Expression of proteins in yeast systems

Jaana Toikkanen
VTT Biotechnology
Expression of proteins in yeast systems

- Introduction to recombinant protein production
- Yeast *Saccharomyces cerevisiae*
- Expression of proteins in *S. cerevisiae*
- Secretory pathway of *S. cerevisiae*
- Attempts to enhance production of secreted proteins in *Saccharomyces cerevisiae*
- Nonconventional yeasts in biotechnology
Protein production steps

- Gene expression in the new host
  - Regulation of gene expression
  - Molecular biology methods

- Posttranslational modifications
  - Translocation to the secretory pathway
  - Folding and modifications
  - Glycosylation

- Protein secretion and intracellular transport
  - The secretory pathway: ER → Golgi → secret. vesicles → PM
  - Secretion to the growth medium

- Downstream processing
  - Chemical character of the protein
  - Protein engineering
  - Recovery & purification

Cell physiology - knowledge and engineering
Changes in physiology caused by protein overproduction
Bioprocess design for optimized production and recovery
Gene technology of protein production

General structure of genes

- Promoters are regulatory units and they may be strong or weak.
- Signal sequences direct the produced protein out of the cell.
- Introns are found in eukaryotic genes.
- cDNA produced from the mRNA does not contain introns.
- Organisms use slightly different codon-preferences.
Introduction to protein production

• Proteins may be intracellular or extracellular

• Intracellular proteins may be produced intracellularly or directed out of the cell by adding a signal sequence or by construction of a fusion protein with an extracellular protein

• In bacteria, proteins may end in the periplasmic space => formation of inclusion bodies

• Production may be enhanced by the use of fusion protein strategies (+introduction of specific protease sites for cutting of partners later)

• PCR has made it much easier to make the constructions needed
Introduction to protein production

• Homologous production = production by the gene donating host organism
• Heterologous production = production by a different host organism
• Classical mutagenesis can be used for increasing homologous expression
• Recombinant production may be facilitated by introducing the gene in an autonomously replicating plasmids or by integrating it to the genome
• Copy number of genes in the organism can be increased
• The degradation of the product by proteases may be controlled by e.g. use of protease defective mutants or by changing the sequence of the protein
Posttranslational modifications

• Almost all cells (except most prokaryotes) glycosylate their (extracellular) proteins. Glycosylation is species specific.

• N-glycosylation occur at Asn in protein sequences Asn-x-Thr/Ser (x not Pro). O-glycosylation occur usually at repeated sequences of Ser and Thr.

• Many chaperons have been found that help proteins to fold during/after translation.

• The signal sequence is cut off before the protein leaves the cell.

• Some proteins contain pro- and prepro- sequences needed to be cut off for activity.

• Some proteins are phosphorylated or modified in other ways.
Recombinant protein biosynthesis

• Optimal conditions for cell growth are typically sub-optimal or incompatible for recombinant protein overproduction

• Protein overproduction:
  Very demanding for energy (ATP) and reducing power (NADPH)
  Depends strongly on microbial growth rate

• Understand metabolic changes resulting from protein overexpression

• Evaluate product formation at various physiological states

• Identify possible bottlenecks

• Address such bottlenecks genetically or by optimizing cultivation conditions
Cellular responses to secretion stress

- **unfolded protein response (UPR)**: induction of genes involved in protein folding and removal from ER
- **attenuation of translation** (observed only in mammalian cells)

**Diagram:**
- **NUCLEUS**: Secreted protein
- **ER**: pdi, Bip
- **SECRETION**: tunikamycin, DTT, Ca++-ionophors, Sec mutations, Heterologous protein
- **SECRETION**

**Key Components:**
- pdi
- Bip
- PDI
What limits protein overexpression? Cellular as well as environmental factors

<table>
<thead>
<tr>
<th>Biological</th>
<th>Chemical</th>
<th>Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microbial Strain</td>
<td>Media Composition</td>
<td>Temperature</td>
</tr>
<tr>
<td></td>
<td>(C, N, P, S, micro-nutrients, surfactants)</td>
<td></td>
</tr>
<tr>
<td>Gene Dosage, Induction</td>
<td>Dissolved O₂, CO₂ Levels</td>
<td>Gas Transfer Rates</td>
</tr>
<tr>
<td>mRNA Stability</td>
<td>Chemical Inducer</td>
<td>Sheer Stress</td>
</tr>
<tr>
<td>Translational Capacity</td>
<td>pH</td>
<td>Cell Morphology</td>
</tr>
<tr>
<td>(ribosomal RNA)</td>
<td></td>
<td>(filaments vs. pellets)</td>
</tr>
<tr>
<td>Protein Folding,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secretion, Degradation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth rate and phase</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Desired properties for protein production host

Genetic methods for strain construction
• vectors for replication and integration
• selective markers
• efficient transformation

Protein production
• availability of strong constitutive and inducible promoter
• codon usage
• posttranslational processing
• product stability
• GRAS status
Desired properties for protein production host

Secretion
• signal sequence

Physiology
• growth conditions
• carbon and energy sources
## Comparison of expression systems

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>E. coli</th>
<th>Yeast</th>
<th>Insect cells</th>
<th>Mammalian cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell growth</td>
<td>rapid (30 min)</td>
<td>rapid (90 min)</td>
<td>slow (18-24 h)</td>
<td>slow (24 h)</td>
</tr>
<tr>
<td>Complexity of growth medium</td>
<td>minimum</td>
<td>minimum</td>
<td>complex</td>
<td>complex</td>
</tr>
<tr>
<td>Cost of growth medium</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>Expression level</td>
<td>high</td>
<td>low-high</td>
<td>low-high</td>
<td>low-moderate</td>
</tr>
<tr>
<td>Extracellular expression</td>
<td>secretion to periplasm</td>
<td>secretion to medium</td>
<td>secretion to medium</td>
<td>secretion to medium</td>
</tr>
<tr>
<td>Posttranslational modifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein folding</td>
<td>refolding usually required</td>
<td>refolding may be required</td>
<td>proper folding</td>
<td>proper folding</td>
</tr>
<tr>
<td>N-linked glycosylation</td>
<td>none</td>
<td>high mannose</td>
<td>simple, no sialic acid</td>
<td>complex</td>
</tr>
<tr>
<td>O-linked glycosylation</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Acetylation</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Acylation</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>gamma-Carboxylation</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
Yeast *Saccharomyces cerevisiae*

- Well known genetics
- Well developed genetic methods
- A model to other eukaryotes
- Full genomic sequence known
- About 6000 genes
- GRAS
- No toxins
- Ability to secrete proteins
- Capacity for posttranslational modifications
Yeast vectors

**YIp** = yeast integrating plasmid, single copy
- for integration into genomic locus

**YCp** = yeast centromeric plasmid
- replicating single copy plasmid
- for *e.g.* complementation studies

**YEp** = yeast episomal plasmid
- replicating multicopy (tens of copies) plasmid
- based on yeast 2µ plasmid
- for recombinant protein production, multicopy suppression studies etc.
Yeast gene technology

Transformation methods:
- spheroplast transformation
- whole cell transformation using LiAc
- electroporation

Selection of transformants:
- complementation of auxotrophic mutations
  - strain carries a mutation in a given gene (ura3, leu2, trp1, his3)
  - corresponding wt copy of the gene (URA3, LEU2, TRP1, HIS3) present in the vector
- dominant selection markers
  - resistance to G418, cycloheximide, formaldehyde, Cu^{2+}
Yeast promoters most commonly used for expression of heterologous genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Induction/Derepression</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC1</td>
<td>Alcohol dehydrogenase I</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PGK</td>
<td>Phosphoglyserate kinase</td>
<td>Fermentable carbon source</td>
<td>-</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Fermentable carbon source</td>
<td>-</td>
</tr>
<tr>
<td>PHO5</td>
<td>Acid phosphatase</td>
<td>Low PI</td>
<td>+</td>
</tr>
<tr>
<td>Gal1 GAL10</td>
<td>Galactokinase</td>
<td>Galactose</td>
<td>-</td>
</tr>
<tr>
<td>SUC2</td>
<td>Invertase</td>
<td>Low glucose</td>
<td>+</td>
</tr>
<tr>
<td>MFα1</td>
<td>Mating pheromone a</td>
<td>MATα</td>
<td>+</td>
</tr>
<tr>
<td>ADRIII</td>
<td>Alcohol dehydrogenase 2</td>
<td>Low glucose</td>
<td>-</td>
</tr>
</tbody>
</table>
\[ \alpha\text{-Amylase units x } 10^{-2} / \text{mg cell dry weight} \]

\[ \text{Cell dry weight (mg/ml)} \]

Legend:
- **SHORT**
- **LONG**
- **MIDDLE**
The graph shows productivity (U/g.h) over time (h) for different categories:

- **"long"** represented by green triangles
- **"middle"** represented by red squares
- **"short"** represented by blue circles

The x-axis represents time in hours (0 to 120), and the y-axis represents productivity in U/g.h. The graph illustrates the productivity profiles for each category over time.
Yeast glycosylation

- Fungal type glycans are composed of mannose only (*S. pombe* galactose)
- *S. cerevisiae* exceptional in making extensive mannosylation - not in all proteins
- Yeast type glycosylation may affect properties of the protein product *e.g.* activity or binding
- $\alpha$-1,3-Man linkage at the chain terminus is immunogenic - foreign to mammals
Bacillus $\alpha$-amylase produced in yeast

Western blot Coomassie stained

Y B Y B

Western blot Coomassie stained
The fusion strategy for enhanced folding and secretion

Yeast MFα1p fusion

FUSION PROTEINS IN OTHER SECRETORY SYSTEMS

Host

Carrier protein

*Escherichia coli*

*Staphylococcus* protein A (SPA)
Maltose binding protein (MBP)
Thioredoxin (DsbA)

*Aspergillus niger*

Glucoamylase (GA)

*Trichoderma reesei*

Cellobiohydrolase I (CBHI)

<table>
<thead>
<tr>
<th>Host</th>
<th>Carrier protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td><em>Staphylococcus</em> protein A (SPA)</td>
</tr>
<tr>
<td><em>Aspergillus niger</em></td>
<td>Glucoamylase (GA)</td>
</tr>
<tr>
<td><em>Trichoderma reesei</em></td>
<td>Cellobiohydrolase I (CBHI)</td>
</tr>
</tbody>
</table>

Diagram:
- Yeast MFα1p fusion
- Carrier part
- Heterologous protein
- KEX2
- α factor
- Potential N-glycosylation site
- KEX1 (yscα)
- KEX2 (yscF)
Yeast secretory pathway
The secretory pathway genes of yeast

V

YPT7
VPS33/SLP1
PTH1/VAM3

N

SEC65
SRP54
SRP101
SSA1-SSA4
SEC61, SSH1
SS1
SEB1,2
SEC62
SEC63
SEC66/SEC71
SEC67/SEC72
SEC70

S

E

YPT7
NYV1

SEC16
SEC19
SEC21
SEC26
SEC27

SEC12
SAR1
SEC13
SEC23
SEC24
SEC31

SEC10, SEC15
MSO1, DSS4, SMY1
SEC17, SEM1, TPM1
SEC18, SCD5
SEC19, MYO2
ACT1

ER

SEC53
SEC59
KAR2

ERD2
SED4
SEC20

VPS1
CHC1
CLC1

G

SEC22
BET1
BOS1
RET1
SEC21
SEC24
SEC31

RET2
RET3
UFE1

SEC14 ANP1 ERD1
PMR1 MNN9
KEX1 VAN1
KEX2 VRG4
STE14 SFT1

ARF1
ARF2
SEC21
ERD2

Sv

SNC1,2
SOS1,2
SEC1, SEC2, SEC3
SEC4, SEC4-GDP
SEC5, SEC6
SEC8, SEC9
SEC10, SEC15
MSO1, DSS4, SMY1
SEC17, SEM1, TPM1
SEC18, SCD5
SEC19, MYO2
ACT1

VMA12
VMA22
KTR1
MPD1
CNE1 PDI1 EUG1 FKB1 ERN1 SHR3 SEC11

GOS1

ARF1
ARF2
SEC21
ERD2

Sv

RET1
VPS1
CHC1
CLC1

SEC22
BET1
BOS1
RET1
SEC21
SEC24
SEC31

RET2
SEC14 ANP1 ERD1
PMR1 MNN9
KEX1 VAN1
KEX2 VRG4
STE14 SFT1

ARF1
ARF2
SEC21
ERD2

Sv
Ways of increasing protein secretion in yeast

Isolation of super secretory mutants
• several isolated
• most do not affect secretion
• almost all are recessive

Fusion of the heterologous protein to a well secreted endogenous protein
• prepro α-factor
• Hsp 150

Deletion of a quality control gene
• CNE1 (Calnexin homologue) -> misfolded proteins

Overexpression of er lumenal folding machinery
• PDI1 (protein disulfide isomerase)
• HAC1

Overexpression of components of the secretory machinery
• SEB1 (ER translocon component)
• SSO1, SSO2 (plasma membrane t-SNARES)

Optimization of production conditions
• Rich medium
• Low temperature
• pH
Nonconventional yeasts in biotechnology

- *Schwannimyces occidentalis*
- *Kluyveromyces lactis*
- *Pichia pastoris*
- *Pichia guilliermondii*
- *Pichia methanolica (Pichia pinus MH4)*
- *Hansenula polymorpha (Pichia angusta)*
- *Yarrowia lipolytica*
- *Arxula adeninivorans*
- *Candida maltosa*
- *Trichosporon*
Kluyveromyces lactis

- Can grow on inexpensive medium containing lactose
- Well developed genetics
- Product yield can be in g/l scale
- Commercially used for chymosin production
- GRAS status
- Ability to secrete high molecular weight proteins
- Protease problems with some proteins
**Pichia pastoris**

- Methylotrophic yeast that grows to high cell densities
- Strong inducible promoters
- Expression from integrated plasmid
- Requirement of MetOH for induction (toxicity and need for explosive-safe fermentors)
- Production system commercially available from Invitrogen
- High production yields
- Both intracellular and secreted protein production
- Requires precise fermentation conditions
- Product stability can be increased by using protease deficient strains or fusion to other stably expressed protein
### Selected proteins expressed in *Pichia pastoris*

<table>
<thead>
<tr>
<th>Protein Expressed</th>
<th>Expression Level (g / l)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enzymes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Proteases and Protease Inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Membrane Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Antigens and Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bm86 tick gut glycoprotein</td>
<td>1.5</td>
<td>Rodriguez, M. <em>et al.</em> (1994) <em>J. Biotechnology</em> 33: 135-146.</td>
</tr>
<tr>
<td><strong>Regulatory Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptokinase (active)</td>
<td>0.08</td>
<td>Hagenson, M.J. <em>et al.</em> (1989) <em>Enzyme Microbial Technology</em> 11: 650-656.</td>
</tr>
</tbody>
</table>
**Hansenula polymorpha**

- Methylotrophic yeast
- Secretes heterologous proteins, in many cases by the leader peptide of the heterologous protein
- Short fermentation times
- High level expression can be obtained on glycerol with certain promoters
- Rarely hyperglycosylation of the protein product
- Thermotolerant, growth optimum 42°
Yarrowia lipolytica

- Efficient and precise integrative transformation
- Strong regulated and constitutive promoters
- Very few compounds may be used as N or C source
Conclusions

Choices to be made:

Homologous or heterologous production?
   Genetic engineering needed in heterologous expression
   Classical mutagenesis (and genetic engineering) usually needed in homologous expression

Intracellular or extracellular?
   Wild type versus recombinant protein

Which production host?
   "Plug and play" a good choice if gene from unknown family

Fermentation system (size)?
Down-stream processing?

Keep it as simple as possible!
If you need help, please contact VTT Biotechnology!

Expression service
- *Trichoderma reesei* and *Aspergillus*
- *E.coli* and *Bacillus*
- *Pichia pastoris* and *Saccharomyces*
- Baculovirus

Fermenters
- 1 - 2000 l for yeast, filamentous fungi and bacteria
- Up to 8 l for baculovirus

Down-stream processing
- Semi-continuous centrifugation
- Drum filter for fungi, Cross flow filtration, Ultrafiltration
- Chromatography (semi-large scale)
- Freeze drying and granulation