Department of BSBE
Indian Institute Of Technology Guwahati

# BT 207

Dr. Sanjukta Patra

# Vectors -3

## Expression vector

Elements of expression vectors

- an origin of replication
- a selectable marker
- and a suitable site for the insertion of a gene like the multiple cloning site

- Usually a plasmid or virus is designed for gene expression in cells.
- The vector is used to introduce a specific gene into a target cell, and can commandeer the cell's mechanism for protein synthesis to produce the protein encoded by the gene.
- The vector is engineered to contain regulatory sequences that act as enhancer and promoter regions and lead to efficient transcription of the gene carried on the expression vector.
- The goal of a well-designed expression vector is the efficient production of protein, achieved by the production of significant amount of stable messenger RNA, which can then be translated into protein.
- The expression of a protein may be tightly controlled, and the protein is only produced in significant quantity when necessary through the use of an inducer, in some systems however the protein may be expressed constitutively.
- *Escherichia coli* is commonly used as the host for protein production, but other cell types may also be used.

## Prokaryotic expression vector

- Promoter - commonly use inducible promoters derived from *lac* operon and the T7 promoter. Other strong promoters used include Trp promoter and Tac-Promoter, which are a hybrid of both the Trp and Lac Operon promoters.
- Ribosome binding site (RBS) - follows the promoter, and promotes efficient translation of the protein of interest.
- Translation initiation site - Shine-Dalgarno sequence enclosed in the RBS, 8 base-pairs upstream of the AUG start codon.

## Eukaryotes expression vector

Eukaryote expression vectors require sequences that encode for:

- Polyadenylation tail: Creates a polyadenylation tail at the end of the transcribed pre-mRNA that protects the mRNA from exonucleases and ensures transcriptional and translational termination: stabilizes mRNA production.
- Minimal UTR length: UTRs contain specific characteristics that may impede transcription or translation, and thus the shortest UTRs or none at all are encoded for in optimal expression vectors.
- Kozak sequence: Vectors should encode for a Kozak sequence in the mRNA, which assembles the ribosome for translation of the mRNA.

**Prokaryotic expression vectors**

**pET 28a**

**Examples** of mammalian **expression vectors** include

- the adenoviral **vectors**
- the pSV
- the pCMV series of plasmid **vectors**
- vaccinia and retroviral **vectors**
- baculovirus

The promoters for cytomegalovirus (CMV) and SV40 are commonly used in mammalian **expression vectors** to drive gene **expression**.

A **shuttle vector** - a plasmid that can propagate in two different host species.
DNA inserted into a shuttle vector can be tested or manipulated in two different cell types.
Advantage - they can be manipulated in *E. coli*, then used in a system which is more difficult or slower to use (e.g. yeast).

1. Shuttle vectors include plasmids that can propagate in eukaryotes and prokaryotes (e.g. both *Saccharomyces cerevisiae* and *Escherichia coli*.

2. In different species of bacteria example - both *E. coli* and *Rhodococcus erythropolis*.

3. Adenovirus shuttle vectors, which can propagate in *E. coli* and mammals.

Shuttle vectors are frequently used to quickly make multiple copies of the gene in *E. coli* (amplification).

Example - yeast shuttle vector. Yeast shuttle vectors have components that allow for replication and selection in both *E. coli* cells and yeast cells.

The *E. coli* component of a yeast shuttle vector includes an origin of replication and a selectable marker, e.g. antibiotic resistance, beta lactamase, beta galactosidase.

The yeast component of a yeast shuttle vector includes an autonomously replicating sequence (ARS), a yeast centromere (CEN), and a yeast selectable marker (e.g. URA3, a gene that encodes an enzyme for uracil synthesis).

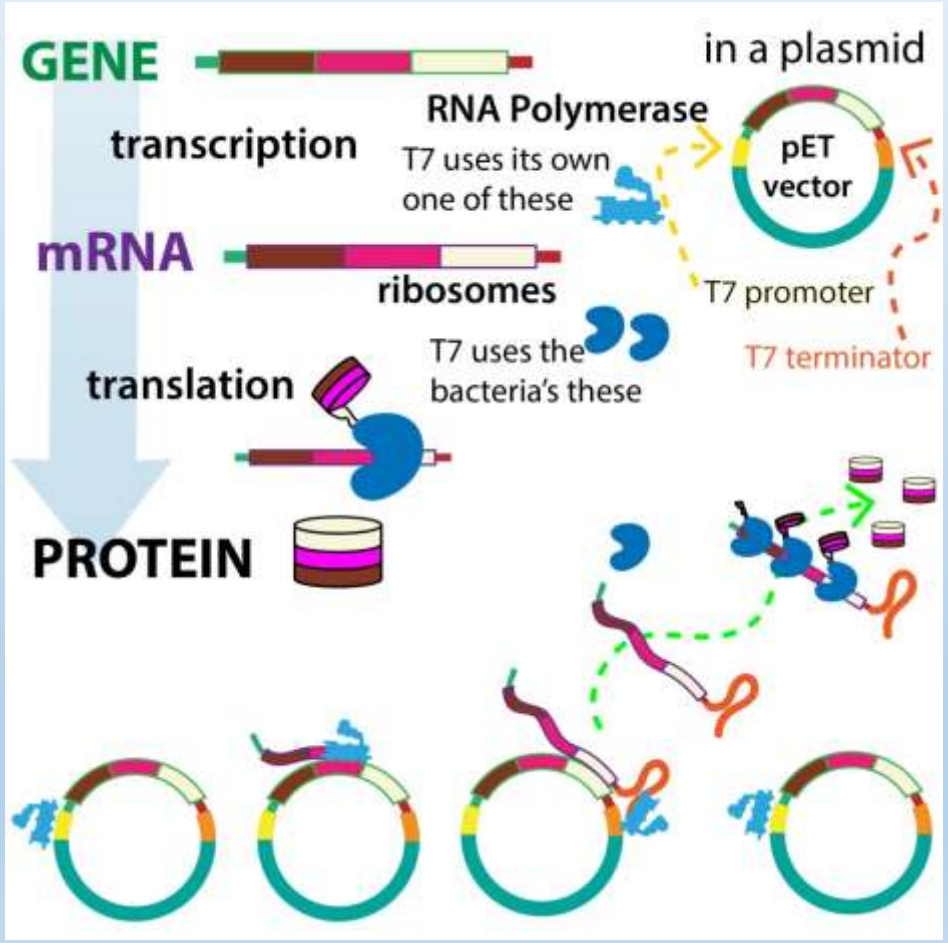
# Eukaryotic and Shuttle Vectors

- Because different organisms use different origins of replication and regulatory signals, different cloning vectors must be used in different species.

- Special cloning vectors can replicate in prokaryotes and in eukaryotes.

- **Shuttle vectors** can replicate in *E. coli* and in other species.

# pET28a EXPRESSION SYSTEM



# EXPRESSION VECTOR COMPONENTS

| Goal | Component |
|---|---|
| Insert cargo into the plasmid<br>and verify the insert sequence accuracy | MCS – restriction sites OR recombination regions 5' and 3' Primer sites for sequence verification |
| Insert plasmid into cells,<br>enable the plasmid to replicate inside the host, & select for cells carrying the plasmid | •Backbone compatible with cloning method<br>• Origin of replication<br>• Selection marker and/or screening marker |
| Transcribe mRNA from the plasmid | •Promoter (constitutive or inducible) operator, terminator |
| Translate mRNA into protein | Ribosome Binding Site, start codon, stop codon |
| Promote proper folding of nascent protein | •co-expression of chaperones<br>•Solubilization tags<br>•custom-designed synthetic RBS<br>•Codon-optimized ORF |
| Detect or Purify target protein | •Epitope tags (His)<br>•reporters (GFP) |

# Why is pET28a a good expression vector?

**pET-28a(+) sequence landmarks**

| | |
|---|---|
| T7 promoter | 370-386 |
| T7 transcription start | 369 |
| His•Tag coding sequence | 270-287 |
| T7•Tag coding sequence | 207-239 |
| Multiple cloning sites (*Bam*H I - *Xho* I) | 158-203 |
| His•Tag coding sequence | 140-157 |
| T7 terminator | 26-72 |
| *lacI* coding sequence | 773-1852 |
| pBR322 origin | 3286 |
| Kan coding sequence | 3995-4807 |
| f1 origin | 4903-5358 |

- ✓ Strong promoter
- ✓ Intact ORF
- ✓ Ribosomal binding site
- ✓ Termination sequence

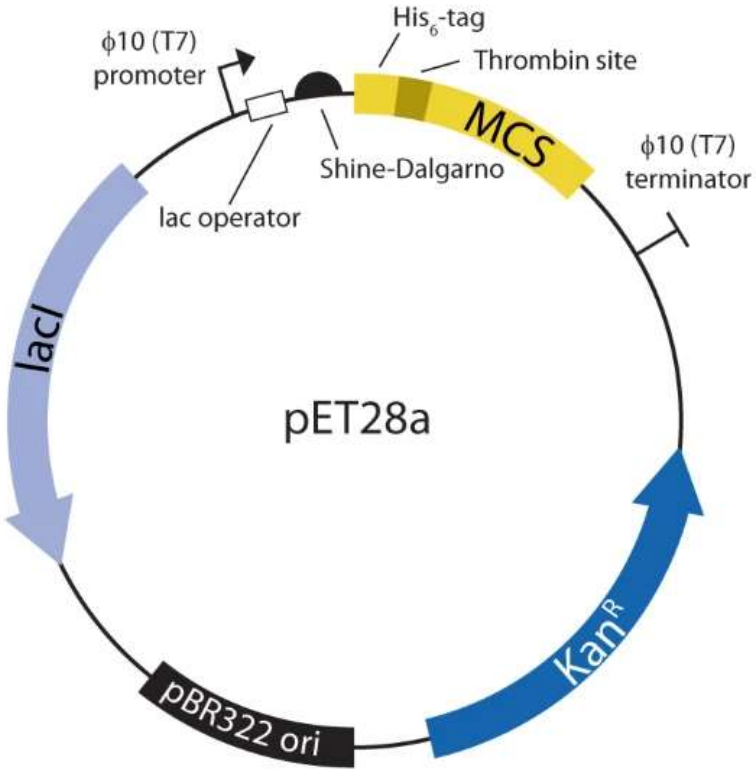
**Strong promoter** is the one which promotes “strong” or “very high level of transcription rate of mRNA” from downstream DNA sequence

- The T7 RNA polymerase (T7 promoter) and the transcription terminator (T7 terminator) when incorporated into the pBR322 backbone led to the pET nomenclature (plasmid for expression by T7 RNA polymerase).

- These expression plasmids support high levels of transcription in strains of *Escherichia coli* that contain a lysogenised DE3 phage fragment encoding the T7 RNA polymerase for making proteins.

- The *lac* operator sequence adjacent to T7 promoter is included to suppress uninduced expression by the lac repressor molecule. Translation initiation is mediated by a Shine–Dalgarno (SD) sequence.

- In a typical experiment, the coding sequence to be expressed is cloned downstream of, and in frame with, the coding sequence for a poly-histidine purification tag ($His_6$) and a thrombin protease recognition site (TPS) included in the pET backbone so that the recombinant protein produced can be easily purified.

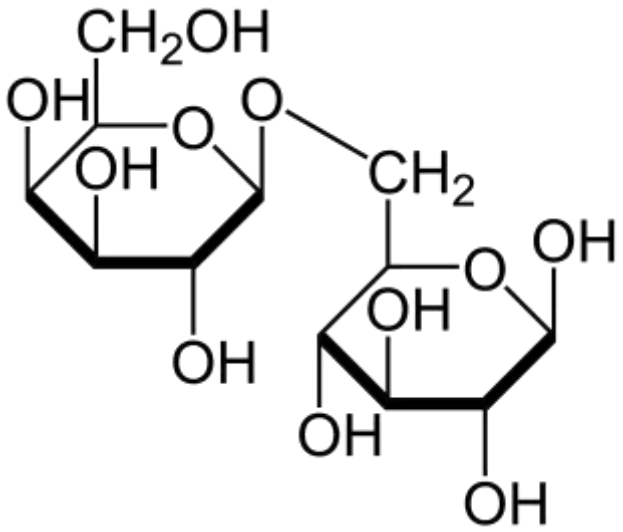


# What controls the expression??

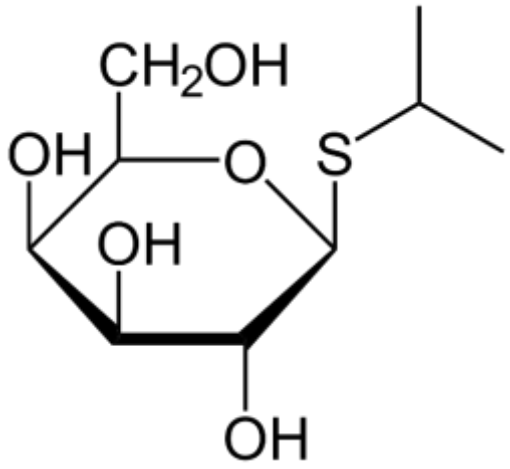
**Normal function-** **NO** protein expression (**LacI** protein represses transcription by blocking T7 RNA polymerase expression)

**Altered function-** protein expression (**IPTG** binds to Lac repressor protein and expresses T7 RNA polymerase for transcription)

Like **allolactose**, **IPTG** binds to the lac repressor and releases the tetrameric repressor from the lac operator in an **allosteric manner**, thereby allowing the transcription of genes in the lac operon!




Allolactose



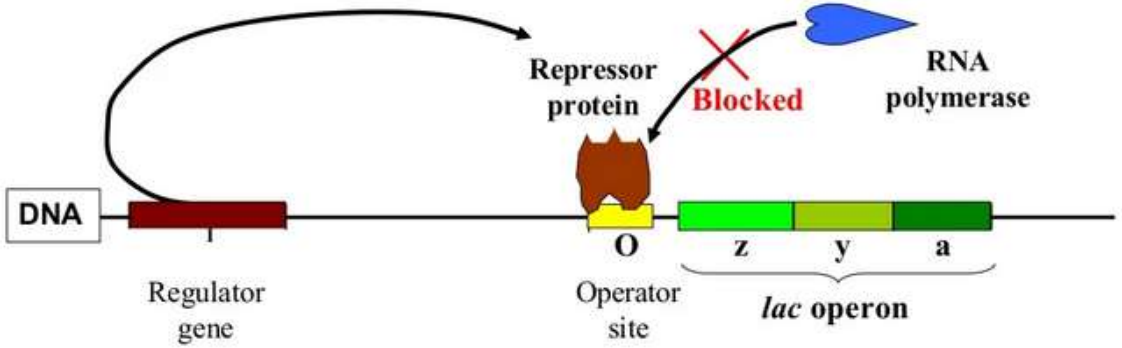

IPTG

Isopropyl β-d-1-thiogalactopyranoside (**IPTG**) is a molecular mimic of **allolactose**, a lactose metabolite that triggers transcription of the lac operon, and it is therefore used to induce protein expression where the gene is under the control of the lac operator

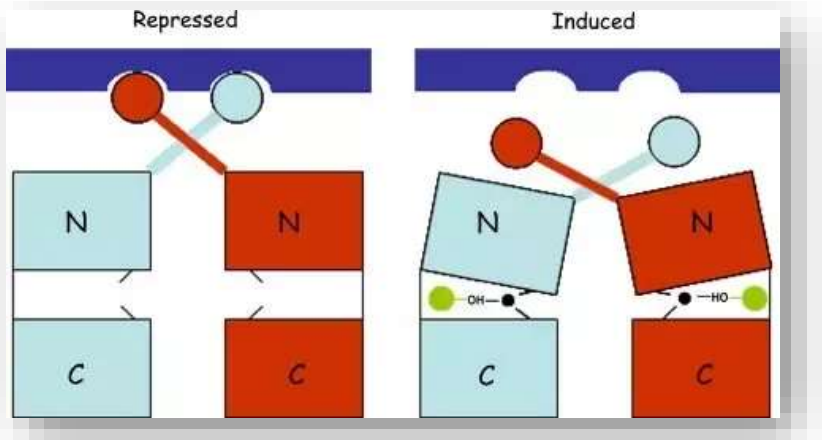
# Regulation of *lac* system

## 1. When lactose is absent

- A repressor protein is continuously synthesised. It sits on a sequence of DNA just in front of the *lac* operon, the **Operator site**
- The **repressor protein** blocks the **Promoter site** where the RNA polymerase settles before it starts transcribing

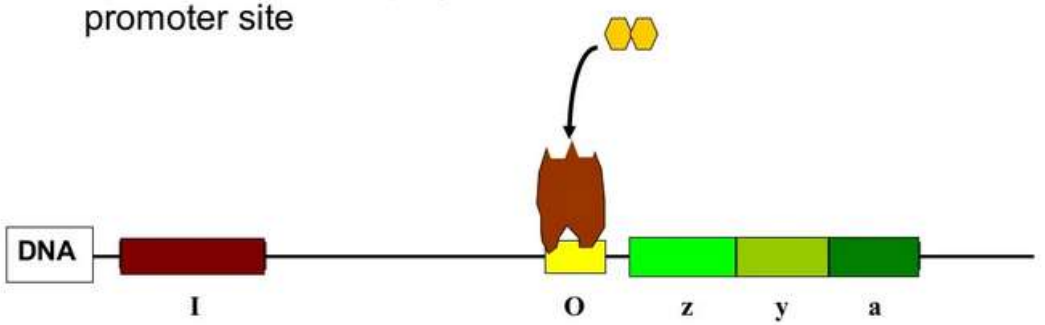



At the molecular level this looks like…



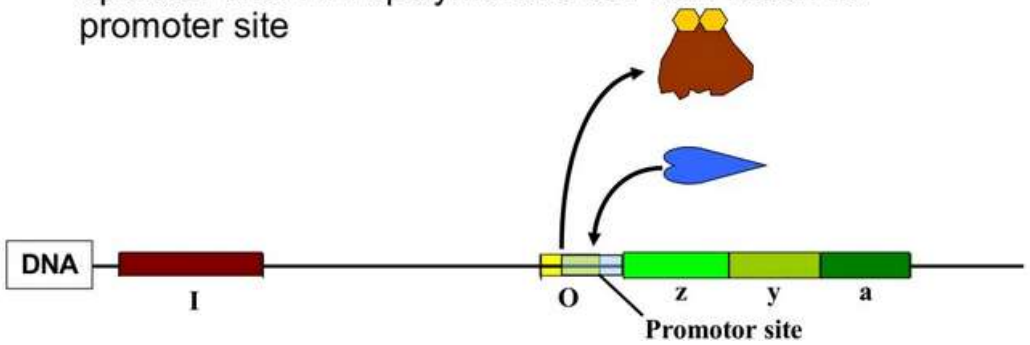

## 2. When lactose is present

- A small amount of a sugar allolactose is formed within the bacterial cell. This fits onto the repressor protein at another active site (**allosteric site**)
- This causes the repressor protein to change its shape (a **conformational change**). It can no longer sit on the operator site. RNA polymerase can now reach its promoter site





- This causes the repressor protein to change its shape (a **conformational change**). It can no longer sit on the operator site. RNA polymerase can now reach its promoter site

# A pictorial representation of overexpression using pET vectors, what exactly goes on inside the host cell

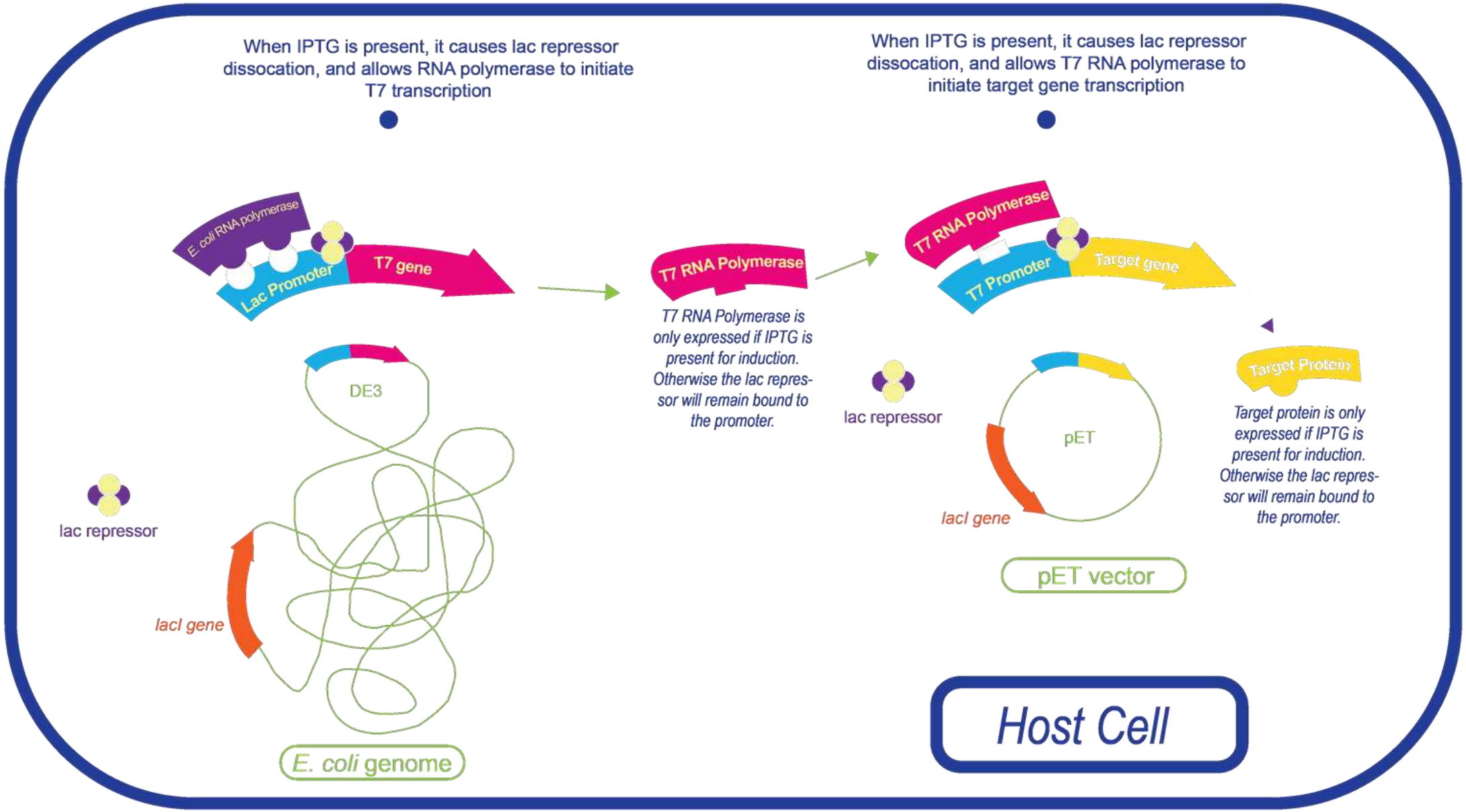
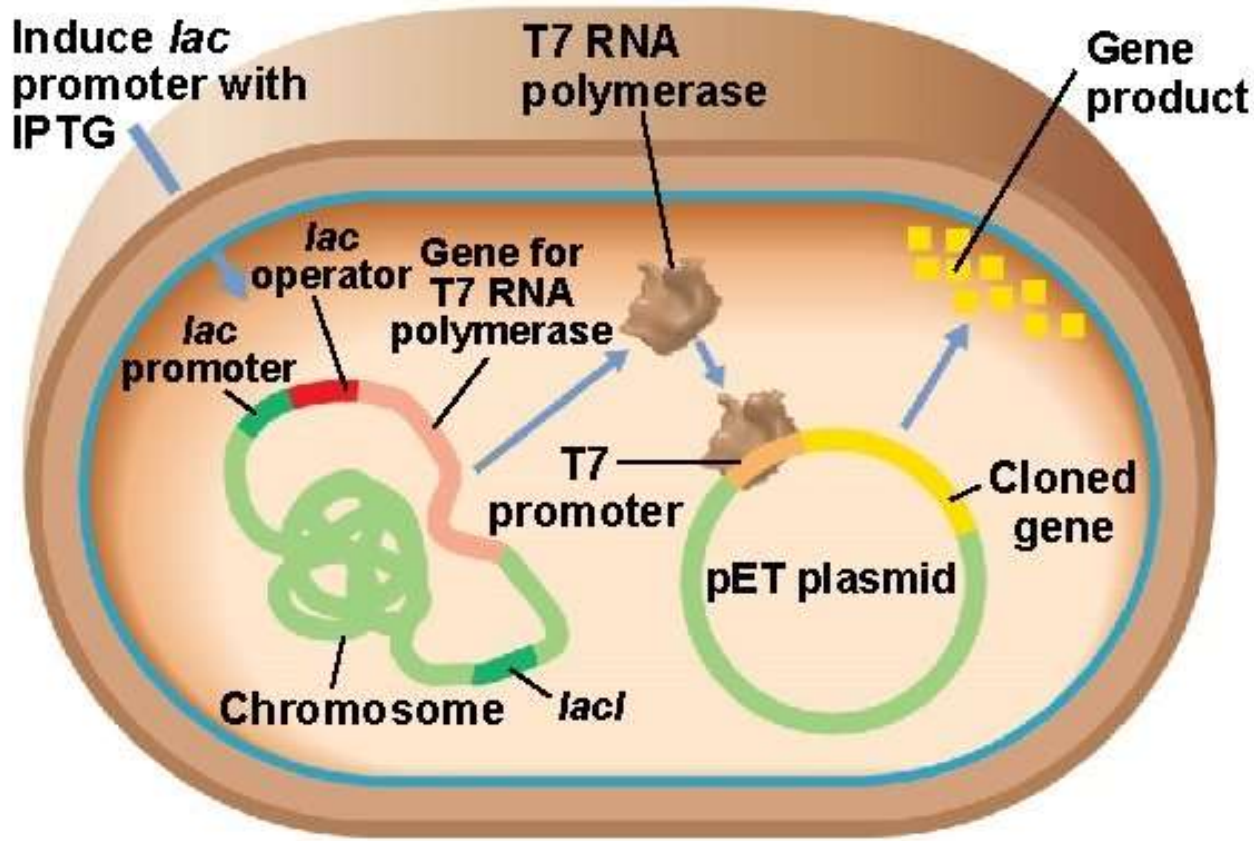
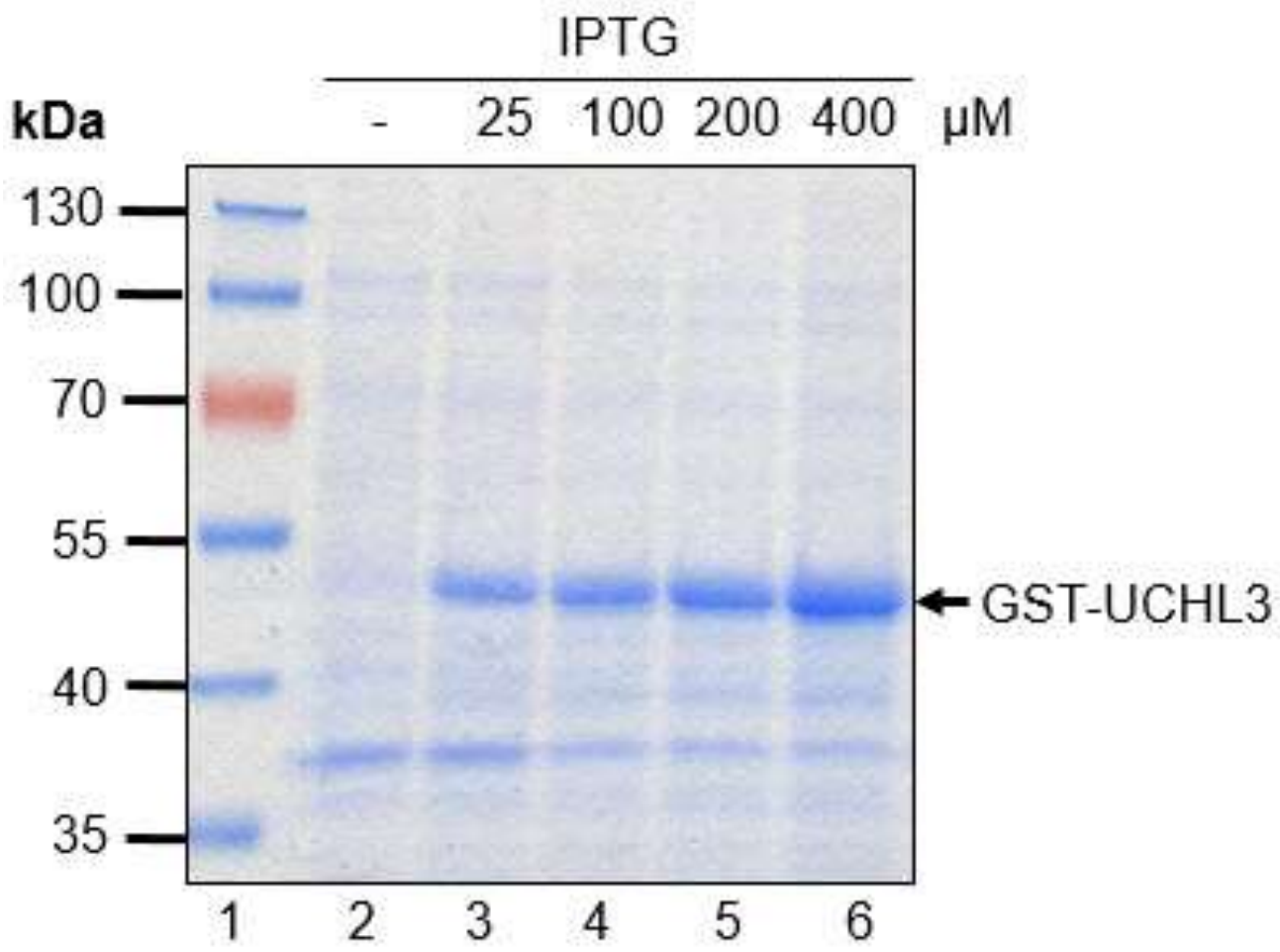


Figure 11.17



# A sample image of IPTG induced expression of protein in *E.coli*

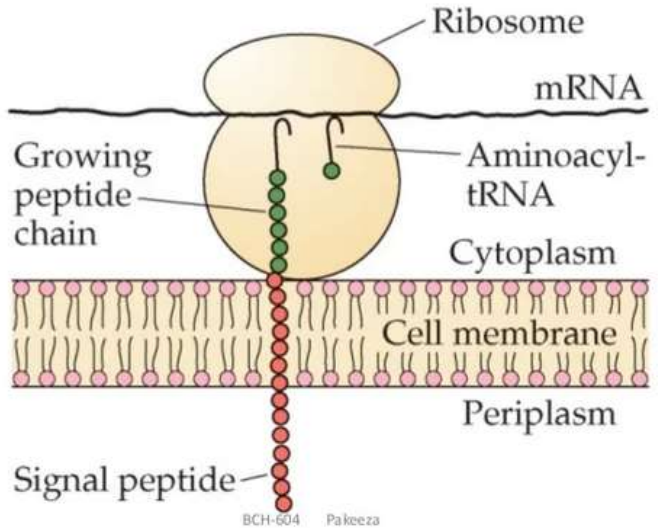



*Coomassie dye stained SDS-PAGE done for visualization of overexpressed proteins in the laboratory*

Expression of GST-UCHL3 in BL21 (DE3) cells was induced by the addition of 25 μM (lane 3), 100 μM (lane 4), 200 μM (lane 5) or 400 μM (lane 6) IPTG at 37 °C for 2 hours. Lane 2 without IPTG.

# What type of proteins can be expressed using pET??

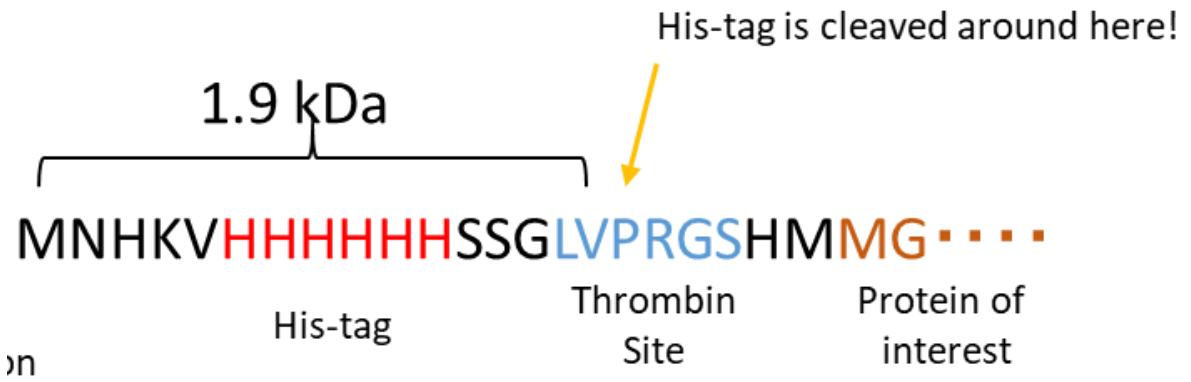
- The signal peptide at the N-terminal end facilitates its export by enabling the protein to pass through the cell membrane.






- GC rich genes can be cloned and expressed
- Prokaryotic proteins
- Fusion proteins
- Industrially important enzymes
- Secretory proteins (with signal peptides)
- Other codon optimized genes can be cloned and expressed

After expression of proteins and subsequent purification, the thrombin site allows enzymatic removal of the unnecessary poly His tract to form the mature protein



## What is the pET advantage??

The pET plasmids have a medium copy number. (~20-25 per cell), which can be helpful because it has optimum copy number that is neither too high to cause huge amounts f proteins being made and ending up in formation of inclusion bodies neither it is too low. It allows for the high expression level of the T7 promoter without overloading the cell with many copies of the plasmid in addition.

## APPLICATION OF EXPRESSION VECTOR

- ✓ ➢ The construction of **expression libraries**.
- ✓ ➢ The **analysis of gene function** at protein level.
- ✓ ➢ The commercial **production of proteins**.
- ✓ ➢ The **production of antibodies**.
- ✓ ➢ For **in vivo studies** of the protein.

## What is the pET disadvantage??

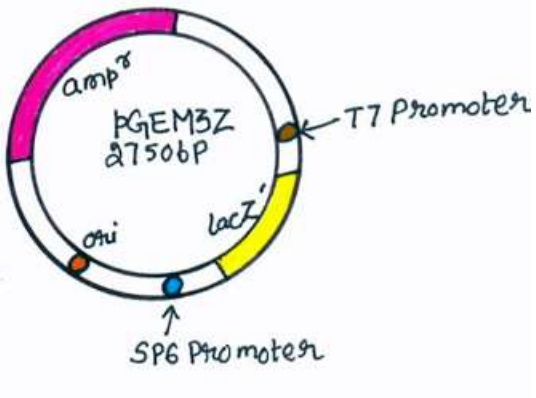
Despite the strong selectivity of the T7 promoter for its phage-encoded polymerase, residual "leaky" expression of very toxic proteins from the basic pET constructs can be sometimes lethal to the cell.

## Challenges in overexpression using pET

- Low expression
- Degradation by bacterial proteases
- Improper folding
- Oxygen limitations

**PGEM3Z plasmid**

- pGEM3Z is having similarity with a pUC vector because of the presence of the amp$^R$ genes and *lacZ'* genes containing a cluster of restriction sites.
- The difference is that pGEM3Z has two additional, short pieces of DNA, each of which acts as the recognition site for attachment of an RNA polymerase enzyme.
- These two promoter sequences lie on either side of the cluster of restriction site used for introduction of new DNA into the pGEM3Z molecule.
- **Advantages:**
- High copy number
- Identification can be achieved by single step by plating on agar medium containing ampicillin and X gal.



# Questions