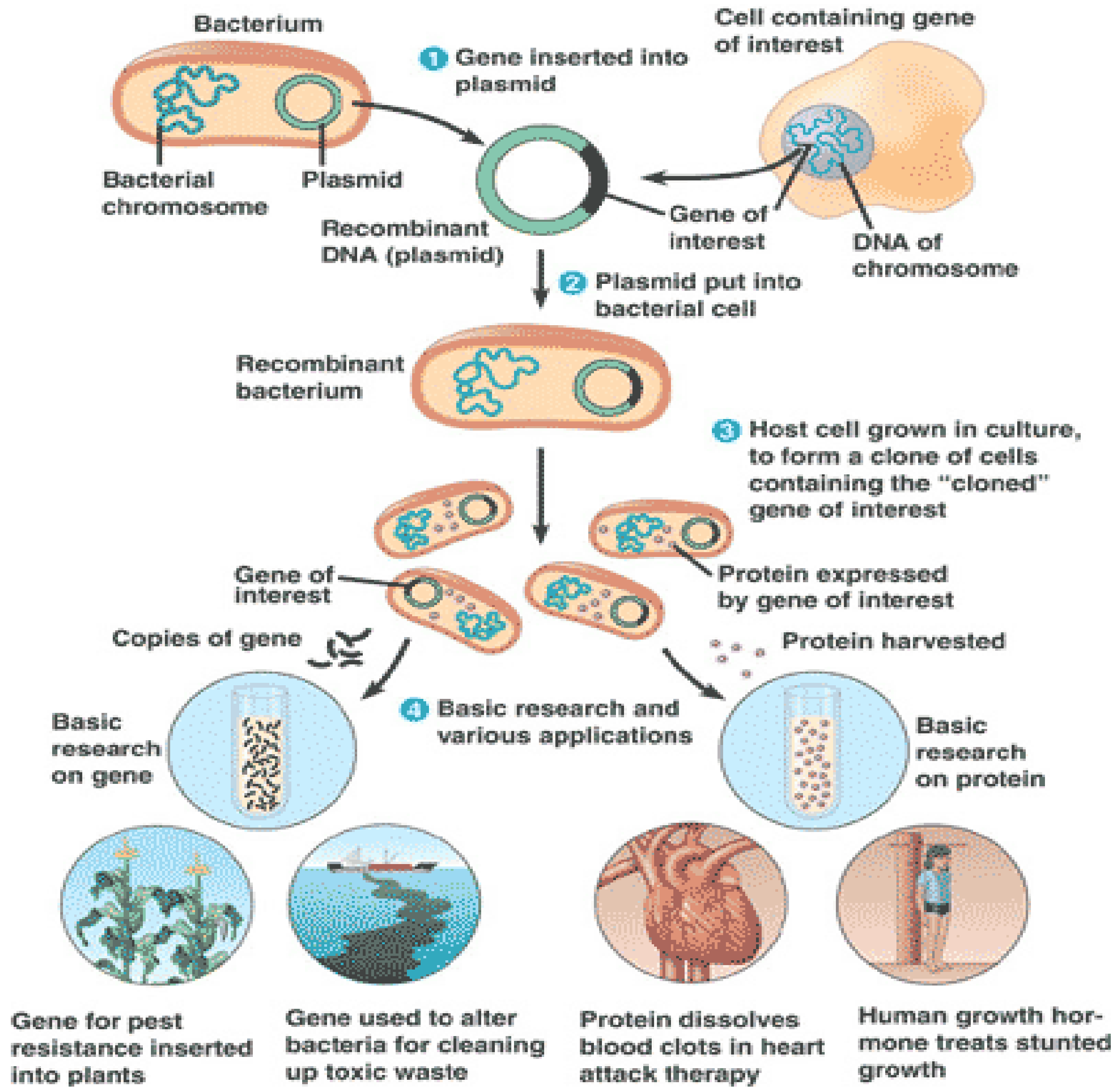
Therapeutic and Industrial Recombinant Proteins

Farid Ataya

Introduction

- Recombinant proteins have been produced for over 30 years
- Applications range from recombinant proteins (mostly enzymes) used in laundry detergents, therapeutic enzymes, replacement therapy for some hormones, antigen-detecting antibodies in cancer therapy, etc.
- For centuries, humans have been using cultured cells since the dawn of civilization like the fermentation of grains by microbes into beer, production of alcohol, vinegar, cheese, bread, etc…
- In the 20th century, the cultured cells are used for medicinal and industrial production of important compounds.
- With the advent of recombinant DNA technology, a large variety of proteins became available.
- Today, increasing number of recombinant proteins (> 170) are used worldwide in medicine. Moreover, protein engineering techniques are used to introduce modifications on such proteins to increase their solubility, stability, activity, etc.



Bacterium
Bacterial chromosome
Plasmid
1 Gene inserted into plasmid
Cell containing gene of interest
Recombinant DNA (plasmid)
Gene of interest
DNA of chromosome
2 Plasmid put into bacterial cell
Recombinant bacterium
3 Host cell grown in culture, to form a clone of cells containing the "cloned" gene of interest
Gene of interest
Protein expressed by gene of interest
Copies of gene
Protein harvested
4 Basic research and various applications
Basic research on gene
Basic research on protein
Gene for pest resistance inserted into plants
Gene used to alter bacteria for cleaning up toxic waste
Protein dissolves blood clots in heart attack therapy
Human growth hormone treats stunted growth

	Bacterium	Yeast	Insect	Mammalian
Prokaryotic protein expression	+++	+	+	+
Eukaryotic protein without PTM	+++	+++	+++	+++
Eukaryotic protein with PTM#	-	++	++	+++
PTM, folding	+	++	+++	+++
PTM, glycosylation	-	++ High mannose	++	+++
PTM, phosphorylation	-	++	++	+++
Turnaround time	Short	Moderate	Long	Long
Cost	Low	Low	Middle	Middle
Yield (mg/L)	1-5mg/l with an average level 3mg/l	0.1-5mg/l with an average level 1mg/l	0.2-50mg/l with an average level 1mg/l	1-1000mg/l (average 50mg/l for antibody, 3mg/l for protein)
Recommended applications	Antigen protein, Protein standards, Functional proteins	Proteins with glycosylation, Vaccine, Secreted form	Proteins with glycosylation, Assay standards, Secreted form	Functional study, PTM study, Assay standards, Characterization

Methods of disrupting cell membrane

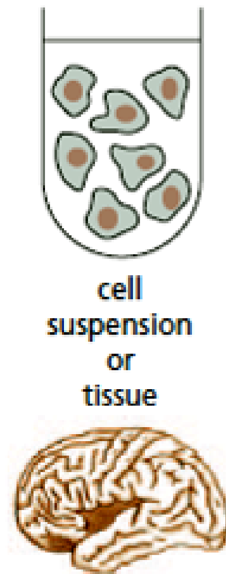
- Breaking of cells can be achieved by
  - Physical methods
    - Mechanical methods
    - Other methods
  - Chemical methods
  - Biochemical methods
  - Thermal methods

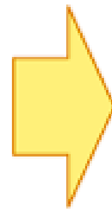
Methods of disrupting cell membrane

Technique	Principle	Advantages (+) / Disadvanatges (-)
Liquid shear pressure (e.g., French press)	Rapid pressure drop by transferring the sample from a chamber at high pressure through an orifice into a chamber at low pressure	+ Fast and efficient, also for large volumes - Causes heating of the sample (cooling is required)
Ultrasonication	Cells disrupted by high frequency sound	+ Simple - Causes heating of the sample, which can be difficult to control by cooling - Proteins may be destroyed by shearing - Noisy - Not for large volumes
Glass bead milling	Agitation of the cells with fine glass beads	+ Useful for cells that are more difficult to disrupt (e.g., yeast) - Somewhat slow and noisy
Osmotic shock	Change from high to low osmotic medium	+ Simple, inexpensive - Only useful for disruption of cells with less robust walls (e.g., animal cells)
Repeated freezing and thawing	Cells disrupted by repeated formation of ice crystals; usually combined with enzymatic lysis	+ Simple, inexpensive + Yields large membrane fragments - Slow - May damage sensitive proteins and dissociate membrane protein complexes - Low yield
Enzymatic lysis	Often used in combination with other techniques, e.g., freeze-thawing or osmotic shock; lysozyme is commonly used to break cell walls of bacteria	+ Gentle + Yields large membrane fragments - Slow - Low yield

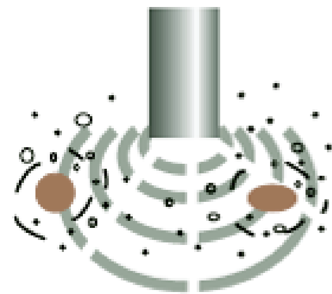
BREAKING CELLS AND TISSUES

The first step in the purification of most proteins is to disrupt tissues and cells in a controlled fashion.

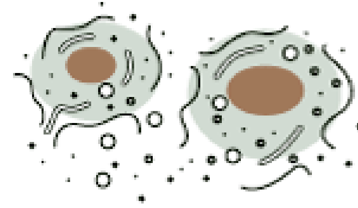




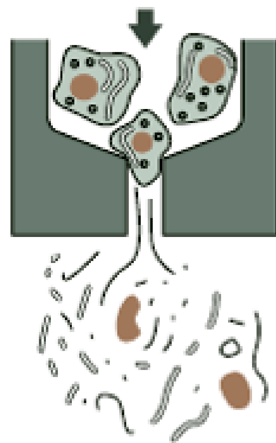
Using gentle mechanical procedures, called homogenization, the plasma membranes of cells can be ruptured so that the cell contents are released. Four commonly used procedures are shown here.



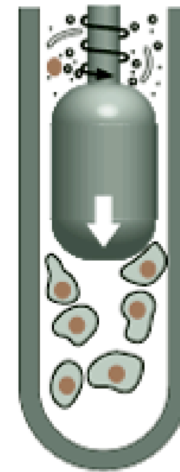
1 Break cells with high-frequency sound (ultrasound).



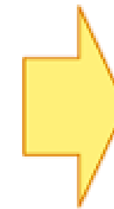
2 Use a mild detergent to make holes in the plasma membrane.



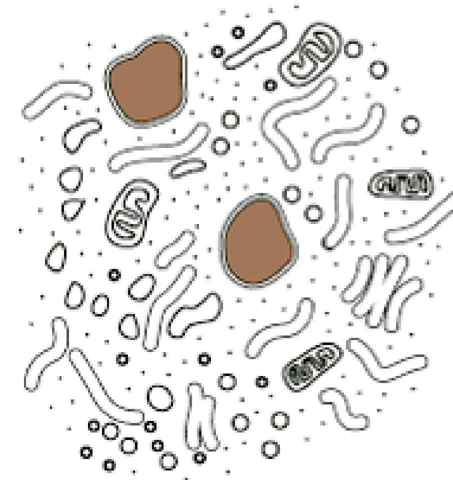
3 Force cells through a small hole using high pressure.



4 Shear cells between a close-fitting rotating plunger and the thick walls of a glass vessel.



The resulting thick soup (called a homogenate or an extract) contains large and small molecules from the cytosol, such as enzymes, ribosomes, and metabolites, as well as all of the membrane-enclosed organelles.



When carefully conducted, homogenization leaves most of the membrane-enclosed organelles largely intact.

Cell disruption

- Chemical disruption – mild method
- Permeabilizing the cell using chemicals
- Solvents – DMSO, Alcohol, Methanol, chloroform creates pores in the cells membrane
- Other chemicals like some antibiotics, chelating agents, chaotropic agents and Detergents can be used

Cell disruption

- Chemical disruption – mild method
- Chelating agent (EDTA) –
  - Binds to divalant cations like Ca^{2+} and Mg^{2+} which are essential for stabilization of outer membrane of cells
  - EDTA removes these cations
  - Cell membrane integrity disturbed
- Might effect co-factors needed for enzyme function

Cell disruption

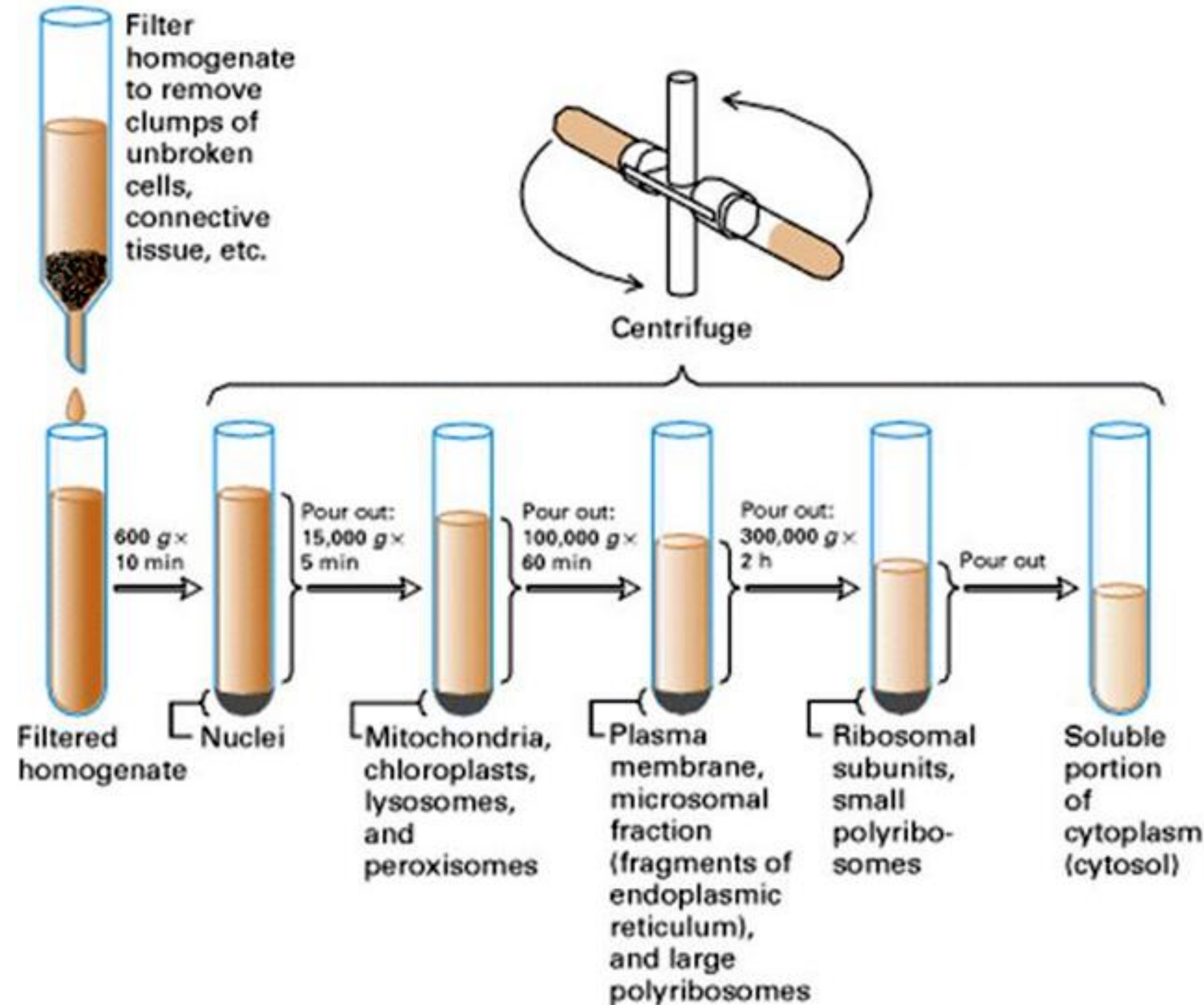
- Chemical disruption – mild method
- Chaotropic agents – guanidine HCl, Urea…
  - Bring hydrophobic compounds in to hydrophilic environment
  - Disrupt hydrophobic interactions of solute
  - Disrupts hydrophilic environment of water (disrupts hydrogen bonds of water)
  - Causes denaturation of proteins in membrane
- Some detergents can also be employed

Cell disruption

- Enzymatic disruption
  - Enzymes can also be used to permeabilize the cells
  - Glyconases, proteases, mannases…
  - Good for periplasmic proteins
  - In combination with EDTA or GuHCl
  - Expensive, saturation problems

- Lysozyme used to lyse G+ve bacteria

Separation of cell components by Differential centrifugation



Recombinant Protein Production

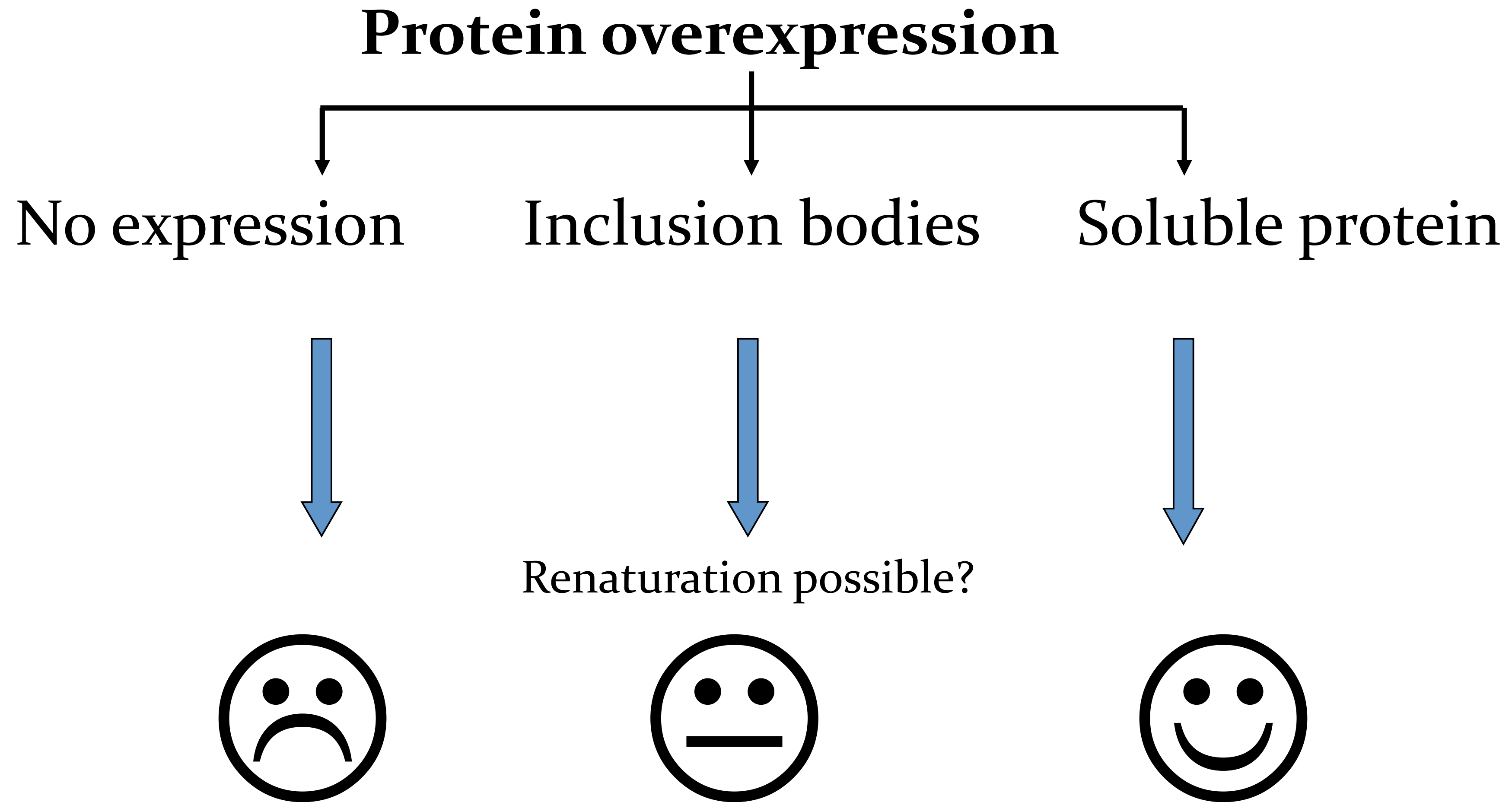
-Why?

- over-expression to get enough amount
- easy purification

-Application

- functional studies
- structural studies
- vaccine/antigen/antibodies
- therapeutic drug
- industrial enzymes for reaction

Protein expression: common problems



Things needed to be considered for recombinant protein expression

1. **How to produce?**

   choose for protein expression system (vector and host)

2. **How to make an expression recombinant DNA construct?**

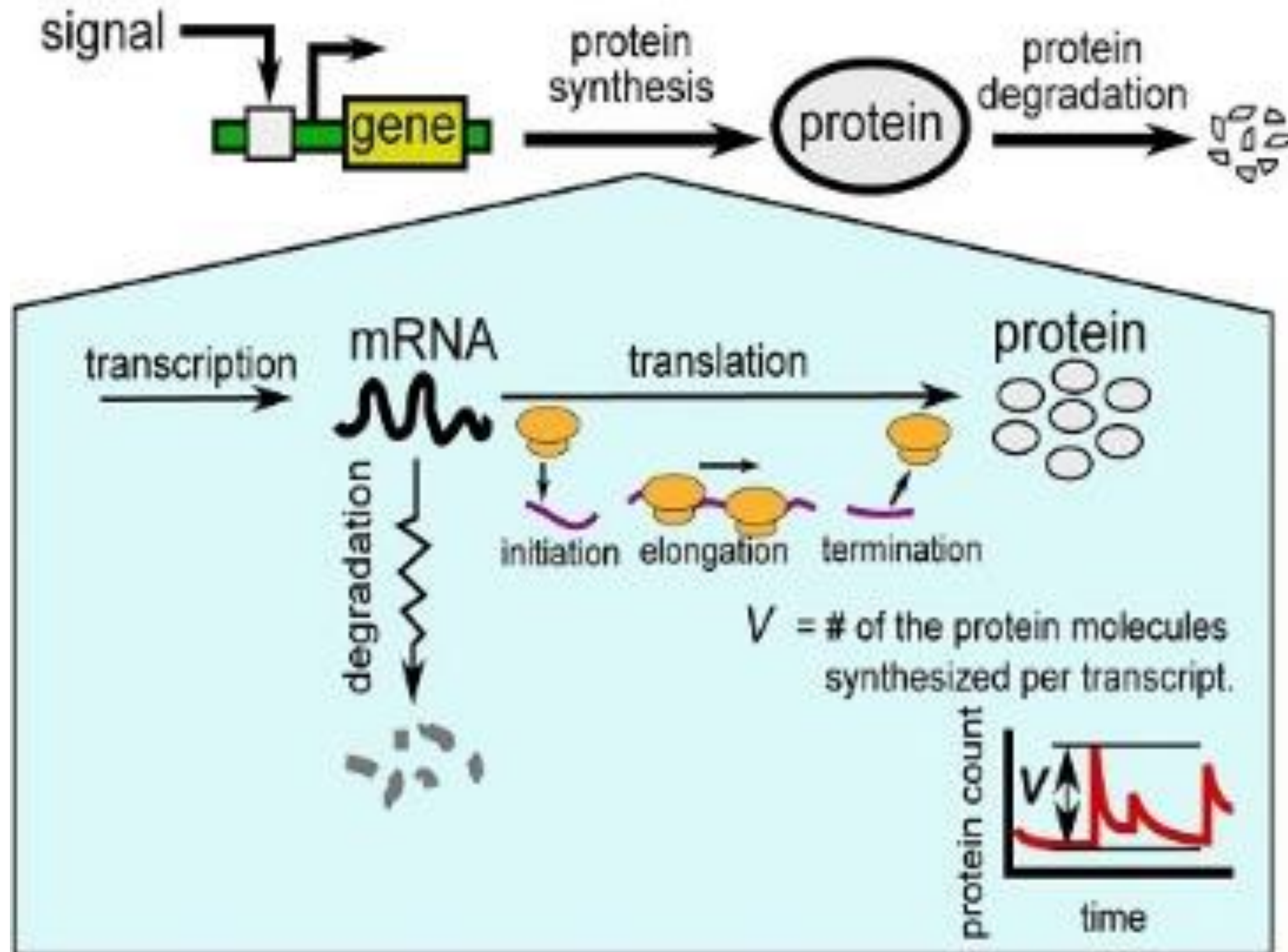
   translational or transcriptional fusion, promoter use (inducible or constitutive)

3. **Where to express?**

   cytosol, periplasm, secretion, inclusion body

4. **Difficulties (protein expression problems)**

Stability of the product



Prokaryotic expression systems-

advantages V disadvantages

- e.g. *Escherichia coli*, *Lactococcus lactis*, *Bacillus* species….
- Widely used for expression of recombinant proteins
- Easy manipulation/transformation, rapid growth, simple nutrient requirements
- Many commercial vectors (*e.g.* Invitrogen, Novagen, Stratagene) available with a variety of N- and C-terminal tags to facilitate purification (*e.g.* His tag, GST fusions, Trx fusions)
- Well suited for proteins to be used for Ab production, structural, functional studies

- Many proteins expressed in inclusion bodies
- No post-translational modifications
- Improper folding of disulphide linked proteins
- Occasional problems with removal of fusion partner
- High endotoxin content with G-ve
  - (Reichelt et al. (2005) Single step protocol to purify recombinant proteins with low endotoxin content. Prot. Expr. Purif.

Where to express the recombinant proteins?

1.Direct expression (cytosol):

E. coli cytoplasm is a reducing environment - difficult to ensure proper disulphide bonds formation.

2. Fusion expression (inclusion body?):

Ensures good translation initiation. Can overcome insolubility and/or instability problems with small peptides. Has purification advantages based on affinity chromatography.

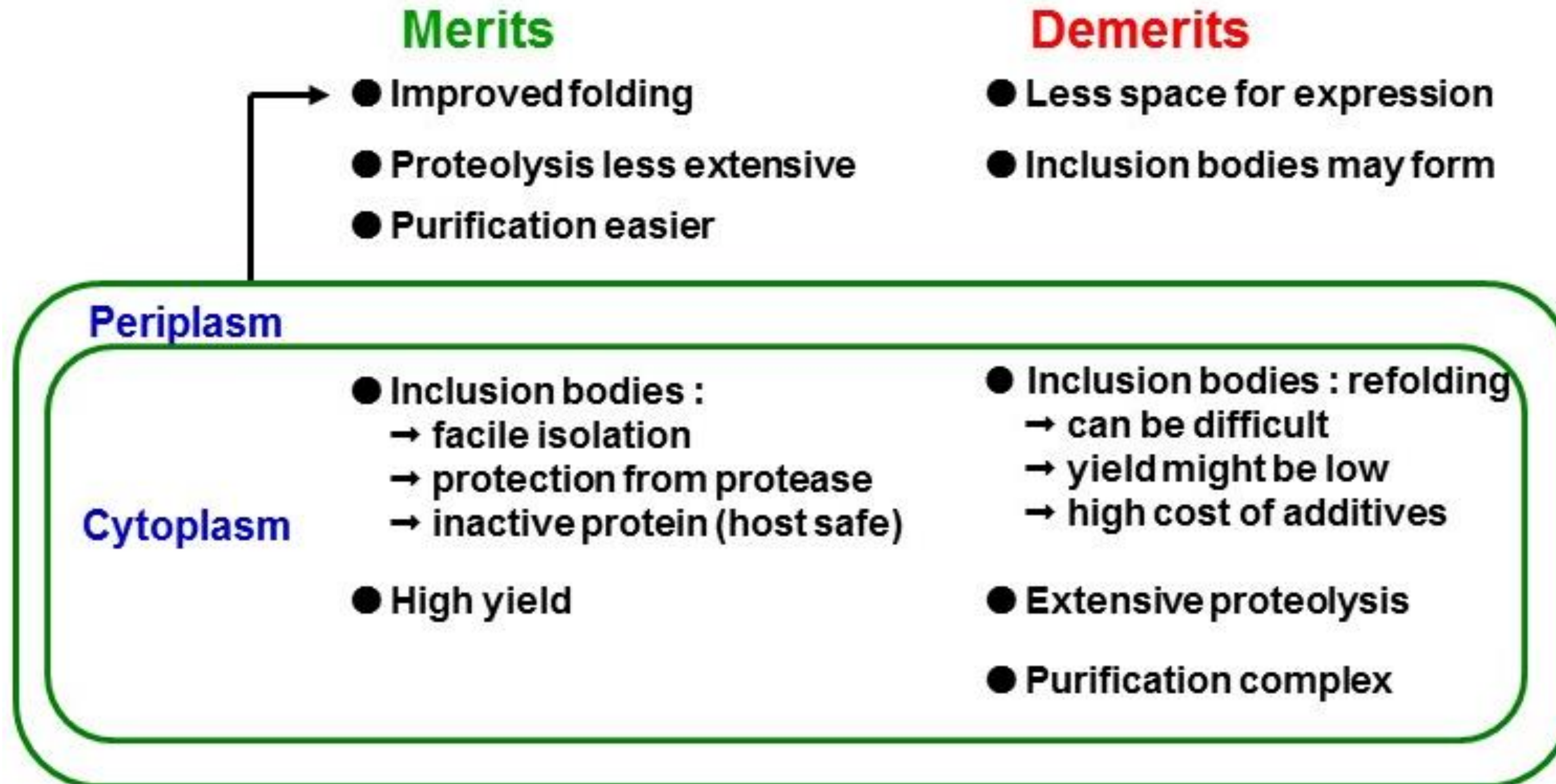
3. Secretion (periplasm or medium):

a fusion alternative when proteins are fused to peptides or proteins targeted for secretion. Periplasm offers a more oxidizing environment, where proteins tend to fold better.

Major drawbacks: limited capacity for secretion (0.1-0.2% total cell protein compared to 10% produced intracellularly) and **inability for posttranslational modifications** of proteins.

Where to express the recombinant proteins?

A intracellular product can be made extracellular by adding leader / signal sequence in recombinant enzyme expression



E. coli

Exporting recombinant proteins in the periplasm

Advantage of protein expression in periplasm:

- suitable for guest proteins containing multiple disulfide.
- low content of proteases reduce the severity of proteolysis.
- low concentration of contaminating proteins.
- So, protein purification is easier.
- Cleavage of signal peptide by the activity of signal peptidase 1.

Disadvantage:

Periplasm is nearly 7 nm wide and Inclusion Bodies may form.

The yield of recombinant proteins are higher in the cytoplasm.

Strategy to choose of protein expression system

The *KEY* idea is the cloned gene must be transcribed and translated most efficiently.

***Expression vector*: *MAXIMIZE GENE EXPRESSION*.**

***Host*: *MAXIMIZE* TURNOVER OF GENE PRODUCTS**

Preventing proteolysis *in vivo* in *E. coli*:

---- Use protease deficient mutants as hosts.

Lon - a major ATP-dependent protease in *E. coli*. Has broad specificity for unfolded or misfolded proteins *in vivo*.

ompT - an outer membrane localized protease. Cleaves at paired basic residues.

degP - periplasmic protease - could inactivate some secreted proteins.

- **BL21(DE3) strain**
  - *lon* and *ompT proteases* deficient
  - Carries a lambda DE3 lysogen, the *lacI* gene and *lacUV5*-driven T7 RNA polymerase

Strain	Derivation	Key Feature(s)	Antibiotic Resistance	Available as Competent Cells
Rosetta	Tuner™	BL21 *lacYZ* deletion,	Cam	yes
Rosetta(DE3)	(B)	Lacks *lon* and *ompT*	Cam	yes
Rosetta(DE3)pLysS		proteases	Cam	yes
Rosetta(DE3)placI			Cam	yes
RosettaBlue™	NovaBlue	*recA, endA, lacI*^{q}	Tet + Cam	yes
RosettaBlue(DE3)	(K-12)	High transformation	Tet + Cam	yes
RosettaBlue(DE3)pLysS		efficiency	Tet + Cam	yes
RosettaBlue(DE3)pLacI			Tet + Cam	yes
Rosetta-gami™	Origami™	*trxB/gor* mutant, greatly	Kan + Tet + Cam	yes
Rosetta-gami(DE3)	(K-12)	facilitates cytoplasmic	Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLysS		disulfide bond formation	Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLacI			Kan + Tet + Cam	yes

Novagen

Features of the BL21-Derived Expression Strains[5]

Expression Strain	Induction	Advantages	Disadvantages
BL21-Gold (DE3) competent cells	IPTG induction of T7 polymerase from *lacUV5* promoter	High-level expression; Ease of induction, direct cloning in expression strain	Leaky expression of T7 polymerase can lead to uninduced expression of potentially toxic proteins.
BL21-Gold (DE3)pLysS competent cells	IPTG induction of T7	Ease of induction, reduced uninduced expression, direct cloning in expression strain	Slight inhibition of induced expression when compared with BL21-Gold (DE3).
BL21 competent cells	Infection with lambda bacteriophage CE6	Tightest control of uninduced expression	Induction not as efficient as DE3 derivatives. Induction (infection) process is more cumbersome.

General problems with heterologous gene expression

***(a) Not enough protein is produced*:**

→ codon usage preferential (rare codon)

→ potential mRNA secondary structure. (5'-end AT content, 3'-end transcriptional terminator)

→toxic gene.

***(b) Enough protein is produced, but it is insoluble*:**

→vary the growth temperature.

→ change growth medium.

→low-copy-number plasmids.

→ selection of promoter.

The *KEY* idea is to slow down the expression rate of protein.

Popular promoters for heterologous protein expression in E. coli

1. **T7 system**. Uses T7 promoters, which require T7 RNA polymerase. T7 RNA polymerase (encoded by T7 gene 1) has stringent specificity for its own promoters. It initiates and elongates chains 5 times faster than *E. coli* RNA Pol and is resistant to Rifampicin (unlike *E. coli* Pol). pET series of vectors (Plasmid for Expression by T7 RNA pol) commercially available by Novagen.
2. **P*lac*.** Negatively regulated by *lacI*. Need for sufficient levels of repressor (*lacI*q and *lacI*q1 alleles on vectors). P*lacUV5* is very popular because its regulation is not dependent on CAP.
3. **P*trp*.** Negatively regulated by *trp*R. Vectors containing this promoter can be transformed into any strain, easy induction by starvation for tryptophan. Not suitable for expression of proteins with high Trp content.
4. **Hybrid promoters** - P*tac* and P*trc*. Induced by IPTG, a lot stronger than P*lac* and P*trp*.
5. **p*BAD*** - induced by arabinose (Invitrogen)

Optimizing cultivation condition

1. Growth temperature.
2. Inducer concentration
3. Pre-induction growth.
4. Post-induction incubation.

Insolubility of heterologous proteins produced in E. coli

Inclusion bodies:

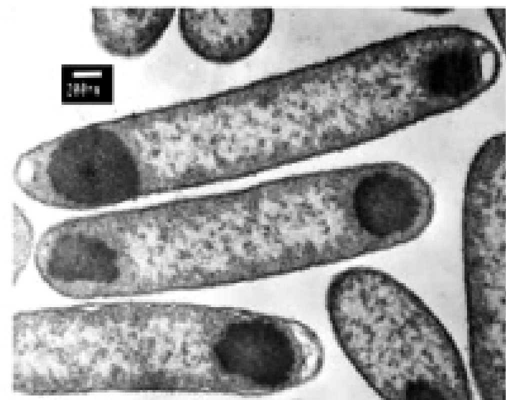
Dense particles, containing precipitated proteins. Their formation depends on protein synthesis rate, growth conditions.

Advantages: proteolysis resistant, big yield, relatively pure, easy to separate.

Disadvantages: inactive product requires in vitro refolding and renaturation

Expression problems

- Protein is insoluble
  - Collect inclusion bodies and refold protein
  - Reduce growth temperature
  - Use heat shock to induce chaperones
  - Use a low/moderate copy number plasmid vector
  - Fuse a periplasmic targetting sequence to N-terminus
  - Co-express chaperones/foldases (PDI…)
  - Try a different fusion partner (e.g. MBP)



Refolding of recombinant proteins

Solubilisation:

High T ^{o} C, detergents, high concentration of inorganic salts or organic solvents all used. The most commonly used organic solutes such as urea or guanidine-HCl often used in the presence of reducing agents (mercaptoethanol or DTT). Solubilized proteins can be purified by ion-exchange chromatography or other conventional methods, prior to refolding.

Refolding:

If no S-S bonds present - remove denaturing agent to allow protein to fold correctly. If S-S bonds present - their formation can be accomplished: by air oxidation, catalysed by trace metal ions; by a mixture of reduced and oxidized thiol compounds - oxidized DTT, reduced DTT; GSSG/GSH; cystine and cysteine, cystamine and cysteamine.

Why bother with recombinant fusion protein or protein engineering?

1. **to minimize proteolysis.**
2. **for efficient and selective purification.**
3. **to optimize translation efficiency.**
4. **for different applications**
   (specific expression scenarios):
   antibody production,
   biochemical experiments,
   structural biology,
   industrial usage.

(protein biotechnology or protein engineering)

Fusion proteins

- Used to avoid problems with digestion of foreign proteins by proteases
- Used to aid purification of foreign proteins, often by affinity chromatography
- Often a rare protease cut site is added to the fusion partner

TABLE 6.2 Some protein fusion systems used to facilitate the purification of foreign proteins in *E. coli* and other host organisms

Fusion partner	Size	Ligand	Elution conditions
ZZ	14 kDa	Immunoglobulin G	Low pH
Histidine tail	6–10 amino acids	Ni^{2+}	Imidazole
Strep tag	10 amino acids	Streptavidin	Iminobiotin
Pinpoint	13 kDa	Streptavidin	Biotin
Maltose-binding protein	40 kDa	Amylose	Maltose
GST	26 kDa	Glutathione	Reduced glutathione
Flag	8 amino acids	Specific monoclonal antibody (MAb)	EDTA or low pH
Poly-arginine	5–6 amino acids	SP-Sephadex	High salt at pH >8.0
c-myc	11 amino acids	Specific MAb	Low pH
S tag	15 amino acids	S fragment of RNase A	Low pH
Calmodulin-binding peptide	26 amino acids	Calmodulin	EGTA and high salt
Cellulose-binding domain	4–20 kDa	Cellulose	Urea or guanidine hydrochloride
Chitin-binding domain	51 amino acids	Chitin	SDS or guanidine hydrochloride
SBP tag	38 amino acids	Streptavidin	Biotin

Affinity and solubility tags for recombinant proteins

Tag	Size (aa)	Comments
His-tag	5–15	Purification under native or denaturing conditions
FLAG	8	Calcium-dependent, mAb-based purification
Streptag II	8	Modified streptavidin, elution with biotin analog
HA-tag	9	Influenza virus hemagglutinin tag, Ab-based purification
Softag1, Softag 3	13, 8	Recognized by polyol-responsive mAb
c-myc	10	mAb-based purification
T7-tag	11–16	mAb-based purification
S-tag	15	S-protein resin affinity purification
Elastin-like peptides	18–320	Protein aggregation by temperature shift, intein used to remove tag
Chitin-binding domain	52	Binds only insoluble chitin (see intein, Table 3)
Thioredoxin	109	Affinity purification with modified resin
Xylanase 10A	163	Cellulose based capture, elution with glucose
Glutathione *S*-transferase	201	Glutathione or GST-Ab affinity
Maltose binding protein	396	Amylose affinity purification
NusA	495	Increased solubility in *E. coli*. Affinity tag needed for purification

[a] Only a few relevant references are included.

Protein Secretion

- Requires a bacterial signal peptide sequence
- In gram negative bacteria like *E. coli*, the protein is secreted to the periplasmic space, as opposed to the media in gram positive bacteria
- Secretion affords some degree of protection against bacterial proteases

TABLE 6.8 Yields of several secreted recombinant proteins produced in different bacteria

Protein	Yield (per liter)	Host bacterium
Hirudin	>3 g	*E. coli*
Human antibody fragment	1–2 g	*E. coli*
Human insulin-like growth factor	8.5 g	*E. coli*
Monoclonal antibody 5T4	700 mg	*E. coli*
Humanized anti-CD18 $F(ab')_2$	2.5 g	*E. coli*
Human epidermal growth factor	325 mg	*E. coli*
Alkaline phosphatase	5.2 g	*E. coli*
Staphylokinase	340 mg	*B. subtilis*
Human proinsulin	1 g	*B. subtilis*
Human calcitonin precursor	2 g	*Staphylococcus carnosus*
Organophosphohydrolase	1.2 g	*Ralstonia eutropha*
Human CD4 receptor	200 mg	*Streptomyces lividans*
Human insulin	100 mg	*Streptomyces lividans*

Advantages and disadvantages for using tags in fusion proteins

Plus factors:

(1) improve protein yield

(2) prevent proteolysis

(3) facilitate protein refolding

(4) protect the antigenicity of the fusion protein

(5) increase solubility

(6) increase the sensitivity of binding assays for tagged ScFv.

Minus factors:

(1) a change in protein conformation (solubility and activity)

(2) lower protein yields

(3) inhibition of enzyme activity

(4) alteration in biological activity

(5) undesired flexibility in structural studies

(6) cleavage/removing the fusion partner requires expensive protease