Recombinant Protein Expression & Purification -- Challenges & Solutions

Liyan Pang, Ph.D.

liyan.pang@genscript.com
Table of Contents

1. Choose Expression System
2. Optimize Protein Expression
3. Protein Refolding
4. Protein Purification
5. Difficult-to-Express Proteins
6. Advantages of GenScript Protein Services
Challenges in Protein Expression

- Soluble expression
- Correct conformation
  - Disulfide bond
  - Chaperonins
  - Refolding
- Production efficiency
- Purity
- Difficult-to-express proteins
  - Membrane protein
  - Toxic protein
  - Easy-to-degrade protein
Protein Expression Impacting Factors

- Sequence
- Expression system
  - Bacterial
  - Yeast
  - Insect
  - Mammalian
- Vector
  - Promoter
  - Tag
- Host strain
- Expression conditions
  - Medium component
  - Temperature
  - Inducer concentration & time
  - Inoculation volume
Protein Expression Common Questions

Protein name
- Full-length or fragment of the protein?
- PCR or gene synthesis?

Obtain DNA
- Expression system: bacterial, yeast, insect or mammalian?
- Expression vector?
- Tag or no tag?
- Which tag?
- Tag placed at N or C-terminus?

Create expression clone
- Expression host?
- The best expression matrix?

Express the protein
- A necessary step?

Small scale test expression

Scale up protein expression
- Purification strategy?
- Refolding?
- Endotoxin removal?

Protein purification

Protein characterization
- QC methods?
Expression Systems

Bacteria
- *E. coli*
  1. “Work horse”
  2. Well established
  3. High expression
  4. Simple genetics
  5. Easy scale up
  6. Speed
  7. Costs
  8. Equipment

Insect
- *Sf9, Sf21, S2, High-5*
  1. PTMs
  2. Soluble proteins
  3. High expressers

Yeast
- *S. cerevisiae, P. pastoris*
  1. PTMs
  2. Soluble proteins
  3. High expresser

Mammalian
- *CHO, HEK, COS*
  1. PTMs
  2. Soluble proteins
  3. Low expresser
  4. Expensive

Cell Free
- *In vitro*
  1. Expensive
  2. Not reproducible
  3. Scalability issues

Before Embarking on a Protein Expression Project
Factors to Consider:

- Protein property
  - MW
  - Disulfide bonds
  - Post-translational modifications
  - Homogeneity

- Intended applications
  - Structural biology
  - Functional assays
  - Therapeutic protein/vaccines
  - Antigens for Ab production
  - Protein-protein interactions

- Yield
- Cost

Which Expression System to Choose?
# Expression System Selection

<table>
<thead>
<tr>
<th>Expression System</th>
<th>Pros</th>
<th>Cons</th>
<th>Intended Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial</strong></td>
<td>Relatively inexpensive</td>
<td>Lack of efficient post-translational modifications</td>
<td>Structural biology</td>
</tr>
<tr>
<td></td>
<td>Simple genetics</td>
<td>Codon usage issues</td>
<td>Functional assays</td>
</tr>
<tr>
<td></td>
<td>Easy to manipulate</td>
<td>Inclusion bodies</td>
<td>Antigen production</td>
</tr>
<tr>
<td></td>
<td>Easy scale up</td>
<td>Low yield and activity for some eukaryotic proteins</td>
<td>Protein-protein interaction (Bacterial &amp; Yeast expression systems recommended)</td>
</tr>
<tr>
<td></td>
<td>Fast expression</td>
<td>Difficult to express higher MW proteins</td>
<td>Therapeutic protein (Yeast &amp; Mammalian expression systems recommended)</td>
</tr>
<tr>
<td><strong>Yeast</strong></td>
<td>Diverse post-translational modifications</td>
<td>Improper glycosylation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low cost of culture media</td>
<td>Excessive glycosylation</td>
<td></td>
</tr>
<tr>
<td><strong>Insect</strong></td>
<td>Good secretion</td>
<td>Long production time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post translational modifications resemble mammalian system</td>
<td>Relative high media costs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suitable for toxic gene products</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mammalian</strong></td>
<td>Comprehensive post-translational modifications</td>
<td>Long production time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Excellent method for the production of bioactive proteins</td>
<td>High media costs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Protein yields relatively lower</td>
<td></td>
</tr>
</tbody>
</table>
# GenScript Protein Standard Services

**From sequence to purified protein - gene synthesis included!**

<table>
<thead>
<tr>
<th>Expression System</th>
<th>Deliverables</th>
<th>Timeline</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BacPower™</strong></td>
<td>3mg purified protein guaranteed</td>
<td>6 - 8 weeks</td>
<td>Staring from $2,200</td>
</tr>
<tr>
<td><strong>InsectPower™</strong></td>
<td>1mg purified protein guaranteed</td>
<td>8 - 10 weeks</td>
<td>Staring from $3,950</td>
</tr>
<tr>
<td><strong>MamPower™</strong></td>
<td>3mg purified recombinant protein or 50mg purified antibody guaranteed</td>
<td>8 - 12 weeks</td>
<td>Staring from $8,499</td>
</tr>
<tr>
<td><strong>YeastHigh™</strong></td>
<td>Customizable production up to 2000L</td>
<td>8 - 10 weeks</td>
<td>Quote</td>
</tr>
</tbody>
</table>

Make Research Easy
Eliminate the guesswork from your protein production work

- Evaluate whether your target protein expresses in your chosen system
- Identify the best expression system for your target protein
  - PROTential™ Standard packages
  - Before scale-up protein production, to avoid waste on your time & valuable resources

- Optimize your protein expression
  - PROTential™ Silver & Gold packages
  - When challenges arise – the most efficient & cost-effective way

One stop service at GenScript: gene synthesis → Subcloning → PROTential™ → Scale up protein production

GenScript’s Solution for Expression Optimization
# PROTential™ - Portfolios

<table>
<thead>
<tr>
<th>Name</th>
<th>Service Type</th>
<th>Expression System(s)</th>
<th>Price</th>
<th>Timeline</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard</strong></td>
<td>Protein Expression Evaluation</td>
<td><em>E. coli</em> (SC1653-B)</td>
<td>Starting from $280</td>
<td>1-2 weeks</td>
<td>Test 1 condition for soluble expression in a customer chosen bacterial strain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Insect (SC1653-I)</td>
<td>Starting from $400</td>
<td>3-4 weeks</td>
<td>Test 1 condition for soluble expression in Sf9, Sf21, S2 or Hi-5 cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mammalian (SC1653-M)</td>
<td>Starting from $500</td>
<td>3-4 weeks</td>
<td>Test 1 condition for soluble expression in CHO or 293 cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 3 Systems (SC1653-3S)</td>
<td>Starting from $950</td>
<td>3-4 weeks</td>
<td>Test soluble expression 1 condition/expression system (<em>E. coli</em>, Insect, Mammalian); total 3 systems</td>
</tr>
</tbody>
</table>
| **Silver**      | Protein Expression Optimization       | *E. coli* (SC1667)        | Login to inquire | 2-3 weeks | • Test 8 different conditions  
• Optimize growth temperature, media components & inducer concentrations  
• Identify the best expression condition with your chosen vector and bacterial strain                                      |
| **Gold**        | Protein Expression Optimization       | *E. coli* (SC1668)        | Login to inquire | 4-8 weeks | • Test 40 different conditions  
• Optimize growth temperatures, media components, inducer concentrations, promoters, host cell strains & fusion partners  
• Robust, industry-first, high-throughput expression and solubility optimization matrix                                      |
Challenges:

- Pilot purification (final yield 1mg/L) - ~28kDa protein;
- The protein can be only purified from the soluble part;
- Large amount of protein with large scale fermentation (1000L) and purification is needed.

Strategies:

1. Expression improvement:
   a. Promoter optimization
   b. Strain optimization
   c. pH optimization
   d. Temperature and induction optimization
   e. Inoculated quantity optimization

2. Recovery rate improvement during purification - Purification condition optimization
## Case Study- Protein Expression Optimization

### Strategy execution: HT expression testing

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X-1</td>
<td>X-2</td>
<td>X-3</td>
<td>X-4</td>
<td>X-5</td>
<td>X-6</td>
</tr>
<tr>
<td>B</td>
<td>X-7</td>
<td>X-8</td>
<td>X-9</td>
<td>X-10</td>
<td>X-11</td>
<td>X-12</td>
</tr>
<tr>
<td>C</td>
<td>X-13</td>
<td>X-14</td>
<td>X-15</td>
<td>X-16</td>
<td>X-17</td>
<td>X-18</td>
</tr>
<tr>
<td>D</td>
<td>X-19</td>
<td>X-20</td>
<td>X-21</td>
<td>X-22</td>
<td>X-23</td>
<td>X-24</td>
</tr>
<tr>
<td>E</td>
<td>X-25</td>
<td>X-26</td>
<td>X-27</td>
<td>X-28</td>
<td>X-29</td>
<td>X-30</td>
</tr>
<tr>
<td>F</td>
<td>X-31</td>
<td>X-32</td>
<td>X-33</td>
<td>X-34</td>
<td>X-35</td>
<td>X-36</td>
</tr>
<tr>
<td>G</td>
<td>X-37</td>
<td>X-38</td>
<td>X-39</td>
<td>X-40</td>
<td>X-41</td>
<td>X-42</td>
</tr>
<tr>
<td>H</td>
<td>X-43</td>
<td>X-44</td>
<td>X-45</td>
<td>X-46</td>
<td>X-47</td>
<td>X-48</td>
</tr>
</tbody>
</table>

Conditions X-1 to X-48
Case Study - Protein Expression Optimization

- 12X as much protein expression yield as original

1 mg/L starting protocol

2 mg/L T7 promoter/induction condition optimization

5 mg/L phoA promoter/induction condition optimization

6 mg/L growth condition optimization (pH)

12 mg/L seeding density optimization
Table of Contents

1. Choose Expression System
2. Optimize Protein Expression
3. Protein Refolding
4. Protein Purification
5. Difficult-to-Express Proteins
6. Advantages of GenScript Protein Services
What is an inclusion body?

- When *E. coli* is transformed to manufacture large amounts of recombinant protein, the protein sometimes forms dense aggregates of insoluble misfolded proteins, known as inclusion bodies.

**Benefit**

- allow high protein concentrations
- protect sensitive proteins from proteolytic (enzymatic) degradation
- protect the cell from any toxic proteins

**Challenge**

- to solubilise and refold this protein into its correct ‘active’ form
Protein Refolding Introduction

- All the information necessary for folding the peptide chain into its native structure is contained in the primary amino acid sequence of the peptide.
- The native form of a protein has the thermodynamically most stable structure.
- There are vastly too many different possible conformations for a protein to fold by a random search.
- A new view of protein folding suggested that there is no single route, but a large ensemble of structures follow a many dimensional funnel to its native structure.
GenScript’s FoldArt™ Technology Overview

- Evaluation of target proteins' biochemical and biophysical properties

- Refolding optimizations
  - Selection of particular refolding strategy based on protein's sequence and the structural properties.
  - Buffer screening: Solutions for the inclusion body will be diluted to 20 different refolding buffers to determine which parameters affect the refolding results.

- Denaturant removal
  Techniques: dilution, dialysis, diafiltration, gel filtration, and chromatography (ion exchange, size exclusion, and affinity)

- Validation
  Refolding results will be validated by SDS-PAGE, HPLC and/or functional assay.
Refolding Buffer

- Refolding conditions must be optimized for each individual protein.
- Important variables are:
  - buffer type
  - pH
  - ionic strength
  - Additives, often in combination (glycerol, redox reagents, saccharides, amino acids, metal ion, detergents, chaperones)

Over 95% of the inclusion bodies can be solubilized and refolded by our proprietary FoldArt™ protein refolding technology
Case Study: Protein Refolding

- Human interleukin – 5: disulfide-bond linked homodimer as active form
Table of Contents

1. Choose Expression System
2. Optimize Protein Expression
3. Protein Refolding
4. Protein Purification
5. Difficult-to-Express Proteins
6. Advantages of GenScript Protein Services
Protein Purification

- Flexible purification methods
  - Affinity column (GST, Ni-NTA, protein A/G/L resins, etc.)
  - Ion exchange
  - Size exclusion
  - Hydrophobic interaction chromatography (HIC)

- Double Tag strategy for big protein isolation
Affinity Chromatography

- **Affinity tags**
  - His
  - GST
  - MBP
  - Flag
  - SUMO

- **Matrices & elution conditions**

- **Methods**
  - Column
  - Batch

- **Tag removal by site-specific protease**
  - TEV
  - Enterokinase
  - SUMO protease

- Add mixture to column. Discard flow through.
- Add wash to column. Discard flow through.
- Add elution buffer to column. Retain flow through.
- Add mixture to batch to bind target molecule.
- Remove mixture.
- Add wash buffer.
- Remove wash buffer.
- Add elution buffer
- Remove elution buffer and retain
What are endotoxins?
- Endotoxins, also known as lipopolysaccharides (LPS), are large molecules found in the outer membrane of Gram-negative bacteria, which elicit strong immune responses in animals.

Detection
- Gel clot method
- Chromogenic method

Removal methods:
- Polymyxin B (PMB) – affinity based ToxinEraser™ (L00338) by GenScript allows highly efficient removal of endotoxin down to 0.1 EU/ml
- Size exclusion chromatography (SEC)
- Ion exchange chromatography (IEC)
## Case Study: Endotoxin Removal

GenScript Endotoxin Removal Services: Endotoxin <= 1 EU/ug; <= 0.1 EU/ug; <= 0.01 EU/ug

<table>
<thead>
<tr>
<th></th>
<th>Protein 1</th>
<th>Protein 2</th>
<th>Protein 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume of resins</td>
<td>3 ml</td>
<td>3 ml</td>
<td>3 ml</td>
</tr>
<tr>
<td>Volume of sample</td>
<td>15 ml</td>
<td>15 ml</td>
<td>15 ml</td>
</tr>
<tr>
<td>Initial endotoxin</td>
<td>500,000 - 2,000,000 EU/ml</td>
<td>&gt; 40,000,000 EU/ml</td>
<td>&gt; 40,000,000 EU/ml</td>
</tr>
<tr>
<td>Final concentration</td>
<td>2.2 mg/ml</td>
<td>1 mg/ml</td>
<td>0.8 mg/ml</td>
</tr>
<tr>
<td>Final endotoxin</td>
<td>64 – 128 EU/ml</td>
<td>20 – 40 EU/ml</td>
<td>12.5 – 25 EU/ml</td>
</tr>
<tr>
<td>Final endotoxin</td>
<td>0.029 – 0.058 EU/μg</td>
<td>0.02 – 0.04 EU/μg</td>
<td>0.016 – 0.032 EU/μg</td>
</tr>
<tr>
<td>Table of Contents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1</strong></td>
<td>Choose Expression System</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>Optimize Protein Expression</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>Protein Refolding</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>Protein Purification</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5</strong></td>
<td>Difficult-to-Express Proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>6</strong></td>
<td>Advantages of GenScript Protein Services</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Difficult-to-Express Proteins

- Membrane protein
- Toxic protein
- Prone-to-degrade protein
Membrane Protein

Definition: Membrane proteins are proteins that interact with biological membranes.

• Targets of over 50% of all modern medicinal drugs.
• 20-30% of all genes in most genomes encode membrane proteins.

❖ Expression
  ▪ Insect cells or mammalian cells;
  ▪ Budded baculovirus or virus-like particle;
  ▪ Cell-free

❖ Purification
  ▪ Detergent screen;
  ▪ Phosphate lipid;
  ▪ Nano-disc;
Gene synthesis and subcloning into transfer vector

Generating the Recombinant Bacmid DNA

Transfecting Insect cells

Amplifying Baculovirus

Expression optimization

Large scale expression and harvest cell pastes

Homogenizing cells with a glass-Teflon tissue homogenizer

Centrifugation at low speed

Centrifugation at ultra speed and re-suspend the resulting pellet

- Deliver recombinant Baculovirus, or cell pastes, or prepared membrane protein to the customer.
- Deliver functional budded virus containing membrane protein to the customer.
- Provide high throughput screening of membrane proteins and deliver expression evaluation report to the customer.
Case Study: Membrane Protein

- 25 L expression of GPCR, Expression Level: 30-40 mg/L
- Deliver the cell pastes in 6 weeks.

WB analysis of the membrane preparation
1. Whole cell lysate of GPCR before sonication
2. Whole cell lysate of GPCR after sonication
3. Supernatant after centrifugation of cell lysate for 10 minutes at 8k rpm
4. Pellet after centrifugation of cell lysate for 10 minutes at 8k rpm
5. Supernatant after ultracentrifugation of cell lysate for 45 min at 42k rpm
6. Pellet after ultracentrifugation of cell lysate for 45 min at 42k rpm
7. Negative control: supernatant after centrifugation of Sf9 cell lysate for 10 minutes at 8k rpm
8. Negative control: pellet after centrifugation of Sf9 cell lysate for 10 minutes at 8k rpm

Antibody: anti-His monoclonal-antibody (Genscript, Cat.No. A00186)
Toxic Protein Expression

Definition: Toxic proteins defined here as proteins that cause cell death or severe cultivation and maintenance defects during the growth phase when their genes were introduced into E. coli strain.

- Mostly due to leaking expression
- ~80% protein growth and expression problems are caused by the toxicity of proteins

Strategies in solving the problem

- Promoter selection
- Suppress basal expression from leaky inducible promoters
- Tight control of plasmid copy numbers
- Protein production as inactive (insoluble) forms
Case Study: Prone-to-Degrade Protein

◆ Inconsistency of measured concentration

◆ Trouble shooting:
  • Transfection methods
  • Cell lysis
  • Purification

◆ Challenge: DNase is only partially responsible for the protein degradation. This protein itself is prone-to-degrade.

◆ Solutions:
  • Remove DNase
  • Add protease inhibitor to every step
  • Optimize buffer components
  • Add protein stabilizers
  • Lyophilization immediately after protein purification
  • Storage temperature

Western blot

Before

After
Table of Contents

1. Choose Expression System
2. Optimize Protein Expression
3. Protein Refolding
4. Protein Purification
5. Difficult-to-Express Proteins
6. Advantages of GenScript Protein Services
Advantages of GenScript Protein Services

- **Core in-house technologies for expression optimization & production efficiency.**

- OptimumGene™ – expression system specific codon optimization
- BacPower™ – increase bacterial soluble expression
- FoldArt™ – ensure appropriate protein refolding
- YeastHigh™ – high copy-number gene selection technique
- BacuVance™ – for protein secretion from baculovirus-infected insect cells
- MamPower™ – technology licensed from NRC for rapid recombinant protein production with high yield
- DoubleTag strategy – for big protein isolation
Advantages of GenScript Protein Services

◆ One-stop service from sequence to purified proteins with large capacity.

- **Guaranteed Protein Expression Package**
  - 3 mg purified soluble protein from $2,200
  - Subcloning
  - Transformation
  - Expression
  - Refolding

- **Protein Expression and Purification Services**
  - Bacteria
  - Yeast
  - Baculovirus/insect cells
  - Mammalian

- **Bioprocessing Services**
  - Mammalian protein expression services
  - Stable cell line development & protein production

- **Large-scale Protein Production Services**
  - Bacterial fermentation up to 1,000 L
  - Yeast fermentation up to 500 L
  - Baculovirus/insect cell production up to grams
  - Mammalian cell production up to grams

Make Research Easy
**Advantages of GenScript Protein Services**

- **Flexible production scales**
- **Fast turn-around time (from sequence to purified protein in as little as 4 weeks)**

### Capacity:

<table>
<thead>
<tr>
<th></th>
<th>Bacteria</th>
<th>Yeast</th>
<th>Baculovirus</th>
<th>Mammalian</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000 L</td>
<td>500 L</td>
<td>100 L</td>
<td></td>
<td>500 L</td>
</tr>
<tr>
<td>Fermentor</td>
<td>Fermentor</td>
<td>Wave™ Mixer</td>
<td>Wave™ Mixer</td>
<td>Sartorius bioreactor Hyclone SUB bioreactor</td>
</tr>
</tbody>
</table>
Variety of GenScript Protein Services

- **Protein Expression & Purification**
  Bacteria, Insect, Mammalian, Yeast, Customized protein services, Fermentation, Transient & Stable cell lines

- **Protein Expression Evaluation & Optimization**
  Small scale expression testing and optimization in bacterial, insect and mammalian expression systems

- **Large Scale Protein Production**
  Upstream & Downstream Process development, fermentation, GLP-compliant Bioprocess Services

- **Chemical Protein Synthesis**
  Alternative method to produce high purity functional proteins for hard-to-express proteins

- **High Throughput Protein Variants**
  Largest high-throughput capacity in the industry, proprietary platforms, 30 days for 1,000 protein variants

- **Structural Biology**
  CrystalPro™ Gene-to-Structure Services, high purity protein preparation, crystal, co-crystal structure determination

- **Other Protein Services**
  Endotoxin removal, codon optimization, custom purification, protein characterization, refolding
GenScript has delivered over 5,000 proteins in four expression systems. Statistics showed 95% success rate for all protein projects.
About GenScript

Gene

Peptide

Protein

Antibody

Discovery Biology

Cell Line

GenScript

Products & Services

Make Research Easy
Register for other webinars in the GenScript Webinar Series @
http://www.genscript.com/webinars.html

June 18, 2014/ 2:00 pm EST
Can CRISPR/Cas9 off-target genomic editing be avoided? Ways to improve target specificity - Maxine Chen, Ph.D

June 25, 2014/ 2:00 pm EST
Building a Synthetic Eukaryotic Genome – Sc2.0 - Leslie Mitchell, Ph.D., NYU Langone Medical Center

Promo Code: BacPowerUP
- Free additional amount of protein from 3mg to 5mg for any BacPower™ Guaranteed Packages
- Valid till 8/15/2014