Protein Cloning and Expression

Why We Need Cloning?

- Exploit the expression capabilities for large scale production.
- Work with the intended protein directly, instead of its orthologs.
- Highly customizable property make it possible to add tags, purify the protein easily.

- Cost-effective.
- No risk of contamination by pathogens present in source material.

Research on the application of protein itself:

- Proteins as therapeutics in their own right.
- Diagnostic molecules.

Proteins as targets:

- Investigation in the involvement in a disease of interest

Process

Main procedures include:

- Isolate gene
- Clone the gene into a vector
- Transfer
- Extract and purify the protein

Isolate the Gene of Interest

Methods to Isolate genes

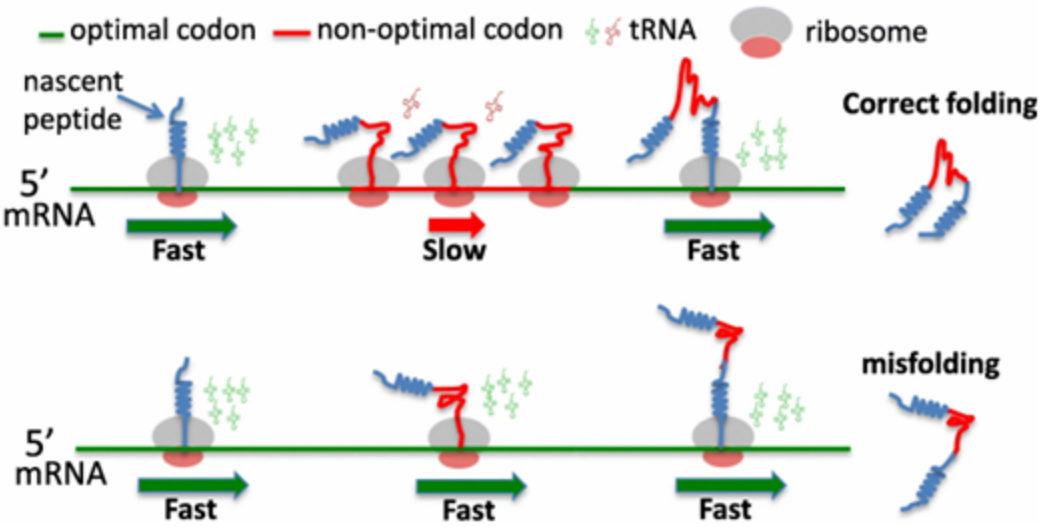
There are two methods to isolate gene from organisms: using cDNA libraries and using genomic DNA libraries.

cDNA library refers to the gene sourced from reverse transcription of mRNA. A very straightforward feature is obtained genes do not have introns and reflect what is transcribed.

Genomic DNA library is another situation: introns are included because genes obtained through this method come from chromosome. Meanwhile, regulatory genomic regions of DNA, along with other non-coding regions are also reserved.

Codon Optimisation

One thing to note is that **tRNA abundance discrepancies**[1] are often observed between the gene source organisms and the expression hosts. This difference further changes the translation rate directly, and may lead to protein misfolding or protein aggreation[2]. Thus, **codon optimisation** is very necessary.



Other factors, such as the possible containing restriction enzyme cutting site in inserted gene, are also the reasons why a codon optimisation is needed.

Clone the gene into a vector

Vector is a circular autonomously replicating DNA element. The most often used vector is the E.coli plasmid.

Components in vectors

There are two kinds of vectors: Expression vector and Cloning vector. The two kinds of vectors often do not co-exist in the same colony, unless extra complex procedures are executed. Differences between expression vectors and cloning vectors are shown below.

Features	Cloning Vector	Expression Vector
Function	Used to cloning, and reserving recombinant DNA	Expressing proteins
Promoters	Weak or None	Strong and controllable[3]. Commonly used promoters include T7 or lac promoter[4]
Tags	N/A	Owing tags such as soluble tag[5] and TEV tag[6]

Common components in vectors include:

- For reproduction: Origin of replication
- For transcription: transcription promoter/ terminator site and ribsome binding site (**Only exists in prokaryotic cells**)
- For translation: start/ stop codon
- Antibiotic resistance gene
- Cloning site
- Tags, expression promoters, etc. (**Only exist in expression vectors**)

Antibiotic marker

Add an antibiotic marker to force the cell to maintain the inserted DNA element.

Choose different antibiotics

Some antibiotics are more strict in terms of selective pressure they put cells under. Use these antibiotics in case of expressing host unfavorable genes. Commonly used antibiotics are ampicillin, kanamycin, chloramphenicol.

Maintain more than one plasmid at one time.

Multiple cloning site (MCS)

MCS is a region in cloning vector which does not exist in the wild type E.coli. Rather, it is engineered by human and **contains numerouse specific DNA sequences recognised by restriction enzymes**. Meanwhile, the original restriction enzyme system in E.coli is destructed. So that it will not degrade the inserted region.

There is a strategy to identify whether a gene sequence has been correctly inserted into MCS called 'blue white selection'.

```
MCS before recombinant -express-> One kind of protein -add specific substrate-> B

MCS with inserted sequence -> Fail to express that kind of protein -> No colour c
```

Restriction enzymes

Restriction enzymes are an array of enzymes recognising specific DNA sequences and hydrolysing the phosphodiester bonds. Various types of cutting ends, such as blunt end, 3'/5' overhang, single-strand overhangs, can be attained after the process of restriction enzymes.

Other cloning strategies

Ligation-independent cloning
Non-enzymatic approaches

Problems may occur

- Protein misfolding/aggregation
- Inclusion body formation
- Loss of production because of toxicity
- Fail to form disulfide bond
- Absence of post-translation modification
- Presence of rare codons
- Product being degraded by proteases

Solution: Expression environment modification

Modify E.coli genes to:

- Create an oxidising environment
- Express cold resist chaperon and prepare for low temperature expression
- Produce tRNA of rare codons
- have mutations in transporters to allow titratable induction of expression

Decorating expressed proteins with signal peptides, which directe peptides to ER or plasma environment

Solution: Modify the protein

- Add soluble tags to prevent aggregation
- Introduce a protease cleavage site can be engineered to remove the tag before/after purification

Solution: Alter experiment conditions

- **Temperature**: Decreasing the expression temperature (low temperature expression) to slow down the expression speed
- **Timing**: Induce expression later than usual (avoid expressing in the log phase)

Solution: Express in a different host

Construct in prokaryotic cell, express in eukaryotic cell

Construct and reproduction:

- In E.coli
- Vector contains origin of replication and selectable markers (antibiotic resistant gene)
- Muted expression promotors

Expression:

- In eukaryotic cell (mammalian cell, yeast cell, and insect cell)
- Send the plasmids into expression cells through transfection
- Activate eukaryotic promoter. Initiate transcription and translation
- Constitutive or inducive

Express in mammalian cell:

Advantages	Disadvantages
Correct glycosylation	Time consuming
Chaperones assist in proper folding	Cell culture can only be performed for a finite amount of time
Cells secret proteins into medium	Costly

Miscellaneous

Express gene at hosts' log phase, unless the expressed protein is toxic.

There are also some kinds of proteins hard to recombinant in E.coli such as *transmemberane proteins*.

Transfer into a host

Transformation is the process by which foreign DNAs are introduced in bacteria cell. This process can be achieved through **chemical** (osmotic pressure) or **electroporation**.

Transient transfection or stable transfection

Transient transfection: do not use an introducer to integrate DNA elements into the host DNA. As cells reproduce, introduced genes will be degraded and 'vanishing'.

Stable transfection: Persistence the transfer of genes through integration into host's chromasome.

Choose an expression system basing on requirements

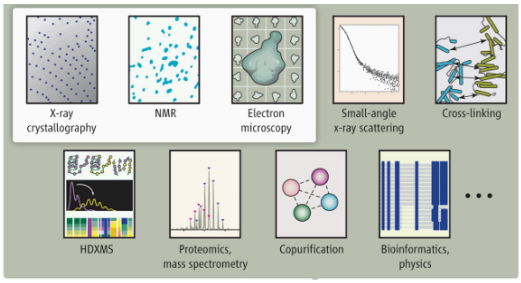
	E. coli	Yeast	Mammalian	In vitro
Cost of growth medium	Low	Low	High	High
Product yield/volume	High	High	Low	High
Folded, functional mammalian protein	Sometimes	Often	Yes	Often
Glycosylation	No	Yes, but different	Full	Yes if mammalian
Time from cloning to expression	Fast	Medium	Slow	Fast
Ease of growth/ avoid contamination	Easy	Easy	Difficult	n/a
Location of recombinant product	Intracellular cytosol/periplasm	Intracellular or extracellular	Extracellular	Solution

In summary, from E.coli to yeast to mammalian cell, the cost and time consumed on culture increases, while biological functions are better reserved.

Problems to concern

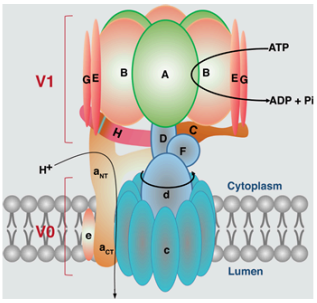
Some things to consider: application

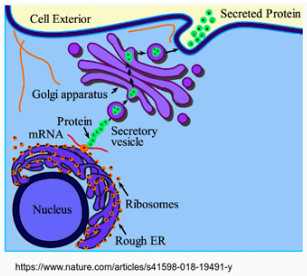
- What will the expressed protein be used for?
- What yield is required?
  - Crystallography and NMR need large quantities
  - Biochemistry and biophysics medium
  - Cryo-EM requirements can be minimal
- Is function important?
  - Usually: yes
  - As an antigen – possibly not



Some things to consider: target characteristics

- Architecture
  - Does the protein have a transmembrane region?
  - Is the protein multidomain?
  - Is it sufficient to express a fragment?
- Construct
  - Any additions likely to affect function?
  - Any detection/purification tags to be removed?
- Compatibility with expression system
  - Toxicity
  - Aggregation tendency
  - Requirement for secretion

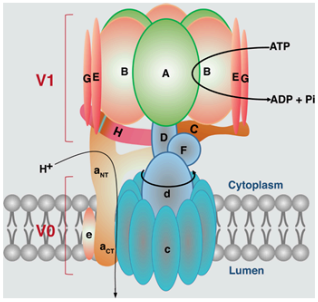


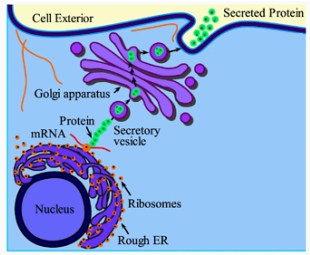


https://www.nature.com/articles/s41598-018-19491-y

Some things to consider: target requirements

- Post-translational modifications
  - Glycosylation
  - Phosphorylation
  - Hydroxylation
  - Sulfation
  - γ-Carboxylation of Glu
  - Proteolysis
- Are there stabilising factors that are needed?
  - Binding partners
  - Ligands
  - Disulfides
  - Chaperones – general or specific

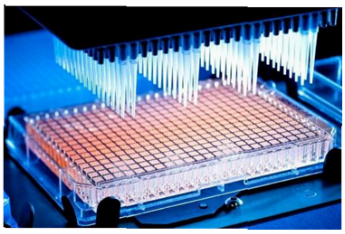




https://www.nature.com/articles/s41598-018-19491-y

Some things to consider: project management

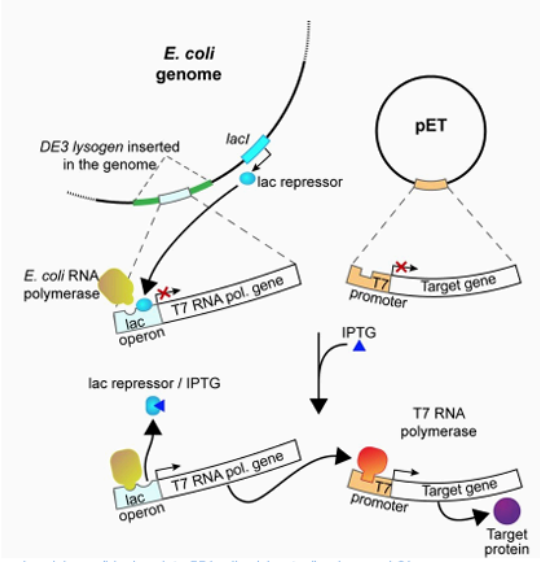
- Is throughput important?
  - Screening a large number of mutations, or domains?
- Is cost important?
- Time-critical?



1. Some tRNAs are used preferentially in highly expressed genes. ↩
2. Nascent peptide chains find their conformation while they are being produced. ↩

3. lacI: lac repressor protein, to stronly repress baseline expression. It is useful to handle toxic expressions so that leaky expression will not kill hosts. ↩

A prokaryotic expression system



https://wisc.pb.unizin.org/biochemistry551online/chapter/background-2/

- In the T7 expression system, T7 RNA polymerase is under the control of the *lac* promoter
- Addition of a mimic of a metabolite of lactose releases the lac repressor protein and de-represses expression of the polymerase
- The polymerase is then able to transcribe the target gene

4. Shown is the prokaryotic expression system regulated by the lac operator. Some promoters might be too expressive, causing nascent nucleic acids sequences do not have enough time to fold then aggregate. ↩
5. Solubility tag: highly stable and quickly folding tags can stimulate the expressed proteins fold while inhibiting aggregation. ↩
6. TEV tag: very specific to protease, allowing remove tags before or after purification. ↩