
MINOtech
BIOTECHNOLOGY
IMBB-FORTH

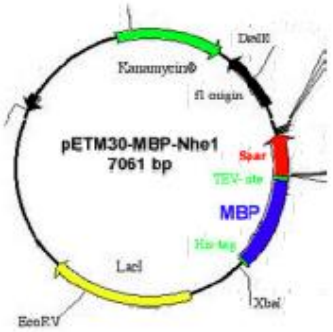
Heterologous protein expression systems

What are they?

- Protein expression systems producing desired polypeptides from recombinant genes.
- Plasmids carry and express the desired recombinant genes
- Expression systems: prokaryotic and eukaryotic microorganisms, cell cultures of plants, mammals or insects, plants, animals, cell-free...

Heterologous Expression: What do we need?

1. DNA of target of interest



2. Expression vector for host system

- promoter
- tags/fusions

3. Host cells

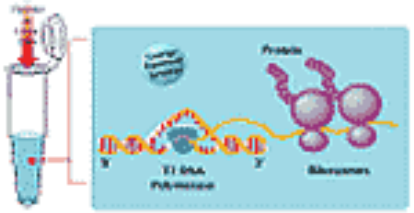
4. Cell growth system

MINOTECH BIOTECHNOLOGY

Protein expression systems

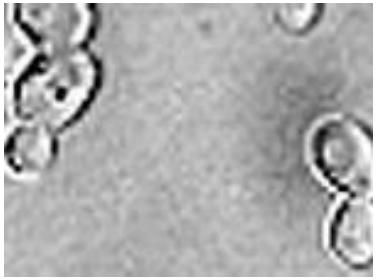
Bacterial

Cell-free



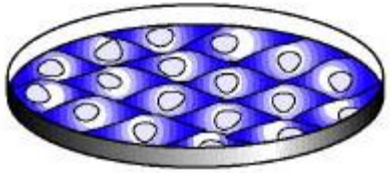
Insect

Yeast



Transgenic plants

Mammalian



Transgenic animals

MINOtech BIOTECHNOLOGY

Heterologous protein expression systems

Main applications:

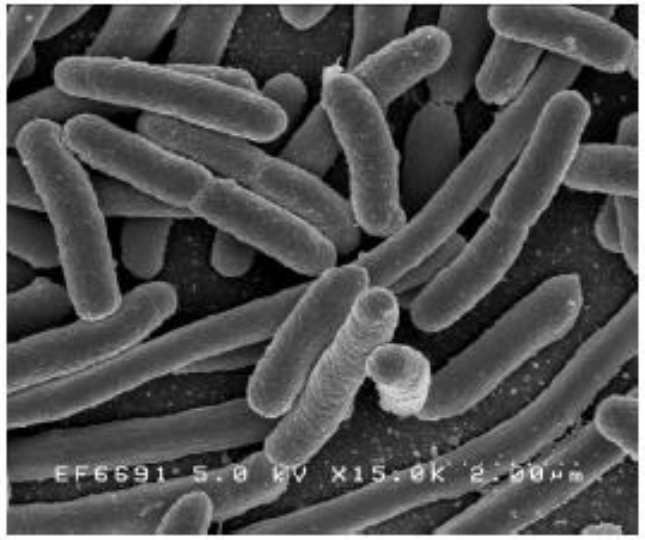
- Production of proteins that are expressed in almost undetectable levels in the wild type source.
- Industrial production of proteins for biomedical applications.
- Characterization of proteins detected by reverse genetic techniques.
- Structural and functional analysis of proteins through protein engineering techniques.
- Production of proteins with novel customized properties.

Heterologous protein expression systems: *E.coli*

Escherichia coli: the best seller of protein expression systems

- ***easy to use, low-cost, fast***

- easy to use cloning and genetic manipulation procedures
- Low cost culture media
- Fast growth and expression



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

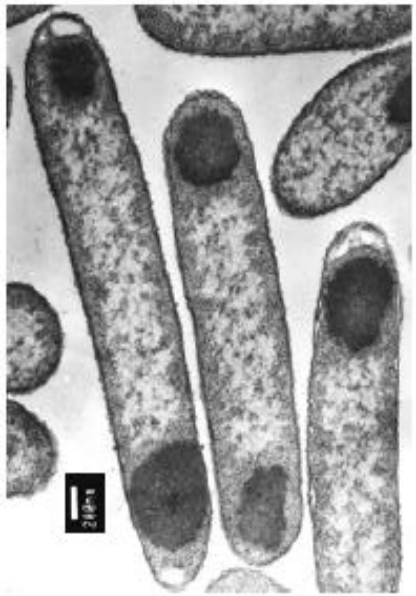
Advantages:

- Fast growth of cell cultures at high density
- Small scale and large expression schemes (few ml to more than 100 lts cultures)
- Small cost.
- High level of expression (<100mg/l of culture).
- Vast number of engineered plasmids and strains.
- Tight control of protein expression (quite helpful for toxic proteins).
- Cytoplasmic, periplasmic and extracellular expression.

Heterologous protein expression systems: *E.coli*

Disadvantages:

- Inefficient secretion of recombinant proteins
- Inefficient disulfide bond formation
- Poor folding of proteins in the cytoplasm (inc. bacterial proteins)
- Inclusion body formation
- In vitro refolding protocols inefficient
- Codon usage different to eukaryotes
- Minimal post-translational modifications
- Endotoxin



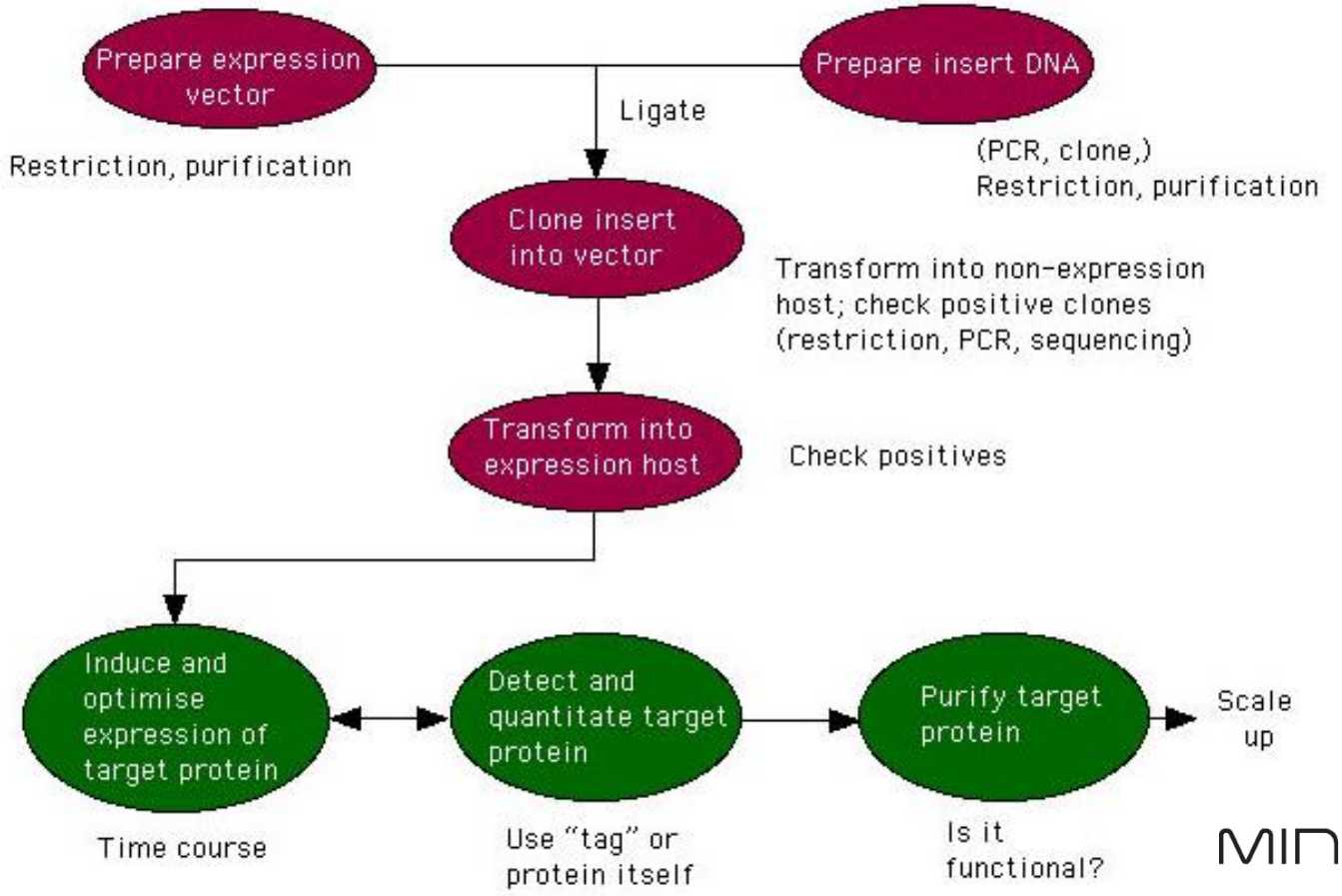
MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Key steps of recombinant protein expression:

- Isolation and amplification of target gene (PCR) or DNA synthesis of the desired sequence
- Cloning into desired expression vector
- Transformation of an *E.coli* expression strain
- Optimization of expression conditions at low scale (media, temperature, concentration of inducer, induction time)
- Identification of the recombinant protein (enzymatic assay, western blot)
- Large scale expression scheme
- Protein purification (Chromatography)

Heterologous protein expression systems: *E.coli*



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

1. Analysis of target gene sequence

- ORF identification (start, stop codon)
- Signal peptides
- Transmembrane domains
- Post-translational modifications
- Rare codons
- Restriction site analysis

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

2. Primer Design

- The gene of interest usually has to be amplified from genomic or vector DNA by PCR (polymerase chain reaction) before it can be cloned into an expression vector. The first step is the design of the necessary primers.
- **Primer length.** Usually a primer length of **18-30 bases** is optimal for most PCR applications. Shorter primers could lead to amplification of nonspecific PCR products.
- **Melting temperature (T_m).** The **specificity** of PCR depends strongly on the melting temperature (T_m) of the primers (the temperature at which half of the primer has annealed to the template). Usually good results are obtained when the T_m's for both primers are similar (within 2-4 °C) and **above 60°C**. The T_m for a primer can be estimated using the following formula:
- $T_m = 2°C * (A + T) + 4°C * (C + G)$
- **GC content.** The GC content of a primer should be between **40** and **60%**.

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Design of the 5'-end primer

●The 5'-end primer overlaps with the 5'-end of the gene of interest and should contain the following elements:

●**Restriction site.** The restriction site should be the same or provide the same sticky end to the first of the restriction enzymes in the multiple cloning site of the vector chosen to clone the gene of interest into. Alternatively, you could pick any restriction enzyme that gives a blunt end upon cleavage (see cloning). Often **Nco I**CC**ATG**G) or **Nde I** (CAT**ATG**) are chosen because the **ATG** within these sites can be used directly to create the ATG start codon and/or the ATG codon for the N-terminal methionine residue (see Utilisation of the Nco I cloning site)

●**5'-extension to the restriction site.** Restriction enzymes cleave DNA much less efficient towards the end of a fragment. A 5' extension of the restriction site with 2-10 bases greatly increases the cleavage efficiency of most enzymes. Data on the effect of the extension length and sequence on the cleavage efficiencies of the most used restriction enzymes can be found in the reference appendix of the New England Biolabs catalogue.

●**Start codon.** A start codon (usually **ATG**) should be included when the gene of interest is not expressed with an N-terminal tag or fusion partner or when an N-terminal methionine residue is present. It should be checked that the start codon and the gene of interest are in frame with an eventual N-terminal tag and/or fusion partner.

●**Overlap with the gene of interest.** The overlap between the primer and the gene of interest should be long enough to give a T_m of 60°C or more (calculated as shown above).

Heterologous protein expression systems: *E.coli*

Design of the 3'-end primer

- The 3'-end primer overlaps with the DNA strand complementory to the 3'-end of the gene of interest and should contain the following elements:
- **Restriction site.** The restriction site should be the same or provide the same sticky end to the second of the restriction enzymes in the multiple cloning site of the vector chosen to clone the gene of interest into. Alternatively, you could pick any restriction enzyme that gives a blunt end upon cleavage (see cloning)
- **5'-extension to the restriction site.** Restriction enzymes cleave DNA much less efficient towards the end of a fragment. A 5' extension of the restriction site with 2-10 bases greatly increases the cleavage efficiency of most enzymes..
- **Stop codon(s).** A stop codon (**TAA** is preferred because it is less prone to read-through than **TAG** and **TGA**) should be included when no C-terminal tag is used. To increase the efficiency of termination it is possible to use 2 or 3 stop codons in series.
- **Overlap with the stand complement to the 3'-end of the gene of interest.** The overlap between the primer and the strand complement to the 3'-end of the gene of interest should be long enough to give a T_m of 60°C or more (calculated as shown above). It should be checked that the gene of interest is in frame with an eventual C-terminal tag.

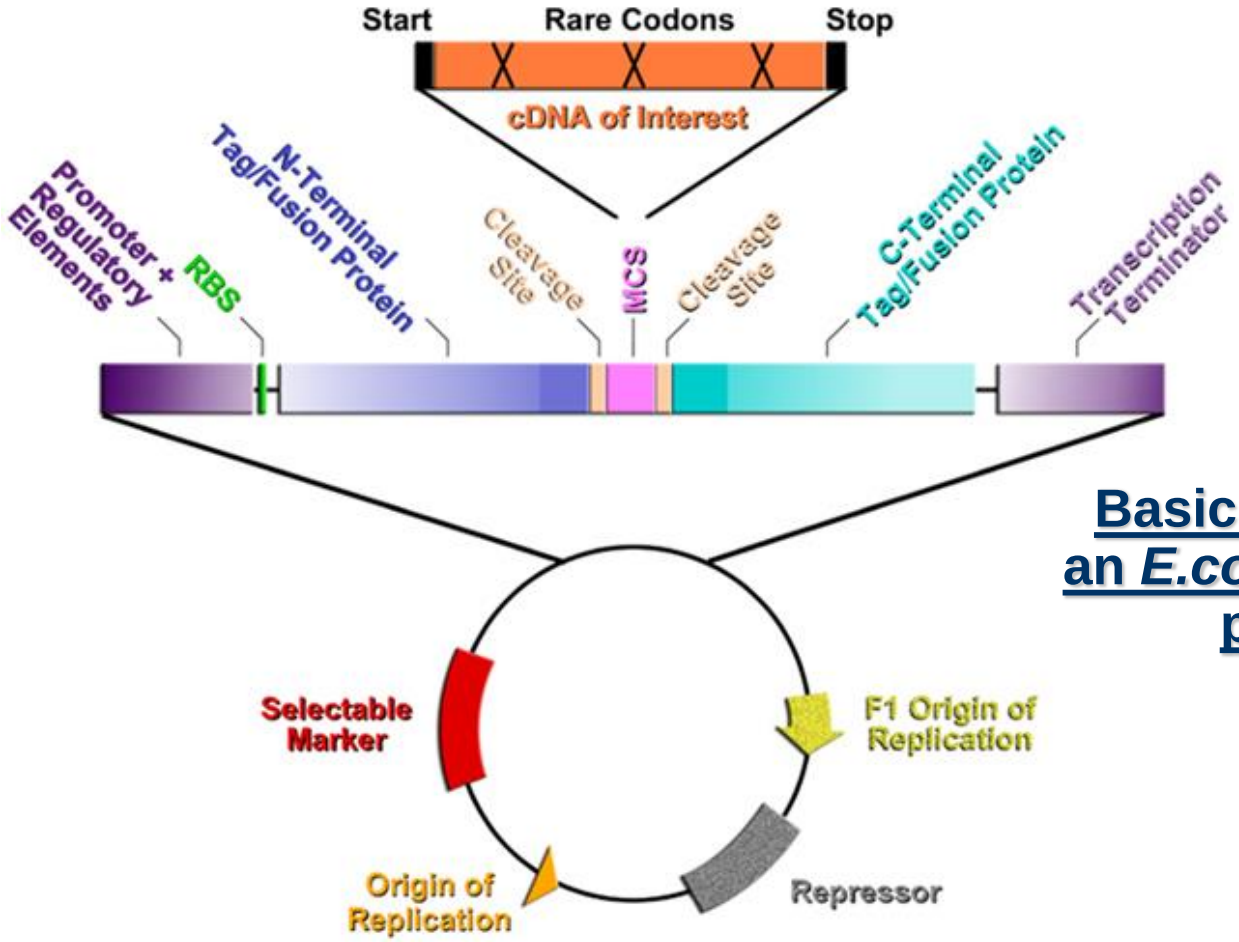
Heterologous protein expression systems: *E.coli*

3.Choosing the right expression plasmid

Basic elements of an *E.coli* expression plasmid:

- Origin of replication (*Ori*) region
- Selectable marker
- Transcriptional promoter
- Unique multiple cloning sites (MCSs)
- Translational initiation regions (TIRs)
- Translational terminator
- Various tags and fusion partners
- Protease cleavage site

Heterologous protein expression systems: *E.coli*

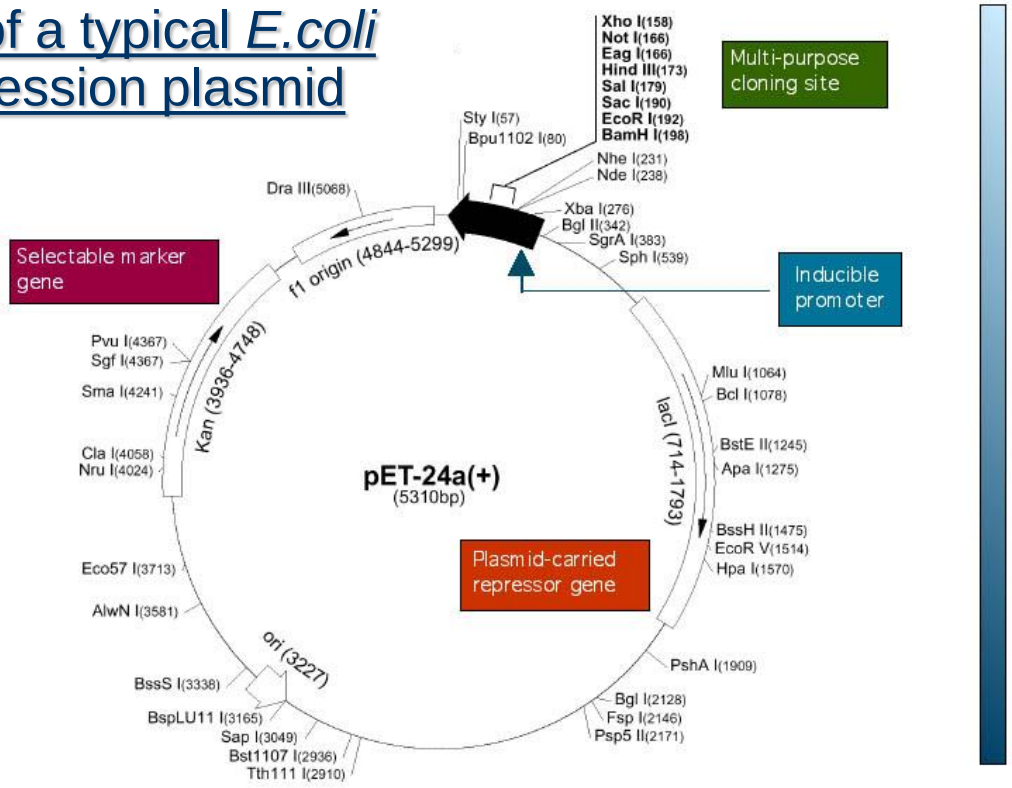


Basic elements of an *E.coli* expression plasmid

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Map of a typical *E.coli* expression plasmid



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

- **Origin of replication (*Ori*) region:**
  - a particular sequence in a plasmid (or genome) at which replication is initiated
  - Narrow or broad host range
  - High- or low-copy number

Plasmid	Replicon	Copy Number
pBR322	pMB1	*15-20*
pACYC	p15A	*18-22*
pSC101	pSC101	*5*
pUC	colE1	*500-700*

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

- **Selective marker(s) :**

Sequences encoding a selectable marker that assures maintenance of the vector in the host

- Ampicillin → Inhibits bacterial cell wall synthesis by binding to one or more of the penicillin binding proteins (PBPs);
- Kanamycin → Interacts with the 30S subunit of prokaryotic ribosomes
- Chloramphenicol → Inhibits peptidyl transferase activity of the bacterial ribosome
- Tetracycline → Interacts with the 30S subunit of prokaryotic ribosomes
- …

Heterologous protein expression systems: *E.coli*

- **Ribosome Binding Site (RBS):**

RBS RBS 5-9 n START

GAAGGAATTCAGGAGCCCTTCACCATG ... ...

- **START codons:**

E. coli uses 77% ATG (AUG), 14% GTG (GUG), 8% TTG (UUG) and a few others

- **STOP codons:**

TAG (UAG), TGA (UGA), TAA (UAA)

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

- **Promoter:**
  - A region of DNA that control the transcription of a particular gene
  - Located upstream of the genes

Good promoter should be:

- Strong promoter. Accumulation of expressed protein up 10 to 30% or more of the total cellular protein.
- Minimal basal expression level. Tight regulation of the promoter
- Easily induced. Simple and cost effective manner

Heterologous protein expression systems: *E.coli*

- **Host's promoters**

2500 in the entire genome of *E. coli* K12 strain

Most frequently used: Plac / Ptac / Ptrc, P_{PBAD}, rha_{PBAD}

- Regulation of expression

- **Promoters from phages**

T7, T3, SP6, T5, P_L

- Highly efficient and specific expression

Heterologous protein expression systems: *E.coli*

Plac, Ptac, Ptrc: Characteristics

	Level of expression (inductor)	Key features
Plac	Low level up to middle (IPTG)	Weak, regulated. Suitable for expression of gene products at very low intracellular level. Comparatively expensive induction.
Ptac Ptrc (trp-lac)	Moderately high (IPTG)	High-level, but lower than T7 system. Regulated expression still possible.Comparatively expensive induction. High basal level.

MINOTECH BIOTECHNOLOGY

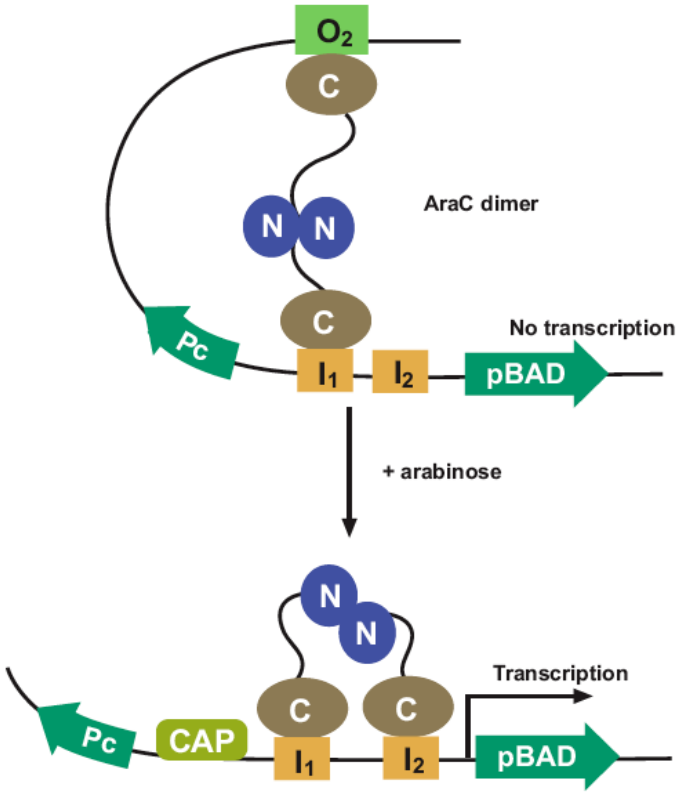
Heterologous protein expression systems: *E.coli*

P_{PBAD}: Regulation

• Arabinose binds to AraC. The protein releases the O2 site and binds the I2 site, which is adjacent to the I1 site. This releases the DNA loop and allows transcription to begin (5).

• The cAMP activator protein (CAP)-cAMP complex binds to the DNA and stimulates binding of AraC to I1 and I2.

Basal expression levels can be repressed by introducing glucose to the growth medium. Glucose acts by lowering cAMP levels, which in turn decreases the binding of CAP. As cAMP levels are lowered, transcriptional activation is decreased. This is ideal when the protein of interest is extremely growth-inhibitive or toxic to the host.



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

PPBAD and *Rha*PBAD

	Level of expression (inductor)	Key features
P_{PBAD}	Variable from low to high level (L-arabinose)	Can fine-tune expression levels in a dose-dependent manner. Tight regulation possible. Low basal level. Inexpensive inducer.
Rha_{PBAD}	Variable from low to high level (L-rhamnose)	Tight regulation. Low basal activity. Relatively expensive inducer.

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

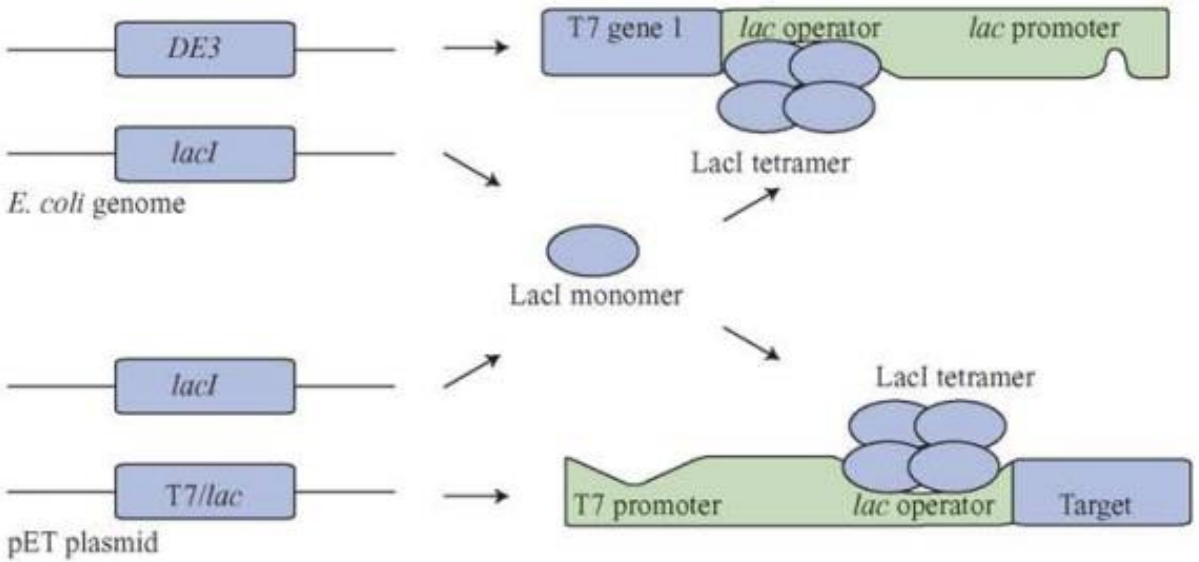
Phage promoters

	Level of expression (inductor)	Key features
T7 T5	Very high (IPTG) High (IPTG)	Utilizes T7 RNA polymerase. Utilizes E. coli RNA polymerase.
P_L	Moderately high (temperature shift)	Temperature-sensitive host required. Less likelihood of "leaky" un-induced expression. Basal level by temperatures below 30°C. Induction at 42°C (repressor inactivated)

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

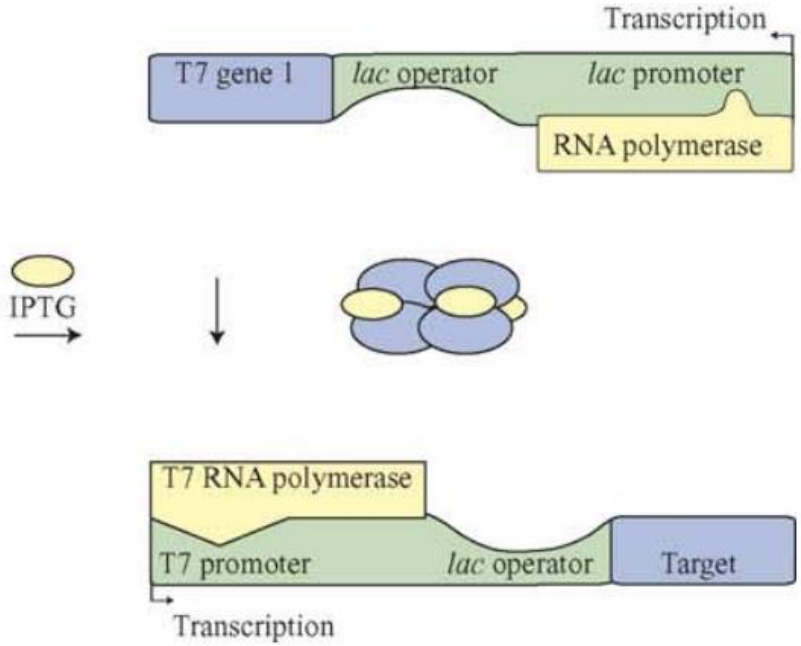
- **pET expression system (T7 promoter)**
  - **IPTG absent**



MINOTECH BIOTECHNOLOGY

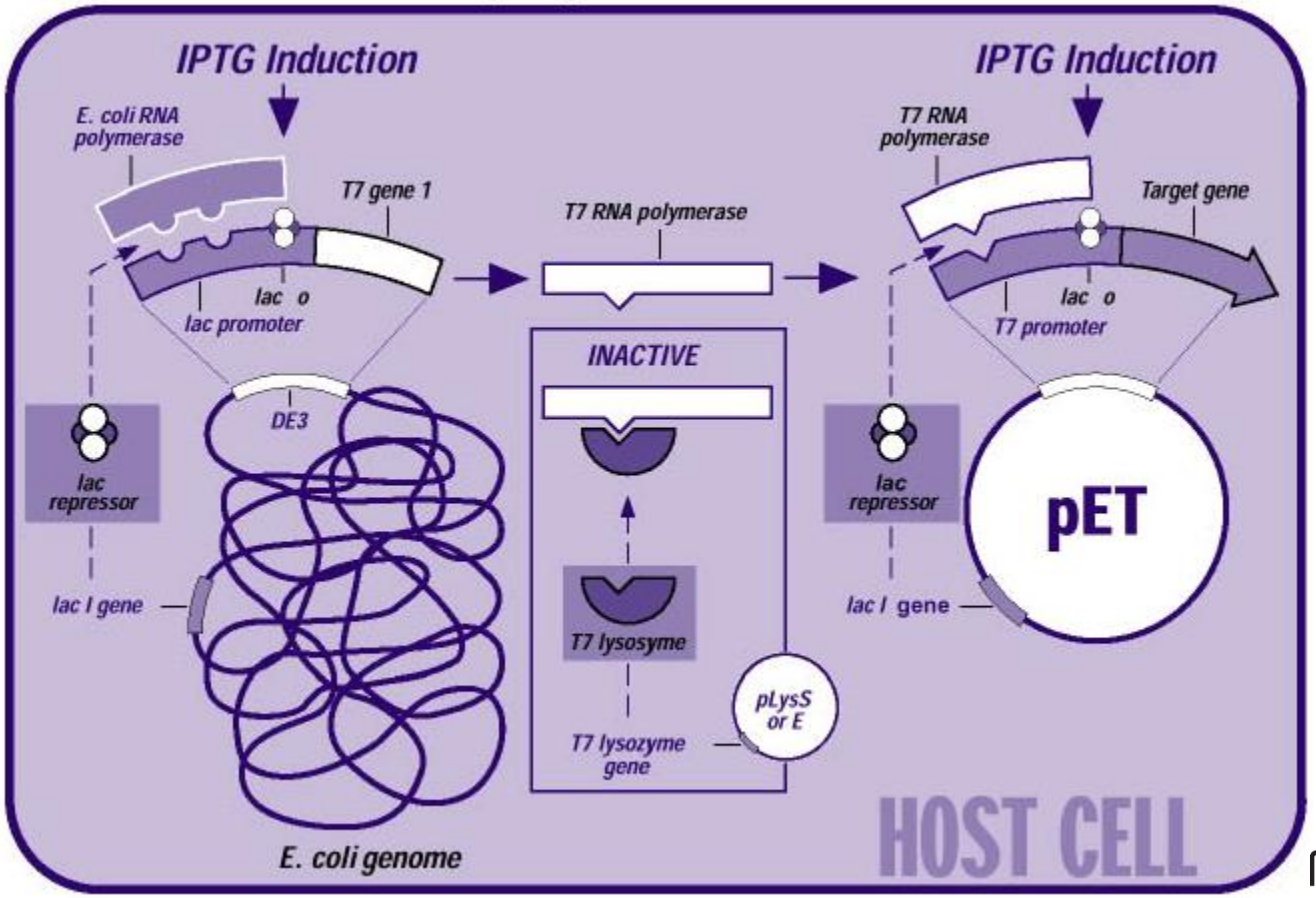
Heterologous protein expression systems: *E.coli*

- **pET expression system (T7 promoter)**
  - **IPTG present**



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

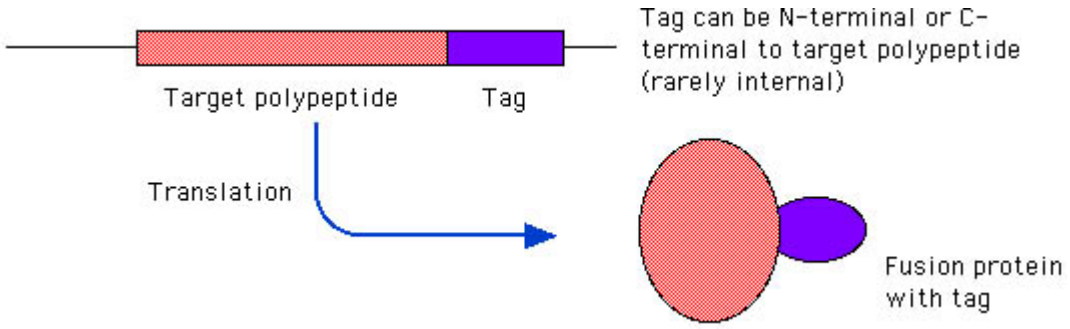


MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Fusion tags

- Regions of coding sequence supplied by the expression vector which are expressed in frame with the desired polypeptide
- Resulting fusion polypeptide retains the functional properties of both the target protein and the tag.
- Tag can be used to 1) detect, 2) purify, 3) direct the expressed protein to a particular cellular localization, 4) assist folding
- Tags can be removed either by specific proteases or as part of the translation process



MINOtech BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Fusion tags

- **Improved solubility (S) -** Fusion of the N-terminus of the target protein to a soluble fusion partner often improves the solubility of the target protein.

- **Improved detection (D) -** Fusion of the target protein to either terminus of a short peptide (epitope tag) or protein which is recognized by an antibody (Western blot analysis) or by biophysical methods (e.g. GFP by fluorescence) facilitates the detection of the resulting protein during expression or purification.

- **Improved purification (P) -** Simple purification schemes have been developed for proteins used at either terminus which bind specifically to affinity resins.

- **Localization (L) -** Tag, usually located on N-terminus of the target protein, which acts as address for sending protein to a specific cellular compartment.

- **Improved Expression (E) -** Fusion of the N-terminus of the target protein to a highly expressed fusion partner results in high level expression of the target protein.

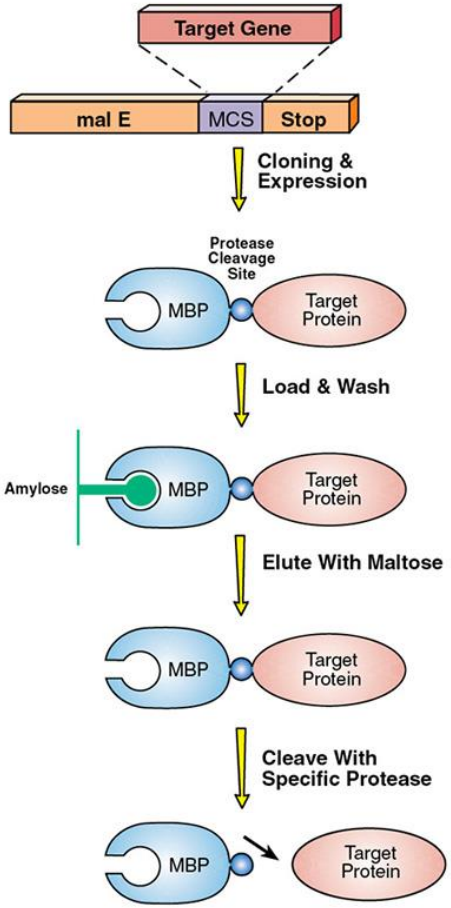
Heterologous protein expression systems: *E.coli*

- Commonly used tags for purification and enhancing the solubility

Tag	Size	Tag placement	Uses
Affinity tags			
His tag	6 or 8 amino acids (aa)	N-, C- or internal	Purification
Strep II tag	8 aa		Purification
T7- tag	11 or 16 aa	N-, internal	Purification
FLAG tag	8 aa	N-, C-	Purification
S-tag	15 aa	N-, C-	Purification
HA tag	9 aa	N-, C-	Purification
c-Myc tag	11 aa	N-, C-	Purification
DHFR	25 kda	N-	Purification
Chitin binding domain	51 aa	N-, C-	Purification
Calmodulin binding domain	26 aa	N-, C-	Purification
Cellulose binding domain	27–129 aa	N-, C-	Purification
Solubility-enhancing tags			
GST	211 (26 kDa)	N-	Enhances solubility & purification
MBP	396 (40 kDa)	N-, C-	Enhances solubility & purification
T7 gene10	260 aa	N-	Enhances solubility & purification
NusA	495 aa (54.8 kDa)	N-	Enhances solubility
Thioredoxin	109 aa	N-, C-	Enhances solubility
SUMO	100 aa	N-	Enhances solubility
Ubiquitin	76 aa	N-	Enhances solubility

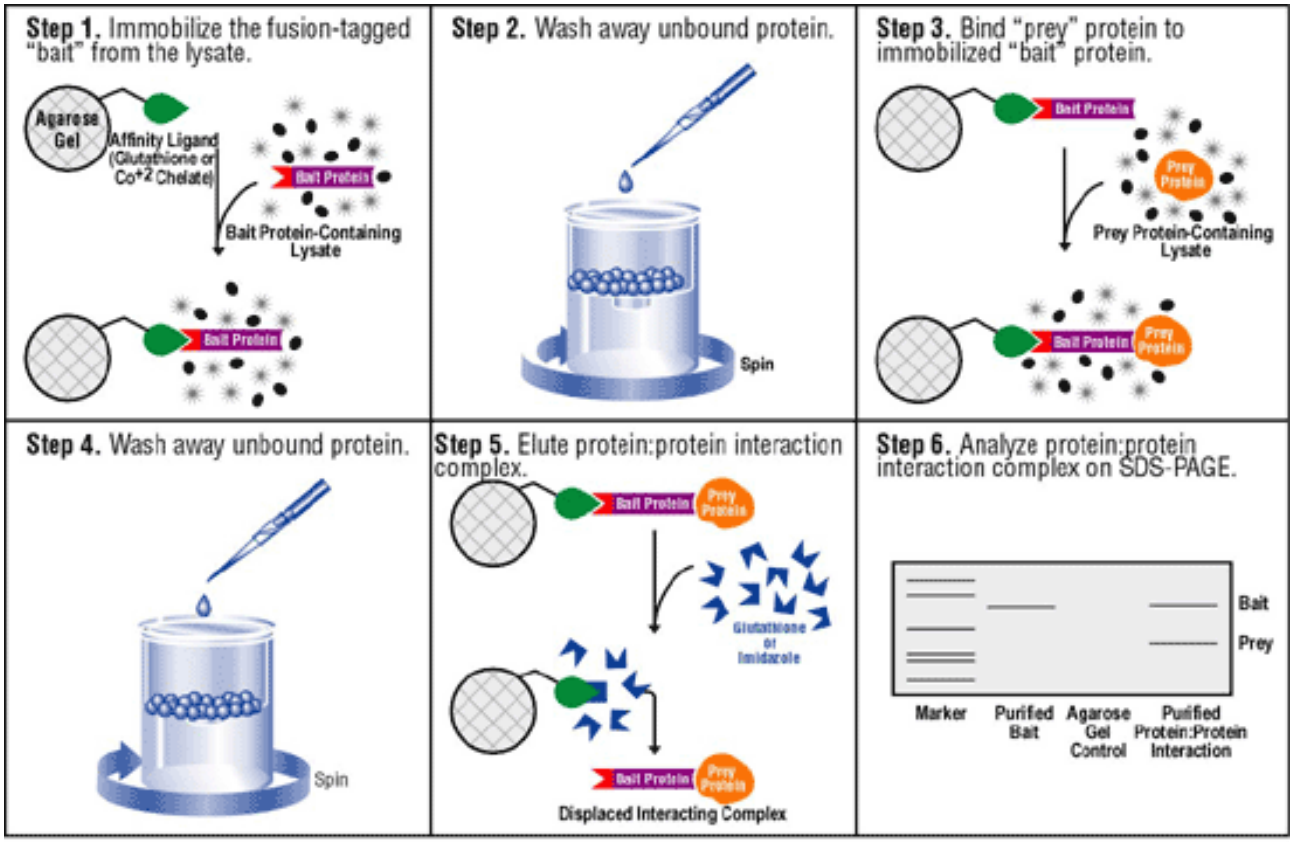
Heterologous protein expression systems: *E.coli*

Maltose Binding Protein tag (NEB pMal workflow)



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*



Protein-protein interaction identification using tags

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

- Protease cleavage enzymes used for cutting off the fused tag

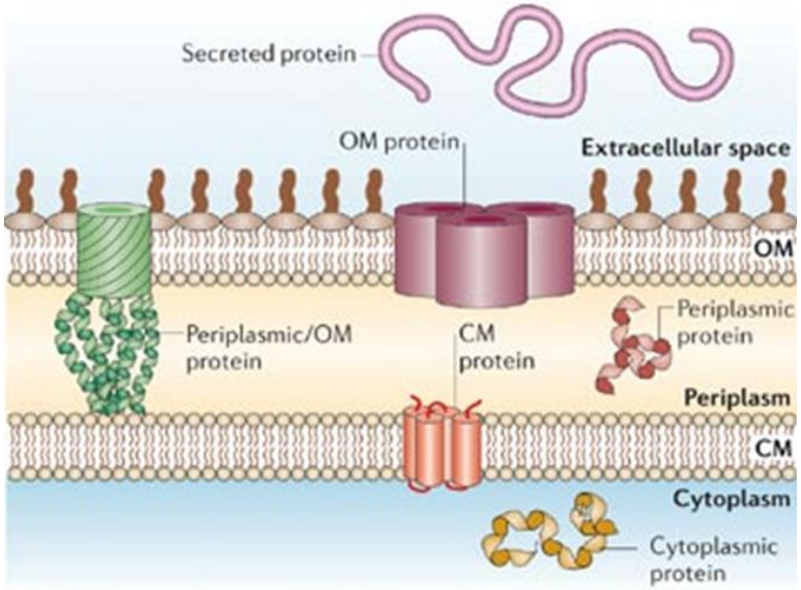
Enzyme	Recognition Site	Comments
Thrombin	LVPR/GS	Active, less specific
Factor Xa	IEGR/	Less specific
Enterokinase	DDDDK/	Very specific Leaves native N-term!
PreScission™	LEVLFQ/GP	Very specific
TEV	ENLYFQ/G(S)	Very specific
TVMV	ETVRFQ/G(S)	Very specific
Ubq/SUMO Pro.	Ubq/SUMO fusion	

Heterologous protein expression systems: *E.coli*

4. Choose the location of the expressed protein

Three cellular locations

- Cytoplasmic expression
- Periplasmic expression
- Extracellular expression



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Cytoplasmic expression

Advantages

- Target protein protected from proteases
- Protein in high purity and concentration
- Simple plasmid constructs

Disadvantages

- Inclusion bodies formation
- Protein insolubility
- Reduction in final protein yield
- Protein is not active

Solutions

- Lower growth temperature
- Control of expression rate
- Alternate *E.coli* strain
- Insert tag
- Chaperones co-expression

MINOtech BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Periplasmic expression

Advantages

- Simple purification
- Less extensive proteolysis
- Improved disulfide bond formation/folding

Disadvantages

- Signal peptide does not always facilitate transport
- Inclusion bodies may form

Solutions

- Co-expression of signal peptidase I
- Co-expression of sec genes
- Insert tag

Heterologous protein expression systems: *E.coli*

Extracellular expression

Advantages

- Simple purification
- Least level of proteolysis
- Improved protein folding
- N-terminus authenticity

Disadvantages

- No secretion usually
- Protein diluted

Solutions

- Fusion to normally secreted protein
- Glycine supplement in medium
- Concentration

MINOtech BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

5. Choose an *E.coli* strain that best fits to our needs

CELL LINE	COMMENTS
General Expression	
BL21	Deficient in ompT and lon proteases
BL21 DE3	T7-based expression
Expression Problems	
RIL/RP	Codon supplements (high AT content/High GC content, respectively)
RILP/Rosetta	Codon supplements (Codons for both high AT and GC content)
Solubility Problems	
Origami	Enhance disulfide bond formation (thx and glut. Reductase mutants)
Tuner	Can finely tune expression using IPTG (mutation in *lacZY*)
Arctic Express	Express Cpn60/Cpn10, cold-adapted chaperones (10 ^{o}C expression)
Labeling	
B834	Selenomethionine labeling for crystallography (*Met* auxotroph)
Membrane/Toxic Prot.	
pLysS	Reduce basal expression by expression of lysozyme
C43	Facilitates soluble expression of toxic and integral membrane proteins

Heterologous protein expression systems: *E.coli*

Troubleshooting

- Most common problems in *E.coli* expression system
  - Codon usage
  - Inclusion bodies

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Codon usage

It refers to differences in the frequency of occurrence of synonymous codons in coding DNA. Usually, the frequency of the codon usage reflects the abundance of their cognate tRNAs. The following problems are often encountered:

- Decreased mRNA stability (by slowing down translation)
- Premature termination of transcription and/or translation, which leads to a variety of truncated protein products
- Frameshifts, deletions and misincorporations (*e.g.* lysine for arginine).
- Inhibition of protein synthesis and cell growth.

.

Expressed levels can be improved by:

- replacing codons that are rarely found in highly expressed *E. coli*g enes with more favorable codons throughout the whole gene
- co-expressing the genes encoding for a number of the rare codon tRNAs (engineered *E.coli* strains)
- making changes in the coding sequence that reduce secondary structure in the translation initiation region.

Heterologous protein expression systems: *E.coli*

Codons associated with translation problems in *E. coli*

AGG	arginine
AGA	arginine
CGG	arginine
CGA	arginine
GGA	glycine
AUA	isoleucine
CUA	leucine
CCC	proline

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Organism	AGG arginine	AGA arginine	CUA leucine	AUA isoleucine	CCC proline
Escherichia coli	1.2	2.1	3.9	4.4	5.5
Homo sapiens	11.4	11.5	6.5	6.9	**20.0**
Drosophila melanogaster	6.4	5.1	8.2	9.2	**18.0**
Caenorhabditis elegans	4.0	**15.4**	8.0	9.7	4.5
Saccharomyces cerevisiae	9.3	**21.3**	13.4	**17.8**	6.8
Plasmodium falciparium	4.1	**20.2**	**15.2**	**33.2**	8.5
Clostridium pasteurianum	2.4	**29.4**	6.2	**50.0**	0.9
Pyrococcus horikoshii	**30.1**	**20.1**	**18.2**	**44.5**	10.2
Thermus aquaticus	14.3	1.3	3.6	1.4	**38.8**
Arabidopsis thaliana	10.9	**18.8**	10.0	12.7	5.3

Table 1

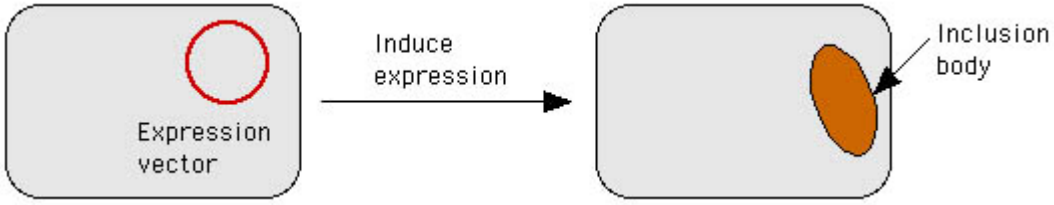
Codon Usage in Various Organisms

Codon frequencies are expressed as codons used per 1000 codons encountered. The arginine codons AGG and AGA are recognized by the same tRNA and should therefore be combined. Codon frequencies of more than 15 codons/1000 codons are shown in bold to help identify a codon bias that may cause problems for high-level expression in *E. coli*. These frequencies are updated regularly. A complete compilation of codon usage of the sequences in the GenBank database can be found at www.kazusa.or.jp/codon/.

Heterologous protein expression systems: *E.coli*

Inclusion bodies

- In many cases the expressed protein is insoluble and accumulates in so-called **inclusion bodies**. This is especially true under conditions of high level expression. Several strategies are available to improve the solubility of the expressed protein.



MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

Solutions against inclusion bodies formation

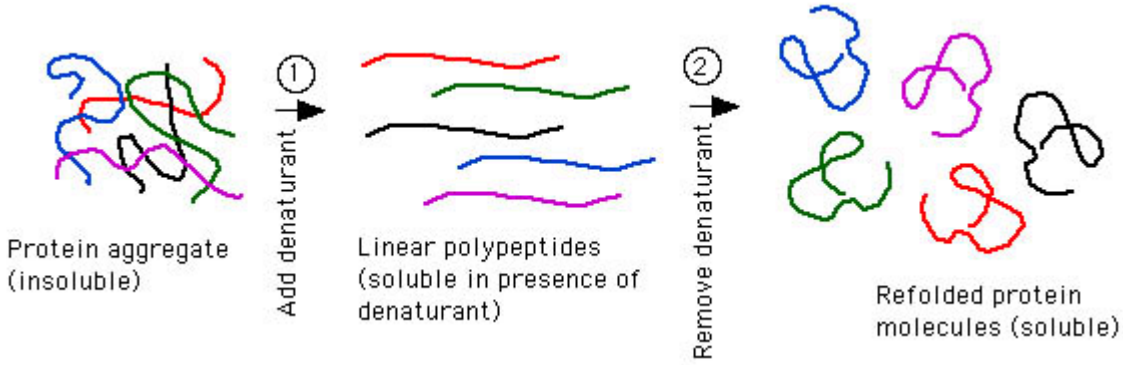
- Reducing the rate of protein synthesis (lowering the growth temperature, using a weaker promoter , using a lower copy number plasmid, lowering the inducer concentration).
- Changing the growth medium
- Co-expression of chaperones and/or foldases (GroES-GroEL,DnaK-DnaJ-GrpE,ClpB, **DsbA** and **DsbC, PDI**)
- Periplasmic expression
- Using specific host strains
- Addition of a fusion partner
- Expression of a fragment of the protein
- *In vitro* denaturation and refolding of the protein

MINOTECH BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

In vitro denaturation and refolding of the protein:

●When despite all efforts the target protein still is expressed in inclusion bodies, then the last resort is to denature and refold the protein *in vitro*. This procedure is carried out in three phases:



●isolation of the inclusion bodies (cell lysis and centrifugation)

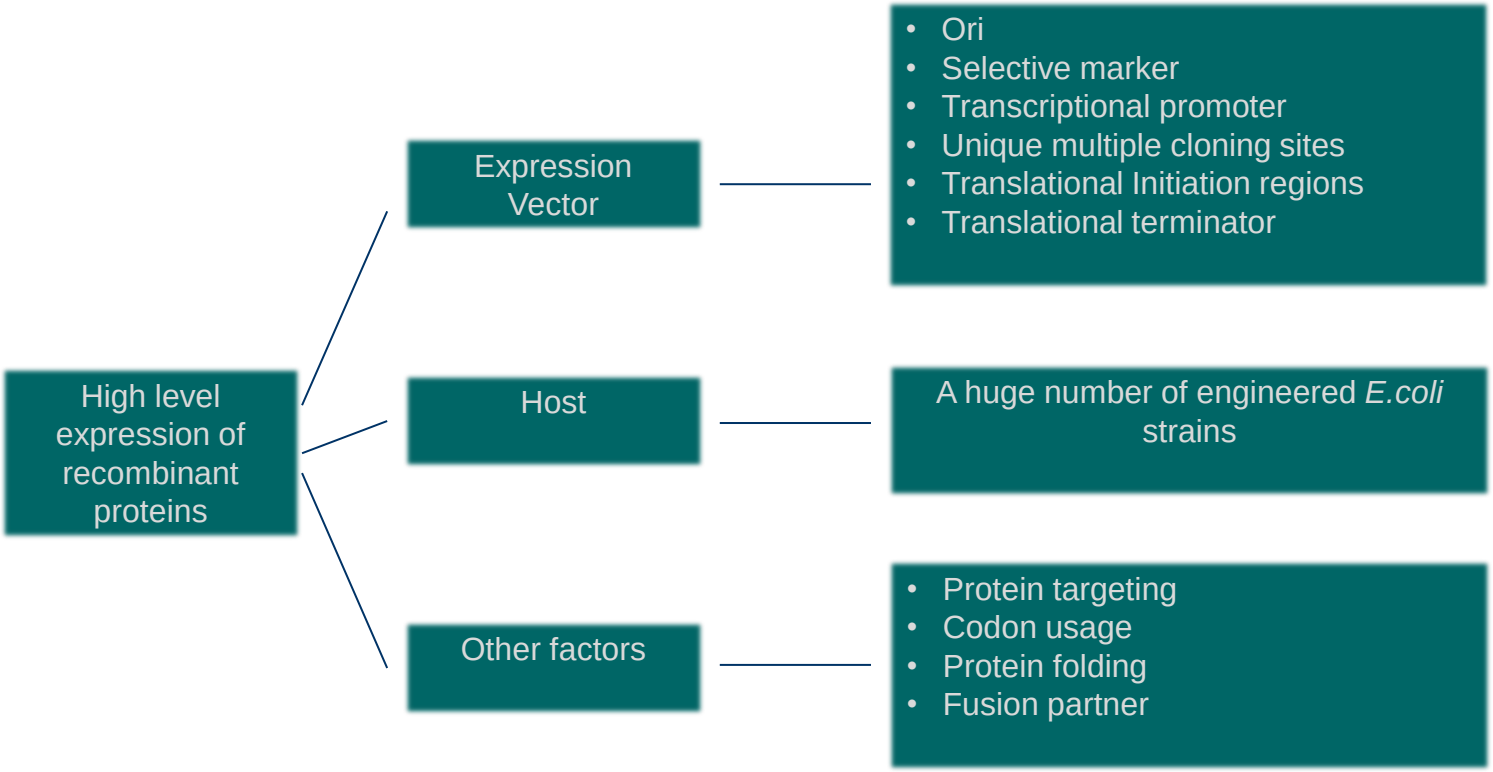
●solubilization and denaturation of the target protein. This is done by the addition of a denaturing agent (usually guanidine or urea) under reducing conditions (*e.g.*20 mM DTT).

●refolding of the protein by removing the denaturating agent using dialysis, dilution or chromatography. For proteins containing disulfide bonds this has to be carried out in the presence of a redox shuttling system *e.g.* reduced and oxidized glutathione.

MINOtech BIOTECHNOLOGY

Heterologous protein expression systems: *E.coli*

What to have in mind for a successful *E.coli* protein expression scheme



MINOTECH BIOTECHNOLOGY

GOOD LUCK

MINOTECH
BIOTECHNOLOGY