Medical Genetics: Past, Present, Future, pages 17–21

THE GENOMIC SEQUENCING TECHNIQUE

George M. Church and Walter Gilbert

Biogen
Cambridge, Massachusetts 02142

In a mammalian cell, the DNA corresponding to any gene sequence is surrounded by DNA corresponding to some million other such sequences. How can we study a specific gene without isolating it away from all others? We have devised a technique that will display the sequence of any gene without previous purification (Church and Gilbert 1983).

Our method extends the Southern blotting technique to a new degree of sensitivity. We use a radioactive probe not just to identify a single restriction fragment of DNA but as a way of end-labeling a unique restriction fragment. This end-labeling by hybridization is analogous to the end-labeling that we do in the conventional DNA sequencing, which enables us to display in any partial digest a series of fragments extending out from the point of labeling as a series of labeled bands separated by size. Although the chemical DNA sequencing techniques usually use only a single labeled phosphate, if the hybridization probe is kept moderately short, on the order of 100 nucleotides, it will serve as a end-label that can display sequence running out several hundred bases beyond the end of the probe.

We take total DNA and cut it with a restriction enzyme, one of whose cuts is close to the sequence of interest. After we perform the chemical sequencing reactions on that total DNA, we electrophorese the denatured DNA through a conventional sequencing gel. We load from 20 to 30 γ of DNA in each lane of the gel.

Then, by electrophoresis perpendicular to the gel plane in low ionic strength buffer, we transfer the DNA out of the gel onto a nylon membrane. Irradiation with ultraviolet light covalently immobilizes the DNA to the membrane. We hybridize the membrane with a single-stranded probe corresponding to a sequence that runs from the restriction cut of interest out about 100 bases into the fragment that we wish to explore. We use two types of probes. For single-stranded DNA probes, the probe is synthesized by sub-cloning an appropriate fragment into phage M13, isolating single-stranded virus, and then using a synthetic primer to prime the synthesis of 100 nucleotides of DNA across part of the insert. The probe is sized and separated from the viral DNA by a brief electrophoresis, isolated from that gel, and used immediately. In these probes, about 100 to 200 bases in length, about one out of 8 phosphates is radioactive, about 10^{10} DPM/γ. Alternatively, we synthesize an RNA probe by sub-cloning into an SP6 vector and synthesizing an RNA transcript with the SP6 RNA polymerase. With an RNA probe, we find that treatment with RNAse is essential to lower the background. We do the hybridization in a special buffer that contains 1% bovine serum albumin, 7% sodium dodecyl sulphate, and 0.5 M sodium phosphate buffer; the wash solutions also contain sodium dodecyl sulphate. The autoradiography takes about two to three days with an intensifying screen.

We need about 1000 times the sensitivity of the usual Southern blot, since the single DNA restriction fragment that one normally visualizes will be broken up by the sequencing reactions into some several hundred fragments. Because we generally cleave at less than 1 hit/500 nucleotides, in such a way as to leave large amounts of the original material unreacted, only a few tenths of a percent of the original restriction fragment will be found in each band. Thus we detect in each band about 2 to 4 femtograms of DNA. The technique will display the sequence extending from the probe out some 300 or 400 bases, depending on exactly how the sequencing gel is run. Since we can determine the sequence of single copy genes in mammalian DNA, of course we can detect sequence from the genome of any lower organism. For more complicated genomes, a preliminary fractionation of the DNA, by running a one-dimensional gel on the restriction enzyme digest and isolating the size fraction of interest, to enrich the DNA a further fraction of 10 will permit one to visualize the DNA even from organisms of great complexity.

The ability to sequence genomic DNA directly has certain uses in itself, although we need to have a cloned fragment or some knowledge of the sequence near the region of interest. Genomic sequencing can then be used to verify that a cloned sequence does correspond to the sequence in the genome of the organism, to detect mutations within cell lines and tissues, and to detect polymorphisms in a population, with ease if they are homozygous, with some difficulty if they are heterozygous. The techniques of probe synthesis, DNA immobilization, and the special washing yield an extremely sensitive Southern blotting method that requires very little DNA.

Another application of this technique rests on the fact that methylated cytosines in DNA do not react with the chemical cleavage reagents. Thus in the C reaction with hydrazine, if the C has been methylated, no corresponding band will appear in the sequencing pattern. This enables us to study the state of methylation of mammalian DNA directly for any gene of interest in any tissue of interest. We are no longer restricted to just those methylations that can be visualized with restriction enzymes.

Most importantly, genomic sequencing gives us the ability to visualize proteins that bind along the genome of mammalian cells. If we create the partial digest not by sequencing chemistry but by using some cleavage agent whose effect on the DNA is modified by the presence of proteins bound to the DNA, we can determine the positions of those proteins. Among such agents are dimethyl sulphate, bleomycin, DNAse, and light, all of whose effects can, either immediately or finally, be interpreted in terms of a breakage pattern in DNA. The frequency of these breakages at individual bases, and thus the intensity of the corresponding bands, can either be suppressed or enhanced by proteins bound to DNA. For example, dimethyl sulphate reacts with the N7 positions of the guanines of DNA, in the major groove. In in vitro experiments, the binding of proteins to DNA sequences can block the attack of dimethyl sulphate on the N7 position, presumably by steric hindrance. One also observes an increased rate of attack, usually at the edge of the protein binding site, which has been interpreted as a hydrophobic pocket enhancing the reaction. Thus we can "footprint" the presence of proteins on DNA in mammalian nuclei and even in whole cells.

With Harry Nick, we used this technique with bacterial cells to visualize a gene on the bacterial chromosome, to examine the lac repressor sitting on the lac operator. We treat the growing bacterial cells with dimethyl sulphate and can see that the lac repressor in vivo makes the same contacts in DNA that were long ago detected in vitro. We can also detect the contacts of the CAP protein to DNA near the lac promoter, binding to the DNA in vivo as it is activated with cyclic AMP. Such experiments do not strain this method at all, but they illustrate that we can use any agent that penetrates cells and, by doing the end-labeling restriction analysis after the DNA has been modified, we can detect the placement of protein within living cells.

With Anne Ephrussi and Susumu Tonegawa, we have been studying the enhancer of the immunoglobulin genes. This enhancer is a DNA sequence that activates the immunoglobulin genes specifically in those cells in which the immunoglobulin molecule is expressed. A characteristic feature of enhancers is that they activate neighboring promoters, a few thousand base pairs either downstream or upstream from the position of the enhancer element, in a tissue-specific manner. For the immunoglobulin genes, the enhancer sequence lies within the long intron that separates the assembled variable region from the constant region. It has been characterized by transformation and deletion mapping experiments that localize it to a few hundred base pairs (Gillies et al. 1983; Banerji et al. 1983). We have studied the enhancer region from the mouse immunoglobulin heavy chain by using dimethyl sulphate either on intact nuclei or on whole cells growing in culture. In examining the guanines in the enhancer region we can see that some are protected against the attack of dimethyl sulphate, while at others the attack of dimethyl sulphate is enhanced. These changes occur only in cells that express immunoglobulins, while cells from other tissues in which the gene is resting do not show these changes. These changes characterize two regions of sequence in the enhancer; these regions are homologous and represent an inverted sequence about 130 bp apart. Thus we infer that we detect the binding of protein to these regions in a tissue-specific manner and hypothesize that this protein is in fact responsible for the increased activity of the immunoglobulin promoter. We detect these contacts in whole cells in vivo. In nuclei, these contacts are visualized in low salt but weaken and disappear as the salt concentration goes up. Homologs of these two sequences can be found in other immuno-globulin genes, for the the mouse kappa gene and the human heavy chain gene. We believe we have detected an important binding site of protein in the immunoglobulin enhancer.

These techniques will have a general use to study the specific regulatory proteins involved in the control of mammalian genes. By being able to detect tissue-specific effects on DNA, and hence the presence of tissue-specific factors, we can predict which regions of the DNA are critical for control. Furthermore, this same detection of protein DNA contacts can be used in vitro to follow the purification of gene-specific factors.

REFERENCES

Banerji et al. (1983) A lymphocyte-specific cellular enhancer is located down stream of the joining region in immunoglobulin heavy chain genes. Cell 33:729.

Church GM, Gilbert, W. (1984). Genomic Sequencing. Proc Natl Acad Sci USA 81:1991.

Gillies et al. (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33:717.