

## Basis of molecular biology and its applications: II. An outline of common laboratory techniques used in molecular biology

R.T. ERASMUS<sup>1,2</sup>, D.P. MURTHY<sup>1</sup> AND B.O. OGUNBANJO<sup>1</sup>

Department of Pathology, University of Papua New Guinea, Port Moresby

This article will briefly discuss the concepts of linkage analysis, restriction fragment length polymorphism (RFLP) and variable number of tandem repeats (VNTR) mapping. Blotting methods as well as nucleic acid hybridization will also be considered. Some other techniques have already been discussed in the first article of the series (1), for example, cloning.

*Linkage analysis* has played an important role in human genetic analysis. If two alleles present in an individual lie on different chromosomes, they will be inherited separately. However, if they are on the same chromosome, there is a high probability that they will be inherited together and cosegregate, though it is not a certainty since recombination during meiosis may cause them to become separated. The closer they are on the same chromosome, the greater the chance of them staying together at meiosis. Such genes will be tightly linked. Thus the study of mutant pedigrees allows the genes to be arranged in linkage groups and the genes within a linkage group to be placed in linear order.

Linkage analysis in humans is done by retrospective examination of pedigrees which is time-consuming, tedious and not always informative. No one instance of cosegregation proves linkage, since this will happen 50% of the time by chance in unlinked alleles. Conversely, one instance of failure to cosegregate does not prove nonlinkage, for it may always be due to recombination. Statistical analysis of the pedigree is usually carried out to calculate the probability that two alleles are linked. Sometimes a single pedigree may not contain enough information, and it may be necessary to combine data from several pedigrees.

The technique of *positional cloning* relies on the principle of linkage analysis to map a heritable allele using an array of DNA polymorphisms. These are specific sequence markers whose chromosomal positions have already been determined using restriction endonucleases, which have the ability to cut DNA molecules at precisely defined internal sites. Restriction endonucleases are obtained from various bacteria. Restriction enzymes bind to DNA, and cut both strands at specific sequences. The sites recognized by restriction enzymes are commonly 4 to 6 nucleotides long and always inverted repeat palindromes (1). Polymorphisms occur, as previously discussed (1), whenever a mutation alters a restriction site in the genome, or inserts or deletes some part of the sequence between two restriction sites. Most of them have been discovered by chance, especially during Southern blot analysis (described later) when bands of different sizes are picked up in DNA from different people following digestion with restriction enzymes. These *restriction fragment length polymorphisms (RFLPs)* are extremely useful in gene mapping since they are inherited according to mendelian laws of genetics. Thus a mutant gene can be mapped by analyzing DNA samples in a pedigree containing the mutation and looking for RFLPs that cosegregate with the mutation. The more tightly linked an RFLP is to the mutant phenotype, the closer on the chromosome it must be. In practice, a sufficiently large family needs to be studied, which may not always be possible, due to noncooperation or deceased members, using several restriction enzymes and probes. The detection of HLA class II polymorphism is valuable in the areas of individual identification, tissue typing for

---

<sup>1</sup> Department of Pathology, Faculty of Medicine, University of Papua New Guinea, PO Box 5623, Boroko, NCD 111, Papua New Guinea

<sup>2</sup> Present address: Department of Chemical Pathology, Faculty of Medicine, University of Transkei, Private Bag X1 UNITRA, Umtata 5100, Eastern Cape, Republic of South Africa

transplantation and establishing genetic susceptibility to specific autoimmune diseases. Polymorphism in the HLA class II region has been identified using serological reagents (HLA-DR and -DQ specificities), by cellular techniques (Dw and DPw specificities) and recently by RFLP. For HLA class II typing, RFLP analysis is based on the presence or absence of polymorphic restriction sites located primarily in the noncoding regions. In combination with the polymerase chain reaction (described later), it has proved to be an invaluable tool in the study of polymorphism and evolution as well as in the analysis of genetic susceptibility to disease. In Papua New Guinea, Serjeantson et al. (2) have used this technique to examine HLA-DRB1 nucleotide sequence polymorphisms in 304 Melanesians from Madang, Rabaul, Goroka and Fiji and found considerable heterogeneity in the distribution of HLA-DRB1 alleles, refuting earlier claims of HLA-DR allele restriction in Melanesians. The same group (3) also examined the gene frequency distributions of HLA-A, -B and -DR and found that, with the exception of a correlation between Reiter's syndrome and B27, other HLA and autoimmune disease associations are markedly absent in Melanesians.

Polymorphism has recently been demonstrated in the circumsporozoite (CS) protein of human malaria parasites. Qari et al. (4) reported that only about 75% of the CS protein gene sequences are similar between 1980 *Plasmodium falciparum* isolates collected from Brazil and those of recent field *P. falciparum* isolates. Using sequencing techniques, the same group of workers (5) have also identified a *Plasmodium vivax*-like human malaria parasite from Madang and the Sepik whose nucleotide sequence of the CS protein gene is similar to a monkey malaria parasite but different from the other known CS protein genes of human malaria parasites.

RFLP mapping has recently been supplemented by *variable number of tandem repeats* (VNTR) mapping. This technique depends on the extreme variability of highly repetitive DNA sequences within the genome. They are further subdivided into minisatellite and microsatellite DNAs, the former having a variable number of tandem repeats of no more than 20 kb long whilst the latter are brief lengths of tandem repeats of very short sequences (collectively termed as VNTR). The

human genome contains about 500 000 of these repeated sequences and differs very considerably between individuals in the number of repeats of the short-core sequences at specific sites. This has led to the concept of *DNA fingerprints*. Because of this polymorphism, these 'fingerprints' can be used to identify individuals almost unequivocally. This is utilized in paternity testing. Since hypervariable bands are inherited in a mendelian fashion, every band in the child's blot (Southern blot, which is described later) must have been contributed by either his mother or true biological father; if bands are seen that do not match those of the mother or alleged father, then the true biological father has not been found. This is relevant to Papua New Guinea, where the practice of adoption is common with natural assimilation into families. In earlier studies on the determination of the mode of inheritance of Melanesian (South-East Asian) ovalocytosis considerable confusion occurred because of this. Similar logic is used in matching a forensic specimen (usually dried blood, semen or hair roots) to the DNA fingerprints of a suspect in a murder or rape case. In the celebrated O.J. Simpson case, forensic evidence of this kind was used by the prosecutor's office. However, one has to be cautious in interpreting the results. Other problems in forensic analysis include band shifting during electrophoresis, the need for strict chain-of-custody documentation to avoid sample substitution, and the lack of knowledge of allele frequencies for fingerprint patterns in some racial and ethnic groups. Some DNA evidence has been thrown out of court because of these problems (6). VNTR probes can also be used to distinguish donor from recipient haematopoietic cells following bone marrow transplantation (7). A DNA match with the donor in the post-transplant sample indicates engraftment.

Since DNA fingerprints behave as a single heterozygous trait each fingerprint band is transmitted on average to half of the offspring, and unlinked bands assort independently. With a microsatellite that is very abundant and extremely polymorphic, it is thus possible to establish linkage between VNTR polymorphism and inherited diseases. VNTR markers are faster and generally more informative than RFLP markers. One disadvantage is their extreme variability. In Papua New Guinea considerable variation in the number and frequency of alleles at

particular loci has been observed using VNTR markers (8).

Since restriction endonucleases recognize only specific DNA sequences, a long DNA molecule will be cut only at the sites where those particular sequences occur and therefore a specific restriction enzyme will always digest the same DNA molecules into the same number of fragments. Such digestion creates, from any given DNA, a particular mixture of double-stranded fragments. Such existing restriction recognition sites may be abolished by single base pair changes, causing an alteration in the length of these fragments which leads to, as mentioned earlier, restriction fragment length polymorphism (RFLP). Blotting techniques are commonly used in molecular biology to identify nucleic acid fragments or proteins.

The blotting methods usually consist of 4 separate steps. The protein or nucleic acid fragments in the sample are first separated by electrophoresis, then transferred on to an easily handled support medium or membrane through a blotting mechanism involving filter paper. Support media commonly used include

nitrocellulose paper, charge-modified nylon sheets or glass fibres. The immobilization of protein or DNA fragments on to any of these membranes facilitates subsequent biochemical analysis. After the DNA fragment or protein is transferred on to the support medium, it is localized by the use of labelled DNA probes or antibodies that specifically bind to the molecule of interest (hybridization). The position of the probe bound to the immobilized target molecule is visualized by autoradiography (in the case of radiolabelled probes).

Three main blotting techniques are used and are called Southern, Northern and Western blotting. Southern blotting (Figure 1), named after its originator, enables the DNA fragments to be separated and identified by means of DNA probes which hybridize to complementary fragments of chromosomal DNA. Before Southern or Northern blotting can be carried out, it is mandatory that cloned DNA or synthetic DNA sequences that are used as probes for the gene are available. Northern blotting is similar to Southern blotting except that fragments of mRNA instead of DNA are separated by

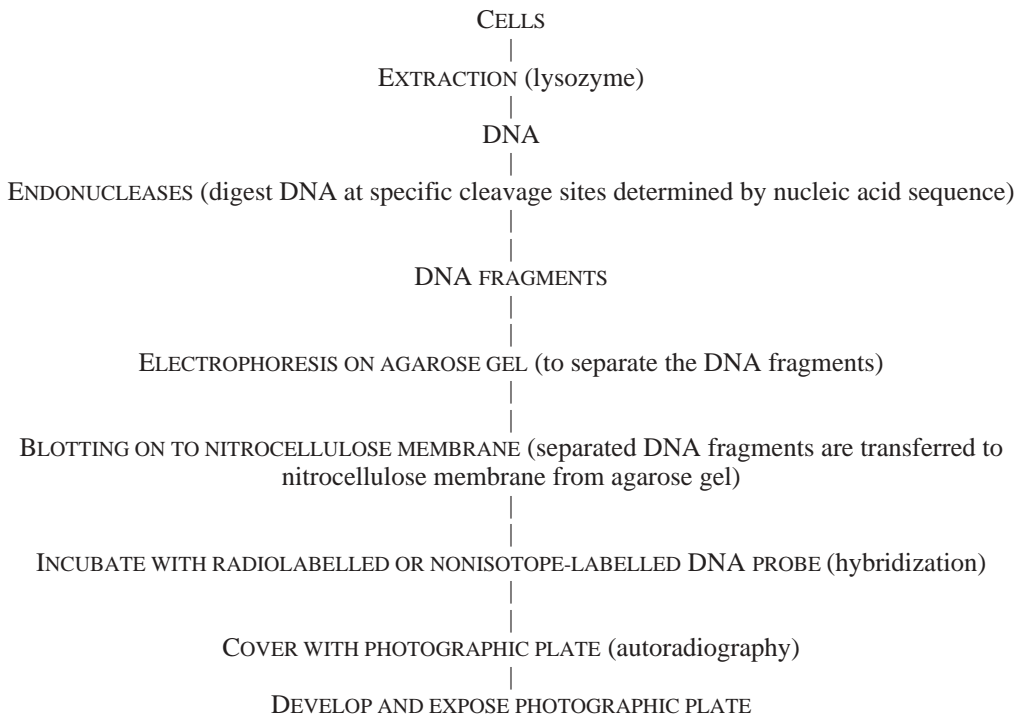


Figure 1. The sequence of steps in Southern blotting.

electrophoresis. After blotting they are identified by hybridization with complementary DNA probes. With this method the transcripts of specific genes can be studied and measured. Only the Southern blotting technique will be described here.

### *Southern blotting*

The preferred sample for this is cellular DNA that has been extracted with minimal shearing. The DNA is subsequently digested with one or more restriction endonucleases to prepare the fragments ranging from 200 base pairs to 150 kilobase pairs in size. This fragment size may change if the DNA has been altered by deletions or base changes. Thus changes in the DNA may add or delete enzyme cutting sites leading to alteration in the size of the fragments produced (RFLP). Specific probes that are distinctive for these sites can be prepared and used for identification of these altered sites. In most instances one restriction enzyme or two in combination are used to prepare the fragments of the required size range.

The digested DNA is subjected to agarose gel electrophoresis to separate the fragments. Once the DNA fragments are separated, they are visualized by staining the gel with ethidium bromide. It is important to include appropriate controls for digestion and electrophoresis. These controls serve as checks for the enzyme activity and also provide molecular weight markers for electrophoresis. Many controls can be obtained commercially.

Before the DNA fragments can be hybridized to a probe, they have to be transferred out of the gel. The transfer is accomplished through capillary action on to a nitrocellulose membrane in the presence of a buffer. This is a slow process and often takes about 14 hours. After blotting, the DNA is immobilized on the membrane by heating. It is now ready for hybridization. Before hybridization the membrane is soaked in the hybridization cocktail without the probe to block the nonspecific binding sites. The probe is then introduced. The whole process takes place inside a sealable plastic bag, from which air bubbles have been removed. Hybridization may take up to 24 hours or more.

Following the hybridization, the membrane is

washed in graded salt solutions. If a radiolabelled probe has been used, the membrane is developed by exposure to X-ray film, a process termed autoradiography. Restriction fragments that contain sequences complementary to the probe are indicated as developed portions on the film. Exposure time varies from hours to days. This period can be lengthened to increase the assay sensitivity but is limited by high background development. With nonisotope-labelled probes, colorimetric detection is achieved directly on the membrane. Colorimetric detection is more rapid than autoradiography.

Interpretation involves the visual inspection of autoradiographic bands which indicate the size and the presence of nucleic acid in the sample hybridizing to the applied probe. The lack of bands in the sample lanes indicates a lack of complementary sequences. The presence of bands at molecular weight levels different from normal samples may indicate a change in the genetic material detected by the probe. Thus deletions, point mutations and gene duplications, leading to a gain or loss of a restriction enzyme cutting site and change in DNA fragment length, can be detected and used for the diagnosis of diseases associated with gene alterations.

Over the last few years several improvements of this method have taken place. For example, commercially manufactured cassettes that maintain consistent exposure of the membrane to a small volume of the hybridization solution are now available for the hybridization step. Nitrocellulose is increasingly being replaced by a less brittle positively charged nylon membrane. Nylon membranes are reported to fix low molecular weight DNA efficiently. They can also be stripped and reprobbed through a greater number of cycles. Some manufacturers have developed alternative forms of blotting such as electrophoretic or vacuum blotting which are less time-consuming than capillary blotting. Many of the steps involved in the blotting process have been partially automated. These steps have no doubt favoured the movement of the use of nucleic acid probes from the research laboratory to clinical diagnostic laboratories. Furthermore, many of the conditions for both Southern and Northern blotting have been streamlined.

The gradual introduction of semiautomated equipment has also reduced some of the artifacts that are associated with these methods. Interpreting the blots has been made less subjective. Depending on the probe, Southern blots can be highly specific and sensitive. In many cases, they may also provide information that cannot be obtained by any other method. A major disadvantage is the time required to perform the assay as well as several dedicated steps requiring manipulation of reagents and materials.

The Southern blotting technique has been used for identifying the genes associated with genetically transmitted diseases such as Huntington’s disease, familial growth hormone deficiency, sickle cell anaemia, adult polycystic kidney disease and cystic fibrosis. It has also been used in the antenatal diagnosis of Duchenne type and Becker’s muscular dystrophy, as well as in the investigation of

malignant disease. The usefulness of this technique is limited when the exact site of the abnormal gene is unknown. Its accuracy is dependent on how close the nucleic acid sequence under investigation is to the gene under study. However, such a limitation does not apply to probes for genes already known to be responsible for the disease.

*Northern blotting* is used to study the products of gene transcription, utilizing DNA probes that hybridize with complementary RNA sequences. This is useful in studying whether the overexpression of certain proteins in neoplastic cells is due to gene amplification or is due to post-transcriptional mechanisms such as reduced cellular excretion.

*Western blotting* allows particular proteins to be identified with specific antibodies used as analytical probes. The mixture containing the protein under investigation is separated by

**TABLE 1**

INFECTIOUS ORGANISMS THAT HAVE BEEN DETECTED BY DNA PROBES

| <b>Viruses</b>                   | <b>Bacteria</b>             | <b>Protozoa</b> |
|----------------------------------|-----------------------------|-----------------|
| Hepatitis A and B                | <i>Shigella</i>             | Trypanosomes    |
| Cytomegalovirus                  | <i>Salmonella</i>           | Schistosomes    |
| Epstein-Barr virus               | <i>Legionella</i>           | Plasmodia       |
| Adenoviruses                     | <i>Campylobacter</i>        |                 |
| Measles                          | <i>Bordetella pertussis</i> |                 |
| Enteroviruses                    | <i>Escherichia coli</i>     |                 |
| Rotavirus                        |                             |                 |
| HIV                              |                             |                 |
| HTLV-I                           |                             |                 |
| Rubella                          |                             |                 |
| Rhinovirus                       |                             |                 |
| Herpes simplex                   |                             |                 |
| Mumps                            |                             |                 |
| HIV human immunodeficiency virus |                             |                 |
| HTLV human T-lymphotropic virus  |                             |                 |

polyacrylamide gel electrophoresis and then transferred to an appropriate medium (e.g. nitrocellulose paper) through a blotting process. The protein under investigation is then identified by its appropriate antibody (the primary antibody) and the antigen-antibody complex is further identified by means of a labelled antibody against the primary antibody. The label can be either isotopic or nonisotopic in nature. Western blotting has been increasingly used for research into the acquired immune deficiency syndrome (AIDS). Serum samples from patients are screened for antibodies against a variety of viral antigens. Due to its specificity, Western blotting has been used to confirm the positive results of enzyme-linked immunosorbent assay (ELISA) tests for the AIDS virus (HIV). This technique is also used in other viral diseases such as hepatitis B (Table 1). In the areas of enzymology and oncology, Western blotting is used to identify different isoenzymes and their pattern of distribution in diseased and healthy tissue. Similarly, tumour markers within the neoplastic cells can be easily identified and localized using appropriate antisera.

#### *Nucleic acid hybridization*

Hybridization between nucleic acids is possible because of the biochemical property of complementary base pairing which allows fragments of known sequences to find matching sequences in an unknown sample. The known sequence, or probe, hybridizes with its complementary sequence and signals the presence of a particular marker in the sample. Central to the hybridization assay is the availability of the probe and sample preparation (extraction of DNA or RNA). Controls are also required for validation of the hybridization assay. A positive sample control, i.e., a known sample containing sequences complementary to the probe, is needed to ensure that the probe is indeed hybridizing under the conditions of the assay. Similarly a negative sample control, in which the sample does not contain sequences complementary to the probe, is also needed to ensure the specificity of the probe for its intended target.

Several types of hybridization assay have been described. All hybridization assays are conceptually similar. In liquid or solution phase hybridization, the sample and the nucleic acid probes interact in solution. Dot blot

hybridization involves the immobilization of sample DNA on a solid matrix. Multiple samples can be applied on to a single matrix. Problems may arise from a weak signal, which may either indicate a small amount of the specific target or large amounts of weakly cross-reacting related sequences which compromise the sensitivity and specificity of the assay. Sandwich hybridization is a modified dot blot hybridization which overcomes the problem of the nonspecific signal generated with crude samples by using two pure nucleic acid probes. One probe (the filter) is immobilized on the membrane and the other labelled probe is adjacent to it. Only samples that interact with these two probes forming a sandwich generate a signal.

The in situ hybridization (ISH) technique, which is based on the principle of nucleic acid hybridization, and the polymerase chain reaction (PCR) are two important and widely used procedures in molecular biology. These form the theme of the third and concluding article of this series on the basics of molecular biology and its medical applications.

#### REFERENCES

- 1 **Erasmus RT, Murthy DP, Ogunbanjo BO.** Basis of molecular biology and its applications: I. Molecular biology in medicine: basic concepts. *PNG Med J* 1996;39:56-66.
- 2 **Serjeantson SW.** HLA and disease in Oceania. *PNG Med J* 1989;32:241-249.
- 3 **Gao X, Bhatia K, Trent RJ, Serjeantson S.** HLA-DR, DQ nucleotide sequence polymorphisms in 5 Melanesian populations. *Tissue Antigens* 1992;40:31-37.
- 4 **Qari SH, Collins WE, Lobel HO, Taylor F, Lal AA.** A study of polymorphism in the circumsporozoite protein of human malaria parasites. *Am J Trop Med Hyg* 1994;50:45-51.
- 5 **Qari SH, Shi YP, Povo MM, Alpers MP, Deloron P, Murphy GS, Harjosuwarno S, Lal AA.** Global occurrence of *Plasmodium vivax*-like human malaria parasite. *J Infect Dis* 1993;168:1485-1489.
- 6 **Lander ES.** DNA fingerprinting on trial. *Nature* 1989;339:501-505.
- 7 **Gatti RA, Nakamura Y, Nussmeier M, Susi E, Shan W, Grody WW.** Informativeness of VNTR genetic markers for detecting chimerism after bone marrow transplantation. *Dis Markers* 1989;7:105-112.
- 8 **Deka R, De Croo S, Yu LM, Ferrell RE.** Variable number of tandem repeat (VNTR) polymorphism at locus D17S5 (YNZ22) in four ethnically defined human populations. *Hum Genet* 1992;90:86-90.