Standard hybridization assays using labeled probes to detect immobilized nucleic acids

Nucleic acid hybridization has long been popularly used to detect nucleic acids that have been immobilized in some way. In some cases, the nucleic acid samples have been isolated from cells and immobilized by binding them to a solid support such as a nylon membrane. That can be done very simply by injecting individual nucleic acid samples onto the membrane so that they are bound at regular intervals (dot blot hybridization). In other cases, the isolated nucleic acids may have been size-fractionated by gel electrophoresis before being transferred to the solid support. Alternatively, the hybridization assay is designed to detect nucleic acids *in situ* within native structures such as cells or chromosomes.

Southern and northern blot hybridization

In Southern hybridization assays, a sample population of purified DNA is digested with one or more restriction endonucleases, generating fragments that are several hundred to many thousands of base pairs in length. The restriction fragments are separated according to size by agarose gel electrophoresis, denatured, and transferred to a nitrocellulose or nylon membrane. Labeled probes are hybridized to the membrane-bound target DNA and the positions of the labeled heteroduplexes are revealed by autoradiography (**Figure 6.15**). Although this method has largely been superseded by PCR assays, it is still used in diagnostic assays that seek to identify large DNA changes (which are difficult to detect by PCR).

Northern blot hybridization is a variant of Southern blotting in which the samples contain undigested size-fractionated RNA instead of DNA. In the past, this method was regularly used to obtain information on which tissues genes were expressed in, and to identify tissue-specific isoforms. However, it has been superseded by RT-PCR and sequencing assays.

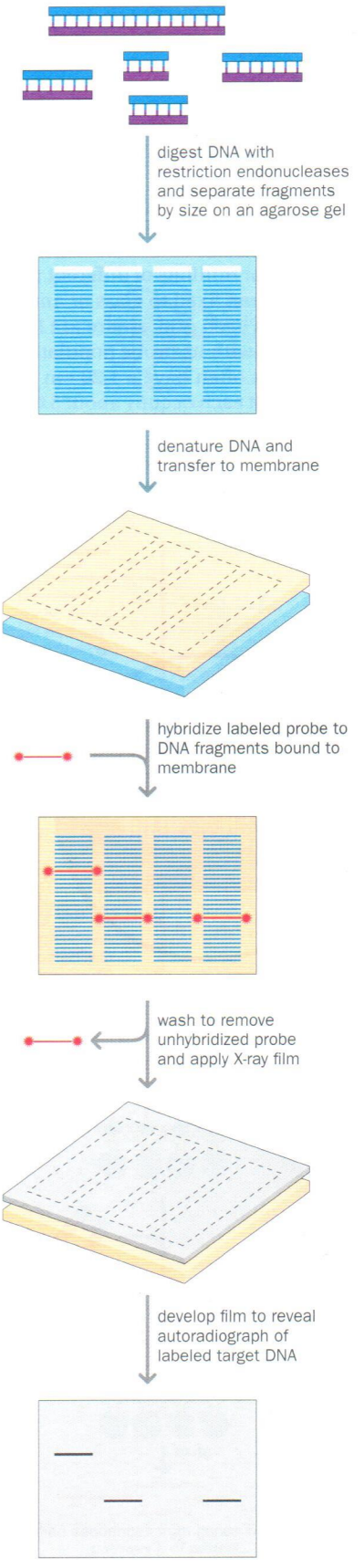


Figure 6.15 Southern blot hybridization. A complex DNA sample is digested with restriction endonucleases. The resulting fragments are applied to an agarose gel and separated by size using electrophoresis. The gel is treated with alkali to denature the DNA fragments and is then placed against a nitrocellulose or nylon membrane. DNA will be transferred from the gel to the membrane, which is then soaked in a solution containing a radiolabeled, single-stranded DNA probe. After hybridization, the membrane is washed to remove excess probe and then dried. The membrane is then placed against an X-ray film and the position of the labeled probe will cause a latent image on the film that can be revealed by development of the autoradiograph as a hybridization band.

Chromosome *in situ* hybridization

In **chromosome *in situ* hybridization**, a suitable labeled DNA probe is hybridized against chromosomal DNA that has been denatured *in situ*. First, an air-dried, microscope-slide chromosome preparation is made, typically using metaphase or prometaphase chromosomes from peripheral blood lymphocytes or lymphoblastoid cell lines. RNA and protein are then removed from the sample by treatment with RNase and proteinase K, and the remaining chromosomal DNA is denatured by exposure to formamide. The denatured DNA is exposed to an added solution containing a labeled nucleic acid probe and overlaid with a coverslip.

Depending on the particular technique, chromosome banding of the chromosomes can be arranged either before or after the hybridization step. As a result, the signal obtained after removal of excess probe can be correlated with the chromosome band pattern in order to identify a map location for the DNA sequences recognized by the probe. Chromosome *in situ* hybridization has been revolutionized by the use of fluorescently labeled probes in fluorescence *in situ* hybridization (FISH) techniques; see **Table 6.3** for cross-references to some figures containing example images.

Tissue *in situ* hybridization

In **tissue *in situ* hybridization**, labeled probes are hybridized against RNA in tissue sections. Very thin tissue sections are cut, either from paraffin wax-embedded tissue blocks or from frozen tissue using a cryostat, and then mounted onto glass slides. A hybridization mix including the probe is applied to the section on the slide and covered with a glass coverslip. Single-stranded antisense RNA probes (riboprobes; see **Box 6.2**) are preferred.

The riboprobes can be labeled with a radioisotope such as ^{33}P or ^{35}S, and the hybridized probes visualized using autoradiographic procedures. The localization of the silver

grains is often visualized using only dark-field microscopy. Here, direct light is not allowed to reach the objective; instead, the illuminating rays of light are directed from the side so that only scattered light enters the microscopic lenses and the signal appears as an illuminated object against a black background. However, better signal detection is possible using bright-field microscopy, in which the image is obtained by direct transmission of light through the sample. Alternatively, probes are subjected to fluorescence labeling and detection is accomplished by fluorescence microscopy.

Microarray hybridization: large-scale parallel hybridization of labeled test-sample nucleic acids to immobilized probes

Innovative and powerful hybridization technologies developed in the early 1990s permit numerous hybridization assays to be simultaneously conducted on a common sample under the same conditions. A DNA or oligonucleotide microarray consists of many thousands or millions of different unlabeled DNA or oligonucleotide probe populations that have been fixed to glass or another suitable surface within a high-density grid format. Within each grid square are large numbers of identical copies of just one probe (a grid square with its probe population is called a **feature**). For example, oligonucleotide microarrays often have a 1.28 cm × 1.28 cm surface that contains very many different features that each occupy about 5 or 10 μm^2 (**Figure 6.16**).

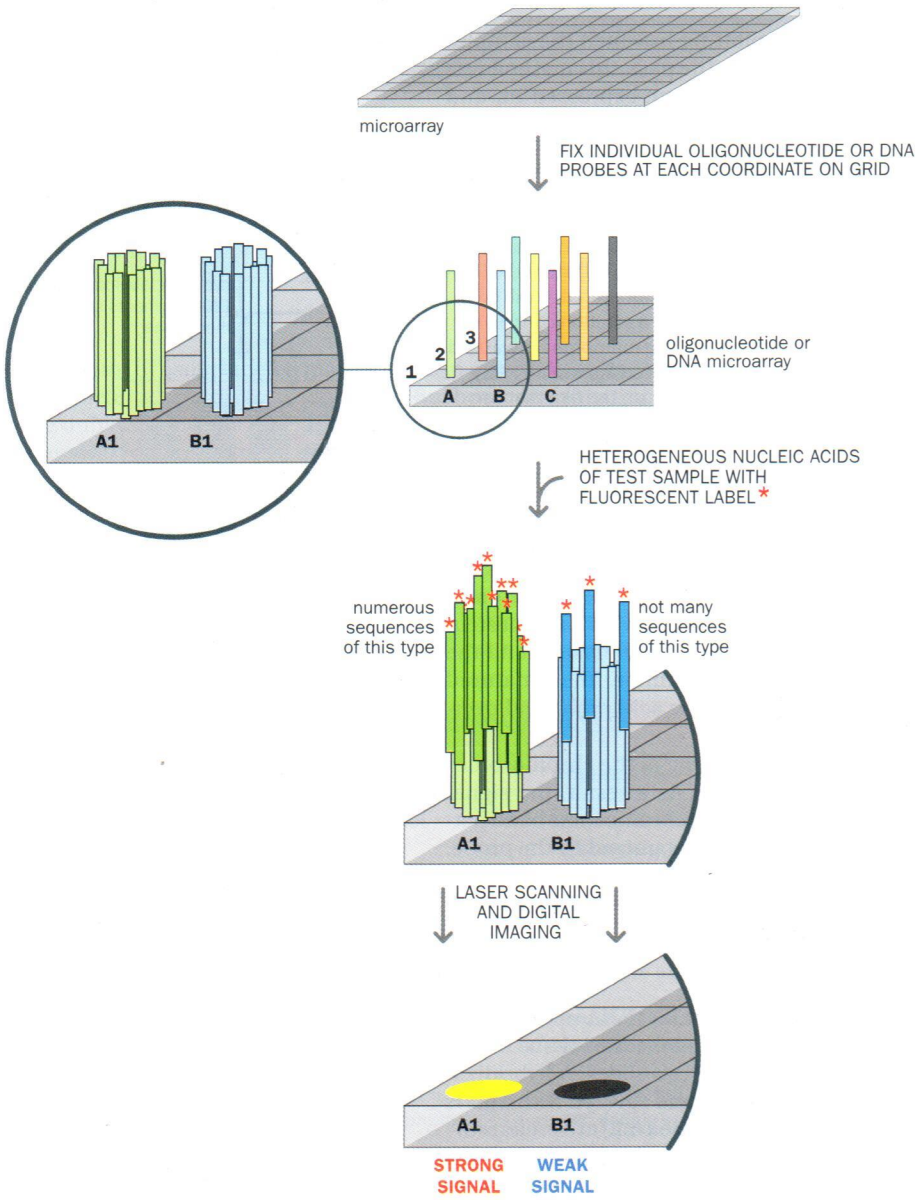


Figure 6.16 Principle of microarray hybridization. A microarray is a solid surface on which molecules can be fixed at specific co-ordinates in a high-density grid format. Oligonucleotide or DNA microarrays have thousands to millions of different synthetic, single-stranded oligonucleotide or DNA probes fixed at specific pre-determined positions in the grid. As shown by the expanded item enclosed within the circle (left), each grid square will have many identical copies of a single type of oligonucleotide or DNA probe (a *feature*). An aqueous test sample containing a heterogeneous collection of labeled DNA fragments or RNA transcripts is denatured and allowed to hybridize with the probes on the array. Some probes (e.g., the A1 feature) may find numerous complementary sequences in the test population, resulting in a strong hybridization signal; for other probes (e.g., the B1 feature) there may be few complementary sequences in the test sample, resulting in a weak hybridization signal. After washing and drying of the microarray, the hybridization signals for the numerous different probes are detected by laser scanning, giving huge amounts of data from a single experiment. (For ease of illustration, we show test-sample nucleic acids with end labels, but normally they would contain labels on internal nucleotides.)

A test sample—an aqueous solution containing a complex population of fluorescently labeled denatured DNA or RNA—is hybridized to the different probe populations on the microarray. After a washing step to remove nonspecific binding of labeled test-sample molecules to the array, the remaining bound fluorescent label is detected using a high-resolution laser scanner. The signal emitted from each feature on the array is analyzed using digital imaging software that converts the fluorescent hybridization signal into one of a palette of colors according to its intensity (see **Figure 6.16**).

Because the intensity of each hybridization signal reflects the amount of labeled molecules that have bound to a feature, microarray hybridization is used to quantitate different sequences in complex test-sample populations such as different samples of genomic DNA or total cellular RNA (or cDNA). As described in later chapters, frequent applications include quantifying different transcripts (expression profiling) and also scanning genomes to look for large-scale deletions and duplications.

Nucleic acid hybridization as a method to selectively purify desired nucleic acid sequences

Up until this point we have considered how nucleic acid hybridization is commonly used as an assay. That is, a well-known nucleic acid or oligonucleotide probe population is used to interrogate some other, less understood nucleic acid population, and the object is simply to get some information about the latter population (we might want to know the sizes of the hybridizing fragments, or their copy number, or where they are expressed in cells and tissues, and so on). But we can also use nucleic acid hybridization for a different purpose: to selectively purify a desired type of nucleic acid sequence. That is made possible by covalently attaching to the probe molecules some other molecule that can bind with very high specificity to "capture molecules" immobilized on a surface.

The most popular approach is to covalently attach biotin to the probe. Biotin, a naturally occurring vitamin (known as vitamin B7 or vitamin H), just happens to have an extraordinarily high affinity for streptavidin, a protein that originates from a *Streptomyces* bacterium. (A homotetramer of streptavidin can bind biotin with a dissociation constant of about 10^{-14} mol/L, one of the strongest noncovalent interactions between natural biological molecules; the strength of the interaction depends on the formation of numerous hydrogen bonds and van der Waals interactions between biotin and streptavidin.)

After biotin-linked probe molecules have been allowed to hybridize to complementary sequences in a test sample, the probe–test-sample heteroduplexes can be captured using magnetized beads coated with streptavidin. The magnetized beads, with attached probe–test-sample heteroduplexes, can then be selectively removed using a magnet and the desired nucleic acids can be eluted (**Figure 6.17**).

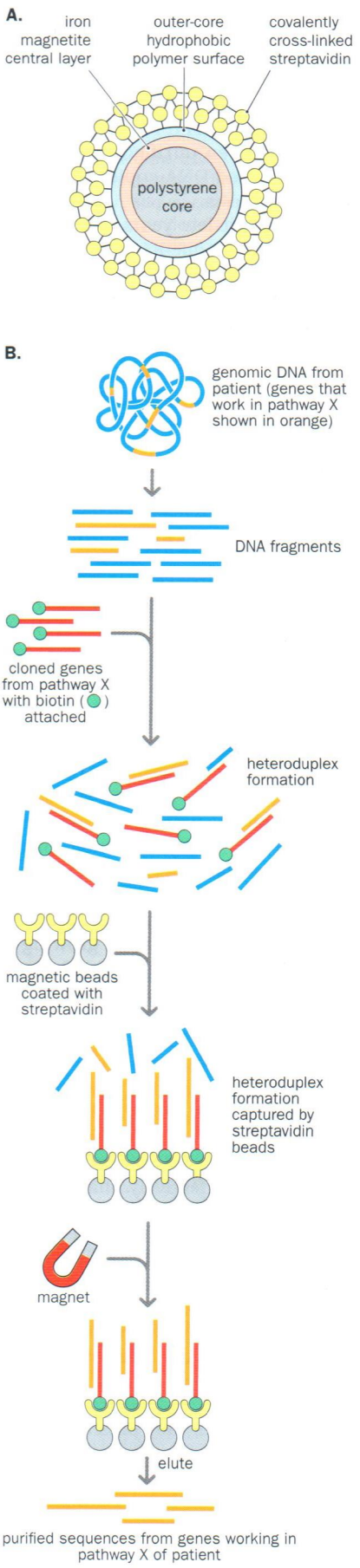


Figure 6.17 Selective purification of desired nucleic acid sequences using nucleic acid hybridization. (**A**) Structure of streptavidin-coated magnetic beads. (**B**) Example of purification. Here we imagine a situation where we wish to purify sequences from a family of many genes that work in some common cell signaling pathway, X, and are considered as possible candidates for having a disease-causing mutation in a group of patients. The idea is to take genomic DNA from individual patients, fragment it, and then mix the fragments with a combination of cloned DNA sequences representing all the normal genes that work in pathway X. The DNA sequences in the mixture are then denatured and allowed to re-anneal, allowing heteroduplexes to form between biotinylated pathway X sequences and complementary sequences from the patient. After streptavidin-coated magnetic beads are added, heteroduplexes with an end-biotin group are bound via the streptavidin to the beads, and can be selectively removed using a magnet. The desired X pathway gene sequences in the patient can be eluted by heating the sample so as to break the hydrogen bonds of the heteroduplex. The resulting purified sequences from each patient can then be investigated by DNA sequencing.

6.4 DNA SEQUENCING PRINCIPLES AND SANGER DIDEOXY SEQUENCING

DNA sequencing means working out the linear sequence of the four bases (A, C, G, and T) in fragments of DNA, usually fragments that have been amplified by PCR or DNA cloning. By working out the base sequence of many short DNA fragments with overlapping DNA sequences, one can get complete sequences for whole genes and whole genomes. The standard method does not distinguish between modified bases and the unmodified version of the base (such as between 5-methylcytosine and cytosine, or between N^6-methyladenine and adenine)—more specialized methods can do that job,

as described for 5-methylcytosine/cytosine in **Box 10.3**. Note that DNA sequencing is also used to infer the linear sequence of bases in RNA molecules (the RNA must first be converted into a complementary DNA sequence using a reverse transcriptase).

Until quite recently, dideoxy DNA sequencing was essentially the only DNA sequencing method that was used (it was first published in 1997 by Fred Sanger, earning him his second Nobel Prize). It relies on amplifying individual DNA sequences that have been purified by DNA cloning or PCR. For each amplified DNA, nested sets of labeled DNA copies are made and then separated according to size using gel electrophoresis.

In the last few years, completely different technologies have allowed massively-parallel DNA sequencing. No attempt is made to obtain the sequence of just a purified DNA component; instead, millions of DNA fragments present in a complex DNA sample are simultaneously sequenced without the need for gel electrophoresis.

Dideoxy DNA sequencing provides highly accurate DNA sequences and is still widely used for investigating specific DNA sequences, such as testing whether individuals have mutations in a particular gene or confirming a suspected mutation. What the newer DNA sequencing technologies offer is a dramatic increase in sequencing capacity and the ability to sequence complex DNA populations, such as genomic DNA sequences, very rapidly. As a result of fast-developing technology, the running costs of DNA sequencing are plummeting and very rapid sequencing of whole exomes and even whole genomes is becoming routine.

The basics of Sanger dideoxy DNA sequencing

Like PCR, dideoxy DNA sequencing uses primers and a DNA polymerase to make DNA copies of specific DNA sequences of interest. To obtain enough DNA for sequencing, the DNA sequences are amplified using PCR (or sometimes by cloning in cells). The resulting purified DNAs are then sequenced, one after another, in individual reactions. Each reaction begins by denaturing a selected purified DNA. A single oligonucleotide primer is then allowed to bind and used to make labeled DNA copies of the desired sequence (using a provided DNA polymerase and the four dNTPs).

Instead of making full-length copies of the sequence, the DNA synthesis reactions are designed to produce a population of DNA fragments sharing a common 5′ end sequence (defined by the primer sequence) but with variable 3′ ends. This is achieved by simultaneously having present the standard dNTP precursors of DNA plus low concentrations of ddNTPs, dideoxynucleotide analogs that differ from a standard deoxynucleotide only in that they lack an OH group at the 3′ carbon of the sugar as well as at the 2′ carbon (**Figure 6.18A**).

DNA synthesis continues smoothly when dNTPs are used, but once a dideoxynucleotide is incorporated into a growing DNA chain, chain synthesis is immediately terminated (the dideoxynucleotide lacks a 3′-OH group to form a phosphodiester bond). To keep the balance tilted toward chain elongation, the ratio of each ddNTP to the corresponding dNTP is set to be about 1:100, so that a dideoxynucleotide is incorporated at only about 1% of the available nucleotide positions.

If we consider competition between ddATP and dATP in the example in **Figure 6.18B**, there are four available positions for nucleotide insertion: opposite the T at nucleotide positions 2, 5, 13, and 16 in the starting DNA. Because the DNA synthesis reaction results in numerous DNA copies, then, by chance, some copies will have a dideoxyA incorporated opposite the T at position 2, some will have a dideoxyA opposite the T at position 5, and so on. Effectively, chain elongation is randomly inhibited, producing sets of DNA strands that have a common 5′ end but variable 3′ ends.

Fluorescent dyes are used to label the DNA. One convenient way of doing this, as shown in **Figure 6.18B**, is to arrange that the four different ddNTPs are labeled with different fluorescent dyes. The reaction products will therefore consist of DNA strands that have a labeled dideoxynucleotide at the 3′ end, carrying a distinctive fluorophore according to the type of base incorporated.

All that remains is to separate the DNA fragments according to size using electrophoresis (see **Box 6.3**), and then to detect the fluorescence signals. The latter is usually achieved during gel electrophoresis: as the migrating DNA fragments reach a certain point in the gel, they pass a laser that excites the fluorophores, causing them to emit fluorescence at distinct wavelengths. The fluorescence signals are recorded, and an output is provided in the form of intensity profiles for the differently colored fluorophores (see **Figure 6.18C** for an example).

Dideoxy DNA sequencing is disadvantaged by relying on gel electrophoresis. It is not amenable, therefore, to full automation (slab polyacrylamide gels were used initially;

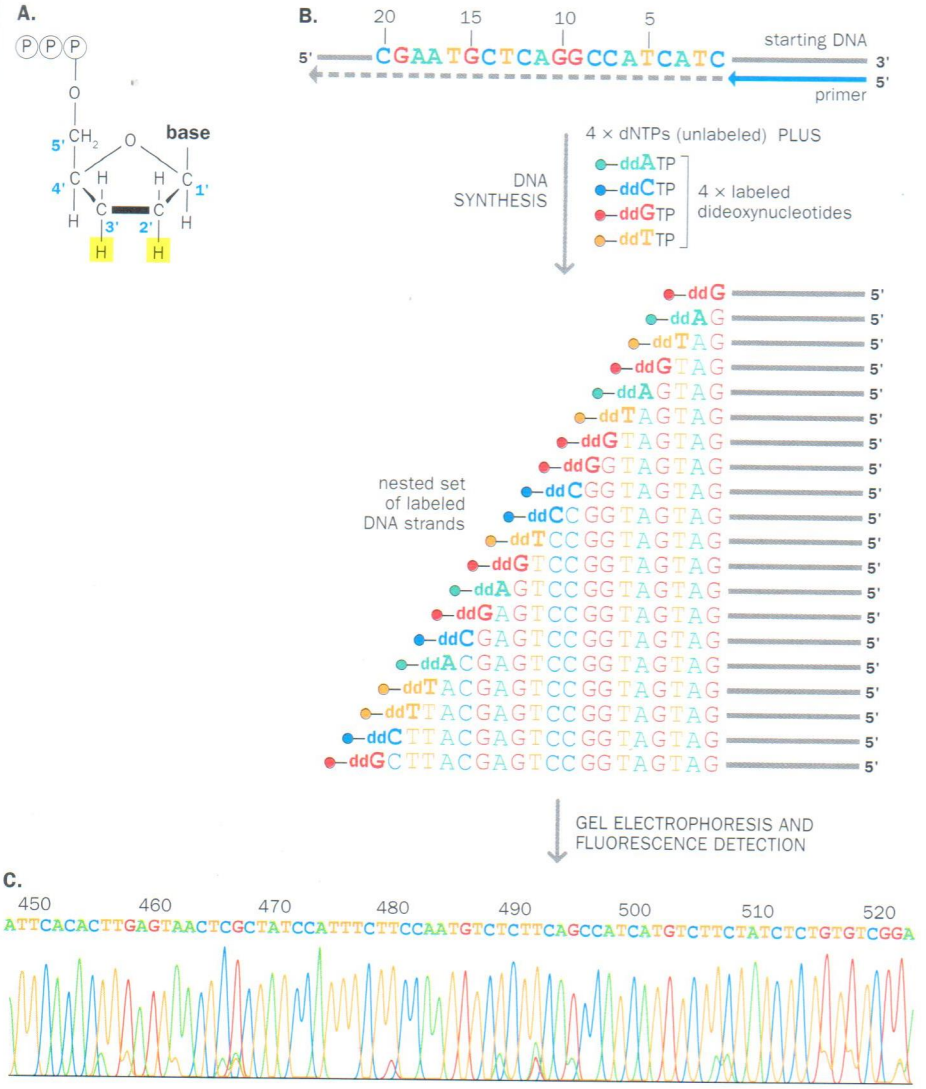


Figure 6.18 Principle of dideoxy sequencing. (**A**) Generalized structure of a 2′,3′ ddNTP. The sugar is *dideoxy*ribose because the hydroxyl groups attached to both carbons 2′ and 3′ of ribose are each replaced by a hydrogen atom (shown by yellow shading). (**B**) In dideoxy sequencing reactions, a DNA polymerase uses an oligonucleotide primer to make complementary sequences from a purified, single-stranded starting DNA. The sequencing reactions include ddNTPs that compete with the standard dNTPs for insertion of a nucleoside monophosphate into the DNA. Different labeling systems can be used but it is convenient to use labeled ddNTPs that have different fluorescent groups according to the type of base, as shown here by colored circles. The DNA copies will have a common 5′ end (defined by the sequencing primer) but variable 3′ ends, depending on where a labeled dideoxynucleotide has been inserted, producing a nested set of DNA fragments that differ by a single nucleotide in length. A series of nested fragments that incrementally differ by one nucleotide from their common 5′ end are fractionated according to size by gel electrophoresis, and the fluorescence signals are recorded and interpreted to produce a linear base sequence. (**C**) Example of DNA sequence output, showing a succession of dye-specific (and therefore base-specific) intensity profiles. This example shows a cDNA sequence from the *PHC3* polyhomeotic gene, provided by E. Tonkin, Newcastle University.

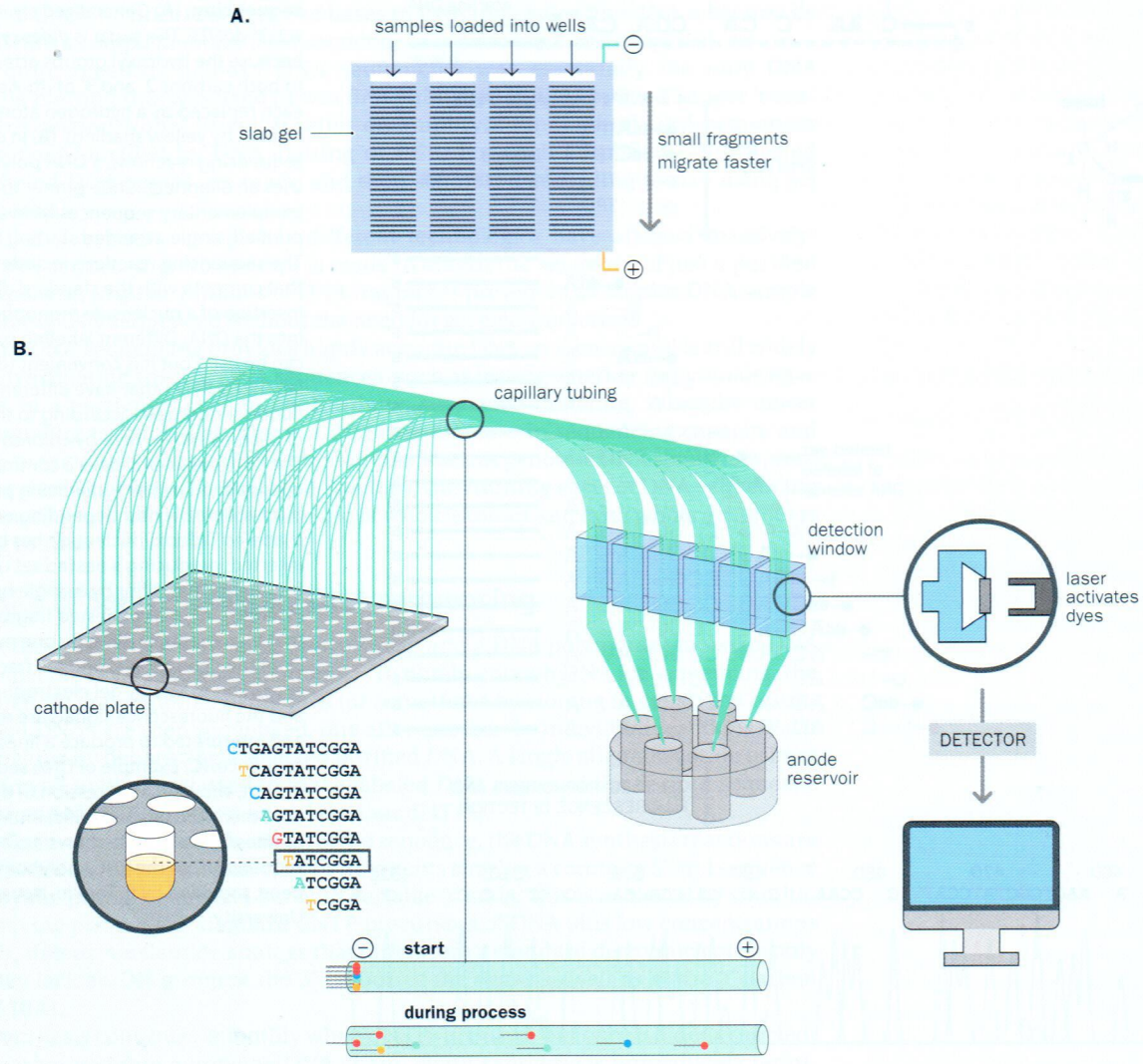
BOX 6.3 SLAB GEL AND CAPILLARY ELECTROPHORESIS FOR SEPARATING NUCLEIC ACIDS ACCORDING TO SIZE

Nucleic acids carry numerous negatively-charged phosphate groups and will migrate toward the positive electrode when placed in an electric field. By arranging that they migrate through a porous gel during electrophoresis, nucleic acid molecules can be fractionated according to size. The porous gel acts as a sieve: small molecules pass easily through the pores of the gel, but larger fragments are more impeded by frictional forces.

Standard gel electrophoresis using agarose gels allows fractionation of moderately large DNA fragments (usually from about 0.1 kb to 20 kb). Pulsed-field gel electrophoresis can be used to separate much larger DNA fragments (up to megabases long). It uses specialized equipment in which the electrical polarity is regularly changed, forcing the DNA molecules to periodically alter their conformation in preparation for migrating in a different direction. Polyacrylamide gel electrophoresis allows superior resolution of smaller nucleic acids (it is usually used to separate fragments in size ranges up to 1 kb), and is used in dideoxy DNA sequencing to separate fragments that differ in length by just a single nucleotide.

In slab gel electrophoresis, individual samples are loaded into cut-out wells at one end of a solid slab of agarose or polyacrylamide gel. They migrate in parallel lanes toward the positive electrode (**Figure 1A**). The separated nucleic acids can be detected in different ways. For example, after the end of an electrophoresis run, the gels can be stained with chemicals such as ethidium bromide or SYBR green that bind to nucleic acids and fluoresce when exposed to ultraviolet light. Sometimes the nucleic acids are labeled with fluorophores prior to electrophoresis, and during electrophoresis a recorder detects fluorescence of individual labeled nucleic acid fragments as they sequentially pass a recorder placed opposite a fixed position in the gel.

The disadvantage of slab gel electrophoresis is that it is labor-intensive. The modern trend is to use capillary gel electrophoresis, which is largely automated. Fluorescently-labeled DNA samples migrate through individual long and very thin tubes containing polyacrylamide gel, and a recorder detects fluorescence emissions as samples pass through a fixed point (**Figure 1B**). Modern Sanger DNA sequencing uses capillary electrophoresis, as do many different types of diagnostic DNA screening methods that we outline in Chapter 20.



Box 6.3 Figure 1 (A) Slab gel electrophoresis and (B) capillary gel electrophoresis. (B, courtesy of the Life Sciences Foundation.)

more modern machines use semi-automated capillary electrophoresis—see **Box 6.3 Figure 1B**). Because gel electrophoresis is not suitable for handling large numbers of samples at a time, dideoxy sequencing can generate a limited amount of sequence data only. It is, therefore, not well suited to genome sequencing. It does, however, provide highly accurate sequences over several hundred bases. In modern times it is often used for analyzing variation over small DNA regions, such as regions encompassing individual exons, and in validating some sequences obtained by newer, high-throughput DNA sequencing methods.

6.5 MASSIVELY-PARALLEL DNA SEQUENCING (NEXT-GENERATION SEQUENCING)

For thirty years, Sanger's dideoxy technique was the standard method of DNA sequencing. Over the years, incremental technical improvements saw radiolabeling replaced by fluorescence labeling, and the development of automated capillary sequencers with increasing numbers of capillaries per machine, but the basic principle was unchanged. Sanger sequencing is a highly reliable way of sequencing a pre-determined stretch of DNA up to around 800 bp in length. Although it was used for obtaining the sequence of