Recombinant protein expression in lower eukaryotes
Yeast
Lecture 3
Advantages of eukaryotic hosts

• Conservation of many cellular processes between different eukaryotic species
  – Protein folding
  – Post translational modifications
  – Secretory pathway (proteins targeted to various organelles or exported for harvesting)
  – DNA replication
  – Cell cycle regulation
Recombinant protein production in eukaryotic cells

• Correct S-S bond formation
• Post-translational modifications
  – O- / N-glycosylation
• Proteolytic processing of a precursor into active form
  – Selected segments of amino acid sequences are removed to yield a functional protein
• Typically recombinant human proteins must be identical to the natural protein
  – Exception e.g. EPO
A general eukaryotic expression vector

- A selectable eukaryotic marker gene
- A eukaryotic promoter sequence
- The appropriate eukaryotic transcriptional and translational stop signal
- A sequence that signals polyadenylation of the transcript messenger RNA (mRNA)
- Origin of replication

- Marker gene and ori for propagation of vector in *E. coli*
Eukaryotic origin of replication

E. coli origin of replication

Eukaryotic selection marker

Coding sequence
Yeast as expression system: Advantages of yeast

- It is easy and cheap to grow on simple media in small vessels and large scale bioreactors
- secretes few of its own proteins -> product can easily be purified
- Tremendous range of vectors and genetic resources available, including promoters and regulatory systems
- Vectors can be maintained as plasmids or integrated
- We know a lot about *Saccharomyces cerevisiae* and also *Pichia pastoris*
- For therapeutics no risks of contaminations with human pathogens
Yeast as expression system: Disadvantages of yeast

- Longer growth time. Doubling time in bacteria is 20-30 min while in yeast it’s 1.5-2 hours at 30°C.
- Hyperglycosylation of secreted glycoproteins can be observed
- Post-translational modifications such as glycosylation may be different when compared to human proteins
- Higher number of recombination events
Saccharomyces cerevisiae as expression system

Vector and expression strategies
Some general facts about *Saccharomyces cerevisiae*

- The classic yeast is *S. cerevisiae*
- Has been extensively used as eukaryotic model system for the elucidation of many cellular processes
- Single cell fungus
- Diploid or haploid
- Reproduces asexually and sexually (genetics)
- 5 μm diameter
- First eukaryote being sequenced (1996)
  - Verified ORFs: 4932, uncharacterized ORFs:866, dubious ORFs: 809
- Generally recognized as safe (GRAS-status)
**S. cerevisiae expression vectors**

- Episomal plasmid vectors or *E. coli*/yeast shuttle plasmids
  - High copy number (20-100 copies per cell) based on 2μ origin of replication
  - Low copy number (1-3) based on CEN/ARS ori
- Integrative vectors
- Yeast artificial chromosomes (YACs)
2μ ori plasmid

- 2μ origin: copy number 25-200
- Bacteria & yeast origin of replication
- Yeast promoter and terminator
- Antibiotic resistance (bacteria)
- Yeast selectable markers
  - G418, zeocin resistance
  - Auxotrophic marker genes: URA3, HIS3, TRP1, LEU2, ADE2
- Can be unstable under conditions of large scale growth (>10 liters)

pYES2
5.9 kb
CEN/ARS based plasmid

- low copy vectors that incorporate part of an ARS along with part of a CEN sequence
- vectors replicate as though they are small independent chromosomes and are typically found as 1 to 3 copies
  - Copy number is stably maintained

- CEN/ARS
- URA3: auxotrophic marker
- Yeast promoter and terminator
- Amp<sup>R</sup>
- origin of replication (*E. coli*)

Centromere is required for segregation of chromosomes during mitosis
ARS is the origin of replication in the yeast genome
Yeast artificial chromosome (YAC)

• The YAC is designed to clone large fragments of DNA (100-1000kb)
• The YAC is maintained as a separate chromosome in the host yeast cell and is highly stable
• Use of YAC:
  – the physical mapping of human genomic DNA
  – the analysis of large transcription units
  – the formation of genomic libraries containing DNA from individual human chromosomes
  – Also for expression of complete metabolic pathways for secondary metabolites
Construction of yeast artificial chromosomes (YAC)

DNA fragment can range from 100 to 1000kb

Telomere sequence required to insure stability of YAC ends
CEN and ARS sequences required for stable maintenance of plasmids

Replication occurs but no proper mitotic segregation

Correct replication and mitotic segregation
Effect of TEL sequences on plasmid maintenance

<table>
<thead>
<tr>
<th>Plasmid with sequence from leu− cell</th>
<th>Transfected cell</th>
<th>Progeny of transfected cell</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal yeast</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two circles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circle with LEU and CEN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circle with ARS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear plasmid lacking TEL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>is unstable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Linear plasmids containing ARS and CEN behave like normal chromosomes if genomic fragment TEL is added to both ends</td>
<td>Yes</td>
<td>Good</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 6-46c, Molecular Cell Biology, Sixth Edition © 2008 W.H. Freeman and Company*
Integration into the chromosome by homologous recombination

- Ability of complementary sequences to align and exchange fragments in a double crossover event
- High efficiency with linear fragments, low efficiency with circular plasmids
- Targeted integration of gene of interest or deletion of interfering gene
- Site-specificity and high recombination frequency of *S. cerevisiae* is major advantage over other yeast systems
  - Competition between HR and NHEJ activities!
Expression in S. cerevisiae
Use of integrative systems

- Yeast integrative plasmid lacks eukaryotic origin of replication!
- Homologous region 40-50 bp on each side typically sufficient
Homologous recombination for integration and deletion

- Homologous region 40-50 bp on each side
- Deletion of target gene
- Integration of gene of interest into genome, together with selection marker
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
- $\alpha_1$ antitrypsin
- Blood coagulation factor XIIIa
Secretion of heterologous proteins
Secretion of heterologous proteins

- Protein secretion is basically the same in all eukaryotic organisms
- Signal sequence of protein is recognized by signal recognition particle
- SRP directs secretory protein to the translocation machinery on the ER-membrane
- Translation and translocation process are coupled
- After translocation proteins undergo:
  - Folding
  - Post-translational modifications
  - Is transported through secretory pathway
  - Released into the medium
Mat-alpha leader sequence driving protein expression

- Matα leader consist of pre- and pro-sequence
- Pre-sequence functions as ER import signal
- The pro-peptide, a hydrophobic protein (binding to chaperones) interrupted by short stretches of charged or polar amino acids, is thought to slow down and ensure proper folding of the entire protein
- Sequence of spacer (S) peptide has influence on processing by dipeptidyl aminopeptidase
- Proteolytic processing occurs during translocation (pre) and when leaving the Golgi (pro)

Kjeldsen T. 2000
### Promoter used for expression in *S. cerevisiae*

**Table 7.1** Promoters for *S. cerevisiae* expression vectors

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Expression conditions</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid phosphatase (<em>PH05</em>)</td>
<td>Phosphate-deficient medium</td>
<td>Inducible</td>
</tr>
<tr>
<td>Alcohol dehydrogenase I (<em>ADHI</em>)</td>
<td>2–5% Glucose</td>
<td>Constitutive</td>
</tr>
<tr>
<td>Alcohol dehydrogenase II (<em>ADHII</em>)</td>
<td>0.1–0.2% Glucose</td>
<td>Inducible</td>
</tr>
<tr>
<td>Cytochrome c₁ (<em>CYC1</em>)</td>
<td>Glucose</td>
<td>Repressible</td>
</tr>
<tr>
<td>Gal-1-P Glc-1-P uridylyltransferase</td>
<td>Galactose</td>
<td>Inducible</td>
</tr>
<tr>
<td>Galactokinase (<em>GAL1</em>)</td>
<td>Galactose</td>
<td>Inducible</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase (<em>GAPD, GAPDH</em>)</td>
<td>2–5% Glucose</td>
<td>Constitutive</td>
</tr>
<tr>
<td>Metallothionein (<em>CUP1</em>)</td>
<td>0.03–0.1 mM copper</td>
<td>Inducible</td>
</tr>
<tr>
<td>Phosphoglycerate kinase (<em>PGK</em>)</td>
<td>2–5% Glucose</td>
<td>Constitutive</td>
</tr>
<tr>
<td>Triose phosphate isomerase (<em>TPI</em>)</td>
<td>2–5% Glucose</td>
<td>Constitutive</td>
</tr>
<tr>
<td>UDP galactose epimerase (<em>GAL10</em>)</td>
<td>Galactose</td>
<td>Inducible</td>
</tr>
</tbody>
</table>
Galactose metabolism & galactose dependent gene expression
Glucose results in the inhibition of inducer formation: glucose represses $GAL2$ and $GAL3$ (Gal2 is also glucose-inactivated). Also the inhibition of the main transcriptional activator protein in the system: $GAL4$ expression is repressed.
Cloning into Yeast: inducible promoters

- **Gal1 promoter**
  - GAL genes – metabolism of galactose
  - Inducible by galactose
  - Repressed in presence of glucose

Ro et al., 2008. BMC Biotechnol. 8:83
Pichia pastoris as expression system

Vector and expression strategies
Comparison of expression system
*S. cerevisiae* – *P. pastoris*

- Problems with production in *S. cerevisiae*:
  - For some proteins production level low
  - Hyperglycosylation (more than 100 mannose residues in N-glycosylation)
  - Sometimes secretion not good -> protein retained in cell wall
  - Produces high amounts of EtOH -> toxic for the cells

- Advantages of production in *Pichia pastoris*:
  - Highly efficient promoter, tightly regulated (alcohol oxidase -> AOX, induced by MeOH)
  - Produces no EtOH -> very high cell density
  - Secretes very few of its own proteins
  - 10-100 fold higher heterologous protein expression levels
Methanol utilization by *P. pastoris*

- *Pichia* is a methylotrophic yeast (can metabolize MeOH)
- 2 genes encoding alcohol oxidase (AOX1 and AOX2)
  - AOX1 responsible for majority of activity
- During methanol oxidation, H$_2$O$_2$ is produced
  - Specialized compartment in the cell (peroxisome)
  - H$_2$O$_2$ is detoxified by catalase
Methanol metabolism in *P. pastoris* 

1. **Generation of reducing equivalents**

2. **Generation of cell constituents**

AOX1 promoter as tool for protein expression

- AOX1 gene inducible by methanol
  - Protein can reach up to 30% total soluble protein in cells grown with methanol
- Expression controlled at transcription level
- Similar to Gal1 promoter: Glucose represses transcription, even in presence of methanol -> often Glycerol used as carbon source
- AOX1 promoter powerful tool to drive protein expression
Only integrative vectors available for expression in Pichia
Integrative plasmid systems for *P. pastoris*

- Lower frequency of homologous recombination than in *S. cerevisiae*, due to higher NHEJ activity
- Strain engineering more tedious/random
Plasmids for expression in *P. pastoris*

### Table 2
Common *P. pastoris* Expression Vectors

<table>
<thead>
<tr>
<th>Vector name</th>
<th>Selectable markers</th>
<th>Feature</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular pHIL-D2</td>
<td>HIS4</td>
<td>NotI sites for AOX1 gene replacement</td>
<td>K. Sreekrishna (personal communication)</td>
</tr>
<tr>
<td>pAO815</td>
<td>HIS4</td>
<td>Expression cassette bounded by BamHI and BglII sites for generation of multicopy expression vector</td>
<td>Thill et al. (1990)</td>
</tr>
<tr>
<td>pPIC3K</td>
<td>HIS4 and kan'</td>
<td>Multiple cloning sites for insertion of foreign genes; G418 selection for multicopy strains</td>
<td>Scorer et al. (1993b)</td>
</tr>
<tr>
<td>pPICZ</td>
<td>ble'</td>
<td>Multiple cloning sites for insertion of foreign genes;</td>
<td>Higgins et al. (1998)</td>
</tr>
</tbody>
</table>

(continues)
Plasmids for expression in *P. pastoris*

**Table 2** (continued)

<table>
<thead>
<tr>
<th>Vector name</th>
<th>Selectable markers</th>
<th>Feature</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pHWO10</td>
<td>HIS4</td>
<td>Zeocin selection for multicopy strains; potential for fusion of foreign protein to His₆ and myc epitope tags</td>
<td>Waterham et al. (1997)</td>
</tr>
<tr>
<td>pGAPZ</td>
<td>ble</td>
<td>Expression controlled by constitutive GAPp</td>
<td>Invitrogen, Carlsbad, CA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expression controlled by constitutive GAPp; multiple cloning site for insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His₆ and myc epitope tags</td>
<td></td>
</tr>
<tr>
<td>Secretion</td>
<td></td>
<td>AOX1p fused to PHO1 secretion signal; XhoI, EcoRI, and BamHI sites available for insertion of foreign genes</td>
<td>K. Sreekrishna (personal communication); Invitrogen</td>
</tr>
<tr>
<td>pHIL-S1</td>
<td>HIS4</td>
<td>AOX1p fused to α-MF prepro signal sequence; XhoI (not unique), EcoRI, NotI, SmaI, and AvrII sites available for insertion of foreign genes; G418 selection for multicopy strains</td>
<td>Scoor et al. (1993b)</td>
</tr>
<tr>
<td>pPIC9K</td>
<td>HIS4 and kar</td>
<td>AOX1p fused to α-MF prepro signal sequence; multiple cloning site for insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His₆ and myc epitope tags</td>
<td>Higgins et al. (1998)</td>
</tr>
<tr>
<td>pPICZa</td>
<td>ble</td>
<td>Expression controlled by constitutive GAPp; GAPp fused to α-MF prepro signal sequence; multiple cloning site for insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His₆ and myc epitope tags</td>
<td>Invitrogen, Carlsbad, CA</td>
</tr>
<tr>
<td>pGAPZa</td>
<td>ble</td>
<td>Expression controlled by constitutive GAPp; GAPp fused to α-MF prepro signal sequence; multiple cloning site for insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His₆ and myc epitope tags</td>
<td></td>
</tr>
</tbody>
</table>
# P. pastoris expression host strains

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Genotype</th>
<th>phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y-11430</td>
<td></td>
<td>Wild-type</td>
</tr>
<tr>
<td>GS115</td>
<td>his4</td>
<td>Mut+ His-</td>
</tr>
<tr>
<td>KM71</td>
<td>aox1::ARG4   his4 arg4</td>
<td>MutS His-</td>
</tr>
<tr>
<td>MC100-3</td>
<td>aox1::ARG4   aox2::HIS4 his4 arg4</td>
<td>Mut- His+</td>
</tr>
<tr>
<td>SMD1168</td>
<td>pep4 his4</td>
<td>Mut+ His- protease deficient</td>
</tr>
</tbody>
</table>

- **Mut^+**: both AOX1 and AOX2 are functional -> methanol as inducer and as carbon source used
- **Mut^S**: only AOX2 functional -> methanol as inducer and just to minor extent as carbon source
- **Mut^-**: none of the AOX are functional -> methanol only as inducer
Other fungal expression systems

- *Hansenula polymorpha*
- *Kluyveromyces lactis*
- *Schizosaccharomyces pombe*
- *Schwanniomyces occidentalis*
- *Yarrowia lipolytica*
- *Trichoderma reesei*
Industrial Enzymes Produced:

• Xylanase – breaks down hemicellulose
  – pulp and paper,
  – textile industry
  – animal feed
• Cellulase – breaks down cellulose
  – animal feed
  – ethanol production
• Genes used: native to Trichoderma
• Modified by Directed Evolution Re-introduced into the fungus, Yield: 60 g/L
Broad host range expression systems

Benefit: Test expression of a protein in a number of expression hosts

<table>
<thead>
<tr>
<th>Module and gene</th>
<th>Donor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration</td>
<td></td>
</tr>
<tr>
<td>25S rDNA</td>
<td>A. adeninivorans</td>
</tr>
<tr>
<td>18S rDNA</td>
<td>A. adeninivorans</td>
</tr>
<tr>
<td>18S rDNA</td>
<td>H. polymorpha</td>
</tr>
<tr>
<td>Replication</td>
<td></td>
</tr>
<tr>
<td>2μm ARS</td>
<td>S. cerevisiae</td>
</tr>
<tr>
<td>ARS1</td>
<td>S. cerevisiae</td>
</tr>
<tr>
<td>HARS</td>
<td>H. polymorpha</td>
</tr>
<tr>
<td>Selection</td>
<td></td>
</tr>
<tr>
<td>URA3</td>
<td>S. cerevisiae</td>
</tr>
<tr>
<td>LEU2</td>
<td>S. cerevisiae</td>
</tr>
<tr>
<td>ALEU12m</td>
<td>A. adeninivorans</td>
</tr>
<tr>
<td>ATRP1m</td>
<td>A. adeninivorans</td>
</tr>
<tr>
<td>HIS4</td>
<td>P. pastoris</td>
</tr>
<tr>
<td>Expression</td>
<td></td>
</tr>
<tr>
<td>MOX promoter</td>
<td>H. polymorpha</td>
</tr>
<tr>
<td>AOX1 promoter</td>
<td>P. pastoris</td>
</tr>
<tr>
<td>TEF1 promoter</td>
<td>A. adeninivorans</td>
</tr>
<tr>
<td>GAA promoter</td>
<td>A. adeninivorans</td>
</tr>
<tr>
<td>RPS7 promoter</td>
<td>Y. lipolytica</td>
</tr>
</tbody>
</table>

HARS, Hansenula ARS.
Advantages of yeasts versus other eukaryotic expression systems

• Less expensive, easier to grow, higher throughput
• Shorter cell cycle time than tissue culture (Human embryonic kidney cells – 1 day doubling)
• Higher protein yield
• Protein pharmaceuticals free of human disease
• Fewer regulations compared to tissue culture
• More extensive genetics
• Transformation/DNA manipulations easier
Disadvantage of yeasts versus other eukaryotic expression systems

- Secretory proteins need to cross the cell wall in yeast!
- Limits secretion of bigger molecules
Summary

• As a unicellular eukaryote, yeast is quick, easy and inexpensive to genetically manipulate and culture

• A wealth of knowledge and tools available for S. cerevisiae, and more recently also other fungal expression systems

• High protein yield and ease of industrial scale-up, make yeast/fungal strains useful for protein production