THE HEBREW UNIVERSITY OF JERUSALEM

The Protein Production Facility
A Training and Research Center

Dr. Tsafi Danieli

Session 1

Overview of recombinant proteins production

Introduction to heterologous expression in E.coli

Focus on Industry vs. Academia Requirements

- Documentation

Standard laboratory practice and documentation in ELNs and SOP logbooks

- Vectors

IP considerations, academic use vs. commercial contracts

- Genes

synthetic vs. natural, fusion tags and proteins (regulations and production considerations)

- Cell lines and strains

Verified and documented for cell-based assays, known cell banks and repositories, certified for therapeutic production; or

- Regents

Documented reagents, cell bank, animal-free process

Rules and regulations

Document and register all procedures and reagents

- Academic procedures and practices will not be suitable for production of biologics
- Research in academic laboratory will be sufficient for proof of concept if the documentation was done under SOP conditions
- Easy to translate and commercialize
- Recommended and mandatory practice or basic research, publications and reproducibility

Strategy depends on the required application

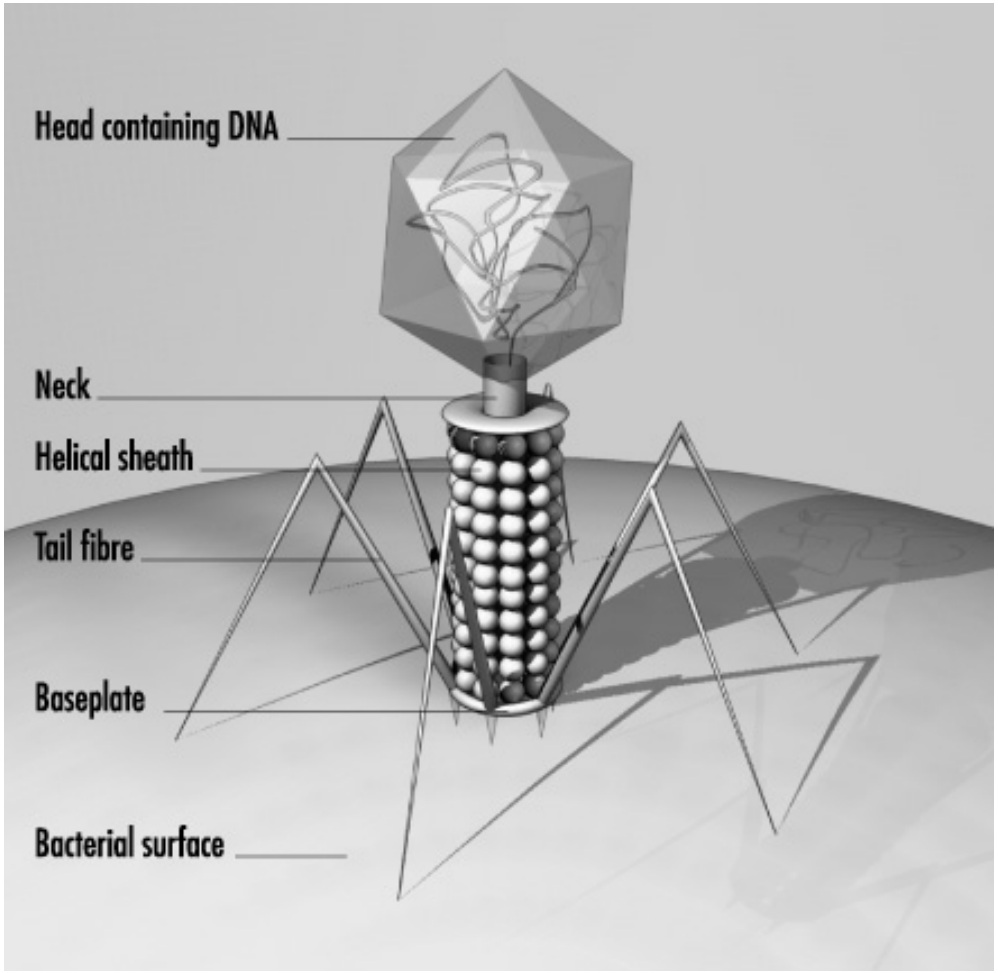
- Activity assays
- Antibodies production
- Protein-protein interaction
- Structure determination
- Therapeutics (Biologics)

Native proteins are often difficult to study

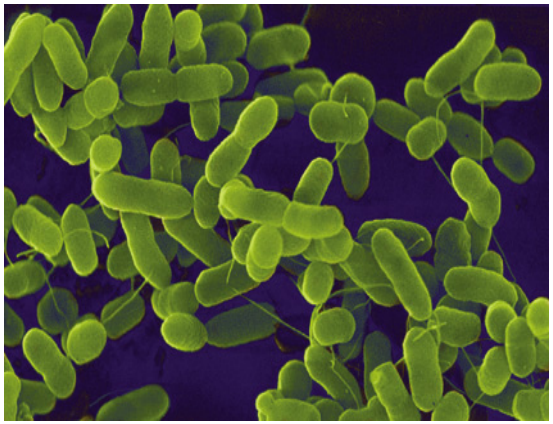
- Limited amounts
- Source unavailable
- Unstable
- Difficult to isolate

Cell Factories

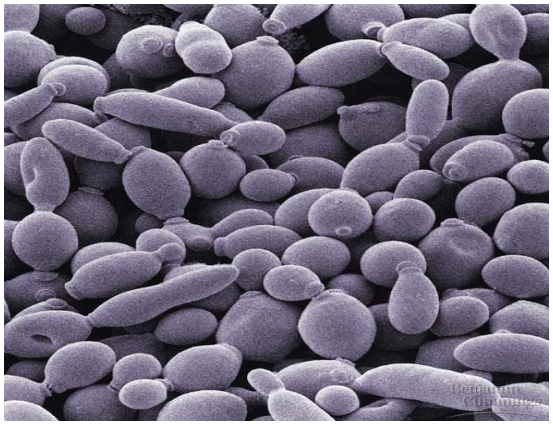
Natural Solutions



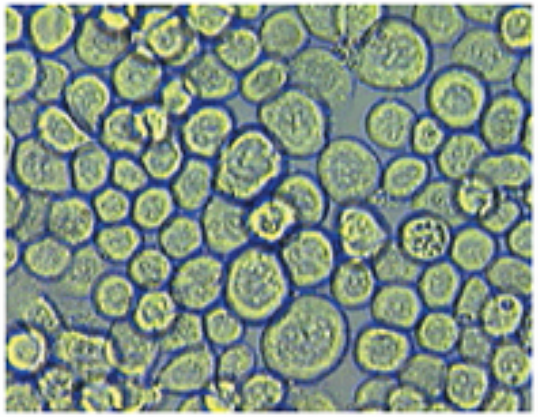
Cell Factories



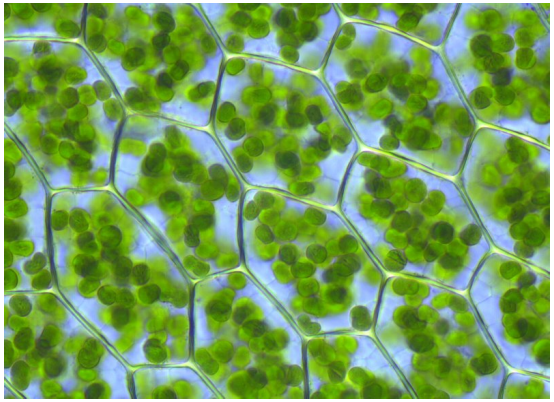
Bacteria E. *coli*



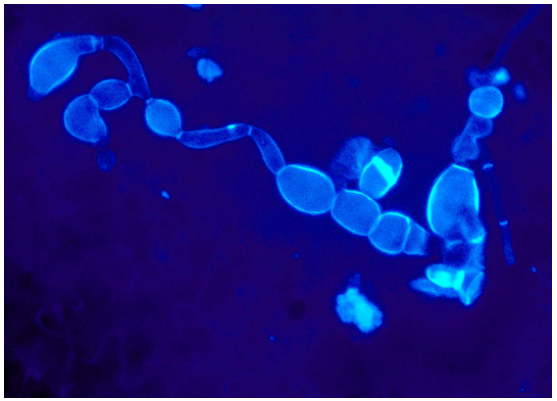
Yeast *s.cerevisiae*



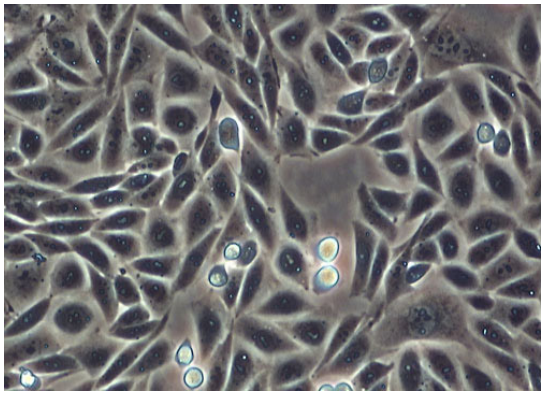
Insect SF9



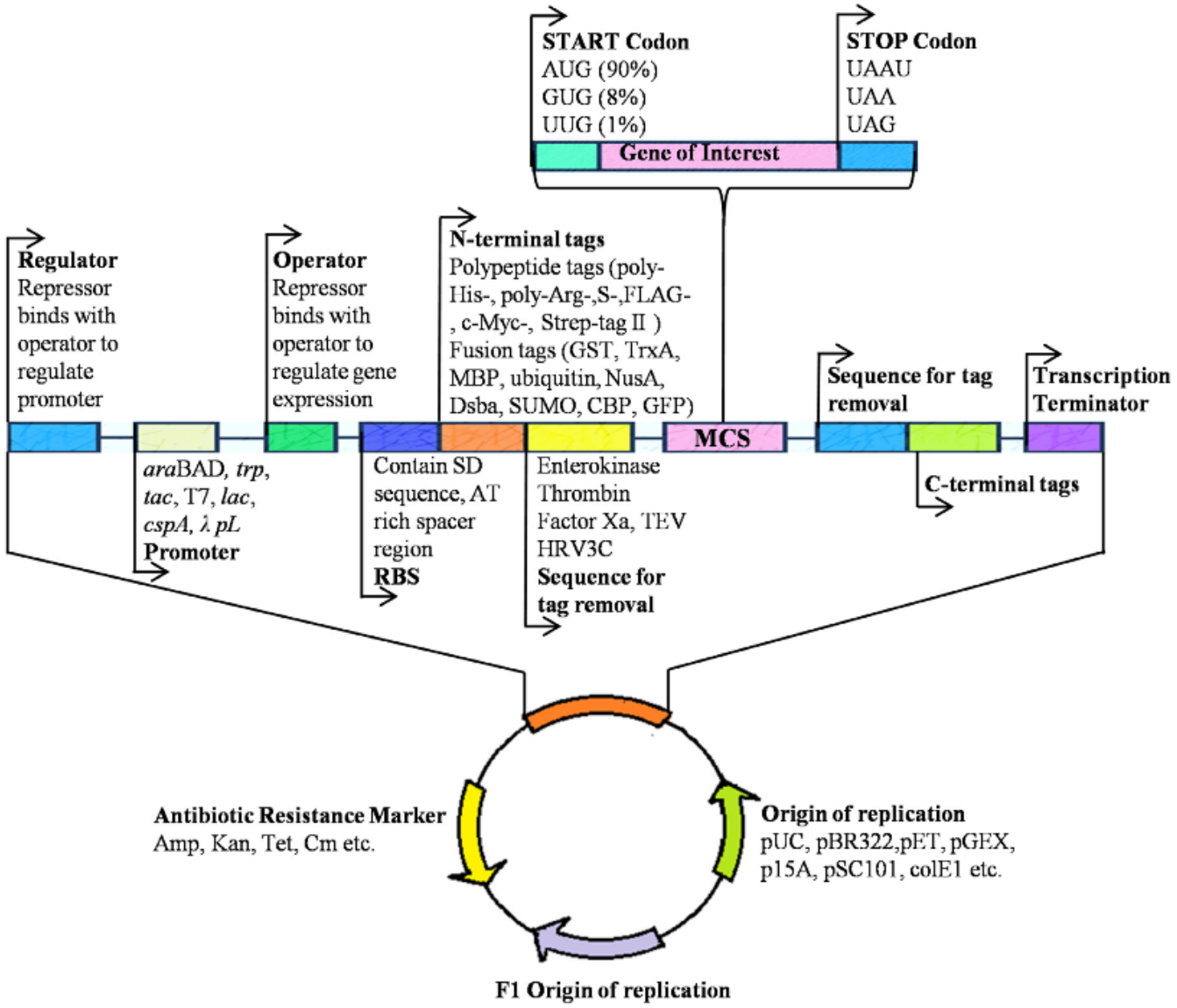
Plant cells



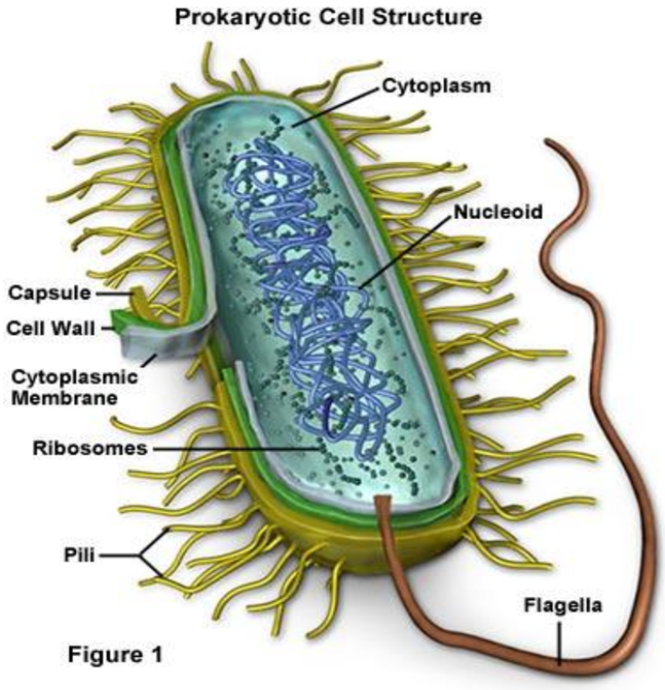
Bacteria B. *subtilis*

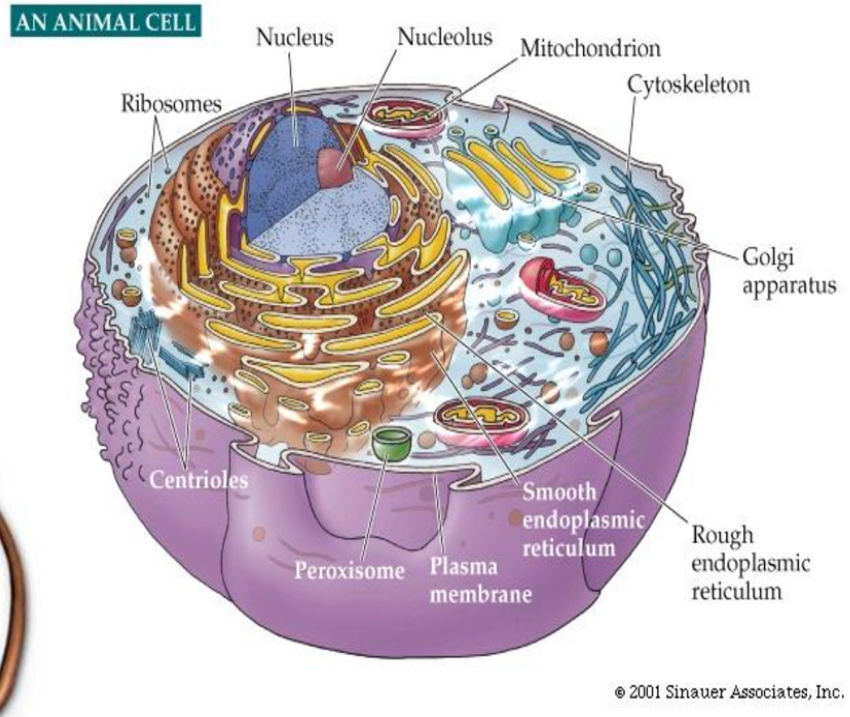


Mammalian CHO


START Codon
AUG (90%)
GUG (8%)
UUG (1%)
STOP Codon
UAAU
UAA
UAG
Gene of Interest
Regulator
Repressor binds with operator to regulate promoter
Operator
Repressor binds with operator to regulate gene expression
N-terminal tags
Polypeptide tags (poly-His-, poly-Arg-,S-,FLAG-, c-Myc-, Strep-tag II)
Fusion tags (GST, TrxA, MBP, ubiquitin, NusA, Dsba, SUMO, CBP, GFP)
Sequence for tag removal
Transcription Terminator
MCS
araBAD, trp, tac, T7, lac, cspA, λ pL
Promoter
Contain SD sequence, AT rich spacer region
RBS
Enterokinase
Thrombin
Factor Xa, TEV
HRV3C
Sequence for tag removal
C-terminal tags
Antibiotic Resistance Marker
Amp, Kan, Tet, Cm etc.
Origin of replication
pUC, pBR322,pET, pGEX, p15A, pSC101, colE1 etc.
F1 Origin of replication

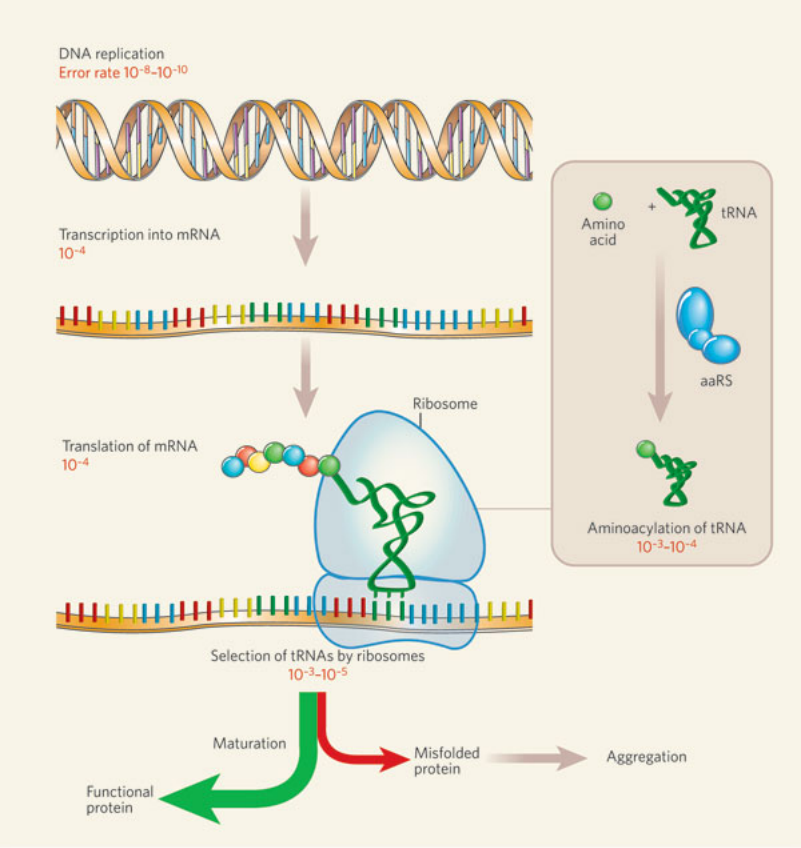
Prokaryotes vs. Eukaryotes





The Protein Pathway in the Host Cell

- Transcription
- Translation
- Modification
- Folding
- Localization
- Interactions
- Degradation



Learn as much as possible about your protein before starting the project

- Eukaryotic/Prokaryotic?
- Secreted/Cytosolic/membrane?
- Contains rare codons?
- Contains cysteines?
- Might be toxic to the host cell?
- Degradable?
- Requires post-translational modifications for activity?
- Part of a complex?

Productivity comparison

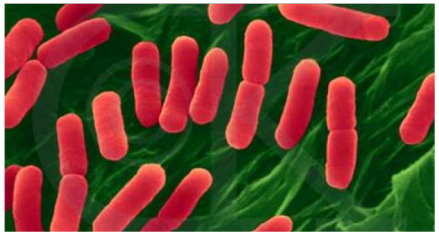
Table I: Comparison of cell-based protein expression systems

Expression System	Ease of Handling and Scale-Up*	Protein Expression Level	Cytotoxic Mammalian Proteins	Percent Yield (Based on Dry Weight)	PTMs†	Applications
Bacterial	****	Up to 10–30 g/L	Yes	1–5%	+	Functional assays Structural analysis Antibody generation Protein interactions
Yeast	***	Up to 30 g/L	Yes	1%	++	Functional assays Structural analysis Antibody generation Protein interactions
Insect	**	Up to 500 mg/L	Yes	30%	+++	Functional assays Structural analysis Antibody generation
Mammalian	*	Under 10 mg/L		<1%	++++	Functional assays Protein interactions Antibody generation
Cell free	*	1–3 mg	Yes	N/A	++	Functional assays Protein interactions

*Most difficult handling: ****; easiest handling: *.
†Very minimal PTM: +; PTM the closest to that in naturally occurred proteins: ++++.

Post Translational Modifications

Post translational modifications	**E. coli**	**Yeast**	**Insect cells**	**mammalian**
N-Acetylation	Yes	Yes	Yes	Yes
Amidation	No	Yes	Yes	Yes
γ–Carboxylation of glutamate	No	No	No	Yes
N-glycosylation	No	Yes	Yes	Yes
O-glycosylation	No	Yes	Yes	Yes
Heterodimer	Yes	Yes	Yes	Yes
Hydroxylation	No	Yes	Yes	Yes
Myristoylation	Possible	Yes	Yes	Yes
Palmitoylation	No	Yes	Yes	Yes
Phosphorilation	Yes	Yes	Yes	Yes
Protein proteolytic process	Signal pept.	Yes	Yes	Yes
Sulfation	No	No	Yes	Yes
N-term Met removal	Partial	Partial	Yes	Yes



E. *coli*

Common applications:

- Proteins production for raising polyclonal antibodies
- Production of non-glycosylated therapeutic proteins
- Structure determination
- Multiple research applications

Pros:

Highest quantity, low cost
Well defined, many vectors and fusion tags

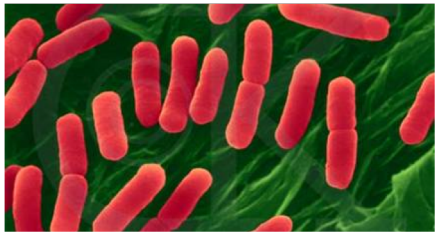
Cons:

Lack of / Different Post translational modifications (PTMs)
Best suited for heterologous proteins smaller than 80kDa,
Often not suited for membrane and secreted proteins
Contains endotoxins

Best candidates for E.coli expression

Best expressed when:

- protein is 80kDa and smaller
- Not part of a complex
- Cytosolic origin (not secreted)
- Known to be soluble
- Single globular domain
- Does not require glycosylations
- If gene if from a different species: check codon bias (order codon optimized gene)



Expression in E. *coli*

- Expression vectors for E.coli
- Gene of interest elements
- Host strains
- Growth conditions

Four required elements for success

Vector

Environment

Gene

Host

Vector

Vector

Growth

conditions

Gene

Host

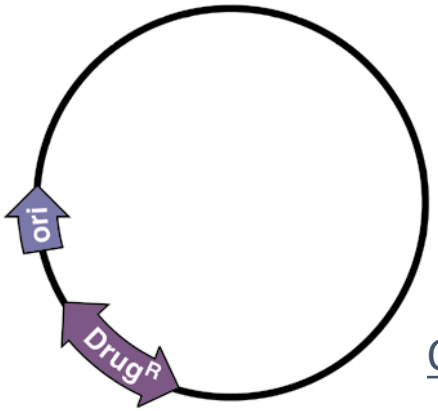
Plasmid Features

E. coli expression vector contains the following elements:

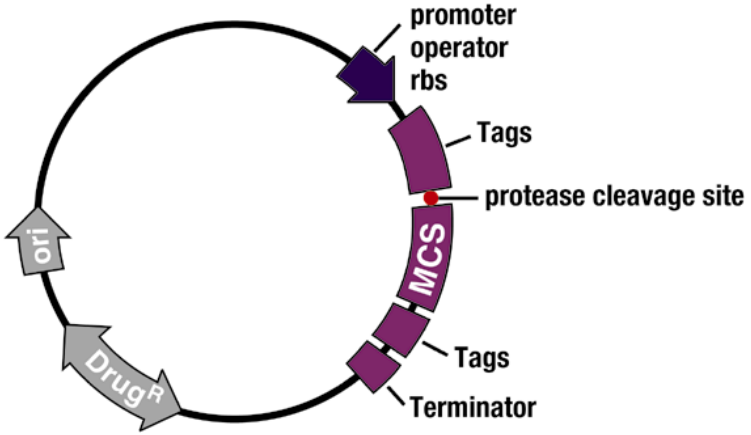
1. Selectable Marker: to ensure maintenance of the vector in the host cell (bla-ampicilin, cat-chloramphenicol, tet-tetracycline, Kanamycin, zeocin etc.)

2. Origin of replication: for independent, extra chromosomal replication .
ColE1 replicon has 10-200 plasmids per cell

3. Promoter: controllable transcriptional promoter, which can be induced for direct production of large amounts of mRNA of the cloned gene.



Cloning vector



Expression vector

Commonly used promoters

Commonly used promoters for the production of recombinant protein along with their properties.

Promoter	Source	Regulation	Induction methods	Limitations
lac	*E. coli*	*lacI, lacIq*	IPTG, Thermal	Expression level low and leaky expression; Lactose cannot be used as inducer in presence of glucose.
trp	*E. coli*	unknown	Trp starvation, IAA	Leaky expression
araBAD	*E. coli*	*araC*	L-Arabinose	Low Availability of vectors; Catabolic repression by glucose
tac, hybrid	*E. coli*	*lacI, lacIqlacId*	IPTG, Thermal	Leaky expression
λ pL		*1 cIts 857*	Thermal	High temperature is required for induction
T7-*lac* operator	*T7*	*LacI q*	IPTG	Leaky expression
phoA	*E. coli*	*phoB, phoR*	Phosphate starvation	Media limitations
cspA	*E. coli*	unknown	Reduced temperature [<200C]	Low temperature causes slow growth of cells.
recA	*E. coli*	*lexA*	Nalidixic acid	Not titratable
trc, hybrid	*E. coli*	*lacI, lacIqlacI[Ts]a, lacIq[Ts]a*	IPTG, Thermal	
cadA	*E. coli*	*cadR*	pH	Uncertainty of induction timing; pH for induction and optimum pH for growth can be far apart.
proU	*E. coli*		Salt	Consumption of salt by cell during growth.

Lac Repressor

Promoter systems:

Lac

Tac

Trc

Also in:

T5

T7

cspA

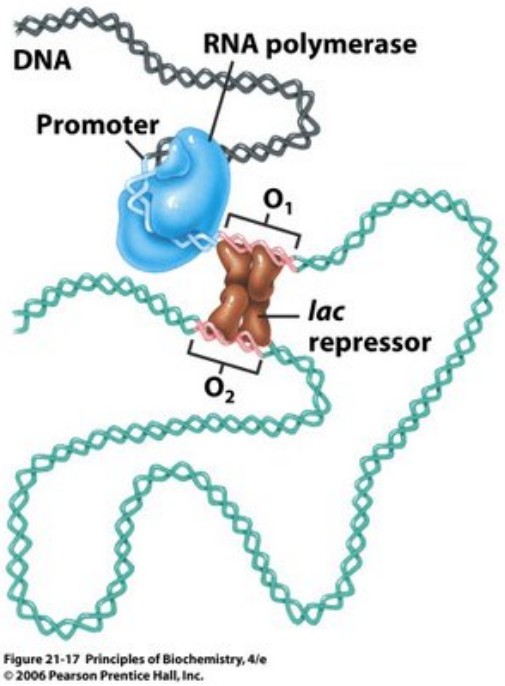


Figure 21-17 Principles of Biochemistry, 4/e
© 2006 Pearson Prentice Hall, Inc.

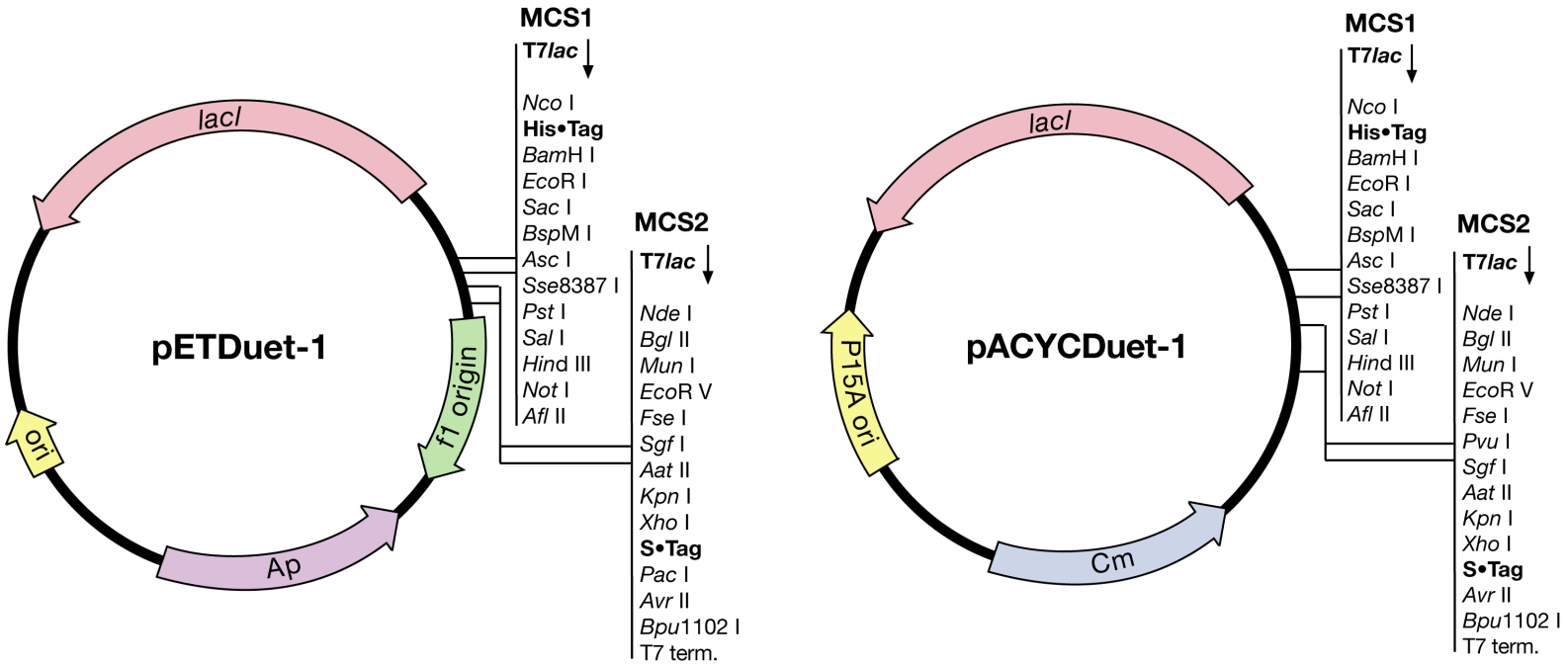
Nature Reviews Microbiology 6, 507-519 (July 2008)

Plasmid Features: compatible ori's

Common plasmid vectors, their origin of replication and copy number.

Common plasmid vectors	ORI	Copy number type
pBR322	pMB1	Low copy (15–20)
pUC	pMB1	High copy (500–700)
pColE1	ColE1	Low copy (15–20)
pBluescript	ColE1	High copy (300–500)
pR6 K	R6 K	Low copy (15–20)
pACYC	p15A	Low copy (10–12)
pGEM	pUC	High copy (300–500)
pGEX	pBR322	Low copy (15–20)
pET	pBR322	Low copy (15–20)
pSC101	pSC101	Low copy (5)
pBAD	pUC	Low copy (10–12)

Plasmids for co-expression



Gene

Vector

Gene

Growth

conditions

Host

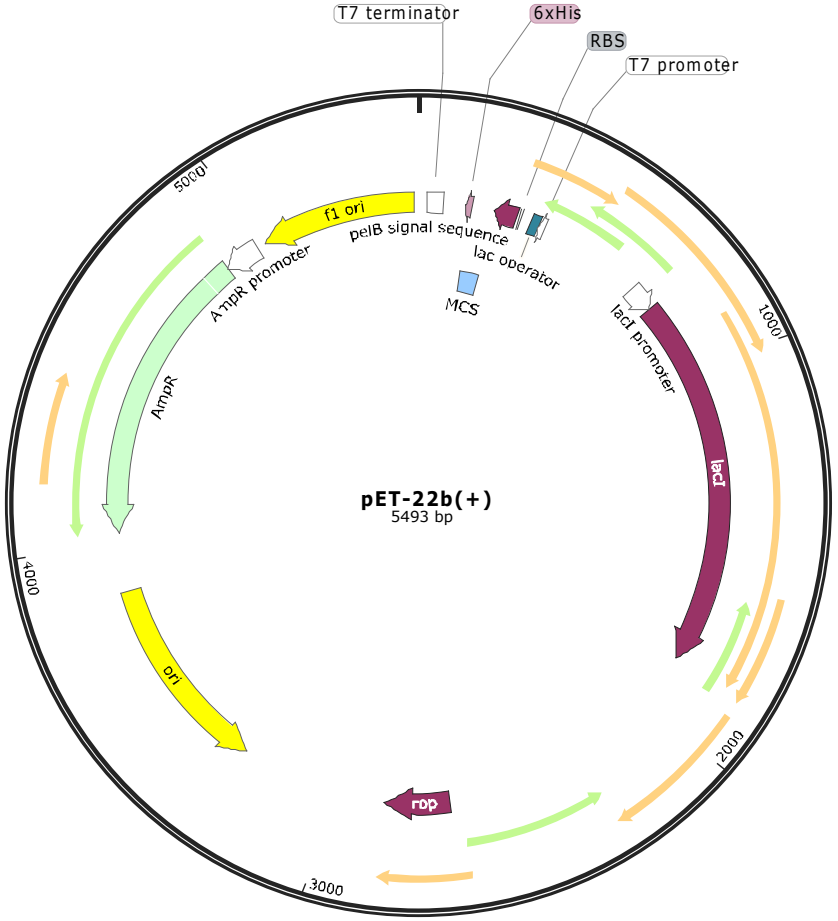
Check sequence of obtained construct

- Check for full ORF and codon suitability
- Distance of initiation ATG from RBS
- GC rich regions following ATG
- Standard procedure: order synthetic gene in vector or for Gibson cloning

Use DNA analysis tools

Tool Sets for DNA Manipulation

1. SnapGene
2. everyVECTOR
3. pDRAW32 DNA analysis software
4. DNA Strider (for Mac users)
5. EnzymeX (for Mac users)
6. APE - A plasmid editor
7. Sequence Manipulation Suite
8. CLC free workbench



http://wolfson.huji.ac.il/expression/rec-dna-web-tools.html

Fusion proteins:

- Increase solubility
- Increase productivity
- Support folding
- Reduce toxicity
- Improve activity
- Facilitates purification

Solves problems such as:

- Aggregation due to misfolding and hydrophobicity
- Low production levels due to poor translation initiation
- Purification of target protein from bacterial contaminants
- Degradation

Fusion Proteins short list:

Fusion protein	Purification method
MBP(can be with N-term His tag)	Amylose/nickel binding
SUMO(with N-term His tag)	Nickel purification
GST	Glutathione binding
Thioredoxin (with N-term His tag)	Nickel Binding
DsbA/C(might be with N-term His tag)	Osmotic shock/Nickel Binding
Lipoyl domain (with N-term His tag)	Nickel column
CBD	Cellulose binding

Fusion protein cleavage methods:

Signals can be removed by chemical or enzymatic methods.
Removal of fusion proteins with cyanogen bromide, or hydroxylamine is effective, but often requires extreme conditions (low pH or high temp.), and is often non-specific.

Enzymatic digestion is the method of choice for soluble fusion protein cleavage. reaction is carried out under relative mild conditions.

Enterokinase:	Asp Asp Asp Asp Lys |
Factor Xa:	Ile Glu/Asp Gly Arg |
TEV:	Glu Asn Leu Tyr Phe Gln |Gly
Thrombin:	Leu Val Pro Arg |Gly Ser
Sumo protease	Full length sumo protein |Gly

Fusion partner should be compatible with the expression problem!

Hydrophobic proteins require highly soluble fusion partner

Maltose-binding protein (MBP)
Glutathion S-tranferase (GST)
Sumo protein
These genes are well expressed, highly soluble, and provide specific characteristics to aid purification.

Problems:

1. Cleavage of fusion partner may cause re-aggregation!
2. Removal of fusion proteins reduces productivity and can be very costly (not suitable for therapeutics)
3. N-terminus partner might cause pre-mature termination

Host

Vector

Growth

conditions

Gene

Host

Host cells for T7 promoters

The pET vector system for protein production.

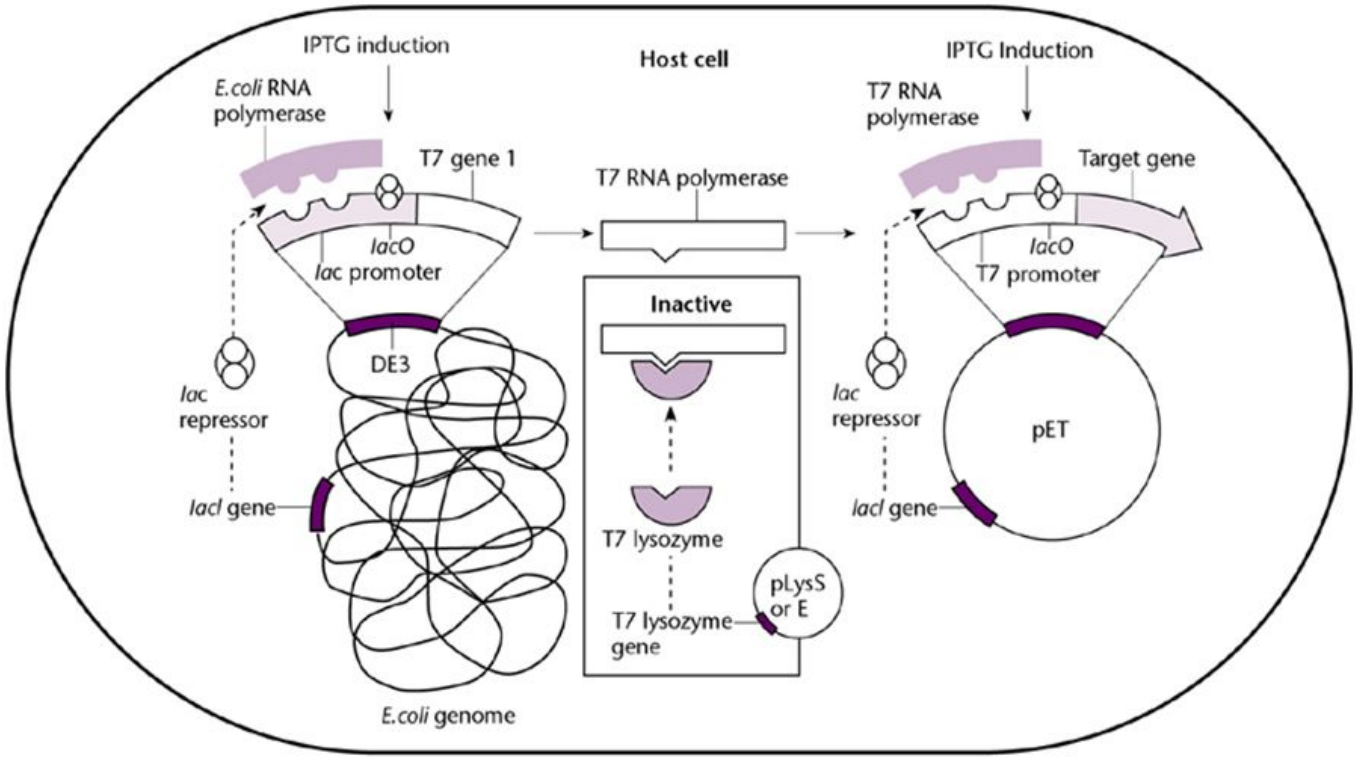
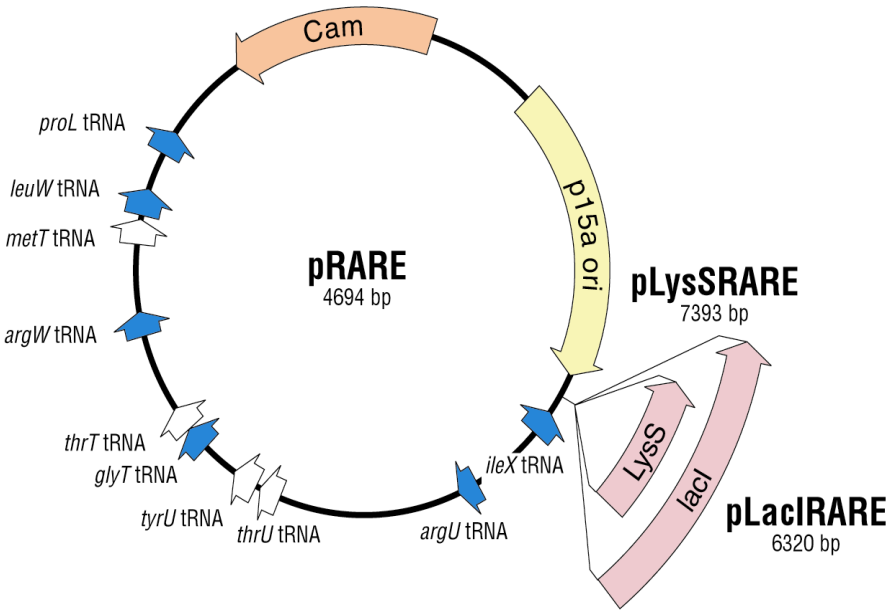


Table 1. Arg, Gly, Ile, Leu and Pro codon usage in *E. coli*

amino acid	codon	fraction in all genes	fraction in Class II
Arg	AGG	0.022	**0.003**
Arg	AGA	0.039	**0.006**
Arg	CGG	0.098	**0.008**
Arg	CGA	0.065	**0.011**
Arg	CGU	0.378	0.643
Arg	CGC	0.398	0.330
Gly	GGG	0.151	0.044
Gly	GGA	0.109	**0.020**
Gly	GGU	0.337	0.508
Gly	GGC	0.403	0.428
Ile	AUA	0.073	**0.006**
Ile	AUU	0.507	0.335
Ile	AUC	0.420	0.659
Leu	UUG	0.129	0.034
Leu	UUA	0.131	0.055
Leu	CUG	0.496	0.767
Leu	CUA	0.037	**0.008**
Leu	CUU	0.104	0.056
Leu	CUC	0.104	0.080
Pro	CCG	0.525	0.719
Pro	CCA	0.191	0.153
Pro	CCU	0.159	0.112
Pro	CCC	0.124	**0.016**

Codon usage is expressed as the fraction of all possible codons for a given amino acid. "All genes" is the fraction represented in all 4,290 coding sequences in the *E. coli* genome (6). "Class II" is the fraction represented in 195 genes highly and continuously expressed during exponential growth (7).

Changing codon bias



In strains such as BL21-Rossetta (Merck/Novagen);
BL21-Codon Plus (Agilent/Stratagene)
AKA: RIL/RP strains

Specialized Host strains

- Rare codons (Rosetta, CodonPlus)
- Disulfide bonds (origami, shuffle)
- Membrane proteins (C41/C43)
- Labile proteins (low proteases)
- Toxic proteins (tight regulation: BL21(AI))
- DNA rearrangements (recA-,low copy Sure2, StbleIII, AbleC, mainly for cloning..)

http://wolfson.huji.ac.il/expression/bac-strains-prot-exp.html

Growth conditions

Vector

Growth
conditions

Gene

Host

Growth conditions may affect:

- Poor cell growth
- Poor expression
- Leaky expression
- Formation of insoluble protein
- Accumulation of truncated protein

Play with The Following Parameters:

1. Medium formulation (rich/minimal; additives)
2. Growth temperature
3. Induction conditions: IPTG levels/at ? OD
4. Induction/harvest time (post induction)
5. Adding supplements to reduce leakiness
6. Adding chemical chaperones
7. Aeration conditions (fermentors)

Common Problems in Protein Expression in E. coli

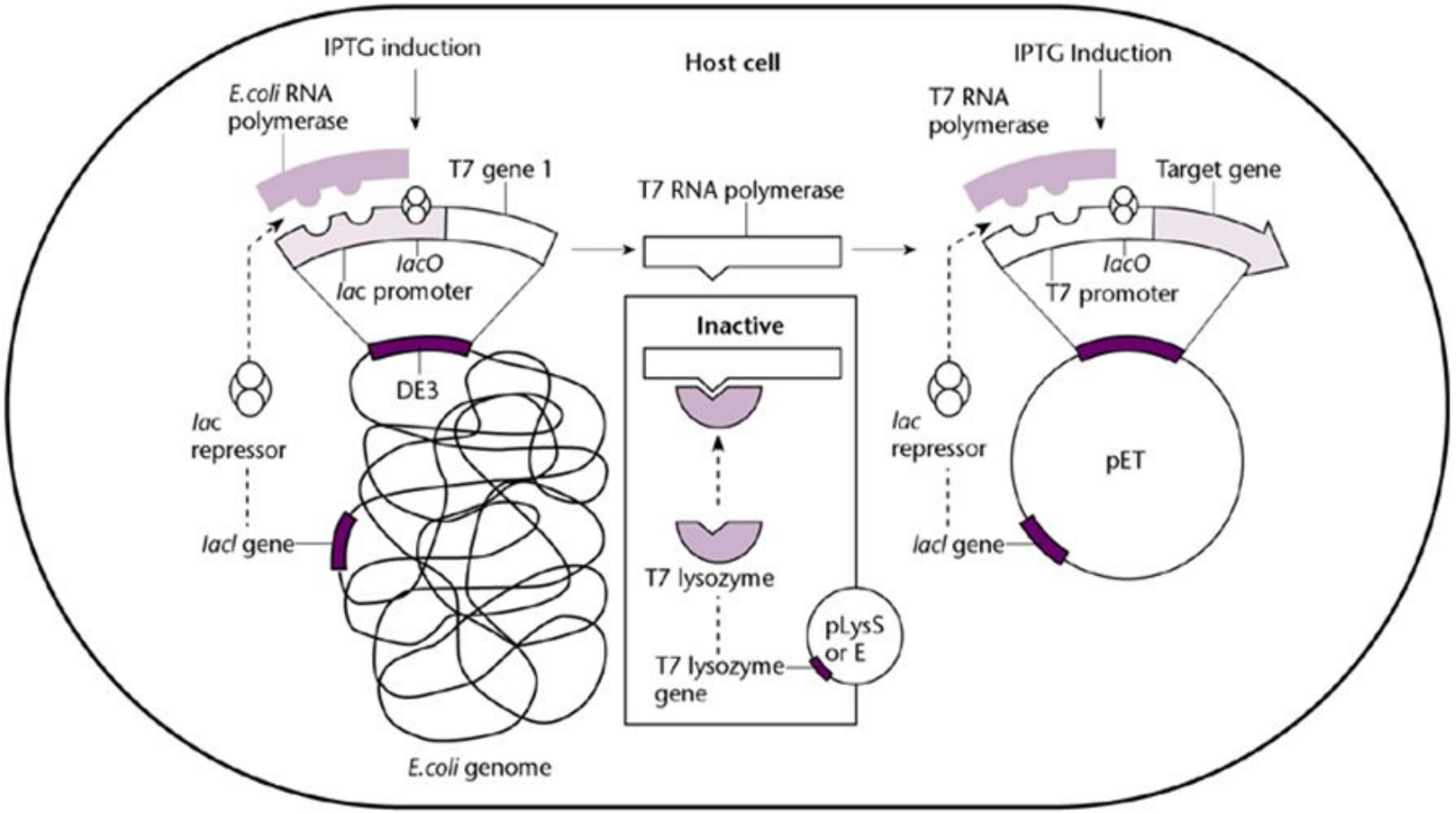
- Aggregations
- Irreproducibility
- Low Productivity
- Truncations
- Degradations
- Inactivity

Observation from the last 20 years of practice

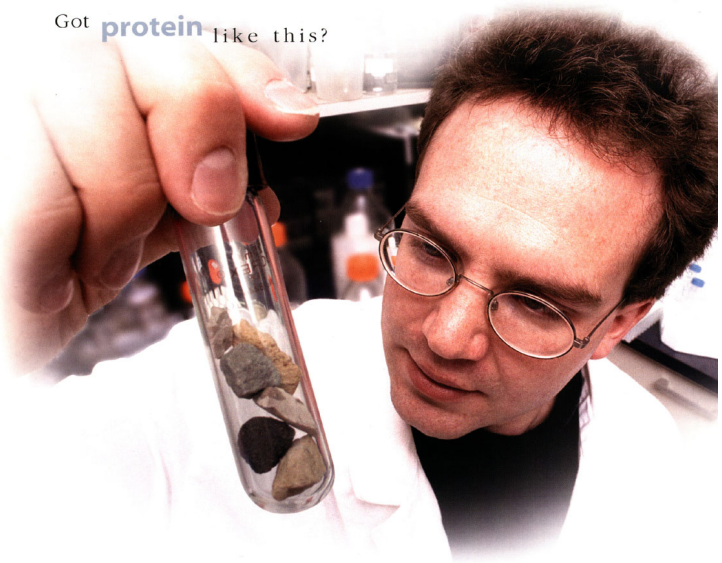
Problem: Irreproducibility and low productivity

Solution: Start from fresh transformation and NOT from glycerol stocks

Lose of transcription mechanism in host



Observation from the last 20 years of practice

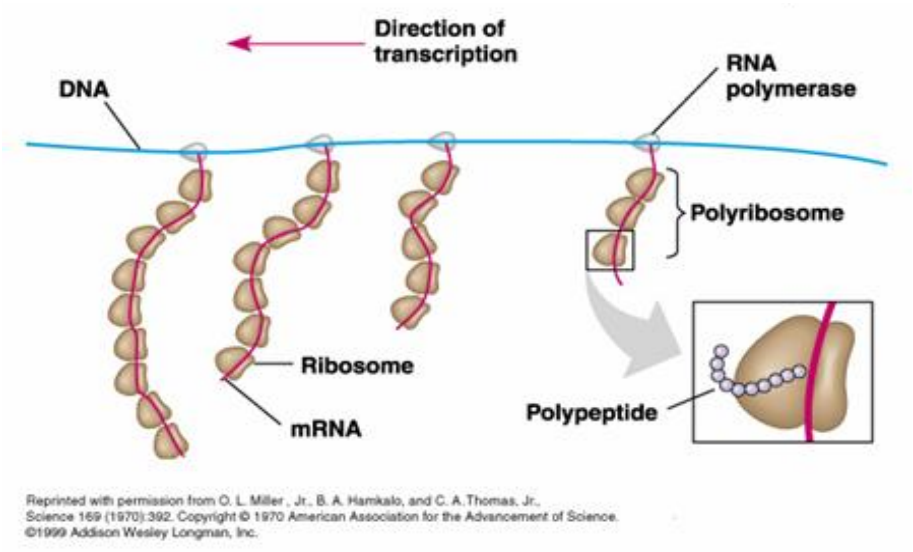


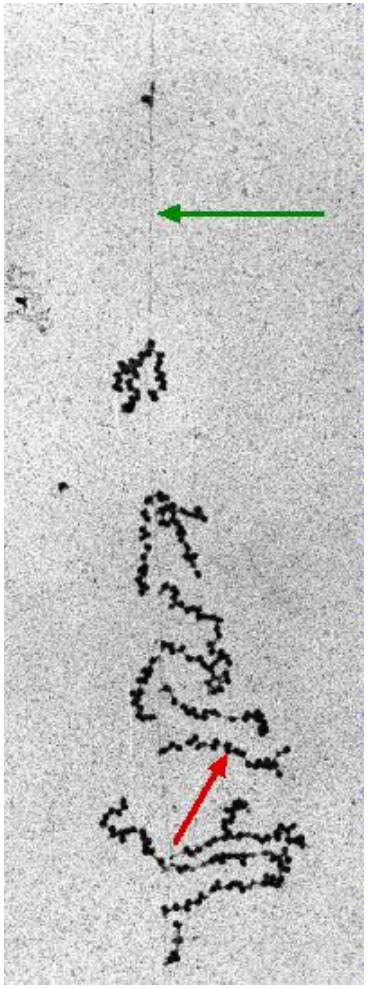
Problem: Aggregates and truncated proteins

Solution: Suppress leaky expression and reduce induction conditions

Coupled transcription translation and the polysome

- A single mRNA molecule usually has many ribosomes traveling along it, in various stages of synthesizing the polypeptide .This complex is called a **polysome**





Fusion proteins:

- Increase solubility
- Increase productivity
- Support folding
- Reduce toxicity
- Improve activity
- Facilitates purification

Solves problems such as:

- Aggregation due to misfolding and hydrophobicity
- Low production levels due to poor translation initiation
- Purification of target protein from bacterial contaminants
- Degradation

Strategies to reduce basal expression in T7 systems

1. Inhibit T7 polymerase with T7 lysozyme
2. Block target promoter with *lac* repressor

Lactose is a contaminant of tryptones from the media:

- Tryptones are tryptic digests of casein
- Casein comes from milk
- Milk contains lactose

Glucose prevents unintended induction!

Glucose Repression	Glucose represses basal expression levels by lowering cAMP levels which in turn decrease the binding of CAP. As cAMP levels are lowered, transcriptional activation is decreased. Add glucose to the culture medium to repress basal expression levels.

Inducing Expression

Classic method of induction:

- Cells are grown in log-phase culture
- IPTG is added at appropriate culture density

Rational for auto-induction:

- Growth in glucose and other carbon sources prevents induction by lactose
- Media with appropriate mixtures of carbon sources allow growth with minimal basal induction, followed by automatic induction at high density

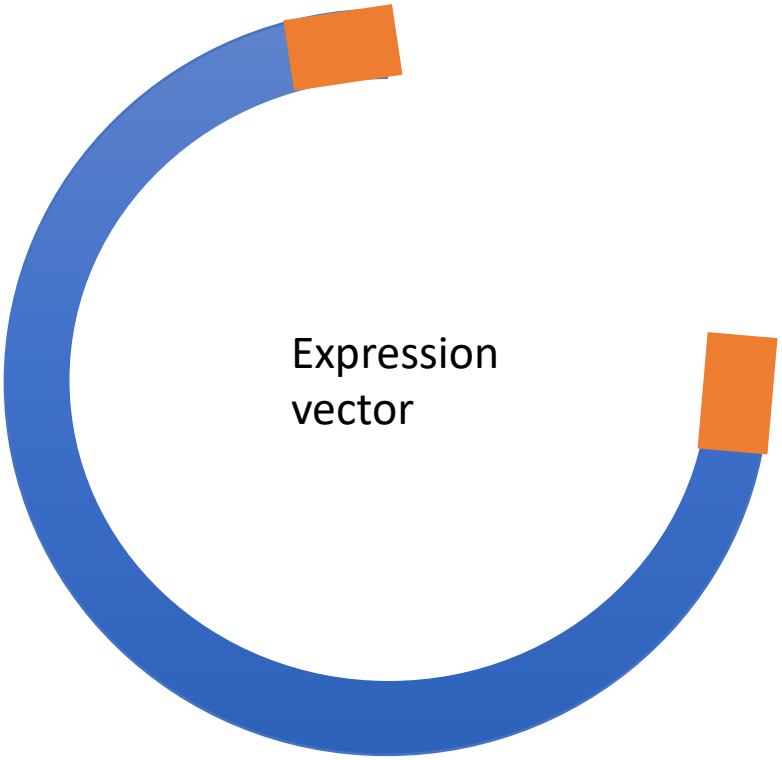
http://wolfson.huji.ac.il/expression/procedures/bacterial/auto-induced-exp.html

F. William Studier; in Protein Expression and Purification (2005) 41:207-234

Our minimal throughput screen

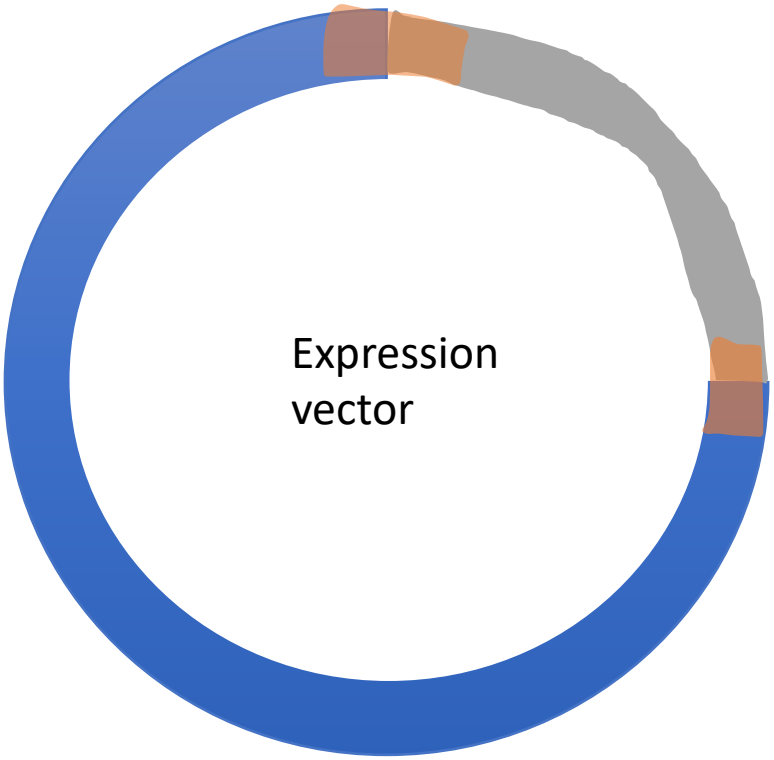
- Clone gene with different fusion partners
- Transform in parallel into several strains
- Pick colonies into an auto induction media
- Screen after over-night growth for Nickel binding

Cloning



Gibson / restriction cloning
PCR from template / Synthetic genes

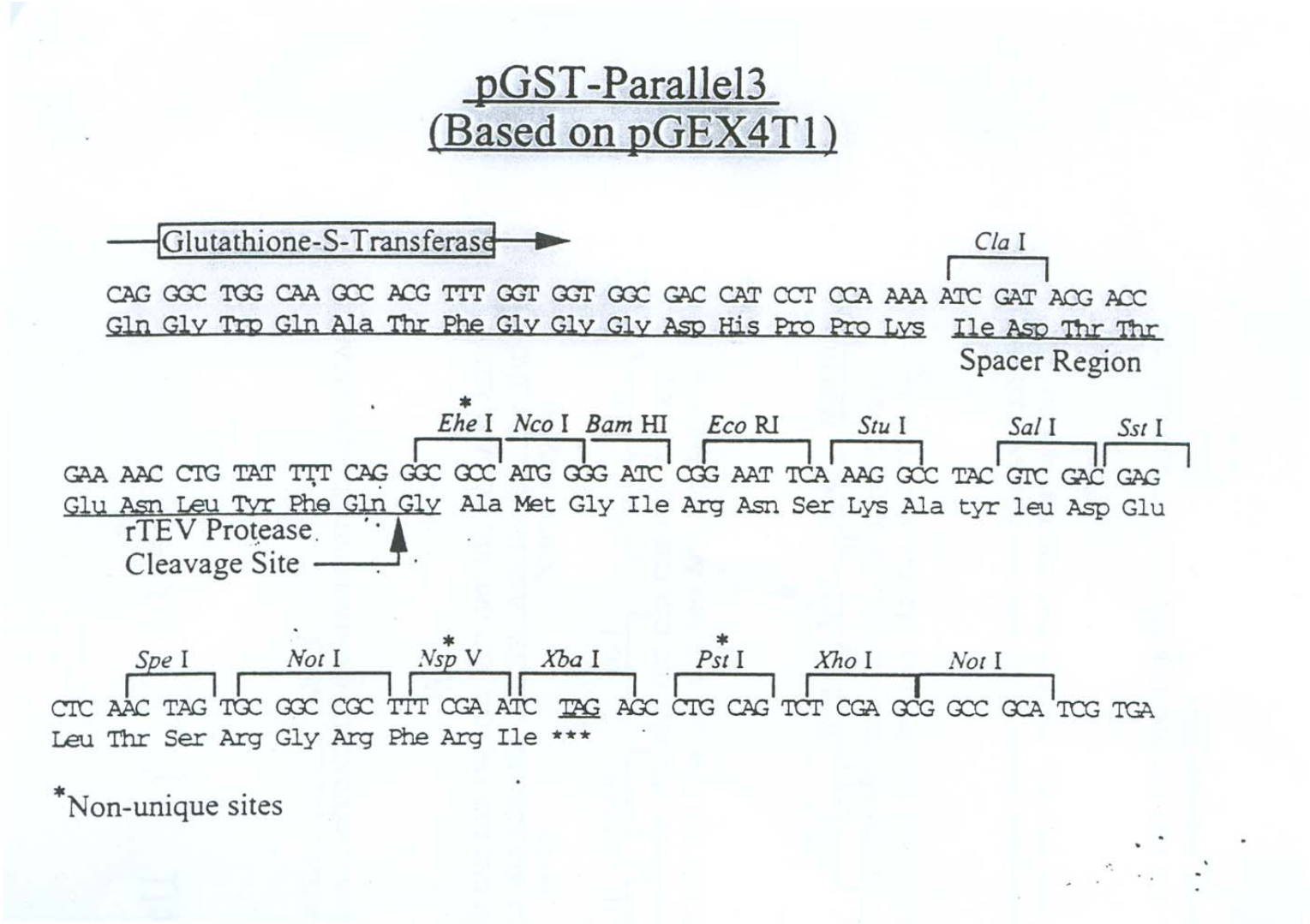
Expression vector requirements



Sequence the vector to verify:

- Correct open reading frame (ORF)
- In frame with N-term or C-term tag
- Stop codon at the end of the reading frame

Check for correct reading frame with N-term



Expressing vectors cassettes examples

Promoters, Fusion tags & proteins, 5’ gene composition, etc.

pET-MBP vector

pET-SUMO vector

pET-PelB-His-tag vector

Mix plasmid with bacteria

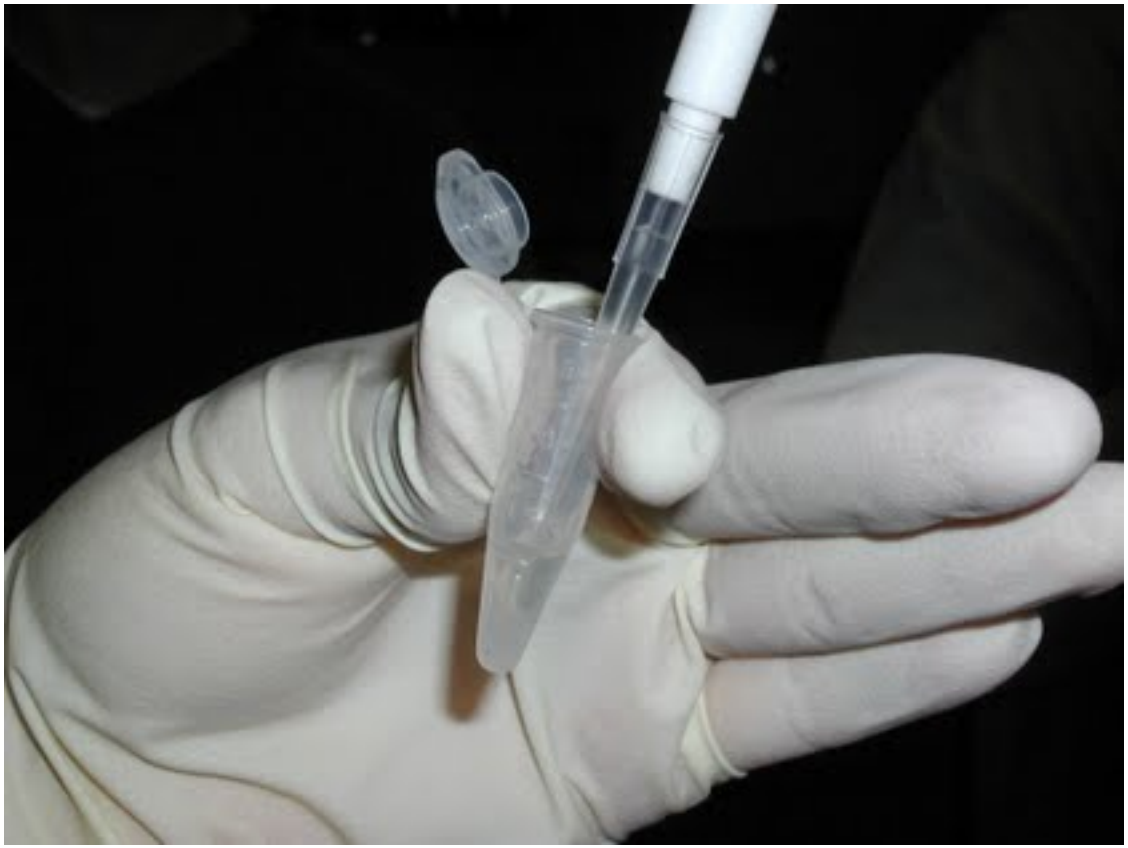
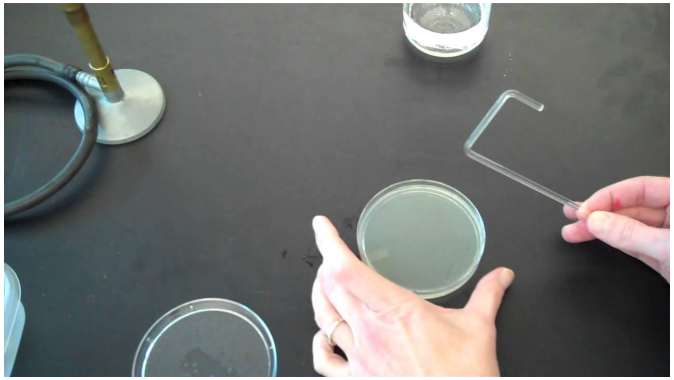
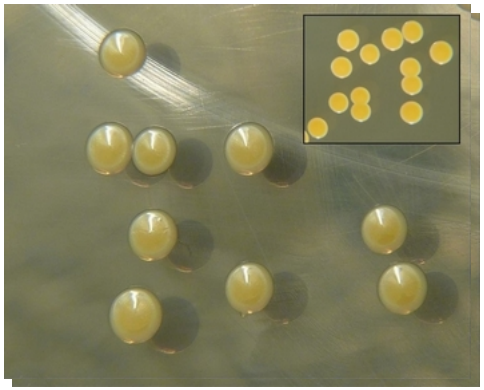


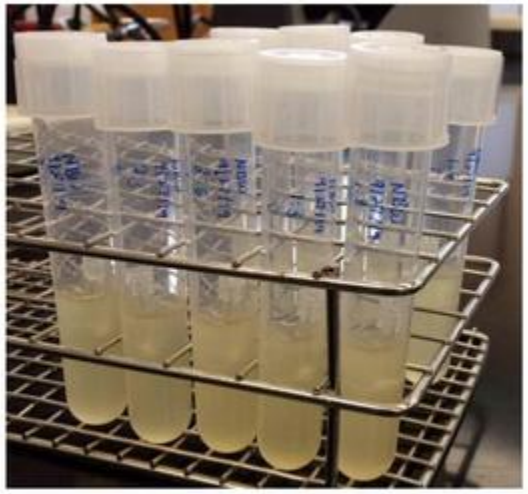
Plate on selective agar plate



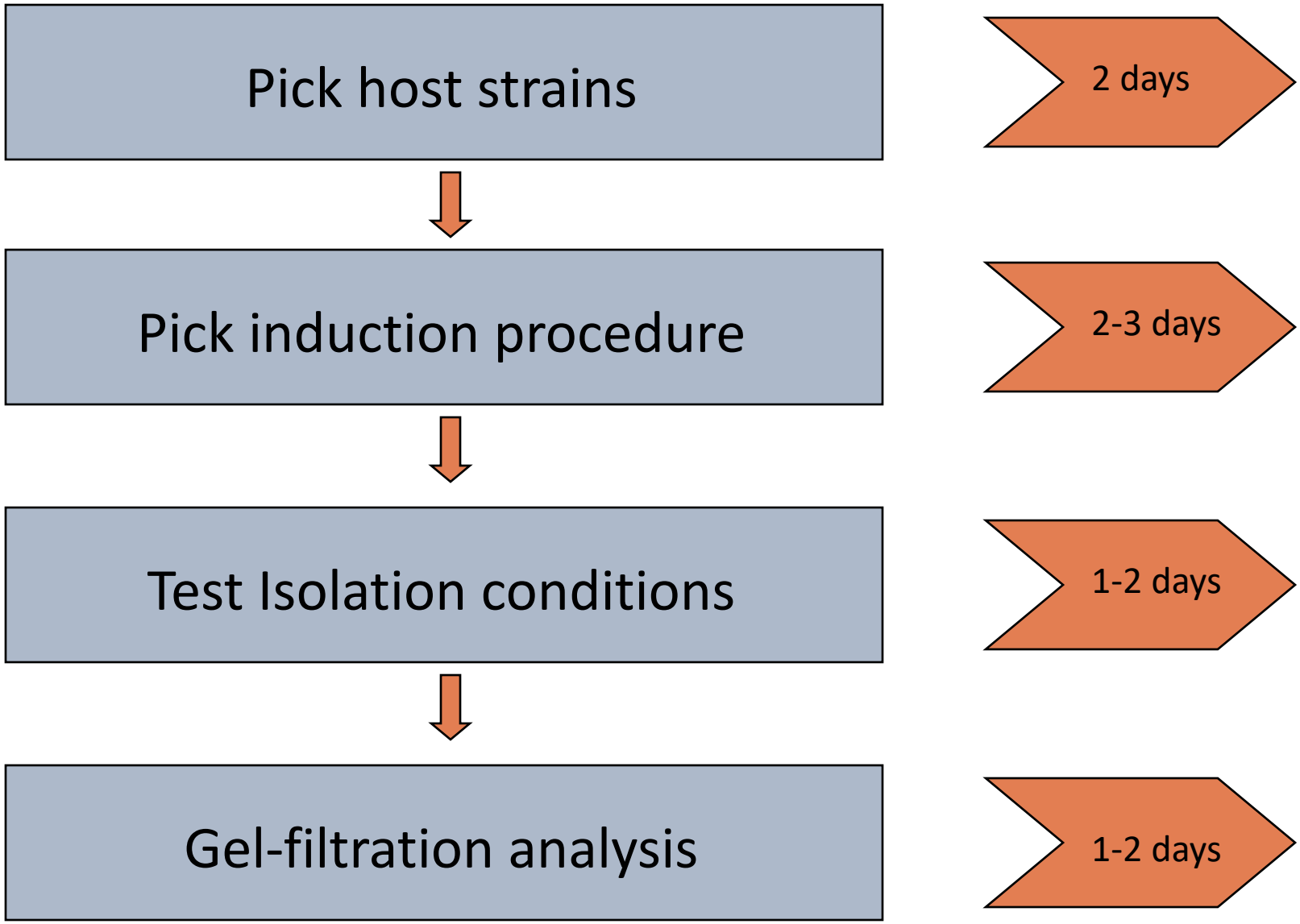
Grow for O/N

Pick a colony into LB containing tube





Grow for 2-16h at 37C


Pick host strains
2 days
Pick induction procedure
2-3 days
Test Isolation conditions
1-2 days
Gel-filtration analysis
1-2 days

Initial screening overview:

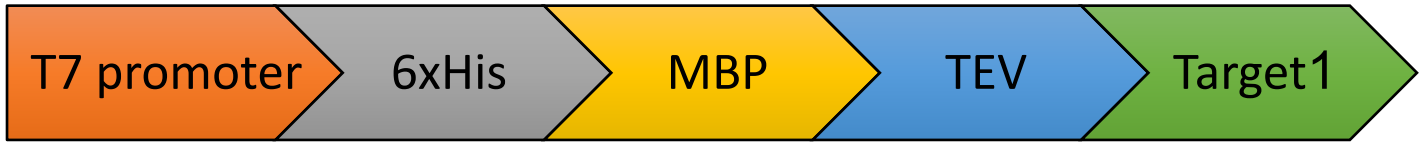
1. Transform construct into several “suitable” bacterial strains
2. Plate strains on selective agar plates
3. Pick a single colony into an auto-induced media snap-cap tube and grow O/N in 30C and 37C shakers
4. Collect samples, lyse (sonication or detergent: BugBuster+ Benzonase nuclease[Novagen] and separate sup from pellet by centrifugation).
5. Load sups Vs. pellets on SDS-PAGE for coomassie and Western analysis or onto small-scale nickel columns

2 constructs x 6 strains x 2 temperature conditions x sup+pellet = 48 samples

http://wolfson.huji.ac.il/expression/procedures/bacterial/auto-induced-exp.html

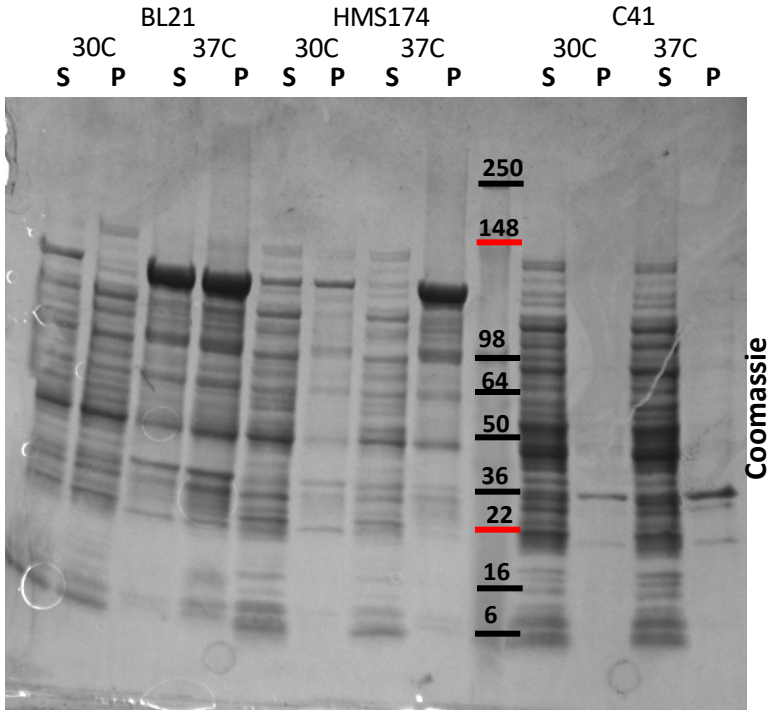
Case study: Expressing Screen of a 100 kDa protein for activity assays (binding)

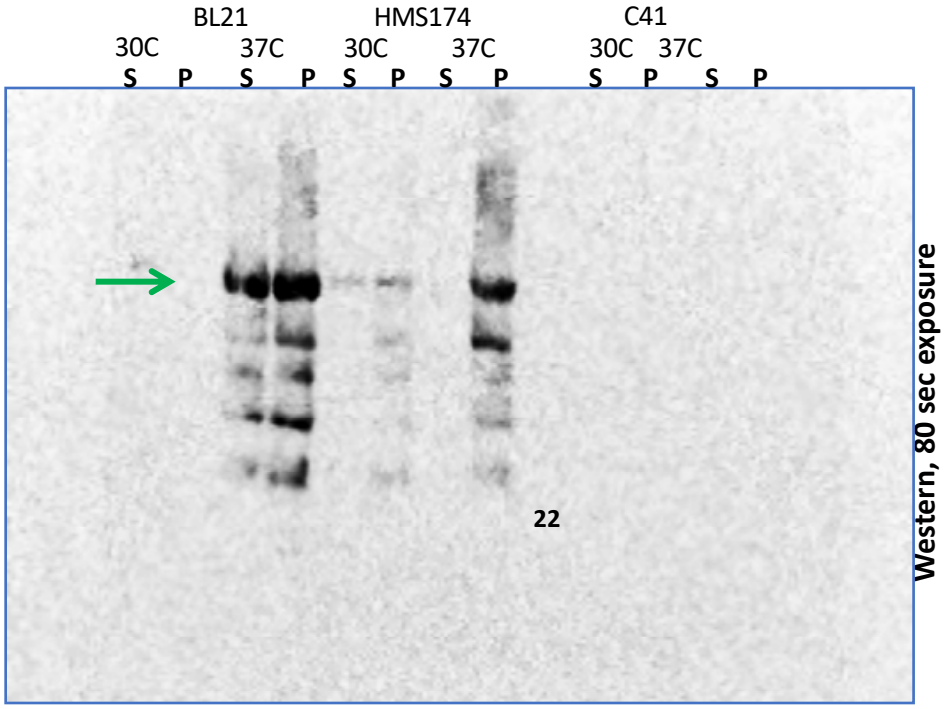
pET-MBP vector



pET-SUMO vector

SUMO-Target1 bacterial selection in auto-induced media





Expected protein size: 130 kDa (Target1 + Sumo)

1st Ab: Mouse anti-His 2nd Ab: HRP goat anti-mouse

Some additional options to save your protein

- Heat-shock + chemical chaperones
- Heat shock +chaperones + fusion protein
- Co-express with a buddy (ligand/receptor)
- Truncate domains (also in HTS in-vitro)
- Denaturation and re-naturation screens

For Induction and Heat Shock procedures:
http://wolfson.huji.ac.il/expression/procedures/bacterial/Induction.Condition.Callib.new.htm

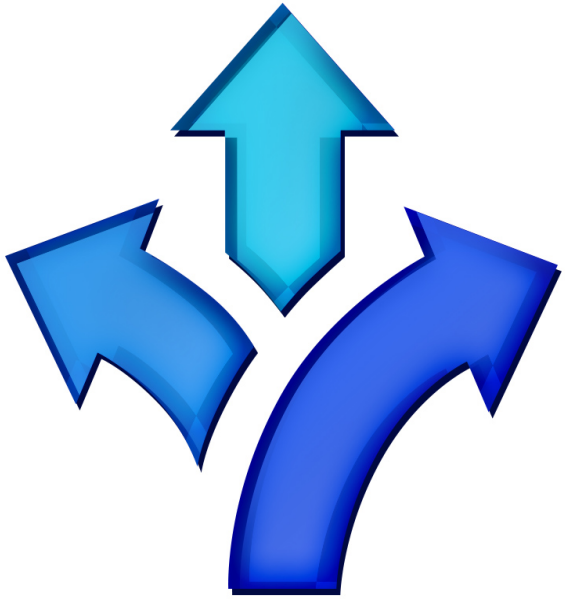
Pareto's Principle - The 80-20 Rule

- 20% of the inputs or activities are responsible for 80% of the outcomes or results.

- This means that for the 20% chance that you might succeed, you should place 80% of your resources.

All depends on how fixed you are on E. *coli*

What next?



- Cloning technologies
- Expression in mammalian cells
- Suggestions?

Aggregations in bacterial cells

Cause of aggregations	Solution 1: Vector	Solution 2: Host Strain	Solution 3: Growth Conditions
lack of correct disulfide bond formation	Fuse with TRX, DsbA, DsbC fusion partners Clone with periplasmic secretion signal (pelB, OmpA)	Use Origami and Shuffle strains	Standard Screen and reduce efficiency of expression conditions
Intrinsically disordered protein	Fuse with MBP, NusA, Sumo lipoyl-domain tag	Membrane-rich strains such as C41 and C43	Standard screen Heat shock with chemical chaperones
Hydrophobic protein	Fuse with solubility enhancing proteins such as MBP, NusA, SUMO	Membrane-rich strains such as C41 and C43	Standard screen Heat shock with chemical chaperones
No appropriate chaperones	Co-express with chaperone containing vectors	Screen BL21, K12, Lemo (NEB) and Arctic (Agilent)	Standard screen Heat shock with chemical chaperones
Protein is natively directed to a sub-cellular localization	Remove localization signal or replace signal with periplasmic secretion signal (pelB, OmpA)	Membrane rich strains as C41 and C43 or Lemo strain	Standard screen Reduce inducer concentration Induce at high OD

Aggregations continues:

Membrane protein	Generate and screen soluble domains	Use membrane-rich strains as C41 and C43, or Lemo strain	Lower induction temperature Reduce inducer concentration Induce at high OD
Protein is part of a complex	Fuse with large fusion proteins such as MBP NusA Co-express with a partner :combination of 2-4 vectors for max 8 proteins	Screen tight induction-regulated strains	Heat shock with chemical chaperones

No / Low Protein Production:

Cause	Solution 1: Vector	Solution 2: Host Strain	Solution 3: Growth Conditions
Toxic protein	Use T7 or Arabinos promoter-• based vectors Tightly regulate induction w/repressor	Check cell growth to • eliminate toxicity Use BL21AI or • BL21(DE3)pLysS/E	Shorten induction time • Add Glucose to suppress leaky • expression
Initiation problems	Re-clone with more A residues • at 5' Shorten distance between RBS • (2-8 nt)and first ATG		
Rare codons	Use synthetic, codon-optimized gene	Use stains supplementing rare codons (Rosetta, Codon +)	Slow translation by reducing temperature
Your gene induces rearrangement and lose of DE3 lysogen	Tightly suppress gene • expression prior to induction Use low-copy ori•	Use recA- strains• (HMS174; BLR)	Start from freshly transformed • bacteria Add Glucose to suppress leaky • expression

Truncated proteins

Cause	Solution 1: Vector	Solution 2: Host Strain	Solution 3: Growth Conditions
Rare codons	Optimize codon usage	Use rare codon strains	Slow elongation by low temp.; low inducer; poor media
Fast-translation of fusion protein	Sub-clone with another fusion partner or avoid N-terminus fusion protein		Slow expression rate with low temp.; low inducer; harvest; poor media short
Degradation	Detect and replace specific protease sites	Low protease strains	Grow and induce at low temp, use protease Inhibitors in lysis media change cell lysis methods lyse cells on ice