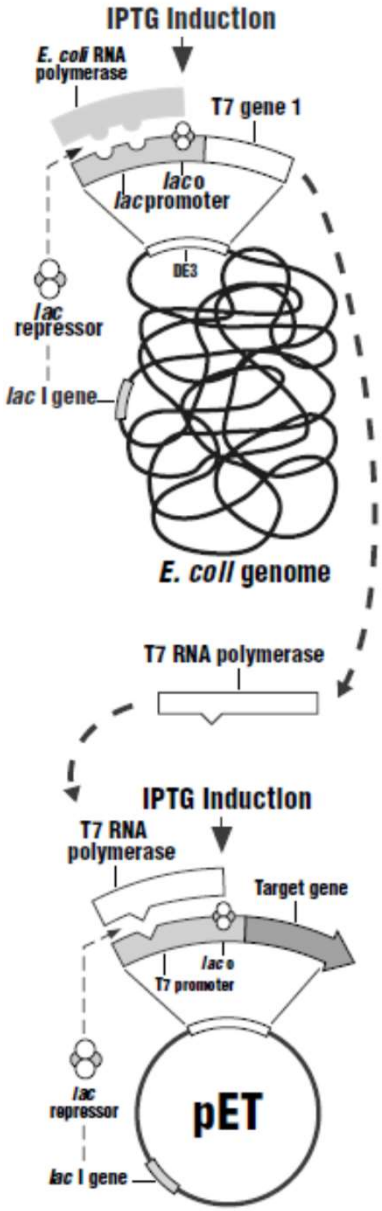
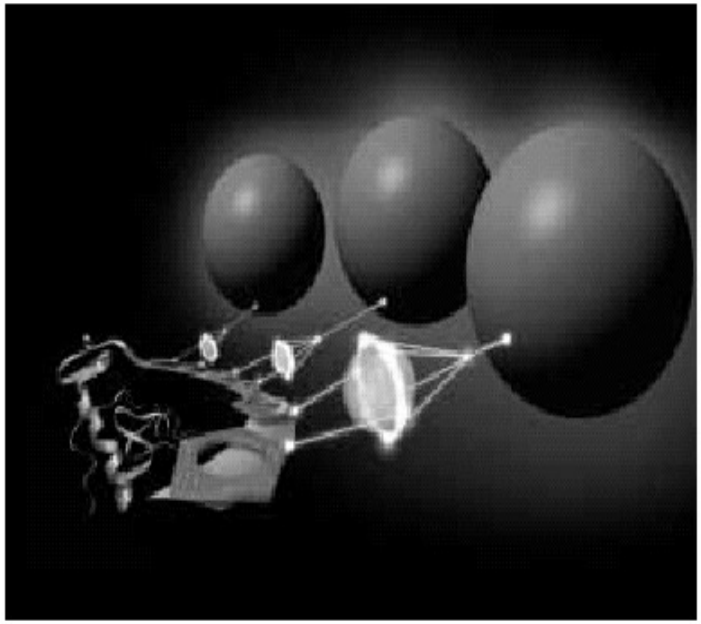


Recombinant protein expression



Recombinant proteins

Recombinant DNA

- artificial DNA sequence created by combining two or more strands of DNA

Recombinant proteins

- proteins obtained by introducing recombinant DNA into heterologous host, where the expression of this DNA occurs

Taking advantage of recombinant proteins

Overexpression and purification of recombinant proteins are essential prerequisite for

- **biochemical characterization of protein function** (determination of kinetic parameters K_m, k_{cat} for enzymes with their substrates, K_i for enzymes with inhibitors and K_d for protein-protein or protein-ligand interactions).

- **analysis of protein structure** (NMR, crystalography).

- **protein engineering (**improvement of protein quality – activity, stability).

- **on an industrial scale there are produced drugs, vaccines and dietary supplements.**

The goal of recombinant technology:

High yield of homogeneous proteins (mg – kg of proteins)

Maintenance of biological activity of proteins

Why to produce recombinant proteins?

Natural source:

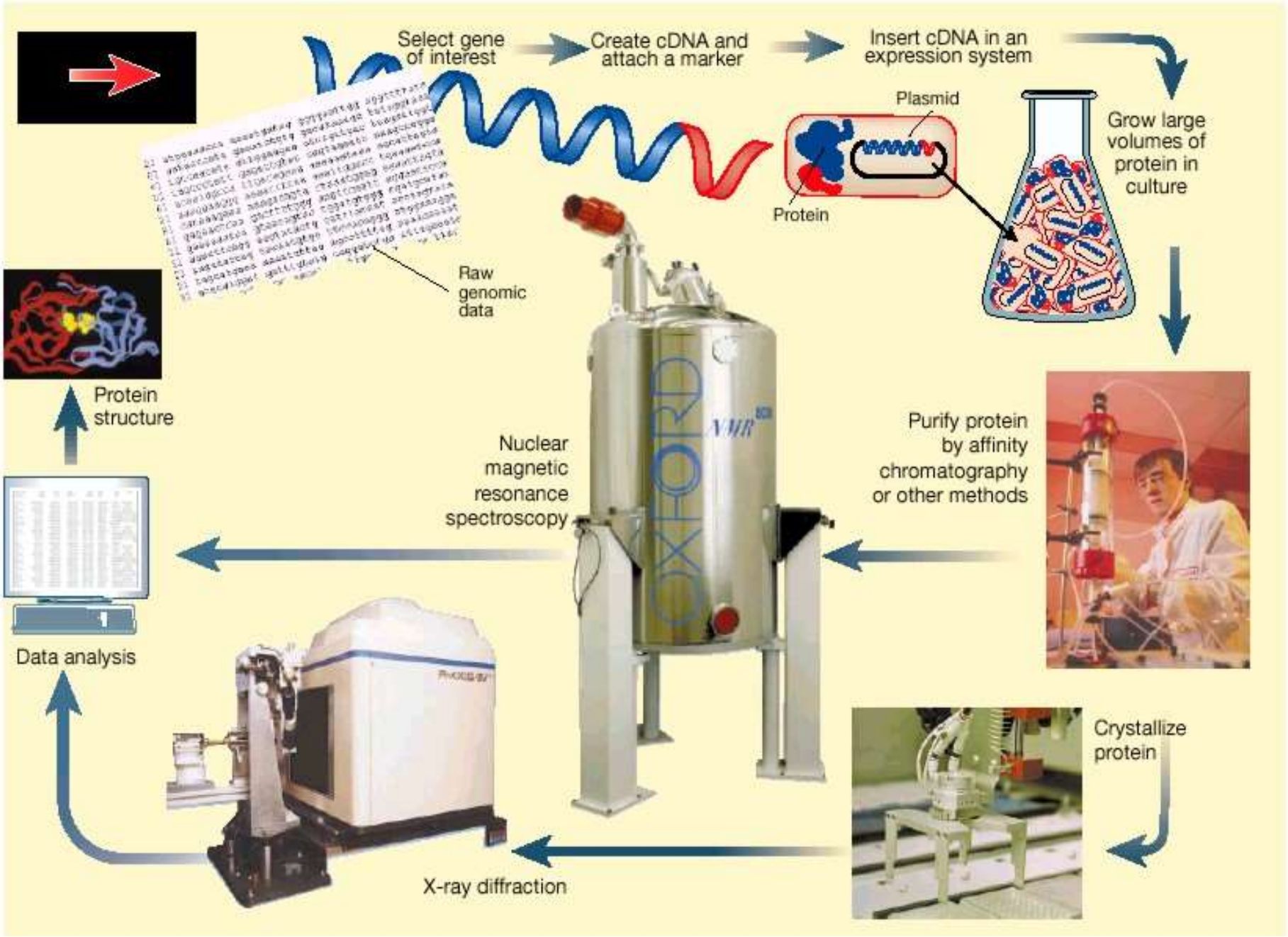
- Difficult to obtain (tissues, organs).
- Difficult to cultivate (bacteria, virusses, tissue cultures).
- Limited expression
- In most cases tough purification of protein

TABLE 1.2. Examples of low-abundance proteins and peptides isolated from natural biological sources

Protein	Source	Yield (µg)	Reference
Multipotential colony-stimulating factor	pokeweed mitogen-stimulated mouse spleen-cell-conditioned medium (10 liters)	1	Cutler et al. (1985)
Human A33 antigen	human colon cancer cell lines (10^{10} cells)	2.5	Catimel et al. (1996)
Platelet-derived growth factor (PDGF)	human serum (200 liters)	180	Heldin et al. (1981)
Granulocyte-colony-stimulating growth factor (G-CSF)	mouse lung-conditioned medium (3 liters)	40	Nicola et al. (1983)
Granulocyte-macrophage colony-stimulating growth factor (GM-CSF)	mouse lung-conditioned medium (3 liters)	12	Burgess et al. (1986)
Coelenterate morphogen	sea anemone (200 kg)	20	Schaller and Bodenmuller (1981)
Peptide YY (PYY)	porcine intestine (4000 kg)	600	Tatemoto (1982)
Tumor necrosis factor (TNF)	HL60 tissue culture medium (18 liters)	20	Wang and Creasy (1985)
Murine transferrin receptor	NS-1 myeloma cells (10^{10} cells)	20	van Driel et al. (1984)
Fibroblast growth factor (FGF)	bovine brain (4 kg)	33	Gospodarowicz et al. (1984)
Transforming growth factor-β (TGF-β)	human placenta (8.8 kg)	47	Frolik et al. (1983)
Human interferon	human leukocyte-conditioned medium (10 liters)	21	Rubinstein et al. (1979)
Muscarinic acetylcholine receptor	porcine cerebrum (600 g)	6	Haga and Haga (1985)
β_2-adrenergic receptor	rat liver (400 g)	2	Graziano et al. (1985)

Adapted, with permission, from Simpson and Nice (1989).

Scheme of recombinant protein technology



Host organisms for recombinant protein expression

- Procaryotic expression systems (E.coli, Baccilus subtillis,…)
- Yeasts (Sacharomyces cerevisiae, Pichia Pastoris)
- Mammalia cells (human embryonic kidney cells- HEK, Chinese hamster ovary cells – CHO)
- Insect cells with baculovirusses
- Expression in vitro – Cell free (lysates from rabbit reticulocyte, extracts from E.coli, extracts from wheat germ)

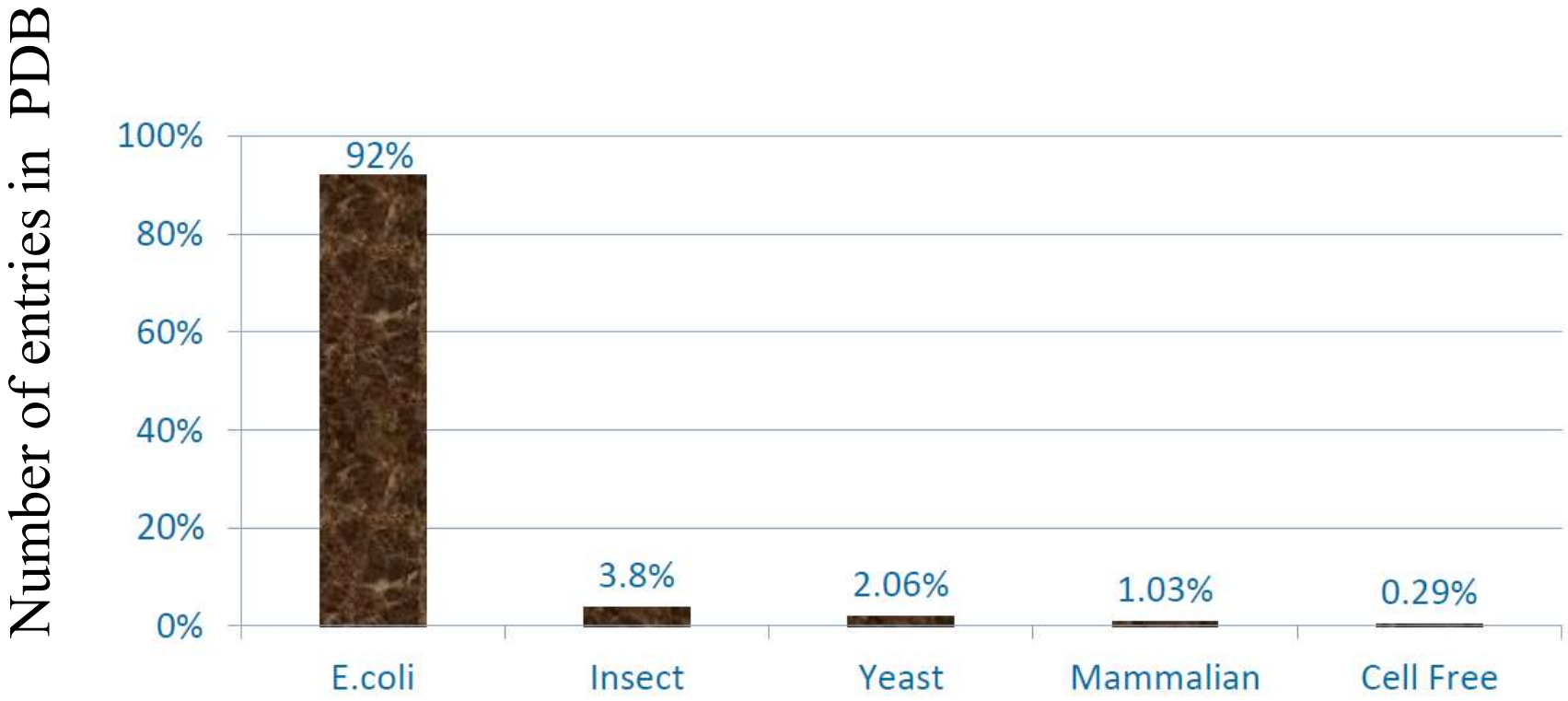
Criteria for selection of expression host organism

	Posttranslation modification					
Expression system	N-glycosylation	O-glycosylation	Phosphorylation	Acetylation	Acylation	Carboxylation
E.coli	missing	-	-	-	-	-
Yeasts	Highly mannosilated glycans	+	+	+	+	-
Insect cells	Simple, without sialization	+	+	+	+	-
Mammalian cells	complex	+	+	+	+	+

Criteria for selection of expression host organism

Expression system	Budget for cultivation	Growth rate	Level of expression	Protein conformation
E. coli	low	high	high	Often refolding needed
Yeasts	low	high	Low/ high	Sometimes refolding needed
Insect cells	high	low	Low / high	Mostly correct
Mammalian cells	high	low	Mostly low	Mostly correct

PDB statistics



Data from 2014

https://www.rcsb.org/stats/distribution-expression-organism-gene

Production of heterologous proteins in *E. coli*

Advantages :

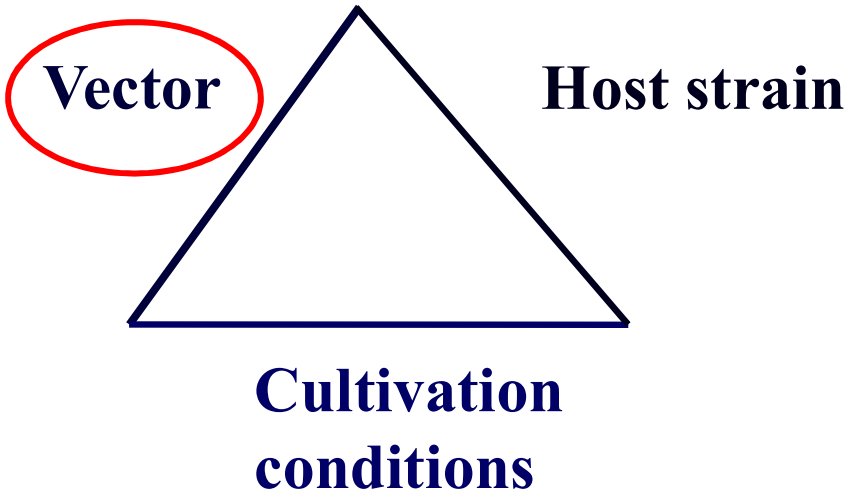
- High yield of recombinant proteins
- Well known genome and proteome -facilitation of gene manipulation
- Design of vectors facilitates cloning and expression of foreign genes
- Rapid growth in an inexpensive medium
- Adaptalibity of system

Production of heterologous proteins in *E. coli*

Disadvantages:

- Requirement of cDNA of target protein
- Absence of eukaryotic posttranslation systems (posttranslation modification)
- Formation of insoluble inclusion bodies
- Limited ability of disulfide bond formation
- Missing secretion mechanism for effective protein releasing into the cultivation medium
- Different codon usage compared to higher eucaryotes
- Contamination of protein by lipopolysacharids (endotoxin)

Expression system for recombinant protein production in *E.coli*

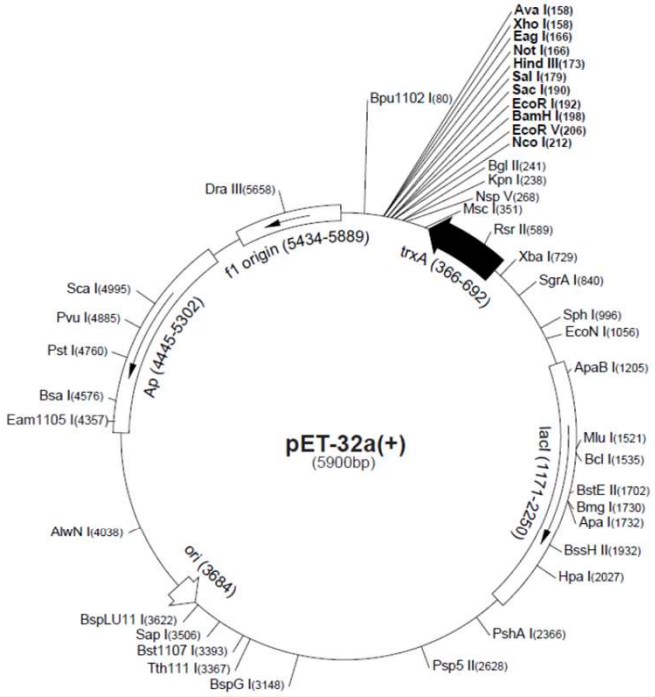


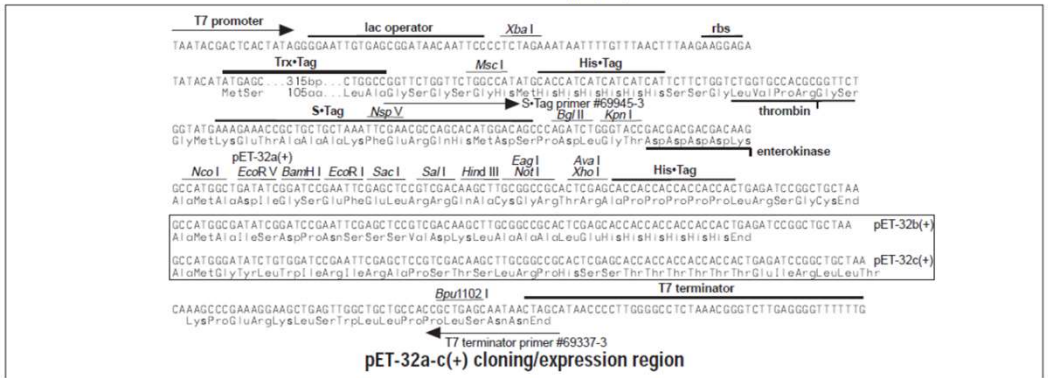
Expression vector

= cloning vector containing necessary regulation sequences for expression of foreign gene inserts

pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3684
bla coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.

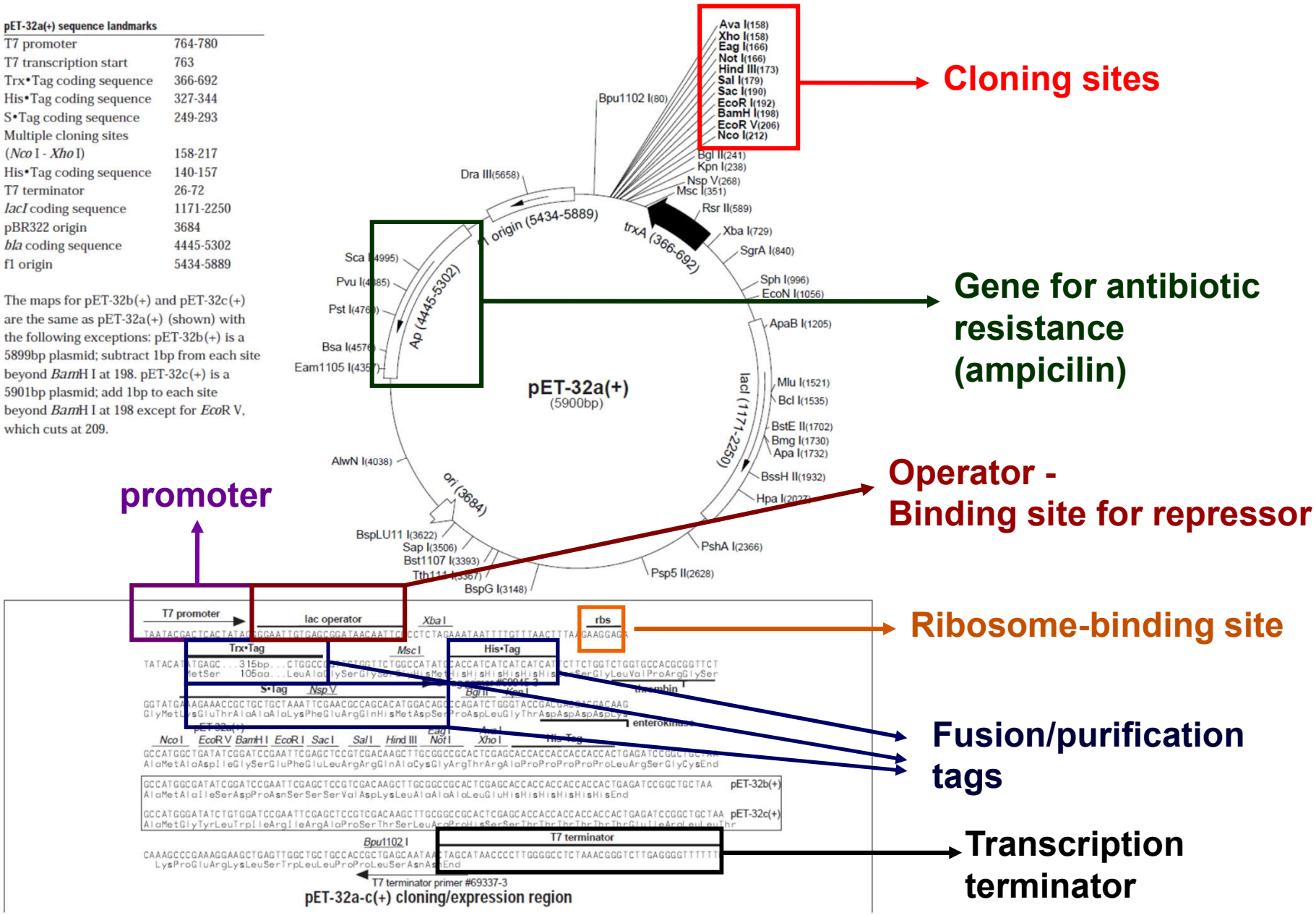




Expression vector configuration

pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3684
bla coding sequence	4445-5302
f1 origin	5434-5889

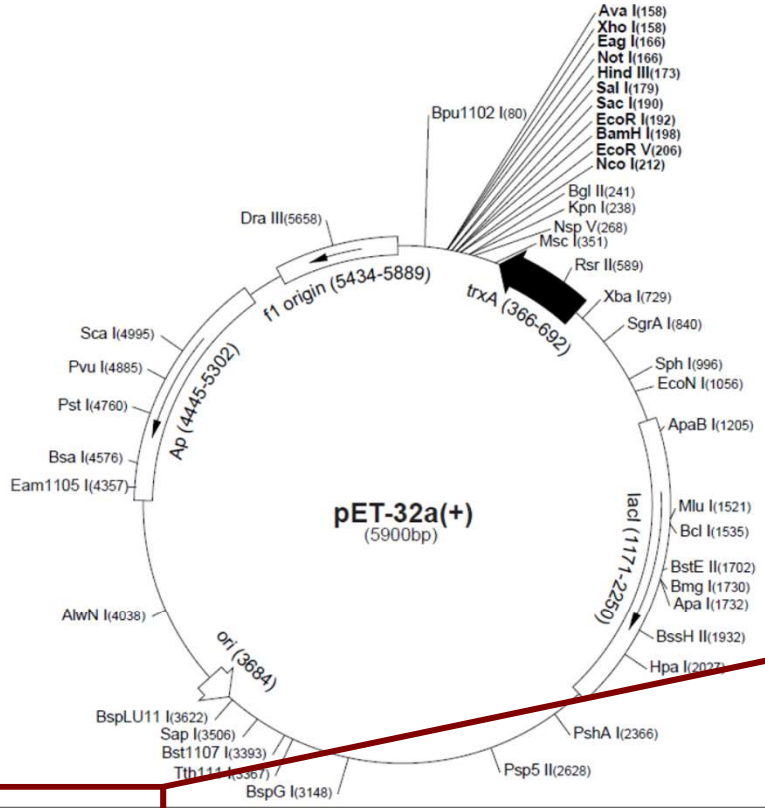
The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.



Expression vector configuration

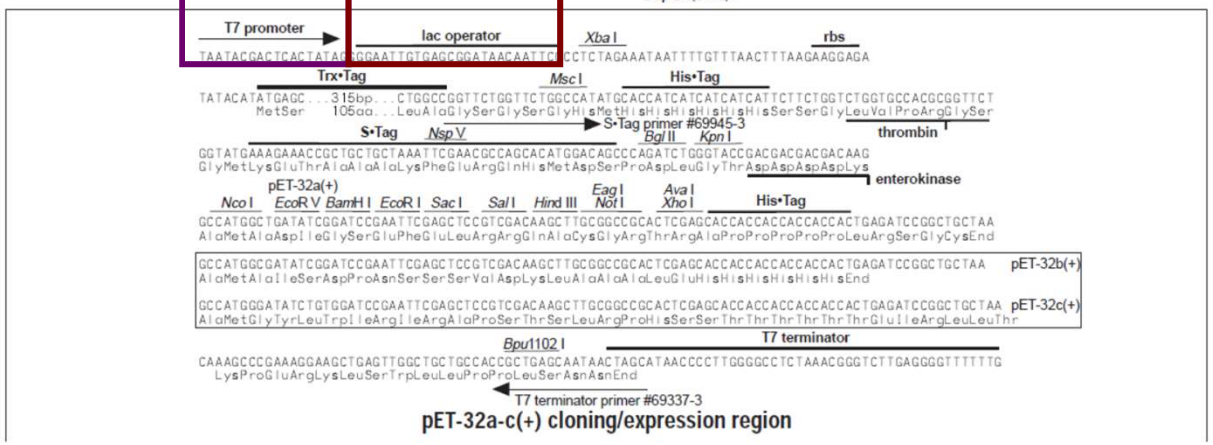
pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3684
bla coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.



promoter

Operator -
Binding site for repressor



pET-32a-c(+) cloning/expression region

Promoter characteristics:

- **Strong promoter** (ptac, ptrp, λpL, pT_7)

Target protein should result in accumulation of protein making up 10-30 % and more of the total cellular protein.

- **Easily transferable to various *E.coli* strains**
- **Simple and cost effective induction**

- Thermal induction (λpL)

- Chemical induction (ptac, ptrp, pT7): IPTG (isopropyl-β-D-thiogalactopyranosid)

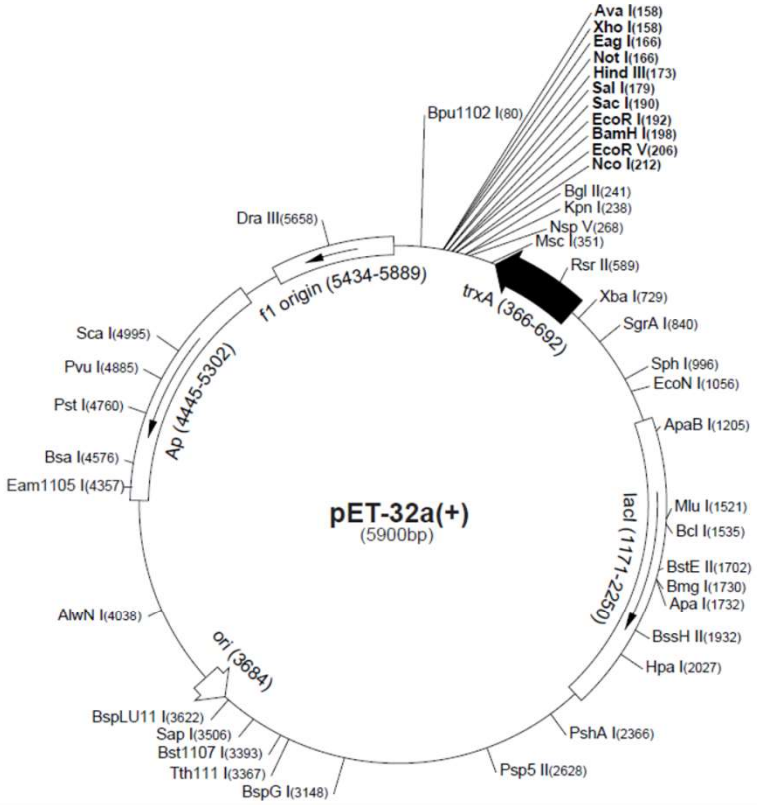
- **It should exhibit minimal level of basal expression**

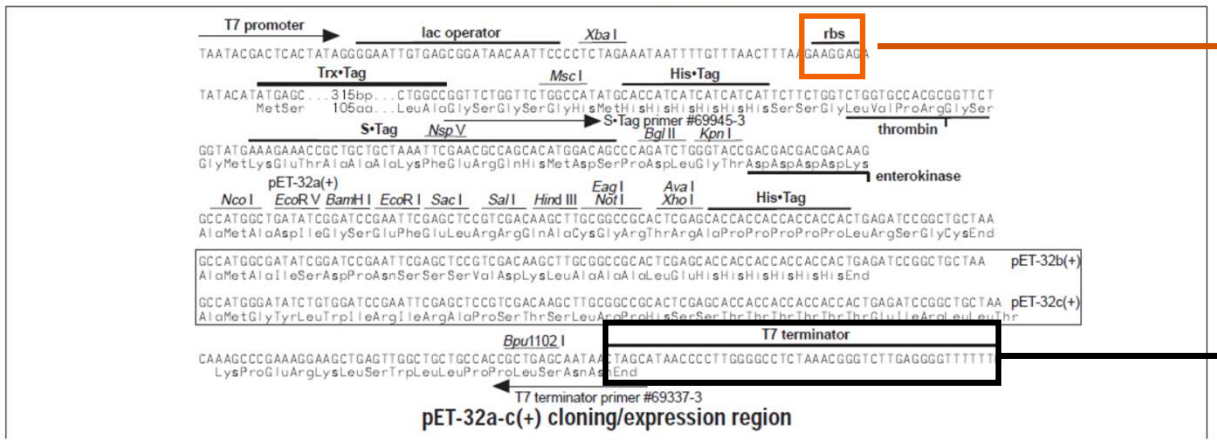
Large-scale gene expression preferably employs cell growth to high density and minimal promoter activity, followed by induction of the promoter. The tight regulation of a promoter is necessary for the production of proteins which may be detrimental to the host cell.

Expression vector configuration

pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3684
bla coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.

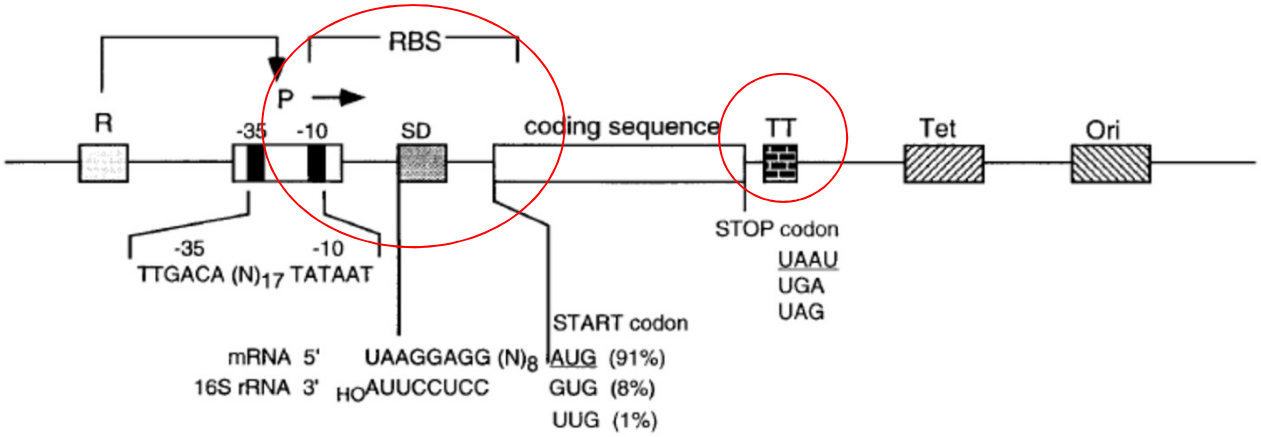




Ribosome-binding site

Transcription terminator

Expression vector configuration



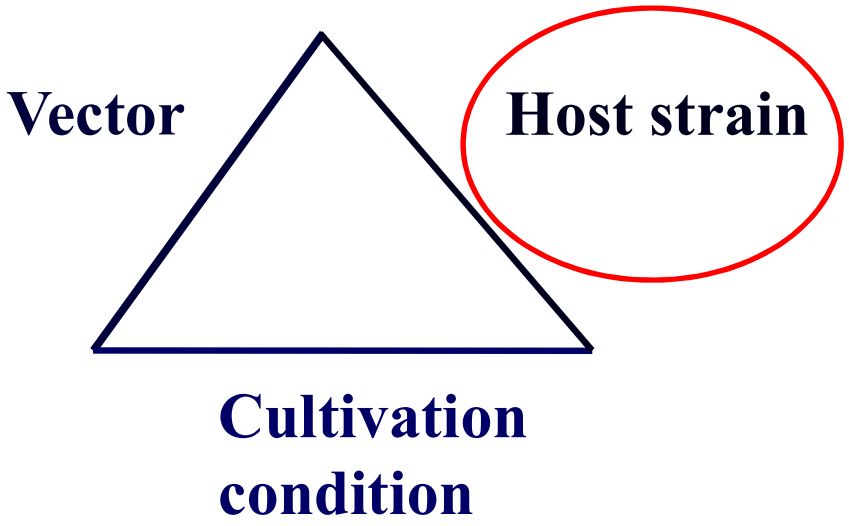
Ribosome-binding site consists of the Shine-Dalgarno (SD) sequence and the translational start codon

Length between SD sequence and start codon is 4-13 nucleotides. These length influences effectivity of translation initiation (optimal length is 4-8 nucleotides), high content of AT base pairs .

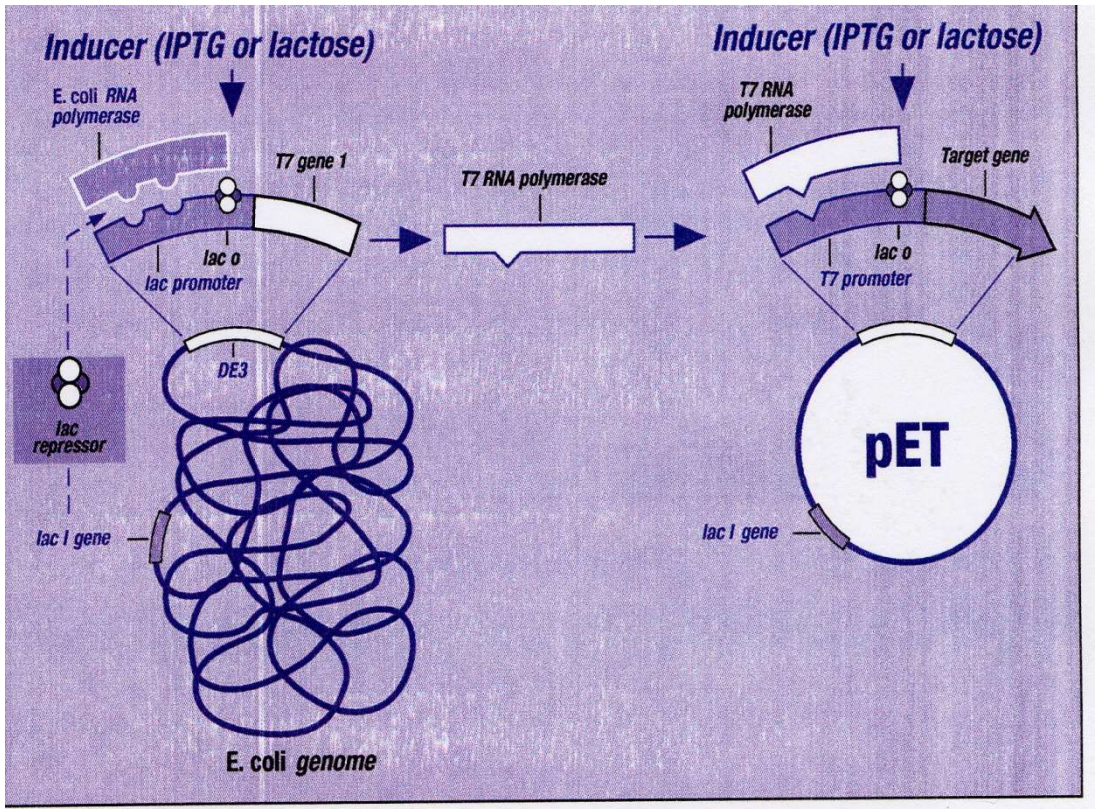
Transcription terminator T_7 term, rrnT1,T2

(preclusion of promoter occlusion, improvement of mRNA stability)

Expression system for recombinant protein production in *E.coli*



Production of recombinant protein in BL21(DE3) *E. coli* host strain



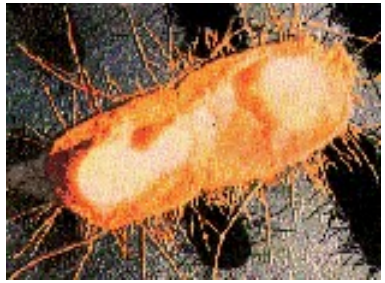
Toxicity of recombinant protein to the host strain

- Toxicity to the host strain is not limited to foreign genes but may also result from the overproduction of the specific native genes.

Proteins that are lethal for the host:

- Recombinant proteins containing hydrophobic regions are toxic for the cells because they associate with the membranes or incorporate to the membrane system of the cell and disturb membrane potential.
- Proteins inactivating ribosomes.

Selection of *E.coli* host strain considering problem with protein toxicity to the host



- Toxicity to the host strain is not limited to foreign genes but may also result from the overproduction of the specific native genes.

- Tight regulation of the expression system

Bacterial strains with various level of the expression regulation are commercially available.

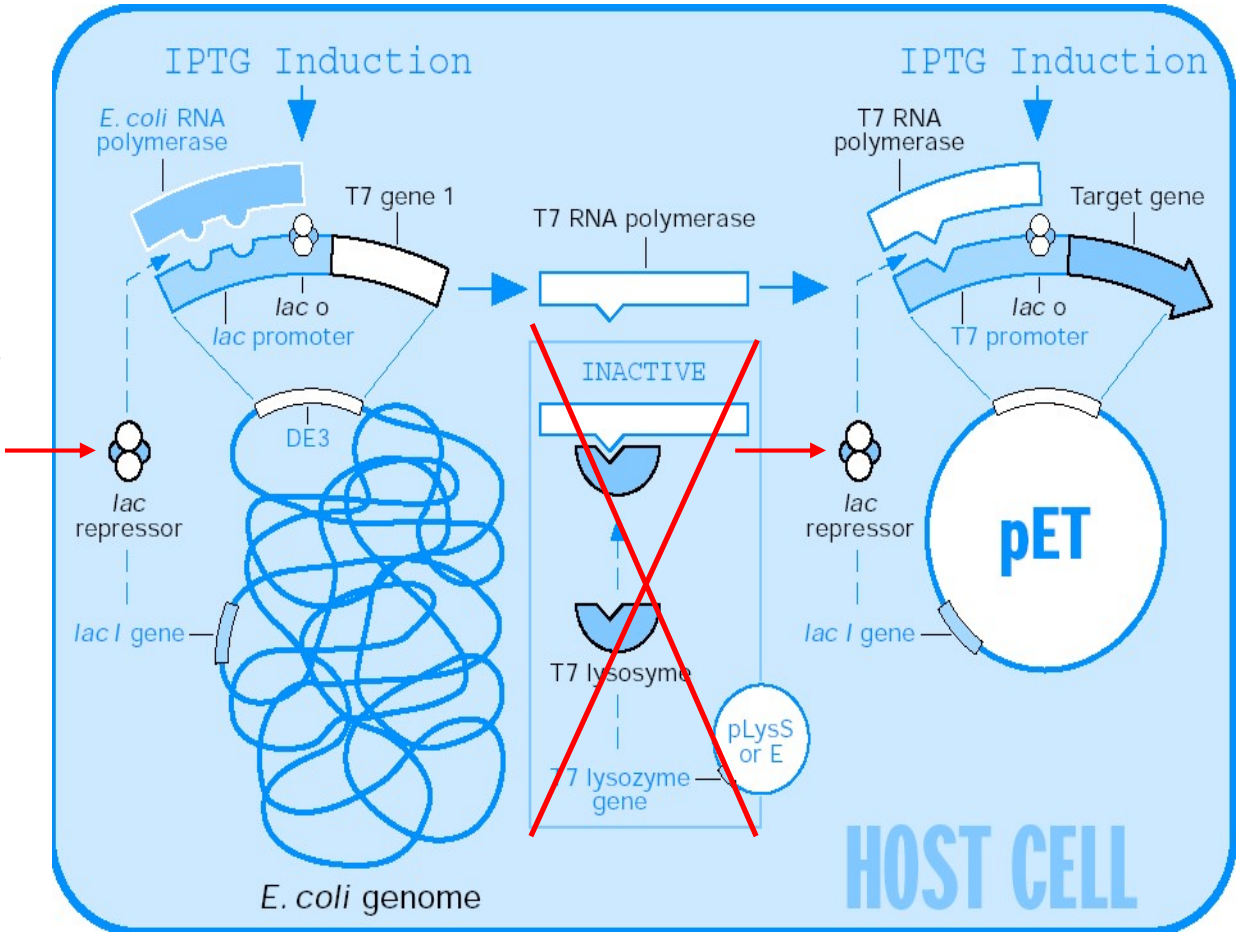
BL21(DE3)
BL21(DE3)pLysS

Different level of expression system regulation

✓**BL21(DE3)**

BL21(DE3)pLysS/E

BL21



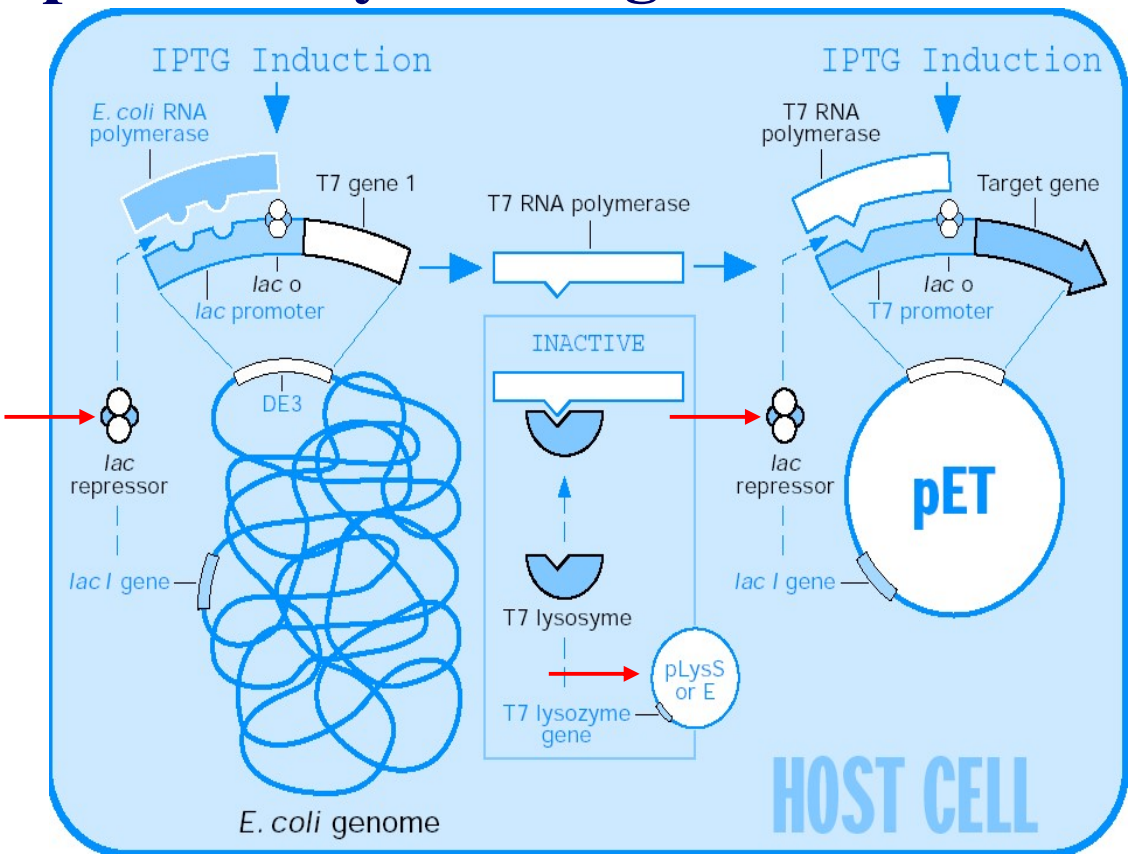
App. 10 % level of basal expression (before induction of expression) of certain gene.

Different level of expression system regulation

BL21(DE3)

✓**BL21(DE3)pLysS/E**

BL21



• pLysS and pLysE plasmids enabling tight regulation of expression system using T7 promoter. These plasmids harbor gene coding for lysozyme. Lysozyme inactivate T7 RNA polymerase to reduce basal expression.

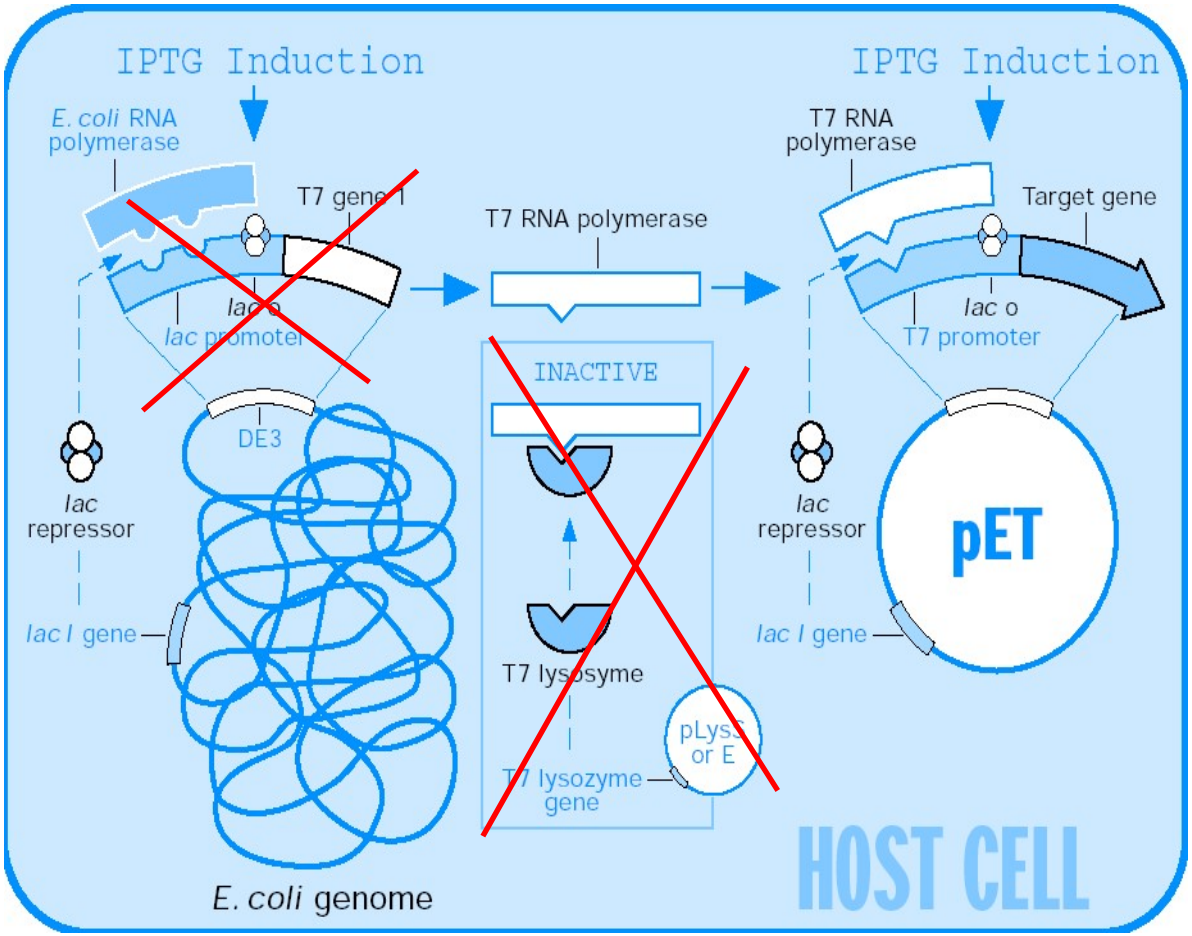
App. 1-3 % level of basal expression (before induction of expression) of certain gene.

Different level of expression system regulation

BL21(DE3)

BL21(DE3)pLysS/E

✓**BL21**



• Induction of expression by CEG bacteriophage infection (gene for T7 RNA polymerase)

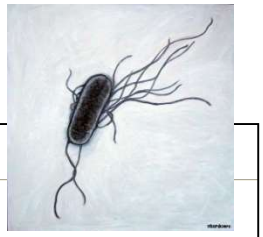
The highest level of repression!!

E. Coli codon usage

• Genes in both prokaryotes and eukaryotes show a nonrandom usage of synonymous codons.

- Codons that are rarely used by *E. coli* may occur in heterologous genes originating from eukaryotes, archebacteria.

- The frequency of use of synonymous codons usually reflects the abundance of their cognate tRNAs.



Escherichia coli K12 **[gbbct]: 14 CDS's (5122 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 19.7(101)	UCU 5.7(29)	UAU 16.8(86)	UGU 5.9(30)
UUC 15.0(77)	UCC 5.5(28)	UAC 14.6(75)	UGC 8.0(41)
UUA 15.2(78)	UCA 7.8(40)	UAA 1.8(9)	UGA 1.0(5)
UUG 11.9(61)	UCG 8.0(41)	UAG 0.0(0)	UGG 10.7(55)
CUU 11.9(61)	CCU 8.4(43)	CAU 15.8(81)	CGU 21.1(108)
CUC 10.5(54)	CCC 6.4(33)	CAC 13.1(67)	CGC 26.0(133)
CUA 5.3(27)	CCA 6.6(34)	CAA 12.1(62)	CGA 4.3(22)
CUG 46.9(240)	CCG 26.7(137)	CAG 27.7(142)	CGG 4.1(21)
AUU 30.5(156)	ACU 8.0(41)	AAU 21.9(112)	AGU 7.2(37)
AUC 18.2(93)	ACC 22.8(117)	AAC 24.4(125)	AGC 16.6(85)
AUA 3.7(19)	ACA 6.4(33)	AAA 33.2(170)	AGA 1.4(7)
AUG 24.8(127)	ACG 11.5(59)	AAG 12.1(62)	AGG 1.6(8)
GUU 16.8(86)	GCU 10.7(55)	GAU 37.9(194)	GGU 21.3(109)
GUC 11.7(60)	GCC 31.6(162)	GAC 20.5(105)	GGC 33.4(171)
GUA 11.5(59)	GCA 21.1(108)	GAA 43.7(224)	GGA 9.2(47)
GUG 26.4(135)	GCG 38.5(197)	GAG 18.4(94)	GGG 8.6(44)

Coding GC 52.35% 1st letter GC 60.82% 2nd letter GC 40.61% 3rd letter GC 55.62%

Arabidopsis thaliana **[gbpln]: 80395 CDS's (31098475 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 21.8(678320)	UCU 25.2(782818)	UAU 14.6(455089)	UGU 10.5(327640)
UUC 20.7(642407)	UCC 11.2(348173)	UAC 13.7(427132)	UGC 7.2(222769)
UUA 12.7(394867)	UCA 18.3(568570)	UAA 0.9(29405)	UGA 1.2(36260)
UUG 20.9(649150)	UCG 9.3(290158)	UAG 0.5(16417)	UGG 12.5(388049)
CUU 24.1(750114)	CCU 18.7(580962)	CAU 13.8(428694)	CGU 9.0(280392)
CUC 16.1(500524)	CCC 5.3(165252)	CAC 8.7(271155)	CGC 3.8(117543)
CUA 9.9(307000)	CCA 16.1(502101)	CAA 19.4(604800)	CGA 6.3(195736)
CUG 9.8(305822)	CCG 8.6(268115)	CAG 15.2(473809)	CGG 4.9(151572)
AUU 21.5(668227)	ACU 17.5(544807)	AAU 22.3(693344)	AGU 14.0(435738)
AUC 18.5(576287)	ACC 10.3(321640)	AAC 20.9(650826)	AGC 11.3(352568)
AUA 12.6(391867)	ACA 15.7(487161)	AAA 30.8(957374)	AGA 19.0(589788)
AUG 24.5(762852)	ACG 7.7(240652)	AAG 32.7(1016176)	AGG 11.0(340922)
GUU 27.2(847061)	GCU 28.3(880808)	GAU 36.6(1139637)	GGU 22.2(689891)
GUC 12.8(397008)	GCC 10.3(321500)	GAC 17.2(535668)	GGC 9.2(284681)
GUA 9.9(308605)	GCA 17.5(543180)	GAA 34.3(1068012)	GGA 24.2(751489)
GUG 17.4(539873)	GCG 9.0(280804)	GAG 32.2(1002594)	GGG 10.2(316620)

Coding GC 44.59% 1st letter GC 50.84% 2nd letter GC 40.54% 3rd letter GC 42.38%

http://www.kazusa.or.ip/codon/

Low-usage codons in *E. coli*

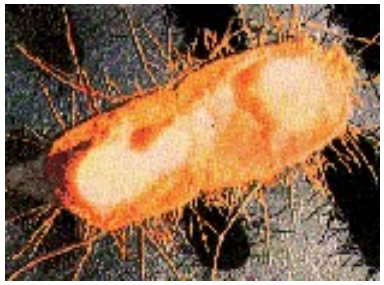
Codon(s)	Amino acid
AGA, AGG, CGA, CGG	Arg
UGU, UGC	Cys
GGA, GGG	Gly
AUA	Ile
CUA, CUC	Leu
CCC, CCU, CCA	Pro
UCA, AGU,UCG, UCC	Ser
ACA	Thr

Makrides, 1996

Expression of low-usage codons containing heterologous genes leads to the translation errors!

- Premature termination of translation (truncated product)
- Open reading frame shift (shift by two amino acids in AGA codon position)
- Change of amino acid – often arginine (AGA codon) for lysine

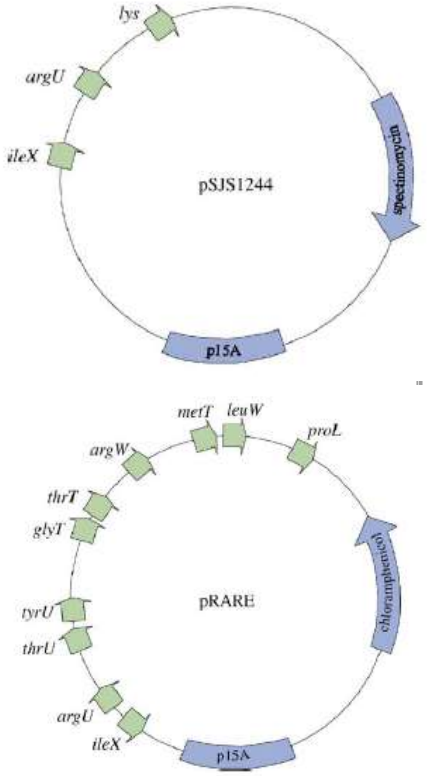
Selection of *E.coli* host strain considering problems with low usage codons



- Comercially available strains that produce tRNA of low usage codons.

•BL21 (DE3) CodonPlus-**RIL**	•**AGG/AGA** (arginine,**R**), **AUA** (isoleucine, **I**) and **CUA** (leucine, **L**)
•BL21 (DE3) CodonPlus-**RP**	•**AGG/AGA** (arginine, **R**) and **CCC** (proline, **P**)
•Rosetta or Rosetta (DE3)	•**AGG/AGA** (arginine, **R**), **CGG** (arginine, **R**), **AUA** (isoleucine, **I**) **CUA** (leucine, **L**)**CCC** (proline), and **GGA** (glycine, **G**)

Plasmids complementing tRNA.



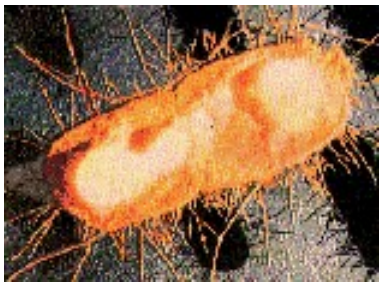
OR: Site directed mutagenesis of low usage codons

Protein degradation

***E.Coli* proteolytic system** includes a large number of proteases that are localized mainly in the cytoplasm, but also in the periplasm, and the inner and outer membranes.

- In- complete polypeptides
- Proteins with amino acid substitutions
- excessively synthesized subunits of multimeric complexes
- Proteins damaged through oxidation or free radical attack
- Foreign, recombinant proteins (proteins < 10 kDa are problematic)

Selection of *E.coli* host strain considering proteolysis of recombinant protein



Protease-deficient host strains

• Mutation eliminating production of proteases and thereby proteolytic degradation of recombinant proteins.

BL21 expression strain deficient in:

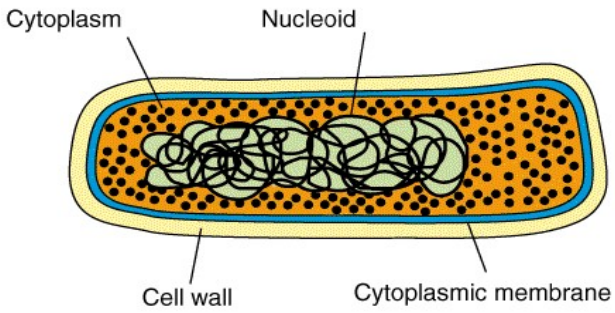
lon cytoplasmatic protease
ompT periplasmatic protease

Targeted protein expression

Cytoplasmatic expression

Periplasmatic expression

Extracellular expression (into the cultivation medium)



Cytoplasmatic expression

- mostly used

Advantages

- High protein yield
- Simplier plasmid constructs
- Inclusion bodies

Disadvantages

- Inclusion bodies
- Reducing enviroment
- Proteolysis
- More complex purification

Inclusion bodies

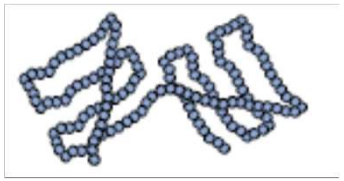
• Insoluble protein aggregates (up to $2\mu m^{3)}$ consisted of native protein of limited solubility, of the unfolded state and partially folded intermediate state of protein.

What causes their formation?

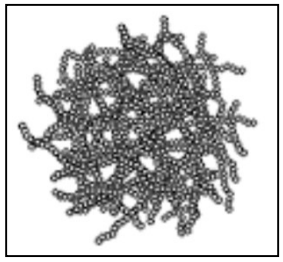
1. Microenvironment of *E.coli* may differ from that of original source in terms of redox potential (reducing environment in E.coli cytoplasm), pH, osmolarity, absence of chaperones, cofactors, lack of post-translational modifications.

2. High level of expression

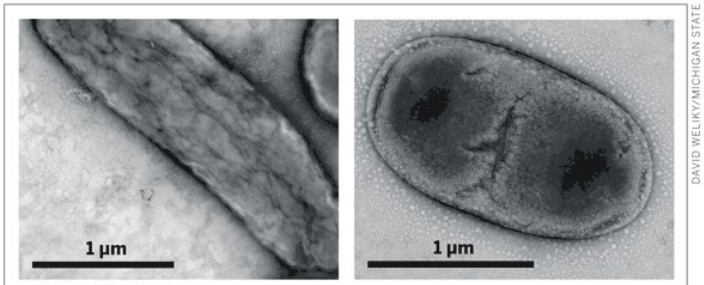
Hydrophobic stretches in the nascent polypeptide are present at high concentrations and associate intramolecularly.



Soluble protein



Inclusion bodies– insoluble protein



AGGREGATES Normal *E. coli* cells (left) can become filled with inclusion bodies (right, large dark spots) when induced to overexpress recombinant protein.

Inclusion bodies

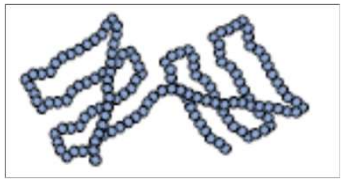
Advantages

- Easy isolation in high purity and concentration
- Protease protection
- Advantage for lethal protein

Disadvantages

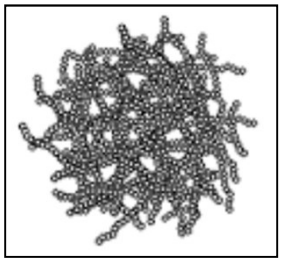
- Protein insolubility
- Refolding to recover protein activity
- Refolded protein may not recover its biological activity
- Reduction of final yield

Inclusion bodies



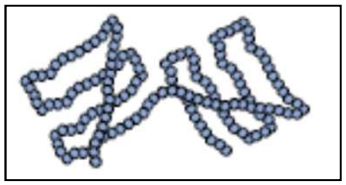
Soluble protein

E.Coli expression



Inclusion bodies

Isolation followed by refolding



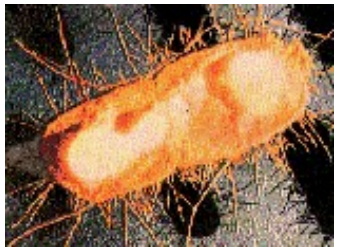
Soluble protein

Modification of expression conditions

E.g.

- Lowering of cultivation temperature
- Co-production of chaperones
- Using of solubility enhancing tags (thioredoxin)
- Selection of *E.coli* strain – e.g. thioredoxin reductase deficient strain

Selection of *E.coli* host strain considering problems with insolubility



• If the protein contains one or more disulfide bonds, folding is stimulated in oxidizing environment in cytoplasm, that is provided by following *E.coli* strains.

AD494	• Mutation in gen for pro thioredoxinreductase (trxB)
Origami	• Two mutation in gen for thioredoxin reductase (trxB) and glutathionreductase (gor)

Periplasmatic expression

- Periplasm contains only 4% of all cellular proteins (app. 100 proteins)
- Transmembrane transport is mediated by N-terminal signal peptide
- Prokaryotic signal peptides succesfully used in in *E.coli* (ompA, ompT from *E.coli*, protein A from *S. Aureus*, endoglukanase z *B.subtilis*)

Advantages

- Simplier purification
- Limited proteolysis
- Improving disulfid bond formation/folding

Disadvantages

- Signal peptide does not always provide transport do periplasm
- Formation of inclusion bodies

Extracelular expression

• Protein secretion into a cultivation medium

•Effective transport through outer membrane is missing in *E.coli (E.coli naturaly secretes limited number of proteins)*.

• The manipulation with transport ways enabling protein secretion is still unsuccesfull.

Advantages

• Minimal contamination by other protein (simplier purification)

• Limited proteolysis

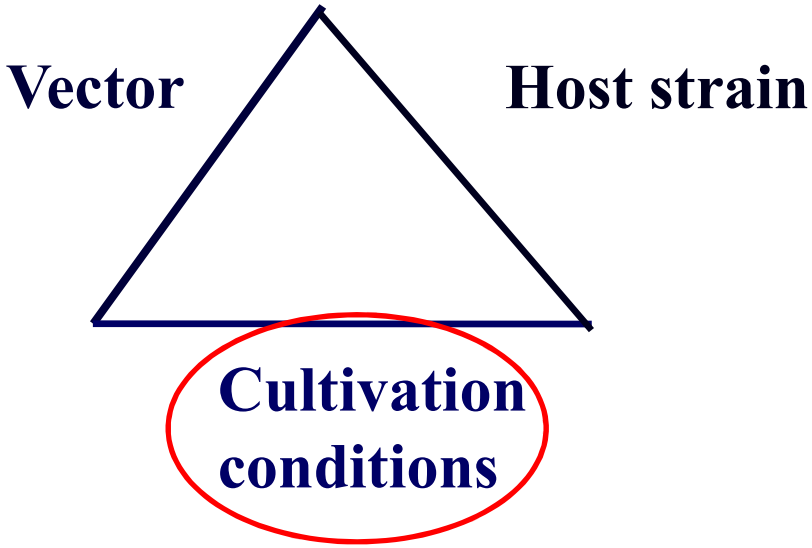
• Folding improvement

Disadvantages

• very low secretion

• highly diluted protein

Expression system for recombinant protein production in *E.coli*

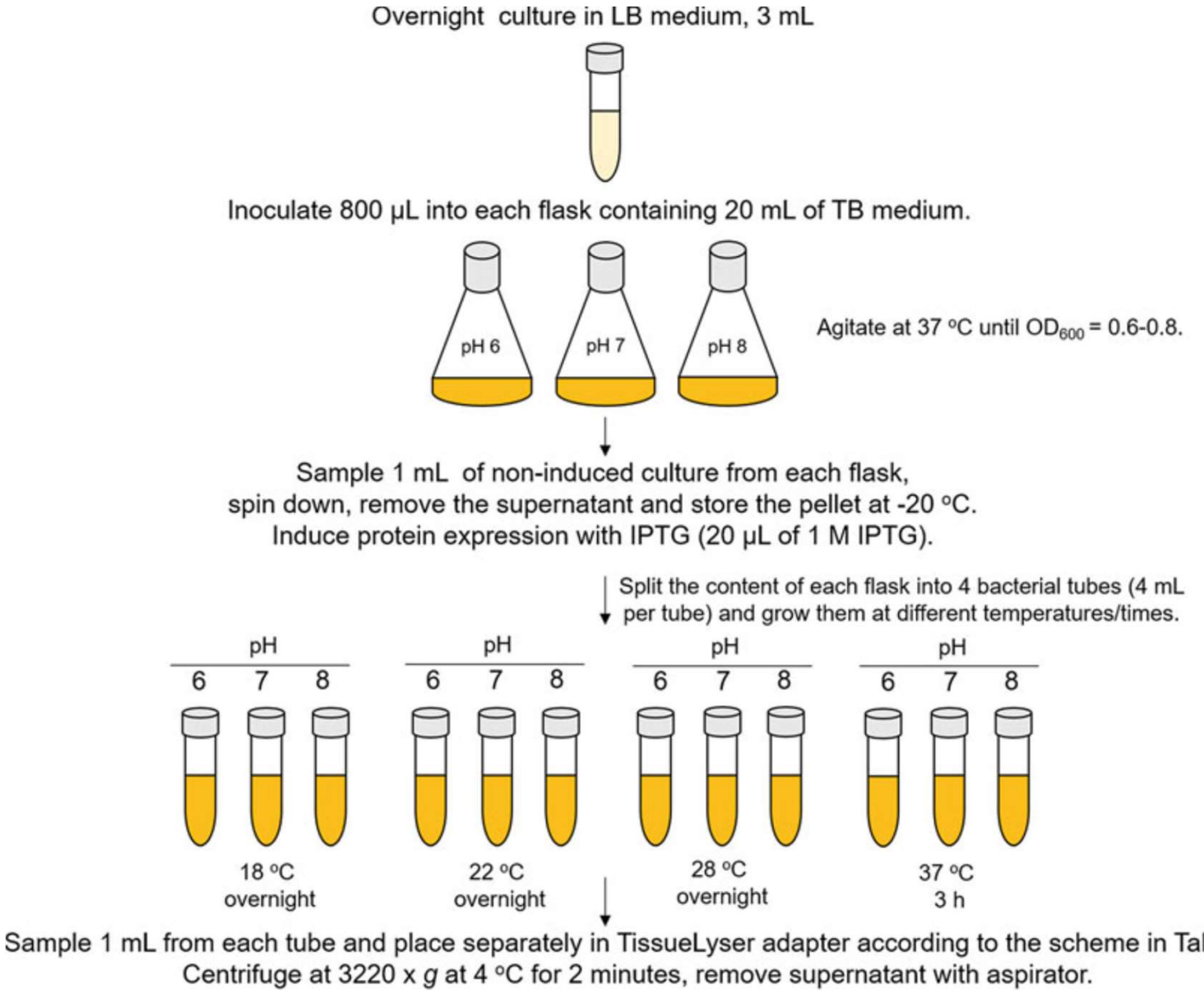


Modification of cultivation conditions

Possibilities for protein solubility enhancement:

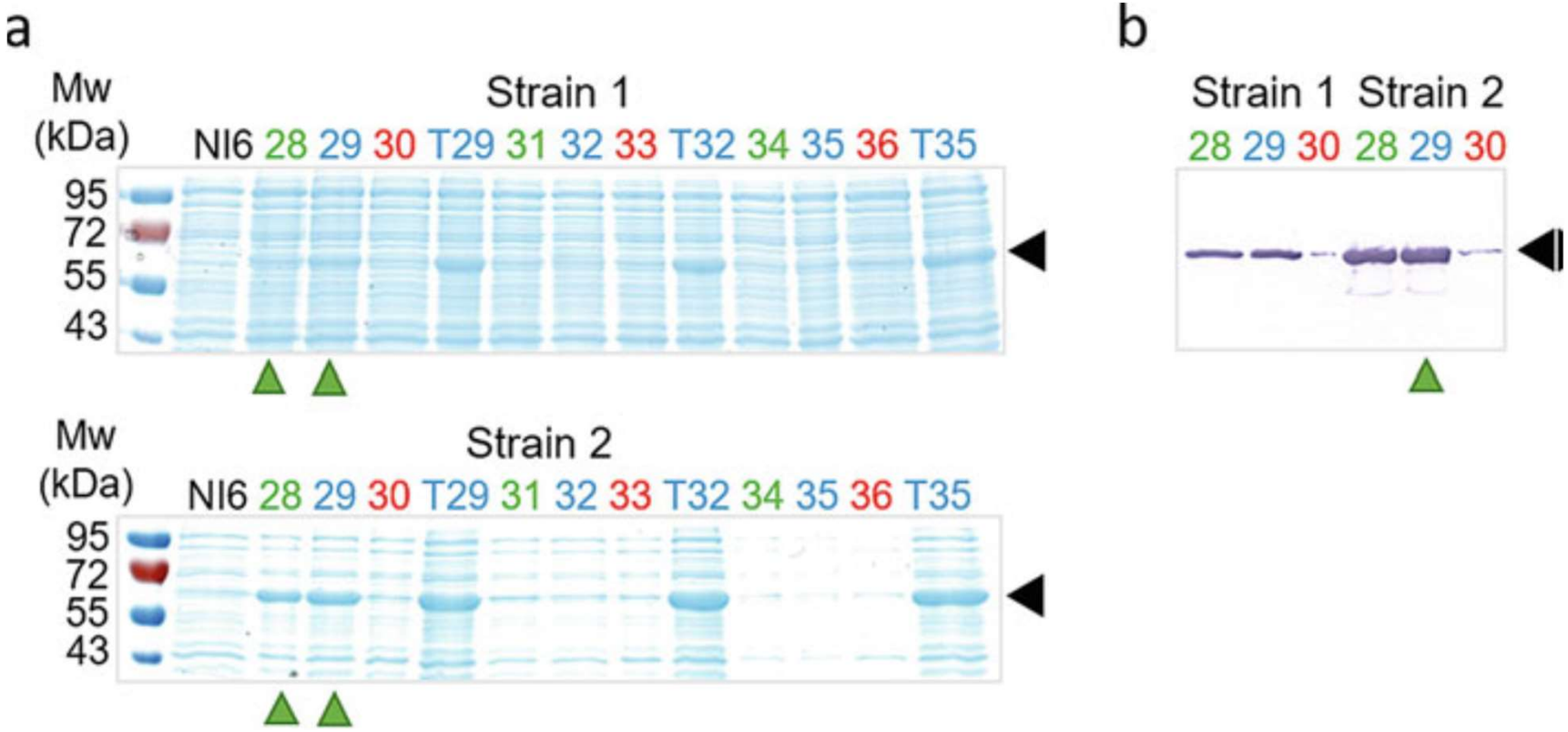
- High cell culture optical density
- Medium composition (pH, addition of specific substrates and cofactors, type of cultivation medium – reach and minimal)
- Temperature optimization for induction of expression.
- Concentration of inducing agent
- Length of induction

Experimental setup for protein expression and solubility trials



A. Smitkowska et al, 2020

Example results from expression and solubility test



E. coli ER2566, TB medium pH 6.0, 3 h induction at 37°C, lysis buffer: MES pH 6.0 (28) or Tris–HCl pH 7.5 (29)

A. Smitkowska et al, 2020

Evaluation of cultivation temperature optimisation

Production of AHP proteins in soluble form (in %)						
t(°C) growth/induction	AHP1	AHP2	AHP3	AHP4	AHP5	AHP6
37°C/28°C	8 %	85 %	100%	0	76 %	0
37°C/22°C	82 %	73 %	100%	0	81 %	51 %
22°C/22°C	71 %	78 %	100%	30 %	81 %	73 %

Production of heterologous proteins in *yeasts*

ADVANTAGES :

- easy gene manipulations
- fast growth into the high densities (fermentor), low price
- ability to posttranslationaly modify expressed protein
- ability of extracellular secretion of produced protein
- ability to produce protein with proper conformation
- expressed protein without contamination by endotoxines

DISADVANTAGES:

- using type of N-oligosacharides structurally different from mammals
- hyperglycosylation

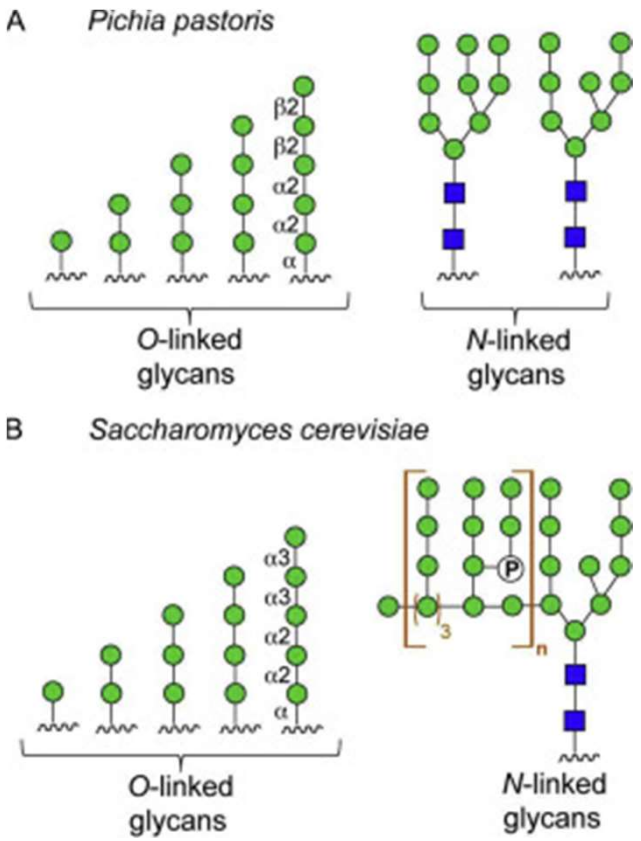
Yeast expression system– *Sacharomyces cerevisiae* *- Pichia pastoris*

- *Sacharomyces cerevisiae* –first yeasts used for recombinant protein production
- *P. pastoris* – most used yeasts expression system

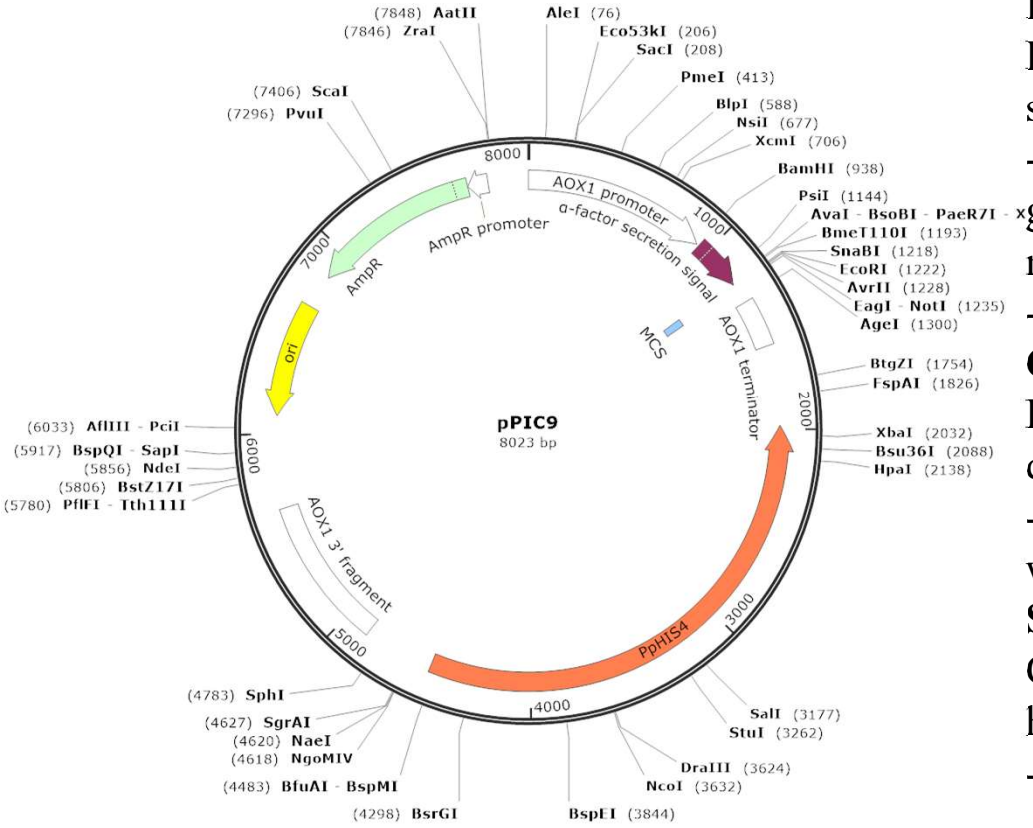
- *P. pastoris* - ability to grow to high densities (*S. cerevisiae* produce high amount of secondary metabolites, which limit to reach high density of the yeast culture)

- *P. pastoris* use different type of N-glycosylation compared to *S. cerevisiae*:
glycans contains 8-17 mannose molecules (linked by α 1,2 bond) in *P. Pastoris*

Glycans contain 50-150 mannose molecules (so called hyperglycosylation, terminal mannoses linked by α 1,3 bond) in *S. cerevisiae*



Vector configuration for expression in *P. pastoris*



- transformation is performed using integrating vectors (plasmids replicating after integration into the chromosome)

Promotors:

Inducible

Promotor of alcohol oxidase 1 (AOX1) gene - strong a strictly regulated

-is fully suppressed when growing on glucose or glycerol and strongly induced when growing on methanol as the carbon source itself

-20-30% of total intracellular protein

Constitutive

Promotor of glyceraldehyde-3 – phosphate dehydrogenase (GAP) gene

-constitutively expressed, and is most induced when growing on glucose

Selection markers: auxotrophic *HIS4*

Complementation of auxotrophic mutation of histidin

- Selection in yeast

Antibiotic resistance

Ampicillin – selection in bacteria

Secretion signals

PHO1 (acid phosphatase) z *P. pastoris*

α-mating factor (α-MF) z *S. cerevisiae*

SUC2 (invertase) z *S. cerevisiae*

Recombinant protein expression in *Pichia pastoris*

Gene cloning into a vector, selection of construct in E.coli, isolation of DNA construct

↓

Construct linearisation

↓

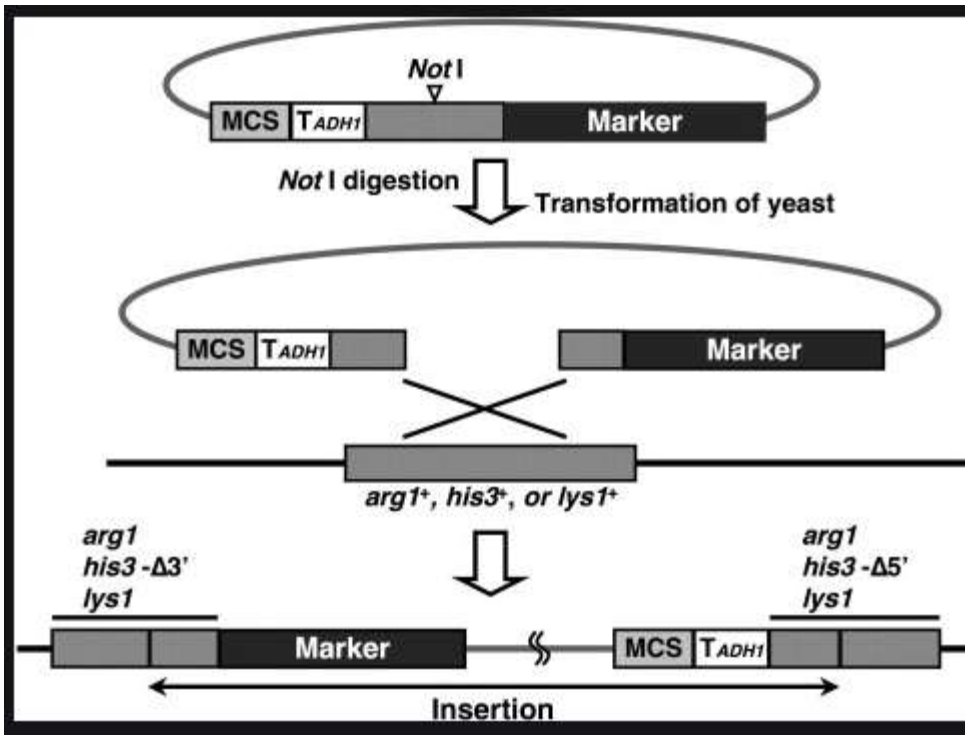
Transformation of competent *P. pastoris*

↓

Selection of transformants

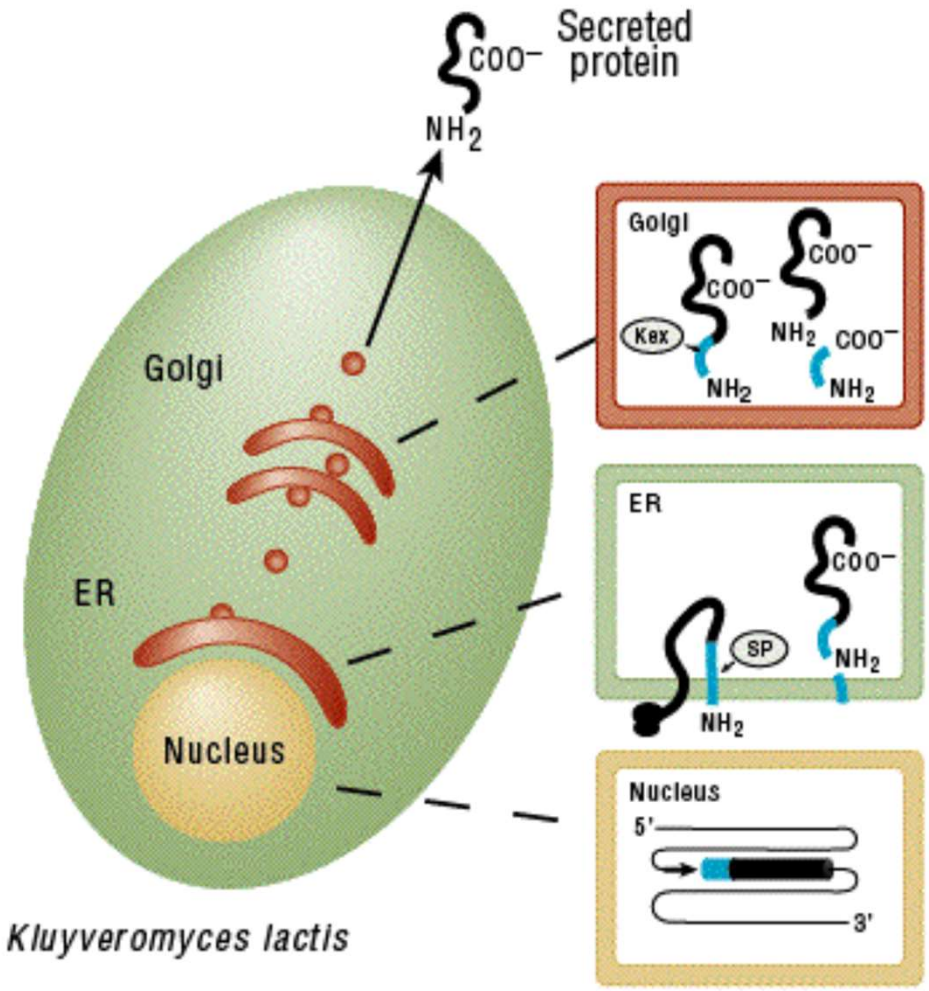
↓

Recombinant protein expression



Protein secretion in yeasts

- Expression in yeasts can be intracellular or extracellular (secretion)



The secretory pathway is very similar to that of higher eukaryotes

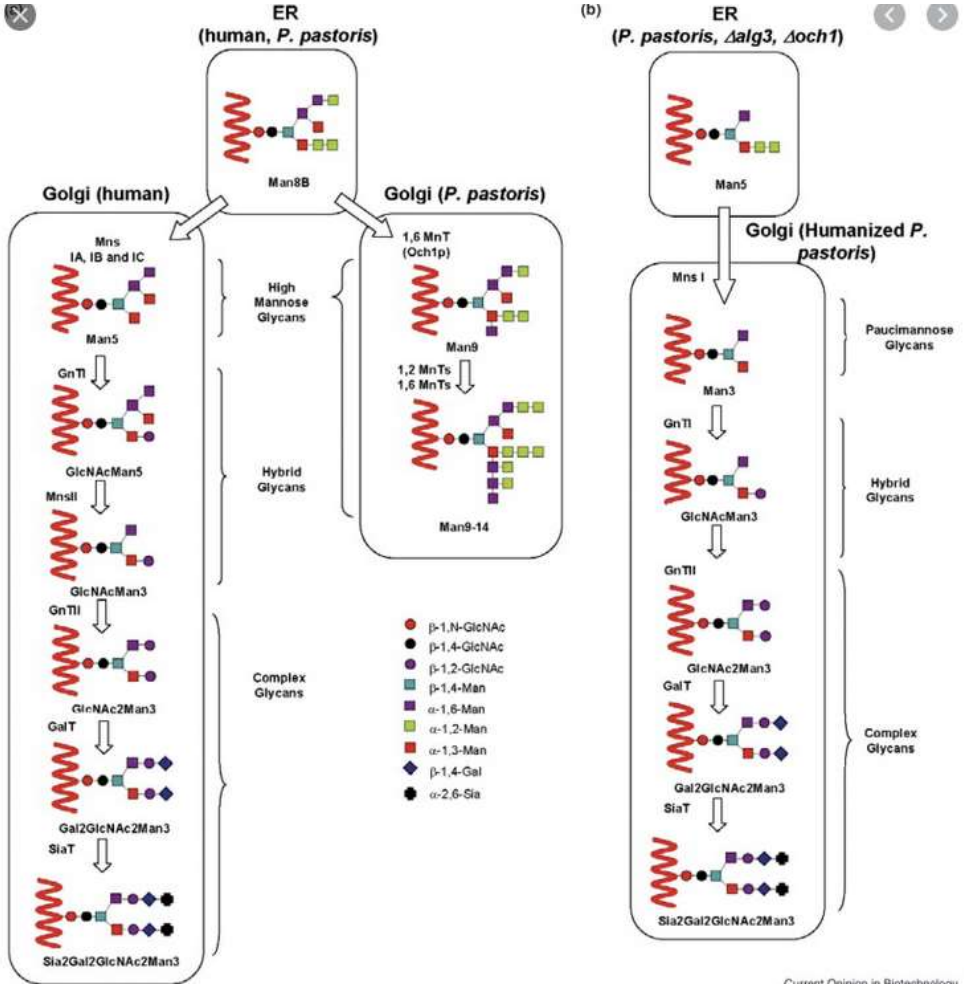
The N-terminal peptide for cotranslational translocation into the ER is cleaved by a signal peptidase.

Examples of signal sequences: Pho5, Suc2 and α-factor

Secretion by α-factor
The protein is expressed in the nucleus together with the signal peptide - α-MF (α-mating factor) domain.
1. The signal peptide in the α-MF domain directs the protein to the ER, where it is cleaved by a signal peptidase.
2.The fusion protein is transported to the Golgi apparatus, where the Kex protease cleaves the remainder of the α-MF domain and the protein is secreted into the medium.

Differences in glycosylation

- High mannose content is immunogenic to humans.
- *P. pastoris* (e.g. strain SuperMan5 and others) with a uniform human N-glycosylation method were developed for the production of therapeutic proteins.
- Additional enzymes: mannosidase I and II, galactose transferase, N-acetylglucosamines and sialic acid
- Expanding use of *P. pastoris*, eg. for antibody expression.



Production of heterologous proteins in insect cells

(insect cells with baculoviruses)

ADVANTAGES:

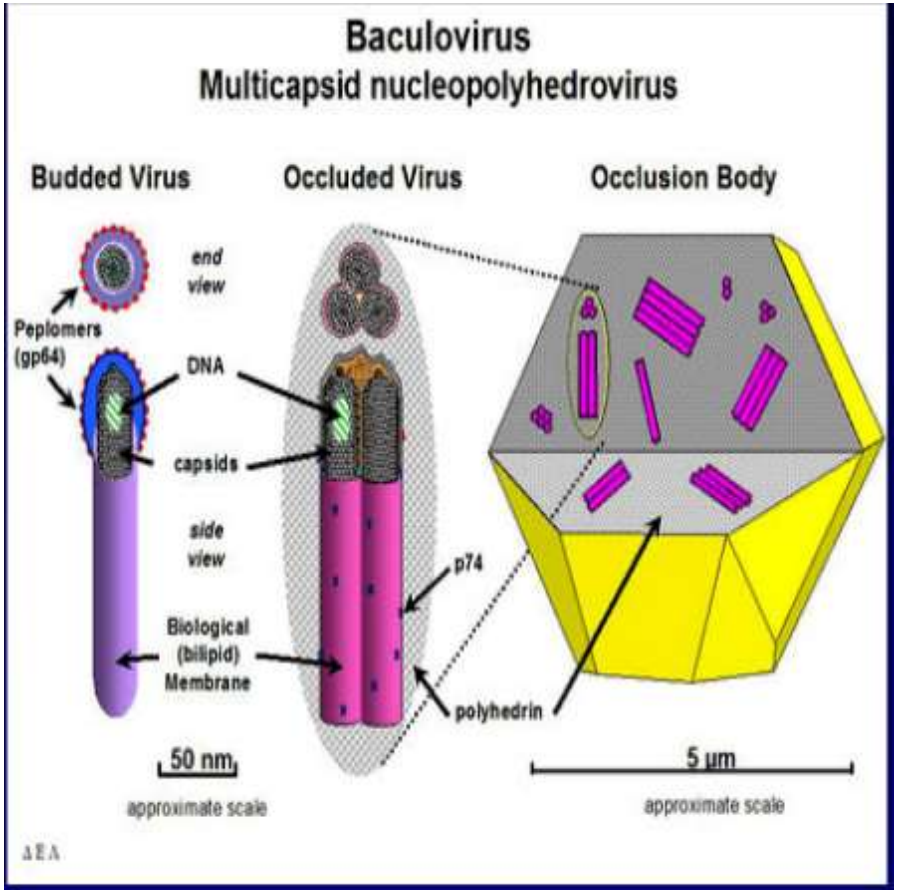
- Ensuring the native conformation of the protein
- Posttranslational modifications
- There is no contamination of the final product with endotoxins
- Protein secretion

DISADVANTAGES:

- Negative effect of baculovirus infection on cell viability
- Heterologous genes are not produced continuously (each expression requires a new infection of cells with baculovirus)
- A method of glycosylation different from mammalian cells
- Lower yield, more time consuming, expensive media, harder to handle (risk of contamination)

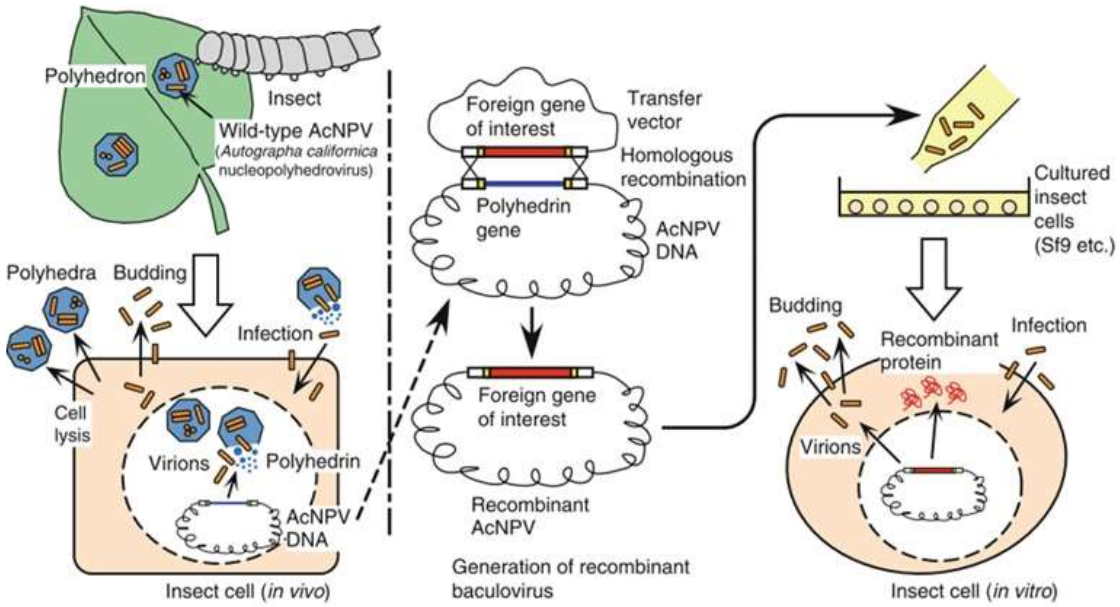
What is baculovirus?

- enveloped, ds DNA virus with rod-shaped capsid
- during the life cycle there are two different forms of "budding virus" and encapsulated virus
- highly species specific - infect only invertebrates
- the most common hosts are immature larval forms of insects
- AcMNPV virus (Autographa californica multiple nuclear polyhedrosis virus) is one of the most studied and used baculoviruses.



Insect cells with baculoviruses expression system

- based on infection of cultured insect cells with a recombinant AcMNPV virus (Autographa californica multiple nuclear polyhedrosis virus) carrying a gene for the production of a target protein.

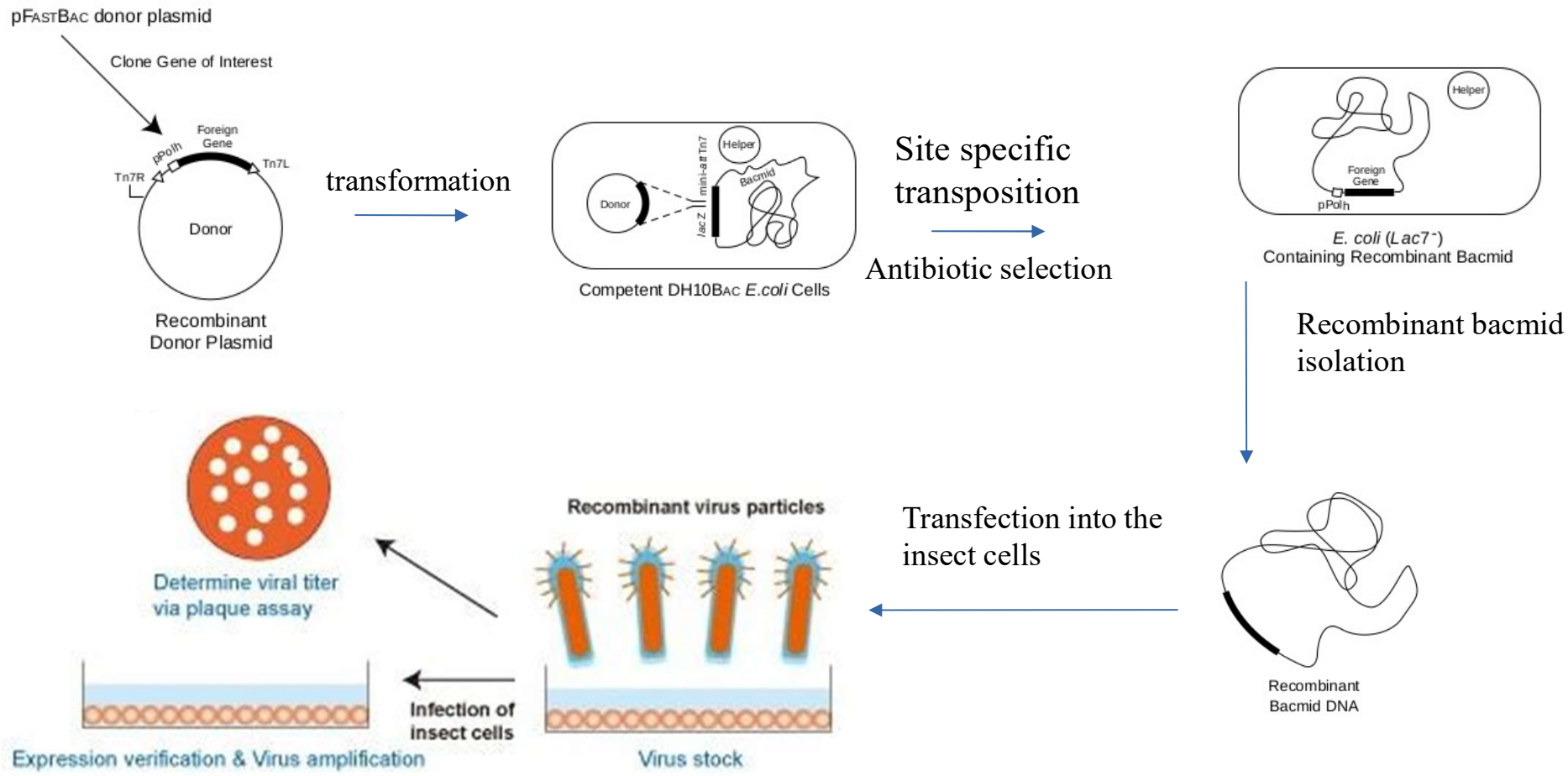


- insect host cells
  - ovarian cells of butterflies of the species Spodoptera frugiperda (Sf9, Sf-21)

- baculovirus vector (so-called bacmid)
  - Large shuttle vector (AcMNPV)
  - contains all the genes necessary for the production of viral particles

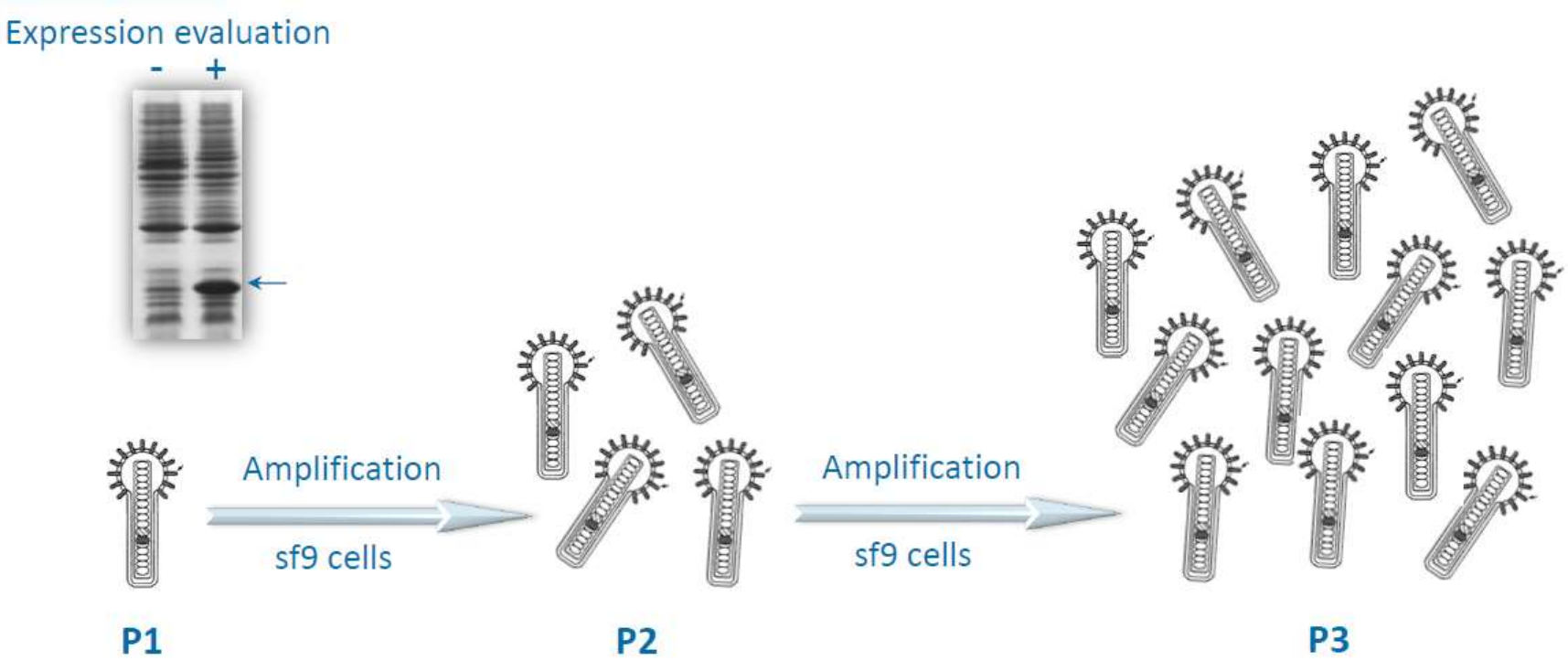
- very strong polyhedrin gene promoter (polyh)

Bac-to-Bac expression system : „from Bacterium to Baculovirus“

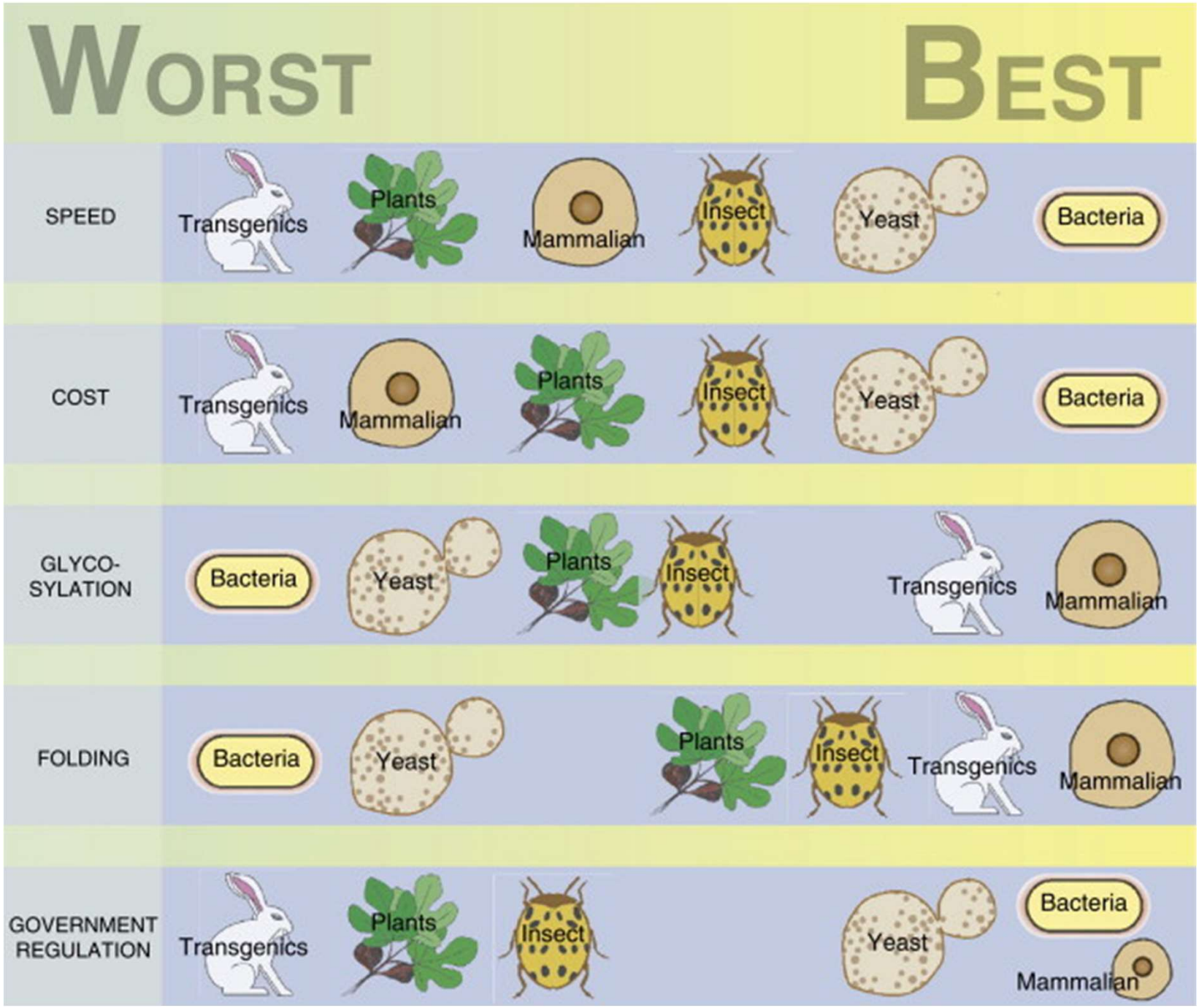


- After successful transfection, recombinant viral particles are recovered from the medium (viral stock) and can be amplified to higher amounts so that they can be used to infect a sufficient number of cells.

Bac-to-Bac expression system: „from Bacterium to Baculovirus“



After two generation a titer of titru 10^6 - 10^8 pfu/ml is reached.


WORST
BEST
SPEED
Transgenics
Plants
Mammalian
Insect
Yeast
Bacteria
COST
Transgenics
Mammalian
Plants
Insect
Yeast
Bacteria
GLYCO-
SYLATION
Bacteria
Yeast
Plants
Insect
Transgenics
Mammalian
FOLDING
Bacteria
Yeast
Plants
Insect
Transgenics
Mammalian
GOVERNMENT
REGULATION
Transgenics
Plants
Insect
Yeast
Bacteria
Mammalian