Genomic and cDNA libraries

- **Library**
  - Contains >=1 copy of all sequences
    - Chromosome library
    - Genomic library
    - cDNA library

- **Step 1: Obtain DNA to be cloned**
  - **Genomic DNA cut into small pieces**
  - **cDNA prepared from mRNA with reverse transcriptase**
- Step 2: Insert DNA fragment into vector
- Step 3: Insert vector into host
- Step 4: Allow host to replicate to high population #
- Step 5: Extract DNA

DNA Libraries

…collections of cloned DNA fragments,

- genomic,
- cDNA (coding sequences).

Genomic Library Construction

- Cloning all fragments of an organisms genome = genomic library
  - Each fragment in a vector transformed into a bacterial cell = Book
  - The collection of all of the clones (thousands) is the genomic library
  - requirement: *random
    * bigger is better

Genomic Library Construction

DNA fractioning

- DNA shearing
- to ensure that DNA fragments are of size suitable for cloning
  - Digest with enzyme with 4bp recognition site
  - One site every 256bp (1/4 x 1/4 x 1/4 x 1/4)
  - Do a partial digestion (not all sites cleaved)
  - Ligate fragments to vector
  - Transform into E. coli

Genomic Sequences and Coverage

$$N = \frac{\ln(1 - P)}{\ln(1 - f)}$$

N = number of clones

P = probability of recovering a sequence,

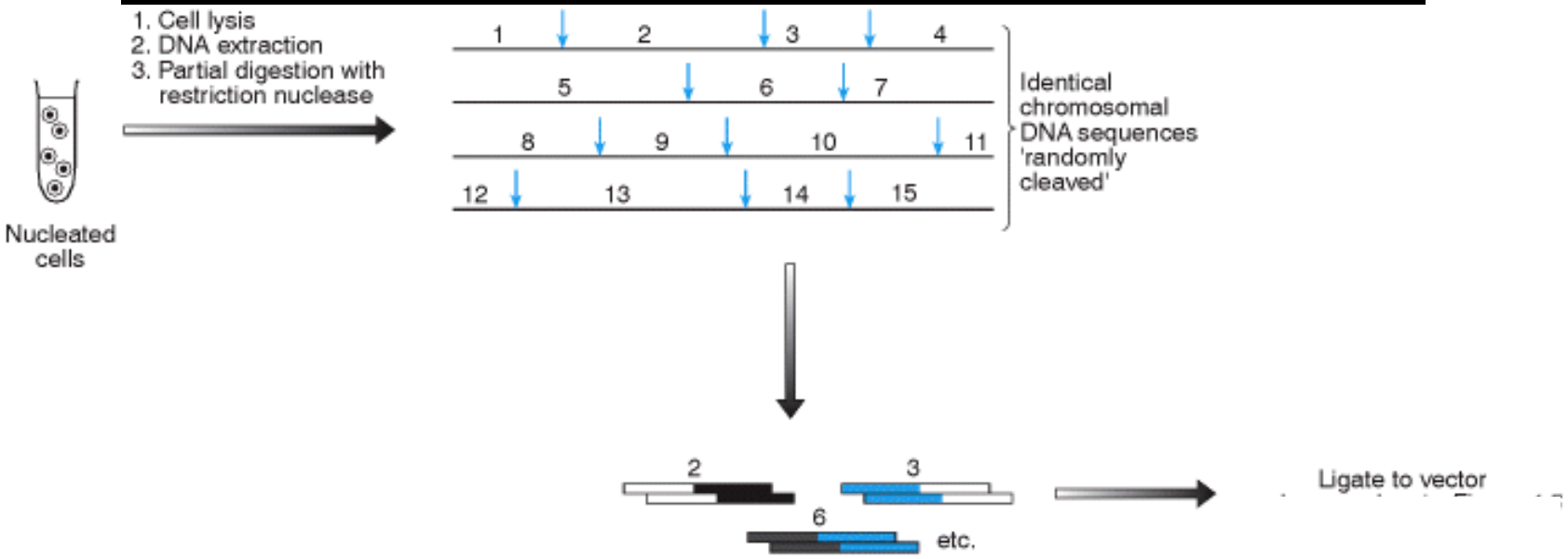
f = fraction of the genome of each clone

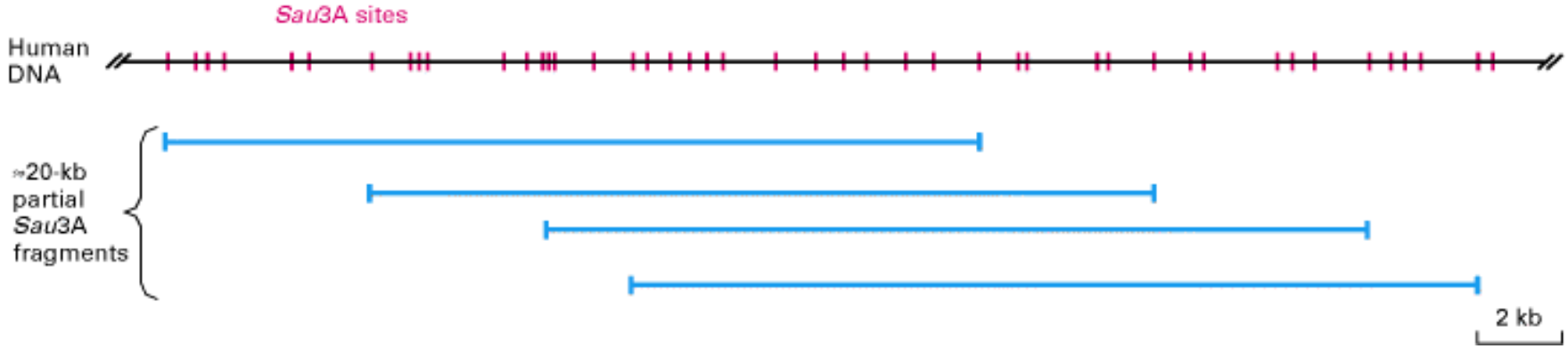
$$f = \frac{\text{genome lenght}}{\text{average lenght of insert}}$$

Probability	Insert lenght	
	15kb	40kb
0.99	860000	320000
0.95	560000	210000
0.9	430000	160000
0.8	300000	115000

Results for the human genome

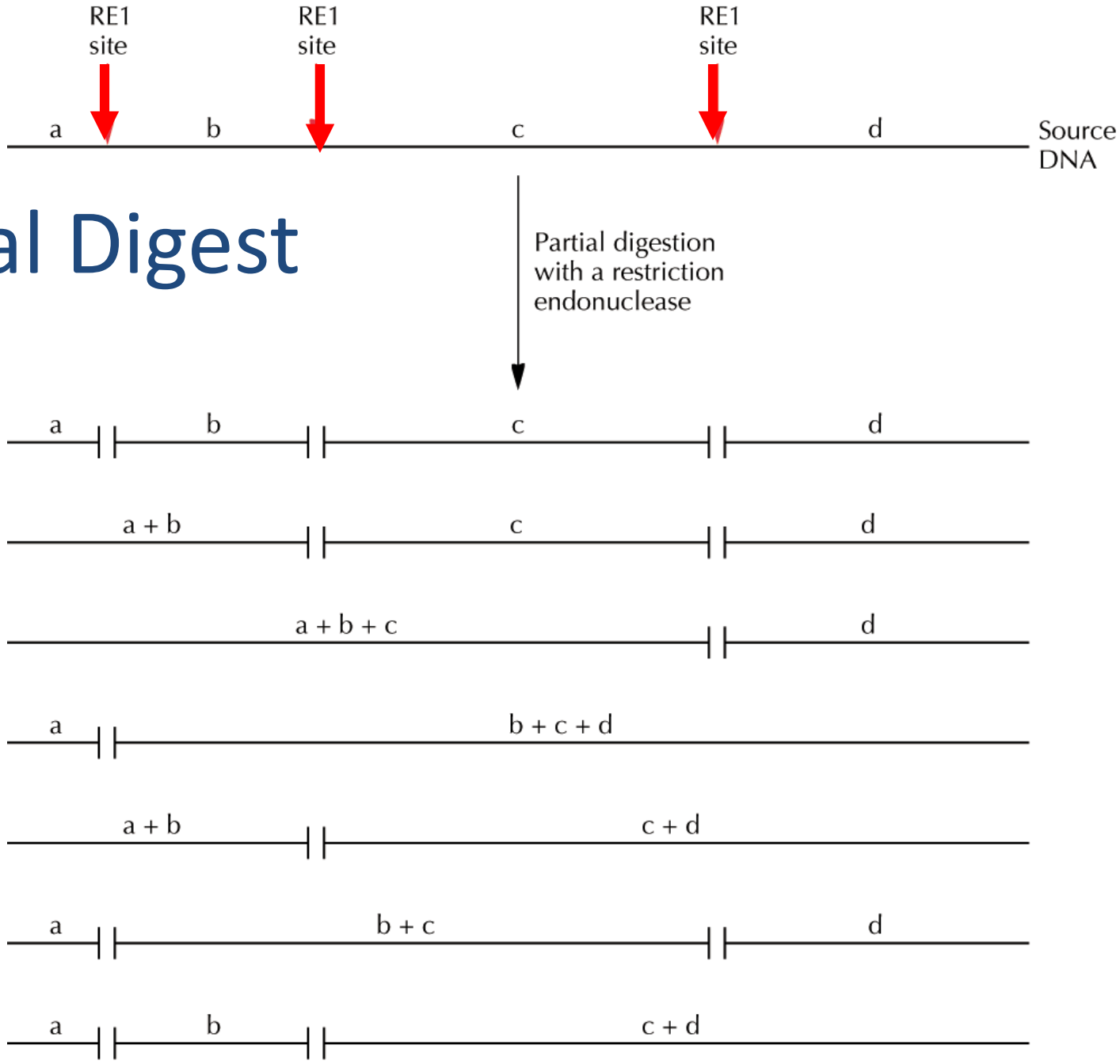
Making a genomic DNA library.



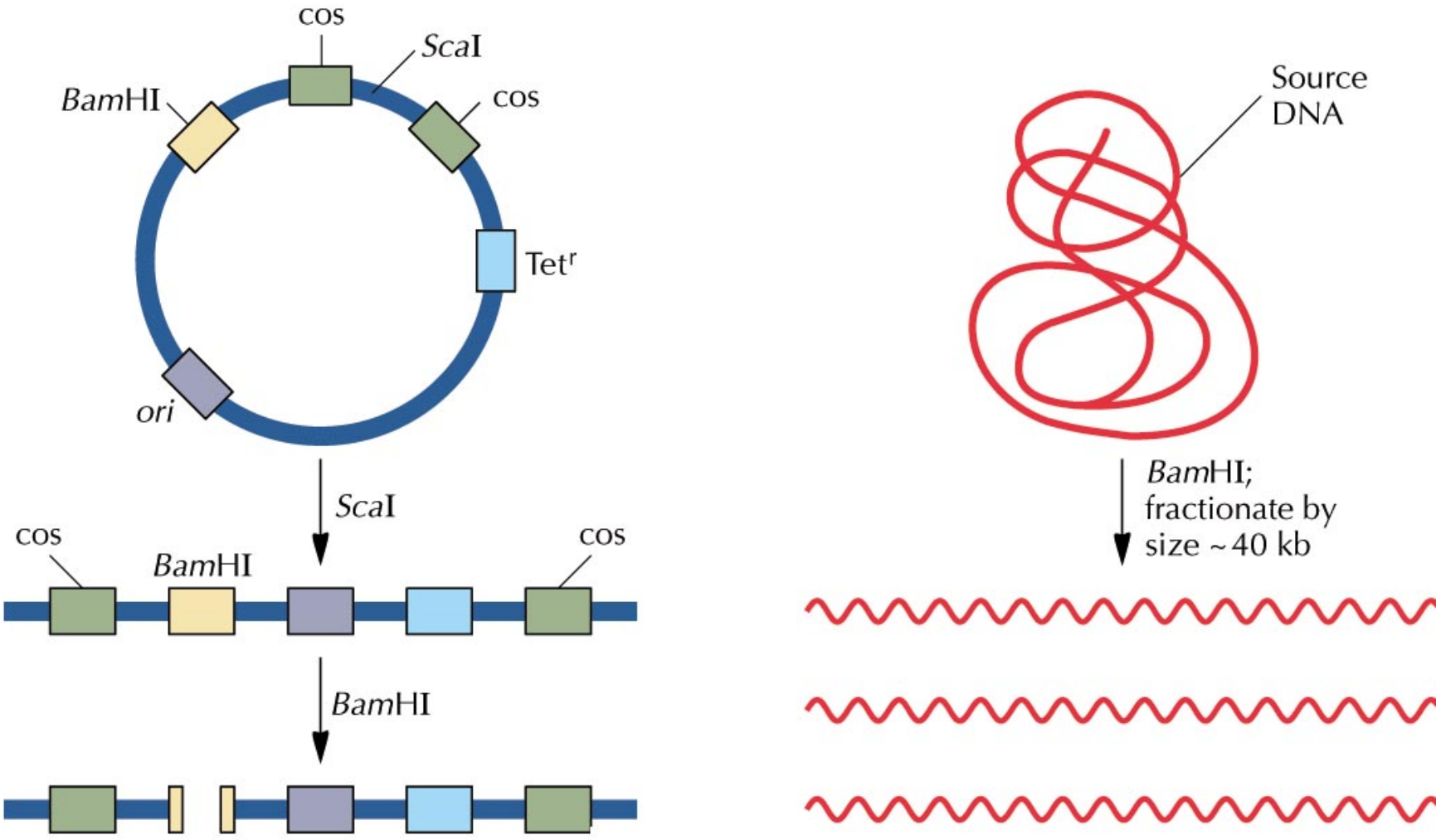


[redacted]ded sticky ends with this sequence on the 5 end of each strand. A hypothetical region of human genomic DNA showing the *Sau*3A recognition sites (red) is shown at the top. Partial digestion of this region of DNA would yield a variety of overlapping fragments (blue) ≈20 kb long. Use of such overlapping fragments increases the probability that all sequences in the genomic DNA will be represented in a λ library.

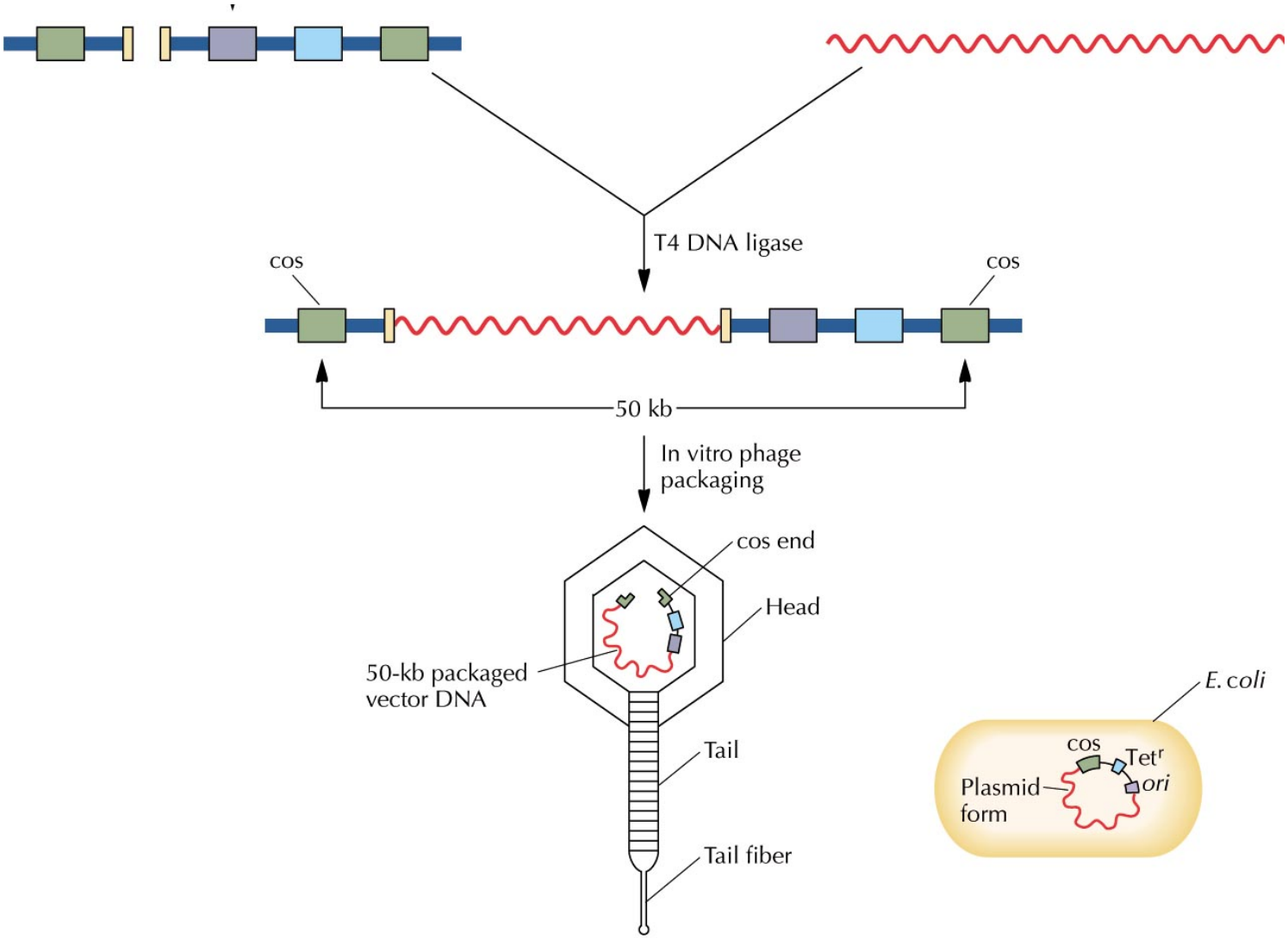
Partial Digest

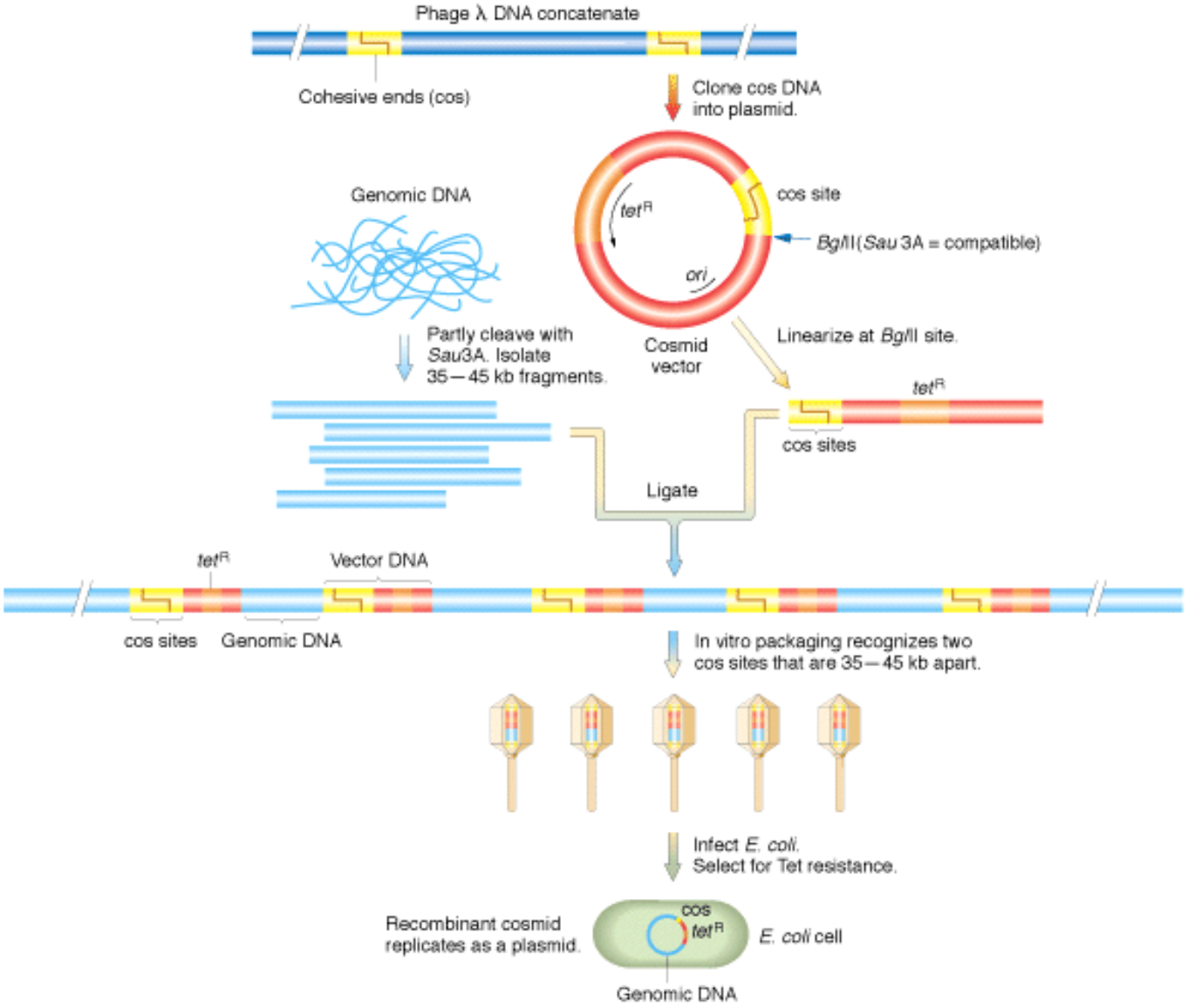


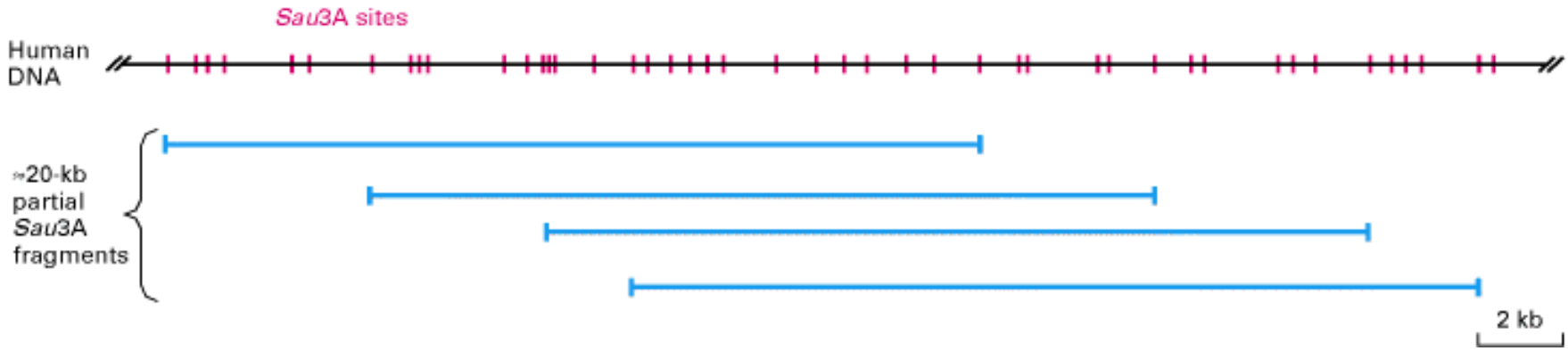
Cosmid Cloning



Cloning in Cosmids

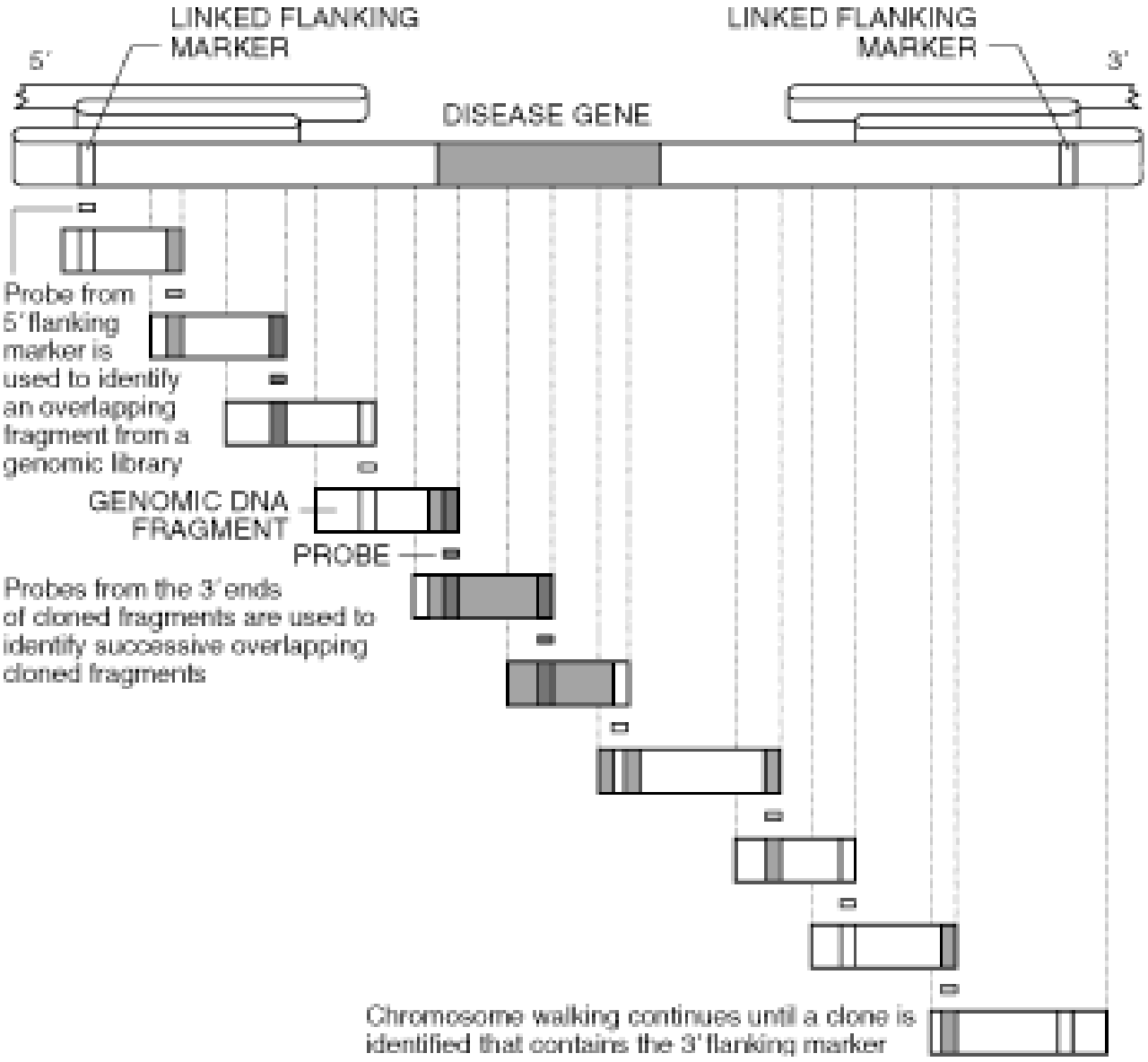



Phage λ DNA concatenate
Cohesive ends (cos)
Clone cos DNA into plasmid.
Genomic DNA
tetR
cos site
BglII (Sau 3A = compatible)
ori
Partly cleave with Sau3A. Isolate 35—45 kb fragments.
Cosmid vector
Linearize at BglII site.
tetR
cos sites
Ligate
tetR
Vector DNA
cos sites
Genomic DNA
In vitro packaging recognizes two cos sites that are 35—45 kb apart.
Infect E. coli. Select for Tet resistance.
Recombinant cosmid replicates as a plasmid.
cos
tetR
E. coli cell
Genomic DNA

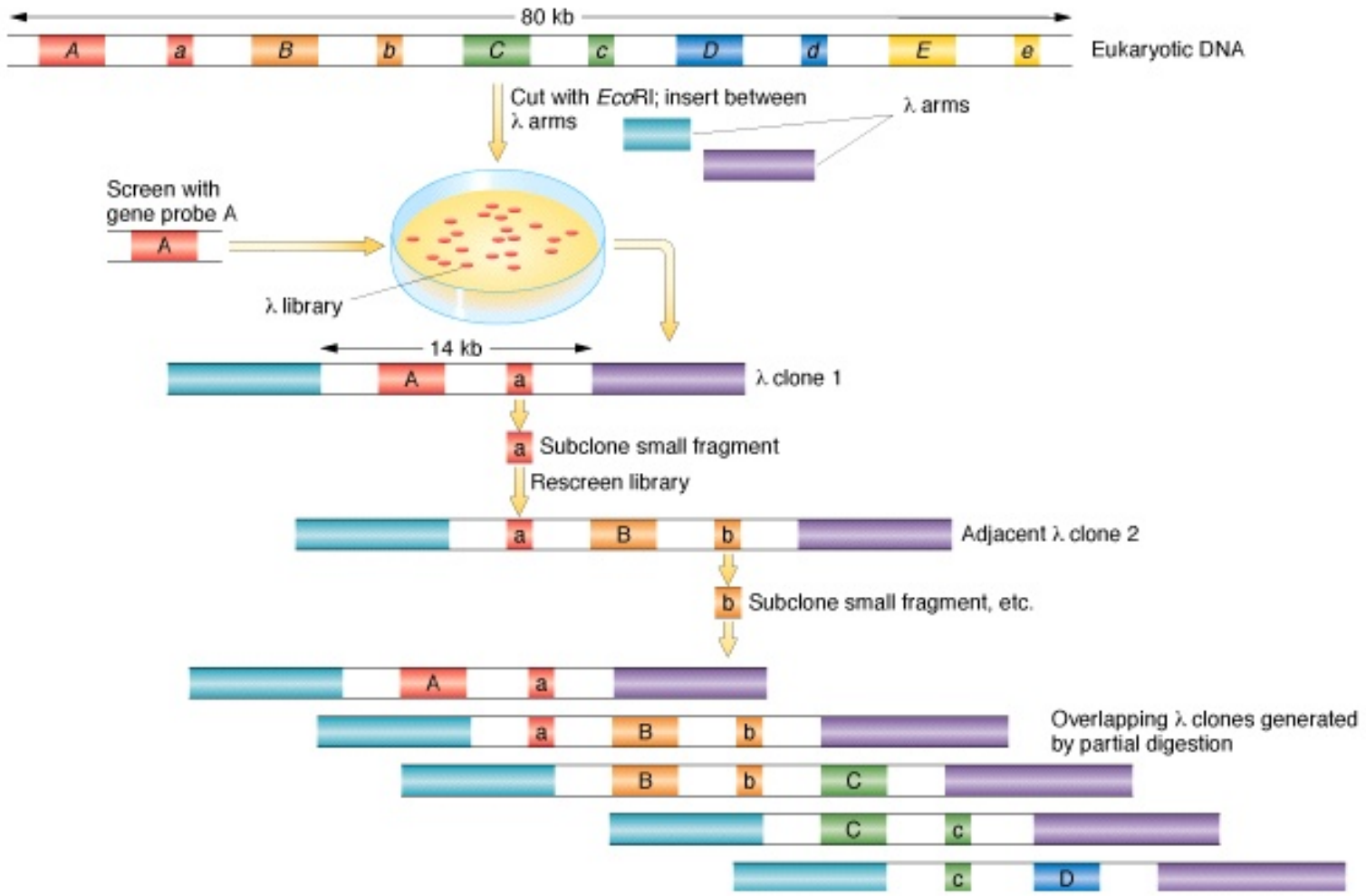


[redacted]

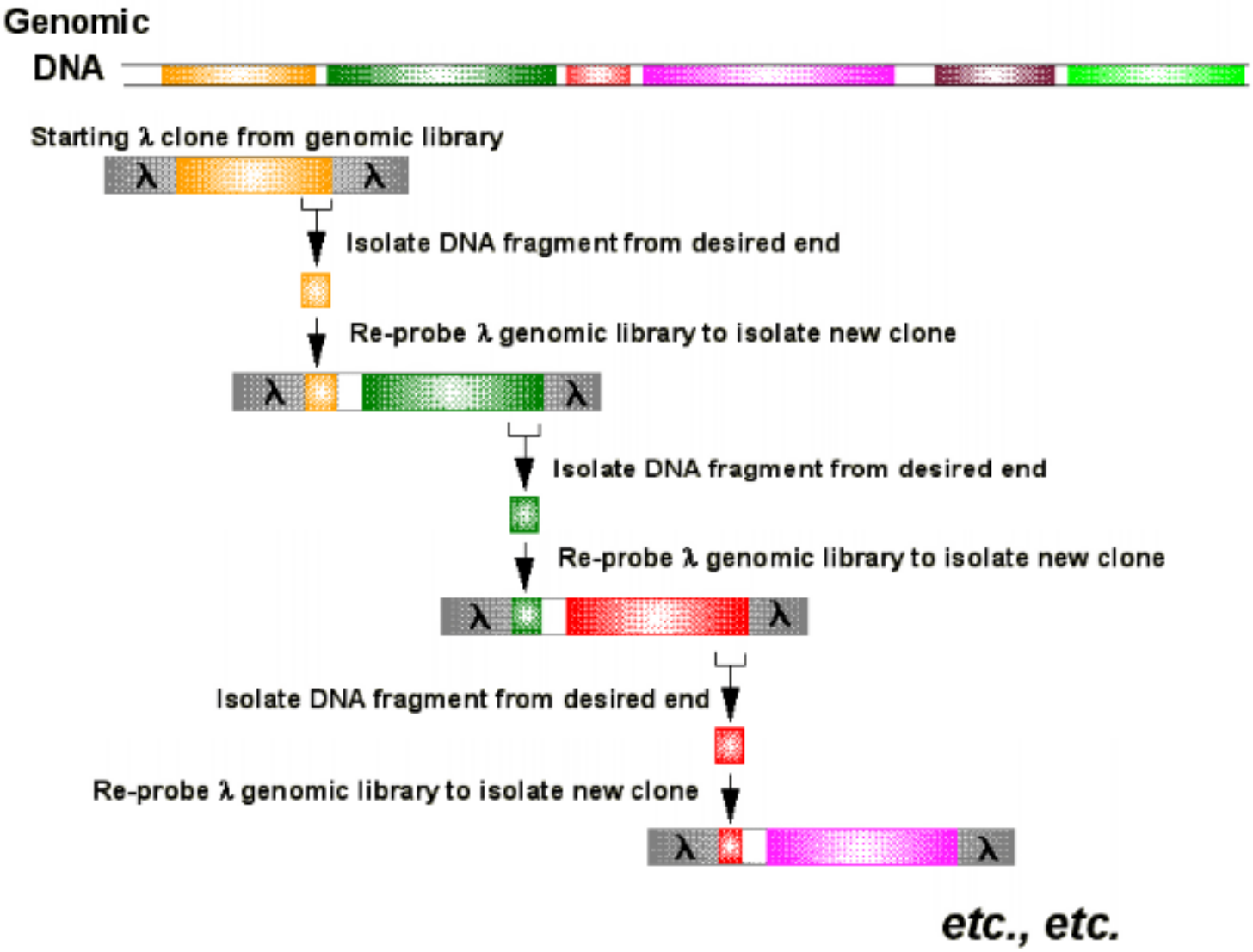
CHROMOSOME WALKING



CHROMOSOME WALKING



Chromosome ‘Walking’

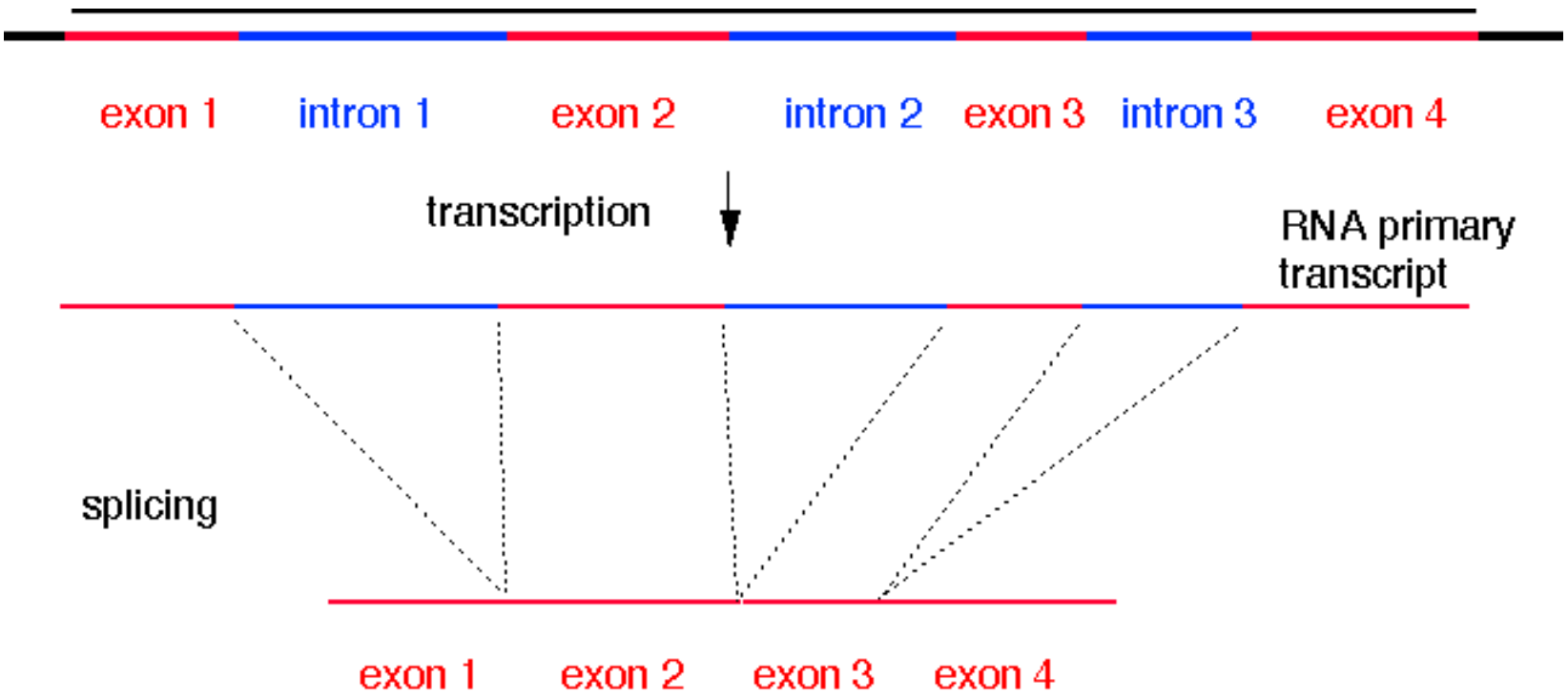


DNA Libraries

…collections of cloned DNA fragments,

- genomic,
- cDNA (coding sequences).

Cloning euykaryotic genes in prkaryotes require special "tricks" because eukaryotic genes have introns which are removed in the nucleus of eukaryotic cells prior to translation



Introns can account for more than 90% of the length of a eukaryotic gene. It is hard to clone very long DNA segments. In addition, intron-containing eukaryotic genes cannot be expressed in a bacterial host because prokaryotes lack splicing apparatus.

To overcome these problems, instead of directly cloning a gene, one can clone cDNA, a DNA copy of gene mRNA.

An enzyme, reverse transcriptase, is used to produce cDNA

cDNA

…DNA synthesized from an mRNA template with the enzyme **reverse transcriptase**.

Reverse Transcriptase

1. RNA dependent, DNA synthesis.

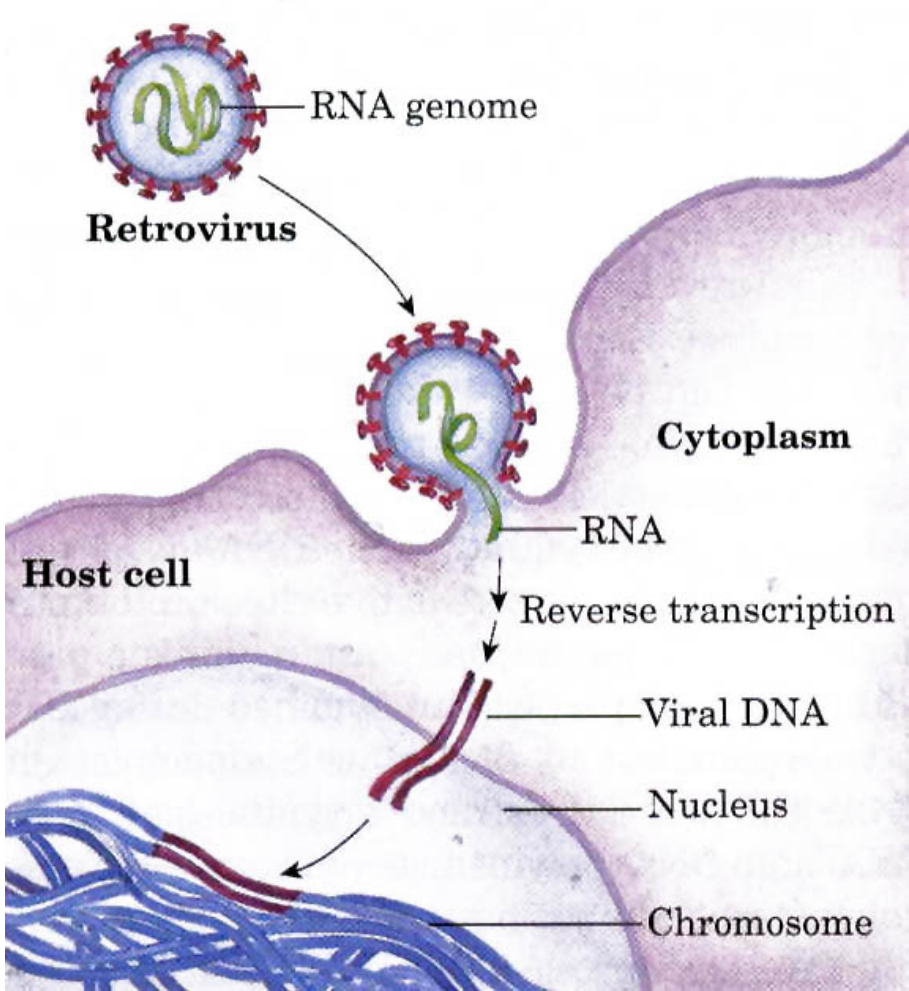
2. RNA Degradation.

3. DNA dependent, DNA Synthesis.

Basically two types:

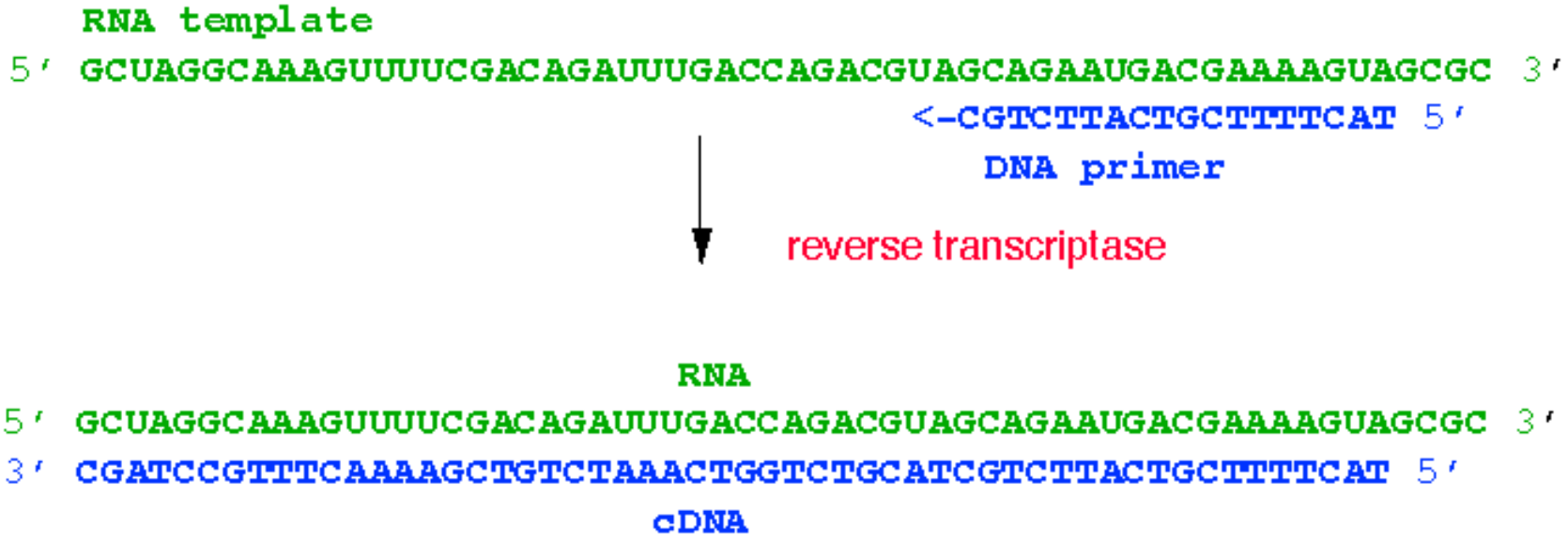
- AMV (aviary myeloblastosis)
- MMLV (Moloney Murine Leukemia Virus)

Error Rate: 1 in 20,000 nucleotides.

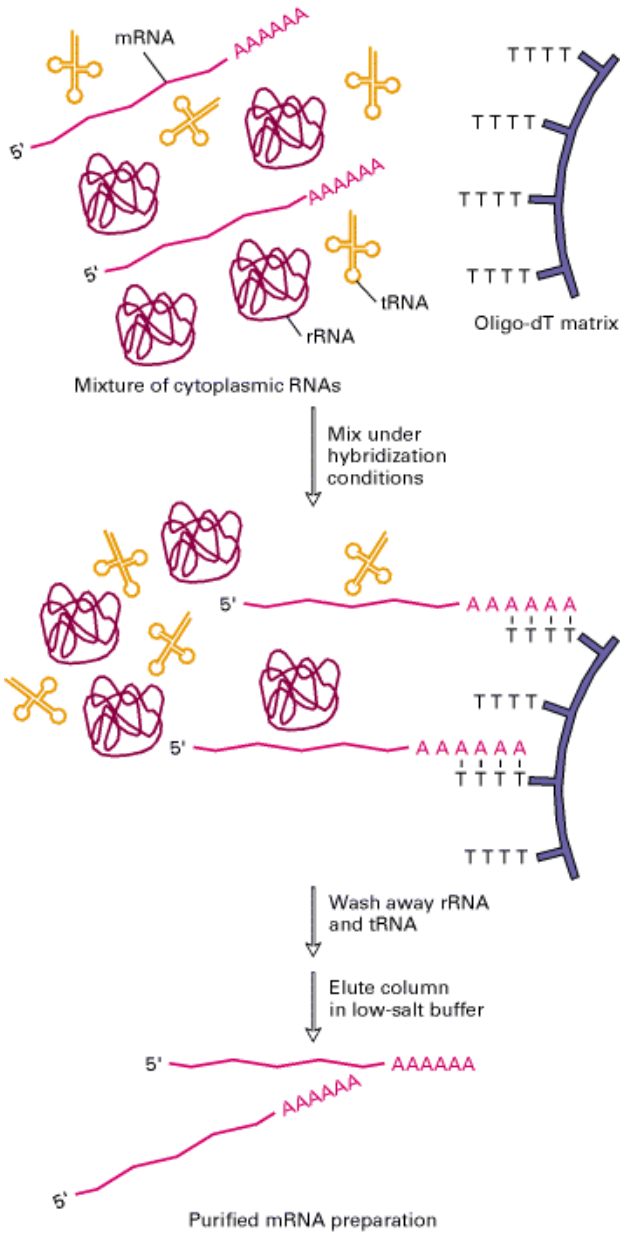


Reverse transcriptase is an RNA-dependent DNA polymerase: it synthesizes a complementary DNA strand on the RNA template.

Similar to other DNA-polymerases, reverse transcriptase needs a PRIMER to initiate DNA synthesis.



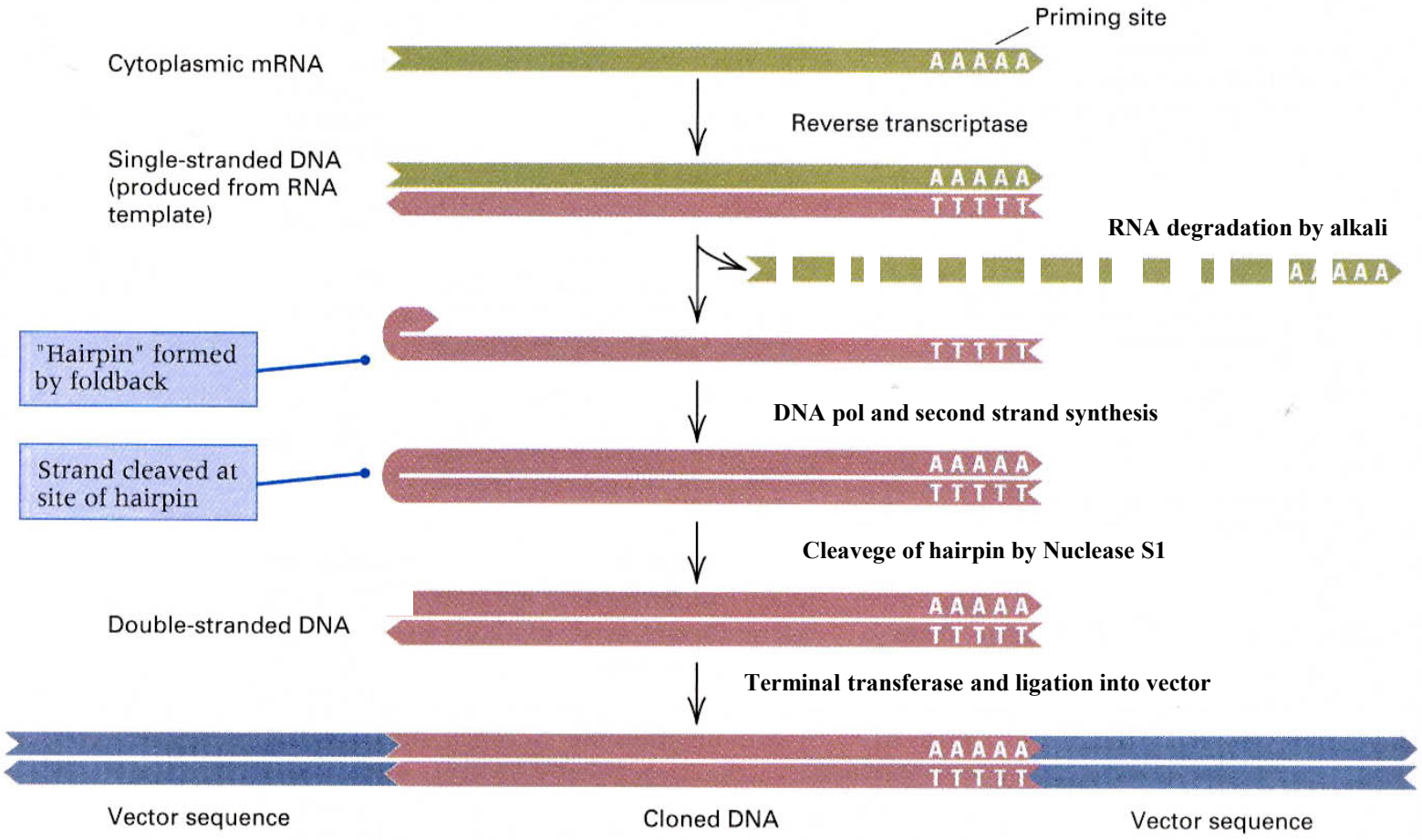
DNA strand synthesized on the RNA template is called **cDNA** (for complementary DNA)

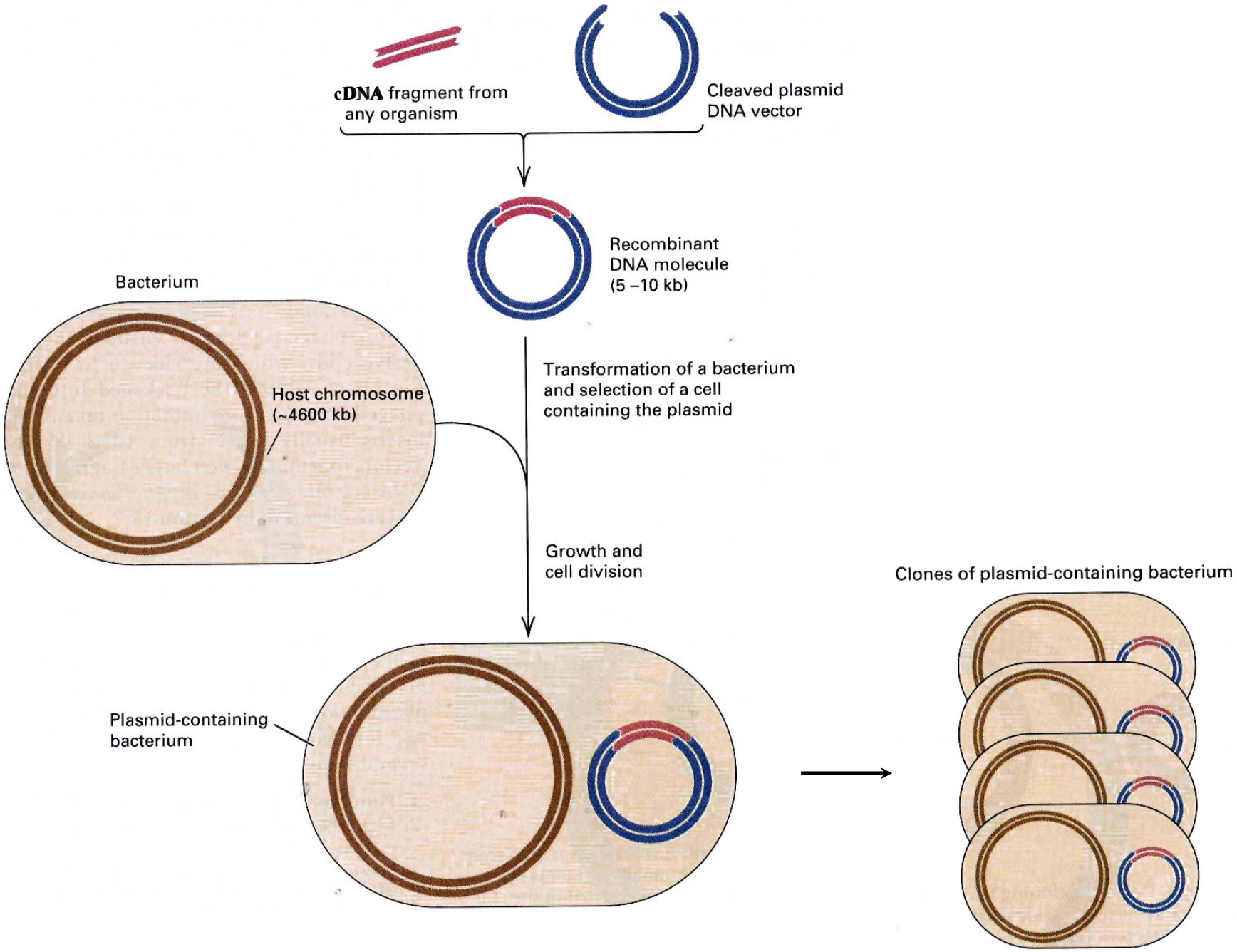


Isolation of mRNA by olig dT affinity chromatography.

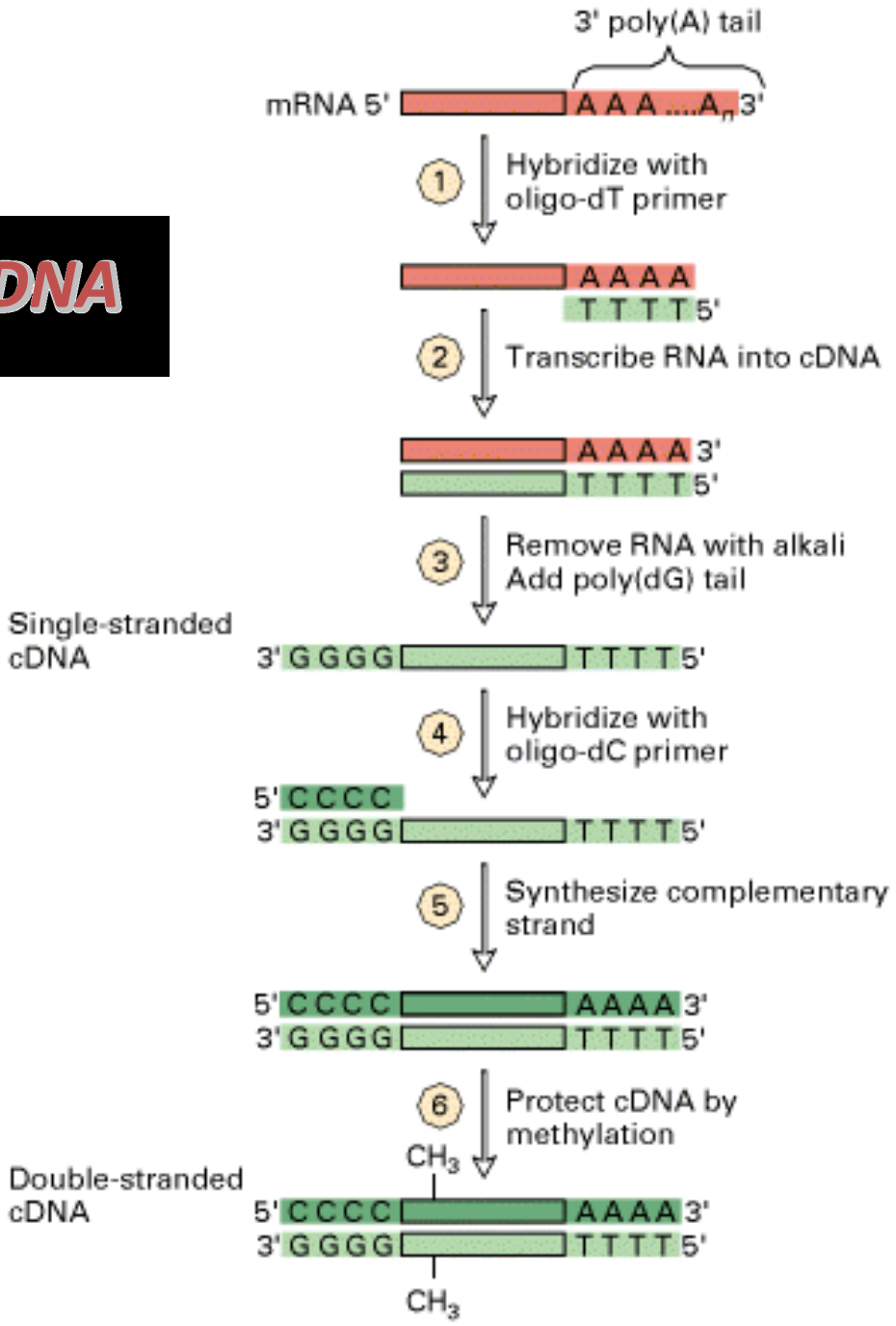
cDNA Construction

in vitro

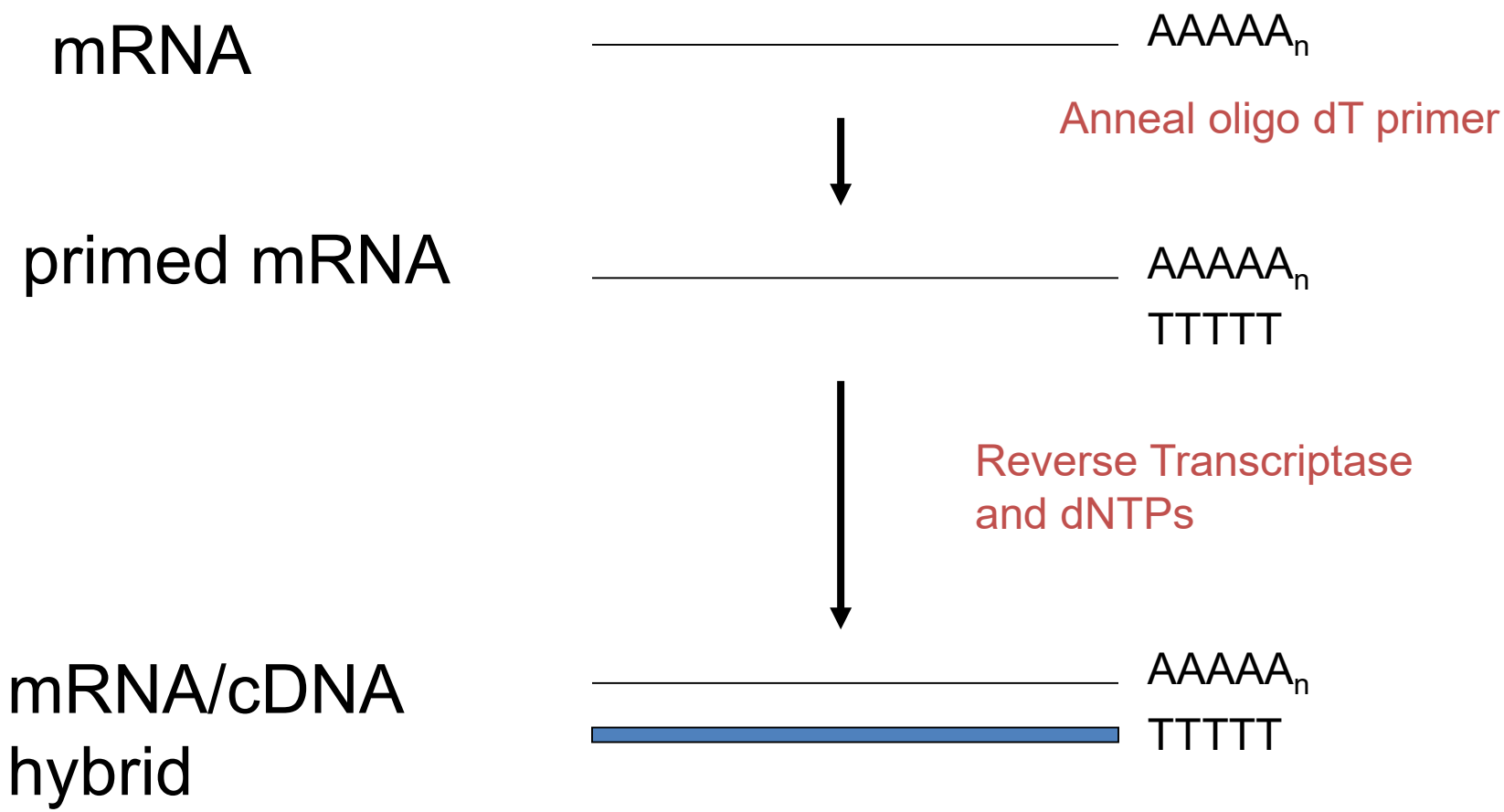



cDNA fragment from any organism
Cleaved plasmid DNA vector
Recombinant DNA molecule (5 –10 kb)
Bacterium
Host chromosome (~4600 kb)
Transformation of a bacterium and selection of a cell containing the plasmid
Growth and cell division
Plasmid-containing bacterium
Clones of plasmid-containing bacterium

Strategy to synthetize double strand cDNA

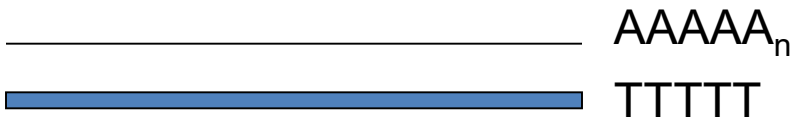


B) cDNA Synthesis



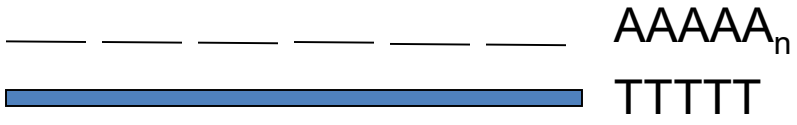
Gubler Hoffman cDNA Synthesis

mRNA/cDNA hybrid



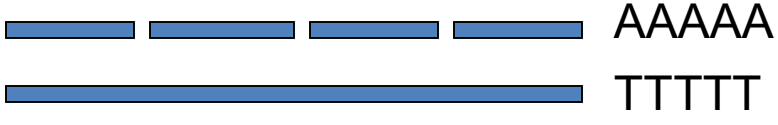
RNase H

nicked RNA



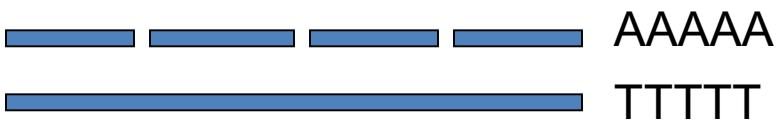
DNA Pol I

nicked RNA used as primers by Pol



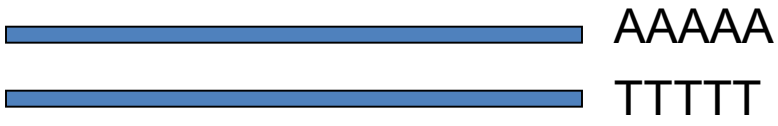
Gubler Hoffman cDNA Synthesis

2nd strand cDNA in pieces



E. coli DNA Ligase

ds cDNA



Clone into vector

cDNA Library

cDNA Libraries

…provide a ‘snap-shot’ of the genes expressed in a particular cell, at a particular time, or under specific condition,

…however, do not provide regulatory sequences.

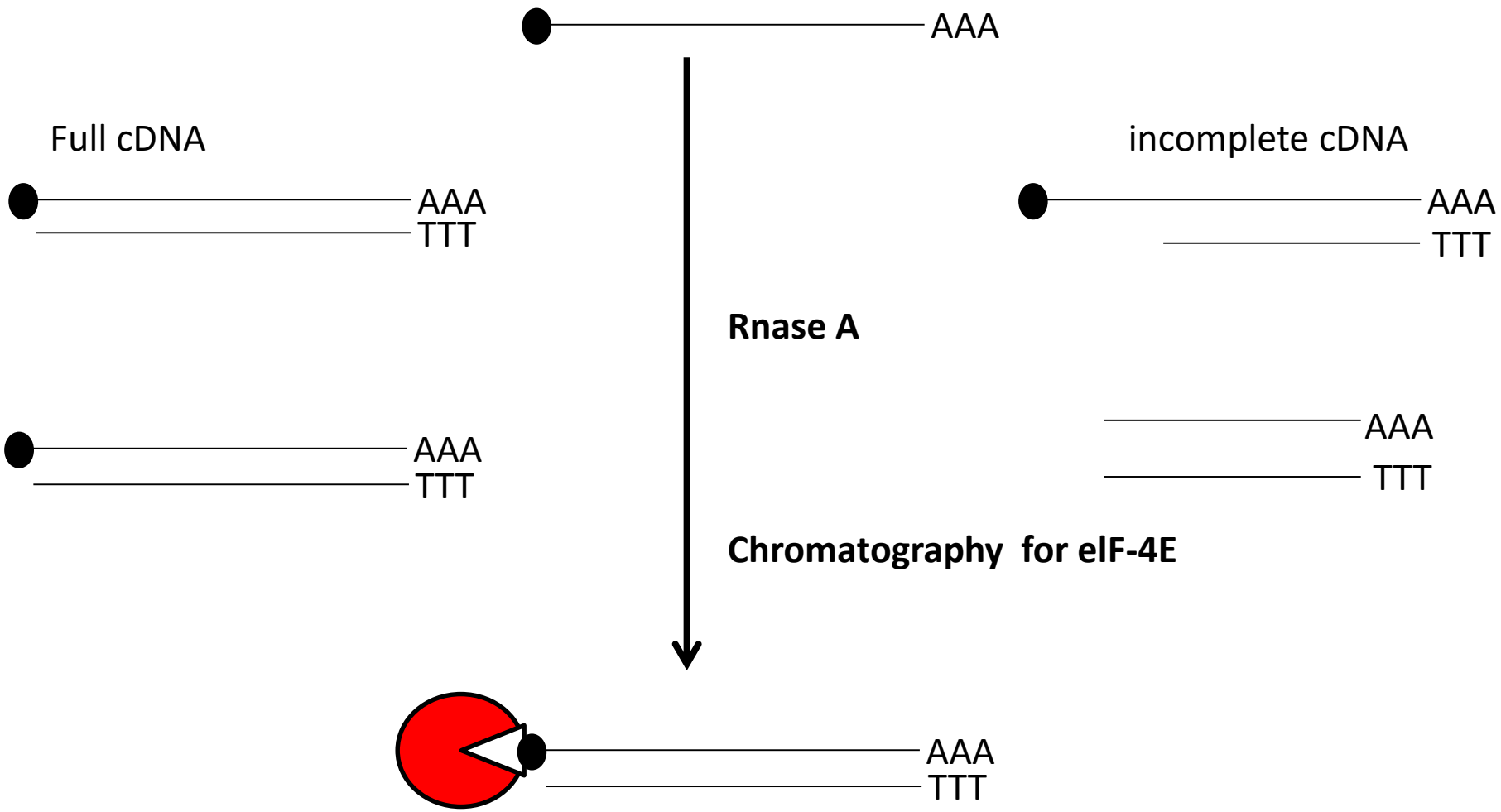
cDNA Libraries Limits:

…using a oligo dT, library has 3’-rich sequences

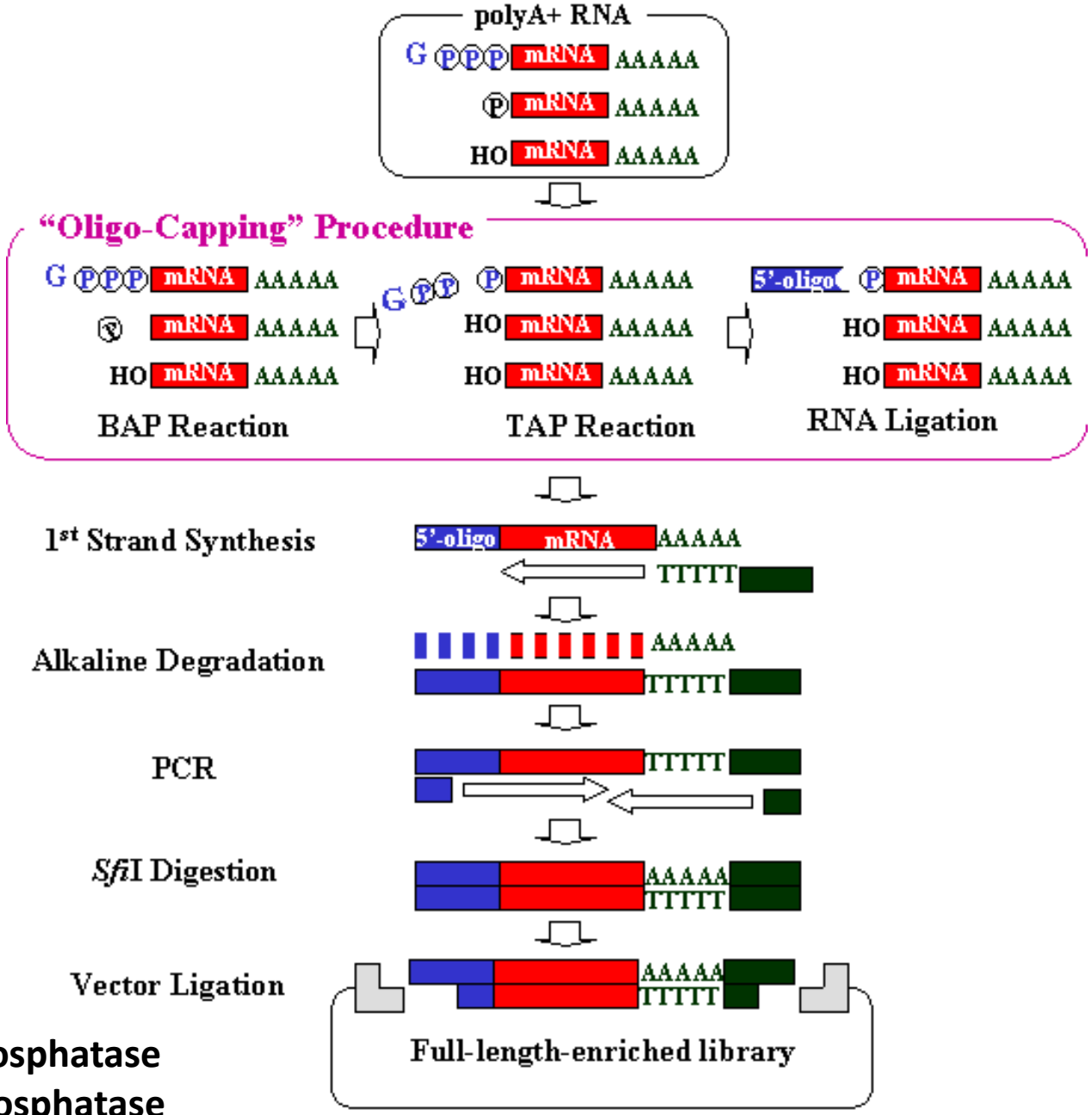
Using random examers it’s possible overcome this drawbacks, but the average length of sequences is reduced.

…hard to isolate long and full-lenght transcripts.

CAPture method:



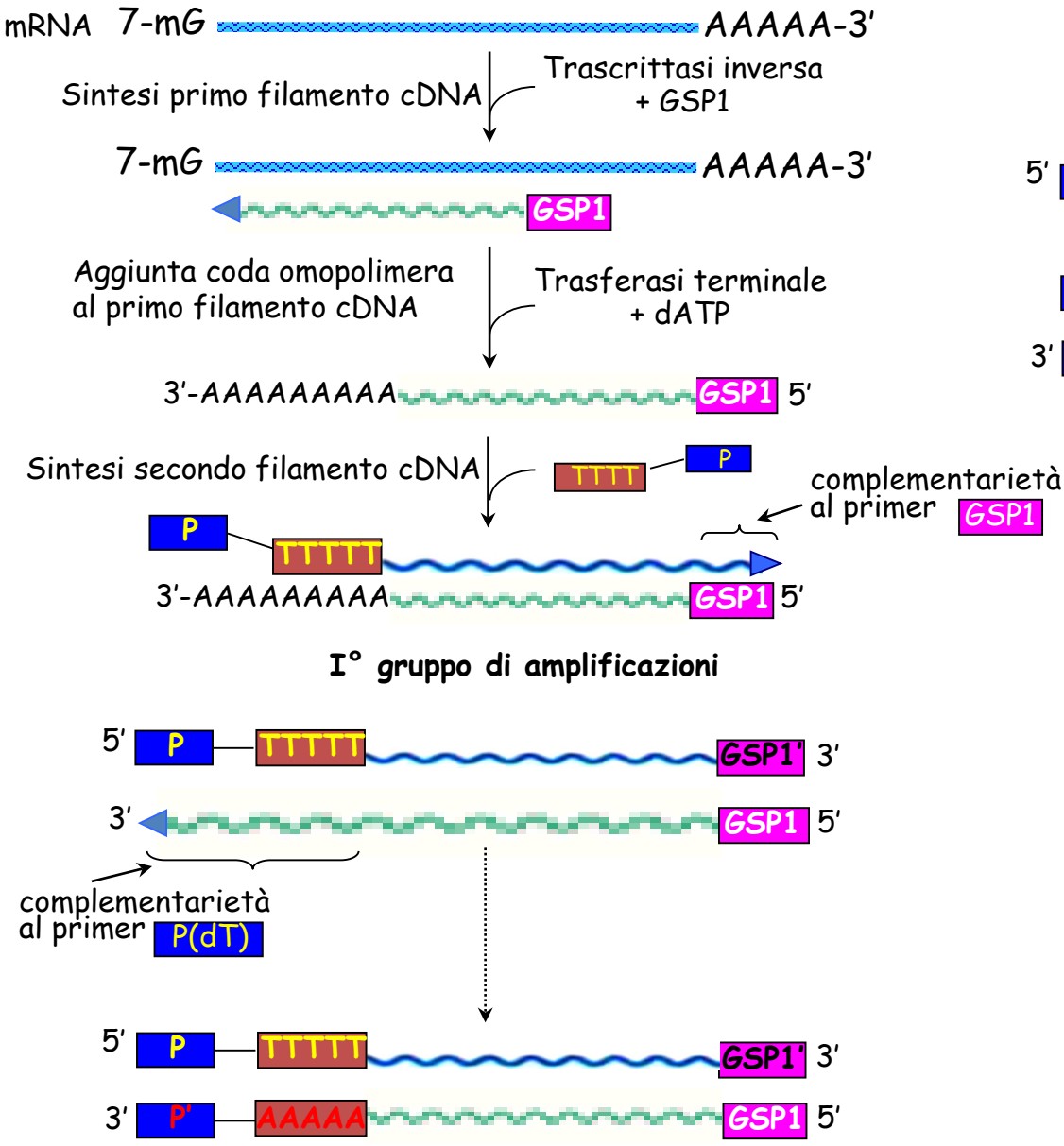
Scheme of Full-Length Library Construction

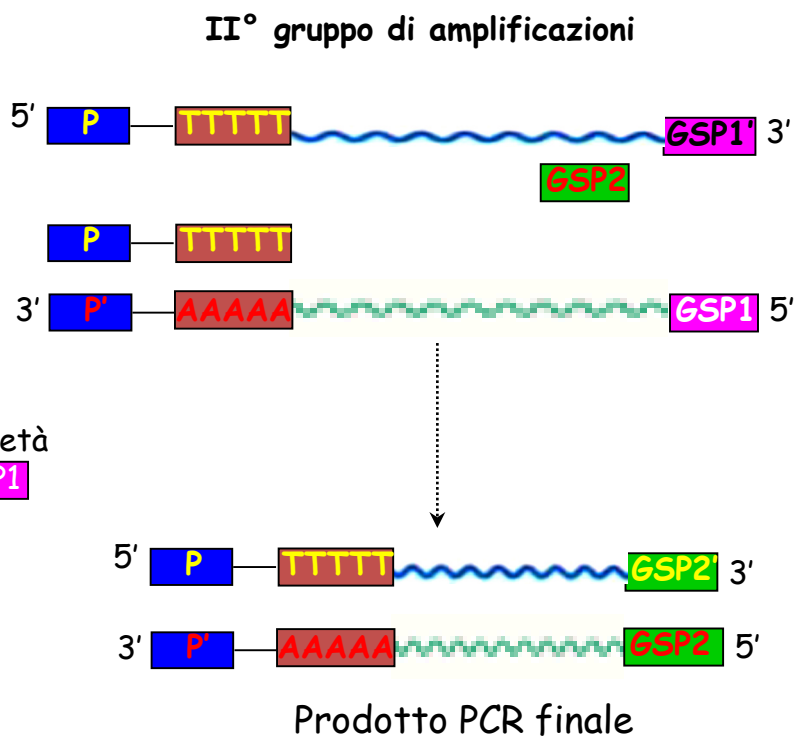


BAP= Alkaline phosphatase

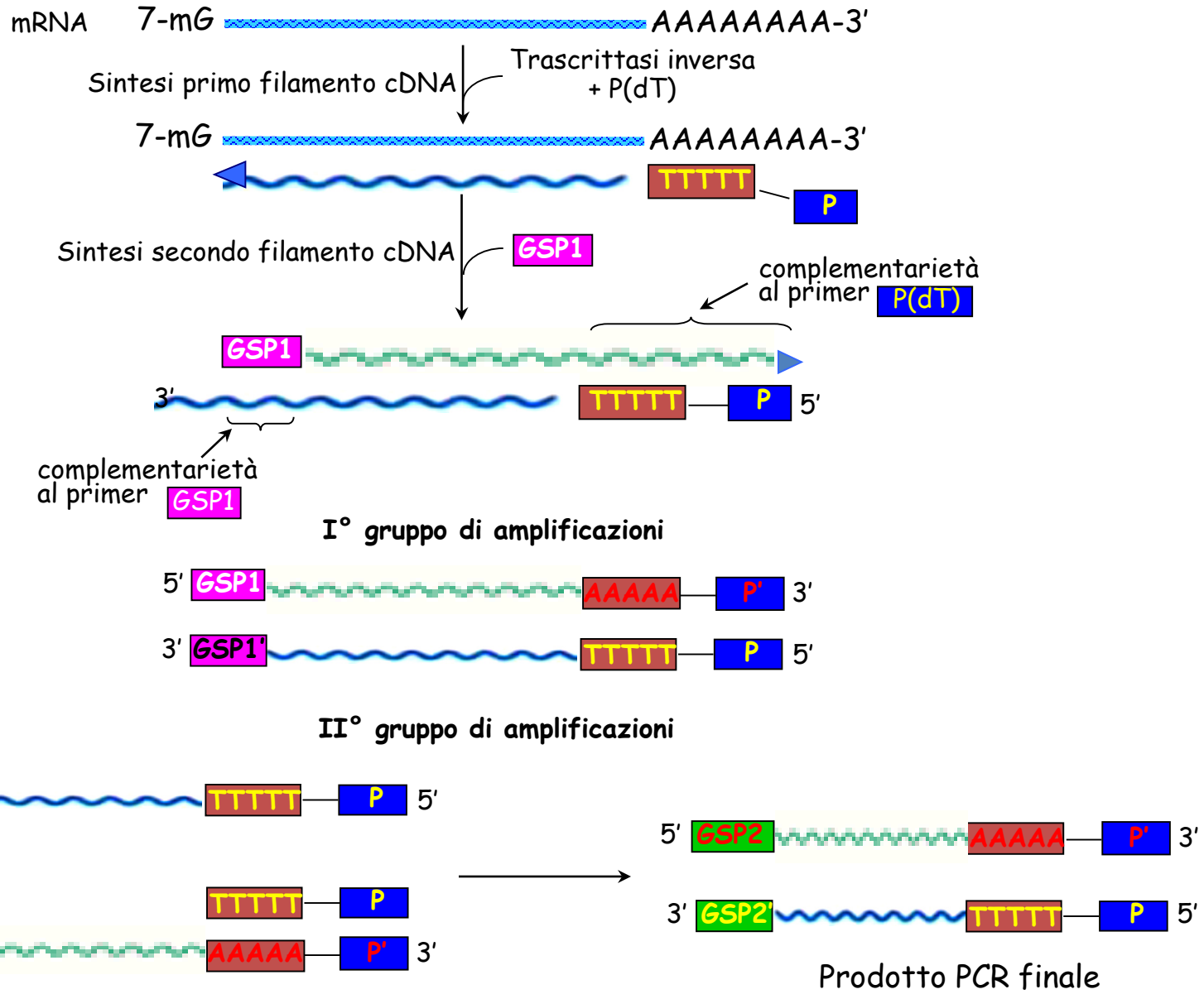
TAP= acid pyrophosphatase

5' - RACE Rapid Amplification Complementary Ends





3' - RACE Rapid Amplification Complementary Ends



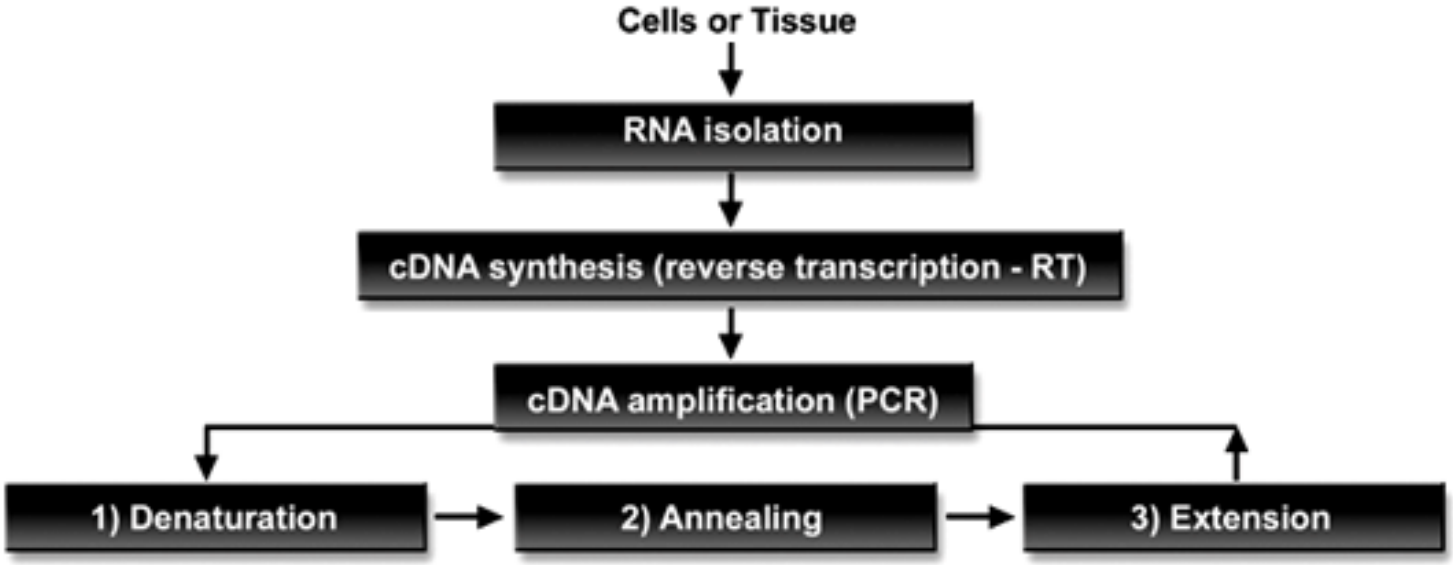
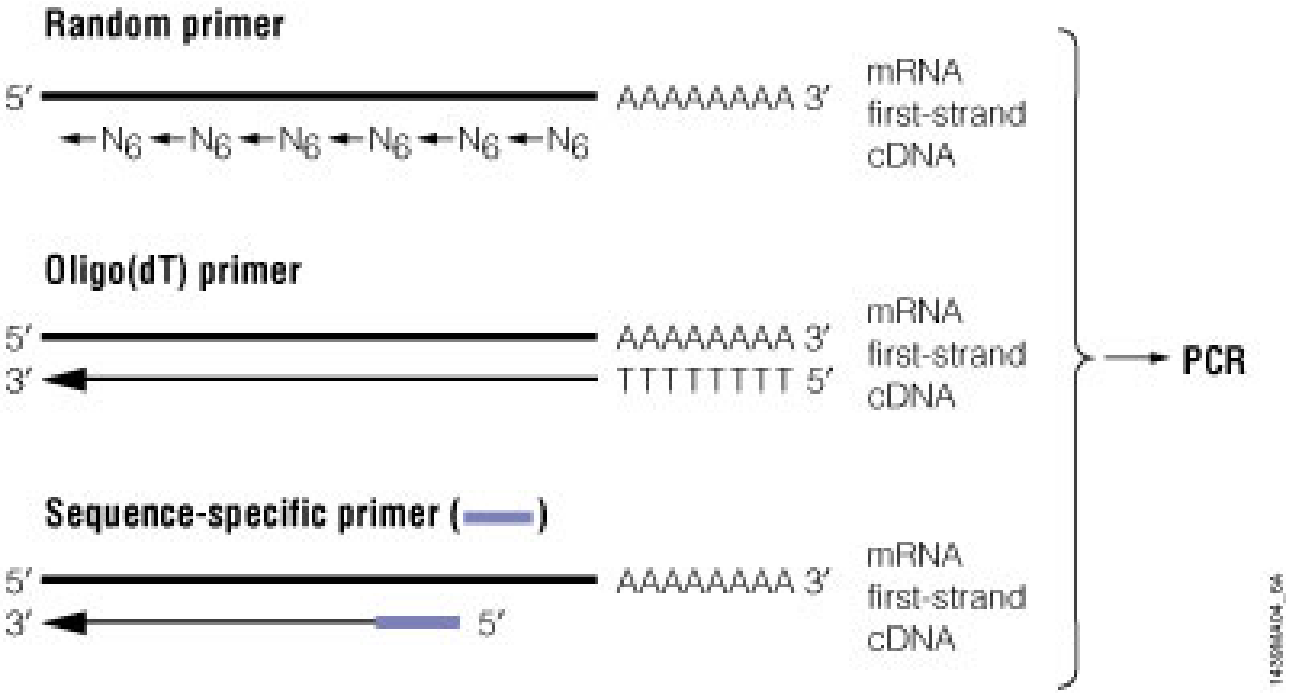


FIGURE 3 - Schematic diagram of RT-PCR showing that RNA isolated from cells or tissue is used as substrate in reverse transcription for synthesis of cDNA that will serve as template for amplification by PCR.

First Strand Synthesis:



o---- Cloning strategies ----o

I. Making DNA "libraries" (from genomic DNA, mRNA "transcriptome")

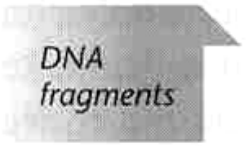
II. Screening to identify a specific clone (the needle in the haystack)

-- by the sequence of the clone

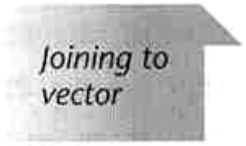
-- by the structure or function of the expressed product of the clone

Course reading: #28
(and 29)

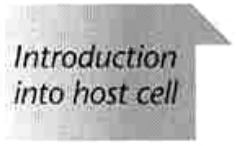
Overview of strategies for cloning genes

1) 

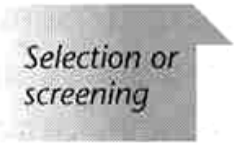
Get DNA

2) 

Ligate to vector

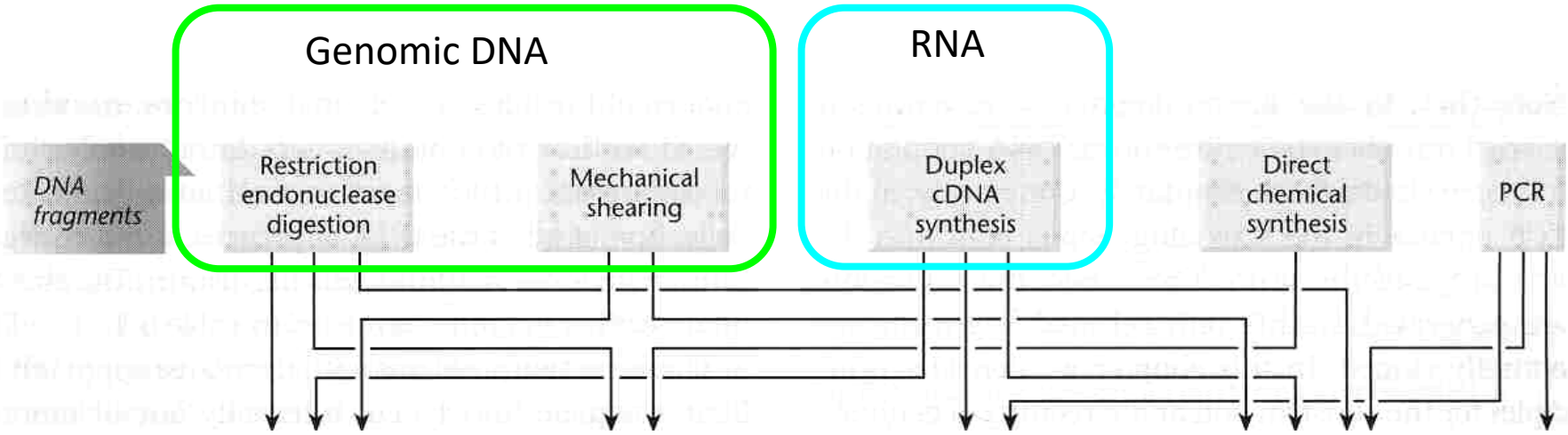
3) 

Transform or transfect

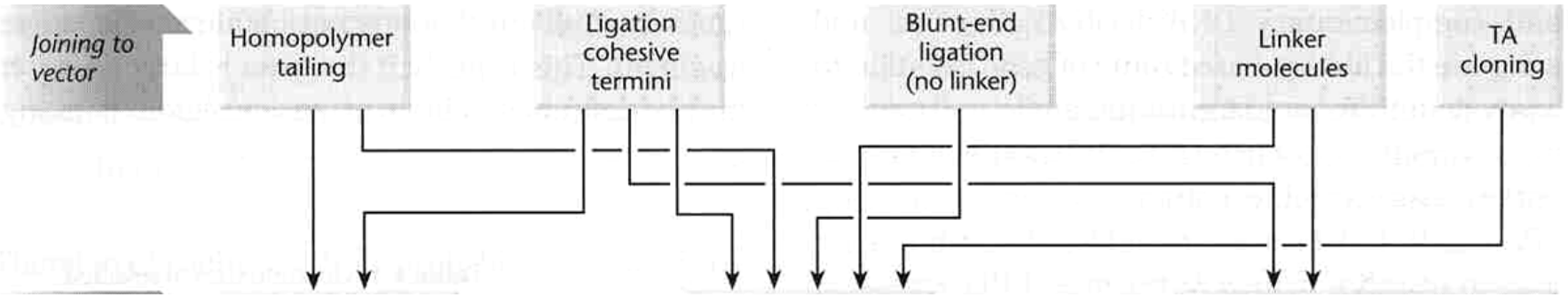
4) 

Look for the gene…

1) Get DNA

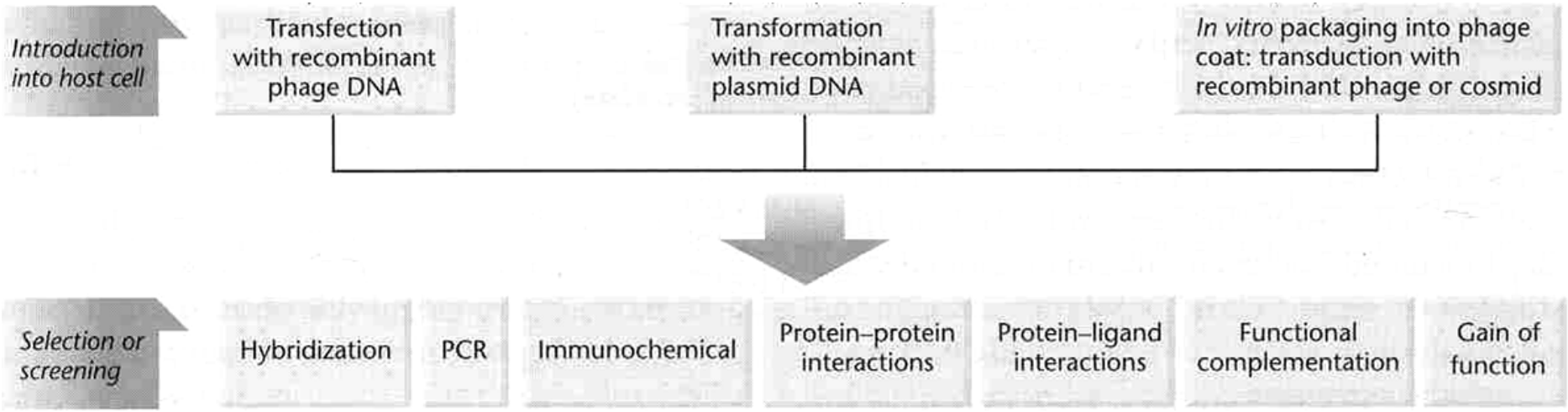

Genomic DNA
RNA
DNA fragments
Restriction endonuclease digestion
Mechanical shearing
Duplex cDNA synthesis
Direct chemical synthesis
PCR

Ligate to vector: how to make this reaction favorable?



This yields a “library”, a representative set of all the pieces of DNA that make up a genome (or all the cDNAs that correspond to the “transcriptome”)

cDNAs from different tissues reflect the different RNA populations that you find in distinct cell types:
Hence “liver” vs. “brain” vs. “heart” cDNA libraries



There are lots of ways to identify a particular gene…

Overview of strategies for cloning genes