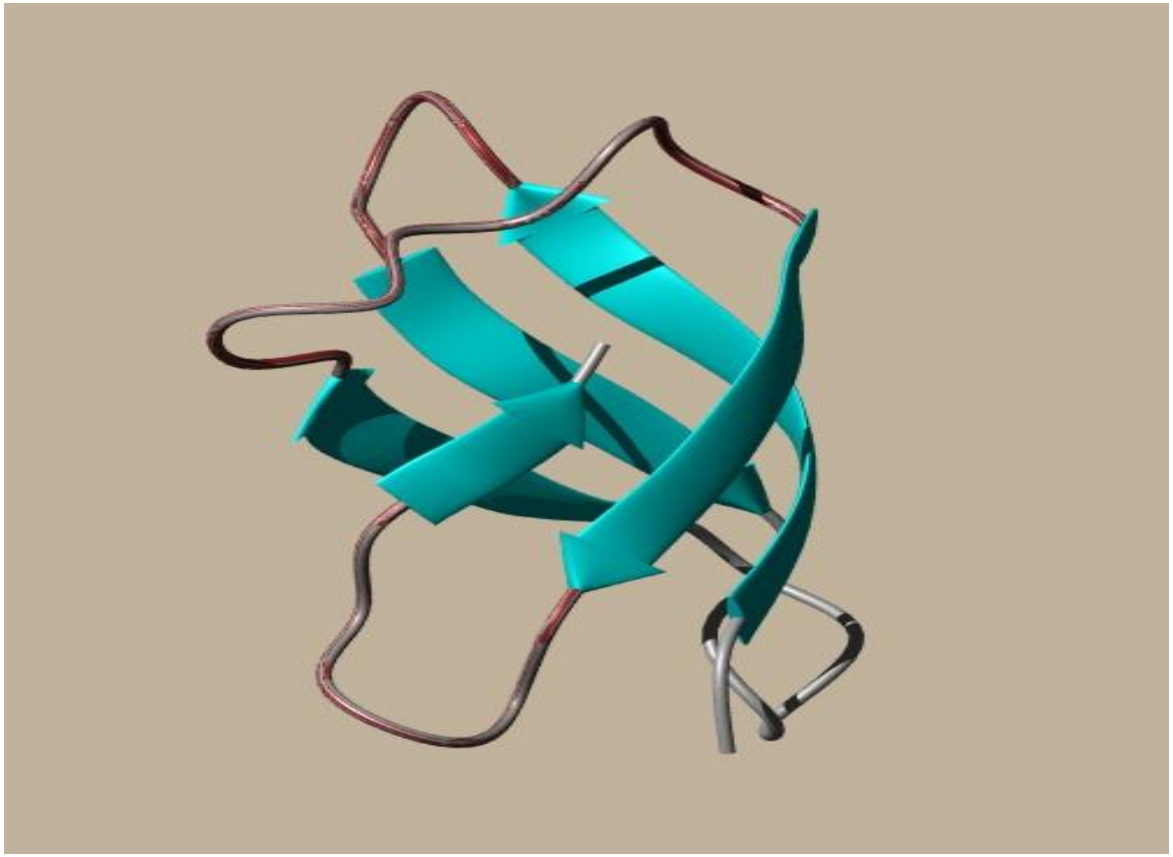
Protein engineering



Why bother with recombinant fusion protein or protein engineering?

1. **to minimize proteolysis.**
2. **for efficient and selective purification.**
3. **to optimize translation efficiency.**
4. **for different applications:**
   antibody production, biochemical experiments, structural biology, industrial usage.

Protein biotechnology or engineering

Definition:

Deliberate design and production of proteins with novel or altered structure and properties, that are not found in natural proteins.

- **To study protein structure and function**
- **Applications in industry (enzymes) and medicine (drugs)**

-- New and improved proteins are always wanted.

Example: Extremophilic proteins have been found in nature (temperatures, salt concentrations, pH values) could be useful.

Outline

1. **Why bother with recombinant fusion protein or protein engineering?**
2. **Principle in recombinant protein expression**
3. **Things need to be considered for recombinant protein expression**

   a. **How to produce?**

   b. How to make an expression recombinant DNA construct?

   c. Where to express?

   d. Difficulties (protein expression problems)

Protein biotechnology or engineering

Definition:

Deliberate design and production of proteins with novel or altered structure and properties, that are not found in natural proteins.

- **To study protein structure and function**
- **Applications in industry (enzymes) and medicine (drugs)**

-- New and improved proteins are always wanted.

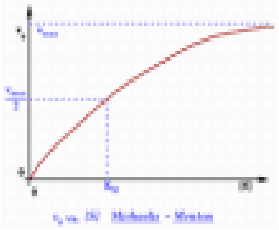
Example: Extremophilic proteins have been found in nature (temperatures, salt concentrations, pH values) could be useful.

Applications

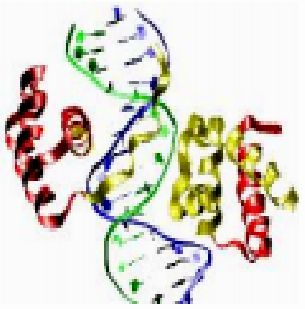
- Functional Studies
- Enzymatic Assays
- Protein-protein interactions
- Protein Ligand Interactions
- Structural Studies

  Protein Crystallography & NMR Structure Determination
- Target Proteins for Rational Drug Design
- Therapeutic Proteins – Preclinical Studies

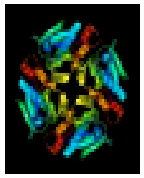
Applications requiring protein expression and purification



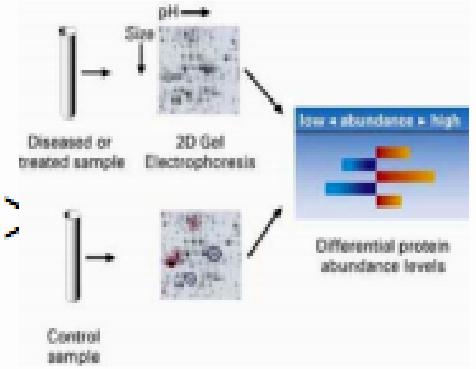
– Enzymology/functional studies (mg)



– Antibody production (µg-mg)



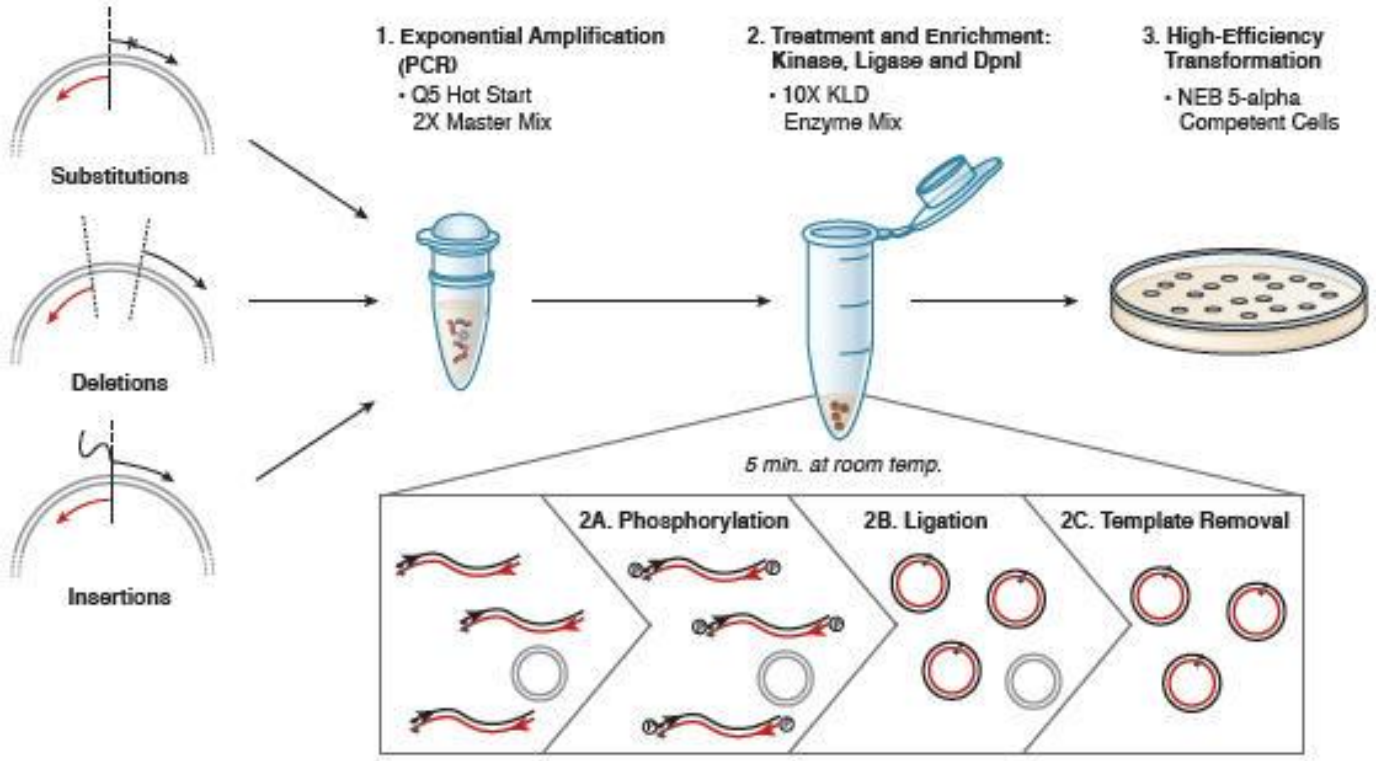
– Structure determination (50 mg or >



– Commercial/therapeutic uses (g)

– *e.g.* Vaccines, immunomodulators

There are two general strategies for protein engineering, **rational design(site directed mutagenesis)** and **directed evolution**.



Site directed mutagenesis

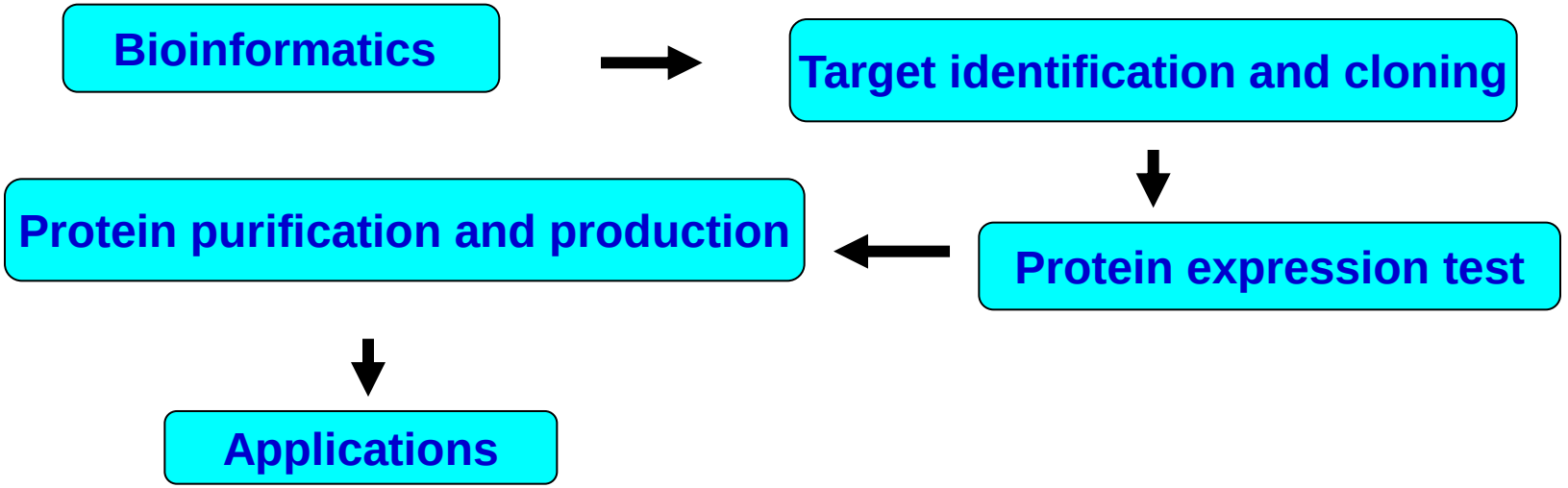
A typical directed-evolution experiment involves three steps:

1- Diversification: The gene encoding the protein of interest is mutated and/or recombined at random to create a large library of gene variants. Techniques commonly used in this step are PCR

2- Selection: The library is tested for the presence of mutants (variants) possessing the desired property using a screen or selection. Screens enable the researcher to identify and isolate high-performing mutants by hand, while selections automatically eliminate all nonfunctional mutants.

3- Amplification: The variants identified in the selection or screen are replicated many fold, enabling researchers to sequence their DNA in order to understand what mutations have occurred

Principle in recombinant protein expression



Things need to be considered for recombinant protein expression

1. **How to produce?**

   choose for protein expression system (vector and host)

2. **How to make an expression recombinant DNA construct?**

   translational or transcriptional fusion, promoter use

3. **Where to express?**

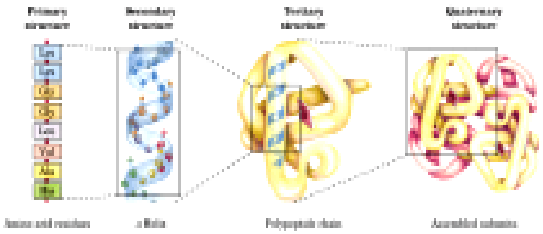
   cytosol, periplasm, secretion, inclusion body

4. **Difficulties (protein expression problems)**

Before you start…..

- Check the literature

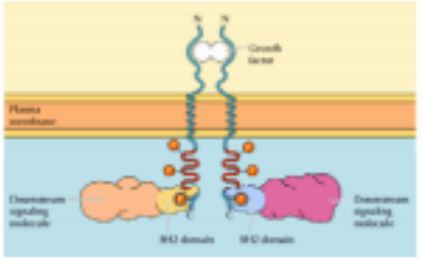
- Any useful predictive bioinformatic information available?
  - www.expasy.org/ **E**xpert **P**rotein **A**nalysis **Sy**stem



- Is the protein membrane-bound or water soluble?
- Is it single domain or multi-domain?
- How to purify and how to identify?
- Do you expect/require PTMs ?
- Do you have an assay for your target protein?

Before you start…..

- Construct design
  - Which cloning protocols?
    - Identify domains, select domains for expression
    - e.g. kinase domain from RTKs for assays and structure based drug discovery



    - Fusion tags ?
  - Which host cell system?
  - Which expression vector?

Which host cell expression system?

- *E. Coli*
- Yeast
- Insect cells
- Mammalian cells
- Cell free

Prokaryotic expression systems-

advantages V disadvantages

- e.g. *Escherichia coli*, *Lactococcus lactis*, *Bacillus* species….
- Widely used for expression of recombinant proteins
- Easy manipulation/transformation, rapid growth, simple nutrient requirements
- Many commercial vectors (*e.g.* Invitrogen, Novagen, Stratagene) available with a variety of N- and C-terminal tags to facilitate purification (*e.g.* His tag, GST fusions, Trx fusions)
- Well suited for proteins to be used for Ab production, structural, functional studies

- Many proteins expressed in inclusion bodies
- No post-translational modifications
- Improper folding of disulphide linked proteins
- Occasional problems with removal of fusion partner
- High endotoxin content with G-ve
  - (Reichelt et al. (2005) Single step protocol to purify recombinant proteins with low endotoxin content. Prot. Expr. Purif.

Characteristic	Bacteria	Yeast	Baculovirus	Mammalian
Cell growth	Rapid (30min)	Rapid (90min)	Slow (18-24h)	Slow (24h)
Med. Complexity	Minimum	Minimum	Complex	Complex
Cost	Low	Low	High	High
Expression	High	Low-High	Low-High	Low-Medium

Post-translational modifications (PTMs)

Protein folding	Not reliable	Usually reliable	Very reliable	Very reliable
N-linked glycosylation	None	High mannose	Simple, no sialic acid	Complex
O-linked glycosylation	N	Y	Y	Y
Phosphorylation	N	Y	Y	Y
Acetylation	N	Y	Y	Y
Acylation	N	Y	Y	Y
γ-carboxylation	N	N	N	Y

Choose of protein expression system

The *KEY* idea is the cloned gene must be transcribed and translated most efficiently.

***Expression vector*: *MAXIMIZE GENE EXPRESSION*.**

***Host*: MINIMIZE TURNOVER OF GENE PRODUCTS (preventing proteolysis *in vivo* in *E. coli)*.**

---- Use protease deficient mutants as hosts.

***Lon* - a major ATP-dependent protease in *E. coli*. Has broad specificity for unfolded or misfolded proteins *in vivo*. *lon* mutants - pleiotropic, but two main phenotypes - mucoidy and UV sensitivity.**

***ompT* - an outer membrane localized protease. Cleaves at paired basic residues.**

***degP* - periplasmic protease - could inactivate some secreted proteins.**

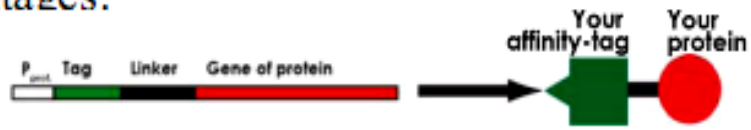
- **BL21(DE3) strain**

Strain	Derivation	Key Feature(s)	Antibiotic Resistance	Available as Competent Cells
Rosetta	Tuner™ (B)	BL21 *lacYZ* deletion, Lacks *lon* and *ompT* proteases	Cam	yes
Rosetta(DE3)			Cam	yes
Rosetta(DE3)pLysS			Cam	yes
Rosetta(DE3)pLacI			Cam	yes
RosettaBlue™	NovaBlue (K-12)	*recA*, *endA*, *lacI*^{q} High transformation efficiency	Tet + Cam	yes
RosettaBlue(DE3)			Tet + Cam	yes
RosettaBlue(DE3)pLysS			Tet + Cam	yes
RosettaBlue(DE3)pLacI			Tet + Cam	yes
Rosetta-gami™	Origami™ (K-12)	*trxB*/*gor* mutant, greatly facilitates cytoplasmic disulfide bond formation	Kan + Tet + Cam	yes
Rosetta-gami(DE3)			Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLysS			Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLacI			Kan + Tet + Cam	yes

Novagen

Protein expression: which vector ?

N- or C-terminal fusions of proteins to short peptides (tags) or to other proteins (fusion partners) offer several potential advantages:



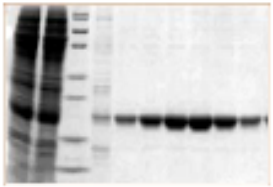
- **Improved expression**

- **Enhanced solubility/folding**
  - E.g. Trx, MBP.
- **Easy detection**

  e.g. Western blot analysis, GFP by fluorescence –

- **Simple purification.**

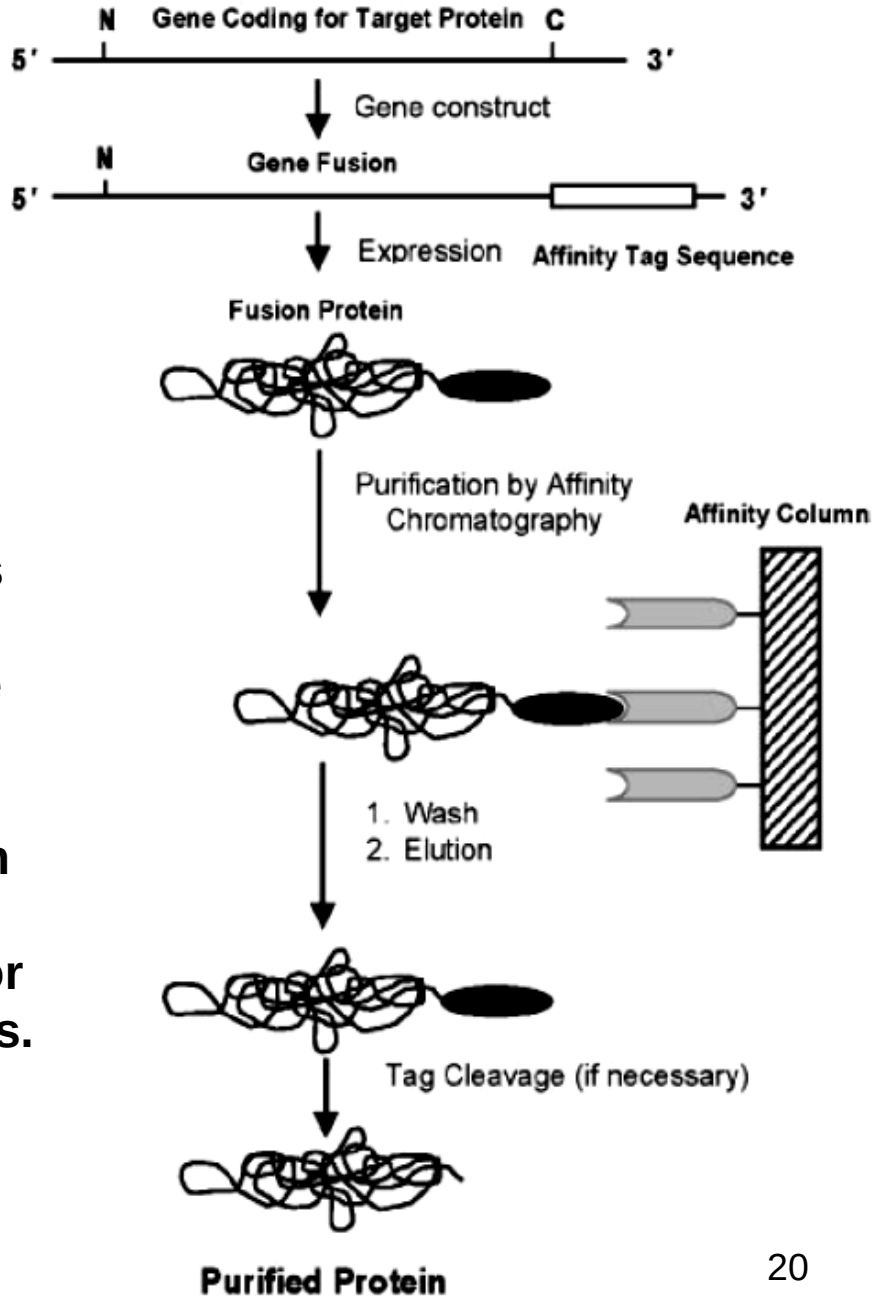
  - fusion partner/tag bind specifically to affinity resins



- Protect from degradation

Increase selectivity of protein purification: (Gene fusion strategies)

Most target protein lack a suitable Affinity ligand usable for capture on a solid matrix. A way to circumvent this obstacle is to genetically fuse the gene encoding the target protein with a gene encoding a purification tag. When the chimeric protein is expressed, the tag allows for specific capture of the fusion protein. This will allow the purification of virtually any protein without any prior knowledge of its biochemical properties.



Advantages and disadvantages for using tags in fusion proteins

Plus factors:

(1)improve protein yield (2) prevent proteolysis (3) facilitate protein refolding (4) protect the antigenicity of the fusion protein and (5) increase solubility

Minus factors:

(1) a change in protein conformation (2) lower protein yields (cleavage may not be complete) (3) inhibition of enzyme activity (4) alteration in biological activity (5) undesired flexibility in structural studies (6) cleavage/removing the fusion partner requires expensive protease (Factor Xa, enterokinase) and (7) toxicity.

Commonly used affinity tag system in recombinant protein expression:

1. **expression and purification of maltose-binding protein fusions. (provides a factor Xa cleavage site).**

2. **expression and purification of Glutathione-S-transferase fusion proteins. (contains either a thrombin cleavage site, a factor Xa cleavage site, or an Asp-Pro acid cleavage site).**
3. **expression and purification of thioredoxin fusion proteins. (provides an enterokinase cleavage site).**
4. **expression and purification of 6X His-tagged proteins.**

Tags and Fusions

Fusion	Size	Purification
GST	26 kDa	Glutathione Sepharose
MBP	40 kDa	Amylose resin
Trx	12 kDa	Thiobond

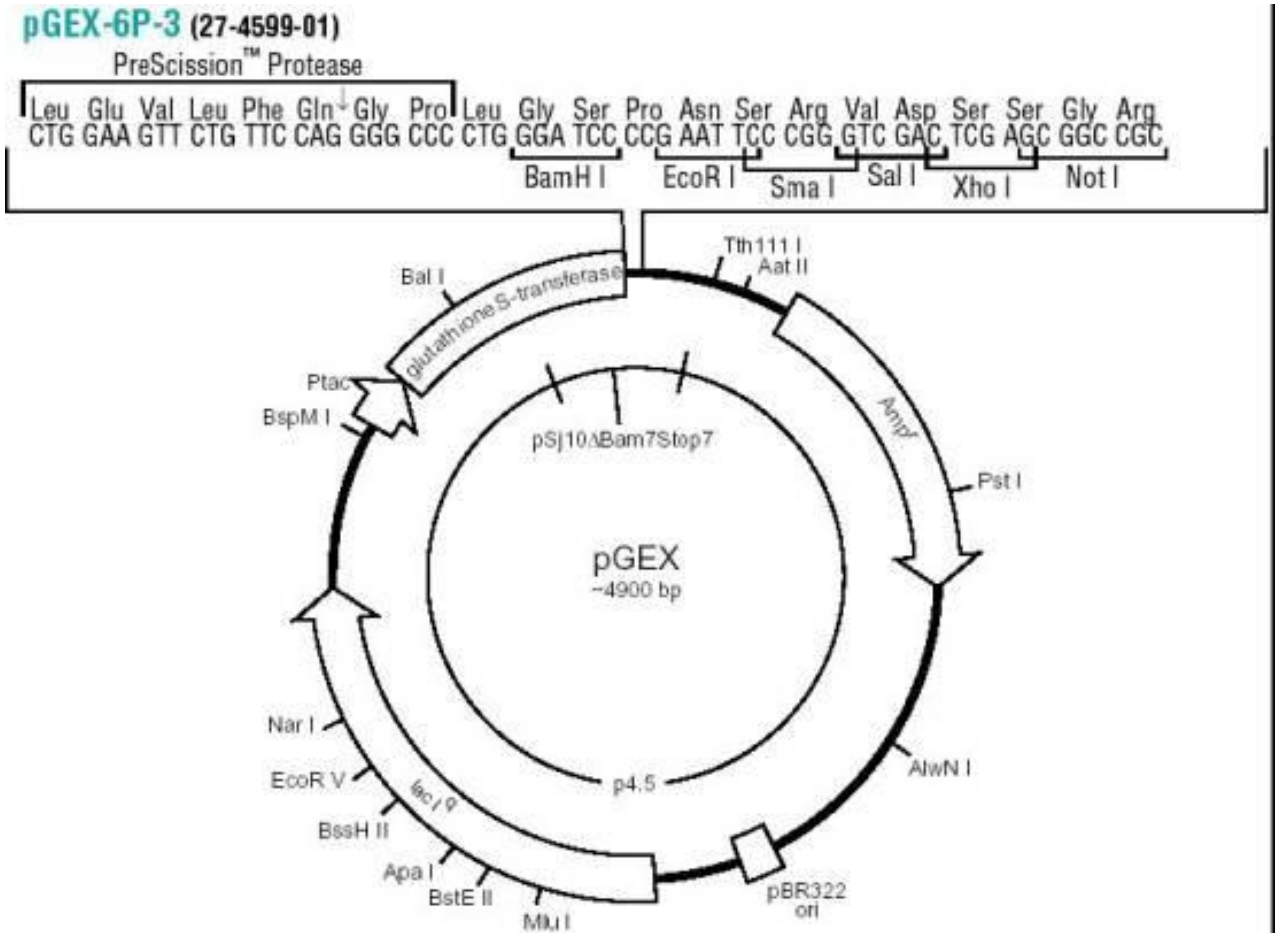
Hunt I, (2005) Protein Exp & Purif 40, 1-22 From gene to protein: a review of new and enabling technologies for multi-parallel protein expression.

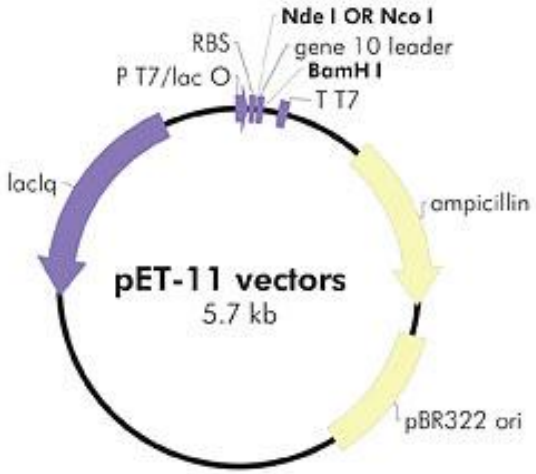
Construct design – fusion cleavage sites

- often included to allow removal of tags/fusions

Enzyme	Recognition site	Comments
Thrombin	LVPR/GS	Less specific
Factor Xa	IEGR/	Less specific
Enterokinase	DDDDK/	Very specific

Waugh DS (2005) Trends Biotechnol.23(6):316-20 Making the most of affinity tags


pGEX-6P-3 (27-4599-01)
PreScission™ Protease
Leu Glu Val Leu Phe Gln↓Gly Pro Leu Gly Ser Pro Asn Ser Arg Val Asp Ser Ser Gly Arg
CTG GAA GTT CTG TTC CAG GGG CCC CTG GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC
BamH I
EcoR I
Sma I
Sal I
Xho I
Not I
Tth111 I
Aat II
Bal I
glutathione S-transferase
Ptac
BspM I
pSj10ΔBam7Stop7
Amp
Pst I
pGEX
~4900 bp
Nar I
EcoR V
lacIq
BssH II
Apa I
BstE II
MluI
p4.5
AlwN I
pBR322
ori



pET-11a–11d Cloning Site Regions
sequence shown (74–130)

pET-11a GAAGGAGATATACAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGC GGA TCC
M A S M T G G Q Q M G R G S
RBS — START — T7 gene 10 leader peptide

pET-11b GAAGGAGATATACAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG GAT CCG
M A S M T G G Q Q M G R D P
RBS — START — T7 gene 10 leader peptide

pET-11c GAAGGAGATATACAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG ATC CGG
M A S M T G G Q Q M G R I R
RBS — START — T7 gene 10 leader peptide

pET-11d GAAGGAGATATACC ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG ATC CGG C
M A S M T G G Q Q M G R I R
RBS — START — T7 gene 10 leader peptide

Feature	Nucleotide Position			
	pET-11a	pET-11b	pET-11c	pET-11d
T7 promoter with *lac* operator	1–43	1–43	1–43	1–43
ribosome binding site (RBS)	74–80	74–80	74–80	74–80
Nde I (pET-11a–c) or Nco I (pET-11d) cloning site	86–91	86–91	86–91	86–91
T7 gene 10 translated leader	89–121	89–121	89–121	88–120
BamH I cloning site	125–130	124–129	123–128	122–127
T7 terminator	199–245	198–244	197–243	196–242
ampicillin resistance (*bla*) ORF	657–1514	656–1513	655–1512	654–1511
pBR322 origin of replication	1665–2332	1664–2331	1663–2330	1662–2329
lacI^{q} repressor ORF	4212–5291	4211–5290	4210–5289	4209–5288

Where to express the recombinant proteins?

1. **Direct expression (cytosol):** *E. coli* cytoplasm is a reducing environment - difficult to ensure proper disulphide bonds formation.

2. **Fusion expression (inclusion body?):** Ensures good translation initiation. Can overcome insolubility and/or instability problems with small peptides. Has purification advantages based on affinity chromatography.

3. **Secretion (periplasm or medium):** a fusion alternative when proteins are fused to peptides or proteins targeted for secretion. Periplasm offers a more oxidizing environment, where proteins tend to fold better.

General problems with heterologus gene expression

***(a) Not enough protein is produced*:**

***(b) Enough protein is produced, but it is insoluble*:**

The *KEY* idea is to slow down the expression rate of protein.

Insolubility of proteins produced in E.coli

Inclusion bodies.

Dense particles, containing precipitated proteins. Their formation depends on protein synthesis rate, growth conditions.

Advantages: proteolysis resistant, big yield, relatively pure, easy to separate.

Disadvantages: inactive product requires in vitro refolding and renaturation

Refolding of recombinant proteins

Solubilisation:

High T ^{0} C, detergents, high concentration of inorganic salts or organic solvents all used. The most commonly used organic solutes such as urea or guanidine-HCl often used in the presence of reducing agents (mercaptoethanol or DTT). Solubilized proteins can be purified by ion-exchange chromatography or other conventional methods, prior to refolding.

Refolding:

If no S-S bonds present - remove denaturing agent to allow protein to fold correctly. If S-S bonds present - their formation can be accomplished: by air oxidation, catalysed by trace metal ions; by a mixture of reduced and oxidized thiol compounds - oxidized DTT, reduced DTT; GSSG/GSH; cystine and cysteine, cystamine and cysteamine.