DNA Vectors

Ms. Bhakti Hirani
MSc – Part I
Dept. of Biotechnology
Asst. Professor, Pillai College

Vectors for cloning large fragments of DNA

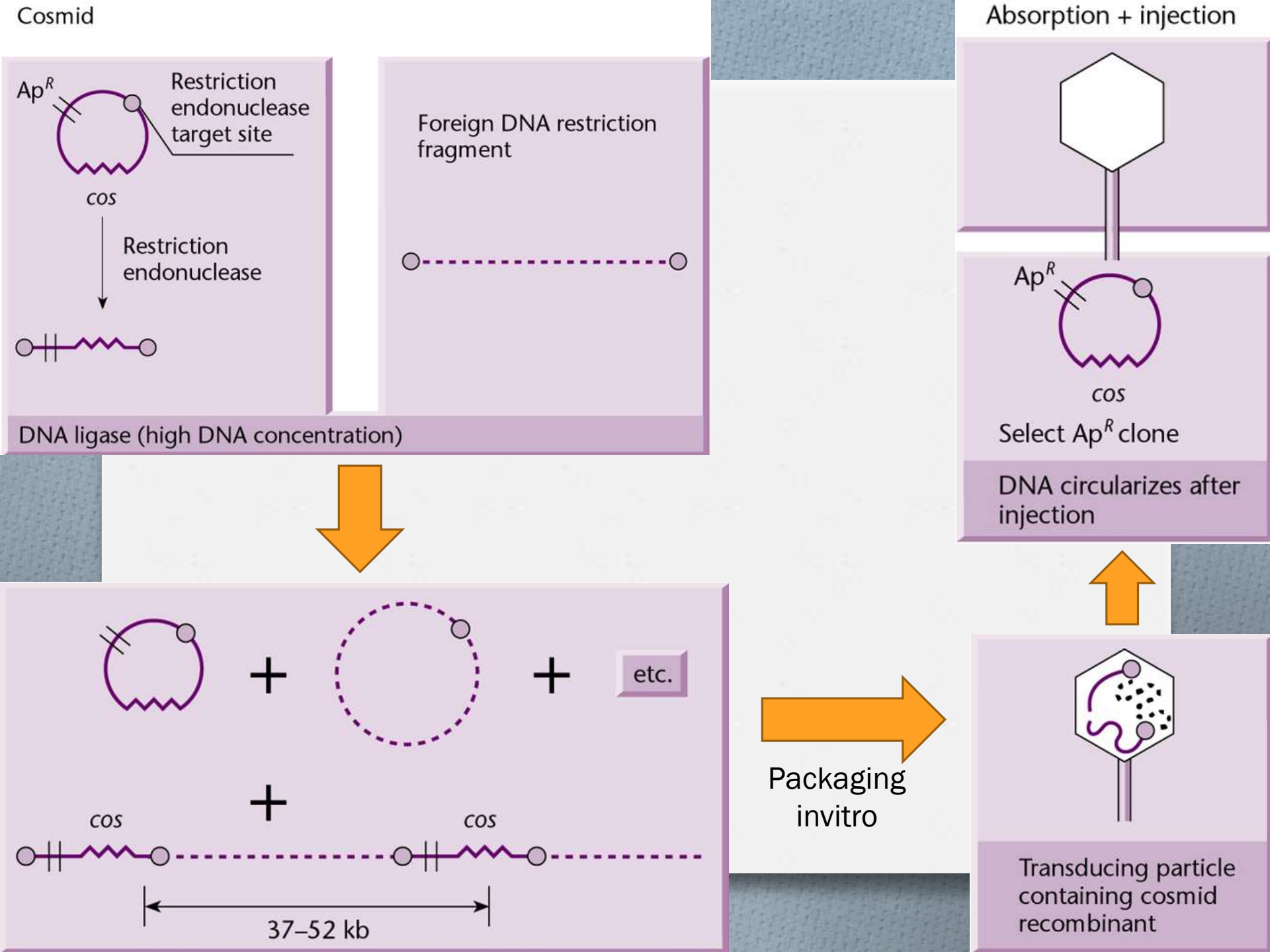
- Concatemers of unit length λ DNA can be packaged if *cos* sites, packaging dependent cleavage are 37-52 kb apart.
- A small region in the proximity of the *cos* sites is required for recognition by the packaging system.
- Plasmids have been constructed – containing fragment of λ DNA including the *cos* sites.
- These plasmids – Cosmids.
- Used as gene cloning vectors in conjunction with *invitro* packaging system.

- Cosmid vector – 5 Kb + 32 to 47 Kb of foreign DNA can be packed into the phage heads.

- After packaging *invitro*, phage is used to infect suitable host.

- Recombinant cosmid DNA is injected & circularizes like phage DNA but replicates as a normal plasmid without expression of any phage functions.

- Transformed cells are selected because of vector drug resistance marker.


Cosmid
Ap^R
Restriction endonuclease target site
cos
Restriction endonuclease
Foreign DNA restriction fragment
DNA ligase (high DNA concentration)
etc.
cos
cos
37–52 kb
Packaging invitro
Absorption + injection
Ap^R
cos
Select Ap^R clone
DNA circularizes after injection
Transducing particle containing cosmid recombinant

- Cosmids – clone large pieces of foreign DNA.
- Used to construct libraries – eukaryotic genome fragments.
- Partial digestion with restriction endonuclease – large fragments.
- Problem!!!!!
- 2 or more fragments join together in ligation reaction – creating clone containing fragments that were not initially adjacent in the genome.
- Incorrect picture of their chromosomal organization.

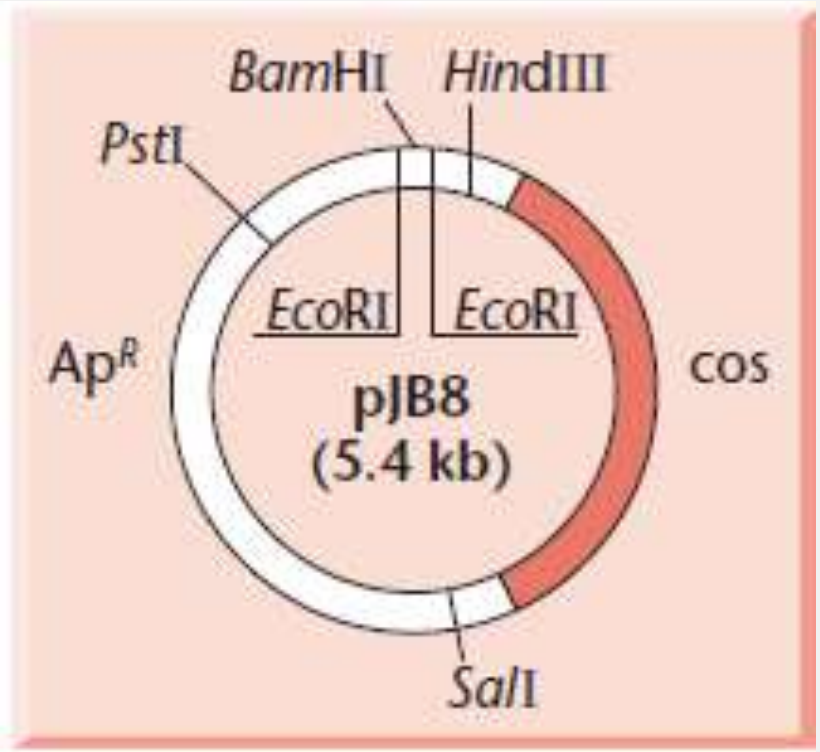
- Problem can be overcome by size fractionation of the partial digest.
- Buttttttt……. Cosmid clones again may be produced that contain non-contiguous DNA fragments ligated to form a single insert.
- This problem can be solved by dephosphorylating foreign DNA fragments – prevent their ligation together.
- This method is sensitive to exact ratio of target to vector DNAs – coz vector to vector ligation can occur.
- Recombinants with duplicated vectors are unstable & break down in host by recombination – resulting in propagation of non-recombinant cosmid vector.

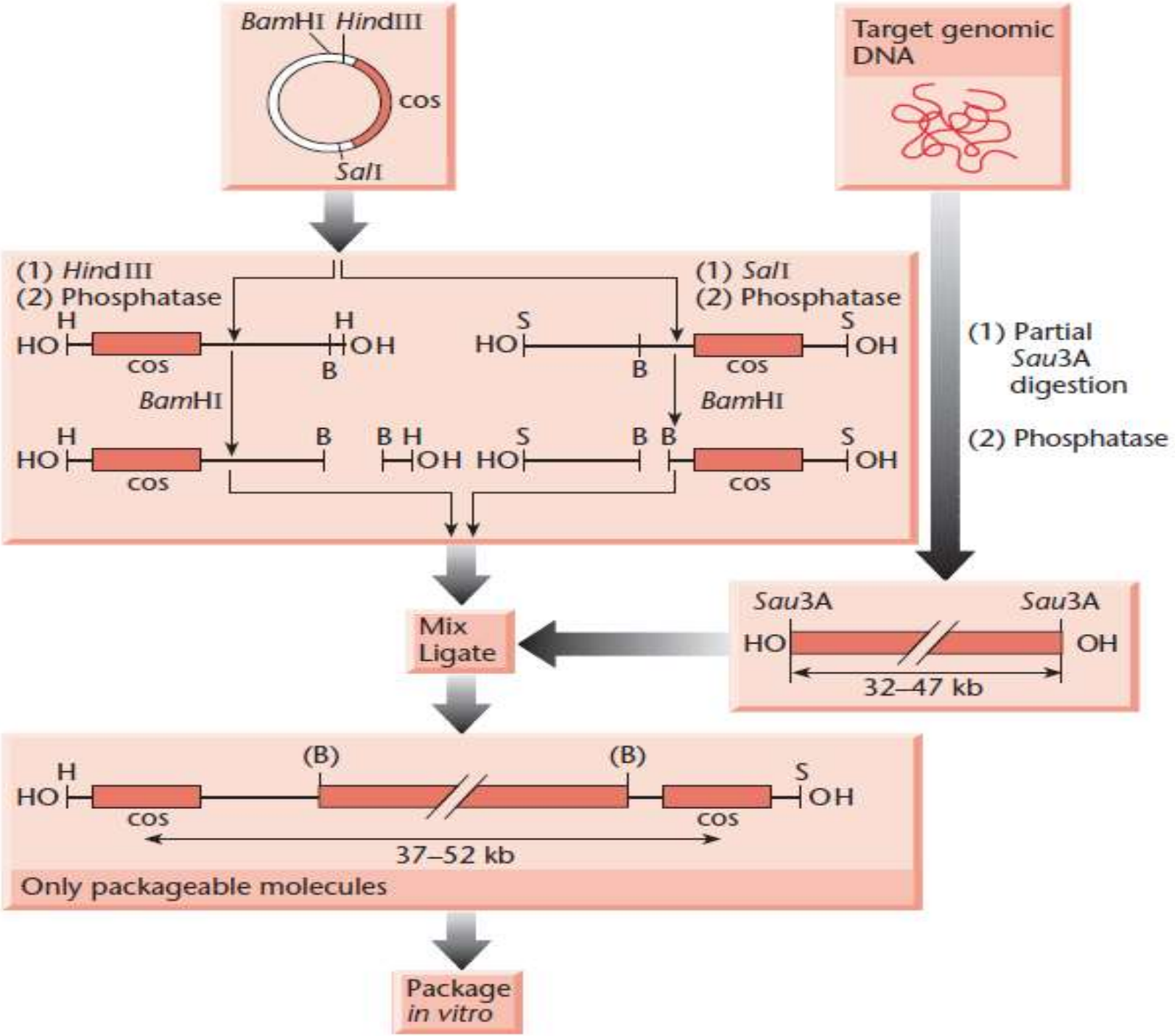
1. Cosmid pJB8

- All such difficulties have been overcome by a novel procedure devised by Ish-Horowicz & Burke (1981).

- Appropriate treatment of pJB8,
- left hand & right hand vector ends are purified – incapable of self ligation & which will accept dephosphorylated foreign DNA.

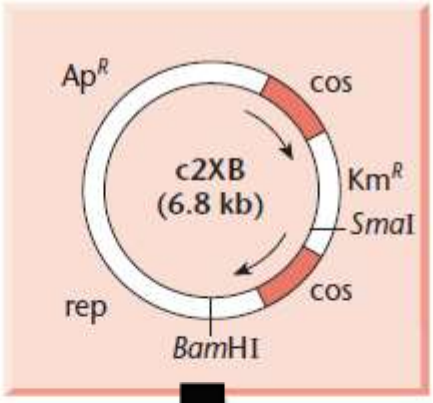
- Thus, no need to size fractionate foreign DNA &
- No Multiple vectors.


BamHI
HindIII
PstI
EcoRI
EcoRI
Ap^R
cos
pJB8
(5.4 kb)
SalI

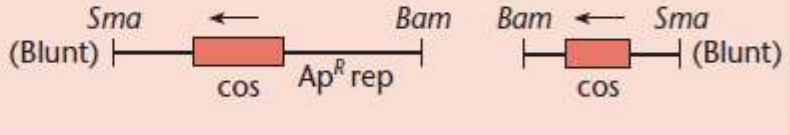

BamHI HindIII
cos
SalI
Target genomic DNA
(1) HindIII
(2) Phosphatase
(1) SalI
(2) Phosphatase
(1) Partial Sau3A digestion
(2) Phosphatase
BamHI
BamHI
Sau3A
Sau3A
Mix Ligate
32–47 kb
37–52 kb
Only packageable molecules
Package in vitro

2. Cosmid C2XB

- Devised by Bates & Swift (1983).
- C2XB carries:
- BamHI insertion site and
- 2 cos sites separated by blunt-end restriction site (Sma I).
- Blunt end ligates inefficiently,
- Thus preventing vector self-ligation in the ligation reaction.

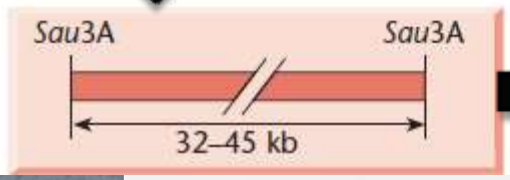


Target genomic DNA

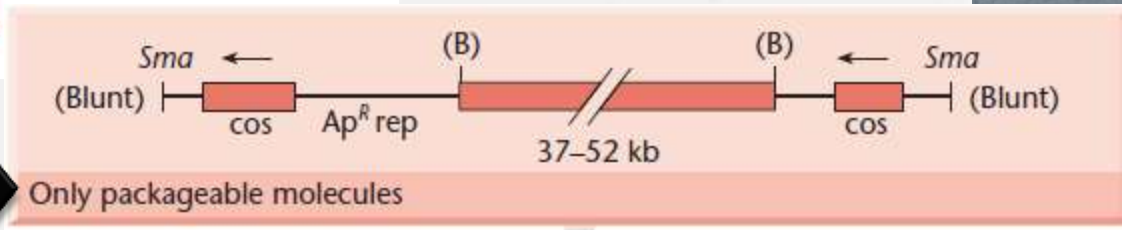


(1) Partial *Sau*3A digestion

(2) Phosphatase



Mix
Ligate



Only packageable molecules

Package
in vitro

3. pWE & sCos series cosmids

Features are:

- Multiple cloning sites – simple cloning using non-sized DNA.
- Phage promoters – flanking cloning site.
- Unique NotI, SacII, SfiI (rare cutters) flank the cloning site to permit removal of the insert from the vector as single fragment.
- Mammalian expression modules encoding dominant selectable markers present – gene transfer to mammalian cells.

4. Phage P1 vector system

- Developed by Sternberg & co-workers.
- Phage P1 is a temperate bacteriophage – mediate generalized transduction.
- They have capacity of 100 kb.
- Capacity – twice that of cosmid but less than YACs.

Phage P1 vector contains:

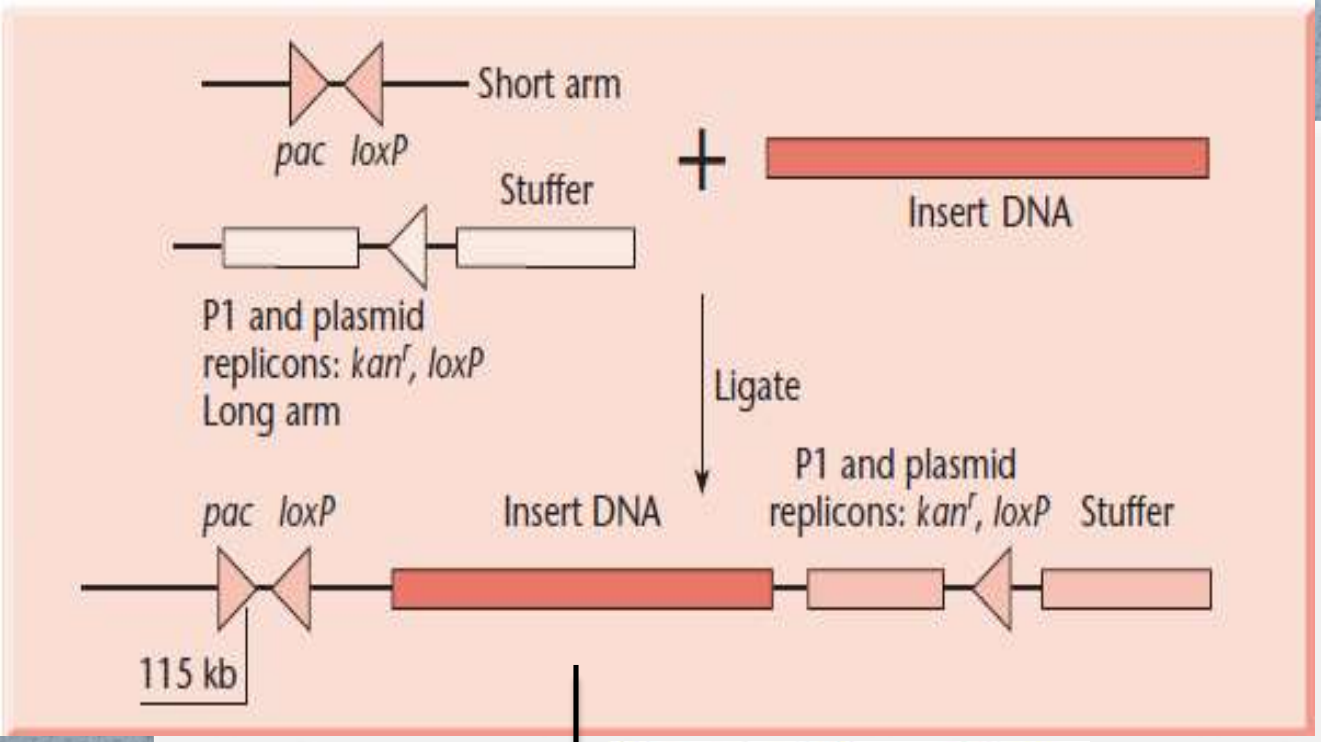
- *pac* site – necessary for invitro packaging of recombinant molecules into phage particles.
- *lox*P site – recognized by phage recombinase, product of phage *cre* gene – circularization of packaged DNA after it has injected into an *E. coli* host expressing recombinase.

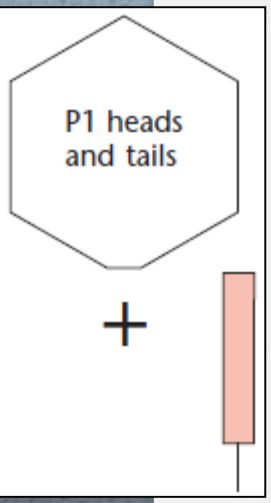
Contd..

- Clones are maintained at low copy number plasmids.
- Selection of the recombinant - Kanamycin resistance marker.
- High copy number can be induced by exploitation of the P1 lytic replicon.

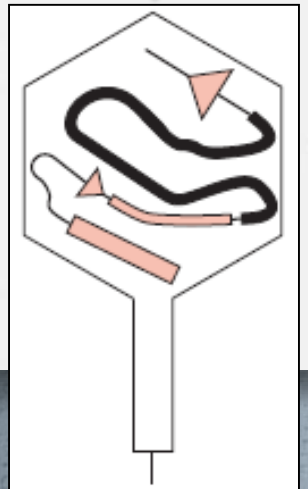
P1 system used construct genomic libraries of:

- Mouse
- Human
- Fission yeast
- Drosophila

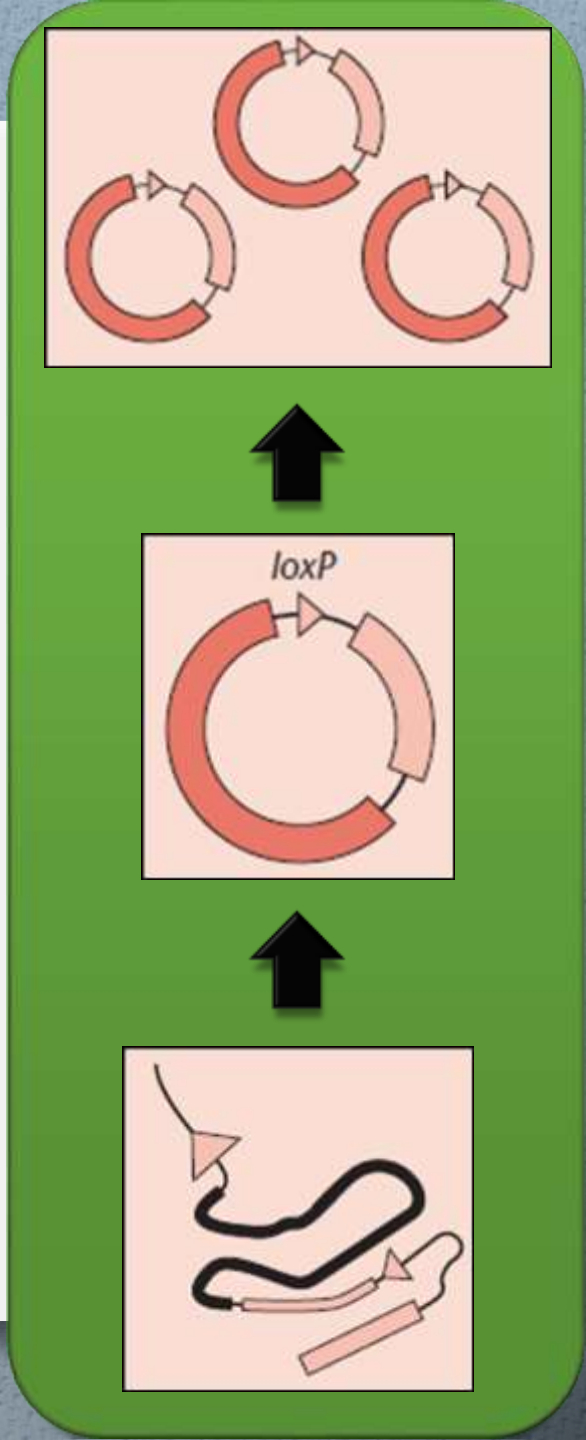




Pac extract cleaves at *pac* site and inserts DNA into P1 heads. Tails are attached

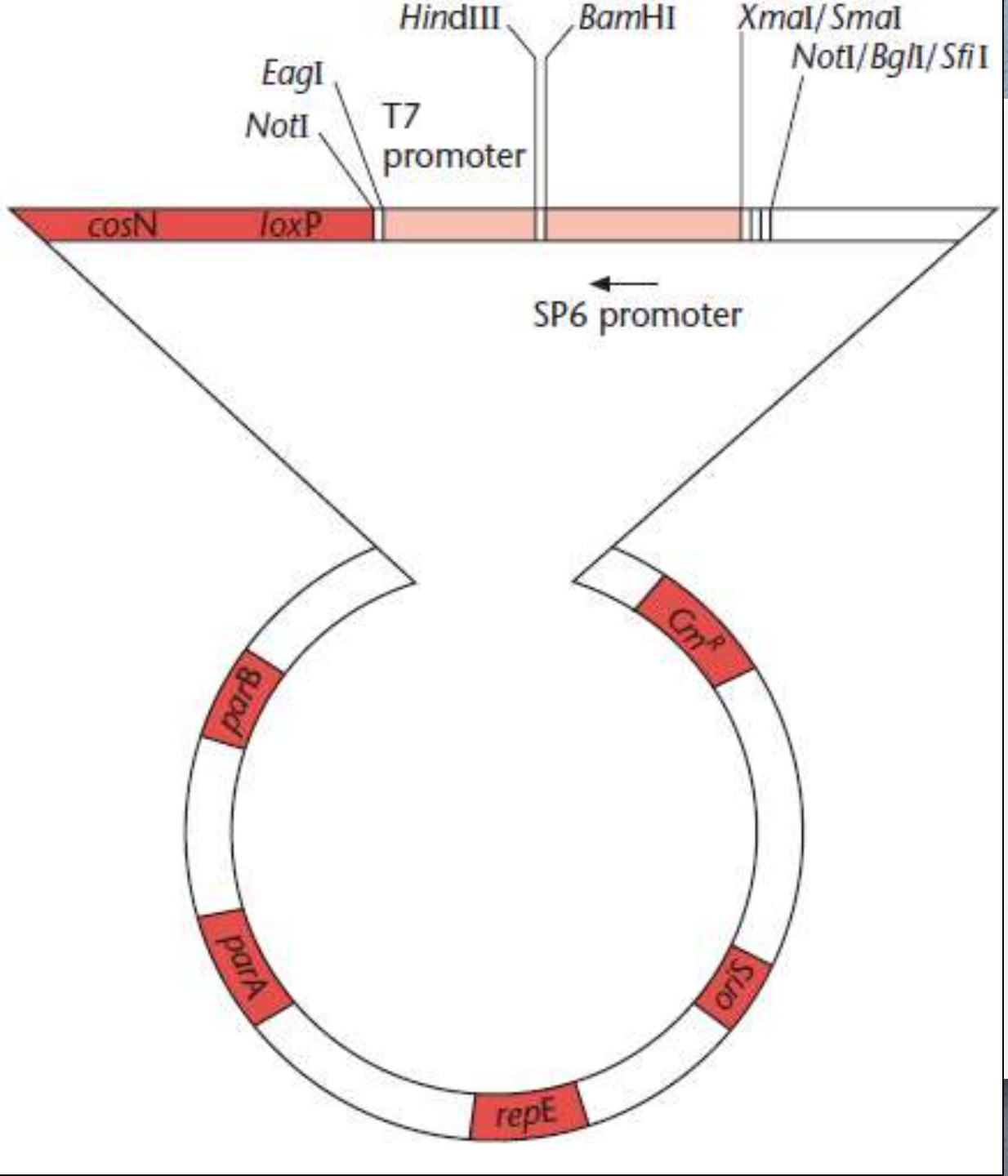


Allow phage to adsorb to host strain and inject DNA into *cre*$^+$ host cell



5. BAC vector

- Devised by Shizuya et al. (1992) – bacterial cloning system for mapping & analysis of complex genomes.
- Bacterial artificial chromosome (BAC) – based on single copy sex factor F of *E.coli*
- BACs – maintain human & plant genomic fragments greater than 300 kb for over 100 generations with high degre of stability.
- Its used to construct genome libraries with an average insert size of 125 kb.


HindIII
BamHI
XmaI/SmaI
NotI/BglI/SfiI
EagI
NotI
T7
promoter
cosN
loxP
SP6 promoter
parB
Cm^R
parA
oriS
repE

BAC vectors includes:

- λ cos N & P1 loxP sites.
- 2 cloning sites (HindIII & BamHI).
- G+C restriction enzyme sites (e.g. SfiI, NotI) for excision of inserts.
- Cloning site is flanked by T7 & SP6 promoters to generate RNA probes.

This BAC can be transformed very easily into *E.coli* very efficiently, thus avoiding packaging extracts that required in Phage P1 vector system.

Contd…

1. pBAC108L – lacked selectable marker for recombinants.

So clones inserts had to be identified by colony hybridization. Its inconvenient now.

2. pBeloBAC11 & pECBAC1 are derivatives of pBAC108L, in which original cloning site is replaced by lac Z gene carrying multiple cloning site.

3. pBeloBAC11 – 2 EcoRI sites

- LacZ gene
- CM^R gene

Contd…

4. pECBAC1 → LacZ gene
5. Improvement of BAC by replacing lacZ gene with sacB gene. Insertional inactivation of sacB gene permits growth of host cell on sucrose containing media – positive selection for inserts. (Hamilton et al. 1996)
6. Further improved BAC by including a site for insertion of a transposon. (Frengen et al. 1999), this enables genomic inserts to be modified after cloning in bacteria – known as retrofitting.
7. Use of retrofitting: simplified introduction of deletions & introduction of reporter genes for use in original host of the genomic DNA.

6. PAC vector

- Devised by Ioannou et al. (1994)
- P1 derived artificial chromosome (PAC) – combining features of P1 & F factor systems.
- PAC vectors are able to handle inserts from 100 – 300 kb range.

Choice of vector

Vector	Host	Insert size
λ phage	*E. coli*	5 – 25 kb
λ cosmid	*E. Coli*	35 – 45 kb
P1 phage	*E. Coli*	70 – 100 kb
PACs	*E. Coli*	100 – 300 kb
BACs	*E. Coli*	< 300 kb
YACs	*S. cerevisiae*	200 – 2000 kb

- 50% of the YACs show structural instability of inserts or are chimeras in which 2 or more DNA fragments have become incorporated into 1 clone.
- These defectives are unsuitable for mapping & sequencing.
- Cosmids too contains same aberrations.
- Other problems arises when DNA being cloned contains tandem arrays of repeated sequences.
- Problem is acute when tandem arrays are several times larger than allowable size of a cosmid insert.

- Advantages of BAC & PAC over YACs n cosmids include:
- Lower level of chimerism
- Ease of library generation
- Ease of manipulation
- Ease in isolation of insert DNA

BAC clones – represent human DNA far more faithfully then their YAC or cosmid counterparts & are excellent substrates for shotgun sequence analysis - accurate contiguous sequence data.

Specialist purpose vectors

Vectors used to make single stranded DNA for sequencing

- A new gene, novel genetic construct – sequencing of the whole or part of that molecule has to be done.
- For sequencing ss DNA is required.
- Originally ss DNA was obtained by cloning the sequence of interest in an M13 vector.
- Nowadays, the sequence of interest is cloned in to a pUC based phagemid vector – M13 ori region + pUC Col E1 origin of replication.
- This vector replicate inside the cell as double stranded molecules.

- ss DNA for sequencing – infecting cultures with a helper phage such as M13Ko7.
- Helper phage has ori of P15A & a Kanamycin resistance gene inserted into M13 ori region 7 mutation in gII gene.
- M13 can replicate on its own.
- In presence of a phagemid bearing a wild type ori, ss phagemid is packaged preferentially & secreted into the culture.
- DNA purified from phagemids can be used directly for sequencing.

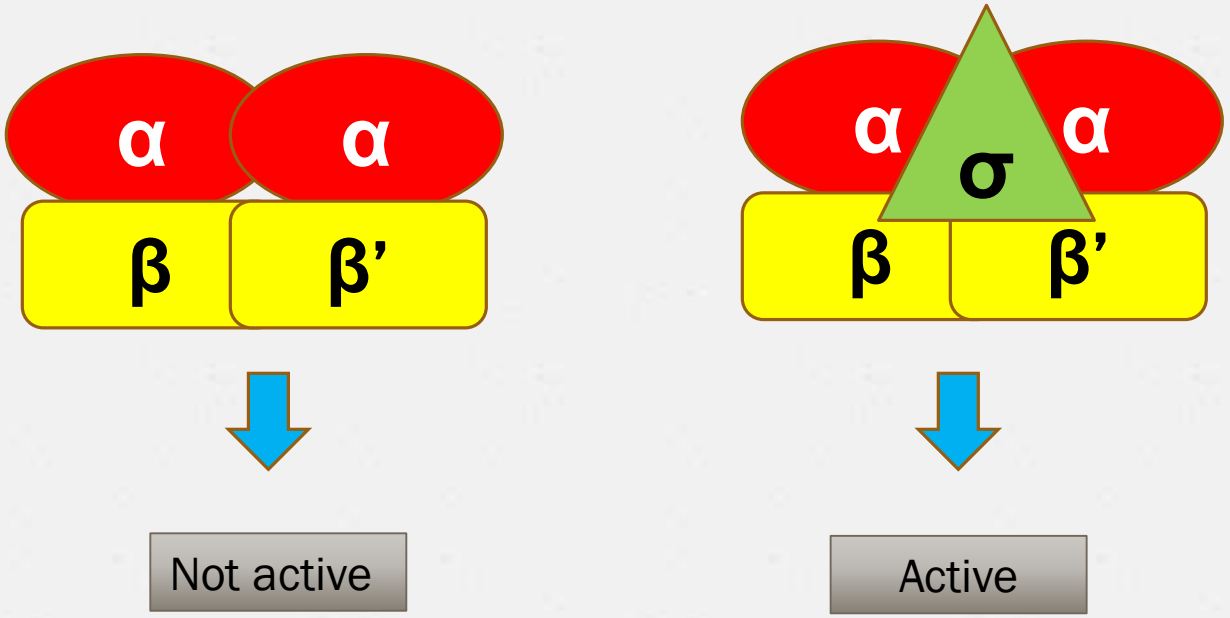
Expression vectors

- Purify gene products.
- Prepare RNA probes from cloned gene

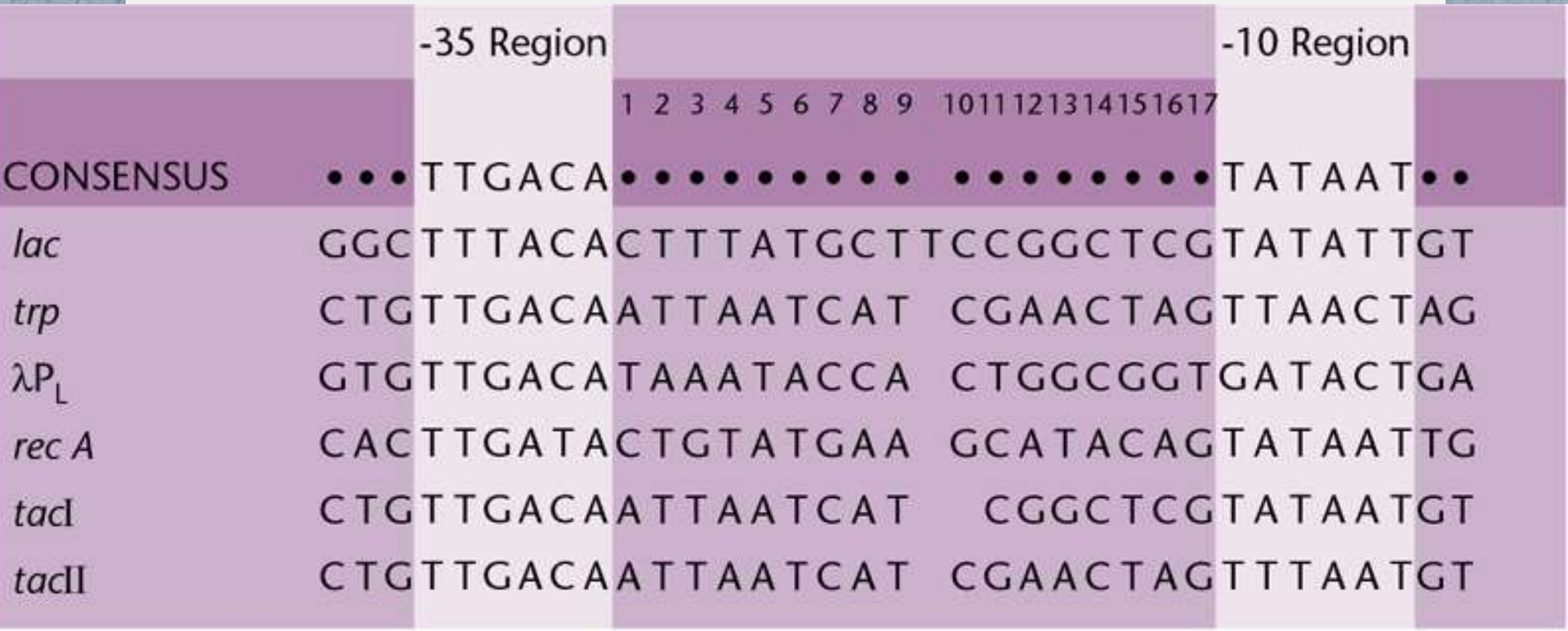
} Transcription of cloned gene is required

- Usually cloned gene is kept under its own promoter.
- Nowadays, promoter specific to the vector is utilized.
- These vector carried promoters are optimized for binding of the E. coli RNA polymerase.
- These can be regulated easily by changes in the growth conditions of the host cell.

- E. coli RNA polymerase is a multi-subunit enzyme



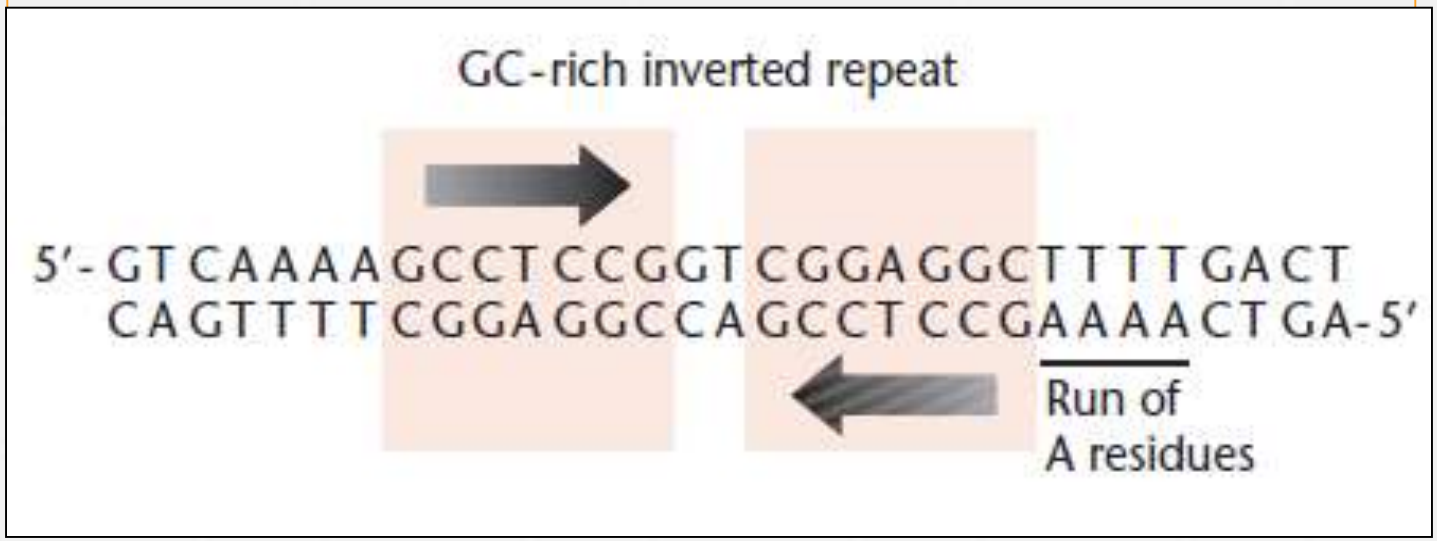
- RNA pol recognize different promoters &
- This recognition depends on the type of σ factors that is attached.
- Most common promoters are recognized by RNA pol with σ^{70}
- The comparison has led to the formulation of a consensus sequence.
- If the transcription start point is assigned position +1 & then consensus sequence consists of the -35 region (5’ TTGACA 3’) & -10 region or Pribnow box (5’ TATAAT 3’)
- RNA pol must bind to both sequences to initiate transcription.
- Strength of the promoter – how many copies of RNA are synthesized per unit time per enzyme molecule depends on how close its sequence is to the consensus.


-35 Region
-10 Region
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
CONSENSUS ••• TTGACA ••••••••• ••••••••• TATAAT ••
lac GGC TTTACA CTTTATGCTTCCGGCTCG TATATTGT
trp CTG TTGACA ATTAATCAT CGAACTAG TTAACTAG
λP_L GTG TTGACA TAAATACCA CTGGCGGT GATACTGA
rec A CAC TTGATA CTGTATGAA GCATACAG TATAATTG
tacI CTG TTGACA ATTAATCAT CGGCTCG TATAATGT
tacII CTG TTGACA ATTAATCAT CGAACTAG TTTAATGT

- Promoter strength is affected by mutations in -35 & -10 region &
- Also bases flanking these region affect promoter activity.
- Distance between -35 & -10 is imp – Promoter was weaker when spacing increased or decreased by 17 bp.

- Upstream elements (UP) located 5’ of the -35 hexamer are A+T rich that increase transcription by interacting with α subunit of RNA polymerase.
- Best UP sequence increased heterologous protein expression from the lac promoter by a factor of 100.

- Transcription once started, it will polymerize ribonucleotides until it encounters a transcription termination site in DNA.
- They are factor dependent & factor independent.
- Factor independent are easy to recognize as they have similar sequences: an inverted repeat followed by a string of A residues resulting in a string of U residues at the 3’ end of the mRNA.



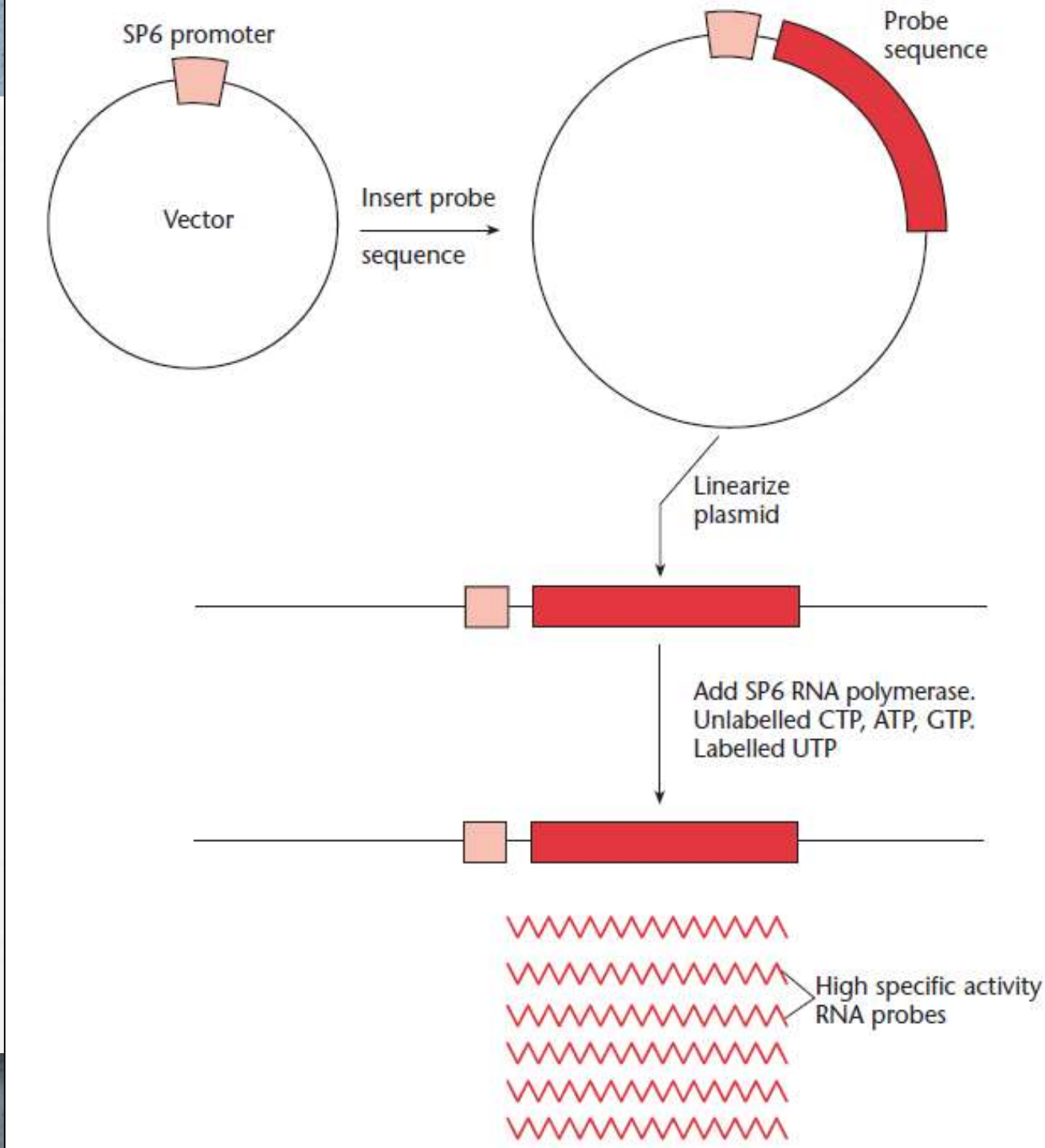
- Factor dependent termination involves interaction with one of the 3 known *E. coli* termination factors.
- Rho (ρ),
- Tau (τ) and
- NusA.

- Expression vectors incorporate factor independent termination sequence downstream from the site of insertion of the cloned gene.

Vectors for making RNA probes

- ss DNA can be use in hybridization experiments
- RNA probes are preferred.
- Coz the rate of hybridization & the stability are far greater for RNA-DNA hybrids compared to DNA-DNA hybrids.

- Relevant gene is cloned in a plasmid vector under the control of a phage promoter.
- After purification, plasmid is linearized with a suitable restriction enzyme &
- Then incubated with a phage RNA polymerase & 4 ribonucleoside triphosphate.
- No transcription termination is required because RNA polymerase will fall off the end of the linearized plasmid.


SP6 promoter
Vector
Insert probe sequence
Probe sequence
Linearize plasmid
Add SP6 RNA polymerase.
Unlabelled CTP, ATP, GTP.
Labelled UTP
High specific activity
RNA probes

- 3 reasons for using a phage promoter:

1. Such promoters are very strong, enabling large amounts of RNA to be made *invitro*.

2. Phage promoter is not recognized by the *E. coli* RNA polymerase & so no transcription will occur inside the cell.

3. RNA polymerases encoded by phages SP6, T7, T3 are much simpler to handle than *E. coli* enzyme, as they are single polypeptide.

- If it is planned to probe RNA or ss DNA sequences – then it is essential to prepare RNA probes corresponding to both strands of the insert.

1. Have 2 different clones – corresponding to 2 orientations of the insert.

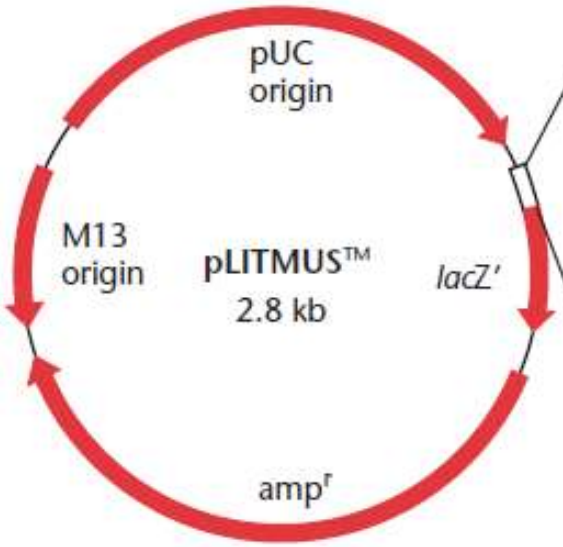
2. Use a cloning vector – insert is placed between 2 different opposing phage promoters (e.g. T7/T3 or T7/SP6) that flank MCS.

- since each of the 2 promoters are recognized by different RNA polymerase, the direction of transcription is determined by the type of polymerase used.

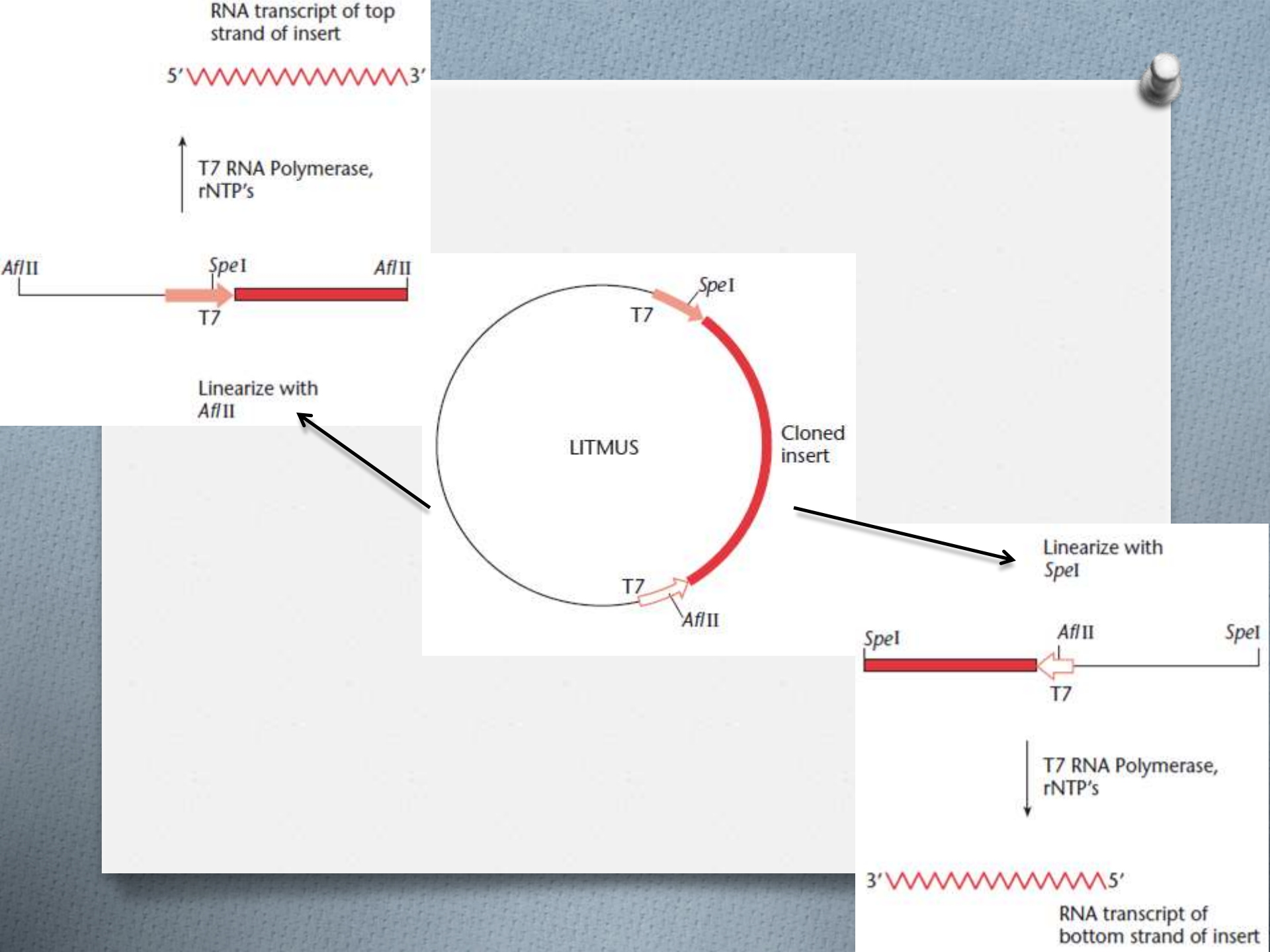
- LITMUS vectors: Polylinker region is flanked by 2 modified T7 RNA polymerase promoters.

- Each promoter contains (SpeI or AflII) that has been engineered into T7 promoter consensus sequence - cleavage with corresponding endonuclease inactivates that promoter.
- Selective unidirectional transcription is achieved by simply inactivating other promoter by digestion with SpeI or AflII prior to *invitro* transcription.
- Efficient labeling of RNA probes demands template be linearized prior to transcription at a site downstream from the insert.
- Cutting at the site within the undesired promoter performs both functions.

(a)



LITMUS 28	LITMUS 29	LITMUS 38	LITMUS 39
*Sna*BI	*Sna*BI	*Sna*BI	*Sna*BI
*Spe*I T7↓	*Spe*I T7↓	*Spe*I T7↓	*Spe*I T7↓
*Bgl*II	*Kpn*I-*Acc*65I	*Apa*I-*Bsp*120I	*Sal*I
*Nsi*I-*Ppu* 10I	*Sac*I	*Mfe*I	*Sph*I
*Bss*HII	*Avr*II	*Ngo*MI	*Bsr*GII
*Bsi*WI	*Xba*I	*Kas*I	*Bsp*EI
*Xho*I	*Age*I	*Hin*dIII	*Mlu*I
*Eco*RI	*Aat*II	*Pst*I	*Eag*I
*Pst*I	*Nco*I	*Eco*RV	*Nhe*I
*Eco*RV	*Hin*dIII	*Bam*HI	*Eco*RI
*Bam*HI	*Bam*HI	*Eco*RI	*Bam*HI
*Hin*dIII	*Eco*RV	*Nhe*I	*Eco*RV
*Nco*I	*Pst*I	*Eag*I	*Pst*I
*Aat*II	*Eco*RI	*Mlu*I	*Hin*dIII
*Age*I	*Xho*I	*Bsp*EII	*Kas*I
*Xba*I	*Bsi*WI	*Bsr*GI	*Ngo*MI
*Avr*II	*Bss*HII	*Sph*I	*Mfe*I
*Sac*I	*Nsi*I-*Ppu*10I	*Sal*I	*Apa*-*Bsp*120I
*Kpn*I-*Acc*65I	*Bgl*II	*Stu*I	*Stu*I
*Stu*I	*Stu*I		
*Afl*II T7↑	*Afl*II T7↑	*Afl*II T7↑	*Afl*II T7↑


RNA transcript of top strand of insert
5′
3′
T7 RNA Polymerase, rNTP's
AflII
SpeI
AflII
T7
Linearize with AflII
SpeI
T7
LITMUS
Cloned insert
T7
AflII
Linearize with SpeI
SpeI
AflII
SpeI
T7
T7 RNA Polymerase, rNTP's
3′
5′
RNA transcript of bottom strand of insert

Vectors for maximizing protein synthesis

- Cloned gene – preceded by a promoter recognized by host cell – detectable synthesis of cloned gene product.
- Protein being synthesized is used to : study its properties or has commercial value.
- Thus protein synthesis has to be maximized.

- Effects of overexpression on host cell should be considered.
- Toxic to host cell even in small amounts.
- E.g. PolA gene product – regulate basic cellular metabolism, cystic fibrosis transmembrane conductance regulator, lentivirus envelope sequences.

- Cloned genes allowed to be expressed

- Rapid selection of mutants that no longer synthesize the toxic protein.

- Even when overexpression of a protein is not toxic to host cell.

⬇

- High level synthesis – metabolic drain on the cell.

⬇

- Slower growth – selection of variants with lower or no expression of cloned gene as this will grow faster.

⬇

- To minimize problem associated with high level of expression

⬇

Use a vector

⬇

- Cloned gene is under the control of regulated promoter.

- Controllable promoters used are: λ P_L , t7, TRC (TAC) or bad.
- trc & tac promoters – hybrid promoters derived from lac & trp promoters.
- trc & tac are stronger promoters.
- Like lac, trc & tac promoters are inducible by lactose & isopropyl-β-D-thiogalactoside (IPTG).
- Vectors using this promoter carry lacO operator & lacI gene – encode repressor.

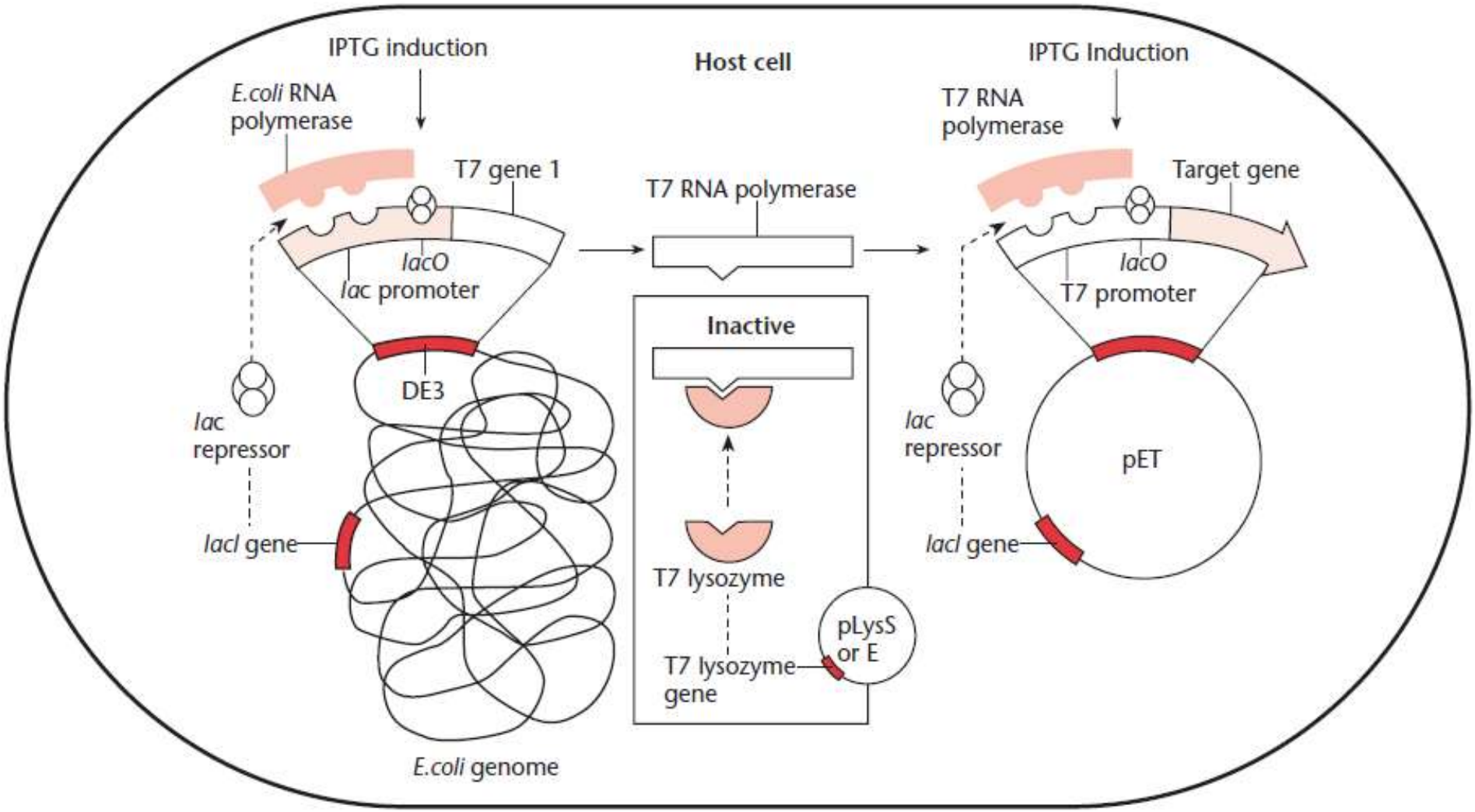
- pET vectors - expression vectors – Phage T7 promoters to regulate synthesis of cloned gene products.
- Source of phage T7 RNA polymerase – *E. coli* – gene 1 of the phage is constructed.
- Gene 1 is cloned downstream of lac promoter in chromosome – phage polymerase will be synthesized after IPTG induction.
- Newly synthesized T7 RNA polymerase - transcribe foreign gene in pET plasmid.

- If protein product is toxic – T7 RNA polymerase.

1. Plasmid compatible with T7 lysS gene is cloned & introduced in host carrying pET plasmid.

It will bind any residual T7 RNA polymerase.

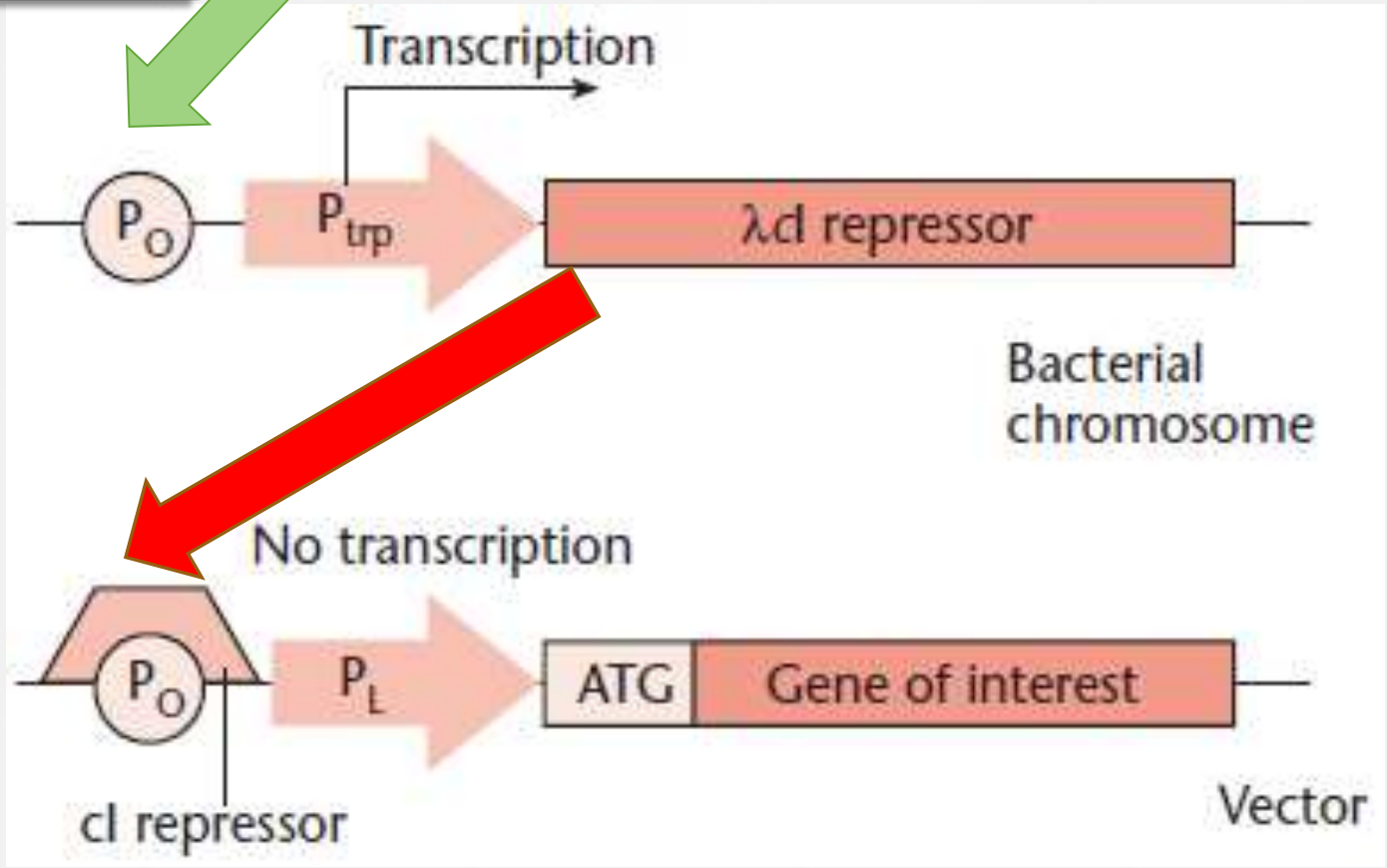
2. Lac operator is placed between T7 promoter & cloned gene – it will reduce transcription of the insert in absence of IPTG.


IPTG induction
E.coli RNA polymerase
T7 gene 1
lacO
lac promoter
DE3
lac repressor
lacI gene
E.coli genome
Host cell
T7 RNA polymerase
Inactive
T7 lysozyme
T7 lysozyme gene
pLysS or E
IPTG Induction
T7 RNA polymerase
Target gene
lacO
T7 promoter
lac repressor
pET
lacI gene

- λ P_L promoter system combines very tight transcriptional control with high level of gene expression.
- Cloned gene is under the control of P_L promoter
- P_L promoter is controlled by CI repressor gene in E. coli host
- CI is under the control of tryptophan (trp) promoter.

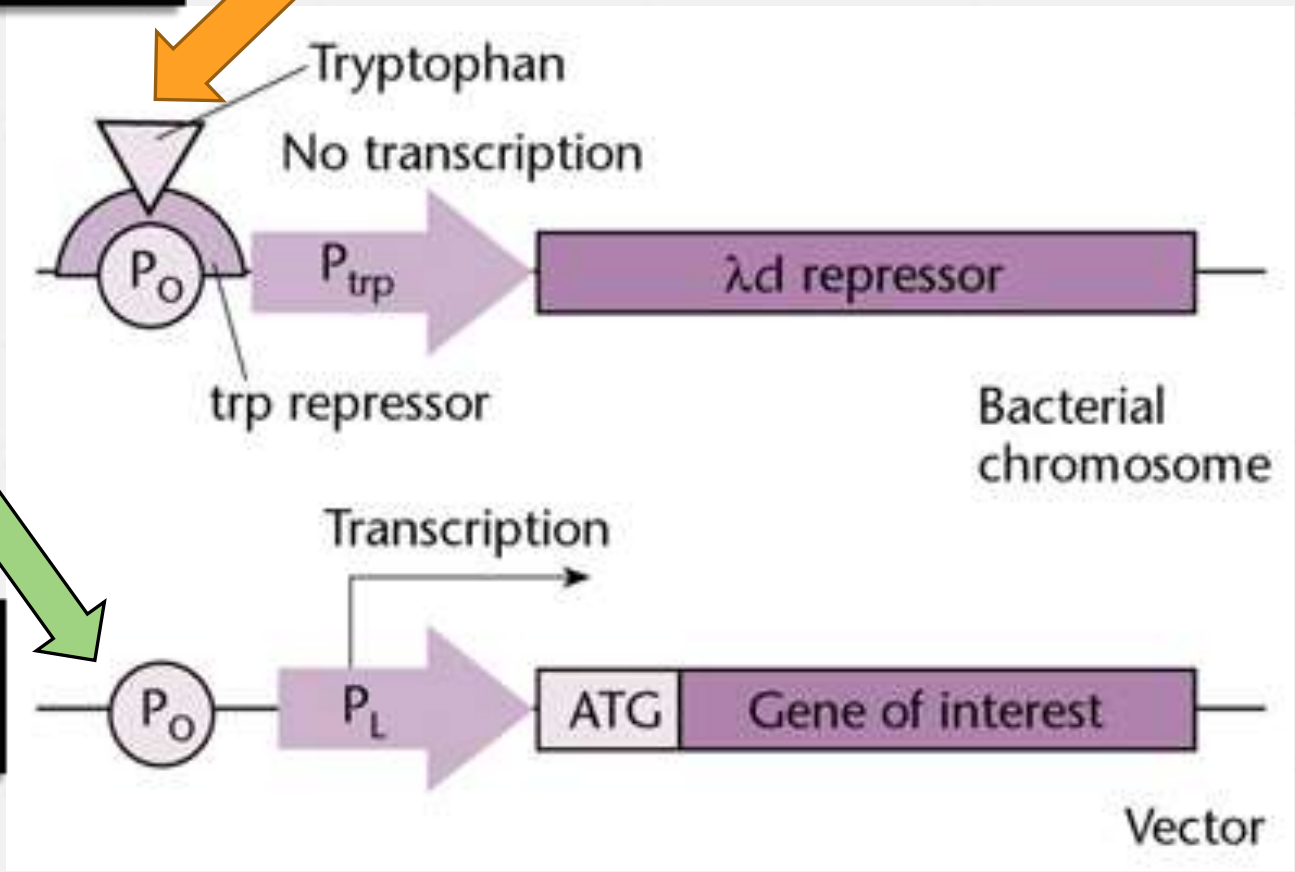
No Tryptophan

No trp repressor



Tryptophan

Trp repressor

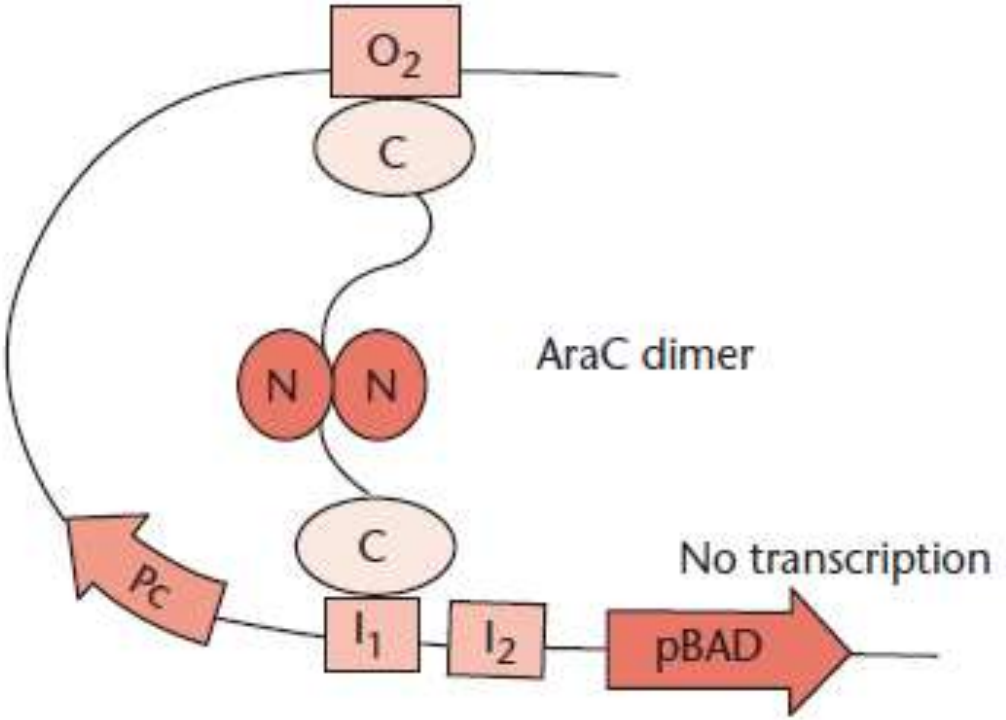


No CI repressor

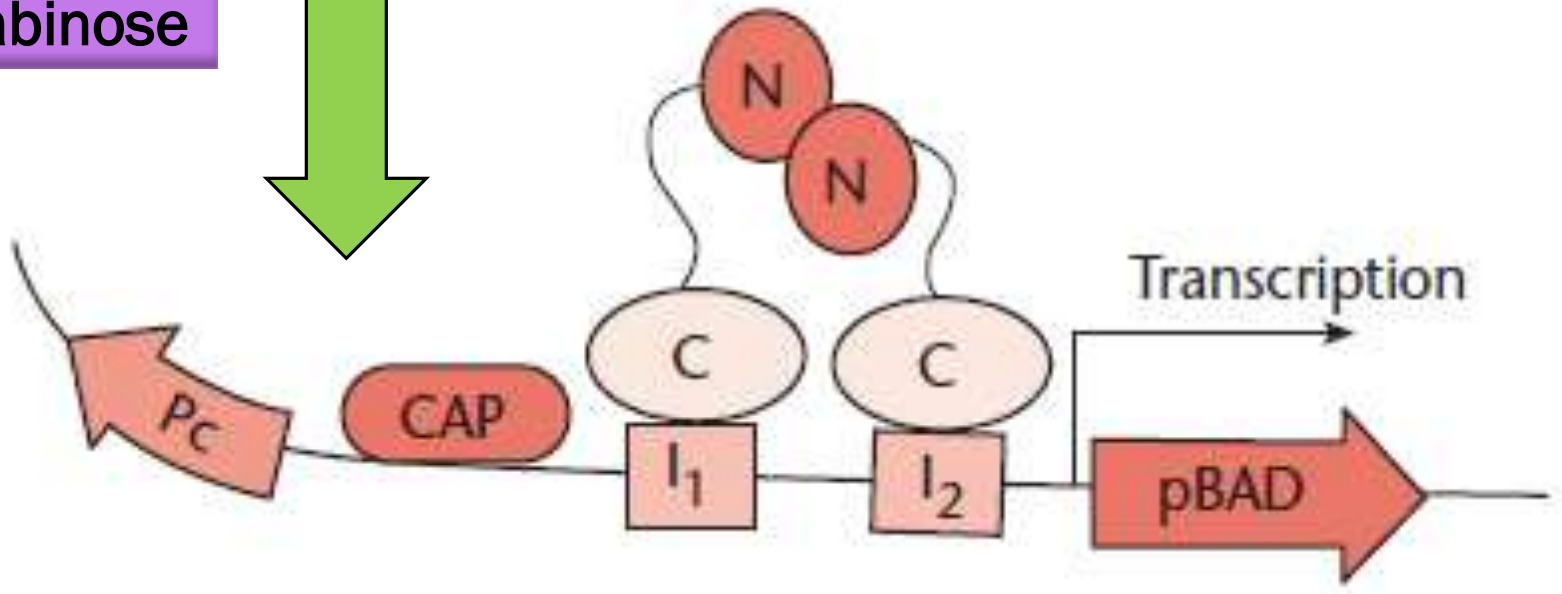
pBAD vectors

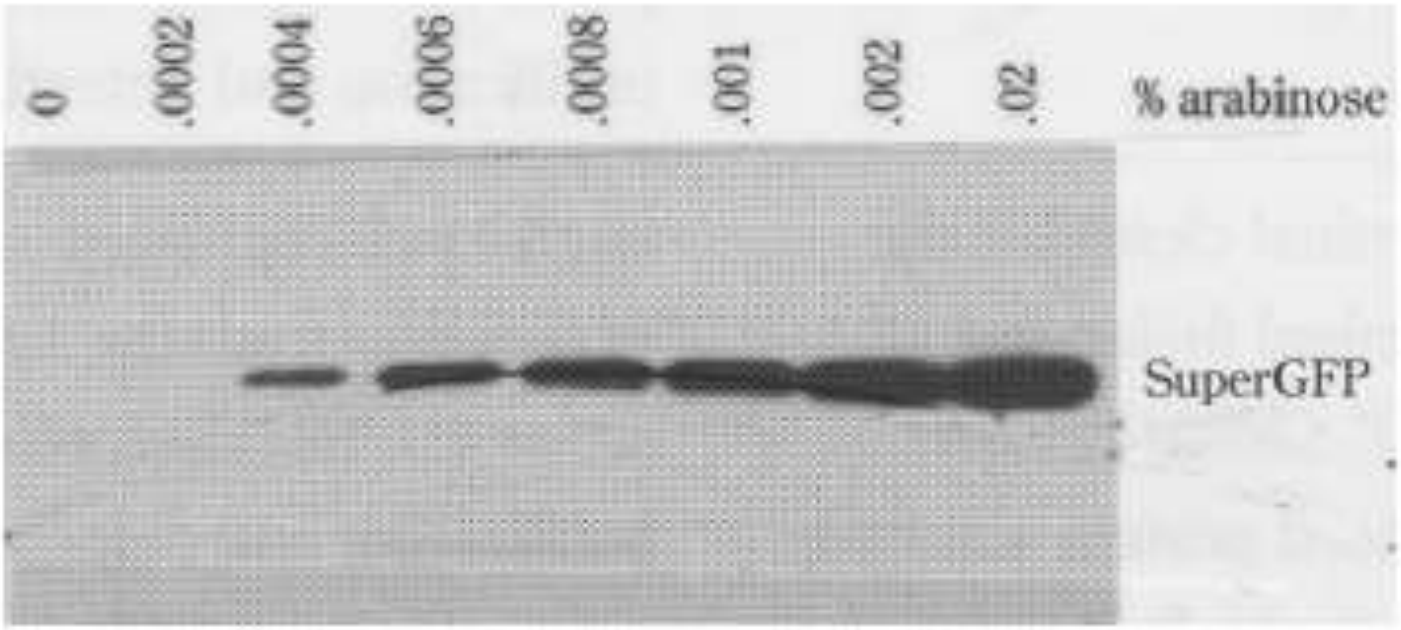
- pBAD vectors – extremely tight control on expression.
- They carry promoter of araBAD (arabinose) operon &
- Gene coding +ve & -ve regulator of this promoter “araC”
- AraC –transcriptional regulator – complex with L-arabinose.
- In the absence of arabinose, AraC binds to the O_2 and I_1 half sites of the araBAD operon – forming 210 bp DNA loop – blocking transcription.

- Arabinose – binds to araC – releasing O_2 site
- AraC binds to I_2 adjacent to the I_1 site – releasing DNA loop – transcription begins.
- Binding of AraC to I_1 and I_2 – activated in presence of cAMP activator protein (CAP) + cAMP.
- If glucose – repression of cAMP synthesis – decreasing expression from araBAD promoter.


O_2
C
N
N
AraC dimer
C
No transcription
Pc
I_1
I_2
pBAD

+ Arabinose


N
N
Transcription
C
C
Pc
CAP
I_1
I_2
pBAD


0
.0002
.0004
.0006
.0008
.001
.002
.02
% arabinose
SuperGFP

Vectors to facilitate protein purification

- Cloning vectors have been engineered so the protein being expressed will be fused to another protein – “Tag”.
- This facilitates in protein purification.
- E.g. Glutathione-S-transferase, MalE (maltose binding) protein & Multiple histidine residues – easily purified by affinity chromatography.

- Tag vectors created - coding sequence for an amino acid sequence cleaved by a specific protease is inserted between the coding sequence for the tag and gene being expressed.

- After purification the tag protein can be cleaved off with the specific protease to leave a normal or nearly normal protein.

- It is possible to include in the tag a protein sequence that can be assayed easily.

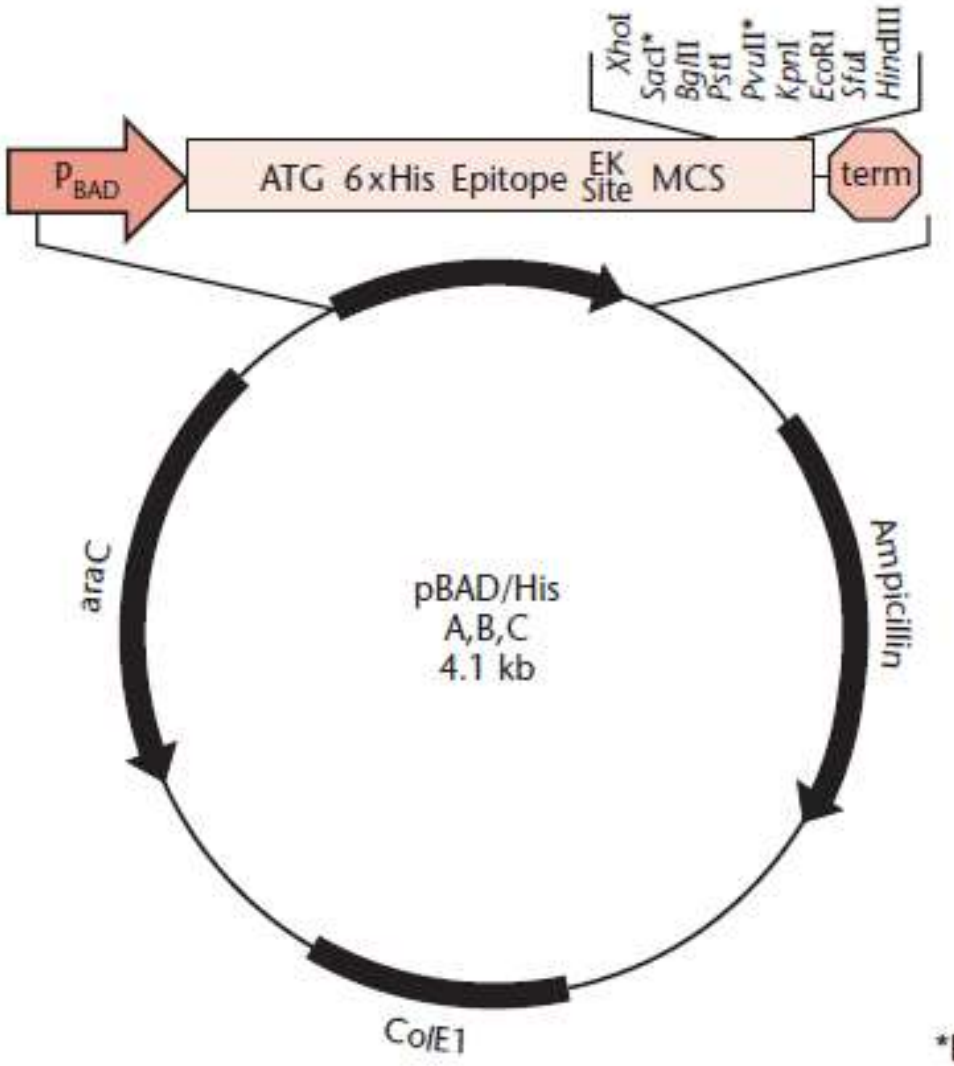
- Permitting the assay of the cloned gene product when its activity is not known or when the usual assay is inconvenient.

Polyhistidine tag

- Gene of interest is first engineered into a vector in which there is a polylinker downstream of six histidine residues and a proteolytic cleavage site.
- The cleavage site is that for enterokinase.
- After induction of synthesis of the fusion protein, the cells are lysed and the viscosity of the lysate is reduced by nuclease treatment.

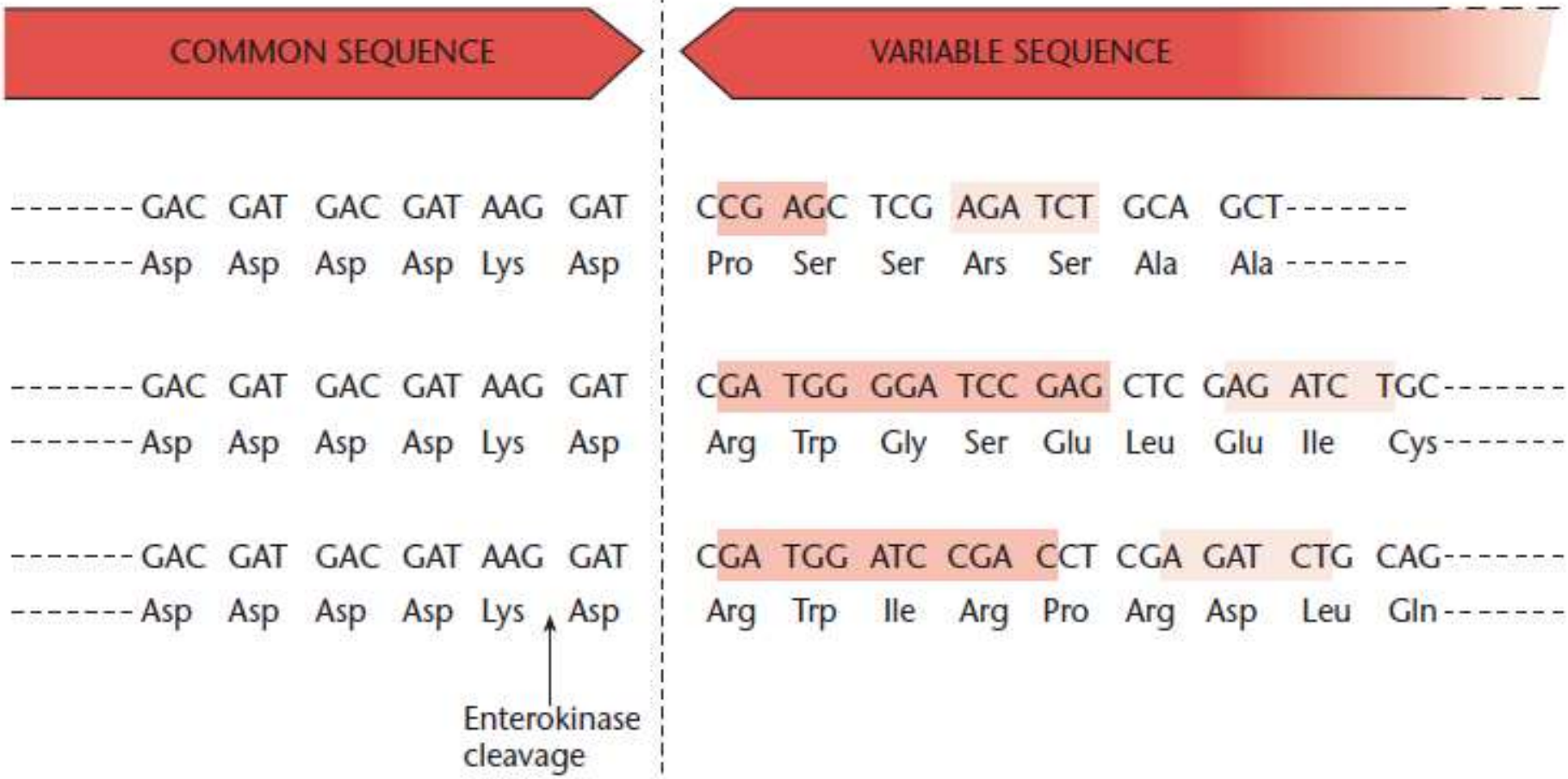
- The lysate is then applied to a column containing immobilized divalent nickel, which selectively binds the polyhistidine tag.

- After washing away any contaminating proteins, the fusion protein is eluted from the column and treated with enterokinase to release the cloned gene product.



*Frame-dependent variations

- For the cloned gene to be expressed correctly, it has to be in the correct translational reading frame.
- This is achieved by having three different vectors, each with a polylinker in a different reading frame.
- Enterokinase recognizes the sequence (Asp)4Lys and cleaves immediately after the lysine residue.
- Therefore, after enterokinase cleavage, the cloned gene protein will have a few extra amino acids at the N terminus.


COMMON SEQUENCE
VARIABLE SEQUENCE
-------GAC GAT GAC GAT AAG GAT
-------Asp Asp Asp Asp Lys Asp
CCG AGC TCG AGA TCT GCA GCT-------
Pro Ser Ser Ars Ser Ala Ala -------
-------GAC GAT GAC GAT AAG GAT
-------Asp Asp Asp Asp Lys Asp
CGA TGG GGA TCC GAG CTC GAG ATC TGC-------
Arg Trp Gly Ser Glu Leu Glu Ile Cys-------
-------GAC GAT GAC GAT AAG GAT
-------Asp Asp Asp Asp Lys Asp
CGA TGG ATC CGA CCT CGA GAT CTG CAG-------
Arg Trp Ile Arg Pro Arg Asp Leu Gln-------
Enterokinase
cleavage

- If desired, the cleavage site and polyhistidines can be synthesized at the C terminus.

- If the cloned gene product itself contains an enterokinase cleavage site, then an alternative protease, such as thrombin or factor Xa, with a different cleavage site can be used.

- To facilitate assay of the fusion proteins, short antibody recognition sequences can be incorporated into the tag between the affinity label and the protease cleavage site.

- These antibodies can be used to detect, by western blotting, fusion proteins carrying the appropriate epitope.

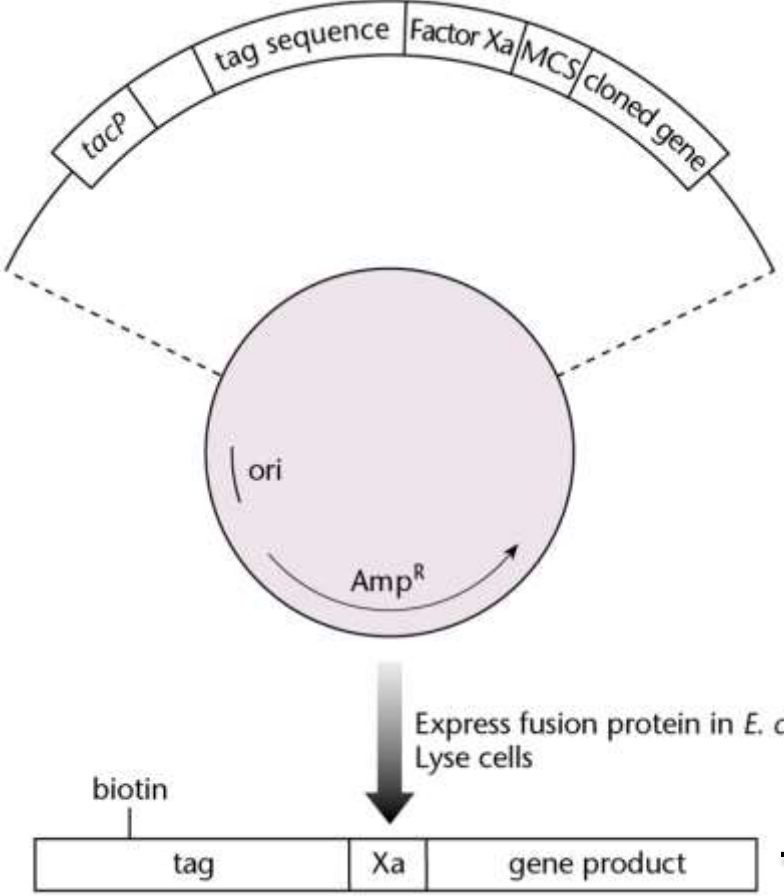
- Note that a polyhistidine tag at the C terminus can function for both assay and purification.

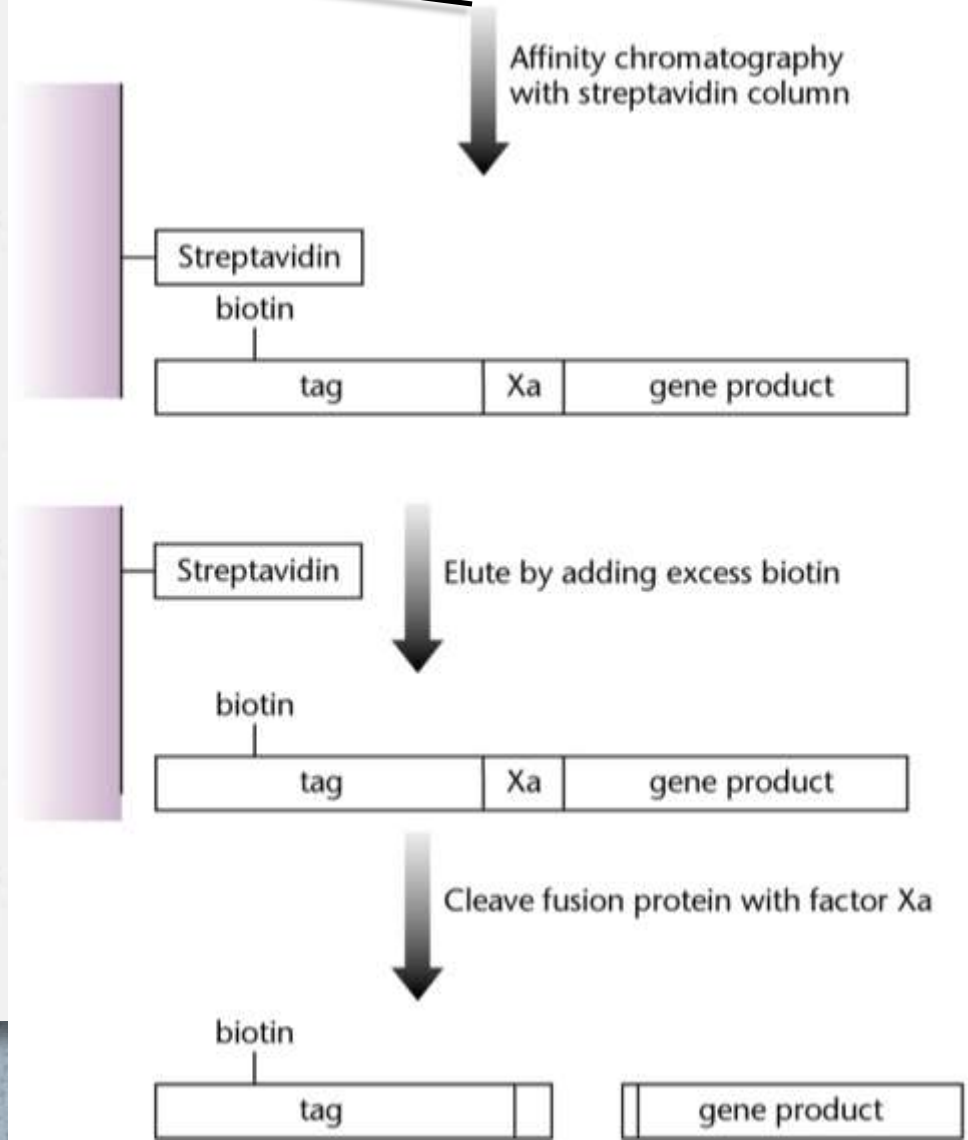
Peptide sequence	Antibody recognition
-Glu-Gln-Lys-Leu-Ile Ser-Glu-Glu-Asp-Leu-	Anti-*myc* antibody
-His-His-His-His-His-His-COOH	Anti-His (C-terminal) antibody
-Gly-Lys-Pro-Ile-Pro-Asn-Pro-Leu-Leu-Gly-Leu-Asp-Ser-Thr-	Anti-V5 antibody

Biotinylated protein

- Biotin is an essential cofactor for a number of carboxylases important in cell metabolism. The biotin in these enzyme complexes is covalently.
- attached at a specific lysine residue of the biotin carboxylase carrier protein.
- Fusions made to a segment of the carrier protein are recognized in *E.coli* by biotin ligase, the product of the *bir*A gene, and biotin is covalently attached in an ATP-dependent reaction.

- The expressed fusion protein can be purified using streptavidin affinity chromatography.
- *E.coli* expresses a single endogenous biotinylated protein, but it does not bind to streptavidin in its native configuration, making the affinity purification highly specific for the recombinant fusion protein.
- The presence of biotin on the fusion protein has an additional advantage: its presence can be detected with enzymes coupled to streptavidin.


tacP
tag sequence
Factor Xa
MCS
cloned gene
ori
Amp^R
Express fusion protein in E. coli
Lyse cells
biotin
tag
Xa
gene product

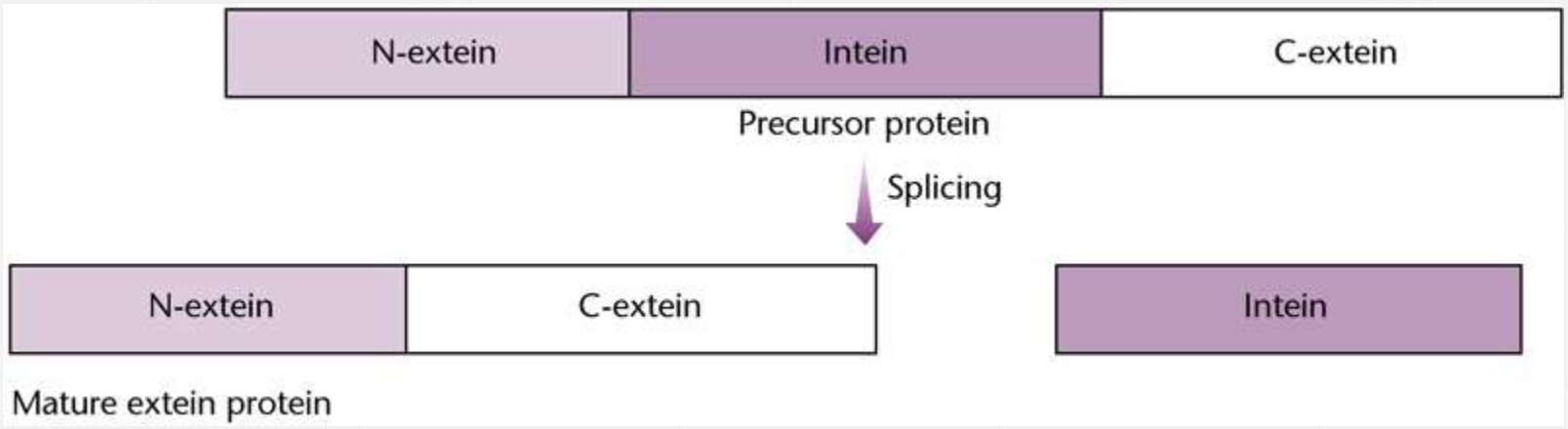

Affinity chromatography
with streptavidin column
Streptavidin
biotin
tag
Xa
gene product
Streptavidin
Elute by adding excess biotin
biotin
tag
Xa
gene product
Cleave fusion protein with factor Xa
biotin
tag
gene product

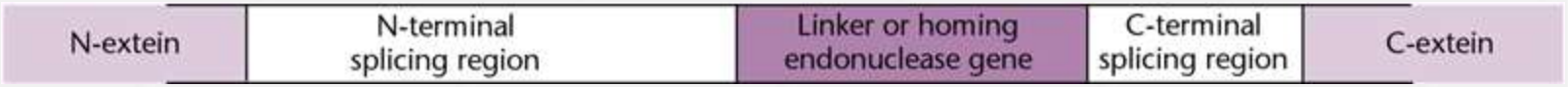
Diasdvantages

- The affinity purification systems described above suffer from the disadvantage that
- a protease is required to separate the target protein from the affinity tag.
- Also, the protease has to be separated from the protein of interest.

Intein splicing

- The system utilizes a protein splicing element, an intein, from the *Saccharomyces cerevisiae VMA1* gene.
- The intein is modified such that it undergoes a self-cleavage reaction at its N terminus at low temperatures in the presence of thiols, such as cysteine, dithiothreitol or β-mercaptoethanol.

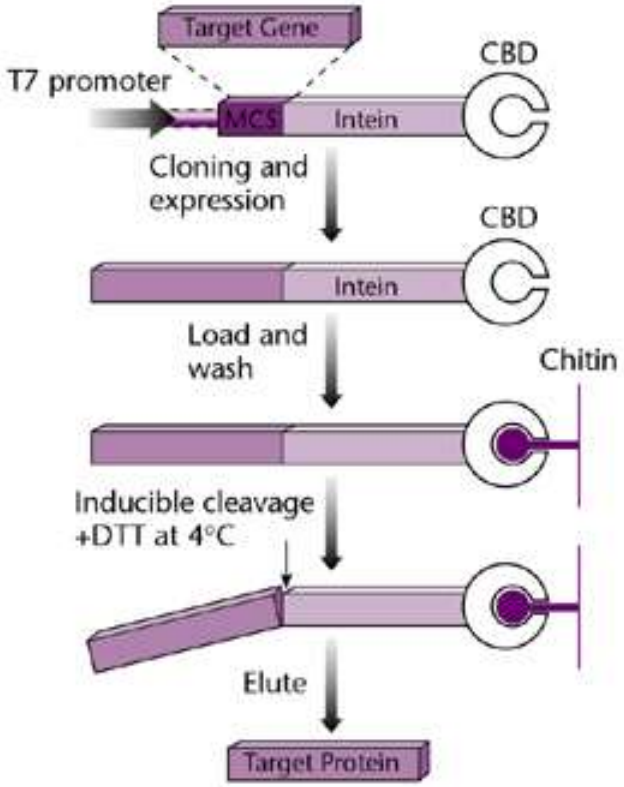

N-extein
Intein
C-extein
Precursor protein
Splicing
N-extein
C-extein
Intein
Mature extein protein

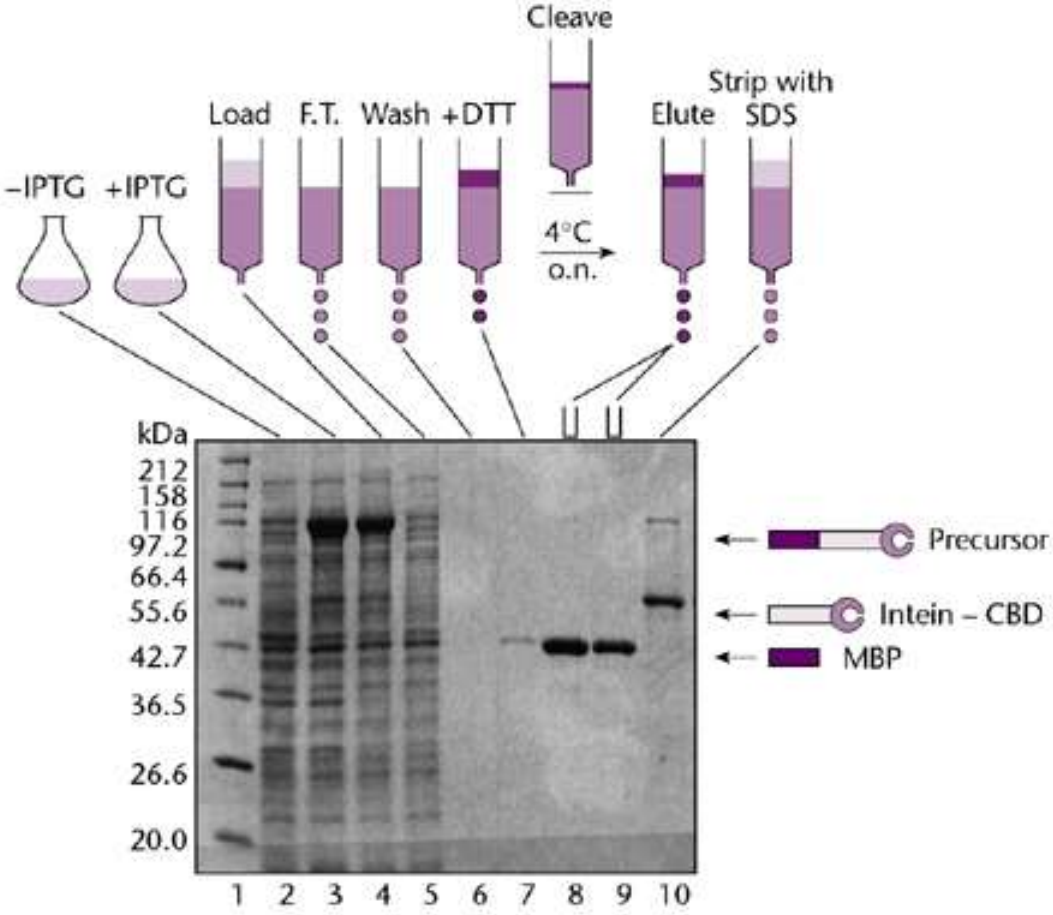

N-extein
N-terminal splicing region
Linker or homing endonuclease gene
C-terminal splicing region
C-extein

- The gene encoding the target protein is inserted into a multiple cloning site (MCS) of a vector to create a fusion between the C terminus of the target gene and the N terminus of the gene encoding the intein.

- DNA encoding a small (5 kDa) chitin-binding domain from *Bacillus circulans* was added to the C terminus of the intein for affinity purification

- The above construct is placed under the control of an IPTG-inducible T7 promoter.
- When crude extracts from induced cells are passed through a chitin column, the fusion protein binds and all contaminating proteins are washed through.
- The fusion is then induced to undergo intein-mediated self-cleavage on the column by incubation with a thiol.
- This releases the target protein, while the intein chitin-binding domain remains bound to the column.

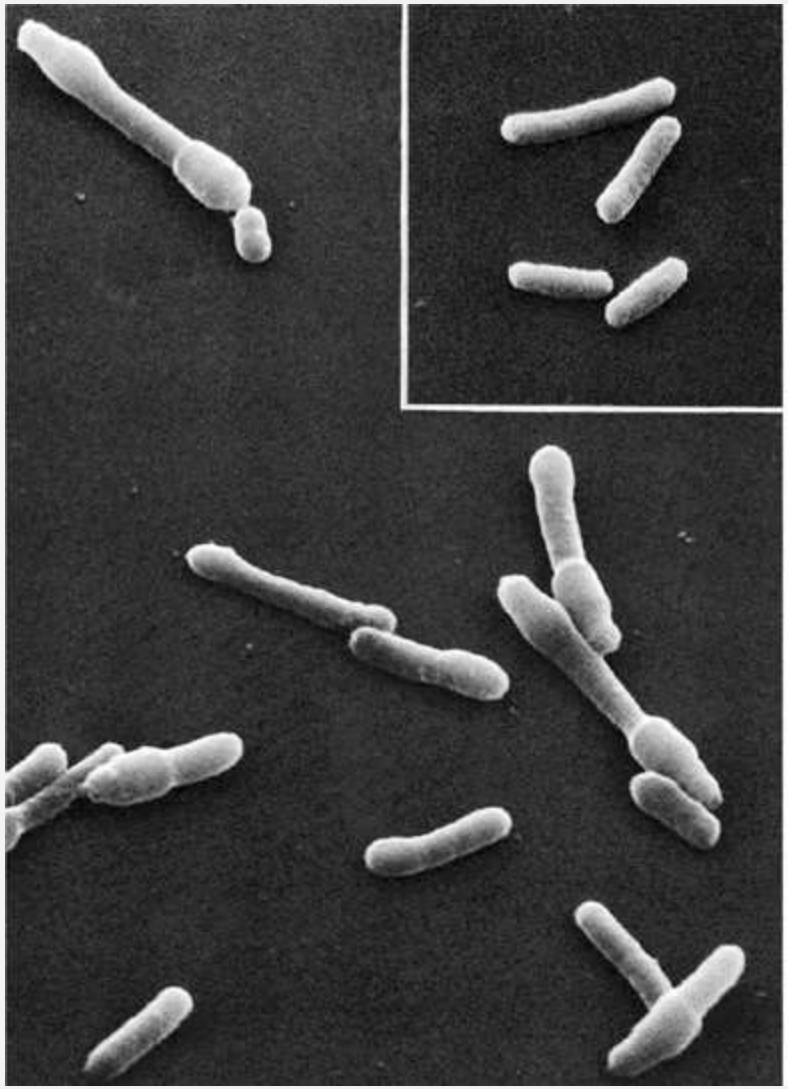


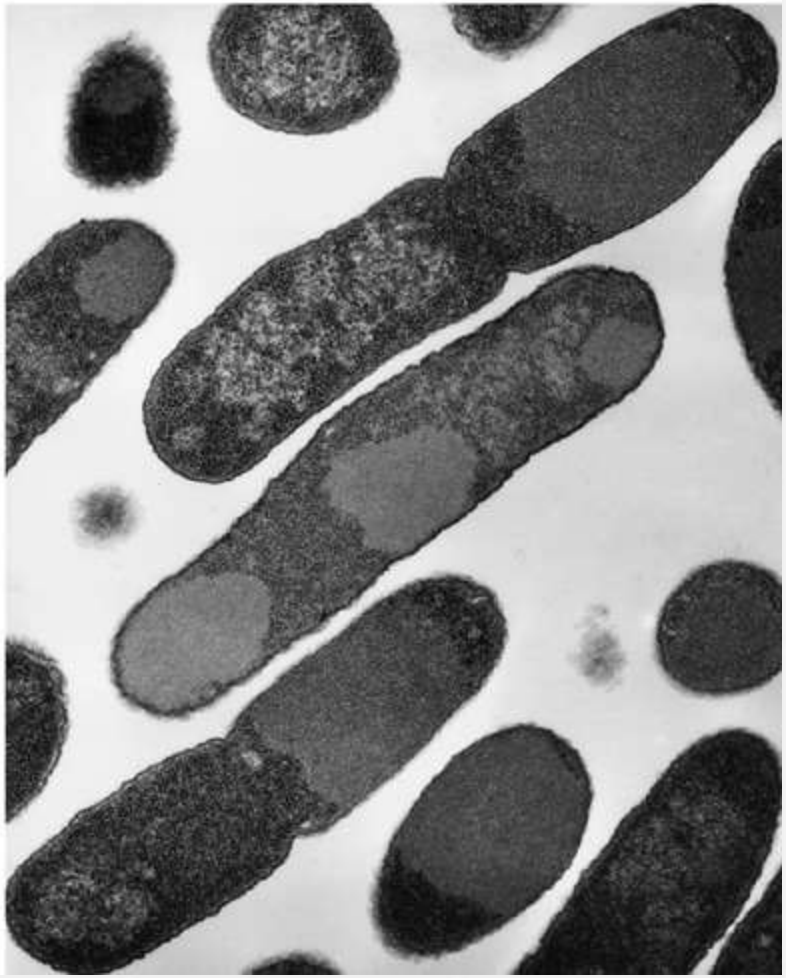


Lane 1: Protein Marker.
Lane 2: Crude extract from uninduced cells.
Lane 3: Crude extract from cells, induced at 15°C for 16 hours.
Lane 4: Clarified crude extract from induced cells.
Lane 5: Chitin column flow through (F.T.).
Lane 6: Chitin column wash.
Lane 7: Quick DTT wash to distribute DTT evenly throughout the chitin column.
Lanes 8-9: Fraction of eluted MBP after stopping column flow and inducing a self-cleavage reaction at 4°C overnight.
Lane 10: SDS stripping of remaining proteins bound to chitin column (mostly the cleaved intein-CBD fusion).

Vectors to promote solubilization of expressed proteins

- One of the problems associated with the overproduction of proteins in *E. coli* is the sequestration of the product into insoluble aggregates or 'inclusion bodies'
- They were first reported in strains overproducing insulin A and B chains.





- At first, their formation was thought to be restricted to the overexpression of heterologous proteins in *E. coli*, but they can form in the presence of high levels of normal *E. coli* proteins, e.g. subunits of RNA polymerase.

- Two parameters that can be manipulated to reduce inclusion-body formation are temperature and growth rate. There are a number of reports which show that lowering the temperature of growth increases the yield of correctly folded, soluble protein.

- Media compositions and pH values that reduce the growth rate also reduce inclusion body formation.

- Renaturation of misfolded proteins can sometimes be achieved following solubilization in guanidinium hydrochloride (Lilie *et al.* 1998).

- Three ‘genetic’ methods of preventing inclusion body formation:

- The host cell is engineered to overproduce a chaperon (e.g. DnaK, GroEL or GroES proteins) in addition to the protein of interest.
- A series of vectors which are compatible with pBR322-type plasmids and which encode the overproduction of chaperons.
- These vectors can be used to test the effect of chaperons on the solubilization of heterologous gene products.
- Even with excess chaperon there is no guarantee of proper folding.

- The second method involves making minor changes to the amino acid sequence of the target protein.
- For example, cysteine-to-serine changes in fibroblast growth factor minimized inclusion-body formation (Rinas *et al.* 1992).

- The third method is derived from the observation that many proteins produced as insoluble aggregates in their native state are synthesized in soluble form as thioredoxin fusion proteins (LaVallie *et al.* 1993).

- More recently, Davis *et al.* (1999) have shown that the NusA and GrpE proteins, as well as bacterioferritin, are even better than thioredoxin at solubilizing proteins expressed at a high level.

- Kapust and Waugh (1999) have reported that the maltose-binding protein is also much better than thioredoxin.

- A series of vectors has been developed in which the gene of interest is cloned into an MCS and the gene product is produced as a thioredoxin fusion protein with an enterokinase cleavage site at the fusion point.

- After synthesis, the fusion protein is released from the producing cells by osmotic shock and purified.

- The desired protein is then released by enterokinase cleavage.
- To simplify the purification of thioredoxin fusion proteins systematically mutated a cluster of surface amino acid residues.
- Residues 30 and 62 were converted to histidine and the modified (‘histidine patch’) thioredoxin could now be purified by affinity chromatography on immobilized divalent nickel.

- An alternative purification method was developed:
- A gene in which a short biotinylation peptide is fused to the N terminus of the thioredoxin gene to generate a new protein called BIOTRX.
- They constructed a vector carrying the BIOTRX gene, with an MCS at the C terminus, and the *bir*A gene. After cloning a gene in the MCS, a fused protein is produced which can be purified by affinity chromatography on streptavidin columns.

- An alternative way of keeping recombinant proteins soluble is to export them to the periplasmic space.
- However, even here they may still be insoluble.

- Barth *et al.* (2000) solved this problem by growing the producing bacteria under osmotic stress (4% NaCl plus 0.5 mol/l sorbitol) in the presence of compatible solutes.

- Compatible solutes are low-molecular-weight osmolytes, such as glycine betaine, that occur naturally in halophilic bacteria and are known to protect proteins at high salt concentrations.

- Adding glycine betaine for the cultivation of *E. coli* under osmotic stress not only allowed the bacteria to grow under these otherwise inhibitory conditions but also produced a periplasmic environment for the generation of correctly folded recombinant proteins.

Vectors to promote protein export

- Gram-negative bacteria such as *E. coli* have a complex wall–membrane structure comprising an inner, cytoplasmic membrane separated from an outer membrane by a cell wall and periplasmic space.

- Secreted proteins may be released into the periplasm or integrated into or transported across the outer membrane.

- In *E. coli* it has been established that protein export through the inner membrane to the periplasm or to the outer membrane is achieved by a universal mechanism known as the general export pathway (GEP).

- This involves the *sec* gene products.

- Proteins that enter the GEP are synthesized in the cytoplasm with a signal sequence at the N terminus.
- This sequence is cleaved by a signal or leader peptidase during transport.
- A signal sequence has three domains: a positively charged amino-terminal region, a hydrophobic core, consisting of five to 15 hydrophobic amino acids, and a leader peptidase cleavage site.
- A signal sequence attached to a normally cytoplasmic protein will direct it to the export pathway.

- Many signal sequences derived from naturally occurring secretory proteins (e.g. OmpA, OmpT, PelB, β-lactamase, alkaline phosphatase and phage M13 gIII)

- support the efficient translocation of heterologous peptides across the inner membrane when fused to their amino termini.

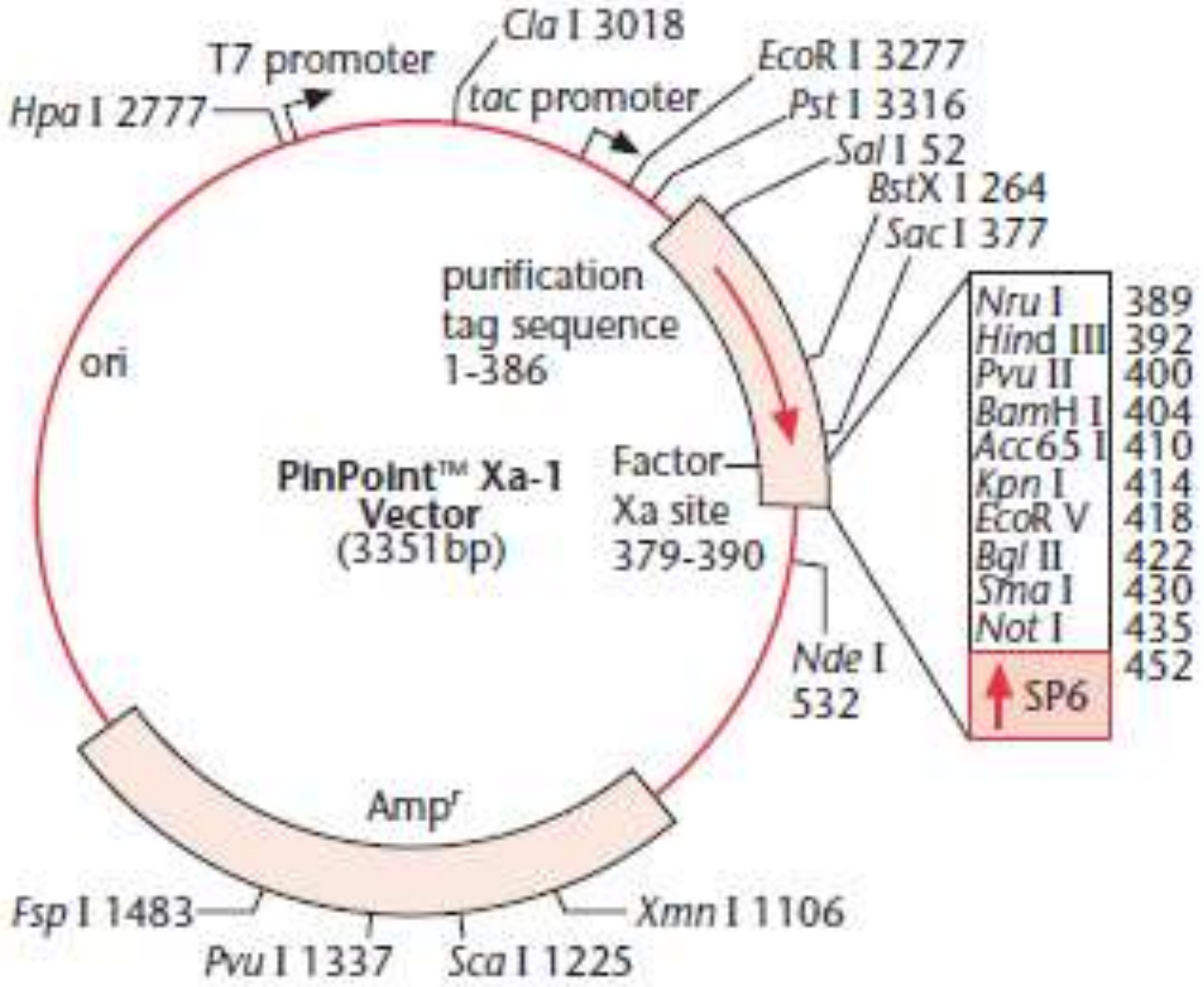
- In some cases, however, the preproteins are not readily exported and either become ‘jammed’ in the inner membrane, accumulate in precursor inclusion bodies or are rapidly degraded within the cytoplasm.

- In practice, it may be necessary to try several signal sequences (Berges *et al.* 1996) and/or overproduce different chaperons to optimize the translocation of a particular heterologous protein

- A first step would be to try the secretion vectors offered by a number of molecular biology suppliers and which are variants of the vectors described above.

- It is possible to engineer proteins such that they are transported through the outer membrane and are secreted into the growth medium. This is achieved by making use of the type I, Sec-independent secretion system.

- The prototype type I system is the haemolysin transport system, which requires a short carboxy-terminal secretion signal, two translocators (HlyB and D), and the outer-membrane protein TolC.
- If the protein of interest is fused to the carboxyterminal secretion signal of a type I-secreted protein, it will be secreted into the medium provided HlyB, HlyD and TolC are expressed as well.
- An alternative presentation of recombinant proteins is to express them on the surface of the bacterial cell using any one of a number of carrier proteins


Cla I 3018
T7 promoter
EcoR I 3277
Hpa I 2777
tac promoter
Pst I 3316
Sal I 52
BstX I 264
Sac I 377
purification
tag sequence
1-386
ori
Nru I 389
Hind III 392
Pvu II 400
BamH I 404
Acc65 I 410
Kpn I 414
EcoR V 418
Bgl II 422
Sma I 430
Not I 435
452
PinPoint™ Xa-1
Vector
(3351bp)
Factor
Xa site
379-390
Nde I
532
SP6
Amp^r
Fsp I 1483
Xmn I 1106
Pvu I 1337
Sca I 1225

THANK YOU