

Techniques in Molecular Biology

(chapter 7)

Nucleic acid separation – gel electrophoresis

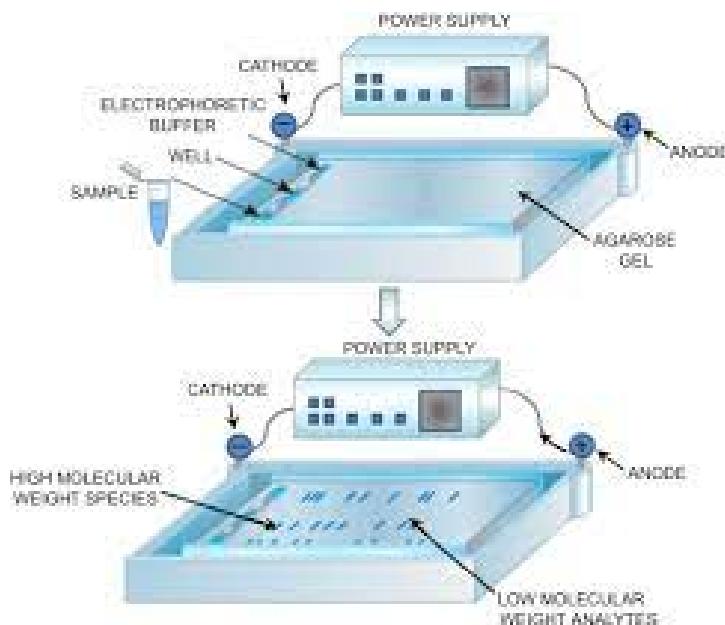
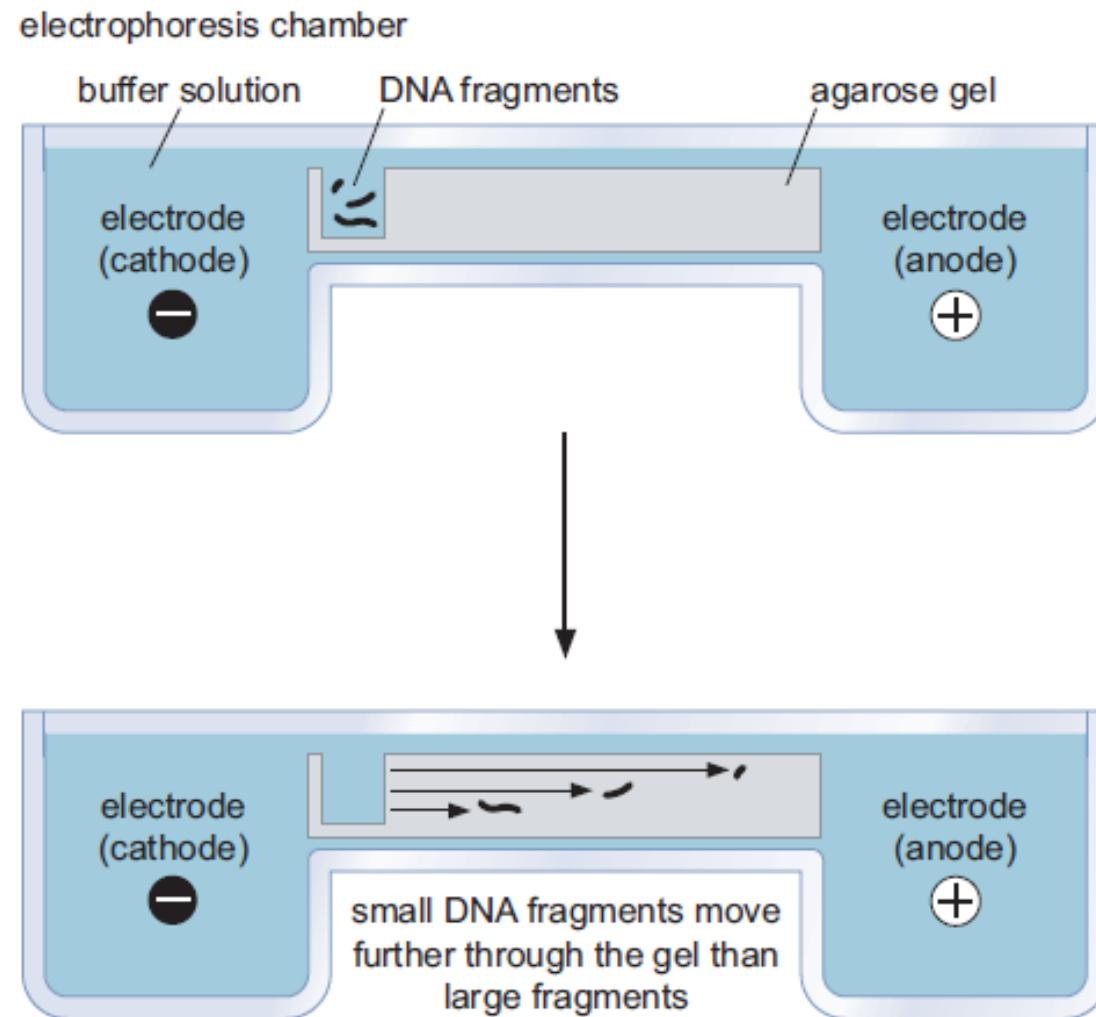
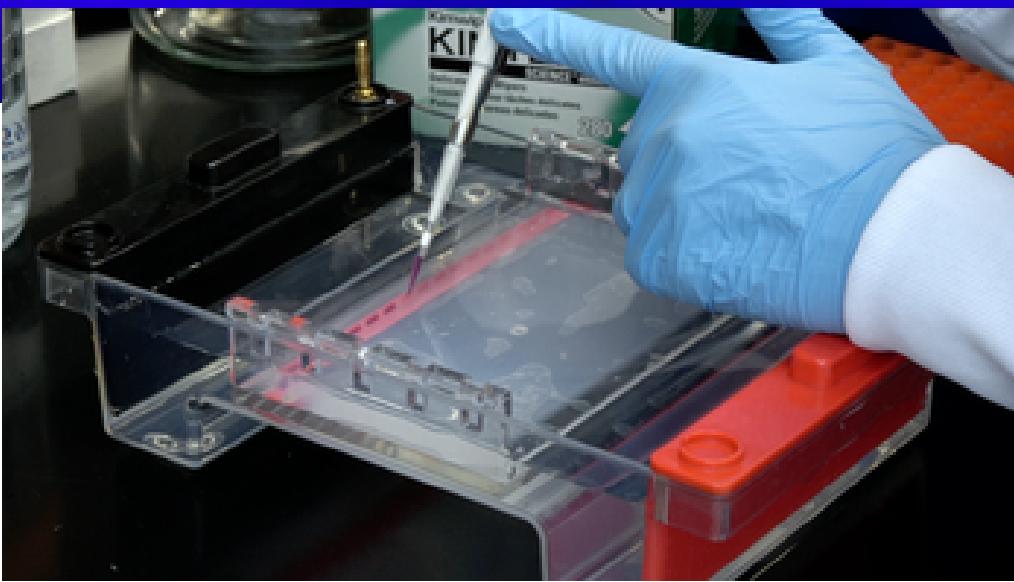
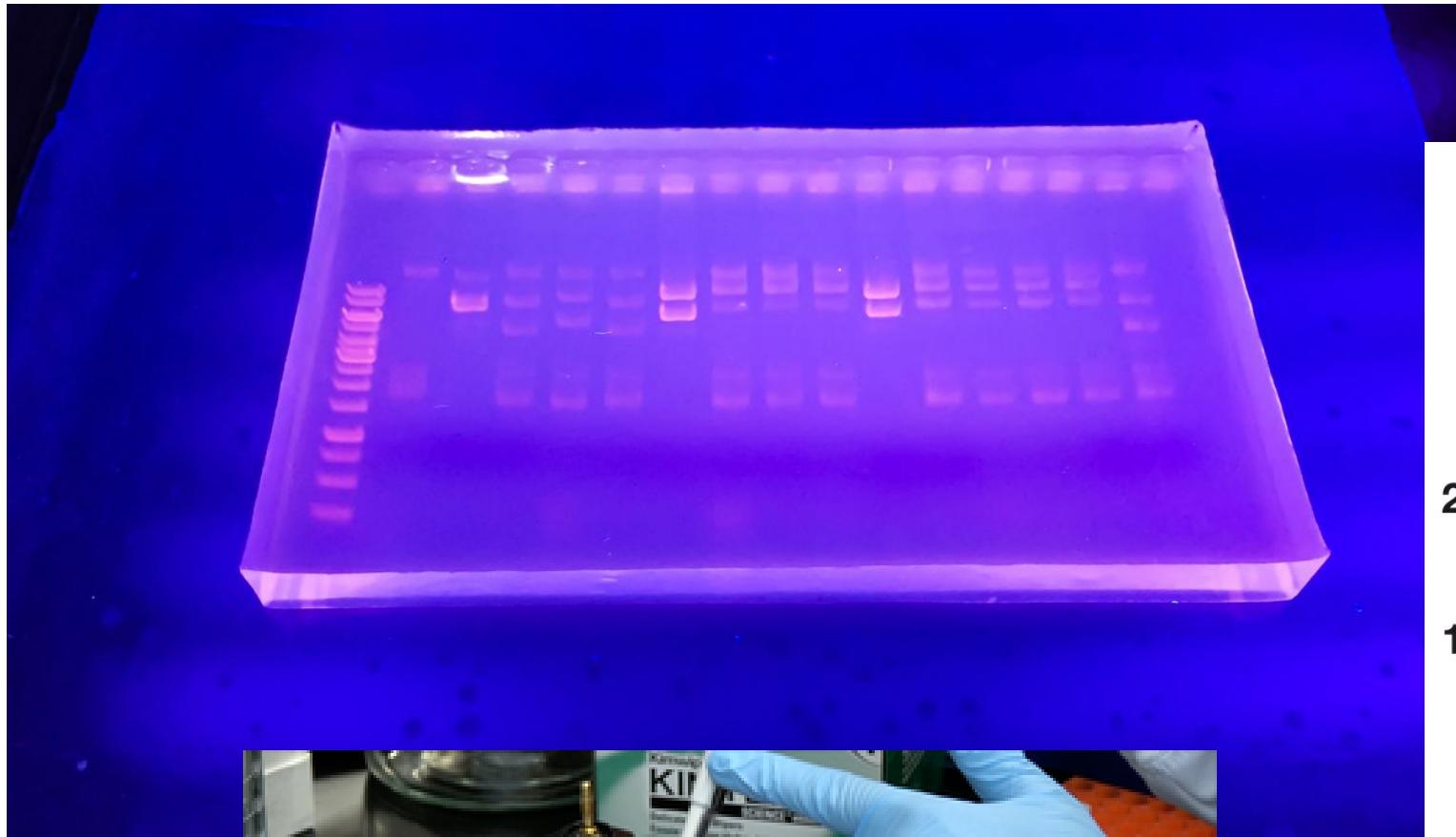
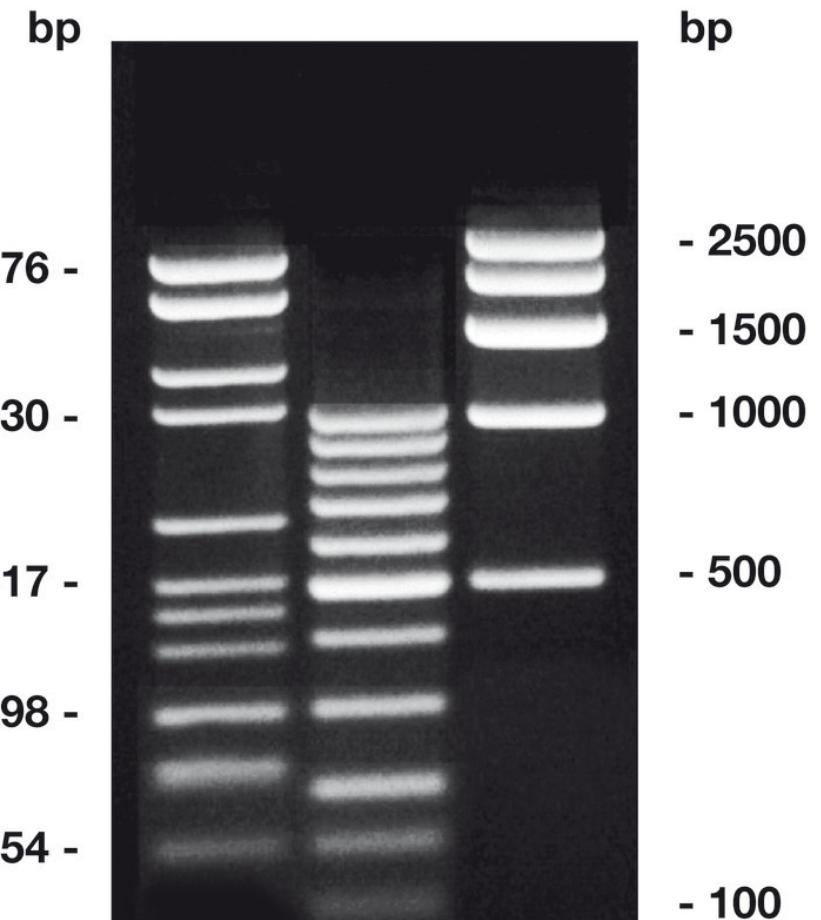


FIGURE 7-1 DNA separation by gel electrophoresis. The figure shows a gel from the side in cross section. The "well" into which the DNA mixture is loaded is indicated at the left, at the top of the gel. This is also the end at which the cathode of the electric field is located, the anode being at the bottom of the gel. As a result, the DNA fragments, which are negatively charged, move through the gel from the top to the bottom. The distance each DNA travels is inversely related to the size of the DNA fragment, as shown. (Adapted, with permission, from Micklos D.A. and Freyer G.A. 2003. *DNA science: A first course*, 2nd ed., p. 114. © Cold Spring Harbor Laboratory Press.)





1.5 % agarose in 1 x TAE



Nucleic acid separation – gel electrophoresis

- Linear DNA molecules are separated according to size
- Subjected to electric field through a gel matrix
 - Agarose gel
 - Polyacrylamide gel
- Visualization
 - **EtBr** – ethidium bromide, **PI**- propidium iodide; red fluorescences; they bind to the DNA
 - **Fluorescent dyes** (FITC- green, TRIC - red, DAPI - blue; it binds to DNA also, TexasRed- red etc)



**Six brilliant colors.
Three species of tropical fish.**

Starfire Red®
Danio Tetra (NEW)

Cosmic Blue®
Danio Tetra (NEW)

Electric Green®
Danio Tetra Barb

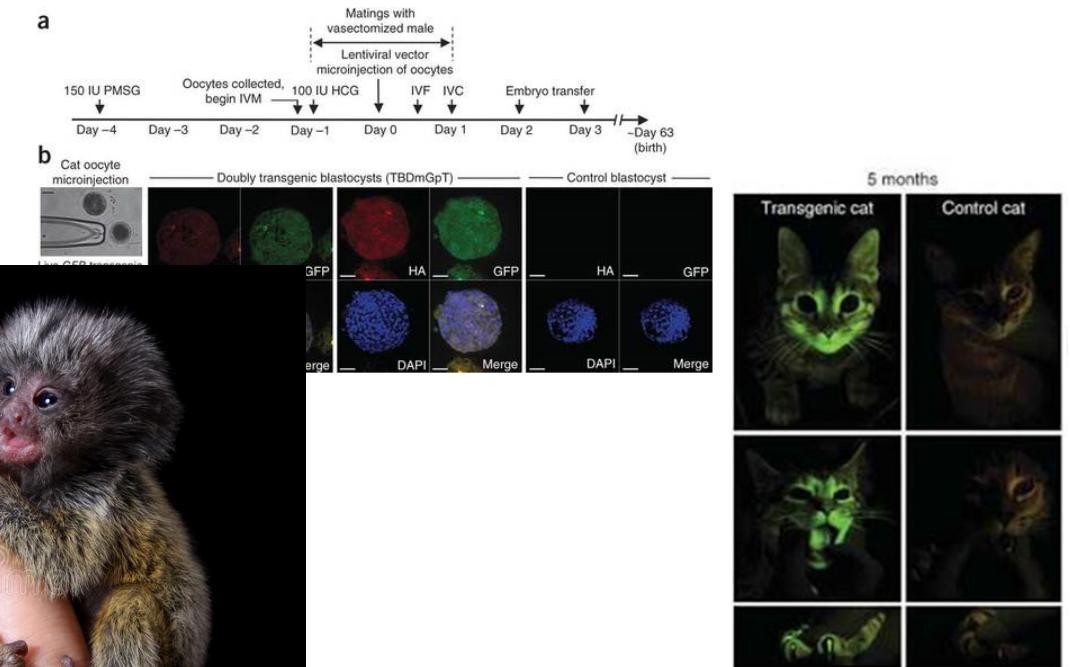
Sunburst Orange®
Danio

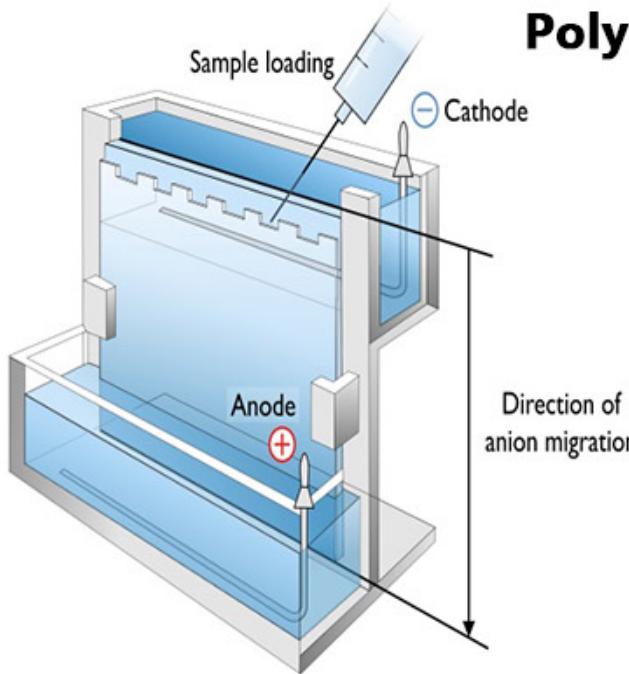
Galactic Purple®
Danio Tetra

Moonrise Pink®
Tetra

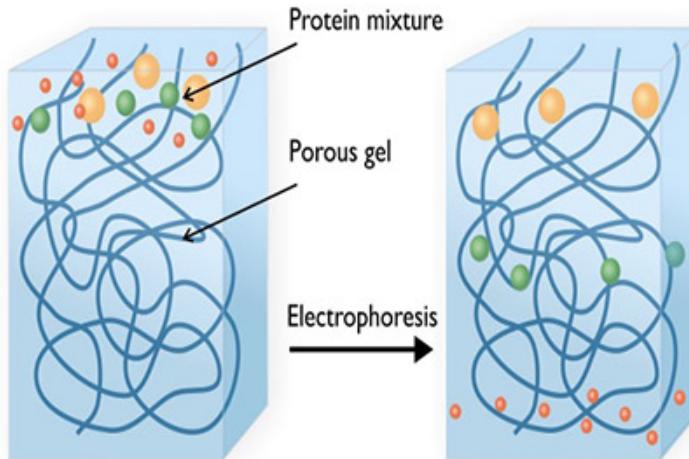


Marmoset monkeys
(*Callithrix jacchus*)

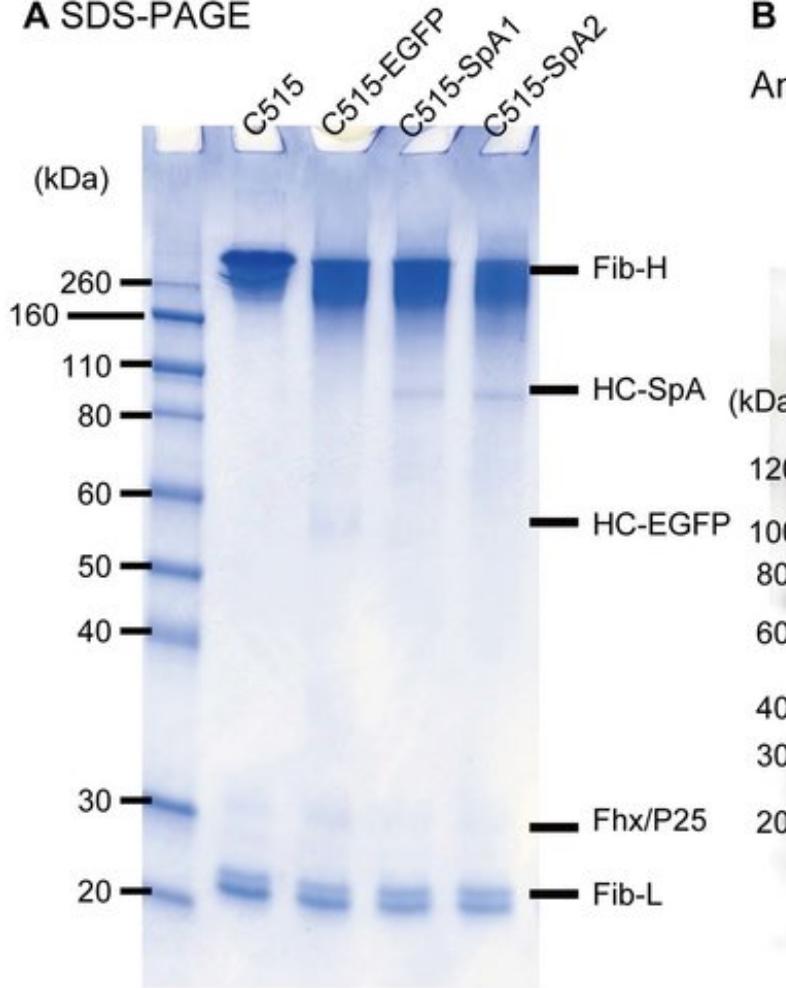




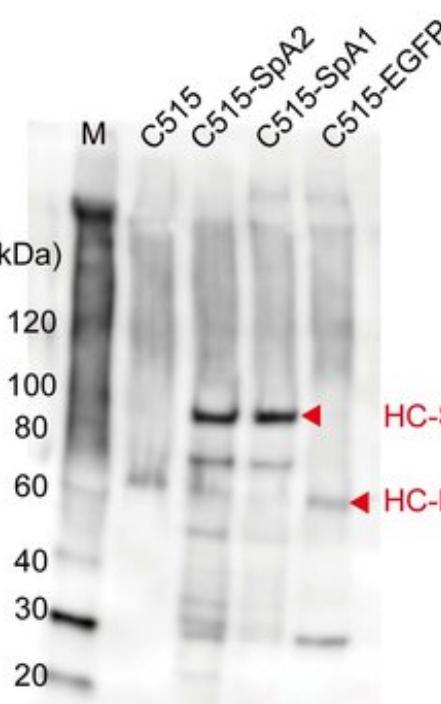
Polyacrylamide Gel Electrophoresis (PAGE)



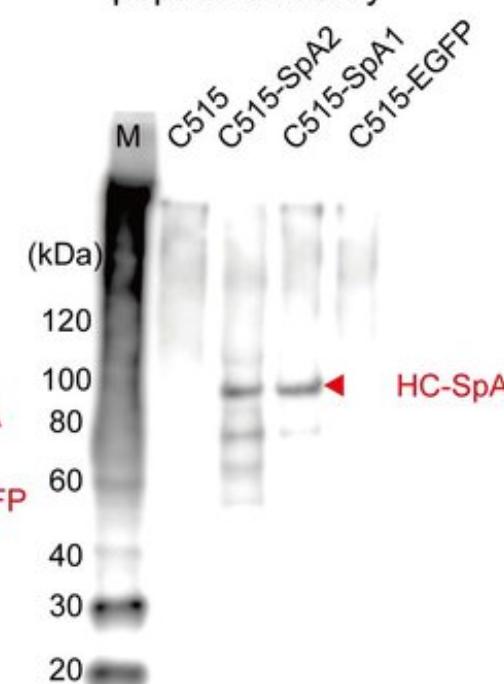
- Electrophoresis through **agarose or polyacrylamide gels** is a standard method used to **separate, identify and purify biopolymers**, since both these gels are porous in nature.
- Polyacrylamide gels are chemically cross-linked gels formed by the polymerization of acrylamide with a cross-linking agent, usually N,N'-methylenebisacrylamide.
- The reaction is a free radical polymerization, usually carried out with ammonium persulfate as the initiator and N,N,N',N'-tetramethylethylenediamine (TEMED) as the catalyst.
- **Polyacrylamide gel electrophoresis (PAGE)** is a technique widely used in biochemistry, forensic chemistry, genetics, molecular biology and biotechnology to **separate biological macromolecules, usually proteins or nucleic acids, according to their electrophoretic mobility**.
- The most commonly used form of polyacrylamide gel electrophoresis is the **Sodium dodecyl sulphate Polyacrylamide gel electrophoresis (SDS- PAGE)** used mostly for the separation of proteins.

A SDS-PAGE**B Western blot**

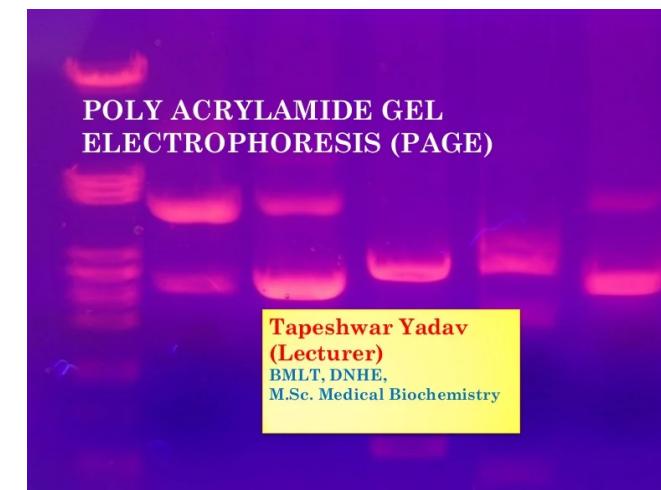
Anti-6xHis tag antiserum



Anti-SpA-specific peptide antibody



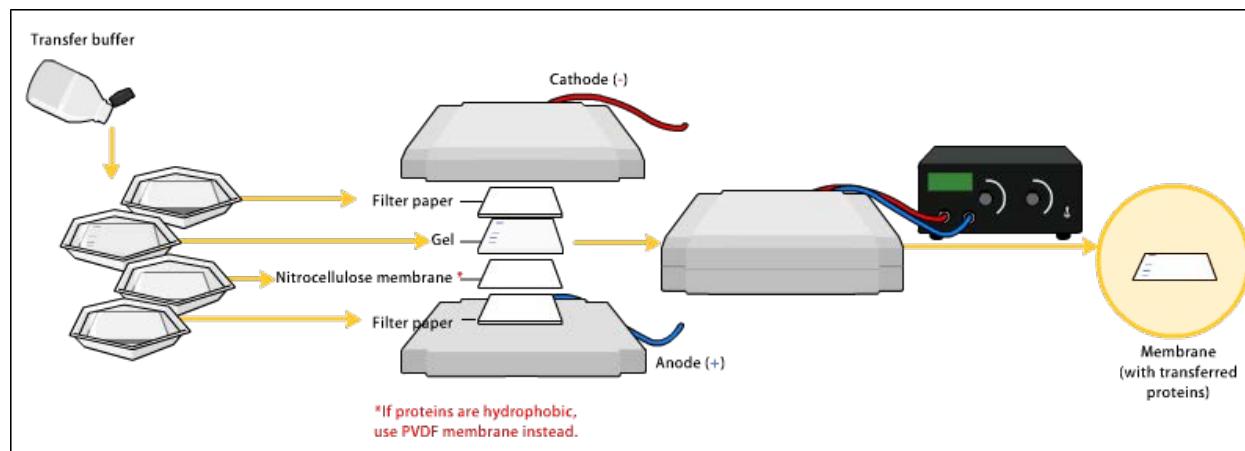
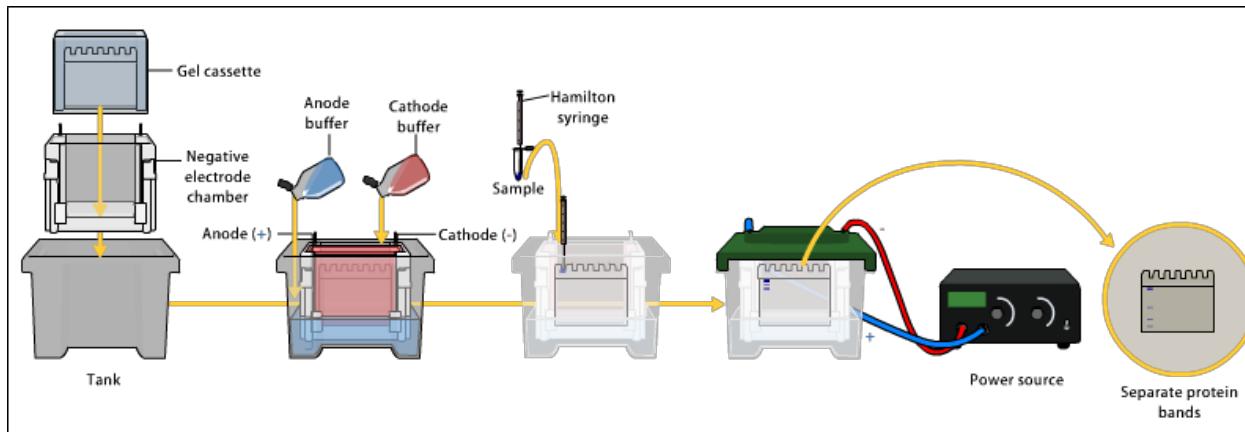
Coomassie brilliant blue is the name of two similar triphenylmethane dyes that were developed for use in the textile industry but are now commonly used for staining proteins in analytical biochemistry.



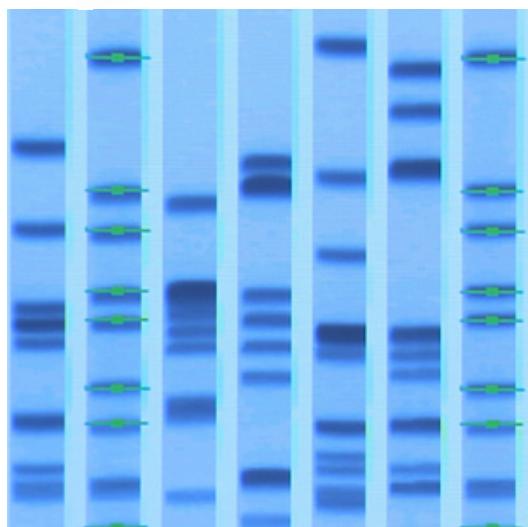
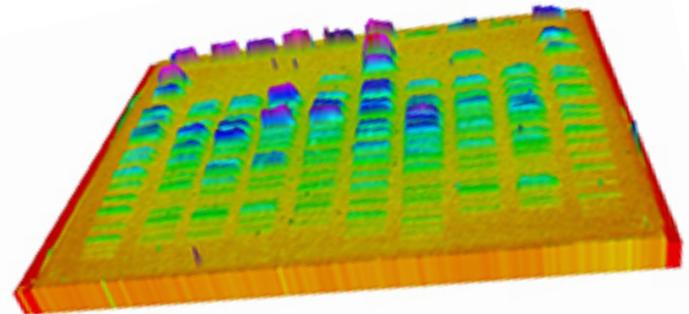
Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting assay of degummed silk proteins. (A) The fibroin proteins of the cocoons of C515 and three transgenic silkworms prepared in this study were analyzed. The cocoons were urea-degummed, and the resultant fibroin fibers were dissolved in 9-M aqueous lithium bromide solution, dialyzed against deionized water, and applied to SDS-PAGE. Fibroin H-chain protein, fibroin L-chain proteins, fhx/P25, and the fusion proteins, HC-SpA and HC-EGFP, were observed. (B) Western blotting analysis of fibroin proteins was performed using anti-6xHis tag antiserum (left) and anti-SpA protein specific peptide antibody (right). Fibroin proteins, such as C515, were not detected by these antibodies. HC-SpA protein, which contains 6xHis-tag and SpA proteins, could be detected with anti-6xHis tag antiserum and anti-SpA peptide antibodies. HC-EGFP protein, which contains EGFP and 6xHis tag, could only be detected using anti 6xHis-tag antiserum. Small amounts of degraded protein, which were not identified by SDS-PAGE, were present.

The western blot (sometimes called the protein immunoblot), or western blotting

is a widely used analytical technique in molecular biology and immunogenetics to detect specific proteins in a sample of tissue homogenate or extract.

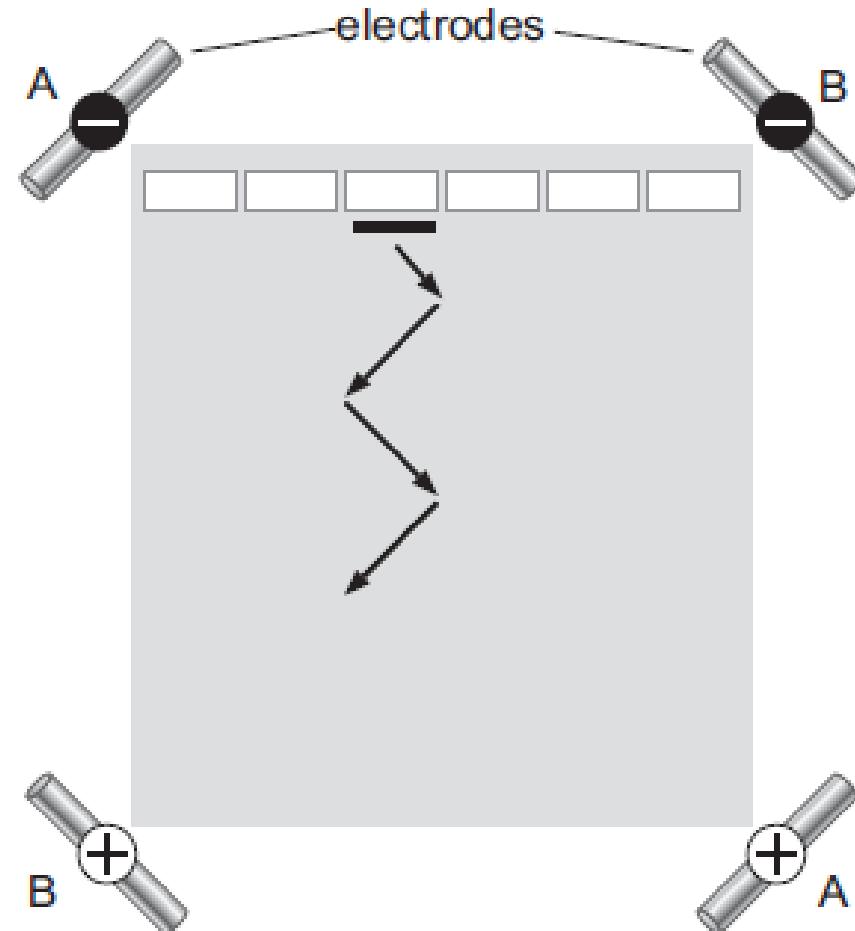


Nucleic acid separation – pulsed-field gel electrophoresis (PFGE)



The gel is stained so that DNA can be seen under ultraviolet (UV) light. A digital camera takes a photograph of the gel and stores the picture in the computer.

FIGURE 7-2 Pulsed-field gel electrophoresis. In this figure, the agarose gel is shown from above with a series of sample wells at the top of the gel. A and B represent two sets of electrodes. These are switched on and off alternately, as described in the text. When A is on, the DNA is driven toward the bottom right corner of the gel, where the anode of that pair is situated. When A is switched off and B is switched on, the DNA moves toward the bottom left corner. The arrows thus show the path followed by the DNA as electrophoresis proceeds. (Adapted, with permission, from Sambrook J. and Russell D.W. 2001. *Molecular cloning: A laboratory manual*, 3rd ed., Fig. 5-7. © Cold Spring Harbor Laboratory Press.)



Nucleic acid separation – pulsed-field gel electrophoresis

- Electric field orientation is changed
- Allows to separate large DNA molecules, 30+ kb to a few Mb

Restriction endonucleases

to cleave DNA at the specified sites

- Cleave DNA at particular sites by recognizing specific sequences
 - Usually **palindromic**

5'-GATCTGACTGATGCGTATGCTAGT-3'
3'-CTAGACTGACTACGCATACGATCA-5'

blunt end

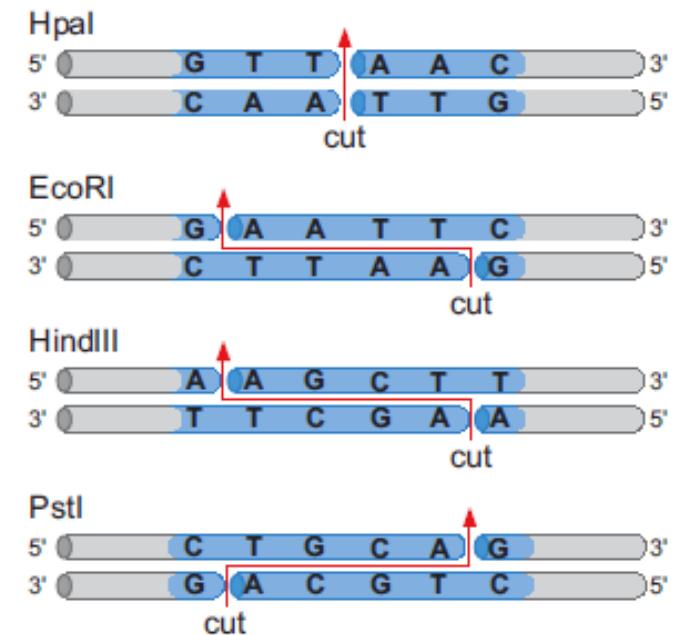


TABLE 7-1 Some Restriction Endonucleases and Their Recognition Sequences

Enzyme	Sequence	Cut Frequency ^a
Sau3A1	5'-GATC-3'	0.25 kb
EcoRI	5'-GAATTTC-3'	4 kb
NotI	5'-GCAGCCGC-3'	65 kb

^aFrequency = 1/4ⁿ, where n is the number of base pairs in the recognition sequence.

FIGURE 7-4 Recognition sequences and cut sites of various endonucleases. As shown, different endonucleases not only recognize different target sites but also cut at different positions within those sites. Thus, molecules with blunt ends or with 5' or 3' overhanging ends can be generated.

Restriction endonucleases

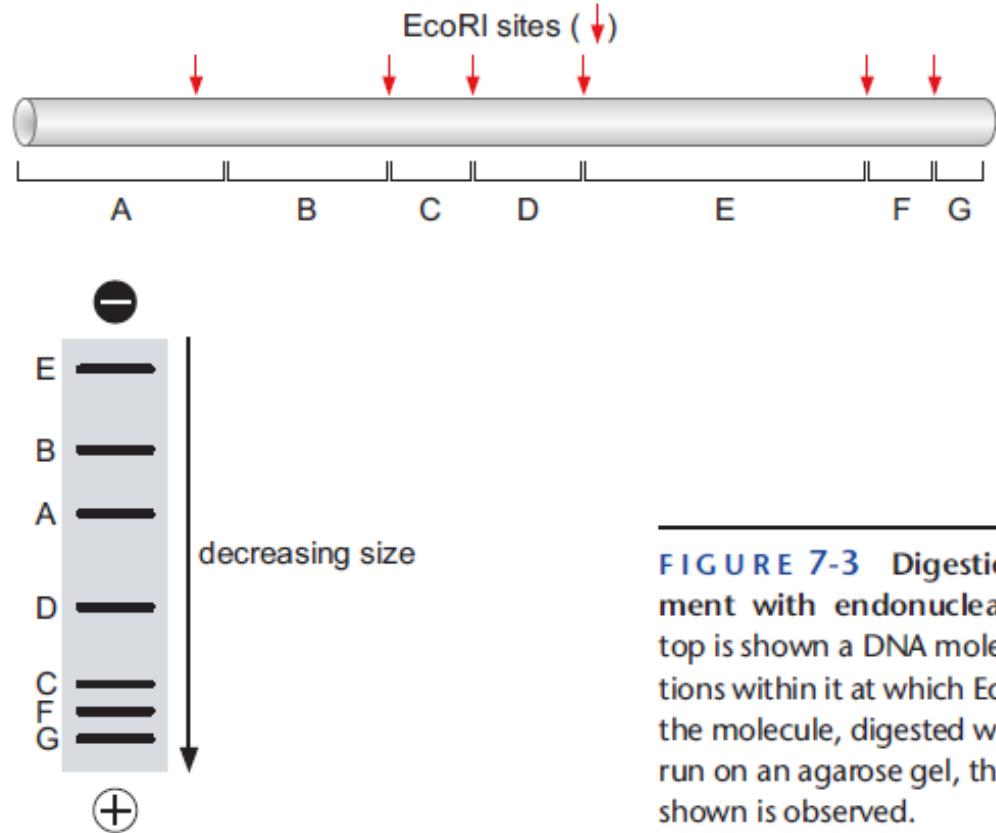


FIGURE 7-3 Digestion of a DNA fragment with endonuclease EcoRI. At the top is shown a DNA molecule and the positions within it at which EcoRI cleaves. When the molecule, digested with that enzyme, is run on an agarose gel, the pattern of bands shown is observed.

Ronald W. Davis (b. 1941) first discovered sticky ends as the product of the action of EcoRI, the restriction endonuclease.

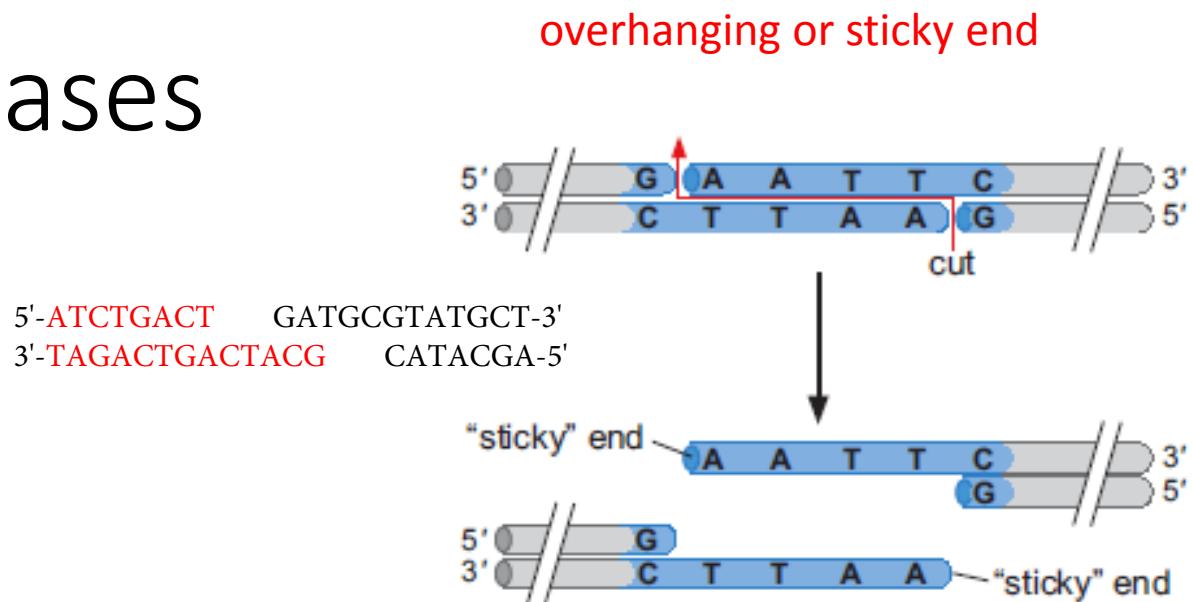


FIGURE 7-5 Cleavage of an EcoRI site. EcoRI cuts the two strands within its recognition site to give 5' overhanging ends. These are called "sticky" ends—they readily adhere to other molecules cut with the same enzyme because they provide complementary single-strand ends that come together through base pairing.

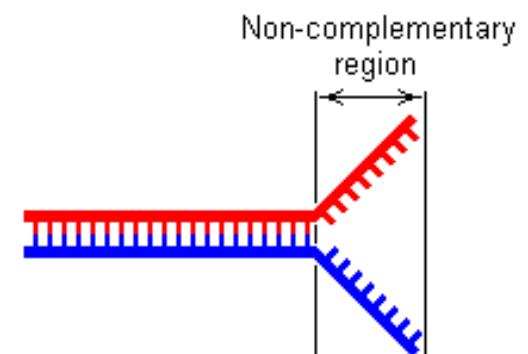
Normally, we see typically adenine pair with thymine, and cytosine with guanine to form a parallel complementary stand.

5'-ATCTGACT-3'
3'-TAGACTGA-5'

A frayed end refers to a region of a double stranded (or other multi-stranded) DNA molecule near the end with a significant proportion of non-complementary sequences; that is, a sequence where nucleotides on the adjacent strands do not match up correctly:

5'-ATCTGACTAGGCA-3'
3'-TAGACTGACTACG-5'

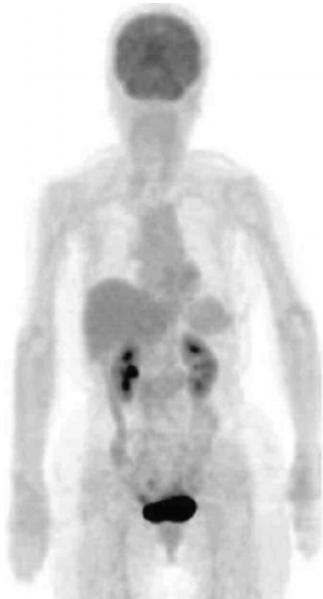
The term "frayed" is used because the incorrectly matched nucleotides tend to avoid bonding, thus appearing similar to the strands in a fraying piece of rope.



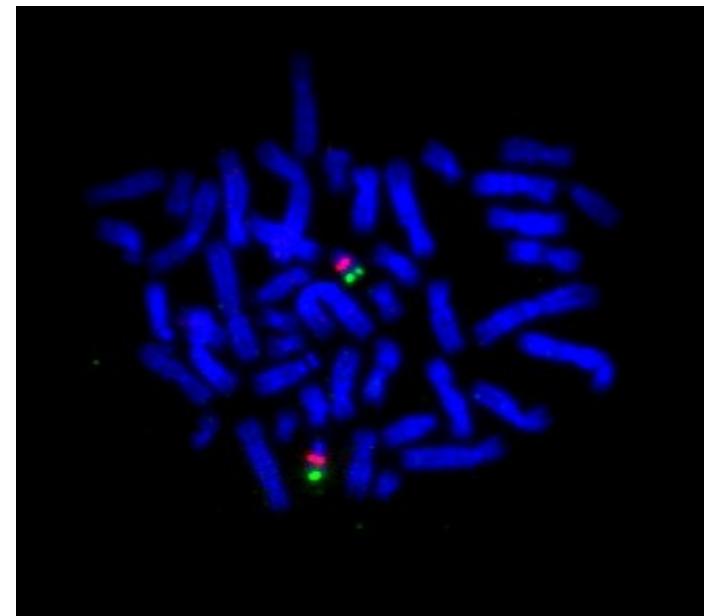
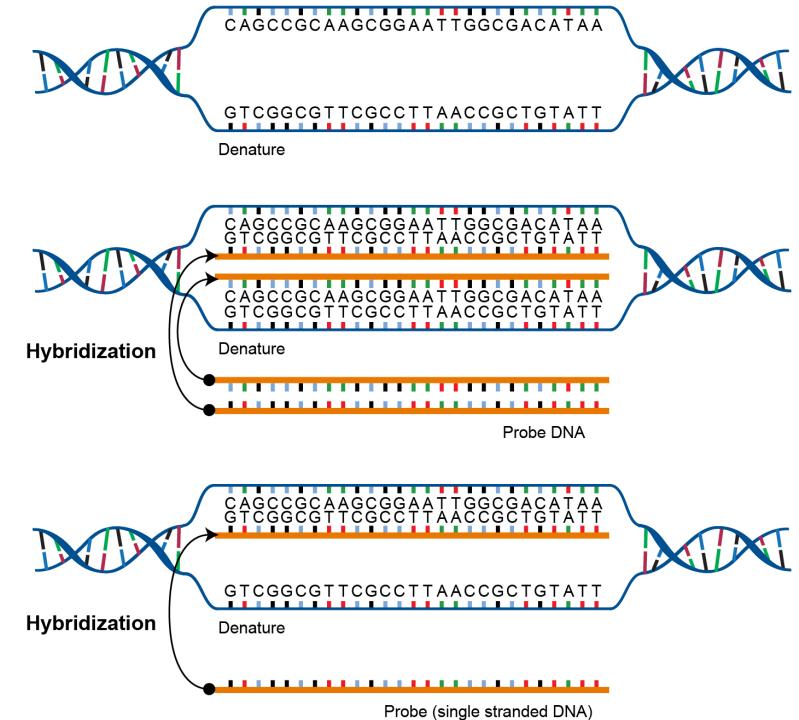
Hybridization

to make the DNA visible

- Process of base pairing between complementary single-stranded polynucleotides is known as hybridization



18-fluoro-deoxsy-glycose-positron-emission-tomography = **FDG-PET**



Fluorescence in situ hybridization method = **FISH-method**

- Fluorescent label

Hybridization

Northwestern blot - to detect interactions between RNA and proteins

- **Southern blot – DNA**

- **Southwestern blot –** to detect specific oligonucleotide probes

for detecting DNA-binding proteins

- **Northern blot – RNA**

- Reverse Northern blot – DNA or RNA is at first immobilized on blotting matrix

- **Western blot – proteins**

- Far-Western blot – use protein-protein interactions to detect specific protein

- Eastern blot – to detect specific posttranslational modifications (PTMs) of proteins

- Far-Eastern blot – detection of lipid-linked oligosaccharides

- **Dot-blot –** the probe will be dotted on the blotting matrix

The capacity of denatured DNA to reanneal (i.e., to re-form base pairs between complementary strands) allows for the formation of hybrid molecules when homologous, denatured DNAs from two different sources are mixed with each other under the appropriate conditions of ionic strength and temperature. This process of base pairing between complementary single-stranded polynucleotides is known as hybridization

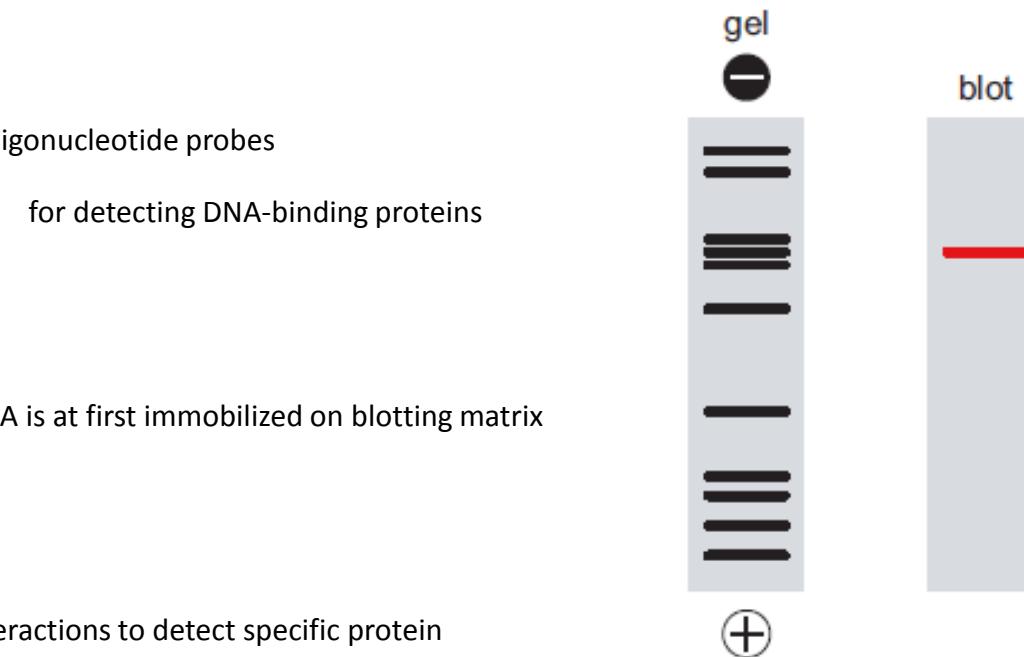
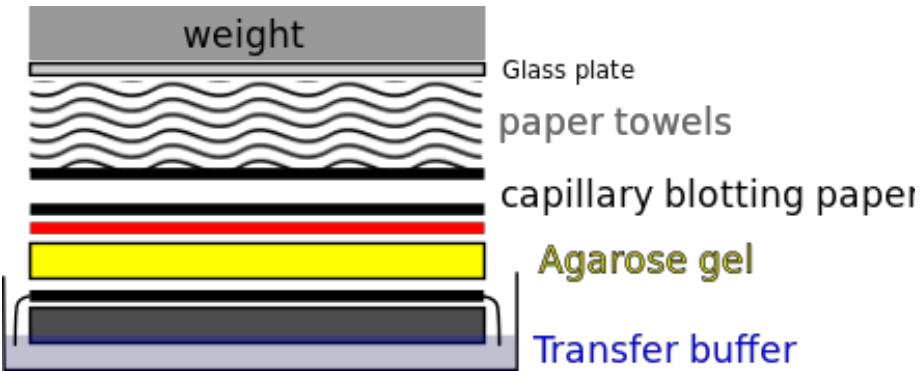
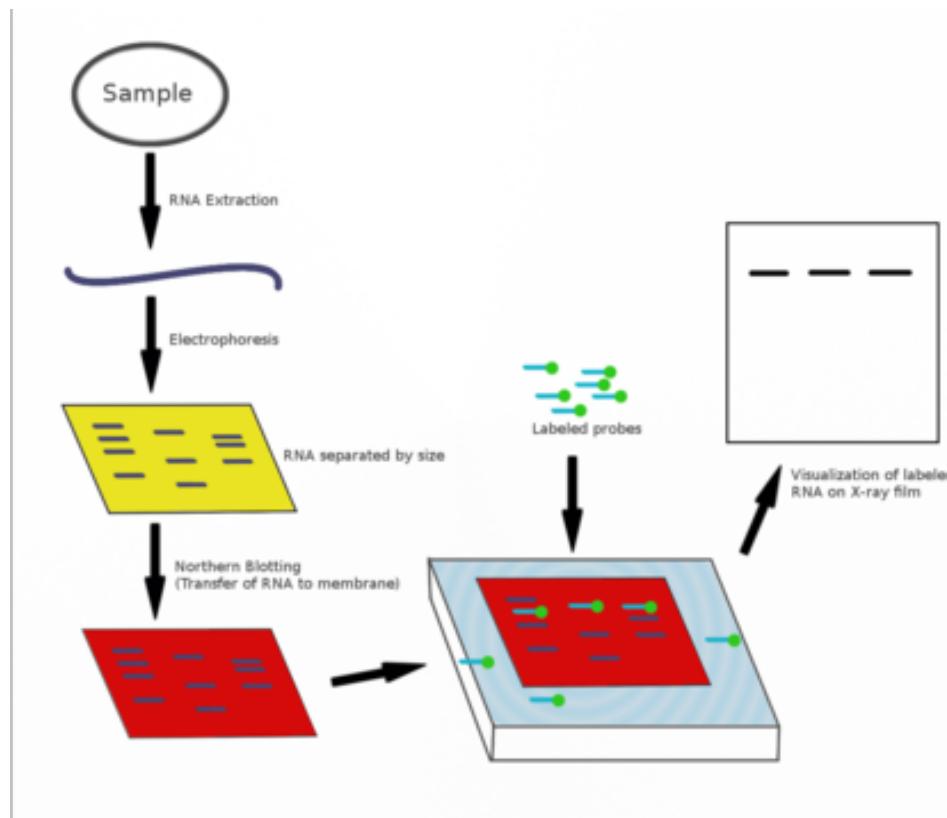


FIGURE 7-6 A Southern blot. DNA fragments, generated by digestion of a DNA molecule by a restriction enzyme, are run out on an agarose gel. Once stained, a pattern of fragments is seen. When transferred to a filter and probed with a DNA fragment homologous to just one sequence in the digested molecule, a single band is seen, corresponding to the position on the gel of the fragment containing that sequence.

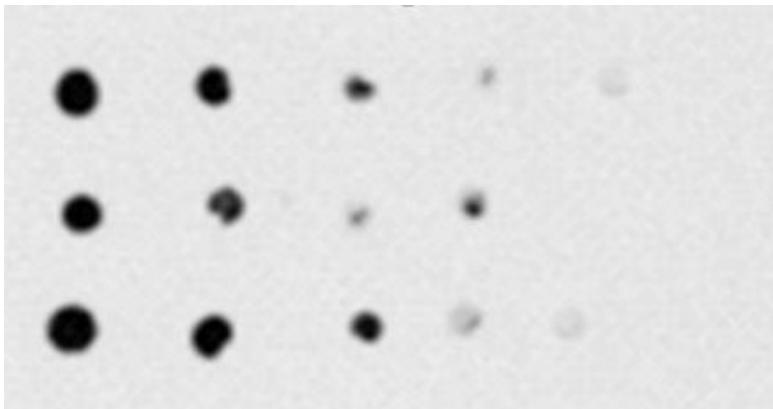
Nothern blot



Capillary blotting system setup for the transfer of RNA from an electrophoresis gel to a blotting membrane.



Dot blot (or slot blot)



A technique in molecular biology used to detect proteins. The **proteins to be detected** are not first separated by electrophoresis. Instead, the **sample is applied directly on a membrane** in a single spot, and the blotting procedure is performed.

Hybridization – microarray

In microarray analysis, the **fixed, unlabeled sequences are called the “probes,”** because **these are known DNA sequences**, are dotted or bind to plastic, whereas the **“target” is composed of amplified, labeled cDNAs generated from the total RNA from a cell or tissue.**

When target sequences are hybridized to the array of probe DNAs, the intensity of the hybridization signal to each DNA species in the array is a measure of the level of expression of the gene in question.

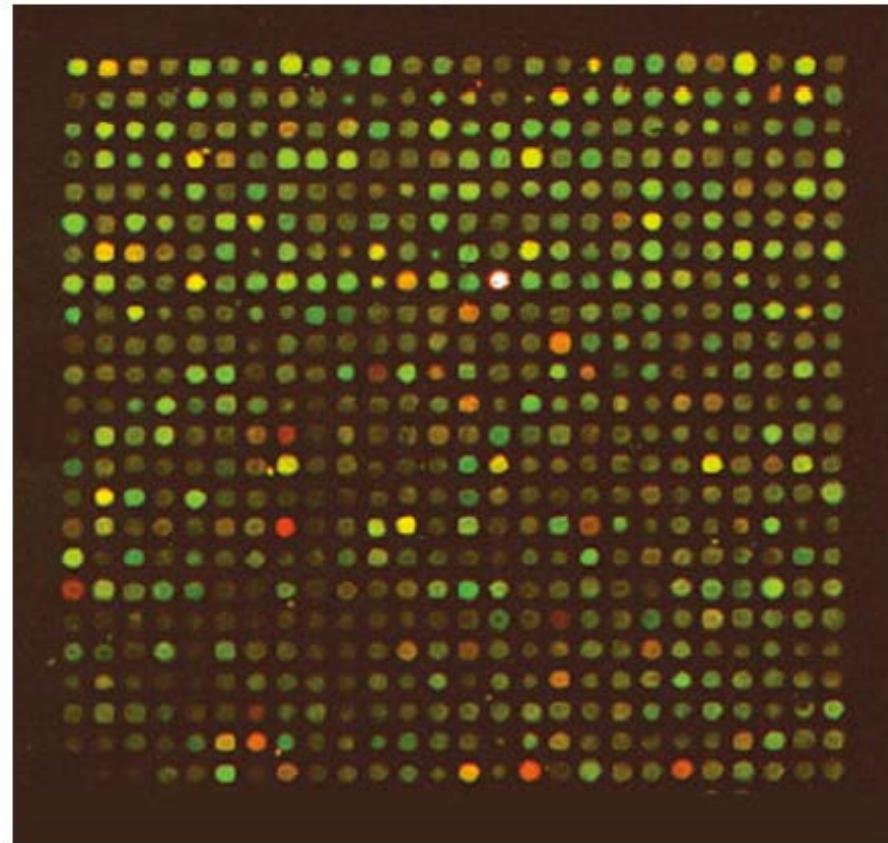


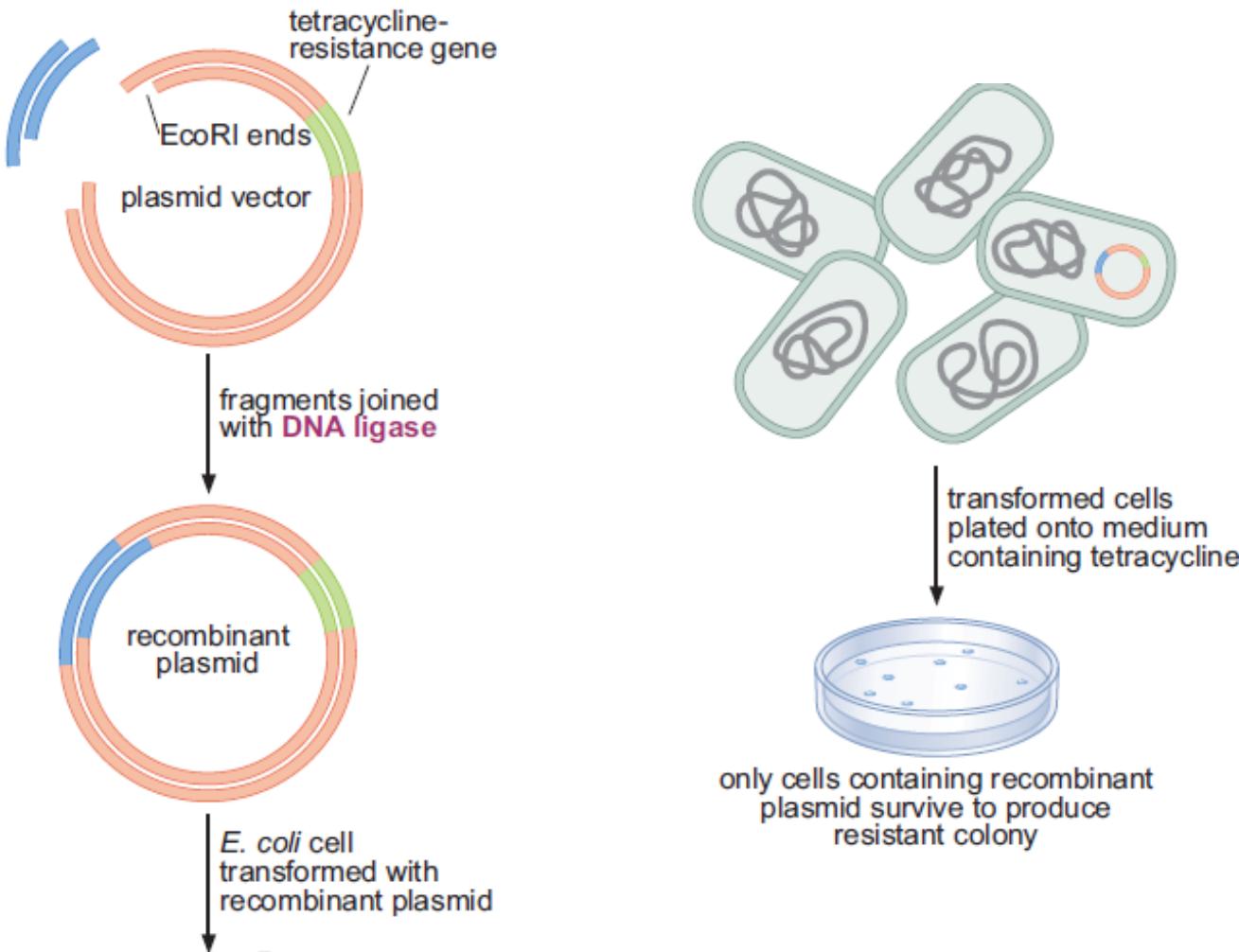
FIGURE 7-7 Microarray grid comparing expression patterns in two tissues (muscles and neurons) in *Caenorhabditis elegans*. Each circle in the grid contains a short DNA segment from the coding region of a single gene in the *C. elegans* genome. RNA was extracted from muscles and neurons, and labeled with fluorescent dyes (red and green, respectively). Thus, the red circles indicate genes expressed in muscle, whereas the green circles reflect genes expressed in neurons. The yellow circles indicate genes expressed in both cell types. It is clear that the two samples express distinct sets of genes. (Courtesy of Stuart Kim, Stanford University.)

The principles of Southern and Northern blot hybridization also underlie **microarray analysis**, which we consider in the Genomics section of this chapter. The availability of complete sequence information has enabled development of this “reverse hybridization” experiment. A microarray is constructed by attaching several hundred to thousands of known DNA sequences to a solid surface, typically a glass or plastic slide.

DNA cloning

- Construction of recombinant DNA molecules
- Vector to propagate the cloned DNA (plasmid)
 - Origin of replication
 - Selection marker
 - Unique restriction sites
- Transformation
 - Host organism can take up DNA from its environment

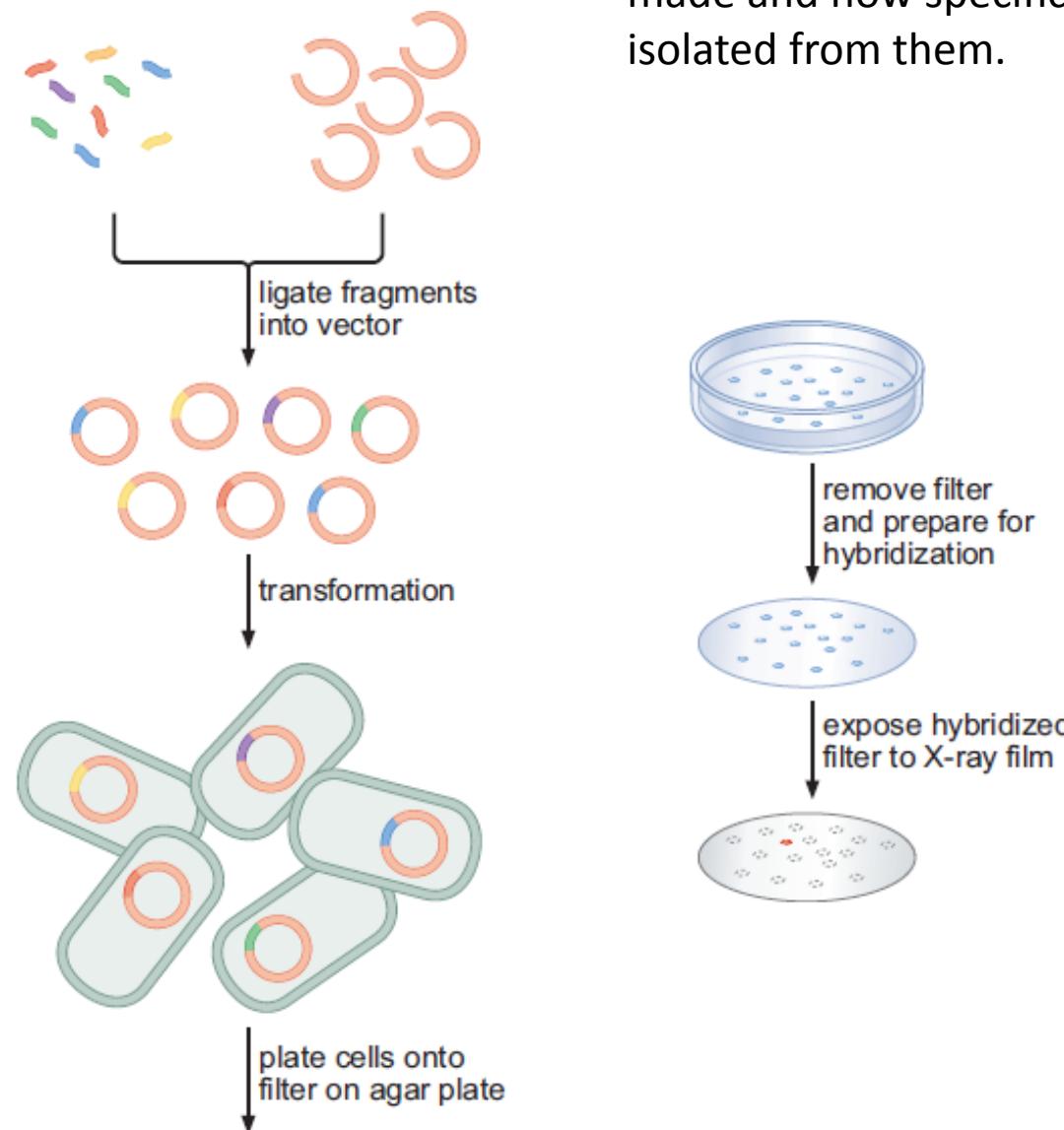
DNA cloning



The ability to construct recombinant DNA molecules and maintain them in cells is called DNA cloning.

FIGURE 7-8 Cloning in a plasmid vector. A fragment of DNA, generated by cleavage with EcoRI, is inserted into the plasmid vector linearized by that same enzyme. Once ligated (see text), the recombinant plasmid is introduced into bacteria by transformation (see text). Cells containing the plasmid can be selected by growth on the agar plates that contain growth media including antibiotic to which the plasmid confers resistance. (Adapted, with permission, from Micklos D.A. and Freyer G.A. 2003. *DNA science: A first course*, 2nd ed., p. 129. © Cold Spring Harbor Laboratory Press.)

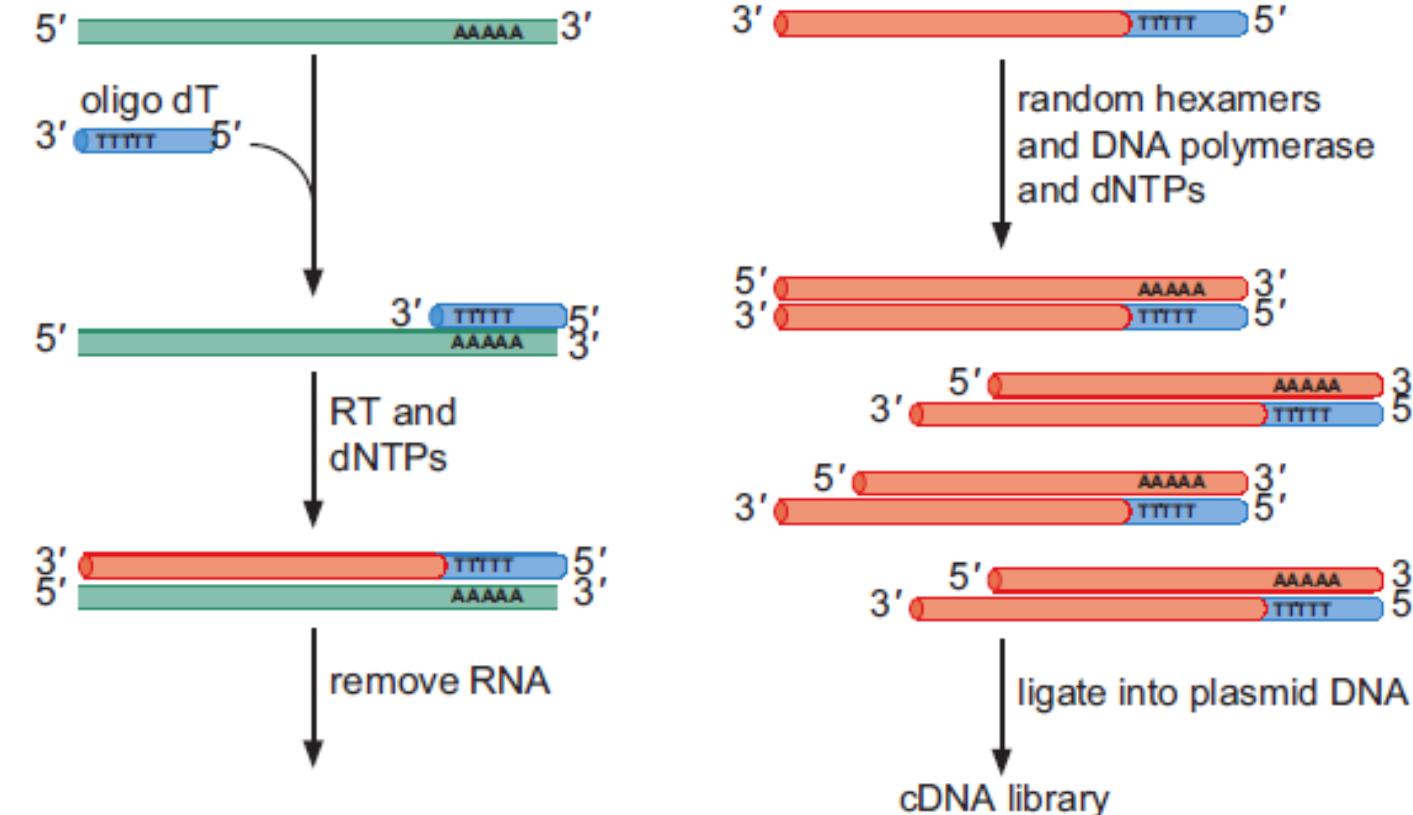
DNA libraries



We can discuss how large collections of such hybrid molecules, called **libraries**, can be created. In a library, a common vector carries many alternative inserts. We describe how libraries are made and how specific DNA segments can be identified and isolated from them.

FIGURE 7-9 Construction and probing of a DNA library. To construct the library, genomic DNA and vector DNA, digested with the same restriction enzyme, are incubated together with ligase. The resulting pool or library of hybrid vectors (each vector carrying a different insert of genomic DNA, represented in a different color) is then introduced into *E. coli*, and the cells are plated onto a filter placed over agar medium. Once colonies have grown, the filter is removed from the plate and prepared for hybridization: cells are lysed, the DNA is denatured, and the filter is incubated with a labeled probe. The clone of interest is identified by autoradiography.

cDNA libraries



To enrich for coding sequences in the library, a cDNA library is created. This is made as shown in Figure 7-10. Instead of starting with genomic DNA, mRNAs are converted into DNA sequences. The process that allows this is called reverse transcription and is performed by a special DNA polymerase (reverse transcriptase) that can make DNA from an RNA template.

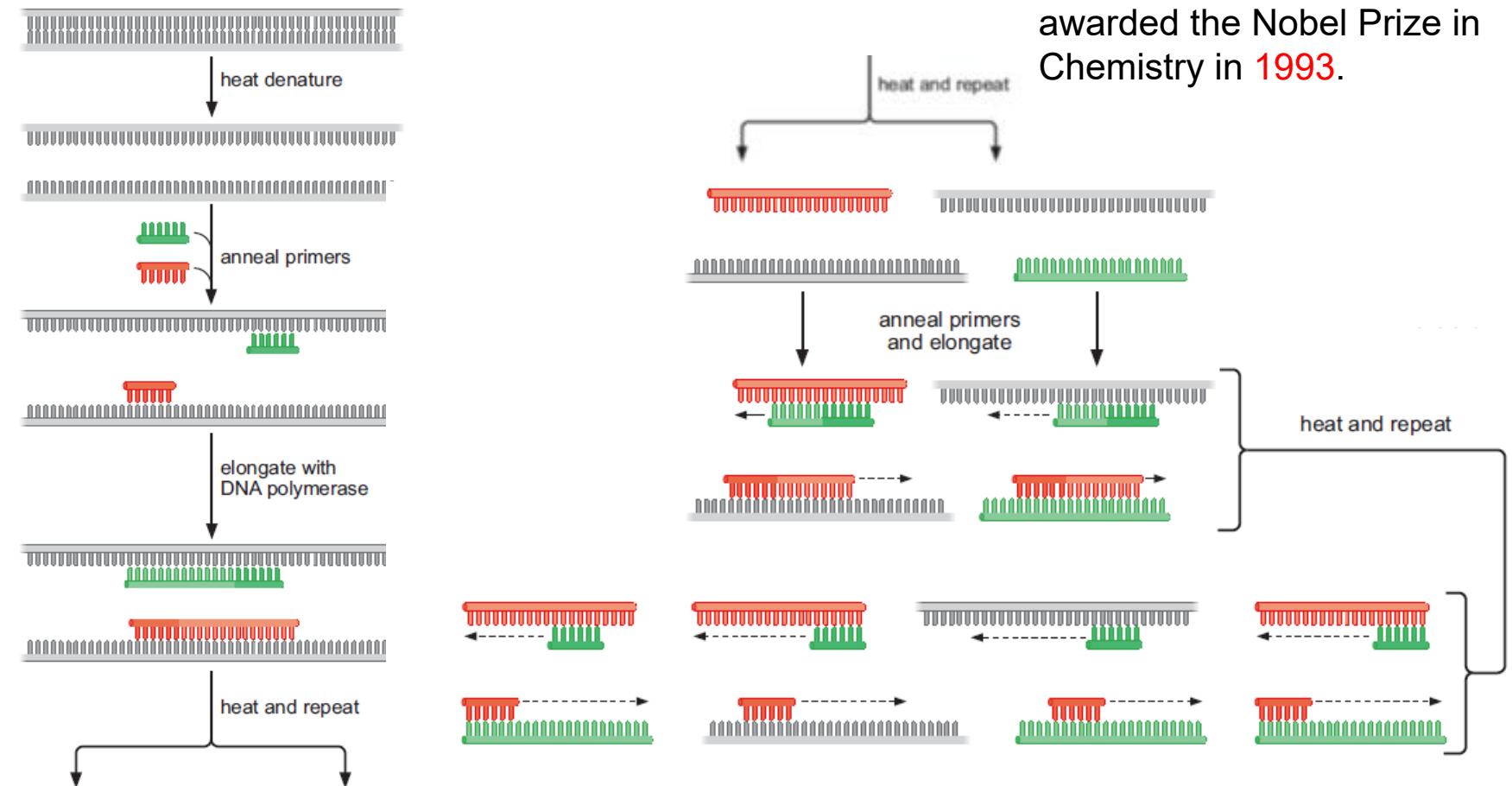
FIGURE 7-10 Construction of a cDNA library. The RNA-dependent DNA polymerase reverse transcriptase (RT) transcribes RNA into DNA (copy, or cDNA). In the first step (first-strand synthesis), oligos of poly-T sequence serve as primers by hybridizing to the poly-A tails of the mRNAs. (cDNA libraries are typically made from eukaryotic cells whose mRNA have poly-A tails at their 3' ends; see Chapter 19.) Reverse transcriptase extends the dT primer to complete a DNA copy of the mRNA template. The product is a duplex composed of one strand of mRNA and its complementary strand of DNA. The RNA strand is removed by treatment with base (NaOH), and the remaining single-stranded DNA now serves as template for the second step (second-strand synthesis). Short random sequences of DNA usually ~6 bp long (called random hexamers) serve as primers by hybridizing to various sequences along the copy DNA template. These primers are then extended by DNA polymerase to create double-stranded DNA products that can be cloned into a plasmid vector (see Fig. 7-8) to create a cDNA library.

PCR – polymerase chain reaction

The Polymerase Chain Reaction Amplifies DNAs by repeated rounds of DNA replication *in vitro* – Its artificial multiplying of DNA repeats

FIGURE 7-12 Polymerase chain reaction (PCR). In the first step of the PCR, the DNA template is denatured by heating and annealed with synthetic oligonucleotide primers (dark orange and dark green) corresponding to the boundaries of the DNA sequence to be amplified. DNA polymerase is then used to copy the single-stranded template by extension from the primers (light orange and light green). In the next step, DNA is once again denatured

PCR procedure is performed entirely biochemically, that is, *in vitro*. PCR uses the enzyme DNA polymerase that directs the synthesis of DNA from deoxynucleotide substrates on a single-stranded DNA template.



PCR was invented in 1983 by the American biochemist **Kary Mullis** (1944-2019) at Cetus Corporation; Mullis and biochemist **Michael Smith** (1932-2000, site-directed mutagenesis), who had developed other essential ways of manipulating DNA, were jointly awarded the Nobel Prize in Chemistry in 1993.

DNA sequencing

- Conventional DNA sequencing
- Separate nested sets of DNA molecules by size
 - Same 5' end, different length due to different 3' ends



Frederick Sanger 1918 - 2013), a pioneer of sequencing. Sanger is one of the few scientists who was awarded two Nobel prizes, one for the sequencing of proteins, and the other for the sequencing of DNA.

The game-changing method for amplifying particular segments of DNA, distinct from cloning and propagation within a host cell, is the polymerase chain reaction (PCR).

In 1958, he was awarded a Nobel Prize in Chemistry "for his work on the structure of proteins, especially that of insulin". In 1980, Walter Gilbert and Sanger shared half of the chemistry prize "for their contributions concerning the determination of base sequences in nucleic acids".

DNA sequencing – the chain-termination method

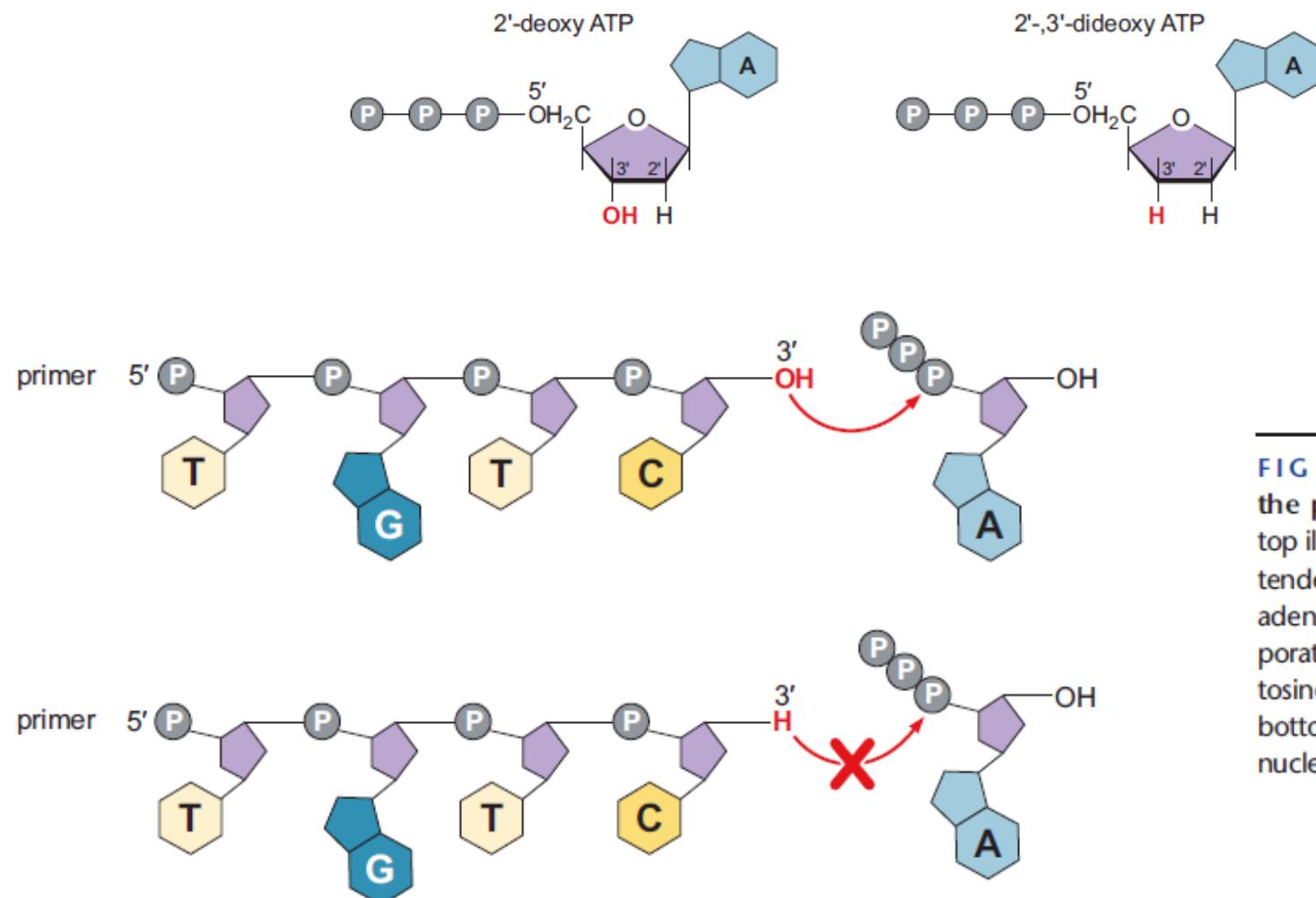


FIGURE 7-13 Dideoxynucleotides used in DNA sequencing. On the left is 2'-deoxy ATP. This can be incorporated into a growing DNA chain and allow another nucleotide to be incorporated directly after it. On the right is 2',3'-dideoxy ATP. This can be incorporated into a growing DNA chain, but once in place it blocks further nucleotides being added to the same chain.

FIGURE 7-14 Chain termination in the presence of dideoxynucleotides. The top illustration shows a DNA chain being extended at the 3' end with addition of an adenine nucleotide onto the previously incorporated cytosine. The presence of dideoxycytosine in the growing chain (shown at the bottom) blocks further addition of incoming nucleotides as described in the text.

The chain-termination method uses special, modified substrates called 2',3'-dideoxynucleotides (ddNTPs), which lack the 3-hydroxyl group on their sugar moiety as well as the 2'-hydroxyl (Fig. 7-13). DNA polymerase will incorporate a 2',3'-dideoxynucleotide at the 3' end of a growing polynucleotide chain, but once incorporated, the lack of a 3'-hydroxyl group prevents the addition of further nucleotides, causing elongation to terminate (Fig. 7-14).

Now suppose that we “spike” (delete or add) a cocktail of the nucleotide substrates with the modified substrate **2',3'-dideoxyguanosine triphosphate (ddGTP)** at a ratio of one ddGTP molecule to 100 2'-deoxy-GTP molecules (dGTP). This will cause DNA synthesis to abort at a frequency of one in 100 times the DNA polymerase encounters a C on the template strand (Fig. 7-15a).

DNA sequencing

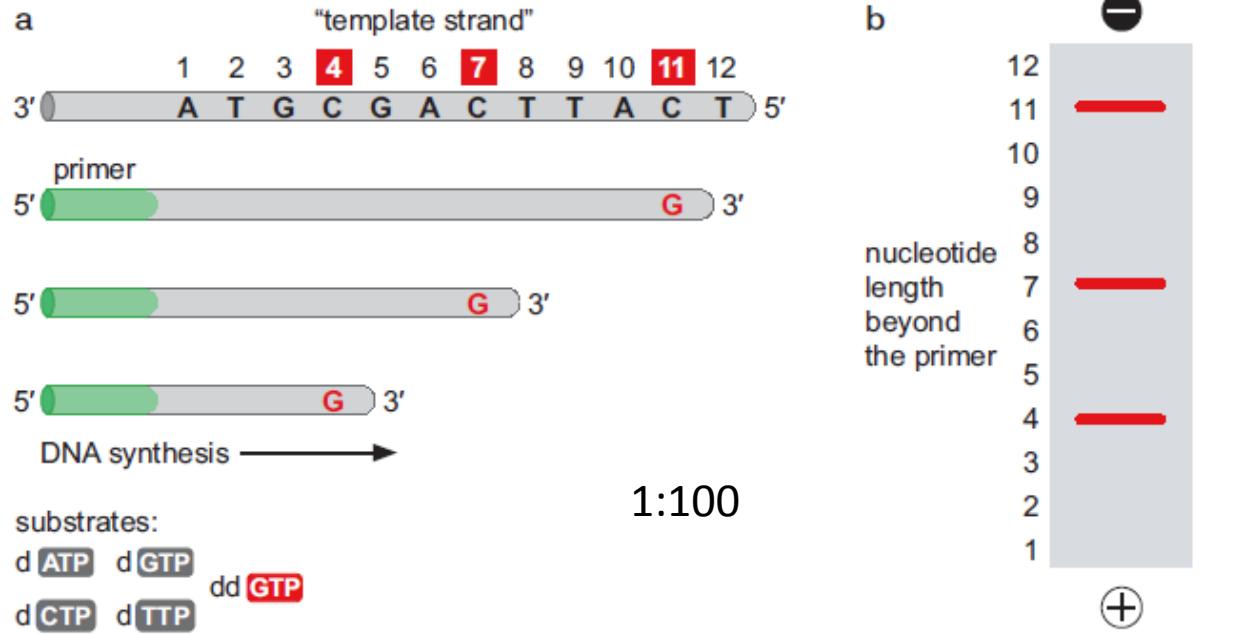


FIGURE 7-15 DNA sequencing by the chain-termination method. As described in the text, chains of different length are synthesized in the presence of dideoxynucleotides. The length of the chains produced depend on the sequence of the DNA template and which dideoxynucleotide is included in the reaction. (a, top) The sequence of the template. In this reaction, all bases are present as deoxynucleotides, but G is present in the dideoxy form as well. Thus, when the elongating chain reaches a C in the template, it will, in some fraction of the molecules, add the ddGTP instead of dGTP. In those cases, chains terminate at that point. (b) Fragments separated on a polyacrylamide gel. The lengths of fragments seen on the gel reveal the positions of cytosines in the template DNA being sequenced in the reaction described.

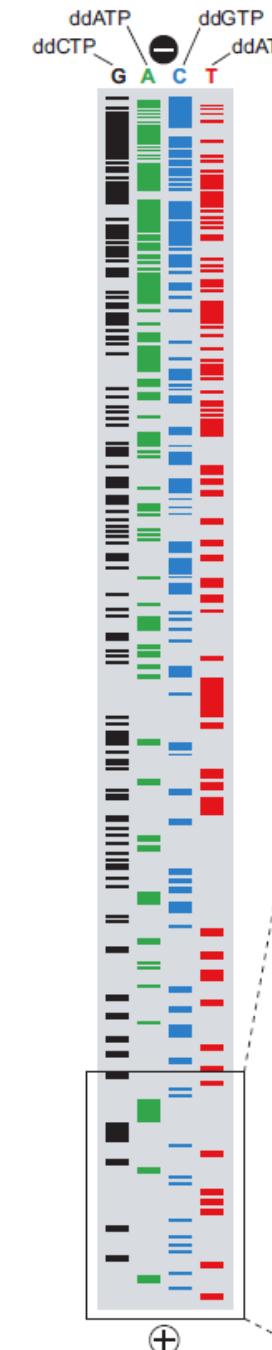
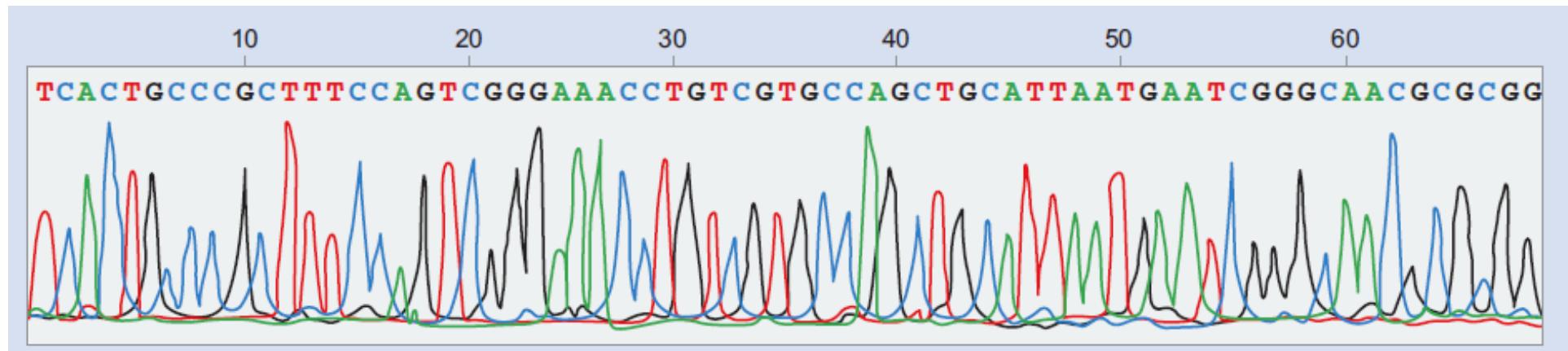


FIGURE 7-16 DNA-sequencing gel. The lengths of DNA chains, terminated with the dideoxynucleotide indicated at the top of each lane, are determined by resolving on a polyacrylamide gel, as shown. Reading the gel from bottom to top gives the 5'-to-3' sequence.

Shotgun DNA sequencing

- Generate plasmid libraries of random fragments from the genome
- Sequence all of them , with ~10x overlay



BOX 7-2 FIGURE 1 DNA sequence readout. In this reaction, as described in the text, fluorescently end-labeled dideoxynucleotides are used, and the chains are separated by column chromatography. The profile of positions of As is represented in green, Ts in red, Gs in black, and Cs in blue.

Shotgun DNA sequencing

The average human chromosome is composed of 150 Mb. Thus, the 600 bp of DNA sequence provided by a typical sequencing reaction represents only 0.0004% of a typical chromosome. Consequently, to determine the complete sequence of the chromosome, it is necessary to generate a large number of sequencing reads from many short DNA fragments (Fig. 7-17). To achieve this goal, DNA is prepared from each of the 23 chromosomes that constitute the human genome and then sheared into small fragments by passage through small-gauge pressurized needles. The collection of small fragments, each derived from individual chromosomes, is then reduced into pools. Typically, two or three pools are constructed for fragments of differing (increasing) sizes—for example, fragments of 1, 5, or 100 kb in length. These fragments are then randomly cloned into bacterial plasmids as we described above to make libraries.

Hartwell et al. Genetics: From Genes to Genomes, 2e,
©The McGraw-Hill Companies Inc.

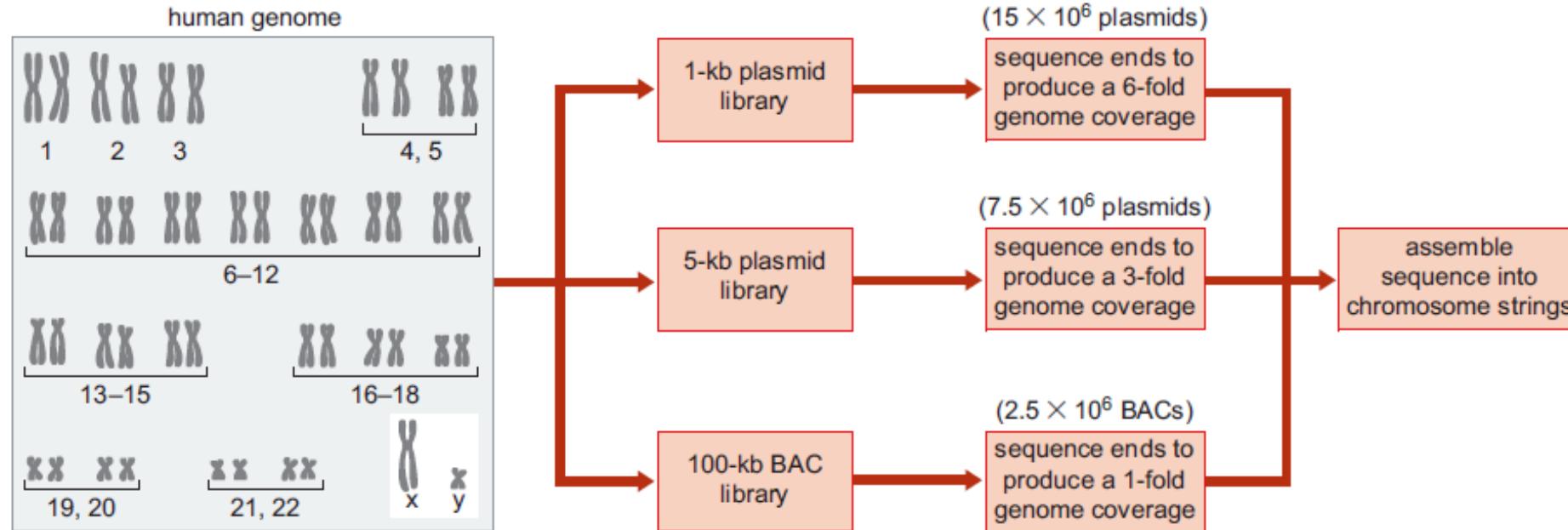


FIGURE 7-17 Strategy for construction and sequencing of whole-genome libraries. Contiguous sequences are determined for the shotgun sequencing of the short genomic DNA fragments. Contigs are extended by the use of end sequences derived from the larger fragments carried in the 5-kb and 100-kb insert clones as described in the text. (Adapted, with permission, from Hartwell L. et al. 2003. *Genetics: From genes to genomes*, 2nd ed., Fig. 10-13. © McGraw-Hill.)

Shotgun DNA sequencing

Sophisticated computer programs have been developed that assemble the short sequences from random shotgun DNAs into larger contiguous sequences called **contigs**. Sequences or “reads” that contain identical sequences are assumed to overlap and are joined to form larger contigs (Fig. 7-18). The sizes of these contigs depend on the amount of sequence obtained—the more sequence, the larger the contigs and the fewer gaps in the sequence.

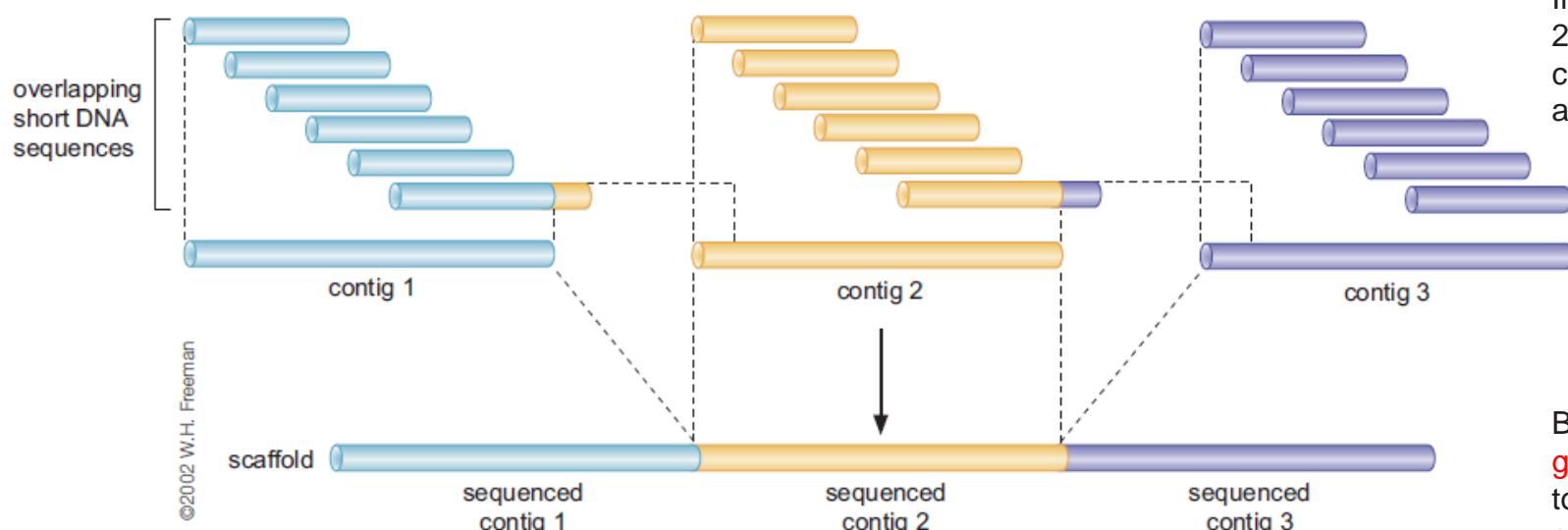


FIGURE 7-18 Contigs are linked by sequencing the ends of large DNA fragments. For example, one end of a random 100-kb genomic DNA fragment might contain sequence matches within contig 1, whereas the other end matches sequences in contig 2. This places the two contigs on a common scaffold. (Adapted, with permission, from Griffiths A.J.F. et al. 2002. *Modern genetics*, 2nd ed., Fig. 9-29b. © W.H. Freeman.)

Individual contigs are typically composed of 50,000–200,000 bp. This is still far short of a typical human chromosome. However, such contigs are useful for analyzing compact genomes.

Because the **human genome contains an average of one gene every 100 kb**, a typical contig is often insufficient to capture an entire gene, let alone a series of linked genes. We now consider **how relatively short contigs are assembled into larger scaffolds** that are typically **1–2 Mb** in length.

Human haploid genome
ca 3.2×10^9 bp

Shotgun sequencing – paired-end

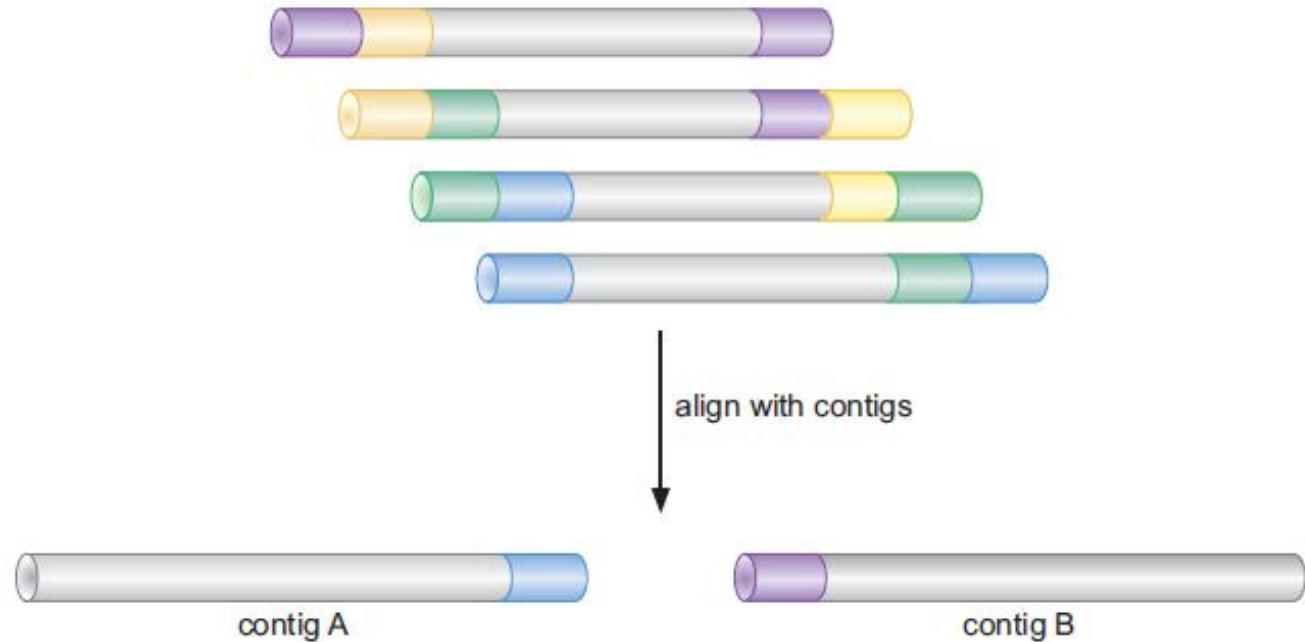
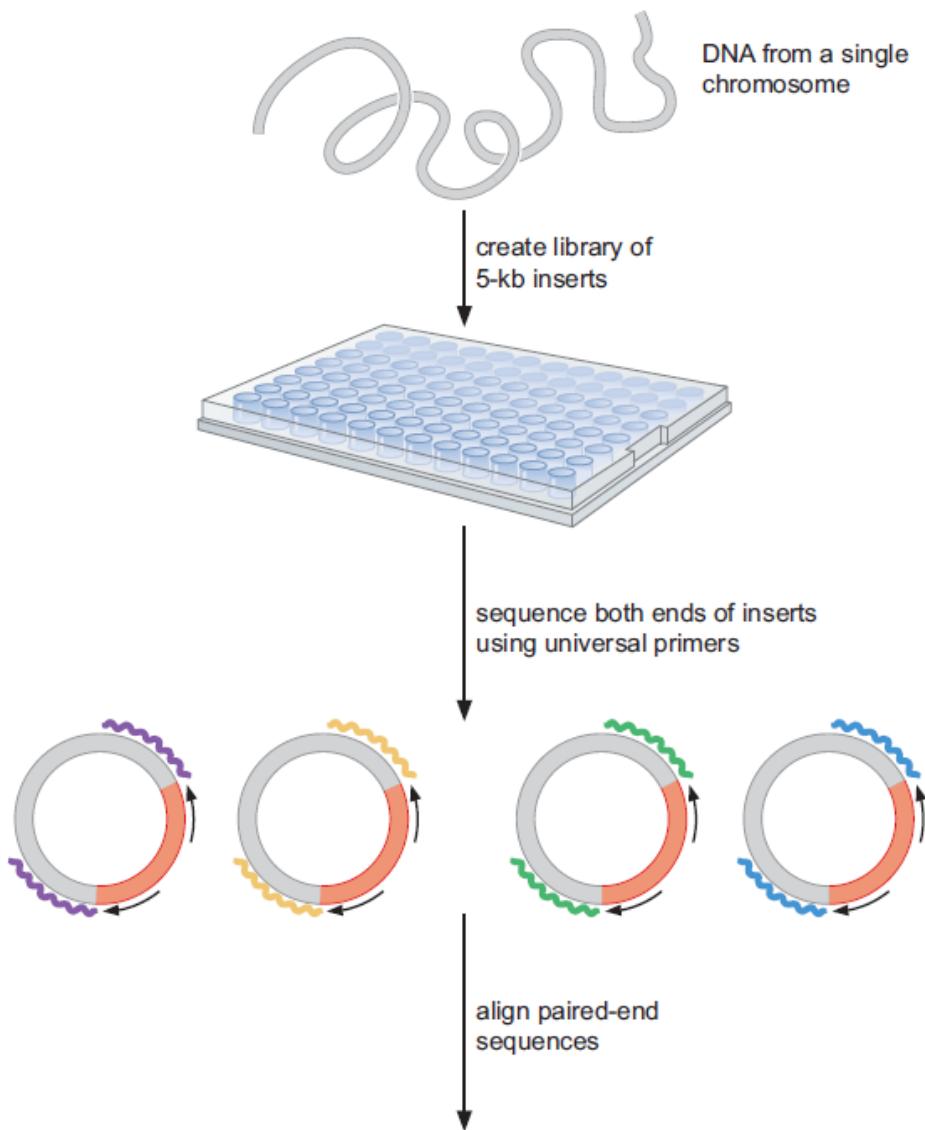


FIGURE 7-19 A “shotgun” library containing random genomic DNA inserts of 5 kb in length. Each well on the plate contains a different insert. Sequences 600 bp in length are determined for both ends of each genomic DNA (color coded). These paired-end sequences are used to align different contigs. In this example, the 5-kb genomic DNA fragment with the blue sequences contains matching sequences with contig A and contig B.



Human genome sequencing

- Initial project cost >300 million USD
- 2nd gen instruments
 - 400 Mb in 4 hours
 - Detects released pyrophosphate
 - 200-250 nt per well
 - 1x coverage of a genome – 30 runs
(2 – 3 weeks on one machine)
 - 10000-30000 USD, today less than 1000 USD

The first human genome sequences were published in nearly complete draft form in February 2000 by the Human Genome Project and Celera Corporation.

Completion of the Human Genome Project's sequencing effort was announced in 2006 with the publication of a draft genome sequence.

At 2018 the diploid genomes of over a million individual humans had been determined using next-generation sequencing.

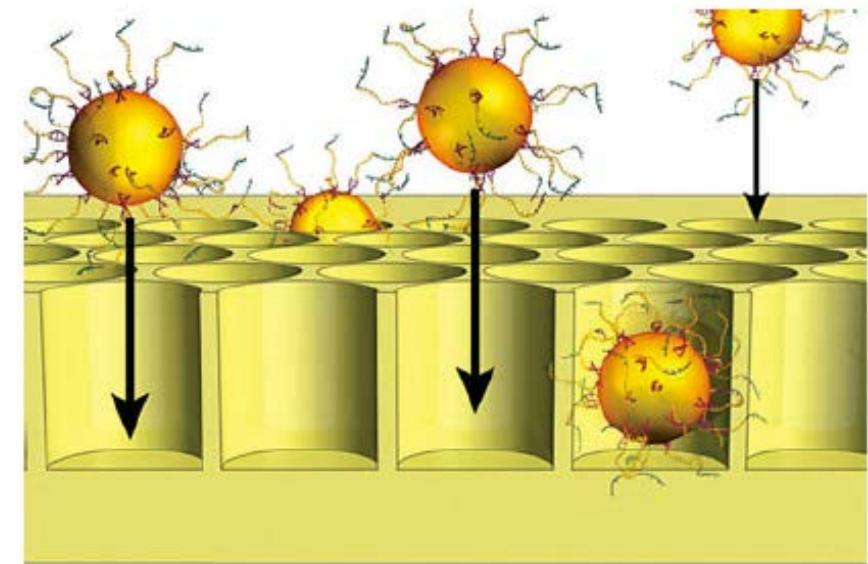
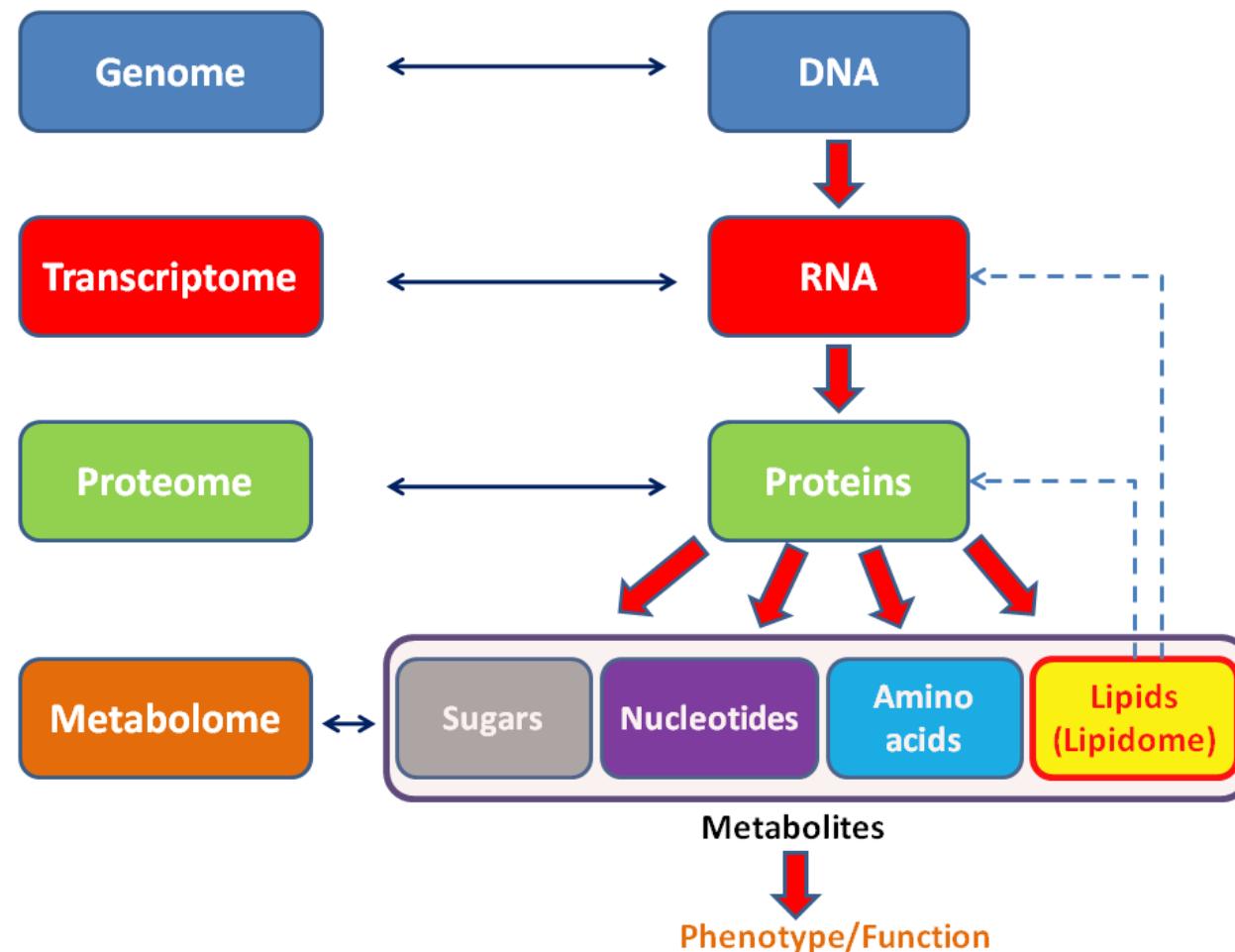


FIGURE 7-20 Cartoon of individual pores in the 454 sequencing apparatus. Each pore contains a small bead with an amplified DNA sequence. Sequential rounds of sequencing are detected by the release of pyrophosphate and light. Further description of the method is given in the text. (Reprinted, with permission, from Margulies M. et al. 2005. *Nature* 437: 376–380, Fig. 1a. © Macmillan.)

- The transcriptome is the set of all RNA molecules in one cell or a population of cells.

It is sometimes used to refer to all RNAs, or just mRNA, depending on the particular experiment. It differs from the exome in that it includes only those RNA molecules found in a specified cell population, and usually includes the amount or concentration of each RNA molecule in addition to the molecular identities.



Tiling array (mosaic array) – a way to study transcriptome

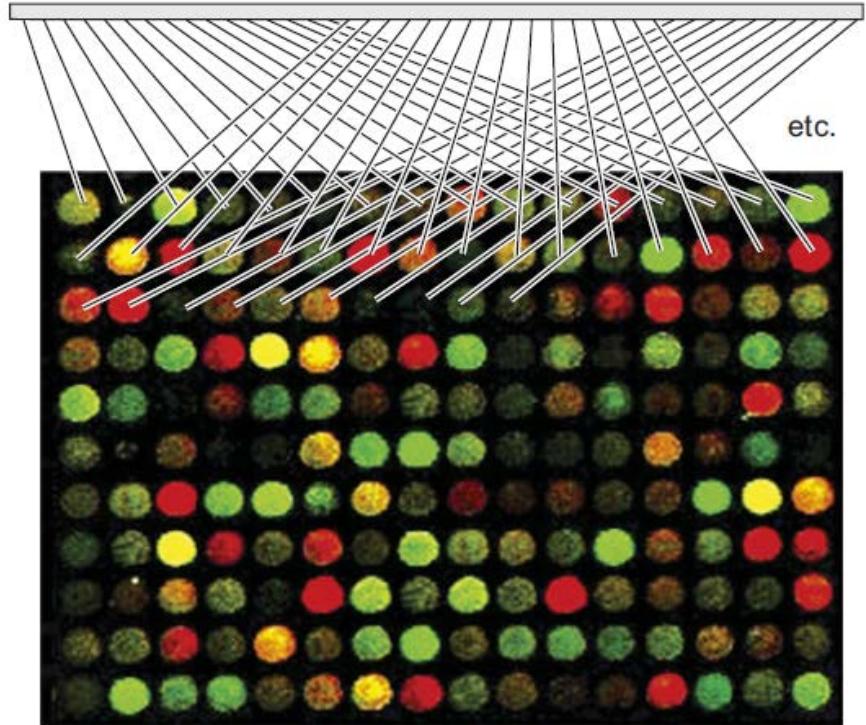
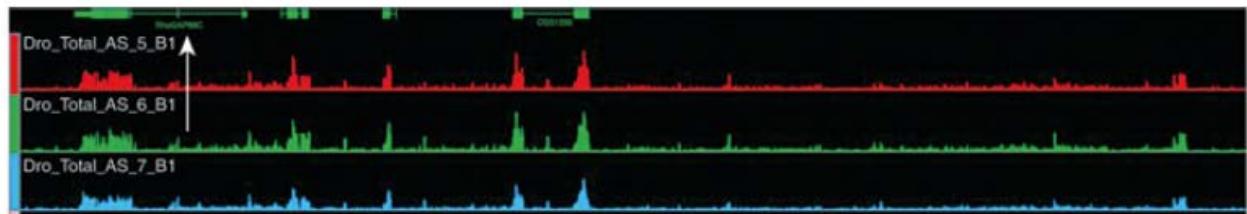


FIGURE 7-22 Whole-genome tiling microarray. The image represents a portion of a tiling array that has been hybridized with fluorescently labeled probes. The grid includes a high number of uniformly spaced DNA probes across a region of interest (e.g., an entire genome).



The **portion of an organism's genome that acts as a template for RNA synthesis is known as the transcriptome**. To identify this portion of the genome, synthetic, single-stranded DNAs of 50 nucleotides in length are spotted on a glass or silicon slide. Typically, one oligonucleotide is produced for every 100–150 bp of DNA sequence in a sequential manner across the genome, resulting in a “tiling array” of DNA sequences.

To visualize the transcriptome, the tiling arrays are hybridized with fluorescently labeled RNA (or cDNA) probes (see Fig. 7-22). These probes might be derived from a specific cell type, such as the tail muscles of the sea squirt tadpole or yeast cells grown in a particular medium. The end result is a series of hybridization signals superimposed on all of the predicted protein-coding sequences across the genome (Fig. 7-23).

FIGURE 7-23 Whole-genome tiling array reveals details of the intron–exon structure of a gene. A 50-kb interval on *Drosophila* chromosome 3 that contains four different genes. The intron–exon structure of each transcription unit is shown at the top of the figure. (White arrow) The large intronic region that might contain a small (“micro–”) exon. Total RNA was extracted from progressively older embryos (red, young; green, older; and blue, still older) and hybridized to the tiling array, which contains 25-nucleotide sequences every 35 bp throughout the entire genome. Strong hybridization signals coincide with the exons, whereas there are weaker signals in the intronic regions. Based on the similar signals in all three colors this gene is expressed at similar levels at all three ages of embryos tested. (Reprinted, with permission, from Manak et al. 2006. *Nat. Genet.* 38: 1151–1158, Fig. 5. © Macmillan.)

Bioinformatics – powerful tool

Genome technologies are effective at identifying genes and determining the structures of their transcription units.

In particular, the Basic Local Alignment Search Tool, or BLAST, algorithm provides a powerful approach for searching, comparing, and aligning either protein or nucleic acid sequences.

A subset of vertebrate regulatory sequences can be identified using variations in the BLAST searches developed for characterizing protein-coding sequences. Cell-specific enhancers contain clustered binding sites for one or more sequence-specific DNA-binding proteins (see Chapter 19). In some cases, this clustering is sufficient for the identification of short stretches of DNA sequence alignment.

BLAST - Basic Local Alignment Search Tool

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.21.559229; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.21.559229; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

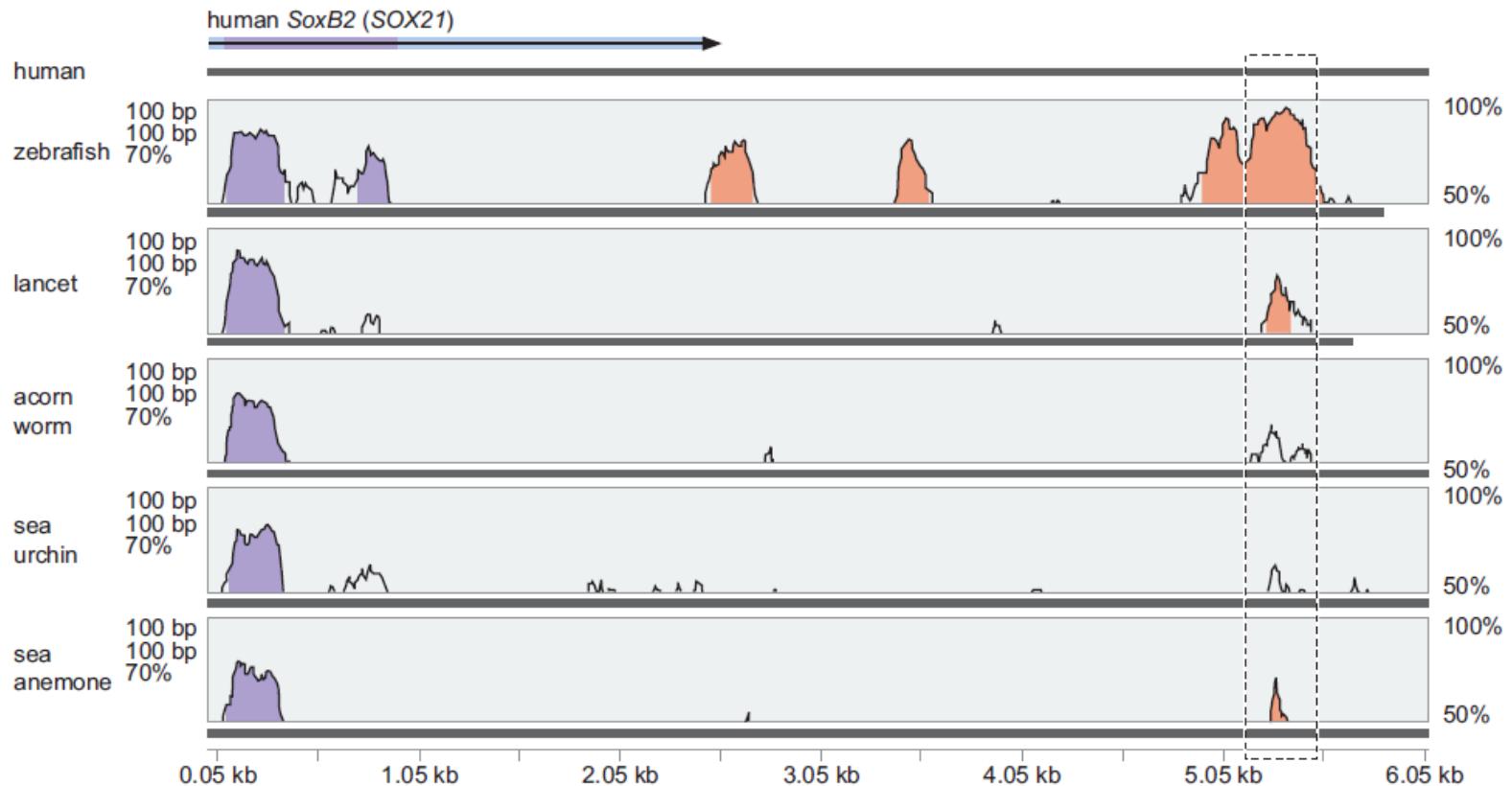


FIGURE 7-24 Comparison of the *SoxB2* gene in divergent animals. The lavender signals correspond to conserved sequences in the 3'-UTR of the *SoxB2* transcription unit. The pink signals indicate conserved sequences that map downstream of the gene. The dashed rectangle identifies enhancers that mediate expression in the nervous system. (Adapted, with permission, from Royo J.L. et al. 2011. *Proc. Natl. Acad. Sci.* 108: 14186–14191, Fig. 1A, p. 14187.)



Analyze first *in silico*, then *in vivo*

Tissue-specific enhancers can also be identified by scanning genomic DNA sequences for potential binding sites of known regulatory proteins. Consider the case of the α -catenin gene, which encodes a cell adhesion molecule.

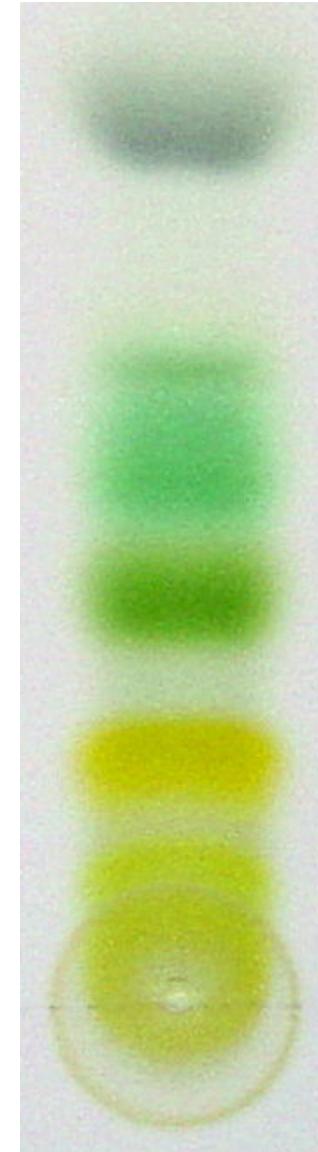
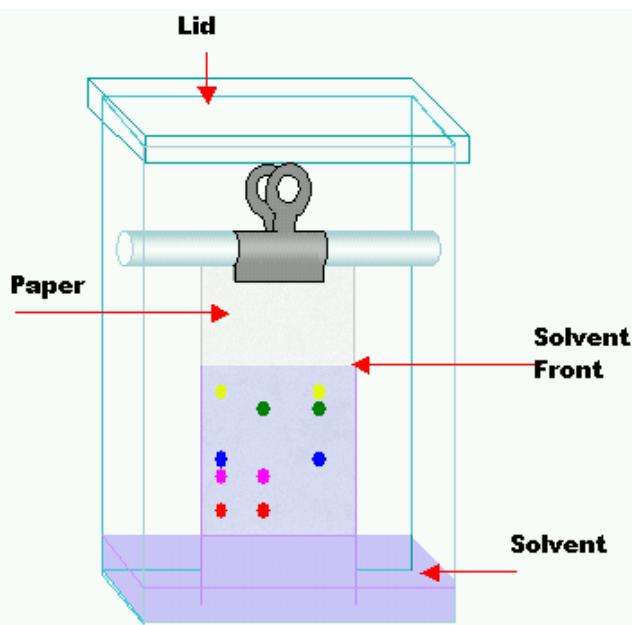
The gene is expressed in several different tissues, but it shows particularly strong expression in heart precursor cells called cardiomyocytes. It was possible to identify a heart-specific enhancer by surveying the flanking and intronic sequences of α -catenin for matches to the binding sites of known heart cell regulatory proteins, including MEF2C, GATA-4, and E47/HAND(Fig. 7-25). Each of these proteins recognizes a spectrum of short sequence motifs of 6–10 bp.



FIGURE 7-25 In silico identification of a heart enhancer. An ~140-bp sequence in the 5'-flanking region of the α -catenin gene is conserved in the mouse, rat, and human genomes. The conserved sequence contains binding sites for three critical regulators of heart differentiation: E47/HAND, MEF2C, and GATA. The mouse sequence has been shown to function as an authentic heart-specific enhancer. In principle, it could be identified by either VISTA alignments (see Chapter 20, Fig. 20-4) or the clustering of heart regulatory proteins. (Portion reprinted, with permission, from Vanpoucke G. et al. 2004. *Nucleic Acids Res.* 32: 4155–4165, Fig. 1. © Oxford University Press.)

Chromatography

In chemical analysis, chromatography is a laboratory technique **for the separation of a mixture into its components**. The mixture is dissolved in a fluid solvent (gas or liquid) called the mobile phase, which carries it through a system (a column, a capillary tube, a plate, or a sheet) on which a material called the stationary phase is fixed.

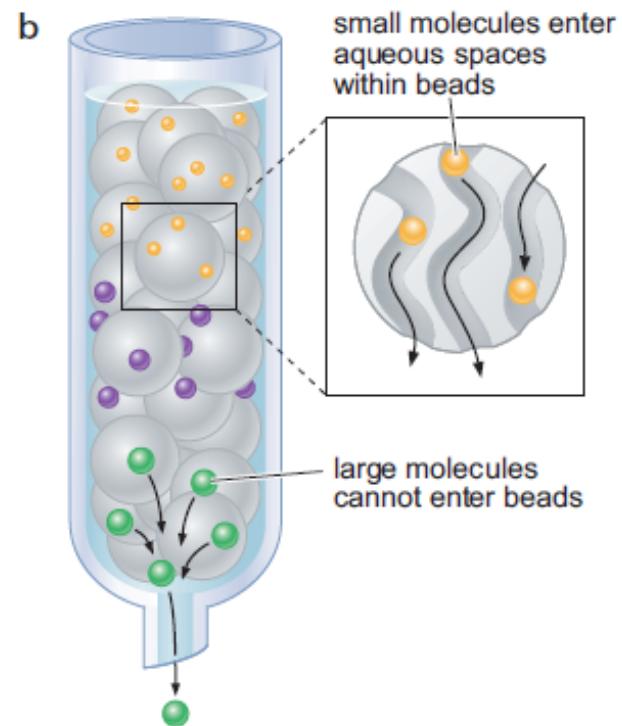
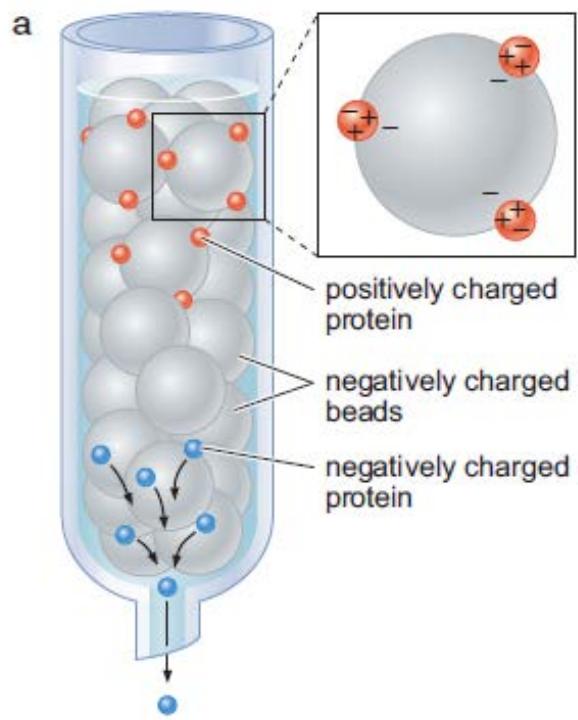


paper chr.
thin-layer chr.
gas chr.
liquid chr.
affinity chr. etc

Protein purification

- Column chromatography
 - Ion-exchange chromatography
 - Gel-filtration (size-exclusion) chromatography
 - Affinity chromatography

Protein purification/separation



Ion-Exchange Chromatography

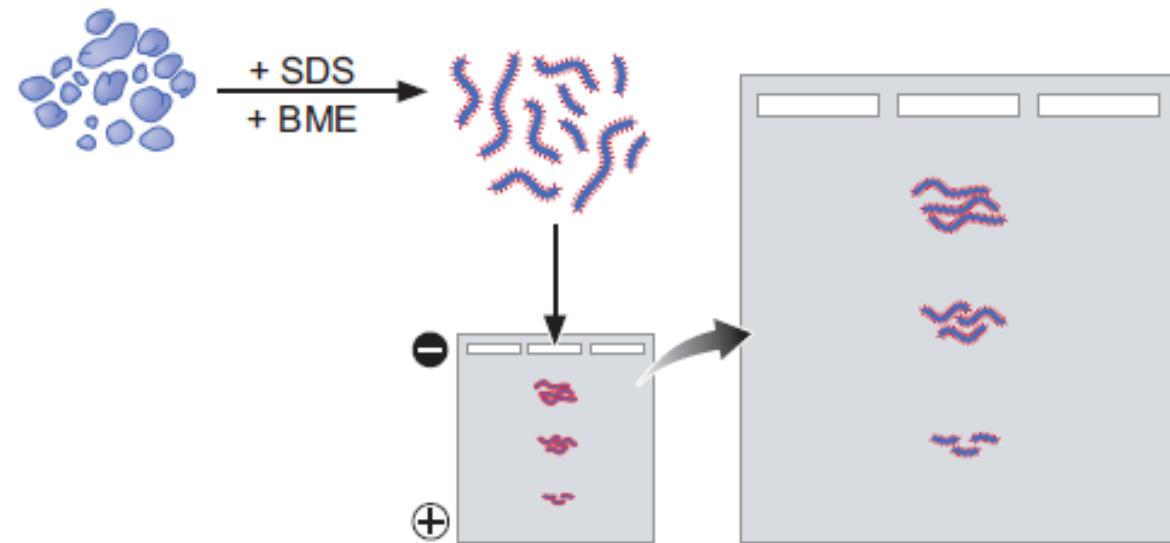
Gel-filtration Chromatography

FIGURE 7-26 Ion-exchange and gel-filtration chromatography. As described in the text, these two commonly used forms of chromatography separate proteins on the basis of their charge and size, respectively. Thus, in each case, a glass tube is packed with beads, and the protein mixture is passed through this matrix. The nature of the beads dictates the basis of protein separation. (a) Ion-exchange chromatography. In this example, the beads are negatively charged. Thus, positively charged proteins bind to them and are retained on the column, whereas negatively charged proteins pass through. Increasing the concentration of salt in the surrounding buffer can elute bound proteins by competing for the negative charges on the column. (b) Gel-filtration chromatography. The beads contain aqueous spaces into which small proteins can pass, slowing down their progress through the column. Larger proteins cannot enter the beads, allowing them to pass more rapidly through the column.

Protein separation by size

Proteins have neither a uniform negative charge nor a uniform structure. Rather, they are constructed from 20 distinct amino acids, some of which are uncharged, some are positively charged, and still others are negatively charged. After electrophoresis, the proteins can be visualized with a stain, such as Coomassie Brilliant Blue, that binds to protein nonspecifically. When the SDS is omitted, electrophoresis can be used to separate proteins according to properties other than molecular weight, such as net charge and isoelectric point.

FIGURE 7-27 SDS gel electrophoresis. A mixture of three proteins of different size are illustrated (much more complex mixtures are usually analyzed). Addition of SDS (shown in red) and β -mercaptoethanol denatures the proteins and provides each with a uniform negative charge. Separation on the basis of size is achieved by electrophoresis.



sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) = SDS-PAGE

BME - β -Mercaptoethanol

Protein detection - immunoblotting

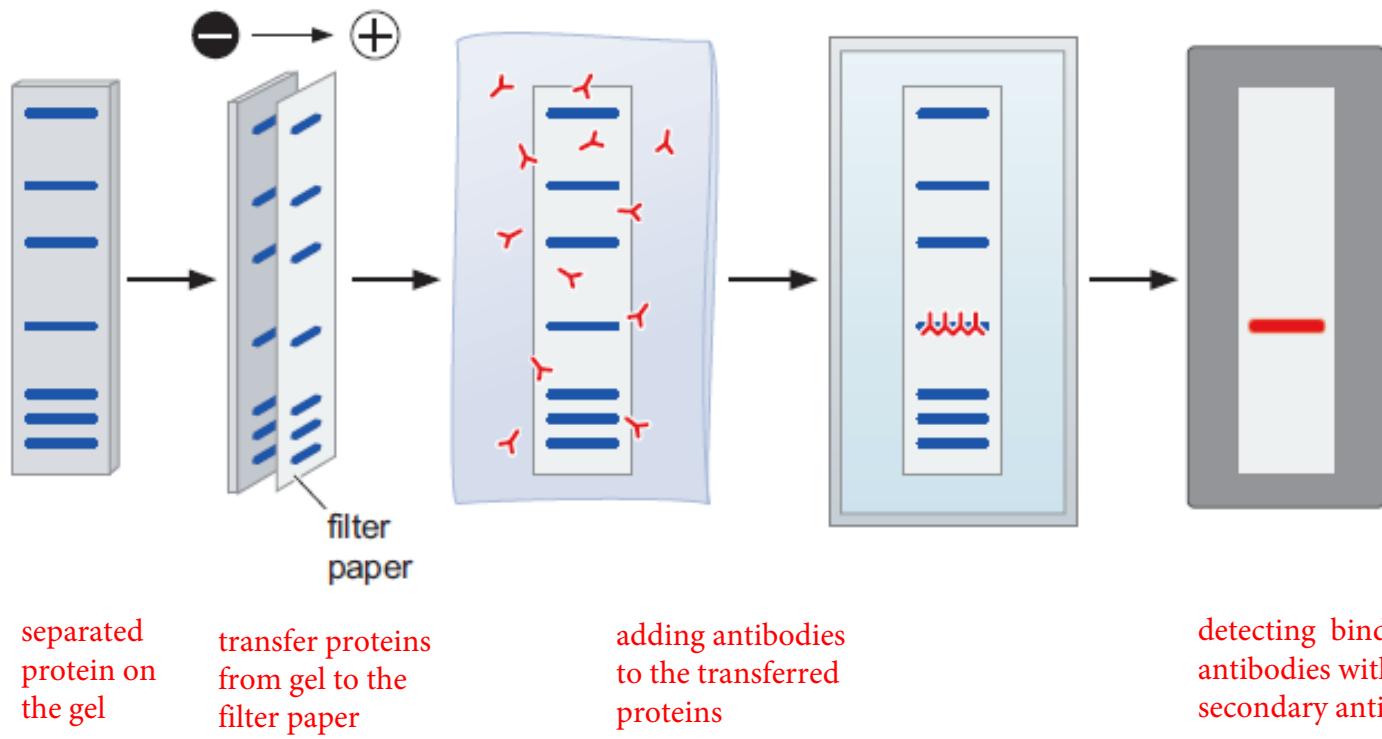


FIGURE 7-28 Immunoblotting. After proteins are separated by electrophoresis, they are transferred to filter paper (again using an electric field) in a manner that retains the same relative position of the proteins. After blocking nonspecific protein-binding sites, antibody to the protein of interest is added to the filter paper. The site of antibody binding is then detected using an attached enzyme that creates light when it acts on its substrate.

Protein sequencing

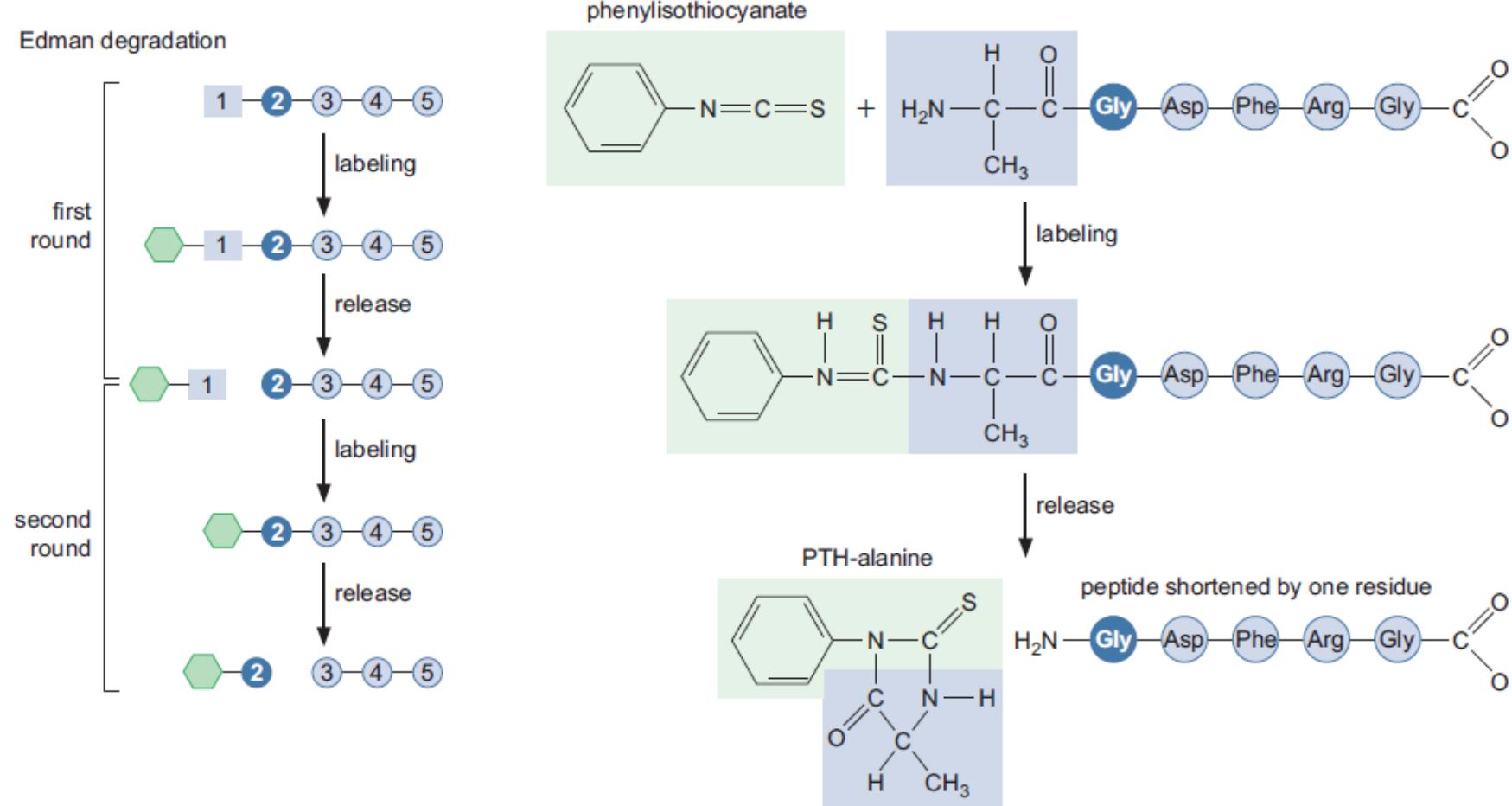
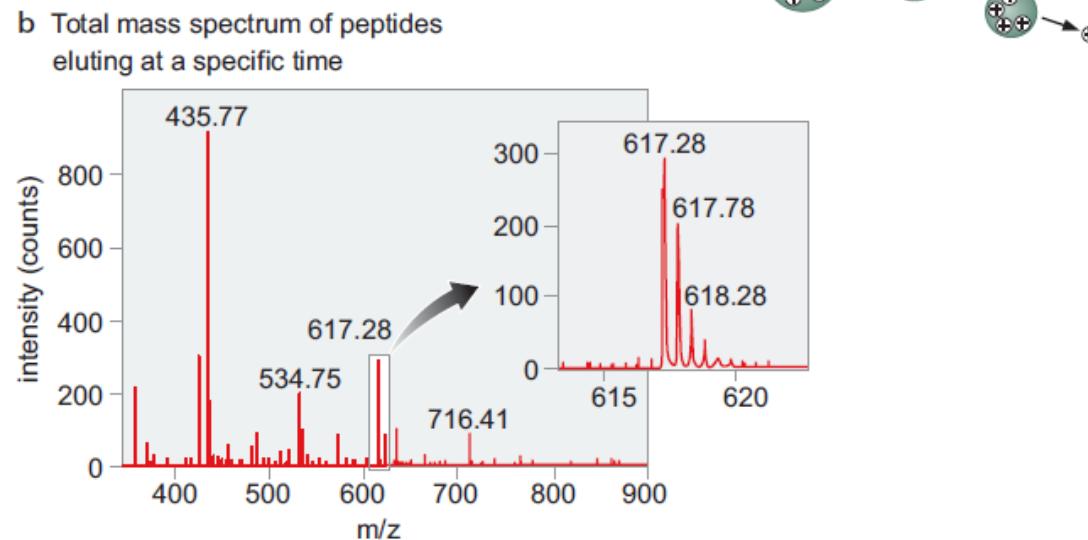
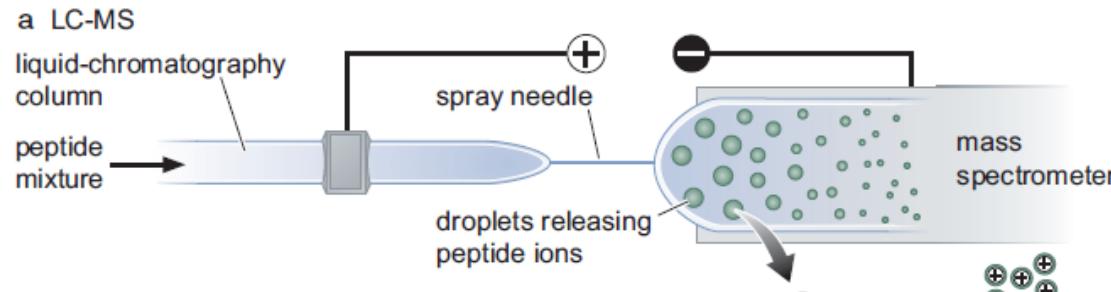
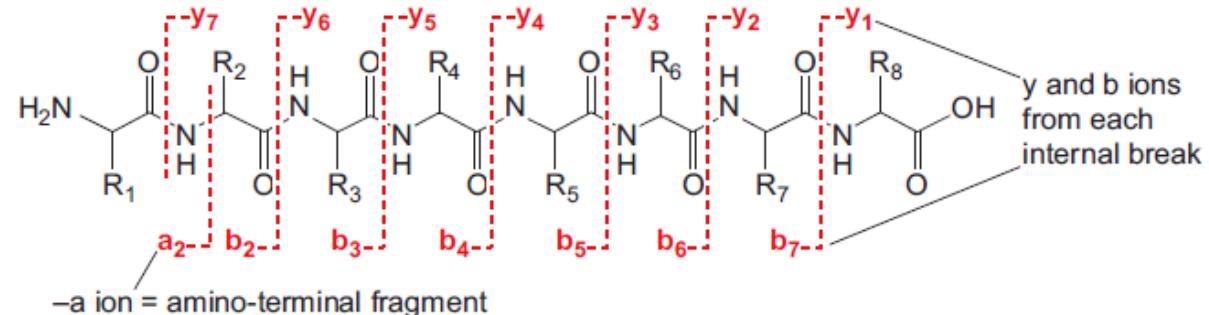


FIGURE 7-29 Protein sequencing by Edman degradation. The amino-terminal residue is labeled and can be removed without hydrolyzing the rest of the peptide. Thus, in each round, one residue is identified, and that residue represents the next one in the sequence of the peptide.

Mass spectrometry



c Common peptide fragments used for MS/MS sequence determination



d Predicted and observed spectra are used to give a confidence score for identification

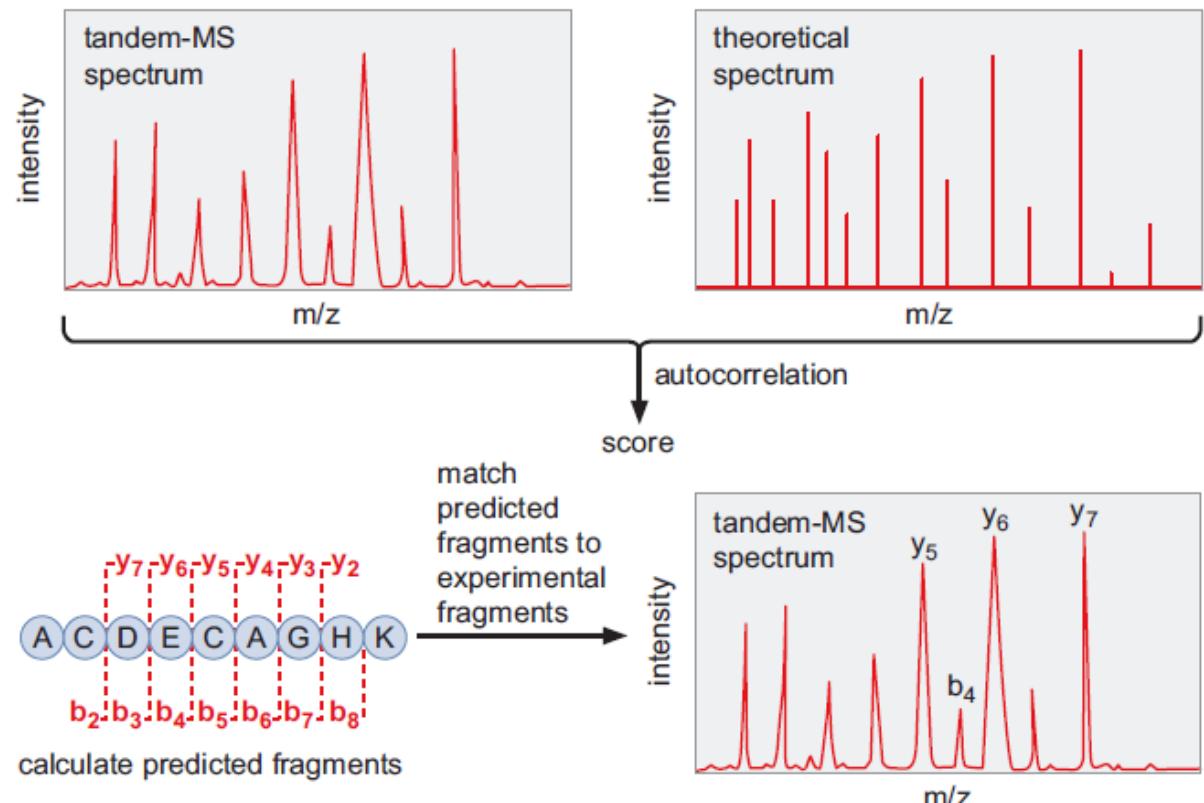
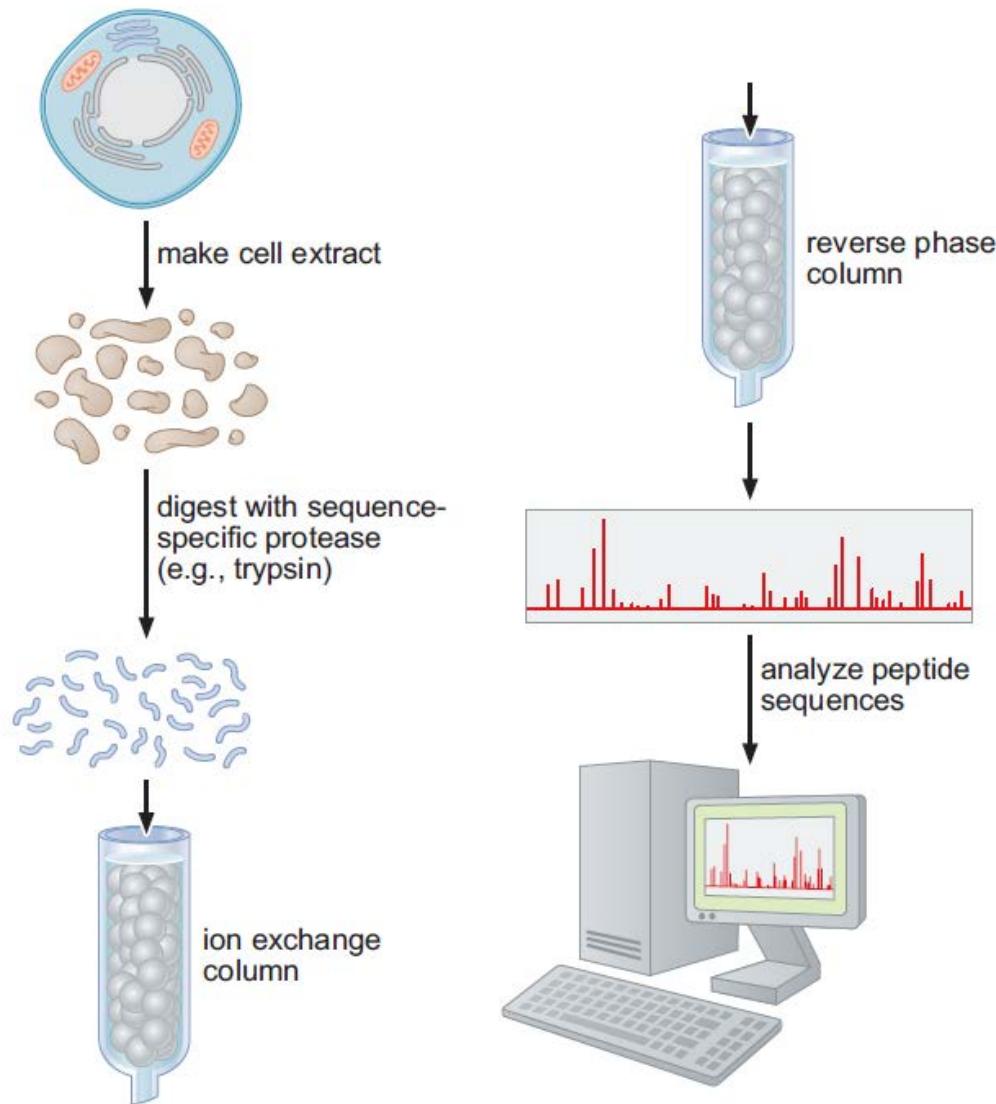


FIGURE 7-30 Using liquid chromatography-MS/MS to analyze the content of a protein mixture. (a) A peptide

Large-scale proteomics – what everybody uses



A powerful method to identify all of the proteins in a complex mixture such as a crude cell extract uses combination of liquid chromatography and mass spectrometry.

FIGURE 7-31 Separation of proteins by liquid chromatography followed by mass spectrometric analysis. The steps of the method are illustrated in the figure and described in the text.

Protein-protein interactions determined by co-precipitation

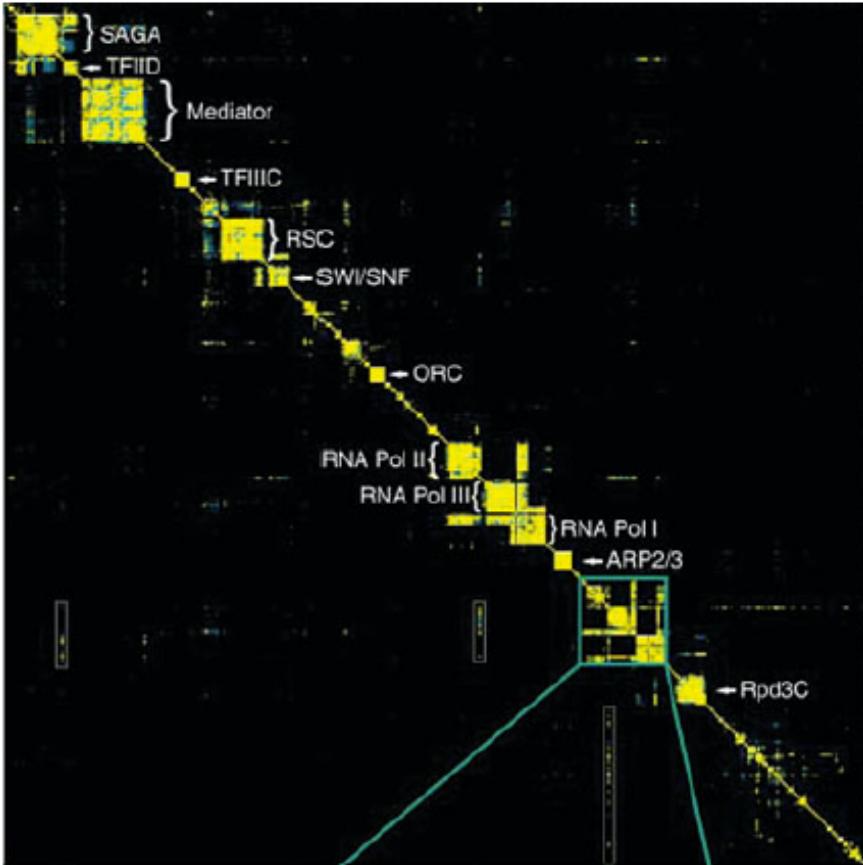


FIGURE 7-32 The physical interactome map of *S. cerevisiae*. Shown here are the results of affinity purification/mass spectrometry studies of all of the proteins in *S. cerevisiae*. The figure is actually composed of a series of columns of boxes indicating which proteins coprecipitated with a given protein. If a protein is coprecipitated with the “tagged” protein, the box is yellow. If not, the box is black. In this view, proteins that are in the same complex have been clustered together on both the vertical and horizontal axes; thus, complexes are observed on the diagonal. A subset of all of the complexes (many of which are discussed elsewhere in the text) are labeled and shown in the image presented here. (Reprinted, with permission, from Collins S.R. et al. 2007. *Mol. Cell. Proteom.* 6: 439–450, Fig. 3b. © American Society for Biochemistry and Molecular Biology.)

DNA-protein interactions determined by mobility shift

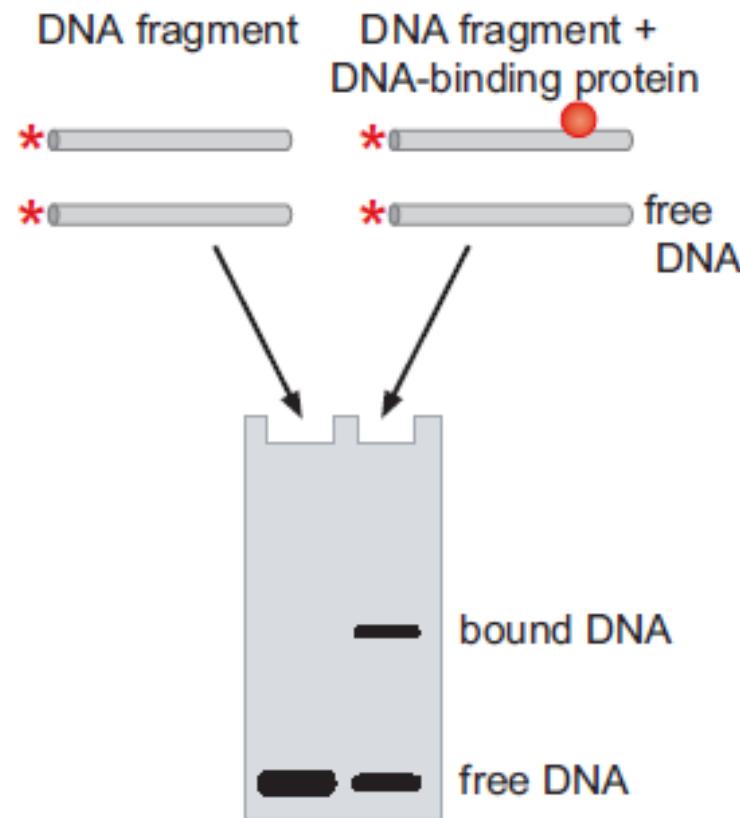


FIGURE 7-33 Electrophoretic mobility-shift assay. The principle of the mobility-shift assay is shown schematically. A protein is mixed with radiolabeled probe DNA containing a binding site for that protein. The mixture is resolved by acrylamide gelelectrophoresis and visualized using autoradiography. DNA not mixed with protein runs as a single band corresponding to the size of the DNA fragment (left lane). In the mixture with the protein, a proportion of the DNA molecules (but not all of them at the concentrations used) binds the DNA molecule. Thus, in the right-hand lane, there is a band corresponding to free DNA and another corresponding to the DNA fragment in complex with the protein.

Nuclease protection footprinting

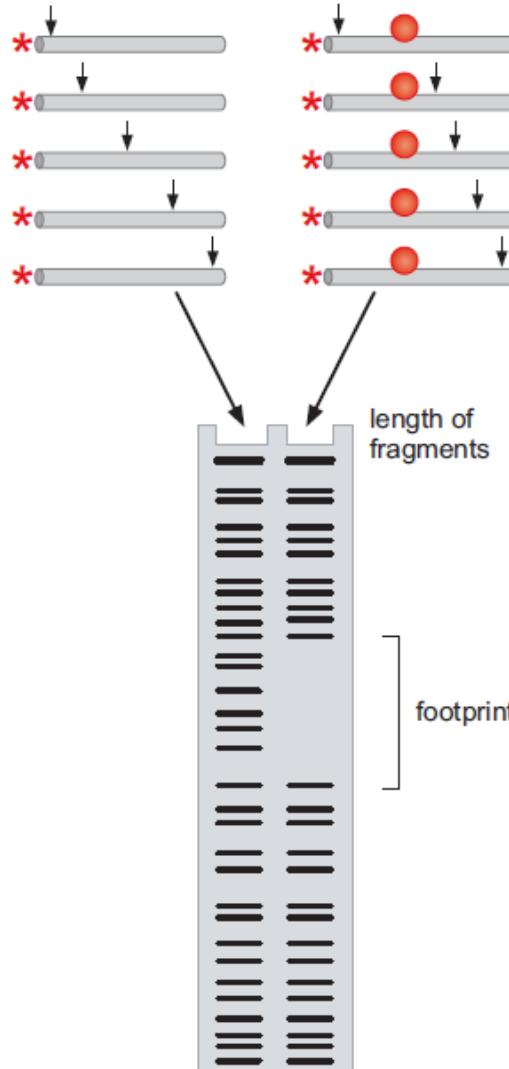
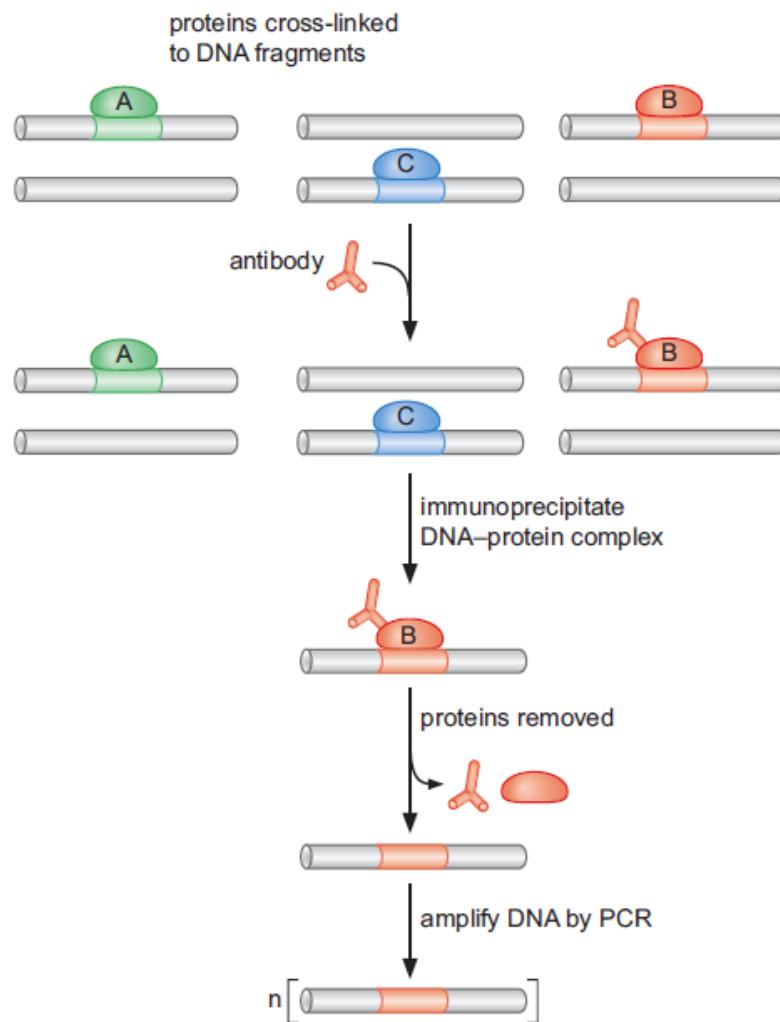


FIGURE 7-34 Nuclease protection footprinting. (Stars) The radioactive labels at the ends of the DNA fragments; (arrows) sites where DNase cuts; (red circles) *Lac* repressor bound to operator. On the left, DNA molecules cut at random by DNase are separated by size using gel electrophoresis. On the right, DNA molecules are first bound to repressor and then subjected to DNase treatment. The “footprint” is indicated on the right. This corresponds to the collection of fragments generated by DNase cutting at sites in free DNA but not in DNA with repressor bound to it. In the latter case, these sites are inaccessible because they are within the operator sequence and hence covered by repressor.

Chromatin immunoprecipitation (ChIP)



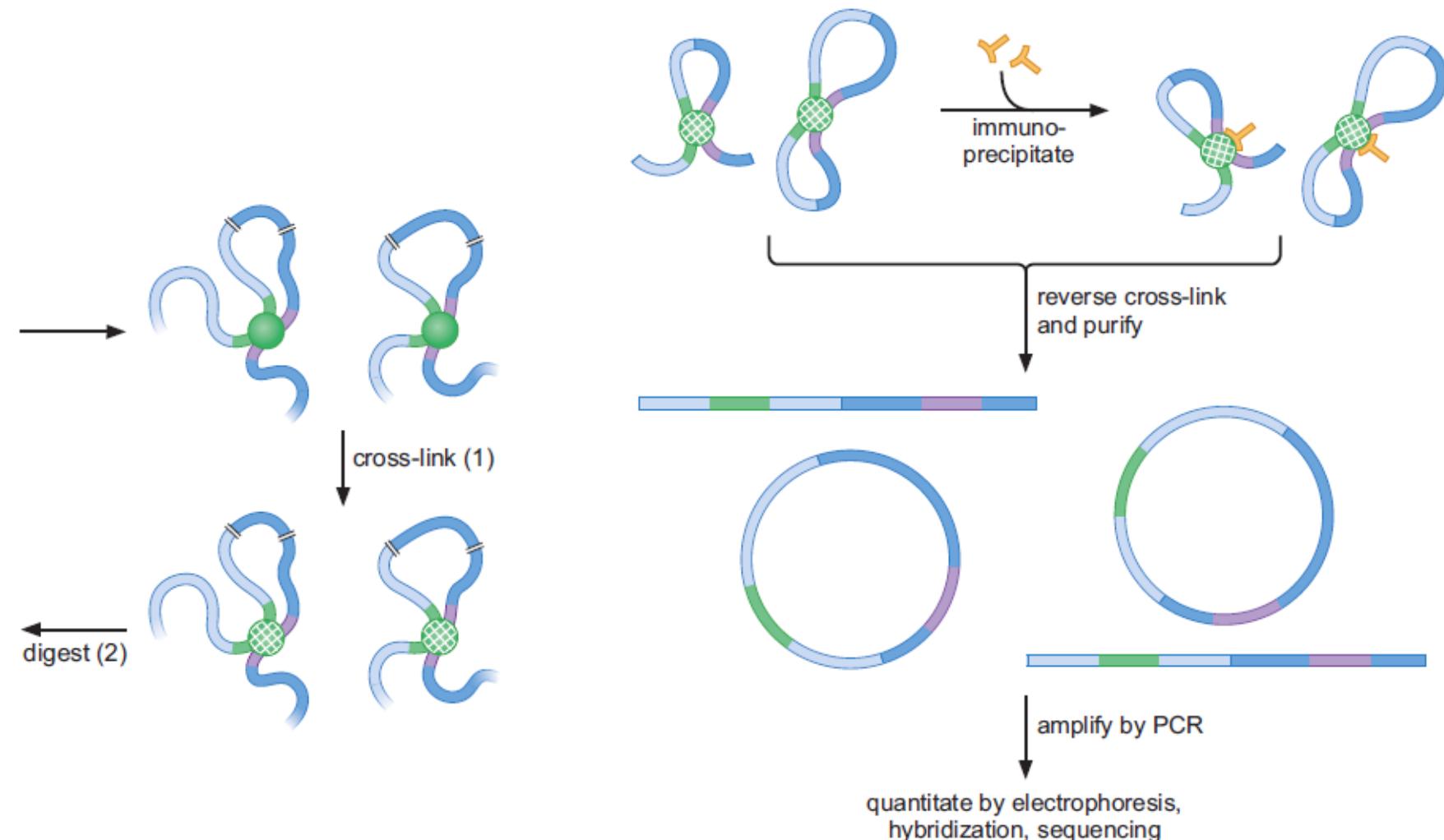
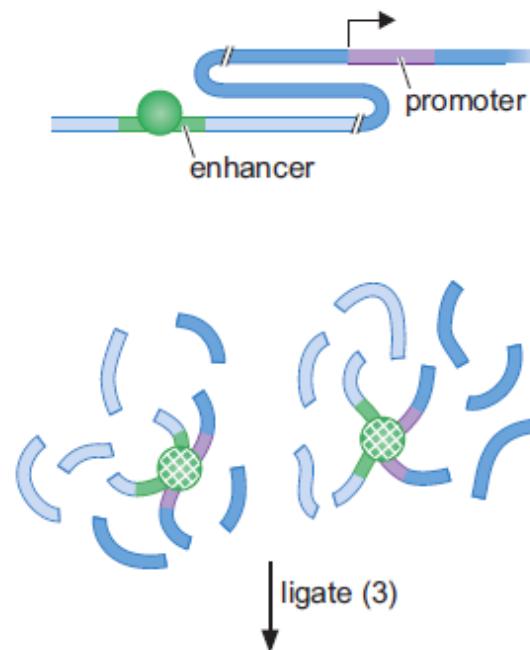
This can be accomplished by one of two basic approaches. To determine if a particular region of DNA (e.g., a promoter) is bound by the protein of interest, PCR can be performed using primers that are targeted to that region. If the protein was bound to that DNA at the time of cross-linking, the sequence will be present in the IP and will be amplified. There are two important controls that are generally included in this assay. First, PCR primers targeting another region of DNA (one to which the protein is known or expected not to bind) are used; in that case, no DNA should be amplified (Fig. 7-35).

FIGURE 7-35 Chromatin immunoprecipitation (ChIP).

Chromosome Conformation Capture

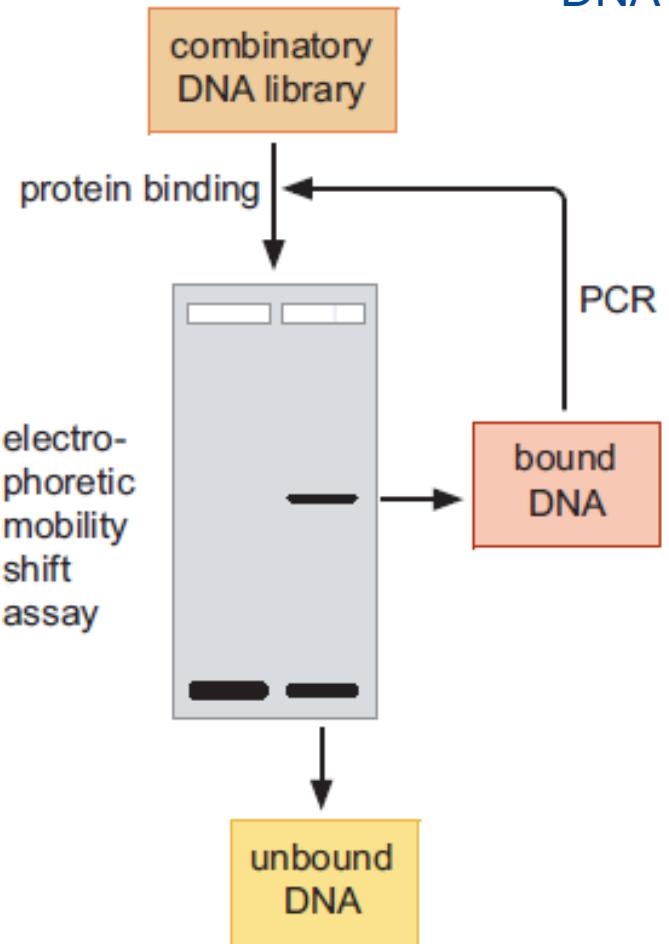
Chromosome conformation capture assays (3C) can be used to detect interactions. The method, illustrated in Figure 7-36, works as follows: the treatment of intact cells with formaldehyde serves to link interacting genomic regions by crosslinking proteins to DNA and proteins to other proteins. The chromatin is then broken up by digestion with restriction endonucleases or by physical disruption, such as sonication. The resulting DNA is subjected to ligation under conditions that favor intramolecular ligation of the associated DNA fragments. At this point, the cross-linking is reversed and the ligation mixture is purified.

FIGURE 7-36 Chromosome Conformation Capture schematic. 3C assays involve three basic steps: (1) interacting chromosome segments are cross-linked with formaldehyde, (2) the DNA is digested, and (3) cross-linked DNA fragments are ligated to produce products that are amplified and can be further analyzed.



SELEX (Systematic Evolution of Ligands by Exponential Enrichment)

In Vitro Selection Can Be Used to Identify a Protein's DNA- or RNA-Binding Site



One powerful approach, called in vitro selection or SELEX (for Systematic Evolution of Ligands by Exponential Enrichment), involves the use of the sequence specificity of the protein to probe a diverse library of oligonucleotides.

By characterizing the enriched DNA, the sequences that bind tightly to the protein can be identified.

FIGURE 7-37 Invitro selection scheme. A combinatorial DNA library in which the middle 10–12 bases are randomized is bound to the protein of interest. Protein-bound DNA is separated from unbound DNA using an EMSA. Bound DNA is eluted from the gel and subjected to PCR using primers directed against constant regions flanking the random regions of the DNA. These sequences are subjected to two to five more cycles of binding and enrichment to identify the highest affinity.

RB69 RegA SELEX logo
11 sites; $Rs = 19.37 \pm 0.68$ bits

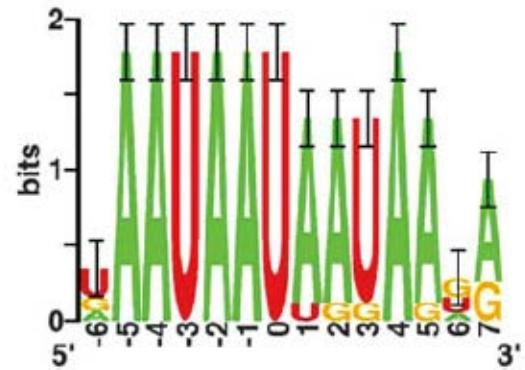
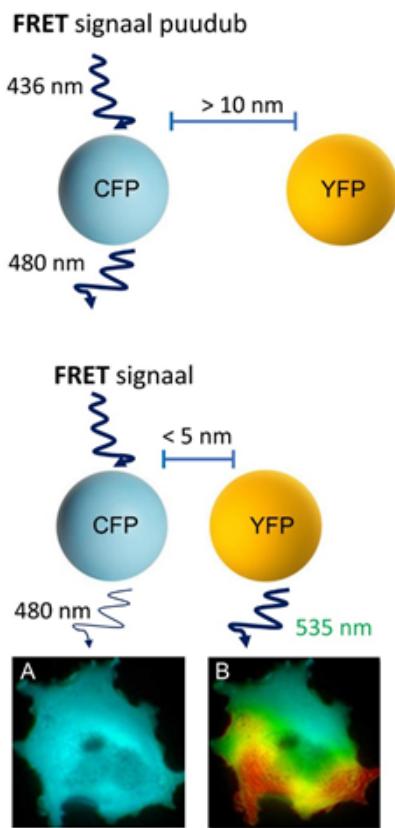
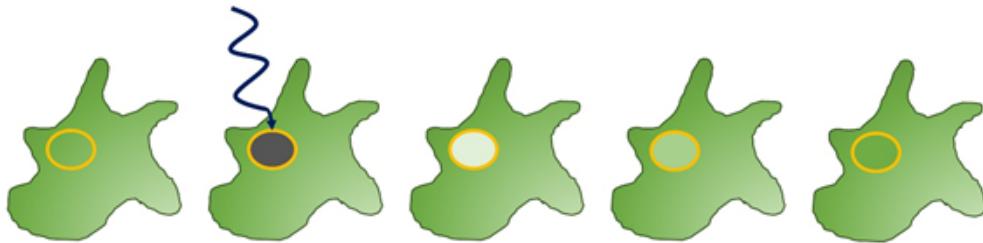


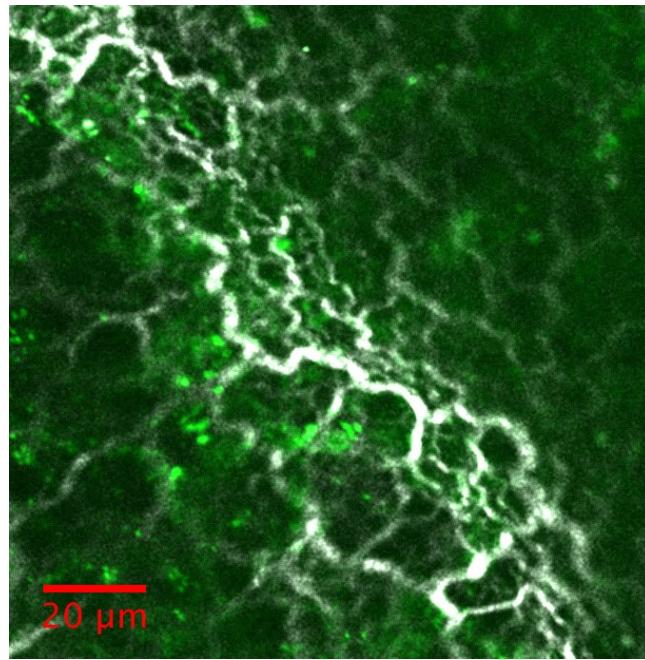
FIGURE 7-38 SELEX sequence logo. In vitro selection was used to isolate RNAs that bind the translational repressor protein RB69 RegA. The image shows the logo of selected sequences. The letter height is proportional to the frequency of each base at that position, with the most frequently occurring base at the top. (Reprinted, with permission, from Dean T.R. et al. 2005. *Virology* 336: 26–36, Fig. 4a. © Elsevier.)



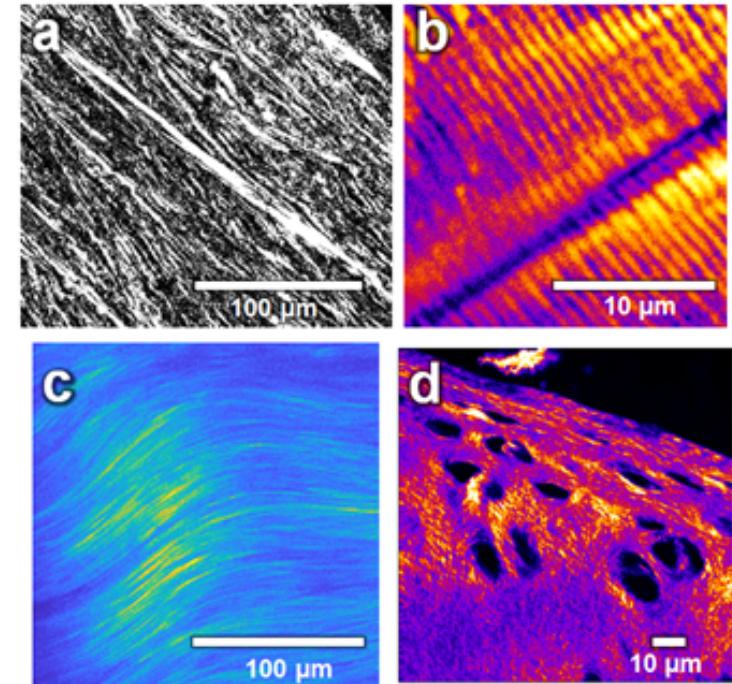
FRET (Förster or fluorescence resonance energy transfer) microscopy



FRAP (Fluorescence recovery after photobleaching microscopy)



Second harmonic image of collagen (shown in white) in liver



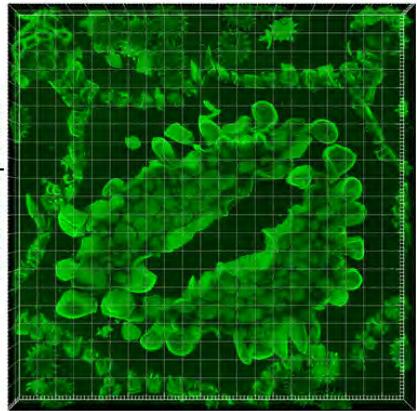
SHIM (second-harmonic imaging microscopy)
 a. human eye cornea; b. zebrafish skeletal muscle myosin; c. cartilage of *Mus musculus*; d. horse cartilage of knee





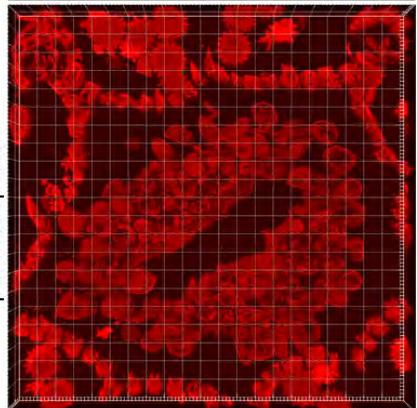
Ülaltvaade

Autofluoresentsens
rohelises spektris

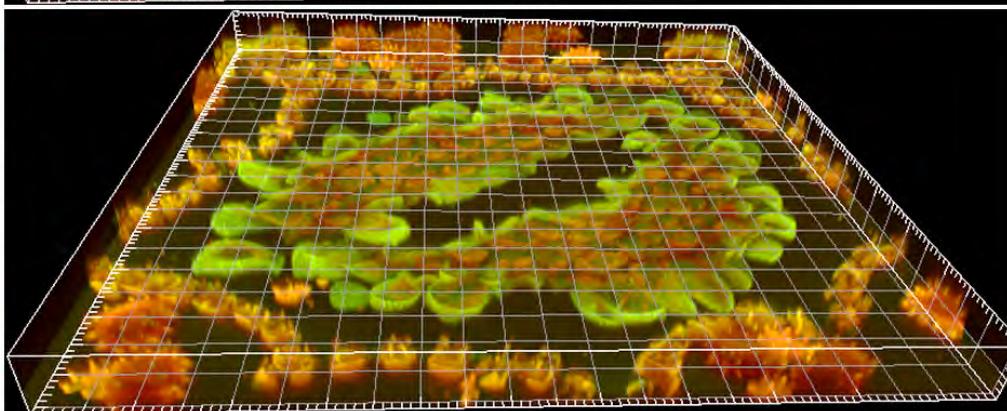
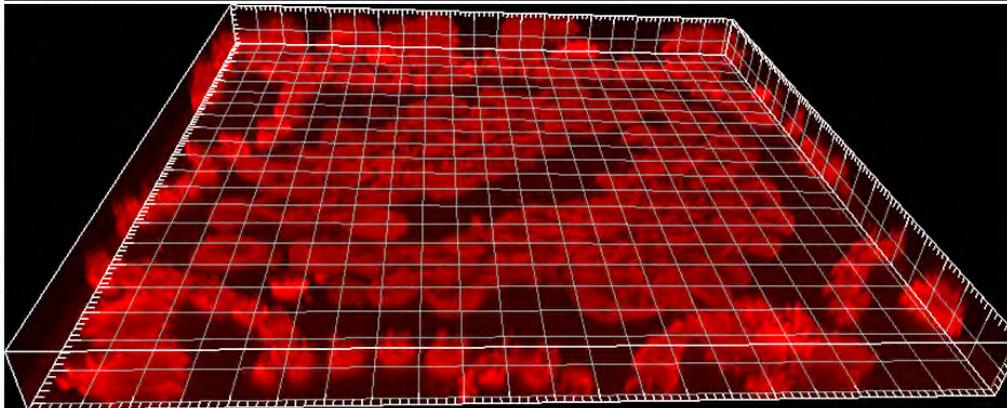
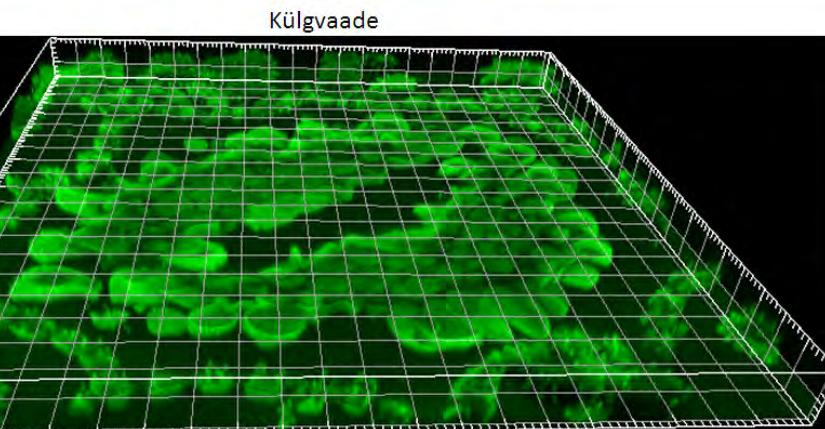
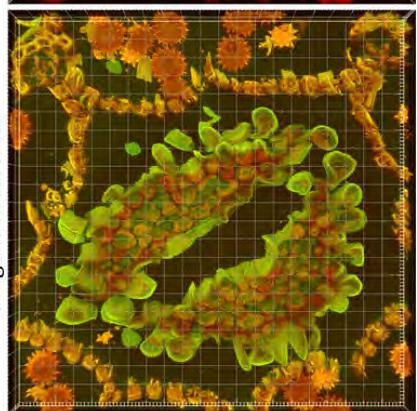


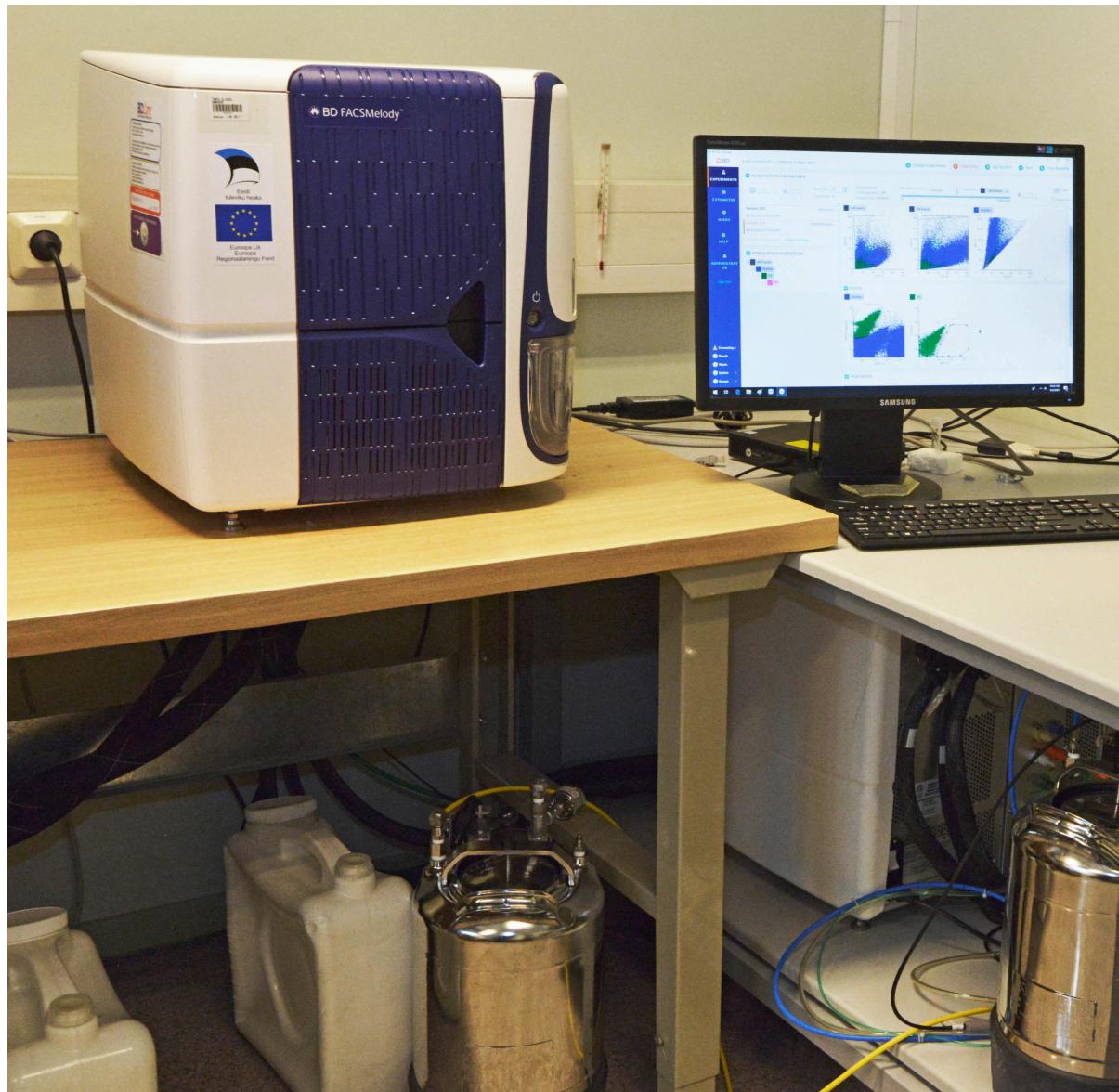
Külgvaade

Autofluoresentsens
punases spektris



Kahe pildi kokku-
langemise tulemus





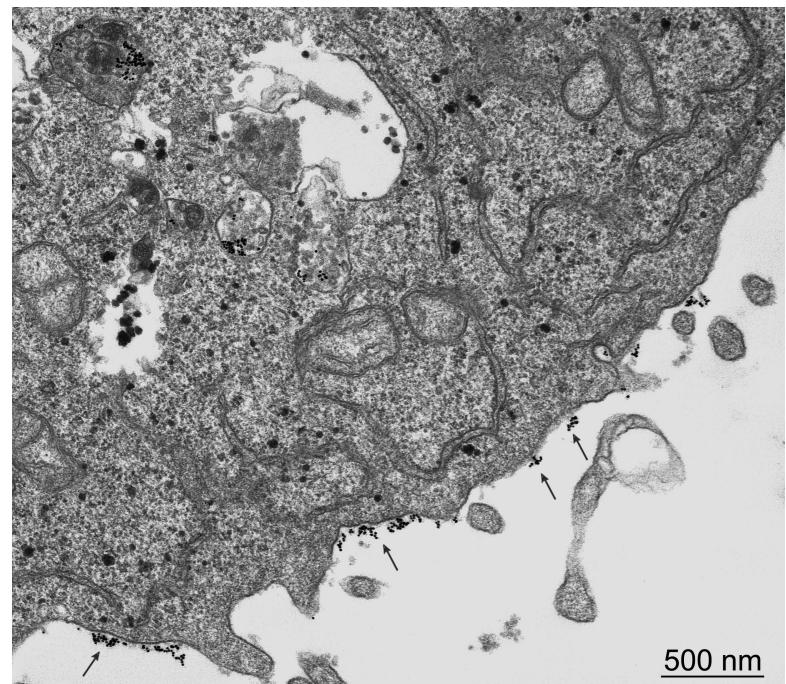


Valgusmikroskoopia -
light microscopy
resolution until 0,2 micrometers

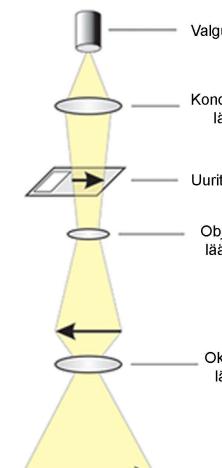


Transmissioon elektronmikroskoopia
transmission electron microscopy
resolution until 0,2 nm

Skaneeriv elektronmikroskoopia-
scanning electron microscopy
resolution until 50 nm

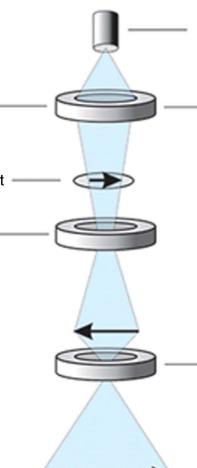


Valgusmikroskoopia



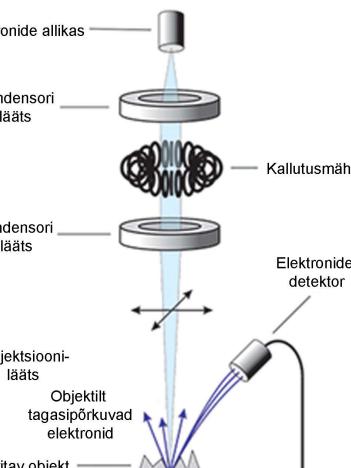
Kujutis objektist on läbi objektiivi silmaga otse jäigitar

Transmissioon-elektronmikroskoopia



Objektiil läbinud elektronidest saadud kujutis kuvatakse fluoresseerulal ekraanil või arvutikuvana

Skaneeriv elektronmikroskoopia



Objektiilt tagasisõeegeldunud elektronide voog kuvatakse arvutikuvana

