MOL.911
Molecular Biotechnology I

Cloning and Expression

Recombinant DNA Technology (Cloning)

Construct proper Vectors

Creating defined Gene Elements – DNA Fragments

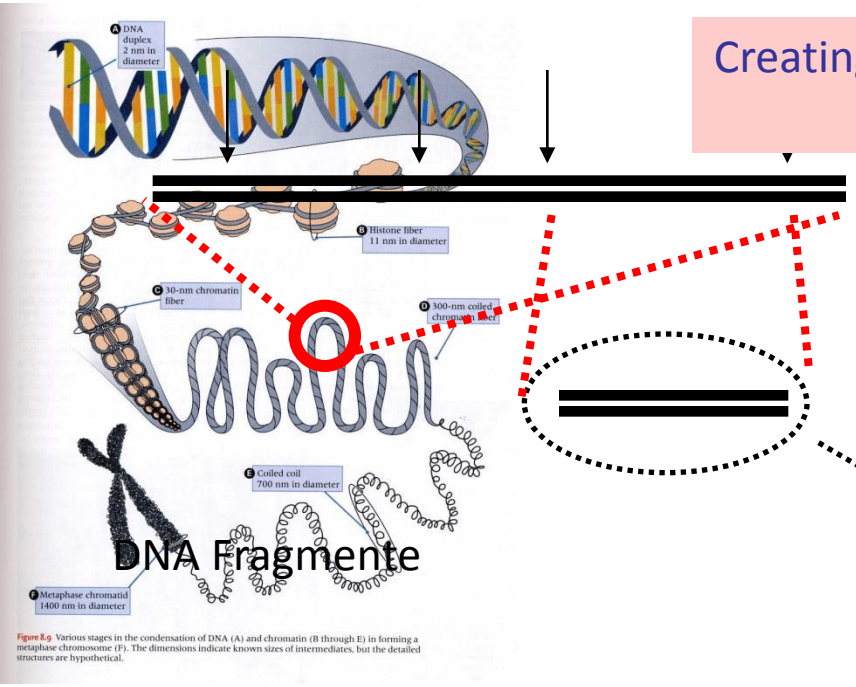
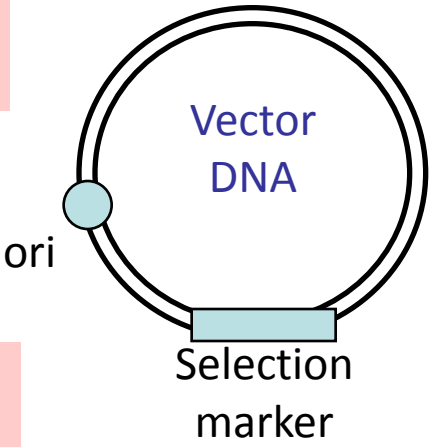
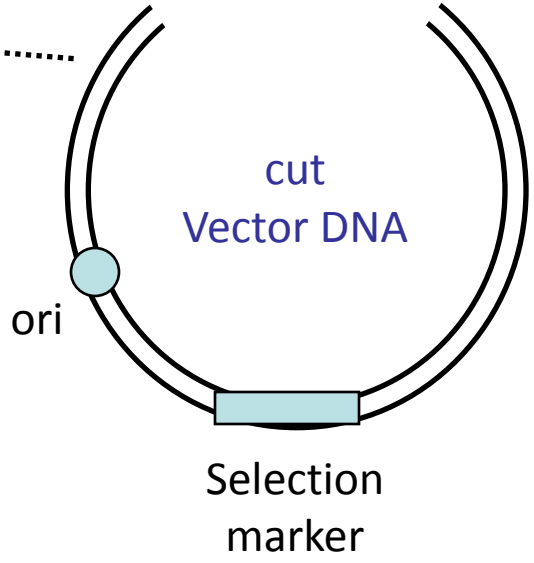


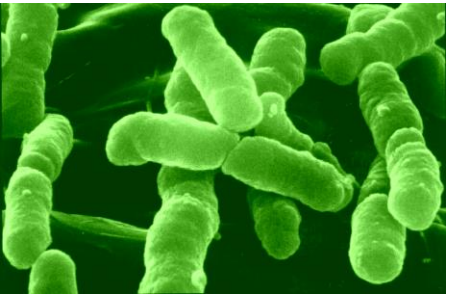
Figure 8.9 Various stages in the condensation of DNA (A) and chromatin (B through E) in forming a metaphase chromosome (F). The dimensions indicate known sizes of intermediates, but the detailed structures are hypothetical.

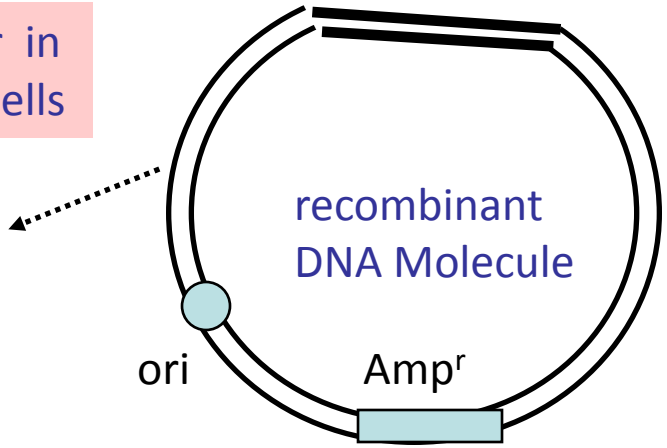


Combine Fragments with Vector

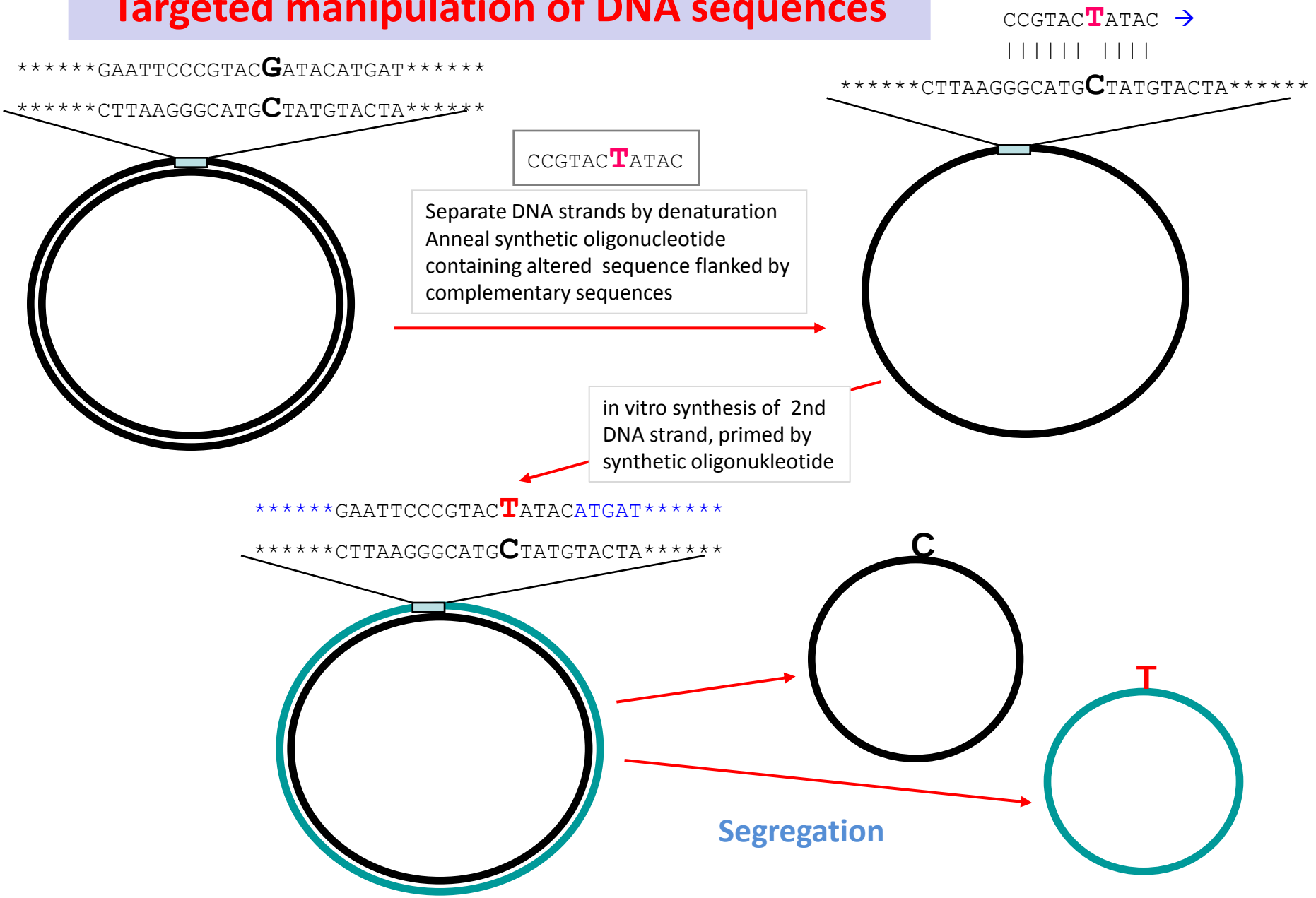


Transfer in Living Cells



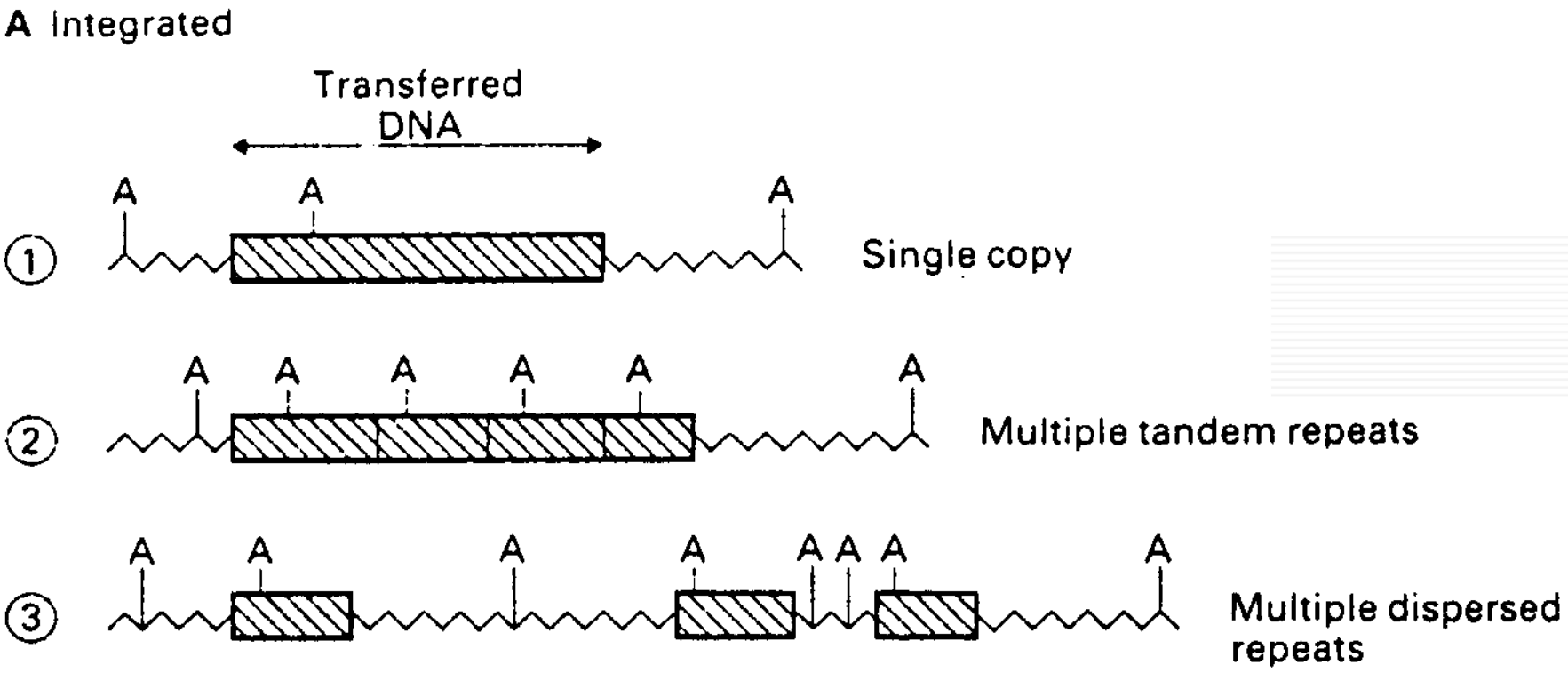


Targeted manipulation of DNA sequences



Autonomous Replication → Plasmids Viruses ARS

Integration into genome Site specific -- Ectopic



Vectors for Recombinant DNA Technology

Plasmids
Autonomous Replication
Integration into genome
Shuttle Plasmids
E.coli → Target host

Cosmids, Bacmids
Plasmid – Bacteriophage Hybrids

Phages
Bacteriophage Lambda

Artificial Chromosomes
YAC

Viruses
Baculovirus – Insect Cells
Retroviruses – Mammalian Cells

Integration – general steps

Gene Replacement (A)

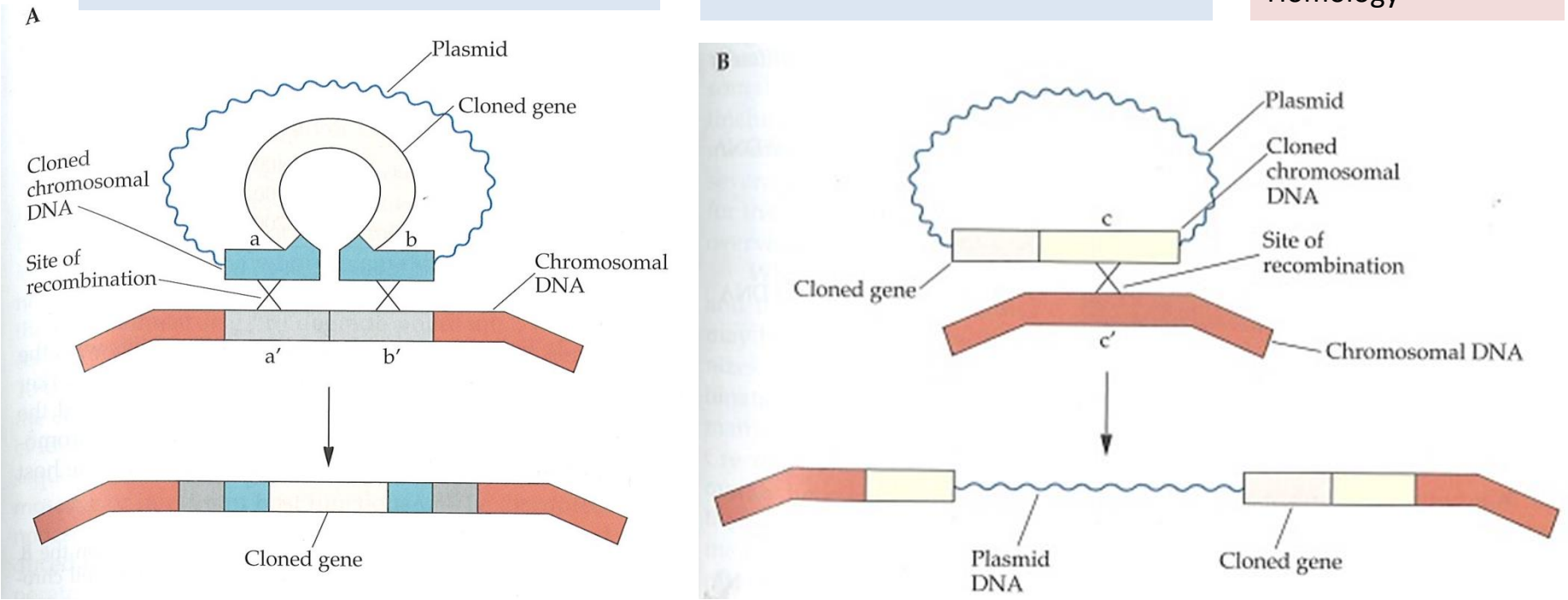
→ Double Cross-over at regions showing sufficient homology
Linearized DNA

Site specific Insertion (B)

→ → Single Cross-over at regions showing sufficient homology

Ectopic Integration

→ Recombination at regions of no (low ?) Homology



Integration of a cloned gene into a chromosomal site. **(A)** The cloned gene has been inserted, on a plasmid, in the middle of the cloned segment of DNA (ab) from the host chromosome. Homologous DNA pairing occurs between plasmid-borne DNA regions a and b and host chromosome DNA regions a´ and b´, respectively. A double cross-over event (x—x) results in the integration of the cloned gene. **(B)** The cloned gene is inserted adjacent to the cloned DNA from the host chromosome (c). Homologous DNA pairing occurs between plasmid DNA region c and host chromosome DNA region c´. A single recombination event (x) within the paired c-c´ DNA region results in the integration of the entire plasmid, including the cloned gene.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Plasmid vectors for bacteria

Bacterial plasmids

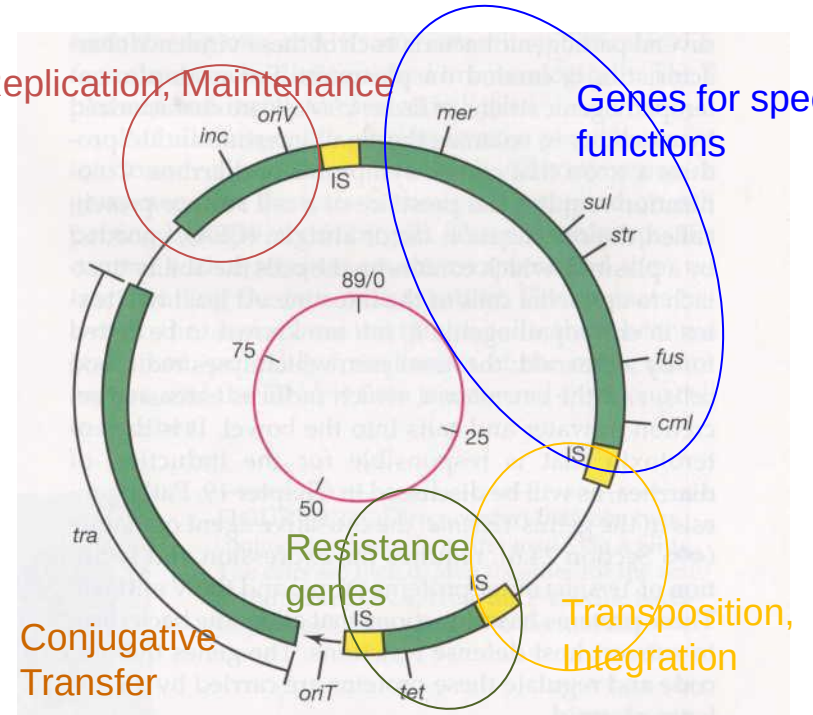


FIGURE 9.21 Genetic map of the resistance plasmid R100. The inner circle shows the size of the plasmid in kilobase pairs. The outer circle shows the location of major antibiotic resistance genes and other key functions: *inc*, incompatibility genes; *oriV*, origin of replication site; *oriT*, origin of conjugative transfer; *mer*, mercuric ion resistance; *sul*, sulfonamide resistance; *tet*, tetracycline resistance; *tra*, transfer functions. The locations of insertion sequences (IS) are also shown.

Replication
- Origin of replication (*oriV*)
- Regulatory functions for replication (*rep, trf*)

→ copy number
→ Host range
→ incompatibility

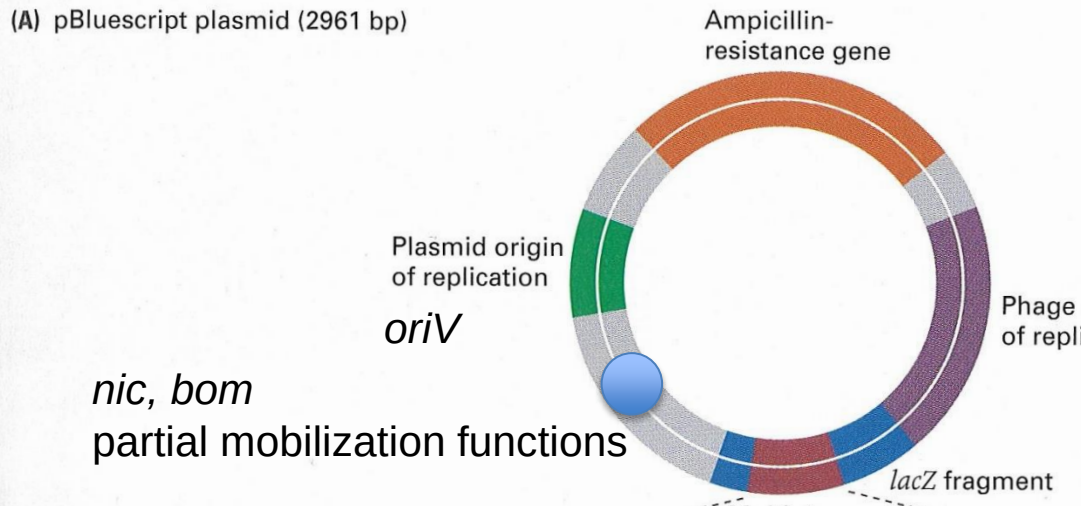
Maintenance
- Partitioning systems (*par*)
- Multimer resolution systems (*mrs*)
- Addiction systems (e.g. *hok-sok*)

→ Stable maintenance of plasmids upon cell division

Conjugative transfer
- Complete Transfer regions (*tra*)
- Mobilization regons (*mob, oriT, nic, bom*)

→ Autonomous In vivo transfer of plasmids
→ In vivo transfer mediated by helper functions

Plasmid vector



nic, bom
partial mobilization functions

ss-specific replication
rolling circle mode

Screening for insert (blue-white)

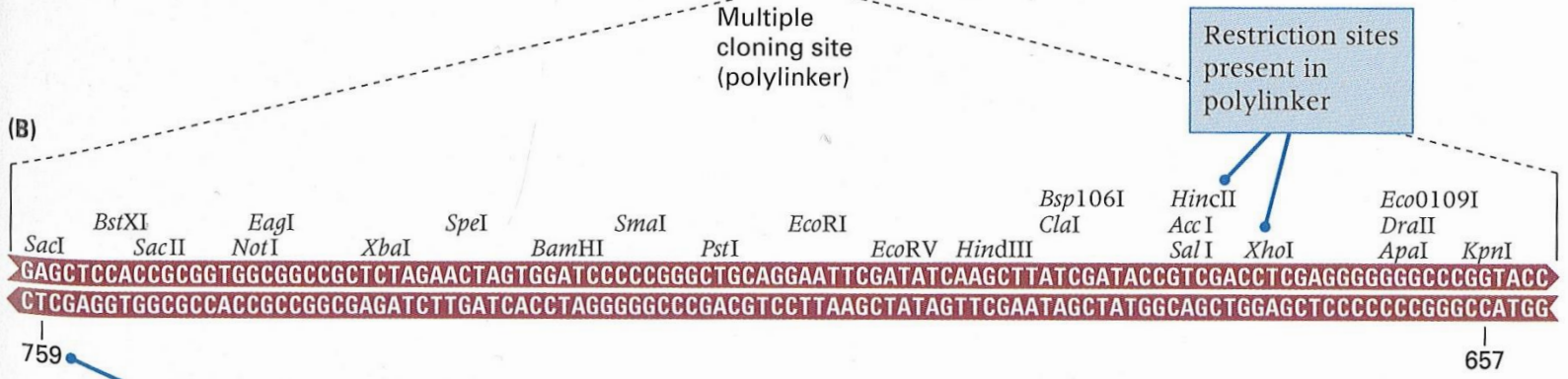


Figure 12.9 (A) Diagram of the cloning vector pBluescript II. It contains a plasmid origin of replication, an ampicillin-resistance gene, a multiple cloning site (polylinker) within a fragment of the *lacZ* gene from *E. coli*, and a bacteriophage origin of replication. (B) Sequence of the multiple cloning 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.]

Taken from: D.L. Hartl, E.W. Jones; GENETICS Analysis of Genes and Genomes, 6th Ed.; Jones and Bartlett

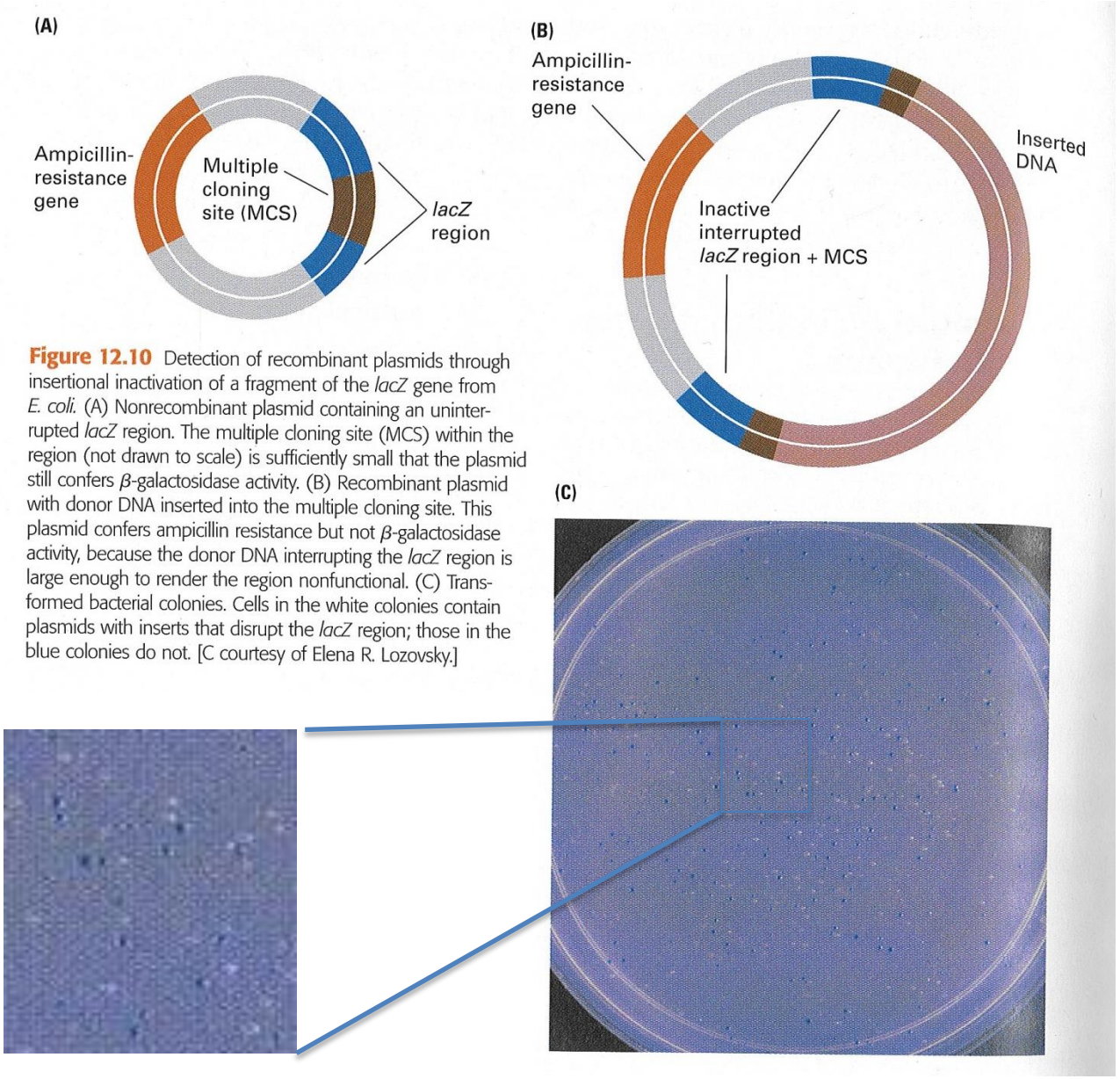


Figure 12.10 Detection of recombinant plasmids through insertional inactivation of a fragment of the *lacZ* gene from *E. coli*. (A) Nonrecombinant plasmid containing an uninterrupted *lacZ* region. The multiple cloning site (MCS) within the region (not drawn to scale) is sufficiently small that the 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.]

Taken from: D.L. Hartl, E.W. Jones; GENETICS Analysis of Genes and Genomes, 6th Ed.; Jones and Bartlett

Bacteriophage Lambda Vectors

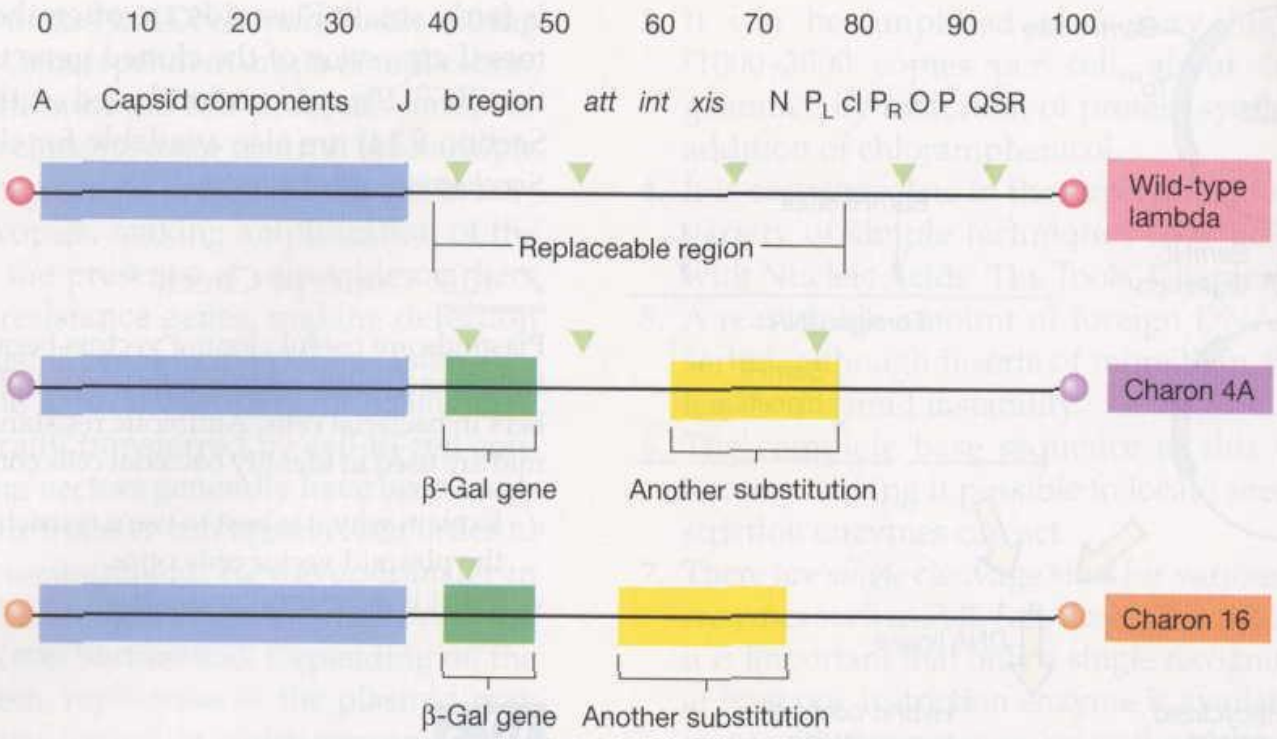


FIGURE 10.3 Molecular cloning with lambda. Abbreviated genetic map of bacteriophage lambda showing the cohesive ends as circles (Figure 8.26). Charon 4A and 16 are both derivatives of lambda, which have various substitutions and deletions in the nonessential region. One of the substitutions in each case 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 *Eco*RI.

Bacteriophage Lambda Vectors

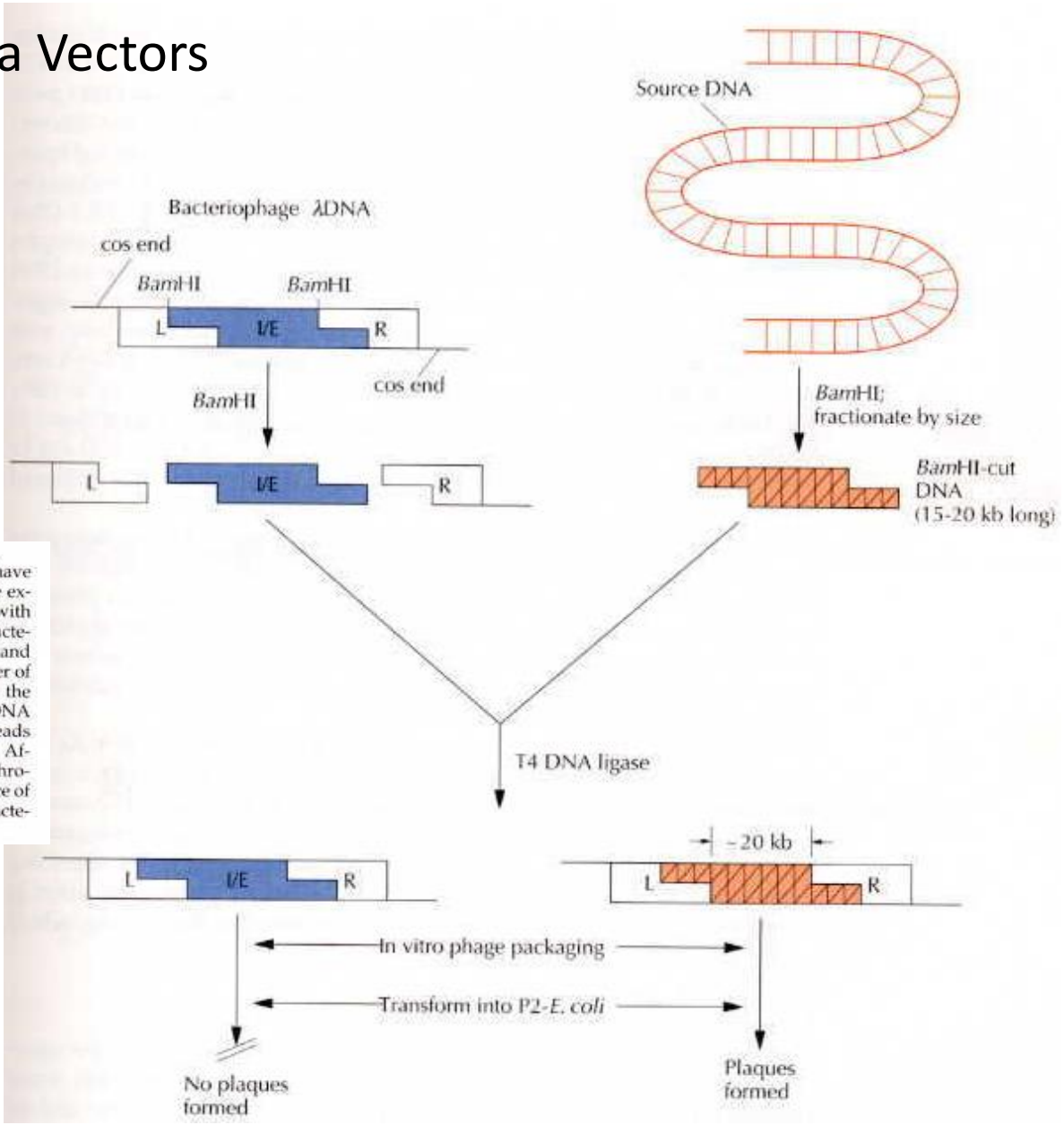


Figure 4.17 Bacteriophage λ cloning system. Bacteriophage λ is engineered to have two *Bam*HI sites that flank the I/E region of the bacteriophage λ genome. The extensions indicate the cos ends of the λ DNA. For cloning, the source DNA is cut with *Bam*HI and fractionated by size to isolate pieces that are 15 to 20 kb long. The bacteriophage λ DNA is also cut with *Bam*HI. The two DNA samples are mixed and treated 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 chromosomes, 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.

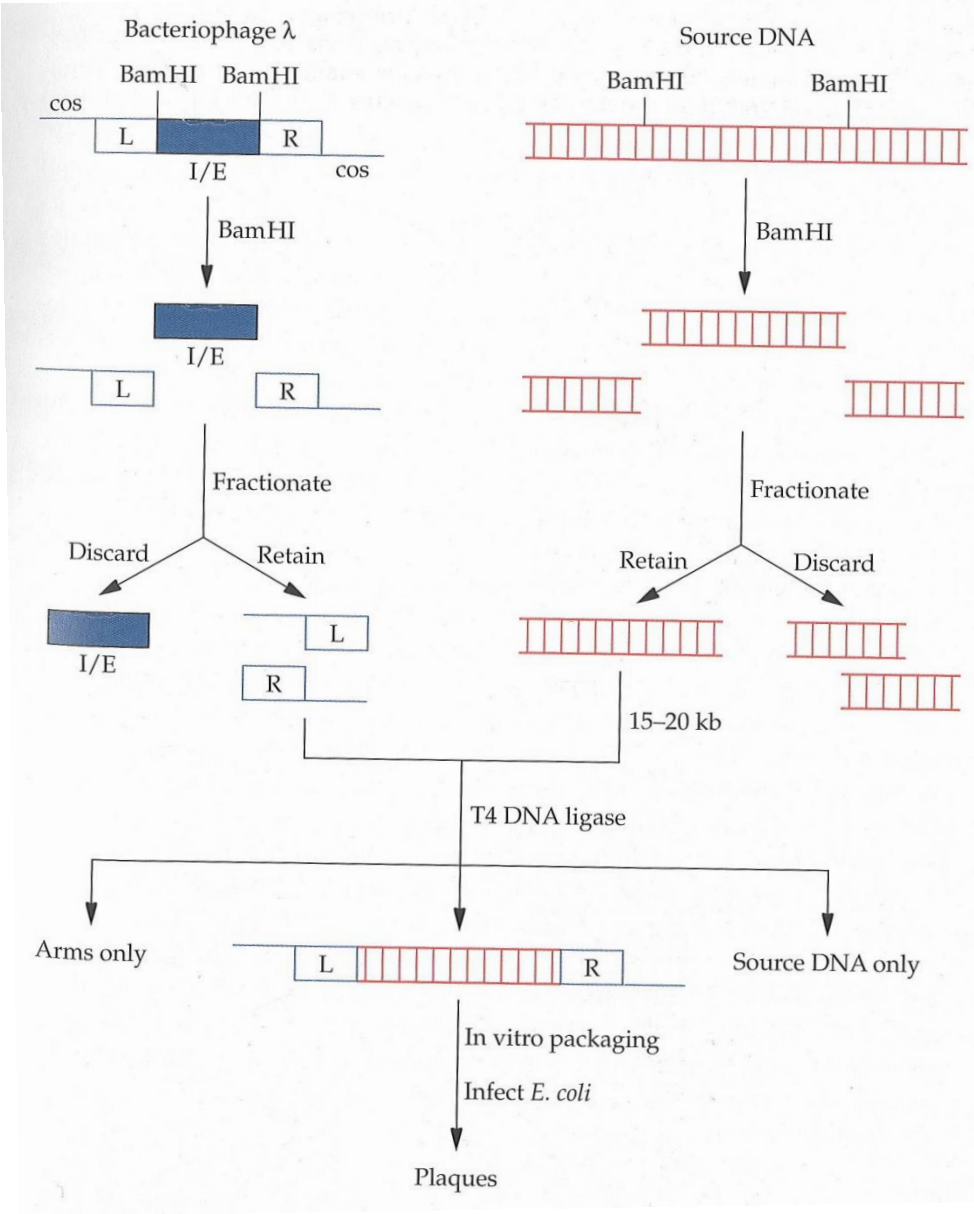
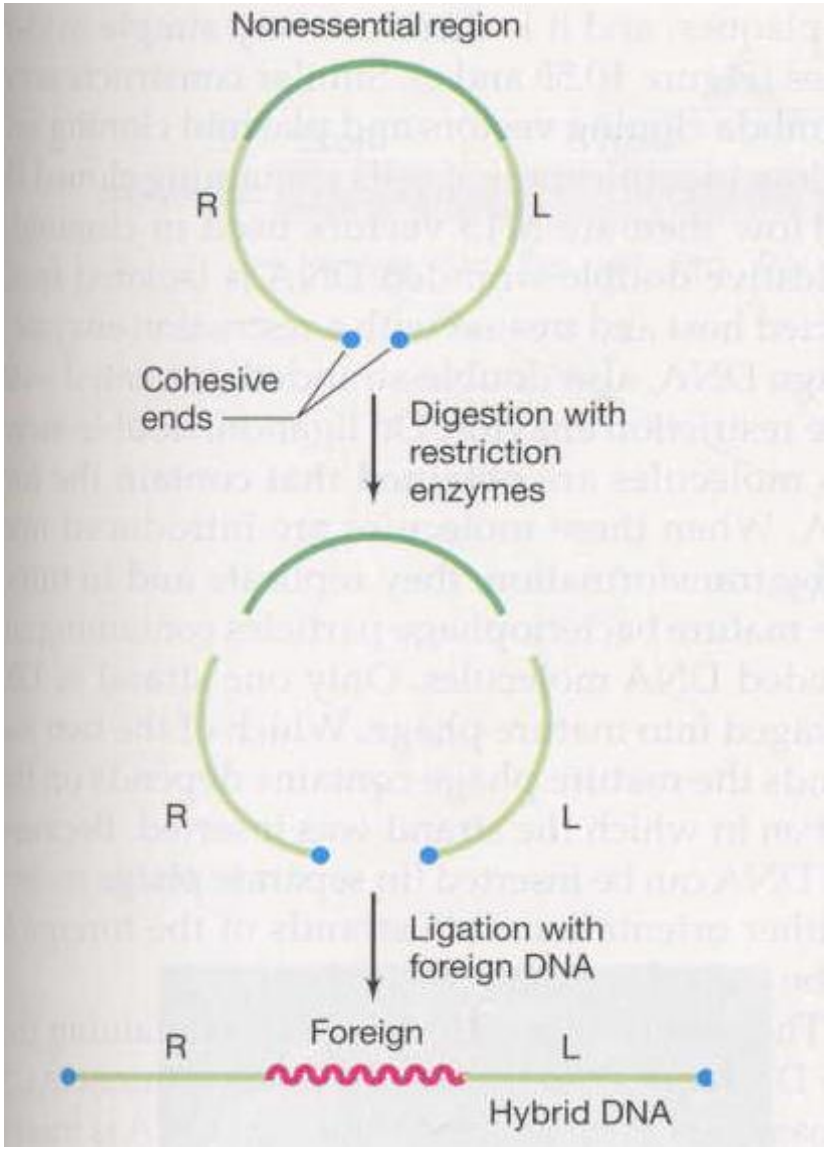


FIGURE 3.31 A bacteriophage λ cloning system. Bacteriophage λ is engineered to have two BamHI sites that flank the I/E region. For cloning, the source DNA is cut with BamHI and fractionated by size to isolate pieces that are about 15 to 20 kb long. The bacteriophage λ DNA is also cut with BamHI, and size fractionation removes the I/E segment. The L and R arms, plus the 15- to 20-kb source DNA molecules, are mixed with T4 DNA ligase. The ligation reaction produces a number of different DNA molecules, including ligated source DNA only, combined L and R arms only, and molecules that have a source DNA molecule flanked by L and R arms. The last molecules are packaged into bacteriophage heads in vitro, and infective particles are formed after the addition of tail assemblies. The recombined bacteriophage λ is perpetuated by infection of *E. coli*. Some 50-kb source DNA ligation products may be packaged into heads, but since this DNA lacks both a functional origin of replication and cos ends, it cannot be perpetuated. Other ligation products are either too small or too large to be packaged. For some bacteriophage λ cloning systems (not shown here), high packaging efficiency is achieved by setting the conditions of ligation to favor concatemer formation to imitate how the phage heads are normally filled.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Cloning in Lambda Vectors



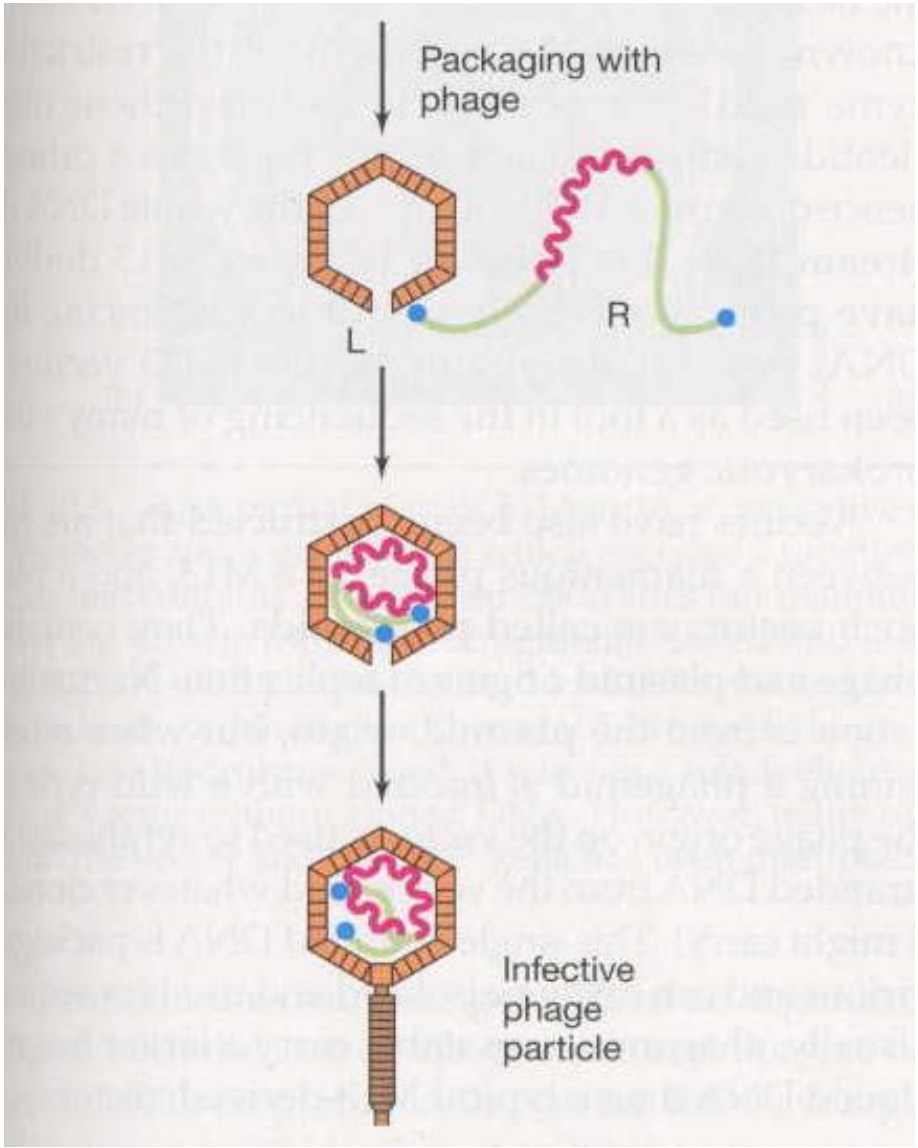
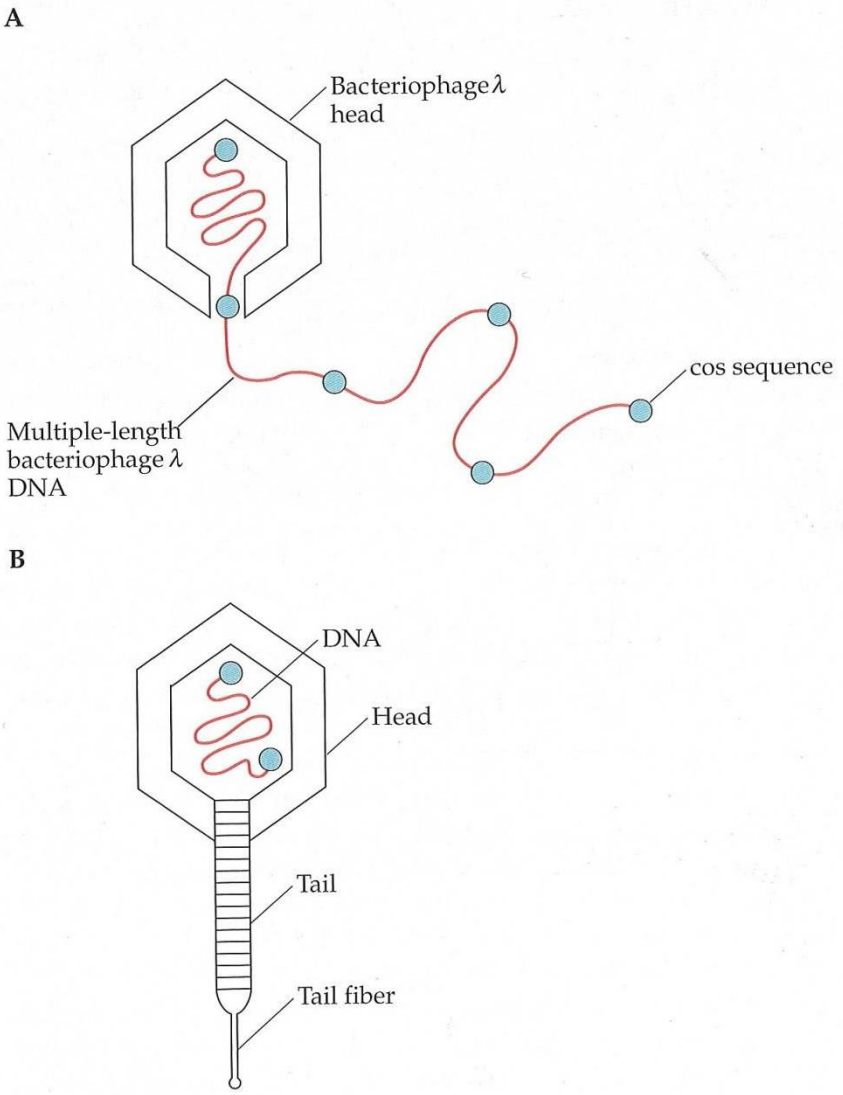


FIGURE 3.30 Packaging of bacteriophage λ DNA into heads during the lytic cycle. **(A)** DNA replication from the circular form of bacteriophage λ creates a linear form that has contiguous, multiple lengths (concatemers) of bacteriophage DNA with units of approximately 50 kb each. **(B)** Each newly assembled head is filled with a 50-kb unit of λ DNA before the tail assembly is attached.



Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Cosmid vectors

Principle: Plasmid DNA Transfer via Phage infection

Resulting recombinant clone Contains self-replicating plasmid

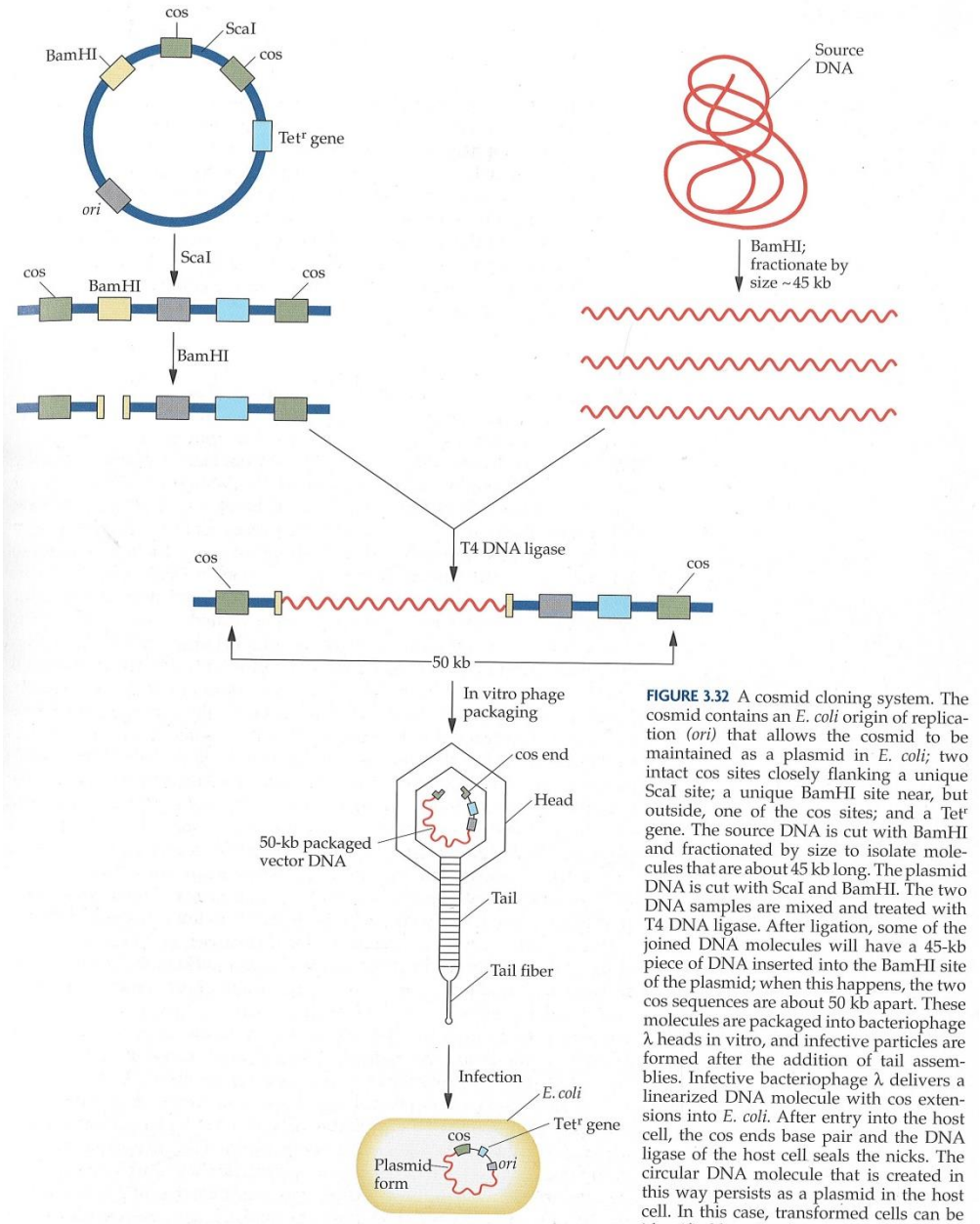
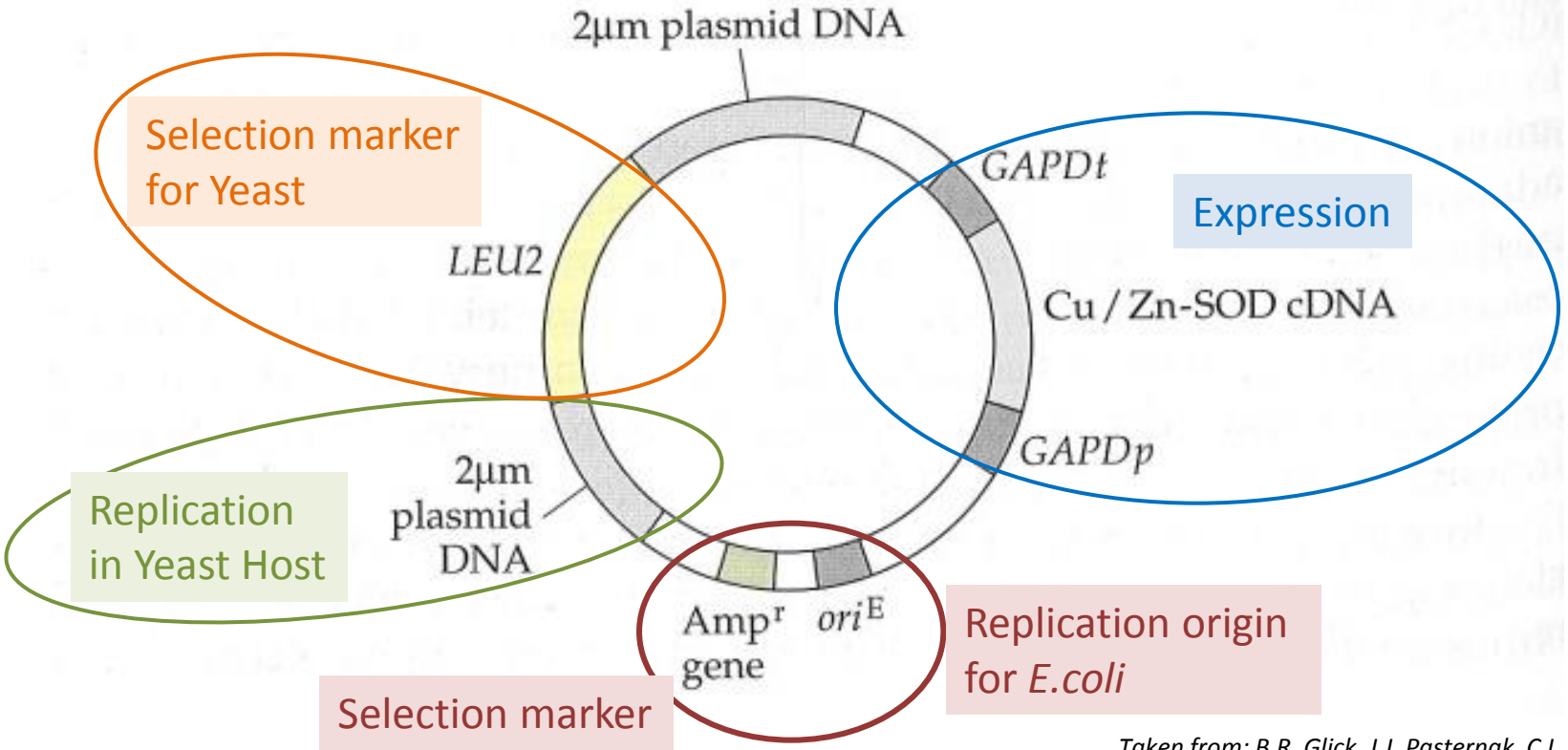


FIGURE 3.32 A cosmid cloning system. The cosmid contains an *E. coli* origin of replication *(ori)* that allows the cosmid to be maintained as a plasmid in *E. coli;* two intact cos sites closely flanking a unique ScaI site; a unique BamHI site near, but outside, one of the cos sites; and a Tet^r gene. The source DNA is cut with BamHI and fractionated by size to isolate molecules that are about 45 kb long. The plasmid DNA is cut with ScaI and BamHI. The two DNA samples are mixed and treated with T4 DNA ligase. After ligation, some of the joined DNA molecules will have a 45-kb piece of DNA inserted into the BamHI site of the plasmid; when this happens, the two cos sequences are about 50 kb apart. These molecules are packaged into bacteriophage λ heads in vitro, and infective particles are formed after the addition of tail assemblies. Infective bacteriophage λ delivers a linearized DNA molecule with cos extensions into *E. coli.* After entry into the host cell, the cos ends base pair and the DNA ligase of the host cell seals the nicks. The circular DNA molecule that is created in this way persists as a plasmid in the host cell. In this case, transformed cells can be identified because they are resistant to the antibiotic tetracycline.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

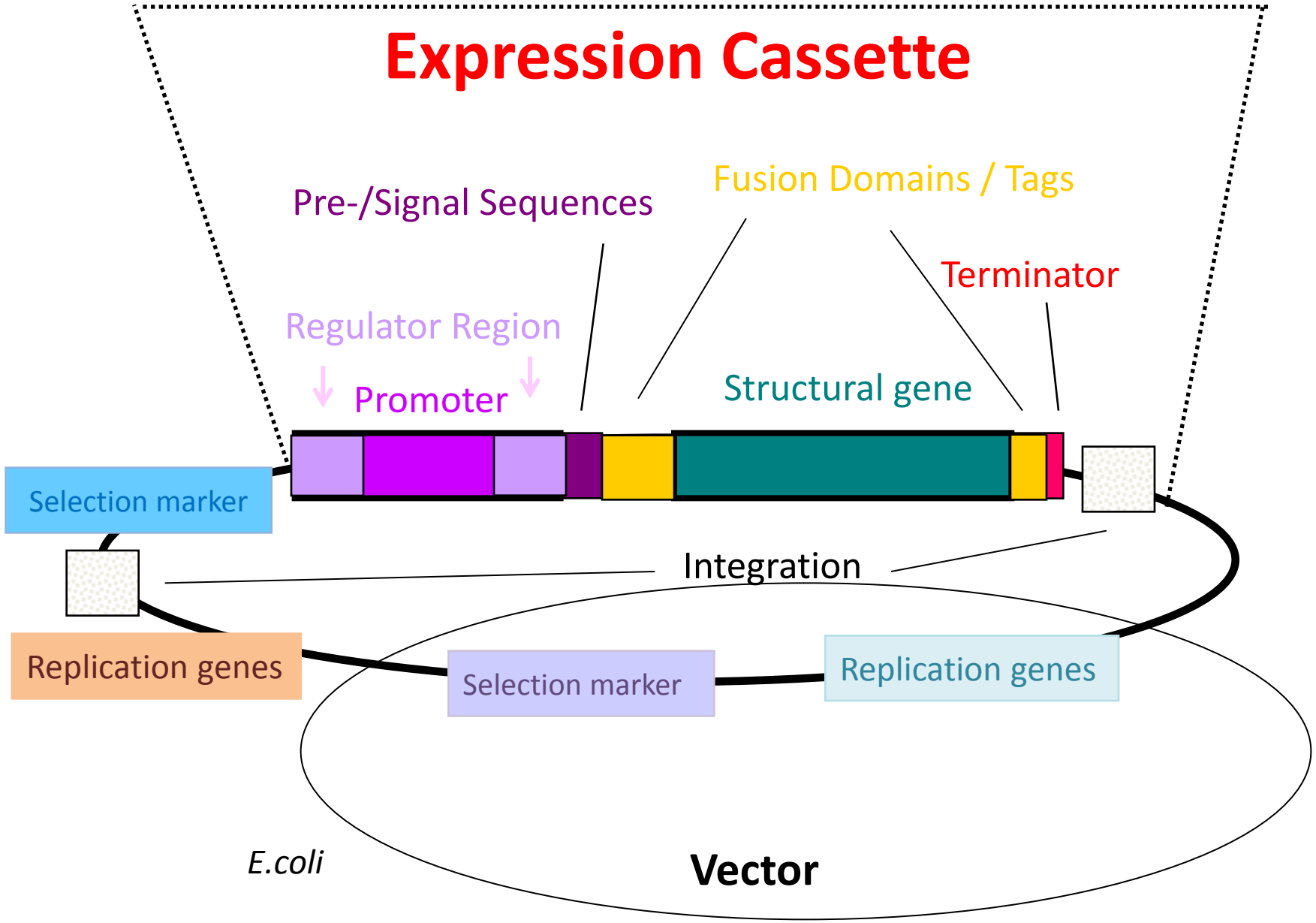
E.coli – Saccharomyces cerevisiae Shuttle Vector

FIGURE 7.7 *S. cerevisiae* expression vector. The cDNA for human Cu/Zn-SOD was cloned between the promoter *(GAPDp)* and termination–polyadenylation sequence *(GAPDt)* of the *S. cerevisiae* glyceraldehyde phosphate dehydrogenase gene. The *LEU2* gene that was cloned between segments of the yeast 2µm plasmid DNA encodes a functional enzyme of the leucine biosynthesis pathway. The yeast origin of replication is included in the 2µm plasmid DNA. The ampicillin resistance (Amp^r) gene and the *E. coli* origin of replication (ori^E) are derived from plasmid pBR322.

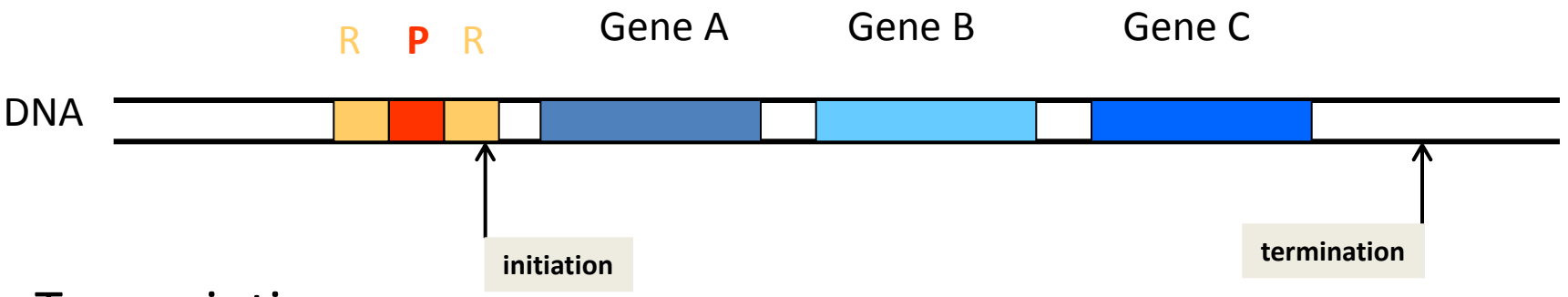


Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

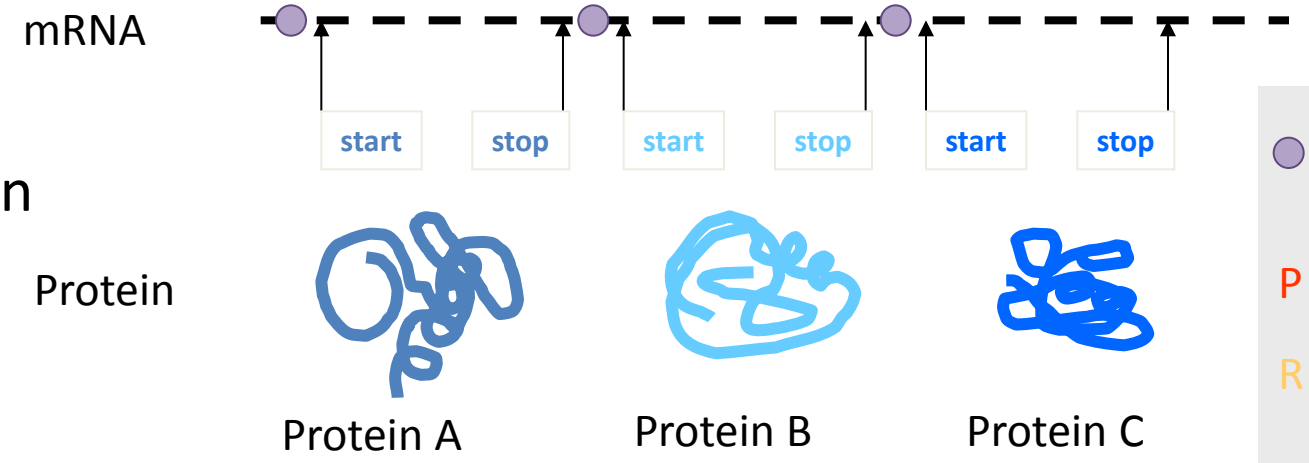
13.10.16


Expression Cassette
Pre-/Signal Sequences
Fusion Domains / Tags
Terminator
Regulator Region
Promoter
Structural gene
Selection marker
Integration
Replication genes
Selection marker
Replication genes
E.coli
Vector

Gene Expression in Prokaryotes



Transcription



Post-translational processing

Gene Expression – Points to consider

Location in Genome → Autonomous replication, Integration

Transcription Initiation → Promoters **Transcription Termination**

Regulatory Systems → positive/negative regulatory systems

Transcript Processing

RNA Structure

mRNA stability

Translation Initiation - Translation elongation

Codon usage

Post-translational modifications

Modification of AA-side chains: Glycosylation, Phosphorylation, etc

Proteolytic Processing

Protein Folding

Disulfide bond formation

Assembly of subunits

Toxicity of gene producs

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, ara*
  - T7, T5, T3
- 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)

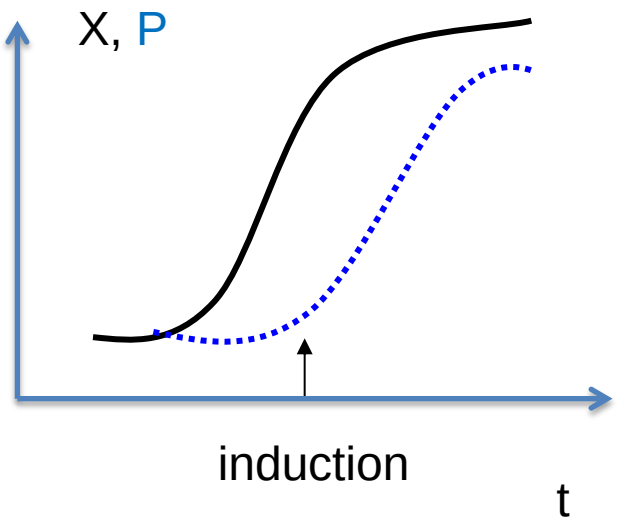
Plasmid copy number and segregation

Regulated Promoters ←→ Constitutive Promoters

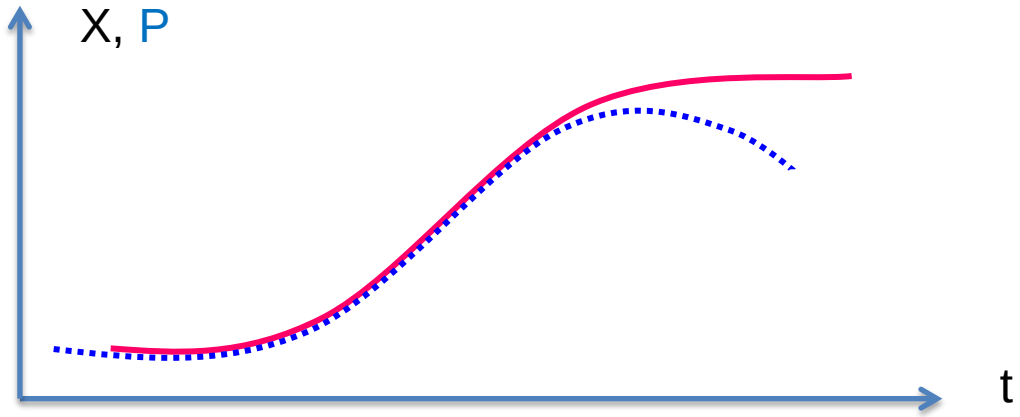
Both systems are used

Preferred Combination: strong Promoters – tightly regulated

Regulated – specific induction



constitutive



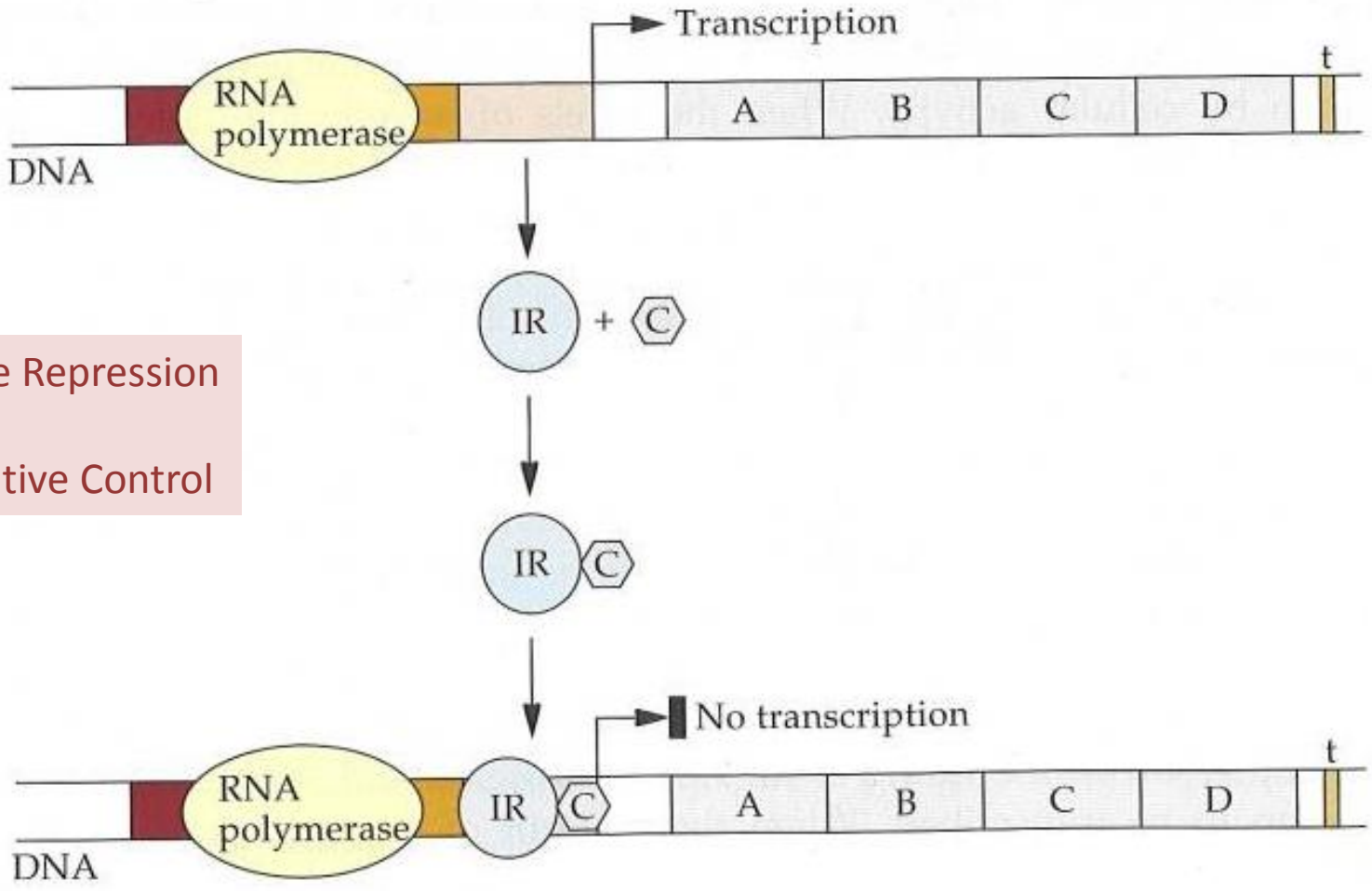
Constitutive promoters: weak to medium activity

Regulated expression:
Separation of growth phase and production phase
→ High specific growth rate in growth phase

Constitutive Expression: growth is impaired due to heavy metabolic load directed to expression of one specific protein
→ Overall low specific grothe rate
→ Lower overall productivity

Regulated Expression in Prokaryotes

FIGURE 2.25 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.



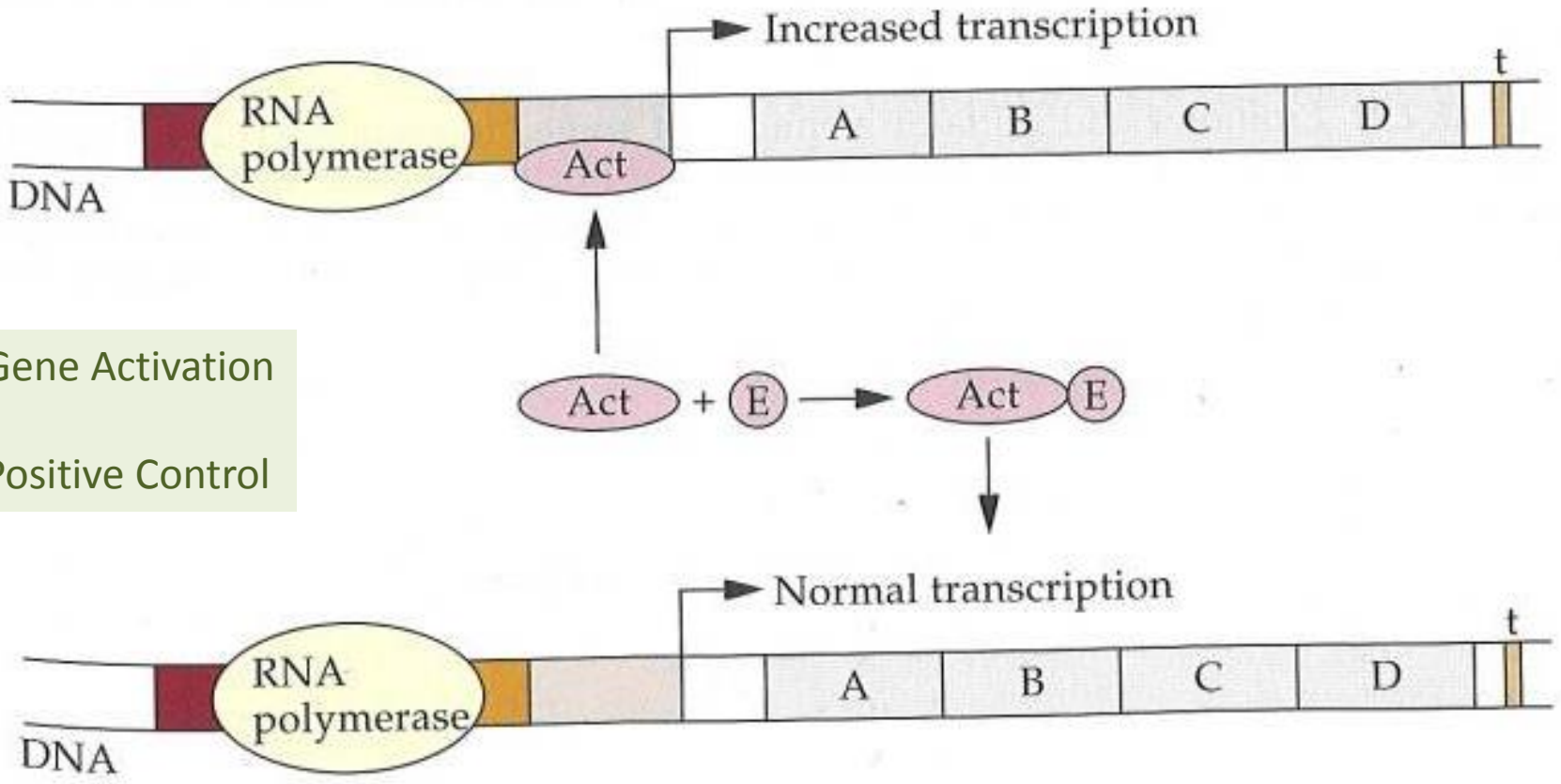
Gene Repression

negative Control

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Regulated Expression in Prokaryotes

FIGURE 2.26 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. When 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.



Gene Activation

Positive Control

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Regulated Expression in Prokaryotes

Complex Initiation System

Multiple regulation elements in concerted action

- Enhancers
- Activators
- Repressing Systems

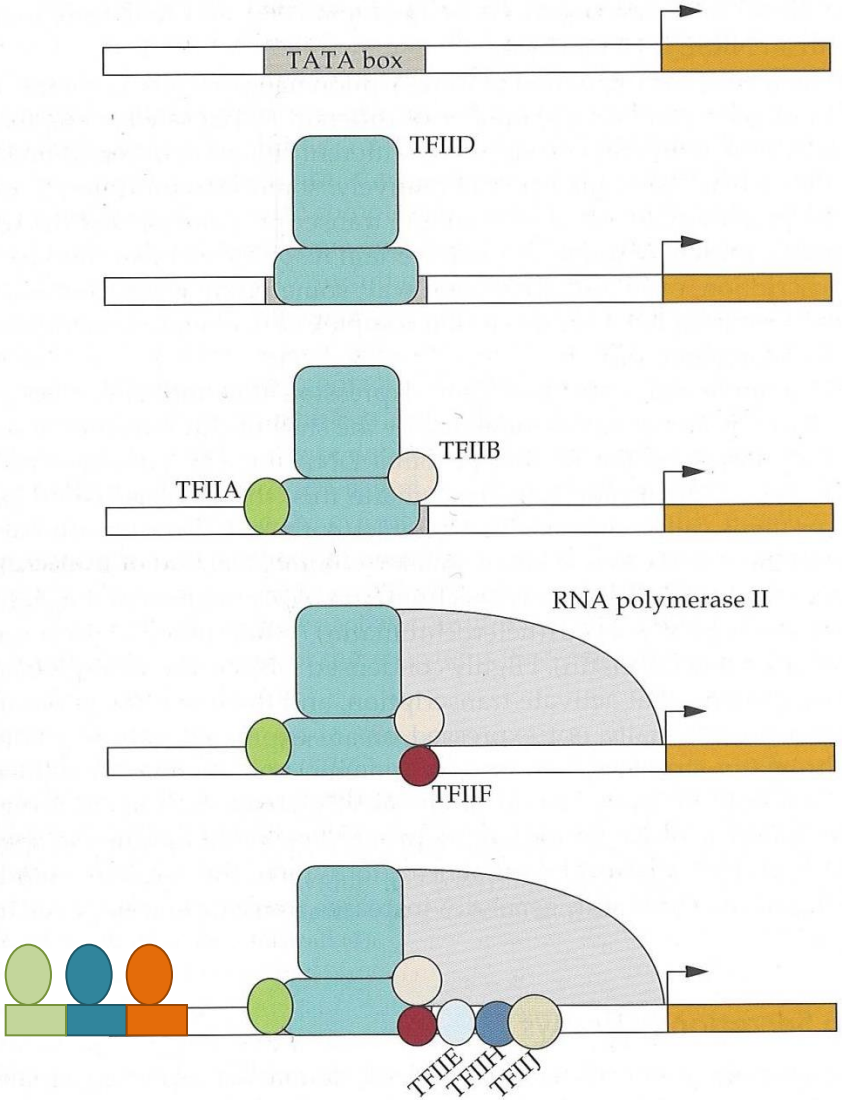
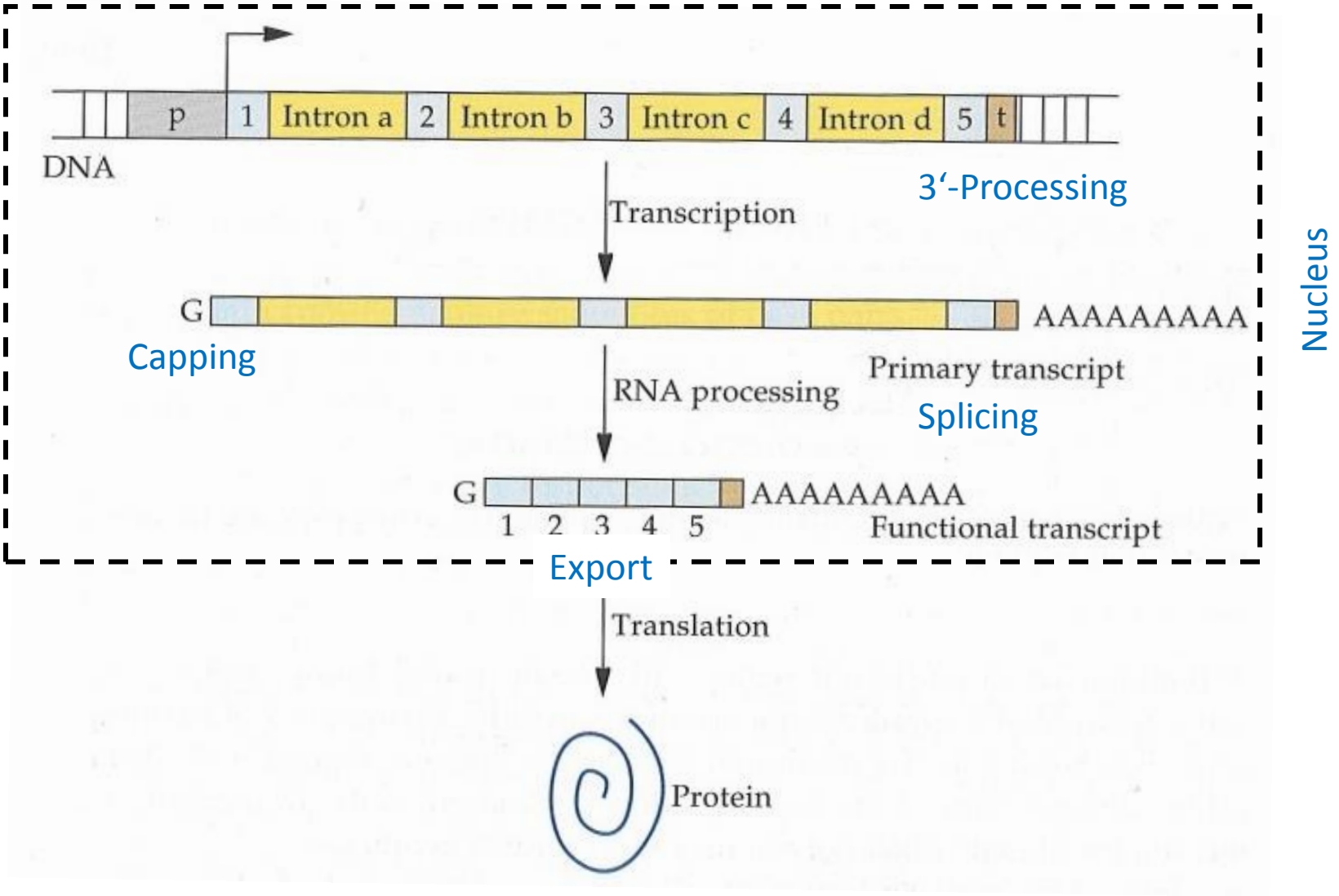


FIGURE 2.28 Formation of an RNA polymerase II transcription initiation complex at a TATA box. Transcription factor TFIID binds to a TATA box, and in sequence, other transcription factors and RNA polymerase II bind to form a protein aggregate that is responsible for initiating transcription. The right-angled arrow indicates the site of initiation and the direction of transcription.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

RNA Processing → Complex Mechanisms



Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Alternative Splicing

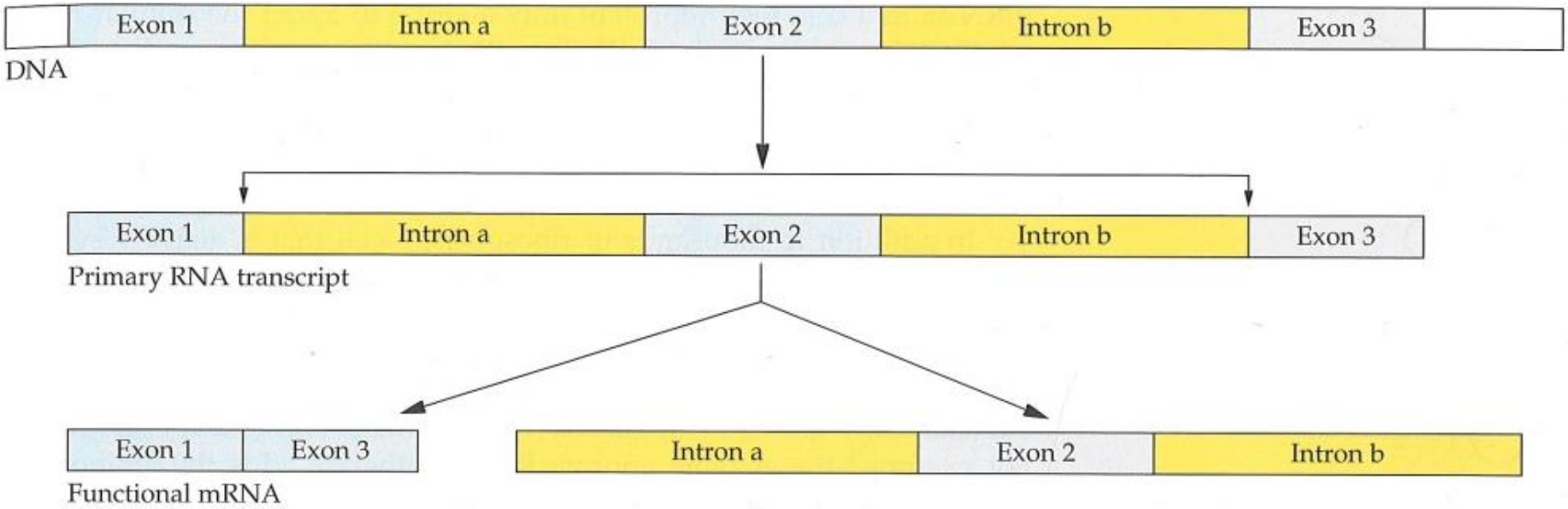


FIGURE 2.13 Alternative splicing of a eukaryotic primary RNA transcript. The bracketing arrows mark the sites that are spliced together after the removal of the intervening RNA region. In this example, exon 2, flanked by introns a and b, is spliced out of the primary transcript, and exons 1 and 3 are spliced together to form a functional mRNA transcript.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Expression Systems for E.coli

Inducible Promoters based on *lacI/lacO* repressor/operator

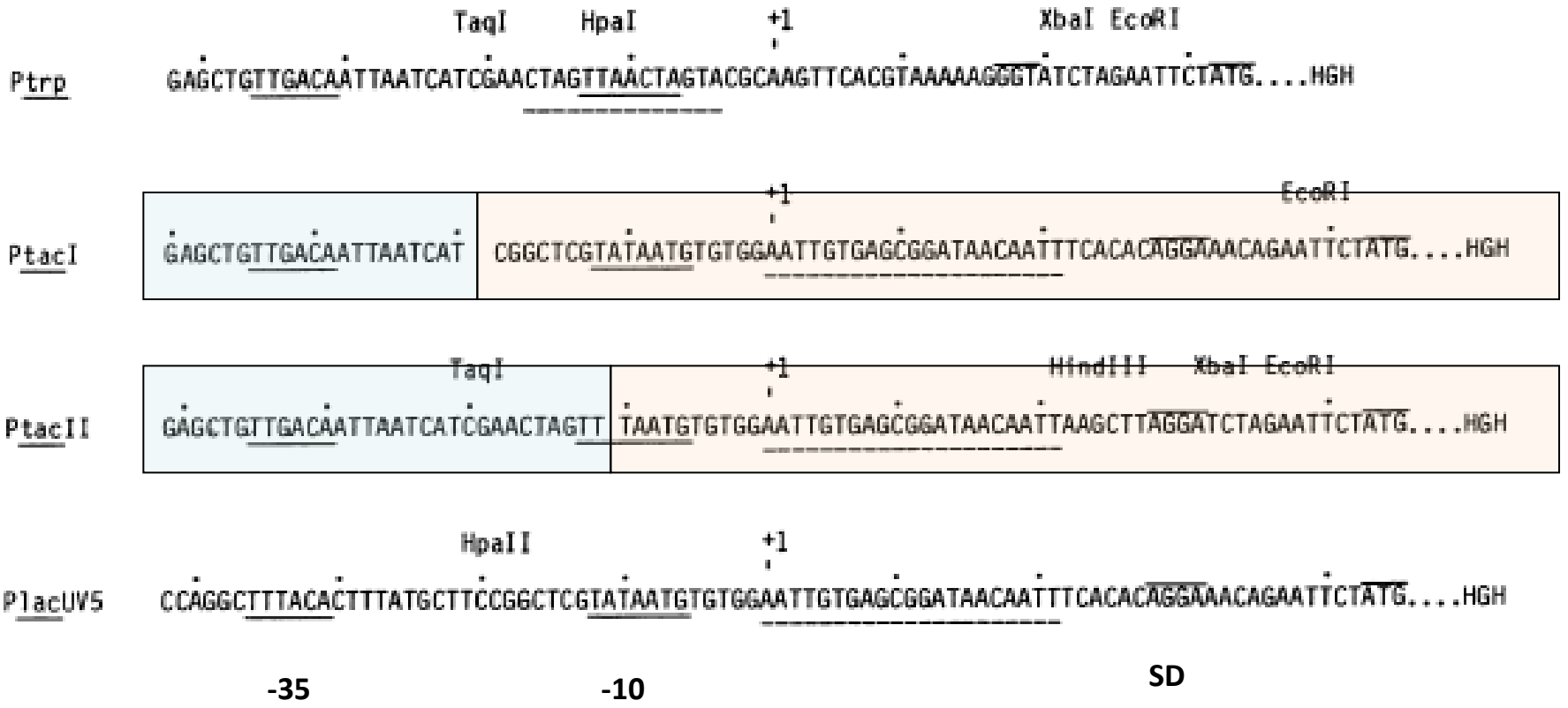
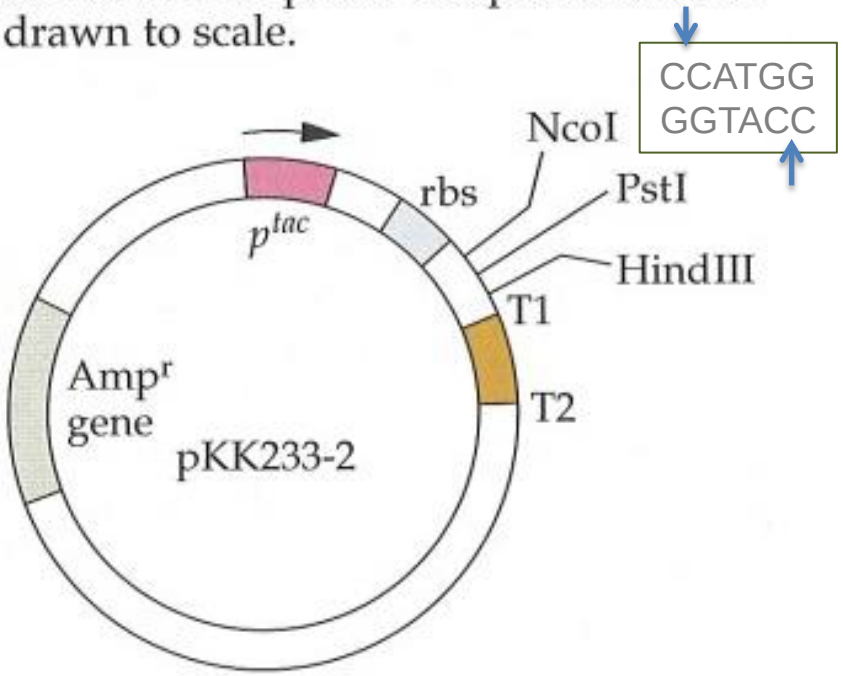


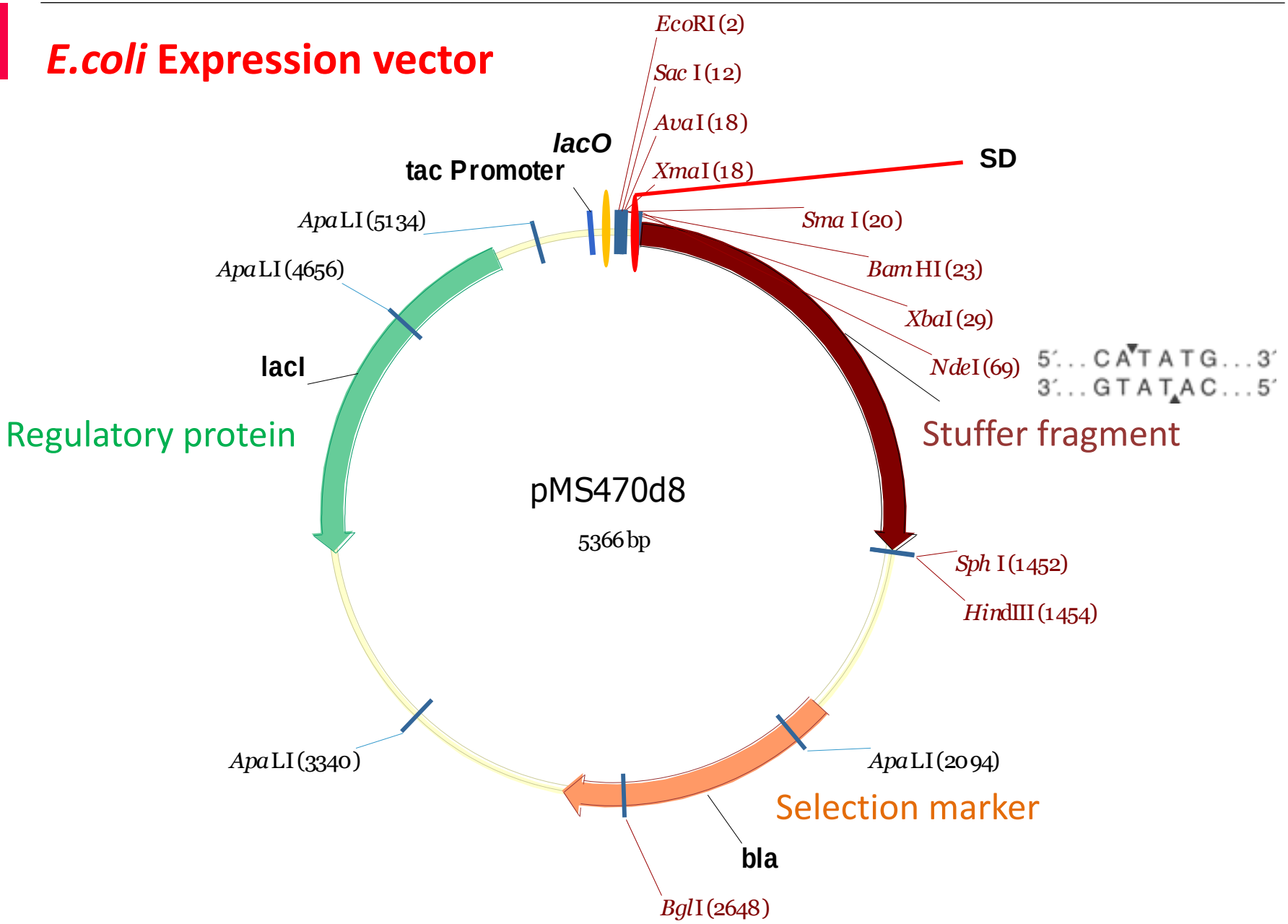
FIGURE 6.19 The expression vector pKK233-2. The plasmid pKK233-2 codes for the ampicillin resistance (Amp^r) gene as a selectable marker gene, the *tac* promoter (p^{tac}), the *lacZ* ribosome-binding site (rbs), three restriction endonuclease cloning sites (NcoI, PstI, and HindIII), and two transcription termination sequences (T1 and T2). The arrow indicates the direction of transcription. The plasmid is not drawn to scale.



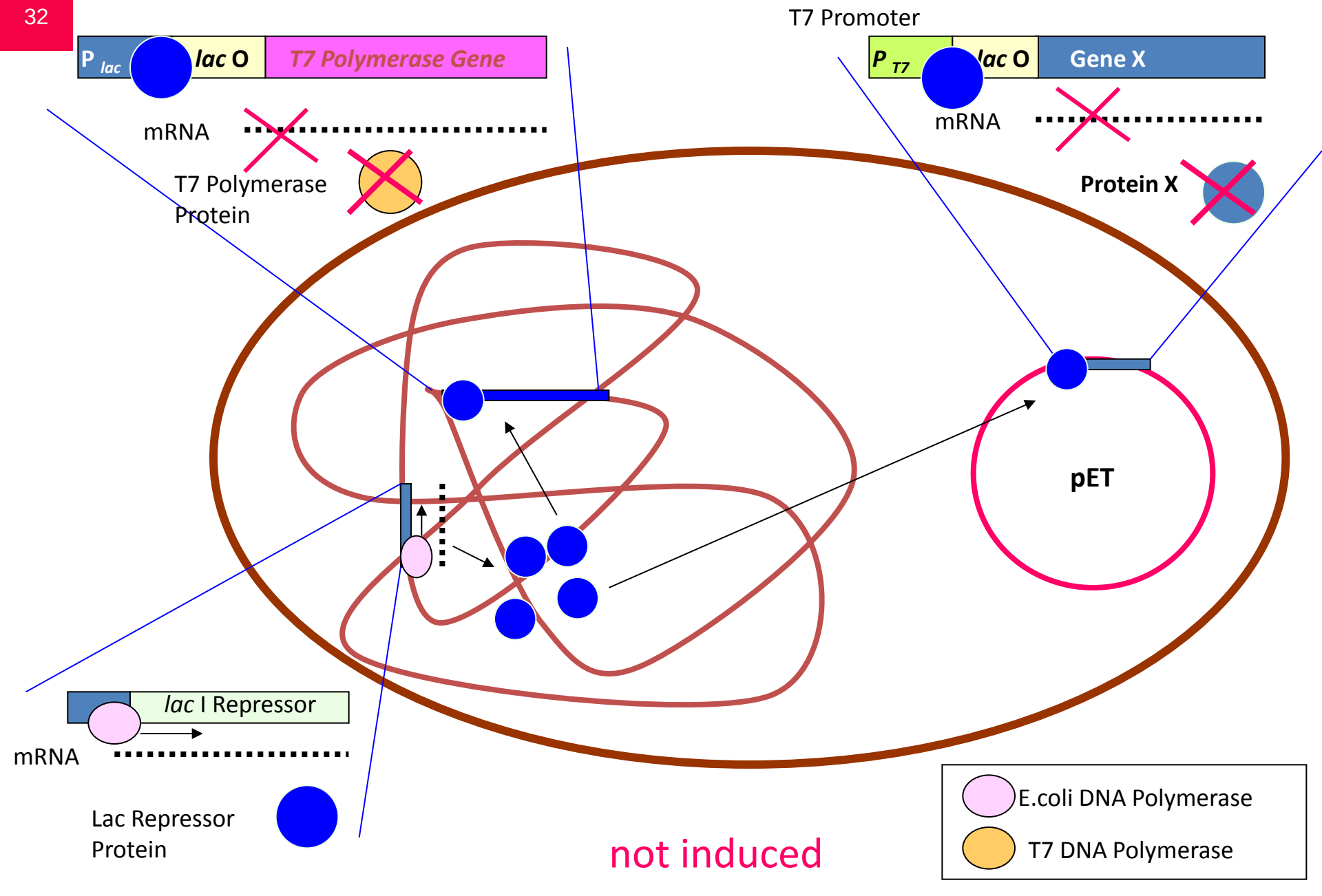
Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

14.10.16

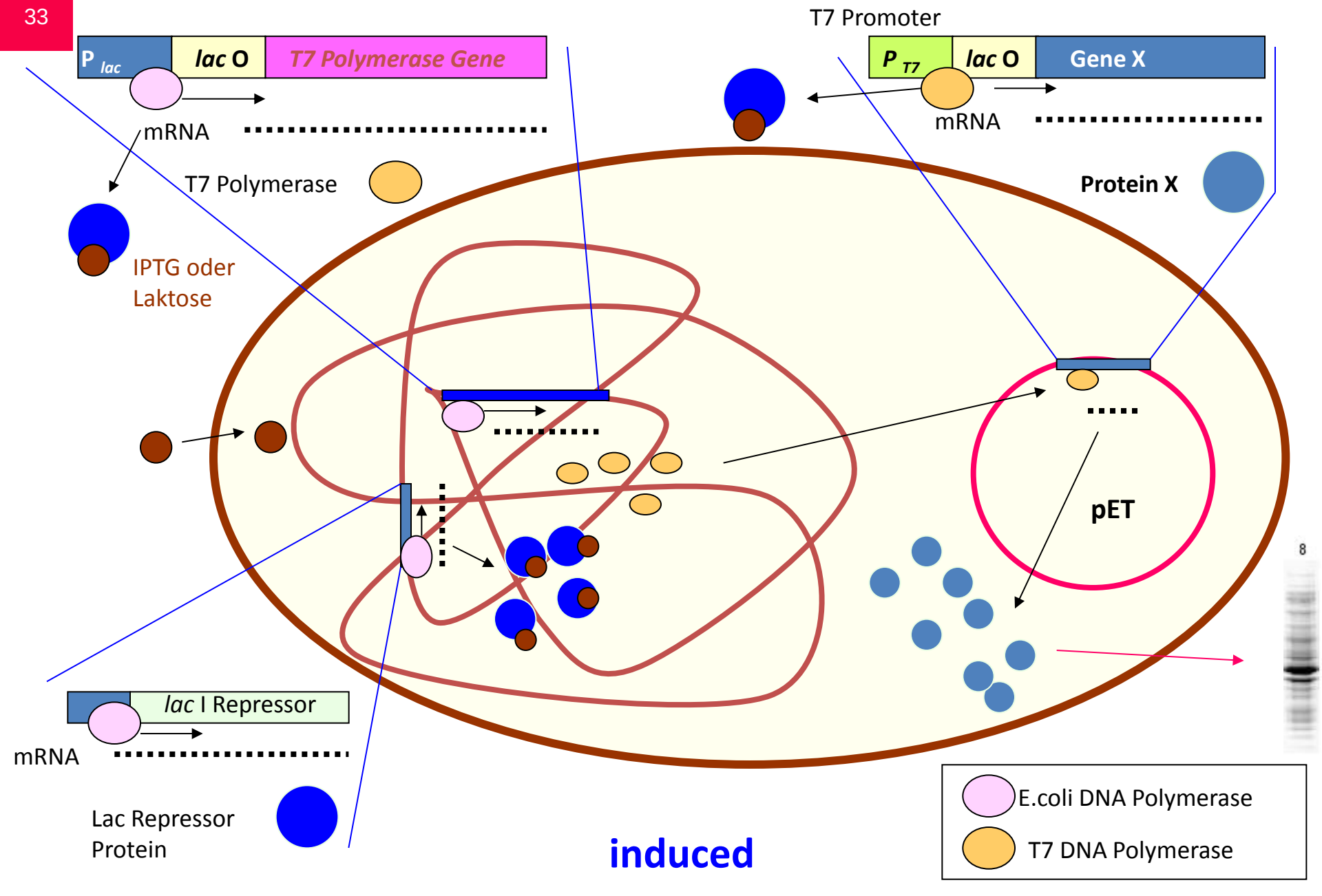
E.coli Expression vector



pET-Expression system



pET-Expression system

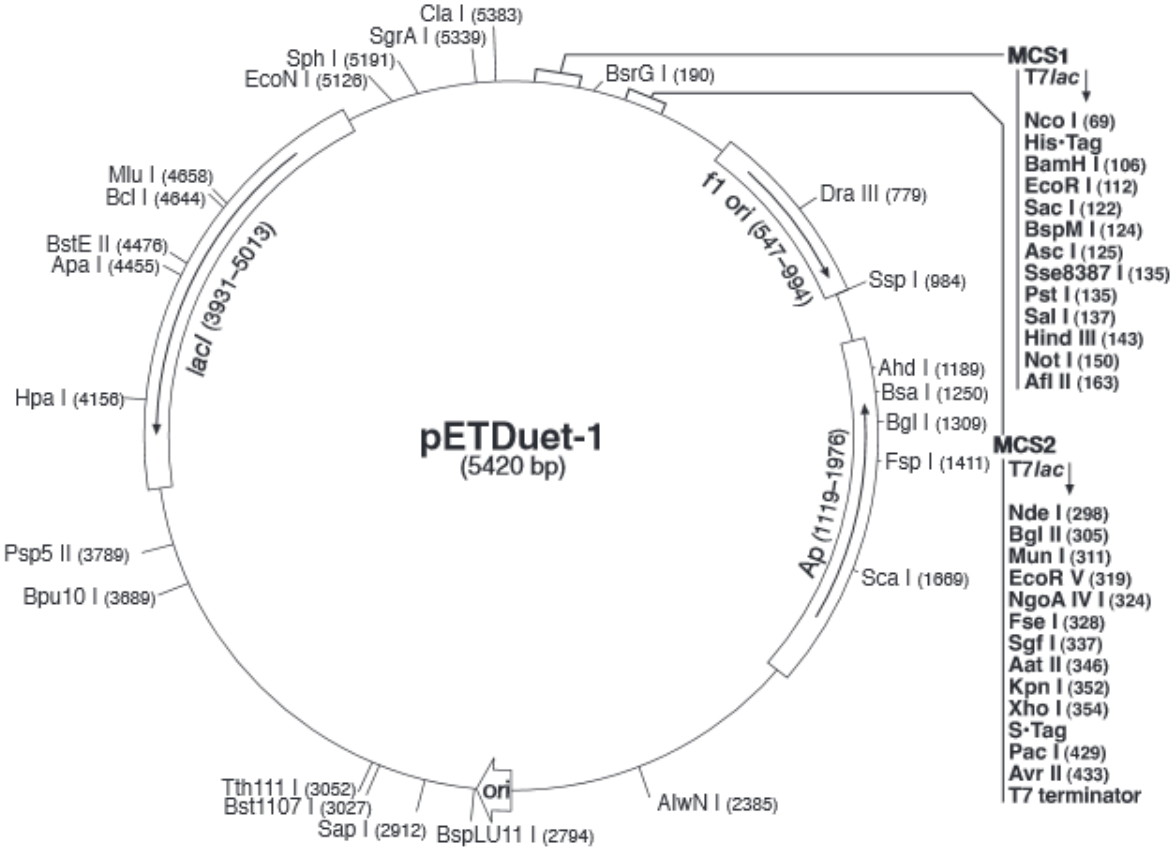


pETDuet-1 Vector

TB337 RevA 0903

	Cat. No.
pETDuet-1 DNA	71146-3
pETDuet-1 sequence landmarks	
T7 promoter-1	5404–5420
T7 transcription start-1	1
His•Tag® coding sequence	83–100
Multiple cloning sites-1 (*Nco* I–*Afl* II)	69–168
T7 promoter-2	214–230
T7 transcription start-2	231
Multiple cloning sites-2 (*Nde* I–*Avr* II)	297–438
S•Tag™ coding sequence	366–410
T7 terminator	462–509
lacI coding sequence	3931–5013
pBR322 origin	2737
bla (Ap) coding sequence	1119–1976
f1 origin	547–994

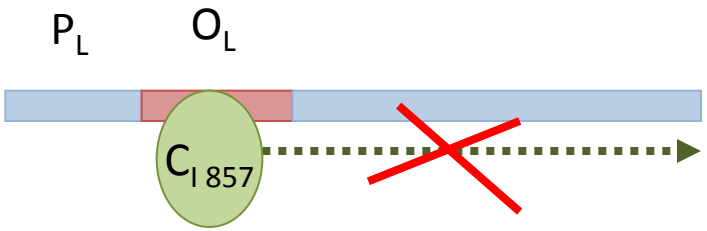
pETDuet™-1 is designed for the coexpression of two target genes. The vector contains two multiple cloning sites (MCS), each of which is preceded by a T7 promoter/*lac* operator and a ribosome binding site (rbs). The vector also carries the pBR322-derived ColE1 replicon, *lacI* gene and ampicillin resistance gene. This vector can be used in combination with pACYCDuet™-1 (Cat. No. 71147-3) in an appropriate host strain for the coexpression of up to 4 target genes. Genes inserted into MCS1 can be sequenced using the pET Upstream Primer (Cat. No. 69214-3) and DuetDOWN1 Primer (Cat. No. 71179-3). Genes inserted into MCS2 can be sequenced using the DuetUP2 Primer (Cat. No. 71180-3) and T7 Terminator Primer (Cat. No. 69337-3).



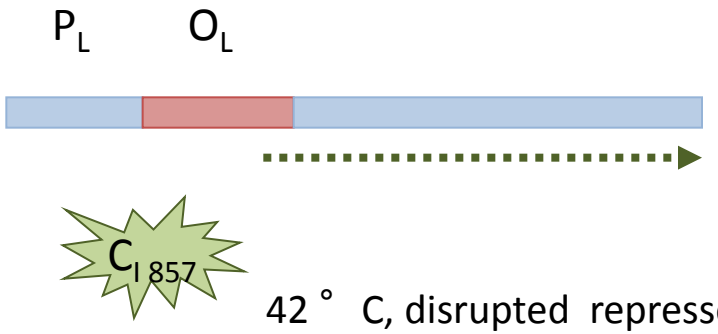
https://www.merckmillipore.com/AT/de/product/pETDuet%E2%84%A2-1-DNA---Novagen,EMD_BIO-71146#anchor_USP

Bacteriophage Lambda Promoters P_L and P_R

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

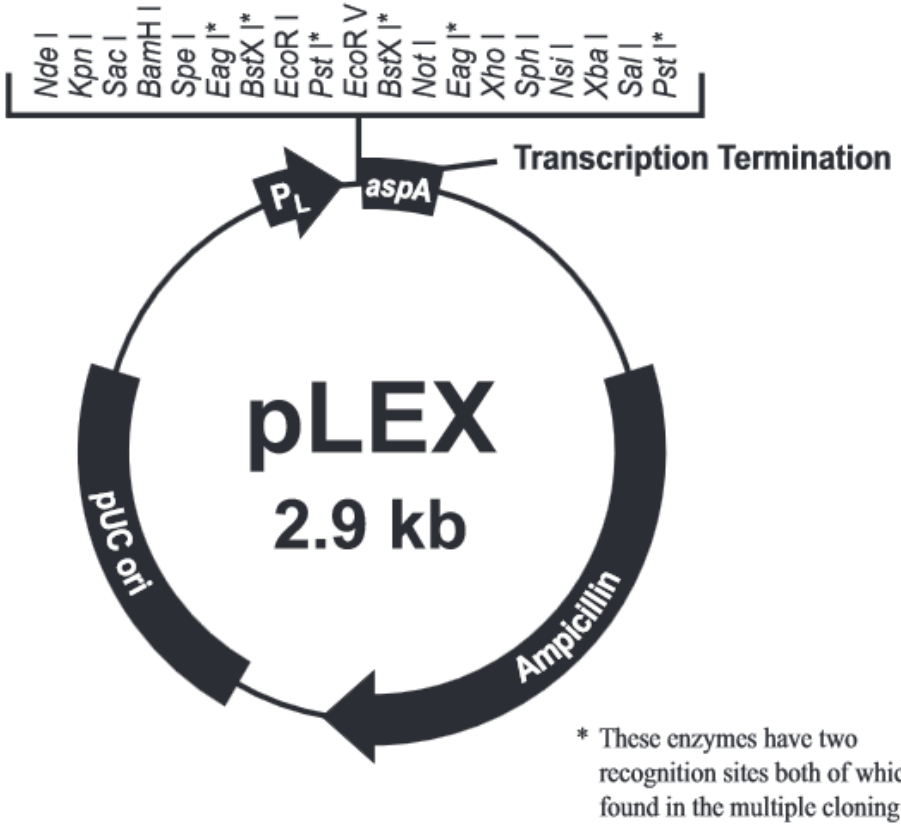


30 ° C, intact repressor



42 ° C, disrupted repressor

P_L based Expression Vector

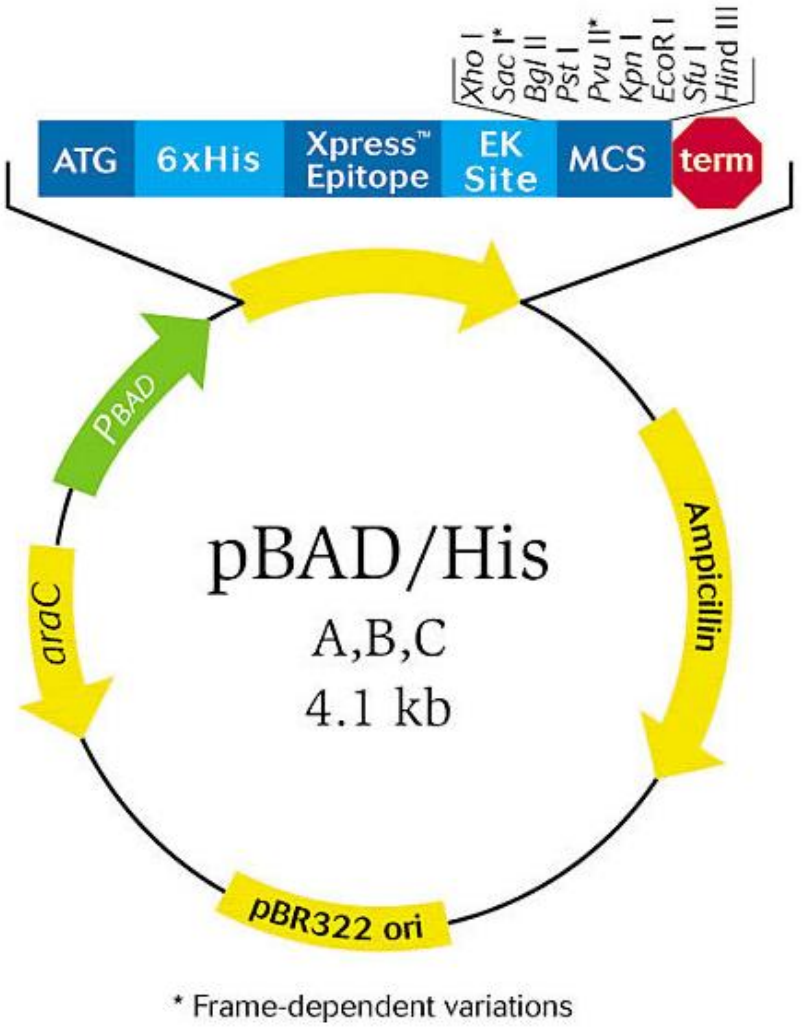


* These enzymes have two recognition sites both of which are found in the multiple cloning site.

http://tools.thermofisher.com/content/sfs/manuals/plex_man.pdf

Arabinose Operon based Expression system

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 *ara*C 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 ColE1 origin for selection and maintenance in E. coli

https://www.thermofisher.com/order/catalog/product/V43001

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 → Target for RNases

Sequence structure → 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....

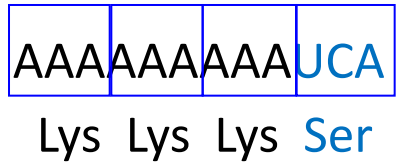
- Secondary structures

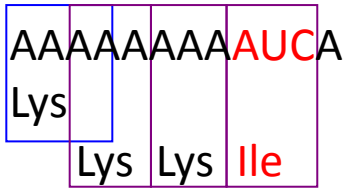
Translation elongation

- Codon usage

- Secondary structures

- Codon structure – translational frameshifting





Translation - Prokaryotes

Shine-Dalgarno (SD) Sequence

```
rRNA    5'-GAUACCAUCCUCCUUA-3'
mRNA              ....GGAGG..(5-7bp)...AUG
```

Influences:

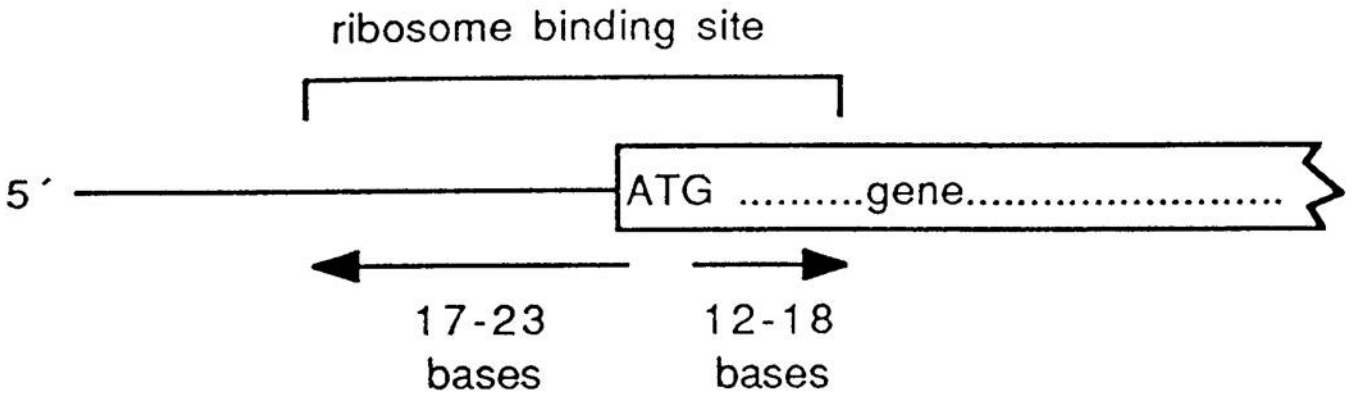
Secondary structure!! SD and AUG in unstructured region

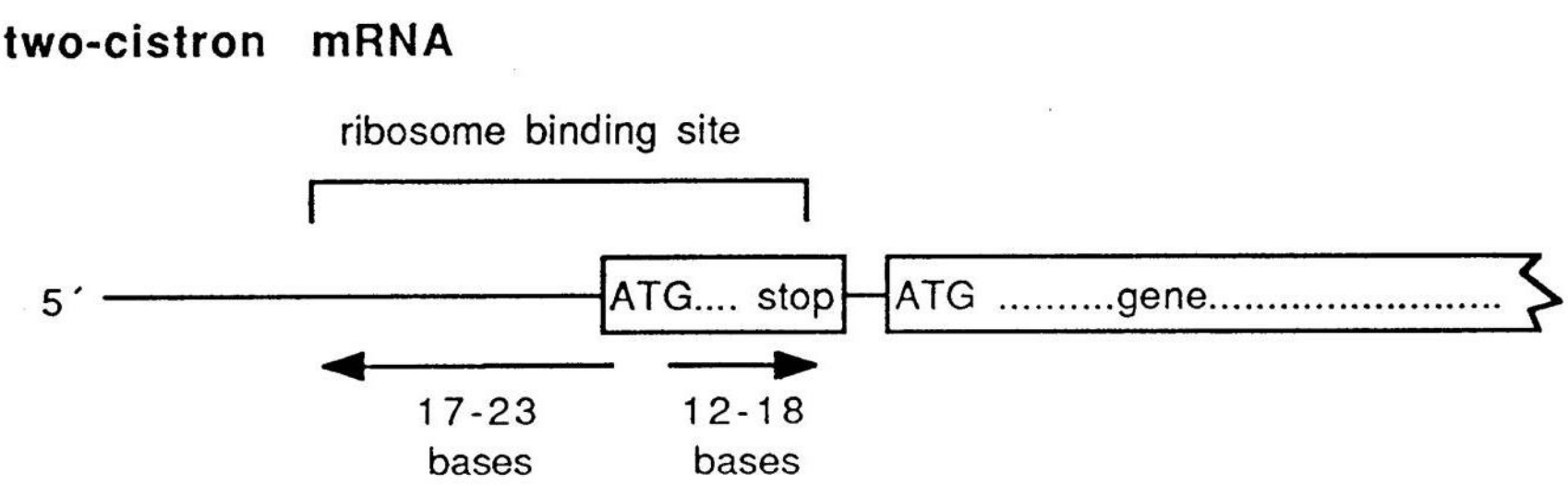
Surrounding of SD and AUG!!!

Start

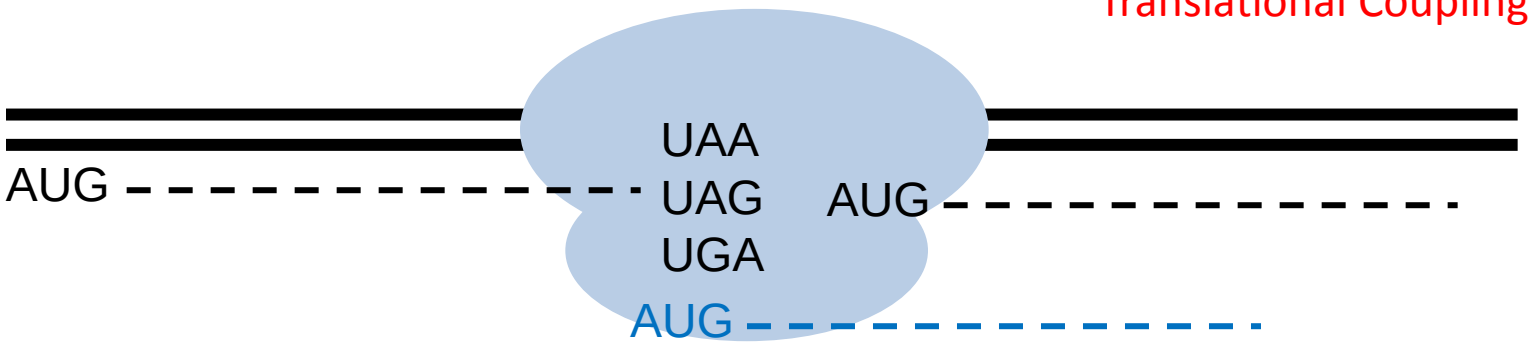
AUG	91%
GUG	8
UUG	1

one-cistron mRNA





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 → not a ribosome binding site,
present with highly expressed genes

……CC$^{A}/_{G}$CCAUGG…… mammalian

…… $^{A}/_{T}A^{A}/_{C}A^{A}/_{C}$AAUGTC$^{T}/_{C}$…… yeast

Translation Initiation

- SD sequence ...AGGAG....

- Secondary structures

Translation elongation

- Codon usage

- Secondary structures

- Codon structure – translational frameshifting

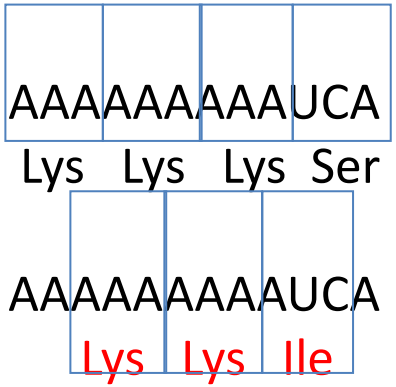


Table 3.2 The genetic code and codon usage in *E. coli* and humans

Codon	Amino acid	Frequency of use in:	
		E. coli	Human
GGG	Glycine	0.13	0.23
GGA	Glycine	0.09	0.26
GGU	Glycine	0.38	0.18
GGC	Glycine	0.40	0.33
GAG	Glutamic acid	0.30	0.59
GAA	Glutamic acid	0.70	0.41
GAU	Aspartic acid	0.59	0.44
GAC	Aspartic acid	0.41	0.56
GUG	Valine	0.34	0.48
GUA	Valine	0.17	0.10
GUU	Valine	0.29	0.17
GUC	Valine	0.20	0.25
GCG	Alanine	0.34	0.10
GCA	Alanine	0.22	0.22
GCU	Alanine	0.19	0.28
GCC	Alanine	0.25	0.40
AAG	Lysine	0.24	0.60
AAA	Lysine	0.76	0.40
AAU	Asparagine	0.39	0.44
AAC	Asparagine	0.61	0.56
AUG	Methionine	1.00	1.00
AUA	Isoleucine	0.07	0.14
AUU	Isoleucine	0.47	0.35
AUC	Isoleucine	0.46	0.51
ACG	Threonine	0.23	0.12
ACA	Threonine	0.12	0.27
ACU	Threonine	0.21	0.23
ACC	Threonine	0.43	0.38
UGG	Tryptophan	1.00	1.00
UGU	Cysteine	0.43	0.42
UGC	Cysteine	0.57	0.58
UGA	Stop	0.30	0.61
UAG	Stop	0.09	0.17
UAA	Stop	0.62	0.22
UAU	Tyrosine	0.53	0.42
UAC	Tyrosine	0.47	0.58
UUU	Phenylalanine	0.51	0.43
UUC	Phenylalanine	0.49	0.57
UCG	Serine	0.13	0.06
UCA	Serine	0.12	0.15
UCU	Serine	0.19	0.17
UCC	Serine	0.17	0.23
AGU	Serine	0.13	0.14
AGC	Serine	0.27	0.25
CGG	Arginine	0.08	0.19
CGA	Arginine	0.05	0.10
CGU	Arginine	0.42	0.09
CGC	Arginine	0.37	0.19
AGG	Arginine	0.03	0.22
AGA	Arginine	0.04	0.21
CAG	Glutamine	0.69	0.73
CAA	Glutamine	0.31	0.27
CAU	Histidine	0.52	0.41
CAC	Histidine	0.48	0.59
CUG	Leucine	0.55	0.43
CUA	Leucine	0.03	0.07
CUU	Leucine	0.10	0.12
CUC	Leucine	0.10	0.20
UUG	Leucine	0.11	0.12
UUA	Leucine	0.11	0.06
CCG	Proline	0.55	0.11
CCA	Proline	0.20	0.27
CCU	Proline	0.16	0.29

Table 3.2 The genetic code and codon usage in *E. coli* and humans

Codon	Amino acid	Frequency of use in:	
		E. coli	Humans
GGG	Glycine	0.13	0.23
GGA	Glycine	0.09	0.26
GGU	Glycine	0.38	0.18
GGC	Glycine	0.40	0.33
GAG	Glutamic acid	0.30	0.59
GAA	Glutamic acid	0.70	0.41
GAU	Aspartic acid	0.59	0.44
GAC	Aspartic acid	0.41	0.56
GUG	Valine	0.34	0.48
GUA	Valine	0.17	0.10
GUU	Valine	0.29	0.17
GUC	Valine	0.20	0.25
GCG	Alanine	0.34	0.10
GCA	Alanine	0.22	0.22
GCU	Alanine	0.19	0.28
GCC	Alanine	0.25	0.40
AAG	Lysine	0.24	0.60
AAA	Lysine	0.76	0.40
AAU	Asparagine	0.39	0.44
AAC	Asparagine	0.61	0.56
AUG	Methionine	1.00	1.00
AUA	Isoleucine	0.07	0.14
AUU	Isoleucine	0.47	0.35
AUC	Isoleucine	0.46	0.51
ACG	Threonine	0.23	0.12
ACA	Threonine	0.12	0.27
ACU	Threonine	0.21	0.23
ACC	Threonine	0.43	0.38
UGG	Tryptophan	1.00	1.00
UGU	Cysteine	0.43	0.42
UGC	Cysteine	0.57	0.58

UGA	Stop	0.30	0.61
UAG	Stop	0.09	0.17
UAA	Stop	0.62	0.22
UAU	Tyrosine	0.53	0.42
UAC	Tyrosine	0.47	0.58
UUU	Phenylalanine	0.51	0.43
UUC	Phenylalanine	0.49	0.57
UCG	Serine	0.13	0.06
UCA	Serine	0.12	0.15
UCU	Serine	0.19	0.17
UCC	Serine	0.17	0.23
AGU	Serine	0.13	0.14
AGC	Serine	0.27	0.25
CGG	Arginine	0.08	0.19
CGA	Arginine	0.05	0.10
CGU	Arginine	0.42	0.09
CGC	Arginine	0.37	0.19
AGG	Arginine	0.03	0.22
AGA	Arginine	0.04	0.21
CAG	Glutamine	0.69	0.73
CAA	Glutamine	0.31	0.27
CAU	Histidine	0.52	0.41
CAC	Histidine	0.48	0.59
CUG	Leucine	0.55	0.43
CUA	Leucine	0.03	0.07
CUU	Leucine	0.10	0.12
CUC	Leucine	0.10	0.20
UUG	Leucine	0.11	0.12
UUA	Leucine	0.11	0.06
CCG	Proline	0.55	0.11
CCA	Proline	0.20	0.27
CCU	Proline	0.16	0.29
CCC	Proline	0.10	0.33

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

Protein Folding

Translation Conditions

- Elongation velocity
- Codon Structure – Pausing
- Domain folding

Disulfide Bond Formation

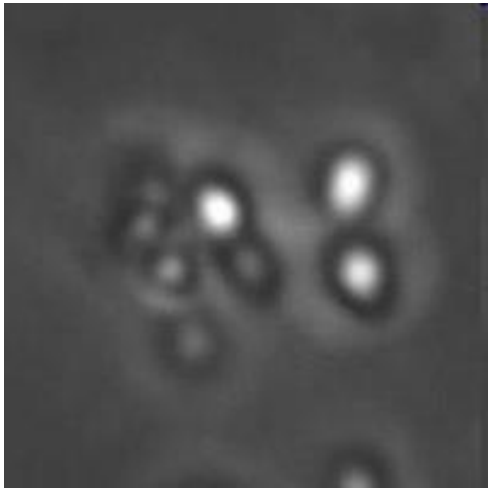
- Redox Conditions
  - *E.coli* Cytosol → bad conditions - reductive
  - *E.coli* Periplasm → optimal conditions - oxidative

Chaperones

Inclusion Body Formation

Expression velocity → Translation

Protein Folding



The Department of Surface Biotechnology with the Center for Surface Biotechnology, Box 577, BMC, 751 23 Uppsala

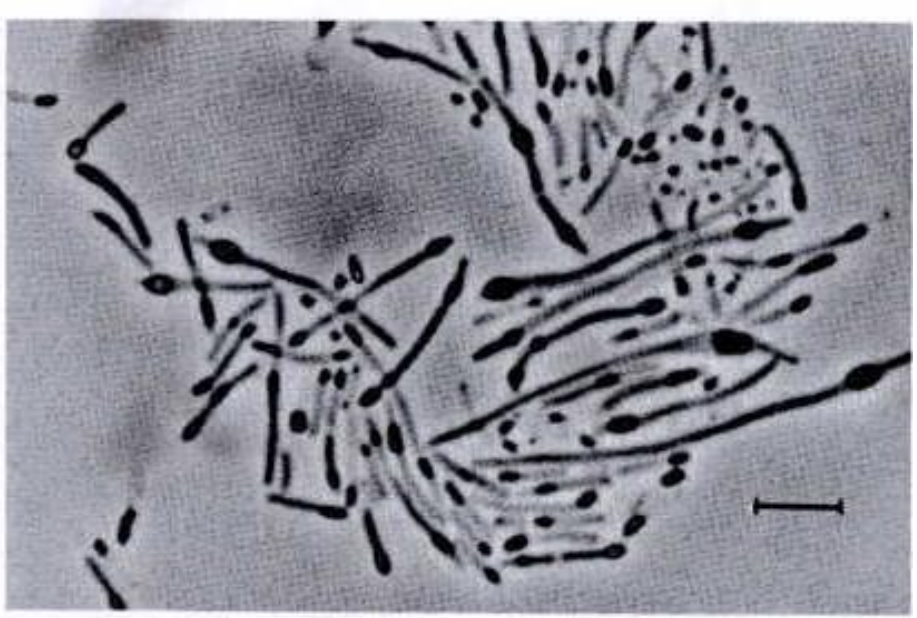
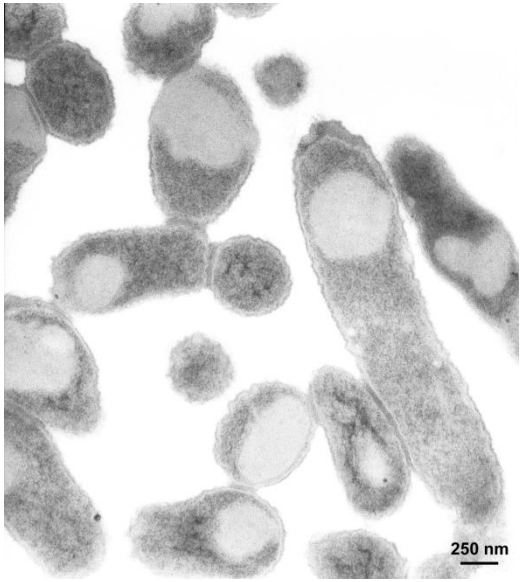


Figure 3 Phase-contrast microscopy of HB101/pBSF2-SD7 cells overproducing hIL-6 (bar equals 2 μm). (From Ref. 61)



www.boku.ac.at/IAM/dn/EM424_23.jpg

20.10.16

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

Post-translational modifications

- Side Chain Modifications
  - Glycosylation, Phosphorylation, Sulfatation, etc.
- Proteolytic Processing
  - ss Cleavage
  - Pro-protein processing
  - N/C-terminal Processing

Posttranslational 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 → ..... X-Pro
- aminopeptidase P: NH_2-X-/Pro
- dipeptidylaminopeptidase I (DAP-I, Cathepsin C) → not at NH_2 Pro/Arg/Lys

protein fusion strategies

- sequence specific proteases
- tags

Table 6.5 α-Amylase gene copy number and activity in *B. subtilis*

Copies/genome	Activity (U/mL of mid-log cells)
2	500
5	2,300
7	3,100
8	3,400
9	4,400
Multicopy plasmid	700

Adapted from Kallio et al., 1987, *Appl. Microbiol. Biotechnol.* **27**:64–71.

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

E. coli HB101 with plasmid:	Plasmid copy number	Relative specific growth rate
None	0	1.00
A	12	0.92
B	24	0.91
C	60	0.87
D	122	0.82
E	408	0.77

Adapted from Seo and Bailey, 1985, *Biotechnol. Bioeng.* 27:1668–1674.

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

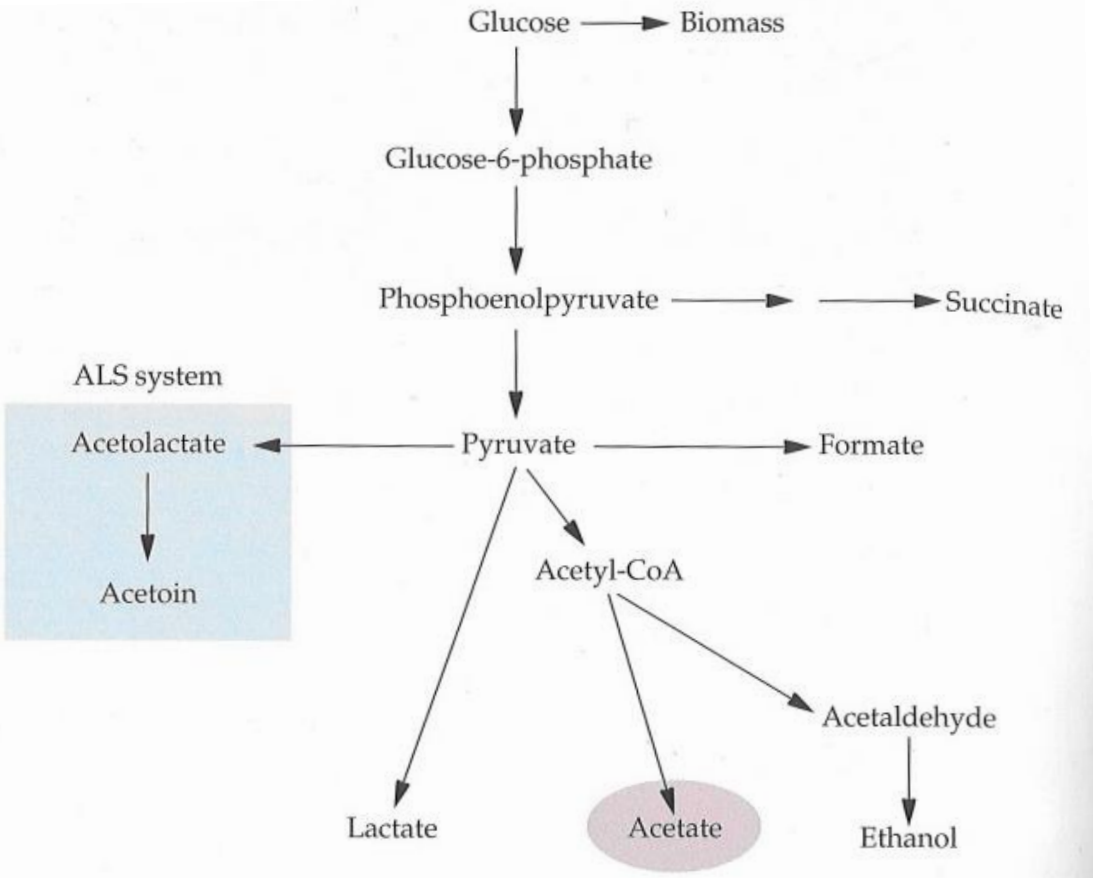


FIGURE 17.7 Schematic representation of the pathways for glucose metabolism in an *E. coli* strain that has been transformed with a plasmid carrying the genes for the protein subunits of acetolactate synthase (ALS). Note that the conversion of glucose to biomass is a multistep process. CoA, coenzyme A.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Gene Fusion Strategies

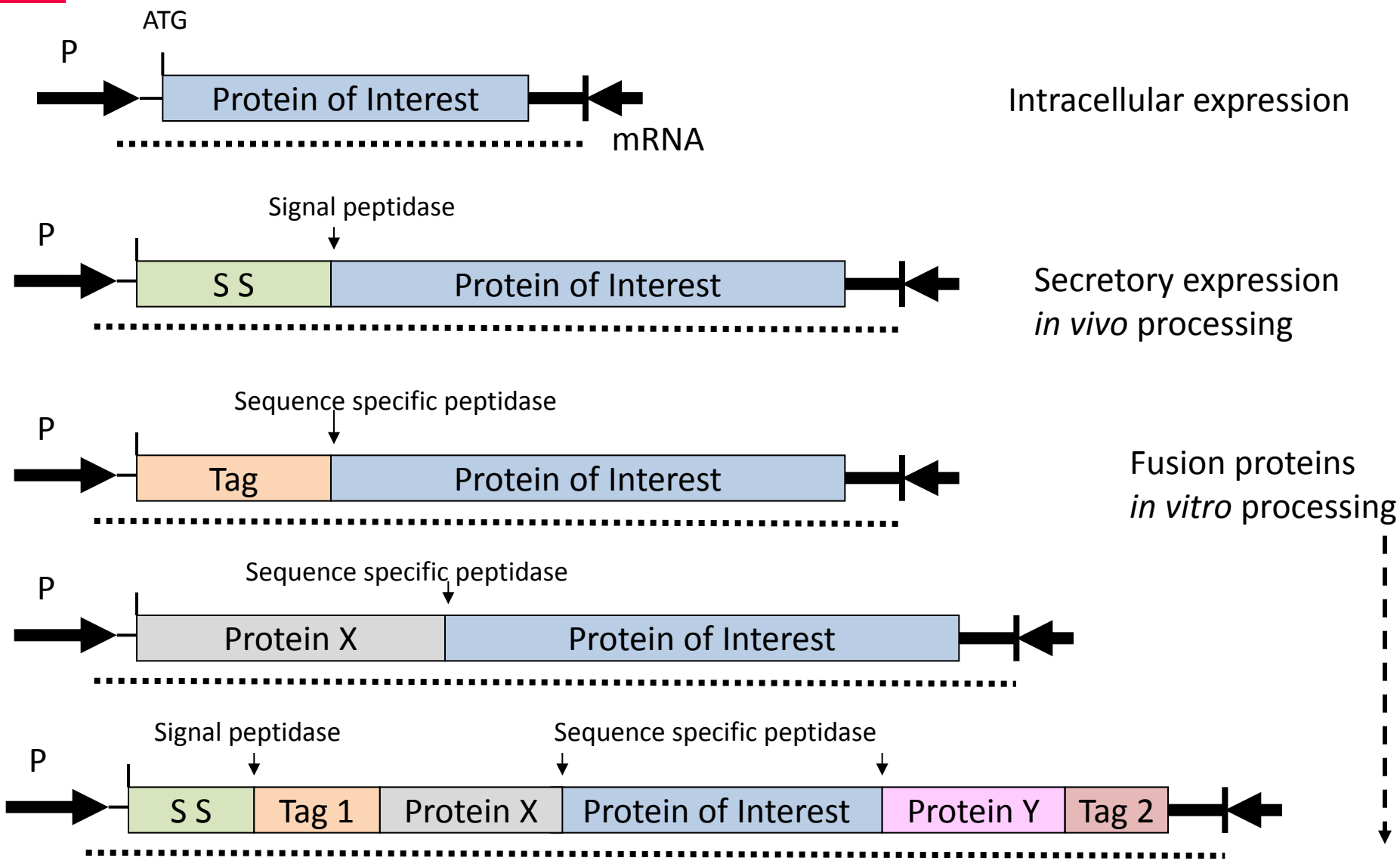


Table 6.3 Some fusion systems used to facilitate the purification of foreign proteins produced in *E. coli*

Fusion partner	Size	Ligand	Elution condition
ZZ	14 kDa	IgG	Low pH
His tail	6–10 aa	Ni^{2+}	Imidazole
Strep-tag	10 aa	Streptavidin	Iminobiotin
PinPoint	13 kDa	Streptavidin	Biotin
MBP	40 kDa	Amylose	Maltose
β-Lactamase	27 kDa	Phenyl-boronate	Borate
GST	25 kDa	Glutathione	Reducing agent
Flag	8 aa	Specific MAb	Low calcium

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

Tag	Residues	Sequence	Size (kDa)
Poly-Arg	5–6 (usually 5)	RRRRR	0.80
Poly-His	2–10 (usually 6)	HHHHHH	0.84
FLAG	8	DYKDDDDK	1.01
Strep-tag II	8	WSHPQFEK	1.06
c-myc	11	EQKLISEEDL	1.20
S-	15	KETAAAKFERQHMDS	1.75
HAT-	19	KDHLIHNVHKEFHAHAHNK	2.31
3x FLAG	22	DYKDHDGDYKDHDIDYKDDDDK	2.73
Calmodulin-binding peptide	26	KRRWKKNFIAVSAANRFKKISSSGAL	2.96
Cellulose-binding domains	27–189	Domains	3.00–20.00
SBP	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP	4.03
Chitin-binding domain	51	TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ	5.59
Glutathione S-transferase	211	Protein	26.00
Maltose-binding protein	396	Protein	40.00

Table 1 Matrices and elution conditions of affinity tags

Affinity tag	Matrix	Elution condition
Poly-Arg	Cation-exchange resin	NaCl linear gradient from 0 to 400 mM at alkaline pH>8.0
Poly-His	Ni^{2+}-NTA, Co^{2+}-CMA (Talon)	Imidazole 20–250 mM or low pH
FLAG	Anti-FLAG monoclonal antibody	pH 3.0 or 2–5 mM EDTA
Strep-tag II	Strep-Tactin (modified streptavidin)	2.5 mM desthiobiotin
c-myc	Monoclonal antibody	Low pH
S	S-fragment of RNaseA	3 M guanidine thiocyanate, 0.2 M citrate pH 2, 3 M magnesium chloride
HAT (natural histidine affinity tag)	Co^{2+}-CMA (Talon)	150 mM imidazole or low pH
Calmodulin-binding peptide	Calmodulin	EGTA or EGTA with 1 M NaCl
Cellulose-binding domain	Cellulose	Family I: guanidine HCl or urea>4 M Family II/III: ethylene glycol
SBP	Streptavidin	2 mM Biotin
Chitin-binding domain	Chitin	Fused with intein: 30–50 mM dithiothreitol, β-mercaptoethanol or cysteine
Glutathione S-transferase	Glutathione	5–10 mM reduced glutathione
Maltose-binding protein	Cross-linked amylose	10 mM maltose

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

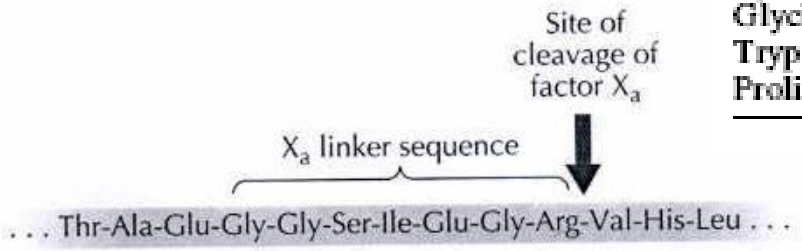


Figure 6.6 Proteolytic cleavage of a fusion protein by blood coagulation factor X_a. The factor X_a recognition sequence (X_a linker sequence) lies between the amino acid sequences of two different proteins. A functional cloned gene protein (with Val at its N terminus) is released after cleavage.

Table 4 Cleavage (%) of enterokinase through densitometry (Hosfield and Lu 1999) based on the amino acid residue X_1. The sequence....-GSDYKDDDDK-X_1-ADQLTEEQIA-... of a GST-calmodulin fusion protein was tested using 5 mg protein digested with 0.2 U of enterokinase for 16 h at 37 °C

Amino acid in position X_1	Cleavage of enterokinase (%)
Alanine	88
Methionine	86
Lysine	85
Leucine	85
Asparagine	85
Phenylalanine	85
Isoleucine	84
Aspartic acid	84
Glutamic acid	80
Glutamine	79
Valine	79
Arginine	78
Threonine	78
Tyrosine	78
Histidine	76
Serine	76
Cysteine	74
Glycine	74
Tryptophan	67
Proline	61

Tag purification strategies

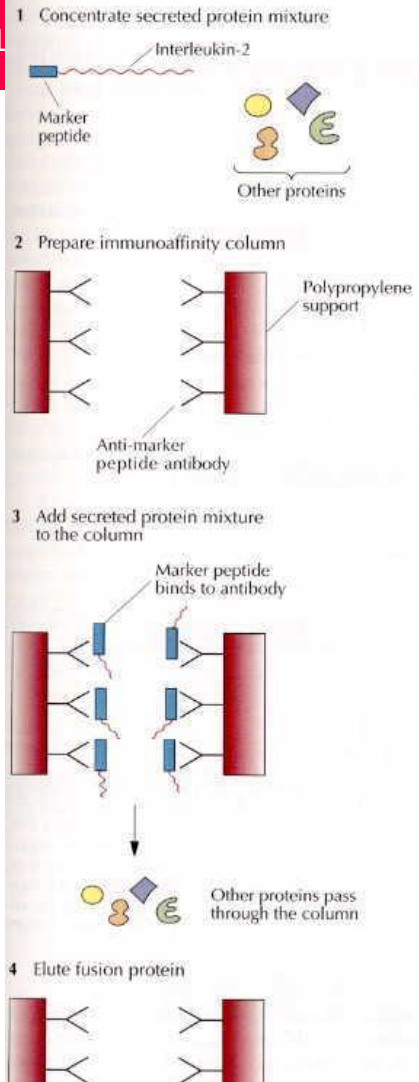
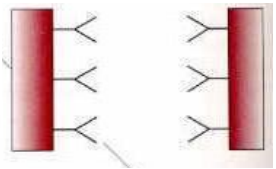
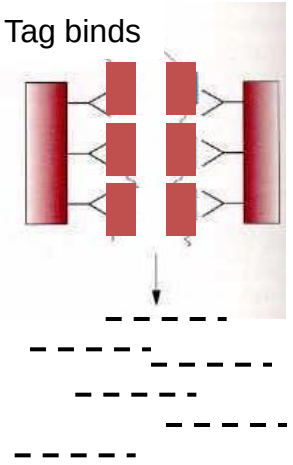


Figure 6.8 Immunoaffinity chromatographic purification of a fusion protein. An antibody that binds to the marker peptide of the fusion protein (anti-marker peptide antibody) is attached to a solid polypropylene support. The secreted proteins are passed through the column containing the bound antibody. 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.

Removal of Tag

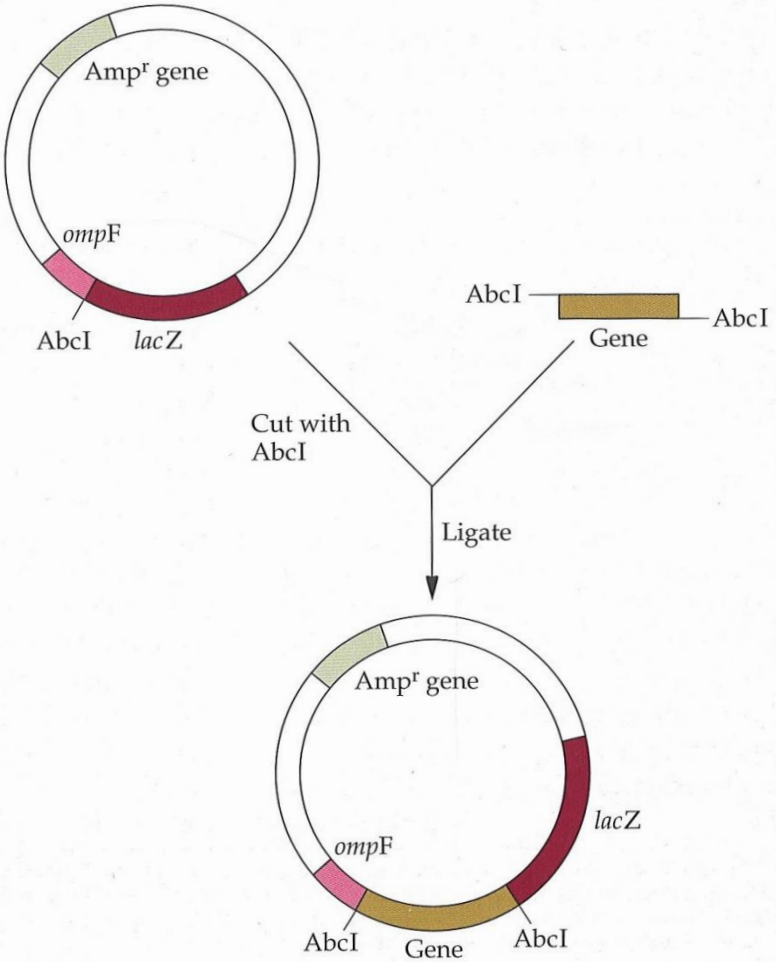


Tag binds



Proteolytic cleavage of Tag

FIGURE 6.11 A fusion protein cloning vector. The plasmid contains an ampicillin resistance (Amp^r) 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 tribrid protein is produced consisting of OmpF–target protein–LacZ.



Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Examples for fusion strategies

For *E.coli*:

Maltose binding protein
Thioredoxin reductase

Generally: well soluble proteins
Well folded proteins

Fusions can help for:

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

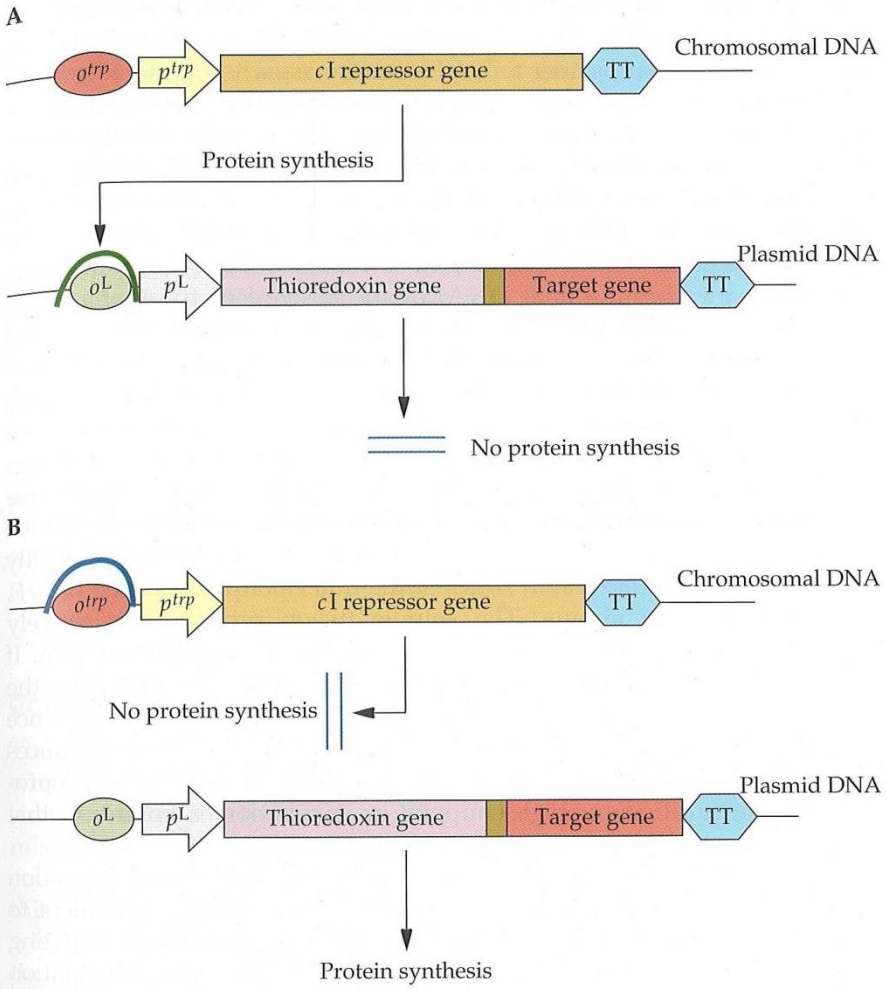


FIGURE 6.21 Regulation of the synthesis of a thioredoxin–target protein fusion in the absence **(A)** or presence **(B)** of tryptophan in the growth medium. The arrows labeled p^{trp} and p^L indicate the direction of transcription. o^{trp}, the operator region where the *trp* repressor protein binds; o^L, the operator region where the *c*I repressor binds; p^{trp}, the *trp* promoter; p^L, the leftward promoter from bacteriophage λ; TT, transcription termination region. The box between the thioredoxin and target genes indicates the DNA region that codes for the peptide that acts as the enterokinase cleavage site; the horseshoes indicate the binding of a repressor protein to its operator region.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Eukaryotic Expression Systems

Fungi – Yeasts

Insect Cells

Plant Cells

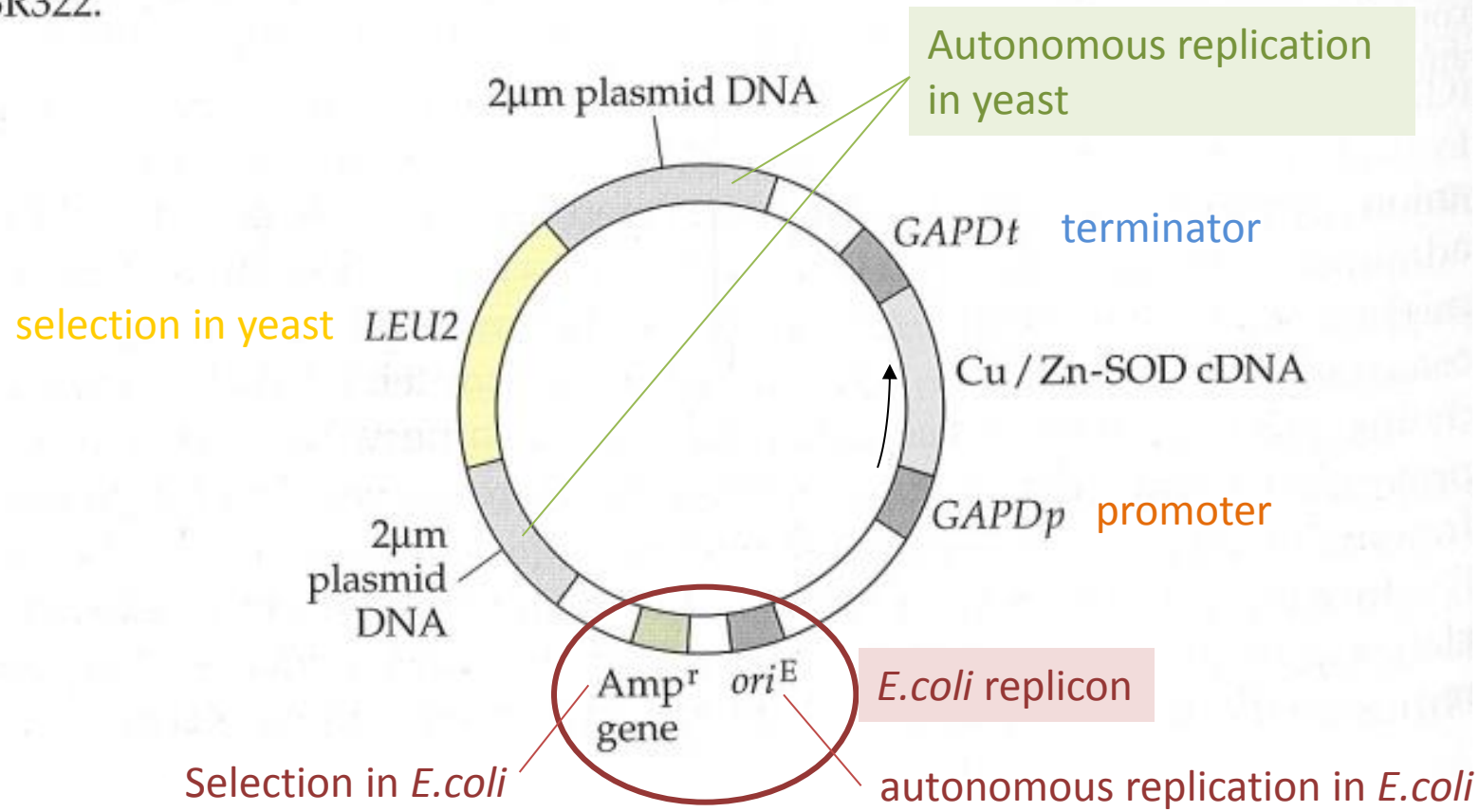
Mammalian Cells

- Mouse
- Hamster
- Avian
- Human

Transgenic Plants

Transgenic Animals

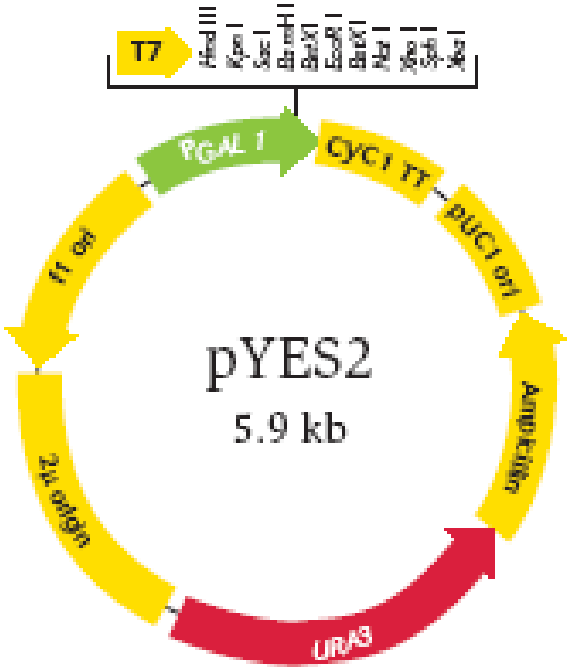
FIGURE 7.7 *S. cerevisiae* expression vector. The cDNA for human Cu/Zn-SOD was cloned between the promoter *(GAPDp)* and termination–polyadenylation sequence *(GAPDt)* of the *S. cerevisiae* glyceraldehyde phosphate dehydrogenase gene. The *LEU2* gene that was cloned between segments of the yeast 2µm plasmid DNA encodes a functional enzyme of the leucine biosynthesis pathway. The yeast origin of replication is included in the 2µm plasmid DNA. The ampicillin resistance (Amp^r) gene and the *E. coli* origin of replication (ori^E) are derived from plasmid pBR322.

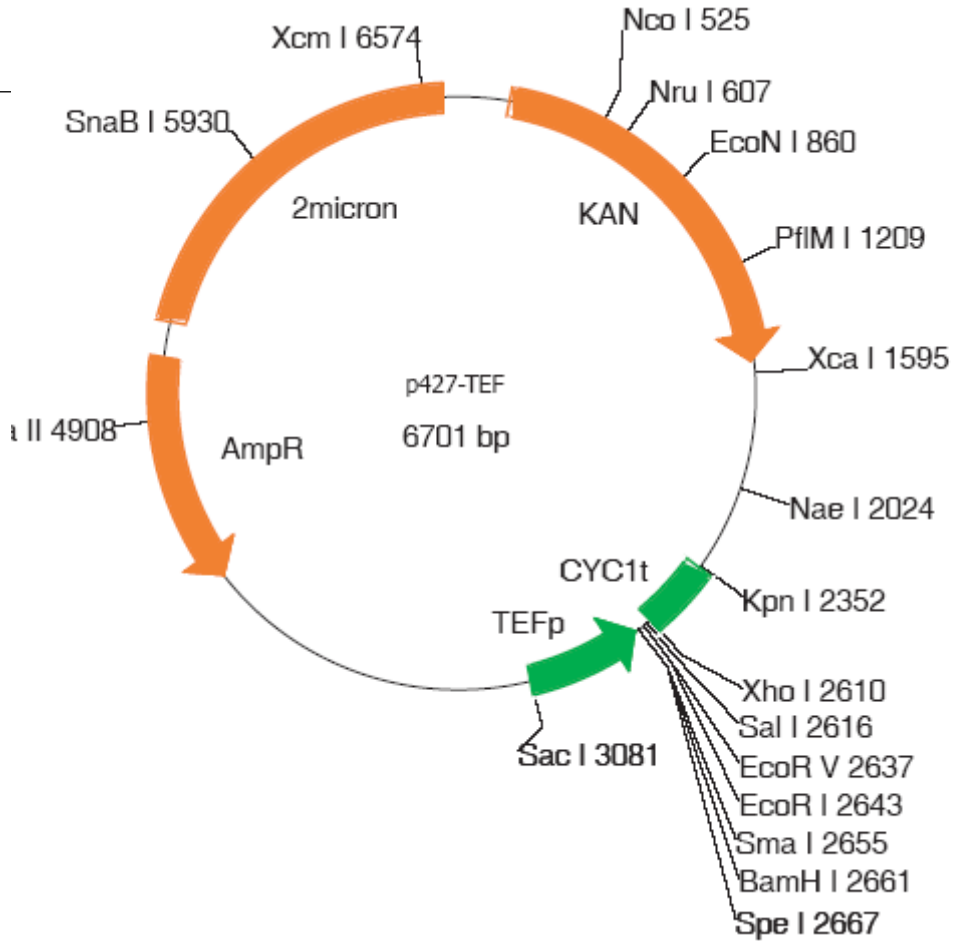


Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

S.cerevisiae Expression vectors

2µ-based multicopy vector





Vector features

TEFp TEF1 promoter (nt 2673-3081)

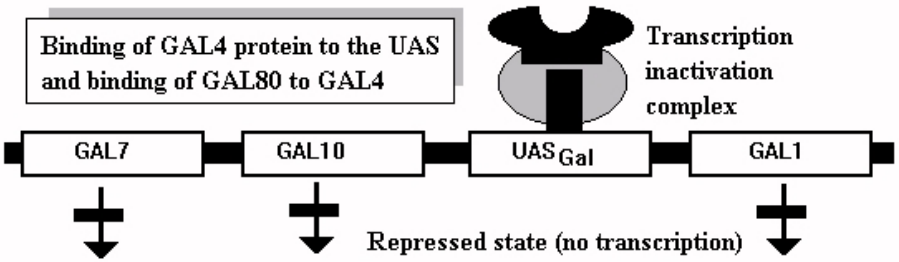
CYC1t S. cerevisiae CYC1 terminator (nt 2352-2610)

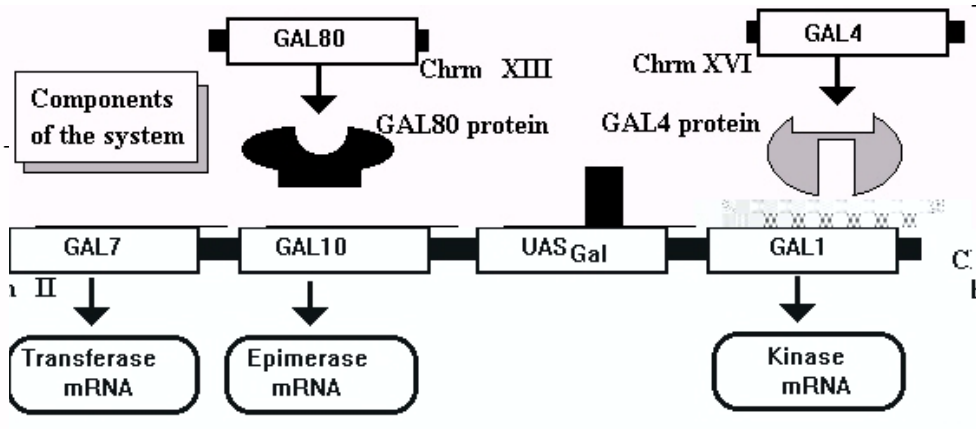
KAN Kanamycine resistance gene (aminoglycoside phosphotransferase), allows selection in yeast using 200 mg/ml G418 (nt 190-1571)

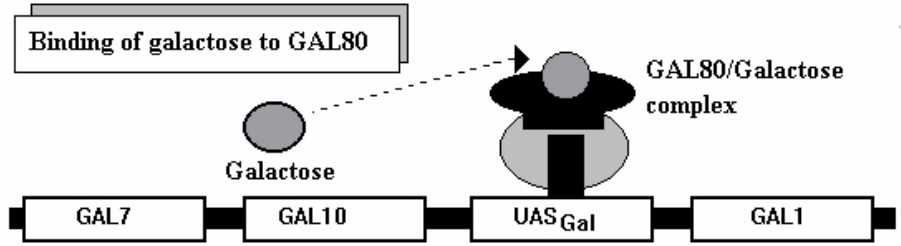
2micron Origin of replication derived from the endogenous yeast 2m circle. Allows propagation of plasmids in yeast at high copy numbers (10-50 copies/cell, nt 5291-6637)

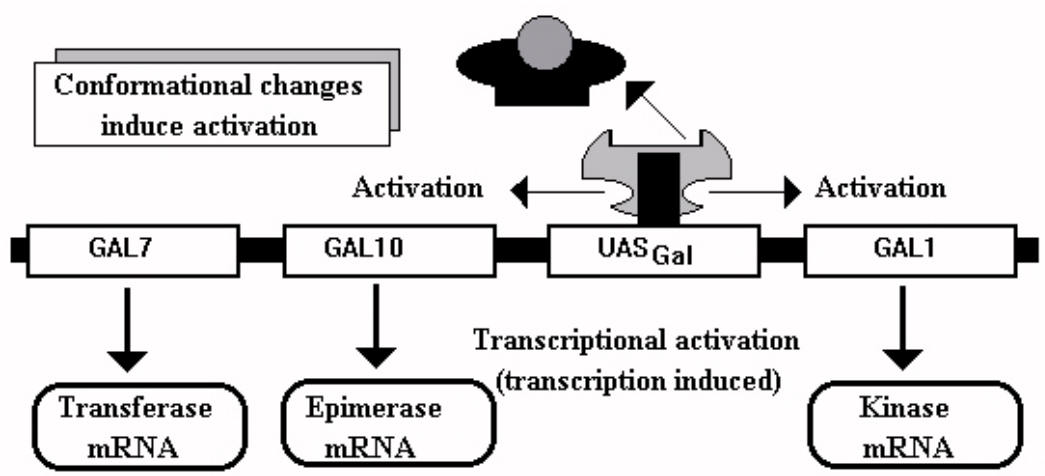
AmpR Ampicillin resistance gene (nt 4300-5158)

GAL4 regulatory system









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.

21.10.16

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, virusal, 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

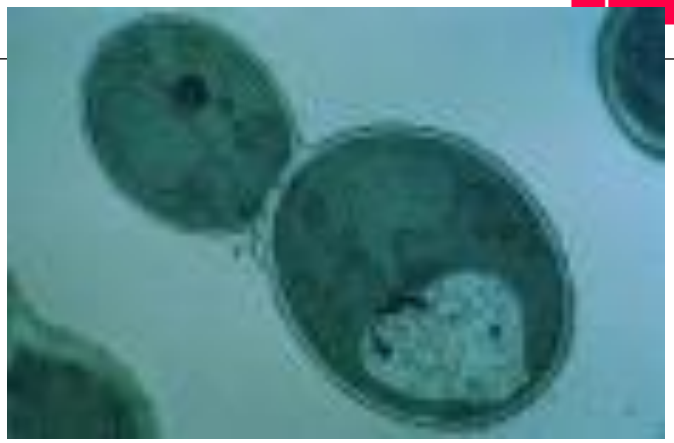
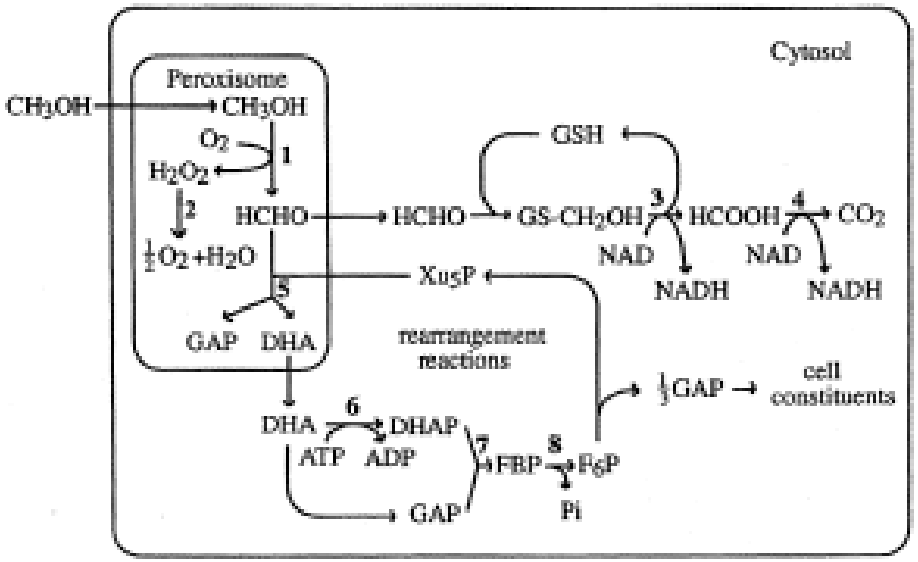


Figure 1 - High Biomass of *Pichia pastoris*



AOX1: strong expression
AOX2: weak expression

S. cerevisiae *P. pastoris*

Pichia expression tools

- **Promoters**

  AOX1, GAP

- **Selection marker**

  ***HIS4*, *ARG4*,** Zeocin^{R}, Blasticidin^{R}, Kanamycin^{R} (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 *Δ his4* 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

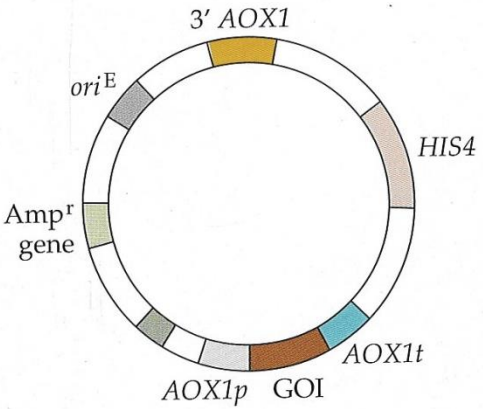
GAP

Tandem repeat multicopy integration

Ectopic integration events

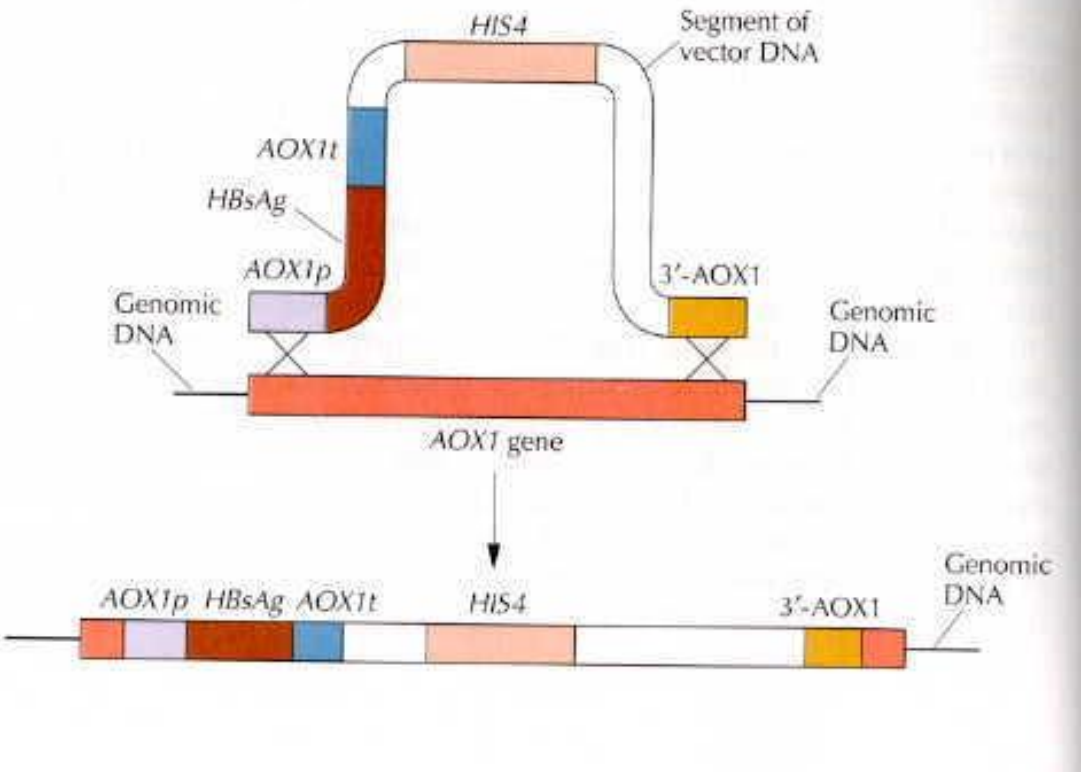
Integration vector for *Pichia pastoris*

FIGURE 7.10 *P. pastoris* integrating expression vector. The gene of interest (GOI) is cloned between the promoter *(AOX1p)* and termination-polyadenylation sequence *(AOX1t)* of the *P. pastoris* alcohol oxidase 1 gene. The *HIS4* gene encodes a functional histidinol dehydrogenase of the histidine biosynthesis pathway. The ampicillin resistance (Amp^r) gene and an origin of replication (ori^E) 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*. A double recombination event between the *AOX1p* and 3′ *AOX1* regions of the vector and the homologous segments of chromosome DNA results in the insertion of the DNA carrying the gene of interest and the *HIS4* gene.



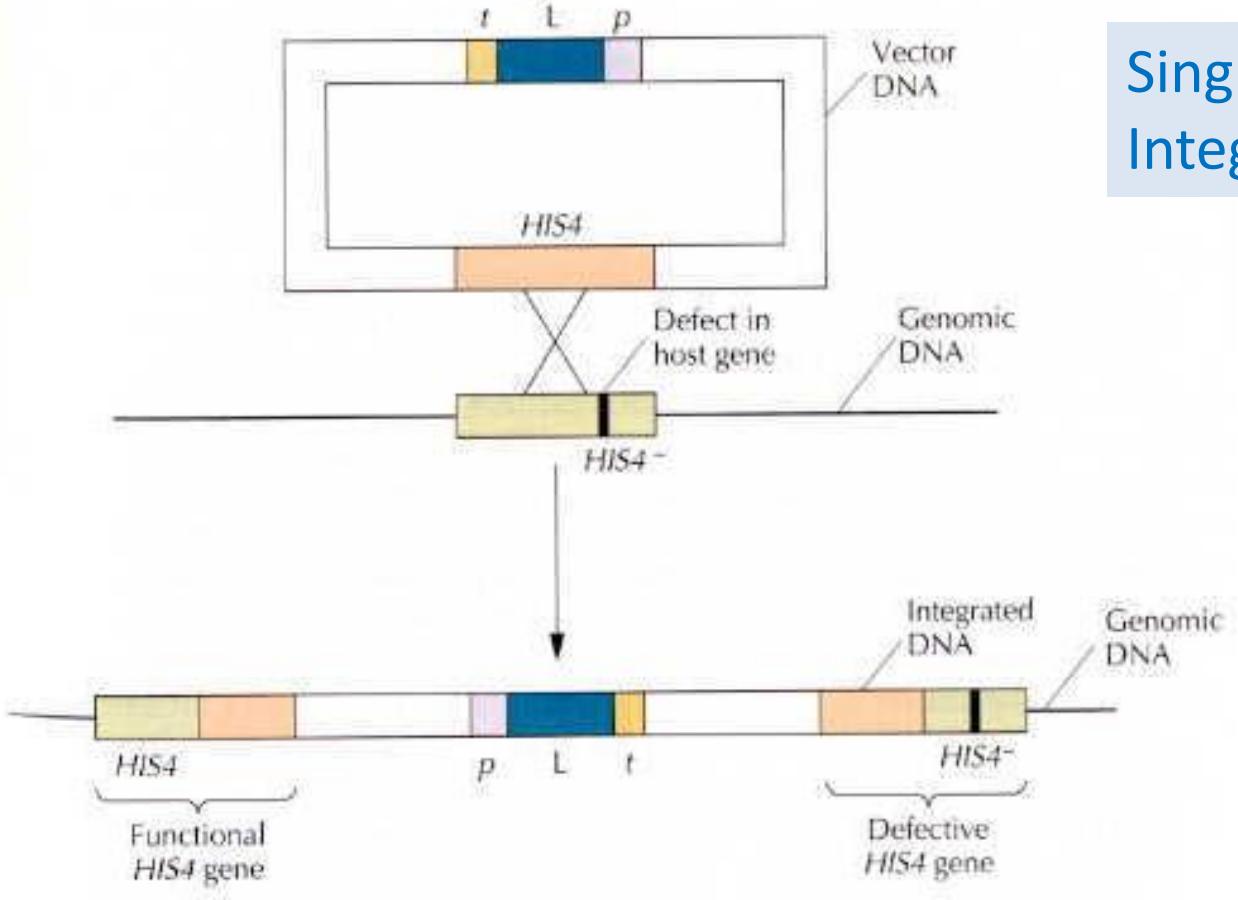
Gene Replacement

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.



Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Figure 7.7 Integration of an expression vector into the defective chromosomal *HIS4*⁻ gene of *P. pastoris*. The input DNA is a plasmid that is first introduced by transformation into a histidine-requiring *P. pastoris* strain. A single crossover within the *HIS4* gene of the plasmid and the *HIS4*⁻ gene of the host cell results in the integration of the entire plasmid, which then is flanked by the functional *HIS4* gene and the defective *HIS4*⁻ gene. The letters *p*, L, and *t* denote the *AOX1* promoter sequence, bovine lysozyme C2 cDNA, and *AOX1* termination–polyadenylation signal sequence, respectively. The black bar in the *HIS4*⁻ gene represents the defective sequence.



Single-site Integration

Single-site Integration

Gene Replacement

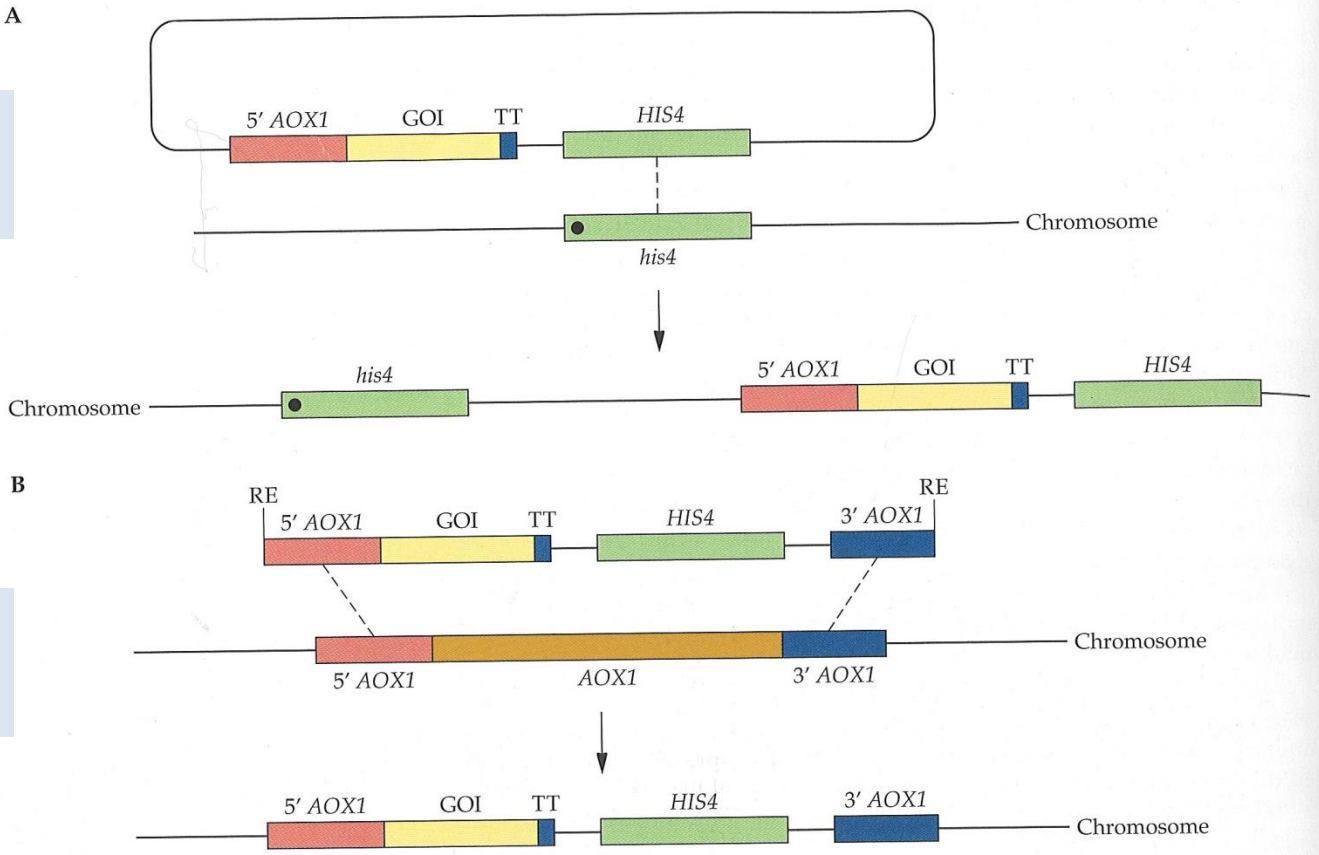
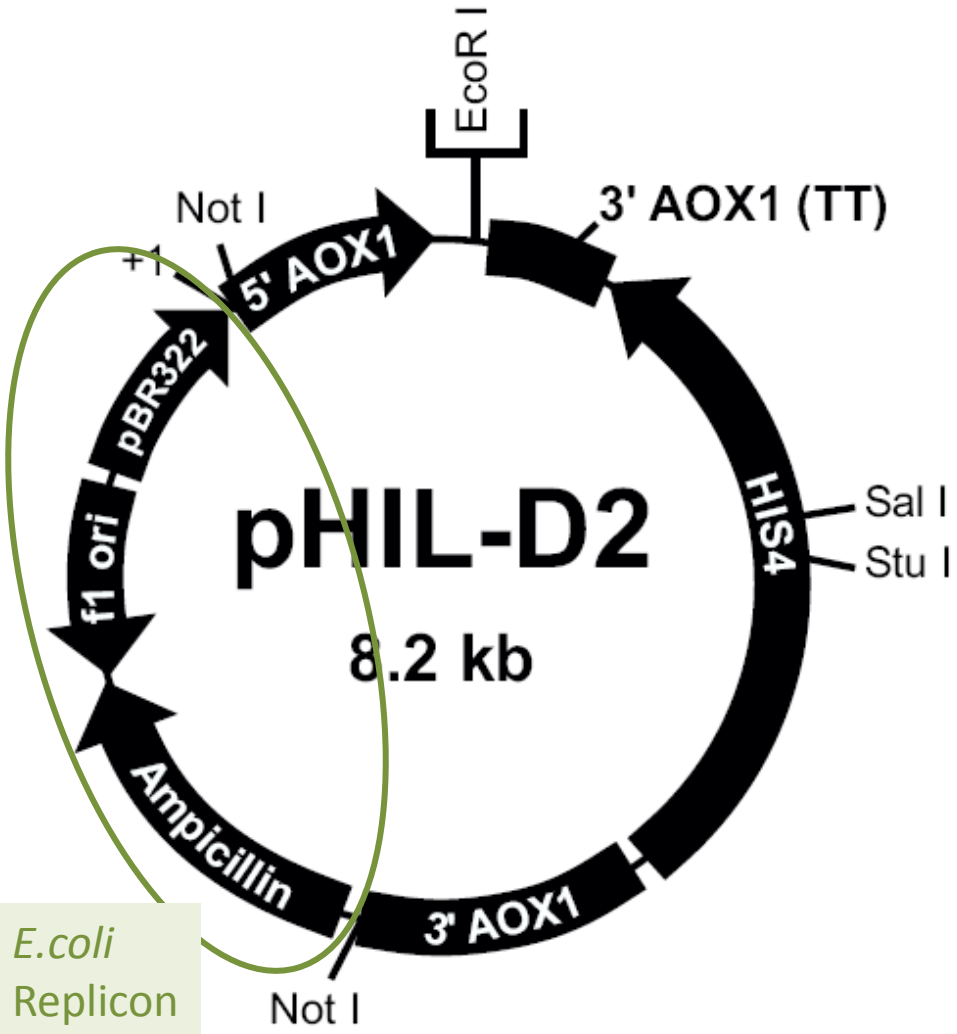


FIGURE 7.11 Integration of DNA into a specific *P. pastoris* chromosome site by single **(A)** or double **(B)** recombination. **(A)** A single recombination (dashed line) between the *HIS4* gene of an intact circular plasmid and a chromosome *his4* mutant gene results in the integration of the entire vector, including the gene of interest (GOI) with the *AOX1* promoter in the 5′ *AOX1* DNA segment and the transcription–polyadenylation sequence from the *AOX1* gene (TT), into the chromosome. The inserted DNA is flanked by recombined mutant *his4* and functional *HIS4* genes. The dot in the *his4* gene represents the mutation. **(B)** A double recombination (dashed lines) between the cloned 5′ *AOX1* and 3′ *AOX1* DNA segments of a restriction endonuclease (RE) linearized DNA fragment from the vector and the corresponding chromosome regions results in the integration of the gene of interest (GOI) with the *AOX1* promoter in the 5′ *AOX1* segment, the termination–polyadenylation sequence from the *AOX1* gene (TT), and a functional *HIS4* gene. The chromosome *AOX1* coding region is lost as a result of the recombination event.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Vector for Intracellular Expression

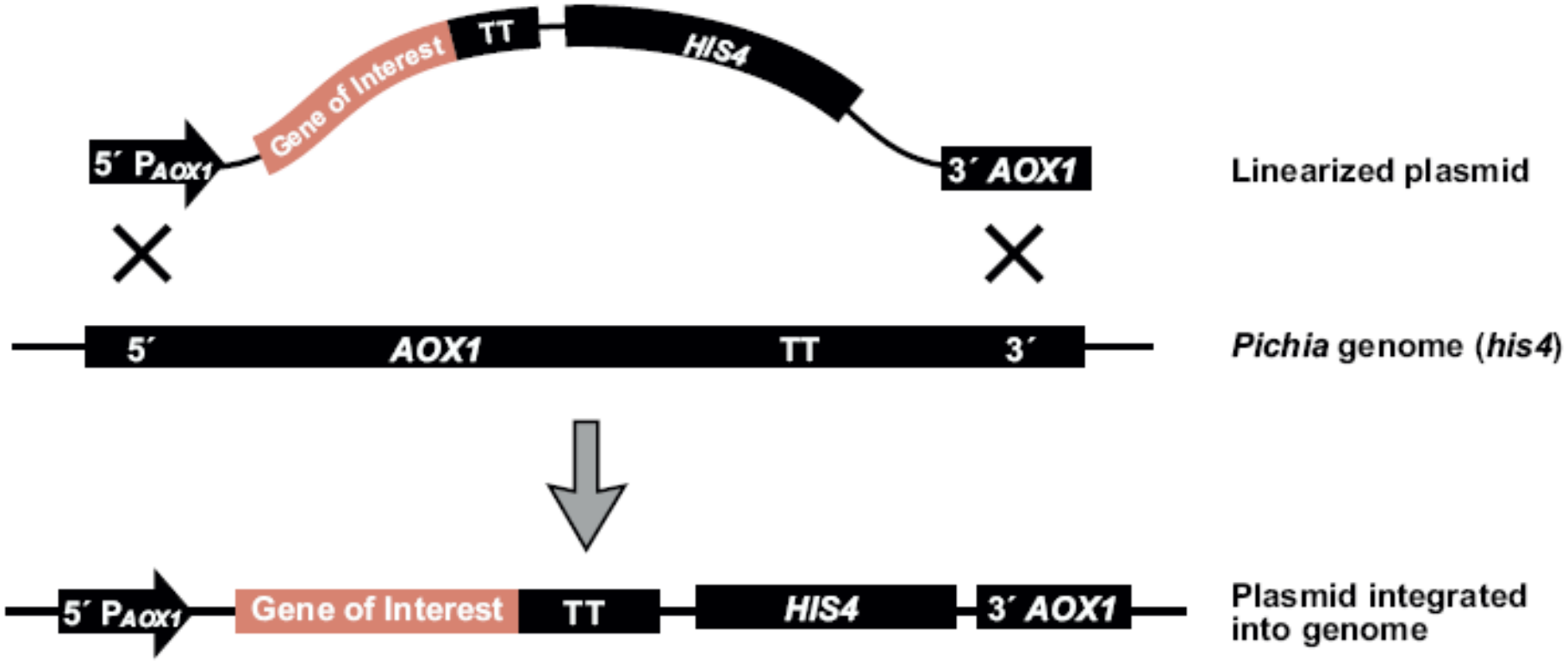


E.coli Replicon

Comments for pHIL-D2:
8209 nucleotides

5' *AOX1* promoter fragment: bases 14-941
5' *AOX1* primer site: bases 868-888
*Eco*R I 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

https://tools.thermofisher.com/content/sfs/manuals/pich_man.pdf



https://tools.thermofisher.com/content/sfs/manuals/pich_man.pdf

Vector for Intracellular Expression

pHILD2

Promoter (app. 900 bp)

785

AOX1 mRNA 5´ end (837)

ACAGGCAATA TATAAACAGA AGGAAGCTGC CCTGTCTTAA ACCTTTTTTT TTATCATCAT TATTAGCTTA

5´ *AOX1* Primer Site (868-888)

CTTTCATAAT TGCGACTGGT TCCAATTGAC AAGCTTTTGA TTTTAACGAC TTTTAACGAC AACTTGAGAA

Remaining 'A' of the native *AOX1* 'ATG'

*Eco*R I

Stop-codon

GATCAAAAAA CAACTAATTA TTCGAAACGA GGAATTCGCC TTAGACATGA CTGTTCCTCA GTTCAAGTTG

3´ *AOX1* Primer Site (1036-1056)

GGCATTACGA GAAGACCGGT CTTGCTAGAT TCTAATCAAG AGGATGTCAG AATGCCATTT GCCTGAGAGA

AOX1 mRNA 3´ end (1127)

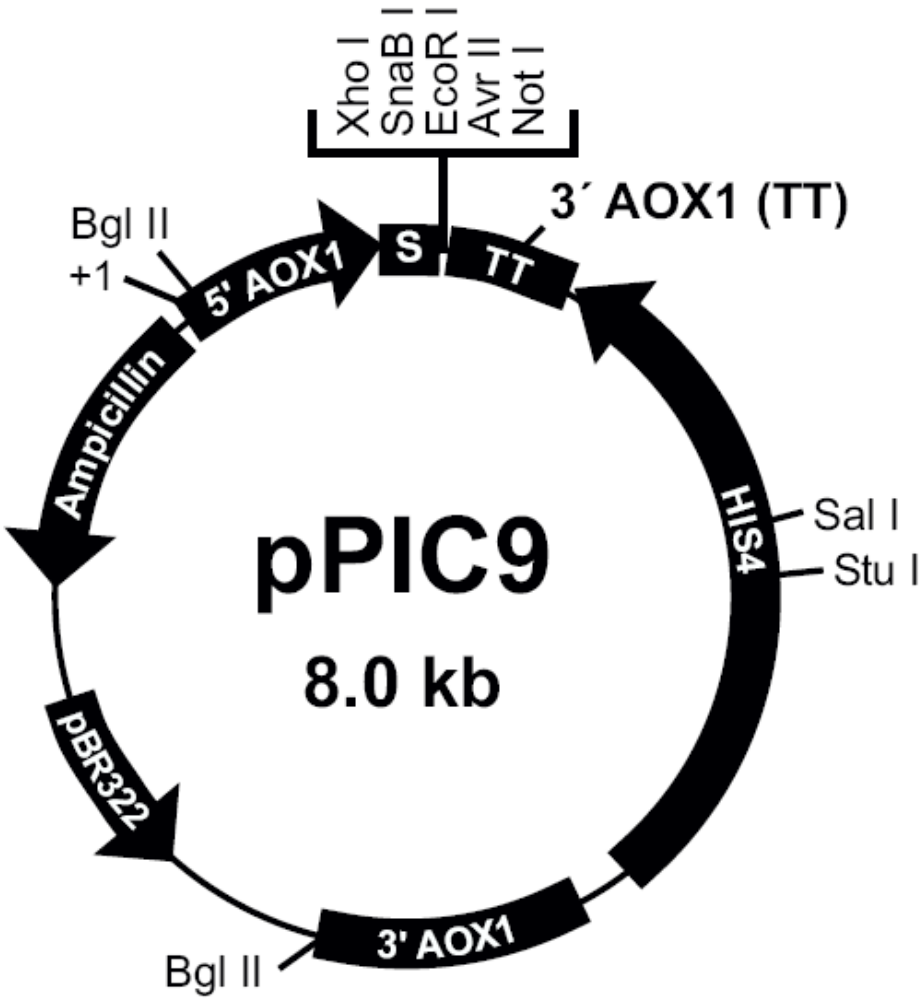
TGCAGGCTTC ATTTTTGATA CTTTTTTATT TGTAACCTAT ATAGTATAGG ATTTTTTTTG TCA

mRNA

- 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.

https://tools.thermofisher.com/content/sfs/manuals/pich_man.pdf

Vector for Secretory Expression



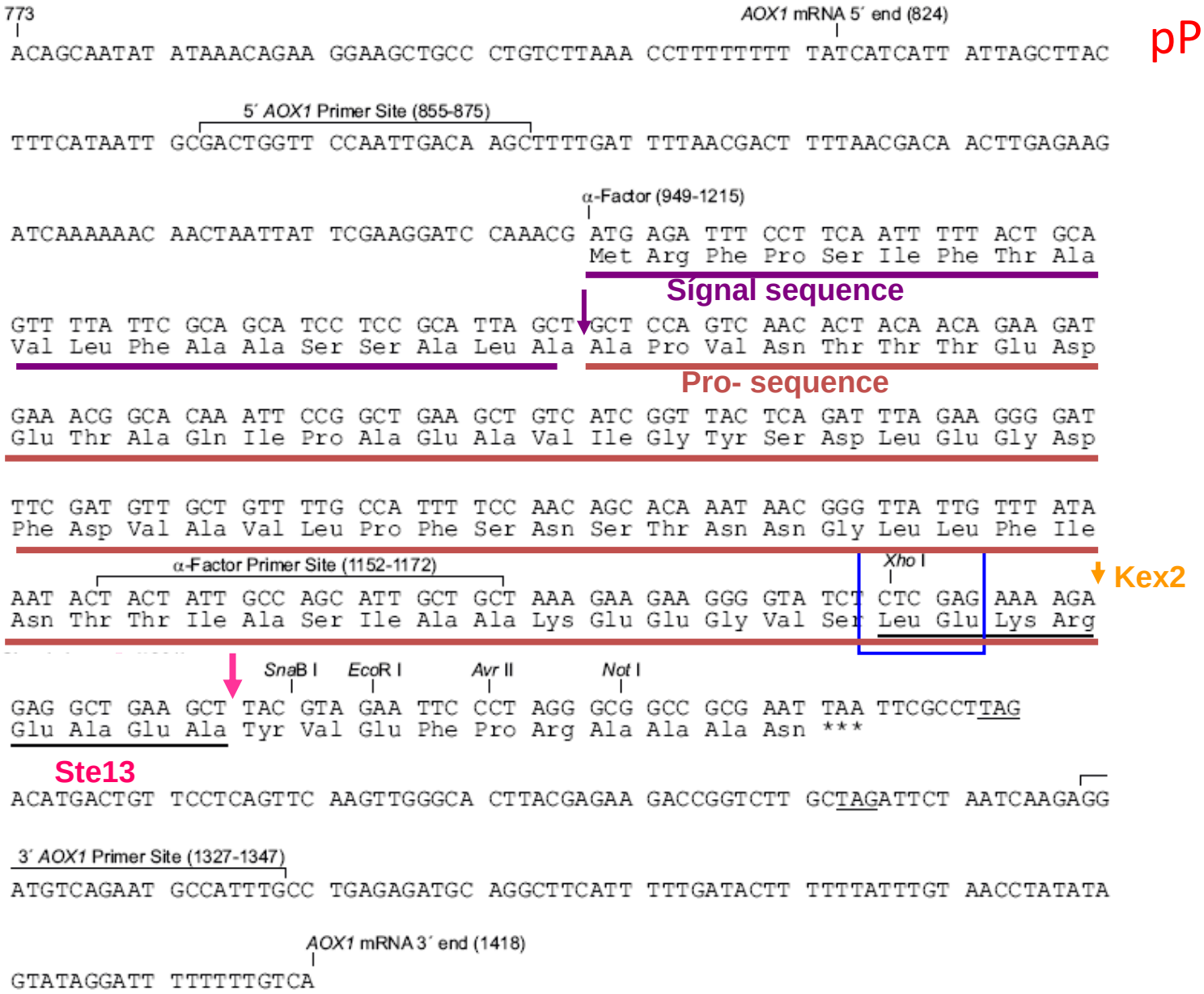
Comments for pPIC9:
8023 nucleotides

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

https://tools.thermofisher.com/content/sfs/manuals/pich_man.pdf

Vector for Secretory Expression

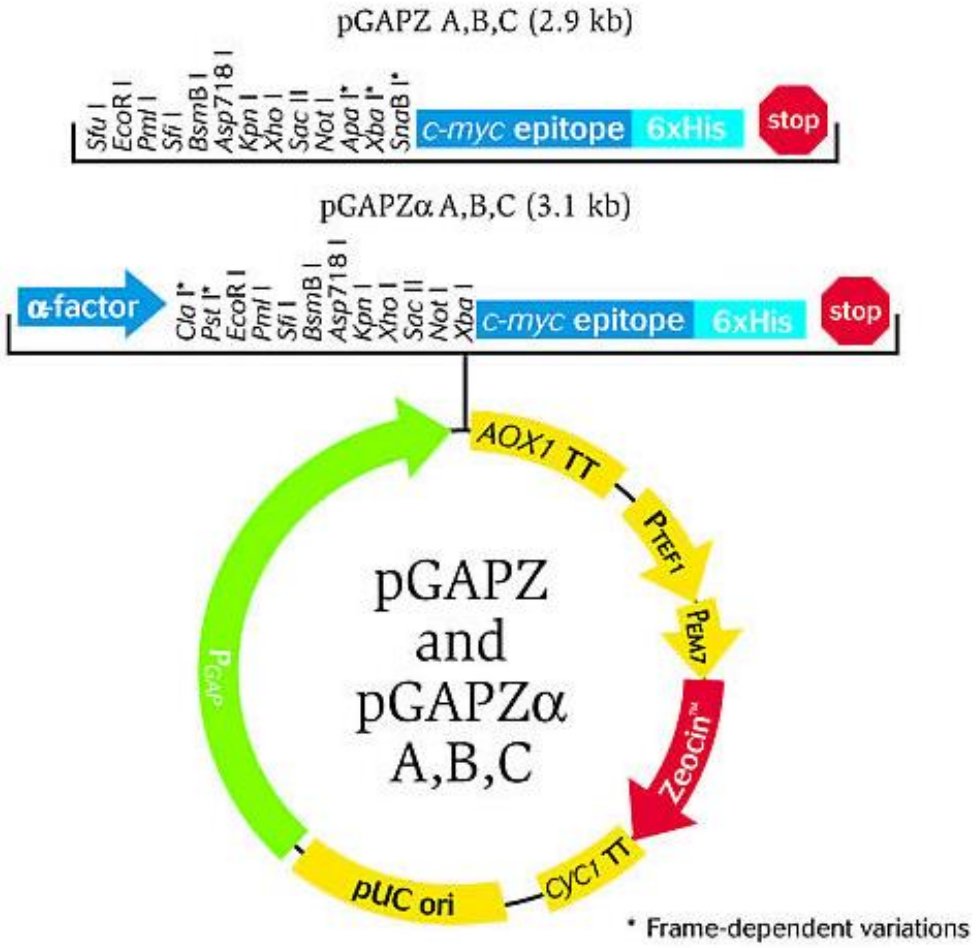
pPIC9



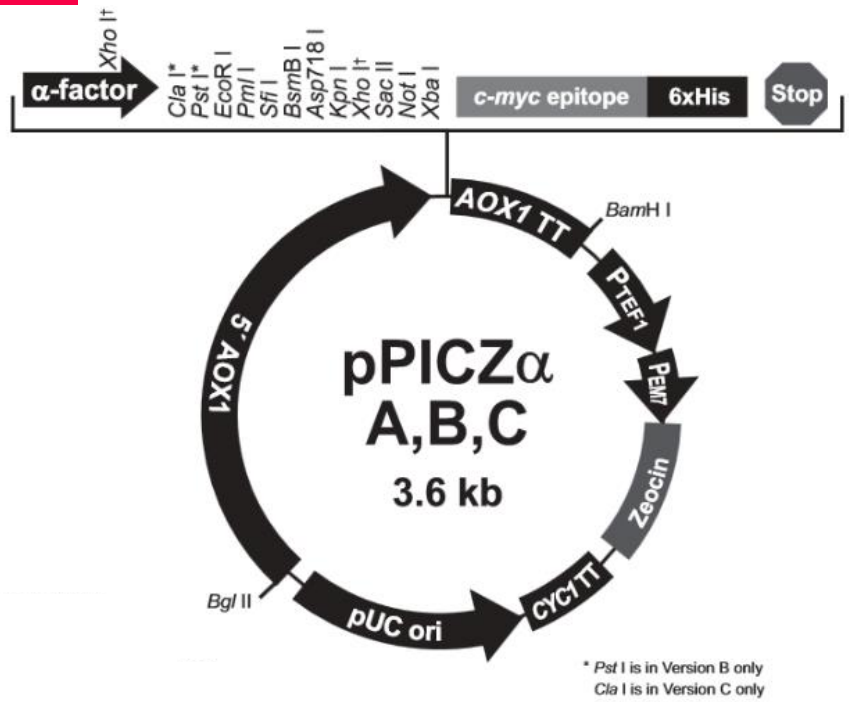
https://tools.thermofisher.com/content/sfs/manuals/pich_man.pdf

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"



https://www.thermofisher.com/order/catalog/product/V20020

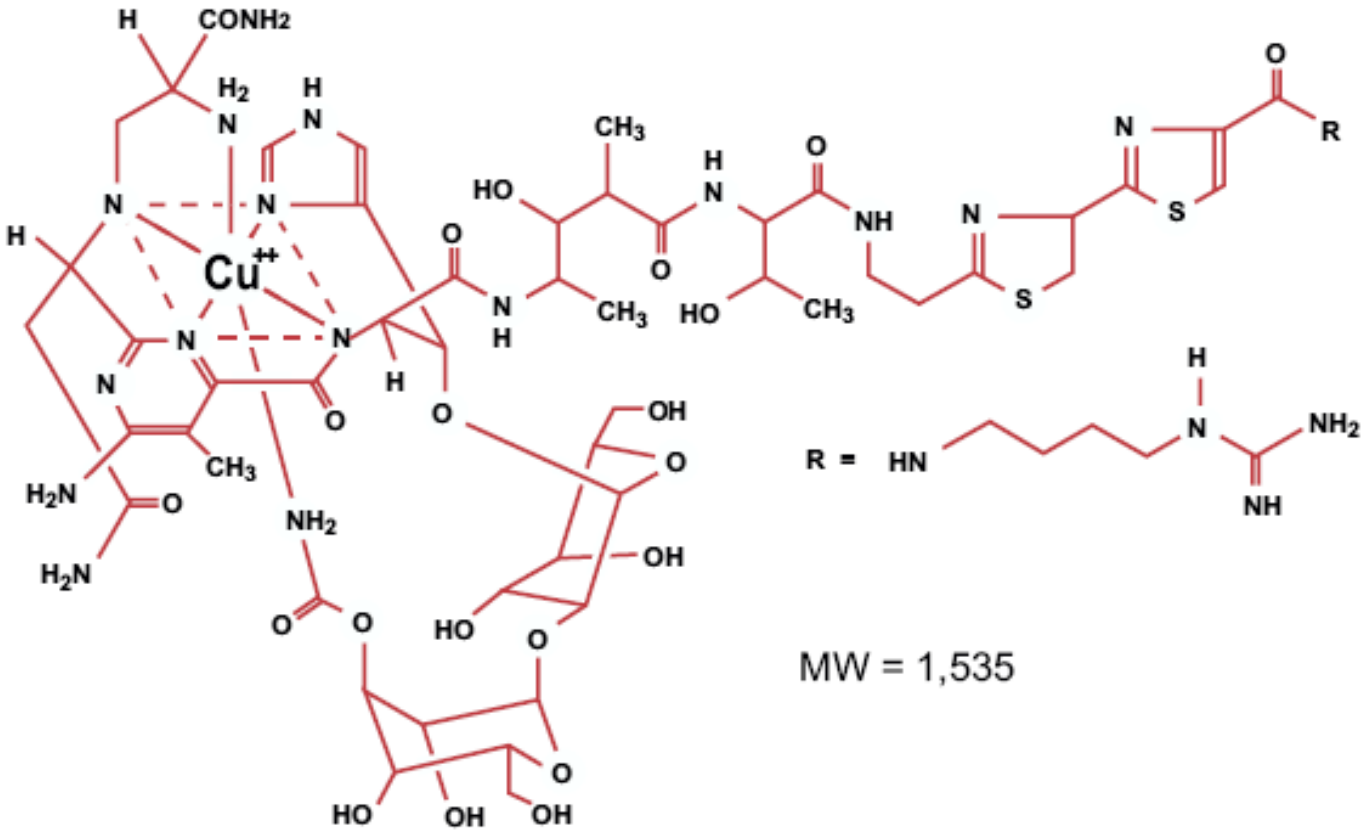


5´ *AOX1* promoter region: bases 1-941
5´ *AOX1* priming site: bases 855-875
α-factor signal sequence: bases 941-1207
Multiple cloning site: bases 1208-1276
c-myc epitope: bases 1275-1304
Polyhistidine (6xHis) tag: bases 1320-1337
3´ *AOX1* priming site: bases 1423-1443
AOX1 transcription termination region: bases 1341-1682
TEF1 promoter: bases 1683-2093
EM7 promoter: bases 2095-2162
Sh ble ORF: bases 2163-2537
CYC1 transcription termination region: bases 2538-2855
pUC origin: bases 2866-3539 (complementary strand)

Feature	Benefit
5´ *AOX1* promoter	A 942 bp fragment containing the *AOX1* promoter that allows methanol-inducible, high-level expression of the gene of interest in *Pichia*. Targets plasmid integration to the *AOX1* locus.
α-factor secretion signal (from *Saccharomyces cerevisiae*)	Allows for efficient secretion of most proteins from *Pichia*.
Multiple cloning site	Allows insertion of your gene into the expression vector.
c-myc epitope (Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu)	Permits detection of your recombinant fusion protein with the Anti-*myc* Antibody or Anti-*myc*-HRP Antibody (Evans *et al.*, 1985). See page viii for ordering information.
C-terminal polyhistidine (6×His) tag	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 (Lindner *et al.*, 1997) and the Anti-His(C-term)-HRP Antibody. See page viii for ordering information.
AOX1 transcription termination (TT) region	Native transcription termination and polyadenylation signal from *AOX1* gene (~260 bp) that permits efficient 3´ mRNA processing, including polyadenylation, for increased mRNA stability.
TEF1 promoter (GenBank accession nos. D12478, D01130)	Transcription elongation factor 1 gene promoter from *Saccharomyces cerevisiae* that drives expression of the Zeocin™ resistance gene in *Pichia*.
EM7 promoter	Synthetic prokaryotic promoter that drives constitutive expression of the Zeocin™ resistance gene in *E. coli*.
Zeocin™ resistance gene (*Sh ble*)	Allows selection of transformants in *E. coli* and *Pichia*.
CYC1 transcription termination region (GenBank accession no. M34014)	3´ end of the *Saccharomyces cerevisiae CYC1* gene that allows efficient 3´ mRNA processing of the Zeocin™ resistance gene for increased stability.
pUC origin	Allows replication and maintenance of the plasmid in *E. coli*.

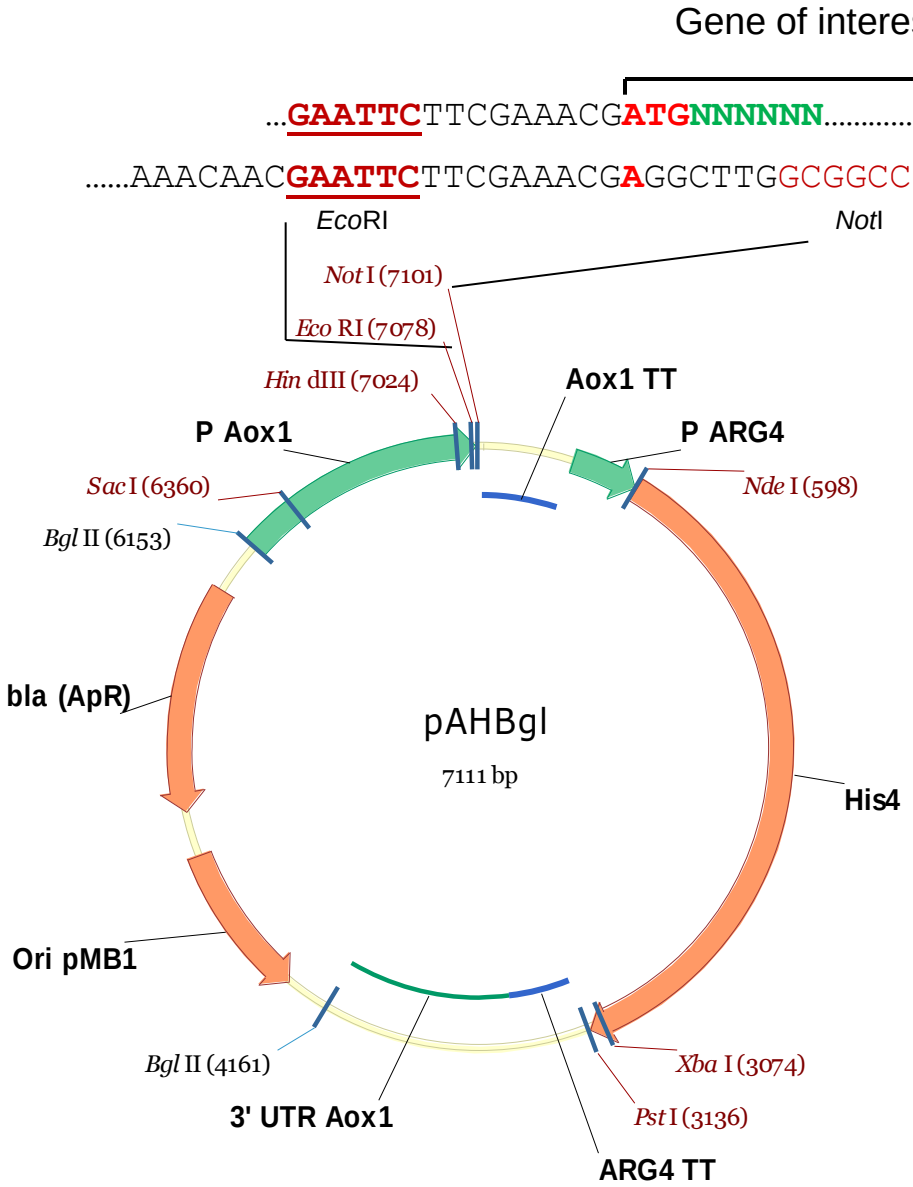
http://tools.thermofisher.com/content/sfs/manuals/ppiczalpha_man.pdf

Zeocin



http://tools.thermofisher.com/content/sfs/manuals/ppiczalpha_man.pdf

P.pastoris vectors for intracellular expression



p A H Bgl

Promoter:

A → *AOX1*

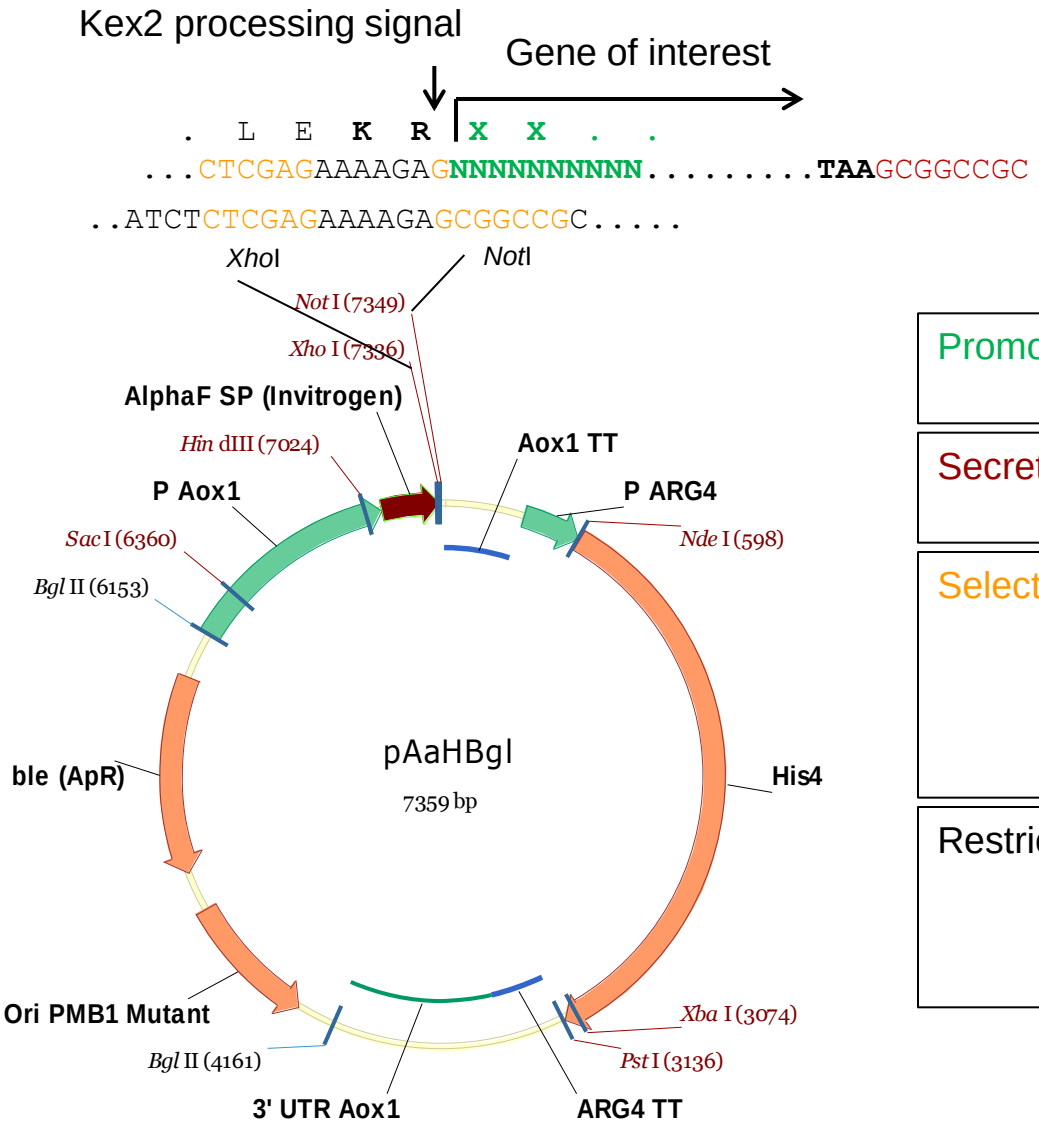
Selection Marker:

H → *HIS4*
Z → Zeocin^R
A → *ARG4*
K → Kanamycin^R

Restriction site:

Bgl → *Bgl*II
Sph → *Sph*I
Swa → *Swa*I

P.pastoris vectors for secretory expression



p A a H Bgl

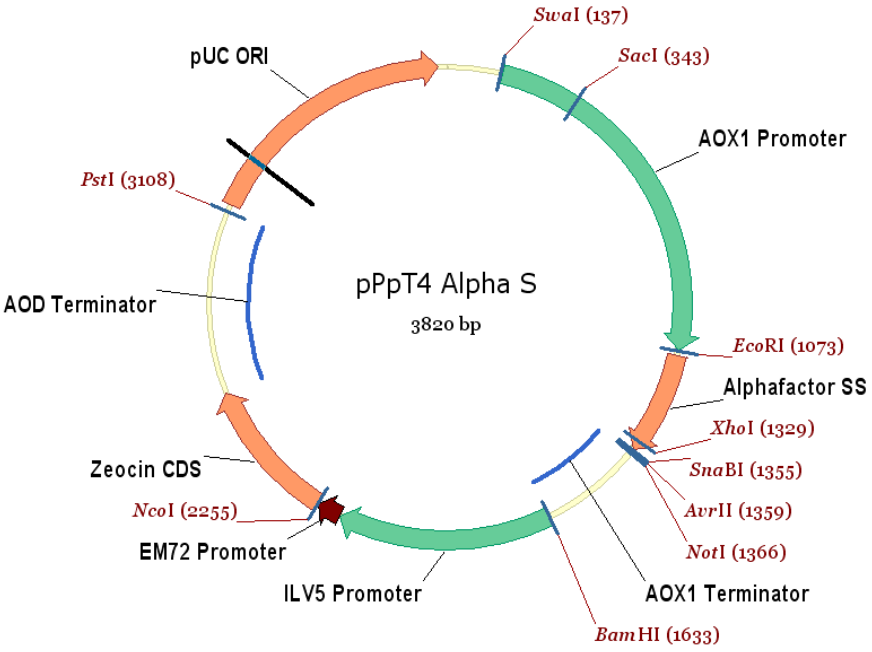
Promoter:
A→ *AOX1*

Secretion Signal:
a → *alpha factor*

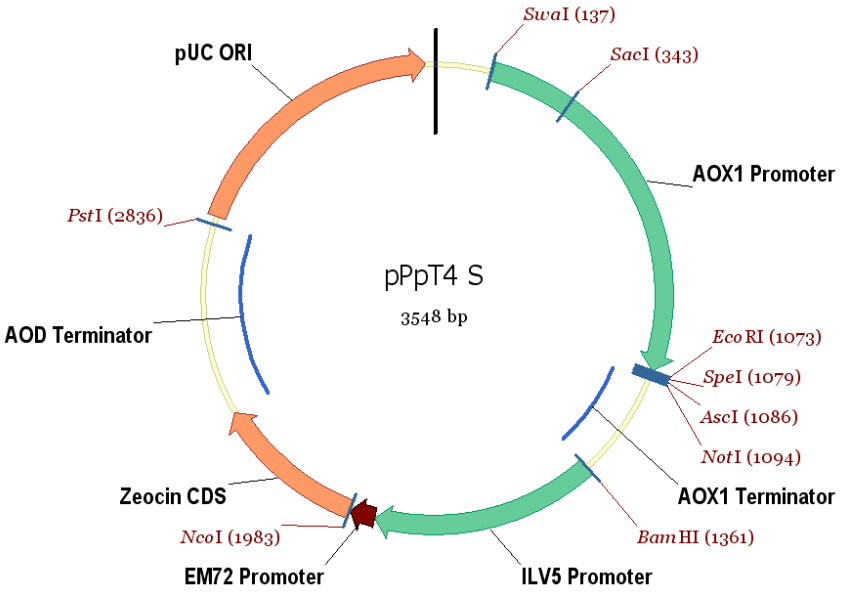
Selection Marker:
H → *HIS4*
Z → Zeocin^R
A → *ARG4*
K → Kanamycin^R

Restriction site:
Bgl →*Bgl*II
Sph →*Sph*I
Swa → *Swa*I

Vectors for multiple Integrations



secretory

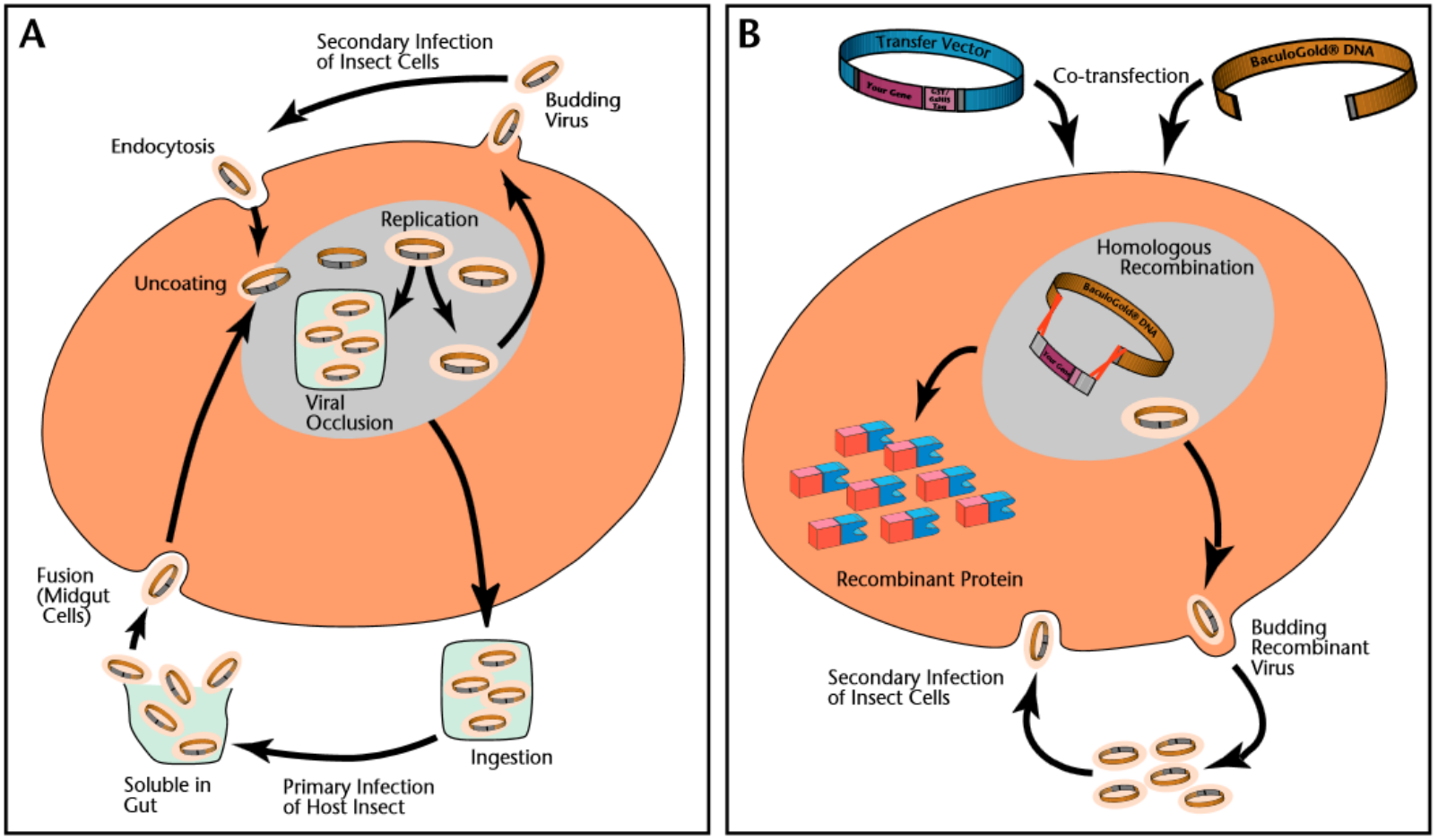


intracellular

Protein expressed	Expression Level (mg/L)	Reference
Bacterial proteins		
Tetanus toxin fragment C	12,000	Clare, J.J. *et al.* (1991) *Bio/Technology* 9: 455–460
α-amylase	2,500	Paifer, E. *et al.* (1994) *Yeast* 10: 1415–1419
T2A peroxidase	2,470	Thomas, L. *et al.* (1998) *Can. J. Microbiol.* 44: 364–372
C. botulinum neurotoxin fragment	78	Smith, L.A. (1998) *Toxicon* 36: 1539–1548
Yeast proteins		
Catalase L	2,300	Calera, J.A. *et al.* (1997) *Infect. Immun.* 65: 4718–4724
Glucoamylase	400	Fierobe, H.-P. *et al.* (1997) *Protein Expr. Purif.* 9: 159–170
Lipase	60	Minning, S. *et al.* (1998) *J. Biotechnol.* 66: 147–156
Plant proteins		
Hydroxynitrile lyase	22,000	Hasslacher, M. *et al.* (1997) *Protein Expr. Purif.* 11: 61–71
Wheat lipid transfer protein	720	Klein, C. *et al.* (1998) *Protein Expr. Purif.* 13: 73–82
Aeroallergen	60	Huecas, S. *et al.* (1999) *Eur. J. Biochem.* 261: 539–546.
Invertebrate proteins		
Hirudin	1,500	Rosenfeld, S.A. *et al.* (1996) *Protein Expr. Purif.* 8: 476–482.
Spider dragline silk protein	663	Fahnestock, S.R. *et al.* (1997) *Appl. Micro. Biotechnol.* 47: 33–39
Honeybee olfactory protein	200	Danty, E. *et al.* (1999) *J. Neuroscience* 19: 7468-7475
Mammalian proteins		
Mouse gelatin	14,800	Werten, M.W. *et al.* (1999) *Yeast* 15: 1087-1096
Porcine carboxypeptidase B	200	Ventura, S. *et al.* (1999) *J. Biol. Chem.* 274: 19925-33
Human tumor necrosis factor	10,000	Sreekrishna, K. *et al.* (1989) *Biochemistry* 28: 4117-4125
Human IGF-1	600	Brierley, R.A. (1998) *Methods Mol. Biol.* 103: 149–177
Human CD38	455	Munshi, C.B. (1997) *Methods Enzymol.* 280: 318–330
15N-Interferon τ	10	Johnson, T.M. *et al.* (1999) *J. Interferon Cytokine Res.* 19: 631-636

Baculovirus Expression system

Autographa californica nuclear polyhedrosis virus (AcNPV)

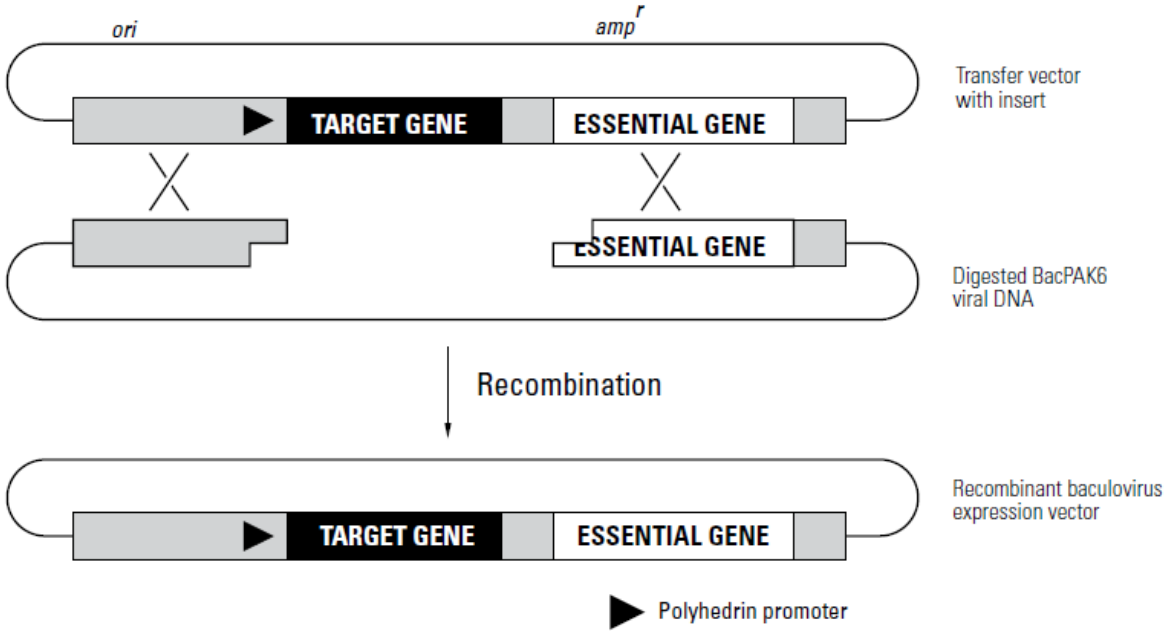


Heterologous genes are cloned into transfer vectors

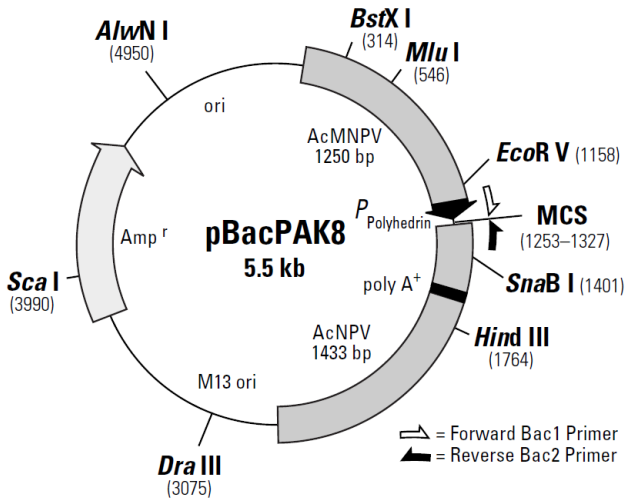
Co-transfection of the transfer vector and AcNPV DNA into *Spodoptera frugiperda* (*Sf*) cells → recombination between homologous sites

https://www.bdbiosciences.com/documents/Baculovirus_vector_system_manual.pdf

Baculovirus Expression System



Transfer Vector



Transfer of a target gene to the Baculovirus expression vector by forced recombination between a transfer vector and BacPAK6 viral DNA.

http://www.clontech.com/AT/Products/Protein_Expression_and_Purification/Baculovirus_Expression

Baculovirus Expression System

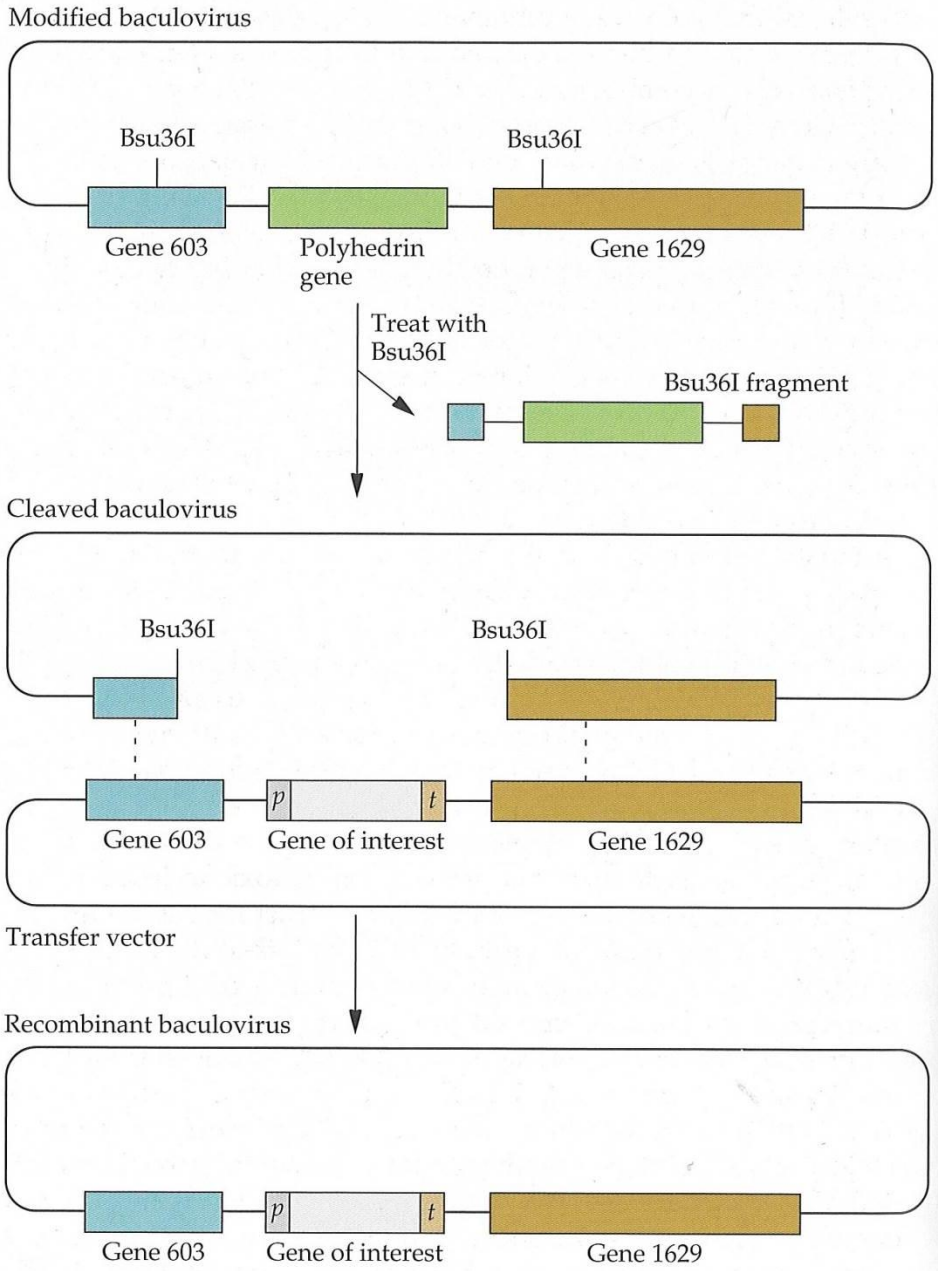
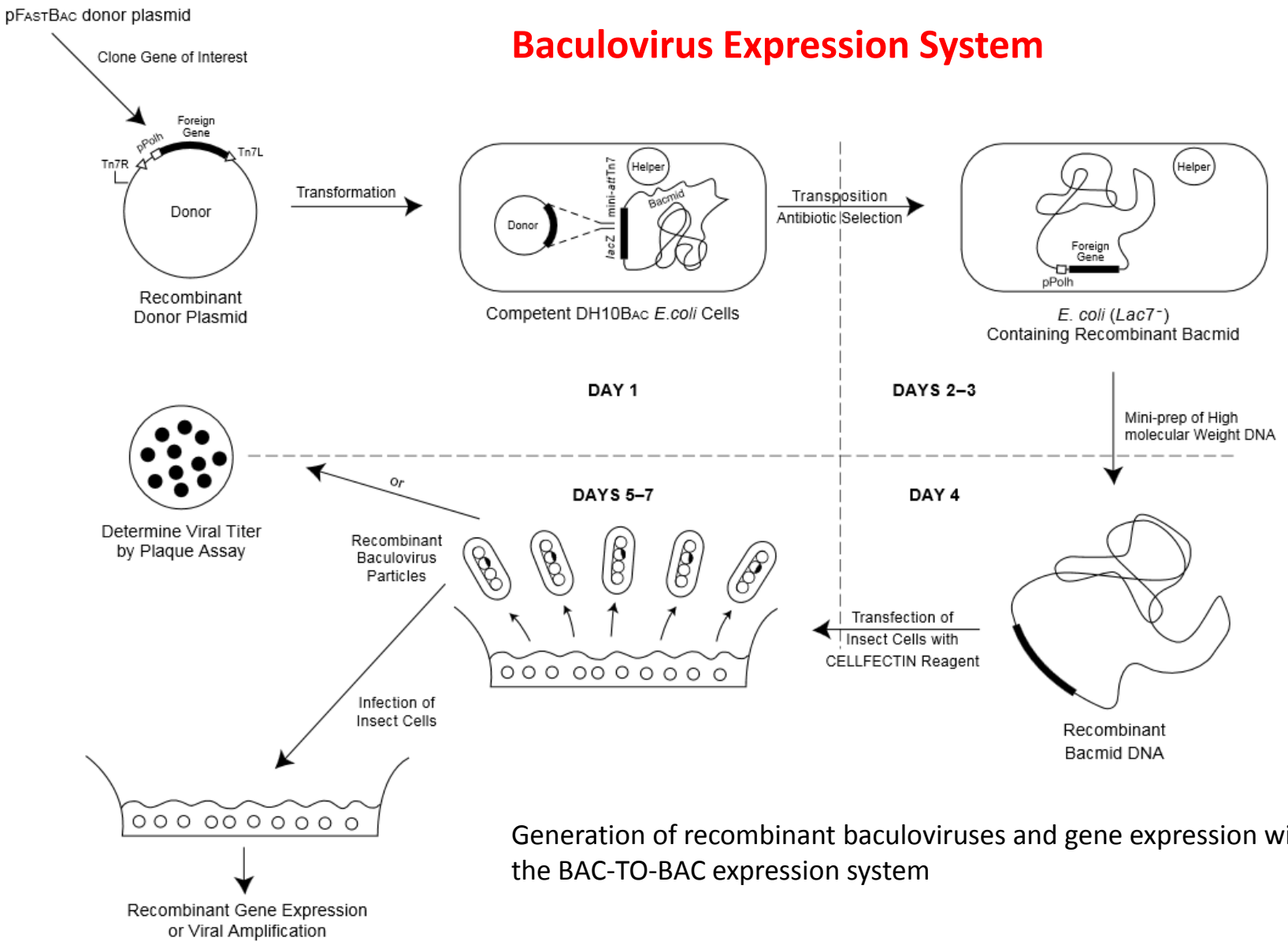


FIGURE 7.18 Production of recombinant baculovirus. Single Bsu36I sites are engineered into gene 603 and a gene (1629) that is essential for AcMNPV replication. These genes flank the polyhedrin gene in the AcMNPV genome. After a baculovirus with two engineered Bsu36I sites is treated with Bsu36I, the segment between the Bsu36I sites is deleted. Insect cells are cotransfected with a Bsu36I-treated baculovirus DNA and a transfer vector with a gene of interest under the control of the promoter *(p)* and terminator *(t)* elements of the polyhedrin gene and the complete sequences of both genes 603 and 1629. A double crossover event (dashed lines) generates a recombinant baculovirus with a functional gene 1629. With this system, almost all of the progeny baculoviruses are recombinant.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Baculovirus Expression System



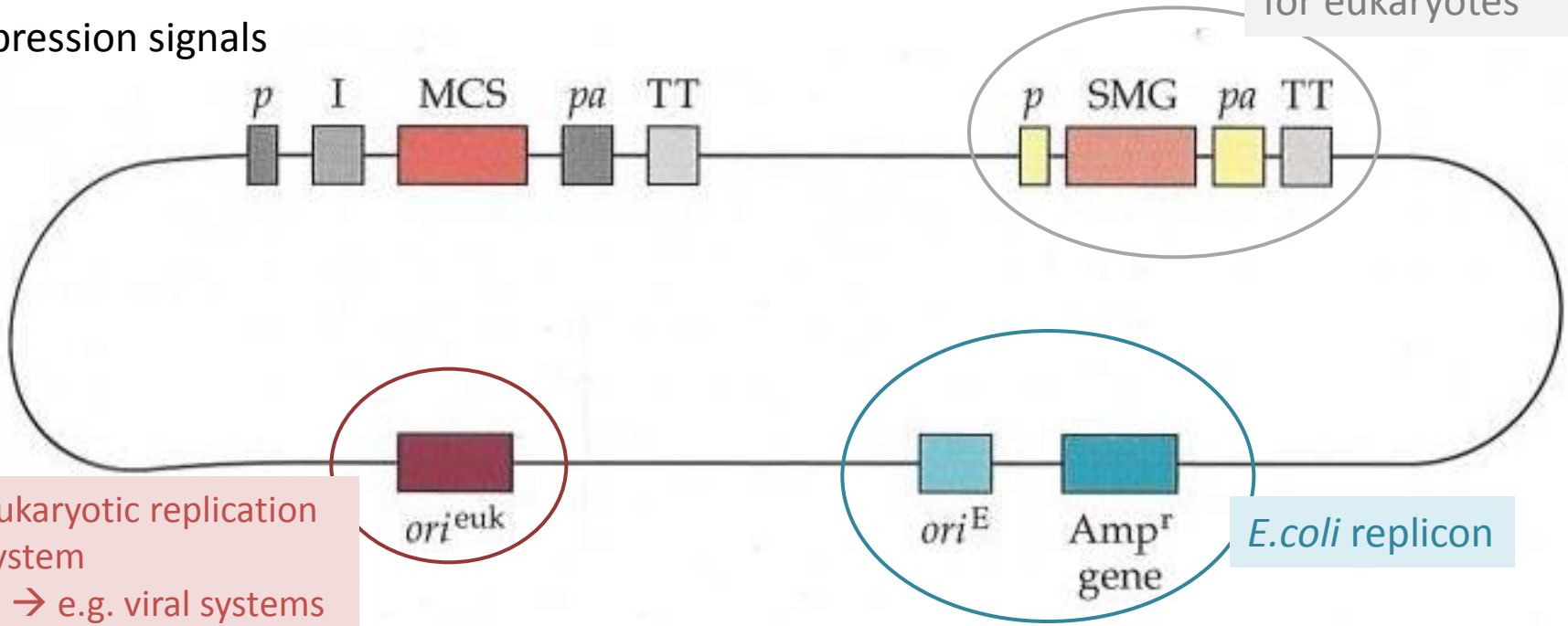
Generation of recombinant baculoviruses and gene expression with the BAC-TO-BAC expression system

https://tools.lifetechnologies.com/content/sfs/manuals/bevtest.pdf

Mammalian Expression System

Selection marker for eukaryotes

Expression signals



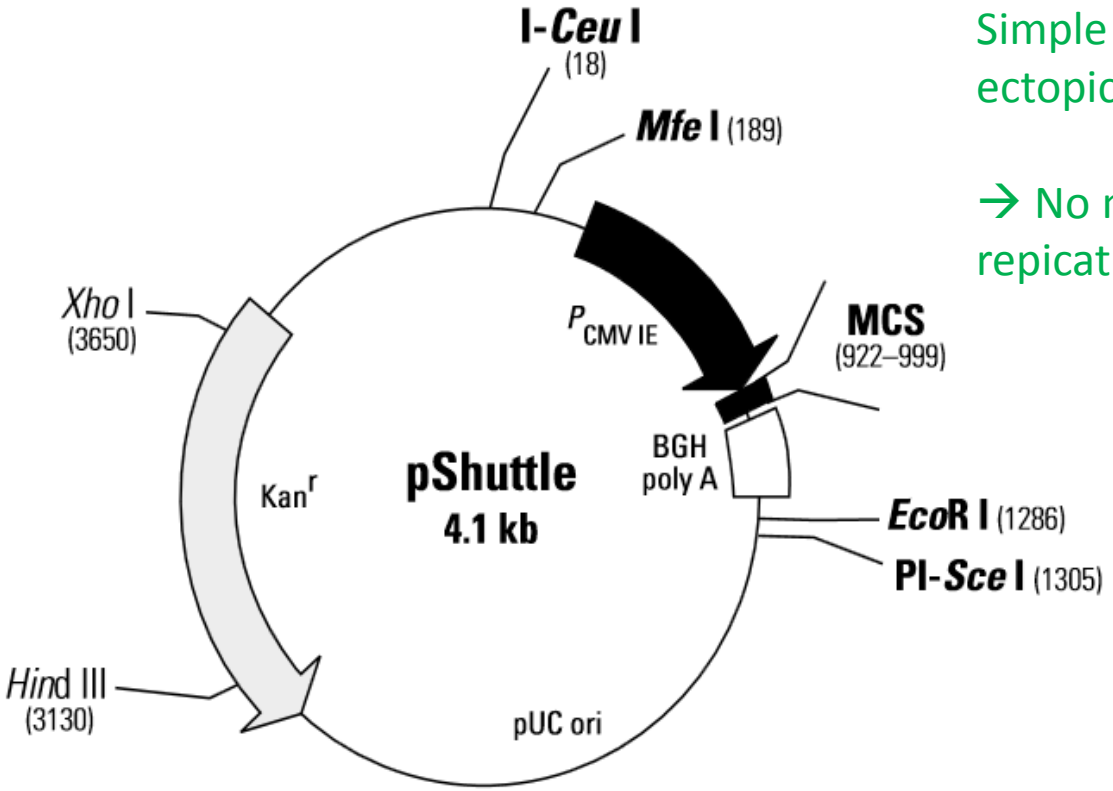
Eukaryotic replication system
→ e.g. viral systems

E.coli replicon

FIGURE 7.23 Generalized mammalian expression vector. The multiple cloning site (MCS) and selectable marker gene (SMG) are under the control of eukaryotic promoter *(p)*, polyadenylation *(pa)*, and termination of transcription (TT) sequences. An intron (I) enhances the production of heterologous protein. Propagation of the vector in *E. coli* and mammalian cells depends on the origins of replication ori^{E} and ori^{euk}, respectively. The ampicillin resistance (Amp^{r}) gene is used for selecting transformed *E. coli*.

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

Mammalian Expression System



Simple Plasmid for ectopic integration

→ No mammalian repication system

pShuttle MCS

920

TGGCTAGCGTTTAAACGGGCCCTCTAGACTCGAGCGGCCGCCACTGTGCTGG

Nhe I Apa I Xba I Xho I Not I BstX I

972

STOP (ORF 1) STOP (ORF 2) STOP (ORF 3)

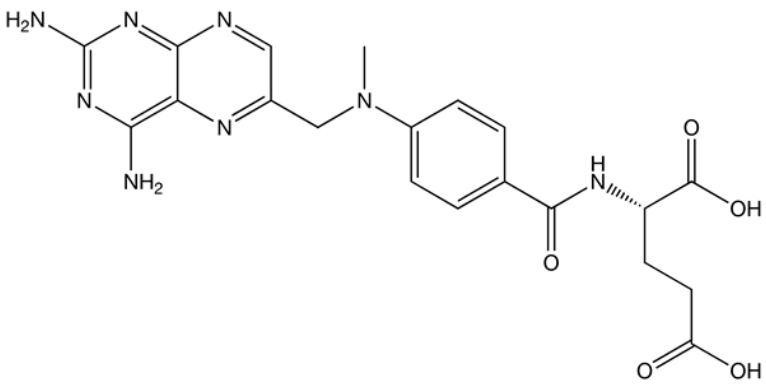
ATGATCCGAGCTCGGTACCAAGCTTAAGTTTAAACCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC

Kpn I Hind III Afl II

Expression strategies for mammalians

Selection needed in order to find clones positioned at transcription-active sites

DHFR:
Selection for high expression with methotrexate
→ Incresed resistance to methodrexate: high probability of high expression of targed protein



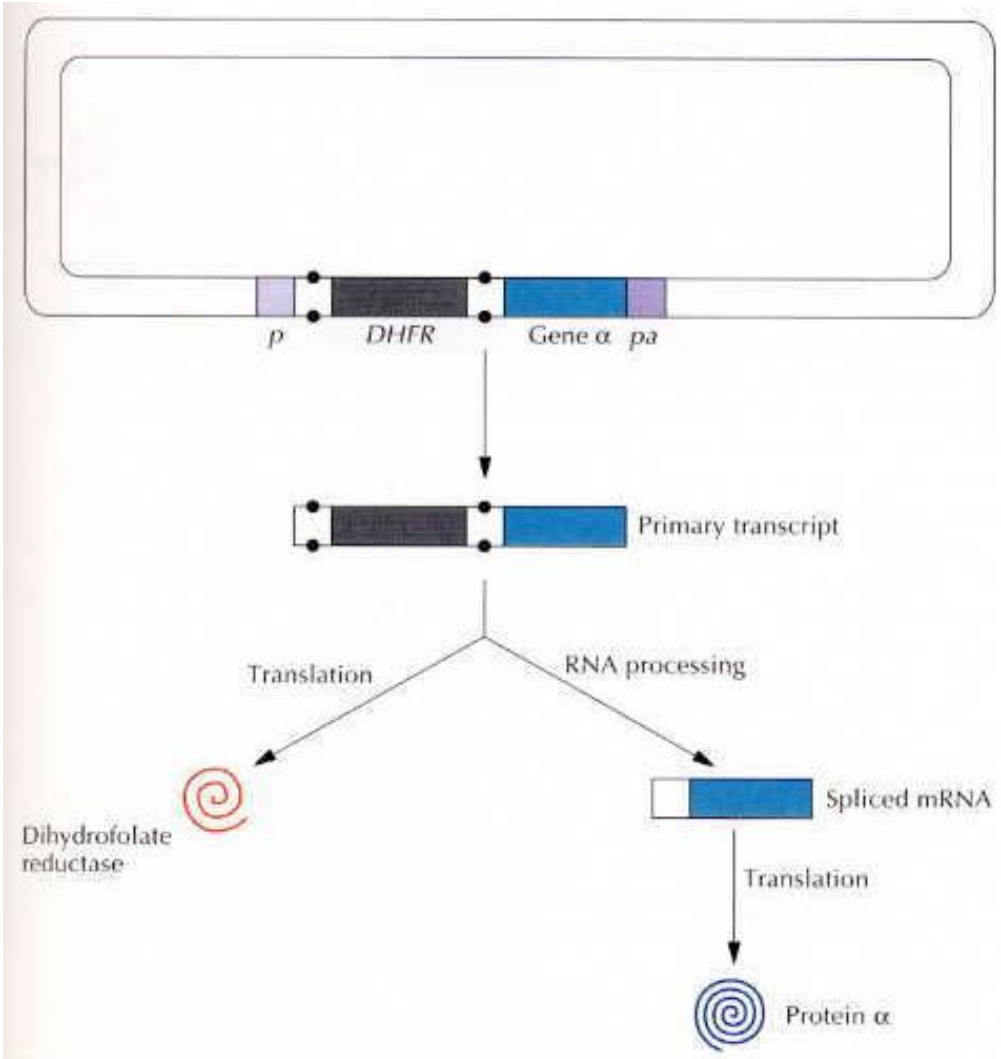
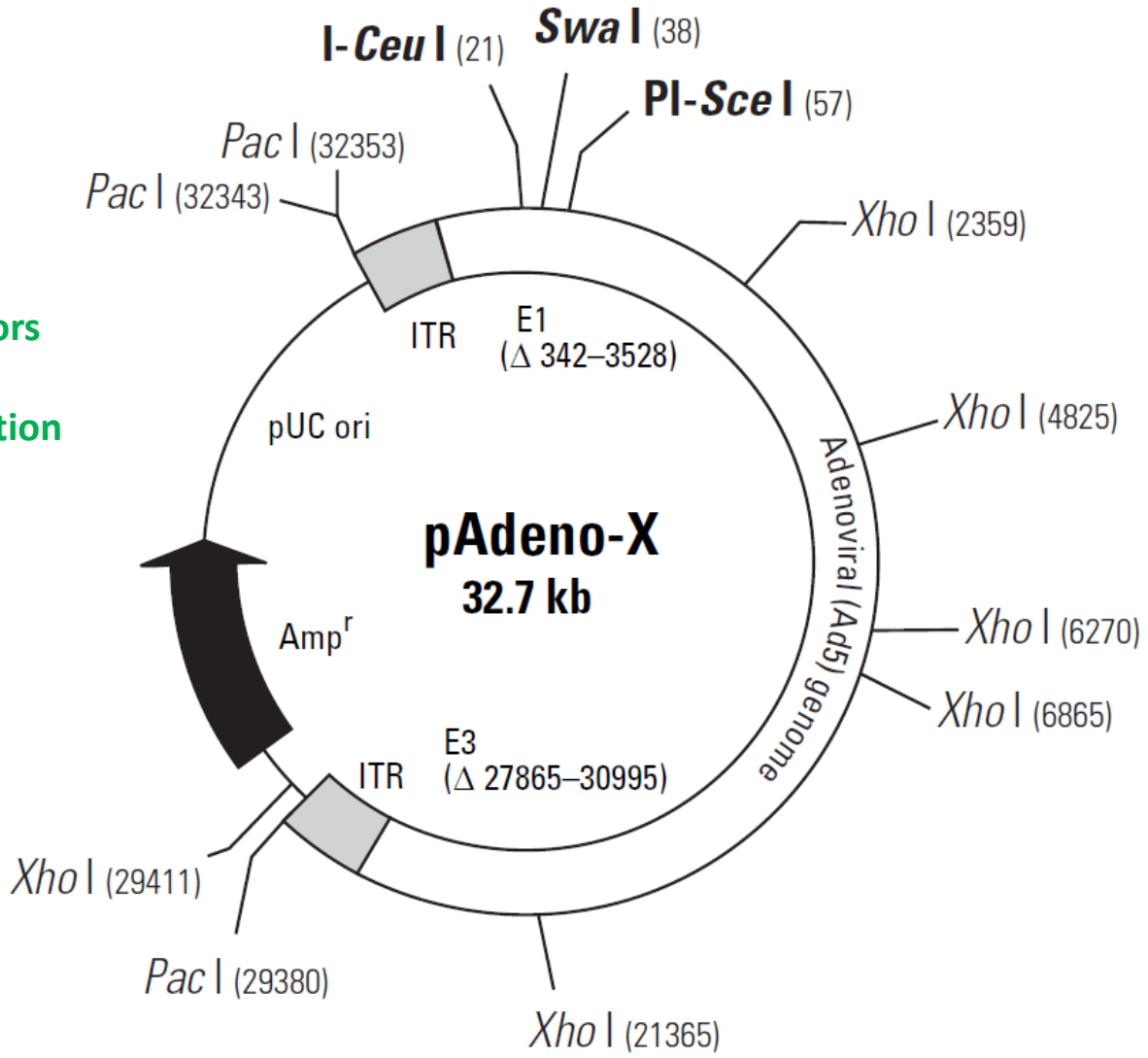


Figure 7.14 Coordinated expression of DHFR and a recombinant protein. A *DHFR* gene is cloned between intron donor and acceptor splice sites (dots) and upstream from a cloned gene (gene *α*). Both the *DHFR* and cloned genes are under the control of eukaryotic promoter (*p*) and polyadenylation (*pa*) sequences. Dihydrofolate reductase and the heterologous protein (protein *α*) are translated from the unspliced (primary) and processed (spliced) transcripts, respectively.

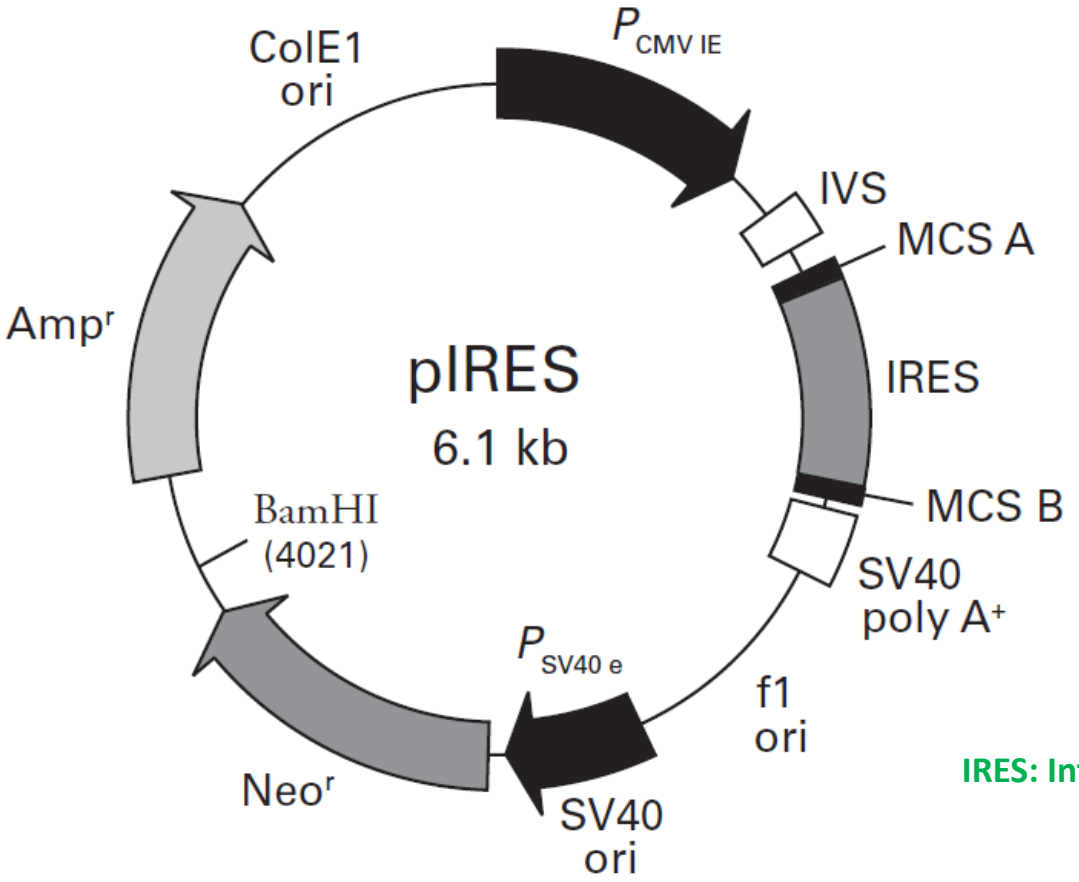
Mammalian Expression System

Adenovirus based Vectors

→ Autonomous replication in mammalian cells



Mammalian Expression System

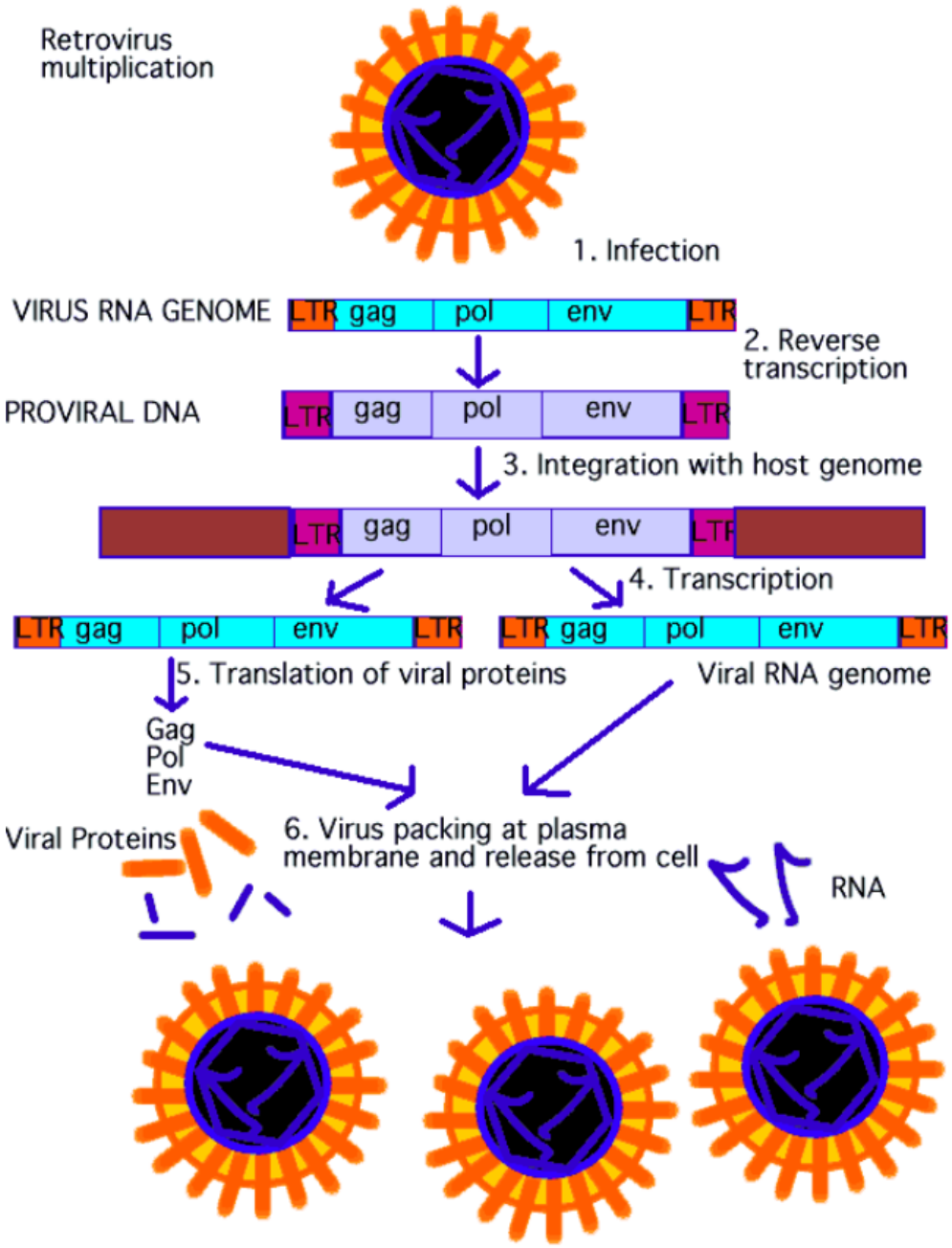


IRES: Internal Ribosomal Entry Site

pIRES is a mammalian expression vector that allows high level expression of two genes of interest from the same bicistronic mRNA transcript. The vector contains the encephalomyocarditis virus (ECMV) internal ribosome entry site (IRES) flanked by two multiple cloning sites (MCS A and B), an arrangement that allows cap-independent translation of the gene cloned into MCS B (1–3). pIRES utilizes a partially disabled IRES sequence that reduces the rate at which the gene cloned into MCS B is translated relative to that of MCS A.

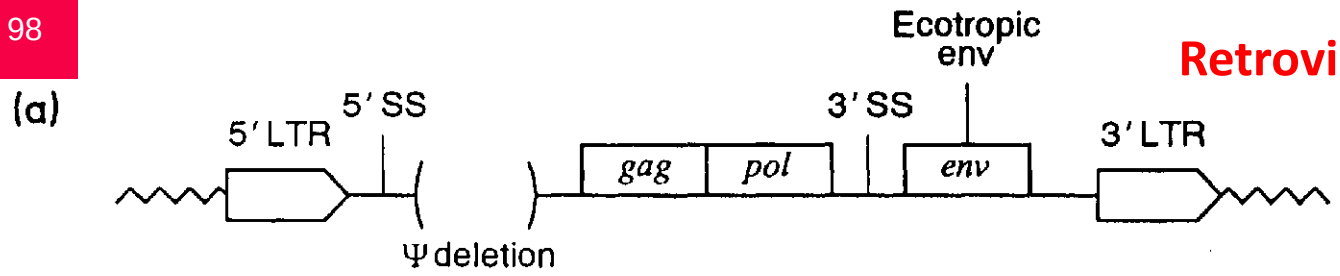
www.clontech.com

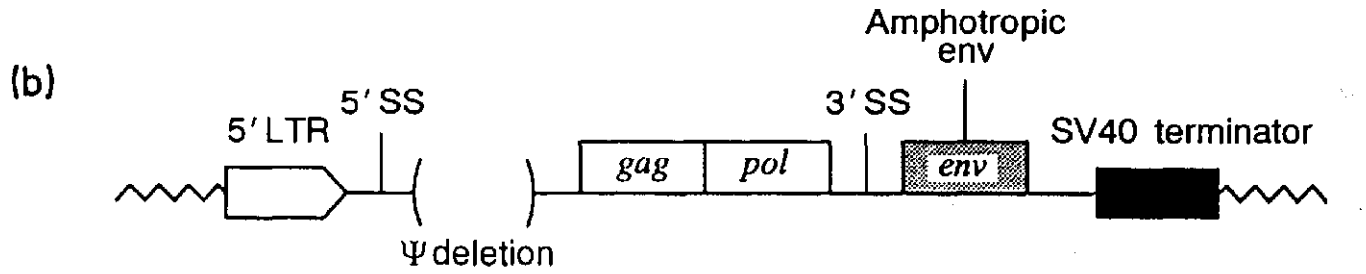
Retroviral vectors

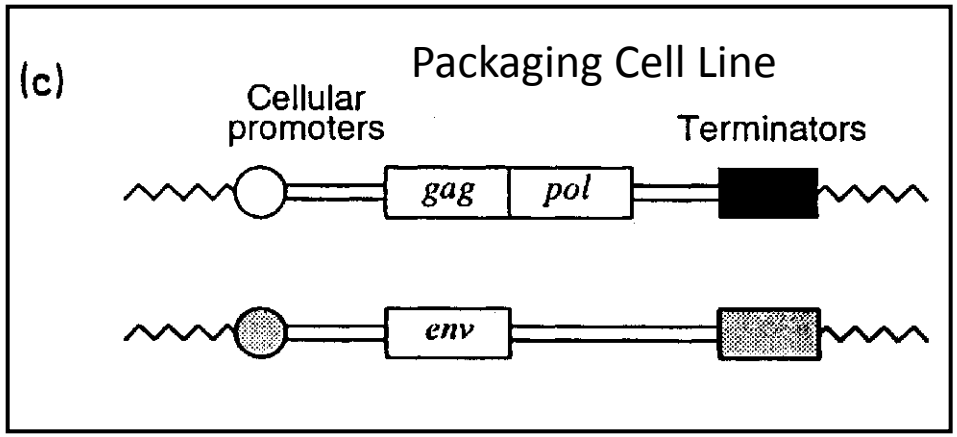


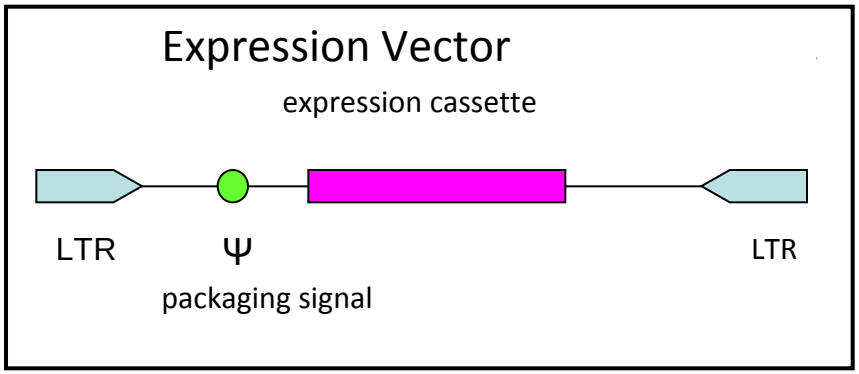
http://staff.jccc.net/pdecell/bio205/webct/bio205home.html

Retroviral expression system

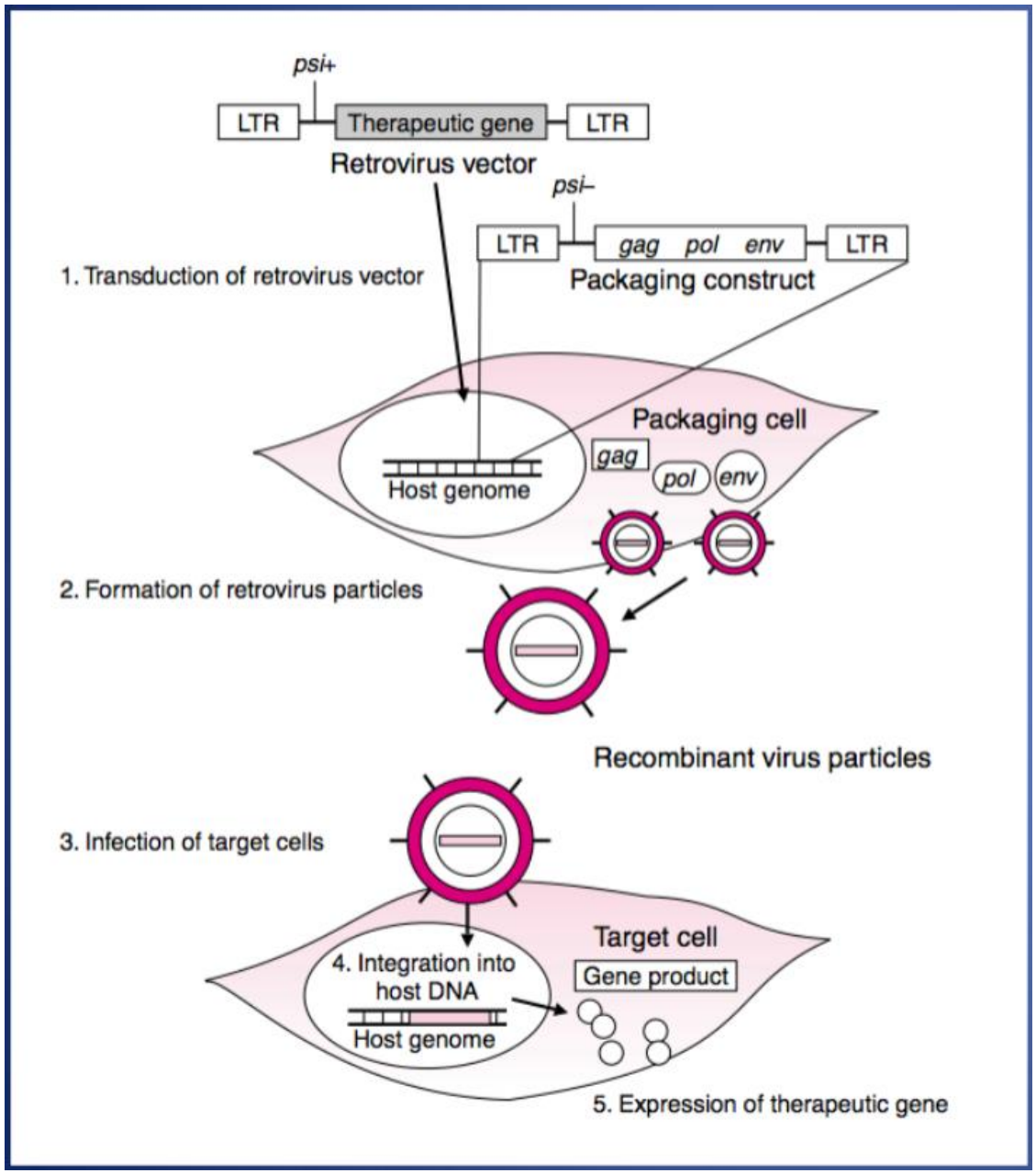








Recombinant retroviral genomes in packaging cell lines

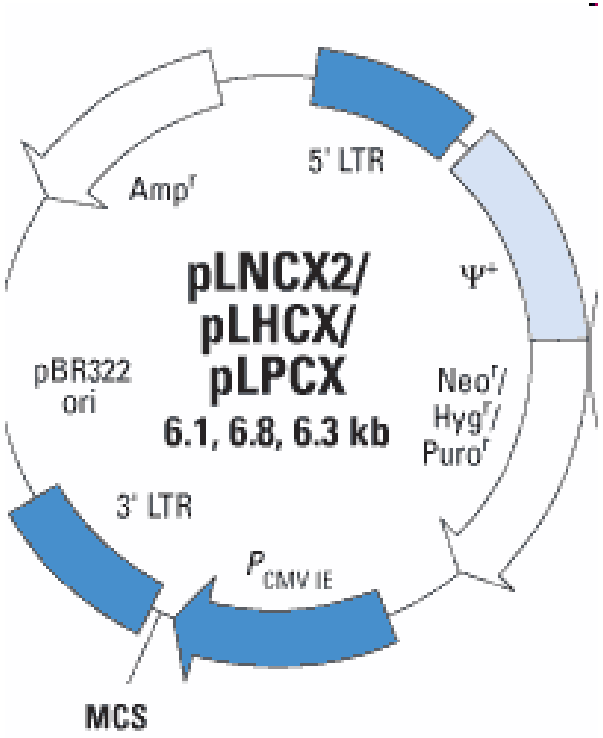


Recombinant Retroviruses

- Can be pseudotyped with various env proteins to broaden tropism
- Stable packaging cells
- Long-term gene expression through integration
- Downside is insertional mutagenesis
- Disadvantage is only infects dividing cells

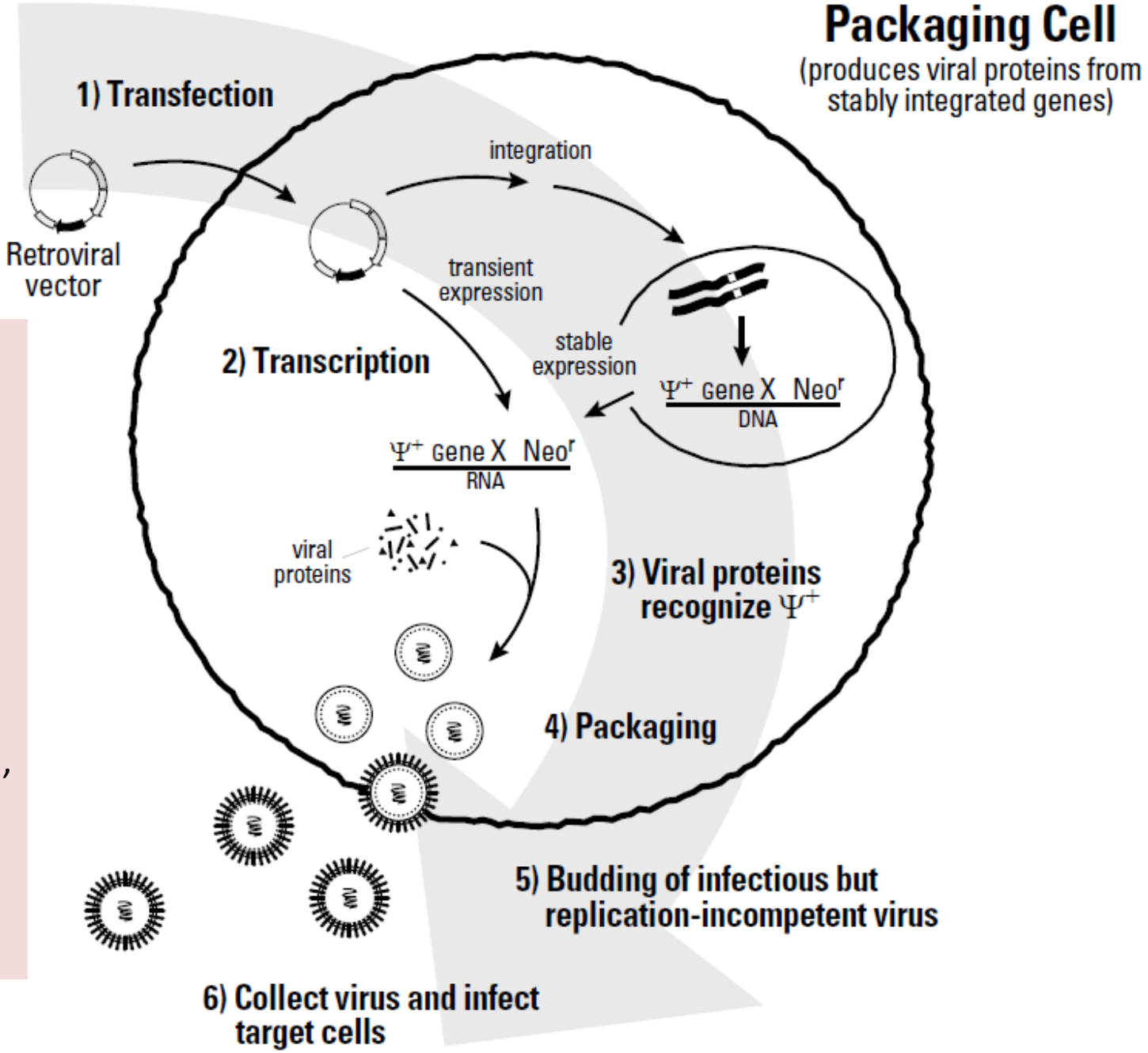
http://www.ohsu.edu/xd/about/services/integrity/upload/IBC_Presentation-Choosing-a-Viral-Vector-System.pdf

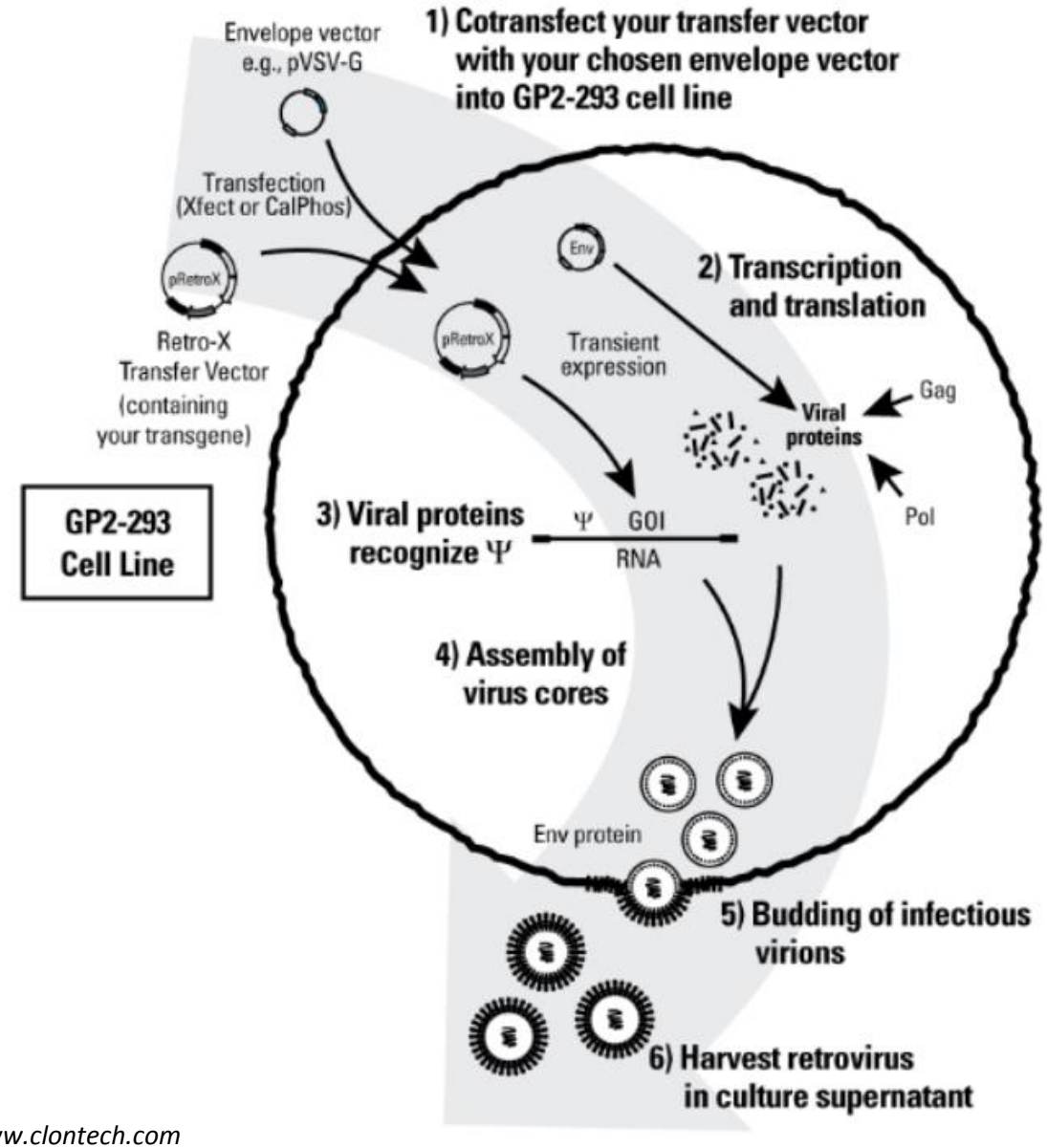
Basic retrovirus vector



Schematic of LRCX Vectors

Virus production in packaging cell lines. The *gag, pol* and *env* genes required for viral production are integrated into the packaging cells genome. The vector provides the viral packaging signal, commonly denoted Ψ+, a target gene, and drug-resistance marker.





Step 1: Cotransfect GP2-293 cells with the Retro-X vector containing your gene of interest (GOI) and an envelope plasmid such as pVSV-G

Step 2: Resulting production of the corresponding recombinant retroviral genome and viral packaging proteins. GP2-293 Cells express gag and pol from genomic locations

Step 3: Recognition of the packaging sequence (Ψ) on the recombinant viral RNA genome by the packaging proteins.

Step 4: Resulting assembly of viral cores, which are transported to the cell membrane.

Step 5: Cores are then enveloped by cellular membrane containing aggregated VSV-G or other envelope proteins. Mature, infectious virions then bud from the cell.

Step 6: Infectious virions are collected in the medium.

NOTE: Although the virions are infectious, they lack several critical genes required for their subsequent replication and production in target cells. Separating the viral proteins and supplying them *in trans* adds a strong measure of safety to virus production, since several low-frequency recombination events would need to occur in order to regenerate a replication-competent viral genome.

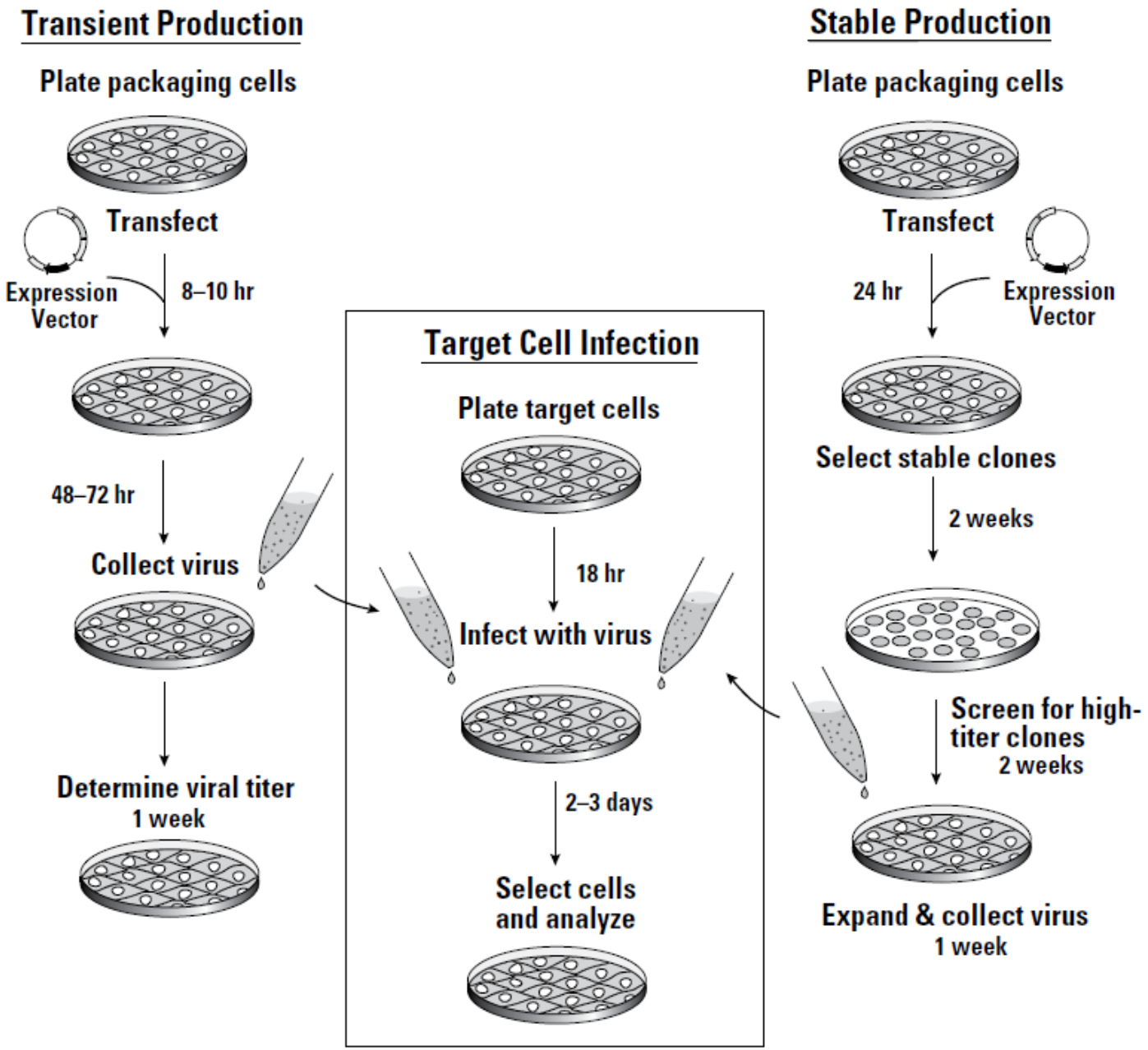
www.clontech.com

Table I: Summary of Retroviral Packaging Cell Lines Available from Clontech

Cell Line	GP2-293	Ampho-Pack-293	EcoPack 2-293	RetroPack PT67
Cat. Nos.	631530, 631512	631505	631507	631510
Cell line origin	HEK 293	HEK 293	HEK 293	NIH/3T3
Tropism	Pantropic, ecotropic, amphotropic, dualtropic	Amphotropic	Ecotropic	Dualtropic
Envelopes	VSV-G, ampho, eco, or 10A1	4170A (ampho)	gap70 (eco)	10A1
Target cells	Wide range of mammalian/ non-mammalian cells	Wide range of mammalian cells	Mouse, rat cells	Wide range of mammalian cells

Table 2: Tropisms Associated with Commonly Used Retroviral Envelopes

Envelope		VSV-G	4070A (Ampho)	gap70 (Eco)	10A1 (Dual)
Tropism		Pantropic	Amphotropic	Ecotropic	Dualtropic
Receptor (target cell)		Unknown[b]	RAM1 (Pit2)	mCAT-1	GALV (Pit1), RAM1 (Pit2)
Possible target cell types[a]	Human	+	+	–	+
	Mouse	+	+	+	+
	Rat	+	+	+	+
	Hamster	+	+/–	–	+
	Cat	+	+	–	+
	Dog	+	+	–	+
	Monkey	+	+	–	+
	Avian	+	–	–	–
	Fish	+	–	–	–
	Insect	+	–	–	–


Transient Production
Plate packaging cells
Transfect
Expression Vector
8–10 hr
48–72 hr
Collect virus
Determine viral titer
1 week
Target Cell Infection
Plate target cells
18 hr
Infect with virus
2–3 days
Select cells and analyze
Stable Production
Plate packaging cells
Transfect
24 hr
Expression Vector
Select stable clones
2 weeks
Screen for high-titer clones
2 weeks
Expand & collect virus
1 week

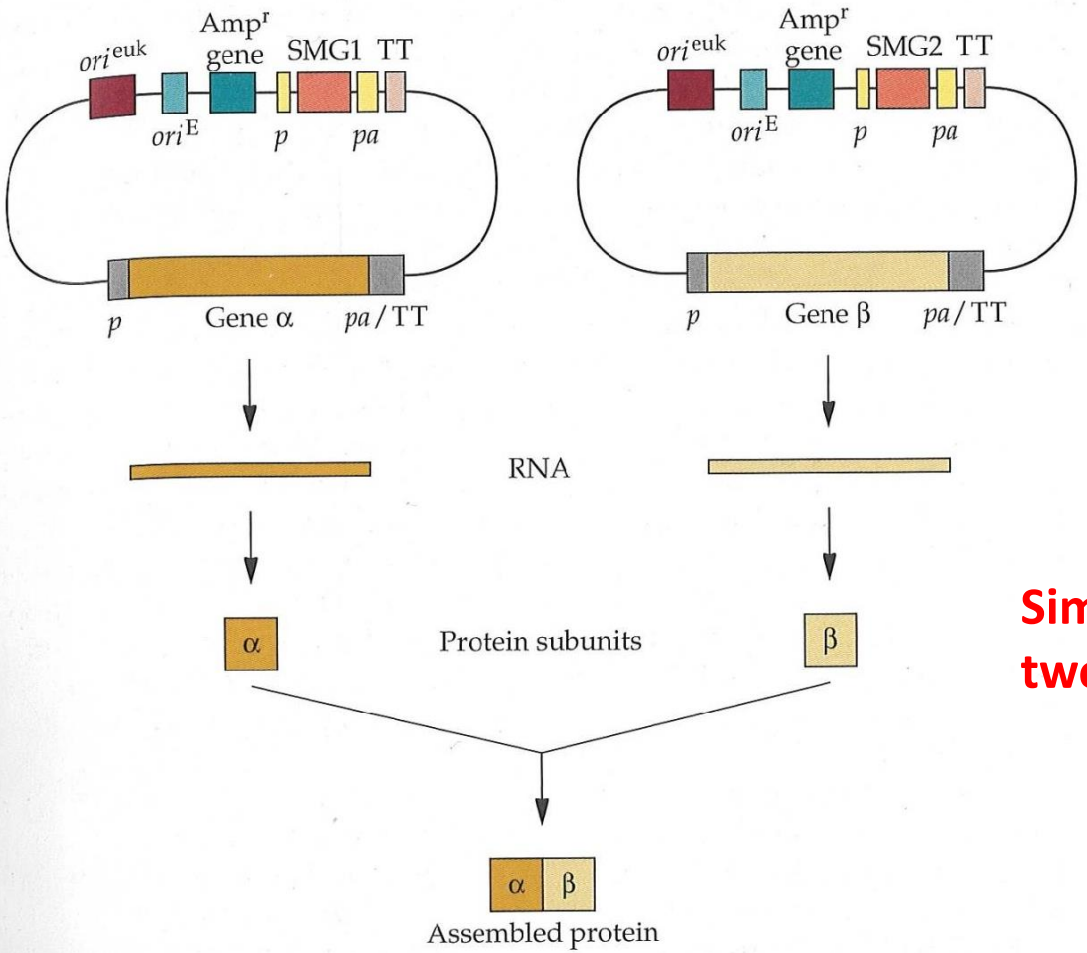


FIGURE 7.25 Two-vector expression system. The cloned genes (gene α and gene β) encode subunits of a protein dimer (αβ). After cotransfection, both subunits (α and β) are synthesized and assembled into a functional protein dimer. Both vectors carry origins of replication for *E. coli* (ori^E) and mammalian cells (ori^{euk}) and a marker (Amp^r) gene for selecting transformed *E. coli*. The selectable marker genes (SMG1 and SMG2) and the cloned genes (gene α and gene β) are each under the control of promoter *(p)*, polyadenylation *(pa)*, and termination of transcription (TT) sequences.

Simultaneous expression of two polypeptides

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

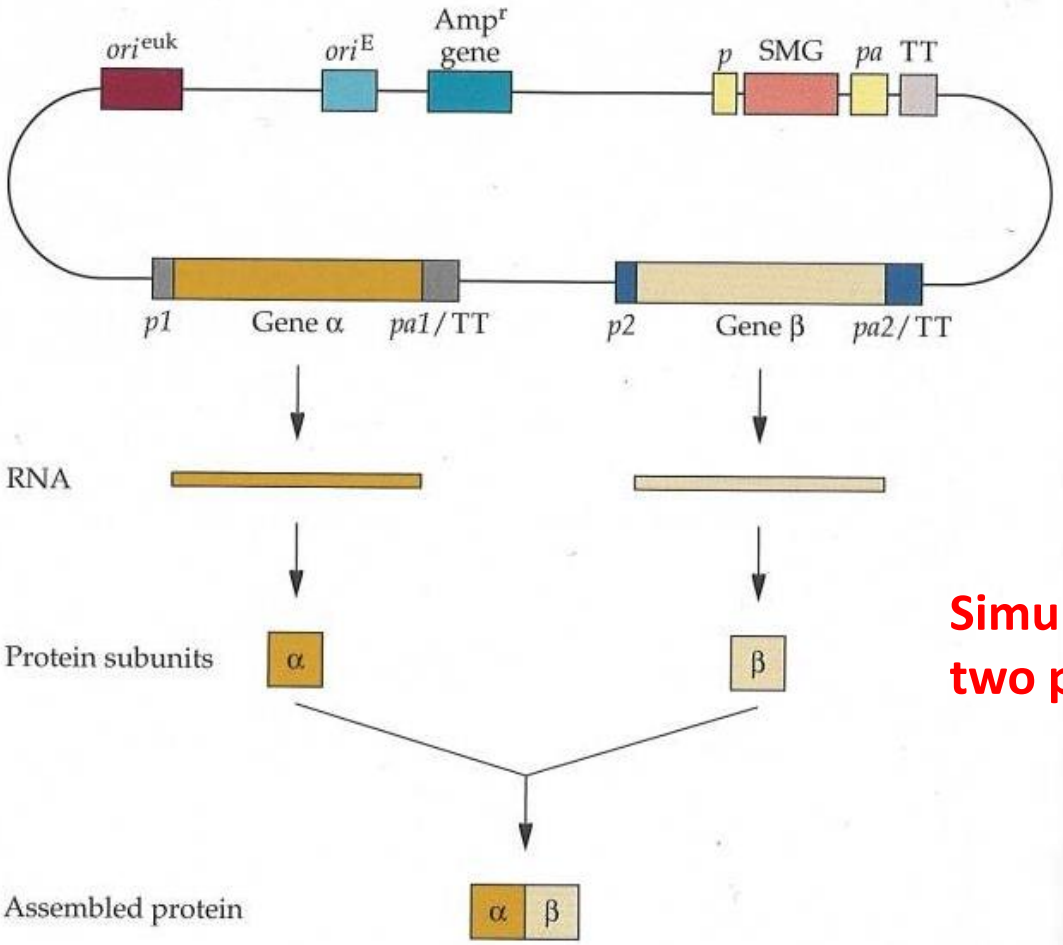


FIGURE 7.26 Two-gene expression vector. The cloned genes (gene α and gene β) encode subunits of a protein dimer (αβ). The cloned genes are inserted into a vector and are under the control of different eukaryotic promoter *(p)*, polyadenylation *(pa)*, and termination of transcription (TT) sequences. Each subunit is translated from a separate mRNA, and a functional protein dimer (αβ) is assembled. The vector has origins of replication for *E. coli* (ori^E) and mammalian cells (ori^{euk}), a marker gene (Amp^r) for selecting transformed *E. coli*, and a selectable marker gene (SMG) that is under the control of eukaryotic promoter *(p)*, polyadenylation *(pa)*, and termination of transcription (TT) sequences.

Simultaneous expression of two polypeptides

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

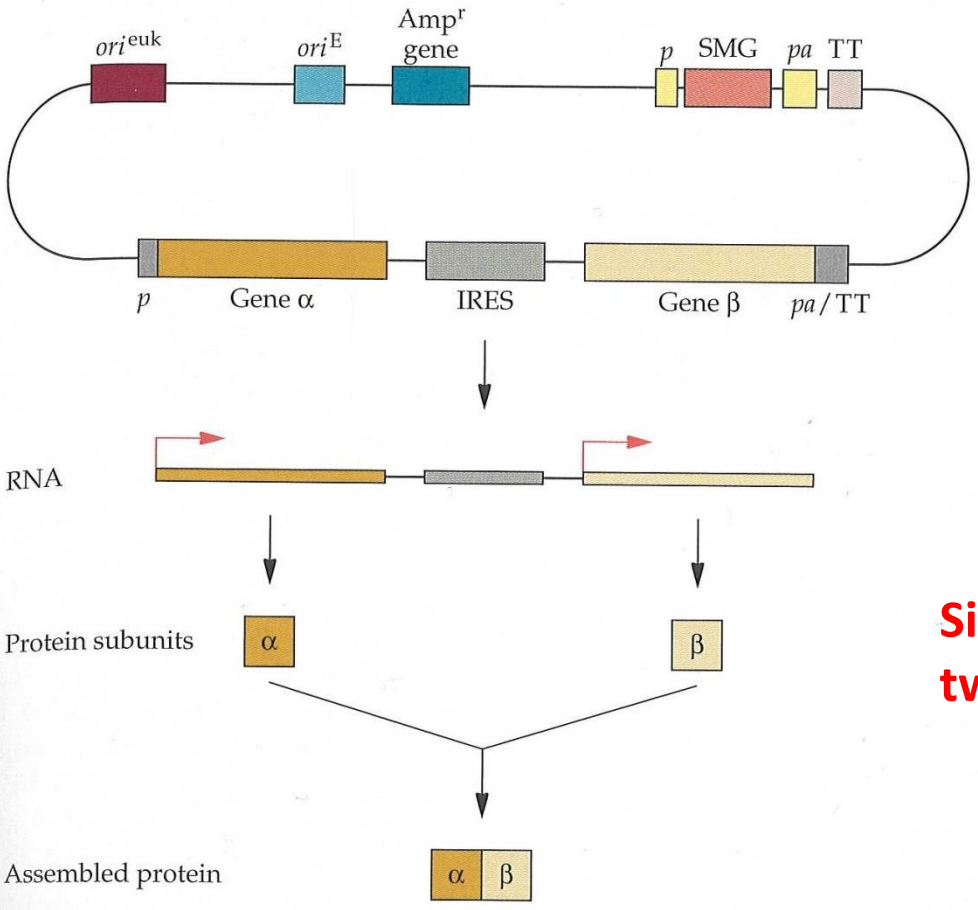


FIGURE 7.27 Bicistronic expression vector. The cloned genes (gene α and gene β) encode subunits of a protein dimer (αβ). Each cloned gene is inserted into a vector on either side of a sequence for an IRES. The two genes and the IRES sequence form a transcription unit under the control of a single eukaryotic promoter *(p)*, polyadenylation *(pa)*, and termination of transcription (TT) sequence. Translation of the mRNA occurs from the 5′ end and internally (right-angled arrows). Both subunits (α and β) are synthesized and assembled into a functional protein dimer (αβ). The vector carries origins of replication for *E. coli* (ori^E) and mammalian cells (ori^{euk}), a selectable marker (Amp^r) for selecting transformed *E. coli*, and a selectable marker gene (SMG) that is under the control of eukaryotic promoter *(p)*, polyadenylation *(pa)*, and termination of transcription (TT) sequences.

IRES:

Internal Ribosome Entry Site

Simultaneous expression of two polypeptides

Taken from: B.R. Glick, J.J. Pasternak, C.L. Patten; Molecular Biotechnology, 4th Ed.; ASM Press

2.10.16