Fusion partners for recombinant soluble protein production in *E. coli*

Keshav Vasanthavada, MSc., MS

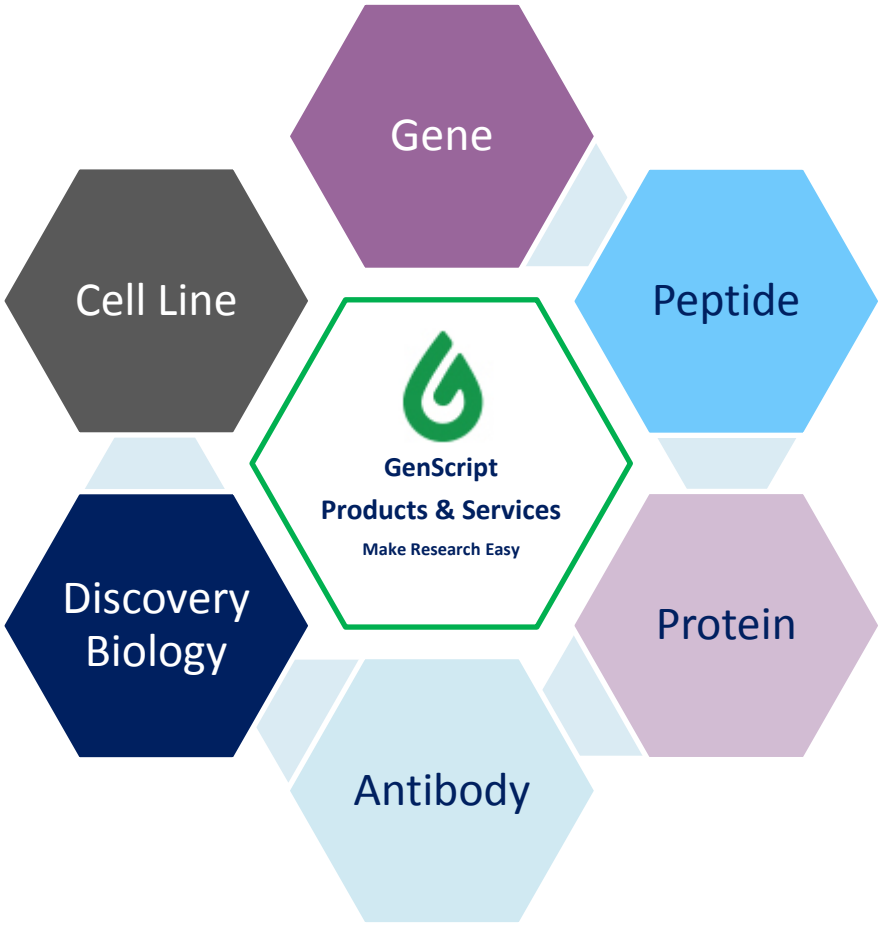
December 3rd, 2014

GenScript
Make Research Easy

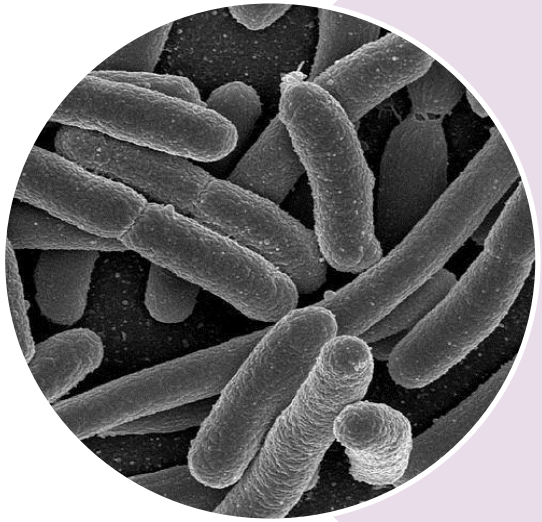
Presentation overview

1. Introduction
2. Recombinant protein solubility
3. Protein expression in *E. coli*
4. Factors that define solubility outcome
5. Fusion partner approach
6. Conclusion

About GenScript

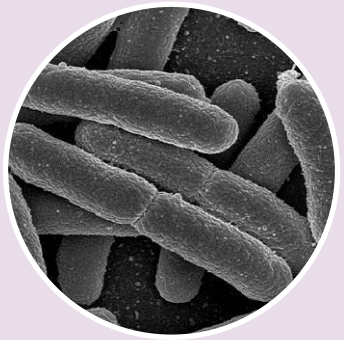


Introduction



- Recombinant protein production (RPP) is essential for functional characterization and structure determination
  - Variety of hosts available for RPP
  - Scientists faced with overwhelming choices
  - *Escherichia coli* [*E. coli*] is “work horse”
  - Successful implementation of bacterial protein production project dependent on several factors
  - Observations & Recommendations for optimizing recombinant protein production in *E. coli* with an emphasis on fusion partners
  - We describe learnings from our collective experience
  - GenScript Production Team has shipped over 5,300 batches of recombinant proteins

Host considerations



Bacteria

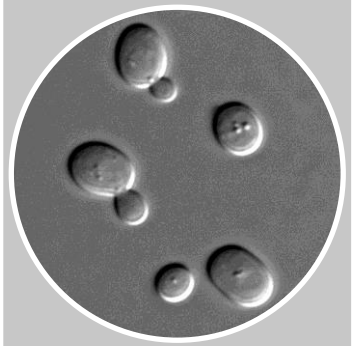
E. coli

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

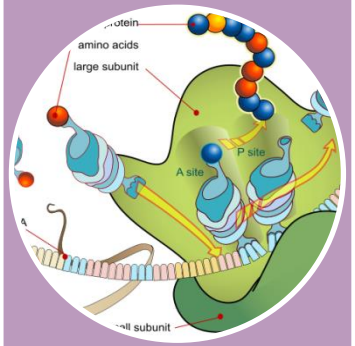
P. pastoris

1. PTMs
2. Soluble proteins
3. High expresser

Mammalian

CHO, HEK, COS

1. PTMs
2. Soluble proteins
3. Low expresser
4. Expensive

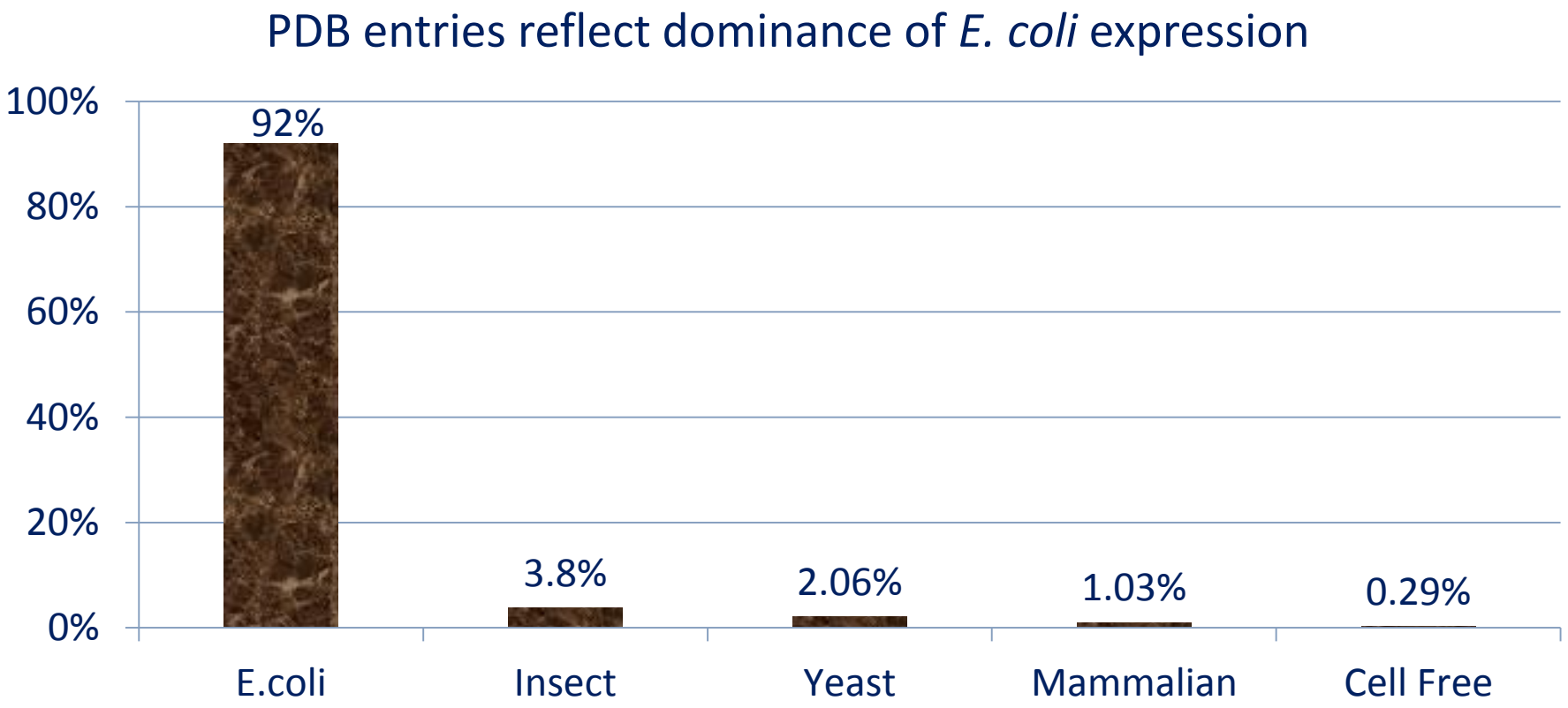


Cell Free

In vitro

1. Expensive
2. Not reproducible
3. Scalability issues

E. coli is the top choice



Number of entries in the PDB by expression system as a percentage of total number of chains with an identifiable expression system, as of April 15, 2014. All values are approximate.

Reference - http://www.rcsb.org/pdb/home/home.do

Common questions

GenScript *E. coli* expression services - I

BacPower™ Guaranteed Bacterial Protein Expression

Customer Provides	Tag options	Amount (mg)	Timeline (weeks)	Purity options	Endotoxin level options	Pricing
Target protein sequence	Small tag (e.g., His)	3-6	6-8	≥75% ≥85% ≥90%	≤1.0EU/μg ≤0.1EU/μg ≤0.01EU/μg	Starting from $2,200
	Large tag (e.g., GST)	5-10				
	No Tag	5-10	9-10			

Why use above service?

Zero-risk, best value packages that can be used for most target proteins under 100kDa. Hassle free service and lots of options to choose from – we would recommend you try this first in case of an E. coli project. Moreover, you can get an instant quote online within seconds

BacPower™ Customized Bacterial Protein Expression

Customer Provides	Deliverables	Delivery Time
Target sequence (DNA or protein)/DNA template/Expression construct	Pure, recombinant protein as specified by the customer	Depending on project complexity, it can be as short as 6 weeks

When use above service?

Best for higher difficulty targets like membrane proteins, toxins, ion channel proteins etc. Also, ideal if you have unique experimental requirements [example – special protocols, your own recommendations, your DNA samples instead of gene synthesis etc]

GenScript *E. coli* expression services - II

Fast gene-to-protein™ – Industry's Fastest Protein Production Service

Customer Provides	Deliverables	Purity Options	Delivery Time
Target protein sequence	• 3mg protein, guaranteed • COA including QC data • Gene sequence report • Expression construct (additional cost)	≥75%	30 days
		≥85%	

When use above service?

Ideal when you have a time-sensitive project and need protein fast for a milestone deadline, manuscript publication, grant renewal or simply to accelerate your research

FragPower™ Guaranteed Antibody Fragments

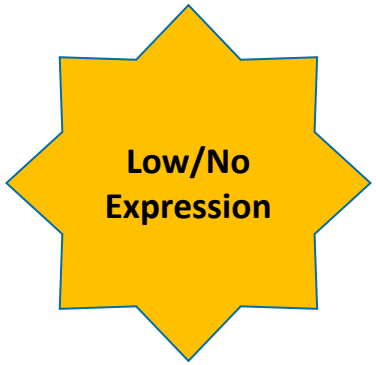
Customer Provides	Fragments	System	Pricing	Timeline (weeks)	Purity options	Deliverables
Target fragment amino acid sequence	scFv	*E. coli*	$2,399	7	≥90%	• Antibody Fragment in buffered solution with the specified amount & purity – Guaranteed • Optimized gene sequence report • QC data
	V_{HH}/V_H	*E. coli*				
	Fab	Insect	$3,399	10		

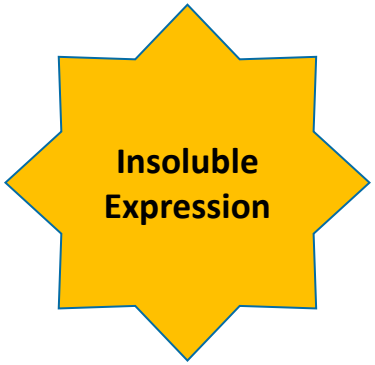
Why use above service?

Cheaper than generating full length monoclonal antibodies. Excellent for research and diagnostic applications that require full antigen-binding capacity

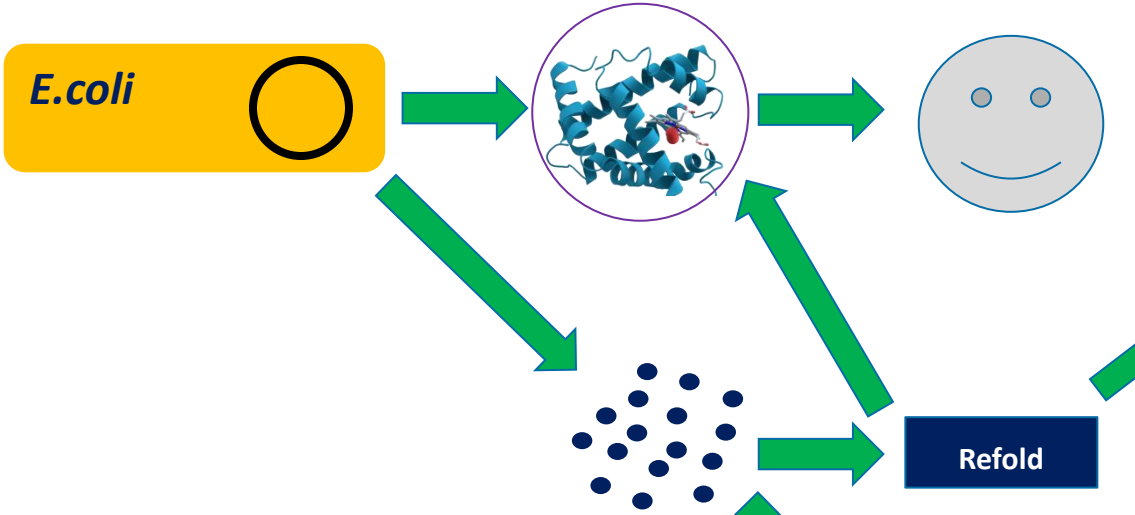
E. coli expression – why & why not?

Advantages	Disadvantages
Simple genetics	Lack of PTM
Easy to manipulate	Codon usage
Inexpensive to culture	Inclusion bodies
Easy to Scale up	Low yield of many eukaryotic proteins
Fast expression	Poor secretion
High yields	High MW proteins difficult



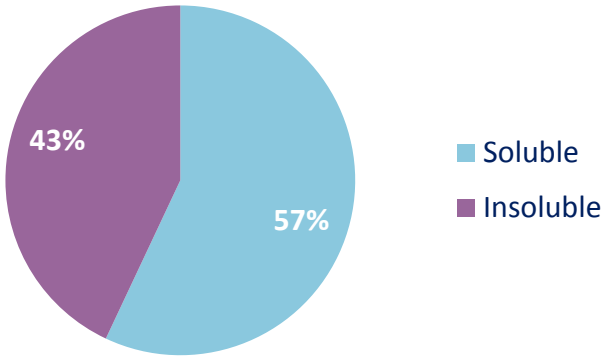


What is the problem? what are the solutions?



No Target Engineering
- Temperature
- Strains
- Media
- Buffers
- Chaperones

~24,000 Targets
NESG Complete



Engineer Target
- Truncate
- Mutate
- Fusion Partner

Goal of a protein production campaign

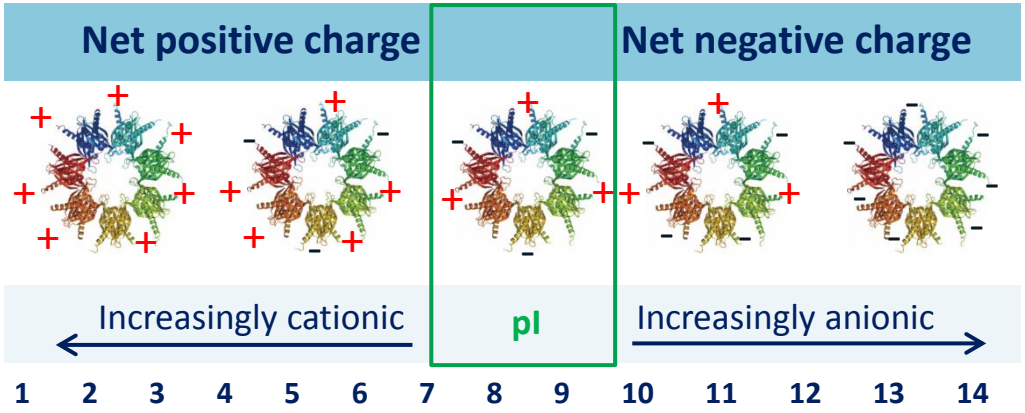
Generation of high yields of soluble protein in a cost-effective and expeditious manner is the single most important determinant of a successful protein production campaign

Relationship between solubility & functionality

- Merely soluble protein need not be conformational, high quality protein
- Even properly folded and fully functional polypeptides aggregate as inclusion bodies so functional protein is also available in the insoluble fraction
- IB formation does not imply loss of bioactivity
- Specific biological activity is the true indicator of the conformational quality of the product and therefore its biotechnological potential
- Enhancing solubility of a recombinant protein does not necessarily enhance yield of functional protein
- Majority of the time, for lack of an easier predictor, solubility is often used as the best indicator of a protein's functional activity
- Also, there is no defined way to make "functional protein" and hence solubility is the closest indicator to functionality

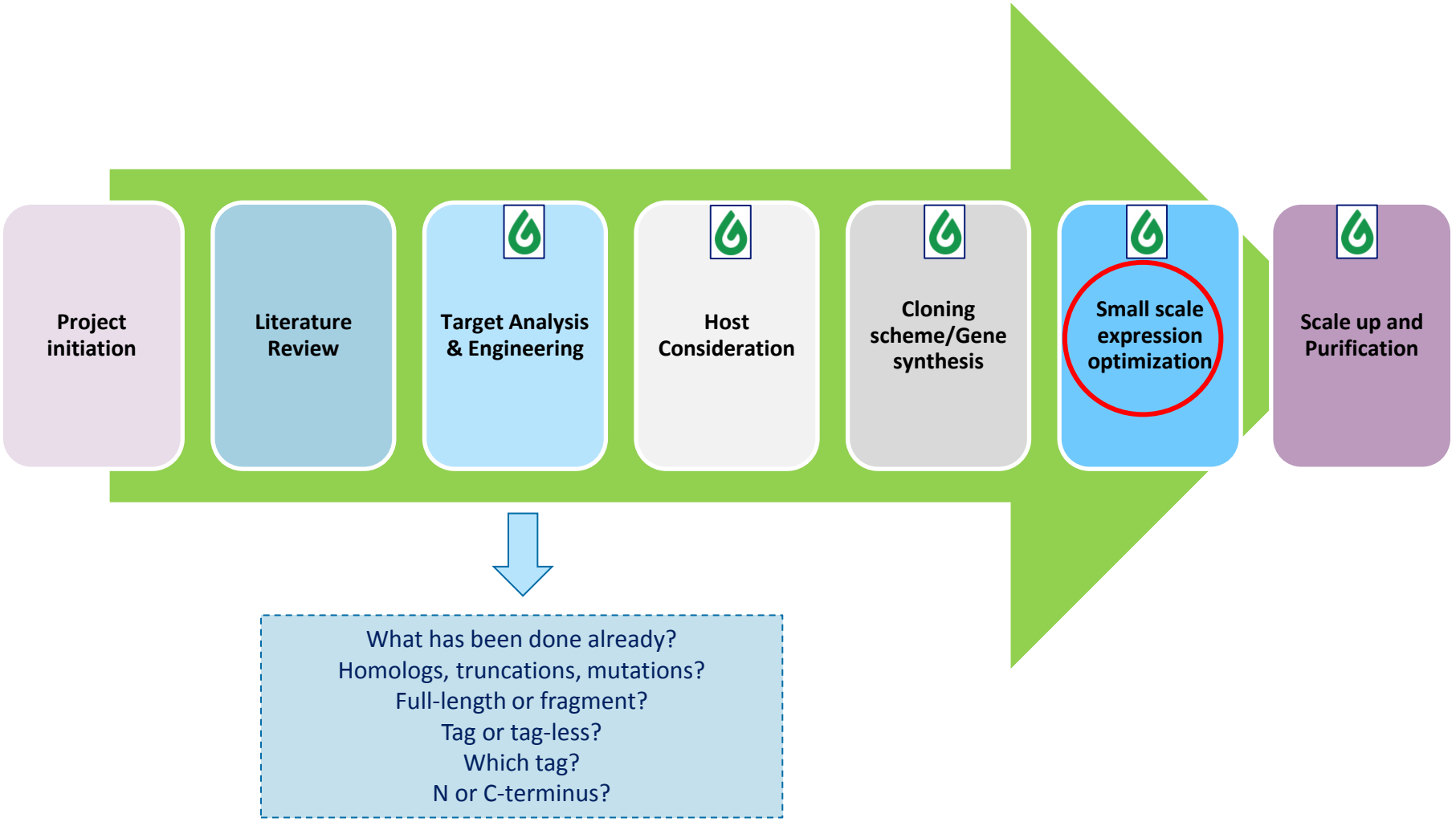
Factors that determine protein solubility

- Charge
  - Based on amino acid sequence and buffer pH
  - Functional groups determine charge at defined pH
    - Amino acid composition: Asp, Glu carry negative charges at neutral pH
    - Lys, Arg carry positive charges at neutral pH
  - At a certain pH the net charge will be 0 [pI]
  - When pH changes, charges on Asp, Glu, Lys, Arg changes
  - No net charge causes proteins to interact among themselves rather than with water molecules - aggregation
  - Charged molecules remain soluble

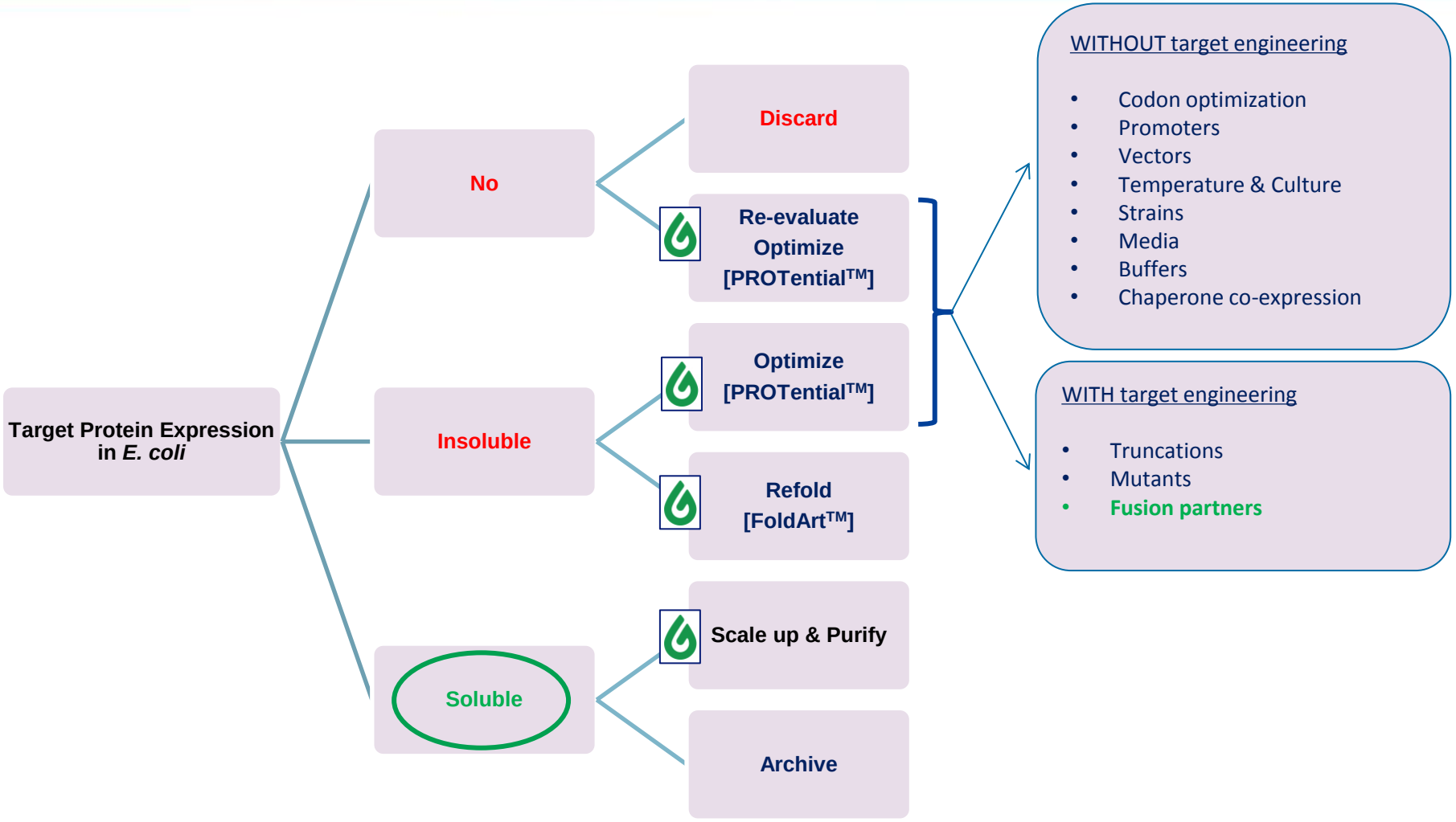


- Turn forming residue content: Pro, Ser, Gly, Asn

Protein expression campaign workflow



Flow chart of a bacterial expression project



Good idea to run small scale expression and test purification

PROTential™ - expression evaluation & optimization

Eliminate the guesswork from your Protein production work

- **What is PROTential™?**
  - PROTential™ is a new protein expression, evaluation and optimization service offered by GenScript.

- **What are the applications of PROTential™?**
  - Evaluate whether your target protein expresses in your chosen system.
  - Identify the best expression system for your target protein.
  - Identify the construct and conditions that give you the most robust soluble expression of your target protein.

- Evaluate before scale-up protein production, to avoid wasting your time & resources

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

PROTential™ portfolio

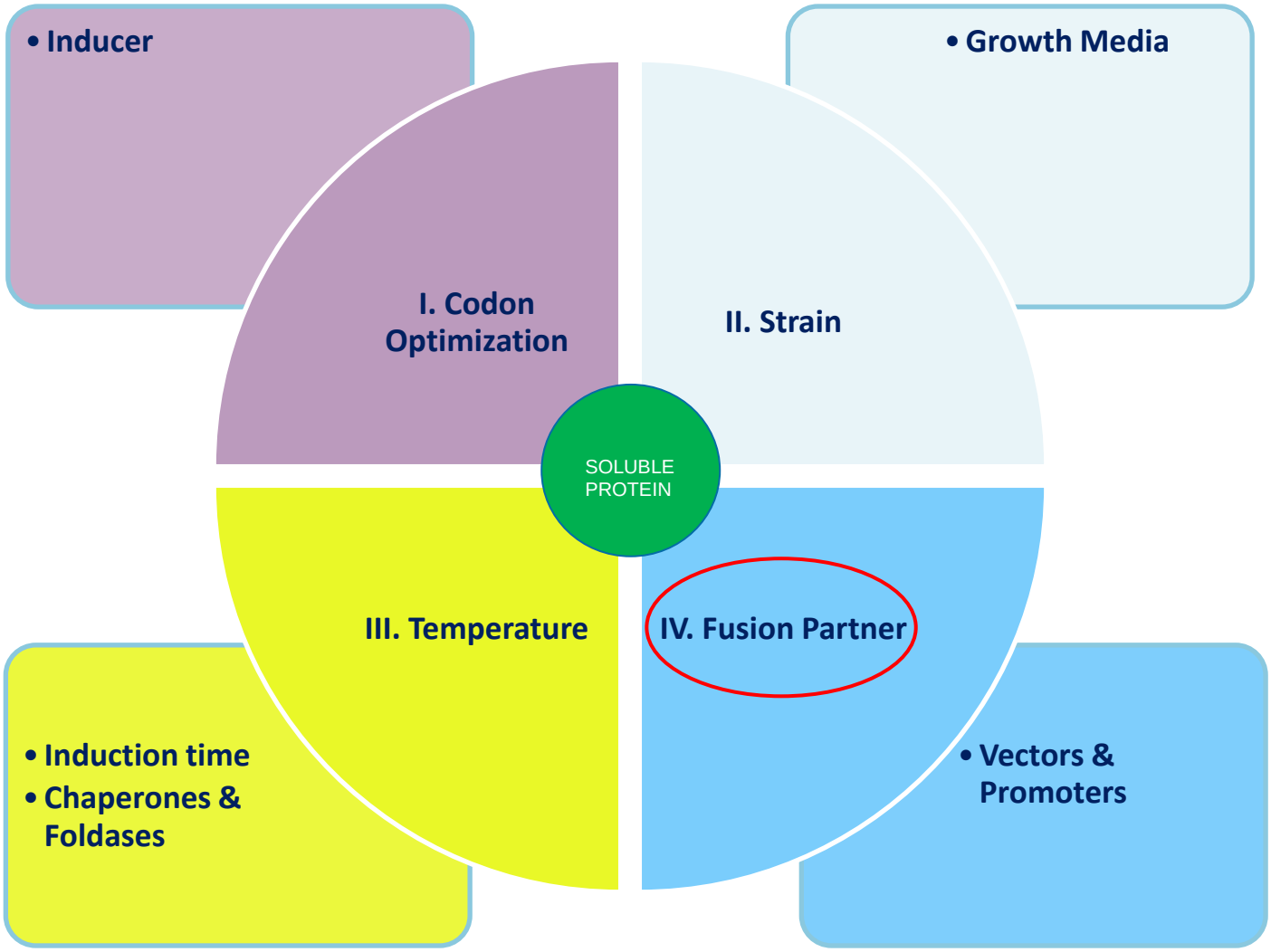
- **PROTential™ Standard Package (starting from $280)**
  - Available for 3 expression systems: bacterial, insect & mammalian
  - Protein expression evaluation & solubility test

- **PROTential™ Silver Package (starting from $1440)**
  - Currently for bacterial expression system only
  - Test 8 conditions by combining temperature, media components & inducer concentration.
  - Allow to identify the best expression condition with your chosen vector & bacterial strain.

- **PROTential™ Gold Package (starting from $4800)**
  - Currently for bacterial expression system only
  - Test 48 conditions, by combining parameters including not only those 3 covered in silver package, but also promoter, host strain, and fusion partner.
  - The most robust and high throughput protein expression & solubility optimization matrix.

- **Add-on item:** 1L bacterial expression with 1 step purification at $600/condition (except for Flag-tagged protein, the cost is $800/condition).

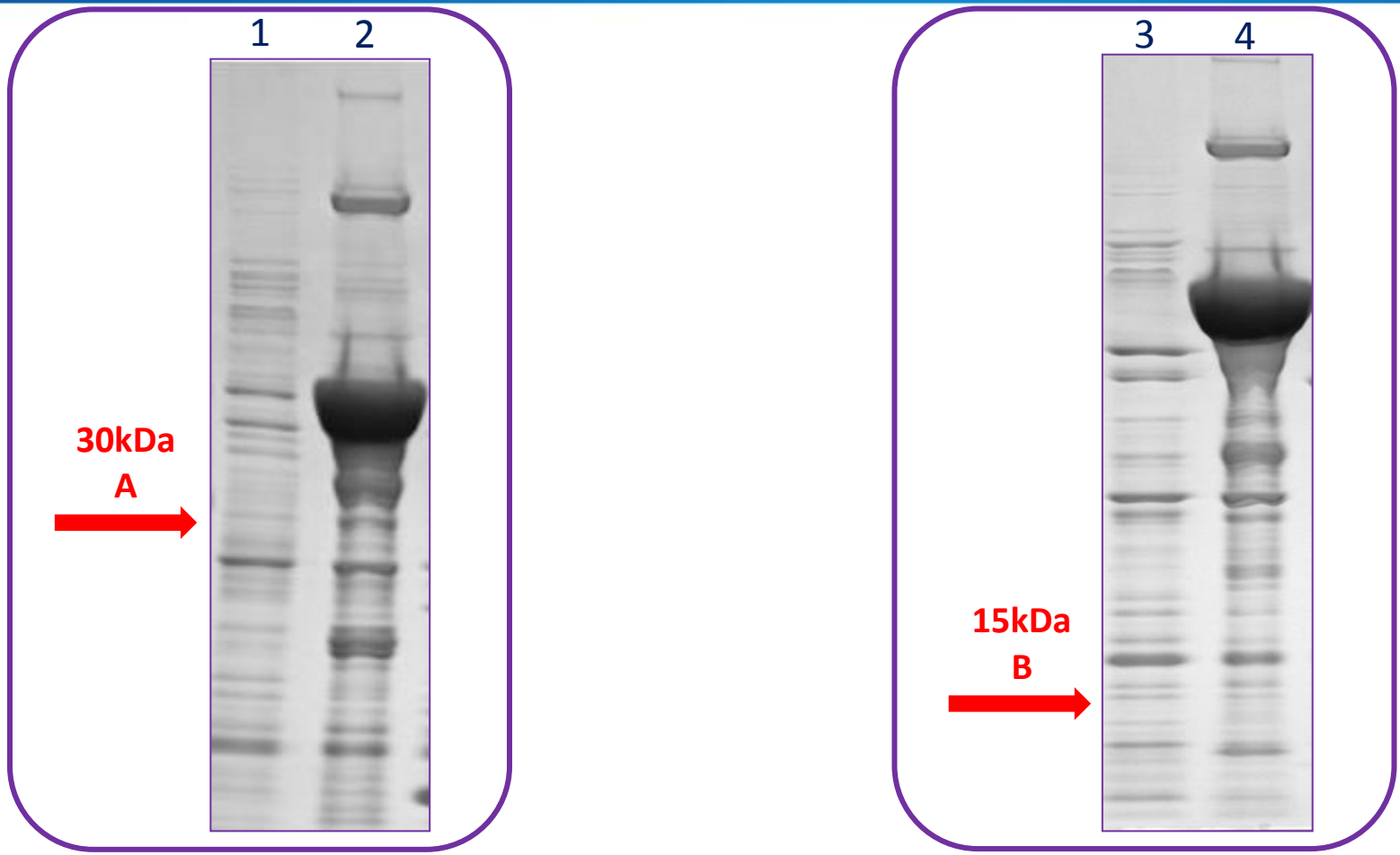
Factors critical to solubility



Fusion partner/affinity tag

Tag	~MW [kDa]	Affinity matrix	Pros	Cons
His	≤1	Nickel	Universal, small, works with native and denaturing conditions	Usually requires more than 1-step elution. Not entirely innocuous
CBP	4	Calmodulin	High specificity	Purity differs greatly
DsbA	23	N/A	Helps with disulfide bond formation	Periplasmic
Flag	1	Anti-Flag	High purity Introduces Enterokinase	Expensive and elution can be a problem occasionally
GST	26	Glutathione	Elution easy, functions as solubilizing tag	Dimerizing nature Leaches occasionally
MBP	43	Amylose	Solubility enhancer	Need longer contact times, Relatively large tag
NusA	55	N/A	Solubility enhancer	Very large tag
Strep	≤1	Streptadivin	High specificity, mild elution conditions	Need longer contact times
SUMO	12	N/A	Solubility enhancer, high fidelity cleavage	Needs affinity tag
Trx	12	N/A	Solubility enhancer	Does not work well with larger MW target proteins

Case study: Importance of fusion partner



FUSION PARTNERS MAKE A DIFFERENCE

GST tag

- ~26kDa
- Solubility enhancer
- *Schistosoma japonicum* parasitic worm
- N-terminus
- Can be used for *E. coli*, mammalian, insect and even yeast expression
- Affinity chromatography to immobilized glutathione
- Eluted with reduced glutathione under non-denaturing conditions
- Resin cheap, high binding capacity
- Non-denaturing, non-reducing conditions
- Easy detection using anti-GST antibodies
- Forms homodimers
- Slow binding hence lowered flow rates

NusA tag

- ~55kDa
- Large, hydrophilic
- Solubility enhancer based on a statistical solubility model
- Transcription anti-termination factor (NusA)
- No affinity matrix so combinatorial tagging is required
- Performed equal to or better than the MBP

Thioredoxin tag

- ~12kDa
- Ubiquitously found from bacteria to mammals
- Functions in cellular proliferation, redox signaling and the inhibition of apoptosis
- Can accumulate to 40% of total cellular protein while remaining fully soluble – translated very efficiently
- Robust folding characteristics make it a successful fusion partner
- N-terminal
- Simple isolation of fusion protein
- Possible that Trx may serve as a covalently joined chaperone protein by keeping folding intermediates of linked heterologous proteins in solution long enough for them to adopt their correct final conformations

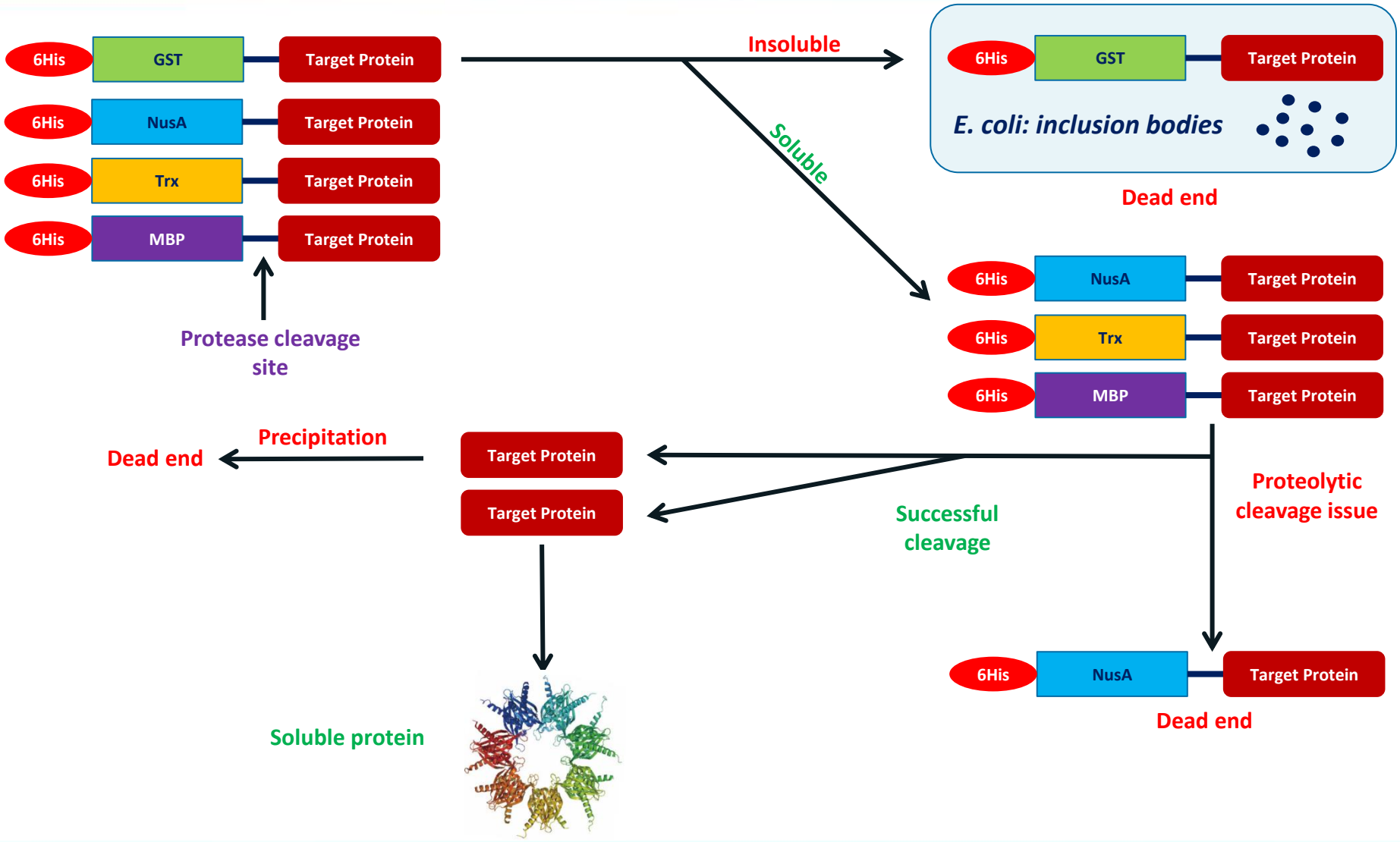
MBP tag

- ~43kDa
- *mal*E gene responsible for the uptake, breakdown and transport of maltodextrin carbohydrate
- Can increase the solubility of over-expressed fusion proteins in *E. coli*, especially of eukaryotic origin
- Common approach is to fuse it to N-terminus of target protein
- Generally used in combination with a small affinity tag
- Stabilizes and protects downstream passenger protein from proteolytic degradation during and after protein synthesis
- Efficient translation initiation
- Might act as chaperone by interactions through a solvent exposed hotspot on its surface which stabilizes the otherwise insoluble passenger protein
- Affinity purification using cross-linked amylose resin
- Typically eluted under non-denaturing conditions using maltose

Disadvantages of fusion partner approach

- Target protein often needs to be removed
- Removal of C-terminal tag tricky
- Target protein cleavage can be an issue due to steric hindrance
- In some cases a certain resin can be fragile and relatively expensive [amylose, FLAG]
- Target protein precipitation post-cleavage [often addressed by optimizing protein purification conditions]
- Some fusion proteins do not bind efficiently to the resin and even when they do, yields can be an issue [occasionally addressed by optimizing protein purification conditions or by change in affinity purification strategy – using His tag instead]
- Aggregation post-cleavage

Multi-tag approach for soluble protein production



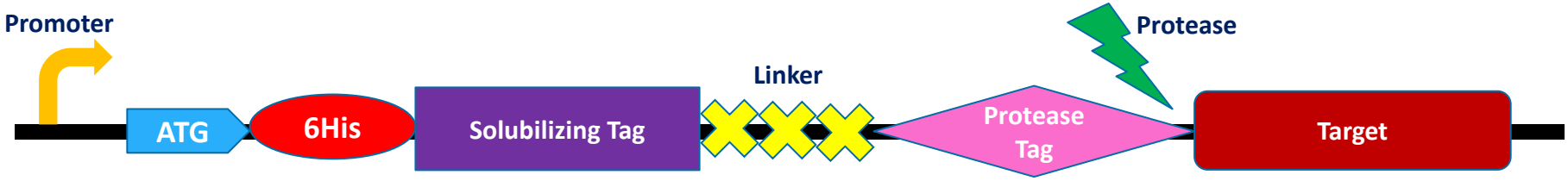
Factors to consider

- **Cloning Strategy**
  - Use LIC or recombination-based cloning strategy for higher throughput & parallel construct generation
  - If possible, create expression vectors with N-terminal fusion/solubility partners for easy cloning scheme
- **Expression**
  - Express both N- and C-terminally affinity tagged constructs [especially His]
  - In many cases at least one version will express
  - 

    If checking for enzyme activity, one version might show higher activity than the other
- **Purification**
  - Note that C-terminally affinity tagged constructs allow you to purify full length proteins
  - You could try double-tagging approach by using one affinity tag [e.g., GST, MBP at the N-ter and 6His at the C-ter to solubilize and purify full length protein at the same time

Conclusion & takeaways

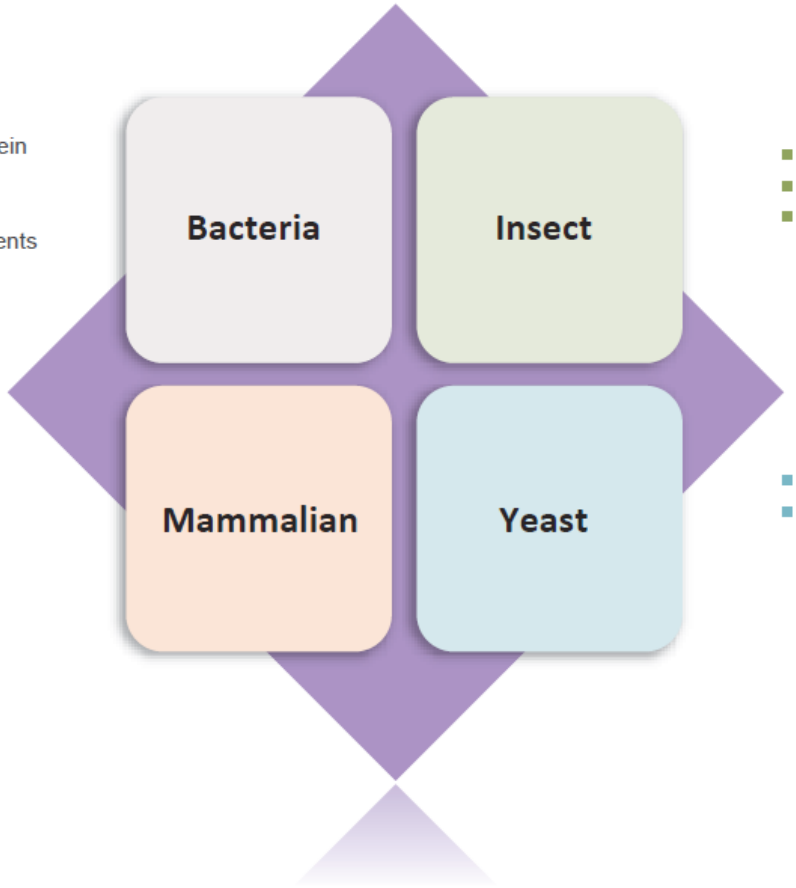
- No simple, global solution to address insolubility issues in *E.coli* expression
- Fusion partners make a difference
- Parallel approaches increase the chance of success
- Suggested approach –
  - Combinatorial tagging at the N-terminus [6His + solubilizing tag]
  - Protease Cleavage site for tag removal after 1st round purification
  - Linker before and/or after protease tag to improve chances of proteolytic cleavage



- Expression levels, activity, purity, homogeneity, stability are important factors that must be considered

GenScript protein services

High quality recombinant proteins produced in a choice of 4 expression systems



- BacPower™ guaranteed protein
- BacPower™ custom protein
- Fast gene-to-protein™
- FragPower™ antibody fragments
- PROTential™ expression

- InsectPower™ guaranteed protein
- BacuVance™ custom protein
- GLP-level insect expression

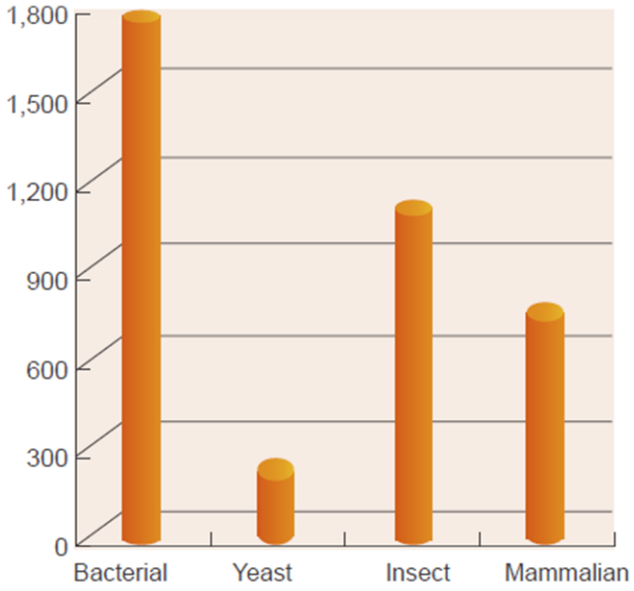
- MamPower™ protein
- MamPower™ antibody
- Transient custom expression
- Stable cell line

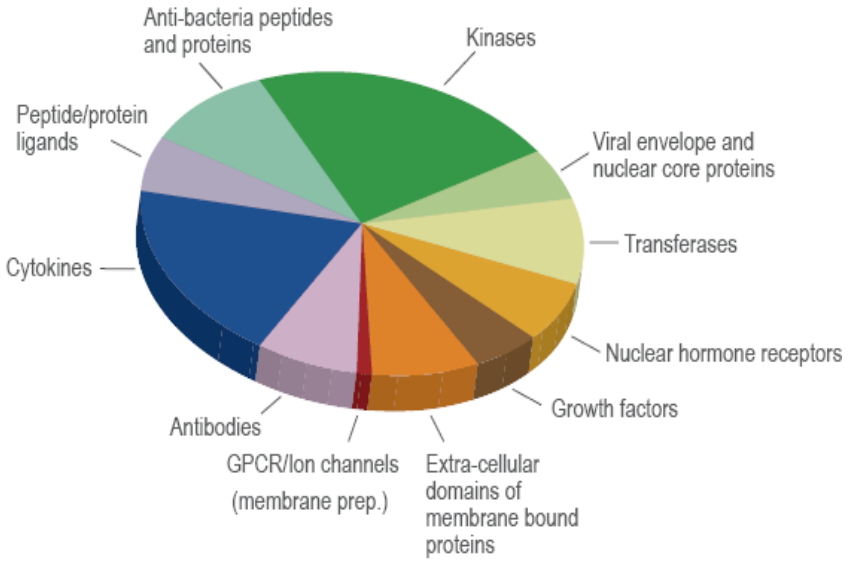
- YeastHIGH™ protein expression
- Large scale production

Email protein@genscript.com with questions or get a quote online

- Host variety
- Purity & tag options
- Major services are comparable or better value than most leading competitors'
- Gene synthesis included in most of our packages
- Refolding [if required] included
- Many guaranteed packages
- Additional protein cost lower
- Industry's fastest protein service
- High Throughput services with our HTP variants service
- Technical support

Experience in protein expression & purification





- GenScript has delivered over **5,300** proteins in four expression systems with **92%** success rate for all protein projects
- GenScript has successfully delivered a variety of difficult proteins, including trans-membrane proteins, co-expression proteins, proteases, kinases, cytokines, antibodies and several other proteins to customers worldwide

Thank you for your participation

We wish you all success in your research

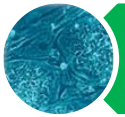
keshav.vasanthavada@genscript.com

Other archived webinars in the GenScript Webinar Series @ http://www.genscript.com/webinars.html

Peptide design strategy: basics, optimization, and application - *Tiffany Gupton Campolongo, Ph.D.*

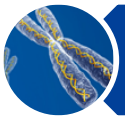
Codon optimization: Why & how to design DNA sequences for optimal soluble protein expression - *Rachel Speer, Ph.D.*

Anti-idiotypic antibodies – A powerful tool for antibody drug studies – *Michelle Parker, Ph.D.*



Recombinant protein applications: choosing the best cytokine, growth factor or chemokine for your experimental system - *Matthew Riolo, Ph.D.*

Identify the optimal protein purification strategy for your recombinant protein production - *Liyan Pang, Ph.D.*



Design high specificity CRISPR/Cas9 gRNAs: principles and tools – *Heidi Huang, PhD*

Biologics drug discovery: Steps to achieving an antibody drug candidate – *Frank Fan, MD, PhD*