC	N	0.51	21.4	AGC	S	0.25	15.2
AAA	K	0.74	35.3	AGA	R	0.07	3.6
AAG	K	0.26	12.4	AGG	R	0.04	2.1
GTT	V	0.28	19.8	GCT	A	0.18	17.1
GTC	V	0.20	14.3	GCC	A	0.26	24.2
GTA	V	0.17	11.6	GCA	A	0.23	21.2
GTG	V	0.35	24.4	GCG	A	0.33	30.1
GAT	D	0.63	32.7	GGT	G	0.35	25.5
GAC	D	0.37	19.2	GGC	G	0.37	27.1
GAA	E	0.68	39.1	GGA	G	0.13	9.5
GAG	E	0.32	18.7	GGG	G	0.15	11.3

Escherichia coli

http://www.kazusa.or.jp/codon/
https://www.genscript.com/tools/codon-frequency-table

How to get around the codon bias problem?

- Optimize the sequence by ordering a synthetic gene
  - Cost have decreased enormously over the last twenty years with the technological advances around gene synthesis (~$0.15-0.20/bp)
- Use of modified bacteria strains
  - BL21-CodonPlus or Rosetta cells
- Co-transformation with a plasmid coding for rare or missing tRNA (here, 12 tRNA)



Elongation rate and folding rate of the polypeptide chain

- **In bacteria**, transcription is closely coupled to translation

  Transcription ~ **60 nucleotides read/s** → Translation ~ 20 assembled amino-acid/s

  In wild-type conditions ribosomes and tRNA are not rate limiting

From a pET system, this coupling no longer exists
because the speed of transcription can exceed 300 nucleotides/s

💣 Translation can be overwhelmed resulting in

- an activation of mRNA degradation pathways (Sørensen & Mørtensen (2005) Microb Cell Fact)
- the formation of secondary structures blocking polypeptide synthesis

= lower expression yields despite a higher number of transcripts

Elongation rate and folding rate of the polypeptide chain

- **In eukaryotes,** the rate of polypeptide synthesis is 5-10 times slower than in prokaryotes, but protein folding is more complex (Goustin et al. (1982); Widman et al. (2000)).

➢ Difficult to obtain eukaryotic proteins in *E. coli* (inclusion bodies, proteolysis, misfolding)

➢ Slow down transcription : weaker promoter, weaker replication origin

➢ Slow down biosynthesis by decreasing the temperature

➢ Improving the quality of transcripts:
- Mutation to increase the stability of mRNA
- Optimise codons
- Test of several homologous sequences

➢ Improve culture conditions (substrate, additives, [inducer])

➢ Test of different strains or organisms

Effect of temperature on the quality of overexpressed proteins

Decreasing the temperature slows down enzymes activity (transcription, translation, proteases), increases the amount of chaperones, while the folding processes are mostly autonomous and do not see their speed affected.

Expression of the bacterial Mg^{2+} transporter CorA in *E.coli*

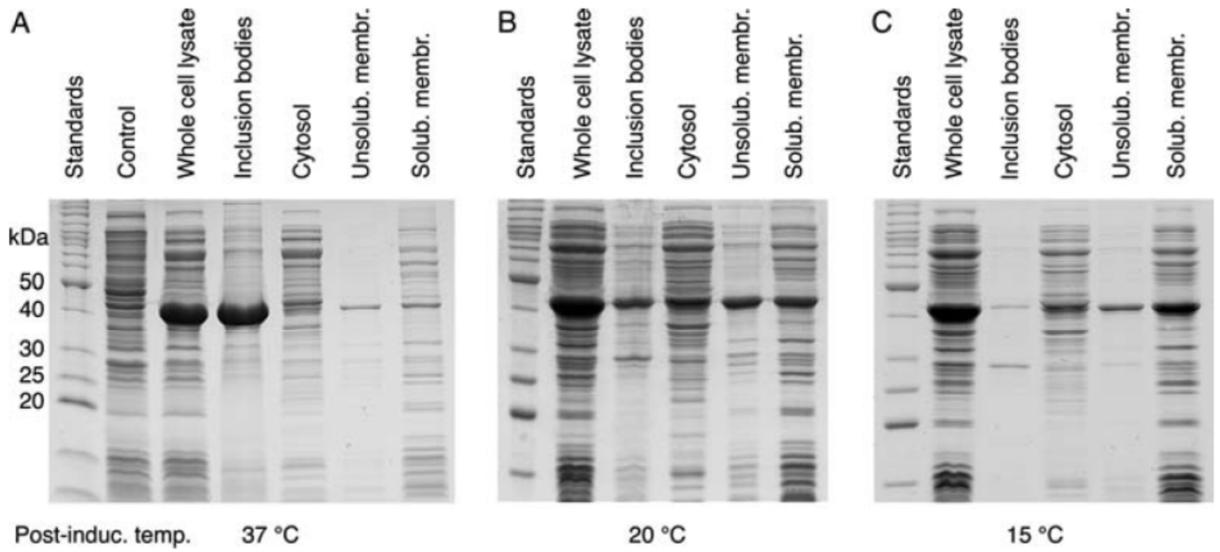
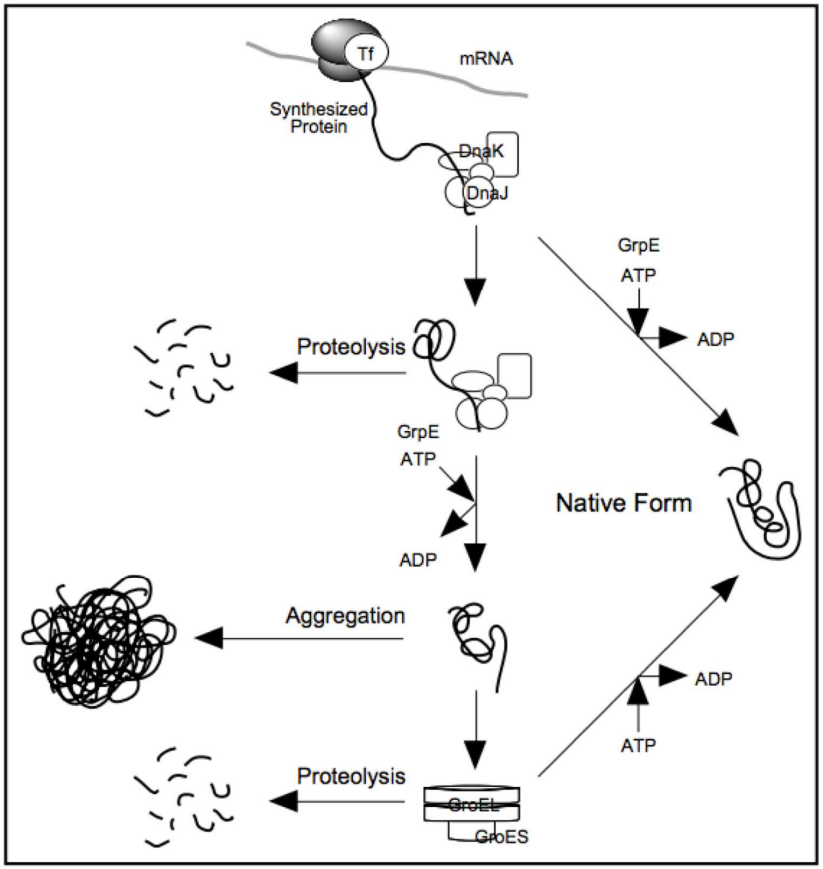


Fig. 1. SDS–PAGE analysis (12%, Coomassie blue-stained) of CorA overexpressed in *E. coli* at (A) 37 °C, (B) 20 °C, and (C) 15 °C. Control lane at 37 °C: uninduced whole cell lysate. The expressed CorA protein migrated as a 40 kDa polypeptide. Inclusion bodies were solubilized in 6 M urea before the sample was loaded onto the gel.

Chen *et al.* Protein Expr Purif (2003)

Role of chaperone proteins



- Chaperone proteins (e.g DnaK/J) bind to nascent polypeptidic chain to prevent aggregation of hydrophobic patches, therefore preventing misfolding. Others directly assist the folding (GrpE, GroEL/S)

Name	Chaperone	Promoter	Company
pG-KJE8	DnaK, DnaJ, GrpE (DnaK system), GroELS	araB and Pzt-1	Takara
pGro7	GroELS	araB	Takara
pKJE7	DnaK, DnaJ, GrpE (DnaK system)	araB	Takara
pG-Tf2	Tig (trigger factor), GroELS	Pzt-1	Takara
pTf16	Tig (trigger factor)	araB	Takara
pBB530	GrpE (DnaK system)	PA1/lacO1	Addgene
pBB535	DnaK, DnaJ (DnaK system)	PA1/lacO1	Addgene
pBB540	GrpE (DnaK system), ClpB (HSP 100 family)	PA1/lacO1	Addgene
pBB542	DnaK, DnaJ, GroESL(Large amounts)	PA1/lacO1	Addgene
pBB550	DnaK, DnaJ, GroESL(Small amounts)	Plac/lacO1	Addgene
pBB872	ibpB, ibpA (small HSPs)	PA1/lacO1	Addgene
pColdI-IV	None (Cold shock protein promoter)	CspA	Takara

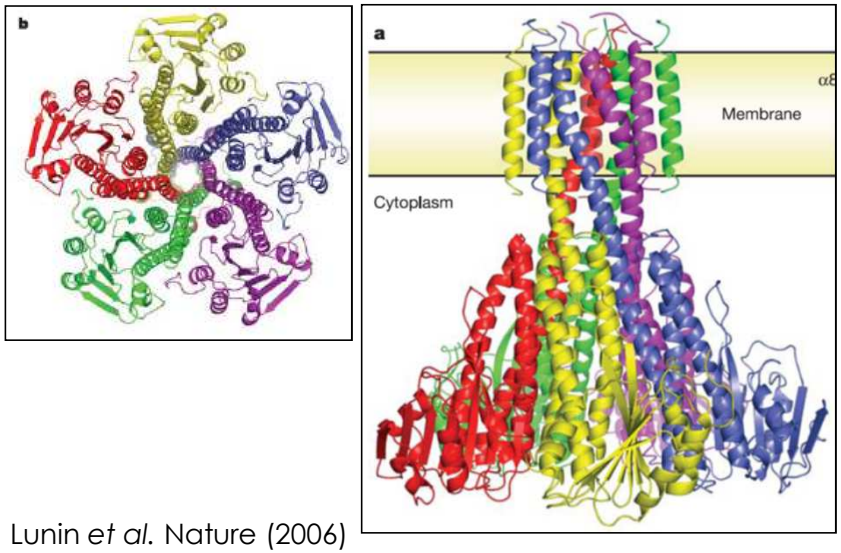
Table 1: Chaperone co-expression systems.

J Microb Biochem Technol 2016, 8:2. http://dx.doi.org/10.4172/1948-5948.1000266

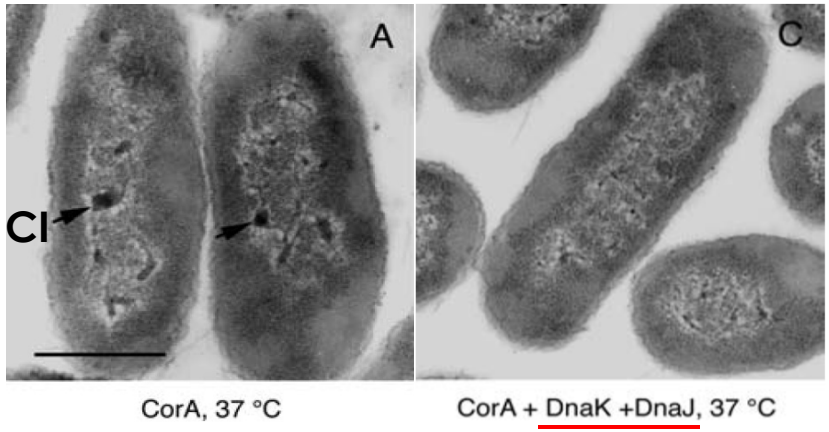
Co-expression of chaperones

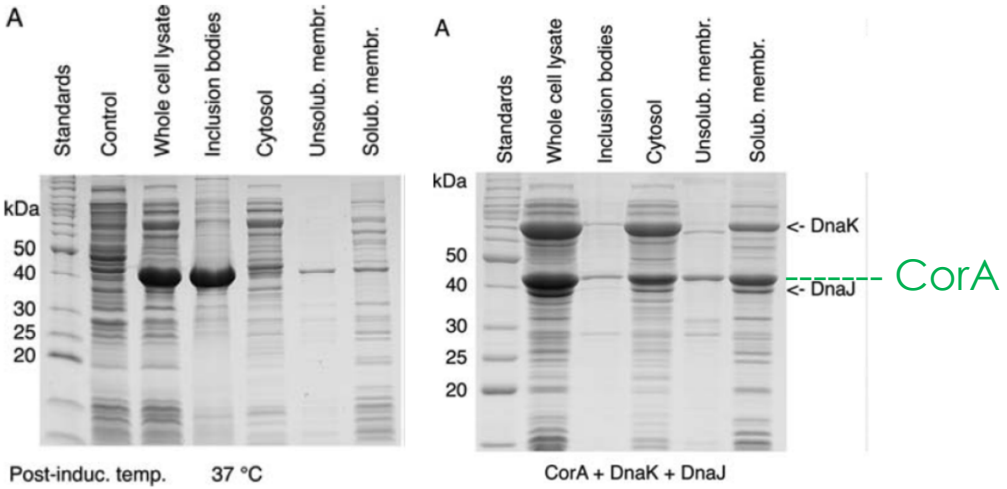
Expression of the bacterial Mg^{2+} transporter CorA in *E.coli*

- Physiology
  - Bacterial voltage-dependent ionic channel,
  - Main entry pathway for Mg^{2+},
- Structure
  - homo-pentamer, 10 TM (5x2),
  - 5 x 37 kDa soit ~185 kDa
- Expression / Purification strategy
  - cDNA from *Thermotoga maritima*,
  - *E. coli*, N-ter hexahistidine tag.



Lunin *et al.* Nature (2006)





Chen *et al.* Protein Expr Purif (2003)

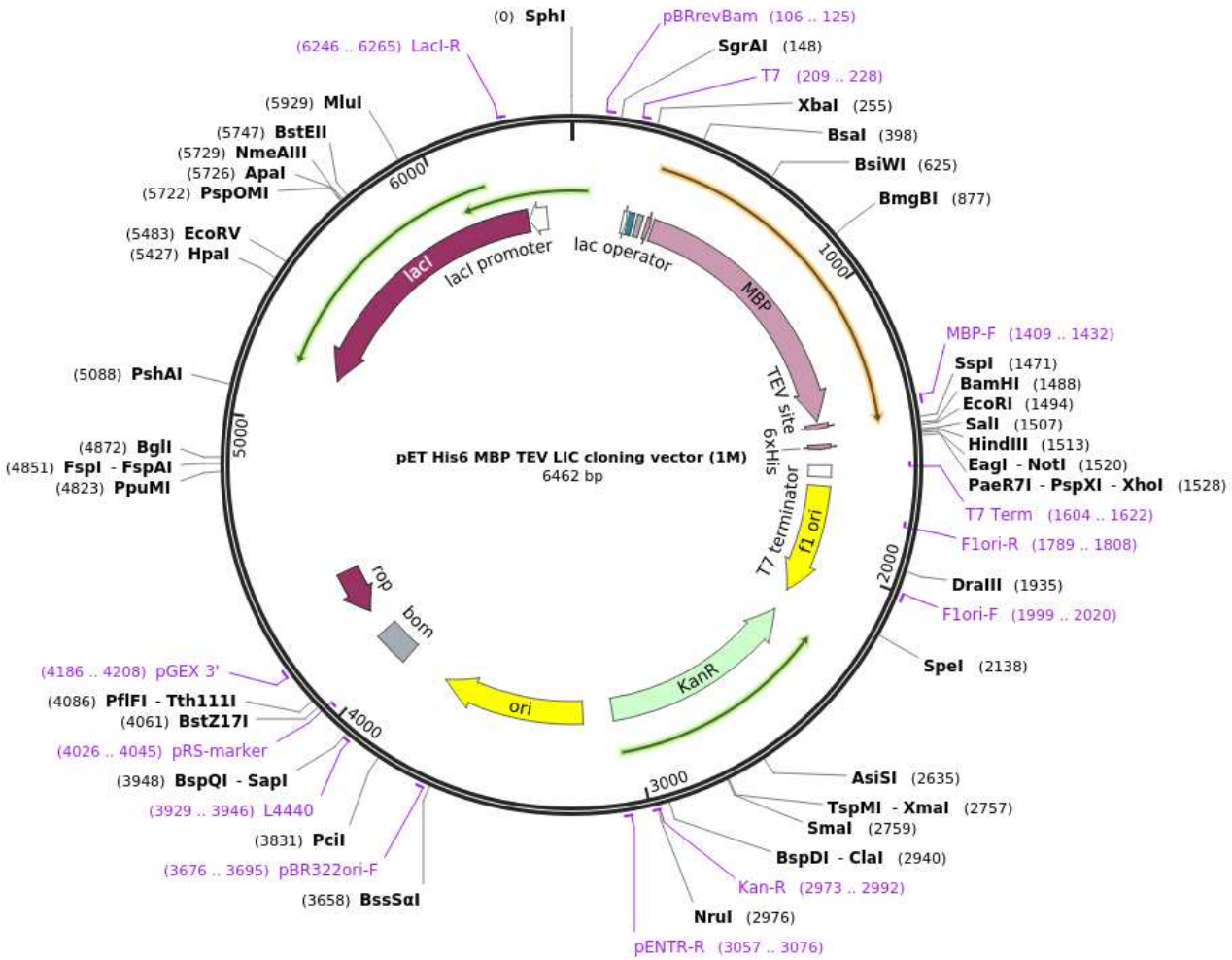
Use of fusion partners: Why?

- To improve the initiation of the translation
- To improve the solubility, help in folding
- To address the protein to the plasma membrane or the periplasm
- To help for purification e.g on affinity chromatography

BUT

- No universal partner… let's try ☺
- Cleavage may occur spontaneous ly during translation between the protein of interest and the fusion partner
- Should be removed so as not to interfere with the activity of the protein of interest

Fusion partners: how to add them ?



MCS for cloning of the sequence of interest

Open Reading Frame

1. ATG CAA TGG GGA AAT GTT ACC AGG TCC GAA CTT ATT GAG GTA AGA CAG ATT TAA
2. A TGC AAT GGG GAA ATG TTA CCA GGT CCG AAC TTA TTG AGG TAA GAC AGA TTT AA
3. AT GCA ATG GGG AAA TGT TAC CAG GTC CGA ACT TAT TGA GGT AAG ACA GAT TTA A

This sequence contains 3 frames but only ORF 1 allows complete translation

In fusion proteins, the sequence coding the fusion partner
and the sequence coding the protein of interest should share the same ORF
in order to obtain the full-length fusion protein

Table 1. Advantages and disadvantages of some commonly used fusion partners

+ Green Fluorescent Protein and derivatives

Tag[a]	Advantages	Disadvantages
GST	Efficient translation initiation	High metabolic burden
	Inexpensive affinity resin	Homodimeric protein
	Mild elution conditions	Does not enhance solubility
MBP	Efficient translation initiation	High metabolic burden
	Inexpensive affinity resin	
	Enhances solubility	
	Mild elution conditions	
NusA	Efficient translation initiation	High metabolic burden
	Enhances solubility	
	Not an affinity tag	
Thioredoxin	Efficient translation initiation	Not an affinity tag[b]
	Enhances solubility	
Ubiquitin	Efficient translation initiation	Not an affinity tag
	Might enhance solubility	
FLAG	Low metabolic burden	Expensive affinity resin
	High specificity	Harsh elution conditions
BAP	Low metabolic burden	Expensive affinity resin
	Mild elution conditions	Variable efficiency of enzymatic biotinylation
	Provides convenient means of immobilizing proteins in a directed orientation	Co-purification of *E. coli* biotin carboxyl carrier protein on affinity resin
		Does not enhance solubility
His_6	Low metabolic burden	Specificity of IMAC is not as high as other affinity methods
	Inexpensive affinity resin	
	Mild elution conditions	
	Tag works under both native and denaturing conditions	Does not enhance solubility
STREP	Low metabolic burden	Expensive affinity resin
	High specificity	Does not enhance solubility
	Mild elution conditions	
SET	Enhances solubility	Not an affinity tag
CBP	Low metabolic burden	Expensive affinity resin
	High specificity	Does not enhance solubility
	Mild elution conditions	
S-tag	Low metabolic burden	Expensive affinity resin
	High specificity	Harsh elution conditions (or on-column cleavage)
		Does not enhance solubility

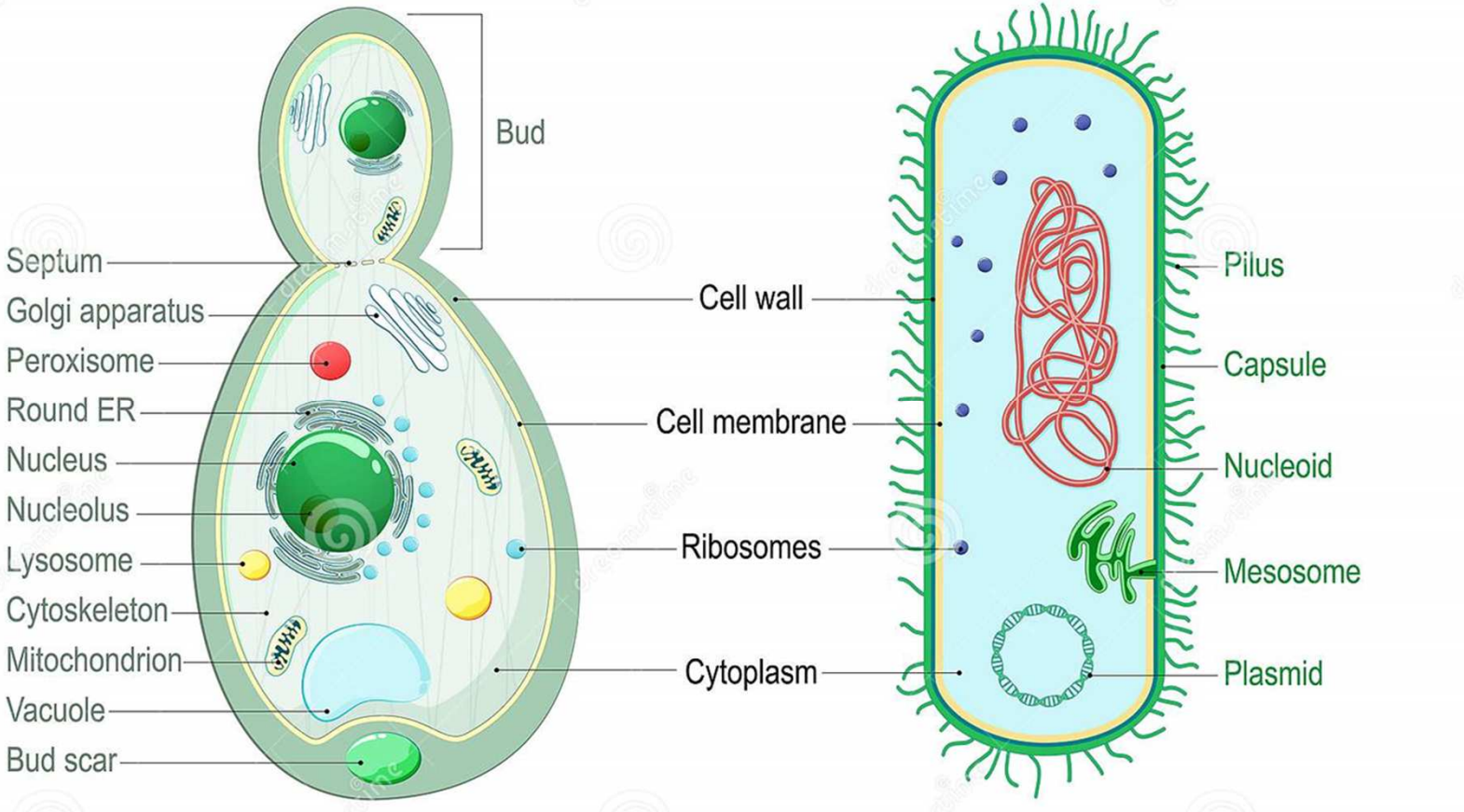
[a]GST, glutathione S-transferase; MBP, maltose-binding protein; NusA, N-utilization substance A; FLAG, FLAG-tag peptide; BAP, biotin acceptor peptide; His_6, hexahistidine tag; STREP, streptavidin-binding peptide; SET, solubility-enhancing tag; CBP, calmodulin-binding peptide.

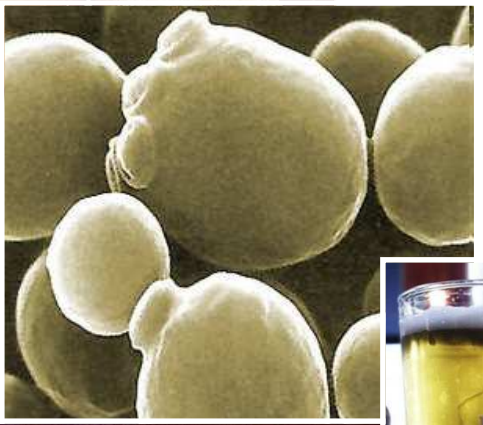
[b]Derivatives of thioredoxin have been engineered to have affinity for immobilized metal ions (His-patch thioredoxin) or avidin/streptavidin [38].

Waugh Trends in biotechnology (2005)

Expression in yeast

Small but with all the qualities of a big one





The baker's yeast *Saccharomyces cerevisiae*

- Also called "brewer's yeast" or "budding yeast"
- Unicellular organism, the simpliest eucaryote
- Numerous applications in the industry (alcohol, bread, soil decontamination) and research (cloning and expression of proteins, model organisms)
- First eukaryote genome fully sequenced in 1996

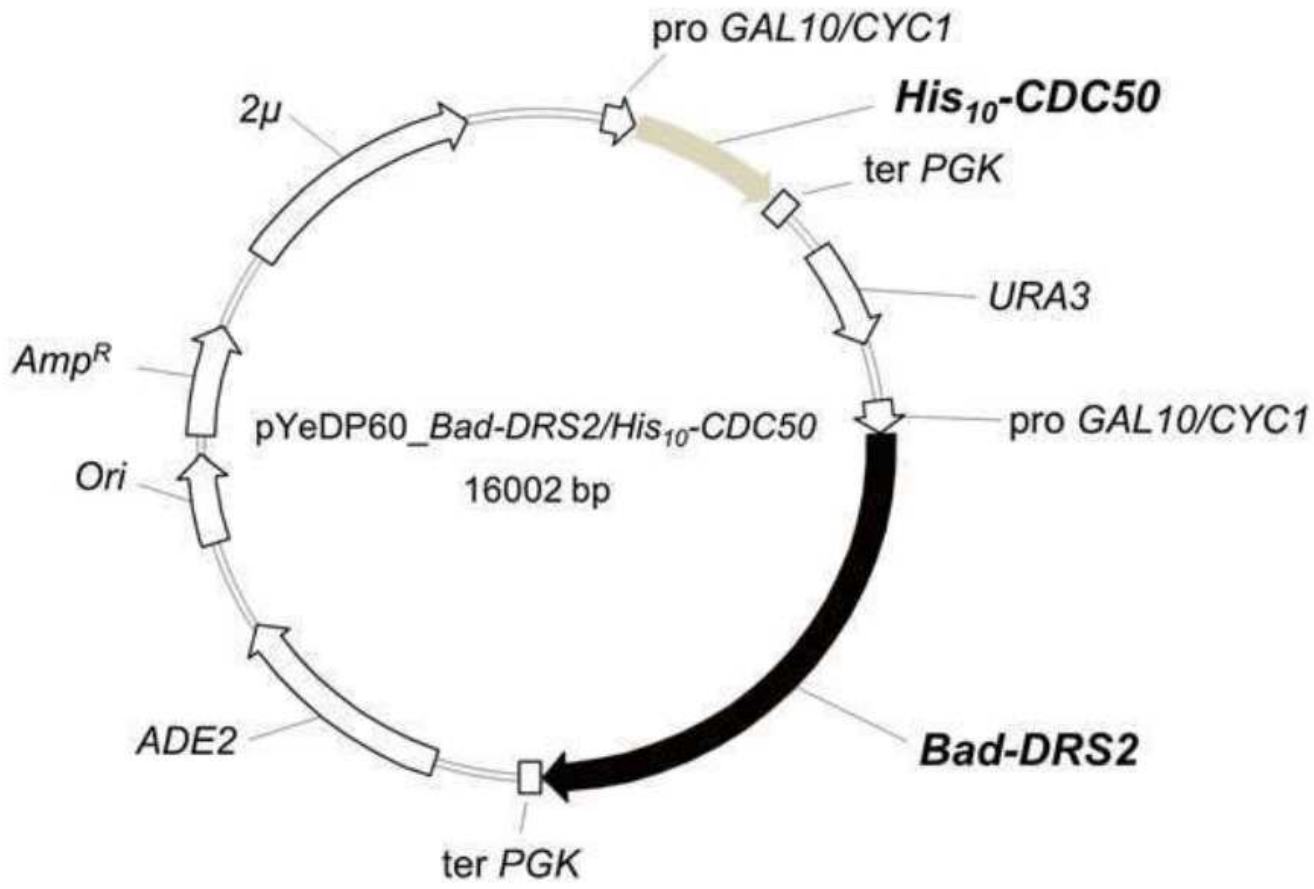
- Use of both constitutive or inducible (GAL10) promoters, on plasmids or after integration in the genome
- Selection by auxotrophy or drug resistance

Pichia pastoris, the methylotrophic yeast

- Also a budding yeast, able to grow in presence of methanol
- As cheap as *S. cerevisiae* to handle
- Can grow at very high density
- First eukaryote genome fully sequenced in 1996

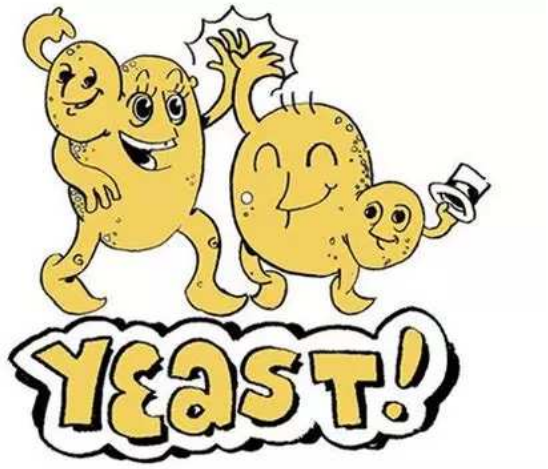
- Use of inducible (AOX1 and AOX2) promoters, on plasmids or after integration in the genome
- Selection by auxotrophy or drug resistance
- Secretion of the recombinant protein can be considered

An example of yeast expression vector



P-Type ATPases: Methods and Protocols, Methods in Molecular Biology, vol. 1377, DOI 10.1007/978-1-4939-3179-8_6

Selection markers confer resistance to drugs or complementation



- Complementation

  Strains deleted for essential gene for growth have been developed for selection. The plasmid contains a gene coding the missing enzyme to restore the metabolic pathway

  - URA3 or ADE2 for nucleotide biosynthesis pathways
  - TRP1, LEU2, HIS3 (…) for essential amino-acid biosynthesis

- Resistance to drug

  As for bacteria, vectors can contains gene confering resistance to toxic drugs (Geneticin (G418), neomycin)

Improvement of expression

Same tips as for bacteria

- Composition of the medium
- Temperature
- Type of promoter
- Use of fusion proteins
- Co-expression of chaperones
- …

Expression in insect cells using the baculovirus system

Spodoptera frugiperda cell lines

- the most common cell line is Sf9
- Baculoviridae: DNA enveloped viruses specific to arthropods
- Use since the early 1980's as the only system combining the advantages of prokaryotic systems in terms of yield and those of mammalian cells in terms of quality and PTM

+ Safety: *As Baculovirus are specific to insect cells, they are nonpathogenic to mammals and plants. It can be used under minimal biosafety conditions.*

Jarvis Methods in Enzymology (2009) https://doi.org/10.1016/S0076-6879(09)63014-7

Why using insect cells combined to baculovirus?

ADVANTAGES

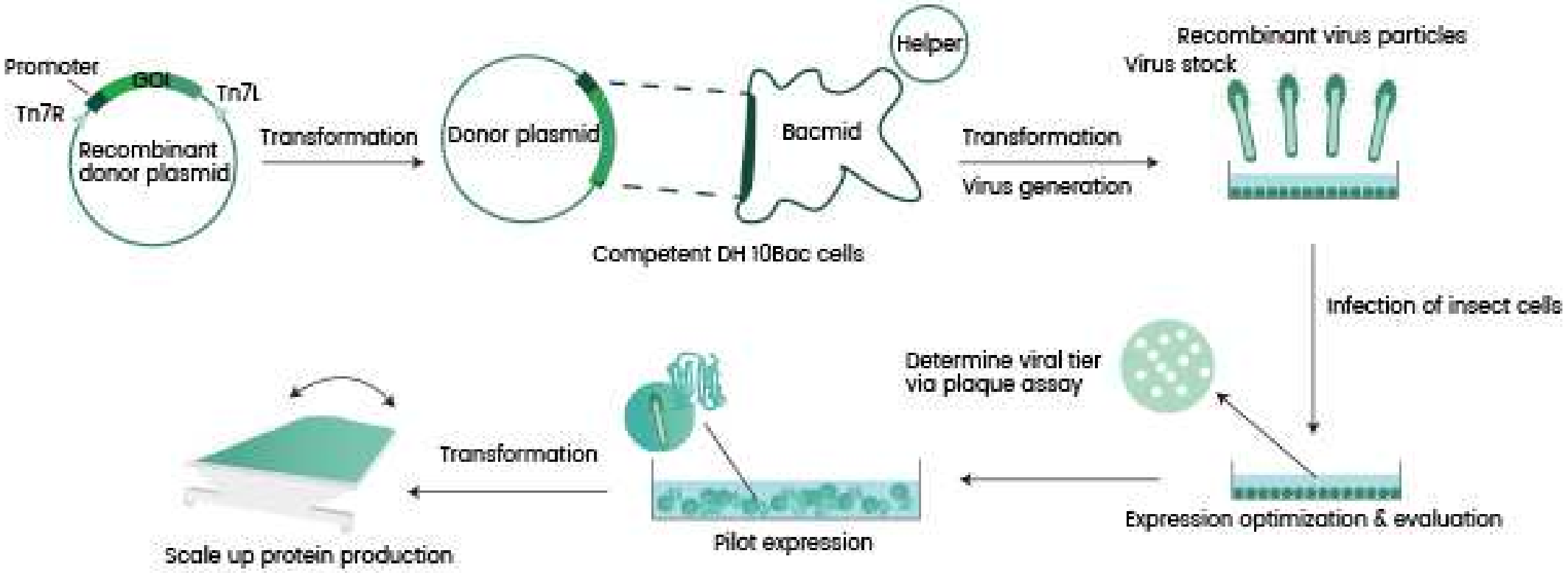
- **Eukaryotic system:** Insect cells carry out PTM similar to that of mammalian cells.
- **Strong expression and good solubility:** Secretion or internal expression.
- **Ease of scale-up:** from a few milliliters to large 20 liters wave cellbags in research labs, upto 500 l in industry. Growth in suspension of Lepidoptera cell lines.
- **MultiBac system:** easy to co-express up to 4-6 proteins,

  maybe more… the limit is the size of the vector.

DISADVANTAGES

- **Technical complexity:** Protein production requires multiple steps, including generating recombinant bacmid from donor vector, and transfecting bacmid into insect cells. These procedures are technical demanding and time consuming.
- Lipid composition of insect cells

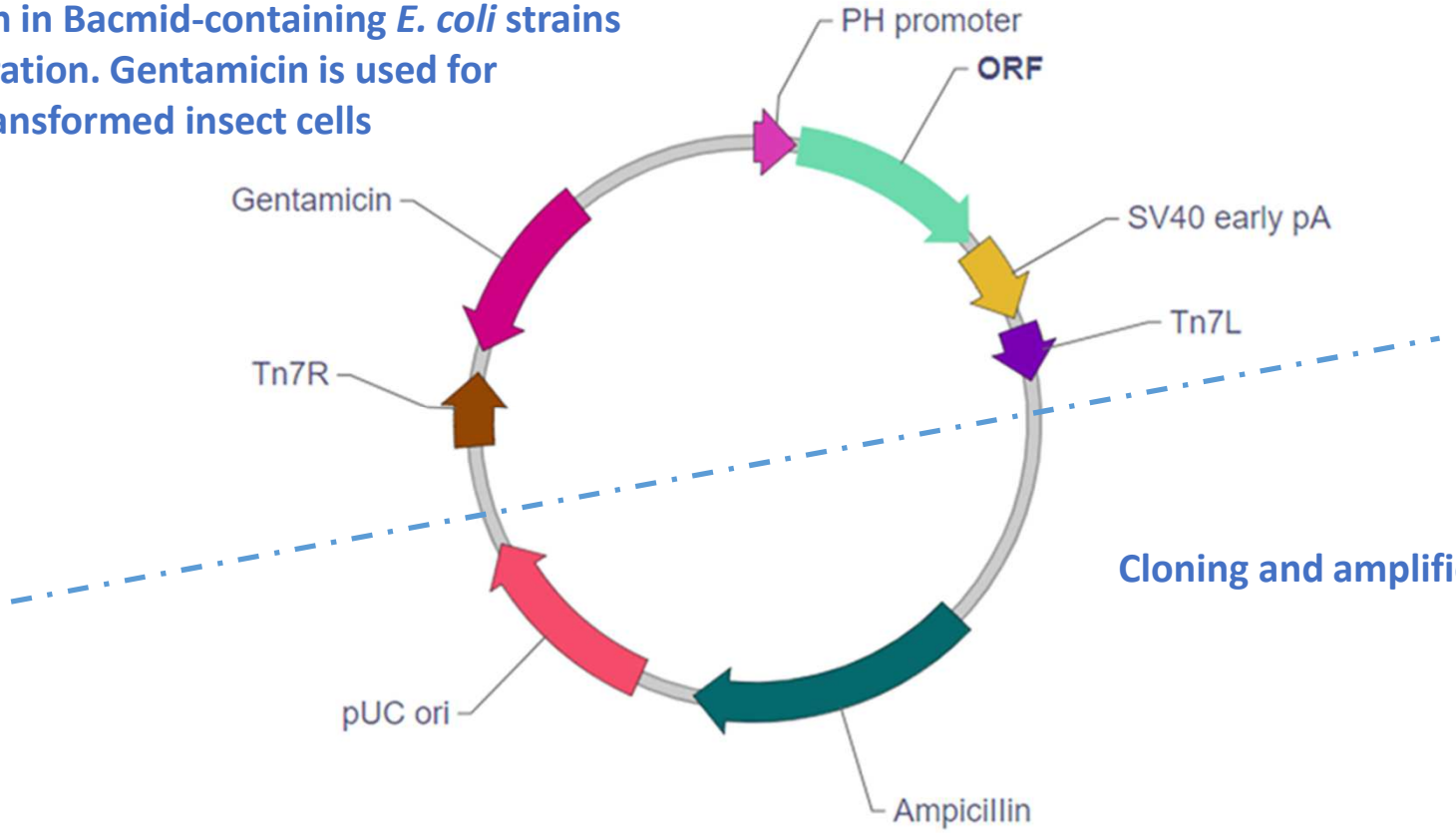
Recombinant protein expression in insect cells



https://www.sinobiological.com/news/insect-cells-for-protein-expression

A typical vector for baculovirus synthesis

Recombination in Bacmid-containing *E. coli* strains for virus generation. Gentamicin is used for selection of transformed insect cells



Cloning and amplification in *E. coli*

https://en.vectorbuilder.com/resources/vector-system/pBV.html

Expression in mammalian cells

Expression in mammalian cells

	Expression System	Pros	Cons
Mammalian cells	General	*Native lipid environment *Native secretory/post-translational pathways Good track record of functional expression	Cost Technical requirements
	Stable integration	Consistent expression levels No need to generate vector	Time required to establish Instability of integration
	Semliki Forest Virus (SFV)	Success in expression-screening studies Efficient infection	Requirement for helper RNA Technically demanding
	Transient transfection	Speed Structure of recombinant rhodopsin	Amount of DNA required Cost of reagents at large scale
	Bacmam	Lack of lysis in mammalian cells Cross-over	Requirement for two different kinds of cells

+ Golden standard for expression of eukaryotic proteins
- Technical demanding, time consuming and expensive
- Variability of the cell lines after selection depending on the marker and on the technique used for transfection

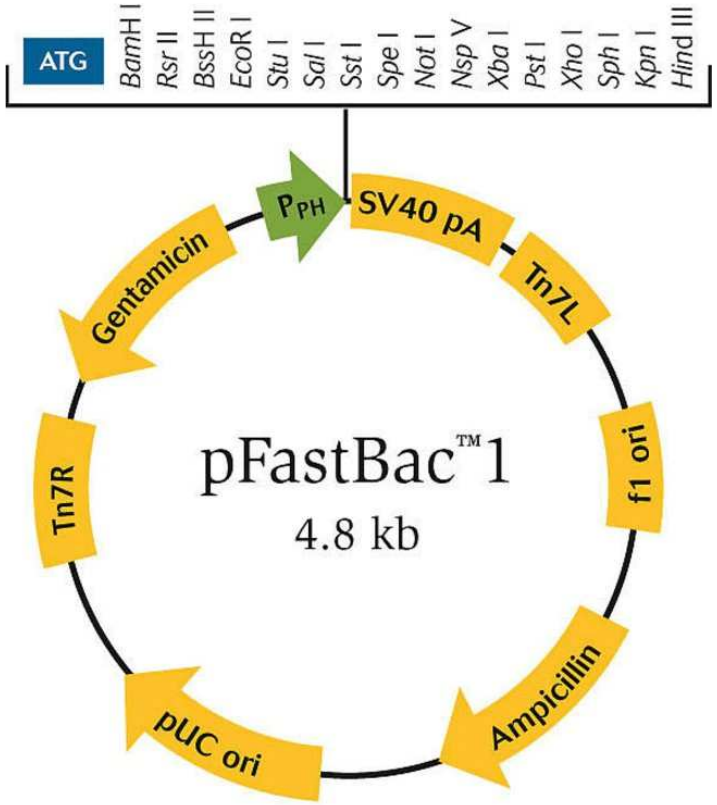
Cell lines and vectors

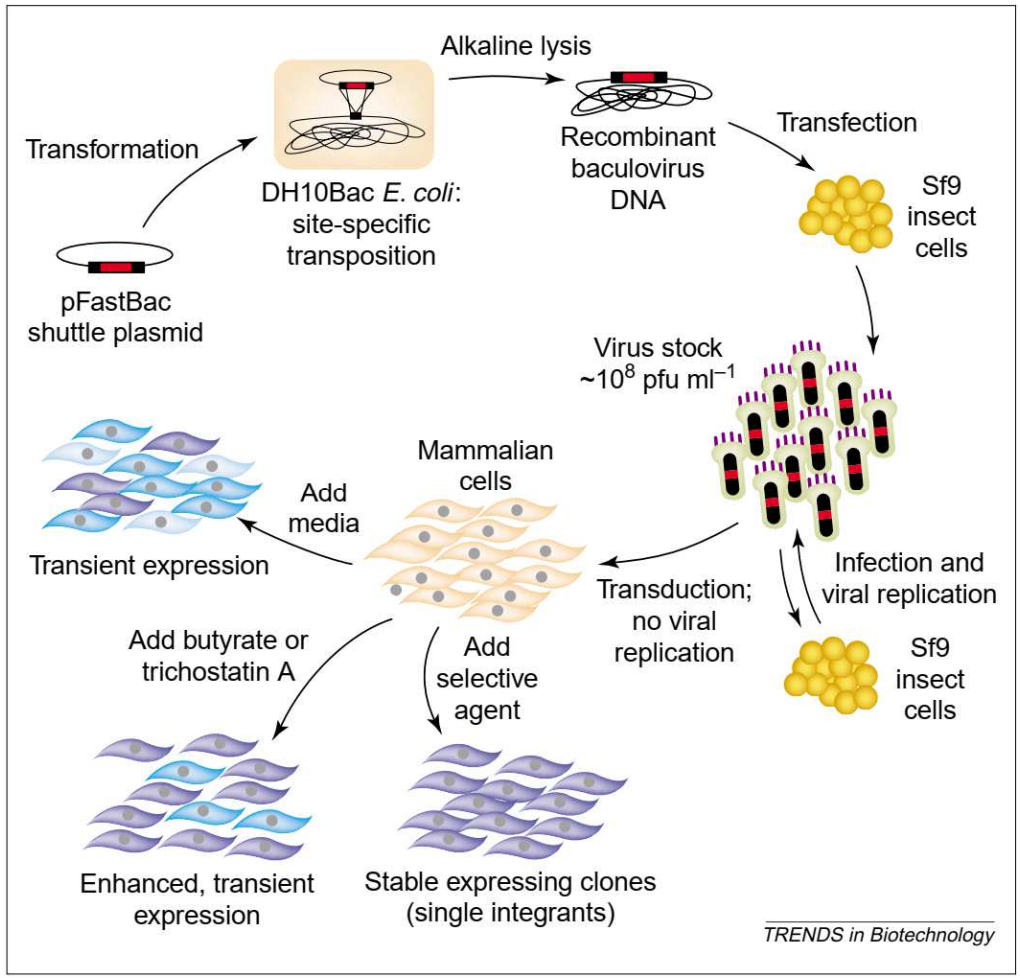
- Several dozen different cell lines immortalized from carcinomas and tumors
- Adherent cells (attachment to plates) or suspension (flasks, wave cell bags, bioreactors) for small-scale to industrial yield of production

- Selection by drug resistance (zeocin, neomycin, G418, blasticidin)
- Large variety of promoters (weak/strong, constitutive/inducible; see below)
- Most of the fusion proteins and affinity tags can be used

Curr Protoc Protein Sci. 1997 Dec; 10(1): 591–5918. doi: 10.1002/0471140864.ps0509s10
J Biol Chem 2021; https://doi-org.insb.bib.cnrs.fr/10.1016/j.jbc.2021.100838

Baculovirus for recombinant protein expression in mammalian cells: The BacMam system





Boyce et al. PNAS 1996; Kost et al. Nat. Biotch (2005); Kost et al. Trends in Biotech (2002); Methods Mol Biol. 2020;2125:205-208. doi: 10.1007/7651_2019_249; https://geneva-biotech.com/bacmam-compatible-cells/

The BacMam system: Pros & Cons

ADVANTAGES

- Baculovirus can be easily manipulated and produced
- Broad cell type specificity, transduction of primary cell types
- Little to no microscopically observable cytopathic effect
- Transient or stable transduction
- Safety: non-replicative in mammalian cells
- Large stable insert capacity (38 kb)
- Simultaneous delivery of multiple genes (MultiBacMam system)

DISADVANTAGES

- Need two cell lines, insect cells for replication of the virus and a mammalian cell line for expression
- Technical demanding, time consuming and expensive

Cell-free expression system

What is cell-free expression?

- Use of a purified machinery for *in vitro* protein synthesis by only providing mRNA (Niremberg et Matthaei, 1961)

BUT

- Secondary products from transcription and translation inhibit the reaction
- ATP is very limiting as the only energy source
- Nucleotides are not stable a long time especially mRNA

Cell-free expression systems for eukaryotic protein production. Endo Y, Sawasaki T. Curr Opin Biotechnol (2006) ; Cell-free gene expression: an expanded repertoire of applications. Silverman AD, Karim AS, Jewett MC. Nat Rev Genet. 2020 Mar;21(3):151-170

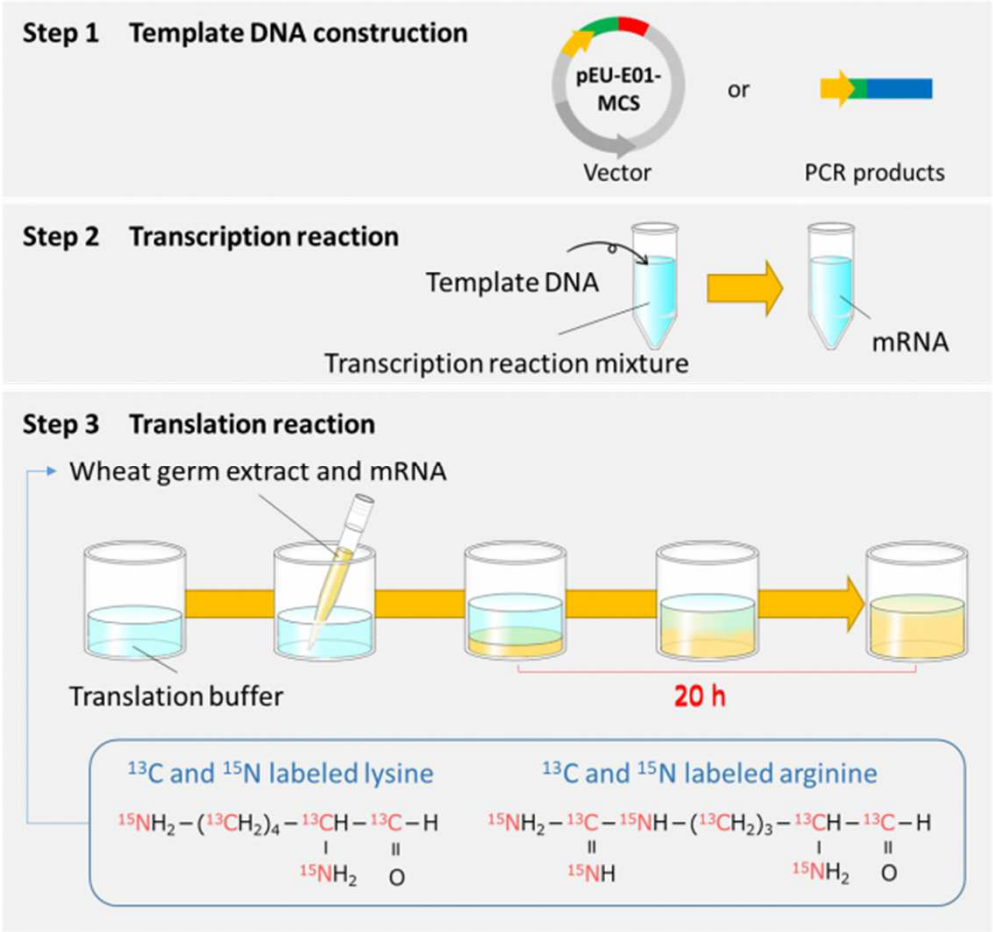
Available systems

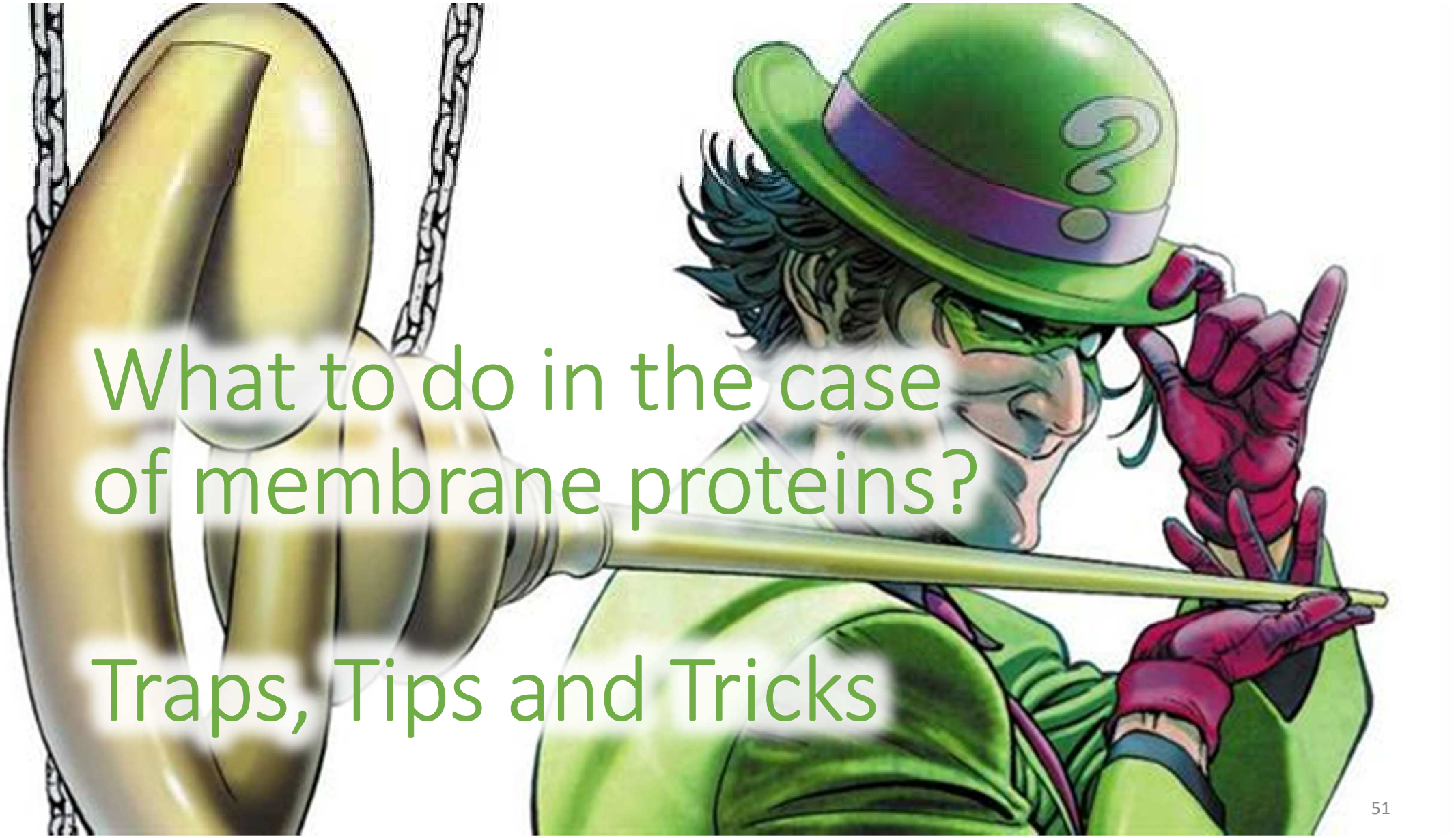
- PURE system from New England Biolabs = pure expression system reconstituted from recombinant protein BUT very expensive
- *E. coli* lysate or extract, commercial or hand-made
- Wheat germ extract, commercial or hand-made. Protocols to get very low amount of translation inhibitors.

- Continuous dialysis systems or “Bilayer” reaction system to get rid of secondary products

Expensive for very large production but very handy for specific labelling

- In cells, all the protein will incorporate the labelled amino-acid whereas in cell-free systems, only the protein of interest will be labelled.
- Recommend for toxic proteins





What to do in the case of membrane proteins?

Traps, Tips and Tricks

Main pitfalls in the heterologous expression of membrane proteins (MPs)

In addition to the above parameters considered so far you should consider

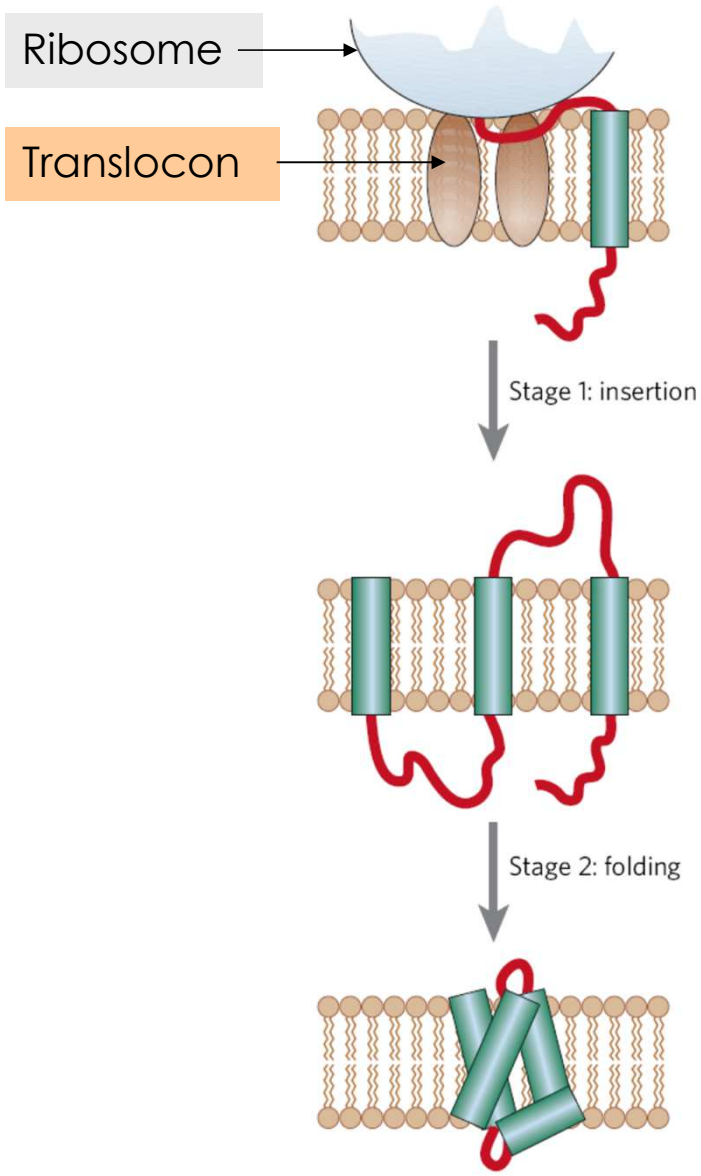
- Adressing to the membranes / to the right compartment or organelle
- Elongation, folding and maturation (PTM?)
- Lipids (type of membrane, cofactors for stabilisation or activation)
- Toxicity

Addressing,

Elongation,

Folding

and maturation of MPs



Addressing and maturation of MPs

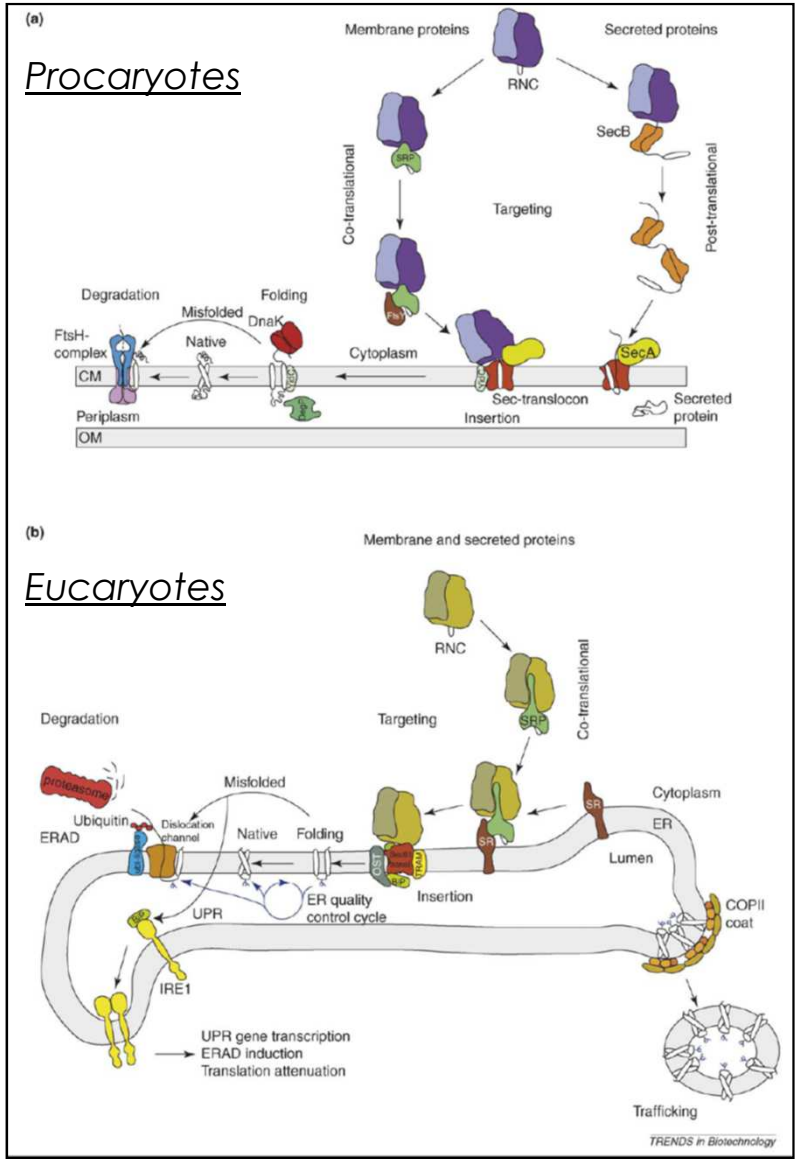
Bacteria

- Expression at the plasma membrane
- **asymmetry in K and R** (von Heijne (1989 & 1992) « *the positive-inside rule* »)
- Maturation is possible in the periplasm

Eucaryote

- Expression in the ER (translocon and bound ribosomes)
- maturation in the lumen and then at the Golgi (**glycosylation**, phosphorylation…)
- **Chaperones** (Kota *et al.* (2005))

- Add charges or signal peptide to improve membrane insertion
- Co-express chaperones (see above, CorA production)
- Use of humanized strains



Wagner *et al.* Trends Biotech. (2006)

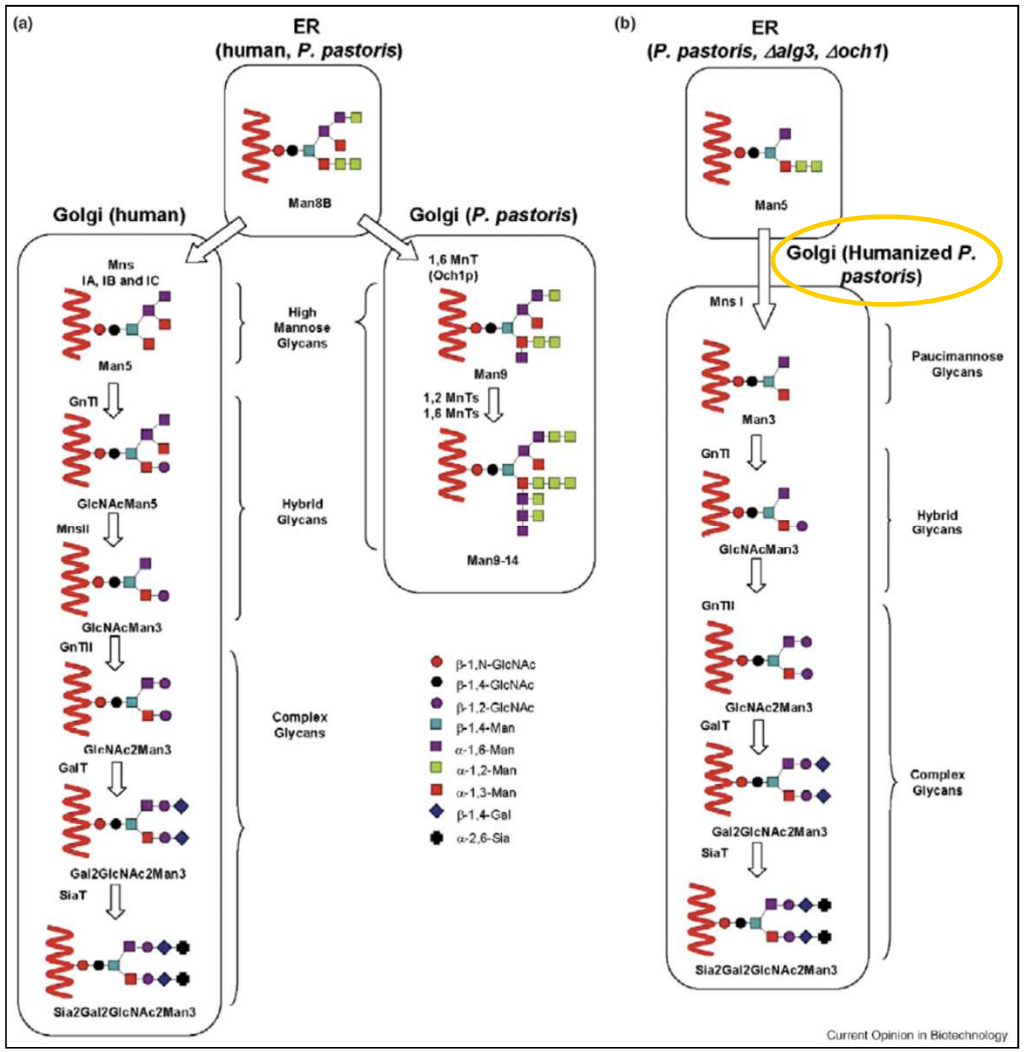
Development of humanized yeast strains for glycosylation of MPs

- Frequent hyperglycosylation to mannose in *S. cerevisiae* and *P. pastoris*

- Humanization by homologous recombinaison

  = addition of N-acétylglucosamine, galactose and sialic acid.

☺ Successful strategy
for IL-1 receptor
and DC-SIGN adhesion proteins



Hamilton *et al.* Science (2003), Science (2006), Cur. Opin. Biotech. (2007), Meth Mol Biol (2015) ; Jacobs *et al.* Nat. Protocols (2009) ; Wang *et al.* Microb Cell Fact (2020)

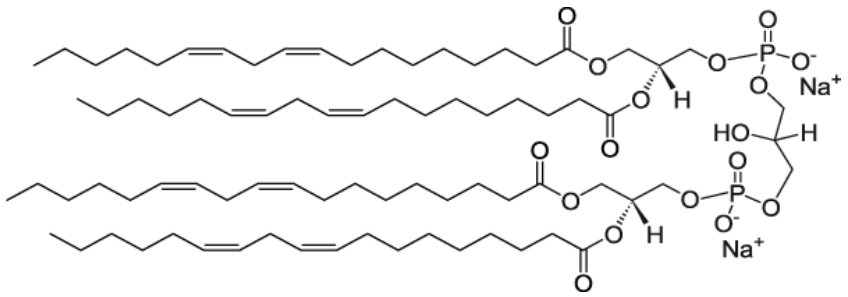
Nature
and composition of membranes

Composition of cell membranes

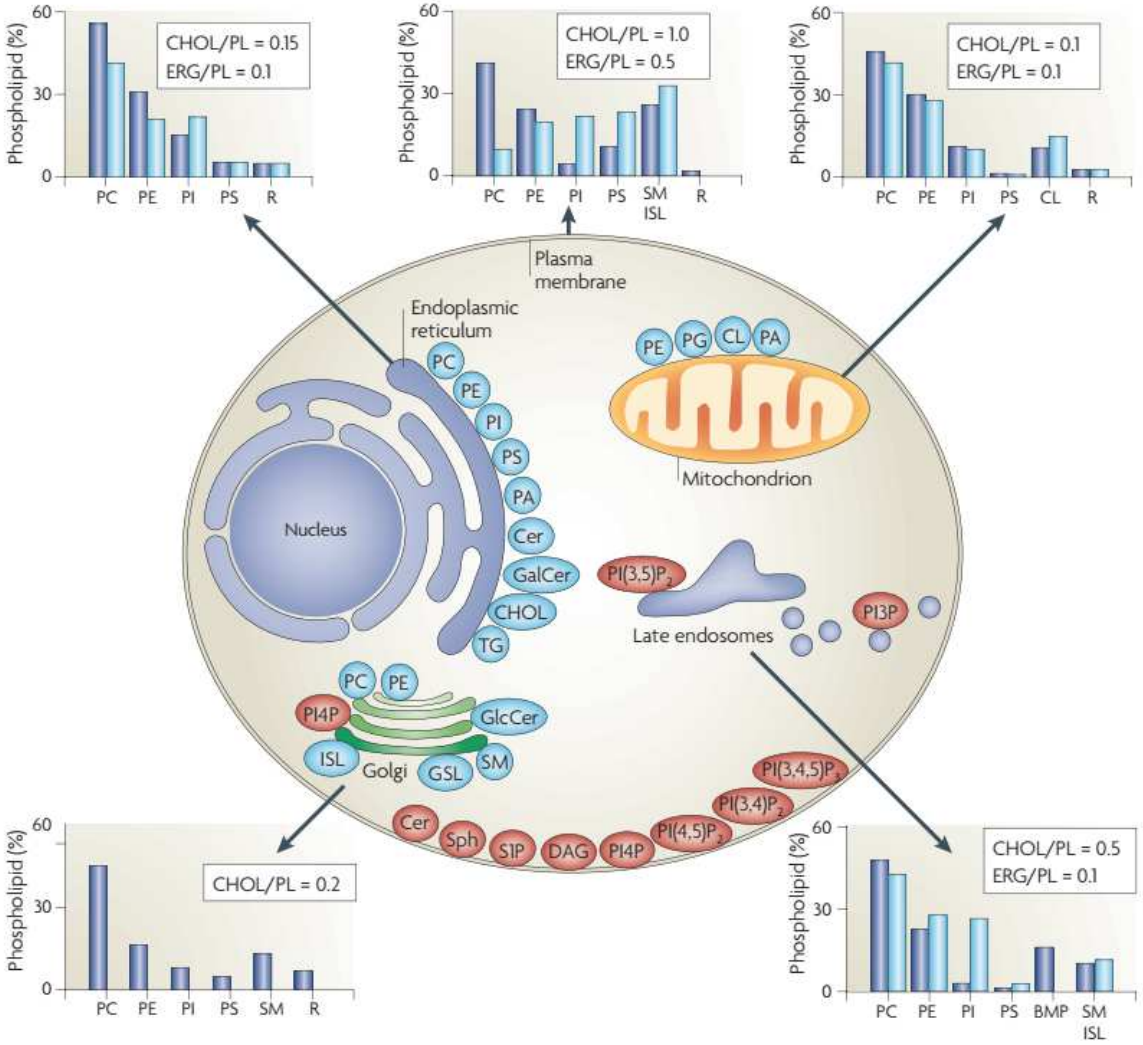
mammals

S. cerevisiae

***E. coli* Total Extract Phospholipid Profile**	
Component	wt/wt%
PE	57.5
PG	15.1
CA	9.8
Unknown	17.6
Total	100.0



Cardiolipin (CA)

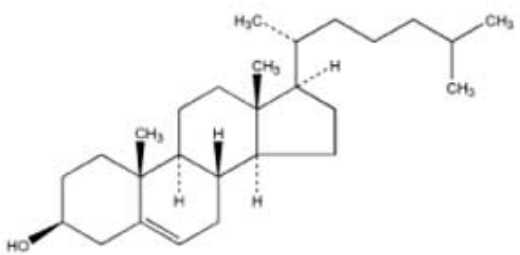


Avanti Polar lipids

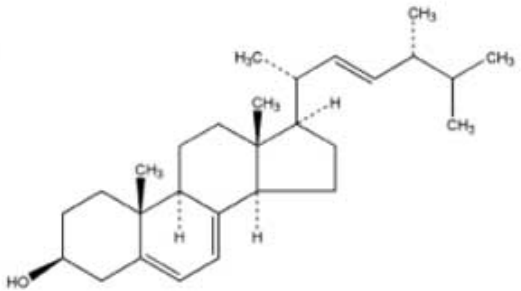
van Meer et al. Nat Rev Mol Cell Biol. 2008 Feb;9(2):112-24. doi: 10.1038/nrm2330.

Activating/stabilising role of some lipids

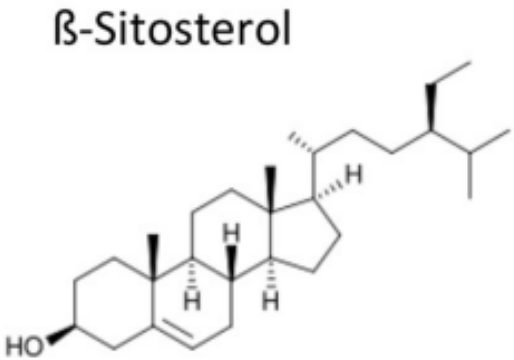
Lipids can act as cofactors, be involved in folding or added to protein as PTM



Cholesterol



Ergosterol



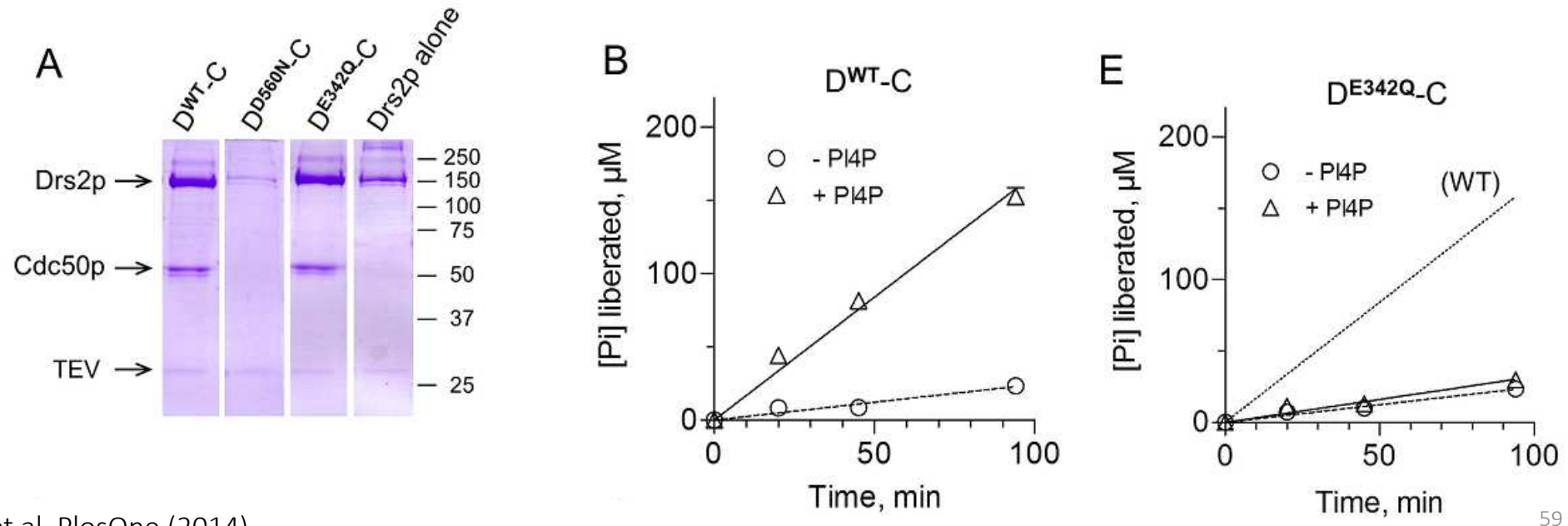
- Choice of an host that contains the required lipid
  - ✓ Cholesterol and SERT receptor expression (see below, Magnani *et al.* (2004)).
  - ✓ The role of PI4P in Drs2/Cdc50 flippase activation (see below, Azouaoui et al. (2014))

- Exogenous lipids can be added during extraction and purification

Phosphatidyl Inositol-4-Phosphate is essential for activation of the yeast Drs2/Cdc50 Phosphatidyl-Serine transporter

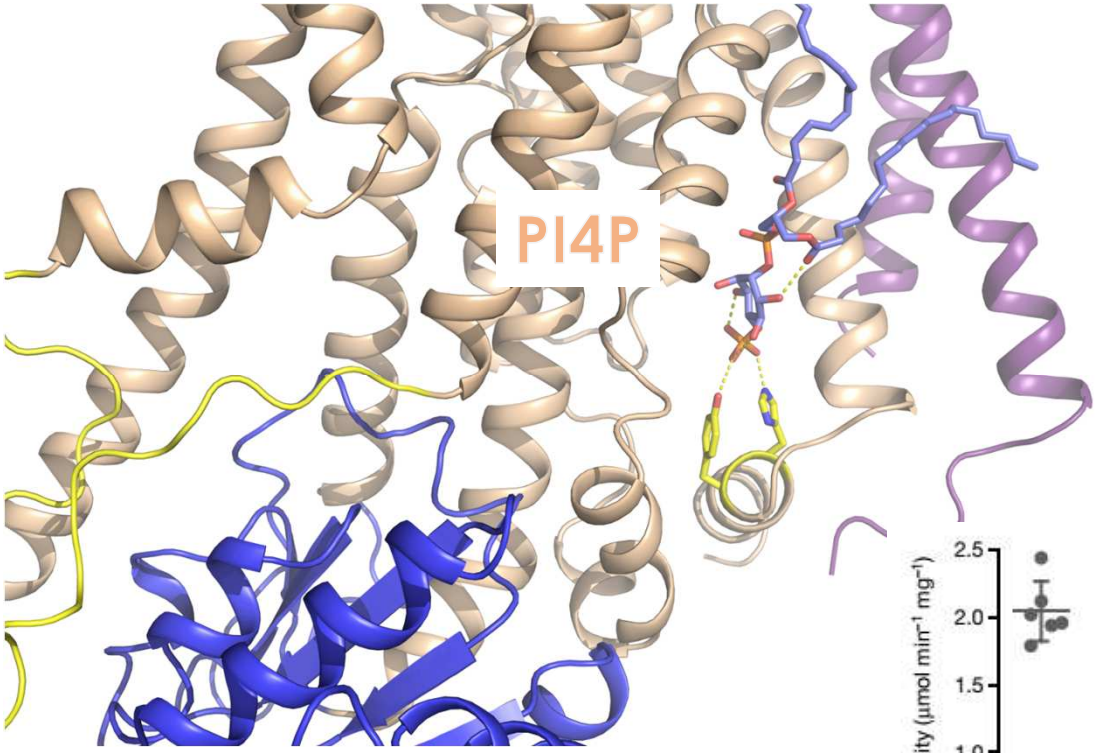
Expression : homologous expression in *S. cerevisiae*, GAL10/CYC1 inducible strong promoter, URA3 and ADE2 auxotrophie selection markers, growth at 28°C then expression at 18°C

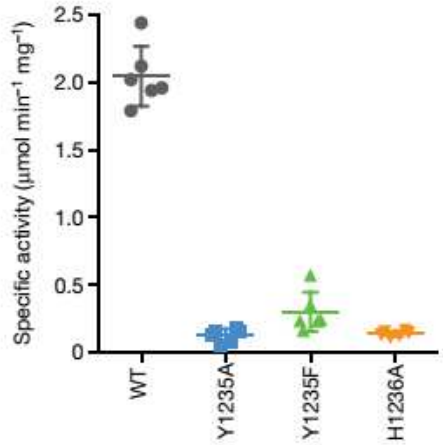
Purification : N-ter biotinylation domaine *on Drs2*, Streptavidin Sepharose affinity resin, elution by TEV proteolysison the resin

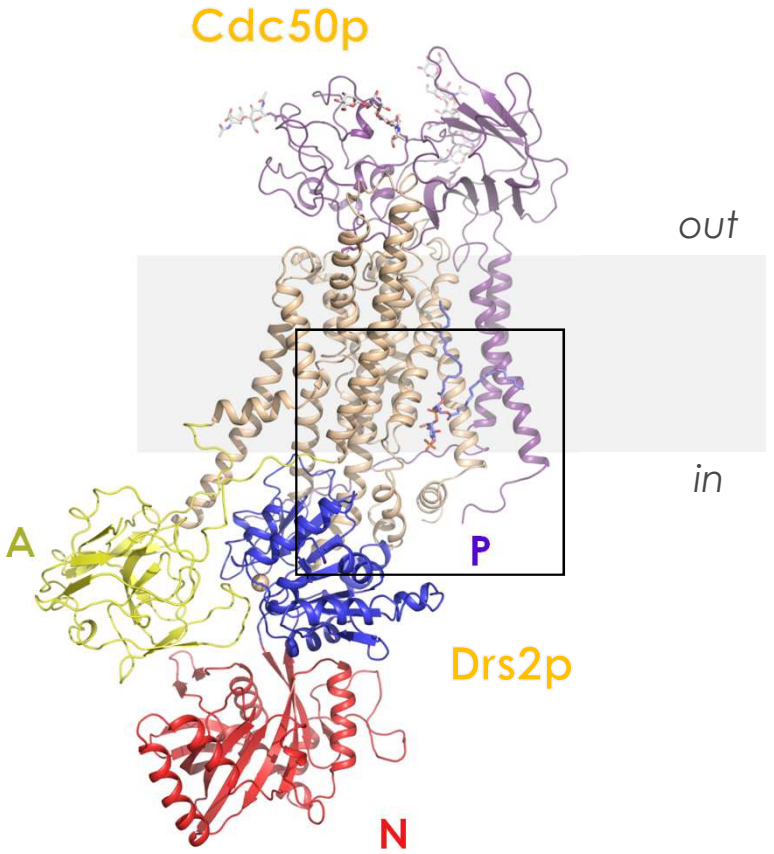


Azouaoui et al. PlosOne (2014)

Localization of the regulatory PI4P binding site







Timcenko et al. (2019) Nature

Nature and composition of membranes

- In procaryotes, only one membrane available, the plasma membrane
- In eucaryotes, all the organelles are available but peptide synthesis machinery can be overwhelmed leading to lower yield of expression.

1. Limited amount of membrane to store neo-synthesized proteins

- In procaryotes, protein are therefore stored as inclusion bodies (IB)
  - extraction of IB and in vitro refolding of MPs, especially β-barrel folded MPs

- Activation of recycling pathways (ERAD, UPR : Griffith et al. (2003)).
  - Lower the amount of neo-synthetize protein to improve the quality (see above)
  - Use of strains with less endogenous proteases
    - *E. coli* Bl21
    - *S. cerevisiae* BJ5457 : pep4Δ, prb1Δ (principales aspartyl- et sérine protéases)
  - **Selection of strains resistant to PM**

Toxicity of recombinant proteins
for host organisms

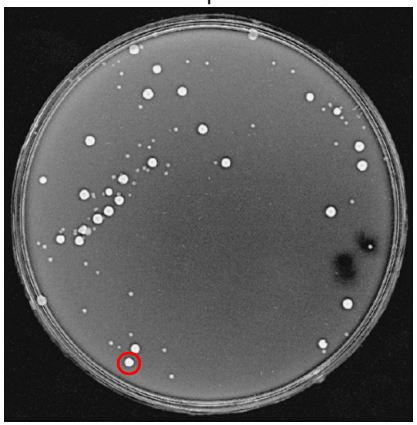
1 - optimisation/selection of resistant clones

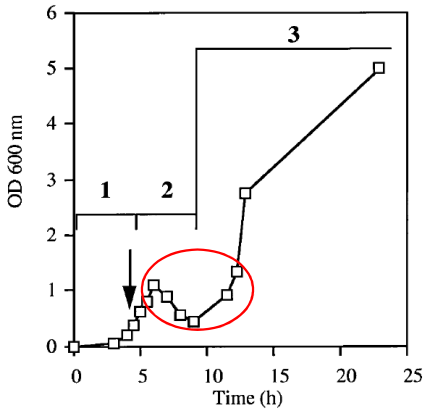
2 – protein engineering

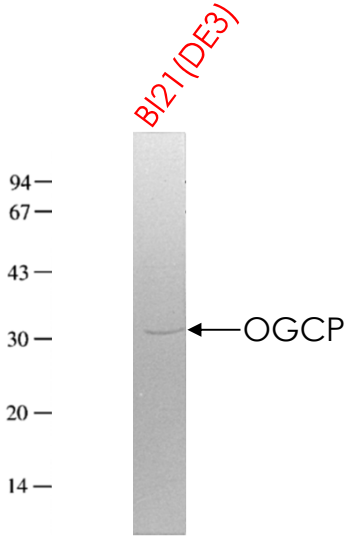
1 - Genetic selection of an *E. coli* clone resistant to expression of mitochondrial oxoglutarate-malate translocase (OGCP)

milieu LB + Amp

BI21(DE3)

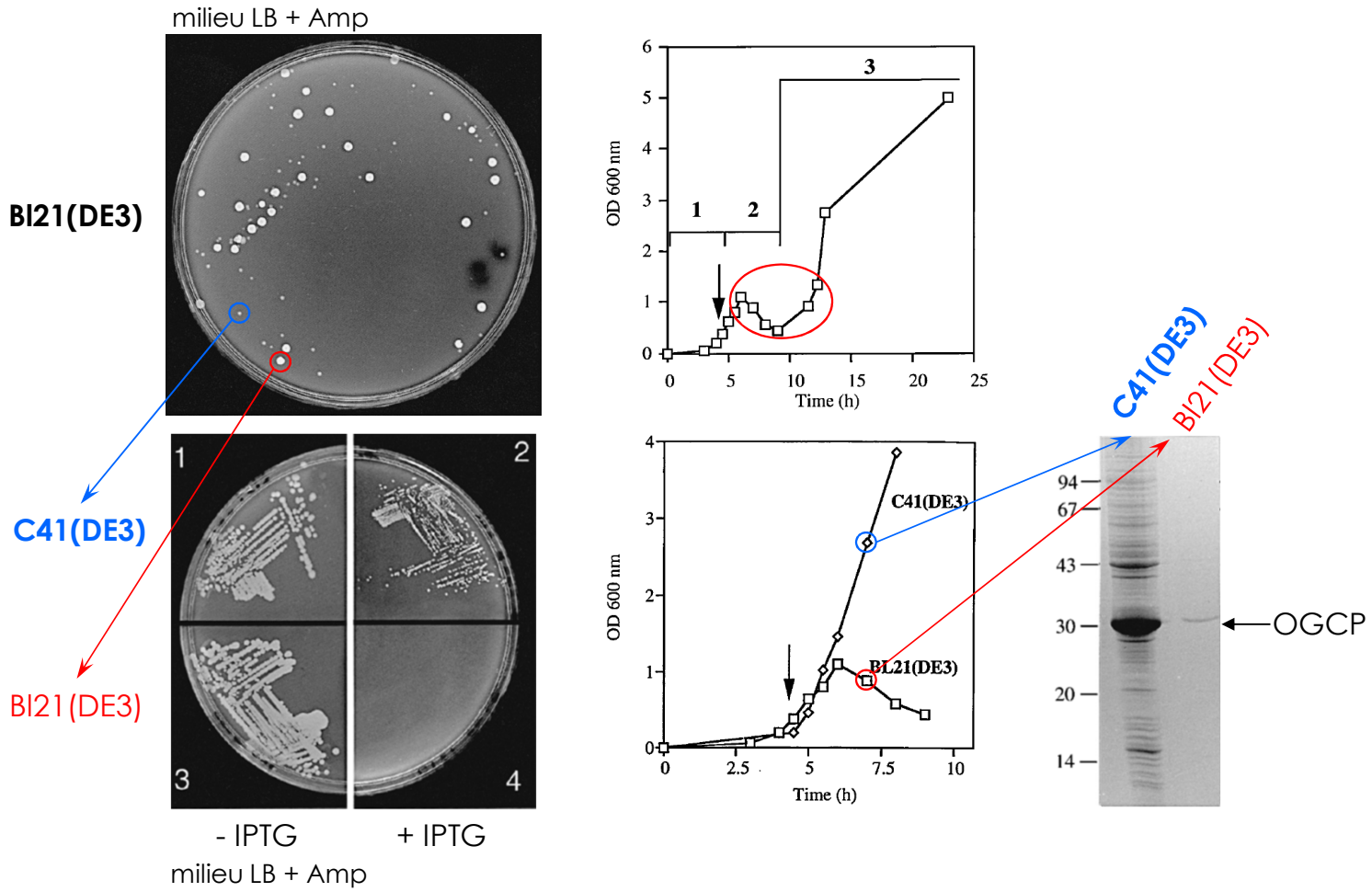






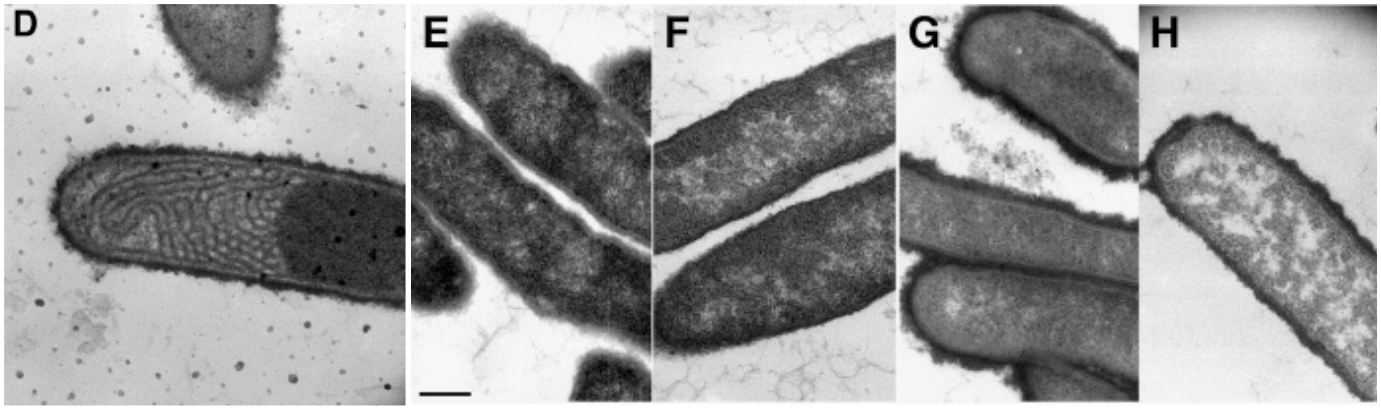
Miroux et al. J. Mol. Biol. (1996)

1 - Genetic selection of an E. coli clone resistant to expression of mitochondrial oxoglutarate-malate translocase (OGCP)



Miroux et al. J. Mol. Biol. (1996)

1 - Genetic selection of an *E. coli* clone resistant to expression of mitochondrial oxoglutarate-malate translocase (OGCP)



- Low amount of IB (Miroux *et al.* (1996) J Mol Biol)
- Development of **an internal network of membranes** from the plasma membrane(Arechaga et al. (2000) FEBS Letter)
- Expression is toxic in ***E. coli*** **Bl21(DE3) parental cells which acquire resistance to Ampicilline and get rid of the expression cassette** (Dumon-Seignovert *et al.* (2004) Prot Exp Purif)
- Spontaneous mutations appears in the T7 RNA polymerase resulting in a lower translation yield (Angius et al. (2018) Sci Rep)

2 – Protein engineering

- Addition fusion partners (see above)
- Mutagenesis
  - Site-directed or random mutagenesis to improve stability of the recombinant protein (Martinez-Molina *et al.* Prot. Sc (2008) ; Vaidehi Trends in Pharm. Sc. (2015))
  - Expression of protein domains
  - Deletion of protein domains involved in aggregation process (hydrophobic patches)
  - Deletion of flexible domain to improve stability and structural homogeneity (e.g cristallisation of RCPG)

2 – Protein engineering

How Can Mutations Thermostabilize G-Protein-Coupled Receptors?

Nagarajan Vaidehi,[1,*] Reinhard Grisshammer,[2] and Christopher G. Tate[3]

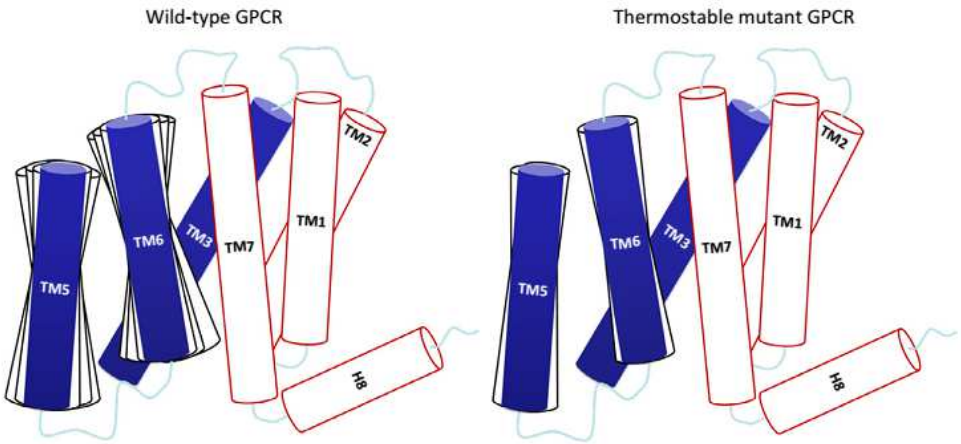


Table 1. Crystal Structures Determined of Thermostabilized GPCRs

Receptor	Mutant Name	Stabilized Conformation	Number of Mutations	G Protein Activation	Thermo-stabilization Methodology	Refs
β_1-Adrenergic receptor (β_1AR)	m23	Inactive	6	Yes	SM[a]	[15]
Adenosine A_{2A} receptor (A_{2A}R)	StaR2	Inactive	8	No	SM	[29]
	GL31	Active-like	4	Weak	SM	[51]
Neurotensin receptor (NTSR1)	GW5	Active-like	6	No	SM	[31]
	ELF	Active-like	3	Yes	SM	[32]
	TM86V	Agonist-bound, inactive	11	Weak	Directed evolution	[26]
Chemokine receptor (CCR5)	–	Inactive	4	No	Rational design	[52]
Free fatty acid receptor (FFA1R)	–	Inactive	4	Weak	SM	[53]
Corticotrophin releasing factor receptor (CRF1R)	–	Inactive	12	ND[a]	SM	[54]
Metabotropic glutamate receptor (mGlu5)	–	Inactive	6	ND	SM	[55]

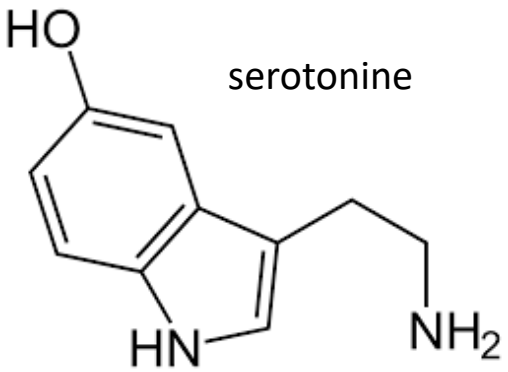
[a]Scanning mutagenesis; ND, not determined.

Vaidehi Trends in Pharm. Sc. (2015)

Case study

The Serotonine transporter (SERT)

The Serotonine transporter (SERT)



serotonine

- Physiology
  - Transport of serotonine, an hormone derivated from tryptophan
  - Na^+ dépendent transport (simport)
  - Mood management (happiness, stress, anxiety, phobias, depression), survival instinct, circadian rhythm
  - Target for amphetamines, ecstasy and cocaïne

- Predicted structure
  - MW ~ 64 kDa (630 aa)
  - 12 transmembrane spans
  - 2 sites of N-glycosylation
  - Role of cholesterol ?
  - Oligomerisation ?

Which strategy to choose for production of SERT?

SERT - 1 - Diversifying approaches

E.coli

Yeast (*Pichia pastoris)*

Baculovirus (BacMan/HEK293)

Mammalian cells

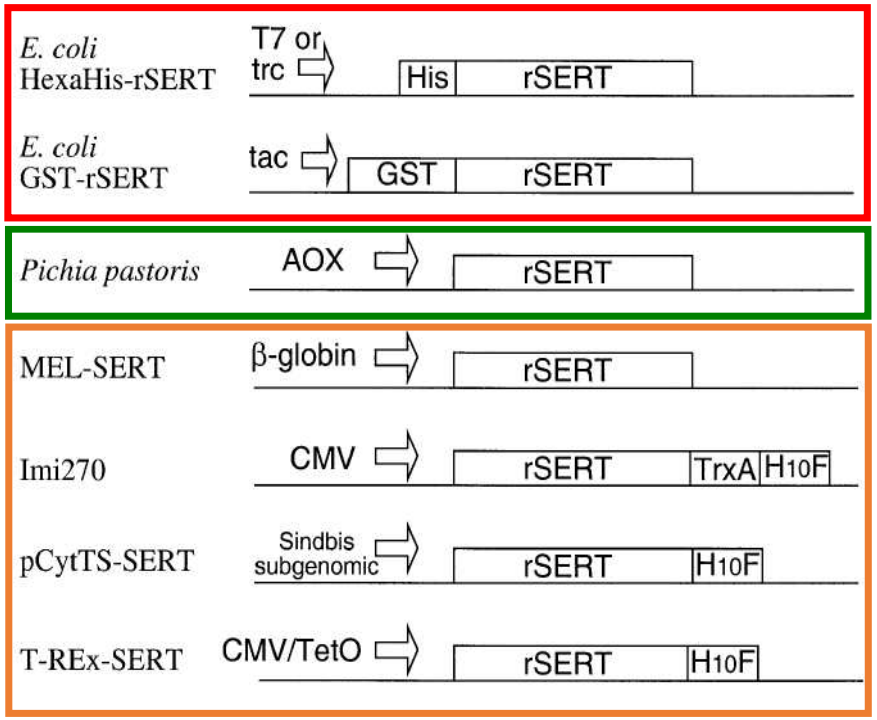
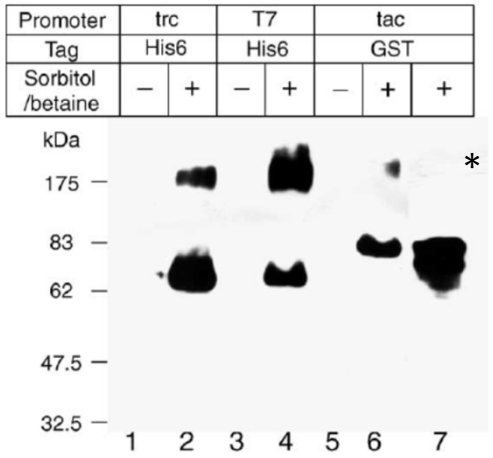


Fig. 1. Schematic diagram of SERT cDNAs expressed in each expression system. Boxes (not to scale) represent open reading frames encoding the rat serotonin transporter (rSERT) and various tags (His, hexa-histidine tag; GST, glutathione *S*-transferase; TrxA, thioredoxin; $H_{10}F$, deca-histidine tag followed by a FLAG tag). The arrows represent the promoter used to expressed SERT in each system. See Section 2 for further details.

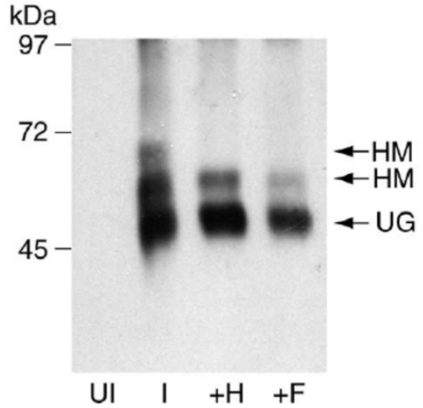
Tate et al. J. Biol. Chem (1999), Tate et al. BBA (2003)

SERT - 1 - Diversifying approaches

a. *E. coli*



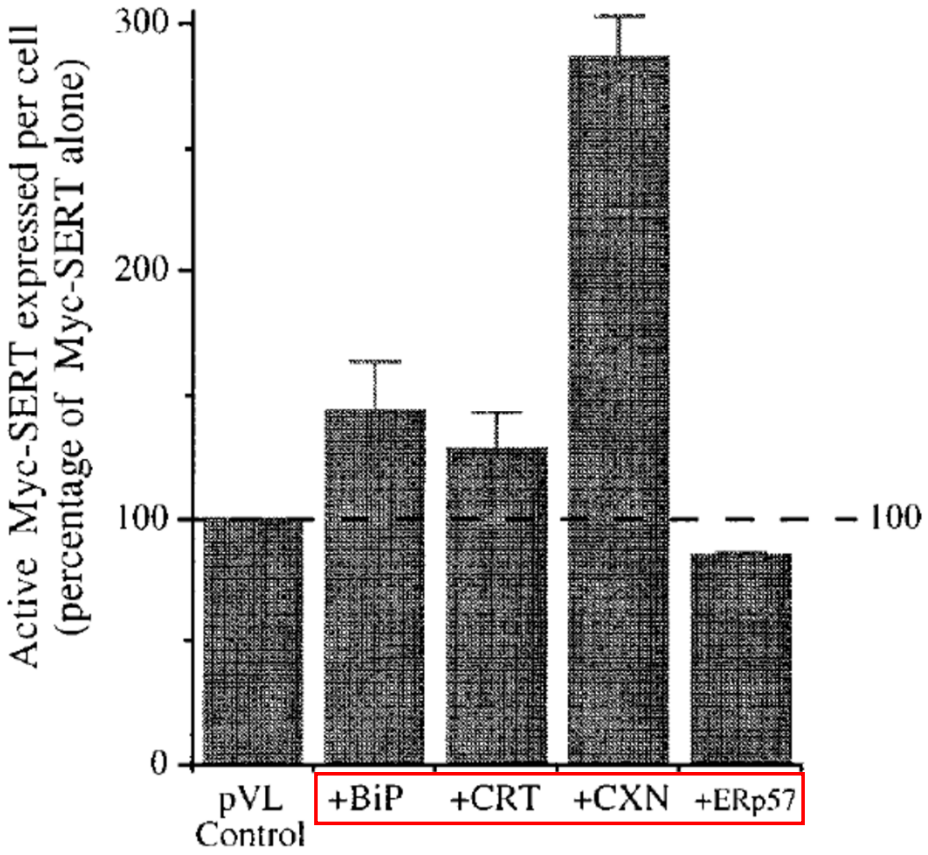
b. *Pichia pastoris*



- *E. coli*
  - 2-3 mg/L
  - Aggregate (*)
  - No activity !
- P. pastoris
  - 2-3 mg/L
  - Glycosylated forms
  - Glycosydases do not remove all sugars = misfolding ?
  - No activity !

Tate et al. BBA (2003)

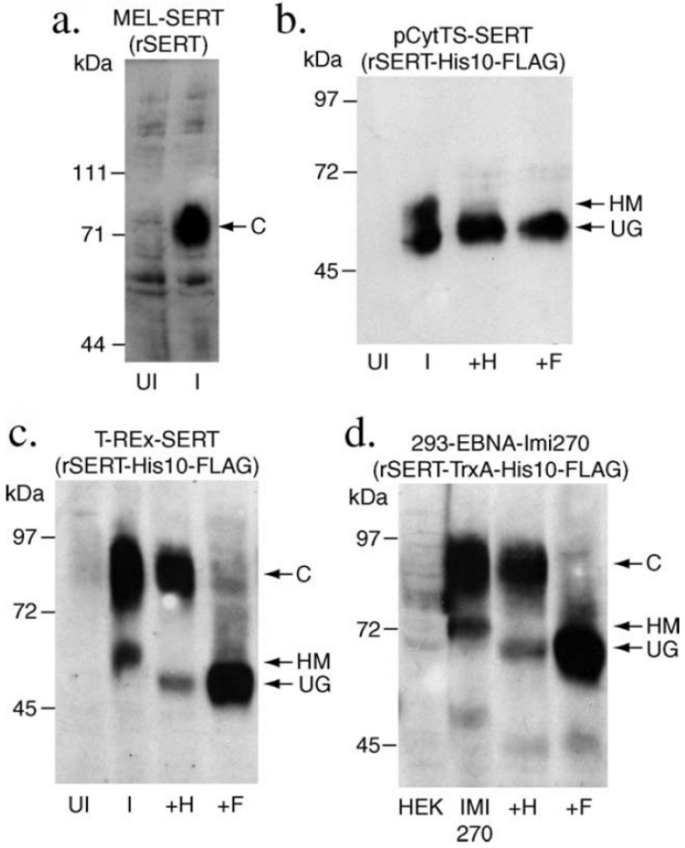
SERT - 2 - Coexpression of chaperones

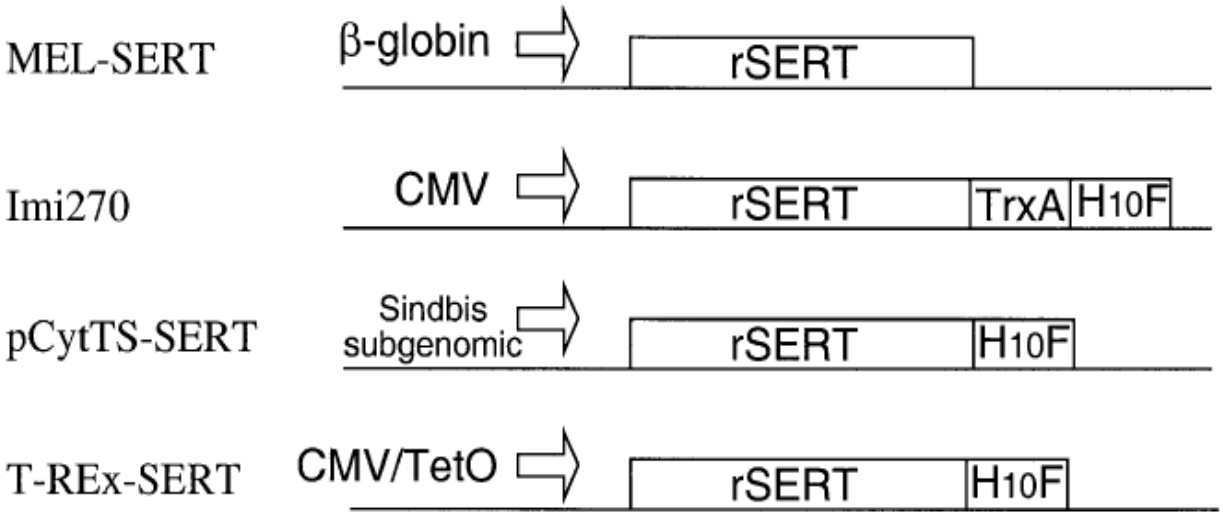


☑Co-expression of **calnexin (CXN)** in baculovirus increases the proportion of active protein by a factor of 3!

Tate et al. J. Biol. Chem (1999)

SERT - 3 – Expression in mammalian cells

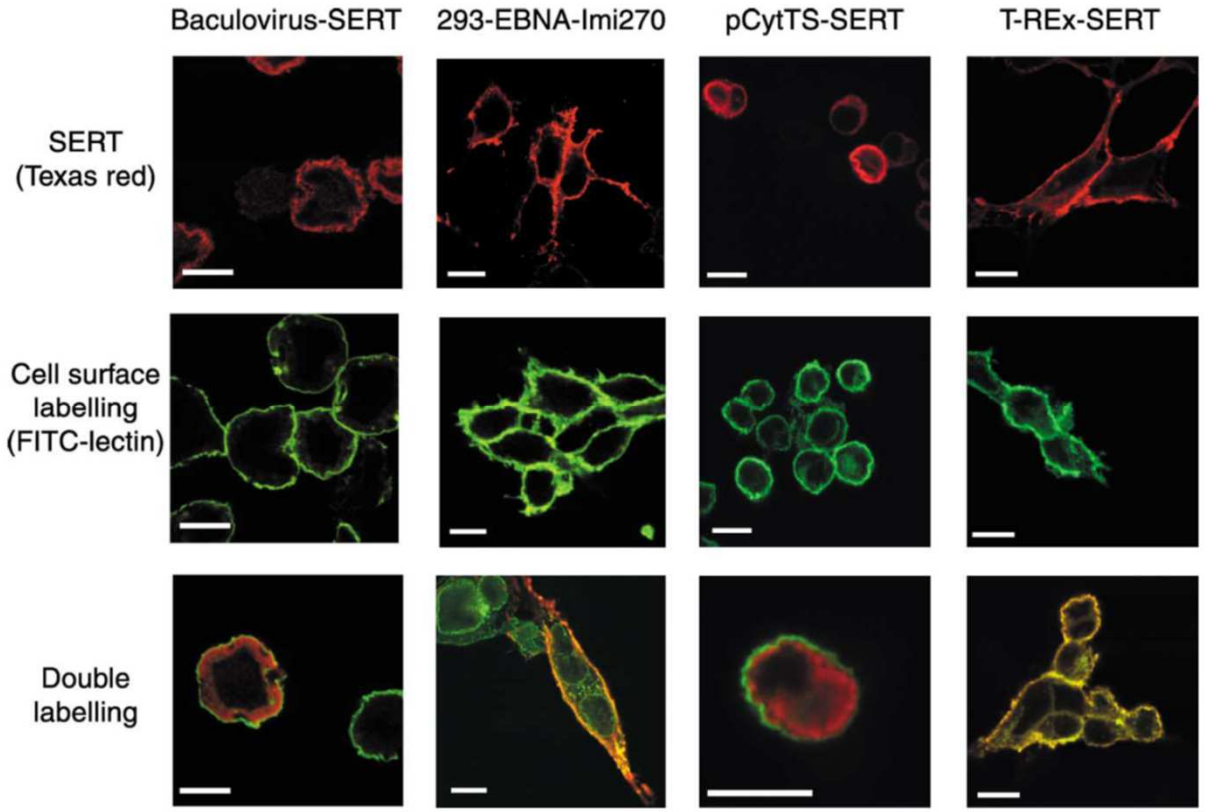




☺ The four systems allow the expression of glycosylated forms of the protein

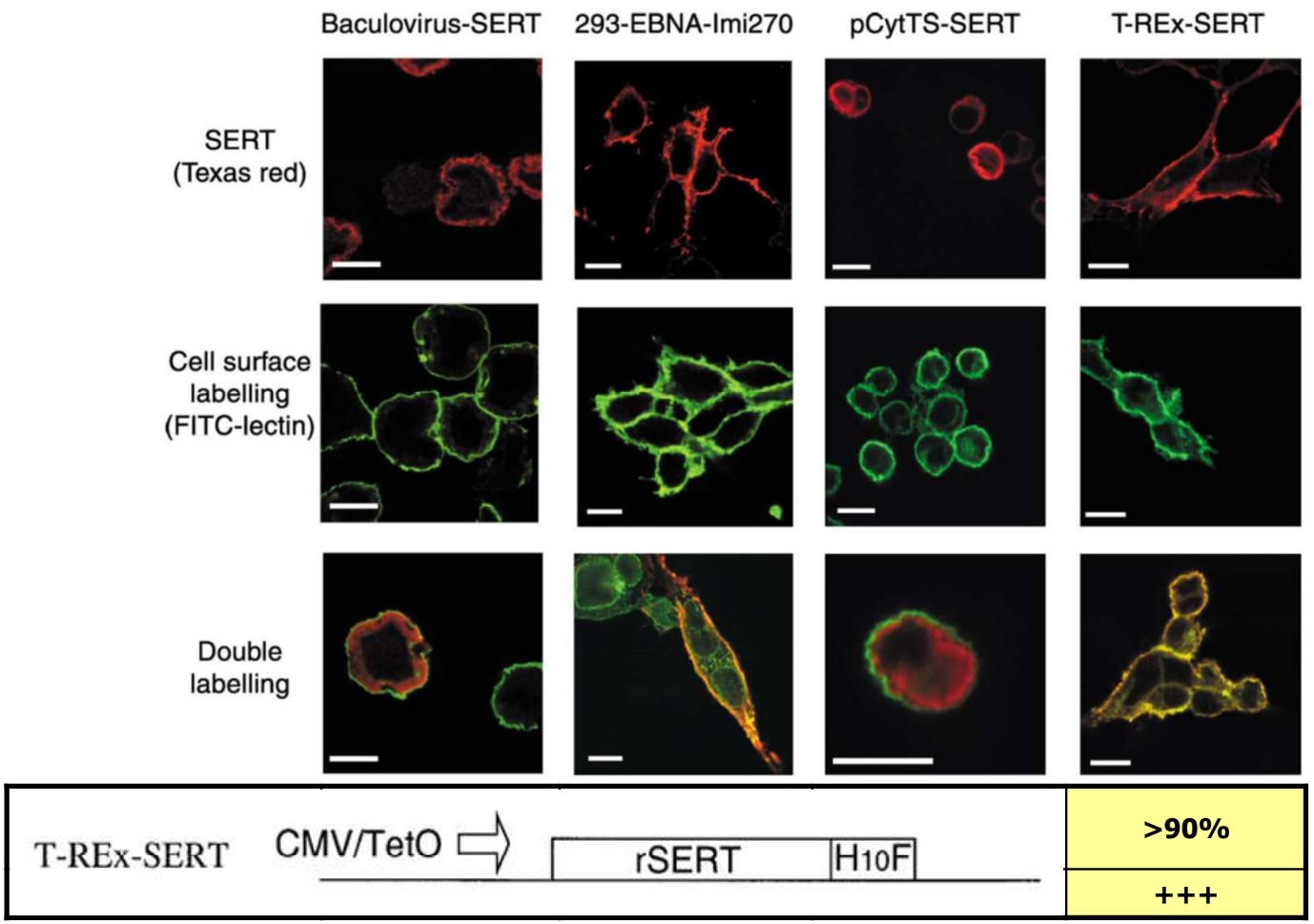
UI, undinduced; I, induced; +H, endoH treated; +F, PNGaseF treated;
UG, unglycosylated form; HM, High-mannose core N-glycans;
C, complex N-glycans

SERT - 4 – Subcellular localization of recombinant protein by confocal microscopy

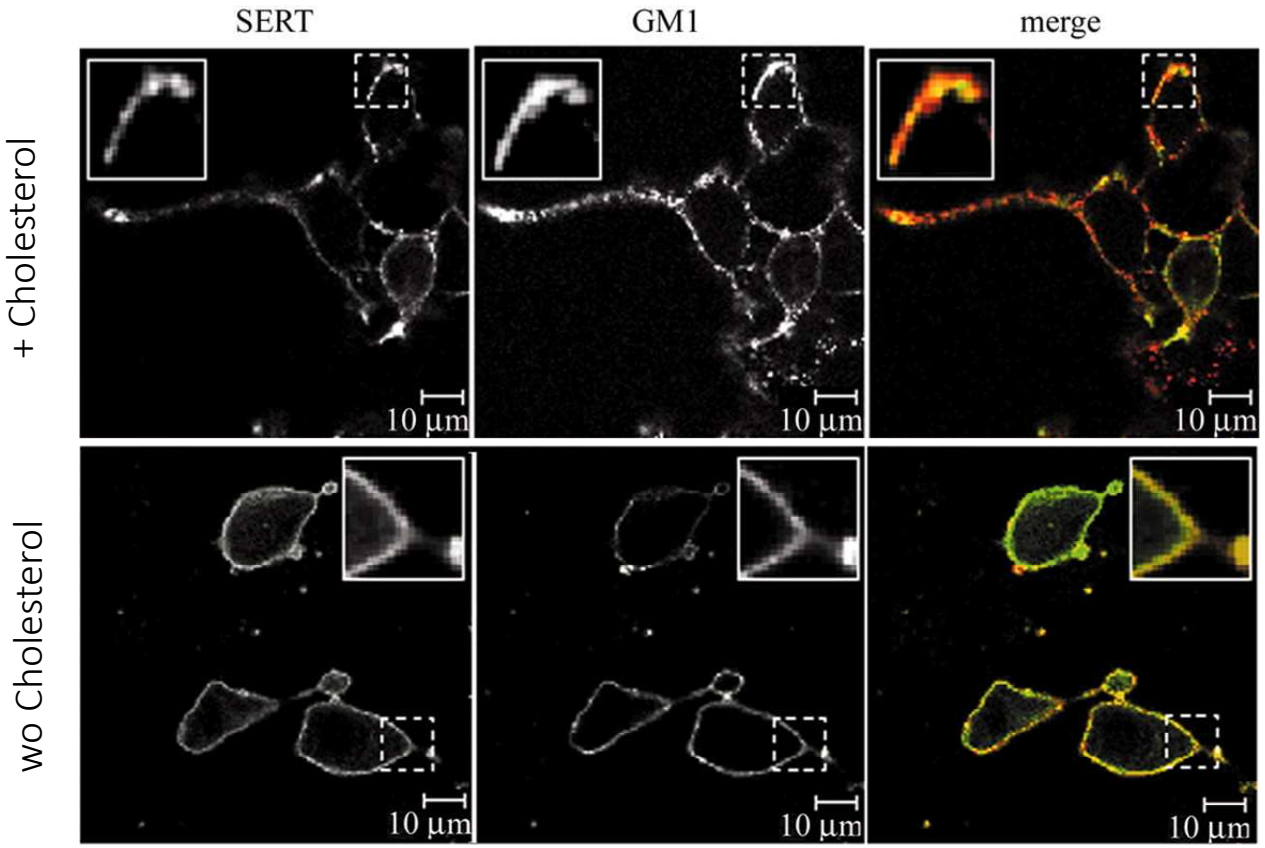


Quantité exprimée en membranes	<10%	30%	<20%	**>90%**
Activité spécifique	+	**+++**	+	**+++**

SERT - 4 – Subcellular localization of recombinant protein by confocal microscopy



SERT - 5 - Nature of lipids & role in addressing



SERT co-localises with GM1 plasma membrane sphingolipid
only in the presence of cholesterol

Magnani et al. J. Biol. Chem (2004)

Expression of SERT transporter – Sum up

- Post-translational modifications: N-glycosylations are essential for functionality
- Cholestérol is essential for addressing and functionality

➢ Expression in mammalian cells (T-Rex system)

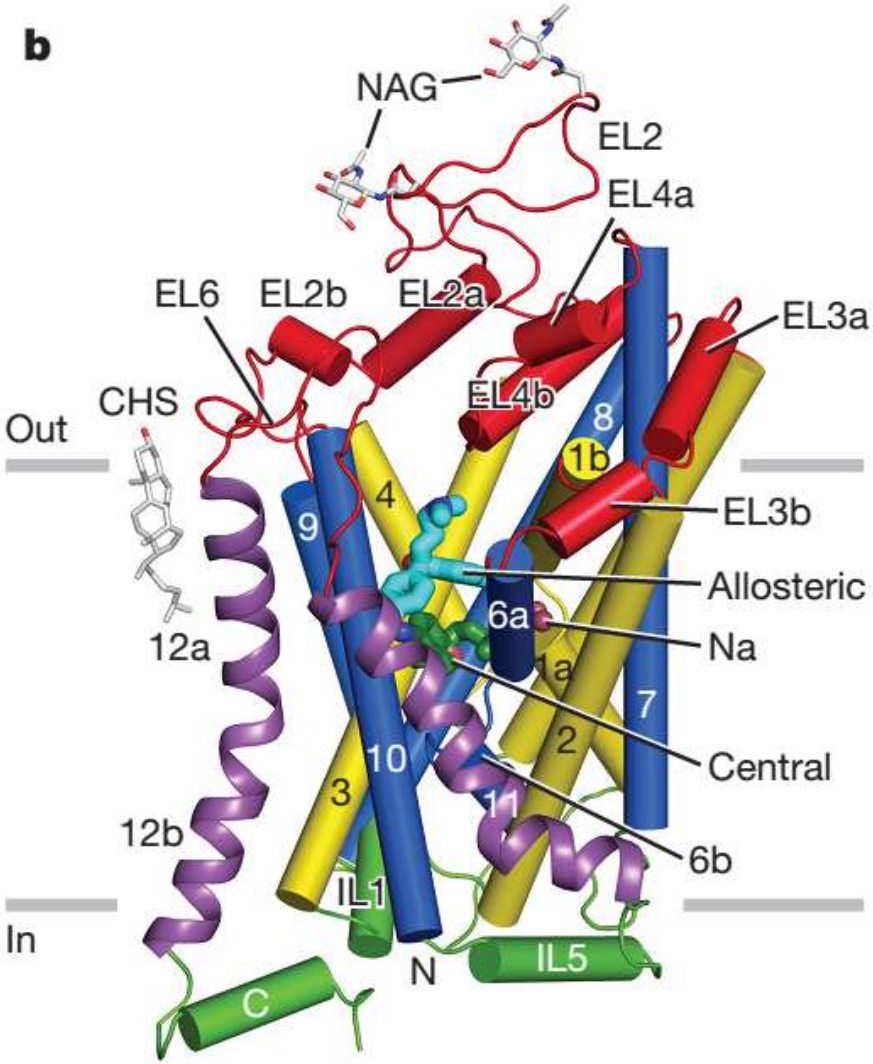
✓ Strong inducible hybrid promoter (CMV+Tet-ON),

✓ Selection of stable cell lines (G418),

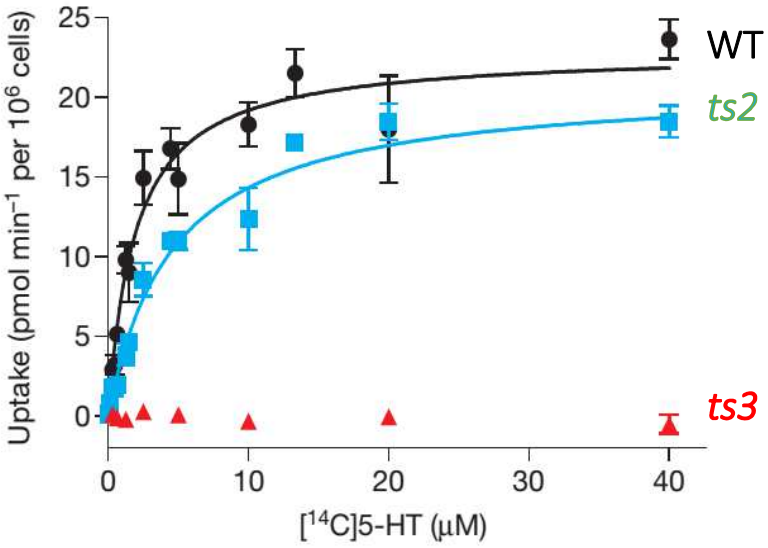
✓ A few inactive proteins expressed (confocal microscopy),

✓ Cells in suspension for culture in bioreactors to improve the yield of production.

X-ray structure of the human serotonin transporter



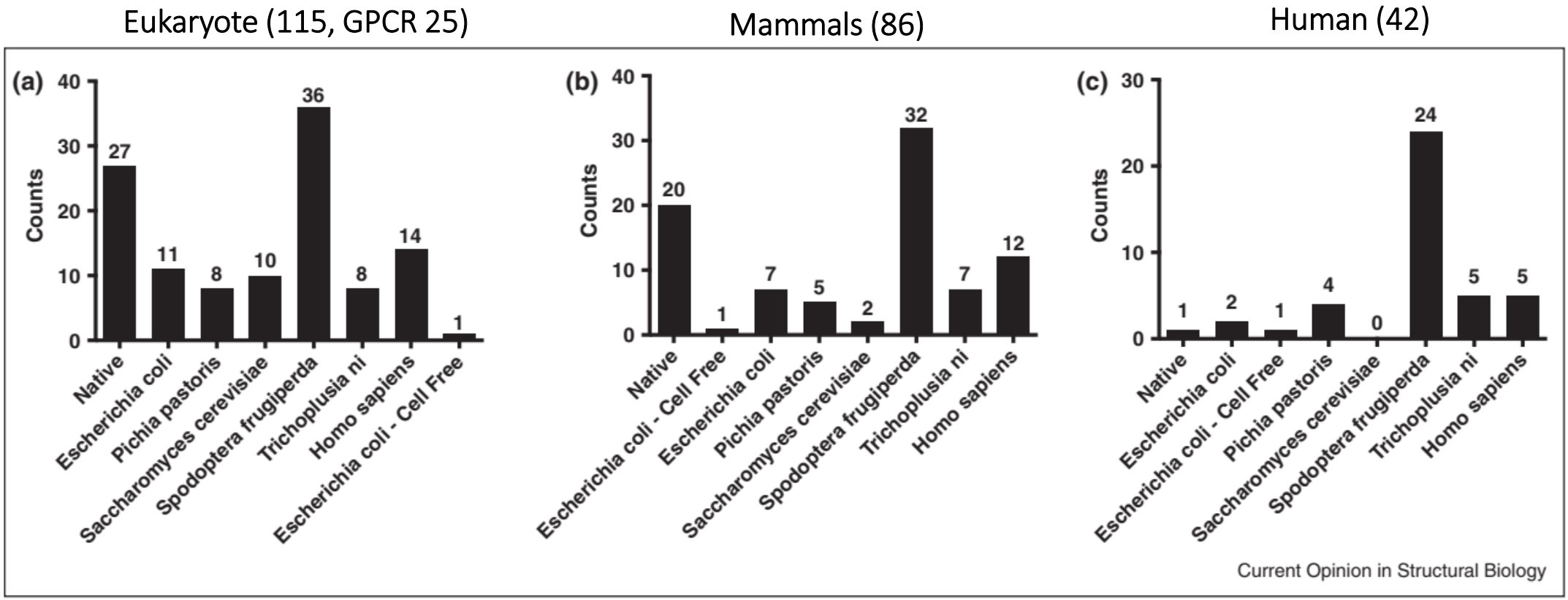
- BacMan/HEK293 cells
- DoDécylMaltoside + Cholesteryl HemiSuccinate (CHS)
- Thermostabilization (*2* ou *3* Ala) / FSEC
- Strep-Tactin affinity chromatography 2 cystéines → Ala
- S-citalopram ou paroxetine



Coleman et al. Nature (2016) 532(7599):334-9.

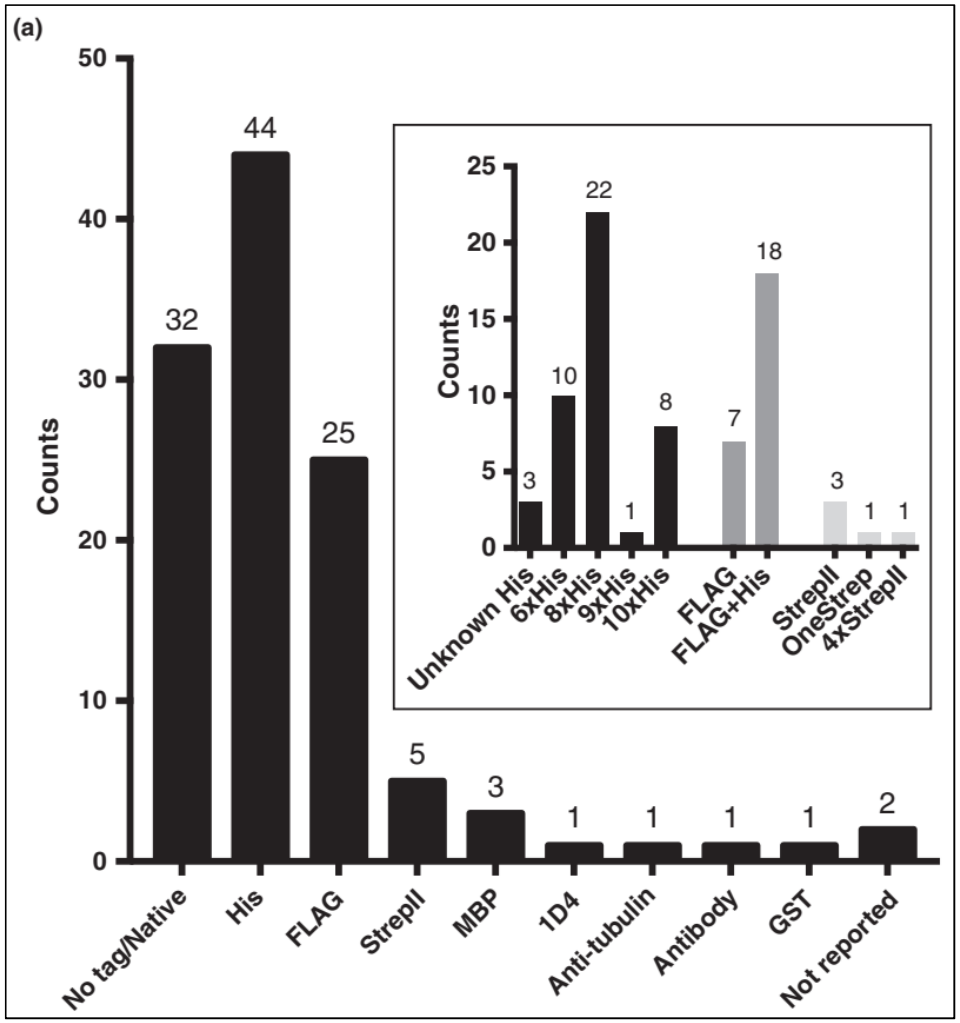
… to conclude

Expression strategies for structural studies of eukaryotic membrane proteins



Lyons et al. Curr Op Struct Biol (2016)

Expression strategies for structural studies of eukaryotic membrane proteins



Lyons et al. Curr Op Struct Biol (2016)


TIME FOR
QUESTIONS