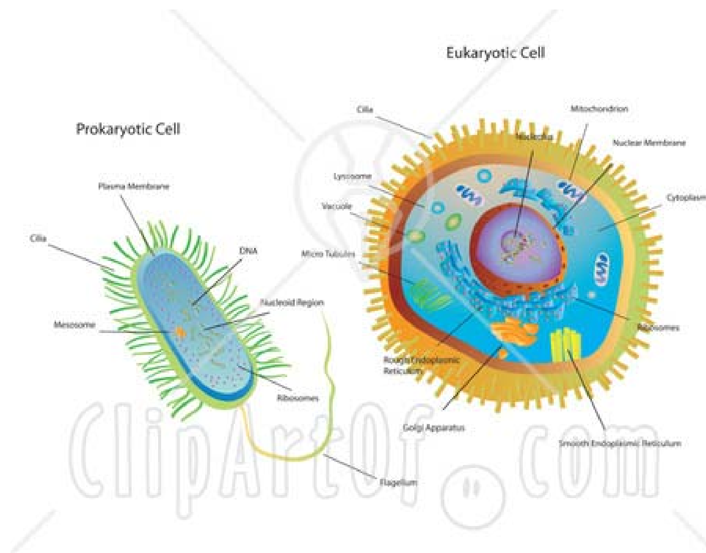
# Heterologous Protein Production in Eukaryotic Cells

# Difference between Eukaryotic and Procaryotic Cells




The genetic information of eukaryotic cell is contained within chromosomes, formed from single, long DNA that contains many genes, introns, many replication origins, one centromere, and two telomeres. DNA is bounded to proteins - histones.

Eukaryotic cells contain nucleus and different organels separated by membrane: ER, Golgi apparatus, lysozomes, mitochondria.

# Expression of Proteins in Eukaryotic Systems

- Yeast

  - *Saccharomyces cerevisiae*
  - *Pichia pastoris*
  - *Schizosaccharomyces pombe*

- Insect Cells – Baculovirus

- Mammalian Cells

There is no universally effective eukaryotic host cell that performs correctly every posttranslation modification.
Testing of different systems

# Advantages of Eukaryotic Systems

Rapid growth (doubling in 90 minutes)

Simple manipulation with the cells

Posttranslation modification

Cheap production

Number of strong constitutive promoters have been used to drive target gene expression

Number of expression vectors and mutated host strains have been prepared

Possibility to direct proteins to secretion (signal sequence for secretion)

Different codon usage (96% of amino acids is encoded only by 25 codons from 61 available combinations)

Lower yield of expressed proteins

# Postranslation modification of proteins

Functions of glycosylation:

- Protection from degradation
- Protein Folding
- Transport and packaging signals
- Cell communication when displayed on the outer membrane as glycocalyx
- About 50% of eukaryotic proteins are glycosylated

**Glycosylation** in ER and Golgi apparatus by specific enzymes (glycosylases and glycosyltransferases)

Attachment of specific sugar to the hydroxyl group of serine or threonine (O-linked glycosylation)

and to the amide group of asparagine (N-linked glycosylation)

# Posttranslation modification of proteins

**Phosphorylation**: mainly on Ser and Tyr, also on Asp and His

- Protein phosphorylation is probably the most important regulatory event. Many enzymes and receptors are switched "on" or "off" by phosphorylation and dephosphorylation.

**Acylation:** the free α-amino group at the N-terminus is blocked by an acetyl

Histones are acetylated and deacetylated on lysine residues, this process is important for regulation of gene expression. Acetylation decreases the interaction of histones with DNA. The condensed chromatin is transformed into a more relaxed structure.

# Posttranslation modification of proteins

**Modification of Lys residue**: acetylation
biotinilation
hydroxylation
methylation
ubiquitinylation - regulation of protein degradation

**Proteolytic cleavage** of protein precursors-regulation of activity
pepsinogen / pepsin
preproinsulin/ insulin

# Posttranslation modification of proteins

**Protein methylation**: Arg and Lys residues: enzymes involved in methylations: lysine and arginine methyl transferases (PRMT)

Methylation of histones, transcription factors, and other activators:

The role in: signal transduction,
mRNA splicing,
transcriptional control (repression and activation),
DNA repair,
protein translocation

# Posttranslation modification of proteins

**Myristoylation**: covalent attachment to the α-amino group of an N-terminal amino acid of the polypeptide (very frequently Gly)

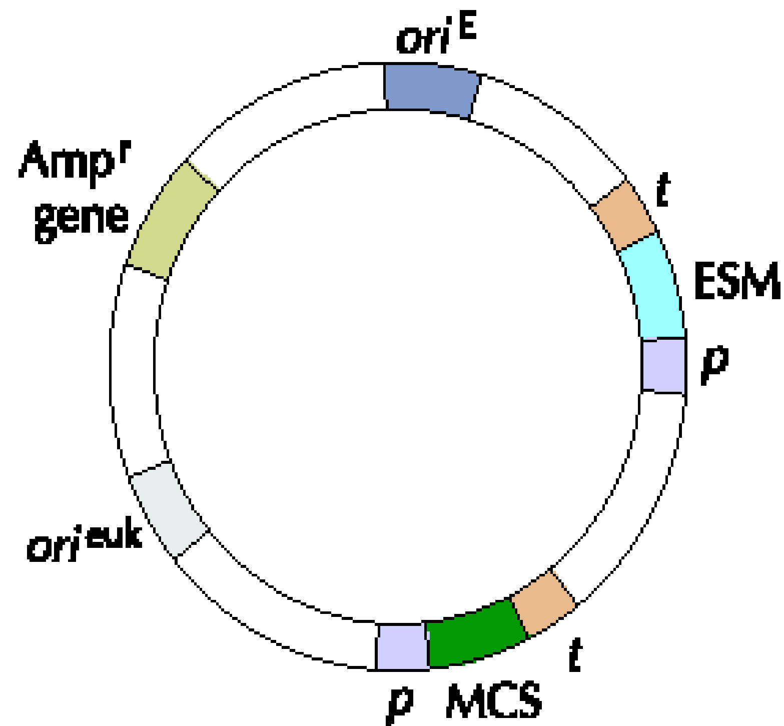
Myristoylation plays a role in membrane targeting and signal transduction in plant responses to environmental stress.

**Palmitoylation:** covalent attachment of fatty acid to cysteine from membrane proteins.

Palmitoylation enhances the hydrophobicity of proteins and contributes to their membrane association; plays a role in subcellular trafficking of proteins between membrane compartments.

# Generalized Eukaryotic Cloning Vector

**Shuttle vector**




Plasmids can be transformed in both, bacteria and yeasts

- p- promoter
- MCS- multiple cloning sites- for a gene of interest
- t- termination and polyadenylation signals
- ESM- eukaryotic selectable marker
- $Amp^r$ -*E. coli* selectable marke
- $ori^E$ -prokaryotic origin of replication
- $ori^{euk-}$ -eukaryotic origin of replication

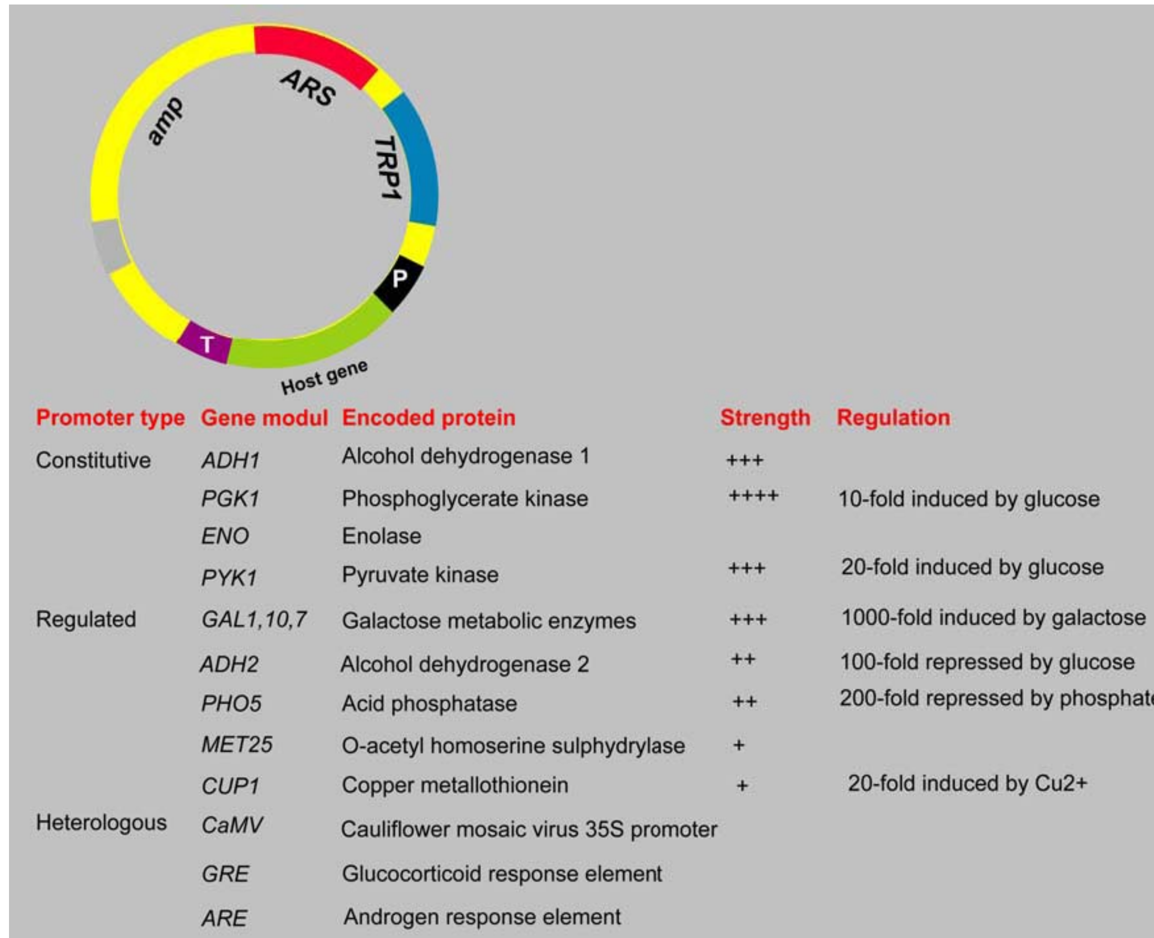
# Selectable markers

| Gene | Enzyme | Selection |
|---|---|---|
| HIS3 | Imidazole glycerolphosphate dehydratase | histidine |
| LEU2 | Isopropylmalate dehydrogenase | leucine |
| LYS2 | Aminoadipate reductase | lysine |
| TRP1 | N-(5´-phosphoribosyl)-anthranilate isomerase | tryptophan |
| URA3 | Orotidine-5´-phosphate decarboxylase | uracil |

Strain containing particular marker (*LEU)* is auxotropic for Leu.
This strain carries a mutation inactivating isopropylmalate dehydrogenase that renders it unable to synthesize leucine and will only be able to grow if Leu can be taken up from the medium during cultivation.

# Promoters

## Tightly, regulatable, inducible promoters are preferred




| Promoter type | Gene modul | Encoded protein | Strength | Regulation |
|---|---|---|---|---|
| Constitutive | *ADH1* | Alcohol dehydrogenase 1 | +++ | |
| | *PGK1* | Phosphoglycerate kinase | ++++ | 10-fold induced by glucose |
| | *ENO* | Enolase | | |
| | *PYK1* | Pyruvate kinase | +++ | 20-fold induced by glucose |
| Regulated | *GAL1,10,7* | Galactose metabolic enzymes | +++ | 1000-fold induced by galactose |
| | *ADH2* | Alcohol dehydrogenase 2 | ++ | 100-fold repressed by glucose |
| | *PHO5* | Acid phosphatase | ++ | 200-fold repressed by phosphate |
| | *MET25* | O-acetyl homoserine sulphydrylase | + | |
| | *CUP1* | Copper metallothionein | + | 20-fold induced by Cu2+ |
| Heterologous | *CaMV* | Cauliflower mosaic virus 35S promoter | | |
| | *GRE* | Glucocorticoid response element | | |
| | *ARE* | Androgen response element | | |

Maximal expression depends also on efficient termination of transcription.
In vectors, the terminator sequence is usually from the same gene as the promoter.

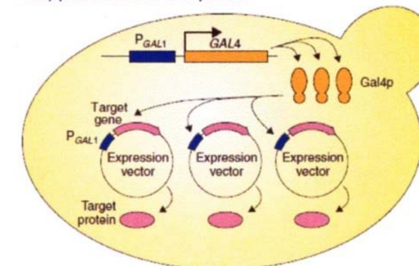
# The *GAL System*

- In yeasts, **the galactose is converted to glucose-6-phosphate by a series of the enzymes** (Leloir pathway). These enzymes are expressed at high level when cells are grown on galactose.
- The transctriptional activator Gal4p regulates expression of galactose catabolism (*GAL*) genes
- Each *GAL* gene contains within promoter one or multiple binding sites for the transcription activator Gal4p
- Binding of Gal4p to the promoter is regulated by the source of carbon
- When cells are grown on glucose, transcription from the GAL4 promoter is down-regulated, there is less Gal4p in the cells and reduced level of activator binding.

The protein production is initiated by switching the cells into a galactose-containing medium.

Synthetic promoters contain multiple Gal4p binding sites

*GAL4 gene* under the *GAL1* promoter control




Induction by galactose results in the production of Gal4p so that more of target gene may be expressed.

# The *CUP1 System*

Copper ions ($Cu^{2+}$ and $Cu^{+}$) are essential at appropriate level but toxic at high level for all living cells. Cells maintain a proper cellular level of copper ions.

At high Cu concentration in *S. cerevisiae,* cooper ion sensing transcription factor Acep1 binds DNA upsream of the *CUP1* gene, encoding methallothionein, and induces its transcription.

The transcription of *CUP1* is rapidly induces by addition of exogenous cooper to the medium (0.5 mM final concentration).

Advantage of *CUP1 system:*

Yeast culture containing *CUP1* expression plasmid can be grown on rich carbon source (glucose).

# Vectors for *S. cerevisiae* expression

Three basic types: Yeast episomal plasmid, YEp

Yeast integrative plasmids, YIp

Yeast centromere plasmid, YCp

# YEp Vector

- Yeast episomal plasmids YEp –a small circular plasmid.

**The vector replicates independently on chromosome via a single origin of replication.**

- The copy number of most YEp plasmids ranges from 10 to 40 copies per cell.

Selectable markers

*LEU2* gene as a selectable marker:

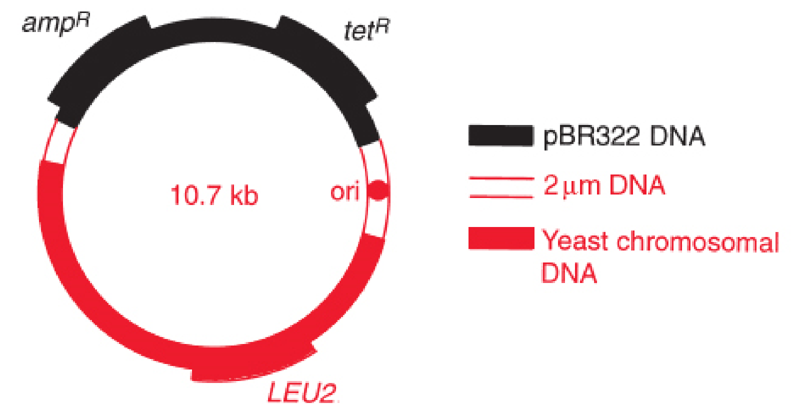
- Most YEp plasmids are relatively unstable and even under conditions of selective growth, only 60 to 95 percent of the cells retain the YEp plasmid.

Although this system is used for small scale expression studies, the use of YEp vectors in large-scale manufacturing is not advisable.

# YEp13 vector

- Example of yeast episomal plasmids

YEp13 is a shuttle vector,
It contains the 2μ origin of replication,
selectable *LEU* gene,
sequences from bacterial vector pBR322.




**This plasmid can be replicated in *E. coli,* this is advantage for recovering of DNA from transformed cells.**

# YIp Vector

Yeast integrating vectors YIp (selectable markers *URA3, LEU2)*
relies on integration into the host chromosome at the locus of the marker, at low frequencies by homologous recombination for survival and replication.

The plasmid does not carry an origin of replication, but contains selectable gene marker and yeast specific transcription and translation control sequences.
Its survival depends on intergration into yeast chromosomal DNA.

Typically, YIp vectors integrate as a single copy. However, methods to integrate multiple copies and stable cell lines with up to 15-20 copies of recombinant gene integrations have been developed for over-expressing specific genes.

# YCp Vectors.

YCp yeast centromere plasmid vectors

Autonomously replicating vectors containing centromere sequences (CEN), and autonomously replicating sequences (ARS).

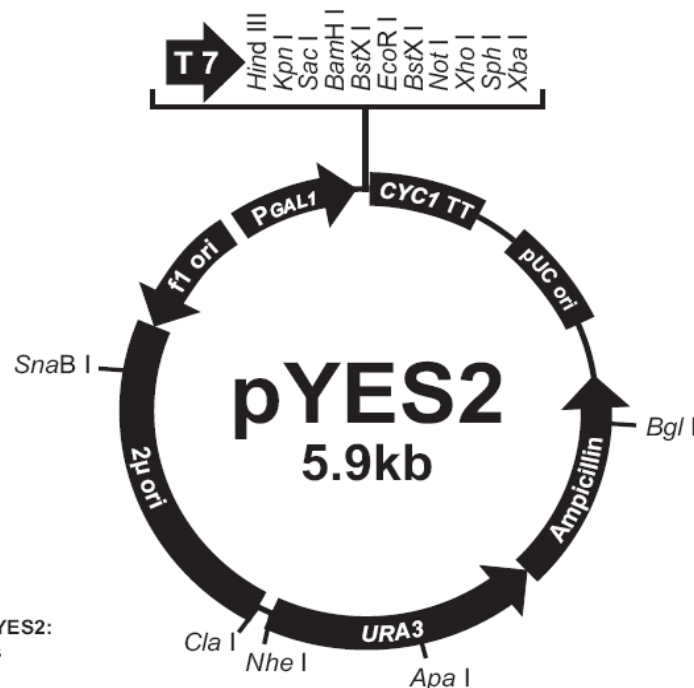
The YCp vectors are typically present at very low copy numbers from 1 to 3 per cell. These vectors are also relatively unstable and not very useful in high level expression but are used as regular cloning vectors (e.g., pYC2, pBM272).

# pYES2

pYES2 is a 5.9 kb vector designed for inducible expression of recombinant proteins in *Saccharomyces cerevisiae*. Features of the vectors allow easy cloning of the gene of interest and selection of transformants by uracil prototrophy.

**Map of pYES2**

The figure below summarizes the features of the pYES2 vector. **The sequence for pYES2 is available for downloading from our Web site (www.invitrogen.com) or from Technical Service (see page 17).**




**Comments for pYES2:**
**5856 nucleotides**

*GAL1* promoter: bases 1-451
T7 promoter/priming site: bases 475-494
Multiple cloning site: bases 501-600
*CYC1* transcription terminator: bases 608-856
pUC origin: bases 1038-1711
Ampicillin resistance gene: bases 1856-2716 (C)
*URA3* gene: bases 2734-3841 (C)
2 micron (μ) origin: bases 3845-5316
f1 origin: bases 5384-5839 (C)
(C) = complementary strand

Yeast *GAL1* promoter for high level inducible protein expression in yeast by galactose and repression by glucose.

A versatile multiple cloning site for simplified cloning.

*CYC1* transcriptional terminator for efficient termination of mRNA.

*URA3* gene for selection of transformants in yeast host strains with a *ura3* genotype.

Ampicillin resistance gene for selection in *E. coli.*

# pYES vectors

| Table 1 - YES™ Vector Collection | | | |
|---|---|---|---|
| Vector | Origin | Selection Marker | Tag |
| pYES2 | 2μ | *URA3* | – |
| pYES2/NT | 2μ | *URA3* | NT Xpress™ |
| pYES2/CT | 2μ | *URA3* | CT V5-6xHis |
| pYES3/CT | 2μ | *TRP1* | CT V5-6xHis |
| pYES6/CT | 2μ | Blasticidin | CT V5-6xHis |
| pYC2/NT | *CEN6/ARSH4* | *URA3* | NT Xpress™ |
| pYC2/CT | *CEN6/ARSH4* | *URA3* | CT V5-6xHis |
| pYC6/CT | *CEN6/ARSH4* | Blasticidin | CT V5-6xHis |

CT = C-Terminal    NT = N-Terminal

# Yeast Transformation

The introduction of DNA into yeasts:

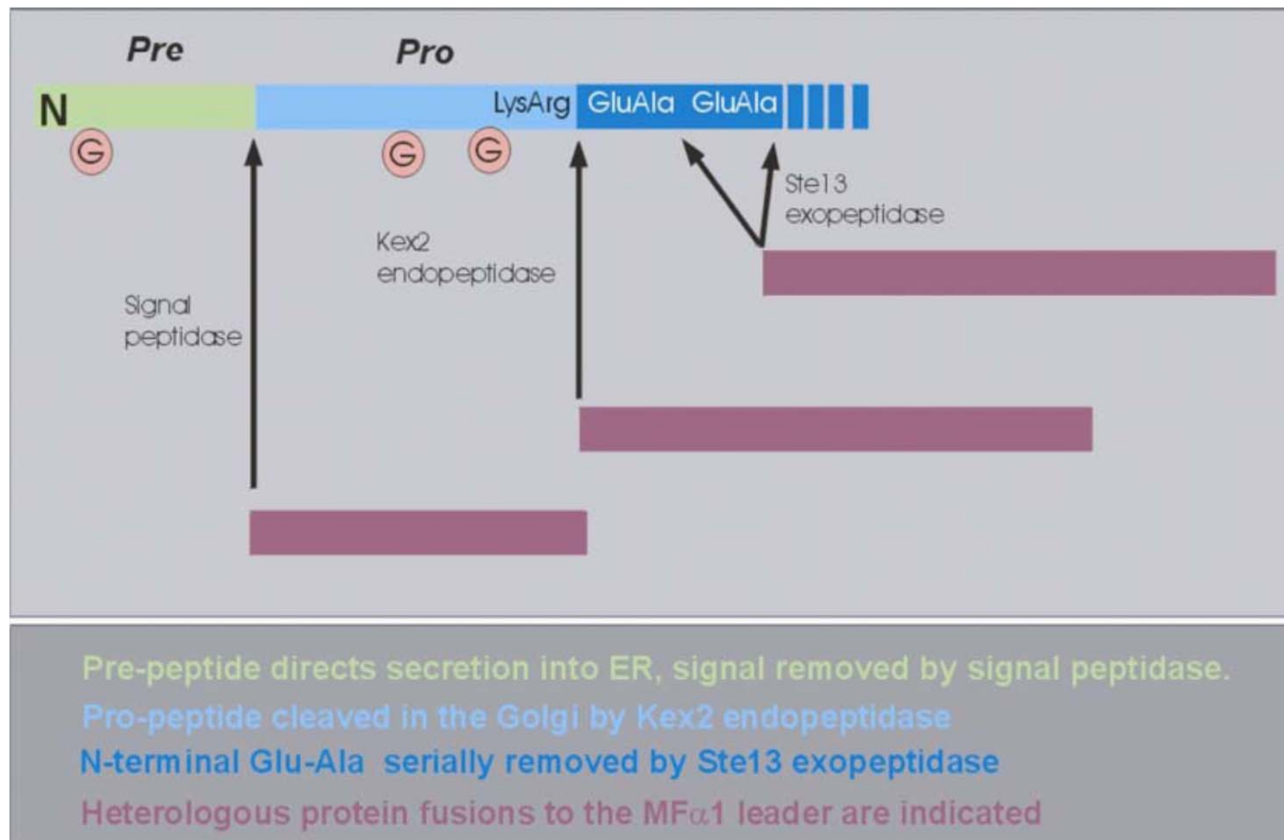
- Electroporation
- Lithium acetate treatment

For animal cells it is termed TRANSFECTION

# Secretion of Heterologous Proteins from Yeast

Gene must encode signal or leader sequence to pass through secretory system

Frequently used signal sequences in *S. cerevisiae* include those derived from invertase (*SUC2*), acid phosphatase (*PHO5*) or α-factor pheromone (MFα1);




Johnston, J.R.: Molecular genetics of yeast - a practical approach. Oxford University Press, Oxford, 1994.

# Expression in *Schizosaccharomyces pombe*

*S. Pombe* is a single-cell eukaryotic organism with many properties similar to higher-eukaryotic organisms: the chromosome structure, cell-cycle control, RNA spicing, codon usage[1].

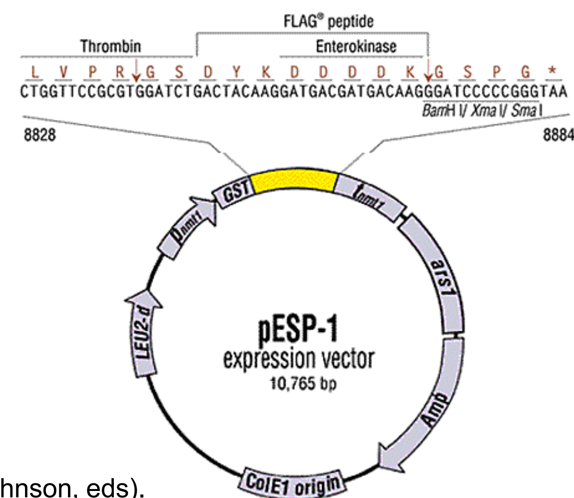
Proteins expressed in *S. pombe* are more likely to be folded properly.

Vectors (e.g. pESP-1 and pESP-2) use usually the *nmt1* promoter (no message in thiamine) for constitutive or induced expression of the gene of interest.

This promoter is active when the cells grow in the absence of thiamine. In the presence of thiamine greater than 0.5 μM thiamine, the promoter is turned off.

The main advantage of the ESP system over other eukaryotic expression systems is quick, easy expression and purification of recombinant proteins.

The *S. pombe* system offers high yield production with options for either inducible or constitutive expression.




http://www.biocompare.com

[1]Sipiczki, M. (1989) In Molecular Biology of Fission Yeast (A. Nisim,P. Young, and B.F. Johnson, eds). Academic Press, San Diego, Calif.

# Inducible Expression in *Pichia pastoris*

- *P. pastoris* is methylotrophic yeast, capable to metabolize methanol as a sole carbon source. *P. pastoris* has two alcohol oxidase genes *AOX1* and *AOX2*, which have a strongly inducible promoter. These enzyme allow to use the methanol as a source of carbon (oxidation of methanol to formaldehyde).
- The gene for the desired protein is present under the control of the *AOX1* promoter, the protein production can be induced by the addition of methanol.
- The expression vector is usually integrated into the genome as single or multiple copies.
- Cell are grown on glycerol (growth on glucose represses *AOX1* transcription even in the presence of methanol) to extremely high cell density prior to addition of methanol.
- *P. pastoris* media (glycerol, methanol, salts, trace elements, and biotin)
- *P. pastoris* has a broad spectrum of optimal growth pH (3.0-7.0).

# *Pichia pastoris* expression vector

*P. Pastoris* is convenient when trying to produce large quantities of protein without expensive equipment.

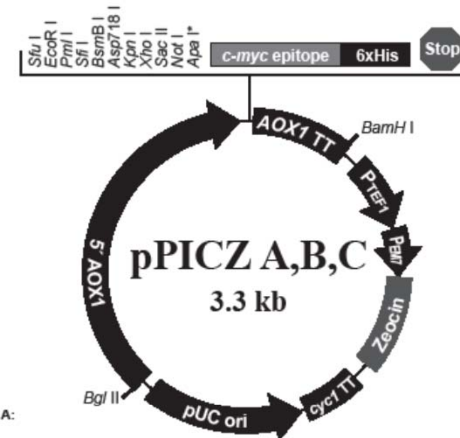
Highly efficient promoter

The preference of *P. pastoris* for respiratory growth is a key characteristic, which allows it to be cultured at high cell densities

**The protein can be secreted**

**The intracellular protein production can be also high.**

Methanol induction expression




**Comments for pPICZ A:**
3329 nucleotides

*5' AOX1* promoter region: bases 1-941
5' end of *AOX1* mRNA: base 824
5' *AOX1* priming site: bases 855-875
Multiple cloning site: bases 932-1011
*c-myc* epitope tag: bases 1012-1044
Polyhistidine tag: bases 1057-1077
3' AOX priming site: bases 1159-1179
3' end of mRNA: base 1250
*AOX1* transcription termination region: bases 1078-1418
Fragment containing TEF1 promoter: bases 1419-1830
EM7 promoter: bases 1831-1898
*Sh ble* ORF: bases 1899-2273
*CYC1* transcription termination region: bases 2274-2591
pUC origin: bases 2602-3275 (complementary strand)

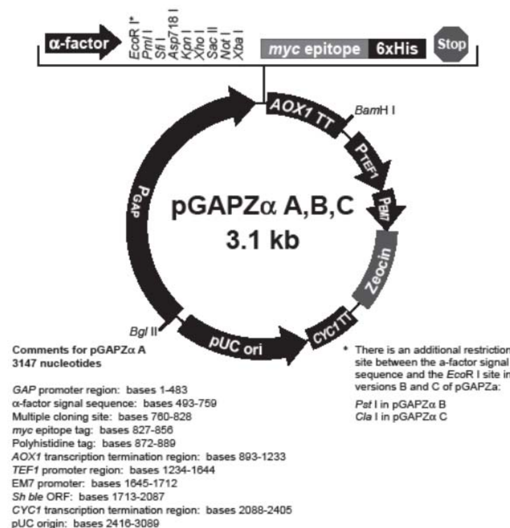
* The restriction site between *Not* I and the *myc* epitope is different in each version of pPICZ:

*Apa* I in pPICZ A
*Xba* I in pPICZ B
*Sna*B I in pPICZ C

Invitrogen™
life technologies

# Constutitive expression in *P. pastoris*

**pGAPZ** was created by replacing the methanol-regulated *AOX1* promoter with the **constitutive, glyceraldehyde-3-phosphate dehydrogenase *(GAP)* promoter** in the backbone of the pPICZ vectors.




Promoter:GAP Vector Type:pGAP

Protein Tag or Fusion:c-Myc Epitope Tag, His Tag (6x His)

Antibiotic Resistance (Bacterial):Zeocin™ (ZeoR)

Expression can be done in either YPD medium (1% yeast extract, 2% peptone,2% glucose) or Yeast Nitrogen Base with 0.5% glucose.

*GAP* promoter is a constitutively expressed promoter.
If the gene product is toxic to the cells, you may have to try an inducible expression system.

http://tools.invitrogen.com/content/sfs/brochures/B-067202_Pichia_Flyer.pdf

# Yeast Expression Systems

Table 1 - Yeast Expression Systems

| System/ Vector | Host | Secretion Signal | Fusion Partner Position | Fusion Partner Purif. | Fusion Partner Epitope | Selectable Marker | Promoter | Inducer | Advantage |
|---|---|---|---|---|---|---|---|---|---|
| *Pichia* Expression System | *P. pastoris* | α-factor or *PHO* | C-term. | 6xHis | *c-myc* | *HIS4*, Zeocin™, or Blasticidin | *AOX1* or *GAP* | methanol (*GAP* is constitutive) | High-level expression, copy number control, suitable for industrial-scale protein production |
| | *P. methanolica* | α-factor | C-term. | 6xHis | V5 | *ADE2* | *AUG1* | methanol | High-level expression |
| YES™ Vector Collection | *S. cerevisiae* | — | N-term. C-term. | 6xHis | Xpress™ V5 | *URA3* or blasticidin | *GAL1* | galactose | High- or low-copy episomal expression |
| pTEF1/Zeo pTEF1/Bsd | *S. cerevisiae* *P. pastoris* *P. methanolica* | — | — | — | — | Zeocin™ *or* Blasticidin | *TEF1* | — | Simplified construction of Zeocin™- or Blasticidin-resistant vectors in yeast |
| SpECTRA™ *S. pombe* Expression System | *S. pombe* | — | C-term. | 6xHis | V5 | *LEU2* | *nmt1* *nmt41* *nmt81* | thiamine | Flexible control of expression levels in *S. pombe* |
| pYD1 | *S. cerevisiae* | *AGA2* | N-term. C-term. | 6xHis | Xpress™ V5 | *TRP1* | *GAL1* | galactose | Displays protein on cell surface |

# Expression in Insect cells

1. Expression using baculoviruses: insect cells are infected with recombinant baculovirus bearing a host gene

2. Stable transformation of insect cells using selective marker, long term expression, convenient for expression of receptors, membrane and glycoproteins

3. Transient expression without selection, convenient for immunodetection, functional studies

# Expression using Baculoviruses

- High yield of glycosylated and processed proteins
- Cells grow in suspension: convenient for up-scaling
- Enables expression of viral proteins and oncoproteins
- Expression of proteins occur in the very late stage of infection

# Baculoviruses

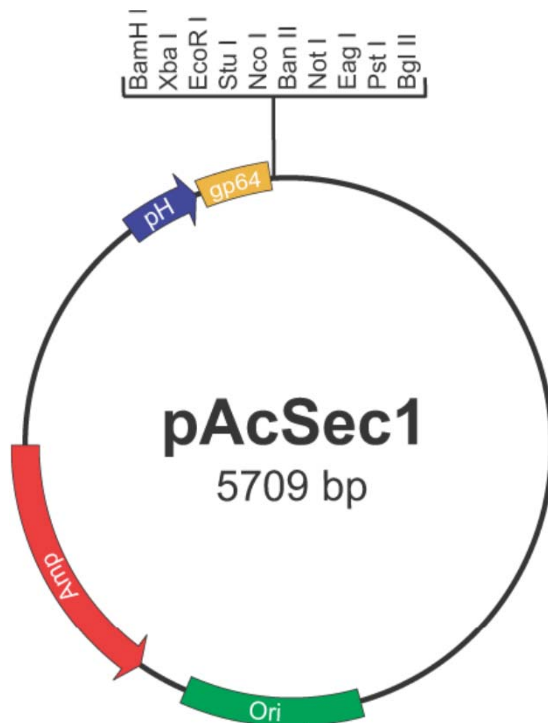
- *Baculoviridae Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) – rod-shape virus.
- It contains double-stranded DNA.

- This DNA could be easily destroyed by exposure to sunlight or by conditions in the host's gut, an infective baculovirus particle (*virion*) is protected by protein coat called a *polyhedrin (p10),* which is highly expressed due to very strong promoters and is released into the environment
- Polyhedrins formed in these cells are dispensable for infectious virus formation.
- Deletion of these genes and replacement with foreign genes convert these viruses into expression vectors for the production of recombinant proteins.

# Transfer vector

*E. coli* based plasmid with AcMNPV DNA (baculovirus)– polyhedrin promoter/terminator + flanking sequences) -> gene of interest cloned downstream of promoter

Plasmid transfer vector contains *POLH* promoter




The pAcSec1 vector contains the signal sequence of acidic glycoprotein gp64, which is the most abundant surface glycoprotein of multiple nuclear polyhedrosis virus AcNPV.

The gp64 signal sequence is one of the most effective leader sequences known in baculovirus.
Polyhedrin promoter (***POLH)*** can be used to drive target gene expression in insect cell lines derived from the moth *Spodoptera frugiperda* (cell line Sf9, Sf21).

When the expressed protein is transported across the cell membrane, the signal peptide is cleaved and the native protein can be easily purified from the culture medium.

# Production of Recombinant viruses

- The transfer plasmid is co-transfected with linearized baculovirus DNA into insect cells and the recombinant virus is produced (target gene is introduced into the baculovirus genome by homologous recombination).
- The recombination process results in the repair of the circular viral DNA and allows viral replication to proceed.
- Recombinant viral formation can be observed microscopically by viewing viral plaques in the insect cells.
- Often the transfer vector contains the *lacZ* gene or other reporter gene, which allows the visual identification of plaques by blue apparence after staining with X-Gal.
- The recombinant virus is isolated using plaque purification
- High titer virus is prepared and used to infect large-scale insect culture for protein production.
- Protein harvesting 4-5 days post-infection

# Expression of Proteins in Insect Cells

Advantage of protein production using baculovirus system:

Production of high level of proteins

Production of glycosylated proteins

Production of protein with other posttranslation modification: myristoylation, palmitoylation, phosporylation

Maintaining of  signal sequences for transport in nucleus, membrane.

Multiple genes can be expressed from a single virus. This allows production of protein complexes.

Disadvantage: Production of recombinant baculovirus can take as long as 4 – 6 weeks and the insect Sf cells grow slowly, media are expensive.