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 cumarin 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

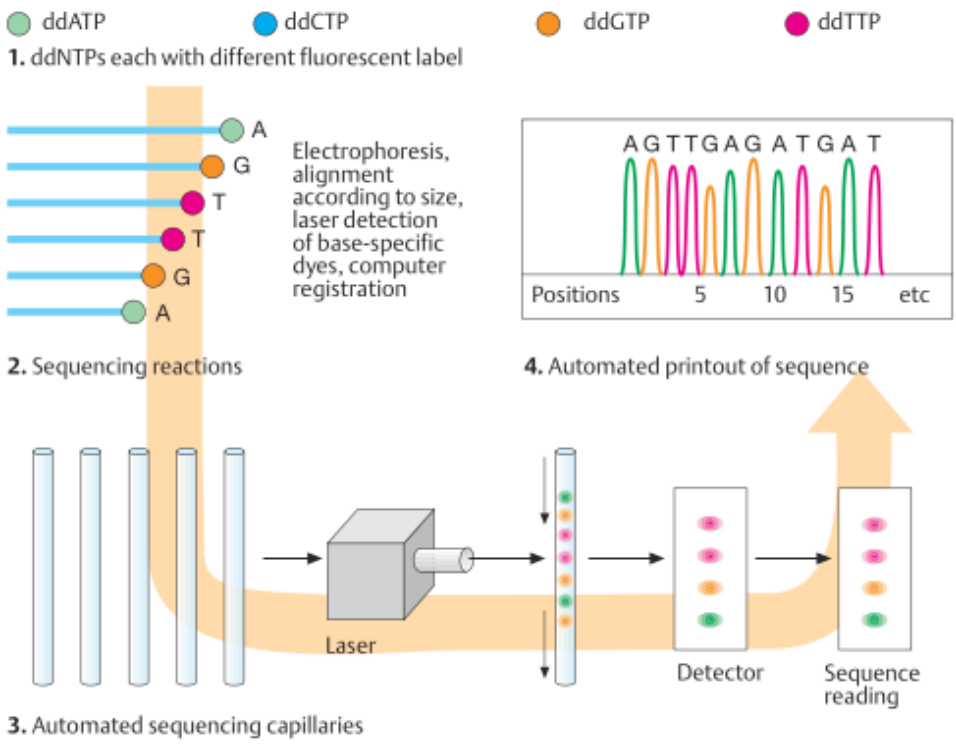
The DNA to be sequenced is contained in vector DNA <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. The primer, a short oligonucleotide with a sequence complementary to the site of attachment on the single-stranded DNA, is used as a starting point. For sequencing short stretches of 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</fn> 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