HCTOA Winter 2006

Professor Bob Goldberg

Lecture #4 - Nuts & Bolts of Genetic Engineering: The Factor VIII Story - From Disease to Gene to Drug - Genetic Engineering in "Action"

THEMES

1. Hemophilia - An inherited "inborn" error!
2. Finding Genes & cDNAs
3. Recombinant DNA Tools - A Review
4. Restriction Enzymes & Maps - The tool of the Gene Engineer
5. Genome vs. cDNA Libraries
6. Making a Genome Library - Overlapping Clones
7. Finding the Factor VIII Gene - Why Gene?
8. Finishing Clones for the Entire Gene - Chromosome Walking
9. DNA Testing - Using Factor VIII Probes to find Carriers - RFLP concept reviewed
10. Hunting for the Factor VIII cDNA - RT-PCR
11. Making a Factor VIII Drug!!

Read Chps 8 & 11 / Textbook

APPLICATIONS OF GENETIC ENGINEERING TECHNOLOGY

3.4 WHAT CAN YOU DO WITH A CLONED GENE? APPLICATIONS OF RECOMBINANT DNA TECHNOLOGY

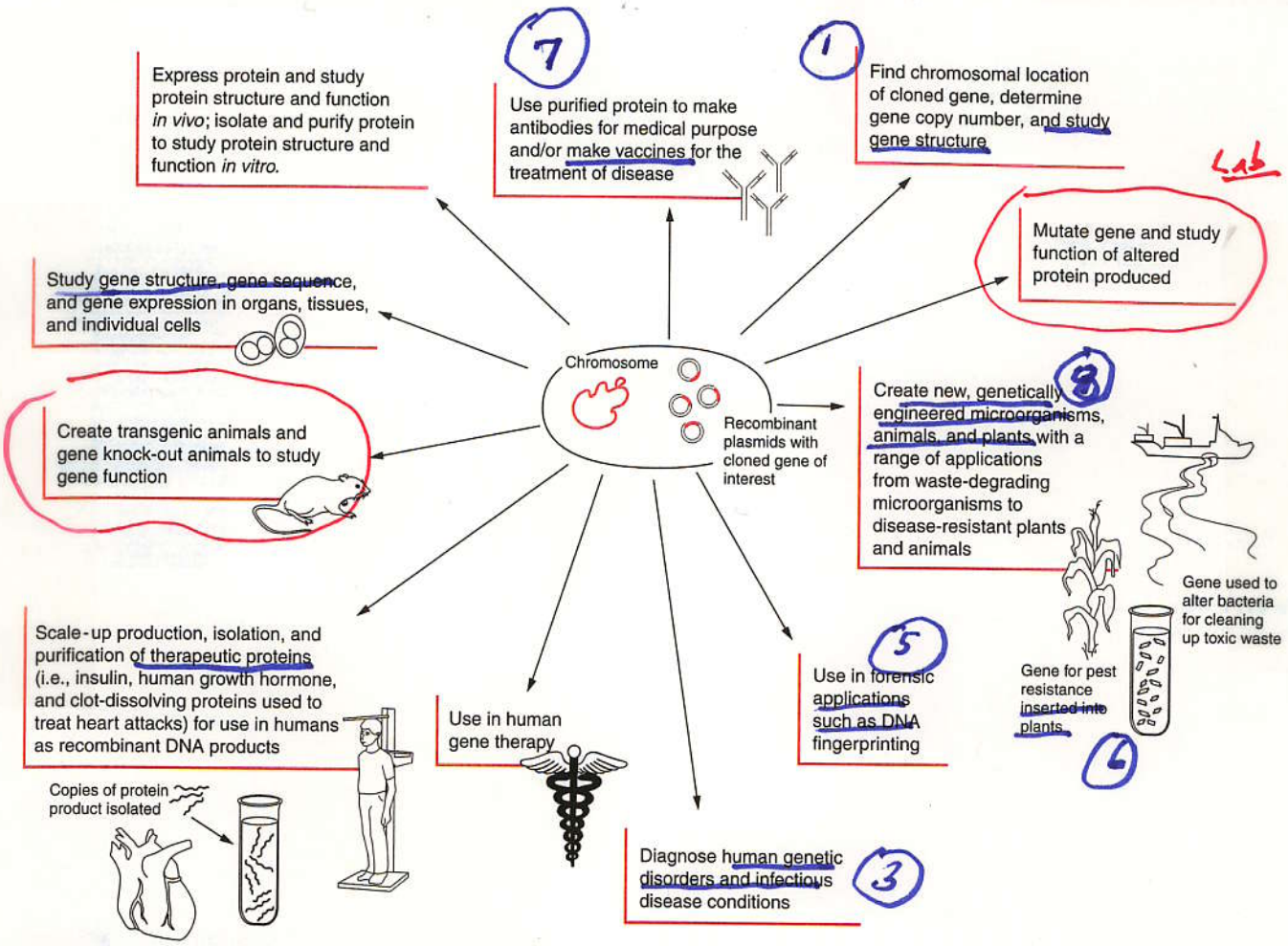
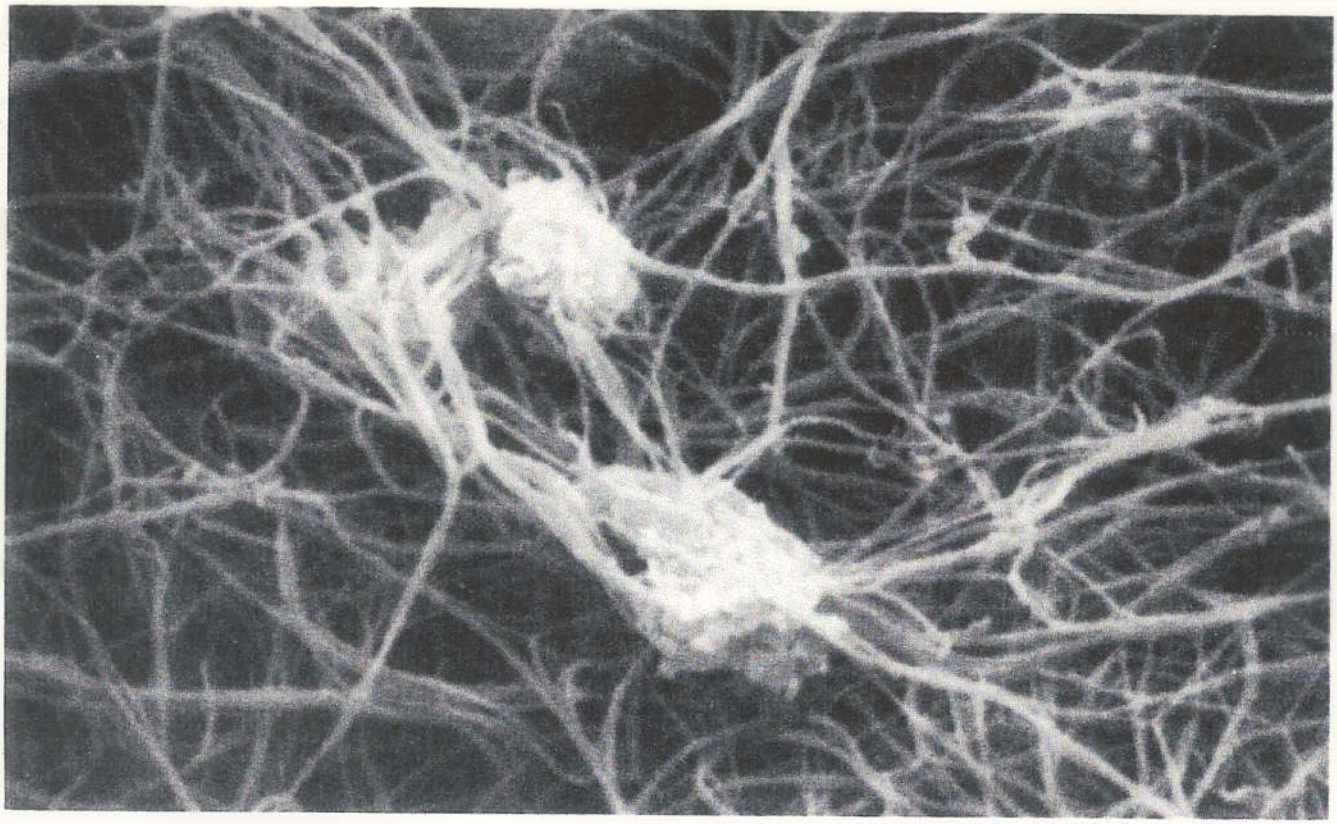


Figure 3.10 Applications of Recombinant DNA Technology

Gene → Drug

FACTOR VIII!

The Molecular Genetics of Hemophilia



FIBRIN STRANDS stabilize a blood clot at the site of a wound by trapping the platelets that form the bulk of the clot. The electron micrograph, which was made by Jon C. Lewis of Wake Forest University, shows a clot formed in a suspension of platelets and fibrin.

A clot in the bloodstream is the result of a complex cascade of enzymatic reactions culminating in the conversion of fibrinogen, a soluble protein, into insoluble fibrin strands. In hemophiliacs a crucial protein in the blood-clotting cascade is either missing or defective.

A case study of cloning genes and mRNAs

http://ehrweb.aaas.org/ehr/books/yourgenes.pdf

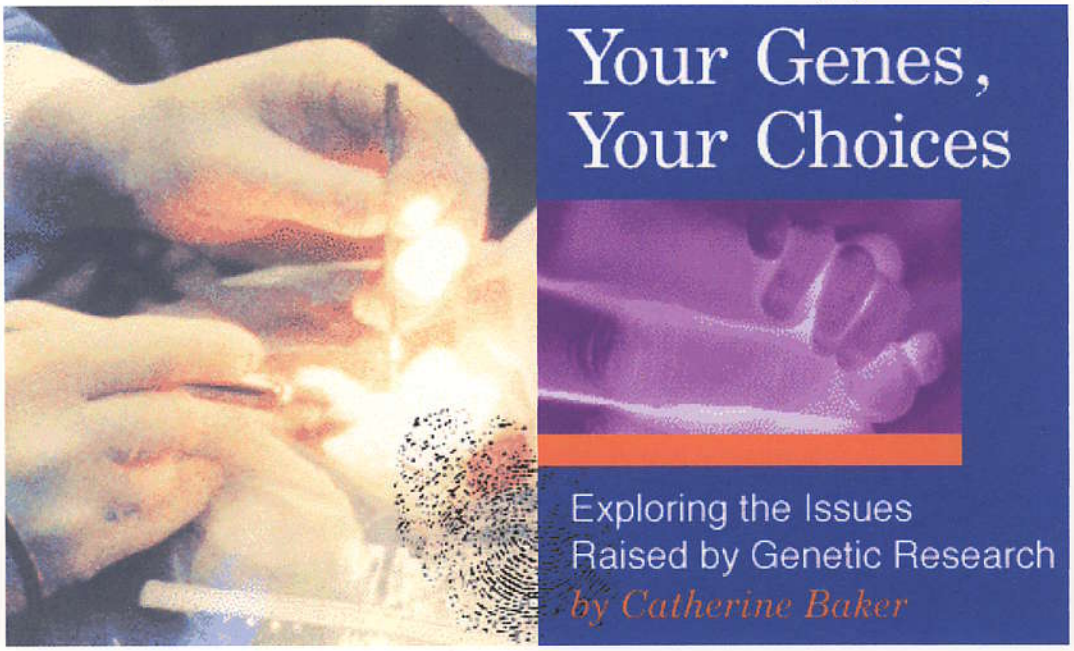


Table of Contents

Your Genes, Your Choices describes the Human Genome Project, the science behind it, and the ethical, legal, and social issues that are raised by the project. This book was written as part of the Science + Literacy for Health project of the American Association for the Advancement of Science (AAAS) and funded by the U.S. Department of Energy.

AAAS has a strong commitment to science literacy and the public understanding of science. Through its Directorate for Education and Human Resources Programs, AAAS has been a leader in identifying and meeting the needs of underrepresented groups in science. Science + Literacy for Health fits into this vision of making science accessible to everyone.

Most people think that science is remote from the work they do, the lives they lead, and the decisions that they make day by day. Nothing could be further from the truth. *Your Genes, Your Choices* points out how the progress of science can potentially "invade" your life in the most direct ways, affecting the choices you make at the grocery store, your own health care and that of your family, and even your reproductive decisions. The connection between science and health is a direct one, and your ability to understand the science behind health affects your ability to understand the issues and the stakes.

Science may seem difficult, because scientists often use technical language to talk about abstract ideas. This book has been written to introduce you to important ideas, but also to convince you that you can understand the basic concepts of science and that it is important to do so.

Most people are curious about the way their bodies work (and the ways they sometimes don't work very well). This curiosity goes beyond immediate concerns about any specific health condition. We hope that *Your Genes, Your Choices* helps to feed that interest.

HEMOPHILIA HAS BEEN KNOWN AS AN INHERITED DISEASE FOR > 2500 Years!

HUMAN GENETICS SIDELIGHT

Hemophilia: Successful Treatment of a Once Deadly Disorder

A small defect in an important gene can cause a fatal human disease. In the past, hemophilia, excess bleeding caused by a defect in blood clotting, was such a disease—often fatal early in life. Before the 1960s, when scientist-physicians developed the first effective treatment, the life expectancy of individuals with hemophilia was about 20 years. Today, hemophiliacs in most of the world have a nearly normal life expectancy. An understanding of the molecular basis of the disease resulted in the development of an effective treatment.

There are two major types of hemophilia. About 80 percent of the individuals with this disease have hemophilia A (classical hemophilia), and about 20 percent have hemophilia B (also called Christmas disease because it was first detected in a patient named Stephen Christmas). Both types of hemophilia are caused by defective genes on the X chromosome, the human chromosome that is present in two copies in females and one copy in males (Chapter 6). Most hemophiliacs are males, because they only need one copy of the defective gene to have the disease. Hemophilia is rare in females, because they need two copies of the defective gene, one on each X chromosome, to have the disorder.

Hemophilia A is sometimes called "royal hemophilia" because of its prevalence in the royal families of Europe. England's Queen Victoria (Figure 1) did not have hemophilia, although she carried the defective gene that causes hemophilia A on one of her X chromosomes. However, she passed the defective gene to two of her daughters—Alice, who transmitted the gene to the imperial families of Russia (see Figure 6.9) and Germany, and Beatrice, who passed the gene to the royal family of Spain—and to her son Prince Leopold, who died at age 31 from hemorrhages after a fall. Several of the queen's grandsons and great-grandsons died early in life because of excess bleeding or hemorrhages after surgery or accidents.

The mode of transmission of hemophilia was probably recognized in ancient civilizations. The Jewish Talmud, which dates to about 400 B.C. and was compiled into a single document in the 4th and 5th centuries A.D., decreed that boys whose older brothers or male cousins had died from excessive bleeding after circumcision were exempt from this procedure.

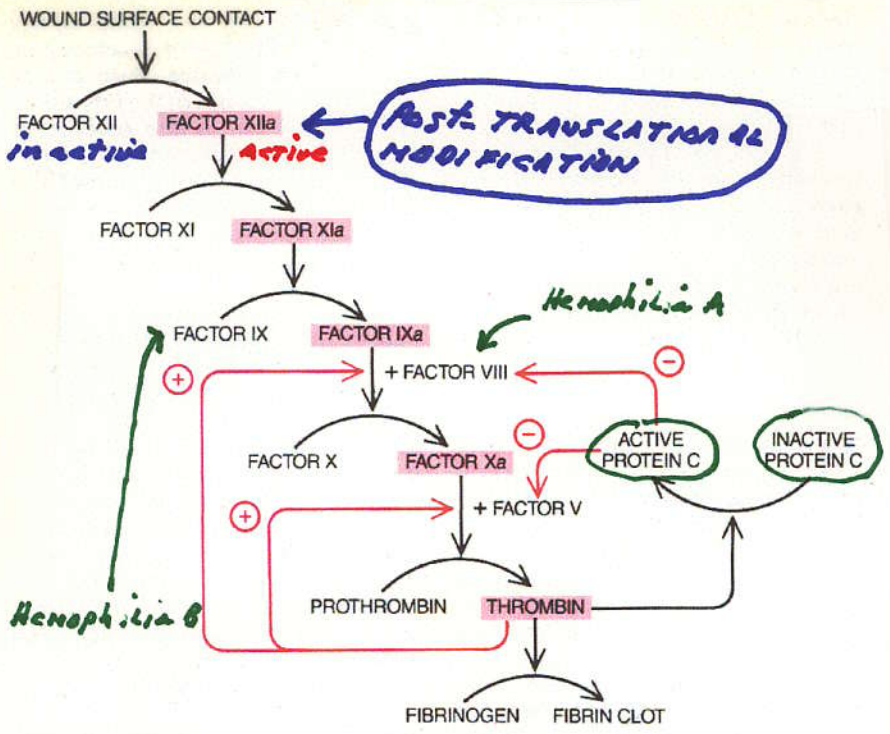
Bible

Hemophilia A and hemophilia B both result from defects in blood coagulation—the cascade of reactions that causes blood to clot at the site of a wound. A simplified version of part of this pathway is shown in Figure 2. Individuals with hemophilia A are deficient in a gene product called factor VIII; those with hemophilia B are lacking factor IX. In the absence of either of these blood-clotting factors, an individual can bleed to death after suffering a small cut or can die from internal hemorrhages after an otherwise minor bruise.

When scientists discovered that hemophilia was caused by the absence of specific blood-clotting factors, they realized that the disease could be treated with transfusions of concentrates of the missing factor. Initially, beginning in the 1960s, the proteins were purified from blood obtained from large numbers of donors. This process was expensive, and the concentrates were either unavailable or were too expensive for use by hemophiliacs in many countries. Fortunately, the advent of genetic engineering brought positive changes. The genes that encode factor VIII and factor IX were both isolated, and each gene was introduced into mammalian cells growing in culture. By this procedure, cell culture lines were produced that synthesize large quantities of either factor VIII or factor IX. The clotting factors are now purified from these cells and used to prepare concentrates for use in transfusions. As a result, both clotting factors are now available in essentially unlimited quantity to treat people suffering from hemophilia.

Figure 1 A portrait of Great Britain's Queen Victoria, her husband Prince Albert, and five of their nine children. Queen Victoria passed the defective gene that is responsible for hemophilia to at least three of her children. They, in turn, passed the gene to the royal families of Germany, Russia, and Spain (see Figure 6.9). The present British royal family is free of hemophilia. They are descendants of Victoria's son King Edward VII, who did not inherit the hemophilia gene from his mother.

HOW DOES BLOOD CLOT AFTER WOUNDING?



CLOTTING CASCADE begins when cell damage at a wound somehow activates the enzyme factor XII; it ends with the conversion of fibrinogen into fibrin by thrombin. At each step an inactive protein is converted into a protease, or protein-cutting enzyme (*color*), which activates the next protein. Some steps require cofactors such as factors VIII and V. The cascade includes positive- and negative-feedback loops (*colored arrows*). Thrombin activates factors VIII and V; it also deactivates them (by activating protein *C*), which helps to halt clotting. Some 85 percent of hemophiliacs lack factor VIII. The rest lack factor IX.

Post translation modification

Eight Proteins/Genes required

1. Factor VII
2. Factor XI
3. Factor IX
4. Factor VIII
5. Factor X
6. Protein C
7. Prothrombin
8. Fibrinogen

What happens if any of these proteins or genes are mutated?

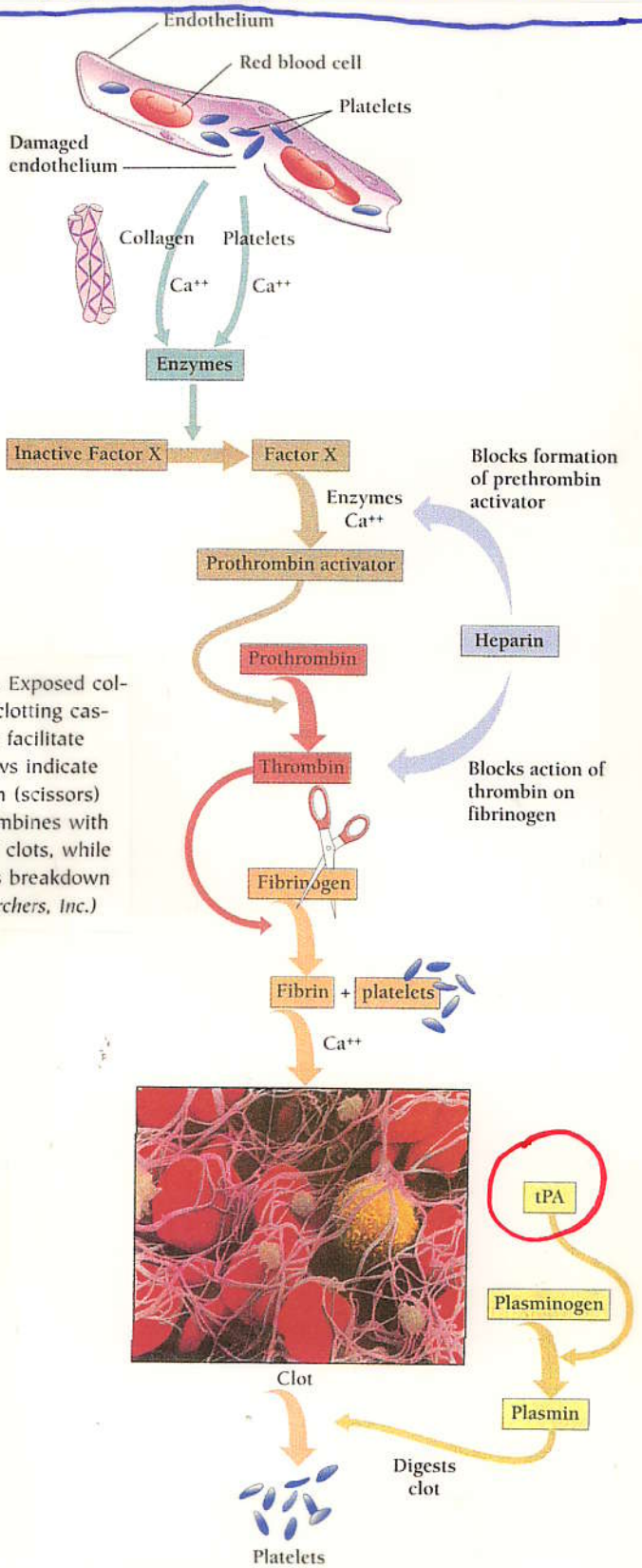
↓

NO BLOOD CLOT!

Anti-Thrombin ??

CASCADE

t PA OR TISSUE PLASMINOGEN ACTIVATOR DISSOLVES CLOTS & IS AN IMPORTANT DRUG TO COUNTER HEART ATTACKS!



ure 40-5 Making and unmaking blood clots. Exposed col-gen or blood platelets trigger the first steps in the "clotting cascade." Thin arrows indicate the work of enzymes that facilitate transformation of one molecule into another. Fat arrows indicate the transformation. For example, the enzyme thrombin (scissors) cuts off a piece of fibrinogen, leaving fibrin, which combines with platelets to form a clot. Heparin prevents formation of clots, while the enzyme tPA (tissue plasminogen activator) triggers breakdown of clots. *(Photo, CNRI/Science Photo Library/Photo Researchers, Inc.)*

t PA IS A Biotech DRUG!

Hemophiliacs Have Mutations in Either Factor VIII or Factor IX Genes

Table 13.2 Some Important Genetic Disorders

Disorder	Symptom	Defect	Dominant/ Recessive	Frequency among Human Births
Cystic fibrosis	Mucus clogs lungs, liver, and pancreas	Failure of chloride ion transport mechanism	Recessive	1/2500 (Caucasians)
Sickle cell anemia	Poor blood circulation	Abnormal hemoglobin molecules	Recessive	1/625 (African Americans)
Tay-Sachs disease	Deterioration of central nervous system in infancy	Defective enzyme (hexosaminidase A)	Recessive	1/3500 (Ashkenazi Jews)
Phenylketonuria	Brain fails to develop in infancy	Defective enzyme (phenylalanine hydroxylase)	Recessive	1/12,000
Hemophilia	Blood fails to clot	Defective blood clotting factor VIII	Sex-linked recessive	1/10,000 (Caucasian males)
Huntington's disease	Brain tissue gradually deteriorates in middle age	Production of an inhibitor of brain cell metabolism	Dominant	1/24,000
Muscular dystrophy (Duchenne)	Muscles waste away	Degradation of myelin coating of nerves stimulating muscles	Sex-linked recessive	1/3700 (males)
Hypercholesterolemia	Excessive cholesterol levels in blood, leading to heart disease	Abnormal form of cholesterol cell surface receptor	Dominant	1/500

Hemophilia A — Defective Factor VIII Gene — 1/10,000 Males

Hemophilia B — Defective Factor IX Gene — 1/30,000 Males

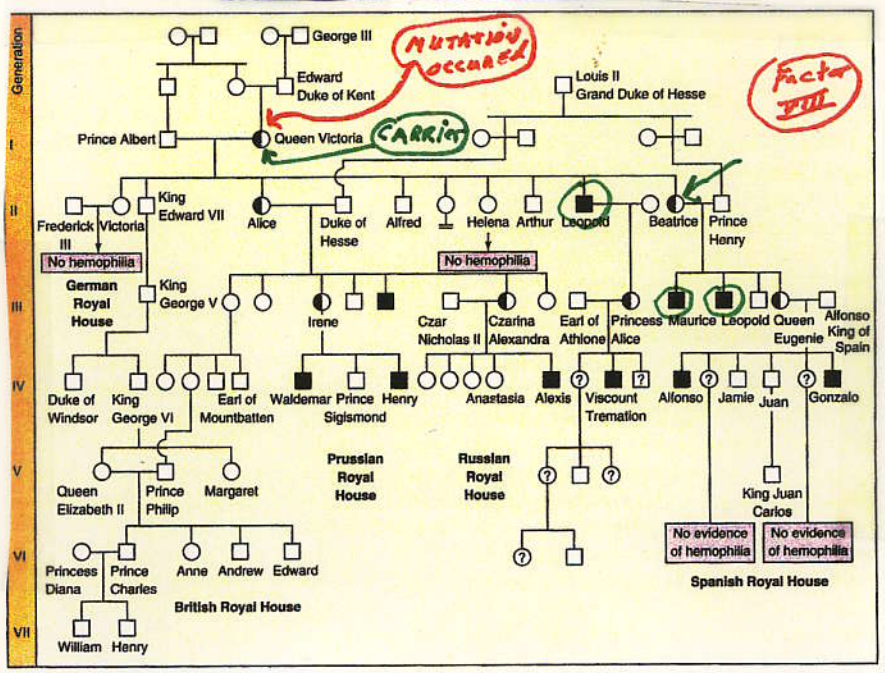
Hypothesis For High Frequency?

Both genes on X-Chromosome

♀ → ♂'s

Hemophilia A & B ARE Sex-Linked GENES

ROYAL HEMOPHILIA Gene

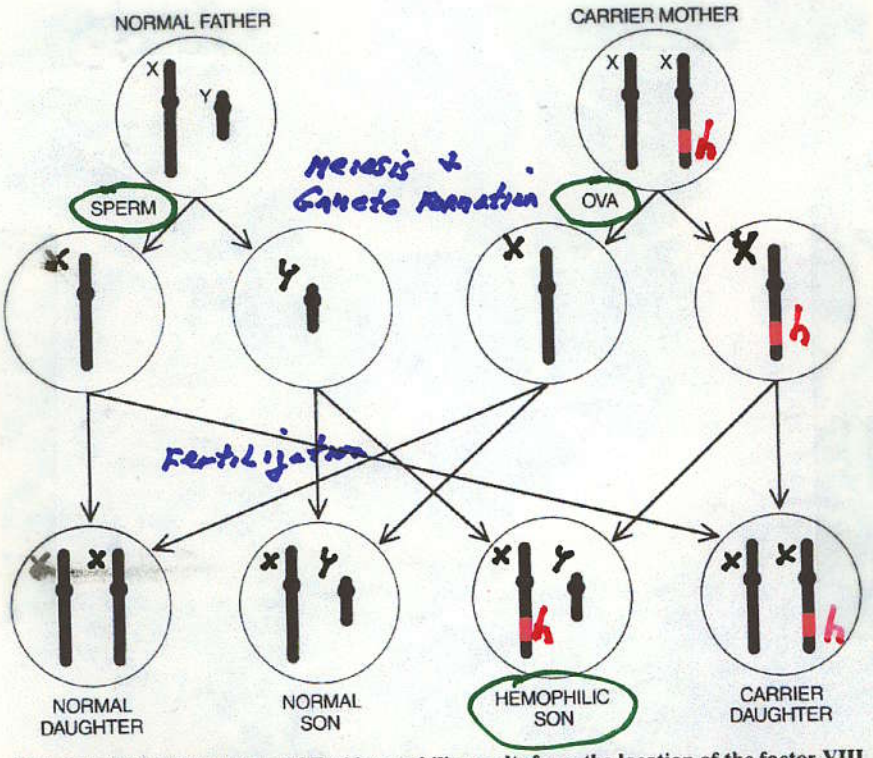


NOTE: ♀ obtain defective Allele from ♀

FIGURE 13.26
The Royal hemophilia pedigree. Queen Victoria's daughter Alice introduced hemophilia into the Russian and Austrian royal houses, and Victoria's daughter Beatrice introduced it into the Spanish royal house. Victoria's son Leopold, himself a victim, also transmitted the disorder in a third line of descent. Half-shaded symbols represent carriers with one normal allele and one defective allele; fully shaded symbols represent affected individuals.

Genes passed on from Mother "CARRIERS" to SONS

HEMOPHILIA A and B Inheritance



SEX-LINKED INHERITANCE of hemophilia results from the location of the factor VIII gene on the X chromosome. A male carrying a mutant factor VIII gene lacks normal factor VIII and is hemophilic. A female carrier is protected by the normal gene on her second X chromosome, but half of her daughters will be carriers and half of her sons will be hemophilic. In the case of a hemophilic father (not shown), his sons will not be hemophilic, because they receive his Y (not his X) chromosome, but his daughters will be carriers.

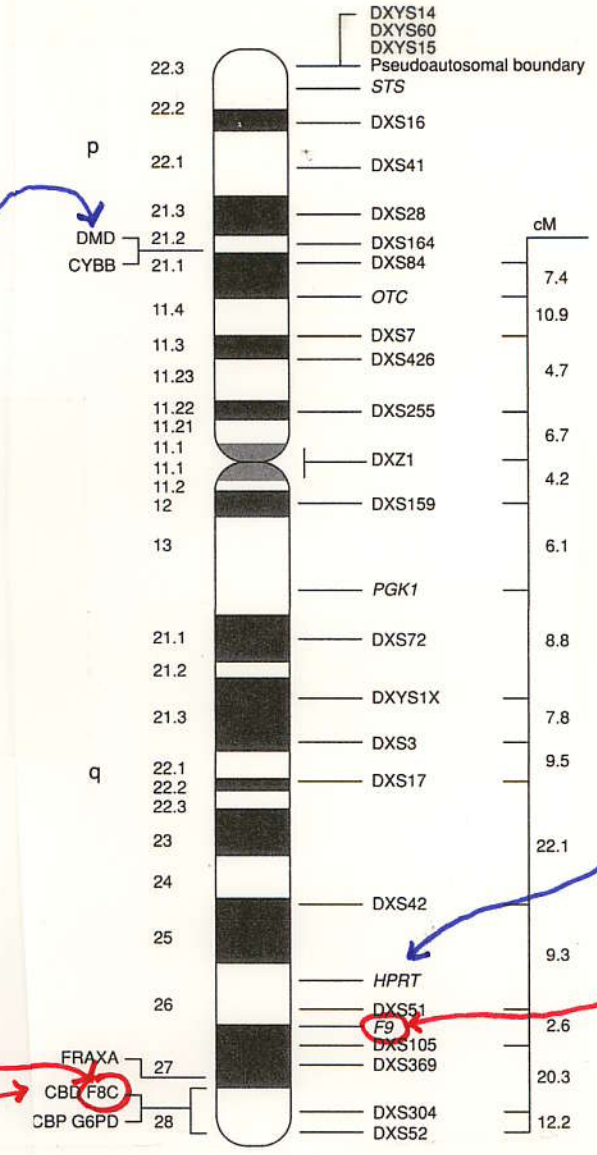
PARENTS 2n
↓ MEIOSIS
GAMETES n
↓ Fertilization
CHILDREN 2n

SEX-LINKED INHERITANCE

♀ CARRIERS → ½ SONS & NO DAUGHTERS!

ONLY ONE X-CHROMOSOME in ♂

FACTOR VIII and FACTOR IX GENES ARE ON THE X-CHROMOSOME



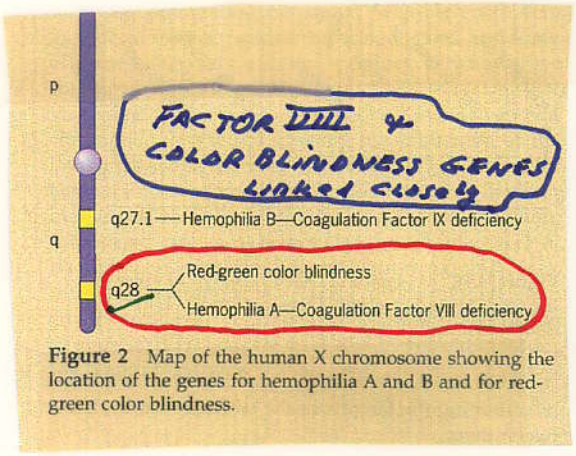


Figure 2 Map of the human X chromosome showing the location of the genes for hemophilia A and B and for red-green color blindness.

MUSCULAR DISTROPHY

NOTE: BANDING PATTERN OF CHROMOSOME

BANDS = MARKERS of CHROMOSOME POSITIONS & ARE CHROMOSOME SPECIFIC?

LESCH-NYHAN SYNDROME

FACTOR IX

FACTOR VIII

HYPOTHESIS FOR BAND SPECIFICITY?

COLOR BLINDNESS

MARKER for

HOW CORRELATE BAND & GENE?

HOW DO BANDS FORM & SHOWN?

CHROMOSOME – DNA SEQUENCE – GENE ORDER

FROM DISEASE TO GENE - USING PROTEIN TO IDENTIFY FACTOR EIGHT GENE

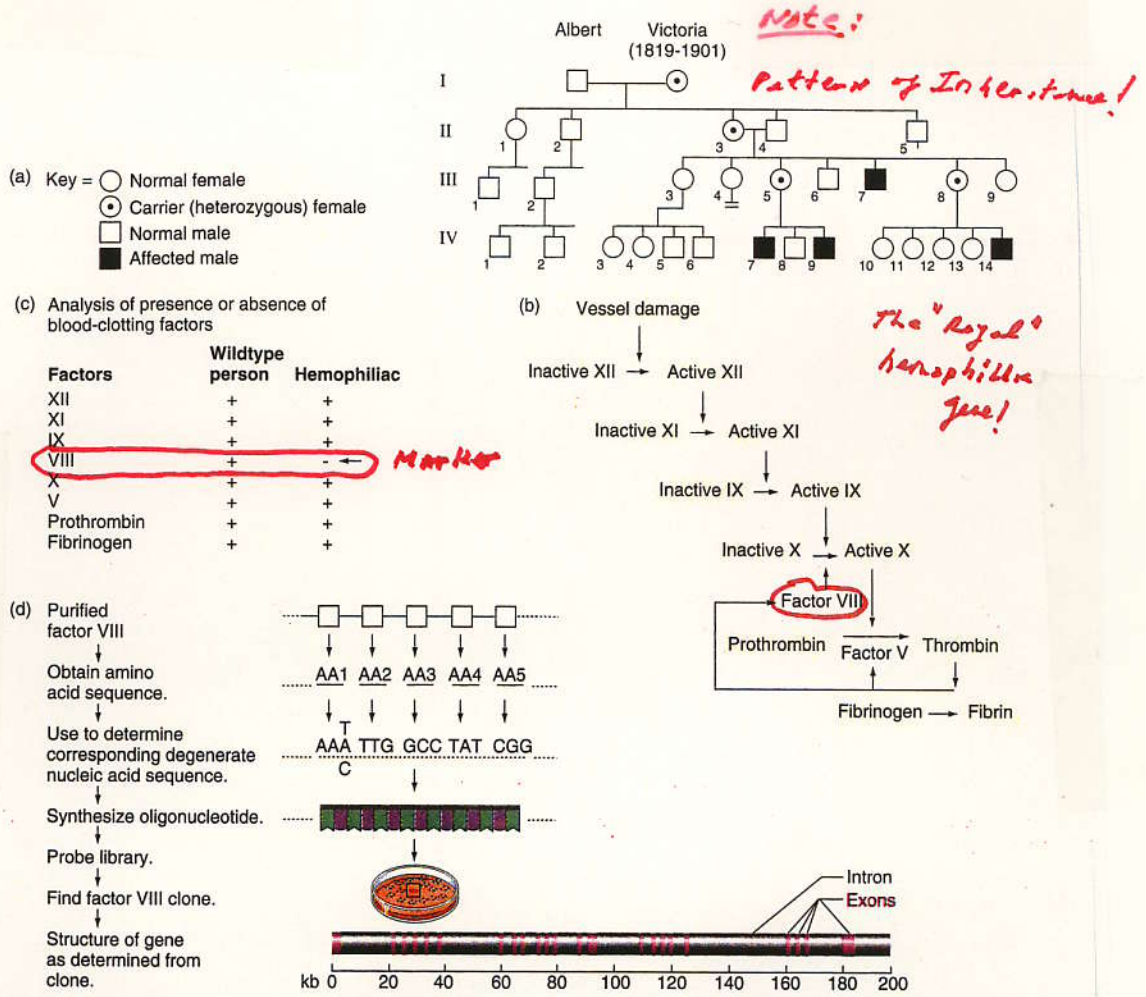


Figure 10.1 How geneticists identified the hemophilia A gene. (a) A pedigree of the royal family descended from Queen Victoria. This family tree uses the standard pedigree symbols. Black boxes represent males with hemophilia. (b) The blood clotting cascade. Vessel damage induces a cascade of enzymatic 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 CLONE A GENE WHEN YOU DON'T KNOW WHERE IT IS EXPRESSED!

What was known about Factor VIII before gene cloned?

1. Blood Protein (but perhaps synthesized elsewhere!)
2. Could be purified in small amounts from 25,000 liters of cow's blood! & pig's blood
3. Short stretch of both proteins sequenced & sequence known
4. Hemophilia A could be treated by blood transfusions from normal individuals ∴ clotting factor in blood.

∴ How to go from protein to gene?

KNOWLEDGE OF THE PROTEIN SEQUENCE AND THE GENETIC CODE MAKES IT POSSIBLE TO IDENTIFY A GENE

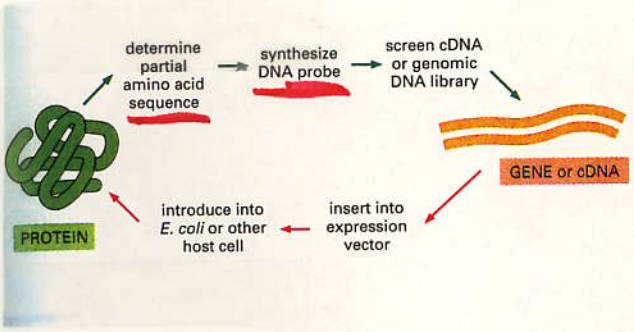


Figure 8–44 Knowledge of the molecular biology of cells makes it possible to experimentally move from gene to protein and from protein to gene. A small quantity of a purified protein is used to obtain a partial amino acid sequence. This provides sequence information that enables the corresponding gene to be cloned from a DNA library. Once the gene has been cloned, its protein-coding sequence can be inserted into an expression vector and used to produce large quantities of the protein from genetically engineered cells.

(1) ∴ PROTEIN → GENE → DRUG

OR

(2) GENE → PROTEIN USING SEQUENCING and Genetic Code

Genomics

GENBANK

2005 – JUST SEQUENCE Everything & Identify Protein – GENBANK HUGE

What is the Purpose of Cloning Genes/mRNAs?
A Review!

1. Isolate specific genes/mRNAs from genome & population of mRNAs.
2. Amplify specific gene/mRNA copies to obtain quantities for study.
3. Study Activity of Gene/What it does & what function does it play in cell?
4. Study Structure of Gene/Sequence of Gene/mRNA - Introns? Exons? Switches?
5. Determine what protein encoded by gene/mRNA
6. Use gene/mRNA to make drugs in bacteria, animals, +/or plants.
7. → Use gene/mRNA as probe to study genetic diseases/gene diversity/map genes
8. → Use gene/mRNA as probe to identify & trace human diseases/pedigrees & DNA fingerprints
9. Use gene/mRNA probe for forensics & DNA identification
10. Use specific genes/mRNAs/switches to engineer organisms genetically

WHAT TOOLS REQUIRED TO CLONE GENES AND MRNAS ?
A REVIEW!

(1) **DNA or mRNA** FROM ORGANISM / CELL TYPE (for mRNA) you WANT TO CLONE --- need to Isolate

(2) **Host Cells** For Vector Replication ---
E. coli (prokaryote) or Yeast (eukaryote) or Mammalian cells (vector 2)

(3) **Vectors** to Replicate DNA / Express Gene Coding Sequence --- PLASMID, VIRUS, COSMID, BAC, YAC

(4) **Enzymes** to cut + join (engineer) DNA Sequences & Synthesize cDNA copies of mRNAs ---
Restriction Enzymes, Ligase, DNA polymerase, Terminal TRANSferase, Reverse TRANSCRIPTASE →
ENZYMES that naturally function in cells

(5) **Probes** to identify Specific genes / sequence of DNA (e.g., switch) / mRNAs --- radioactive DNA/RNA probes, antibody probes

OR **DNA SEQUENCING** Machines + Large database (ity.!)

What are the properties of restriction enzymes? A review

(1) present only in bacteria & have a defense function

(2) Bind double-stranded DNA molecules only - linear & circular

(3) Recognize a specific DNA sequence - 4, 6, or 8 bp (Probability)

(4) DNA Recognition Sequence = palindrome or sequence that is the same when "read" from either direction - i.e., strand of DNA.

5' ↓ G A A | T T C 3'
3' C T T | A A G 5'
↑ where break bond
AXIS

(5) Some enzymes produce single-stranded complementary ("sticky") ends by digesting phosphodiester bonds within recognition sequence - bases that can anneal!

Palindrome: 5' ↓ G A A T T C 3' / 3' C T T A A G ↑ 5'

Heat →

sticky ends ↓ ↓

G
C T T A A + A A T T C

(2 fragments)

Because recognition sites are present in all DNAs - DNA from different "sources" can be joined together

(6) Restriction Enzymes Recognize all double-stranded DNA

(7) # of restriction sites ∝ to Genome Size

Bacteria < HUMAN
# #

(8) Order of restriction sites reflects DNA sequence
∴ unique DNA sequences have unique orders of restriction sites ∴ used for diagnostics - Markers!

ARE NON-RANDOM!!

(Map)

RESTRICTION ENZYMES RECOGNIZE Specific Sites

TABLE 7-1 Examples of the Actions of Restriction Endonucleases

Source Microorganism	Enzyme*	Recognition Site (↓)†		Number of Cuts (kb)‡ λ (50)	Ad2 (36)	SV40 (5.2)	pBR322 (4.3)
Arthrobacter luteus	*Alu*I	AG↓CT	4bp	143	158	34	14
Thermus aquaticus	*Taq*I	T↓CGA	4bp	121	50	1	13
Haemophilus parahaemolyticus	*Hph*I	GGTGA + 5	5bp	168	99	4	18
Haemophilus gallinarum	*Hga*I	GACGC + 8	5bp	102	87	0	12
Escherichia coli	*Eco*RI	G↓AATTC	6bp	5	5	1	1
Haemophilus influenzae	*Hin*dIII	A↓AGCTT	6bp	6	12	6	1
Nocardia otitiscaviaruns	*Not*I	GC↓GGCCGC	8bp	0	7	0	0
Streptomyces fimbriatus	*Sfi*I	GGCCN$_4$↓NGGCC		0	3	1	0

(2) (1)

← Genome size
← Fragment #

*Enzymes are named with abbreviations of the bacterial strains from which they are isolated; the Roman numeral indicates the enzyme's priority of discovery in that strain (for example, *Alu*I was the first restriction enzyme to be isolated from *Arthrobacter luteus*).

†Recognition sequences are written 5′ → 3′ (only one strand is given) with the cleavage site indicated by an arrow. For example, G↓GATCC is an abbreviation for

```
      ↓
(5′)GGATCC(3′)
(3′)CCTAGG(5′)
      ↑
```

The cleavage site for *Hph*I and *Hga*I occurs five or eight bases away from the recognition sequence; N indicates any base.

‡These columns list the number of cleavage sites recognized by specific endonucleases on the DNA of bacteriophage λ (λ), adenovirus type 2 (Ad2), simian virus 40 (SV40), and an *E. coli* plasmid (pBR322). The sizes of the DNAs in kilobases (kb) are in parentheses. Note that the actual number of cuts in these sequences deviates from the expected number in random sequences, which would be given by $L/4^n$, where n is the length of the site recognized by an enzyme and L is the length of the sequence.

SOURCE: R. J. Roberts, 1988, *Nuc. Acids Res.* 16(supp):r271.

(1) # Sites ∝ Genome Size

(2) Smaller DNA Recognition Sequence → Larger Total # DNA Fragments

How many Human DNA fragments are produced when Human Genome is digested with EcoRI?

(1) Human Genome Has 3×10^9 bp of DNA

(2) EcoRI recognizes the sequence 5'GAATTC3'

(3) Assume Human Genome Has 50% G+C & 50% A+T bases

(4) What is probability of EcoRI site in Human DNA?

$P = (0.25)^6 = 2.44 \times 10^{-4}$
or 1 site every 4096 bp ($1/2.44 \times 10^{-4}$)

(5) How many EcoRI sites are there in the human genome on the basis of chance?

Sites $= (2.44 \times 10^{-4})(3 \times 10^9) = 732{,}000$

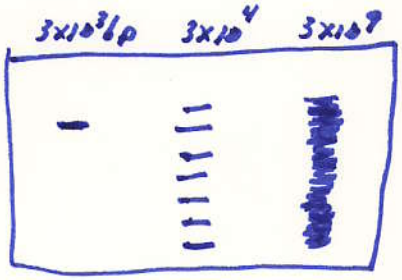
(6) How many EcoRI fragments?

$732{,}000 + 1 = 732{,}001!$ Why!? A Linear DNA is digested into TWO DNA fragments. ∴ # fragments = # sites + 1!

Could you visualize (see!) any one fragment on an electrophoresis gel?

How many fragments are produced with smaller genomes?

(1)	3×10^3 bp	~1	Specific Bands
(2)	3×10^4 bp	~7	Specific Bands
(3)	3×10^5 bp	~70	Specific Bands
(4)	3×10^6 bp	~700	Smear
(5)	3×10^7 bp	~7000	Smear
(6)	3×10^8 bp	~70,000	Smear
(7)		~700,000	Smear

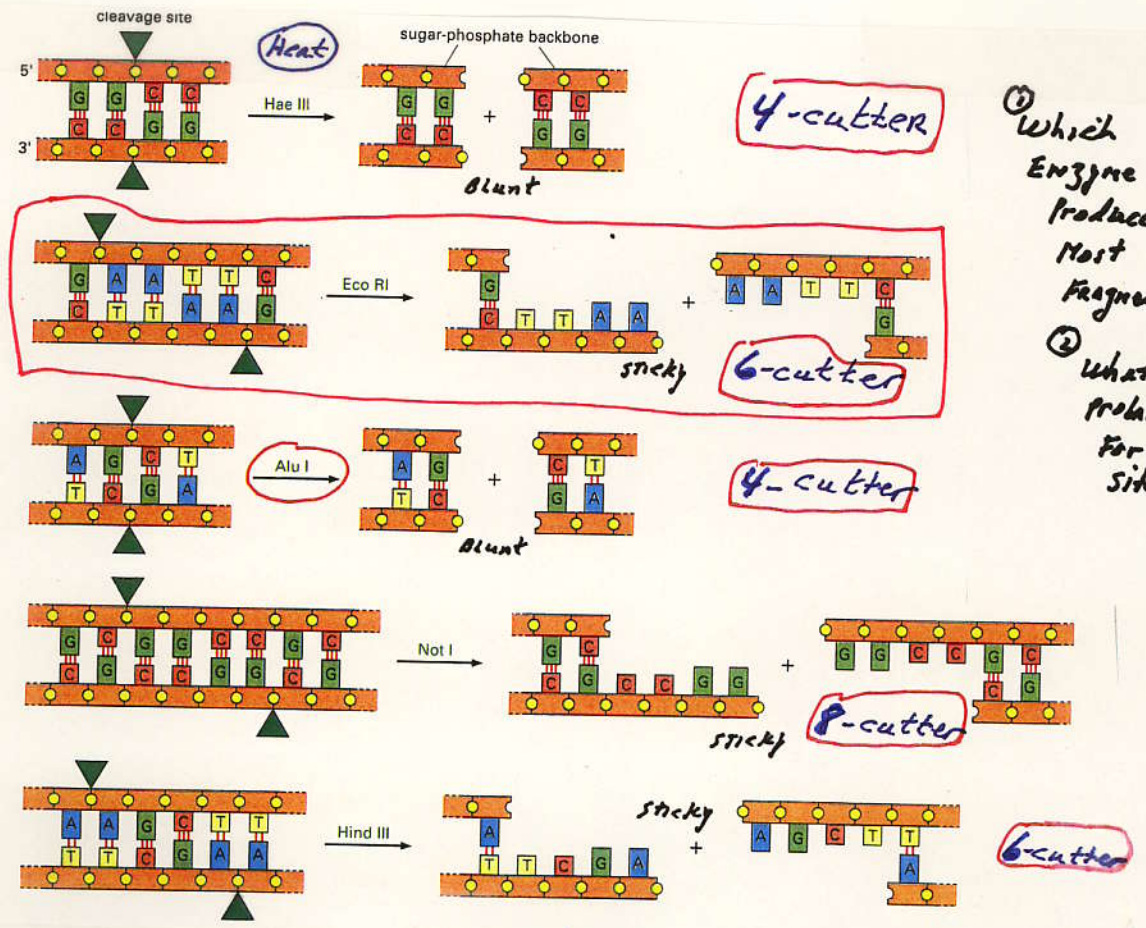


too many fragments to see individual ones!

? How "see" specific Human DNA fragment without PCR or Cloning?

NOTE: Sites are not distributed randomly in a genome - this is Theoretical

MANY RESTRICTION ENZYMES LEAD TO "STICKY END" FORMATION



Note! PALindrome

① Which Enzyme Produces Most Fragments.

② What is Probability For A Site?

Figure 10–2 The nucleotide sequences recognized and cut by five widely used restriction nucleases. As shown, the target

Need For/To CLONE PCR Product

RESTRICTION ENZYMES HAVE MANY USES IN GENETIC ENGINEERING & GENE STUDY

(1) **Cloning / Recombinant DNA** - Creating recombinant DNA Molecule
Moving parts of Genes (switches, introns, exons)

(2) **Mapping Clones / Genes / Chromosomes**

Maps provide guide posts - Mark positions in gene, plasmid, chromosome, genome, etc. Unique Sequence → Unique Map
Land Marks for DNA segments

(3) **DIAgnosis**

Specific Genes / Alleles (e.g., Normal vs. Disease Gene (RFLP))
Identity / Forensics e.g., Crime, Paternity, Lineage
Presence of Pathogens e.g., detect specific strain of bacteria

(4) **Ecology**

Species Identity
Tracing Races to Geography
Movement of Endangered Species

Contrast VNTRs - STRs

(5) **Anthropology**

Human Lineages
Population Diversity
Presence of Specific Pathogens

PROVIDES SPECIFIC FRAGMENT IDENTITY

CAN BE USED IN COMBINATION with PCR

"UPSTAGED" BY PCR → specific Band but ALSO use RFLP's

MAPS

RESTRICTION ENZYME SITES ARE SEQUENCED BASED AND ARE ESSENTIAL FOR GENE & GENOME MAPPING AND DNA TESTING/IDENTITY

(1) Map Genes, Chromosomes, Genome

(2) Maps of Genes CAN be used to:
a. Study & Manipulate Gene Regions (eg, Switch)
b. Cut out & Clone Specific Gene Regions
c. Diagnosis/Identify Disease Genes/Specific Genes

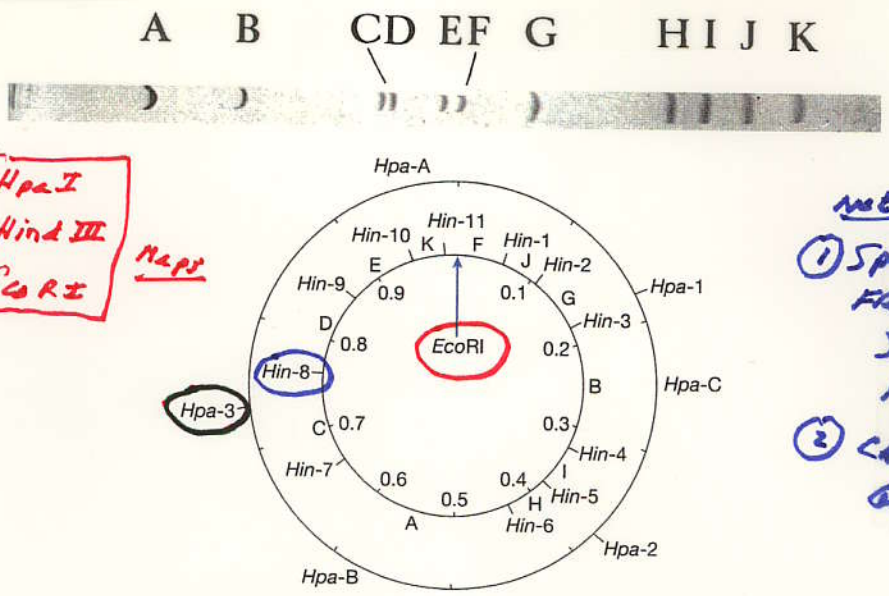
(3) Maps of Chromosomes can be used to:
a. Mark-Map Gene Locations
b. Identify Specific Chromosomes (eg, Y chromosome)
c. Identify Regions containing known Genes from other Studies - Markers for Genes

(4) Maps of Genome can be used to:
a. Start Sequencing Entire Genome - know where fragment being Sequenced is!
b. Create Recombinant Vectors using Vector Genome Map!

BASIS OF ALL Gene Manipulation & Engineering - Need Maps to know where you are!

Mapping Requires Cloned DNA Molecules - It is done after recombinant DNAs created or generated from DNA sequence!

RESTRICTION SITES ARE FOUND AT SPECIFIC LOCATIONS IN GENOME



HpaI
Hind III
EcoRI
Maps

Note:
(1) Specific DNA Fragments Mark Specific Locations in Genome!
(2) Can be used as Gene & Genome Markers

Cleavage Map of the SV40 Genome

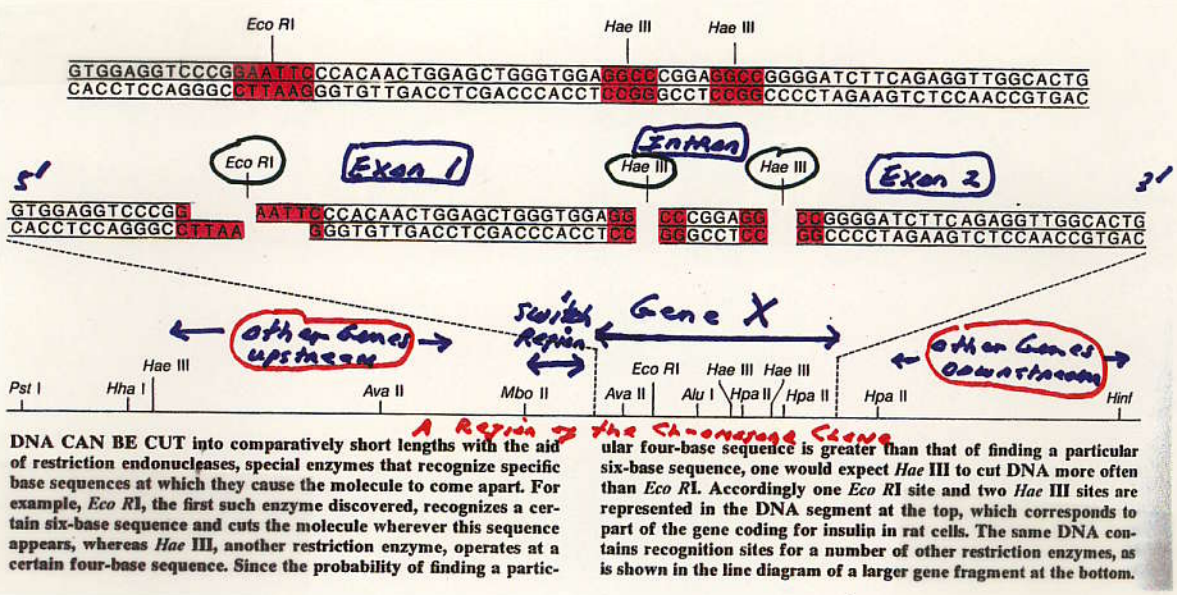
(*Top*) DNA fragments generated from SV40 DNA digested with *Hin*dII restriction endonuclease separated by electrophoresis. (Reprinted, with permission, from Danna K. and Nathans D. 1971. Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc. Natl. Acad. Sci. 68:* 2913–2917.) (*Bottom*) A cleavage map of the SV40 genome. Fragments created by *Hin*dII digestion are labeled A to K, corresponding to fragments seen on the gel above. (Reprinted, with permission, from Danna K., Sack G.H., and Nathans D. 1973. Studies of simian virus 40 DNA. VII. A cleavage map of the SV40 genome. *J. Mol. Biol. 78:* 363–376.)

RESTRICTION SITES REFLECT DNA SEQUENCE ∴ PATTERN IS GENE & GENOME SPECIFIC!

STARTING POINT - CLONES! IF KNOW DNA SEQUENCE - RESTRICTION MAP FOLLOWS (QUIZ #2!)

ISOLATING THE MODULES
Yo! It's in the Sequence!

A RESTRICTION MAP PROVIDES GUIDEPOSTS FOR IDENTIFYING AND MANIPULATING Genes



GENE X

DNA CAN BE CUT into comparatively short lengths with the aid of restriction endonucleases, special enzymes that recognize specific base sequences at which they cause the molecule to come apart. For example, *Eco RI*, the first such enzyme discovered, recognizes a certain six-base sequence and cuts the molecule wherever this sequence appears, whereas *Hae* III, another restriction enzyme, operates at a certain four-base sequence. Since the probability of finding a particular four-base sequence is greater than that of finding a particular six-base sequence, one would expect *Hae* III to cut DNA more often than *Eco RI*. Accordingly one *Eco RI* site and two *Hae* III sites are represented in the DNA segment at the top, which corresponds to part of the gene coding for insulin in rat cells. The same DNA contains recognition sites for a number of other restriction enzymes, as is shown in the line diagram of a larger gene fragment at the bottom.

1. Isolating Switches + Terminators
2. Isolating Coding Regions
3. Making Chimeric Genes with "Mix/Match" Parts FROM Different Genes
4. Identifying Specific Genes / FORMS of Gene e.g., disease gene!

RESTRICTION MAPS GENERATED FROM SEQUENCE OF Gene/Genome & knowledge of Restriction ENZYME Site OR GENERATED AS A PUZZLE WITHOUT THE DNA SEQUENCE

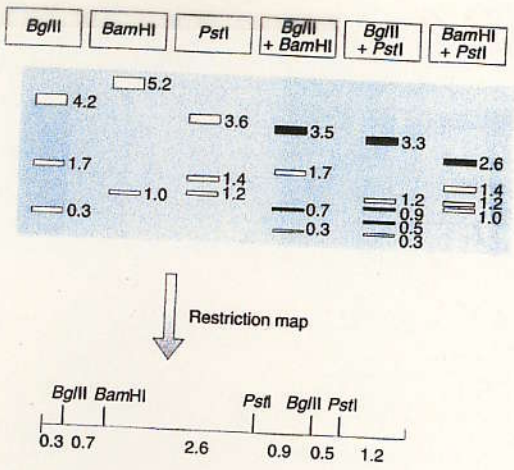


Figure 4.9: Generating a restriction map.

The size patterns from double digests provide information on the relative locations of restriction sites. The example shows size fractionation by agarose gel electrophoresis of restriction fragments following incubation of a 6.2 kb DNA fragment with the indicated enzymes. New bands in the double digests (i.e. not found in the original single digests) are indicated by black boxes. In the *Bgl*II + *Bam*HI double digest, the original 1.7 kb and 0.3 kb bands from the *Bgl*II digest alone are maintained, suggesting that these fragments do not have a *Bam*HI site, while the 4.2 kb *Bgl*II fragment is replaced by 3.5 kb and 0.7 kb fragments, suggesting that there is a *Bam*HI site within 0.7 kb from one end of the 4.2 kb *Bgl*II fragment. Similarly, in the *Bam*HI + *Pst* I double digest, the 1.4 kb and 1.2 kb fragments seen in the *Pst*I digest alone are maintained, suggesting that they lack a *Bam*HI site, while the 3.6 kb *Pst*I fragment is replaced by a 2.6 kb + 1.0 kb fragment, as a result of possession of an internal *Bam*HI site located 1.0 kb from one end. By comparing all three patterns of double digestion, the restriction map at the bottom can be deduced. *Note* that restriction mapping is often helped by the use of partial digests and also by end-labeling (Section 5.1.1).

Direct DNA Sequencing Has Replaced Making Restriction Maps

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 – ⓐ Transformation, ⓑ infection

EASily Purified from host cell + manipulated
(e.g.) ⓐ plasmid, ⓑ 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 genetic engineering - Can do anything!

Purpose of all vectors - Clone / Manipulate DNA/cDNA

VECTORS USED IN GENETIC ENGINEERING HAVE SIMILAR CONCEPTUAL PROPERTIES BUT ARE USED IN DIFFERENT SITUATIONS

TABLE 3.2 A COMPARISON OF DNA VECTORS AND THEIR APPLICATIONS

Vector Type	Maximum Insert Size (kb)	Applications	Limitations
Bacterial plasmid vectors (circular)	~6–12	DNA cloning, protein expression, subcloning, direct sequencing of insert DNA	Restricted insert size; limited expression of proteins; copy number problems; replication restricted to bacteria
Bacteriophage vectors (linear)	~25	cDNA, genomic and expression libraries	Packaging limits DNA insert size; host replication problems
Cosmid (circular)	~35	cDNA and genomic libraries, cloning large DNA fragments	Phage packaging restrictions; not ideal for protein expression; cannot be replicated in mammalian cells
Bacterial artificial chromosome (circular)	~300	Genomic libraries, cloning large DNA fragments	Replication restricted to bacteria; cannot be used for protein expression
Yeast artificial chromosome (circular)	200–1,000 (1 megabase)	Genomic libraries, cloning large DNA fragments	Must be grown in yeast; cannot be used in bacteria
Ti vector (circular)	Varies depending on type of Ti vector used	Gene transfer in plants	Limited to use in plant cells only; number of restriction sites randomly distributed; large size of vector not easily manipulated.

Plasmids vs. Bacteriophage Vectors

TWO COMMON PLASMID VECTORS

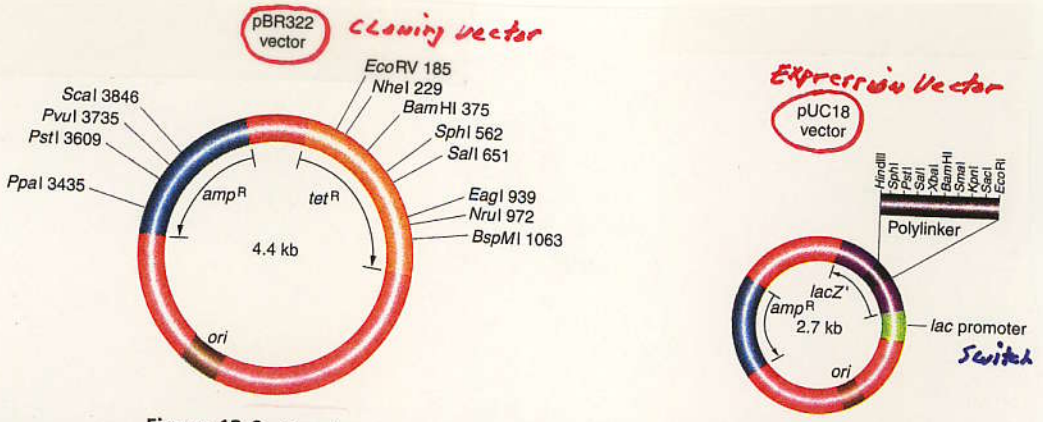


Figure 12-6 Two plasmids designed as vectors for DNA cloning, showing general structure and restriction sites. Insertion into pBR322 is detected by inactivation of one drug-resistance gene (tet^R), indicated by the Tet^S (sensitive) phenotype. Insertion into pUC18 is detected by inactivation of the β-galactosidase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

Newer - More Sophisticated
21st century Vectors
Exist - DO Everything!

Built From Modules!

CAN USE ANTIBIOTIC RESISTANCE AND COLOR TO SCREEN FOR RECOMBINANT PLASMIDS

pUC 18 vector

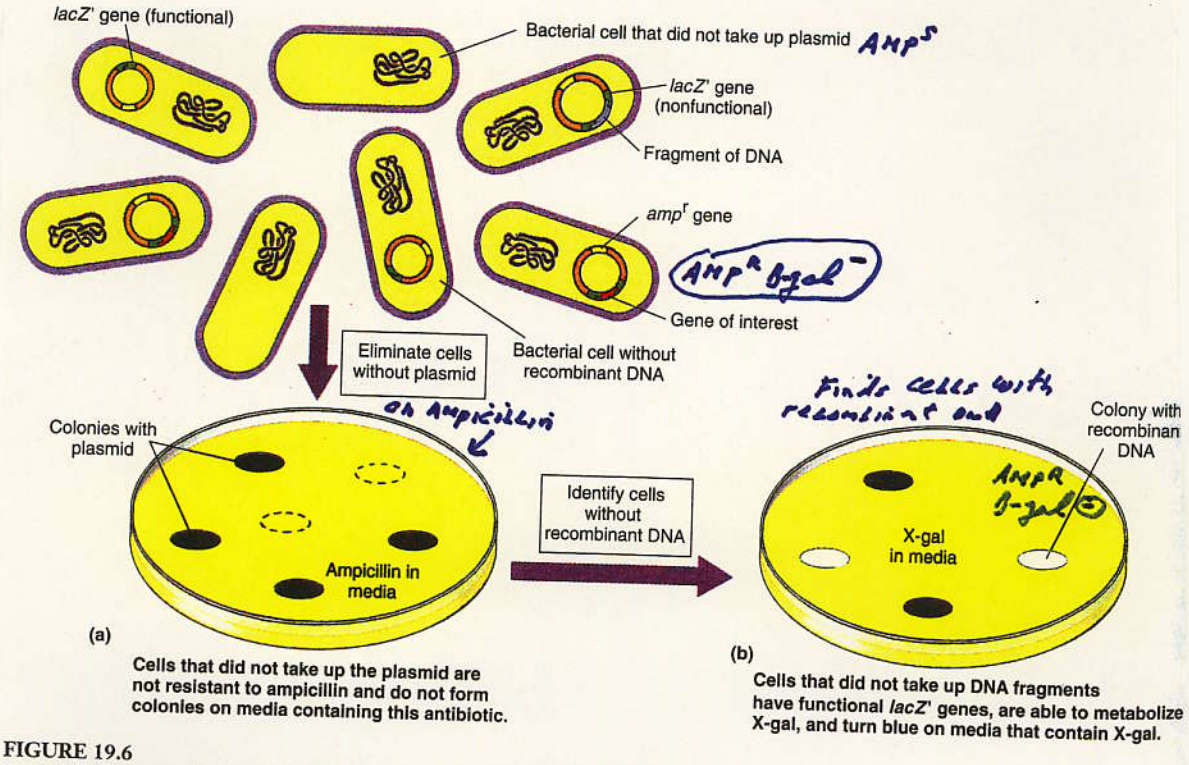


FIGURE 19.6

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

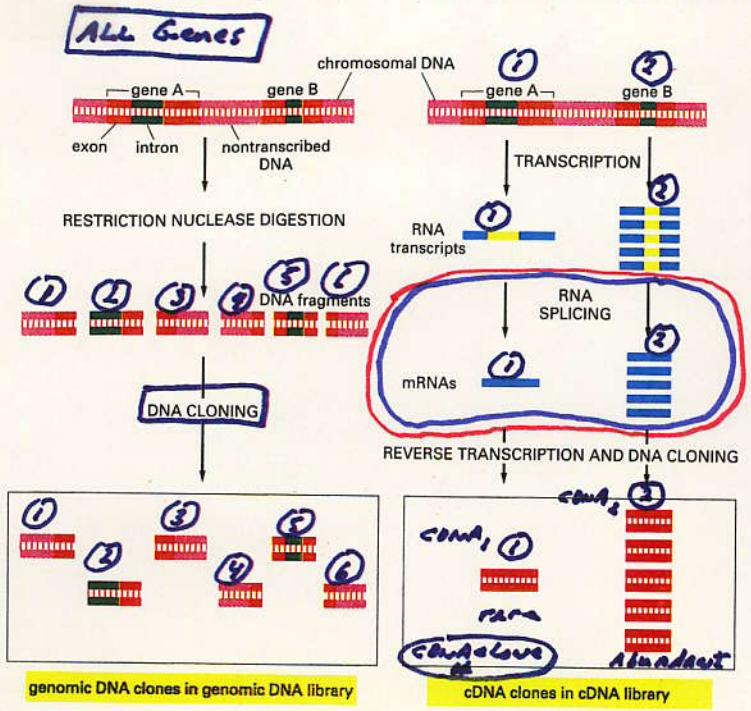


Figure 8–35 The differences between cDNA clones and genomic DNA clones derived from the same region of DNA. In this example gene A is infrequently transcribed, whereas gene B is frequently transcribed, and both genes contain introns *(green)*. In the genomic DNA library, both the introns and the nontranscribed 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 DNAx → 10 DNAx in library
1 DNAy → 1 DNAy in library

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 mRNAx → 1 cDNAx
100 mRNAy → 100 cDNAy

EASIER to find Abundant cDNA 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

cDNA Clones

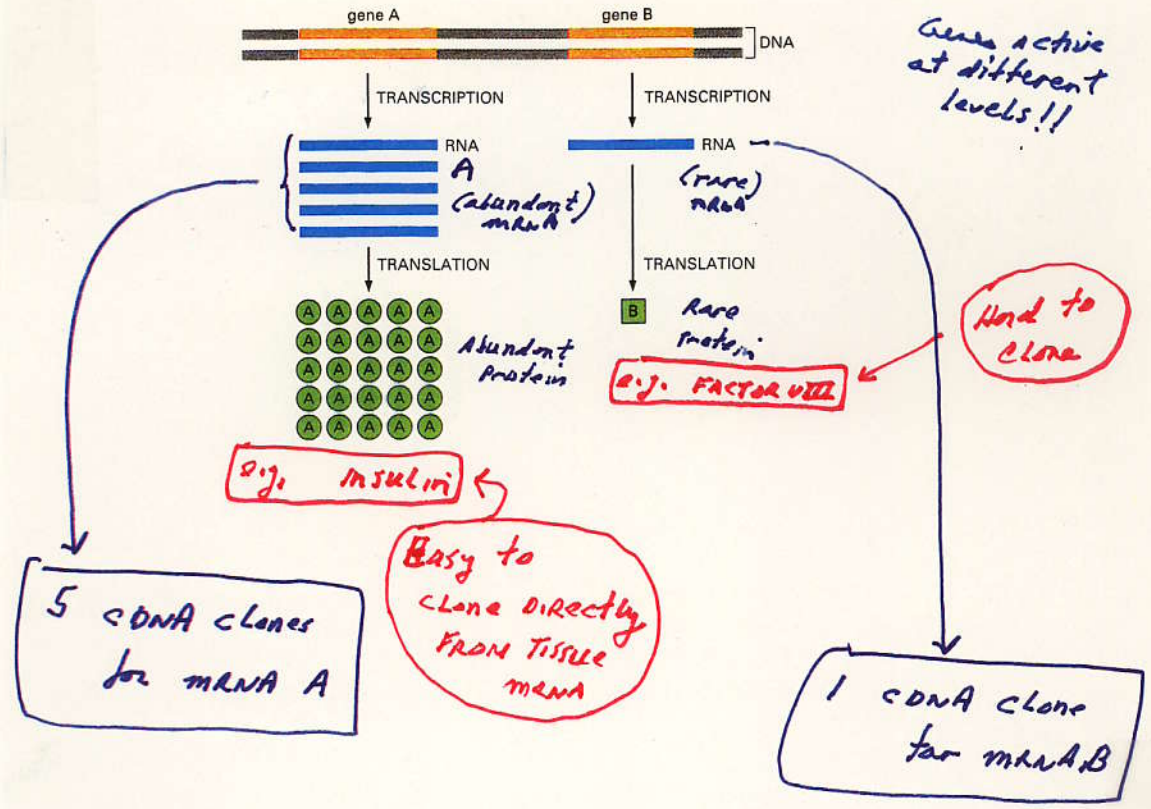
① ONLY Coding sequences ∴ useful to identify proteins
② Subset of gene sequences - what genes active in specific cell - time of development - Tumor? Profile of Active Genes
③ useful to identify Specific "genes"
④ For Drugs/Bacteria expression

Need FOR Genome Sequencing

Need for EST Sequencing

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

Genes Active in Organ X



∴ 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

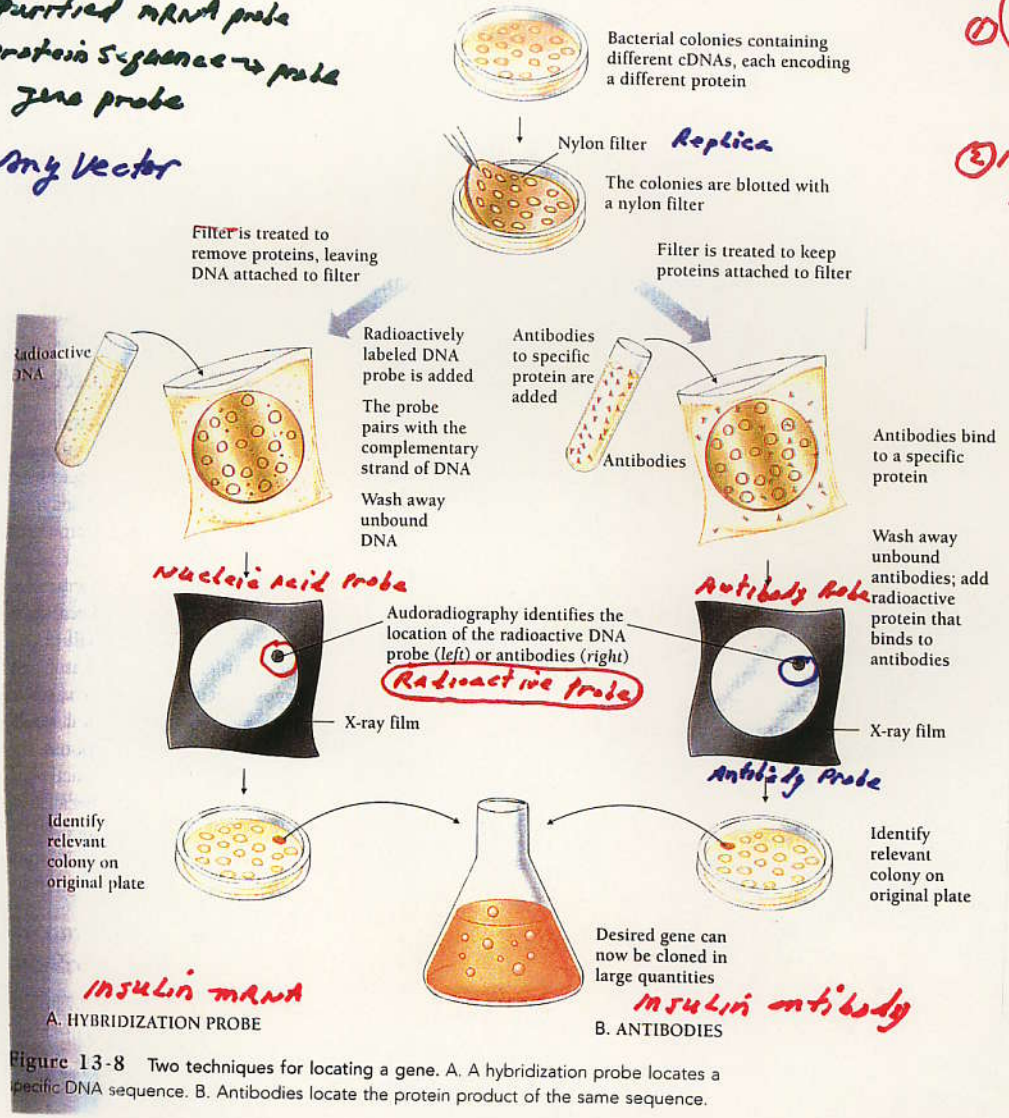
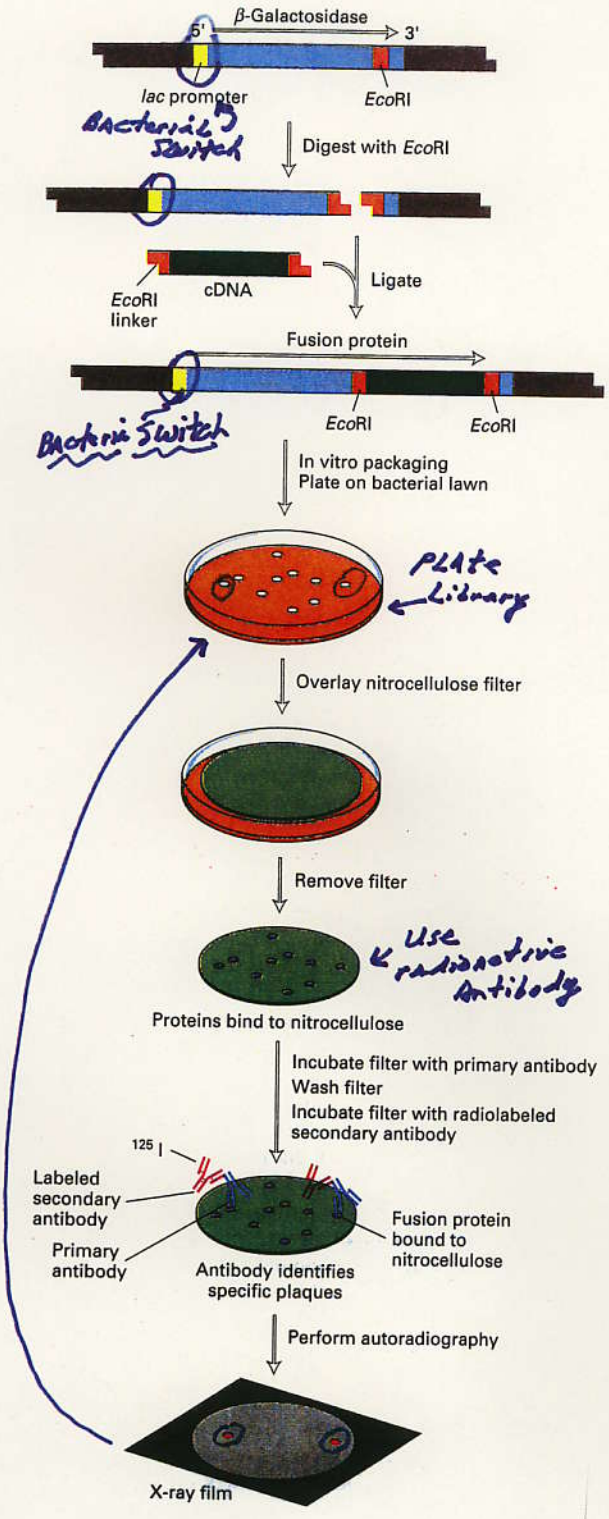


Figure 13-8 Two techniques for locating a gene. A. A hybridization probe locates a specific DNA sequence. B. Antibodies locate the protein product of the same sequence.

Identification of Insulin cDNA clone FROM PANCREAS mRNA Library

200 → 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.]

THE PROBLEM

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

1980's

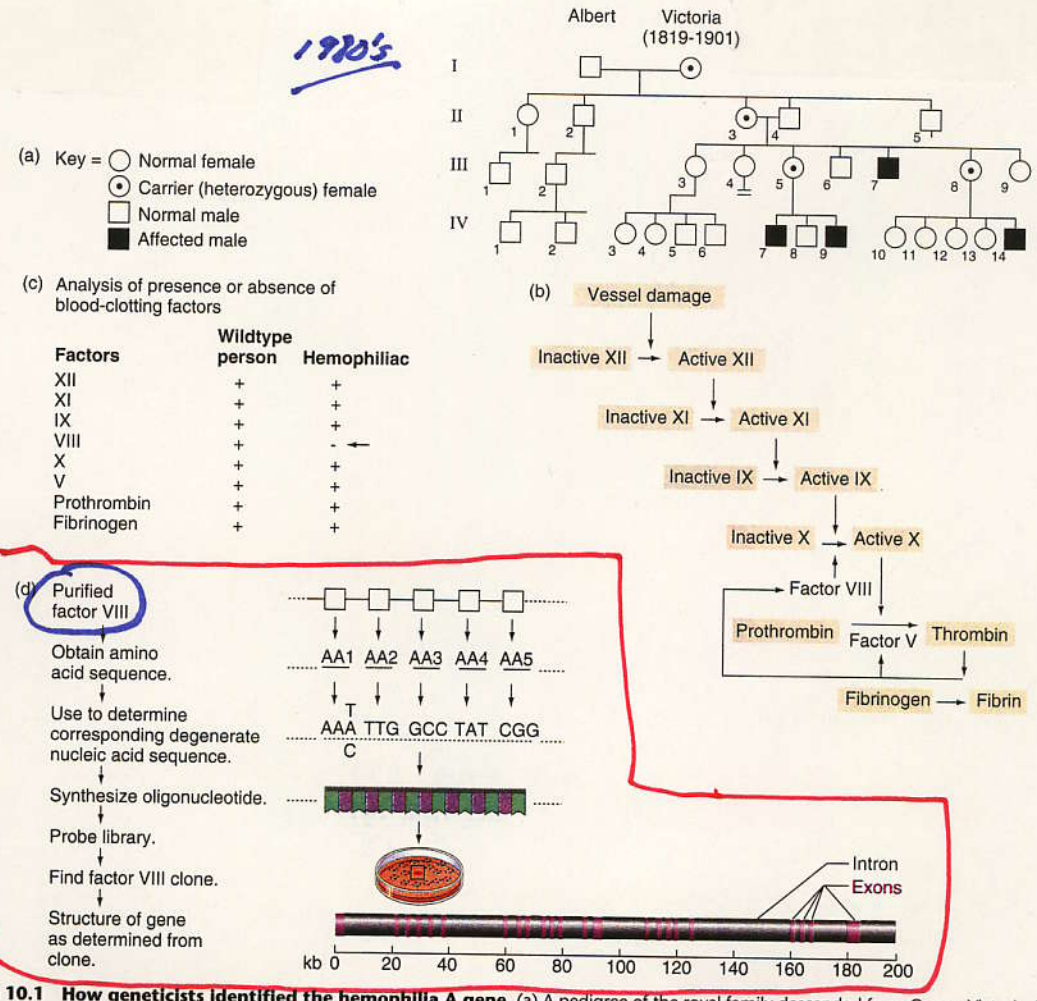


Figure 10.1 How geneticists identified the hemophilia A gene. (a) A pedigree of the royal family descended from Queen Victoria. This family tree uses the standard pedigree symbols. Black boxes represent males with hemophilia. (b) The blood clotting cascade. Vessel damage induces a cascade of enzymatic 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?

Protein → Gene → mRNA → DNA!

FACTOR VIII PROTEIN → GENE USING GENOME LIBRARY

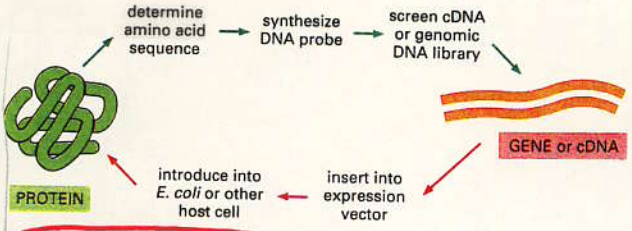


Figure 10–28 Knowledge of the molecular biology of cells makes it possible to experimentally move from gene to protein and from protein to gene. A small quantity of a purified protein is used to obtain a partial amino acid sequence. This provides sequence information that enables the corresponding gene 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

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 → Database
2. Probe from cDNA/Switch
3. Probe from pure mRNA
4. Synthetic Probe from translated DNA Sequence ← ↓ Genetic Code

cDNA

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

Can't use Antibody - don't know where gene active (3D) ∴ can't make cDNA library

STEPS REQUIRED TO CLONE FACTOR VIII GENE AND cDNA

① Make Genome Library Because Factor VIII GENE in genome!

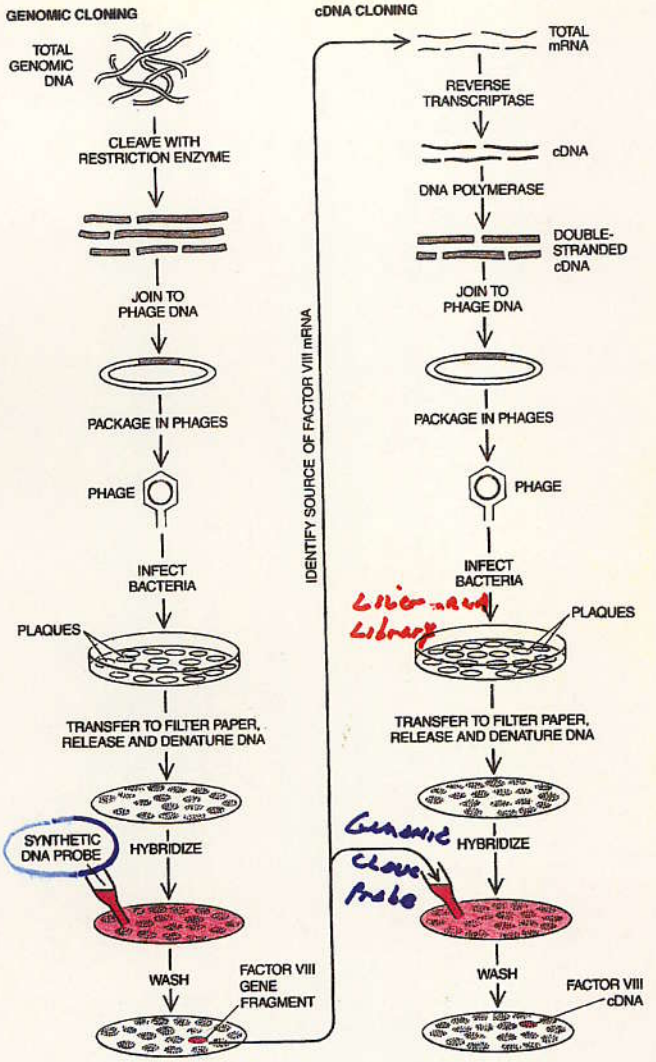
② Purify protein from Blood – that's where it works (wasn't know where made)

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 or re 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 TEAT Hemophilia A!

USING BACTERIOPHAGES AS VECTORS TO CLONE THE HUMAN GENOME

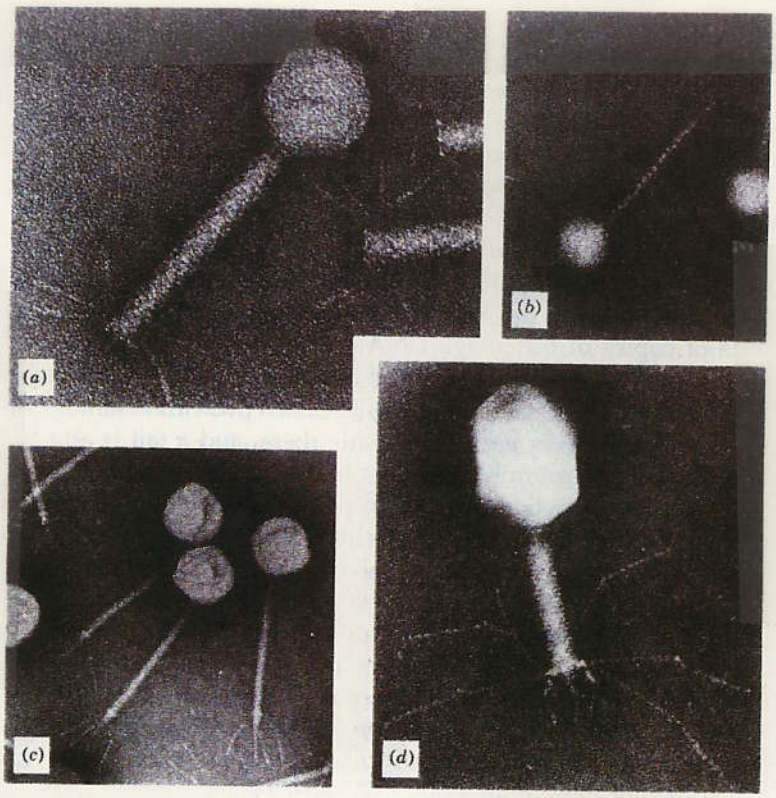
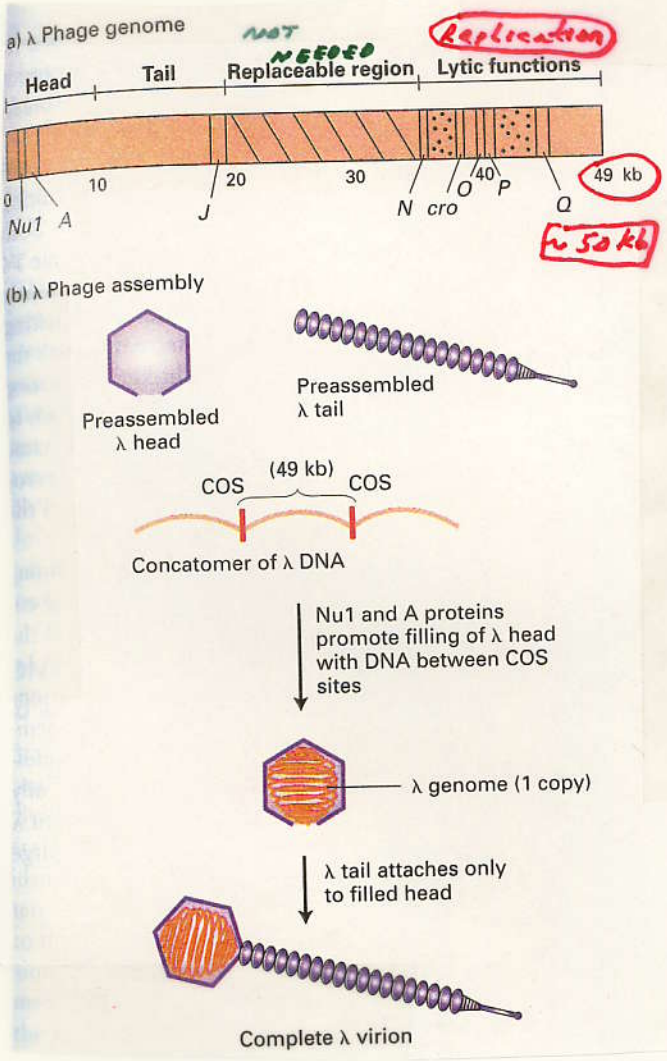


Figure 6-5. Electron Micrographs of Bacteriophages. **(a)** Bacteriophage P2, magnification 226,000 times. **(b)** Bacteriophage lambda, magnification 109,000 times. **(c)** Bacteriophage T5, magnification 91,000 times. **(d)** Bacteriophage T4, magnification 180,000 times. (Photomicrographs courtesy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

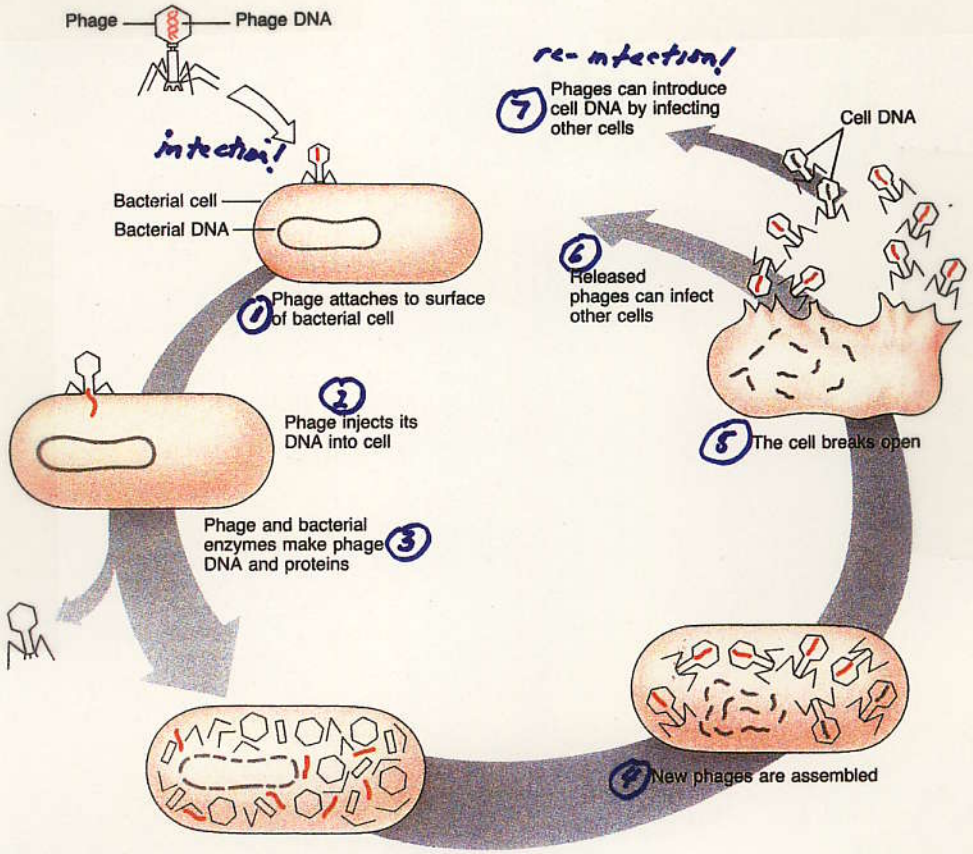
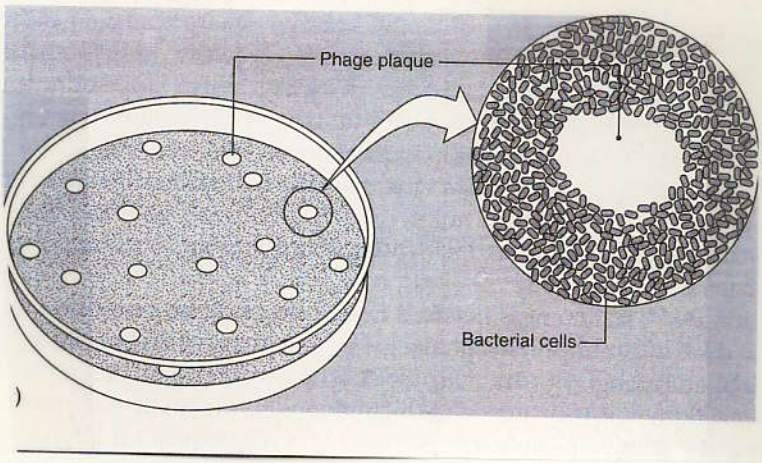
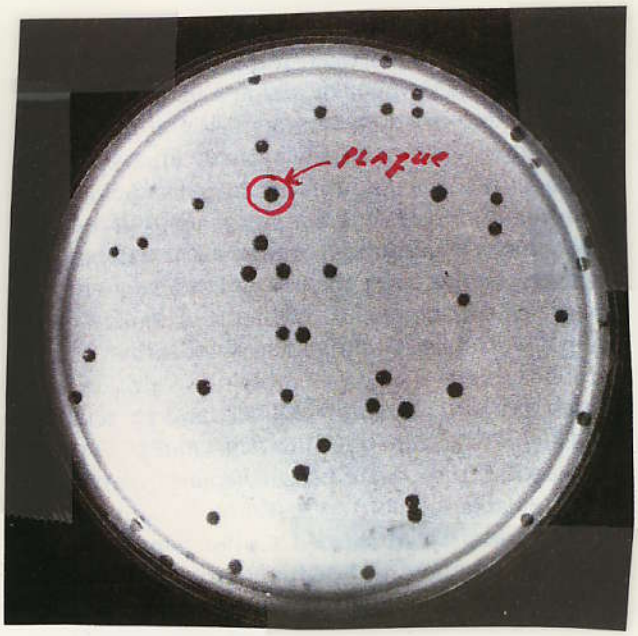


Figure 4.4 Events that occur when a phage infects a bacterial cell.

LYSED CELLS CAN BE SEEN AS CLEAR PLAQUES ON AGAR PLATES





EACH PLAQUE IS A VIRUS CLONE REPRESENTING ONE VIRAL infection!

CLONING THE HUMAN GENOME AND SCREENING FOR THE FACTOR VIII GENE

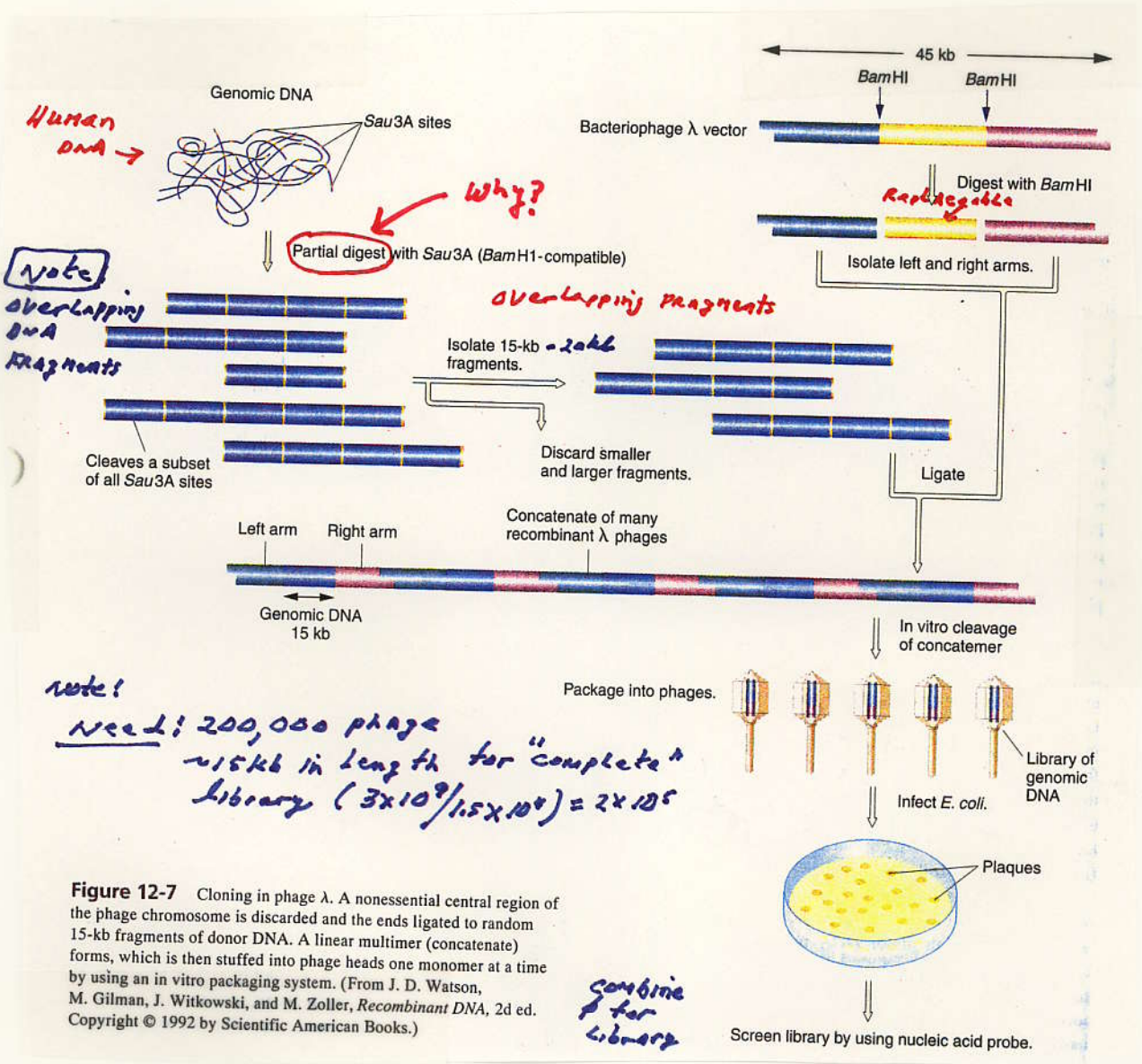


Figure 12-7 Cloning in phage λ. A nonessential central region of the phage chromosome is discarded and the ends ligated to random 15-kb fragments of donor DNA. A linear multimer (concatenate) forms, which is then stuffed into phage heads one monomer at a time by using an in vitro packaging system. (From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

note!
Need 200,000 phage ~15kb in length for "complete" library ($3\times10^9/1.5\times10^4$) = 2×10^5

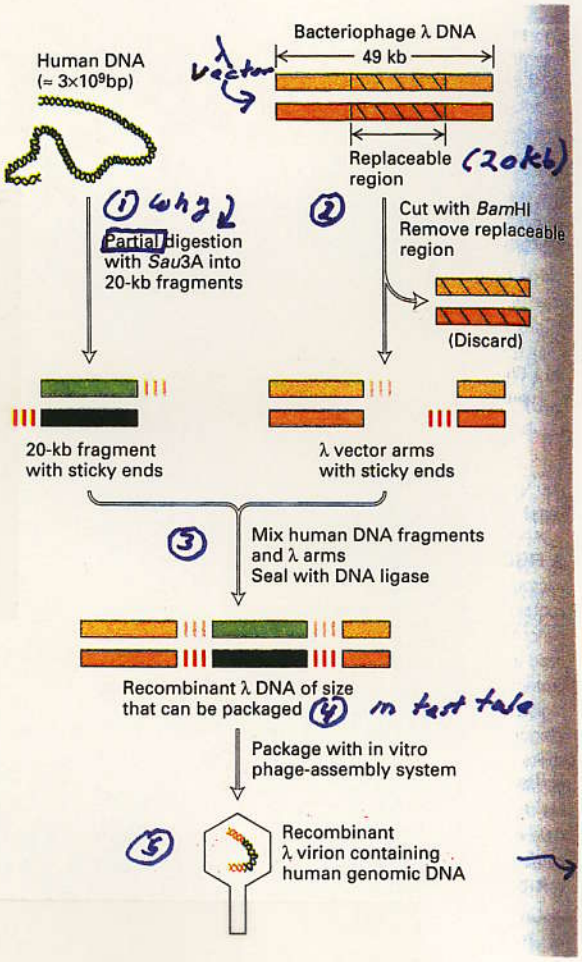
combine for library

Why overlapping fragments?

USING λ VIRUSES AS A VECTOR TO CLONE HUMAN GENOME

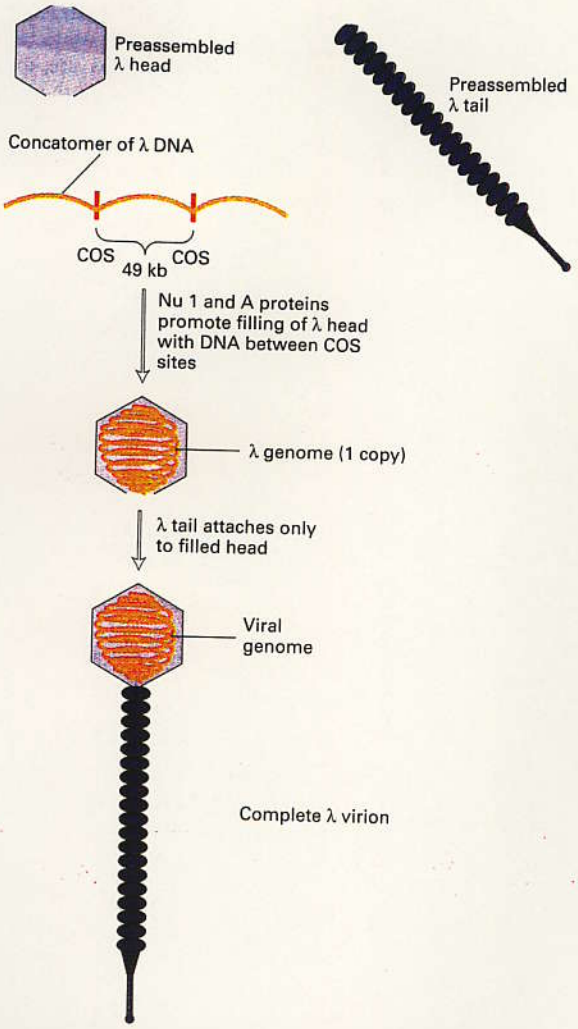
A COMPLETE HUMAN GENOME LIBRARY in λ phage vector

① Large DNA Fragments
② Keep Genes in tact
③ Keep Clones in Library as few as possible to Find Gene – Easier to hunt thru 600,000 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.

λ VIRUS CAN BE SELF-ASSEMBLED IN A TEST TUBE



▲ **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.

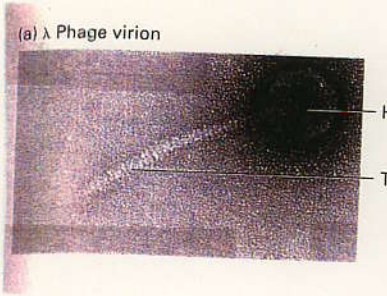
ALSO PACKAGED NATURALLY IN E. coli CELLS

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

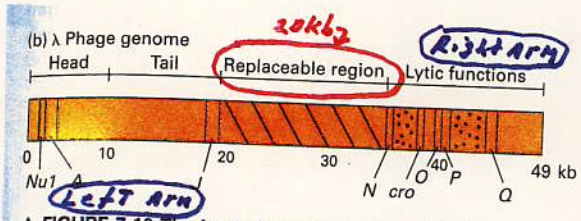
ADVANTAGES over PLASMIDS

MAX Delbruck — Father of phage!

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!

Note: Restriction Enzymes "Fight" Viral infection!

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! 20 kb

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!

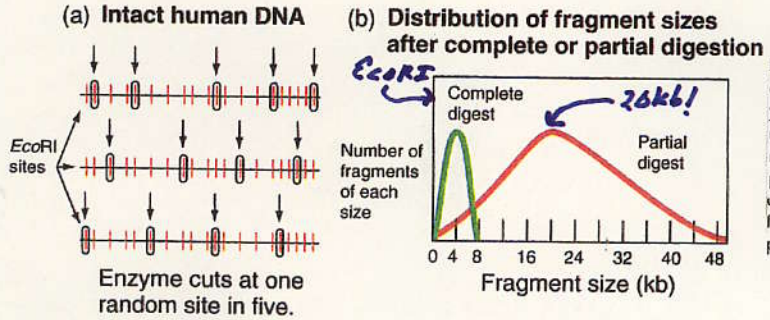
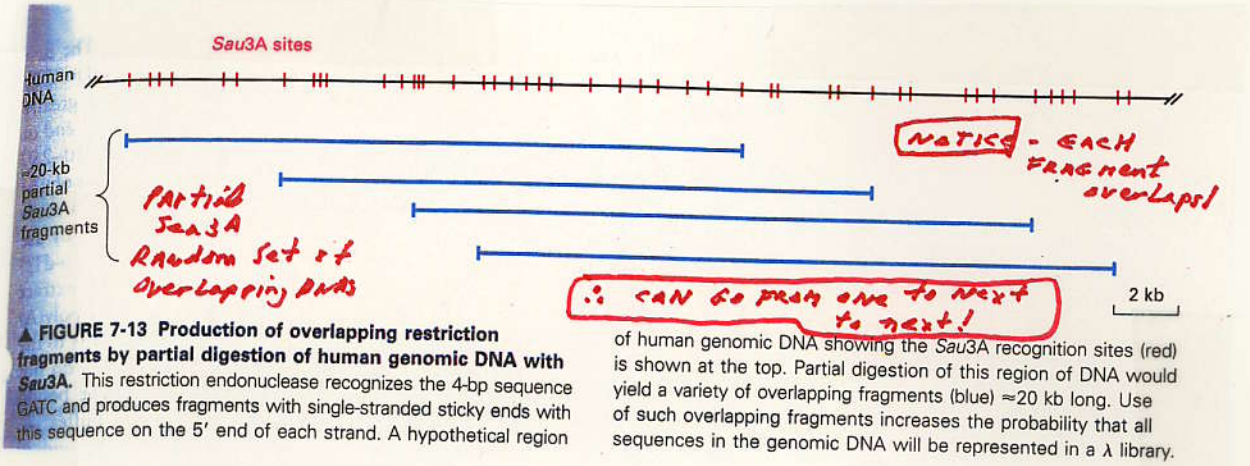


Figure 8.4 Comparison of results from partial and complete digests. (a) By reducing the time available for the reaction to occur, you can ensure that an enzyme actually cuts only a subset of the total recognition sites within a DNA sample. In this example, the chosen reaction time allowed only 1/5 of all *EcoRI* 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!!

CONSTRUCTING A HUMAN GENOME LIBRARY BY PARTIAL DIGESTION CREATES A SET OF OVERLAPPING DNA FRAGMENTS/CLONES



▲ **FIGURE 7-13 Production of overlapping restriction fragments by partial digestion of human genomic DNA with *Sau*3A.** This restriction endonuclease recognizes the 4-bp sequence GATC and produces fragments with single-stranded 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.

∴ WOULD AN overlapping set for each of the 24 chromosomes ALLOWING clones to be ordered FROM beginning to end by restriction mapping BECAUSE EACH Chromosome contains one DNA MOLECULE!

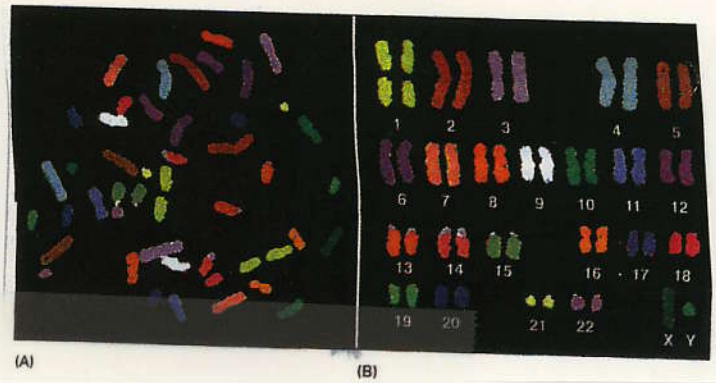
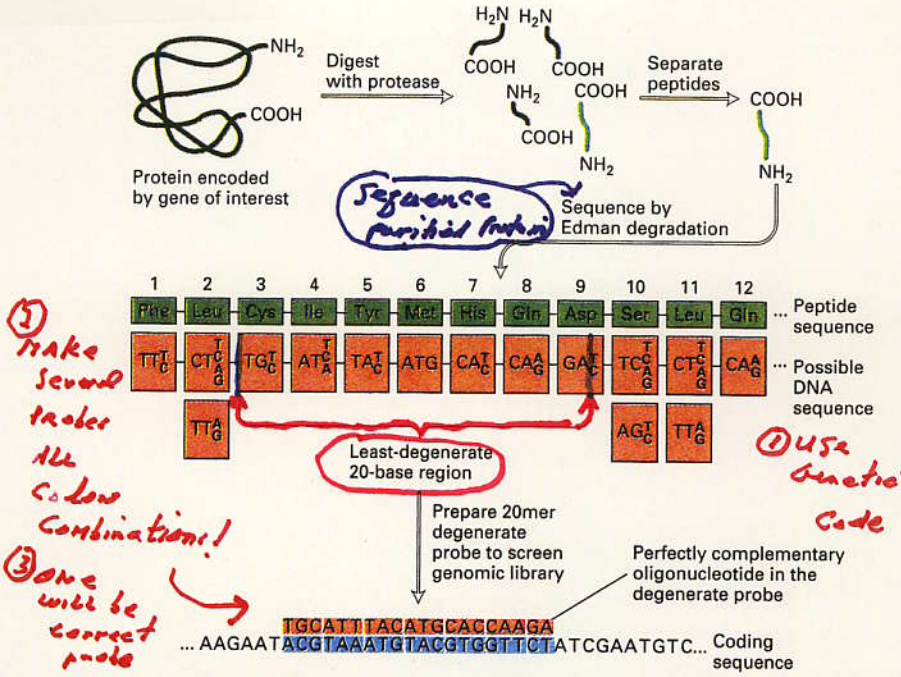


Figure 8–2 Human chromosomes. (A) The chromosomes as visualized as they originally spilled from the lysed cell. (B) The same chromosomes artificially lined up in order. This arrangement of the full chromosome set is called a karyotype. (From E. Schröck et al., *Science* 273:494–497, 1996.)

FACTOR VIII Protein → Gene

USING THE FACTOR VIII PROTEIN SEQUENCE AND GENETIC CODE AS A GUIDE TO SYNTHESIZE A FACTOR VIII Probe



▲ FIGURE 7-19 Designing oligonucleotide probes based on protein sequence. An isolated protein is digested with a selective protease such as trypsin, which specifically cleaves peptide bonds on the carboxy-terminal side of lysine and arginine residues. The resulting peptides are separated, and several are partially sequenced from their N-terminus by sequential Edman degradation. The determined sequences then are analyzed to identify the 6- or 7-aa region that can be encoded by the smallest number of possible DNA sequences. Because of the degeneracy of the genetic code, the 12-aa sequence (light green) shown here theoretically could be encoded by any of the DNA triplets below it, with the possible alternative bases at the same position indicated. For example, Phe-1 is encoded by TTT or TTC; Leu-2 is encoded by one of six possible triplets (CTT, CTC, CTA, CTG, TTA, or TTG). The region with the least degeneracy for a sequence of 20 bases (20-mer) is indicated by the red bracket. There are 48 possible DNA sequences in this 20-base region that could encode the peptide sequence 3–9. Since the actual sequence of the gene is unknown, a degenerate 20-mer probe consisting of a mixture of all the possible 20-base oligonucleotides is prepared. If a cDNA or genomic library is screened with this degenerate probe, the one oligonucleotide that is perfectly complementary to the actual coding sequence (blue) will hybridize to it.

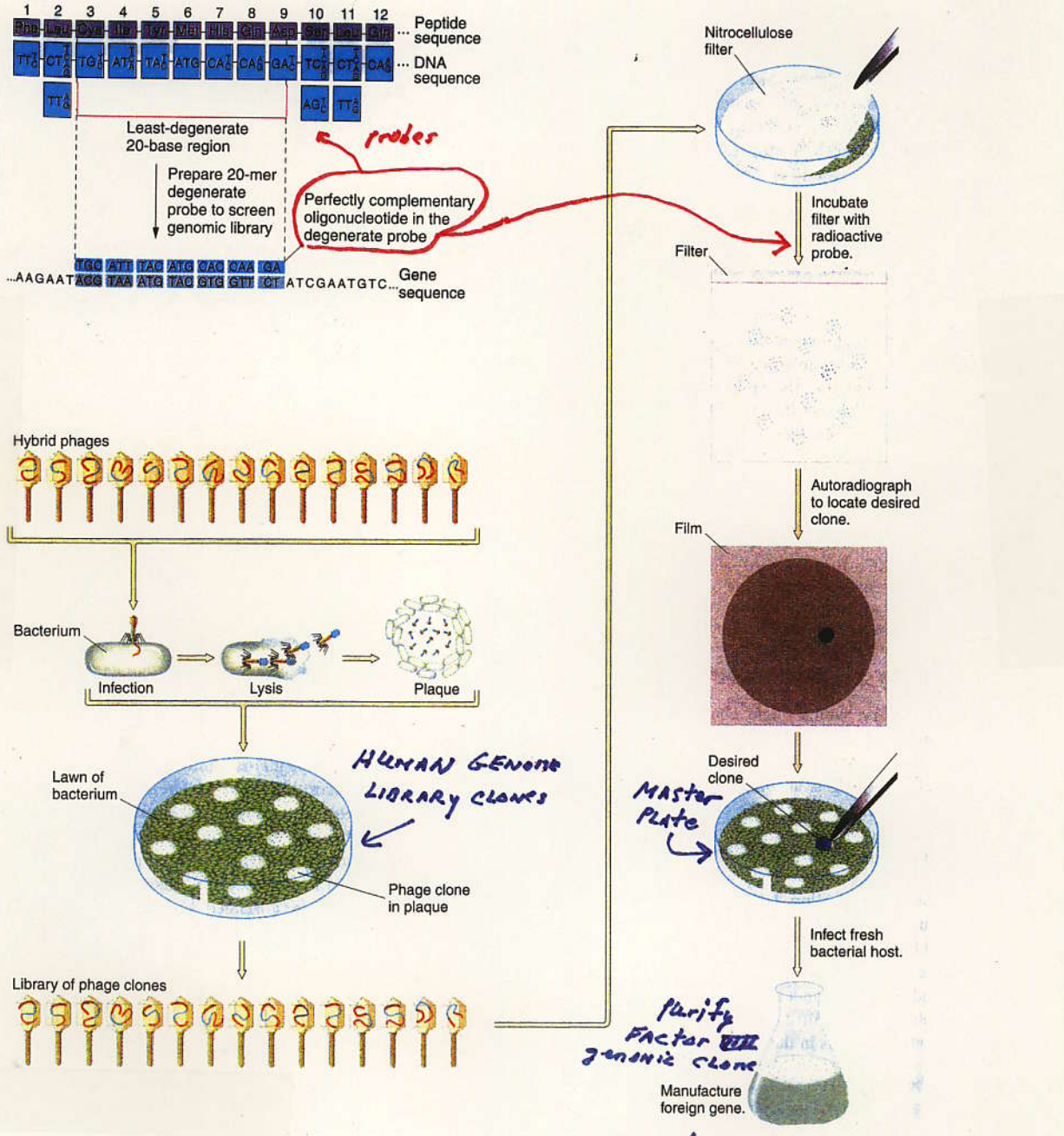
How many combination of Probes?

2 × 3 × 2 × 2 × 2 = 48

Synthetic Probes!!

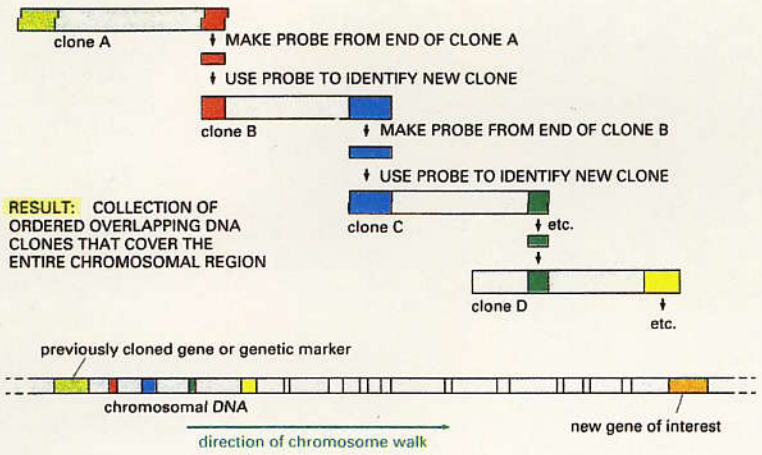
FINDING THE FACTOR VIII GENE
OR PART OF GENE!!

FACTOR VIII Protein Sequence
↳ Synthetic DNA Probe



SEQUENCE TO SEE IF IT MATCHES PROBE/MATRIX

CHROMOSOME WALKING CAN BE USED TO FIND ALL PARTS OF FACTOR VIII Gene



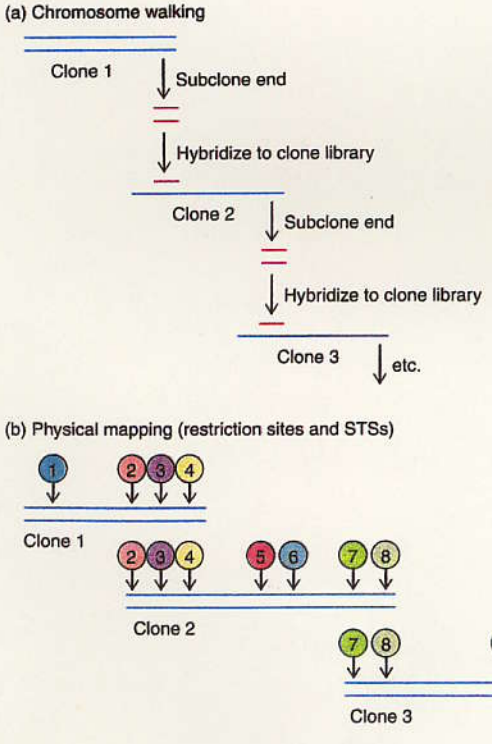
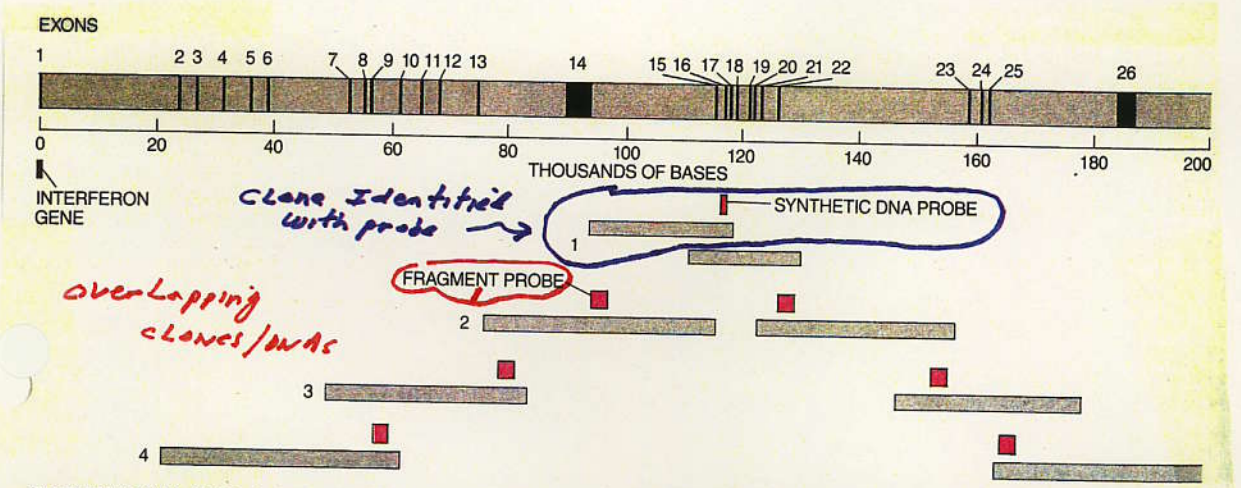


Figure 24.18 Mapping by chromosome walking. (a) Chromosome walking. To start the walk, choose a cloned piece of DNA (clone 1) and subclone one end of it. Then use this small end piece (red) as a probe to identify an overlapping clone (clone 2) in a library. Repeating the process, subclone the far end of clone 2 to generate a probe to identify yet another overlapping clone (clone 3). Repeat this cycle as many times as needed to build a set of overlapping clones spanning large stretches of DNA. (b) Physical mapping of restriction sites or STSs in each clone allows one to align the overlapping DNAs and build a map of the whole contig.

Align using Restriction Maps/Sequence of Each Clone

THE RESULT - THE FACTOR VIII GENE IS HUGE - 186,000 bp - The Probe IDENTIFIED A CLONE CONTAINING ONLY one part of Gene !!!



TREMENDOUS SIZE of the factor VIII gene, the largest gene cloned to date, forced workers to apply a cloning technique called chromosome walking. The factor VIII gene is 186,000 bases long. In contrast the interferon gene, which was cloned in 1980, incorporates only about 600 bases. Because the factor VIII gene is too large to fit into a single phage, segments of it were found in different plaques in the genomic library. When the library was screened with a synthetic DNA probe, the probe hybridized with overlapping segments (*1*). Pieces of the segments then served as probes to rescreen the library and identify further segments (*2*). By repeating this procedure nearly all of the gene was identified (*3, 4*). (Its beginning was found once factor VIII cDNA was available as a probe.) Less than one-twentieth of the gene consists of exons, or coding sequences (*black bands*); the 26 exons are separated by 25 introns.

HOW FIND CLONES WITH Rest of Gene? Key Question!

Remember - the library contains overlapping DNA clones :: CAN use one part of First clone to Re-screen library & "WALK" to other gene regions - using restriction maps & sequencing as guides!

SEQUENCE ~~~~~> GENBANK

WALKING UP AND DOWN GENES AND CHROMOSOMES

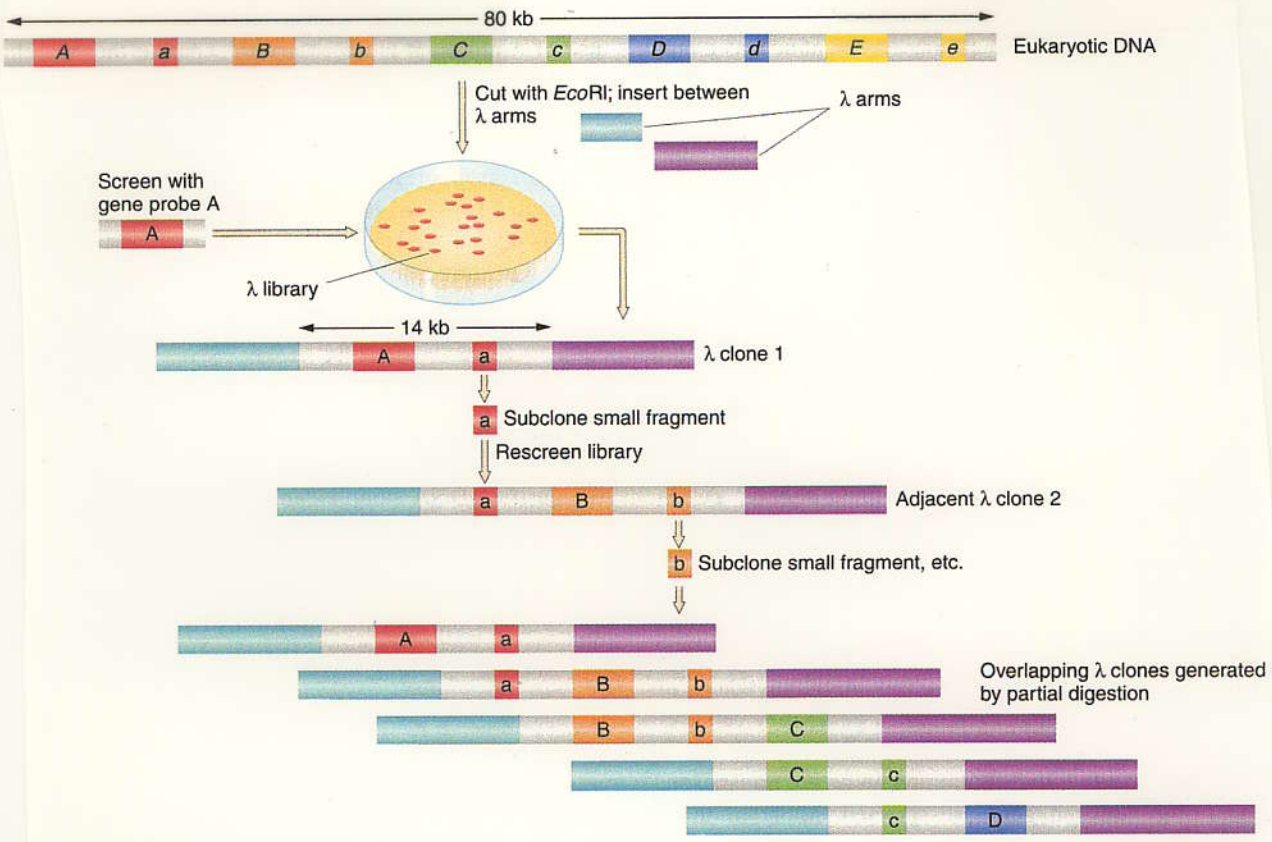


Figure 12-15 Chromosome walking. One recombinant phage obtained from a phage library made by the partial *Eco*RI digest of a eukaryotic genome can be used to isolate another recombinant phage containing a neighboring segment of eukaryotic DNA, as described in the text. (From J. D. Watson, J. Tooze, and D. T. Kurtz, *Recombinant DNA: A Short Course.* Copyright © 1983 by W. H. Freeman and Company.)

BASIS OF GENOME PROJECTS + WHOLE GENOME SEQUENCING

→ HOW KNOW FIND COMPLETE FACTOR WITH GENE?

CAN WALK DOWN AN ENTIRE CHROMOSOME & OBTAIN AN ENTIRE SET OF OVERLAPPING CLONES CONTAINING EVERY GENE IN CHROMOSOME

1. Used to Sequence Human Genome
2. Used to Map Genes to Chromosomes
3. Used FOR MARKERS (RFLPs) to identify & follow Disease Genes

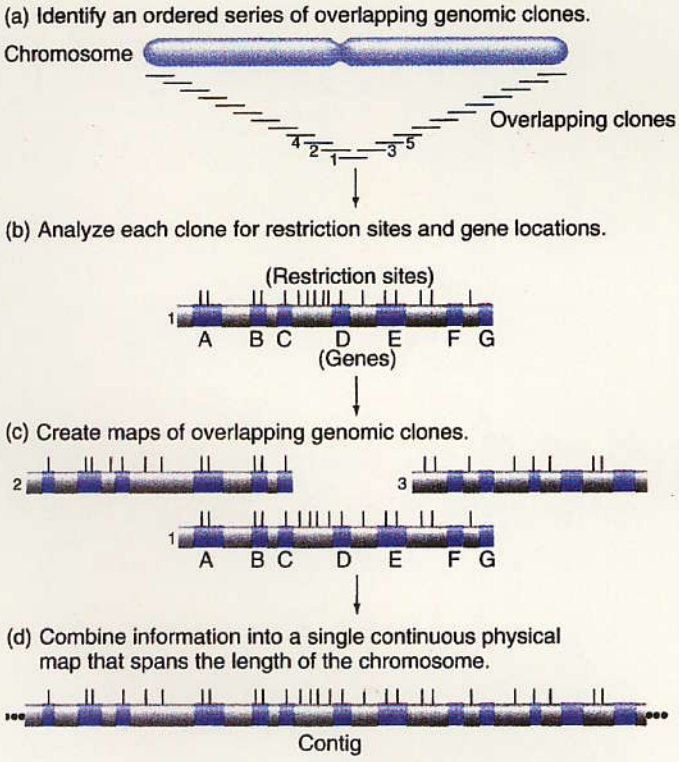
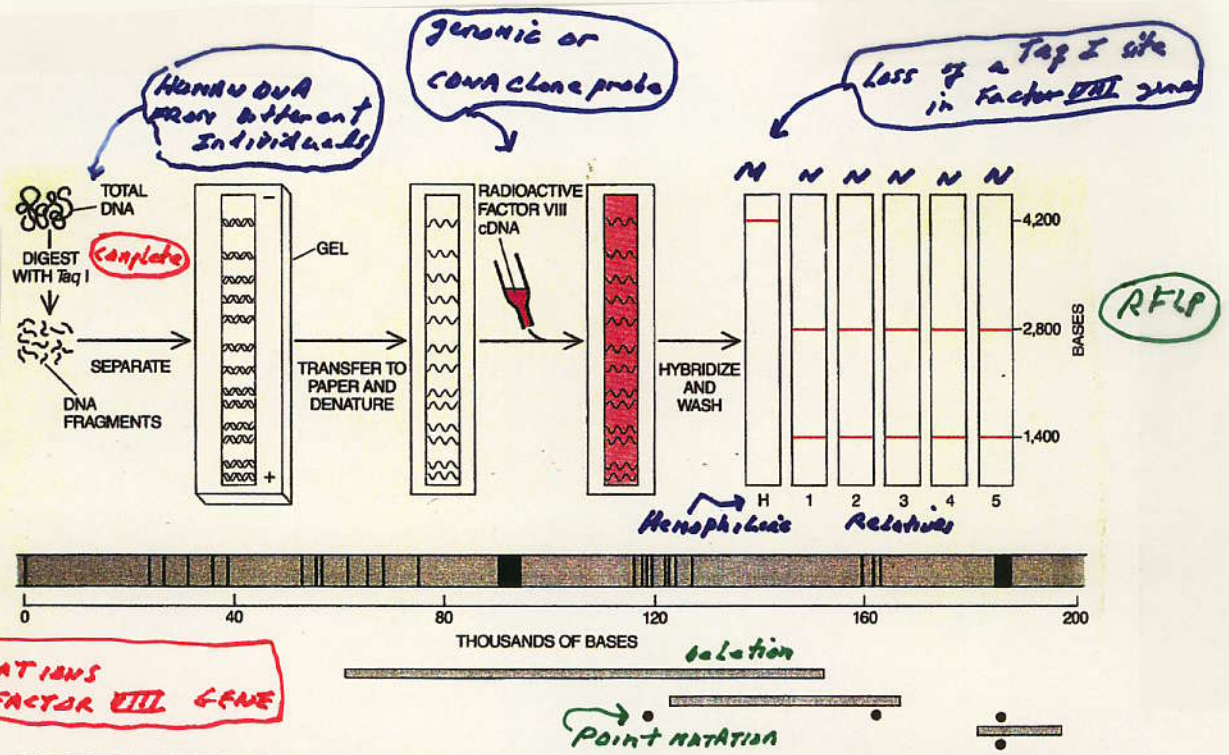


Figure 10.5 Building a whole-chromosome physical map. (a) To produce a whole-chromosome physical map, you first order a set of overlapping genomic clones that extend from one end of the chromosome to the other. Subsequent figures describe various methods of obtaining this ordered set of clones. (b) You next map the restriction sites of each clone in the set through restriction analysis, and analyze individual restriction fragments in other ways, such as Northern blot analysis, to identify transcription units. (c) Computers overlay the different types of maps for each clone onto the overlapping clones to obtain a continuous map. (d) The result is a single continuous map extending the length of the chromosome.

There are 24 sets of Clones FOR HUMAN Genome

22 Autosomes + X chromosome + Y Chromosome

FACTOR VIII GENE PROBES/SEQUENCE CAN BE USED TO CHARACTERIZE MUTANT GENES & DO DNA TESTING FOR CARRIER!

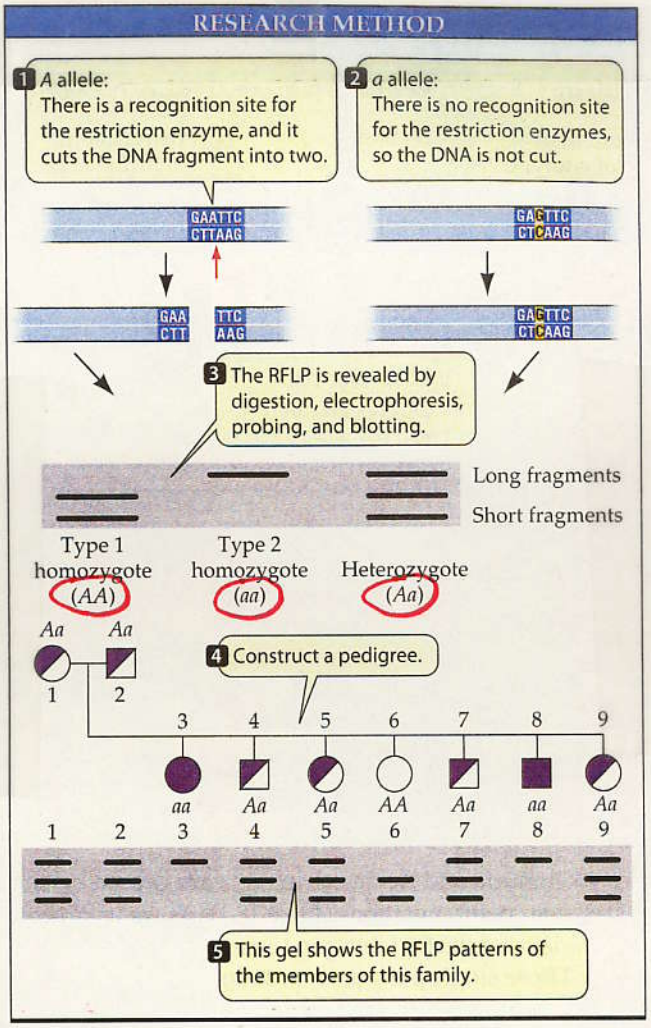


HEMOPHILIA-CAUSING MUTATIONS in the factor VIII gene can be detected by Southern blotting (*top*) if they happen to change the way the gene is fragmented by a restriction enzyme. DNA from blood cells is cut into millions of fragments, in this case with the enzyme *TaqI*. The fragments are separated according to size by electrophoresis, unraveled into single strands and blotted onto filter paper. The filter is bathed in a solution of radioactive factor VIII cDNA, which hybridizes only with fragments of the factor VIII gene. The size of the hybridizing fragments is revealed by exposing X-ray film to the filter. In the example shown here a point mutation in the factor VIII gene of a hemophiliac (*H*) has eliminated a *TaqI* cleavage site. The 2,800- and 1,400-base fragments on the blot patterns of his relatives (*1–5*) are replaced by a single, uncut 4,200-base fragment. So far seven different mutations have been located on hemophilic factor VIII genes (*bottom*). Four are point mutations, or changes of a single base (*dots*); three are extensive deletions (*bars*).

USE DNA Gel Blots & Factor VIII Probes to Investigate Presence of Mutant Alleles in FAMILIES (CARRIERS)

ONCE gene or cDNA identified!

MUTATIONS IN RESTRICTION ENZYME SITES IN GENE LEAD TO POLYMORPHISMS



18.7 RFLP Mapping
Restriction fragment length polymorphisms are differences in DNA sequences that serve as genetic markers. More than 1,000 such markers have been described for the human genome.

(1) (A)

(a) VARIANT

(2) BLOT →

a
A

ON BLOT

(3) FOLLOW GENE WITH PEDIGREE!

THESE ARE VISUALIZED USING GENE-SPECIFIC PROBES & DNA GEL BLOTS

GENE VARIABILITY AT THE DNA LEVEL SAME AS BIG/SMALL TOMATOES!

HOW IS A SPECIFIC GENE DETECTED IN GENOME?

DNA CAN BE TRANSFERRED "in situ" to PAPER & ANNEALED with RADIOACTIVE ANALOG

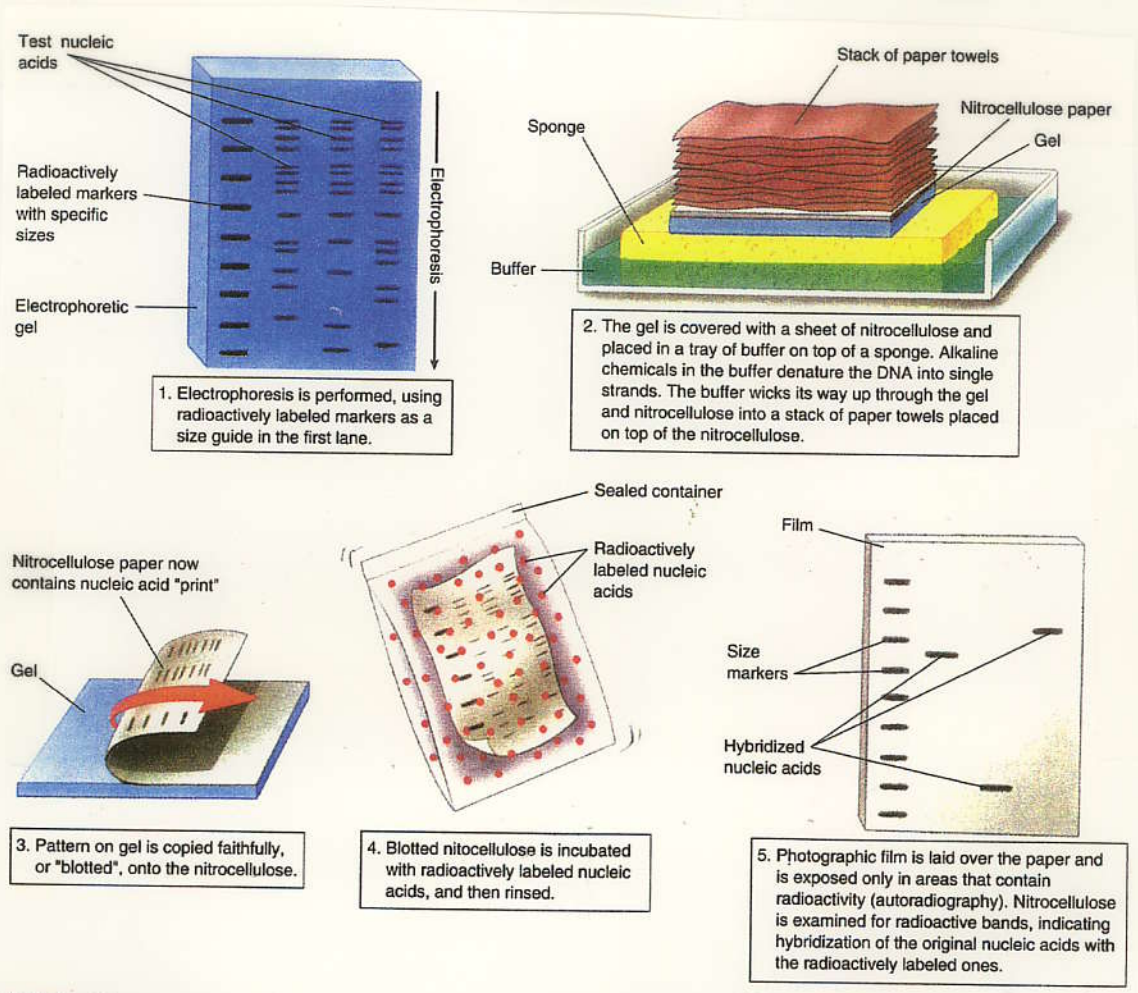


FIGURE 19.9
The Southern blot procedure. E. M. Southern developed this procedure in 1975 to enable DNA fragments of interest to be visualized in a complex sample containing many other fragments of similar size. The DNA is separated on a gel, then transferred ("blotted") onto a solid support medium such as nitrocellulose paper or a nylon membrane. It is then incubated with a radioactive single-strand copy of the gene of interest, which hybridizes to the blot at the location(s) where there is a fragment with a complementary sequence. The positions of radioactive bands on the blot identify the fragments of interest.

PROBE REPRESENTS A CLONED FRAGMENT FROM GENOME WITH A UNIQUE SEQUENCE!

USE GENE PROBE TO TEST FOR CARRIERS

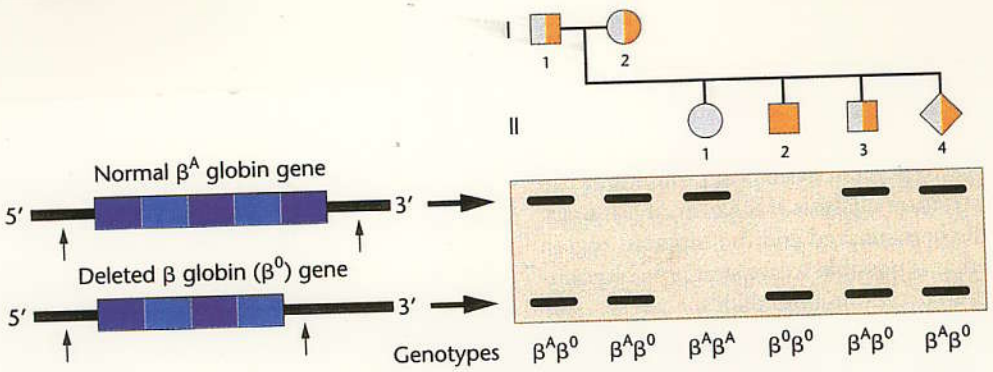


FIGURE 21.7 Diagnosis of β-thalassemia caused by a partial deletion of the β-globin gene. The family pedigree is shown positioned above each individual's genotype on a Southern blot. The normal β-globin gene (β^A) contains three exons and two introns. The deleted β-globin gene (β^0) has the third exon deleted. Arrows indicate the cutting sites for restriction enzymes used in this analysis. The normal gene produces a larger fragment (shown as the top row of fragments on the Southern blot); the smaller fragments produced by the deleted gene are represented at the bottom of the gel. The genotype of each individual in the pedigree can be determined from the pattern of bands on the blot, and these are shown below the blot.

RFLP = Restriction Fragment Length Polymorphism

Use Gene Probe to Test for Disease Gene Pre-Natally

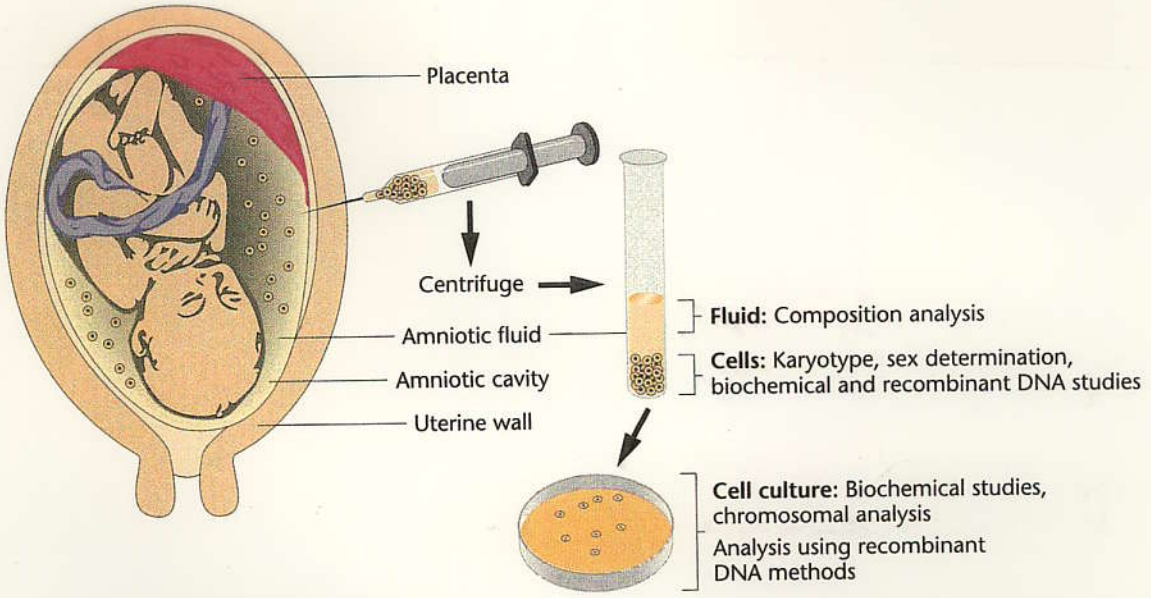


FIGURE 21.6 The technique of amniocentesis. The position of the fetus is first determined by ultrasound, and then a needle is inserted through the abdominal and uterine wall to recover fluid and fetal cells for cytogenetic and/or biochemical analysis.

MAKING THE DRUG

NEED cDNA Not Gene

FACTOR VIII GENE CAN BE USED TO FIND OUT WHERE IT IS ACTIVE USING RNA BLOTS

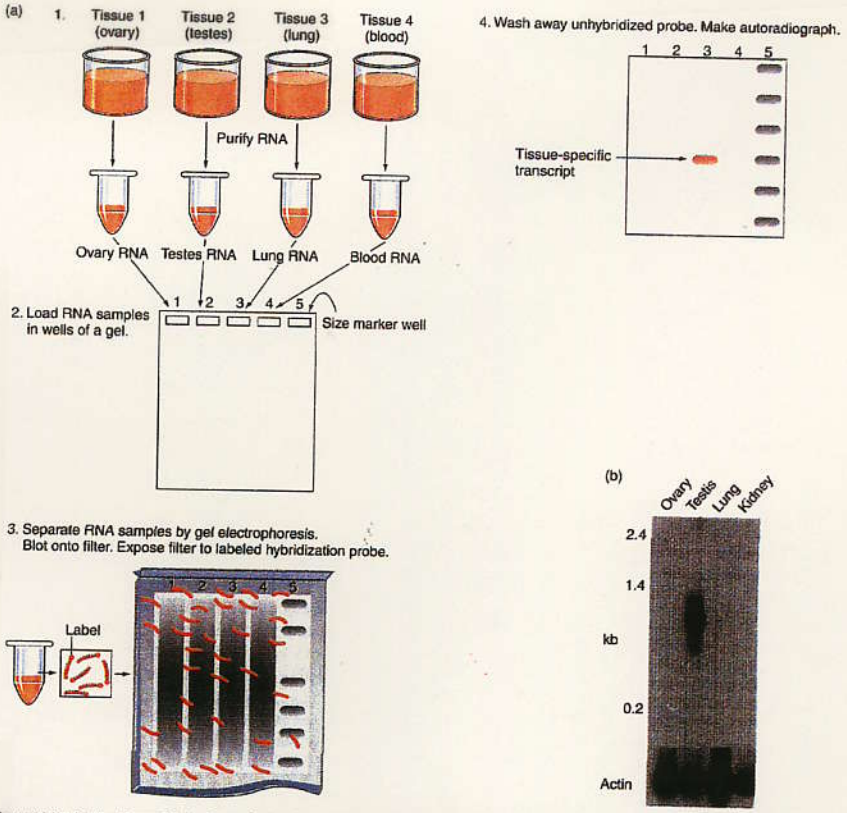
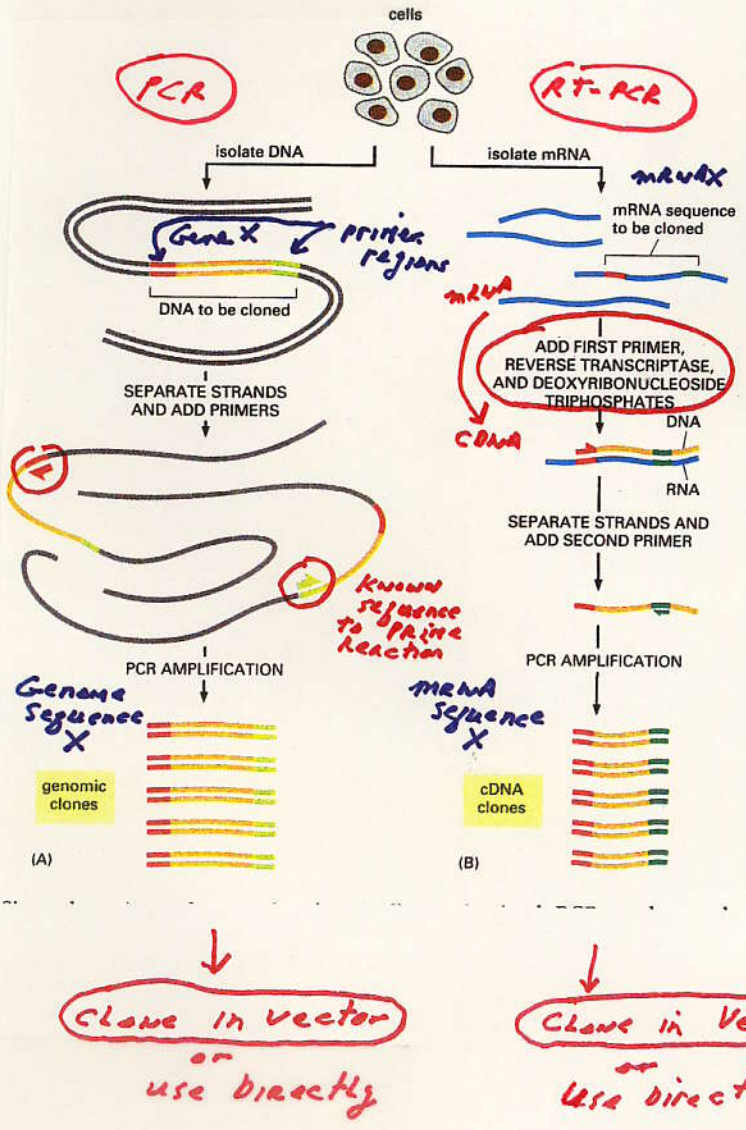


Figure 10.16 Northern blots: Snapshots of gene expression. (a) The protocol. (1) Purify RNA from each tissue to be examined for expression of the gene under investigation; here since you are looking at the *SRY* candidate for the testes-determining factor, the tissues to be examined are ovary, testes, lung, and blood. (2) Make an agarose gel and load each of the four RNA samples into a different well and load a fifth well with RNA size markers. Now subject the gel to an electric current that causes the RNA in each sample to migrate along a lane toward the bottom of the gel. The mobility of each RNA transcript in a sample depends on its size: smaller RNAs move faster, while larger RNAs migrate more slowly. When the smallest RNAs reach the bottom of the gel, turn off the current. Staining the RNAs in each lane would produce a smear reflecting the presence of so many RNAs of different sizes that they cannot be resolved from each other. (3) Blot the RNA within the gel and fix it to a filter so that each RNA molecule retains its position relative to all the other molecules. Expose the filter to labeled probe and allow the label to hybridize for several hours. (4) Wash away unhybridized probe. Place the filter on a film for autoradiography. Develop the film. You will see bands only in those lanes containing a tissue where the gene represented by the probe has been expressed. (b) Northern blot results obtained using the pY53.3 clone as a probe. This clone contains the *SRY* gene. The results show that *SRY* is expressed in the testes, but not the ovary, lung, or kidney. This result makes SRY a good candidate for the *TDF* locus. In a control experiment, researchers probed an identical blot with the same RNA samples using a clone containing the actin gene. As expected, a band of the same size appears in every lane. This control demonstrates the integrity of the RNA samples used in this study.

FACTOR VIII GENE is HIGHLY ACTIVE IN LIVER!

FACTOR VIII GENE SEQUENCE CAN BE USED TO GUIDE PRIMER synthesis FOR USE IN PCR TO AMPLIFY FACTOR VIII GENE FROM GENOME OR mRNA FROM LIVER



Clone in vector or use directly

Clone in Vector or use directly

Figure 10–23 Use of PCR to obtain a genomic or cDNA clone. (A) To obtain a genomic clone using PCR, chromosomal DNA is first purified from cells. PCR primers that flank the stretch of DNA to be cloned are added, and many cycles of the PCR reaction are completed (see Figure 10–22). Since only the DNA between (and including) the primers is amplified, PCR provides a way to obtain selectively a short stretch of chromosomal DNA in an effectively pure form. (B) To use PCR to obtain a cDNA clone of a gene, mRNA is first purified from cells. The first primer is then added to the population of mRNAs, and reverse transcriptase is used to make a complementary DNA strand. The second primer is then added, and the single-stranded DNA molecule is amplified through many cycles of PCR, as shown in Figure 10–22.

FACTOR VIII GENE PROBE CAN BE USED TO VISUALIZE ITS LOCATION ON CHROMOSOMES

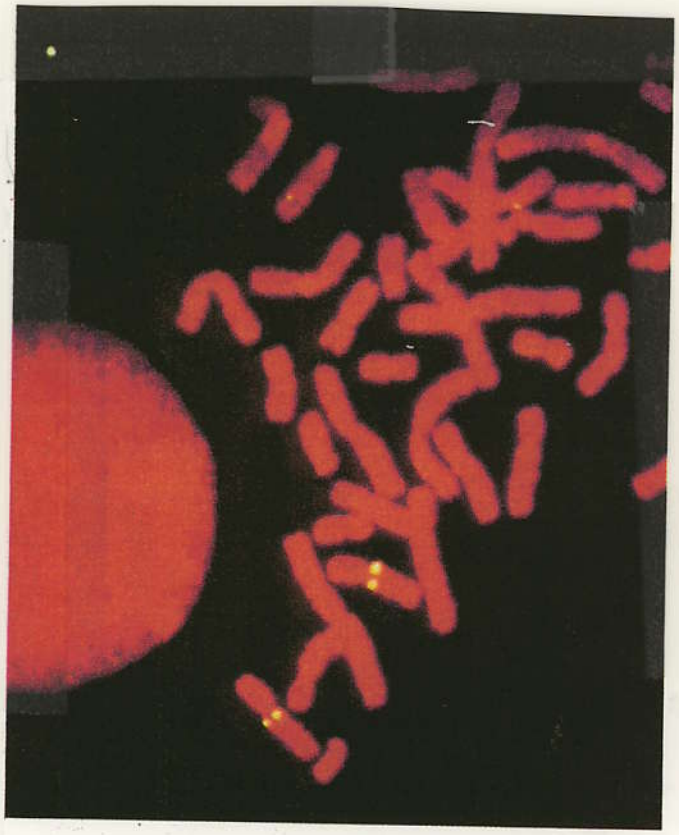
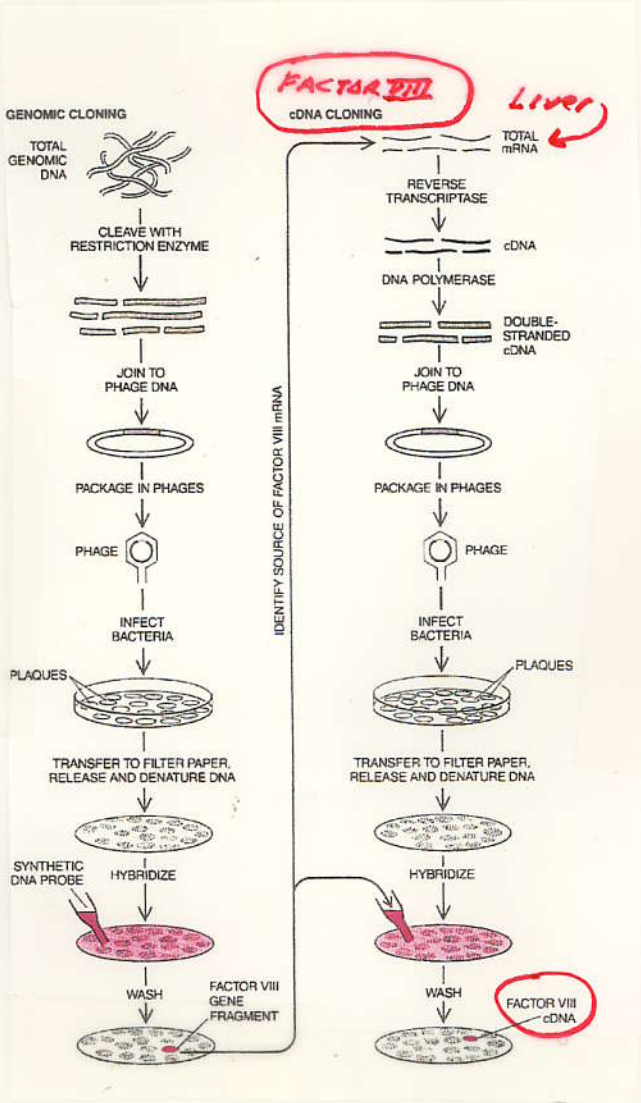


Figure 14-9 FISH analysis. Chromosomes probed in situ with a fluorescent probe specific for a gene present in a single copy in each chromosome set — in this case, a muscle protein. Only one locus shows a fluorescent spot corresponding to the probe bound to the muscle protein gene. (From P. Lichter et al., "High-Resolution Mapping of Human Chromosome 11 by in Situ Hybridization with Cosmid Clones," *Science* 247, 1990, 64.)

+ Analyze Gene Structure

USING FACTOR VIII GENE PROBE TO IDENTIFY FACTOR VIII cDNA CLONE

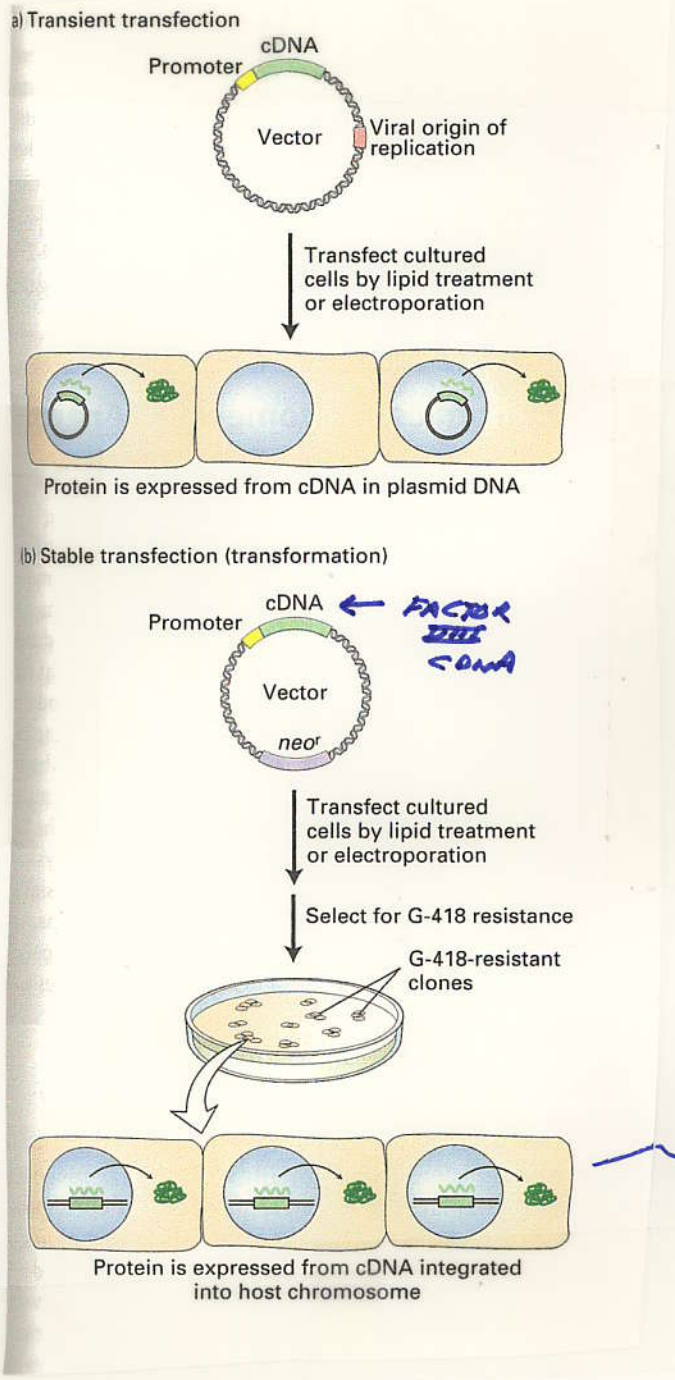


FACTOR VIII cDNA

→ USE TO MAKE DRUG!

A FACTOR VIII DRUG/"CURE"

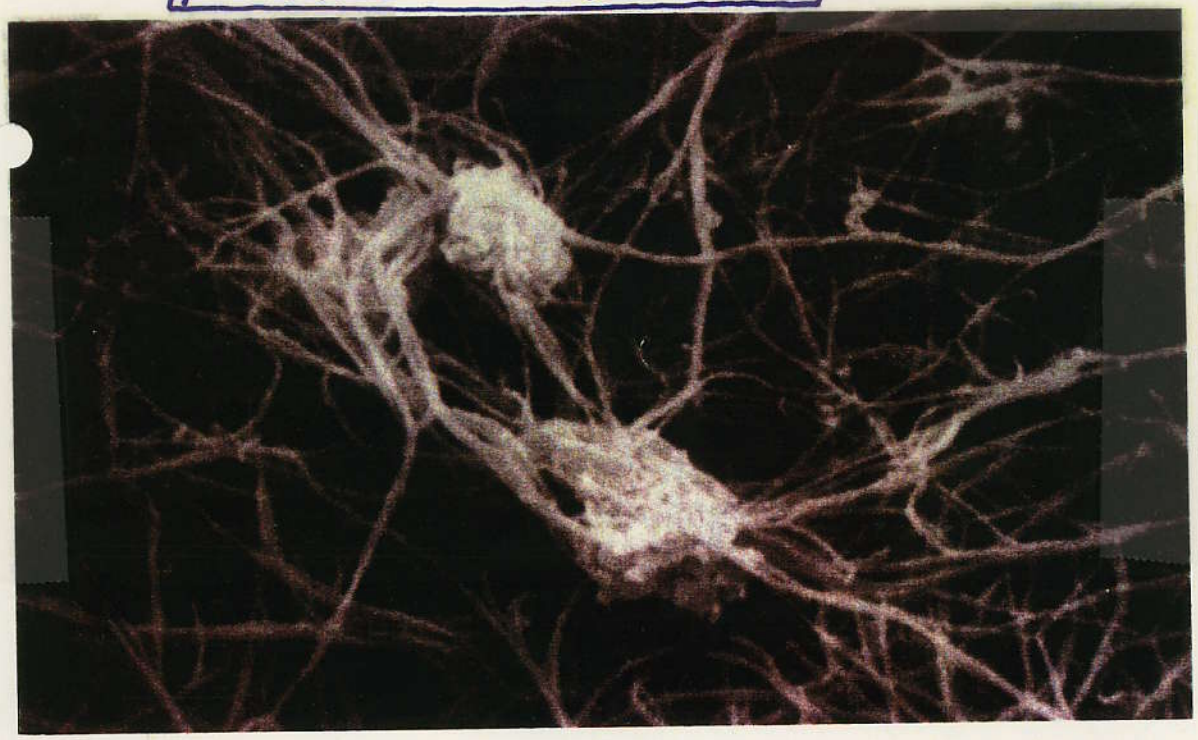
MAKING FACTOR VIII IN MAMMALIAN CELLS



Purify FACTOR VIII PROTEIN!

USING FACTOR VIII TO TREAT HEMOPHILIA

FORMATION OF A BLOOD CLOT



FIBRIN STRANDS stabilize a blood clot at the site of a wound by trapping the platelets that form the bulk of the clot. The electron ›icrograph, which was made by Jon C. Lewis of Wake Forest Uni- .rsity, shows a clot formed in a suspension of platelets and fibrin.

A clot in the bloodstream is the result of a complex cascade of enzymatic reactions culminating in the conversion of fibrinogen, a soluble protein, into insoluble fibrin strands. In hemophiliacs a crucial protein in the blood-clotting cascade is either missing or defective.

A TRIUMPH of Genetic Engineering

RECOMBINANT FACTOR VIII

Factor VIII

Active Ingredients: Antihemophilic Factor (Human)
Pronunciation: an tee hee moe fil' ik fak tir
Representative Names: AHF (Human), AHG, Alphanate, Factor VIII, Hemofil M, Humate-P, Koate-HP, Monoclate-P, Profilate HP

Who is this for?

Your doctor has ordered antihemophilic factor (human), an antihemophilic factor, to help your blood to clot. The drug will be either injected directly into your vein or added to an intravenous fluid that will drip through a needle or catheter placed in your vein for approximately 5-10 minutes. It will be given as often as your doctor determines you need it, possibly as often as every other day. Antihemophilic factor (human), a substance naturally produced in your body, activates substances in your blood to form clots and decrease bleeding episodes. This medication is sometimes prescribed for other uses; ask your doctor or pharmacist for more information. Your health care provider (doctor, nurse, or pharmacist) may measure the effectiveness and side effects of your treatment using laboratory tests and physical examinations. It is important to keep all appointments with your doctor and the laboratory. The length of treatment depends on how your symptoms respond to the medication.

FACTOR VIII STORY – SUMMARY

1. Purify small amounts of FACTOR VIII
2. Obtain complete or partial amino acid sequence
3. Use the Genetic Code & synthesize short DNA probes
4. Isolate Factor VIII DNA fragments complementary to probe using Genome Library
5. Determine if Have Entire Gene – If not, "walk" to obtain overlapping DNA fragments that collectively contain the entire FACTOR VIII Gene
6. Sequence DNA fragments to find Entire Factor VIII Gene Sequence – Compare with Protein Sequence
7. Use a Factor VIII Genome Probe to find the body tissue/cell type where Factor VIII Gene Expressed
8. Make cDNA Library from this tissue/cell type & Isolate FACTOR VIII cDNA clone.
9. SEQUENCE FACTOR VIII cDNA clone & Compare with Gene Sequence – Describe anatomy of Factor VIII gene – introns, exons, switches
10. Use Factor VIII cDNA as probe to find RFLP Marker for Disease Gene
11. Use FACTOR VIII cDNA to make Factor VIII in mammalian cells for use as drug (SK)