The Journal of Laryngology and Otology
March 1998, Vol. **112,** pp. 230–234

Basic techniques in molecular genetics

ANGELA BROOMFIELD, DAVID BOURN

This article covers some of the more important techniques in molecular genetics (particularly in diagnostic work). The aim is to give an idea of the principles underlying these techniques rather than recipes for use in the laboratory. Further descriptions of these methods can be found in the texts suggested below.

Basic techniques in DNA analysis

Electrophoresis of DNA fragments

Linear DNA carries a uniform negative charge along the length of a fragment owing to the negatively-charged phosphate groups along the backbone of the double helix. Therefore, in a suitable gel matrix, DNA fragments can be separated by size by subjecting the gel to an electric current. This process is called electrophoresis. The DNA fragments will migrate towards the positive electrode, and the extent of migration is determined by the length of the DNA. The migration of large fragments is retarded by the gel matrix more than that of small fragments, ie.the smaller the fragment of DNA, the faster it will move in the gel. Two gel types are commonly used, agarose and polyacrylamide. The choice of matrix depends on the size of the fragments to be separated and on the desired degree of resolution, ie. the capability to separate fragments of a similar size.

Agarose gel electrophoresis

Agarose is extracted from seaweed and produces a jelly-like material. The agarose is melted and poured into a mould of the appropriate dimensions and a comb used to provide wells for the DNA samples. The gel is run horizontally immersed in a suitable buffer. Agarose gels can be used in the separation of DNA fragments ranging in size from less than 100 base pairs (bp) to over 20 kilobases (kb). More specialised methods are required to separate larger DNA fragments as electrophoretic mobility is no longer proportional to fragment size beyond a certain length (approximately 30–50 kb). Electrophoresis of very large fragments is possible by pulsed field gel electrophoresis (PFGE). The basis of resolution is the reorientation of large fragments as the direction of electric field changes. The field direction changes in 'pulses' and the larger the fragment, the longer it takes to re-orientate.

Polyacrylamide gel electrophoresis

Polyacrylamide gels are usually used to separate small DNA fragments, as the pore size in the gel matrix is smaller than that achieveable with agarose gels. A chemical reaction is used to produce cross-linked acrylamide polymers which can separate fragments of several hundred bp to a resolution of one base pair. Polyacrylamide gels are poured between glass plates separated by spacers (usually of 1 mm thickness or less) and run vertically with buffer tanks at top and bottom allowing the flow of current. Again a comb is used to form wells in which to load the DNA samples.

Visualization of DNA fragments

DNA separated in a gel can be visualized under ultraviolet light (UV), by prior staining with ethidium bromide. Ethidium bromide is a dye which inserts itself into the DNA double helix and is used to detect DNA fragments, particularly in agarose gels. Visualization of DNA fragments requires a UV light source, which causes ethidium bromide to fluoresce orange.

Silver staining can be used to detect smaller amounts of DNA than those seen with ethidium bromide. Positively charged silver ions associate with the negatively charged DNA and with appropriate treatment are deposited as silver. This technique is used in staining polyacrylamide gels. Labelling of DNA fragments can also be accomplished by enzymatically incorporating either radioactive or fluorescent nucleotides. Radioactive labelling (usually with ^{35}S or ^{32}P) followed by exposure to X-ray film allows the detection of very small amounts of DNA. Increasingly, fluorescent labelling followed by electrophoresis and detection on a semi-automated analyser is the method of choice for rapid DNA analysis.

Fragmentation of DNA

Human DNA is most conveniently obtained from lymphocytes. Despite some inevitable shearing of

From the Regional Genetics Service, Birmingham Women's Hospital, Edgbaston, Birmingham B15 2TG.

the DNA during the extraction process due to the fragility of chromosomal DNA, the fragments produced are still too long to be of manageable length for analysis. Smaller DNA fragments can be produced by sonication or other physical disruption. However, the size and number of fragments produced by these methods will be random. A more specific method uses restriction endonucleases (or restriction enzymes). These enzymes cut DNA at defined sequences, and the specific restriction enzyme used determines at which particular sites and how frequently the DNA is cut. The type of restriction enzyme usually used in molecular analysis has the property of cutting at a short palindromic sequence, ie. one that reads the same 5′ to 3′ on either DNA strand. For example the enzyme *Eco* RI (*Eco* because the enzyme was isolated from the bacterium *E. coli*) cuts at the sequence.

5′GAATTC3′
3′CTTAAG5′

The enzyme cuts both DNA strands within the target site, generating either 'blunt' ends if the cut is at the same position on each strand or 'sticky' ends if the two strands are cut at different positions within the recognition site. Usually these enzymes have recognition sites of 4–6 bp, and hence by chance occur rather often in a haploid human genome of three thousand million bases. DNA digested by such enzymes is cut into many thousands of fragments, and since the enzymes cut at specific sites, in most cases the same enzyme will generate identical fragment sizes from DNA from different individuals.

DNA sequencing

One of the most important advances in genetics in recent years has been the development of methods for determining the nucleotide sequence of DNA fragments. Most sequencing protocols in current use are based on the dideoxy chain termination procedure. The DNA of interest is required as a single stranded template. This can be achieved in a number of ways including cloning the DNA into a specialized vector (see cloning section below) or increasingly through the direct use of polymerase chain reaction (PCR) products (see PCR section below). DNA synthesis for sequencing using the single-stranded template DNA is performed in the presence of normal nucleotides (dNTPs) and dideoxynucleotides (ddNTPs). The ddNTPs are analogues of normal dNTPs, but they lack a hydroxyl group at the 3′ carbon position. They can be incorporated into the growing DNA chain, but are not suitable templates for further phosphodiester bond formation, resulting in termination of synthesis. In a reaction containing ddATP, all of the newly synthesized DNA strands will terminate at an A, in a reaction containing ddTTP at a T and so on. The proportion of ddNTPs to dNTPs is controlled, so that incorporation of a ddNTP occurs at random positions, so in effect a ladder of differently sized DNA fragments varying by one base pair is generated, each fragment having a common 5′ end. The fragments are separated using polyacrylamide gel electrophoresis, and detected through the incorporation of radioactive or fluorescent nucleotides into the newly-synthesized DNA strand (Figure 1). Either four separate reactions are performed, one for each base, or the dideoxynucleotides are labelled with four different fluorescent colours.

Analysis of single genes

The major problem for gene analysis either in a research or diagnostic context is the difficulty of studying one specific region of the genome against the background of other genes and the non-coding majority of the genome. Three possible approaches to this problem are DNA cloning, DNA hybridization and PCR.

DNA cloning

Cloning as a general term is the process of obtaining exact replicas from a single cell or organism. In molecular biology terms it refers to the propagation, usually in a micro-organism, of a particular fragment of inserted foreign DNA. The DNA of interest is attached to a vector, a small DNA molecule capable of propagation in the host micro-organism (typically a bacterium or a yeast). Both the vectors used and the host species have been modified enormously to make cloning technically easier. One strategy is to insert foreign DNA fragments into bacterial plasmid DNA. Plasmids are small, circular molecules of double stranded DNA. They are

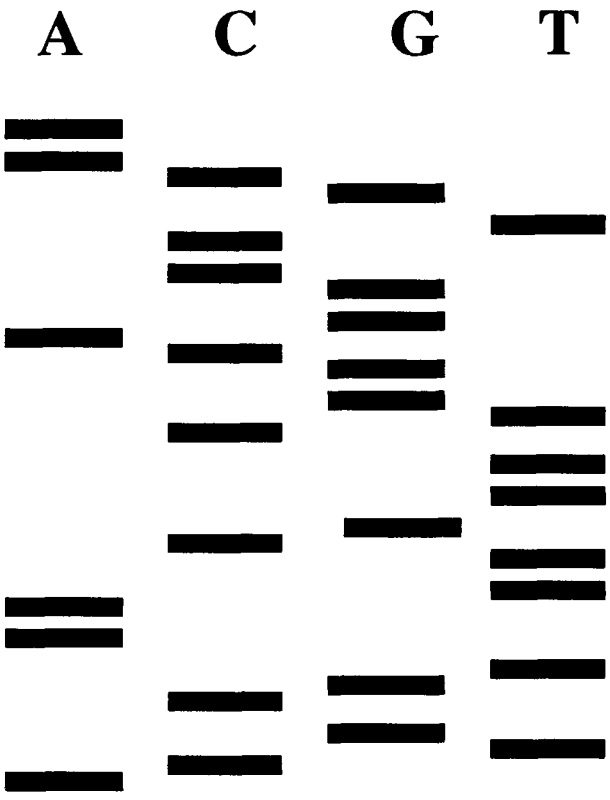


FIG. 1

Schematic representation of a sequencing gel. This schematic example illustrates the sequence, read from bottom to top: 5′-ACTGCGTAATTCGTTCTGGCAGGCCTGCAA-3′.

https://doi.org/10.1017/S0022215100158256 Published online by Cambridge University Press

separate from the bacterial chromosome and occur naturally in some strains of bacteria. The DNA for cloning and the plasmid can be cut by the same restriction enzyme, generating complementary single-stranded ends. These complementary ends allow the foreign DNA fragments to be joined to the plasmid DNA, the breaks being re-sealed by the enzyme DNA ligase.

If an entire genome is cut into fragments and joined in this way to a vector a genomic library is generated which hopefully contains a representation of all sequences in the genome, including the gene of interest. Many ingenious strategies have been devised to select for the actual DNA of interest against the background of a library, which is the crucial step in cloning, but these are outside the scope of this article (see Old and Primrose (1994) for some examples). Once a bacterium has been selected which contains the gene of interest it can be propagated indefinitely and the foreign DNA can be isolated for study. Cloning is a laborious, time-consuming and uncertain process and as far as routine diagnostic work is concerned it is little used once a human disease gene has been isolated.

DNA hybridization

Hybridization is the formation of a double-stranded DNA molecule between one single-stranded fragment and another. Techniques using DNA hybridization are possible because of the complementary nature of double-stranded DNA. Any single-stranded DNA fragment can (under the correct conditions) hybridize to denatured DNA provided they represent opposite strands of the same sequence, ie. they are complementary. A suitable 'probe' (usually a single-stranded DNA fragment of a few hundred bases in length) can be labelled radioactively or fluorescently and used to specifically hybridize to a complementary sequence even against the background of the rest of a complex genome. Two applications of this type of technique are briefly discussed below.

Fluorescence in situ *hybridization (FISH)*

FISH is performed using short single-stranded probes that have been fluorescently labelled. These are hybridized directly to denatured chromosome preparations that have been fixed on slides. The probe will hybridize to the region of the chromosome where that particular DNA fragment has its complementary sequence, and this is visualized using a microscope able to detect fluorescence. This immensely powerful technique can therefore be used to rapidly map newly cloned genes to their respective chromosomes, or to detect clinically significant chromosome rearrangements at a level of resolution far better than that possible with conventional cytogenetics. FISH is becoming increasingly important both as a research tool and in diagnostic cytogenetics.

Southern blotting

Southern blotting was developed (by Professor E. Southern) to allow the convenient hybridization of radioactively-labelled DNA probes to genomic DNA cut into fragments by restriction enzymes. The digested DNA sample is electrophoresed on an agarose gel to separate by size all of the many thousands of fragments generated by cutting a large genome with a restriction enzyme. The separated nucleic acid molecules are denatured and the resulting single-stranded molecules are transferred from the gel to a nylon membrane, which is more robust and therefore easier to handle in subsequent procedures.

The transfer is accomplished by placing the gel on a solid support over a reservoir of a salt solution. A wick is placed into this solution and the gel is placed onto the wick. The nylon membrane is positioned directly on top of the gel, and absorbent material, placed on top of the membrane, used to draw the salt solution through the gel and the membrane by capillary action. The DNA fragments are drawn out of the gel onto the membrane and fixed there in the same relative positions as they were on the gel. After Southern blotting, a radioactively-labelled probe with a complementary sequence to the region of interest is hybridized to the nylon membrane. The radioactive band can be revealed by exposing the membrane to X-ray film, producing an autoradiograph. The band or bands detected represent the sequence complementary to the probe against the background of all the rest of the genome, and is specific to the gene of interest.

This technique has many potential applications. One of these is the detection of restriction fragment length polymorphisms (RFLPs). Complete digestion of total genomic DNA with a restriction enzyme generates fragments as a result of cutting wherever the recognition sequence occurs for that particular enzyme. If every individual had a completely identical DNA sequence, total DNA would always be cut by a particular restriction enzyme into exactly the same number and size of fragments. However, because each individual has a slightly varying DNA sequence, this is not the case. The DNA coding for polypeptides accounts for only about three per cent of total DNA, and it is largely these regions which are strongly conserved by natural selection during evolution. Most of the remainder of the genome can accumulate harmless changes in nucleotide sequences during evolution. If one of these changes either creates or destroys a restriction enzyme recognition site or changes the distance between two sites, the same DNA probe will detect different sizes of fragments in different individuals, and this is termed a restriction fragment length polymorphism. RFLPs have been used in mapping human genes, or in tracking genes within a family, for example, by seeing whether a disease gene segregates with a particular RFLP in affected families. (If a gene is located physically close to a RFLP, even though there is no functional consequence of the RFLP, it will be inherited along with the gene, and hence can

https://doi.org/10.1017/S0022215100158256 Published online by Cambridge University Press

be used to track the gene within a family. The concept of genetic linkage and gene tracking will be described in detail in the next article in this series.) A particularly useful version of this technique makes use of regions of the genome made up of short tandem repeats, termed variable number of tandem repeats or VNTRs, because these regions will almost always differ (and therefore be informative) between unrelated individuals. In a few cases, the mutations that cause genetic disease will either create or destroy a restriction enzyme recognition site and so can be detected directly by RFLP analysis, eg. sickle-cell anaemia, where the same mutation occurs in all affected individuals, and abolishes a recognition site for the enzyme *Mst* II.

It should be noted that similar transfer methods for transferring protein and RNA from gels onto membranes are known as Western blotting and Northern blotting respectively.

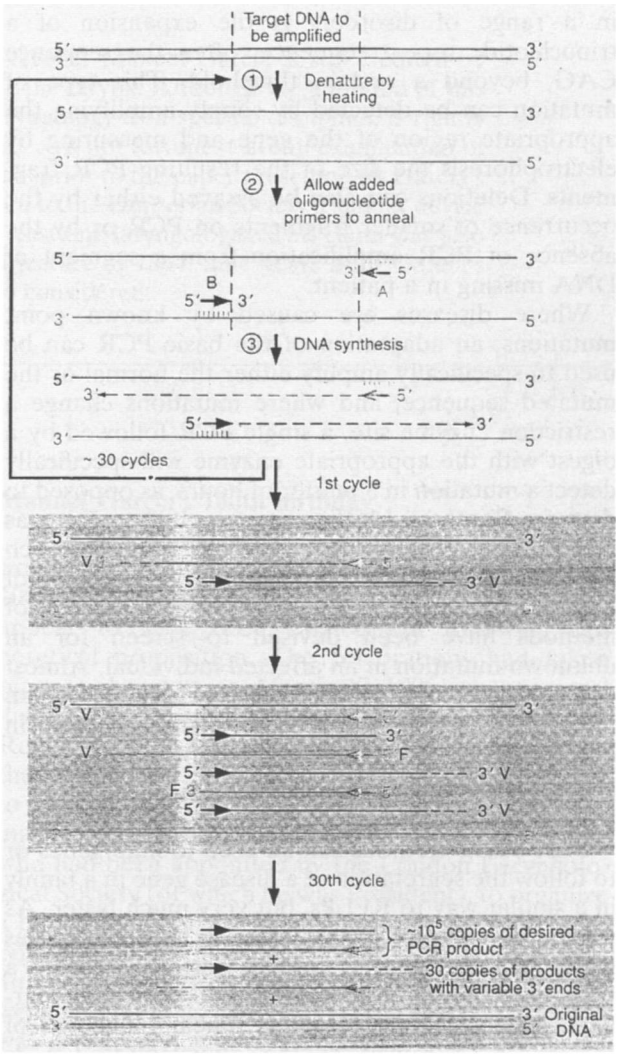
The polymerase chain reaction (PCR)

It is no exaggeration to say that the PCR has revolutionized a field which tends to be fast moving at the best of times. The power of PCR is that it allows analysis of single genes, or parts of genes, against the whole genetic background in a very short space of time. During a typical PCR which requires an average time of about three hours, a single copy of any gene under study can be uniquely amplified to millions of copies. It is now the case that diagnostic testing for most diseases involves the PCR.

PCR methodology

The basic technique relies on the use of specific oligonucleotide primers. These are short single-stranded DNA molecules, usually between 17 and 25 bases in length. The crucial requirement is that these primers are complementary to unique sequences on opposite strands flanking the region of interest (see Figure 2). The PCR reaction includes these two primers, the DNA template (for diagnostic tests, a small amount of genomic DNA from the patient), dNTPs (as new copies of the template DNA will be synthesized) and an enzyme, DNA polymerase. The reaction mix is subjected to repeated cycles of denaturation by heating to 95°C (to make the target DNA and any reaction products single-stranded), annealing (allowing the primers to hybridize to the target DNA) at a temperature which is balanced such that the short primers can hybridize, but only to the specific target sequence, and a final step at 72°C, the optimum temperature for the polymerase to synthesize new DNA.

The main problem with this method originally was that the DNA polymerase was inactivated after each reaction cycle, due to the temperature of the denaturing step. However this problem was overcome by the fortuitous discovery of a bacterium (*Thermus aquaticus*) found in the hot springs of Yellowstone National Park. DNA polymerase extracted from this bacterium, known as *Taq* polymerase, is naturally stable at high temperature and this therefore eliminated the need to add extra polymerase after each cycle. Cycling is achieved using programmable heating blocks so once a reaction is started no further hands-on time is required until the products are ready to be analysed by agarose or polyacrylamide gel electrophoresis. The principle behind the PCR is as outlined in Figure 2. The PCR primers anneal to the denatured target DNA and are used by the *Taq* polymerase as the start point for DNA synthesis. In the first cycle the polymerase will synthesize new strands of varying length before dissociating from the DNA. As shown in Figure 2, after subsequent rounds of replication, the newly synthesized DNA, with an end



Oligonucleotide primers A and B are complementary to DNA sequences located on opposite DNA strands and flanking the region to be amplified. Annealed primers are incorporated into the newly synthesized DNA strands. The first cycle will result in two new DNA strands whose 5′ end is fixed by the position of the oligonucleotide primer but whose 3′ end is variable ('ragged' 3′ ends). The two new strands can serve in turn as templates for synthesis of complementary strands of the desired length (the 5′ ends are defined by the primer and the 3′ ends are fixed because synthesis cannot proceed past the terminus of the opposing primer). After a few cycles, the desired fixed length product begins to predominate.

FIG. 2

PCR is an *in vitro* method for amplifying DNA sequences using defined oligonucleotide primers.

Figure and legend reproduced from Strachan and Read (1996) with kind permission of Bios Scientific Publishing.

https://doi.org/10.1017/S0022215100158256 Published online by Cambridge University Press

defined by the position of the primer, will become the majority target for further rounds of PCR, and exponential amplification of these specific defined fragments will result.

Applications of the PCR are numerous and form an integral part of many molecular biology protocols. An illustration of the clinical applicability of PCR based methods is given in the following discussion on mutation analysis and gene tracking.

Mutation detection in disease causing genes

Some genetic disorders may be caused by the same mutation in the same gene, or by a small number of different mutations. One class of mutation occurring in a range of disorders is the expansion of a trinucleotide repeat sequence, often the sequence CAG, beyond a certain threshold. This type of mutation can be detected by simply amplifying the appropriate region of the gene and measuring by electrophoresis the size of the resulting PCR fragments. Deletions can also be assayed either by the occurrence of smaller fragments on PCR or by the absence of PCR amplification from a segment of DNA missing in a patient.

Where diseases are caused by known point mutations, an adaptation of the basic PCR can be used to specifically amplify either the normal or the mutated sequence; and where mutations change a restriction enzyme site, a single PCR followed by a digest with the appropriate enzyme will specifically detect a mutation in a matter of hours, as opposed to days for Southern blotting or even longer if it was necessary to clone and sequence material from each patient. Some disorders are caused by a spectrum of different mutations, and in these cases a number of methods have been devised to screen for an unknown mutation in an affected individual. Almost without exception, however, these methods use an initial PCR step to amplify DNA from the gene in question from each individual patient.

Gene mapping/tracking

PCR-based methods can be used to map genes or to follow the segretation of a disease gene in a family in a similar way to RFLPs, but very much faster. As will be described in detail in the next article in this series, linkage analysis can be performed using a PCR-based method exploiting 'microsatellite' markers. Microsatellites are short repeat sequences of 1–5 bases (a repeat of the dinucleotide CA is frequently used) that occur with highly variable copy numbers of the repeat sequence (which is thought to have no functional significance) in different individuals. They are abundant, dispersed throughout the genome, are highly informative and easy to type usually using radioactively- or fluorescently-labelled PCR fragments.

Forensic applications

PCR has proved to be an invaluable tool in forensic science. The sensitivity of the reaction (material can be amplified from a single cell, and from degraded DNA of some age) makes it possible to amplify otherwise uselessly minute quantities of DNA. Samples collected at the scene of a crime such as blood, hair, semen or skin, can be analysed, and even if only tiny amounts of fragmented DNA can be recovered, it is still possible to carry out tests that can exclude or implicate a person in a crime. The technique of DNA fingerprinting, often based on Southern blotting and VNTR (see above) can be adapted to PCR-based methods, detecting highly variable, usually non-coding regions of DNA that are unique to each individual. This can be invaluable in criminal cases or in establishing the exact biological relationships between individuals.

Conclusions

Some of the more important techniques now routinely employed in research and diagnosis have been discussed. The development of molecular biology has often been described as 'explosive' and to put this in perspective, one might consider some of the historical landmarks. Frederick Sanger showed that a protein had a precisely defined amino acid sequence by determining it in insulin in 1953 for which he was awarded the Nobel prize. Nearly a quarter of a century later, in 1977, he described his method of DNA sequencing with chain-terminating inhibitor to win his second Nobel prize. It has been a decade since Mullis described the polymerase chain reaction and now its use can only be described as ubiquitous in research and diagnostic molecular biology. The methods described here have been used to clone many genes of clinical importance and their use in diagnosis is rapidly and continually expanding to assume an important role in clinical practice.

Bibliography

Elles, R. (Ed.) (1996) *Methods in Molecular Medicine: Molecular diagnosis of Genetic Disease* Humana Press Inc. Totowa, New Jersey.

Latchman, D. (Ed.) (1997) *Basic Molecular and Cell Biology*, 3rd Edition. British Medical Journal Publishing Group, London.

Old, R. W., Primrose, S. B. (1994) *Principles of Gene Manipulation*, 5th Edition. Blackwell Scientific Publications, Oxford.

Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) *Molecular Cloning- A Laboratory Manual*, 2nd Edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbour.

Strachan, T., Read, A. P. (1996) *Human Molecular Genetics*, 1st Edition. Bios Scientific Publishers, Oxford.