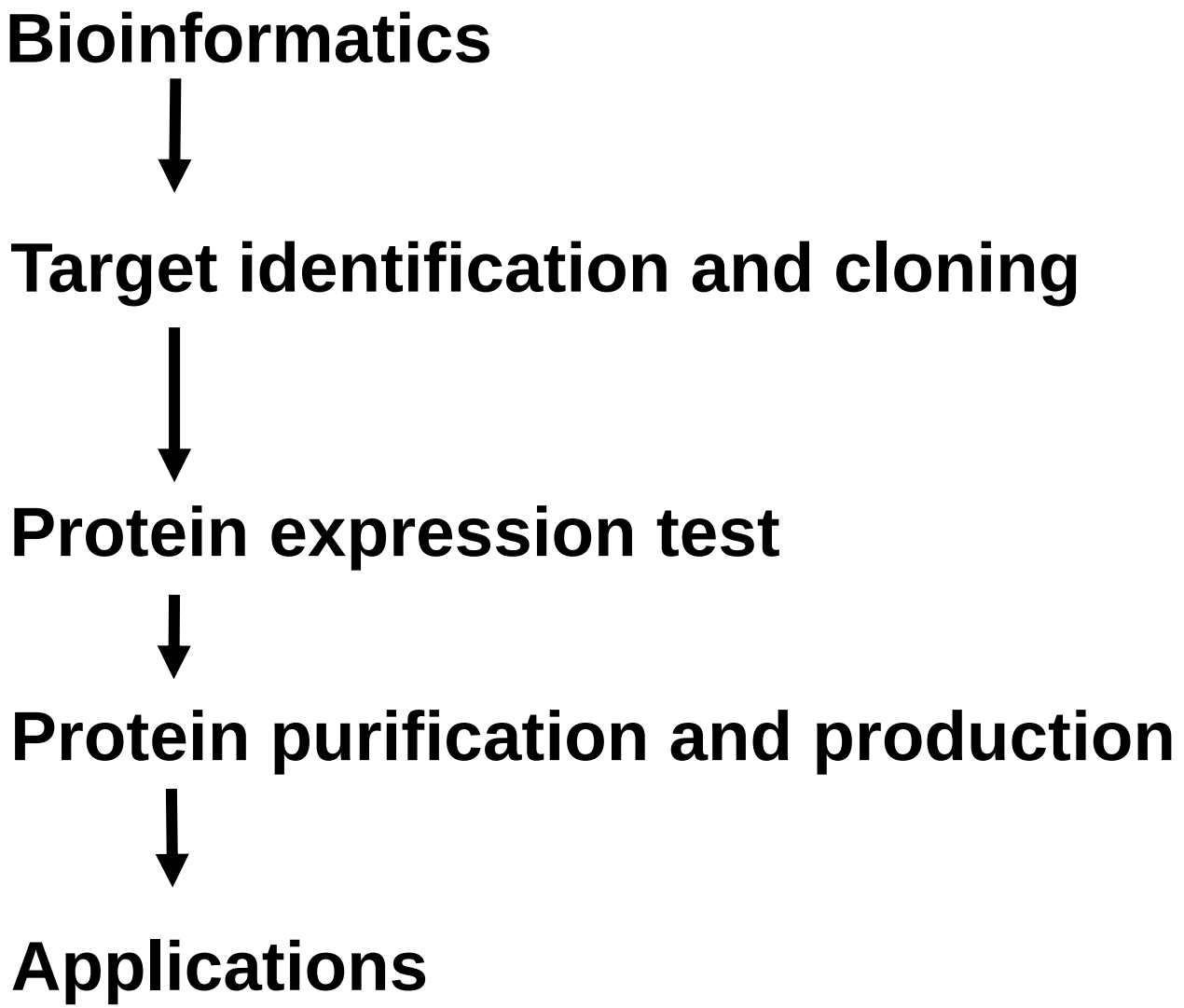
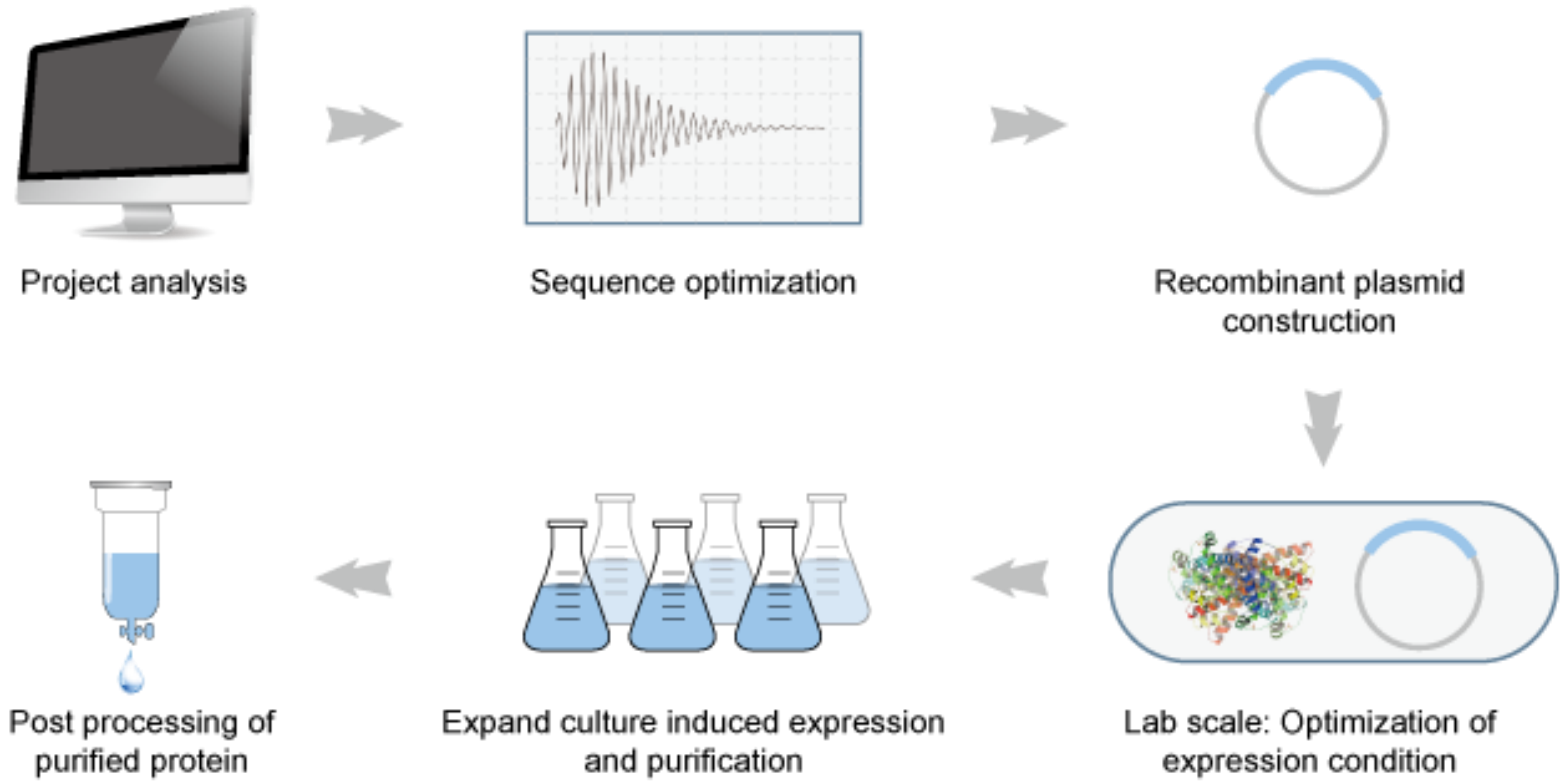
Protein expression and purification

Why to express proteins

- Determine the amino acid sequence (if no gene known)
- Make antibodies
- Identification by mass spectroscopy
- Structural analysis (x-ray crystallography & NMR spectroscopy)
- Enzyme function
- Interaction partners
- Biochemistry/biophysics (phosphorylation, regulation,...)
- Functional studies (e.g. localization)
- Pharmaceutical intervention

Principle in recombinant protein expression




Project analysis
Sequence optimization
Recombinant plasmid construction
Lab scale: Optimization of expression condition
Expand culture induced expression and purification
Post processing of purified protein

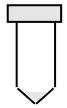
Steps in recombinant protein purification

1. Design expression plasmid, transform, select
2. Grow culture of positive clone, induce expression
3. Lyse cells
4. Centrifuge to isolate protein-containing fraction
5. Column Chromatography—collect fractions
6. Assess purity on SDS-PAGE

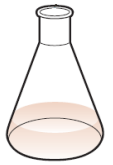
Protein expression and purification series workflow

Streak Cells

Overnight culture



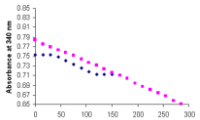
Subculture, monitor, and induce



Harvest and lyse cells

Purify

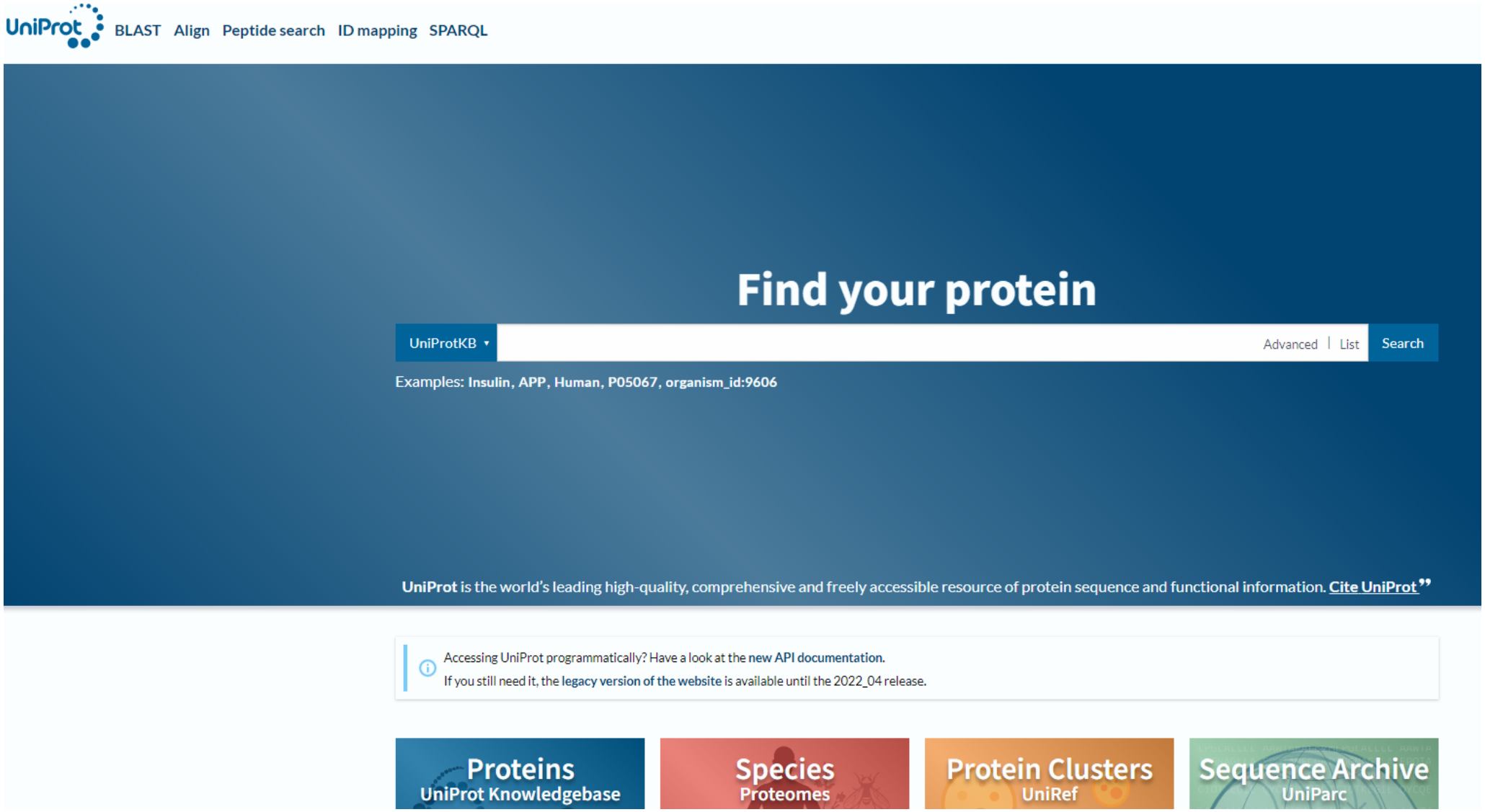
Analyze



Things to consider

- Which protein construct to express
- Expression host (bacterial, insect cell, yeast, or mammalian)
- Cell line for expression
- Promoter for induction of protein production
- Codon optimization (for mammalian proteins expressed in bacteria)
- Cloning method
- May require expression as a fusion protein
- May require co-expression with molecular chaperones
- Where to express: cytosol, periplasm, secretion, inclusion body
- Affinity tag(s) for purification & protease cleavage site to remove the tag
- Purification protocol & buffer to keep protein active

https://www.uniprot.org/



B0FWC7 · COX1_AEDAE

Cytochrome c oxidase subunit 1 · **Aedes aegypti (Yellowfever mosquito) (Culex aegypti)** · EC:7.1.1.9 · **Gene:** mt:CoI (COI, COX1) · 514 amino acids · Ev

Function

Component of the cytochrome c oxidase, the last enzyme in the mitochondrial electron transport chain which drives oxidative phosphorylation. The dehydrogenase (complex II, CII), ubiquinol-cytochrome c oxidoreductase (cytochrome b-c1 complex, complex III, CIII) and cytochrome c oxidase (cor succinate to molecular oxygen, creating an electrochemical gradient over the inner membrane that drives transmembrane transport and the ATP syr catalyzes the reduction of oxygen to water. Electrons originating from reduced cytochrome c in the intermembrane space (IMS) are transferred via t the active site in subunit 1, a binuclear center (BNC) formed by heme A3 and copper B (CU(B)). The BNC reduces molecular oxygen to 2 water molec mitochondrial matrix. By Similarity

Catalytic Activity

4 Fe(II)-[cytochrome c] + 8 H^+(in) + O2 = 4 Fe(III)-[cytochrome c] + 4 H^+(out) + 2 H2O By Similarity

This reaction proceeds in the forward direction. By Similarity

EC:7.1.1.9 (**UniProtKB** | **ENZYME** | **Rhea**)

Source: Rhea 11436

Fe(II)-[cytochrome *c*]	H^+ (in)	O_2	Fe(III)-[cytochrome *c*]

Engineering of protein construct

- Entire protein/mosaic protein
- If protein is of bacterial origin, express in bacteria
- If protein is of non-bacterial origin, because of post-translational modification in non-bacterial cells, may need to express in higher organisms:
  - **Bacteria → Yeast → insect cells → Mammalian cell lines**

- May need to express as a fusion protein or require codon-optimization

Expression hosts

Bacteria
+ Easy, great over-expression, low protease activity, no post-translational modifications
- Protein solubility, lack of post-translational modifications

Eukaryotes
+ Protein solubility, post-translational modifications
- Expensive, low yield, proteases, time consuming

Isolate protein from native source
+ Protein solubility, authenticity
- Expense/effort, yield, slaughter-houses

Bacteria, Yeast, Insect, Mammal

Expression hosts

	Bacteria	Yeast	Mammalian
Contamination risk	Low	Low	High
Cost of growth medium	Low	Low	High
Product titer (concentration)	High	High	Low
Folding	Sometimes	Probably	Yes
Glycosylation	No	Yes, but different pattern	Full
Relative ease to grow	Easy	Easy	Difficult
Relative ease of recovery	Difficult	Easy	Easy
Deposition of product	Intracellular	Intracellular or extracellular	Extracellular
Product	Intracellular	Often secreted into media	Secreted

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: minimize turnover of gene products (e.g. preventing proteolysis *in vivo* in *E. coli*).

Bacterial DE3 cells

Hosts for protein expression

DE3

1) Host-encoded lac Repressor represses host RNAPol transcription of T7 RNAPol from lac promoter.
2) IPTG induction knocks lac Repressor off and allow host RNAPol to transcribe T7RNAPol.
3) T7 RNAPol transcribe gene from T7 promoter on plasmid.
4) Lon/OmpT protease deficient.

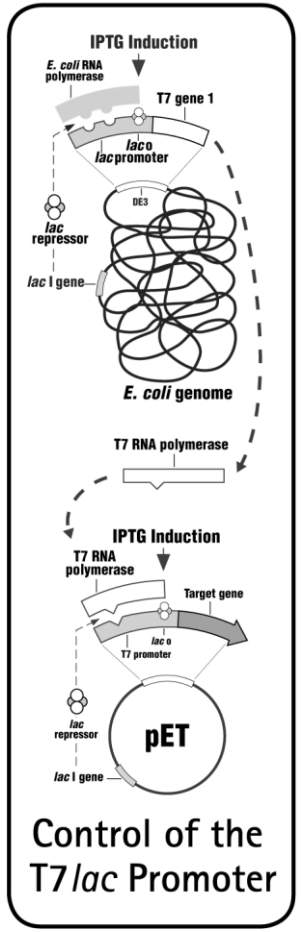
DE3_pLysS

1) DE3 strains have leaky expression, which leads to problems if expressed protein is toxic.
2) Plasmid-encoded T7 lysozyme inhibits T7 RNAPol and decrease leaky expression.

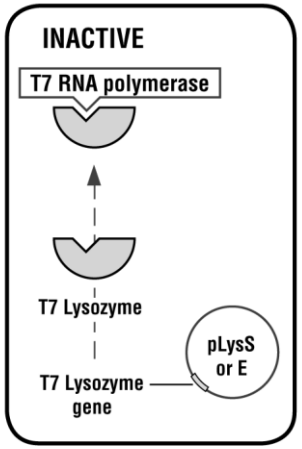
Host for cloning

1) These hosts lack T7 RNAPol , and so are suitable for plasmid amplification and not protein expression.

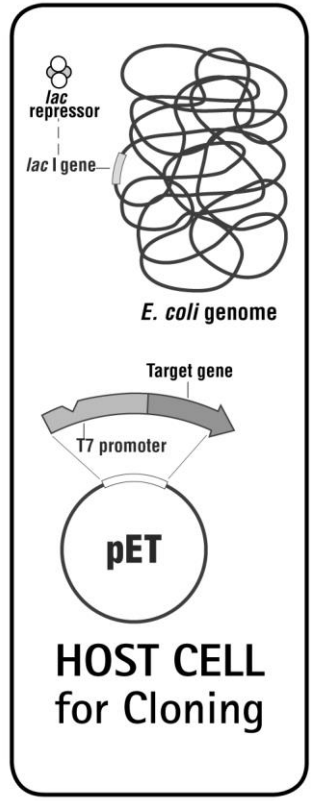
a)DE3



b)DE3_pLysS



c)Host for cloning



Isopropyl β-D-1-thiogalactopyranoside (IPTG) = allolactose mimic

Codon usage

Organism	AGG arginine	AGA arginine	CUA leucine	AUA isoleucine	CCC proline
Escherichia coli	1.2	2.1	3.9	4.4	5.5
Homo sapiens	11.4	11.5	6.5	6.9	**20.0**
Drosophila melanogaster	6.4	5.1	8.2	9.2	**18.0**
Caenorhabditis elegans	4.0	**15.4**	8.0	9.7	4.5
Saccharomyces cerevisiae	9.3	**21.3**	13.4	**17.8**	6.8
Plasmodium falciparum	4.1	**20.2**	**15.2**	**33.2**	8.5
Clostridium pasteurianum	2.4	**29.4**	6.2	**50.0**	0.9
Pyrococcus horikoshii	**30.1**	**20.1**	**18.2**	**44.5**	10.2
Thermus aquaticus	14.3	1.3	3.6	1.4	**38.8**
Arabidopsis thaliana	10.9	**18.8**	10.0	12.7	5.3

Table 1

Codon Usage in Various Organisms

Codon frequencies are expressed as codons used per 1000 codons encountered. The arginine codons AGG and AGA are recognized by the same tRNA and should therefore be combined. Codon frequencies of more than 15 codons/1000 codons are shown in bold to help identify a codon bias that may cause problems for high-level expression in *E. coli*. These frequencies are updated regularly. A complete compilation of codon usage of the sequences in the GenBank database can be found at www.kazusa.or.jp/codon/.

CCA(P) 0.84%	CGA(R) 0.35%	CAA(Q) 1.54%	CTA(L) 0.39%
CCG(P) 2.34%	CGG(R) 0.54%	CAG(Q) 2.90%	CTG(L) 5.33%
CCT(P) 0.69%	CGT(R) 2.10%	CAT(H) 1.28%	CTT(L) 1.10%
CCC(P) 0.55%	CGC(R) 2.22%	CAC(H) 0.96%	CTC(L) 1.12%
GCA(A) 2.03%	GGA(G) 0.79%	GAA(E) 3.97%	GTA(V) 1.09%
GCG(A) 3.40%	GGG(G) 1.10%	GAG(E) 1.79%	GTG(V) 2.63%
GCT(A) 1.53%	GGT(G) 2.48%	GAT(D) 3.21%	GTT(V) 1.83%
GCC(A) 2.57%	GGC(G) 2.97%	GAC(D) 1.91%	GTC(V) 1.53%
ACA(T) 0.69%	AGA(R) 0.20%	AAA(K) 3.36%	ATA(I) 0.42%
ACG(T) 1.44%	AGG(R) 0.11%	AAG(K) 1.03%	ATG(M) 2.79%
ACT(T) 0.89%	AGT(S) 0.86%	AAT(N) 1.74%	ATT(I) 3.05%
ACC(T) 2.34%	AGC(S) 1.60%	AAC(N) 2.14%	ATC(I) 2.52%
TCA(S) 0.70%	TGA(*) 0.09%	TAA(*) 0.20%	TTA(L) 1.38%
TCG(S) 0.89%	TGG(W) 1.54%	TAG(*) 0.02%	TTG(L) 1.37%
TCT(S) 0.84%	TGT(C) 0.51%	TAT(Y) 1.61%	TTT(F) 2.23%
TCC(S) 0.86%	TGC(C) 0.64%	TAC(Y) 1.23%	TTC(F) 1.66%

Rare codons in *E. coli*

Amino acid	Codon	Fraction in all genes[6]	Fraction in Class II[7]
Arg	AGG	0.022	0.003
Arg	AGA	0.039	0.006
Arg	CGG	0.098	0.008
Arg	CGA	0.065	0.011
Arg	CGU	0.378	0.643
Arg	CGC	0.398	0.330
Gly	GGG	0.151	0.044
Gly	GGA	0.109	0.020
Gly	GGU	0.337	0.508
Gly	GGC	0.403	0.428
Ile	AUA	0.073	0.006
Ile	AUU	0.507	0.335
Ile	AUC	0.420	0.659
Leu	UUG	0.129	0.034
Leu	UUA	0.131	0.055
Leu	CUG	0.496	0.767
Leu	CUA	0.037	0.008
Leu	CUU	0.104	0.056
Leu	CUC	0.104	0.080
Pro	CCG	0.525	0.719
Pro	CCA	0.191	0.153
Pro	CCU	0.159	0.112
Pro	CCC	0.124	0.016

6. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) *Nucl. Acids Res.* **29**, 292.

7. Henaut, A. and Danchin, A. (1996) in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, Vol. 2, (Neidhardt, F., Curtiss III, R., Ingraham, J., Lin, E., Low, B., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M., and Umbarger, H., eds), pp. 2047–2066, American Society for Microbiology, Washington, DC.

Codon usage

Organism	AGG arginine	AGA arginine	CUA leucine	AUA isoleucine	CCC proline
Escherichia coli	1.2	2.1	3.9	4.4	5.5
Homo sapiens	11.4	11.5	6.5	6.9	**20.0**
Drosophila melanogaster	6.4	5.1	8.2	9.2	**18.0**
Caenorhabditis elegans	4.0	**15.4**	8.0	9.7	4.5
Saccharomyces cerevisiae	9.3	**21.3**	13.4	**17.8**	6.8
Plasmodium falciparium	4.1	**20.2**	**15.2**	**33.2**	8.5
Clostridium pasteurianum	2.4	**29.4**	6.2	**50.0**	0.9
Pyrococcus horikoshii	**30.1**	**20.1**	**18.2**	**44.5**	10.2
Thermus aquaticus	14.3	1.3	3.6	1.4	**38.8**
Arabidopsis thaliana	10.9	**18.8**	10.0	12.7	5.3

Table 1

Codon Usage in Various Organisms

Codon frequencies are expressed as codons used per 1000 codons encountered. The arginine codons AGG and AGA are recognized by the same tRNA and should therefore be combined. Codon frequencies of more than 15 codons/1000 codons are shown in bold to help identify a codon bias that may cause problems for high-level expression in *E. coli*. These frequencies are updated regularly. A complete compilation of codon usage of the sequences in the GenBank database can be found at www.kazusa.or.jp/codon/.

Codon usage regulates the speed of translation elongation, resulting in non-uniform ribosome decoding rates on mRNAs during translation.

Codon usage plays an important role in regulating protein folding and function in both prokaryotic and eukaryotic organisms.

Rare codons in *E. coli*

Amino acid	Codon	Fraction in all genes[6]	Fraction in Class II[7]
Arg	AGG	0.022	0.003
Arg	AGA	0.039	0.006
Arg	CGG	0.098	0.008
Arg	CGA	0.065	0.011
Arg	CGU	0.378	0.643
Arg	CGC	0.398	0.330
Gly	GGG	0.151	0.044
Gly	GGA	0.109	0.020
Gly	GGU	0.337	0.508
Gly	GGC	0.403	0.428
Ile	AUA	0.073	0.006
Ile	AUU	0.507	0.335
Ile	AUC	0.420	0.659
Leu	UUG	0.129	0.034
Leu	UUA	0.131	0.055
Leu	CUG	0.496	0.767
Leu	CUA	0.037	0.008
Leu	CUU	0.104	0.056
Leu	CUC	0.104	0.080
Pro	CCG	0.525	0.719
Pro	CCA	0.191	0.153
Pro	CCU	0.159	0.112
Pro	CCC	0.124	0.016

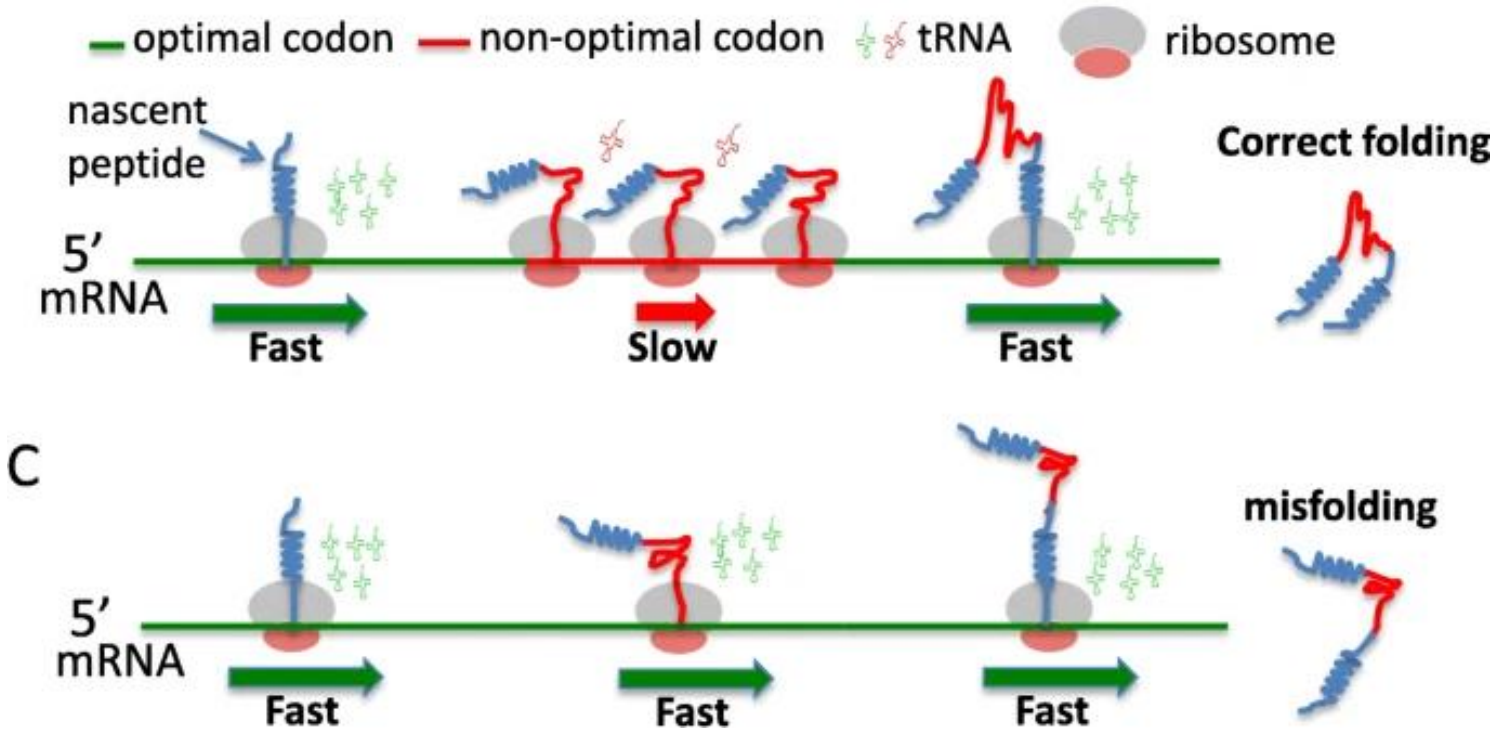
6. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) *Nucl. Acids Res.* 29, 292.
7. Henaut, A. and Danchin, A. (1996) in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, Vol. 2, (Neidhardt, F., Curtiss III, R., Ingraham, J., Lin, E., Low, B., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M., and Umbarger, H., eds), pp. 2047–2066, American Society for Microbiology, Washington, DC.

Codon usage

Rare codons cause ribosome stalling on an mRNA during translation, which, in eukaryotes, can result in premature translation termination mediated by termination factors.

Preferred codons speed up rate of translation elongation. Rare codons slow translation elongation.

Replacing rare codons with common ones can result in modest decreases of protein activity or solubility.



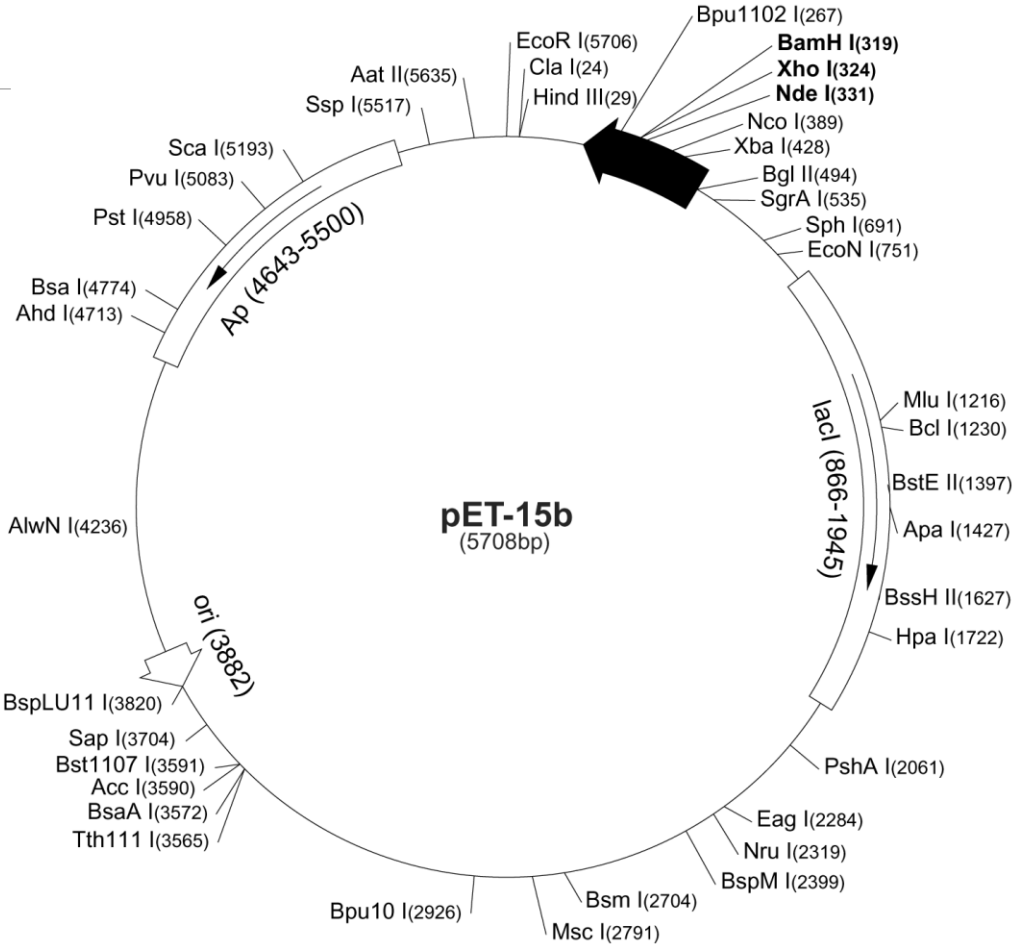
https://doi.org/10.1186/s12964-020-00642-6

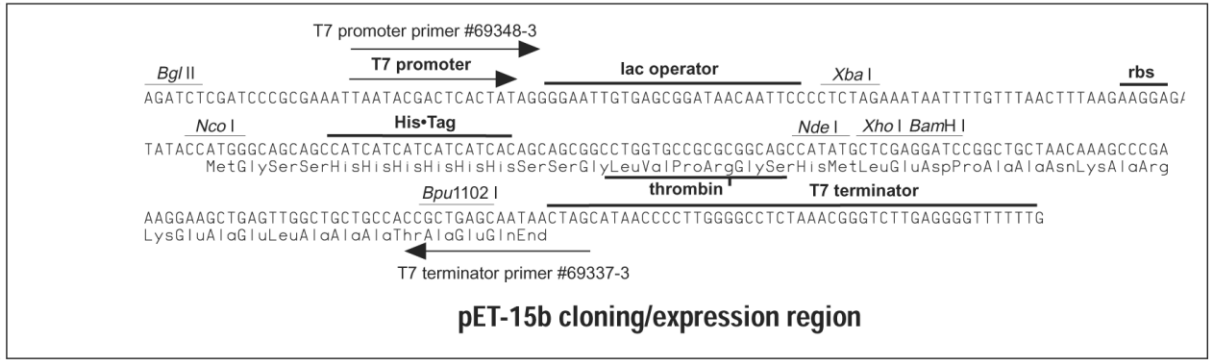
Expression vector

pET vector

- Ap = ampicillin resistance
- ori = ColE1/pBR322 origin of replication
- lacI = lac repressor; bind lacO, until IPTG induction
- T7P = T7 Polymerase promoter
- lacO = lac operator where lac repressor binds
- ➡ = multiple cloning site

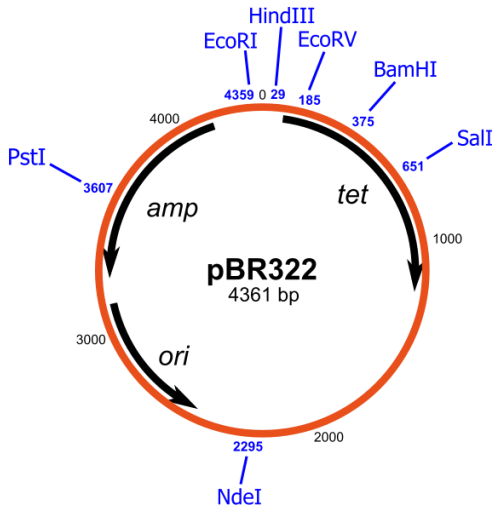
pET-15b sequence landmarks	
T7 promoter	453-469
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde* I - *Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500

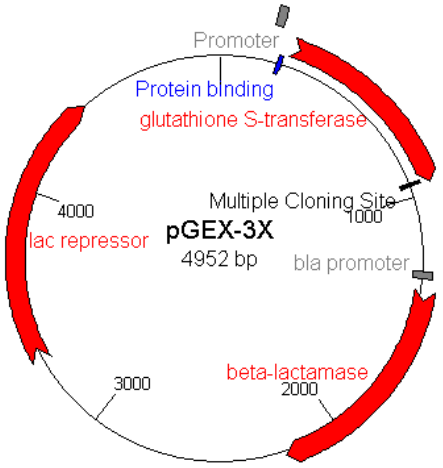




Cloning vector and expression vector

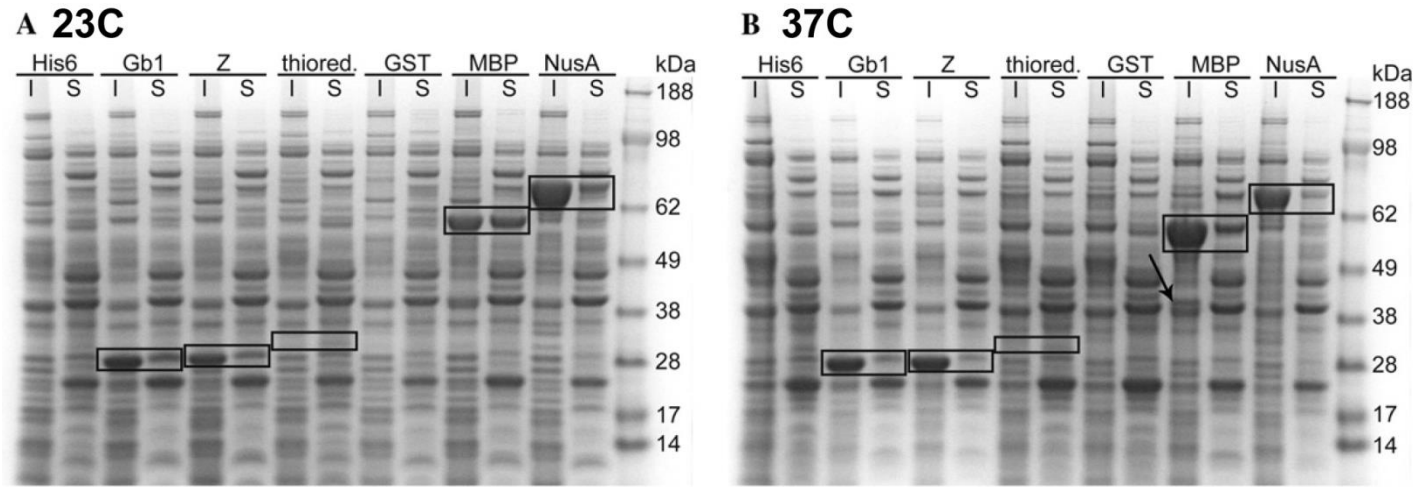
- Expression vectors are specially designed cloning vectors.
- Aid in the transcription and protein translation process of the desired DNA fragment.
- Contain all features of a typical cloning vector.
- But have promoters, enhancers, termination sequence, initiation sites, stop codon, etc., to help in protein synthesis.





Selection of expression vector & fusion partner

- Expression in insect and mammalian cells are expensive and time consuming.
- A feasible method is required for expression in bacterial cells as a first choice.
- Expression difficulty in bacterial cells may be overcome by expression as a fusion protein.
- **Expression at lower temperature improves solubility.**



Niiranen et al, 2007

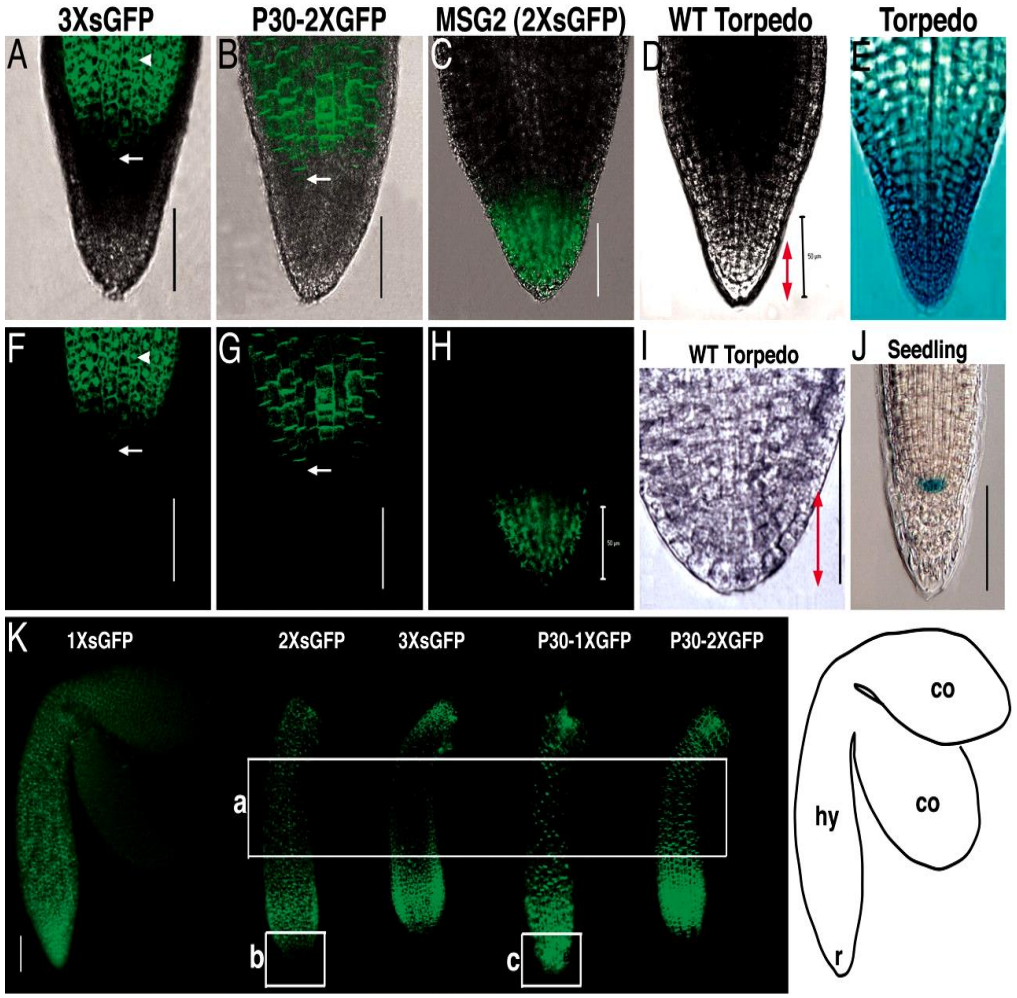
Architecture of reporter gene constructs

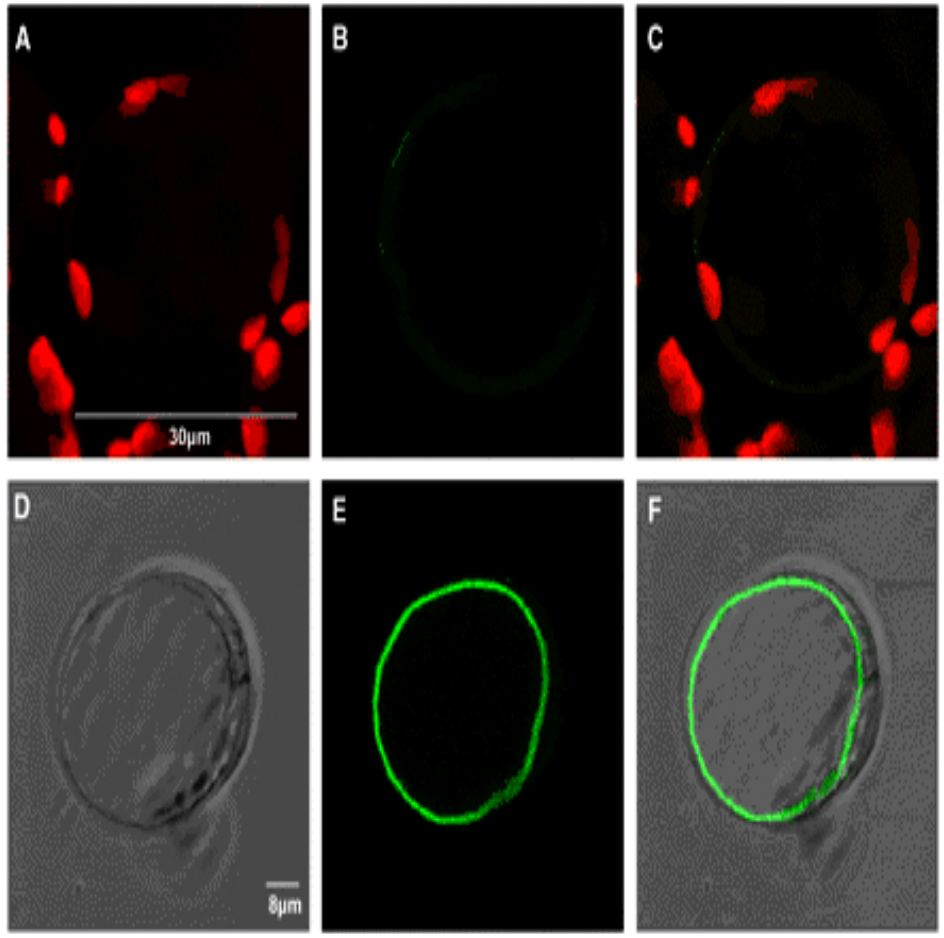
Translational reporter

Translational reporters

- In-frame **gene fusions** between GFP and a gene of interest.
- Includes the entire genomic locus of a gene (5’ upstream region, exons, introns, 3 UTR).
- GFP can be inserted at any point in the open reading frame, preferably at a site that does not disrupt protein function or topology.

Boulin *et al*., 2006


3XsGFP
P30-2XGFP
MSG2 (2XsGFP)
WT Torpedo
Torpedo
A
B
C
D
E
F
G
H
I
WT Torpedo
J
Seedling
K
1XsGFP
2XsGFP
3XsGFP
P30-1XGFP
P30-2XGFP
a
b
c
co
co
hy
r


A
B
C
30µm
D
E
F
8µm

Translational fusion

Assume the restriction site identified in the gene which you want to express is a *BamH*I site. Digest with *BamH*I to obtain:

GATCCXXXXXXXXXXX

GYYYYYYYYYYYYY

↓

Treat with Klenow fragment to fill in the unpaired bases to obtain:

GATCCXXXXXXXXXXX

CTAGYYYYYYYYYYYY

↓

Determine the proper reading frame of the gene. Assume the coding sequence of the filled-in fragment should read:

GA TCC XXX XXX XXX

↓

Determine which restriction endonuclease should be used to digest an expression vector pSKF301 in order to allow expression of the fusion protein. For this example, StuI is required to yield:

ccatg gat cat atg tta aca gat atc aag gGA TCC XXX XXX

pSKF301 (carrier sequence) Fusion gene

Where to express the recombinant proteins

- Direct expression (cytosol)

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

- 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.

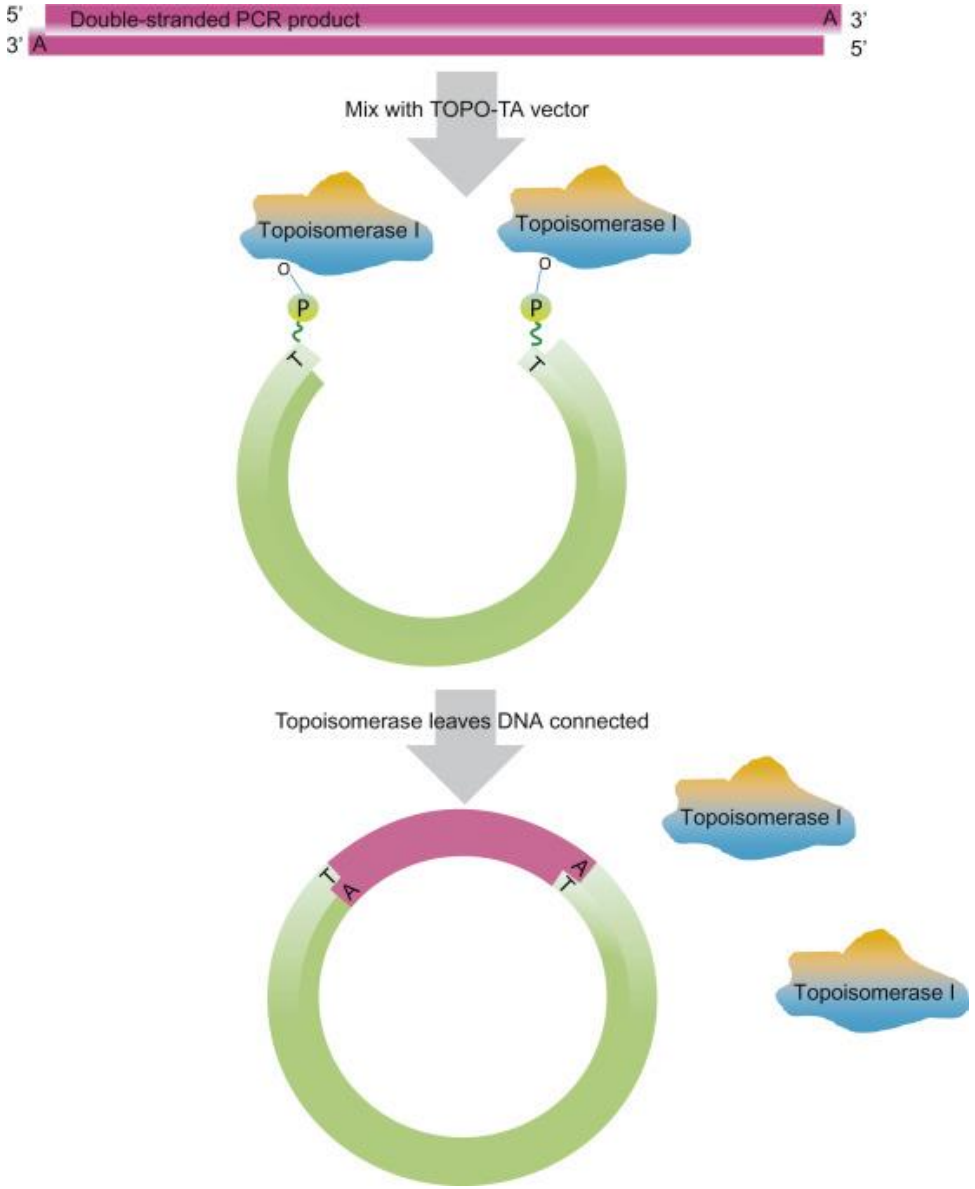
  Inclusion body: tiny aggregated proteins found freely suspended and floating within cytoplasm due to overexpression of heterologous proteins.

- Secretion (periplasm or medium)

  A fusion alternative when proteins are fused to peptides or proteins targeted for secretion.

Cloning methods

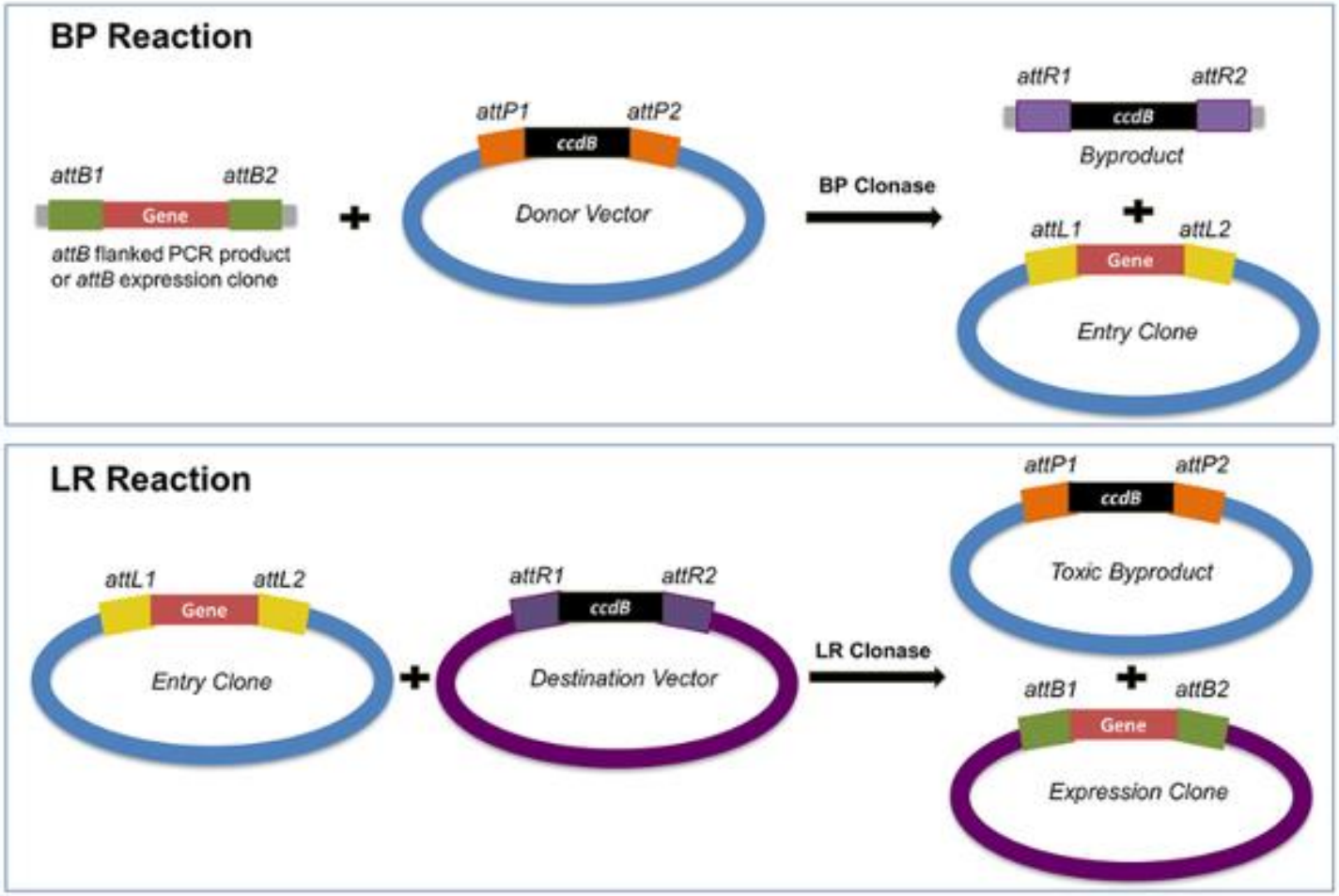
TOPO cloning



Clark *et al*., 2019

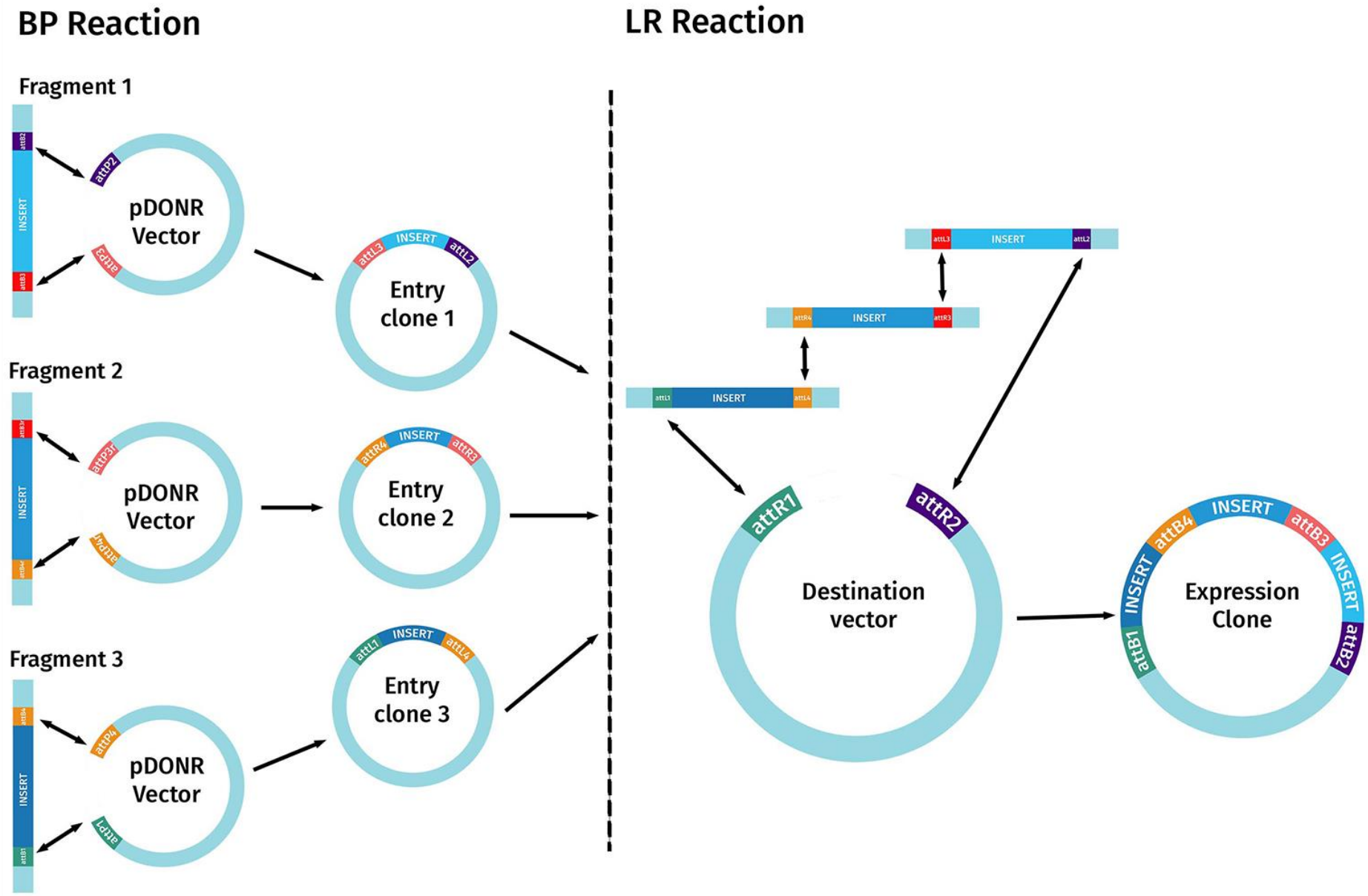
Cloning methods

Gateway cloning



Cloning methods

Gateway cloning



Transformation

- Addition of plasmid DNA to cells, followed by cold- and heat-shock allows plasmid to enter through the small holes in the cell wall.
- Amplify the number of cells in SOC media.
- Plate on LB agar (+ antibiotic) to select for transformed cells only (antibiotic resistance is conferred by gene encoded by on plasmid).

- Before starting, confirm that you can make a significant quantity of soluble protein.
- Small scale solubility experiments are very important and typically will involve varying inducer concentration, expression temperature, expression construct, etc.

Each protein is unique – must exploit differences

Particular affinities	**GST, 6xHis, antibodies**
Solubility	**$(NH_4)_2SO_4$, PEG precip.**
Charge	**ion exchange**
Hydrophobicity	**hydrophobic chromatography**
Size	**gel exclusion**
Iso-electric point	**iso-electric focusing**
Thermal stability	**alter temp.**

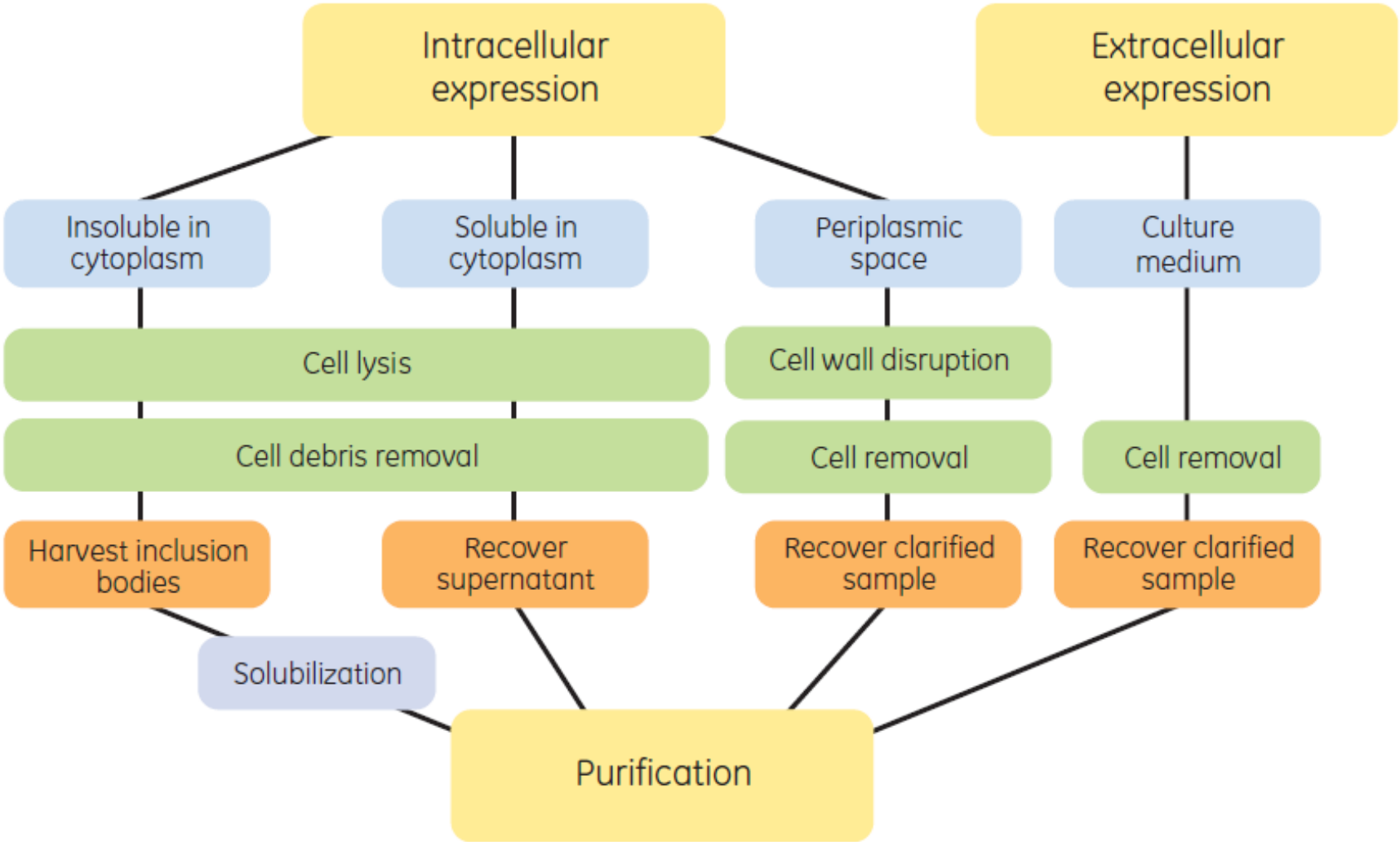


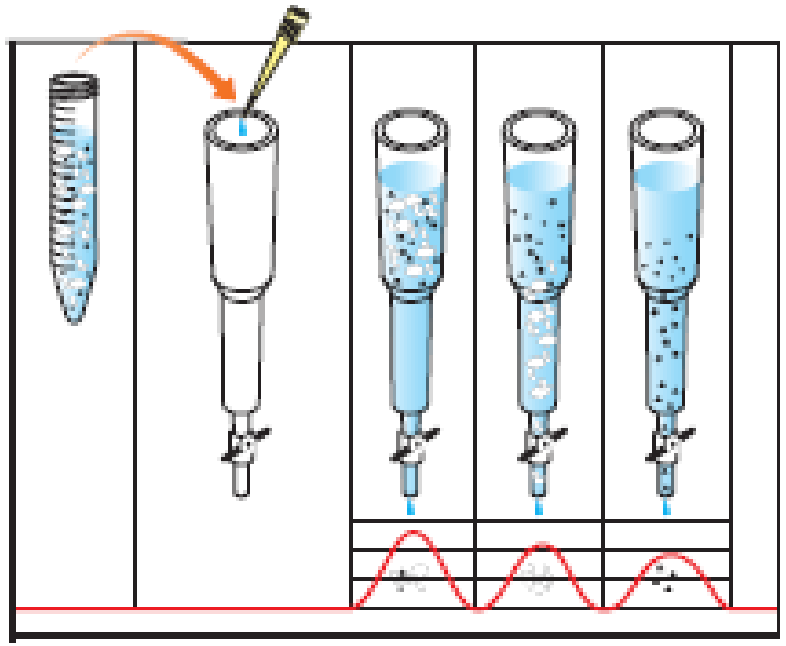
Fig 1.2. Overview of sample preparation.

Recovery

- Separation of protein from other molecules

Purification

- Separation of the protein of interest from other proteins



Cell lysis

Rupture cell wall / plasma membrane >> release contents (organelles, proteins…)

- Physical means
- Sonication
- Osmotic shock

Chromatography basics

- **Mobile phase** (solvent and the molecules to be separated)
- **Stationary phase** (through which the mobile phase travels)
  - paper (in paper chromatography)
  - glass, resin, or ceramic beads (in column chromatography)
- Molecules travel through the stationary phase at different rates because of their chemistry.

Performing the chromatographic separation

- **Gravity Chromatography**
  - Spin Column Chromatography

- **Chromatography Instrumentation**
  - Small scale
  - Bio-manufacturing scale (bioreactors)

Protein purification – column chromatography

- Protein mixture applied to column
- **Solvent** (buffer) applied to top, flowed through column
- Different proteins interact with matrix to different extents, flow at different rates
- Proteins collected separately in different **fractions.**

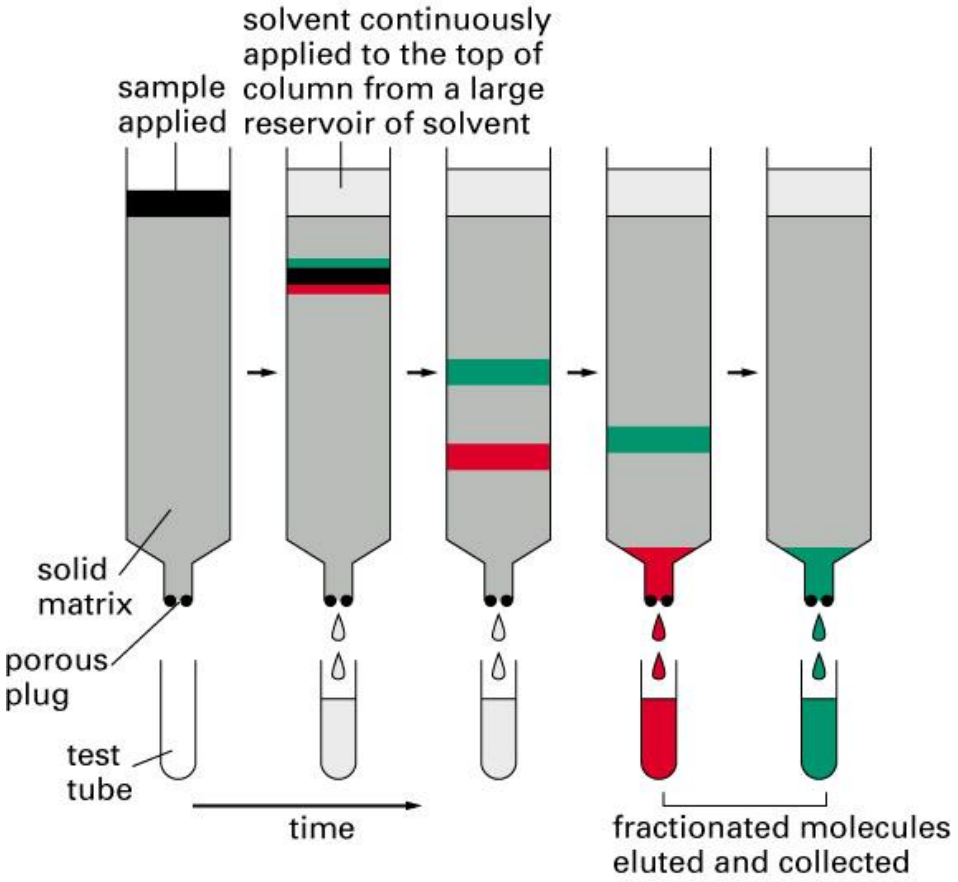
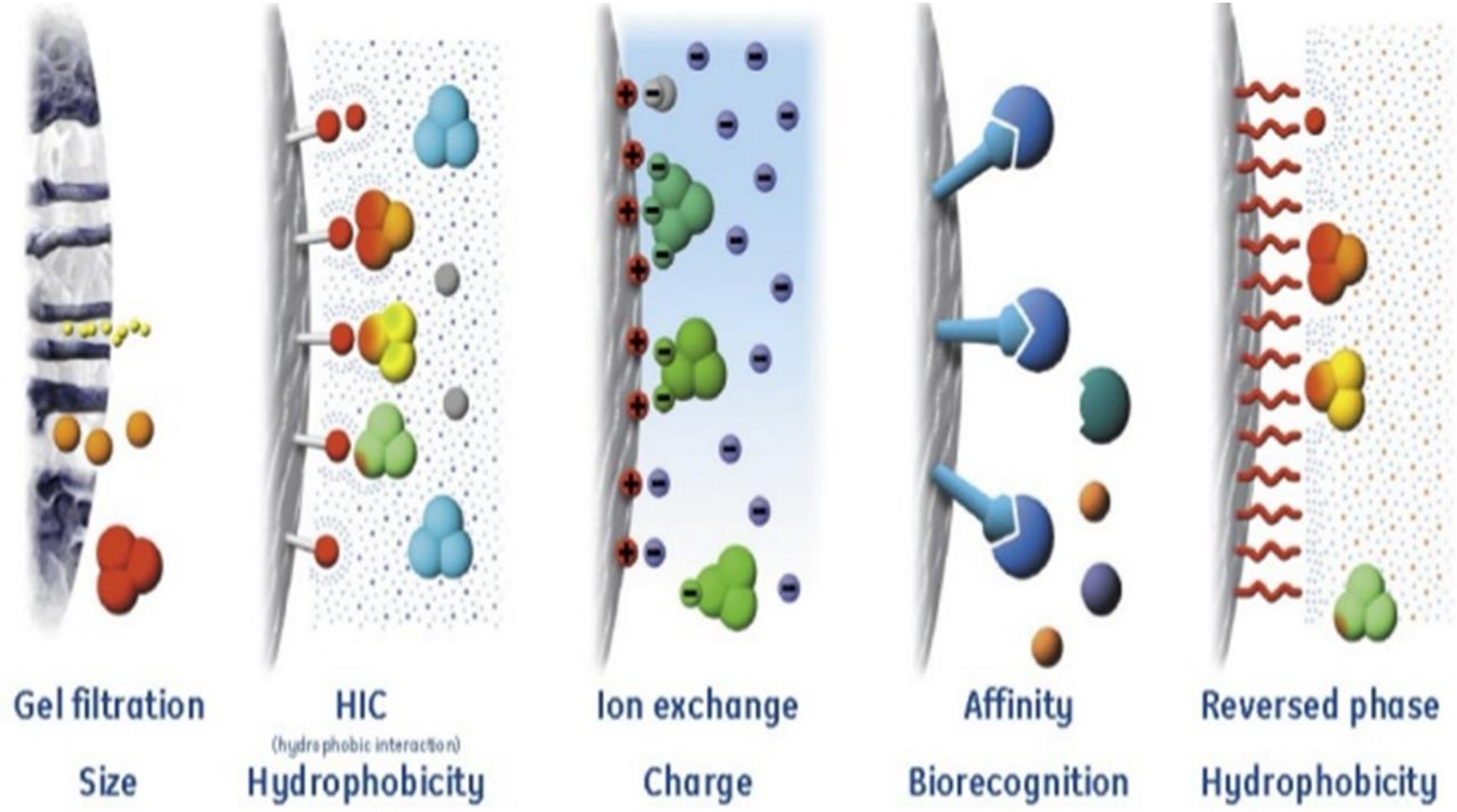


Figure 8–10. Molecular Biology of the Cell, 4th Edition.

Column chromatography

Molecules can be separated on the basis of:

- Size—gel filtration
- Charge—ion exchange
- Specific binding—affinity



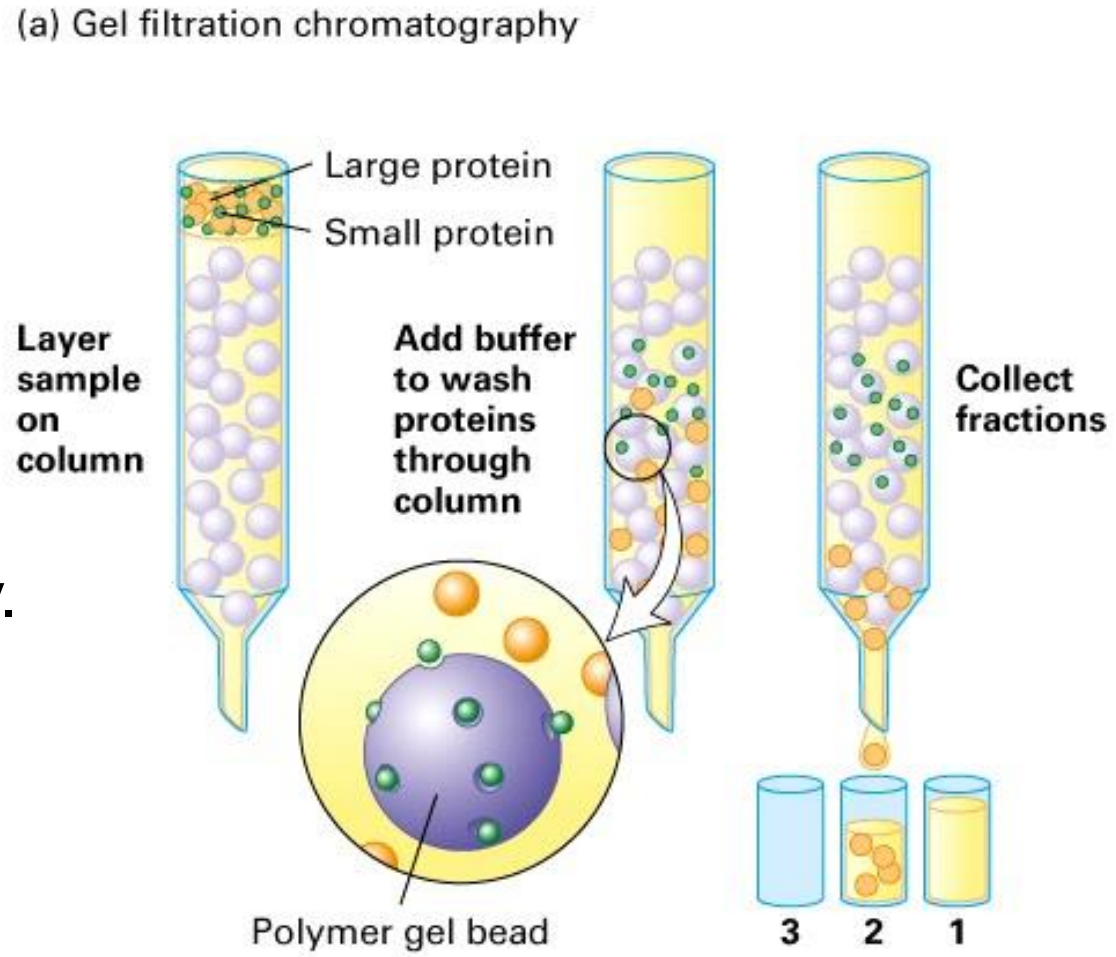
Gel filtration chromatography - separation by size

Beads have **different size pores**

As column flows:

• **Large proteins** excluded from pores, therefore flow rapidly.

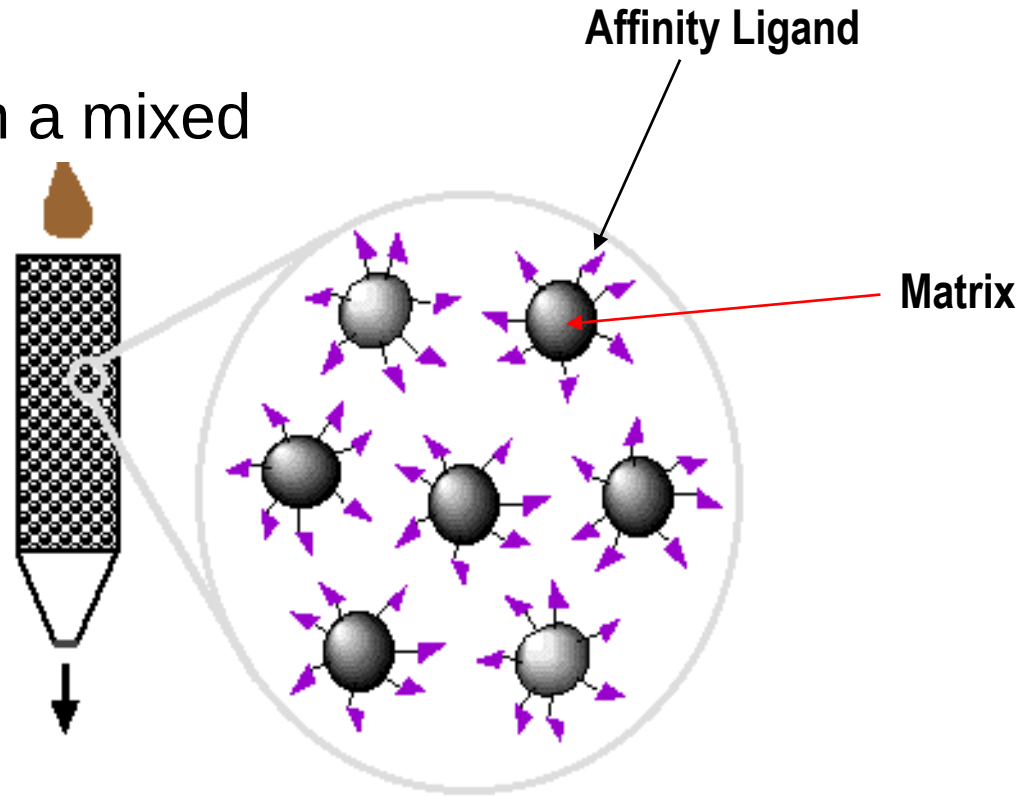
• **Small proteins** enter pores and flow slowly.



Affinity chromatography

- A technique enabling purification of a biomolecule with respect to biological **function** or individual **chemical structure**.

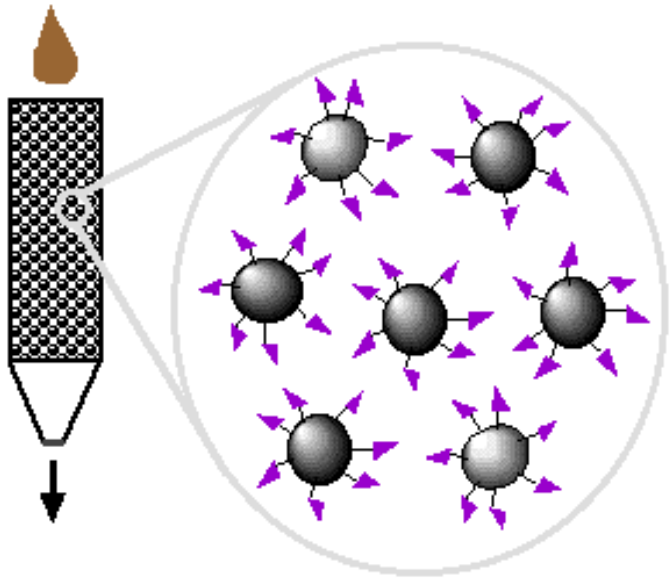
- Designed to purify a **particular molecule** from a mixed sample.



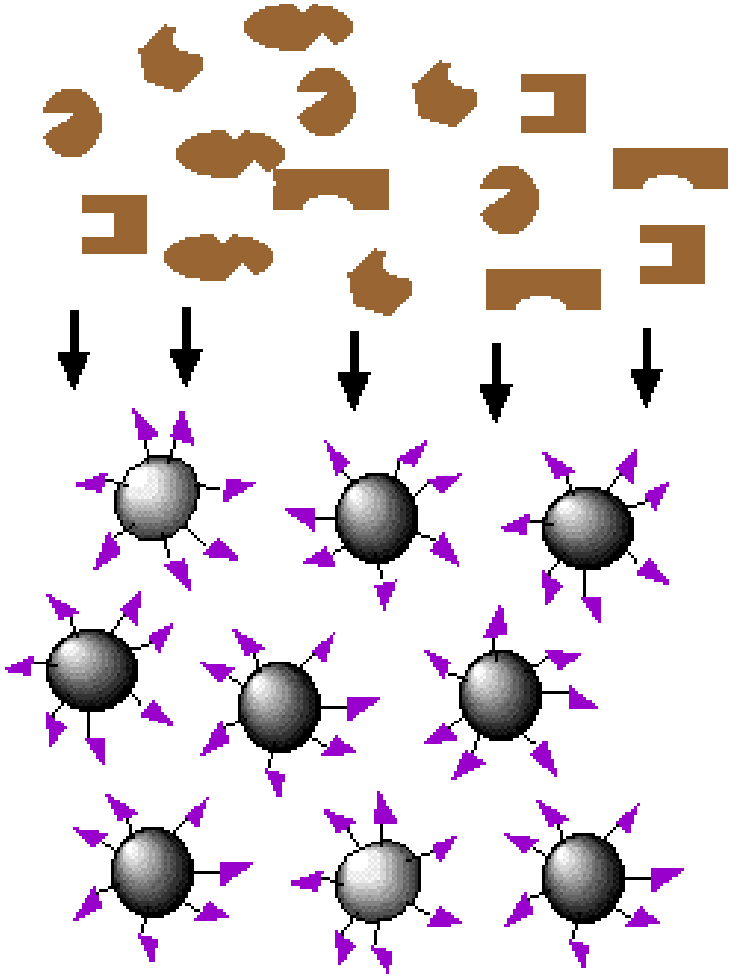
Examples of tags and ligands

- **His-tag**
- FLAG™ peptide
- *Strep*-tag
- **GST tag**
- Maltose binding protein fusion (MBP)
- Calmodulin binding protein fusion

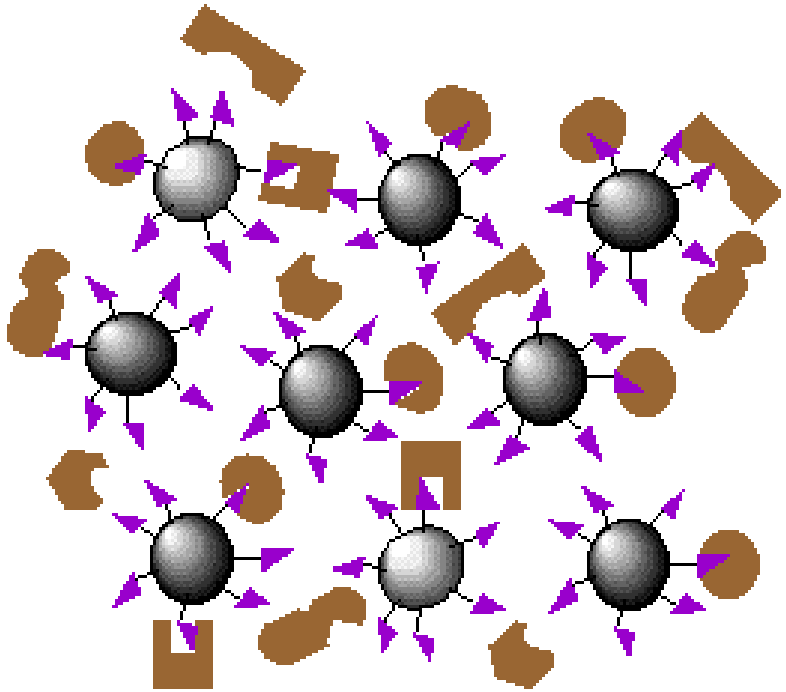
Step 1. Loading affinity column.



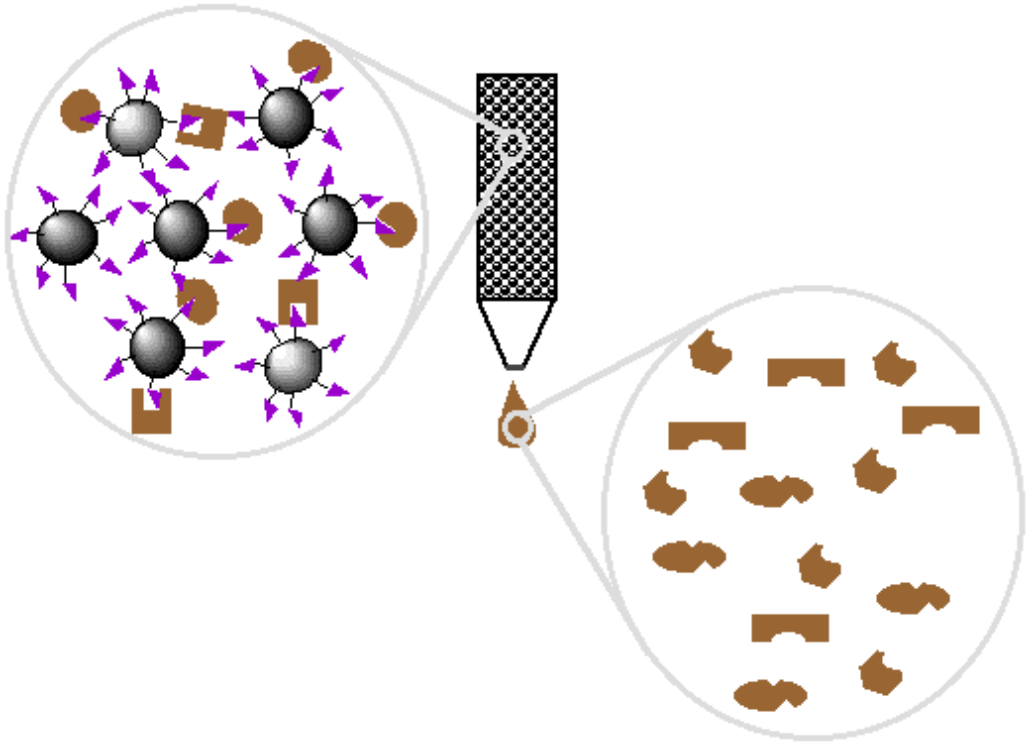
Step 2. Proteins sieve through matrix of affinity beads.



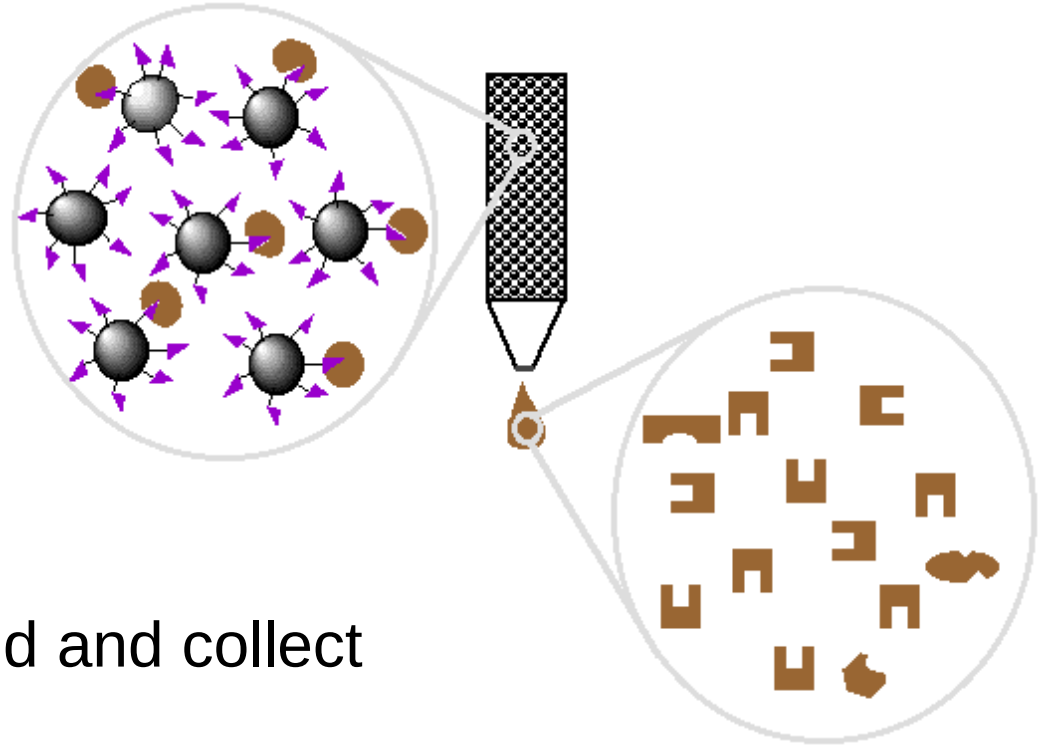
Step 3.
Proteins interact with affinity ligand with some binding **loosely** and others **tightly.**



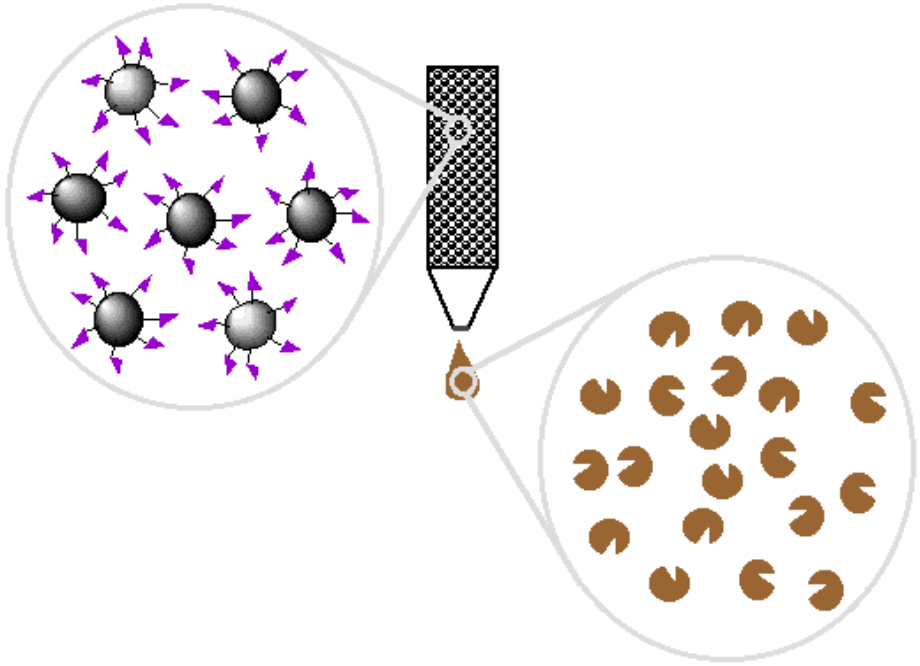
Step 4. Wash off proteins that do not bind.



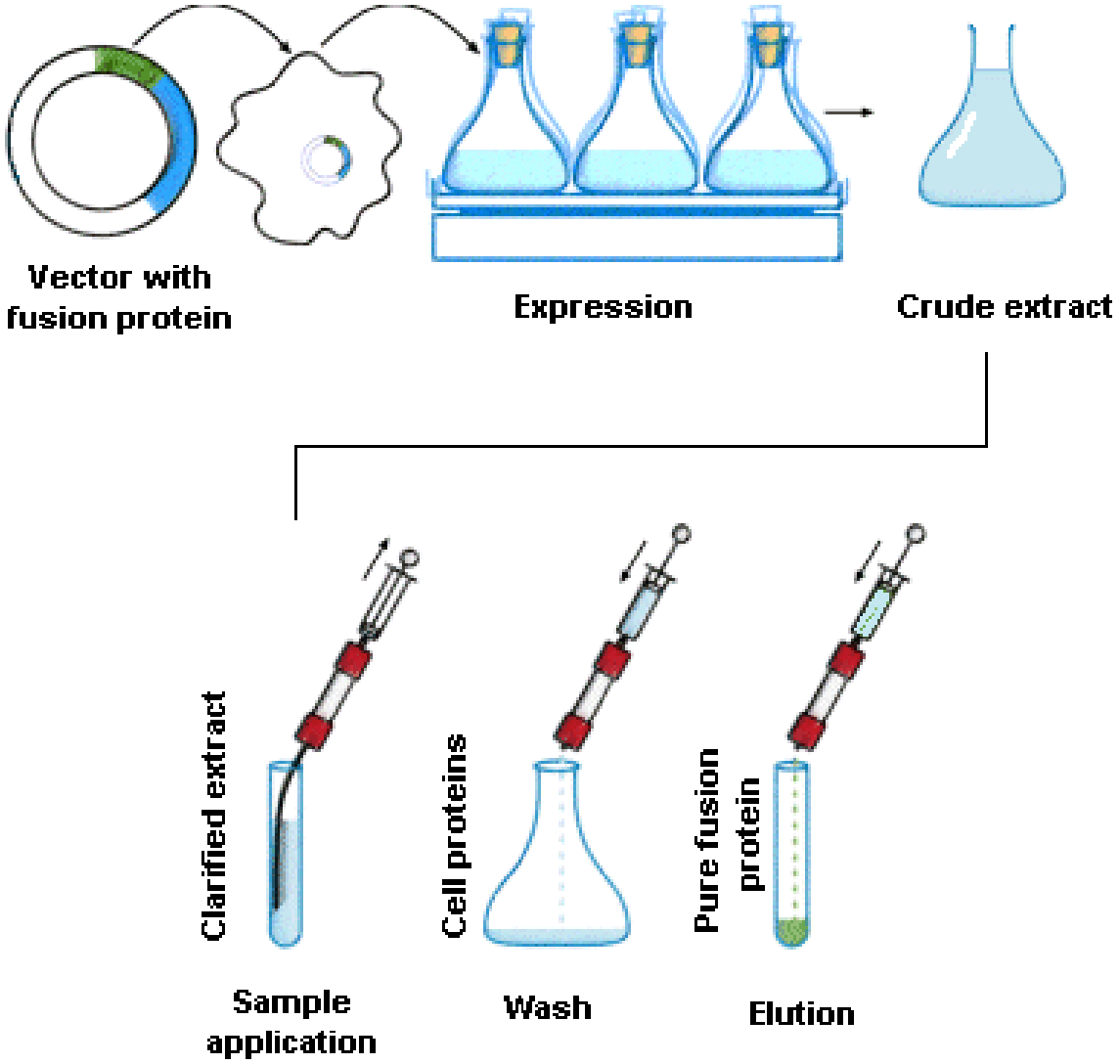
Step 5. Wash off proteins that bind loosely.



Step 6. Elute proteins that bind tightly to ligand and collect purified protein of interest.

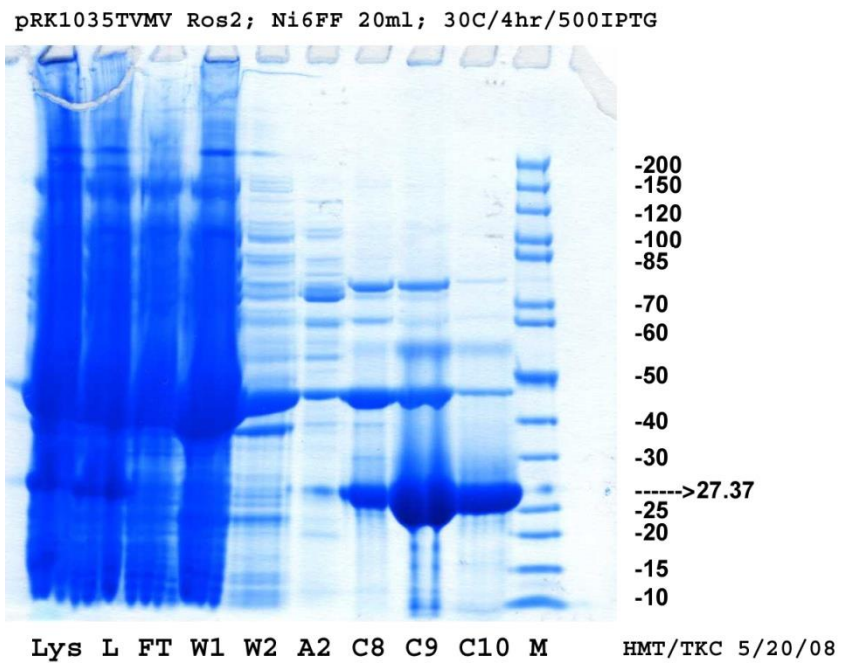


Affinity chromatography applied to recombinant proteins

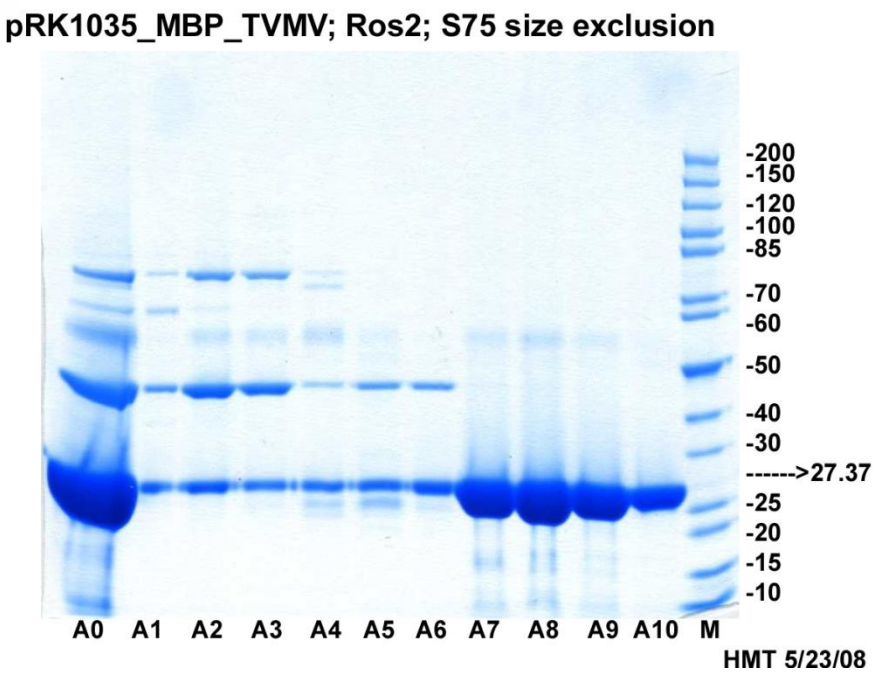


Different column, different results

1) Nickel column



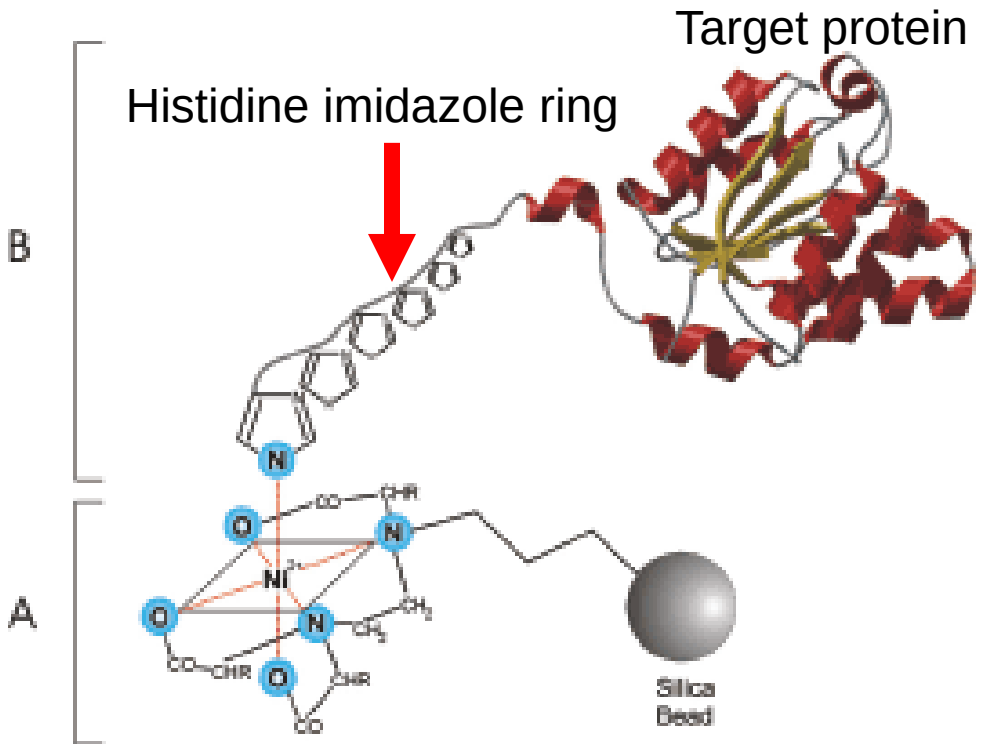
2) Superdex 75 26/60



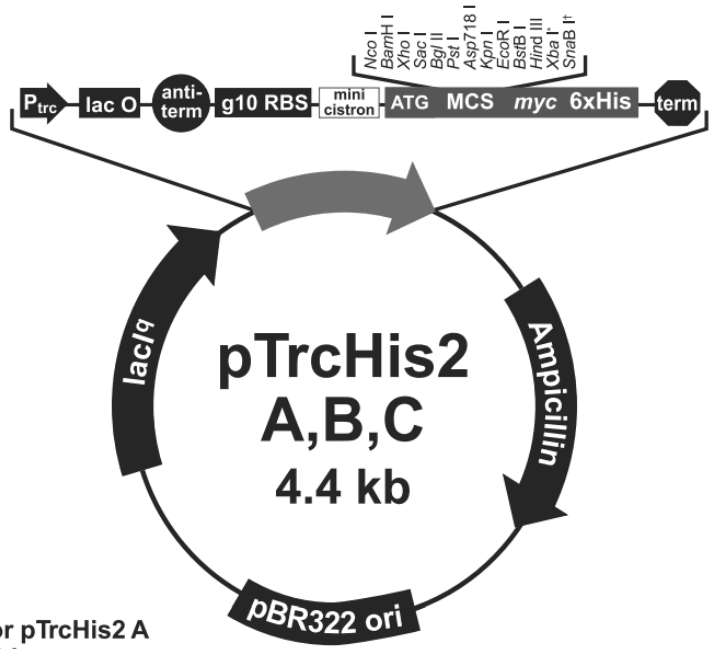
IMAC: Immobilized Metal Affinity Chromatography (His6 purification)

- Relies on the interaction between multiple electron donors on the affinity tag with a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}) chelated to a solid-phase support.

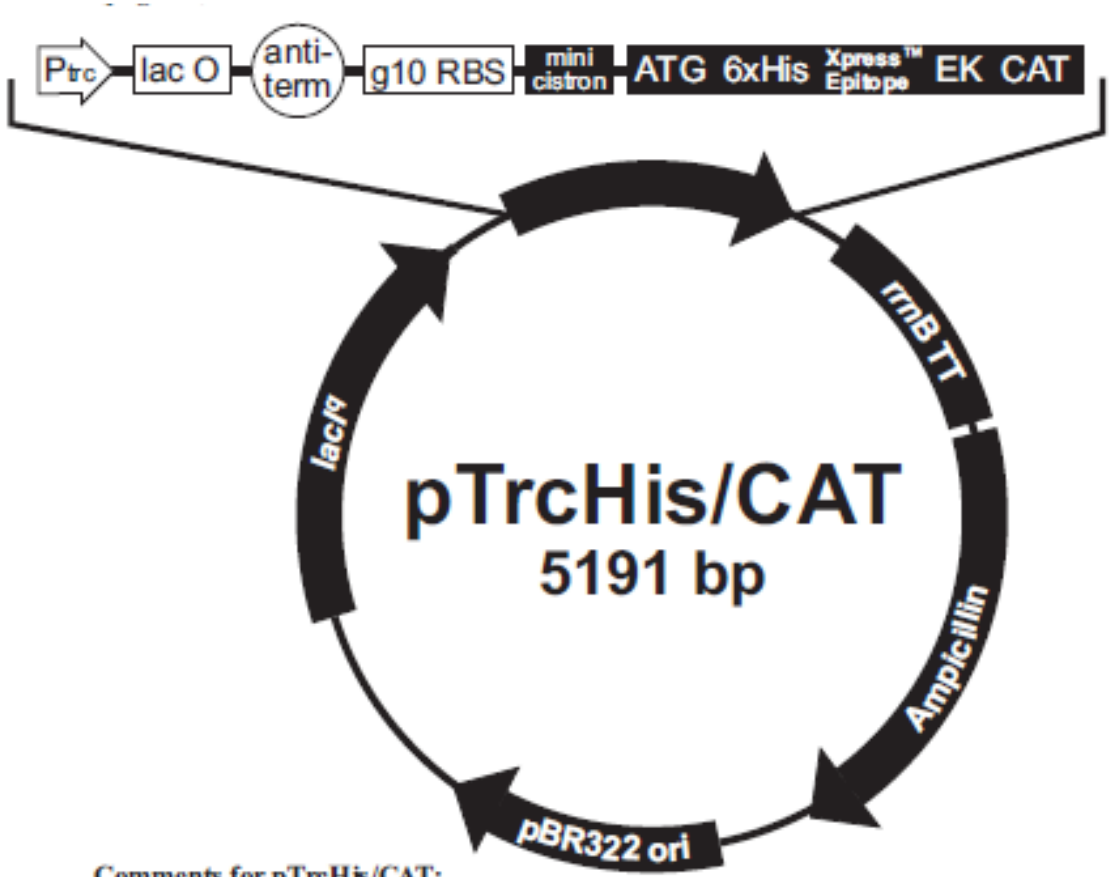
The affinity tag is typically polyhistidine, ranging 6–12 residues in length fused to the N- or C-terminus of the target, where 6-His is most common and the electron donor is the histidine imidazole ring.



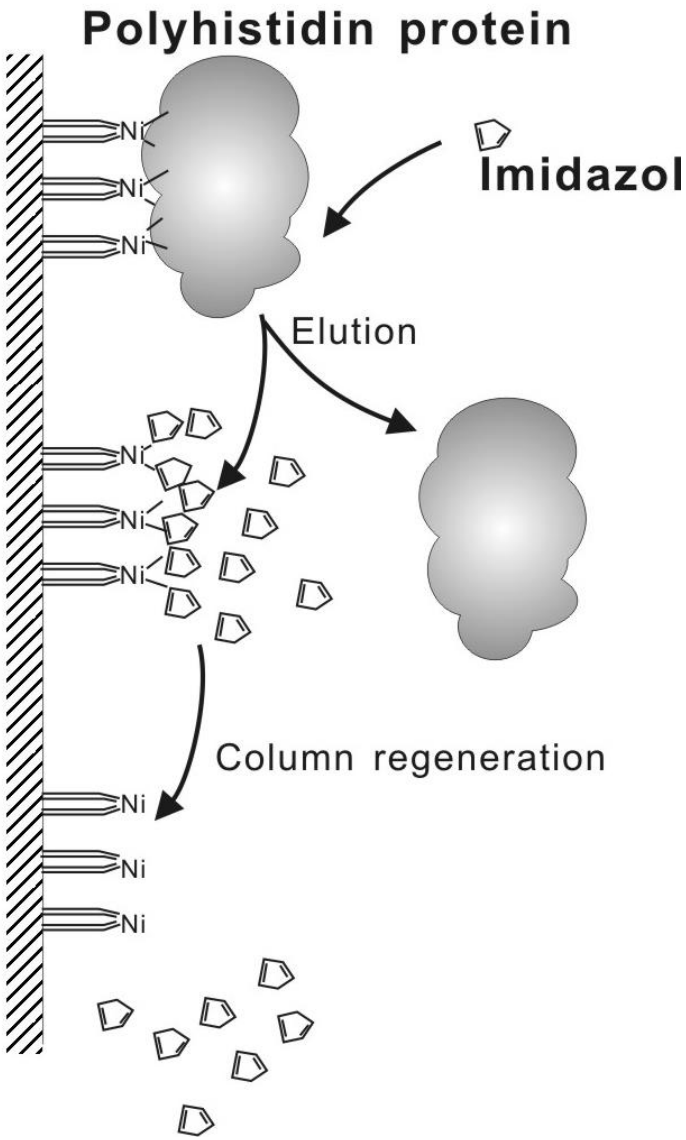
Expression vectors with His tag

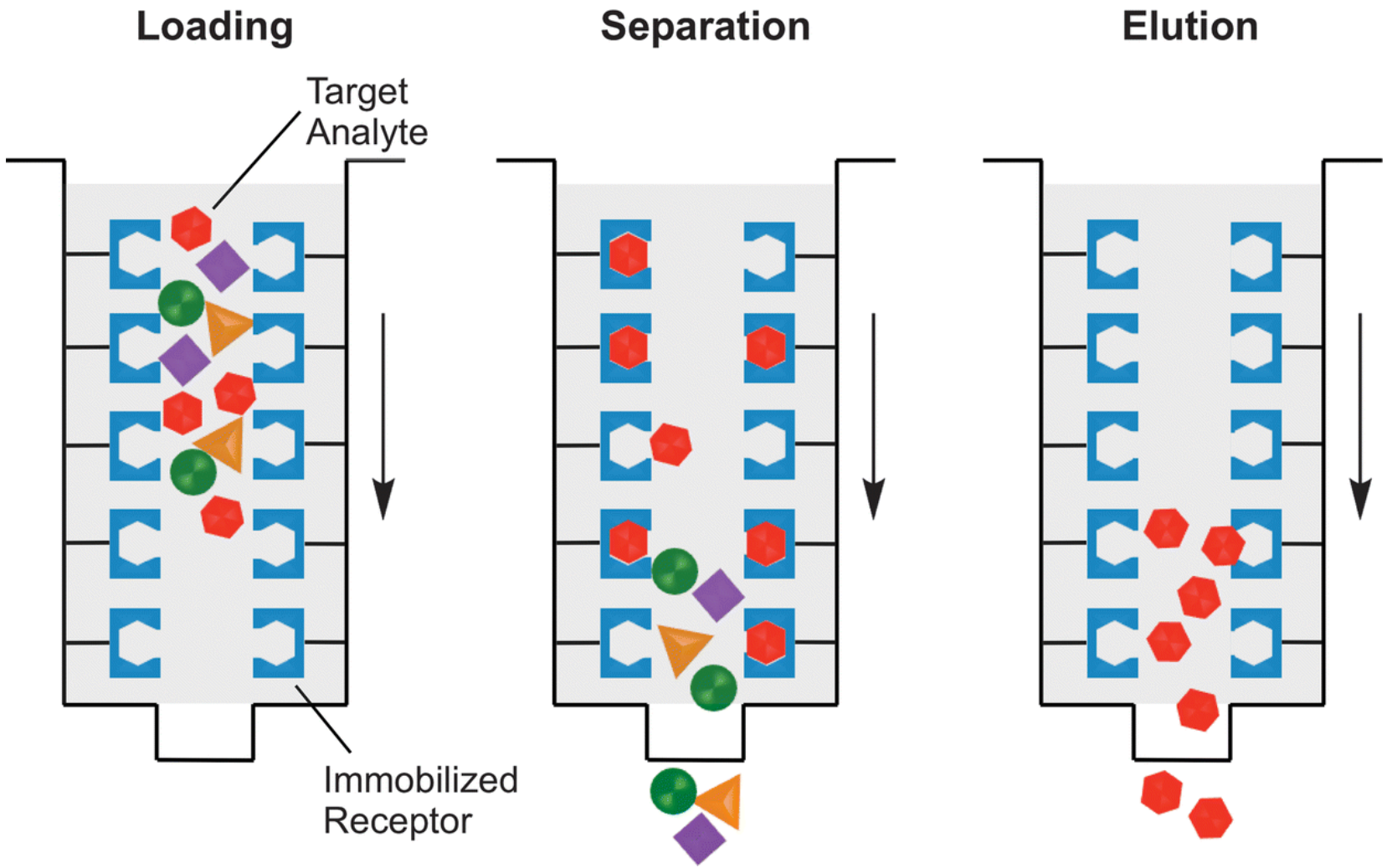


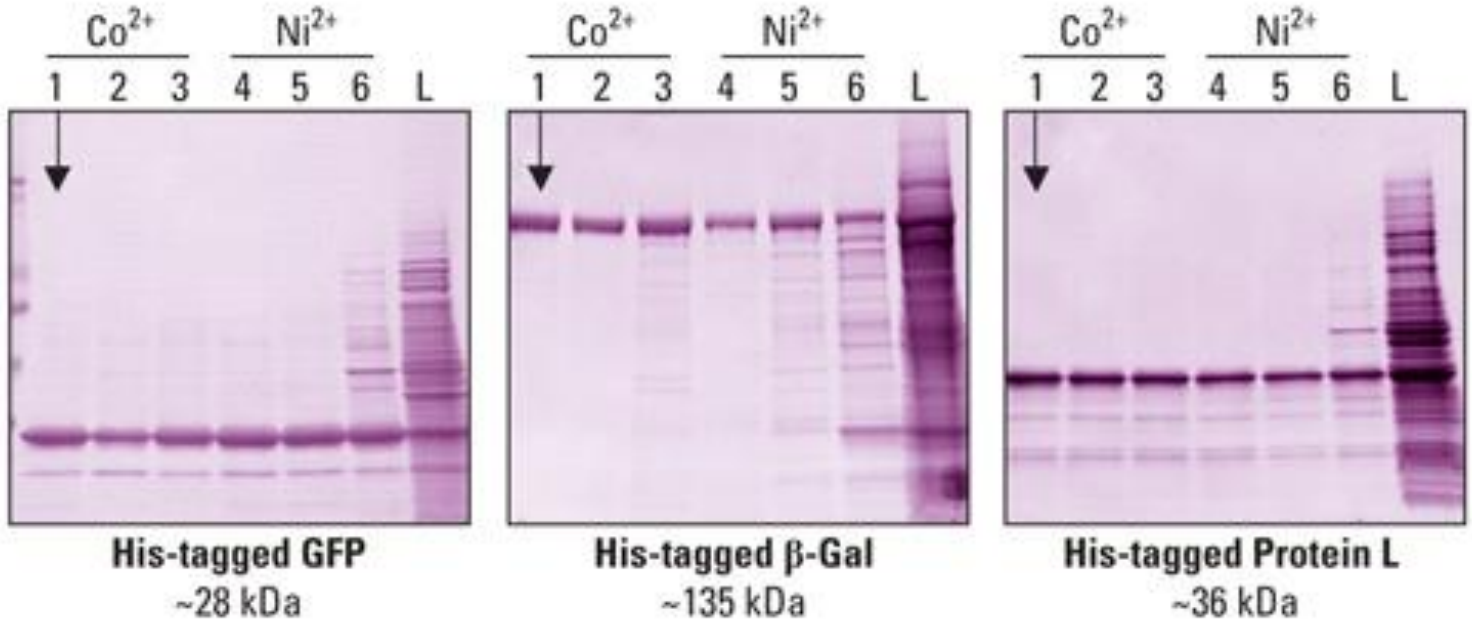
Comments for pTrcHis2 A
4406 nucleotides



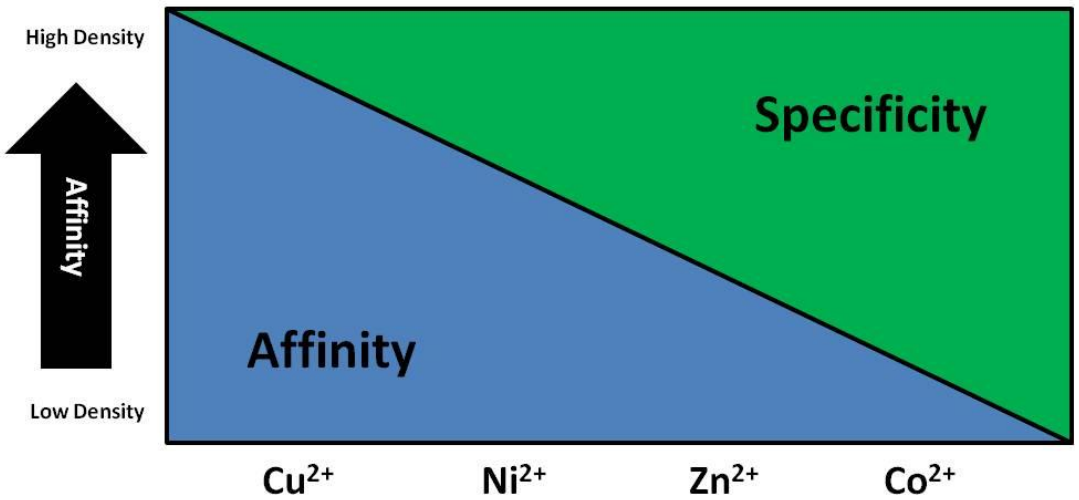
Comments for pTrcHis/CAT:
5191 nucleotides


Polyhistidin protein
Ni
Ni
Ni
Imidazol
Elution
Ni
Ni
Ni
Column regeneration
Ni
Ni
Ni


Loading
Target
Analyte
Immobilized
Receptor
Separation
Elution



Metal Ion Affinity vs. Specificity for His-Tag Protein Purification



IMAC advantages

- **The small size of the affinity ligand**. Due to the small size, it has minimal effects on the folding of the protein.

- If the his-tag is placed on the n-terminal end of the protein, **it can easily be removed using an endoprotease**.

- Another advantage of using his-tag purification methods is that polyhistidine tags **can bind proteins under both native and denaturing conditions**.

- The use of denaturing conditions becomes important when proteins are found in inclusion bodies and must be denatured so they can be solubilized.

IMAC disadvantages

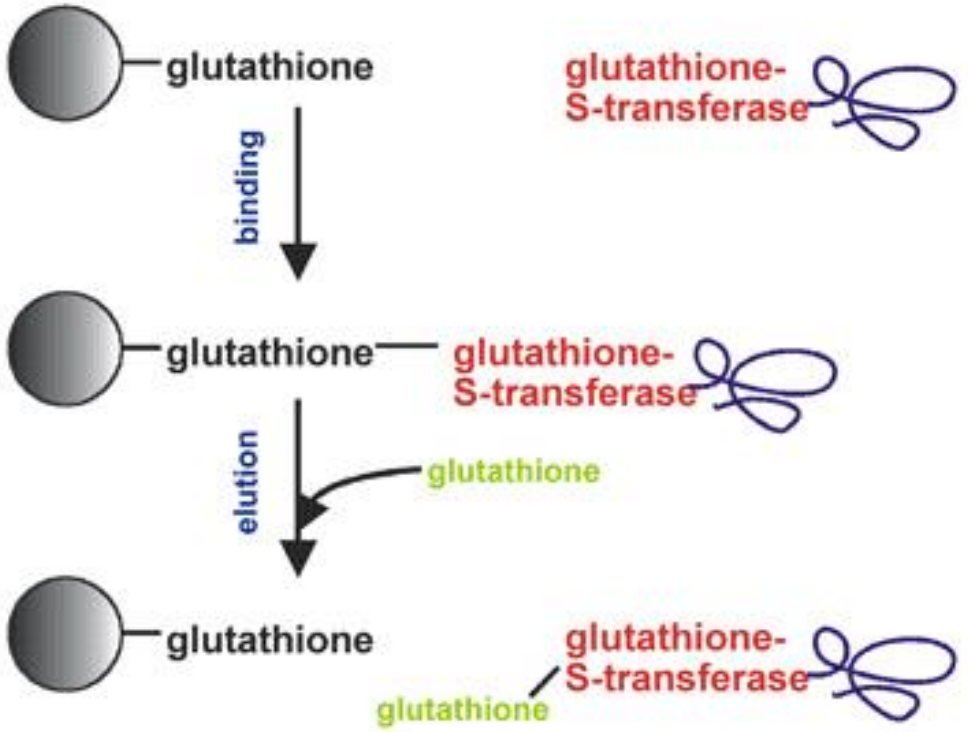
Potential degradation of the histag
When a few histidine residues are proteolytically degraded, the affinity of the tagged protein is greatly reduced leading to a decrease in the protein yield.

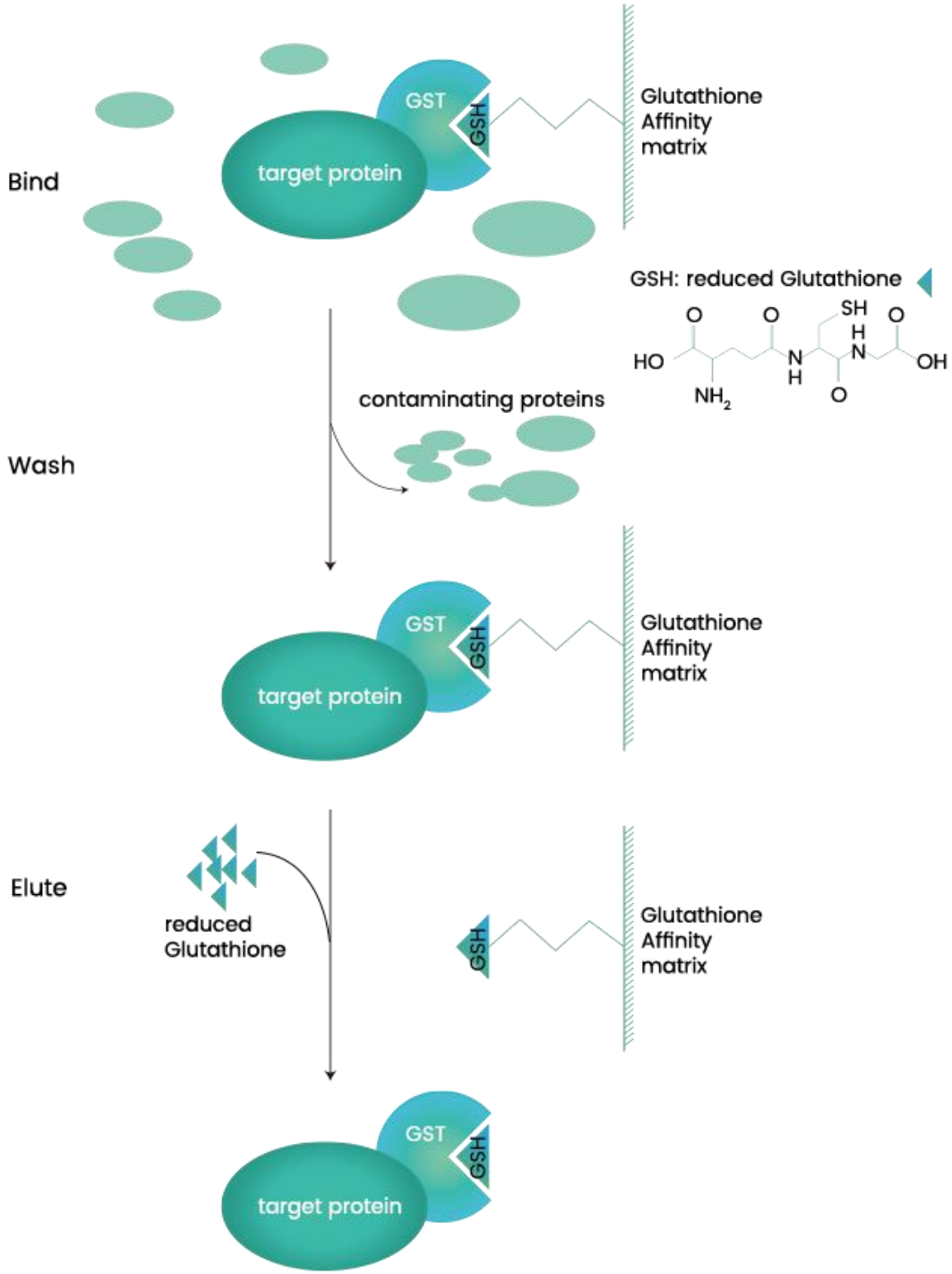
Dimer and tetramer formation,
Once a protein has a his-tag added to its structure, it has the potential to form dimers and tetramers in the presence of metal ions. While this is often not a large problem, it can lead to inaccurate molecular mass estimates of the tagged protein.

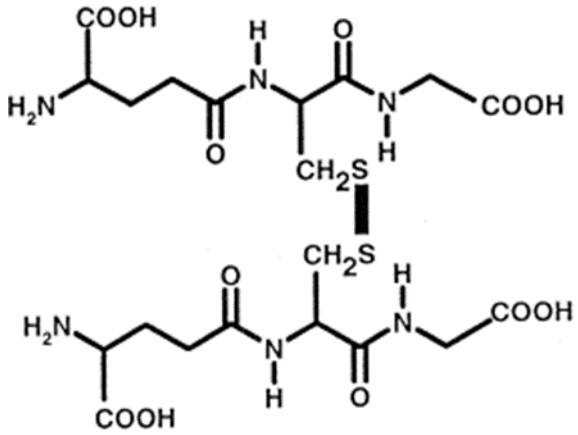
Coelution of other histidine-containing proteins.
Coelution of proteins that naturally have two or more adjacent histidine residues.

Glutathione S-transferase (GST) Gene Fusion System

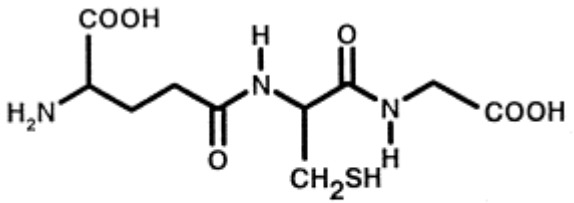
- A versatile system for the expression, purification, and detection of fusion proteins produced in *E. coli*.
- Based on inducible, high-level expression of genes or gene fragments as fusions with *Schistosoma japonicum* GST.
- The protein accumulates within the cell's cytoplasm.







Glutathione Oxidized, GSSG

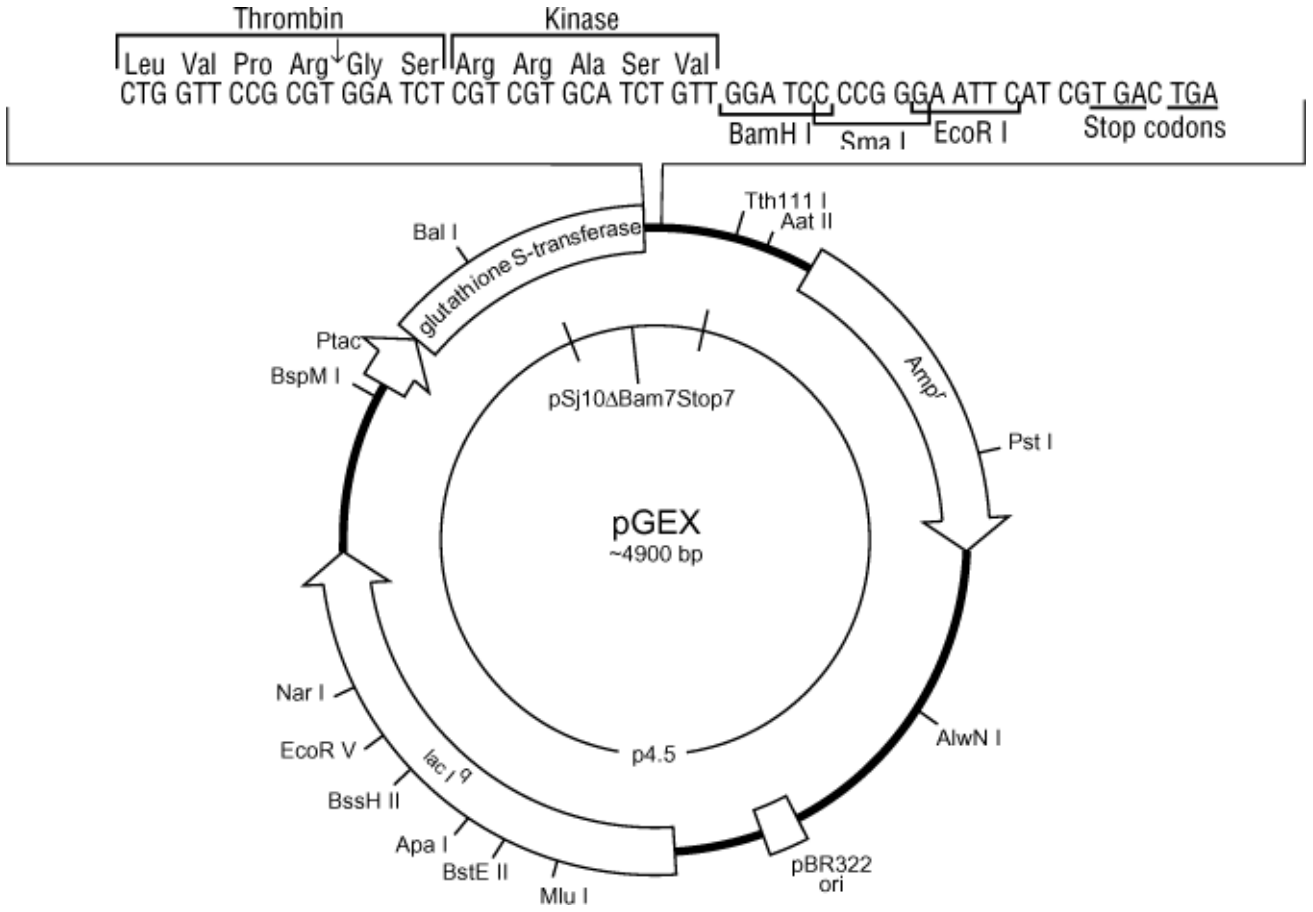


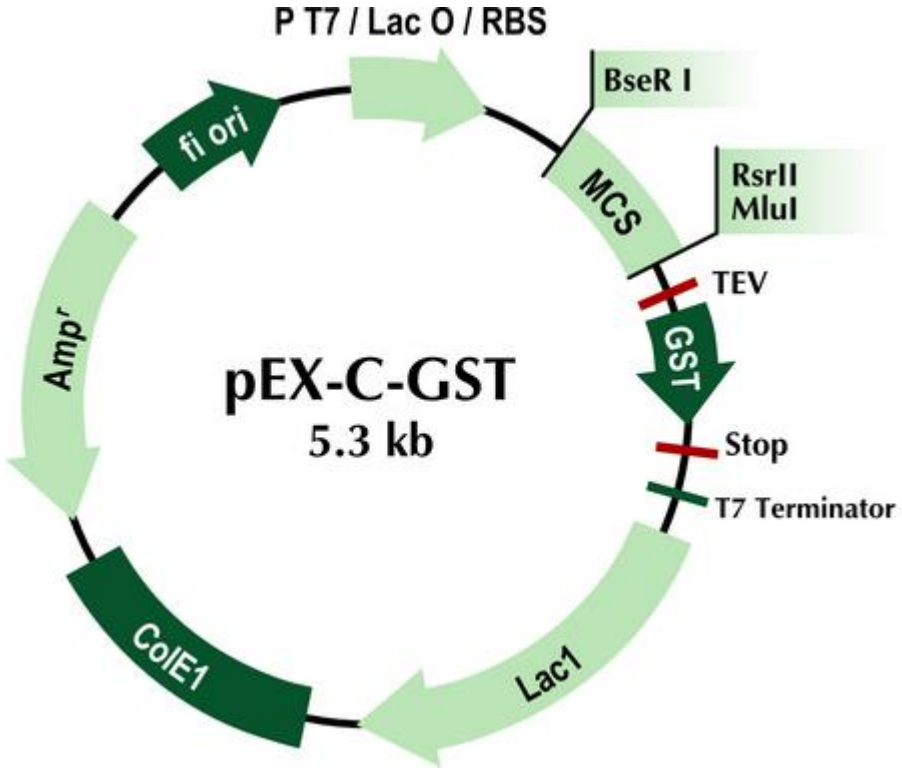
Glutathione Reduced, GSH

https://www.sinobiological.com/resource/protein-review/gst-tag-protein-expression

pGEX plasmid:

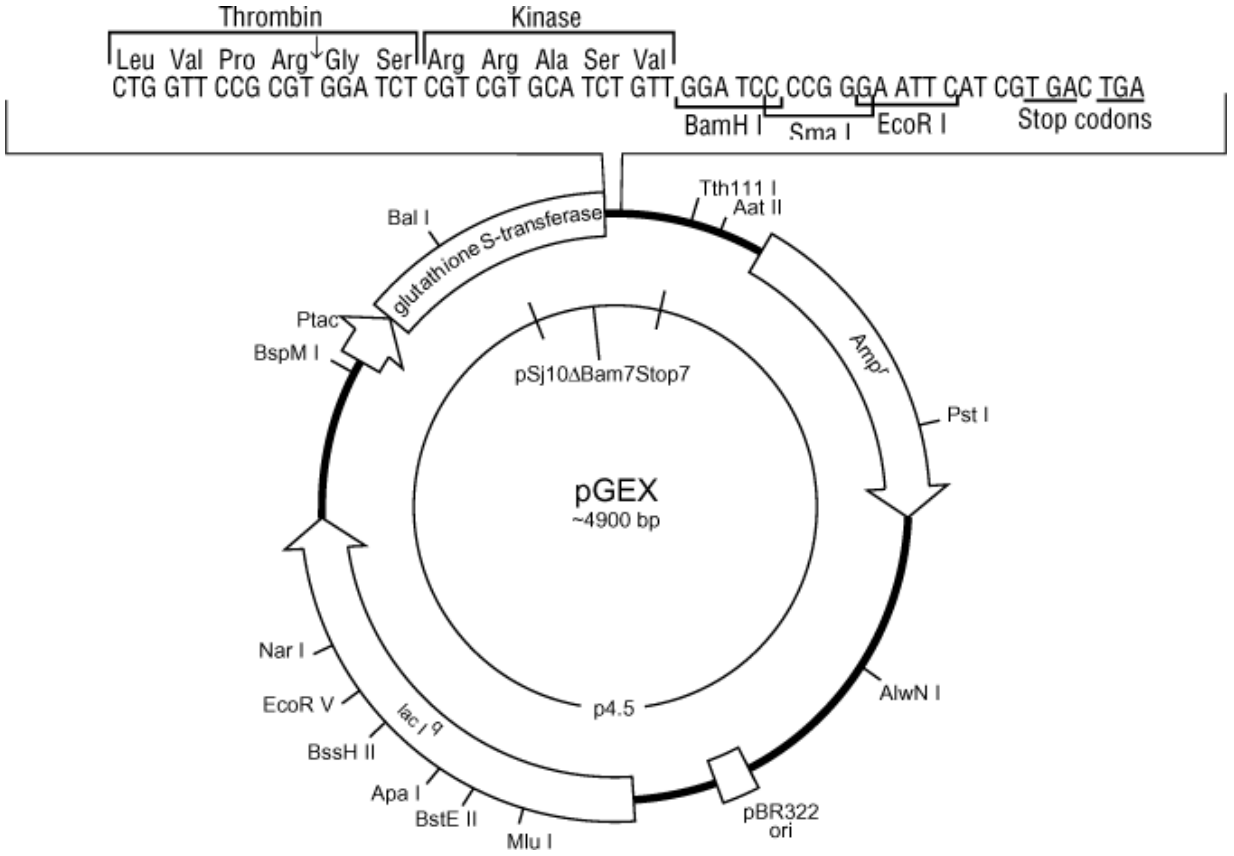
- Gene encoding affinity tag-glutathione S tranferase (GST)





- Cloning the gene or gene fragment into a pGEX expression vector pGEX vectors GST fusion proteins are constructed by inserting a gene or gene fragment into **the multiple cloning site of one of the pGEX vector**.

- Expression is induced by the lactose analog isopropyl b-D thiogalactoside (IPTG).



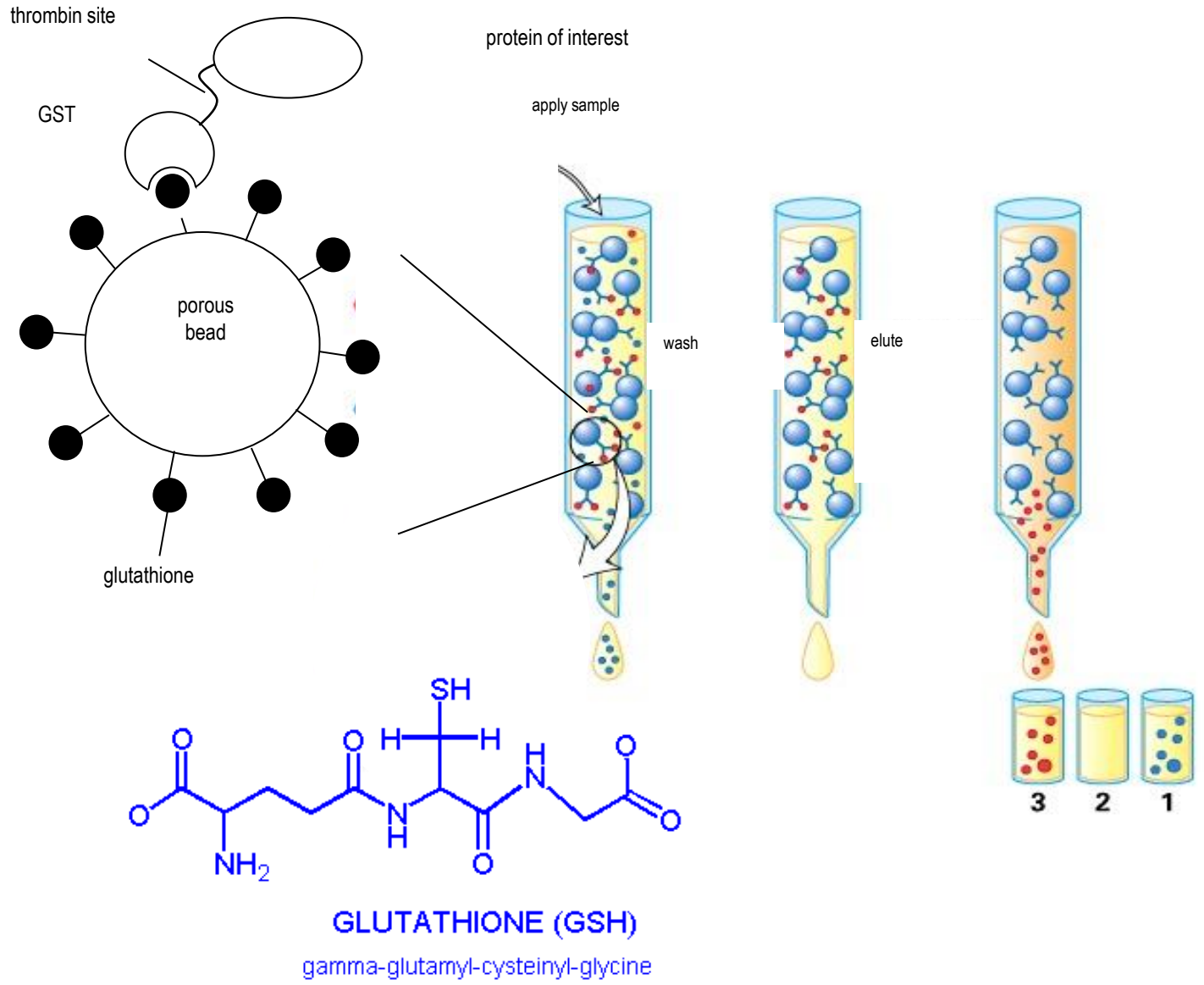
All pGEX vectors are also engineered with an **internal lac I^q gene**.
The lac I^q gene product is a repressor protein that binds to the operator region of the tac promoter, preventing expression until induction by IPTG, thus maintaining tight control over expression of the insert.

Affinity chromatography: separation by biological binding interactions

GST-tagged proteins bind to gluthatione on beads

Non-specifically or weakly bound proteins washed off

GST-tagged proteins **eluted with glutathione** (competitor) or thrombin (protease)



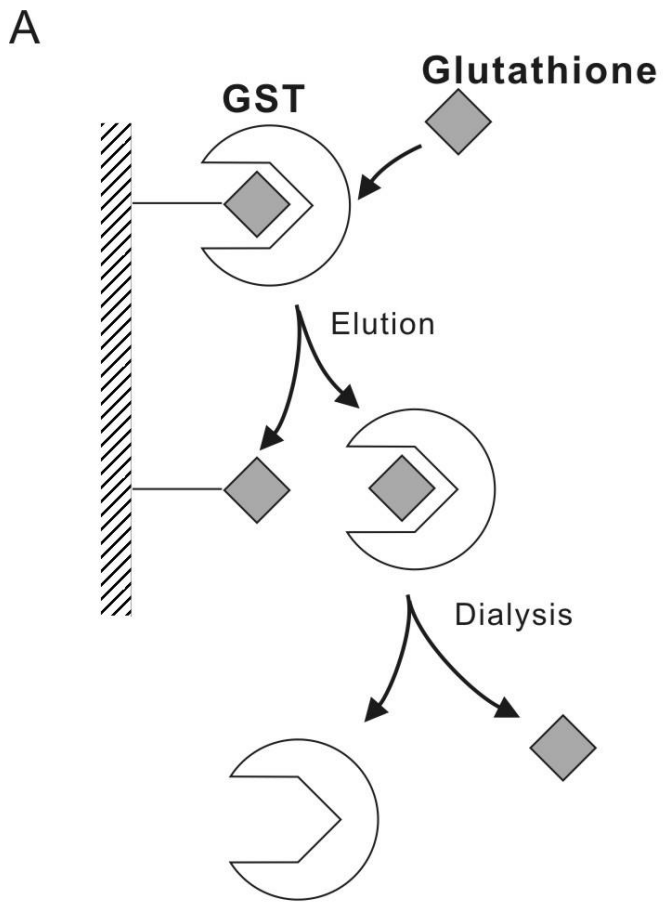
GST fusion proteins are purified from bacterial lysates by affinity chromatography **using immobilized glutathione**.

GST fusion proteins are captured by **the affinity medium**, and impurities are removed by washing.

Fusion proteins are **eluted under mild, non-denaturing conditions** using reduced glutathione.

The purification process preserves protein antigenicity and function.

If desired, **cleavage of the protein from GST can be achieved** using a site-specific protease whose recognition sequence is located immediately upstream from the multiple cloning site on the pGEX plasmids.



Advantages and disadvantages for using tags in fusion proteins

Plus:

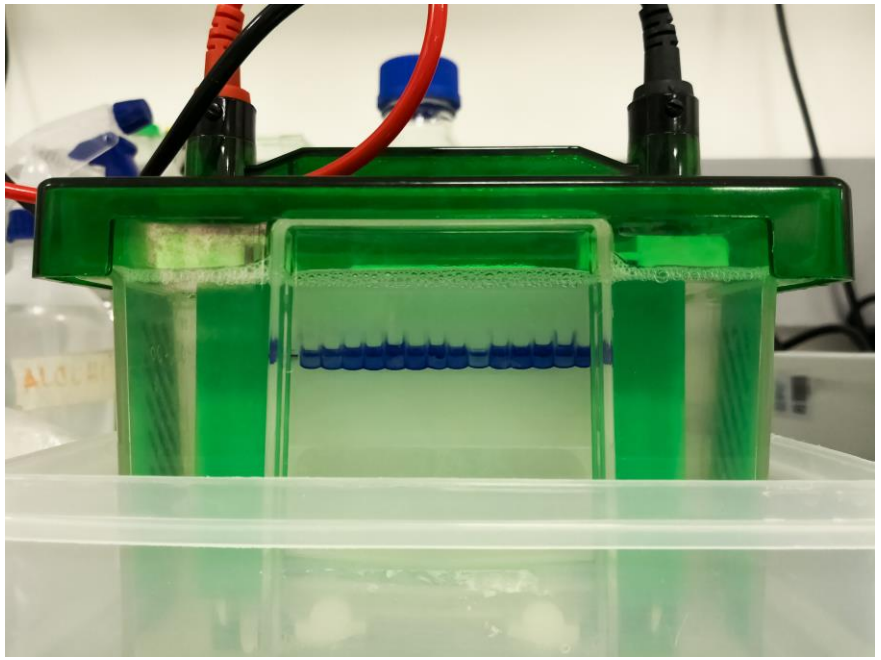
improve protein yield, prevent proteolysis, facilitate protein refolding, protect the antigenicity of the fusion protein, increase solubility

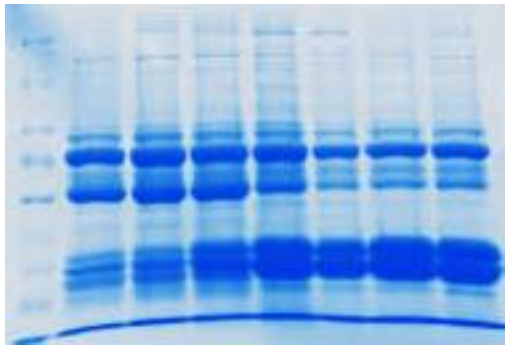
Minus:

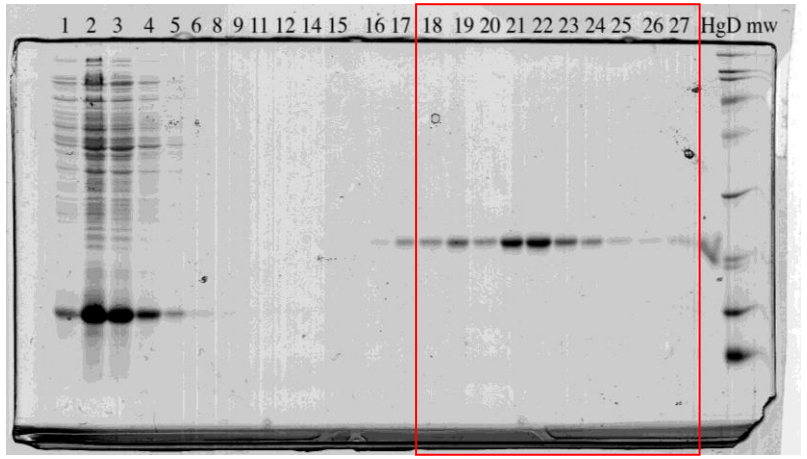
a change in protein conformation (solubility and activity), lower protein yields (cleavage may not be complete), inhibition of enzyme activity, alteration in biological activity, undesired flexibility in structural studies

Purity test

SDS-PAGE



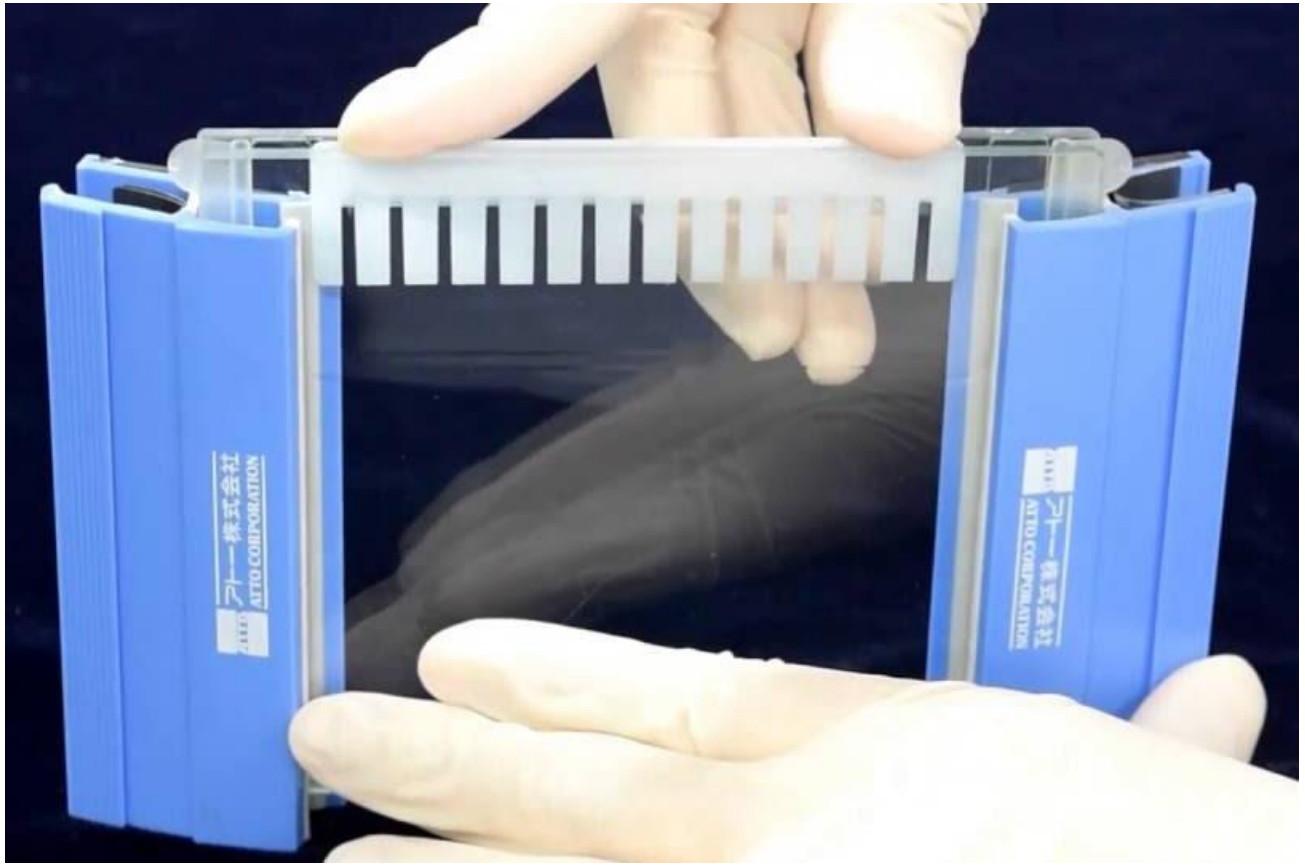




Mass spectrometry
N-terminal sequencing, etc.

SDS-PAGE

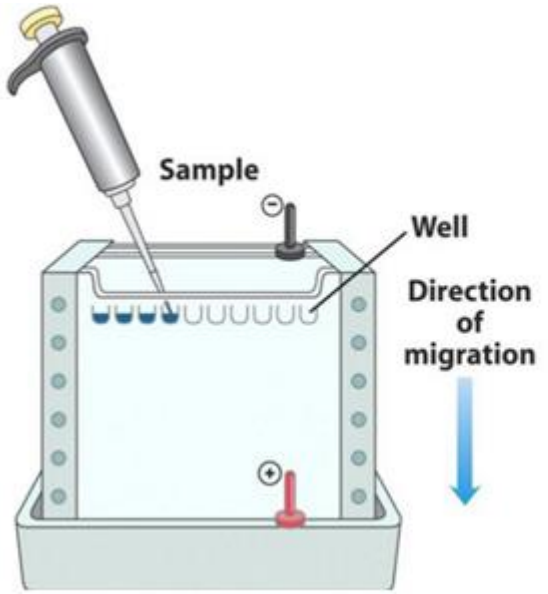
SDS: Sodium Dodecyl Sulfate
PAGE: PolyAcrylamide Gel Electrophoresis

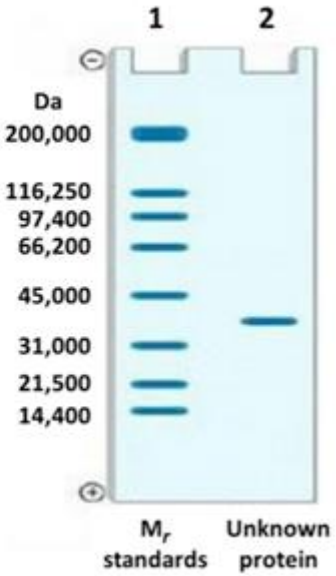


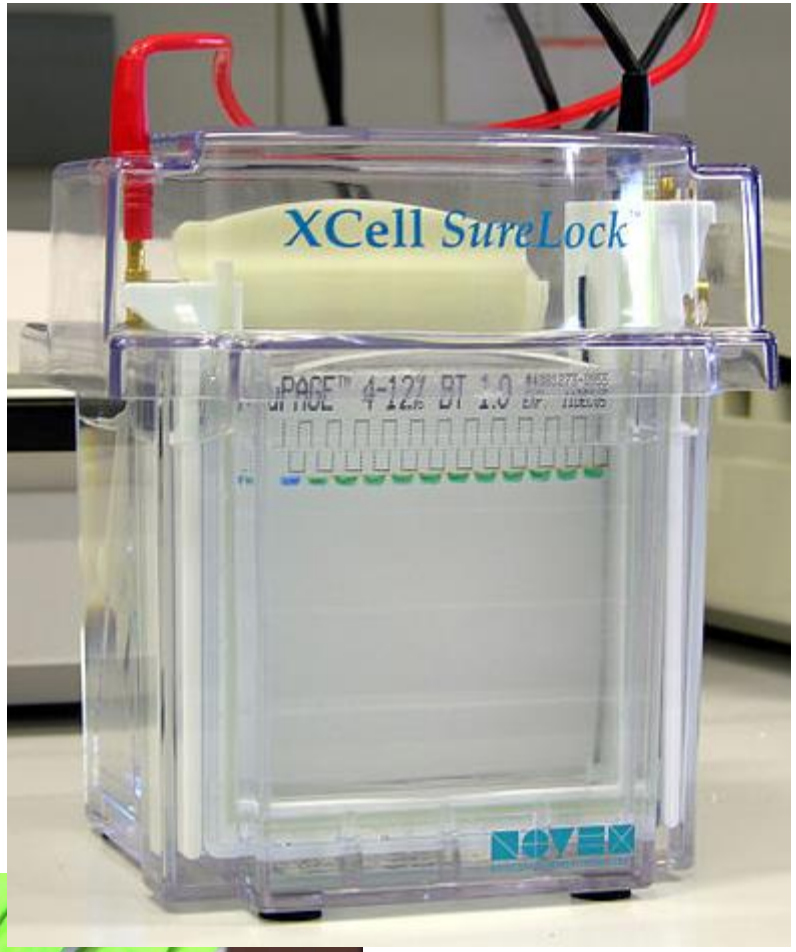
SDS-PAGE

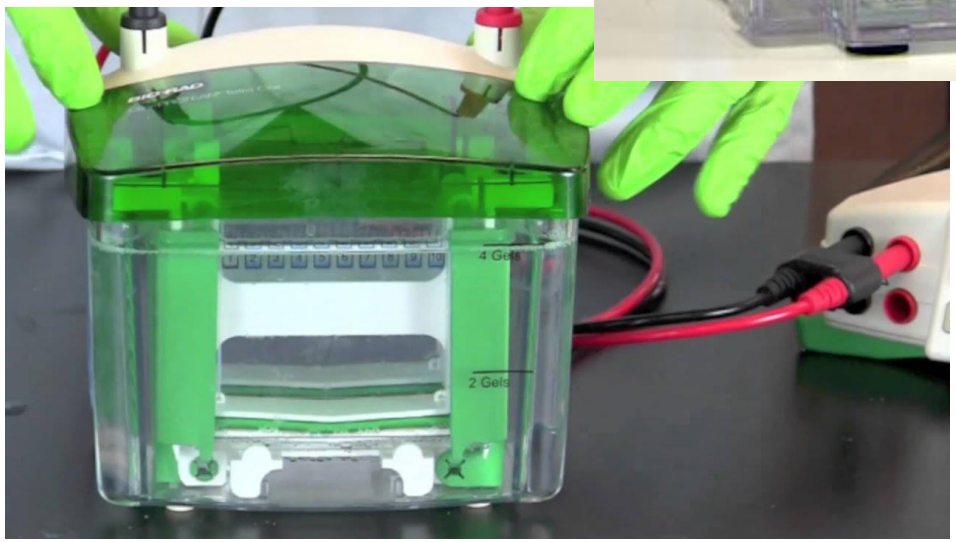
SDS: Sodium Dodecyl Sulfate
PAGE: PolyAcrylamide Gel Electrophoresis

- Separate proteins with molecular masses
- Protein sample + dye ➔ electrophoresis
  ➔ stain with silver staining/Coomassie

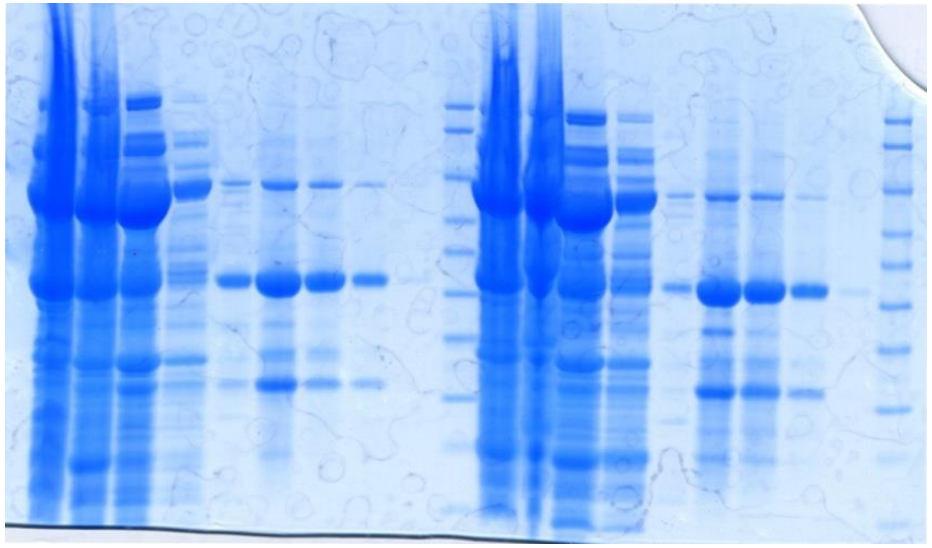


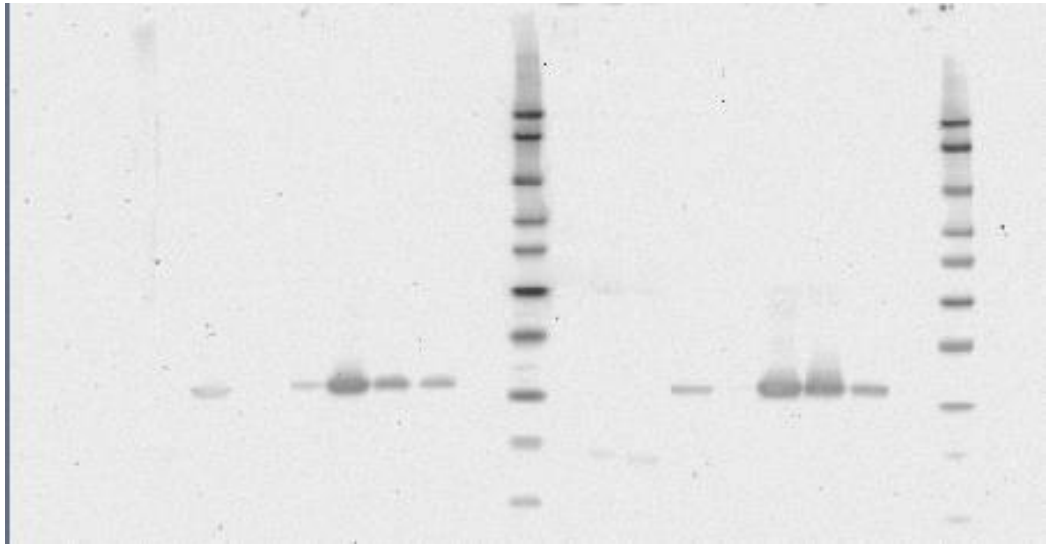






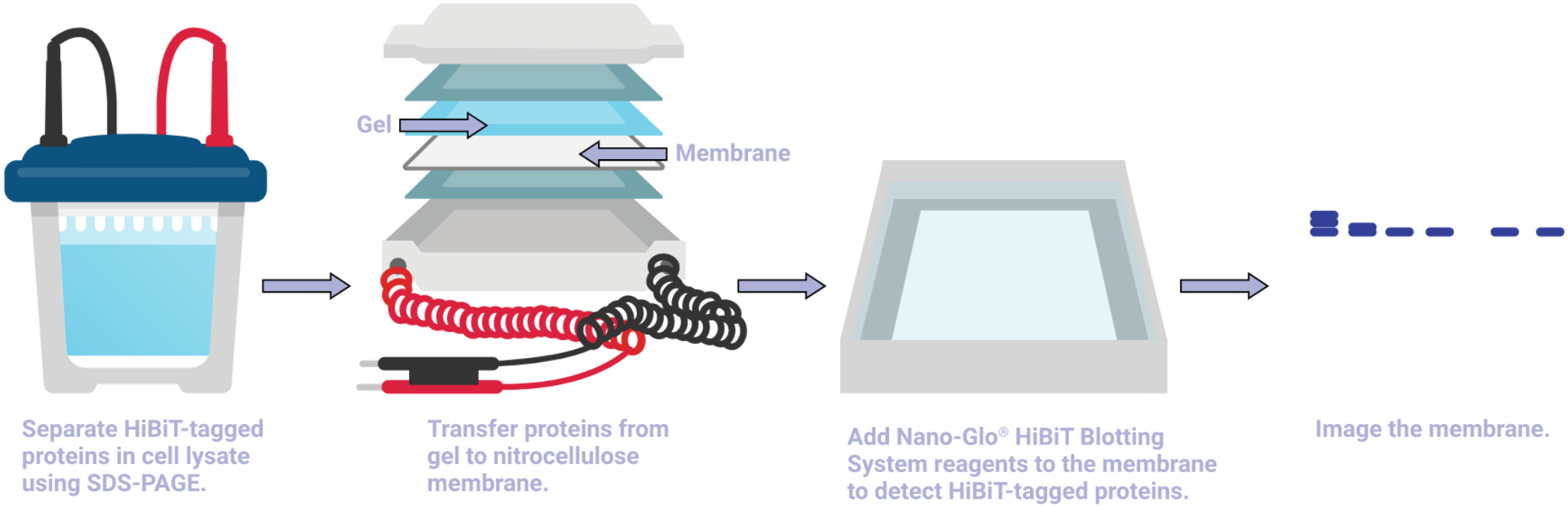
Confirming expressed protein by Western Blotting



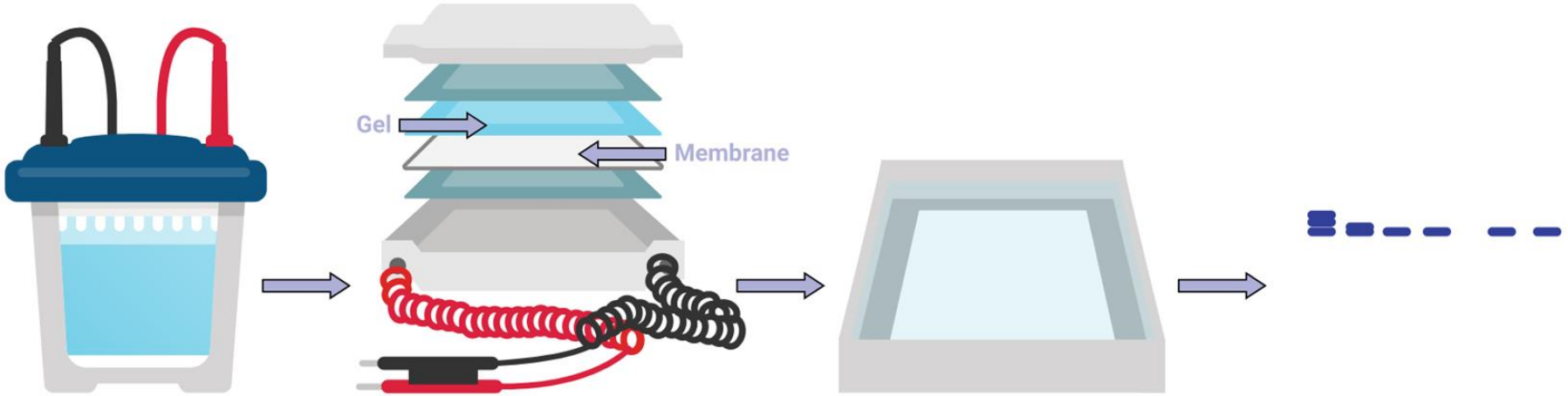


Western Blot

Target specific proteins from SDS-PAGE to a membrane (e.g. nitrocellulose) using specific *antibodies*.



https://worldwide.promega.com/resources/technologies/hibit-protein-tagging-system/faster-western-blotting/

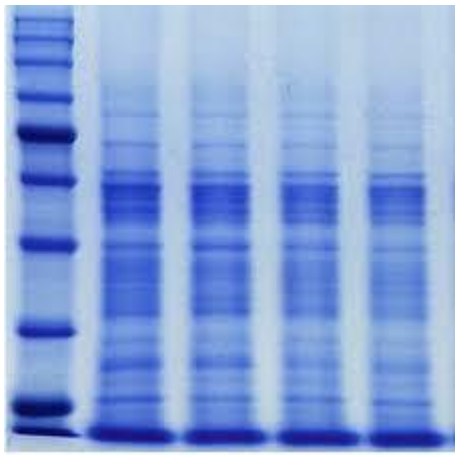


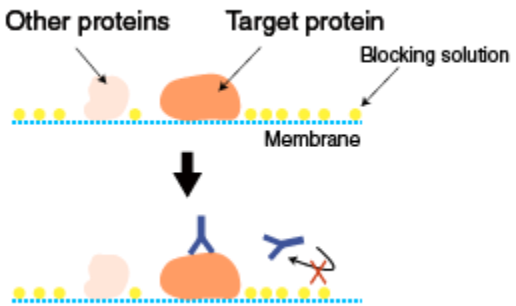
Separate HiBiT-tagged proteins in cell lysate using SDS-PAGE.

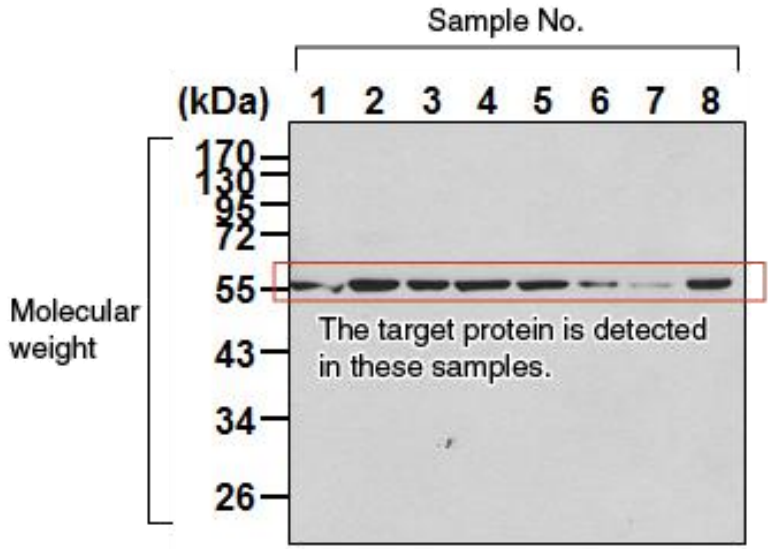
Transfer proteins from gel to nitrocellulose membrane.

Add Nano-Glo® HiBiT Blotting System reagents to the membrane to detect HiBiT-tagged proteins.

Image the membrane.







General problems with heterologus gene expression

Not enough protein is produced:

- Codon usage preferential (rare codon)
- Potential mRNA secondary structure.(5’-end, atcontent, 3’-en, transcriptional terminator)
- Toxic gene.

Enough protein is produced, but it is insoluble:

- Vary the growth temperature.
- Change fermentation medium.
- Low-copy-number plasmas.
- Selection of promoter.

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

Optimizing transcription of the cloned gene

- Genetic fusion to strong promoters (transcriptional fusion).
- Increased gene dosage (utilize the gene’s
- Own promoter with the gene on a high-copy plasmid).
- Potential problem with toxic genes and available methods for efficient repression.
- Solutions to potential problems with premature termination and mRNA instability.

Optimizing translation of the cloned gene

- Sequence determinants for translation initiation (Shine-Delargo sequence).
- Translational fusion vectors.
- Potential problem with biased codon usage.
- Enhancing the stability of protein products.

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

Baculovirus expression system

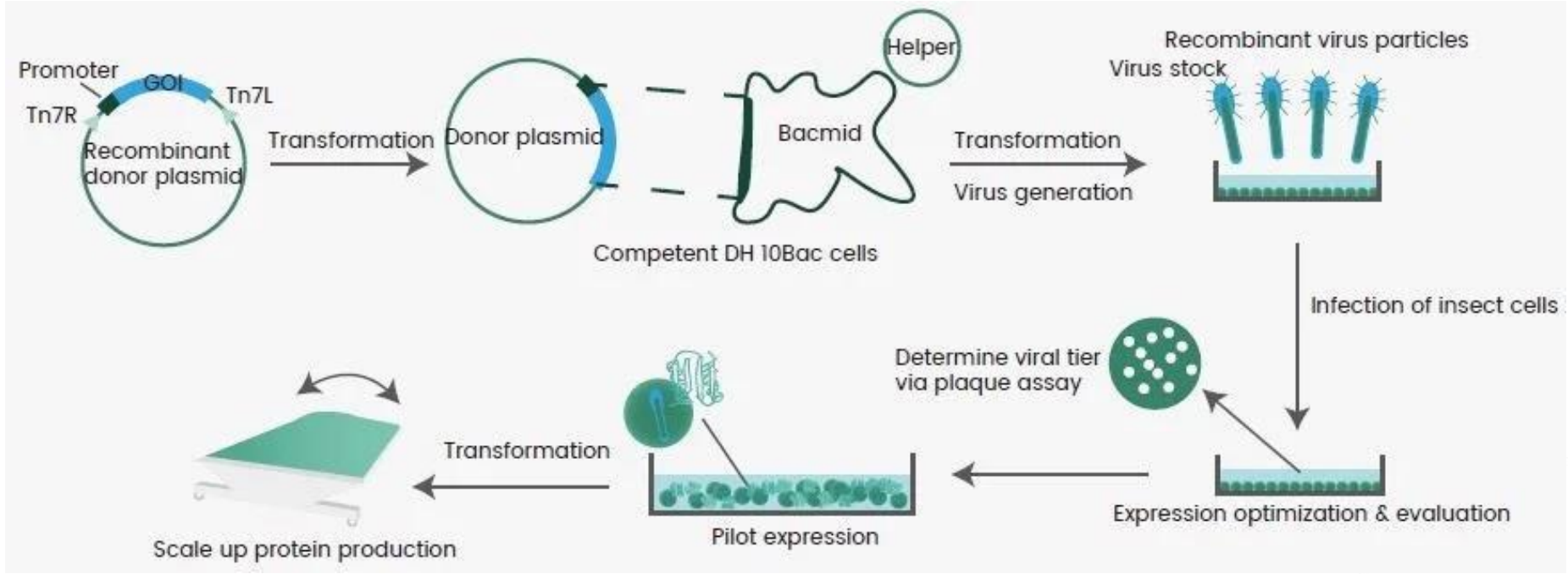
Baculovirus

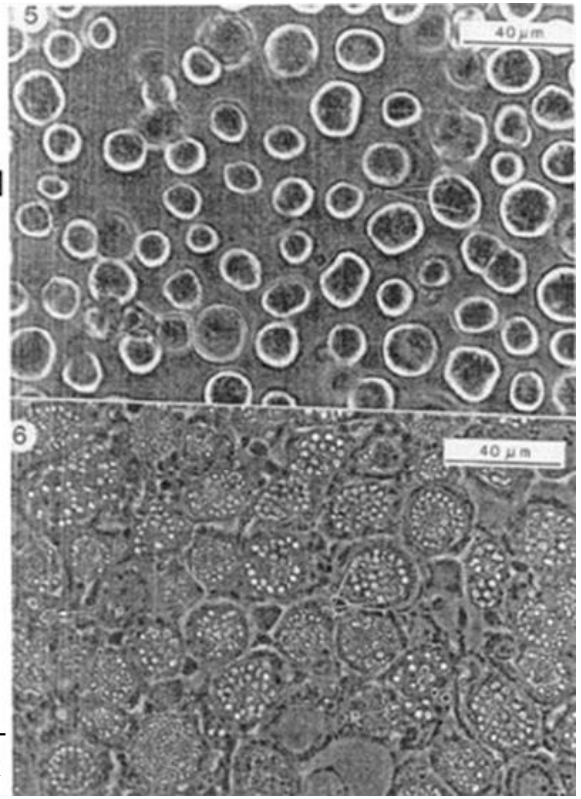
- Baculovirus are present in invertebrates primarily insect species
- They are not infectious for vertebrates and plants
- Genome is covalently closed circular double stranded of 134 kbp
- Can accommodate large fragments of foreign DNA
- 2 groups based on their structures:
  - Nucleopolyhedroviruses (NPV)
  - Granuloviruses

Insect cells

- Baculovirus infects lepidopteran insects and insect cell lines
- Commonly used cell lines are sf9 & sf21 derived from the pupal ovarian tissue of the fall army worm *Spodoptera frugiperda.*
- Recombinant baculovirus have become widely used as vectors to express heterologous genes in cultured insect cells and insect larvae.

- Heterologous genes placed under the transcriptional control of the strong polyhedrin promoter of the *Autographa californica polyhedrosis virus* (AcNPV) >> fuse into a baculovirus shuttle vector (bacmid).





Sf9 non-infected

Sf9 infected

https://www.news-medical.net/whitepaper/20220509/Expressing-recombinant-proteins-with-a-baculovirus-insect-cell-system.aspx