Eukaryotic Gene Expression: Basics & Benefits

P N RANGARAJAN

Lecture 30

Eukaryotic protein expression systems - I

Eukaryotic protein expression systems-I (lecture 30)

Protein expression in yeast and insect cells

Eukaryotic protein expression systems-II (lecture 31)

Protein expression in mammalian cells

Cell-free protein expression systems

Eukaryotic protein expression systems-III (lecture 32)

Production of recombinant proteins in plants and farm animals

---

Human gene therapy (lecture 33)

Biopharmaceuticals

Blood factors (Factor VIII and Factor IX)

Thrombolytic agents (tissue plasminogen activator)

Hormones (insulin, glucagon, growth hormone, gonadotrophins)

Haematopoietic growth factors (Erythropoietin, colony stimulating factors)

Interferons (Interferons-α, -β, -γ)

Interleukin-based products (Interleukin-2)

Vaccines (Hepatitis B virus surface antigen, Human papilloma virus surface antigens)

Monoclonal antibodies (Various)

Additional products (tumour necrosis factor, therapeutic enzymes)

Recombinant DNA therapeutics sector represents the core of the human medical biotechnology industry, worth over $32 billion in 2003.

The rDNA therapeutics sector :

- >110 companies
- >80 therapeutics in clinical development
- ~73 marketed products

Top ten recombinant therapeutic proteins and their global sales in 2003

Product /marketing company	2003 ($million)
Procrit (epoetin alfa)/Johnson & Johnson	3,986
Epogen (epoetin alfa)/Amgen	2,435
Neupogen (filgrastim)/Amgen	1,268
Neulasta (pegfilgrastim)/Amgen	1,255
Novolin (insulin systemic)/Novo Nordisk	2,235
Avonex (interferon beta-1a)/Biogen IDEC	1,170
PEG-Intron A franchise (pegylated interferon alpha)/Schering Plough	1,851
Enbrel (etanercept)/Amgen	1,300
Aranesp (darbepoetin alfa)/Amgen	1,544
NeoRecormon (epoetin-beta)/Roche	1,318

The recombinant protein therapeutics market is valued at $52,150 million in 2010.

Adapted from: Nature Biotechnology 22, 1513 - 1519 (2004)

Some Commercialized Recombinant Biologicals and their Expression systems

Insulin	E.coli	Blood coagulation Factors VII, VIII,IX	BHK cells
Interferon alpha, beta, gamma	E.coli		CHO cells
Interleukin 2	E.coli	Erythropoetin	CHO cells
Plasminogen activator	E.coli	FSH	CHO cells
Tumor Necrosis factor	E.coli	LH	CHO cells
Growth Hormone	E.coli	Goinadotropin	CHO cells
Calcitonin	E.coli	Interferon beta	CHO cells
		Tissue Plasminogen activator	CHO cells
Glucagon	S.cerevisiae	Glucocerebrocidase	CHO cells
Eutropin (hGH deriv)	S.cerevisiae	Interleukin 11-agonist	ROMI 8866
Platelet Derived Growth Factor	S.cerevisiae		(human cell line)
Hepatitis B Vaccine	S.cerevisiae		
Hepatitis B Vaccine	P.pastoris		
Insulin	P.pastoris		
Streptokinase	P.pastoris		

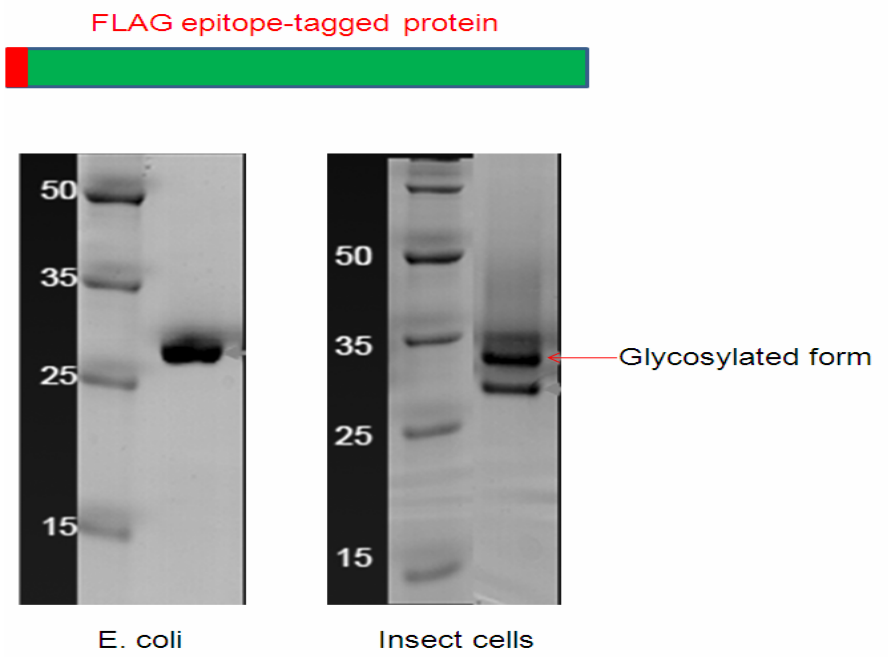
Bacterial expression systems

Advantages

- Grow quickly (8-12 hrs to produce protein)
- High yields (50-500 mg/L)
- Low cost of media
- Low fermentor costs

Disadvantages

- Difficulty expressing large proteins (>50 kD)
- Eukaryotic proteins are sometimes toxic
- Can't handle disulphide-bonded proteins
- No glycosylation or signal peptide removal



Yeast expression systems

Advantages

- Grow quickly (12-24 hrs to produce protein)
- Very high yields (50-5000 mg/L)
- Low cost of media (simple media constituents)
- Low fermentor costs
- Can express large proteins (>50 kD)
- Glycosylation & signal peptide removal
- Has chaperonins to help fold “tough” prtns
- Can handle S-S rich proteins

Baculovirus Systems

Advantages

- Can express large proteins (>50 kD)
- Correct glycosylation & signal peptide removal
- Has chaperonins to help fold “tough” proteins
- Very high yields, cheap

Disadvantages

- Grow very slowly (10-12 days for set-up)
- Cell culture is only sustainable for 4-5 days
- Set-up is time consuming, not as simple as yeast

Mammalian Systems

Advantages

- Can express large proteins (>50 kD)
- Correct glycosylation & signal peptide removal, generates authentic proteins
- Has chaperonins to help fold “tough” proteins

Disadvantages

- Selection takes time (weeks for set-up)
- Cell culture is only sustainable for limited period of time
- Set-up is very time consuming, costly, modest yields

Prokaryotic systems are generally cheaper, but…

Eukaryotic proteins produced in bacteria may be unstable or lack biological activity due to lack of posttranslational modifications or correct assembly

Possess unacceptable contaminants after purification

Yeast expression systems

Making recombinant proteins in yeast cells

Saccharomyces cerevisiae

Pichia pastoris

Saccharomyces cerevisiae

A single cell

Well characterized genetically and physiologically

Can be readily grown in both small vessels and large scale bioreactors

Several strong promoters have been isolated and characterized

Carry out many post-translational modifications (phosphorylation, glycosylation and targeting)

Readily grown in small and large scale bioreactors

secretes few proteins, the product can easily be purified generally recognized as safe (GRAS)

Selection Markers commonly used in yeast vectors

- *ARG4*
- *HIS4*
- *LEU2*
- *TRP1*
- *URA3*

S. cerevisiae expression vectors

Integrative vectors (YIp)

Autonomously replicating high copy-number vectors (YEp)

Autonomously replicating low copy-number vectors (YCp)

Yeast artificial chromosomes (YACs)

YIp Vectors

The YIp integrative vectors do not replicate autonomously, but integrate into the genome at low frequencies by homologous recombination.

The YIp vectors typically integrate as a single copy.

Strains transformed with YIp plasmids are extremely stable, even in the absence of selective pressure.

YEp Vectors

The YEp yeast episomal plasmid vectors replicate autonomously because of the presence of a segment of the yeast 2 μ plasmid that serves as an origin of replication (2 μ *ori*).

The 2 μ *ori* is responsible for the high copy-number and high frequency of transformation of YEp vectors.

Most YEp plasmids are relatively unstable, being lost in approximately 10-2 or more cells after each generation.

Even under conditions of selective growth, only 60% to 95% of the cells retain the YEp plasmid.

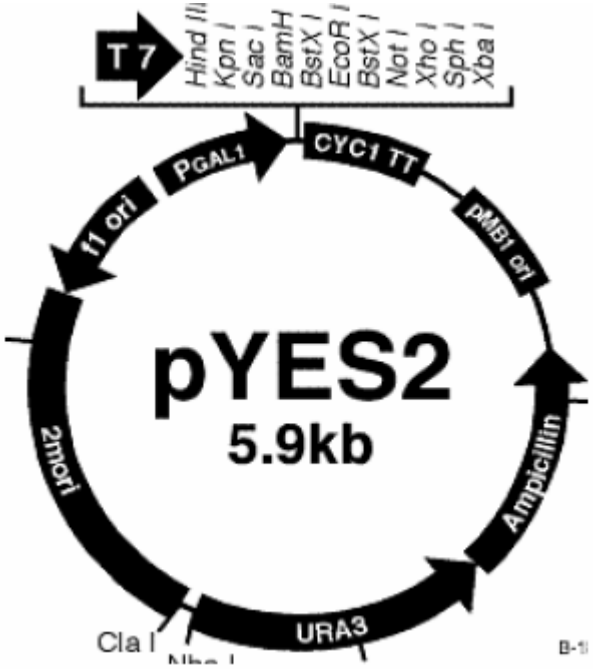
YCp Vectors

The YCp yeast centromere plasmid vectors are 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, and possibly more, and are lost in approximately 10^{-2} cells per generation without selective pressure.

YRp vectors, containing *ARS* but lacking functional *CEN* elements, transform yeast at high frequencies, but are lost at too high a frequency, over 10% per generation, making them undesirable for general vectors.

Commonly used commercial yeast episomal vectors
For protein expression in *S. cerevisiae*



PGal1	galactose-inducible promoter (yeast)
cyc1 TT	transcription terminator
pMB1 ori	*E. coli* origin of replication (pUC)
Amp^r	*E. coli* selectable marker
URA3	yeast selectable marker (ura3 host)
2μ ori	yeast origin of replication
f1 ori	ssDNA origin of replication
T7	phage promoter (in vitro transcription)

A galactose-inducible S. cerevisiae expression vector

Available form Invitrogen

pESC vectors available from Agilent

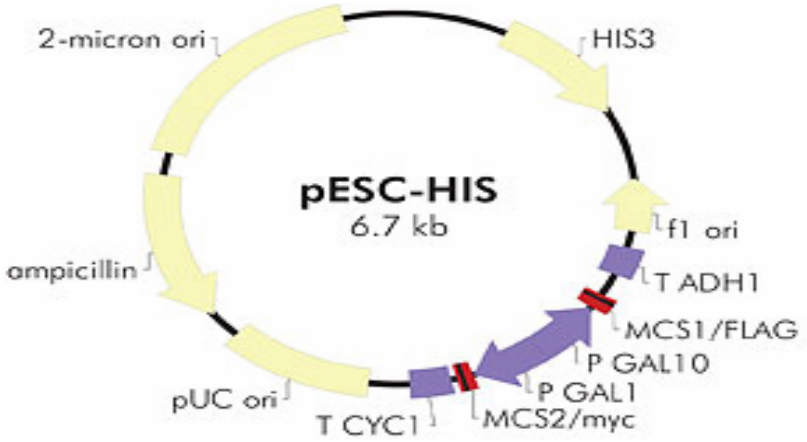
The pESC vectors are a series of epitope-tagged vectors designed for expression and functional analysis of eukaryotic genes in the yeast *S. cerevisiae*.

These vectors contain the *GAL1* and *GAL10* yeast promoters in opposing orientation.

With these vectors one or two cloned genes can be introduced into a yeast host strain under the control of a repressible promoter.

pESC-HIS
pESC-LEU
pESC-TRP
pESC-URA

Each of these pESC vectors contains one of four different yeast-selectable markers (*HIS3, TRP1, LEU2,* or *URA3*) in the same vector backbone, which allows expression and epitope-tagging analysis of two different genes in a single yeast cell



pESC-HIS Multiple Cloning Site 1 Region
(sequence shown 2255–2338, bottom strand)

EcoR I Not I Spe I Cla I Sac I Pac I

GAA TTC AAC GCT CAC TAA AGG GCG GCC GCA CTA GTA TCG ATG GAT TAC AAG GAT GAC GAC GAT AAG ATC TGA GCTCTTAATTAA

M D Y K D D D D K I STOP

FLAG epitope

pESC-HIS Multiple Cloning Site 2 Region
(sequence shown 3011–3108, top strand)

BamH I Apa I Sfi I Sal I

G GAT CCG TAA TAC GAC TCA CTA TAG GGC CCG GGC GTC GAC...

(STOP)

Xho I

...ATG GAA CAG AAG TTG ATT TCC GAA GAA GAC CTC GAG TAA GCTTGGTACCGCGGCTAGC

M E Q K L I S E E D L E STOP

myc epitope

Feature	Nucleotide Position
yeast *HIS3* selection marker ORF	504–1163
f1 origin of ss-DNA replication	1558–1864
yeast *ADH1* terminator	1960–2124
multiple cloning site 1	2255–2338
FLAG tag	2273–2299
yeast *GAL1/GAL10* divergent promoters	2343–3009
multiple cloning site 2	3011–3086
c-myc tag	3051–3086
yeast *CYC1* terminator	3113–3302
pUC origin of replication	3939–4156
ampicillin resistance (*bla*) ORF	4307–5164
2μ yeast origin of replication	5298–6453

http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductData&PageID=596

Promoters for *S. cerevisiae* expression vectors

Promoter	Expression conditions	Status
Acid phosphatase (*PH05*)	Phosphate-deficient medium	Inducible
Alcohol dehydrogenase I (*ADHI*)	2–5% Glucose	Constitutive
Alcohol dehydrogenase II (*ADHII*)	0.1–0.2% Glucose	Inducible
Cytochrome c_1 (*CYC1*)	Glucose	Repressible
Gal-1-P Glc-1-P uridyltransferase	Galactose	Inducible
Galactokinase (*GAL1*)	Galactose	Inducible
Glyceraldehyde-3-phosphate dehydrogenase (*GAPD*, *GAPDH*)	2–5% Glucose	Constitutive
Metallothionein (*CUP1*)	0.03–0.1 mM copper	Inducible
Phosphoglycerate kinase (*PGK*)	2–5% Glucose	Constitutive
Triose phosphate isomerase (*TPI*)	2–5% Glucose	Constitutive
UDP galactose epimerase (*GAL10*)	Galactose	Inducible

Recombinant proteins produced in *S. cerevisiae*

VACCINES
- Hepatitis B virus surface antigen
- Malaria circumsporozoite protein
- HIV-1 envelope protein

DIAGNOSTICS
- Hepatitis C virus protein
- HIV-1 antigens

HUMAN THERAPEUTIC AGENTS
- Epidermal growth factor
- Insulin
- Insulin-like growth factor
- Platelet-derived growth factor
- Proinsulin
- Fibroblast growth factor
- Granulocyte-macrophage colony-stimulating factor
- α_1 antitrypsin
- Blood coagulation factor XIIIa

Recombinant protein production in methylotrophic yeasts

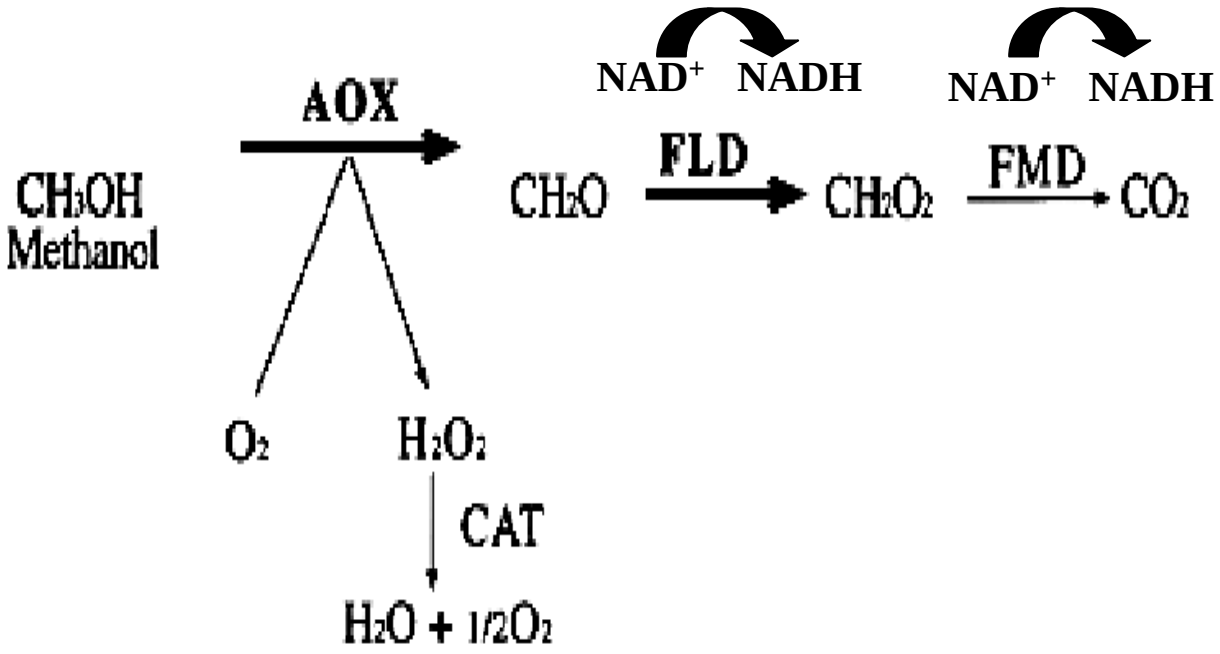
Four known genera: *Pichia*, *Hansenula*, *Candida* and *Torulopsis*

Can grow on methanol, as a sole carbon and energy source

Harbor a highly efficient and regulated metabolic pathway for methanol, an otherwise toxic compound

The genes coding for methanol utilization pathway are under glucose repression and are induced to the maximum level by the cognate substrate i.e. methanol

Metabolism of methanol in methylotrophic yeasts.



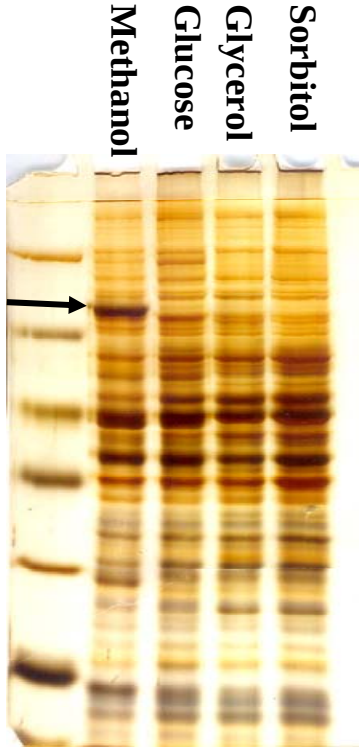
AOX, alcohol oxidase; CAT, catalase; FLD, formaldehyde dehydrogenase; FMD, formate dehydrogenase.

Johnson *et al*., Genetics, Vol. 151, 1379-1391, April 1999.

Key features of Pichia pastoris

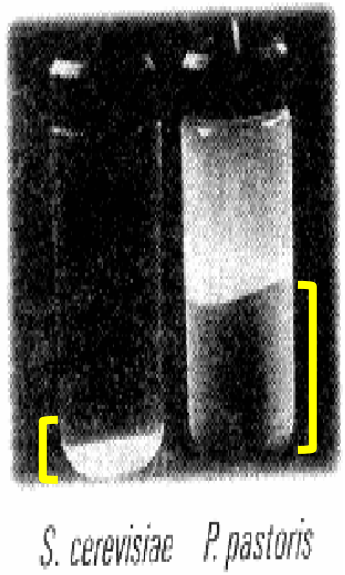
- **Methylotrophic budding yeast with a proven record for recombinant protein production.**
- **In the absence of glucose or glycerol, it can utilize methanol as the sole carbon source.**
- **The alcohol oxidase promoter controls the expression of alcohol oxidase, which catalyzes the first step in methanol metabolism.**
- **Typically 30% of total soluble protein in methanol-induced cells is alcohol oxidase.**
- **The cell density of *Pichia pastoris* can be 10 times greater than that of *Saccharomyces cerevisiae*.**

A



High levels of induction

B

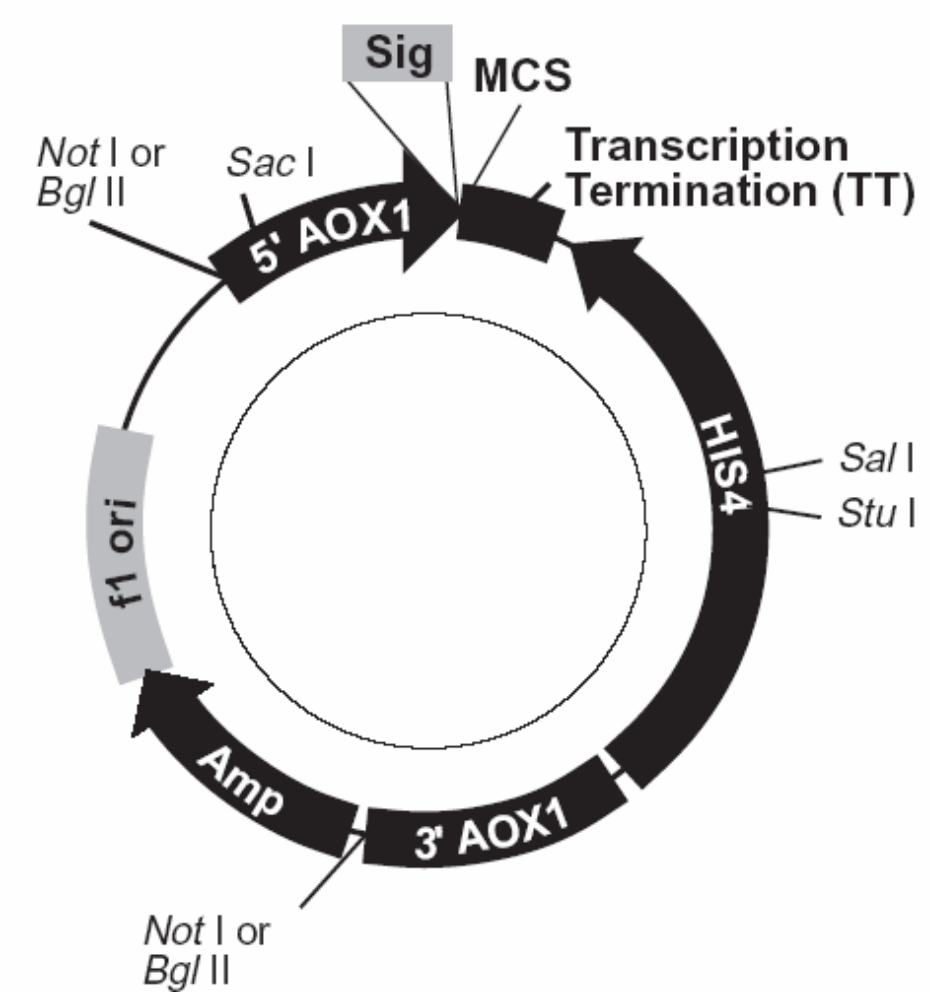


High Biomass

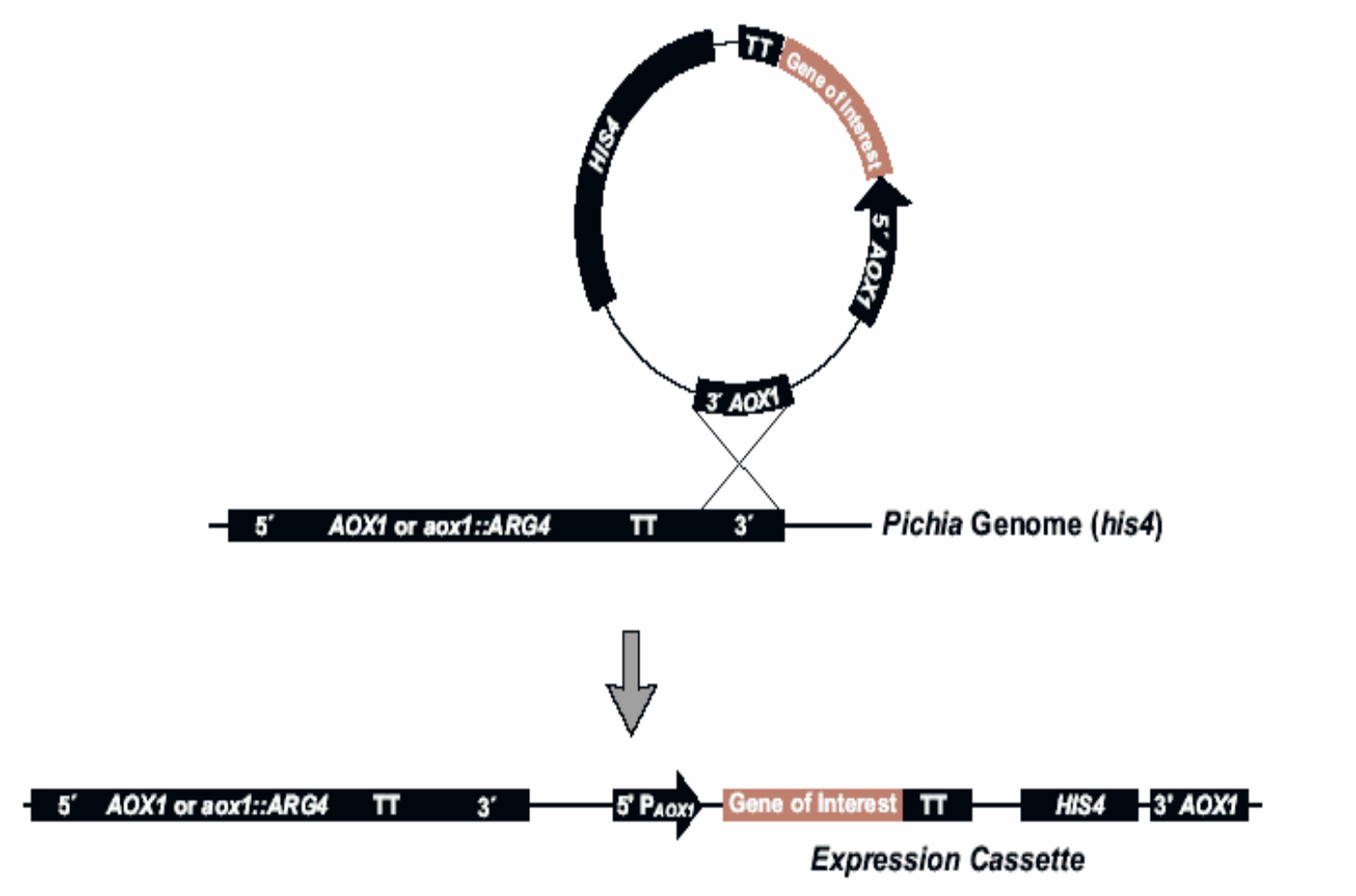
Unique features of *Pichia pastoris*.

A. High level expression of Alcohol oxidase enzyme (arrow) in methanol grown cells.

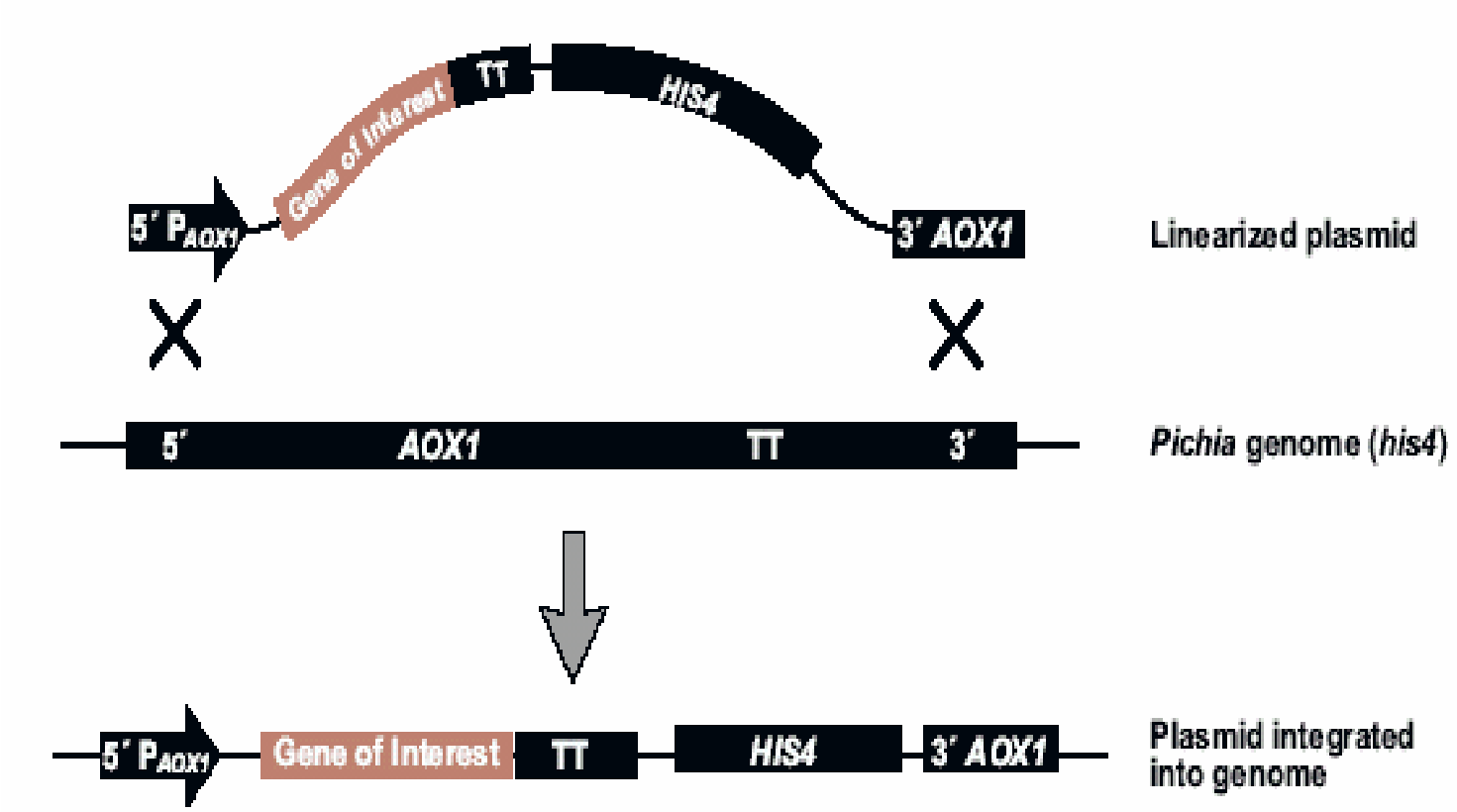
B. *P. pastoris* can grow to very high cell densities compared to *S. cerevisiae.*



A generic methanol-inducible P. pastoris expression vector



Genomic integration of the gene of interest by homologous recombination



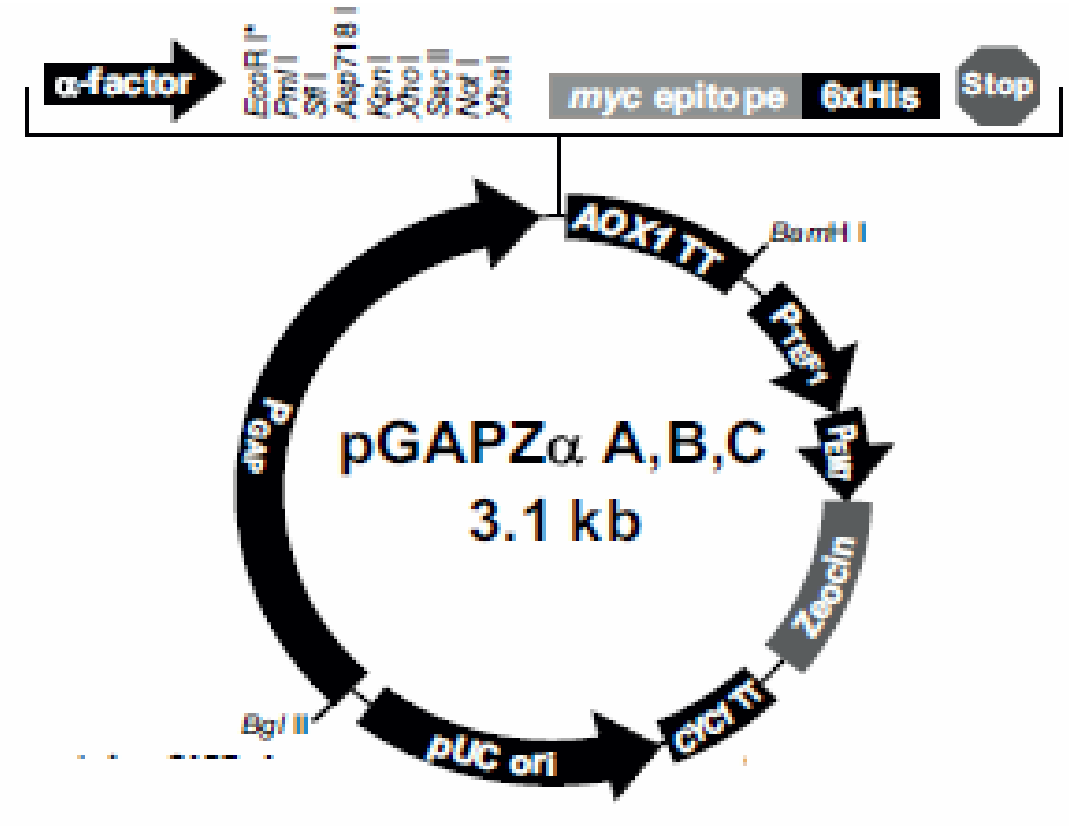
Genomic integration of the expression cassette involving double cross over event

pGAP vectors

- Strong constitutive promoter
- High transcription on D-glucose,
- Moderate transcription on Glycerol
- Low transcription on Methanol

pGAPZ A, B, and C
pGAPZα A, B, and C

Pichia expression vectors for constitutive expression and purification of recombinant proteins



www.invitrogen.com

Host strains routinely used for recombinant protein production in *Pichia pastoris*

Strain	Genotype	Application
GS115	*his4*	Selection of expression vectors containing *HIS4*
X-33	wild-type	Selection of Zeocin™-resistant expression vectors
KM71	*his4, aox1::ARG4, arg4*	Selection of expression vectors containing *HIS4* to generate strains with Mut^S phenotype
KM71H	*aox1::ARG4, arg4*	Selection of Zeocin™-resistant expression vectors to generate strains with Mut^S phenotype
SMD1168	*his4, pep4*	Selection of expression vectors containing *HIS4* to generate strains without protease A activity
SMD1168H	*pep4*	Selection of Zeocin™-resistant expression vectors to generate strains without protease A activity

Recombinant proteins produced in *P. pastoris*

Protein expressed	Expression Level (mg/L)
Bacterial proteins	
Tetanus toxin fragment C	12,000
α-amylase	2,500
T2A peroxidase	2,470
C. botulinum neurotoxin fragment	78
Yeast proteins	
Catalase L	2,300
Glucoamylase	400
Lipase	60
Plant proteins	
Hydroxynitrile lyase	22,000
Wheat lipid transfer protein	720
Aeroallergen	60
Invertebrate proteins	
Hirudin	1,500
Spider dragline silk protein	663
Honeybee olfactory protein	200
Mammalian proteins	
Mouse gelatin	14,800
Porcine carboxypeptidase B	200
Human tumor necrosis factor	10,000
Human IGF-1	600
Human CD38	455
15N-Interferon τ	10

RECOMBINANT HEPATITIS B VACCINE

Biological E Limited, Hyderabad (22.12.2004)

Indian Immunologicals Limited, Hyderabad (24.9.2006)

Development of glycoengineered yeast strains

Hamilton SR, Gerngross TU. Curr Opin Biotechnol. 2007 Oct;18(5):387-92.
Glycosylation engineering in yeast: the advent of fully humanized yeast..

http://www.glycofi.com/

They have generated a library of yeast strains that have been engineered to perform specific human glycosylation at high fidelity.

GlycoFi owns or controls over 60 issued and pending patents in the U.S. and abroad relating to glycosylation engineering and the production of human glycoproteins with proper glycosylation in yeast and fungi.

Making recombinant proteins in insect cells

What is Baculovirus?

Baculoviruses are enveloped, double- stranded DNA (circular, supercoiled) viruses with rodshaped nucleocapsids

Baculovirus life cycle involves two distinct forms of viruses, Budded Virus [BV] and Occluded Virus [OV]

BV consists of a single nucleocapsid enveloped by GP64, a virus derived lycoprotein, and host membrane proteins

OV consists of multiple nucleocapsids embedded in a protein matrix (polyhedrin matrix)

The most extensively studied baculovirus strain *is Autographa californica multiple nuclear* polyhedrosis virus (AcMNPV).

AcMNPV only infects larval lepidopterans

Baculovirus life cycle

- Early Phase (0-6 h PI)
- Virus enters cells by endocytosis
- Nucleocapsids migrate to nucleus
- Viral DNA is released
- Early gene expression starts

- Late Phase (6-24 h PI)
- Extensive DNA replication
- Progeny nucleocapsids leave nucleus and acquire envelope as they leave cytoplasm
- Production of budded virus

- Very Late Phase
- Decrease in the formation of budded virus
- Nucleocapsids acquire envelopes inside nucleus to form MNPVs
- MNPVs are embedded in a matrix made predominantly of the polyhedrin protein and form occlusion bodies

Baculovirus Expression Vector System (BEVS)

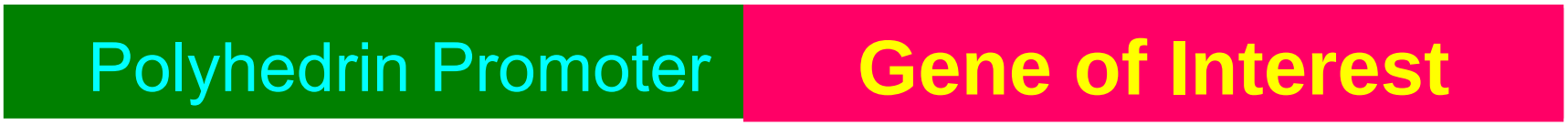
BEVS was pioneered by Dr. Max D. Summers, and Dr. Gale Smith in 1982

BEVS is based on replacement of a very late, non-essential, viral gene (*polyhedrin), with a gene of interest*

Most of the transfer vectors use either early (*Ie1) or very late (p10, pPolyh) promoters*

Modified and linearized AcMNPV DNA revolutionized the BEVS

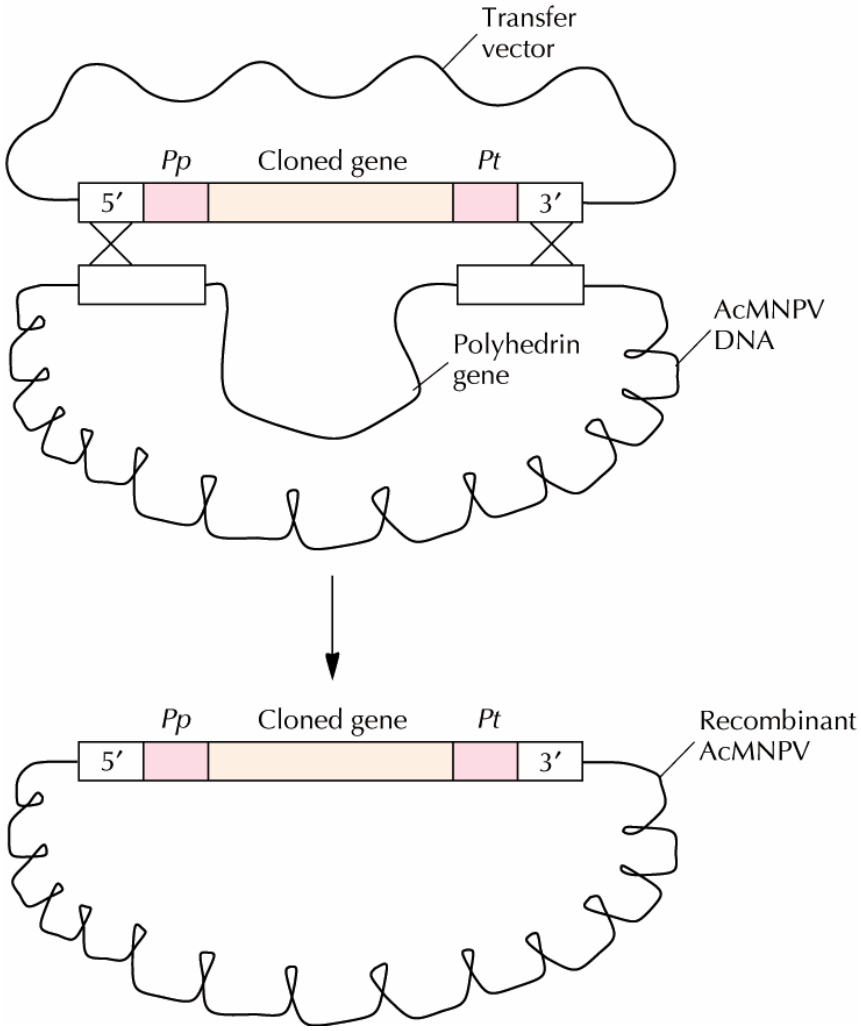
BEVS allows rapid cloning and expression of recombinant proteins in insect cells (Sf9, Sf21, Hi5)



Foreign gene cloned into a transfer vector based on *E. coli* plasmid that carries a segment of the DNA from AcNPV

Co-transfected along with ds-baculovirus DNA into insect cells

Homologous recombination of the transfer vector with insert DNA with viral genome leads to the cloned gene being transferred into the AcNPV DNA.



Baculovirus (AcMNPV) Cloning Process

Baculovirus –based protein expression

Higher level of gene expression (up to 50% of total cellular protein), in most cases, soluble and functionally active

Permits post-translational modifications

Disulphide bonds and proper folding

N-and O-linked glycosylation

Signal peptide cleavage

Easy to scale-up, insect cells are simple to maintain as suspension culture compared to mammalian cells

Inexpensive compared to other eukaryotic expression systems

A number of baculovirus expression systems are commercially available

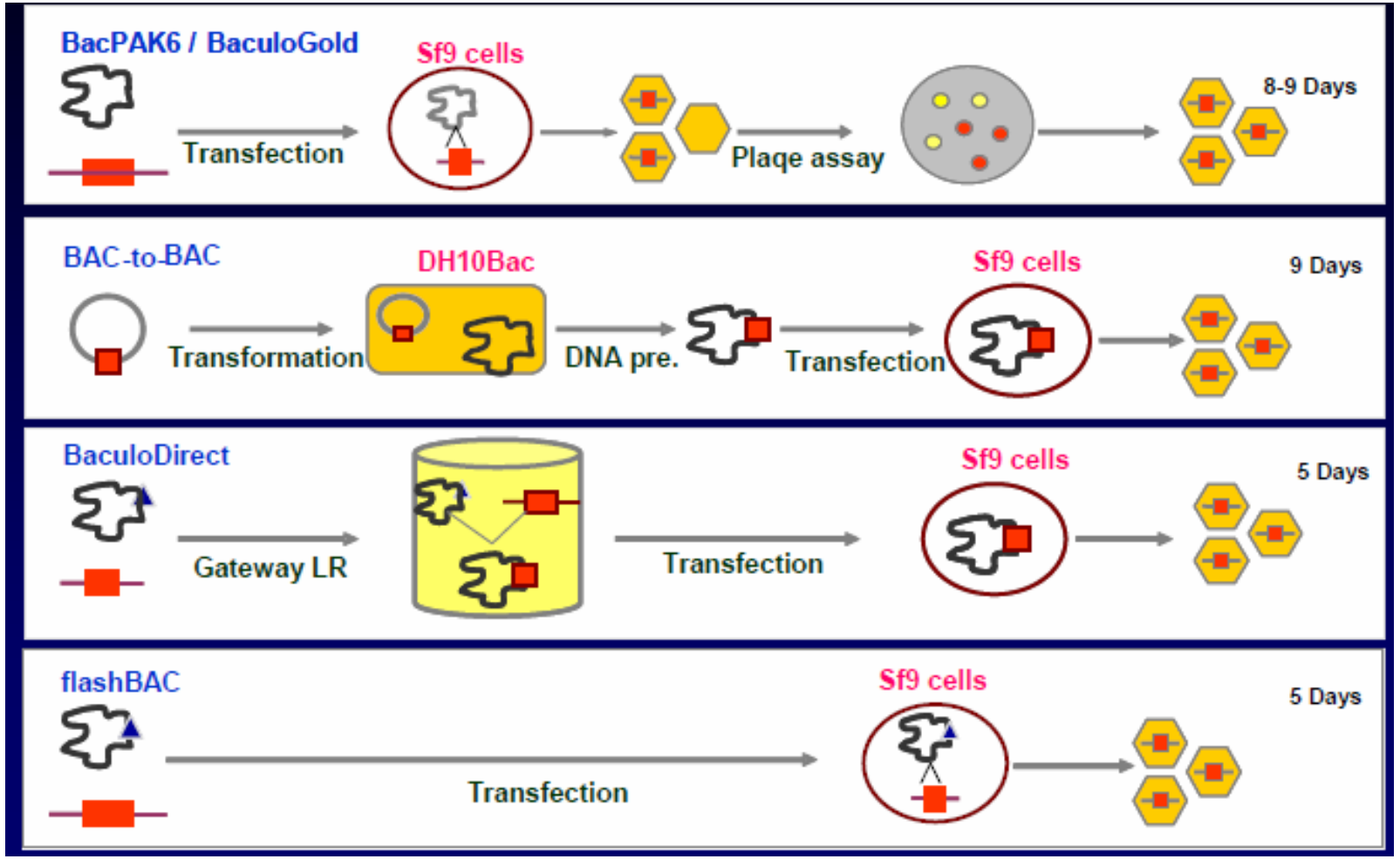
Bac-to-Bac (Invitrogen™)

BacPAK6/BaculoGold (BD Biosciences/Clonetech)

BaculoDirect™ (Invitrogen™)

flashBAC™/BacMagic (EMD/OET/Nextgen)

Comparison of Baculovirus Expression Systems



Commonly used insect cells for protein production

Cell line	Origin	Use
Sf-9	*Spodoptera frugiperda* (Pupal ovarian tissue)	Recombinant baculovirus production Intracellular protein expression Plaque assay
Sf-21	*Spodoptera frugiperda* (Pupal ovarian tissue)	Intracellular protein production Secretion of recombinant protein
Hi-5	*Trichoplusia ni* (Ovarian cells)	Secretion of recombinant protein
Tri-Ex	Sf-9 derivative	Intracellular protein production

Recombinant proteins produced in BEVS

- α-Interferon
- Adenosine deaminase
- Anthrax antigen
- β-Amyloid precursor protein
- β-Interferon
- Bovine rhodopsin
- Bluetongue virus neutralization antigen
- Cystic fibrosis transmembrane conductance regulator
- Dengue virus type 1 antigen
- Erythropoietin
- G-protein-coupled receptors
- HIV-1 envelope protein
- HSV capsid proteins
- Human alkaline phosphatase
- Human DNA polymerase α
- Human pancreatic lipase
- Influenza virus hemagglutinin
- Interleukin-2
- Lassa virus protein
- Malaria proteins
- Mouse monoclonal antibodies
- Multidrug transporter protein
- Poliovirus proteins
- Pseudorabies virus glycoprotein 50
- Rabies virus glycoprotein
- Respiratory syncytial virus antigen
- Simian rotavirus capsid antigen
- Tissue plasminogen activator

Eukaryotic expression costs ca. 5-10 times that of E.coli !

Expression System Selection:

Choice depends on size and character of protein

•Large proteins (>100 kDa)?	Choose eukaryote
•Small proteins (<30 kDa)?	Choose *E.coli*
•Glycosylation essential?	Choose baculovirus or mammalian host
•Post-translational modifications essential?	Choose yeast, baculovirus or mammalian host

Reichert, J.M. & Paquette, C. Therapeutic recombinant proteins: trends in US approvals 1982-2002. Curr. Opin. Mol. Ther. 5, 139–147 (2003).

Alex K Pavlou & Janice M Reichert. Recombinant protein therapeutics—success rates, market trends and values to 2010 Nature Biotechnology 22, 1513 - 1519 (2004)

Reichert, J.M. Therapeutic monoclonal antibodies: trends in development and approval in the US. Curr. Opin. Mol. Ther. 4: 110–118 (2002).

Web sites of:

Invitrogen, Clontech, Novagen, Stratagene

www.invitrogen.com/proteinexpression