

Automated DNA Sequencing

Large-scale DNA sequencing requires automated procedures based on fluorescence labeling of DNA and suitable detection systems. In general, a fluorescent label can be used either directly or indirectly. Direct fluorescent labels, as used in automated sequencing, are fluorophores. These are molecules that emit a distinct fluorescent color when exposed to UV light of a specific wavelength. Examples of fluorophores used in sequencing are fluorescein, which fluoresces pale green when exposed to a wavelength of 494 nm; rhodamine, which fluoresces red at 555 nm; and aminomethyl coumarin acetic acid, which fluoresces blue at 399 nm. In addition, a combination of different fluorophores can be used to produce a fourth color. Thus, each of the four bases can be distinctly labeled.

Another approach is to use PCR-amplified products (thermal cycle sequencing). This has the advantage that double-stranded rather than single-stranded DNA can be used as the starting material. And since small amounts of template DNA are sufficient, the DNA to be sequenced does not have to be cloned beforehand.

Thermal cycle sequencing

The DNA to be sequenced is contained in vector DNA <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. The primer, a short oligonucleotide with a sequence complementary to the site of attachment on the single-stranded DNA, is used as a starting point. For sequencing short stretches of DNA, a universal primer is sufficient. This is an oligonucleotide that will bind to vector DNA adjacent to the DNA to be sequenced. However, if the latter is longer than about 750 bp, only part of it will be sequenced. Therefore, additional internal primers are required. These anneal to different sites and amplify the DNA in a series of contiguous, overlapping chain

termination experiments <fn>Rosenthal, N.: Fine structure of a gene–DNA sequencing. New Eng. J. Med. 332 :589–591, 1995</fn>. Here, each primer determines which region of the template DNA is being sequenced. In thermal cycle sequencing <fn>Strachan, T., Read, A.P.: Human Molecular Genetics. 2 nd ed. Bios Scientific Publishers, Oxford, 1999.</fn>, only one primer is used to carry out PCR reactions, each with one dideoxynucleotide (ddA, ddT, ddG, or ddC) in the reaction mixture. This generates a series of different chain-terminated strands, each dependent on the position of the particular nucleotide base where the chain is being terminated <fn>Wilson, R.K., et al.: Development of an automated procedure for fluorescent DNA sequencing. Genomics 6 :626–636, 1990.</fn>. After many cycles and with electrophoresis, the sequence can be read as shown in the previous plate. One advantage of thermal cycle sequencing is that double-stranded DNA can be used as starting material.

Automated DNA sequencing (principle)

Automated DNA sequencing involves four fluorophores, one for each of the four nucleotide bases. The resulting fluorescent signal is recorded at a fixed point when DNA passes through a capillary containing an electrophoretic gel. The base-specific fluorescent labels are attached to appropriate dideoxynucleotide triphosphates (ddNTP). Each ddNTP is labeled with a different color, e.g., ddATP green, ddCTP blue, ddGTP yellow, and ddTTP red <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. (The actual colors for each nucleotide may be different.) All chains terminated at an adenine (A) will yield a green signal; all chains terminated at a cytosine (C) will yield a blue signal, and so on. The sequencing reactions based on this kind of chain termination at labeled nucleotides <fn>Rosenthal, N.: Fine structure of a gene–DNA sequencing. New Eng. J. Med. 332

:589–591, 1995 are carried out automatically in sequencing capillaries <fn>Strachan, T., Read, A.P.: Human Molecular Genetics. 2 nd ed. Bios Scientific Publishers, Oxford, 1999.</fn>. The electrophoretic migration of the ddNTP-labeled chains in the gel in the capillary pass in front of a laser beam focused on a fixed position. The laser induces a fluorescent signal that is dependent on the specific label representing one of the four nucleotides. The sequence is electronically read and recorded and is visualized as alternating peaks in one of the four colors, representing the alternating nucleotides in their sequence positions. In practice the peaks do not necessarily show the same maximal intensity as in the schematic diagram shown here. (Illustration based on Brown, 1999, and Strachan and Read, 1999).



Automated DNA sequencing

Genome sequencing

Knowledge of the nucleotide sequence of a gene provides important information about its structure, function, and evolutionary relationship to other similar genes in the same or different organisms. Thus, the development in the 1970s of relatively simple methods for sequencing DNA has had a great impact on genetics. Two basic methods for DNA sequencing have been developed: a chemical cleavage method (A. M. Maxam and W. Gilbert, 1977) and an enzymatic method (F. Sanger, 1981). A brief outline of the underlying principles follows.

Sequencing by chemical degradation

This method utilizes base-specific cleavage of DNA by certain chemicals. Four different chemicals are used in four reactions, one for each base. Each reaction produces a set of DNA fragments of different sizes. The sizes of the fragments in a reaction mixture are determined by positions in the DNA of the nucleotide that has been cleaved. A double-stranded or singlestranded fragment of DNA to be sequenced is processed to obtain a single strand labeled with a radioactive isotope at the 5' end

Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.

This DNA strand is treated with one of the four chemicals for one of the four reactions. Here the reaction at guanine sites(G) by dimethylsulfate (DMS) is shown. Dimethyl sulfate attaches a methyl group to the purine ring of G nucleotides. The amount of DMS used is limited so that on average just one G nucleotide per strand is methylated, not the others (shown here in four different positions of G). When a second chemical, piperidine, is added, the nucleotide purine ring is removed and the DNA molecule is cleaved at the phosphodiester bond just upstream of the site without the base. The overall procedure results in a set of labeled fragments of defined sizes according to the positions of G in the DNA sample being sequenced. Similar reactions are carried out for the other three bases (A, T, and C, not shown). The four reaction mixtures, one for each of the bases, are run in separate lanes of a polyacrylamide gel electrophoresis. Each of the four lanes represents one of the four bases G, A, T, or C. The smallest fragment will migrate the farthest downward, the next a little less far, etc. One can then read the sequence in the direction opposite to migration to obtain the sequence in the 5' to 3' direction (here TAGTCGCAGTACCGTA).

Sequencing by chain termination

This method, now much more widely used than the chemical cleavage method, rests on the principle that DNA synthesis is terminated when instead of a normal deoxynucleotide (dATP, dTTP, dGTP, dCTP), a dideoxynucleotide (ddATP, ddTTP, ddGTP, ddCTP) is used. A dideoxynucleotide (ddNTP) is an analogue of the normal dNTP. It differs by lack of a hydroxyl group at the 3' carbon position. When a dideoxynucleotide is incorporated during DNA synthesis, no bond between its 3' position and the next nucleotide is possible because the ddNTP lacks the 3' hydroxyl group. Thus, synthesis of the new chain is terminated at this site. The DNA fragment to be sequenced has to be single-stranded <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. DNA synthesis is initiated using a primer and one of the four ddNTPs labeled with ^{32}P in the phosphate groups or, for automated sequencing, with a fluorophore (see next plate). Here an example of chain termination using ddATP is shown <fn>Strachan, T., Read, A.P.: Human Molecular Genetics. 2nd ed. Bios scientific Publishers,</fn>. Wherever an adenine (A) occurs in the sequence, the dideoxyadenine triphosphate will cause termination of the new DNA chain being synthesized. This will produce a set of different DNA fragments whose sizes are determined by the positions of the adenine residues occurring in the fragment to be sequenced. Similar reactions are done for the other three nucleotides. The four parallel reactions will yield a set of fragments with defined sizes according to the positions of the nucleotides where the new DNA synthesis has been terminated. The fragments are separated according to size by gel electrophoresis as in the chemical method. The sequence gel is read in the direction from small fragments to large fragments to derive the nucleotide sequence in the 5' to 3' direction. An example of an actual sequencing gel is shown between panel A and B.