HC 70A Winter 2004

Professor Bob Goldberg

Lecture #4 – Nuts & Bolts of Genetic Engineering

THE FACTOR VIII STORY

Themes / Concepts

1. How to identify a specific gene / cDNA clone
2. Hemophilia
3. Inheritance of Hemophilia
4. Finding Genes + cDNAs
5. Nuts & Bolts of Cloning – Tools, Restriction Enzymes, Vectors
6. Libraries
7. Making Genome Libraries / Overlapping Fragments
8. Protein Sequence to Synthetic Probes
9. Walking to Find Genes
10. Using Probe to Find Carriers / RFLPs
11. Finding cDNA + Making Drugs

STOP 2/3/04

THINKING ABOUT THE CONSEQUENCES OF GMOs

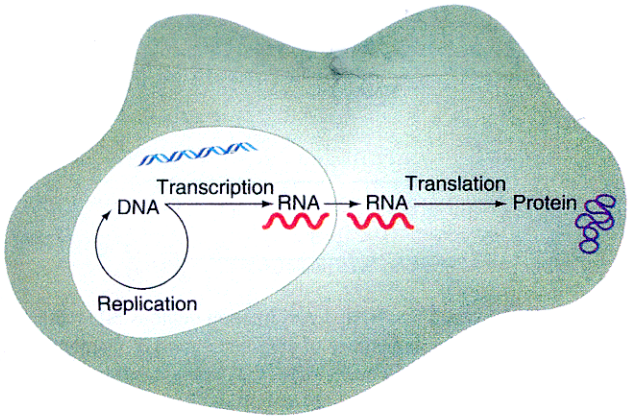


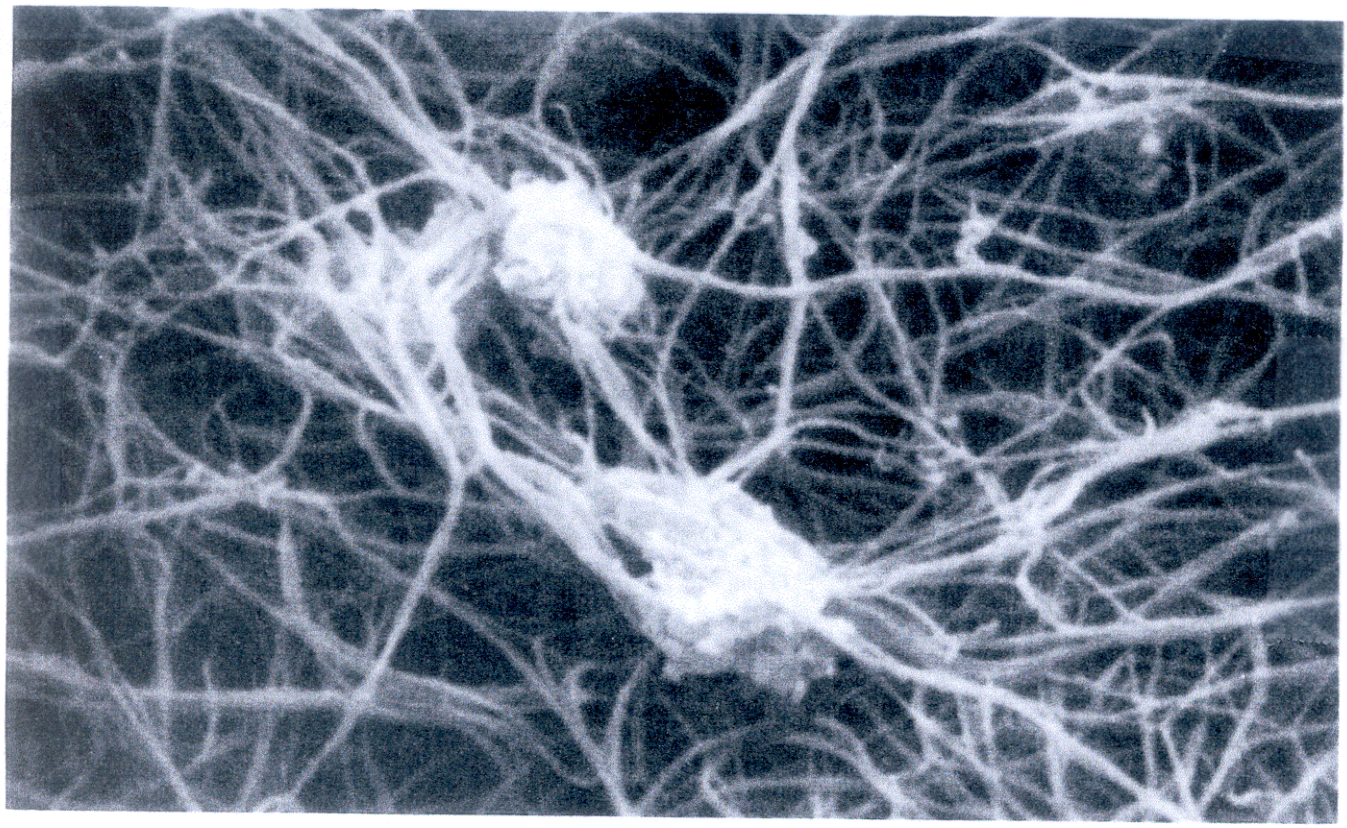
Figure 10-2 The three processes of information transfer: replication, transcription, and translation.

Need Science-based Questions & Science-based solutions

There is NO HOCUS POCUS all hypotheses are testable!!

1. What is a Gene?
2. What is the Anatomy of a Gene?
3. How does the Gene Replicate?
4. How does the Gene direct Synthesis of a Protein?
5. Does the Gene work independently of other Genes?
6. What is the Sequence & Structure of the Protein?
7. How does it work in Cell?
8. Does the Protein Structure imply any potential "Harm"
9. Does the Gene Change the organism? Fitness?

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

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.

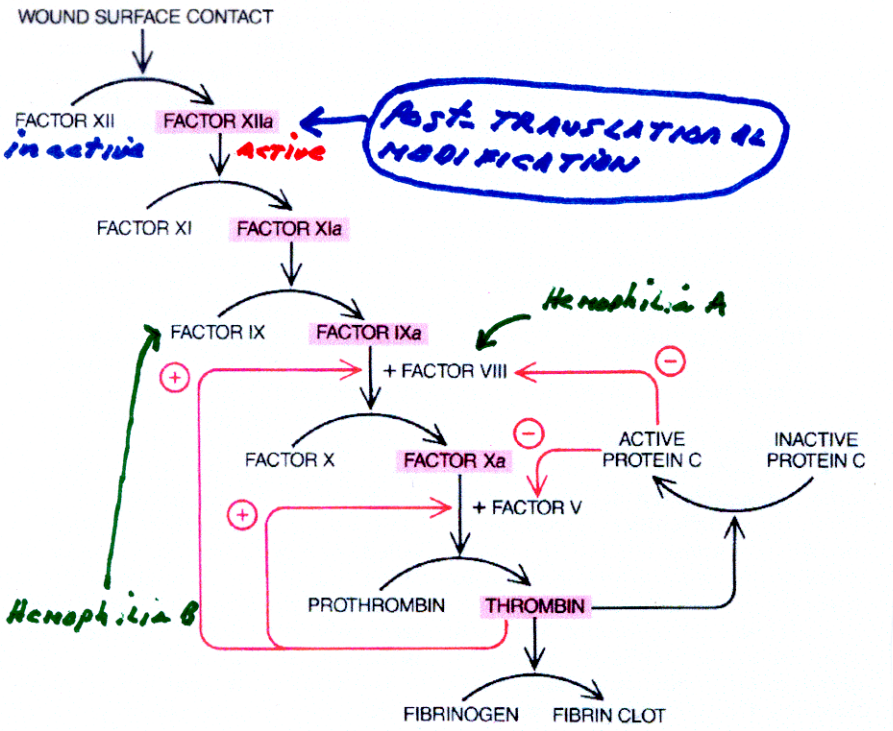
Bible

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.

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!

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

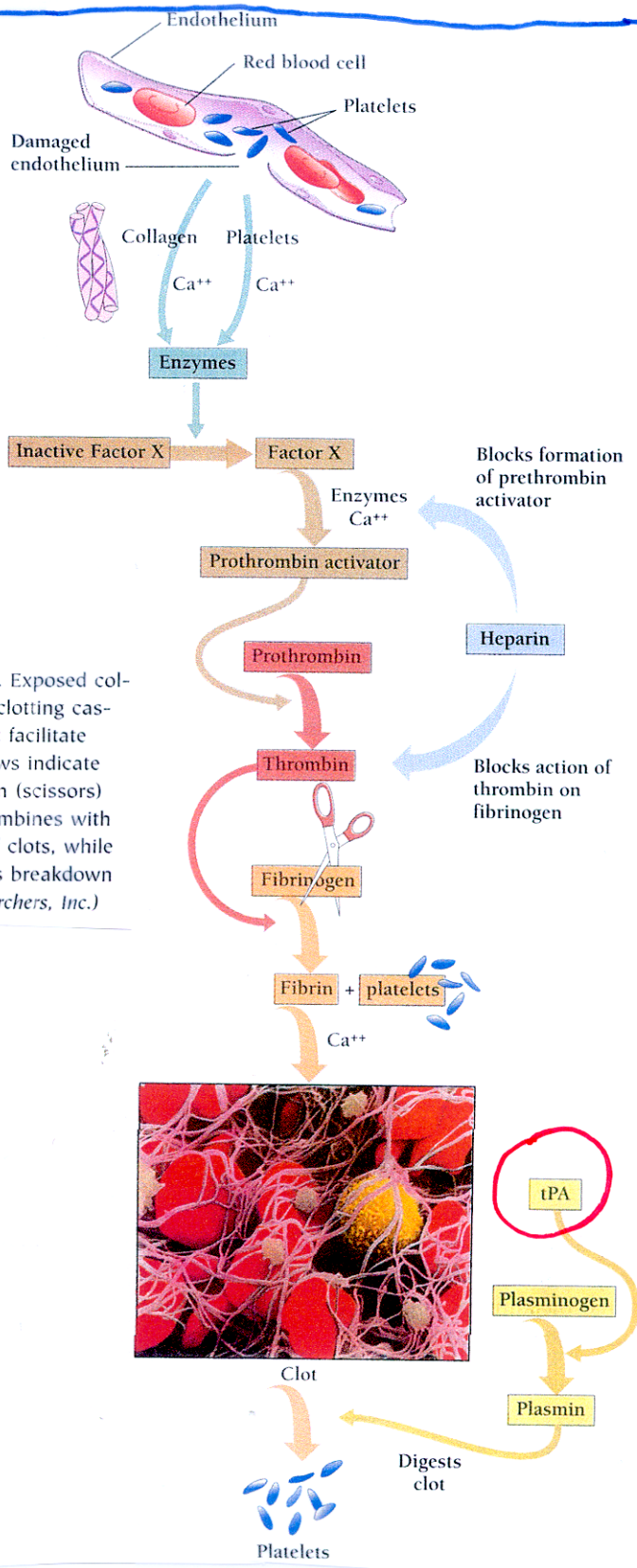


Figure 40-5 **Making and unmaking blood clots.** Exposed collagen or blood platelets trigger the first steps in the "clotting cascade." Thin arrows indicate the work of enzymes that facilitate transformation of one molecule into another. Fat arrows indicate the transformation. For example, the enzyme 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?

Hemophilia A & B ARE Sex-Linked GENES

ROYAL HEMOPHILIA Gene

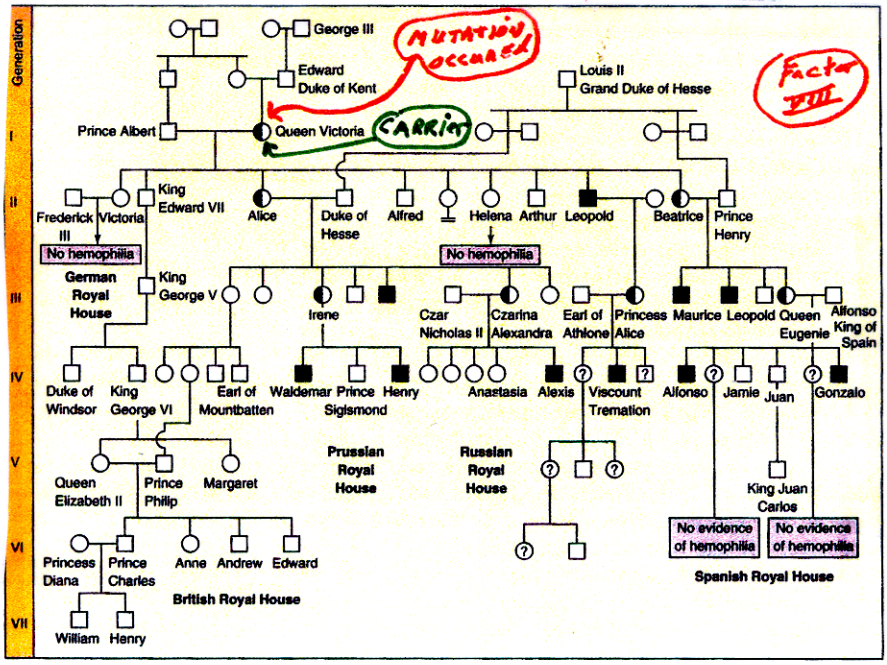
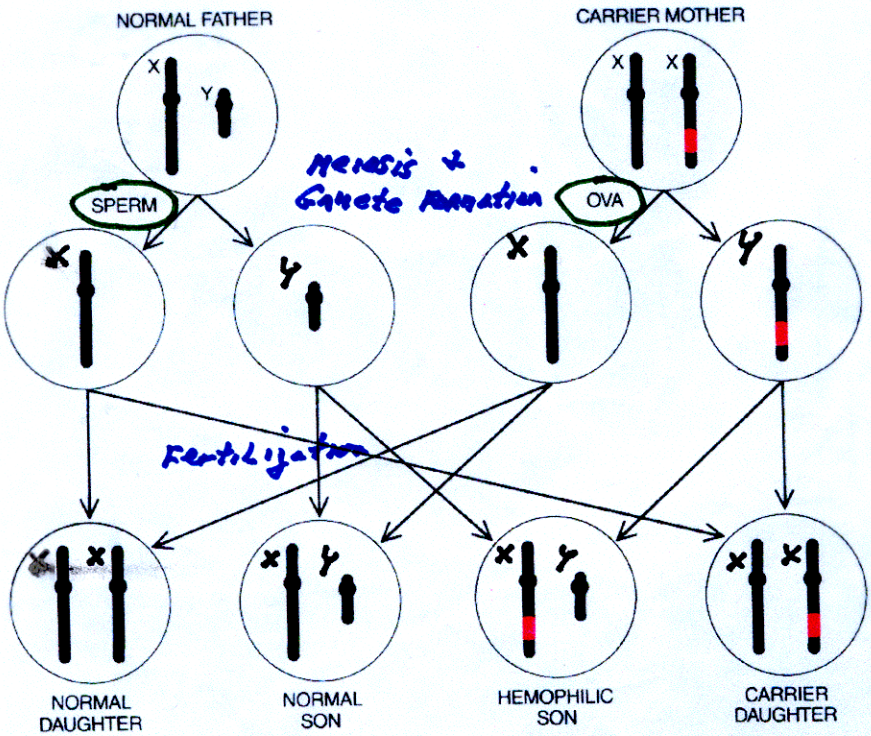


FIGURE 13.26

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



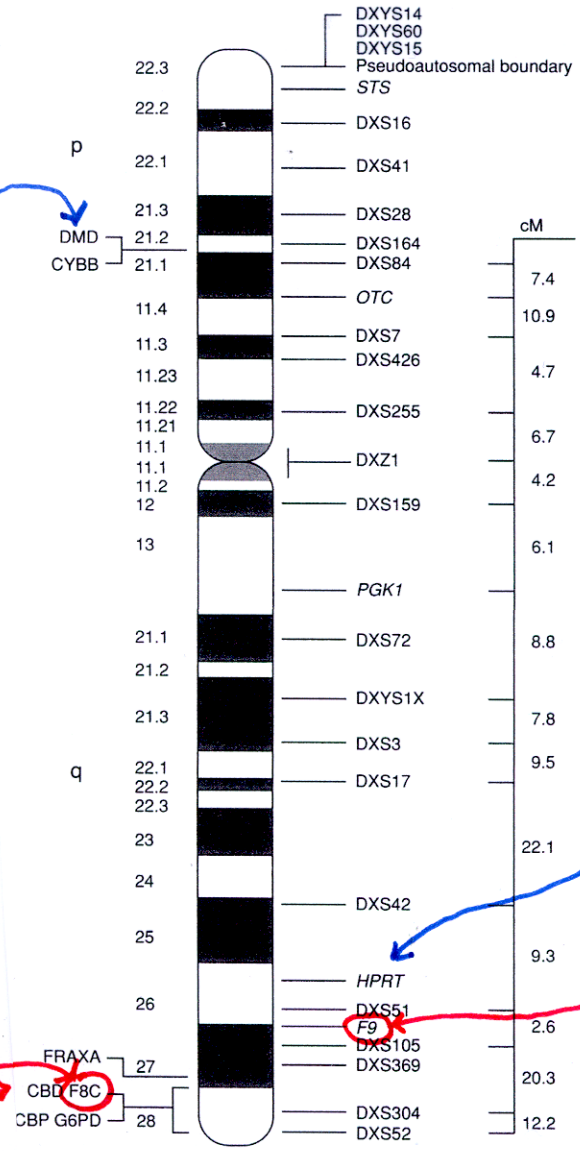
SEX-LINKED INHERITANCE of hemophilia results from the location of the factor VIII gene on the X chromosome. A male carrying a mutant factor VIII gene lacks normal factor VIII and is hemophilic. A female carrier is protected by the normal gene on her second X chromosome, but half of her daughters will 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!

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



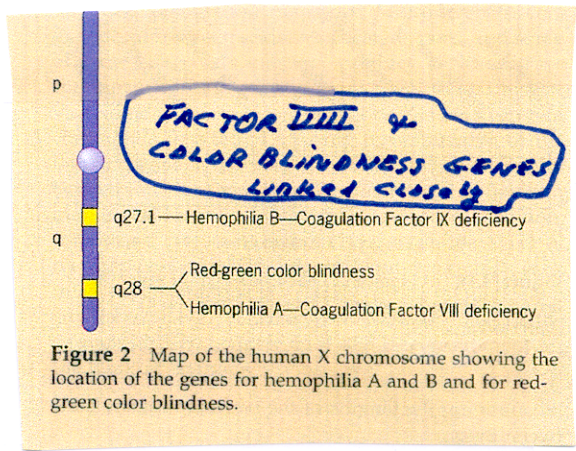


Figure 2 Map of the human X chromosome showing the location of the genes for hemophilia A and B and for red-green color blindness.

FACTOR VIII & COLOR BLINDNESS GENES LINKED CLOSELY

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

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.

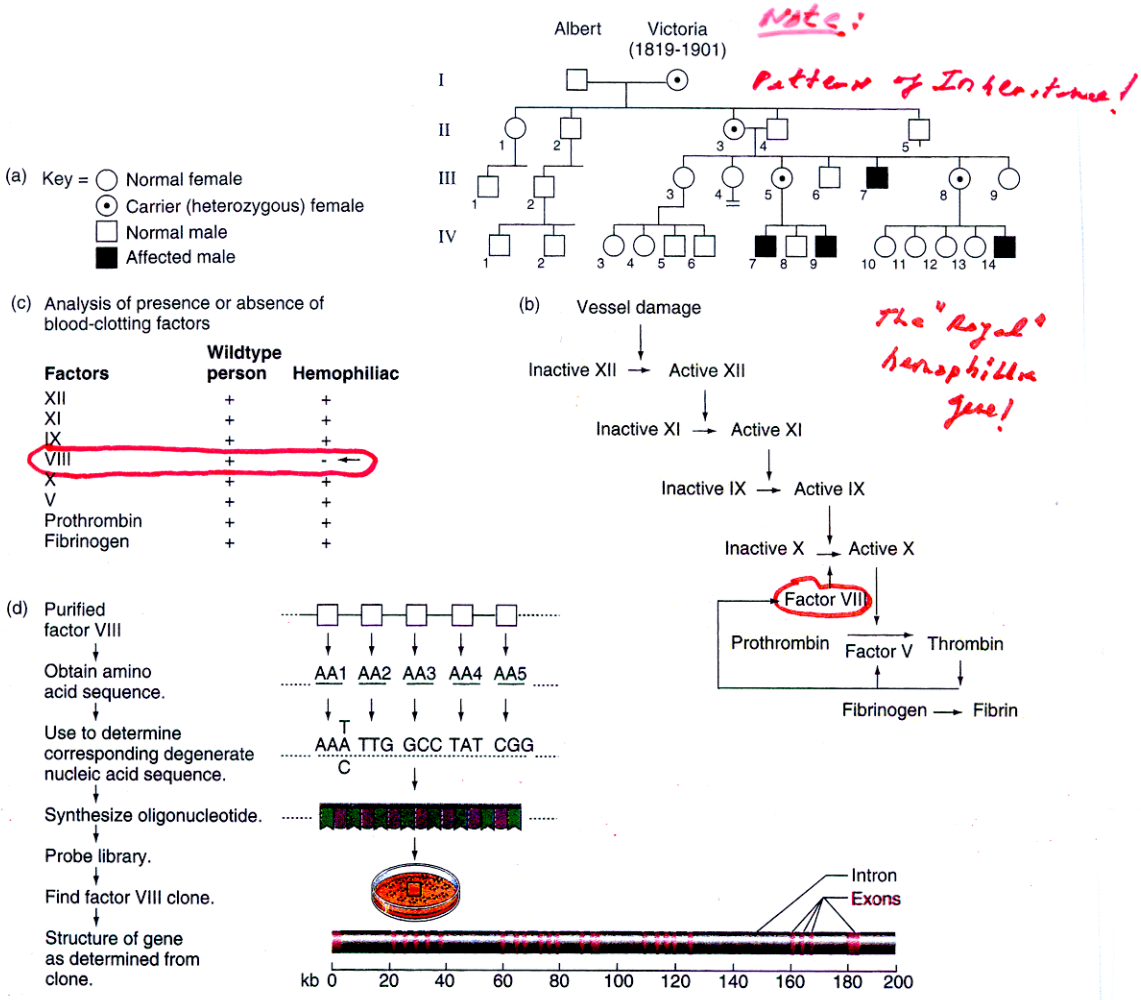


Figure 10.1 How geneticists identified the hemophilia A gene. (a) A pedigree of the royal family descended from Queen Victoria. This family tree uses the standard pedigree symbols. Black boxes represent males with hemophilia. (b) The blood clotting cascade. Vessel damage induces a cascade of enzymatic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

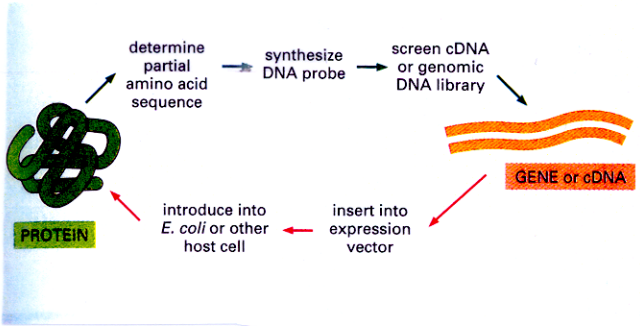


Figure 8–44 Knowledge of the molecular biology of cells makes it possible to experimentally move from gene to protein and from protein to gene. A small quantity of a purified protein is used to obtain a partial amino acid sequence. This provides sequence information that enables the corresponding gene 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.

∴ PROTEIN → GENE

OR

GENE → PROTEIN
USING SEQUENCING
and Genetic Code

GENBANK

2004 – 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

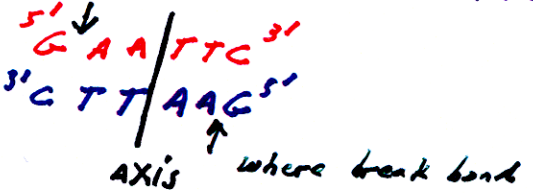
What are the properties of restriction enzymes?
A review

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

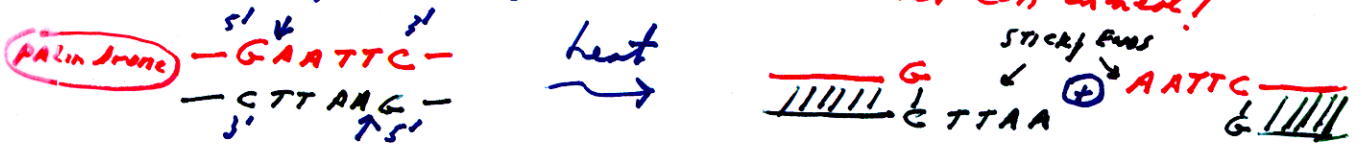
(2) Bind double-stranded DNA molecules only - linear & circular

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

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



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



Because recognition sites are present in all DNAs - DNA from different "sources" CAN be joined together (2 fragments)

(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!!

There ARE MANY DIFFERENT RESTRICTION ENZYMES

TABLE 7-1 Selected Restriction Endonucleases and Their Restriction-Site Sequences

Source Microorganism	Enzyme*	Recognition Site (↓)†	Ends Produced
Arthrobacter luteus	*Alu*I	AG↓CT 4 4p	Blunt
Bacillus amyloliquefaciens H	*Bam*HI	G↓GATCC 6	Sticky
Escherichia coli	*Eco*RI	G↓AATTC 6	Sticky
Haemophilus gallinarum	*Hga*I	GACGC+5↓ 5	‡
Haemophilus influenzae	*Hind*III	A↓AGCTT 6	Sticky
Haemophilus parahaemolyticus	*Hph*I	GGTGA+8↓ 5	‡
Nocardia otitiscaviaruns	*Not*I	GC↓GGCCGC 8	Sticky
Staphylococcus aureus 3A	*Sau*3AI	↓GATC 4	Sticky
Serratia marcesens	*Sma*I	CCC↓GGG 6	Blunt
Thermus aquaticus	*Taq*I	T↓CGA 4	Sticky

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, AluI was the first restriction enzyme to be isolated from Arthrobacter luteus*).*

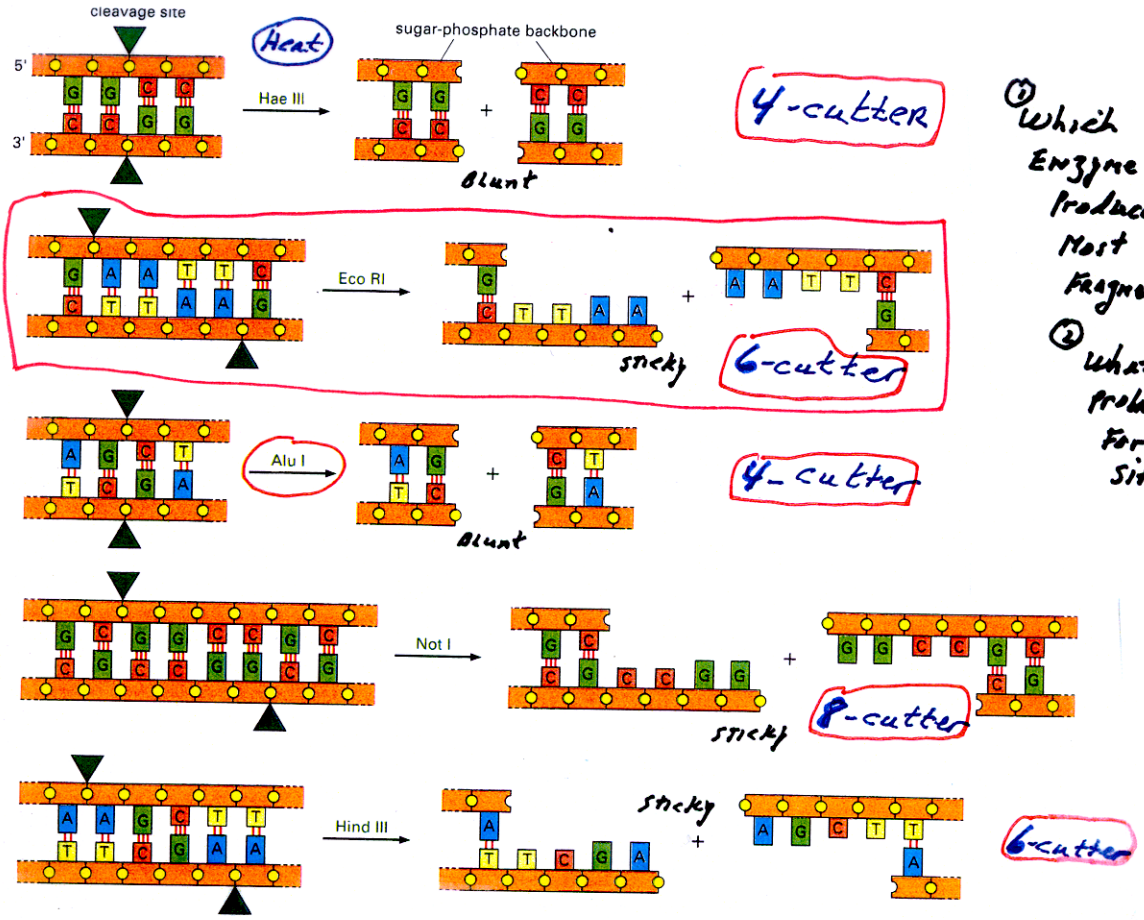
†*Recognition sequences are written 5′→3′ (only one strand is given), with the cleavage site indicated by an arrow. Enzymes producing blunt ends cut both strands at the indicated site; those producing stick ends make staggered cuts, with cleavage occurring between the same nucleotides in each strand as shown in Figure 7-5a.*

‡*The cleavage sites for* HphI *and* HgaI *occur several nucleotides away from the recognition sequence.* HgaI *cuts five nucleotides 3′ to the GACGC sequence on the top strand and ten nucleotides 5′ to the complementary GTGCG sequence on the bottom strand.* HphI *cuts eight nucleotides 3′ to the GGTGA sequence on the top strand and seven nucleotides 5′ to the complementary CCACT sequence on the bottom strand.*

SOURCE: R. J. Roberts, 1988, *Nucl. Acids Res.* 16(suppl):271.

Frequency of sites $\propto \frac{1}{\#\text{ bases recognized}}$

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

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

(5) **Anthropology**

Human Lineages
Population Diversity
Presence of Specific Pathogens

PROVIDES SPECIFIC FRAGMENT IDENTITY

CAN BE USED IN COMBINATION with PCR

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

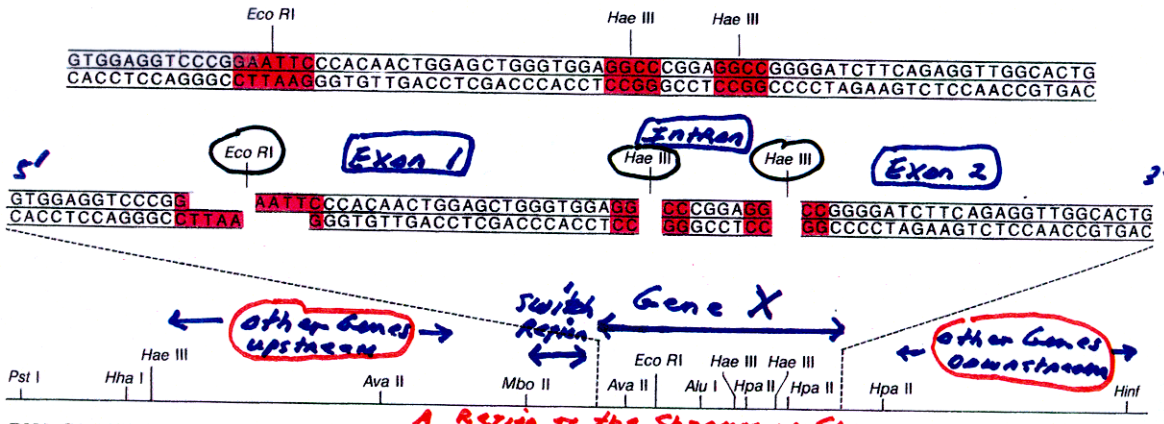
(3) Maps of Chromosomes can be used to:
a. Mark-Map Gene Locations
b. Identify Specific Chromosomes (eg, Y chromosome)
c. Identify Regions Containing Known Genes from other Studies - Markers for Genes

(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!

A RESTRICTION MAP PROVIDES GUIDEPOSTS FOR IDENTIFYING AND MANIPULATING Genes.



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

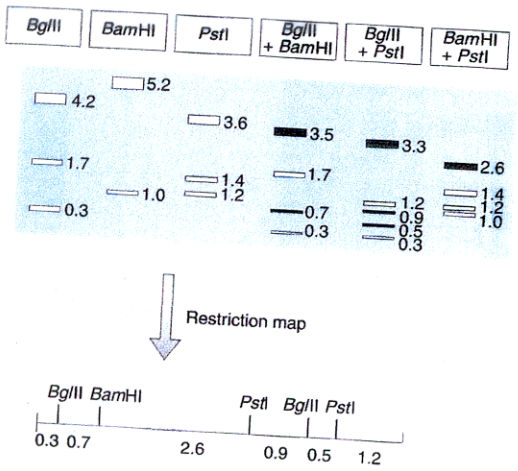


Figure 4.9: Generating a restriction map.

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