• Activity
• Cellular localization
• Protein-protein interaction
• Structure-Function relations
• Rational drug design
• Therapeutics
Proteins are often difficult to study

1. Limited amounts
2. Source unavailable
3. Unstable
4. Difficult to isolate
Cell Factories

Bacteria *E. coli*

Yeast *s. cerevisiae*

Insect SF-9

Plant cells

Bacteria *B. subtilis*

Mammalian CHO
Cell Factories setups
Tricking the host

Requirements for expression in heterologues systems
Natural solutions
Natural solutions

Hepatitis C virus

Coronavirus

Herpes virus

Bird flu virus

Smallpox virus

Influenza virus
Anatomy of an expression vector

http://wolfson.huji.ac.il/expression/vector/vec-anat.html
E. Coli: The most popular expression system

**Used for:**
- Vaccines and therapeutics
- Structural studies (X-ray and NMR)
- Protein function studies

**Pros:**
- Highest quantity, low cost
- Simple training requirements
- Well defined, many vectors and fusion tags

**Cons:**
- Aggregations, no PTMs, endotoxins
- Not suitable for large proteins, limited with membrane proteins
Used for:
• Vaccines and therapeutics
• Protein function studies
• Food supplements

Pros:
• Can accommodate large proteins >45KDa
• No endotoxins are produced (unlike *E.coli*),
• Low cost
• Well established and characterized host (genome and proteases)

Cons:
• Significant initial investments (host selection; training)
• Not efficient in high % of proteins
• Hypermannosylated and non homogenous glycosylation
Used for:

- Vaccines
- Structural and protein characterization studies
- Agriculture applications (veterinary vaccines/pesticides)

Pros:

- Viral vectors can contain large genes, and multiple proteins
- Good for complexes: multiple infection
- A variety of vectors fusion proteins, tags and hosts

Cons:

- Long implementation and training
- High cost media and plastics
- Secreted proteins might cause allergic response (paucimannose)
Mammalian cells (CHO)

**Used for:**
- Production of secreted proteins for therapeutics
- Structural and protein characterization studies
- Protein-protein interactions and localizations

**Pros:**
- Regulated systems for therapeutics
- Suitable glycosylation and PTMs for therapeutics

**Cons:**
- Low volumetric productivity
- High cost of media, serum, tissue culture plastics and reagents
- Complex purification of cytosolic and secreted proteins
- Requires comprehensive training and time consuming
Basic considerations in heterologous protein production

1. Required DNA and vector elements
2. Host requirements
3. Prokaryotic vs. Eukaryotic hosts
4. Expression example in E. coli
1. Required DNA elements

- ORF = Start codon (ATG) + Stop codon (TAA)
- Promoter: Transcription initiation & regulation
- RBS: Translation regulatory element
- Termination signal
- Delivery and maintenance in host

All elements should be host-compatible!
2. Host requirements

Or: Why are some cellular factories better for your protein?

Quantity

Time scale = Efficiency

Simplicity

Cost

Stability and robustness
3. Prokaryotes vs. Eukaryotes

Prokaryotic Cell (Bacteria)

Eukaryotic Cell (Plant)

Eukaryotic Cell (Animal)
The protein pathway in the host machinery

- Transcription
- Translation
- Modification
- Folding
- Destination
- Interactions
- Degradation

Diagram:

- DNA replication
  - Error rate $10^{-8}$-$10^{-10}$
- Transcription into mRNA
  - $10^{-4}$
- Translation of mRNA
  - $10^{-4}$
- Ribosome
- Selection of tRNAs by ribosomes
  - $10^{-3}$-$10^{-5}$
- Amiroylation of tRNA
  - $10^{-3}$-$10^{-4}$
- Amino acid + tRNA → aaRS
- Functional protein
- Maturation → Misfolded protein → Aggregation
Differences in translation machinery

### Prokaryotic
- rRNA:
  - 23S (2900 rNTs)
  - 5S (120 rNTs)
  - 16S (1500 rNTs)
- Proteins: Total 31
- Subunits: 50S
- Assembled ribosomes: 70S

### Eukaryotic (vertebrate)
- rRNA:
  - 28S (4800 rNTs)
  - 5.8S
  - 5S (120 rNTs)
  - 18S (1900 rNTs)
- Proteins: Total 50
- Subunits: 60S
- Assembled ribosomes: 80S
Coupled transcription-translation in bacteria
A single mRNA molecule usually has many ribosomes traveling along it, in various stages of synthesizing the polypeptide. This complex is called a polysome.
Transcription and translation in Eukaryotes

1. RNA is transcribed from a DNA template.

2. In eukaryotes, the RNA transcript (pre-mRNA) is spliced and modified to produce mRNA, which moves from the nucleus to the cytoplasm.

3. mRNA leaves the nucleus and attaches to the ribosome.

4. Each amino acid attaches to its proper tRNA with the help of a specific enzyme and ATP.

5. A succession of tRNAs add their amino acids to the polypeptide chain as the mRNA moves through the ribosome, one codon at a time. (When completed, the polypeptide is released from the ribosome.)

©1999 Addison Wesley Longman, Inc.
Protein secretion and translocation

Protein Secretion Pathways

- SEC Pathway: Post-translational
- SRP Pathway: Co-translational
- TAT Pathway: Post-translational

- SecYEG Translocon
- TAT Translocon

Cytoplasm
Periplasm
Extracellular
Sub cellular localization in eukaryotes

1. mRNA translation in the cytosol.
2. ER signal sequence recognition.
3. Transport through the Golgi complex.
4. Sorting to the plasma membrane, lysosome, or endoplasmic reticulum (ER).
5. Peroxisome targeting.

SECRETORY PATHWAY

Nucleus

Ribosomes

Outer nuclear membrane

Inner nuclear membrane

Nuclear pore

Cytosol

mRNA

Cytosolic protein

Targeting sequence

Rough endoplasmic reticulum

Golgi complex

Plasma membrane

Lysosome

Mitochondrion

Intermembrane space

Matrix

Inner membrane

Stroma

Thylakoids

Chloroplast

Membrane

Matrix

Inner membrane

Outer membrane

Peroxisome

Membrane

Matrix

Inner membrane

Outer membrane
Molecular chaperones and chaperonins differ in eukaryotes and prokaryotes
Ab production = quantity
Immunization = q&q
Antibodies & drugs = q&q

Activity assays = quality
Localization = quality
Protein interactions = quality

High quality requires homogeneous PTMs, proper folding
<table>
<thead>
<tr>
<th>Host</th>
<th>Strategy</th>
<th>Produce g/liter</th>
<th>Expression % total cell protein</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>• Inclusion bodies</td>
<td>0.5-5</td>
<td>5%-25%</td>
<td>Common strategy</td>
</tr>
<tr>
<td></td>
<td>• Soluble proteins</td>
<td>0.5-5</td>
<td>5%-25%</td>
<td>Minority of prot.</td>
</tr>
<tr>
<td></td>
<td>• Secreted</td>
<td>&lt;0.5</td>
<td>0.3%-4% (80% pure)</td>
<td>Periplasm</td>
</tr>
<tr>
<td></td>
<td>• Fusion proteins</td>
<td>1-3</td>
<td>5%-15%</td>
<td>Stability/solubility</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>• Secretion</td>
<td>&lt;0.25</td>
<td>&lt;1% (30-60% pure)</td>
<td>Normally secreted</td>
</tr>
<tr>
<td></td>
<td>• Intracellular</td>
<td>1-4</td>
<td>5%-20%</td>
<td>Soluble</td>
</tr>
<tr>
<td></td>
<td>• Intra-inclusion</td>
<td>1-4</td>
<td>5%-20%</td>
<td>Aggregates</td>
</tr>
<tr>
<td>Insect cells</td>
<td>• Baculovirus</td>
<td>0.001-0.5</td>
<td>&lt;30%</td>
<td>Secreted and intracellular</td>
</tr>
<tr>
<td>Mammalian cells</td>
<td>• Secretion</td>
<td>0.001-0.1</td>
<td>&lt;0.5%</td>
<td>Low productivity</td>
</tr>
<tr>
<td></td>
<td>• Intracellular</td>
<td>0.001</td>
<td>&lt;0.1%</td>
<td>Complex PTM</td>
</tr>
<tr>
<td>Post translational modifications</td>
<td>E. coli</td>
<td>Yeast</td>
<td>Insect cells</td>
<td>mammalian</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>---------</td>
<td>-------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>N-Acetylation</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Amidation</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>γ–Carboxylation of glutamate</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N-glycosylation</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>O-glycosylation</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Heterodimer</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hydroxylation</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Myristoylation</td>
<td>Possible</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Palmitoylation</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Phosphorilation</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Protein proteolytic process</td>
<td>Signal pept.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sulfation</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N-term Met removal</td>
<td>Partial</td>
<td>Partial</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Modifications can change your protein

Glycosylations can influence:

• Solubility
• Stability
• Protein folding

Oligosaccharides will affect:

• Specific activity
• Antigenicity / immunogenicity
• Pharmacokinetics
• Clearance

Details on the protein will determine expression system
5. Expression outline in *E. coli*
Expression vector for *E. coli*

- **PT5**
- **IacO**
- **IacO**
- **RBS**
- **ATG**
- **6xHis**
- **MCS**
- **Stop Codons**

**Ampicillin**

- pQE-30
- pQE-31
- pQE-32

3.4 kb
Transcription induction

NO IPTG

LacI protein → Gene of interest → STOPs protein production

ADD IN IPTG

LacI protein → IPTG

LacI protein → Gene of interest → STARTs protein production
Mix plasmid with bacteria
Plate on selective plate
Pick a colony and induce expression
Lyse cells and run on SDS-PAGE
Scale according to initial yield
Go to Mario
Successful expression depends on the protein characteristics.

From DNA structure to RNA transcription via Protein translation into Protein modification and accumulation.

Each step can affect the production of your protein.
Learn as much as possible about your protein!!

- Eukaryotic? Prokaryotic?
- Secreted? Cytosolic? membrane?
- Contains rare codons?
- Contains cysteins?
- Might be toxic to the host cell?
- Degradable?
- Requires post-translational modifications for activity?
Next week’s options:

• Getting deeper into bacterial expression systems?
• Mainly expression troubleshooting?
• Tags and fusion protein options?
• Other systems: Insect or Mammalian cells?
• Cloning strategies?