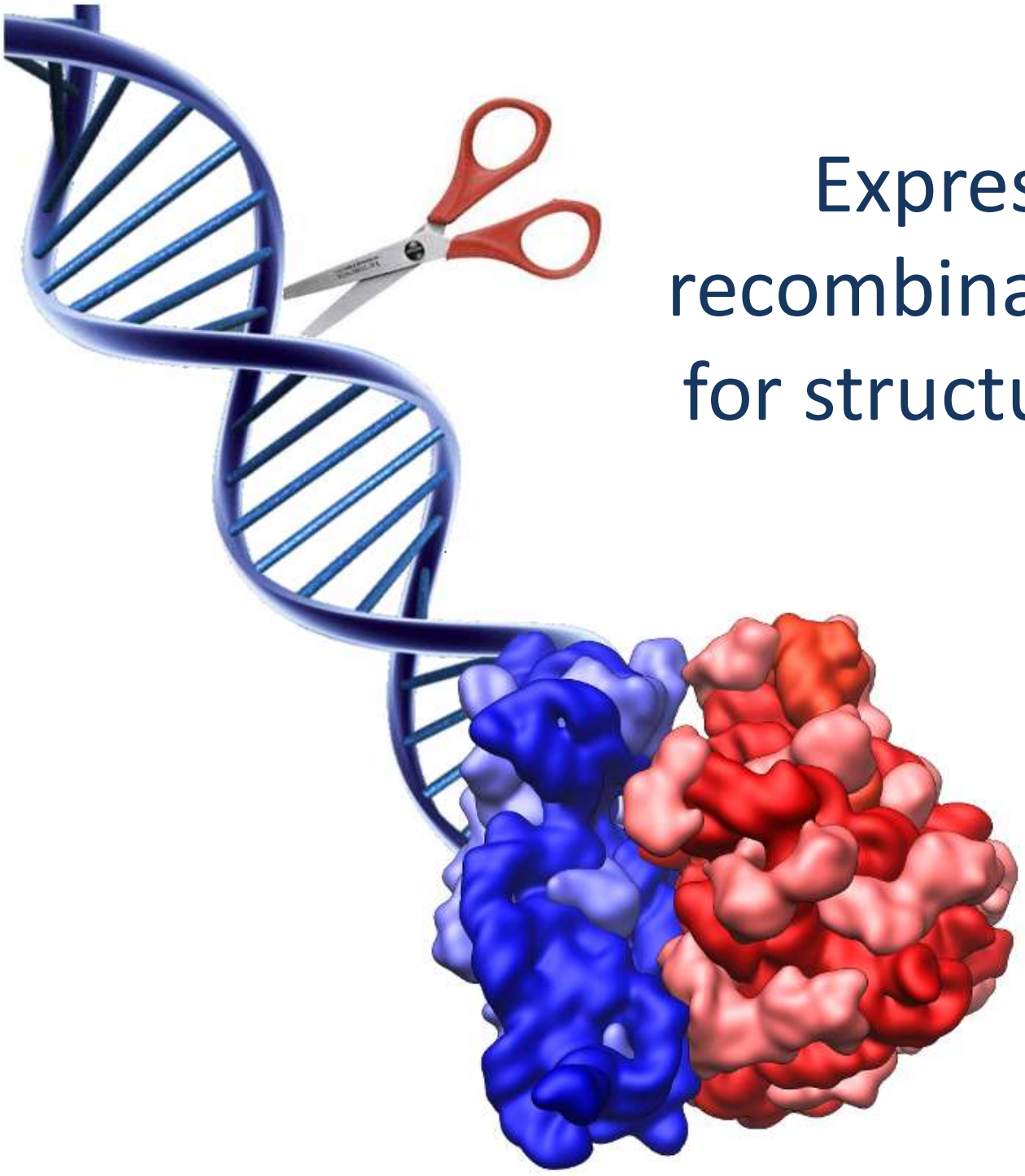


Expression of recombinant proteins for structural studies

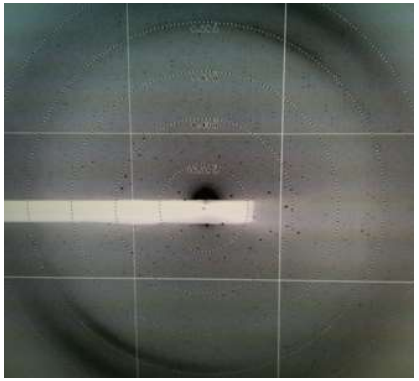
Corso di Biocristallografia e Microscopia Elettronica

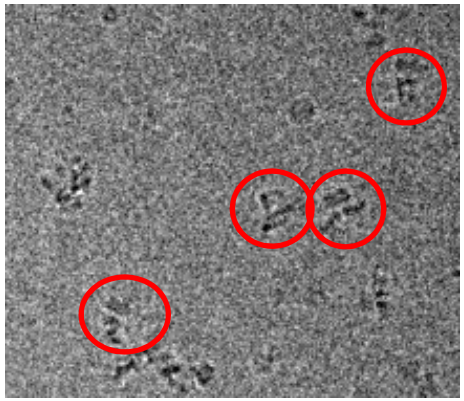
rdezorzi@units.it

Proteins suitable for structural studies

<u>Crystallography</u>:

- LARGE amounts of protein for crystallization experiments! (>5mg)
- High purity to facilitate crystallization
- Stability



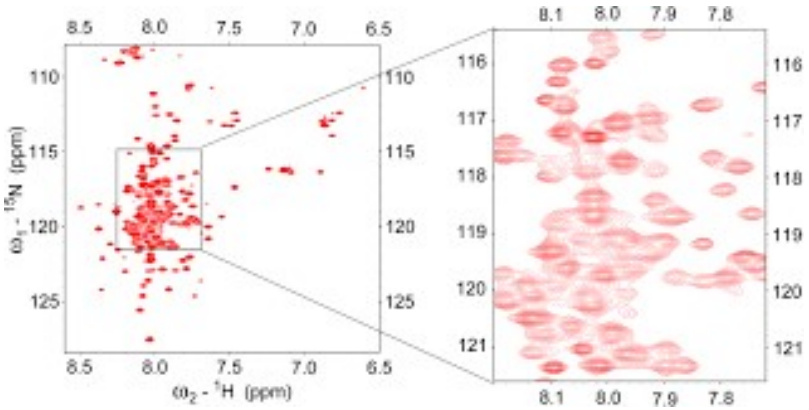


<u>Electron Microscopy</u>:

- High purity! And conformational homogeneity
- Stability in solution
- Suitable buffer

<u>NMR</u>:

- Isotopic labelling
- Stability
- Conformational homogeneity



Why recombinant proteins?

In the old days…

First structure of myoglobin (Kendrew & Perutz, Nobel 1962) determined from crystals of the protein extracted from whale meat. Myoglobin is abundant in meat!!

Today…

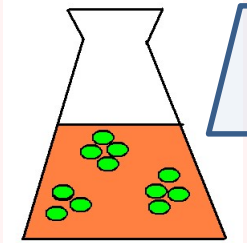
Growing interest for proteins with very low expression, but important cellular functions (enzymes! regulatory proteins! membrane proteins!).
Natural sources cannot provide enough protein for structural studies.

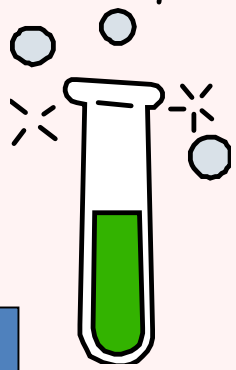
For crystallization of some proteins, engineering is required for stabilization, purification, mutational studies…

For NMR studies, engineering required for isotopic labeling.

Bioinfor matics

Molecular biology

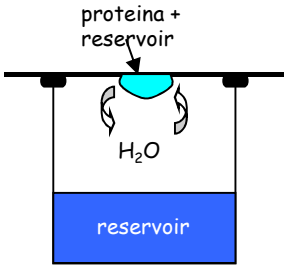


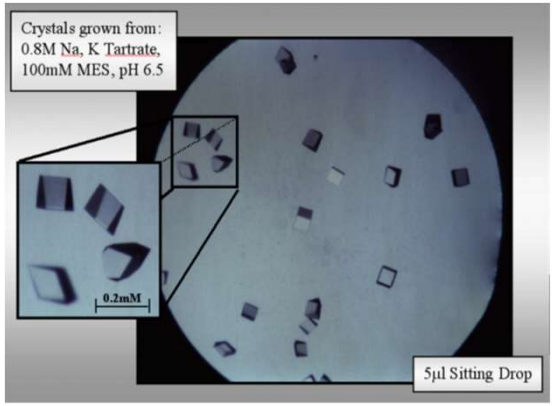


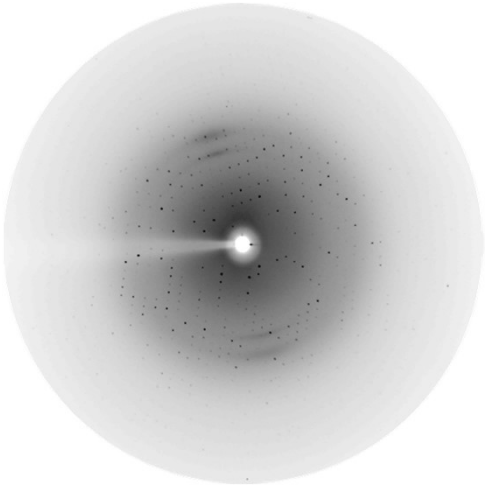
Expression and purification

Chemistry

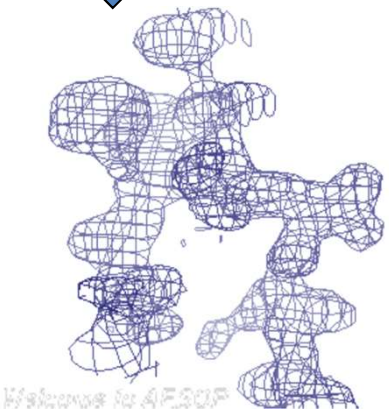
Crystallization



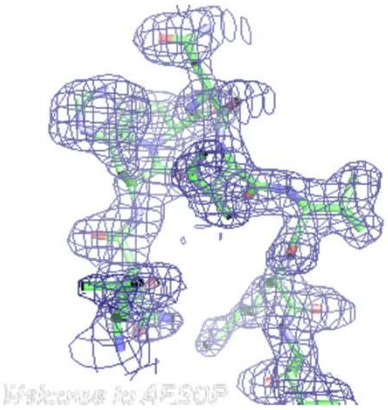




Diffraction data collection



Phasing



Structure refinement

10 days to
10 years

Preliminary knowledge

Expasy
Swiss Bioinformatics Resource Portal

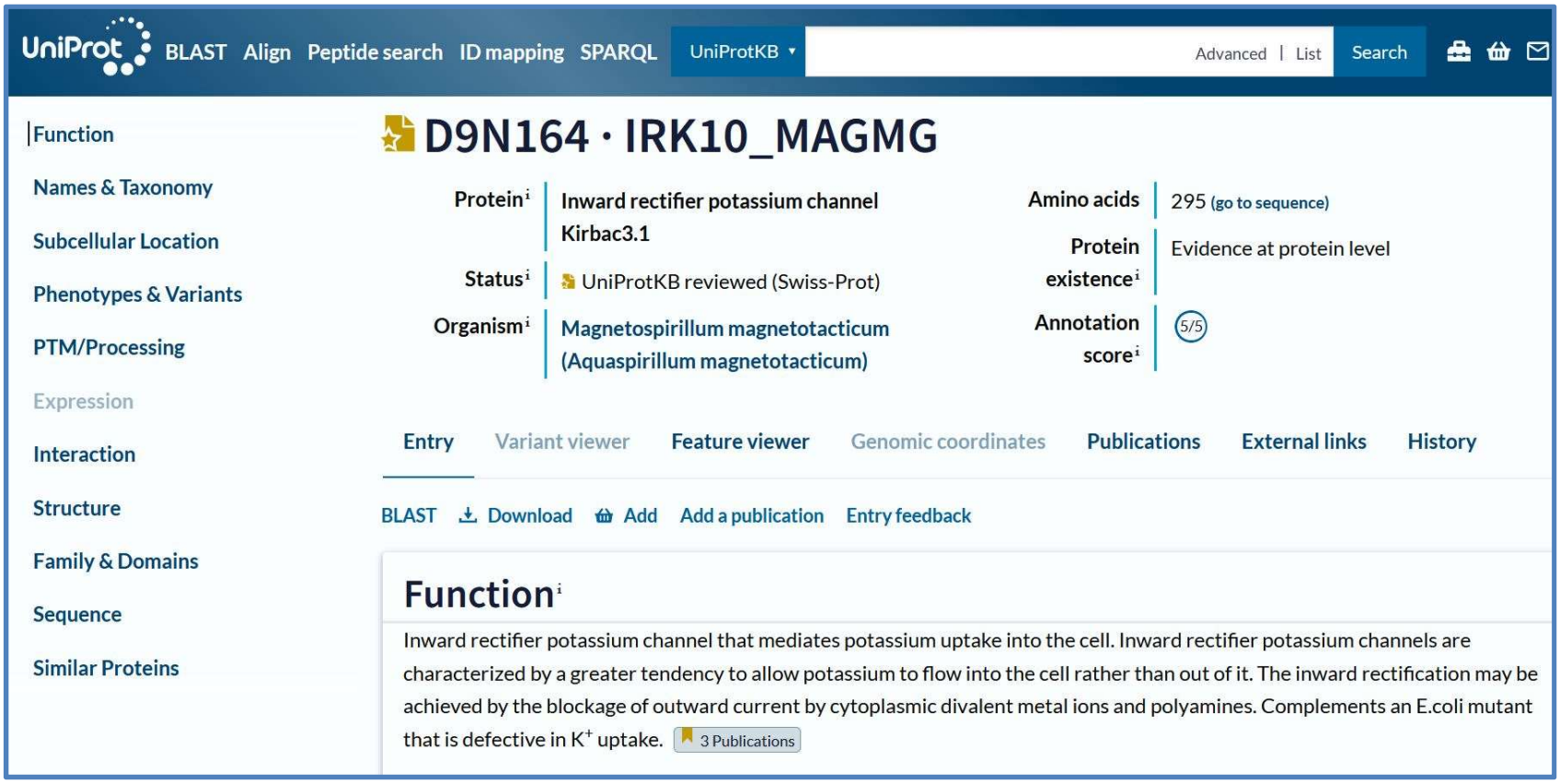
http://www.expasy.org/

UniProt

http://www.uniprot.org/

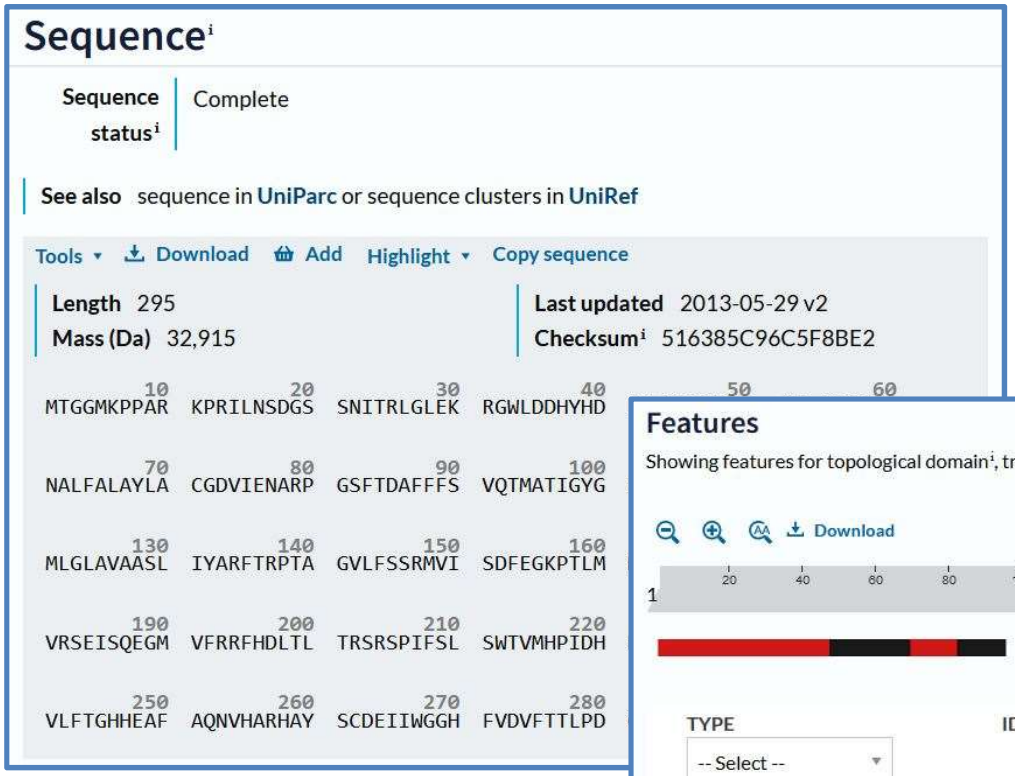
- **Function of the protein.**

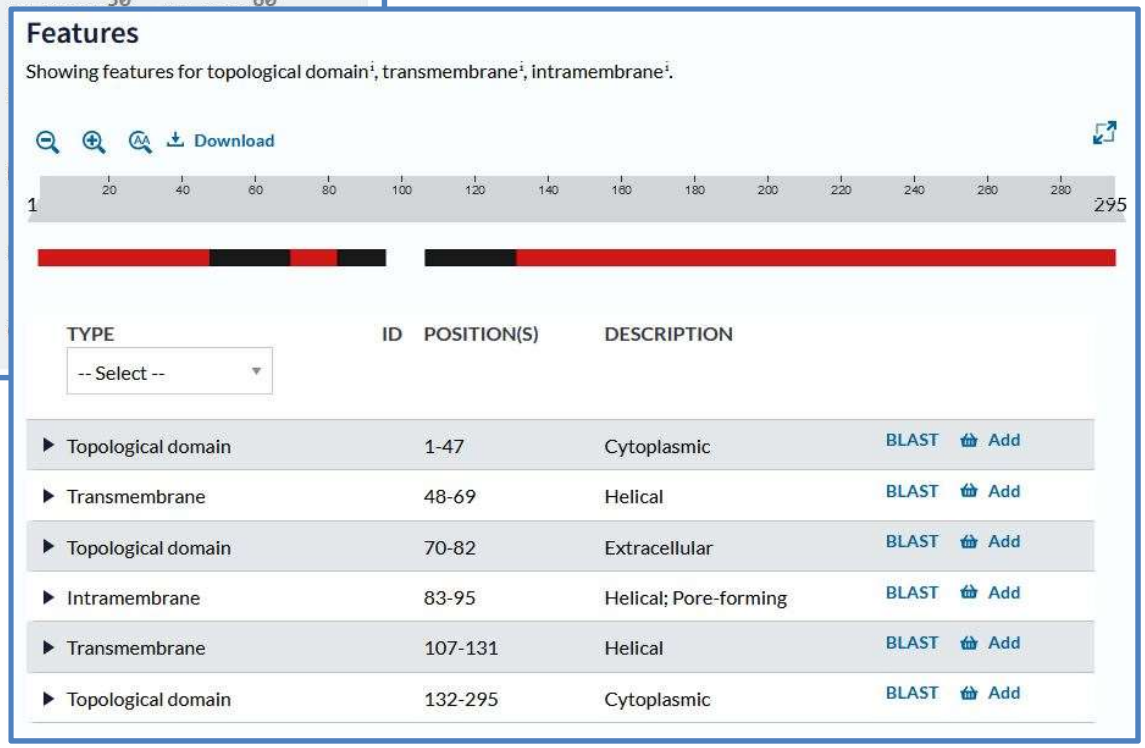
Prokaryotic? Eukaryotic? Family?
Localization: cytoplasm, membrane, periplasm?
Expression levels in vivo
Interactions with other proteins



Preliminary knowledge

- **Physical and chemical properties of the protein.**





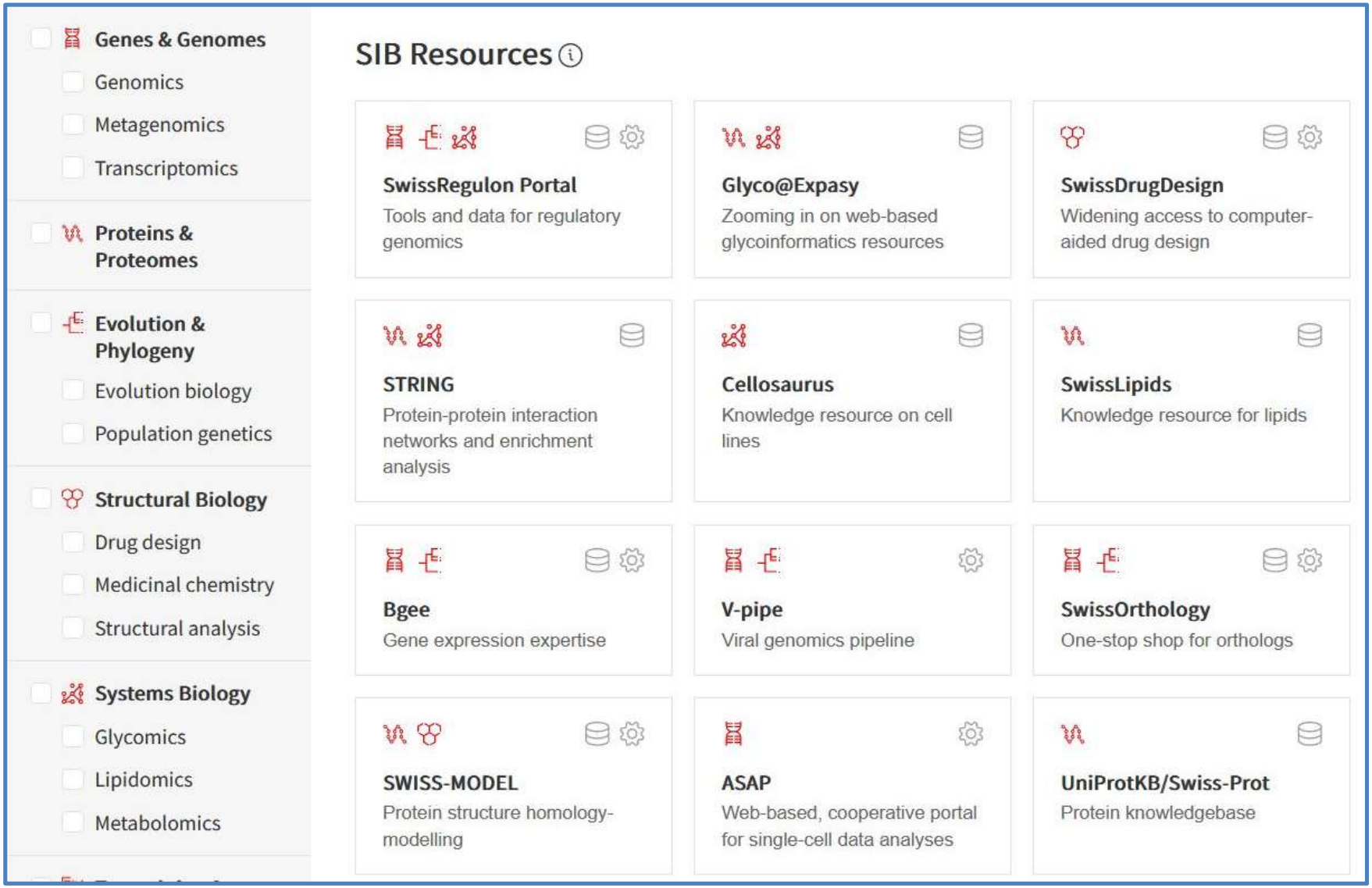
Sequence
Number of residues and MW
pI and hydrophobicity
Post-translational modifications (glycosilation…)

Preliminary knowledge

Expasy

Swiss Bioinformatics Resource Portal

http://www.expasy.org/



Preliminary knowledge

- **Physical and chemical properties of the protein.**

Expasy — Compute pI/Mw — Home | Contact

Compute pI/Mw

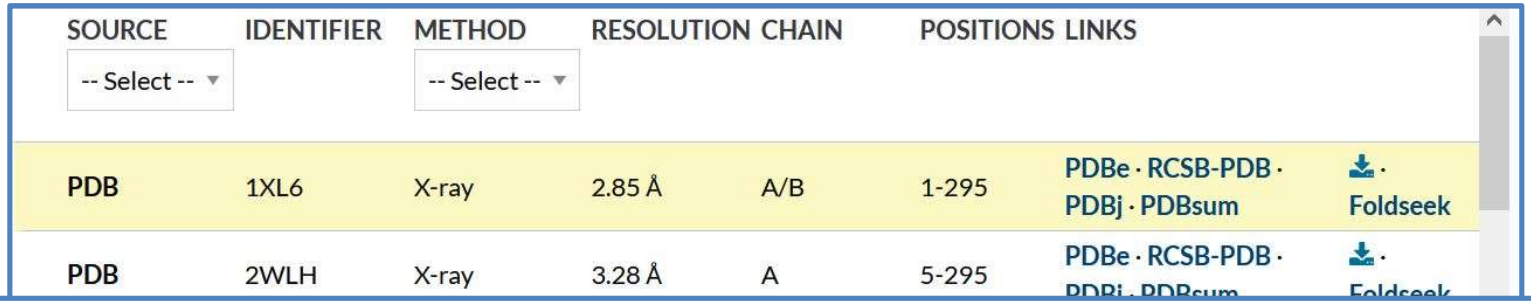
Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MTGGMKPPAR KPRILNSDGS SNITRLGLEK RGWLDDHYHD LLTVSWPVFI TLITGLYLVT

        70         80         90        100        110        120
NALFALAYLA CGDVIENARP GSFTDAFFFS VQTMATIGYG KLIPIGPLAN TLVTLEALCG

       130        140        150        160        170        180
MLGLAVAASL IYARFTRPTA GVLFSSRMVI SDFEGKPTLM MRLANLRIEQ IIEADVHLVL

       190        200        210        220        230        240
VRSEISQEGM VFRRFHDLTL TRSRSPIFSL SWTVMHPIDH HSPIYGETDE TLRNSHSEFL

       250        260        270        280        290
VLFTGHHEAF AQNVHARHAY SCDEIIWGGH FVDVFTTLPD GRRALDLGKF HEIAQ
```

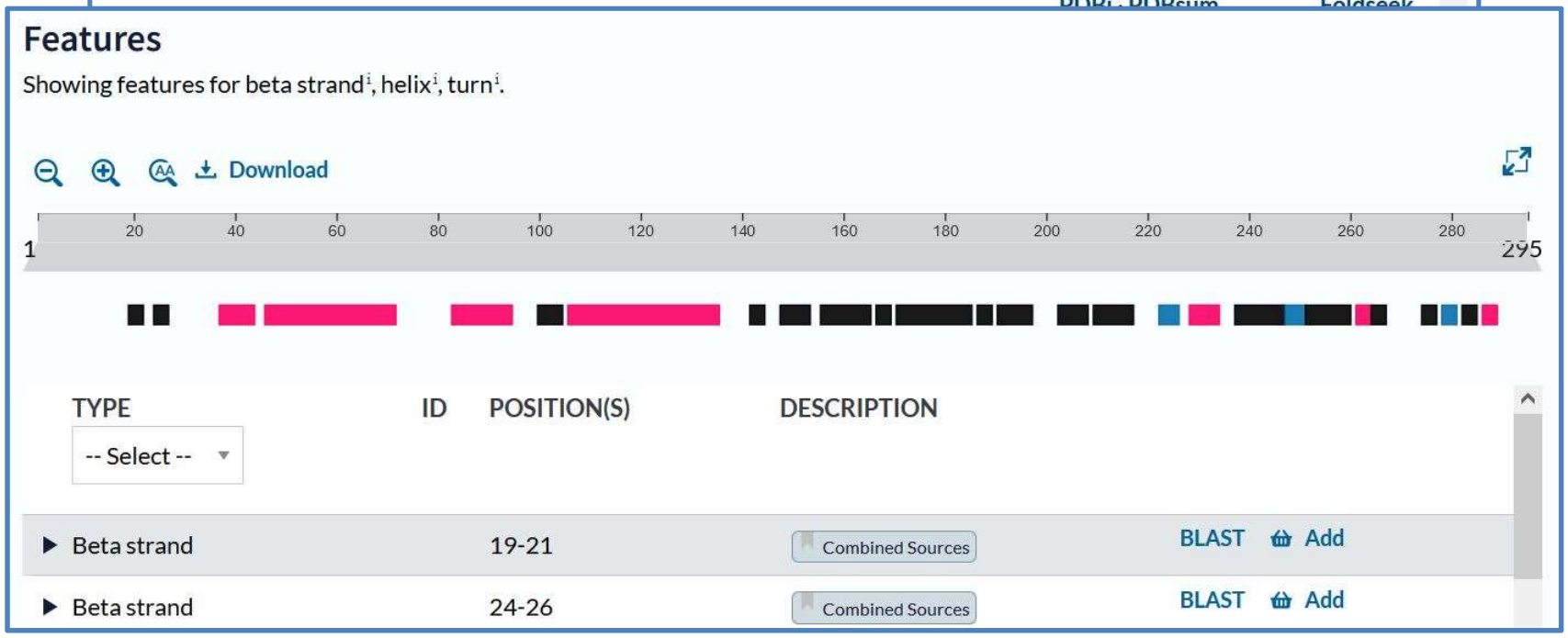
Theoretical pI/Mw: 6.38 / 32914.94

Preliminary knowledge

- **Structure.** Homologous proteins of known structure
  Domains and oligomerization state
  Structure prediction (α-helices, β-strands, flexible loops, unstructured domains)



SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
-- Select --		-- Select --				
PDB	1XL6	X-ray	2.85 Å	A/B	1-295	PDBe · RCSB-PDB · PDBj · PDBsum · Foldseek
PDB	2WLH	X-ray	3.28 Å	A	5-295	PDBe · RCSB-PDB · PDBj · PDBsum · Foldseek



Features

Showing features for beta strand[i], helix[i], turn[i].

Download

1 20 40 60 80 100 120 140 160 180 200 220 240 260 280 295

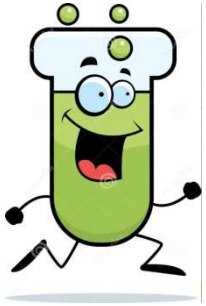
TYPE	ID	POSITION(S)	DESCRIPTION	
-- Select --				
▶ Beta strand		19-21	Combined Sources	BLAST Add
▶ Beta strand		24-26	Combined Sources	BLAST Add

Expression of recombinant proteins

Choice of the expression system

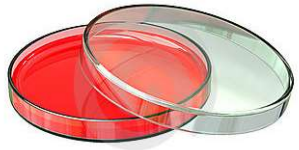
Choice of host and vector

Construct design



Cloning and vector ligation

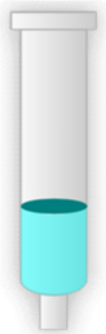
Colony selection



Sequencing

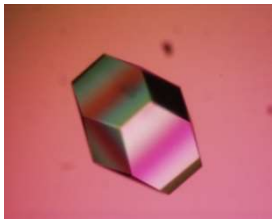
Expression protocol optimization

Protein purification



Sequence and activity check

Crystallization

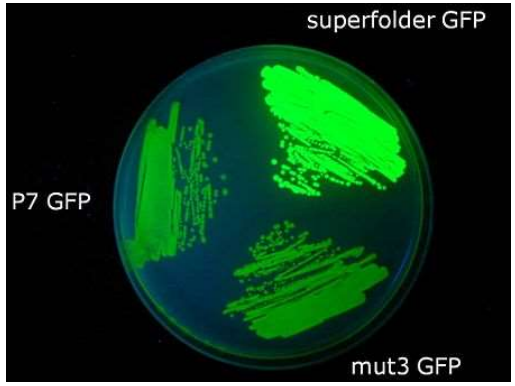


Expression systems

PROKARYOTIC HOSTS

(strains of E.Coli)

- No post-translational modifications
- Small proteins
- Cheap
- Good expression levels



CELL-FREE SYSTEMS

- Template + various cell extracts (commercially available from *E.Coli* to Human!)
- Post-translational modifications
- Easy for isotopic labelling
- Fast but expensive

EUKARYOTIC HOSTS

Yeast (S.Cerevisiae, P.Pastoris):

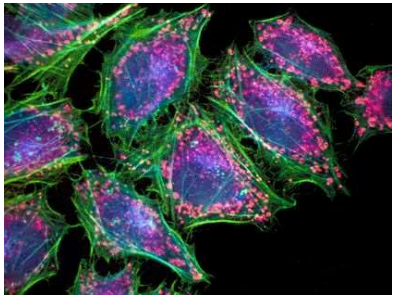
- Eukaryotic proteins with some post-translational modifications
- Cheap and easy

Baculovirus-Insect cells (Sf9, Sf21, Drosophila S2):

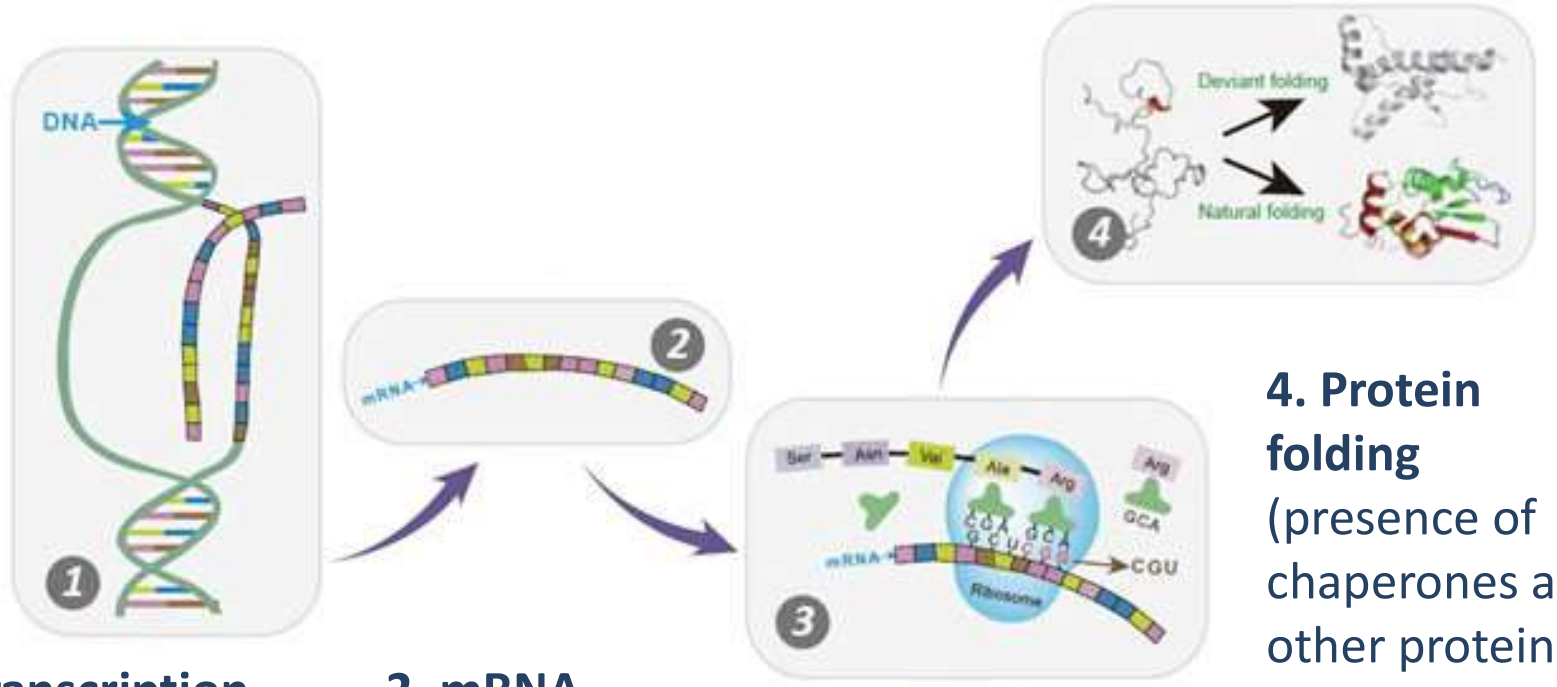
- Post-translational modifications
- No selection antibiotics
- Expensive!

Mammalian (CHO, 293, HeLa):

- All post-translational modifications
- No selection antibiotics
- Expensive & time consuming



Protein expression steps



1. Transcription (start and termination codons, protein binding sites, regulatory elements…)

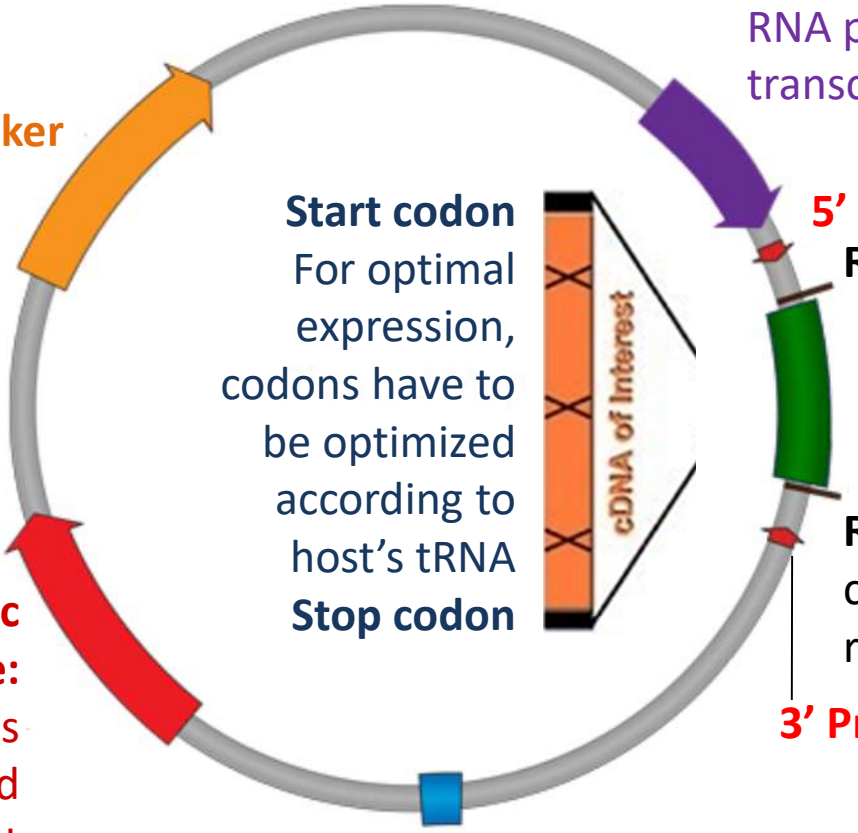
2. mRNA processing (splicing, mRNA secondary structure, stability of mRNA…)

3. Translation (codon usage specific for the host, tRNAs, ribosomal binding sites…)

4. Protein folding (presence of chaperones and other proteins that help folding, temperature, protease degradation…)

Expression vector

Contains genetic information to produce protein, replicates independently from the host's chromosomal DNA



Promoter: controls binding of RNA polymerase and transcription factors

Selectable marker

Start codon
For optimal expression, codons have to be optimized according to host's tRNA
Stop codon

5' Primer Site

Restriction Site

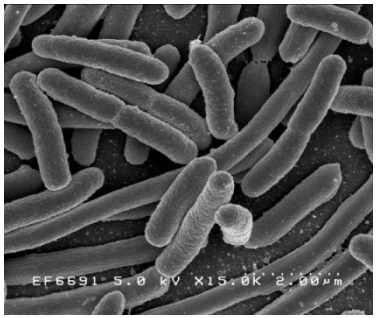
Inserted gene: in the Multiple Cloning Site (MCS)

Restriction Site: cleavage site for restriction enzymes

3' Primer Site

Antibiotic Resistance Gene: to select cells containing the plasmid and help them beat competition of other organisms

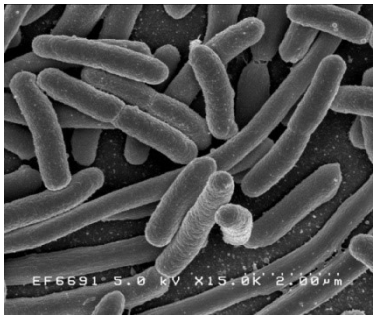
Origin of Replication (ORI): controls plasmid replication and number of copies



E.Coli strains /1

E.Coli strains is the host generally used for cloning and plasmid replication.

Strain	Natural resistance	Primary use	Genotype
DH10B	Streptomycin	General cloning and storage, blue-white screening.	*F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-*
DH5α	-	General cloning and storage, blue-white screening.	*F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K-m_K+), λ-*
Top10	Streptomycin	General cloning and storage, blue-white screening.	*F- mrcA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara,leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ-*
XL1 Blue	Tetracycline	Blue-white screening, routine cloning.	*endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K-m_K+)*
XL10 Gold	Tetracycline and Chloramphenicol	Cloning and propagation of large plasmids, high competency.	*endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB lacI^q ZΔM15 Tn10(Tet^R Amy Cm^R)]*



E.Coli strains /1

E.Coli strains is the host generally used for cloning and plasmid replication.

Strain	Natural resistance	Pr…	Genotype
DH			*F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-*
		General cloning and storage, blue-wh…	*F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K-m_K+), λ-*
Top10	…tomycin	Ge… …e, …g.	*F- mrcA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara,leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ-*
XL1 B…		…e-white screening, routine cloning.	*endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K-m_K+)*
X… Gold		…oning and propagation of large plasmids, high competency.	*endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB lacI^q ZΔM15 Tn10(Tet^R Amy Cm^R)]*

Endonuclease mutation: improves plasmid yield

Mutations in galactose metabolism pathway: cells cannot grow only on galactose

Lac operon mutations

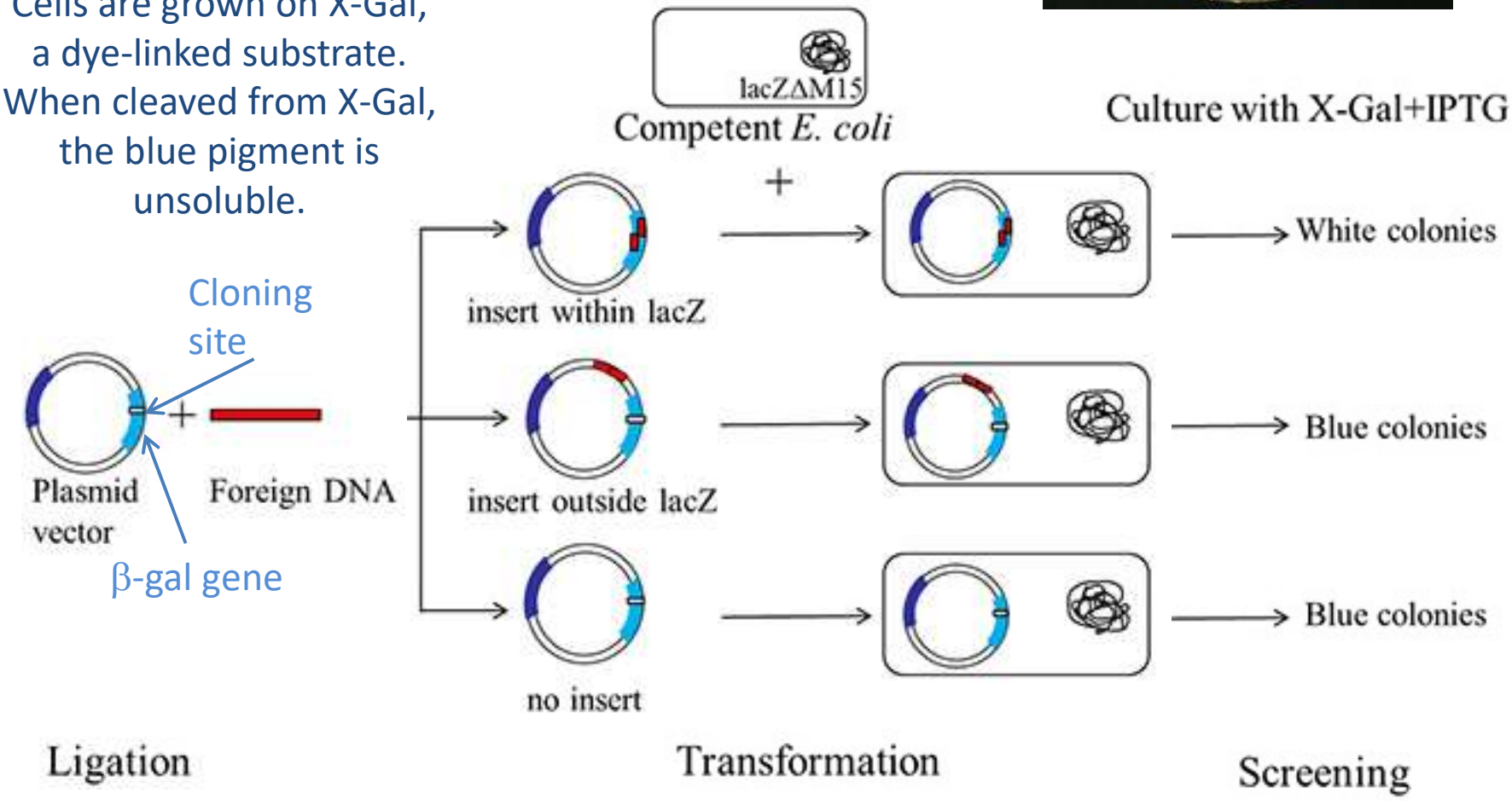
Mutations that reduce plasmid recombination: increased plasmid stability

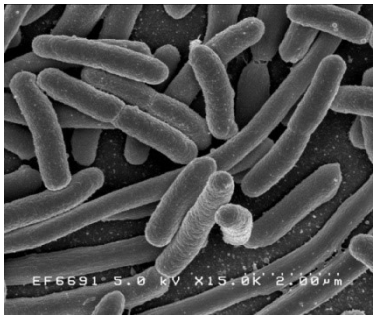
Blue-white screening

To select colonies that contain a plasmid with the insertion of the gene of interest in the correct position.

Uses an *E.Coli* strain deficient for β-galactosidase function (lacZΔM15) and a vector that can complement the β-gal function.

Cells are grown on X-Gal, a dye-linked substrate. When cleaved from X-Gal, the blue pigment is unsoluble.



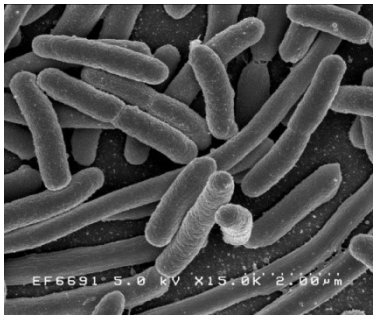


E.Coli hosts

- Exogenous protein expression takes up cell resources
- Some exogenous proteins are toxic/unsoluble/enzymatically active

⇩

- Strains with antibiotic resistance to help selection and reduce competition
- Strains with reduced proteolitic resources
- Expression during exponential growth phase, when cells are alive and healty
- Promoters with tight expression control to reduce basal expression, particularly for toxic products
- Co-expression of chaperones or other factors that assist folding
- Strains with tRNA expression that helps with rare codons



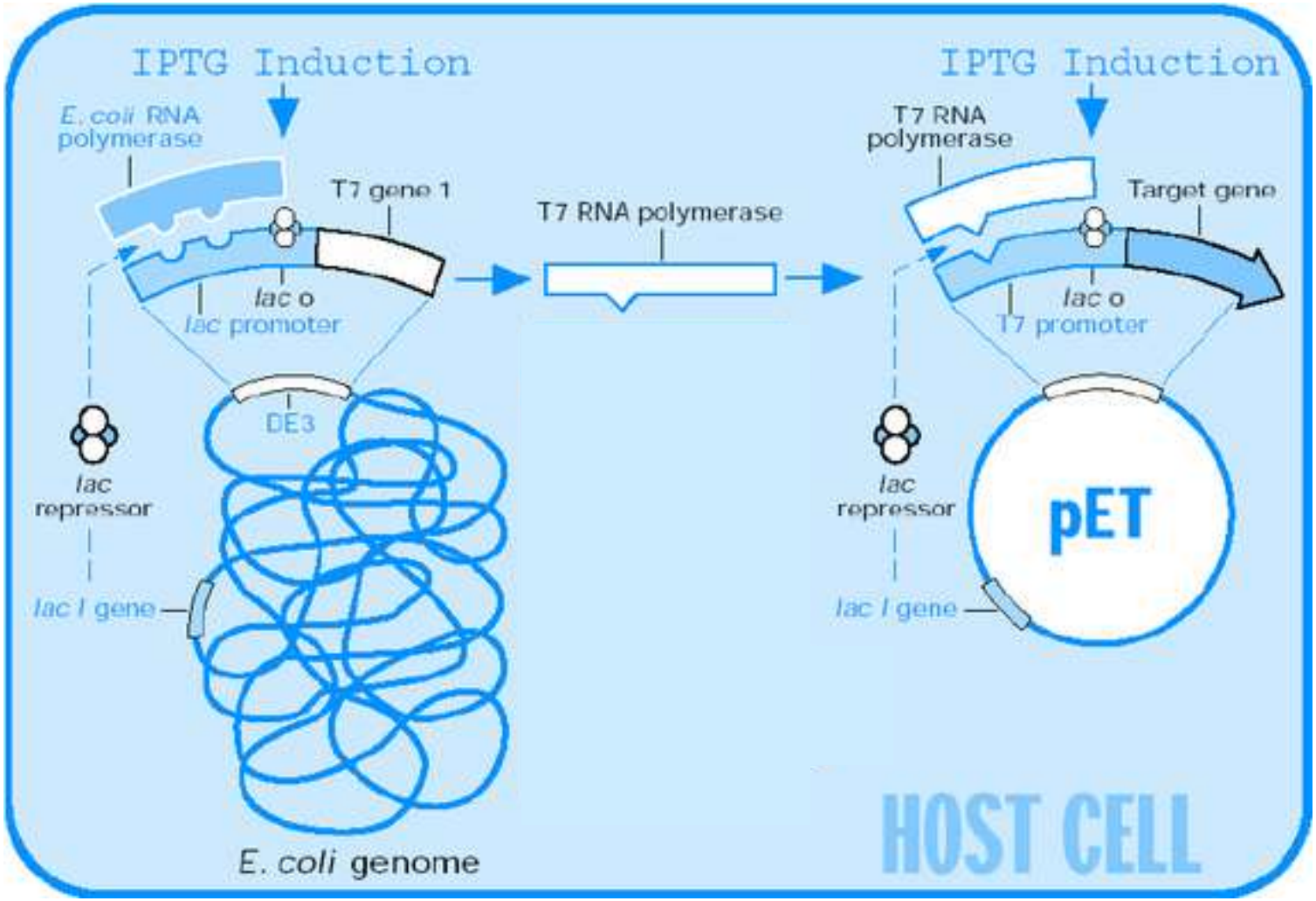
E.Coli strains /2

70% of structures in PDB are from recombinant proteins expressed in *E.Coli*

Strain	Natural resistance	Primary use	Genotype
BL21(DE3)	-	General protein expression.	*F- ompT Ion hsdSB(rB- mB-) gal dcm (DE3)*
BL21(DE3) pLysS	Chloramphenicol	Expression of toxic proteins.	*F- ompT Ion hsdSB(rB- mB-) gal dcm (DE3) pLysS(CamR)*
BL21(DE3) pLysE	Chloramphenicol	Expression of toxic proteins.	*F- ompT Ion hsdSB(rB- mB-) gal dcm (DE3) pLysE(CamR)*
BL21 star (DE3)	-	General expression, not recommended for toxic proteins.	*F- ompT Ion hsdSB(rB- mB-) gal dcm rne131 (DE3)*
Rosetta2 (DE3)	Chloramphenicol	Expression of eukaryotic proteins.	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (CamR)*
Lemo21 (DE3)	Chloramphenicol	Expression of toxic, insoluble or membrane proteins.	*fhuA2 [Ion] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo (CamR)*
Origami2 (DE3)	Streptomycin and tetracyclin	Expression of insoluble proteins.	*Δ(ara,leu)7697 ΔlacX74 ΔphoA Pvull phoR ara D139 ahpC galE galK rpsL F'[lac+ laclq pro] (DE3) gor52::Tn10 trxB (StrR, TetR)*

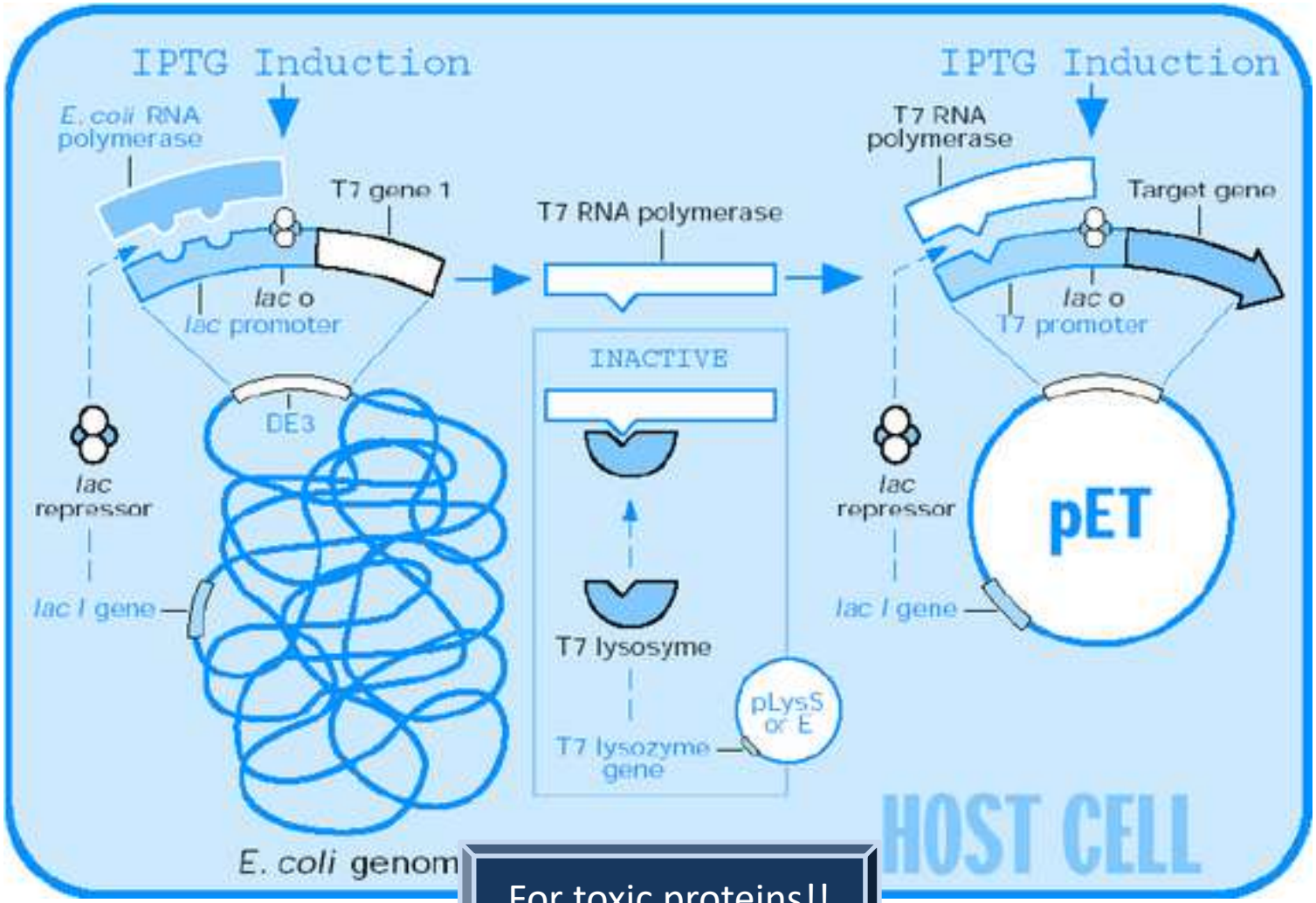
An example of bacterial expression system for toxic proteins:

pET vector in *E.Coli* BL21(DE3) and BL21(DE3)pLysE



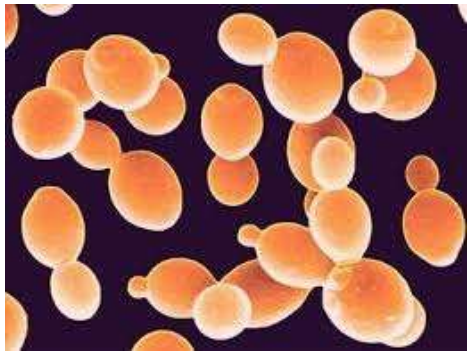
An example of bacterial expression system for toxic proteins:

pET vector in *E.Coli* BL21(DE3) and BL21(DE3)pLysE



For toxic proteins!!

Yeast expression systems



Simplest eukaryotic system: combines advantages of eukaryotic protein expression with ease-of-use and low cost of prokaryotes

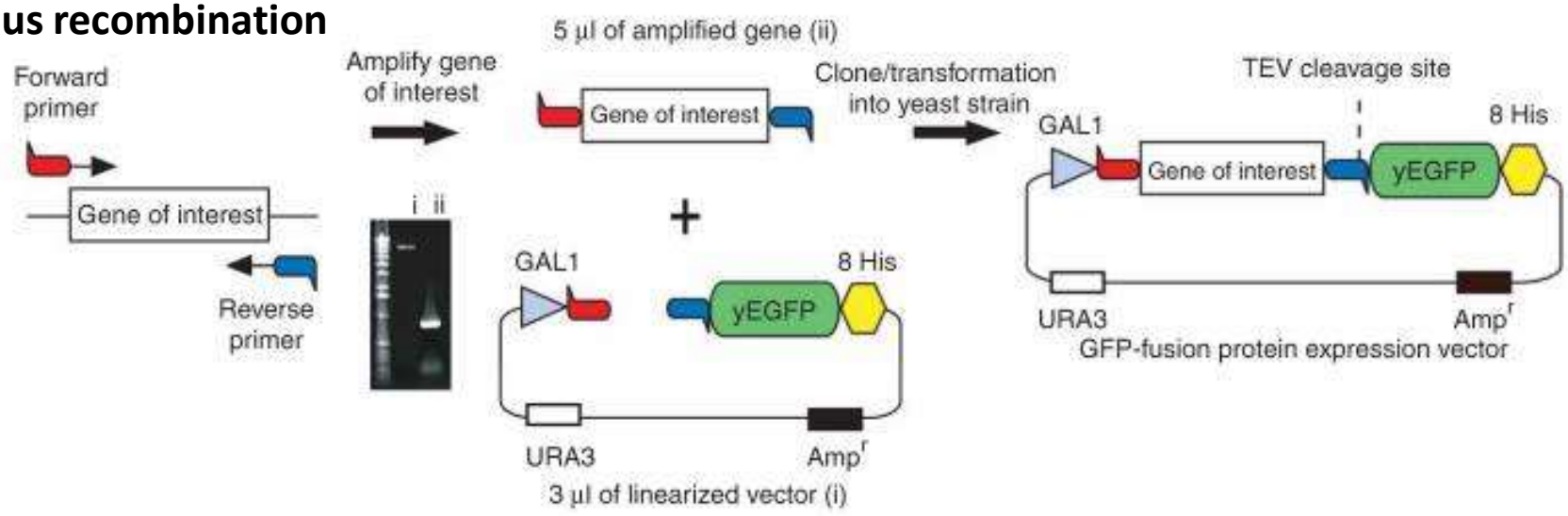
Advantages:

- Rapid growth
- Known genetics
- Versatile DNA transformation system
- Efficient homologous recombination system
- Ease to shuttle between yeast and *E.Coli* during cloning
- Auxotrophic selection
- Galactose inducible vectors available

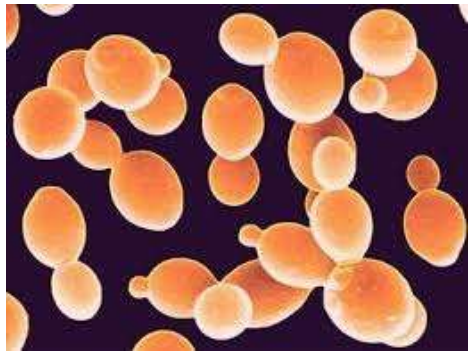
Auxotrophic selection

A strain deficient for the synthesis of a crucial compound (His, Ura, Leu, Lys, Trp, Met) is transformed with a plasmid containing a gene that complements the host's auxotrophy and is grown on media deficient for the compound

Homologous recombination



Yeast expression systems



Yeast species more frequently used: *Saccharomices cerevisiae, Pichia pastoris, Saccharomices pombe, …*

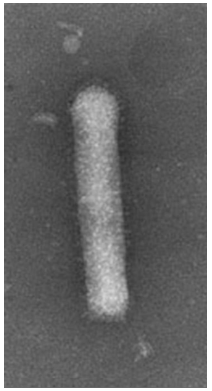
S. cerevisiae:

- Known as budding yeast or baker's yeast
- Easy system to grow; low-cost; well-known genome
- Many mutants or engineered strains available
- **Gal promoter**: galactose inducible system, using glucose as inhibitor
- Easy to control trough auxotrophic selection (with weak promoters before the auxotrophy gene that allows high copy number)
- Eukaryotic post-translational modifications partially available

Pichia pastoris:

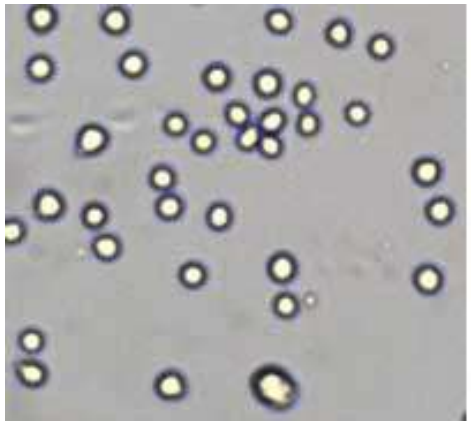
- Can grow to very high cell densities, increasing the amount of protein that can be produced by 1L of media
- **Methylotrophic**: grows in conditions that would kill other microrganisms – less contamination
- **AOX1 promoter**: strong, inducible with methanol and repressed by glucose/glycerol
- Has eukaryotic co- and post- translational systems that allow partial processing of eukaryotic proteins
- Has a lipid composition similar to higher eukaryotes, particularly beneficial for membrane proteins
- Allows integration of exogenous DNA into genome to produce stable clones

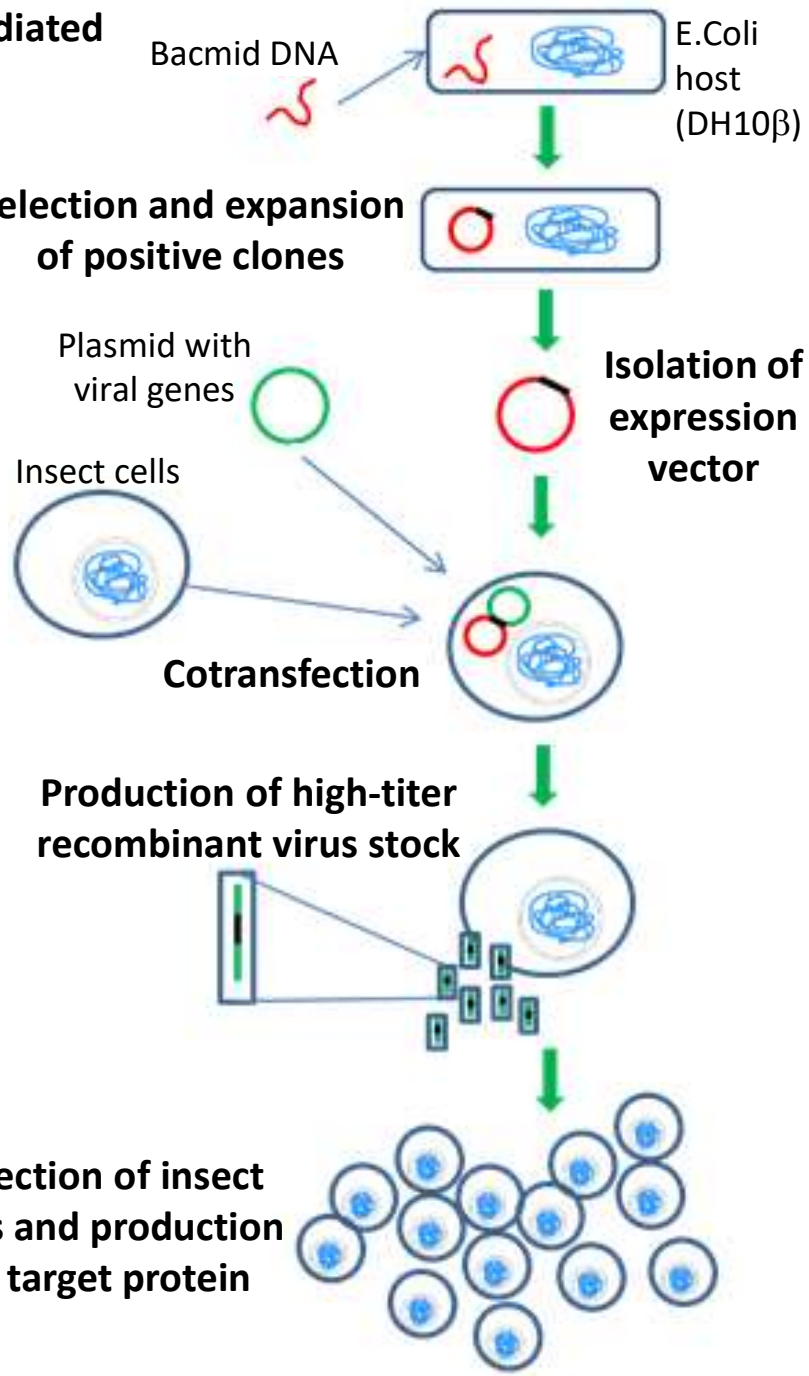
Baculovirus/ Insect cells expression system



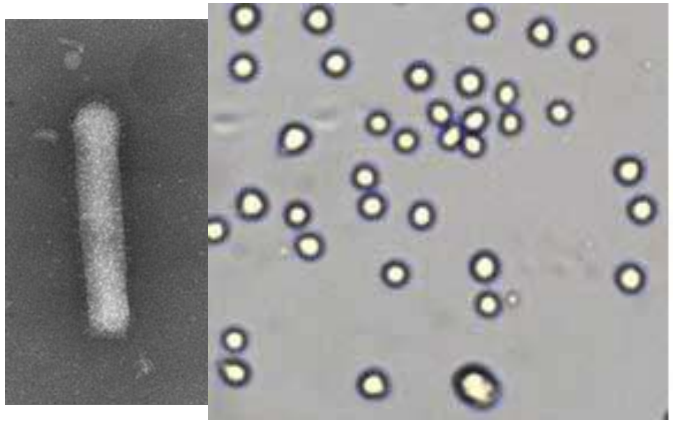
Baculovirus: rod-shaped DNA-viruses that replicate in nuclei of insect cells, natural and specific insect pathogens
Double stranded circular genome ≈80-180 kbp

Insect cells or insect larvae: cells sensitive to baculoviruses and able to express the target protein
Mostly used: sf9, sf21 and High-Five





Baculovirus/Insect cells expression system



Baculoviruses have 4 development phases: immediate-early, delayed early, late, very late. Expression profile changes according to phase. Polyhedrin and P10 proteins, important for virus release, are expressed in late phase.

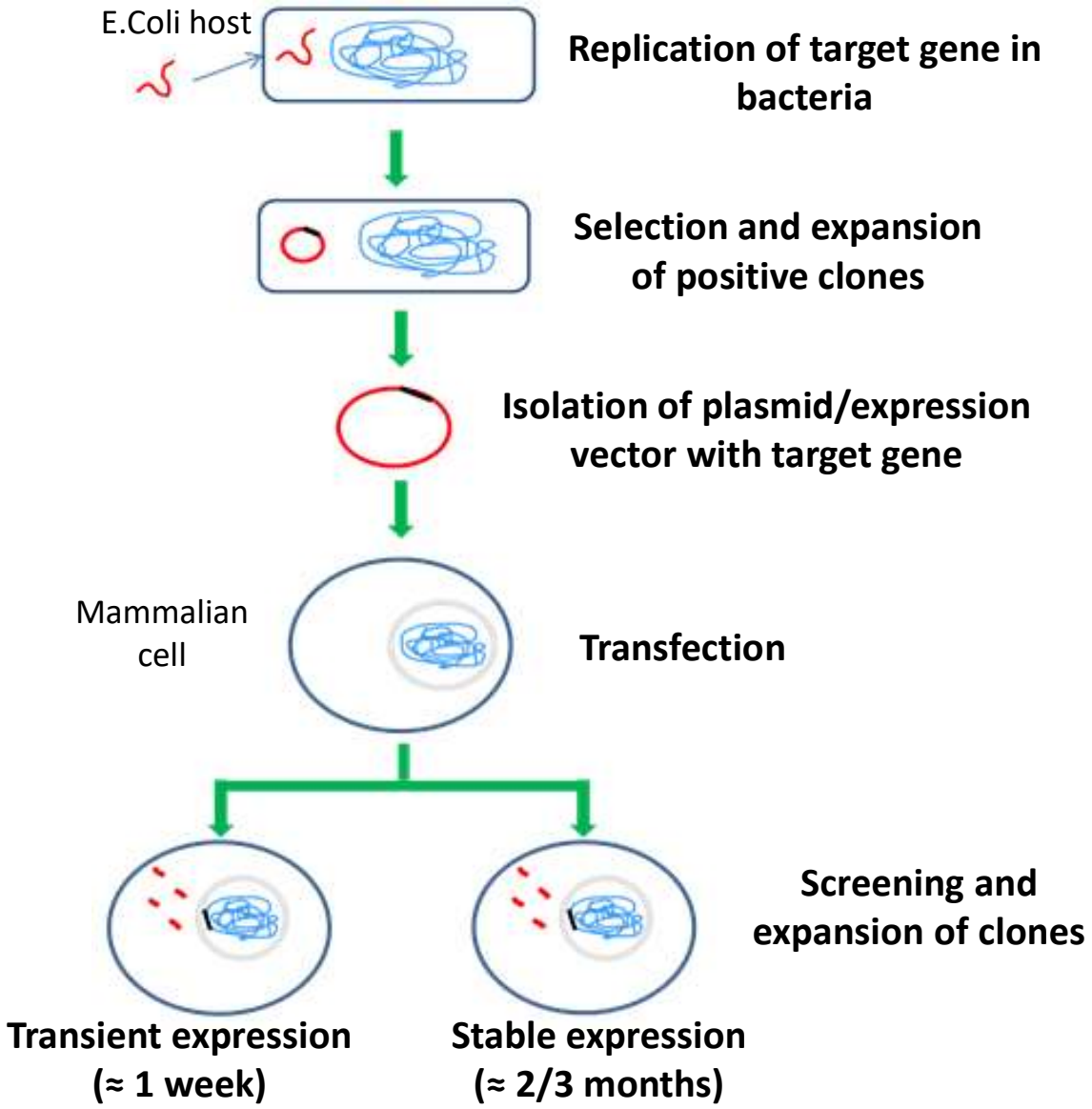
Promoters of Polyhedrin/P10 used to express exogenous proteins with high levels of expression. Expression in late phase particularly beneficial for toxic proteins. For glycosylated proteins, use other promoters.

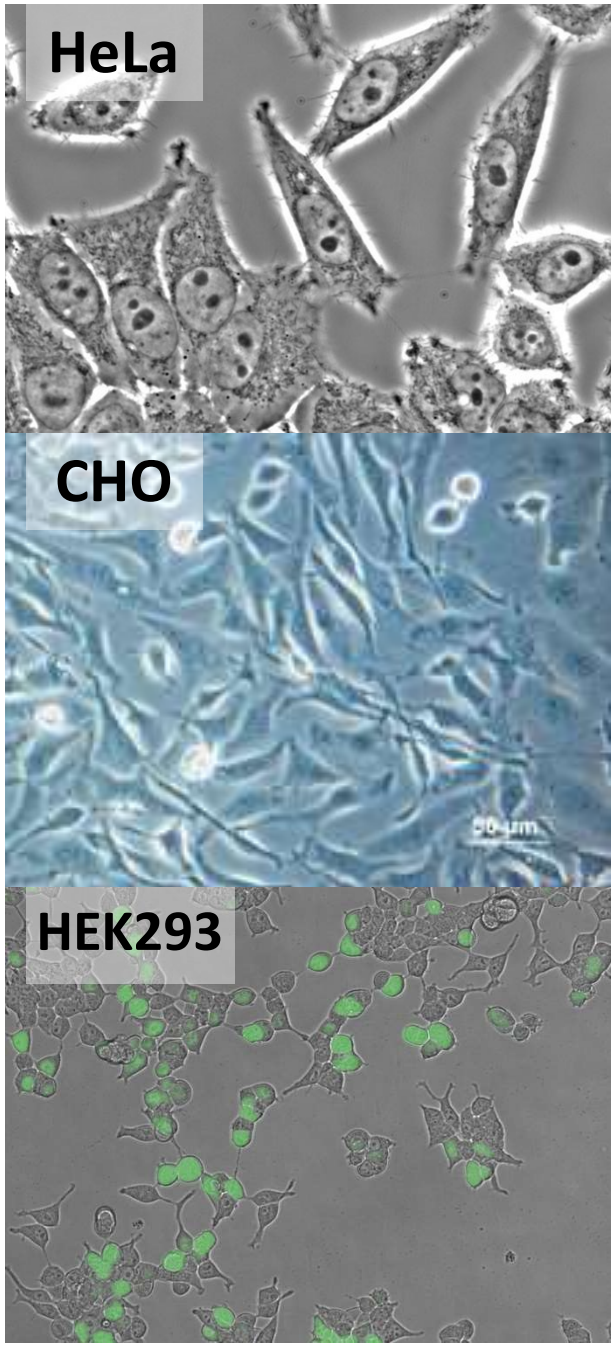
Production of glycosylated proteins. But glycosylation pattern differs from mammalian.

Use of signal peptides to secrete or direct into membrane, and allow proper folding. No splicing.

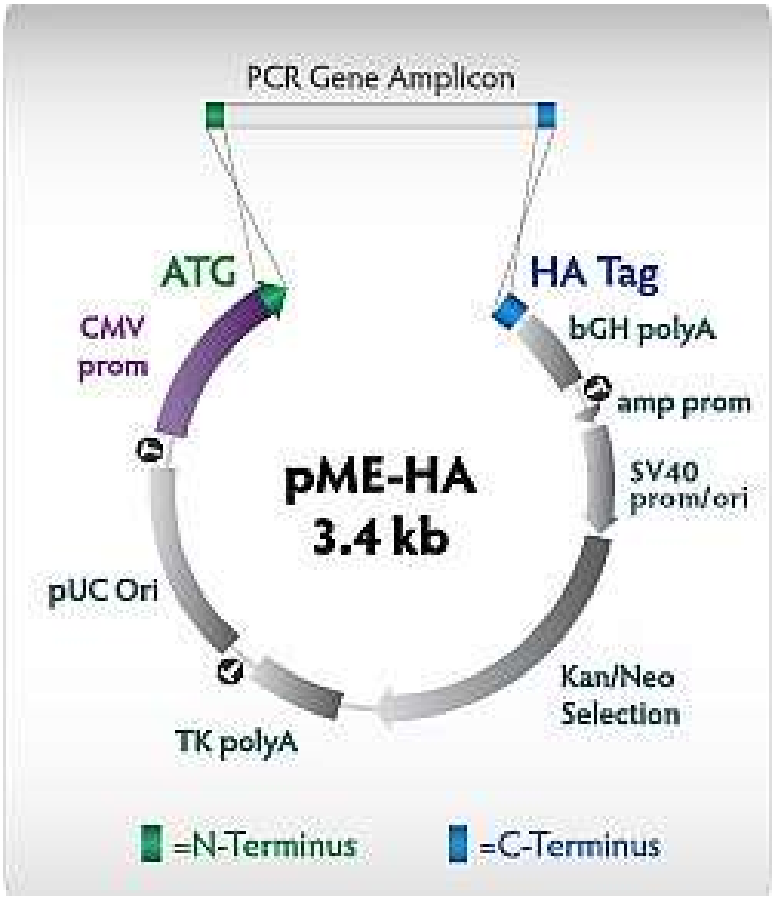
MultiBac: larger vectors that contain multiple protein genes, allows simultaneous expression of more proteins. With MultiBac or cotransfection of multiple bacmides is possible to assemble large eukaryotic complexes.

Mammalian expression systems





Mammalian expression systems



In mammalian vectors there is no ORI for replication, but a viral and/or bacterial ORI can be inserted. Antibiotic selection can be inserted to replicate the gene in bacteria.

First decision: TRANSIENT vs STABLE CELL LINE

For stable cell lines, taget gene is incorporated into the mammalian genome, requires more clone selection (longer time). In transient expression, gene is not incorporated but remains on a plasmid, faster method, but requires re-transfections.

Second decision: CONSTITUTIVE vs INDUCIBLE EXPRESSION

Inducible promoters derived from viruses (e.g. CMV promoter for tetracycline-inducible expression system) are available. Strength of promoter should be chosen according to protein toxicity: strong promoters are not advisable for constitutive expression of toxic proteins

Cell-free expression systems

Used for:

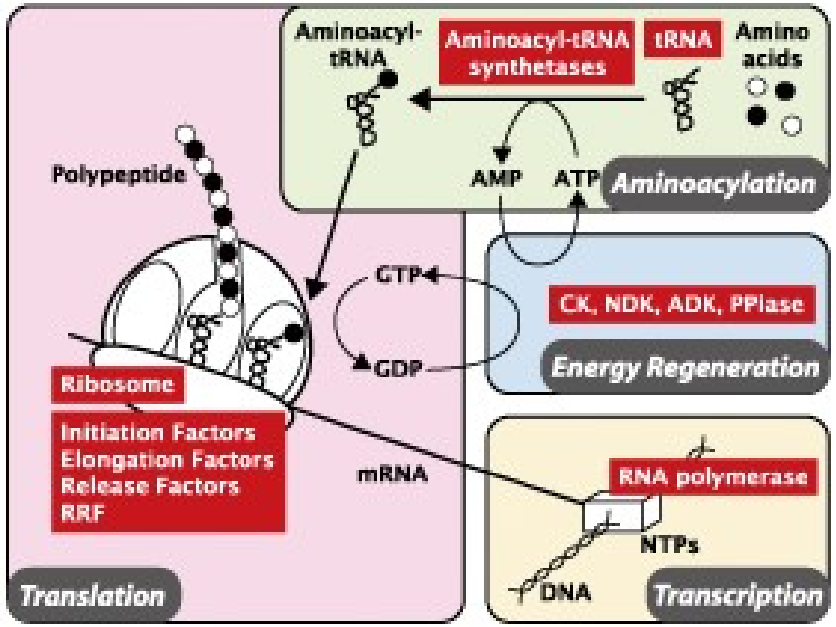
1) Proteins affecting physiology of the host cell
2) NMR studies with isotopic enrichment (and no need for purification!)
3) MAD phasing with Se-Met

- ✓ Ribosomes
- ✓ Initiation/eleongation/termination factors
- ✓ tRNAs
- ✓ Aminoacids
- ✓ Aminoacyl-tRNA synthases
- ✓ ATP, GTP and other energy sources
- ✓ Enzymatic cofactors
- ✓ Cellular components essential for protein folding (chaperones etc.)
- ✓ RNA polymerase

Template: **mRNA** or **DNA**

CELL EXTRACT transcription and translation molecular machineries

Protein



Components isolated from *E.coli*

Cell-free expression systems

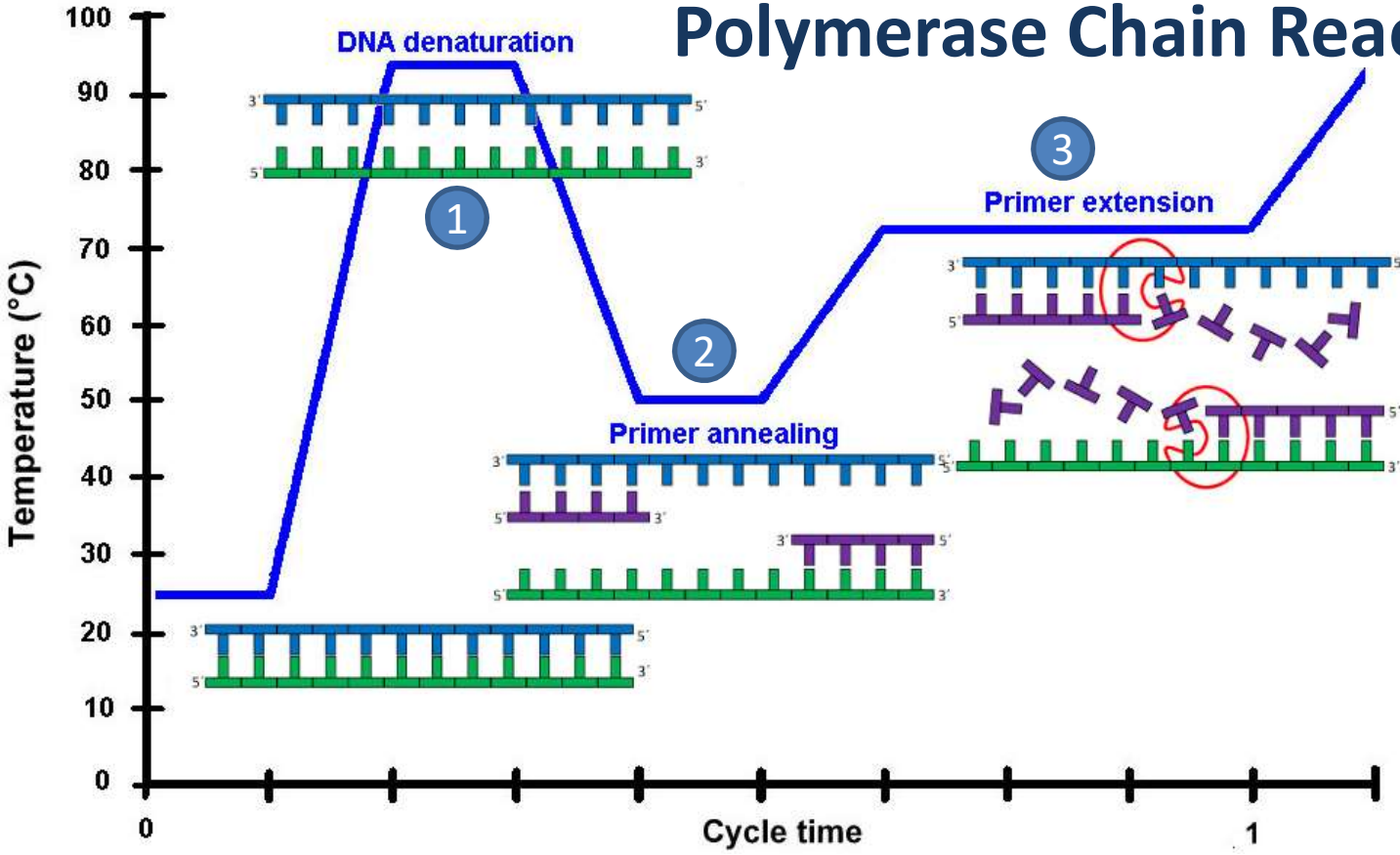
- **E.Coli**
  - ✓ High yield and easy. Low cost. Can express membrane proteins.
  - ✗ Limitation of prokaryotic systems. Post-translational folding (solubility problems).
- **Rabbit Reticulocyte**
  - ✓ Mammalian system. Co-translational folding. Can express membrane proteins.
  - ✗ No protein glycosylation. Low yield. Expensive.
- **Wheat germ**
  - ✓ High yield. Low cost. Can be stored dried (1.5 yrs). Co-translational folding. Suitable for large proteins.
  - ✗ No mammalian post-translational modifications. No membrane proteins.
- **Insect**
  - ✓ Suitable for large proteins. Partial glycosylation.
  - ✗ Not complete mammalian post-translational modifications.
- **Human**
  - ✓ Optimal co- and post-translational modifications for human proteins. Co-translational folding. Production of completely functional proteins.
  - ✗ Low yield. Expensive.

Cloning techniques

Used to **“cut-and-paste”** DNA of target gene into desired vector

- Use of restriction enzymes and ligases
- Use of homologous recombination
- Ligase-independent techniques
- Restriction-free techniques

Polymerase Chain Reaction (PCR)

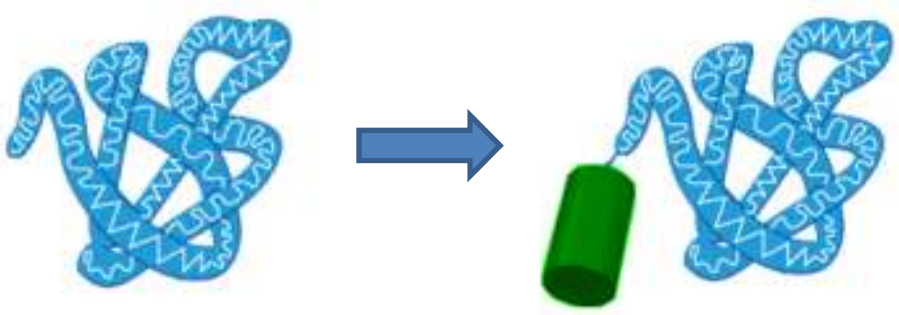


40-50 Cycles:

1) Denaturation of the DNA, from double helix to single strands (T=95°C)
2) Primer annealing to single strand DNA (T=50-65°C)
3) Primer extension by thermostable polymerase, i.e. *Taq* (T=72°C)

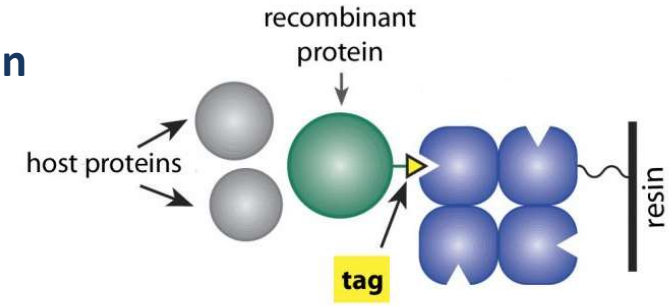
Construct design

1. Insertion of tags



Used for:

- ✓ **Purification**



- ✓ **Stability vs proteases**
- ✓ **Solubility and folding**
- ✓ **Reporting (Western Blot and fluorescent proteins)**

Typical tags for protein expression:

His-tag : 6-10 **histidine tail** that binds to **Ni^{2+}** or **Co^{2+}**; does not affect folding, used for affinity purification and blotting

MBP-tag : Maltose Binding Protein, increases solubility of proteins and improves folding, used for affinity purification

SUMO-tag : Small Ubiquitin-related Modifier, promotes folding and stability

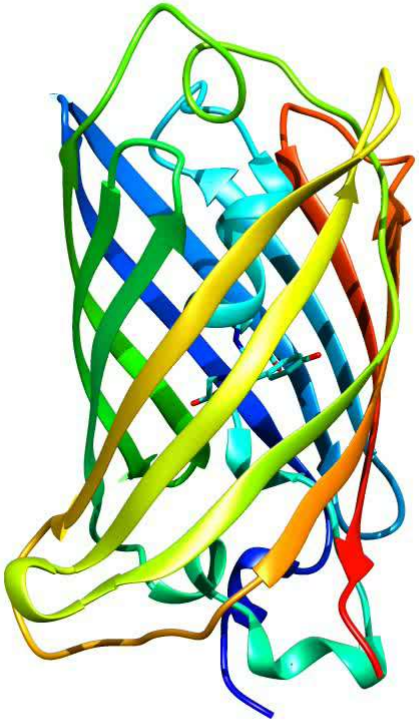
Strep-tag : Streptavidin, a protein that has high affinity for **biotin**, used for affinity purification

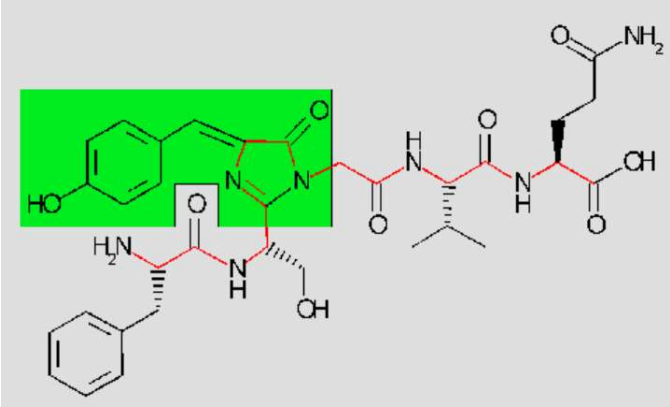
GST-tag : Glutathione S-Transferase that binds **glutathione** with high affinity and increases protein stability to proteases, used for affinity purification

FLAG-tag : Octapeptide that binds strongly to commercially available **antibodies**

GFP-tag : Green Fluorescent Protein tag allows to test expression

Green Fluorescent Protein





Derived from *Aquorea victoria*, a jellyfish.

Does not need cofactors, substrates or enzymatic activity to display fluorescence, but requires O_2 during maturation.

β-barrel of 11 β-sheets; fluorescence depends on **oxidation** of **Ser-Tyr-Gly** in the center of the barrel.

Easy to express as exogenous protein, non-toxic.

Antibodies against GFP are available. Used also for *in-vivo* expression studies.

MW: ≈ 27k Da, 238 aa

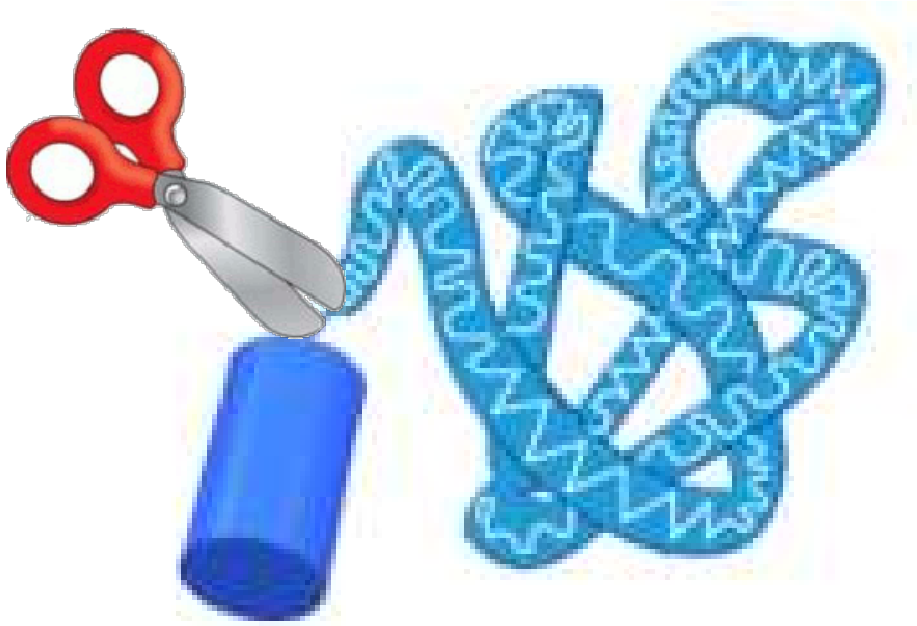
Fluorescence: λ_{ex} = 395-488 nm, λ_{em} = 499-519 nm (different for mutants, e.g. YFP, CFP, BFP, and other fluorescent proteins, e.g. RFP, mCherry)

Construct design

2. Introduction of cleavage sites

Tags are useful for purification, but may hamper crystallization due to flexibility and heterogeneity of conformations.

Proteases are used to cleave the tag without affecting the protein conformation and purity. The ideal protease is **not present in the host organism**, has a **specific recognition sequence**, and such sequence is **not present in the protein sequence**.



Specific protein sequence to be recognized by a protease enzyme

Thrombin

cleaves the sequence: LVPR/GS

TEV-protease

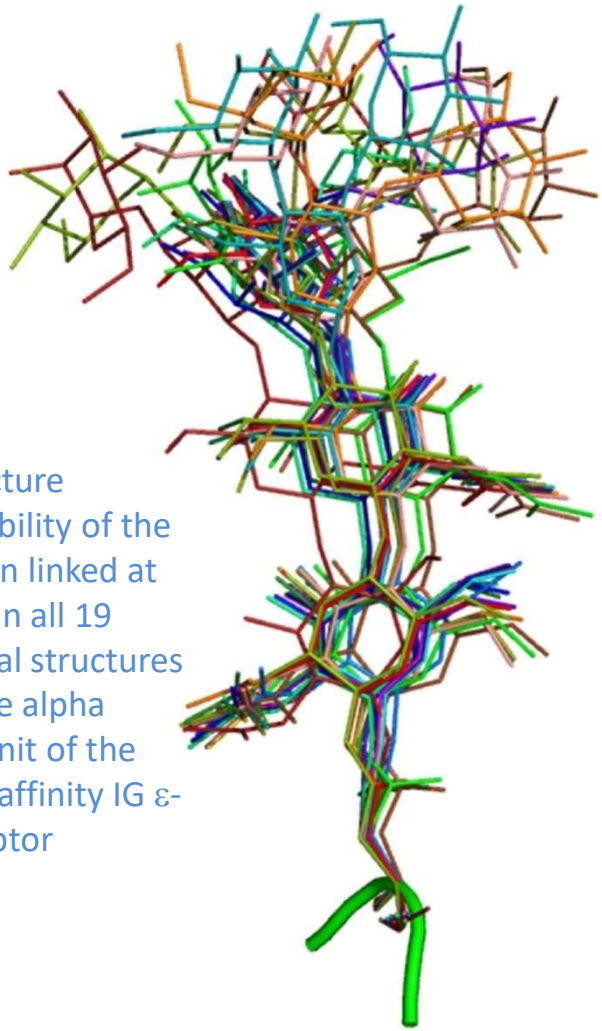
cleaves the sequence: ENLYFQ/G

PreScission protease

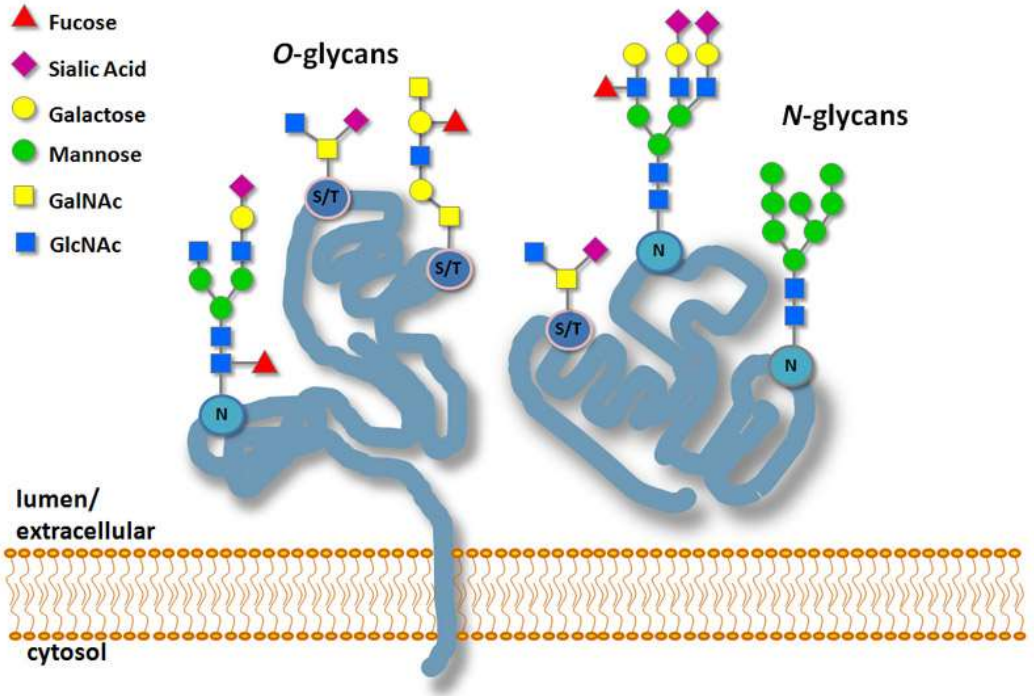
Rhinovirus 3C Protease, cleaves the sequence: LEVLFQ/GP

Construct design

3. Removal of glycosylation sites



Structure variability of the glycan linked at N42 in all 19 crystal structures of the alpha subunit of the high affinity IG ε-receptor



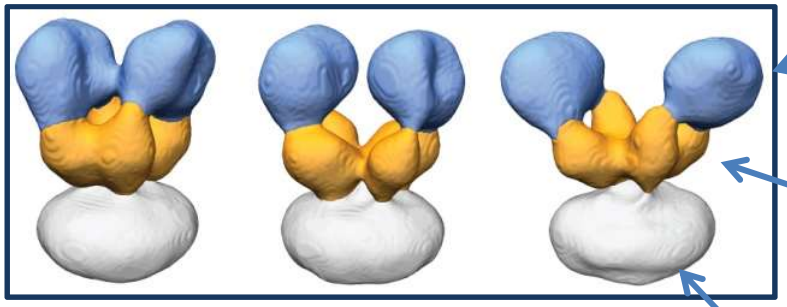
Mutation of residues involved in post-translational modification enhances homogeneity of the sample

An alternative strategy is post-expression enzymatic deglycosylation

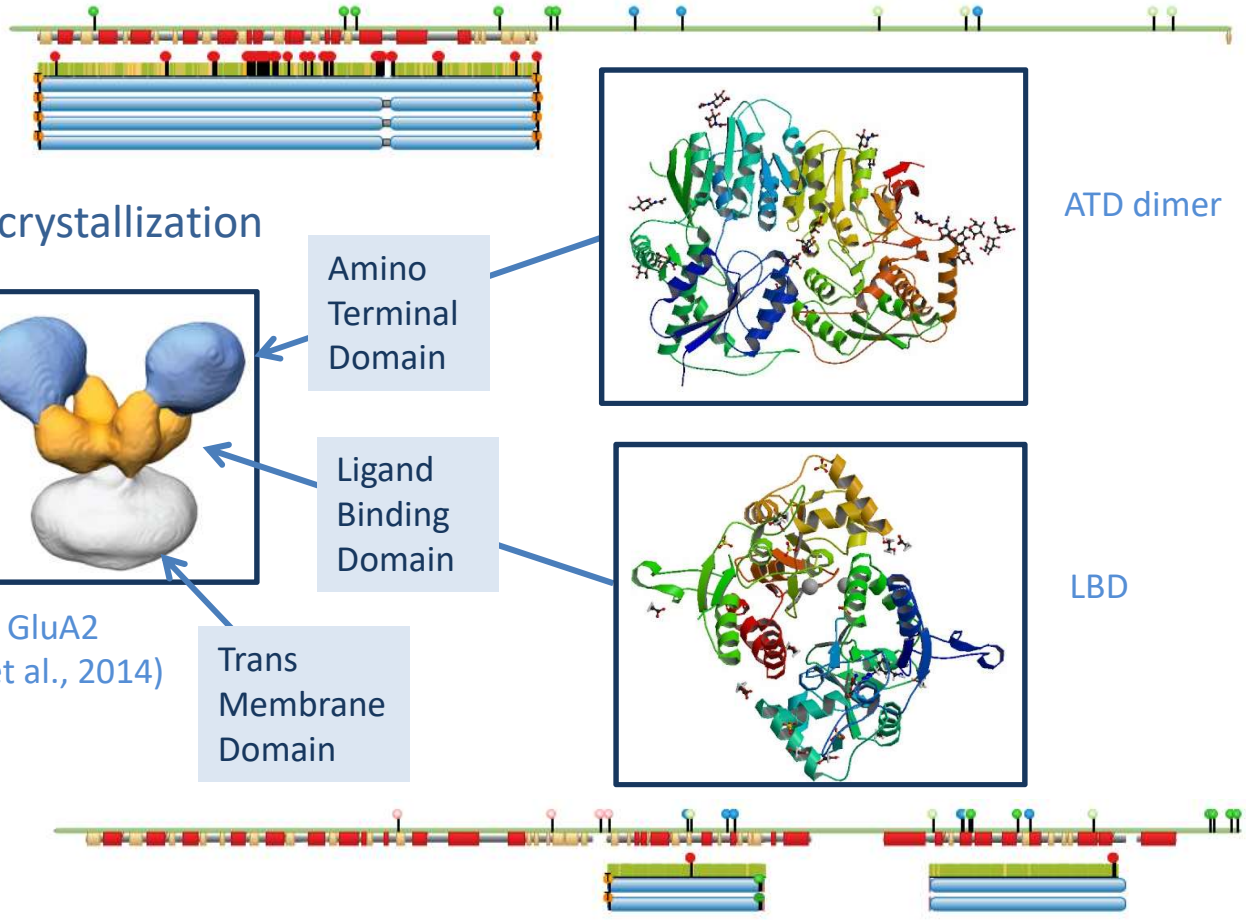
Construct design

4. Removal of flexible domains or membrane domains

The presence of flexible domains increases heterogeneity of the protein sample and hampers crystallization



Three distinct desensitized state GluA2 structures from EM (Meyerson et al., 2014)



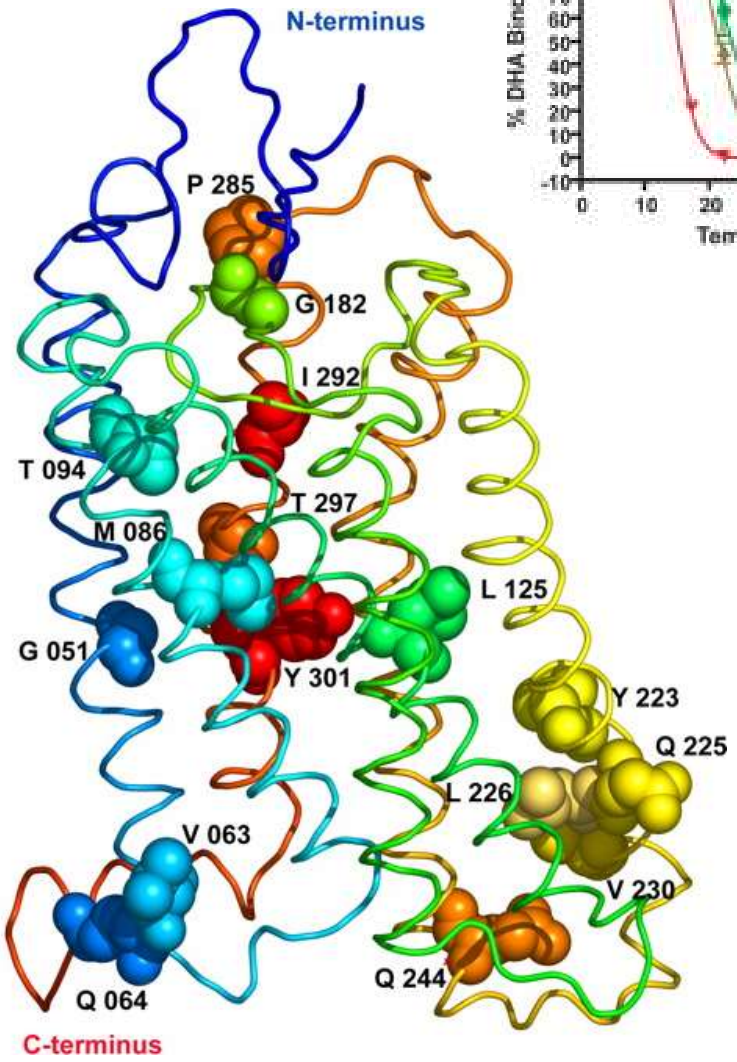
Also:
removal of flexible C-term or N-term sequences, shortening of linkers

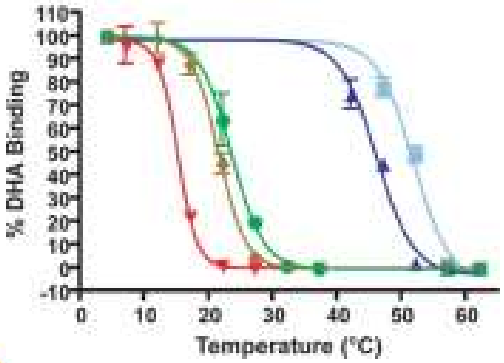
Construct design

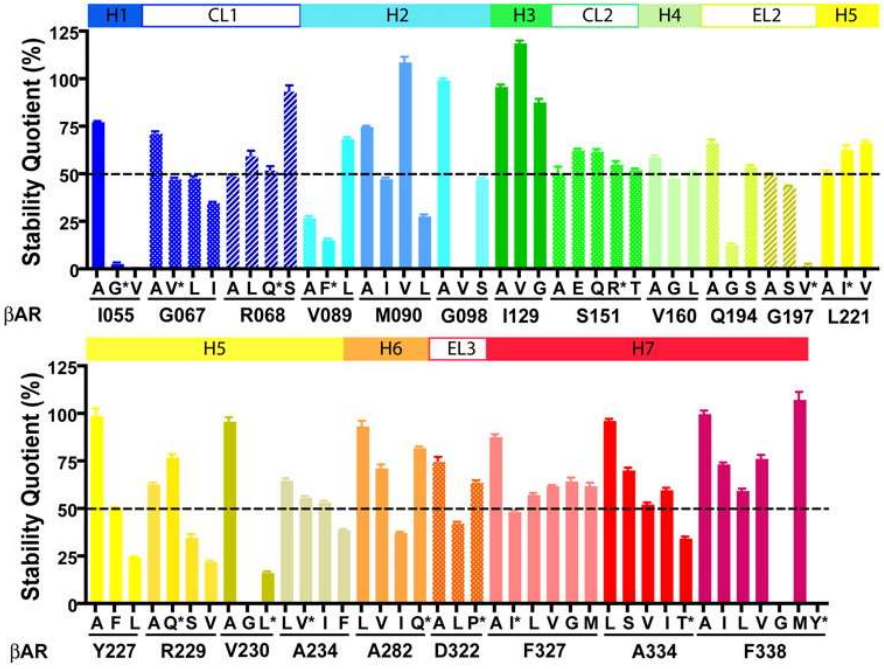
5. Thermostabilization

Mutation of residues to improve stability of the protein

30min @32°C and
Radioligand binding assay



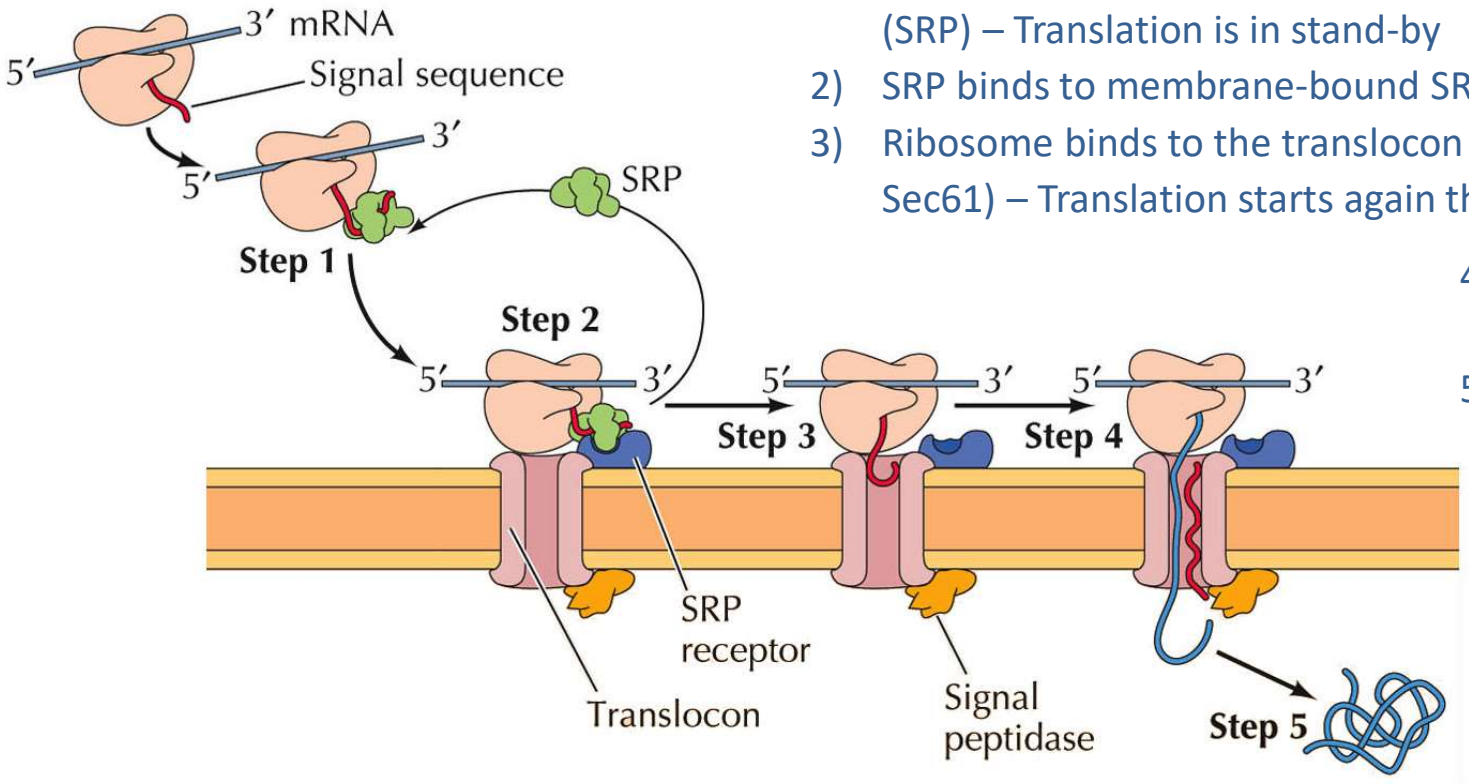




Construct design

6. Introduction of signal peptide

For secreted proteins or integral membrane proteins, 5-30 residues sequence at the N-term that direct the newly synthesized protein to the plasma membrane (prokaryotes) or to the endoplasmic reticulum (eukaryotes)



1) Signal peptide recognized by Signal Recognition Particle (SRP) – Translation is in stand-by
2) SRP binds to membrane-bound SRP receptor
3) Ribosome binds to the translocon complex (SecYEG or Sec61) – Translation starts again through the channel
4) Signal peptide is removed
5) Protein secreted

Construct design

7. Removal of phosphorylation sites

Unlike glycosylation, phosphorylation doesn't increase degrees of freedom, but it might increase heterogeneity in the sample. Heterogeneity is an issue for crystallization, but also for purification.

8. Cross-linking – Introduction of Cysteine residues

Mutation of selected residues to Cys to allow the formation of inter/intramolecular disulfide bonds. Disulfide bonds increase rigidity of the overall structure.

9. Co-expression of proteins for binding of prosthetic groups

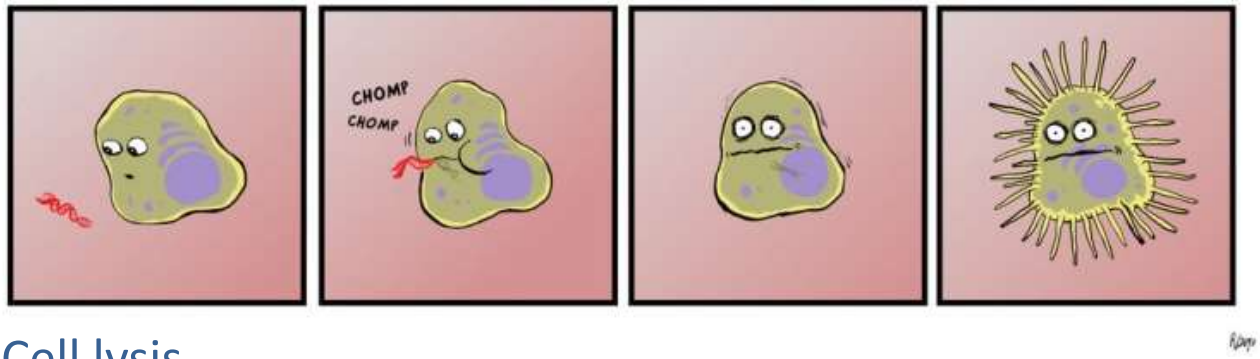
If a prosthetic group is present in the target protein, the co-expression of proteins involved in the synthesis and binding of the prosthetic group might be required for correct folding.

10. Co-expression of chaperones

Chaperones are proteins that *in vivo* assist folding of other proteins. Recombinant expression of proteins may require co-expression of chaperones if (1) not enough chaperones are expressed or (2) the host system does not posses suitable chaperones.

Expression protocol in bacteria

Plasmid design and preparation → Transformation → Colony selection



Colony selection → Culture growth

- Medium
- Temperature
- OD of induction
- Antibiotics
- Additives

Culture growth → Expression induction

- Temperature
- Concentration of inducing agent
- Length of induction

Expression induction → Cell lysis (Sonication or French press)

For the expression of large amounts of proteins, expression conditions need to be carefully optimized.

Test expression before scale up!

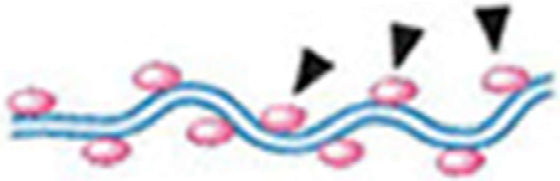
Cost of expression has to be taken into account, too!!

Expression in inclusion bodies

INCLUSION BODIES: dense aggregates of misfolded protein present in the cytoplasm, formed when protein is unsoluble or due to the lack of folding machinery within the host cell. Formed in *E.coli*, but also in yeast, insect cells or mammalian cells.

1. **Isolation of inclusion bodies**: mechanical disruption of cell and removal of soluble fractions (centrifugation, filtration or sucrose gradient separation)

2. **Removal of contaminants**: improves refolding and removes proteases; detergent, sucrose or urea (< 4M) wash



3. **Solubilization**: using chaotropic agents (urea or guanidinium chloride, 4-8 M). Usually, reducing agents are added to prevent wrong disulfide bond formation

4. **REFOLDING**

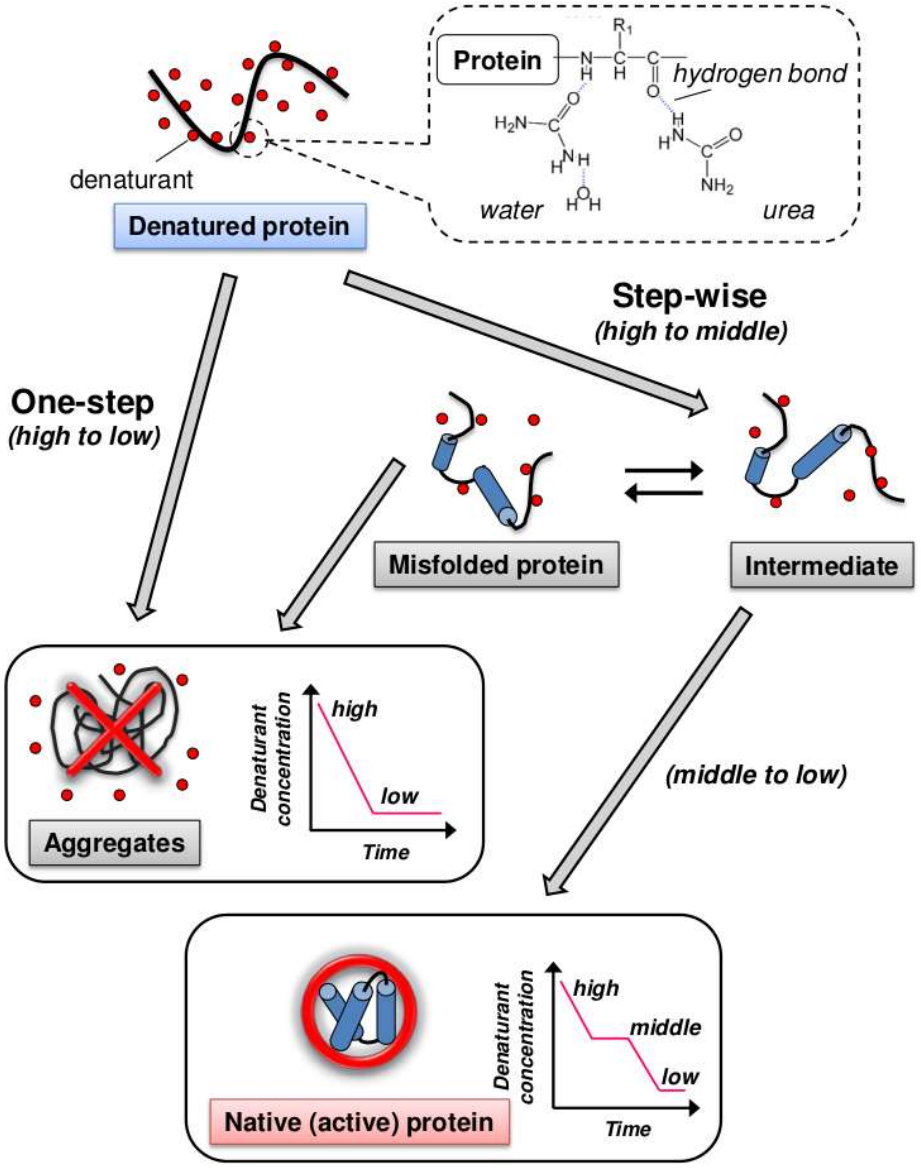
Advantages:

- Large amount of protein
- Lower contamination and higher protein purity

Disadvantages:

- Refolding not always successful
- Requires assessment of native conformation

Refolding



KINETICS:
Slow chaotropic agent removal allows protein to fold in native conformation

THERMODYNAMICS:
Stabilization of the native conformation increases yield of protein with correct folding

Different methods for refolding:

- Dialysis for a slow urea removal
- Dilution of urea
- Refolding on column

Additives:

- Reducing agents
- Stabilizing molecules (e.g. Arginine, glycerol, salts, PEG…)
- Substrates, ligands to aid refolding

References

- Expression systems and vectors: Addgene, “Plasmids 101: A Desktop Resource”, September 2023 (4th Edition), **www.addgene.org**
- Yeast expression systems: Drew D. *et al.*, “GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*”, **Nat. Protoc. 2008**, *3(5)*:784–798; Parker J.L. & Newstead S., “Method to increase the yield of eukaryotic membrane protein expression in *Saccharomyces cerevisiae* for structural and functional studies.”, **Protein Sci. 2014**, *23*:1309-14; Byrne B., “*Pichia pastoris* as an expression host for membrane protein structural biology.”, **Curr. Opin. Struct. Biol. 2015**, *32*:9-17.
- Baculovirus expression systems: van Oers M.M., “Opportunities and challenges for the baculovirus expression system”, **J. Invertebr. Pathol. 2011**, *107*:S3-15; Barford D. *et al.*, “Baculovirus expression: tackling the complexity challenge”, **Curr. Opin. Struct. Biol. 2013**, *23(3)*:357-64.
- Mammalian expression systems: Andréll J. & Tate C.G., “Overexpression of membrane proteins in mammalian cells for structural studies”, **Mol. Membr. Biol. 2013**, *30(1)*:52-63.
- Cell-free expression systems: **Thermofisher website** https://www.thermofisher.com/it/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/cell-free-protein-expression.html; Endo Y. & Sawasaki T., “Cell-free expression systems for eukaryotic protein production”, **Curr. Opin. Biotech. 2006**, *17*:373-380.
- GFP crystal structure: Ormö M. *et al.*, “Crystal structure of the Aequorea victoria green fluorescent protein”, **Science 1996**, *273(5280)*:1392-5.
- Thermostabilization: Serrano-Vega M.J. *et al.*, *“*Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form.*”* **PNAS 2008**, 105(3):877-82.
- Refolding: Middelberg A.P.J., “Preparative protein refolding”, **Trends in Biotech. 2002**, *20(10)*:437-443; Palmer I. & Wingfield P.T., “Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from *Escherichia coli*”, **Curr. Protoc. Protein Sci. 2012**, Chapter 6, Unit 6.3