PROPERTIES OF VECTORS TO CLONE/MANIPULATE DNA
A REVIEW!

Replication origin – Ability to replicate in Prokaryote +/or Eukaryotic cells (ORi)

Selectable/Distinguish from Non-Recombinant Host Vectors – ⓐ Antibiotic gene, ⓑ Color Marker gene, ⓒ ability to infect cells/package in virus

Unique/Single Restriction Sites for Cloning & Selection
Entire Sequence & Map available

EASily Re-introduced into host cells – ② TRANSFORMation, ① infection

EASily Purified from host cell & manipulated (e.g. ② plasmid, ① virus)

ALL are Genetically Engineered to meet cloning Experiment Needs – plasmid, expression plasmid, expression plasmid for animals, virus, plasmid/virus hybrid

E.G. – Plasmids Engineered to be small, have selectable markers, & unique cloning sites

There ARE A VARIETY OF VECTORS ALL ENGINEERED!

TABLE 8.2 Various Vectors and the Size of the Inserts They Carry

Vector	Form of Vector	Host	Typical Carrying Capacity (Size of Insert Accepted)	Major Uses
Plasmid	Double-stranded circular DNA	*E. coli*	Up to 15 kb	cDNA libraries; subcloning
Bacteriophage lambda	Virus (linear DNA)	*E. coli*	Up to 25 kb	Genomic and cDNA libraries
Cosmid	Double-stranded circular DNA	*E. coli*	30–45 kb	Genomic libraries
Phagemid	Virus convertible to plasmid	*E. coli*	Up to 12 kb	cDNA and genomic libraries
Bacteriophage P1	Virus (circular DNA)	*E. coli*	70–90 kb	Genomic libraries
BAC	Bacterial artificial chromosome	*E. coli*	100–500 kb	Genomic libraries
YAC	Yeast, artificial chromosome	Yeast	250–1000 kb (1 megabase)	Genomic libraries

Used in LABS TODAY

① PLASMIDS generally used for routine cloning + sequencing + cDNA libraries

② Genomic libraries usually made with VIRUS, BAC, or YAC vectors that can carry VERY LONG DNA segments (to keep # different clones in library small)

③ Vectors can be plasmids, VIRUSES, or artificial chromosomes or combinations of these vectors! With gene engineering - Can do anything!

Purpose of all Vectors - Clone / Manipulate DNA/cDNA

TWO COMMON PLASMID VECTORS

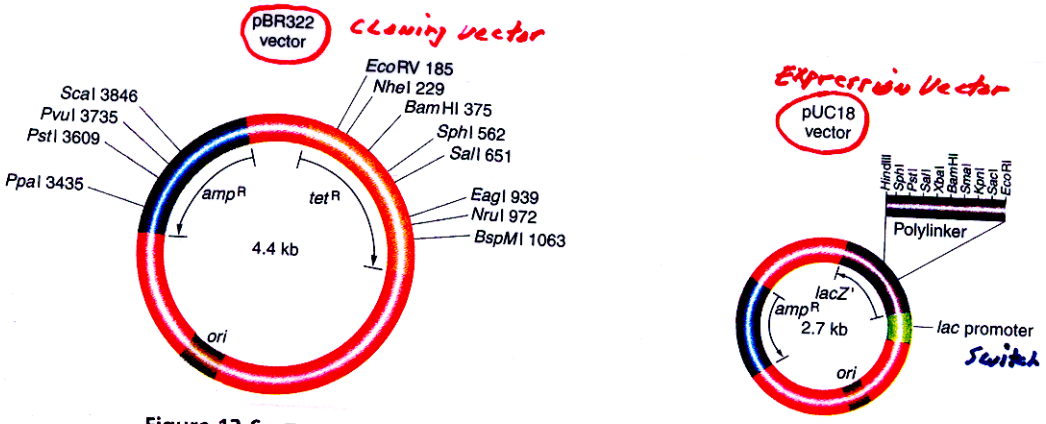


Figure 12-6 Two plasmids designed as vectors for DNA cloning, showing general structure and restriction sites. Insertion into pBR322 is detected by inactivation of one drug-resistance gene (tet^R), indicated by the Tet^S (sensitive) phenotype. Insertion into pUC18 is detected by inactivation of the β-galactosidase function of Z', resulting in an inability to convert the artificial substrate X-Gal into a blue dye.

CAN USE ANTIBIOTIC RESISTANCE AND COLOR TO SCREEN FOR RECOMBINANT PLASMIDS

pUC 18 vector

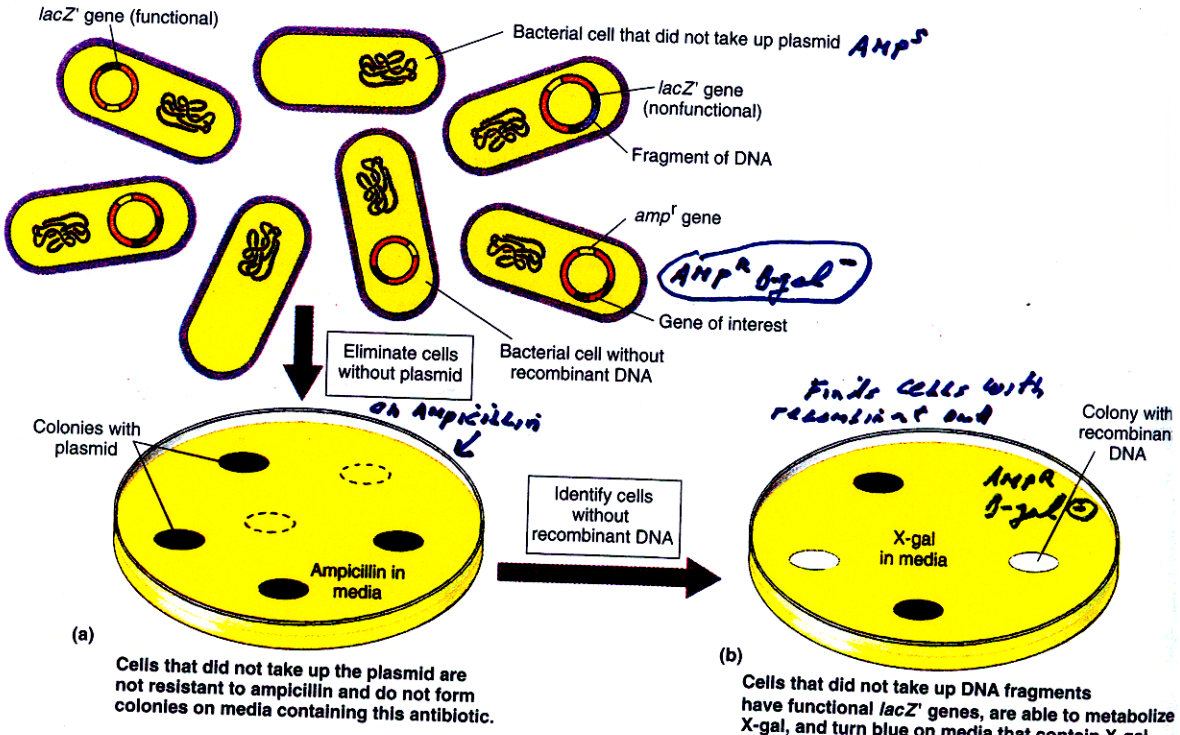


FIGURE 19.6

Stage 4-I: Using antibiotic resistance and X-gal as preliminary screens of restriction fragment clones. Bacteria are transformed with recombinant plasmids that contain a gene (*amp*^{r}) that confers resistance to the antibiotic ampicillin and a gene (*lacZ'*) that is required to produce β-galactosidase, the enzyme which enables the cells to metabolize the sugar X-gal. (*a*) Only those bacteria that have incorporated a plasmid will be resistant to ampicillin and will grow on a medium that contains the antibiotic. (*b*) Ampicillin-resistant bacteria will be able to metabolize X-gal if their plasmid does *not* contain a DNA fragment inserted in the *lacZ'* gene; such bacteria will turn blue when grown on a medium containing X-gal. Bacteria with a plasmid that has a DNA fragment inserted within the *lacZ'* gene will not be able to metabolize X-gal and, therefore, will remain colorless in the presence of X-gal.

X-gal → Blue color

β-galactosidase enzyme

NORMALLY

Lactose → galactose + glucose

MILK sugar

Libraries - A Review!

What are the differences between genomic and cDNA libraries?

A library is a collection of individual DNA clones

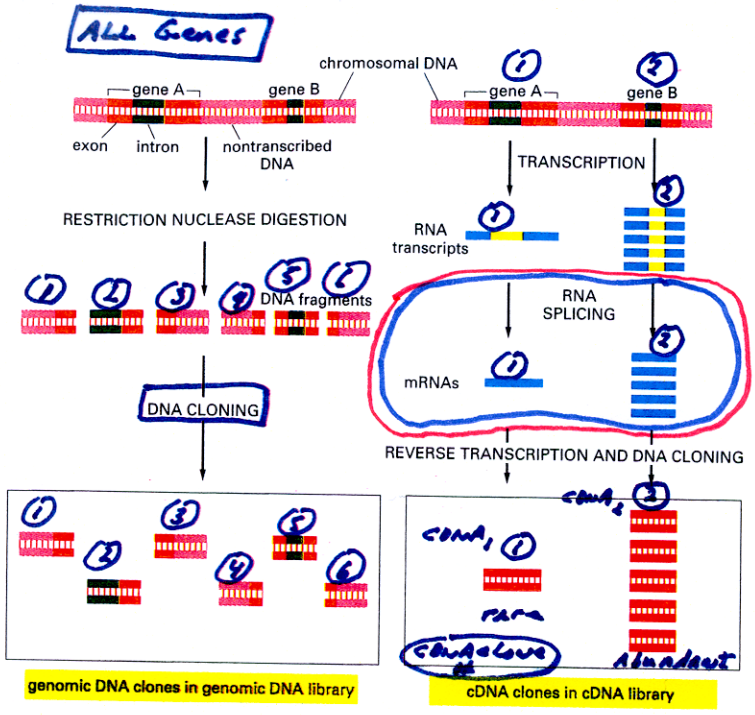


Figure 8–35 The differences between cDNA clones and genomic DNA clones derived from the same region of DNA. In this example gene A is infrequently transcribed, whereas gene B is frequently transcribed, and both genes contain introns *(green)*. In the genomic DNA library, both the introns and the nontranscribed DNA *(pink)* are included in the clones, and most clones contain, at most, only part of the coding sequence of a gene *(red)*. In the cDNA clones the intron sequences *(yellow)* have been removed by RNA splicing during the formation of the mRNA *(blue)*, and a continuous coding sequence is therefore present in each clone. Because gene B is transcribed more abundantly than gene A in the cells from which the cDNA library was made, it is represented much more frequently than A in the cDNA library. In contrast, A and B are in principle represented equally in the genomic DNA library.

① ALL Genes in Library
② EACH Gene/DNA represented equally in Genome Library

1 DNA$_X$ → 10 DNA$_X$ in library
1 DNA$_Y$ → 1 DNA$_Y$ in library

Genomic Clones

① ALL sequences in Genome → genes/switches
② Complete Gene - Exons + introns to understand Gene Structure + Evolution + Mutation/Diseases
③ Needed For Genome Sequencing Projects

① ONLY mRNAs present in specific cells/organs in Library
② Subset of genes in genome
③ cDNA clones not present equally - present in proportion to amount of mRNA sequence in cell

1 mRNA$_X$ → 1 cDNA$_X$
100 mRNA$_Y$ → 100 cDNA$_Y$

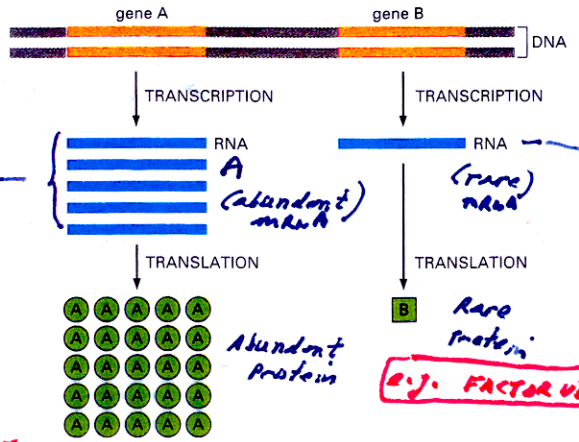
EASier to find Abundant cDNA in Library!

cDNA Clones

① ONLY Coding sequences i.e. useful to identify proteins
② Subset of gene sequences - what genes active in specific cell - time of development - Tumor? Profile of active Genes
③ useful to identify Specific "genes"
④ For Drugs / Bacteria expression

cDNA CLONES REPRESENT GENE CODING SEQUENCES AND ARE PRESENT IN PROPORTION TO mRNAs in Cell

Genes Active in Organ X



Genes active at different levels!!

Hard to clone

Easy to clone directly from tissue mRNA

5 cDNA clones for mRNA A

1 cDNA clone for mRNA B

∴ Find cDNA clones for Abundant mRNAs more frequently than ~ Rare mRNA

What's consequence for screening specific cDNA clones from cDNA library?

IF KNOW where Gene is Active → CAN Isolate cDNA CLONE FROM LIBRARY Made from TISSUE/CELL mRNA ~ e.g. Insulin

SELECTING A SPECIFIC cDNA CLONE FROM A cDNA Library

Using Nucleic Acid +/or Antibody Probes

Nucleic Acid Probes
① purified mRNA probe
② protein sequence → probe
③ gene probe

Any Vector

Plate cDNA Library

Need Purified Protein
① Antibody Probe
② Need Expression Vector

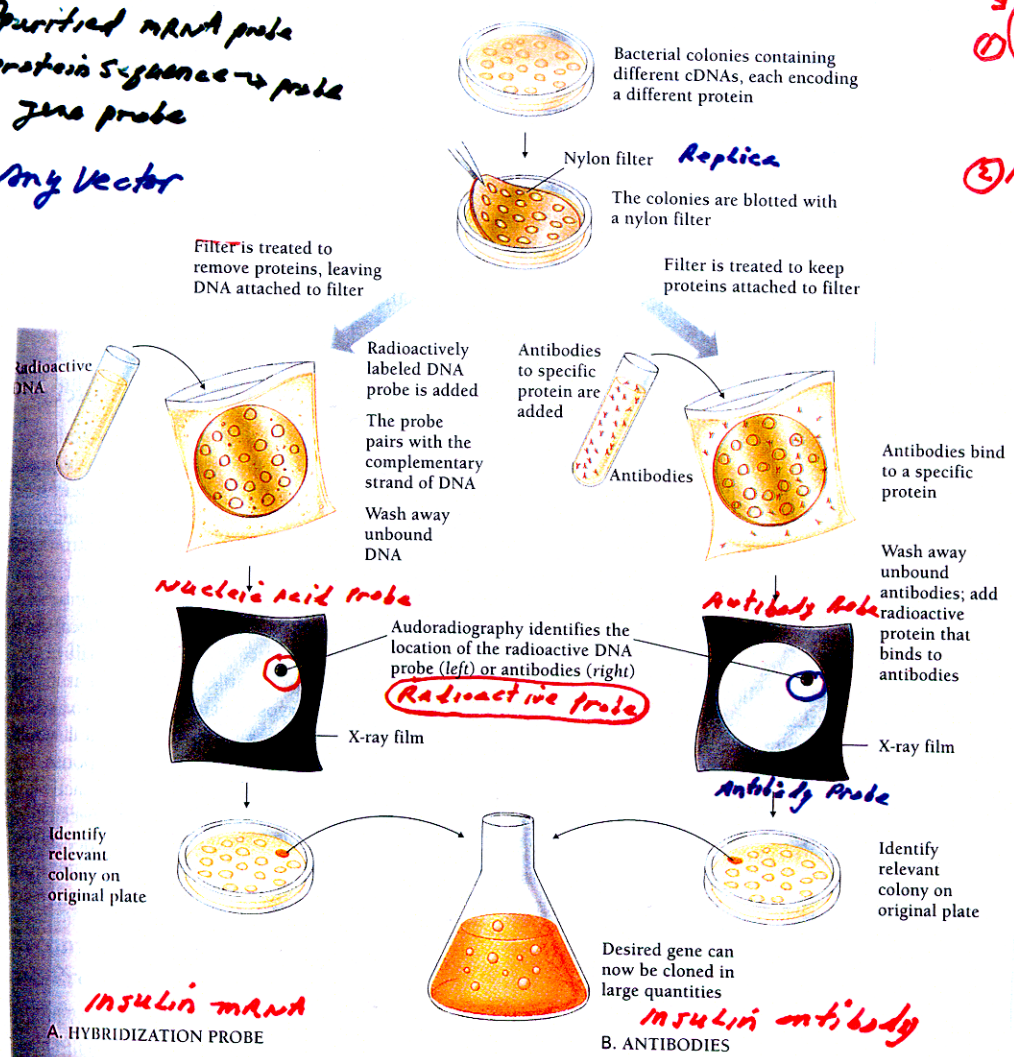
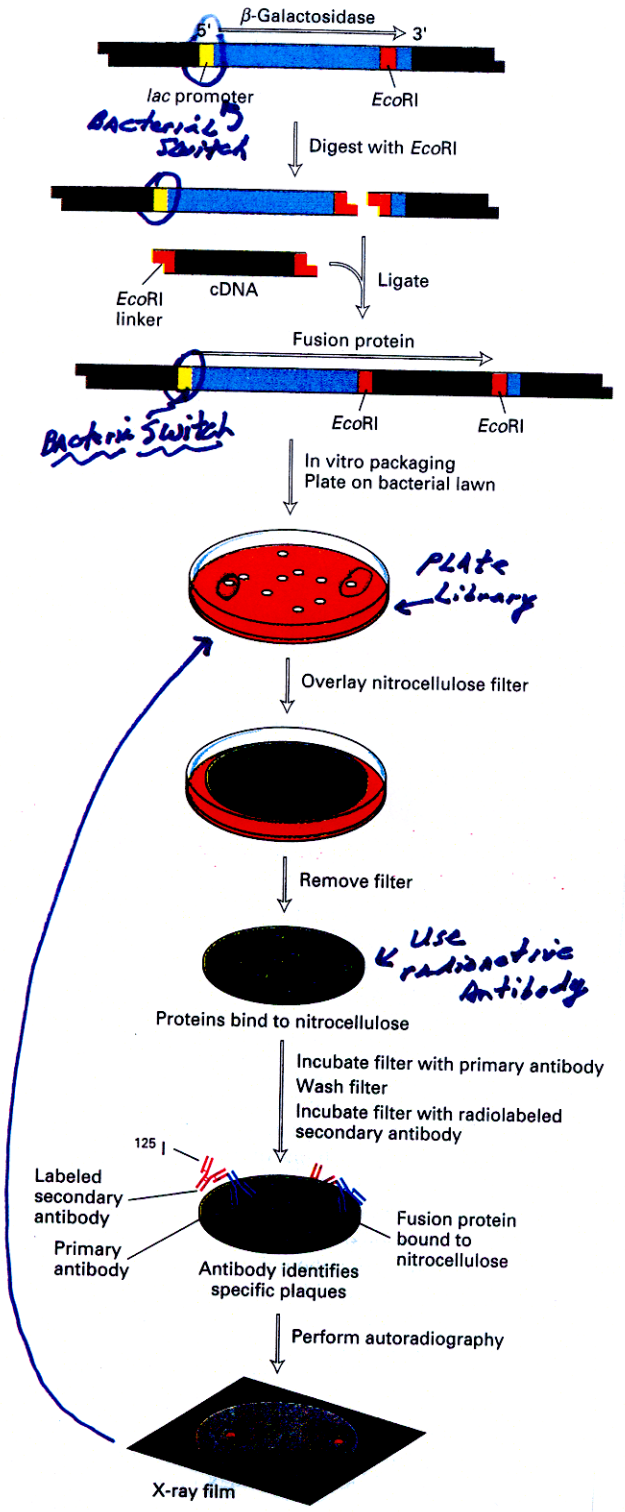


Figure 13-8 Two techniques for locating a gene. A. A hybridization probe locates a specific DNA sequence. B. Antibodies locate the protein product of the same sequence.

Identification of Insulin cDNA clone from pancreas mRNA Library

2004 → OR sequence 10,000's of Clones!
If protein sequence known can find cDNA

USING ANTIBODIES TO SELECT A SPECIFIC cDNA CLONE REQUIRES AN EXPRESSION VECTOR



Need to Purify Protein → Induce Antibodies to Protein in Rabbit

EXAMPLE!
INSULIN ANTIBODY
↓
PANCREAS cDNA Library

◀ **FIGURE 7-21 Use of λ expression cloning to identify a cloned DNA based on binding of the encoded protein to a specific antibody.** The λgt11 vector was engineered to express the *E. coli* protein β-galactosidase at high levels. The only *Eco*RI recognition site (red) in this vector lies near the 3′ end of the β-galactosidase gene. If a cDNA (green), or protein-coding fragment of genomic DNA, is inserted into this *Eco*RI site in the correct orientation and proper reading frame, it will be expressed as a fusion protein in which most of the β-galactosidase sequence is at the N-terminal end and the protein sequence encoded by the inserted DNA is at the C-terminal end. Plaques resulting from infection with recombinant λgtll contain high concentrations of such fusion proteins. These proteins can be transferred and bound to a replica filter, which then is incubated with a monoclonal primary antibody (blue) that recognizes the protein of interest. Rinsing the filter washes away antibody molecules that are not bound to the specific fusion protein attached to the filter. Bound antibody usually is detected by incubating the filter with a second radiolabeled antibody (dark red) that binds to the primary antibody. Any signals that appear on the autoradiogram are used to locate plaques on the master plate containing the gene of interest. [Adapted from J. D. Watson et al., 1992, *Recombinant DNA*, 2d ed., Scientific American Books.]

FOR FACTOR VIII - NOT KNOWN WHERE GENE IS EXPRESSED ∴ MUST USE GENOME LIBRARY

1980's

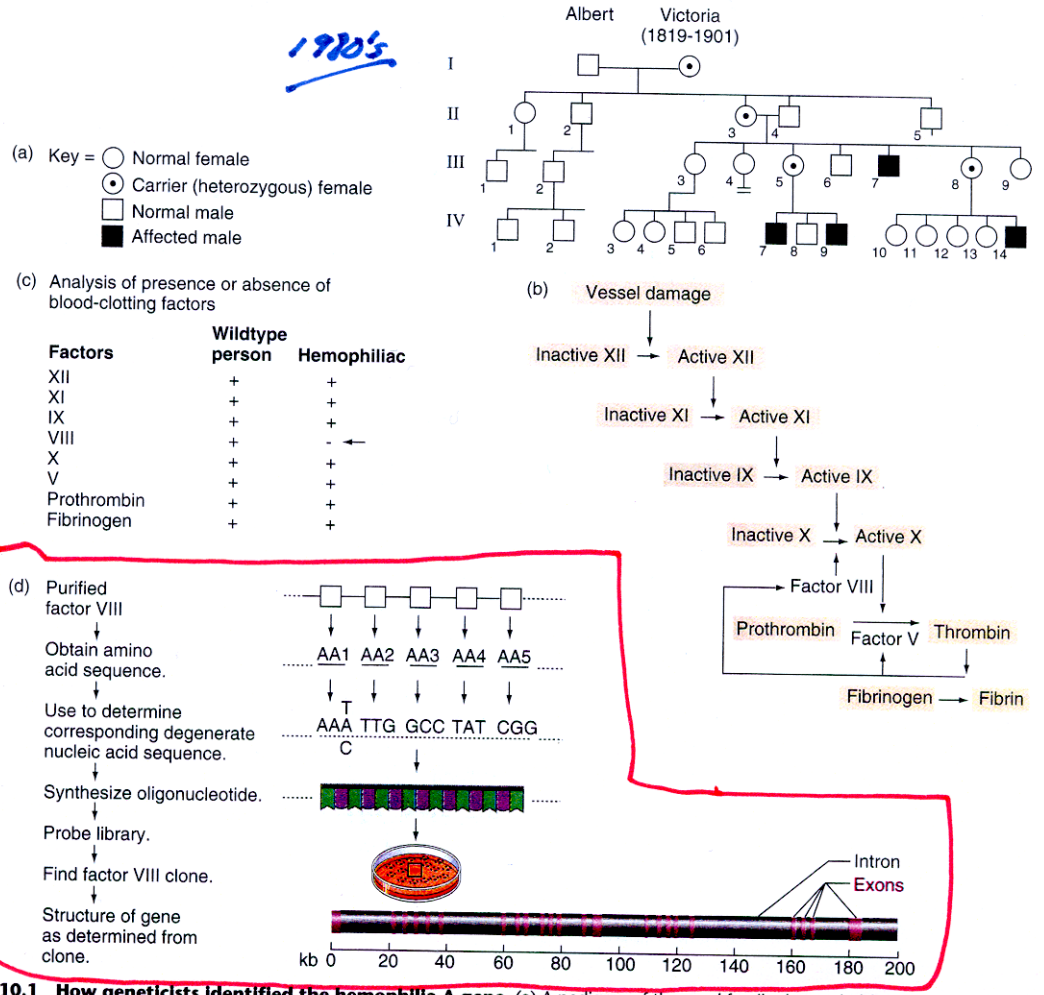


Figure 10.1 How geneticists identified the hemophilia A gene. (a) A pedigree of the royal family descended from Queen Victoria. This family tree uses the standard pedigree symbols. Black boxes represent males with hemophilia. (b) The blood clotting cascade. Vessel damage induces a cascade of enzymatic events that convert inactive factors to active factors. The cascade results in the transformation of fibrinogen to fibrin and the formation of a clot. (c) Many hemophiliac patients do not have an active form of Factor VIII. Blood tests can determine the presence or absence of the active form of each factor involved in the clotting cascade. The results of such analyses show that hemophiliacs, such as those found in Queen Victoria's pedigree, lack an active Factor VIII in their blood. (d) Starting with purified Factor VIII, scientists isolated DNA clones containing the *Factor VIII* gene. Researchers determined the amino-acid sequence of purified protein. Knowledge of this sequence enabled them to synthesize a degenerate oligonucleotide. They then used the oligonucleotide as a probe to screen a genomic library for clones containing all or parts of the gene. Finally, they sequenced the positive clones (that is, the clones with which the probe hybridizes) to determine the structure and coding sequence of the *Factor VIII* gene.

HOW find Gene & cDNA?

FACTOR VIII PROTEIN → GENE USING GENOME LIBRARY

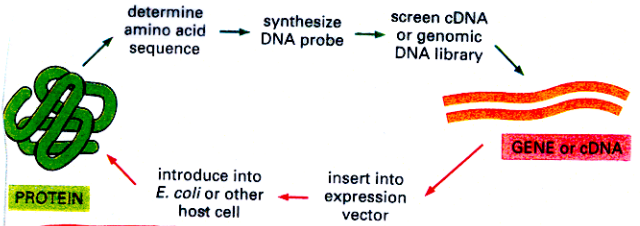


Figure 10–28 Knowledge of the molecular biology of cells makes it possible to experimentally move from gene to protein and from protein to gene. A small quantity of a purified protein is used to obtain a partial amino acid sequence. This provides sequence information that enables the corresponding gene to be cloned from a DNA library (see Figure 10–18). Once the gene has been cloned, its protein-coding sequence can be used to design a DNA that can then be used to produce large quantities of the protein from genetically engineered cells (see Figure 10–27).

FACTOR VIII

FROM PROTEIN TO GENE

Isolate protein on the basis of its molecular function (e.g., enzymatic or hormonal activity)

↓

Determine partial amino acid sequence of the protein

↓

Synthesize oligonucleotides that correspond to portions of the amino acid sequence

↓

Use oligonucleotides as probes to select cDNA or genomic clone encoding the protein from library

↓

Sequence isolated gene

FROM GENE TO PROTEIN

Isolate genomic clone corresponding to an altered trait in mutants (e.g., nutritional auxotrophy, inherited disease, developmental defect)

↓

Use genomic DNA to isolate a cDNA for the mRNA encoded by the gene

↓

Sequence the cDNA to deduce amino acid sequence of the encoded protein

↓

Compare deduced amino acid sequence with that of known proteins to gain insight into function of the protein

↓

Use expression vector to produce the encoded protein

* STOP

Pure Protein
↓
gene from Library

Gene Clone
↓
cDNA
↓
protein in expression vector

Gradually fill GenBank to identify by direct Sequencing

SCReening Library

Genome

1. Sequence → Data base
2. Probe from cDNA/Switch
3. Probe from pure mRNA
4. Synthetic Probe from translated DNA Sequence ↓ Genetic Code

cDNA

1. Sequence → Data base
2. pure mRNA probe
3. Synthetic probe from translated protein sequence / genetic code
4. Exon probe
5. Antibody probe using expression vector

STEPS REQUIRED TO CLONE FACTOR VIII GENE AND cDNA

① Make Genome Library Because Factor VIII GENE in genome!

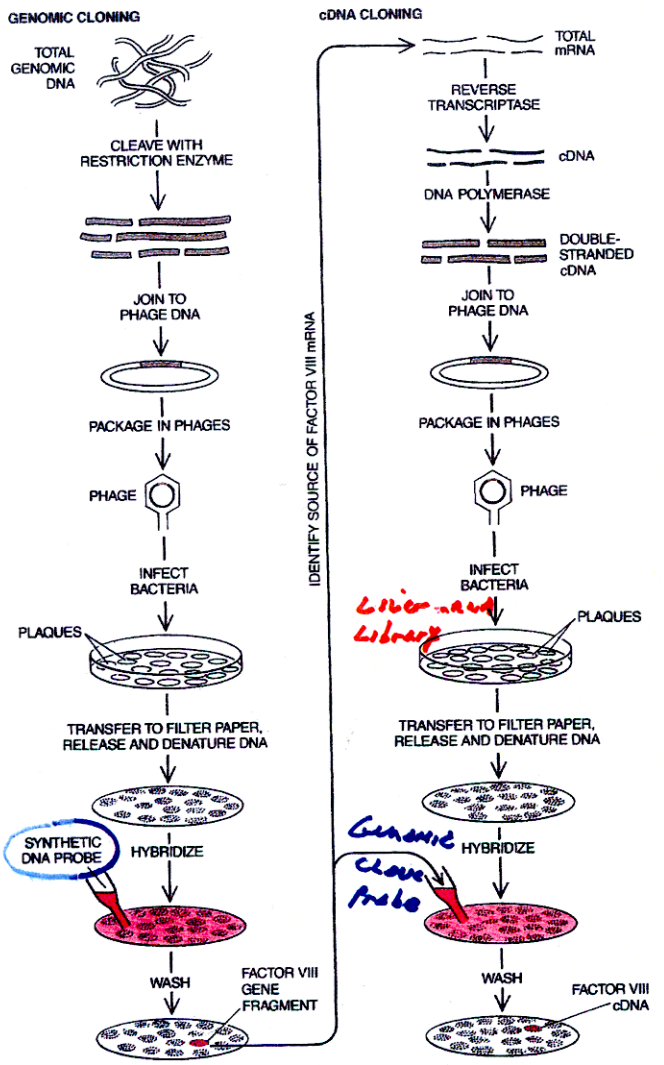
② Purify protein from Blood – That's where it works (wasn't know where made)

③ Reverse translate using the genetic code a portion of the protein sequence

④ Synthesize a DNA probe → complementary to Factor VIII gene corresponding to protein sequence

⑤ Screen Genome Library

Entire Gene in the clone?



GENE CLONING involves finding a specific gene among thousands in a human cell. The standard method, if one knows which cells make the desired protein, is to screen a copy DNA (cDNA) library derived by reverse transcription from the messenger RNA (mRNA) of those cells (*right*). In looking for the factor VIII gene, however, the authors did not know where the protein is produced. Hence they screened the entire human genome (*left*). Chromosomal DNA fragments were joined to the DNA of the bacterial virus phage lambda. Each phage contained one human DNA fragment; each phage multiplied and formed a plaque in a distinct region of a bacterial culture. To identify the plaque containing the factor VIII gene, the phages were blotted onto filter paper and broken open to release their DNA. The DNA was exposed to a radioactive probe: a small piece of synthetic DNA encoding part of factor VIII. The probe hybridized with part of the factor VIII gene, thereby labeling it. To produce factor VIII in cultured cells, it was still necessary to make factor VIII cDNA, which lacks the introns (noncoding sequences) that complicate the full gene. Now fragments of the cloned gene could serve as reliable probes, first for identifying cells that make factor VIII mRNA and then for finding factor VIII cDNA in the cDNA library.

① Use Gene probe to screen cDNA Library for Factor VIII cDNA clone

② How know what mRNA to use to make cDNA library?

③ Use gene probe to probe RNA Blots containing mRNA from all major organs (liver, kidney, blood, etc.) –

④ Find Factor VIII mRNA in liver – Made, Liver – Secreted into Blood

Why need cDNA?

STORY continued.

WANT cDNA to MANUFACTURE FACTOR VIII AS A DRUG TO TREAT Hemophilia A!

USING BACTERIOPHAGES AS VECTORS TO CLONE THE HUMAN GENOME

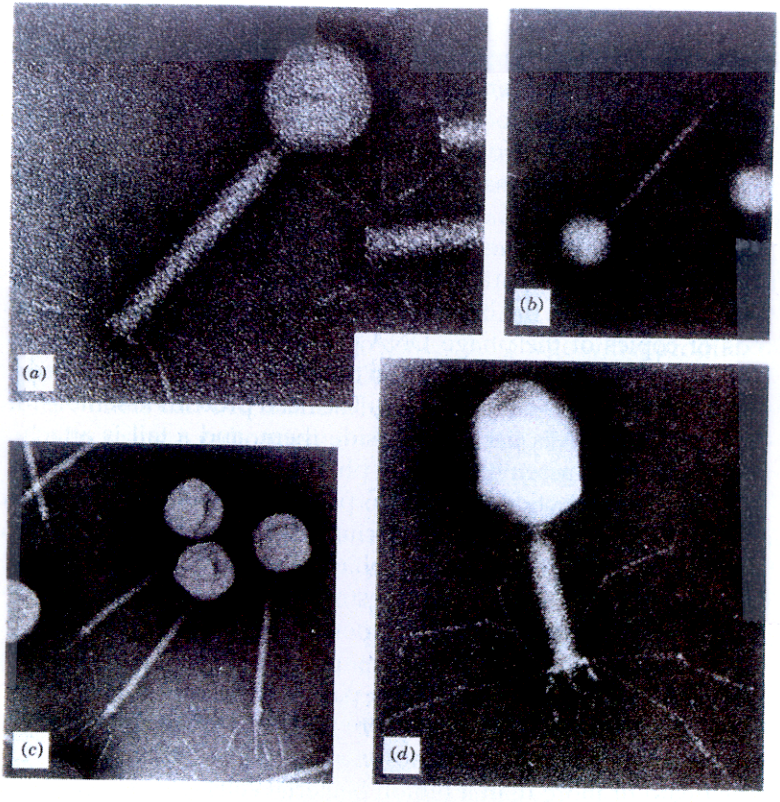
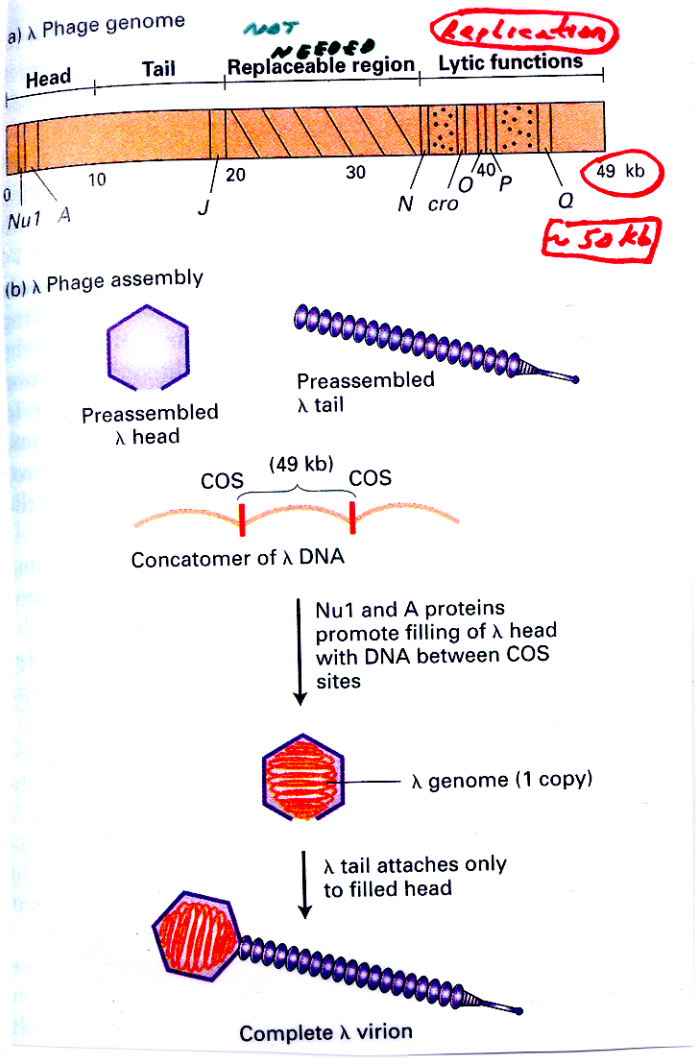


Figure 6-5. Electron Micrographs of Bacteriophages. **(a)** Bacteriophage P2, magnification 226,000 times. **(b)** Bacteriophage lambda, magnification 109,000 times. **(c)** Bacteriophage T5, magnification 91,000 times. **(d)** Bacteriophage T4, magnification 180,000 times. (Photomicrographs courtesy of Robley Williams, University of California, Berkeley.)

STRUCTURE OF THE λ PHAGE AND ITS GENOME



▲ FIGURE 9-14 The bacteriophage λ genome and packaging of bacteriophage λ DNA. (a) Simplified map of the λ phage genome. There are about 60 genes in the λ genome, only a few of which are shown in this diagram. Genes encoding proteins required for assembly of the head and tail are located at the left end; those encoding additional proteins required for the lytic cycle, at the right end. Some regions of the genome can be replaced by exogenous DNA (diagonal lines) or deleted (dotted) without affecting the ability of λ phage to infect host cells and assemble new virions. Up to ≈25 kb of exogenous DNA can be stably inserted between the *J* and *N* genes. (b) In vivo assembly of λ virions. Heads and tails are formed from multiple copies of several different λ proteins. During the late stage of λ infection, long DNA molecules called *concatomers* are formed; these multimeric molecules consist of multiple copies of the 49-kb λ genome linked end to end and separated by COS sites (red), protein-binding nucleotide sequences that occur once in each copy of the λ genome. Binding of λ head proteins Nu1 and A to COS sites promotes insertion of the DNA segment between two adjacent COS sites into an empty head. After the heads are filled with DNA, assembled λ tails are attached, producing complete λ virions capable of infecting *E. coli* cells.

λ Phage infects E. coli & destroys (lyses) cells

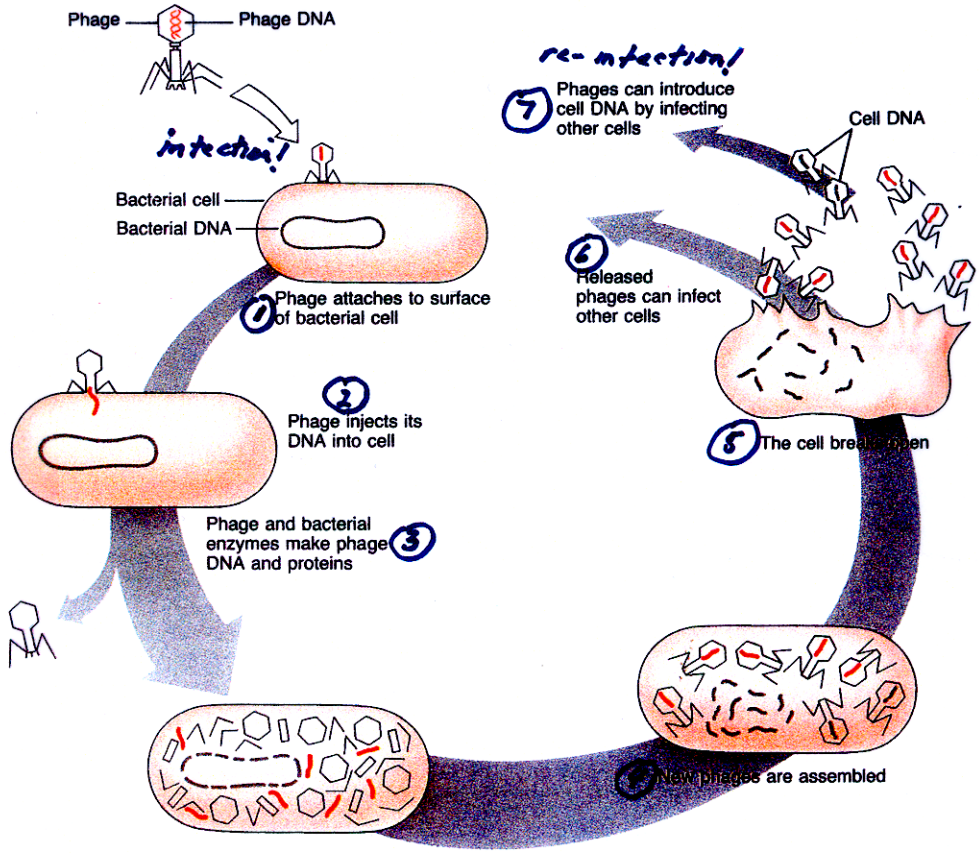
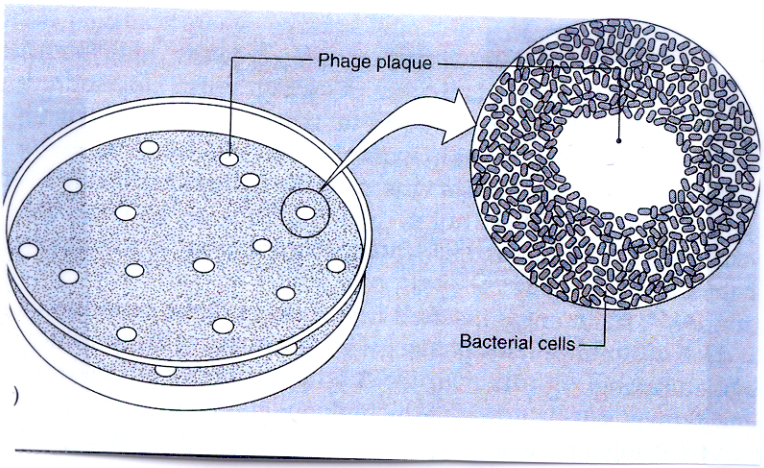
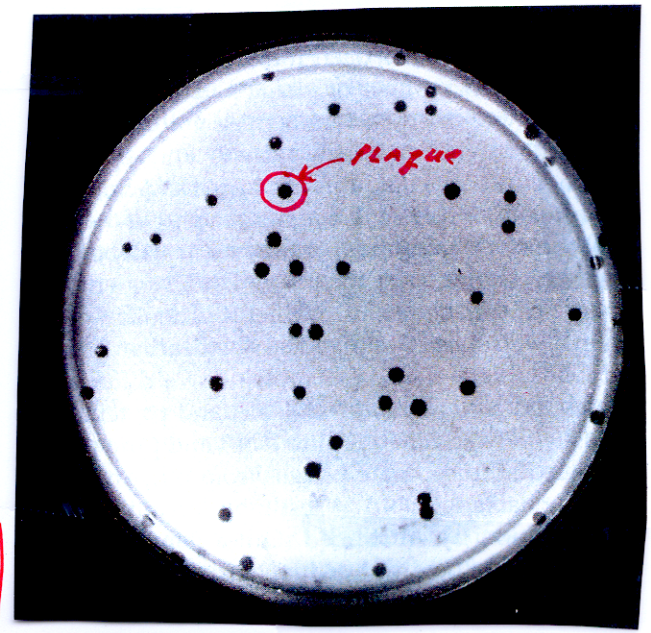


Figure 4.4 Events that occur when a phage infects a bacterial cell.

LYSED CELLS CAN BE SEEN AS CLEAR PLAQUES ON AGAR PLATES





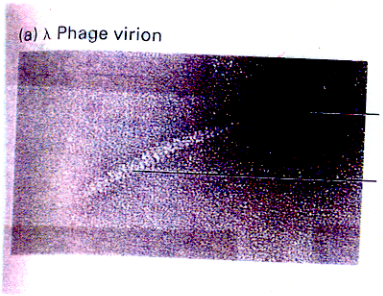
EACH PLAQUE IS A VIRUS CLONE REPRESENTING ONE VIRAL infection!

USING THE LAMBDA (λ) BACTERIA VIRUS AS A VECTOR (E. coli virus)

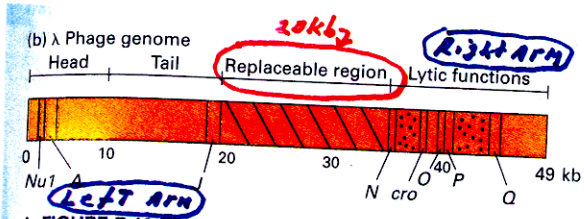
MAX Delbruck
Father of phage!

Advantages over PLASMIDS

1. Use natural Infection Process – much higher efficiency of getting DNA into bacteria cells :: more clones/μg DNA + Easier to use in Lab
2. CAN CLONE Long DNA Segments. Excellent for genome libraries. Need fewer clones for whole Genome!
3. CAN CLONE DNA in VIRUS genome + Self-Assemble virus (DNA + proteins) in test tube!!



VIRUS contains Protein + DNA
1. Head contains DNA
2. Tail binds bacteria cell + injects DNA genome inside



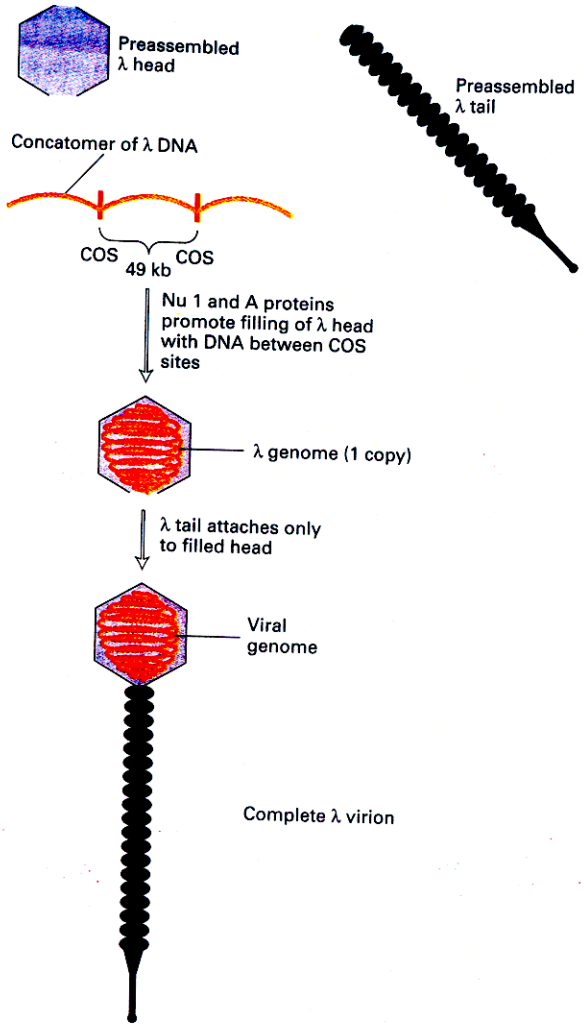
▲ **FIGURE 7-10 The bacteriophage genome.** (a) Electron micrograph of bacteriophage λ virion. The genome is contained within the head. (b) Simplified map of the λ phage genome. Genes encoding proteins required for assembly of the head and tail map at the left end; those encoding additional proteins required for the lytic cycle map at the right end. Some regions of the genome can be replaced by exogenous DNA (diagonal lines) or deleted (dotted area) without affecting the ability of λ phage to infect host cells and assemble new virions, permitting insertion of up to ≈25 kb of exogenous DNA between the *J* and *N* genes. There are about 60 genes on the λ genome. Only a few individual genes are shown in this diagram. Small numbers indicate positions in kilobases (kb). [Photograph courtesy of R. Duda and R. Hendrix.]

One of First Genomes to be Sequenced!

λ Genome
1. 50,000 bp
2. Genes Needed for Replication on Right side or "ARM"
3. Genes Needed for Proteins – Head + Tail on Left "Arm"
4. Genes in middle not needed & can be replaced with foreign DNA! 20kb

Note: Restriction Enzymes "Fight" Viral infection!

λ VIRUS CAN BE SELF-ASSEMBLED IN A TEST TUBE



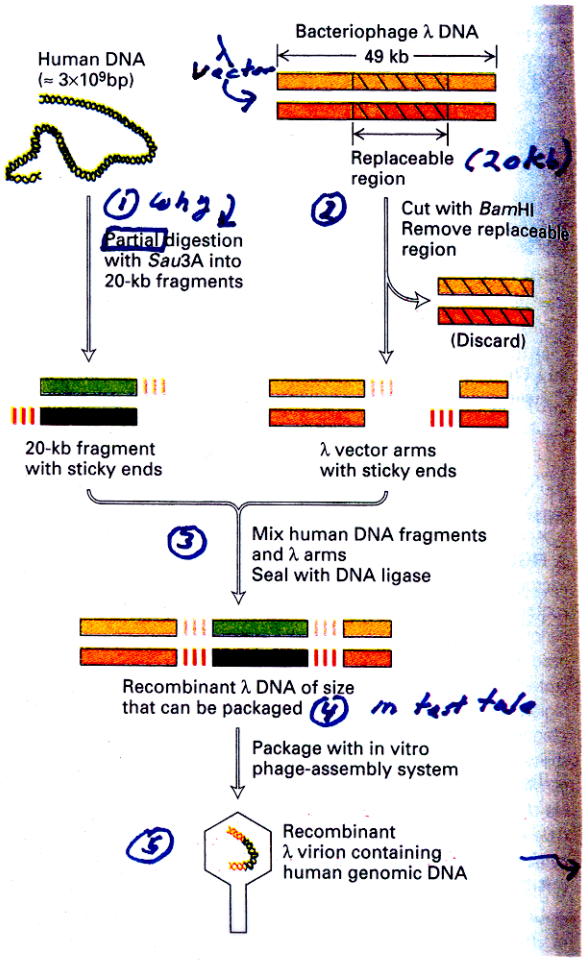
ALSO PACKAGED NATURALLY IN E. coli CELLS

▲ **FIGURE 7-11 Assembly of bacteriophage λ virions.** Empty heads and tails are assembled from multiple copies of several different λ proteins. During the late stage of λ infection, long DNA molecules called *concatomers* are formed; these multimeric molecules consist of copies of the λ genome linked end to end and separated by COS sites (red), a protein-binding nucleotide sequence that occurs once in each copy of the λ genome. Binding of the λ proteins Nu1 and A to COS sites promotes insertion of the DNA between two adjacent COS sites into an empty head. After the heads are filled with DNA, preassembled λ tails are attached, producing complete λ virions capable of infecting *E. coli* cells.

USING λ VIRUSES AS A VECTOR TO CLONE HUMAN GENOME

A COMPLETE HUMAN GENOME LIBRARY in λ phage vector

1. Large DNA fragments
2. Keep Genes intact
3. Keep Clones in Library as Few as possible to Find Gene – Easier to hunt thru 60,000 clones than 5,000,000!!!



▲ **FIGURE 7-12 Construction of a genomic library of human DNA in a bacteriophage λ vector.** The nonessential regions in the right half of the λ genome (dotted areas in Figure 7-10b) usually are deleted to maximize the size of the exogenous DNA fragment that can be inserted. Then the λ DNA is treated to remove the central replaceable region. In this example, the replaceable region is cut out with *Bam*HI, and the total DNA from human cells is partially digested with *Sau*3A. These two restriction enzymes produce fragments with complementary sticky ends (red lines). The λ vector arms and ≈20-kb genomic fragments are mixed, ligated, and packaged in vitro to produce recombinant λ phage virions, which are plated on a lawn of *E. coli* cells. In the diagrams of DNA regions, light and dark shades of the same color indicate complementary strands.

CLONING THE HUMAN GENOME AND SCREENING FOR THE FACTOR VIII GENE

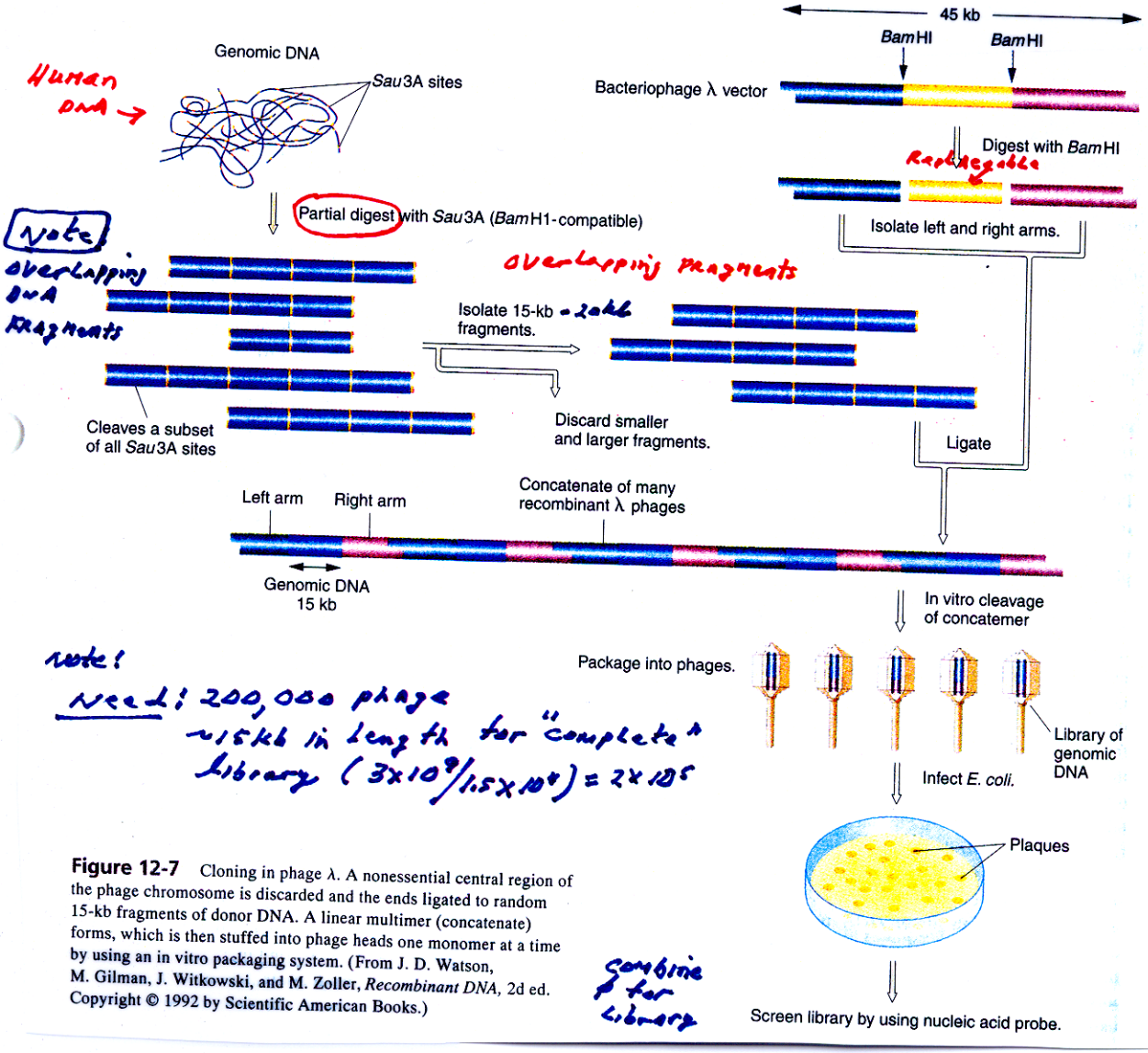


Figure 12-7 Cloning in phage λ. A nonessential central region of the phage chromosome is discarded and the ends ligated to random 15-kb fragments of donor DNA. A linear multimer (concatenate) forms, which is then stuffed into phage heads one monomer at a time by using an in vitro packaging system. (From J. D. Watson, M. Gilman, J. Witkowski, and M. Zoller, *Recombinant DNA*, 2d ed. Copyright © 1992 by Scientific American Books.)

What is the purpose of Partial Digestion of HUMAN DNA?

Sau 3A = 4bp = 5' GATC 3' ∴ 1 site every 280bp if digest to completion ≈ 1×10^7 DNA fragments

EcoRI = 6bp = 5' GAATTC 3' ∴ 1 site every 3100bp if digest to completion (cleave every site) ≈ 972,000 DNA fragments!

∴

1. Complete Digestion produces fragments that are too small to clone in λ virus (need 20kb)
2. Complete digestion would create huge genome libraries with large # clones to screen
3. Complete digestion would break up genes on different DNA fragments – particularly if human genes big – ∴ would have one gene on many different clones – parts separated!
4. Complete digestion provides no way to find neighbors of clones in genome – what's next to gene in chromosome!

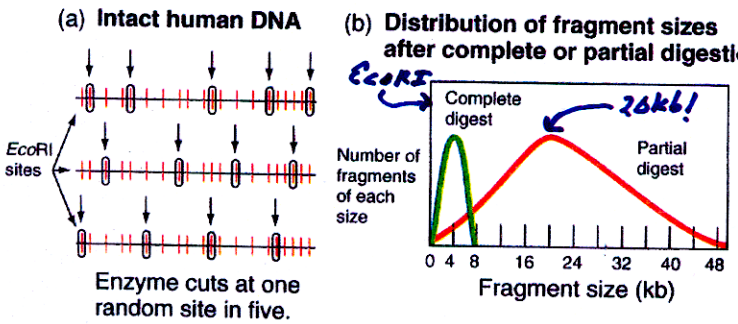


Figure 8.4 Comparison of results from partial and complete digests. (a) By reducing the time available for the reaction to occur, you can ensure that an enzyme actually cuts only a subset of the total recognition sites within a DNA sample. In this example, the chosen reaction time allowed only 1/5 of all *Eco*RI sites to be cut. The particular 20% of sites at which the cuts occur is totally random and different even on identical DNA molecules. (b) Most of the restriction fragments produced by partial digestion are larger than those produced by complete digestion with the same restriction enzyme.

PARTIAL DIGESTION PRODUCES A SERIES OF LARGE, OVERLAPPING DNA FRAGMENTS/CLONES!

CAN CONNECT ONE CLONE WITH ANOTHER!! BUILD UP CLONES OF EACH CHROMOSOME!!