Vectors for Recombinant DNA Technology

Plasmids
- Autonomous Replicaton
- Integration into genome
- Shuttle Plasmids
  - E.Coli → Target host

Phages
- Bacteriophage Lambda

Viruses
- Baculovirus – Insect Cells
- Retroviruses – Mammalian Cells

Cosmids, Bacmids
- Plasmid – Bacteriophage Hybrids

Artificial Chromosomes
- YAC,
Integration into genome

Site specific -- Ectopic

Autonomous Replication

Plasmids  Viruses  ARS

A Integrated

Transferred DNA

① A A A A A A A A
     Single copy

② A A A A A A A A A A
     Multiple tandem repeats

③ A A A A A A A A A A
     Multiple dispersed repeats
Gene Replacement
→ Double Cross-over at regions showing sufficient homology

Site specific Insertion
→ → Single Cross-over at regions showing sufficient homology

Ectopic Integration
→ Recombination at regions of no (low ?) Homology
Figure 13.9 (A) Diagram of the cloning vector pBluescript II. It contains a plasmid origin of replication, an ampicillin-resistance gene, a multiple cloning site (polylinker) within a fragment of the lacZ gene from E. coli, and a bacteriophage origin of replication. (B) Sequence of the multiple cloning site showing the unique restriction sites at which the vector can be opened for the insertion of DNA fragments. The numbers 657 and 759 refer to the position of the base pairs in the complete sequence of pBluescript. [Courtesy of Stratagene Cloning Systems, La Jolla, CA.]
Figure 13.10 Detection of recombinant plasmids through insertional inactivation of a fragment of the lacZ gene from E. coli. (A) Nonrecombinant plasmid containing an uninterrupted lacZ region. The multiple cloning site (MCS) within the region (not drawn to scale) is sufficiently small that the plasmid still confers β-galactosidase activity. (B) Recombinant plasmid with donor DNA inserted into the multiple cloning site. This plasmid confers ampicillin resistance but not β-galactosidase activity, because the donor DNA interrupting the lacZ region is large enough to render the region nonfunctional. (C) Transformed bacterial colonies. Cells in the white colonies contain plasmids with inserts that disrupt the lacZ region; those in the blue colonies do not. [C courtesy of Elena R. Lozovsky.]
Bakteriophage Lambda Vectors

FIGURE 10.3 Molecular cloning with lambda. Abbreviated genetic map of bacteriophage lambda showing the cohesive ends as circles (Figure 8.26). Charon 4A and 16 are both derivatives of lambda, which have various substitutions and deletions in the nonessential region. One of the substitutions in each case is a gene (β-Gal) that codes for the enzyme β-galactosidase, which permits detection of clones containing this phage. Whereas the wild-type lambda genome is 48.5 kilobase pairs, that for Charon 4A is 45.4 and that for Charon 16 is 41.7 kilobase pairs. The arrows (▼) shown above the maps of each phage indicate the sites recognized by the restriction enzyme EcoRI.
Figure 4.17  Bacteriophage λ cloning system. Bacteriophage λ is engineered to have two BamHI sites that flank the I/E region of the bacteriophage λ genome. The extensions indicate the cos ends of the λ DNA. For cloning, the source DNA is cut with 
BamHI and fractionated by size to isolate pieces that are 15 to 20 kb long. The bacteriophage λ DNA is also cut with BamHI. The two DNA samples are mixed and mixed with T4 DNA ligase. The ligation reaction mixture will contain a number of different DNA molecules, including (1) reconstituted bacteriophage λ and (2) the bacteriophage λ L and R regions with a 20-kb piece of DNA from the source DNA instead of the I/E region. These molecules are packaged into bacteriophage λ heads in vitro, and infective particles are formed after the addition of tail assemblies. After infection of E. coli cells that have P2 bacteriophage DNA integrated in their chromosome, only the molecules with the R and L regions and a cloned ~20-kb piece of DNA can replicate and form infectious bacteriophage λ. In this way, only the bacteriophage λ containing a DNA insert are perpetuated.
Cloning in Lambda Vectors

Nonessential region

Digestion with restriction enzymes

Ligation with foreign DNA

Packaging with phage

Infective phage particle
In vitro packaging into phage vesicles

Figure 4.16  Packaging of bacteriophage λ DNA into heads during the lytic cycle. A. DNA replication from the circular form of bacteriophage λ creates a linear form that is composed of contiguous, multiple lengths of bacteriophage DNA of approximately 50 kb each. B. Each newly assembled head is filled with a 50-kb unit of λ DNA before the tail assembly is attached.
Cosmid vectors

Principle: Plasmid DNA Transfer via Phage infection

Resulting recombinant clone Contains self-replicating plasmid
**E.coli – Saccharomyces cerevisiae Shuttle Vector**

*Figure 7.4* S. cerevisiae expression vector. The cDNA for human Cu/Zn-SOD has been cloned between the promoter (GAPDp) and termination–polyadenylation sequence (GAPDt) of the S. cerevisiae glyceraldehyde phosphate dehydrogenase gene. The LEU2 gene, cloned into the middle of the yeast 2µm plasmid DNA, encodes an enzyme of the leucine biosynthesis pathway. The yeast origin of replication is included in the 2µm plasmid sequences. The ampicillin resistance gene (Amp\(^r\)) and the *E. coli* origin of replication (ori\(^E\)) are derived from plasmid pBR322.
Expression Cassette

- Promoter
- Structural gene
- Pre-/Signal Sequences
- Fusion Domains / Tags
- Regulator Region
- Terminator
- Integration
- Selection marker
- Replication genes

E. coli

Vector
Gene Expression in Prokaryotes

**DNA**

- **DNA**
- **R** (Ribosome Binding Site)
- **P** (Promoter)
- **R** (Regulatory Region, e.g. Operator)

**Transcription**

- **initiation**
- **termination**

**mRNA**

- **Start**
- **Stop**

**Translation**

- **Protein A**
- **Protein B**
- **Protein C**

**Post-translational processing**

- **Ribosome Binding Site (Shine Dalgarno)**
- **Promoter (P)**
- **Regulatory Region (e.g. Operator)**
Gene Expression – Points to consider

Location in Genome → Autonomous replication, Integration

Transkription Initiation → Promoters  Transkription Termination

Regulatory Systems → positive/negative regulatory systems

Transkript Processing

Translation Initiation

RNA Structure
  Codon usage
  mRNA Stability

Post-translational modifications
  Modification of AA-side chains: Glycosylation, Phosphorylation, etc
  Proteolytic Processing

Protein Folding
  Disulfide bond formation

Assembly of subunits

Toxicity of gene products

Protein Degradation

Localization
  Intracellular
  Periplasmic
  Extracellular
  Membrane associated
  Organelle specific
  Surface display
Heterologous expression in prokaryotes – *E.coli*

**Transcription**

- constitutive promoters
- regulated promoters
  - lambda $p_L$, $p_R$
  - *lac, trp, tac, trc*
  - T7
- termination
  - rrnB (T1,T2), trpt
  - Lambda N gene (premature termination)

**m-RNA stability**

**Translation**

- Initiation – SD sequence \(...AGGAG...,\)
- elongation – codon usage

**Proteolysis**

- Lon, Clp, *htpR* (heat shock regulatory protein)

**Plasmid copy number and segregation**
Regulated Promoters ↔ Constitutive Promoters

Both systems are used

Preferred Combination: strong Promoters – tightly regulated

Constitutive promoters: weak to medium activity
Regulated Expression in Prokaryotes

Gene Repression
negative Control

Figure 3.21 Induction of the off state for transcription of a bacterial operon. The binding of a corepressor molecule (C) to an inactive repressor protein (IR) changes the conformation of the repressor protein. The corepressor–repressor protein complex (IR-C) binds to the operator region and blocks transcription of the operon by RNA polymerase.
Regulated Expression in Prokaryotes

Figure 3.22  Activation and deactivation of a bacterial operon. An activator protein (Act) binds to an activating site and enhances the rate of transcription of the operon. An effector molecule (E) binds to the activator protein. The Act–E complex does not bind to the activating site. The rate of transcription of the operon is diminished when the activating site is not occupied by the activating protein.
Regulated Expression in Eukaryotes

Complex Initiation System

Enhancer
Activators
Repressing Systems

Figure 3.24 Formation of an RNA polymerase II transcription initiation complex at a TATA box. Transcription factor TFIIH binds to a TATA box, and, in sequence, other transcription factors and RNA polymerase II bind to form a protein aggregate that is responsible for initiating transcription. The right-angled arrow designates the site of initiation and direction of transcription.
RNA Processing → Complex Mechanisms

- Capping
- Splicing
- 3′-Processing
- Export

DNA → Transcription → RNA processing → Splicing → Functional transcript → Translation → Protein

Nucleus → Cytoplasm
Figure 3.13 Alternative splicing of a eukaryotic primary RNA transcript. The bracketing arrows mark the sites that are spliced together after the removal of the intervening DNA region. In this example, exon 2 flanked by introns 1 and 2 is spliced out of the primary transcript and exons 1 and 3 are spliced together to form a functional mRNA transcript.
Figure 6.14 The expression vector pKK233-2. The plasmid pKK233-2 codes for the ampicillin resistance gene (Amp<sup>r</sup>) as a selectable marker gene, the tac promoter (ptac), the lacZ ribosome binding site (rbs), three restriction endonuclease cloning sites (NcoI, PstI, and HindIII), and two transcription termination sequences (T1 and T2). The arrow indicates the direction of transcription. The plasmid is not drawn to scale.
Expression Systems for E.coli

Inducible Promoters based on $lacI/lacO$ repressor/operator
pET-Expression system

- **P_{lac}**
- **lac O**
- **T7 Polymerase Gene**

- mRNA
- T7 Polymerase
- Protein

- **P_{T7}**
- **lac O**
- **Gene X**

- mRNA
- Protein X

- **not induced**

- Lac Repressor Protein
- E.coli DNA Polymerase
- T7 DNA Polymerase
pET-Expression system

- **lac** Promoter (P_{lac})
- **T7 Polymerase Gene**
- **IPTG oder Laktose**
- **T7 Promoter (P_{T7})**
- **lac O Gene X**
- **mRNA**
- **T7 Polymerase**
- **Protein X**
- **lac I Repressor**
- **E.coli DNA Polymerase**
- **T7 DNA Polymerase**
- **Protein X**
- **induced**
pETDuet™-1 is designed for the coexpression of two target genes. The vector contains two multiple cloning sites (MCS), each of which is preceded by a T7 promoter/lac operator and a ribosome binding site (rbs). The vector also carries the pBR322-derived ColE1 replicon, lacI gene and ampicillin resistance gene.
Bacteriophage Lambda Promoters $P_L$ and $P_R$

- **$C_{I\ 857}$** Lambda Repressor $C_I$
  - thermosensitive mutant

- **PL** $O_L$  
  - 30 °C, intact repressor

- **42 °C, disrupted repressor**

- **P_L based Expression Vector**

- **pLEX**
  - 2.9 kb
  - ColE1
  - Ampicillin
  - V1.0-150123

*These enzymes have two recognition sites found only in the multiple cloning site*
The pBAD Expression System is based on the araBAD operon which controls the arabinose metabolic pathway in E. coli. It allows you to precisely modulate heterologous expression to levels that are optimal for recovering high yields of your protein of interest.

The pBAD/His vector offers the following key features:

The PBAD promoter and the araC gene product for regulated expression of the gene of interest
N-terminal polyhistidine tag for rapid purification of fusion proteins using ProBond™ resin
Anti-Xpress™ epitope for detection of fusion proteins with the Anti-Xpress™ Antibody
Enterokinase cleavage site to facilitate removal of the fusion partner
Multiple cloning site in three reading frames to simplify subcloning in frame with the N-terminal polyhistidine tag
Ampicillin resistance gene and CoIE1 origin for selection and maintenance in E. coli
Heterologous expression in prokaryotes – *E.coli*

Transcription
- regulated promoters
  - lambda p_L, p_R
  - *lac, trp, tac, trc, araBAD*
  - T7
- termination
  - rrnB (T1,T2), *trpt*
  - Lambda N gene (premature termination)

m-RNA stability

Translation
- Initiation – SD sequence ...AGGAG....
- elongation – codon usage

Protein Folding

Proteolysis
- Lon, Clp, *htpR* (heat shock regulatory protein)

Posttranslational Processing

Plasmid copy number and segregation
m-RNA Stability

RNA has programmed half life
  no good information available on factors determining decay

Secondary structures $\rightarrow$ Target for RNases

Sequence structure $\rightarrow$ determines secondary structure and
  accessibility to RNases
Heterologous expression in prokaryotes – *E.coli*

**Transcription**
- regulated promoters
  - lambda $p_L$, $p_R$
  - *lac, trp, tac, trc, araBAD*  
  - T7
- termination
  - rrnB (T1,T2), trpt
  - Lambda N gene (premature termination)

**m-RNA stability**

**Translation**
- Initiation – SD sequence ...AGGAG....
- elongation – codon usage

**Protein Folding**

**Proteolysis**
- Lon, Clp, *htpR* (heat shock regulatory protein)

**Posttranslational Processing**

**Plasmid copy number and segregation**
Translation Initiation

- SD sequence ...AGGAG....
- Secondary structures

Translation elongation

- Codon usage
- Secondary structures
- Codon structure – translational frameshifting

![Diagram of codon structure showing Lys, Lys, Lys, Ser followed by Lys, Lys, Ile]
Translation - Prokaryotes

Shine-Dalgarno (SD) Sequence

<table>
<thead>
<tr>
<th>rRNA</th>
<th>5′-GAUACCAUCCUCCUUA-3′</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA</td>
<td>...GGAGG...(5-7bp)...AUG</td>
</tr>
</tbody>
</table>

Influences:

Secondary structure!! SD and AUG in unstructured region

Surrounding of SD and AUG!!!

Start

<table>
<thead>
<tr>
<th></th>
<th>AUG</th>
<th>GUG</th>
<th>UUG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>91%</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>
Translational Coupling

One-cistron mRNA

ribosome binding site

ATG ..........gene.........................

5'  

17-23 bases  12-18 bases

Two-cistron mRNA

ribosome binding site

ATG.... stop

5'

17-23 bases

ATG ..........gene.........................

12-18 bases

Translational Coupling
Translation - Eukaryotes

Start Codon

mRNA 5'−CAP......AUG
CAP structure essential for efficient translation initiation

Influences on Translation efficiency:

Surrounding of AUG!!

Kozak Consensus

........CC^A/GCC^AUGG...... mammalian

........^A/TA^A/CA^A/CA^AUGTC^T/C....... yeast
Translation Initiation

- SD sequence  ...AGGAG....

- Secondary structures

Translation elongation

- Codon usage

- Secondary structures

- Codon structure – translational frameshifting
<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino acid</th>
<th>Frequency of use in:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>E. coli</td>
</tr>
<tr>
<td>GGG</td>
<td>Glycine</td>
<td>0.13</td>
</tr>
<tr>
<td>GGA</td>
<td>Glycine</td>
<td>0.09</td>
</tr>
<tr>
<td>GCC</td>
<td>Glycine</td>
<td>0.38</td>
</tr>
<tr>
<td>GGU</td>
<td>Glycine</td>
<td>0.40</td>
</tr>
<tr>
<td>GAC</td>
<td>Glutamic acid</td>
<td>0.30</td>
</tr>
<tr>
<td>GAA</td>
<td>Glutamic acid</td>
<td>0.70</td>
</tr>
<tr>
<td>GGU</td>
<td>Aspartic acid</td>
<td>0.59</td>
</tr>
<tr>
<td>GAC</td>
<td>Aspartic acid</td>
<td>0.41</td>
</tr>
<tr>
<td>GUG</td>
<td>Valine</td>
<td>0.34</td>
</tr>
<tr>
<td>GUA</td>
<td>Valine</td>
<td>0.17</td>
</tr>
<tr>
<td>GUC</td>
<td>Valine</td>
<td>0.29</td>
</tr>
<tr>
<td>GCA</td>
<td>Valine</td>
<td>0.20</td>
</tr>
<tr>
<td>GCC</td>
<td>Alanine</td>
<td>0.34</td>
</tr>
<tr>
<td>GCU</td>
<td>Alanine</td>
<td>0.22</td>
</tr>
<tr>
<td>GGU</td>
<td>Alanine</td>
<td>0.39</td>
</tr>
<tr>
<td>GGC</td>
<td>Alanine</td>
<td>0.25</td>
</tr>
<tr>
<td>AAG</td>
<td>Lysine</td>
<td>0.24</td>
</tr>
<tr>
<td>AAA</td>
<td>Lysine</td>
<td>0.76</td>
</tr>
<tr>
<td>AAU</td>
<td>Asparagine</td>
<td>0.39</td>
</tr>
<tr>
<td>AAC</td>
<td>Asparagine</td>
<td>0.61</td>
</tr>
<tr>
<td>AUG</td>
<td>Methionine</td>
<td>1.00</td>
</tr>
<tr>
<td>AUU</td>
<td>Threonine</td>
<td>0.07</td>
</tr>
<tr>
<td>AUA</td>
<td>Threonine</td>
<td>0.47</td>
</tr>
<tr>
<td>AUC</td>
<td>Threonine</td>
<td>0.46</td>
</tr>
<tr>
<td>AGC</td>
<td>Threonine</td>
<td>0.23</td>
</tr>
<tr>
<td>ACA</td>
<td>Threonine</td>
<td>0.12</td>
</tr>
<tr>
<td>ACC</td>
<td>Threonine</td>
<td>0.21</td>
</tr>
<tr>
<td>UGG</td>
<td>Threonine</td>
<td>0.43</td>
</tr>
<tr>
<td>UGU</td>
<td>Tryptophan</td>
<td>1.00</td>
</tr>
<tr>
<td>UGC</td>
<td>Cysteine</td>
<td>0.43</td>
</tr>
<tr>
<td>UGA</td>
<td>Cysteine</td>
<td>0.57</td>
</tr>
<tr>
<td>UAA</td>
<td>Stop</td>
<td>0.00</td>
</tr>
<tr>
<td>UAG</td>
<td>Stop</td>
<td>0.09</td>
</tr>
<tr>
<td>UAU</td>
<td>Stop</td>
<td>0.62</td>
</tr>
<tr>
<td>UAC</td>
<td>Tyrosine</td>
<td>0.53</td>
</tr>
<tr>
<td>UUC</td>
<td>Tyrosine</td>
<td>0.47</td>
</tr>
<tr>
<td>UUU</td>
<td>Phenylalanine</td>
<td>0.51</td>
</tr>
<tr>
<td>UUC</td>
<td>Phenylalanine</td>
<td>0.49</td>
</tr>
<tr>
<td>UCC</td>
<td>Serine</td>
<td>0.13</td>
</tr>
<tr>
<td>UCA</td>
<td>Serine</td>
<td>0.12</td>
</tr>
<tr>
<td>UCU</td>
<td>Serine</td>
<td>0.19</td>
</tr>
<tr>
<td>UCC</td>
<td>Serine</td>
<td>0.17</td>
</tr>
<tr>
<td>AGU</td>
<td>Serine</td>
<td>0.13</td>
</tr>
<tr>
<td>AGC</td>
<td>Serine</td>
<td>0.27</td>
</tr>
<tr>
<td>CGG</td>
<td>Arginine</td>
<td>0.08</td>
</tr>
<tr>
<td>CGA</td>
<td>Arginine</td>
<td>0.05</td>
</tr>
<tr>
<td>CGU</td>
<td>Arginine</td>
<td>0.02</td>
</tr>
<tr>
<td>CCC</td>
<td>Arginine</td>
<td>0.07</td>
</tr>
<tr>
<td>UGC</td>
<td>Arginine</td>
<td>0.07</td>
</tr>
<tr>
<td>CGG</td>
<td>Arginine</td>
<td>0.05</td>
</tr>
<tr>
<td>CGU</td>
<td>Arginine</td>
<td>0.02</td>
</tr>
<tr>
<td>CCC</td>
<td>Arginine</td>
<td>0.07</td>
</tr>
<tr>
<td>Codon</td>
<td>Amino acid</td>
<td>E. coli</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>GGG</td>
<td>Glycine</td>
<td>0.13</td>
</tr>
<tr>
<td>GGA</td>
<td>Glycine</td>
<td>0.09</td>
</tr>
<tr>
<td>GGU</td>
<td>Glycine</td>
<td>0.38</td>
</tr>
<tr>
<td>GGC</td>
<td>Glycine</td>
<td>0.40</td>
</tr>
<tr>
<td>GAG</td>
<td>Glutamic acid</td>
<td>0.30</td>
</tr>
<tr>
<td>GAA</td>
<td>Glutamic acid</td>
<td>0.70</td>
</tr>
<tr>
<td>GAU</td>
<td>Aspartic acid</td>
<td>0.59</td>
</tr>
<tr>
<td>GAC</td>
<td>Aspartic acid</td>
<td>0.41</td>
</tr>
<tr>
<td>GUG</td>
<td>Valine</td>
<td>0.34</td>
</tr>
<tr>
<td>GUA</td>
<td>Valine</td>
<td>0.17</td>
</tr>
<tr>
<td>GUU</td>
<td>Valine</td>
<td>0.29</td>
</tr>
<tr>
<td>GUC</td>
<td>Valine</td>
<td>0.20</td>
</tr>
<tr>
<td>GCG</td>
<td>Alanine</td>
<td>0.34</td>
</tr>
<tr>
<td>GCA</td>
<td>Alanine</td>
<td>0.22</td>
</tr>
<tr>
<td>GCU</td>
<td>Alanine</td>
<td>0.19</td>
</tr>
<tr>
<td>GCC</td>
<td>Alanine</td>
<td>0.25</td>
</tr>
<tr>
<td>AAG</td>
<td>Lysine</td>
<td>0.24</td>
</tr>
<tr>
<td>AAA</td>
<td>Lysine</td>
<td>0.76</td>
</tr>
<tr>
<td>AAU</td>
<td>Asparagine</td>
<td>0.39</td>
</tr>
<tr>
<td>AAC</td>
<td>Asparagine</td>
<td>0.61</td>
</tr>
<tr>
<td>AUG</td>
<td>Methionine</td>
<td>1.00</td>
</tr>
<tr>
<td>AUA</td>
<td>Isoleucine</td>
<td>0.07</td>
</tr>
<tr>
<td>AUU</td>
<td>Isoleucine</td>
<td>0.47</td>
</tr>
<tr>
<td>AUC</td>
<td>Isoleucine</td>
<td>0.46</td>
</tr>
<tr>
<td>ACG</td>
<td>Threonine</td>
<td>0.23</td>
</tr>
<tr>
<td>ACA</td>
<td>Threonine</td>
<td>0.12</td>
</tr>
<tr>
<td>ACU</td>
<td>Threonine</td>
<td>0.21</td>
</tr>
<tr>
<td>ACC</td>
<td>Threonine</td>
<td>0.43</td>
</tr>
<tr>
<td>UGG</td>
<td>Tryptophan</td>
<td>1.00</td>
</tr>
<tr>
<td>UGU</td>
<td>Cysteine</td>
<td>0.43</td>
</tr>
<tr>
<td>UGA</td>
<td>Cysteine</td>
<td>0.57</td>
</tr>
<tr>
<td>Codon</td>
<td>Amino Acid</td>
<td>Frequency</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>UGA</td>
<td>Stop</td>
<td>0.30</td>
</tr>
<tr>
<td>UAG</td>
<td>Stop</td>
<td>0.09</td>
</tr>
<tr>
<td>UAA</td>
<td>Stop</td>
<td>0.62</td>
</tr>
<tr>
<td>UAU</td>
<td>Tyrosine</td>
<td>0.53</td>
</tr>
<tr>
<td>UAC</td>
<td>Tyrosine</td>
<td>0.47</td>
</tr>
<tr>
<td>UUU</td>
<td>Phenylalanine</td>
<td>0.51</td>
</tr>
<tr>
<td>UUC</td>
<td>Phenylalanine</td>
<td>0.49</td>
</tr>
<tr>
<td>UCG</td>
<td>Serine</td>
<td>0.13</td>
</tr>
<tr>
<td>UCA</td>
<td>Serine</td>
<td>0.12</td>
</tr>
<tr>
<td>UCU</td>
<td>Serine</td>
<td>0.19</td>
</tr>
<tr>
<td>UCC</td>
<td>Serine</td>
<td>0.17</td>
</tr>
<tr>
<td>AGU</td>
<td>Serine</td>
<td>0.13</td>
</tr>
<tr>
<td>AGC</td>
<td>Serine</td>
<td>0.27</td>
</tr>
<tr>
<td>CGG</td>
<td>Arginine</td>
<td>0.08</td>
</tr>
<tr>
<td>CGA</td>
<td>Arginine</td>
<td>0.05</td>
</tr>
<tr>
<td>CGU</td>
<td>Arginine</td>
<td>0.42</td>
</tr>
<tr>
<td>CGC</td>
<td>Arginine</td>
<td>0.37</td>
</tr>
<tr>
<td>AGG</td>
<td>Arginine</td>
<td>0.03</td>
</tr>
<tr>
<td>AGA</td>
<td>Arginine</td>
<td>0.04</td>
</tr>
<tr>
<td>CAG</td>
<td>Glutamine</td>
<td>0.69</td>
</tr>
<tr>
<td>CAA</td>
<td>Glutamine</td>
<td>0.31</td>
</tr>
<tr>
<td>CAU</td>
<td>Histidine</td>
<td>0.52</td>
</tr>
<tr>
<td>CAC</td>
<td>Histidine</td>
<td>0.48</td>
</tr>
<tr>
<td>CUG</td>
<td>Leucine</td>
<td>0.55</td>
</tr>
<tr>
<td>CUA</td>
<td>Leucine</td>
<td>0.03</td>
</tr>
<tr>
<td>CUU</td>
<td>Leucine</td>
<td>0.10</td>
</tr>
<tr>
<td>CUC</td>
<td>Leucine</td>
<td>0.10</td>
</tr>
<tr>
<td>UUG</td>
<td>Leucine</td>
<td>0.11</td>
</tr>
<tr>
<td>UUA</td>
<td>Leucine</td>
<td>0.11</td>
</tr>
<tr>
<td>CCG</td>
<td>Proline</td>
<td>0.55</td>
</tr>
<tr>
<td>CCA</td>
<td>Proline</td>
<td>0.20</td>
</tr>
<tr>
<td>CCU</td>
<td>Proline</td>
<td>0.16</td>
</tr>
<tr>
<td>CCC</td>
<td>Proline</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Heterologous expression in prokaryotes – *E.coli*

Transcription

regulated promoters
- lambda $p_L$, $p_R$
- *lac, trp, tac, trc, araBAD*
- T7

termination
- rRN (T1,T2), trpt
- Lambda N gene (premature termination)

m-RNA stability

Translation

Initiation – SD sequence ...AGGAG....

elaboration – codon usage

**Protein Folding**

**Proteolysis**

- Lon, Clp, *htpR* (heat shock regulatory protein)

**Posttranslational Processing**

- Plasmid copy number and segregation
Protein Folding

Translation Conditions

Elongation velocity
Codon Structure – Pausing
Domain folding

Disulide Bond Formation
Redox Conditions
E.coli Cytosol $\rightarrow$ bad conditions - reductive
E.coli Periplasm $\rightarrow$ optimal conditions - oxidative

Chaperones
Inclusion Body Formation

Expression velocity $\rightarrow$ Translation

Protein Folding

The Department of Surface Biotechnology with the Center for Surface Biotechnology, Box 577, BMC, 751 23 Uppsala
Heterologous expression in prokaryotes – *E.coli*

Transcription
- regulated promoters
  - lambda p_L, p_R
  - lac, trp, tac, trc, araBAD
  - T7
- termination
  - rrnB (T1,T2), trpt
  - Lambda N gene (premature termination)

m-RNA stability
Translation
- Initiation – SD sequence ...
- elongation – codon usage

Protein Folding
Proteolysis
- Lon, Clp, *htpR* (heat shock regulatory protein)

Posttranslational Processing
Plasmid copy number and segregation
Post-translational modifications
  Side Chain Modifications
    Glycosylation, Phosphorylation, Sulfatation, etc.
  Proteolytic Processing
    ss Cleavage
    Pro-protein processing
    N/C-terminal Processing

Post-translational Processing in prokaryotes – *E.coli*

N-terminal processing – the problem of Met
  f-Met deformylase
  methionine aminopeptidase (MAP) of *E.coli*
  peptidase M (*S. typhimurium*)
  aminopeptidase M: Exopeptidase $\rightarrow$ ...... X-Pro
  aminopeptidase P: $\text{NH}_2$-X-/Pro
  dipeptidylaminopeptidase I (DAP-I, Cathepsin C) $\rightarrow$ not at $\text{NH}_2$ Pro/Arg/Lys

protein fusion strategies
  sequence specific proteases
  tags
<table>
<thead>
<tr>
<th>Copies/genome</th>
<th>Activity (U/mL of mid-log cells)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>500</td>
</tr>
<tr>
<td>5</td>
<td>2,300</td>
</tr>
<tr>
<td>7</td>
<td>3,100</td>
</tr>
<tr>
<td>8</td>
<td>3,400</td>
</tr>
<tr>
<td>9</td>
<td>4,400</td>
</tr>
<tr>
<td>Multicopy plasmid</td>
<td>700</td>
</tr>
</tbody>
</table>

Table 6.6  Effect of plasmid copy number on host cell growth rate

<table>
<thead>
<tr>
<th>E. coli HB101 with plasmid:</th>
<th>Plasmid copy number</th>
<th>Relative specific growth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>A</td>
<td>12</td>
<td>0.92</td>
</tr>
<tr>
<td>B</td>
<td>24</td>
<td>0.91</td>
</tr>
<tr>
<td>C</td>
<td>60</td>
<td>0.87</td>
</tr>
<tr>
<td>D</td>
<td>122</td>
<td>0.82</td>
</tr>
<tr>
<td>E</td>
<td>408</td>
<td>0.77</td>
</tr>
</tbody>
</table>


The different plasmids, designated A, B, C, D and E, encode only β-lactamase and are all the same size. The growth rates were normalized to the growth rate value for E. coli HB101 without a plasmid.
Metabolic load

Figure 6.17  Schematic representation of the pathways for glucose metabolism in an *E. coli* strain that has been transformed with a plasmid carrying the genes for acetolactate synthase (ALS).
Gene Fusion Strategies

Intracellular expression

Secretory expression

Fusion proteins

ATG

Protein of Interest

mRNA

Signal peptidase

Sequence specific peptidase

Tag

Protein of Interest

Protein X

Protein Y

Tag 1

Tag 2

Secretory expression

in vivo processing

Fusion proteins

in vitro processing

Secretory expression

in vivo processing
### Table 6.3 Some fusion systems used to facilitate the purification of foreign proteins produced in E. coli

<table>
<thead>
<tr>
<th>Fusion partner</th>
<th>Size</th>
<th>Ligand</th>
<th>Elution condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZZ</td>
<td>14 kDa</td>
<td>IgG</td>
<td>Low pH</td>
</tr>
<tr>
<td>His tail</td>
<td>6–10 aa</td>
<td>Ni²⁺</td>
<td>Imidazole</td>
</tr>
<tr>
<td>Strep-tag</td>
<td>10 aa</td>
<td>Streptavidin</td>
<td>Iminobiotin</td>
</tr>
<tr>
<td>PinPoint</td>
<td>13 kDa</td>
<td>Streptavidin</td>
<td>Biotin</td>
</tr>
<tr>
<td>MBP</td>
<td>40 kDa</td>
<td>Amylose</td>
<td>Maltose</td>
</tr>
<tr>
<td>β-Lactamase</td>
<td>27 kDa</td>
<td>Phenyl-boronate</td>
<td>Borate</td>
</tr>
<tr>
<td>GST</td>
<td>25 kDa</td>
<td>Glutathione</td>
<td>Reducing agent</td>
</tr>
<tr>
<td>Flag</td>
<td>8 aa</td>
<td>Specific MAb</td>
<td>Low calcium</td>
</tr>
</tbody>
</table>

Adapted from Nygren et al., 1994, Trends Biotechnol. 12:184–188.

Abbreviations: aa, amino acids; kDa, kilodaltons; ZZ, a fragment of Staphylococcus aureus protein A; His, histidine; Strep-tag, a peptide with affinity for streptavidin; PinPoint, a protein fragment which is

### Table 2 Sequence and size of affinity tags

<table>
<thead>
<tr>
<th>Tag</th>
<th>Residues</th>
<th>Sequence</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly-Arg</td>
<td>5–6</td>
<td>RRRRR</td>
<td>0.80</td>
</tr>
<tr>
<td>Poly-His</td>
<td>2–10</td>
<td>HHHHHH</td>
<td>0.84</td>
</tr>
<tr>
<td>FLAG</td>
<td>8</td>
<td>DYYKDDDK</td>
<td>1.01</td>
</tr>
<tr>
<td>Strep-tag II</td>
<td>8</td>
<td>WSHFPQFEK</td>
<td>1.06</td>
</tr>
<tr>
<td>c-myc</td>
<td>11</td>
<td>EQKLISEEDL</td>
<td>1.20</td>
</tr>
<tr>
<td>S-</td>
<td>15</td>
<td>KETAAKPERQHMDS</td>
<td>1.75</td>
</tr>
<tr>
<td>HAT-</td>
<td>19</td>
<td>KDHLIHNVHKEFHAHHK</td>
<td>2.31</td>
</tr>
<tr>
<td>3x FLAG</td>
<td>22</td>
<td>DYYKDHGDYKDHIDYKDDDDK</td>
<td>2.73</td>
</tr>
<tr>
<td>Calmodulin-binding peptide</td>
<td>26</td>
<td>KRRWKKNHIAVSAANRFFKISSSGL</td>
<td>2.96</td>
</tr>
<tr>
<td>Cellulose-binding domains</td>
<td>27–189</td>
<td>Domains</td>
<td>3.00–20.00</td>
</tr>
<tr>
<td>SBP</td>
<td>38</td>
<td>MDEKTGGWRGCGHVVEGLAGELEQLRARLEHHPQGQREP</td>
<td>4.03</td>
</tr>
<tr>
<td>Chitin-binding domain</td>
<td>51</td>
<td>TNPFGVSAAQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPNSVPAWQLQ</td>
<td>5.59</td>
</tr>
<tr>
<td>Glutathione S-transferase</td>
<td>211</td>
<td>Protein</td>
<td>26.00</td>
</tr>
<tr>
<td>Maltose-binding protein</td>
<td>396</td>
<td>Protein</td>
<td>40.00</td>
</tr>
<tr>
<td>Affinity tag</td>
<td>Matrix</td>
<td>Elution condition</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------------------</td>
<td>--------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Poly-Arg</td>
<td>Cation-exchange resin</td>
<td>NaCl linear gradient from 0 to 400 mM at alkaline pH&gt;8.0</td>
<td></td>
</tr>
<tr>
<td>Poly-His</td>
<td>Ni(^{2+})-NTA, Co(^{2+})-CMA (Talon)</td>
<td>Imidazole 20-250 mM or low pH</td>
<td></td>
</tr>
<tr>
<td>FLAG</td>
<td>Anti-FLAG monoclonal antibody</td>
<td>pH 3.0 or 2–5 mM EDTA</td>
<td></td>
</tr>
<tr>
<td>Strep-tag II</td>
<td>Strep-Tactin (modified streptavidin)</td>
<td>2.5 mM desfthiobiotin</td>
<td></td>
</tr>
<tr>
<td>c-myc</td>
<td>Monoclonal antibody</td>
<td>Low pH</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>S-fragment of RNaseA</td>
<td>3 M guanidine thiocyanate,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2 M citrate pH 2, 3 M magnesium chloride</td>
<td></td>
</tr>
<tr>
<td>HAT (natural histidine affinity tag)</td>
<td>Co(^{2+})-CMA (Talon)</td>
<td>150 mM imidazole or low pH</td>
<td></td>
</tr>
<tr>
<td>Calmodulin-binding peptide</td>
<td>Calmodulin</td>
<td>EGTA or EGTA with 1 M NaCl</td>
<td></td>
</tr>
<tr>
<td>Cellulose-binding domain</td>
<td>Cellulose</td>
<td>Family I: guanidine HCl or urea&gt;4 M</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Family II/III: ethylene glycol</td>
<td></td>
</tr>
<tr>
<td>SBP</td>
<td>Streptavidin</td>
<td>2 mM Biotin</td>
<td></td>
</tr>
<tr>
<td>Chitin-binding domain</td>
<td>Chitin</td>
<td>Fused with intein: 30–50 mM dithiothreitol,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta)-mercaptoethanol or cysteine</td>
<td></td>
</tr>
<tr>
<td>Glutathione S-transferase</td>
<td>Glutathione</td>
<td>5–10 mM reduced glutathione</td>
<td></td>
</tr>
<tr>
<td>Maltose-binding protein</td>
<td>Cross-linked amylose</td>
<td>10 mM maltose</td>
<td></td>
</tr>
</tbody>
</table>
Cleavage of Tags:

- **Enterokinase**
  - D-D-D-D-K-X1

- **TEV protease**
  - E-X-X-Y-X-Q-S

- **α-thrombin**
  - X4-X3-P-R[K]-X1'-X2
  - L - V-P-R- G - S

---

**Table 4** Cleavage (%) of enterokinase through densitometry (Hosfield and Lu, 1999) based on the amino acid residue X1. The sequence...-GSDKKDDDKK-X1-ADQLTERQIA-... of a GST-calmodulin fusion protein was tested using 5 mg protein digested with 0.2 Uof enterokinase for 16 h at 37 °C.

<table>
<thead>
<tr>
<th>Amino acid in position X1</th>
<th>Cleavage of enterokinase (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>88</td>
</tr>
<tr>
<td>Methionine</td>
<td>86</td>
</tr>
<tr>
<td>Lysine</td>
<td>85</td>
</tr>
<tr>
<td>Leucine</td>
<td>85</td>
</tr>
<tr>
<td>Asparagine</td>
<td>85</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>85</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>84</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>84</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>80</td>
</tr>
<tr>
<td>Glutamine</td>
<td>79</td>
</tr>
<tr>
<td>Valine</td>
<td>79</td>
</tr>
<tr>
<td>Arginine</td>
<td>78</td>
</tr>
<tr>
<td>Threonine</td>
<td>78</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>78</td>
</tr>
<tr>
<td>Histidine</td>
<td>76</td>
</tr>
<tr>
<td>Serine</td>
<td>76</td>
</tr>
<tr>
<td>Cysteine</td>
<td>74</td>
</tr>
<tr>
<td>Glycine</td>
<td>74</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>67</td>
</tr>
<tr>
<td>Proline</td>
<td>61</td>
</tr>
</tbody>
</table>

**Figure 6.6** Proteolytic cleavage of a fusion protein by blood coagulation factor Xa. The factor Xa recognition sequence (Xa linker sequence) lies between the amino acid sequences of two different proteins. A functional cloned gene protein (with Val at its N terminus) is released after cleavage.
Tag purification strategies

**Figure 6.8** Immunaoaffinity chromatographic purification of a fusion protein. An antibody that binds to the marker peptide of the fusion protein (anti-marker peptide antibody) is attached to a solid polypropylene support. The secreted proteins are passed through the column containing the bound antibody. The marker peptide portion of the fusion protein is bound to the antibody, and the other proteins pass through. The immunopurified fusion protein can then be selectively eluted from the column.

1. Concentrate secreted protein mixture
2. Prepare immunaoaffinity column
3. Add secreted protein mixture to the column
4. Elute fusion protein

Removal of Tag

Proteolytic cleavage of Tag
Figure 6.7 Fusion protein cloning vector. The plasmid contains an ampicillin resistance (Amp') gene as the selectable marker, a DNA sequence encoding the N-terminal segment of the E. coli outer membrane protein (ompF), a restriction endonuclease site (AbcI) for cloning, and a truncated β-galactosidase gene (lacZ). The cloned gene (Gene) is inserted into the AbcI site. After transcription and translation, a tridrid protein is produced.
Examples for fusion strategies

For E.coli:

Maltose binding protein
Thioredoxin reductase

Generally: well soluble proteins
Well folded proteins

Fusions can hep for:

Translation initiation
Folding
Protein detection: Antibodies against
Fusion partner (also with small tags)
Eukaryotic Expression Systems

- Fungi – Yeasts
- Insect Cells
- Plant Cells
- Mammalian Cells
  - Mouse
  - Hamster
  - Avian
  - Human
- Transgenic Plants
- Transgenic Animals
Figure 7.4  *S. cerevisiae* expression vector. The cDNA for human Cu/Zn-SOD has been cloned between the promoter (GAPD$p$) and termination–polyadenylation sequence (GAPD$t$) of the *S. cerevisiae* glyceraldehyde phosphate dehydrogenase gene. The *LEU2* gene, cloned into the middle of the yeast 2μm plasmid DNA, encodes an enzyme of the leucine biosynthesis pathway. The yeast origin of replication is included in the 2μm plasmid sequences. The ampicillin resistance gene (Amp$^r$) and the *E. coli* origin of replication (ori$^E$) are derived from plasmid pBR322.

**Diagram:**
- Autonomous replication in yeast
- Terminator
- Promoter
- Selection in yeast
- Selection in *E. coli*
- Autonomous replication in *E. coli*
**S. cerevisiae**

**Expression vectors**

2µ-based multicopy vector

---

**Figure 3 - pYES2 vector**

**Vector features**

- **TEFp** TEF1 promoter (nt 2673-3081)
- **CYC1t** S. cerevisiae CYC1 terminator (nt 2352-2610)
- **KAN** Kanamycine resistance gene (aminoglycoside phosphotransferase), allows selection in yeast using 200 mg/ml G418 (nt 190-1571)
- **2micron** Origin of replication derived from the endogenous yeast 2m circle. Allows propagation of plasmids in yeast at high copy numbers (10-50 copies/cell, nt 5291-6637)
- **AmpR** Ampicillin resistance gene (nt 4300-5158)
Induction of the Gal genes in yeast. The constitutively expressed proteins GAL80 and GAL4 regulate expression of several genes required for the conversion of galactose to glucose-6-phosphate. The product of the GAL4 gene binds to the Upstream Activating Sequence (UAS-GAL). The activating effect of GAL4 is repressed by GAL80. In the presence of galactose, a metabolic product is formed which releases GAL80p from GAL4p and then activation of the GAL genes cluster occurs.
Protein Expression in *Pichia pastoris*

- Methylotrophic yeast
  - Two alcohol oxidase genes: *AOX1, AOX2*
  - AOX1: 5% of total mRNA, 30% of total protein
- Well established commercial expression system
- More than 300 proteins successfully expressed
  - (bacterial, viral, fungal, plant, protozoan, invertebrate, vertebrate → 120 human proteins)
- High cell density fermentation (>100 g/L) on simple media
- No switch to anaerobic fermentation (ethanol problem)
- Stable integration into host chromosome
- Intracellular and secretory production capacities
- Advantages of a eukaryotic host cell – but simple system
  - Glycosylation (N-linked, high-mannose type)
  - Post-translational processing
**P. pastoris**

Expression system

**Figure 1 - High Biomass of**

*Pichia pastoris*

AOX1: strong expression

AOX2: weak expression

*S. cerevisiae  P. pastoris*
Pichia expression tools

- **Promoters**
  - AOX1, GAP

- **Selection marker**
  - HIS4, ARG4, ZeocinR, BlasticidinR, KanamycinR (G418)

- **Signal sequences**
  - PHO1, alpha-Factor

- **Host strains**
  - X-33 (wt), GS115 (his4), KM 71 (aox1::arg4 his4),
  - KM7IH (aox1::arg4), SMD1168 (pep4 his4), SMD1168H (pep4)
  - CBS 7435 (WZ or Δaox1 or Δhis4 knockouts)
Integration in *Pichia pastoris*

Gene replacement at AOX1
phenotype: Mut$_S$

Single cross-over integration of circular molecules
*AOX1* (5′ and 3′ regions)
*HIS4*
*GAP*

Tandem repeat multicopy integration

Ectopic integration events
**Integration vector for *Pichia pastoris***

**Gene Replacement**

*Figure 7.5*  *P. pastoris* integrating expression vector. The HBsAg gene is cloned between the promoter (AOX1p) and termination-polyadenylation sequence (AOX1t) of the *P. pastoris* alcohol oxidase 1 gene. The HIS4 gene encodes histidinol dehydrogenase, which is an enzyme in the histidine biosynthesis pathway. An origin of replication from *P. pastoris* (ori^P^) is included, as are both the ampicillin resistance gene (Amp^r^) and an origin of replication (ori^E^) that function in *E. coli*. The segment marked 3'-AOX1 is a piece of DNA from the 3' end of the alcohol oxidase 1 gene of *P. pastoris*. The joined right-angled arrows indicate the DNA region that will be integrated into the *P. pastoris* genome.

*Figure 7.6* Integration of part of an expression vector into the alcohol oxidase 1 gene of *P. pastoris*. The double crossover event occurs within the AOX1p and 3'-AOX1 DNA segments (shown at the top). This event results in the integration of the input DNA into the genomic DNA and the loss of most of the alcohol oxidase 1 gene (AOX1) from the host chromosome (shown at the bottom). The HIS4 gene product enables cells with integrated DNA to grow on medium lacking histidine. In the presence of methanol, the AOX1p region drives the transcription of the HBsAg gene. The AOX1t segment provides transcription termination and polyadenylation signals for the HBsAg gene.
Single-site Integration
Vector for Intracellular Expression

pHIL-D2
8.2 kb

http://tools.invitrogen.com/content/sfs/manuals/pich_man.pdf

Comments for pHIL-D2:
8209 nucleotides

5' AOX1 promoter fragment: bases 14-941
5' AOX1 primer site: bases 868-888
EcoRI Site: bases 956-961
3' AOX1 primer site: bases 1036-1056
3' AOX1 transcription termination (TT) fragment: bases 963-1295
HIS4 ORF: bases 4223-1689
3' AOX1 fragment: bases 4578-5334
Ampicillin resistance gene: bases 5686-6546
f1 origin of replication: bases 7043-6588
pBR322 origin: bases 7138-7757
Linearized plasmid

Pichia genome (his4)

Plasmid integrated into genome
Vector for Intracellular Expression

For pHIL-D2, the fragment containing the gene of interest should have a Kozak consensus sequence for proper translation initiation, although this requirement is not as stringent in yeast. For example, **ACC ATG G** is a Kozak consensus sequence, where the ATG corresponds to the initiating ATG for your gene of interest (Cavener and Stuart, 1991; Kozak, 1987; Kozak, 1990).

Shorter, 5′ untranslated leaders reportedly work better in AOX1 expression. In pHIL-D2, make the untranslated region as short as possible when cloning your gene.
Vector for Secretory Expression

pPIC9
8.0 kb

Comments for pPIC9:
8023 nucleotides

5′ AOX1 promoter fragment: bases 1-948
5′ AOX1 primer site: bases 855-875
α-Factor secretion signal(s): bases 949-1215
α-Factor primer site: bases 1152-1172
Multiple Cloning Site: bases 1192-1241
3′ AOX1 primer site: bases 1327-1347
3′ AOX1 transcription
termination (TT): bases 1253-1586
HIS4 ORF: bases 4514-1980
3′ AOX1 fragment: bases 4870-5626
pBR322 origin: bases 6708-6034
Ampicillin resistance gene: bases 7713-6853
Vector for Secretory Expression

**Signal sequence**

\[ \text{GTT TTA TTC GCA GCA TCC TCC GCA TTA GCT} \]
\[ \text{GCT CCA GTC ACO ACT ACA ACA GAA GAT Val Leu Phe Ala Ala Ser Ser Ala Leu Ala} \]

**Pro-sequence**

\[ \text{GAA ACG GCA CAA ATT CCG GCT GAA GCT GTC ATC GCT TAC TCA GAT TTA GAA GGG GAT Glu Thr Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp} \]

**\( \alpha \)-Fator (949-1215)**

\[ \text{ATCAAAAAAC AACTAATTAT TCGAAGGATC CAAACG ATG AGA TTT CCT TCA ATT TTT ACT GCA Met Arg Phe Pro Ser Ile Phe Thr Ala} \]

**Ste13**

\[ \text{ACATGACTGT TCCTCAGTT GAGTTGGGCA CTTAGAGAAA AACCGGTTCTT GCTAGATTCT AATCAAGAGG} \]

**3' \( AOX1 \) Primer Site (1327-1347)**

\[ \text{ATGTCAGAAT GCCATTTGCC TGAGAGATGC AGGCCTCATT TTGATACCT TTTTATTTGT AACCTATATA} \]

**\( \alpha \)-Factor Primer Site (1152-1172)**

\[ \text{AAT ACT ACT ATT GCC AGC ATC GCT GCT AAA GAA GAA GGG GTA TCT CTC GAG AAA AGA Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Ala Ile} \]

**\( \alpha \)-OX1 Primer Site (855-875)**

\[ \text{TTC GAT GCT GTT GTT TCG CCA TTT TCC AAC AGC ACA AAT AAC GGG TTA TTG TTT ATA Phe Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Gly Leu Leu Phe Ile} \]

**\( Kex2 \)**

\[ \text{GAG GCT GAA GCT TAC GTA GAA TTC CCT AGG GCC GCC GCG AAT TAA TCGCCTTAG Glu Ala Glu Ala Tyr Val Glu Phe Pro Arg Ala Ala Ala Asn ***} \]

**Not1**

\[ \text{GAG GCT GAA GCT TAC GTA GAA TTC CCT AGG GCC GCC GCG AAT TAA TCGCCTTAG Glu Ala Glu Ala Tyr Val Glu Phe Pro Arg Ala Ala Ala Asn ***} \]

**EcoRI**

\[ \text{GAG GCT GAA GCT TAC GTA GAA TTC CCT AGG GCC GCC GCG AAT TAA TCGCCTTAG Glu Ala Glu Ala Tyr Val Glu Phe Pro Arg Ala Ala Ala Asn ***} \]

**NsiI**

\[ \text{GAG GCT GAA GCT TAC GTA GAA TTC CCT AGG GCC GCC GCG AAT TAA TCGCCTTAG Glu Ala Glu Ala Tyr Val Glu Phe Pro Arg Ala Ala Ala Asn ***} \]
Resistance selection in *Pichia pastoris*, multiple integration and secretion

- $P_{GAP}$
- $AOX1$ TT
- Zeo$^R$
- C-myc Epitope
- 6xHis
- alpha-factor
- ColE1 ori
- Multicopy Integration “in vivo”
Comments for pPICZα A
3593 nucleotides

<table>
<thead>
<tr>
<th>Feature</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′ AOXI</td>
<td>A 942 bp fragment containing the AOXI promoter that allows methanol-inducible, high-level expression in Pichia. Targets plasmid integration to the AOXI locus.</td>
</tr>
<tr>
<td>Native Saccharomyces cerevisiae α-factor secretion signal</td>
<td>Allows for efficient secretion of most proteins from Pichia.</td>
</tr>
<tr>
<td>Multiple cloning site with 10 unique restriction sites</td>
<td>Allows insertion of your gene into the expression vector</td>
</tr>
<tr>
<td>C-terminal myc epitope tag</td>
<td>(Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu-Asn) Permits detection of the fusion protein by the Anti-myc Antibody or Anti-myc-HRP Antibody (see page viii for ordering information) (Evan et al., 1985)</td>
</tr>
<tr>
<td>C-terminal polyhistidine tag</td>
<td>Permits purification of your recombinant fusion protein on metal-chelating resin such as ProBond™. In addition, the C-terminal polyhistidine tag is the epitope for the Anti-His(C-term) Antibody and the Anti-His(C-term)-HRP Antibody (see page viii for ordering information) (Lindner et al., 1997)</td>
</tr>
<tr>
<td>AOXI Transcription Termination (TT)</td>
<td>Native transcription termination and polyadenylation signal from AOXI gene (~260 bp) that permits efficient 3′ mRNA processing, including polyadenylation, for increased mRNA stability</td>
</tr>
<tr>
<td>TEF1 promoter</td>
<td>Transcription elongation factor 1 gene promoter from Saccharomyces cerevisiae that drives expression of the Sh ble gene in Pichia, conferring Zeocin™ resistance (GenBank Acc. no. D12478, D01130)</td>
</tr>
<tr>
<td>EM7 (synthetic prokaryotic promoter)</td>
<td>Constitutive promoter that drives expression of the Sh ble gene in E. coli, conferring Zeocin™ resistance</td>
</tr>
<tr>
<td>Sh ble gene (Streptococcaceae hindustanus ble gene)</td>
<td>Zeocin™ resistance gene</td>
</tr>
<tr>
<td>CYC1 transcription termination region</td>
<td>3′ end of the Saccharomyces cerevisiae CYC1 gene that allows efficient 3′ mRNA processing of the Sh ble gene for increased stability (GenBank Acc. no. M34014)</td>
</tr>
<tr>
<td>pUC origin</td>
<td>Allows replication and maintenance of the plasmid in E. coli</td>
</tr>
<tr>
<td>Sac I, Pme I, BsuX I</td>
<td>Unique restriction sites that permit linearization of the vectors at the AOX1 locus for efficient integration into the Pichia genome</td>
</tr>
</tbody>
</table>

5′ AOX1 promoter region: bases 1-941
5′ AOX1 priming site: bases 855-875
α-factor signal sequence: bases 941-1207
α-factor priming site: bases 1144-1164
Multiple cloning site: bases 1208-1276
c-myc epitope: bases 1275-1304
Polyhistidine (6xHis) tag: bases 1320-1337
Sh ble ORF: bases 2163-2537
CYC1 transcription termination region: bases 2538-2855
pUC origin: bases 2866-3539 (complementary strand)
Zeocin

MW = 1,535
**P.pastoris vectors for intracellular expression**

Gene of interest

<table>
<thead>
<tr>
<th>EcoRI</th>
<th>NotI (7101)</th>
</tr>
</thead>
</table>

`...GAATTC TTGAAACG ATGNNNNN.........TAAGCGGCGGC...`

| NotI |  |
|------| |

`.....AAACAAAGAATTC TTGAAACG AGGCTTGCGGCGGCGG......`

<table>
<thead>
<tr>
<th>Promoter:</th>
<th>A → AOX1</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Selection Marker:</th>
<th>H → HIS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z → Zeocin&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>A → ARG4</td>
<td></td>
</tr>
<tr>
<td>K → Kanamycin&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Restriction site:</th>
<th>Bgl → BglII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sph → SphI</td>
<td></td>
</tr>
<tr>
<td>Swa → SwaI</td>
<td></td>
</tr>
</tbody>
</table>

**pAHBgl** 7111 bp

- bla (ApR)
- P Aox1
- 3' UTR Aox1
- Ori pMB1
- ARG4 TT
- His4
- Ori pMB1
- Bgl II (6153)
- Sac I (6360)
- Pst I (3136)
- Hin dIII (7024)
- Nde I (598)
- Not I (7101)
- Eco RI (7078)
**P. pastoris** vectors for secretory expression

Kex2 processing signal

Gene of interest

---

Promoter:

\[ A \rightarrow AOX1 \]

Secretion Signal:

\[ a \rightarrow \text{alpha factor} \]

Selection Marker:

- \( H \rightarrow \text{HIS4} \)
- \( Z \rightarrow \text{Zeocin}^R \)
- \( A \rightarrow \text{ARG4} \)
- \( K \rightarrow \text{Kanamycin}^R \)

Restriction site:

- \( Bgl \rightarrow \text{BglII} \)
- \( Sph \rightarrow \text{Spel} \)
- \( Swa \rightarrow \text{SwaI} \)
Vectors for multiple Integrations

secretory

intracellular
<table>
<thead>
<tr>
<th>Protein expressed</th>
<th>Expression Level (mg/L)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Yeast proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Plant proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Invertebrate proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mammalian proteins</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Baculovirus Expression System

Figure 7.11 Production of recombinant baculovirus. Single Bsr36I sites are engineered into gene 603 and a gene (ORF1629) that is essential for baculovirus replication in insect cells (essential gene) of the AcMNPV genome. These genes flank the polyhedrin gene of the AcMNPV genome. After a baculovirus with two engineered Bsr36I sites is treated with Bsr36I, the segment between the Bsr36I sites is deleted. After transfection of an insect cell carrying a Bsr36I-treated baculovirus with a transfer vector that has a cloned gene under the control of the promoter (p) and terminator (t) elements of the polyhedrin gene and the complete sequence of both gene 603 and the essential gene, a double crossover event (dashed lines) generates a recombinant baculovirus with a functional essential gene. This system produces up to 99% recombinant baculoviruses.
Baculovirus Expression System

1. **Transformation**
   - pFastBac™ donor plasmid
   - Clone Gene of Interest
   - Recombinant Donor Plasmid

2. **DAY 1**
   - Transposition
   - Antibiotic Selection
   - E. coli (LacZ⁺) containing Recombinant Bacmid

3. **DAY 2–3**
   - Mini-prep of high molecular weight DNA

4. **DAY 4**
   - Transfection of Insect Cells with Cellfectin® Reagent
   - Recombinant Bacmid DNA

5. **DAY 5–7**
   - Recombinant Baculovirus Particles
   - Infection of Insect Cells

6. **Determine Viral Titer by Plaque Assay**

7. **Recombinant Gene Expression or Viral Amplification**
Mammalian Expression System

**Figure 7.13** Generalized mammalian expression vector. The multiple cloning site (mcs) and selectable marker gene (sm) are under the control of eukaryotic promoter (p) and polyadenylation (pa) sequences. Propagation of the vector in *E. coli* and mammalian cells depends on the origins of replication, ori*E* and ori*euk*, respectively. A marker gene (Amp*) can be used for selecting transformed *E. coli*.

- **E.coli repicon**
- **Eukaryotic replication system → e.g. viral systems**
- **selection marker for eukaryote**
- **Expression signals**
Mammalian Expression System

Simple Plasmid for ectopic integration
Fig. 30. Recombinant retroviral genomes in packaging cell lines. The factors required to rescue defective viral genomes (retroviral vector) are supplemented in turn. They symbolically represent essential components for viral expression.
Expression strategies

DHFR:
Selection for high expression with methotrexate
Simultaneous expression of two polypeptides

Figure 7.15  Two-vector expression system. The cloned genes (gene α and gene β) encode subunits of a protein dimer (αβ). After cotransfection, both subunits (α and β) are synthesized and assembled into a functional protein dimer. Both vectors carry origins of replication for E. coli (oriB) and mammalian cells (oriEuk), a marker gene (AmpR) for selecting transformed E. coli; and eukaryotic promoter (p) and polyadenylation (poly) sequences that control a selectable marker gene (sm) and each of the cloned genes (gene α and gene β).
Simultaneous expression of two polypeptides

Figure 7.16  Two-gene expression vector. The cloned genes (gene α and gene β) encode subunits of a protein dimer (αβ). Each cloned gene is inserted into a vector as part of its own transcription unit under the control of a eukaryotic promoter (p) and polyadenylation sequence (pa). Each subunit is translated from a separate mRNA, and a functional protein dimer (αβ) is assembled. Each vector carries origins of replication for E. coli (oriE) and mammalian cells (oriEuk), a marker gene (Amp') for selecting transformed E. coli, and a selectable marker gene (sm) that is under the control of a eukaryotic promoter (p) and sequence polyadenylation (pa):
Simultaneous expression of two polypeptides

IRES:
Internal Ribosome Entry Site