

DNA Sequencing:

Sanger Method (Dideoxynucleotide chain termination)

Sanger sequencing is a DNA sequencing method in which target DNA is denatured and annealed to an oligonucleotide primer, which is then extended by DNA polymerase using a mixture of deoxynucleotide triphosphates (normal dNTPs) and chain-terminating dideoxynucleotide triphosphates (ddNTPs). ddNTPs lack the 3' OH group to which the next dNTP of the growing DNA chain is added. Without the 3' OH, no more nucleotides can be added, and DNA polymerase falls off. The resulting newly synthesized DNA chains will be a mixture of lengths, depending on how long the chain was when a ddNTP was randomly incorporated.

Manual DNA sequencing example:

- First, anneal the primer to the DNA template (must be single stranded):

5' -GAATGTCCTTTCTCTAAG
3' -GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCAAG-5'

- Then split the sample into four aliquots including the following nucleotides:

"G" tube: **All four** dNTPs, one of which is radiolabeled, **plus ddGTP** (low concentration)

"A" tube: All four dNTPs, one of which is radiolabeled, **plus ddATP**

"T" tube: All four dNTPs, one of which is radiolabeled, **plus ddTTP**

"C" tube: All four dNTPs, one of which is radiolabeled, **plus ddCTP**

- When a DNA polymerase (e.g. Klenow fragment) is added to the tubes, the synthetic reaction proceeds until, by chance, a dideoxynucleotide is incorporated instead of a deoxynucleotide. This is a "**chain termination**" event, because there is a 3' H instead of a 3' OH group. Since the synthesized DNA is labeled (classically with ³⁵S-dATP), the *products* can be detected and *distinguished* from the template.

Note that the higher the concentration of the ddNTP in the reaction, the shorter the products will be, hence, you will get sequence CLOSER to your primer. With lower concentrations of ddNTP, chain termination will be less likely, and you will get longer products (sequence further AWAY from the primer).

If, for example, we were to look only at the "G" reaction, there would be a mixture of the following products of synthesis:

5' -GAATGTCCTTTCTCTAAGTCCTAA**G**
3' -GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCAAG-5'

5' -GAATGTCCTTTCTCTAAGTCCTAAGTCCTCC**G**
3' -GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCAAG-5'

5' -GAATGTCCTTTCTCTAAGTCCTAAGTCCTCCG**G**
3' -GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCAAG-5'

5' -GAATGTCCTTTCTCTAAGTCCTAAGTCCTCCGGAT**G**
3' -GGAGACTTACAGGAAAGAGATTCAGGATTCAGGAGGCCTACCATGAAGATCAAG-5'

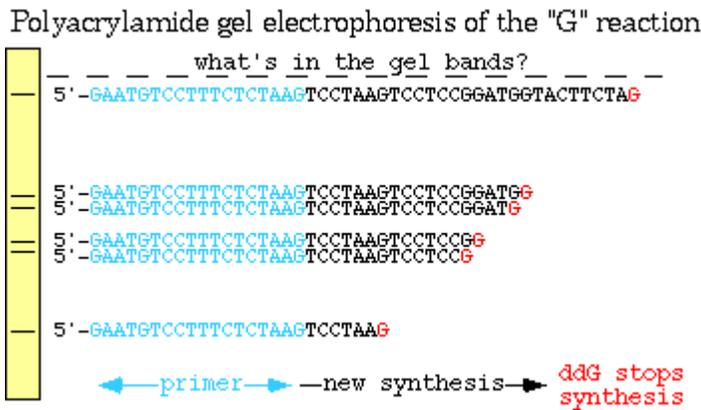
5' -GAATGCCTTTCTCTAAGTCCTAAGTCCTCCGGATG**G**
 3' -GGAGACTTACAGGAAAGAGATTTCAGGATTCAGGAGGCCTACCATGAAGATCAAG-5'

5' -GAATGCCTTTCTCTAAGTCCTAAGTCCTCCGGATGGTACTTCTA**G**
 3' -GGAGACTTACAGGAAAGAGATTTCAGGATTCAGGAGGCCTACCATGAAGATCAAG-5'

(and so on, if the DNA being sequenced continues to the right)

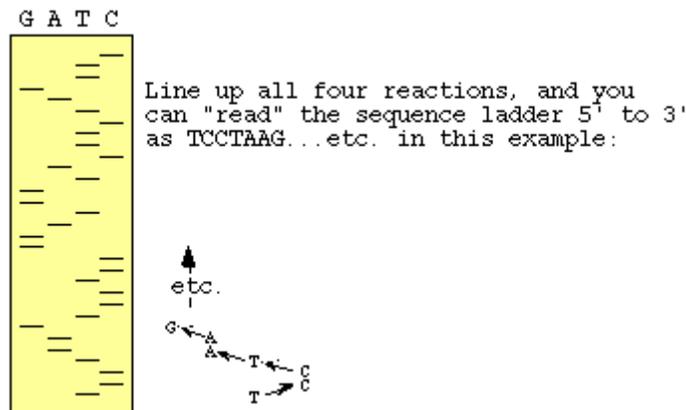
Each newly synthesized strand at some point had a ddGTP incorporated instead of dGTP. Chain termination then occurred (no more polymerization). Because ddGTP incorporation is random, all possible lengths of DNA that *end in G* are produced.

These products are denatured into single stranded DNA molecules and run on a polyacrylamide/urea gel. (Polyacrylamide gels, unlike agarose, allow resolution of DNA molecules that differ in size by only one nucleotide.) The gel is dried onto chromatography paper and exposed to X-ray film. Since the template strand is not radioactively labeled, it does not generate a band on the X-ray film. Only the labeled top strands generate bands, which would look like this:



As you can see from this one reaction (the "G" reaction) the chain termination events produce individual bands on a gel. The chain terminations closest to the primer generate the smallest DNA molecules (which migrate further down the gel), and chain terminations further from the primer generate larger DNA molecules (which are slower on the gel and therefore remain nearer to the top).

When similar chain termination reactions are run for each nucleotide, the four reactions can be run next to each other, and the sequence of the DNA can be read off of the "ladder" of bands, 5' to 3' sequence being read from bottom to top:



The resolution of the gel electrophoresis is very important in DNA sequencing. Molecules that are 50, 100, or 200 bases in length must be separable from molecules that are 51, 101, or 201 bases in length (respectively). To accomplish this:

- Polyacrylamide, not agarose, is used
- The gels must be quite large so that the molecules migrate further and are better resolved.
- Samples are denatured before they are loaded, and the gels must contain a high concentration of urea (7 to 8 molar) to prevent folding of the molecules and formation of secondary structures by hydrogen bonding that would alter the mobility of the molecule.
- The gels are run at higher temperature (about 50 C), also to prevent H bond formation.



{Note that this example is for demonstration only: you can't really obtain usable sequence information that close to the end of the primer because few termination events will have occurred so soon. If you increased the ddNTP:dNTP ratio, you would get more sequence close to the primer, but make it more difficult to read sequence 200 to 300 nucleotides further down, because most of the synthetic products would have terminated earlier.}

Sequence on gel at left:

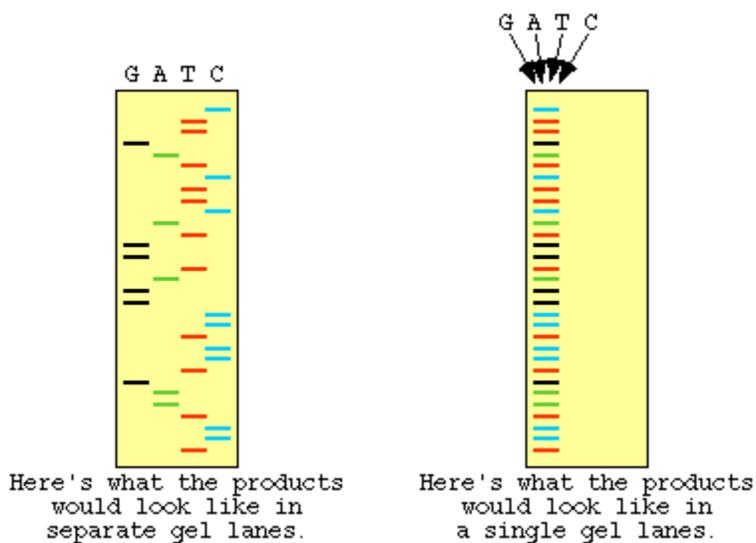
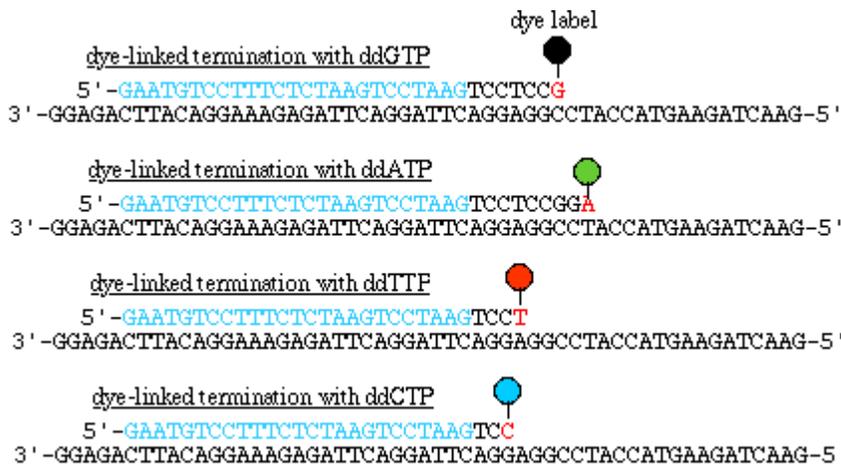
{TG}TACAACCTTTACTATGGCGTGACACCTAAATTATAGGCAGAAA...

Automated DNA sequencing:

Dye termination sequencing

Most DNA sequencing is now automated. The Sanger method chain termination reactions are still used, but pouring, running, & reading polyacrylamide gels has been replaced by automated methods. Instead of labeling the products of all 4 sequencing reactions the same (with a radioactive deoxynucleotide), each dideoxynucleotide is labeled with a *different fluorescent* marker. When excited with a laser, the 4 different kinds of products are detected and the fluorescence intensity translated into a data “peak”.

Thus all four chain termination reactions can be performed in the same tube, and run on a single lane on a gel. A machine scans the lane with a laser. The wavelength of fluorescence from the label conjugated to the ddNTPs can be interpreted by the machine as an indication of which reaction (ddG, ddA, ddT, or ddC) a particular DNA band came from.



The fluorescence output is stored in the form of a **chromatogram**:

