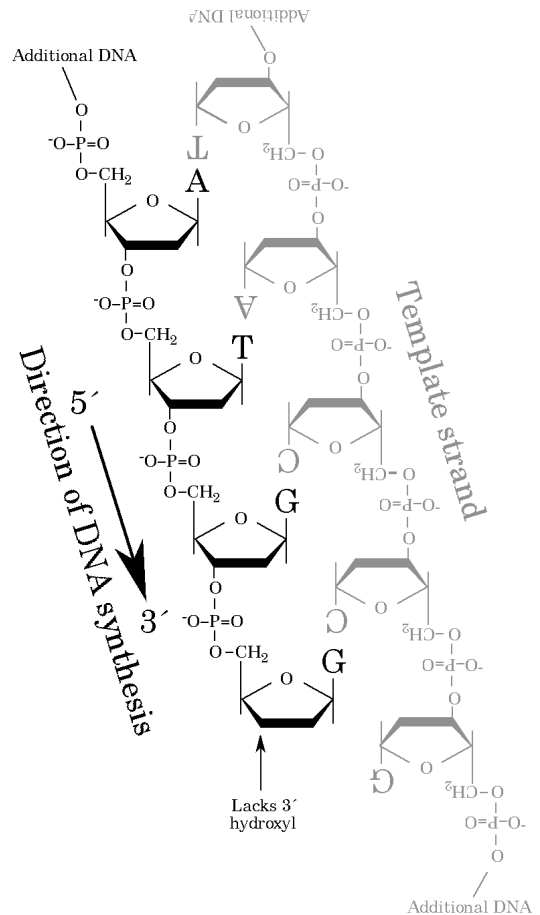


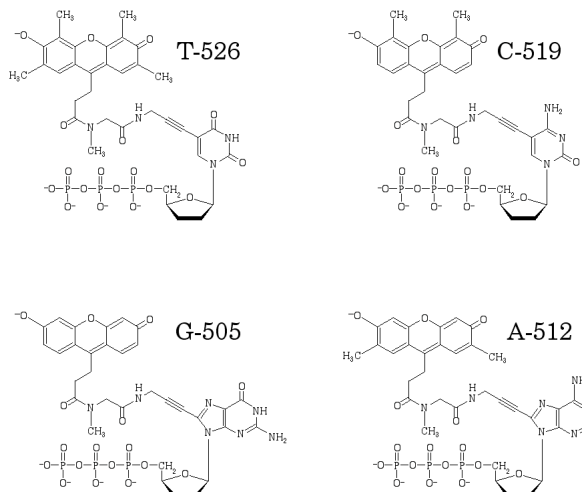
Sequencing DNA

DNA sequencing was invented in the late 1970s. It is currently far easier to sequence DNA than to sequence either RNA or proteins. DNA sequencing involves a modification of a normal cellular DNA replication process. As with most DNA synthetic reactions, sequencing reactions require an oligonucleotide to use as a primer, a template to use for synthesizing a complementary strand, and dNTPs to use as substrates for synthesis. However, the DNA sequencing reaction also requires a way of identifying each base added to the growing DNA strand.

The commonly used methods for DNA sequencing all involve the use of “chain terminators”, which are modified nucleotides that can be incorporated into the new DNA strand, but do not permit continued synthesis of DNA. The chain-terminators lack a 3'-hydroxyl group (in the example below, dideoxyGTP was inserted instead of dGTP, terminating DNA synthesis). If the sequencing reaction is set up such that the identity of the chain terminator nucleotide added is known, it is possible to identify the nucleotide at each position.



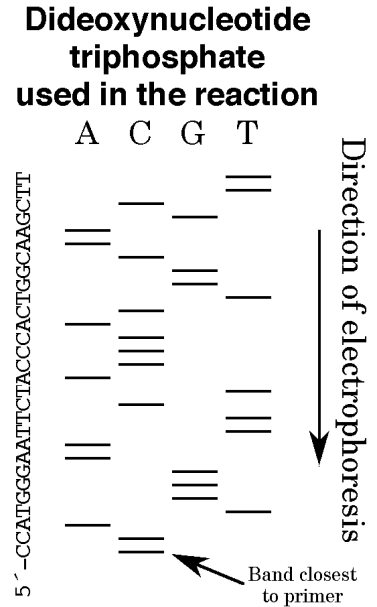
One method being used more and more frequently is the use of fluorescently labeled chain terminators. If each type (*i.e.* A, C, G, and T) of nucleotide has a different fluorescent label (such as those shown at right), the sequence can be determined very readily using an automated system. Variations on this method are being used in the genome sequencing projects currently in progress. This method is also used by sequencing services (you send your DNA sample to a lab, and they tell you what the sequence is without you having to actually do the work). The fluorescent-tag system, however, requires expensive equipment.



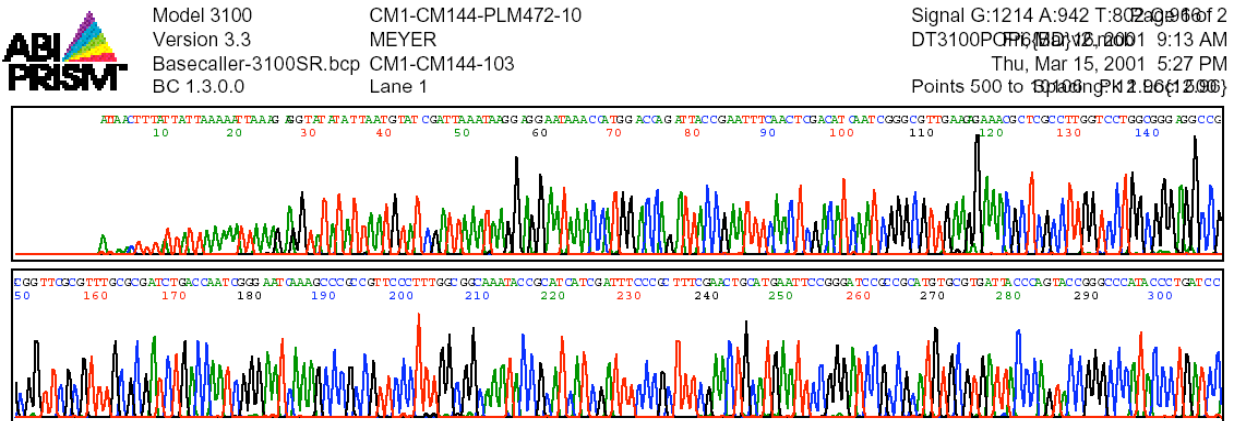
Laboratories that perform small scale sequencing usually use a somewhat older method in which four separate reactions are run, one for each of the possible chain

terminator nucleotides; in each case, the newly synthesized DNA is radioactively labeled so that it can be detected (usually by exposing the gel to film). After the reaction is complete, the DNA is run on a polyacrylamide gel; sequencing gels can separate DNA based on differences in size of a single nucleotide.

A representation of a sequencing gel autoradiograph (*i.e.* the developed film after an incubation of the gel with the film) is shown at right. Four reactions were run using the same template DNA, but with only one dideoxynucleotide. The gel is separating the DNA fragments based on their size; as usual, the smallest bands are at the bottom. The sequence can be read directly from the gel; simply start at the bottom, and for each band, note the lane in which the band occurred (the sequence derived from this “experiment” is shown on the left side of the figure).



The semi-automated sequencing machines use a generally similar method; the difference is that, instead of stopping the gel partway through the separation (as is shown here), the bands are detected (based on the fluorescent tags on the ddNTPs) as each one runs off the end of the gel. The result is a sequencing “electropherogram” (an example of which is shown below), which is a tracing that shows the elution of the different fragments corresponding to the sequence.



Depending on how well the reaction runs, on the homogeneity of the gel, on the quality of the template DNA, and on the exact method used, a single sequencing reaction can yield anywhere from 100 to 1000 bases of information. A cDNA of 1000 bp would usually require 2 to 4 reactions (using different primer positions) to read the entire sequence. The human genome (3×10^9 bp of unique sequence) would require at least 3×10^6 sequencing reactions; in practice, it required many more than that, in order to sequence each part of the genome more than once.

The “gel” shown above is an idealized example. Some sequencing gels actually look similar to this; most, however, have artifacts of various types, which make reading the sequence somewhat more challenging. Artifacts can be due to inhomogeneities

in the gel (if some lanes run faster than others, it can be difficult to read the sequence). If some copies of the DNA template have holes (or other problems that force the polymerase to stop synthesis), the gel may show bands in all of the lanes (making it difficult to decide which is the correct base). In addition, the sequence itself may cause problems: the GC base-pair binds more tightly than the AT base-pair; sequences with high GC content are especially subject to artifacts where the polymerase has trouble synthesizing the new strand, or where the DNA forms secondary structure while running on the gel which tends to alter its speed of migration through the gel. While experienced researchers can usually compensate for these problems, determining the sequence of unknown DNA samples can be quite difficult. Verifying a known or expected sequence (for example, when checking for the presence of point mutations following a site-directed mutagenesis experiment) is usually much less difficult, but can still be subject to errors.