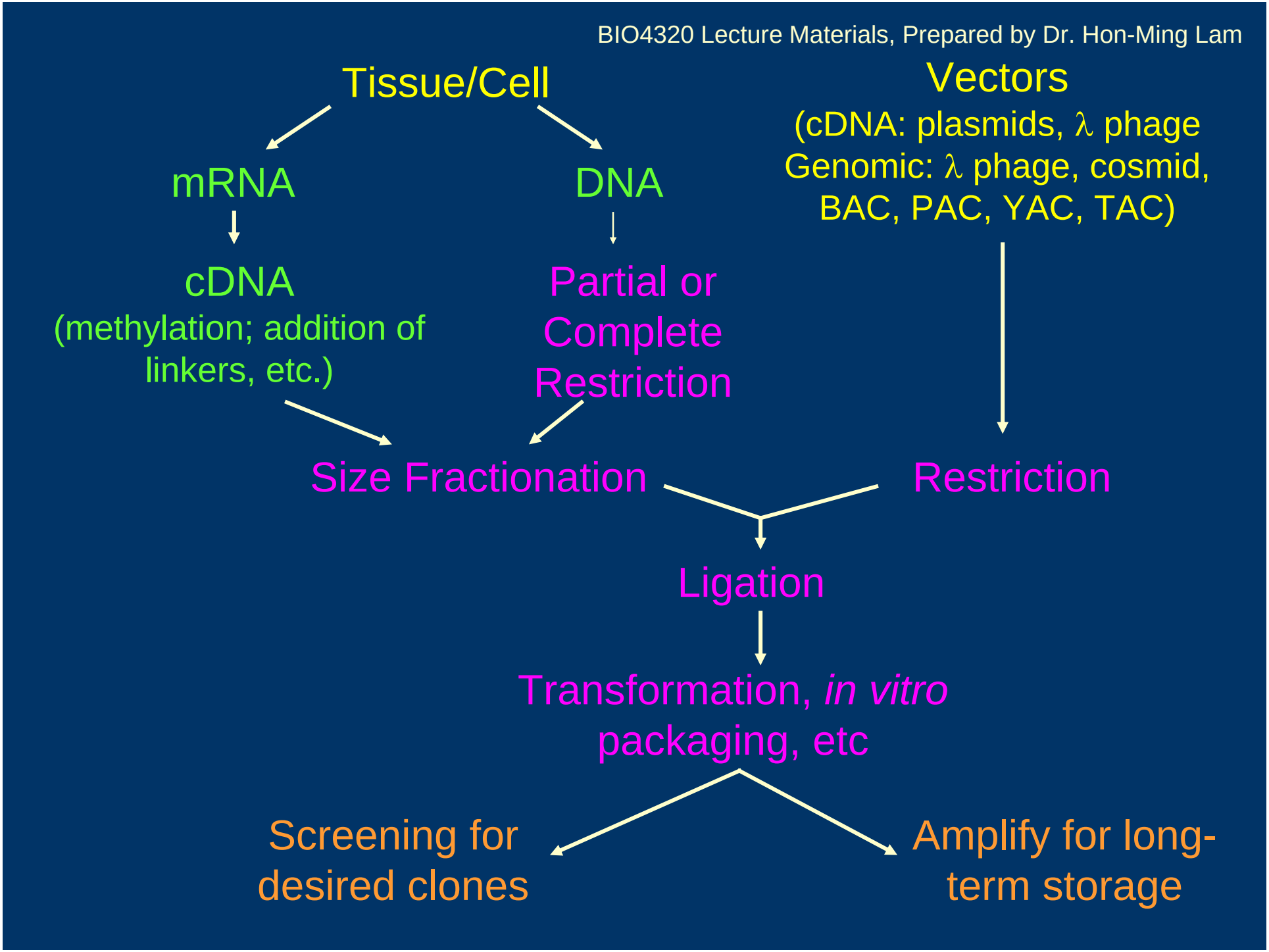
Construction and Screening of Gene Libraries

Further Readings:

"Genome II" by TA Brown, Ch. 4;

"Current Protocols in Molecular Biology" by Ausubel et al., Ch. 5 and 6;

PNAS 88:1731-35


Tissue/Cell
mRNA
DNA
cDNA
(methylation; addition of linkers, etc.)
Partial or Complete Restriction
Vectors
(cDNA: plasmids, λ phage
Genomic: λ phage, cosmid, BAC, PAC, YAC, TAC)
Size Fractionation
Restriction
Ligation
Transformation, *in vitro* packaging, etc
Screening for desired clones
Amplify for long-term storage

Genomic Libraries

- from genomic DNA
- frequency of hits independent of gene expression levels
- may contain promoters and introns
- cannot express in heterologous system
- useful for genome analysis, map-based cloning, promoter studies, etc

cDNA Libraries

- reverse transcription of mRNA
- dependent
- no promoters or introns
- expression is feasible if linked to a suitable promoter
- useful for analysis of coding regions and gene functions

Representation and Randomness

$Q = (1 – 1/n)^N$

$P = 1 – Q$

$P = 1 – (1 – 1/n)^N$

$(1 – 1/n)^N = 1 – P$

$\ln (1 – 1/n)^N = \ln (1 – P)$

$N = \ln (1-P) / \ln (1-1/n)$

Q: probability of a clone not in a random library

P: probability of including any DNA sequence in a random library

N: independent recombinant clones to be screened

n: total number of clones needed to cover the total genome if the clones are not overlapping (total size of genome divided by the average size of a single cloned fragment)

Representation and Randomness

Genomic Libraries

$N = \ln (1 - P) / \ln (1 - 1/n)$

P: probability of including any DNA sequence in a random library

N: independent recombinant clones to be screened

n: total number of clones needed to cover the total genome if the clones are not overlapping (total size of genome divided by the average size of a single cloned fragment)

cDNA Libraries

$N = \ln (1 - P) / \ln (1 - m/T)$

P: probability that each mRNA will be represented once

N: independent recombinant clones to be screened

m: number of molecules of the rarest mRNA in a cell

T: total number of mRNA molecules in a cell

Representation and Randomness

Genomic Libraries

$N = \ln (1 - P) / \ln (1 - 1/n)$

To have 99% chance of getting a desired sequence, screen 4.6 times the total number of base pairs

cDNA Libraries

$N = \ln (1 - P) / \ln (1 - m/T)$

Since it is hard to estimated the total number of mRNA molecules and the number of rarest mRNA molecules, it is difficult to predict the exact number of clones to be screened. In general, to have at least one copy of every mRNA, 500,000 to 1,000,000 independent cDNA clones should be screened

Vectors for Genomic Libraries

More details can be found here:
http://batzerlab.lsu.edu/Hum_Mol_Gen_Lectures/Hum%20Mol%20Genet%20Lecture%203%20Feburary%203,%202004.ppt

Vector	Insert Size	Remarks
YAC (yeast artificial chromosome)	230 - 1700 kb (length of natural yeast chromosome) Average: 400-700 kb	• Propagate in *Saccharomyces cerevisiae*; • Three major elements: centromere for nuclear division; telomeres for marking the end of the chromosome; origins of replication for initiation of new DNA synthesis when the chromosome divides • Previously an important tool to map complex genomes; • Problems: chimera, instability (rearrangement)

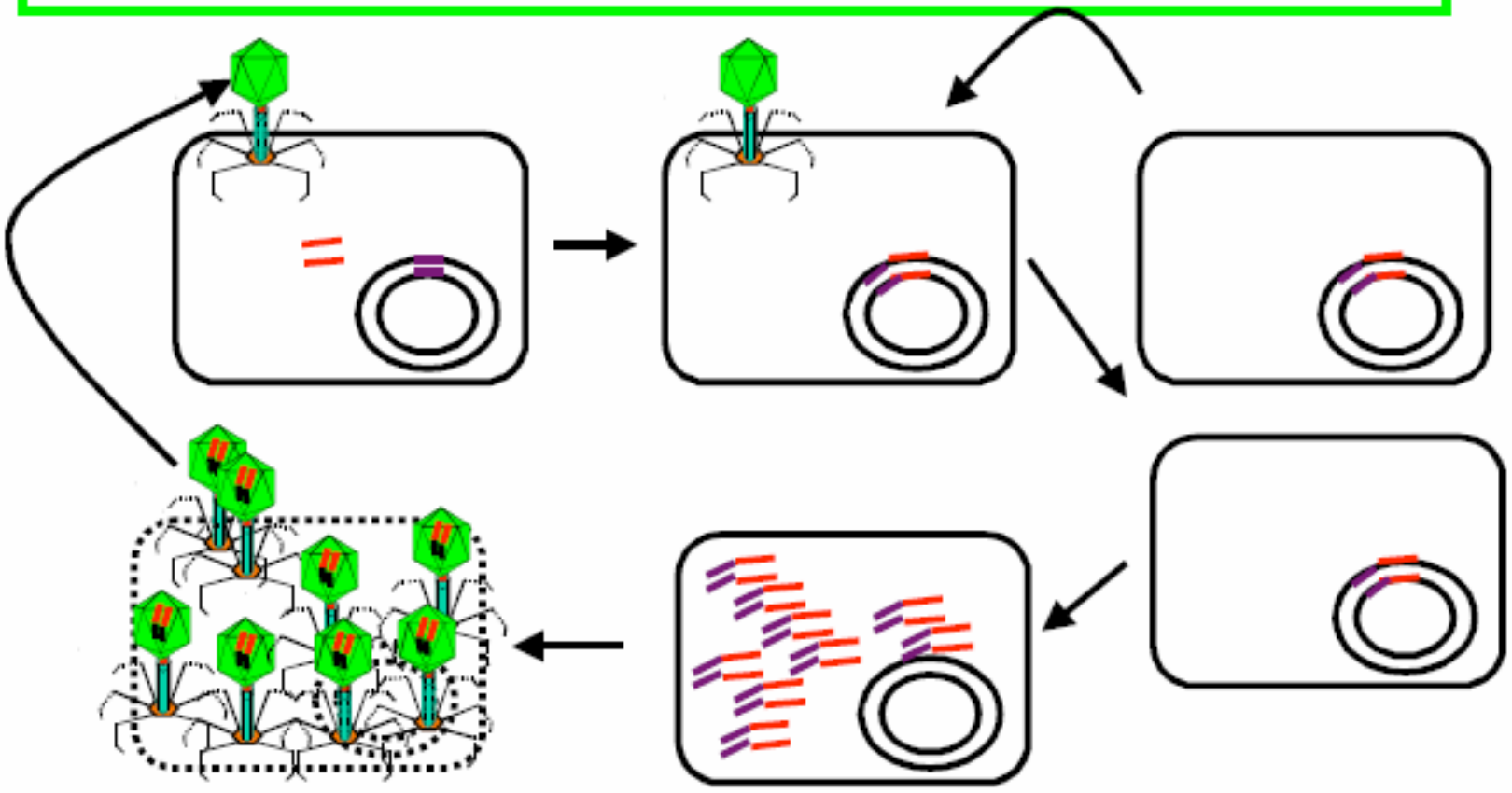
Vectors for Genomic Libraries

Vector	Insert Size	Remarks
BAC (bacterial artificial chromosome)	Up to 300 kb Average: 100 kb	• Plasmid vector containing the F factor replicon; • One copy per bacterial cell
Bacteriophage P1	Maximum about 100 kb	• Deletion version of a natural phage genome • P1 phage genome is about 110 kb • Efficient packaging system • *pac* cleavage site for recognition • P1 plamsid replicon and inducible P1 lytic replicon • *loxP* site for Cre action
PAC (P1-derived artificial chromosome)	Similar to BAC	• A combination of BAC and P1 features
TAC (Transformable artificial chromosome)	Similar to P1	• With P1 plasmid replicon (single copy in *E. coli*) and Ri plasmid replicon (single copy in *Agrobacterium tumefaciens*) • With T-DNA border and can transform plant directly

Sources: www.genetics.wisc.edu/courses/spring04/466/files/bacteria2.pdf

Specialized transduction

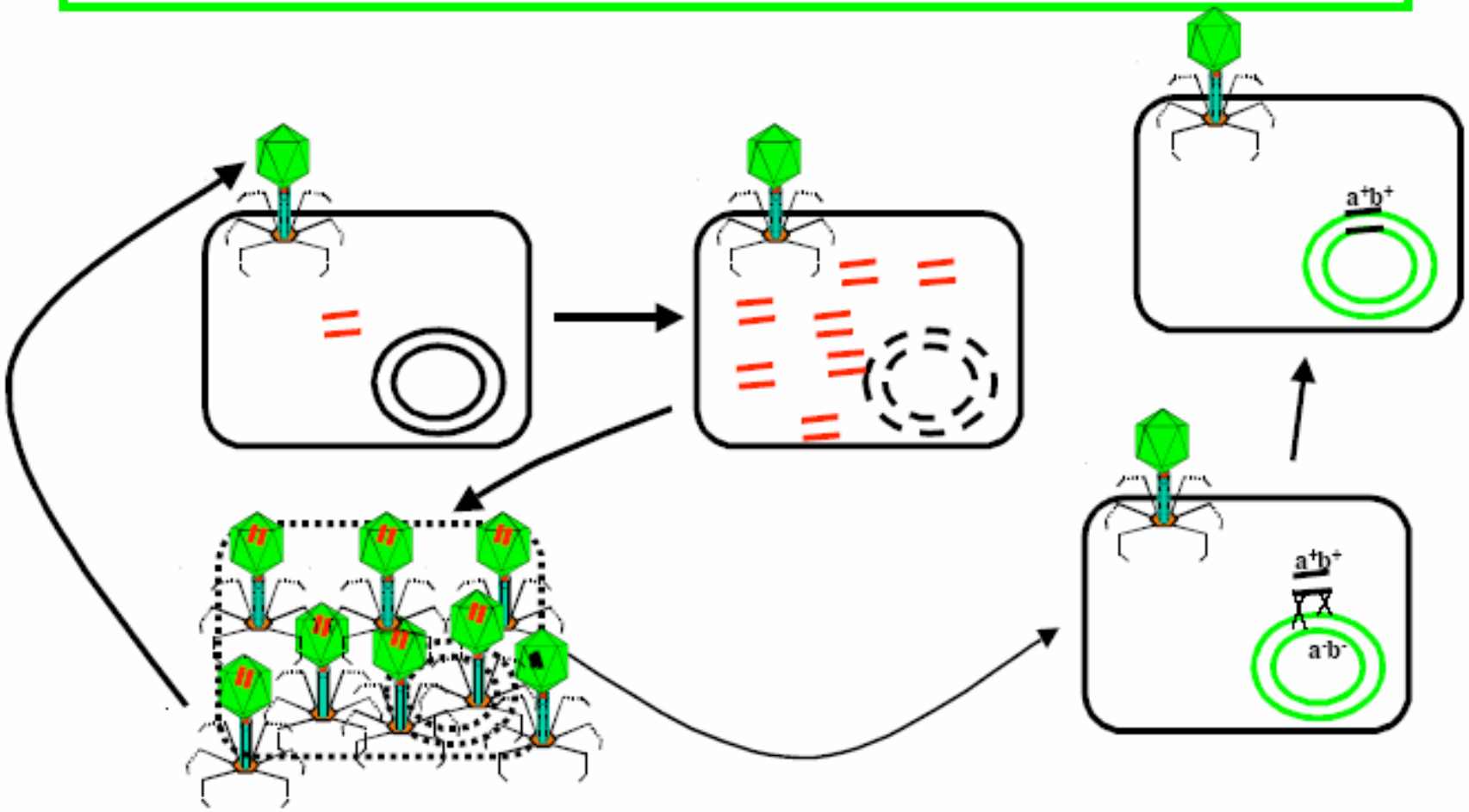
- **Imprecise excision of an integrated lysogenic phage (prophage) ==> ==> specialized transduction (information carried by the phage is restricted by prophage integration site)**



Sources: www.genetics.wisc.edu/courses/spring04/466/files/bacteria2.pdf

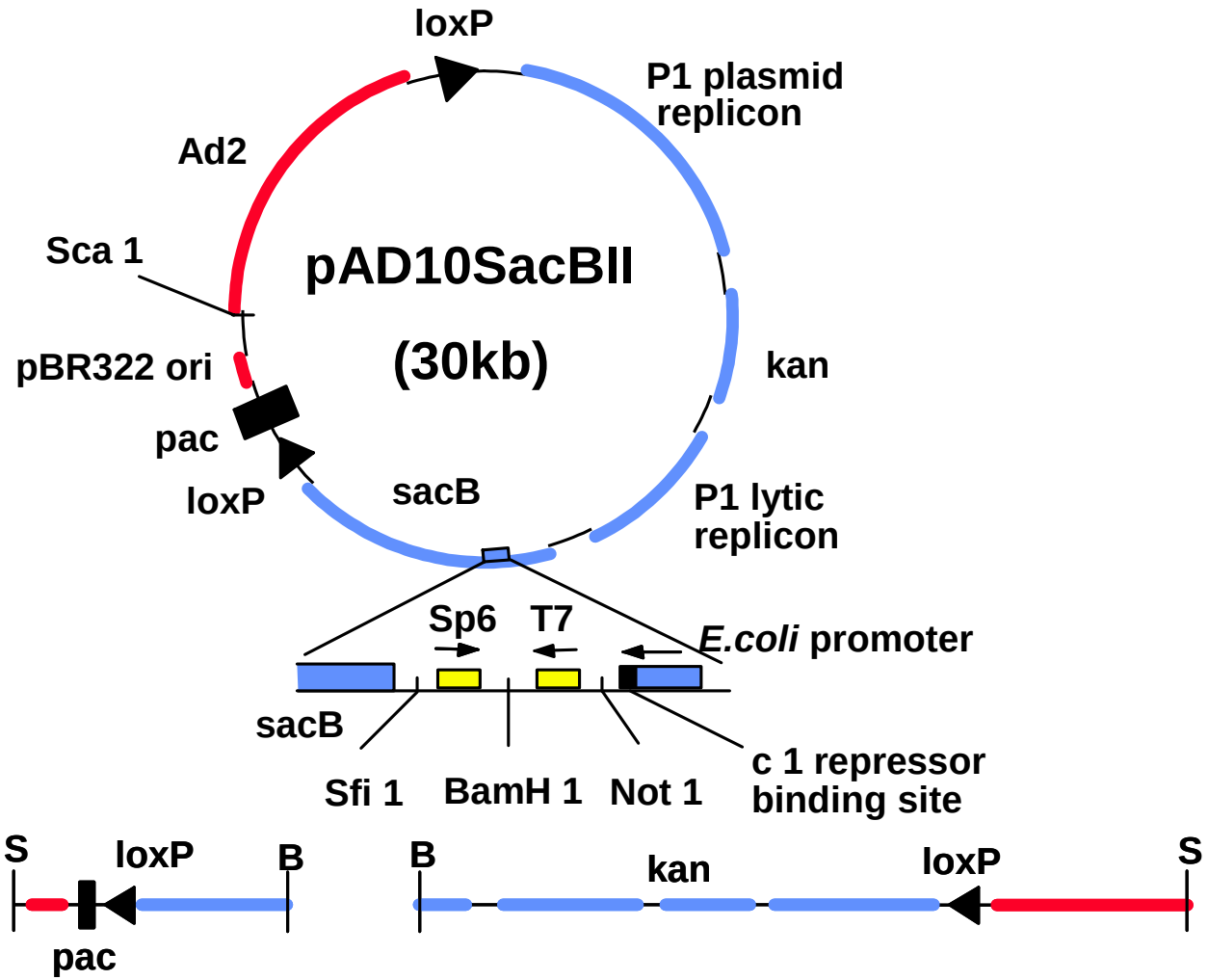
Generalized transduction

- During the lytic cycle, a phage particle packages a random piece of genomic DNA. That particle can inject its DNA into another cell
- Example: P1 phage of *E. coli* is a very efficient transducer



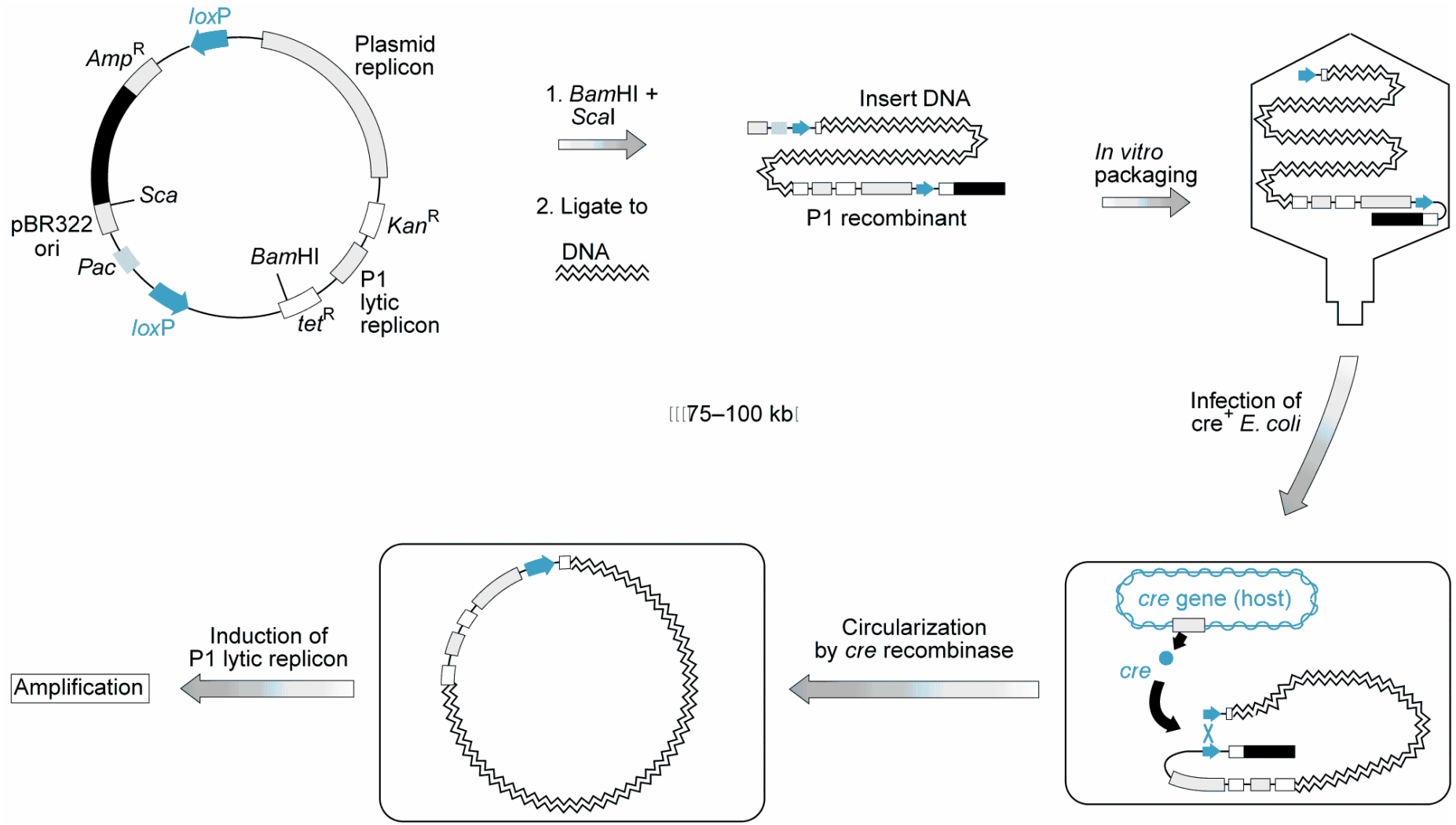
Sources: http://batzerlab.lsu.edu/Hum_Mol_Gen_Lectures/Hum%20Mol%20Genet%20Lecture%203%20February%201,%202005.ppt

P1 vector for the construction of recombinants by P1 packaging



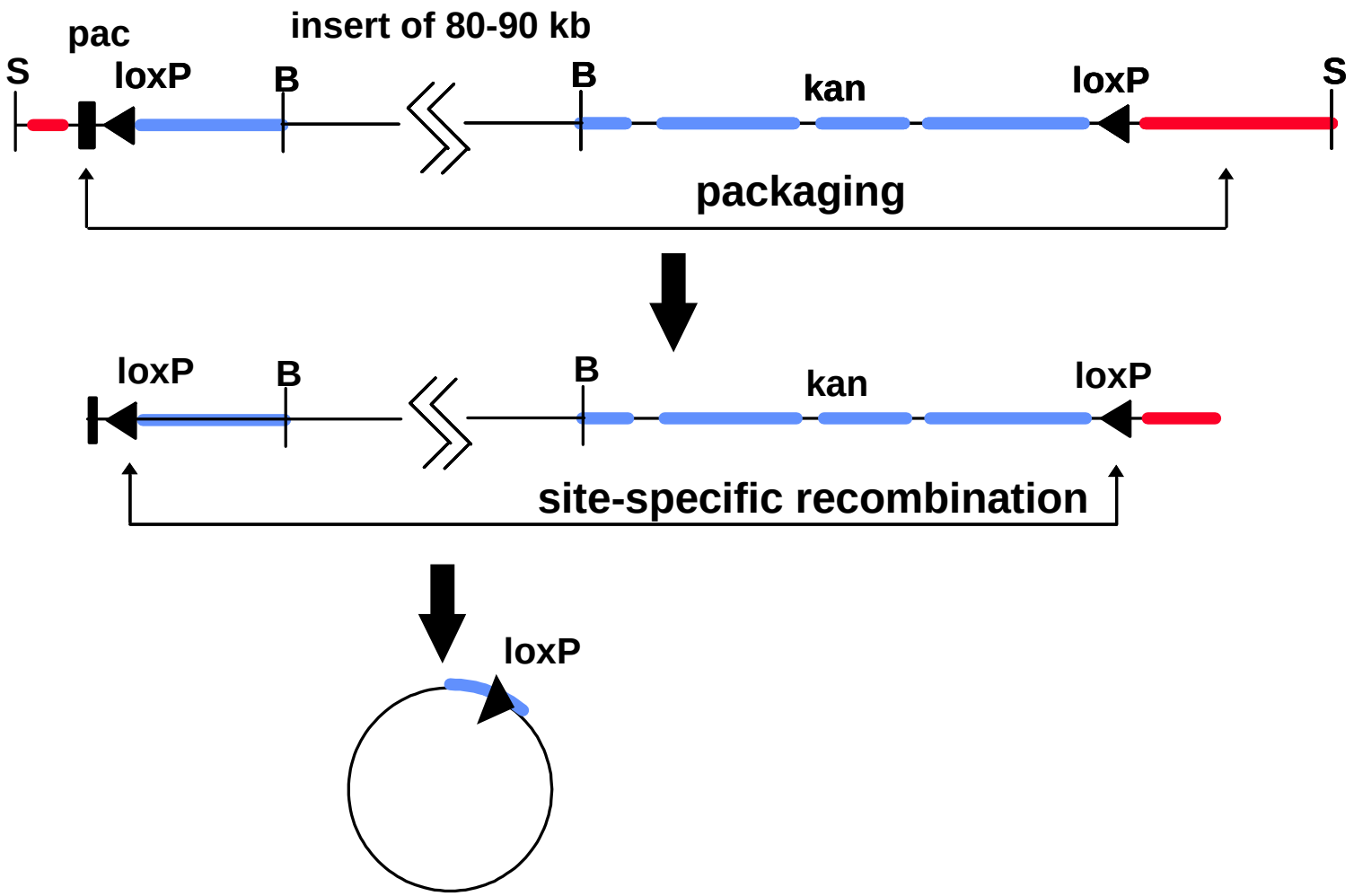
Pierce et al, Proc. Natl. Acad. Sci.U.S.A. (1992) 89, 2056-2060

Sources: http://batzerlab.lsu.edu/Hum_Mol_Gen_Lectures/Hum%20Mol%20Genet%20Lecture%203%20February%201,%202005.ppt



Sources: http://batzerlab.lsu.edu/Hum_Mol_Gen_Lectures/Hum%20Mol%20Genet%20Lecture%203%20February%201,%202005.ppt

Constructing P1 recombinants using packaging extracts



Sources: http://batzerlab.lsu.edu/Hum_Mol_Gen_Lectures/Hum%20Mol%20Genet%20Lecture%203%20February%201,%202005.ppt

P1 Derived Artificial Chromosomes (PACs)

- **Electroporation Based System**
  - **Large insert size**
  - **Low copy number origin for propagation**
  - **High copy origin for DNA production**
  - **Negative selection against non-recombinants**
  - **Very stable inserts**

Vectors for Genomic Libraries

Vector	Insert Size	Remarks
λ Phages	Up to 20-30 kb	• Genome size of λ phages is about 47 kb; • Packaging system is efficient and can handle a total size of 78-105% of the λ genome; • Replacement vector system is usually employed; • Pre-digested arms are commercially available for library constructions; • Useful for study of individual genes

Vectors for Genomic Libraries

Vector	Insert Size	Remarks
Cosmid	35-45 kb	• Plasmid contain the *cos* site of λ phage and hence can use λ phage packaging system; • Propagate in *E. coli* as plasmids; • Useful for subcloning of DNA inserts from YAC, BAC, PAC, etc.
Fosmids	Similar to cosmid	• Contain F plasmid origin of replication and λ *cos* site; • Low copy number and hence more stable

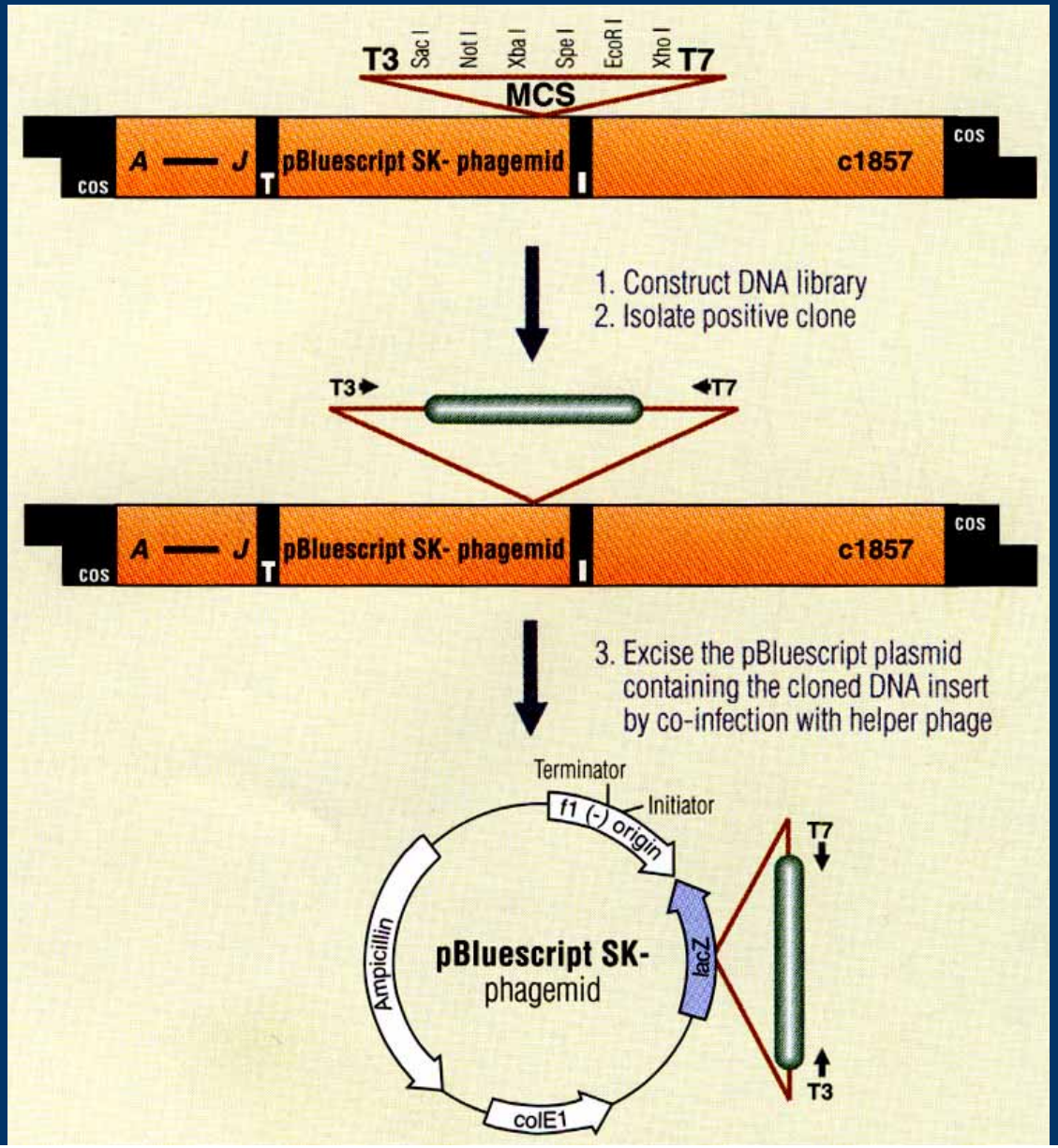
Vectors for cDNA Libraries

Vector	Insert Size	Remarks
λ Phages	Up to 20-30 kb (for replacement vectors) and 10-15 kb (for insertion vectors)	• Maximum size for mRNA is about 8 kb, hence the capacity of DNA insert is not a major concern here; • Insertion vector system is usually employed; • Useful for study of individual genes and their putative functions • Efficient packaging system, easy for gene transfer into *E. coli* cells, more representative than plasmid libraries, subcloning and subsequent DNA manipulation processes are less convenient than plasmid systems
Bacterial plasmids	Up to 10-15 kb	• Relatively easy to transform *E. coli* cells although may not be as efficient as the λ phage system for large scale gene transfer; • Less representative than λ phage libraries, subcloning and subsequent DNA manipulation processes are more convenient than the λ phage systems

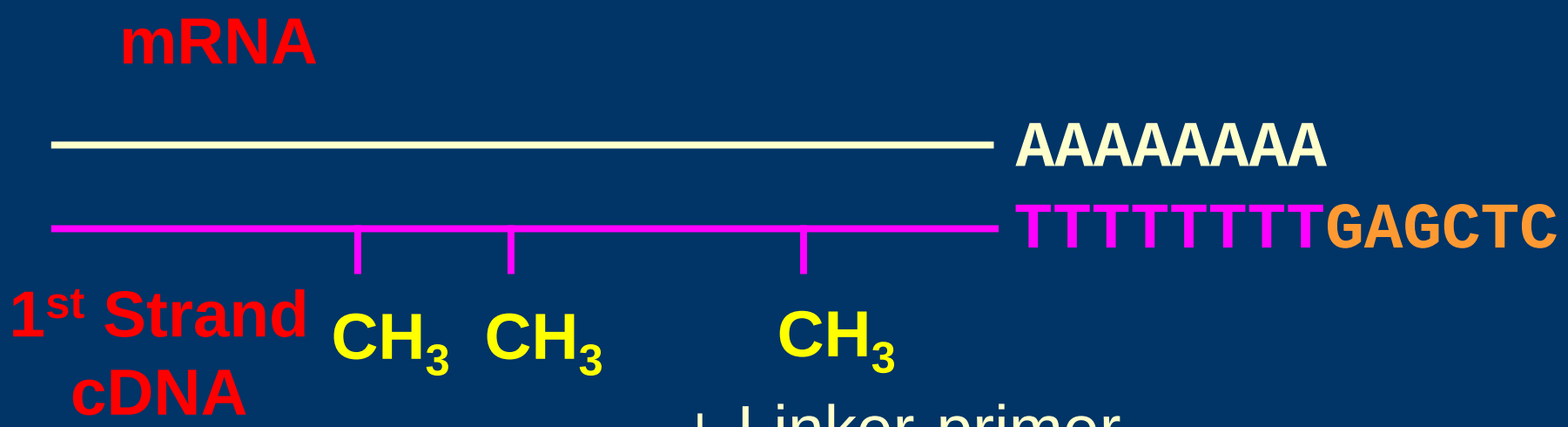
Combining the Advantage of the λ Phage and Plasmid Systems

- Embed a plasmid vector in a λ vector
- *In vitro* package, transfect, propagate, and screen as λ phages
- Excise the plasmid for the λ vector and propagate and manipulate as plasmids subsequently, e.g.
  - Excise by filamentous helper phages (e.g. Strategene)
  - Excise by the Cre-*lox* system (Clontech)

λZAP Library (Stategene)



First Strand Synthesis



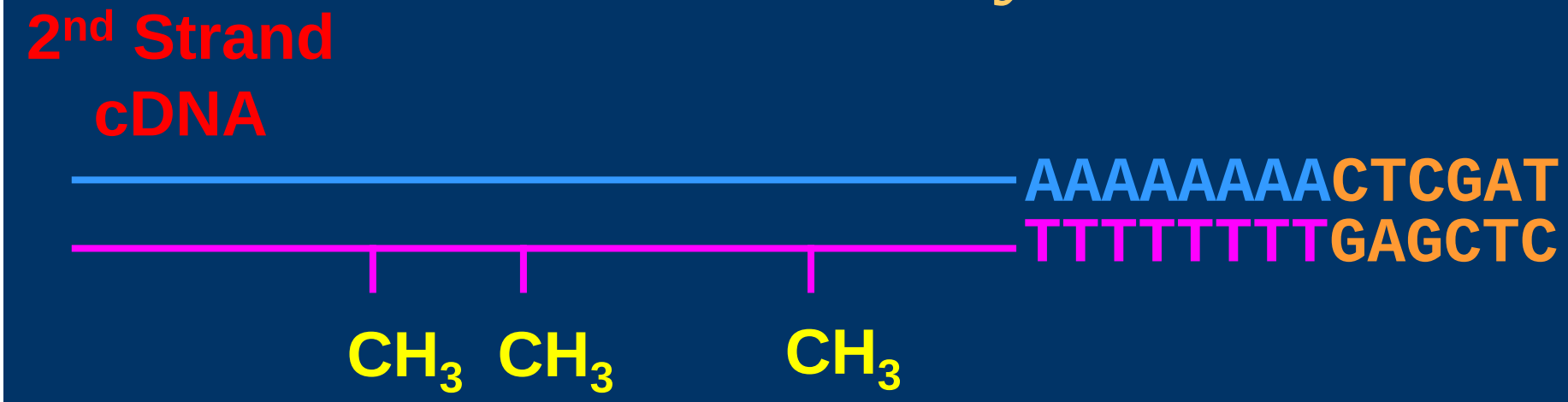
+ Linker-primer

Reverse transcriptase (no RNase activities)

5-methyl dCTP, dATP, dGTP, dTTP

- linker-primer: reverse transcription, restriction site (*Xho*I)
- 5-methyl dCTP: protect internal sites

Second Strand Synthesis

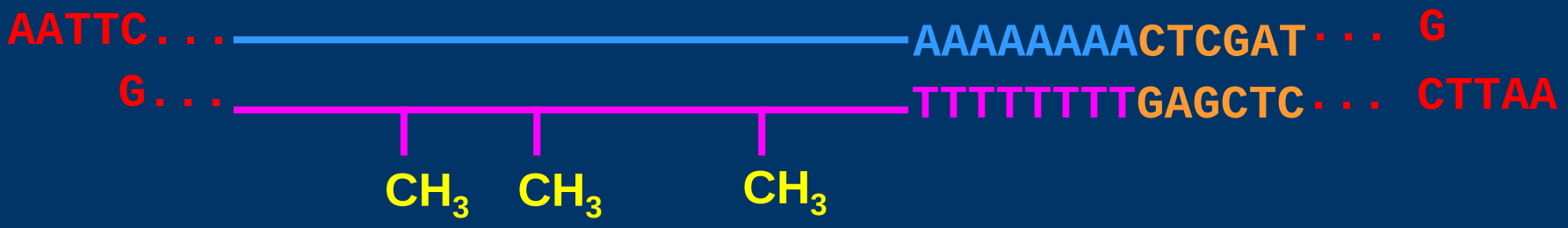


+ RNase H and DNA polymerase I; dNTPs

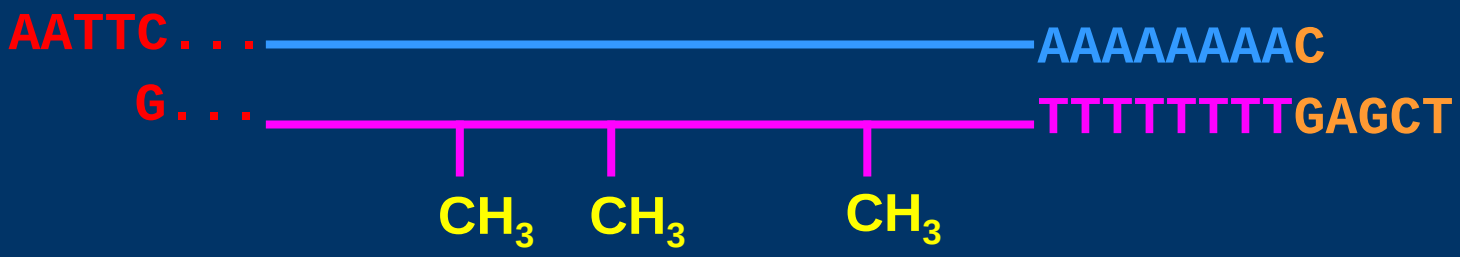
- RNase H: remove mRNA in DNA-RNA hybrid
- DNA Polymerase I: synthesis 2nd strand
- Both enzymes added simultaneously
- When RNase H starts to degrade the mRNA, residual RNA fragments may act as primers for initiation of DNA synthesis
- DNA Polymerase 1 fills the gaps by nick translation

Addition of Adapter

+ *Eco*RI adapter; ligase



+ *Xho*I

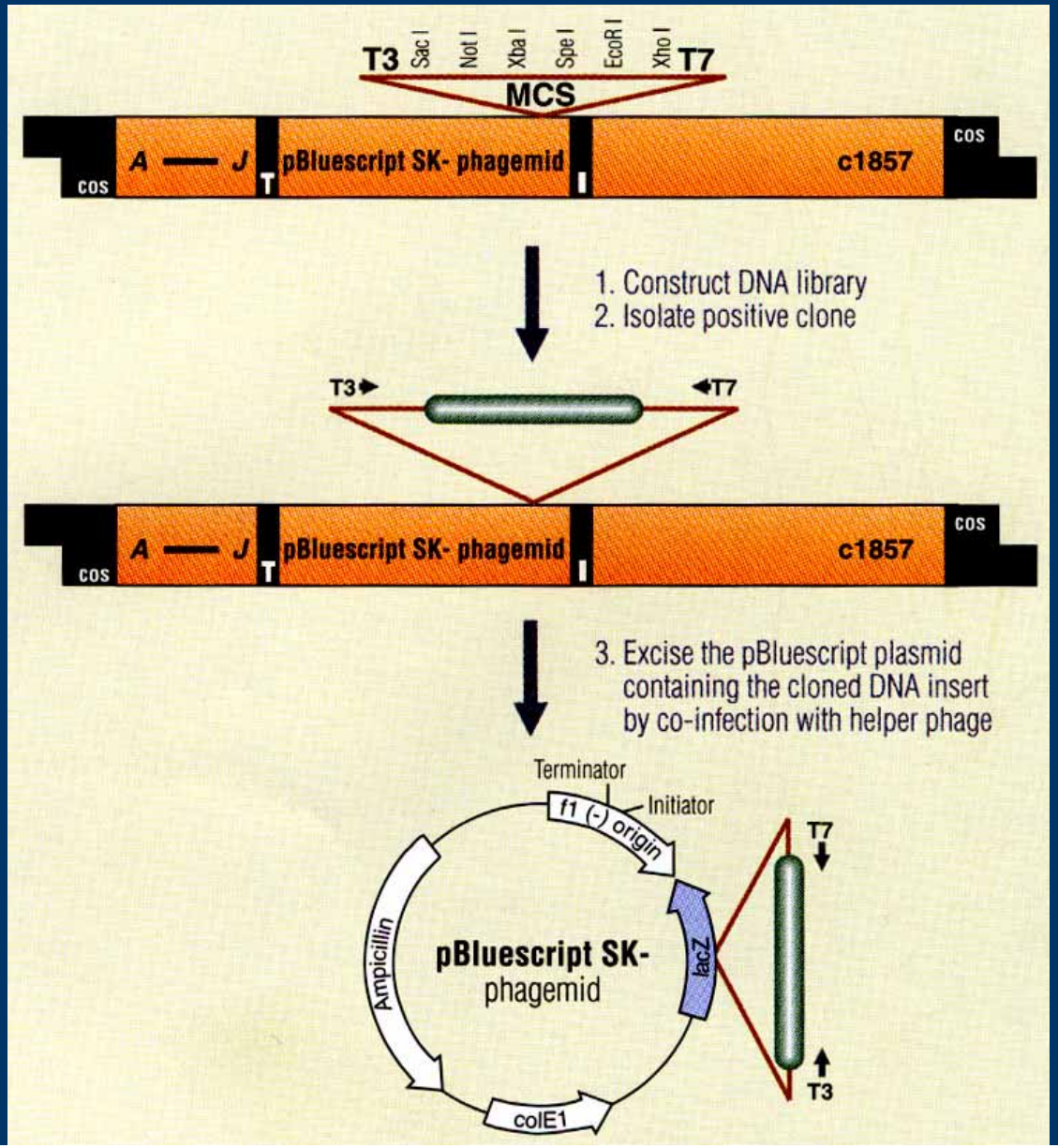


- *EcoRI* Adapter and *Xho*I linker: directional cloning

Primary Library Synthesis

- Size fractionation
- Ligation to λ arms
- *In vitro* package (packaging extract should be McrA$^-$, McrB$^-$, and Mrr$^-$ to prevent digestion of hemimethylated DNA)
- Host for λ phage infection, e.g. XL1-Blue MRF':
  - Restriction deficient Δ(*mcrA*)183, Δ(*mcrBC*-hsdSMR-mrr)173; *supE* to allow propagation of helper filamentous phages (with amber mutation); *recA*$^-$ to prevent recombination
  - F' plasmid:
    - ΔM15*lacZ* for blue-white screening
    - lacI^q to prevent uncontrolled expression of fusion protein
    - F pili to allow infection of helper filamentous phages

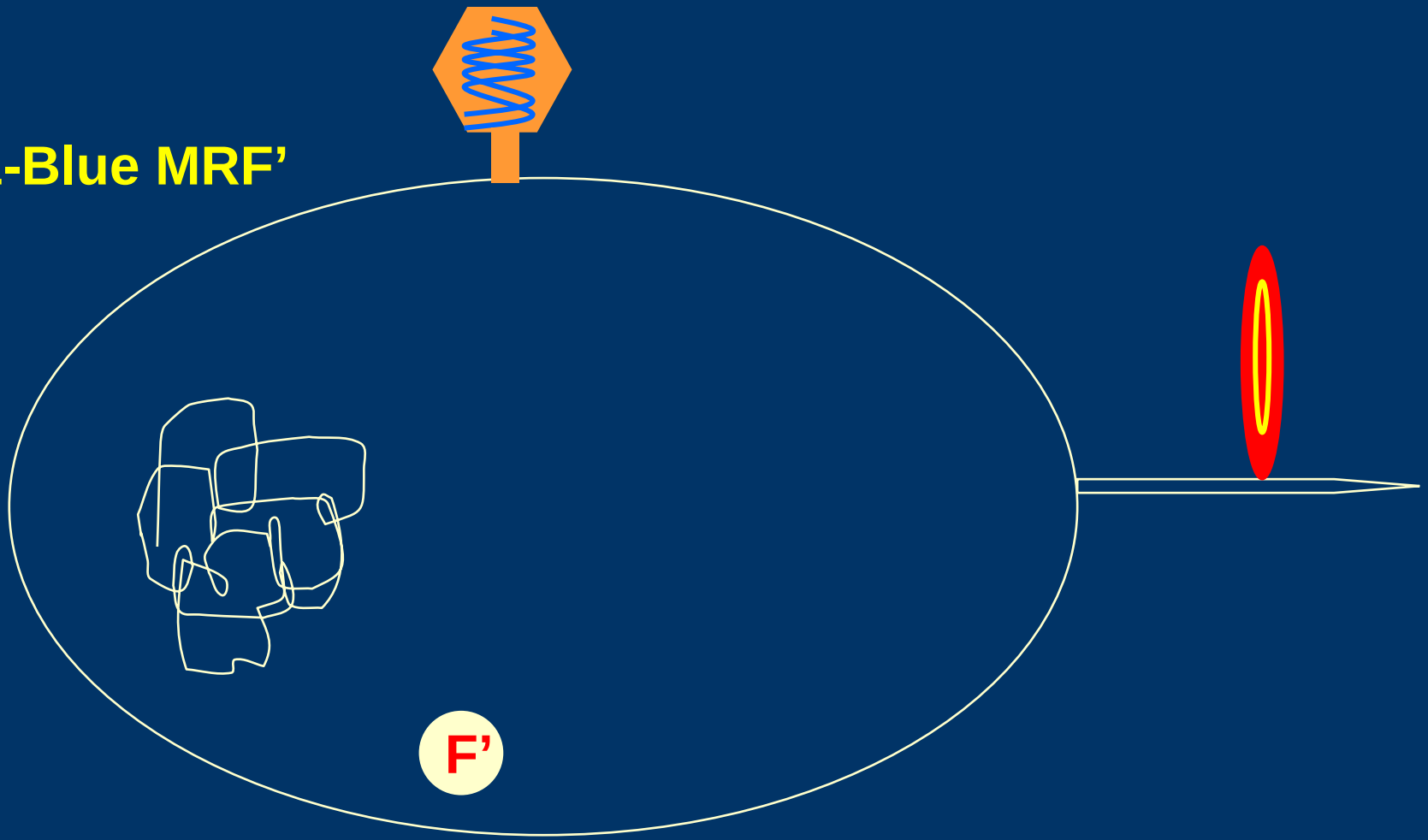
λZAP Library (Stategene)



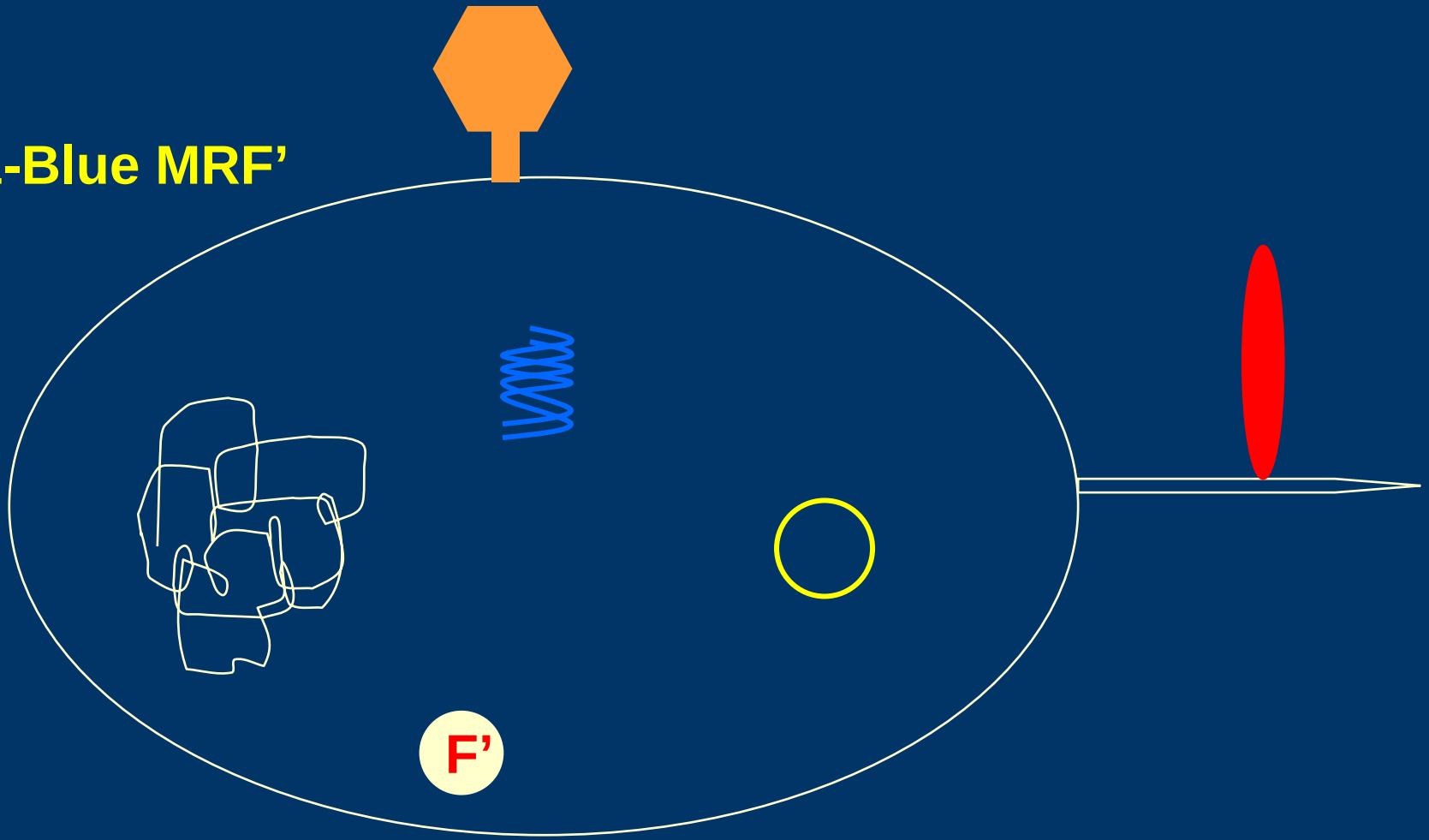
Excision of pBluescript Plasmid

- First host XL1-Blue MRF': F'plasmid provides F pili for f1 filamentous phage attachment; it is a suppressing strain so that f1 helper phages which carry an amber mutation can pack DNA
- The λ ZAP: contains initiation and termination signal sequences for f1 packing flanking pBluescript sequence
- Second host: SOLR, a non-suppressing strain to stop helper phage propagation; a λ phage resistant strain to prevent λ phage propagation

XL1-Blue MRF’

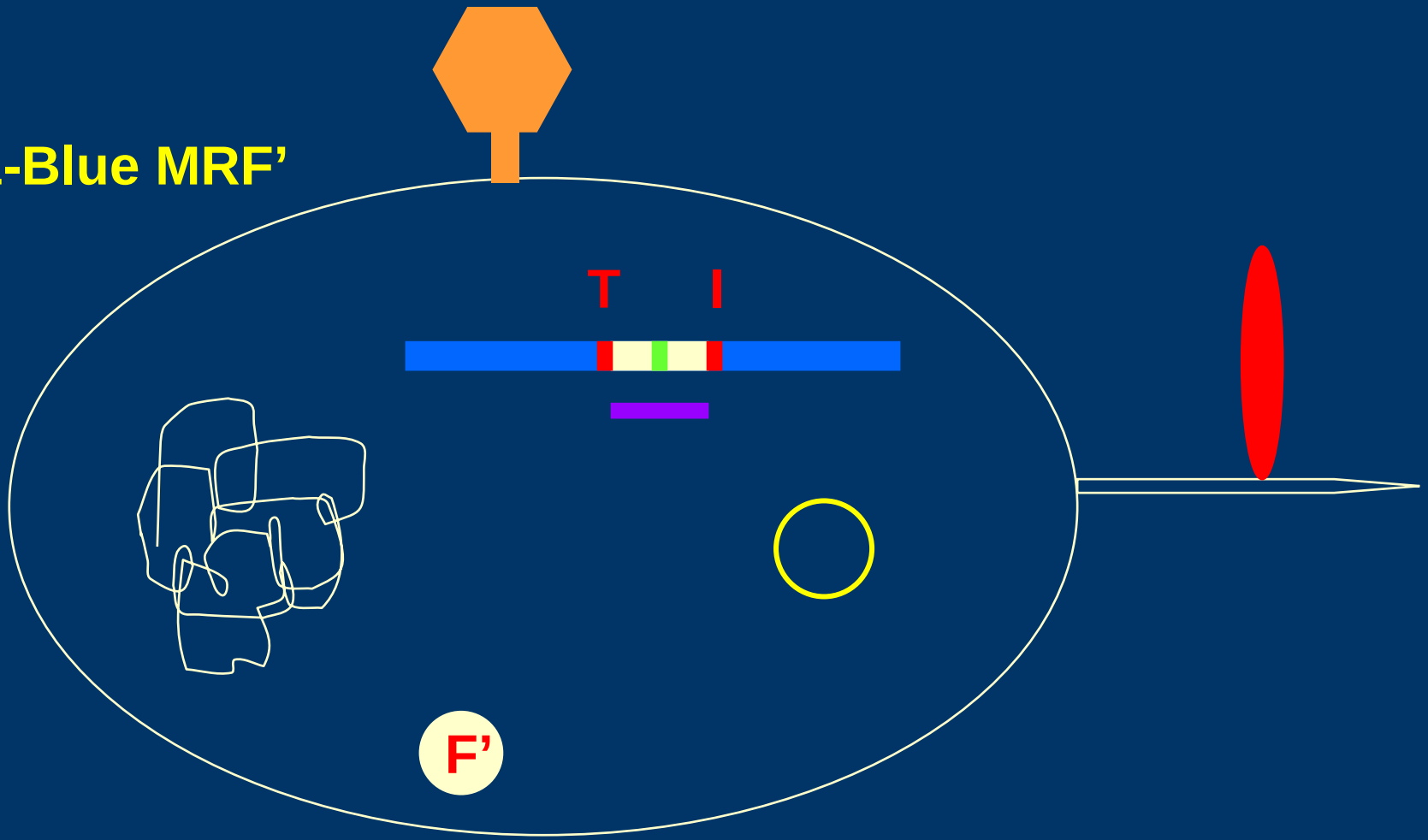


XL1-Blue MRF'



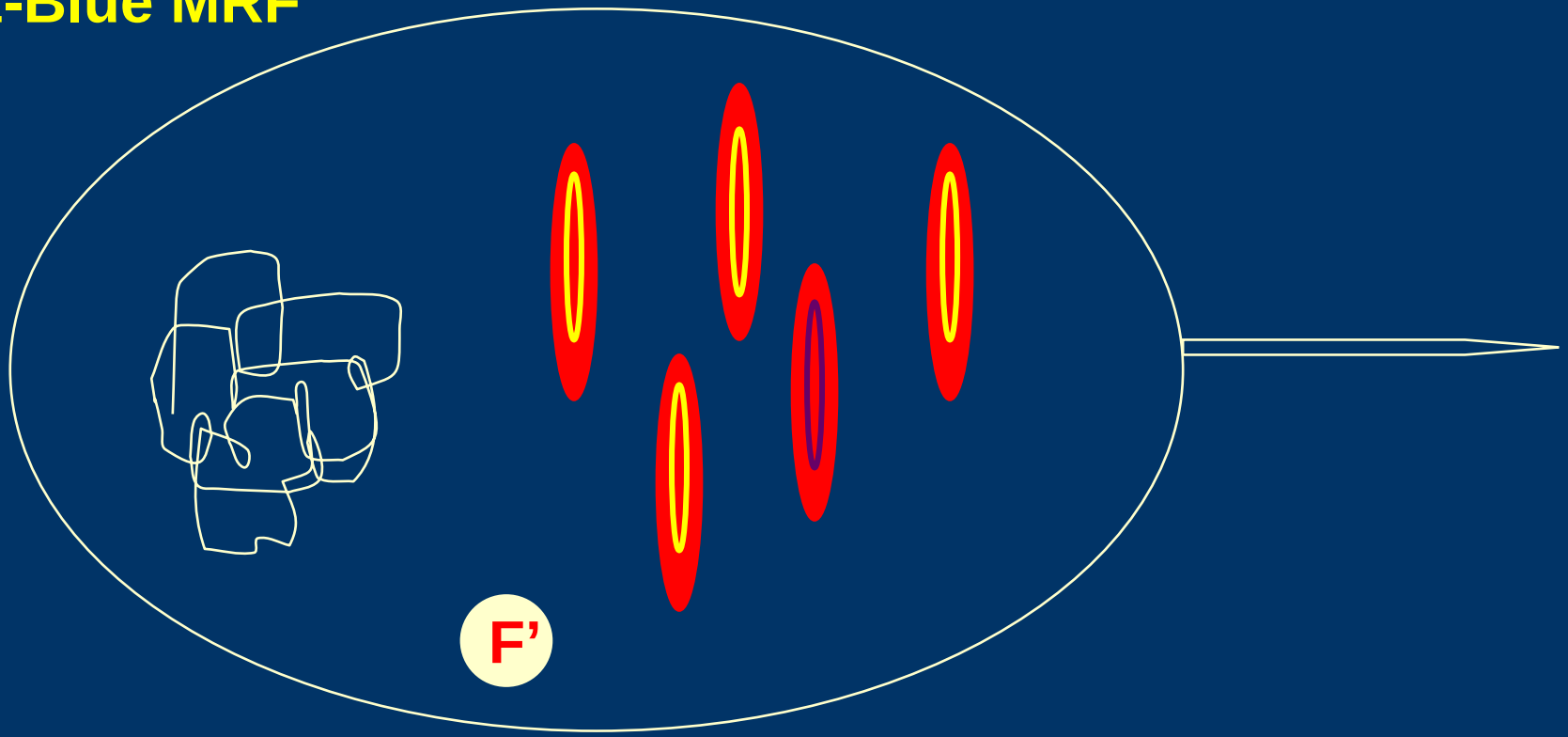
- Co-transfection of λZAP and fl help phage

XL1-Blue MRF'



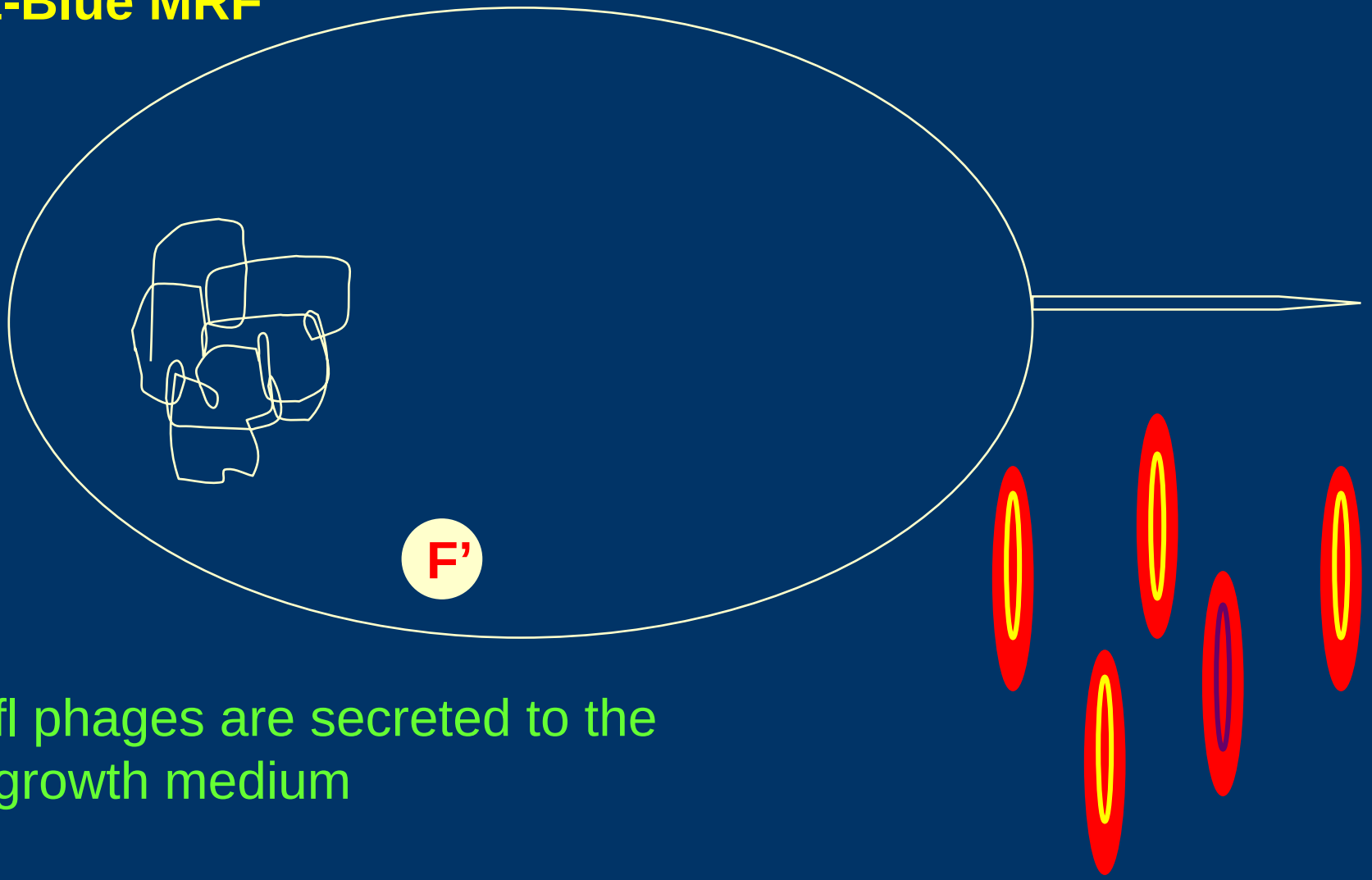
- The λZAP phage clone replicates in the host cell
- The fl helper phage protein nicks the fl replication initiation site I on λ ZAP and replication continues until reaching the fl replication termination site T

XL1-Blue MRF'



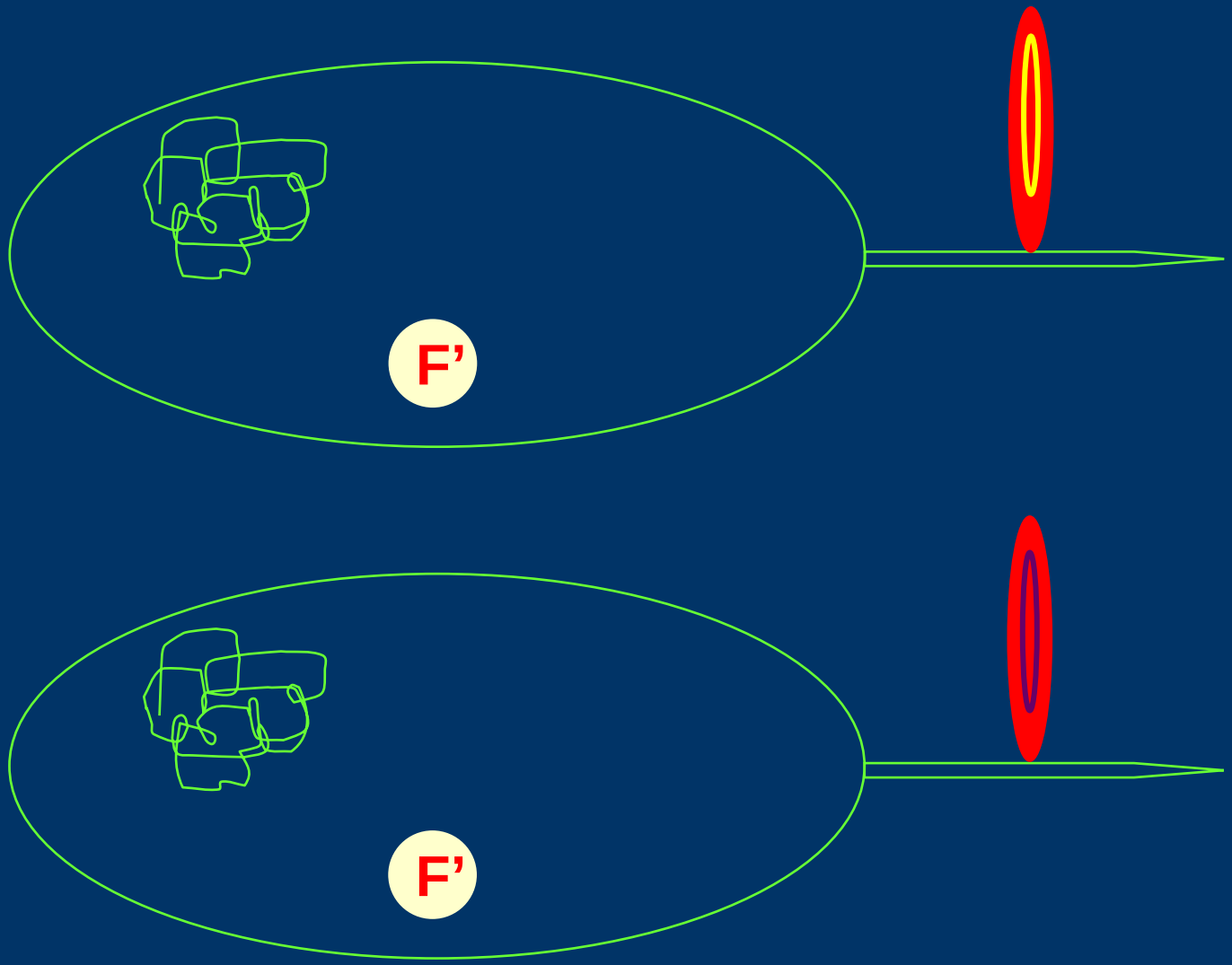
- fl phage coats indiscriminately pack DNA molecules containing the fl replication origin; both fl genome and pBluescript fragments are packed

XL1-Blue MRF'



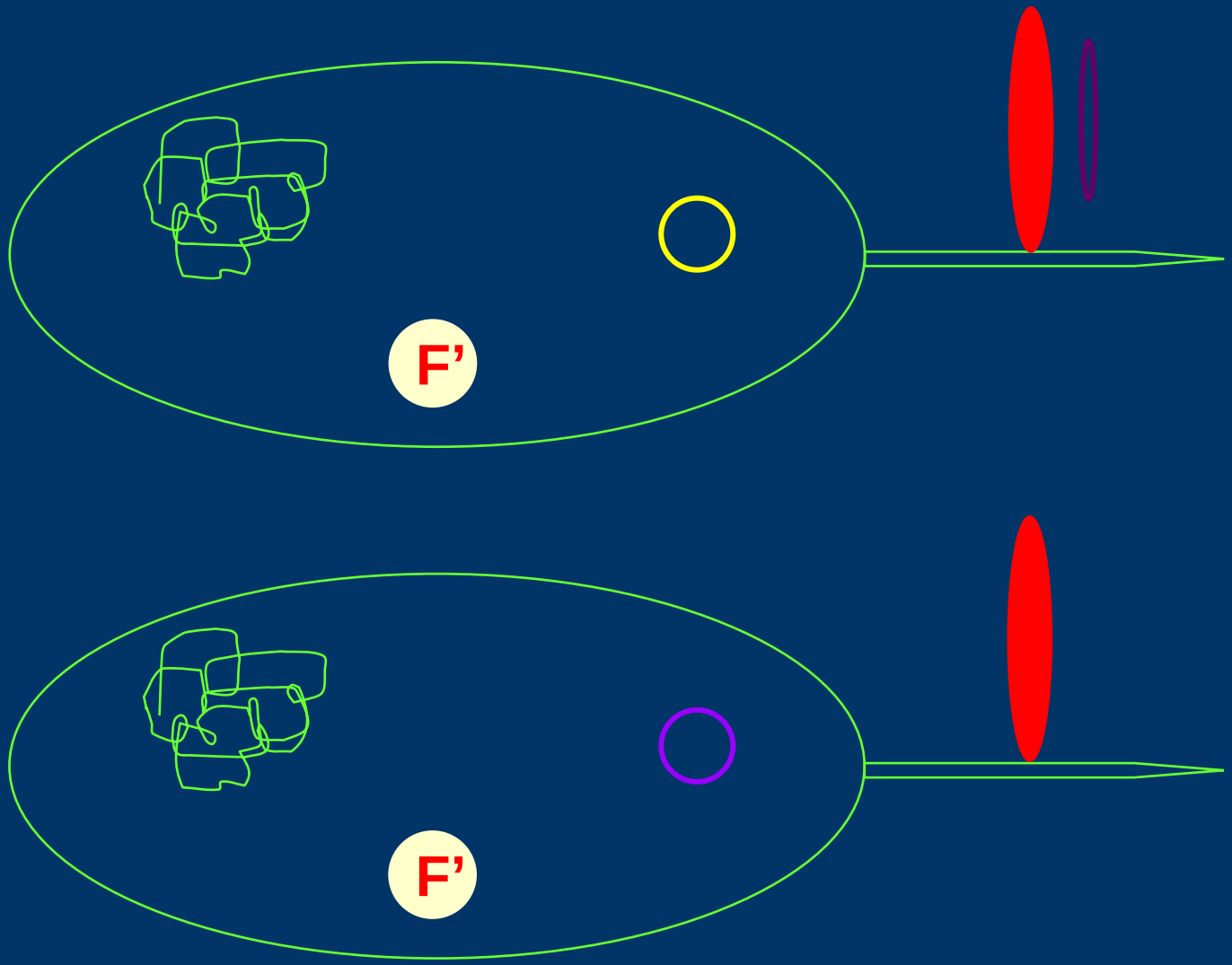
- fl phages are secreted to the growth medium

SOLR



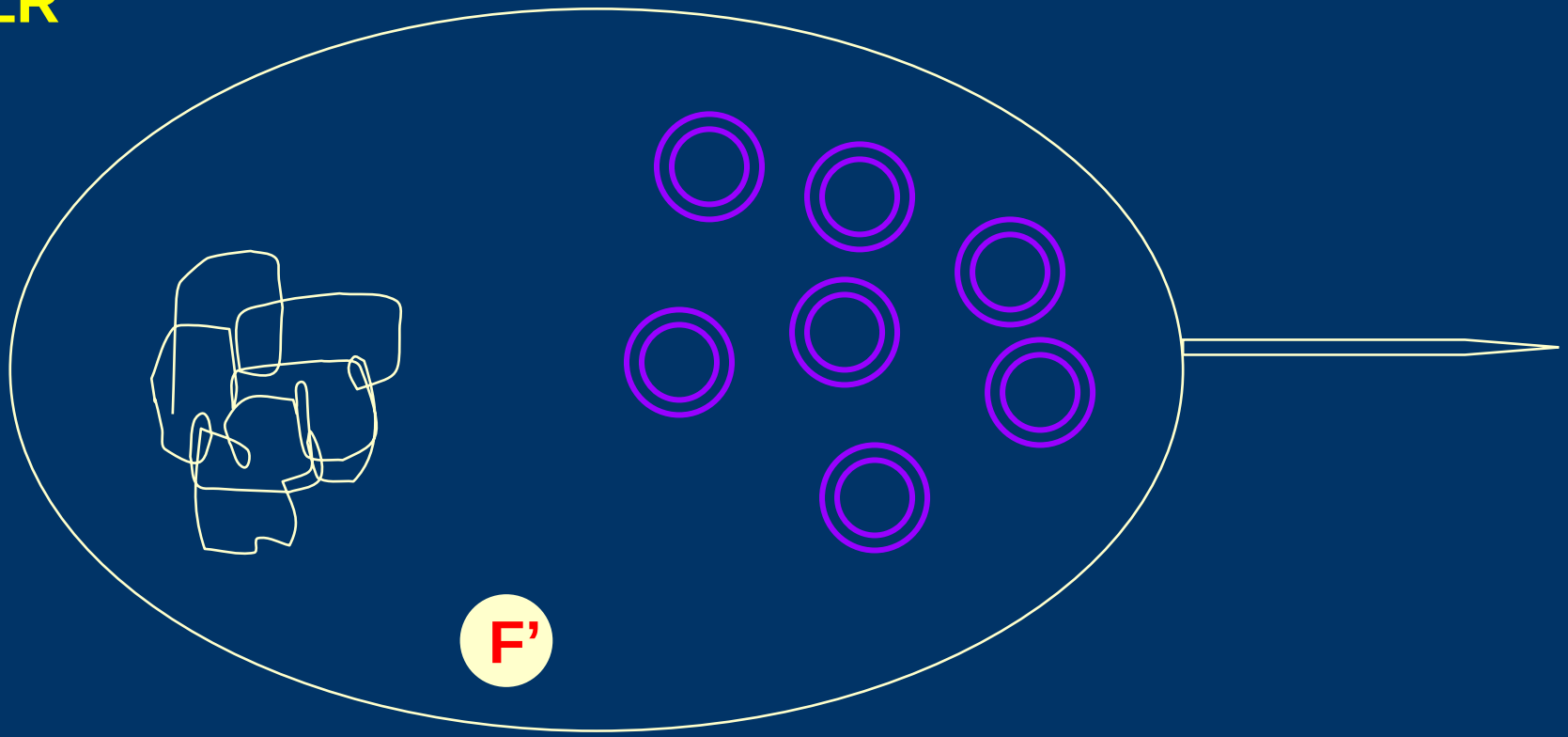
- fl phages infect a new host SOLR (λ^R, suppressor free)

SOLR



- fl phages infect a new host SOLR (λ^R, suppressor free)

SOLR



- fl genome (contains an amber mutation) cannot propagate in the suppressor free SOLR and pBluescript containing cells can be selected by the ampicillin resistance marker

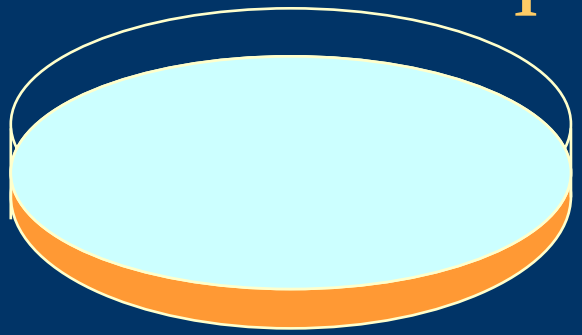
Screening for the Right Clones

- By PCR or hybridization if DNA sequence information of the target gene or homologous genes, or the amino acid sequence of the target gene product is available
- By hybridization if DNA fragments of the target gene or homologous genes are available
- By functional assays, e.g.
  - Yeast functional complementation
  - Microinjection

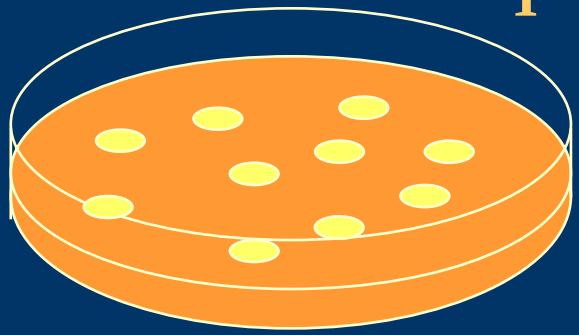
Screening by Hybridization

- Plaque/colony lifting
- Making labeled probes
- Hybridization
  - Pre-hybridization
  - Hybridization
  - Washing
- Detection
  - Radioactivity
  - Luminescence
  - Florescence
  - Chemiluminescence
- Rescuing clones and further analysis

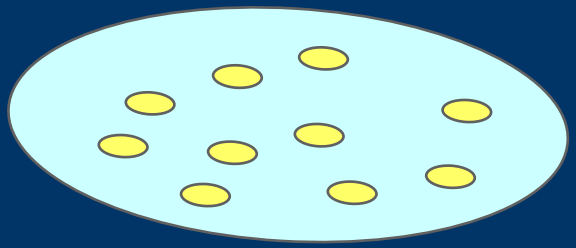
Plaque/Colony Lifting



Plaque/Colony Lifting



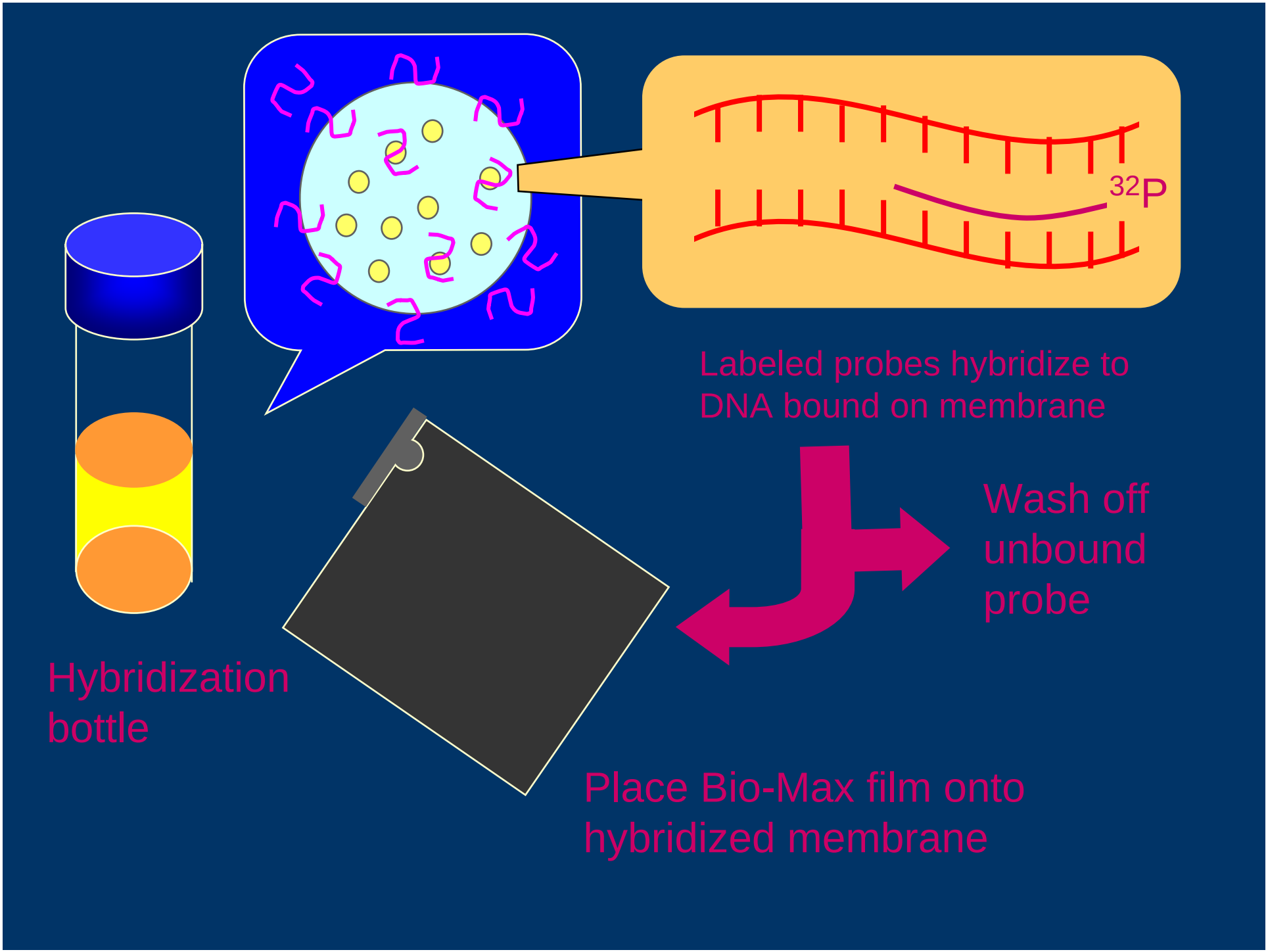
Master plate

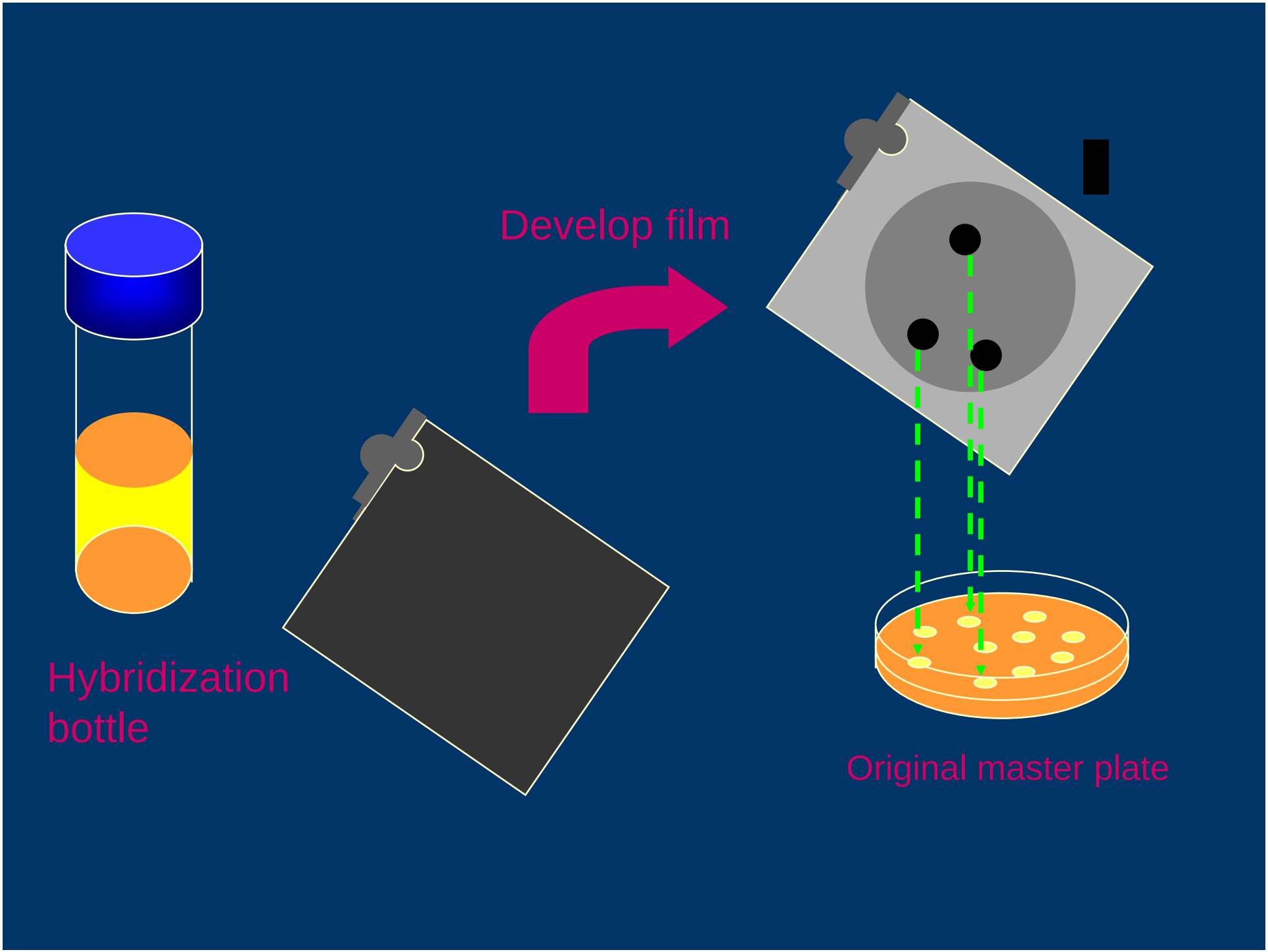


Replica membrane

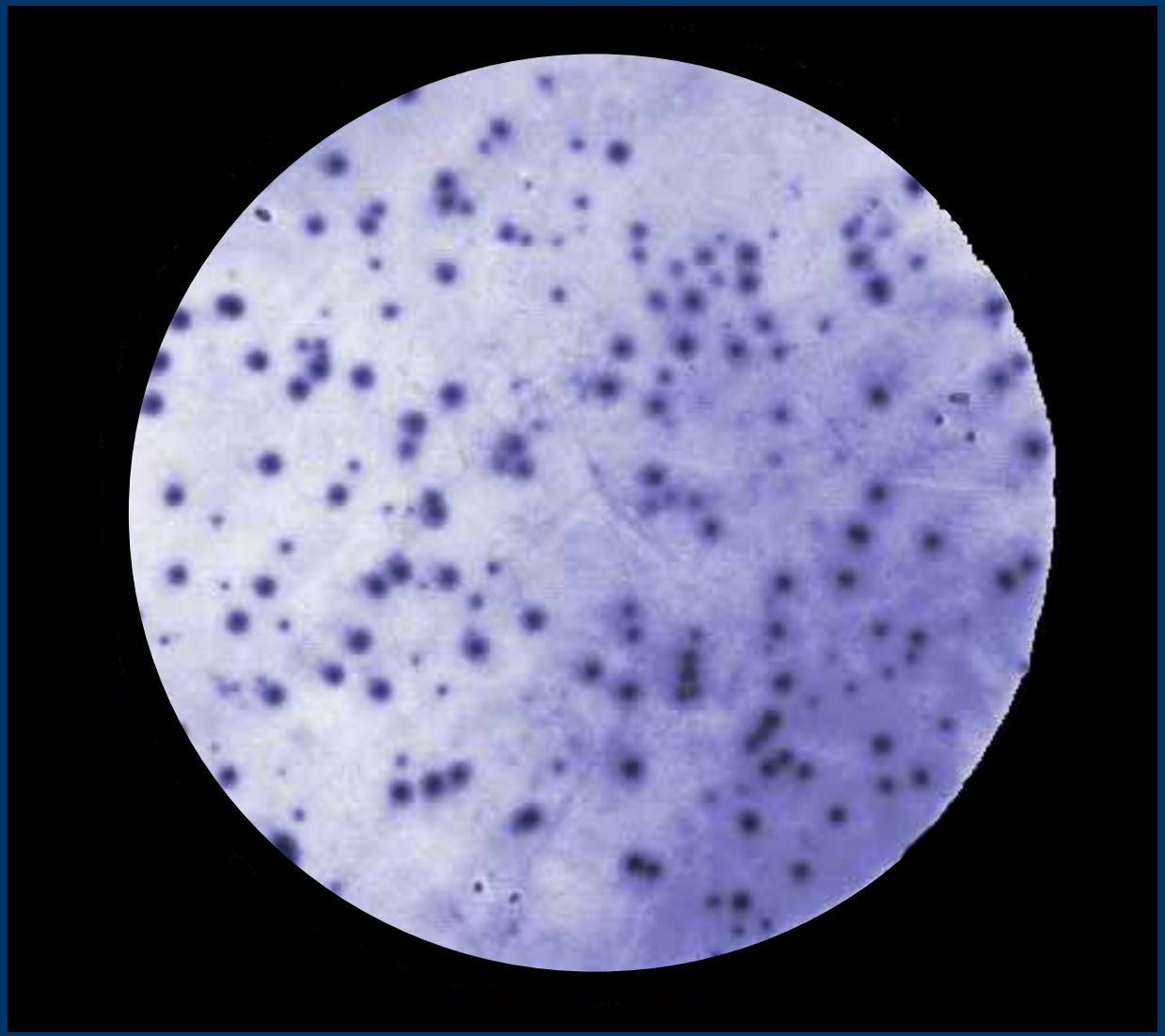
Labeled nucleic acid probes (radioactive or non-radioactive)

Hybridization


^{32}P
Labeled probes hybridize to DNA bound on membrane
Wash off unbound probe
Hybridization bottle
Place Bio-Max film onto hybridized membrane


Develop film
Hybridization bottle
Original master plate

The Result of 2nd Round Screening

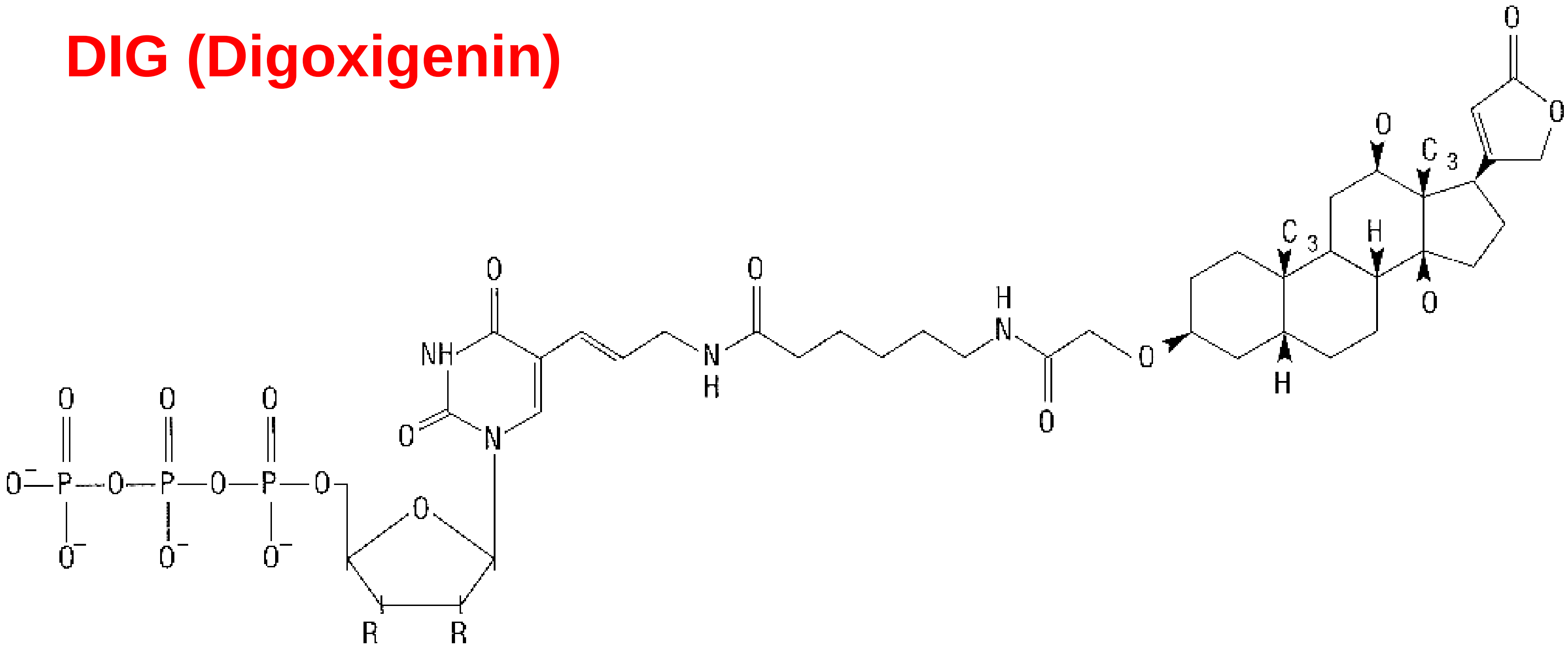


Making Probes

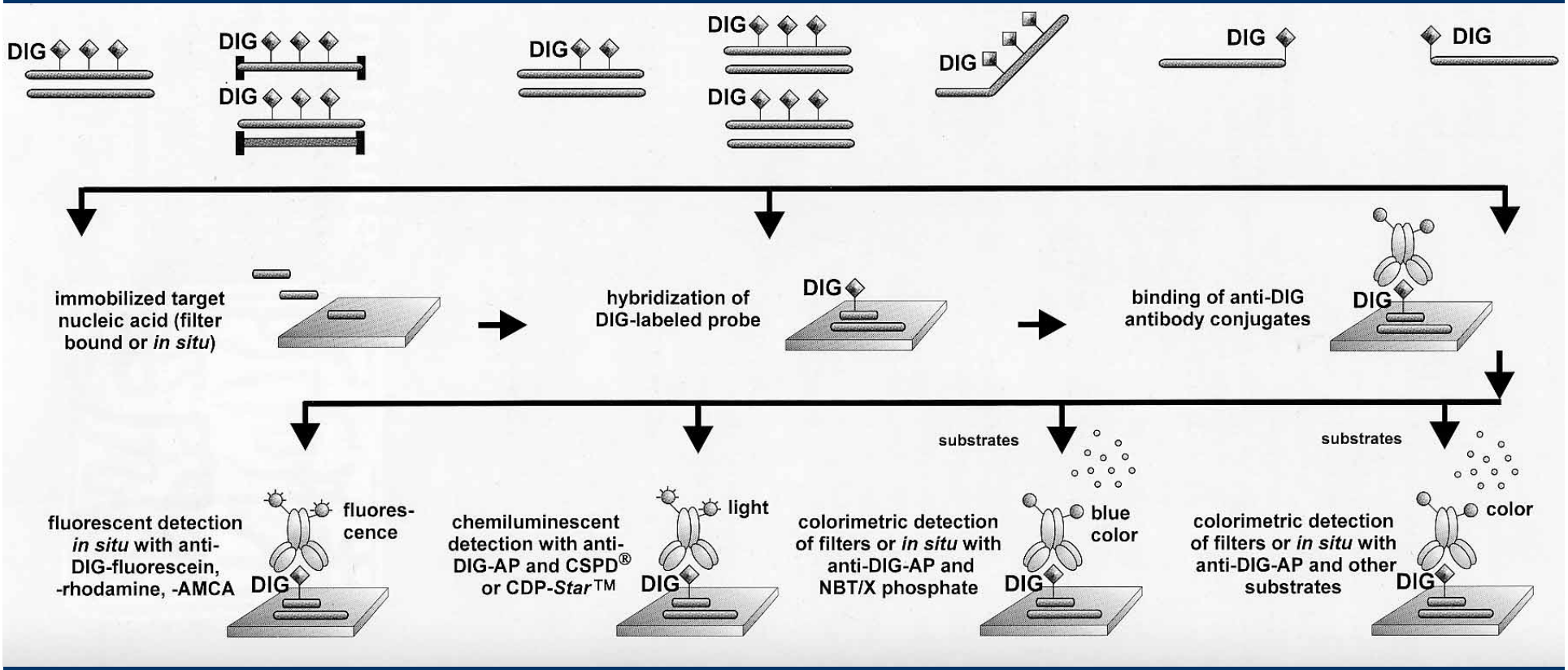
- Methods of Detection
  - Radioactive (e.g. P^{32}, P^{33}, S^{35})
  - Non-radioactive (e.g. biotinylation, horseradish peroxidase, Digoxigenin)

Example of Non-Radioactive Probes (from Boehringer Mannheim)

DIG (Digoxigenin)



Example of Non-Radioactive Probes (from Boehringer Mannheim)



Making Probes

- Methods of Synthesis
  - DNA
    - Nick translation *(E. coli* polymerase I)
    - Random priming (Klenow)
    - Random labeling by PCR
    - 5'-end labeling (DNA kinase)
    - 3'-end labeling (T4 DNA polymerase, terminal transferase, modified T7 DNA polymerase, Klenow, reverse transcriptase)

Nick Translation

Nick Translation

- Nicks created by DNaseI

Nick Translation

- Gaps created by *E. coli* DNA polymerase I (5’ to 3’ exonuclease activities)

Nick Translation

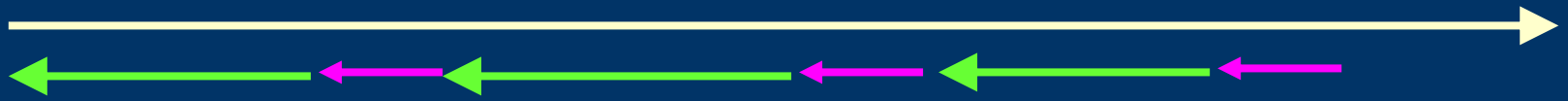
- Gaps filled by *E. coli* DNA polymerase I in the presence of dNTPs and labeled deoxynucleotides (5’ to 3’ polymerase activities, proofreading by 3’ to 5’ exonuclease activities)

Nick Translation

- Reaction repeats to give uniform labeling

Random Priming

Random Priming



- Heat denaturing
- + Random primers (hexamers);

  + Klenow, dNTPs, and labeled deoxynucleotides

Making Probes

- Methods of Synthesis
  - RNA
    - Random labeling by *in vitro* transcription (DNA dependent RNA polymerase, e.g. Sp6, T3, T7 RNA polymerase)
    - 3’-end labeling (DNA independent RNA polymerase + ATP by polyA tail; RNA ligase)

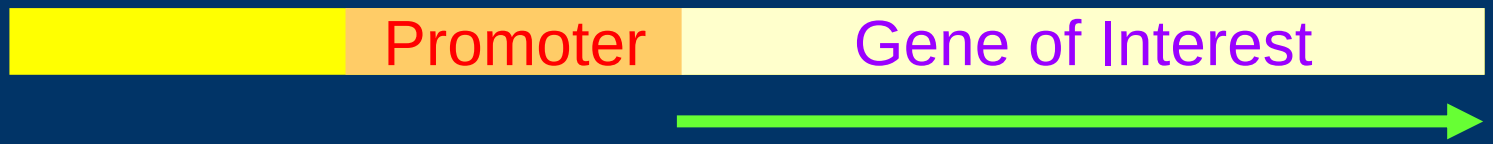
Making Riboprobes

- Clone gene of interest under the control of a strong, specific, and inducible promoter (e.g. Sp6, T3, T7)

Making Riboprobes

- Cut with a restriction enzyme to generate a blunt end or 5’ overhang at the 3’ end of the gene of interest

Making Riboprobes



- + DNA-dependent RNA polymerase corresponding to the promoter used, in the presence of NTPs and labeled nucleotides, to generate randomly labeled cRNA probes

Hybridization and Washing

- Prehybridization
  - for reduction of noise to signal ratio
- Hybridization
  - usually carried out in solutions of high ionic strength to maximize annealing rate
  - 20-25°C below Tm
  - smaller volume the better: fast reassociation and less probes
  - to minimize background, hybridize for the shortest time with minimum probes
- Washing
  - to remove non-specific binding
  - can control the stringency by altering temperature or $[Na^+]$

Conditions for Hybridization and Washing

- For cloning of a specific gene, use high stringency conditions for hybridization (high temperature, with formamide) and washing (high temperature, low NaCl)
- For cloning of homologous genes (or members of a gene family), use low stringency conditions for hybridization (low temperature, without formamide) and washing (low temperature, high NaCl)

Factors Affecting Hybrid Stability

Ionic Strength	Tm increases 16.6^{o}C each 10-fold increase in monovalent cations (Na^{+}) between 0.01 to 0.40 M NaCl
Base Composition	AT base pairs are less stable than GC base pairs in aqueous solutions containing NaCl
Destabilizing Agents	Each 1% of formamide reduces the Tm by about 0.6^{o}C for a DNA-DNA hybrid
Mismatched base pairs	Tm is reduced by 1^{o}C for each 1% of mismatching
Duplex length	Negligible effect with probes >500 bp

Tm: Transition Mid-Point; Melting Temperature

- Temperature at which 50% of the nucleotides of DNA-DNA, DNA-RNA or RNA-RNA duplex are hybridized.

For Hybrids > 100 Nucleotides

- DNA-DNA
  - $Tm=81.5^{o}C + 16.6\log_{10}[Na^{+}] + 0.41(\%G+C) - 0.63(\%\ formamide) - (600/l)$
- DNA-RNA
  - $Tm=79.8^{o}C + 18.5\log_{10}[Na^{+}] + 0.58(\%G+C) + 11.8(\%G+C)^{2} - 0.5(\%\ formamide) - (820/l)$
- RNA-RNA
  - $Tm=79.8^{o}C + 18.5\log_{10}[Na^{+}] + 0.58(\%G+C) + 11.8(\%G+C)^{2} - 0.35(\%\ formamide) - (820/l)$

For Hybrids > 100 Nucleotides

- l = the length of the hybrid in base pairs
- [Na^+] between 0.01 M to 0.4 M
- %G+C between 30% to 75%
- Stability: RNA-RNA > RNA-DNA > DNA-DNA
- Tm of a double-stranded DNA decreases by 1-1.5^oC with every 1% decrease in homology
- pH between 5-9; when outside this range, stability of DNA decreases drastically due to protonation or deprotonation of the bases

For Oligos 14-70 nucleotides

- $Tm = 81.5^{o}C + 16.6(\log_{10}[Na^{+}]) + 0.41(\%G+C) - (600/N)$
- N=number of nucleotides

For Oligos < 18 nucleotides

- Tm = (#A+T x 2) + (#G+C x 4)

Factors Affecting Hybridization Rate

Ionic Strength	Optimal rate at 1.5 M Na^+
Base Composition	Little effect
Destabilizing Agents	50% formamide: no effect; other concentrations: reduced
Mismatched base pairs	Each 10% of mismatching reduces rate by a factor of two
Duplex length	Directly proportional to duplex length
Temperature	Maximum rate: 20-25^{o}C below Tm for DNA-DNA hybrids, 10-15^{o}C below Tm for DNA-RNA hybrids
Viscosity	Increase rate of membrane hybridization; 10% dextran sulfate increases rate by factor of ten
Probe complexity	Repetitive sequences increase the rate
pH	Little effect between pH 5.0 to 9.0

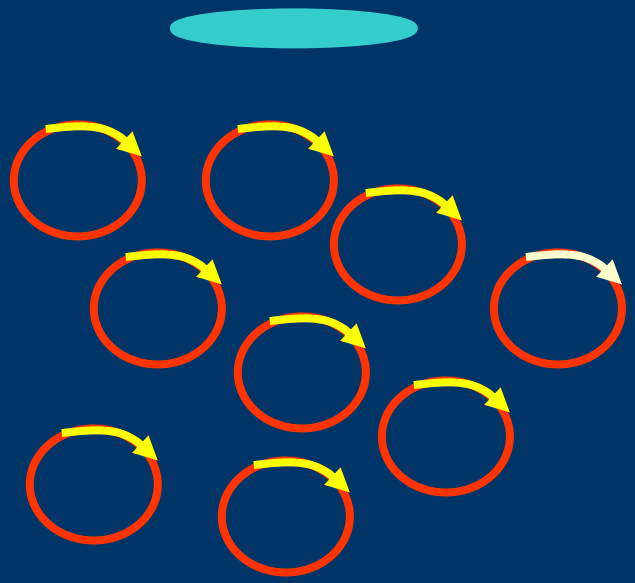
Screening for the Right Clones

- By PCR or hybridization if DNA sequence information of the target gene or homologous genes, or the amino acid sequence of the target gene product is available
- By hybridization if DNA fragments of the target gene or homologous genes are available
- By functional assays, e.g.
  - Yeast functional complementation
  - Microinjection

Screening by Yeast Functional Complementation

- *Saccharomyces cerevisiae* is a simple eukaryotic system
- Relatively easy to grow and perform genetic manipulation
- Mutants, expression vector, transformation protocol, etc. are available

Cloning by Yeast Functional Complementation



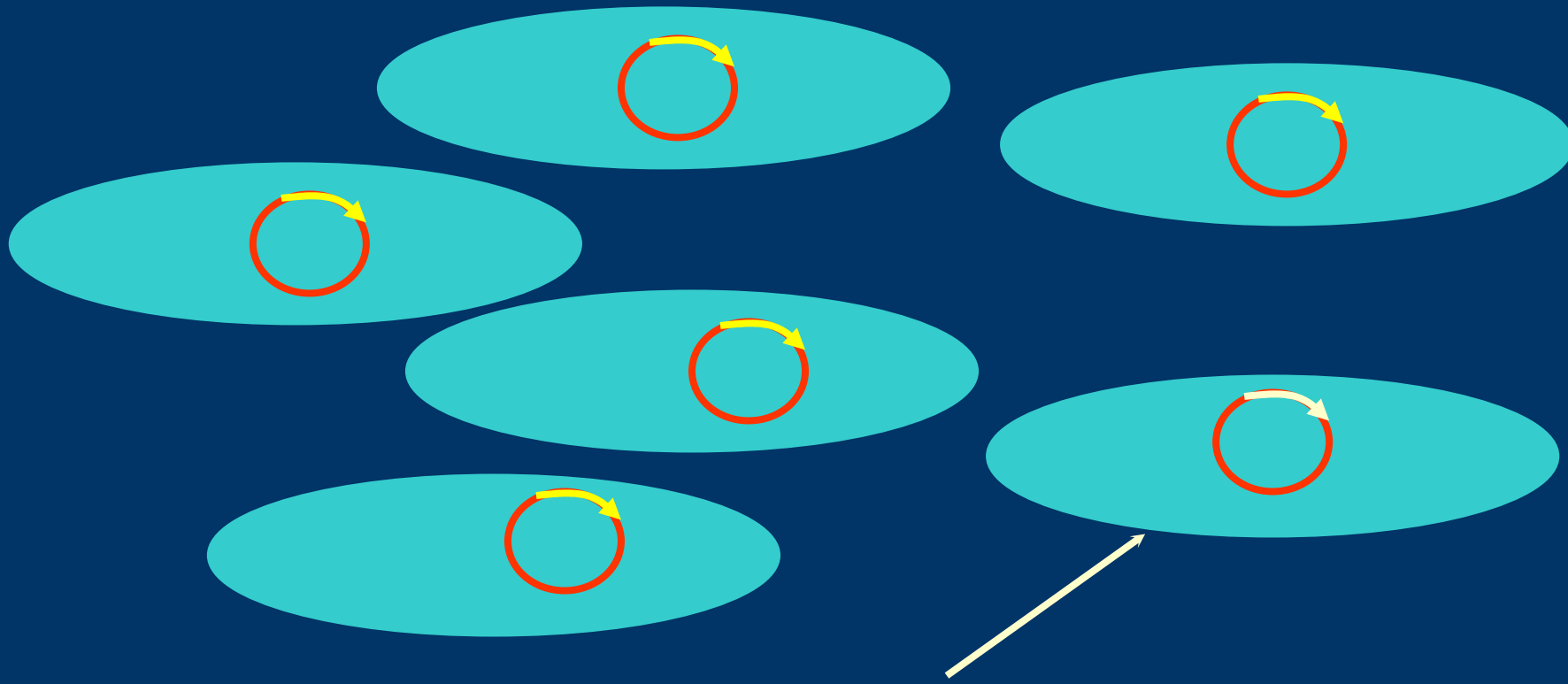
A yeast mutant which is defective in a function of your interest

A library containing cDNAs of the organism of interest is inserted in a yeast vector;

cDNAs express under an inducible yeast expression promoter

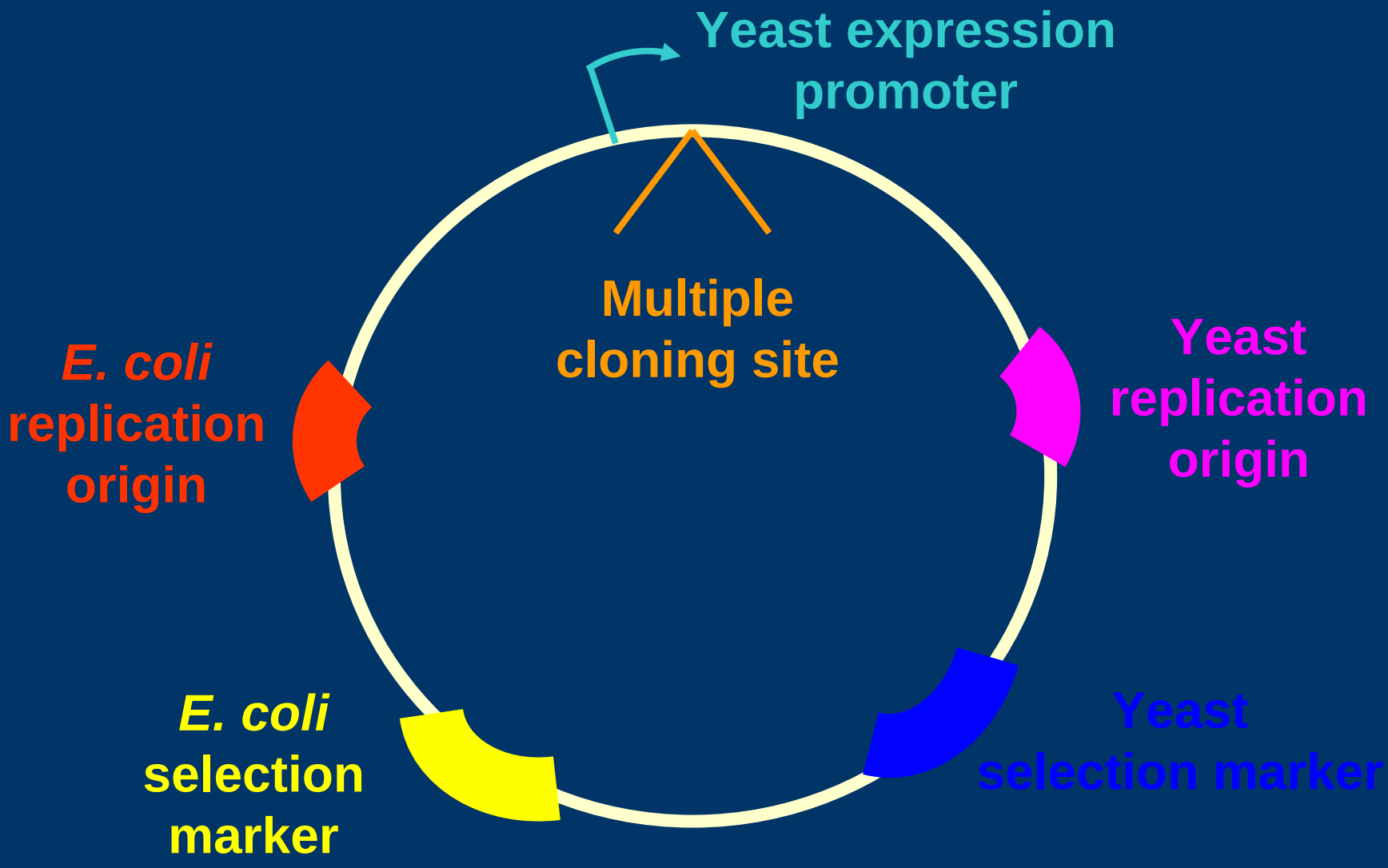
Yeast expression promoter

Cloning by Yeast Functional Complementation

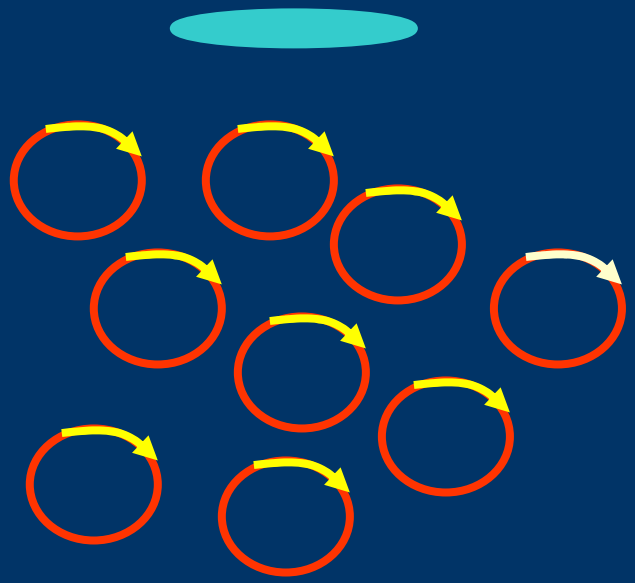


Yeast mutant is “complemented”
when the transgene expression is induced

Designing a Suitable Vector



A Case Study: Cloning of Potassium Channels



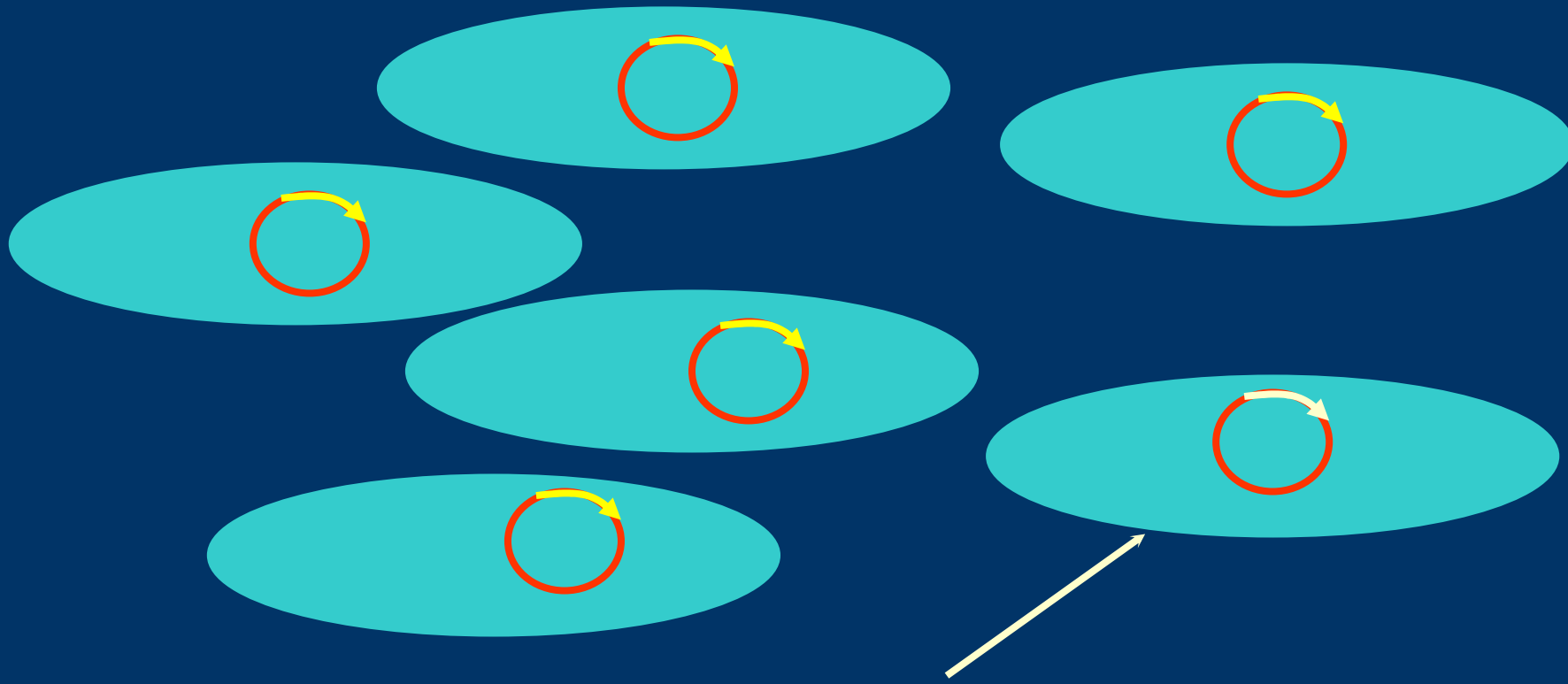
A yeast mutant which cannot survive on K^+-limiting medium

A plant cDNA library constructed in a yeast vector;

Yeast expression promoter

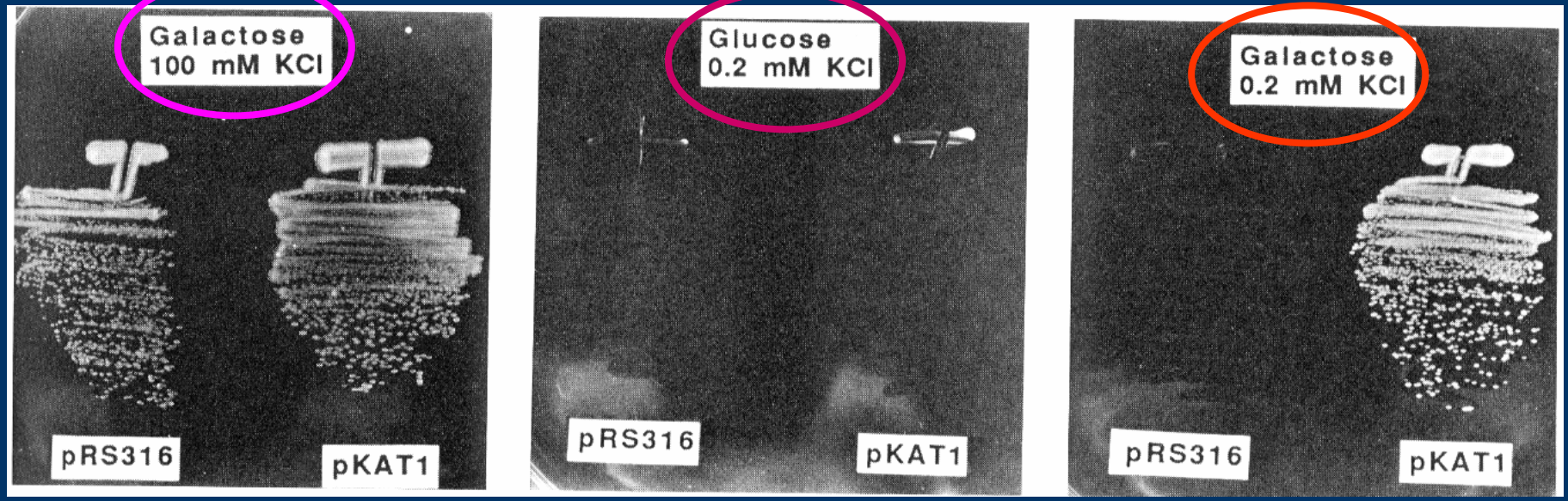
Plant cDNAs express under an inducible yeast expression promoter

A Case Study: Cloning of Potassium Channels



Survives in K^+-limiting medium
when the transgene expression is induced

A Case Study: Cloning of Potassium Channels



Mutant **Transformant** **Transgene not induced**

Only the transformant survives on K^+-limiting medium

Both survive on K^+-rich medium

Anderson et al., 1992