Gene Expression Systems in Prokaryotes and Eukaryotes

- Expression studies
- Expression in Prokaryotes (Bacteria)
- Expression in Eukaryotes

Gene Expression Systems in Prokaryotes and Eukaryotes

Expression studies:

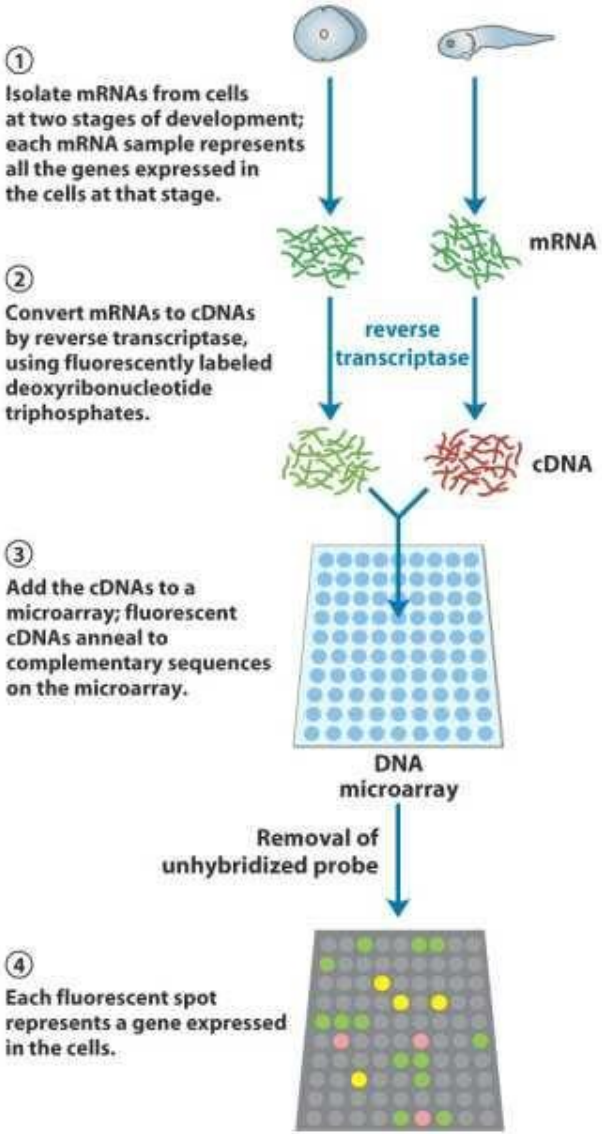
1. Analyzing Transcription
   - Northern blot
   - Micro array
   - real-time PCR
   - Primer extension

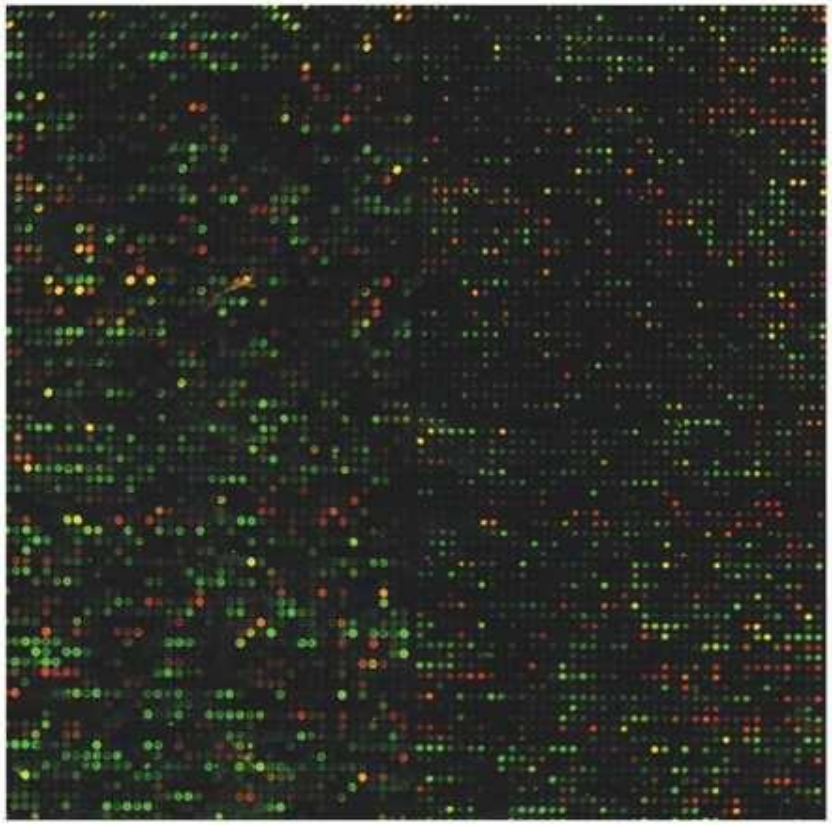
2. In vivo Expresion studies
   Use of report genes to study regulatory elements

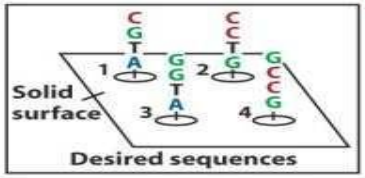
3. Analyzing Translation
   - Western blot - immuno assays
   - 2D electrophoresis
   - proteomics

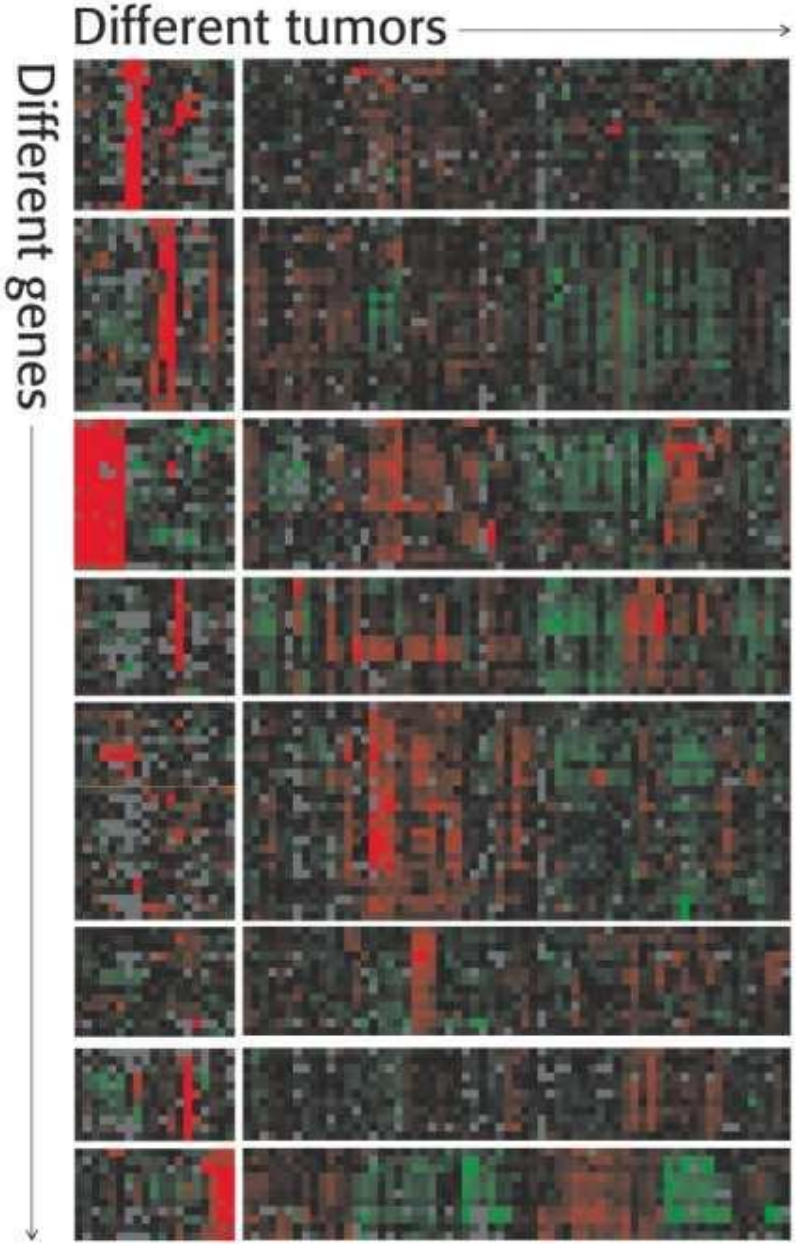
Studying Transcription

Microarray technique – DNA chips



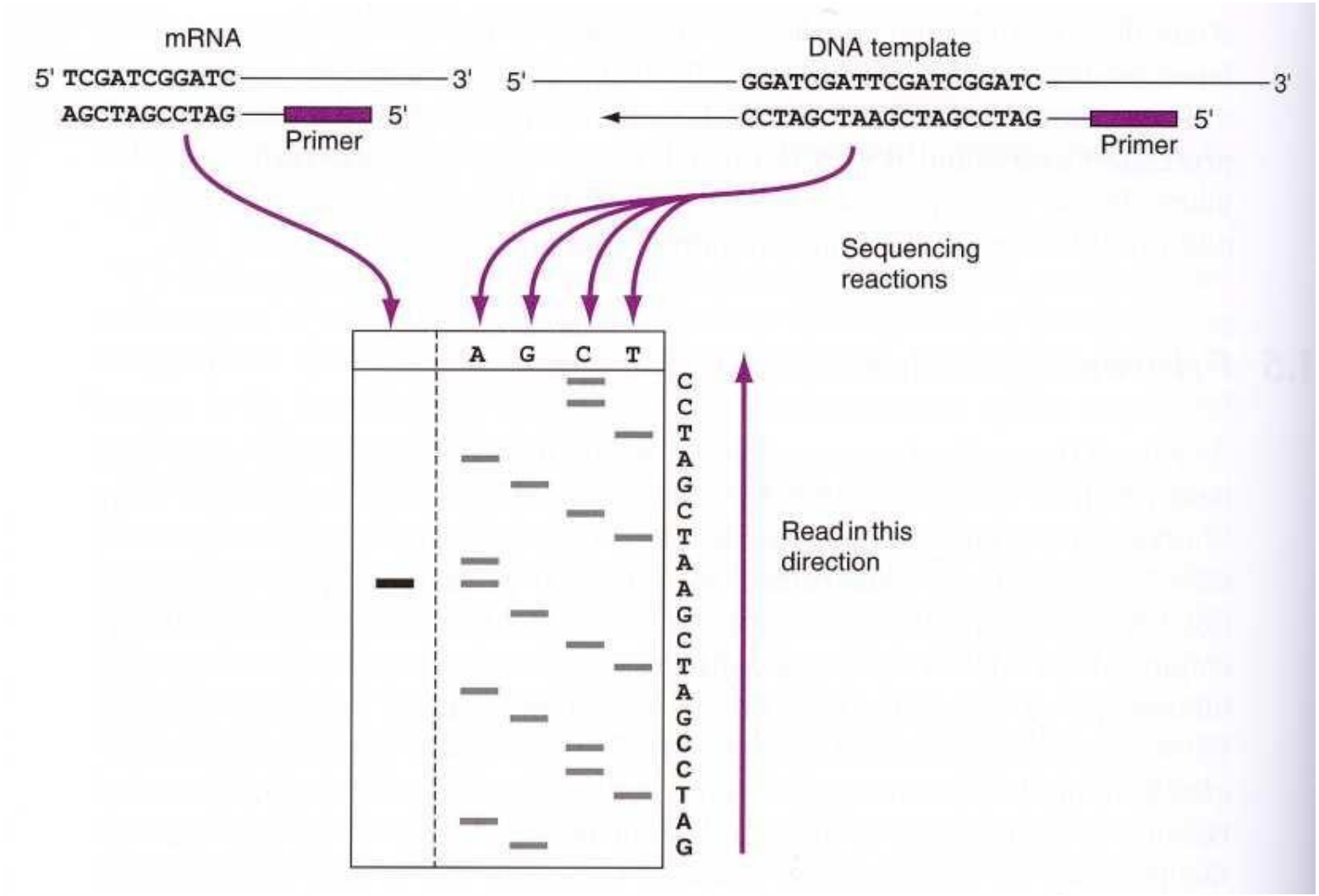





Different tumors
Different genes

Studying Transcription

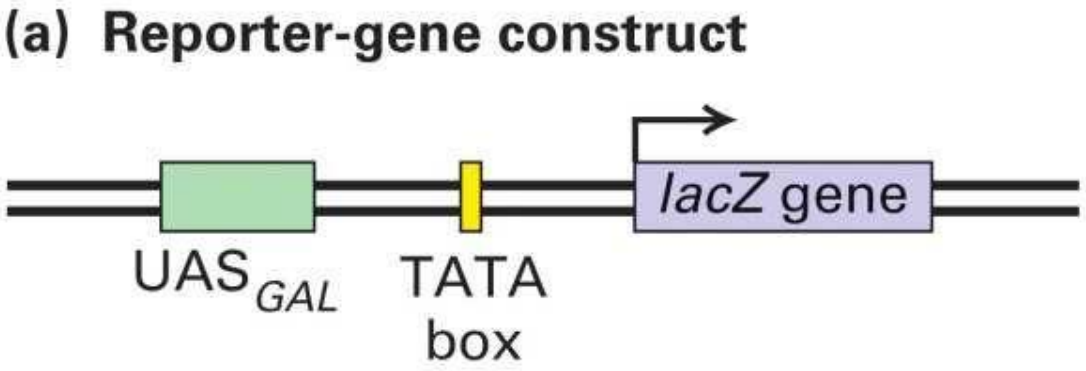
Primer Extension



Promoter Studies

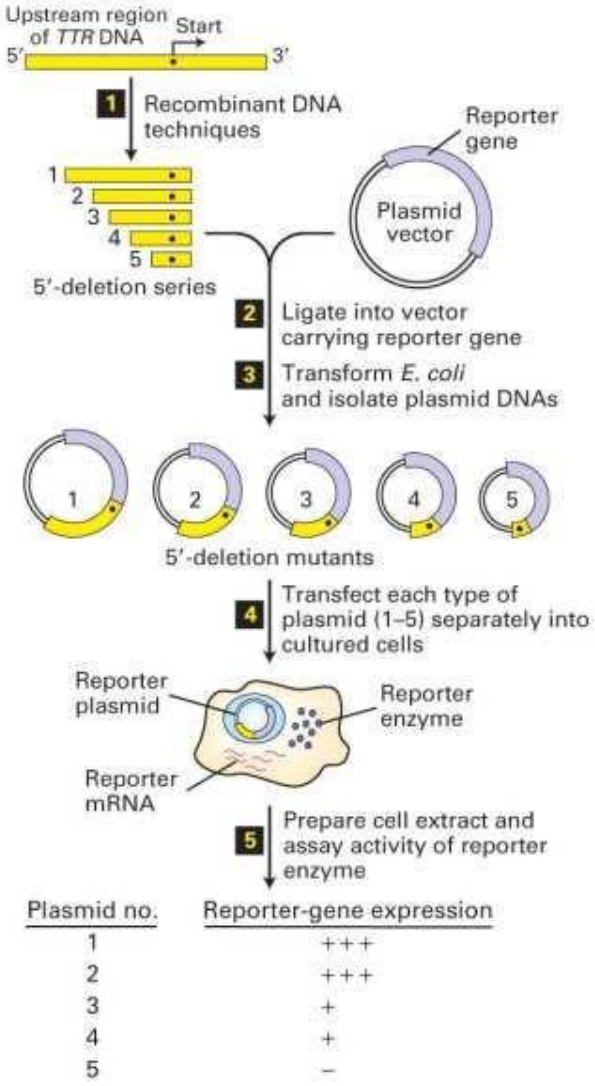
Used reporter genes:

- *Lac Z*
- *GFP*
- *Luciferase*



Promoter

Promoter studies by using reporter genes



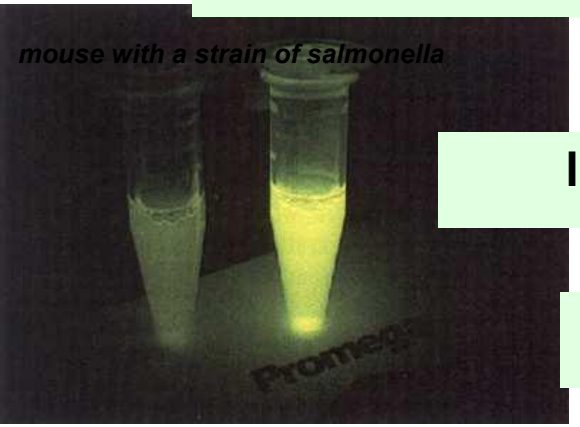
Plasmid no.	Reporter-gene expression
1	+++
2	+++
3	+
4	+
5	–

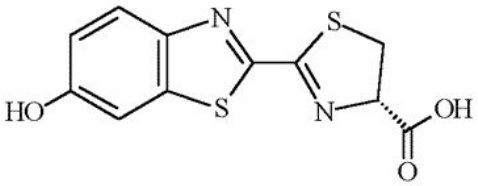
Luciferase (luc) systems

firefly species **Photinus pyralis**

Expressed luciferase catalyses

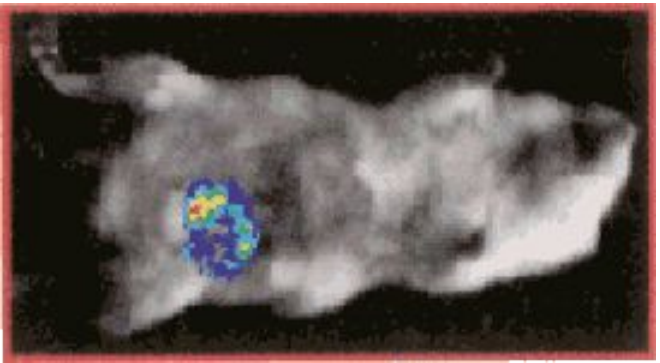
oxidation of compounds called luciferans
(ATP-dependent process)





luciferans emit fluorescense

luminometer measurement

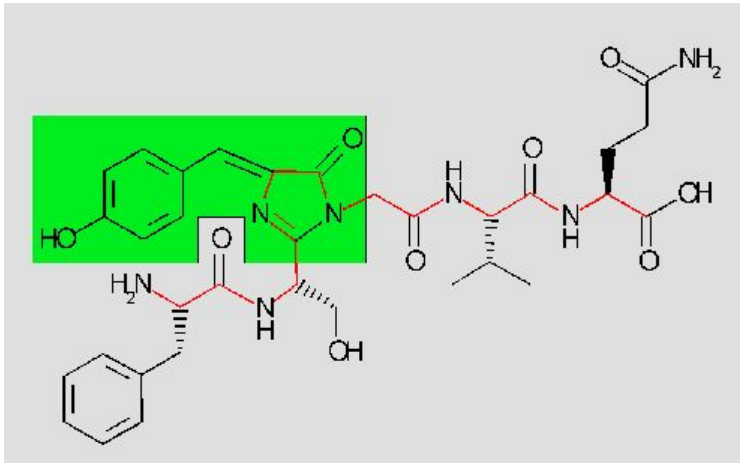


Mice are injected with LUC+ salmonellas.
Sensitive digital cameras allow **non-invasive detection.**
For GT vectors pics look the same

Green fluorescent protein (GFP)

autofluorescent protein from **Pacific Northwest jellyfish Aequorea victoria**

GFP is an extremely stable protein of 238 amino acids with unique post-translationally created and covalently-attached chromophore from oxidised residues 65-67, Ser-Tyr-Gly



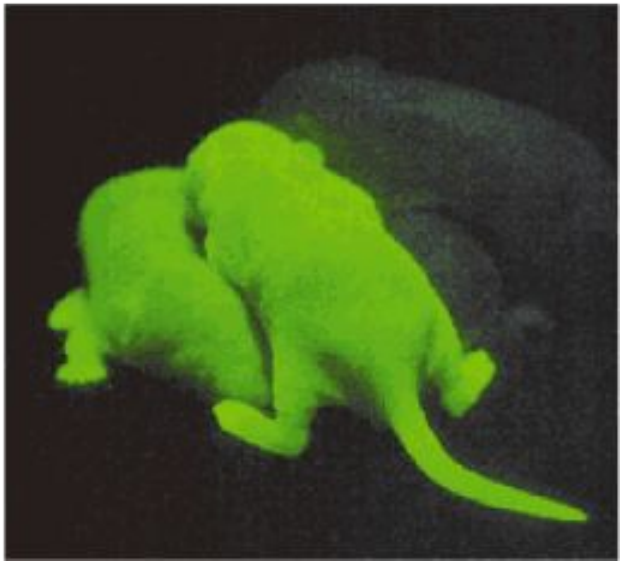
ultraviolet light causes GFP to autofluoresce In a bright green color

Jellyfish do nothing with UV, The activate GFP by aequorin (Ca++ activated, biolumuniscent helper)

GFP expression is harmless for cells and animals

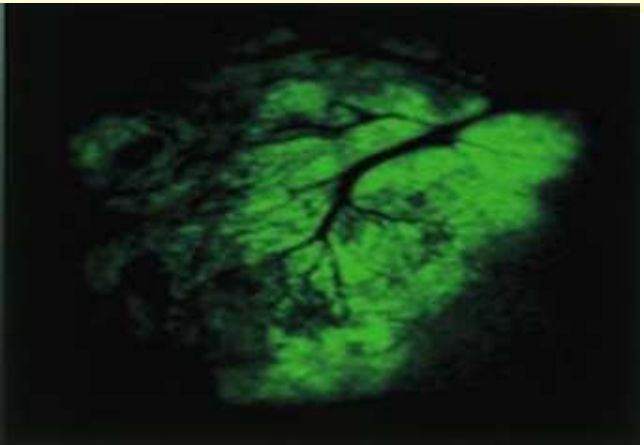
White Light

Fluorescent Light



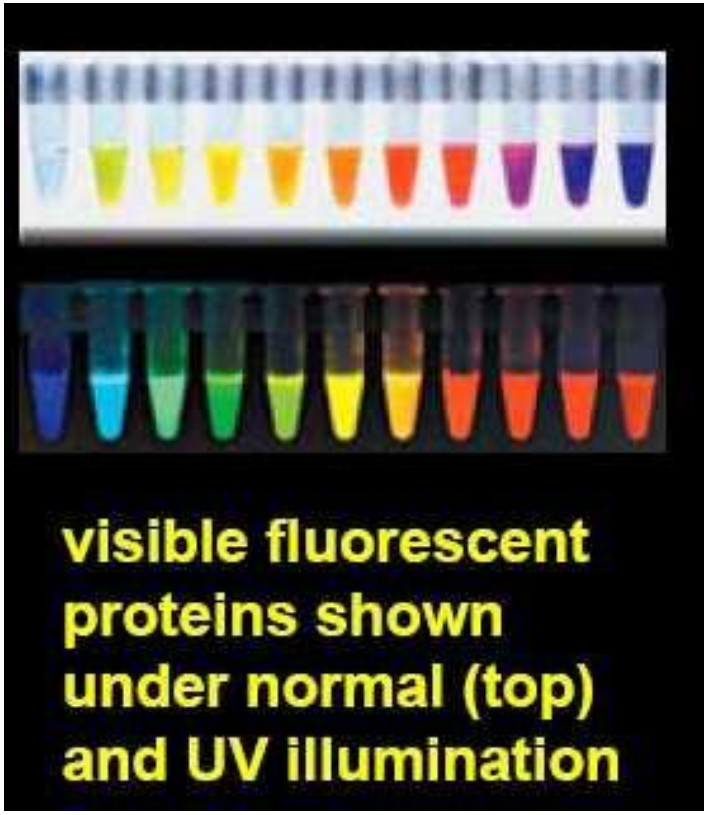
GFP transgenic mice from
Osaka University
(Masaru Okabe)

GFP construct could be used for construct tracking in living organism



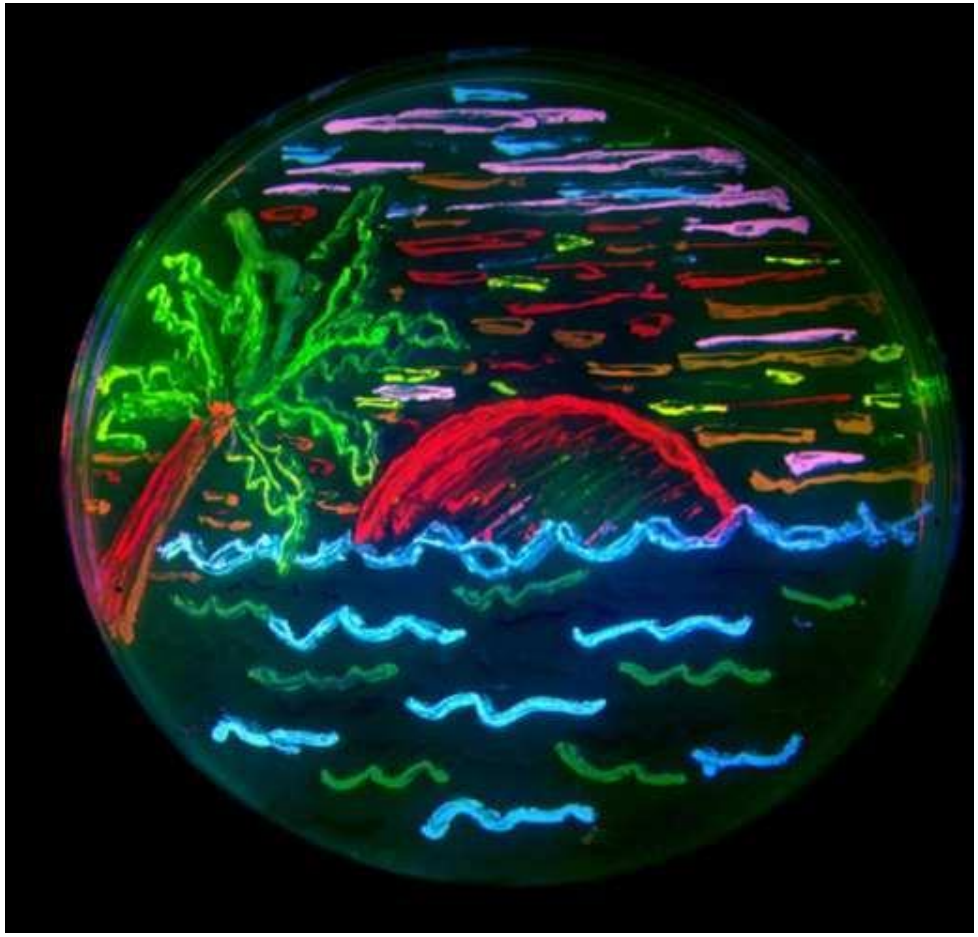
**GFP labelled image of a human tumor.
Vessel on the tumor surface
are visible in black**

Many more fluorescent proteins are engineered

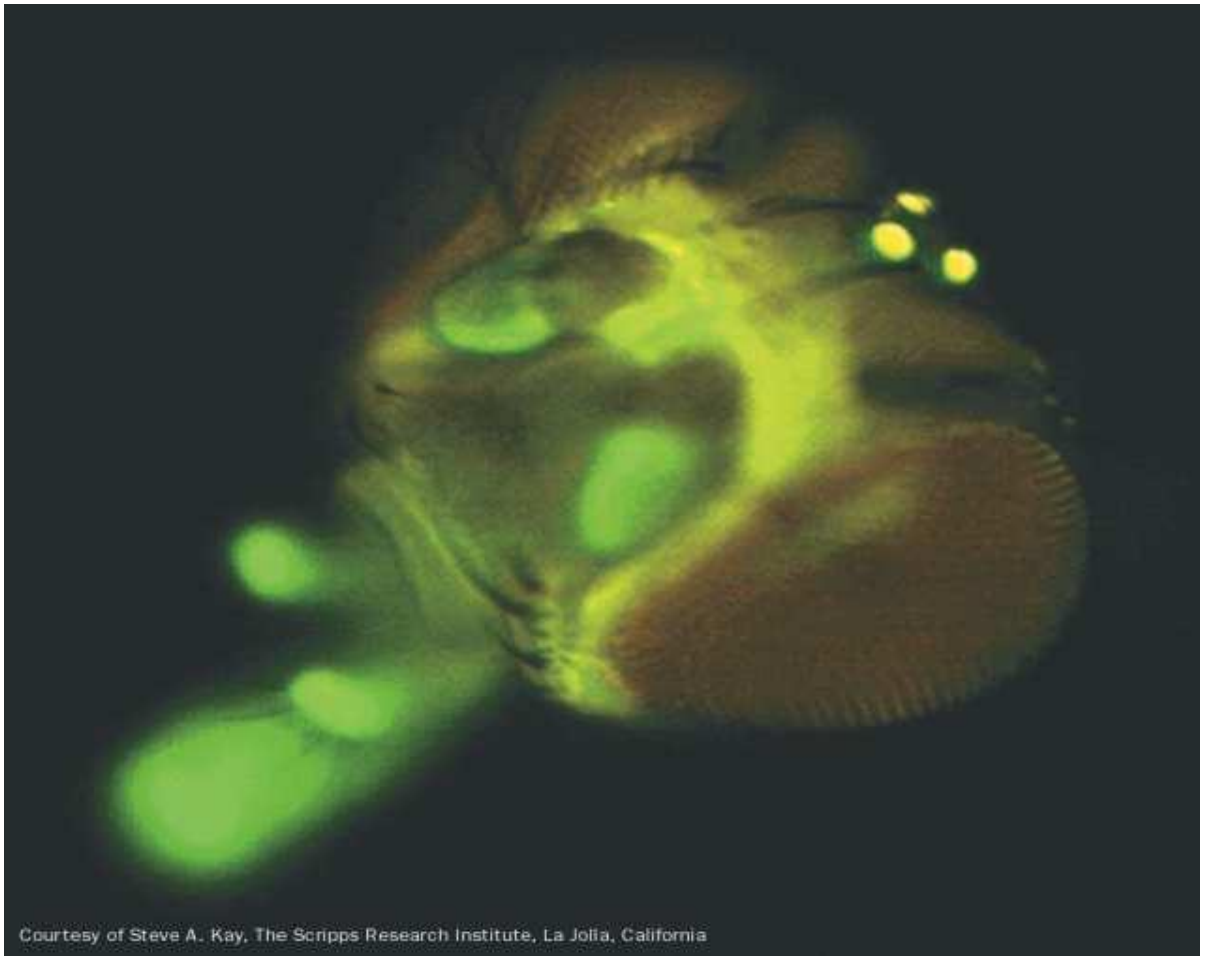


Chem. & Eng. News, Oct. 25, 2004

Engineered proteins are covering all the spectrum

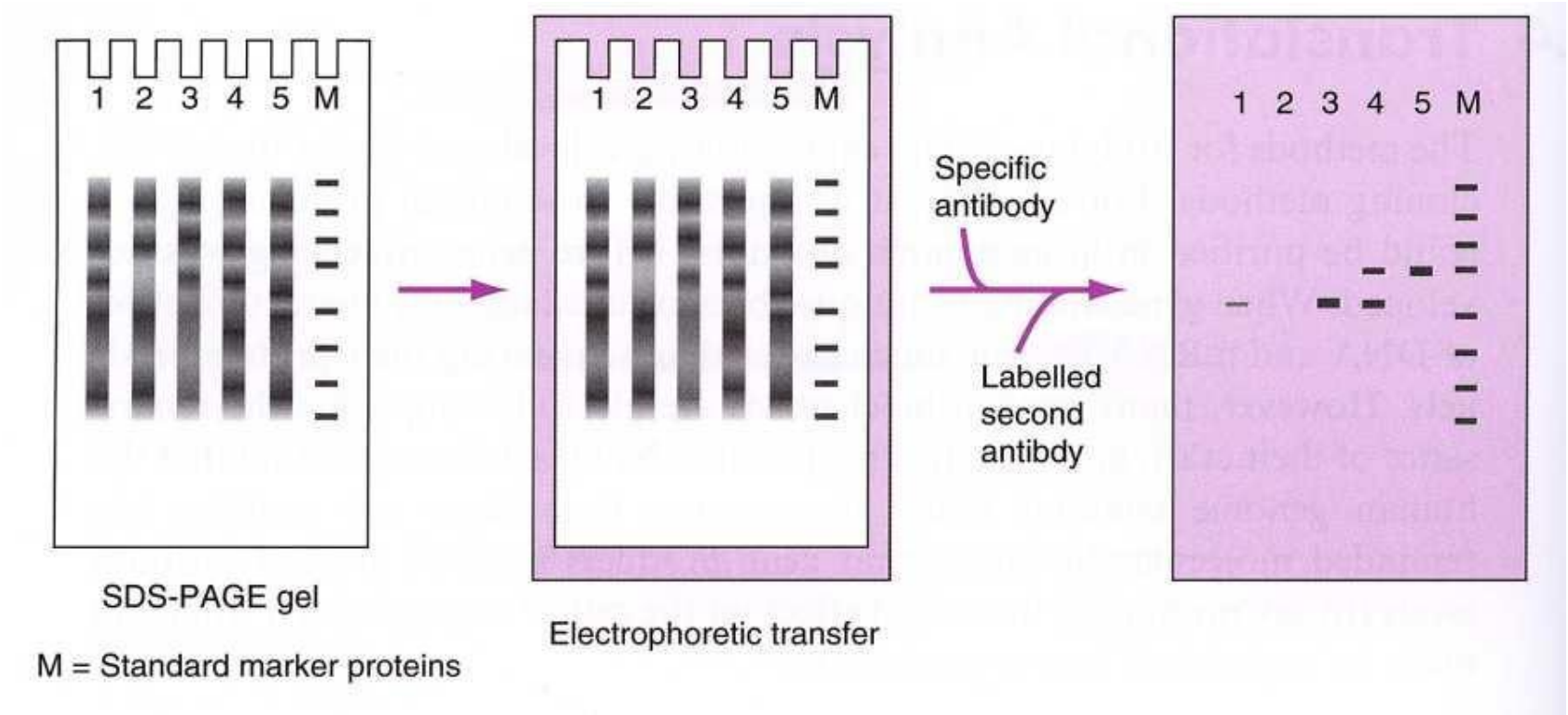


San Diego beach scene drawn with living bacteria expressing 8 different colors of fluorescent proteins.

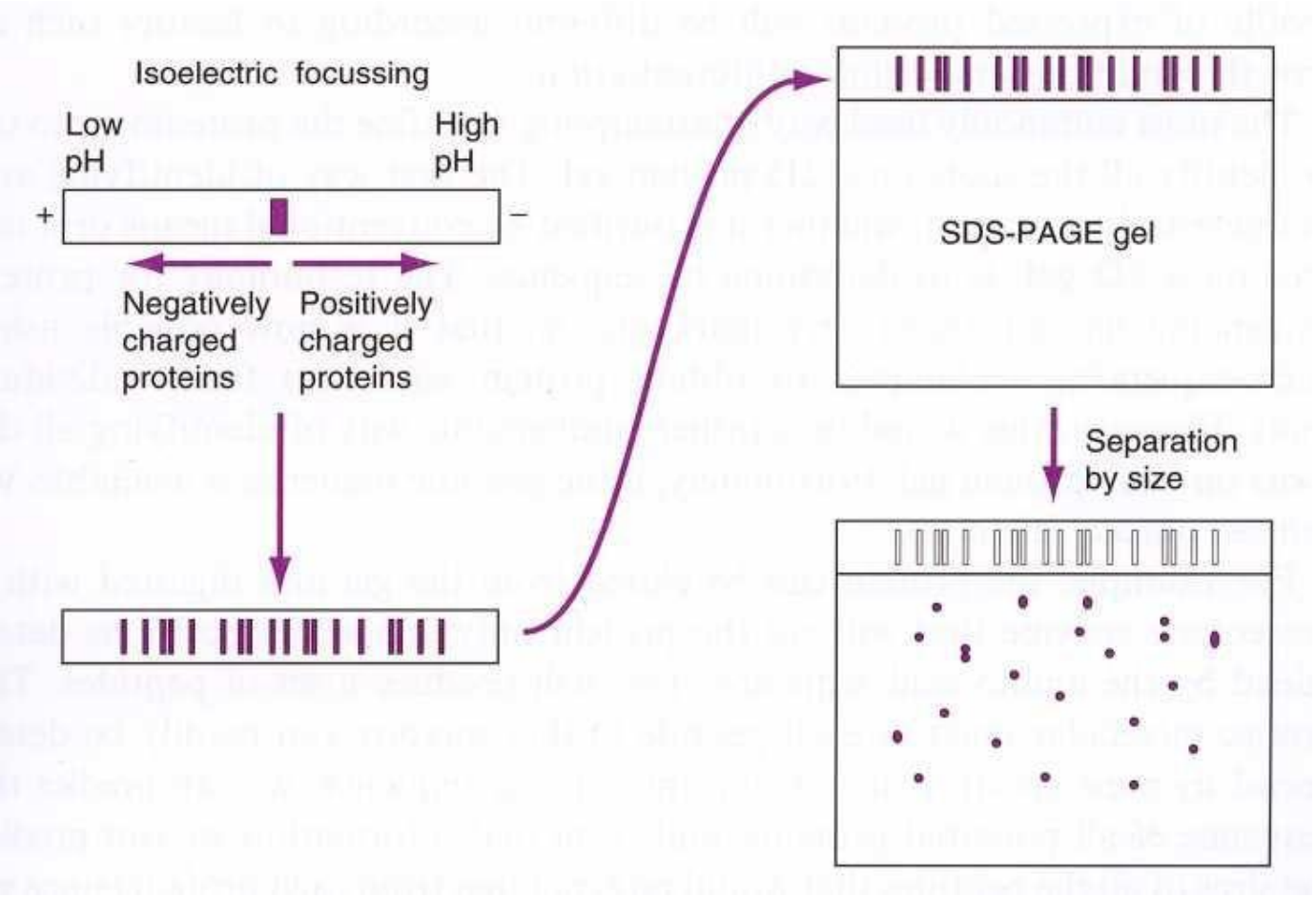


Use of green fluorescent protein (GFP) as a reporter gene.

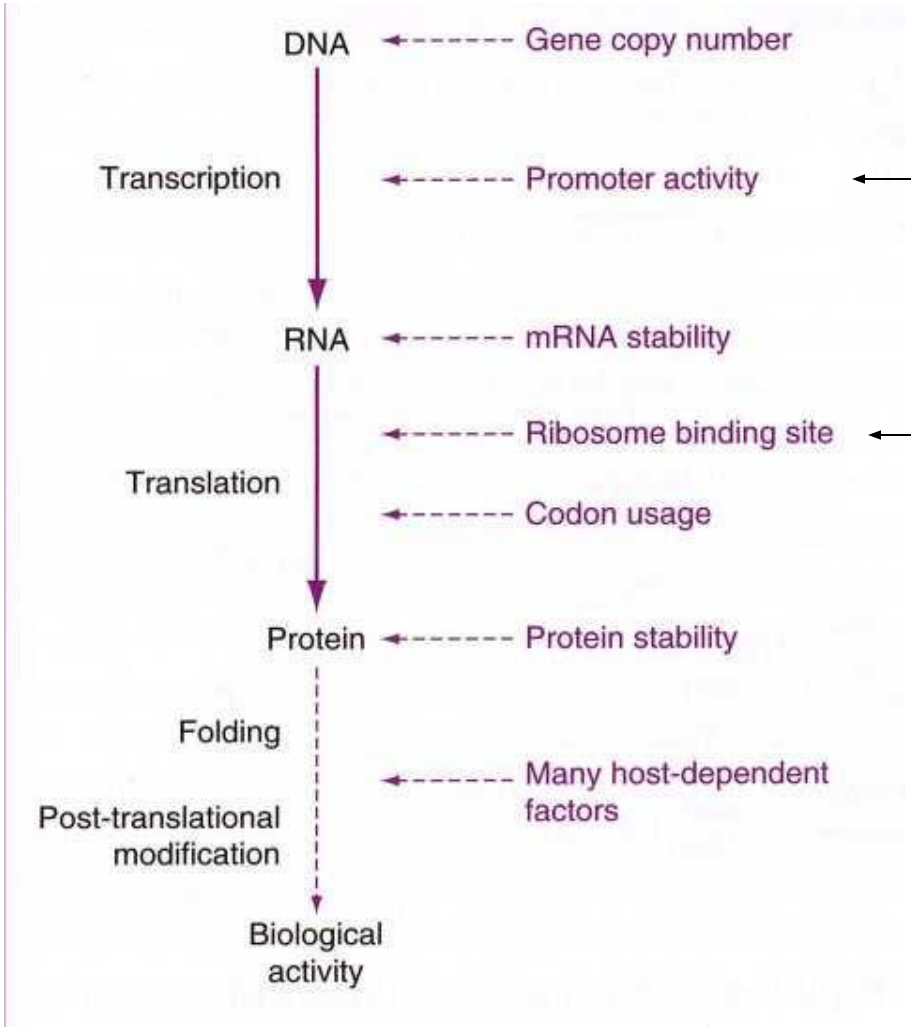
Analyzing Translation – Western Blot



2 D Electrophoresis



Gene Expression



Gene Expression

Gene copy number:

1. Plasmid copy number:

The copy-number of a plasmid in the cell is determined by regulating the **initiation of plasmid replication**.

The initiation of plasmid replication may be controlled by:

- the amount of available primer (RNA)
- the amount of essential replication proteins
- the function of essential replication proteins.

2. Gene dosage -> number of genes integrated into chromosome

- prokaryotic systems -> i.e. Transposons, phages, recombinantion
- mainly eukaryotic systems

Table 1. Origins of Replication and Copy Numbers of Various Plasmids and Cosmids

DNA construct	Origin of replication	Copy number	Classification
Plasmids			
pUC vectors	pMB1*	500–700	high copy
pBluescript® vectors	ColE1	300–500	high copy
pGEM® vectors	pMB1*	300–400	high copy
pTZ vectors	pMB1*	>1000	high copy
pBR322 and derivatives	pMB1*	15–20	low copy
pACYC and derivatives	p15A	10–12	low copy
pSC101 and derivatives	pSC101	~5	very low copy
Cosmids			
SuperCos	pMB1	10–20	low copy
pWE15	ColE1	10–20	low copy

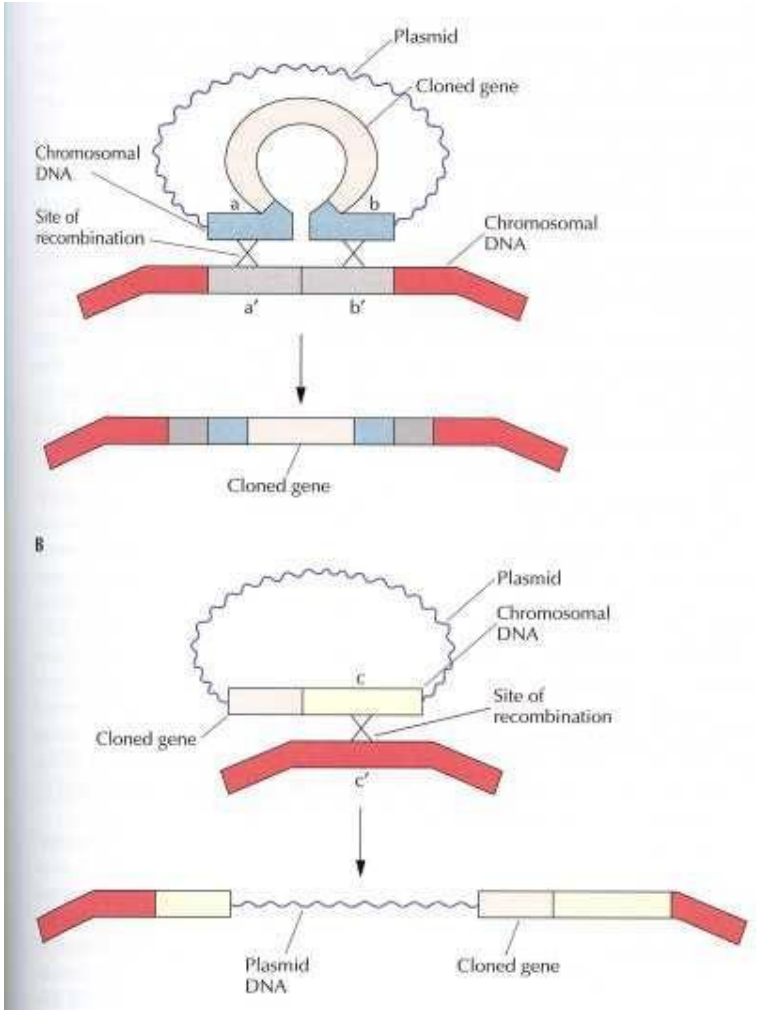
* The pMB1 origin of replication is closely related to that of ColE1 and falls in the same incompatibility group. The high-copy plasmids listed here contain mutated versions of this origin.

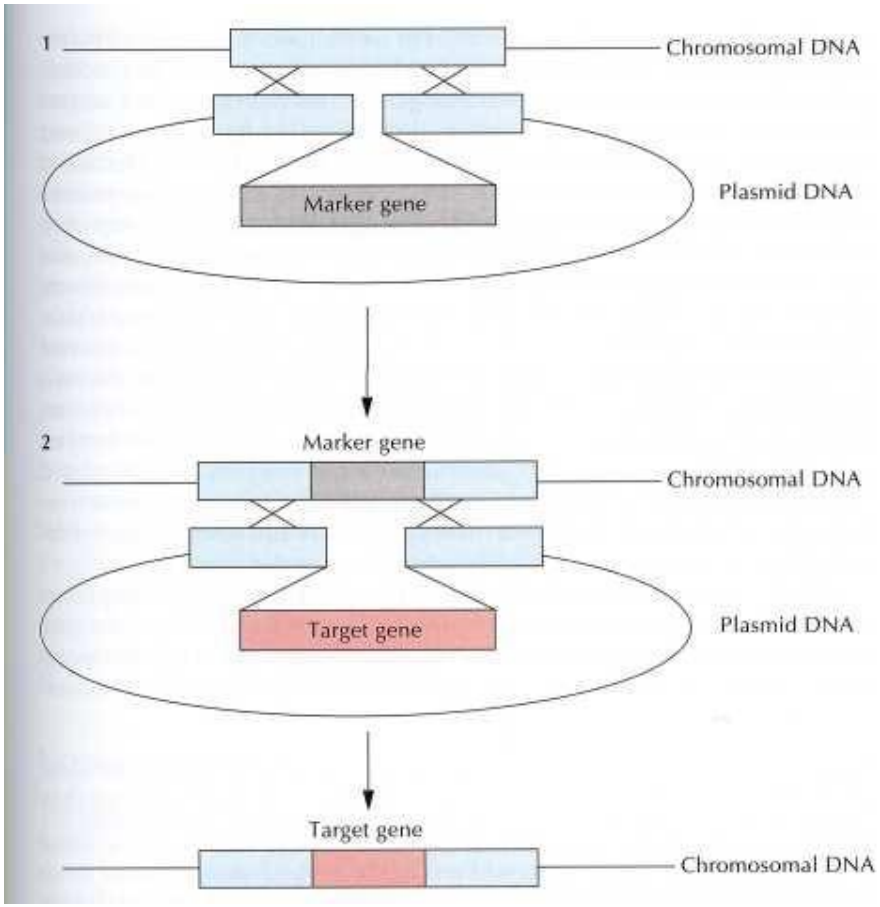
Incompatibility of plasmids:

Not all plasmids are able to coexist in the same cell.
Plasmids which have the same replication control functions are incompatible, and are assigned to the same incompatibility group (inc group).
Plasmids of one incompatibility group are related to each other, but cannot survive together in the same bacterial cell, as only different kinds of plasmids are compatible.

Ensures that we can make libraries -> just one plasmid taken up by one cell

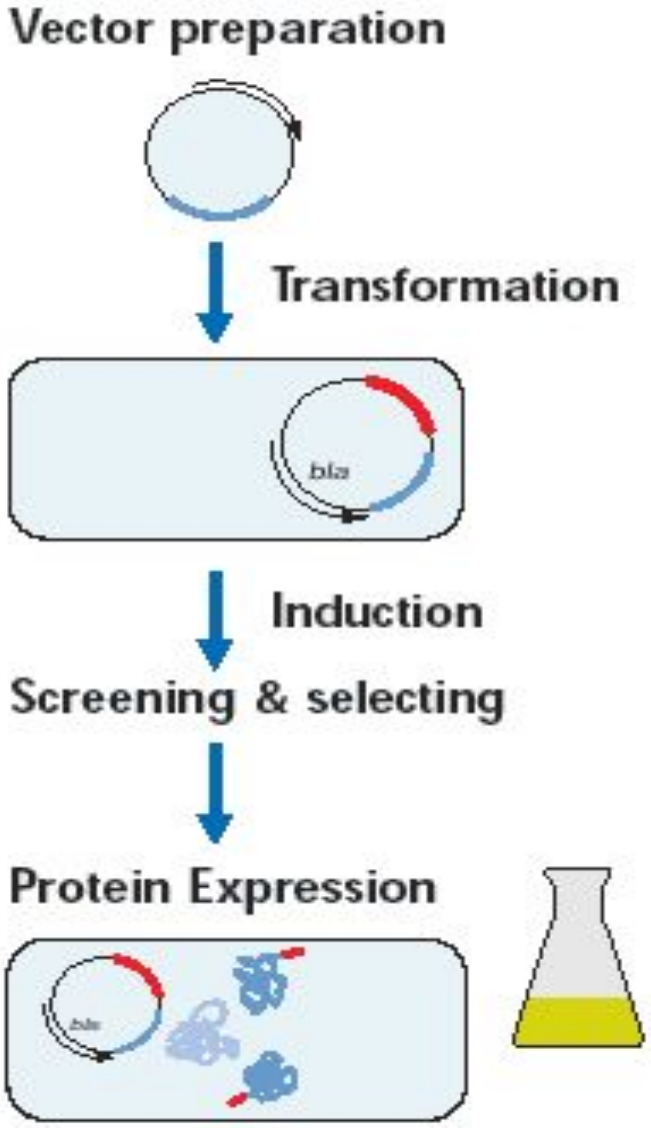
Homologous integration into chromosome





Insertion on *Bacillus subtilis* chromosome

Protein expression in prokaryotic systems



So, this new story would be about vectors again.

Bacterial expression vectors have some distinct features:

Inducible promoter systems;

Protein fusions including fused tags;

General advices for one who wants to produce gene expression in prokaryotes

Most obvious and common mistakes:

1. Do not forget to cut out the intron

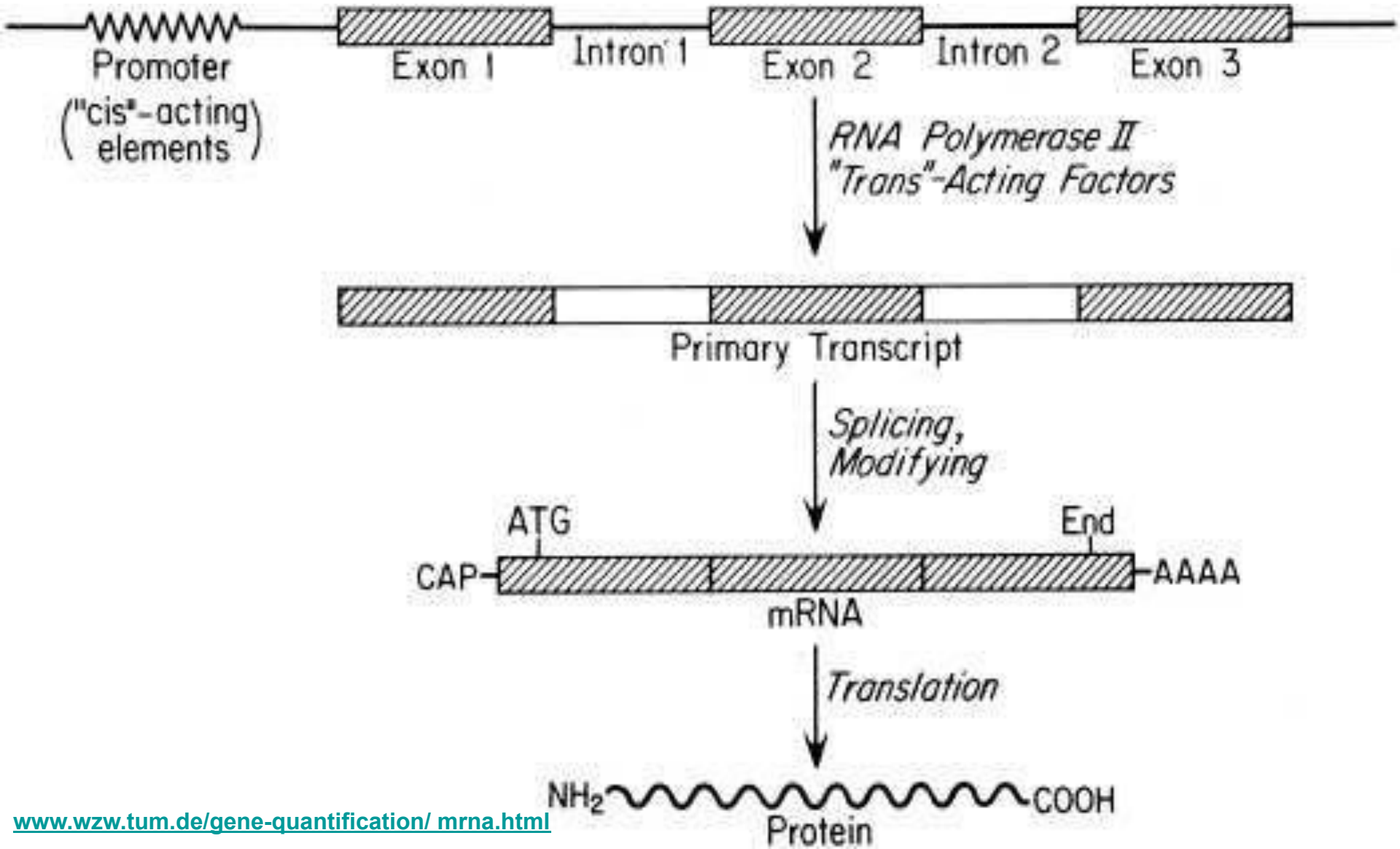
2. Check orientation of insert

3. Do fusions with something In-frame

4. No Post-translation modification = no product activity

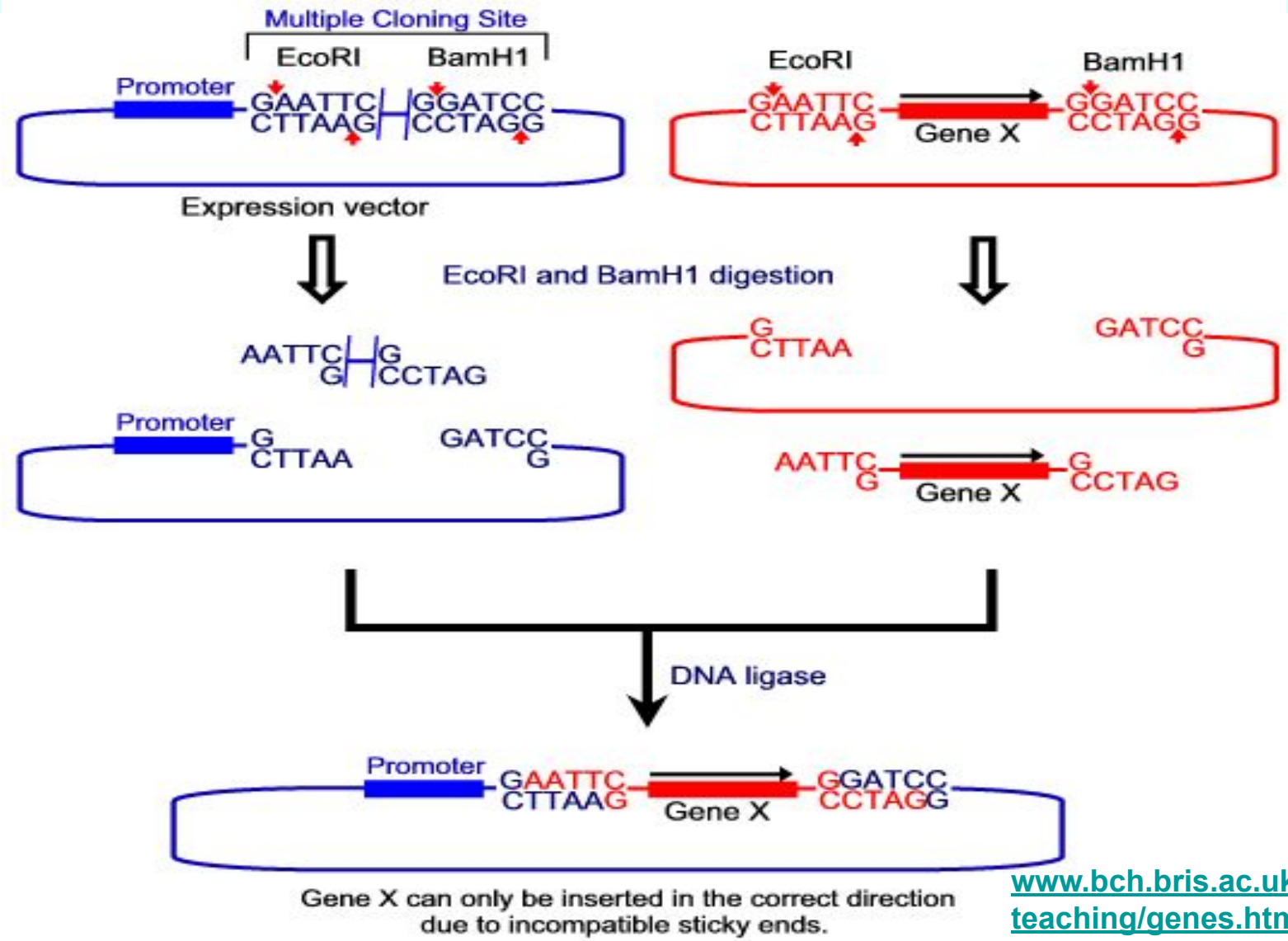
Introns

Not an issue when you clone a cDNA



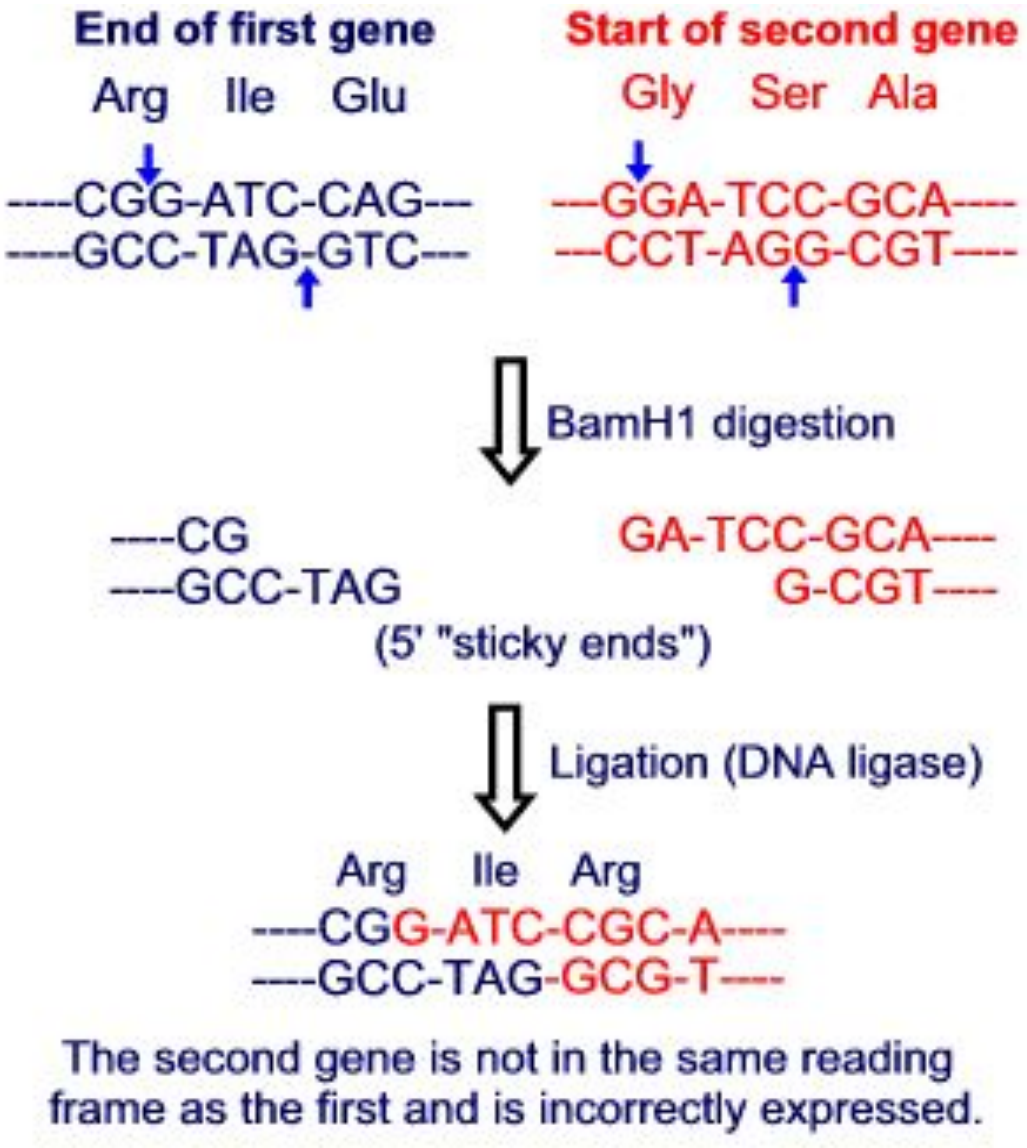
www.wzw.tum.de/gene-quantification/ mrna.html

Orientation of insert (could go backward, if cloned with same-type sticky ends) – use incompatible sticky ends



www.bch.bris.ac.uk/staff/ pfdg/ teaching/genes.htm

Fusion proteins.



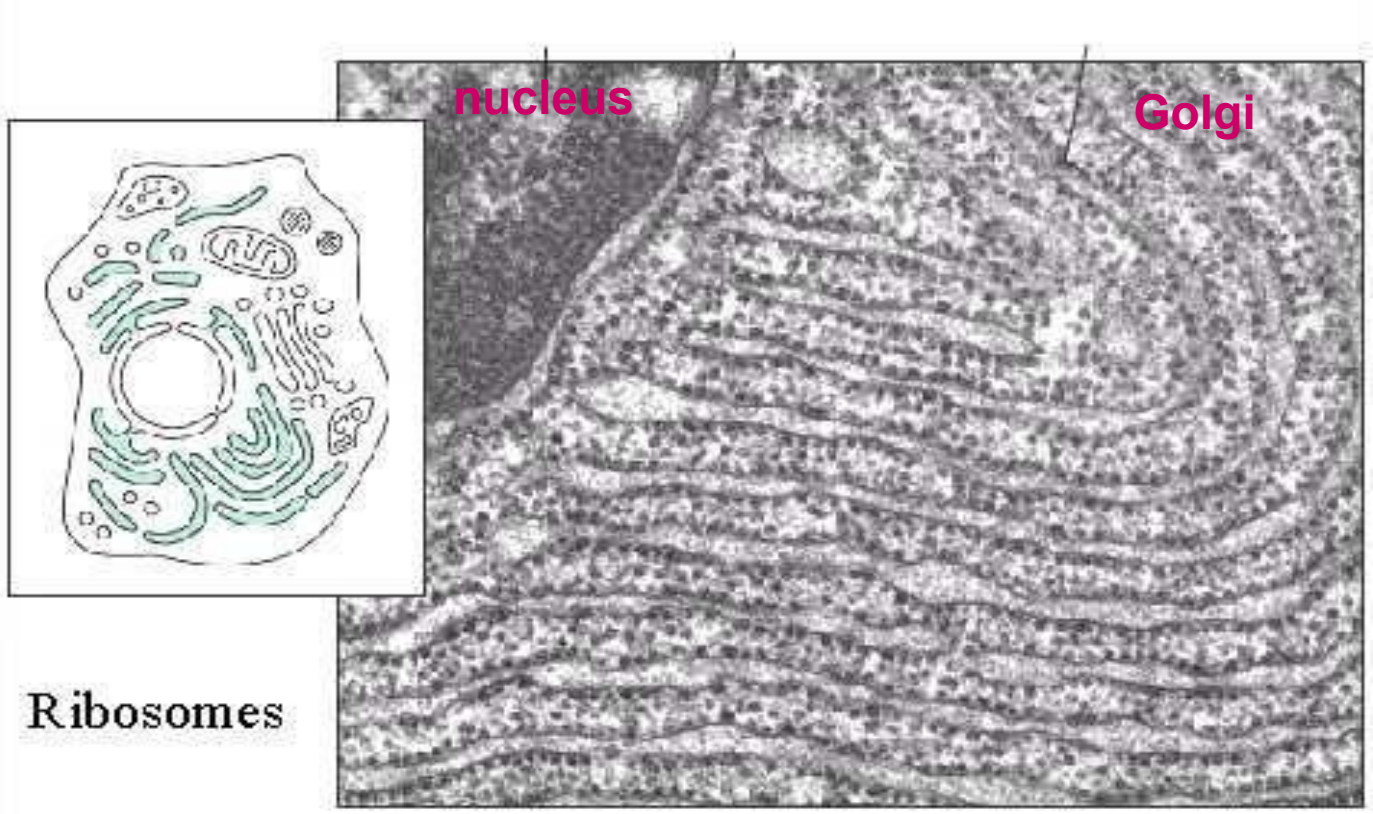
When expressing a fusion proteins, ensure that both of them are in the same reading frame

www.bch.bris.ac.uk/staff/pfdg/teaching/genes.htm

PostTranslational modification

Eukaryotic cells have Golgi system

Prokaryotic cells do not have it



Synthesis of secreted and membrane proteins

Efficiency of expression in E.coli

Dependent of:

1. **Type of transcription promoter and terminator**
2. **Affinity of mRNA and prokaryotic ribosome**
3. **Amount of copies of transgene and its localization (chromosome or plasmid)**
4. **Cellular localisation of the protein end-product**
5. **Efficiency of translation in the host organism**
6. **Stability of protein product in the host organism**

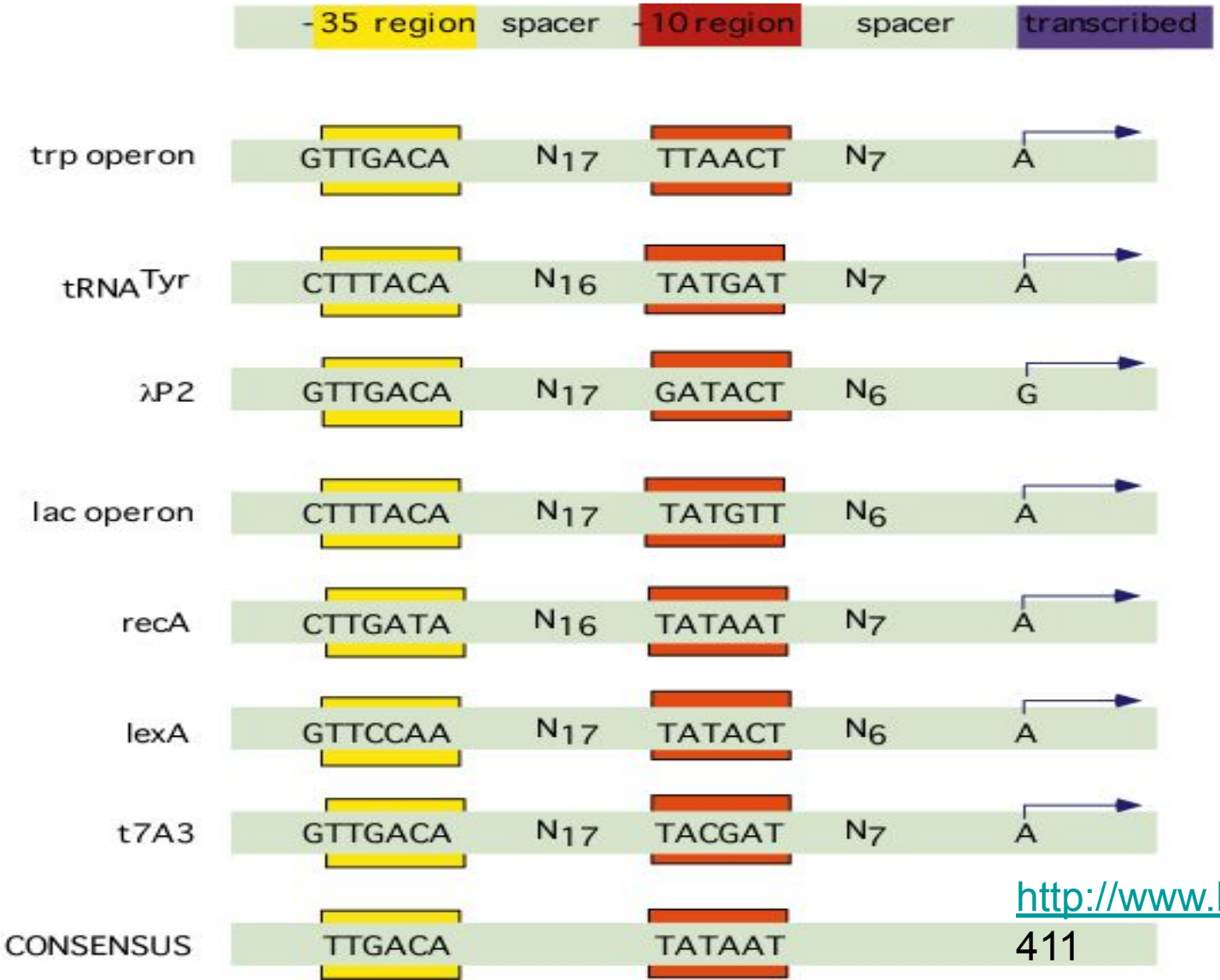
Systems could be optimized on gene to gene basis.
No universal strategy possible

Factors affecting transcription

1. Promoters (including regulated ones)
PROKARYOTIC!!!!

2. Terminators
PROKARYOTIC!!!!

Variations between prokaryotic promoters are minimal



http://www.blc.arizona.edu/marty/411

Factors affecting translation

1. Ribosome binding site (RBS)

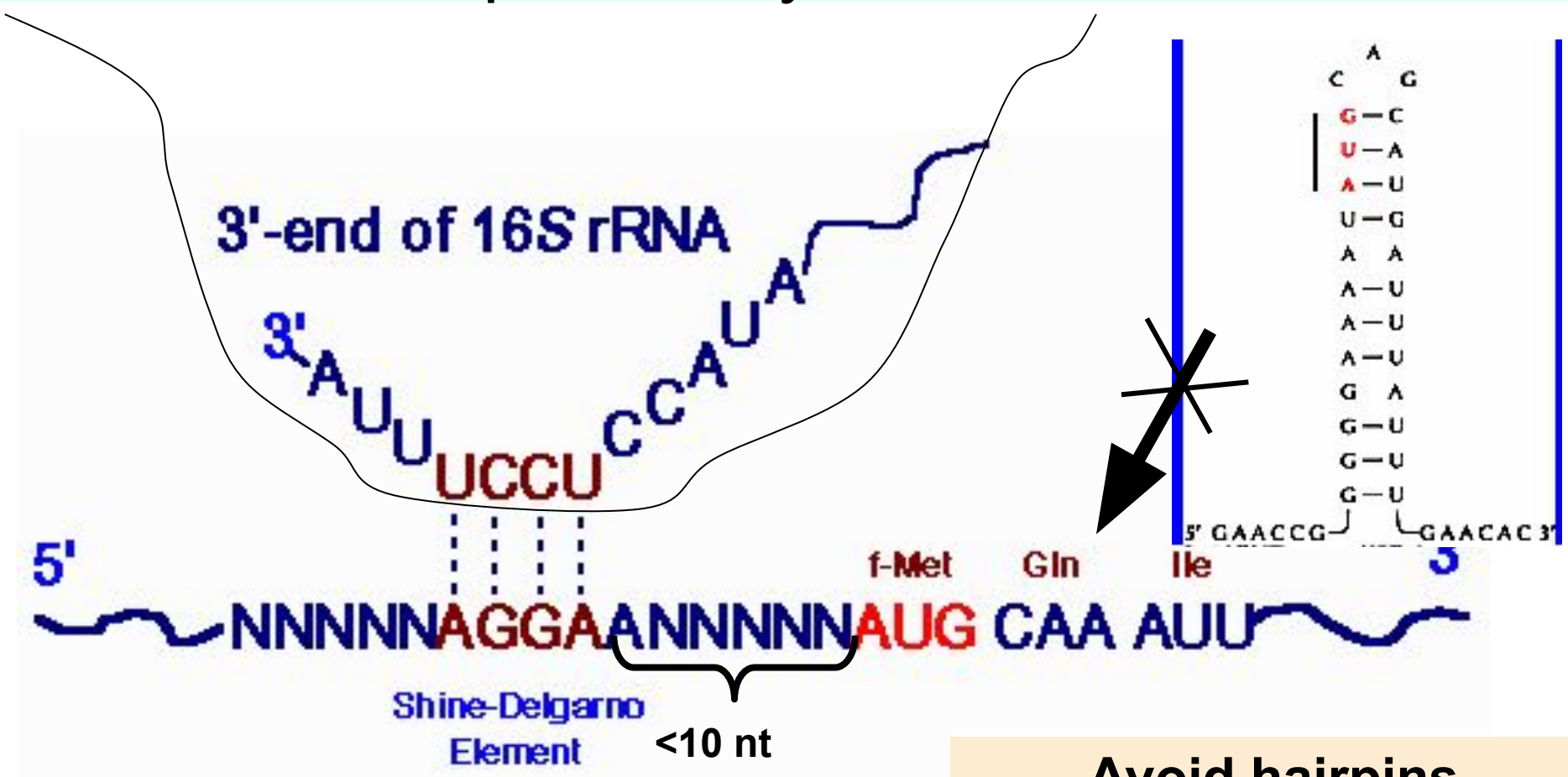
2. Codon bias

3. Stability of the transcript

Ribosome binding site (RBS) =

translation initiation site

complimentary to 16S rRNA



Examining the second codon;
better AAA – lysin (13.9% of all E.coli genes).
Expression can vary 15 times.

Avoid hairpins
on 5' end of gene
(minimize GC content)

Codon Usage in E. coli & humans

Codon	Amino acid	Frequency of use in:	
		E. coli	Humans
GAG	Glutamic acid	0.30	0.59
GAA	Glutamic acid	0.70	0.41
CGG	Arginine	0.08	0.19
CGA	Arginine	0.05	0.10
CGU	Arginine	0.42	0.09
CGC	Arginine	0.37	0.19
AGG	Arginine	0.03	0.22
AGA	Arginine	0.04	0.21
CCG	Proline	0.55	0.11
CCA	Proline	0.20	0.27
CCU	Proline	0.16	0.29
CCC	Proline	0.10	0.33
UGA	Stop	0.30	0.61
UAG	Stop	0.09	0.17
UAA	Stop	0.62	0.22

Codon Optimization Strategies

- Chemically synthesize new gene
  - Alter sequence of the gene of interest to match donor codons to the codons most frequently used in host organism
- Express in different host
  - choose host with better matching codon usage
- Use an engineered host cell that overexpresses low abundance tRNAs

Commercial *E. coli* strains encode for a number of the rare codon genes

BL21 (DE3) CodonPlus-RIL (AT-rich compatible)	arginine (**AGG**, **AGA**), isoleucine (**AUA**) and leucine (**CUA**)
BL21 (DE3) CodonPlus-RP (GC-rich compatible)	arginine (**AGG**, **AGA**) and proline (**CCC**)
(AT-rich compatible) Rosetta or Rosetta (DE3)	**AGG/AGA** (arginine), **CGG** (arginine), **AUA** (isoleucine) **CUA** (leucine)**CCC** (proline), and **GGA** (glycine)

Mitochondria and chloroplast genes

Alterations in the Standard Genetic Code in Mitochondria

CODON	Standard Code: Nuclear-Encoded Proteins	Mammals	Mitochondria: *Drosophila*	Mitochondria: *Neurospora*	Mitochondria: Yeasts	Mitochondria: Plants
UGA	Stop	Trp	Trp	Trp	Trp	Stop
AGA, AGG	Arg	Stop	Ser	Arg	Arg	Arg
AUA	Ile	Met	Met	Ile	Met	Ile
AUU	Ile	Met	Met	Met	Met	Ile
CUU, CUC, CUA, CUG	Leu	Leu	Leu	Leu	Thr	Leu

Factors affecting protein stability

1. Overall level of protease activity in bacterial cells

2. N-terminal amino acid affects protein half-life

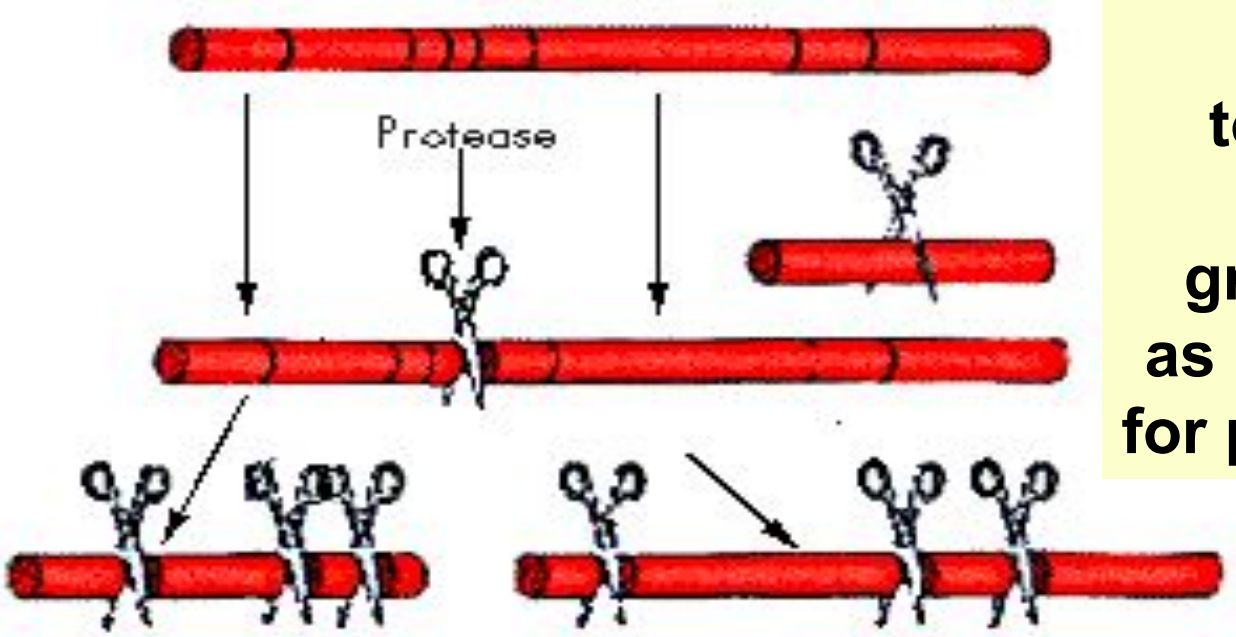
3. Internal regions containing clusters of certain amino acids can increase proteolysis

P proline
E glutamic acid
S serine
T threonine

…. Mutate PEST aminoacids….

Protease-deficient host strains

BL21, the work horse of *E. coli* expression, is deficient in two proteases encoded by the *lon* (cytoplasmic) and *omp*T (periplasmic) genes.



It is dangerous to kill proteases, it makes E.coli grow much slowly as proteases needed for proper metabolism

Inducible bacterial promoters

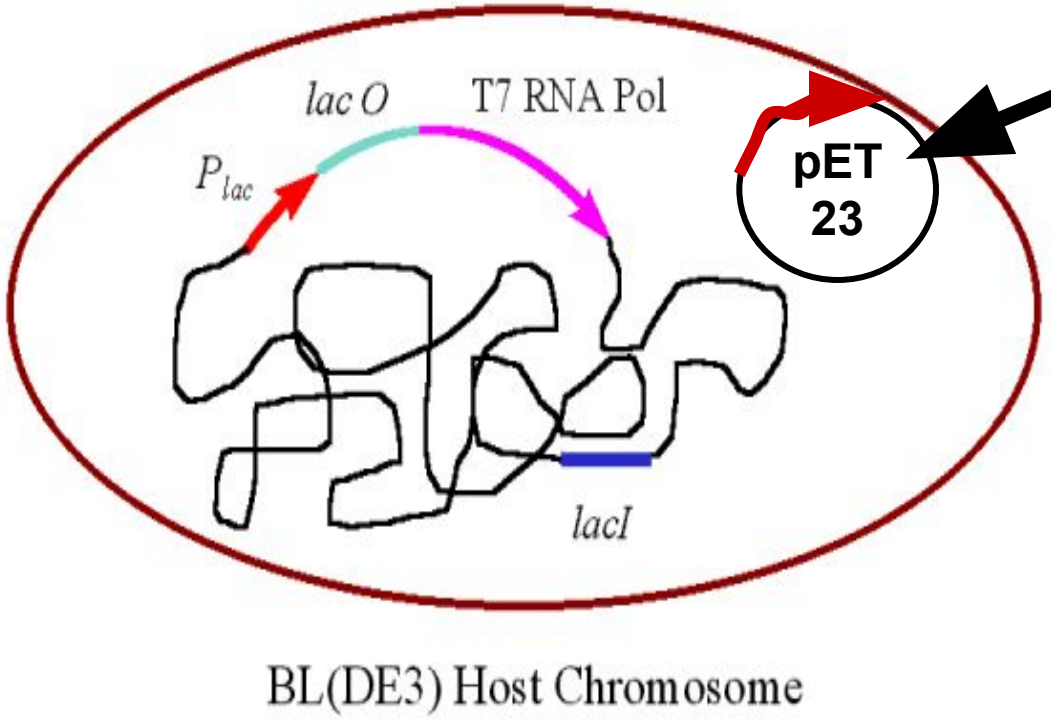
Why not to use constitutive, always strong promoter?

Because recombinant (alien) protein is often toxic for bacterial cell.

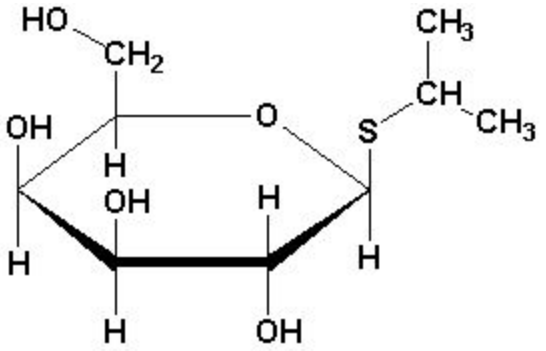
Bacteria tend to expel harmful plasmids

Induction

BL(DE3) inducible system and pET vectors (invented in 1984 by Bill Studier, on sale by Novagen)



Gene of interest is expressed from strong T7 promoter



Isopropyl Thiogalactoside (IPTG)

1) T7 RNA polymerase gene is integrated in chromosome under the control of a *lac* promoter and operator

2) lactose analogue, IPTG, causes the host to produce T7 RNA polymerase

•3) The *E. coli* host genome also carries the *lacI* (repressor) gene

Why repressor gene and gene of interest are expressed from different DNA molecules?

Repressor gene expressed from chromosome;
Gene of Interest expressed from plasmid

If too high repressor □ no transcription
(you need to increase expensive IPTG)

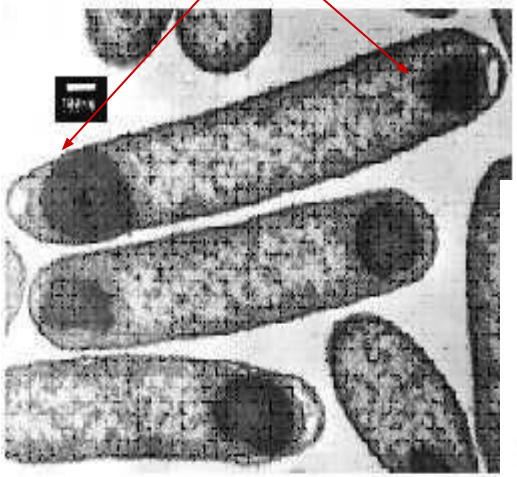
If too low repressor □ promoter is leaky
(active without IPTG)

Repressor is in chromosome,
because there it is best kept controlled there
(no plasmid loss, not too high expression)

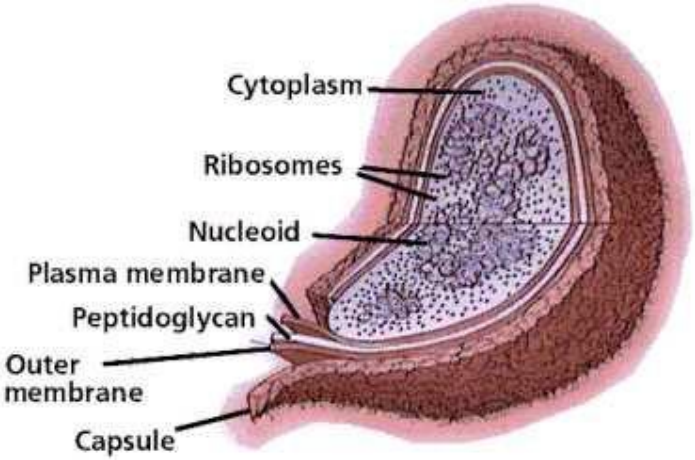
Where your expressed protein will be located?

Secreted (!!)
E.Coli
can not do that

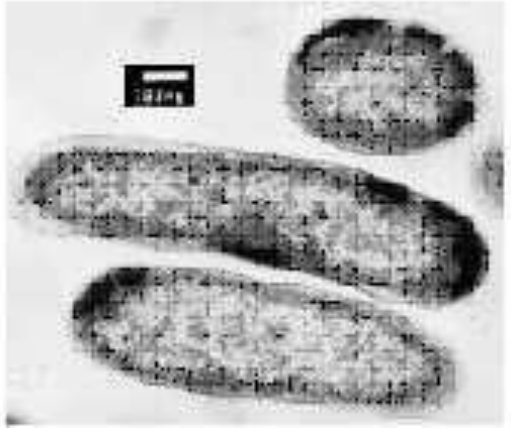
Inclusion bodies (insoluble)



Cytoplasm (soluble)



Periplasmatic space (soluble or insoluble)



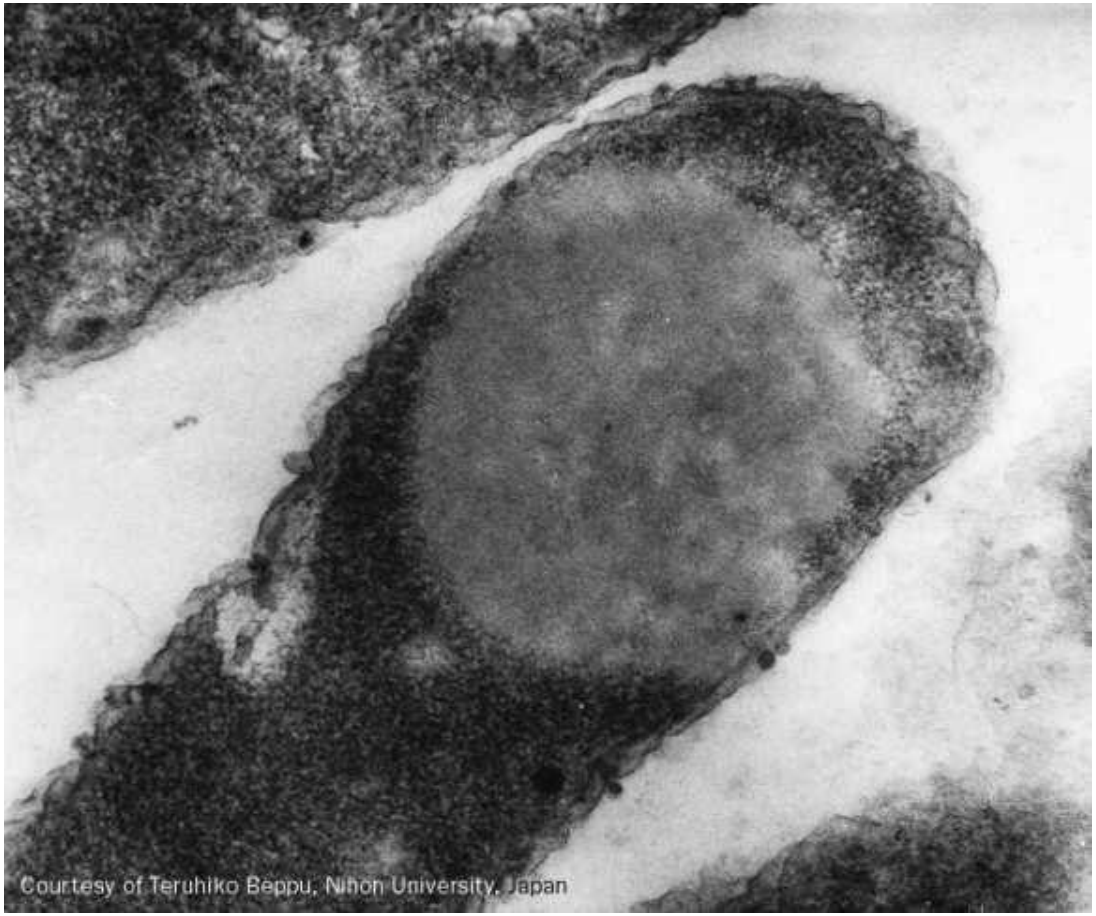
1. Inclusion bodies (most common case)

-- Inclusion bodies are formed through the accumulation of folding intermediates rather than from the native or unfolded proteins.

-- It is not possible to predict which proteins will be produced as inclusion bodies.

-- Production of inclusion bodies not dependent on the origin of protein, the used promoters, the hydrophobicity of target proteins...

Protein Folding



Electron micrograph of an inclusion body of the protein prochymosin in an *E. coli* cell

Good side of inclusion bodies

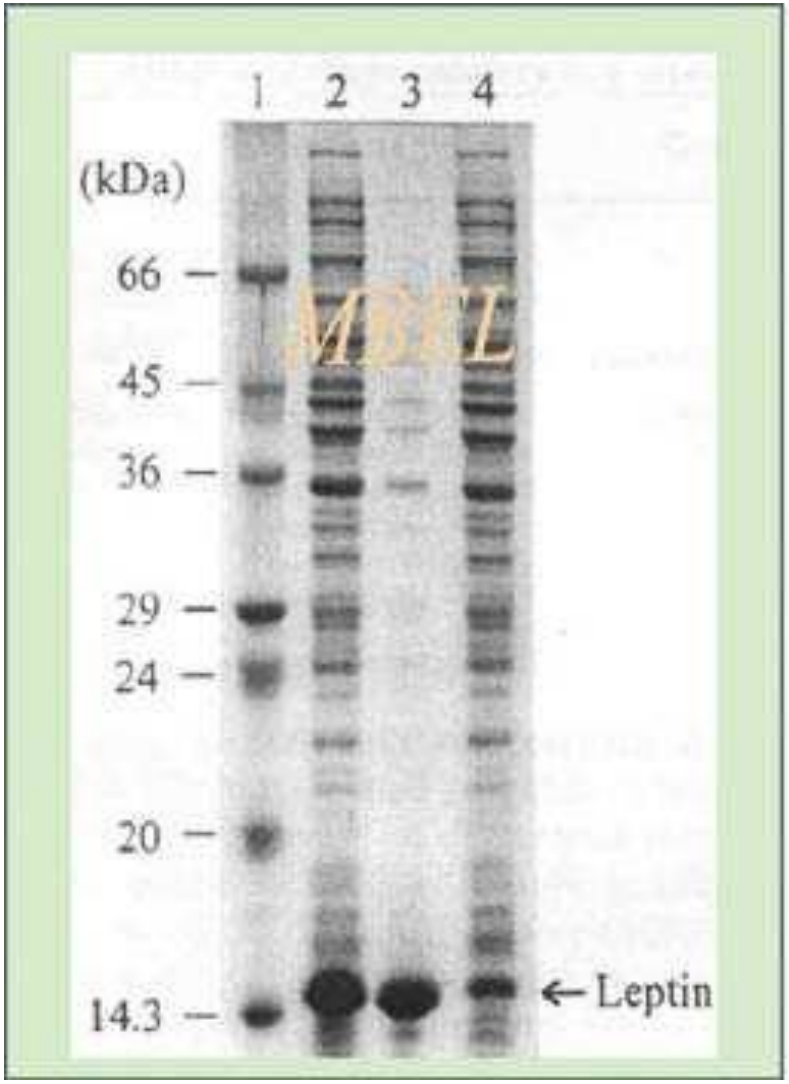
1) **inclusion bodies can be accumulated in the cytoplasm to much higher level (greater than 25%) than production as soluble form;**

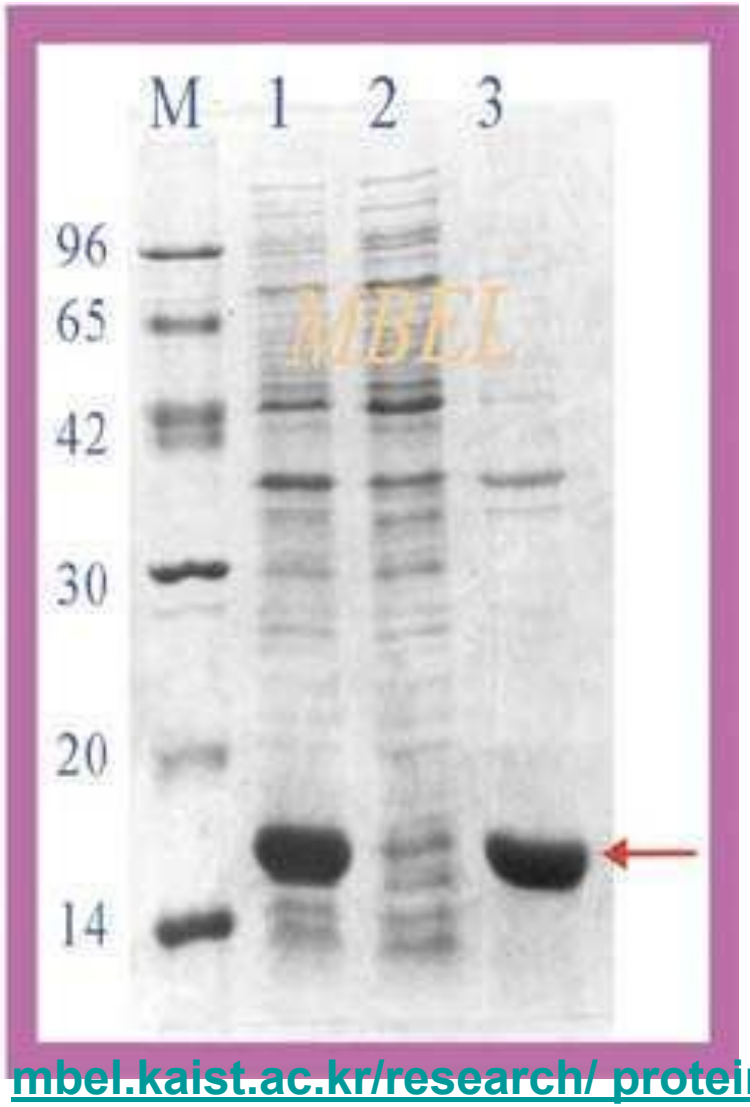
2) inclusion bodies is initially isolated in a highly purified, solid, and concentrated state by simple physical operation (centrifugation).

3) inclusion bodies have no biological activity. For toxic proteins it may be the only one available;

4) inclusion bodies are resistant to proteolysis That results in the high yield of protein production.

SDS-PAGE analysis of recombinant protein produced as inclusion body





hG-CSF

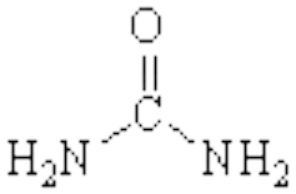
mbel.kaist.ac.kr/research/_protein_en1.html

Recovery of proteins from inclusion bodies

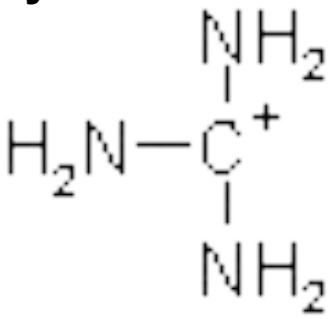
Is not a straightforward process, but road of trials and errors

Solubilization

Choice of solubilizing agents, e.g., urea, guanidine HCl, or detergents, plays a key role in solubilization efficiency



Urea



Guandinium

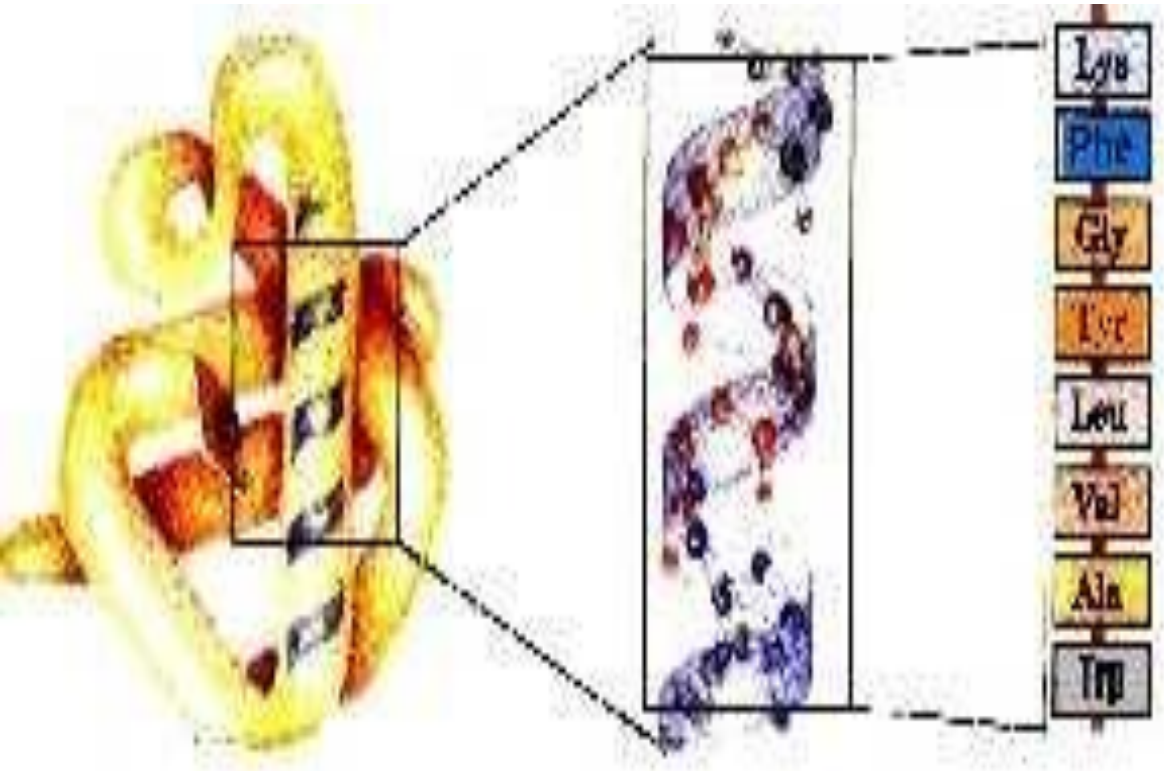
Refolding

-- Refolding is initiated by reducing concentration of denaturant used to solubilize IBs.

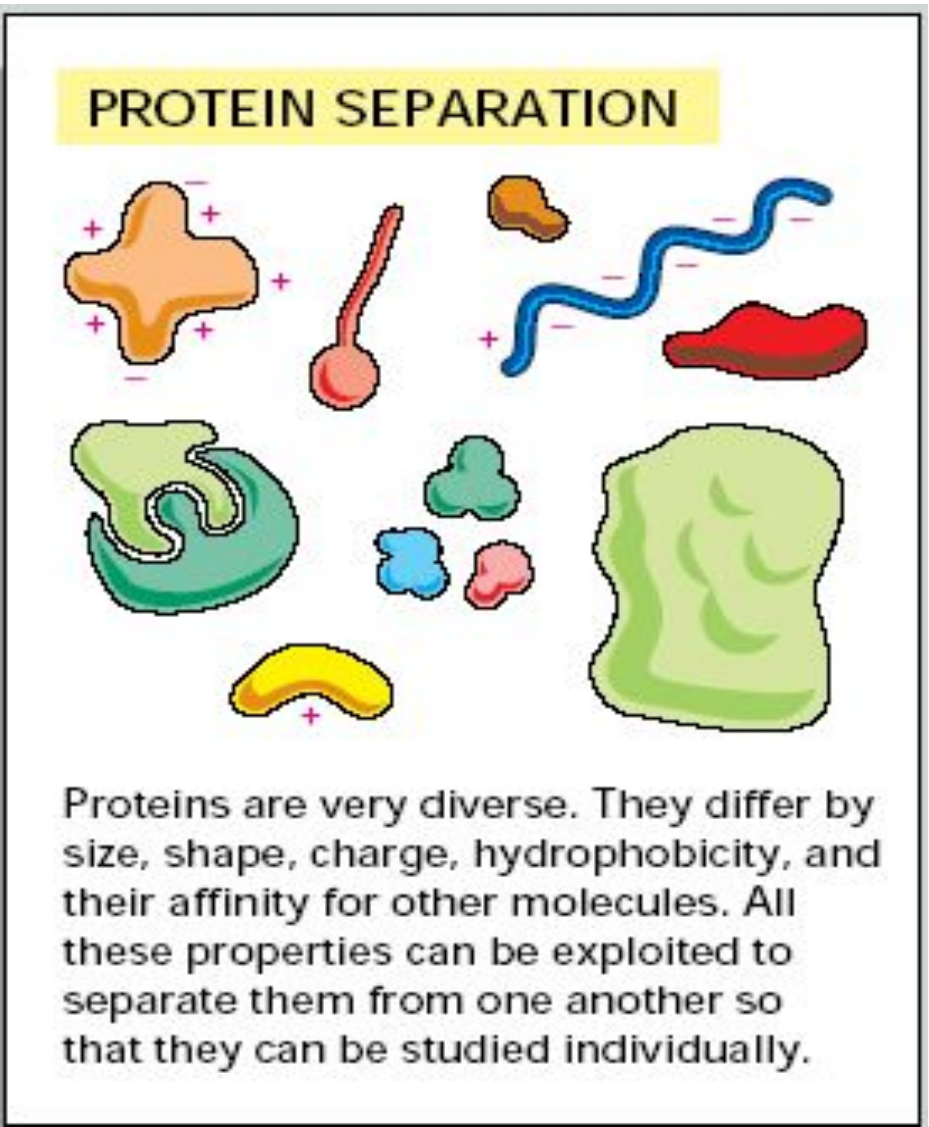
-- Refolding competes with other reactions, such as misfolding and aggregation (both are leading to bad results)

-- Chaperones are helpful in refolding (including chemical chaperones)

Question of questions – how to purify your protein?



Diversity of proteins could be exploited



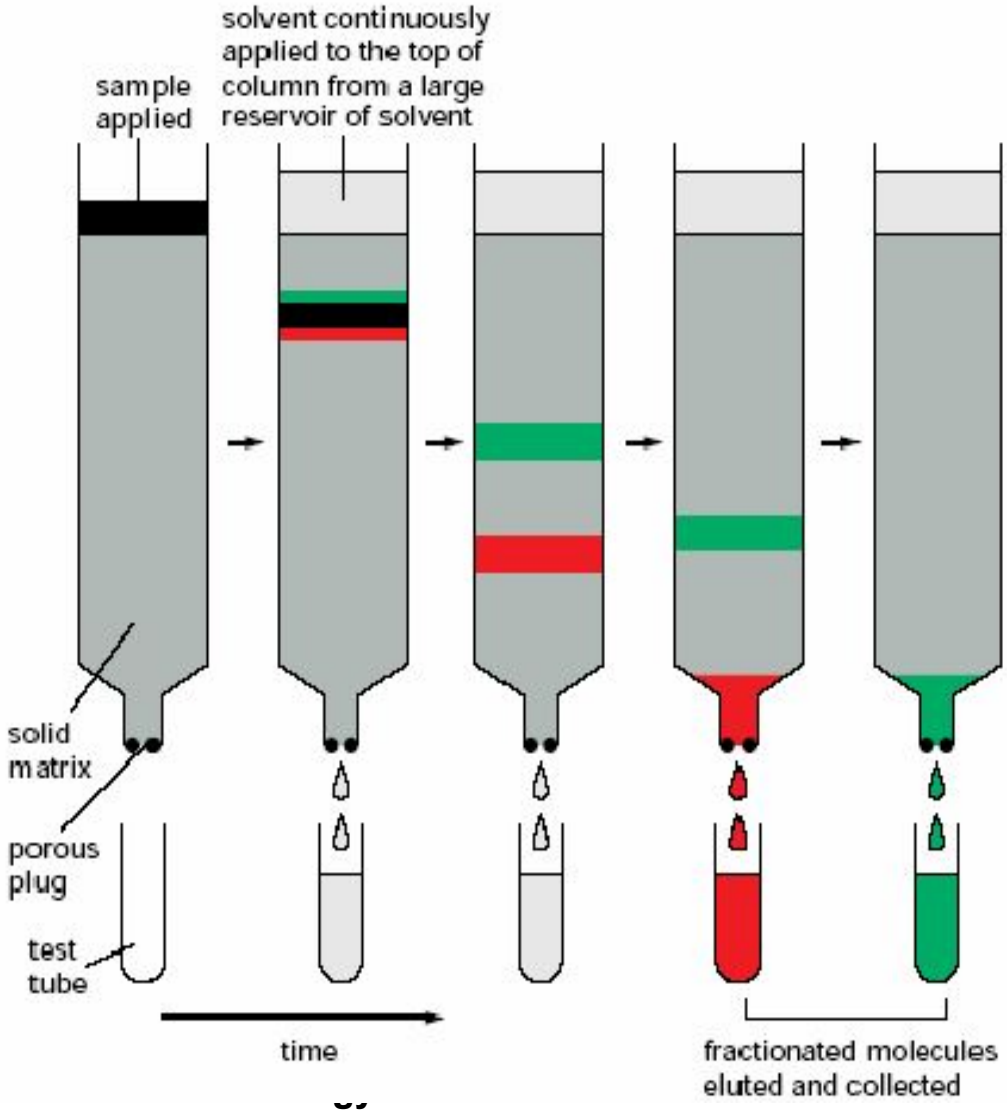
Column chromatography

Matrix particles usually packed in the column in the form of small beads.

A protein purification strategy might employ in turn each of the three kinds of matrix described below,

with a final protein purification Of up to 10,000-fold.

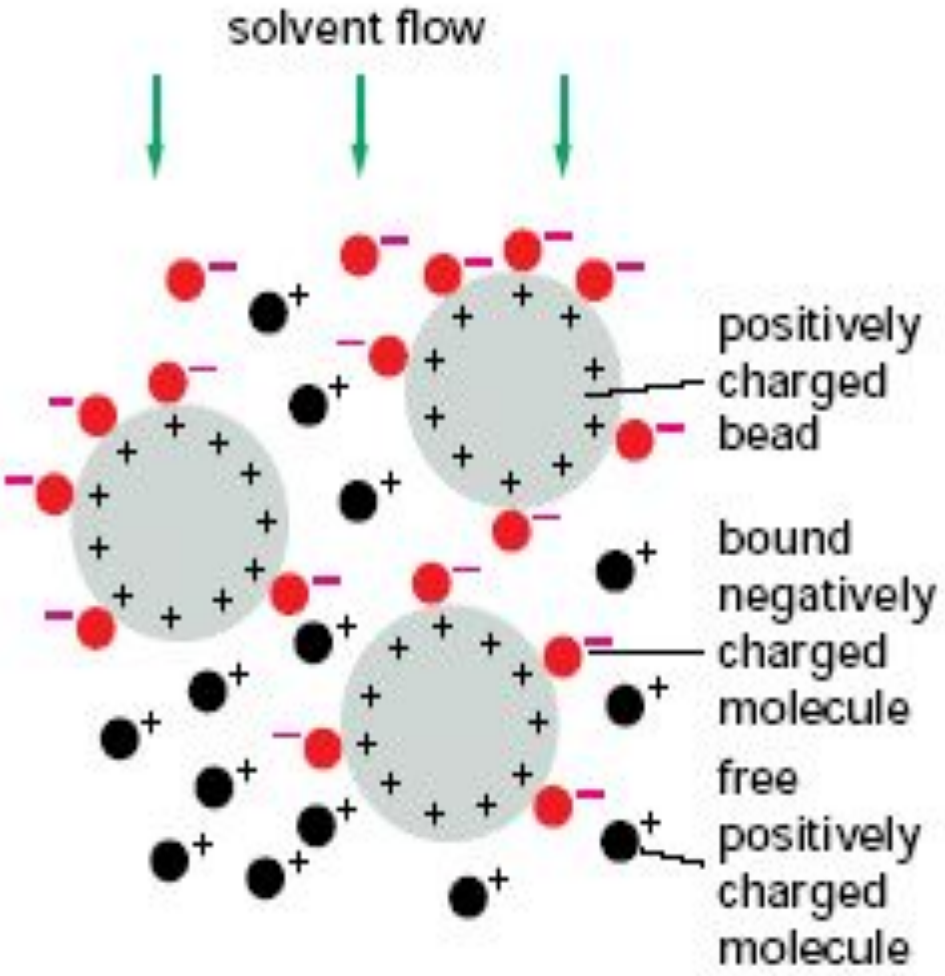
Column chromatography



Different proteins are retarded to different extents by their interaction with the matrix, they can be collected separately as they flow out from the bottom.

According to the choice of matrix, proteins can be separated according to
-- their charge,
-- their hydrophobicity,
-- their size,
-- their ability to bind to particular chemical groups (!!)

(A) ION-EXCHANGE CHROMATOGRAPHY

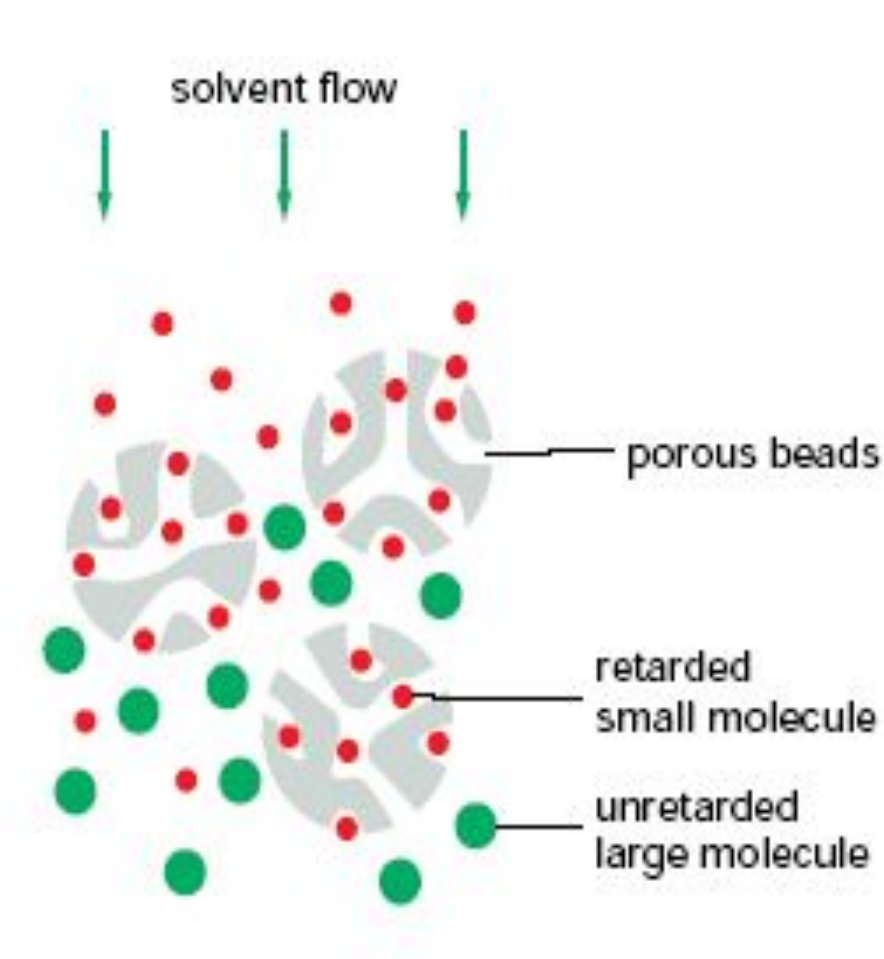


Ion-exchange columns
are packed with small beads that carry positive or negative charges retarding proteins of the opposite charge.

The association between a protein and the matrix depends on the pH and ionic strength of the solution passing down the column.

These can be varied in a controlled way to achieve an effective separation.

(B) GEL-FILTRATION CHROMATOGRAPHY



Gel-filtration columns
separate proteins
according to their size
on tiny porous beads.

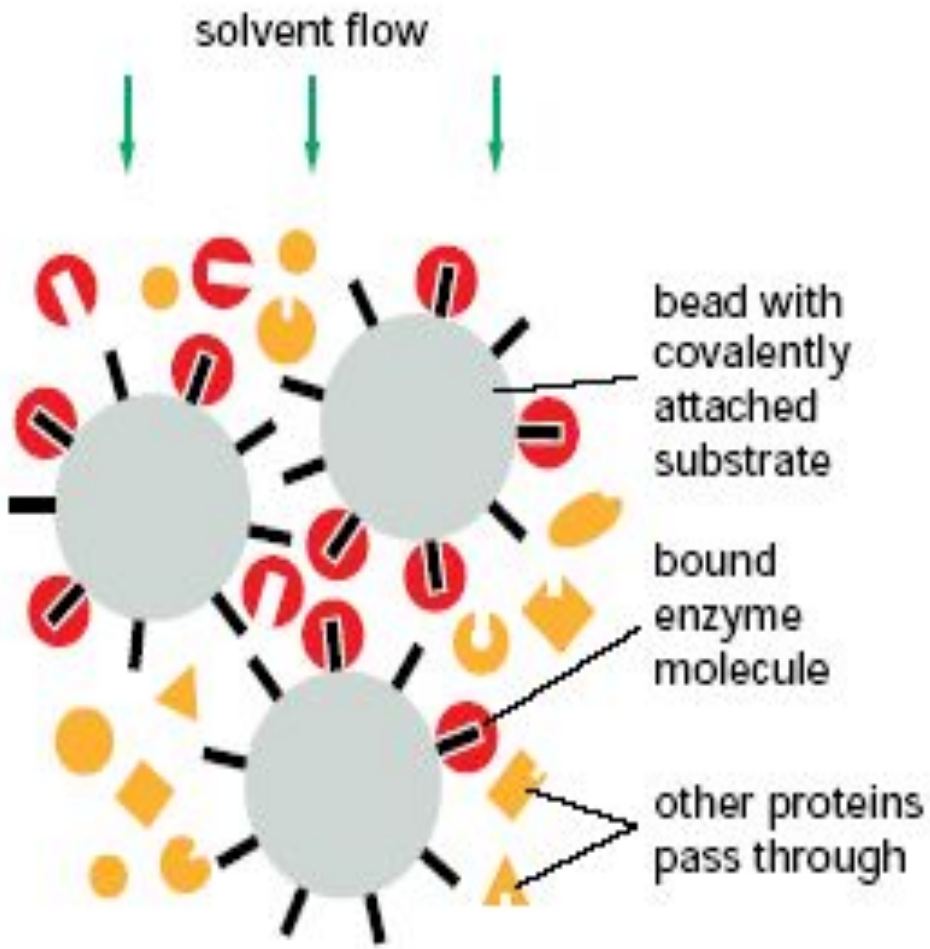
Protein molecules
that are small enough to enter
the holes in the beads
are delayed and travel more slowly
through the column.

Proteins that cannot enter the
beads are washed out of the column
first.

Such columns also
allow an estimate of protein size.

Essential Cell Biology:
An Introduction to the Molecular Biology of the Cell

(C) AFFINITY CHROMATOGRAPHY



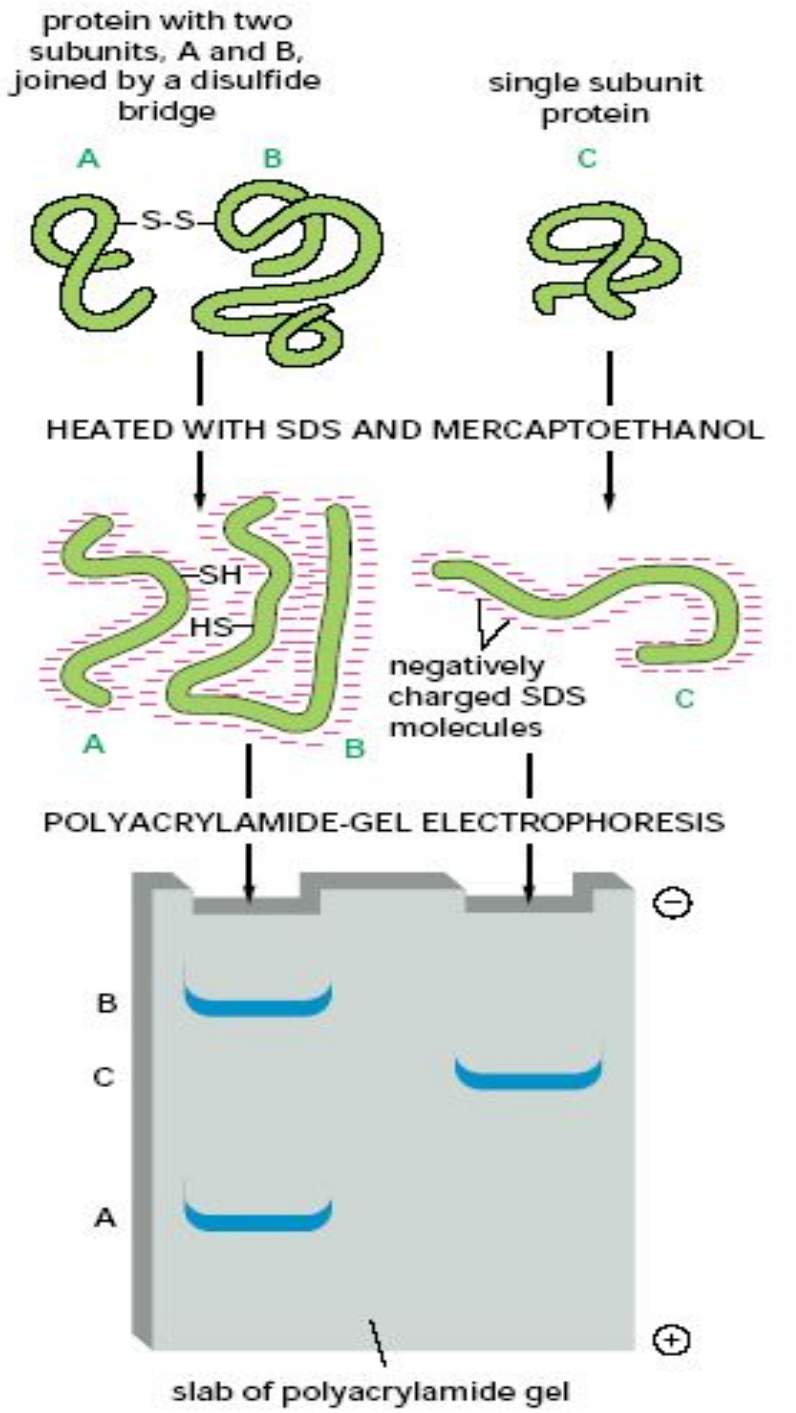
Affinity columns
contain a matrix
covalently coupled to a molecule
that interacts specifically
with the protein of interest

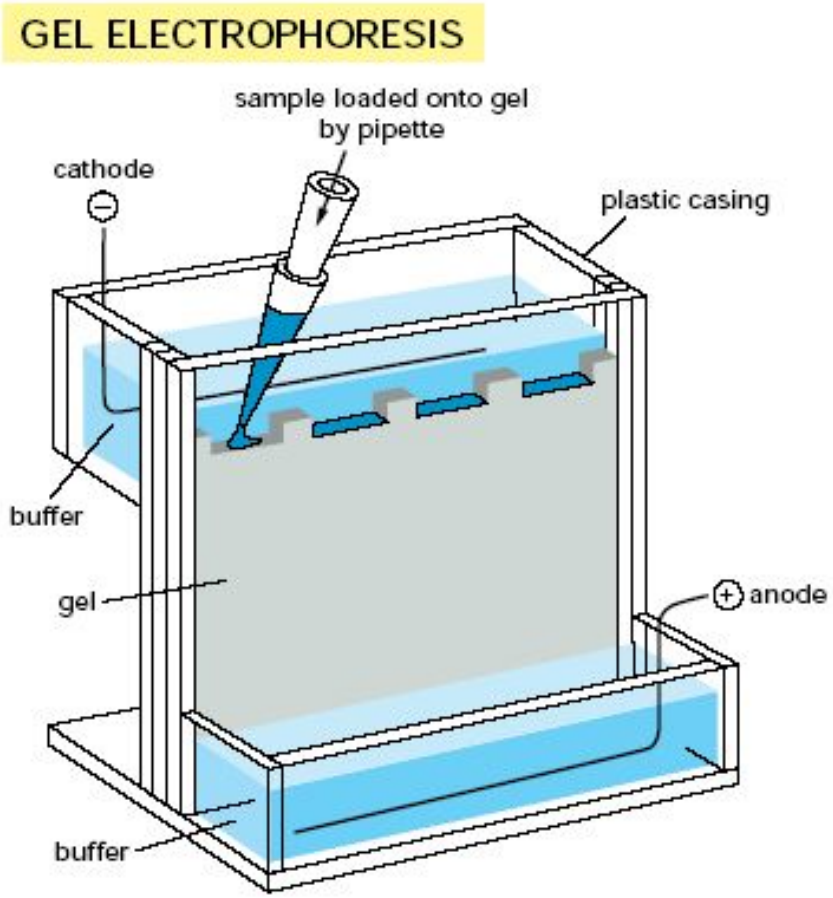
(e.g., an antibody, or an enzyme substrate).

Proteins that bind specifically
to such a column
can finally be released
by a pH change or
by concentrated salt solutions,
and they emerge highly purified.

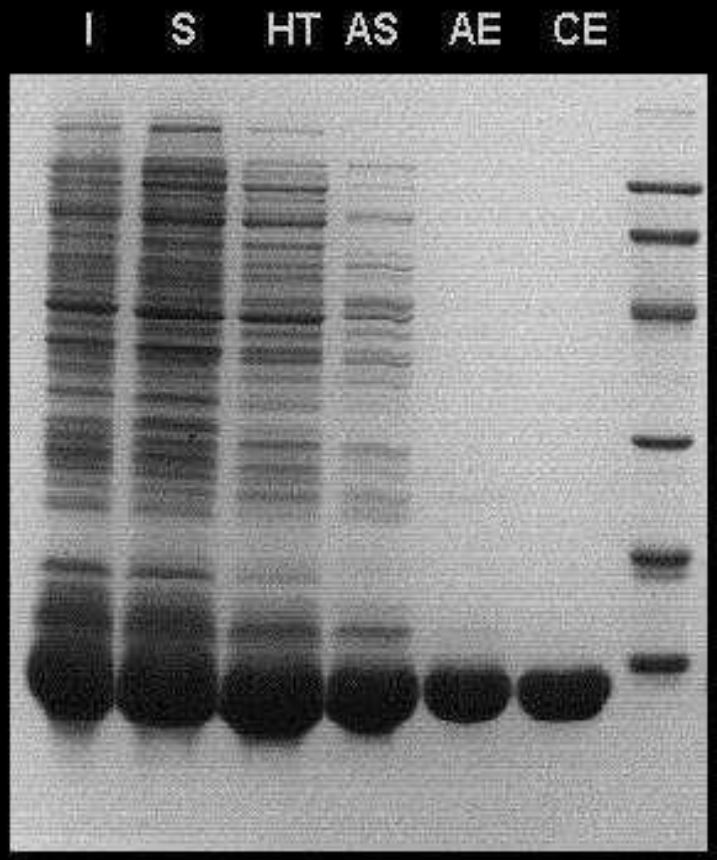
Essential Cell Biology:
An Introduction to the Molecular Biology of the Cell

Protein electrophoresis





Expression and Purification of Protein D



kDa

97
66
45
31
21.5
14.5

Legend

I: Induced (4h @ 37°C)
S: Supernatant after sonication
HT: Supernatant after heat treatment (50°C)
AS: Ammonium sulfate pellet
AE: Eluate from anion-exchange column
CE: Eluate from cation-exchange column

Result

50 mg/l of pure protein D

www.unizh.ch/.../Teaching_slide_shows/ Lambda/sld015.htm

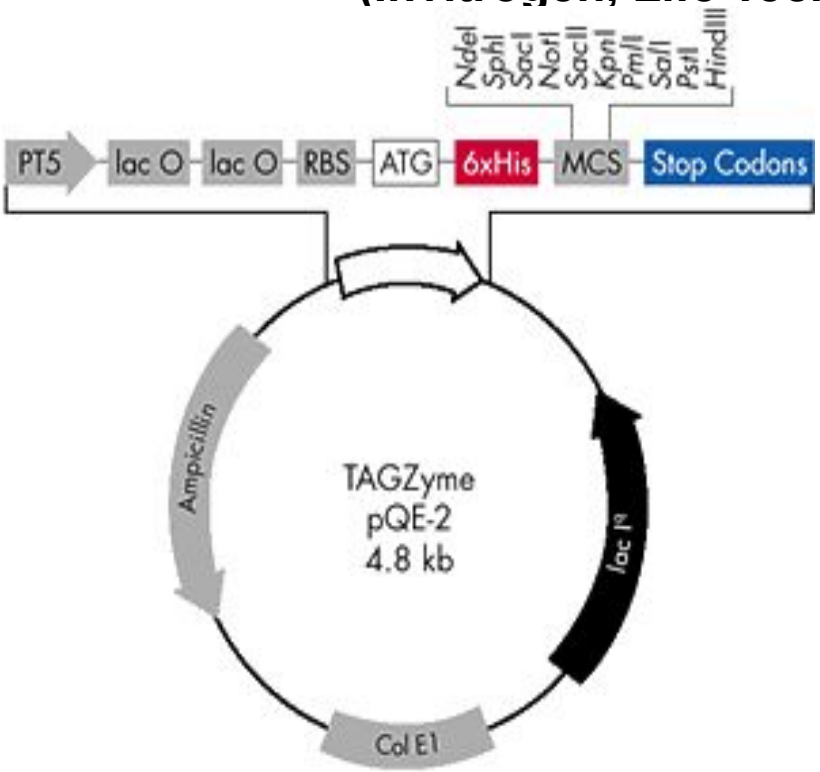
Fusion proteins

- increase production level
- facilitate purification (taq)
- detection of expression (GFP fusion)
- Redirection of proteins (secretion -> signal peptidases)
- Surface display (for screening of libraries)
- Tandem arrays (for small peptides, toxic proteins,..)

Most widely used purification strategy – to produce your protein as a fusion with something easily purifyable

6xHIS Tag

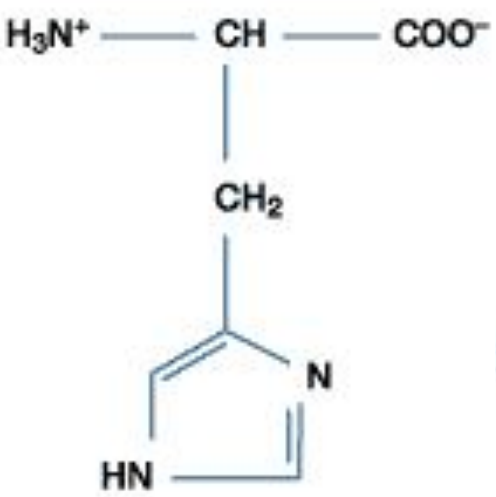
(Invitrogen, Life Technologies, Novagen, QIAGEN):



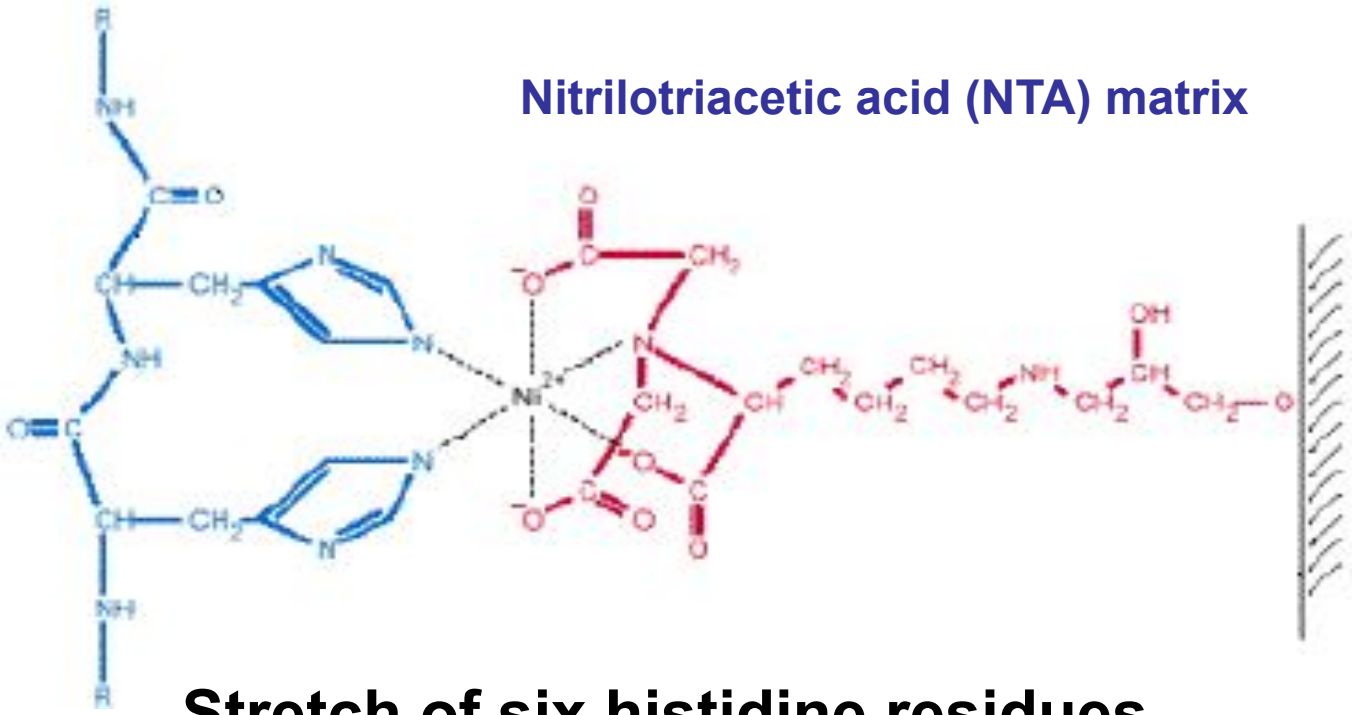
1. This small addition rarely affects protein structure to a significant degree

2. Interaction so strong, it tolerates denaturing conditions (could be used for inclusion bodies purification)

Histidine: a charged aminoacid



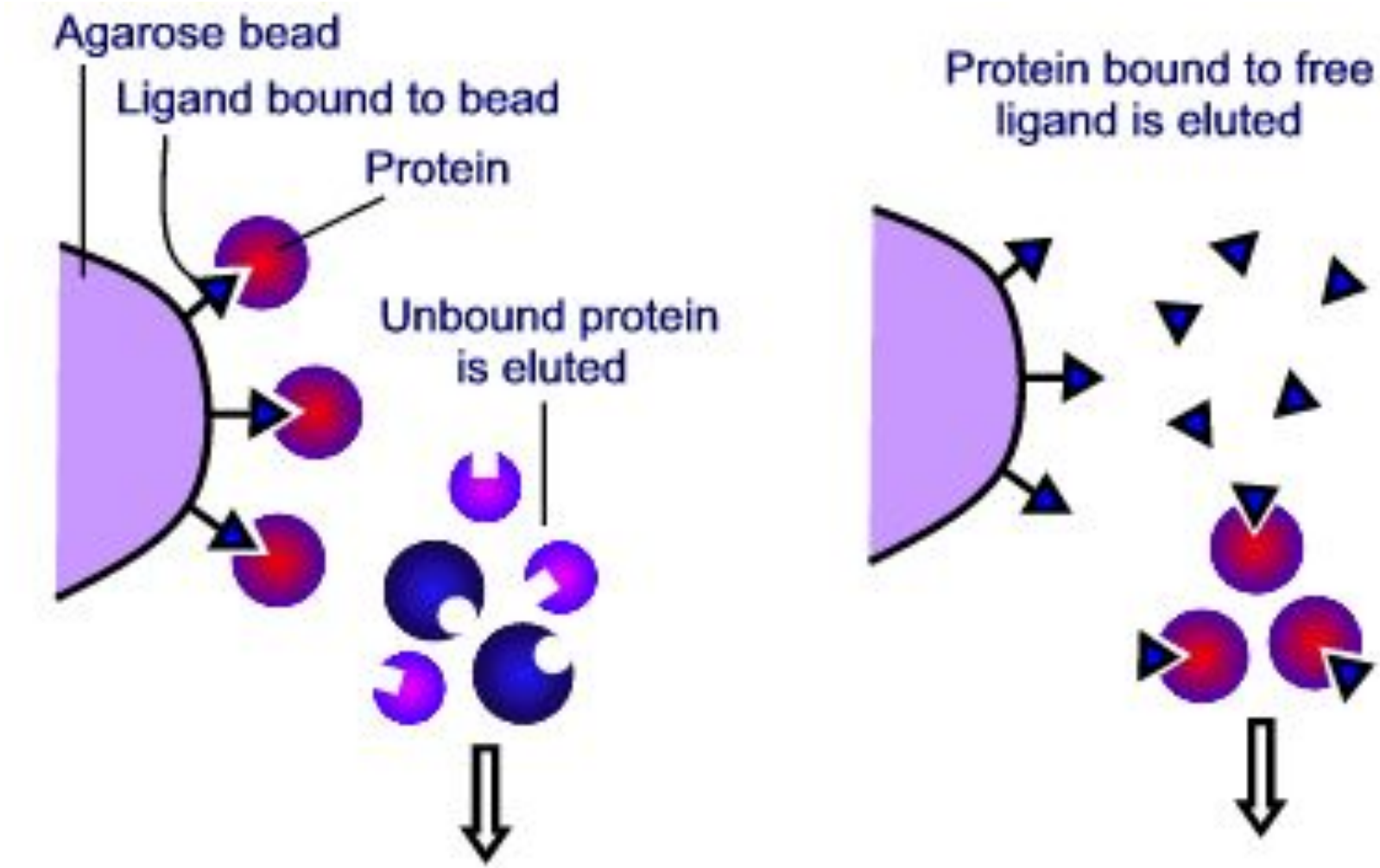
Histidine



Stretch of six histidine residues interacts with nickel ion that is tightly bound to a NTA matrix

The affinity of this interaction is very high
which allows protein purification to 95% in a single step.

GST – fusion. Principle is the same. Binds to glutation



FUNCTIONS OF GST

Require strong binding to glutathione

CATALYTIC FUNCTIONS:

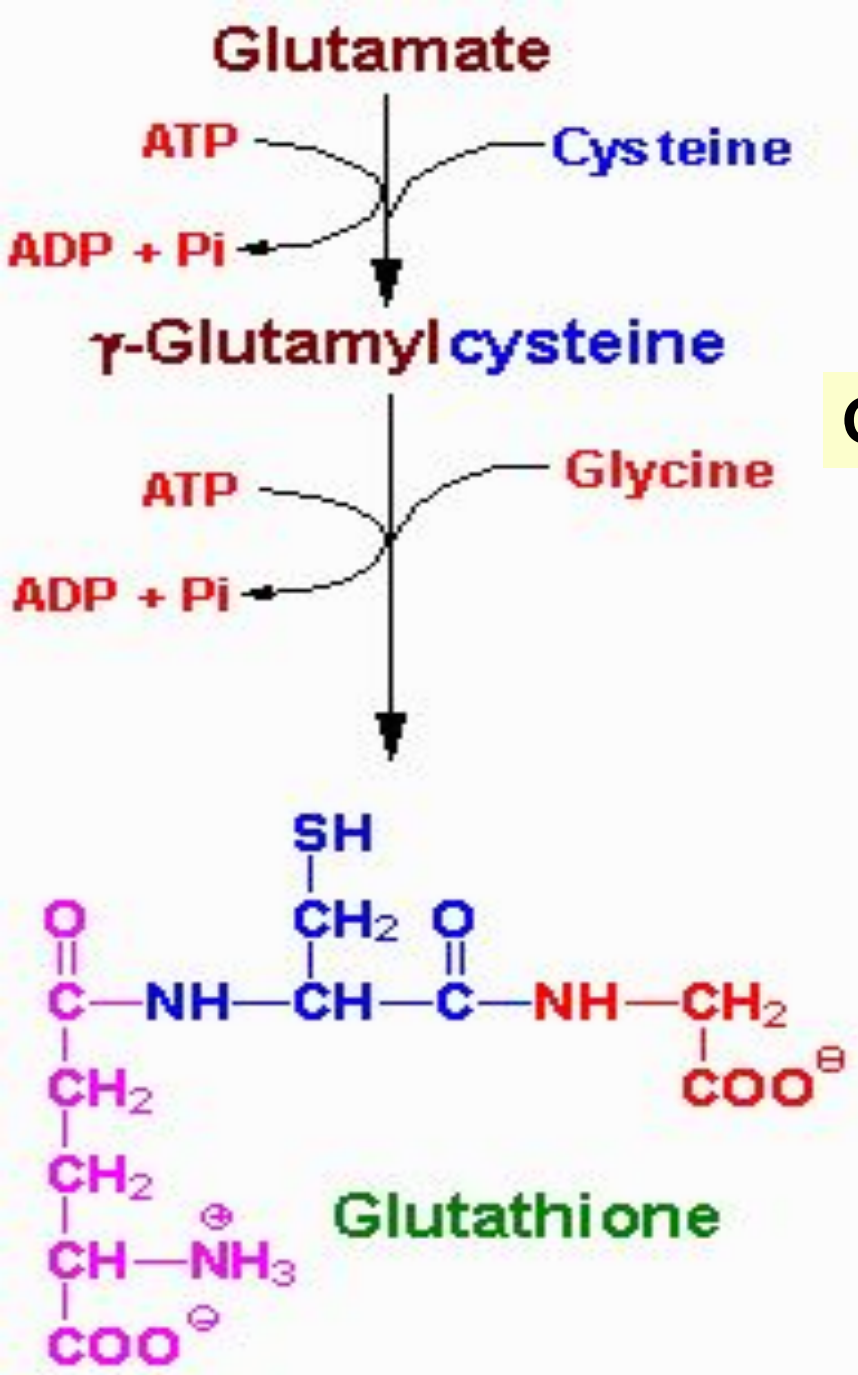
- Glutathione S-transferase activity
- Glutathione Peroxidase II activity

GSTs function catalytically to conjugate glutathione (GSH) with a wide variety of electrophilic substrates

---

- Reversible biding and transport of several organic compounds
- Irreversible binding and transport of several electrophilic compounds including some carcinogen

Glutathione



GST from Schistosoma japonicum

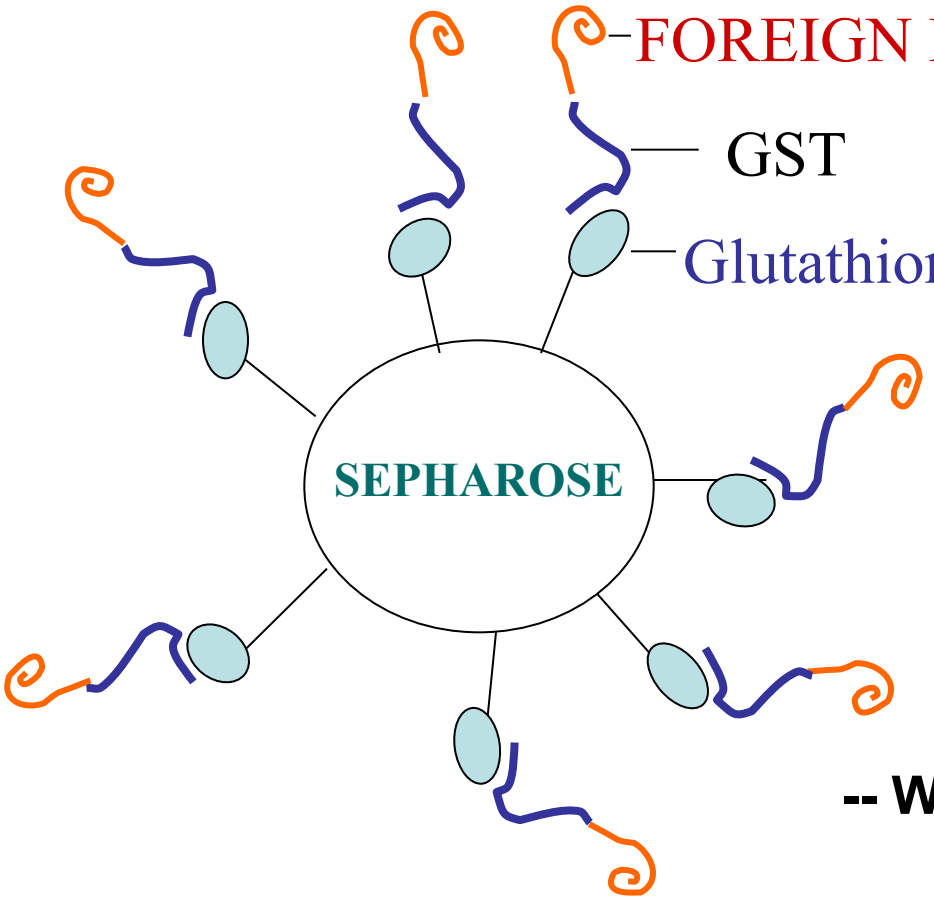
26 kDa tag

1) Keeps fusion proteins soluble

2) Used for fusion purification

3) Used for protein detection with GST antibody

FUSION PROTEIN BOUND TO GLUTATHIONE SEPHAROSE



Purification is simple :

-- WASH COLUMN EXTENSIVELY

-- ELUTE WITH REDUCED GLUTATHIONE

-- RESULTS IN PURE GST FUSION PROTEIN

Table 2b Comparison of rare codons in *E. coli*

Organism	AGG arginine	AGA arginine	CUA leucine	AUA isoleucine	CCC proline	GGA glycine
Bacterial hosts						
Escherichia coli B	2.1	2.4	3.4	5.0	2.4	8.2
E. coli K12	1.2	2.1	3.9	4.3	5.5	7.9
Anabaena sp.	2.6	8.3	14.0	8.3	13.0	12.4
Bacillus megaterium	2.7	9.1	10.9	10.3	2.9	26.2
Bacillus subtilis	3.9	10.5	4.8	9.3	3.3	21.8
Caulobactert crescentus CB 15	2.2	0.8	1.4	0.6	18.7	4.3
Methylobacterium extorquens	1.2	0.6	0.8	0.3	18.7	3.7
Staphylococcus carnosus	0.4	7.9	4.6	9.8	0.8	16.7
Streptomyces lividans	4.0	1.1	0.6	0.8	22.7	6.5
Some organisms						
Arabidopsis thaliana	10.9	18.9	9.9	12.6	5.3	24.2
Caenorhabditis elegans	4.0	15.4	7.9	9.4	4.4	31.6
Clostridium tetani E88	5.0	25.5	11.2	67.4	1.8	34.6
Drosophila melanogaster	6.3	5.2	8.2	9.5	18.0	17.7
Homo sapiens	11.9	12.0	7.2	7.4	19.9	16.5
H. sapiens Mitochondriom	0.4	0.4	70.4	44.5	33.9	18.9
Plasmodium falciparum	3.3	17.0	5.3	40.8	3.0	20.0
Pichia pastoris	6.6	20.2	10.9	11.7	6.7	19.1
Picrophilus torridus DSM9790	22.9	16.2	7.8	30.3	2.6	18.1
Saccharomyces cerevisiae	9.3	21.3	13.4	17.8	6.8	10.9

The codon usages of bacteria used for heterologous protein expression and some organisms are listed (for *Bacillus brevis* not available). Codon frequencies are expressed as codons used per 1,000 codons encountered. The arginine codons AGG and AGA are recognized by the same tRNA and should therefore be combined. Codon frequency of more than 15 codons/1,000 codons may cause problems for high-level expression in *E. coli*. A complete summary of codon usages can be found at http://www.kazusa.or.jp/codon/

Some problems of production in *E. coli*

Table 2a Some problems of heterologous protein production in *E. coli* and possible solutions

Symptom	Possible problem	A collection of solutions
Cell death or no colonies	Toxic protein, high basal expression	More stringent control over basal expression Tightly controlled promoter system Weaker promoter Lowering temperature Lowering inducer concentration
Insoluble disulfide protein (inclusion bodies)	Reduction of disulfide bonds	Minimize reduction in cytoplasm Accumulation in the periplasm
Insoluble protein (inclusion bodies)	Too much expression	Attenuate expression by: weaker promoter, lowering temperature, lowering inducer concentration, decrease plasmid copy number, fusion of a hydrophilic affinity tag
No activity	Misfolded protein, affinity tag can decrease activity	Minimize reduction in cytoplasm Accumulation in the periplasm Attenuate expression Change affinity tag
No protein, truncated protein	*E. coli* codon usage (codon bias)	Supply rare tRNAs Stronger promoter Increase plasmid copy number Lower temperature Tightly controlled promoter system

Nevertheless, another bacterial host than *E. coli* could also solve the problem

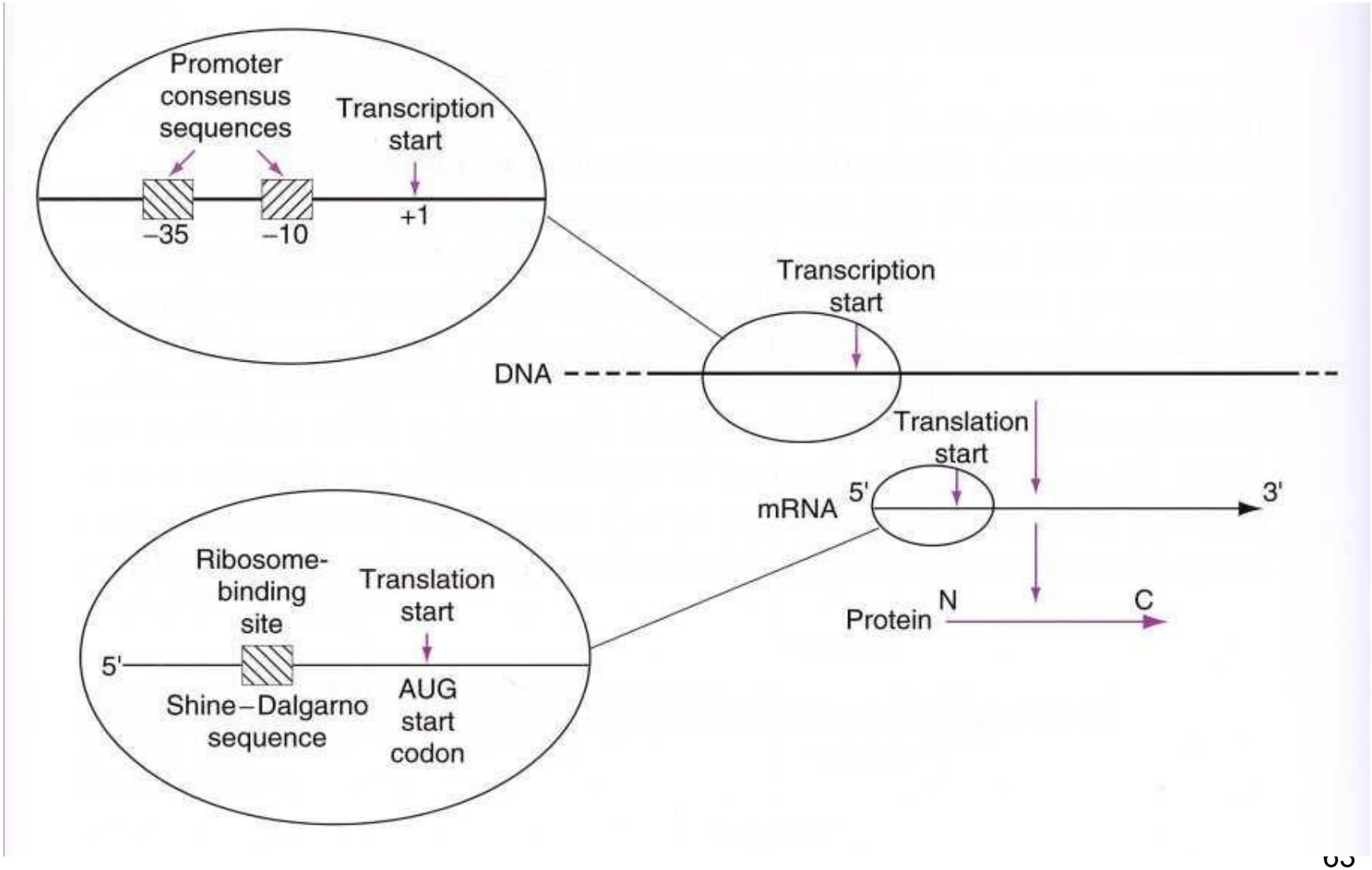
Some *E.coli* expression host considerations

Table 1 Some *E. coli* strains most frequently used for heterologous protein production and their key features

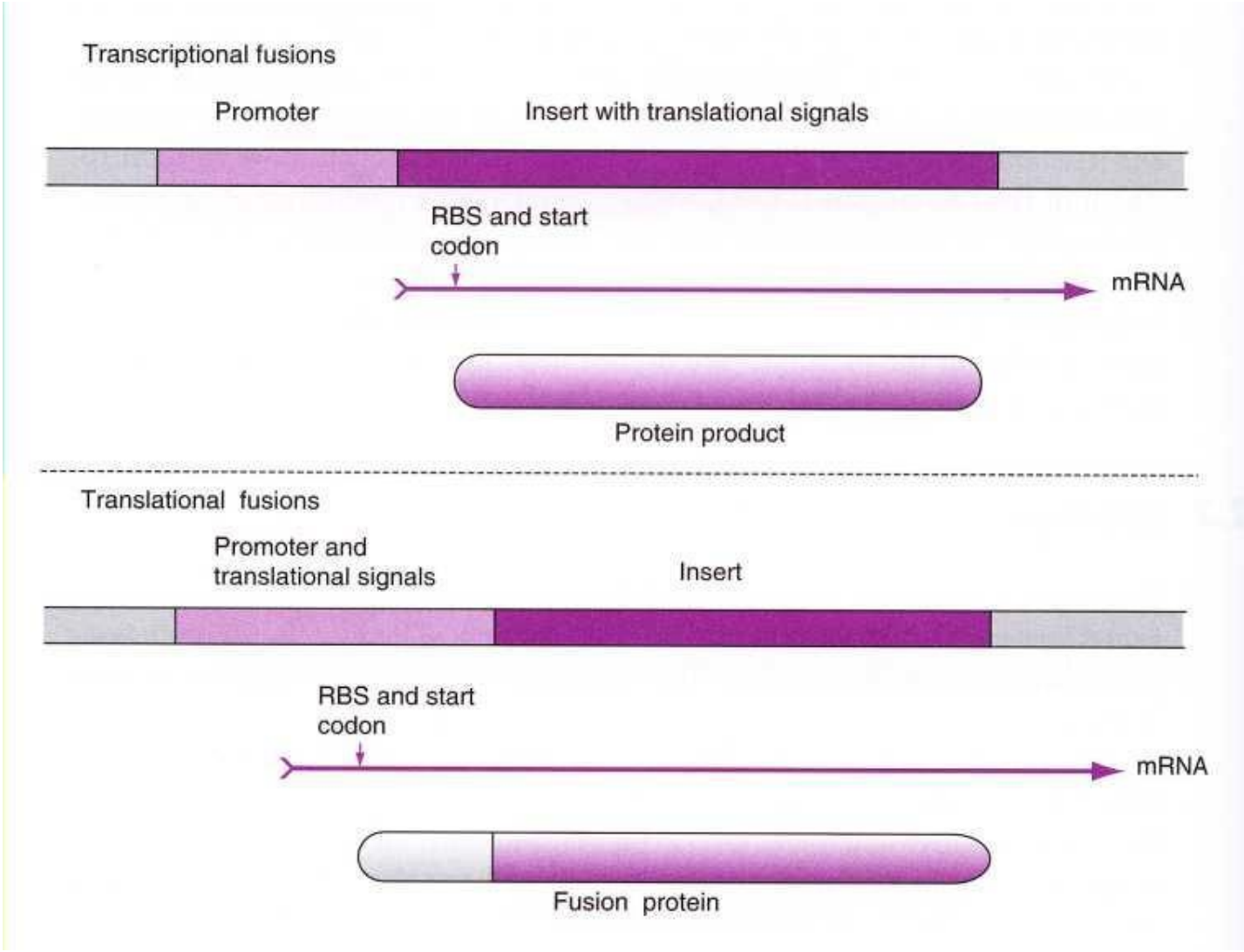
E. coli strain	Derivation	Key features
AD494	K-12	*trxB* mutant; facilitates cytoplasmic disulfide bond formation
BL21	B834	Deficient in *lon* and *ompT* proteases
BL21 trxB	BL21	*trxB* mutant; facilitates cytoplasmic disulfide bond formation; deficient in *lon* and *ompT* proteases
BL21 CodonPlus-RIL	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AGG, AGA, AUA, CUA; deficient in *lon* and *ompT* proteases.
BL21 CodonPlus-RP	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AGG, AGA, CCC; deficient in *lon* and *ompT* proteases.
BLR	BL21	*recA* mutant; stabilizes tandem repeats; deficient in *lon* and *ompT* proteases
B834	B strain	Met auxotroph; ^{35}S-met labeling
C41	BL21	Mutant designed for expression of membrane proteins
C43	BL21	Double mutant designed for expression of membrane proteins
HMS174	K-12	*recA* mutant; Rif resistance
JM 83	K-12	Usable for secretion of recombinant proteins into the periplasm
Origami	K-12	*trxB/gor* mutant; greatly facilitates cytoplasmic disulfide bond formation
Origami B	BL21	*trxB/gor* mutant; greatly facilitates cytoplasmic disulfide bond formation; deficient in *lon* and *ompT* proteases
Rosetta	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AUA, AGG, AGA, CGG, CUA, CCC, and GGA; deficient in *lon* and *ompT* proteases
Rosetta-gami	BL21	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AUA, AGG, AGA, CGG, CUA, CCC, and GGA; deficient in *lon* and *ompT* proteases; *trxB/gor* mutant; greatly facilitates cytoplasmic disulfide bond formation

Most strains are also available as DE3 and DE3 pLysS strains. Strains are commercially available from different manufacturers

Principal factors in bacterial expression



Type of expression vectors



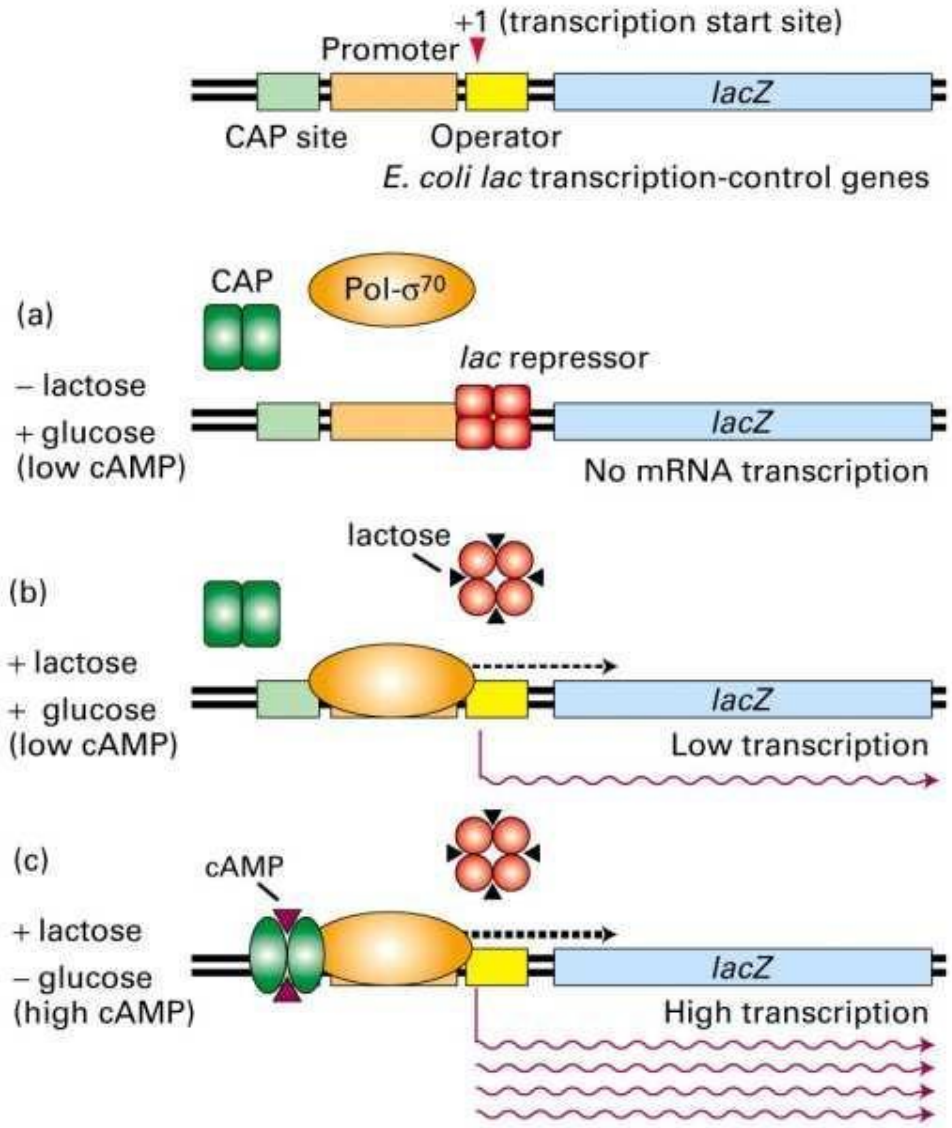
Initiation of Transcription
Promoters for Expression in Prokaryotes

- In Escherichia coli

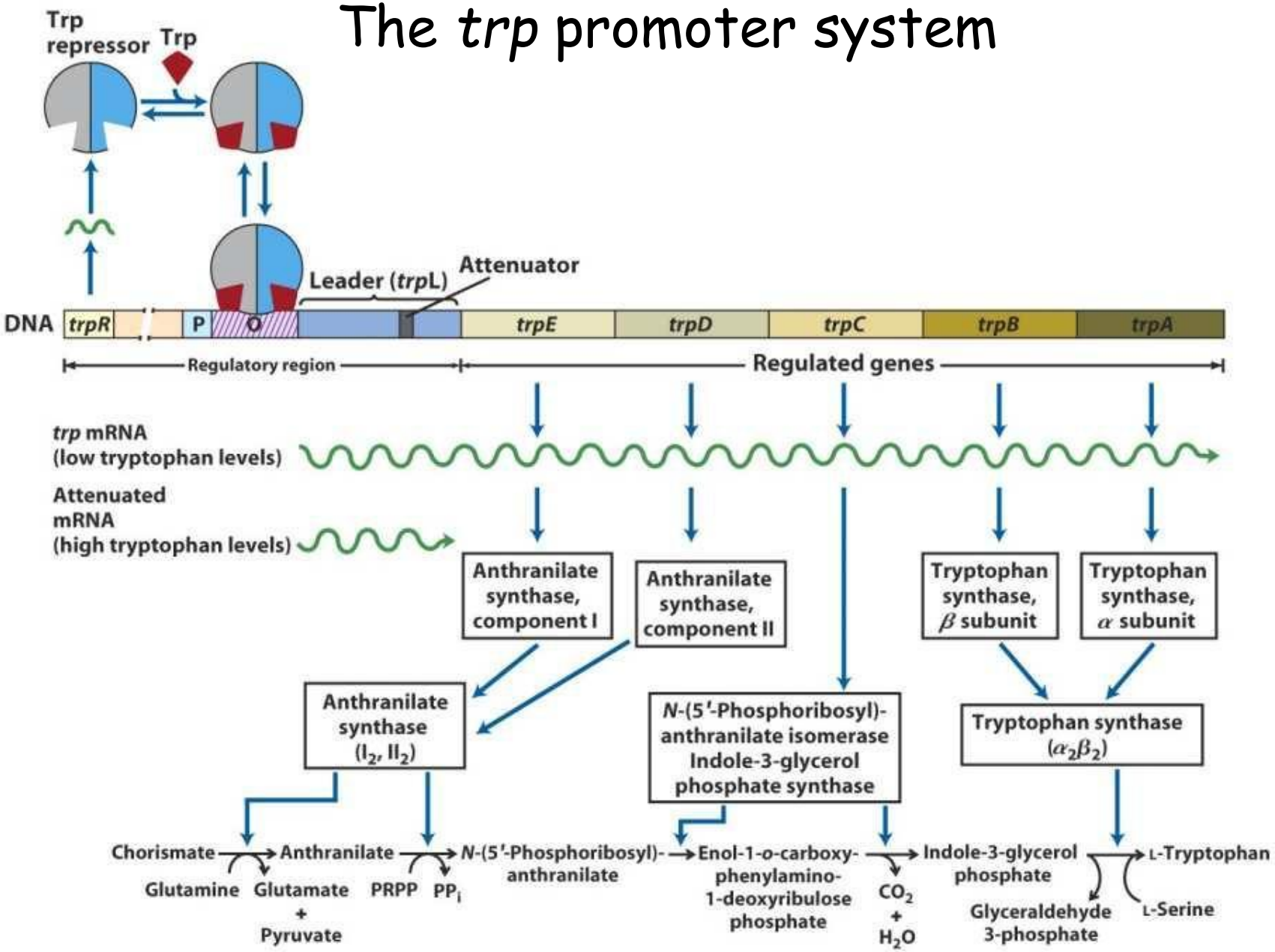
  - *Lac* system - p*lac*
  - *Trp* system
  - synthetic systems – p*tac*, p*trc*

- In Bacillus

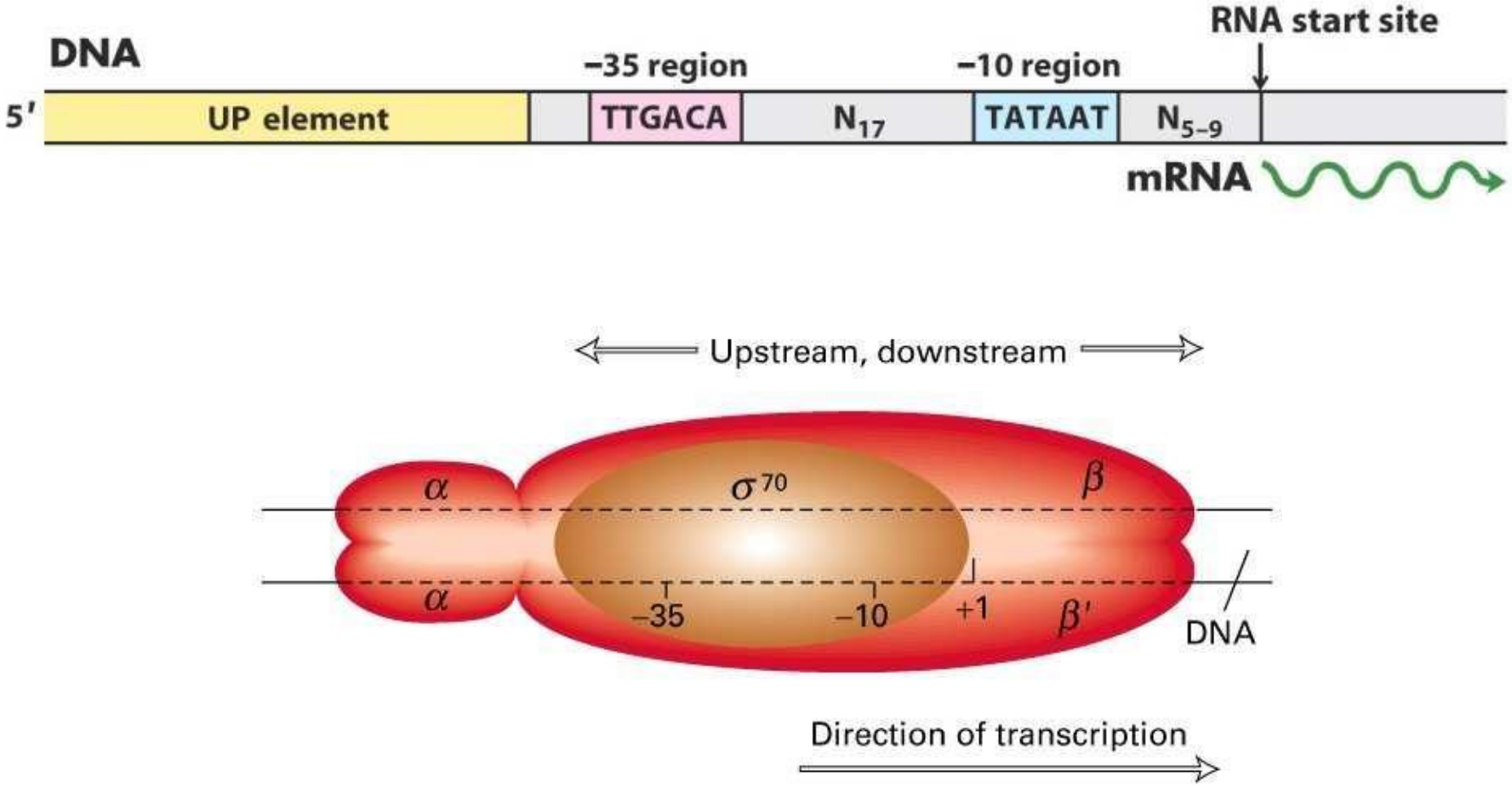
The *Lac* promoter System



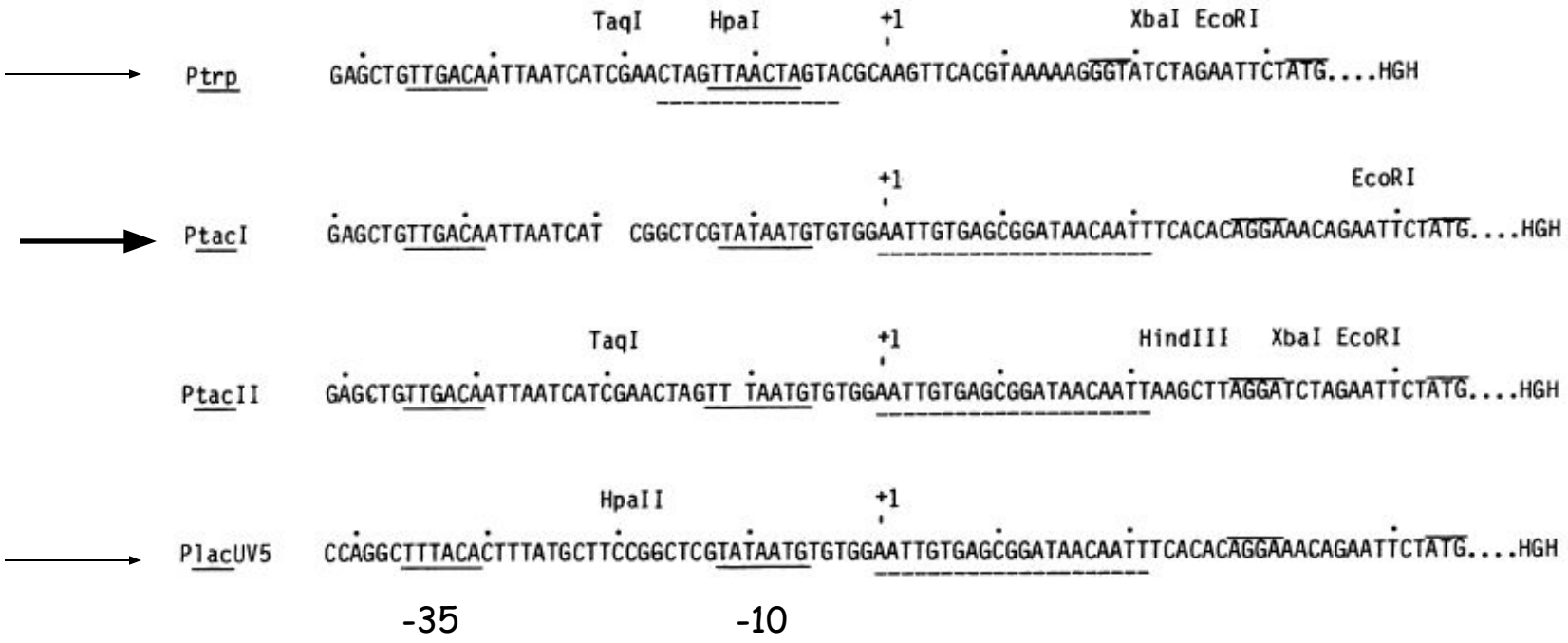
The *trp* promoter system



E. coli Promoter Sites



Synthetic *E. coli* promoters



p*tac* -> -35 box from p*trp* + -10 box from p*lac* -> p*t+ac*

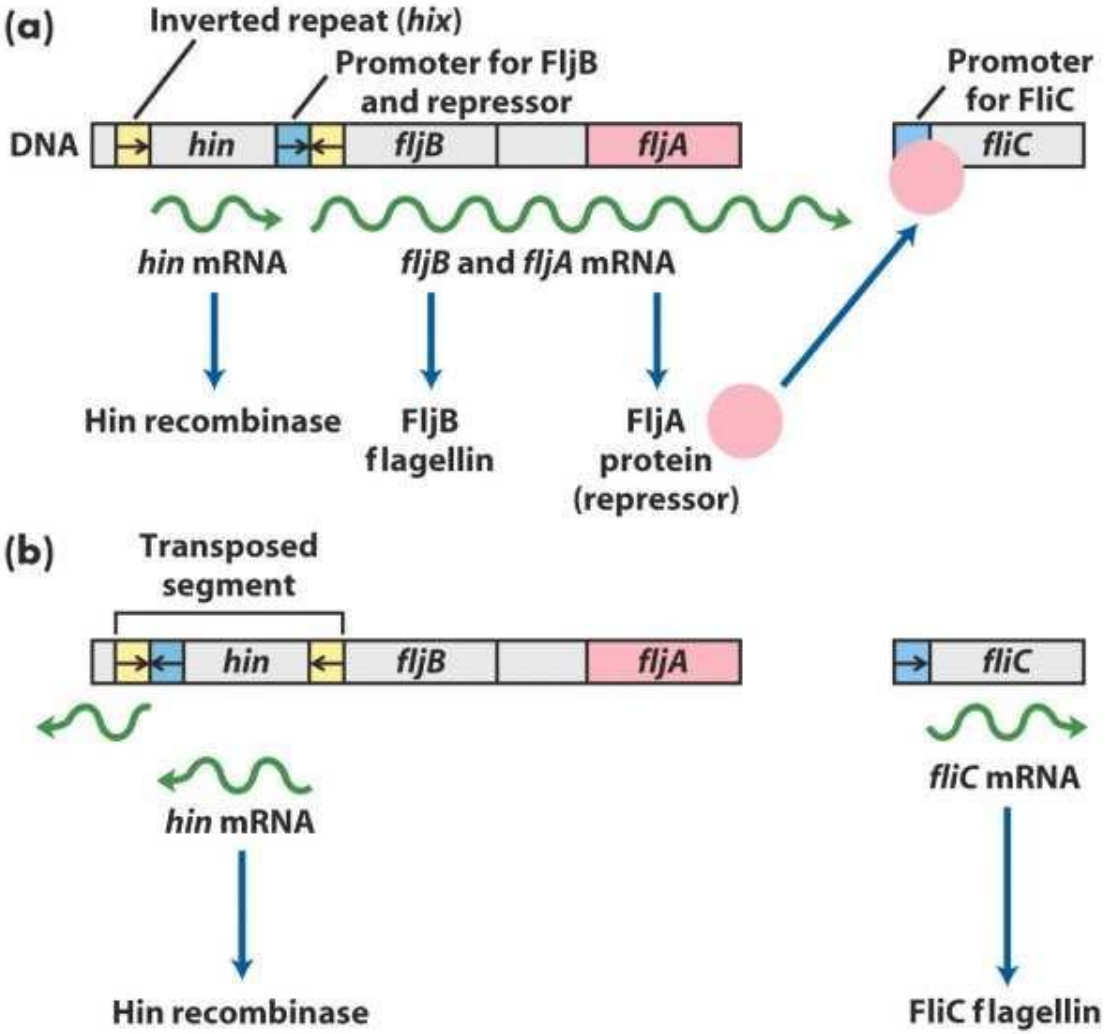
TABLE I Commonly Used Promoter Systems

Promoter	Repression requirements	Induction	Benefits	Drawbacks
lac/*lacUV5*[a]	*lac I*^{q}	IPTG (lactose)	Strong promoter, easily induced	Leakage
trp	*trp R*	trp starvation or β-IAA addition	Strong promoter	Difficult to repress, induction toxicity
tac[b]	*lac I*^{q}	IPTG (lactose)	Very strong promoter, easily induced	Leakage
P_L, P_R[c]	phage repressor (cI857)	Heat shock or nalidixic acid	Very strong promoter	Leakage, induction of SOS and heat shock responses, slow growth
tac or *lac* and phage *att* sites[d]	lysogen λ*xis*$^-$cI857	Short heat shock inverts promoter	Strong promoters, no leakage	Efficiency, timing of inversion may vary
T7[e]	T7 lysozyme (pLys) or F′ for infection	IPTG induces T7 pol expression T7 phage infection	Strong promoter, easily induced No leakage	Lysozyme toxicity in cells, some leakage may occur, Limited to engineered strains
phoA[f]	*phoR*	Phosphate starvation	Inexpensive, growth regulated induction	May activate SOS response

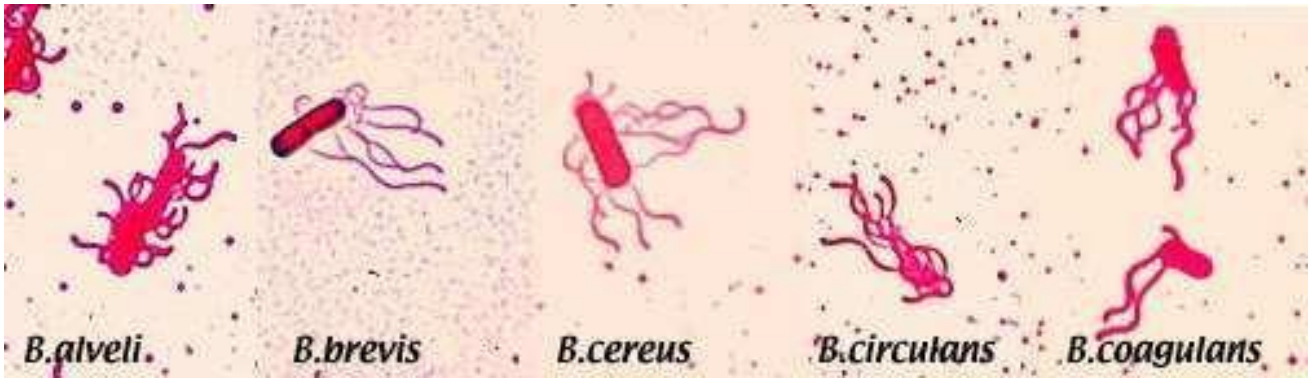
[a] pSL301, Invitrogen.
[b] pKK223-3, Pharmacia; pPROK-1, Clontech.
[c] pPL-lambda, Pharmacia.
[d] pNH series, Stratagene.
[e] pET series, Novagen.
[f] pBAce, Stratagene.

Inverted Promoter System (from Salmonella)

-> for very toxic proteins



Bacillus



Flagellar stains of various species of *Bacillus* from CDC

In 1872, Ferdinand Cohn, a student of Robert Koch, recognized and named the bacterium ***Bacillus subtilis*.**

The organism was made to represent a large and diverse genus of Bacteria, ***Bacillus,*** and was placed in the family ***Bacillaceae*.**

The family's distinguishing feature is the production of **endospores**, which are highly refractile resting structures formed within the bacterial cells. Since this time, members of the genus *Bacillus* are characterized as Gram-positive, rod-shaped, aerobic or facultative, endospore-forming bacteria.

Bacillus

- **Antibiotic Producers**: *B. brevis* (e.g. gramicidin, tyrothricin), *B. cereus* (e.g. cerexin, zwittermicin), *B. circulans* (e.g. circulin), *B. laterosporus* (e.g. laterosporin), *B. licheniformis* (e.g. bacitracin), *B. polymyxa* (e.g. polymyxin, colistin), *B. pumilus* (e.g. pumulin) *B. subtilis* (e.g. polymyxin, difficidin, subtilin, mycobacillin).
- **Pathogens of Insects**: *B. larvae, B. lentimorbis,* and *B. popilliae* are invasive pathogens. *B. thuringiensis* forms a parasporal crystal that is toxic to beetles.
- **Pathogens of Animals**: *B. anthracis, and B. cereus. B. alvei, B. megaterium, B. coagulans, B. laterosporus, B. subtilis, B. sphaericus, B. circulans, B. brevis, B. licheniformis, B. macerans, B. pumilus,* and *B. thuringiensis* have been isolated from human infections.

- The Genus *Bacillus* includes two bacteria of significant medical importance, *B. anthracis,* the causative agent of anthrax, and *B. cereus*, which causes food poisoning. Nonanthrax *Bacillus* species can also cause a wide variety of other infections, and they are being recognized with increasing frequency as pathogens in humans.

Bacillus

- ***Bacillus* strains used as production organisms:**
  - *B. subtilis*
  - *B. brevis*
  - *B. licheniformis*

- **Transformation systems:**
  - **via competent cells** (during transition from vegetative cells -> sporulation, cell can take up DNA (ss) when population reaches a metabolic state called competence)
  - **protoplast**
  - **bacteriophage-mediated transduction**

- **Vectors:**
  - **replicating plasmids** (pUB110, pE194, pC194, pHP13, shuttle vectors)
    -> replicating plasmids with temperature-sensitive origin of replication (replication stops above certain temp. -> pE194 stops above 45°C)
  - **integrative vectors** (normally shuttle vectors)

- **Promoters:**
  - ***aprE* promoter** -> induction with onset of sporulation
  - **amylase promoter** -> growth-phase and nutrition regulated promoter (induction at end of exponential growth + repression by glucose)
  - ***sacB* promoter** (levansurase) -> not regulated
  - ***spac* promoter** -> hybrid promoter (subtilis phage + lac operator) -> induction with IPTG

Bacillus as expression host

Table 1
Examples of Expression of Heterologous and Homologous Proteins in Bacilli

	Origin	Yield	*Bacillus* sp.	References
Eukaryotic genes				
Epidermal growth factor	Human	3 g/liter	*B. brevis*	Udaka and Yamagata (1993a)
Growth hormone	Human	0.2 g/liter	*B. brevis*	Udaka and Yamagata (1993a)
Growth hormone	Human	0.2 g/liter	*B. subtilis*	Honjo *et al.* (1987)
Growth hormone	Tuna	0.24 g/liter	*B. brevis*	Udaka and Yamagata (1993a)
IL–3	Human	0.10 g/liter	*B. licheniformis*	van Leen *et al.* (1991)
Pepsinogen	Swine	0.01 g/liter	*B. brevis*	Takao *et al.* (1989)
Salivary α-amylase	Human	0.06 g/liter	*B. brevis*	Udaka and Yamagata (1993a)
Isomerase (disulfide)	*Humicola insolens*	0.3 g/liter	*B. brevis*	Udaka and Yamagata (1993a)
Taka-amylase	*Aspergillus orizae*	0.02 g/liter	*B. brevis*	Udaka and Yamagata (1993a)
Bacterial genes				
α-amylase	*B. licheniformis*	3.5 g/liter	*B. brevis*	Udaka and Yamagata (1993)
α-amylase	*B. licheniformis*	1.0 g/liter	*B. subtilis*	Sloma *et al.* (1988)
β-lactamase	*B. licheniformis*	140m g/liter	*B. subtilis*	Yoshimura *et al.* (1986)
Cholera toxin B subunit	*Vibrio cholerae*	1.4 g/liter	*B. brevis*	Udaka and Yamagata (1993a)
Pertussis toxin S1	*B. pertussis*	100m g/liter	*B. subtilis*	Saris *et al.* (1990)
Pertussis toxin S4	*B. pertussis*	0.5m g/liter	*B. subtilis*	Saris *et al.* (1990)
Protein A	*S. aureus*	1 g/liter	*B. subtilis*	Fahnestock and Fisher (1986)
Subtilisin	*B. amyloliquefaciens*	60 mg/liter	*B. subtilis*	Ferrari *et al.* (1993)

Bacillus as expression host

Table 2
Advantages and Disadvantages of *Bacillus* Expression

Advantages	Disadvantages
Efficient transformation via natural competence	Industrial strains not easy to transform
B. subtilis is genetically well characterized.	
Laboratory strains easy to manipulate	
B. subtilis genome fully sequenced	
Availability of numerous plasmids	Poor efficiency of transformation of ligation mixtures.
Capable of secreting proteins into the medium	Poor secretion of most heterologous proteins; degradation problems
Availability of several signal sequences	
Well-established fermentation protocol	Genetics and physiology of the fermentation not well understood
Can grow efficiently on inexpensive feedstock	
GRAS status of its products	

Products produced in Prokaryotic Systems

- Restriction Endonucleases -> produced in *E. coli*
- L- Ascorbic Acid (Vitamin C) -> recombinant *Erwinia herbicola* (gram-negative bacterium)
- Synthesis of Indigo (blue pigment -> dye cotton /jeans) -> produced in *E. coli*
- Amino Acids -> produced in *Corynebacterium glutamicum* (gram-positive bacterium)
- Lipases (laundry industry) -> from *Pseudomonas alcaligenes* produced in *Pseudomonas alcaligenes*
- Antibiotica (most of them from *Streptomyces*, other gram-positive bacteria, fungi) -> produced in recombinant *Streptomyces* and fungi (*Penicillium*)
- Biopolymers (PHB -> biodegradable plastics) -> produced in *E. coli* (stabilized with parB)

Expression in Eukaryotic Systems

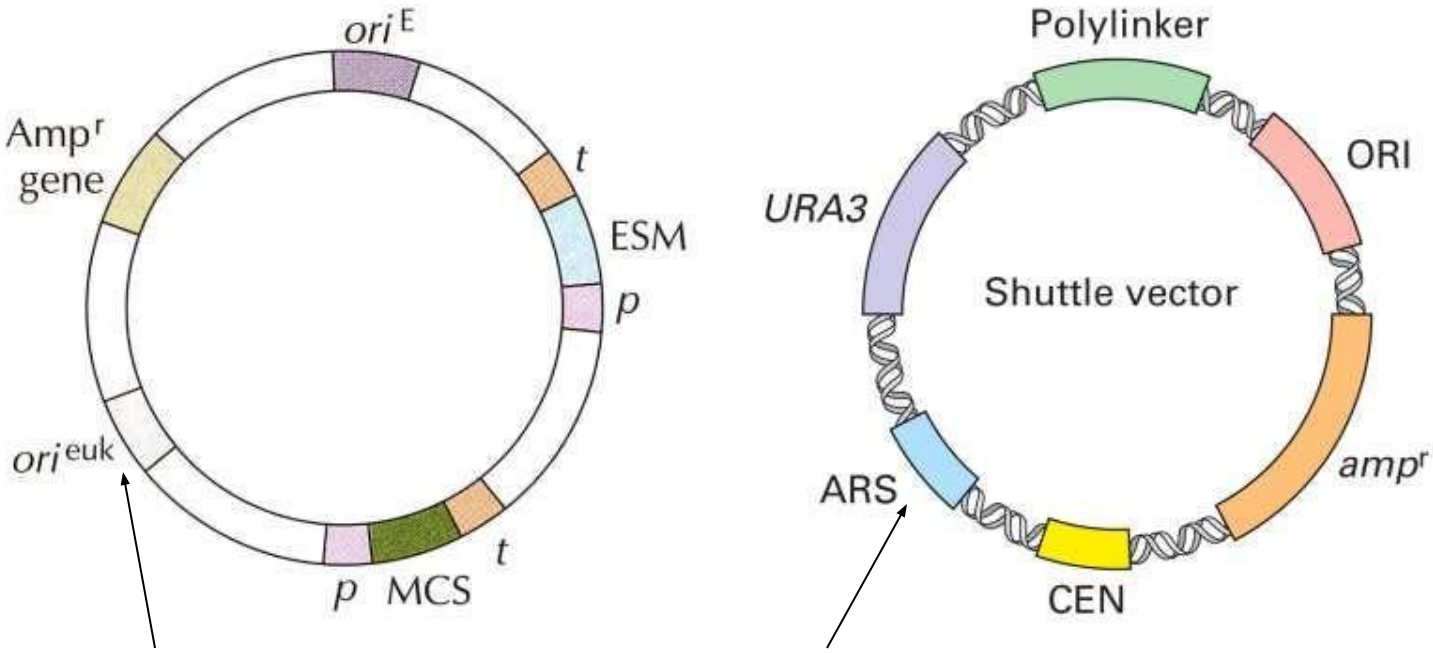
- Yeast

  - *Saccharomyces cerevisiae* (baker's yeast)
  - *Pichia pastoris*

- Insect Cells – Baculovirus

- Mammalian Cells

Expression in Yeast



Autonomous replicating vectors -> shuttle vectors

Expression in Saccharomyces cerevisiae

Autonomous replicating systems

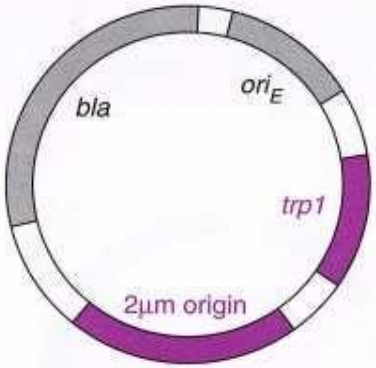


Figure 6.15 Structure of a yeast episomal vector: *bla* = beta-lactamase (ampicillin resistance); *ori* = origin of replication in *E. coli*; *trp1* = selectable marker in *S. cerevisiae* auxotrophs; 2 μm origin = origin of replication in *S. cerevisiae*

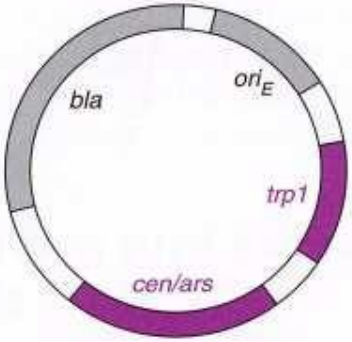


Figure 6.16 Structure of a yeast centromere vector: *bla* = beta-lactamase (ampicillin resistance); *ori* = origin of replication in *E. coli; trp1* = selectable marker in *S. cerevisiae* auxotrophs; *cen/ars* = centromere and autonomously replicating sequence, providing an origin of replication in *S. cerevisiae*

Expression in Saccharomyces cerevisiae

Integrative systems

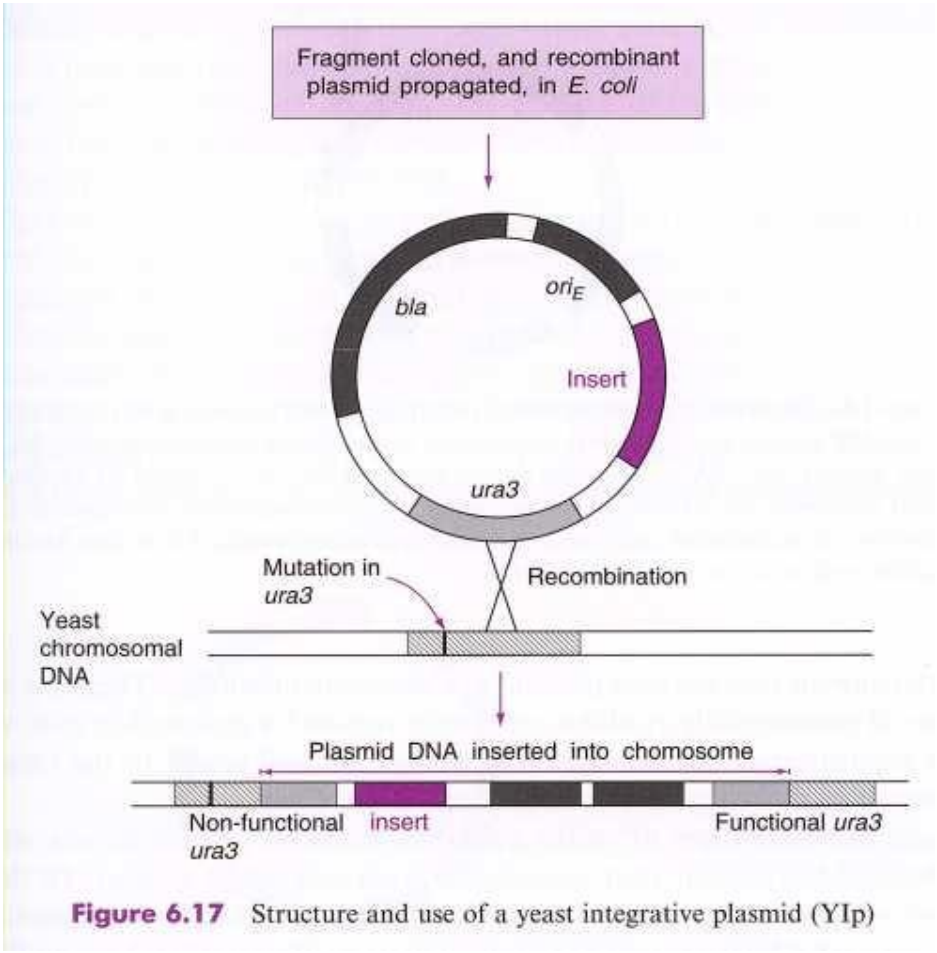
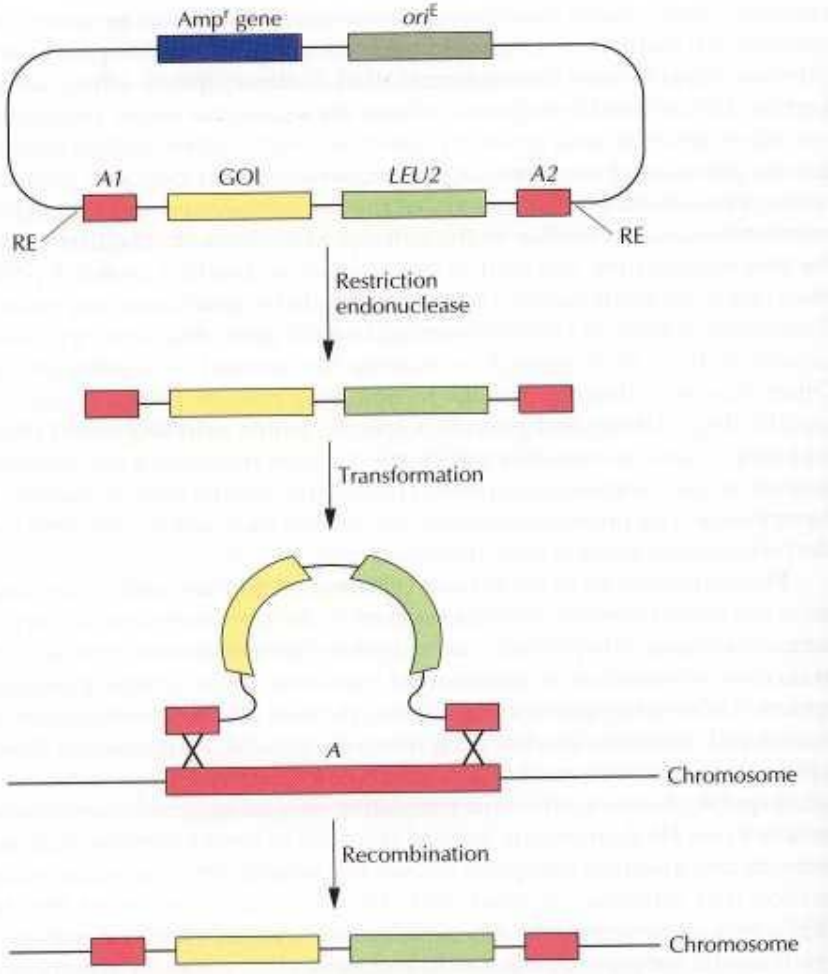


Figure 6.17 Structure and use of a yeast integrative plasmid (YIp)



Probability for integration higher with linear fragments !

Expression in Saccharomyces cerevisiae

TABLE II Examples of Vectors for Yeast Expression

Plasmid type	Copy number	Selectable marker[a]	Comments
Integrative (YIp)	One	*URA3, LEU2*	Stable integration into chromosome
Multi-integrative	One–five	*URA3, LEU2*	Stable integration into chromosome
Centromere (YCp)	One–two	*TRP1, URA3, LEU2, HIS3*	Stable autonomously replicating from *ARS* element or 2-μm origin
ARS (YRp)	Moderate	*TRP1*	Autonomously replicating but unstable
2 μm (YEp)	Moderate	*LEU2, URA3, HIS3*	Autonomously replicating but fairly stable
2 μm (YEp)	High	*LEU2-d, POT*	Autonomously replicating, stable

[a] Examples of the common selectable markers available for each vector type; other combinations of selectable markers and vector type have been developed in individual laboratories.

Expression in Saccharomyces cerevisiae

Table 7.1 Promoters for *S. cerevisiae* expression vectors

Promoter	Expression conditions	Status
Acid phosphatase (*PH05*)	Phosphate-deficient medium	Inducible
Alcohol dehydrogenase I (*ADHI*)	2–5% Glucose	Constitutive
Alcohol dehydrogenase II (*ADHII*)	0.1–0.2% Glucose	Inducible
Cytochrome c_1 (*CYC1*)	Glucose	Repressible
Gal-1-P Glc-1-P uridyltransferase	Galactose	Inducible
Galactokinase (*GAL1*)	Galactose	Inducible
Glyceraldehyde-3-phosphate dehydrogenase (*GAPD, GAPDH*)	2–5% Glucose	Constitutive
Metallothionein (*CUP1*)	0.03–0.1 mM copper	Inducible
Phosphoglycerate kinase (*PGK*)	2–5% Glucose	Constitutive
Triose phosphate isomerase (*TPI*)	2–5% Glucose	Constitutive
UDP galactose epimerase (*GAL10*)	Galactose	Inducible

Yeast are efficient secretors !

Secretory expression preferred if:

-> if product toxic

-> if many S-S bonds need to be closed

TABLE I Partial List of Human Proteins Produced in *Saccharomyces cerevisiae* at High Levels[a,b]

Protein (molecular weight)	Reference/source
Cytoplasmic expression	
Cu, Zn superoxide dismutase (15.7 kDa)	Hallewell *et al.*, 1987, 1991
γ-Interferon (16.8 kDa)	Derynck *et al.*, 1983; Fieschko *et al.*, 1987
Fibroblast growth factor (18 kDa)	Barr *et al.*, 1988; G. McKnight, pers. comm.
Hepatitis B surface antigen (22 kDa)	Hitzeman *et al.*, 1983a; Miyanohara *et al.*, 1983; McAleer *et al.*, 1984; Bitter and Egan, 1984; Bitter *et al.*, 1988; Jacobs *et al.*, 1989
Hepatitis B core antigen (22 kDa)	Kniskern *et al.*, 1986
Lipocortin V (35 kDa)	M. Irani, pers. comm.
Gα,1 subunit (41 kDa)	T. Jones and V. L. MacKay, unpublished
α-1-Antitrypsin (44.7 kDa)	Cabezón *et al.*, 1984; Casolaro *et al.*, 1987
Platelet-derived endothelial cell growth factor (45 kDa)	Finnis *et al.*, 1992
Hemoglobin (62 kDa)	Wagenbach *et al.*, 1991; Coghlan *et al.*, 1992; Ogden *et al.*, 1991, 1992
cAMP phosphodiesterase isozyme IV (77 kDa)	McHale *et al.*, 1991
Coagulation factor XIIIa (83 kDa)	Bishop *et al.*, 1990
HIV-1 reverse transcriptase (117 kDa)	Bathurst *et al.*, 1990
Secretory expression	
Epidermal growth factor (5.5 kDa)	Brake *et al.*, 1984; George-Nascimento *et al.*, 1988
Insulin precursors (6 kDa)	Thim *et al.*, 1986
Insulin-like growth factor I (7.5 kDa)	Bayne *et al.*, 1988; Steube *et al.*, 1991
Parathyroid hormone (9.4 kDa)	Gabrielsen *et al.*, 1990
Granulocyte-macrophage colony-stimulating factor (14 kDa)	Miyajima *et al.*, 1986; Ernst, 1988
Lysozyme (14.7 kDa)	Jigami *et al.*, 1986; Taniyama *et al.*, 1988; Ichikawa *et al.*, 1989
Interleukin-1α, β (17 kDa)	Baldari *et al.*, 1987; Ernst, 1988; Livi *et al.*, 1990
α-Interferon (20 kDa)	Hitzeman *et al.*, 1983b; Singh *et al.*, 1984; Bitter *et al.*, 1984; Mellor *et al.*, 1985; Chang *et al.*, 1986; Zsebo *et al.*, 1986
Growth hormone (22 kDa)	Hitzeman *et al.*, 1984; Tokunaga *et al.*, 1985; Hiramatsu *et al.*, 1990
Interleukin-6 (22 kDa)	Guisez *et al.*, 1991
Erythropoietin (24 kDa)	Elliott *et al.*, 1989
Platelet-derived growth factor (30 kDa)	Kelly *et al.*, 1985; Östman *et al.*, 1989
Thrombin zymogens (36–69 kDa)	H. Han and V. L. MacKay, unpublished
Single-chain urokinase (47 kDa)	Melnick *et al.*, 1990
β1–4 Galactosyltransferase (48 kDa)	Krezdorn *et al.*, 1993
Chimeric L6 antibody Fab (48 kDa)	Horwitz *et al.*, 1988; Better and Horwitz, 1989

(*Continues*)

Expression in *S. cerevisiae* – *Pichia pastoris*

Problems with production in S. cerevisiae:

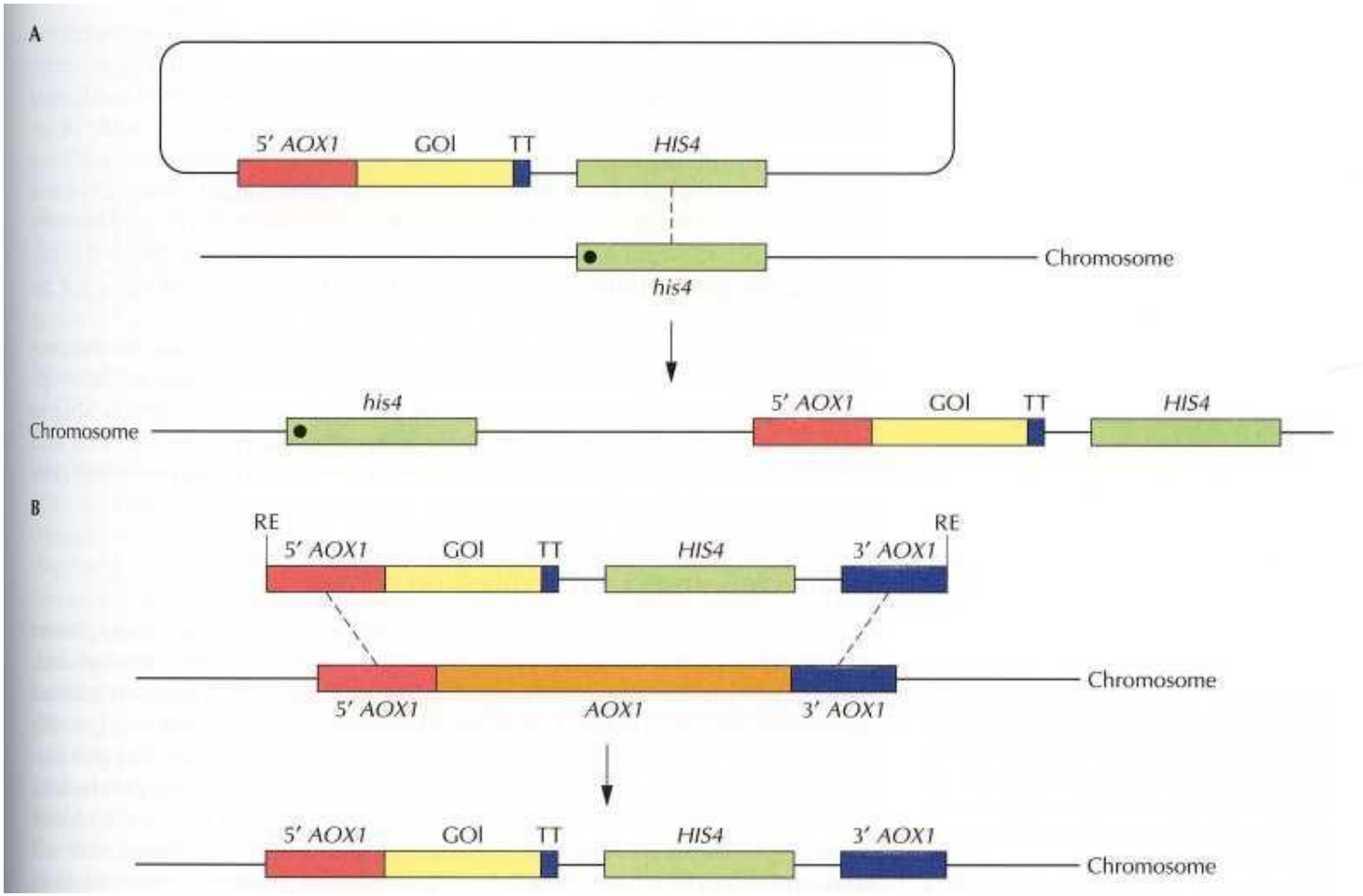
- For some proteins production level low
- Hyperglycosylation (more than 100 mannose residues in N-glycosylation)
- Sometimes secretion not good -> protein stack in cells (periplasma)
- S. cerevisiae produces high amount of EtOH -> toxic for the cells -> effects level of production

Advantages of production in *Pichia pastoris*:

- Highly efficient promoter, tightly regulated (alcohol oxidase -> AOX, induced by MeOH)
- Produces no EtOH -> very high cell density -> secretion very efficient
- Secretes very few proteins -> simplification of purification of secreted proteins

Expression in *Pichia pastoris*

Integrative systems



Expression in *Pichia pastoris*

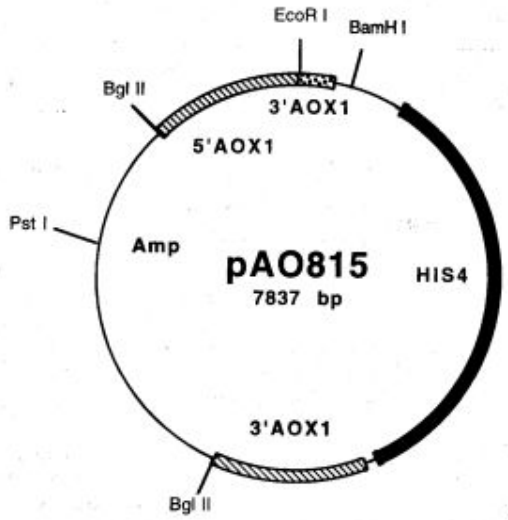


Figure 2 *P. pastoris* expression vector pAO815. The thin line represents sequences from *E. coli* plasmid pBR322 (with permission from Cregg and Higgins, 1995).

Table 2
Common *P. pastoris* Expression Vectors

Vector name	Selectable markers	Feature	Reference
Intracellular			
pHIL-D2	HIS4	NotI sites for *AOX1* gene replacement	K. Sreekrishna (personal communication)
pAO815	*HIS4*	Expression cassette bounded by *Bam*HI and *Bgl*II sites for generation of multicopy expression vector	Thill *et al.* (1990)
pPIC3K	*HIS4* and *kan*^{r}	Multiple cloning sites for insertion of foreign genes; G418 selection for multicopy strains	Scorer *et al.* (1993b)
pPICZ	*ble*^{r}	Multiple cloning sites for insertion of foreign genes;	Higgins *et al.* (1998)

(continues)

Expression in *Pichia pastoris*

Table 4
Advantages and Disadvantages of the *P. pastoris* Expression System

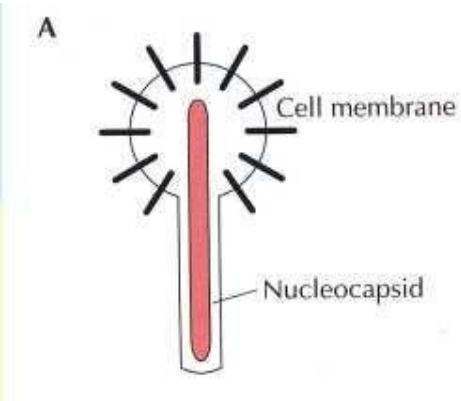
Advantages	Disadvantages
Culturing	
Rapid growth rate	
High-cell density: >100 g (dcw)/liter	
Clean medium composed of salts, biotin, and carbon source	
Easily scaled up to large-volume, high-density fermentor cultures	Fermentor culturing often needed to achieve high level of foreign protein
Molecular genetics	
Classical genetic methods available	
Molecular methods similar to *S. cerevisiae*	Range of vectors limited
Stable integrated expression vectors	
Promoters	
AOX1p—strong, tightly regulated, and easily controlled	
GAPp—strong constitutive	
Expression	
Eukaryotic environment aids folding of higher eukaryotic foreign proteins	
High expression levels	
Secretion	Improper posttranslational modifications
Proper posttranslational modifications	
Sulfhydril bond formation	Native signals not always processed
Signal sequence processing	
Folding	Some proteins misfold and become stuck in secretory pathway
Glycosylation	
High levels—g/liter fermentor cultures	Lower eukaryotic (high mannose)-type glycosylation
Few yeast proteins in medium—high initial purity of foreign protein	Proteases in medium degrade some foreign proteins

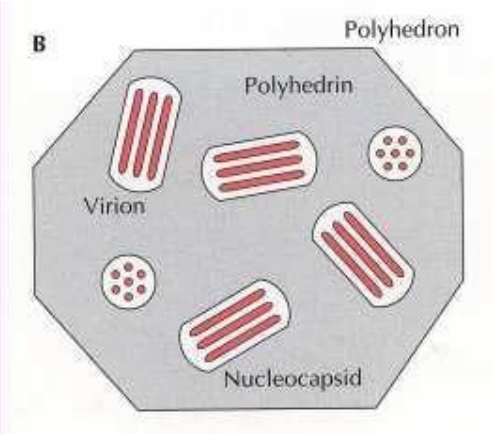
Expression in Insect cells

- Baculovirus:

-> infects invertebrates (insects)
-> in infection cycle 2 forms of baculovirus are formed:
-> single virus particle
-> in protein matrix (polyhedron) trapped clusters of viruses

-> during late stage of infection massive amount of polyhedron produced -> strong promoter
-> polyhedron not required for virus production
-> **polyhedron promoter** optimal for heterologous protein production in insect cells





Expression in Insect cells

- Baculovirus:

-> ***Autographa californica*** **multiple nuclear polyhedrosis virus** (**AcMNPV**) many used as expression vector

-> **Production of recombinant baculovirus:**

1. create a transfer vector (*E. coli* based plasmid with AcMNPV DNA – polyhedrin promoter/terminator + flanking sequences) -> gene of interest cloned downstream of promoter

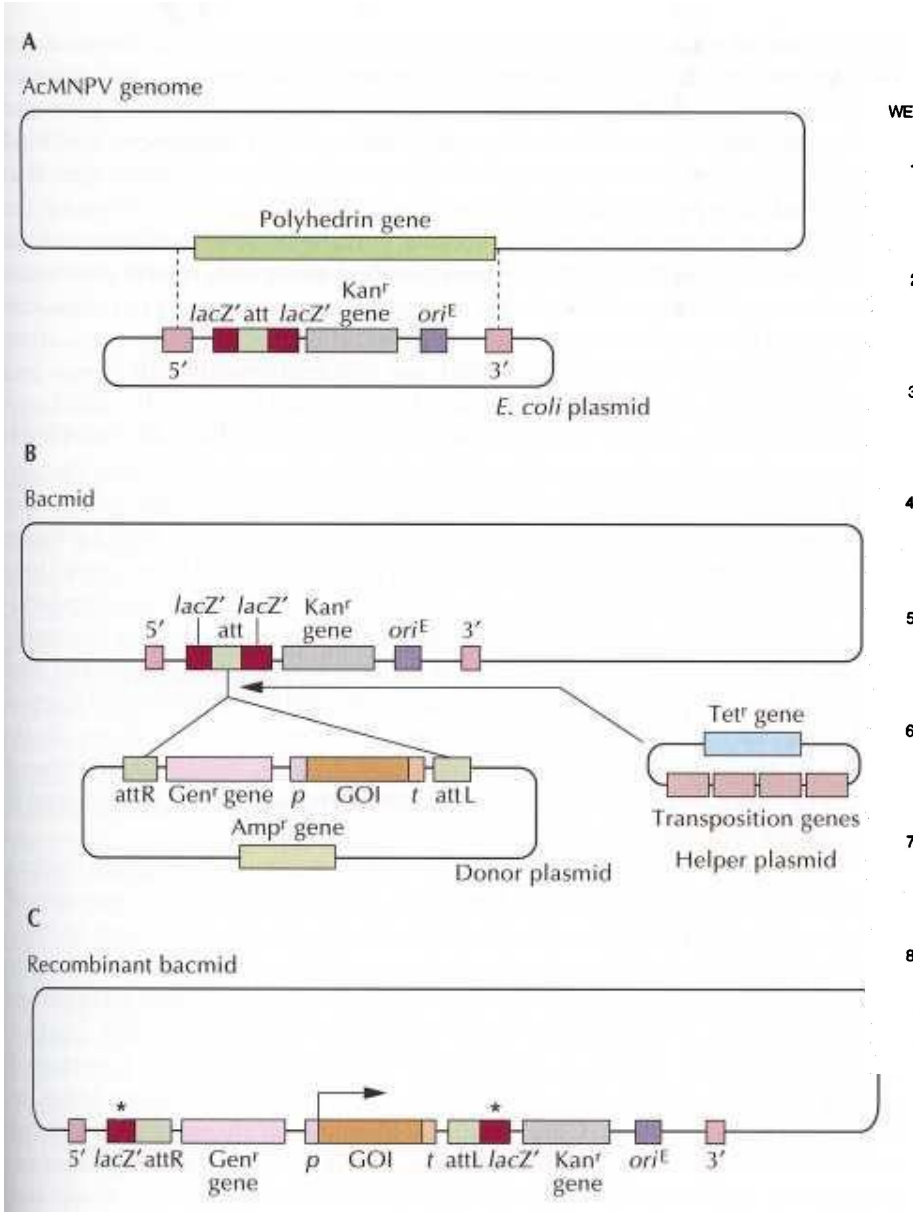
2. Insect cells are cotransfected with virus (AcMNPV) + transfer vector
   -> in some double infected cells -> double crossover event (recombination)
   -> produce recombinant virus (bacmid -> *E. coli* - insect cell baculovirus shuttle vector)

   -> cells infected with recombinant virus -> produce plaques (lack of polyhedrin)

3. DNA hydridisation + PCR used to identify recombinant virus

4. Infection of insect cells with concentrated stock of verified recombinant virus
   -> 4-5 days later protein harvested

Baculovirus expression system



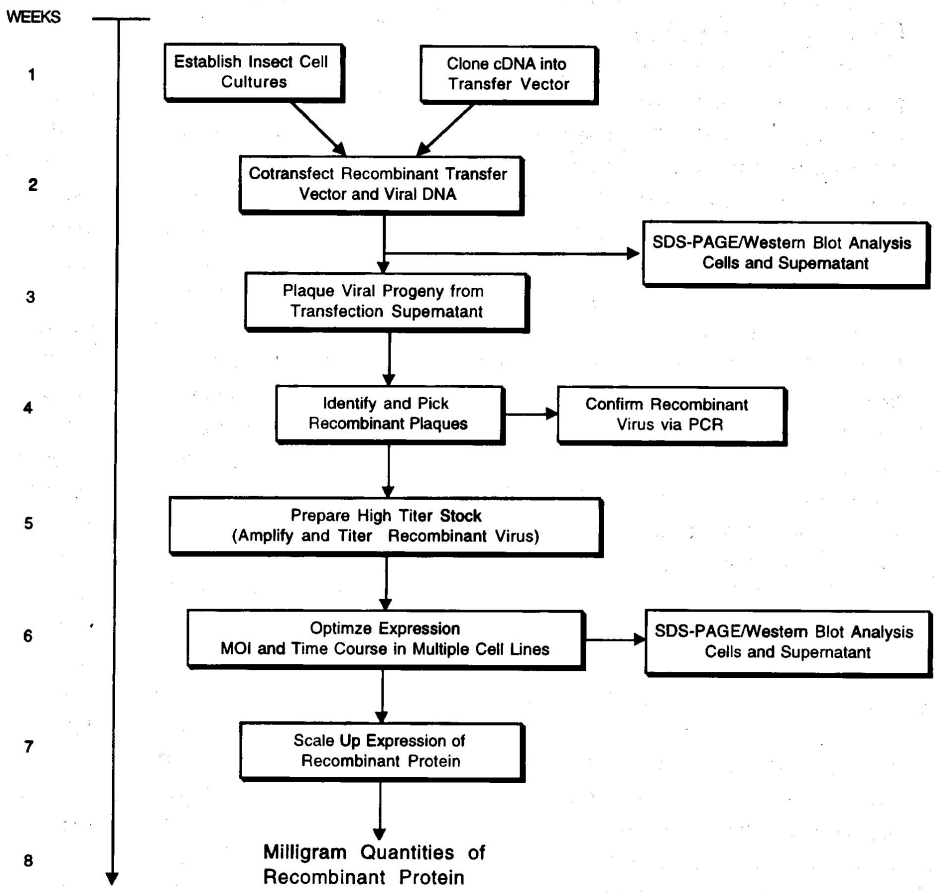


Figure 2 Flow chart of BEVS outlining the generation and characterization of recombinant baculoviruses.

Baculovirus expression system

Why this system?

1. Insect cells have almost the same posttranslational modifications as mammalian cells
2. Higher expression level than mammalian cells

Table 4
Heterologous Protein Expression

Protein	Origin	Reference
β-Hexoaminidase B	Human	Pennybacker *et al.* (1997)
Myosin light chain kinase	Human	Lin *et al.* (1997)
DNA helicase II	Human	Zhang and Grosse (1997)
Bullous pemphigoid antigen 2	Human	Masunaga *et al.* (1997)
Lymphotoxin-α1β2 complex	Human	Williams-Abbott *et al.* (1997)
Werner's syndrome (WRN) helicase	Human	Suzuki *et al.* (1997)
Angiostatin	Murine	Wu *et al.* (1997)
Serotonin 5-HT7 receptor	Murine	Obosi *et al.* (1997)
Replication factor C (hRFC)	Human	Cai *et al.* (1997)
Endothelial nitric oxide (eNOS)	Bovine	Ju *et al.* (1997)
Growth hormone receptor	Human	Bieth *et al.* (1997)
MAL	—	Puertollano *et al.* (1997)
Leptin receptor	—	Devos *et al.* (1997)
Insulin receptor substrate-1	Human	Algenstaedt *et al.* (1997)
Manganese-dependent superoxide dismutase	Human	Wright *et al.* (1997)
Parvovirus VP2 capsid protein	Porcine	Sedlik *et al.* (1997)
Varicella-zoster virus (VZV) glycoproteins E and I (heterodimer)	—	Kimura *et al.* (1997)
Hawaii calicivirus capsid protein	—	Green *et al.* (1997)
Cardiac Ca^{2+} pump and phospholamban	Canine	Autry and Jones (1997)
Type IV collagenase/gelatinase	Human	George *et al.* (1997)
Eukaryotic initiation factor-2B (eIF-2B)	Human	Fabian *et al.* (1997)
Cellobiohydrolase	Fungal	von Ossowski *et al.* (1997)
Respiratory syncytial virus fusion protein	—	Parrington *et al.* (1997)

Mammalian cell expression system

1. Why do we use that system?
   -> to get full complement of posttranslational modifications on proteins

2. Developed cell lines:
   -> short term (transient) expression -> autonomous replicating systems -> viral origins (SV40)
      - African green monkey kidney (COS)
      - baby hamster kidney (BHK)
      - human embryonic kidney (HEK-239)

   -> long term (stable) expression -> integration into chromosome -> viral origins
      - chinese hamster ovary (CHO)

Mammalian cell expression system

Table 3
Comparison of Sindbis Virus and Other Viral Expression Systems

	Sindbis virus	Retrovirus	Adenovirus	Baculovirus
Expression	Transient	Stable	Transient	Transient
Expression level	High	Depends on integration site	High	High
Construction	Plasmid construction	Requires use of cell line	Requires recombination	Requires recombination and plaque purification
Tropism	Wide	Depends on packaging cell line	Variable	Insect cells
Infection efficiency	High	Variable	Variable	High
Host protein shutoff	Yes	No	Yes	Yes

Gene expression in mammalian cell lines

A convenient alternative for setting up mammalian cell facilities – get a comprehensive service from us. We will achieve stable expression of the gene of your interest in mammalian cells.

Customer provides:

- Mammalian vector with the gene (cDNA) to be expressed. We accept plasmid and retroviral vectors
- Sequence of the gene and map of the construct for transfection
- Cell line or information about the cell line to be transfected.

Our service includes:

- Transfection of the cells. In case of a retroviral vector, virus production and cell infection
- Antibiotic selection and generation of stable transfected (infected) cell clones. At least 10 independent clones will be selected and grown
- Quantitative assay of the gene (cDNA) expression level in each transfected clone by RNA isolation followed by Northern hybridisation and/or RT-PCR
- Selection of the best expressing clone
- Cell freezing and depositing
- Duration: 3-6 months (depending on the cell growth rate), allow 1month in addition if the cell line is not available in our collections

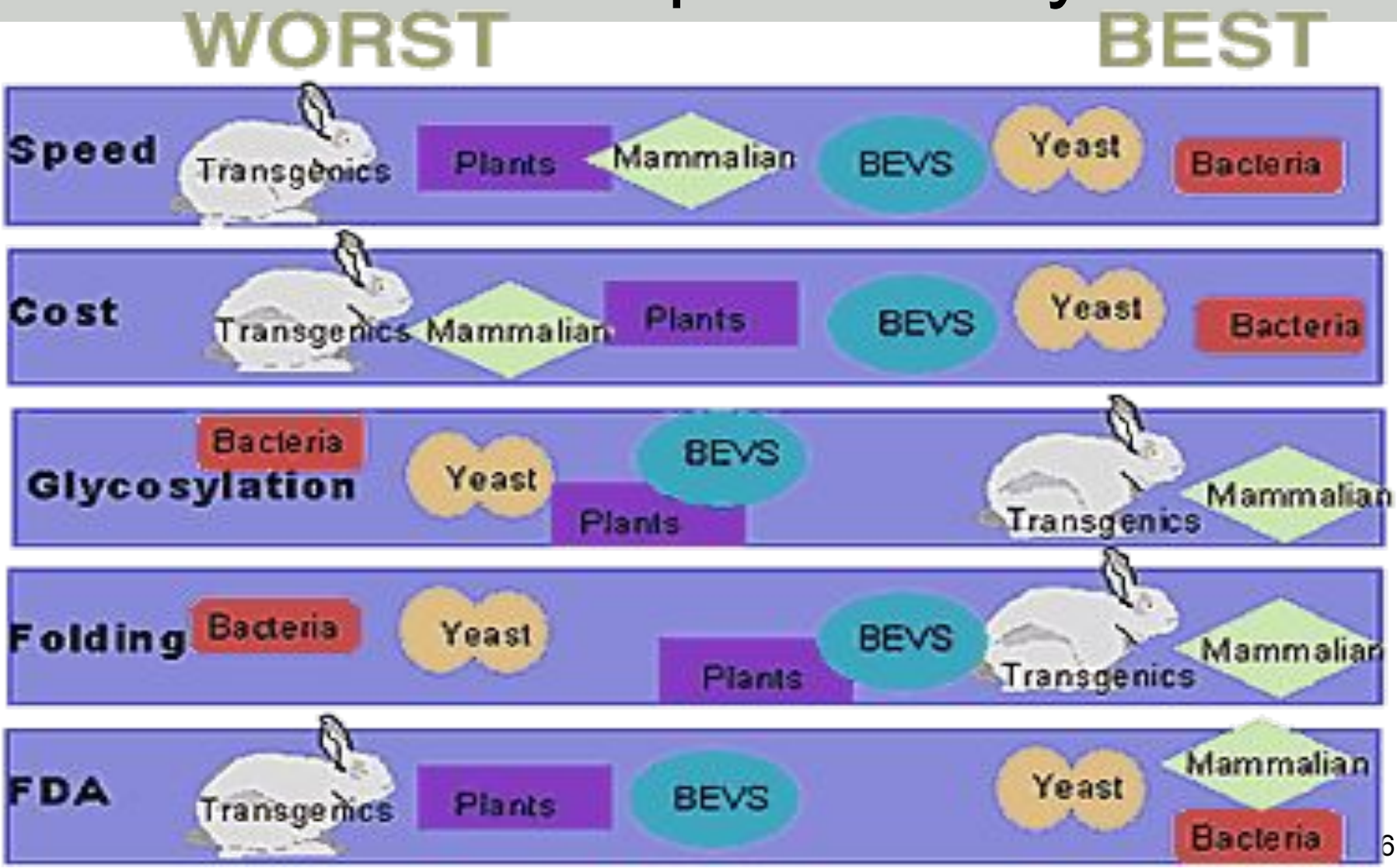
Customer receives:

- Detailed report on experiments and data obtained.
- Two vials of transfected cells (the best expressing clone)
- We will deposit the transfected cells in our collection as a precaution against accidental loss of the clone.

Price guide:

Price per transfection and selection of at least 10 clones: £3500.

Competitiveness of different expression systems



http://www.proteinsciences.com/technology/pix/best_worse.gif